From Comparative Genomics to Synthetic Biology: Using Ancestral Gene Reconstruction Approaches to Test Hypotheses Regarding Proximate Mechanisms in our Evolutionary History

by Jennifer Baker

B.A. in Anthropology, May 1991, Syracuse University
M.A. in Biology, May 1998, University of Colorado, Colorado Springs
M.Phil. in Hominid Paleobiology, May 2013, The George Washington University

A Dissertation submitted to

The Faculty of
The Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 17, 2015

Dissertation directed by

Bernard Wood
University Professor of Human Origins

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

The Columbian College of Arts and Sciences of The George Washington University certifies that Jennifer Baker has passed the Final Examination for the degree of Doctor of Philosophy as of March 17, 2015. This is the final and approved form of the dissertation

From Comparative Genomics to Synthetic Biology: Using Ancestral Gene Reconstruction Approaches to Test Hypotheses Regarding Proximate Mechanisms in our Evolutionary History

Jennifer Baker

Dissertation Research Committee:

Bernard A. Wood, University Professor of Humans Origins, Dissertation Director

Thomas Maynard, Associate Research Professor of Pharmacology & Physiology, Committee Member

Chet C. Sherwood, Professor of Anthropology, Committee Member

Dedication

This work is dedicated to my lovely family. To my parents, my sister, my husband, and my children, who have supported me as I've walked through life. Most of all this is dedicated to Jay who has walked along by my side.

Acknowledgments

I am very grateful for the many people who have helped me complete this dissertation project. First, many thanks to my dissertation committee of Dr. Brenda Bradley, Dr. Derek Wildman, Dr. Thomas Maynard, and especially Dr. Chet Sherwood for their support as I completed my dissertation. I would like to express my gratitude to my primary advisor, Dr. Bernard Wood, for his guidance and encouragement over the past five years and for providing me with the opportunity to complete my PhD. His mentorship has made this a rewarding and thoughtful journey.

I would like to thank Dr. Thomas Maynard and Beverly Karpinski Oakley at the GW Institute for Neuroscience. Dr. Maynard provided constant intellectual guidance as I moved from a simple idea to a completed research project and Beverly Oakley was an endless source of help, friendship, and advice throughout every step of this project.

I would like to thank Dr. Joseph Bielawski and Dr. Katherine Dunn, both of whom have contributed in countless ways to the success of this dissertation. Their generosity of time and knowledge ultimately moved this project toward completion.

I am very fortunate to be part of the CASHP community at GWU and have received support from many people, especially Serena Bianchi, Dr. Habiba Chirchir, Dr. Aida Gómez-Robles, Chrisandra Kufeldt, and Dr. Tetyana Duka, all of who were unreserved in their guidance, inspiration, and friendship.

My dissertation research was funded by a GWU Presidential Merit Fellowship in Hominid Paleobiology, a Sigma Xi Grant-In-Aid of Research, a National Science Foundation Doctoral Dissertation Research Improvement Grant (#1455625), and a Wenner-Gren Foundation Dissertation Fieldwork Grant (#8735).

I would like to thank my family above all else for their love, encouragement, friendship and faith.

Abstract of Dissertation

From Comparative Genomics to Synthetic Biology: Using Ancestral Gene Reconstruction Approaches to Test Hypotheses Regarding Proximate Mechanisms in our Evolutionary History

At its core human evolutionary biology seeks to answer the question of how the defining characteristics of modern humans evolved, such as large-brains, obligatory bipedal gait, extended juvenile period, and increased longevity. Traditional fossil-based research uses morphology to infer behavior and life history and only recently have researchers been able to make predictions regarding the effect of modifications to the DNA and proteins of our forbearers. Using these innovative methods we investigated the molecular evolution of a superfamily of transcription factors called the Nuclear Receptors. The patterns of sequence evolution observed in our bioinformatic analyses suggest a shift in the intensity of selection pressure occurred on *NR2C1*, a gene that plays a role early in embryonic stem cell proliferation and neuronal differentiation. Methods are now available to reconstruct ancestral DNA and its corresponding protein sequences and thus generate testable hypotheses about the functional evolution of genes on specific lineages. These methods allowed us to analyze how modifications to the modern human version of *NR2C1* affected the ability of an embryonic stem cell to remain in its proliferative state. We began by creating three different copies of our gene of interest: the human copy, the chimpanzee copy, and the ancestral copy of *NR2C1* for the inferred last common ancestor of chimpanzee and modern humans. Inserting these three different gene variants into mouse embryonic stem cells that have had NR2C1 knocked down allowed us to quantitatively analyze the transcriptional and regulatory functions of *NR2C1*.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: DISSERTATION INTRODUCTION

Background, Specific Aims & Hypotheses

Comparative evolutionary genomics

What makes us human? Philosophers contemplated answers to this question long before Huxley (1863) and Darwin (1871) postulated the evolutionary relatedness of humans and the African great apes based on similar morphological characteristics. It was a little over one hundred years later that molecular data began to lend credence to their hypothesis by investigators such as Sarich &Wilson (1967), Goodman (1983), and Sibley & Alquist (1984). This continues today as recent advances in DNA sequencing and bioinformatic software have allowed for the exponential increase in genomic data collection (Mardis, 2008; 2011) and provides a deeper understanding of the evolutionary molecular relationships between our closest living relatives and ourselves (Opazo et al., 2006; Reich et al., 2010; Jameson et al., 2011; Perelman et al., 2011).

Although there are a spectacular assortment of organisms living on the planet today, they each share exactly the same four simple building blocks of DNA- adenine (A), thymine (T), cytosine (C) and guanine (G)- that contain all of the information necessary to build an organism. Exactly which protein a given gene produces depends upon the sequence in which the four building blocks are arranged along DNA's double-helix structure (Nirenberg, 1965; Nirenberg et al., 1965). This genetic information is recorded on each of two strands- and when fresh cells are produced these strands detach and new strands are synthesized from the templates in an almost perfect manner (Watson

and Crick, 1953a; 1953b; 1953c). However, the replication is not always perfect and sometimes the wrong nucleotide is added to the strand. These substitutions are passed down in subsequent generations and because of this it is possible to determine which species share the same mutation by looking at a piece of DNA. These shared mutations allow us to generate hypotheses about the evolutionary history of a set of DNA sequences.

The field of comparative evolutionary genomics uses this evolutionary history to investigate how changes in the structure and content of genomes have contributed to the evolution of life on Earth. Comparing the genomic sequences from different species with the complete human reference sequence enables the identification of regions of similarities and differences and provides a powerful tool for studying the structure and function of genes, genetic architecture, and evolutionary adaptations among organisms. Applying statistical models of evolution to these comparative datasets enables the identification of conserved features that have been preserved for millions of years (Ward and Kellis, 2012; Alföldi and Lindblad-Toh, 2013). These models also allow the identification of genes that have seen a shift in selection pressure and such genes may be responsible for the evolution of the unique aspects of the modern human phenotype (Sabeti et al., 2002; Nielsen, R. et al., 2005; Sabeti et al., 2006; Akey, 2009; Sabeti, 2011; Enard et al., 2014; Wilde et al., 2014). We used these techniques to investigate the molecular evolution of an important and well-studied group of proteins directly involved with numerous physiological functions, the nuclear receptors.

The molecular evolution of the nuclear receptors

Chapter 2 describes our use of comparative genomic methods to investigate the 48 human nuclear receptors (NRs). The NRs act as transcription factors that regulate gene expression and are involved in numerous aspects of development, behavior, endocrine signaling, and reproduction (Robinson-Rechavi et al., 2001; Bain et al., 2007).

Using previously published phylogenetic trees based on morphological (Lockwood et al., 2004; Diogo and Wood, 2011) and genetic (Ruvolo, 1997; Wildman et al., 2003) data, we used codon models (Yang, 1997; Yang, 2007), to investigate the direction and intensity of natural selection acting on DNA and amino acid alignments, estimate the optimal parameters of the sequence alignments, and thus suggest genes for expanded sampling or further analysis with ancestral reconstruction (AGR) studies. The advantage of using a codon model over a DNA or amino acid model was demonstrated in the 1970s and 1980s when researchers showed that the rate of nucleotide substitutions in a protein that results in a change in an amino acid (i.e., nonsynonymous, or dN), can be compared with the rate of silent (i.e., synonymous, or dS) changes (Miyata et al., 1979; Czelusniak et al., 1982). The ratio of these changes (dN/dS, and abbreviated ω [omega]) provides a quantifiable signal of the direction and strength of selective pressure applied to the coding region of the gene (Li et al., 1985). For example, when the number of nonsynonymous amino acid substitutions equals the number of synonymous amino acid substitutions ($\omega = 1$) it indicates that evolution is neutral. Most of the time natural selection results in the removal of deleterious mutations and $\omega < 1$, which suggests purifying selection. When $\omega > 1$ it suggests that positive selection has occurred in the evolutionary history of the gene (Yang and Bielawski, 2000).

Using this methodology the patterns of sequence evolution in nuclear receptors were investigated within two primate clades (human-chimpanzee and the great apes), and on three specific primate branches: the branch leading to the great apes, to the human-chimpanzee clade, and to the human lineage, with the intention of identifying genes that would be good candidates for ancestral gene reconstruction (AGR).

NR2C1 expression patterns during mouse embryonic development

Chapter 3 provides a detailed embryonic characterization of *Nr2c1*, the gene we chose for an AGR study. This was the logical choice for laboratory analyses, not only because its patterns of sequence change suggest there was a shift in evolutionary pressure on the human lineage and the human-chimpanzee clade, but also because of its purported biological roles. Studies indicate *Nr2c1* plays an important role in maintaining pluripotentiality in stem cells as it was shown to regulate several genes (e.g., *Oct4* and *Nanog*) associated with this function (Park et al., 2007) and it is associated with neural differentiation during brain development (Shyr et al., 2009).

Nr2c1 belongs to a subtype of NRs referred to as orphan receptors for which the endogenous ligand has yet to be identified (Lee and Chang, 1995). Because the orphan receptors are the most ancient of the NRs, they likely play a role in early embryonic development as well as in the differentiation of embryonic cells (Laudet, 1997; Enmark and Gustafsson, 2013), however the precise functions performed by this gene are not yet fully understood and thus far detailed expression patterns of *Nr2c1* in the mouse head during early embryonic development have not been reported.

Chapter 3 provides spatial and temporal expression patterns of *Nr2c1* for embryonic stem cells (ESC), as well as from embryonic day (E) 9.5 to 17.5, a longer time range from that previously reported (Young et al., 1998), with the goal of improving our understanding of the biological activity of this gene. In order to provide an overview of *Nr2c1* expression at the earliest stages of neurogenesis, we used whole mount immunohistochemistry (WMIHC) to investigate E9.5 and E10.5 embryos. Next, the structural expression of *Nr2c1* in the developing brain was established by using semi-quantitative qPCR to measure transcript levels in undifferentiated ES cells as a comparative standard, in whole embryos at E8.5 and E9.5, as well as in microdissected regions from E8.5-E10.5 embryos. The results of these two assays demonstrated that *Nr2c1* expression is robust in the embryonic OE, so we used fluorescent IHC to look more thoroughly at the spatial and temporal expression of this protein.

Chapter 3 details the results of these expression analyses in the developing mouse at time periods that correspond to neural proliferation in modern humans (Reichert, 2009). We used the mouse as a model organism because of its similarity to modern humans in genetics, anatomy, and physiology (Chen et al., 2012) and because it allowed for a careful evaluation of temporal and spatial specificity of *Nr2c1* during neurogenesis, something that cannot be done using modern humans or other primates. These experiments examined whether *Nr2c1* is present in two specific well-characterized neural stem cell populations (i.e., telencephalon and the olfactory epithelium [OE]) in the developing mouse embryo at developmental stages that correspond to when neural proliferation occurs in modern human neurogenesis (Downs and Davies, 1993; Ikeda et al., 2007).

Lastly, chapter 3 provides the results of our tests to determine whether *Nr2c1* is selectively expressed in a subset of progenitor cells. We investigated whether it is co-expressed with three stem cell markers: *Pax7*, which identifies the pluripotential olfactory stem cell population in the OE; *Ascl1*, which identifies a rapidly dividing transit-amplifying precursor population, and *Ncam*, which identifies post-mitotic olfactory neurons.

Functional analysis of *NR2C1* gene variants using ancestral gene reconstruction methods

Chapter 4 reports the results of our AGR experiments. While comparative genomic analysis suggests a shift in selection pressure occurred in the human lineage on *NR2C1*, a signature of positive selection is inadequate to conclude that specific nucleotide substitutions are linked with phenotypic adaptations (Barrett and Hoekstra, 2011). In order to establish the latter, the gold standard is to use biophysical assays to identify any functional modifications arising from changes to individual amino acids. We can now reconstruct ancestral DNA and its corresponding protein sequences and generate testable hypotheses about the functional evolution of genes on specific phylogenetic lineages.

These methods were used to analyze how modifications to the modern human version of *NR2C1* may have influenced pluripotentiality in embryonic stem cells. We created three different copies of our gene of interest: human, chimpanzee, and the inferred ancestral copy of *NR2C1* for the hypothetical last common ancestor (LCA) of chimpanzee and modern humans, inserted them into separate mouse embryonic stem cells

that have had the endogenous form of *NR2C1* knocked down, and analyzed the transcriptional and regulatory functions of the different variants of *NR2C1*.

The question the study addresses is whether, or not, the amino acid substitutions seen only in modern humans affect these interactions (*i.e*. the modern human version of *NR2C1* maintains the pluripotentiality of the embryonic stem cells for a longer time relative to the chimpanzee *NR2C1* or to the version inferred for the LCA). Any improvement in the efficiency of *NR2C1* as a transcription factor could result in an increase in pluripotentiality efficiency. Understanding modern human neural development at its earliest stages may help us to better understand one of the proximate mechanisms underlying brain evolution.

Undoubtedly more than one genetic, regulatory, or epigenetic modification occurred in the modern human brain, but by using careful candidate gene approaches it is possible to narrow down the places to look for clues to this aspect of brain evolution and ask questions such as: What genes keep the earliest stem cells in modern humans in their pluripotent state? What allowed modern humans to proliferate more neurons and thus evolve a larger brain? At some point during modern human evolutionary history genomic modifications occurred that kept neural proliferating regions active for a longer period of time and enabled the proliferative cells to divide faster than the proliferative cells of either chimpanzee or of the LCA. Investigating whether or not the substitutions that occurred on the modern human variant of *NR2C1* alter the proliferative capacity of human neural progenitors was at the heart of this part of the thesis project.

The information above helped develop of the following testable hypotheses:

Specific Aim: To determine whether the modern human *NR2C1* (*hNR2C1*) has quantitatively different ability to maintain pluripotentiality in stem cell populations, relative to the inferred ancestral *NR2C1* (*aNR2C1*) and the chimpanzee *NR2C1* (*cNR2C1*) genes, and to quantitatively analyze whether the amino acid substitutions that occurred solely on the modern human lineage altered the efficiency of the gene to act as a transcriptional activator of key stem cell regulatory genes.

Hypothesis 1: If the amino acid substitutions on *hNR2C1* resulted in an alteration in its function as a transcriptional activator of stem cell differentiation, then *hNR2C1* should have different abilities to regulate expression of key stem cell pluripotentiality factors, including *Oct-3/4* and *Nanog*, relative to *aNR2C1* and *cNR2C1*

Hypothesis 2: If the amino acid substitutions on *hNR2C1* resulted in an alteration in its ability to maintain pluripotentiality, *aNR2C1* and *cNR2C1* should have different abilities to maintain pluripotentiality in embryonic stem cell populations.

CHAPTER 2: EVOLUTIONARY STASIS AND DYNAMISM IN HOMINID NUCLEAR RECEPTOR GENES

ABSTRACT

One of the goals of evolutionary biology is to identify the genes that played key roles in the evolution of the modern human phenotype. Genes encoding nuclear receptors (NRs) are attractive candidates for two reasons. First, they are involved in development, reproduction, endocrine signaling, and behavior. Second, because they regulate gene expression by acting as intracellular transcription factor, their evolution has the potential to influence the phenotype. Several NRs have already been implicated in the evolution of modern humans, but the overall prevalence of evolution by natural selection within this group has not yet been assessed. We used a comparative method based on widely-used models of codon evolution to examine NR evolution in primates by examining multiple sequence alignments in human, chimpanzee, gorilla, orangutan, baboon, macaque, marmoset, bushbaby, rat, mouse, dog, and cow. The patterns of sequence evolution observed in our analyses suggest a shift in the intensity of selection pressure occurred on a small subset of genes: *NR2C1* and *PGR* in the human-chimpanzee clade and on *RORA*, *NR2C1,* and *ESRRB* in the great ape clade. These genes can be considered candidates for further investigation via ancestral gene reconstruction using synthetic biology techniques.

Keywords

Nuclear receptors; molecular evolution; functional divergence; codon models; nonsynonymous/synonymous rate ratio (ω)

INTRODUCTION

Human evolutionary biology seeks to answer the question of how the defining characteristics of modern humans evolved, such as our large-brains, upright posture, obligatory bipedal gait, longevity, and extended juvenile period. Traditional research uses morphology and artifacts recovered from archaeological sites to infer anatomy and behavior (e.g., Alemseged et al., 2006; Tryon et al., 2008; Jungers et al., 2009a; 2009b; Braun et al., 2010; Ward et al., 2011) but molecular evolutionary studies and complementary functional analyses hold the best prospect of reconstructing the genetic changes that underlie the evolution of phenotypic variation (Wood, 1996; Allman et al., 2010; Boddy et al., 2012; Sherwood and Duka, 2012).

In 1963 Pauling and Zuckerkandl predicted that the time would come when scientists could use molecular sequences of extant species to reconstruct the genes and proteins of ancestral species (Pauling and Zuckerkandl, 1963). The discovery of the genetic code (Matthaei and Nirenberg, 1961; Nirenberg et al., 1963) allowed for the eventual development of methods for detecting adaptive evolution using genetic data (Goldman and Yang, 1994a; Muse and Gaut, 1994) and helped transform the field of evolutionary biology. Predicting the ancestral morphotype is a key concept in phylogeny reconstruction (Wood, 2010; Wood and Harrison, 2012) and ancestral gene reconstruction (AGR) can provide insights into the evolutionary history of a phylogenetic lineage at the molecular level (Thornton, 2001; Thornton et al., 2003). While questions in evolutionary biology are often difficult to answer, AGR has been a useful approach for testing hypotheses about the functional impact of sequence changes on a protein's evolutionary history (Chang, 2003; Thornton, 2004) and can provide insight into the

evolution of protein function (Chang, 2002; Ugalde et al., 2004; Bridgham et al., 2009; Brayer et al., 2011; Eick and Thornton, 2011; Harms and Thornton, 2010; 2013); when the fossil record alone cannot provide such information. While not the only method for identifying evolutionarily interesting genes, codons models have been especially fruitful for identifying sites for AGR (Ivarsson et al., 2003; Sawyer et al., 2005; Norrgard et al., 2006; Weinberger et al., 2009). In this study, we applied a comparative method based on models of codon evolution to investigate the molecular evolution of a superfamily of transcription factors called the Nuclear Receptors (NRs) to identify suitable candidates for ancestral gene reconstruction studies.

Nuclear receptors diversified from more ancestral gene forms early in animal evolution, more than 400 million years ago (Owen and Zelent, 2000; Thornton, 2001) and function as intracellular transcription factors capable of regulating gene expression in the presence of molecules that have an affinity for lipids. Some are regulated by distinct ligands (e.g., steroid hormones), by interactions with other proteins or DNA, or via post-translational modifications (Gronemeyer et al., 2004). In their capacity as transcription factors, they control many aspects of development, metabolism, reproduction, endocrine signaling, and even carcinogenesis (Kohn et al., 2012), and their direct involvement in numerous physiological functions has motivated considerable research into their role in the evolution of hormone-mediated traits (Ketterson and Atwell, 2009). We chose to focus on the evolution of the nuclear hormone receptors in primates because many of them mediate responses to environmental stressors and play a central role in determining reproductive states and social conditions (Beehner et al., 2009; Bernstein, 2010; Gesquiere et al., 2008; Higham et al., 2012; Nunn, 1999).

The evolutionary history of transcription factors includes cases of adaptive evolution of interactions critical to the evolution of development and the regulation of vertebrate physiology (Thornton et al., 2003; Wagner and Lynch, 2008; Bridgham et al., 2009; Carroll et al., 2011; Eick and Thornton, 2011; Finnigan et al., 2012). Because this family of receptors has been highly conserved for millions of years, an increase in the number of substitutions along a given lineage is often viewed as consistent with an underlying adaptive event. However, amino acid changes also accumulate by neutral processes, and it is challenging to differentiate adaptive substitutions from the background of neutral changes. There are a handful of studies of primate NRs indicating that the dynamics of nonsynonymous evolution can differ significantly from neutral expectations (Krasowski et al., 2005; Williamson et al., 2007; Chen et al., 2008). Such inferences are a statistical challenge because sequence divergence among primate NRs occurs at characteristically low levels. A full understanding of the functional divergence of primate NRs requires linking structurally-aware evolutionary studies with experimental approaches that can investigate the functional and evolutionary significance of specific amino acid substitutions (Ugalde et al., 2004; Brayer et al., 2011; Kratzer et al., 2014).

Modern humans possess 48 NRs (Robinson-Rechavi et al., 2001) and early research grouped them according to their ligands or DNA binding properties, but more recently they have been grouped phylogenetically according to inferred synapomorphies (Laudet, 1997; Nuclear Receptors Nomenclature Committee, 1999). This system resulted in a nomenclature with six subfamilies known as NR 1 through 6 (Germain et al., 2006). Most NRs share a modular structure comprised of the following six [A-F] regions: [A/B]

a variable N-terminal domain (NTD), [C] a DNA binding domain (DBD), [D] a flexible hinge domain, [E] a ligand binding domain (LBD) and [F] a C-terminal domain (CTD); the latter is absent from some NRs. The DBD is the most evolutionarily conserved region, with most of the variation among NRs residing within the NTD. The LBD can also vary considerably and deep within this domain lies a receptor-specific hydrophobic ligand-binding pocket (Zhang et al., 2004). Integral to the process of target gene activation is the association of DNA-bound NRs with coregulators (typically a subunit of a larger multi-protein complex). There are at least two coregulator-recruiting activation domains for target gene activation. The first, AF-1, is located within the NTD and is ligand independent. The second, AF-2, is ligand dependent and located in the LBD (Germain et al., 2006). The structural organization of the nuclear receptors is found in Figure 1.

Regulation of gene expression by NRs is typically induced by either endogenous or exogenous ligands, but there are some NRs, termed orphan receptors, for which the ligand has yet to be identified or may not exist (Benoit et al., 2006; Enmark and Gustafsson, 1996). Although the coregulators of NRs can have a variety of functions (e.g., histone modification, chromatin remodeling, controlling the pre-initiation complex), they are essential for regulating the expression of the target gene (Wärnmark et al., 2003) and are labeled as coactivators or corepressors, depending on whether they activate or repress transcription. The majority of coactivators are known to interact solely in either the AF-1 region of the NTD or in the AF-2 region of the LBD, however a small number have the capacity to bind in both regions (Glass and Rosenfeld, 2000) and some coregulators have the ability to function as both coactivators and corepressors (Rogatsky

et al., 2001). Lastly, posttranscriptional events such as phosphorylation also affect NR activity (Weigel, 1996).

For the reasons set out above, NRs are attractive candidates for investigation in the context of understanding the recent evolution of the modern human phenotype. However, it is not feasible to experimentally assess the functional consequences of even a subset of amino acids in each of the primate NRs. Because such an assessment involves ancestral sequence reconstruction and experimental evaluation of phenotype, this makes a comprehensive survey prohibitive. Alternatively, molecular evolutionary modeling and analysis, which is feasible on this scale, can provide information that can be used to identify certain genes as candidates for further investigation. For example, Chen et al., (2008) used the results of a published genome-wide scan (Nielsen et al., 2005) to select the progesterone receptor gene for expanded sampling and a structurally-aware evolutionary analysis. Here we describe a large-scale evolutionary modeling and analysis of primate NRs with a focus on detecting cases of hominid-specific (*i.e.* the great apes, to include humans) alterations in the intensity of natural selection. The purpose is twofold: first, to characterize rates and patterns of primate NR evolution with respect to its major structural domains, and second, to identify the best candidates for experimental investigation of human-specific adaptive evolution. While both tasks involve model-based hypothesis tests about the intensity of natural selection pressure, in the latter case the objective is to use the signal in the data to generate novel hypotheses about patterns of NR evolution that could possibly be related to the distinctive aspects of the modern human phenotype.

Methods

Origin and processing of DNA sequences

The UCSC Genome Browser (Kent et al., 2002) contains 146 NR sequences for each of the 12 mammalian lineages (Figure 2) included in this study. There are more than 48 humans NRs in this database because splice variants are included. We downloaded all the NR sequences (Date of download: March 25, 2014) and obtained a provisional alignment for each using MAFFT with default settings (Multiple Alignment using Fast Fourier Transform, 6.864) (Katoh et al., 2002). These data were then filtered in a two-step process. In the first step, all of the alignments were inspected for the presence of in-frame stop codons within well-aligned regions. From each set of alignments corresponding to the splice variants of a single locus, we selected the largest alignment that did not contain in-frame stop codons in any of the sequences. This yielded a set of 48 multiple sequence alignments (1 alignment per each human NR encoding gene). In the second step, the selected 48 alignments were visually inspected and, where necessary, manually adjusted to improve the alignment, to maintain the reading frame, or to exclude poorly aligned regions. These 48 alignments are the basis for all of the evolutionary analyses in this study. The details of the original 146 sequences are provided in Supplemental Note 1 and the classification of the species and genomes used in the comparative genomic analysis is presented in Supplemental Table 1. The final 48 alignments were deposited in the DRYAD data repository (datadryad.org [*add an accession number(s) here*]).

Overview of modeling framework

Modeling and analysis of NR sequence evolution employed codon models, which simultaneously include information about evolutionary processes at both the DNA and protein levels (Goldman and Yang, 1994; Muse and Gaut, 1994). We employed the simplified formulation of Goldman and Yang (1994), which is summarized by

$$q_{ij} = \begin{cases} \pi_j \text{ if synonymous transversion,} \\ \kappa\pi_j \text{ if synonymous transition,} \\ \omega\pi_j \text{ if nonsynonymous transversion,} \\ \kappa\omega\pi_j \text{ if nonsynonymous transition,} \\ 0 \text{ if codons differ by >1 nucleotide} \end{cases}$$

,where κ is the DNA transition to transversion rate ratio, ω is the nonsynonymous to synonymous rate ratio, and π_j is the steady state frequency of the j^{th} codon. Note that Goldman and Yang (1994) differ from Muse and Gaut (1994) by modeling the transition probability between codons as proportional to the target codon frequency (π_j). Muse and Gaut (1994) model the transition probability as proportional to the equilibrium frequency of the target nucleotide at the codon position that has changed. As part of our robustness analyses (described below) we also employed the Muse and Gaut (1994) strategy. Hereafter, we will refer to this as the "MG94-style" codon model.

Our primary interest in this study is the ω parameter, which provides an index of the strength and direction of natural selection pressure acting on a protein (Bielawski, 2013). When nonsynonymous substitutions are selected against, their substitution rate will, on average, be less than the synonymous rate and accordingly this purifying selection is characterized by $\omega < 1$. $\omega = 1$ is consistent with neutral evolution, where

neither type of substitution is selected against. Lastly, a ratio of the rates of nonsynonymous to synonymous substitutions greater than one ($\omega > 1$) is evidence of positive Darwinian selection. However, when molecular adaptation occurs by just a few, or just a single, amino acid substitutions it will not be possible to detect $\omega > 1$ (Bielawski et al., 2004). For this reason functional divergence at the molecular level is also identified by detecting changes in the distribution of ω rather than $\omega > 1$ (Forsberg and Christiansen, 2003; Bielawski and Yang, 2004).

Maximum likelihood (ML) was used to estimate the values for the codon model parameters, including ω. The only exception was the equilibrium codon frequencies, which can be reliably estimated from the empirical frequencies of each alignment (Bielawski, 2013). The likelihood ratio test (LRT) was used to test a variety of different hypotheses (*e.g.*, H_0: $\omega \leq 1$ versus H_A: $\omega > 1$). The LRT combines parameter estimation with a formal comparison of two competing hypotheses. When the null hypothesis is true, the likelihood ratio is approximately χ^2 distributed with degrees of freedom equal to the difference in the number of parameters between the two models (Anisimova et al., 2001). The codeml program from version 4.8 of the PAML package (Yang, 2007) was used for likelihood calculation and parameter estimation.

Fixed effect codon models for structurally aware modeling of selective pressures

Structural and functional interactions at the level of the folded protein yield differences among codon sites in the intensity of natural selection (ω), and sometimes other aspects of the substitution process (Bao et al., 2007; 2008). When relevant information is available *a priori*, it can be built into a codon model as a fixed effect (FE)

partition of an alignment, where different partitions are allowed to have heterogeneous evolutionary dynamics (Yang and Swanson, 2002; Bao et al., 2007). We applied this analytical framework to investigate the rates and patterns of primate NR evolution with respect to its major structural domains. As we were interested in characterizing primate NR evolution, we compared results obtained by using the full dataset to those obtained by excluding the non-primate lineages (rat, mouse, dog, and cow) from these analyses. We then mapped each site in an alignment to (*i*) the NTD, (*ii*) the DBD, (*iii*) the flexible hinge domain, and (*iv*) the LBD using the Conserved Domain Database (Marchler-Bauer et al., 2011). We excluded the CTD from the analysis, as it is not present in every NR. These mappings are included in the files deposited in the DRYAD data repository *(accession number*). With these structural domains specified as a fixed effect in a codon model, we fit the models to the data and employed LRTs to test for heterogeneity among domains in selection intensity (ω), overall rate of evolution via a branch-length scale parameter (c), transition-transversion ratio (κ), and equilibrium codon frequencies (π_j) (Bao et al. 2007).

Fixed effect models for temporal variation in selective pressure

The intensity of natural selection also can change across evolutionary history (Anisimova and Liberles, 2012). This can be investigated with branch models, which are useful for testing hypotheses about when selection intensity has changed over time because they have different ω parameters for different branches of a tree (Yang, 1998). Note that in each case a codon model will employ two independent ω parameters: one for the branch or clade of interest (called the foreground, and denoted ω_{FG}) and the other

parameter for all other branches in the tree (called the background, and denoted ω_{BG}). In these models the ω parameter is estimated as an average over all the sites in an alignment. The LRT was used to test each hypothesis against the null hypothesis of no change in selection pressure (H_0: $\omega_{FG} = \omega_{BG}$).

In this study we tested five *a priori* hypotheses. Three hypotheses represent episodic models of evolution. The first episodic hypothesis is that the intensity of selection pressure changed at the origin of the great apes (H_1 in Figure 2A). The second is that the intensity of selection pressure changed during the origin of the human-chimpanzee clade (H_2 in Figure 2B) (Note: Although this is the human-chimpanzee/bonobo clade in phylogeny, our data analysis was based only on the chimpanzee genome sequence from the (CSAC 2.1.4/panTro4) downloaded from the UCSC browser.) The third is that a change occurred during the origin of modern humans (H_3 in Figure 2C). The intensity of selection pressure returns to ancestral levels after the event of interest in H_1 and H_2; however, as modern humans are the crown group this is not the case in H_3. The final two hypotheses represent models for long-term shifts in selection pressure. In the fourth hypothesis a shift in the intensity of selection pressure occurred at the origin of the great apes and persisted in the descendant lineages (H_4 in Figure 2D). In the fifth hypothesis a shift occurred at the origin of the human-chimpanzee clade and it persisted in both the human and chimpanzee lineages (H_5 in Figure 2E).

Random effect codon models for among site variation in selective pressures

The structurally-aware fixed effect models described above are appropriate for investigating differences in the evolutionary dynamics of different domains. However, any aspect of the evolutionary process that is not captured within the fixed site-partitions of those models remains untested. To further investigate variation in selection intensity among sites in NR genes we employed a suggested set of "M-series" codon models: M0, M1a, M2a, M3, M7, and M8), whose properties have been extensively tested via simulation (*e.g.*, Anisimova et al., 2001; Wong et al., 2004; Bao et al., 2008; Bielawski, 2013). With the exception of M0, these are mixture models that employ a statistical distribution to allow among-site variation in ω without having to specify fixed-effect partitions of sites. A full description of each model is presented in Supplemental Note 2. These models permit three LRTs. The first, M0 vs. M3, is a formal test for variable selection pressure among sites. The second, M1a vs. M2a, is a test for a fraction of sites under positive selection. The third, M7 vs. M8, provides an alternative test for sites under positive selection. The third test is more powerful than the second, but it incurs higher false positive rates and greater sensitivity to misspecification of nuisance parameters. Note that model M3 is based on an unconstrained discrete distribution for ω having a pre-specified number of categories. LRTs were also used to evaluate the number of categories for the M3 distribution.

Codon models for spatial and temporal variation in selective pressure

All the models so far have permitted selection pressure to vary among sites or over time, but never both at the same time. This last set of models was used to test for a fraction of sites where the intensity of selection changed in the tree. The change point in

the tree is modeled as a fixed effect, and we evaluated the same five hypotheses shown in Figure 2. Unlike changes in the tree, variation among sites is modeled as a random effect using one of several different discrete distributions.

Models for an episodic shift (Figure 2A-C) at a fraction of sites are referred to as "branch-site" models (Yang and Nielsen, 2002; Zhang et al., 2005). There are two such models, referred to as Model A and Model B. Model A differs from Model B by placing additional restrictions on its ω parameters. A full description of each model is presented in Supplemental Note 2. Three LRTs are possible: (*i*) M1a *vs*. Model A ($\omega_{FG} > 1$), (*ii*) Model A ($\omega_{FG} = 1$) *vs*. Model A ($\omega_{FG} > 1$), and (*iii*) M3 *vs*. Model B. The first two are intended as formal tests for an episode of positive selection at a fraction of sites, and the third is a test for any episodic change in selection at a fraction of sites (*i.e*., the episode need not involve $\omega_{FG} > 1$).

Models for a fraction of sites experiencing a long-term shift for an entire clade (Figure 2D&E) are commonly referred to as "clade-site" models (Bielawski and Yang, 2004). There are two such models, referred to as Model C and Model D. Model C differs from Model D by additional restrictions placed on its ω parameters. Two LRTs are possible: (*i*) M2a_rel vs. Model C, and (*ii*) M3 vs. Model D. The former is intended as a test for a fraction of sites under positive selection across an entire clade. The latter is a generalized test for a long term-shift in the intensity of selection. A full description of each model is presented in Supplemental Note 2.

Statistical considerations

Because we used the LRT as a tool to survey all 48 sequence alignments, the single tests significance level does not provide adequate control over the probability of making one or more type I errors. However, because a subgroup of genes could be related by an evolutionary event (*e.g.*, all of the genes that experience a shift in selection pressure at the origin of the great apes) such genes can be viewed as a family, and we performed selective inference of this subgroup by using the false discovery rate criterion (Benjamini and Hochberg, 1995). The false discovery rate (FDR) is the expected proportion of false positives among the genes discovered to comprise a family, and we controlled the FDR for a given family of genes by using the method of Storey (2002).

All the codon models used in this study are necessarily simplified representations of the true evolutionary process. The hope is that the models will be able to capture the major features of sequence evolution while being robust to the inevitable departure of the NR sequences from their assumptions. Hence, prior to interpreting the modeling results for specific genes we carried out a series of reliability and robustness analyses. First, at least two co-authors independently inspected the alignments and compared their assessments, as incorrect positional homology can impact some of the inferences (Schneider et al., 2009; Fletcher and Yang, 2010). Second, we used the GARD-MBP method (Kosakovsky Pond et al., 2006) to test for evidence of within-gene recombination events, as these can negatively impact some LRTs (Anisimova et al., 2003). GARD-MBP is powerful, as well as being robust to shifts in selection pressure and sites evolving under positive selection (Bay and Bielawski, 2011). Third, because our analyses were carried out using the assumed species tree, we also estimated gene trees by using RAXML (Stamatakis, 2014) and re-analyzed the data under that topology. Fourth, we re-

analyzed the data using a model where the transition probability was proportional to the equilibrium value of the target nucleotide (MG94-style codon model via codeml). Lastly, we employed a rigorous method to quantify the uncertainty in the parameter estimates. Although ML estimates can provide important insights into the process of evolution, they sometimes have estimation errors. The problem is especially difficult when regularity conditions are not met, which can happen in some datasets and not others. Rather than rely on regularity conditions, we used a novel method (Mingrone et al., in prep) based on the non-parametric bootstrap to assess the uncertainty in the parameter estimates.

Results

Characterizing NR domain evolution in primates and mammals

Prior to the domain-specific investigation, we assessed the overall level of divergence represented by our set of 12 mammalian lineages. As expected, only moderate levels of sequence divergence were observed across these lineages. The median tree length (across the 48 NR genes) computed over the entire evolutionary history of the 12 mammalian lineages was just 1.27 substitutions per codon. Codons have three nucleotide sites, and thus saturation occurs when there are >3 substitutions along any branch of the tree. In Figure 3, which summarizes the full distribution of tree lengths, 75% of genes have a tree length (summed over all 12 lineages) of less than 1.5 substitutions per codon. This distribution is not unusual for gene families such as the NRs, whose role as transcription factors regulating the expression of genes involved in such conserved biological roles as homeostasis and embryonic development necessitates functional constraint. Partitioning total sequence divergence into synonymous and

nonsynonymous divergence revealed, as expected, that the majority of change between these NRs was synonymous (Figure 3). The low levels of nonsynonymous change among NRs means that inferring changes in selection pressure over hominid evolutionary history is a challenging statistical task.

We further investigated the extent to which NR functional constraints were associated with their major structural domains. In two cases (*NR0B1* and *NR0B2*) we were unable to assign each site to the four structural domains (NTD, DBD, hinge domain, LBD), so they were excluded from these particular analyses. We fit a fully heterogeneous codon model (denoted FE1 in (Bao et al., 2007) to the remaining 46 genes in order to estimate the amount of among-domain variability in (*i*) the d_N/d_S ratio, ω_i; (*ii*) the transition to transversion rate-ratio, κ_i; (*iii*) a branch length scale factor, c_i; and (*iv*) codon bias, π_j's, for each of the *i* different structural partitions. Although this model is very likely over-parameterized for gene-specific inferences, our purpose is to investigate average properties of domain evolution; hence, we aggregated the parameter values estimated for each domain over the 46 NRs (Figure 4). These results indicated that NR domain evolution is relatively homogenous for both the transition to transversion rate-ratio and for %GC3 (Figure 4A and 4B). Note that we use %GC at third positions ($\%GC3_i$) as a proxy for the codon bias parameter (π_j's) in the model. In contrast, the intensity of selection pressure differs substantially among domains; the DBD and the LBD are highly constrained by purifying selection and exhibit almost no nonsynonymous changes over the entire evolutionary history of these sequences (Figure 4C). The estimated branch length scale parameters (c_i's) were used to compute domain-specific estimates of tree length, which are presented in Figure 4D. As tree length is largely

homogenous among domains, differences in ω_i among partitions (Figure 4C) are predominantly due to differences in the non-synonymous rate rather than synonymous rate differences among domains. Results in Figure 4 are based on the complete dataset, which includes non-primate mammals. Primate specific estimates were very similar, but with higher estimation errors due to removal of the non-primate sequences from the data.

Lastly, we used the LRT to formally assess model complexity for each gene (Bao et al., 2007). As expected, the models selected for individual genes were less complex than the full FE1 model used above (*i.e.*, in no case was FE1 retained). The most common factor that showed significant variation among domains was ω (36 genes in total), which was expected based on Figure 4. The next most common factor was the branch length scale parameter (25 genes), which was also expected based on Figure 4. However, even among these 25 genes the effect-size for the branch length estimates tended to be small. Indeed, considering their large estimation errors it is not surprising that testing the MLEs for among-domain heterogeneity in tree length does not produce a significant result (ANOVA: F =2.24, p=0.09), which is consistent with the major differences between domains being due to differences in the nonsynonymous rate (Figure 4). A full description of model selection via the LRT for these data, and the results, are presented in Supplemental Note 3.

Survey of temporal variation in selection pressure

If NRs played a key role in the evolution of hominids there might be a large-scale association between the nonsynonymous changes in NRs and individual hominid divergence events. We investigated this possibility by using 'branch' models to test for

changes in the intensity of selection at (*i*) the origin of the great apes, (*ii*) the origin of the human-chimpanzee clade, and (*iii*) along the human branch. We employed LRTs to formally test five different models for evolutionary change at these points: three episodic models (Figure 2A-C) and two shift models (Figure 2D&E). Because our goal was to determine if there might be sub-groups of genes having such associations, the inference of genes related by an evolutionary event was based on controlling the expected proportion of false discoveries within a family of significant hypothesis tests (*i.e.*, FDR control via the *q*-value) (Storey et al. 2002).

Very little signal was recovered for an association with the episodic models (Table 1). Only three significant LRTs were obtained for H_1, and none had a *q*-value < 0.05. The distribution of *q*-values is not surprising because the change from ω_{BG} to ω_{FG} was due to the absence of changes along the FG branch (which could reflect sampling errors when divergence is low), and in two cases the estimated difference between omegas was small (Table 1). No significant LRTs were obtained for H_2. Seven significant LRTs were obtained for H_3, but only one gene (*NR2C1*) had a *q*-value < 0.05. Note that the direction of change between ω_{BG} and ω_{FG} was consistently positive under H_3, and generally larger in magnitude than under H_1, but a high rate of false discovery cannot be ruled out for this set of genes. The small number of significant genes might be due to the statistical difficulty of attempting inferences about the intensity of selection along a single branch as an average over all the sites in the data. Based on these analyses, we do not find evidence for sub-groups of NR genes whose evolution is associated with episodes within hominid evolution as only one gene, *NR2C1* under H_3, had a significant *q*-value.

We did identify a sub-group of genes experiencing a long-term shift in the intensity

of selection during their evolutionary history (Table 1). Using a q-value < 0.05 as the selection criterion, seven genes experienced a shift at the origin of the great apes (Table 1, H_4), and four genes at the origin of the human-chimpanzee clade (Table 1, H_5). In all but one case the inferred direction of change from ω_{BG} to ω_{FG} was an increase in ω. This result is consistent with several evolutionary models. First, an increase in ω could arise due to a generalized relaxation of selective constraints across all sites. Given the functional significance and general evolutionary conservation of NRs, this seems unlikely. Second, the elevated ω could be due to the presence of a fraction of sites under positive selection (*i.e.*, ω >1) in the FG branches of these models. As the models used here averages ω over sites, averaging ω over both constrained and positively selected sites is expected to elevate ω without causing it to exceed one. A third scenario is that there was a radical reorganization of the distribution of selective constraints among sites, and we detected this as an overall shift in ω. Regardless of the cause, these data suggest a subgroup of genes have evolutionary dynamics that are associated in some way with the evolution within the great ape and human-chimpanzee clades.

Survey of among site variation in selective pressures

Among site variation in selection pressure was investigated by using random effect codon models (i.e. the 'Site Models"), which do not require *a priori* categorization of the sites. The hypothesis of among-site variation in ω was formally tested in all NR genes via the LRT of M0 vs. M3(k=2). All models are fully explained in Supplemental Note 2. This test was significant in 43 of the 48 genes (Supplemental Note 4). We expect that all functional NRs will have such among site variation, as the mature protein has a complex

relationship between structure and function. However, we also expect that biologically realistic models will sometimes be rejected in cases of low sequence divergence simply due to the limited information content of such data. Thus we did not apply FDR control to this set of LRTs. Rather, we used this LRT as a criterion to filter out the non-significant datasets prior to applying more complex models (*i.e.*, all subsequent analyses were restricted to the 43 genes that passed this LRT). Among these 43 genes, the distribution of the MLEs (Figure 5) confirmed the expectation that the majority of sites in NRs are subject to strong selective constraints.

We employed two additional LRTs to test each gene for the presence of a fraction of sites evolving under positive selection. No significant results were obtained under the more conservative LRT (M1a vs. M2a). Given the low sequence divergence of these data, we followed this test with the more powerful LRT of M7 vs. M8. This test yielded three significant results; however, none had a q-value < 0.05. Full results for these analyses are provided in Supplemental Note 4. Note that these two LRTs are best suited to detect diversifying selection, which is characterized by rapid nonsynonymous evolution at a fraction of sites over long periods of evolutionary history (Zhang and Rosenberg, 2002). Genes subject to diversifying selection are typically associated with host-pathogen coevolution (*e.g.*, Fitch et al., 1991; Hughes and Nei, 1988; Tanaka and Nei, 1989) or coevolution mediated by sexual conflict (*e.g.*, Rooney and Zhang, 1999; Tsaur et al., 2001; Wyckoff et al., 2000). Bielawski et al. (2004) pointed out that many proteins have been adaptively modified via a few, or just one, amino acid substitution. They referred to this latter scenario as the "globin model" because the globin family represents a well-described case of such evolution (Czelusniak et al., 1982, Perutz, 1983).

The LRTs for positive selection used here are not well suited for detecting sites where short-term bursts of nonsynonymous change are driven by positive selection (Yang and Bielawski, 2000). Given that NR genes are highly conserved, and are not directly involved in strong coevolutionary arms races, their adaptive divergence is more likely to follow the globin model than diversifying selection. For this reason it is not surprising that we did not find evidence for long term positive selection within the NRs.

Testing for lineage-specific episodes of adaptive evolution at subsets of sites within NR sequences

Given that NRs tend to be highly conserved, and are thus unlikely to be subject to long-term positive selection, we used branch-site codon models to formally test for a fraction of sites having a limited episode of positive selection at the origin of the great apes (H_1), the origin of the human-chimpanzee clade (H_2), or along the human lineage (H_3) (Figure 2A-C). Despite using two alternative LRTs for each hypothesis, there was little evidence of episodic evolution. One gene was significant for an episode of temporal change in selection pressure at the origin of the great apes, no genes were significant at the origin of human-chimpanzee clade, and four genes were significant for a change along the human lineage, but only one of these (*NR1D1*) had a q-value < 0.05 (Table 2). These results are consistent with our previous tests under the branch codon models (Table 1), in that we did not find evidence for groups of NR genes having episodes of evolution uniquely associated with hominid evolution.

The signal in *NR1D1* for H_3 was strong despite a small estimate for the fraction of sites having experienced an episodic change of selection intensity (~1%). *NR1D1* was

also the only gene that has a significant LRT under the more conservative LRT-1, although the *q*-value was 0.065 (Table 2). It seems that the intensity for positive selection (ω_{FG}>1) at a few sites appears to be strong (Table 2), as the MLE for ω under both Model A and B was 99. However, this value represents a boundary on the parameter space, and such MLEs are known to have considerable estimation errors. To further investigate the reliability of both the statistical tests and the MLEs, we carried out a series of robustness analyses under Model B (*q*-value = 0.023). First, we estimated a gene tree for *NR1D1* and found that it was identical to the organismal tree employed in our initial analyses. Second, we employed GARD (Kosakovsky Pond et al., 2006) to test for recombination, which can be a source of false positives (Anisimova et al., 2003), and found no evidence for it within this gene. Third, we re-tested M3(*k*=2) vs. Model B (LRT-2) by modeling transition probabilities between codons as proportional to the target nucleotide (*sensu* Muse and Gaut, 1994) rather than proportional to the target codon (*sensu* Goldman and Yang, 1994b). Under the MG94-style model, LRT-2 was significant for H_3 and the MLEs were highly consistent with the original analyses (Table 3). Lastly, we quantified the uncertainty in the MLEs of the original analyses using a novel method based on non-parametric bootstrapping (Mingrone et al. in prep). Unfortunately, the estimated distribution for the fraction of sites under episodic evolution (p_{FG}) was strongly bimodal, with considerable density at zero. This suggests instabilities in the estimate of p_{FG} and the associated ω parameters, which can happen when statistical regularity conditions are not met (Mingrone et al. in prep). This means that the conditions necessary for reliable MLEs and LRTs cannot be assumed for this case despite the good results for our other robustness analyses. Thus, we do not view *NR1D1* as a current candidate for further

study via ancestral sequence reconstruction methods in laboratory. However, because the statistical problems uncovered here could be ameliorated through additional sampling (Mingrone et al. in prep) we suggest this gene is a candidate for expanded sampling of sequences and re-analysis.

Genes in Table 2 with significant LRT *p*-values (> 0.05) also represent additional candidates for expanded sampling of sequences and re-analysis. Testing hypotheses about the intensity of natural selection pressure along a single branch for a small subset of sites is a very challenging statistical task, especially for highly conserved data. If these genes have evolved under an episodic model of positive selection, the statistical tests could also benefit from increasing the amount of data. Indeed, an analysis of the *PGR* gene based on a larger sample of sequences recovered strong signal for positive selection (Chen et al., 2008). As *PGR* had the weakest signal of those presented in Table 2 for H_3, we suggest that *NR2C1* and *PPARG* warrant further investigation. Such expanded sampling, especially when coupled with experimental evaluation of the effect on phenotype, is beyond the scope of the present survey-based study.

Testing for long term shifts in the intensity of selection pressure at subsets of sites within NR sequences

In contrast to our analyses of episodic evolution, using the clade-site models we uncovered evidence in some lineages that for a non-trivial sub-group of genes having a shift in selection pressure. We carried out alternative LRTs for a shift at the origin of the great apes (H_4). Nine genes were significant for LRT-1 (M2a-rel vs. Model C), with five having a *q*-value < 0.05 (Table 4). Ten genes were significant for LRT-2 (M3[*k*=2] vs

Model D[k=2]), but only four had a q-value < 0.05 (Table 4). These four genes also had a q-value < 0.05 for LRT-1, suggesting that the signal for a long-term shift is strong among this subset (*RORA*, *RARG*, *ESRRB*, and *NR2C1*). The fraction of sites subject to a shift (p_{SHIFT} in Table 4) ranged from ~11% (*RORA*) to 90% (*RARG*) under Model D. Note that in *NR2C1* the MLEs differed substantially between Models C and D. This is not surprising given that considerable uncertainty is expected for highly conserved genes such as NRs. Furthermore, the p_{SHIFT} of 90% for *RARG* is extreme. Although this could represent a radical reorganization of the distribution of selective constraints, it is also possible that it is simply an MLE with a large estimation error, or a failure to meet statistical regularity conditions for *RARG*. The results for *RORA* also are noteworthy, as the estimate of ω for the great apes (ω_{FG} in Table 4) is consistent with positive selection.

Further interpretation of the MLEs for any of the genes identified here must be treated with caution without additional assessment (*e.g.*, robustness analyses and bootstrapping MLEs). Therefore, we carried out a series of robustness analyses for the five genes that had a q-value < 0.05 under H_4 for at least one of the LRTs. None of these genes exhibited evidence of recombination, as inferred under the GARD method (Supplemental Table 2). All five genes had highly consistent MLEs and LRT results under the MG94-style codon models, as well as under gene trees that differed from the organismal tree (Table 5): the gene tree differed from the assumed organismal tree for all genes other than *NR2C1*. The highly consistent signal across alternative analyses for H_4 is strong evidence for a long-term shift in selection pressure in these genes at the origin of the great apes. The MLE distributions or *RORA*, *ESRRB* and *NR2C1* were unimodal and bell-shaped (allowing for boundaries on the parameter space such as non-negative

frequencies). As expected, the MLE distributions revealed considerable uncertainty associated with the point estimates for these parameters (Table 5). Nonetheless, the distributions corroborate the signal for a fraction of sites subject to divergent selection pressure among the great apes in those three genes ($p_{SHIFT} > 0$), as well as the signal for positive selection within *RORA* (Table 5).

The extreme value of 90% originally observed for p_{SHIFT} within *RARG* was robust (Table 5) to alternative codon model (MG-style) and alternative tree topology (gene tree) models. Incomplete lineage sorting can arise from many factors, such as unequal rates of evolution, gene flow, recombination, and gene loss or gene duplication. Because of these events, gene trees and species trees are often inconsistent (Fitch, 1970; Goodman et al., 1979) and the problems stemming from this are well known (Avise et al., 1983; Pamilo and Nei, 1988; Takahata, 1989; Maddison, 2006; Rosenberg, 2003). Thus, we conducted a series of alternative optimizations using different initial model values, and although the results presented in Tables 4 and 5 were confirmed as representing the maximum peak in likelihood, there were suboptimal peaks where the size of p_{SHIFT} was not as extreme (data not shown). To determine if those maximum likelihood estimates had good statistical properties we used non-parametric bootstrapping to infer their distributions. Analysis of small datasets can result in parameter estimates with high variability. Moreover the estimates, and the LRTs, can be negatively impacted when the conditions required for maximum likelihood estimation do not hold. To assess the quality of the estimates we used non-parametric bootstrapping of codon sites to generate a number of datasets. For each bootstrap dataset we calculated maximum likelihood parameter estimates, resulting in a distribution of the estimates. Such distributions that are well modeled by a normal

distribution are an indication of reliable parameter estimation. The estimated distribution for p_{SHIFT} was bimodal, which indicates instabilities in the MLEs for *RARG* and *ESRRA*. We used multiple analysis using different initial parameter values to confirm that the bimodal distribution is not due to suboptimal peaks in likelihood. This indicates that statistical regularity conditions might not have been met, and we are forced to conclude that the LRTs, as well as the MLEs, for *RARG* and *ESRRA* cannot be considered reliable. Interestingly, estimates of p_{SHIFT} for *ESRRA* were also very large as compared to the other genes (although not as extreme as *RARG*). Sampling additional sequences, with complete reanalysis, will be required to assess if *RARG* (and possibly *ESRRA*) have indeed experienced a radical re-organization of selection pressure at the origin of the great apes. Alternatively, we do suggest that *RORA*, *NR2C1* and *ESRRB* are excellent candidates for further investigation via ancestral gene reconstruction by gene synthesis and molecular analysis in the laboratory.

We also carried out two LRTs to look for evidence of a shift in selection pressure at the origin of the human-chimpanzee clade (H_5). Eight genes were significant for LRT-1, with five having a q-value < 0.05 (Table 4). Eight genes were also significant for LRT-2, but only three had a q-value < 0.05 (Table 4). These three genes also had a q-value < 0.05 for LRT-1, suggesting that the signal for a long-term shift is strong among this subset (*NR2C1*, *NR2E3* and *PGR*), and the MLEs for these three genes indicate positive selection since the origin of the human-chimpanzee clade at a substantial fraction of sites (22-35%). One gene, *NR2C1*, also had a q-value < 0.05 under H_4, but the MLEs did not indicate positive selection in that case. As the foreground branches of H_5 are a subset of the foreground branches of H_4, it is possible that a significant result for *NR2C1* under H_4

reflects the strong signal for positive selection along the H_5 subset of branches. In such a scenario ω_{FG} under H_4 would have been averaged over some branches that are not under positive selection, thereby yielding an estimate for $\omega_{FG} < 1$. Four of the five genes having a q-value < 0.05 the MLEs also indicate a shift from purifying selection to positive selection at the origin of the human-chimpanzee clade (Table 4). The reason that Models C and D detected this while Models A and B did not is due to their different assumptions about the process of evolution made by these models. Models C and D assume sites experience a shift in selection pressure independent of any selection pressure acting on sites that show no evidence of such a shift, whereas Models A and B assume that sites experience a shift in selection pressure from an ancestral level that is equivalent to the level of selection at sites with no evidence of a shift. As with all our other NRs analyses, further interpretation of the MLEs under Models C and D must be treated with caution without additional assessment of robustness and bootstrapping of the MLEs.

One gene within this set (*NR2E3*) had a significant signal for recombination, as inferred under the GARD method. As the inference of positive selection can be systematically impacted by a history of recombination (Anisimova et al., 2003), and because recombination methods such as GARD are relatively robust to a history of positive selection (Bay and Bielawski, 2011), the results for *NR2E3* must be treated with caution despite being very robust to alternative models (Table 6). Interestingly, the estimated gene tree for *NR2E3* was so different from the organismal tree that it was not possible to use it to test H_5. *NR2E3* was the only gene where this occurred, and one consequence of recombination can be a large difference between the gene and organismal tree. It is worth noting that recombination for *NR2E3* was only marginally significant (*p*

= 0.032 for one breakpoint within the gene), and that recombination analyses were corrected for multiple testing within a gene (via GARD) but not for multiple tests across genes. Because recombination detection methods can sometimes be misled by systematic differences in branch lengths among different regions within a gene, and because we have detected such evolution in NRs (Figure 4C&D), it is possible that the signal for positive selection within *NR2E3* is real. The best approach to understanding the potential impact of recombination on this gene is to collect additional sequence data and carry out a more exhaustive investigation of recombination using several methods, including Bayesian methods for joint analysis of recombination and selection (*e.g.*, Wilson and McVean, 2006).

The bootstrapping analyses revealed bimodal MLE distributions in two genes (*NR1DI* and *ESRRB*) suggestive of parameter estimate instabilities (Table 6). Thus, we cannot assume reliability of the MLEs and LRTs in these two cases, and therefore will not be considered further with respect to H_5. The remaining three genes (*NR2C1*, *PGR*, and *NR2E3*) met the expectation for unimodal and approximately bell-shaped MLE distributions, but because *NR2E3* showed a signal consistent with recombination, it will not be evaluated further. The remaining two genes had parameter estimates suggesting positive selection in the human-chimpanzee clade that were consistent across alternative codon models (models C and D), modeling frameworks (GY94 vs. MG94) and tree topologies (gene vs. organism) (Table 6). Although bootstrapping revealed substantial uncertainties in the point estimates for the MLEs, it also corroborated the signal for a fraction of sites subject to divergent selection pressure among the human-chimpanzee clade ($p_{SHIFT} > 0$). While the bootstrap distributions revealed substantial density for ω_{FG}

> 1, it was not 100% (*NR2C1: 90/94%* ω_{FG} >1; *PGR:* 95/98% ω_{FG} >1). Taking the results of all of the analyses together, we conclude that the signal for functional divergence in these genes is strong, particularly because changes in the distribution of ω, even when $\omega < 1$, is an indicator of functional divergence at the molecular level (Forsberg and Christiansen, 2003; Bielawski and Yang, 2004). We therefore suggest that *NR2C1* and *PGR* are candidates for further investigation via ancestral state reconstruction, gene synthesis, and molecular analysis in the laboratory.

When the H_4 and H_5 results are combined, a total of 15 genes have at least one significant LRT. Nine of these had a *q*-value < 0.05 for at least one of the hypotheses. Genes with *q*-values > 0.05 were excluded to control for the expected false positive rate among those genes "discovered" to be related by an evolutionary event (H_4 and H_5). Four genes were excluded due to statistical considerations: *RARG* and *ESRRA* (under H_4) and *ESRRB* and *NR1D1* (under H_5). *NR2E3* was excluded due to potential issues with recombination. We are left with a subset of four genes (*NR2C1*, *PGR, ESRRB,* and *RORA*) comprise a "family" whose evolution is statistically associated with the origin of the great ape and the human-chimpanzee clades, and we suggest that the signals detected within these genes are sufficient to warrant further laboratory-based investigations. The six genes with a *p*-value < 0.05 but a *q*-value > 0.05 are not necessarily false positives (*ESR2*, *NR1H2*, *NR3C2*, *NR4A3*, *NR5A2*, *PPARG,* and *THRA*). Although we do not view the signal in those six genes as sufficient to warrant laboratory based investigations at this time, we do think it is sufficient to justify collecting additional sequence data followed by a full reanalysis. Five genes (*NR1D1*, *NR2E3*, *RARG* and *ESRRA* and *ESRRB*) also warrant additional sequence data and reanalysis; *NR2E3* to confirm the signal for

recombination and positive selection, and the remaining five to increase the chance that the statistical regularity conditions will be met.

Discussion

Molecular evolutionary methods are now available to help identify patterns of sequence change, make predictions about the functional consequences of modifications to the DNA of our forbearers, as well as to reconstruct ancestral DNA to test hypotheses about the functional evolution of genes in specific lineages. However, exploring the phenotypic consequences of alterations to the genotype is a challenging undertaking and requires a commitment of precious resources- namely time and money- and therefore it is a critical to chose wisely the gene(s) to investigate. Since our results indicate that purifying selection acted on the majority of this gene-set throughout primate evolution, the genes which do display evidence of a shift in selection pressure- especially in the human lineage for those interested in human evolution- are attractive candidates for further investigation in the lab using AGR methods. Those genes that had a signal in the human or human-chimpanzee clade are considered further and are discussed below.

NR2C1

The data from the clade-site models (Table 4) suggest that *NR2C1* experienced a long-term shift in the intensity of selection pressure at the origin of the human-chimpanzee clade (H_5). In our survey of temporal variation in selection pressure using the branch model, *NR2C1* was the only gene with a significant *q*-value under H_3 and was also significant under H_4 and H_5. In all three hypotheses the data suggest that ω

increased between ω_{BG} and ω_{FG}; thus this genes has sites where the rate of nonsynonymous substitution increased in humans and chimpanzees.

NR2C1 is an orphan receptor that binds to DNA and by doing so regulates gene expression (Lee et al., 2002). First isolated from modern human testis and prostate (Chang and Kokontis, 1988; Anderson et al., 2012) it was originally named the second testicular receptor (TR2), but *NR2C1* also helps regulate estrogen receptor activity in mammary glands (Hu, 2002; Shyr et al., 2009) as well as retinoic acid and retinoid X receptors (Lee et al., 1998). *NR2C1*'s expression in embryonic stem cells implies it also plays a key role in embryonic development (Hu, 2002; Shyr et al., 2009) for it is known to regulate several genes associated with maintaining pluripotentiality in stem cells such as *Oct4* and *Nanog* (Park et al., 2006). It is expressed in the developing mouse brain, and its expression has been associated with neuronal differentiation in a mouse ES cell model (Shyr et al., 2009) and because large brains are a hallmark of the later stages of human evolution, any evidence of positive selection in a gene involved with neural differentiation is intriguing. Understanding the differences in molecular interactions at critical stages of neural development between modern humans, chimpanzees, and the inferred LCA might offer insights into modern human brain evolution.

Using the Conserved Domain Database (CDD) (Marchler-Bauer et al., 2011) we located the specific regions in which the human specific amino acid substitutions occurred (Figure 6). The first three of the human-specific amino acid substitutions (V242M, T254A, S274N) are located between the DBD and the LBD, in the putative hinge region. The final amino acid substitution (G514S) is located in the LBD; however, as an orphan receptor its ligand (if any) remains unknown. Glycine (G) and Serine (S) are

both very small, hydropathy-neutral, nonpolar amino acids. Glycine is unique because it lacks a bulky side chain, leading to conformational flexibility that allows it to rest in tight turns in structures and allows the glycine backbone to bind to phosphates. Serine has the ability to lie either on the surface of a protein or within the interior. It is uncharged with a moderately reactive hydroxyl group that can form hydrogen bonds with other polar substrates. Serine can potentially be phosphorylated by a number of protein kinases, thus the addition of this amino acid means a potential gain of a new regulatory site. We used a computational method (NetPhos 2.0; Blom 1999) to verify that this could indeed represent a phosphorylation site and thus have a functional impact on this gene (score = 0.962). However, the ultimate test requires experimental validation in the laboratory, so we suggest that *NR2C1* is a good candidate gene for AGR, and the G514S amino acid in particular is very interesting. A new amino acid in this location could disrupt the ability to bind a ligand or otherwise regulate its activity. As *NR2C1* is classified as an orphan receptor, however, it is currently not known what ligand binds the receptor, or if there is one at all. If *NR2C1* functions as a ligand-independent transcription factor, it is possible that the LBD may simply act as a modulator or transcriptional activator. Thus an amino acid change in the LBD could possibly modulate the transcriptional activity of the protein, either with, or without, ligand binding.

PGR

Although the molecular evolution of the progesterone receptor gene, *PGR,* has been previously described (Nielsen et al., 2005; Chen et al., 2008), because it is a nuclear receptor we were interested in its rate of evolution compared to the rates of other nuclear

receptors. The results were interesting, but not conclusive, under the branch model; this gene had a significant *p*-value, but not a significant *q*-value under hypotheses H_3 and H_5- the two tests that analyzed the human lineage and the human-chimpanzee clade. However, by using the clade-site models we detected significant evidence within *PGR* for a long-term shift in the intensity of selection pressure at the origin of the human-chimpanzee clade (H_5).

Progesterone plays a central role in the biology of childbirth (Csapo, 1961; Graham and Clarke, 1997; Henson, 1998) and progesterone-receptor interaction is crucial for establishing and maintaining a pregnancy (Challis et al., 2000). Differences between modern humans and other primates in the timing of pregnancy and in the mechanics of childbirth have long been a focus of attention in paleoanthropology. There are two isoforms of the progesterone receptor, *PGR*-A and *PGR*-B, and the isoform produced depends on the promoter involved (Giangrande and McDonnell, 1998; Kastner et al., 1990). A change in isoform has been implicated in the initiation of human parturition (Mesiano, 2004; Oh et al., 2005; Merlino et al., 2007). The effects of progesterone are controlled by interactions between *PGR* and its coregulators (Lonard and O'Malley, 2007) in the AF domains (2 in isoform A and 3 in isoform B as well as in the inhibitory function (IF) regions that repress transcription (C. Chen et al., 2008; Sartorius et al., 1994). Isoforms A and B display differential coregulator interactions (Tetel et al., 1999), with *PGR*-A having a higher affinity than *PGR*-B for the corepressor SMRT when PR is in the presence of antagonists, while having a lower affinity than PRG-B in the presence of agonists SRC-1 and SRC-2 (Giangrande and McDonnell, 1998; Heneghan et al., 2007; Molenda-Figueira et al., 2008). This cofactor interaction is crucial for the initiation of

labor (Condon et al., 2003), so it is noteworthy that 6 of the 7 amino acid substitutions in the human *PGR* gene variant in relation to the great apes occurred in the N-terminal domain of *PGR* where cofactor binding takes place (Figure 6) (Chen et al., 2008).

NR2E3

Although a recent analysis of mammalian phototransduction genes using branch-site tests did not identify *NR2E3* as a gene evolving under positive selection pressure (Invergo et al. 2013), when we applied clade-site models we did detect a significant indication that this gene may have experienced a long-term shift in the intensity of selection pressure at the origin of the human-chimpanzee clade.

As a photoreceptor cell-specific nuclear receptor, *NR2E3* is a potential gene for an AGR study because it is involved in the development of the eye (Kobayashi et al., 1999). *NR2E3* is expressed exclusively in the retina (Milam et al., 2002) and acts as a transcription factor to activate rod development and repress cone development, ultimately playing a central regulatory role in signaling pathways fundamental to the photoreceptor cell function (Chen et al., 2005). Within the primate clade the eye is uniquely specialized (Kaas, 2005; Ross and Martin 2007; Kaas, 2013). For example, the evolution of trichromatic vision in catarrhines (Kamilar, Heesy, and Bradley 2013) and the appearance of blue eyes in modern humans (Kobayashi et al., 1999; Kobayashi and Kohshima, 2001; Tomasello et al., 1999; 2007 Bradley et al., 2009). This gene has five human-specific substitutions (Figure 6B).

This gene has five human-specific substitutions (Figure 6B). The first substitution, indicated as T1M, appears to be the creation of a human-specific start codon,

in is what is most likely the 5'UTR region in all of the other non-human species (annotated as an N-terminal truncation p.Met1_Leu8del relative to the human form). However, there is the possibility that the "T" in the other species is an alternative start codon (ACG), as translation at non-AUG codons are possible, but extremely rare and ACG is one of the codons noted as having this capacity (Peabody, 1989). In the absence of experimental evidence of the precise translational start site it is not clear whether this is an N-terminal deletion or a substitution, therefore we have used the simpler latter nomenclature in Figure 6B.

The biological roles with which this gene are involved clearly make this interesting. Although the MLEs suggest the possibility of positive selection in the human-chimpanzee clade, bootstrapping revealed considerable uncertainty about the point estimates for their ω value. Further assessments of *NR2E3* molecular evolution are clearly warranted. It will be particularly important to (i) to further investigate a potential history of recombination (*i.e.* verify signal, estimate the number of events, and evaluate the scope of the sequence exchanges), and (ii) to collect more data to verify if positive selection indeed had a role in its evolution during the evolution of chimps and humans. With the addition of more sequences, this gene might be best analyzed with a full Bayesian method that jointly models selection pressure and recombination. (*e.g.*, Wilson and McVean, 2006). If positive selection can be verified, this gene would then be a good candidate for AGR.

NR3C2

The mineralcorticoid receptor (*NR3C2*) had a significant *q*-value using the branch

model under H_4, and significant *p*-values under H_3 and H_5 and at the origin of the human-chimpanzee clade; but, critically, it did not have a *q*-value < 0.05 under those latter evolutionary scenarios. *NR3C2* binds both glucocorticoids (cortisol) as well as mineralocorticoids (aldosterone), it plays a role in electrolyte and fluid regulation through its interactions with its ligand (Pippal and Fuller, 2008), and is crucial to the initiation of the stress response (De Kloet et al., 2005; Joëls et al., 2008). Stress activates the secretion of corticosteroids from the adrenal glands in the form of cortisol in humans, which return back to the brain and begin working after binding to the glucocorticoid (GR) and mineralocorticoid (MR) receptors (de Kloet et al., 2005; Horst et al., 2012). While both receptors mediate the effect of corticosteroids on emotion and cognition, the MR is involved in the preliminary phases of memory such as evaluating information (Lupien and McEwen, 1997; Oitzl et al., 2010). The MR also mediates actions critical for stress responsiveness and behavioral adaptation (De Kloet, 2000; De Kloet et al., 1998), traits which clearly distinguish us from our closest living relatives.

There are human-specific substitutions in this genes, and Figure 6D indicates that they all occurred in the N-terminal domain where changes in cofactor binding could lead to a different gene transcription pattern in humans. This region contains multiple activation functions, AF1a, AF1b, and a middle transactivation domain (Fischer et al., 2010), as well as an inhibitory function (Tallec et al., 2003) resulting in a complex transactivation system that enables coactivator and corepressor binding and could have an effect on the way this gene functions. Given the attention this gene has already attracted in the literature, we suggest that further collection of sequence data is warranted. Regardless of whether this gene ultimately turns out to be a good candidate for AGR,

understanding its evolution in the context of hormone–receptor interactions could add new insights to the large body of literature already published regarding hormones and primate behavior (Emery Thompson et al., 2010; Markham et al., 2014; Muller and Wrangham, 2004; Sapolsky, 2005; Surbeck et al., 2012).

RORA

Our data indicated that *RORA* displays evidence of a long-term shift in selection pressure in the great ape clade. This gene offers protection to neurons and glial cells from the degenerative effects of oxidative stress (Boukhtouche et al., 2006; Jolly et al., 2012) and is widely expressed in a number of brain regions including the cerebral cortex, hypothalamus, and thalamus (Ino, 2004). A succession of recent papers have implicated this gene in a variety of disorders linking *RORA* with attention-deficit hyperactivity disorder (Neale et al., 2008), bipolar disorder (Le-Niculescu et al., 2009), depression (Garriock et al., 2010; Terracciano et al., 2010), autism (Sarachana et al., 2011) and posttraumatic stress disorder (Logue et al., 2012; Amstadter et al., 2013). Such a clear connection to brain activity makes this gene an interesting candidate for AGR.

ESRRB

Our data indicated that *ESRRB* also displays evidence of a long-term shift in selection pressure in the great ape clade. This gene may regulate *ESR1* transcription (Bombail et al., 2010) and, while its exact function remains unknown, it is interesting because of its biological connections with cancer. *ESRRB* seems to be important in maintaining a hypoxic environment, which is essential for tumor cell growth and

persistence, by interacting with and inducing hypoxia-inducible factor (Ao et al., 2008), making this gene a very attractive target for cancer therapy. In fact, *ESRRB* has been implicated in other cancer research and has been shown to suppress prostate (Cheung et al., 2005; Yu et al., 2008) and breast (Sengupta et al., 2014) cancer cells. Because of their importance for modern human health, nuclear receptors are one of the most common pharmaceutical targets and yet they are susceptible to cross-reactivity and 'off target' pharmacology. Thus far, paleoanthropological research into ancient hormone receptors has informed drug development by uncovering the shared structural identity of ancestral NRs and has improved our understanding of drug selectivity (Fiorucci et al., 2012; Kohn et al., 2012).

Conclusion

Employing ancestral gene reconstruction methods to understand the *proximate* mechanisms (i.e., the differences in molecular interactions between modern humans, chimpanzees, and the inferred LCA) could provide clues to the *ultimate* mechanisms under-pinning the factors that underlie the evolution of modern humans. Our investigation into the molecular evolution of the nuclear receptors is part of a more general effort to identify genes suitable for ancestral gene reconstruction studies. Our data suggest that within the human chimpanzee clade *NR2C1* and *PGR* are two candidates for ancestral gene reconstruction analyses and *RORA*, *NR2C1, AND ESRRB* are candidates within the great ape clade

Figure Legends:

Figure 1. Nuclear receptor structure. NRs have six regions called A-F. The A/B region contains activation function 1 (AF-1). Region C is the highly conserved DNA binding domain (DBD). Region D is a flexible hinge region that connects the DBD to region E, the ligand binding domain (LBD) containing activation function 2 (AF-2). The C-terminal region F varies in length between the different NRs and is nonexistent in some of the genes.

Figure 2A-E. Phylogenetic relationships of the twelve mammalian lineages included in this study and the alternative hypotheses for branch-site and clade analyses. Each alternative hypothesis is tested against the null hypothesis: H_0: $\omega M = \omega GA = \omega HC = \omega H$

Figure 3. Distribution of maximum likelihood estimates of total tree length (t), the nonsynonymous rate (d_N), and the synonymous rate (d_S) under codon model M0. The scale for the total tree length is the number of substitutions per codon site, which consists of three nucleotides. The scale for both the nonsynonymous and synonymous rates is the number of substitutions per single nucleotide site of the relevant type. The distributions are summarized as box and whisker plots. Data points outside the upper and lower fences are interpreted as outliers.

Figure 4. Summary of the variability in the evolutionary process among domains of NR proteins as inferred from codon models where the domain structure was specified as a

fixed-effect in the model. Data are from 46 NR genes partitioned into four categories: N-terminal domain (NTD); DNA binding domain (DBD); flexible hinge domain (HD); and the Ligand binding domain (LBD). Codon model FE1 was used to obtain partition specific estimates (A) transition-transversion ratio; (B) and equilibrium codon frequencies, which are approximated by %GC3 in this figure; (C) selection intensity; and (D) the overall rate of evolution (via a branch-length scale parameter). Each figure summarizes the distribution of the estimates over 46 NR genes. The dotted lines on the boundary of the grey band are equivalent to the upper and lower fences in a box and whisker plot, and are referred to as the upper and lower adjacent values. Values beyond the upper and lower adjacent values in the plot are considered outliers and for clarity are not displayed in these plots.

Figure 5. Distribution of the parameter estimates for model M3 among 48 NR proteins. The p_i parameters are the maximum likelihood estimates for the proportions of two site classes in this model. The ω_i parameters are the maximum likelihood estimates of the d_N/d_S ratio for the two site classes in this model. The distributions are summarized as box and whisker plots. Data points outside the upper and lower fences are interpreted as outliers.

Figure 6. Domain analysis of human specific amino acid substitutions compared to the great apes on key genes with evidence of a variation in selection pressure in the human-chimpanzee clade.

CHAPTER 3: WIDESPREAD EXPRESSION OF NR2C1 SUGGESTS A POTENTIAL ROLE AS A MODULATOR OF NEUROECTODERMALLY-DERIVED TISSUES

Abstract

A comparative evolutionary assessment of all 48 human nuclear receptors found only one gene, *Nr2c1,* a steroid/thyroid superfamily orphan receptor, with a pattern of sequence change consistent with the hypothesis that selection pressure changed within the human-chimpanzee clade. Previous *in situ* analyses lacked the dynamic range and spatial resolution necessary to characterize this gene from early to mid embryonic gestation, when many nuclear receptors, including *Nr2c1*, may be critical for development. We investigated the pattern of *Nr2c1* expression during mouse embryogenesis from embryonic day (E) 9.5 to (E) 17.5 using fluorescent immunohistochemistry, whole mount immunohistochemistry, and quantitative RT-PCR (qPCR). We found evidence of robust *Nr2c1* expression in the olfactory epithelium (OE) during embryonic development. It was also expressed in subsets of neural progenitors such as *Pax7*-positive slowly-dividing stem cells, as well as in *Ascl1*-positive rapidly-dividing intermediate progenitors. *Nr2c1* is also transiently expressed in the olfactory bulb at early stages, coincident with *Nestin*-positive neuroepithelial stem cells, although this expression diminishes with age. We did not observe robust expression of this gene in mature neurons, but modest levels of expression were observed in the newly-born neurons of the migratory mass adjacent to the olfactory epithelium. In addition, we detected mosaic gene expression in the mouse head including Scarpa's ganglion as well

as supporting cells for cranial sensory structures (e.g., retinal-pigmented epithelial (RPE) cells and cells surrounding the mystacial vibrissae). Mosaic gene expression was also detected in a subset of craniofacial mesenchymal cells in developing cranial bones and teeth. The timing and distribution of the expression suggests that *Nr2c1* is associated with the early genesis of mammalian sensory specializations and craniofacial mesenchymal stem cells. Thus, *Nr2c1* may be a candidate for mediating adaptive changes in both craniofacial sensory specializations such as olfaction and in mesenchymally-derived craniofacial structures, such as the face and dentition.

Keywords

Nr2c1, TR2, olfactory epithelium, *Pax7*, *Ncam*, *Ascl1*, craniofacial mesenchyme

Introduction

Nuclear receptors (NRs), which emerged in animals more than 400 million years ago (Owen and Zelent, 2000; Thornton, 2001), function as intracellular transcription factors capable of regulating gene expression in the presence of molecules that have an affinity for lipids. In their capacity as transcription factors, nuclear receptors control many aspects of development, reproduction, endocrine signaling, behavior, and even carcinogenesis (Kohn et al., 2012), and their direct involvement in numerous physiological functions makes them a significant and well-studied group of proteins (Ketterson et al., 2009). Research into the molecular evolutionary history of the NRs indicates strong purifying selection acted on the majority of this gene-set throughout primate evolution. However, a recent investigation of the patterns of NR sequence evolution in primates found that *Nr2c1* is one of the only genes in this family that underwent a shift in selection pressure specifically on the human-chimpanzee clade (Baker et al., 2015, in prep).

Nr2c1 belongs to a subtype of NRs referred to as orphan receptors, for which the endogenous ligand has yet to be identified or does not exist (Enmark and Gustafsson, 1996; Benoit et al., 2006). These orphan receptor NRs share close evolutionary ties to the ligand-binding NRs, and many have been shown to bind to DNA as homodimers, or as a heterodimers with other NRs, to regulate gene expression (Lee et al., 2002) in a ligand-independent manner. Because orphan receptors are the most ancient of the NRs they likely play a role in early embryonic development as well as in the differentiation of primitive cells (Enmark and Gustafsson, 1996; Laudet, 1997). This fact, together with the

noted change in selection pressure on *Nr2c1* in the human-chimpanzee clade, makes this gene a potentially evolutionarily important gene in modern human lineage.

Nr2c1 was initially isolated from modern human testis and prostate (Chang and Kokontis, 1988; Anderson et al., 2011) and based on sequence homology and its similar biological activity (Chang et al., 1994; Giguere, 1999) it is often grouped together with the closely related *NR2C2* gene. Using the mouse as a model organism due to their similarity to modern humans in genetics, anatomy, and physiology (Chen et al., 2012) allowed for a careful evaluation of temporal and spatial specificity of *NR2C1* during neurogenesis, which cannot be performed in modern human or other primates. Early research in mice suggests that *Nr2c1* may have a functional role in regulating stem cell differentiation during early development (Hu et al., 2002), but its known functions in the body are still being explored. Gene expression patterns provide clues to biological function and studies demonstrate that *Nr2c1* is present in the active neural proliferating zones (Lee et al., 1996) and is prominently expressed in the sensory nerve organs (Young et al., 1998), but thus far detailed expression patterns of *Nr2c1* in the mouse head during early embryonic development were mostly determined using radioactive *in situ* hybridization. This method has both a low dynamic range and low spatial resolution, which makes it difficult to identify relative levels of expression and therefore, lacks the ability to assign expression to specific cellular and tissue compartments. Using more specific assays to investigate whether *Nr2c1* is selective for a particular structure or cell type during mouse embryonic development can provide insight to its function.

To better understand *Nr2c1* expression and to provide further insight into its potential roles during development, we investigated the expression of both mRNA and

protein during mouse development across developmental stages by fluorescent immunohistochemistry (IHC), whole mount immunohistochemistry (WMIHC), and quantitative RT-PCR (qPCR). Our study, which provides spatial and temporal expression patterns of *Nr2c1* from embryonic day (E) 9.5 to 17.5, a longer time range than previously reported, is intended to enhance our understanding of its biological activity.

Materials and Methods

All animal procedures were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) at George Washington University.

RNA extraction, cDNA synthesis, and quantitative RT-PCR (qPCR)

We evaluated the expression levels of *Nr2c1* within different brain regions at three stages of development with qPCR, which allows for accurate, sensitive, and fast quantification of nucleic acids. GAPDH served as a control to calibrate the cDNA template used in the qPCR assays and consistent expression levels across different developmental stages were observed. RNA was extracted from whole embryos and microdissected embryo compartments, and cDNA was prepared as per Maynard (2013). Tissues were homogenized in Trizol (Invitrogen), chloroform extracted, and RNA precipitated with 0.5 mL of isopropyl alcohol per 1 mL of Trizol. Purified RNA was quantified, and genomic DNA removed using DNA-free Turbo DNAse treatment (Ambion). First-strand cDNA pools were generated from RNA using ImPrompII reverse transcriptase (Promega) with random hexamer primers. Assays were prepared using an epMotion liquid handling system, by combining 1.25 nM of forward and reverse primers,

cDNA, and 1x SSO-FAST qPCR master mix (Bio-Rad). Quantitative analysis was performed using a CFX384 qPCR platform (Bio-Rad). For additional rigor, all analyses were performed with two independent primers for *Nr2c1*, but because the two primers gave equivalent results in all cases, we only report results from the first set here (using the "*Nr2c1*-A" set). *Gapdh* served as a control and relative gene expression was calculated by 2^ddCT. Primer sequences are shown in Table 1.

Whole Mount Immunohistochemistry

Mouse embryos, ages E9.5 and E10.5, were dissected, fixed o/n at 4°C in 4% PFA, and subsequently dehydrated in MeOH. The endogenous peroxidase activity was blocked by immersion in a MeOH/H_2O_2 solution. *Nr2c1* rabbit anti-human polyclonal antibody (LSBio) and TR2 rabbit polyclonal IgG antibody (Santa Cruz) primary antibodies were incubated for three days at 4°C. The primary antibody was washed away by rinsing in PBST and then the embryos were incubated with HRP-conjugated anti-rabbit secondary antibody diluted in block o/n at 4°C. The secondary antibody was washed away with PBST and developed in DAB/NiCl/PBST solution activated with H_2O_2. Embryos were imaged after dehydration and clearing in BABB. At least five embryos of each stage were stained and analyzed, all with consistent results.

Fluorescent IHC Staining

Mouse embryos were dissected, fixed with 4% PFA, for 1h, cryoprotected by infiltrating with 30% sucrose, embedded in OCT embedding compound (Sakura), and quick-frozen in liquid nitrogen-cooled 2-methylbutane. Specimens were cryosectioned at 12 μm and placed onto histological slides. The fluorescent staining procedure began by

rehydrating the slides in buffer and blocking the tissue with 2% BSA. *Nr2c1* rabbit anti-human polyclonal (LSBio) and TR2 rabbit polyclonal IgG (Santa Cruz) primary antibodies were incubated o/n at 4°C with *Nestin, Pax7, Ncam,* or *Ascl1* as different comarkers. Depending on the second primary antibody, the secondary antibodies used were either AlexaFluor 488 or 546 goat-anti-rabbit IgG (Molecular Probes) or goat-anti-rat IgG (Molecular Probes), in blocking solution. Assays were completed in both microscope channels to verify results. DAPI was applied and slides mounted in a Mowiol mounting solution. To ensure that our antibody reagents were robust and specific, we performed standard immunohistochemistry negative controls (immunostaining without either primary or secondary antibody) on representative sections at all ages; we saw no evidence of non-specific staining in our tissue sections. We confirmed the expression pattern of *Nr2c1* on representative sections with a second, independently-generated antibody to *Nr2c1* that was prepared against a different isotope of the protein.

Results

Initial data indicate olfactory epithelium expression

Whole mount immunohistochemistry (WMIHC) of E9.5 and E10.5 embryos provide an overview of *Nr2c1* expression at the earliest stages of neurogenesis. The data from the E9.5 embryo reveals *Nr2c1* is expressed in the head, branchial arches, and limb buds (Figure 1A). This labeling is most robust in the anterior aspect of the telencephalic vesicle and the dorsal aspect of the mesencephalon. By E10.5, *Nr2c1* displays more restricted expression in the anterior telencephalon and along the dorsal mesencephalon (Figure 1B). Robust expression is still apparent in the limb buds, which are significantly

more developed at this stage. Weaker expression is evident in the mandibular and maxillary pharyngeal arches, and the dorsal neural tube in the trunk. While the olfactory epithelium (OE) is modest at E9.5, distinct concentrations of *Nr2c1* expression are seen in the nascent olfactory placode (Figure 1A), and expression persists in the invaginating olfactory pit at E10.5 (Figure 1B-D). For consistency, *Nr2c1* expression is shown in red throughout the entire manuscript. All areas with detectible *Nr2c1* expression in the E9.5 and E10.5 embryos are areas undergoing proliferation and all are either neural structures or are structures that contain significant numbers of neural crest-derived mesenchymal cells.

In order to ascertain the structural expression of *Nr2c1* in the developing brain, we used semi-quantitative qPCR to measure transcript levels in whole embryos at E8.5 and E9.5, as well as in microdissected regions from E8.5-E10.5 embryos (Figure 1E). To assess whether *Nr2c1* expression is concentrated in specific embryonic regions, we dissected the neural plate/folds from E8.5, and in an E9.5 embryo we microdissected levels corresponding to key segemental boundaries: the telencephalon/mesencephalon, mesencephalon/rhombencephalon, and posterior boundary of the rhombencephalon (i.e., posterior to the otic vesicle). Expression of all but the rhombencephalic segment was at levels approximating that observed in whole E9.5 embryos. To further confirm and quantify expression in the olfactory placode/pit we examined expression in olfactory tissues from E9.5 and E10.5 embryos that were microdissected and enzymatically separated into the component tissues (LaMantia 2000). This analysis showed that expression of *Nr2c1* is significantly increased in the olfactory epithelium (OE), relative to the forebrain or the mesenchyme, in the E9.5 embryo. Expression in the E10.5 embryo

was also significantly increased in the olfactory epithelium (Figure 1E) when compared with the E 9.5 embryo. This is consistent with the pattern of immunostaining in the whole mount embryo (Figure 1C).

Nr2c1 is expressed in a subset of olfactory stem cells

Because our analysis indicated that *Nr2c1* expression is robust in the embryonic OE, we decided to look more closely at the spatial and temporal expression of this protein in the OE using fluorescent IHC in embryos from developmental ages E10.5-E17.5. As recent studies have identified specific markers that identify several key olfactory stem cell niches (Tucker et al. 2010), immunolocalization allows us to investigate whether *Nr2c1* is uniquely associated with a specific class of olfactory stem cell. As noted earlier, *Nr2c1* is expressed in the E10.5 OE, and it is detected most robustly on the lateral side (Figure 2A), consistent with what was observed in whole mount immunolabeling (Figure 1C).

To determine whether *Nr2c1* is selectively expressed in a subset of progenitor cells, we investigated whether it is co-expressed with three stem cell markers: *Pax7*, which identifies the pluripotential olfactory stem cell population in the OE; *Ascl1*, which identifies a rapidly dividing transit-amplifying precursor population, and *Ncam*, which identifies post-mitotic olfactory neurons. At E10.5, we see a small fraction of the cells that robustly express *Nr2c1* co-expressing *Pax7* (Figure 2B, C, E), and significantly more cells co-expressing *Nr2c1* and *Ascl1* (Figure 2G, H, J). In contrast, the *Ncam*-expressing immature neurons do not robustly express *Nr2c1* (Figure 2K-Q). Thus, at E10.5, robust *Nr2c1* expression appears to be found in a subset of olfactory stem cells, although not all *Nr2c1* expressing cells appear to be co-labelled with these stem cell markers.

As olfactory development proceeds, the olfactory placode grows in size and morphological complexity. At E11.5, *Nr2c1* is still expressed throughout the OE, although its expression is less robust in the lateral aspect (Figure 3A, F, K). There is also a substantial shift in the localization of the stem cell populations identified by *Pax7* and *Ascl1* between E10.5 and E11.5. In the larger E11.5 OE, the *Pax7* population is clearly separated from the *Ascl1* population, with the former clearly on the lateral aspect of the OE (Figure 3B), and the latter spatially separated along the medial aspect (Figure 3G). We found that *Nr2c1* is no longer robustly associated with the *Pax7* labeled olfactory stem cell pool (Figure 3D, E, asterisks), although the occasional cell weakly expresses nuclear localized *Nr2c1* (Figure 3D, E, arrow). In contrast, numerous cells along the medial aspect of the OE co-express *Nr2c1* and *Ascl1* (Figure 3I, J, arrows). Intriguingly the medial population that co-expresses *Nr2c1* and *Ascl1* does not show the robust nuclear localization of *Nr2c1* observed at E10.5; instead *Nr2c1* in this subset appears to be found primarily in the cytoplasm. Most *Ncam* positive cells do not express *Nr2c1* (e.g. Figure 3N, O, asterisk), although a small number can be found that co-express *Nr2c1* at E11.5 (Figure 3N, O, arrow), although this expression is cytoplasmic, and appears to be relatively weak in comparison to the robust co-localized expression in the *Ascl1* population.

Later expression suggests widespread craniofacial expression

At later developmental stages, although *Nr2c1* expression is still present throughout the OE, it is also more widespread. Robustly labeled cells can be found throughout the entire OE at E17.5 (Figure 4A). Expression is observed in both the thicker, presumptive neural olfactory epithelium, as well as in the thinner, presumptive

respiratory-type epithelium (Figure 4A, B). The expression of *Nr2c1* at these later stages also appears to be primarily cytoplasmic, unlike the expression observed at earlier stages. As the olfactory epithelium retains an active stem cell pool that is capable of regenerating both neural and respiratory epithelium into adulthood, we investigated whether *Nr2c1* could still be found co-expressed in subsets of cells that express markers indicative of stem cells in older OE samples. We found that *Nr2c1* is robustly expressed in small subsets of *Ascl1*-positive cells that are found in discrete patches within the E16.5 OE (Figure 4C), suggesting that *Nr2c1* is still coincident with this key marker for rapidly dividing transit-amplifying precursor populations

Finally, we investigated whether these patterns of *Nr2c1* expression were also observed in the VNO, the specialized domain of the OE that contains pheromone receptors. We found that *Nr2c1* expression is observed in the VNO at E14.5 (Figure 4D), and E17.5 (Figure 4E), although its expression in the VNO at both stages is less robust than it is in the adjacent olfactory epithelium (OE).

Nr2c1 is expressed in craniofacial mesenchyme and tooth primordia

At E17.5, modest expression is also apparent in the craniofacial mesenchyme, such as the developing septal bone adjacent to the OE and VNO at E14.5 and E17.5 (Labeled as cMES in Figure 4D, E). This expression is particularly robust in a subset of ventral facial bones, including the aforementioned septal and frontal bones (Figure 4F, G). Although expression is most robust at later stages, labeled mesenchymal cells are apparent from the earliest stages we examined, and at E10.5 cells expressing *Nr2c1* were found within in the craniofacial mesenchyme, most robustly on the lateral side of the OE (Figure 2A).

In addition, *Nr2c1* is expressed in the developing teeth (Figure 5A). There is robust labeling in the dental pulp (Figure 5B), and *Nr2c1* expression is co-localized with a subset of *Nestin*-positive presumptive dental mesenchymal stem cells (arrows). *Nr2c1* expression is also observed in a subset of presumptive odontoblasts at the margin of the dental pulp (Figure 5C) as well as in the mesenchymal stem cells and presumptive osteoblasts in the developing alveolar bone (Figure 5D).

Brief Nr2c1 expression in the early stages of central nervous system development

At E11.5 there is clear *Nr2c1* expression in the ventricular zone of the olfactory bulb rudiment (Figure 6A), but expression diminishes in more caudal sections, and only low levels of expression are observed in the anterior cortex. By E14.5 *Nr2c1* protein expression in these regions is greatly reduced and no significant expression is evident in the olfactory bulb (Figure 6B) and the anteriormost regions of the cortex. Only diffuse staining was apparent in more caudal regions of the brain, including the cortex and subpallium/thalamus between E10.5 and E14.5.

Nr2c1 is expressed in supporting cells around cranial sensory specializations

In addition to the expression observed in the olfactory epithelium, we also observed robust expression of *Nr2c1* in several other embryonic sensory structures. *Nr2c1* is apparent in the developing otic placode at E10.5 (Figure 6), as well as in the adjacent otic ganglia. *Nr2c1* expression in the eye, which is robust, is recognizable at all of the stages we examined. At E14.5, the developing eye shows distinctive *Nr2c1* expression at high levels at the margin of the retina, coincident with the presumptive retinal pigmented epithelium. (Figure 6D and inset). Expression in the retina is largely

absent, although there is some low-level diffuse signal along the innermost layer. Finally, *Nr2c1* is heavily expressed in the supporting cells of the vibrissae (Figure 5E, F). Thus, *Nr2c1* expression is present in numerous sensory specializations, although not exclusively in their neural components.

Discussion

Previous research associated *Nr2c1*with embryonic stem (ES) cells and neural development (Hu et al., 2002; Lee et al., 1996; Shyr et al. 2009). Our research demonstrates that this gene is also associated with a wider array of stem cells, particularly those in the OE. *Nr2c1* is expressed in the olfactory epithelium during all of the embryonic ages we studied, although its expression changes as the OE develops. Initially it is expressed robustly in the nuclei of the *Pax7*- and *Ascl1*-positive stem cell populations, but as development proceeds its expression becomes primarily cytoplasmic. *Nr2c1* is expressed in other epithelial sites, generally in locations and in cell populations that contain significant numbers of stem cells associated with sensory specializations. For example, the support cells of the vibrissae harbor a significant population of stem cells that remain pluripotent well into adulthood (Oshima 2001). In some cases, the close identification with undifferentiated and/or stem cell populations is transient (e.g., expression in the nascent olfactory bulb is not maintained as development proceeds). In other cases, however, expression in sensory structures that maintain resident stem cell populations is maintained into late embryogenesis (e.g., the persistent expression of *Nr2c1* in the vibrissae and olfactory epithelium). In the case of the olfactory epithelium, however, the association between *Nr2c1* expression and pluripotentiality is not entirely

clear: while our co-localization analysis clearly shows that *Nr2c1* identifies a subpopulation of pluripotent stem cells at E10.5, its expression becomes more widespread by E17.5.

It is not clear whether the variability in subcellular localization we have observed indicates a change in *Nr2c1* function. At the earliest stages of OE development, we observed clear nuclear localization of *Nr2c1*, however, at later stages, and in most other tissues, we observed cytoplasmic localization of the protein. It is possible that this shift may indicate a relative change in transcriptional activity. Alternatively, the change in subcellular localization may indicate either the presence or absence of an unidentified cofactor or ligand.

We have also established the expression of *Nr2c1* in the craniofacial mesenchyme, including in developing teeth and in a subset of craniofacial bones. As a significant subset of the mesenchymal population involved in dental development is neural crest-derived (Sharpe 2001), it is possible that robust *Nr2c1* expression reflects the significant numbers of neuroectodermally-derived stem cells in these regions. Similarly, the cranial neural crest contributes significantly to the mesenchymal population that forms craniofacial bones and other mesenchymally-derived structures (including vascular smooth muscle and pericytes) particularly in the anterior and ventral regions of the head.

Contrary to our initial expectations, however, we did not find significant expression of *Nr2c1* in the developing CNS at these stages. While we saw clear *Nr2c1* expression in the nascent olfactory bulb at E10.5, this expression was not maintained as development proceeded. We did not observe significant levels of *Nr2c1* expression in more caudal regions of the brain at any gestational age, despite the presence of numerous

Nestin-expressing neural stem cells. Thus, our observations are consistent with *Nr2c1* playing a role in the development of neural crest and placodal neuroepithelium, but not in the development of the central nervous system.

Conclusion

As these experiments were performed in mice, relating these results to human evolution remains a challenge, but our results suggest the possibility that *Nr2c1* may be involved in regulating the pluripotentiality of wide array of cells involved in special sensory modalities and in neural crest-derived craniofacial structures. Evolutionary changes in this gene in the human-chimpanzee clade suggest that *Nr2c1* may be a candidate for mediating adaptive changes in both olfaction and in mesenchymally-derived craniofacial structures, such as the face and dentition, that distinguish modern humans and chimpanzees.

Figure Legends

Figure 1. Expression of *Nr2c1* in the mouse embryo at mid-gestation. **A.** Whole-mount IHC of an E9.5 embryo demonstrates widespread expression of *Nr2c1* in the anterior telencephalon, mandibular and maxillary pharyngeal arches, and dorsal neural tube, as well as in the midbrain. **B.** *Nr2c1* expression in an E10.5 embryo is more restricted in the anterior telencephalon. Expression is still visible in the mandibular and maxillary pharyengeal arches, and to a lesser extent, in the dorsal neural tube and midbrain. Expression is also robust in the developing limb buds. **C.** A higher-magnification view of the E10.5 embryo shown in A reveals concentrated *Nr2c1* expression in the olfactory epithelium, especially on the lateral side. **D.** Fluorescent immunolabeling of *Nr2c1* of a sagittal cryosection from an E10.5 embryo validates that *Nr2c1* expression is robustly expressed in the olfactory epithelium. **E.** qPCR of microdissected embryos confirms *Nr2c1* is robustly expressed in olfactory epithelium at E10.5. Relative expression of *Nr2c1* was assessed for whole embryos (E8.5 and E9.5), microdissected embryos (E8.5 anterior neural plate/folds, and E9.5 embryos dissected at boundaries of the telencephalon, mesencephalon, and rhombencephalon), and the microdissected and enzymatically separated tissue compartments of the frontonasal process (the olfactory epithelium, underlying mesenchyme, and forebrain neuroepithelium of an E9.5 embryo, and isolated olfactory epithelium from an E10.5 embryo). Key: P: posterior; A: anterior; NP: neural plate; TEn: telencephalon; MEn: mesencephalon; REn: rhombencephalon; OE: olfactory epithelium; MES: mesenchyme; FB: forebrain

Figure 2. Fluorescent immunolocalization of *Nr2c1* and markers of stem cell populations in mouse olfactory epithelium at E10.5 shows that a subset of *Nr2c1*-expressing cells are olfactory stem cells. **A-E.** Co-expression of *Nr2c1* (red) and *Pax7* (green), a marker for pluripotent stem cells in the OE. A subset of *Nr2c1*-expressing cells coexpress *Pax7* (arrows, D-E). **F-J.** Co-expression of *Nr2c1* (red) and *Ascl1*, a marker for a rapidly dividing transit-amplifying precursor population in the OE. Numerous *Nr2c1*-expressing cells co-express *Ascl1* (arrows, I,J). K-P. Co-expression of *Nr2c1* (red) and *Ncam* (green), a marker for post-mitotic neurons. Some *Ncam*-positive neurons in the migratory mass show *Nr2c1* expression (L-N, arrows), although this expression is less robust than the expression in the adjacent OE (L). **O-Q.** Within the OE, cells that robustly express *Nr2c1* are not co-labeled with *Ncam* (P-Q, arrows), and *Ncam*-labeled cells express *Nr2c1* at low levels (P-Q, arrows). D-E, I-J, M-N, and P-Q are confocal optical sections (63x, 2 μm)

Figure 3. Fluorescent immunolocalization of *Nr2c1* and markers of stem cell populations in mouse olfactory epithelium at E11.5 shows significant co-localization with *Ascl1*. **A-E.** Coexpression of *Nr2c1* and *Pax7* in the E11.5 OE. *Nr2c1* expression is most robust in the medial aspect, while *Pax7* is primarily in the lateral OE and lateral mesenchyme. Previous experiments have shown that the lateral OE cells that express modest levels of *Pax7* are stem cells. Most *Pax7*-expressing cells in the OE do not express *Nr2c1* (asterisk, D-E), although there are a small number of double-labeled cells with modest *Nr2c1* expression. **F-J.** Co-expression of *Nr2c1* and *Ascl1* shows that most *Ascl1* expressing cells co-express cytoplasmically-localized *Nr2c1* (arrows, I,J).

K-O. Co-expression of *Nr2c1* and *Ncam* shows that there is some co-expression in the OE. Some *Ncam*-positive neurons in the OE have modest levels of cytoplasmic *Nr2c1* expression (N-O, arrow), although *Nr2c1* is absent from others (N-O, asterisks). D-E, I-J, and N-O are confocal optical sections (63x, 2 μm)

Figure 4. *Nr2c1* is expressed in the OE, the vomeronasal organ (VNO), and craniofacial mesenchyme (cMES) at E14.5-17.5. **A.** *Nr2c1* is robustly expressed throughout the entire OE at E17.5. **B.** Higher-magnification image of E17.5 OE shows clear expression in both the thicker (presumptive neural) and thinner (presumptive respiratory) epithelium. **C.** *Nr2c1* and *Ascl1* appear co-localized in small patches of the OE. Inset: 1.5x magnifcation and channel separation. **D.** Expression of *Nr2c1* in the vomeronasal organ (VNO) at E14.5; expression of *Nr2c1* is present, but at lower levels in the VNO and adjacent cMES relative to the adjacent OE. **E.** Expression of *Nr2c1* is diminished in the VNO at E17.5, relative to the adjacent cMES and OE. **F-G.** Expression of *Nr2c1* in presumptive craniofacial bones at E17.5. (F) and (G) are 5x magnifications of the tiled micrograph in (A), corresponding to the locations marked by asterisks, illustrating expression in the ventral septal bone and ventral skull vault, respectively.

Figure 5. Expression of *Nr2c1* in a developing tooth. **A.** Section of E17.5 upper incisor, stained for *Nr2c1* (red) and *Nestin* (green). **B-D.** 4x magnification of specific regions of A, illustrating specific features. **B.** *Nr2c1* shows robust labeling in the dental pulp. *Nr2c1* expression is found in the vicinity of, and co-localized with a subset of

Nestin-positive presumptive dental mesenchymal stem cells (arrows). **C.** Expression in a subset of presumptive odontoblasts at the margin of the dental pulp. **D.** Expression in presumptive osteoblasts and mesenchymal stem cells in the developing alveolar bone.

Figure 6. *Nr2c1* is expressed in the nascent olfactory bulb and a subset of sensory organs. **A.** *Nr2c1* is robustly expressed at the ventricular surface of the olfactory bulb at E11.5, coincident with the apical endfeet of the *Nestin*-positive neuroepithelial stem cells (radial glia). **B.** Expression of *Nr2c1* is absent from the olfactory bulb at E14.5, although numerous *Nestin*-positive radial glia are present. **C.** *Nr2c1* is expressed in the placodally-derived neuroepithelium of the otic vesicle at E10.5, as well as in the adjacent otic ganglia. **D.** *Nr2c1* is robustly expressed in the presumptive retinal pigment epithelia of the eye at E14.5. Inset: 5x magnification. **E-F.** Expression of *Nr2c1* in support cells of the facial vibrissae. E. Low-magnification image across multiple vibrissae shows that *Nr2c1* labeled cells have distinctive morphologies, depending on the depth of the cross-section of the vibrissae. F. Confocal optical section illustrating support cells.

CHAPTER 4: AN INVESTIGATION OF EVOLUTIONARILY IMPOSED DIFFERENCES IN THE MAINTENANCE OF PLURIPOTENTIALITY USING ANCESTRAL GENE RECONSTRUCTION METHODS

ABSTRACT

Explaining how the defining characteristics of our species evolved is at the core of human evolutionary biology. Fossil-based research primarily uses morphology to generate and test hypotheses about taxonomy, phylogeny, behavior, and life history, but recently tools have been developed that allow researchers to generate hypotheses about evolutionary history using evidence drawn from DNA and proteins. It is now possible to reconstruct ancestral DNA sequences and generate testable hypotheses about the evolution of genes on specific phylogenetic lineages. Large brains are a hallmark of modern humans and although there has been extensive research exploring the possible forces that drove an increase in brain size during human evolution, less is known about the proximate molecular mechanisms that underlie human-specific aspects of brain development. A recent comparative genomic analysis of the nuclear receptor family suggested that *Nr2c1* displays evidence of adaptive evolution in the human-chimpanzee clade as demonstrated by a significant increase in nonsynonymous substitutions compared to synonymous substitutions. *Nr2c1*, also known as the testicular receptor 2, is expressed in blastocysts and embryonic stem cells. Manipulations of expression in cultured embryonic stem (ES) cells suggest that *Nr2c1* is involved with neural differentiation, and it can act as a transcriptional activator of two pluripotency factors

required for self renewal of embryonic stem cells, *Oct4* and *Nanog*. Because a difference in neuronal proliferation is one of the key characteristics that vary among mammals, *Nr2c1* is an attractive candidate gene for ancestral gene reconstruction (AGR) studies. In this study, we have used AGR to assess the possibility that the evolutionary substitutions in the sequence of Nr2c1 modulate a key aspect of brain development, namely the pluripotentiality of stem cell populations. By co-transfecting ES cells with an siRNA and with vectors that express either the human, chimpanzee, or the synthetically generated gene reconstruction version of the human-chimpanzee last common ancestor, of *Nr2c1*, we were able to test whether the isoforms transcriptionally-activate differential expression of pluripotentiality genes, and whether they differentially support the self renewal of ES cells. However, we found no evidence that primate versions of *Nr2c1* can appropriately restore expression of *Oct4* and *Nanog*, nor was there any evidence that the primate forms of *Nr2c1* are able to modulate self-renewal in a significant fashion. Either the effects of *Nr2c1* modulation are too subtle to be detected using the current assays, or *Nr2c1* may influence a different subpopulation of stem cells.

Keywords

Ancestral gene reconstruction, pluripotent, embryonic stem cell, *Nr2c1*

Introduction

Evolutionary changes in the capacity for innovation, understanding of mental states, abstract thought, and symbolism in modern humans have been linked with the enlargement and reorganization of the brain (Moss And Young, 1960; Holloway, 1968; Foley and Lee, 1991; Leigh, 2004; Sherwood et al., 2006; Preuss, 2011; Buckner and Krienen, 2013; Rilling, 2014). We can infer brain size and structure based on endocranial casts, which provide impressions of the external brain surface (Holloway et al., 2004; Falk et al., 2000), but the fossil record does not preserve the anatomy of structures underneath the brain's outer surface, nor can it provide clues about cellular microstructure or gene and protein expression. Artifacts have been used as a proxy for the cognitive abilities of extinct hominins (Alemseged et al., 2006; Tryon et al., 2008; Jungers et al., 2009a; 2009b; Braun et al., 2010; Ward et al., 2011), but comparative biology and complementary functional analyses hold the best prospect of reconstructing the evolutionary history and the genetic basis of the neuroanatomical changes that underlie the unique aspects of modern human behavior (Wood, 1996; Wood, 2010; Allman et al., 2011; Boddy et al., 2012; Sherwood and Duka, 2012).

The average modern human brain is approximately 1330 g, while it is estimated that the brain of the last common ancestor (LCA) of modern humans and chimpanzees/bonobos was approximately 300-400 g (Sherwood et al., 2008; Montgomery et al., 2010; Boddy et al., 2012). This increase in overall brain size, which results from an increase in the number of neurons rather than to substantial increases in neuronal size (Kaas, 2008; Lent et al., 2012), was accompanied by other modifications to

brain structure and function, such as changes to modularity and to the connectivity of different brain regions during development (Sherwood et al., 2006; Preuss, 2011; Buckner and Krienen, 2013; Gómez-Robles et al., 2014).

There has been extensive research exploring the possible forces that drove the increase in the size of the human brain, with hypotheses ranging from energetic trade-offs to cognitive benefits (Montgomery et al., 2010; Isler and Van Schaik, 2014). Researchers have also identified some of the genes involved with human brain evolution (Geschwind and Konopka, 2012; Enard, 2014) phenotypic changes to brain folding (Lewitus et al., 2013; Zilles et al., 2013), cortical thickness (Lewitus et al., 2014) and neuronal number (Fonseca-Azevedo and Herculano-Houzel, 2012). This body of work suggests it is likely there are evolutionary pressures that modulate the complex network of intercellular signaling events regulating the generation of neurons in the primate brain, both globally to influence overall brain size, and regionally to influence the relative size and/or morphology of specific brain regions.

The number of neurons is determined in large part by the number and type of proliferative cell divisions of the neural stem cell pool. The neurons of the central nervous system (CNS), as well as the neural crest, all begin as highly proliferative, pluripotent neuroepithelial cells within the neural plate that undergo symmetric cell divisions to generate similarly proliferative and pluripotent progeny. As neurogenesis proceeds, the pace of proliferation decreases and neural stem cells increasingly undergo asymmetric cell divisions that generate more restricted neural stem cell populations, as well as committed neurons and glial progenitors (Florio and Huttner, 2014). Thus, the number of neurons in the neocortex, as well as its cellular diversity, may be regulated in

large part by how long the neural stem cell pool retains pluripotentiality (Geschwind and Rakic, 2013). The evolutionary and developmental constraints imposed by this progressive restriction of pluripotentiality is likely more significant in nervous system development than it is during the development of other organs or tissues, as the proliferation of CNS neurons is almost entirely complete before birth, whereas most other organs and tissues retain pluripotent stem cell populations into adulthood.

Comparative genomics as a method for identifying candidate genes

In an attempt to understand the molecular mechanisms involved with brain evolution, previous research has examined the molecular evolution of genes linked to variation in brain size in developmental disorders (e.g., investigating microcephaly genes to provide a clue to brain size evolution). Comparative genomic analyses of microcephaly genes suggest that *ASPM*, *CDK5RAP2*, *CENPJ* and *MCPH1* underwent positive selection during the evolution of human and non-human primates (Zhang, 2003; Evans et al., 2004; Kouprina et al., 2004; Wang and Su, 2004; Evans et al., 2006; Ali and Meier, 2008; Montgomery et al., 2011). Candidate genes have been proposed that may modulate the proliferative ability of cells in the developing brain, including energetic factors and modulators of cell-cycle (Preuss et al., 2004; Johnson et al., 2009; Montgomery et al., 2011; Vallender, 2011; Konopka et al., 2012; Somel et al., 2013; Enard, 2014; Fontenot and Konopka, 2014), but only recently have functional genomic experiments begun to test the adaptive effects of these genetic modifications (Pulvers et al., 2010; Shi et al., 2013; Boyd et al. 2015; Florio et al. 2015). A recent comparative genomic analysis of a family of transcription factors, the highly conserved nuclear receptors (NRs), found that the pattern of sequence evolution on one of these genes,

Nr2c1, is consistent with the hypothesis that this gene underwent positive selection in the human-chimpanzee clade.

Nr2c1, testicular receptor 2

Nr2c1 belongs to a subtype of NRs known as orphan receptors for which the endogenous ligand has yet to be identified (Lee and Chang, 1995). Because orphan receptors are the most ancient of the NRs, they likely play a role in early embryonic development as well as in the differentiation of embryonic cells (Enmark and Gustafssson,1996; Laudet, 1997). Acting as a transcription factor, *Nr2c1* binds to DNA to regulate gene expression (Lee et al., 2002). Originally named the testicular receptor because it was first isolated from modern human testis and prostate (Chang and Kokontis, 1988; Anderson et al., 2012), its expression in embryonic stem cells and in pluripotent cell culture lines indicates it plays a role in early embryonic development (Hu et al., 2002). It has been implicated as one of a handful of genes that regulate the pluripotentiality of stem cell populations in the embryo, and, in particular, in neural stem cells (Lee and Chang, 1995; Hu et al., 2002; Lee et al., 2002; Shyr et al., 2009). In addition, *Nr2c1* has been shown to regulate the expression of *Oct4* and *Nanog*, two transcription factors that are essential for maintaining the pluripotentiality of embryonic stem cells (Pikarsky et al., 1994; Niwa et al., 2000; Boiani et al., 2002). Thus, because its patterns of sequence change are suggestive of positive selection and it has a potential role in maintaining the pluripotentiality of stem cells and in neural differentiation (Lee and Chang, 1995; Hu et al., 2002; Lee et al., 2002; Shyr et al., 2009) this gene may be relevant for human evolutionary neurobiology. However, a signature of positive selection from bioinformatic analysis is not sufficient to conclude that specific nucleotide

substitutions are linked with phenotypic adaptations (Barrett and Hoekstra, 2011). In order to demonstrate this, it is necessary to use gene synthesis and molecular analysis to identify the functional modifications arising from changes to individual amino acids. We can now reconstruct ancestral DNA and its corresponding protein sequences and generate testable hypotheses about the functional evolution of genes on specific phylogenetic lineages.

Ancestral gene reconstruction (AGR)

In 1963 Linus Pauling and Emil Zuckerkandl predicted that the time would come when scientists could reconstruct the genes and proteins of ancestral species using the molecular sequences of extant species (Pauling and Zuckerkandl, 1963). We now have methods that can detect a shift in evolutionary pressure on specific genes (Goldman and Yang, 1994; Muse and Gaut, 1994) and functional assays that use ancestral sequence reconstruction methods (Chang 2002; Ugalde et al., 2004; Bridgham et al., 2009; Brayer et al., 2011; Eick and Thornton, 2011; Akanuma et al. 2013). Predicting ancestral character states, whether phenotypic or genetic, is fundamental to phylogeny reconstruction (Wood, 2010). Ancestral gene reconstruction (AGR) can be used to recover the molecular ancestral character states and provide insights regarding the evolutionary history of a phylogenetic lineage (Harms and Thornton, 2010; Eick and Thornton, 2011). Hypotheses about primate and mammalian relationships in the form of phylogenetic trees, together with genetic data, allows researchers to infer the ancestral sequences and generate hypotheses about the functional impact of sequence changes in a protein's evolutionary history- something not possible with fossil evidence (Ortlund et al., 2007; Carroll et al., 2008; Chen et al., 2008; Finnigan, 2012).

In the current study, we set out to determine if the *Nr2c1* amino acid substitutions unique to the terminal modern human lineage alter the ability of the gene to maintain pluripotentiality relative to either the chimpanzee *Nr2c1,* or the inferred *Nr2c1* of the hypothetical last common ancestor (LCA) of humans and chimpanzees-bonobos. By heterologously expressing these proteins in embryonic stem cells, we tested whether these sequences differ in their ability to maintain pluripotentiality, or in their relative ability to regulate transcripts associated with pluripotentiality (i.e., *Oct4* and *Nanog*). Any change in the efficiency of *Nr2c1* as a transcriptional activator of pluripotentiality genes, or any change in its ability to maintain pluripotentiality in a stem cell pool, could be evolutionarily significant, as such a change, even if relatively small, could underlie substantial changes in overall brain size.

Results and Discussion

Analysis of Primate Nr2c1 and Ancestral Sequence Reconstruction

We inferred the *aNr2c1* sequence by aligning *Nr2c1* sequences from closely related extant primates, using the mouse and rat as outgroups. The sequence for the ancestral protein was reconstructed by maximum likelihood using the aligned sequence data and previously published phylogenetic trees based on morphological (Lockwood et al., 2004; Diogo and Wood, 2011) and genetic (Ruvolo, 1997; Wildman et al., 2003) data, which allowed for the subsequent synthesis of the ancestral protein (Supplemental Note 1).

Although the ancestral sequence reconstruction allowed us to identify specific amino acid substitutions that define the human, chimpanzee, and ancestral sequences, we

first wanted to exclude a potential alternative hypothesis that could lessen the validity of the sequence reconstruction. The sequencing of the human genome and transcriptome has been thorough, and included numerous individuals, but the sequencing of the non-human primates has generally relied on deep sequencing of the genomic DNA of a small cohort of individuals, without further confirmation by sequencing individual transcripts. To exclude the possibility that some of the amino acid substitutions identified as evolutionarily significant might actually be polymorphisms present in the small cohort of individual chimpanzees included in the sequencing project, or errors in the prediction of the mRNA sequences, we sequenced *Nr2c1* transcripts from multiple chimpanzees. Using PCR to amplify four overlapping fragments of chimpanzee *Nr2c1* that included all of the key amino acid changes included in the ancestral sequence reconstruction, we did not find evidence of that any of the amino acid substitutions represented in the chimpanzee or AGR sequence are likely representative of non-synonymous single nucleotide polymorphisms (SNPs). All of our identified amino acid substitutions were confirmed in all four individuals (Figure 1). While our cohort of chimpanzee transcripts was modest in size, and therefore we cannot exclude the possibility of polymorphisms in *Nr2c1*, the concurrence of the primary sequence in at least nine individuals (the four individuals sequenced here, plus the five individuals included in the genome sequencing project) suggests that any non-synonymous SNPs are likely relatively rare in the population. We also confirmed that a three base indel, leading to the deletion of a serine (S) in the human transcript, was present in the chimpanzee and in the four other primate species we sequenced (Figure 2).

To confirm that the human specific amino acid substitutions were not polymorphisms, we queried the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) to identify known polymorphisms in the human *Nr2c1* coding sequence. None of the 52 reported polymorphisms correspond to any amino acid substitutions identified as human-specific in our previous informatics analyses (Figure 1). Given all of the above, we felt confident that the amino acid sequences for human, chimpanzee, and our inferred LCA were adequate for functional analysis.

We created plasmid vectors that express human (*hNr2c1*), chimpanzee (*cNr2c1*), and the inferred LCA (*aNr2c1*) under the control of the composite CMV/chicken beta actin (CAG) enhancer/promoter. The CAG promoter permits expression of constructs at high levels in a wide range of cell types, including ES cells and other stem cells. We used a PCR-cloned open reading frame (ORF) for the human sequence; but for the chimpanzee and AGR sequence, we created the ORFs *de novo*, with "humanized codon usage" (back-translating the amino acid sequence to a DNA sequence, selecting each codon according to its average frequency in human cDNAs), and generating the ORF by gene synthesis. These plasmids also contain the neomycin acyltransferase gene linked to the *Nr2c1* ORF with an internal ribosome entry site (IRES) (an "IRES-neo cassette"), which allows for the bicistronic expression of both *Nr2c1* and the G418-selectable neomycin gene from the same transcript. In order to reduce the expression of endogenous mouse *Nr2c1* gene (*mNr2c1*) in transfected cells we also generated an siRNA knockdown vector, which expresses a short hairpin RNA fragment, containing sequence complementary to the mouse *Nr2c1* mRNA, This siRNA vector contains a variant of red fluorescent protein (mCherry) and a hygromycin phosphotransferase gene ("IRES-Hyg cassette") driven by

the CAG promoter, allowing for the selection and identification of transfected cells. Using this approach, we transfected mouse ES cells with the siRNA construct to knock down endogenous expression and replaced its expression by co-transfecting either with *hNr2c1*, *cNr2c1*, or *aNr2c1*.

The effect of the repression of TR2 using semi-quantitative qPCR

To test the transcriptional response of our knockdown/replacement strategy, we transfected cohorts of ES cells with five combinations of plasmid vectors. The "nonsense" control cohort uses a control (nonsense) siRNA and an empty neo-cassette containing expression vector; these cells should express normal levels of mouse *Nr2c1*, and no introduced *Nr2c1*. The "knockdown" cohort uses the siRNA construct to reduce expression of endogenous *mNr2c1*, and an empty neo-cassette containing expression vector. The "*hNr2c1*", "*cNr2c1*", and "*aNr2c1*" cohorts all use the siRNA construct to reduce endogenous *mNr2c1* expression, and either the human, chimpanzee, or AGR expression vectors.

After transfecting the ES cells with the appropriate expression vector, we plated the cells in standard ES cell growth promoting conditions – on a feeder layer of mitotically-inactive STO cells, and the addition of Leukemia Inhibitory Factor (LIF) – which supports the maintenance of self-renewing ES cell populations. After 24 hours, co-transfected cells were selected by the addition of G418 and hygromycin. After 18-21 days clonal colonies of ES cells expressing both plasmids could be seen on each plate. We selected 6 colonies from each plasmid combination for expression analysis; these clones were dissociated, replated, grown for 14-18 days and then pooled for analysis.

We performed qRT-PCR to quantify the effect of the repression of *Nr2c1* in ESCs on *Oct4* and *Nanog* expression. Our qPCR analysis showed a 75% decrease in *Nr2c1* expression in the knockdown cohort, thus confirming *Nr2c1* was knocked down by our siRNA construct (Figure 3A). As predicted from previously published reports (Shyr et al., 2009), we also confirmed that two pluripotentiality genes, *Oct4* and *Nanog*, had significantly reduced expression in the knockdown cohort by 76%. We then evaluated whether overexpression of the *Nr2c1* variants could rescue the reduced expression of *Oct4* and *Nanog* following the knockdown of the endogenous *mNr2c1* transcript. We found that *Oct4* was rescued in the ancestral gene variant, reaching levels above that observed in control-transfected cells. We did not observe any rescue in the chimpanzee or human variants (Figure 3B), as the expression levels of *Oct4* and *Nanog* remained at the levels observed in the knockdown cohort, suggesting there is a differential ability of the variants to rescue *Oct4*, a gene central to self-renewal or differentiation. Because only the ancestral sequence of *Nr2c1* rescued *Oct4*, and none of the *Nr2c1* variants rescued *Nanog* (Figure 3C), it appears that there are substantial differences between the ability of mouse and primate sequences to transcriptionally activate these two pluripotency genes. It also suggests that *aNr2c1* may have a different level of activity relative to its human and chimpanzee counterparts, one that may actually be closer in function to *mNR2C1* as the ancestral form behaved the most like the negative control, which is supposed to be representative of the normal *mNR2C1*.

Colony formation as a test for pluripotentiality

As our knockdown of *Nr2c1* led to a significant reduction in *Oct4* and *Nanog*, both of which are key transcription factors that have been shown to be central for

regulating pluripotentiality and self-renewal in ES cells, we predicted that knockdown of endogenous *mNr2c1* and its replacement with its primate variants would result in a detectable change in the pluripotentiality of transfected ES cells. As ES cells must remain pluripotent in order to undergo self-renewal under their standard culture conditions, the ability of ES cells to form new clonal colonies is considered to be a reliable proxy of their pluripotentiality.

Starting with 12 clonal ES cultures for each plasmid combination, we initiated a set of repeated assays to assess the "clone forming ability" of each ES cell population. At each stage, 2,000 cells of each replicate were replated onto a single well of a 6-well tissue culture dish under standard ES cell culture conditions (on a STO feeder layer, with LIF). After 18-21 days, some of these 2,000 cells will be able to self-renew, and each cell will generate a new clone. We assayed the number of clones formed for each replicate, then harvested and dissociated 10 random clones from each of the five experiments for replating at 2,000 cells per well and the entire process was repeated. Any change in pluripotentiality would be expected to result in a decrease (or increase) in the fraction of cells that are capable of self-renewal, and therefore decrease (or increase) the number of clones formed. Furthermore, by repeating this process iteratively, even small (but persistent) changes in pluripotentiality would be expected to concatenate.

We assayed each of the replicates across 4 passages, using this method of repeated passages of 2,000 cells every 18-21 days. Of the initial 60 replicates, 58 survived through the duration of the experiment (two did not survive the initial passage from a single selected clone and are not included in any analysis). To assess the concatenated effects of *Nr2c1* knockdown, we assayed the number of colonies formed

per 2,000 cells at passage 4, and also assessed the average colony size by counting the number of cells present in a sample of 10 colonies (Figure 4). Despite the differential expression of two key pluripotentiality factors (*Oct4*, *Nanog*) in each of the ESC populations, we did not observe that any of the plasmid combinations led to a significant change in clone forming ability (Figure 4). The average colony size of *Nr2c1* knockdown ES cells appears to be marginally reduced (P=0.04 by Student's T-test vs. nonsense control). This reduction in colony size was rescued by co-transfection with *hNr2c1*, *cNr2c1*, and *aNr2c1*. While the average colony size of the *aNr2c1*-transfected cells was the highest, no statistical difference was apparent by one-way ANOVA. Thus, our assays indicate that reduced *Nr2c1* levels are associated with reduced proliferation, and provide an intriguing hint that *aNr2c1* may have a more potent ability to rescue this proliferation defect than *hNr2c1* and *cNr2c1*. However, statistically conclusive evidence of the difference in proliferative potential of the *Nr2c1* variants was not seen in this assay, nor was there evidence of a statistically detectable change in self-renewal (clone-forming) ability (Figure 4).

The failure of our assay to detect changes in self-renewal may be due to one of several factors. First, it is possible that this specific assay, which focused on pluripotentiality/self-renewal on a per-cell basis, may not have sufficient sensitivity to detect modest changes in self-renewal potential. Similar assays have been used to assay potentiality in the published literature (Chambers et al., 2003; Councill et al., 2006), however, it is possible that the success of these assays is due to the use of a more robust challenge that even more significantly reduces expression of pluripotentiality factors, or induces differentiation of the ESC pool. It is possible that an assay that focuses on other

measures of self-renewal, such as an "extinction assay" that forces ES cell cultures to undergo repeated passages at short intervals, may be able to detect disruptions in proliferation that are not obvious in our assay.

Alternatively, it is possible that the primary function of *Nr2c1* is to regulate self-renewal and/or pluripotentiality in populations of stem cells other than ESCs. Our previous analysis of *Nr2c1* expression in the mouse embryo illustrated that it is robustly expressed primarily in placodally-derived neuroectodermal stem cells (e.g. olfactory epithelia), and presumptive neural crest derived mesenchymal stem cells. While previous experiments have shown that manipulations of *Nr2c1* can alter some characteristics of ES cells, it is possible that its effects on pluripotentiality and self-renewal would be more apparent in another stem cell population, such as in a culture of neural stem cells (NSCs).

Finally, it is possible that our results simply support the null hypothesis that *Nr2c1*, of any variant, is neither essential for, nor a modulator of, stem cell pluripotentiality. This seems unlikely given the results of previous research as well as our observation of robust *Nr2c1* expression in subpopulations of mouse stem cells (Baker et al., in prep).

Methods

Sequence validation of primate Nr2c1 cDNA sequence.

In order to verify the amino acid substitutions we observed in chimpanzees were not polymorphisms, we partially sequenced 4 cDNA samples prepared from cortex samples. RNA was extracted by Trizol extraction (Invitrogen), and contaminating

genomic DNA was removed by DNAse digestion (DNAfree Turbo, Ambion). cDNA was generated by random-hexamer-primed first strand cDNA synthesis, using ImPromII Reverse Transcriptase (Promega). We PCR amplified four independent amplicons that overlapped the key polymorphisms identified in the AGR analysis. Amplification of each cDNA was performed using standard Taq polymerase (Qiagen), with PCR primers listed in Table 1. Each PCR product was gel purified, quantified, and sent for sequencing using both the forward and reverse primers used for the initial PCR. Sequences were aligned using Sequencher software and each chromatogram was manually inspected to validate each point where sequences were divergent. To further validate that amino acid substitutions were unique to primates, we sequenced amplicons from additional primate samples from the following species: *Pan troglodytes* (3), *Sympalangus syndactylus* (1), *Papio anubis* (1), *Macaca mulatta* (4). As a reference and for comparison, we also sequenced a human *Nr2c1* coding sequence from a commercially available plasmid clone containing the *Nr2c1* ORF in the shuttle vector pFN21A (Promega).

Ancestral sequence reconstruction

The likelihood (Empirical Bayes) reconstruction of the ancestral sequence was generated with the PAML4 software per (Yang, 1997; Yang, 2007). The ancestral sequences for the human, chimpanzee, and the inferred LCA are presented in Supplemental File 1.

Creation of expression vectors

We generated plasmid expression vectors containing the open reading frame of the human, chimpanzee, and inferred ancestral sequence reconstruction of *Nr2c1*, by

cloning these vectors into the pCINeo4 plasmid, which contains the composite CMV-chicken beta actin enhancer/promoter from pCAGG, the multicloning-site and EMCV-IRES sequence from pIRES2-EGFP (Clontech), and a neomycin coding frame. The human ORF was cloned into the SalI-BamHI sites of this vector by PCR amplifying the ORF of the human *Nr2c1* clone described above, with a forward primer that adds a SalI site and canonical Kozak's consensus site (GTCGACCACC) at the 5′ end of the ORF, immediately preceding the ATG start codon, and adding a BamHI site to the 3' end, in frame with a human influenza hemagglutinin (HA) epitope tag in the vector. Chimp and Ancestral sequence expression vectors were generated by creating human codon-optimized sequences corresponding to their respective amino acid sequences, with similar SalI/Kozak's consensus sites and BamHI sites added to the 5' and 3' ends, respectively. These DNA fragments were synthesized (GeneBlocks, IDT Incorporated) and cloned in a similar fashion into the SalI and BamHI sites of pCINeo. An siRNA knockdown vector was generated by cloning a synthetic oligonucleotide into a short-hairpin RNA expression vector derived from pSilencer (Invitrogen). This siRNA cassette was then subcloned into the backbone of an expression vector (pSCH) containing the composite CMV-chicken beta actin (pCAGG) enhancer/promoter fused to a mCherry-IRES-Hygromycin cassette. Control plasmids containing the pCINeo plasmid without the *Nr2c1* insert, and containing the pSCH siRNA expression plasmid with a nonsense-sequence insert, were also generated. All plasmids were fully sequence-verified before use. Endotoxin-free DNA preparations were made of each vector for use in ES cell experiments (E.Z.N.A. Plasmid Maxi Kit, Omega Bio-tek).

ES cell cultures

Mouse ES cells (E14Tg2a, ATCC) were cultured on a feeder layer in ES cell media containing Dulbecco's Minimum Essential Medium (Invitrogen), supplemented with 15% Fetal Calf Serum (HyClone), 0.1 mM 2-mercaptoethanol, Penicillin/Streptomycin/Amphotercin-B (Anti-Anti, Invitrogen), and 100 U/ml LIF (Enzo). Feeder layers were generated from confluent cultures of STO fibroblast cells (ATCC) by mitotically inactivating the cells with a 2 h treatment with 10 μg/ml mitomycin C and passaging the cells at a 1:2 ratio onto gelatin-coated tissue culture plastic dishes. Stock ES cells, were routinely maintained by trypsinizing and passaging confluent plates of ES cells at a 1:4 subculture ratio every 3-4 days.

Transfection of ES cells and clonal cultures

To transfect ES cells with expression vector plasmids, confluent plates of ES cells were trypsinized, washed twice in PBS, and re-suspended in OptiMem (Invitrogen). 10×10^6 ES cells were re-suspended in 300 μl aliquoted into a 4 mM cuvette, along with 10 μg each of two separate plasmids (a pCINeo plasmid expressing a variant of *Nr2c1* or control, and a pSCH plasmid expressing the siRNA knockdown or the nonsense control). A total of 5 plasmid conditions were assayed and annotated as follows: Nonsense (Nonsense siRNA plasmid + control pCINeo plasmid), Knockdown (knockdown siRNA plasmid + control pCINeo plasmid), Human (knockdown siRNA plasmid + pCINeo containing *hNr2c1* ORF), Chimpanzee (knockdown siRNA plasmid + pCINeo containing *cNr2c1* ORF), and Ancestral (knockdown siRNA plasmid + pCINeo containing the inferred *aNr2c1* ORF). The cells were electroporated with 2 pulses of 500V for 1ms 10 million cells with 5 μg DNA using a square-pulse electroporator (BTX). Electroporated

cells were plated onto feeder layers with LIF containing media in 6-well plates, as described above. After 24 hours, ES cell media was changed to fresh media containing both 300 µg/ml Hygromycin and 300 µg/ml G418, to select for ES cells co-transfected with both expression vectors.

After 18-21 days, single selected clones were apparent on the ES cell plates; we selected single clones, dissociated and counted the total cell numbers, and plated *a calculated* 2,000 cells/well into individual wells of 6-well plates, in the same selection media (ES media with Hygromycin and G418). These clonal ES cell cultures were re-propagated every 18-21 days by selecting 10 clones (chosen from a set of contiguous clones in a single field to minimize any potential selection bias), and then this set of cells was again dissociated in trypsin, counted on a hemocytometer, and replated at 2,000 cells/well to facilitate analysis over multiple generations.

qPCR analysis of expression

Trizol-extracted RNA was obtained from pools of ESCs grown during the first passage. RNA was collected for 6 independent clonal lines for each of the 5 plasmid conditions. cDNA was generated as above, and quantitative real-time PCR was performed to assess expression levels for mouse *Nr2c1*, *Oct4*, and *Nanog* using the primers listed in Table 2. Expression was normalized using *Gapdh* primers as an internal control (Table 2). Reactions were assembled using an EpMotion 5070 liquid handling system (Eppendorf, Hauppauge, NY) that combines forward and reverse gene-specific primers, with 7.5 µl of SsoFast EvaGreen Supermix (Bio-Rad, Hercules, CA) in a 14 µl reaction. qPCR analysis was performed using a CFX-384 Real-Time PCR Detection

System. Statistical differences were assessed by Student's T-test and a one-way anova of the normalized CT values (delta-CT) for each independent analysis.

FIGURE LEGENDS

Figure 1. Validated substitutions in chimpanzees and humans. Amino acid substitutions were verified in humans using dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) and by Sanger sequencing in chimpanzees.

Figure 2. Sequencing of *Nr2c1* transcript cDNAs from human and primate cortex. Example of amino acid sequence showing the human specific deletion in the N-terminal domain

Figure 3. Effect of *Nr2c1* knockdown on pluripotentiality-related gene expression in ES cells. A) ES cells transfected with a m*Nr2c1* knockdown vector show significantly reduced *Nr2c1* expression relative to ES cells receiving a nonsense control vector (24%, $p<0.01$). B) Co-transfection of *aNr2c1* with knockdown vector rescues expression of Oct4 (130%), but *hNr2c1* and *cNr2c1* do not. C) Normal expression of Nanog is not rescued by co-transfection with any of the *Nr2c1* plasmids.

Figure 4. Use of colony formation as a test for pluripotentiality and proliferation shows that knockdown of endogenous m*Nr2c1* leads to non-statistically significant differences for human, chimp, and ancestral *Nr2c1*. A) Assay of colony formation fails to show any statistically significant differences in colony formation for any condition after 4 passages. B) Knockdown of endogenous *Nr2c1* does significantly reduce proliferation, as measured by average colony size ($p=0.04$ by Student's T-test. Co-transfection of either human, chimpanzee, or ancestral *Nr2c1* rescue this reduction in proliferation. Each of these

rescues rises above control levels, although none is statistically significant relative to control ($P>0.1$ by one-way ANOVA).

CHAPTER 5: DISSERTATION CONCLUSION

Project Summary

One of the main goals of human evolutionary biology is to generate and test hypotheses that attempt to explain how the defining characteristics of our species evolved. Inferences regarding taxonomy, phylogeny, behavior, and life history have traditionally been drawn from archaeology and paleontology, but the advent of new computational and biological tools have allowed researchers to generate hypotheses about evolutionary history using evidence drawn from DNA and proteins. Today it is possible to recreate ancestral DNA sequences and test hypotheses about the evolution of genes on individual phylogenetic lineages. The overall goal of this dissertation was to generate and test hypotheses about specific genes, the nuclear receptors, which may have been important in our evolutionary history. There were three parts to this dissertation. First, using comparative genomic methods we found an evolutionary signal in a large amount of genetic data and genes with evidence of a shift in the intensity of selection pressure were identified. Second, a detailed gene expression analysis was undertaken for one gene, *NR2C1*, from the list of candidate genes. Third, ancestral gene reconstruction was used to assess the impact of human-specific amino acid substitutions in *NR2C1* on proliferation and self-renewal in embryonic stem cells.

The molecular evolution of the nuclear receptors

Chapter 2 reports on the results of a comparative genomic investigation of the nuclear receptors. Comparing genomic sequences from different species to the complete human reference sequence allows for the identification of regions of similarities and differences within gene families and for the identification of genes that display a shift in selection pressure throughout their evolutionary history. We used these comparative genomic techniques to investigate the molecular evolution of an important group of proteins directly involved with many physiological functions, the nuclear receptors, the importance of which was reviewed in detail in chapter 2.

Nuclear receptors, which emerged in the animal lineage approximately 400 million years ago, act to regulate gene expression, thereby controlling many aspects of development, reproduction, endocrine signaling, and behavior. Using a comparative genomics approach based on widely-used models of codon evolution to study the evolutionary history of the nuclear receptors, we asked whether adaptive evolution occurred on the human, human-chimpanzee, and the great ape branches of the Primates by examining multiple sequence alignments from human, chimpanzee, gorilla, orangutan, baboon, macaque, marmoset, bushbaby, rat, mouse, dog, and cow. The results of the analyses suggest a shift in the intensity of selection pressure occurred on a small subset of genes: *NR2C1* and *PGR* in the human-chimpanzee clade and on *RORA, NR2C1,* and *ESRRB* in the great ape clade. These genes can be considered candidates for further investigation via ancestral gene reconstruction using synthetic biology techniques. One gene, *NR2C1*, featured in both clades and initial reports suggested that not only was it expressed in the prefrontal cortex, but it was also involved in maintaining the self-renewing state of embryonic stem cells and in early embryonic neuronal differentiation.

All of these observations made *NR2C1* an attractive candidate for further investigation using AGR methods.

NR2C1expression patterns during mouse development

Chapter 3 looks at the gene expression patterns of *NR2C1* in the embryonic mouse. *NR2C1* belongs to a subtype of NRs referred to as orphan receptors for which the endogenous ligand has yet to be identified. Because the orphan receptors are the most ancient of the NRs it is probable they play a role in early embryonic development as well as in the differentiation of primitive cells, however the exact role played by this gene is not fully understood and detailed expression patterns of *NR2C1* in the mouse head during early embryonic development have not been reported. Previous *in situ* analyses provide approximate clues as to its expression, but they lack the dynamic range and spatial resolution necessary to characterize *NR2C1* from early to mid embryonic gestation, when many nuclear receptors, including potentially *NR2C1* may be critical for development. Therefore, chapter 3 provides the results of assays to establish the patterns of *Nr2c1* expression during mouse embryogenesis from embryonic day (E) 9.5 to (E) 17.5 using fluorescent immunohistochemistry, whole mount immunohistochemistry, and real-time quantitative PCR. We observed that *NR2C1* is robustly expressed in the olfactory epithelium during all ages of embryonic development studied, including in subsets of neural progenitors such as *Ascl1* (rapidly dividing intermediate progenitors) as well as *Pax7* (slowly dividing stem cells). Additionally, we found that *Nr2c1* is transiently expressed in the olfactory bulb at early stages, coincident with *Nestin*-positive neuroepithelial stem cells and demonstrate that this expression is reduced with age.

There was no observable robust *NR2C1* expression in mature neurons; however, modest levels of expression were observed in the newly born neurons of the migratory mass adjacent to the olfactory epithelium. Furthermore, locations of mosaic gene expression in the mouse head including Scarpa's ganglion as well as supporting cells of cranial sensory structures: retinal-pigmented epithelial (RPE) cells, and cells surrounding the mystacial vibrissae, and in a subset of craniofacial mesenchymal cells in developing cranial bones and teeth. These locations suggest that *Nr2c1* expression in mammals is associated with the early genesis of cranial sensory specializations and craniofacial mesenchymal stem cells. Thus, it is feasible that evolutionary changes in *Nr2c1* may be a candidate for mediating adaptive changes in both craniofacial sensory specializations such as olfaction, or morphological changes in mesenchymally-derived craniofacial structures, such as facial bones and teeth, as well as in transit amplifying intermediate progenitors in the olfactory epithelium.

Assessing the evolutionary significance of NR2C1amino acid modifications in the human lineage using ES cells

Large brains are a hallmark of modern humans and although there has been extensive research exploring the ultimate mechanisms that drove the increase in brain size during human evolution, less is known about the proximate molecular mechanisms that determine human-specific brain development. The comparative genomic analysis of the nuclear receptor family presented in chapter 2 demonstrates *Nr2c1* displays evidence of adaptive evolution in the human-chimpanzee clade as demonstrated by a statistically

significant increase in nonsynonymous substitution rates compared to synonymous substitution rates.

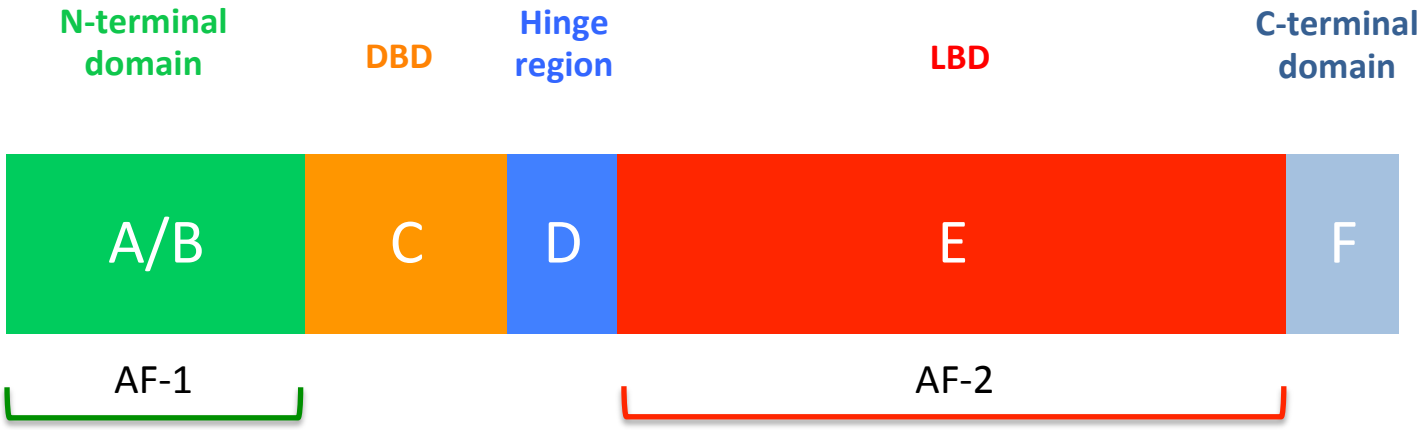
Chapter 4 presents data collected using AGR methods to assess whether the amino acid substitutions of *NR2C1* that occurred in the modern human lineage resulted in a change in the self-renewing state of stem cells, or whether there was any change in its role as a transcriptional activator of two pluripotency factors required for self renewal of embryonic stem cells, *Oct4* and *Nanog*. *Nr2c1* is expressed in blastocysts and embryonic stem cells and manipulations of expression in cultured embryonic stem cells suggest that *Nr2c1* is involved with neural differentiation and acts as a transcriptional activator of *Oct4* and *Nanog*. Modifications in neuronal proliferation will lead to changes in brain size, making *Nr2c1* an attractive candidate gene for ancestral gene reconstruction studies.

Our research used AGR methods to evaluate whether the human-specific amino acid substitutions on *Nr2c1* changed the pluripotentiality of stem cell populations. By transfecting knockdown ES cells with vectors that express either the human, chimpanzee, or the synthetically generated ancestral gene reconstruction version, of *Nr2c1*, we were able to test whether the gene variants differentially support the self renewal of ES cells and whether they transcriptionally-activate differential expression of pluripotentiality genes. Only the ancestral variant of *Nr2c1* recovered *Oct4* expression, while none of the gene variants recovered *Nanog*. Although the chimp variant displayed the highest numbers of stem cell colonies and the ancestral variants had the largest sized colonies, none of the primate forms of *Nr2c1* modulated self-renewal or proliferation in a significant fashion. Either the effects of *Nr2c1* modulation are too subtle to be detected

using the current assays, or *Nr2c1* may influence a different subpopulation of stem cells such as the transit amplifiers that were identified in chapter 3 of this research project.

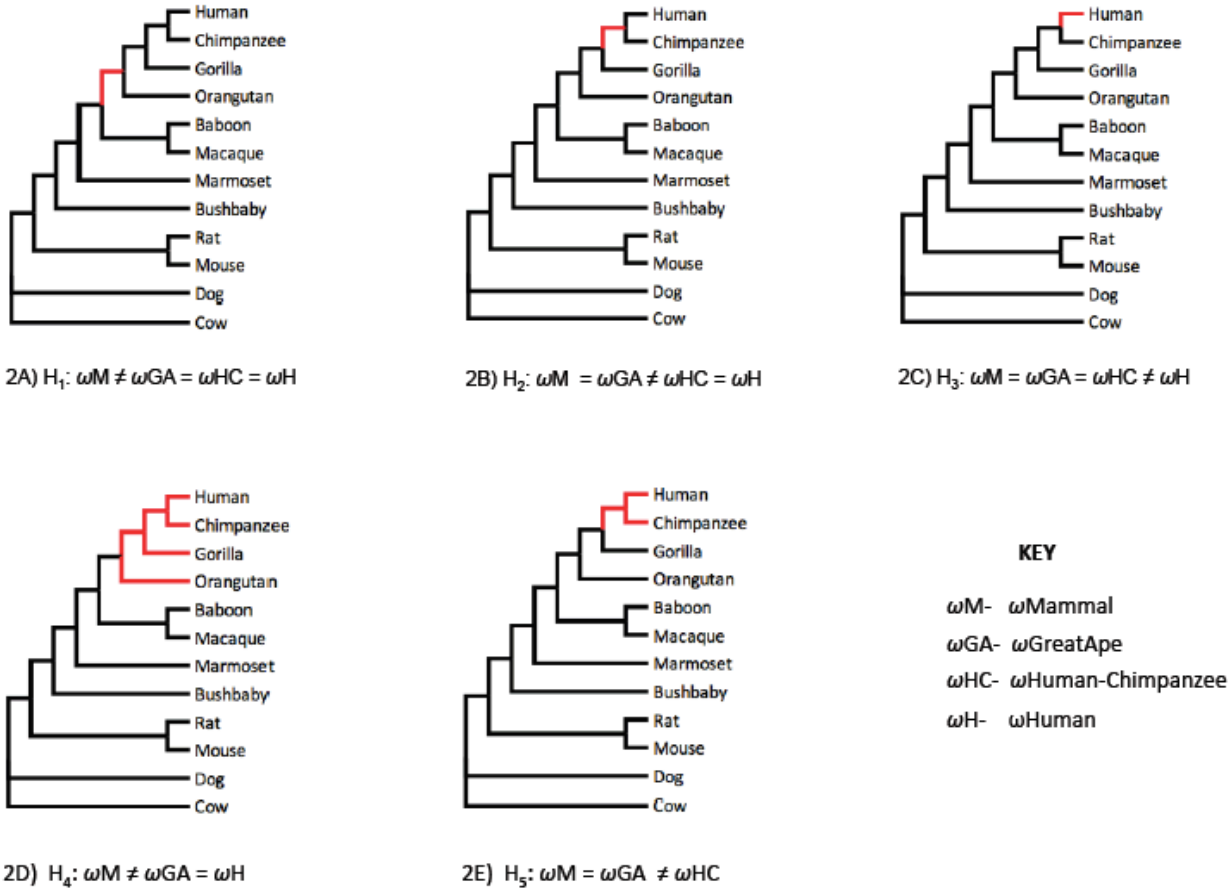
Overall significance

Our analysis suggests that adaptive evolution may have acted on *NR2C1*, but the only way to know for sure if modern human-specific amino acid substitutions resulted in functional changes to the protein is with biophysical analyses. This project investigated the molecular interactions that take place at the earliest stages of brain development. It was an attempt to understand the proximate mechanisms at play during this crucial time period as well as how these molecular interactions may differ from our closest living relative and from the hypothetical inferred last common ancestor. This research attempted to bridge the gap between pattern and process in human evolution and sought to identify one of the mechanisms that underlie the evolution of the large human brain that we can trace, albeit imperfectly, in the fossil evidence of our own lineage.

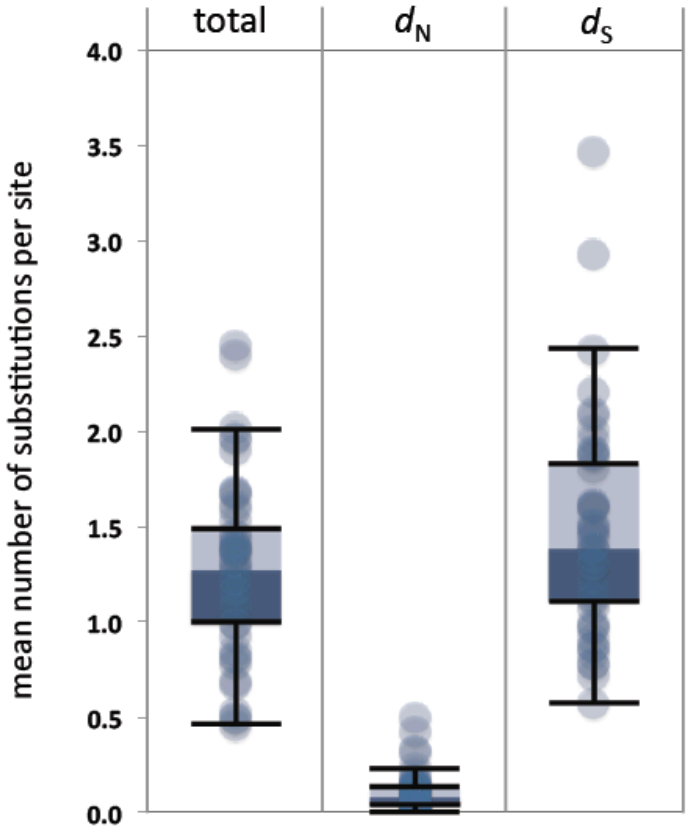


DBD: DNA binding domain
LBD: Ligand binding domain
AF-1: Activation function 1
AF-2: Activation function 2

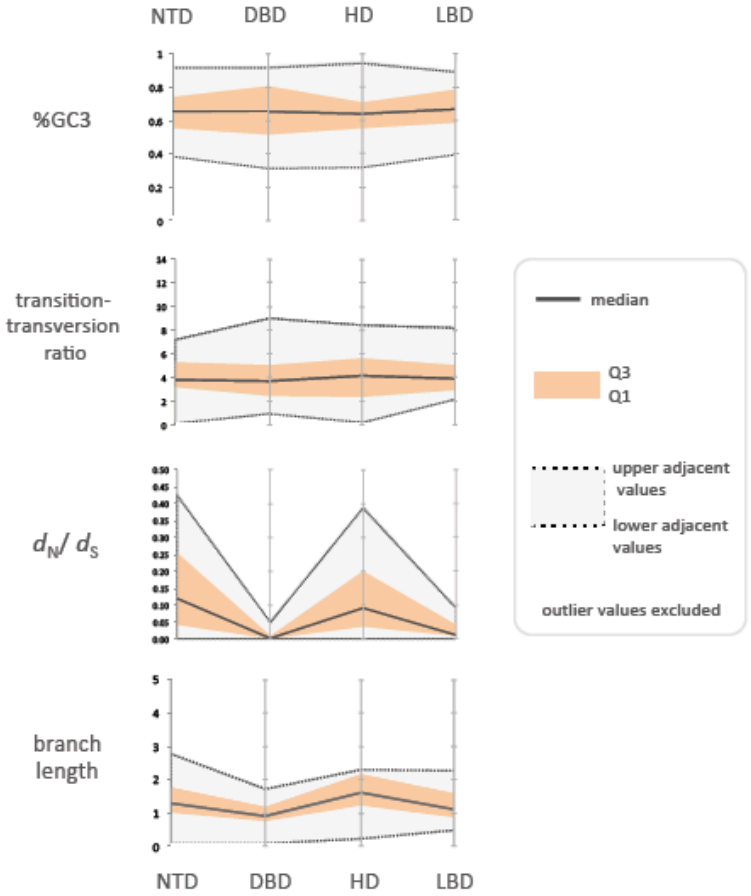
Chapter 2, Figure 1. Nuclear Receptor Structure



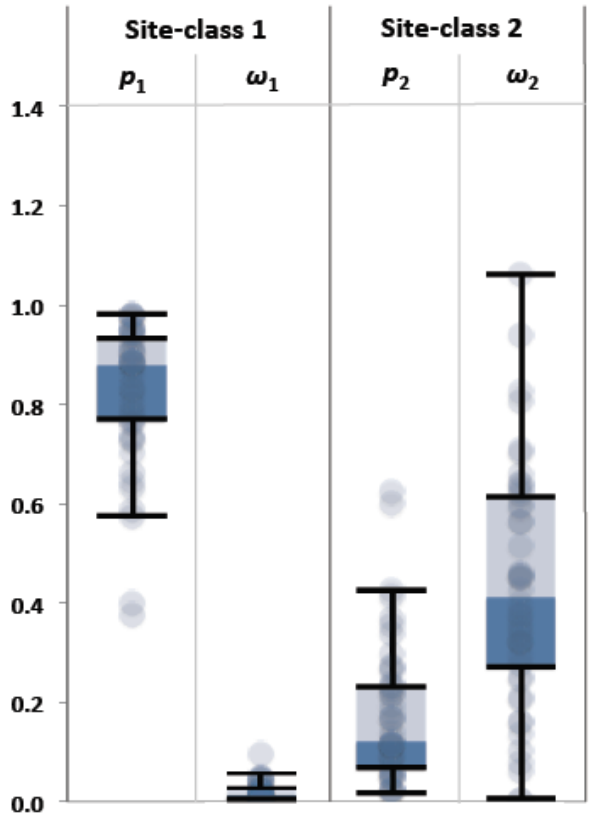
Chapter 2, Figure 2. Phylogenetic relationships of the mammalian lineages studied



Chapter 2, Figure 3. Distribution of maximum likelihood estimates



Chapter 2, Figure 4. Variability in the evolutionary process among NR domains



Chapter 2, Figure 5. Parameter estimates for model M3 among 48 NR proteins

NR2C1

Substitution	Domain
V242M	Hinge
T254A	Hinge
S274N	Hinge
G514S	LBD

PGR

Substitution	Domain
A8V	NTD
S28A	NTD
G37E	NTD
P52A	NTD
S183C	NTD
P266T	NTD
I498V	LBD

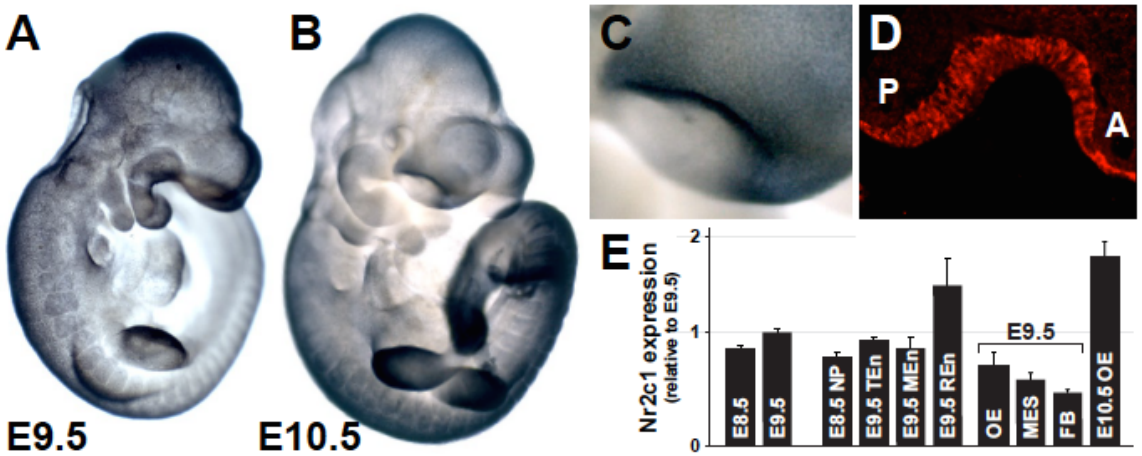
NR2E3

Substitution	Domain
T1M	NTD
R131H	Hinge
P141S	Hinge
P155R	Hinge
R304Q	LBD

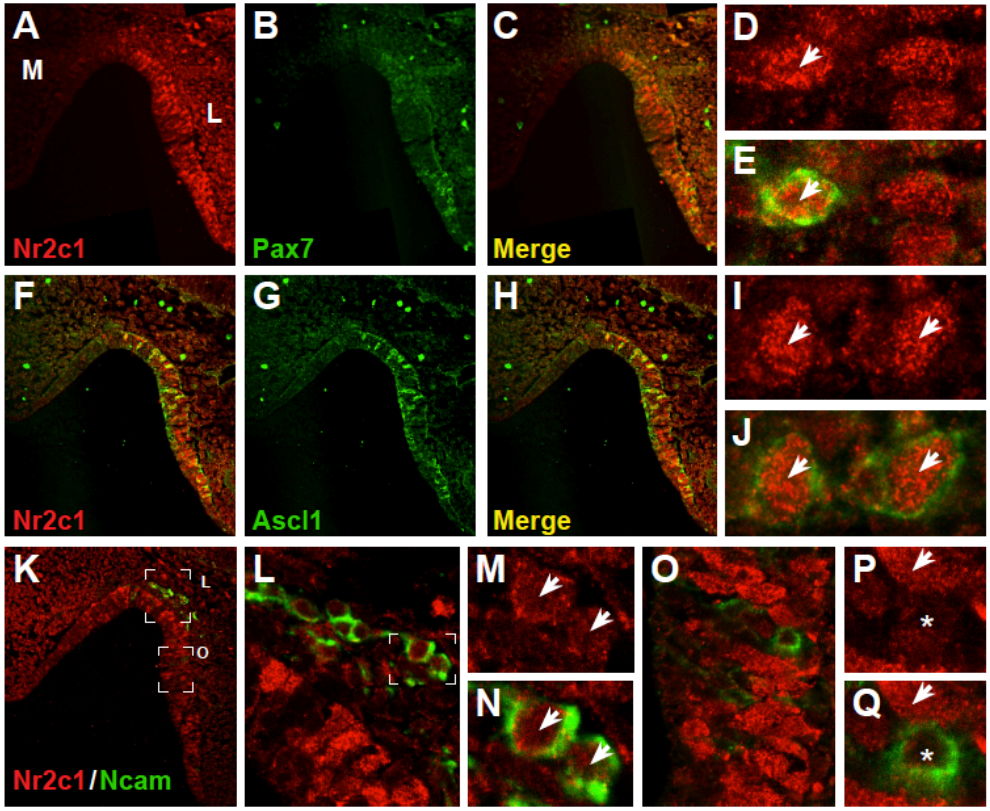
NR3C2

Substitution	Domain
H28R	NTD
H68Q	NTD
R159T	NTD
I180V	NTD

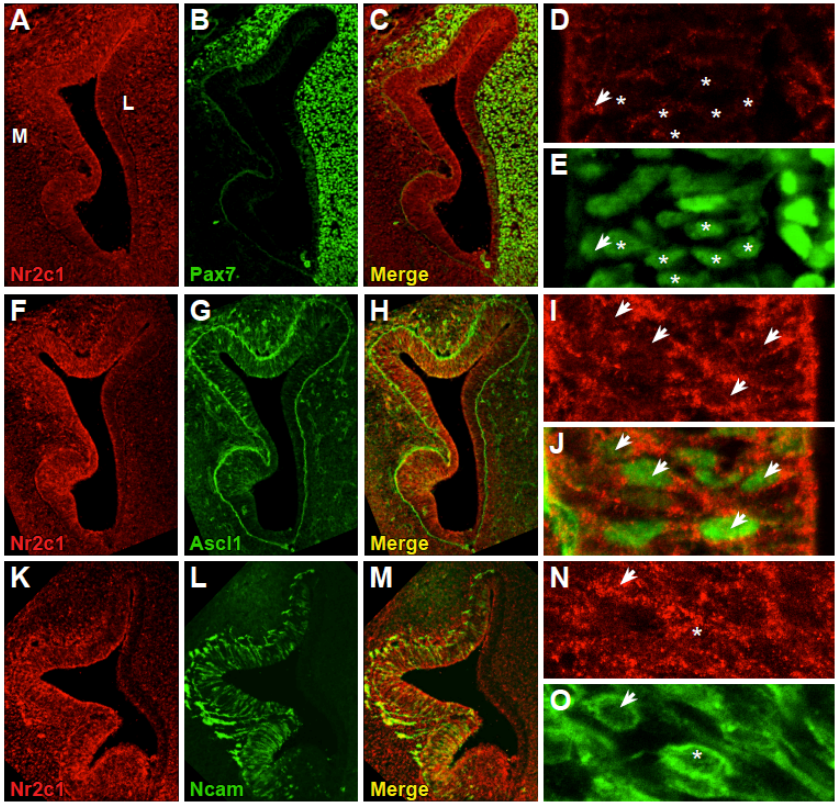
Chapter 2, Figure 6. Amino acid substitutions and their domain locations



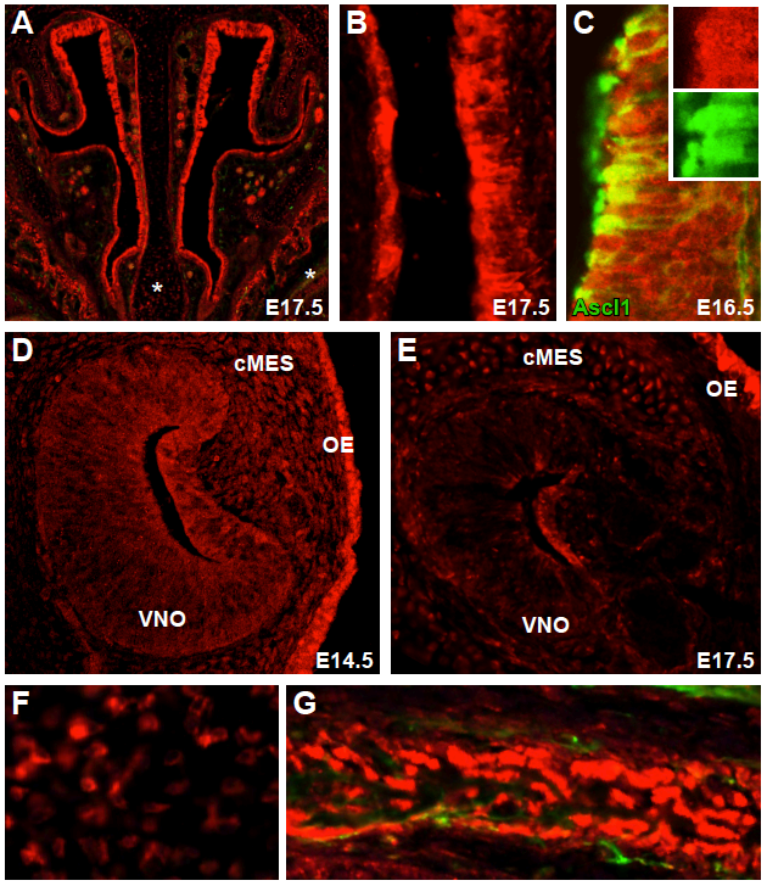
Chapter 3, Figure 1. Expression of *Nr2c1* in the mouse embryo at mid-gestation



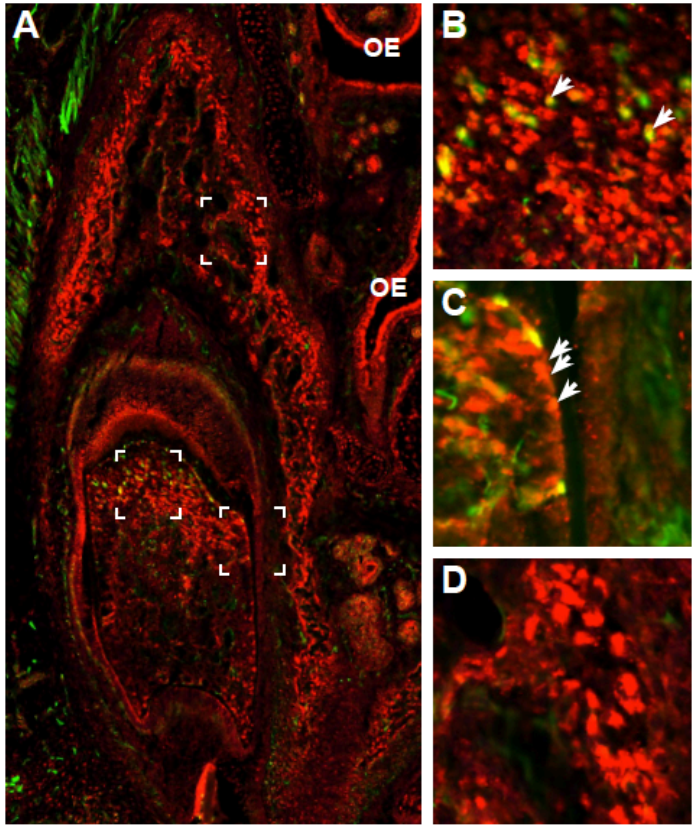
Chapter 3, Figure 2. *Nr2c1* and markers of stem cell populations in mouse OE at E10.5



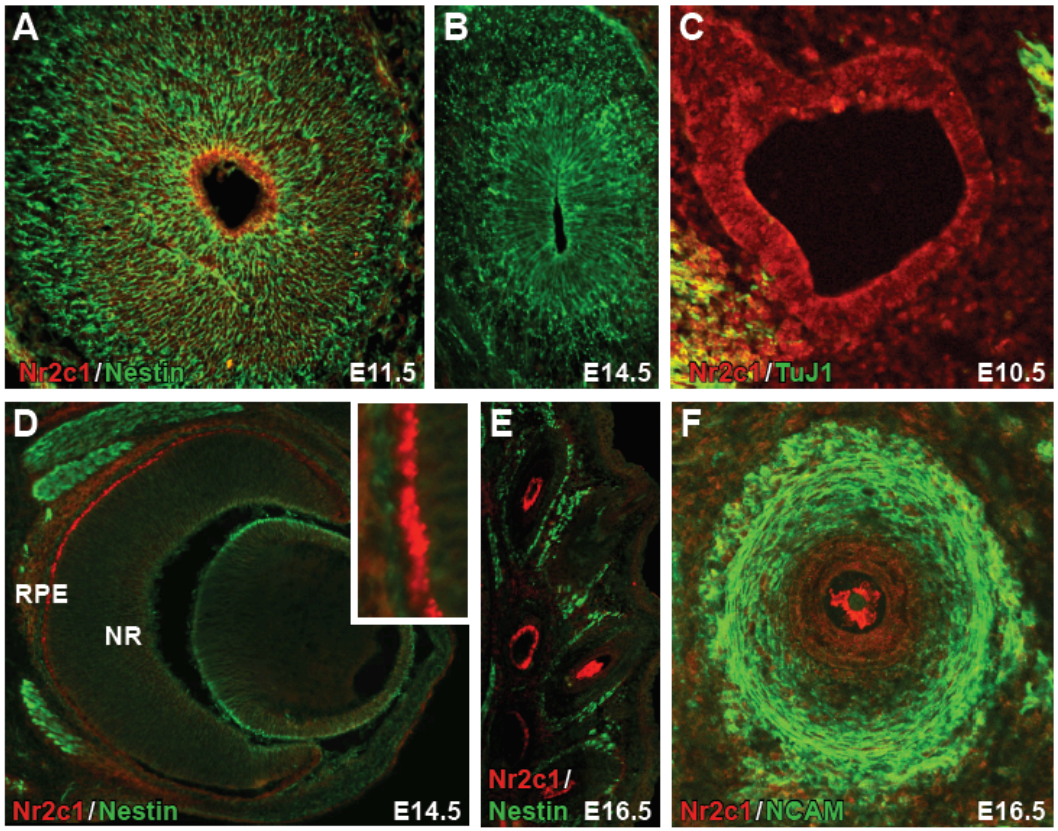
Chapter 3, Figure 3. *Nr2c1* and markers of stem cell populations in mouse OE at E11.5



Chapter 3, Figure 4. *Nr2c1* expression



Chapter 3, Figure 5. Expression of *Nr2c1* in a developing tooth



Chapter 3, Figure 6. *Nr2c1* expression in the nascent olfactory bulb and sensory organs

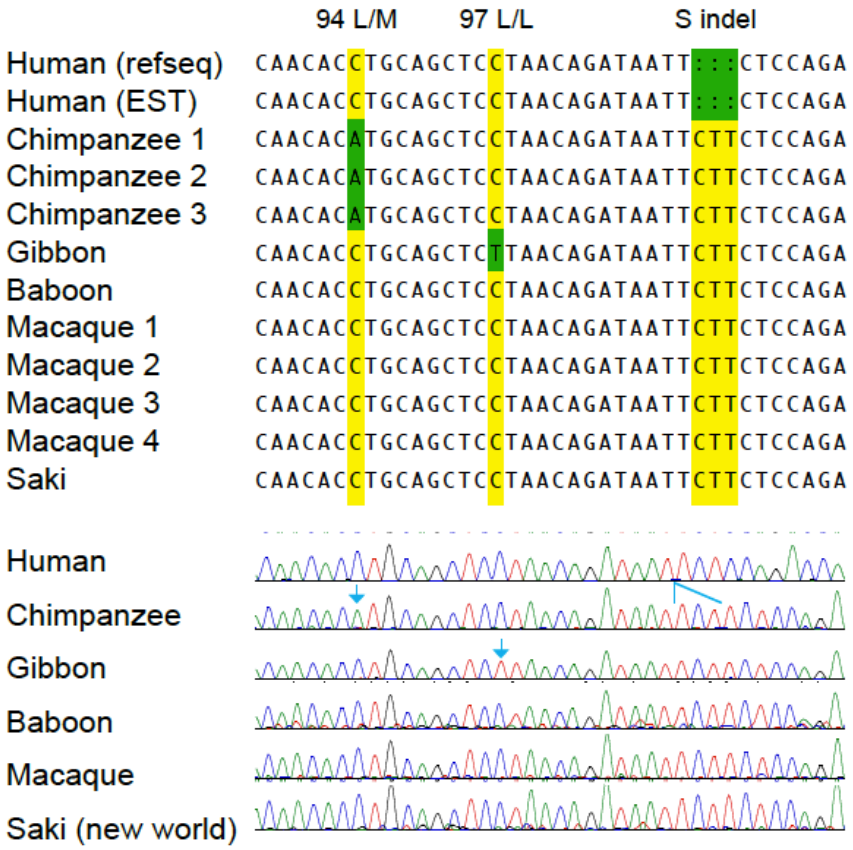
NR2C1 Human substitutions

Substitution	Domain
S101-102INDEL	NTD
V242M	Hinge
T254A	Hinge
S274N	Hinge
G514S	LBD

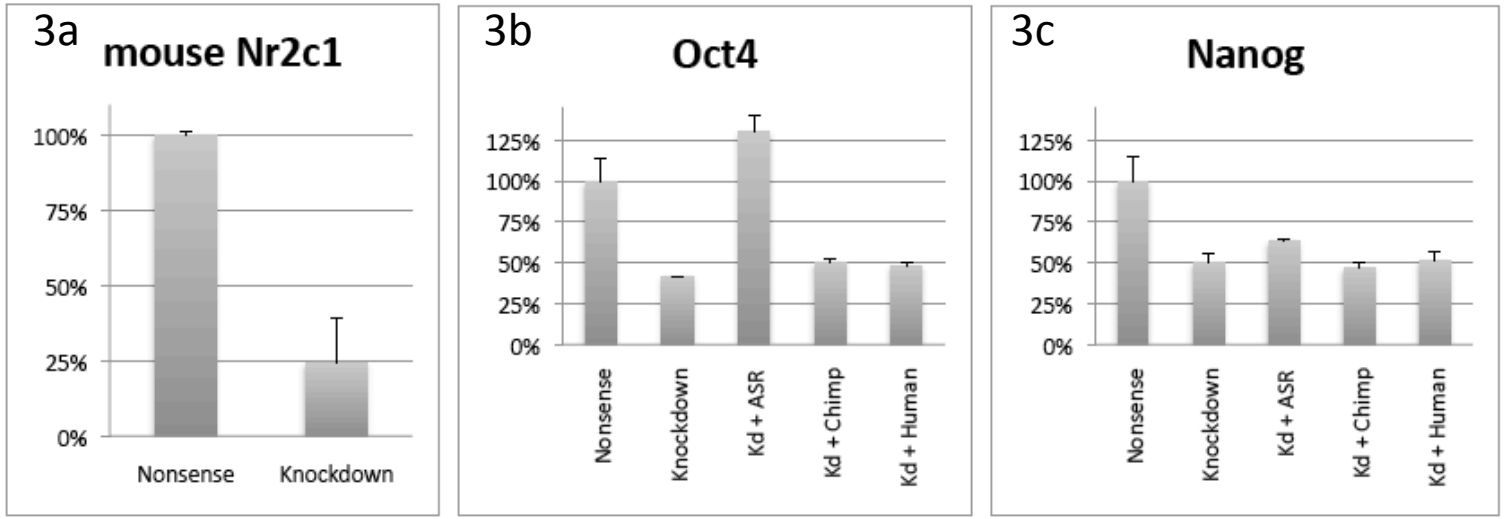
NR2C1 Chimpanzee substitutions

Substitution	Domain
H37R	NTD
N38T	NTD
L94M	NTD
S148A	DBD
T225A	Hinge

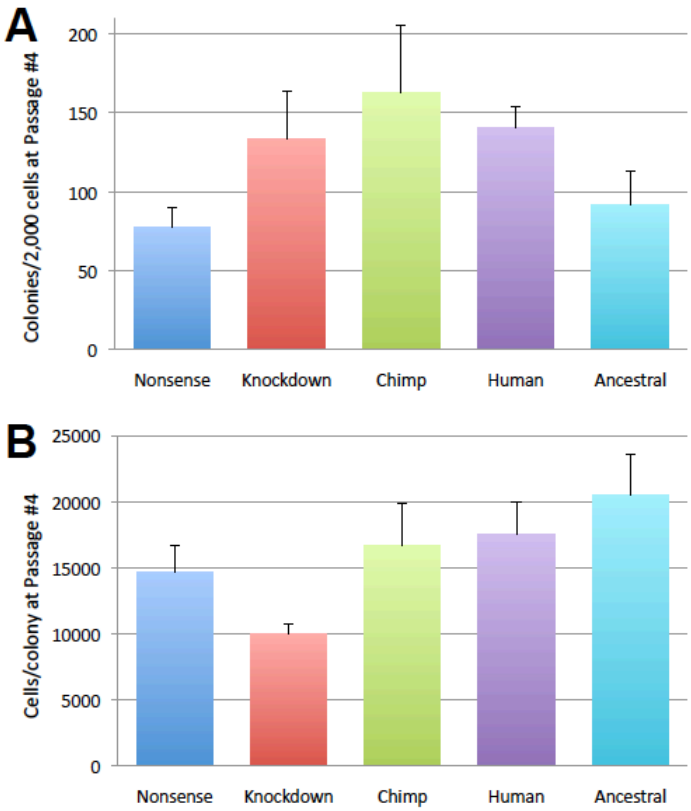
Chapter 4, Figure 1. Validated substitutions in chimpanzees and humans



Chapter 4, Figure 2. Sequencing of *Nr2c1* transcript cDNAs from primate cortex



Chapter 4, Figure 3. Effect of *Nr2c1* knockdown on pluripotentiality-related expression



Chapter 4, Figure 4. Colony formation as a test for pluripotentiality and proliferation

Chapter 2, Table 1. Significant LRTs for pre-specified models of temporal change in selection pressure before and after controlling the FDR

	Gene	$2\delta l$	p	q	ω_{BG}	ω_{FG}	Δ
Episodic models:							
$\mathbf{H_1}$	*NR0B2*	6.502	0.011	0.131	0.207	0	-
	VDR	4.556	0.033	0.131	0.087	0	-
	PPARD	4.250	0.039	0.131	0.039	0	-
$\mathbf{H_2}$	none	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
$\mathbf{H_3}$	***NR2C1***	10.30	0.001	**0.040**	0.093	999	+
	NR3C2	8.020	0.005	0.069	0.099	0.952	+
	PGR	7.155	0.007	0.079	0.195	1.366	+
	ESRRB	6.938	0.008	0.084	0.031	0.174	+
	PPARG	6.204	0.013	0.096	0.045	0.291	+
	NR2E3	5.704	0.017	0.101	0.09	0.532	+
	NR0B1	4.352	0.037	0.162	0.245	999	+
Shift models:							
$\mathbf{H_4}$	***RORA***	45.140	1.83e-11	**1.83e-10**	0.028	0.338	+
	ESRRB	13.116	0.0003	**0.0015**	0.027	0.119	+
	NR2C1	10.342	0.0013	**0.0034**	0.088	0.346	+
	RARG	10.280	0.0013	**0.0034**	0.016	0.083	+
	NR5A2	5.560	0.0184	**0.0367**	0.061	0	-
	ESRRA	5.114	0.0237	**0.0396**	0.018	0.074	+
	NR3C2	4.522	0.0335	**0.0478**	0.097	0.192	+
$\mathbf{H_5}$	***NR2C1***	13.384	0.0002	**0.0061**	0.090	0.752	+
	PGR	11.450	0.0007	**0.0093**	0.003	0.957	+
	NR2E3	9.760	0.0018	**0.0214**	0.088	0.698	+
	ESRRB	7.856	0.0051	**0.0405**	0.030	0.161	+
	PPARG	4.624	0.0315	0.1583	0.045	0.208	+
	NR3C2	4.236	0.0396	0.1583	0.099	0.386	+

Notes: The foreground (FG) and background (FG) ω parameters for branches are fully specified for each hypothesis in Figure 2. The null model for all LRTs assumes homogenous selection pressure over branches ($\omega_{BG} = \omega_{FG}$). df =1 for each LRT. The q-value is the expected proportion of false discoveries expected if the single-test p-value is used to control the FDR. Δ denotes the sign of the difference between ω_{BG} and ω_{FG}.

Chapter2, Table 2. Significant LRTs for a subset of sites having experienced an episodic alteration of selection pressure

Gene	$2\delta l$	p	q	p_i	ω_i
H_1: Great Ape					
LRT1: Model A (ω_{FG}=1) vs. Model A (ω_{FG}>1)					
none	n.a.	n.a.	n.a.	n.a.	n.a.
LRT2: M3(k=2) vs Model B					
NR0B2	7.07	0.0292	0.7012	$p_{2(a+b)}$ = (0.35+0.65)	ω_{FG} = 0.0
				p_1 = 0.00	ω_1 = 0.59
				p_0= 0.00	ω_0 = 0.05
H_2: Human-Chimpanzee					
LRT1: Model A (ω_{FG}=1) vs. Model A (ω_{FG}>1)					
none	n.a.	n.a.	n.a.	n.a.	n.a.
LRT2: M3(k=2) vs. Model B					
none	n.a.	n.a.	n.a.	n.a.	n.a.
H_3: Human					
LRT1: Model A (ω_{FG}=1) vs. Model A (ω_{FG}>1)					
NR1D1	10.27	0.0013	0.0650	$P_{FG(a+b)}$ = 0.002	ω_{FG} = 99
				p_1 = 0.05	[ω_1 = 1]
				p_0= 0.95	ω_0 = 0.04
LRT2: M3(k=2) vs. Model B					
NR1D1	**15.25**	**0.0005**	**0.0234**	$P_{FG(a+b)}$ = 0.01	ω_{FG} = 99
				p_1 = 0.05	ω_1 = 0.86
				p_0= 0.94	ω_0 = 0.04
PPARG	9.71	0.0077	0.1867	$P_{FG(a+b)}$ = 0.01	ω_{FG} = 35
				p_1 = 0.11	ω_1 = 0.46
				p_0= 0.88	ω_0 = 0.0
NR2C1	8.70	0.0129	0.2063	$P_{FG(a+b)}$ = (0.24+0.76)	ω_{FG} =99
				p_1 = 0.00	ω_1 = 0.33
				p_0= 0.00	ω_0 = 0.02
PGR	7.67	0.0215	0.2586	$P_{FG(a+b)}$ = (0.12+0.05)	ω_{FG} = 6.23
				p_1 = 0.24	ω_1 = 0.58
				p_0= 0.59	ω_0 = 0.05

Notes: Genes having a q-value < 0.05 are shown in bold. The foreground (FG) branches are fully specified for each hypothesis in Figure 2. The null model for all LRTs assumes homogenous selection pressure for all branches ($\omega_{BG} = \omega_{FG}$). LRT 1 has df =1 . LRT 2 has df =2. The q-value is the expected proportion of false discoveries expected if the single-test p-value is used is used as the boundary to control the FDR. The parameter p represents the proportion of sites subject to a change in selection intensity.

Chapter 2, Table 3. Robustness analyses for episodic evolution at a subset of sites

Gene	H_i	analysis	p_{FG}	ω_{FG}	p_1	ω_1	p_0	ω_0	LRT
NR1 D1	H_3	original	0.002	99	0.054	0.86	0.944	0.04	p=0.0005
		gene tree	match	match	match	match	match	match	n.a.
		MG94	0.002	99	0.067	0.77	0.931	0.04	p=0.0005
		bootstrap	[0-0.141][b]	99*	[0.019-0.116]	[0.315-1.581]	[0.79-0.982]	[0.017-0.051]	n.a.

Notes: Robustness and bootstrapping analyses were carried out for genes having at least one q-value < 0.05. For analyses under the gene tree, a "match" designation indicates that the gene tree was identical to the organismal tree. [b] Bimodal distribution * distribution was at the bound of 99

Chapter 2, Table 4. Significant LRTs for a subset of sites having experienced a shift in the intensity of selection pressure

Gene	$2\delta l$	p	q	p_{SHIFT}	ω_{BG}	ω_{FG}
H_4: Great Ape Clade						
LRT 1: M2a-rel vs. Model C						
RORA	**47.02**	**7.04e-12**	**1.55E-10**	**0.11**	**0.26**	**3.40**
RARG	**15.88**	**6.76e-05**	**0.0007**	**0.90**	**0.00**	**0.08**
ESRRB	**11.89**	**0.0006**	**0.0041**	**0.27**	**0.11**	**0.44**
NR2C1	**9.52**	**0.0020**	**0.0112**	**0.38**	**0.23**	**0.91**
ESRRA	**6.66**	**0.0098**	**0.0433**	**0.75**	**0.00**	**0.08**
PGR	5.67	0.0173	0.0633	0.71	0.05	0.18
NR5A2	4.59	0.0321	0.0946	0.30	0.21	0.00
NR4A3	4.47	0.0344	0.0946	0.19	0.39	0.10
THRA	4.02	0.0449	0.1098	0.025	1.91	0.00
LRT 2: M3(k=2) vs Model D(k=2)						
RORA	**47.02**	**7.04e-12**	**1.69e-10**	**0.11**	**0.26**	**3.40**
RARG	**15.88**	**6.76e-05**	**0.0008**	**0.90**	**0.00**	**0.08**
ESRRB	**11.89**	**0.0006**	**0.0045**	**0.28**	**0.11**	**0.44**
NR2C1	**9.59**	**0.0020**	**0.0117**	**0.81**	**0.03**	**0.28**
ESRRA	5.86	0.0155	0.0690	0.75	0.00	0.08
PGR	5.67	0.0173	0.0690	0.71	0.05	0.18
NR5A2	4.57	0.0325	0.0967	0.28	0.23	0.00
NR4A3	4.47	0.0344	0.0967	0.19	0.39	0.10
NR1H2	4.38	0.0363	0.0967	0.89	0.02	0.00
NR3C2	4.19	0.0407	0.0976	0.88	0.04	0.12
H_5: Human-Chimpanzee Clade						
LRT 1: M2a-rel vs. Model C						
NR2C1	**12.57**	**0.0004**	**0.0126**	**0.36**	**0.25**	**2.06**
NR1D1	**8.70**	**0.0032**	**0.0384**	**0.002**	**0.00**	**126.89**
PGR	**8.47**	**0.0036**	**0.0384**	**0.30**	**0.54**	**2.57**
NR2E3	**7.90**	**0.0049**	**0.0395**	**0.22**	**0.41**	**3.39**
ESRRB	**7.23**	**0.0071**	**0.0458**	**0.27**	**0.12**	**0.61**
PPARG	5.01	0.0251	0.1341	0.12	0.43	1.86
ESR2	4.07	0.0437	0.1820	0.74	0.04	0.40
NR3C2	4.00	0.0455	0.1820	0.12	0.52	2.71
LRT 2: M3(k=2) vs Model D(k=2)						
NR2C1	**12.55**	**0.0004**	**0.0111**	**0.35**	**0.26**	**2.13**
NR2E3	**9.38**	**0.0022**	**0.0306**	**0.22**	**0.41**	**3.39**
PGR	**8.47**	**0.0036**	**0.0336**	**0.30**	**0.54**	**2.57**
ESRRB	7.23	0.0071	0.0500	0.27	0.12	0.61
PPARG	5.01	0.0251	0.1408	0.12	0.43	1.86
ESR2	4.12	0.0424	0.1629	0.79	0.05	0.40

NR3C2	4.03	0.0448	0.1629	0.11	0.54	2.72
NR1D1	3.96	0.0465	0.1629	0.06	0.81	4.88

Note: Genes having a q-value < 0.05 are shown in bold. The foreground (FG) branches are fully specified for each hypothesis in Figures 2. The null model for all LRTs assumes homogenous selection pressure for all branches ($\omega_{BG} = \omega_{FG}$). LRT 1 has df =1 . LRT 2 has df =1. The q-value is the expected proportion of false discoveries expected if the single-test p-value is used as the boundary to control the FDR. pSHIFT is the fraction of sites subject to a shift in selection pressure.

Chapter 2, Table 5. Robustness of inferences about long-term shifts in the intensity of selection pressure at the origin of the great apes (H_4)

Gene	analysis	Model C				Model D			
		p_{SHIFT}	ω_{BG}	ω_{FG}	LRT	p_{SHIFT}	ω_{BG}	ω_{FG}	LRT
RORA	original	0.11	0.26	3.40	p=7.0e-12	0.11	0.26	3.40	p=7.0e-12
	MG94	0.11	0.30	3.90	p=8.0e-12	0.11	0.30	3.90	p=8.0e-12
	gene tree	0.12	0.24	3.03	p=1.9e-11	0.12	0.24	3.03	p=1.9e-11
	bootstrap	[0.04-0.17]	[0.04-0.42]	[0.83-9.53]	n.a.	[0.06-0.17]	[0.11-0.43]	[1.44-8.32]	n.a.
RARG	original	0.90	0	0.08	p=6.8e-05	0.90	0	0.08	p=6.8e-05
	MG94	0.89	0	0.10	p=7.1e-05	0.89	0	0.10	p=7.1e-05
	gene tree	0.90	0	0.08	p=6.8e-05	0.90	0	0.08	p=6.8e-05
	bootstrap	[0.02-0.96][b]	[0-6e-3]	[0.02-3.06]	n.a.	[0.77-0.99]	[0-8e-3]	[0.03-0.15]	n.a.
ESRRB	original	0.27	0.11	0.44	p=0.0006	0.27	0.11	0.44	p=0.0006
	MG94	0.27	0.14	0.59	p=0.0007	0.27	0.15	0.59	p=0.0007
	gene tree	0.27	0.11	0.44	p=0.0006	0.27	0.11	0.44	p=0.0006
	bootstrap	[0.14-0.75]	[0-0.16]	[0.07-1.00]	n.a.	[0.14-0.45]	[0.04-0.17]	[0.19-1.07]	n.a.
NR2C1	original	0.38	0.23	0.91	p=0.002	0.81	0.03	0.28	p=0.002
	MG94	0.37	0.25	1.02	p=0.001	0.36	0.27	1.0	p=0.002
	gene tree	match	match	match	n.a.	match	match	match	n.a.
	bootstrap	[0.05-0.71]	[5e-3-0.36]	[0.21-2.86]	n.a.	[0.65-0.94]	[0-0.05]	[0-0.70]	n.a.
ESRRA	original	0.75	0	0.08	p=0.010	0.75	0	0.08	p=0.015

MG94	0.69	0	0.09	p=0.010	0.69	0	0.09	p=0.014
gene tree	0.75	0	0.08	p=0.008	0.75	0	0.08	p=0.013
bootstrap	[0.24-1]	[0-0.03]	[0-0.21]	n.a.	[0.162-1][b]	[0-0.02]	[0-0.35]	n.a.

Notes: Robustness and bootstrapping analyses were carried out for genes having at least one q-value < 0.05. For analyses under the gene tree, a "match" designation indicates that the gene tree was identical to the organismal tree. [b] Bimodal distribution

Chapter 2, Table 6. Robustness of inferences about long-term shifts in the intensity of selection pressure at the origin of the human-chimpanzee clade (H_5)

		Model C				Model D			
Gene	**analysis**	p_{SHIFT}	ω_{BG}	ω_{FG}	**LRT**	p_{SHIFT}	ω_{BG}	ω_{FG}	**LRT**
NR2C1	original	0.36	0.25	2.06	p=0.0004	0.35	0.26	2.13	p=0.0004
	MG94	0.37	0.25	1.02	p=0.0015	0.36	0.27	1.04	p=0.0017
	gene tree	match	match	match	n.a.	match	match	match	n.a.
	bootstrap	[0.17-0.57]	[0.11-0.42]	[0.31-8.32]	n.a.	[0.18-0.47]	[0.18-0.41]	[0.4-9.41]	n.a.
NR1D1	original	0.002	0	126.89	p=0.0032	0.06	0.81	4.88	p=0.0465
	MG94	0.002	0	162.68	p=0.0057	0.07	0.74	4.19	p=0.0511
	gene tree	match	match	match	n.a.	match	match	match	n.a.
	bootstrap	[0-8e-3]	[0-2.12]	[8.84-99]	n.a.	[0.02-0.13]	[0.38-1.43]	[0-22.31]	n.a.
PGR	original	0.30	0.54	2.57	p=0.0036	0.30	0.54	2.57	p=0.0036
	MG94	0.30	0.59	2.70	p=0.0049	0.30	0.60	2.68	p=0.0049
	gene tree	0.31	0.54	2.54	p=0.0011	0.31	0.54	2.54	p=0.0038
	bootstrap	[0.18-0.5]	[0.22, 0.7]	[0.65-5.11]	n.a.	[0.22-0.42]	[0.41-0.68]	[0.77-8.37]	n.a.
NR2E3	original	0.22	0.41	3.39	p=0.0049	0.22	0.41	3.39	p=0.0022
	MG94	0.21	0.45	3.97	p=0.0051	0.21	0.45	3.94	p=0.0015
	gene tree	***	***	***	n.a.	***	***	***	n.a.
	bootstrap	[0.13-0.36]	[0.26-0.71]	[0.36-99]	n.a.	[0.11-0.34]	[0.27-0.69]	[0.6-13.98]	n.a.
ESRRB	original	0.27	0.12	0.61	p=0.0071	0.27	0.12	0.61	p=0.0071
	MG94	0.26	0.17	0.80	p=0.0081	0.26	0.17	0.80	p=0.0081
	gene tree	0.27	0.12	0.61	p=0.0076	0.27	0.13	0.61	p=0.0076
	bootstrap	[0.15-0.84][b]	[0-0.21]	[0.07-1.92]	n.a.	[0.1-0.40]	[0.06-0.22]	[0.07-1.84]	n.a.

Notes: Robustness and bootstrapping analyses were carried out for genes having at least one *q*-value < 0.05. For analyses under the gene tree, a "match" designation indicates that the gene tree was identical to the organismal tree. Analyses that were impossible because the gene tree topology prevented specification of H_5 are indicated by "***". [b] Bimodal distribution

Chapter 3, Table 1. Sequences of oligonucleotide primers used for qPCR analysis (listed as synthesized, in 5'-3' orientation).

Transcript	Forward primer	Reverse primer	Amplicon size (bp)
Gapdh	CTGACGTGCCGCCTGGAGAAA	GTTGGGGGCCGAGTTGGGATAGG	346
Nr2c1 (A)	CTGCGGTGGGGATATGCCTGTG	GCTACGTGCGACTCGCTCAGCAG	252
Nr2c1 (B)	GACAAAGCATCAGGGCGTCATTAC	GGCGTTGCGGCTAGAGGGCTTC	333

Chapter4, Table 1. Oligonucleotide primers used for partial sequencing of *Nr2c1* from primate cDNA (listed in 5' – 3' orientation)

	Forward primer	Reverse primer
Oct4	GCCCGAAGCCCTCCCTACAGCAG	GGAAAGGCCTCGCCCTCAGGAAAAG
Nanog	CCTCTTCAAGGCAGCCCTGATTCTTC	TTCTGCCACCGCTTGCACTTCATC
Gapdh	CTGACGTGCCGCCTGGAGAAA	GTTGGGGGCCGAGTTGGGATAGG

Chapter 4, Table 2. Oligonucleotide primers used for qPCR analysis of expression (listed in 5' – 3' orientation)

	Forward primer	Reverse primer
3c	AATTATTGAACAACAGATGGGAGAG	ATCCTTTGCAGCCTTCACAAGTTAC
4a	CTGGCCAGGCAAGATTCCACTC	TTGCAGTTAATGGGCTACGAAGGTC
5a	TAGCCCATTAACTGCAACTCCAAC	TCCTTCCATGCCCGCTACTGAG
6b	GCACTGGGCACTTTCGATTCCTTC	AAGCTGGCAATCTGAGTAGTAGTCTG

References

Akanuma S, Nakajima Y, Yokobori S, Kimuraa M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A., 2013. Experimental evidence for the thermophilicity of ancestral life. Proc Natl Acad Sci USA 110(27):11067–11072.

Akey, J.M., 2009. Constructing genomic maps of positive selection in humans: where do we go from here? Genome research. 19, 711–722.

Alemseged, Z., Spoor, F., Kimbel, W.H., Bobe, R., Geraads, D., Reed, D., Wynn, J.G., 2006. A juvenile early hominin skeleton from Dikika, Ethiopia. Nature. 443, 296–301.

Alföldi, J., Lindblad-Toh, K., 2013. Comparative genomics as a tool to understand evolution and disease. Genome research. 23, 1063–1068.

Ali, F., Meier, R., 2008. Positive selection in ASPM is correlated with cerebral cortex evolution across primates but not with whole-brain size. Molecular biology and evolution. 25, 2247–2250.

Allman J.M., Tetreault N.A., Hakeem A.Y., Manaye K.F., Semendeferi K., Erwin J.M., Park S., Goubert V., Hof P.R., 2010. The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. Brain Struct Funct. 214(5-6):495-517.

Amstadter, A.B., Sumner, J.A., Acierno, R., Ruggiero, K.J., Koenen, K.C., Kilpatrick, D.G., Galea, S., Gelernter, J., 2013. Support for association of RORA variant and post traumatic stress symptoms in a population-based study of hurricane exposed adults. Mol. Psychiatr. 18, 1148–1149.

Anderson, A.M., Carter, K.W., Anderson, D., Wise, M.J., 2012. Coexpression of nuclear receptors and histone methylation modifying genes in the testis: implications for endocrine disruptor modes of action. PloS one. 7, e34158.

Anderson AM, Carter KW, Anderson D, Wise MJ (2012) Coexpression of Nuclear Receptors and Histone Methylation Modifying Genes in the Testis: Implications for Endocrine Disruptor Modes of Action. PLoS ONE 7(4): e34158.

Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Mol. Biol. Evol. 18, 1585–1592.

Anisimova, M. and D. Liberles 2012. Detecting and understanding natural selection, in Codon Evolution: mechanisms and models, eds. Cannarozzi G, Schneider A., Oxford University Press.

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics. 164, 1229–1236.

Ao, A., Wang, H., Kamarajugadda, S., Lu, J., 2008. Involvement of estrogen-related receptors in transcriptional response to hypoxia and growth of solid tumors. Proc. Natl. Acad. Sci. U.S.A. 105, 7821–7826.

Avise, J.C., Shapira, J.F., Daniel, S.W., Aquadro, C.F., Lansman, R.A., 1983. Mitochondrial DNA differentiation during the speciation process in Peromyscus. Mol. Biol. Evol. 1, 38–56.

Bain, D.L., Heneghan, A.F., Connaghan-Jones, K.D., Miura, M.T., 2007. Nuclear receptor structure: implications for function. Annual review of physiology. 69,

201–220.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E., Gottlieb, D.I., 1995. Embryonic stem cells express neuronal properties in vitro. Dev. Biol. 168, 342–357.

Balmer CW, LaMantia AS. Noses and neurons: induction, morphogenesis, and neuronal differentiation in the peripheral olfactory pathway. Dev Dyn. 2005;234:464–481.

Bao, L., Gu, H., Dunn, K.A., Bielawski, J.P., 2007. Methods for selecting fixed-effect models for heterogeneous codon evolution, with comments on their application to gene and genome data. BMC Evol. Biol. 7. Suppl. 1, S5.

Bao, L., Gu, H., Dunn, K.A., Bielawski, J.P., 2008. Likelihood-based clustering (LiBaC) for codon models, a method for grouping sites according to similarities in the underlying process of evolution. Mol. Biol. Evol. 25, 1995–2007.

Barrett, R.D.H., Hoekstra, H.E., 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature reviews. Genetics. 12, 767–780

Bay, R.A., Bielawski, J.P., 2011. Recombination detection under evolutionary scenarios relevant to functional divergence. J. Mol. Evol. 73, 273–286.

Beehner, J.C., Gesquiere, L., Seyfarth, R.M., Cheney, D.L., Alberts, S.C., Altmann, J., 2009. Testosterone related to age and life-history stages in male baboons and geladas. Horm. Behav. 56, 472–480.

Benjamini, Y., & Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B. Met. 289-300.

Benoit G, Cooney A, Giguere V, Ingraham H, Lazar M, Muscat G, Perlmann T, Renaud JP, Schwabe J, Sladek F, Tsai MJ, Laudet V. 2006. International Union of Pharmacology. LXVI. Orphan nuclear receptors. Pharmacol. Rev. 58(4):798–836.

Bernstein, R.M., 2010. The big and small of it: how body size evolves. Am. J. Phys. Anthropol. 143 Suppl 51, 46–62.

Bielawski, J.P., 2013. Detecting the signatures of adaptive evolution in protein-coding genes. Curr. Protoc. Mol. Biol. Chapter 19, Unit 19.1-21.

Bielawski, J.P., Dunn, K.A., Sabehi, G., Béjà, O., 2004. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. Proc. Natl. Acad. Sci. U.S.A. 101, 14824–14829.

Bielawski, J., Yang, Z., 2004. A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. J. Mol. Evol. 59(1):121-32.

Blom, N., Gammeltoft, S., Brunak, S., 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294, 1351–1362.

Boddy AM, McGowen MR, Sherwood CC, Grossman LI, Goodman M, Wildman DE. 2012. Comparative analysis of encephalization in mammals reveals relaxed constraints on anthropoid primate and cetacean brain scaling. 25, 981–994.

Boiani, M., Eckardt, S., Schöler, H.R., McLaughlin, K.J., 2002. Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes & development. 16, 1209–1219.

Bombail, V., Collins, F., Brown, P., Saunders, P.T.K., 2010. Modulation of ER alpha transcriptional activity by the orphan nuclear receptor ERR beta and evidence for differential effects of long- and short-form splice variants. Mol. Cell. Endocrinol. 314, 53–61.

Boukhtouche, F., Vodjdani, G., Jarvis, C.I., Bakouche, J., Staels, B., Mallet, J., Mariani,

J., Lemaigre-Dubreuil, Y., Brugg, B., 2006. Human retinoic acid receptor-related orphan receptor alpha1 overexpression protects neurones against oxidative stress-induced apoptosis. J. Neurochem. 96, 1778–1789.

Bradley, B.J., Pedersen, A., Mundy, N.I., 2009. Blue eyes in lemurs and humans: same phenotype, different genetic mechanism. Am. J. Phys. Anthropol. 139, 269–273.

Braun, D.R., Harris, J.W.K., Levin, N.E., McCoy, J.T., Herries, A.I.R., Bamford, M.K., Bishop, L.C., Richmond, B.G., Kibunjia, M., 2010. Early hominin diet included diverse terrestrial and aquatic animals 1.95 Ma in East Turkana, Kenya. Proceedings of the National Academy of Sciences of the United States of America. 107, 10002–10007.

Brayer, K.J., Lynch, V.J., Wagner, G.P., 2011. Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proceedings of the National Academy of Sciences 108, E414–20.

Bridgham, J.T., Ortlund, E.A., Thornton, J.W., 2009. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461, 515–519.

Buckner, R.L., Krienen, F.M., 2013. The evolution of distributed association networks in the human brain. Trends in cognitive sciences. 17, 648–665.

Carroll, S.M., Bridgham, J.T., Thornton, J.W., 2008. Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. Molecular biology and evolution. 25, 2643–2652.

Carroll, S.M., Ortlund, E.A., Thornton, J.W., 2011. Mechanisms for the Evolution of a Derived Function in the Ancestral Glucocorticoid Receptor. PLoS Genet. 7,

e1002117.

Challis JRG, Matthews, S.G., Gibb, W., Lye, S.J., 2000. Endocrine and paracrine regulation of birth at term and preterm. Endocr. Rev. 21, 514–550.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., Smith, A., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell. 113, 643–655.

Chang, B.S.W., 2002. Recreating a functional ancestral archosaur visual pigment. Mol. Biol. Evol. 19, 1483–1489.

Chang B.S.W., Jönsson K., Kazmi M.A., Donoghue M.J., Sakmar T.P., 2002. Recreating a functional ancestral archosaur visual pigment. Mol Biol Evol 19(9):1483–1489.

Chang C, Da Silva SL, Ideta R, Lee Y, Yeh S, Burbach JP. 1994. Human and rat TR4 orphan receptors specify a subclass of the steroid receptor superfamily. Proc Natl Acad Sci U S A. Jun 21;91(13):6040-4.

Chang, C., Kokontis, J., 1988. Identification of a new member of the steroid receptor super-family by cloning and sequence analysis. Biochemical and biophysical research communications. 155, 971–977.

Chen, C., Opazo, J.C., Erez, O., Uddin, M., Santolaya-Forgas, J., Goodman, M., Grossman, L.I., Romero, R., Wildman, D.E., 2008. The human progesterone receptor shows evidence of adaptive evolution associated with its ability to act as a transcription factor. Mol. Phylogenet. Evol. 47, 637–649.

Chen, J., McKay, R.M., Parada, L.F., 2012. Malignant glioma: lessons from genomics, mouse models, and stem cells. Cell. 149, 36–47.

Chen, J., Rattner, A., Nathans, J., 2005. The rod photoreceptor-specific nuclear receptor Nr2e3 represses transcription of multiple cone-specific genes. J. Neurosci. 25, 118–129.

Cheung, C.P., Yu, S., Wong, K.B., Chan, L.W., Lai, F.M.M., Wang, X., Suetsugi, M., Chen, S., Chan, F.L., 2005. Expression and functional study of estrogen receptor-related receptors in human prostatic cells and tissues. J. Clin. Endocrinol. Metab. 90, 1830–1844.

Condon, J.C., Jeyasuria, P., Faust, J.M., Wilson, J.W., Mendelson, C.R., 2003. A decline in the levels of progesterone receptor coactivators in the pregnant uterus at term may antagonize progesterone receptor function and contribute to the initiation of parturition. Proc. Natl. Acad. Sci. U.S.A. 100, 9518–9523.

Councill, J.H., Tucker, E.S., Haskell, G.T., Maynard, T.M., Meechan, D.W., Hamer, R.M., Lieberman, J.A., LaMantia, A.S., 2006. Limited influence of olanzapine on adult forebrain neural precursors in vitro. Neuroscience. 140, 111–122.

Csapo, A.I., 1961. The onset of labour. Lancet. 2, 277–280.

Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J., Tashian, R.E., 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature. 298, 297–300

Darwin C. 1871. The Descent of Man, and Selection in Relation to Sex. New York: D Appleton and Company.

De Kloet, E.R., 2000. Stress in the brain. Eur. J. Pharmacol. 405, 187–198.

De Kloet, E.R., Joëls, M., Holsboer, F., 2005. Stress and the brain: from adaptation to disease. Nat. Rev. Neurosci. 6, 463–475.

De Kloet, E.R., Vreugdenhil, E., Oitzl, M.S., Joëls, M., 1998. Brain corticosteroid receptor balance in health and disease. Endocr. Rev. 19, 269–301.

Diogo, R., Wood, B., 2011. Soft-tissue anatomy of the primates: phylogenetic analyses based on the muscles of the head, neck, pectoral region and upper limb, with notes on the evolution of these muscles. Journal of anatomy. 219, 273–359.

Domínguez-Rodrigo, M., Pickering, T.R., Semaw, S., Rogers, M.J., 2005. Cutmarked bones from Pliocene archaeological sites at Gona, Afar, Ethiopia: implications for the function of the world's oldest stone tools. J. Hum. Evol. 48, 109–121.

Downs, K.M., Davies, T., 1993. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. Development (Cambridge, England). 118, 1255–1266.

Eick, G.N., Thornton, J.W., 2011. Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. Molecular and cellular endocrinology. 334, 31–38.

Emery Thompson, M., Muller, M.N., Kahlenberg, S.M., Wrangham, R.W., 2010. Dynamics of social and energetic stress in wild female chimpanzees. Horm. Behav. 58, 440–449.

Enard, D., Messer, P.W., Petrov, D.A., 2014. Genome-wide signals of positive selection in human evolution. Genome research. 24, 885–895.

Enard, W., 2014. Comparative genomics of brain size evolution. Frontiers in human neuroscience. 8, 345.

Enmark, E., Gustafsson, J.Å., 1996. Orphan nuclear receptors--the first eight years. Mol. Endocrinol. 10, 1293–1307.

Evans, P.D., Anderson, J.R., Vallender, E.J., Choi, S.S., Lahn, B.T., 2004.

Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. Human molecular genetics. 13, 1139–1145.

Evans, P.D., Vallender, E.J., Lahn, B.T., 2006. Molecular evolution of the brain size regulator genes CDK5RAP2 and CENPJ. Gene. 375, 75–79.

Falk, D., Redmond, J.C., Guyer, J., Conroy, C., Recheis, W., Weber, G.W., Seidler, H., 2000. Early hominid brain evolution: a new look at old endocasts. Journal of human evolution. 38, 695–717.

Ferraro, J.V., Plummer, T.W., Pobiner, B.L., Oliver, J.S., Bishop, L.C., Braun, D.R., Ditchfield, P.W., Seaman, J.W., Binetti, K.M., Seaman, J.W., Hertel, F., Potts, R., 2013. Earliest archaeological evidence of persistent hominin carnivory. PLoS ONE. 8, e62174.

Finnigan, G.C., Hanson-Smith, V., Stevens, T.H., Thornton, J.W., 2012. Evolution of increased complexity in a molecular machine. Nature. 481, 360–364.

Fiorucci, S., Distrutti, E., Bifulco, G., D'Auria, M.V., Zampella, A., 2012. Marine sponge steroids as nuclear receptor ligands. Trends Pharmacol. Sci.. 33, 591–601.

Fischer, K., Kelly, S.M., Watt, K., Price, N.C., McEwan, I.J., 2010. Conformation of the mineralocorticoid receptor N-terminal domain: evidence for induced and stable structure. Mol. Endocrinol. 24, 1935–1948.

Fitch, W. M., 1970. Distinguishing homologous and analogous proteins, Syst. Zool. 19, 99-113.

Fitch, W.M., Leiter, J.M., Li, X.Q., Palese, P., 1991. Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. U.S.A. 88, 4270–4274.

Fletcher, W., Yang, Z., 2010. The effect of insertions, deletions, and alignment errors on

the branch-site test of positive selection. Mol. Biol. Evol. 27, 2257–2267.

Florio, M., Huttner, W.B., 2014. Neural progenitors, neurogenesis and the evolution of the neocortex. Development (Cambridge, England). 141, 2182–2194.

Foley, R.A., Lee, P.C., 1991. Ecology and energetics of encephalization in hominid evolution. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 334, 223–31– discussion 232.

Fonseca-Azevedo, K., Herculano-Houzel, S., 2012. Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. Proceedings of the National Academy of Sciences of the United States of America. 109, 18571–18576.

Fontenot, M., Konopka, G., 2014. Molecular networks and the evolution of human cognitive specializations. Current opinion in genetics & development. 29, 52–59.

Forsberg, R., Christiansen, F.B., 2003. A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. Mol. Biol. Evol. 20, 1252–1259.

Garriock, H.A., Kraft, J.B., Shyn, S.I., Peters, E.J., Yokoyama, J.S., Jenkins, G.D., Reinalda, M.S., Slager, S.L., McGrath, P.J., Hamilton, S.P., 2010. A genomewide association study of citalopram response in major depressive disorder. Biol. Psychiat. 67, 133–138.

Germain, P., Staels, B., Dacquet, C., Spedding, M., Laudet, V., 2006. Overview of nomenclature of nuclear receptors. Pharm. Rev. 58, 685–704.

Geschwind, D.H., Konopka, G., 2012. Neuroscience: Genes and human brain evolution. Nature. 486, 481–482.

Geschwind DH, Rakic P., 2013. Cortical evolution: judge the brain by its cover. Neuron. 80, 633-47.

Gesquiere, L.R., Khan, M., Shek, L., Wango, T.L., Wango, E.O., Alberts, S.C., Altmann, J., 2008. Coping with a challenging environment: effects of seasonal variability and reproductive status on glucocorticoid concentrations of female baboons (Papio cynocephalus). Horm. Behav. 54, 410–416.

Giangrande, P.H., McDonnell, D.P., 1999. The A and B isoforms of the human progesterone receptor: two functionally different transcription factors encoded by a single gene. Recent Prog. Horm. Res. 54, 291–314.

Giguere V. 1999. Orphan nuclear receptors: from gene to function. *Endocr Rev* 20:689–725.

Glass, C.K., Rosenfeld, M.G., 2000. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev. 14, 121–141.

Goldman, N and Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 11, 725–736.

Gómez-Robles, A., Hopkins, W.D., Sherwood, C.C., 2014. Modular structure facilitates mosaic evolution of the brain in chimpanzees and humans. Nature communications. 5, 4469.

Goodman, M., Czelusniak, J., Moore, G.W., Romero-Herrera, A.E., Matsuda, G. 1979. Fitting the gene linage into its species lineage. A parsimony strategy illustrated by cladograms constructed from globin sequences, Syst. Zool. 28, 132-163.

Graham, J.D., Clarke, C.L., 1997. Physiological action of progesterone in target tissues.

Endocr. Rev. 18, 502–519.

Gronemeyer, H., Gustafsson, J.-A., Laudet, V., 2004. Principles for modulation of the nuclear receptor superfamily. Nat. Rev. Drug Discov. 3, 950–964.

Hansen DV, Lui JH, Flandin P, Yoshikawa K, Rubenstein JL, Alvarez-Buylla A, Kriegstein AR. 2013. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat Neurosci* 16(11):1576-87.

Harms, M.J., Thornton, J.W., 2010. Analyzing protein structure and function using ancestral gene reconstruction. Current opinion in structural biology. 20, 360–366.

Harms, M.J., Thornton, J.W., 2013. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. 14(8), 559-571.

Heneghan, A.F., Connaghan-Jones, K.D., Miura, M.T., 2007. Coactivator assembly at the promoter: efficient recruitment of SRC2 is coupled to cooperative DNA binding by the progesterone receptor. Biochem. 46(39), 11023-11032.

Henson, M.C., 1998. Pregnancy maintenance and the regulation of placental progesterone biosynthesis in the baboon. Hum. Reprod. Update. 4(4), 389-405.

Higham, J.P., Heistermann, M., Saggau, C., 2012. Sexual signalling in female crested macaques and the evolution of primate fertility signals. BMC Evol. Biol. 12:89.

Holloway, R.L., 1968. The evolution of the primate brain: some aspects of quantitative relations. Brain research. 7, 121–172.

Holloway R.L., Broadfield D.C., Yuan M.S., 2004. The Human Fossil Record, Volume 3, Brain Endocasts-The Paleoneurological Evidence (eds Schwartz JH, Tattersal I). New York: Wiley.

Holsboer, F., 2000. The corticosteroid receptor hypothesis of depression.

Neuropsychopharmacol. 23(5), 477-501.

Horst, ter, J.P., van der Mark, M.H., Arp, M., Berger, S., De Kloet, E.R., Oitzl, M.S., 2012. Stress or no stress: mineralocorticoid receptors in the forebrain regulate behavioral adaptation. Neurobiol. Learn. Mem. 98, 33–40.

Hu, Y.C., 2002. Suppression of Estrogen Receptor-mediated Transcription and Cell Growth by Interaction with TR2 Orphan Receptor. J. Biol. Chem. 277, 33571–33579.

Hu, Y.C., Shyr, C.R., Che, W., Mu, X.M., Kim, E., Chang, C., 2002. Suppression of estrogen receptor-mediated transcription and cell growth by interaction with TR2 orphan receptor. The Journal of biological chemistry. 277, 33571–33579.

Hughes, A.L., Nei, M., 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 335, 167–170.

Huxley TH. 1863. Evidence as to Man's Place in Nature. London: Williams and Norgate

Ikeda, K., Ookawara, S., Sato, S., Ando, Z.-I., Kageyama, R., Kawakami, K., 2007. Six1 is essential for early neurogenesis in the development of olfactory epithelium. Developmental biology. 311, 53–68

Ino, H., 2004. Immunohistochemical characterization of the orphan nuclear receptor ROR alpha in the mouse nervous system. J. Histochem. Cytochem. 52, 311–323.

Invergo, B.M., Montanucci, L., Laayouni, H., Bertranpetit, J., 2013. A system-level, molecular evolutionary analysis of mammalian phototransduction. BMC Evol. Biol. 13:52.

Isler, K., Van Schaik, C.P., 2014. How humans evolved large brains: comparative

evidence. Evolutionary anthropology. 23, 65–75.

Ivarsson, Y., 2002. Identification of Residues in Glutathione Transferase Capable of Driving Functional Diversification in Evolution. A novel approach to protein redesign. J. Biol. Chem. 278, 8733–8738.

Jameson, N.M., Hou, Z.-C., Sterner, K.N., Weckle, A., Goodman, M., Steiper, M.E., Wildman, D.E., 2011. Genomic data reject the hypothesis of a prosimian primate clade. Journal of human evolution. 61, 295–305.

Johnson, M.B., Kawasawa, Y.I., Mason, C.E., Krsnik, Z., Coppola, G., Bogdanović, D., Geschwind, D.H., Mane, S.M., State, M.W., Sestan, N., 2009. Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron. 62, 494–509.

Joëls, M., Karst, H., DeRijk, R., De Kloet, E.R., 2008. The coming out of the brain mineralocorticoid receptor. Trends. Neurosci. 31(1), 1-7.

Jolly, S., Journiac, N., Naudet, F., Gautheron, V., Mariani, J., Vernet-der Garabedian, B., 2011. Cell-autonomous and non-cell-autonomous neuroprotective functions of RORα in neurons and astrocytes during hypoxia. J. Neurosci. 31, 14314–14323.

Jungers, W.L., Harcourt-Smith, W.E.H., Wunderlich, R.E., Tocheri, M.W., Larson, S.G., Sutikna, T., Due, R.A., Morwood, M.J., 2009a. The foot of Homo floresiensis. Nature. 459, 81–84.

Jungers, W.L., Larson, S.G., Harcourt-Smith, W., Morwood, M.J., Sutikna, T., Due Awe, R., Djubiantono, T., 2009b. Descriptions of the lower limb skeleton of Homo floresiensis. Journal of human evolution. 57, 538–554.

Kaas, J.H., 2005. The Evolution of Visual Cortex in Primates, in The Primate Visual System: A Comparative Approach (ed J. Kremers), John Wiley & Sons, Ltd, Chichester, UK.

Kaas, J.H., 2008. The evolution of the complex sensory and motor systems of the human brain. Brain research bulletin. 75, 384–390.

Kaas JH. 2013. The evolution of the visual system in primates. In: Werner J, Chalupa L, editors. The visual neurosciences. Cambridge: MIT Press. p 1563–1572.

Kamilar, J.M., Heesy, C.P., Bradley, B.J., 2013. Did trichromatic color vision and red hair color coevolve in primates? Am. J. Primatol. 75, 740–751.

Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., Chambon, P., 1990. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. EMBO J. 9, 1603–1614.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14), 3059-3066.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., 2002. The human genome browser at UCSC. Genome Res. 12(6), 996-1066.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., 1989. Identification of the cystic fibrosis gene: genetic analysis. Science. 245(4922), 1073-1080.

Ketterson, E.D., Atwell, J.W., and J.W. McGlothlin. 2009. Phenotypic integration and independence: hormones, performance, and response to environmental change. Integrative and Comparative Biology 49:365-379.

Kimbel, W.H., 1984. Variation in the pattern of cranial venous sinuses and hominid phylogeny. American journal of physical anthropology. 63, 243–263.

Klok, M.D., Alt, S.R., Lafitte, A.I., Turner, J.D., 2011. Decreased expression of mineralocorticoid receptor mRNA and its splice variants in postmortem brain regions of patients with major depressive disorder. J. Psychiatr. Res. 45(7), 871-878.

Kobayashi, H., Kohshima, S., 2001. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J. Hum. Evol.. 40, 419–435.

Kobayashi, M., Takezawa, S.I., Hara, K., Yu, R.T., Umesono, Y., Agata, K., Taniwaki, M., Yasuda, K., Umesono, K., 1999. Identification of a photoreceptor cell-specific nuclear receptor. Proc. Natl. Acad. Sci. U.S.A.. 96, 4814–4819.

Kohn JA, Deshpande K, Ortlund EA. 2012. Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. J Biol Chem, 287, 16267-16275.

Konopka, G., Friedrich, T., Davis-Turak, J., Winden, K., Oldham, M.C., Gao, F., Chen, L., Wang, G.-Z., Luo, R., Preuss, T.M., Geschwind, D.H., 2012. Human-specific transcriptional networks in the brain. Neuron. 75, 601–617.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics. 22, 3096–3098.

Kouprina, N., Pavlicek, A., Mochida, G.H., Solomon, G., Gersch, W., Yoon, Y.-H., Collura, R., Ruvolo, M., Barrett, J.C., Woods, C.G., Walsh, C.A., Jurka, J.,

Larionov, V., 2004. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. PLoS biology. 2, E126.

Krasowski, M.D., Yasuda, K., Hagey, L.R., Schuetz, E.G., 2005. Evolutionary selection across the nuclear hormone receptor superfamily with a focus on the NR1I subfamily (vitamin D, pregnane X, and constitutive androstane receptors). Nucl. Recept. 3, 2.

Kratzer, J.T., Lanaspa, M.A., Murphy, M.N., Cicerchi, C., Graves, C.L., Tipton, P.A., Ortlund, E.A., Johnson, R.J., Gaucher, E.A., 2014. Evolutionary history and metabolic insights of ancient mammalian uricases. Proc. Natl. Acad. Sci. U.S.A. 111, 3763–3768.

LaMantia AS, Bhasin N, Rhodes K, Heemskerk J. 2000. Mesenchymal/epithelial induction mediates olfactory pathway formation. Neuron 28:411–425.

Laudet, V., 1997. Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. J. Mol. Endocrinol. 19(3), 207-226.

Lee, H.J., Chang, C., 1995. Identification of human TR2 orphan receptor response element in the transcriptional initiation site of the simian virus 40 major late promoter. The Journal of biological chemistry. 270, 5434–5440.

Lee, Y.-F., Lee, H.-J., Chang, C., 2002. Recent advances in the TR2 and TR4 orphan receptors of the nuclear receptor superfamily. J. Steroid Biochem. 81, 291–308.

Lee, Y.F., Young, W.J., Burbach, J.P., Chang, C., 1998. Negative feedback control of the retinoid-retinoic acid/retinoid X receptor pathway by the human TR4 orphan receptor, a member of the steroid receptor superfamily. The J. Biol. Chem. 273,

13437–13443.

Leigh, S.R., 2004. Brain growth, life history, and cognition in primate and human evolution. American journal of primatology. 62, 139–164.

Le-Niculescu, H., Patel, S.D., Bhat, M., Kuczenski, R., Faraone, S.V., Tsuang, M.T., McMahon, F.J., Schork, N.J., Nurnberger, J.I., Niculescu, A.B., 2009. Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. Am. J. Med. Genet. B. 150B, 155–181.

Lent, R., Azevedo, F.A.C., Andrade-Moraes, C.H., Pinto, A.V.O., 2012. How many neurons do you have? Some dogmas of quantitative neuroscience under revision. The European journal of neuroscience. 35, 1–9.

Lepre, C.J., Roche, H., Kent, D.V., Harmand, S., Quinn, R.L., Brugal, J.-P., Texier, P.-J., Lenoble, A., Feibel, C.S., 2011. An earlier origin for the Acheulian. Nature. 477, 82–85.

Lewitus, E., Kelava, I., Huttner, W.B., 2013. Conical expansion of the outer subventricular zone and the role of neocortical folding in evolution and development. Frontiers in human neuroscience. 7, 424.

Lewitus, E., Kelava, I., Kalinka, A.T., Tomancak, P., Huttner, W.B., 2014. An adaptive threshold in mammalian neocortical evolution. PLoS biology. 12, e1002000.

Li, W.H., Wu, C.I., Luo, C.C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Molecular biology and evolution. 2, 150–174

Lockwood, C.A., Kimbel, W.H., Lynch, J.M., 2004. Morphometrics and hominoid

phylogeny: Support for a chimpanzee-human clade and differentiation among great ape subspecies. Proceedings of the National Academy of Sciences of the United States of America. 101, 4356–4360.

Logue, M.W., Baldwin, C., Guffanti, G., Melista, E., Wolf, E.J., Reardon, A.F., Uddin, M., Wildman, D., Galea, S., Koenen, K.C., Miller, M.W., 2012. A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. Mol. Psychiatr. 18, 937–942.

Lonard, D.M., O'Malley, B.W., 2007. Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. Mol. Cell. 27(5). 691-700.

Lonard DM, O'Malley BW. 2009. Emerging roles of the ubiquitin proteasome system in nuclear hormone receptor signaling. Prog Mol Biol Transl Sci 87:117-35.

Lupien, S.J., McEwen, B.S., 1997. The acute effects of corticosteroids on cognition: integration of animal and human model studies. Brain Res. Brain Res. Rev. 24, 1–27.

Maddison, W., Knowles, L., 2006. Inferring Phylogeny Despite Incomplete Lineage Sorting. Syst. Biol. 55, 21–30.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Lu, F., Marchler, G.H., Mullokandov, M., Omelchenko, M.V., Robertson, C.L., Song, J.S., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Zheng, C., Bryant, S.H., 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, D225–9.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends in genetics : TIG. 24, 133–141.

Mardis, E.R., 2011. A decade's perspective on DNA sequencing technology. Nature. 470, 198–203.

Markham, A.C., Santymire, R.M., Lonsdorf, E.V., Heintz, M.R., Lipende, I., Murray, C.M., 2014. Rank effects on social stress in lactating chimpanzees. Anim Behav 87, 195–202.

Matthaei, J.H., Nirenberg, M.W., 1961. Characteristics and stabilization of DNAase-sensitive protein synthesis in E. coli extracts. Proc. Natl. Acad. Sci. U.S.A. 47, 1580–1588.

Maynard T, Gopalakrishna D, Meechan D, Paronett E, Newbern J, LaMantia, A. 2013. 22q11 Gene dosage establishes an adaptive range for sonic hedgehog and retinoic acid signaling during early development *Hum Mol Genet* 22(2):300-312.

Merlino, A.A., Welsh, T.N., Tan, H., Yi, L.J., Cannon, V., Mercer, B.M., Mesiano, S., 2007. Nuclear Progesterone Receptors in the Human Pregnancy Myometrium: Evidence that Parturition Involves Functional Progesterone Withdrawal Mediated by Increased Expression of Progesterone Receptor-A. J. Clin. Endocrinol. Metab. 92, 1927–1933.

Mesiano, S., 2004. Myometrial progesterone responsiveness and the control of human parturition. J. Soc. Gynecol. Invest. 25(1), 5-13.

Michalczyk, K., Ziman, M. 2005. *Nestin* structure and predicted function in cellular cytoskeletal organisation. Histol Histopathol 20 (2): 665–71.

Milam, A.H., Rose, L., Cideciyan, A.V., Barakat, M.R., Tang, W.-X., Gupta, N., Aleman,

T.S., Wright, A.F., Stone, E.M., Sheffield, V.C., Jacobson, S.G., 2002. The nuclear receptor NR2E3 plays a role in human retinal photoreceptor differentiation and degeneration. Proc. Natl. Acad. Sci. U.S.A. 99, 473–478.

Mingrone, J.R., Susko, E., Bielawski, J.P., 2015. SBA for Detecting Selection Pressure at Amino Acid Site. Mol. Biol. Evol. in prep.

Miyata, T., Miyazawa, S., Yasunaga, T., 1979. Two types of amino acid substitutions in protein evolution. Journal of molecular evolution. 12, 219–236.

Molenda-Figueira, H.A., Murphy, S.D., Shea, K.L., Siegal, N.K., Zhao, Y., Chadwick, J.G., Denner, L.A., Tetel, M.J., 2008. Steroid receptor coactivator-1 from brain physically interacts differentially with steroid receptor subtypes. Endocrinology. 149, 5272–5279.

Montgomery, S.H., Capellini, I., Barton, R.A., Mundy, N.I., 2010. Reconstructing the ups and downs of primate brain evolution: implications for adaptive hypotheses and Homo floresiensis. BMC Biology. 8(9):1-19.

Montgomery, S.H., Capellini, I., Venditti, C., Barton, R.A., Mundy, N.I., 2011. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. Molecular biology and evolution. 28, 625–638.

Moss, M.L., Young, R.W., 1960. A functional approach to craniology. American journal of physical anthropology. 18, 281–292.

Muller, M.N., Wrangham, R.W., 2004. Dominance, cortisol and stress in wild chimpanzees (*Pan troglodytes schweinfurthii*). Behav. Ecol. Sociobiol. 55: 332-340.

Muse, S.V. and Gaut, B.S., 1994. A likelihood approach for comparing synonymous and

nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol. 11(5):715-24.

Neale, B.M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J.B., Vasquez, A.A., Asherson, P., Chen, W., Banaschewski, T., Buitelaar, J., Ebstein, R., Gill, M., Miranda, A., Oades, R.D., Roeyers, H., Rothenberger, A., Sergeant, J., Steinhausen, H.C., Sonuga-Barke, E., Mulas, F., Taylor, E., Laird, N., Lange, C., Daly, M., Faraone, S.V., 2008. Genome-wide association scan of attention deficit hyperactivity disorder. Am. J. Med. Genet. B. Neuropsychiatr. Genet. 147B(8):1337–1344.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, J Sninsky J, Adams MD, Cargill M. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biology 3: doi: 10.1371/journal.pbio.0030170.

Nirenberg, M., 1965. Protein synthesis and the RNA code. Harvey lectures. 59, 155–185.

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., O'Neal, C., 1965. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proceedings of the National Academy of Sciences of the United States of America. 53, 1161–1168.

Nirenberg, M.W., Matthaei, J.H., Jones, O.W., Martin, R.G., Barondes, S.H., 1963. Approximation of genetic code via cell-free protein synthesis directed by template RNA. Fed. Proc. 22, 55–61

Niwa, H., Miyazaki, J., Smith, A.G., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nature genetics. 24,

372–376.

Norrgard, M.A., Ivarsson, Y., Tars, K., Mannervik, B., 2006. Alternative mutations of a positively selected residue elicit gain or loss of functionalities in enzyme evolution. Proc. Natl. Acad. Sci. U.S.A. 103, 4876–4881.

Nuclear Receptors Nomenclature Committee, 1999. A unified nomenclature system for the nuclear receptor superfamily. Cell. 97, 161–163.

Nunn, C.L., 1999. The evolution of exaggerated sexual swellings in primates and the graded-signal hypothesis. Anim. Behav. 58(2), 229-246

Oh, S.-Y., Kim, C.J., Park, I., Romero, R., Sohn, Y.-K., Moon, K.C., Yoon, B.H., 2005. Progesterone receptor isoform (A/B) ratio of human fetal membranes increases during term parturition. Am. J. Obstet. Gynecol. 193, 1156–1160.

Oitzl, M.S., Champagne, D.L., van der Veen, R., de Kloet, E.R., 2010. Brain development under stress: hypotheses of glucocorticoid actions revisited. Neurosci Biobehav Rev 34, 853–866.

Opazo, J.C., Wildman, D.E., Prychitko, T., Johnson, R.M., Goodman, M., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Molecular phylogenetics and evolution. 40, 274–280.

Ortlund, E.A., Bridgham, J.T., Redinbo, M.R., Thornton, J.W., 2007. Crystal structure of an ancient protein: evolution by conformational epistasis. Science (New York, N.Y.). 317, 1544–1548.

Owen, G.I., Zelent, A., 2000. Origins and evolutionary diversification of the nuclear receptor superfamily. Cell. Mol. Life Sci. 57(5):809-27.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol.

Evol. 5, 568–583.

Park, S.W., Hu, X., Gupta, P., Lin, Y.-P., Ha, S.G., Wei, L.-N., 2007. SUMOylation of Tr2 orphan receptor involves Pml and fine-tunes *Oct4* expression in stem cells. Nature structural & molecular biology. 14, 68–75.

Pauling, L., Itano, H.A., 1949. Sickle cell anemia, a molecular disease. Science 109, 443.

Pauling L., Zuckerkandl E., 1963. Chemical paleogenetics: Molecular 'restoration studies' of extinct forms of life. Acta Chem Scand 17:9–16.

Peabody, D.S., 1989. Translation initiation at non-AUG triplets in mammalian cells. The J. Biol. Chem.. 264, 5031–5035.

Perelman, P., Johnson, W.E., Roos, C., Seuánez, H.N., Horvath, J.E., Moreira, M.A.M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M.P.C., Silva, A., O'Brien, S.J., Pecon-Slattery, J., 2011. A molecular phylogeny of living primates. PLoS genetics. 7, e1001342.

Perutz, M. F. 1983. Species adaptation in a protein molecule. Mol. Biol. Evol. 1, 1-28.

Pikarsky, E., Sharir, H., Ben-Shushan, E., Bergman, Y., 1994. Retinoic acid represses Oct-3/4 gene expression through several retinoic acid-responsive elements located in the promoter-enhancer region. Molecular and cellular biology. 14, 1026–1038.

Pippal, J.B., Fuller, P.J., 2008. Structure–function relationships in the mineralocorticoid receptor. J. Mol. Endocrinol. 41(6), 405-413.

Preuss, T.M., Cáceres, M., Oldham, M.C., Geschwind, D.H., 2004. Human brain evolution: insights from microarrays. Nature reviews. Genetics. 5, 850–860.

Preuss, T.M., 2011. The human brain: rewired and running hot. Annals of the New York Academy of Sciences. 1225 Suppl 1, E182–91.

Pulvers, J.N., Bryk, J., Fish, J.L., Wilsch-Bräuninger, M., Arai, Y., Schreier, D., Naumann, R., Helppi, J., Habermann, B., Vogt, J., Nitsch, R., Tóth, A., Enard, W., Pääbo, S., Huttner, W.B., 2010. Mutations in mouse Aspm (abnormal spindle-like microcephaly associated) cause not only microcephaly but also major defects in the germline. Proceedings of the National Academy of Sciences of the United States of America. 107, 16595–16600.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., Maricic, T., Good, J.M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E.E., Stoneking, M., Richards, M., Talamo, S., Shunkov, M.V., Derevianko, A.P., Hublin, J.-J., Kelso, J., Slatkin, M., Pääbo, S., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature. 468, 1053–1060.

Reichert, H., 2009. Evolutionary conservation of mechanisms for neural regionalization, proliferation and interconnection in brain development. Biology letters. 5, 112–116.

Rilling, J.K., 2014. Comparative primate neuroimaging: insights into human brain evolution. Trends in cognitive sciences. 18, 46–55.

Robinson-Rechavi, M., Carpentier, A.-S., Duffraisse, M., Laudet, V., 2001. How many nuclear hormone receptors are there in the human genome? Trends Genet. 17, 554–556.

Rogatsky, I., Zarember, K.A., Yamamoto, K.R., 2001. Factor recruitment and TIF2/GRIP1 corepressor activity at a collagenase-3 response element that mediates regulation by phorbol esters and hormones. EMBO J. 20(21), 6071-6083.

Rooney, A.P., Zhang, J., 1999. Rapid evolution of a primate sperm protein: relaxation of

functional constraint or positive Darwinian selection? Mol. Biol. Evol. 16, 706–710.

Rosenberg, N.A., 2002. The Probability of Topological Concordance of Gene Trees and Species Trees. Theor. Popul. Biol. 61, 225–247.

Ross CF and Martin RD. 2006. The role of vision in the origin and evolution of primates. In: Preuss TM. Evolution of Nervous Systems Volume 4: The Evolution of Primate Nervous Systems. Oxford: Elsevier. 59-78.

Ruvolo, M., 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. Molecular biology and evolution. 14, 248–265.

Sabeti, P., 2011. Genome-sequencing anniversary. The landscape of human evolution. Science (New York, N.Y.). 331, 690–690.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. 419, 832–837.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., Lander, E.S., 2006. Positive natural selection in the human lineage. Science (New York, N.Y.). 312, 1614–1620.

Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. Science. 308, 648–652.

Sarachana, T., Xu, M., Wu, R.-C., Hu, V.W., 2011. Sex hormones in autism: androgens

and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. PloS One. 6, e17116.

Sartorius, C.A., Melville, M.Y., Hovland, A.R., 1994. A third transactivation function (AF3) of human progesterone receptors located in the unique N-terminal segment of the B-isoform. Mol. Endocrinol. 8(10), 1347-1360.

Sawyer, S.L., Wu, L.I., Emerman, M., Malik, H.S., 2005. Positive selection of primate TRIM5 identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. U.S.A. 102, 2832–2837.

Schneider, A., Souvorov, A., Sabath, N., Landan, G., Gonnet, G.H., Graur, D., 2009. Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. Genome Biol. Evol. 1, 114–118.

Semaw, S., Renne, P., Harris, J.W., Feibel, C.S., Bernor, R.L., Fesseha, N., Mowbray, K., 1997. 2.5-million-year-old stone tools from Gona, Ethiopia. Nature. 385, 333–336.

Sengupta, D., Bhargava, D.K., Dixit, A., Sahoo, B.S., Biswas, S., Biswas, G., Mishra, S.K., 2014. ERRβ signalling through FST and BCAS2 inhibits cellular proliferation in breast cancer cells. Br. J. Cancer 110, 2144–2158.

Sherwood, C.C., Duka, T., 2012. Now that we've got the map, where are we going? Moving from gene candidate lists to function in studies of brain evolution. Brain Behav. Evol. 80, 167–169.

Sherwood, C.C., Stimpson, C.D., Raghanti, M.A., Wildman, D.E., Uddin, M., Grossman, L.I., Goodman, M., Redmond, J.C., Bonar, C.J., Erwin, J.M., Hof, P.R., 2006. Evolution of increased glia-neuron ratios in the human frontal cortex. Proceedings of the National Academy of Sciences of the United States of America. 103, 13606–

13611.

Sherwood, C.C., Subiaul, F., Zawidzki, T.W., 2008. A natural history of the human mind: tracing evolutionary changes in brain and cognition. Journal of anatomy. 212, 426–454.

Shi, L., Li, M., Lin, Q., Qi, X., Su, B., 2013. Functional divergence of the brain-size regulating gene MCPH1 during primate evolution and the origin of humans. BMC biology. 11, 62.

Shyr, C.R., Kang, H.Y., Tsai, M.Y., Liu, N.C., Ku, P.Y., Huang, K.E., Chang, C., 2009. Roles of testicular orphan nuclear receptors 2 and 4 in early embryonic development and embryonic stem cells. Endocrinology. 150, 2454–2462.

Sibley CG, Ahlquist, JE. 1984. The phylogeny of the hominoid primates, as indicated by DNA–DNA hybridization. J Mol Evol 20, 2–15.

Somel, M., Liu, X., Khaitovich, P., 2013. Human brain evolution: transcripts, metabolites and their regulators. Nature reviews. Neuroscience. 14, 112–127.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312–1313.

Storey, J. D., 2002. A direct approach to false discovery rates. J. Roy. Sta.t Soc. B. 64(3), 479-498.

Surbeck, M., Deschner, T., Weltring, A., Hohmann, G., 2012. Social correlates of variation in urinary cortisol in wild male bonobos (Pan paniscus). Horm. Behav. 62(1), 27-35.

Takahata, N., 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics. 122, 957–966.

Tallec, L.P.-L., Kirsh, O., Lecomte, M.-C., Viengchareun, S., Zennaro, M.-C., Dejean, A., Lombès, M., 2003. Protein inhibitor of activated signal transducer and activator of transcription 1 interacts with the N-terminal domain of mineralocorticoid receptor and represses its transcriptional activity: implication of small ubiquitin-related modifier 1 modification. Mol. Endocrinol. 17, 2529–2542.

Tanaka, T., Nei, M., 1989. Positive darwinian selection observed at the variable-region genes of immunoglobulins. Mol. Biol. Evol. 6, 447–459.

Terracciano, A., Tanaka, T., Sutin, A.R., Sanna, S., Deiana, B., Lai, S., Uda, M., Schlessinger, D., Abecasis, G.R., Ferrucci, L., Costa, P.T., 2010. Genome-wide association scan of trait depression. Biol. Psychiat. 68, 811–817.

Tetel, M.J., Giangrande, P.H., Leonhardt, S.A., McDonnell, D.P., Edwards, D.P., 1999. Hormone-dependent interaction between the amino- and carboxyl-terminal domains of progesterone receptor in vitro and in vivo. Mol. Endocrinol. 13, 910–924.

Thornton, J.W., 2001. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc. Natl. Acad. Sci. U.S.A. 98, 5671–5676.

Thornton, J.W., 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. 5, 366–375.

Thornton, J.W., Need, E., Crews, D., 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science. 301, 1714–1717.

Tomasello, M., Hare, B., Agnetta, B., 1999. Chimpanzees, Pan troglodytes, follow gaze direction geometrically. Anim. Behav. 58, 769–777.

Tomasello, M., Hare, B., Lehmann, H., Call, J., 2007. Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J. Hum. Evol. 52, 314–320.

Tryon, C., Roach, N., Logan, M., 2008. The Middle Stone Age of the northern Kenyan Rift: age and context of new archaeological sites from the Kapedo Tuffs. J. Hum. Evol. 55, 652–664.

Tsaur, S.C., Ting, C.T., Wu, C.I., 2001. Sex in Drosophila mauritiana: a very high level of amino acid polymorphism in a male reproductive protein gene, Acp26Aa. Mol. Biol. Evol. 18, 22–26.

Tucker ES, Lehtinen MK, Maynard T, Zirlinger M, Dulac C, Rawson N, Pevny L, Lamantia AS. 2010. Proliferative and transcriptional identity of distinct classes of neural precursors in the mammalian olfactory epithelium. Development. 137(15):2471-81.

Ugalde, J.A., Chang, B.S.W., Matz, M.V., 2004. Evolution of coral pigments recreated. Science. 305, 1433–1433.

Vallender, E.J., 2011. Comparative genetic approaches to the evolution of human brain and behavior. American journal of human biology : the official journal of the Human Biology Council. 23, 53–64.

Varki, A., Altheide, T.K., 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res. 15, 1746–1758.

Wagner, G.P., Lynch, V.J., 2008. The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. 23, 377–385.

Wang, Y.Q., Su, B., 2004. Molecular evolution of microcephalin, a gene determining

human brain size. Human molecular genetics. 13, 1131–1137.

Ward, C.V., Kimbel, W.H., Johanson, D.C., 2011. Complete Fourth Metatarsal and Arches in the Foot of Australopithecus afarensis. Science. 331, 750–753.

Ward, L.D., Kellis, M., 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science (New York, N.Y.). 337, 1675–1678.

Wärnmark, A., Treuter, E., Wright, A.P.H., Gustafsson, J.-A., 2003. Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. Mol. Endocrinol. 17, 1901–1909.

Watson, J.D., Crick, F.H., 1953a. Genetical implications of the structure of deoxyribonucleic acid. Nature. 171, 964–967.

Watson, J.D., Crick, F.H., 1953b. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 171, 737–738.

Watson, J.D., Crick, F.H., 1953c. The structure of DNA. Cold Spring Harbor symposia on quantitative biology. 18, 123–131

Weigel, N.L., 1996. Steroid hormone receptors and their regulation by phosphorylation. Biochem. J. 319 (Pt 3), 657–667.

Weinberger, H., Moran, Y., Gordon, D., Turkov, M., Kahn, R., Gurevitz, M., 2010. Positions under Positive Selection--Key for Selectivity and Potency of Scorpion -Toxins. Mol Biol Evol. 27, 1025–1034.

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., Hollfelder, N., Potekhina, I.D., Schier, W., Thomas, M.G., Burger, J., 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans

during the last 5,000 y. Proceedings of the National Academy of Sciences of the United States of America. 111, 4832–4837.

Wildman, D.E., Uddin, M., Liu, G., Grossman, L.I., Goodman, M., 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. Proceedings of the National Academy of Sciences of the United States of America. 100, 7181–7188.

Williamson, S.H., Hubisz, M.J., Clark, A.G., Payseur, B.A., Bustamante, C.D., Nielsen, R., 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet. 3, e90.

Wilson, D.J., McVean, G., 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics. 172, 1411–1425.

Wong, W.S.W., Yang, Z., Goldman, N., Nielsen, R., 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics. 168, 1041–1051.

Wood, B., 1996. Hominid palaeobiology: have studies of comparative development come of age? American journal of physical anthropology. 99, 9–15.

Wood, B., 2010. Reconstructing human evolution: Achievements, challenges, and opportunities. Proc. Natl. Acad. Sci. U.S.A. 107, 8902–8909.

Wood, B., Harrison, T., 2012. The evolutionary context of the first hominins. Nature. 470, 347–352.

Wyckoff, G.J., Wang, W., Wu, C.I., 2000. Rapid evolution of male reproductive genes in the descent of man. Nature. 403(6767), 304-309.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum

likelihood. Computer applications in the biosciences : CABIOS. 13, 555–556.

Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15, 568–573.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution. 24, 1586–1591.

Yang, Z., Bielawski, J., 2000. Statistical methods for detecting molecular adaptation. Trends in ecology & evolution. 15, 496–503.

Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908–917.

Yang, Z., Swanson, W.J., 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. 19(1), 49-57.

Young, W.J., Lee, Y.F., Smith, S.M., Chang, C., 1998. A bidirectional regulation between the TR2/TR4 orphan receptors (TR2/TR4) and the ciliary neurotrophic factor (CNTF) signaling pathway. The Journal of biological chemistry. 273, 20877–20885.

Yu, S., Wong, Y.C., Wang, X.H., Ling, M.T., Ng, C.F., Chen, S., Chan, F.L., 2008. Orphan nuclear receptor estrogen-related receptor-beta suppresses in vitro and in vivo growth of prostate cancer cells via p21(WAF1/CIP1) induction and as a potential therapeutic target in prostate cancer. Oncogene. 27, 3313–3328.

Zhang, J., 2003. Evolution of the human ASPM gene, a major determinant of brain size. Genetics. 165, 2063–2070.

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood

method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472–2479.

Zhang, J., Rosenberg, H.F., 2002. Diversifying selection of the tumor-growth promoter angiogenin in primate evolution. Mol. Biol. Evol. 19, 438–445.

Zhang, Z., Burch, P.E., Cooney, A.J., Lanz, R.B., Pereira, F.A., Wu, J., Gibbs, R.A., Weinstock, G., Wheeler, D.A., 2004. Genomic analysis of the nuclear receptor family: new insights into structure, regulation, and evolution from the rat genome. Genome Res. 14, 580–590.

Zilles, K., Palomero-Gallagher, N., Amunts, K., 2013. Development of cortical folding during evolution and ontogeny. Trends in neurosciences. 36, 275–284.