Supplementary Tables and Figures

"Genome-wide Analysis and Functional Prediction of the Estrogen-Regulated Transcriptional Response in the Mouse Uterus"

Vasquez *et al.* (2019)

This document contains the following supplementary information: Page

1) Supplementary Tables

See the individual Excel spreadsheets containing Supplementary Tables S1 through S6 included with this paper. Supplementary Table S7 is shown below.

Supplemental Table S1: RNA-seq Analysis of Protein Coding genes (mRNA).
Ovariectomized wild-type (WT) C57BL/6J animals were treated with 100 ng of E2 in corn oil for 0.5, 1, 2, and 6 hour (hr) or vehicle control (0 hr). RNA was isolated from whole uteri. PolyA RNA species were selected for strand-specific sequencing and differential expression analysis. Each sheet summarizes the results for cuffdiff analysis of differential gene expression for the following comparisons: (**1**) 0hr vs. 0.5hr (**2**) 0hr vs. 1hr, (**3**) 0hr vs. 2hr, and (**4**) 0hr vs. 6hr. Headings are as follows: (**A**) test_id = A unique identifier describing the transcript, gene, primary transcript, or CDS being tested, (**B**) gene_id = the gene_name(s) or gene_id(s) being tested, (**C**) gene = gene name, (**D**) locus = genomic coordinates for easy browsing to the genes or transcripts being tested, (**E**) sample_1 = the first sample being tested, (**F**) sample_2 = the second sample being tested, (**G**) status = OK (test successful), (**H**) value_1 = FPKM of sample 1, (**I**) value_2= FPKM of sample 2, (**J**) log2(fold_change), (**K**) test_stat = x(**L**) p_value = uncorrected p-value of the test statistic, (**M**) q_value = FDR-adjusted p-value of the test statistic, and (**N**) significant = can be either "yes" or "no", depending on whether p is greater than the FDR after Benjamini-Hochberg correction for multiple-testing.

Supplemental Table S2: RNA-seq Analysis of Long Noncoding RNAs (lncRNA).
Ovariectomized wild-type (WT) C57BL/6J animals were treated with 100 ng of E2 in corn oil for 0.5, 1, 2, and 6 hour (hr) or vehicle control (0 hr). RNA was isolated from whole uteri. PolyA RNA species were selected for strand-specific sequencing and differential expression analysis. Each sheet summarizes the results for cuffdiff analysis of differential gene expression for the following comparisons: (**1**) 0hr vs. 0.5hr (**2**) 0hr vs. 1hr, (**3**) 0hr vs. 2hr, and (**4**) 0hr vs. 6hr. Headings are as follows: (**A**) test_id = A unique identifier describing the transcript, gene, primary transcript, or CDS being tested, (**B**) gene_id = the gene_name(s) or gene_id(s) being tested, (**C**) gene = gene name, (**D**) locus = genomic coordinates for easy browsing to the genes or transcripts being tested, (**E**) sample_1 = the first sample being tested, (**F**) sample_2 = the second sample being tested, (**G**) status = OK (test successful), (**H**) value_1 = FPKM of sample 1, (**I**) value_2= FPKM of sample 2, (**J**) log2(fold_change), (**K**) test_stat = x(**L**) p_value = uncorrected p-value of the test statistic, (**M**) q_value = FDR-adjusted p-value of the test statistic, and (**N**) significant = can be either "yes" or "no", depending on whether p is greater than the FDR after Benjamini-Hochberg correction for multiple-testing.

Supplemental Table S3: Biological Processes Enriched in E2-regulated mRNAs.
Gene Set Enrichment Analysis of E2-regulated mRNAs in the mouse uterus. The universe of ontology terms for biological processes with the associated normalized enrichment scores at each time point of E2 treatment. Headings are as follows: (**A**) ID= gene ontology identification, (**B**) Biological Pathway, (**C**) NES_0.5hr = normalized enrichment score for GO term at 0.5 hour, (**D**) NES_1hr = normalized enrichment score for GO term at 1 hour, (**E**) NES_2hr = normalized enrichment score for GO term at 2 hours, and (**F**) NES_6hr = normalized enrichment score for GO term at 6 hour.

Supplemental Table S4: Molecular Functions Enriched in E2-regulated mRNAs.
Gene Set Enrichment Analysis of E2-regulated mRNAs in the mouse uterus. The universe of ontology terms for molecular functions with the associated normalized enrichment scores at each time point of E2 treatment. Headings are as follows: (**A**) ID= gene ontology identification, (**B**) Molecular Function, (**C**) NES_0.5hr = normalized enrichment score for GO term at 0.5 hour, (D) NES_1hr = normalized enrichment score for GO term at 1 hour, (**E**) NES_2hr = normalized enrichment score for GO term at 2 hours, and (**F**) NES_6hr = normalized enrichment score for GO term at 6 hour.

Supplemental Table S5: Genomic Region Enrichment of Annotations (GREAT) LncRNA-mRNA Associations.
Version 3.0.0. Species assembly mm10. Association rule: Basal+extension: 5000 bp upstream, 1000 bp downstream, 1000000 bp max extension, curated regulatory domains included. Headings are as follows: (**A**) target = genomic annotation associated to test region (**B**) Test Region (Distance to target).

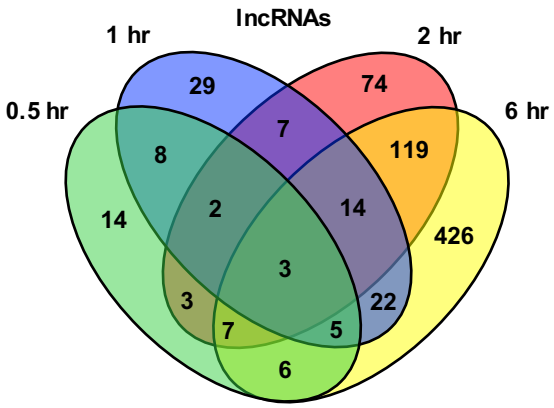
Supplemental Table S6: LncRNA Conservation Analysis.
Summary of lncRNA conservation analysis that identified 20 E2-regulated mouse lncRNAs with human homologs. Headings are as follows: (**A**)Transcript = transcript identifier in mm10, (**B**) Locus = chromosomal region in mm10, (**C**) Gene = gene name for mouse, (**D**) Length = total length of chromosomal region, (**E**) chrom_mapped = chromosome region mapped to human (hg38), (F) Score = BLAST score, (**G**) Evalue = The Expectation value or Expect value represents the number of different alignments with scores equivalent to or better than S that is expected to occur in a database search by chance. The lower the E value, the more significant the score and the alignment. (**H**) Identities = Percentage of homology between both fragments (experimental/reference), (**I**) Gaps = total number of gaps allowed during alignment. (**J**) Position conserved = Neighboring genes conserved across mouse-human, (**K**) Human homolog = gene name for human homolog.

Supplementary Table S7. E2-regulated LncRNAs with Functions in Cancer
See Supplementary Table S7 on the next page.

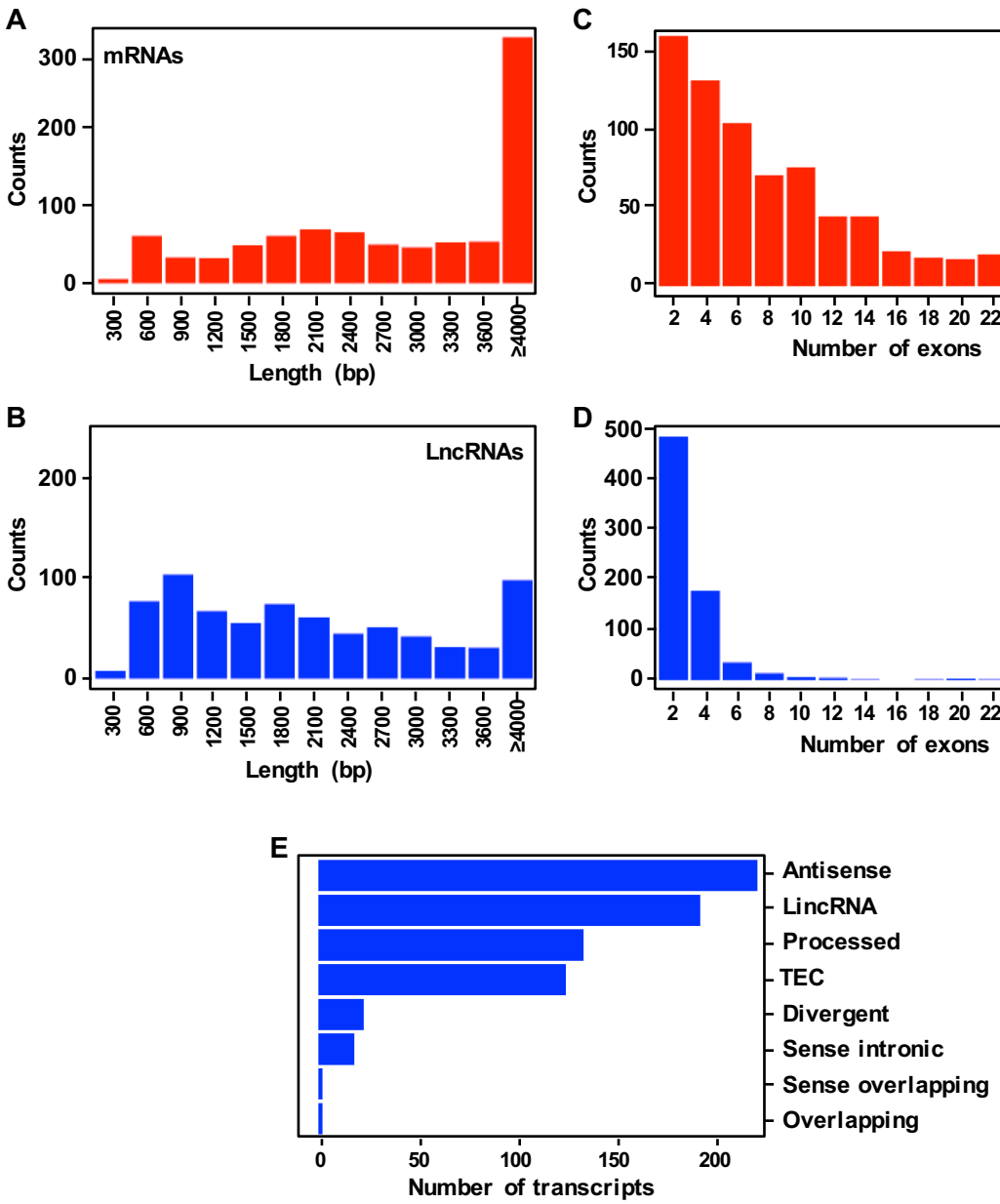
Supplementary Table S7. E2-regulated LncRNAs with Functions in Cancer
Representative studies exploring the role of lncRNAs in various cancer models.

LncRNA	Cell Type	Aberrant Phenotype
GAS5	ovarian cancer	overexpression inhibits proliferation, colony formation, and apoptosis [1]
ZFAS1	hepatoma	overexpression promotes cell invasion and tumor metastasis *in vitro* and *in vivo* [2]
SNHG5	osteosarcoma	overexpression promotes growth and metastasis [3]
	hepatoma	overexpression represses cell apoptosis, induces cell cycle progression, and promotes proliferation, invasion, and migration *in vitro* [4]
	colorectal adenocarcinoma	overexpression enhances proliferation, metastasis, migration and inhibits apoptosis [5]
SNHG8	hepatocellular carcinoma	overexpression accelerates cell proliferation, colony formation, invasion and migration [6]
	non-small cell lung cancer	overexpression promotes growth of non-small cell lung tumors [7]
SNHG18	glioma cancer	overexpression promotes radioresistance and growth in tumors in vivo [8]
SNHG15	colorectal adenocarcinoma	overexpression promotes cell proliferation, colony formation and invasion in vitro, and tumor growth *in vivo* [9]
	thyroid cancer	overexpression promotes cell proliferation, migration, and invasion [10]
SNHG14	glioma cancer	overexpression inhibits cell viability, reduces cell invasion, and induces apoptosis [11]
	breast cancer HER2 positive subtype	overexpression promotes cell proliferation, invasion and trastuzumab resistance [12]
SNHG12	colorectal adenocarcinoma	overexpression promotes cell cycle progression and inhibits apoptosis [13]
SNHG1	colorectal adenocarcinoma	overexpression promotes cell proliferation [14]
MIR22HG	human hepatocellular carcinoma	overexpression suppresses proliferation, invasion, and metastasis *in vitro* and *in vivo* [15]
MIR17HG	glioma cancer	knockdown in combination with FXR1 reduces proliferation, migration and invasion abilities, and increases apoptosis [16]
KCNQ1OT1	lung adenocarcinoma	knockdown depresses proliferation and invasion and promotes apoptosis [17]
	non-small cell lung cancer	overexpression inhibits tumor growth *in vivo* [18]
	breast cancer luminal subtype	overexpression promotes tumor growth *in vivo* [19]
H19	endometrial cancer	overexpression promotes cell proliferation [20]
	uterine leiomyoma	overexpression promotes cell proliferation [21]
FTX	hepatocellular carcinoma	overexpression increases proliferation and promotes Warburg effect [22]
	hepatocellular carcinoma	overexpression promotes proliferation and cell cycle progression [23]
DNM3OS	ovarian cancer	knockdown results in altered epithelial-to-mesenchymal-linked genes/pathways, promotes mesenchymal-to-epithelial transition, and reduces cell migration and invasion [24]
	esophageal squamous cell carcinoma	overexpression confers radioresistance [25]

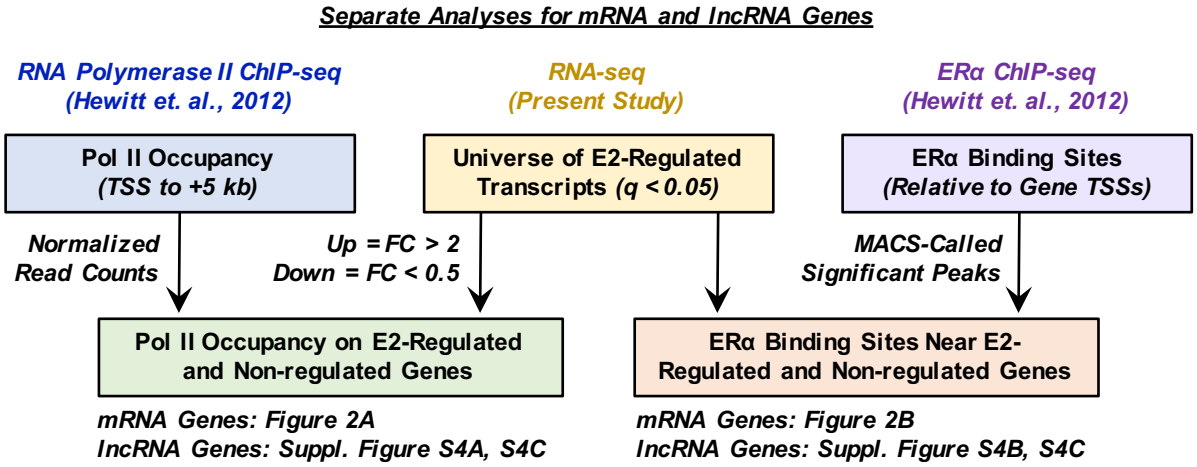
2) Supplementary Figures



Supplementary Figure S1. Venn Diagram Representation of E2-regulated LncRNAs in the Mouse Uterus.
Comparison of the numbers of lncRNAs regulated by E2 at 0.5, 1, 2, and 6 hours relative to 0 hour (hr).

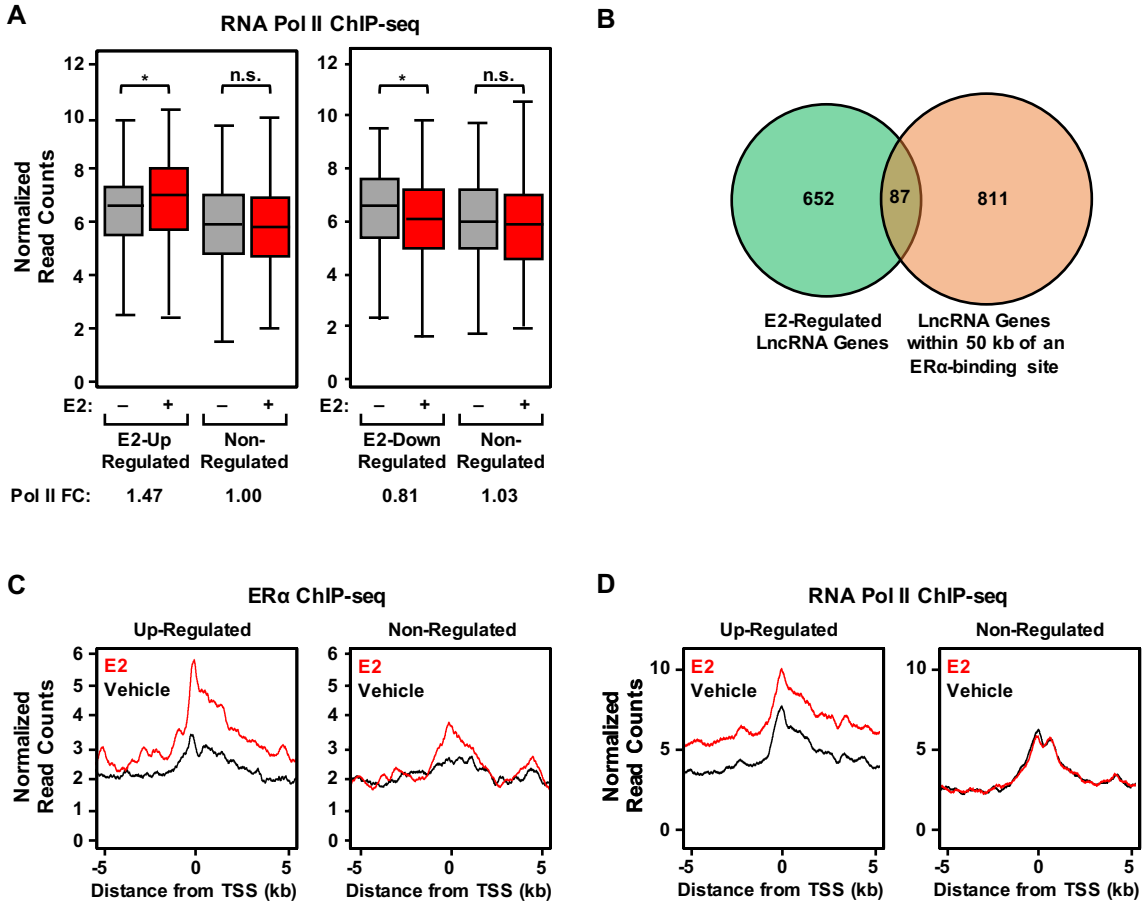


Supplementary Figure S2. Molecular Features of E2-regulated mRNAs and LncRNAs.
(A) Distribution of mRNA length. **(B)** Distribution of lncRNA length. **(C)** Distribution of exon content in mRNAs. **(D)** Distribution of exon content in lncRNAs. **(E)** Biotypes of E2-regulated lncRNAs as defined by GENCODE and ENSEMBL.



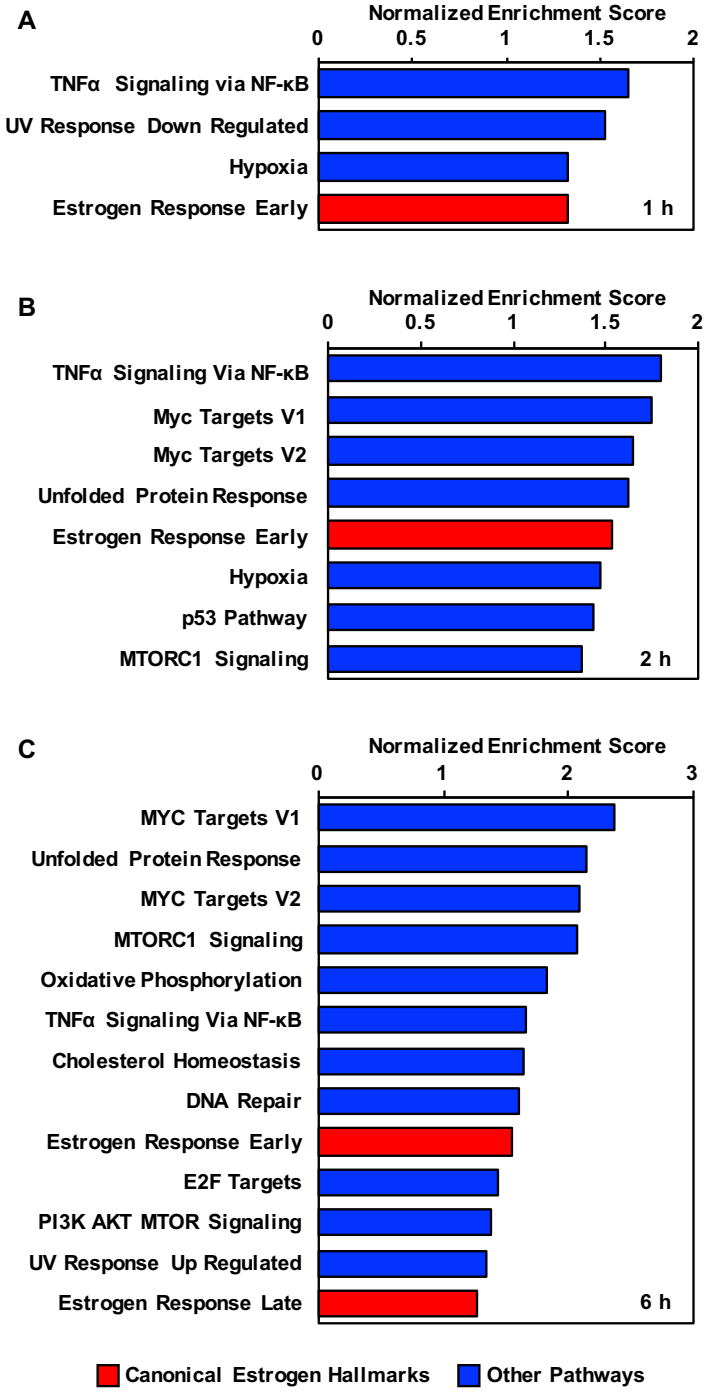
Supplementary Figure S3. Flow Chart for the Integration of RNA-seq and ChIP-seq Datasets.

The genome-wide enrichment profiles for RNA Polymerase II (Pol II) and ERα were determined previously by chromatin immunoprecipitation sequencing (ChIP-seq) (Hewitt et. al., 2012) [26]. The raw data were processed and integrated with the RNA-seq data generated herein as in as described Materials and Methods. For these analyses, we determined the following for the set of E2-regulated mRNA and lncRNA genes (determined separately): (1) Pol II enrichment at the promoters (TSS to +5 kb) and (2) the location of the nearest ERα binding site. This allowed us to determine (1) the Pol II occupancy on E2-regulated and non-regulated genes, expressed as normalized read counts, and (2) the fraction of E2-regulated genes with ERα binding sites located within 20 kb or 50 kb, as well as enrichment of ERα at the promoters (-5 kb to +5 kb) of E2-regulated and non-regulated genes.

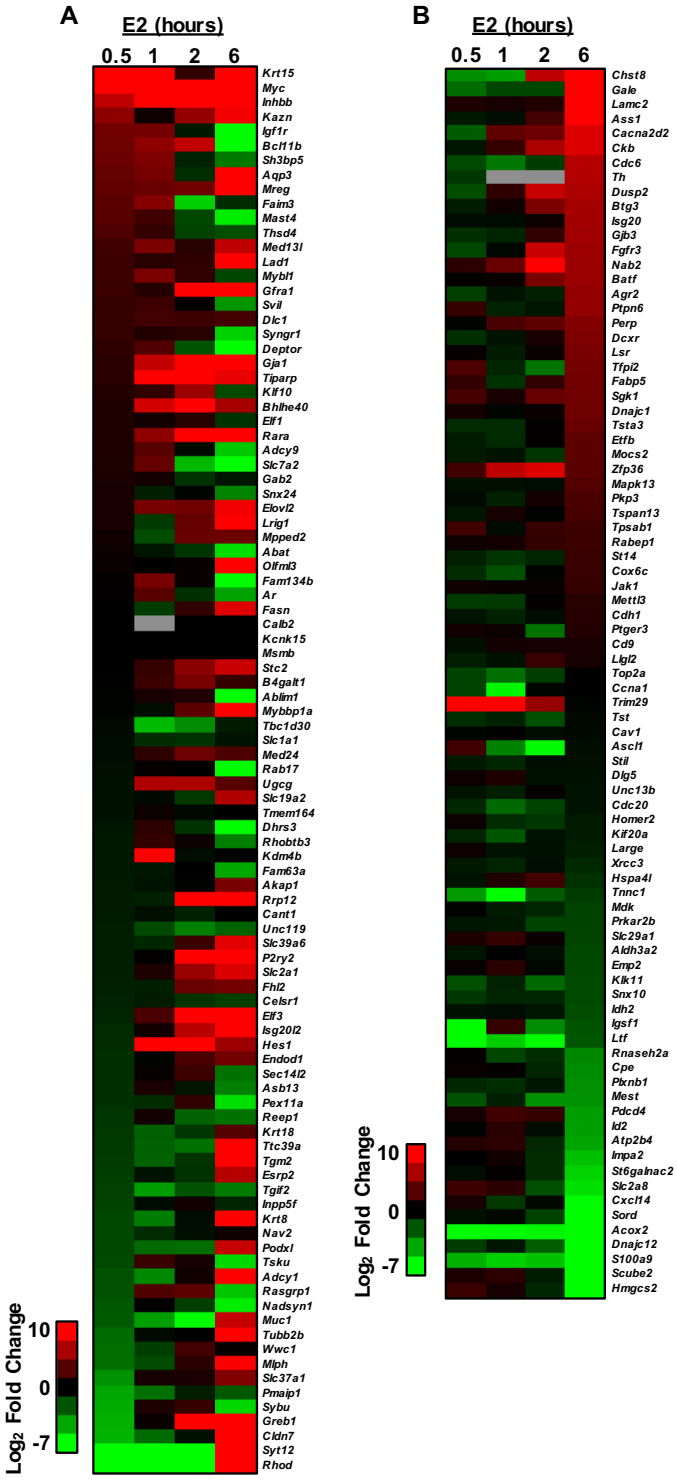


Supplemental Figure S4. Occupancy of RNA Polymerase II and ERα near E2-regulated LncRNA Genes.

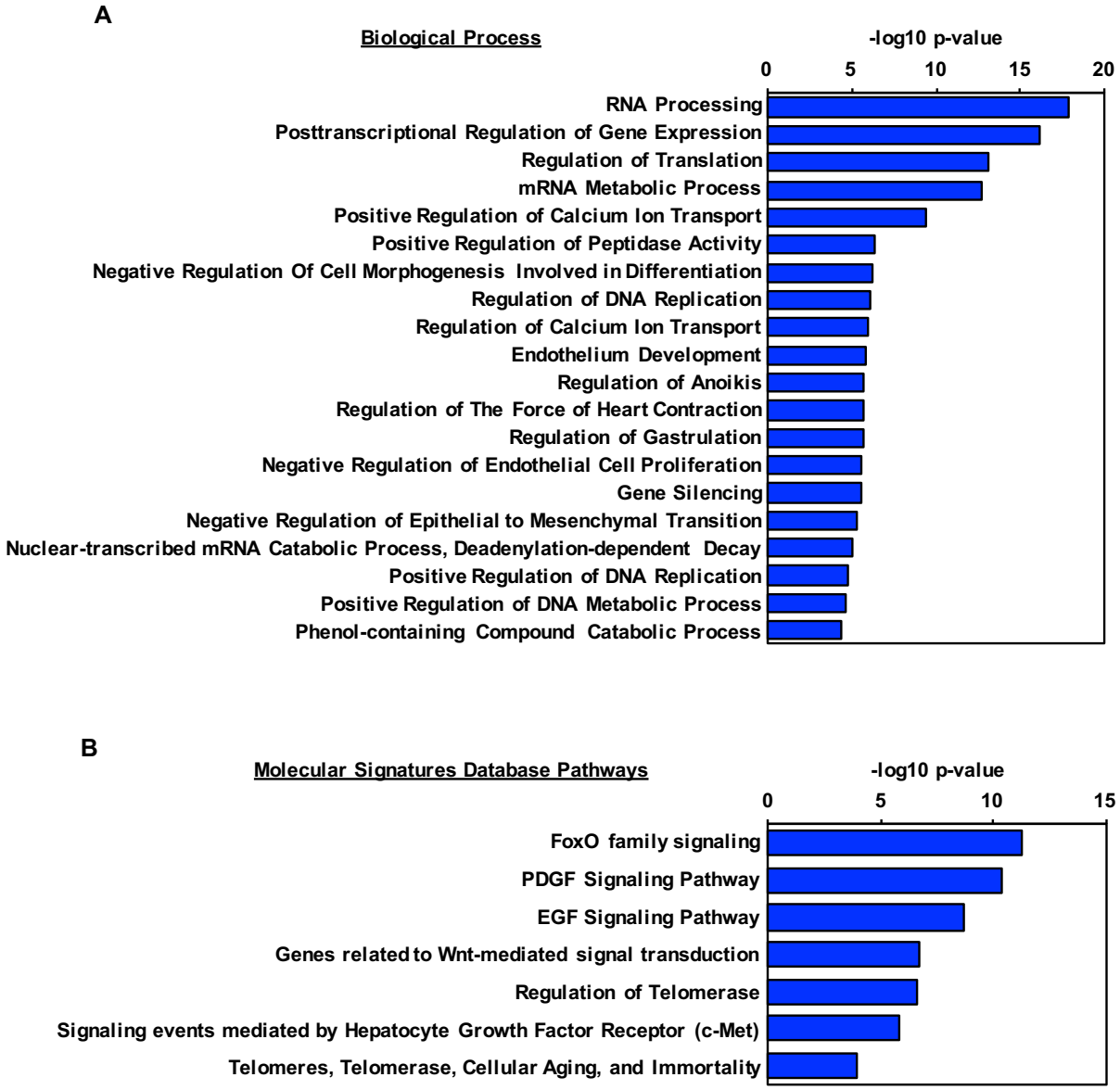
Pol II and ERα occupancy on E2-regulated and non-regulated lncRNA genes in vehicle- and E2-treated ovariectomized mouse uteri, determined as shown in Supplementary Figure S3 using ChIP-seq data from Hewitt et al. (2012) [26] and RNA-seq data generated as described herein. **(A)** The enrichment of Pol II at the promoters (TSS to +5 kb; determined by Pol II ChIP-seq) of E2-upregulated, E2-downregulated, and non-regulated genes [upregulated = fold change (FC) > 2, downregulated = FC < 0.5; determined by RNA-seq) in uteri collected from vehicle- or E2-treated mice (grey and red boxes, respectively) was determined and expressed in box plots for each condition. Significance was determined by a Wilcoxon Rank Sum test with p-values indicated (asterisk, upregulated p = 0.00337, downregulated p = 0.0007392; n.s., not significant). "Pol II FC" = the fold change in Pol II ChIP-seq normalized read counts for vehicle versus E2. **(B)** Comparison of lncRNA genes regulated by E2 at all time points tested as determined by RNA-seq (green circle) with lncRNA genes residing within 50 kb of an ERα-binding site determined by ChIP-seq in E2-treated mouse uteri (orange circle). **(C and D)** Metaplot representations of ERα (panel C) and RNA Pol II (panel D) ChIP-seq enrichment at the promoters (-5 kb to +5 kb) of E2-upregulated and nonregulated lncRNA genes.



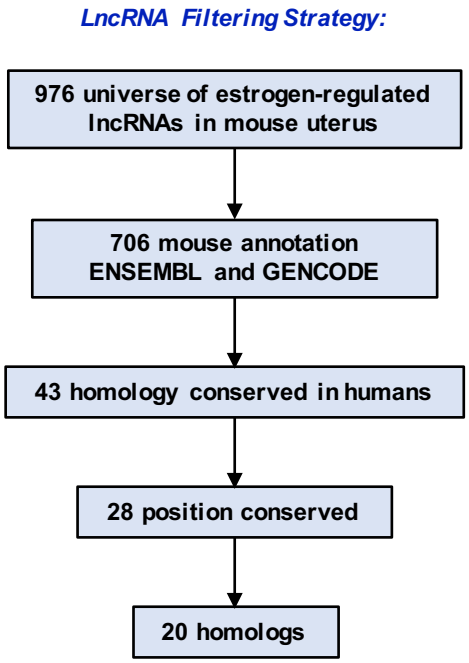
Supplementary Figure S5. Gene Set Enrichment Analysis (GSEA) of Differentially Regulated Protein-coding Genes.
GSEA analysis identified enriched pathways associated with genes regulated by E2 at **(A)** 1 hour, **(B)** 2 hours, and **(C)** 6 hours.



Supplementary Figure S6. Gene Set Enrichment Analysis of Genes Regulated by E2.
Heat map representation of genes categorized in hallmarks of **(A)** "Early" and **(B)** "Late" E2 response. Genes are ranked by maximum expression at 0.5, and 6 hours respectively.



Supplementary Figure S7. Genomic Regions Enrichment of Annotations Tool (GREAT) Analysis of E2-regulated LncRNAs.
(A) Biological processes **(B)** Molecular Signatures Database (MsigDB) Pathways.



Supplementary Figure S8. Filtering strategy for conservation lncRNAs.
Flowchart illustrating each level of filtering starting from a universe of 976 mouse lncRNAs, each level depicts one stage of the filtering process ending with 20 genes with human homologs.

3) Supplementary References

1. Li J, Yang C, Li Y, Chen A, Li L, You Z. LncRNA GAS5 suppresses ovarian cancer by inducing inflammasome formation. *Biosci Rep* 2017.
2. Li T, Xie J, Shen C, Cheng D, Shi Y, Wu Z, Deng X, Chen H, Shen B, Peng C, Li H, Zhan Q, et al. Amplification of long noncoding RNA ZFAS1 promotes metastasis in hepatocellular carcinoma. *Cancer Res* 2015; 75:3181-3191.
3. Ju C, Zhou R, Sun J, Zhang F, Tang X, Chen KK, Zhao J, Lan X, Lin S, Zhang Z, Lv XB. LncRNA SNHG5 promotes the progression of osteosarcoma by sponging the miR-212-3p/SGK3 axis. *Cancer Cell Int* 2018; 18:141.
4. Li Y, Guo D, Zhao Y, Ren M, Lu G, Wang Y, Zhang J, Mi C, He S, Lu X. Long non-coding RNA SNHG5 promotes human hepatocellular carcinoma progression by regulating miR-26a-5p/GSK3beta signal pathway. *Cell Death Dis* 2018; 9:888.
5. Zhang M, Li Y, Wang H, Yu W, Lin S, Guo J. LncRNA SNHG5 affects cell proliferation, metastasis and migration of colorectal cancer through regulating miR-132-3p/CREB5. *Cancer Biol Ther* 2019; 20:524-536.
6. Dong J, Teng F, Guo W, Yang J, Ding G, Fu Z. LncRNA SNHG8 promotes the tumorigenesis and metastasis by sponging miR-149-5p and predicts tumor recurrence in hepatocellular carcinoma. *Cell Physiol Biochem* 2018; 51:2262-2274.
7. Chen C, Zhang Z, Li J, Sun Y. SNHG8 is identified as a key regulator in non-small-cell lung cancer progression sponging to miR-542-3p by targeting CCND1/CDK6. *Onco Targets Ther* 2018; 11:6081-6090.
8. Zheng R, Yao Q, Ren C, Liu Y, Yang H, Xie G, Du S, Yang K, Yuan Y. Upregulation of long noncoding RNA small nucleolar RNA host gene 18 promotes radioresistance of glioma by repressing semaphorin 5A. *Int J Radiat Oncol Biol Phys* 2016; 96:877-887.
9. Saeinasab M, Bahrami AR, Gonzalez J, Marchese FP, Martinez D, Mowla SJ, Matin MM, Huarte M. SNHG15 is a bifunctional MYC-regulated noncoding locus encoding a lncRNA that promotes cell proliferation, invasion and drug resistance in colorectal cancer by interacting with AIF. *J Exp Clin Cancer Res* 2019; 38:172.
10. Liu Y, Li J, Li F, Li M, Shao Y, Wu L. SNHG15 functions as a tumor suppressor in thyroid cancer. *J Cell Biochem* 2019; 120:6120-6126.
11. Wang Q, Teng Y, Wang R, Deng D, You Y, Peng Y, Shao N, Zhi F. The long non-coding RNA SNHG14 inhibits cell proliferation and invasion and promotes apoptosis by sponging miR-92a-3p in glioma. *Oncotarget* 2018; 9:12112-12124.
12. Dong H, Wang W, Mo S, Liu Q, Chen X, Chen R, Zhang Y, Zou K, Ye M, He X, Zhang F, Han J, et al. Long non-coding RNA SNHG14 induces trastuzumab resistance of breast cancer via regulating PABPC1 expression through H3K27 acetylation. *J Cell Mol Med* 2018; 22:4935-4947.
13. Wang JZ, Xu CL, Wu H, Shen SJ. LncRNA SNHG12 promotes cell growth and inhibits cell apoptosis in colorectal cancer cells. *Braz J Med Biol Res* 2017; 50:e6079.

14. Tian T, Qiu R, Qiu X. SNHG1 promotes cell proliferation by acting as a sponge of miR-145 in colorectal cancer. *Oncotarget* 2018; 9:2128-2139.

15. Zhang DY, Zou XJ, Cao CH, Zhang T, Lei L, Qi XL, Liu L, Wu DH. Identification and functional characterization of long non-coding RNA MIR22HG as a tumor suppressor for hepatocellular carcinoma. *Theranostics* 2018; 8:3751-3765.

16. Cao S, Zheng J, Liu X, Liu Y, Ruan X, Ma J, Liu L, Wang D, Yang C, Cai H, Li Z, Feng Z, et al. FXR1 promotes the malignant biological behavior of glioma cells via stabilizing MIR17HG. *J Exp Clin Cancer Res* 2019; 38:37.

17. Ren K, Xu R, Huang J, Zhao J, Shi W. Knockdown of long non-coding RNA KCNQ1OT1 depressed chemoresistance to paclitaxel in lung adenocarcinoma. *Cancer Chemother Pharmacol* 2017; 80:243-250.

18. Sun X, Xin Y, Wang M, Li S, Miao S, Xuan Y, Wang Y, Lu T, Liu J, Jiao W. Overexpression of long non-coding RNA KCNQ1OT1 is related to good prognosis via inhibiting cell proliferation in non-small cell lung cancer. *Thorac Cancer* 2018; 9:523-531.

19. Feng W, Wang C, Liang C, Yang H, Chen D, Yu X, Zhao W, Geng D, Li S, Chen Z, Sun M. The dysregulated expression of KCNQ1OT1 and its interaction with downstream factors miR-145/CCNE2 in breast cancer cells. *Cell Physiol Biochem* 2018; 49:432-446.

20. Zhang L, Wang DL, Yu P. LncRNA H19 regulates the expression of its target gene HOXA10 in endometrial carcinoma through competing with miR-612. *Eur Rev Med Pharmacol Sci* 2018; 22:4820-4827.

21. Cao T, Jiang Y, Wang Z, Zhang N, Al-Hendy A, Mamillapalli R, Kallen AN, Kodaman P, Taylor HS, Li D, Huang Y. H19 lncRNA identified as a master regulator of genes that drive uterine leiomyomas. *Oncogene* 2019; 38:5356-5366.

22. Li X, Zhao Q, Qi J, Wang W, Zhang D, Li Z, Qin C. lncRNA Ftx promotes aerobic glycolysis and tumor progression through the PPARgamma pathway in hepatocellular carcinoma. *Int J Oncol* 2018; 53:551-566.

23. Liu Z, Dou C, Yao B, Xu M, Ding L, Wang Y, Jia Y, Li Q, Zhang H, Tu K, Song T, Liu Q. Ftx non coding RNA-derived miR-545 promotes cell proliferation by targeting RIG-I in hepatocellular carcinoma. *Oncotarget* 2016; 7:25350-25365.

24. Mitra R, Chen X, Greenawalt EJ, Maulik U, Jiang W, Zhao Z, Eischen CM. Decoding critical long non-coding RNA in ovarian cancer epithelial-to-mesenchymal transition. *Nat Commun* 2017; 8:1604.

25. Zhang H, Hua Y, Jiang Z, Yue J, Shi M, Zhen X, Zhang X, Yang L, Zhou R, Wu S. Cancer-associated fibroblast-promoted lncRNA DNM3OS confers radioresistance by regulating DNA damage response in esophageal squamous cell carcinoma. *Clin Cancer Res* 2019; 25:1989-2000.

26. Hewitt SC, Li L, Grimm SA, Chen Y, Liu L, Li Y, Bushel PR, Fargo D, Korach KS. Research resource: whole-genome estrogen receptor alpha binding in mouse uterine tissue revealed by ChIP-seq. *Mol Endocrinol* 2012; 26:887-898.