DEVELOPMENTAL BIOLOGY

A genome scale transcriptional regulatory model of the human placenta

Alison Paquette[1,2]*, Kylia Ahuna[3], Yeon Mi Hwang[4], Jocelynn Pearl[5], Hanna Liao[1], Paul Shannon[4], Leena Kadam[3], Samantha Lapehn[2], Matthew Bucher[3], Ryan Roper[4], Cory Funk[4], James MacDonald[1], Theo Bammler[1], Priyanka Baloni[4]†, Heather Brockway[6]‡, W. Alex Mason[7], Nicole Bush[8], Kaja Z. Lewinn[8], Catherine J. Karr[1], John Stamatoyannopoulos[1,9], Louis J. Muglia[10,11], Helen Jones[12], Yoel Sadovsky[13,14], Leslie Myatt[3], Sheela Sathyanarayana[1,2], Nathan D. Price[4,15]

Gene regulation is essential to placental function and fetal development. We built a genome-scale transcriptional regulatory network (TRN) of the human placenta using digital genomic footprinting and transcriptomic data. We integrated 475 transcriptomes and 12 DNase hypersensitivity datasets from placental samples to globally and quantitatively map transcription factor (TF)–target gene interactions. In an independent dataset, the TRN model predicted target gene expression with an out-of-sample R^2 greater than 0.25 for 73% of target genes. We performed siRNA knockdowns of four TFs and achieved concordance between the predicted gene targets in our TRN and differences in expression of knockdowns with an accuracy of >0.7 for three of the four TFs. Our final model contained 113,158 interactions across 391 TFs and 7712 target genes and is publicly available. We identified 29 TFs which were significantly enriched as regulators for genes previously associated with preterm birth, and eight of these TFs were decreased in preterm placentas.

INTRODUCTION

The placenta is an essential organ with multiple functions including regulating gas, nutrient, and waste transport and providing immunological defense (1). It also performs endocrine functions, produces key hormones, and coordinates the passage of hormonal and paracrine signals between the mother and developing fetus (2). Thus, the placenta influences fetal development, growth, and health (3). This complex organ is composed of various types of trophoblast cells, including extravillous trophoblasts, cytotrophoblasts, and the syncytiotrophoblasts. Syncytiotrophoblasts are a large terminally differentiated, multinucleated cell type that is the primary site of nutrient uptake/transport as well as peptide and steroid hormone synthesis. The placenta is composed of additional cell types such as stromal cells, fibroblasts, endothelial cells, and various immune cells (4). With its many cell types and highly specialized functions, the placenta has a unique transcriptome (5) which includes numerous genes expressed exclusively within this tissue. The placental transcriptome changes throughout gestation (6, 7) and in response to cues from the maternal environment (8). In this way, the placental transcriptome reflects the underlying physiological functions of the organ. Characterizing transcriptomic differences between normal and pathological placental tissue can provide insight into the molecular mechanisms involved in placental pathophysiology. Perturbations to the placental transcriptome are linked to pregnancy pathologies including preterm birth (9) and preeclampsia (10). Previous studies of individual genes have revealed several temporally regulated and cell type–specific transcription factors (TFs) that play key roles in discrete developmental stages and functions of the placenta. Throughout pregnancy, TFs including *GCM1*, *GATA2/3*, and *CREB* emerge as unique regulators of specific trophoblast subtypes (11). These TFs regulate key placental processes such as invasion and nutrient transport that may be linked to adverse placental function and prenatal complications. Further, understanding these transcriptional changes in normal and pathological placentas can provide insights into the molecular mechanisms underlying gestational pathophysiology.

There is a large and active field of research by the genome science community involving the generation and analysis of transcriptomic data using transcriptional regulatory networks (TRNs) (12–15), which map TFs to the target genes they regulate (16). A TRN is made up of nodes (TFs and their target genes) and edges, which explain which TFs regulate which target genes. We have defined TRN edges based on TF binding within regions of open chromatin and relationships in gene expression between TFs and target genes. TRNs are organized into units known as "regulons," that are composed of each individual TF and the target genes they regulate. Our team has previously constructed TRNs by integrating information about chromatin accessibility captured through deoxyribonuclease I hypersensitive sites sequencing (DNase-seq) or Assay for transposase-accessible chromatin using sequencing and TF binding using sequence motifs from a compendium of TFs and RNA sequencing (RNA-seq) (17). We have generated a similar network in the human brain (18), which revealed TFs that were regulated by viral abundance and associated with Alzheimer's disease (19), TFs that drive gene expression changes associated with bipolar disorder and schizophrenia (18), regulatory functions for single-nucleotide polymorphisms (SNPs) associated with genetic risks of these

[1]University of Washington, Seattle, WA, USA. [2]Seattle Children's Research Institute, Seattle, WA, USA. [3]Oregon Health and Sciences University, Portland, OR, USA. [4]Institute for Systems Biology, Seattle, WA, USA. [5]Tune Therapeutics, Seattle, WA, USA. [6]Department of Physiology and Aging, University of Florida, Gainesville, FL, USA. [7]University of Tennessee Health Sciences Center, Memphis, TN, USA. [8]University of California San Francisco, San Francisco, CA, USA. [9]The Altius Institute, Seattle, WA, USA. [10]The Burroughs Wellcome Fund, Research Triangle Park, NC, USA. [11]Cincinnati Children's Hospital Medical Center and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. [12]University of Florida, Gainesville, FL, USA. [13]Magee Womens Research Institute, Pittsburgh, PA, USA. [14]University of Pittsburgh, Pittsburgh, PA, USA. [15]Thorne HealthTech, New York City, NY, USA.
*Corresponding author. Email: alison.paquette@seattlechildrens.org
†Present address: School of Health Sciences, Purdue University, West Lafayette, IN, USA.
‡Present address: Center for Scientific Review, National Institutes of Health, Bethesda, MD, USA.

disorders (*18*), and TF–target gene relationships underlying gene expression changes related to Huntington's disease (*20*). Our brain TRN model has demonstrated that quantifying transcriptional networks has the potential to illuminate the etiology of a wide variety of pathological disorders. In the present work, we construct a placental-specific TRN, based on its important role in fetal development and distinct transcriptional regulation compared to other tissues.

Other network biology-based approaches have been previously deployed to gain insight into placental function and prenatal disease, such as Weighted gene coexpression network analysis (WGCNA). WGCNA is a network strategy that uses coexpression clustering to identify coregulated gene clusters based on correlated expression (*21*). A WGCNA of the human placental transcriptome collected at delivery revealed gene clusters involved in core placental functions that were reproducible across trimesters (in mice) (*22*). An independent WGCNA constructed on a large number of term placenta transcriptomes revealed modules associated with fetal growth restriction (FGR) (*23*). TRNs are distinct from other network databases, such as the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) protein interaction database, which collects and integrates physical interactions and coexpression data with curated text mining of scientific literature, because this network is not tissue specific and does not include any information about transcriptional regulation (*24*). The TRN is also distinct from tissue-specific networks such as the HumanBase database, which is a compendium of tissue-specific genome-scale functional networks created by integrating hundreds of publicly available RNA-seq datasets, and interaction networks defined using data-driven Bayesian methodologies (*25*). Our placental TRN implements constraints of coexpression, TF binding, and chromatin accessibility using highly curated placental datasets, therefore is more conservative. Combining coexpression analyses with chromatin accessibility data is essential for a deeper understanding of the underlying transcriptional mechanisms of regulation. A placenta TRN constructed using DNase-seq data derived from first-trimester placental tissue has been used to model the transformation of embryonic stem cells to trophoblast cells, providing key insights into TFs which regulate genes with regions of open chromatin in this context (*26*). An accurate model of the placental transcriptome at the end of gestation, enhanced with DNase-seq data, will provide valuable insight into TF–target gene interactions and serve as a tool for placental gene expression analyses.

Our goal was to construct the first placental-specific genome scale TRN using DNase-seq data and curated TF binding motif databases paired with RNA-seq data compiled from 475 placental samples. We applied a machine learning approach to evaluate the accuracy of the model at predicting gene expression and performed experimental validation of transcriptional regulation of a core group of TFs in isolated villous cytotrophoblasts. Because previous transcriptomic studies identified fetal sex as a critical biological variable that affects the placental transcriptome (*27*) and perinatal outcomes (*28*), we also characterized sex-specific differences in TF–target gene interactions within the placenta. We hypothesized that our network could reveal core TFs that regulated genes whose placental expression was altered in the context of prematurity. To test this hypothesis, we reanalyzed a list of genes whose placental expression was previously associated with preterm birth in an independent analysis (*29*) to identify the TFs that regulated these genes. We also identified specific TFs that regulated specific placental cell types, based on signatures from recently published single-cell RNA-seq data (*30*). These examples demonstrate how this approach can be broadly applied to other types placental omics data (e.g., DNA methylation data), providing researchers a different way to interpret lists of candidate genes and identify transcriptional regulators.

RESULTS

Reconstruction of a transcriptional regulatory network model for the human placenta

Our placental TRN was constructed using transcriptomic data generated from 475 placental samples collected in two independent birth cohorts: (i) participants from Magee Womens Research Institute (MWRI) in Pittsburgh, PA and at the University of Tennessee Medical Center in Memphis, TN; and (ii) participants from the Conditions Affecting Cognitive Development and Learning in Early Childhood (CANDLE) study. DNase-seq data were generated in 12 term human placentas collected as part of the Oregon Health and Sciences Placental Biorepository. Covariate information for all placental samples used within the model can be found in Table 1. We excluded samples from pregnancies with medical indications for preterm delivery, as well as multifetal gestations to capture the placental transcriptome of infants without medically complicated deliveries. Overall, the demographics of RNA-seq samples in this dataset was 46% Black, 52% white, and 2% other race groups. There were some intrinsic differences in the demographic and clinical information between the cohorts. The CANDLE cohort predominantly self-identified as Black (56%); whereas the samples from MWRI predominantly self-identified as white (85%). The mean gestational age of the MWRI samples was 39.6 weeks, and the mean gestational age of the CANDLE cohort was 39.2 weeks. A subset (3%) of the samples in the CANDLE cohort were born spontaneously prematurely (<37 weeks), and we have previously published transcriptomic signatures from these spontaneous preterm deliveries (*29*). These spontaneous preterm samples were included in the model to capture variation in the placental transcriptome related to gestational age. This is important for future studies that may use this model to study a variety of pregnancy pathologies and with samples from varying gestational ages.

Model construction and evaluation in testing dataset

To construct our genome scale TRN, genomic footprints derived from placental DNase-seq data (Fig. 1A) were paired with known TF binding motifs (Fig. 1B) to generate putative edges between TFs and target genes (Fig. 1C). Prior work by our group demonstrated that the digital genomic footprinting workflow used in this analysis achieved a Matthews correlation coefficient of 0.44 when compared to chromatin immunoprecipitation sequencing (ChIP-seq) data in matched cell types (*17*). These results indicated that footprinting data alone did not improve accuracy but did improve robustness of predictions. There is sparse ChIP-seq data for TFs available within placental tissue, which limits our ability to perform this type of calculation (*17*). Instead, we rely on prediction of target gene expression in an independent holdout dataset to assess model accuracy. Next, we generated a coexpression network of TFs and target genes in our training dataset using these putative edges (Fig. 1D). We then performed a series of Lasso regression models to evaluate how different parameters of the model influenced the overall model size and ability to predict gene expression in our test dataset (Fig. 1E). We constructed six models using Lasso regression with different parameters for our training dataset,

Table 1. Covariate data from cohorts used for RNA-seq and DNase-seq data. Exact values are not reported for cells with <5 participants. BMI, body mass index.

|	MWRI–RNA-seq			CANDLE–RNA-seq			OHSU–DNase-seq		
	Males (52)	**Females (52)**	**Total (104)**	**Males (176)**	**Females (176)**	**Total (371)**	**Males (6)**	**Females (6)**	**Total (12)**
Delivery method (*N*)									
C-section	15	14	29	66	74	140	6	6	12
Vaginal	37	38	75	110	121	231	0	0	0
Maternal race (*N*)									
Black	7	4	11	99	110	209	0	<5	<5
White	41	47	88	75	82	157	5	5	10
Other	<5	<5	5	<5	<5	5	<5	0	<5
Gestational age (weeks)									
Range	37.43–41.71	37–41.14	37–41.71	36–41.43	34.29–41.29	34.29–41.43	38.57–39.29	38.29–39.57	38.29–39.57
Median	39.64	39.86	39.86	39.29	39.14	39.29	39.07	39.29	39.14
Birth weight (grams)									
Range	2755–4801	2830–4348	2755–4801	2407–4685	2097–4410	2097–4685	3280–4167	3220–4465	3220–4465
Median	3520.5	3359.5	3477.5	3420	3250	3311	3545	3657.5	3565
Maternal BMI (kg/m²)									
Range	17.33–36.94	18.19–37.57	17.38–37.57	14.2–52.4	14.2–56.8	14.2–56.8	19.3–25	18.3–25.27	18.3–25.27
Median	23.39	23.02	23.03	25.85	25.70	25.80	23.91	23.06	23.82
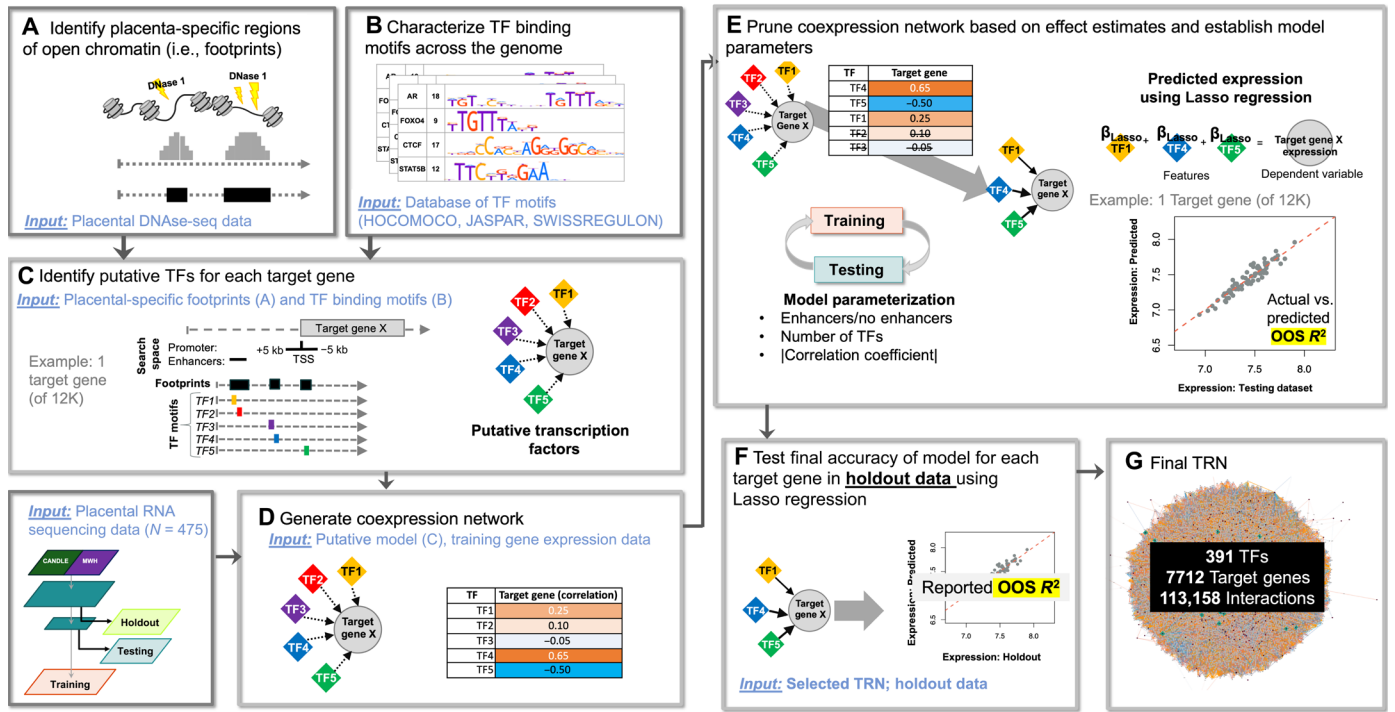
Fig. 1. Overview of data and strategy used to construct TRN. (**A**) First, regions of open chromatin, i.e., footprints are identified from placental DNase-seq data. (**B**) Concurrently, TF binding motifs were curated from publicly available databases. (**C**) These data were overlayed (C) to identify putative regulators (TFs) of each target gene. (**D**) We then generated a coexpression network within these putative TF–target gene relationships using RNA-seq data. (**E**) We evaluated different parameters to prune the model in the training data. Lasso regression was used with selected TFs for each target gene to predict expression, and then the actual versus predicted expression was used to evaluate the model. (**F**) The final model was tested in our holdout data to generate final accuracy for each gene. (**G**) Only target genes we could predict with high confidence were included in our final model. TSS, transcriptional start site.

which are summarized in table S1. The accuracy of each model was calculated as the R^2 value of the correlation coefficient (R) between the actual and predicted gene expression in the dataset generated from Lasso regression (out of sample R^2 or OOS R^2). The full range of predictions for each individual gene is depicted in fig. S1.

The effects of three parameter choices were evaluated: (i) including or excluding enhancers (model 1 versus model 2), (ii) changing the correlation cutoff (model 1 versus models 3 and 4), and (iii) changing the number of TFs (model 1 versus models 5 and 6) (see table S1 and Fig. 1E). We compared these models to the "reference model" (model 1), where we used the median correlation coefficient and included only the top 15 TFs that were correlated with expression of each target gene in our model. Overall, the two parameters that altered the model the most were (i) varying the cutoff for the correlation coefficient and (ii) the number of TFs included in each model (fig. S1). There were minimal differences in the overall model size or accuracy based on varying the maximum number of TFs in the model. Models with a more stringent cutoff contained fewer TFs, target genes, and interactions (models 4 and 5). Model 4 used a higher correlation coefficient cutoff (correlation coefficient > |0.5|) and overall had the highest accuracy of all the models tested (98% of samples had OOS $R^2 > 0.25$ in test dataset). There was no significant difference in the OOS R^2 values between models constructed with and without enhancers ($P = 1$, Tukey test; see table S2). We constructed our final model with enhancers to provide complete coverage of potential transcriptional regulators. We elected to construct our final model using a correlation cutoff of 0.25, as this model was among the more comprehensive in terms of interactions and TFs and had the highest accuracy outside of model 4. Thus, only this one model was tested against the final hold out dataset (which was never included in the process of model building and selection), minimizing overfitting from model selection.

We also generated a series of "null" models that we compared to our initial model (model 1) to estimate (i) how our TRN performed compared to the background null distribution and (ii) that the addition of TF binding sites and TFs with strong correlations with target genes improved the accuracy of our Lasso regression models. These three null models (described in more detail in table S3) included models in the presence or absence of TF motifs with and without correlation coefficients above the absolute value of the median correlation cutoff. We evaluated the accuracy of these models to predict expression in our test dataset again using Lasso regression, and we compared the results of these Lasso regression models to the results from model 1 (fig. S1). All three of these null models were significantly less accurate at predicting gene expression in the test dataset than any of our actual models ($P < 0.05$, Tukey test; table S2). The null models including TFs with weak correlations with their target genes (<median correlation coefficient; models 7 and 9) had the lowest accuracy of predicting gene expression in the test dataset (mean OOS R^2, 0.2871 and 0.079; fig. S1 and table S3). Model 8, which predicted gene expression using TFs without binding sites at the specific region but with stronger correlations with target genes (>median correlation coefficient), had comparable accuracy to model 7 (OOS R^2, 0.2831) but was still significantly lower than all of the primary models (Tukey test; see table S2 and fig. S1).

Computational validation of model accuracy in holdout and independent dataset

After model parameterization in the training dataset, we then combined the training and testing dataset ($N = 382$) and used this as input data to calculate our final accuracy in the holdout dataset ($N = 93$), again using Lasso regression (Fig. 1F). This step is critical as a guard against overfitting as this holdout dataset was completely separate and not used during the construction of the initial variations in the model and was selected randomly from samples at the beginning of the project (see Fig. 1 and fig. S2). In the holdout data, 88% of TF–target gene correlations were statistically significant after adjusting for multiple comparisons [false discovery rate (FDR) <0.05], and 83% of TF–target gene relationships were highly correlated with a correlation coefficient of >|0.25| (Fig. 2A). Across all genes, Lasso regression using the TFs identified in our model predicted target gene expression with a mean OOS R^2 of 0.39, with 73.35% of target genes achieving an OOS $R^2 > 0.25$. We constructed our final TRN using only target genes which we were able to predict with an OOS $R^2 > 0.25$ in our holdout dataset, removing 2802 genes from our final model. All remaining target gene Lasso regression models were also statistically significant after adjusting for multiple comparisons (FDR adjusted $P < 0.05$). We separately evaluated the reproducibility of our model in an independent placental transcriptomic dataset from the CANDLE cohort ($N = 366$) that was generated after model construction had been completed. In this separate dataset, 94.7% of TF–target gene correlations were statistically significant after adjustment for multiple comparisons (FDR <0.05), and 72.34% of TF–target gene relationships were highly correlated with a correlation coefficient of >|0.25| (see fig. S3). In this independent dataset, Lasso regression using the TFs predicted target gene expression with a mean OOS R^2 of 0.40, with 73.0% of target genes achieving an OOS R^2 of >0.25. Thus, results were comparable in the original holdout dataset and the newer independent dataset.

Model description

Our final robust model contained 7712 genes, with a subset of this model including 3213 genes (41.7%) with OOS R^2 prediction accuracy in our holdout dataset >0.5 (calculated from Lasso regression; Fig. 2B). Our final model was composed of 113,158 interactions between 391 TFs and 7712 target genes, including 75,014 positively correlated interactions and 38,144 negatively correlated interactions (Fig. 3A). In our full model, the effect estimates (correlation coefficients) ranged from −0.75 to −0.25 for the negative regulators and 0.25 to 0.93 for the positive regulators. The median regulon size (target genes per TF) was 156 and was heavily tailed with 75 "hub" TFs which regulated >500 genes (Fig. 3B). Network centrality measures are provided in table S4. The eight most central TFs in our model based on eigenvector centrality (measure of influence of node within a network for each TF) were *SP1*, *SMAD4*, *SP3*, *CTCF*, *STAT5B*, *NR2C2*, *YBX1*, and *CLOCK*. Our model contained 25 TFs that were placenta-specific or placenta enriched, based on definitions from the Human Protein Atlas (*31*), including known placental transcriptional regulators such as *GCM1*, *DLX3*, *DLX4*, and *DLX5* (see table S5). The largest placental enriched transcriptional regulator was *TFAP2A*, which regulated 1255 different target genes.

Experimental validation of model accuracy

We performed an small interfering RNA (siRNA) knockdown of four TFs (*GRHL2*, *CREB3L2*, *ESRRG*, and *GCM1*) in primary villous trophoblast cells. These four genes were selected because they were characterized as "placenta-enriched genes" from the Human Protein Atlas (*31*) and because they regulated >125 target genes in our model. TF knockdown was performed in seven female and six male samples
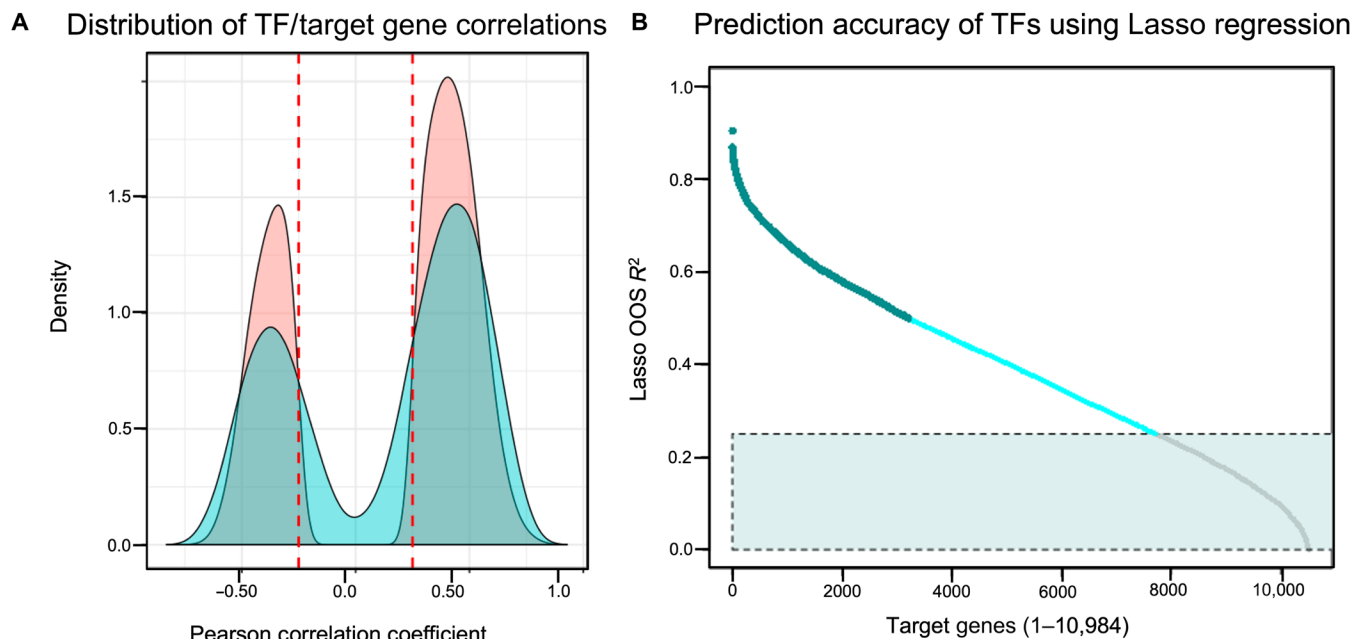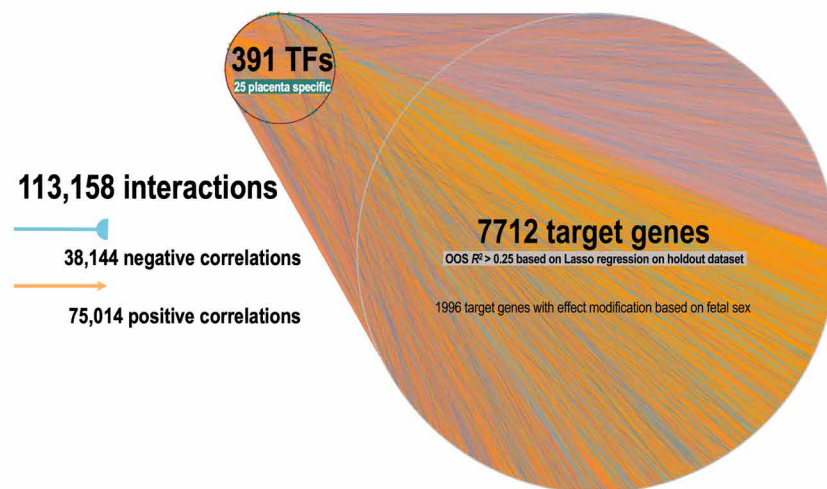
A Distribution of TF/target gene correlations   **B** Prediction accuracy of TFs using Lasso regression

Fig. 2. Accuracy of TRN model to predict target gene expression in holdout dataset. (**A**) Distribution of Pearson correlation coefficients in training + testing dataset ($N = 382$ samples, pink) and holdout dataset ($N = 93$, blue). Red lines indicate 0.25, which was the cutoff in the training dataset. (**B**) Distribution of out of sample accuracy (OOS R^2) for each of the 10,984 genes calculated in our holdout dataset. Light blue box indicates genes with OOS $R^2 < 0.25$, which were genes not included in our final model. Dark blue genes were high confidence OOS R^2.

A Global Summary of TRN   **B** Network connectivity of 391 TFs
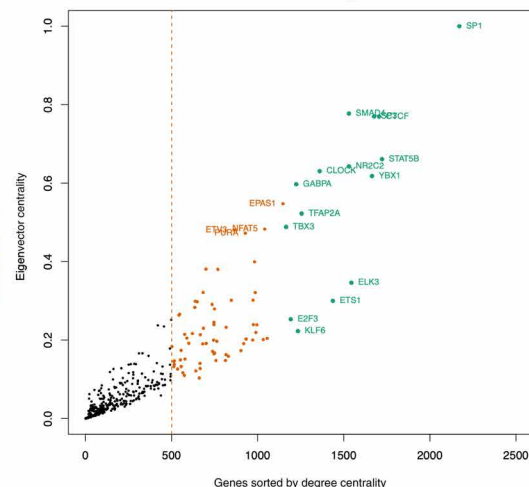
Fig. 3. Summary of genome scale TRN. (**A**) Network diagram and (**B**) network connectivity of 391 TFs based on their target genes. Orange line indicates hub TFs with >500 genes. Top 15 TFs based on degree centrality are shown in green

(with the exception of *GCM1*, which had four male samples, and *GRHL2*, which had five male samples). Gene expression for predicted target genes was quantified through RNA-seq, and differentially expressed genes (DEGs) (unadjusted $P < 0.05$) are summarized in fig. S4. We generated model accuracy values by comparing number of instances of "true positives" and "true negatives" out of the total number of target genes based on the directionality of the knockdown (log fold change of the knockdown versus control) with the direction of the correlation coefficient (positive or negative) between TF and target

gene (Fig. 4), with the understanding that knockdown of TFs which were positive regulators in the model should result in decreased expression (negative log fold change), i.e., true positives, and knockdowns of TFs that were negative regulators in the model should result in increased expression (positive log fold change), i.e., true negatives. We achieved an accuracy of 0.96 for *GRHL2*, 0.80 for *GCM1*, 0.71 for *ESRRG*, and 0.32 for *CREB3L2*. Statistical significance was determined using a one-sided Fisher's exact test (*GRHL2*, $P = 5.65 \times 10^{-4}$; *CREB3L3*, $P = 1$; *ESRRG*, $P = 7.09 \times 10^{-4}$; *GCM1*, $P = 0.08$). We also
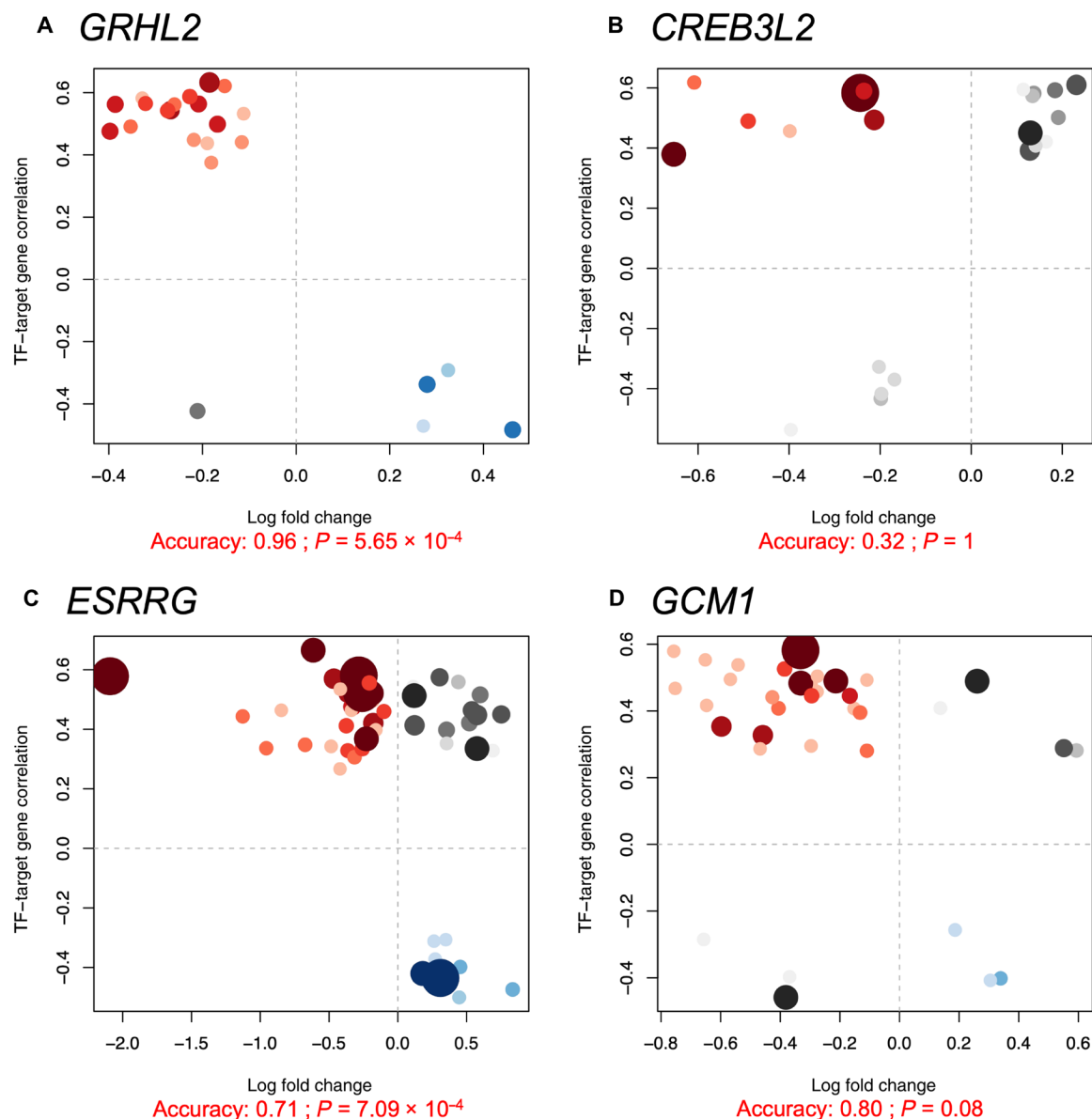
Fig. 4. Experimental validation for assessment of final model accuracy. Results of Experimental validation for (**A**) GRHL2, (**B**) CREB3L2, (**C**) ESRRG, and (**D**) GCM1. X axis represents the log fold change of knockout versus control, and the Y axis represents correlation between TF and target gene. Model accuracy was calculated as the number of positively regulated genes with a negative log fold change (i.e., true positives, shown in red) added to the number of negatively regulated target genes with a positive log fold change (i.e., true negatives, shown in blue) divided by the total number of target genes for each sample. Genes that were not concordant are shaded in gray. The size and shading of each target gene represent the relative rank (1 to 15) of the TF as a regulator in our TRN.

analyzed the predictive accuracy and significance of our model within the experimental data in a sex-stratified analysis. In the female siRNA knockdown samples, we achieved a stronger validation of *GRHL2* (accuracy = 1, $P = 0.015$) and *ESRGG* (accuracy = 0.88, $P = 0.018$), and we generally had lower accuracy in male samples (fig. S5). Overall, our experimental knockdowns revealed that the TRN provided accurate insight into TF regulation for three of our four selected TFs.

Identification of sex-specific differences in transcriptional regulation

We examined effect modification by fetal sex in all TF–target gene interactions captured in our model by running separate linear regression

models between each target gene (dependent variable) and TF (independent variable) with fetal sex as a statistical interaction term. Interactions were considered statistically significant based on the FDR of the interaction term <0.05. We also calculated the difference in effect estimates in these models by performing sex-stratified linear regression models in males ($N = 247$) and females ($N = 228$). We observed 5361 TF–target gene relationships between 254 TFs and 1996 target genes that were significantly different between male and females (FDR adjusted, <0.05; table S6). Six of these TFs and 81 of the target genes were specifically located on the X chromosome, and the rest of the interactions were within autosomal regions. The TFs with the biggest sex-specific differences were *ELK3*, *SP11*, *ETV5*,

ETS2, and *TWIST1*. We compared the difference in estimates between males and females and the direction of the estimates between the TF–target gene relationships. We observed four distinct types of sex-specific interactions, as shown in Fig. 5. We found that 291 positive TF–target gene relationships were stronger in males than in females (quadrant 1), 3681 positive TF–target gene relationships were stronger in females than in males (quadrant 2), 1261 negative TF–target gene relationships were stronger in females than in males (quadrant 3), and 120 negative TF–target gene relationships that were stronger in males than in females. (quadrant 4) The top three biggest differences for each quadrant are depicted in fig. S6. Overall, these statistically significant interaction terms indicate global effect modification by fetal sex and suggest differences in regulation of gene expression between female and male placentas that can be identified in our robust model.

Examination of TFs enriched as regulators of specific placental cell types

Next, we integrated our TRN with recently published placental cell type signatures derived from single-cell RNA-seq data (*30*) to provide additional insight into how individual TFs may regulate specific placental cell types with different roles in placental function. Campbell *et al.* (*30*) identified 19 fetal placental cell types from single-cell RNA-seq data, and we used the cell type signatures derived from Seurat
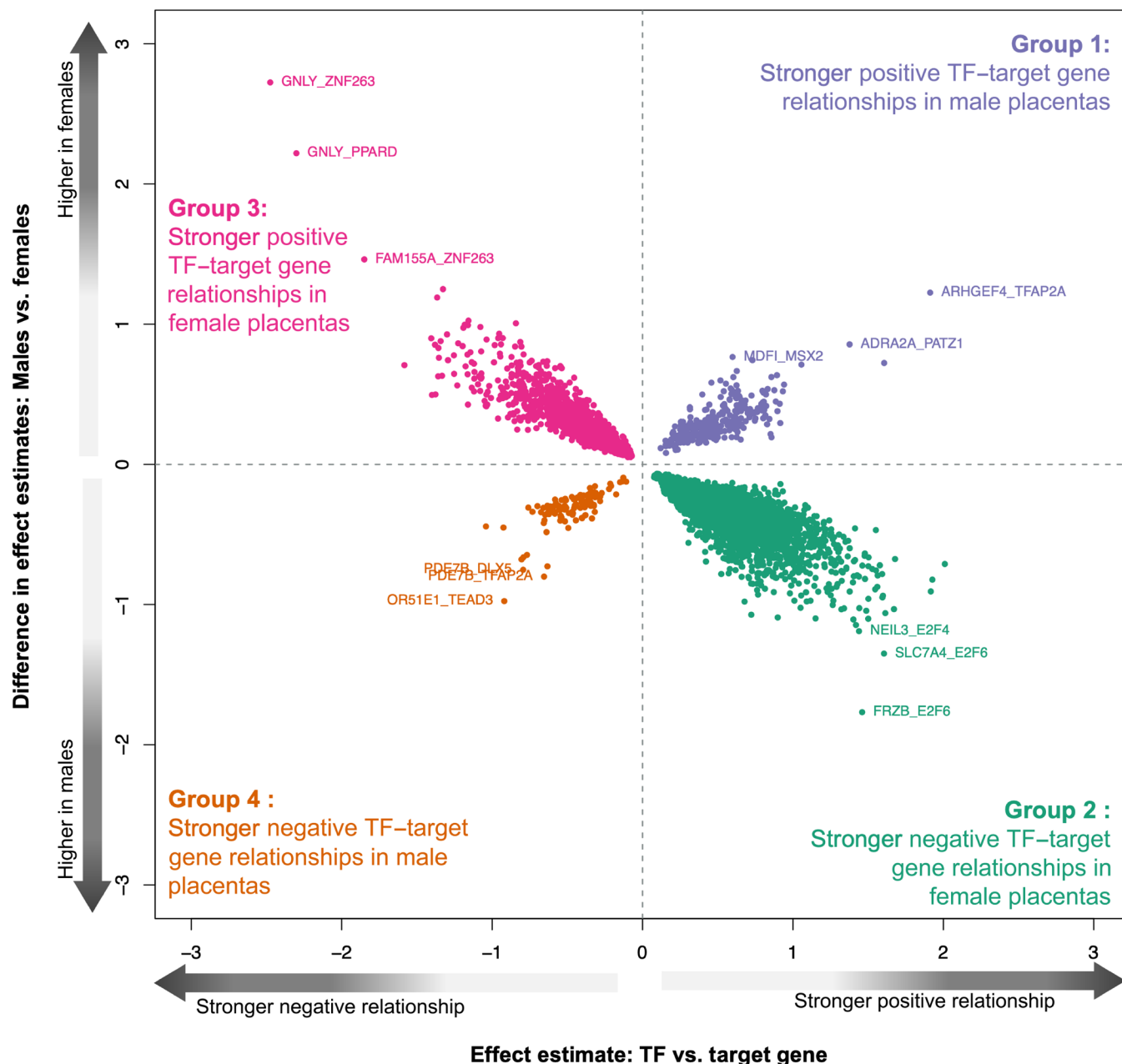

Fig. 5. Plot depicting the 5361 sex-specific interactions in the placental TRN, divided into four distinct interaction categories. The *x* axis represents TF–target gene estimate values from linear regression analysis in the whole dataset. The *y* axis represents the difference in estimates between male and female samples.

(presented in Campbell *et al.*'s supplementary table 2) for TF enrichment tests. We identified enriched TFs using overrepresentation analysis via hypergeometric tests, with the regulons used in lieu of a gene ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) gene set. Overall, there were 159 TFs that were enriched for these cell type markers in one or more placental cell line (FDR < 0.05,

Fisher's exact test), which are presented visually in Fig. 6. We observed a pattern of distinct TFs regulating different types of immune subpopulations (CD14[+] monocytes, Hoffbauer cells, and granzyme B (GZMB) natural killer cells) and a different subset of TFs regulating trophoblast-derived cell types (cytotrophoblasts and proliferative cytotrophoblasts). Fibroblasts and proliferative cytotrophoblast cell
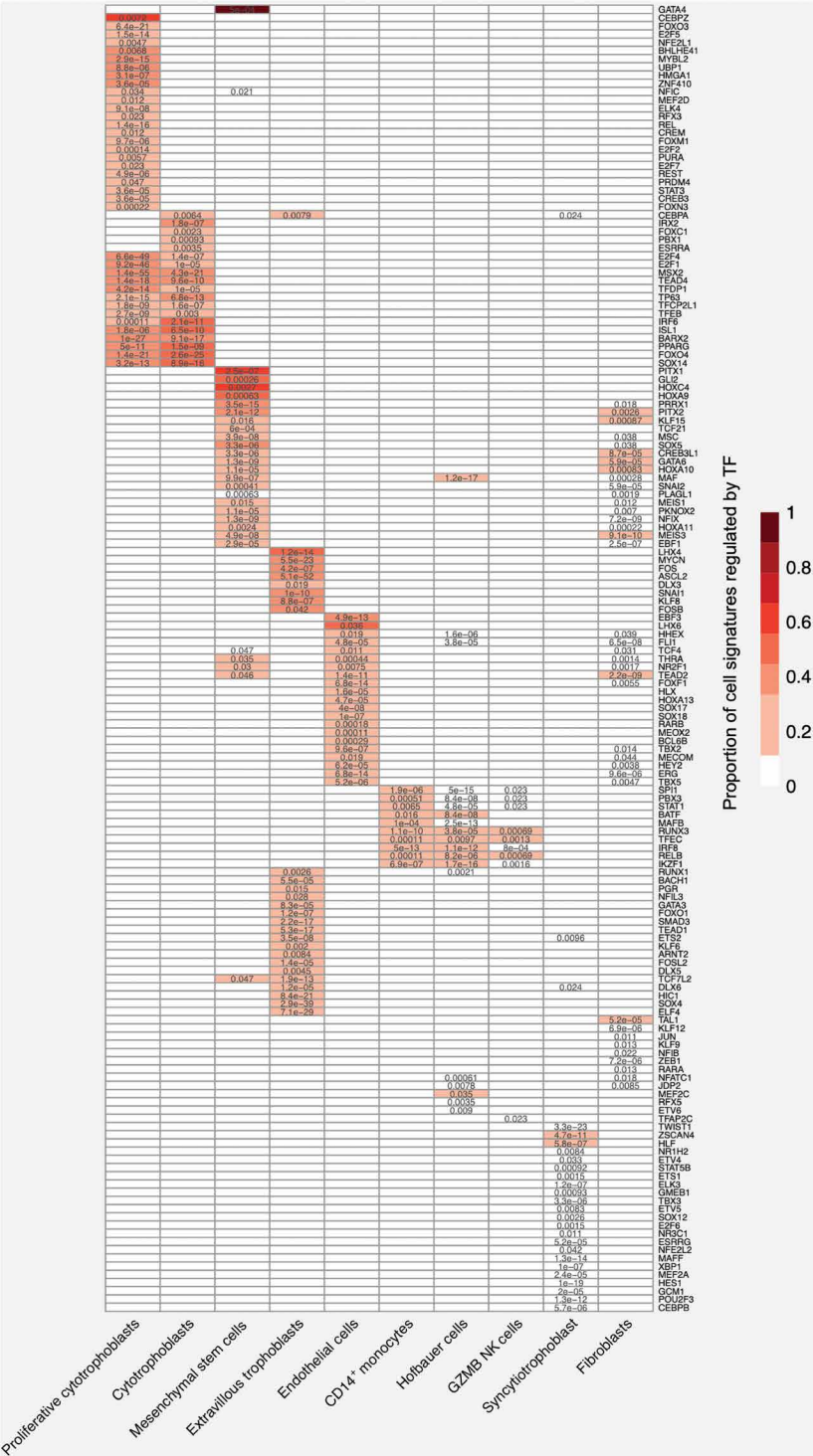
Fig. 6. Heatmap of all TFs enriched for one or more placental cell type signature as defined by Campbell *et al.* (*30*), based on overrepresentation-based analysis (FDR <0.05; Fisher's exact test). The color represents the proportion of total TF target genes that were cell type signatures, and we report *P* values adjusted for FDR. NK cells, natural killer cells.

signatures had the highest number of enriched TFs (38 TFs). Fourteen TFs were enriched for both cytotrophoblasts and proliferative cytotrophoblasts. Extravillous trophoblasts (EVTs) and syncytiotrophoblasts were generally enriched for different TFs, with the exception of the TF *CEBPA*. Syncytiotrophoblasts had the highest number of unique TFs (*22*), and the placental-specific TF *TWIST1* was the most significantly enriched TF in this cell type. EVTs had the second highest number of uniquely enriched TFs (*21*), and the most enriched TFs for EVTs was the placenta-specific TF *ASCL2*. Fibroblasts and proliferative cytotrophoblast cell signatures had the highest number of enriched TFs (38 TFs). These results may be related to the underlying proportions of cell types present in our RNA-seq data.

Applying the placental TRN to discover transcriptional regulators of genes associated with spontaneous preterm birth

Ultimately, the goal of the placental TRN is to gain insight into transcriptional regulation in the context of prenatal exposures or placental pathologies. As a proof of concept, we used the placental TRN to identify key TFs that regulated the group of 961 genes whose placental gene expression was associated with preterm birth, which we established in an independent study of spontaneous preterm birth (sPTB) using clinical definitions and adjustment for confounding variables (*29*). These 961 genes were regulated by a total of 301 TFs. Twenty-nine TFs were significantly enriched for these preterm birth associated genes (FDR < 0.05, Fisher's exact test; see table S7). Together, these 29 TFs were responsible for regulation of 47.55% of the differences in expression observed. In addition, 31 of the 961 transcriptomic signature were TFs that regulated downstream genes that were also significantly associated with preterm birth. Table 2 lists the eight TFs that were both decreased in relation to preterm birth [FDR < 0.05 (*29*)] and were enriched for PTB genes (FDR < 0.05, Fisher's exact test). A network of these TF–target gene relationships revealed that these eight TFs together positively regulated 304 genes that were also deceased in preterm placentas and negatively regulated 30 genes that were increased in preterm placentas (Fig. 7). Of note, the TFs *SPI1* and *RUNX1* emerged as upstream regulators of these TFs based on hierarchical network mapping. These data demonstrate that our placental TRN can be used to identify regulatory pathways driving differential gene expression in placentas from preterm births.

Applications of the placental TRN

Placental omics data can be used to quantify perturbations that influence gene regulation and physiological activity, providing insight into the underlying etiology related to outcomes or exposures (*32*). Many researchers studying the placenta have used omics data to generate lists of genes which are associated with prenatal exposures as well as infant or child health outcomes, as summarized in our recent review (*33*). These genes are often contextualized using pathway analysis approaches within KEGG or GO gene sets. The TF enrichment analysis we have performed using regulons is complementary to using other defined gene sets such as KEGG or GO. TRNs are focused on the question: "What TFs regulate these genes within the placenta?" There are other applications that provide an opportunity to query TF binding, such as Chip-X enrichment analysis (*34*), which are based on aggregation of ChIP-seq data that are not tissue or cell type specific. Our TRN provides a placental tissue–specific version of this information that can be specifically queried to investigate individual TFs or target genes of interest or to perform an overrepresentation analysis, akin to the analysis described above using genes associated with premature birth. We have presented the full results of the genome-scale model in the format of a searchable web tool, PlacentalTRN, which is available at https://paquettelab.org/placentaltrn. This model has 3 features: (i) a target gene search feature, (ii) a TF search feature, and (iii) an enrichment test. The search tools allow users to search the model for specific genes or specific TFs and evaluate model accuracy and predicted targets on a case-by-case basis. The enrichment test tool allows users to import a list of genes (such as the list of preterm birth genes presented here) and identify TFs which regulate these genes. For those interested in accessing the TRN independently of the web interface, we provide the entire TRN in table S8. Our intention is that other placental biology researchers will be able to easily access the TRN and identify transcriptional regulators of their own gene lists of interest by using this database.

DISCUSSION

In this study, we generated a placenta-specific TRN using established genomic footprinting techniques (*35*) and high dimensional RNA-seq data from 475 placental transcriptomes. Our network captures 113,158 interactions between 391 TFs and 7712 target genes.

Table 2. TFs significantly associated with sPTB and enriched for genes associated with sPTB. FC, fold change.

| TF | Association with sPTB* | | | Enrichment test results | | |
	Log FC	FDR adjusted *P*	*N* target Genes	*N* target genes Associated with sPTB	Proportion of PTB target genes	FDR†
SPI1	−0.61	8.92×10^{-04}	996	208	0.21	2.23×10^{-14}
KLF9	−0.53	3.60×10^{-03}	888	172	0.19	7.61×10^{-09}
ETS2	−0.29	6.78×10^{-03}	904	160	0.18	1.25×10^{-05}
RUNX1	−0.32	3.84×10^{-02}	650	133	0.20	2.70×10^{-08}
RARA	−0.43	4.72×10^{-04}	479	127	0.26	3.48×10^{-16}
JDP2	−0.3	2.59×10^{-02}	434	98	0.23	2.70×10^{-08}
FOXO1	−0.28	4.90×10^{-02}	402	89	0.22	3.96×10^{-07}
IRF9	−0.26	1.26×10^{-02}	170	46	0.27	2.87×10^{-06}

*Results described in (*29*).   †*P* values were adjusted using the Benjamini Hochberg method to calculate the FDR.
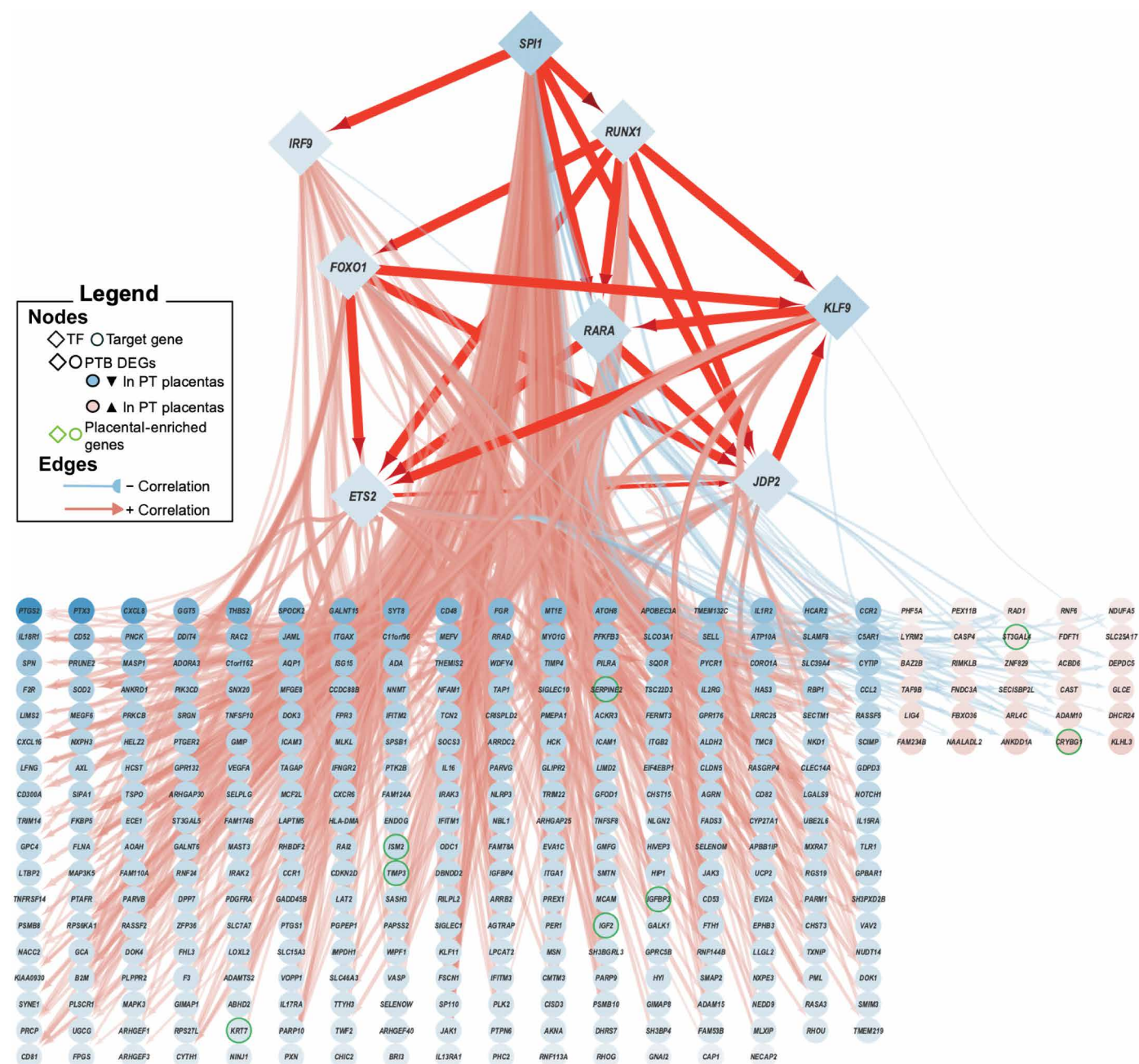
Fig. 7. TRN subnetwork depicting enriched TFs (*P* < 0.1; Fisher's exact test) and their corresponding genes that were previously associated with preterm birth (PT). Diamonds are TFs. Color of gene indicates directionality of log fold change in association with preterm birth. Placental-specific TFs, defined by the Human Protein Atlas (*31*), are outlined in green. Color of edge indicates correlation between TF and target gene, and the number represents correlation coefficient. Transcriptional regulation between enriched TFs is highlighted by red dotted edges.

Our model was validated against a separate transcriptomic dataset and follow-up experiments for several key regulators. Our study revealed that the TFs *SP1*, *ZNF148*, *SP3*, *SMAD4*, and *CLOCK* were master transcriptional regulators of placental gene expression within the TRN. Our study also identified 25 TFs unique to the placenta, with *TFAP2A* emerging as the largest regulon, and highlighted a subset of 5361 TF–target gene interactions that displayed sexual dimorphisms between male and female samples, revealing more transcriptional regulation occurring within female placentas. We used

our TRN to identify specific TFs that were enriched as regulators of signatures of specific placental cell types. Last, we applied our model to identify 29 TFs that were enriched as regulators of genes whose placental expression was previously associated with preterm birth, including eight TFs that were also decreased in preterm birth, which may indicate clinically modifiable targets. This model captures genomic regulation from placentas collected at the end of pregnancy, which is generally when placental samples are readily available and typically collected for large cohort studies. To facilitate additional

studies, we have made our model publicly available as a web accessible resource (https://paquettelab.org/placentaltrn/).

The major advancement of this study is the generation of a large and sophisticated TRN model using DNase-seq data within key placental tissues at term and the validation of this model using gene expression data captured within a large and diverse cohort. Our model also complements other models of placental transcription constructed using gene coexpression network analysis approaches but provides individual gene level resolution, which enhances interpretability. An independent gene coexpression network of the human placenta constructed using WGCNA on transcriptomic data revealed five hub modules associated with fetal growth restriction (FGR) (*23*). One of the core genes of the module associated with both FGR and fetal overgrowth was *CREB3*, which was a hub TF that regulated 738 target genes within our model. Our model builds upon this by inclusion of information about chromatin accessibility using DNase-seq. DNase-seq data (without coexpression analysis) have been used to construct a gene regulatory model of early placental transcriptional regulation (*26*). This study mapped genes involved in differentiation of human embryonic stem cells to trophoblast cells in mice and identified 20 key TFs involved in placental differentiation through centrality analysis (*26*). Many of the reported genes in this study that were linked to placental development are hub TFs in our model, including *SMAD4*, *ETS1*, *RUNX1*, *CREB1*, *SMAD3*, *PPARA*, and *JUND*. We did not include the subset of established TFs without motifs as defined by Lambert *et al.* (*36*). We surmised that there were complex mechanisms by which these TFs act to regulate gene expression (including acting as cofactors). Comparison of the accuracy of null models to our primary model highlights that constraining models based on TF motifs within regions of open chromatin and strong TF–target gene correlations results in higher accuracy of TFs at predicting target gene expression compared to models with weak correlations and/or no binding sites.

Our model also captures regulation of key TFs identified in other studies from human placental data collected at term. Another WGCNA-based analysis revealed that a large proportion of the placental transcriptome is organized into modules of coexpressed genes that were involved in core placental functions and were reproducible across trimesters (*22*). Our model (generated at the end of gestation) is concordant with previously generated network-based analyses earlier in gestation, suggesting that these TFs are essential not only to placental development but also to placental function across gestation. However, there are instances where established TF–target gene interactions were not included in our network. For example, several studies in placental cell lines have demonstrated that TF GCM1 is a positive regulator of the *PGF* gene (encoding placental growth factor) (*37*, *38*), and while we identified binding sites and a positive correlation in our initial model, this relationship was not strong enough to be included in the final model based on our criteria. Thus, it is important to recognize that not all TF–target gene interactions are captured in our network, due to type II error. We note that *PGF* and *GCM1* relationships are also not captured in other network tools such as HumanBase or the STRING database, indicating that this is a common issue with network approaches. We elected to use Lasso regression–based approaches for each target gene individually, as opposed to clustering-based approaches, for enhanced interpretability and decreased chances of capturing indirect and false-positive interactions (*39*). WGCNA and clustering-based approaches may be more appropriate in other contexts, and researchers can apply our TRN to

their own coexpression clusters by using web application to identify regulons that are enriched for their lists of coexpressed genes.

Many of the highest-ranking core TFs in our model, including *SP1*, *SP3*, *SMAD4*, and *TFAP2A*, are essential regulators of placental function. *SP1* and *SP3* had the highest eigenvector and degree centrality in our model. SP TFs are involved in placental glucocorticoid signaling (*40*), and placental expression of these TFs has also been linked to preeclampsia (*41*). *SMAD4* had the second highest eigenvector centrality value and is a signaling protein in the transforming growth factor (TGF) family which is activated in response to TGF-β. This signaling pathway plays a crucial role in regulation of trophoblast proliferation, differentiation, and invasion, as was reviewed by Adu-Gyamfi *et al.* (*42*). *TFAP2A* regulated 1255 target genes and was the largest placental-specific TF based on both eigenvector centrality and degree centrality. *TFAP2A* encodes the activating enhancer binding protein 2-alpha (AP-2α), a TF that is involved in several key processes unique to the placenta (*43*). *TFAP2A* expression is detectable throughout the placenta, including extravillous cytotrophoblasts, cytotrophoblasts, and syncytiotrophoblasts. AP-2α promotes the epidermal growth factor–dependent invasion of the human trophoblast (*44*) and induces human CRH gene expression (*45*). We were able to recapitulate some of these published findings of *TFAP2A* regulation. In another study, the TFs specificity protein 1 (SP1) and specificity protein 3 (SP3) regulate expression of *TFAP2A* in trophoblast cells (*46*). In our model, *SP1* and SP3 were included in the top 15 TFs regulating the expression of *TFAP2A*. *TFAP2A* is a positive regulator of *CRH* in our model, in alignment with work showing that AP-2 induced *CRH* gene expression (*45*). Thus, our TRN reveals the significance of this TF for placental function as well as some core TFs that regulate many genes within the placenta.

Sex differences between male and female placentas may contribute in part to differences in growth rates of male and female fetuses and differential responses to perturbations in the in utero environment and manifestation of disease later in life (*47*). In our analysis, we identified 5361 TF–target gene relationships with effect modification based on fetal sex. Many of these TF–target gene differences may be related to hormonal changes, as we observed several differences related to nuclear hormone receptors that are responsive to sex hormones, including *ESR1*, *ESSRA*, *ESRRG*, and *PGR*. We observed a general trend in our interaction terms where both negative and positive TF–target gene relationships were stronger in female placentas compared to male placentas. An aggregated meta-analysis of microarray data revealed 140 differentially expressed genes in male and females samples, with the majority residing in autosomal regions (*27*). This meta-analysis also identified several TFs that were enriched for these genes, including *NFATC2*, *AHR*, *NKX3*, *HIF1A*, and *REL* (*27*). A number of these TFs were associated in our analysis with sex differences in expression of target genes, such as the *AHR*, which was associated with 111 TF–target gene relationships that were significantly different between males and females. On the basis of these findings, we hypothesize that there may be stronger transcriptional control of gene expression in female compared to male placentas. These differences may be driven by sex-specific differential response to stress hormones, environmental chemicals, and nutrition, as reviewed by Rosenfeld (*48*). This is important because there is variation in pregnancy outcomes related to fetal sex (*49*), such as an increased rate of preterm birth in male infants (*50*). Further analysis needs to be performed on these relationships, including factors that modify these relationships and the role of genomic imprinting.

We used our TRN to perform an enrichment analysis and identify specific TFs that regulated transcriptomic signatures attributed to 19 specific fetal placental cell types identified by Campbell *et al.* (*30*). Our analysis revealed unique TFs that were enriched as regulators of different types of trophoblast cells. This is likely related to the specific regulatory roles of trophoblast TFs that are unique to specific lineages of placental cells as summarized by Knofler *et al.* (*51*). For example, *ASCL2* was the most significantly enriched TF for extravillous trophoblasts in our model, and this TF is known to be a key regulatory TF unique to extravillous trophoblasts and their progenitors (*51*). The TF *TWIST1* is involved in the process of trophoblast syncytialization (*52*), and we found this TF to be uniquely enriched in synctiotrophoblasts in our samples. Overall, this enrichment analysis using the regulon models provided additional insight into the regulatory roles of specific TFs within individual placental cell types.

To this end, we used the TRN to identify 29 different TFs enriched for genes associated with spontaneous preterm birth identified in an independent study (*29*), including eight TFs (*SPI1*, *KLF9*, *ETS2*, *RUNX1*, *RARA*, *JDP2*, *FOXO1*, and *IRF9*) whose placental gene expression decreased in preterm placentas compared to term placentas. Developmental biology studies have revealed that these TFs play crucial roles in placental development. Both *FOXO1* and *KLF9* have been linked to the process of decidualization (*53*, *54*), which forms the basis of the maternal-fetal interface. *RUNX1* and *RARA* are required for trophoblast differentiation (*55*, *56*), and *ETS2* is essential for trophoblast cell renewal (*57*). This body of work suggests that the network level changes we observed, including decreased expression of these TFs and their downstream genes, may reflect foundational differences in the development of the placenta in preterm infants. TFs that were decreased in preterm placentas, including *FOXO1*, *JDP2*, and *KLF9*, were involved in transcriptional regulation of the progesterone receptor (encoded by *PGR*). Progesterone is essential for the maintenance of pregnancy and timing of parturition (*58*). Kruppel-like TF 9 (encoded by *KLF9*) influences the length of parturition through regulation of *PGR* in myometrial tissue in mice (*59*), and this relationship was confirmed in our placental TRN as well. Forkhead Box O1 (encoded by *FOXO1*) is regulated by progesterone in myometrial cells (*60*), and *PGR* also regulated the *FOXO1* TF in our TRN. This body of work confirms the important regulatory role of progesterone in preterm birth and indicates specific mechanisms of transcriptional regulation. Several of these TFs have also previously been linked to preterm birth in other studies. Both *FOXO1* and *ETS2* were differentially expressed in preterm placentas compared to term placental control samples in a case-control transcriptomic study (*61*). Another network biology–based study that used a protein-protein interaction network identified *RUNX1* as regulator of genes containing SNPS associated with sPTB (*62*). The retinoic acid receptor *RARA* was decreased in sPTB placentas and regulated 127 downstream genes that were also decreased in preterm placentas in our TRN. It is well established that serum levels of retinoic acid are decreased in preterm placentas (*63*), which has been confirmed in a recent metabolomics study (*64*). Ultimately, the 29 TFs that regulated sPTB genes could also represent potential clinical targets for preterm birth, as we reveal that downstream gene expression differences related to preterm birth can be regulated by these TFs. For example, our analysis revealed that several TFs involved in progesterone signaling decreased in sPTB, as well as the retinoic acid receptor. There are ongoing clinical studies investigating the use of progesterone (*65*–*67*) and retinoic acid (*68*,

69) in the treatment and prevention of preterm birth. TFs are attractive druggable targets in the context of disease and have broadly been used in other diseases including cancer and autoimmune disease (*70*).

Our model of placental transcriptional regulation should be interpreted with limitations in mind. We use TF expression as a proxy for TF activity, which does not account for posttranscriptional or posttranslational modifications. Common to many similar studies, our results may be confounded due temporal and spatial disconnects, cellular heterogeneity, and differences in proliferation (*71*). Our results may be confounded due to cellular heterogeneity, as TRNs reflect epigenetic regulation dependent on regions of open chromatin and DNA methylation which is intrinsically cell but not tissue type specific. Our results are constructed from bulk RNA-seq of placental tissue which is a composite of different cell types. This can create false-positive relationships and dilute signal strengths between known TFs and target genes. For each target gene, we present only the top 15 TFs that are best predicted to regulate expression. Thus, we do not represent all the TFs with binding sites within that region, contributing to type II error in our models. Our model was constrained to regions of open chromatin identified through DNase-seq. One limitation of this assay is the potential for DNase cleavage bias, which can obscure specific genomic locations and would also lead to type II error in our analysis (*72*). Our analyses are composed of data solely from placentas at the end of gestation from relatively nonpathological pregnancies without preeclampsia and so may be inappropriate for researchers studying earlier developmental time points and also may not represent perturbed TF–target gene interactions related to those pathologies. Our transcriptome data were generated from placentas of infants delivered either vaginally or through cesarean section, and our DNase-seq data were collected from placentas after cesarean section. The influence of labor and delivery method on DNase-seq data are not well studied, but there are known influences on the placental transcriptome (*73*). It is essential that studies investigating placental transcriptional differences consider the role of labor type and delivery method as a potential confounding variable before downstream network analysis. We applied the widely used ComBat algorithm (*74*) to adjust for batch effects within our two datasets, which may remove biological variation in the data, so it is possible that the TF–target gene relationships we present in our models may have type II error. As always, it will be important to further test the model in an entirely separate validation cohort collected from different participants as they become available. In the future, we hope to build upon this model by including temporal resolution of time series changes that occur in the placental transcriptome and to apply to this model to other clinically relevant endpoints such as preeclampsia or gestational diabetes with specially designed cohorts designed to capture transcriptomic signatures in these populations.

Our model makes an important contribution to the field of placental biology as the first publicly available placental TRN, as well as one of the first models of transcriptional regulation of any organ with an accessible and searchable database that allows researchers to perform their own analysis. We applied rigorous methods to test the accuracy of our model, including an independent validation dataset which is the gold standard in machine learning (*75*, *76*). To our knowledge, our model is composed of the largest placental transcriptomic dataset constructed to date. We also provide an innovative assessment of sex differences in TF–target gene relationships in our model, which further captures the importance of fetal sex as a biological variable in placental omics research. Last, we have tested our

TRN using isolated villous cytotrophoblasts collected from placentas of nonpathological women delivering at term, which are more reflective of the term placental transcriptome than commercial placental cell lines which are commonly used for transcriptomic analyses. Our data could potentially be used in the future to provide insight into transcriptional mechanisms related to pregnancy complications and identify pharmacological interventions to target specific TFs that alter placental function.

MATERIALS AND METHODS

Sample information
To construct the TRN, we integrated transcriptomic and chromatin accessibility data from placental tissue generated from three different sources as described below.

MWRI sample
Tissue was collected from nonpathological placentas from pregnant individuals who delivered at the Magee Womens Hospital in Pittsburgh, PA. Inclusion criteria included singleton pregnancies with signed informed consent and gestational age ≥ 37 weeks. Exclusion criteria included pregnancy complications including FGR, chromosomal abnormalities, preeclampsia, gestational diabetes, chorioamnionitis, antepartum hemorrhage, and a previous history of preterm birth. Placental biopsies (5 mm^3) were sharply dissected within 30 to 40 min after delivery from a region of the placenta midway between the cord insertion and the placental margin between the chorionic and basal plates and free of any macroscopic lesions, as we previously detailed (77). Tissue biopsies were placed in RNAlater for 48 hours in 4°C and then stored in −80°C until processed for RNA-seq. Directional RNA-seq was performed by the Genomics, Epigenomics, and Sequencing Core at the University of Cincinnati. Approximately 50 mg tissue was homogenized in 0.8 ml lysis/binding buffer from the mirVana miRNA Isolation Kit (Thermo Fisher Scientific, Grand Island, NY) by using FastPrep-24 5G homogenizer (MP Biomedicals, Solon, OH). Total RNA extraction was conducted according to the manufacturer's protocol. RNA quality was examined on an Agilent 2100 Bioanalyzer (Santa Clara, CA). Ribosomal RNA was depleted using the Ribo-Zero Gold kit (Illumina, San Diego, CA) with 1 μg of total RNA as input for the SMARTer Apollo NGS library prep system (Takara, Mountain View, CA). The library for RNA-seq was prepared by using the NEBNext Ultra II Directional RNA Library Prep kit (New England BioLabs, Ipswich, MA), and the sequencing was run on an Illumina High-Seq2000 with paired end reads, 101-bp read length, and at 50 million read depth.

CANDLE samples
Placental tissue was collected from nonpathological placental tissue from women from Shelby County Tennessee as part of the CAN-DLE study. RNA-seq data were generated as part of the ECHO-PATHWAYS consortium (78). Inclusion criteria included singleton pregnancies enrolled in CANDLE with signed informed consent. Exclusion criteria included infants categorized as having severe pregnancy complications recorded in the CANDLE medical records, including confirmed clinical chorioamnionitis, preeclampsia, oligohydramnios, placental abruption, infraction or previa, and fetal chromosomal abnormalities. Methods of placental sample collection and RNA-seq processing for this cohort have been previously described (79). Briefly, within 15 min of delivery, a piece of placental villous tissue approximately 2 cm by 0.5 cm by 0.5 cm was dissected from the placental parenchyma. The tissue cubes were stored in 20 ml of RNAlater

and refrigerated at 4°C overnight (at least 8 hours but no more than 24 hours). Following this, each cube was transferred to an individual cryovial containing fresh RNAlater and stored at −80°C. The fetal villous tissue was manually dissected and cleared of maternal decidua. RNA was isolated using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Germantown, MD) according to the manufacturer's recommended protocol. RNA purity was assessed by measuring optical density at 260/230 nm (OD$_{260/230}$) and OD$_{260/260}$ ratios with a NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). RNA integrity was determined with a Bioanalyzer 2100 using RNA 6000 Nanochips (Agilent, Santa Clara, CA). cDNA libraries were prepared from 1 μg of total RNA using the TruSeq Stranded mRNA kit (Illumina, San Diego, CA) and the Sciclone NGSx Workstation (PerkinElmer, Waltham, MA). Before cDNA library construction, ribosomal RNA was removed by means of polyadenylation (Poly-A) enrichment. Each library was uniquely barcoded and subsequently amplified using a total of 13 cycles of polymerase chain reaction. Library concentrations were quantified using Qubit fluorometric quantitation (Life Technologies, Carlsbad, CA). Average fragment size and overall quality were evaluated with the DNA1000 assay on an Agilent 2100 Bioanalyzer. Each library was sequenced to approximately a depth of 30 million reads on an Illumina HiSeq sequencer.

Oregon Health Sciences Center Placental Repository
Twelve placental villi samples were derived from nonpathological pregnancies from individuals recruited into the Oregon Health Sciences Center Placental Repository (OHSU). Inclusion criteria included singleton pregnancies delivered via c-section at OHSU Hospital Labor and Delivery Unit with signed informed consent and gestational age ≥ 37 weeks. Exclusion criteria included maternal body mass index > 25, maternal age < 18 or > 40, any current pregnancy complications, and current smokers. Samples were specifically selected to be equally matched across fetal sexes. Placentas were collected and weighed immediately following cesarean section. Five random samples of tissue (~80 grams) were collected from each placenta and placed in phosphate-buffered saline (PBS) to be transported back to the lab. The chorionic plate and decidua were removed from each randomly isolated placental sample, leaving only villous tissue, which was thoroughly rinsed in PBS to remove excess blood. Primary cytotrophoblasts were isolated from villous tissue using a protocol adapted from Eis et al. (80), using trypsin/DNase digestion followed by density gradient purification. Isolated cytotrophoblast cells (5 to 10 × 10^6 cells/ml) were then frozen in freezing media [10% dimethyl sulfoxide in fetal bovine serum (FBS)] and stored in liquid nitrogen until usage. Placentas were collected into a tissue repository under a protocol approved by the Oregon Health and Sciences University Institutional Review Board with informed consent from the patients. All tissues and clinical data were deidentified before being made available to the investigative team.

Quantifying placental gene expression in the CANDLE AND MWRI placentas
A flowchart describing sample preprocessing is described in fig. S2B. At the time of TRN construction, 475 CANDLE placental samples had been sequenced. For the RNA-seq samples from MWRI and the CANDLE cohort, transcript abundances were estimated using the quantification program Kallisto (80, 81). Initial quality control was performed on each RNA-seq sample using FastQC (82). Length scaled counts were condensed to Ensembl Gene IDs using the Bioconductor TXImport package (83). Genes were then filtered to remove

low-expressing genes using the "FilterByExprs" function of EdgeR, which, by default, removes genes with <10 counts [0.5 counts per million (CPM)] in 70% of samples (*84*). For the dataset from MWRI, this included 17,574 genes, and from the CANDLE samples, this included 16,440 genes. Datasets were then normalized using the trimmed mean of *M*, and then precision weights were generated for log CPM values using the voom transformation. The datasets were then merged by all genes common across both datasets, generating a final transcriptomic dataset of 14,837 genes in 475 samples. Confounding related to both RNA-seq batch and intrinsic differences between the cohorts was eliminated using the ComBat algorithm from the Bioconductor SVA package (*74*), which applies an empirical Bayes method to estimate location and scale parameters for each batch, which are then removed to provide a consistent mean and variance for each gene, across all batches. This approach has successfully been used to integrate microarray datasets (*85*), as well as RNA-seq datasets (*86*). RNA-seq batch was set as the "batch" variable to be adjusted for (the CANDLE cohort was sequenced in four batches, and the MWRI dataset was sequenced as one batch), and maternal race and gestational age were included as covariate variables in the model matrix, to ensure that differences related to these variables would be retained in the final dataset. Batch effects were evaluated using principal components analysis (see fig. S7). For TRN construction, the data were split into a holdout dataset (20% of samples, $N = 93$), and 20% of the remaining samples was used for the testing dataset ($N = 75$), with 80% used for the construction of the initial TRN ($N = 307$). After model splitting but before model construction, we tested to ensure that there were no statistical differences between the training and holdout datasets in relation to the distribution of covariates including batch/cohort, labor status maternal race, and fetal sex (categorical, tested using chi-square tests), as well as gestational age and birth weight (continuous, *t* tests). In the final model, all samples were used ($N = 475$ samples). An additional 366 CANDLE placental transcriptomes were sequenced after model construction and were processed using the same pipeline for use as an independent validation dataset.

DNase hypersensitivity data generation

DNase-seq data were generated from the samples collected at OHSU. The Altius Institute used an established protocol for tissue processing developed as part of the Encyclopedia of DNA Elements (ENCODE) project (*87*). Samples processed using the (DNAse hypersensitivity (DHS) -pseudonano assay & nuclei were sterile-filtered (0.25 to 6.4 million nuclei per sample), with minimal cellular debris, and treated with three doses of DNase (40, 60, and 80 U) followed by a stop buffer including ribonuclease A, followed by treatment with proteinase K. The samples had a signal portion of tags (spot) score > 0.5 except for one sample (which had a score of 0.34), which is a measure of enrichment, reflecting the fraction of reads that fall within peak calling regions. A score > 0.4 is indicative of high-quality DNase sequencing data meeting ENCODE quality guidelines. Peak regions of DNase-seq data were identified using F-seq (*88*), a density estimator for high-throughput sequence tags, using a threshold parameter of 2.5, which was selected based on previously reported performance metrics (*72*). The peak regions and DNase-seq data were both used to detect genomic "footprints" (i.e., regions of accessible DNA) using Hmm-based IdeNtification of Transcription factor footprints (*89*), a method based on hidden Markov models.

TF motifs and single gene model construction

For each gene in our network, we identified putative TFs through integration of the DNase-seq data and TF binding motif information. TFs were defined by the Human Transcription Factor Database which had motifs from these TF databases (667 TFs) (*36*). TF motifs were identified using three established databases: JASPAR (*90*), HOmo sapiens COmprehensive MOdel COllection (*91*), and Swissregulon (*92*). We first matched these binding sites with the footprints in the placenta (described above) using Find Individual Motif Occurrences, applying the default 10-4 match threshold (version 5.0.5) (*93*). The motifs were aggregated into position weight matrices within the R package "MotifDB" (*94*). We assigned motifs to their TFs as well as those in the same DNA binding domain family using the TFClass database, as established in our previous work (*17*). In our analysis, we did not consider the subset of TFs classified by Lambert *et al.* (*36*) with no known motifs. We included both promoter and enhancer regions in our model. The promoter region was defined as ±5 kb from the transcriptional start site, which has been shown to maximize target gene prediction in TRNs constructed with footprinting data (*95*). The enhancer regions were identified using genehancer (version 4.11). Genehancer is a database that integrates reported enhancers from multiple sources, including ENCODE and Ensembl, and uses a combinatorial likelihood method to define genes based on tissue coexpression correlations, enhancer-targeted TFs, expression quantitative trait loci, and chromatin conformation capture (*96*). We included all enhancer regions defined by Genehancer as "elite," indicating that they had more than two evidence sources, and those defined as "placental-specific" based on Genehancer analysis of tissue-specific genes.

TRN construction and parameterization
Constructing and evaluating model in training data

We integrated the data described above to generate a model consisting of nodes and edges. A "putative edge" was identified between a TF and a target gene if a motif for that TF was present in a region of open chromatin (i.e., footprint) in the gene's promoter or enhancer (see Fig. 1). We generated a coexpression matrix for each putative edge. We ranked each target gene based on the absolute value of the Pearson correlation coefficient. We pruned our model through an initial "baseline model" (model 1), including only TFs with > median absolute value of the overall Pearson correlation coefficient for each target gene and then including only the top 15 TFs based on absolute Pearson correlation coefficient. We selected this initial module size of 15 as an initial cutoff based on similar analyses performed in the construction of a TRN constructed in the context of T helper cell type 17 differentiation (*97*), and we tested other cutoffs (10 and 20 TFs). All models were constructed using Lasso regression within the "glmnet" R package (version 2.0-16). First, we constructed the model using the "cv.glmnet" function using the default parameters, which performs a 10-fold cross-validation on a grid of 100 values of lambda and identifies a lambda minimum value from the mean squared error of these models. This lambda minimum was used to predict gene expression first in the test dataset (during model construction) and then in the holdout dataset. To evaluate the accuracy of our initial TRN and select model parameters for our full model, we examined how the expression of the TFs in our model predicted gene expression of each target gene in our testing dataset by comparing the overall OOS R^2 values. These values were calculated as the squared estimate of the correlation between the predicted versus

actual expression values generated from Lasso regression. We selected the best model to use in the holdout dataset based on these thresholds.

Null model

We created three null TRNs to evaluate the validity of our model to distinguish between signal and noise, which we constructed in our training and evaluated in testing dataset. The first null model (model 7) was constructed from 15 randomly selected genes with motifs but with correlations that were less than the median correlation coefficient. In parallel to our model constructed above, we generated a network consisting of "null putative edges" defined as interactions between a TF and a target gene if a motif for that TF was not present in a region of open chromatin (i.e., footprint) in the genes promoter or enhancer. We generated a coexpression matrix for each "null putative edge" within our training dataset. Our second null model (model 8) was constructed of 15 randomly selected genes with no motifs but with correlation coefficients > the median value in our true dataset. Our third null model (model 9) was constructed of 15 randomly selected genes with correlation coefficients less than the median value in our true dataset and with no motifs. In some instances, 15 genes were not available that met these criteria, and in this case, we included all the genes available.

Final model construction and analysis

After we established specific model parameters (i.e., enhancers, numbers of TFs, and correlation coefficient cutoff) in our training and testing dataset, we combined these two datasets and then tested the model in our final holdout dataset using the best parameters as identified from our original models. We did not include target genes with an OOS R^2 accuracy < 0.25. Network connectivity scores were calculated using the R package "igraph" (version 1.2.4.1). To examine effect modification in TF–target gene relationships by fetal sex, we ran linear regression models for each target gene–TF interaction in our TRN, with the target gene as the dependent variable and each one of the TFs as a separate independent variable and with fetal sex as an interaction term. Models were considered statistically significantly different by sex if the FDR adjusted P value of the interaction term was <0.05. We adjusted for the total number of linear regressions run to test all interaction terms ($N = 113{,}158$). We ran fetal sex-stratified linear regression models in male samples ($N = 247$) and female samples ($N = 228$) to calculate the coefficient estimate and differences in these estimates, which was used to evaluate the directionality of these relationships.

Experimental validation

We performed knockdowns of selected TFs in primary villous trophoblast cells from placental samples collected through the OHSU placental biorepository. Placenta-specific TFs were identified using the Human Protein atlas (*31*), and then four TFs were selected on the basis of module size (>125 target genes) and median gene expression (>median reads per million). Separate transfections were performed using the ON-Target SMARTpool siRNAs (Dharmacon) to induce silencing of four genes: *CREB2L2* (catalog no. L-019357-02-0005), *ESRRG* (catalog no. L-003405-00-0005), *GCM1* (catalog no. L-011491-00-0005), and *GRHL2* (catalog no. L-014515-02-0005) with a nontargeting negative control (catalog no. D-001810-10-05). siRNAs were diluted in nuclease-free water and stored at a 50 μM stock concentration. These transfections were each performed from villous trophoblasts collected from elective *c*-sections at OHSU as described above.

For transfection, cytotrophoblast cells were rapidly thawed in a 37°C water bath and immediately diluted in Iscove's modified Dulbecco's medium [25 mM glucose, 4 mM glutamine, and 1 mM pyruvate (Gibco #12440-046) supplemented with 10% FBS (Corning #35-010-CV) and 1% penicillin/streptomycin (Gibco #15140-122)] termed as complete media. Cells were centrifuged at 1000*g* for 10 min and suspended in fresh complete media. The cells were plated at 2×10^6 cells per well concentration in 12-well culture plates and cultured at 37°C, 5% CO_2. After 24 hours, cells were transiently transfected using the above siRNAs and Lipofectamine (Invitrogen #11668-019) as per the manufacturer's protocol and incubated for another 24 hours. At the end of 24 hours (total of 48 hours from the start of culture), the transfection media was replaced with complete media, and cells were cultured for another 24 hours. After 72 hours of total culture time (24 hours following previous complete media change), the cells were lysed in 700 μl of Qiazol, and RNA was isolated using the Zymo Direct-zol Mini-prep kit and reagents (Zymo, catalog no. RD2051). RNA integrity was determined with a bioanalyzer, and only RNA samples with an RNA integrity number > 8.6 were used for RNA-seq analysis. cDNA libraries were prepared from 1 μg of total RNA using the Illumina Stranded Total RNA library kit. Each library was sequenced to approximately a depth of 50 million reads on an Illumina NovaSeq 6000 S4 (2 ×100 read length) sequencer by the OHSU sequencing core.

We used a similar pipeline to process the RNA-seq data as was used for the human placental data, as described above. Transcript abundances were estimated using Kallisto, and length scaled counts were condensed to Ensembl Gene IDs using the Bioconductor TXImport package (*83*). Genes were then filtered to remove low-expressing genes using the FilterByExprs function of EdgeR (*84*). Differentially expressed genes between siRNA knockdowns versus negative scrambled controls were identified using generalized linear models implemented within EdgeR (*84*), including only target genes that were predicted to be expressed by each TF (i.e., for *GCM1*, we analyzed 240 genes; for *CREB3L2*, we analyzed 425 genes; for *ESR1*, we analyzed only 676 genes; and for *GRHL2*, we analyzed 283 genes). Dispersion parameters were estimated using the Cox-Reid method, and then differentially expressed genes were identified using the using quasi-likelihood *F* tests, which is more robust and produces more reliable error rates than a likelihood ratio test when the number of replicates is small (*98*). For this screen, we considered genes statistically significant using an unadjusted $P < 0.05$. These significant genes were used to calculate model accuracy and statistical significance. Model accuracy was calculated as the number of positively regulated genes with a negative log fold change (i.e., true positives) added to the number of negatively regulated target genes with a positive log fold change (i.e., true negatives) divided by the total number of target genes for each sample. Model accuracy reflects these ratios. Statistical significance was assessed using a one-sided Fisher's exact test on the number of genes that were true positives and true negatives versus other genes, and we report *P* values from this test.

TF module enrichment analyses

We performed two separate TF enrichment analyses (*99*) in which we considered the TFs and all the genes they regulate (i.e., regulons) as gene groups (analogous to predefined pathways such as KEGG or GO) and compared if the two gene lists of interest were enriched within specific regulons using one sided Fisher's exact test. The first

analysis involved cell signatures of fetal placental-specific cell types derived from single-cell RNA-seq data of the placenta and identified by Seurat by Campbell *et al.* (*30*). Top marker gene expression is reported as a Bonferroni adjusted *P* of <0.05 for this paper (*30*) . The second analysis involved differentially expressed genes that were significantly associated with spontaneous preterm birth identified in a combined cohort analysis of RNA-seq data from the CANDLE cohort as well as a different cohort (Global Alliance to Prevent Prematurity and Stillbirth) (*29*). In the original study, genes were considered statistically significantly associated with preterm birth using an FDR < 0.05. We are using the subset of 961 genes significant from the binary analysis comparing preterm to term. For both analyses, we considered the 7712 target genes captured within our TRN as the background list of genes for overrepresentation analyses, based on the rationale underlying pathway enrichment analysis that nonspecific background lists bias overrepresentation tests (*100*, *101*). We only included TFs in our enrichment tests which regulated three or more differentially expressed genes. For both of these exploratory analyses, we considered a TF to be a key regulator if its target genes were overrepresented based on an FDR < 0.05.

We used Benjamini and Hochberg's method to compute the FDR using *P* values (*102*) for all analyses. All code is publicly available in the following github repository (https://github.com/AlisonPaquette/PlacentalTRN_ManuscriptCode/). All analyses were performed in R (version 3.6.1), and data were visualized using Cytoscape (version 3.9.1). The TRN was developed using the R Shiny Web Application.

Supplementary Materials

This PDF file includes:
Figs. S1 to S7
Tables S1 to S3, S5, and S7
Legends for tables S4, S6, and S8

Other Supplementary Material for this manuscript includes the following:
Tables S4, S6, and S8

REFERENCES AND NOTES

1. G. J. Burton, E. Jauniaux, What is the placenta? *Am. J. Obstet. Gynecol.* **213**, S6–S8 (2015).
2. S. Mesiano, CHAPTER 11 - The endocrinology of human pregnancy and fetoplacental neuroendocrine development A2 - Strauss, Jerome F. in *Yen and Jaffe's Reproductive Endocrinology (Sixth Edition)*, R. L. Barbieri, Ed. (W.B. Saunders, 2009; https://sciencedirect.com/science/article/pii/B9781416049074000115), pp. 249–281.
3. L. Myatt, Placental adaptive responses and fetal programming. *J. Physiol.* **572**, 25–30 (2006).
4. Y. Wang, S. Zhao, *Vascular Biology of the Placenta* (Morgan & Claypool Life Sciences, 2010).
5. J. Saben, Y. Zhong, S. McKelvey, N. K. Dajani, A. Andres, T. M. Badger, H. Gomez-Acevedo, K. Shankar, A comprehensive analysis of the human placenta transcriptome. *Placenta* **35**, 125–131 (2014).
6. V. Sitras, C. Fenton, R. Paulssen, Å. Vårtun, G. Acharya, Differences in gene expression between first and third trimester human placenta: A microarray study. *PLOS ONE* **7**, e33294 (2012).
7. Y. C. Lim, J. Li, Y. Ni, Q. Liang, J. Zhang, G. S. H. Yeo, J. Lyu, S. Jin, C. Ding, A complex association between DNA methylation and gene expression in human placenta at first and third trimesters. *PLOS ONE* **12**, e0181155–e0181155 (2017).
8. B. Cox, K. Leavey, U. Nosi, F. Wong, J. Kingdom, Placental transcriptome in development and pathology: Expression, function, and methods of analysis. *Am. J. Obstet. Gynecol.* **213**, S138–S151 (2015).
9. H. R. Eidem, W. E. Ackerman, K. L. McGary, P. Abbot, A. Rokas, Gestational tissue transcriptomics in term and preterm human pregnancies: A systematic review and meta-analysis. *BMC Med. Genomics* **8**, 27 (2015).
10. K. Leavey, S. A. Bainbridge, B. J. Cox, Large scale aggregate microarray analysis reveals three distinct molecular subclasses of human preeclampsia. *PLOS ONE* **10**, e0116508 (2015).
11. M. Knöfler, S. Haider, L. Saleh, J. Pollheimer, T. K. J. B. Gamage, J. James, Human placenta and trophoblast development: Key molecular mechanisms and model systems. *Cell. Mol. Life Sci.* **76**, 3479–3496 (2019).
12. M. Rebeiz, N. H. Patel, V. F. Hinman, Unraveling the tangled skein: The evolution of transcriptional regulatory networks in development. *Annu. Rev. Genom. Hum. Genet.* **16**, 103–131 (2015).
13. V. Y. Muley, R. König, Human transcriptional gene regulatory network compiled from 14 data resources. *Biochimie* **193**, 115–125 (2022).
14. J. P. Faria, R. Overbeek, F. Xia, M. Rocha, I. Rocha, C. S. Henry, Genome-scale bacterial transcriptional regulatory networks: Reconstruction and integrated analysis with metabolic models. *Brief. Bioinform.* **15**, 592–611 (2014).
15. T. R. Hughes, C. G. de Boer, Mapping yeast transcriptional networks. *Genetics* **195**, 9–36 (2013).
16. B. He, K. Tan, Understanding transcriptional regulatory networks using computational models. *Curr. Opin. Genet. Dev.* **37**, 101–108 (2016).
17. C. C. Funk, A. M. Casella, S. Jung, M. A. Richards, A. Rodriguez, P. Shannon, R. Donovan-Maiye, B. Heavner, K. Chard, Y. Xiao, G. Glusman, N. Ertekin-Taner, T. E. Golde, A. Toga, L. Hood, J. D. Van Horn, C. Kesselman, I. Foster, R. Madduri, N. D. Price, S. A. Ament, Atlas of transcription factor binding sites from ENCODE DNase hypersensitivity data across 27 tissue types. *Cell Rep.* **32**, 108029 (2020).
18. J. R. Pearl, C. Colantuoni, D. E. Bergey, C. C. Funk, P. Shannon, B. Basu, A. M. Casella, R. T. Oshone, L. Hood, N. D. Price, S. A. Ament, Genome-scale transcriptional regulatory network models of psychiatric and neurodegenerative disorders. *Cell Syst.* **8**, 122–135.e7 (2019).
19. B. Readhead, J.-V. Haure-Mirande, C. C. Funk, M. A. Richards, P. Shannon, V. Haroutunian, M. Sano, W. S. Liang, N. D. Beckmann, N. D. Price, E. M. Reiman, E. E. Schadt, M. E. Ehrlich, S. Gandy, J. T. Dudley, Multiscale analysis of independent Alzheimer's cohorts finds disruption of molecular, genetic, and clinical networks by human herpesvirus. *Neuron* **99**, 64–82.e7 (2018).
20. S. A. Ament, J. R. Pearl, J. P. Cantle, R. M. Bragg, P. J. Skene, S. R. Coffey, D. E. Bergey, V. C. Wheeler, M. E. MacDonald, N. S. Baliga, J. Rosinski, L. E. Hood, J. B. Carroll, N. D. Price, Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. *Mol. Syst. Biol.* **14**, e7435–e7435 (2018).
21. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics.* **9**, 559 (2008).
22. S. Buckberry, T. Bianco-Miotto, S. J. Bent, V. Clifton, C. Shoubridge, K. Shankar, C. T. Roberts, Placental transcriptome co-expression analysis reveals conserved regulatory programs across gestation. *BMC Genomics* **18**, 10 (2017).
23. M. A. Deyssenroth, S. Peng, K. Hao, L. Lambertini, C. J. Marsit, J. Chen, Whole-transcriptome analysis delineates the human placenta gene network and its associations with fetal growth. *BMC Genomics* **18**, 520 (2017).
24. D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
25. C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, R. Zhang, B. M. Hartmann, E. Zaslavsky, S. C. Sealfon, D. I. Chasman, G. A. FitzGerald, K. Dolinski, T. Grosser, O. G. Troyanskaya, Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).
26. Y. Liu, D. Ding, H. Liu, X. Sun, The accessible chromatin landscape during conversion of human embryonic stem cells to trophoblast by bone morphogenetic protein 4. *Biol. Reprod.* **96**, 1267–1278 (2017).
27. S. Buckberry, T. Bianco-Miotto, S. J. Bent, G. A. Dekker, C. T. Roberts, Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Mol. Hum. Reprod.* **20**, 810–819 (2014).
28. S. Lorente-Pozo, A. Parra-Llorca, B. Torres, I. Torres-Cuevas, A. Nuñez-Ramiro, M. Cernada, A. García-Robles, M. Vento, Influence of sex on gestational complications, fetal-to-neonatal transition, and postnatal adaptation. *Front. Pediatr.* **6**, 63 (2018).
29. A. G. Paquette, J. Macdonald, T. Bammler, D. B. Day, C. T. Loftus, E. Buth, W. A. Mason, N. R. Bush, K. Z. Lewinn, C. Marsit, J. A. Litch, M. Gravett, D. A. Enquobahrie, S. Sathyanarayana, Placental transcriptomic signatures of spontaneous preterm birth. *Am. J. Obstet. Gynecol.* **228**, 73.e1–73.e18 (2023).
30. K. A. Campbell, J. A. Colacino, M. Puttabyatappa, J. F. Dou, E. R. Elkin, S. S. Hammoud, S. E. Domino, D. C. Dolinoy, J. M. Goodrich, V. Loch-Caruso, V. Padmanabhan, K. M. Bakulski, Placental cell type deconvolution reveals that cell proportions drive preeclampsia gene expression differences. *Commun. Biol.* **6**, 264 (2023).
31. M. Uhlen, P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Björling, F. Ponten, Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
32. T. M. Everson, C. J. Marsit, Integrating -omics approaches into human population-based studies of prenatal and early-life exposures. *Curr. Environ. Health Rep.* **5**, 328–337 (2018).
33. S. Lapehn, A. G. Paquette, The placental epigenome as a molecular link between prenatal exposures and fetal health outcomes through the DOHaD hypothesis. *Curr. Environ. Health Rep.* **9**, 490–501 (2022).
34. A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, A. Ma'ayan, ChEA: Transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).

35. C. C. Funk, S. Jung, M. A. Richards, A. Rodriguez, P. Shannon, R. Donovan, B. Heavner, K. Chard, Y. Xiao, G. Glusman, N. Erteskin-Taner, T. Golde, A. Toga, L. Hood, J. D. Van Horn, C. Kesselman, I. Foster, S. Ament, R. Madduri, N. D. Price, Atlas of transcription factor binding sites from ENCODE DNase hypersensitivity data across 27 tissue types. bioRxiv 252023 [Preprint] (2018). doi:10.1101/252023.

36. S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, M. T. Weirauch, The human transcription factors. Cell 172, 650–665 (2018).

37. D. Baczyk, S. Drewlo, L. Proctor, C. Dunk, S. Lye, J. Kingdom, Glial cell missing-1 transcription factor is required for the differentiation of the human trophoblast. Cell Death Differ. 16, 719–727 (2009).

38. Y.-H. Chiu, M.-R. Yang, L.-J. Wang, M.-H. Chen, G.-D. Chang, H. Chen, New insights into the regulation of placental growth factor gene expression by the transcription factors GCM1 and DLX3 in human placenta. J. Biol. Chem. 293, 9801–9811 (2018).

39. F. Maleki, K. Ovens, D. J. Hogan, A. J. Kusalik, Gene set analysis: Challenges, opportunities, and future research. Front. Genet. 11, 654 (2020).

40. J. N. Li, Y. C. Ge, Z. Yang, C. M. Guo, T. Duan, L. Myatt, H. Guan, K. Yang, K. Sun, The Sp1 transcription factor is crucial for the expression of 11β-hydroxysteroid dehydrogenase type 2 in human placental trophoblasts. J. Clin. Endocrinol. Metabol. 96, E899–E907 (2011).

41. S. Sõber, M. Reiman, T. Kikas, K. Rull, R. Inno, P. Vaas, P. Teesalu, J. M. L. Marti, P. Mattila, M. Laan, Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. Sci. Rep. 5, 13336–13336 (2015).

42. E. A. Adu-Gyamfi, Y.-B. Ding, Y.-X. Wang, Regulation of placentation by the transforming growth factor beta superfamily†. Biol. Reprod. 102, 18–26 (2020).

43. Y.-H. Cheng, B. J. Aronow, S. Hossain, B. Trapnell, S. Kong, S. Handwerger, Critical role for transcription factor AP-2alpha in human trophoblast differentiation. Physiol. Genomics 18, 99–107 (2004).

44. K. Biadasiewicz, S. Sonderegger, P. Haslinger, S. Haider, L. Saleh, C. Fiala, J. Pollheimer, M. Knöfler, Transcription factor AP-2α promotes EGF-dependent invasion of human trophoblast. Endocrinology 152, 1458–1469 (2011).

45. Y.-H. Cheng, S. Handwerger, AP-2alpha modulates human corticotropin-releasing hormone gene expression in the placenta by direct protein-protein interaction. Mol. Cell. Endocrinol. 191, 127–136 (2002).

46. M. Li, R. E. Kellems, Sp1 and Sp3 are important regulators of AP-2gamma gene transcription. Biol. Reprod. 69, 1220–1230 (2003).

47. A. Gabory, T. J. Roseboom, T. Moore, L. G. Moore, C. Junien, Placental contribution to the origins of sexual dimorphism in health and diseases: Sex chromosomes and epigenetics. Biol. Sex Differ. 4, 5–5 (2013).

48. C. S. Rosenfeld, Sex-specific placental responses in fetal development. Endocrinology 156, 3422–3434 (2015).

49. J. G. Eriksson, E. Kajantie, C. Osmond, K. Thornburg, D. J. P. Barker, Boys live dangerously in the womb. Am. J. Hum. Biol. 22, 330–335 (2010).

50. J. Zeitlin, M.-J. Saurel-Cubizolles, J. De Mouzon, L. Rivera, P.-Y. Ancel, B. Blondel, M. Kaminski, Fetal sex and preterm birth: Are males at greater risk? Hum. Reprod. 17, 2762–2768 (2002).

51. M. Knofler, R. Vasicek, M. Schreiber, Key regulatory transcription factors involved in placental trophoblast development—A review. Placenta 22, S83–S92 (2001).

52. X. Lu, Y. He, C. Zhu, H. Wang, S. Chen, H.-Y. Lin, Twist1 is involved in trophoblast syncytialization by regulating GCM1. Placenta 39, 45–54 (2016).

53. M. Pavličev, G. P. Wagner, A. R. Chavan, K. Owens, J. Maziarz, C. Dunn-Fletcher, S. G. Kallapur, L. Muglia, H. Jones, Single-cell transcriptomics of the human placenta: Inferring the cell communication network of the maternal-fetal interface. Genome Res. 27, 349–361 (2017).

54. N. J. Sakabe, I. Aneas, N. Knoblauch, D. R. Sobreira, N. Clark, C. Paz, C. Horth, R. Ziffra, H. Kaur, X. Liu, R. Anderson, J. Morrison, V. C. Cheung, C. Grotegut, T. E. Reddy, B. Jacobsson, M. Hallman, K. Teramo, A. Murtha, J. Kessler, W. Grobman, G. Zhang, L. J. Muglia, S. Rana, V. J. Lynch, G. E. Crawford, C. Ober, X. He, M. A. Nóbrega, Transcriptome and regulatory maps of decidua-derived stromal cells inform gene discovery in preterm birth. Sci. Adv. 6, eabc8696 (2020).

55. A. Stephanou, N. J. Sarlis, R. Richards, S. Handwerger, Expression of retinoic acid receptor subtypes and cellular retinoic acid binding protein-II mRNAs during differentiation of human trophoblast cells. Biochem. Biophys. Res. Commun. 202, 772–780 (1994).

56. A. Kannan, J. R. Beal, A. M. Neff, M. K. Bagchi, I. C. Bagchi, Runx1 regulates critical factors that control uterine angiogenesis and trophoblast differentiation during placental development. PNAS Nexus 2, pgad215 (2023).

57. F. Wen, J. A. Tynan, G. Cecena, R. Williams, J. Múnera, G. Mavrothalassitis, R. G. Oshima, Ets2 is required for trophoblast stem cell self-renewal. Dev. Biol. 312, 284–299 (2007).

58. R. Romero, L. Yeo, P. Chaemsaithong, T. Chaiworapongsa, S. S. Hassan, Progesterone to prevent spontaneous preterm birth. Semin. Fetal and Neonatal Med. 19, 15–26 (2014).

59. J. M. P. Pabona, D. Zhang, D. S. Ginsburg, F. A. Simmen, R. C. M. Simmen, Prolonged pregnancy in women is associated with attenuated myometrial expression of progesterone receptor co-regulator Krüppel-like Factor 9. J. Clin. Endocrinol. Metab. 100, 166–174 (2015).

60. M. Lappas, Forkhead box O1 (FOXO1) in pregnant human myometrial cells: A role as a pro-inflammatory mediator in human parturition. J. Reprod. Immunol. 99, 24–32 (2013).

61. H. R. Eidem, D. C. Rinker, W. E. Ackerman IV, I. A. Buhimschi, C. S. Buhimschi, C. Dunn-Fletcher, S. G. Kallapur, M. Pavličev, L. J. Muglia, P. Abbot, A. Rokas, Comparing human and macaque placental transcriptomes to disentangle preterm birth pathology from gestational age effects. Placenta 41, 74–82 (2016).

62. D. Brubaker, Y. Liu, J. Wang, H. Tan, G. Zhang, B. Jacobsson, L. Muglia, S. Mesiano, M. R. Chance, Finding lost genes in GWAS via integrative-omics analysis reveals novel sub-networks associated with preterm birth. Hum. Mol. Genet. 25, 5254–5264 (2016).

63. R. B. Brandt, D. G. Mueller, J. R. Schroeder, K. E. Guyer, B. V. Kirkpatrick, N. E. Hutcher, F. E. Ehrlich, Serum vitamin A in premature and term neonates. J. Pediatr. 92, 101–104 (1978).

64. Y.-A. You, S.-Y. Hwang, S.-M. Kim, S. Park, G.-I. Lee, S. Park, A. Ansari, J. Lee, Y. Kwon, Y.-J. Kim, Identification of indicators for preterm birth using retinoid metabolites. Metabolites. 11, 443 (2021).

65. J. E. Norman, Progesterone and preterm birth. Int. J. Gynecol. Obstet. 150, 24–30 (2020).

66. D. B. Nelson, D. D. McIntire, J. McDonald, J. Gard, P. Turrichi, K. J. Leveno, 17-alpha Hydroxyprogesterone caproate did not reduce the rate of recurrent preterm birth in a prospective cohort study. Am. J. Obstet. Gynecol. 216, 600.e1–600.e9 (2017).

67. P. J. Meis, M. Klebanoff, E. Thom, M. P. Dombrowski, B. Sibai, A. H. Moawad, C. Y. Spong, J. C. Hauth, M. Miodovnik, M. W. Varner, K. J. Leveno, S. N. Caritis, J. D. Iams, R. J. Wapner, D. Conway, M. J. O'Sullivan, M. Carpenter, B. Mercer, S. M. Ramin, J. M. Thorp, A. M. Peaceman, Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. N. Engl. J. Med. 348, 2379–2385 (2003).

68. H. Mactier, Vitamin A for preterm infants; where are we now? Semin. Fetal Neonatal Med. 18, 166–171 (2013).

69. B. A. Darlow, P. J. Graham, M. X. Rojas-Reyes, Vitamin A supplementation to prevent mortality and short- and long-term morbidity in very low birth weight infants. Cochrane Database Syst. Rev. 2016, CD000501 (2016).

70. T. R. Butt, S. K. Karathanasis, Transcription factors as drug targets: Opportunities for therapeutic selectivity. Gene Expr. 4, 319–336 (1995).

71. C. Buccitelli, M. Selbach, mRNAs, proteins and the emerging principles of gene expression control. Nat. Rev. Genet. 21, 630–644 (2020).

72. H. Koohy, T. A. Down, M. Spivakov, T. Hubbard, A comparison of peak callers used for DNase-seq data. PLOS ONE 9, e96303 (2014).

73. P. Kothiyal, K. Schulkers, X. Liu, S. Hazrati, T. Vilboux, L. M. Gomez, K. Huddleston, W. S. W. Wong, J. E. Niederhuber, T. P. Conrads, G. L. Maxwell, S. K. Hourigan, Differences in maternal gene expression in Cesarean section delivery compared with vaginal delivery. Sci. Rep. 10, 17797 (2020).

74. J. T. Leek, J. D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis. PLOS Genet. 3, 1724–1735 (2007).

75. J. Sung, Y. Wang, S. Chandrasekaran, D. M. Witten, N. D. Price, Molecular signatures from omics data: From chaos to consensus. Biotechnol. J. 7, 946–957 (2012).

76. D. Chicco, Ten quick tips for machine learning in computational biology. BioData Min. 10, 35 (2017).

77. S. M. Wyatt, F. T. Kraus, C.-R. Roh, U. Elchalal, D. M. Nelson, Y. Sadovsky, The correlation between sampling site and gene expression in the term human placenta. Placenta 26, 372–379 (2005).

78. K. Z. LeWinn, C. J. Karr, M. Hazlehurst, K. Carroll, C. Loftus, R. Nguyen, E. Barrett, S. H. Swan, A. A. Szpiro, A. Paquette, P. Moore, E. Spalt, L. Younglove, A. Sullivan, T. Colburn, N. Byington, L. Sims Taylor, S. Moe, S. Wang, A. Cordeiro, A. Mattias, J. Powell, T. Johnson, A. Norona-Zhou, A. Mason, N. R. Bush, S. Sathyanarayana, Cohort profile: The ECHO prenatal and early childhood pathways to health consortium (ECHO-PATHWAYS). BMJ Open 12, e064288 (2022).

79. A. G. Paquette, J. MacDonald, S. Lapehn, T. Bammler, L. Kruger, D. B. Day, N. D. Price, C. Loftus, K. Kannan, C. Marsit, W. A. Mason, N. R. Bush, K. Z. LeWinn, D. A. Enquobahrie, B. Prasad, C. J. Karr, S. Sathyanarayana, A comprehensive assessment of associations between prenatal phthalate exposure and the placental transcriptomic landscape. Environ. Health Perspect. 129, 097003 (2021).

80. A. L. Eis, D. E. Brockman, J. S. Pollock, L. Myatt, Immunohistochemical localization of endothelial nitric oxide synthase in human villous and extravillous trophoblast populations and expression during syncytiotrophoblast formation in vitro. Placenta 16, 113–126 (1995).

81. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. Nat. Biotech. 34, 525–527 (2016).

82. J. Brown, M. Pirrung, L. A. McCue, FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33, 3137–3139 (2017).

83. C. Soneson, M. Love, M. Robinson, Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. F1000Research 4, 1521 (2015).

84. Y. Chen, A. Lun, G. Smyth, From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research 5, 1438 (2016).

85. A. G. Paquette, H. M. Brockway, N. D. Price, L. J. Muglia, Comparative transcriptomic analysis of human placentae at term and preterm delivery. *Biol. Reprod.* **98**, 89–101 (2018).

86. A. Lachmann, D. Torre, A. B. Keenan, K. M. Jagodnik, H. J. Lee, L. Wang, M. C. Silverstein, A. Ma'ayan, Massive mining of publicly available RNA-seq data from human and mouse. *Nat. Commun.* **9**, 1366 (2018).

87. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

88. A. P. Boyle, J. Guinney, G. E. Crawford, T. S. Furey, F-Seq: A feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**, 2537–2538 (2008).

89. E. G. Gusmao, C. Dieterich, M. Zenke, I. G. Costa, Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. *Bioinformatics* **30**, 3143–3151 (2014).

90. A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Chèneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018).

91. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, V. J. Makeev, HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

92. M. Pachkov, I. Erb, N. Molina, E. van Nimwegen, SwissRegulon: A database of genome-wide annotations of regulatory sites. *Nucleic Acids Res.* **35**, D127–D131 (2007).

93. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

94. Paul Shannon, MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs (2017).

95. C. L. Plaisier, S. O'Brien, B. Bernard, S. Reynolds, Z. Simon, C. M. Toledo, Y. Ding, D. J. Reiss, P. J. Paddison, N. S. Baliga, Causal mechanistic regulatory network for glioblastoma deciphered using systems genetics network analysis. *Cell Syst.* **3**, 172–186 (2016).

96. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, bax028 (2017).

97. E. R. Miraldi, M. Pokrovskii, A. Watters, D. M. Castro, N. De Veaux, J. A. Hall, J.-Y. Lee, M. Ciofani, A. Madar, N. Carriero, D. R. Littman, R. Bonneau, Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. *Genome Res.* **29**, 449–463 (2019).

98. A. T. L. Lun, Y. Chen, G. K. Smyth, It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. *Methods Mol. Biol.* **1418**, 391–416 (2016).

99. S. Falcon, R. Gentleman, Hypergeometric testing used for gene set enrichment analysis in *Bioconductor Case Studies*, F. Hahne, W. Huber, R. Gentleman, S. Falcon, Eds. (Springer New York, 2008; https://doi.org/10.1007/978-0-387-77240-0_14), pp. 207–220.

100. K. Wijesooriya, S. A. Jadaan, K. L. Perera, T. Kaur, M. Ziemann, Urgent need for consistent standards in functional enrichment analysis. *PLOS Comput. Biol.* **18**, e1009935 (2022).

101. J. A. Timmons, K. J. Szkop, I. J. Gallagher, Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* **16**, 186 (2015).

102. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B.* **57**, 289–300 (1995).

103. A. Paquette, Placental TRN Manuscript Code (2023), (available at https://doi.org/10.5281/zenodo.7741718).