Cell-type phylogenetics and the origin of endometrial stromal cells

Kin, Koryu; Nnamani, Mauris C.; Lynch, Vincent J.; Michaelides, Elias; Wagner, Günter P.

Published in:
Cell Reports

DOI:
10.1016/j.celrep.2015.01.062

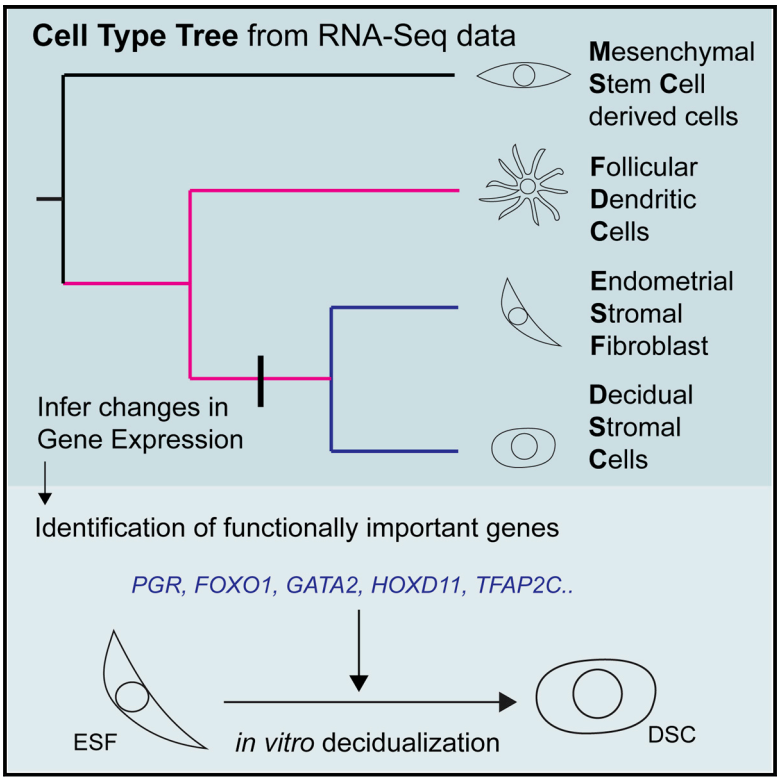
Publication date:
2015

Resource

Cell Reports

Cell-type Phylogenetics and the Origin of Endometrial Stromal Cells

Graphical Abstract



Authors

Koryu Kin, Mauris C. Nnamani, ..., Elias Michaelides, Günter P. Wagner

Correspondence

gunter.wagner@yale.edu

In Brief

Kin et al. apply a phylogenetic approach using RNA-seq data to infer the relationship of a uterine cell type, endometrial stromal cells (ESCs), to other related cell types. This approach led to the discovery of several genes that are essential for the function of ESCs, confirmed by RNAi gene knockdown.

Highlights

- Phylogenetic relationships of endometrial stromal cells (ESCs) are inferred
- An immune cell type, follicular dendritic cell, is found to be closely related to ESCs
- Gene-expression changes are reconstructed on the inferred cell-type tree
- Knockdown assays confirm the function of genes inferred to be important for ESCs

Accession Numbers

GSE63733

Kin et al., 2015, Cell Reports *10*, 1398–1409
March 3, 2015
http://dx.doi.org/10.1016/j.celrep.2015.01.062

OPEN ACCESS
CellPress

Cell Reports
Resource

Cell-type Phylogenetics and the Origin of Endometrial Stromal Cells

Koryu Kin,[1,2] Mauris C. Nnamani,[1,2] Vincent J. Lynch,[1,2,3] Elias Michaelides,[4] and Günter P. Wagner[1,2,*]
[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA
[2]Yale Systems Biology Institute, Yale University, New Haven, CT 06516, USA
[3]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[4]Department of Surgery, Yale University, New Haven, CT 06519, USA
*Correspondence: gunter.wagner@yale.edu
http://dx.doi.org/10.1016/j.celrep.2015.01.062

SUMMARY

A challenge of genome annotation is the identification of genes performing specific biological functions. Here, we propose a phylogenetic approach that utilizes RNA-seq data to infer the historical relationships among cell types and to trace the pattern of gene-expression changes on the tree. The hypothesis is that gene-expression changes coincidental with the origin of a cell type will be important for the function of the derived cell type. We apply this approach to the endometrial stromal cells (ESCs), which are critical for the initiation and maintenance of pregnancy. Our approach identified well-known regulators of ESCs, *PGR* and *FOXO1*, as well as genes not yet implicated in female fertility, including *GATA2* and *TFAP2C*. Knockdown analysis confirmed that they are essential for ESC differentiation. We conclude that phylogenetic analysis of cell transcriptomes is a powerful tool for discovery of genes performing cell-type-specific functions.

INTRODUCTION

One essential aspect of animal development is cellular differentiation. It is known that this process often proceeds in a hierarchical manner, where totipotent cells sequentially commit to fates of more-restricted developmental potential (Graf and Enver, 2009). Thus, the relationship of cell types in ontogeny is expected to form a tree-like structure, although it is also possible that the relationship among cell types can be better represented as networks of alternative developmental pathways.

A possible evolutionary explanation for the hierarchical relationships among cell types is the so-called "sister-cell-type model" proposed by Detlev Arendt (Arendt, 2008). According to this model, novel cell types arise as pairs (sister cell types) from an ancestral cell type by sub-specialization. If we assume this mode of cell type origination to be true, then the evolutionary relatedness of cell types is expected to be, at least initially, congruent with the ontogenetic hierarchy of cellular differentiation. This is so because, according to the model, the development of sister cell types is the same up to the last stages of differentiation (Arendt, 2008). In contrast, if new cell types frequently evolve by combining regulatory modules from distantly related cell types, there would be no expectation of a hierarchical set of relationships among cell types. In this paper, we call the hypothetical tree-like relationship of cell types in ontogeny and phylogeny a "cell-type tree" and reconstruct the cell-type tree from transcriptomic data.

Hierarchical developmental relationships among cell types have traditionally been elucidated through a series of laborious experiments involving in vitro differentiation of cell types from various stem cells (Bryder et al., 2006; Pronk et al., 2007; Villadsen et al., 2007). In recent years, with the advent of technologies to obtain genome-wide gene-expression data such as microarray or RNA-seq, attempts have been made to characterize the relationships among cell types using high-throughput transcriptomic information (Alizadeh et al., 2000; Novershtern et al., 2011; Sugino et al., 2006). With genome-wide gene-expression data and a phylogenetic hypothesis about the relationship among the cell types in hand, we are able to identify a series of gene-expression gain and loss events during the evolution of the cell types. These events can be reconstructed with standard methods of ancestral state reconstruction (Cunningham et al., 1998). Moreover, the inferred gene recruitment events (i.e., gain of gene expression) are expected to identify functionally important genes that are essential in the derived cell types. Here, we demonstrate that this approach is an effective way of discovering genes functionally relevant to a particular cell type. Our model system is the development and evolution of the human endometrial stromal cells, the endometrial stromal fibroblast and the decidual stromal cell, of the mammalian uterus.

Endometrial stromal fibroblasts (ESFs) are a cell type present in the uterus of eutherian mammals. In many species, they undergo a characteristic cellular transformation called decidualization, either spontaneously during the sexual cycle (Emera et al., 2012) or upon pregnancy, and become decidual stromal cells (DSCs) (Gellersen et al., 2007; Ramathal et al., 2010). Decidualization is essential for the successful implantation of embryos with invasive placentation as well as the maintenance of pregnancy. DSCs have various functional roles such as the regulation of trophoblast invasion, modulation of maternal immune and inflammatory reactions, and control of tissue remodeling of the endometrium (Gellersen et al., 2007; Gellersen and Brosens,

2014). DSC is known to be a derived cell type of placental mammals (Kin et al., 2014; Mess and Carter, 2006). In contrast, ESFs are present in the gray short-tailed opossum, *Monodelphis domestica*, a basal marsupial and thus an outgroup taxon to placental mammals (Kin et al., 2014). Therefore, ESFs are both the ontogenetic precursor of decidual cells as well as phylogenetically ancestral to decidual cells. Here, we focus on identifying a cell type that is closely related to endometrial stromal cells (ESCs) (a collective designation for ESFs and DSCs) in order to reconstruct gene-recruitment events involved in the origin of these cell types important in human and mammalian fertility.

ESCs are derived from mesenchymal stem cells (Aghajanova et al., 2010; García-Pacheco et al., 2001) or perivascular cells (Spitzer et al., 2012). Two other cell types, also derived from mesenchymal stem cells, have previously been proposed to be related to ESCs: myofibroblasts (Oliver et al., 1999) and follicular dendritic cells (Dunn et al., 2003; Muñoz-Fernández et al., 2006). To elucidate which of these two cell types are more closely related to ESCs, we collected RNA-seq data from human ESCs, myofibroblasts, and follicular dendritic cells, as well as two other cell types derived from mesenchymal stem cells, and performed a phylogenetic analysis of their transcriptomes. Finally, we used RNAi experiments to test genes inferred to be recruited in the origin of ESCs and found that the majority of them are essential for endometrial fibroblast differentiation.

RESULTS

RNA-Seq Data Reject Sister-Cell-type Relationship between ESCs and MFBs

Previous published work suggested two candidate cell types as closely related to ESCs: myofibroblast cells (Oliver et al., 1999) and follicular dendritic cells (Muñoz-Fernández et al., 2006). To assess which of the two cell types is more closely related to ESCs, we obtained RNA-seq data for six different mesenchymal cell types: chondrocytes (CHONs), myometrial cells (MYOs), myofibroblasts (MFBs), follicular dendritic cells (FDCs), ESFs, and DSCs. Five of them are previously established cell lines that were isolated and immortalized from normal human tissues (see Experimental Procedures). FDCs were isolated in our lab from human tonsils following a previously published protocol (Muñoz-Fernández et al., 2006). The identity of cells obtained from tonsils was confirmed by marker expression (Figure S1; Muñoz-Fernández et al., 2006). Using the Illumina RNA-seq technology, on average, 61.5 million (34–110 million) 35-bp sequence reads were obtained from mRNA isolated from each cell type. On average, 69% of sequence reads were mapped uniquely to known features (Table S1), and reads mapped to multiple locations or to locations with no known features were discarded from the following analyses. We also limited our analyses to protein-coding genes. We normalized the data by calculating *transcripts per million* (TPM) values from gene counts and transcript lengths (Wagner et al., 2012).

We first performed hierarchical clustering with bootstrap resampling on the RNA-seq data of six cell types. We took the square root of TPM values and used those values for the clustering. Pearson correlation coefficients between biological replicates were consistently above 0.96, indicating there is no obvious problematic sample (Figure 1A, heatmap). The result (Figure 1A) clearly rejects the hypothesis that MFBs are closely related to ESCs. MFB clusters with MYOs with 100% bootstrap support. This result is also reflected in the PCA, where ESFs and DSCs are clearly separated along PC1 from most mesenchymal-stem-cell-derived cell types with the exception of FDCs (Figure 1B). FDCs are close to ESCs on PC1 but separated from them on PC2. This arrangement is also reflected in the cluster analysis, where FDCs cluster with CHONs in a weakly supported cluster.

One potential problem with comparing complete transcriptomes is that the transcriptome similarities may reflect functional similarities, like the expression of contractile proteins in unrelated contractile cells, rather than evolutionary or developmental relatedness. Specifically, the finding that MFB is clustering with MYO could be due to the fact that both are contractile cells. To test whether the relationships revealed in Figure 1A reflect evolutionary/developmental relationships or functional similarity, we limited the data to transcription factor genes (TFs) (TF list taken from Ravasi et al., 2010). With only TFs, ESCs (DSC+ESF) clustered with FDCs with a high bootstrap support value (94%), whereas MFBs clustered again with MYOs (Figure 1C). On the PCA plot, FDCs moved slightly closer to ESCs on PC2 (Figure 1D). To test the robustness of these results, we repeated the analysis using two other lists of transcription factors (GO: 0003700 = sequence-specific DNA-binding transcription factor activity; Ravasi et al., 2010; Vaquerizas et al., 2009). These data gave essentially the same result (Figure S2). Overall, the clustering results clearly reject the hypothesis that MFBs are sister to ESCs, contra Oliver et al. (1999). The results further suggest that FDCs could be related to ESCs. To further evaluate the robustness of this result, we turned to phylogenetic methods, rather than clustering, and first explored the amount of tree structure in our data set.

Cell-type Transcriptome Data Have Significant Tree Structure

In order to apply phylogenetic methods, such as maximum parsimony, on the transcriptomic data, we first transformed quantitative expression data into qualitative (expressed/non-expressed) data. We operationally classified the genes as expressed if TPM > 3 and non-expressed if TPM < 3. This operational criterion is based on a model of transcript-abundance distribution previously developed (Hebenstreit et al., 2011; Wagner et al., 2013). The genomic distribution of H3K4me3, chromatin modification marks for active promoters, is also consistent with the classification. Genes classified as "ON" (mid-low: 3–44.8 TPM; high: >44.8 TPM) in DSCs had much stronger association with H3K4me3 marks compared to genes classified as "OFF" (Figure 2).

A convenient method of assessing the structure of distance data is the so-called δ statistic (Holland et al., 2002). The δ statistic is a measure of the "treeness" of distance data. The δ value varies between zero and one, where $\delta = 0$ indicates perfect tree structure and $\delta = 1$ indicates a perfect network without tree structure (Figure 3A). The δ statistic is calculated from a tetrad of cell types, and each tetrad has a unique δ value.

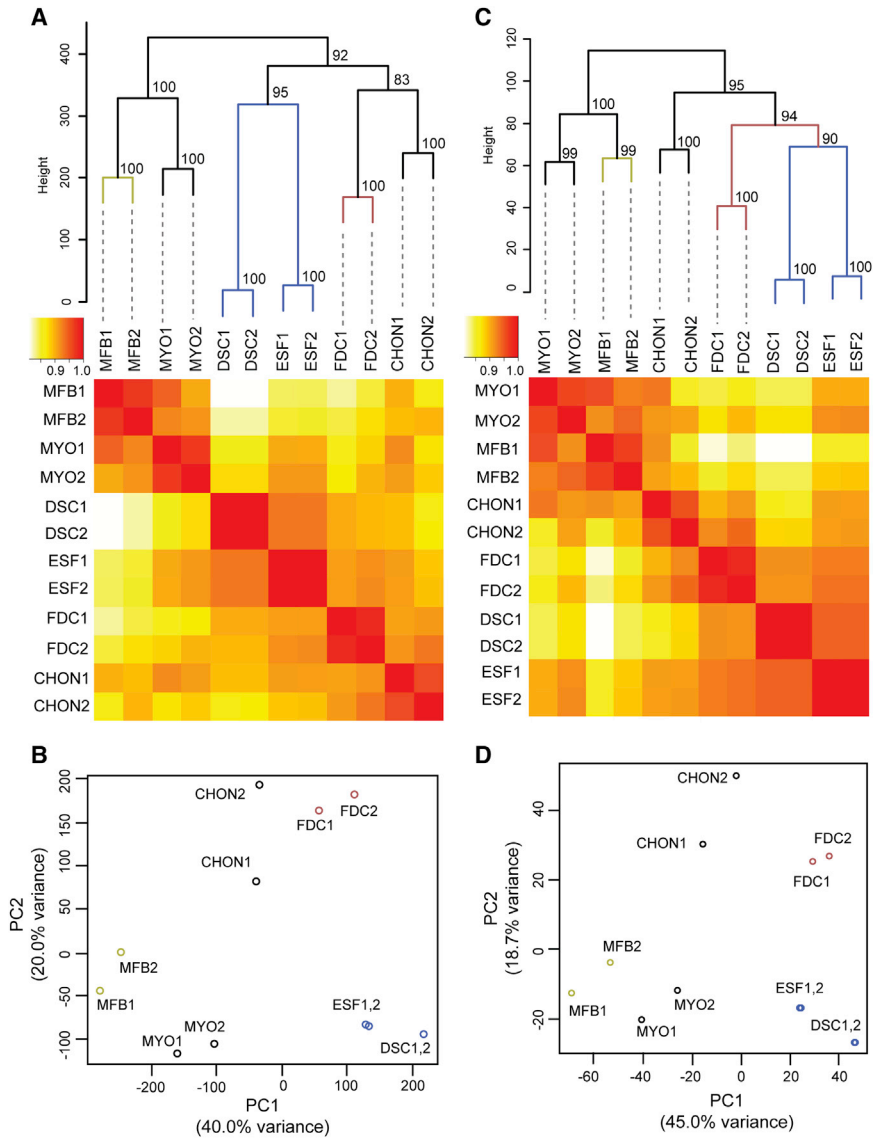


Figure 1. Hierarchical Clustering of RNA-Seq Data Refutes Sister-Cell-type Relationship between Endometrial Stromal Cells and Myofibroblasts

(A) Hierarchical clustering of RNA-seq data of six cell types using all protein-coding genes. The values on the nodes are bootstrap support values obtained by pvclust. Branches of ESFs/DSCs, FDCs, and MFBs are colored blue, red, and brown. The heatmap below the dendrogram indicates Pearson's correlations of square root TPM values among samples, with the color key at the top-left corner of the heatmap.

(B) Principal-component analysis of RNA-seq data of six cell types. Principal-component scores of RNA-seq data are plotted on PC1 and PC2. Circles representing ESFs/DSCs, FDCs, and MFBs are colored blue, red, and brown, respectively.

(C) Hierarchical clustering of RNA-seq data of six cell types using only transcription factor genes (Ravasi et al., 2010) with the heatmap. The values on the nodes are bootstrap support values. Branches are colored as in (A).

(D) Principal-component analysis of RNA-seq data of six cell types using only transcription factor genes. Principal-component scores of RNA-seq data are plotted on PC1 and PC2. Circles are colored as in (B).

See also Table S1 and Figure S2.

For our transcriptome data, we calculated the Hamming distance among all pairs of samples (n = 12, with two replicate samples for each cell type). We then calculated the δ value for each tetrad of samples (i.e., a total of 495 tetrads). The frequency distribution of δ has a mode at the smallest bin (0–0.025) and has a long tail extending to 0.95 (Figure 3B). The set of all tetrads was then filtered into two subsets. One is what we call the "replicate set," which contains all pairs of replicates for two cell types each. There are 15 such replicate tetrads. These replicate tetrads allow us to measure the non-treeness (amount by which δ is larger than 0) due to experimental noise. The other set includes all the tetrads that have only one replicate per cell type in each set of four samples.

The average δ value for pairs of replicate samples is 0.057 and indicates that, on average, technical noise does not contribute much to the δ values of compared cell types. The average δ value for the cell type set of tetrads excluding replicates is 0.36. Holland et al. (2002) suggest to calculate the δ_x value, i.e., the average δ value of all tetrads that include a particular cell type x. The comparison of δ_x values allows one to identify cell types that do not fit the tree structure. The δ_x values vary between 0.307 and 0.405 (Figure 3C), which is typical for tree-like data with unbalanced tree structure (Figure 3D; Holland et al., 2002). There is no cell type that has a considerably larger δ_x value than the others, as would be expected for operational taxonomic units (OTUs) (in this case cell types) that resulted from recombination (hybridization) of distantly related cells. In order to assess statistically whether our δ-value data are significantly less than 1 (i.e., has significant tree signal), we performed a jackknife procedure on log-transformed δ data with δ values larger than 0.3 (at smaller δ values, the jackknife process leads to artifacts). The distribution of p values as a function of δ values is given in Figure 3E. In our data, the minimal p values estimated started to rise considerably for $\delta > 0.7$ and are consistently larger than $p = 0.05$ for $\delta > 0.8$. We conclude that, in our data, only δ values below 0.7 can be significantly smaller than 1. To assess the overall tree structure of our data, we recorded the fraction of δ values less than 0.7. The cumulative δ value distribution for all cell types is shown in Figure 3F. We find that 84% of all the tetrads in our cell type set have a $\delta < 0.7$. The available evidence supports the assumption that the transcriptome data have tree structure comparable to simulated data with known tree structure (Holland

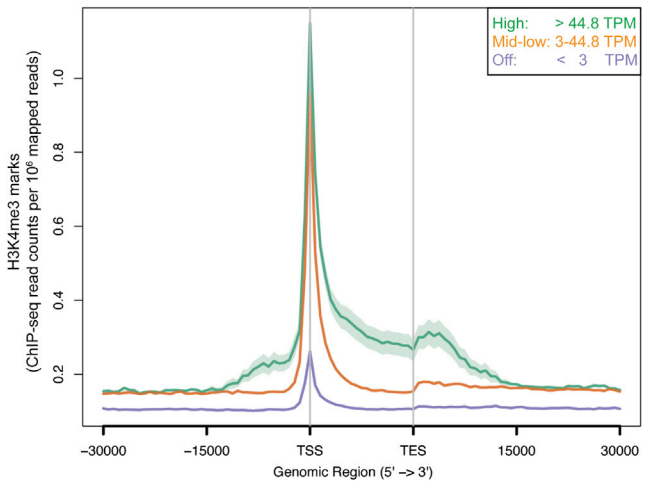


Figure 2. Genes below the Expression Threshold of 3 TPM Have Weak Epigenetic Promoter Mark, H3K4me3

Average ChIP-seq profiles for H3k4me3 read enrichment in DSCs are shown. The x axis represents genomic regions from 30 kb upstream of transcriptional start sites (TSSs) to 30 kb downstream of transcriptional end sites (TES), and the y axis represents the level of H3K4me3 marks. Reads were filtered by expression levels (TPM values) determined from transcriptomic data of DSCs. First two classifications were made (ON and OFF genes) as described in Experimental Procedures. Specifically, off genes were determined as genes with less than 3 TPM, whereas ON (expressed) genes, >3 TPM. All expressed genes were further divided into two classes: medium–lowly expressed and highly expressed genes. The range of TPM values for the three classifications were OFF genes (<3 TPM), mid-low expression (3 TPM–44.8 TPM), and high expression levels (>44.8 TPM). These data suggest that the operational criterion for non-expressed genes of <3TPM is statistically associated with low H3K4me3 signal.

et al., 2002). This approach thus fails to reject the prediction of the sister-cell-type model that the transcriptome data have significant tree structure.

ESCs Are Related to FDCs

Maximum parsimony tree reconstruction was performed on the transcriptomic data, transformed into expressed and non-expressed as described above, with bootstrap resampling to obtain support values on each node. The reconstructed tree was generally well supported (Figure 4A), and the topology of the tree is identical to that of the hierarchical clustering dendrogram obtained using only TF genes. Again, MFB was clearly separated from the ESCs. The tree also supports a relationship between ESCs (ESC = ESF + DSC) and FDCs with a moderate bootstrap value (88.5%).

ESCs Evolved Role in Cell-Cell Signaling and Leukocyte Immunity

The inferred relationship between ESCs and FDCs imply a history of gene activation and suppression during the evolutionary or ontogenetic differentiation of these cell types. An elementary inference suggests that the genes inferred to have been acquired during the cell differentiation process are likely to be important to the derived function of these cells. We first explored this implication with a gene ontology analysis.

We performed maximum parsimony ancestral character state reconstruction to infer the gene-expression changes associated with the evolution of ESCs. The number of gene-expression changes unambiguously reconstructed on each branch of the cell type tree is shown in Figure 4B. We paid particular attention to genes whose gene expression state changes on three branches related to the clade of FDCs, ESFs, and DSCs: (1) the branch to the clade formed by FDCs, ESFs, and DSCs; (2) the branch to the clade formed by ESFs and DSCs; and (3) the branch to DSCs. The overall result is summarized in Figure 4B, and the lists of all gene-expression changes, including the ones reconstructed only by ACCTRAN (ACC) or DELTRAN (DEL) reconstruction algorithm, found on these branches are shown in Tables S2, S3, and S4.

GO term-enrichment analyses on the genes inferred to have been recruited at the branch uniting FDCs and ESCs, branch *a* in Figure 4B, share genes related to cell migration (GO: 0030334, 2000145, and 0051270). Genes overrepresented in the lineage leading to ESCs (branch *b*) are enriched for genes with developmental functions, cell-cell signaling, and leukocyte immunity (GO: 0009653, 0007267, and 0002443). This list of genes is also enriched for genes with reproductive defects in knockout mice (Table S5). The lineage of DSCs (branch *c*) has genes recruited that are involved in hormone metabolism, gonadal development, and regulation of developmental processes (GO: 0042445, 0008406, and 0051094). The comprehensive lists of enriched GO terms can be found in Tables S6, S7, and S8.

TFs Recruited to ESCs Are Necessary for Decidualization

In Table 1, we list the top 15 (in terms of gene expression) out of 28 TFs inferred to be recruited during the evolution of ESCs. The list of recruited genes includes known regulators of decidualization, *FOXO1*, *HOXA11*, and *PGR* (progesterone receptor), as well as genes that have not been implicated in ESC biology. We performed RNAi-mediated gene knockdown of these 15 TF genes in cultured human endometrial cells. The RNAi reduced the expression levels of target genes by 40%–98% (Figure S3). Among the TFs tested, *HOXA11* and *HOXD8* gene expression was not consistently knocked down with the siRNA we used, so they were removed from further analyses (data not shown). To assess knockdown effects on decidualization, we measured RNA expression of two decidual marker genes (*PRL* and *IGFBP1*). As a reference, we also knocked down genes that were recruited on other branches of the cell-type tree. Specifically, *EMX2* and *FOXF1* were recruited on the branch leading to the FDC-ESC clade (branch *a* in Figure 4) and *SOX6* and *IKZF2* were recruited on the branch leading to the CHON-(ESC-FDC) clade (Table 1). For some TFs, especially *HOXD9*, *HOXD10*, and *SALL1*, the knockdown had measurable effects, but the outcome was variable among replicates for unknown reasons. Nevertheless, knocking down 8 out of 13 ESC-recruited TFs consistently decreased *PRL* expression whereas only one out of four non-ESC-recruited TFs showed consistent *PRL* decrease upon knockdown (Figure 5A). Similarly, knocking down 5 out of 13 ESC-recruited TFs consistently decreased *IGFBP1* expression, whereas one out of four non-ESC-recruited genes consistently

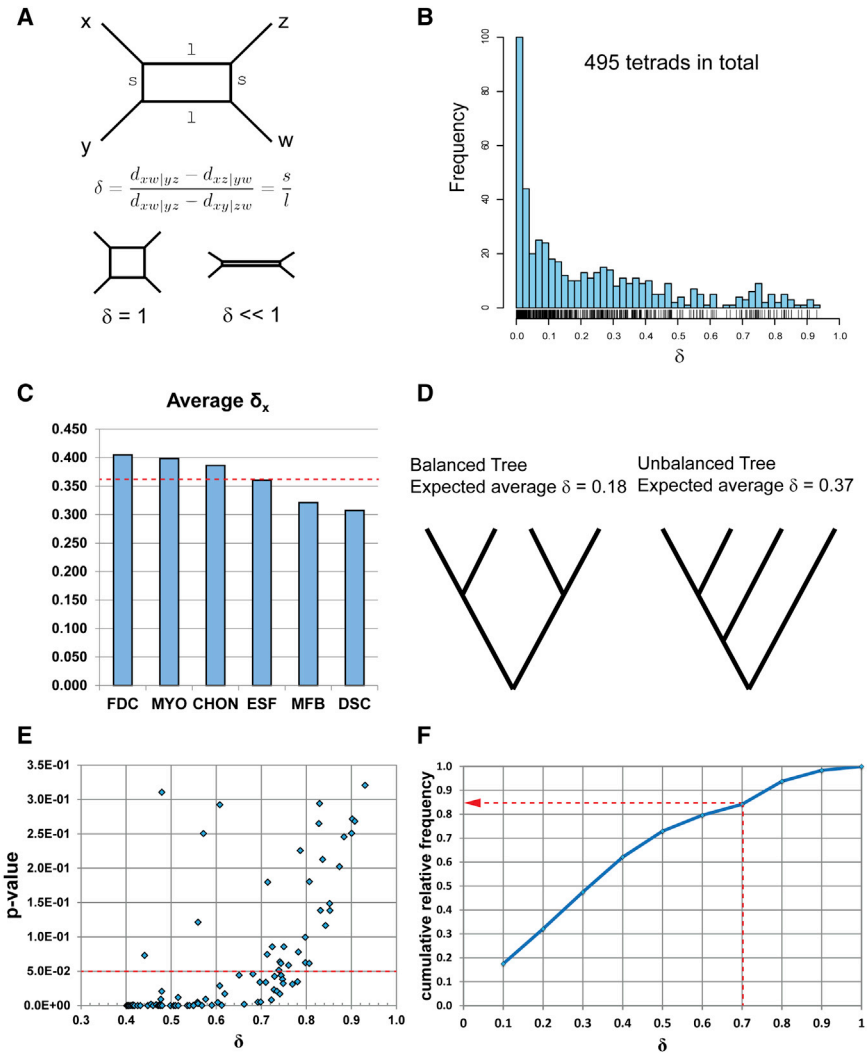


Figure 3. RNA-Seq Data Contain Significant Tree Structure

(A) Schematic for explaining the treeness metric, δ. When δ = 1, the relationship among the tetrad should be represented as a network. When delta is significantly smaller than 1, the relationship among the tetrad can be regarded as more tree-like.

(B) The frequency distribution of δ for all 495 tetrads with the "rugplot" showing where actual δ values fall on the x axis.

(C) The delta plot for the cell types used in this study. The average δ value among all cell types excluding replicates (0.36) is shown by the red dashed line. This distribution of delta values is expected for trees with unbalanced tree structure.

(D) Schematic showing balanced and unbalanced tree topology. Theoretically expected average delta value is higher for unbalanced trees (0.37) than for balanced ones (0.18).

(E) A plot of delta values and p values as calculated by the jackknife method. The dashed line shows the threshold of p = 0.05.

(F) Cumulative delta plot of all cell types. It is shown that 84% of tetrads falls under delta = 0.7 as indicated by the red dashed arrow.

decreased *IGFBP1* upon knockdown (Figure 5B). This difference may be due to higher-average gene-expression levels for ESC-recruited genes (22.16 TPM compared to 8.72 TPM for non-ESC genes), so we selected bottom seven ESC-recruited TFs from the ranked list so that the average TPM value is approximately equal to non-ESC TFs (8.26 versus 8.72) and created boxplots of *PRL* (Figure 5C) and *IGFBP1* (Figure 5D) expression levels relative to negative control upon gene knockdown. The plots show that the decrease of marker gene expression tends to be larger for ESC-recruited TFs than for genes recruited earlier in cell evolution.

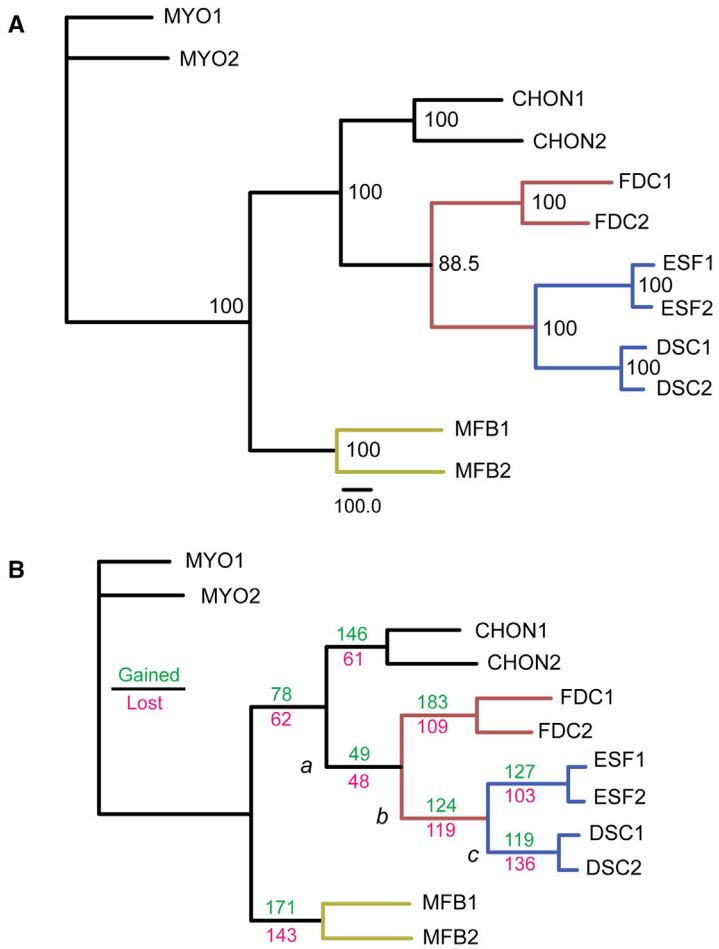
FOXO1 and *PGR* Are Hubs of Decidual Gene Regulation

We further investigated the regulatory relationships among the eight ESC-recruited transcription factors by examining the effects of RNAi knockdown on other TFs (Figure S4). The results are summarized in the heatmap in Figure 5E. There are some interesting features in this heatmap, which include (1) *FOXO1* is activated by almost all other TFs tested; (2) *PGR* knockdown significantly reduced the expression levels of *HOXD11*, *FOXO1*, and *GATA2*, suggesting the involvement of these TFs downstream of the progesterone-signaling pathway. *PGR* is suppressed by most other TFs; (3) *TFAP2C* is the only TF whose knockdown reduced the expression level of *PGR*. Its knockdown also reduced the expression of *FOXO1* and *GATA2*, and these genes in turn suppress the expression of *TFAP2C*. These results suggest a multi-layered co-regulatory network in DSCs (Figure 5F). The top layer genes *PGR* and *TFAP2C* actively regulate the bottom layer genes *GATA2*, *FOXO1*, and *HOXD11*, whereas the bottom layer genes suppress the top layer genes and thus form negative-feedback loops. *FOXO1* acts as a network hub and receives inputs from all the other genes in the network.

DISCUSSION

Interpretation of Cell-type Trees and the Origin of Decidual Cells

We interpret the reconstruction of the cell-type tree (Figure 4) as a hypothesis about the evolutionary history of cell-type origination (Arendt 2008). Each branching point on the cell-type tree implies an inferred novelty where a new pair of sister cell types originated from an ancestral cell type. This interpretation is analogous to that of gene trees among paralog genes and also so-called "character trees," which reconstruct evolutionary differentiation of repeated organs as, for instance, in the evolution of eye types in arthropods (Oakley et al., 2007). In our reconstruction, we infer two well-supported cell-type-origination



Recruited Branch	GO Term	Description	P-value	Enrichment
a (ESC-FDC)	GO:0030334	regulation of cell migration	5.04E-05	5.92
	GO:2000145	regulation of cell motility	7.17E-05	5.63
	GO:0051270	regulation of cellular component movement	1.15E-04	5.26
b (ESC)	GO:0009653	anatomical structure morphogenesis	7.95E-07	3.03
	GO:0007267	cell-cell signaling	9.96E-06	3.56
	GO:0002443	leukocyte mediated immunity	9.59E-05	10.90
c (DSC)	GO:0042445	hormone metabolic process	7.01E-06	8.05
	GO:0008406	gonad development	3.02E-05	10.02
	GO:0051094	positive regulation of developmental process	1.30E-04	2.76

Figure 4. Cell-type Tree Reconstructed by Maximum Parsimony and Inferred Changes in Gene Expression

(A) Cell-type tree as reconstructed by maximum parsimony. Values on the nodes are bootstrap values (% of instances in which the node appeared in 1,000 bootstrap replicates). The scale bar corresponds to 100 changes.

(B) The numbers of unambiguous gene expression changes on each branch of the cell-type tree. The number of gene expression gained is indicated in green above each branch, and the number of gene expression lost is indicated in pink below each branch. The numbers shown here represent conservative estimates and only include genes that were unambiguously reconstructed through maximum parsimony reconstruction. Three branches of particular interest are labeled as *a* (the branch uniting the ESC-FDC clade), *b* (the branch uniting the ESC clade), and *c* (the branch uniting the DSC clade). The table below shows three representative GO terms (ranked by p values) enriched in the list of genes recruited (gene expression gained) on the labeled branches.

See also Tables S2, S3, S4, S5, S6, S7, and S8 for specific list of genes recruited on the three branches and the lists of enriched GO terms and KO terms.

events. One separates FDCs and ESCs (ESC = ESF + DSC), and the second separates ESFs and DSCs. There is no support for a close relationship of ESCs and MFB cells.

The close relationship between FDCs and ESCs does not prove that FDC is in fact the sister cell type to ESCs, i.e., the most closely related cell type, because we could not exhaustively sample all human cell types. Nevertheless, the result suggests that ESCs can be considered as a specialized immune regulatory cell type, a role consistent with their function of providing an immune tolerant environment for the allogenic fetus (Erlebacher, 2013; Haig, 1993). This interpretation is also supported by the gene ontology analysis of the genes inferred to be recruited into ESCs. This set is enriched for genes involved in cell-cell signaling and the regulation of leukocyte immunity (Figure 4B).

Because the present analysis was only done with cells from one species, human, it is not possible to associate the inferred cell-type-origination events with particular lineages in mammalian history. But comparison with data from other species can constrain the phylogenetic timing of these events. For the split between ESFs and DSCs, the consensus view is that DSCs originated prior the radiation of placental mammals and after the most-recent common ancestor of therians, i.e., in the stem lineage of placental mammals (Kin et al., 2014; Mess and Carter, 2006). DSCs have only been described from placental mammals (Mossman, 1987) and are absent from the marsupial *Monodelphis domestica* (Kin et al., 2014). Reports about the endometrium from other marsupials show that there is no direct interface between the trophoblast and the endometrial stroma in any marsupial (reviewed in Wagner et al., 2014). The decidual cell is a shared derived (synapomorphic) character of placental mammals.

Given that *Monodelphis* has both ESFs (Kin et al., 2014) as well as FDCs (K.K., unpublished data), the split between FDCs and ESCs likely happened prior to the most-recent common ancestor of therians (before the lineage split between marsupial and placental mammals). But at this point, it is impossible to more precisely identify the time in phylogeny when this event happened.

Table 1. Average Gene-Expression Levels of TFs in ESFs-DSCs Selected for RNAi Assay

Recruitment Type	Ensembl Gene ID	Gene Name	Avr. ESF-DSC TPM
ESC	ENSG00000088881	*EBF4*	57.84
ESC	ENSG00000128710	*HOXD10*	56.64
ESC	ENSG00000103449	*SALL1*	38.67
ESC	ENSG00000150907	*FOXO1*	33.82
ESC	ENSG00000082175	*PGR*	27.26
ESC	ENSG00000005073	*HOXA11*	24.49
ESC	ENSG00000128709	*HOXD9*	16.03
ESC	ENSG00000179348	*GATA2*	12.83
ESC	ENSG00000173917	*HOXB2*	9.34
ESC	ENSG00000128713	*HOXD11*	7.66
ESC	ENSG00000175879	*HOXD8*	7.50
ESC	ENSG00000134532	*SOX5*	7.41
ESC	ENSG00000087510	*TFAP2C*	7.41
ESC	ENSG00000153234	*NR4A2*	6.83
ESC	ENSG00000198945	*L3MBTL3*	6.31
ESC-FDC	ENSG00000103241	*FOXF1*	11.92
ESC-FDC	ENSG00000170370	*EMX2*	10.00
CHON-(ESC-FDC)	ENSG00000110693	*SOX6*	8.96
CHON-(ESC-FDC)	ENSG00000030419	*IKZF2*	4.00

Tree Structure in the RNA-Seq Data of Cell Types

One of our goals in the present study was to assess whether the relationship of cell types should be represented as a tree or a network. The assumption of treeness is true for gene trees as long as there is no recombination among genes, but when recombination occurs, networks become the best representation of the relationship rather than trees. Whether or not we can treat the relationships among cell-type transcriptomes as trees is an issue that was unexplored until recently. It is true that, in some cases, it is known that sequential expression of transcription factors creates a tree-like hierarchical differentiation pattern during the process of cell differentiation (see Graf and Enver, 2009 for a review). However, it is also possible that large-scale recruitment of gene-regulatory modules occurs frequently, which would make the relationship among transcriptomes more network-like than tree-like. We explored this issue by applying the technique called δ-plot, which was originally developed to assess the treeness for phylogenetic analyses (Holland et al., 2002). We found no evidence of significant "recombination" events among the cell types we used in the present study and concluded that the RNA-seq data of the six cell types contain tree structure. Whether or not this conclusion holds true for other groups of cell types is an open question that warrants future studies. Interestingly, a few recent studies approached this issue using public data sets generated from large-scale sequencing projects such as ENCODE or FANTOM (Liang et al., 2015; Nair et al., 2014). In Nair et al. (2014), the authors used ChIP-seq histone modification data from ENCODE to reconstruct cell-type trees under the assumptions very similar to ours, although they did not explicitly test whether reconstructing cell-type trees can be justified with their data. In Liang et al. (2015), the authors developed a statistical model for calculating probability distributions of δ and applied the model to ENCODE and FANTOM RNA-seq data. They found, similar to us, that the RNA-seq data contain significant tree structures. The fact that different kinds (ChIP-seq data in Nair et al., 2014 as opposed to RNA-seq data in Liang et al., 2015 and this study) and scales of data support the tree-like relationship of cell types implies wide applicability of the cell-type tree model.

Methodological Considerations for Reconstructing Cell-type Trees

An interesting finding was that the inferred relationship of cell types differed when using different subsets of genes for clustering. When using all protein-coding genes for clustering, FDCs and CHONs clustered, whereas FDCs clustered with ESFs and DSCs when using only TFs. The discrepancy is not unexpected given that there are broadly two classes of genes contributing to cellular phenotypes: "realizer" or "effector" genes and regulatory genes (Erwin and Davidson, 2009; Graf and Enver, 2009; Wagner, 2007). The former are represented by enzymes, cytoskeletal genes, extracellular matrix protein genes, etc., and are directly responsible for the physiological phenotype of the cell. In contrast, the latter is represented by transcription factors and co-factors, and its effect on the cellular phenotype is mediated through realizer genes they are regulating. Thus, regulatory genes are more indirect and "abstract" in their relation to the function of cells. Functional significance of regulatory genes can, in principle, change by changing which realizer genes they regulate through modification of cis-regulatory elements of their target genes. On the other hand, the expression patterns of realizer genes are more likely to show convergent similarity due to shared function rather than shared developmental or evolutionary history. An example to illustrate this point is the myoepithelial cells in the mammalian breast glands. They express a similar set of contractile proteins as smooth muscle cells and serve as contractile cells during lactation. In spite of their functional resemblance to smooth muscle cells, myoepithelial cells differ from smooth muscle cells by the lack of transcription factors such as myocardin and others (Li et al., 2006) and are derived mammary gland epithelial cells.

In the human genome, TFs constitute less than 10% of all protein-coding genes (Vaquerizas et al., 2009) and the vast majority of genes can be considered as realizer genes. Besides, the dynamic range of realizer genes is generally much larger than that of TFs (in our data, the maximum TPM value for TF genes was 4,159.13 as opposed to the maximum TPM value of 57,000.16 for all protein-coding genes). Therefore, in the transcriptomic analysis, signals from TFs can be easily overridden by those from realizer genes, and the results thus are influenced by functional similarity rather than historical relatedness. For these reasons, we regard the clustering results obtained from TF-gene data as a better representation of the historical relationships of cell types. This interpretation is confirmed by the results from our maximum parsimony analysis, which relies on the pattern of shared derived gene expression (synapomorphy) rather than overall similarity, given the same topology as hierarchical clustering using only TF genes. We thus conclude that

FDCs are the most closely related cell types to ESCs among the cell types compared here.

Cell-type Phylogenetics Reliably Identify Decidual Regulatory Genes

One of the advantages of using maximum parsimony to compare transcriptomes is that it allows inferences about changes in gene expression associated with inferred cell-type origination events. Using the reconstructed cell-type tree, we inferred a list of genes gained or lost on the branch leading to ESCs (= ESF and DSC). The hypothesis is that the list of genes gained should be enriched for genes necessary to perform the derived function of the ESCs. Consistent with this hypothesis, the gene set is enriched for mouse knockout phenotypes with reproductive defects (Table S8). Also, we directly tested this hypothesis by knockdown experiments of a sample of those TFs and monitored the expression of molecular markers for decidualization, *PRL* and *IGFBP1*. The expression of these decidualization markers decreased in 8 out of 13 cases for *PRL* and 5 out of 13 cases in *IGFBP1*. Five TFs, *PGR*, *FOXO1*, *GATA2*, *TFAP2C*, and *HOXD11*, showed a consistent decrease in both *PRL* and *IGFBP1* expression upon knockdown. The roles of *PGR* and *FOXO1* in decidualization have been well studied (Gellersen and Brosens, 2014). *GATA2* is known to be expressed in murine DSCs (Rubel et al., 2012), and its downregulation through hypermethylation has been recently linked to endometriosis (Dyson et al., 2014). *HOXD11* is also known to be expressed in ESCs, but its functional role in decidualization was unknown until recently, when Raines et al. (2013) generated triple knockout mice of *HOXD9*, *-10*, and *-11*. The *HOXD9*, *-10*, and *-11* mutant mice are infertile and display significantly reduced stromal components in the uterus, although individual knockout of *HOXD9*, *-10*, or *-11* did not result in any phenotypic defects in reproductive tract. All these HOXD genes were also identified as potential decidual genes through our phylogenetic analysis. The functional redundancy among HOXD genes may also contribute to the variable results of knockdown for *HOXD9* and *10* genes. *TFAP2C* is known to be important for trophectoderm development (Kuckenberg et al., 2012), but no role in decidualization has been documented. It is interesting that, in this study, *TFAP2C* was found to be not regulated by *PGR*. Given that the activity of AP2 is modulated by cAMP/PKA signaling (García et al., 1999), it is possible that *TFAP2C* acts as a mediator of the cAMP-signaling pathway in decidualization. These results not only show that decidual regulatory genes can be discovered from cell-type phylogenetic analysis but also imply that the sister-cell-type model is biologically meaningful.

A History of Cellular Innovations

In Figure 6, we summarize the broad biological implications of our findings. The cell-type family consisting of FDCs, ESFs, and DSCs is characterized by the acquisition of genes regulating and contributing to cell migration. ESCs likely originated through acquisition of progesterone responsiveness and changes in cell-cell signaling and the regulation of leukocyte-mediated immunity. Finally, decidual cells derived through the acquisition of genes involved in gonad development and hormone metabolism.

The implication of this study is not limited to the field of endometrial biology. What we have shown here is that phylogenetic analysis of cell-type relationships can be an effective discovery tool for genes with cell-type-specific functions. As the amount of RNA-seq data from different cell types rapidly increases, we have to find a good way to represent, organize, and extract information from such data. Given that the cell types are the products of both developmental and evolutionary histories, we think that the phylogenetic method, which has been developed to infer historical relationships, has a great potential to become an important discovery tool for cell and developmental biologists.

EXPERIMENTAL PROCEDURES

Transcriptome Data Acquisition

Human ESFs (ATCC; cat. no. CRL-4003), CHONs (ATCC; cat. no. CRL-2847), and MFBs (ATCC; cat. no. CRL-2854) were purchased from American Type Culture Collection (ATCC). MYOs were obtained from Urogynecology Research Laboratory at University of Texas Southwestern Medical Center. Each type of cell was cultured following the instructions of the suppliers. Specifically, ESFs were grown in DMEM supplemented with 10% charcoal-stripped calf-serum (Hyclone) and 1% antibiotic/antimycotic (ABAM; GIBCO), CHONs were cultured in DMEM with 0.1 mg/ml G-418 and 10% FBS, MFBs were cultured in DMEM with 1% ABAM and 5% FBS, and MYOs were cultured in DMEM/F12 with 1% ABAM.

Human FDCs were freshly isolated from human tonsils obtained through routine tonsillectomy following the protocol described in Muñoz-Fernández et al. (2006). De-identified fresh tonsillar tissue was obtained after routine tonsillectomy from the pathology department, after routine gross inspection. Use of this tissue was approved by the Yale Human Investigation Committee, protocol no. 1007007149. The tonsils were thoroughly washed in PBS solution and cut into small pieces and finely minced in a small volume of RPMI 1640 medium with 1× ABAM. The suspension was put in a solution of 0.25% trypsin and 0.5 mM EDTA for 15 min at 37°C, and the reaction was stopped by adding cold RPMI 1640 with 20% FCS. The suspension was filtered through 70-μm nylon cell strainer (BD Falcon) and centrifuged at 425 × *g* for 10 min. The supernatant was discarded, and the cell pellet was suspended in RPMI 1640 and centrifuged on Ficoll-Paque (Pharmacia Biotech) for 20 min at 600 × *g*. Cells were collected from the interface, suspended in PBS, and washed. This suspension was incubated in culture flasks for 1 hr at 37°C in complete RPMI 1640 with 10% FCS to allow macrophages and granulocytes to adhere to the flask. The supernatant cells were washed and incubated in fibroblast medium with 1× ABAM. After overnight incubation to allow adherent cells to attach to the flask, lymphocytes in the supernatant were discarded. Fibroblast medium was then replaced and changed twice a week. The identity of cells isolated from tonsils was confirmed by RT-PCR of total RNA isolated from cultured cells with the primer sets described in Muñoz-Fernández et al. (2006) except for those for *CD13* (*ANPEP*) and *CD21* (*CR2*). For amplifying *CD13* and *CD21*, we used the following forward and reverse primers: *CD13* forward = AACCTCATCCAGGCAGTGAC; *CD13* reverse = GCCTGGGTCAT CAGGAACTA; *CD21* forward = ACACATGAGGGAACCTGGAG; and *CD21* reverse: AGTGAACGGGATCTGCAAAC. See also Figure S1.

To induce decidualization in human ESCs, the cells were treated with 0.5 mM 8-Br-cAMP (Sigma) and 0.5 μM of the progesterone analog medroxyprogesterone acetate (MPA) for 48 hr in DMEM supplemented with 2% charcoal-stripped calf-serum. Total RNA was extracted using the RNeasy Midi RNA-extraction kit (QIAGEN) followed by on-column DNase I treatment. Total RNA quality was assayed with a Bioanalyzer 2100 (Agilent) and found to be of excellent quality. Aliquots from the total RNA samples were sequenced using the Illumina Genome Analyzer II platform, following the protocol suggested by Illumina for sequencing of cDNA samples. Sequence reads were mapped to the human (GRCh37.69) cDNA builds at Ensembl with TopHat2 (Trapnell et al., 2009); two mismatches were allowed and reads aligning to more than one cDNA were discarded.

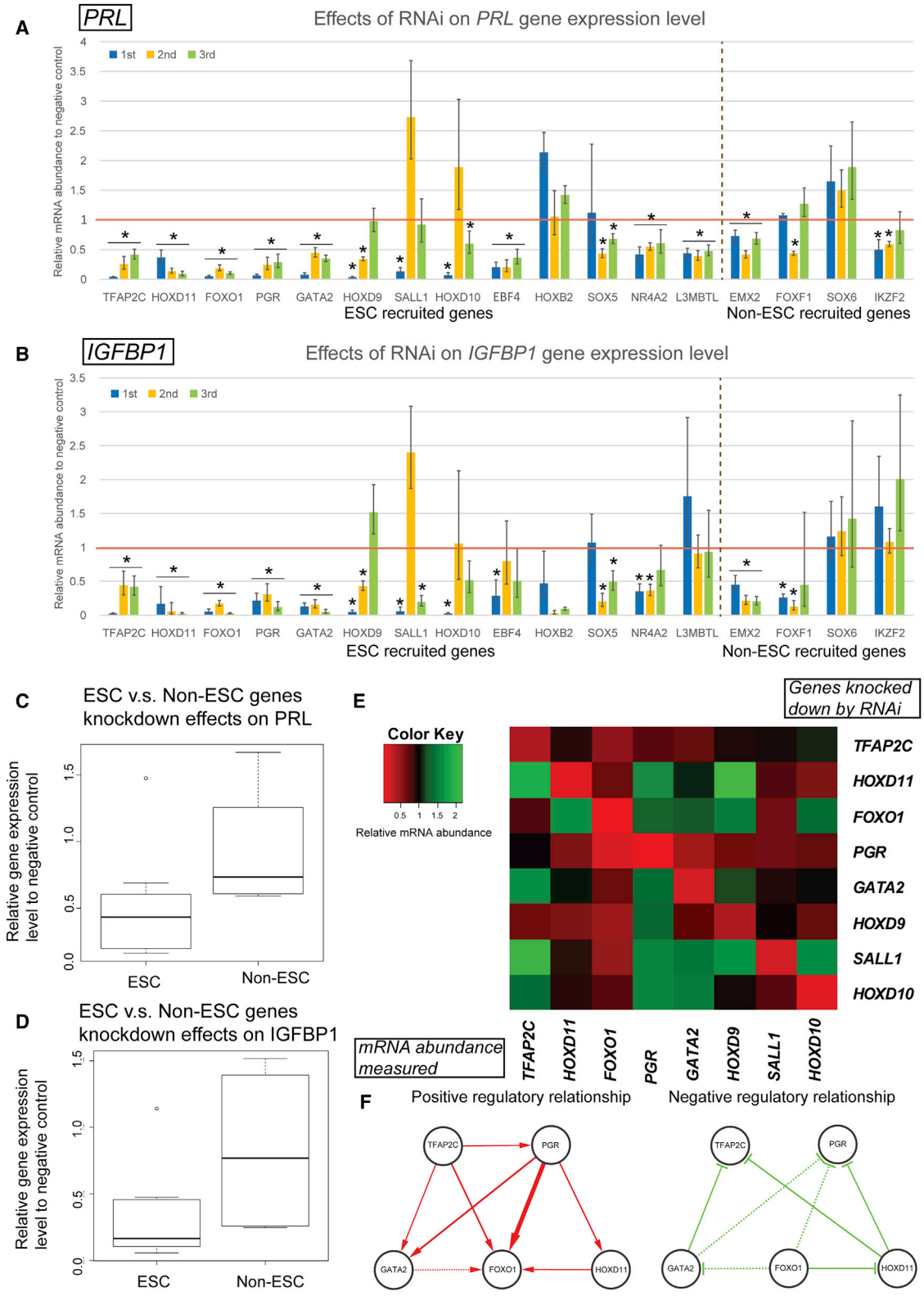


Figure 5. RNAi Knockdown of Several ESC-Recruited Genes Affect In Vitro Decidualization

(A and B) Effects of gene knockdown of 17 transcription factors on *PRL* and *IGFBP1* gene-expression level, respectively. The mRNA abundances in siRNA-introduced cells relative to negative control in three independent sets of experiment are shown separately as first, second, and third. The vertical dashed line

(legend continued on next page)

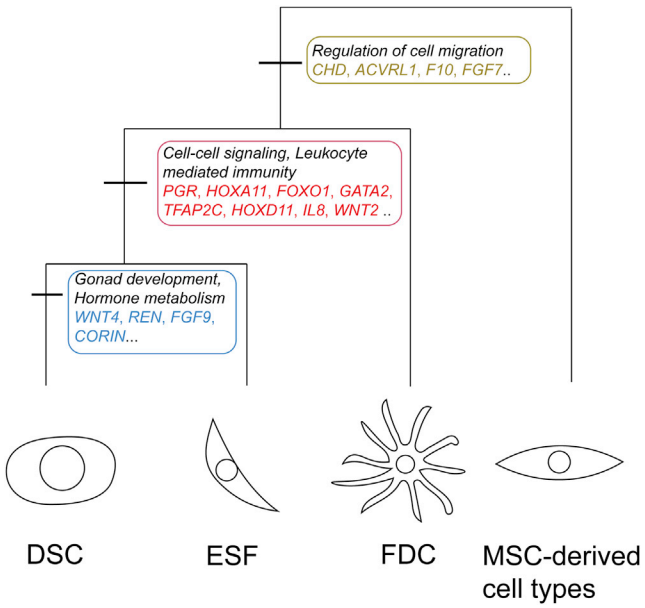


Figure 6. A Schematic Model Summarizing the Findings of the Present Study

The gene-expression reconstructions on the cell-type tree suggest the history of cellular innovations in the evolution of these cell-type families. Most notable is the acquisition of leukocyte regulatory activity with the origin of endometrial stromal cells and the acquisition of genes related to hormone metabolism with the origin of decidual cells.

Transcriptomic Data Analysis

The TPM values (Wagner et al., 2012) were calculated for data normalization and were used in the following analyses. The transcript length information was obtained from the Ensemble database with Biomart. When a gene has multiple transcripts, a median length of all transcripts for the gene was used. Hierarchical clustering with bootstrap analyses was done using the pvclust package (Suzuki and Shimodaira, 2006) in R (R Development Core Team, 2012). Principal-component analysis was also performed with R. For the purpose of phylogenetic analyses, TPM values were transformed into binary values (1 or 0), representing presence or absence of gene expression. Specifically, genes with TPM values above 3 were called present and those with TPM values less than 3 were called absent based on our previous finding (Wagner et al., 2013). The resulting data matrices were used for calculating distances for tree-likeness test. For the tree-likeness test, four sets of RNA-seq data, which are collectively called "tetrad," were chosen out of the 12 RNA-seq data sets, and pairwise hamming distance was calculated for all pairs in the tetrads. Assuming that the four-point condition is met (Holland et al., 2002), the network representing the relationship of the tetrad can be uniquely derived. We calculate the value, δ, which is a proxy for the tree-likeness. If δ equals 1, the network does not have tree structure at all, whereas if δ equals zero, the network should be represented as a tree. We estimated the probability of δ being significantly smaller than 1 by using jackknife statistics. Maximum parsimony phylogenetic reconstruction of cell-type tree was performed with PAUP* 4.0b10 (Swofford, 2003). Maximum parsimony ancestral reconstruction was performed with the reconstructed tree, setting the MYO type as an outgroup, using PAUP* 4.0b10.

Gene Function Annotation

GO term enrichment analyses were performed with a web-based tool GOrilla (Eden et al., 2009). Genes recruited into endometrial expression were annotated based on their mouse knockout phenotypes using data available at the Mouse Gene Informatics (MGI) database. Enrichments, p values (hypergeometric), and FDR q values were calculated using VLAD (http://proto.informatics.jax.org/prototypes/vlad/).

Processing of H3K4me3 ChIP-Seq Data

Sequence reads were aligned to the human reference genome (hg19) using the ultra-fast short DNA sequence aligner Bowtie (Langmead and Salzberg, 2012; Langmead et al., 2009). Sequencing depth for ChIP-seq samples and input averaged 34.5 million and 32 million reads, respectively, per biological sample with >76% overall uniquely aligned reads. Only uniquely aligned reads were used for further analysis. Visualization of reads at functional genomic regions was obtained by ngs.plot, a genomic database-integrating software, following author's recommendations (Shen et al., 2014).

siRNA Knockdown

siRNAs for *TFAP2C* (MU-005238-00), *HOXA11* (MU-012108-01), *HOXD9* (MU-012494-00), *HOXD10* (MU-011696-01), *HOXD11* (MU-013095-00), *GATA2* (MU-009024-00), *PGR* (MU-003433-01), *SALL1* (MU-006560-01), and *FOXO1* (MU-003006-03) were purchased from GE Healthcare (siGENOME; Dharmacon), and siRNAs for *EBF4* (s226921), *HOXB2* (s6792), *HOXD8* (s6852), *SOX5* (s13303), *NR4A2* (s9787), *L3MBTL* (s39037), *EMX2* (s4668), *FOXF1* (s5221), *SOX6* (s30968), and *IKZF2* (s22420) were purchased from Life Technologies (Silencer Select; Ambion). As negative control, we used ON-TARGETplus Non-targeting Pool siRNAs (D-001810-10-05; Dharmacon), which are supposed to have minimal targets in the human transcriptome. We followed a protocol developed for human ESCs by Yale Molecular Discovery Center. Specifically, stock siRNA solution was diluted to make 100 nM working solution. 100 μl of the working solution was added to each well of 24-well plates and then mixed with transfection mix (1:100 dilution of RNAiMax in Optimem) and incubated for 20 min. 300 μl of 15,000 ESCs was added to each well in the growth media described above. After incubating cells for 48 hr, the media was changed to differentiation media (DMEM supplemented with 2% charcoal-stripped calf-serum, 0.5 mM 8-Br-cAMP [Sigma], and 0.5 μM of MPA). Cells were incubated for additional 48 hr and then processed for RNA isolation with RNeasy kit (QIAGEN). cDNA was synthesized with High Capacity cDNA Reverse Transcription Kit (Invitrogen). Taqman probes for real-time PCR (Applied Biosystems) for *TFAP2C* (Hs00231476_m1), *HOXA11* (Hs00194149_m1), *HOXD11* (Hs00360798_m1), *FOXO1* (Hs01054576_m1), *PGR* (Hs01556702_m1), *GATA2* (Hs00231119_m1), *HOXD9* (Hs00610725_g1), *SALL1* (Hs00231307_m1), *HOXD10* (Hs00157974_m1), *EBF4* (Hs00325662_m1), *HOXB2* (Hs00609873_g1), *HOXD8* (Hs00980336_g1), *SOX5*

separates ESC TFs and non-ESC TFs. The red line represents the relative expression level of 1, which means no effect. Asterisks indicate that the relative expression is significantly smaller than 1 ($p < 0.05$; one tailed Welch's t test between delta Ct values of negative control and test samples). Error bars represent 2 SEM.

(C and D) Boxplots showing knockdown effects of ESC-recruited (left) and non-ESC-recruited (right) genes on *PRL* (C) and *IGFBP1* (D). From ESC-recruited genes, seven TFs were selected so that the average TPM values (8.26) match that of four non-ESC TFs (8.72). We first took the geometric mean of the relative expression levels in three biological replicates for each gene and then created boxplots for each group.

(E) Heatmap showing knockdown effects among eight TFs. The genes knocked down are shown in row, and the genes whose expression levels were examined are shown in column. The color key is given at the top left corner of the heatmap: green means upregulation and red means downregulation following knockdown.

(F) Schematic showing positive and negative co-regulatory relationships of five TFs that showed consistent downregulation of *PRL* and *IGFBP1*. The widths of edges are proportional to the strength of regulation as revealed by RNAi knockdown assays. Dotted lines represent marginally significant interactions (p value < 0.06).

See also Figures S3 and S4.

(Hs00753050_s1), *NR4A2* (Hs00428691_m1), *L3MBTL* (Hs00287133_m1), *EMX2* (Hs00244574_m1), *FOXF1* (Hs00230962_m1), *SOX6* (Hs00264525_m1), and *IKZF2* (Hs00212361_m1) were purchased and used for real-time PCR experiments.

ACCESSION NUMBERS

All the raw .fastq RNA-seq data have been deposited in the NCBI Gene Expression Omnibus database and are accessible through GEO Series accession number GSE63733.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.01.062.

ACKNOWLEDGMENTS

Financial support has been provided by a grant from the John Templeton Fund (grant no. 54860) and the Yale University Science Development Fund.

Received: July 18, 2014
Revised: December 23, 2014
Accepted: January 28, 2015
Published: February 26, 2015

REFERENCES

Aghajanova, L., Horcajadas, J.A., Esteban, F.J., and Giudice, L.C. (2010). The bone marrow-derived human mesenchymal stem cell: potential progenitor of the endometrial stromal fibroblast. Biol. Reprod. *82*, 1076–1087.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503–511.

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. Nat. Rev. Genet. *9*, 868–882.

Bryder, D., Rossi, D.J., and Weissman, I.L. (2006). Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. Am. J. Pathol. *169*, 338–346.

Cunningham, C.W., Omland, K.E., and Oakley, T.H. (1998). Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. *13*, 361–366.

Dunn, C.L., Kelly, R.W., and Critchley, H.O. (2003). Decidualization of the human endometrial stromal cell: an enigmatic transformation. Reprod. Biomed. Online *7*, 151–161.

Dyson, M.T., Roqueiro, D., Monsivais, D., Ercan, C.M., Pavone, M.E., Brooks, D.C., Kakinuma, T., Ono, M., Jafari, N., Dai, Y., and Bulun, S.E. (2014). Genome-wide DNA methylation analysis predicts an epigenetic switch for GATA factor expression in endometriosis. PLoS Genet. *10*, e1004158.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Emera, D., Romero, R., and Wagner, G. (2012). The evolution of menstruation: a new model for genetic assimilation: explaining molecular origins of maternal responses to fetal invasiveness. BioEssays: news and reviews in molecular, cellular and developmental biology *34*, 26–35.

Erlebacher, A. (2013). Immunology of the maternal-fetal interface. Annu. Rev. Immunol. *31*, 387–411.

Erwin, D.H., and Davidson, E.H. (2009). The evolution of hierarchical gene regulatory networks. Nat. Rev. Genet. *10*, 141–148.

García, M.A., Campillos, M., Marina, A., Valdivieso, F., and Vázquez, J. (1999). Transcription factor AP-2 activity is modulated by protein kinase A-mediated phosphorylation. FEBS Lett. *444*, 27–31.

García-Pacheco, J.M., Oliver, C., Kimatrai, M., Blanco, F.J., and Olivares, E.G. (2001). Human decidual stromal cells express CD34 and STRO-1 and are related to bone marrow stromal precursors. Mol. Hum. Reprod. *7*, 1151–1157.

Gellersen, B., and Brosens, J.J. (2014). Cyclic decidualization of the human endometrium in reproductive health and failure. Endocr. Rev. *35*, 851–905.

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.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Haig, D. (1993). Genetic conflicts in human pregnancy. Q. Rev. Biol. *68*, 495–532.

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*, 497.

Holland, B.R., Huber, K.T., Dress, A., and Moulton, V. (2002). Delta plots: a tool for analyzing phylogenetic distance data. Mol. Biol. Evol. *19*, 2051–2059.

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.

Kuckenberg, P., Kubaczka, C., and Schorle, H. (2012). The role of transcription factor Tcfap2c/TFAP2C in trophectoderm development. Reprod. Biomed. Online *25*, 12–20.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Li, S., Chang, S., Qi, X., Richardson, J.A., and Olson, E.N. (2006). Requirement of a myocardin-related transcription factor for development of mammary myoepithelial cells. Mol. Cell. Biol. *26*, 5797–5808.

Liang, C., Forrest, A.R., and Wagner, G.P.; FANTOM Consortium (2015). The statistical geometry of transcriptome divergence in cell-type evolution and cancer. Nat. Commun. *6*, 6066.

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.

Mossman, H.W. (1987). Vertebrate Fetal Membranes (Rutgers University Press).

Muñoz-Fernández, R., Blanco, F.J., Frecha, C., Martín, F., Kimatrai, M., Abadía-Molina, A.C., García-Pacheco, J.M., and Olivares, E.G. (2006). Follicular dendritic cells are related to bone marrow stromal cell progenitors and to myofibroblasts. J. Immunol. *177*, 280–289.

Nair, N.U., Lin, Y., Manasovska, A., Antic, J., Grnarova, P., Sahu, A.D., Bucher, P., and Moret, B.M. (2014). Study of cell differentiation by phylogenetic analysis using histone modification data. BMC Bioinformatics *15*, 269.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

Oakley, T.H., Plachetzki, D.C., and Rivera, A.S. (2007). Furcation, field-splitting, and the evolutionary origins of novelty in arthropod photoreceptors. Arthropod Struct. Dev. *36*, 386–400.

Oliver, C., Montes, M.J., Galindo, J.A., Ruiz, C., and Olivares, E.G. (1999). Human decidual stromal cells express alpha-smooth muscle actin and show ultrastructural similarities with myofibroblasts. Hum. Reprod. *14*, 1599–1605.

Pronk, C.J., Rossi, D.J., Månsson, R., Attema, J.L., Norddahl, G.L., Chan, C.K., Sigvardsson, M., Weissman, I.L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell *1*, 428–442.

R Development Core Team (2012). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

Raines, A.M., Adam, M., Magella, B., Meyer, S.E., Grimes, H.L., Dey, S.K., and Potter, S.S. (2013). Recombineering-based dissection of flanking and paralogous Hox gene functions in mouse reproductive tracts. Development *140*, 2942–2952.

Ramathal, C.Y., Bagchi, I.C., Taylor, R.N., and Bagchi, M.K. (2010). Endometrial decidualization: of mice and men. Semin. Reprod. Med. *28*, 17–26.

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell *140*, 744–752.

Rubel, C.A., Franco, H.L., Jeong, J.W., Lydon, J.P., and DeMayo, F.J. (2012). GATA2 is expressed at critical times in the mouse uterus during pregnancy. Gene Expr. Patterns *12*, 196–203.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Spitzer, T.L., Rojas, A., Zelenko, Z., Aghajanova, L., Erikson, D.W., Barragan, F., Meyer, M., Tamaresis, J.S., Hamilton, A.E., Irwin, J.C., and Giudice, L.C. (2012). Perivascular human endometrial mesenchymal stem cells express pathways relevant to self-renewal, lineage specification, and functional phenotype. Biol. Reprod. *86*, 58.

Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. *9*, 99–107.

Suzuki, R., and Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics *22*, 1540–1542.

Swofford, D.L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony. Version 4 (Sunderland, Massachusetts: Sinauer Associates).

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. *10*, 252–263.

Villadsen, R., Fridriksdottir, A.J., Rønnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M.A., Bissell, M.J., and Petersen, O.W. (2007). Evidence for a stem cell hierarchy in the adult human breast. J. Cell Biol. *177*, 87–101.

Wagner, G.P. (2007). The developmental genetics of homology. Nat. Rev. Genet. *8*, 473–479.

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.

Wagner, G.P., Kin, K., Muglia, L., and Pavlicev, M. (2014). Evolution of mammalian pregnancy and the origin of the decidual stromal cell. Int. J. Dev. Biol. *58*, 117–126.