

# Changes in Gene Expression Associated with Reproductive Maturation in Wild Female Baboons

Courtney C. Babbitt<sup>1,2,\*</sup>, Jenny Tung<sup>3</sup>, Gregory A. Wray<sup>1,2,3</sup>, and Susan C. Alberts<sup>1,2,3,4</sup>

<sup>1</sup>Department of Biology, Duke University

<sup>2</sup>Institute for Genome Sciences & Policy, Duke University

<sup>3</sup>Department of Evolutionary Anthropology, Duke University

<sup>4</sup>Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya

\*Corresponding author: E-mail: courtney.babbitt@duke.edu.

**Accepted:** 6 December 2011

## Abstract

Changes in gene expression during development play an important role in shaping morphological and behavioral differences, including between humans and nonhuman primates. Although many of the most striking developmental changes occur during early development, reproductive maturation represents another critical window in primate life history. However, this process is difficult to study at the molecular level in natural primate populations. Here, we took advantage of ovarian samples made available through an unusual episode of human–wildlife conflict to identify genes that are important in this process. Specifically, we used RNA sequencing (RNA-Seq) to compare genome-wide gene expression patterns in the ovarian tissue of juvenile and adult female baboons from Amboseli National Park, Kenya. We combined this information with prior evidence of selection occurring on two primate lineages (human and chimpanzee). We found that in cases in which genes were both differentially expressed over the course of ovarian maturation and also linked to lineage-specific selection this selective signature was much more likely to occur in regulatory regions than in coding regions. These results suggest that adaptive change in the development of the primate ovary may be largely driven at the mechanistic level by selection on gene regulation, potentially in relationship to the physiology or timing of female reproductive maturation.

**Key words:** RNA-Seq, gene expression, wild primate population.

## Transcriptome Analysis and Life History in a Wild Primate Population

Nonhuman primates are valuable sources of insight into human evolution. Until recently, however, such insight was limited by the dearth of genetic resources for most primate species. In addition, studies of primates in their natural habitats, while rich in behavioral and ecological detail, have rarely included extensive genetic or genomic components. This situation is changing now that genomic resources are increasingly available, and gene regulatory studies of captive primates have set the stage (reviewed in Tung et al. 2010). However, we still know relatively little about variation in gene expression in wild primates.

Collecting functional genomic data on such systems could provide important context for the evolution of gene regulation in humans. Specifically, studying changes in gene expression during maturational milestones in nonhuman primates

may provide insight into the loci that contributed to shifts in developmental timing and physiology during human evolution (Uddin et al. 2008; Somel et al. 2009; Gunz et al. 2010). Some examples of these shifts include relatively late menarche in human hunter-gatherers compared with nonhuman primates (reviewed in Blurton Jones et al. 1999); a skeletal growth spurt that accompanies female maturation in humans that appears to be absent in nonhuman primates (Bogin and Smith 1996); and short interbirth intervals in humans relative to body size (reviewed in Mace 2000). Circumstantial evidence suggests a role for gene regulation in these changes. Indeed, sequence-based analyses have revealed that the regulatory regions of many development-related genes have undergone positive selection within primates (Haygood et al. 2010) and that rapidly evolving regulatory regions near duplicated genes in humans are enriched for genes related to pregnancy and reproduction (Kostka et al. 2010).

© The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.

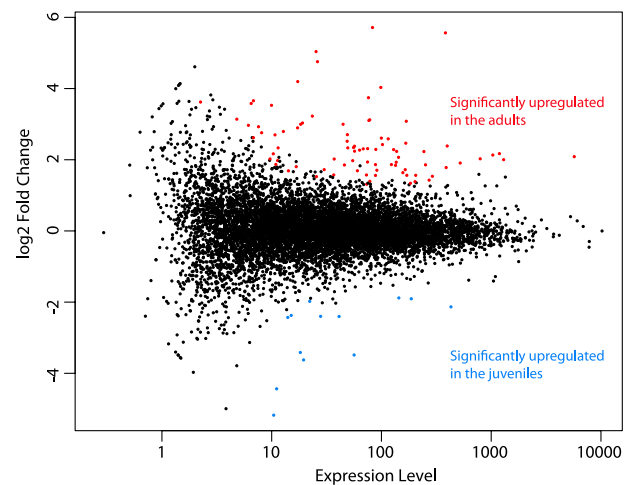
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Yellow baboons (*Papio cynocephalus*) are close human relatives (~94% sequence similarity; see Silva and Kondrashov 2002) that, like humans, are large-bodied terrestrial savanna primates with long life histories and nonseasonal reproduction. They also inhabit African savanna environments similar to those relevant for early humans (Potts 1998; Behrensmeyer 2006). Yellow baboons have been the subjects of extensive study in the wild (Altmann SA and Altmann J 1970; Jolly 1993; Rhine et al. 2000; Buchan et al. 2003; Wasser et al. 2004; Alberts et al. 2006), including in the Amboseli basin of Kenya where individually recognized baboons have been monitored since 1971 (Altmann SA and Altmann J 1970; Buchan et al. 2003; Alberts et al. 2006). This system therefore presents an exceptional opportunity to integrate functional genomic data sets with detailed life history information about the same animals.

Here, we take advantage of life history and behavioral data from the Amboseli baboon population, combined with an unusual circumstance in which we were able to collect fresh tissue from seven known females (four premenarcheal juveniles and three multiparous adults). Six of these seven females died in an episode of conflict with the local human population (the Maasai community in Amboseli) that perceived the baboons as a threat to their livestock; the seventh died of natural causes a few days later. The bodies of all seven females were collected within a few hours of their death, with the help of the Maasai community. We used these data and samples to investigate gene expression changes related to the onset of sexual maturity in females and to examine differential expression in maturity-related genes among genes inferred to have evolved under lineage-specific selection in primates. We focused specifically on expression differences in the ovary, an organ that plays a central role in reproductive maturation. We present a genome-wide analysis of ovarian gene expression changes in these seven female baboons from this natural population using RNA-Seq.

### Expression Differences by Life History Stage

RNA-Seq libraries were made using ovarian RNA from three adult and four juvenile females (supplementary fig. S1 and table S1, Supplementary Material online). We obtained ~15 million reads per individual (supplementary table S1, Supplementary Material online), and we measured the expression of a total of 9,770 genes in the baboon ovary. Ninety-seven genes (~1% of genes in the data set) were differentially expressed between the juveniles and the adults (False discovery rate [FDR]–adjusted  $P$  value < 0.05) (fig. 1). This result is consistent with the expectation that intraspecific differential expression, particularly within a population and within sex, is likely to be less common than interspecific differential expression between different primate species (Babbitt et al.



**FIG. 1.**—MA plot of the normalized data. Each dot represents a single gene, and significantly differentially expressed genes are colored by higher expression levels in adults (red) or juveniles (blue).

2010; Blehman et al. 2010; Xu et al. 2010). Of the differentially expressed genes, 79 were upregulated in the adults and 18 were upregulated in the juveniles. This imbalance in upregulated expression toward the adult females was expected, as the adult ovary is much more metabolically active than the premenarcheal ovary (reviewed in McGee and Hsueh 2000).

To evaluate the global effect of maturation stage on gene expression variation, we performed a principal components (PCs) analysis. The first three PCs in this analysis explained ~67% of variation in the gene expression data (supplementary fig. S3, Supplementary Material online). None of these PCs clearly differentiated adult and juvenile tissues, although PC2 exhibited the strongest (albeit nonsignificant) relationship with life history stage (Mann–Whitney test,  $W = 11$ ,  $P$  value = 0.1143). In contrast, when we examined only those genes that were significantly differentially expressed ( $n = 97$ ), PC1 alone explained 70% of the variation in the gene expression data. PC1 also exhibited a trend toward higher values for juveniles than for adults (Mann–Whitney test,  $W = 0$ ,  $P$  value = 0.05714 and supplementary fig. S3, Supplementary Material online).

Little is known about ovarian gene expression in human or mouse models during either the premenarcheal stage or in noncycling adult tissue, as most studies concerning ovarian gene expression have focused on embryonic sex specification (Nef et al. 2005; Small et al. 2005), fertility disorders (reviewed in Matzuk and Lamb 2008), or cancer states (e.g., Wang et al. 1999; Welsh et al. 2001; Haviv and Campbell 2002; Adib et al. 2004). To explore patterns in expression differences between these life history stages, we performed categorical enrichment analyses using the GO (Gene Ontology Consortium 2000) and PANTHER (Mi et al. 2005) ontology databases. The enrichments were performed in

**Table 1**

PANTHER Biological Process Categorical Enrichments

PANTHER Biological Process Category	P Value	Total Occurrence
Signal transduction	$1.58 \times 10^{-09}$	1,359
Cell-surface receptor-mediated signal transduction	$2.95 \times 10^{-08}$	558
Cell communication	$1.73 \times 10^{-07}$	435
Immunity and defense	$9.16 \times 10^{-07}$	497
Ligand-mediated signaling	$7.77 \times 10^{-06}$	111
Neuronal activities	$4.08 \times 10^{-05}$	201
Cell motility	0	151
G-protein-mediated signaling	0	228
Other neuronal activity	0	64
Cytokine- and chemokine-mediated signaling pathway	0	74
Developmental processes	0	903
B-cell-mediated and antibody-mediated immunity	0	28
Skeletal development	0	59
Interferon-mediated immunity	0	20
Homeostasis	0.01	89
Macrophage-mediated immunity	0.01	34
Cell adhesion	0.01	252
Extracellular matrix protein-mediated signaling	0.01	39
Ectoderm development	0.01	272
Blood circulation and gas exchange	0.01	21
Neurogenesis	0.01	250
Mesoderm development	0.01	251
Cell adhesion-mediated signaling	0.01	139
Cytokine/chemokine-mediated immunity	0.01	27
Angiogenesis	0.02	38
Detoxification	0.02	38
Fatty acid metabolism	0.02	85
Anion transport	0.02	25
MHCII-mediated immunity	0.02	16
Other receptor-mediated signaling pathway	0.02	81
Extracellular transport and import	0.03	32
Sensory perception	0.03	96
JAK-STAT cascade	0.03	30
Natural killer cell mediated immunity	0.03	11

NOTE.—Categorical enrichments were performed using a Wilcoxon rank test. The right-hand column shows the total number of genes evaluated. Categories that evaluated fewer than ten genes are not shown. Categories in white have a B-H corrected *P* value < 0.05 (Benjamini and Hochberg 1995), and categories in gray have a nominal *P* value < 0.05.

two ways: first, using the absolute rankings of gene expression differences between adults and juveniles, regardless of the direction of the difference; and second, using only genes that were more highly expressed in the more metabolically active adult tissue (table 1, supplementary tables S2 and S3, Supplementary Material online).

Several patterns emerged from these analyses. First, we identified a number of ontology categories generally associated with blood, including “immunity and defense” and “angiogenesis” (table 1). The cortex of the ovary becomes highly vascularized after the onset of maturity (Redmer and Reynolds 1996; Abulafia and Sherer 2000), a maturational process that could account for some of the observed enrichments. In addition, follicular development in the mature ovary is correlated with increased inflammation (reviewed in Bukovsky and Caudle 2008). In keeping with this change, we identified cytokine, chemokine, and macrophage-related immune activities among the significant categories of genes

that show increased expression in the adults (supplementary table S3, Supplementary Material online). Second, and perhaps unsurprisingly, genes involved in developmental processes (i.e., developmental processes and mesoderm development) tended to be enriched for differential expression (table 1 and supplementary table S2, Supplementary Material online). These enrichments emphasize that the physiological distinctions between the mainly quiescent juvenile ovary and the mature ovary are likely related, at least in part, to differences in gene regulation.

At the level of individual genes, we found a significant upregulation in the adult ovary of genes essential for ovarian function and folliculogenesis (table 2), including genes such as *VGF* (VGF nerve growth factor inducible), *MMP19* (matrix metalloproteinase-19), and *ADAMTS1* (a disintegrin and metalloproteinase motif 1) (fig. 2). *MMP19* and *ADAMTS1* function to remodel the extracellular matrix as follicles develop (Jo and Curry 2004; Brown et al. 2010). The role of

**Table 2**

Differentially Expressed Genes (FDR-adjusted *P* value < 0.05) in the Adult and Juvenile Baboon Ovary

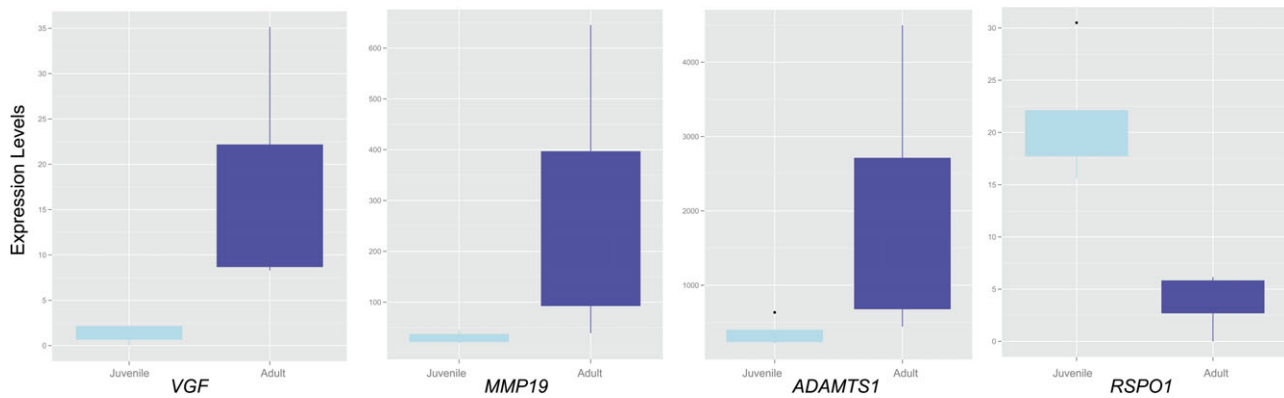
Gene ID	Log FC	<i>P</i> Value	<i>P</i> Value FDR	Gene ID	Log FC	<i>P</i> Value	<i>P</i> Value FDR
<i>serpina3</i>	5.56053945	2.44 × 10 <sup>-16</sup>	2.39 × 10 <sup>-12</sup>	<i>EPO</i>	-2.3597099	6.37 × 10 <sup>-05</sup>	0.01219865
<i>ADAMTS4</i>	5.09716829	5.79 × 10 <sup>-15</sup>	2.83 × 10 <sup>-11</sup>	<i>SLC7A8</i>	2.85383446	6.49 × 10 <sup>-05</sup>	0.01219865
<i>REN</i>	3.98944937	5.39 × 10 <sup>-12</sup>	1.76 × 10 <sup>-08</sup>	<i>ERRFI1</i>	2.13531073	7.99 × 10 <sup>-05</sup>	0.01472884
<i>TFPI2</i>	5.09331391	1.82 × 10 <sup>-11</sup>	3.58 × 10 <sup>-08</sup>	<i>MYOC</i>	-4.5033477	8.42 × 10 <sup>-05</sup>	0.0152368
<i>ADAMTS6</i>	4.69173617	1.83 × 10 <sup>-11</sup>	3.58 × 10 <sup>-08</sup>	<i>Mmp1</i>	8.22694892	8.90 × 10 <sup>-05</sup>	0.01581064
<i>Melk</i>	3.76852295	2.12 × 10 <sup>-10</sup>	3.44 × 10 <sup>-07</sup>	<i>rspo1</i>	-2.7082948	0.00010143	0.01769646
<i>LRG1</i>	4.68207316	4.18 × 10 <sup>-10</sup>	5.83 × 10 <sup>-07</sup>	<i>HIF1A</i>	2.20582485	0.00010339	0.01772116
<i>FABP4</i>	9.28927774	3.47 × 10 <sup>-09</sup>	4.24 × 10 <sup>-06</sup>	<i>rpl21</i>	3.86520992	0.0001073	0.01807402
<i>CH25H</i>	3.20323737	1.81 × 10 <sup>-07</sup>	0.00018065	<i>OSMR</i>	2.19076016	0.00011098	0.01837741
<i>SOST</i>	-4.8295275	1.85 × 10 <sup>-07</sup>	0.00018065	<i>TIMP1</i>	2.00766916	0.00011652	0.01897307
<i>IL1RL1</i>	9.32559106	3.14 × 10 <sup>-07</sup>	0.0002791	<i>CYP21A2</i>	2.59714864	0.00012208	0.0193343
<i>F3</i>	2.72902443	4.14 × 10 <sup>-07</sup>	0.00033246	<i>PAPPA</i>	3.66989958	0.00012459	0.0193343
<i>RGS2</i>	2.94854719	4.42 × 10 <sup>-07</sup>	0.00033246	<i>Adamts1</i>	-3.7455767	0.0001265	0.0193343
<i>GALNT9</i>	-3.4824072	6.16 × 10 <sup>-07</sup>	0.00042976	<i>Trem1</i>	1.97618174	0.0001301	0.0193343
<i>TNFAIP6</i>	9.06613349	6.82 × 10 <sup>-07</sup>	0.00044438	<i>RAB38</i>	-2.3456431	0.00013012	0.0193343
<i>CRTAC1</i>	-3.1329928	7.44 × 10 <sup>-07</sup>	0.00045431	<i>Sgip1</i>	2.17468855	0.00013061	0.0193343
<i>C19orf26</i>	2.79022659	1.08 × 10 <sup>-06</sup>	0.00062056	<i>DDX21</i>	2.13304855	0.00013343	0.01945687
<i>LdhA</i>	2.80233251	1.91 × 10 <sup>-06</sup>	0.00102984	<i>f2rl1</i>	3.0581654	0.00014459	0.0207748
<i>stc1</i>	2.95769089	2.00 × 10 <sup>-06</sup>	0.00102984	<i>SBNO2</i>	2.8690604	0.00015474	0.02160406
<i>Gdf15</i>	2.63705264	2.61 × 10 <sup>-06</sup>	0.00127693	<i>S100A8</i>	3.74690504	0.00015479	0.02160406
<i>FCER1G</i>	3.18705038	2.77 × 10 <sup>-06</sup>	0.00129071	<i>DST</i>	3.67689535	0.00016231	0.02233414
<i>VGF</i>	3.53951033	3.72 × 10 <sup>-06</sup>	0.00165083	<i>AADAC</i>	7.97969292	0.00016903	0.02293631
<i>MMP19</i>	2.49571312	5.03 × 10 <sup>-06</sup>	0.00213603	<i>CHI3L1</i>	2.26915607	0.00017834	0.0238048
<i>FosI2</i>	2.56131942	5.46 × 10 <sup>-06</sup>	0.0022307	<i>H6pd</i>	2.1574967	0.0001803	0.0238048
<i>SERPINE1</i>	3.00295543	6.06 × 10 <sup>-06</sup>	0.00226587	<i>ADAMTS16</i>	-2.5741504	0.00018334	0.02388255
<i>S100A9</i>	2.80444185	6.24 × 10 <sup>-06</sup>	0.00226587	<i>HLA-DQB1</i>	3.33877719	0.00019359	0.02488701
<i>ADPRHL1</i>	3.46982396	6.26 × 10 <sup>-06</sup>	0.00226587	<i>CTSG</i>	2.8539012	0.00021891	0.02777611
<i>Cd163l1</i>	2.93708735	7.60 × 10 <sup>-06</sup>	0.00265144	<i>RPF2</i>	2.32131206	0.00022495	0.02817639
<i>socs3</i>	2.49813522	9.29 × 10 <sup>-06</sup>	0.00306778	<i>Cd48</i>	7.59418959	0.00023965	0.02877211
<i>ifi30</i>	2.43485712	9.42 × 10 <sup>-06</sup>	0.00306778	<i>tnfrsf11b</i>	2.68764743	0.0002414	0.02877211
<i>CHGB</i>	2.68694397	1.04 × 10 <sup>-05</sup>	0.00327812	<i>C10orf10</i>	1.97584087	0.00024149	0.02877211
<i>Cntn4</i>	6.69229158	1.41 × 10 <sup>-05</sup>	0.00430415	<i>KCNN4</i>	-8.0092773	0.00024933	0.02934901
<i>Il1r1</i>	2.38670209	1.64 × 10 <sup>-05</sup>	0.00484285	<i>IL8</i>	4.44362712	0.00025339	0.02947116
<i>AG2</i>	7.88467942	1.87 × 10 <sup>-05</sup>	0.00507962	<i>ZFP36</i>	1.96860745	0.00026324	0.02998838
<i>GADD45A</i>	2.30721084	1.89 × 10 <sup>-05</sup>	0.00507962	<i>DLK1</i>	1.87823533	0.00026397	0.02998838
<i>LMO1</i>	2.31100372	1.97 × 10 <sup>-05</sup>	0.00507962	<i>GALR3</i>	-3.0699664	0.00027127	0.0304633
<i>TNFAIP3</i>	-3.1164769	2.00 × 10 <sup>-05</sup>	0.00507962	<i>ANKRD31</i>	7.59845327	0.00027525	0.03055914
<i>ANKRD1</i>	7.90497765	2.01 × 10 <sup>-05</sup>	0.00507962	<i>TRIB1</i>	2.3659021	0.00029898	0.03274912
<i>gpr84</i>	2.20041148	2.08 × 10 <sup>-05</sup>	0.00507962	<i>apol3</i>	2.42154371	0.00030168	0.03274912
<i>NUP35</i>	2.25111635	2.11 × 10 <sup>-05</sup>	0.00507962	<i>PPARGC1A</i>	-7.848901	0.00031593	0.03366667
<i>LCNL1</i>	-8.1713233	2.13 × 10 <sup>-05</sup>	0.00507962	<i>PTCHD1</i>	2.54755259	0.00031702	0.03366667
<i>cebpd</i>	2.21602916	2.28 × 10 <sup>-05</sup>	0.00530536	<i>EFNA5</i>	-2.683381	0.00037376	0.03926515
<i>NR5A2</i>	3.01054291	2.37 × 10 <sup>-05</sup>	0.00537637	<i>LGALS3</i>	1.95933307	0.00038655	0.04017603
<i>TMEM49</i>	2.23411164	3.46 × 10 <sup>-05</sup>	0.00767944	<i>c2cd4c</i>	1.86851938	0.00039556	0.04042629
<i>GZMB</i>	4.67164885	3.73 × 10 <sup>-05</sup>	0.00810511	<i>NFIL3</i>	-3.0702127	0.00039723	0.04042629
<i>SLC16A10</i>	6.39312229	4.21 × 10 <sup>-05</sup>	0.00894971	<i>WNT6</i>	-1.9481788	0.0004573	0.04606028
<i>ptgds</i>	-2.2794722	4.40 × 10 <sup>-05</sup>	0.00913917	<i>CAMP</i>	7.71716682	0.00047286	0.04714096
<i>HPGDS</i>	8.25515497	5.22 × 10 <sup>-05</sup>	0.01062847	<i>ism1</i>	1.96334052	0.00050099	0.04944076
<i>CD163</i>	2.06718683	5.90 × 10 <sup>-05</sup>	0.01153785				

NOTE.—FC means fold change.

*VGF* is less clear, but its essential role has been demonstrated in *VGF*<sup>-/-</sup> mice, which produce many primary follicles but few mature follicles (Hahm et al. 1999; Jethwa and Ebling 2008). Fewer genes are upregulated in the juveniles; however, one intriguing example is *RSPO1* (R-spondin1), which is known to be critical for early human ovary development and

specification (Tomaselli et al. 2011). Our data indicate that it continues to be expressed until the stages right before puberty (fig. 2).

Changes in gene regulation could reflect differences in alternative splicing and exon usage between juveniles and adults in addition to changes in transcript abundance



**Fig. 2.**—Boxplot diagrams of four representative differentially expressed genes involved in ovarian function and folliculogenesis. Juvenile expression data are in light blue, and adult expression data are in dark blue.

(e.g., Barberan-Soler and Zahler 2008; Revil et al. 2010). To investigate this possibility, we looked for differential exon expression (FDR-adjusted  $P$  value  $< 0.05$ )—a proxy for alternative splicing in a transcriptome without alternative splicing gene models—in genes with more than one exon. Specifically, we identified cases in which at least one exon, but not all exons, were differentially expressed (supplementary table S4, Supplementary Material online). To avoid false positives due to limited power, if one exon was differentially expressed, we relaxed the FDR-adjusted  $P$  value for differential expression to 0.15. Thus, evidence for exon-specific differential expression required relatively strong evidence for differential expression in at least one exon and a relative absence of evidence for differential expression in at least one other exon. Twenty-four genes exhibited this pattern, including *STC* (stanniocalcin) and *GCLC* (gamma-glutamylcysteine synthetase, catalytic subunit), both of which are thought to be important in ovarian development and function (Paciga et al. 2002; Luderer et al. 2003; Luo et al. 2004; Hoang et al. 2009).

### Differential Expression in the Ovary and Lineage-Specific Selection in Primate Noncoding Regions

Many of the genes expressed in juvenile and adult baboon ovaries are also likely to be expressed in juvenile and adult ovaries of other primates, including humans. Thus, genes that we identified as differentially expressed across life history stages in baboons might be informative for identifying genes important in female life history evolution in humans or in primates more generally. To gain insight into the patterns of natural selection that may have acted on such genes, we therefore integrated the novel functional data from this study with evidence for selection in primates from previous studies.

We obtained estimates of positive selection on the lineage leading to humans for protein-coding regions from

Nielsen et al. (2005) and for putative regulatory regions 5 kb upstream of genes from Haygood et al. (2007). Both studies took a similar approach to identify selective targets: Specifically, they compared the rate of nucleotide evolution in the focal region (protein-coding regions in Nielsen et al. 2005 and upstream regulatory regions in Haygood et al. 2007) with the rate of nucleotide evolution in a nearby region thought to be evolving neutrally (the general approach is reviewed in Yang and Bielawski 2000). An elevated rate of nucleotide evolution in the focal region relative to the nearby neutral region was interpreted as a signature of adaptive change. Likelihood ratio tests were then used to identify cases in which these rates differed across different branches of a species tree; we identified possible targets of lineage-specific selection by locating elevated rates of evolution in protein-coding or regulatory regions that occurred only on specific branches of the tree.

Combining our data with results from these studies (Nielsen et al. 2005; Haygood et al. 2007), we identified 225 genes that were both included in Haygood et al. (2007) and were differentially expressed in this study ( $P < 0.05$  for differential expression; we relaxed this threshold to increase the sensitivity of this analysis). Of these 225 genes, we found 19 differentially expressed genes that were associated with signatures of selection in noncoding regions on the human lineage ( $P < 0.05$  for the test for selection). In contrast, we found that none of our differentially expressed genes overlapped with signatures of positive selection in coding regions (of a total of 35 genes that were differentially expressed in this study and were included in Nielsen et al. (2005)). We did not observe a significant enrichment of ovarian differentially expressed genes among genes with a history of positive selection on the human branch. However, the target of selection in genes that were both differentially expressed between reproductively mature and immature ovarian tissue and also exhibited evidence for selection in the lineage leading to humans, was much more likely to have been a gene regulatory region than a coding

**Table 3**

The Overlap Get of Genes That 1) Show Significant *P* Values for Selection in Noncoding Regions in Both Humans and Chimpanzees and 2) Also Show Evidence for Significant Differential Expression by Life History Stage in the Baboon Ovary Gene Expression Data

Gene ID	Selection <i>P</i> Value Human Noncoding	Selection <i>P</i> Value Chimpanzee Noncoding	<i>P</i> Value Differential Expression Baboons
<i>serpina3</i>	0.01	0.04	$2.44 \times 10^{-16}$
<i>CHGA</i>	0.01	0	0.01
<i>LMO1</i>	0.01	0.03	$2.00 \times 10^{-05}$
<i>OSMR</i>	0	0.03	0
<i>DRG1</i>	0.03	0.05	0.02
<i>CAMP</i>	0	0	0
<i>dusp5</i>	0	0.03	0.03
<i>pfkfb3</i>	0.03	0	0
<i>SCUBE3</i>	0.01	0.04	0.04
<i>vwa2</i>	0.02	0.01	0.04

region (Fisher's exact test,  $P = 2.367 \times 10^{-08}$ ). If historical selection pressures on these loci were related to female maturation, changes in gene regulation may therefore have played an important role in the evolution of these traits in humans.

Genes expressed in reproductive tissue tend to be rapidly evolving exhibiting signatures of selection in multiple lineages (reviewed in Swanson and Vacquier 2002). We therefore examined whether differentially expressed genes were likely to be members of this rapidly evolving class or if they were specific to selection on the human branch. We asked whether noncoding regions that appear to have been positively selected on the chimpanzee (*Pan troglodytes*) lineage (Haygood et al. 2007) were similarly enriched for differential expression. We observed a similar number of differentially expressed genes by life history stage that correspond to positively selected regulatory regions in chimpanzees (21 in chimpanzees vs. the 19 seen in humans). Interestingly, 10 of these regions are shared between the two species, significantly more than expected by chance (hypergeometric test,  $P = 7.595 \times 10^{-18}$ ; table 3). These results suggest that positive selection on the specific aspects of ovarian maturation controlled by these genes may be a general characteristic of primate evolution. Indeed, genes involved in reproductive and immune pathways that evolved under selection in humans are often also under selection in other primates (reviewed in Vallender and Lahn 2004) and in mammals more generally (Kosiol et al. 2008). Our data suggest that this shared pattern of positive selection may apply to regulatory regions of reproductively important genes as well.

## Conclusion

The timing of female sexual maturity is one of many life-history traits that have shifted during primate and human

evolution, probably in response to selection. Our results suggest that there has been repeated selection on the *cis*-regulatory regions of some sexual maturity-related genes in multiple primate lineages. These loci are therefore of special interest in relationship to phenotypic evolution during reproductive maturation. Thus, examining the overlap of signatures of selection and differential gene expression from samples obtained from natural populations may serve as a useful filter for identifying loci of particular evolutionary or phenotypic interest. Although such opportunities will be uncommon, they promise to enrich our ability to interpret the phenotypic relevance of sequence-based signatures of selection.

## Materials and Methods

### Study Subjects

Samples used in this study were obtained from seven healthy females from the Amboseli baboon population in Kenya (supplementary fig. S1 and table S1, Supplementary Material online) and retrieved within 5–8 h of death. Tissue was stored in RNAlater (Ambion, Austin, TX) and transported to  $-20^{\circ}\text{C}$  storage in Nairobi within 24 h. Upon transport to the United States, samples were stored at  $-80^{\circ}\text{C}$ . To minimize the effects of cell type heterogeneity in the ovary, we sampled from the lateral ovarian cortex.

### Sample Preparation and Sequencing

Four micrograms of total RNA were isolated for each sample using an RNeasy kit (Qiagen, Valencia, CA) (supplementary table S1, Supplementary Material online) and used as input for the mRNA-Seq 8-Sample Prep Kit (Illumina, San Diego, CA). Libraries were sequenced on an Illumina GAIIx (one lane per sample) at the Yale University Keck Sequencing Core Facility. Approximately 15 million of 75 bp sequences resulted from each lane of sequencing.

### Baboon Gene Models

The current publicly available baboon genome assembly (Pham\_1.0, 20 November 2008) contains 387,373 linear scaffolds with approximately  $5.3\times$  coverage of the genome but has not yet been assembled into chromosomes (<http://www.hgsc.bcm.tmc.edu/project-species-p-Papio%20hamadryas.hgsc>). We mapped the RNA-Seq reads to the subset of these scaffolds (134,448 scaffolds with mean length of 20,246 bp) that mapped unambiguously to the macaque genome (Mmul\_051212, rhemac2) using lastz (Harris 2007). Overall, the subset covered 94.9% of the current rhesus macaque assembly. Gene models were obtained by mapping human RefSeq exons to the baboon genome with lastz in Galaxy (Taylor et al. 2007) with a 90% similarity cutoff based on previous estimates of human–baboon sequence conservation (Silva and Kondrashov 2002).

## Mapping Reads, Data Normalization, and Patterns of Differential Gene Expression

The RNA-Seq reads were mapped to the baboon scaffolds using “bowtie” (Langmead et al. 2009). Reads were defined as being within exon models using HTSeq (<http://www.huber.embl.de/users/anders/HTSeq/doc/overview.html>). Gene counts are the sum of the exon expression counts. The overall distributions of read counts were similar across all individuals and, more importantly, were not different between juveniles and adults, our primary axis of comparison (supplementary fig. S2, Supplementary Material online). Both exon counts and gene counts were normalized using the edgeR package (Robinson et al. 2010) in R (R Development Core Team 2008).

To evaluate the effect of maturation stage on specific genes, we used a generalized linear model with a negative binomial error structure to model variation in gene expression for each gene. Gene expression counts represented the response variable, and life history stage was modeled as a binary explanatory variable (juvenile or adult). We eliminated seven genes from this analysis that exhibited a significant relationship between gene expression and admixture-related genetic background as well as a relationship with life history stage (admixture between *P. cynocephalus* and a sister taxon, *P. anubis*, has previously been documented in this population and presented a possible confounder: Alberts and Altmann 2001; Tung et al. 2008). FDR corrections for multiple comparisons were performed using the Benjamini–Hochberg method (Benjamini and Hochberg 1995) at an FDR of 5% (fig. 1).

## Categorical Enrichment Analyses and Alternative Exon Usage

To determine functional categorical enrichment for the differentially expressed genes, we employed the PANTHER (HMM Library Version 6.0) (Mi et al. 2005) and GO (Gene Ontology Consortium 2000) databases and computed enrichment scores using Wilcoxon rank tests. Our background set of genes was composed only of genes measured in this study.

## Supplementary Material

Supplementary figures S1–S3 and tables S1–S4 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

Raphael Mututua, Serah Sayialel, Moonyoi ole Parsetau, Lenkai ole Rikoyan, and Kinyua Warutere contributed skillfully and in difficult circumstances to collecting the samples and the long-term data used in this analysis. We thank the Office of the President, Republic of Kenya, the Kenya

Wildlife Service, its Amboseli staff and Wardens, the National Museums of Kenya, and the members of the Amboseli-Longido pastoralist communities for permission to work in Amboseli. We thank Jeanne Altmann for providing access to long-term data and contributing samples, Baylor College of Medicine (Human Genome Sequencing Center) for the preliminary draft assembly of the baboon genome, Ralph Wiedmann and Olivier Fedrigo for aid in data analysis, Adam Pfefferle for assistance in data generation, and David Garfield for comments on the manuscript. This work was supported by the Duke Primate Genomics Initiative; National Science Foundation (NSF) BCS-0323553, NSF DEB 0919200, National Institute on Aging (NIA) R01AG034513-01, and NIA P01-AG031719 to S.C.A.; NSF-BCS-08-27552 (HOMINID) to G.A.W.; and NSF DEB-0846286 to S.C.A. and G.A.W.

## Literature Cited

- Abulafia O, Sherer DM. 2000. Angiogenesis of the ovary. *Am J Obstet Gynecol.* 182:240–246.
- Adib TR, et al. 2004. Predicting biomarkers for ovarian cancer using gene-expression microarrays. *Br J Cancer.* 90:686–692.
- Alberts SC, Altmann J. 2001. Immigration and hybridization patterns of yellow and anubis baboons in and around Amboseli, Kenya. *Am J Primatol.* 53:139–154.
- Alberts SC, Buchan JC, Altmann J. 2006. Sexual selection in wild baboons: from mating opportunities to paternity success. *Anim Behav.* 72:1177–1196.
- Altmann SA, Altmann J. 1970. Baboon ecology: African field research. Chicago: University of Chicago Press and Basel (Switzerland): S. Karger (Bibliotheca Primatologica, no. 12).
- Babbitt CC, et al. 2010. Both noncoding and protein-coding RNAs contribute to gene expression evolution in the primate brain. *Genome Biol Evol.* 2010:67–79.
- Barberan-Soler S, Zahler AM. 2008. Alternative splicing regulation during *C. elegans* development: splicing factors as regulated targets. *PLoS Genet.* 4:e1000001.
- Behrensmeyer AK. 2006. Climate change and human evolution. *Science* 311:476–478.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 57:289–300.
- Blekhman R, et al. 2010. Sex-specific and lineage-specific alternative splicing in primates. *Genome Res.* 20:180–189.
- Blurton Jones N, Hawkes K, O’Connell JF. 1999. Some current ideas about the evolution of human life history. In: Lee PC, editor. *Comparative primate socioecology*. Cambridge: Cambridge University Press. p. 140–166.
- Bogin B, Smith BH. 1996. Evolution of the human life cycle. *Am J Hum Biol.* 8:703–716.
- Brown HM, et al. 2010. Adamts1 cleavage of versican mediates essential structural remodeling of the ovarian follicle and cumulus-oocyte matrix during ovulation in mice. *Biol Reprod.* 83:549–557.
- Buchan JC, Alberts SC, Silk JB, Altmann J. 2003. True paternal care in a multi-male primate society. *Nature* 425:179–181.
- Bukovsky A, Caudle MR. 2008. Immune physiology of the mammalian ovary—a review. *Am J Reprod Immunol.* 59:12–26.
- Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat Genet.* 25:25–29.

- Gunz P, Neubauer S, Maureille B, Hublin J. 2010. Brain development after birth differs between Neanderthals and modern humans. *Curr Biol*. 20:R921–R922.
- Hahm S, et al. 1999. Targeted deletion of the Vgf gene indicates that the encoded secretory peptide precursor plays a novel role in the regulation of energy balance. *Neuron* 23:537–548.
- Harris RS. 2007. Improved pairwise alignment of genomic DNA. *Computer Science and Engineering*. State College (PA): Pennsylvania State University. p. 74.
- Haviv I, Campbell IG. 2002. DNA microarrays for assessing ovarian cancer gene expression. *Mol Cell Endocrinol*. 191:121–126.
- Haygood R, Babbitt CC, Fedrigo O, Wray GA. 2010. Contrasts between adaptive coding and non-coding changes during human evolution. *Proc Natl Acad Sci U S A*. 107:7853–7857.
- Haygood R, et al. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet*. 39:1140–1144.
- Hoang YD, Nakamura BN, Luderer U. 2009. Follicle-stimulating hormone and estradiol interact to stimulate glutathione synthesis in rat ovarian follicles and granulosa cells. *Biol Reprod*. 81:636–646.
- Jethwa PH, Ebling FJ. 2008. Role of VGF-derived peptides in the control of food intake, body weight and reproduction. *Neuroendocrinology* 88:80–87.
- Jo M, Curry TE Jr. 2004. Regulation of matrix metalloproteinase-19 messenger RNA expression in the rat ovary. *Biol Reprod*. 71:1796–1806.
- Jolly CJ. 1993. Species, subspecies, and baboon systematics. In: Kimbel WH, Martin LB, editors. *Species, species concepts, and primate evolution*. New York: Plenum Press. p. 67–107.
- Kosiol C, et al. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet*. 4(8):e1000144.
- Kostka D, Hahn MW, Pollard KS. 2010. Noncoding sequences near duplicated genes evolve rapidly. *Genome Biol Evol*. 2:518–533.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10:R25.
- Luderer U, Diaz D, Faustman EM, Kavanagh TJ. 2003. Localization of glutamate cysteine ligase subunit mRNA within the rat ovary and relationship to follicular apoptosis. *Mol Reprod Dev*. 65:254–261.
- Luo CW, Kawamura K, Klein C, Hsueh AJ. 2004. Paracrine regulation of ovarian granulosa cell differentiation by stanniocalcin (stc) 1: mediation through specific stc1 receptors. *Mol Endocrinol*. 18:2085–2096.
- Mace R. 2000. Evolutionary ecology of human life history. *Anim Behav*. 59:1–10.
- Matzuk MM, Lamb DJ. 2008. The biology of infertility: research advances and clinical challenges. *Nat Med*. 14:1197–1213.
- McGee EA, Hsueh AJ. 2000. Initial and cyclic recruitment of ovarian follicles. *Endocr Rev*. 21:200–214.
- Mi HY, et al. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res*. 33:D284–D288.
- Nef S, et al. 2005. Gene expression during sex determination reveals a robust female genetic program at the onset of ovarian development. *Dev Biol*. 287:361–377.
- Nielsen R, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*. 3:976–985.
- Paciga M, Watson AJ, DiMattia GE, Wagner GF. 2002. Ovarian stanniocalcin is structurally unique in mammals and its production and release are regulated through the luteinizing hormone receptor. *Endocrinology* 143:3925–3934.
- Potts R. 1998. Variability selection in hominid evolution. *Evol Anthropol*. 7:81–96.
- R Development Core Team. 2008. R: a language and environment for statistical computing [Internet]. Version 2.11.0. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org>.
- Redmer DA, Reynolds LP. 1996. Angiogenesis in the ovary. *Rev Reprod*. 1:182–192.
- Revil T, et al. 2010. Alternative splicing is frequent during early embryonic development in mouse. *BMC Genomics* 11:399.
- Rhine RJ, Norton GW, Wasser SK. 2000. Lifetime reproductive success, longevity, and reproductive life history of female yellow baboons (*Papio cynocephalus*) of Mikumi National Park, Tanzania. *Am J Primatol*. 51:229–241.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Silva JC, Kondrashov AS. 2002. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. *Trends Genet*. 18:544–547.
- Small CL, et al. 2005. Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biol Reprod*. 72:492–501.
- Somel M, et al. 2009. Transcriptional neoteny in the human brain. *Proc Natl Acad Sci U S A*. 106:5743–5748.
- Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev*. 3:137–144.
- Taylor J, Schenck I, Blankenburg D, Nekrutenko A. 2007. Using galaxy to perform large-scale interactive data analyses. *Curr Protoc Bioinformatics*. 19:10.15.11–10.15.25.
- Tomaselli S, et al. 2011. Human RSPO1/R-spondin1 is expressed during early ovary development and augments beta-catenin signaling. *PLoS One* 6:e16366.
- Tung J, Alberts SC, Wray GA. 2010. Evolutionary genetics in wild primates: combining genetic approaches with field studies of natural populations. *Trends Genet*. 26:353–362.
- Tung J, et al. 2008. Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Mol Ecol*. 17:1998–2011.
- Uddin M, et al. 2008. Distinct genomic signatures of adaptation in pre- and postnatal environments during human evolution. *Proc Natl Acad Sci U S A*. 105:3215–3220.
- Vallender EJ, Lahn BT. 2004. Positive selection on the human genome. *Hum Mol Genet*. 13:R245–R254.
- Wang K, et al. 1999. Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene* 229:101–108.
- Wasser SK, Norton GW, Kleindorfer S, Rhine RJ. 2004. Population trend alters the effects of maternal dominance rank on lifetime reproductive success in yellow baboons (*Papio cynocephalus*). *Behav Ecol Sociobiol*. 56:338–345.
- Welsh JB, et al. 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci U S A*. 98:1176–1181.
- Xu AG, et al. 2010. Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by RNA-Seq. *PLoS Comput Biol*. 6:e1000843.
- Yang Z, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 15:496–503.

**Associate editor:** Michael Purugganan