

# Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex

Matthew B Johnson<sup>1-3,9</sup>, Peter P Wang<sup>1-3,9</sup>, Kutay D Atabay<sup>1-3</sup>, Elisabeth A Murphy<sup>1-3</sup>, Ryan N Doan<sup>1-3</sup>, Jonathan L Hecht<sup>4,5</sup> & Christopher A Walsh<sup>1-3,6-8</sup>

The human cerebral cortex depends for its normal development and size on a precisely controlled balance between self-renewal and differentiation of diverse neural progenitor cells. Specialized progenitors that are common in humans but virtually absent in rodents, called outer radial glia (ORG), have been suggested to be crucial to the evolutionary expansion of the human cortex. We combined progenitor subtype-specific sorting with transcriptome-wide RNA sequencing to identify genes enriched in human ORG, which included targets of the transcription factor neurogenin and previously uncharacterized, evolutionarily dynamic long noncoding RNAs. Activating the neurogenin pathway in ferret progenitors promoted delamination and outward migration. Finally, single-cell transcriptional profiling in human, ferret and mouse revealed more cells coexpressing proneural neurogenin targets in human than in other species, suggesting greater neuronal lineage commitment and differentiation of self-renewing progenitors. Thus, we find that the abundance of human ORG is paralleled by increased transcriptional heterogeneity of cortical progenitors.

The neurons of the cerebral cortex are generated from several diverse types of neural progenitor cells whose molecular controls and lineage relationships are still not well understood. In the embryonic mouse brain, two primary progenitor subtypes that produce the excitatory projection neurons of the neocortex are clearly distinguishable by their germinal zone location, morphology, gene expression and lineage potential<sup>1</sup>. Radial glial cells (RGC), a progenitor subtype shared by all mammals, are highly polarized, epithelial-like progenitors whose cell bodies reside in the ventricular zone (VZ) and possess apical processes integrated into the ventricular surface while a radial process extends basally to the pial basement membrane. RGC are multipotent and self-renewing, undergoing mitosis at the ventricular surface and sequentially producing excitatory neurons of all cortical layers, as well as glial lineages<sup>2</sup>. RGC primarily produce neurons indirectly, generating intermediate progenitors (IP) that, in contrast to RGC, are multipolar, nonepithelial cells, basally located in the subventricular zone (SVZ), with limited capacity for self-renewal and restricted in fate to producing neurons<sup>3</sup>.

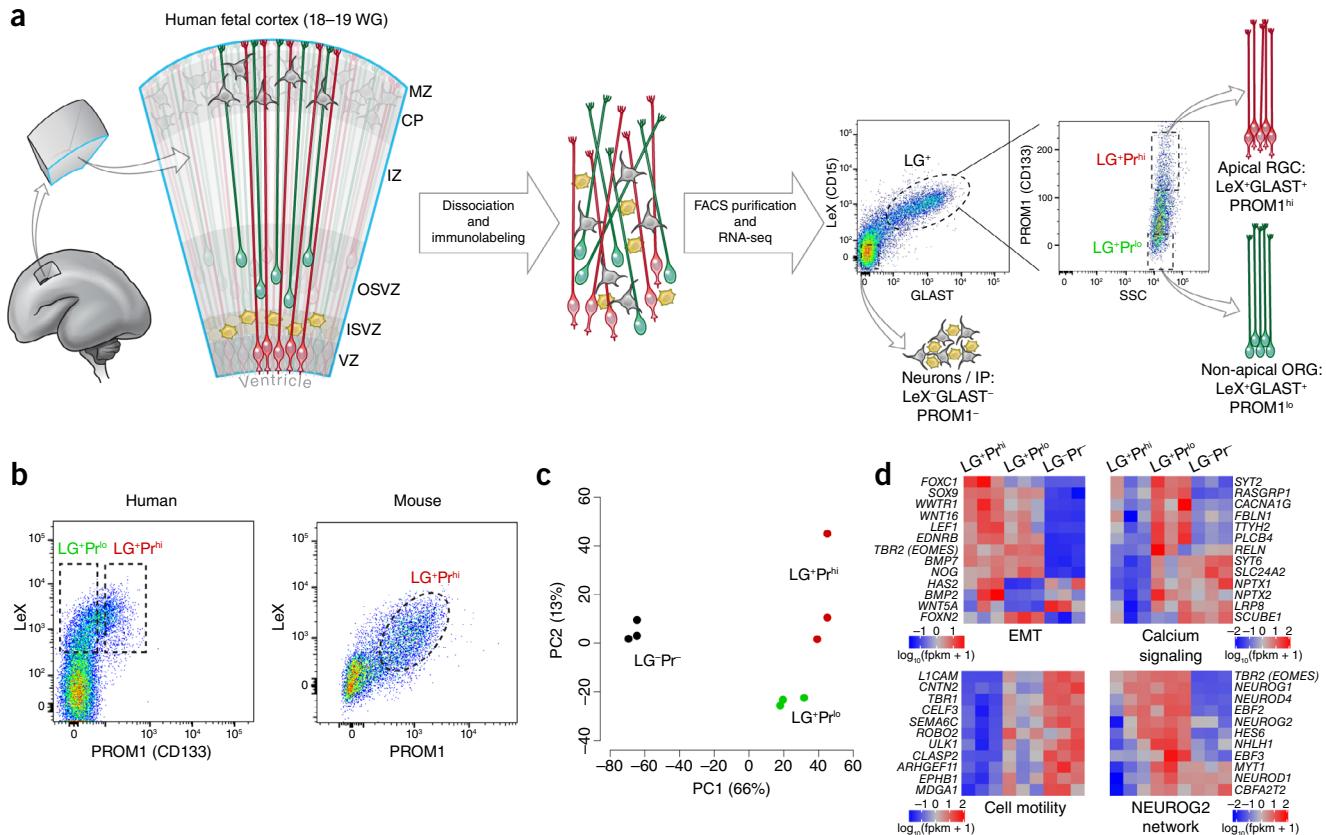
In contrast to those of rodents, primate cortical germinal zones, especially those of the human, are more complex, as exemplified by the dramatic expansion and subdivision of the SVZ into inner and outer compartments containing heterogeneous populations of progenitors with diverse morphological and molecular characteristics<sup>4-7</sup>. Most notably, basal or outer radial glial cells (ORG), which are abundant in the human fetal cortex but rare in the mouse, display characteristics of

both RGC and IP: although delaminated and basally located in the SVZ, ORG retain a radial process that frequently reaches the pial surface, and express many of the transcription factors and cytoskeletal markers of apical RGC<sup>5,6</sup>. Despite undergoing mitosis in the SVZ, ORG appear less restricted than multipolar IP in their self-renewal capacity and lineage potential: they can divide symmetrically to produce two daughter ORG<sup>8</sup> and give rise to both neurons and astrocytes<sup>9</sup>. Finally, a subset of ORG coexpresses TBR2 (refs. 6,9,10), the proneural transcription factor associated predominantly with IP in rodents. Unfortunately, although these findings highlight the need for a more detailed molecular characterization of ORG and other human progenitor subtypes, the paucity of ORG in the mouse has presented a barrier to better understanding their molecular and cellular identity.

Recent applications of high-throughput technologies have produced extensive transcriptome-wide atlases of gene expression in the human fetal brain, providing valuable insights into the evolution of human cortical neurogenesis and patterning<sup>11-15</sup>. Surprisingly, however, these studies so far have not uncovered a distinctive transcriptional signature of the expanded human outer SVZ or of the ORG that reside there. The remarkable cellular heterogeneity of the human germinal zones may obscure such a signal, since comparisons of bulk tissue samples collected by microdissection are limited to producing an average gene expression profile of the many cell types present in the samples. The SVZ in particular harbors various subtypes of radial and non-radial

<sup>1</sup>Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA. <sup>2</sup>Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, Massachusetts, USA. <sup>3</sup>Howard Hughes Medical Institute, Boston Children's Hospital, Boston, Massachusetts, USA. <sup>4</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA. <sup>5</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. <sup>6</sup>Department of Pediatrics, Harvard Medical School, Boston, Massachusetts, USA. <sup>7</sup>Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA. <sup>8</sup>Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>9</sup>These authors contributed equally to this work. Correspondence should be addressed to C.A.W. (christopher.walsh@childrens.harvard.edu).

Received 27 October 2014; accepted 24 February 2015; published online 3 March 2015; doi:10.1038/nn.3980



**Figure 1** Transcriptional profiling of isolated human radial glial cells distinguishes apical from non-apical subpopulations. **(a)** Workflow and strategy for FACS isolation of human RGC subpopulations by cell surface marker expression. LeX and GLAST are used as pan-RGC markers ( $LG^+$ ) while PROM1 is used to select apical ( $Pr^{hi}$ ) and non-apical ( $Pr^{lo}$ ) subpopulations from within the  $LG^+$  pool. SSC, side scatter. **(b)** Human  $LG^+$  cells are predominantly (~80%)  $LG^+Pr^{lo}$ , whereas >95% of mouse  $LG^+$  cells are  $LG^+Pr^{hi}$ , consistent with the relative abundance of ORG in humans and their paucity in mouse. **(c)** Principal component (PC) analysis of transcriptome-wide gene expression estimates, in fragments per kilobase per million reads (fpkm), across three biological replicates of FACS-separated subpopulations reveals major gene expression differences between  $LG^+$  progenitors and  $LG^-$  cells (first PC, x axis), as well as between  $LG^+Pr^{hi}$  apical and  $LG^+Pr^{lo}$  non-apical radial glial subtypes (second PC, y axis). **(d)** Differential expression between  $LG^+Pr^{hi}$  apical and  $LG^+Pr^{lo}$  non-apical RGC subpopulations included genes involved in calcium signaling, epithelial-to-mesenchymal transition (EMT) and cell migration and motility, as well as several members of a proneural transcription factor network regulated by the transcription factor NEUROG2.

progenitors, radially migrating projection neurons, and tangentially migrating interneurons originating from separate progenitor pools in the ventral telencephalic germinal zones.

In the present study, we dissected the cellular heterogeneity of the fetal human cortex by first isolating the cell populations of interest from dissociated tissue and then applying single-cell gene expression profiling. Using this approach, we found hundreds of genes that were specifically enriched in apical RGC, and a smaller but distinct transcriptional signature of human ORG. The ORG transcriptional profile was dominated by proneural transcription factors of the neurogenin pathway, indicating that, at the population level, ORG represent a distinct reservoir of neuronal lineage-committed, self-renewing radial progenitors. We overexpressed the neurogenin pathway in the developing cortex of the ferret, a carnivore with abundant ORG, and confirmed that this pathway has a conserved function in processes critical to ORG formation, including delamination from the ventricular neuroepithelium and migration into the SVZ. A comparative single-cell transcriptional analysis of human, ferret and mouse progenitors confirmed that neurogenin pathway-expressing cells were more abundant among human progenitors and less common in mice, which lack a large ORG subpopulation. More generally, single-cell profiling revealed a surprising transcriptional heterogeneity of human and, to a lesser degree, ferret cortical progenitors, which we propose reflects an extended, more

graded transcriptional transition from RGC to IP in these species, characterized by a large proportion of cells coexpressing classic markers of both self-renewing RGC and neuronal lineage-committed IP. Finally, comparative genomic analysis of several previously undescribed human long noncoding RNA (lncRNA) genes indicated that many of these loci were potentially present in the common ancestor of human, ferret and mouse but show enriched expression in human ORG accompanied by greater genomic sequence divergence in rodents.

## RESULTS

## Purification and RNA-seq analysis of human ORG

We used the differential expression of surface markers to separate cortical progenitor subtypes using fluorescence-activated cell sorting (FACS) before RNA-seq (**Fig. 1a**). Human apical RGC, the epithelial progenitor subtype, express LeX (CD15) and GLAST (SLC1A3)<sup>16–18</sup>, as well as prominin (PROM1; CD133) on their apical surfaces<sup>19,20</sup>. ORG express LeX and GLAST but lack apical proteins, including PROM1 (ref. 5). IP and neurons lack all three markers. Therefore, we separated for RNA-seq analysis LeX and GLAST double-positive (LG<sup>+</sup>) cells showing the top (LG<sup>+</sup>Pr<sup>hi</sup>) and bottom (LG<sup>+</sup>Pr<sup>lo</sup>) 5–10% of PROM1 signal intensity to enrich for apical RGC and non-apical ORG, respectively, as well as cells negative for all three markers (LG<sup>-</sup>Pr<sup>-</sup>), comprising IP and neurons, among other cells (**Fig. 1a** and **Table 1**).

**Table 1 Cell populations immunolabeled and collected by FACS from human fetal cortex for RNA-seq analysis**

Sorted pool	Antigenicity	Enriched markers (qRT-PCR)	Cell type
LG <sup>+Pr<sup>hi</sup></sup>	LeX <sup>+</sup> , Glast <sup>+</sup> , PROM1-high	PAX6, SOX2, VIM, NES, BLBP, PARD3 (PAR3), TJP1 (ZO1), MPP5 (PALS)	Apical RGC
LG <sup>+Pr<sup>lo</sup></sup>	LeX <sup>+</sup> , Glast <sup>+</sup> , PROM1-low	PAX6, SOX2, VIM, NES, BLBP	Non-apical RGC, including ORG
LG <sup>-Pr<sup>-</sup></sup>	LeX <sup>-</sup> , Glast <sup>-</sup> , PROM1 <sup>-</sup>	TUBB3 (TUJ1), DCX, MEF2C, RBFOX3 (NEUN)	IP, neurons

Dissociated fetal human cortical tissue was immunolabeled for radial glial progenitor markers LeX (CD15), GLAST (SLC1A3) and PROM1 ( prominin; CD133). LeX and Glast are detected on most RGC<sup>16–18</sup>, while PROM1 is specific to the apical membrane domain found at the ventricular surface<sup>19,20</sup>.

Several lines of evidence confirm that our sorting approach enriches for RGC while separating apical from non-apical subpopulations. First, quantitative real-time reverse transcription PCR (qRT-PCR) confirmed that both LG<sup>+Pr<sup>hi</sup></sup> and LG<sup>+Pr<sup>lo</sup></sup> cells were enriched for markers of neural progenitors and radial glia, while being depleted for neuronal genes, compared to the LG<sup>-Pr<sup>-</sup></sup> population (**Supplementary Fig. 1a**). LG<sup>+Pr<sup>hi</sup></sup> cells showed enrichment for mRNAs encoding PROM1 and other apical membrane proteins as compared to LG<sup>+Pr<sup>lo</sup></sup> cells (**Supplementary Fig. 1a**). LeX<sup>+</sup> cells proliferated *in vitro* to produce neurospheres that showed SOX2 immunoreactivity and were serially passaged at clonal density, consistent with neural stem cell behavior<sup>16</sup> (**Supplementary Fig. 1b**). Furthermore, in RGC sorted from embryonic mouse cortex, PROM1 largely overlapped with both LeX and GLAST, with few LG<sup>+Pr<sup>lo</sup></sup> cells detected, confirming the scarcity of non-apical ORG in mouse (**Fig. 1b**). Finally, since non-apical multipolar IP lack PROM1, the absence of a significant LG<sup>+Pr<sup>lo</sup></sup> population in the mouse also corroborates the absence of LeX and GLAST on IP, as these would appear LG<sup>+Pr<sup>lo</sup></sup> by FACS. Thus, our method provides an opportunity to assay transcriptome-wide differences between, as well as heterogeneity within, human progenitor subtypes.

RNA-seq of the three FACS-enriched cell populations from three biological replicates 18–19 weeks of gestation (WG) (**Supplementary Table 1**) identified ~3,500 known genes, as well as ~250 new, non-reference loci, with significantly different expression (false discovery rate <5%, fragments per kilobase per million reads fpkm >1). Principal component analysis indicated that the greatest proportion of variability between samples reflected the differences between the LG<sup>+</sup> RGC and LG<sup>-</sup> cells, but the second principal component highlighted differences between the LG<sup>+Pr<sup>hi</sup></sup> apical and LG<sup>+Pr<sup>lo</sup></sup> non-apical subpopulations, indicating a distinct ORG transcriptional signature (**Fig. 1c**). Gene set enrichment analysis further demonstrated the radial glial progenitor nature of the LG<sup>+</sup> population: relative to the LG<sup>-Pr<sup>-</sup></sup> pool, LG<sup>+</sup> cells were enriched for genes involved in cell cycle regulation, DNA replication, extracellular matrix, and growth factor pathways critical for RGC maintenance and neurogenesis (**Supplementary Fig. 2**). Notably, LG<sup>+</sup> enriched genes included integrin signaling and basement membrane components, such as laminins, consistent with both LG<sup>+Pr<sup>hi</sup></sup> and LG<sup>+Pr<sup>lo</sup></sup> subpopulations maintaining radial processes contacting the pial basement membrane, as has been shown for ORG<sup>5,6,9</sup>. In all, 552 genes significantly differed between LG<sup>+Pr<sup>hi</sup></sup> and LG<sup>+Pr<sup>lo</sup></sup> cells, with 79 of these genes specifically enriched or depleted in LG<sup>+Pr<sup>lo</sup></sup> non-apical ORG (**Fig. 1d** and **Supplementary Fig. 3**). Among the genes upregulated in LG<sup>+Pr<sup>lo</sup></sup> ORG, six transcription factors—HES6, NEUROD4 (Atoh3, Math3), NHLH1 (HEN1, NSCL1), NEUROD1, CBFA2T2 (Mtgr1) and MYT1—are all downstream of the critical regulatory gene NEUROG2, encoding neurogenin 2 (refs. 21,22), which in mouse cortex and chick spinal cord initiates delamination and neuronal lineage commitment of neural precursors<sup>23–25</sup>. Notably, NEUROG2 itself and two

more early markers of neuronal fate commitment, TBR2 (EOMES)<sup>23</sup> and BTG2 (Tis21)<sup>26</sup>, were all highly expressed in both LG<sup>+Pr<sup>hi</sup></sup> and LG<sup>+Pr<sup>lo</sup></sup> subpopulations (**Fig. 1d**). Given that NEUROG2 expression is transient in mouse apical progenitors during their transition from RGC to TBR2<sup>+</sup> IP<sup>23,27</sup>, we sought to validate the expression of NEUROG2 in apical and non-apical RGC and to test its function in a model species with abundant cortical ORG.

### NEUROG2 function in RGC of the gyrencephalic ferret cortex

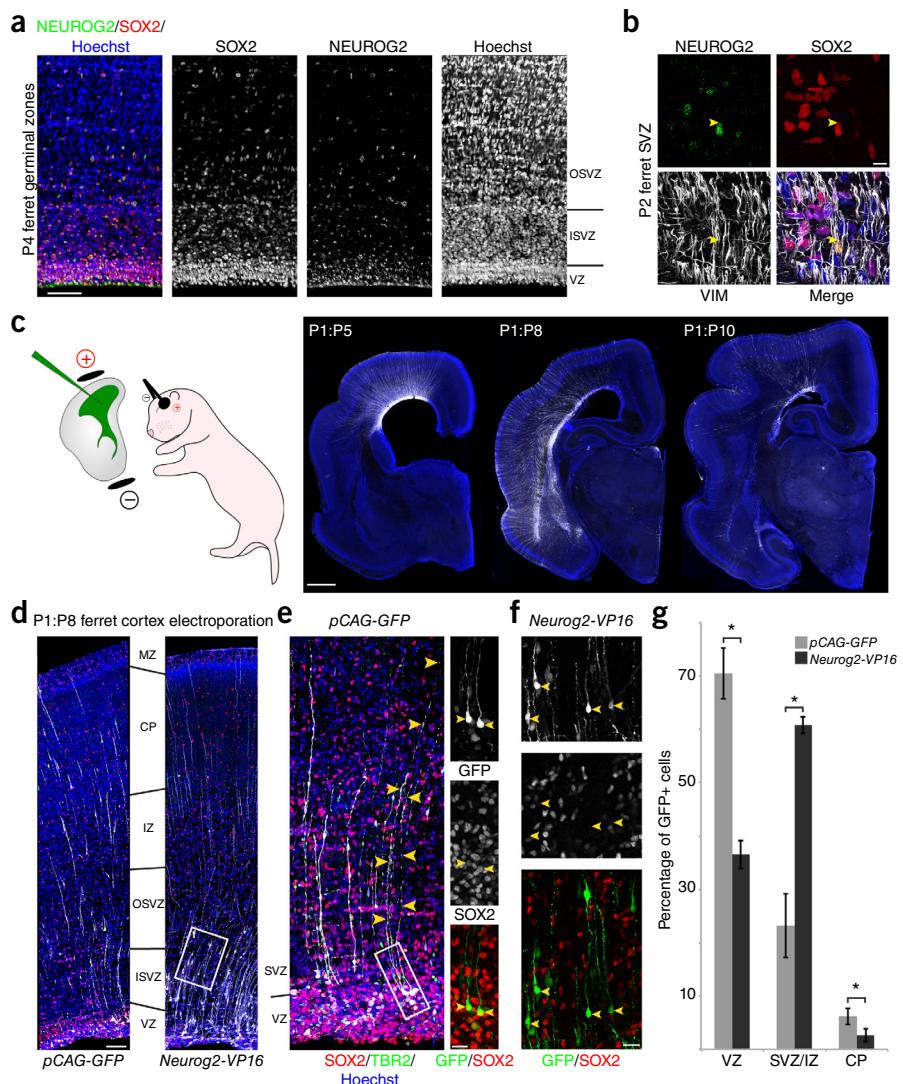
The developing ferret and human cerebral cortices share several key features, including stereotyped sulci and gyri, a dramatically expanded subventricular zone (SVZ), and an abundance of ORG, making the ferret an attractive model for the study of cortical neurogenesis<sup>9,28,29</sup>. We first confirmed the expression of NEUROG2 in ferret RGC (**Fig. 2a,b**), identifying a marked ‘salt-and-pepper’ pattern of NEUROG2 immunoreactivity in both the VZ and SVZ. We then used *in vivo* electroporation of the newborn ferret dorsal cortical VZ to express a NEUROG2-VP16 fusion protein, which links the DNA-binding domain of NEUROG2 to the VP16 constitutive transcriptional activator domain<sup>25</sup>, thus activating all direct downstream targets of NEUROG2 in ferret apical RGC. Following delivery of the *Neurog2-VP16* expression construct, we allowed ferret kits to develop for up to 10 days post-electroporation (DPE), during which time many neurons of the upper cortical layers are generated<sup>30–32</sup> (**Fig. 2c**). In both control (*pCAG-GFP*) and *Neurog2-VP16* brains, we observed numerous GFP<sup>+</sup>SOX2<sup>+</sup> and GFP<sup>+</sup>TBR2<sup>+</sup> cells in the VZ, including cells in the basal VZ and inner SVZ with a characteristic ORG morphology (**Fig. 2d,e**), as well as many GFP<sup>+</sup> cells in the SVZ and intermediate zone (IZ), with a small number reaching the cortical plate (CP) after longer survival times (up to 10 DPE) (**Fig. 2f**). At 7–9 DPE, NEUROG2-VP16 induced a significant shift in the proportion of GFP<sup>+</sup> cells from the VZ to the SVZ/IZ and a concomitant reduction in the proportion of GFP<sup>+</sup> cells coexpressing SOX2 (**Fig. 2g**), with most NEUROG2-VP16<sup>+</sup> cells displaying the morphology of radially migrating postmitotic neurons in the outer SVZ and IZ.

In addition, we FACS-purified electroporated cells from the ferret cortex and performed qRT-PCR analysis of ORG-enriched candidate genes identified in humans. We found that nearly all human ORG-enriched NEUROG2 downstream targets were highly upregulated by the NEUROG2-VP16 construct in ferret, relative to their expression in control GFP-expressing cells, while Sox2 was repressed (**Supplementary Fig. 4**). Notably, NEUROG2-VP16 expression in ferret RGC *in vivo* also resulted in increased expression of the ferret orthologs of several other human ORG-enriched genes, including *Gadd45g* and *Ttyh2* (**Supplementary Fig. 4**), further suggesting that NEUROG2 is a critical regulator of a conserved radial progenitor development program in species with abundant ORG. Taken together, our ferret functional experiments demonstrate a conserved role for NEUROG2 transcriptional targets in driving delamination from the ventricular neuroepithelium, which is a key step in the production of ORG, while additional downstream effectors initiate repression of Sox2 and activation of a neuronal differentiation program, including radial migration, as previously described in mice. Future studies will be required to identify the specific factors downstream of NEUROG2 that regulate neuroepithelial integration and elucidate the molecular mechanisms that permit a subset of NEUROG2-expressing ferret and human RGC to remain integrated in the VZ while others detach and migrate into the SVZ.

### Single-cell analysis of species-specific RGC heterogeneity

Both our human RNA-seq and ferret immunofluorescence data demonstrate NEUROG2<sup>+</sup> RGC subpopulations in both the VZ and SVZ, intermingled with NEUROG2<sup>-</sup> progenitors, exemplifying the heterogeneity

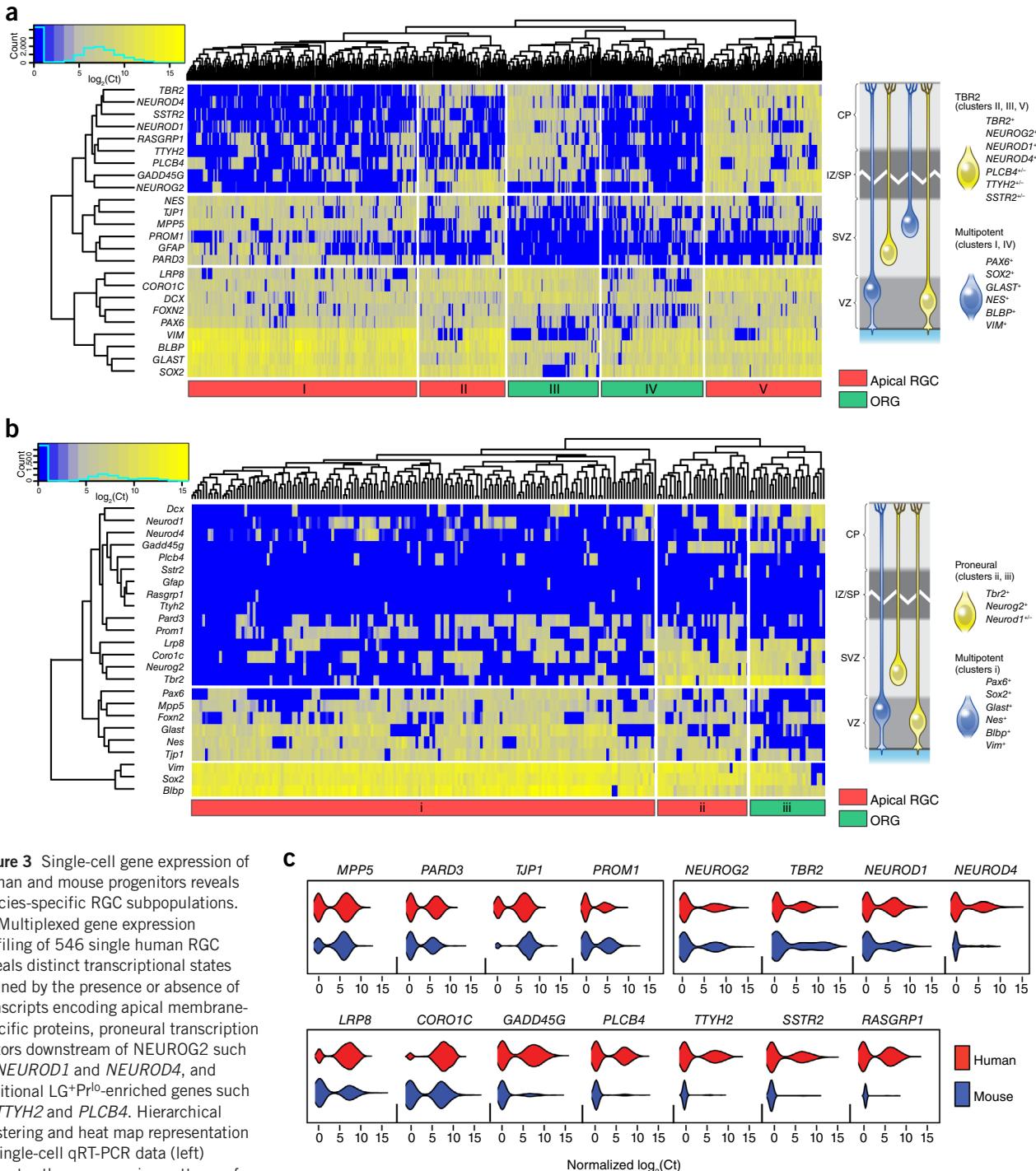
**Figure 2** NEUROG2 regulates progenitor morphology and molecular identity in the developing cortex of the gyrencephalic ferret. **(a)** Coexpression of RGC marker SOX2 and proneural marker NEUROG2 in developing ferret cortex. Numerous SOX2<sup>+</sup>NEUROG2<sup>+</sup> cells are found in both the VZ and SVZ at early postnatal ages (shown here, postnatal day (P) 4), when these germinal zones are populated respectively by many apical and non-apical RGC-producing neurons destined for the upper cortical layers. **(b)** Higher magnification of SOX2<sup>+</sup>NEUROG2<sup>+</sup> progenitors at P2 in the SVZ coexpressing the RGC neurofilament protein vimentin (VIM), which labels the basal radial process (yellow arrowheads). **(c)** Genetic manipulation of apical RGC is achieved by *in vivo* intraventricular injection and electroporation in neonatal ferret kits. In the cortex of kits electroporated at P1 and examined at P5 (P1:P5), P8 (P1:P8) or P10 (P1:P10), GFP<sup>+</sup> cells are observed throughout the developing cortical wall, with GFP<sup>+</sup> radial fibers extending from the germinal zones to the pial surface and newborn neurons migrating through the SVZ and IZ into the cortical plate (CP). **(d)** Littermates electroporated with NEUROG2 gain-of-function (*pCAG-Neurog2-VP16*) or GFP control (*pCAG-GFP*) expression constructs, analyzed at 7 DPE (P1:P8) for the distribution of GFP<sup>+</sup> cells and their coexpression of SOX2. ISVZ, inner SVZ; OSVZ, outer SVZ; MZ, marginal zone. **(e)** At all survival timepoints, we identified numerous GFP<sup>+</sup>SOX2<sup>+</sup> apical RGC in the VZ as well as occasional GFP<sup>+</sup>SOX2<sup>+</sup> ORG with somata at the VZ/SVZ border (insets, magnified from boxed area) and radial fibers extending through the SVZ/IZ towards the pial surface (yellow arrowheads). **(f)** Higher magnification of the boxed area from **d** shows GFP<sup>+</sup> cells in the SVZ/IZ that are SOX2<sup>-</sup> with the morphology of radially migrating newborn neurons. **(g)** At 7–9 DPE, NEUROG2 gain of function induced a significant shift of GFP<sup>+</sup> cells from the VZ into the SVZ/IZ compared to the control (\* $P < 0.05$ , paired *t*-test; exact *P* values: VZ, 0.026; SVZ/IZ, 0.026; CP, 0.034;  $n = 3$  animals per condition, 3 or 4 brain sections counted per animal; data represented as mean  $\pm$  s.e.m.), with a concomitant loss of RGC morphology and SOX2 expression, demonstrating that the NEUROG2 proneural network promotes delamination of daughter cells from the ventricular surface, migration into the SVZ and neuronal differentiation. Scale bars: 50  $\mu$ m (**a**), 10  $\mu$ m (**b**), 1 mm (**c**), 100  $\mu$ m (**d**), 20  $\mu$ m (**e,f**).



that confounds population-level transcriptome comparisons, so we turned to single-cell analysis to compare the subpopulations of radial progenitors in human, ferret, and mouse. We first sorted RGC from human fetal cortex ( $n = 6$ , 16–21 WG; **Supplementary Table 1**) into 96-well plates and performed microfluidics-based, highly multiplexed, single-cell qRT-PCR to simultaneously assay several dozen genes, which included markers for all RGC (*PAX6*, *SOX2*, *GLAST*, *BLBP* (*FABP7*), *VIM*, *NES*) and apical RGC (*PROM1*, *PARD3*, *MPP5*, *TJPI*); proneural neurogenin-pathway transcription factors; and additional validated LG<sup>+</sup>Pr<sup>lo</sup>, ORG-enriched genes (**Supplementary Fig. 3**). Among 546 sorted single human progenitors, *PAX6*, *SOX2*, *GLAST*, *BLBP* and *VIM* were detected in  $93 \pm 3\%$  of cells (**Supplementary Fig. 5a**), confirming their radial glial identity. Hierarchical clustering revealed several distinct transcriptional states characterized by the combinatorial expression of apical markers and proneural factors, such that cells fell into one of four main subpopulations. We refer to these as apical/multipotent, apical/proneural, non-apical/multipotent and non-apical/proneural (**Fig. 3a** and **Supplementary Fig. 5b**). Apical RGC subpopulations (clusters I, II, V in **Fig. 3a**;  $71 \pm 8\%$  detection rate for apical genes) could

be divided into apical/multipotent (I) and apical/proneural (II and V) subpopulations on the basis of their lesser ( $21 \pm 5\%$ ) or greater ( $81 \pm 7\%$ ) expression of proneural neurogenin-pathway transcription factors, respectively. ORG-like non-apical subpopulations (clusters III and IV) with lower detection rates of apical complex transcripts ( $31 \pm 6\%$ ) were similarly subdivided according to lower or higher rates of proneural gene expression ( $22 \pm 7\%$  versus  $65 \pm 15\%$ ). Notably, most non-apical/proneural cells were NEUROG2<sup>-</sup> (cluster III), consistent with observations in the mouse that NEUROG2 represses apical identity and is then downregulated upon delamination<sup>23</sup>. Finally, both apical and non-apical proneural RGC were further subdivided by expression of other LG<sup>+</sup>Pr<sup>lo</sup>-enriched genes (*TTYH2*, *PLCB4*, *SSTR2*, *RASGRP1*) that define additional transcriptional heterogeneity among human cortical progenitors. These results demonstrate significant multigenic transcriptional diversity within cortical radial glial progenitors and are characteristic of the previously unappreciated heterogeneity recently revealed by single-cell analyses in other, non-neuronal stem cell niches<sup>33,34</sup>.

In contrast to the human analysis, single RGC from the embryonic day (E) 16–17 mouse cortex showed fewer distinct transcriptional states



**Figure 3** Single-cell gene expression of human and mouse progenitors reveals species-specific RGC subpopulations. (a) Multiplexed gene expression profiling of 546 single human RGC reveals distinct transcriptional states defined by the presence or absence of transcripts encoding apical membrane-specific proteins, proneural transcription factors downstream of NEUROG2 such as *NEUROD1* and *NEUROD4*, and additional LG<sup>+</sup>Pr<sup>lo</sup>-enriched genes such as *TTYH2* and *PLCB4*. Hierarchical clustering and heat map representation of single-cell qRT-PCR data (left) indicates the coexpression patterns of

these genes. At right, a schematic representation of the four main subpopulations of RGC identified. Multipotent RGC (blue) are found as subsets of both the apical (cluster I) and non-apical RGC (cluster IV), as are the proneural *NEUROG2*<sup>+</sup>*TBR2*<sup>+</sup> RGC (clusters II, III, V). In addition, proneural RGC can be further subdivided according to their expression of downstream factors and additional LG<sup>+</sup>Pr<sup>lo</sup>-enriched genes (for example, compare clusters II and V). Ct, cycle threshold. (b) The same genes assayed in 226 RGC from E16–E17 mouse cortex yield only three subpopulations: apical multipotent (i), apical proneural (ii) and non-apical proneural (iii). Schematic representation of these subpopulations (right) highlights the main species differences: the mouse has fewer non-apical cells overall; few, if any, multipotent (*NEUROG2*<sup>+</sup>*TBR2*<sup>+</sup>) non-apical cells, suggesting the absence of a significant subpopulation of proliferative ORG; and very few cells expressing other human subset-enriched genes (for example, *TTYH1*, *PLCB4*). (c) Violin plots of gene expression distributions, with number of single cells on the y axis, for apical complex, NEUROG2 network and ORG-enriched genes in human and mouse single RGC. Several ORG-enriched genes appear to be abundantly expressed in subsets of human but not mouse RGC, including *NEUROD4*, *GADD45G*, *PLCB4*, *TTYH2*, *SSTR2*, *RASGRP1*.

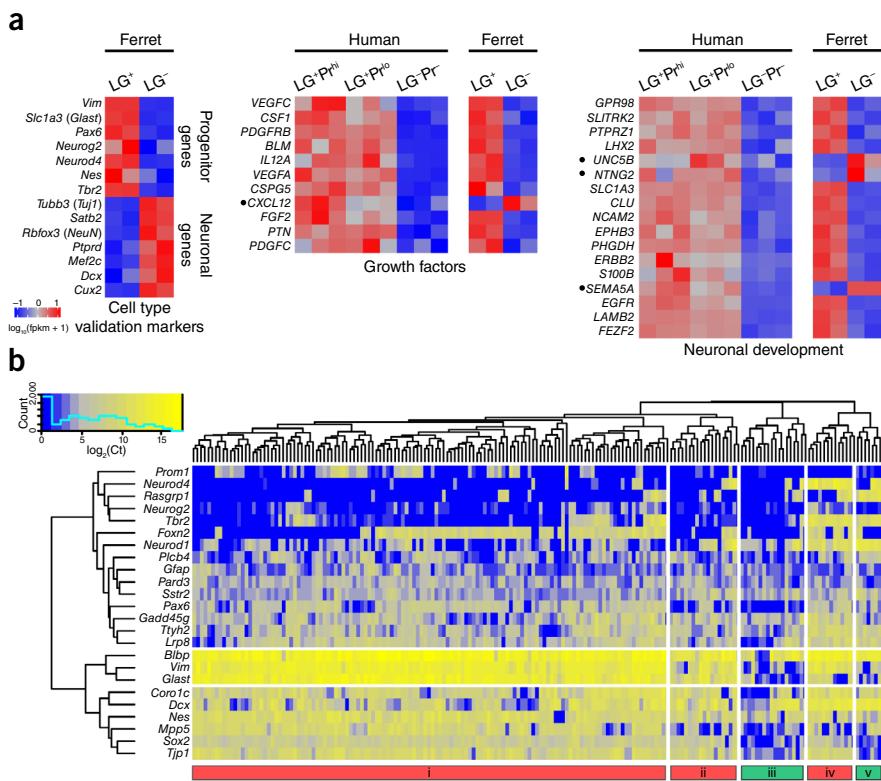
(Fig. 3b and Supplementary Fig. 5a) and rarely expressed many of the genes that defined subsets of human cells (Fig. 3b,c). Among 226 mouse single cells, 93 ± 8% expressed the RGC markers *Sox2*, *Vim*, *Blbp* and *Glast*, and most (88%) of cells also expressed some apical complex

genes, confirming that ORG are rare in the mouse. Hierarchical clustering yielded only three subpopulations (Fig. 3b), corresponding to the apical/multipotent (cluster i), apical/proneural (ii) and non-apical/proneural (iii) subsets observed in the human cortex. Although a

**Figure 4** Population-level whole-transcriptome RNA-seq and single-cell expression analysis of ferret RGC. **(a)** Expression heat maps of known progenitor and neuronal marker genes, as well as selected human RGC-enriched gene sets, from LG<sup>+</sup> and LG<sup>-</sup> cells isolated by FACS from the P2 developing ferret cortex ( $n = 2$ ). Enrichment of classic RGC markers and a high degree of similarity between gene sets enriched in human and ferret LG<sup>+</sup> cells validate the use of LeX and Glast to select RGC from the developing ferret cortex. Notably, however, several genes (CXCL12, UNC5B, NTNG2, SEMA5A) show distinct expression patterns (black bullets) between the two species, suggesting that certain growth factor and other pathways may be expressed in a species-specific manner in RGC. **(b)** Single-cell gene expression profiling of 185 single ferret LG<sup>+</sup> progenitors was performed using the same gene panel as shown in **Figure 3** for human and mouse RGC. As in humans, a substantial fraction of ferret cells in clusters i, iv and v coexpress both RGC markers and are *Tbr2*<sup>+</sup>/*Neurog2*<sup>+</sup>, consistent with our immunohistochemical analysis of NEUROG2 expression in the ferret (**Fig. 2a**) and suggesting this proneural RGC transcriptional state is conserved. As in human RGC, a subset of these proneural cells also express the downstream factors NEUROD1 and NEUROD4. However, the orthologs of some human ORG-enriched genes (for example, *Rasgrp1*) are expressed in fewer ferret RGC, while others (for example, *Plcb4*, *Sstr2*, *Gadd45g*, *Ttyh2*) appear more homogenous across all cells. *Foxn2*, which was detected in nearly all human RGC, appears to mark a distinct subpopulation of ferret apical RGC (clusters i and iv). Overall, while ferret RGC exhibit more diversity of transcriptional states than mouse and generally more similarity to human, they are nonetheless distinct in their relative proportions and composition.

substantial subset of mouse RGC coexpressed proneural transcription factors, the proportion of cells was significantly smaller than in human (27% in mouse versus 47% in human;  $P = 9.57 \times 10^{-8}$ , Fisher's exact test). Importantly, the absence of an appreciable non-apical/multipotent subpopulation (human cluster IV) suggests a critical species difference in the proliferative potential of ORG, which could underlie the paucity of ORG in the mouse. Most notably, orthologs of human ORG-enriched genes that contributed markedly to human RGC heterogeneity, including *Plcb4*, *Gadd45g*, *Ttyh2*, *Rasgrp1* and *Sstr2*, were detected in a rare and uncorrelated minority of mouse RGC (<10%, compared to 45–55% in human) (**Fig. 3b,c**), further highlighting the species specificity of RGC transcriptional heterogeneity.

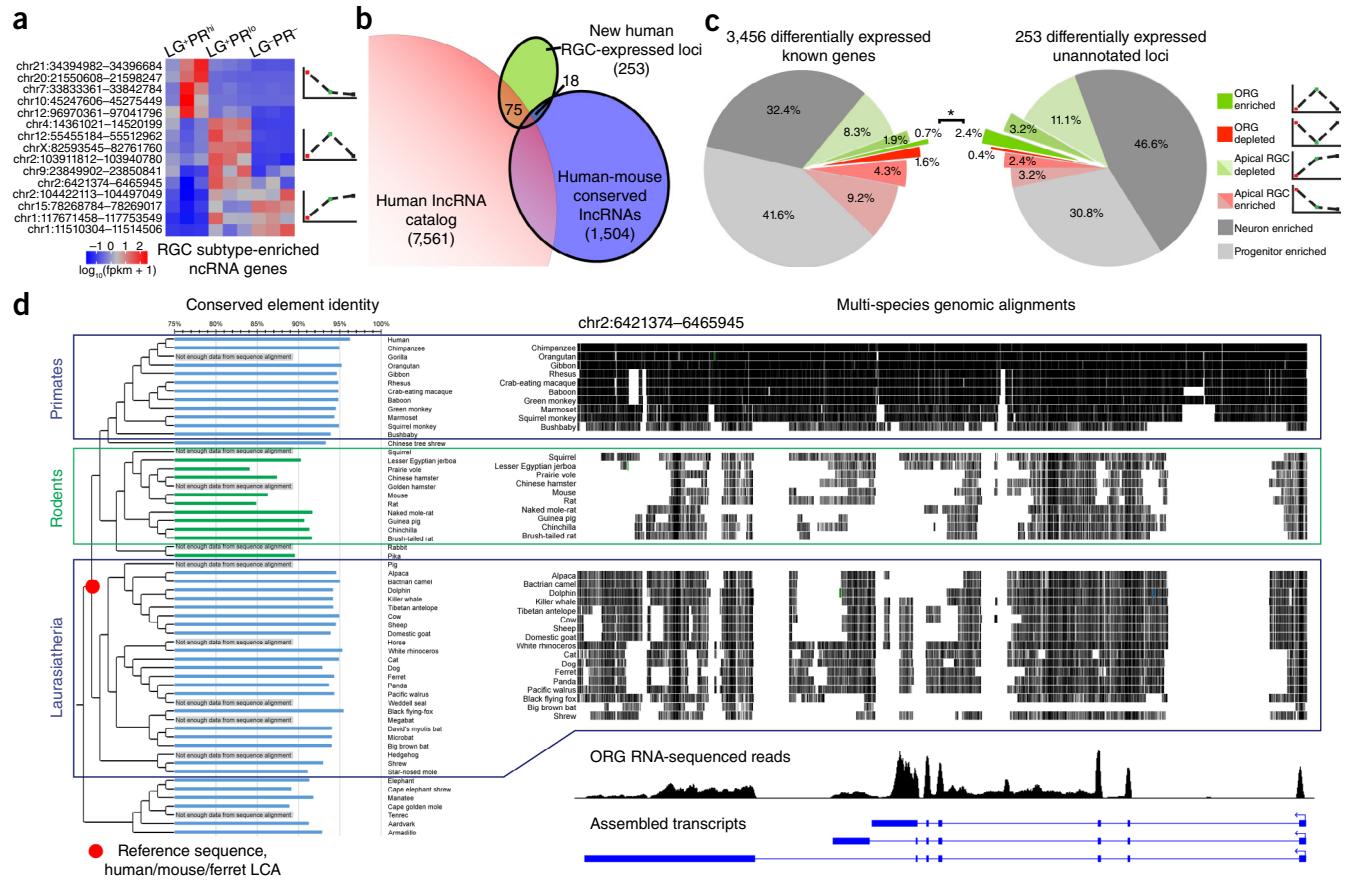
Finally, we performed RNA-seq and single-cell profiling of ferret radial glial progenitors and found that they shared some of the key transcriptional states of human RGC. In the absence of working antibodies against ferret prominin, we first validated LeX and Glast antibodies by immunohistochemistry in ferret brain sections, as well as by FACS (data not shown), then collected LG<sup>+</sup> and LG<sup>-</sup> cells from neonatal ferret cortex, at which time middle and upper layers of cortex are being generated<sup>30,31</sup>, roughly corresponding to mouse E16–17 or human 16–20 WG, and performed population-level RNA-seq. Ferret LG<sup>+</sup> cells were enriched for most previously described RGC marker genes and showed transcriptome-wide expression patterns similar to those of LG<sup>+</sup> cells from human and mouse cortex (**Fig. 4a**). Having verified that LG<sup>+</sup> cells comprise a substantial proportion of ferret RGC, we performed single-cell profiling on 185 single LG<sup>+</sup> cells from the neonatal ferret cortex. We found a degree of heterogeneity intermediate between that of the human and mouse progenitors (**Fig. 4b**). Most sorted ferret cells expressed classic RGC markers, confirming the specificity of the sorting, while a subset coexpressed *Tbr2* (ferret clusters iv and v, and a number of cells in cluster i), and a



smaller subset of those were positive for *Neurod4* and *Neurod1* (clusters iv and v). The ferret homologs of some human ORG-enriched genes (for example, *Rasgrp1*) were preferentially expressed in these proneural cells, as in human, whereas the homologs of most human ORG genes (such as *Ttyh2*, *Sstr2*) were more homogeneously expressed in most ferret cells and, conversely, some genes (such as *Foxn2*) were heterogeneously expressed in subsets of cells that did not correspond to those seen in human. Finally, we noted that a greater proportion of ferret cells than human cells expressed the gliogenic marker *Gfap*, which is consistent with evidence that ferret ORG have astrogliogenic potential<sup>9</sup>. Taken together, our single-cell analyses from human, ferret and mouse implicate many genes acting in a coordinated network that may be responsible for the evolution of novel progenitor transcriptional states critical for human cortical development.

### Long noncoding transcripts enriched in human ORG

Given the species differences in RGC subpopulations revealed by our single-cell analysis, we next searched our RNA-seq data for transcriptional influences on species differences in RGC molecular identity. We identified candidate nonconserved RNA transcripts that included lncRNAs, which are evolutionarily dynamic, frequently lack human-mouse homology<sup>35</sup> and are involved in critical neural developmental processes such as progenitor pluripotency, neurogenesis and epithelial–mesenchymal transition<sup>36–38</sup>. We compared 253 differentially expressed unannotated loci (**Fig. 5a**) to two published human lncRNA catalogs<sup>39,40</sup> and identified 75 loci overlapping putative human lncRNAs<sup>39</sup>, while only 18 loci matched reported human-mouse conserved lncRNAs<sup>40</sup> (**Fig. 5b** and **Table 2**), suggesting that the human RGC subtype-specific transcripts we have identified include many lncRNAs that lack any homologous mouse transcripts. Unexpectedly, we found that a much greater proportion of the loci, as compared to



**Figure 5** Transcripts detected by RNA-seq include previously unknown lncRNAs with distinct RGC subtype expression patterns and evolutionary conservation. (a) Differential expression patterns of selected noncoding RNA loci, including several new multixon lncRNAs (Table 3). Schematics at right represent differential expression patterns of distinct groups of transcripts. (b) Intersection of 253 differentially expressed non-reference loci from our RNA-seq analysis with previous catalogs of human noncoding RNA genes revealed a number of reported human lncRNAs<sup>39</sup> and a smaller number of human-mouse conserved lncRNAs<sup>40</sup>. (c) A significantly greater proportion of the differentially expressed transcripts were specifically enriched in LG<sup>+</sup>Pr<sup>lo</sup> non-apical progenitors (2.4%), as compared to known genes (0.7%; \*P = 0.012, Fisher's exact test), implicating this evolutionarily dynamic gene class in the regulation of the ORG progenitor subpopulation, which is greatly expanded in humans. (d) Comparative genomics analysis of our ORG-enriched lncRNAs was performed by comparing conserved elements from within each genomic locus from 58 species to a computed ancestral sequence for the laurusatherian LCA of human, ferret and mouse (see Table 3). Here we show an example from a human ORG-enriched lncRNA gene located on chromosome 2. Left, percentage identity of each species' conserved elements from this locus to the LCA sequence; rodents (green) show a highly divergent sequence compared to both primates and other more distantly related groups, including carnivores. Right, multispecies genomic alignments to the same human locus for primates, rodents and other laurusatherian species (top), with the human ORG RNA-seq reads (middle) and assembled transcripts (bottom), illustrating the greater sequence divergence of rodents compared to either nonhuman primates or other more distant laurusatherian species. A list of species names appears in Supplementary Table 3.

known genes, were specifically enriched in ORG (Fig. 5c; 2.4% versus 0.7%; P = 0.012, Fisher's exact test), suggesting that lncRNAs are especially relevant to the molecular identity and function of the ORG subpopulation in humans. By manual inspection, we determined that although a few of the transcripts reflected incomplete annotations of known genes (for example, alternative transcription start sites or untranslated regions), most resembled bona fide unannotated genes, many of which showed multiple exons and alternative splicing (Table 3 and Supplementary Fig. 6) and none of which have, to our knowledge, previously been reported in cortical development.

Few known lncRNAs that are functionally essential<sup>37</sup> or have been transcriptionally profiled<sup>41</sup> in mouse brain development were detected in human, and those that were conserved displayed species-specific expression patterns, further illustrating the dynamic evolutionary changes in lncRNAs. Of 18 lncRNAs recently knocked out in mice<sup>37</sup>, we identified only two orthologs with appreciable expression in human developing cortex (Table 2). Similarly, of 15 mouse IP-enriched lncRNAs<sup>41</sup>, only two, MIAT and RMST,

showed appreciable expression in human fetal cortex, and even these showed cell-type enrichment patterns distinct from those in mouse (Table 2 and Supplementary Fig. 7). Within lncRNAs conserved between human and mouse, we found several, including *LINC-PINT*, *TUNAR*, *CRNDE* and *MIR22HG*, that were depleted in mouse RGC but enriched in human apical and outer RGC, suggesting potentially distinct functions in cortical development (Supplementary Fig. 7). Thus, the dynamic patterns of lncRNA expression in RGC subtypes and their notable lack of conservation are consistent with the highly species- and also cell type-specific expression of lncRNAs in other contexts and suggest that this transcript class is unusually dynamic in its evolutionary relationship to cortical development.

To probe the evolutionary history of ORG-enriched lncRNAs, we performed comparative genomic analysis, specifically evaluating their presence in a common mammalian ancestor and their conservation in gyrencephalic mammals such as the ferret and nonhuman primates compared to rodents. We extracted conserved elements within the newly identified lncRNA genomic loci and compared their percentage

**Table 2** Differentially expressed known lncRNA genes

Human gene symbol	Human locus (hg19)	Human expression pattern	Mouse ortholog expression	Comments
<i>H19</i>	chr11:2016405–2019065	Apical RGC-enriched	RGC-enriched	Imprinted, maternally expressed tumor suppressor
<i>CRNDE</i>	chr16:54952777–54963101	Apical RGC-enriched	Not expressed	Knockout mouse made but no phenotype reported <sup>37</sup>
<i>MIR22HG</i>	chr17:1614797–1619571	Apical RGC-enriched	No differential expression	
<i>LINC00643</i>	chr14:62570095–62606691	ORG-depleted	Neuron-enriched	
<i>TUNAR</i>	chr14:96343108–96391908	ORG-depleted	Neuron-enriched	Regulates pluripotency and neural lineage commitment in mouse
<i>LINC00599</i>	chr8:9753778–9767085	Neuron-enriched	Neuron-enriched	Immediately adjacent to <i>MIR124-1</i>
<i>MIAT</i>	chr22:27042391–27176170	Neuron-enriched	No differential expression	Enriched in mouse TBR2 <sup>+</sup> IP <sup>41</sup>
<i>LINC-PINT</i>	chr7:130628918–130794831	RGC-enriched	Neuron-enriched	Knockout mouse displays general growth defect <sup>37</sup>
<i>RMST</i>	chr12:97856520–97958754	RGC-enriched	n.d.	Enriched in mouse TBR2 <sup>+</sup> IP <sup>41</sup>

Expression patterns of known lncRNA genes with conserved mouse orthologs, most of which show species-specific expression patterns. Human and mouse orthologous pairs were assayed in FACS-purified progenitor and neuron populations by qRT-PCR (see **Supplementary Fig. 7**). n.d., no data.

identity to a reconstructed last common ancestor (LCA) of human, mouse and ferret. Both primates and more distant nonrodent species indeed shared greater identity to the LCA conserved sequences than did rodents (**Fig. 5d** and **Table 3**). These findings are consistent with the interpretation that functional transcripts were present at these loci in the lauriasiatherian LCA and are either highly divergent or lost in the rodent lineage. Thus, many of these newly described human ORG-enriched lncRNAs show comparative patterns of sequence conservation that parallel levels of gyration, being more highly conserved in many larger-brained gyrencephalic species, including other primates and ferrets, and more highly divergent in non-gyrencephalic rodents, consistent with the suggestion that gyrencephaly is an ancestral mammalian trait<sup>42,43</sup>.

## DISCUSSION

Using a combined FACS enrichment and transcriptional profiling strategy, we identified a molecular signature of human ORG comprising hundreds of known genes and newly identified transcripts. Among ORG-enriched genes, we observed a notable over-representation of a well-known transcription factor network, controlled by the critical regulatory factor NEUROG2, and used ferrets to confirm that this transcription factor network drives key steps in ORG production—specifically, delamination from the ventricular neuroepithelium and migration into the SVZ. However, both our human RNA-seq and our ferret immunohistochemical data indicated heterogeneity of expression of NEUROG2 itself within both apical RGC and ORG, and our human single-cell data showed remarkably diverse transcriptional states within both apical RGC and ORG, characterized by the combinatorial expression patterns of classic progenitor markers, proneural transcription factors and new ORG-enriched candidates such as *RASGRP1*, *TTYH2* and *SSTR2*. This heterogeneity was markedly simplified in mouse, consistent with the paucity of ORG in that species, but was more evident in ferret single progenitors, which included a substantial subpopulation of NEUROG2 target-expressing RGC. Finally, we describe gene loci putatively encoding lncRNAs, including several loci with enriched expression in human ORG. Several of these ORG-enriched lncRNA loci show comparative genomic evidence of having been present in the LCA of humans and ferrets, which also possess abundant ORG and are

gyrencephalic, but greatly diverged during rodent evolution, suggesting that these transcripts may be expressed in other species with expanded SVZ progenitor populations. Taken together, our population-level and single-cell transcriptional data show an intriguing correlation between mature cortical size and structure and the heterogeneity of the progenitors that create this structure during development.

Recent studies have provided evidence for or against functional heterogeneity within mouse RGC, particularly with respect to the transcription factors *CUX2* and *FEZF2*, respectively<sup>44,45</sup>. Our human RGC subtype-specific RNA-seq data confirmed that *FEZF2* was highly enriched in both LG<sup>+Pr<sup>hi</sup></sup> and LG<sup>+Pr<sup>lo</sup></sup> RGC subpopulations relative to LG<sup>-</sup> neurons. In contrast, *CUX2* showed a highly significant ( $P = 1.33 \times 10^{-85}$ ), >50-fold enrichment in LG<sup>-</sup> neurons relative to LG<sup>+Pr<sup>hi</sup></sup> apical RGC, with a more modest but still highly significant ( $P = 1.86 \times 10^{-32}$ ) ~8-fold enrichment in LG<sup>+Pr<sup>lo</sup></sup> ORG relative to LG<sup>+Pr<sup>hi</sup></sup> apical RGC. These patterns are consistent with the interpretation that *FEZF2* is expressed in both apical and outer RGC, with no significant difference between the progenitor subsets illustrated in **Figure 3**, whereas *CUX2* is most likely enriched in the *NEUROG2<sup>+</sup>* proneural subsets, consistent with this factor's role in upper layer neuronal morphogenesis. We note, however, that the human specimens available for our studies were from the second half of the second trimester, during the later stages of upper-layer neurogenesis, and that earlier human fetal cortical samples would be required to specifically contrast the expression or function of these two transcription factors in early versus late human radial glial progenitors.

Two previous studies have reported gene expression profiles of the human VZ, ISVZ and OSVZ, using laser capture-assisvto separate the germinal zones from each other and from the postmitotic IZ and CP compartments<sup>13,14</sup>, with one of these studies also directly contrasting human and mouse germinal zones<sup>13</sup>, and another recent study explored the transcriptional signature of human RGC and the differences in gene expression between human and mouse progenitors<sup>15</sup>. The genes reported by these studies as OSVZ-enriched or human RGC-enriched we found to be expressed either in apical RGC or in both apical RGC and ORG (**Supplementary Table 2**), consistent with their being radial glial markers. Conversely, few of the ORG-enriched genes found by our methods were captured by previous studies, highlighting the ability of sorted cell

**Table 3 Expression and conservation of newly identified noncoding transcripts**

Locus (hg19)	Peak expression	Primate identity	Rodent identity	Laurasiatherian identity	Comments
chr1:11510304–11514506	ORG and neurons	96%	93%	95%	"PTCHD2-OS1," ~25 kb upstream of <i>PTCHD2</i>
chr1:117671458–117753549 <sup>a</sup>	ORG and neurons	91%	0%	0%	"VTCN1-OS1," spliced transcript on the opposite strand overlapping the 3' UTR of <i>VTCN1</i> , which is not expressed
chr2:104422113–104497049 <sup>b</sup>	ORG and neurons	97%	93%	91%	Expressed from a bidirectional promoter shared with another new RGC-enriched lncRNA ( <b>Supplementary Fig. 6a</b> )
chr15:78268784–78269017	ORG and neurons	n.d.	n.d.	n.d.	Single-exon transcript maps to a low-complexity region, ~20 kb downstream of apical RGC-enriched <i>TBC1D2B</i>
chr2:6421374–6465945 <sup>b</sup>	ORG	95%	89%	94%	Multiexon, alternatively spliced lncRNA locus ( <b>Supplementary Fig. 6b</b> )
chr2:103911812–103940780	ORG	96%	90%	95%	Spliced transcript, ~475 kb from nearest RefSeq gene ( <i>TMEM182</i> )
chr4:14361021–14520199 <sup>b</sup>	ORG	92%	91%	92%	Opposite strand overlapping the 3' end of another lncRNA, <i>LINCO0504</i> , which is not expressed
chr9:23849902–23850841	ORG	94%	89%	94%	Likely <i>ELAVL2</i> alternative TSS
chr12:55455184–55512962	ORG	97%	90%	95%	Spliced transcript ~30 kb downstream of ORG-enriched <i>NEUROD4</i>
chrX:82593545–82761760	ORG	96%	91%	95%	"POU3F4-OS1," transcript ~1.5 kb upstream on the opposite strand
chr7:33833361–33842784 <sup>b</sup>	Apical RGC	n.d.	n.d.	n.d.	Apical-specific, intergenic (>100 kb) spliced transcript
chr10:45247606–45275449	Apical RGC	94%	86%	94%	Apical-specific, intergenic (>30 kb) spliced transcript ( <b>Supplementary Fig. 6c</b> )
chr12:96970361–97041796 <sup>b</sup>	Apical RGC	96%	91%	96%	Overlaps a locus now called <i>CFAP54</i> ; coding potential uncertain
chr20:21550608–21598247 <sup>b,c</sup>	Apical RGC	97%	92%	96%	"NKX2-2OS" (LOC101929625)
chr21:34394982–34396684	Apical RGC	95%	92%	91%	Unspliced transcript ~1.5 kb upstream of <i>OLIG2</i> TSS; possible <i>OLIG2</i> alternative TSS

<sup>a</sup>No homologous conserved elements outside of primates; may have arisen after split with bushbaby (Galagidae). <sup>b</sup>Overlaps previously reported human lncRNA<sup>39</sup>. <sup>c</sup>Overlaps previously reported human-mouse conserved transcript<sup>40</sup>.

Previously uncharacterized transcripts with differential expression between apical and non-apical RGC subtypes are listed with their cell type of peak expression. Some transcripts overlap previously reported putative human lncRNAs<sup>39,40</sup>, but most are completely unannotated. Manual inspection of RNA-seq reads in their genomic context revealed multiple classes of transcripts, including unannotated alternative transcription start sites (5' UTR exons) and antisense transcripts of known protein-coding genes, as well as many multi-exon intergenic transcripts. For each lncRNA, the identity columns show the results of a comparative genomics analysis of each locus to a computed common ancestor of human, ferret and mouse, indicating that, for many loci, sequence homology was preferentially lost along the rodent lineage, potentially contributing to the evolutionary loss of the ORG cell subpopulation in mouse cortex (for further illustration of an example, see **Fig. 5d**). TSS, transcription start site; n.d., no data.

populations to reveal cell type-specific expression patterns. One previous transcriptional analysis that also identified some of the human ORG-enriched genes found in our study was performed on single progenitors from the embryonic mouse cortex (**Supplementary Table 2**)<sup>46</sup>. Notably, those authors showed by *in situ* hybridization that some human ORG-enriched genes, which they described as labeling a novel progenitor subpopulation intermediate between classic RGC and IP, were expressed in a thin band at the VZ–SVZ border in E14 mouse cortex, suggesting that these cells were indeed transitioning from RGC to IP. These data are consistent with our interpretation that the human ORG transcriptional signature reflects an abundance of cells within the ORG population persisting in just such a transitional state. Remarkably, the comparison of our data with those of Kawaguchi *et al.*<sup>46</sup> suggests that the embryonic mouse cortex may have a cell type that is analogous to the human ORG, but that differs both in morphology, having already retracted its radial fiber, and in position, residing between the VZ and the SVZ rather than superficial to the zone of classic TBR2<sup>+</sup> multipolar IP.

Overall, our data show that human radial glial progenitors, as well as to a lesser extent those of the gyrencephalic ferret, differ most strikingly from mouse RGC in the 'gradedness' of their transition from NEUROG2<sup>-</sup> neuroepithelial RGC to delaminated, multipolar, neuronal lineage-committed IP. Live-imaging studies of the embryonic mouse cerebral cortex have consistently shown that daughter cells from the abventricular mitoses of classic RGC concurrently delaminate, retract their radial fibers, lose PAX6 expression, gain TBR2 expression and

migrate into the SVZ. In contrast, our transcriptional analysis of human ORG and our unbiased single-cell profiling of hundreds of RGC from human and ferret show that these cells exist in a surprising number of transcriptional transitional states between classic RGC and IP.

*Note added in proof:* In work published while this paper was in the press, Florio *et al.*<sup>47</sup> used a complementary approach to similarly isolate and transcriptionally contrast human and mouse apical and outer radial glial progenitors.

## METHODS

Methods and any associated references are available in the online version of the paper.

*Note:* Any Supplementary Information and Source Data files are available in the online version of the paper.

**Accession codes.** Gene Expression Omnibus: GSE66217.

## ACKNOWLEDGMENTS

We thank J. Partlow for coordinating human tissue protocols, D. Gonzalez for animal protocol and technical experimental assistance, S. Lazo-Kallanian for single-cell FACS assistance and all members of the Walsh laboratory for comments and discussion. The *Neurog2-VP16* construct was a generous gift from C. Schuurmans (University of Calgary). This work was supported by grants to C.A.W. from the US National Institutes of Neurological Disease and Stroke (R01 NS032457) and the Paul G. Allen Family Foundation. M.B.J. was supported by a fellowship from the Nancy Lurie Marks Family Foundation. P.P.W. was supported by the Stuart H.Q. &

Victoria Quan Fellowship at Harvard Medical School. Single-cell expression profiling experiments were performed at the Molecular Genetics Core at Boston Children's Hospital (BCH IDDRC, P30 HD18655). Transcriptome analysis was performed using Harvard Medical School's Orchestra high-performance computing cluster, which is partially supported by US National Institutes of Health grant NCRR 1S10RR028832-01. C.A.W. is an Investigator of the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

M.B.J., P.P.W. and R.N.D. designed and conducted experiments and analyzed data. K.D.A. and E.A.M. performed experiments and analyzed data. J.L.H. procured and examined human tissue samples. M.B.J., P.P.W. and C.A.W. interpreted the data and wrote the manuscript.

## COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Florio, M. & Huttner, W.B. Neural progenitors, neurogenesis and the evolution of the neocortex. *Development* **141**, 2182–2194 (2014).
2. Gorski, J.A. *et al.* Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J. Neurosci.* **22**, 6309–6314 (2002).
3. Kowalczyk, T. *et al.* Intermediate neuronal progenitors (basal progenitors) produce pyramidal-projection neurons for all layers of cerebral cortex. *Cereb. Cortex* **19**, 2439–2450 (2009).
4. Smart, I.H.M., Dehay, C., Giroud, P., Berland, M. & Kennedy, H. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb. Cortex* **12**, 37–53 (2002).
5. Fietz, S.A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* **13**, 690–699 (2010).
6. Hansen, D.V., Lui, J.H., Parker, P.R.L. & Kriegstein, A.R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
7. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
8. Gertz, C.C., Lui, J.H., LaMonica, B.E., Wang, X. & Kriegstein, A.R. Diverse behaviors of outer radial glia in developing ferret and human cortex. *J. Neurosci.* **34**, 2559–2570 (2014).
9. Reillo, I., de Juan Romero, C., García-Cabezas, M.Á. & Borrell, V. A role for intermediate radial glia in the tangential expansion of the mammalian cerebral cortex. *Cereb. Cortex* **21**, 1674–1694 (2011).
10. Martínez-Cerdeño, V. *et al.* Comparative analysis of the subventricular zone in rat, ferret and macaque: evidence for an outer subventricular zone in rodents. *PLoS ONE* **7**, e30178 (2012).
11. Johnson, M.B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
12. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
13. Fietz, S.A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc. Natl. Acad. Sci. USA* **109**, 11836–11841 (2012).
14. Miller, J.A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
15. Lui, J.H. *et al.* Radial glia require PDGFD-PDGFR $\beta$  signalling in human but not mouse neocortex. *Nature* **515**, 264–268 (2014).
16. Capela, A. & Temple, S. LeX is expressed by principle progenitor cells in the embryonic nervous system, is secreted into their environment and binds Wnt-1. *Dev. Biol.* **291**, 300–313 (2006).
17. Mo, Z. *et al.* Human cortical neurons originate from radial glia and neuron-restricted progenitors. *J. Neurosci.* **27**, 4132–4145 (2007).
18. Shibata, T. *et al.* Glutamate transporter GLAST is expressed in the radial glia-astrocyte lineage of developing mouse spinal cord. *J. Neurosci.* **17**, 9212–9219 (1997).
19. Weigmann, A., Corbeil, D., Hellwig, A. & Huttner, W.B. Prominin, a novel microvilli-specific polytopic membrane protein of the apical surface of epithelial cells, is targeted to plasmalemmal protrusions of non-epithelial cells. *Proc. Natl. Acad. Sci. USA* **94**, 12425–12430 (1997).
20. Uchida, N. *et al.* Direct isolation of human central nervous system stem cells. *Proc. Natl. Acad. Sci. USA* **97**, 14720–14725 (2000).
21. Seo, S., Lim, J.-W., Yellajoshyula, D., Chang, L.-W. & Kroll, K.L. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *EMBO J.* **26**, 5093–5108 (2007).
22. Gohlke, J.M. *et al.* Characterization of the proneural gene regulatory network during mouse telencephalon development. *BMC Biol.* **6**, 15 (2008).
23. Ochiai, W. *et al.* Periventricular notch activation and asymmetric Ngn2 and Tbr2 expression in pair-generated neocortical daughter cells. *Mol. Cell. Neurosci.* **40**, 225–233 (2009).
24. Rousso, D.L. *et al.* Foxp-mediated suppression of N-cadherin regulates neuroepithelial character and progenitor maintenance in the CNS. *Neuron* **74**, 314–330 (2012).
25. Kovach, C. *et al.* Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex. *Cereb. Cortex* **23**, 1884–1900 (2013).
26. Iacopetti, P. *et al.* Expression of the antiproliferative gene TIS21 at the onset of neurogenesis identifies single neuroepithelial cells that switch from proliferative to neuron-generating division. *Proc. Natl. Acad. Sci. USA* **96**, 4639–4644 (1999).
27. Kawaguchi, A. *et al.* Differential expression of Pax6 and Ngn2 between pair-generated cortical neurons. *J. Neurosci. Res.* **78**, 784–795 (2004).
28. Reid, C.B., Tavazoie, S.F. & Walsh, C.A. Clonal dispersion and evidence for asymmetric cell division in ferret cortex. *Development* **124**, 2441–2450 (1997).
29. Ware, M.L., Tavazoie, S.F., Reid, C.B. & Walsh, C.A. Coexistence of widespread clones and large radial clones in early embryonic ferret cortex. *Cereb. Cortex* **9**, 636–645 (1999).
30. Jackson, C.A., Peduzzi, J.D. & Hickey, T.L. Visual cortex development in the ferret. I. Genesis and migration of visual cortical neurons. *J. Neurosci.* **9**, 1242–1253 (1989).
31. Noctor, S.C., Scholnicoff, N.J. & Julian, S.L. Histogenesis of ferret somatosensory cortex. *J. Comp. Neurol.* **387**, 179–193 (1997).
32. Borrell, V. In vivo gene delivery to the postnatal ferret cerebral cortex by DNA electroporation. *J. Neurosci. Methods* **186**, 186–195 (2010).
33. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
34. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
35. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
36. Ng, S.-Y., Bogu, G.K., Soh, B.S. & Stanton, L.W. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol. Cell* **51**, 349–359 (2013).
37. Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**, e01749 (2013).
38. Hu, P. *et al.* LncRNA expression signatures of twist-induced epithelial-to-mesenchymal transition in MCF10A cells. *Cell. Signal.* **26**, 83–93 (2014).
39. Cabilio, M.N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915–1927 (2011).
40. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
41. Aprea, J. *et al.* Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J.* **32**, 3145–3160 (2013).
42. Kelava, I. *et al.* Abundant occurrence of basal radial glia in the subventricular zone of embryonic neocortex of a lissencephalic primate, the common marmoset *Callithrix jacchus*. *Cereb. Cortex* **22**, 469–481 (2012).
43. Lewitus, E., Kelava, I., Kalinka, A.T., Tomancak, P. & Huttner, W.B. An adaptive threshold in mammalian neocortical evolution. *PLoS Biol.* **12**, e1002000 (2014).
44. Franco, S.J. *et al.* Fate-restricted neural progenitors in the mammalian cerebral cortex. *Science* **337**, 746–749 (2012).
45. Guo, C. *et al.* Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. *Neuron* **80**, 1167–1174 (2013).
46. Kawaguchi, A. *et al.* Single-cell gene profiling defines differential progenitor sub-classes in mammalian neurogenesis. *Development* **135**, 3113–3124 (2008).
47. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* doi:10.1126/science.aaa1975 (26 February 2015).

## ONLINE METHODS

**Human tissue specimens and processing.** Research performed on samples of human origin was conducted according to protocols approved under expedited category 5 with waiver of consent (45 CFR 46.110) by the institutional review boards of Beth Israel Deaconess Medical Center and Boston Children's Hospital. Fetal brain tissue was received after release from clinical pathology, with a maximum post-mortem interval of 4 h. Cases with known anomalies were excluded. Gestational ages were determined using fetal foot length. Tissue was transported in HBSS medium on ice to the laboratory for research processing.

**Purification of cortical progenitors.** Cortical tissue was separated from remaining brain tissue in ice-cold HBSS medium and manually disrupted using a sterile razor blade down to ~1-mm<sup>3</sup> pieces. The tissue was then dissociated into a single-cell suspension using the trypsin Neural Dissociation Kit (Miltenyi Biotec) according to manufacturer's instructions. Cells were placed into FACS pre-sort medium (Neurobasal medium, 0.25% HEPES, 0.5% FBS, rhEGF, rhFGF) for labeling with cell surface antibodies. Cells were labeled in aliquots of 500 µl containing up to 40 million cells with anti-CD15-FITC (BD Biosciences 560997) at 1:10,000, anti-GLAST-PE (Miltenyi Biotec 130-098-804) at 1:10,000, and anti-CD133-APC (Miltenyi Biotec 130-098-829) at 1:1,000 for 30 min at 4 °C and washed twice with pre-sort medium before FACS. Alternatively, minced tissue was cryopreserved before enzymatic dissociation by storing in HBSS plus 10% DMSO, cooled gradually in a cryochamber to -80 °C overnight and transferred to -150 °C for long-term storage.

**RNA isolation, processing and sequencing.** Cells were sorted directly into RNA stabilizing lysis buffer, followed by total RNA extraction (Qiagen). Next-generation sequencing libraries were prepared using Illumina TruSeq v2 according to manufacturer's instructions and sequencing was performed on an Illumina HighSeq 2000. Data were analyzed primarily with the Tuxedo software suite (bowtie/tophat/cufflinks/cummerbund)<sup>48</sup> using the hg19 genome and UCSC KnownGene transcriptome references. Additional R/Bioconductor packages were used for principal component analysis, clustering and the generation of heat maps. Gene set enrichment analyses were performed using DAVID (<http://david.abcc.ncifcrf.gov/>) and comparison of non-reference cufflinks transcripts to published lncRNA catalogs was done in Galaxy (<http://main.g2.bx.psu.edu/>).

**Ferret electroporation.** Timed-pregnant ferrets (*Mustela putorius furo*) were obtained from Marshall Bioresources. Neonatal ferret kits were anesthetized with 5% and maintained at 3% isoflurane using a nose cone during the entire procedure. A small incision was made on the skin at the dorsomedial part of the head using a surgical blade and a hole was opened anterior to the bregma on the left side of the skull, above the lateral ventricle, using an insulin syringe needle. 3–5 µl of DNA construct (1 µg/µl) was injected into the lateral ventricle using a pulled glass micropipette inserted through the craniotomy and the overlying cortical wall. 150-V electric pulses were passed five times at 1-s intervals using paddle electrodes positioned outside the animal's head. The skin incision was closed using VetBond (3M) tissue adhesive and kits were returned to the nest after recovering from anesthesia. Kits were deeply anesthetized before transcardial perfusion with cold PBS and 4% PFA, and brains were extracted and placed in 4% PFA at 4° overnight before processing for immunohistochemistry.

**Immunohistochemistry.** Ferret brains were embedded in 4% low-melting-point agarose and sectioned at 70 µm on a vibrating microtome. Sections were washed in cold 0.1 M PB followed by antigen retrieval in 10 mM citric acid (pH 6.0) plus 0.05% Tween-20 at 80 °C for 30 min. Sections were then cooled to room temperature and washed in cold 0.1 M PB. Sections were blocked for at least 1 h at room temperature (10% normal donkey serum, 0.1% Triton X-100, 0.2% gelatin in PBS). Primary antibodies were incubated at 4 °C in 0.2× blocking buffer for at least 48 h. Primary antibodies included chicken anti-vimentin 1:250 (Abcam

ab24525), chicken anti-Tbr2 1:250 (Millipore AB15894), rabbit anti-GFP 1:1,000 (Abcam ab290), goat anti-Sox2 1:250 (Santa Cruz sc-17320). Sections were washed in PBS and then incubated for 2 h in 0.2× blocking buffer containing Alexa Fluor secondary antibodies (Life Technologies). Slices were then rinsed and coverslipped with Fluoromount-G (Southern Biotech) containing Hoechst 1:1,000 (Roche). Images were obtained with a Zeiss LSM700 confocal microscope and Leica MZ16 F fluorescence stereomicroscope. For quantification, tiled confocal images spanning the entire cortical wall were captured at 20×, stitched, and exported to Photoshop. Three to four pairs of coronal sections from three control (GFP)-electroporated and three NEUROG2-VP16-electroporated hemispheres, matched for the level of section and spatial extent of the electroporation, were imaged and the images were then coded and quantified blind to experimental condition. The Hoechst nuclear counterstain was used to demarcate the borders between VZ, SVZ, IZ and CP, and the numbers of GFP<sup>+</sup> cell bodies in each zone counted. The percentages of GFP<sup>+</sup> cells in each zone were calculated separately for each image and averaged across the images for each brain. The averages for each replicate were then compared across experimental conditions using a paired Student's *t*-test.

**Single-cell mRNA expression profiling.** After cell labeling, single cells were sorted by FACS into skirted 96-well PCR plates containing Pre-Amplification solution (Cells Direct kit, Life Technologies) and appropriate mixtures of Taqman assays (for human and mouse) or validated primer pairs (for ferret). Plates were transported on ice and briefly centrifuged before pre-amplification (94 °C 10 min, 50 °C 60 min, 94 °C 30 s, 50 °C 3 min × 28 cycles). Target-specific cDNA from single cells was harvested, screened for expression of housekeeping genes ACTB and GAPDH and then loaded onto a Biomark chip (Fluidigm) for expression profiling with the panel of qRT-PCR assays. Expression data were processed and analyzed using the Singular Analysis Toolset (Fluidigm) and gplots packages in R. Hierarchical clustering was performed using complete linkage based on Euclidean distance and clusters of cells were defined by cutting the single-cell dendrogram at the same height for all three species.

**Comparative genomics analysis of new lncRNA loci.** Comparative evolutionary analysis of lncRNAs was performed using a modified version of the recently published "forward genomics" approach<sup>49</sup>. Briefly, multisequence fasta files were generated for all conserved regions located within the lncRNAs using existing 100-way vertebrate multiple alignment files available from the UCSC genome browser. Next we generated ancestral sequences for the common ancestor of human, mice and ferrets using the prequel algorithm (--keep-gaps --no-probs --msa-format PHYLIP), part of the PHAST tools<sup>50</sup>. The percentage identities of sequences from all species were determined by alignment to the corresponding ancestral sequence using Needleall, part of the EMBOSS tools<sup>51</sup>. Species with low quality or missing sequence information were excluded from the analysis. Finally, the number of identical bases from all regions within each lncRNA were calculated to yield the percentage identity to the common ancestor.

**Statistics.** No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those generally employed in the field and are comparable to those reported in previous publications<sup>11,13</sup>. For parametric analyses, data distribution was assumed to be normal, but this was not formally tested.

A Supplementary Methods Checklist is available.

- 48. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
- 49. Hiller, M. *et al.* A 'forward genomics' approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Reports* **2**, 817–823 (2012).
- 50. Hubisz, M.J., Pollard, K.S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
- 51. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).