Progenitor Cell Diversity and Function in the Developing Cerebral Cortex

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Wang, Peter Pei. 2015. Progenitor Cell Diversity and Function in the Developing Cerebral Cortex. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467353">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17467353</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Progenitor Cell Diversity and Function in the Developing Cerebral Cortex

A dissertation presented

by

Peter Pei Wang

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

April 2015
Progenitor cell diversity and function in the developing cerebral cortex

Abstract

The human cerebral cortex, the largest structure of our brain, is the seat of our most highly developed cognitive functions. Its normal development, proper function, and ultimate size depend on a precisely controlled balance between self-renewal and differentiation of neural progenitor cells that reside in distinct germinal zones during development. Compared to other species, human cortical progenitor cells are exceptionally diverse in both cellular morphology and gene expression profile.

In particular, previous work has identified specialized progenitors, called outer radial glia (ORG), that are especially common in humans yet virtually absent in rodents and other species with relatively smaller brains. This has led to the suggestion that ORG may represent targets of developmental mechanisms underlying the rapid evolutionary expansion of the human cortex. However, these cells have not been sufficiently profiled at the transcriptional level. To address this question, we combined cell type-specific sorting of cortical progenitor cells with transcriptome-wide RNA-sequencing (RNA-seq) to identify genes enriched in human ORG, which included several targets of the transcription factor Neurogenin as well as a surprising number of previously unknown, evolutionarily dynamic long noncoding RNAs (lncRNAs). We showed that activating the Neurogenin pathway in cortical progenitors of the ferret, a
species with abundant ORG, promotes delamination and outward migration. We then used single-cell transcriptional profiling to compare patterns in human, ferret, and mouse progenitors, and found that a larger proportion of human RGC co-express proneural Neurogenin targets, suggesting greater self-renewal of neuronal lineage-committed progenitors in humans. Finally, comparative genomic analysis of several novel, human ORG-enriched long noncoding RNA genes indicated that many of these loci, while potentially present in the common ancestor of human, ferret, and mouse, show highly distinct patterns of ORG expression accompanied by greater genomic sequence divergence in rodents.

Taken together, we find that the expansion of the ORG subpopulation and increased cortical size in humans is paralleled by increased transcriptional diversity of human RGC. Furthermore, we identify coding and noncoding genes that may be involved in human cortical progenitor identity, function, and evolution.
Table of Contents

Abstract ........................................................................................................................................... iii

Table of Contents .............................................................................................................................. v

Acknowledgements .......................................................................................................................... vi

Dedication .......................................................................................................................................... viii

Chapter 1: Introduction .................................................................................................................... 1

Chapter 2: Purification and transcriptional analysis of human outer radial glia ................................................. 26

Chapter 3: Single-cell analysis of transcriptional heterogeneity in cortical progenitor cells .............................................. 61

Chapter 4: Discussion ....................................................................................................................... 87

Appendix: Supplemental Data and Tables ...................................................................................... 99
Acknowledgements

Without doubt, graduate school and completing my Ph.D. has been the most difficult, rewarding, and exciting thing that I’ve ever done in my life. It would not have been possible without the unending support, patience, and guidance of my friends, family, and colleagues.

First and foremost, I thank my advisor, Christopher A. Walsh who has served as a mentor and role model these past years. He is deeply wise, curious about everything, intellectually fearless, and a generous collaborator. He is a kind and loyal friend, colleague, and mentor. Essentially, he is everything that I hope to be not only as a scientist but also as a person. I will never forget his many jewels of wisdom and I will always be indebted to him for shaping me into the scientist that I have become.

I thank my dissertation advisory committee: John Flanagan, Maxwell Heiman, and Samara Reck-Peterson for their unwavering guidance and support. In particular, thank you to my Chair, John, who has shepherded me through each stage of the graduate school metamorphosis from Qualifying Exam to DAC to Examine Committee and has always advocated for me. I would like to thank my thesis examination committee: John Flanagan, Isaac Chiu, Jeffrey Macklis, and Myriam Heiman for the generous contribution of their time and advice.

I have also had the great fortune of being surrounded by amazing labmates who are not only wonderful colleagues but have become close friends. Chief among them is Matthew Johnson, who is my close comrade in arms on all of the work presented here. We not only turned this work from dream to reality but also had to fight a lot of battles together. In addition, I thank all of the members of the Walsh Lab through the years but
particularly the members of my bay (i.e. “swag bay”) who offered invaluable guidance and advice in both science and life through the years: Jenny Yang, Chiara Manzini, Maria Lehtinen, Kutay Deniz Atabay, Elisabeth Murphy, Byoung-II Bae, and Mollie Woodworth. I also thank all of my graduate school classmates for their unending support and comic relief during purple shirt lunches. Science is just too hard and ultimately not worth it if you cannot do it with friends.

Most importantly, I would like to thank my family. To Heather Haines, who is my partner in all things, and to my parents and brother, who gave me everything, none of this would have been possible without your love.
To my family:

Bill, Yan, Alex, and Heather
Chapter 1

Introduction
Overview

Above all others, the brain is the organ that makes us uniquely human. Studying the human brain likely holds the key to understanding many of our most important philosophical questions such as the nature of our intellect, consciousness, and our sense of individuality. Unfortunately, the human brain is also the most complex organ in our bodies – consisting of at least 80 billion neurons with trillions of precisely made connections (Azevedo et al., 2009). Understanding how this structure is made during development is a longstanding and central challenge of neuroscience. To do this will require deep knowledge of the brain at all levels ranging from molecules and genes to cells, circuits, and systems. This task is further confounded by the fact that the brain continues to change throughout adult life and is prone to degeneration over time. Nonetheless, collectively explaining how the brain works (if at all possible) will likely be one of the greatest scientific achievements of our time.

Unique characteristics of the human cerebral cortex

The six-layered neocortex is the largest and evolutionarily newest part of the mammalian brain exhibiting a remarkable diversity of size and shape suggesting that it has been a constant target of evolutionary selection (Bae et al., 2015). As a result, the human cortex contains many anatomical features that render it distinct from other species. These include its relatively large size (compared to body size), the appearance of complex and distinctive cortical folds (gyrification), the elaboration of the upper cell layers (layers II and III), the appearance of new neuronal subtypes, the development of human-specific patterns of neuronal connectivity (e.g. for controlling speech and manual
dexterity), left-right asymmetry (*e.g.* expansion of the temporal lobe), and pronounced specialization of functional areas (Allman et al., 2002; Hill and Walsh, 2005; Kaas, 2000; Rilling and Insel, 1999a; 1999b). Many of these features are closely linked to some of our most advanced cognitive abilities, including speech and language, executive function, and intellect (Abrahams et al., 2007; Preuss et al., 2004; Tramo et al., 1998). Indeed, developmental disorders that disrupt proper formation of the cerebral cortex and its connectivity often lead to developmental delay along with intellectual or social impairment (Stoner et al., 2014; White and Hilgetag, 2011).

In recent decades, considerable emphasis has been placed on studying the mechanistic underpinnings of human cerebral cortex development under the overarching hypothesis that elucidating the mechanisms of cortex formation will lead to better understanding its normal function, disease states, and evolution. Though great strides have been made using animal model systems, many aspects of human cerebral cortex development cannot be replicated in model systems. For example, rodents, an extremely popular model system, lack subdivided upper cortical layers, certain functional areas found in primates, and are naturally smooth-brained, or lissencephalic (Liao and Zhang, 2008; Mrzljak et al., 1988; Preuss, 1995). Even monkeys lack some of the specialized areas found in the human cortex, particularly within prefrontal regions (Allman et al., 2002; Preuss et al., 2004). As a result, there are critical limitations to what we can learn from studying cerebral cortex development in animal model systems. Thus, these inherent human-specific structural complexities compounded by the scarcity of developing human brain tissue available for research have left many of the
cellular and molecular mechanisms that drive human cortex development and function largely mysterious.

**Development of the cerebral cortex**

*Cortical neurogenesis*

Proper development of the cerebral cortex requires the precise coordination of likely hundreds of cell types and the temporal and spatial regulation of many thousands of genes. The mature neocortex is a six-layered structure that consists primarily of only two major intermingled cell types: neurons and glia. Cortical neurons can be further classified as either excitatory projection neurons, which are born in the dorsal telencephalon and often make long-distance axonal connections with both contralateral and ipsilateral targets, or inhibitory interneurons, which are mostly if not all born in the ventral telencephalon and are more locally restricted in their connectivity (Anderson et al., 2002; Molyneaux et al., 2007; Wonders and Anderson, 2006). Both projection and interneurons can be further subdivided into distinct subtypes based on their histological positioning (both laminar and areal), cellular morphology, connectivity, electrophysiological properties, and gene expression (DeFelipe et al., 2013; Mallamaci and Stoykova, 2006; Sorensen et al., 2015). Work presented in this dissertation will primarily focus on the developmental mechanisms governing projection neurons that arise from germinal zones in the dorsal telencephalon adjacent to the lateral ventricles. Projection neurons destined to reside in different layers of the cortex are generated sequentially in the classically described “inside-out” fashion, whereby neurons destined for superficial cortical layers are born later during neurogenesis and migrate past deep-
layer cells (Angevine and Sidman, 1961). Generally, deep-layer neurons (layers VI and V) send ipsilateral axons toward subcortical targets in the thalamus, hindbrain, and spinal cord while upper-layer neurons (layers II and III) send axons toward targets in the contralateral hemisphere (Greig et al., 2013). In addition to birthdate and laminar position, areal position also influences the identity of cortical projection neurons, as exemplified by sharp changes in laminar composition observed at major functional boundaries such as between primary motor and somatosensory areas (Mallamaci and Stoykova, 2006; O'Leary et al., 2007).

Cortical progenitor cells

Development of the six-layered mammalian neocortex involves precise regulation of the balance between proliferation and differentiation within each of several distinct pools of progenitor cells in the developing germinal zones (Florio and Huttner, 2014). These neuroepithelial cells are first specified from anterior neural tube cells based on morphogen gradients of FGF, BMP, WNT, SHH, and other diffusible signals (Jessell and Sanes, 2000; Monuki and Walsh, 2001; Rallu et al., 2002; Rash and Grove, 2007; Vaccarino et al., 1999). Prior to the onset of neurogenesis, and through mechanisms that are still poorly understood, these early neuroepithelial progenitors undergo a stable transition to become multipotent radial glial cells (RGC) (Götz and Huttner, 2005).

RGC reside apically near the ventricular surface of the dorsal telencephalon and are the primary progenitor source of neocortical excitatory projection neurons and glia during cortical neurogenesis and gliogenesis (Haubensak et al., 2004; Rakic, 1988). They also provide a radial scaffold guiding the migration of newborn neurons from the
germinal zones into the maturing cortical plate (CP) (Rakic, 1988). The neurogenic and cell cycle properties of RGC and their precursors have been studied extensively in the rodent ventricular zone (VZ), where adherens junctions between adjacent RGC comprise the structural integrity of the ventricular wall. Early in cortical neurogenesis, these epithelial-like precursors proliferate through symmetric cell divisions to produce the initial RGC population of the nascent cortex (Takahashi et al., 1999). Subsequently, RGC transition to an asymmetric mode of division, in which each mitosis produces one self-renewed RGC that remains tethered to the VZ and one daughter cell that delaminates (Noctor et al., 2004). Delaminated RGC lose adherens junction contacts and retract both their apical and basal (radial) processes. These now multipolar cells or intermediate progenitors (IP) migrate basally and accumulate to form a second germinal zone called the subventricular zone (SVZ) (Noctor et al., 2007). In rodents, the delamination and repositioning of IP is closely correlated with a reduction in proliferative capacity and onset of neuronal lineage commitment. The vast majority of IP undergo only a limited number of additional cell cycles to produce post-mitotic neurons, which then migrate radially along the RGC basal fibers to populate the CP (Kowalczyk et al., 2009; Noctor et al., 2004). Several groups have postulated that the emergence of IP as a transient amplifying progenitor pool was critical for the evolutionary expansion of neocortical size (Martínez-Cerdeño and Noctor, 2006; Pontious et al., 2008).

Each of these morphological and functional transitions, from early neuroepithelial precursor to RGC to IP, is accompanied by important changes in gene expression. RGC during peak neurogenesis can be identified by their expression of transcription factors (e.g. Pax6 and Sox2), cytoskeletal proteins (e.g. Nestin, Vimentin, and BLBP), cell
surface proteins, (e.g. GLAST), and members of the adherens junction complex (e.g. cadherins and Prominin-1), which are specifically localized to the apical membrane domain of polarized (epithelial-like) cell types (Arai et al., 2005; de Juan Romero and Borrell, 2015; Hatakeyama et al., 2014; Kim et al., 2010; Lehtinen and Walsh, 2011; Muzio and Mallamaci, 2003). When RGC divide asymmetrically during neurogenesis, one daughter cell inherits the basal fiber and remains a proliferating RGC, while the other typically activates expression of the bHLH transcription factor Neurog2, downregulates E-cadherin, delaminates from the adherens junction belt, and migrates basally through the VZ into the SVZ (Das and Storey, 2014; Imayoshi and Kageyama, 2014; Sugimori et al., 2007). There, as Neurog2 expression diminishes, expression of the IP critical regulatory gene Tbr2 is activated, and RGC transcription factors such as PAX6 and SOX2 are repressed (Kovach et al., 2013; Lacomme et al., 2012). One or two terminal cell divisions yield post-mitotic neurons, which then turn off Tbr2 expression and activate neuronal differentiation programs as they migrate into the CP (Imayoshi and Kageyama, 2014; Schuurmans and Guillemot, 2002).

Several classes of genes have already been identified as critical for RGC maintenance and function. The transcription factors Sox2 and Pax6 have been shown to be essential for multipotency and stemness of neural progenitor cells (NPC) in vivo (Bani-Yaghoub et al., 2006; Cavallaro et al., 2008; Englund et al., 2005; Hutton and Pevny, 2011; Muzio and Mallamaci, 2003). In addition to transcription factors, regulation of the mitotic spindle in apical RGC plays a critical role in determining the cell fate of RGC progeny. Mitotic spindle orientation dictates the cleavage plane and affects the distribution of cytoplasmic determinants such as Numb, Notch, and Minibrain, which are
regulators of symmetric vs. asymmetric division of progenitors (Bond et al., 2005; Chenn and McConnell, 1995; Kosodo et al., 2004; Lizarraga et al., 2010; Rhyu et al., 1994; Siller and Doe, 2009; Wang et al., 2014). Finally, the apical complex and primary cilia found at the apical surface serve as a critical signaling center for reception of growth factors and signals such as IGFII present in the cerebral spinal fluid filled lateral ventricles (Kim et al., 2010; Lehtinen and Walsh, 2011; Lehtinen et al., 2011).

**The outer subventricular zone and outer radial glia**

Recently, considerable attention has focused on the observation that the SVZ germinal zone is dramatically expanded in many non-rodent species such as carnivores, primates, and humans. In many of these species, fibers known as the inner fiber layer often run tangentially and histologically divide the SVZ into inner (ISVZ) and outer (OSVZ) compartments (Betizeau et al., 2013; Smart et al., 2002). Furthermore, a new subtype of cortical progenitor termed basal or outer radial glia (ORG) resides in the OSVZ; intriguingly, they are more abundant in the SVZ of gyrencephalic carnivores, non-human primates, and humans compared to smaller-brained, lissencephalic rodents (Fietz et al., 2010; Hansen et al., 2014; Wang et al., 2011). ORG express characteristics of both classical RGC and IP: they are delaminated, lack apical complexes, and located in the SVZ, yet they remain polarized by retaining a basal process that frequently reaches the pial surface (Gertz et al., 2014; Reillo et al., 2011). Although they express most of the transcription factor and cytoskeletal markers of RGC, they lack expression of the apical complex proteins, and a fraction of them also co-express the IP marker TBR2 (Martínez-Cerdeño et al., 2012). Furthermore, their proliferative and lineage
capacity appear to be less restricted than those of IP: they can divide symmetrically to produce two daughter ORG, and may give rise to both neurons and astrocytes late in neurogenesis (Gertz et al., 2014; LaMonica et al., 2013; Reillo et al., 2011). After their initial discovery, some suggested that ORG might be required for cortical gyrification by serving as an additional pool of neurogenic cells in larger-brained species, by providing additional radial scaffolding to mediate long-distance tangential migration of newborn neurons, or by providing additional tension required for cortical “buckling” (Reillo et al., 2011; Ronan and Fletcher, 2014; Ventura-Antunes et al., 2013). However, more recent comparative histological studies have found abundant ORG in lissencephalic species suggesting that ORG may be necessary but are not sufficient for gyrification (García-Moreno et al., 2012; Hevner and Haydar, 2012; Kelava et al., 2012). Nonetheless, the potential significance of this progenitor subtype in the development and evolution of a more complex cortical architecture is still an open question of great interest. However, the paucity of this cell type in the developing mouse neocortex represents a considerable barrier to the study of their molecular and cellular properties as genetic studies of cortical development have traditionally been performed in rodent models. Thus, the molecular characteristics of ORG, particularly with respect to classic apical RGC, remain largely unexplored.

_Ferret as a model for cortical development_

Ferrets are an excellent model system to study many aspects of cortical development that are lacking in rodents. They have a highly stereotyped, gyrified cortex with distinct and subdivided upper cortical layers reminiscent of primates and humans
(Knutsen et al., 2013; Poluch and Juliano, 2015; Reillo and Borrell, 2012; Rowell et al., 2010). Their neurogenic period is also prolonged compared with rodents, beginning during mid-embryogenesis (approximately embryonic day 21) and finishing after birth by the second postnatal week (approximately postnatal day 10), with neuronal migration complete by the fourth postnatal week (approximately postnatal day 30) (Noctor et al., 1997). Critically, they have an expanded OSVZ with abundant ORG throughout neurogenesis (Fietz et al., 2010). Finally, ferrets are also experimentally facile as they generally have relatively large litter sizes (8-12 kits) and are born at mid-cortical neurogenesis when upper layer neurons are generated (Noctor et al., 1997). Recently, we and other groups have developed methods for genetic manipulation of postnatal ferret cortical progenitor cells in vivo by intraventricular injection of DNA constructs followed by electroporation resulting in the expression of transgenes in apical RGC and their progeny (Borrell, 2010; Kawasaki et al., 2012; 2013). This allows us to begin studying the functional role of genes in a gyrified brain with abundant ORG and anatomical characteristics more similar to primates and humans.

Gene expression profiling and cortical development

Transcriptomic studies of the developing cerebral cortex

Over the past decade, many groups have spent considerable effort performing transcriptomic studies of the cerebral cortex in a multitude of species including rodents, non-human primates, and humans (Ayoub et al., 2011; Belgard et al., 2011; Kang et al., 2014; Lein et al., 2006; Molnár and Belgard, 2012; Ng et al., 2009). The ambitious overarching goal of these studies has been to comprehensively define the
transcriptomes of brain cells present within different germinal zones and layers across many areas and regions within different species at different time points. These efforts were often successful at discovering genes important for brain development and function and have been a tremendous resource to the neuroscience community. However, they are not without limitations. Initially, these studies largely depended on microarray technologies to perform genome-wide expression analysis. While these assays are unbiased, they are limited in their ability to detect novel genes, alternative splice variants, and noncoding RNAs (Pan et al., 2008; Wang et al., 2008). More recent efforts have begun to utilize high-throughput RNA sequencing (RNA-seq) to overcome these limitations. Nonetheless, in virtually all of these studies, RNA-seq transcriptome analysis was performed on either whole or microdissected tissue rather than purified cell populations of interest (Fietz et al., 2012; Lui et al., 2014). While this may not significantly confound the analysis of homogenous samples such as the cortical ventricular zone, which primarily consists of apical RGC, most other zones or layers of the cortex are extremely heterogeneous, consisting of at least several major subtypes of progenitors, neurons, and glia. Thus, these datasets are, at best, a population-level expression analysis averaged over all these cell types. Unsurprisingly, finding unique molecular signatures of cell types buried within highly heterogeneous regions of tissue, such as ORG within the human OSVZ, has been a very difficult task, even from carefully microdissected tissue. Overcoming this challenge likely requires either extremely deep sequencing of heterogeneous tissues followed by validation using secondary methods such as in situ hybridization, or purification of cell types out of heterogeneous tissues prior to transcriptomic analysis.
**Single-cell analysis of transcription**

The advent of single-cell approaches to analyze gene expression of dozens of genes simultaneously or indeed across the entire transcriptome promises to further unravel cellular heterogeneity within tissues (Islam et al., 2011; Kawaguchi et al., 2008; Tang et al., 2009). Traditional bulk gene expression analysis of tissues, or even purified cell populations, is limited by the critical assumption that the thousands or millions of cells assayed are essentially homogenous, masking heterogeneity and even potentially introducing misleading correlations between samples (Stegle et al., 2015). These limitations arise from the transcriptional substructure of a cell population, which is especially complex and largely unknown for the developing human brain. Single-cell technologies are particularly pertinent, since recent studies have suggested that classical approaches of distinguishing neural cells, both progenitors and neurons, based primarily on their positioning or morphology may provide an incomplete picture of their full complement (Muñoz-Manchado et al., 2015; Pollen et al., 2014). Indeed, single-cell analysis of gene expression patterns has already been shown to identify additional subtypes of cells in several tissues and systems including cancer cells, the lung, and the adult mouse brain (Dalerba et al., 2011; Patel et al., 2014; Ramsköld et al., 2012; Shalek et al., 2013; Treutlein et al., 2014). Furthermore, these techniques have also been employed to characterize the distinct transcriptional states undertaken by cells during highly dynamic processes such as differentiation and cell fate transitions (Faddah et al., 2013; Shalek et al., 2014; Trapnell et al., 2014).
**Noncoding RNAs**

Recent decades have seen a growing appreciation for both short and long noncoding RNAs (lncRNAs) as a critical class of regulators of gene expression, particularly during development. LncRNAs are typically defined as non-protein coding genes of greater than 200 nucleotides. Similar to mRNAs, they are transcribed by RNA polymerase II and often undergo RNA processing such as alternative splicing and polyadenylation, but they are generally localized in the nucleus (Flynn and Chang, 2014). Their mechanisms of action vary from acting as scaffolds for protein-protein interactions to serving as decoys for miRNA-mediated degradation and signals for epigenetic silencing of DNA elements (Fatica and Bozzoni, 2013; Ramos et al., 2015). Transcriptional profiling of lncRNAs has shown that they are expressed at much lower levels than protein coding mRNAs. Interestingly, they are often highly differentially expressed between cell types, to the point that some have suggested they may be more cell-type specific than mRNA expression (Aprea et al., 2013). Furthermore, initial cross-species analyses of lncRNA expression have shown that they are often poorly conserved between species suggesting that they be important mechanisms of evolution (Necsulea et al., 2014). Unfortunately, these unique features have also made them difficult to functionally characterize, and thus the function of most lncRNAs is presently unknown. A recent study, which generated knockout mice for 18 lncRNAs, showed that loss of lncRNA loci resulted in phenotypes ranging from general growth defects to tissue-specific abnormalities in the lung, heart, and brain. Indeed, knockout of *linc-Brm1b* led to a decrease in IP of the developing mouse cortex with a concomitant reduction in upper layer thickness in the cortical plate (Sauvageau et al., 2013). While
these studies suggest that lncRNAs may be a relevant class of developmental regulators, their expression patterns and functional roles in human cerebral cortical development remains uncharacterized.

**Conclusions**

Any comprehensive and mechanistic understanding of the human brain, or the cortex specifically, will require knowing the genes at work during its development. Differential expression of genes allows cells that largely share the same genome or DNA to arrive at different cell fates and functions. While the Human Genome Project has identified virtually all the genes encoded within the genome, we know far less about how different suites of genes are utilized to determine cellular identity in the body. Elucidating how various transcriptional programs drive cellular identity is not only critical to understanding normal physiological function, but also how physiological function may be changed during disease states. Finally, many of these genetic programs of cellular identity and function may be critical targets of evolutionary selection responsible for dramatic changes in brain structure and function.

This study focuses on: 1) characterizing the transcriptional programs of different subtypes of human cortical progenitor cells with an emphasis on ORG progenitors; 2) functional validation of ORG-enriched transcriptional programs in neonatal ferrets, a gyrencephalic species with abundant ORG, as a model; 3) probing transcriptional heterogeneity of human cortical progenitors by single-cell methods and comparing them with other species including mouse and ferret.
It should be noted that during the course of this dissertation, a separate research group conducted similar experiments to purify human and mouse cortical progenitors for comparative transcriptome analysis. Their transcriptome analysis was presented in a coordinated publication and largely validates the findings of our study (Florio et al., 2015).
REFERENCES


Betizeau, M., Cortay, V., Patti, D., Pfister, S., Gautier, E., Bellemín-Ménard, A.,


Fietz, S.A., Kelava, I., Vogt, J., Wilsch-Bräuninger, M., Stenzel, D., Fish, J.L., Corbeil,


Hevner, R.F., and Haydar, T.F. (2012). The (not necessarily) convoluted role of basal


neural epithelium giving rise to striate and extrastriate cortex in the monkey. Cerebral Cortex 12, 37–53.


Chapter 2

Purification and transcriptional analysis of human outer radial glia
Adapted from a manuscript in Nature Neuroscience (March 3, 2015)

Single Cell Analysis Reveals Transcriptional Heterogeneity of Neural Progenitors in Human Cortex

Matthew B. Johnson1,*, Peter P. Wang1,*, Kutay D. Atabay1, Elisabeth A. Murphy1, Ryan N. Doan1, Jonathan L. Hecht2, Christopher A. Walsh1,3

1Division of Genetics and Genomics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children’s Hospital, Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

2Department of Pathology, Beth Israel Deaconess Medical Center and Department of Pathology, Harvard Medical School, Boston, MA

*These authors contributed equally to this work.

3Address for correspondence:
Christopher A. Walsh
Division of Genetics and Genomics
Boston Children’s Hospital
Center for Life Sciences 15062
300 Longwood Avenue, BCH3150
Boston, MA 02115
christopher.walsh@childrens.harvard.edu
AUTHOR CONTRIBUTIONS

Matthew B. Johnson was an equally contributing first author on this manuscript. Together, we designed and conducted all experiments as well as analyzed and interpreted the data. Kutay D. Atabay assisted with developing the cell sorting strategy, validation of cell sorting populations by immunofluorescence, and postnatal ferret surgical experiments. Elisabeth A. Murphy assisted with postnatal ferret surgical experiments and immunostaining analysis. Ryan N. Doan performed the evolutionary analysis of noncoding RNA genes. Jonathan L. Hecht procured and examined human tissue samples. Christopher A. Walsh is the corresponding author and oversaw and guided all experimentation. Matthew B. Johnson, Christopher A. Walsh, and myself wrote the manuscript.
INTRODUCTION

In contrast to 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 (Betizeau et al., 2013; Fietz et al., 2010; Hansen et al., 2014; Smart et al., 2002). Most notably, basal or outer radial glial cells (ORG), which are highly abundant in the human fetal cortex but rare in the mouse, intriguingly 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 (Fietz et al., 2010; Hansen et al., 2014). 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, and give rise to both neurons and astrocytes (Gertz et al., 2014; Reillo et al., 2011). Finally, a subset of ORG co-expresses TBR2, the proneural transcription factor associated predominantly with IP in rodents (Hansen et al., 2014; Martínez-Cerdeño et al., 2012; Reillo et al., 2011). 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 significant barrier to better understanding their molecular and cellular identity.

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 high-throughput RNA-seq (Fietz et al., 2012; Johnson et al., 2009; Kang et al., 2012).
2014; Lui et al., 2014; Miller et al., 2014). Using this approach, we found hundreds of genes that are specifically enriched in apical RGC, and a smaller but distinct transcriptional signature of human ORG that included both protein coding genes and noncoding RNAs. Interestingly, 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. Finally, comparative genomic analysis of several novel long noncoding RNA (IncRNA) genes indicated that many of these loci, while potentially present in the common ancestor of human, ferret, and mouse, 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 cell surface markers to separate cortical progenitor subtypes using fluorescence-activated cell sorting (FACS) prior to RNA-seq. Human apical RGC, the epithelial progenitor subtype, express LeX (CD15) and GLAST (SLC1A3), as well as prominin (PROM1; CD133) on their apical surface (Capela and Temple, 2006; Mo et al., 2007; Shibata et al., 1997; Uchida et al., 2000; Weigmann et al., 1997). ORG express LeX and GLAST, but lack apical proteins including PROM1
(Fietz et al., 2010). Intermediate progenitors (IP) and neurons lack all three markers. Therefore, we separated LeX and GLAST positive (LG⁺) cells showing the top (LG⁺Pr⁺) and bottom (LG⁺Pr⁻) 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⁻Pr⁻) comprising IP and neurons, among other cells, for RNA-seq analysis (Fig. 1a and Table 1).

Multiple 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⁺Pr⁺ and LG⁺Pr⁻ cells were enriched for markers of neural progenitors and radial glia, while being depleted for neuronal genes, compared to the LG⁻Pr⁻ population (Supplementary Fig. 1a). LG⁺Pr⁺ cells showed enrichment for mRNAs encoding PROM1 and other apical membrane proteins compared to LG⁺Pr⁻ cells (Supplementary Fig. 1a). LeX⁺ cells proliferated in vitro to produce neurospheres that showed SOX2 immunoreactivity and were serially passaged at clonal density, consistent with neural stem cell behavior (Supplementary Fig. 1b) (Capela and Temple, 2006). Furthermore, in RGC sorted from embryonic mouse cortex, PROM1 highly overlapped with both LeX and GLAST, with few LG⁺Pr⁻ 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⁺Pr⁻ population in the mouse also corroborates the absence of LeX and GLAST on IP, as these would appear to be LG⁺Pr⁻ by FACS. Thus, our FACS-based method provides a unique opportunity to assay transcriptome-wide differences between, as well as heterogeneity within, human progenitor subtypes.
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. WG, weeks of gestation. b, Human LG+ cells are predominantly LG+Pr^lo (~80%) whereas virtually all mouse LG+ cells are LG+Pr^hi (>95%) 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 (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.
Table 1 | Cell populations immunolabeled and collected by FACS from human fetal cortex for RNA-seq analysis. 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 the vast majority of radial glial cells, while Prom1 is specific to the apical membrane domain found at the ventricular surface (Capela and Temple, 2006; Mo et al., 2007; Shibata et al., 1997; Uchida et al., 2000; Weigmann et al., 1997).

<table>
<thead>
<tr>
<th>Sorted Pool</th>
<th>Antigenicity</th>
<th>Enriched Markers (qRT-PCR)</th>
<th>Cell Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>LG⁺Pr⁺↑</td>
<td>LeX⁺, Glast⁺, Prom1-high</td>
<td>PAX6, SOX2, VIM, NES, BLBP, PAR3 (PAR3), TJP1 (ZO1), MPP5 (PALS)</td>
<td>apical radial glial cells</td>
</tr>
<tr>
<td>LG⁺Pr⁺⁻</td>
<td>LeX⁺, Glast⁺, Prom1-low</td>
<td>PAX6, SOX2, VIM, NES, BLBP</td>
<td>non-apical radial glial cells, including ORG</td>
</tr>
<tr>
<td>LG⁻Pr⁻⁻</td>
<td>LeX⁻, Glast⁻, Prom1⁻⁻</td>
<td>TUJ1, DCX, MEF2C, NEUN</td>
<td>intermediate progenitors, neurons</td>
</tr>
</tbody>
</table>
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 millions reads [fpkm] > 1). Principal component analysis indicated that the greatest proportion of variability between samples reflected the differences between the LG\(^+\) RGC and LG\(^-\) cells, but the second principal component highlighted differences between the LG\(^+\)Pr\(^{hi}\) apical and LG\(^+\)Pr\(^{lo}\) 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\(^+\) population: relative to the LG\(^-\)Pr\(^-\) pool, LG\(^+\) 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\(^+\) enriched genes included integrin signaling and basement membrane components, such as laminins, consistent with both LG\(^+\)Pr\(^{hi}\) and LG\(^+\)Pr\(^{lo}\) subpopulations maintaining radial processes contacting the pial basement membrane, as has been shown for ORG (Fietz et al., 2010; Hansen et al., 2014; Reillo et al., 2011). In all, 552 genes significantly differed between LG\(^+\)Pr\(^{hi}\) and LG\(^+\)Pr\(^{lo}\) cells, with 79 of these genes specifically enriched or depleted in LG\(^+\)Pr\(^{lo}\) non-apical ORG (Fig. 1d and Supplementary Fig. 3). Interestingly, among the genes upregulated in LG\(^+\)Pr\(^{lo}\) ORG, six transcription factors (TFs) – \textit{HES6}, \textit{NEUROD4 (Atoh3, Math3)}, \textit{NHLH1 (HEN1, NSCL1)}, \textit{NEUROD1}, \textit{CBFA2T2 (Mtgr1)}, and \textit{MYT1} – are all downstream of the critical regulatory gene \textit{NEUROG2}, which in mouse cortex and chick spinal cord initiates delamination and neuronal lineage commitment of neural precursors (Gohlke et al., 2008; Kovach et al.,
2013; Ochiai et al., 2009; Rousso et al., 2012; Seo et al., 2007). Interestingly, NEUROG2 itself and two additional early markers of neuronal fate commitment – TBR2 (EOMES) and BTG2 (Tis21) – were all highly expressed in both LG Prhi and LG Prlo subpopulations (Fig. 1d) (Iacopetti et al., 1999; Ochiai et al., 2009). Given that NEUROG2 expression is transient in mouse apical progenitors during their transition from RGC to TBR2+ IP, 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 (Kawaguchi et al., 2004; Ochiai et al., 2009).

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 (Reid et al., 1997; Reillo et al., 2011; Ware et al., 1999). 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, thus activating all direct downstream targets of NEUROG2 in ferret apical RGC (Kovach et al., 2013). Following delivery of the Neurog2-VP16 expression construct, we allowed ferret kits to develop for up to ten days post-electroporation (DPE), during which time many neurons of the upper cortical layers are generated (Fig. 2c) (Borrell, 2010; Jackson et al., 1989; Noctor et al., 1997). In both control (pCAG-GFP) and Neurog2-VP16 brains, we observed numerous
Figure 2 | NEUROG2 regulates progenitor morphology and molecular identity in the developing cortex of the gyrencephalic ferret. a, Co-expression of RGC marker SOX2 and proneural marker NEUROG2 in developing ferret cortex. Numerous SOX2⁺NEUROG2⁺ 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 large numbers of apical and non-apical RGC producing neurons destined for the upper cortical layers. b, Higher magnification of SOX2⁺NEUROG2⁺ progenitors at P2 in the SVZ co-expressing the RGC neurofilament protein Vimentin (VIM), which labels the basal radial process (yellow arrowhead). 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⁺ cells are observed throughout the developing cortical wall, with GFP⁺ radial fibers extending from the germinal zones to the pial surface and newborn neurons migrating through the SVZ and intermediate zone (IZ) into the cortical plate (CP). Scale bars: 50 µm (a), 10 µm (b), 1mm (c)
Figure 2 | NEUROG2 regulates progenitor morphology and molecular identity in the developing cortex of the gyrencephalic ferret. **d**, Littermates electroporated with NEUROG2 gain-of-function (pCAG-Neurog2-VP16) or GFP control (pCAG-GFP) expression constructs, analyzed at seven DPE (P1:P8) for the distribution of GFP+ cells and their coexpression of SOX2. ISVZ, inner SVZ; OSVZ, outer SVZ; MZ, marginal zone. **e**, At all survival time-points, we identified numerous GFP+SOX2+ apical RGC in the VZ as well as occasional GFP+SOX2+ ORG with soma 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+ cells in the SVZ/IZ that are SOX2-negative with the morphology of radially migrating newborn neurons. **g**, At 7 to 9 DPE, NEUROG2 gain-of-function induced a significant shift of GFP+ cells from the VZ into the SVZ/IZ compared to the control (asterisk denotes p < 0.05, paired t-test; exact p-values: VZ=0.026, SVZ/IZ=0.026, CP=0.034; n=3 animals per condition; 3-4 brain sections counted per animal; data represented as mean ± SEM), 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: 100 µm (d), 20 µm (e,f).
GFP^SOX2^ and GFP^TBR2^ cells in the VZ, including cells in the basal VZ and inner SVZ with a characteristic ORG morphology (Fig. 2d,e); as well many GFP^+ 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). After 7-9 DPE, NEUROG2-VP16 induced a significant shift in the proportion of GFP^+ cells from the VZ to the SVZ/IZ, and a concomitant reduction in the proportion of GFP^+ cells co-expressing SOX2 (Fig. 2g), with the majority of NEUROG2-VP16+ 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, and found that nearly all human ORG-enriched NEUROG2 downstream targets were highly upregulated by the NEUROG2-VP16 construct in ferret, relative to 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. Altogether, 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.

Novel long noncoding transcripts enriched in human ORG

Given the species differences in RGC subpopulations revealed by our single-cell analysis (Discussed in Chapter 3), we next searched our RNA-seq data for transcriptional influences on species differences in RGC molecular identity, identifying candidate non-conserved RNA transcripts including IncRNAs, which are evolutionarily dynamic, frequently lacking human-mouse homology, and are involved in critical neural developmental processes such as progenitor pluripotency, neurogenesis, and epithelial-mesenchymal transition (Hu et al., 2014; Iyer et al., 2015; Necsulea et al., 2014; Ng et al., 2013; Sauvageau et al., 2013). We compared 253 differentially expressed unannotated loci (Fig. 3a) to two published human IncRNA catalogs and identified 75 loci overlapping putative human IncRNAs, while only 18 loci matched reported human-mouse conserved IncRNAs (Fig. 3b and Table 2), suggesting that our human RGC subtype-specific novel transcripts include numerous novel IncRNAs that may lack any homologous mouse transcripts (Cabili et al., 2011; Guttman et al., 2009). Surprisingly, we found that a much greater proportion of differentially expressed novel loci were specifically enriched in ORG, compared to known genes (Fig. 3c; 2.4% vs. 0.7%; \( p = 0.012 \), Fisher’s exact test), suggesting that IncRNAs are especially relevant to the molecular identity and function of the ORG subpopulation in humans. By manual inspection, we determined that although a few novel transcripts reflected incomplete
Figure 3 | Novel transcripts detected by RNA-seq include previously unknown IncRNAs with distinct RGC subtype expression patterns and evolutionary conservation. a, Differential expression patterns of selected ncRNA loci, including several novel multi-exon IncRNAs (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 non-coding RNA genes revealed a number of reported human IncRNAs (Cabili et al., 2011), and a smaller number of human-mouse conserved IncRNAs (Guttman et al., 2009). c, A significantly greater proportion of novel differentially expressed transcripts were specifically enriched in LG+Prlo non-apical progenitors (2.4%), 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 novel ORG-enriched IncRNAs was performed by comparing conserved elements from within each genomic locus from 58 species to a computed ancestral sequence for the Laurasiatherian last common ancestor (LCA) of human, ferret, and mouse, the three species examined in this study (see Table 3). Here, we show a detailed example from a human ORG-enriched IncRNA gene located on chromosome 2. At left is shown the percent identity of each species’ conserved elements from this locus to the LCA sequence, which demonstrates that rodents (highlighted in green) show a highly divergent sequence compared to both primates and other more distantly related groups, including carnivores. The panel at right illustrates multi-species genomic alignments to the same human locus for primates, rodents, and other Laurasiatherian species (top), with the human ORG RNA-seq reads (middle) and assembled transcripts (bottom), illustrating the greater sequence divergence of rodents compared to either non-human primates or other more distant Laurasiatherian species. Similar results were found for several other ORG-enriched IncRNA transcripts, as summarized in Table 3 (% ID columns).
Figure 3 (Continued)
Figure 3 (Continued)
**Table 2** | **Differentially expressed known IncRNA genes.** Expression patterns of known IncRNA 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 (Supplementary Fig. 6).
<table>
<thead>
<tr>
<th>Known IncRNA genes</th>
<th>Locus (hg19)</th>
<th>Human Expression Pattern</th>
<th>Mouse Expression Pattern</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>H19 (H19)</td>
<td>chr11:2016405-2019065</td>
<td>apical RGC-enriched</td>
<td>RGC-enriched</td>
<td>imprinted maternally-expressed tumor suppressor</td>
</tr>
<tr>
<td>CRNDE (Crnde)</td>
<td>chr16:5495277-54963101</td>
<td>apical RGC-enriched</td>
<td>not expressed</td>
<td>knockout mouse made but no phenotype reported (Sauvageau et al.)</td>
</tr>
<tr>
<td>MIR22HG (Mir22hg)</td>
<td>chr17:1614797-1619571</td>
<td>apical RGC-enriched</td>
<td>no differential expression</td>
<td></td>
</tr>
<tr>
<td>LINKD0643 (1700086119Rik)</td>
<td>chr14:612570095-62606691</td>
<td>ORG-depleted</td>
<td>neuron-enriched</td>
<td></td>
</tr>
<tr>
<td>TUNAR (Tunar)</td>
<td>chr14:96134330-96339908</td>
<td>ORG-depleted</td>
<td>neuron-enriched</td>
<td></td>
</tr>
<tr>
<td>LINC00599 (A93601012Rik (?))</td>
<td>chr8:97537556-97670685</td>
<td>neuron-enriched</td>
<td>neuron-enriched</td>
<td>immediately adjacent to MIR124-1</td>
</tr>
<tr>
<td>MIAT (Miat)</td>
<td>chr2:27042391-27176170</td>
<td>neuron-enriched</td>
<td>no differential expression</td>
<td>enriched in mouse TBR2+ intermediate progenitors (Aprea et al.)</td>
</tr>
<tr>
<td>LINCHPINT (linc-HPint)</td>
<td>chr7:130628916-130794831</td>
<td>neuron-enriched</td>
<td>neuron-enriched</td>
<td>knockout mouse displays general growth defect (Sauvageau et al.)</td>
</tr>
<tr>
<td>RMST (Rmst)</td>
<td>chr12:97856020-97998754</td>
<td>RGC-enriched</td>
<td>n.d.</td>
<td>enriched in mouse TBR2+ IP (Aprea et al.)</td>
</tr>
</tbody>
</table>
annotations of known genes (e.g., alternative transcription start sites or untranslated regions), the majority resemble bona fide unannotated genes, many of which show multiple exons and alternative splicing (Table 3 and Supplementary Fig. 5), and none of which have previously been reported in cortical development.

Few known lncRNAs that are functionally essential or have been transcriptionally profiled 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 (Aprea et al., 2013; Sauvageau et al., 2013). Of 18 lncRNAs recently knocked out in mice we identified only two orthologs with appreciable expression in human developing cortex (Table 2). Similarly, of 15 mouse IP-enriched lncRNAs, 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) (Aprea et al., 2013). Within conserved lncRNAs between human and mouse, we find several – including LINC-PINT, TUNAR, CRNDE, and MIR22HG – which are depleted in mouse RGC but enriched in human apical and outer RGC suggesting potentially distinct functions in cortical development (Supplementary Fig. 6). 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.
Table 3 | Expression and conservation of novel non-coding transcripts. 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 IncRNAs but most are completely unannotated (Cabili et al., 2011; Guttman et al., 2009). Manual inspection of RNA-seq reads in their genomic context reveals 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 IncRNA, “% ID” 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. 3d).
Table 3 (Continued)

<table>
<thead>
<tr>
<th>Locus (hg19)</th>
<th>Peak Expression</th>
<th>% ID Primates</th>
<th>% ID Rodents</th>
<th>% ID Laurasiatheria</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr1:13510304-13514506 ORG and neurons</td>
<td>98%</td>
<td>93%</td>
<td>99%</td>
<td>&quot;PTCHD2-OS1&quot;, novel transcript ~25 kb upstream of PTCHD2</td>
<td></td>
</tr>
<tr>
<td>chr1:11767148-11775354 ORG and neurons</td>
<td>92%</td>
<td>0%</td>
<td>0%</td>
<td>&quot;VRTN1-OS1&quot;, novel spliced transcript on the opposite strand overlapping the 3'UTR of VRTN1, which is not expressed expressed from a bidirectional promoter shared with another novel, RGC-enriched lincRNA (see Extended Data Fig. 5d)</td>
<td></td>
</tr>
<tr>
<td>* chr1:104422313-10449049 ORG and neurons</td>
<td>97%</td>
<td>93%</td>
<td>93%</td>
<td>single-exon transcript maps to a low-complexity region ~20 kb downstream of apical RGC-enriched BAC1D8B</td>
<td></td>
</tr>
<tr>
<td>chr15:78268764-78269007 ORG and neurons</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chr2:6421374-645945 ORG G</td>
<td>99%</td>
<td>89%</td>
<td>94%</td>
<td>novel multi-exon, alternatively spliced lincRNA locus (see Extended Data Fig. 5b)</td>
<td></td>
</tr>
<tr>
<td>* chr2:103911812-103940780 ORG G</td>
<td>92%</td>
<td>91%</td>
<td>92%</td>
<td>opposite strand overlapping the 3' end of another lincRNA locus LINC00504, which is not expressed</td>
<td></td>
</tr>
<tr>
<td>* chr3:284990-285084 ORG G</td>
<td>94%</td>
<td>89%</td>
<td>94%</td>
<td>Elavl2 alternative TSS</td>
<td></td>
</tr>
<tr>
<td>chr5:5455-55512962 ORG G</td>
<td>97%</td>
<td>90%</td>
<td>93%</td>
<td>novel spliced transcript ~30 kb downstream of ORG-enriched NEUROD4</td>
<td></td>
</tr>
<tr>
<td>chr8:8253954-82547170 ORG G</td>
<td>96%</td>
<td>91%</td>
<td>95%</td>
<td>&quot;POU3F4-OS1&quot;, novel transcript *1.5 kb upstream on the opposite strand</td>
<td></td>
</tr>
<tr>
<td>* chr7:3833361-3841278 ORG apical RGC</td>
<td>n.d.</td>
<td>n.d.</td>
<td>n.d.</td>
<td>apical-specific intergenic (&gt;100 kb) spliced transcript</td>
<td></td>
</tr>
<tr>
<td>* chr10:45247606-45275449 ORG G</td>
<td>94%</td>
<td>86%</td>
<td>94%</td>
<td>apical-specific intergenic (40 kb) spliced transcript (see Extended Data Fig. 5c)</td>
<td></td>
</tr>
<tr>
<td>* chr12:36970381-37041796 ORG G</td>
<td>98%</td>
<td>91%</td>
<td>98%</td>
<td>overlaps the C12orf55/63 locus, now known as CRAP54; &quot;cilia and flagella associated 54&quot;; coding potential uncertain</td>
<td></td>
</tr>
<tr>
<td>* A chr20:21550608-21558347 ORG apical RGC</td>
<td>97%</td>
<td>92%</td>
<td>98%</td>
<td>&quot;NXX2-20S&quot; (LOC101929625)</td>
<td></td>
</tr>
<tr>
<td>* chr21:34394392-34396884 ORG apical RGC</td>
<td>93%</td>
<td>92%</td>
<td>93%</td>
<td>unspliced transcript *3.5 kb upstream of OLIG2 TSS; possible novel OLIG2 alternative TSS</td>
<td></td>
</tr>
</tbody>
</table>

§ no homologous region outside of primates may have arisen after split with bushbaby
* overlaps previously reported human IncRNA (Cabili et al., 2011)
# overlaps previously reported human–mouse conserved transcript (Guttman et al., 2009)
TSS, transcription start site; IP, intermediate progenitors
To probe the evolutionary history of ORG-enriched IncRNAs, 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 IncRNA genomic loci and compared their percentage 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 3d and Table 3). These findings are consistent with the interpretation that functional transcripts were present at these loci in the laurasiatherian LCA and are either highly divergent or lost in the rodent lineage. Thus, many of these newly described human ORG-enriched IncRNAs show comparative patterns of sequence conservation that parallel levels of gyrification, 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 (Kelava et al., 2012; Lewitus et al., 2014).

CONCLUSION

Using a combined FACS enrichment and transcriptional profiling strategy, we identified a molecular signature of human ORG comprising hundreds of known genes and novel transcripts. Among ORG-enriched genes, we observed a notable overrepresentation 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. Finally, we describe novel gene loci, putatively encoding IncRNAs, including several loci with enriched expression in human ORG. Several of these ORG-enriched IncRNA 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.
EXPERIMENTAL 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 two to four hours following release from clinical pathology. 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 \( \sim 1\text{-mm}^3 \) 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” media (Neurobasal media, 0.25% HEPES, 0.5% FBS, rhEGF, rhFGF) for labeling with cell surface antibodies. Cells were labeled in aliquots of 500ul 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 minutes at +4°C and washed twice with pre-sort media before FACS. Alternatively, minced tissue was cryopreserved prior to enzymatic dissociation by storing in HBSS + 10% DMSO, cooled gradually in a cryochamber to -80°C overnight, and transferred to -150°C for long-term storage.
**RNA Isolation, Processing, and RNA 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) (Trapnell et al., 2012) using the hg19 genome and UCSC KnownGene transcriptome references. Additional R/Bioconductor packages were used for principal component analysis, clustering, and the generation of heatmaps. 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 utilizing 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. 3-5µl of DNA construct (1µg/µl) was injected into the lateral ventricle using a pulled glass micropipette inserted through the craniotomy and the overlying cortical wall. 150V electric pulses were passed 5 times at 1s 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 prior to
transcardial perfusion with cold PBS and 4% PFA, and brains were extracted and placed in 4% PFA at +4° overnight prior to processing for immunohistochemistry.

**Immunohistochemistry.** Ferret brains were embedded in 4% low-melting-point agarose and sectioned at 70µm on a vibrating microtome. Sections were washed in cold 0.1M PB followed by antigen retrieval in 10mM citric acid (pH 6.0) + 0.05% Tween-20 at 80°C for 30 minutes. Sections were then cooled to room temperature and washed in cold 0.1M PB. Sections were blocked for at least 1 hour 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 0.2X blocking buffer for at least 48 hours. 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 two hours in 0.2X blocking buffer containing AlexaFluor 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 Zeiss LSM700 confocal microscope and Leica MZ16 F fluorescence stereomicroscope. For quantification, tiled confocal images spanning the entire cortical wall were captured at 20x, stitched, and exported to Photoshop. Three to four pairs of coronal sections from three control (GFP) 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+ cell bodies in each zone counted. The percentages of GFP+...
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.

**Comparative Genomics Analysis of Novel IncRNA Loci.** Comparative evolutionary analysis of IncRNAs was performed using a modified version of the recently published “Forward Genomics” approach (Hiller et al., 2012). Briefly, multi-sequence fasta files were generated for all conserved regions located within novel IncRNAs 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 (Hubisz et al., 2011). The percent identities of sequences from all species were determined by alignment to the corresponding ancestral sequence using Needleall, part of the EMBoss tools (Rice et al., 2000). Species with low quality or missing sequence information were excluded from the analysis. Finally, the number of identical bases from all regions within each IncRNA were calculated to yield the %ID to the common ancestor.

**Statistics.** No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those generally employed in the field and are comparable to those reported in previous publications (Fietz et al., 2012; Johnson et al., 2009). For parametric analyses, data distribution was assumed to be normal but this was not formally tested.

**Accession Codes.** RNA sequencing data are available from the Gene Expression Omnibus, GSE66217.
Acknowledgements

We thank J. Partlow for coordinating human tissue protocols; D. Gonzalez for animal protocol and technical experimental assistance; Suzan Lazo-Kallanian for single-cell FACS assistance; and all members of the Walsh lab for comments and discussion. The Neurog2-VP16 construct was a generous gift from Carol Schuurmans (University of Calgary). This work was supported by grants to C.A.W. from the 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. is a Stuart H.Q. & Victoria Quan Fellow at Harvard Medical School. Transcriptome analysis was performed using Harvard Medical School’s Orchestra high-performance computing cluster, which is partially supported by NIH grant NCRR 1S10RR028832-01. C.A.W. is a Distinguished Investigator of the Paul G. Allen Family Foundation, and an Investigator of the Howard Hughes Medical Institute.
REFERENCES


Chapter 3

Single-cell analysis of transcriptional heterogeneity in cortical progenitor cells
Adapted from a manuscript in Nature Neuroscience (March 3, 2015)

Single Cell Analysis Reveals Transcriptional Heterogeneity of Neural Progenitors in the Human Cortex

Matthew B. Johnson¹,*, Peter P. Wang¹,*, Kutay D. Atabay¹, Elisabeth A. Murphy¹, Ryan N. Doan¹, Jonathan L. Hecht², Christopher A. Walsh¹,³

¹Division of Genetics and Genomics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children’s Hospital, Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

²Department of Pathology, Beth Israel Deaconess Medical Center and Department of Pathology, Harvard Medical School, Boston, MA

*These authors contributed equally to this work.

³Address for correspondence:
Christopher A. Walsh
Division of Genetics and Genomics
Boston Children’s Hospital
Center for Life Sciences 15062
300 Longwood Avenue, BCH3150
Boston, MA 02115
christopher.walsh@childrens.harvard.edu
AUTHOR CONTRIBUTIONS

Matthew B. Johnson was an equally contributing first author on this manuscript. Together, we designed and conducted all experiments as well as analyzed and interpreted the data. Kutay D. Atabay assisted with developing the cell sorting strategy, validation of cell sorting populations by immunofluorescence, and postnatal ferret surgical experiments. Jonathan L. Hecht procured and examined human tissue samples. Christopher A. Walsh is the corresponding author and oversaw and guided all experimentation. Matthew B. Johnson, Christopher A. Walsh, and myself wrote the manuscript. Data on single-cell RNA sequencing of cortical progenitors is currently unpublished work and the bioinformatic analysis was performed in collaboration with Peter V. Kharchenko and Jean Fan (Center for Biomedical Informatics, Harvard Medical School).
INTRODUCTION

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. 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 progenitors, radially migrating projection neurons, and tangentially migrating interneurons originating from separate progenitor pools in the ventral telencephalic germinal zones (Anderson et al., 2002; Wonders and Anderson, 2006). Our progenitor sorting strategy (discussed in Chapter 2) provided the opportunity identify a molecular signature of human ORG but also to further probe transcriptional heterogeneity of neural progenitor cells generally in the developing cortex. Here, our single-cell profiling by single-cell multiplex qRT-PCR and RNA-seq has 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 co-expressing classic markers of both self-renewing RGC and neuronal lineage-committed IP.
RESULTS

Single-cell analysis of species-specific RGC heterogeneity

Both our human RNA-seq and ferret immunofluorescence data (discussed in Chapter 2) demonstrate \textit{NEUROG2$^+$} RGC subpopulations in both the VZ and SVZ, intermingled with \textit{NEUROG2$^-$} progenitors, exemplifying the heterogeneity that confounds population-level transcriptome comparisons. Thus, 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 that included markers for all RGC (\textit{PAX6}, \textit{SOX2}, \textit{GLAST}, \textit{BLBP (FABP7)}, \textit{VIM}, \textit{NES}) and apical RGC (\textit{PROM1}, \textit{PARD3}, \textit{MPP5}, \textit{TJP1}); proneural Neurogenin pathway TFs; and additional validated LG$^+$Pr$^{lo}$, ORG-enriched genes (Supplementary Fig. 3). Among 546 sorted single human progenitors, \textit{PAX6}, \textit{SOX2}, \textit{GLAST}, \textit{BLBP}, and \textit{VIM} were detected in 93±3% of cells (Supplementary Fig. 7a) 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 which we refer to as apical/multipotent; apical/proneural; non-apical/multipotent; and non-apical/proneural (Fig. 4a and Supplementary Fig. 7b). Apical RGC subpopulations (clusters I, II, V in Fig. 4a; 71±8% detection rate for apical genes) could be divided into apical/multipotent (I) and apical/proneural (II and V) subpopulations based on their lesser (21±5%) or greater (81±7%) expression of proneural Neurogenin pathway TFs, respectively.
Figure 4 | 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^Pr^lo-enriched genes such as TTYH2 and PLCB4. Hierarchical clustering and heatmap representation of single-cell qRT-PCR data (left) indicates the co-expression 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^TBR2^+ RGC (clusters II, III, V). In addition, proneural RGC can be further subdivided according to their expression of downstream factors and additional LG^Pr^lo-enriched genes (e.g., compare clusters II and V). Ct, cycle threshold b, The same genes assayed (as Figure 4.1) 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 major species differences, namely, the mouse has fewer non-apical cells overall; few if any multipotent (NEUROG2^TBR2^−) non-apical cells, suggesting the absence of a significant subpopulation of proliferative ORG; and very few cells expressing other human subset-enriched genes (e.g., TTYH1, PLCB4). c, Violin plots of gene expression distributions, with numbers 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.
Figure 4 (Continued)
Figure 4 (Continued)
ORG-like non-apical subpopulations (clusters III and IV) with lower detection rates of apical complex transcripts (31±6%) were similarly subdivided according to lower or higher rates of proneural gene expression (22±7% vs. 65±15%). Notably, most non-apical/proneural cells were NEUROG2⁺ (cluster III), consistent with observations in the mouse that NEUROG2 represses apical identity and is then downregulated upon delamination (Ochiai et al., 2009). Finally, both apical and non-apical proneural RGC were further subdivided by expression of other LG⁺Prlo-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-neural stem cell niches (Jaitin et al., 2014; Treutlein et al., 2014).

In contrast to the human analysis, single RGC from the embryonic day (E)16-17 mouse cortex showed fewer distinct transcriptional states (Fig. 4b and Supplementary Fig. 7a), and rarely expressed many of the genes that defined subsets of human cells (Fig. 4b,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. 4b), corresponding to the apical/multipotent (cluster i), apical/proneural (ii), and non-apical/proneural (iii) subsets observed in the human cortex. Although a substantial subset of mouse RGC co-expressed proneural TFs, the proportion of cells was significantly smaller than in human (27% in mouse vs. 47% in human; p=9.57E⁻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. 4b,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 share 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+ and LG− cells from neonatal ferret cortex, at which time middle and upper layers of cortex are being generated, roughly corresponding to mouse E16-17 or human 16-20 WG, and performed population-level RNA-seq (Jackson et al., 1989; Noctor et al., 1997). Ferret LG+ cells were enriched for most previously described RGC marker genes, and showed transcriptome-wide expression patterns similar to LG+ cells from human and mouse cortex (Fig. 5a). Having validated that LG+ cells comprise a substantial proportion of ferret RGC, we performed single-cell profiling on 185 single LG+ cells from the neonatal ferret cortex, and interestingly found an intermediate degree of heterogeneity compared to the human and mouse progenitors (Fig. 5b).
Figure 5 | Population-level whole-transcriptome RNA-seq and single-cell expression analysis of ferret RGC. a, Expression heatmaps of known progenitor and neuronal marker genes, as well as selected human RGC-enriched gene sets, from LG⁺ and LG⁻ 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⁺ 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⁺ progenitors was performed using the same gene panel as shown in Figure 4 for human and mouse RGC. As in humans, a substantial fraction of ferret cells in clusters i, iv, and v co-express both RGC markers and are Tbr2⁺Neurog2⁺, consistent with our immunohistochemical analysis of NEUROG2 expression in the ferret (Fig. 2a) and suggesting this “proneural” RGC transcriptional state is conserved. Similar to 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 (e.g. Rasgrp1) are expressed in fewer ferret RGC, while others (e.g. Plcb4, Sstr2, Gadd45g, Ttyh2) appear more homogenous across all cells. Interestingly, 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.
Figure 5 (Continued)
Figure 5 (Continued)
The vast majority of sorted ferret cells expressed classic RGC markers, confirming the specificity of the sorting, while a subset co-expressed 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). Interestingly, the ferret homologs of some human ORG-enriched genes (e.g., Rasgrp1) were preferentially expressed in these proneural cells, as in human, whereas most human ORG genes were more homogeneously expressed in most ferret cells (Ttyh2, Sstr2), and conversely, some genes were heterogeneously expressed in novel subsets of cells that did not correspond to those seen in human (Foxn2). Finally, we noted that a greater proportion of ferret cells expressed the gliogenic marker Gfap compared to human cells, which is consistent with evidence that ferret ORG have astrogliogenic potential (Reillo et al., 2011). Taken together, our single-cell analyses from human, ferret, and mouse implicate a large number of genes acting in a coordinated network that may be responsible for the evolution of novel progenitor transcriptional states critical for human cortical development.

Transcriptome-wide gene expression analysis of single human RGC

Single cell RNA-seq has become a powerful tool in the study of cellular transcriptional states particularly within heterogeneous tissues. Recent studies have successfully utilized relatively low-coverage RNA-seq of single cells to reconstruct cell lineages, identify novel cell types, define cellular heterogeneity, and discover critical
signaling pathways within tissues (Hashimshony et al., 2014; Pollen et al., 2014; Shalek et al., 2013; Treutlein et al., 2014; Zeisel et al., 2015). Thus far, these studies have largely examined tissues or collections of cells with distinct morphologies and readily discernable gene expression differences between major cell types. Our FACS strategy provides a unique opportunity to purify human RGC and to identify potentially novel transcriptional states within this cell type. To address this question, we modified our FACS workflow by first purifying LG⁺ RGC, at the population level, from an 18 WG human fetal cortex prior to single cell capture using the Fluidigm C1 platform and RNA-seq (Fig 6a). We successfully captured 215 single LG⁺ RGC and sequenced each cell at an average depth of >2 million reads per cell (data not shown).

The characterization of novel heterogeneity within a cell population by unbiased methods requires identification of gene expression differences that serve as critical drivers of biologically meaningful transcriptional states. However, at the single-cell level, estimating gene expression presents a particularly challenging signal to noise problem. Among single-cells, expression values of any particular gene measured by RNA-seq can vary drastically from cell to cell due to a combination of technical noise, gene dropout events, amplification bias, the inherent stochastic nature of transcription, as well as bone fide differences in the biological states of individual cells. To address these additional potential sources of error, we utilized a Bayesian approach to single-cell gene expression estimates. Briefly, we first performed extensive quality control using cell-specific error models to quantify differences in measurement quality of individual cells and to generate probabilistic models of gene expression based on prior expression magnitudes. Then, we corrected for batch effects, amplification outliers, drop-out (gene
Figure 6 | Transcriptional heterogeneity of human RGC assayed by single-cell RNA-seq. 

a, Workflow and strategy for FACS isolation of human RGC for single-cell RNAseq by cell surface marker expression. LeX and GLAST are used as pan-RGC markers (LG⁺) prior to single-cell capture and sequencing. WG, weeks of gestation 
b, Single-cell RNA-seq of 198 human RGC followed by weighted principal component analysis (PCA) of gene ontology (GO) pathway clusters reveals distinct transcriptional states defined by cell cycle status. Unsupervised hierarchical clustering and heatmap representation of PCA scores (left) indicates major RGC subpopulations featuring interphase cells, M-phase, and S-phase cells. At right, a schematic representation of the single RGC mapped onto the cell cycle showing cells transitioning through all cell cycle phases. 

c, Weighted PCA analysis of single RGC transcriptomes after normalizing for cell cycle GO pathways reveals apical RGC and ORG subpopulations. Unsupervised hierarchical clustering and heatmap representation of PCA scores (left) reveals distinct transcriptional signatures, which include differential expression of known apical RGC markers (e.g. TJP1, MMP5, PARD3, GFAP, VIM) and ORG-enriched genes (e.g. PPP1R17, PLCB4, NEUROD4, NEUROD1, SSTR2, TTYH2) shown at right.
Figure 6 (Continued)
Figure 6 (Continued)

198 LG+ Single-cell Transcriptomes

M Phase Cells

S Phase Cells

Interphase Cells

Apical RGC

GO Pathways

Gene sets

Transitioning Cells?

Apical RGC

GO Pathways

Gene sets

ORG

198 LG+ Single-cell Transcriptomes

Apical RGC
detection failures), and gene lengths in order generate robust measurements of gene expression in single-cells (Kharchenko et al., 2014).

Of 215 sequenced single cells, 198 transcriptomes passed quality filtering. Instead of using individual gene expression measurements to drive unsupervised hierarchal clustering, we utilized a gene-set and pathways based approach. Individual gene measurements, which often vary on orders of magnitude from cell to cell, can lead to uninformative or potentially misleading clustering of cells. Thus, we hypothesized that bone fide transcriptional states would result from coordinated expression of multiple genes within a given pathway. Thus, we used gene ontology (GO) and pre-defined functional pathways, such as KEGG pathways, to drive clustering of single-cell transcriptomes using weighted principal component analysis (PCA) (Bailey, 2012). For each pre-defined GO or KEGG pathway, we first tested whether the amount of variance explained by the first few principal components significantly exceeded genome-wide background expectation and then performed unsupervised hierarchical clustering to define RGC heterogeneity. Using this method, we found that cell cycle pathways are the most prominent drivers of RGC subpopulations with interphase cells distinctly separating away from M and S phase pathway expressing cells (Fig 6b). Importantly, while our pathway based approach revealed distinct transcriptional states of human RGC, hierarchical clustering by ad hoc examination of individual genes previously described in cortical development such as areal specification genes (e.g. FGF receptors, COUP-TF, etc.) and progenitor identity genes (e.g. PAX6, EMX, etc.) did not robustly drive distinct clustering of single RGC transcriptomes (data not shown).
To probe RGC heterogeneity beyond cell cycle status, we normalized and removed cell cycle-related GO pathways and repeated the analysis. Interestingly, we found that hierarchical clustering of cells, after normalizing for cell cycle, identifies three subpopulations of RGC, two of which feature transcriptional signatures of apical RGC and ORG (Fig 6c). Within the apical RGC subpopulation, cells express apical complex genes (e.g. TJP1, MMP5, PARD3) and the RGC markers GFAP and VIM (Fig 6c top right). On the other hand, within the ORG subpopulation, cells express several genes identified as ORG-enriched by bulk RNA-seq (Chapter 2) including PPP1R17, PLCB4, NEUROD4, NEUROD1, SSTR2, and TTHY2 (Fig 6c bottom right). Taken together, these data suggest that aside from cell cycle status, apical-basal positioning is likely one of the major drivers of RGC heterogeneity in the developing human cortex.

CONCLUSIONS

Overall, our data show that human radial glial progenitors, and to a lesser extent those of the gyrencephalic ferret, differ most strikingly from mouse RGC in the “gradedness” of their transition from NEUROG2-negative 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.
EXPERIMENTAL 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 two to four hours following release from clinical pathology. 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-mm³ 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” media (Neurobasal media, 0.25% HEPES, 0.5% FBS, rhEGF, rhFGF) for labeling with cell surface antibodies. Cells were labeled in aliquots of 500ul 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 minutes at +4°C and washed twice with pre-sort media before FACS. Alternatively, minced tissue was cryopreserved prior to enzymatic dissociation by storing in HBSS + 10% DMSO, cooled gradually in a cryochamber to -80°C overnight, and transferred to -150°C for long-term storage.
Bulk RNA Isolation, Processing, and RNA 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) (Trapnell et al., 2012) using the hg19 genome and UCSC KnownGene transcriptome references. Additional R/Bioconductor packages were used for principle component analysis, clustering, and the generation of heatmaps. 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/).

Single Cell mRNA Expression Profiling. Following 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 spun down before pre-amplification (94°C 10 minutes, 50°C 60 minutes, 94°C 30 seconds, 50°C 3 minutes x 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 was 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.

**Single Cell RNA-seq and Data Analysis.** Following cell labeling, cells were sorted by FACS into “pre-sort” media (Neurobasal media, 0.25% HEPES, 0.5% FBS, rhEGF, rhFGF). Cell capture was performed using the Fluidigm C1 platform according to manufacturer’s instructions. cDNA was generated using SMARTer low input RNA kit (Clonetech) and next-generation sequencing libraries were prepared using Illumina TruSeq v2 according to manufacturer’s instructions followed by sequencing on an Illumina HighSeq 2000. Sequencing reads were mapped using the hg19 genome and UCSC KnownGene transcriptome references followed by quality control using cell-specific error models to quantify differences in measurement quality of individual cells and then corrected for batch effects, amplification outliers, gene drop-out, and gene lengths (Kharchenko et al., 2014). Weighted PCA and unsupervised hierarchical clustering was performed using R/Bioconductor packages and custom scripts which will be made available upon publication.
REFERENCES


Chapter 4

Discussion
Summary

This dissertation research aimed to further characterize the transcriptional programs of different human cortical progenitor subtypes with emphasis on recently discovered ORG progenitors. Using a combined FACS enrichment and transcriptional profiling strategy, we identified a molecular signature of human ORG comprising hundreds of known genes and novel transcripts. Most notably, we observed a highly significant enrichment of a well-known transcription factor network, regulated by the critical regulatory factor NEUROG2, in ORG, and used ferrets to confirm the role of this transcription factor network in regulating key steps in ORG production, specifically delamination from the ventricular neuroepithelium and migration into the SVZ. 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 novel 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 novel gene loci, putatively encoding IncRNAs, including several loci with enriched expression in human ORG. Several of these ORG-enriched IncRNA 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. Altogether, our population level and single-cell transcriptional data intriguingly show a correlation between mature cortical size and structure and the heterogeneity of the progenitors that create this structure during development.

Transitions and lineage commitment of neural progenitors

Previous studies have assayed the molecular heterogeneity of ferret, macaque, and human neural progenitors by multi-channel immunofluorescence and have found that, depending on the species, the age examined, and the RGC markers tested, between ~20% and 60% of non-apical RGC co-express the rodent IP-enriched transcription factor TBR2 (Fietz et al., 2010; Hansen et al., 2014; Martínez-Cerdeño et al., 2012). TBR2+ RGC are much more abundant in species with an expanded OSVZ, consistent with the paucity of ORG in the mouse. However, in each case, although the proportion of TBR2+ RGC was greater in the OSVZ, double-positive progenitors were nonetheless detected in all germinal zones, suggesting that regardless of apical-basal location, a subset of RGC express markers of commitment to a neurogenic lineage.

Here, we expanded this characterization beyond the three or four channels available by traditional immunofluorescence staining. We observe at least four distinct subtypes of human RGC based on single-cell transcriptional signatures: apical multipotent; apical proneural; non-apical multipotent; and non-apical proneural. Importantly, this “proneural” subtype, as defined by expression of the classic marker of neuronal lineage commitment TBR2 and its direct upstream transcriptional activator NEUROG2, can be further subdivided according to co-expression of additional genes including the
downstream proneural factors *NEUROD1* and *NEUROD4*. We further demonstrate that the most proneural subtype is more prevalent in human RGC compared to embryonic mouse cortex at a comparable developmental stage, and that several additional factors co-expressed in subsets of human RGC are virtually absent from mouse progenitors. Taken together, these data point to both quantitative and qualitative species differences in the transcriptional programs regulating NPC diversity and function. Importantly, this increase in multipotent neuronal lineage committed progenitors (*i.e.* ORG) might serve to multiply neurogenic output during development and thus may have been critical for the expansion of human cortical size.

**Functional heterogeneity of cortical progenitors**

In contrast to the classical model of temporally induced fate-restriction of RGC, recent studies have provided evidence both for and against functional heterogeneity within mouse RGC, particularly with respect to the transcription factors *CUX2* and *FEZF2*, respectively (Franco et al., 2012; Guo et al., 2013). In particular, Franco et al. have suggested that *CUX2* may mark a distinct population of lineage-restricted RGC that remain largely quiescent during early neurogenesis and only produce upper-layer neurons. Since these studies were both conducted in mice, the existence of these or other populations of fate-restricted RGC in the developing human cortex is largely unknown. Our human RGC subtype-specific RNA-seq data confirm that *FEZF2* is highly enriched in both LG\(^+\)Pr\(^{hi}\) and LG\(^+\)Pr\(^{lo}\) RGC subpopulations relative to LG\(^-\) neurons. In contrast, *CUX2* shows a highly significant (*p = 1.33E\(^{-85}\)\), >50-fold enrichment in LG\(^-\) neurons relative to LG\(^+\)Pr\(^{hi}\) apical RGC, with a more modest but still highly significant (*p
= 1.86E−32), ~8-fold enrichment in LG^+Pr^lo ORG relative to LG^+Pr^hi 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 Fig. 4; whereas CUX2 is most likely enriched in the NEUROG2+ proneural subsets, consistent with this factor’s role in upper layer neuronal morphogenesis. However, it is important to note that the human specimens available for our studies were from the second half of the second trimester (16 to 21 WG), during the later stages of upper-layer neurogenesis. Thus, to fully investigate potential lineage-restricted RGC analogous to the CUX2+ population in the mouse, we would require earlier human fetal cortical samples to specifically contrast the expression of these transcription factors in early versus late human radial glial progenitors and also functional experiments in human cortical cells such as genetic lineage tracing.

In addition to size and surface area expansion, the human cortex also exhibits dramatically increased and more complex cortical functional areas, particularly within prefrontal regions. Due to the scarcity of human fetal tissue suitable for transcriptional studies, many of the genes that are required for the formation of these functional areas are either unknown or insufficiently studied. Unfortunately, due to mechanical damage, tissue samples collected in this dissertation were generally of insufficient quality to allow for regional or areal identification. Furthermore, none of the samples were sufficiently intact to allow for microdissection of distinct cortical areas prior to processing. Theoretically, single-cell transcriptome profiling might overcome these limitations by allowing for unbiased clustering of cells into groups representing distinct functional areas. However, thus far, our data has been unable to clearly distinguish these clusters
likely due to a combination of factors. Few individual genes and pathways have been clearly shown in human cortical progenitors to specify regional or areal identity thus rendering hierarchical clustering based on these transcripts largely uninformative. Furthermore, while our sample of 198 single RGC transcriptomes is large compared to previously published datasets, it nonetheless represents an extremely small sample of all cortical progenitors in the brain, which likely number in the millions or tens of millions of cells. Thus, de novo discovery of biologically informative progenitor clusters may indeed require transcriptional profiling of many thousands or tens of thousands of single-cells, which at present is technically challenging and often cost prohibitive. Indeed, new advances in single-cell technologies and sequencing may be required before these challenges may be adequately addressed. Finally, analytical tools such as transcriptional motif and network analysis must be developed or adapted to single-cell analysis to properly reconstruct either known or potentially novel aspects population substructures.

**Cell type specific transcriptional profiling of cortical cells**  
Two previous studies have reported gene expression profiles of the human VZ, ISVZ, and OSVZ, using laser capture-assisted microdissection to separate the germinal zones from each other and from the postmitotic IZ and CP compartments, with one of these studies also directly contrasting human and mouse germinal zones, and another recent study explored the transcriptional signature of human RGC and the differences in gene expression between human and mouse progenitors (Fietz et al., 2012; Lui et al., 2014; Miller et al., 2014). The genes reported by these studies as OSVZ-enriched or
human RGC-enriched we find to be expressed either in apical RGC or in both apical
RGC as well as ORG (Supplementary Table 2), consistent with their being radial glial
markers. On the other hand, few of the ORG-enriched genes we find by our methods
were captured by previous studies, highlighting the ability of sorted cell populations to
reveal cell type-specific expression patterns. One prior transcriptional analysis that also
identified a number of the human ORG-enriched genes found in our study was
performed on single progenitors from the embryonic mouse cortex (Supplementary
Table 2) (Kawaguchi et al., 2008). Notably, those authors showed by in situ
hybridization that a number of 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. suggests that
the embryonic mouse cortex may have an analogous cell type to the human ORG, but
which differs in both 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+ multipolar IP (Kawaguchi et al., 2008). Finally, it should be noted that
during the course of this dissertation, a separate research group conducted similar
experiments to purify human and mouse cortical progenitors, including ORG, for
comparative transcriptome analysis. Their transcriptome analysis was presented in a
coordinated publication and largely validates our findings (Florio et al., 2015).
Species-specific differences in gene expression

With the completion of the Human Genome Project in 2003, we now have a blueprint of all nucleotides and genes contained within the human genome. It can be argued that this information literally represents what makes us human, yet has still failed to explain much of our unique human features and phenotypes, particularly with respect to the brain. Thus, the answer does not lie in the sequence of human DNA alone. After all, we share a tremendous amount of sequence conservation at both the amino acid and nucleotide levels with chimpanzees, an observation that dates back decades and has been confirmed by modern genome sequencing studies (Chimpanzee Sequencing and Analysis Consortium, 2005; Varki and Altheide, 2005). In their influential 1975 paper, King and Wilson discuss this very question and concluded that changes in gene expression and transcriptional regulation are likely the key to the evolution of unique human features (King and Wilson, 1975).

According to this hypothesis, small differences in the time of activation or in the level activity of a single gene could in principle influence considerably the systems controlling embryonic development... Most important for the future study of human evolution would be the demonstration of differences between apes and humans in the timing of gene expression during development, particularly during the development of adaptively crucial organ systems such as the brain.

King and Wilson, 1975

This underscores the importance of finding the unique transcriptional programs of human cells. It is unlikely that King and Wilson would have foreseen the saltatory changes in technology that now enable us to routinely assay gene expression across the entire genomes of thousands of cells. It is also unlikely that they would have predicted our newfound appreciation for novel players in the regulation of gene expression including short and long noncoding RNAs, tissue-specific alternative
splicing, and epigenetics (Clark et al., 2007; Iyer et al., 2015; Pan et al., 2008). Nonetheless, the challenge remains the same. Understanding how evolutionary forces have given rise to the tremendous diversity of cell types and connections within the human brain will require continued comparative spatiotemporal transcriptomic studies at the cell type or subtype level followed by functional studies of these unique transcriptional signatures. Along those lines, here we have shown that pronounced differences in transcriptional heterogeneity exist between human and mouse cortical progenitors and also identify many IncRNAs that exist in human but not mouse RGC.

**Implications on disease**

Disruptions of cortical development and gyrification can result in a number of developmental disorders including lissencephaly (thickened cortex with simplified or absent gyrification), polymicrogyria (overproduction of small, fused gyri), and microcephaly (head circumference more than two standard deviations below average) (Barkovich et al., 2012; Manzini and Walsh, 2011). Genetic cortical malformation disorders tend to be individually rare, and thus their incidence is difficult to estimate. Collectively, however, these disorders are a significant source of serious pediatric health problems. In one study, the incidence of cortical malformations at a pediatric neurology outpatient clinic was approximately eight per 1,000 new patients, with the majority of cases exhibiting intellectual disability and/or seizures (de Wit et al., 2008). Other studies estimate at least 15% of childhood epilepsy, and a similar proportion of intellectual disability, can be attributable to cortical dysplasia, including gyrification disorders (Barkovich et al., 2012; de Wit et al., 2008; Walsh, 1999). Many causative
mutations for cortical malformations have been mapped to genes critical for cortical progenitor function including cell cycle regulation, centrosome function, epigenetic regulation, and early neuronal migration (Gleeson et al., 1999; Manzini and Walsh, 2011; Walsh and Engle, 2010; Yang et al., 2012).

Our understanding of cortical malformation disorders has often been hampered by the fact that, in many cases, rodent models fail to robustly recapitulate the human disease phenotypes. In some cases, this is due to fundamental differences between rodent and human cortical development. For example, the mouse is naturally smooth brained and therefore cannot adequately model human lissencephaly disorders. In other cases, such as microcephaly, the lack of abundant neurogenic ORG in mouse compared to human might explain the differences in phenotypes. Nonetheless, the exact cellular and functional contribution of ORG to human cortical malformation disorders is still unknown. As a first step towards this goal, we have provided subtype specific transcriptional profiling of human cortical progenitors, which will serve as a resource for future studies of cortical malformation. Furthermore, we have demonstrated that the ferret can be a powerful experimental model for studying both normal and abnormal cortical development using gene perturbation by transient transfection of cortical progenitors. Going forward, the advent of highly efficient genome editing technologies such as TALENs, Zinc finger nucleases (ZFNs), and the CRISPR-Cas9 system will allow for more specific and efficient genetic studies of cortical development in the ferret. Using these technologies, it will now be possible to study loss of function, gain of function, or patient-specific mutations of cortical malformation genes in an animal model with a relatively large cortex, abundant ORG, and gyrification.
REFERENCES


Appendix

Supplementary Data and Tables
Adapted from a manuscript in Nature Neuroscience (March 3, 2015)

Single Cell Analysis Reveals Transcriptional Heterogeneity of Neural Progenitors in the Human Cortex

Matthew B. Johnson¹, Peter P. Wang¹,*, Kutay D. Atabay¹, Elisabeth A. Murphy¹, Ryan N. Doan¹, Jonathan L. Hecht², Christopher A. Walsh¹,³

¹Division of Genetics and Genomics, Manton Center for Orphan Disease Research, and Howard Hughes Medical Institute, Boston Children’s Hospital, Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA, and Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

²Department of Pathology, Beth Israel Deaconess Medical Center and Department of Pathology, Harvard Medical School, Boston, MA

*These authors contributed equally to this work.

³Address for correspondence:
Christopher A. Walsh
Division of Genetics and Genomics
Boston Children’s Hospital
Center for Life Sciences 15062
300 Longwood Avenue, BCH3150
Boston, MA 02115
christopher.walsh@childrens.harvard.edu
AUTHOR CONTRIBUTIONS

Matthew B. Johnson was an equally contributing first author on this manuscript. Together, we designed and conducted all experiments as well as analyzed and interpreted the data. Kutay D. Atabay assisted with developing the cell sorting strategy, validation of cell sorting populations by immunofluorescence, and postnatal ferret surgical experiments. Elisabeth A. Murphy assisted with postnatal ferret surgical experiments and immunostaining analysis. Ryan N. Doan performed the evolutionary analysis of noncoding RNA genes. Jonathan L. Hecht procured and examined human tissue samples. Christopher A. Walsh is the corresponding author and oversaw and guided all experimentation. Matthew B. Johnson, Christopher A. Walsh, and myself wrote the manuscript.
Supplementary Figure 1 related to Figure 1 | Isolation of human RGC subpopulations by FACS. a, Both LG^Pr^hi and LG^Pr^lo subpopulations are enriched for known RGC-expressed genes (GFAP, VIM, GLAST, PAX6, SOX2, BLBP), and depleted for neuronal markers (DCX, TUJ1, NeuN, MEF2C). The LG^Pr^hi subpopulation was enriched relative to the LG^Pr^lo subpopulation for PROM1 transcript as well as three other transcripts encoding apical membrane domain-specific proteins (PARD3 [Par3], TJP1 [ZO-1], MPP5 [Pals]). Data represents four biological replicates (mean ± SEM) ranging from 16 WG to 23 WG. b, Primary neurospheres derived from LeX^- and LeX^+ cells sorted from dissociated human fetal cortex. Neurospheres were serially passaged at clonal density and immunolabeled for RGC marker SOX2.
**Supplementary Figure 2**

**Supplementary Figure 2 related to Figure 1 | Gene set enrichment in human RGC subpopulations.** Gene set enrichment analysis confirmed the RGC progenitor nature of both the LG^Pr^hi and LG^Pr^lo subpopulations, with enrichment of important progenitor signaling pathways (e.g. Wnt/Bmp/Tgf) and gene ontology terms (cell cycle control, neural development) in both subpopulations relative to LG^Pr^- neurons and other cell types.
Supplementary Figure 3 related to Figures 1 and 4 | Selected LG^Pr^b^-enriched candidate non-apical RGC genes validated by qRT-PCR in independent biological replicates of FACS-purified human fetal RGC. Relative expression levels in the LG^Pr^b^-subpopulation compared to LG^Pr^h^-after normalization to housekeeping genes ACTB and GAPDH. Data represents six biological replicates (mean ± SEM) ranging from 16 WG to 23 WG (asterisk denotes p < 0.05, paired t-test; n=6, max p=0.045, all others were lower).
Supplementary Figure 4 related to Figure 2 | Upregulation of proneural neurogenin targets in NEUROG2-VP16 electroporated ferret cells. In vivo delivery of GFP control and NEUROG2-VP16 constructs to ferret apical RGCs was performed by intraventricular injection and electroporation in neonatal ferret kits (n=2 per condition at postnatal day 1) as described in Figure 2. After 48 hours post-electroporation, electroporated cells were isolated for qRT-PCR analysis by enzymatic dissociation and FACS using their GFP fluorescence. Relative to GFP+ control electroporated cells, NEUROG2-VP16 expressing cells showed upregulation of many previously described NEUROG2 effector genes including Cbfa2t2, Foxn2, Foxp2, Hes6, Myt1, Neurod1, Neurod4, Neurog1, and Nhlih1, and down-regulation of Sox2. In addition, we also tested expression of ferret orthologs of human ORG-enriched genes and found that several including Gadd45g, Ttyh2, Sstr2, and Plcb4 were also upregulated in NEUROG2-VP16 cells compared to controls.
Supplementary Figure 5 related to Figure 3 | Differential expression of novel unannotated IncRNAs in human RGC subtypes. RNA-seq reads displayed in genomic context for the LG+Pr<sup>hi</sup> apical RGC (red), LG+Pr<sup>lo</sup> ORG (green), and LG–Pr– cells (black). Novel transcripts assembled from the RNA-seq data are shown in blue, and previously catalogued IncRNA transcripts are shown in brown (Cabili et al., 2011).

a, Two intergenic IncRNAs on chromosome 2 with distinct expression patterns in the human fetal cortex share a bidirectional promoter and overlap at their 5’ ends. The plus-strand IncRNA is enriched in apical RGC, whereas the minus-strand IncRNA is relatively enriched in ORG and neurons. Blue boxed region highlights the overlapping transcription start sites (TSS), and is enlarged below. Black arrows indicate read peaks from each strand’s TSS. Bottom part of (a) shows the promoter at higher magnification, with expression levels of the two IncRNAs (in FPKM) plotted at right. b, Example of an ORG-enriched IncRNA. Multiple alternatively spliced isoforms of this multi-exon locus are expressed in all cell types assayed, but are significantly enriched in the LG+Pr<sup>lo</sup> non-apical subpopulation. A partial transcript overlapping the 5’ end of the locus was previously detected by ultra-high depth RNA sequencing; our data demonstrate that even low-abundance transcripts can be captured and fully reconstructed from an order of magnitude fewer reads when RNA is sequenced from the specific cell types that express the gene, rather than from heterogeneous bulk tissue (Cabili et al., 2011). c, Example of a novel apical RGC-specific intergenic transcript not detected by previous deep-sequencing experiments.
Supplementary Figure 5 (Continued)
Supplementary Figure 6 related to Figure 3 | Differential enrichment of lncRNAs in human and mouse RGC populations. We performed qRT-PCR of several conserved lncRNAs in FACS-purified human (n=4 biological replicates ranging from 16 WG to 23 WG) and mouse RGC populations (n=3 from E15.5) comparing human ORG (LG+PrLo) and apical RGC (LG+PrHi) with neurons (LG-Pr-) and mouse RGC (L+Pr+) with neurons (L-Pr-) (mean ± SEM). We find that several conserved lncRNAs including LINC-PINT, TUNAR, CRNDE, MIR22HG are enriched in human RGC progenitor populations but depleted in mouse RGC suggesting potentially divergent roles in human radial progenitor evolution and function.
Supplementary Figure 7 related to Figure 4 | Single-cell expression profiles of human and mouse RGC. a, Violin plots of RGC marker gene expression in human and mouse single sorted RGC reveals largely similar pattern of gene expression for RGC markers including SOX2, VIM, GLAST, BLBP, PAX6, NES. Interestingly, significant numbers of human RGC express GFAP and DCX but these genes are nearly absent in mouse RGC. b, Principle component analysis of 546 human (left) and 226 mouse (right) single RGC indicates distinct distributions of transcriptional states in human compared to mouse RGC. Here, “apical” is defined by expression of at least two of the four apical complex marker transcripts, and “proneural” by expression of at least two of the four Neurogenin pathway genes. In both species, the first PC (x-axis) reflects the proneural+/− dimension, with “multipotent” (presumptively pre-Neurogenin-pathway-expressing) RGC tending towards the left (red and blue cells) and proneural RGC on the right (black and green cells). Human cortex contains a greater proportion of proneural RGC, whereas mouse has fewer proneural cells which are less distinct, as indicated by the greater overlap of black and red cells in the mouse. In addition, human cortex displays far more non-apical (blue and green) cells than mouse, which again are more distinct from the apical (red and black) cells along the second PC (y-axis). In contrast, mouse non-apical RGC (blue and green) are scarce and not transcriptionally distinct from apical cells, as indicated by the lack of separation along the y-axis.
## Supplementary Table 1

**Supplementary Table 1 | List of human fetal specimens collected, gestational ages, and studies performed.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Age (WG)</th>
<th>Studies Performed</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB018</td>
<td>23</td>
<td>Candidate validation (qRT-PCR)</td>
</tr>
<tr>
<td>FB044</td>
<td>16</td>
<td>Candidate validation (qRT-PCR)</td>
</tr>
<tr>
<td>FB025</td>
<td>23</td>
<td>Candidate validation (qRT-PCR)</td>
</tr>
<tr>
<td>FB035</td>
<td>18</td>
<td>RNA-seq</td>
</tr>
<tr>
<td>FB036</td>
<td>18</td>
<td>Neurosphere culturing</td>
</tr>
<tr>
<td>FB031</td>
<td>19</td>
<td>Candidate validation (qRT-PCR)</td>
</tr>
<tr>
<td>FB032</td>
<td>18</td>
<td>RNA-seq</td>
</tr>
<tr>
<td>FB033</td>
<td>19</td>
<td>RNA-seq</td>
</tr>
<tr>
<td>FB040</td>
<td>20</td>
<td>Candidate validation (qRT-PCR), Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB043</td>
<td>17</td>
<td>Candidate validation (qRT-PCR), Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB049</td>
<td>20</td>
<td>Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB049</td>
<td>16</td>
<td>Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB053</td>
<td>21</td>
<td>Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB066</td>
<td>18</td>
<td>Single-cell expression profiling (Biomark)</td>
</tr>
<tr>
<td>FB068</td>
<td>20</td>
<td>Single-cell expression profiling (Biomark)</td>
</tr>
</tbody>
</table>
Supplementary Table 2 | Comparison of the present RNA-seq data to previous transcriptome studies of fetal cortical progenitors.

Several recent studies have examined the transcriptional signature of human fetal germinal zones using manually or laser capture-assisted microdissection techniques (Fietz et al., 2012; Lui et al., 2014; Miller et al., 2014). Here we list the central findings of several of these papers and the differential expression patterns we observe for these genes using our FACS-RNA-seq strategy (middle column). We also report the expression levels (fpkm) of these genes in ferret RGC (right column), noting that many genes are conserved in their progenitor expression but some, such as PDGFD, are not. Most notable in this analysis is the absence of ORG-enriched genes from previous transcriptome assays of the human OSVZ where most ORG are located (Lui et al., 2014; Miller et al., 2014). We attribute this discrepancy to the highly heterogeneous cellular composition of the OSVZ, which in addition to ORG harbors multipolar intermediate progenitors, radially migrating postmitotic neurons generated in both the VZ and SVZ, and tangentially migrating interneurons originated from the ganglionic eminences. Furthermore, our single-cell data demonstrate additional transcriptional heterogeneity even within ORG, which further confounds efforts to profile these cells from bulk tissue samples. Thus none of the ORG-enriched genes identified in our current study have previously been associated with this cell type by other methods. Remarkably, however, at least 8 genes that we found as having significant or trending human ORG enrichment were previously described in a single-cell expression microarray analysis of the embryonic mouse cortex (genes marked in red bold text in the bottom section) (Kawaguchi et al., 2008). These authors also showed by in situ hybridization in E14 mouse cortex that several human ORG-enriched genes are expressed in a narrow band of cells at the VZ-SVZ border in mouse, in contrast to the OSVZ location of most ORG in human and other ORG-abundant species. We interpret these results as indicating that some human ORG-enriched genes are also expressed in mouse progenitors during the transition from VZ RGC to SVZ IP, as has been clearly demonstrated for Neurog2, thus further supporting our conclusion that the ORG transcriptional signature reflects a transitional developmental state (Ochiai et al., 2009).
### Supplementary Table 2 (Continued)

#### Genes expressed in human but not mouse RGC
*(Lui, Nowakowski et al., 2014)*

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Human Progenitor Expression Pattern (FACS-RNA-seq)</th>
<th>Ferret RGC Expression Level (fpkm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABHD3</td>
<td>pan-RGC enriched</td>
<td>0.8</td>
</tr>
<tr>
<td>ASAP3</td>
<td>pan-RGC enriched</td>
<td>14.4</td>
</tr>
<tr>
<td>BMP7</td>
<td>pan-RGC enriched</td>
<td>25.7</td>
</tr>
<tr>
<td>C5</td>
<td>pan-RGC enriched</td>
<td>0.2</td>
</tr>
<tr>
<td>C8orf4</td>
<td>pan-RGC enriched</td>
<td>22.9</td>
</tr>
<tr>
<td>FAM107A</td>
<td>pan-RGC enriched</td>
<td>1.9</td>
</tr>
<tr>
<td>FOXN4</td>
<td>pan-RGC enriched</td>
<td>9.7</td>
</tr>
<tr>
<td>ITGA2</td>
<td>apical RGC &gt; ORG</td>
<td>30.2</td>
</tr>
<tr>
<td>LRIG3</td>
<td>pan-RGC enriched</td>
<td>7.1</td>
</tr>
<tr>
<td>LRRC17</td>
<td>pan-RGC enriched</td>
<td>0.4</td>
</tr>
<tr>
<td>PAM</td>
<td>pan-RGC enriched</td>
<td>22.9</td>
</tr>
<tr>
<td>PDGFD</td>
<td>pan-RGC enriched</td>
<td>0.1</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>apical RGC &gt; ORG</td>
<td>14.9</td>
</tr>
<tr>
<td>PDLIM3</td>
<td>pan-RGC enriched</td>
<td>17.7</td>
</tr>
<tr>
<td>RFTN2</td>
<td>pan-RGC enriched</td>
<td>43.0</td>
</tr>
<tr>
<td>SLC2A10</td>
<td>pan-RGC enriched</td>
<td>3.1</td>
</tr>
<tr>
<td>SP110</td>
<td>pan-RGC enriched</td>
<td>0.3</td>
</tr>
<tr>
<td>STOX1</td>
<td>pan-RGC enriched</td>
<td>20.8</td>
</tr>
<tr>
<td>ZC3HAV1</td>
<td>pan-RGC enriched</td>
<td>2.8</td>
</tr>
</tbody>
</table>

*Human OSVZ-enriched genes
*(Miller, Ding et al., 2014)*

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Human Progenitor Expression Pattern (FACS-RNA-seq)</th>
<th>Ferret RGC Expression Level (fpkm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRP3</td>
<td>neuron-enriched</td>
<td>34.7</td>
</tr>
<tr>
<td>HNRNPA3</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>HNRNPH3</td>
<td>no significant difference</td>
<td>151.3</td>
</tr>
<tr>
<td>MT1F</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>MT1G</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>MT1H</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>POLR2J2</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>PSMC3IP</td>
<td>no significant difference</td>
<td>30.7</td>
</tr>
</tbody>
</table>

*Human OSVZ-enriched genes
*(Fietz et al., 2012)*

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Human Progenitor Expression Pattern (FACS-RNA-seq)</th>
<th>Ferret RGC Expression Level (fpkm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankrd23</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>C1orf111</td>
<td>neuron-enriched</td>
<td>0.5</td>
</tr>
</tbody>
</table>

*n.a., not annotated*
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Human Progenitor Expression Pattern (FACS-RNA-seq)</th>
<th>Ferret RGC Expression Level (fpkm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerkl</td>
<td>pan-RGC enriched</td>
<td>0.6</td>
</tr>
<tr>
<td>Cldn11</td>
<td>apical RGC &gt; ORG</td>
<td>1.1</td>
</tr>
<tr>
<td>Dcbld1</td>
<td>no significant difference</td>
<td>0.0</td>
</tr>
<tr>
<td>Dclk3</td>
<td>strong_late_ON</td>
<td>0.1</td>
</tr>
<tr>
<td>Fkbp11</td>
<td>no significant difference</td>
<td>8.4</td>
</tr>
<tr>
<td>GIGY2</td>
<td>no significant difference</td>
<td>18.7</td>
</tr>
<tr>
<td>Itih4</td>
<td>no significant difference</td>
<td>0.0</td>
</tr>
<tr>
<td>KIAA1324</td>
<td>neuron-enriched</td>
<td>16.1</td>
</tr>
<tr>
<td>KIAA1324L</td>
<td>neuron-enriched</td>
<td>n.a.</td>
</tr>
<tr>
<td>Lims2</td>
<td>apical RGC &gt; ORG</td>
<td>0.7</td>
</tr>
<tr>
<td>Nipsnap1</td>
<td>no significant difference</td>
<td>42.6</td>
</tr>
<tr>
<td>Olig1</td>
<td>apical RGC &gt; ORG &gt; neurons</td>
<td>n.a.</td>
</tr>
<tr>
<td>Parp10</td>
<td>no significant difference</td>
<td>n.a.</td>
</tr>
<tr>
<td>Pcbp2</td>
<td>no significant difference</td>
<td>541.3</td>
</tr>
<tr>
<td>Prss12</td>
<td>strong_late_ON</td>
<td>11.7</td>
</tr>
<tr>
<td>Sox10</td>
<td>apical RGC &gt; ORG</td>
<td>1.0</td>
</tr>
<tr>
<td>Tmem88</td>
<td>neuron-enriched</td>
<td>n.a.</td>
</tr>
<tr>
<td>TPGS2</td>
<td>no significant difference</td>
<td>20.1</td>
</tr>
</tbody>
</table>

n.a., not annotated

**Embryonic mouse cortex single-cell profiling "cluster II/III genes"**

(Kawaguchi et al., 2008)
## Supplementary Figure 3

### Supplementary Table 3: Species used for multi-species alignment

<table>
<thead>
<tr>
<th>Common name</th>
<th>Latin binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primate subset</strong></td>
<td></td>
</tr>
<tr>
<td>Baboon</td>
<td>Papio hamadryas</td>
</tr>
<tr>
<td>Bushbaby</td>
<td>Cercopithecus aethiops</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>Pan troglodytes</td>
</tr>
<tr>
<td>Crab-eating macaque</td>
<td>Macaca fascicularis</td>
</tr>
<tr>
<td>Gibbon</td>
<td>Nomascus leucogenys</td>
</tr>
<tr>
<td>Orangutan</td>
<td>Pongo pygmaeus abelii</td>
</tr>
<tr>
<td>Rhesus</td>
<td>Macaca mulatta</td>
</tr>
<tr>
<td>Squirrel monkey</td>
<td>Saimiri boliviensis</td>
</tr>
<tr>
<td><strong>Euarchontoglires subset</strong></td>
<td></td>
</tr>
<tr>
<td>Brush-tailed rat</td>
<td>Octodon degus</td>
</tr>
<tr>
<td>Chinchilla</td>
<td>Chinchilla longica</td>
</tr>
<tr>
<td>Chinese hamster</td>
<td>Cricetulus griseus</td>
</tr>
<tr>
<td>Chinese tree shrew</td>
<td>Tupai a chinensis</td>
</tr>
<tr>
<td>Golden hamster</td>
<td>Mesocricetus auratus</td>
</tr>
<tr>
<td>Guinea pig</td>
<td>Cavia porcellus</td>
</tr>
<tr>
<td>Lesser Egyptian jerboa</td>
<td>Larus jaculus</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
</tr>
<tr>
<td>Naked mole-rat</td>
<td>Heterocephalus glaber</td>
</tr>
<tr>
<td>Pika</td>
<td>Ochotona princeps</td>
</tr>
<tr>
<td>Prairie vole</td>
<td>Microtus ochrogaster</td>
</tr>
<tr>
<td>Rabbit</td>
<td>Oryctolagus cuniculus</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>Squirrel</td>
<td>Spermophilus tridecemlineatus</td>
</tr>
<tr>
<td><strong>Laurasiatheria subset</strong></td>
<td></td>
</tr>
<tr>
<td>Alpaca</td>
<td>Vicugna pacos</td>
</tr>
<tr>
<td>Bactrian camel</td>
<td>Camelus ferus</td>
</tr>
<tr>
<td>Big brown bat</td>
<td>Eptesicus fuscus</td>
</tr>
<tr>
<td>Black flying-fox</td>
<td>Pteropus alecto</td>
</tr>
<tr>
<td>Cat</td>
<td>Felis catus</td>
</tr>
<tr>
<td>Cow</td>
<td>Bovidae</td>
</tr>
<tr>
<td>David's myotis bat</td>
<td>Myotis davidii</td>
</tr>
<tr>
<td>Dog</td>
<td>Canis lupus familiaris</td>
</tr>
<tr>
<td>Dolphin</td>
<td>Tursiops truncatus</td>
</tr>
<tr>
<td>Domestic goat</td>
<td>Capra hircus</td>
</tr>
<tr>
<td>Ferret</td>
<td>Mustela putorius furo</td>
</tr>
<tr>
<td>Hedgehog</td>
<td>Erinaceus europaeus</td>
</tr>
<tr>
<td>Horse</td>
<td>Equus caballus</td>
</tr>
<tr>
<td>Killer whale</td>
<td>Ursus arctos</td>
</tr>
<tr>
<td>Megabat</td>
<td>Pteropus vampyrus</td>
</tr>
<tr>
<td>Microbat</td>
<td>Myotis lucifugus</td>
</tr>
<tr>
<td>Pacific walrus</td>
<td>Odobenus rosmarus divergens</td>
</tr>
<tr>
<td>Panda</td>
<td>Ailurpoda melanoleuca</td>
</tr>
<tr>
<td>Pig</td>
<td>Sus scrofa</td>
</tr>
<tr>
<td>Sheep</td>
<td>Ovis aries</td>
</tr>
<tr>
<td>Shrew</td>
<td>Sorex araneus</td>
</tr>
<tr>
<td>Star-nosed mole</td>
<td>Condylura cristata</td>
</tr>
<tr>
<td>Tibetan antelope</td>
<td>Pantholops hodgsoni</td>
</tr>
<tr>
<td>Weddell seal</td>
<td>Leptonychotes weddelli</td>
</tr>
<tr>
<td>White rhinoceros</td>
<td>Ceratotherium simum</td>
</tr>
<tr>
<td><strong>Afrotheria subset</strong></td>
<td></td>
</tr>
<tr>
<td>Aardvark</td>
<td>Orycteropus afer afer</td>
</tr>
<tr>
<td>Cape elephant shrew</td>
<td>Elephas africanus</td>
</tr>
<tr>
<td>Cape golden mole</td>
<td>Chrysochloris asiatica</td>
</tr>
<tr>
<td>Elephant</td>
<td>Loxodonta africana</td>
</tr>
<tr>
<td>Manatee</td>
<td>Trichechus manatus latirostris</td>
</tr>
<tr>
<td>Tenrec</td>
<td>Echinops telfair</td>
</tr>
<tr>
<td><strong>Mammal subset</strong></td>
<td></td>
</tr>
<tr>
<td>Armadillo</td>
<td>Dasypus novemcinctus</td>
</tr>
</tbody>
</table>
REFERENCES


