Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy

Journal Article

Author(s):
Lynch, Vincent J.; Nnamani, Mauris C.; Kapusta, Aurélie; Brayer, Kathryn; Plaza, Silvia L.; Mazur, Eric C.; Emera, Deena; Sheikh, Shehzad Z.; Grützner, Frank; Bauersachs, Stefan; Graf, Alexander; Young, Steven L.; Lieb, Jason D.; DeMayo, Francesco J.; Feschotte, Cédric; Wagner, Günter P.

Publication date:
2015-02-03

Permanent link:
https://doi.org/10.3929/ethz-a-010581094

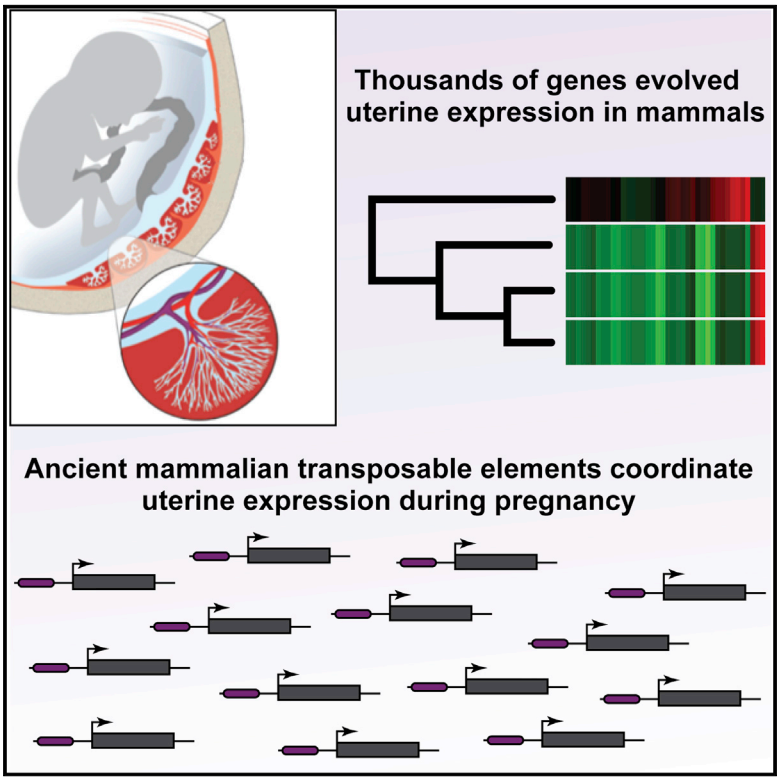
Originally published in:
Cell Reports 10(4), https://doi.org/10.1016/j.celrep.2014.12.052

Article

Cell Reports

Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy

Graphical Abstract



Authors

Vincent J. Lynch, Mauris C. Nnamani, ..., Cédric Feschotte, Günter P. Wagner

Correspondence

vjlynch@uchicago.edu

In Brief

How morphological novelties originate is a major unanswered question in developmental and evolutionary biology. Lynch et al. demonstrate that thousands of genes evolved uterine expression during the evolution of mammalian pregnancy and that recruitment of new gene expression in the uterus was likely mediated by ancient mammalian transposable elements.

Highlights

- Thousands of genes gained uterine expression during the origins of pregnancy
- Genes that mediate maternal-fetal immunotolerance evolved expression in Eutherians
- Ancient transposable elements donated *cis*-regulatory elements to recruited genes
- Ancient transposable elements coordinate the uterine progesterone response

Accession Numbers

GSE57714
GSE30708
GSE29553
GSE21046
GSE43667
GSE48862
GSE61793
GSM1011119

Lynch et al., 2015, Cell Reports *10*, 551–561
February 3, 2015
http://dx.doi.org/10.1016/j.celrep.2014.12.052

Cell Reports
Article

Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy

Vincent J. Lynch,[1,*] Mauris C. Nnamani,[2] Aurélie Kapusta,[3] Kathryn Brayer,[2,10] Silvia L. Plaza,[2] Erik C. Mazur,[8] Deena Emera,[2,11] Shehzad Z. Sheikh,[4] Frank Grützner,[5] Stefan Bauersachs,[6,12] Alexander Graf,[6] Steven L. Young,[7] Jason D. Lieb,[1] Francesco J. DeMayo,[8,9] Cédric Feschotte,[3] and Günter P. Wagner[2]

[1]Department of Human Genetics, The University of Chicago, 920 East 58th Street, CLSC 319C, Chicago, IL 60637, USA
[2]Yale Systems Biology Institute and Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA
[3]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA
[4]Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[5]The Robinson Institute, School of Molecular and Biomedical Sciences, University of Adelaide, Adelaide, SA 5005, Australia
[6]Laboratory for Functional Genome Analysis (LAFUGA), Gene Center, LMU Munich, Feodor Lynen Strasse 25, 81377 Munich, Germany
[7]Department of Obstetrics and Gynecology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27705, USA
[8]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA
[9]Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX 77030, USA
[10]Present address: Department of Internal Medicine, MSC08 4630, University of New Mexico, 2325 Camino de Saludo NE, CRF 121, Albuquerque, NM 87131-0001, USA
[11]Present address: Department of Genetics, Yale University Medical School, 336 Cedar Street, New Haven, CT 06510, USA
[12]Present address: Animal Physiology, Institute of Agricultural Sciences, ETH Zurich, Universitaetstrasse 2, 8092 Zurich, Switzerland
*Correspondence: vjlynch@uchicago.edu
http://dx.doi.org/10.1016/j.celrep.2014.12.052

SUMMARY

A major challenge in biology is determining how evolutionarily novel characters originate; however, mechanistic explanations for the origin of new characters are almost completely unknown. The evolution of pregnancy is an excellent system in which to study the origin of novelties because mammals preserve stages in the transition from egg laying to live birth. To determine the molecular bases of this transition, we characterized the pregnant/gravid uterine transcriptome from tetrapods to trace the evolutionary history of uterine gene expression. We show that thousands of genes evolved endometrial expression during the origins of mammalian pregnancy, including genes that mediate maternal-fetal communication and immunotolerance. Furthermore, thousands of *cis*-regulatory elements that mediate decidualization and cell-type identity in decidualized stromal cells are derived from ancient mammalian transposable elements (TEs). Our results indicate that one of the defining mammalian novelties evolved from DNA sequences derived from ancient mammalian TEs co-opted into hormone-responsive regulatory elements distributed throughout the genome.

INTRODUCTION

A major challenge in biology is determining the genetic and molecular mechanisms that underlie phenotypic differences between species (Wagner and Lynch, 2010). While comparative studies in a few organisms such as plants, insects, birds, fish, and rodents have identified the molecular basis for the loss (and much more rarely the gain) of some complex traits (Hoekstra et al., 2006; Lang et al., 2012; Smith et al., 2013; Wang et al., 2011), the molecular mechanisms that underlie the divergence of morphological characters are almost entirely unknown—particularly for the origin of evolutionarily novel phenotypes ("novelties"). Among the most significant barriers to developing mechanistic explanations for the origin of evolutionary novelties is the lack of transitional forms among extant lineages and experimental systems amenable to detailed functional studies in non-model organisms.

Mammals are an ideal system in which to explore the molecular mechanisms that underlie the evolution of novelties because numerous genomic and experimental resources have been developed for mammals and their slow-evolving genomes facilitate tracing the origins of the novel functional DNA sequences (Lowe and Haussler, 2012; Mikkelsen et al., 2007; Warren et al., 2008). In addition, extant mammals span several major transitions in the origins of major morphological, developmental, and physiological novelties, including mammary glands, the cochlea, placentation, and pregnancy. Monotremes such as the platypus and echidna, for example, are oviparous and lay thin, poorly mineralized eggs that hatch about 10 days after laying (Hill, 1936), but the embryo is retained in the uterus for 10–22 days, during which time it is nourished by maternal secretions delivered through a simple yolk-sac placenta (Hughes and Hall, 1998; Renfree and Shaw, 2001). Live birth (viviparity) evolved in the Therian mammals, the lineage that includes marsupial and Eutherian ("placental") mammals, with the loss of the mineralized eggshell and yolk, and an elaboration of the placenta. Therian mammals, however, have

dramatically different reproductive life histories. In marsupials, pregnancies are relatively short (mean 25 days), and in all but one lineage (macropods), gestation is completed within the span of a single estrous cycle (Renfree and Shaw, 2001). In contrast, Eutherian mammals have evolved prolonged pregnancies that can last up to 670 days (mean 131 days) and that interrupt the estrous cycle (Hughes and Hall, 1998; Renfree and Shaw, 2001).

An essential step in the establishment and maintenance of pregnancy in many Eutherian mammals is the differentiation (decidualization) of endometrial stromal fibroblasts (ESFs) into decidual stromal cells (DSCs) in response to progesterone, the second messenger cyclic AMP (cAMP), and in some species to fetal signals (Gellersen and Brosens, 2003; Gellersen et al., 2007). Decidualization induces large-scale gene regulatory, cellular, and physiological reprogramming in the endometrium, leading to dramatic gene expression changes, the influx of immunosuppressive immune cells, vascular remodeling, and secretory transformation of uterine glands (Aghajanova et al., 2011; Gellersen et al., 2007; Giudice, 2003). Decidualization evolved in the stem lineage of Eutherian mammals (Kin et al., 2014; Mess and Carter, 2006) and underlies the suite of traits that support prolonged pregnancy in Eutherians, including direct implantation of the blastocyst and trophoblast into maternal endometrium, pronounced maternal recognition of pregnancy, maternal-fetal communication, and maternal immunotolerance of the antigenically distinct fetus.

Many of the genes that underlie the origins of maternal provisioning, viviparity, decidualization, and the collection of innovations that characterize prolonged pregnancy in Eutherian mammals likely evolved to be expressed at the fetomaternal interface coincident with the origins of pregnancy. Thus, identifying these genes and determining the mechanisms by which they became expressed in the endometrium will reveal some of the crucial molecular mechanisms underlying the evolution of pregnancy. Here we use ancestral transcriptome reconstruction and functional genomics methods to delineate the evolutionary history of gene expression in the endometrium and trace the molecular origin of the regulatory landscape that orchestrates decidualization. We show that thousands of genes evolved endometrial expression coincident with the evolution of pregnancy, including the recruitment of genes that play essential roles in the establishment of fetal immunotolerance and maternal-fetal communication. Furthermore, we found that 194 distinct families of ancient mammalian transposable elements (TEs) are enriched within *cis*-regulatory elements active in human DSCs and that many of these TEs donated binding sites for transcription factors (TFs) that establish cell-type identity and progesterone responsiveness to DSCs. These data indicate that a defining evolutionary novelty in mammals evolved through the recruitment and loss of ancient genes into an evolutionarily novel tissue, primarily via TE-mediated origination of a new *cis*-regulatory landscape in endometrial stromal cells.

RESULTS

Endometrial Gene Expression Profiling

We used high-throughput Illumina sequencing (RNA-seq) to identify transcribed genes in the endometrium during pregnancy in five Eutherian mammals (dog, cow, horse, pig, and armadillo), a marsupial (short-tailed opossum), and a Monotreme (platypus) and combined these data with existing gene expression data from the decidualized Rhesus monkey endometrium (Liu et al., 2012), decidualized mouse endometrium (McConaha et al., 2011), pregnant lizard uterus (Brandley et al., 2012), chicken uterus (Chan et al., 2010), and frog uterus (Chan et al., 2009). We also used RNA-seq to identify transcribed genes in human decidualized endometrial stromal cells (DSCs) in culture and combined these data with existing gene expression data from human decidual natural killer (dNK) cells (Hanna et al., 2006), decidual macrophage cells (dMP) (Svensson et al., 2011), and decidual endothelial cells (dECs). Our complete dataset includes expression information for 19,641 protein-coding genes from 14 species, as well as all the major cell types found in the human endometrium during pregnancy; this sampling allows us to infer the lineage in which each gene evolved endometrial expression as well as the specific cell type(s) in which each gene is expressed.

Thousands of Genes Were Recruited into and Lost from Endometrial Expression in Early Mammals

We used a model-based method to transform continuous transcript abundance estimates from RNA-seq datasets into discrete character states (transcribed or not) and parsimony to reconstruct ancestral endometrial transcriptomes (Li et al., 2010; Wagner et al., 2012, 2013). We found that parsimony inferred 500 genes gained and 296 lost endometrial expression in the Mammalian stem-lineage, 1,167 genes gained and 239 lost endometrial expression in the Therian stem lineage, and 835 genes gained and 185 lost endometrial expression in the Eutherian stem lineage (Figure 1A; Table S1). We could assign the expression of 89% (2,218 of 2,480) of recruited genes to one or more of the major cell types found in the endometrium during pregnancy (Figures 1B and S1). Significantly more recruited genes were uniquely expressed in DSC than dNK (485 versus 96; $p = 2.90 \times 10^{-189}$, binomial test), dMP (485 versus 55; $p = 4.36 \times 10^{-269}$, binomial test), or dEC (485 versus 62; $p = 8.13 \times 10^{-292}$, binomial test), suggesting recruitment of genes into DSC expression played a particularly important role in the evolution of pregnancy.

Recruited Genes Are Enriched in Immune, Signaling, and Reproductive Functions

We annotated each gene that evolved endometrial expression in the Mammalian, Therian, and Eutherian stem lineage by its expression domains in mouse to infer from which organ and tissue systems the expression of these genes were enriched and therefore likely recruited. We found that the expression of recruited genes was enriched in 33 distinct anatomical systems (false discovery rate [FDR] = 1%; Table S2) and that most genes were likely recruited from neural systems such as the brain (1083 genes), from digestive systems such as the gut (717 genes), and from the hemolymphoid system (661 genes; Figure 2A).

We also annotated genes that were recruited into endometrial expression by their Gene Ontology (GO) terms and mouse knockout phenotypes to infer the functional consequences of their recruitment. We found that Mammalian, Therian, and Eutherian recruited genes were enriched (FDR = 1%) in 87,

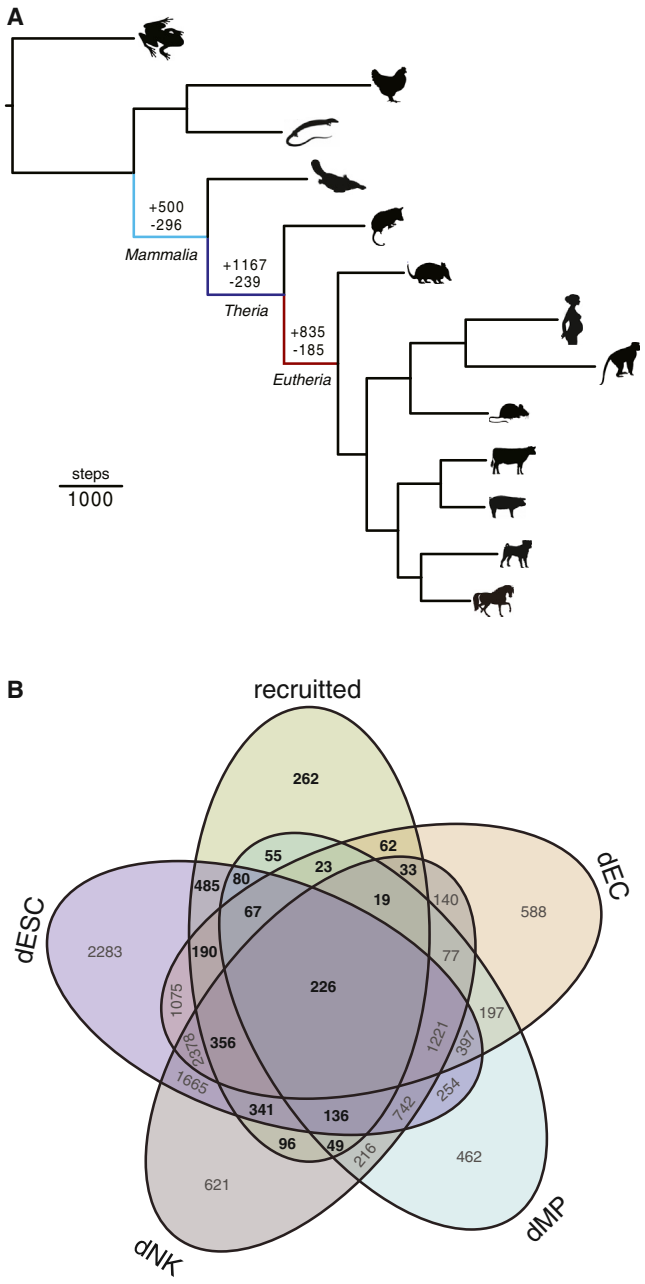


Figure 1. Evolution of the Endometrial Transcriptome in Tetrapods

(A) Parsimony reconstruction of gene expression gain and loss events in the tetrapod endometrium. The numbers above branches indicate the number of genes recruited into (+) or lost (–) from endometrial expression in the stem lineages of Mammalia (light blue), Theria (blue), and Eutheria (red). Branch lengths are proportional to gene expression recruitment and loss events inferred by Wagner parsimony (see inset scale bar).

(B) Expression of recruited genes in the major endometrial cell types. Recruited, total number of genes recruited into endometrial expression in the stem lineages of Mammalia, Theria, and Eutheria (n = 2,502).

480, and 288 knockout phenotypes (Figures 2B and 2C; Table S3) and in 71, 93, and 28 GO terms (Figures 2D and 2E; Table S4), respectively, most of which are related to metabolic processes, immune responses, cell-cell signaling, cell-cell communication, and uterine defects. Among the most significantly enriched mouse knockout phenotypes, for example, are "absent uterine NK cells" (55.50-fold, hypergeometric $p = 2.33 \times 10^{-7}$, FDR $q = 1.23 \times 10^{-5}$), "abnormal professional antigen presenting cell physiology" (2.84-fold, hypergeometric $p = 6.11 \times 10^{-16}$, FDR $q = 1.94 \times 10^{-13}$), "abnormal female reproductive system morphology" (1.98-fold, hypergeometric $p = 9.96 \times 10^{-5}$, FDR $q = 1.79 \times 10^{-3}$), and "decreased interleukin-1 secretion" (101-fold, hypergeometric $p = 1.46 \times 10^{-4}$, FDR $q = 6.83 \times 10^{-3}$). The most enriched GO terms included "positive regulation of leukocyte activation" (2.85-fold, hypergeometric $p = 8.30 \times 10^{-7}$, FDR $q = 1.91 \times 10^{-3}$), "GTPase activator activity" (2.58-fold, hypergeometric $p = 1.79 \times 10^{-3}$, FDR $q = 4.92 \times 10^{-2}$), and "response to cytokine" (1.90-fold, hypergeometric $p = 3.32 \times 10^{-4}$, FDR $q = 4.31 \times 10^{-2}$). These data indicate that many of the pathways essential for implantation, such as regulation of GTPase activity (Grewal et al., 2010), and the establishment of maternal-fetal immunotolerance, such as the recruitment of uterine NK into the endometrium during pregnancy (Hanna et al., 2006), and maternal-fetal signaling through cytokines evolved in the stem-lineage of Therian and Eutherian mammals.

Pervasive Loss of Ion Transporters from Endometrial Expression in Early Mammals

Next we annotated genes that lost endometrial expression by their mouse knockout phenotypes and GO terms to infer the functional consequences of their loss. Genes that lost endometrial expression in the Mammalian, Therian, and Eutherian stem lineages were significantly enriched (FDR = 1%) in 133, 11, and 15 knockout phenotypes (Figures S2A and S2B) and in 115, 16, and 46 GO terms (Figures S2C and S2D), respectively, most of which were related to ion transport and behavioral defects associated with impaired ion transport in the nervous system. Unexpectedly, the most significantly enriched mouse knockout phenotypes are related to abnormal behavior, for example, "abnormal anxiety-related response" (5.04-fold, hypergeometric $p = 7.53 \times 10^{-9}$, FDR $q = 2.13 \times 10^{-5}$), and "abnormal chemical nociception" (13.19-fold, hypergeometric $p = 1.02 \times 10^{-6}$, FDR $q = 2.55 \times 10^{-4}$), whereas the most enriched GO terms included "transporter activity" (2.48-fold, hypergeometric $p = 4.00 \times 10^{-8}$, FDR $q = 3.13 \times 10^{-5}$), "positive regulation of secretion" (3.73-fold, hypergeometric $p = 9.58 \times 10^{-5}$, FDR $q = 0.037$), and "transporter activity" (2.50-fold, hypergeometric $p = 8.46 \times 10^{-6}$, FDR $q = 0.004$). These data suggest that the enrichment of abnormal behavioral phenotypes in the mouse knockout data is related to the loss of ion channels with pleiotropic functions in behavioral regulation in the nervous and reproductive systems.

Consistent with the mouse knockout phenotype data and GO analyses, we found that numerous ion transporters and ligand-gated and voltage-gated ion channels lost endometrial expression during the evolution of mammalian pregnancy (Figure S2E). Among the ion transporters and ion channels that lost endometrial expression are several that have previously been shown to be important for mineralization of avian eggshells, including *ATP2B2*, *SLC12A5*, *SLC12A8*, *SLC26A9*, and *TRPV5* (Brionne et al., 2014; Jonchère et al., 2010, 2012). These data suggest

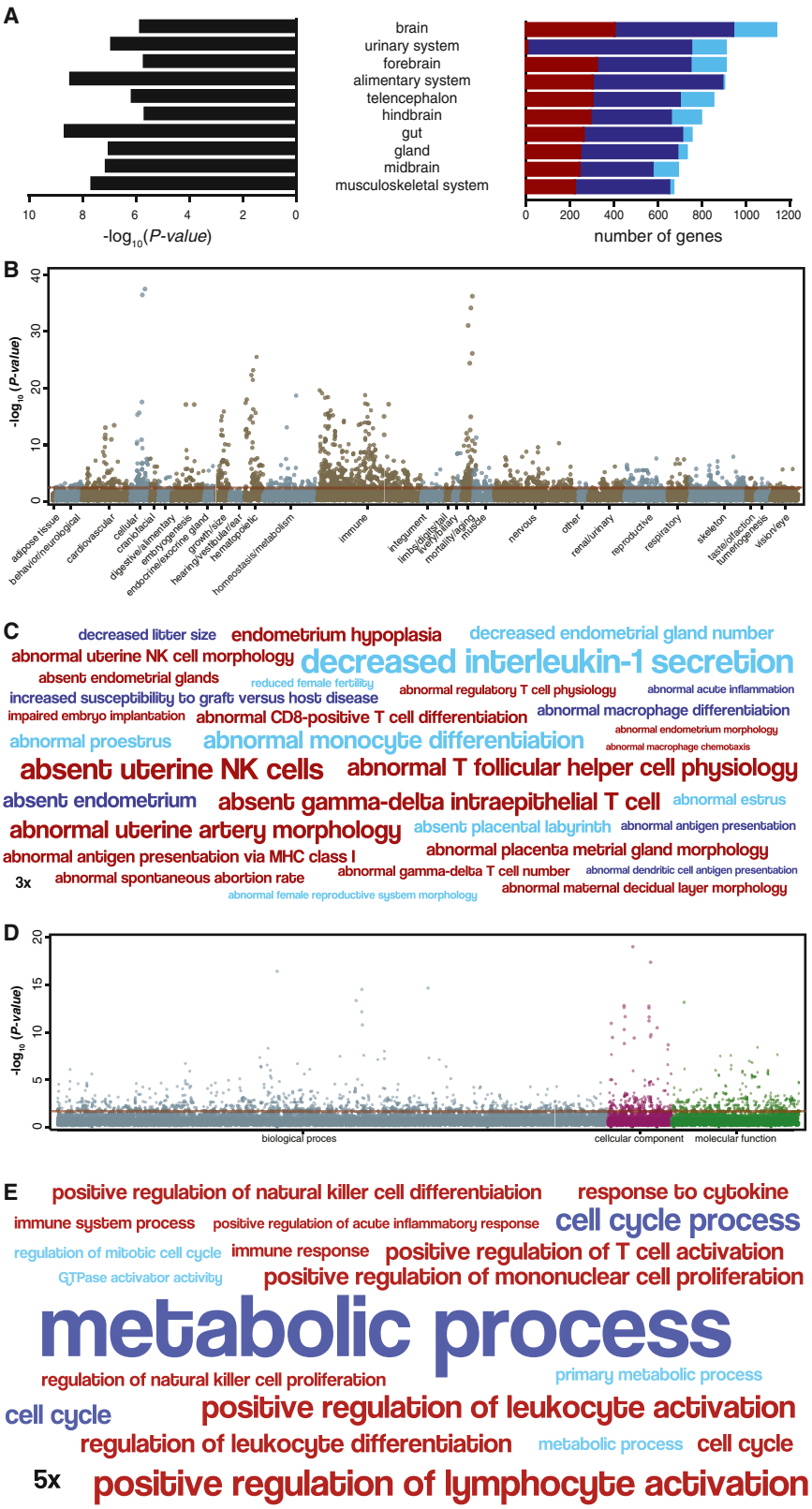


Figure 2. Recruited Genes Are Enriched in Immune, Signaling, and Reproductive Functions

(A) Top 10 anatomical systems in which the expression of recruited genes is enriched. Anatomical system, center; -$\log_{10}$ p value of enrichment (hypergeometric), left. Stacked bar chart shows the number of genes recruited into endometrial expression in the Mammalian (light blue), Therian (blue), and Eutherian (red) stem lineage that are expressed in each anatomical system.

(B) Manhattan plot of -$\log_{10}$ p values (hypergeometric test) for mouse knockout phenotypes enriched among recruited genes. Phenotypes are grouped by anatomical system. Horizontal red line indicates the dataset-wide false discovery rate (FDR q value = 0.1).

(C) Word cloud of mouse knockout phenotypes enriched among genes recruited into endometrial expression in the Mammalian (light blue), Therian (blue), and Eutherian (red) stem-lineage. Abnormal phenotypes are scaled to the $\log_2$ enrichment of that term (see inset scale).

(D) Manhattan plot of -$\log_{10}$ p values (hypergeometric test) for GO terms enriched among recruited genes. Go terms are grouped into biological process, cellular component, and molecular function. Horizontal red line indicates the dataset-wide false discovery rate (FDR q value = 0.1).

(E) Word cloud of GO terms enriched among genes recruited into endometrial expression in the Mammalian (light blue), Therian (blue), and Eutherian (red) stem lineage. GO terms are scaled to the $\log_2$ enrichment of that term (see inset scale).

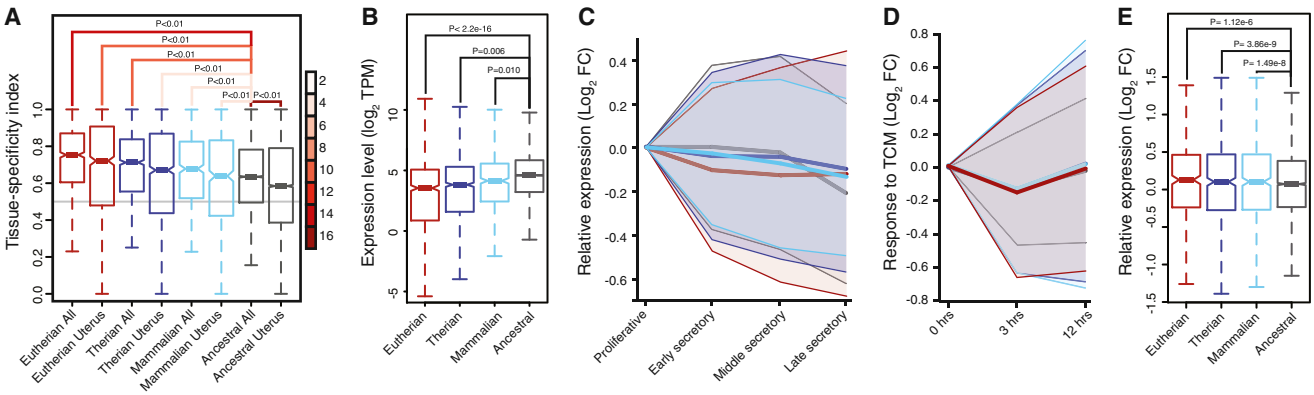


Figure 3. Expression Dynamics of Recruited Genes

(A) Tissue specificity of recruited and ancestrally expressed genes across 27 tissues (all) and how specific the expression those genes are in the uterus. 0, not tissue specific; 1, tissue specific; p values shown for paired t tests, lines connecting comparisons are colored by the t value of the comparison (see scale at right).

(B) Expression level (log_2 transcripts per million [TPM]) of recruited and ancestrally expressed genes; p values shown for paired t tests.

(C) Relative expression (log_2 fold change [FC] of secretory samples relative to proliferative samples) of recruited and ancestrally expressed genes throughout the human menstrual cycle. Mean ± variance.

(D) Response of recruited and ancestrally expressed genes expressed in human endometrial stromal cells to trophoblast conditioned media (TCM); log_2 fold change (FC) in gene expression 3 and 12 hr after treatment relative to control conditioned media (0 hr). Mean ± variance.

(E) Relative expression (log_2 fold change [FC]) of recruited and ancestrally expressed genes in the endometria of women with unexplained infertility compared to fertile controls; p values shown for F test for variance.

the loss of ion channels from endometrial expression in early mammals is associated with the reduced mineralization of eggshells in the mammalian stem lineage (Hill, 1936) and the complete loss of the eggshell in the Therian stem-lineage (Renfree and Shaw, 2001).

Recruited Genes Are More Dynamically Expressed than Ancestrally Expressed Genes

Next, we asked whether recruited genes had different expression dynamics compared with ancestrally expressed genes using several existing endometrial gene expression datasets (Talbi et al., 2006; Hess et al., 2007; Altmäe et al., 2010). We found a progressive increase in the tissue specificity of recruited genes, such that more recently recruited genes were significantly more tissue specific as well as more specifically expressed in the uterus than more anciently recruited and ancestrally expressed genes (Figure 3A). Transcripts of more recently recruited genes were also expressed at significantly lower levels, but had a greater range of expression, than more anciently recruited and ancestrally expressed genes (Figure 3B). To determine whether ancestral and recruited genes are differentially regulated during the reproductive cycle, we compared their expression levels in proliferative, early secretory, middle secretory, and late secretory phase human endometria. We found that Therian and Eutherian recruited genes had significantly greater variance in expression levels compared with ancestrally expressed genes throughout the menstrual cycle but that Mammalian recruited genes had similar expression dynamics as ancestrally expressed genes (Figure 3C). These data indicate that Therian and Eutherian recruited genes are more strongly differentially regulated during the menstrual cycle than mammalian recruited and ancestrally expressed genes. Similarly, recruited genes had significantly greater variance in expression levels in decidualized human endometrial cells 3 and 12 hr after treatment with trophoblast-conditioned media than ancestrally expressed genes (Figure 3D), indicating that they are more responsive to fetal signals than ancestrally expressed genes. Finally, recruited genes were more misregulated than ancestrally expressed genes during the window of implantation in the endometria of women with unexplained infertility compared with fertile controls (Figure 3E), consistent with an important role for recruited genes in the establishment of pregnancy.

Endometrial Regulatory Elements Are Enriched in Ancient Mammalian TEs

Previous studies have shown that TEs can act as hormone responsive regulatory elements in DSCs (Gerlo et al., 2006; Lynch et al., 2011; Emera and Wagner, 2012), suggesting that Mammalian-, Therian-, and Eutherian-specific TEs (hereafter "ancient mammalian" TEs) may have played a role in the recruitment of genes in endometrial expression in early mammals. To identify regulatory elements that may be derived from ancient mammalian TEs (AncMam-TEs) we mapped enhancers (H3K27ac chromatin immunoprecipitation sequencing [ChIP-seq]), promoters (H3K4me3 ChIP-seq), and regions of open chromatin (FAIRE-seq and DNaseI-seq) in cAMP/progesterone-treated human DSCs. We then intersected the location of these regulatory elements with the location of AncMam-TEs across the human genome (hg19).

We found that 59.9% of DNaseI-seq peaks, 30.0% of FAIRE-seq peaks, 57.7% of H3K27ac peaks, and 31.5% of H3K4me3 peaks overlapped AncMam-TE (Figure 4A), most of which were Mammalian- or Eutherian-specific (Figure 4B; Table S5). We also identified 194 AncMam-TEs families that were significantly enriched within *cis*-regulatory elements compared to their genomic abundances (Figure 4C; Table S6), 89.2% (173 of 194)

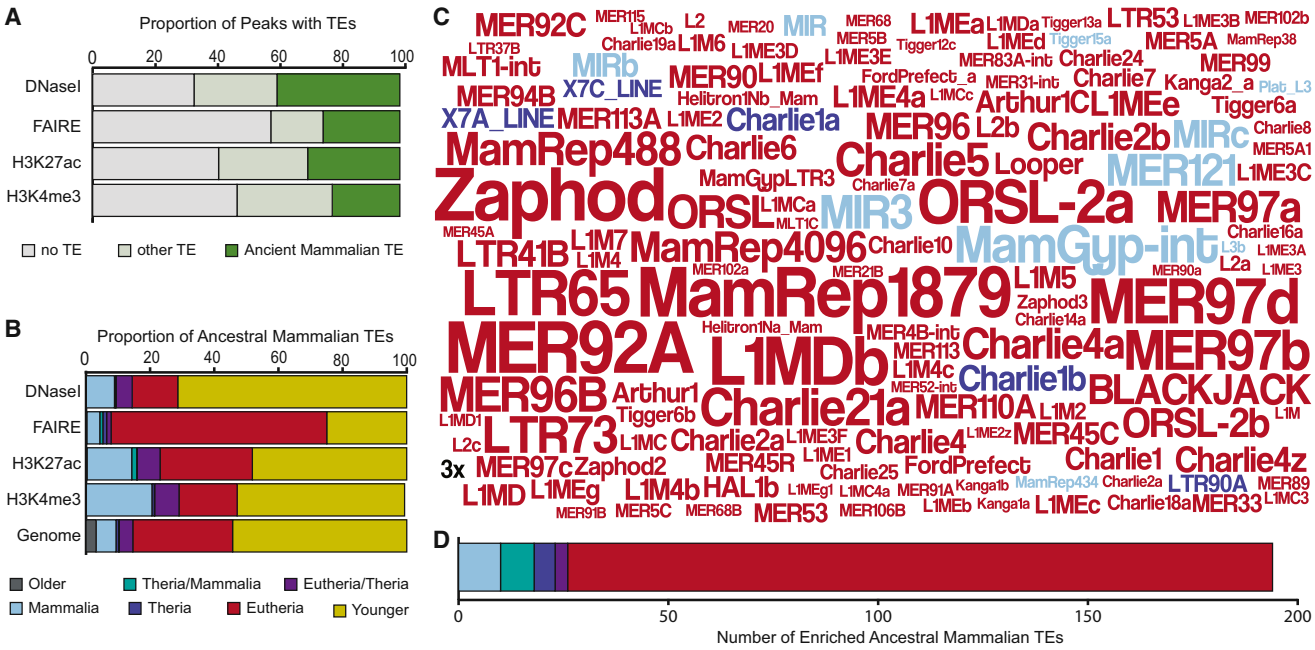


Figure 4. Ancient Mammalian TEs Are Enriched in Regulatory Elements Active in Decidualized Human Endometrial Stromal Cells

(A) Proportion of DNaseI-seq, FAIRE-seq, H3K27ac-seq, and H3K4me3-seq peaks that do not overlap annotated TEs (gray), overlap "ancient mammalian" TEs, i.e., those that integrated into the genome in the stem lineage of Mammalia, Thera, or Eutheria (green), and other TEs (light green).

(B) Proportion of TEs in DNaseI-, FAIRE-, H3K27ac-, and H3K4me3-seq peaks and the human genome (hg18) by age class.

(C) Word cloud of the 150 most enriched TEs among ancient mammalian elements, in DNaseI-, FAIRE-, H3K27ac-, and H3K4me3-seq peaks. The size of the transposon names corresponds to its enrichment (see inset 3-fold scale). Colors indicate lineage in which the TE originated: Mammalian-specific (light blue), Therian or Mammalian (green), Therian specific (blue), Eutherian or Therian (purple), Eutherian specific (red).

(D) Number of enriched TEs among ancient mammalian elements in DNaseI-, FAIRE-, H3K27ac-, and H3K4me3-seq peaks by age class.

of which are Eutherian specific (Figure 4D). Among the enriched AncMam-TEs are 930 copies of MER20 (1.98-fold, $p = 2.79 \times 10^{-8}$) a Eutherian-specific hAT-Charlie DNA transposon we have previously shown to function frequently as cAMP/progesterone-responsive *cis*-regulatory element in hDSCs (Lynch et al., 2011) and 377 individual "exapted" TEs previously shown to have evolved under strong purifying selection near developmental genes (Lowe et al., 2007). Among the exapted repeats are 25 copies of MER121 (5.42-fold, $p = 0$), one of the most constrained TE families in mammalian genomes (Kamal et al., 2006). These data indicate that AncMam-TEs, particularly Eutherian-specific TEs, played an important role in the genesis of *cis*-regulatory elements in DSCs coincident with the origins of decidualization.

Ancient Mammalian TEs Are Associated with Progesterone-Responsive Genes

Our observation that some AncMam-TE families are enriched within active regulatory elements in DSCs suggests they may play a role in orchestrating the transcriptional response of ESFs cells to cAMP and progesterone. To test this hypothesis, we used RNA-seq (Figure 5A) to compare the expression levels of genes associated with AncMam-TE containing *cis*-regulatory elements and genes without such association between undifferentiated ESFs and decidualized DSCs (Figure 5B). We found that genes associated with AncMam-TE containing regulatory elements were no more strongly differentially regulated upon cAMP/progesterone-induced decidualization than genes without such associations when we compared all genes expressed in DSCs (Figure 5C). However, we found that unambiguously recruited genes associated with AncMam-TE containing regulatory elements were more strongly differentially regulated upon decidualization than recruited genes without AncMam-TE containing regulatory elements (Figure 5C). Recruited genes associated with AncMam-TE containing regulatory elements were also more strongly differentially regulated upon decidualization than ancestrally expressed genes associated with AncMam-TE containing regulatory elements. These results suggest that AncMam-TEs play a role in orchestrating the transcriptional response of DSCs to cAMP/progesterone, particularly for those genes that evolved endometrial expression during the origins of pregnancy and decidualization.

Ancient Mammalian TEs Mediate Decidualization and Endometrial Cell-type Identity

Next, we generated ChIP-seq data for the progesterone receptor (PGR), the principle transcriptional effector of progesterone signaling and decidualization, from cAMP/Progesterone treated DSCs and integrated this dataset with previously published ChIP-seq data for 132 TFs to identify TF binding sites (TFBSs) enriched in AncMam-TEs compared with non-TE derived regions of the genome (Eijkelenboom et al., 2013; Lo and Matthews, 2012; Tewari et al., 2012; Wang et al., 2012). We identified

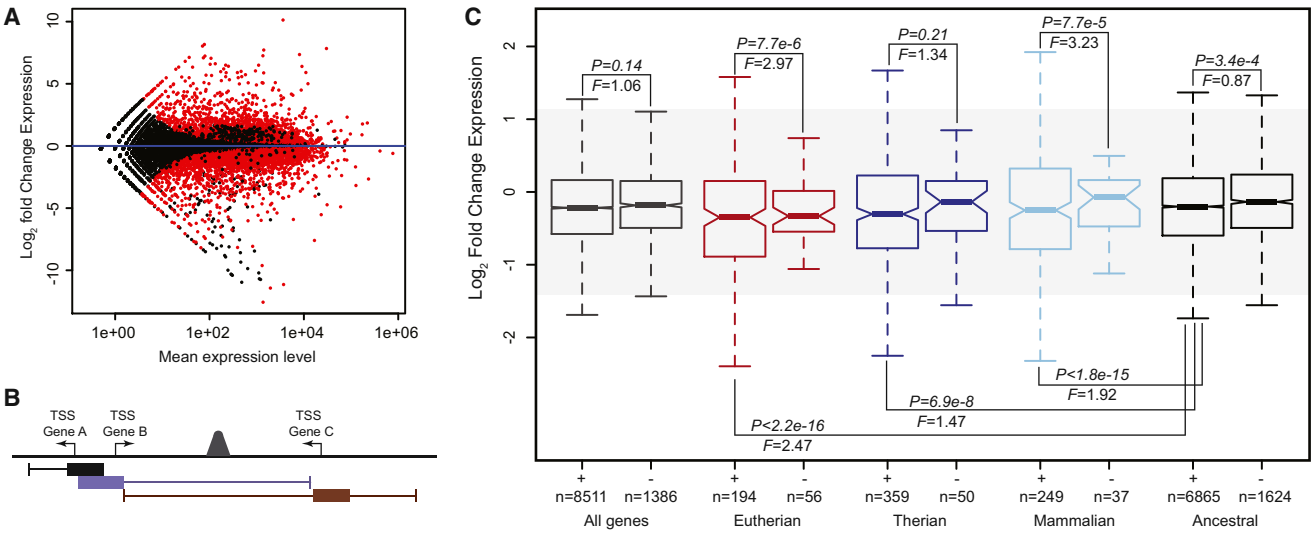


Figure 5. Genes Associated with Ancient Mammalian TE-Derived Regulatory Elements Are More Strongly Differentially Regulated during Decidualization than Genes without TE-Derived Regulatory Elements

(A) MA-plot showing changes in gene expression 48hrs after treatment of human endometrial stromal cells with cAMP/Progesterone. Red dots indicate significantly differentially expressed genes (n = 7,337, N = 16,551).

(B) Cartoon of the GREAT association rule used to associate regulatory elements with nearby genes: regulatory element peaks located +5,000 to −1,000 bp of transcription start sites (TSSs) were defined as the basal regulatory domain (filled boxes represented below each TSS) for each gene (regardless of other nearby genes). Basal regulatory domains were extended in both directions to the nearest gene's basal domain (thin lines) but no more than 100 kb. Under this association rule, for example, the regulatory element peak located between Gene B and Gene C would be associated with these genes but not Gene A.

(C) Recruited genes associated with ancient mammalian TE-derived regulatory elements (+) are more strongly differentially regulated upon cAMP/progesterone treatment (decidualization) than genes without TE-derived regulatory elements (−). Genes are grouped into all genes expressed in DSCs (gray), Eutherian recruited genes (red), Therian recruited genes (blue), Mammalian recruited genes (light blue), and ancestrally expressed genes (black). F is the ratio of variances from a two-sample F test, and p is the significance of the two-sample F test.

50 TFBSs that were significantly enriched (FDR = 5%) in the AncMam-TE-derived portions of regulatory elements (Figure 6A; Table S7). Among the TFs with binding sites enriched in AncMam-TEs are those that mediate responses to hormones such as PGR (10-fold enrichment, $p < 1.0 \times 10^{-5}$) (Conneely et al., 2001; Lydon et al., 1995), NR4A1 (2.6-fold, $p = 6.6 \times 10^{-4}$) (Jiang et al., 2011), and ERRA (3-fold, $p = 1.0 \times 10^{-10}$) (Bombail et al., 2010), TFs essential for the successful establishment of pregnancy such as c-MYC (1.3-fold, $p = 5.6 \times 10^{-21}$) (Davis et al., 1993), AHR (1.5-fold, $p = 4.0 \times 10^{-4}$) (Hernández-Ochoa et al., 2009), and STAT1 (1.2-fold, $p = 7.5 \times 10^{-9}$) (Christian et al., 2001), as well as general TFs and numerous chromatin remodeling factors (Figure 6B). De novo motif discovery also identified many enriched motifs in the AncMam-TE-derived regions of regulatory element peaks compared with TE-free regulatory elements, many of which match consensus binding sites for TFs that function as master regulators of DSC cell-type identity including FOX, COUP-TF, HOX, and AHR/ARNT (Figure 6C). These data indicate that TEs contributed binding sites for TFs that directly mediate the response of stromal cells to progesterone (PGR) and that establish cell-type identity in endometrial cells (HOX and FOX) during the origins of decidualization.

Ancient Mammalian TEs Donated Functional PGR Binding Sites to the Genome

Our observations that AncMam-TEs are enriched in PGR binding sites and are associated with genes that are strongly differentially expressed in DSCs upon decidualization prompted us to explore AncMam-TE derived PGR binding sites in greater detail. We found that nearly all (1,721 of 1,777) of the AncMam-TE-derived PGR binding sites were found within Mammalian- or Eutherian-specific TEs (Figure 7A; Table S8) and that 53.7% of the AncMam-TEs that are enriched in PGR binding sites were Eutherian-specific (Figure 7B; Table S8). We also found that all genes associated with AncMam-TE-derived PGR binding sites were more strongly upregulated and more strongly differentially regulated upon decidualization than genes not associated with AncMam-TE-derived PGR binding sites (Figure 7C). Similarly, recruited and ancestrally expressed genes associated with AncMam-TE-derived PGR binding sites were more strongly upregulated and more strongly differentially regulated than recruited and ancestral genes not associated with AncMam-TE-derived PGR binding sites or than recruited and ancestral genes without TE-derived regulatory elements (Figure 7C).

These data suggest that AncMam-TE derived PGR binding sites orchestrate most of the transcriptional response of DSCs to progesterone. To test further this hypothesis, we used previously published data to compare the effects of PGR knockdown in cAMP/progesterone-treated DSCs (Pabona et al., 2012) on the expression of genes associated with AncMam-TE derived regulatory elements and PGR binding sites to genes without such associations. We found that ancestrally expressed and recruited genes associated with AncMam-TE-derived PGR binding sites were more strongly misregulated by PGR knockdown than genes

Figure 6. Ancient Mammalian TEs Are Enriched in Binding Sites for TFs that Mediate Hormone Responsiveness and Endometrial Cell Identity

(A) Distribution of enriched (red) and depleted (blue) TFBSs in ancient TE-derived segments of FAIRE-, DNaseI-, H3K27ac-, and H3K4me3-seq peaks compared with TF ChIP-seq peaks with no TE overlap. Fifty and 75 TFBSs were significantly enriched and depleted, respectively, at FDR = 5%.

(B) Word cloud of TFBSs enriched in ancient mammalian TE segments of ChIP-seq peaks compared to peaks with no TE overlap. Colors indicate TFs that mediate hormone responses (purple), remodel chromatin (light purple), have known functions in endometrial cells or that mediate immune responses (green), or with general regulatory functions (light green). Data shown for $\geq$1.2-fold enriched TFs at FDR = 5%.

(C) Word cloud of TF motifs enriched in ancient mammalian TE segments of ChIP-seq peaks compared with peaks with no TE overlap. Colors indicate motifs enriched in FAIRE- (red), DNaseI- (tan), H3K27ac- (dark blue), and H3K4me3-seq peaks (light blue). Data shown for $\geq$3-fold enriched motifs at FDR = 1×10^{-4} and E1 $\times 10^{-4}$.

without AncMam-TE-derived PGR binding sites (Figure 7D). These data are consistent with an important role for TE-derived PGR binding sites in the transcriptional response of ESFs to cAMP/progesterone and therefore decidualization into DSCs.

DISCUSSION

A major challenge in developmental evolution is determining the genetic changes that underlie morphological differences between species, particularly the changes that have generated evolutionary novelties such as mammalian pregnancy (Wagner and Lynch, 2010). While it is clear that the evolution of gene regulatory networks is an essential part of developmental evolution, the mechanisms that underlie regulatory network divergence are debated (Carroll, 2005; 2008; Lynch and Wagner, 2008; Prud'homme et al., 2007; Wagner and Lynch, 2010). Comparative studies in a few well-studied organisms suggest that gain and loss of *cis*-regulatory elements provides the molecular basis for the divergence of morphological traits between species (Hoekstra et al., 2006; Lang et al., 2012; Smith et al., 2013; Wang et al., 2011), but whether the gradual gene-by-gene gain and loss of individual regulatory elements is sufficient to explain the origin of new gene regulatory networks and therefore the origin of evolutionary novelties is not clear.

TE-mediated rewiring of gene regulatory networks is an attractive alternative to the gene-by-gene origination of *cis*-regulatory elements because it provides a mechanism to rapidly distribute nearly identical copies of regulatory elements across the genome that are capable of responding to the same input signals (Britten and Davidson, 1969; Davidson and Britten, 1979; McClintock, 1984). Thus far, however, there is little evidence that TEs contribute to the genesis of novel regulatory networks rather than contributing to the turnover of regulatory elements within an existing gene regulatory network. Our results indicate that a major mechanism in the origin of pregnancy was the recruitment of genes that were ancestrally expressed in other organ and tissue systems into endometrial expression, which imparted the endometrium with new functions such as immunoregulation and maternal-fetal signaling. Our data also suggest that TEs that amplified prior to the divergence of Eutherian

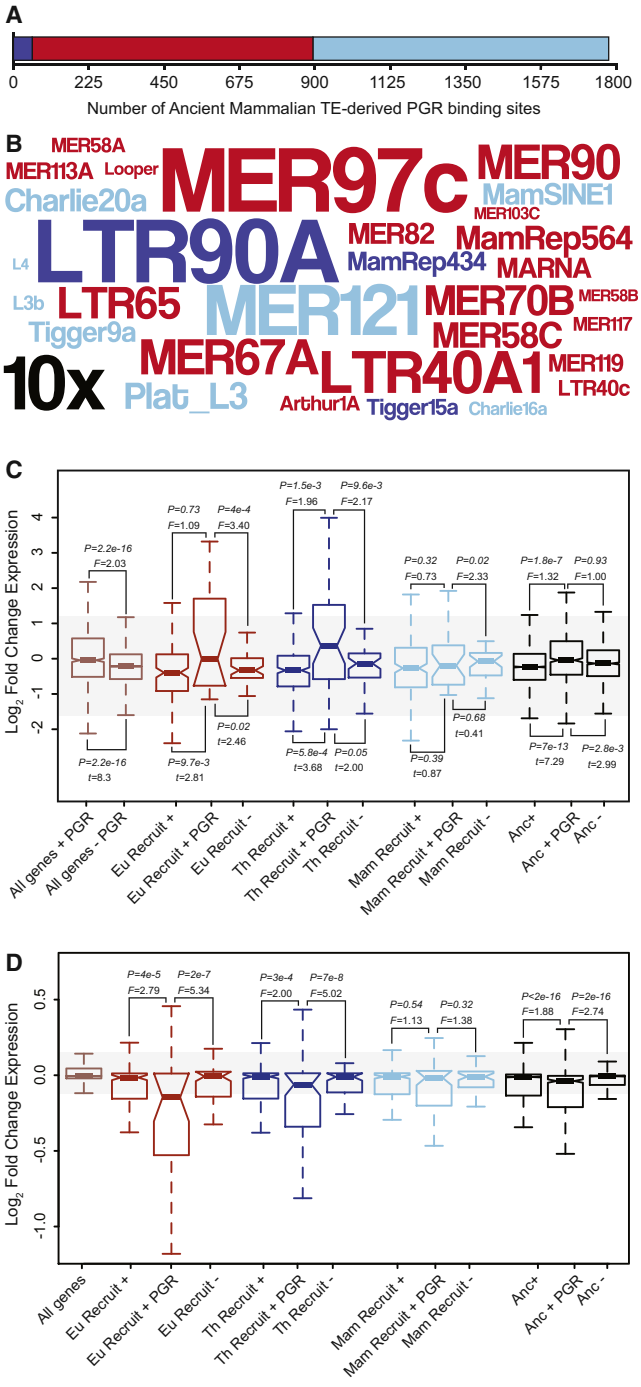


Figure 7. Ancient Mammalian TEs Globally Remodeled PGR Binding Site Architecture across the Genome

(A) The number of PGR ChIP-seq peaks that contain Mammalian-specific (light blue), Therian-specific (blue), and Eutherian-specific (red) TEs in primary cultures of hESCs treated with 10 nM 17β-estradiol, 100 nM medroxyprogesterone acetate, and 1 mM 8-bromo-cAMP.

(B) Word cloud of ancient mammalian TEs enriched within PGR ChIP-seq peaks. Colors follow (A). Inset scale (10×) shows 10-fold enrichment. Only elements with ≥2-fold enrichment are shown.

(C) Genes associated with ancient mammalian TE-derived PGR binding sites (+ PGR) are more strongly differentially regulated in hESCs upon decidualization than genes associated with ancient mammalian TE as a whole (+) and genes without TE-derived regulatory elements (−). Eu Recruit, Eutherian recruited gene; Th Recruit, Therian recruited gene; Mam Recruit, Mammalian recruited gene; Anc, ancestrally expressed gene. F is the ratio of variances from a two-sample F test.

(D) Genes associated with ancient mammalian TE-derived PGR biding sites (+ PGR) are more strongly dysregulated by PGR knockdown in DSCs than either genes associated with ancient mammalian TE as a whole (+) and genes without TE-derived regulatory elements (−). F is the ratio of variances from a two-sample F test.

mammals played a central role in recruiting these genes into endometrial expression and thereby in the origin of decidualization because they apparently deposited binding sites for master transcriptional regulators of endometrial stromal cell-type identity and progesterone responsiveness to numerous genes across the genome.

Taken together, our data suggest that novel gene regulatory networks and cell-type identities can evolve through large-scale genome-wide changes rather than gradual gene-by-gene changes (Goldschmidt, 1940). TEs may play a particularly important role in this process because they provide a mechanism to coordinately regulate the expression of numerous genes to the same stimuli upon their integration into multiple locations in the genome, alleviating the need for the de novo evolution of *cis*-regulatory elements capable of directing stereotyped responses to the same stimuli one gene at a time across the genome (Britten and Davidson, 1969; Davidson and Britten, 1979; Feschotte, 2008; McClintock, 1984).

EXPERIMENTAL PROCEDURES

Transcriptome Sequencing

Total RNA from endometrial samples from midstage pregnant platypus, opossum, armadillo, and dog, and day 14–18 postimplantation cow, horse, and pig were dissected was extracted using the QIAGEN RNA-Easy Midi RNA-extraction kit followed by on-column DNase treatment (QIAGEN). RNA samples were sequenced using the Illumina Genome Analyzer II platform (75 bp reads), following the protocol suggested by Illumina for sequencing of cDNA samples. mRNA-seq reads were quality controlled following standard methods and species specific reads mapped to the human (GRCh37), macaque (MMUL_1), mouse (NCBIM37), dog (BROADD2), cow (UMD3.1), horse (EquCab2), pig (Sscrofa9), armadillo (dasNov2), opossum (monDom5), platypus (OANA5), and chicken (WASHUC2) gene builds at Ensembl using Bowtie with default parameters.

Parsimony Reconstruction of Gene Expression Gain/Loss

We used parsimony to reconstruct gene expression gains and losses using the method implemented in the *pars* program from PhyML (v.2.4.4). We used Mesquite (v.2.75) to identify the number of genes that most parsimoniously gained or lost endometrial expression; expression was classified as most parsimoniously a gain if a gene was inferred as not expressed the ancestral node (state 0) but inferred likely expressed in a descendent node (state 1/[0/1]) and vice versa for the classification of a loss from endometrial expression.

ChIP-Seq, DNaseI-Seq, and FAIRE-Seq Data Generation

See the Supplemental Experimental Procedures.

Identification of TE-Containing Regulatory Elements

To identify regulatory elements derived from TEs, we first intersected regulatory element peaks from the analyses described above with human TE annotation of the hg18 assembly (files from the U.C.S.C genome browser, with Repeat Masker v.3.2.7 and repbase libraries released on 20050112) using

BEDTools. The ratio between counts (in-set divided by in-genome) is used to estimate whether a given TE is enriched or depleted. Significance of enrichment was inferred on TE counts with three standard statistical tests (binomial, hypergeometric, and Poisson models), but figures are built on length since it is more representative of a given TE contribution.

Region-Gene Associations, Motif, and TFBS Finding

We used the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) to associated regulatory elements with nearby genes using the default association rules. To identify ancestral and recruited genes that were expressed in human endometrial stromal cells (hESCs), we intersected ancestral and recruited genes sets with the set of genes expressed in hESCs. To identify sequence motifs that were enriched in "TE-derived" regulatory elements, we used a de novo motif discovery approach. First we used CisFinder webserver (http://lgsun.grc.nia.nih.gov/CisFinder/) to identify overrepresented short DNA motifs in TE-derived regulatory elements, the top 10 scoring motifs from CisFinder were annotated using the STAMP webserver (http://www.benoslab.pitt.edu/stamp/) with TRANSFAC (families labeled) motifs.

TE and Regulatory Element Correlations

See the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The mRNA-seq data reported in this article have been deposited in NCBI GEO under accession numbers GSE57714, GSE30708, GSE29553, GSE21046, GSE43667, and GSE48862. The ChIP-seq and DNaseI-seq data have been deposited under accession number GSE61793, and the FAIRE-seq data have been deposited under accession number GSM1011119.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, eight tables, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.12.052.

ACKNOWLEDGMENTS

This work was funded by a University of Chicago new lab startup package and a Burroughs Wellcome Preterm Birth Initiative grant (V.J.L.) and a grant from the John Templeton Foundation, number 12793 Genetics and the Origin of Organismal Complexity (G.P.W). The generation of mRNA-seq data from cow, pig, and horse was supported by a German Ministry for Education and Research grant (BMBF, FUGATO-plus, COMPENDIUM) to S.B. We thank R.W. Truman (National Hansen's Disease Program/U.S. National Institutes of Allergy and Infectious Diseases IAA-2646) and K. Smith for the generous gifts of pregnant armadillo and opossum uterus and R. Bjornson and N. Carriero for assistance with RNA-seq read mapping. This work was also supported by the NIH (grant number R01GM077582 to C.F). We also thank C. Ober for comments on an earlier version of this manuscript and A. Smit for assistance for TE age classifications.

Received: February 21, 2014
Revised: November 14, 2014
Accepted: December 22, 2014
Published: January 29, 2015

REFERENCES

Aghajanova, L., Tatsumi, K., Horcajadas, J.A., Zamah, A.M., Esteban, F.J., Herndon, C.N., Conti, M., and Giudice, L.C. (2011). Unique transcriptome, pathways, and networks in the human endometrial fibroblast response to progesterone in endometriosis. Biol. Reprod. *84*, 801–815.

Altmäe, S., Martínez-Conejero, J.A., Salumets, A., Simón, C., Horcajadas, J.A., and Stavreus-Evers, A. (2010). Endometrial gene expression analysis at the time of embryo implantation in women with unexplained infertility. Mol. Hum. Reprod. *16*, 178–187.

Bombail, V., Gibson, D.A., Collins, F., MacPherson, S., Critchley, H.O.D., and Saunders, P.T.K. (2010). A Role for the orphan nuclear receptor estrogen-related receptor alpha in endometrial stromal cell decidualization and expression of genes implicated in energy metabolism. J. Clin. Endocrinol. Metab. *95*, E224–E228.

Brandley, M.C., Young, R.L., Warren, D.L., Thompson, M.B., and Wagner, G.P. (2012). Uterine gene expression in the live-bearing lizard, Chalcides ocellatus, reveals convergence of squamate reptile and mammalian pregnancy mechanisms. Genome Biol. Evol. *4*, 394–411.

Brionne, A., Nys, Y., Hennequet-Antier, C., and Gautron, J. (2014). Hen uterine gene expression profiling during eggshell formation reveals putative proteins involved in the supply of minerals or in the shell mineralization process. BMC Genomics *15*, 220.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349–357.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. PLoS Biol. *3*, e245.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Chan, E.T., Quon, G.T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R.A., Aubin, J., Ratcliffe, M.J., Wilde, A., Brudno, M., et al. (2009). Conservation of core gene expression in vertebrate tissues. J. Biol. *8*, 33.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science *327*, 302–305.

Christian, M., Marangos, P., Mak, I., McVey, J., Barker, F., White, J., and Brosens, J.J. (2001). Interferon-gamma modulates prolactin and tissue factor expression in differentiating human endometrial stromal cells. Endocrinology *142*, 3142–3151.

Conneely, O.M., Mulac-Jericevic, B., Lydon, J.P., and De Mayo, F.J. (2001). Reproductive functions of the progesterone receptor isoforms: lessons from knock-out mice. Mol. Cell. Endocrinol. *179*, 97–103.

Davidson, E.H., and Britten, R.J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science *204*, 1052–1059.

Davis, A.C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes Dev. *7*, 671–682.

Eijkelenboom, A., Mokry, M., de Wit, E., Smits, L.M., Polderman, P.E., van Triest, M.H., van Boxtel, R., Schulze, A., de Laat, W., Cuppen, E., and Burgering, B.M. (2013). Genome-wide analysis of FOXO3 mediated transcription regulation through RNA polymerase II profiling. Mol. Syst. Biol. *9*, 638.

Emera, D., and Wagner, G.P. (2012). Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc. Natl. Acad. Sci. USA *109*, 11246–11251.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. *9*, 397–405.

Gellersen, B., and Brosens, J. (2003). Cyclic AMP and progesterone receptor cross-talk in human endometrium: a decidualizing affair. J. Endocrinol. *178*, 357–372.

Gellersen, B., Brosens, I.A., and Brosens, J.J. (2007). Decidualization of the human endometrium: mechanisms, functions, and clinical perspectives. Semin. Reprod. Med. *25*, 445–453.

Gerlo, S., Davis, J.R.E., Mager, D.L., and Kooijman, R. (2006). Prolactin in man: a tale of two promoters. BioEssays *28*, 1051–1055.

Giudice, L.C. (2003). Elucidating endometrial function in the post-genomic era. Hum. Reprod. Update *9*, 223–235.

Goldschmidt, R. (1940). The Material Basis of Evolution (New Haven, CT: Yale University Press).

Grewal, S., Carver, J., Ridley, A.J., and Mardon, H.J. (2010). Human endometrial stromal cell rho GTPases have opposing roles in regulating focal adhesion turnover and embryo invasion in vitro. Biol. Reprod. *83*, 75–82.

Hanna, J., Goldman-Wohl, D., Hamani, Y., Avraham, I., Greenfield, C., Natanson-Yaron, S., Prus, D., Cohen-Daniel, L., Arnon, T.I., Manaster, I., et al. (2006). Decidual NK cells regulate key developmental processes at the human fetal-maternal interface. Nat. Med. *12*, 1065–1074.

Hernández-Ochoa, I., Karman, B.N., and Flaws, J.A. (2009). The role of the aryl hydrocarbon receptor in the female reproductive system. Biochem. Pharmacol. *77*, 547–559.

Hess, A.P., Hamilton, A.E., Talbi, S., Dosiou, C., Nyegaard, M., Nayak, N., Genbecev-Krtolica, O., Mavrogianis, P., Ferrer, K., Kruessel, J., et al. (2007). Decidual stromal cell response to paracrine signals from the trophoblast: amplification of immune and angiogenic modulators. Biol. Reprod. *76*, 102–117.

Hill, J.P. (1936). V. The Development of the Monotremata.—Part I. The Histology of the Oviduct during Gestation. By Catherine J. Hill. B. Sc., Ph. D. Part II. The Structure of the Egg-shell. Trans. Zool. Soc. Lond. *21*, 413–476.

Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., and Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. Science *313*, 101–104.

Hughes, R.L., and Hall, L.S. (1998). Early development and embryology of the platypus. Philos. Trans. R. Soc. Lond. B Biol. Sci. *353*, 1101–1114.

Jiang, Y., Hu, Y., Zhao, J., Zhen, X., Yan, G., and Sun, H. (2011). The orphan nuclear receptor Nur77 regulates decidual prolactin expression in human endometrial stromal cells. Biochem. Biophys. Res. Commun. *404*, 628–633.

Jonchère, V., Réhault-Godbert, S., Hennequet-Antier, C., Cabau, C., Sibut, V., Cogburn, L.A., Nys, Y., and Gautron, J. (2010). Gene expression profiling to identify eggshell proteins involved in physical defense of the chicken egg. BMC Genomics *11*, 57.

Jonchère, V., Brionne, A., Gautron, J., and Nys, Y. (2012). Identification of uterine ion transporters for mineralisation precursors of the avian eggshell. BMC Physiol. *12*, 10.

Kamal, M., Xie, X., and Lander, E.S. (2006). A large family of ancient repeat elements in the human genome is under strong selection. Proc. Natl. Acad. Sci. USA *103*, 2740–2745.

Kin, K., Maziarz, J., and Wagner, G.P. (2014). Immunohistological study of the endometrial stromal fibroblasts in the opossum, Monodelphis domestica: evidence for homology with eutherian stromal fibroblasts. Biol. Reprod. *90*, 111.

Lang, M., Murat, S., Clark, A.G., Gouppil, G., Blais, C., Matzkin, L.M., Guittard, E., Yoshiyama-Yanagawa, T., Kataoka, H., Niwa, R., et al. (2012). Mutations in the neverland gene turned Drosophila pachea into an obligate specialist species. Science *337*, 1658–1661.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics *26*, 493–500.

Liu, J.-L., Liang, X.-H., Su, R.-W., Lei, W., Jia, B., Feng, X.-H., Li, Z.-X., and Yang, Z.-M. (2012). Combined analysis of microRNome and 3′-UTRome reveals a species-specific regulation of progesterone receptor expression in the endometrium of rhesus monkey. J. Biol. Chem. *287*, 13899–13910.

Lo, R., and Matthews, J. (2012). High-resolution genome-wide mapping of AHR and ARNT binding sites by ChIP-Seq. Toxicol. Sci. *130*, 349–361.

Lowe, C.B., and Haussler, D. (2012). 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. PLoS ONE *7*, e43128.

Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc. Natl. Acad. Sci. USA *104*, 8005–8010.

Lydon, J.P., DeMayo, F.J., Funk, C.R., Mani, S.K., Hughes, A.R., Montgomery, C.A., Jr., Shyamala, G., Conneely, O.M., and O'Malley, B.W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. Genes Dev. *9*, 2266–2278.

Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. Evolution *62*, 2131–2154.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. *43*, 1154–1159.

McClintock, B. (1984). The significance of responses of the genome to challenge. Science *226*, 792–801.

McConaha, M.E., Eckstrum, K., An, J., Steinle, J.J., and Bany, B.M. (2011). Microarray assessment of the influence of the conceptus on gene expression in the mouse uterus during decidualization. Reproduction *141*, 511–527.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mess, A., and Carter, A.M. (2006). Evolutionary transformations of fetal membrane characters in Eutheria with special reference to Afrotheria. J. Exp. Zoolog. B Mol. Dev. Evol. *306*, 140–163.

Mikkelsen, T.S., Wakefield, M.J., Aken, B., Amemiya, C.T., Chang, J.L., Duke, S., Garber, M., Gentles, A.J., Goodstadt, L., Heger, A., et al.; Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team (2007). Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature *447*, 167–177.

Pabona, J.M.P., Simmen, F.A., Nikiforov, M.A., Zhuang, D., Shankar, K., Velarde, M.C., Zelenko, Z., Giudice, L.C., and Simmen, R.C.M. (2012). Krüppel-like factor 9 and progesterone receptor coregulation of decidualizing endometrial stromal cells: implications for the pathogenesis of endometriosis. J. Clin. Endocrinol. Metab. *97*, E376–E392.

Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. USA *104* (1), 8605–8612.

Renfree, M., and Shaw, G. (2001). Reproduction in Monotremes and Marsupials (Chichester, UK: John Wiley & Sons, Ltd).

Smith, S.D., Wang, S., and Rausher, M.D. (2013). Functional evolution of an anthocyanin pathway enzyme during a flower color transition. Mol. Biol. Evol. *30*, 602–612.

Svensson, J., Jenmalm, M.C., Matussek, A., Geffers, R., Berg, G., and Ernerudh, J. (2011). Macrophages at the fetal-maternal interface express markers of alternative activation and are induced by M-CSF and IL-10. J. Immunol. *187*, 3671–3682.

Talbi, S., Hamilton, A.E., Vo, K.C., Tulac, S., Overgaard, M.T., Dosiou, C., Le Shay, N., Nezhat, C.N., Kempson, R., Lessey, B.A., et al. (2006). Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. Endocrinology *147*, 1097–1121.

Tewari, A.K., Yardimci, G.G., Shibata, Y., Sheffield, N.C., Song, L., Taylor, B.S., Georgiev, S.G., Coetzee, G.A., Ohler, U., Furey, T.S., et al. (2012). Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity. Genome Biol. *13*, R88.

Wagner, G.P., and Lynch, V.J. (2010). Evolutionary novelties. Curr. Biol. *20*, R48–R52.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. *131*, 281–285.

Wagner, G.P., Kin, K., and Lynch, V.J. (2013). A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. *132*, 159–164.

Wang, Z., Young, R.L., Xue, H., and Wagner, G.P. (2011). Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds. Nature *477*, 583–586.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y., et al. (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. *22*, 1798–1812.

Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grützner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., et al. (2008). Genome analysis of the platypus reveals unique signatures of evolution. Nature *453*, 175–183.

Cell Reports

Supplemental Information

Ancient Transposable Elements Transformed the Uterine Regulatory Landscape and Transcriptome during the Evolution of Mammalian Pregnancy

Vincent J. Lynch, Mauris C. Nnamani, Aurélie Kapusta, Kathryn Brayer, Silvia L. Plaza, Erik C. Mazur, Deena Emera, Shehzad Z. Sheikh, Frank Grützner, Stefan Bauersachs, Alexander Graf, Steven L. Young, Jason D. Lieb, Francesco J. DeMayo, Cédric Feschotte, and Günter P. Wagner

EXPERIMENTAL PROCEDURES

Cell culture: Human endometrial stromal cells (dESC) were grown in phenol red free DMEM, supplemented with 5% charcoal-stripped calf-serum, 1% antibiotic/antimycotic (ABAM), and 1x ITS+. At 80% confluency, cells were decidualized with the addition of 0.5 mM 8-Br-cAMP (Sigma) and 1mM medroxyprogesterone acetate (Sigma) to the cell culture media for 48-72 hours. Deciduaization of the dESC was confirmed by monitoring the expression of prolactin (PRL), which is only expressed in decidualized dESC, by RT-qPCR.

High-throughput transcriptome sequencing: Endometrial samples from mid-stage pregnant platypus, opossum, armadillo and dog, and day 14-18 post-implantation cow, horse, and pig were dissected to remove myometrial and placental tissue, and washed in ice-cold PBS to remove blood cells. Total RNA was extracted from cleaned endometrial tissue and dESC using the Qiagen RNA-Easy Midi RNA-extraction kit followed by on-column DNase treatment (Qiagen). Total RNA quality was assayed with a Bioanalyzer 2100 (Agilent) and found to be of excellent quality for all samples. Aliquots from the total RNA samples were sequenced using the Illumina Genome Analyzer II platform (75-bp reads), following the protocol suggested by Illumina for sequencing of cDNA samples. Two to four biological replicates each were sequenced for all samples except platypus for which only a single sample was available. mRNA-Seq reads were quality controlled following standard methods and species specific reads mapped to the human (GRCh37), macaque (MMUL_1), mouse (NCBIM37), dog (BROADD2), cow (UMD3.1), horse (EquCab2), pig (Sscrofa9), armadillo (dasNov2), opossum (monDom5), platypus (OANA5), and chicken (WASHUC2) gene builds at Ensembl using Bowtie with default parameters.

Gene expression in other cell types: Gene expression data for cell-types found in the pregnant human endometrium were obtained from previously published microarray datasets for human decidual natural killer cells (dNK; GSE5172), human decidual macrophage cells (dMP; GSE30595), and human decidual endothelial cells (dEC; GSE41946). We used the transcript presence/absence calls calculated by each study to classify genes as expressed or not in these cell types.

Gene expression calling: We used a model-based method to classify gene as expressed based on the number of transcripts per million (TPM) in the total RNA-seq dataset (Li et al., 2010; Wagner et al., 2012; 2013). Read counts were normalized by total estimated transcript

number, expressed in transcripts per million transcripts or TPM. This normalization is invariant with respect to Trimmed Mean of M-values normalization since the correction factor affects both the numerator as well as the denominator of the TPM value (Li et al., 2010; Wagner et al., 2012; 2013). The distribution of transcript abundances was fitted to a model consisting of a discretized exponential distribution, represented transcripts from repressed genes, and a negative binomial distribution, representing the distribution of abundance values of actively transcribed genes using R. The model suggests that genes with a TPM≤2 are likely from transcriptionally suppressed genes (Nagamatsu et al., 2009; Wagner et al., 2012; 2013). This threshold is consistent with one obtained by comparing the transcript abundance with the chromatin state of the respective gene (Hebenstreit et al., 2011; Taglauer et al., 2009) and we thus classified genes with TPM>2 as expressed and those with TPM≤2 as not expressed in the human endometrial stromal fibroblast data. Because the genomes of the other species we generated RNA-Seq data for are incomplete we used a hybrid approach to infer which genes in those species were expressed, we first identified all genes expressed in human endometrial stromal fibroblast with TPM>2 and then averaged the raw reads mapped per gene per million mapped reads (RMPM) for these genes. This average RMPM value (RMPM=1) was used as the expression cutoff for all non-human species.

Parsimony reconstruction of gene expression gain/loss: To identify genes that evolved endometrial expression in mammals we combined genes expressed in the human dESC, dNK, dMP, and dEC cells, and used the expressed gene calls for each species (described above). We used Wagner parsimony to reconstruct gene expression gains and losses using the method implemented in the *pars* program from PhyML (v2.4.4). We used Mesquite (v2.75) to identify the number of genes that most parsimoniously gained or lost endometrial expression; expression was classified as most parsimoniously a gain if a gene was inferred as not expressed the ancestral node (state 0) but inferred likely expressed in a descendent node (state 1/[0/1]) and *vice versa* for the classification of a loss from endometrial expression. We also used Mesquite (v2.75) to identify specific genes that unambiguously gained or lost endometrial expression, expression was classified as an unambiguous gain if a gene was not inferred as expressed the ancestral node (state 0) but inferred as expressed in a descendent node (state 1) and *vice versa* for the classification of a loss from endometrial expression. This set of genes was used for all examples discussed in the main text and shown in Figures 3-5.

Gene expression dynamics (Data show in Figure 2): We used mRNA-Seq data

generated for 27 tissues availale from the Human Protein Atlas dataset (http://www.proteinatlas.org/about/download) to calculate the tissue specificity index for recruited and ancestrally expressed genes. Following Yanai et al. (2005) the tissue specificity index (τ) is defined as:

$$\tau = \frac{\sum_{i=1}^{N}(1 - x_i)}{N - 1},$$

where N is the number of tissues (N=27), and x_i is the expression level of each gene normalized by the expression level of the tissue in which that gene is most highly expressed.

To characterize the expression of ancestrally expressed and recruited genes throughout the human menstrual cycle we used Affymetrix Human Genome U133 Plus 2.0 Array data from human endometrium sampled during the proliferative, early secretory, mid secretory, and late secretory phases of the menstrual cycle generated by (GSE4888). We used Affymetrix Human Genome U133 Plus 2.0 Array data from human endometrium samples treated with trophoblast conditioned media generated by (GSE5809) to determine the transcriptional response of ancestrally expressed and recruited genes to paracrine signals from the trophoblast. We used Agilent-014850 Whole Human Genome Microarray 4x44K G4112F data from the endometrium of humans with unexplained infertility and normal fertile controls generated by (GSE16532) to determine if ancestrally expressed and recruited genes were differentially expressed in the endometria of infertile women. Gene expression levels from these datasets were analyzed using the GEO2R implantation of the limma R packages from the Bioconductor project.

Gene function annotation: Genes recruited into endometrial expression in the Mammalian, Therian, and Eutherian stem-lineage were annotated based on their mouse knockout phenotypes and Gene Ontologies (GO) using data available at the Mouse Gene Informatics (MGI) database. Enrichments were calculated using VLAD: http://proto.informatics.jax.org/prototypes/vlad/.

H3K4me3 and H3K27ac Chromatin Immunoprecipitation-sequencing (ChIP-Seq): Human endometrium stromal cells were grown and decidualized as described above, 48 hours after cAMP/Progesterone treatment cells were harvested, resuspended and chemically cross-

linked in suspension by the addition of 1% fresh formaldehyde with 10 minutes incubation at room temperature on a rotator. Cross-linking was quenched and cells were harvested by centrifugation and washed twice in cold PBS. Nuclei were expected and Sonication was performed using a Misonix S4000 sonicator with 431A cup horn. Fragment size between 200bp and 500bp was obtained and confirmed by gel electrophoresis.

Immunoprecipitation was performed using commercially available anti-H3K4me3 (Invitrogen cat# 49-1005) and anti-H3K27ac (Millipore cat# 200551) antibody. Briefly, between 15-150 mg of isolated chromosomal DNA was incubated with 10 μg of antibody (anti-H3K4me3 or anti H3K27ac) coupled to the proteinG Dynabeads (Invitrogen). Antibody-bead complex was prepared following manufactures instructions. Chromatin-antibody-bead complex was incubated overnight at 4^{o}C in 1X ChIP Dilution buffer (0.02% SDS, 2.2% Triton X-100, 2.4mM EDTA, 33.4mM Tris pH 8.1, 334mM NaCl) supplemented with protease and phosphatase inhibitors. After incubation the complex was washed 3x with IP wash buffer (NaCl) (100mM Tris pH8.0, 500mM NaCl, 1% NP-40, 1% deoxycholic acid) followed by 2X with IP wash buffer (LiCl) (100mM Tris pH8.0, 500mM LiCl, 1% NP-40, 1% deoxycholic acid) with 3 min rotation, and once with 1ml TE buffer. Chromatin was eluted in 50mM Tris pH8.0, 10 mM EDTA, 1% SDS by incubation at 65^{o}C with agitation. The eluted DNA was incubated at 65°C overnight to reverse the cross-links. Following incubation, the immunoprecipitated DNA was treated sequentially with RNase A and Proteinase K and was then desalted using the QIAquick PCR purification kit (Qiagen). ChIP library preparation and high-throughput sequencing were performed on an Illumina Genome *Analyzer II* platform by following the protocol suggested by Illumina for sequencing chromosomal DNA. Sequencing performed were two biological replicates at 1x75bp strand specific for both H3K4me3 and H3K27ac, and an additional two biological replicates at 2x36bp for H3K4me3 by the Yale Center for Genome Analysis.

To control for the uneven background distribution, ChIP-Seq data was processed using a two-sample analysis model where both a ChIP sample and a negative control (input) were sequenced for each biological sample. With this strategy, the peak calling and FDR are determined by reversing the control and treatment data. Sequence reads were aligned to the human reference genome (hg19) using the ultra-fast short DNA sequence aligner Bowtie. Sequencing depth for ChIP-Seq samples and input averaged 34.5 million and 32 million reads respectively per biological sample with >76% overall alignment rate. For reads mapping and

peak calling we used the Model-based Analysis of ChIP-Seq (MACS, coupled with PeakSplitter as described for analyzing histone modification markers).

DNaseI-Seq data generation: Human endometrium stromal cells (hESC) were grown in steroid-depleted DMEM, supplemented with 5% charcoal-stripped calf-serum and 1% antibiotic/antimycotic (ABAM). At 100% confluency, cells were induced to decidualized DSC by treatment with 0.5 mM 8-Br-cAMP (Sigma) and 1mM medroxyprogesterone acetate (Sigma) for 48-72 hours. Deciduaization of the hESC was confirmed by monitoring the specific gene expression of prolactin (PRL) by RT-qPCR. Chromatin was isolated as previously described. Chromatin was digested with 2-3 units of DNAse I (Roche cat#04716728001) at 4^{o}C for 60 minutes. Reaction was quenched with 2x stop solution (1% SDS in 50mM EDTA) for 10 min at room temperature followed by a 1 min centrifugation. Following centrifugation the digested chromatin was treated sequentially with Proteinase K and RNAse A and was then desalted using the QIAquick PCR purification kit (Qiagen). Library preparation and high-throughput sequencing were performed on an Illumina Genome *Analyzer II* platform by following the protocol suggested by Illumina for sequencing chromosomal DNA. Sequencing was performed for two biological replicates at 1x75bp strand specific by the Yale Center for Genome Analysis.

FAIRE-Seq data generation: Human endometrium was obtained under a UNC IRB-approved protocol by office suction curettage on normal women between the age of 18 and 38, all of whom had regular, cyclic menses with an intermenstrual interval of 25-34 days. All normal subjects and lacked signs or symptoms of chronic diseases, including endometriosis and none were using hormonal medications. Portions of endometrial biopsies were frozen in liquid nitrogen in the clinic and transported to the laboratory where they were stored at -80°C until further use, while the remaining tissue was fixed in 10% buffered formalin for paraffin embedding and sectioning. Cycle phase was confirmed by examining hematoxylin and eosin stained sections using standard criteria. FAIRE was performed as previously described (Simon et al. 2012). Sequencing was performed using 36- or 50-bp single-end reads (Illumina GAIIx or HiSeq 2000). Reads were filtered using TagDust and aligned to the reference human genome (hg19) with Bowtie using default parameters.

PGR ChIP-Seq: PGR and Input ChIP were performed by Active Motif, Inc. (Carlsbad, CA) on HESCs isolated from six volunteer endometrial biopsy specimens obtained under a BCM IRB-approved protocol. The six primary cultures of hESCs were treated with 10nM 17β-

estradiol, 100nM medroxyprogesterone acetate, and 1mM 8-bromo-cAMP (all from Sigma-Aldrich Co., St. Louis, MO) for 72 hours. The hESCs were pooled and DNA was isolated before amplification using the Illumina ChIP-SeqDNASample Prep Kit. Briefly, DNA ends were polished and 59-phosphorylated using T4 DNA polymerase, Klenow polymerase, and T4 polynucleotide kinase. After addition of 3'-adenine to blunt ends using Klenow fragment (3'-5' exo minus), Illumina genomic adapters were ligated and the sample was size fractionated to 175-225bp on a 2% agarose gel. After 18 cycles of amplification using Phusion polymerase, resultant DNA libraries were tested by RT-qPCR for amplification quality. DNA libraries were sequenced on an Illumina platform and aligned to the human genome (GRCh37/hg19, February 2009) using Eland software. Aligns were extended *in silico* to 110-200bp and assigned to 32bp-bins along the genome. The resulting histograms were stored as Binary Analysis Results (BAR) files. Peaks were determined by applying a threshold of 18 (5 consecutive bins containing 0.18 aligns) and storing the results as Browser Extendable Data (BED) files. The model-based analysis of ChIP-Sequencing (MACS) algorithm was used to find peaks by normalizaing PGR ChIP against Input control with a cutoff of p-value < 10^{-10}.

Identification of Transposable Element-containing regulatory elements: To Identify regulatory elements derived from transposable elements (TEs), we first intersected FAIRE-, DNase-, H3K27ac- and H3K4me3-Seq peaks from the analyses described above with human TE annotation of the hg18 assembly (files from the U.C.S.C genome browser, with Repeat Masker v3.2.7 and repbase libraries released on 20050112) using BEDTools (intersectBed with options wa and wb) (Quinlan and Hall 2010). Prior to intersection, we verified TE lineages using manually curated lineage information in recent Repeat Masker databases (http://www.repeatmasker.org) as well as using the UCSC genome browser (BLAT searches and the "Vertebrate Multiz Alignment & Conservation (100 Species)" track). We removed non-transposable element annotations and converted the TE annotation (Repeat Masker .out file) to bed format.

Intersection files were then parsed using a custom perl script to (i) evaluate TE content of each regulatory element peak and (ii) determine the enrichment (or depletion) of each family of TE relative to that genomic abundance of that family. (i) TE content is defined by the intersection length of TE annotations and peak coordinates (corrected for overlaps between TEs). If TE content is more than 10bp, peaks are classified as 'TE-containing'. The counts of TE fragments are corrected using the interrupted repeats detection of Repeat Masker (for example, if a TE is fragmented in two because of a deletion or an insertion, it will be counted only one

time. This correction doesn't account for inversions or complex rearrangements, but is more accurate than the basic fragment number). (ii) The proportion of each family of TE is estimated in 'TE-containing' peaks, with 100% corresponding to the total amount (counts or length) of all different TEs in the analyzed set. For each TE, this proportion of counts or length is compared to the genomic abundance of that family (with 100% corresponding to the total amount of these same TEs in the genome). The ratio between counts (in-set divided by in-genome) is used to estimate if a given TE is enriched or depleted. Significance of enrichment was inferred on TE counts with three standard statistical tests (binomial, hypergeometric, and Poisson models), but figures are built on length since it is more representative of a given TE contribution.

Region-Gene associations: We used the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) to associated regulatory elements with nearby genes using the default association rules. To identify ancestral and recruited genes that were expressed in hESC we intersected ancestral and recruited genes sets with the set of genes expressed in hESCs.

De novo Motif and TFBS finding: To identify sequence motifs that were enriched in 'TE-derived' regulatory elements we used a *de novo* motif discovery approach. First we used CisFinder webserver (http://lgsun.grc.nia.nih.gov/CisFinder/) to identify over-represented short DNA motifs in 'TE-derived' regulatory elements using FDR=$1x10^{-4}$, counting motifs once per sequence, enrichment≥3, match threshold≥0.9, clustering motifs by similarity, using the "clustered" option. The top 10 scoring motifs from CisFinder were annotated using the STAMP webserver (htpp://www.benoslab.pitt.edu/stamp/) with TRANSFAC (families labeled), the Pearson correleation coefficient based column comparison metric, ungapped Smith-Waterman alignment, and iterative refinement. Previously published ChIP-Seq data for 128 transcription factors (TFs) generated by ENCODE as well as published ChIP-Seq data for AHR and ARNT (Lo and Matthews, 2012), FOXO3 (Eijkelenboom et al., 2013) were used to identify TFs with binding sites enriched within TEs compared to non-TE containing regions of the genome.

TE & Regulatory Element Correlations: Our observation that specific TEs are enriched within PGR-, FAIRE-, DNase-, H3K27ac, and H3K4me3-Seq peaks could result from a genome-wide bias for TEs to be located within regulatory elements. To explore this possibility we used a suite of permutation tests implemented in the GenometriCorr R package (Favorov et al., 2012) to determine if TEs and regulatory elements are generally located closer to each other than expected given uniform (random) genomic distribution and if so if they intersect more often than

expected. For each regulatory element dataset we calculated the correlation between the location of the peak and transposons across the genome significance of the correlation was obtained by permuting the location of peaks across the genome 1000 times. To infer the expected number of PGR-, FAIRE-, DNase-, H3K27ac, and H3K4me3-Seq peaks that contain transposable element we calculated the median number of permuted peak/transposon intersections to generate enrichment estimates; empirical *P*-values for enrichment estimates were calculated from distribution of the number of peaks in each permuted dataset that overlapped transposable elements.

We found a statistically significant positive correlation between the absolute distance between FAIRE-Seq ($P<0.001$), DNase-Seq ($P<0.001$), H3K27ac-Seq ($P<0.001$) and H3K4me3-Seq ($P<0.001$) peaks and TEs, such that these distances were closer than expected (i.e. in the lower tail of the permuted distributions) given a random distribution of peaks relative to TEs. However, FAIRE-Seq (enrichment: 0.73; $P<0.001$), DNase-Seq (enrichment: 0.64; $P<0.001$), H3K27ac-Seq (enrichment: 0.82; $P<0.001$), and H3K4me3-Seq (enrichment: 0.64; $P<0.001$) peaks contained significantly fewer transposable elements than expected given a random distribution. These results indicate that while transposons are generally found close to regulatory elements they occur less frequently than expected within regulatory elements, thus our observation that specific TEs are enriched within FAIRE-, DNase-, H3K27ac-, and H3K4me3-Seq peaks does not result from a general bias for TEs to be found in regulatory elements.

Favorov, A., Mularoni, L., Cope, L.M., Medvedeva, Y., Mironov, A.A., Makeev, V.J., and Wheelan, S.J. (2012). Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput. Biol. *8*, e1002529.

Hebenstreit, D., Fang, M., Gu, M., Charoensawan, V., van Oudenaarden, A., and Teichmann, S.A. (2011). RNA sequencing reveals two major classes of gene expression levels in metazoan cells. Mol. Syst. Biol. *7*.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics *26*, 493–500.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. *131*, 281–285.

Wagner, G.P., Kin, K., and Lynch, V.J. (2013). A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. *132*, 159–164.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics *21*, 650–659.

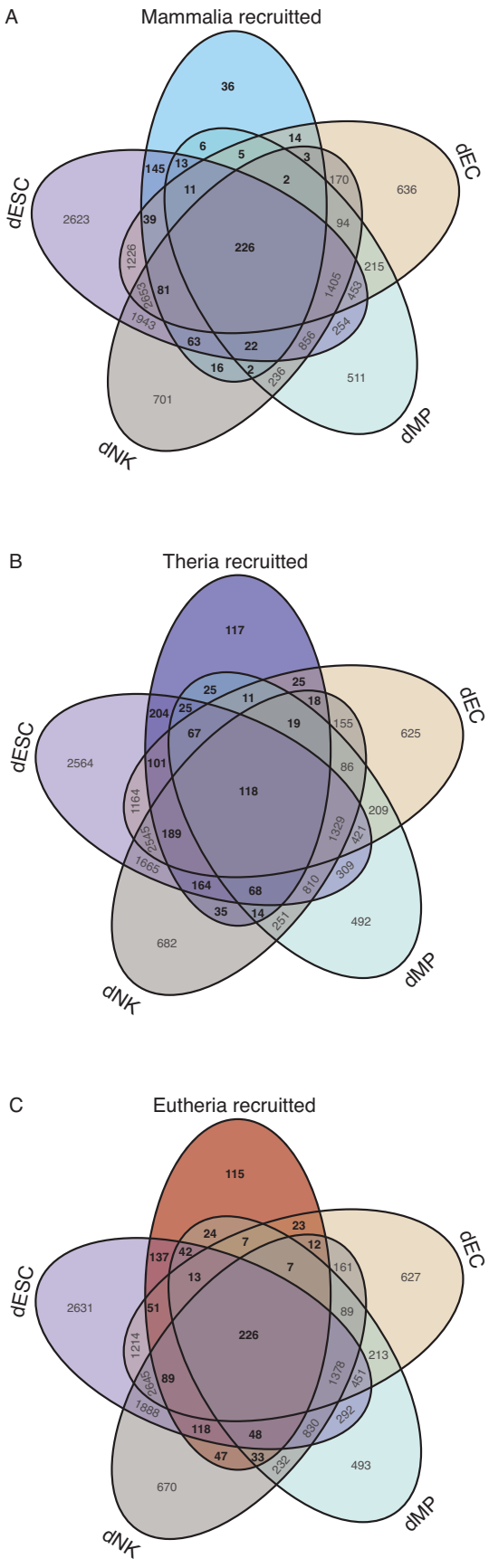
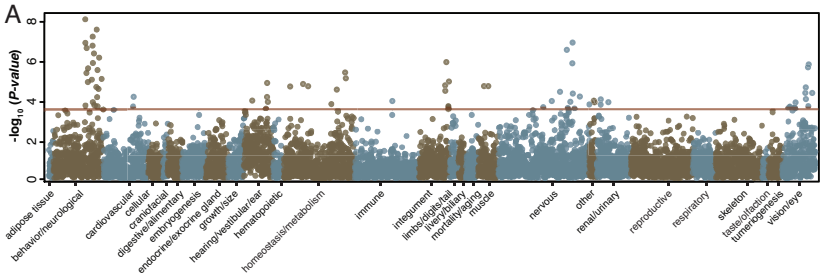
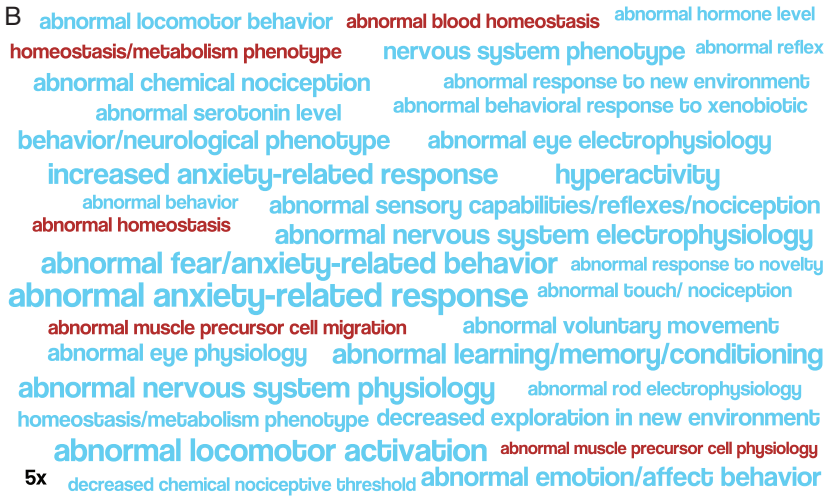
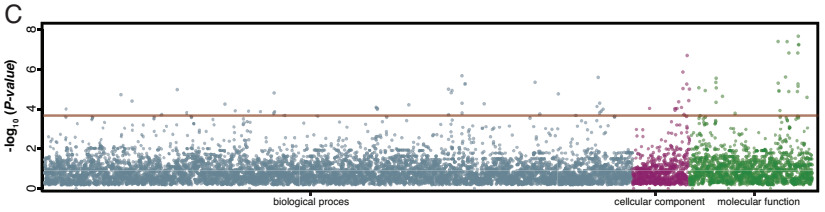
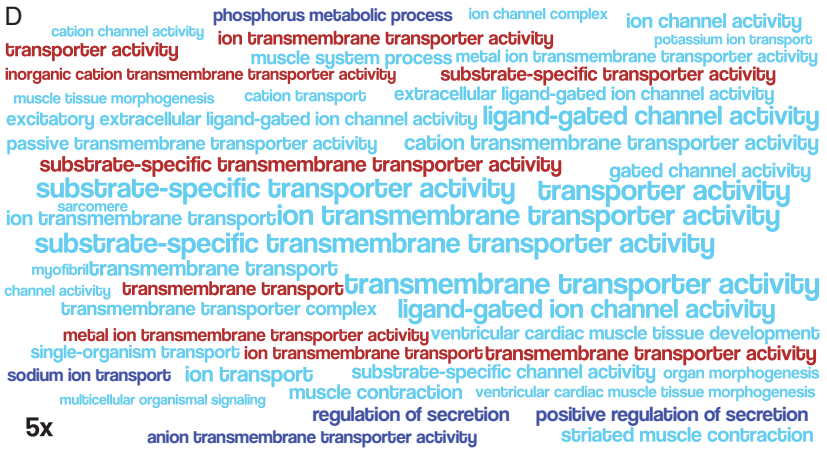


Fig. S1. Recruited genes are preferentially expressed in decidualized endometrial stromal cells, related to Figure 1. Venn diagram showing in which of the four major cell-types found in pregnant endometrium recruited genes are expressed. dESC, cAMP/MPA decidualized endometrial stromal cells; dNK, decidual natural killer cells; dMP, decidual macrophage; dEC decidual endothelial cells.


A
-log10 (P-value)
adipose tissue
behavior/neurological
cardiovascular
cellular
craniofacial
digestive/alimentary
embryogenesis
endocrine/exocrine gland
growth/size
hearing/vestibular/ear
hematopoietic
homeostasis/metabolism
immune
integument
limbs/digits/tail
liver/biliary
mortality/aging
muscle
nervous
other
renal/urinary
reproductive
respiratory
skeleton
taste/olfaction
tumorigenesis
vision/eye


B
abnormal locomotor behavior
abnormal blood homeostasis
abnormal hormone level
homeostasis/metabolism phenotype
nervous system phenotype
abnormal reflex
abnormal chemical nociception
abnormal response to new environment
abnormal behavioral response to xenobiotic
abnormal serotonin level
behavior/neurological phenotype
abnormal eye electrophysiology
increased anxiety-related response
hyperactivity
abnormal behavior
abnormal sensory capabilities/reflexes/nociception
abnormal homeostasis
abnormal nervous system electrophysiology
abnormal fear/anxiety-related behavior
abnormal response to novelty
abnormal anxiety-related response
abnormal touch/ nociception
abnormal muscle precursor cell migration
abnormal voluntary movement
abnormal eye physiology
abnormal learning/memory/conditioning
abnormal nervous system physiology
abnormal rod electrophysiology
homeostasis/metabolism phenotype
decreased exploration in new environment
abnormal locomotor activation
abnormal muscle precursor cell physiology
5x
decreased chemical nociceptive threshold
abnormal emotion/affect behavior


C
-log10 (P-value)
biological proces
cellcular component
molecular function


D
phosphorus metabolic process
ion channel complex
ion channel activity
cation channel activity
ion transmembrane transporter activity
potassium ion transport
transporter activity
muscle system process
metal ion transmembrane transporter activity
inorganic cation transmembrane transporter activity
substrate-specific transporter activity
muscle tissue morphogenesis
cation transport
extracellular ligand-gated ion channel activity
excitatory extracellular ligand-gated ion channel activity
ligand-gated channel activity
passive transmembrane transporter activity
cation transmembrane transporter activity
substrate-specific transmembrane transporter activity
gated channel activity
substrate-specific transporter activity
transporter activity
sarcomere
ion transmembrane transport
ion transmembrane transporter activity
substrate-specific transmembrane transporter activity
myofibril
transmembrane transport
channel activity
transmembrane transport
transmembrane transporter activity
transmembrane transporter complex
ligand-gated ion channel activity
metal ion transmembrane transporter activity
ventricular cardiac muscle tissue development
single-organism transport
ion transmembrane transport
transmembrane transporter activity
sodium ion transport
ion transport
substrate-specific channel activity
organ morphogenesis
multicellular organismal signaling
muscle contraction
ventricular cardiac muscle tissue morphogenesis
5x
regulation of secretion
positive regulation of secretion
anion transmembrane transporter activity
striated muscle contraction

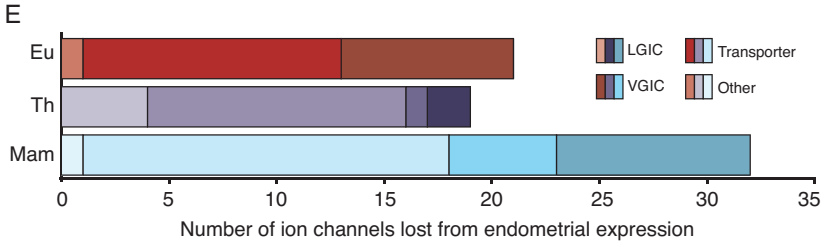

E
Eu
Th
Mam
LGIC
Transporter
VGIC
Other
0
5
10
15
20
25
30
35
Number of ion channels lost from endometrial expression

Fig. S2. Genes that lost endometrial expression are enriched in transporter and ion channel functions, related to Figure 2.

(A) Manhattan plot of -$\log_{10}$ *P*-values (hypergeometric test) for moue knockout phenotypes enriched among genes that lost endometrial expression in early mammals. Phenotypes are grouped by anatomical system. Horizontal red line indicates the dataset-wide false discovery rate (FDR q-value=0.1).

(B) Word cloud of mouse knockout phenotypes enriched among genes that lost endometrial expression in the Mammalian (light blue), Therian (blue), and Eutherian (red) stem-lineage. Abnormal phenotypes are scaled to the $\log_2$ enrichment of that term (see inset scale).

(C) Manhattan plot of -$\log_{10}$ *P*-values (hypergeometric test) for Gene Ontology (GO) terns enriched among genes that lost endometrial expression in early mammals. Go terms are grouped into biological process, cellular component, and molecular function. Horizontal red line indicates the dataset-wide false discovery rate (FDR q-value=0.1).

(D) Word cloud of GO terms enriched among genes that lost endometrial expression in the Mammalian (light blue), Therian (blue), and Eutherian (red) stem-lineage. GO terms are scaled to the $\log_2$ enrichment of that term (see inset scale).

(E) Loss of ion channels from endometrial expression in the Eutherian (Eu), Therian (Th), and Mammalian (Mam) stem-lineages. Stacked bar chart shows the number of ligand gated ion channels (LGIC), voltage gated ion channels (VGIC), transporters, and other kinds (Other) ion channels that lost endometrial expression in each lineage.