Cis-Regulatory Differences Explaining Evolved Levels of Endometrial Invasibility in Eutherian Mammals

Yasir Suhail^{1,3}, Jamie D. Maziarz^{2,3}, Anasuya Dighe^{2,3}, Gunter Wagner^{2,3,5,6,*}, Kshitiz^{1,3,4,*}

1 Department of Biomedical Engineering, University of Connecticut Health, Farmington, CT

2. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT

3. Cancer Systems Biology Center (CaSB@Yale), Yale University, West Haven, CT

4. Center for Cell Analysis and Modeling, University of Connecticut Health, Farmington, CT

4. Department of Obstetrics, Gynecology and Reproductive Sciences, Yale Medical School, New Haven, CT

5. Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI

* Corresponding Authors: kshitiz@uchc.edu, gunter.wagner@yale.edu

Abstract

Eutherian (placental) mammals exhibit great differences in the degree of placental invasion into the maternal endometrium, with humans being on the most invasive end. Previously, we have shown that these differences in invasiveness is largely controlled by the stromal fibroblasts of the maternal endometrium, with secondary effect on stroma of other tissues resulting in correlated differences in cancer malignancy. Here, we present a statistical investigation of the second dogma linking the phenotypic and transcriptional differences to the genomic changes across species, revealing the regulatory genomic sequence differences underlying these inter-species differences. We show that gain or loss of specific transcription factor binding site sequences are connected to the inter-species gene-expression differences in a statistically significant manner, with a particularly larger effect on stromal genes related to invasibility. We also uncover transcriptional factors differentially regulating genes related to pro- and anti- invasible property of stroma. This work extends the understanding of inter-species differences in stromal invasion to the causal genomic sequence differences paving new avenues to target stromal characteristics to regulate placental, or cancer invasion.

Introduction

Placentation during pregnancy presents a remarkable phenotypic diversity between different eutherian (placental) mammals¹. This variation is often attributed to the fact that placentation involves the interaction of genetically distinct maternal and fetal tissues²⁻⁴. The maternal-fetal interface during placentation is classified anatomically by the depth of invasion into the endometrium as epitheliochorial (e.g. cattle, dugong), endotheliochorial (e.g. dogs, cats and many carnivores), and hemochorial (e.g. humans and other primates, and rodents)⁵⁻⁷. Even among the species with invasive hemochorial placentation, there is large variation in the extent to which placental trophoblasts invade and interact with the maternal tissue^{8,9}. Invasiveness during pregnancy is also correlated with the rate of malignancy among mammals, with significantly lower rate of melanoma metastasis observed among cows, and horses compared to carnivores and humans¹⁰.

For long, it was posited that the increased stromal invasion among hemochorial placentation (including in humans) is due to higher invasive capability of placental trophoblasts^{11,12}. Genomically informed phylogeny construction showed that the epitheliochorial placentation has evolved more recently from the ancestral hemochorial placentation, which is present in the stem lineage of eutherian mammals⁶. We have hypothesized, and experimentally validated that the evolution of epitheliochorial placentation is in part a stromal characteristic, which has resulted in the stromal fibroblasts to attain increased resistance to invasion from placental trophoblasts¹³. This hypothesis, termed the Evolved Levels of Invasibility (ELI) posits that stromal invasibility itself is an evolved phenotype, resulting in higher resistance to trophoblast invasion within bovines (and other species with epitheliochorial placentation). Notably, its secondary manifestation in other stromal tissues also limit cancer dissemination, suggesting a heritable,

genomic basis for the establishment of the phenotype^{13,14}. In this paper, we explore the extent to which genomic changes explain the variation in stromal invasibility in a larger cohort of mammalian species.

Specifically, we characterized the gain or loss of transcription factor binding sites (TFBS) in the promoter region of genes to explain the variation in gene expression between these species. We found that differences in TFBS number could explain many times higher fraction of variance in gene expression for ELI correlated genes compared to other genes, providing further confidence that stromal invasibility is a selected, and inherited phenotype. We also found transcription factors which antagonistically act as potential enhancers of pro-invasive genes, and repressors of anti-invasive stromal genes across eutherians. Because the ELI predicted stromal resistance is secondarily manifested in other stromal tissues, our current work charts a specific inherited genomic basis for the acquired stromal resistance to placental invasion, as well as cancer malignancy.

Results

Identifying Gene Expression Signatures Correlated with Placental Invasibility

The ELI hypothesis states that bovine stromal fibroblasts, both in endometrium and skin, have evolved to be more resistant to trophoblast and cancer cell invasion than their human counterparts¹³ (**Figure 1A**). To explore the evolutionary history of genes that have changed along with non-invasive epitheliochorial placentation, we expanded the number of eutherian species in our transcriptomic analysis. We isolated and cultured the stromal fibroblasts from endometrial tissues of rabbit, guinea pig, rat, cat, horse, humans, sheep, dog, and cow. This selection of species comprise of two eutherian clades, the Euarchontoglires (primates, rodents and others), and Laurasiatheria (ungulates, carnivores and others), together forming the clade of Boreoeutheria⁵. This balanced phylogeny contained species with both

invasive placentation (human, rabbit, rat, and guinea pig), and less-invasive endotheliochorial or epitheliochorial placentation (cat, dog, horse, sheep, and cow) (**Figure 1B**).

Figure 1. Finding genes correlated with the Evolved Levels of Invasability (ELI) of the endometrial stroma across eutherian mammals. (A) Schematic explaining the ELI principle, showing that among the epitheliochorial placental mammals, the stroma has evolved to resist trophoblast invasion with secondary manifestation in other tissues resulting in decreased cancer malignancy. (B) Phylogenetic

hierarchical representation of selected eutherian species whose endometrial stromal fibroblasts (ESF) were analyzed in the manuscript; The tree is balanced for invasive and less-invasive placental species. Numbers denote the extent of placental invasion ranked from 0 to 3. (C) Statistical significance of the expression of individual genes for correlation with invasibility among different species, evaluated using a simple linear model (x-axis) and a phylogenetic linear model (y-axis). Larger p-values under the phylogenetic linear model, but similar ordering of the genes under the two methods is visible. (D) Individual genes for adjusted p-value (FDR, x-axis) for correlation to placental invasibility against the ratio of the mean expression of the same gene in hemochorial and endotheliochorial species against non-invasive epitheliochorial species. (E) Illustration of the linear fit test used to probe whether the gene expression of a particular gene (in this case, PKD1) in ESFs across species is correlated with the invasibility. (F) Illustration of a gene whose expression is anti-correlated with invasibility. (G and H) Heatmaps of expression across species for the genes most correlated with the invasibility phenotype across species, either positively (ELI^{up} genes), and negatively (ELI^{dn} genes). (I) Schematic showing the gain and loss of binding sites in the promoter regions of a given gene in different species, and its effect on gene expression; b is a coefficient estimating the extent of effect gain of a binding site has on gene expression. (J) The distribution of p-values for individual transcription factor binding site motifs, whose occurrence is tested for an effect on inter-species gene expression differences. The over-representation of low p-values beyond the distribution expected by chance (red line) is evidence of a real effect.

We then created a rough invasibility index based on the anatomical presentation of the fetal-maternal interface in these species^{5,15}. Specifically, species with epitheliochorial placentation were given a rank of 0 (cow, sheep, horse), carnivores with endotheliochorial placentation were ranked 1 (cat and dog), rodents with hemochorial placentation ranked as 2 (rabbit, guinea pig and rats), while humans, which along with primates have a particularly invasive hemochorial placentation were ranked 3 (**Figure 1B**). Our objective was to identify genes that correlate with the placental invasibility phenotype. We statistically tested the correlation (or anti-correlation) of gene expression changes in all the species with the placental invasion phenotype with both a naïve linear, and a phylogenetic linear model (see Methods) (**Figure 1C, Supplementary Table 1**). The phylogenetic linear model corrects for a correlation that arise due to shared descent of gene expression and placental invisibility phenotype¹⁶. As expected, the phylogenetic linear model showed less significant p-values than the naïve model. Nevertheless, the same

genes show high correlations in either linear model (**Figure 1C**). Therefore, either model can be used to prioritize candidate genes correlated with invasibility. Notably, we found that overall more genes have stromal expression positively correlated to invasive placentation, rather than negatively-correlated, suggesting that the recent evolution of epitheliochorial placentation is associated with an overall downregulation of genes affecting invasibility (**Figure 1D**).

Figure 1E shows an example of a gene with higher expression correlated with increased placental invasiveness (defined as ELI^{up} gene), while **Figure 1H** shows an example of a gene whose expression is anti-correlated to invasibility (defined as ELI^{dn} gene). The ELI^{up} genes are purportedly pro-invasible as their stromal expression increases along with increase in placental invasion, while ELI^{dn} genes are purportedly anti-invasible, as they increase in expression with reduced stromal invasion. **Figures 1G-H** show the top ten genes ranked in order of their positive and negative regression respectively, of their expression and the extent of invasion in each species. We selected a balanced set of 200 genes (all had $FDR < 0.05$), with 100 most statistically significant genes each, that were correlated (ELI^{up}) or anti-correlated (ELI^{dn}) with the invasive phenotype. This gene set, termed ELI gene set, is a rank ordered set of genes which show consistent differences across species from the least invasive to the most invasive placentation. This gene set was used to identify candidate transcription factors that regulate stromal invasibility by correlating transcription factor binding site abundance near ELI genes with invasibility phenotype across eutherians.

Linear Model Relating Transcription Factor Binding Site Abundance to Gene Expression

We used a linear regression model to estimate the extent to which cis-regulatory transcription factor binding site numbers could statistically explain gene expression differences across species. Using gene

expression quantified as transcripts per million (TPM), of 8639 genes with 1 to 1 orthology, we first calculated their square root TPM values to decrease the influence of extremely high expression values. We then subtracted the gene-wise mean across all species from the square-root of the gene expression. This scaled expression measure was fit to a linear model with the number of transcription factor binding sites in promoter regions as the predictor variable (**Figure 1I**). Statistical significance and the size of the regression coefficient was recorded for each binding site motif.

Testing for 572 transcription factors for which consensus binding site motifs are available in the JASPAR database¹⁷, we identified 187 TFs (for a threshold of FDR < .05) which could explain a significant portion of the inter-species differences in gene expression in endometrial stromal cells, as calculated across all genes (**Figure 1J**). A large fraction of TFs statistically influence inter-species gene expression variation indicating that the gene expression profile of endometrial stromal cells has evolved, in part, through changes in cis-regulatory elements (**Figure 1J**).

Identification of cis-Regulatory Elements Related to Inter-Species Variation in Stromal Gene Expression

We found that β , the regression coefficient describing the extent to which TFBS abundance could explain variation in gene expression, was positive for most TFs (**Figure 2A**). A positive β indicates the gain of TFBS is correlated with increased expression of the target gene, suggesting that most TFs acted as positive regulators of gene expression (i.e., activating TFs) among endometrial stromal fibroblasts (ESFs). The range of value of β was similar for positive and negative regression coefficients (**Figure 2A**). We found

that for many TFs, $|\beta|$ was above 0.025 for the whole transcriptome, with high significance ($\text{FDR} < 10^{-20}$) (**Figure 2B, Supplementary Table 2**).

Averaged β over all genes did not reflect the large effect a single transcription factor may have on the expression of individual genes (**Figure 2B-C**). As an example, for many genes the regression coefficient $|\beta|$ for the transcription factor MGA could be 20 times higher than the β averaged over all genes,

Figure 2. Global analysis of the genomic basis of the modulation of gene expression in endometrial stromal fibroblasts in differential placental invasion among eutherians. (A) Histogram for the effect size (beta) found for different transcription factor binding site (TFBS) motifs that were found to be statistically significant (FDR < .05). Beta represents the amount the scaled (square root of TPM) expression changes with one binding site, on average. Positive beta implies an enhancer effect, while negative beta implies a repressor effect on gene expression of downstream genes. (B) Transcription factors whose putative binding sites affect the variation in expression across selected eutherian species. Each point represents a single transcription factor binding motif. The statistical significance is plotted in terms of the

FDR on the y-axis, while the effect size beta (increase or decrease in expression) for additional binding sites is plotted on the x-axis; Statistically significant motifs are labeled green. (C) Illustration of the effect of downstream effect of MGA binding sites on the expression of individual genes. Here, each point represents a gene whose expression was independently tested for a linear relationship to the number of binding MGA binding sites. The gene-specific effect size (beta) and significance (FDR) are plotted on the x- and y-axis respectively. (D) Distribution of the number of binding site motifs matched in the promoter regions of genes in different species, shown here for the MGA binding site motif. (E) Illustration of the effect of MGA binding sites on the expression of LRCH3, across different species listed in the order of placental invasiveness. (F) The R-squared (fraction of variance explained) for the most statistically significant transcription factors affecting global ESF gene expression across species. The colors correspond to the effect size, with blue for an overall increase of gene expression with the particular TF binding in its promoter region, and orange for an overall decreasing effect for additional binding. (G) Transcription factor binding sites whose occurrence is related to the expression of genes experimentally tested¹³ for their role in modulating invasibility. Each point represents a binding site sequence motif, with the mean variance explained on the y-axis and the number of genes with p-value < .05 for the correlation on the x-axis.

demonstrating specificity for particular target genes (**Figure 2C**). The inter-species distribution of the number of binding sites for the same transcription factor for all genes showed enrichment for specific TFBS was high for a small number of genes among hemochorial species (**Figure 2D**). As an example Figure 2E shows the expression of the target gene LRCH3, as a function of the number of binding sites for MGA (**Figure 2E**).

We then identified the top transcription factors with the highest R^2 values (**Figure 2F**). These TFs consisted of various nuclear receptors including NR2F6, a steroid receptor which is also a repressor of IL17, the Estrogen Receptor ESR1, as well as SREBF1, the sterol regulatory element binding transcription factor 1¹⁸. NR2F6 co-regulates and inhibits thyroid receptor¹⁹, and Oxytocin Receptor (OXTR)^{20,21}, while ESR1 directly regulates OXTR expression²², and is considered essential for uterine receptivity in various ruminants^{23,24}. In addition, among the top candidates an important factor with a negative β value was FOXA2, the Forkhead Box A2, an essential transcription factor for fertility and uterine function^{25,26}. In

addition, other top transcription factors with high β values were Max Dimerization Protein MGA, a negative regulator of MYC pathway and implicated in endometrial cancers^{27,28}, PAX5 involved in gene regulation in many tissues.

Previously we experimentally demonstrated that human stromal fibroblasts could be rendered more resistant to trophoblast invasion by modifying gene expression to match that of bovine stromal fibroblasts¹³. We tested whether inter-species variation in these experimentally validated ELI genes could be explained by cis-regulatory evolutionary changes (**Figure 2G**). Indeed we found that among 10 experimentally validated genes (ACVR1, BMP4, CD44, DDR2, LPIN1, MGAT5, NCOA3, STARD7, TGFB1, WNT5B), many TFs contributed to their inter-species expression differences. These included homeobox transcription factors like PITX1, and GSC, sterol binding factors like SREBF1, SREBF2, and NR4A1, and nuclear signal transducers like NFKB2, and STAT6. Notably, the experimentally validated genes were positively-correlated with the invasibility phenotype, and accordingly the regulating TFs also had positive β values.

TFBS Statistically Explain Expression Variation of Genes Correlated with Stromal Invasibility

Our transcriptomic analysis had revealed many genes whose expression is correlated with the placental invasion phenotype, i.e. the ELI^{up} and ELI^{dn} gene sets (**Figures 1C, G-H**). The ELI framework posits that, within eutherians, resistance to placental invasion has emerged by changes in the stromal fibroblasts¹³. Having shown in the previous section that inter-species variation in gene expression is correlated with TFBS abundance, we focus here on the regulation of the ELI gene set (**Figure 3A**). To account for the possible systemic inter-species variation in the number of binding sites of different transcription factors,

we chose the top 100 significant genes each from both ELI-correlated (purportedly pro-invasible, ELI^{up}), and ELI-anti-correlated (purportedly anti-invasible, ELI^{dn}) gene sets.

Modeling specifically for this balanced ELI-specific set of 200 genes, we found that TFBS could statistically explain a much higher fraction of gene expression differences than for all 1-1 orthologous genes (**Figure 3A-B**). For many transcription factors, both the absolute β values, as well as the fraction of variance explained (**Figures 2C, 3C**), were substantially higher than for the whole gene set, with most TFs exhibiting a positive β value. Many of these TFs were related to the myosin enhancer family of transcription factors, with all major MEF TFs, e.g. MEF2A-D, showing significant effects. In addition, we found that GATA transcription factors were highly enriched among the TFs showing significant effects. These included GATA1 and GATA2 which are classically referred to as hemotopoeitic TFs²⁹, and GATA5 and GATA6³⁰, which are classically referred to as cardiac specific TFs. Crucially, GATA1 and GATA2 are key TFs upregulated in cancer stromal fibroblasts, and induce increased cancer dissemination³¹. All 4 GATA factors are also important in trophoblast development³², with GATA2 acting as a key TF regulating decidualization^{14,33}. These data highlight genomically encoded transcriptional regulation of decidual genes in species with increased endometrium-trophoblast interaction. Notably, we found relatively few TFs with significant negative β values, but they included key developmental TFs like Pax6, SOX10, and SIX2.

Figure 3. Genomic basis for the regulation of genes correlated with stromal invasibility among eutherian species. (A) The distribution of p-values for the effect of individual transcription factor binding site motifs, tested for the effect on expression of ELI-related genes, with the red line representing the distribution expected by chance. (B) The transcription factors whose binding sites affect the expression of genes purportedly involved in the

invasibility of the endometrium (the ELI-related genes). Statistical significance (y-axis, FDR) with the effect size (x-axis, beta). (C) The fraction of variance (R-squared) of the expression across genes explained only by the occurrence of a particular binding motif aggregated over a balanced set of the most significant 100 ELI^{up} and 100 ELI^{dn} genes, shown here for the most significant transcription factor binding motifs. The colors of the bars represent the effect size, green being higher. (D) Illustration of the linear fit used to calculate the statistical significance and effect size for an affected gene, UACA; Shown are the changes in gene expression across species with increasing binding sites for the transcription factor GATA2; Species are ordered on the basis of their extent of placental invasion. (E) Specific genes likely to be regulated by GATA2. The fraction of variance explained (y-axis), and the effect size (x-axis, beta) are calculated for a linear model relating the (scaled) gene expression across species for a single gene to the number of GATA2 binding site motifs in the promoter regions among the different species. (F) ELI-correlated (ELI^{up}) and (G) ELI anti-correlated (ELI^{dn}) genes whose variance is most explained by the number of specific binding site sequences. Beta is the common effect size found for the binding site sequence found for all ELI genes. For individual transcription factor-gene pair, a large extent of inter-species variance in the direction (or against the direction) of placental invasiveness could be explained by the cis-regulatory TFBS elements.

Surprisingly, we found that the transcription factors displaying the highest statistical significance showed R² value 50 times higher for similar calculations aggregated over all genes (**Figures 2C, 3B-C**). Individual genes likely to be involved in regulating placental invasibility (such as UACA) can be seen to be affected by differences in the number of binding sites of particular TFs (such as GATA2) (**Figure 3D**). Further, as for the whole transcriptome, an aggregated R² value for a transcription factor over all ELI genes masked the substantial effect of TFBS in explaining the substantial variance in gene expression for individual downstream genes (**Figure 3E**). As an example, binding sites for GATA2 could explain more than 50% of the inter-species variance for many genes. Interestingly GATA2 is acting either as an enhancer, or a repressor of gene expression (**Figure 3E**). Top TF-gene pairs for TFs which act either as enhancers ($\beta > 0$), or repressors ($\beta < 0$) showed that for many ELI genes, a substantial extent of the inter-species transcriptional difference is explained by changes in cis-regulatory binding sites (**Figure 3F-G, Supplementary Table 3**). Overall, these data strongly suggest that hard-coded cis-regulatory genomic changes can significantly explain changes in gene expression associated with increased stromal vulnerability, and stromal resistance to placental invasion.

Evolution of Increased Resistance to Placental Invasion is Associated with Distinct Patterns of Genomic Regulation

Phylogenetic analysis has settled the debate over the evolution of non-invasive epitheliochorial placentation in eutherians, which is now considered to have evolved after an ancestral invasive hemochorial placentation⁶. However, anatomical reports suggest that among primates, the placentation has continued to evolve into a very aggressively invading form^{8,34}, suggesting continued evolution among invasive placentation. We asked whether cis-regulatory genomic changes could specifically explain gain or loss of the phenotype of placental invasibility. Towards this objective, we tested if ELI genes that correlate with increased (or decreased) placental invasion are differently influenced by TFBS.

We used top 100 genes which were the most positively correlated (ELI^{up} genes), and the top 100 genes most anti-correlated (ELI^{dn} genes) with the invasive placental phenotype within the stroma (FDR < 0.05 for all selected genes). Separate treatment of either gene-sets in our model could be influenced by systematic inter-species differences in TFBS frequencies, because the ranked order of ELI^{up} (or ELI^{dn}) genes is correlated with the placental invasion phenotype, and hence the species, by definition. To avoid this artefact, we normalized the frequencies of each TFBS across the species. We found that a higher number of TFs showed gene expression enhancing effects for ELI^{dn} genes, while for ELI^{up} genes, relatively more TFs showed repressive effects on gene expression (**Figure 4A**).

Figure 4. Differences in the TFBS explained regulation of genes correlated (ELI^{up}), or anti-correlated (ELI^{dn}) with placental invasiveness in endometrial stromal fibroblasts. (A) Histograms of the effect sizes (beta) of different binding sites, when tested on only ELI^{dn} and ELI^{up} genes. Only effect sizes for TFBSs that are statistically significant (p-value > 0) in the respective genes (ELI^{dn} or ELI^{up}) (B) Scatter plot shows the effect sizes (beta) of individual TFs for ELI^{dn} (y-axis) and ELI^{up} (x-axis) genes. Names of some selected TFBS's are labeled. Selected genes with a high fraction of variance explained by TFBS motif occurrences are shown as examples in the four tables next to each quadrant.

For each TF, we analyzed the effect size for either gene sets (ELI^{up} and ELI^{dn}) after normalizing for TFBS occurrence across species (**Figure 4B**). We found that while few TFs act as either enhancers or repressors for both ELI^{up} and ELI^{dn} genes (quadrants I, and III), there were also TFs which acted as antagonistic

regulators for ELI^{up} and ELI^{dn} genes (quadrants II, and IV). Tables show examples of TFs which regulate a given ELI^{up}, and ELI^{dn} gene either as an activator or a repressor. Indeed, for many genes, a large fraction of their inter-species variation is explained by a single TF. The statistical significance of the regulation by each TF for individual ELI genes, as well as for the two sets of 100 ELI^{up} and ELI^{dn} genes are listed in Supplementary Table 3.

Quadrant I lists TFs activating both ELI^{up} and ELI^{dn} genes, and includes many GATA factors, and MEF2 factors. All GATA factors, and MEF2 factors are key TFs regulating human and mouse ESF decidualization. We speculate that these TFs may regulate resistive genes in ESFs of epitheliochorial placentation, and also genes which are present in the more resistive decidualized fibroblasts in hemochorial species. Other TFs also suggest a link to decidualization and preparation for pregnancy including estrogen responsive factor, HOXC10³⁵, and early growth response gene 1 (EGR1)³⁶. EGR1 is induced by placenta growth factor and is an inducer of hypoxia inducible factor 1 (HIF1 α) necessary for early implantation³⁷. Quadrant III shows TFs which acted as repressors for both ELI^{up} and ELI^{dn} genes.

In contrast, quadrant II shows TFs which act as repressors for ELI^{up} genes, and activators for ELI^{dn} genes. Since ELI^{dn} genes are increased in species with epitheliochorial placentation, these TFs may act as important regulators promoting stromal resistance, by both activating resistive genes, as well as downregulating genes which potentially promote invasion. These TFs include POU2F2/3, SRY, Nkx3.2, and TBX19. Quadrant IV lists TFs, including SIX3, and E2F2 acting as antagonistic repressors of pro-invasibility ELI^{up} and activators of anti-invasibility ELI^{dn} genes.

Overall, these analyses provide a map of genomically encoded changes in stromal gene expression that have accompanied placental invasion across species. Presence of antagonistic transcriptional regulators,

which act oppositely on pro-invasibility ELI^{up} and anti-invasibility ELI^{dn} genes suggest that the vastly different phenotypic outcomes related to placentation are achieved partially by shared mechanisms. For example, through direct genomic binding, TBX19, acts as an activator for MRPS10, and a repressor for ANGEL1 associated with increased placental invasion. These antagonistic transcription factors also present attractive targets to limit invasive processes by gene knockdowns.

Discussion

The maternal fetal interface presents a highly diverse physiology among placental mammals with large inter-species differences in the extent of placental invasion into the endometrium, even though the central function of this interface of nutrient transfer from the mother to the fetus is broadly conserved³⁸. Phylogenetic analysis has proven that the non-invasive epitheliochorial placentation is a more recent innovation than the invasive hemochorial placentation^{6,39,40}. This has led to the rejection of the long held hypothesis that the hemochorial placentation, also present in humans, had evolved recently. We have advanced a hypothesis, termed “ELI” (Evolved Levels of Invasibility) stating that evolution of non-invasive epitheliochorial placentation was driven by changes in the endometrial stromal genotype, i.e. less invasive placentation is caused by higher stromal resistance to invasion¹³. Experimental validation proved that human stromal invasibility can be modulated by regulation of gene expression similar to that in epitheliochorial mammals. Here, we used an ordinal metric of placental invasiveness to identify candidate genes which are correlated with invasibility among eutherians using transcriptomic data of ESFs from 8 species, balanced for invasive and non-invasive placentation. Both a naïve linear model, and a phylogenetic adjustment to the model identified similar genes, which we termed as ELI^{up} and ELI^{dn} genes according to their correlation with placental invasibility phenotype.

Because the invasibility phenotype varies between species, one expects that genomic differences would inform inter-species transcriptomic differences in ESF between those presenting invasive and non-invasive placentation. Further support of a genomic basis for the phenotype of stromal invasibility was its secondary manifestation in other tissues, e.g. skin, wherein epitheliochorial species present lower rate of malignancy^{10,41}, which we have also experimentally validated¹³. In this work, we explored an important aspect of hard-coded genomic regulation, the abundance of transcription factor binding sites in promoter regions, and asked how well these explain the transcriptomic differences between eutherian species. We found that TFBS for individual transcription factors could, to a limited extent, explain inter-species variation in gene expression.

Indeed, TFBS explained a much larger fraction of inter-species variation in ELI^{up} and ELI^{dn} genes. For individual TF/target gene pairs, the cis-regulatory elements themselves could explain a large portion of expression changes along the trajectory of the evolution of reduced placental invasion. TFs explaining large ELI-correlated transcriptomic differences belonged to myocyte enhancer binding factor 2 (MEF2) families, and GATA-binding factor families. The binding sites for these TFs were also more abundant in species with stronger placental invasion, e.g. rodents and humans, with a concomitant effect on increased downstream gene expression. We speculate that this may be due to a more tissue specific phenomenon, decidualization, which occurs primarily in hemochorial species. MEF2 proteins, in particular MEF2B and MEF2D, are central players in muscle cell differentiation^{42,43}, cardiac hypertrophy, fibrosis⁴⁴, and stroma induced epithelial to mesenchymal transition of cancer⁴⁵. Similarly, GATA factors are also important for the endometrial decidualization program. GATA2 in fact, is a central factor in endothelial cell differentiation⁴⁶ and decidual differentiation in humans and mice^{14,33}. The stroma of hemochorial species may be more prepared to interact with invasive trophoblasts, and among primates, also be participants

in their endothelial transformation. All GATA factors found to have significant influence on invasibility in our study are also crucial for trophoblast development in hemochorial species³². Indeed, GATA2 has been reported to regulate genes associated with decidualization, and decidua-trophoblast interactions⁴⁷, including placental lactogen synthesis^{48,49}, proliferin⁴⁷, and prolactin-like protein-A⁵⁰. GATA5/6 are also important for cardiogenesis, similar to MEF2A and MEF2B. Whether these transcriptional regulators act as pro-invasive factors will need to be experimentally tested. Interestingly, GATA2 gene is alternatively spliced in non-placental species, with the -3.9kb regulatory element of the gene absent in non-eutherians³². Epitheliochorial placentation among ruminants is a more recent derivation than the ancestral hemochorial placentation. Therefore it is worth noting that it not the gain of binding sites for MEF2, GATAs and other TFs, but in fact the loss of these binding sites is found in species with reduced stromal invasion.

In addition to the TFs which acted as enhancers for pro-invasibility genes, we found many TFs which acted antagonistically on ELI^{up} and ELI^{dn} genes. Many of these antagonistic TFs (Figure 4D, quadrant II) acted as enhancers for anti-invasibility (ELI^{dn}) genes, while also acting as repressors for the pro-invasibility (ELI^{up}) genes. What do these TFs tell us about the evolution of epitheliochorial placentation? Because evolutionarily, epitheliochorial placentation is more recent, our analysis suggests that both, ELI^{dn} and ELI^{up} genes have recruited these transcription factor binding sites, but their effects on ELI^{dn} and ELI^{up} genes are opposite. Indeed, we also found examples of antagonism wherein certain TFs (e.g. E2F2 in Figure 4D, quadrant IV) act as repressors of ELI^{dn} genes, while activating ELI^{up} genes. Here, the target genes have lost binding sites for these TFs with the evolution of epitheliochorial placentation. We underline that while the effect of these antagonistic TFs is opposite upon the expression of ELI^{up} and ELI^{dn} genes, the effect on the phenotype of invasibility is consistent. TFs in quadrant II are expected to be

anti- invasibility and those in quadrant IV are expected to be pro- invasibility through their regulatory effect on ELI^{up} and ELI^{dn} genes within stroma. Our analysis presents attractive candidates to test for their effect on stromal invasibility. This analysis also provides a cue to understand the transcriptional mechanisms which have accompanied recent increase in placental invasiveness among the primates.

The ELI hypothesis posits that the stromal compartments of ruminants are particularly resistant to collective cell invasion, by trophoblast or by malignant cancer cells (e.g. skin). Here we have identified a number of candidate transcription factor genes that likely regulate invasibility of endometrial stromal cells, and potentially other stromal fibroblasts regulating cancer dissemination. Understanding the molecular functions of these transcription factors in stromal cell – trophoblast interaction will reveal the mechanisms that enable stromal cells to resist invasion by both, placental as well as cancer cells.

Methods

Endometrial stromal fibroblasts were isolated and sequenced from 9 species and gene expressions calculated for each sample. A set of 8,639 orthologous gene sets with members in each of the species were identified for analysis. The expression of these genes was calculated in transcripts per million (TPM).

Finding genes related to the modulation of placental invasibility

We obtained the gene expression $e_{g,s}$ of gene g in species s in transcripts per million. For each species, we also rated the degree of placental invasibility r_s . In order to find the stromal genes likely involved in modulating the degree of invasibility, we fit the linear model

$$r_s = \gamma_g e_{g,s} + \epsilon_{g,s},$$

where $\epsilon_{g,s}$ is the residual. This is fitted independently for each gene, and positive γ_g represent genes that are up-regulated in species with more invasive placenta, while negative γ_g indicates the opposite. The statistical significance is calculated for the null hypothesis of $\gamma_g = 0$. The p-values for all genes are corrected for multiple testing using the false discovery rate method. In addition, this linear model was also fitted using a phylogenetic linear model, with an assumption of Brownian motion processes producing both the traits (gene expression and species invasibility). The phylogenetic linear model was fitted using the phylolm package¹⁶.

Genomic explanation of transcriptomic variation

For fitting the variation in gene expression between species to the cis-regulatory elements, we first calculated a scaled version of the gene expression by first calculating the square roots and then

subtracting the mean expression of each gene across all species. Using the gene expression $e_{g,s}$ for gene g in species s , we calculate the scaled gene expression as

$$h_{g,s} = \sqrt{e_{g,s}} - \frac{1}{N_{\text{Species}}} \sum_{s'} \sqrt{e_{g,s'}}$$

Known eukaryotic transcription factor binding site motifs were downloaded from the JASPAR database¹⁷. Genomic sequence FASTA files were downloaded from the Ensembl database⁵¹. We identified genomic regions 5kb upstream to 1kb downstream of the translation start site of each gene as purported promoter regions. We searched for matches to each known binding site motif in each of these promoter regions using the FIMO package⁵² using default parameters. The number of statistically significant matches (p-value $< 10^{-4}$) for each binding site motif in each gene's promoter regions were assembled in a matrix. We represent the number of matches to the binding site motif of transcription factor t in the promoter region of gene g in species s as $n_{t,g,s}$. Then, independently for each transcription factor binding site motif, a linear model was fit for a set of all genes (or all ELI genes). The linear model related the gene expression between species to the differing number of binding site matches found in the promoter regions of the orthologs of the same gene across all 9 species. The inter-species variation in gene expression was modeled to relate to the occurrences of the transcription factor binding sites as

$$h_{g,s} = \beta_t n_{t,g,s} + \epsilon_{t,g,s}$$

where $\epsilon_{t,g,s}$ is the residual and β_t is the fitted coefficient for a particular transcription factor binding site motif t . A model is fitten independently for each motif t . Again, a p-value is obtained for the null hypothesis $\beta_t = 0$. The coefficient (β_t) of this linear regression model, and its statistical significance (p-value) was recorded for each transcription factor motif.

References

- 1 Ramsey, E. M. *The Placenta: Human and Animal*. (Praeger Inc., 1982).
- 2 Carter, A. M. Evolution of placental function in mammals: the molecular basis of gas and nutrient transfer, hormone secretion, and immune responses. *Physiol Rev* **92**, 1543-1576, doi:10.1152/physrev.00040.2011 (2012).
- 3 Mess, A. Placental Evolution within the Supraordinal Clades of Eutheria with the Perspective of Alternative Animal Models for Human Placentation. *Advances in Biology*, 639274, doi:10.1155/2014/639274 (2014).
- 4 Soma, H. *et al.* Review: Exploration of placentation from human beings to ocean-living species. *Placenta* **34 Suppl**, S17-23, doi:10.1016/j.placenta.2012.11.021 (2013).
- 5 Gundling, W. E., Jr. & Wildman, D. E. A review of inter- and intraspecific variation in the eutherian placenta. *Philos Trans R Soc Lond B Biol Sci* **370**, 20140072, doi:10.1098/rstb.2014.0072 (2015).
- 6 Wildman, D. E. *et al.* Evolution of the mammalian placenta revealed by phylogenetic analysis. *Proc Natl Acad Sci U S A* **103**, 3203-3208, doi:10.1073/pnas.0511344103 (2006).
- 7 Flenker, H. [Orthology and pathology of the human placenta. 2. Pathology of the placenta]. *Fortschr Med* **92**, 381-386 (1974).
- 8 Carter, A. M. & Pijnenborg, R. Evolution of invasive placentation with special reference to non-human primates. *Best Pract Res Clin Obstet Gynaecol* **25**, 249-257, doi:10.1016/j.bpobgyn.2010.10.010 (2011).
- 9 Silva, J. F. & Serakides, R. Intrauterine trophoblast migration: A comparative view of humans and rodents. *Cell Adh Migr* **10**, 88-110, doi:10.1080/19336918.2015.1120397 (2016).
- 10 D'Souza, A. W. & Wagner, G. P. Malignant cancer and invasive placentation: A case for positive pleiotropy between endometrial and malignancy phenotypes. *Evol Med Public Health* **2014**, 136-145, doi:10.1093/emph/eou022 (2014).
- 11 Costanzo, V., Bardelli, A., Siena, S. & Abrignani, S. Exploring the links between cancer and placenta development. *Open Biol* **8**, doi:10.1098/rsob.180081 (2018).
- 12 Murray, M. J. & Lessey, B. A. Embryo implantation and tumor metastasis: common pathways of invasion and angiogenesis. *Semin Reprod Endocrinol* **17**, 275-290, doi:10.1055/s-2007-1016235 (1999).
- 13 Kshitiz *et al.* Evolution of placental invasion and cancer metastasis are causally linked. *Nat Ecol Evol* **3**, 1743-1753, doi:10.1038/s41559-019-1046-4 (2019).
- 14 Kin, K., Nnamani, M. C., Lynch, V. J., Michaelides, E. & Wagner, G. P. Cell-type phylogenetics and the origin of endometrial stromal cells. *Cell Rep* **10**, 1398-1409, doi:10.1016/j.celrep.2015.01.062 (2015).
- 15 Garratt, M., Gaillard, J. M., Brooks, R. C. & Lemaitre, J. F. Diversification of the eutherian placenta is associated with changes in the pace of life. *Proc Natl Acad Sci U S A* **110**, 7760-7765, doi:10.1073/pnas.1305018110 (2013).
- 16 Ho, L. & Ane, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst Biol* **63**, 397-408, doi:10.1093/sysbio/syu005 (2014).
- 17 Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D1284, doi:10.1093/nar/gkx1188 (2018).
- 18 Wang, X., Sato, R., Brown, M. S., Hua, X. & Goldstein, J. L. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. *Cell* **77**, 53-62, doi:10.1016/0092-8674(94)90234-8 (1994).
- 19 Zelenko, Z., Aghajanova, L., Irwin, J. C. & Giudice, L. C. Nuclear receptor, coregulator signaling, and chromatin remodeling pathways suggest involvement of the epigenome in the steroid hormone response of endometrium and abnormalities in endometriosis. *Reprod Sci* **19**, 152-162, doi:10.1177/1933719111415546 (2012).
- 20 Chu, K. & Zingg, H. H. The nuclear orphan receptors COUP-TFII and Ear-2 act as silencers of the human oxytocin gene promoter. *J Mol Endocrinol* **19**, 163-172, doi:10.1677/jme.0.0190163 (1997).

- 21 *Systems Biology of Livestock sciences*. (John Wiley & Sons, 2011).
- 22 Bazer, F. W., Burghardt, R. C., Johnson, G. A., Spencer, T. E. & Wu, G. Interferons and progesterone for establishment and maintenance of pregnancy: interactions among novel cell signaling pathways. *Reprod Biol* **8**, 179-211, doi:10.1016/s1642-431x(12)60012-6 (2008).
- 23 Tan, J., Paria, B. C., Dey, S. K. & Das, S. K. Differential uterine expression of estrogen and progesterone receptors correlates with uterine preparation for implantation and decidualization in the mouse. *Endocrinology* **140**, 5310-5321, doi:10.1210/endo.140.11.7148 (1999).
- 24 Spencer, T. E. *et al.* Effects of recombinant ovine interferon tau, placental lactogen, and growth hormone on the ovine uterus. *Biol Reprod* **61**, 1409-1418, doi:10.1095/biolreprod61.6.1409 (1999).
- 25 Bazer, F. W. *et al.* Novel pathways for implantation and establishment and maintenance of pregnancy in mammals. *Mol Hum Reprod* **16**, 135-152, doi:10.1093/molehr/gap095 (2010).
- 26 Kelleher, A. M. *et al.* Forkhead box a2 (FOXA2) is essential for uterine function and fertility. *Proc Natl Acad Sci U S A* **114**, E1018-E1026, doi:10.1073/pnas.1618433114 (2017).
- 27 Walker, C. J. *et al.* MAX Mutations in Endometrial Cancer: Clinicopathologic Associations and Recurrent MAX p.His28Arg Functional Characterization. *J Natl Cancer Inst* **110**, 517-526, doi:10.1093/jnci/djx238 (2018).
- 28 Llabata, P. *et al.* Multi-Omics Analysis Identifies MGA as a Negative Regulator of the MYC Pathway in Lung Adenocarcinoma. *Mol Cancer Res* **18**, 574-584, doi:10.1158/1541-7786.MCR-19-0657 (2020).
- 29 Pimkin, M. *et al.* Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res* **24**, 1932-1944, doi:10.1101/gr.164178.113 (2014).
- 30 Reiter, J. F. *et al.* Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes Dev* **13**, 2983-2995, doi:10.1101/gad.13.22.2983 (1999).
- 31 Siletz, A., Kniazeva, E., Jeruss, J. S. & Shea, L. D. Transcription factor networks in invasion-promoting breast carcinoma-associated fibroblasts. *Cancer Microenviron* **6**, 91-107, doi:10.1007/s12307-012-0121-z (2013).
- 32 Bai, H., Sakurai, T., Godkin, J. D. & Imakawa, K. Expression and potential role of GATA factors in trophoblast development. *J Reprod Dev* **59**, 1-6, doi:10.1262/jrd.2012-100 (2013).
- 33 Rubel, C. A. *et al.* A Gata2-Dependent Transcription Network Regulates Uterine Progesterone Responsiveness and Endometrial Function. *Cell Rep* **17**, 1414-1425, doi:10.1016/j.celrep.2016.09.093 (2016).
- 34 Carter, A. M. Comparative studies of placentation and immunology in non-human primates suggest a scenario for the evolution of deep trophoblast invasion and an explanation for human pregnancy disorders. *Reproduction* **141**, 391-396, doi:10.1530/REP-10-0530 (2011).
- 35 Ansari, K. I., Hussain, I., Kasiri, S. & Mandal, S. S. HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4. *J Mol Endocrinol* **48**, 61-75, doi:10.1530/JME-11-0078 (2012).
- 36 Guo, B. *et al.* Expression, regulation and function of Egr1 during implantation and decidualization in mice. *Cell Cycle* **13**, 2626-2640, doi:10.4161/15384101.2014.943581 (2014).
- 37 Patel, N. & Kalra, V. K. Placenta growth factor-induced early growth response 1 (Egr-1) regulates hypoxia-inducible factor-1alpha (HIF-1alpha) in endothelial cells. *J Biol Chem* **285**, 20570-20579, doi:10.1074/jbc.M110.119495 (2010).
- 38 Martin, R. D. Evolution of Placentation in Primates: Implications of Mammalian Phylogeny. *Evolutionary Biology* **35**, 125-145 (2008).
- 39 Mess, A. & Carter, A. M. Evolutionary transformation of fetal membrane characters in Eutheria with special reference to Afrotheria. *Journal of Experimental Zoology Part B* **306B**, doi:10.1002/jez.b.21079 (2005).
- 40 Elliot, M. G. & Crespi, B. J. Phylogenetic evidence for early hemochorial placentation in eutheria. *Placenta* **30**, 949-967, doi:10.1016/j.placenta.2009.08.004 (2009).

- 41 Wagner, G. P., Kshitiz & Levchenko, A. Comments on Boddy et al., 2020: Available data suggests positive relationship between placental invasion and malignancy. *Evolution, Medicine, & Public Health*, eoaa024, doi:10.1093/emph/eoaa024 (2020).
- 42 Karamboulas, C. *et al.* Disruption of MEF2 activity in cardiomyoblasts inhibits cardiomyogenesis. *J Cell Sci* **119**, 4315-4321, doi:10.1242/jcs.03186 (2006).
- 43 Molkentin, J. D. *et al.* MEF2B is a potent transactivator expressed in early myogenic lineages. *Mol Cell Biol* **16**, 3814-3824, doi:10.1128/mcb.16.7.3814 (1996).
- 44 Gossett, L. A., Kelvin, D. J., Sternberg, E. A. & Olson, E. N. A new myocyte-specific enhancer-binding factor that recognizes a conserved element associated with multiple muscle-specific genes. *Mol Cell Biol* **9**, 5022-5033, doi:10.1128/mcb.9.11.5022 (1989).
- 45 Su, L. *et al.* MEF2D Transduces Microenvironment Stimuli to ZEB1 to Promote Epithelial-Mesenchymal Transition and Metastasis in Colorectal Cancer. *Cancer Res* **76**, 5054-5067, doi:10.1158/0008-5472.CAN-16-0246 (2016).
- 46 Kshitiz *et al.* Matrix rigidity controls endothelial differentiation and morphogenesis of cardiac precursors. *Sci Signal* **5**, ra41, doi:10.1126/scisignal.2003002 (2012).
- 47 Ma, G. T. *et al.* GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. *Development* **124**, 907-914 (1997).
- 48 Liang, R., Limesand, S. W. & Anthony, R. V. Structure and transcriptional regulation of the ovine placental lactogen gene. *Eur J Biochem* **265**, 883-895, doi:10.1046/j.1432-1327.1999.00790.x (1999).
- 49 Kim, G. S. *et al.* Identification of trophoblast-specific binding sites for GATA-2 that are essential for rat placental lactogen-I gene expression. *Biotechnol Lett* **31**, 1173-1181, doi:10.1007/s10529-009-9994-4 (2009).
- 50 Ma, G. T. & Linzer, D. I. GATA-2 restricts prolactin-like protein A expression to secondary trophoblast giant cells in the mouse. *Biol Reprod* **63**, 570-574, doi:10.1095/biolreprod63.2.570 (2000).
- 51 Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res* **48**, D682-D688, doi:10.1093/nar/gkz966 (2020).
- 52 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

Supplementary Table 1: Scoring genes to the Evolved Levels of Invasibility (ELI), in terms of correlation and p-values of gene expression with endometrial invasibility.

Supplementary Table 2: Results of statistical testing each transcription factor binding site (TFBS) for their effects on the global inter-species gene expression profile.

Supplementary Table 3: Transcription factors found to be regulating pro-invasive (ELI-up) and anti-invasive (ELI-dn) genes.