Genome Biology and Evolution Advance Access published December 8, 2011 doi:10.1093/gbe/evr134

Changes in gene expression associated with reproductive maturation in wild female baboons

Courtney C. Babbitt[1,2*], Jenny Tung[1,3‡], Gregory A. Wray[1,2,3], Susan C. Alberts[1,2,3,4]

1. Department of Biology, Duke University, Durham, NC, 27708, USA
2. Institute for Genome Sciences & Policy, Duke University, Durham, NC, 27708, USA
3. Department of Evolutionary Anthropology, Duke University, Durham, NC, 27708, USA
4. Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya.

* Corresponding Author:
Courtney C. Babbitt
Mailing address: Box 90338
Dept. of Biology
Durham, NC, 27708
Phone: 919-668-6249
Email: courtney.babbitt@duke.edu

Abstract

Changes in gene expression during development play an important role in shaping morphological and behavioral differences, including between humans and nonhuman primates. While 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.

Keywords: 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 to non-human 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).

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 non-seasonal 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 and Altmann 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 and Altmann 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 (Figure S1, Table S1). We obtained ~15 million reads per individual (Table S1) and we measured the expression of a total of 9770 genes in the baboon ovary. Ninety-seven genes (~1% of genes in the data set) were differentially expressed between the juveniles and the adults (FDR adjusted p-value < 0.05) (Figure 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. 2010; Blekhman et al. 2010; Xu et al. 2010). Of the differentially expressed genes, 79 were upregulated in the adults and 18 upregulated in the juveniles. This imbalance in upregulated expression towards the adult females was expected, as the adult ovary is much more metabolically active than the pre-menarcheal ovary (reviewed in McGee and Hsueh 2000).

To evaluate the global effect of maturation stage on gene expression variation, we performed a principal components analysis. The first three PCs in this analysis explained

~67% of variation in the gene expression data (Figure S3). 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 towards higher values for juveniles than for adults (Mann-Whitney test, W = 0, p-value = 0.05714 and Figure S3).

Little is known about ovarian gene expression in human or mouse models during either the premenarcheal stage or in non-cycling 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 (The Gene Ontology Consortium 2000) and PANTHER (Mi et al. 2005) ontology databases. The enrichments were performed in 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, Tables S2 and S3).

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 (Table S3). Secondly, 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 S2). 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) (Figure 2). *MMP19* and *ADAMTS1* function to remodel the extracellular matrix as follicles develop (Jo and Curry 2004; Brown et al. 2010). The role of *VGF* is less clear, but its essential role has been demonstrated in *VGF* -\- 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 (Figure 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 (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 (Table S4). 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 and colleagues (Nielsen et al. 2005) and for putative regulatory regions 5 kilobases (kb) upstream of genes from Haygood and colleagues (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) to 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 (out 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 region (Fisher's Exact test, p=2.367e-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, ten of these regions are shared between the two species, significantly more than expected by chance (hypergeometric test, p=7.595e-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 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 (Figure S1 and Table S1), and retrieved within 5-8 hours of death. Tissue was stored in RNAlater (Ambion, Austin, Texas) and transported to -20°C storage in Nairobi within 24 hours. Upon transport to the United States, samples were stored at -80°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)(Table S1), 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. ~15 million 75 base pair 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 5.3x 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 base pairs) 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 (Figure S2). 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). False discovery rate corrections for multiple comparisons were performed using the Benjamini-Hochberg method (Benjamini and Hochberg 1995) at an FDR of 5% (Figure 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 (The Gene Ontology Consortium 2000) databases and computed enrichment scores using Wilcoxon-rank tests. Our background set of genes was comprised only of genes measured in this study.

Acknowledgements

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; NSF BCS-0323553,

NSF DEB 0919200, 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.

References

Abulafia, O., and D. M. Sherer. 2000. Angiogenesis of the ovary. Am J Obstet Gynecol 182:240-246.

Adib, T. R. et al. 2004. Predicting biomarkers for ovarian cancer using gene-expression microarrays. Br J Cancer 90:686-692.

Alberts, S. C., and J. Altmann. 2001. Immigration and hybridization patterns of yellow and anubis baboons in and around amboseli, kenya. Am J Primatol 53:139-154.

Alberts, S. C.J. C. Buchan, and J. Altmann. 2006. Sexual selection in wild baboons: From mating opportunities to paternity success. Animal Behaviour 72:1177-1196.

Altmann, S. A., and J. Altmann. 1970. Baboon ecology: African field research. S. Karger.

Babbitt, C. C. et al. 2010. Both noncoding and protein-coding rnas contribute to gene expression evolution in the primate brain. Genome Biology and Evolution 2010:67-79.

Barberan-Soler, S., and A. M. Zahler. 2008. Alternative splicing regulation during c. Elegans development: Splicing factors as regulated targets. PLoS Genet 4:e1000001.

Behrensmeyer, A. K. 2006. Climate change and human evolution. Science 311:476-478.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B-Methodological 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. B.K. Hawkes, and J. F. O'Connell. 1999. Some current ideas about the evolution of human life history. Pp. 140-166 in P. C. Lee, ed. Comparative primate socioecology. Cambridge University Press, Cambridge.

Bogin, B., and B. H. Smith. 1996. Evolution of the human life cycle. American Journal of Human Biology 8:703-716.

Brown, H. M. 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, J. C.S. C. AlbertsJ. B. Silk, and J. Altmann. 2003. True paternal care in a multi-male primate society. Nature 425:179-181.

Bukovsky, A., and M. R. Caudle. 2008. Immune physiology of the mammalian ovary - a review. Am J Reprod Immunol 59:12-26.

361 Gunz, P.S. NeubauerB. Maureille, and J. Hublin. 2010. Brain development after birth
362 differs between neanderthals and modern humans. Current Biology 20:R921-
363 R922.
364 Hahm, S. et al. 1999. Targeted deletion of the vgf gene indicates that the encoded
365 secretory peptide precursor plays a novel role in the regulation of energy
366 balance. Neuron 23:537-548.
367 Harris, R. S. 2007. Improved pairwise alignment of genomic DNA. Pp. 74. Computer
368 Science and Engineering. Pennsylvania State University.
369 Haviv, I., and I. G. Campbell. 2002. DNA microarrays for assessing ovarian cancer
370 gene expression. Mol Cell Endocrinol 191:121-126.
371 Haygood, R.C. C. BabbittO. Fedrigo, and G. A. Wray. 2010. Contrasts between
372 adaptive coding and non-coding changes during human evolution.
373 Proceedings of the National Academy of Sciences of the United States of
374 America 107:7853-7857.
375 Haygood, R. et al. 2007. Promoter regions of many neural- and nutrition-related
376 genes have experienced positive selection during human evolution. Nature
377 Genetics 39:1140-1144.
378 Hoang, Y. D.B. N. Nakamura, and U. Luderer. 2009. Follicle-stimulating hormone and
379 estradiol interact to stimulate glutathione synthesis in rat ovarian follicles
380 and granulosa cells. Biol Reprod 81:636-646.
381 Jethwa, P. H., and F. J. Ebling. 2008. Role of vgf-derived peptides in the control of
382 food intake, body weight and reproduction. Neuroendocrinology 88:80-87.
383 Jo, M., and T. E. Curry, Jr. 2004. Regulation of matrix metalloproteinase-19
384 messenger rna expression in the rat ovary. Biol Reprod 71:1796-1806.
385 Jolly, C. J. 1993. Species, subspecies, and baboon systematics. Pp. 67–107 in W. H.
386 Kimbel, and L. B. Martin, eds. Species, species concepts, and primate
387 evolution. Plenum Press, New York.
388 Kosiol, C. et al. 2008. Patterns of positive selection in six mammalian genomes. PLos
389 Genetics 4.
390 Kostka, D.M. W. Hahn, and K. S. Pollard. 2010. Noncoding sequences near duplicated
391 genes evolve rapidly. Genome Biol Evol 2:518-533.
392 Langmead, B.C. TrapnellM. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-
393 efficient alignment of short DNA sequences to the human genome. Genome
394 Biol 10:R25.
395 Luderer, U.D. DiazE. M. Faustman, and T. J. Kavanagh. 2003. Localization of
396 glutamate cysteine ligase subunit mrna within the rat ovary and relationship
397 to follicular apoptosis. Mol Reprod Dev 65:254-261.
398 Luo, C. W.K. KawamuraC. Klein, and A. J. Hsueh. 2004. Paracrine regulation of
399 ovarian granulosa cell differentiation by stanniocalcin (stc) 1: Mediation
400 through specific stc1 receptors. Mol Endocrinol 18:2085-2096.
401 Mace, R. 2000. Evolutionary ecology of human life history. Anim Behav 59:1-10.
402 Matzuk, M. M., and D. J. Lamb. 2008. The biology of infertility: Research advances
403 and clinical challenges. Nat Med 14:1197-1213.
404 McGee, E. A., and A. J. Hsueh. 2000. Initial and cyclic recruitment of ovarian follicles.
405 Endocr Rev 21:200-214.

406 Mi, H. Y. et al. 2005. The panther database of protein families, subfamilies, functions
407 and pathways. Nucleic Acids Research 33:D284-D288.
408 Nef, S. et al. 2005. Gene expression during sex determination reveals a robust female
409 genetic program at the onset of ovarian development. Dev Biol 287:361-377.
410 Nielsen, R. et al. 2005. A scan for positively selected genes in the genomes of humans
411 and chimpanzees. Plos Biology 3:976-985.
412 Paciga, M.A. J. WatsonG. E. DiMattia, and G. F. Wagner. 2002. Ovarian stanniocalcin is
413 structurally unique in mammals and its production and release are regulated
414 through the luteinizing hormone receptor. Endocrinology 143:3925-3934.
415 Potts, R. 1998. Variability selection in hominid evolution. Evol. Anthropology 7:81-
416 96.
417 R Development Core Team. 2008. R: A language and environment for statistical
418 computing. Pp. http://www.R-project.org. R Foundation for Statistical
419 Computing, Vienna, Austria.
420 Redmer, D. A., and L. P. Reynolds. 1996. Angiogenesis in the ovary. Rev Reprod
421 1:182-192.
422 Revil, T. et al. 2010. Alternative splicing is frequent during early embryonic
423 development in mouse. BMC Genomics 11:399.
424 Rhine, R. J.G. W. Norton, and S. K. Wasser. 2000. Lifetime reproductive success,
425 longevity, and reproductive life history of female yellow baboons (papio
426 cynocephalus) of mikumi national park, tanzania. Am J Primatol 51:229-241.
427 Robinson, M. D.D. J. McCarthy, and G. K. Smyth. 2010. Edger: A bioconductor package
428 for differential expression analysis of digital gene expression data.
429 Bioinformatics 26:139-140.
430 Silva, J. C., and A. S. Kondrashov. 2002. Patterns in spontaneous mutation revealed
431 by human-baboon sequence comparison. Trends Genet 18:544-547.
432 Small, C. L. et al. 2005. Profiling gene expression during the differentiation and
433 development of the murine embryonic gonad. Biol Reprod 72:492-501.
434 Somel, M. et al. 2009. Transcriptional neoteny in the human brain. Proceedings of
435 the National Academy of Sciences of the United States of America 106:5743-
436 5748.
437 Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive
438 proteins. Nature reviews 3:137-144.
439 Taylor, J.I. SchenckD. Blankenburg, and A. Nekrutenko. 2007. Using galaxy to
440 perform large-scale interactive data analyses. Curr Protoc Bioinformatics
441 19:10.15.11-10.15.25.
442 The Gene Ontology Consortium. 2000. Gene ontology: Tool for the unification of
443 biology. Nature Genetics 25:25-29.
444 Tomaselli, S. et al. 2011. Human rspo1/r-spondin1 is expressed during early ovary
445 development and augments beta-catenin signaling. PLoS One 6:e16366.
446 Tung, J.S. C. Alberts, and G. A. Wray. 2010. Evolutionary genetics in wild primates:
447 Combining genetic approaches with field studies of natural populations.
448 Trends Genet 26:353-362.
449 Tung, J. et al. 2008. Genetic evidence reveals temporal change in hybridization
450 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, E. J., and B. T. Lahn. 2004. Positive selection on the human genome. Human Molecular Genetics 13:R245-R254.

Wang, K. et al. 1999. Monitoring gene expression profile changes in ovarian carcinomas using cdna microarray. Gene 229:101-108.

Wasser, S. K.G. W. NortonS. Kleindorfer, and R. J. Rhine. 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, J. B. 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, A. G. et al. 2010. Intergenic and repeat transcription in human, chimpanzee and macaque brains measured by rna-seq. PLoS Comput Biol 6:e1000843.

Yang, Z., and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends in Ecology & Evolution 15:496-503.

Tables

Table 1. PANTHER Biological Process categorical enrichments. 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 10 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.

Table 2. Differentially expressed genes (FDR adjusted p-value < 0.05) in the adult and juvenile baboon ovary.

Table 3. The overlap set 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.

Table S1. Individual information and sample data.

Table S2. All categorical enrichments using both the PANTHER and GO ontologies of the absolute differences in expression between the juveniles and adults. The right-hand column shows the total number of genes evaluated in that category. Categories that included less than 10 genes total were analyzed but are not shown.

Table S3. All categorical enrichments using both the PANTHER and GO ontologies of genes ranked by upregulated expression in the adults. The right-hand column shows the total number of genes evaluated in that category. Categories that included less than 10 genes total were analyzed but are not shown.

Table S4. Genes that exhibit evidence for potential differences in isoform usage (based on inclusion of at least one differentially expressed exon at an FDR adjusted p-value < 0.05 and at least one exon with no evidence for differential expression at an FDR adjusted p-value > 0.15).

Figures

Figure 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).

Figure 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.

Figure S1. Pedigree of the individuals included in this study. Individuals included are labeled in green, with adult females in dark green and juveniles in light green. Although individuals are closely related, the closest pairs of relatives are in different age classes, making our comparisons between age classes conservative with respect to relatedness. The individual male labeled in white is the unidentified father of VEL.

Figure S2. Comparison of transcript expression level distributions in the adults and juveniles. To ensure that we were sampling from similar distributions for both the adults and the juveniles, we plotted a density distribution of the mean of normalized count data for the adults (blue) and the juveniles (red) (K-S test, $D = 0.1037$, p-value = 0.7904).

Figure S3. The first three principal components of the normalized ovarian gene expression data. Adults are plotted in blue and juveniles are plotted in red. A) First three PCs using the full gene expression data set. B) First three PCs of only the genes that were differentially expressed between the juveniles and the adults.

Table 1. PANTHER Biological Process categorical enrichments

PANTHER Biological Process category	*p-value*	*total occurence*
Signal transduction	1.58E-09	1359
Cell surface receptor mediated signal transduction	2.95E-08	558
Cell communication	1.73E-07	435
Immunity and defense	9.16E-07	497
Ligand-mediated signaling	7.77E-06	111
Neuronal activities	4.08E-05	201
Cell motility	0.0003074	151
G-protein mediated signaling	0.0005295	228
Other neuronal activity	0.0008809	64
Cytokine and chemokine mediated signaling pathway	0.001396	74
Developmental processes	0.001563	903
B-cell- and antibody-mediated immunity	0.001653	28
Skeletal development	0.004008	59
Interferon-mediated immunity	0.004157	20
Homeostasis	0.005757	89
Macrophage-mediated immunity	0.00814	34
Cell adhesion	0.008292	252
Extracellular matrix protein-mediated signaling	0.0083	39
Ectoderm development	0.00986	272
Blood circulation and gas exchange	0.01002	21
Neurogenesis	0.01072	250
Mesoderm development	0.01156	251
Cell adhesion-mediated signaling	0.0126	139
Cytokine/chemokine mediated immunity	0.01312	27
Angiogenesis	0.01572	38
Detoxification	0.01719	38
Fatty acid metabolism	0.01925	85
Anion transport	0.0204	25
MHCII-mediated immunity	0.02078	16
Other receptor mediated signaling pathway	0.02398	81
Extracellular transport and import	0.0256	32
Sensory perception	0.02823	96
JAK-STAT cascade	0.02933	30
Natural killer cell mediated immunity	0.03117	11

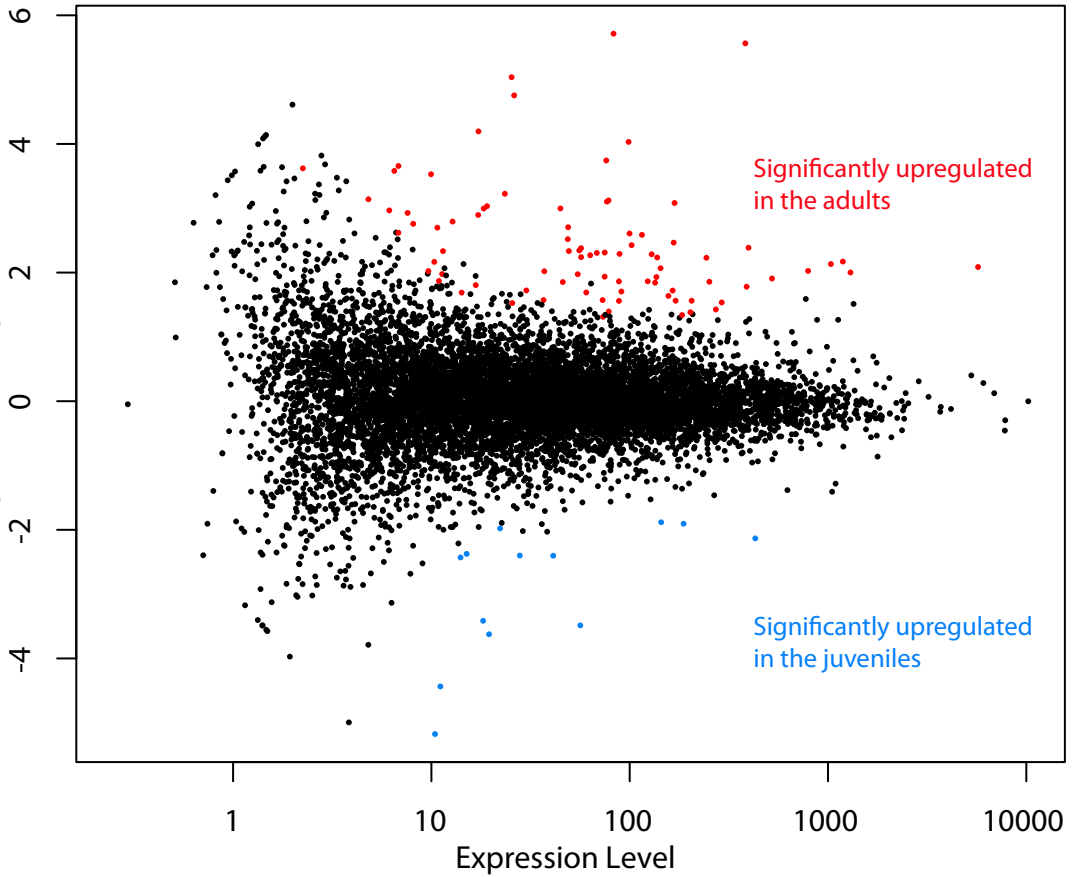
Table 2. Differentially expressed genes (FDR adjusted p-value < 0.05) in the adult and juvenile baboon ovary

Table 2. Differentially expressed genes (FDR adjusted p-value < 0.05) in the adult and juvenile baboon ovary

Gene ID	log FC	p-value	p-value FDR	Gene ID	log FC	p-value	p-value FDR
serpina3	5.56053945	2.44E-16	2.39E-12	EPO	-2.3597099	6.37E-05	0.01219865
ADAMTS4	5.09716829	5.79E-15	2.83E-11	SLC7A8	2.85383446	6.49E-05	0.01219865
REN	3.98944937	5.39E-12	1.76E-08	ERRFI1	2.13531073	7.99E-05	0.01472884
TFPI2	5.09331391	1.82E-11	3.58E-08	MYOC	-4.5033477	8.42E-05	0.0152368
ADAMTS6	4.69173617	1.83E-11	3.58E-08	Mmp1	8.22694892	8.90E-05	0.01581064
Melk	3.76852295	2.12E-10	3.44E-07	rspo1	-2.7082948	0.00010143	0.01769646
LRG1	4.68207316	4.18E-10	5.83E-07	HIF1A	2.20582485	0.00010339	0.01772116
FABP4	9.28927774	3.47E-09	4.24E-06	rpl21	3.86520992	0.0001073	0.01807402
CH25H	3.20323737	1.81E-07	0.00018065	OSMR	2.19076016	0.00011098	0.01837741
SOST	-4.8295275	1.85E-07	0.00018065	TIMP1	2.00766916	0.00011652	0.01897307
IL1RL1	9.32559106	3.14E-07	0.0002791	CYP21A2	2.59714864	0.00012208	0.0193343
F3	2.72902443	4.14E-07	0.00033246	PAPPA	3.66989958	0.00012459	0.0193343
RGS2	2.94854719	4.42E-07	0.00033246	Adamts1	-3.7455767	0.0001265	0.0193343
GALNT9	-3.4824072	6.16E-07	0.00042976	Trem1	1.97618174	0.0001301	0.0193343
TNFAIP6	9.06613349	6.82E-07	0.00044438	RAB38	-2.3456431	0.00013012	0.0193343
CRTAC1	-3.1329928	7.44E-07	0.00045431	Sgip1	2.17468855	0.00013061	0.0193343
C19orf26	2.79022659	1.08E-06	0.00062056	DDX21	2.13304855	0.00013343	0.01945687
LdhA	2.80233251	1.91E-06	0.00102984	f2rl1	3.0581654	0.00014459	0.0207748
stc1	2.95769089	2.00E-06	0.00102984	SBNO2	2.8690604	0.00015474	0.02160406
Gdf15	2.63705264	2.61E-06	0.00127693	S100A8	3.74690504	0.00015479	0.02160406
FCER1G	3.18705038	2.77E-06	0.00129071	DST	3.67689535	0.00016231	0.02233414
VGF	3.53951033	3.72E-06	0.00165083	AADAC	7.97969292	0.00016903	0.02293631
MMP19	2.49571312	5.03E-06	0.00213603	CHI3L1	2.26915607	0.00017834	0.0238048
Fosl2	2.56131942	5.46E-06	0.00222307	H6pd	2.1574967	0.0001803	0.0238048
SERPINE1	3.00295543	6.06E-06	0.00226587	ADAMTS16	-2.5741504	0.00018334	0.02388255
S100A9	2.80444185	6.24E-06	0.00226587	HLA-DQB1	3.33877719	0.00019359	0.02488701
ADPRHL1	3.46982396	6.26E-06	0.00226587	CTSG	2.8539012	0.00021891	0.02777611
Cd163l1	2.93708735	7.60E-06	0.00265144	RPF2	2.32131206	0.00022495	0.02817639
socs3	2.49813522	9.29E-06	0.00306778	Cd48	7.59418959	0.00023965	0.02877211
ifi30	2.43485712	9.42E-06	0.00306778	tnfrsf11b	2.68764743	0.0002414	0.02877211
CHGB	2.68694397	1.04E-05	0.00327812	C10orf10	1.97584087	0.00024149	0.02877211
Cntn4	6.69229158	1.41E-05	0.00430415	KCNN4	-8.0092773	0.00024933	0.02934901
Il1r1	2.38670209	1.64E-05	0.00484285	IL8	4.44362712	0.00025339	0.02947116
AG2	7.88467942	1.87E-05	0.00507962	ZFP36	1.96860745	0.00026324	0.02998838
GADD45A	2.30721084	1.89E-05	0.00507962	DLK1	1.87823533	0.00026397	0.02998838
LMO1	2.31100372	1.97E-05	0.00507962	GALR3	-3.0699664	0.00027127	0.0304633
TNFAIP3	-3.1164769	2.00E-05	0.00507962	ANKRD31	7.59845327	0.00027525	0.03055914
ANKRD1	7.90497765	2.01E-05	0.00507962	TRIB1	2.3659021	0.00029898	0.03274912
gpr84	2.20041148	2.08E-05	0.00507962	apol3	2.42154371	0.00030168	0.03274912
NUP35	2.25111635	2.11E-05	0.00507962	PPARGC1A	-7.848901	0.00031593	0.03366667
LCNL1	-8.1713233	2.13E-05	0.00507962	PTCHD1	2.54755259	0.00031702	0.03366667
cebpd	2.21602916	2.28E-05	0.00530536	EFNA5	-2.683381	0.00037376	0.03926515
NR5A2	3.01054291	2.37E-05	0.00537637	LGALS3	1.95933307	0.00038655	0.04017603
TMEM49	2.23411164	3.46E-05	0.00767944	c2cd4c	1.86851938	0.00039556	0.04042629
GZMB	4.67164885	3.73E-05	0.00810511	NFIL3	-3.0702127	0.00039723	0.04042629
SLC16A10	6.39312229	4.21E-05	0.00894971	WNT6	-1.9481788	0.0004573	0.04606028
ptgds	-2.2794722	4.40E-05	0.00913917	CAMP	7.71716682	0.00047286	0.04714096
HPGDS	8.25515497	5.22E-05	0.01062847	ism1	1.96334052	0.00050099	0.04944076
CD163	2.06718683	5.90E-05	0.01153785				

Table 3. Overlap of genes showing noncoding selection on both the human and chimpanzee branch that also show evidence of differential expression by age in the baboon ovary

Gene ID	selection pvalue human noncoding	selection pvalue chimpazee noncoding	p-value diff. expression baboons
serpina3	0.0112343	0.0370808	2.44E-16
CHGA	0.01111	0.00016589	0.00831165
LMO1	0.00568249	0.0295031	2.00E-05
OSMR	0.00401283	0.0294791	0.00011098
DRG1	0.0255383	0.052671	0.01727405
CAMP	0.00076639	0.0011117	0.00047286
dusp5	0.00014616	0.0303341	0.02844818
pfkfb3	0.0260481	0.00239918	0.0038824
SCUBE3	0.009715	0.0404337	0.03614428
vwa2	0.0214515	0.0129528	0.0350298


6
4
2
0
-2
-4
1
10
100
1000
10000
Expression Level
Significantly upregulated
in the adults
Significantly upregulated
in the juveniles

ure 1

Figure 1