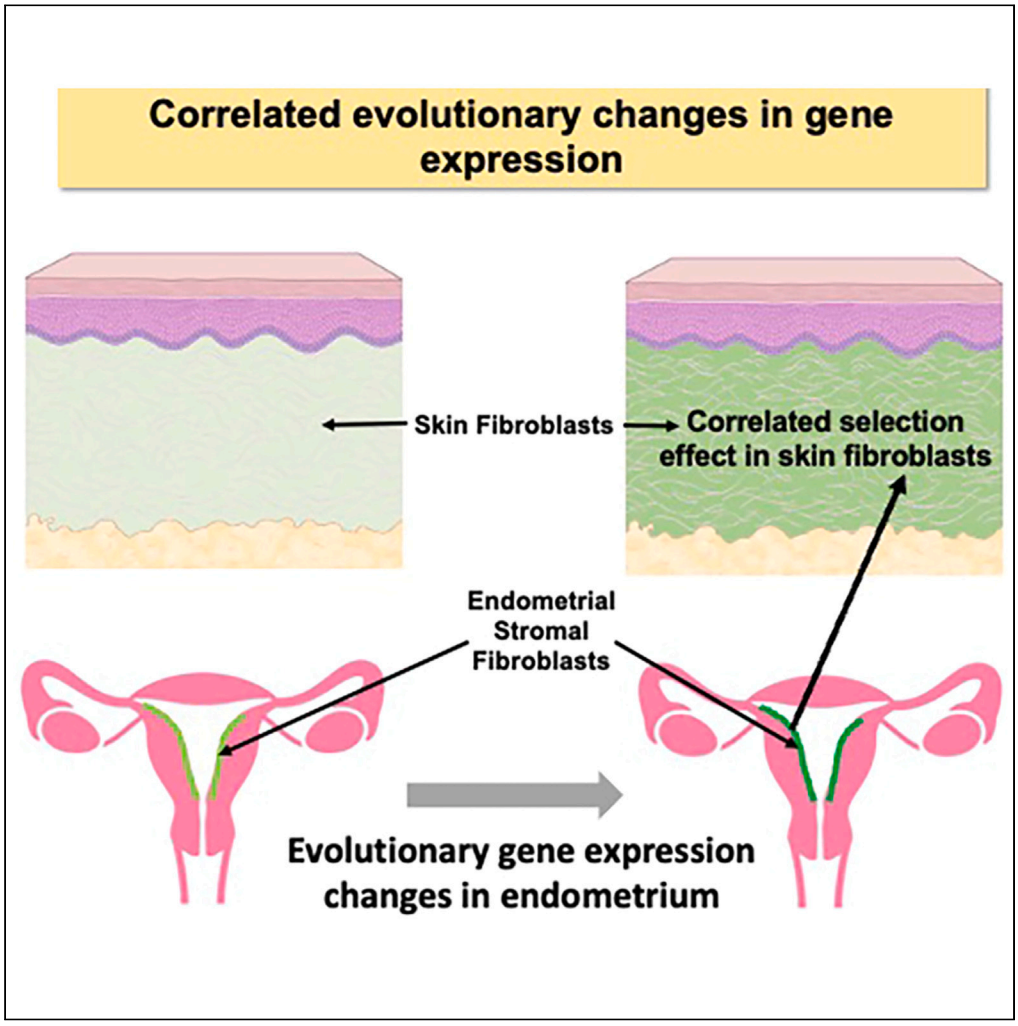
iScience

CellPress
OPEN ACCESS

Article

Experimental and phylogenetic evidence for correlated gene expression evolution in endometrial and skin fibroblasts



Anasuya Dighe, Jamie Maziarz, Arig Ibrahim-Hashim, Robert A. Gatenby, Kshitiz, Andre Levchenko, Günter P. Wagner

gunter.wagner@yale.edu

Highlights

Evolutionary gene expression changes in SFs and ESFs are highly correlated

Selection on gene expression in SFs leads to correlated gene expression changes in ESFs

These findings explain the co-evolution between placenta type and cancer malignancy

Dighe et al., iScience 27, 108593
January 19, 2024
https://doi.org/10.1016/j.isci.2023.108593

iScience

CellPress
OPEN ACCESS

Article

Experimental and phylogenetic evidence for correlated gene expression evolution in endometrial and skin fibroblasts

Anasuya Dighe,[1,2] Jamie Maziarz,[1,2] Arig Ibrahim-Hashim,[3] Robert A. Gatenby,[3] Kshitiz,[4] Andre Levchenko,[2,5] and Günter P. Wagner[1,2,6,7,*]

SUMMARY

Gene expression change is a dominant mode of evolution. Mutations, however, can affect gene expression in multiple cell types. Therefore, gene expression evolution in one cell type can lead to similar gene expression changes in another cell type. Here, we test this hypothesis by investigating dermal skin fibroblasts (SFs) and uterine endometrial stromal fibroblasts (ESFs). The comparative dataset consists of transcriptomes from cultured SF and ESF of nine mammalian species. We find that evolutionary changes in gene expression in SF and ESF are highly correlated. The experimental dataset derives from a SCID mouse strain selected for slow cancer growth leading to substantial gene expression changes in SFs. We compared the gene expression profiles of SF with that of ESF and found a significant correlation between them. We discuss the implications of these findings for the evolutionary correlation between placental invasiveness and vulnerability to metastatic cancer.

INTRODUCTION

Cell types are the fundamental building blocks of multicellular organisms.[1,2] In evolution, cell types arise through the gradual divergence of the gene regulatory networks of different cell populations, eventually leading to distinct although overlapping gene expression profiles.[3–6] More closely related cell types are more likely to have similar gene regulatory networks than more distantly related cell types. Therefore, closely related cell types are likely to be affected by the same mutations leading to correlated patterns of genetic variation in gene expression,[4] and, as a consequence, are expected to evolve similar gene expression changes among species.[7] In quantitative genetics, this phenomenon is called "correlated selection response."[8,9] An example are fibroblastic cells, which exist in many tissues with tissue-specific identities like stellate cells in the liver, or endometrial stromal fibroblasts (ESFs) in the uterus. Nevertheless, fibroblasts share phenotype defining gene expression programs, such as those for remodeling the extracellular matrix (ECM). Because this ECM modifying program is shared among fibroblastic cell types, mutations have the potential to affect the phenotype of different fibroblastic cell types. Here, we test this prediction with respect to gene expression in the dermal (skin) fibroblast (SF) and the uterine ESF. Specifically, we test two complementary predictions: (1) that, in evolution, gene expression in SF and ESF tends to change together, in particular for ECM-related genes; and 2) that selection on gene expression in one cell type, e.g., the SF, leads to similar gene expression changes in another cell type, for instance the ESF.

We use two orthogonal datasets to test these predictions. For the first prediction, we use a comparative transcriptomic dataset from 9 mammalian species from the Boreoeutheria clade of eutherian mammals and perform phylogenetic tests of correlated gene expression evolution. To test the second prediction, we use a mouse strain resulting from an artificial selection experiment on SCID mice, where male mice were selected for lower rate of cancer growth in an induced cancer model.[10]

The focus on SF and ESF is motivated by their role in explaining the correlation between the structure of the fetal-maternal interface and the vulnerability of a species to cancer malignancy.[11–13] Species, from Eutheria (placental mammals *sensu stricto*), with less invasive placenta have endometrial cells that are more resistant to invasion by trophoblast cells than that of species with invasive placenta.[12,14,15] And species with less invasive placenta tend to be less vulnerable to cancer metastasis than species with invasive placenta, for instance humans.[16] This model has been called ELI for "Evolved Levels of Invasibility," where the term "invasibility" denotes the vulnerability of a tissue to invasion by some cell type.[17] This term is well established in ecology, where it describes the vulnerability of a habitat to be invaded by foreign species.[18,19] In the case of tissues, the invasive cell type can be a cancer cell population, the extra villous trophoblast cells of the placenta or a

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
[2]Systems Biology Institute, Yale University, West Haven, CT, USA
[3]Moffitt Cancer Center, Tampa, FL, USA
[4]Biomedical Engineering, University of Connecticut, Farmington, CT, USA
[5]Department of Biomedical Engineering, Yale University, New Haven, CT, USA
[6]Department of Evolutionary Biology, University of Vienna, Djerassi Platz 1, Vienna A-1030, Austria
[7]Lead contact
*Correspondence: gunter.wagner@yale.edu
https://doi.org/10.1016/j.isci.2023.108593

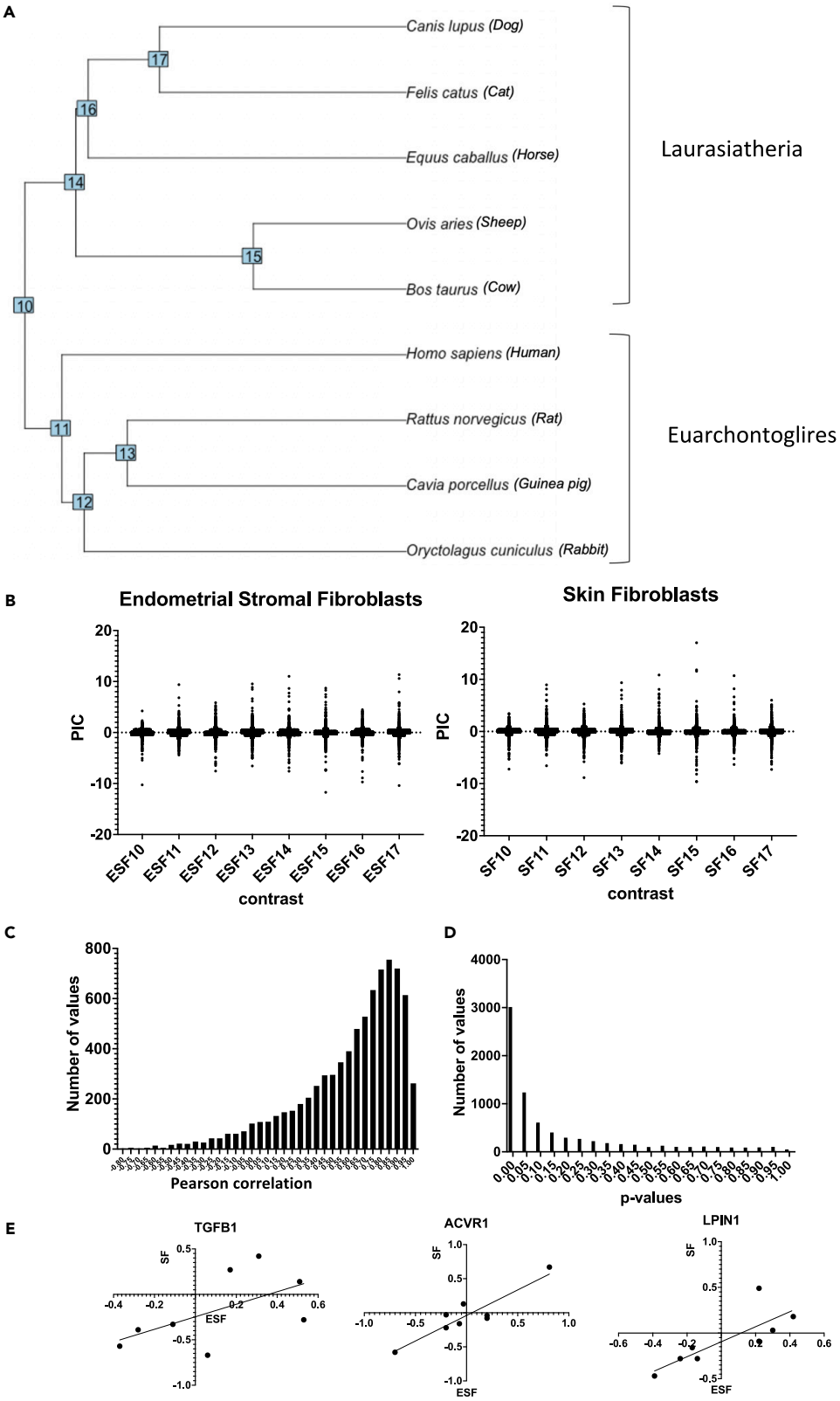

A
Canis lupus (Dog)
Felis catus (Cat)
Equus caballus (Horse)
Ovis aries (Sheep)
Bos taurus (Cow)
Laurasiatheria
Homo sapiens (Human)
Rattus norvegicus (Rat)
Cavia porcellus (Guinea pig)
Oryctolagus cuniculus (Rabbit)
Euarchontoglires
17
16
14
15
10
11
13
12
B
Endometrial Stromal Fibroblasts
Skin Fibroblasts
PIC
contrast
ESF10
ESF11
ESF12
ESF13
ESF14
ESF15
ESF16
ESF17
SF10
SF11
SF12
SF13
SF14
SF15
SF16
SF17
C
Number of values
Pearson correlation
D
Number of values
p-values
E
TGFB1
ACVR1
LPIN1
SF
ESF

Figure 1. Comparative test of correlated gene expression evolution between SF and ESF

(A) phylogenetic relationships of species included in the comparison.

(B) Distribution of PIC values for ESF (left) and SF (right) for each contrast in the phylogenetic tree (A).

(C) Distribution of correlation values among phylogenetic independent contrasts for gene expression changes in SF and ESF (see M&M for details).

(D) Distribution of p values (t-test) of the between PIC correlations showing strong bias toward small values.

(E) Examples of genes with high PIC value correlations. The x axis is always the PIC in ESF, the y axis are the PIC values for SF. The scientific names for the clades represented by the internal nodes are: 10: Boreoeutheria; 11: Euarchontoglires; 12: Glires; 13; Rodentia; 14: Laurasiatheria; 15: Bovidae; 16: Ferungulata; 17: Carnivora.

leukocyte. In brief, the model claims that the nature of the fetal-maternal interface is determined by the invasibility of the uterine endometrium (reviewed in the study by Wagner et al.[16]) and that the evolution of endometrial invasibility by placental cells will lead to correlated changes in the invasibility of somatic tissues, thus leading to a corresponding change in invasibility by cancer cells of other parts of the body. The correlation of invasibility phenotypes has been tested experimentally,[12,20] but the idea that correlated gene expression evolution among these cell types explains this relationship has not been tested. Here, we present two orthogonal datasets to assess the validity of this assumption, one phylogenetic and the other experimental.

The phylogenetic test assesses whether, in evolution, gene expression changes in one cell type are correlated with corresponding gene expression changes in the other cell type. For this, we take advantage of transcriptomic data from cultured SFs and ESFs from 9 mammalian species described previously.[20–22] To test for correlated gene expression evolution, we use phylogenetic independent contrasts (PICs).[23] PICs estimate the amount of evolutionary change along independent parts of the phylogenetic tree. We find a strong signal of correlated gene expression evolution supporting a core assumption of the ELI model of cancer malignancy evolution, namely that evolution of gene expression in the uterus is correlated with gene expression evolution in the other fibroblasts. The direction of dependency, from uterus cells to skin cells, is motivated by the large differences in placental phenotype among mammalian species, suggesting strong selection and fast evolution of endometrial invasibility.[24]

For the experimental test, we focus on an SCID mouse line that was produced by selecting male mice for slow tumor growth of subcutaneously injected Lewis lung carcinoma cells.[10] In this experiment, the authors achieved a significant delay in the size increase of the experimental tumor. The design of the experiment ensured that the slower tumor growth was caused by changes in the host organism rather than changes to the tumor cells. Since SCID mice have no adaptive immune system, the likely cell type responsible for the slower tumor growth are mesenchymal cells, in particular SF. In fact, mice with slower growing tumors have a much more collagenous dermal layer of the skin. We thus infer that artificial selection for tumor growth affected gene expression in SFs (as confirmed by our transcriptomic data in the following section) and ask whether there are correlated changes in the gene expression profile of ESF. We find that gene expression differences between ESFs from wild type and evolved SCID mice are correlated with the evolved changes in SFs.

RESULTS

Phylogenetic test for correlated gene expression evolution

We analyzed gene expression by RNA sequencing in cultured SFs and ESFs from 9 mammalian species representing two sub-clades of the Boreoeutheria, namely the Laurasiatheria represented by dog, cat, horse, cow, and sheep, as well as the Euarchontoglires represented by human, rat, guinea pig, and rabbit (Figure 1A). These animals differ greatly in their fetal-maternal interface, where all Euarchontoglires have a hemochorial placenta (most invasive) and the Laurasiatheria representing less invasive phenotypes, such as epitheliochorial placenta in sheep, cow, and horse as well as endotheliochorial in carnivores.[25] In this taxon sample, we expect that the ESF is the cell type under the selection, driven by evolution of different placental phenotypes, and changes in SF are mostly caused by correlated selection effects, since the function of the skin does not seem to be directly related to placental phenotype. This hypothesis can be assessed by observing the estimated amounts of evolutionary change in ESFs compared to SFs, where ESFs are expected to display larger evolutionary changes than SFs if selection is on ESF gene expression. This prediction is based on the quantitative genetic theory of correlated selection response.[8] This theory says that if two quantitative traits, say X and Y (e.g., expression of a gene in ESF = X and expression of the gene in SF = Y), are genetically correlated, any directional selection on one trait, say X, will lead to changes in the other character as well, say Y, even though Y is not itself under directional selection. The change in Y is called "correlated selection response." However, the size of the correlated selection response in Y will be determined by the degree of correlation between X and Y, where lower correlations lead to smaller change in Y, assuming that X and Y have similar heritabilities.

We calculated the PICs for each gene in our taxon sample. Specifically, for each gene, we calculated the 8 independent contrasts in both cell types, which estimate evolutionary gene expression changes in SF and ESF in corresponding parts of the phylogenetic tree. For each gene, we have a dataset of 8 PICs for SF and 8 corresponding PICs for ESF. We then calculated the correlation of the corresponding PICs for each gene. Figure 1B shows the distribution of the PIC values over all the genes stratified for the 8 contrasts on the phylogenetic trees. Inspection of the results revealed a number of genes with essentially zero evolutionary change, which nevertheless had high nominal correlation values. To avoid artifactually high average correlations, we filtered the genes by eliminating those genes which had less than 1% of the maximal standard deviation of PIC values found among all genes, retaining 7844 one-to-one orthologs. In Figure 1C, we display the distribution of correlation coefficients between corresponding PICs for ESF and SF after eliminating genes with very small amounts of evolutionary change. This distribution is heavily biased toward 1, with a mode within the interval (0.8, 0.85) representing 10% of the genes, and with 73% of

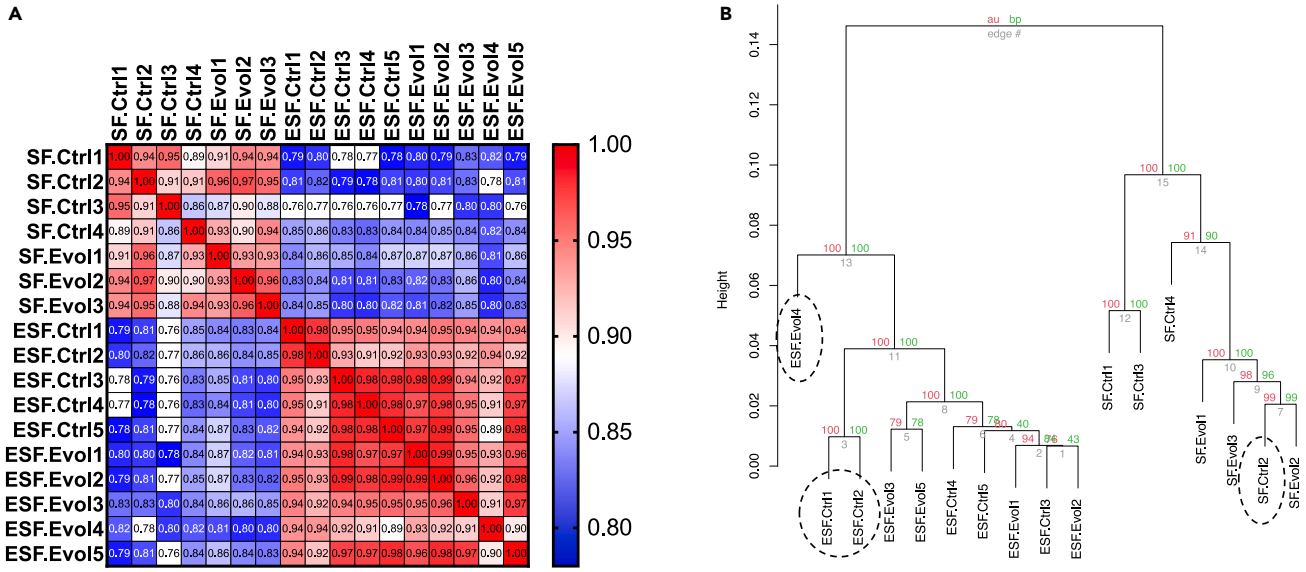


Figure 2. Consistency analysis of all samples included in this study

(A) Heatmap of Spearman correlation coefficients of the transcriptomic profiles; (B) cluster analysis of samples identifying samples that were excluded from the analysis as indicated by dashed ovals. SF and ESF samples are well separated, as expected from different cell types. In the SF samples, one Ctrl sample is nested within the evolved samples and has been eliminated. In the ESF samples, one evolved and two control samples are divergent from the rest and have been excluded from further analysis.

genes with nominal correlations greater than or equal 0.5. The distribution of p values associated with these correlations (Figure 1D) allows us to estimate the number of genes that violate the null hypothesis of no correlation. About 72% of the evolving 1-1 orthologous genes are inferred to have correlated gene expression evolution, which is within the expected range based on other datasets.[7] Examples of highly correlated genes in our data are TGFB1, ACVR1, and LPIN1 (Figure 1E). In order to assess whether contrasts in ESF are greater than that of SF, we looked at the median slope which is 0.66, indicating that the contrasts for SF tend to be smaller than that for ESF. This is consistent with the assumption that most of the changes in ESF gene expression are under direct selection and that the corresponding gene expression changes in SF are due to correlated effects (see introduction above).

We calculated the Gene Ontology (GO) enrichment of highly correlated genes for a strict criterion ($p < 0.001$), and found that the corresponding false discovery rate q-values of enriched GO categories are higher than 0.18, i.e., no strong enrichment of genes in any GO category (Table S1). This result suggests that a broad set of genes are correlated in their evolutionary dynamic rather than genes from specific functional categories.

Overall, the pattern of gene expression divergence within the Boreoeutheria taxon sample is correlated between ESF and SF with, on average, a larger change in ESF than SF. This pattern is consistent with a model of correlated gene expression evolution in which gene expression in ESF is under directional selection and changes in SF are due to correlated selection responses.

While the comparative data analyzed previously supports the notion that gene expression evolution in ESF and SF is correlated, we have not shown that gene expression is correlated because of underlying genetic correlations. An alternative model would be that correlated gene expression evolution in these two cell types is due to co-selection on both cell types, i.e., that selection on gene expression for say ESF co-occurs with selection on gene expression in the other cell type, e.g., SF. In order to address this issue, we analyzed a dataset where selection was only applied to one cell type and ask whether this leads to a correlated change in gene expression in the other cell type.

Experimental test for correlated gene expression evolution

SCID mice are a strain of mice with defective adaptive immune system[26] and are used extensively in cancer research. In a previous study, we used the SCID strain to select for slow tumor growth and deposited this evolved strain in the Charles River frozen embryo collection (MCC SCID line).[10] For this study, we obtained resurrected animals from this strain and bred them to form a colony at Yale University. We call these mice the SCID-EVOL strain. In addition, we obtained wild-type SCID mice from Charles River (Fox Chase SCID Beige Mouse, CB17.Cg-Prkdc^{scid}Lyst$^{bg\text{-}J}$/Crl) and established a colony that we call SCID-CTRL or SCID wild type. We first performed a transcriptomic analysis of SF isolated from the back skin. Data from four SCID-CTRL and three SCID-EVOL animals were obtained. Next, we isolated ESF in order to test whether the uterus also exhibits gene expression changes due to correlated selection responses. To that end, we monitored female SCID-CTRL and SCID-EVOL mice for their ovarian cycle stage via daily vaginal lavage and harvested uterine tissue when the vaginal lavage indicated the presence of large numbers of keratinized squamous epithelial cells, indicating estrus or metestrus I stage of the cycle. ESFs were isolated through differential adhesion and grown to near confluency before RNA extraction and transcriptome sequencing.

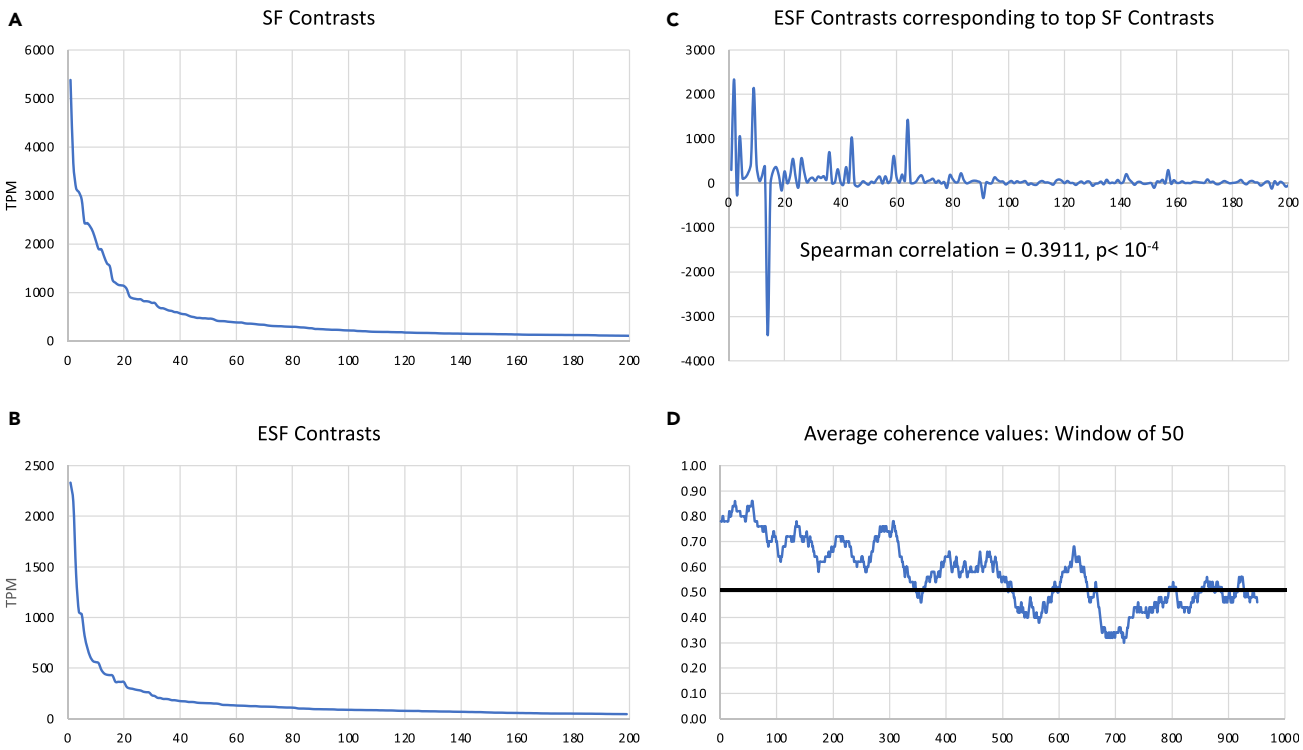


Figure 3. Contrast profiles for different cell types

In each panel, the x axis is the order number of the genes sorted by various criteria.

(A) Contrasts between evolved and control samples for SF samples.

(B) Contrasts between evolved and control samples for ESF samples.

(C) Contrasts between evolved and control samples for ESF where the genes are ordered by the size of SF contrasts (i.e., genes are arranged as in A).

(D) Coherence statistic calculated over a sliding window of size 50. Random expectation of 0.5 is indicated by a thick black line.

QC analysis was performed on the expression profiles of protein coding nuclear genes that were expressed above the gene expression threshold of 3 TPM in at least one sample. The non-parametric Spearman correlation matrix (Figure 2A) shows that SF and ESF samples are well separated as expected for different cell types. The cluster analysis of the samples (Figure 2B) reveals that one of the control SF samples (SFCtrl2) clusters with the evolved samples for unknown reasons. We eliminated this sample from further analysis. Among the ESF samples, we identified one sample from an evolved animal to be strongly distinct in their gene expression profile from all other samples (ESFEvol4) and two wild-type samples (ESFCtrl1 and ESFCtrl2), which are different from other control samples. The latter two samples were from littermates where one individual showed nasal secretions, suggesting that these two individuals may have been infected. All three outlier samples have been eliminated from further analysis, leaving seven ESF samples and six SF samples included in the analysis.

Figure 3A shows the distribution of the top 200 contrasts (=avTPM(Evol) - avTPM(Ctrl)) from SF samples. The highest contrast has been recorded at about 5400 TPM for the thrombospondin 1 (*Thbs1*) gene. The contrasts rapidly fall below 1000 TPM after 20 genes and approach 100 TPM after the top 200 genes. The largest contrast in ESF is less than that for SF, at 2338 TPM for *Acta2*, smooth muscle alpha2 actin, and also rapidly decreases to levels less than 500 TPM after the top 10 genes. Hence, the average of the top contrasts for ESF is smaller (230 TPM) than that of SF (475 TPM), as expected, because the change in SF is due to the direct selection on tumor growth, and the change in ESF is expected to be due to correlated responses. Note that selection was on males, which makes it unlikely that selection on uterine cells played a role.

The ESF contrasts of the genes that constitute the top 200 SF contrasts are highly variable, as expected (Figure 3C). To assess whether the ESF contrasts are more similar to SF contrasts than expected by chance, we first calculated the non-parametric Spearman correlation coefficient and found it to be strongly positive ($\rho = 0.391$, $p < 10^{-4}$), consistent with a correlated selection response. To further assess whether there was in fact a correlated selection response, we calculated the relative frequency of contrasts that have the same sign in ESF as they have in SF for the top 1,000 SF contrasts. We call this measure "coherence" and calculate the coherence value over a sliding window of 50 genes (Figure 2D). The coherence measure starts between 0.8 and 0.9 among the genes with the highest SF contrast and slowly decreases to the random expectation of 0.5 after about 500–600 genes. The binomial test probability for coherence of 0.8 in 50 trials is $p = 1.19 \times 10^{-5}$, also confirming correlated gene expression change.

Genes with large gene expression contrast in SF and a congruent contrast in ESF are largely related to collagenous ECM and cytoskeleton. For collagenous ECM, these are *Thbs1*, the collagen genes *Col1a1*, *Col1a2*, and *Col3a1*, and secreted protein acidic and cysteine rich, *Sparc*,

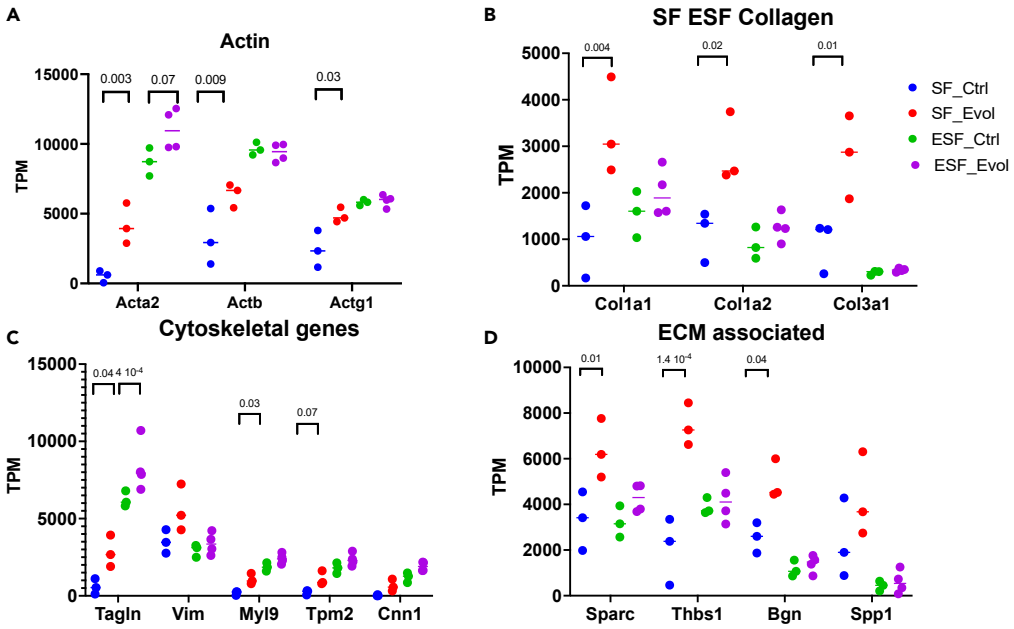


Figure 4. Examples of genes with large SF contrast: (A) actin genes, (B) collagen genes, (C) other cytoskeletal genes, (D) genes for ECM-associated proteins.

as well as biglycan, *Bgn* (Figures 4A and 4C). For cytoskeletal genes, these include *Acta2*, transgelin, *Tagln*, and vimentin, *Vim* (Figures 4B and 4D).

These results support the hypothesis that the correlated evolution of gene expression in SFs and ESFs reported previously can, at least in part, be due to genetic/mutational correlations in gene expression in these two cell types.

DISCUSSION

Because cell types originate in evolution (mostly) by sister cell type divergence, they share parts of their gene regulatory network.[2,4,6,7] One consequence of this model is that mutations that affect gene expression in one cell type tend to have, on average, some degree of correlated effects on gene expression in other cell types. If that is the case, quantitative genetic theory predicts that the evolution of gene expression in one cell type should lead to correlated gene expression changes in other cell types. We have tested this prediction here with a comparative phylogenetic and an experimental approach. In both cases, the results are consistent with a model in which evolutionary changes in gene expression in one cell type are correlated with corresponding gene expression changes in a related cell type.

In the case of our comparative data, the species compared have been selected because they represent different phenotypes at the fetal-maternal interface. The taxon sample includes species with hemochorial placenta, i.e., species where the placenta erodes the maternal tissue and comes into direct contact with maternal blood, such a humans and rodents. On the other hand, there are species with none or only limited removal of maternal tissue as in the epitheliochorial and the endotheliochorial placenta found in cow, sheep, and horse. In the case of comparative data, the estimation of correlations between variables is complicated because of the influence of the phylogenetic relatedness among the species.[23] The solution is to decompose the differences among species into phylogenetically independent contrasts by estimating the amount of evolutionary change along independent segments of the phylogenetic tree. The standard method to do that assumes a Brownian motion process, where the variance among outcomes is increasing linearly through time. The problem with applying this model to gene expression data is that it has been shown that gene expression divergence saturates, i.e., does not increase after a period of time.[27,28] Data from tissue transcriptomes suggest that saturation sets in at about 40 Mio years of additive time of divergence, i.e., the sum of time along both lines of descent since the most recent common ancestor.[28] All but two of the independent contrasts in the phylogenetic tree connecting our taxon sample are estimated to be more than 40 Mio years. The effect of including contrasts with longer divergence time is that the contrasts are normalized by larger expected standard deviations. As a consequence, the PICs calculated are underestimated. However, the corresponding contrasts (gene expression in the two cell types over the same part of the tree) are underestimated by the same factor with the effect of decreasing the calculated co-variance but do not artifactually lead to covariance that is not in the data. For the same reason, the variance among contrasts is also underestimated, and thus the overall effect on the correlation, which is the ratio of the co-variance over the square-root of the product of variances, is approximately canceling out. Based on this reasoning, we propose that PIC is an appropriate method to estimate phylogenetic correlations for gene expression data, despite the violation of the assumptions of the model of Brownian motion.

Experimental evidence suggests that the difference between invasive and non-invasive placentation is due to the endometrium, i.e., due to evolved differences in the maternal tissue. For instance, ectopic pregnancies in pigs lead to invasive implantation while in the uterus the placentation is non-invasive, i.e., epitheliochorial.[14,15] We also have previously shown *in vitro* that ESFs from the cow are resisting invasion (i.e., are less invasible) by trophoblast cells, while human ESFs are much more permissive to invasion, i.e., are more invasible.[12]

In the case of our experimental dataset, the selection was on the growth rate of experimental tumors, which turned out to change SFs. The involvement of SFs is expected since SCID mice have no functioning adaptive immune system[26] and the tumor cells were injected under the skin. At this site, SFs are forming the cancer-associated stroma.[29] There was no selection on ESFs, since the selective breeding was performed on males and males do not have ESFs. Nevertheless, we found that genes which in SFs show a large contrast between wild type and evolved populations also show a correlated response in ESFs. The magnitude of the contrasts in SFs is larger than that in ESFs, which is consistent with the prediction that the changes in ESFs are due to correlated selection responses to changes in SFs rather than direct selection on ESFs.

In both of our datasets, the (natural or artificial) selection was on different types of fibroblasts. In the comparative dataset, we assumed that species differences arose, in part, by selection on the endometrial stroma, i.e., the ESF, and the correlated response was on the SF. The rationale was the fact that the maternal cells determine to a large degree the degree of placental invasion and the species compared differ from hemochorial to epitheliochorial placentation. This assumption is confirmed by the larger inferred evolutionary changes in ESFs compared to that in SFs. In the experimental dataset, the artificial selection resulted in gene expression in SF, as shown by the large gene expression differences in SFs while the ESFs changes were smaller. As argued in the introduction, the cell phenotype determining features of fibroblasts are the expression of ECM molecules and that affecting cell shape (cytoskeletal genes). Consistent with this rationale, large differences in the expression of ECM and cytoskeletal genes are detected in ESFs in the comparative data (Figure S1) and SFs in the experimental data (Figure 4). The degree of expression correlation between cell types differs, as expected for two reasons. In the comparative data, tens of million years of evolution has shaped the transcriptomes of these cells, with additional independent selection forces likely acting on both cell types. On the other hand, the selection response in the short-term selection experiment depends on the spectrum of alleles coincidentally segregating in the founding population. Nevertheless, a broad consistency can be detected.

The results discussed in the previous paragraphs suggest that correlated gene expression evolution is pervasive and needs to be taken into account when studying gene expression evolution.[7] This fact is particularly important when considering the hypothesized correlation between cancer malignancy and the nature of the fetal-maternal interface.[11] Eutherian species with more invasive placenta tend to be more vulnerable to cancer malignancy (e.g., humans), if tumor origination rate and malignancy rate are distinguished.[13] A number of genes have been shown to mediate this effect. One example is the evolution of the *CD44* gene which is expressed in much higher levels in humans and the rhesus monkey than in other eutherian species tested.[21] CD44 is a hyaluronic acid receptor and its expression in the tumor stroma has been implicated in cancer malignancy.[30,31] Similarly, we have shown that GATA2 is a pro-invasibility transcription factor, meaning that a knockout of *GATA2* in human SFs and ESFs leads to lower invasibility by both trophoblast as well as cancer cells.[20] Similar results were obtained for *TFDP1*.[20] The finding that the evolution of gene expression in ESFs and SFs is correlated, and the fact that the same genes expressed in ESFs and SFs affect both placental and cancer invasion, support a model where evolutionary changes at the fetal-maternal interface can affect the disease vulnerability of the respective species, in particular diseases with a similar cell biology as embryo implantation such as cancer malignancy.

Limitations of the study

While our comparative dataset shows robust evidence for correlated evolutionary gene expression change, two limitations need to be acknowledged. Due to difficulties obtaining tissue samples from non-model organism, the skin samples are somewhat heterogeneous with respect to body region (for instance, cow samples are from the face because the rest of hide is commercially valuable), sex, and age. Also, the taxon sample as a whole is limited in size due to the difficulties to obtain uterus samples with live cells. The experimental data from SCID mice are able to show a phenotypic response but it does not allow us to identify the causal polymorphisms underlying the correlated selection response.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - RNA isolation and sequencing
  - Transcript-based abundances using RNAseq data
  - Calculation of phylogenetic independent contrasts (PIC)
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108593.

ACKNOWLEDGMENTS

This research was funded by a grant from the NCI, U54-CA209992 to A.L.

AUTHOR CONTRIBUTIONS

A.D. performed transcriptomic analysis and co-wrote the paper, J.M. isolated and grew the cells and harvested RNA for sequencing, A.I.-H. and R.A.G. created the evolved SCID mouse line, Kshitiz characterized the cell lines, A.L. was the PI of the NIH grant, and G.P.W. conceived the study, performed the animal work, analyzed the data, and wrote the manuscript. All authors edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

Received: August 3, 2023
Revised: October 12, 2023
Accepted: November 27, 2023
Published: November 29, 2023

REFERENCES

1. Vickaryous, M.K., and Hall, B.K. (2006). Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. Biol. Rev. Camb. Philos. Soc. *81*, 425–455.
2. Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecualr studies. Nat. Rev. Genet. *9*, 868–882.
3. Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., and Wagner, G.P. (2016). The origin and evolution of cell types. Nat. Rev. Genet. *17*, 744–757.
4. Musser, J.M., and Wagner, G.P. (2015). Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called "species signal. J. Exp. Zool. B Mol. Dev. Evol. *324*, 588–604.
5. Liang, C., Forrest, A.R.R., and Wagner, G.P.; FANTOM Consortium (2015). The statistical geometry of transcriptome divergence in cell-type evolution and cancer. Nat. Commun. *6*, 6066.
6. DiFrisco, J., Love, A.C., and Wagner, G.P. (2023). The hierarchical basis of serial homology and evolutionary novelty. J. Morphol. *284*, e21531.
7. Liang, C., Musser, J.M., Cloutier, A., Prum, R.O., and Wagner, G.P. (2018). Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes. Genome Biol. Evol. *10*, 538–552.
8. Falconer, D.S., and Mackay, T.F.C. (1996). Introduction to Quantitative Genetics (Longman).
9. Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain : body size allometry. Evolution *33*, 402–416.
10. Ibrahim-Hashim, A., Luddy, K., Abrahams, D., Enriquez-Navas, P., Damagaci, S., Yao, J., Chen, T., Bui, M., Gillies, R.J., O'Farrel, C., and Richards, C.L. (2021). et al.Artificial selection for host resistance to tumour growth and subsequent cancer cell adaptations: an evolutionary arms race. Br. J. Cancer *124*, 455–465.
11. D'Souza, A.W., and Wagner, G.P. (2014). Malignant cancer and invasive placentation: A case for positive pleiotropy between endometrial and malignancy phenotypes. Evol. Med. Public Health *2014*, 136–145.
12. Kshitiz, K., Afzal, J., Maziarz, J.D., Hamidzadeh, A., Liang, C., Erkenbrack, E.M., Kim, H.N., Haeger, J.D., Pfarrer, C., Hoang, T., et al. (2019). Evolution of placental invasion and cancer metastasis are causally linked. Nat. Ecol. Evol. *3*, 1743–1753.
13. Wagner, G.P., Kshitiz, and Levchenko, A. (2020). Comments on Boddy et al. 2020: Available data suggest positive relationship between placental invasion and malignancy. Evol. Med. Public Health *2020*, 211–214.
14. Samuel, C.A. (1971). The development of pig trophoblast in ectopic sites. J. Reprod. Fertil. *27*, 494–495.
15. Samuel, C.A., and Perry, J.S. (1972). The ultrastructure of pig trophoblast rransplanted to an ectopic site in the uterine wall. J. Anat. *113*, 139–149.
16. Wagner, G.P., Kshitiz, Dighe, A., Dighe, A., and Levchenko, A. (2022). The Coevolution of Placentation and Cancer. Annu. Rev. Anim. Biosci. *10*, 259–279.
17. Suhail, Y., Afzal, J., and Kshitiz. (2021). Evolved Resistance to Placental Invasion Secondarily Confers Increased Survival in Melanoma Patients. J. Clin. Med. *10*, 595.
18. Lonsdale, W.M. (1999). Global patterns of plant invasions and the concept of invasibility. Ecology *80*, 1522–1536.
19. Richardson, D.M., and Pyšek, P. (2006). Plant invasions: merging the concepts of species invasiveness and community invasibility. Prog. Phys. Geogr. Earth Environ. *30*, 409–431.
20. Suhail, Y., Maziarz, J.D., Novin, A., Dighe, A., Afzal, J., Wagner, G., and Kshitiz. (2022). Tracing the cis-regulatory changes underlying the endometrial control of placental invasion. Proc. Natl. Acad. Sci. USA *119*, e2111256119.
21. Ma, X., Dighe, A., Maziarz, J., Neumann, E., Erkenbrack, E., Hei, Y.Y., Liu, Y., Suhail, Y., Kshitiz, Pak, I., et al. (2022). Evolution of higher mesenchymal CD44 expression in the human lineage: A gene linked to cancer malignancy. Evol. Med. Public Health *10*, 447–462.
22. Ba, Q., Hei, Y., Dighe, A., Li, W., Maziarz, J., Pak, I., Wang, S., Wagner, G.P., and Liu, Y. (2022). Proteotype coevolution and quantitative diversity across 11 mammalian species. Sci. Adv. *8*, eabn0756.
23. Felsenstein, J. (1985). Phylogenies and the Comparative Method. Am. Nat. *125*, 1–15.
24. Ramsey, E.M. (1982). The Placenta. Human and Animal (Praeger Inc).
25. Elliot, M.G., and Crespi, B.J. (2009). Phylogenetic evidence for early hemochorial placentation in eutheria. Placenta *30*, 949–967.
26. Bosma, M.J., and Carroll, A.M. (1991). The SCID mouse mutant: definition, characterization, and potential uses. Annu. Rev. Immunol. *9*, 323–350.
27. Kin, K., Maziarz, J., Chavan, A.R., Kamat, M., Vasudevan, S., Birt, A., Emera, D., Lynch, V.J., Ott, T.L., Pavlicev, M., and Wagner, G.P. (2016). The Transcriptomic Evolution of Mammalian Pregnancy: Gene Expression Innovations in Endometrial Stromal Fibroblasts. Genome Biol. Evol. *8*, 2459–2473.

28. Chen, J., Swofford, R., Johnson, J., Cummings, B.B., Rogel, N., Lindblad-Toh, K., Haerty, W., Palma, F.d., and Regev, A. (2019). A quantitative framework for characterizing the evolutionary history of mammalian gene expression. Genome Res. *29*, 53–63.
29. Barbazán, J., and Matic Vignjevic, D. (2019). Cancer associated fibroblasts: is the force the path to the dark side? Curr. Opin. Cell Biol. *56*, 71–79.
30. Senbanjo, L.T., and Chellaiah, M.A. (2017). CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front. Cell Dev. Biol. *5*, 18.
31. Chen, C., Zhao, S., Karnad, A., and Freeman, J.W. (2018). The biology and role of CD44 in cancer progression: therapeutic implications. J. Hematol. Oncol. *11*, 64.
32. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.
33. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.
34. 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.
35. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., et al. (2019). Ensembl 2019. Nucleic Acids Res. *47*, D745–D751.
36. Harvey, P.H., and Pagel, M.D. (1991). The Comparative Method in Evolutionary Biology (Oxford University Press).
37. Garland, T., Harvey, P.H., and Ives, A.R. (1992). Procedures for the Analysis of Comparative Data Using Phylogenetically Independent Contrasts. Syst. Biol. *41*, 18–32.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Bulk RNASeq of SF & ESF (SCID Ctrl and Evol)	This paper	PRJNA1006037
Bulk RNASeq of ESF	NCBI BioProject	PRJNA564062
Bulk RNASeq of SF	NCBI BioProject	PRJNA674672
Experimental models: Cell lines		
Immortalized Human Endometrial Stromal Fibroblast (ATCC-CRL-4003)	Gil Mor group(Wayne State University)	HumanESF
Human BJ-5Ta Skin Fibroblasts (ATCC-CRL-4001)	ATCC BJ-5Ta	Human SF
SCID Ctrl (or WT) Mice	Charles River (Fox Chase SCID Beige mouse) CB17.CgPrkdc^{scid}Lyst^{bg-J}/Crl	SCID_Ctrl Mice
SCID Evol Mice	Charles River Frozen Emryo Collection	SCID Evol Mice
Software and algorithms		
R Ape library for calculating PIC	Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).[32]	https://doi.org/10.1093/bioinformatics/btg412
Kallisto for transcript-based abundance	Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).[33]	https://doi.org/10.1038/nbt.3519

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact: Dr. Günter Wagner (gunter.wagner@yale.edu, guenter.wagner@univie.ac.at).

Materials availability

This study did not generate new unique reagents.

Data and code availability

(1) Bulk RNASeq data (for SCID Ctrl and EVOL) have been deposited at NCBI SRA and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
(2) This article does not report any original code.
(3) Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mouse lines: SCID mice from a strain selected for slow tumor growth[10] were obtained from Charles River frozen embryo collection and bred to form a colony at Yale University (IACUC Protocol 2021–11483). We call these mice the SCID-EVOL strain. Wild type SCID mice were obtained from Charles River (Fox Chase SCID Beige Mouse, CB17.Cg-Prkdc^{scid}Lyst^{bg-J}/Crl). We call these mice SCID-CTRL.

Ovarian cycle staging: Female SCID-CTRL and SCID-EVOL mice were monitored for their ovarian cycle stage via vaginal lavage. Lavage samples were spread onto glass slides, fixed with methnol and stained with hematoxylin and eosin. Animals were sacrificed when the vaginal lavage indicated the presence of large numbers of keratinized squamous epithelial cells, indicating estrus or metestrus I stage of the cycle.

Isolation of human endometrial stromal cells and skin fibroblasts: An immortalized human ESF cell line was obtained from the Gil Mor group (ATCC-CRL-4003). Human SFs (BJ5ta) were purchased from ATCC (CRL-4001).

Isolation of skin fibroblasts: Cow (*Bos taurus*), dog (*Canis lupus*), cat (*Felis catus*), guinea pig (*Cavia porcellus*), horse (*Equus caballus*), rabbit (*Oryctolagus cuniculus*), sheep (*Ovis aries*), rat (*Rattus norvegicus*), SCID-CTRL and SCID-EVOL SFs were obtained from fresh skin tissue

samples.[22] Briefly, a piece of skin was collected and shaved, washed in phosphate-buffered saline (PBS) buffer, and cut into pieces of approximately 1.0 cm^2. The dermis was separated from the epidermis enzymatically (30 min in 0.25% trypsin buffer at 37°C, followed by dissociation buffer [collagenase (1 mg/mL), dispase (1 mg/mL), and deoxyribonuclease I (400 μg/mL)] for 45 min at 37°C. Epidermis was removed, and 2-mm pieces of dermis were transferred to a 12-well plate and covered with media. Fibroblasts emerged from the tissue samples and grew to confluency. Fibroblast cell cultures were established in 10-cm dishes with high glucose Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 5% CO2.

Isolation of endometrial stromal fibroblasts: Cow (*Bos taurus*), dog (*Canis lupus*), cat (*Felis catus*), guinea pig (*Cavia porcellus*), horse (*Equus caballus*), rabbit (*Oryctolagus cuniculus*), sheep (*Ovis aries*), rat (*Rattus norvegicus*), SCID-CTRL and SCID-EVOL ESFs were obtained from fresh uterus samples.[21] 2–3 mm uterus fragments were created using a scalpel and digested with 0.25% Trypsin–EDTA for 35 min at 37°C, followed by dissociation buffer (1 mg/ml collagenase, 1 mg/ml Dispase, 400 μg/mL DNase I) for 45 min at 37°C. Cell clumps were homogenized by passage through a 22-gauge syringe. To remove remaining clumps the sample was passed through a 40-μm nylon mesh filter. A single-cell suspension was transferred to T25 flasks containing fresh growth medium. To facilitate enrichment of fibroblasts versus epithelial cells, media were exchanged in each well after 15 min to remove unattached cells. Cells were grown to confluency and sub-passaged into two T25 flasks. Immunohistochemistry was used to test for abundance of vimentin (Santa Cruz, sc-6260) and cytokeratin (Abcam, ab9377) to validate fibroblast identity.

Cell culture: ESFs were grown in phenol red free DMEM/F12 with high glucose (25 mM), supplemented with 10% charcoal-stripped calf serum (Gibco) and 1% antibiotic/antimycotic (Gibco). Human SFs BJ5ta (ATCC) cells were cultured in 80% DMEM and 20% MEM supplemented with 10% FBS, 1% antibiotic/antimycotic and 0.01 mg ml^{-1} hygromycin. All other SFs were cultured in DMEM with high glucose supplemented with 10% FBS.

METHOD DETAILS

RNA isolation and sequencing

RNA was isolated using RNeasy micro kit (QIAGEN) and resuspended in 15 μL of water. The Yale Center for Genome Analysis ran samples on the Agilent Bioanalyzer 2100 to determine RNA quality, prepared mRNA libraries and sequenced on Illumina HiSeq2500 to generate 30–40 million reads per sample (single-end 75 base pair reads).

Transcript-based abundances using RNAseq data

Bulk RNAseq data obtained were quantified using the transcript-based quantification approach as given in the program 'kallisto'.[33] Here, reads are aligned to a reference transcriptome using a fast hashing of k-mers together with a directed de Bruijn graph of the transcriptome. This rapid quantification technique produces transcript-wise abundances which are then normalized and mapped to individual genes and ultimately reported in terms of TPM.[34] The Ensembl[35] gene annotation model was used and raw sequence reads (single-end 75 bp) for ESFs and SFs from human (*Homo sapiens*), cow (*Bos taurus*), dog (*Canis lupus*), cat (*Felis catus*), guinea pig (*Cavia porcellus*), horse (*Equus caballus*), rabbit (*Oryctolagus cuniculus*), sheep (*Ovis aries*) and rat (*Rattus norvegicus*) were aligned to GRCh38.p13, ARS-UCD1.2, CanFam3.1, Felis_catus_9.0, Cavpor3.0, EquCab3.0, OryCun2.0, Oar_v3.1 and Rnor_6.0 reference transcriptome assemblies. Further, to facilitate gene expression across species, a one-to-one ortholog dataset consisting of 8639 genes was formulated using the BioMart tool on Ensembl.

For quantifying gene expression from SCID RNAseq data (Ctrl and Evol), GRCm39 (GCA_000001635.9) mouse reference transcriptome from Ensembl was used.

Calculation of phylogenetic independent contrasts (PIC)

Phylogenetically independent contrasts (PIC)[23,36,37] is a method for removing phylogenetic information from a character dataset with an aim to reduce the phylogenetic interdependence of data points, which can violate the assumptions of many statistical tests. The most broadly used methods to account for phylogenetic structure is the method of PICs where observations on N species are transformed into N-1 contrasts (differences). PIC values were computed by R package APE[32] (Analysis of Phylogenetics and Evolution) using the function "pic" where the gene expression values (in TPM) were iteratively fed into the pic function to obtain an array of resultant PICs for each gene across 8 internal nodes in the phylogenetic tree used in this study.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in R. Details of statistical analyses performed can be found in corresponding Results, Figure legends and Tables.