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.

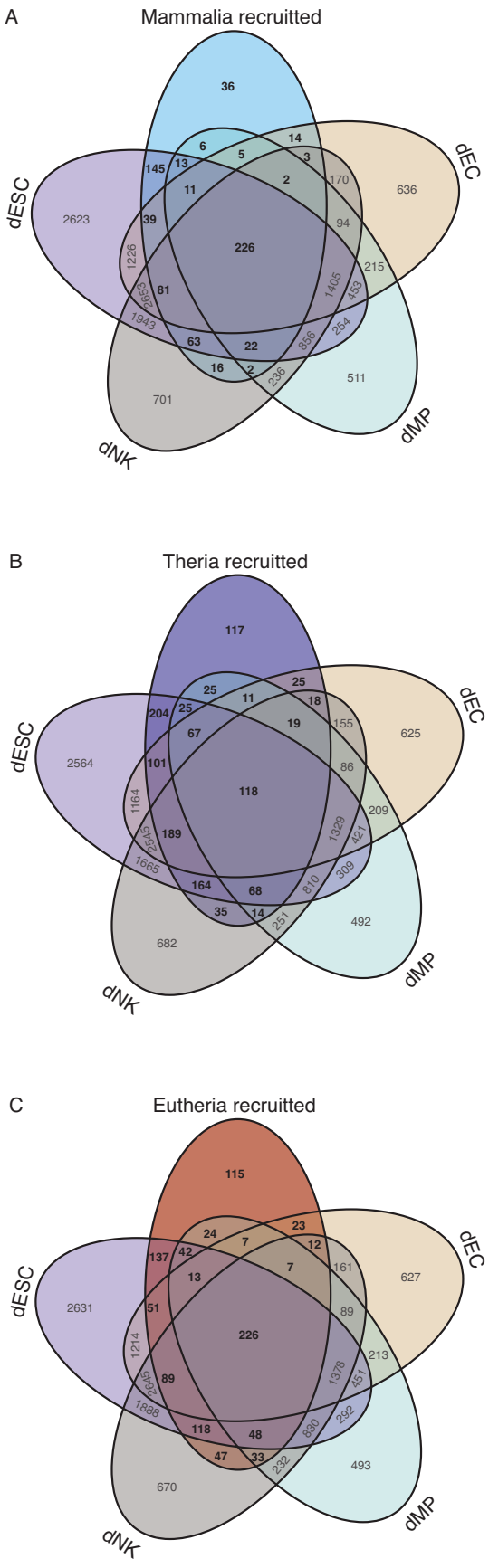
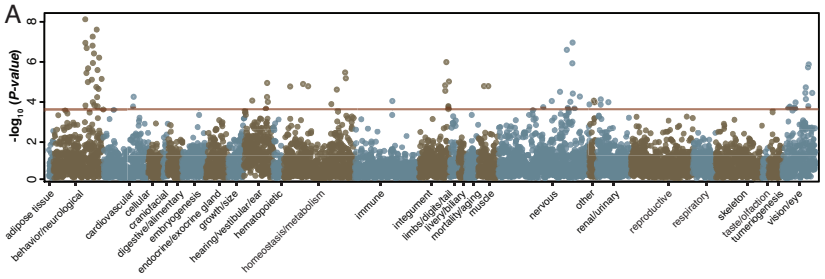
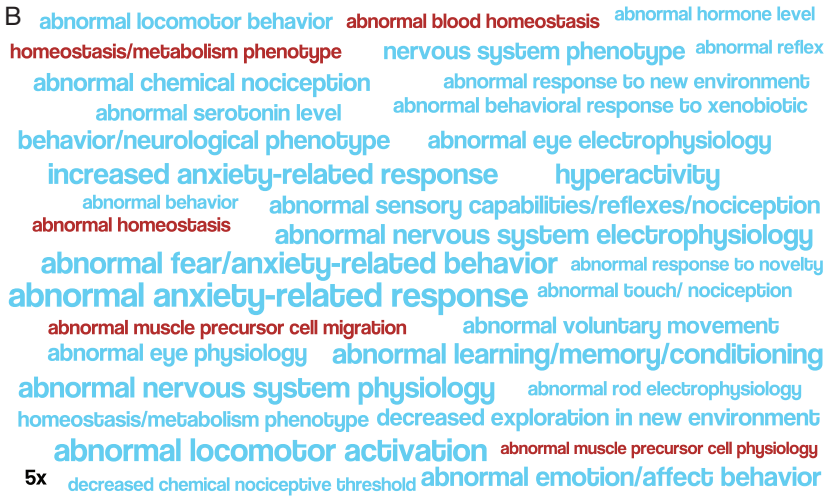
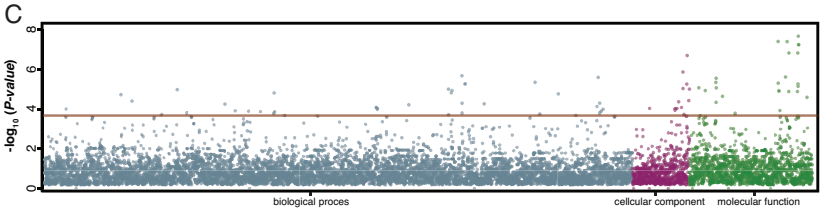
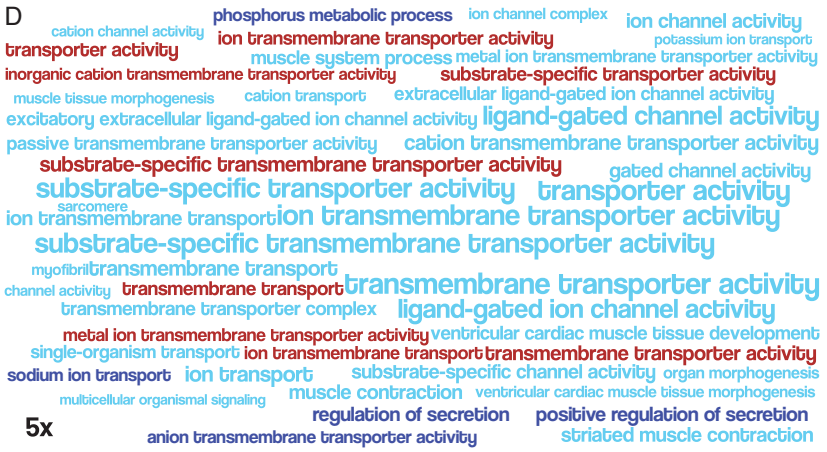


Fig. S1. Recruited genes are preferentially expressed in decidualized endometrial stromal cells, related to Figure 1. Venn diagram showing in which of the four major cell-types found in pregnant endometrium recruited genes are expressed. dESC, cAMP/MPA decidualized endometrial stromal cells; dNK, decidual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

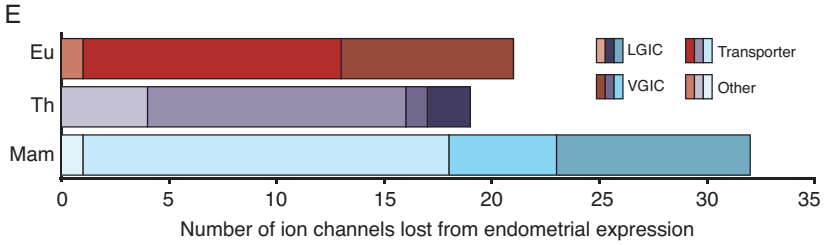

E
Eu
Th
Mam
LGIC
Transporter
VGIC
Other
0
5
10
15
20
25
30
35
Number of ion channels lost from endometrial expression

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.