SUPPLEMENTAL MATERIAL

Supplemental Methods:

Patients

Microarray analysis of human placenta and decidua samples were described earlier[1]. The clinical biomarkers (CTproAVP, MRproADM, MRproANP, CRP) used for clustering analysis were described earlier[2, 3]. The study comprises of patient samples from bio-bank collection at Oslo University Hospital, Oslo, Norway, approved by the Regional Committee of Medical Research Ethics in South-Eastern Norway. The cohort consists of placental and decidual tissues collected following caesarean sections from 56 preeclamptic women (PE) and 28 women with normotensive and uncomplicated pregnancies. The PE group was subdivided into PE + IUGR (n = 14) and PE without IUGR (n = 42). Furthermore, the PE group was divided into early-onset PE (delivery <34 gestational week) and late-onset PE (delivery ≥34 gestational week)[4]. Placenta biopsies were collected according to a strict protocol, following stat-of-the-art sampling and storage[5]. In brief, biopsies were collected from the center of a macroscopically normal looking central cotyledon within 5 minutes of delivery by study personnel. Biopsies were snap-frozen in liquid nitrogen and stored at –80°C until further processing. Decidual tissue was collected through vacuum suctioning of the placental bed[6]. The levels of serum PlGF and sFLT1 of patients from the second preeclamptic cohort used for correlation analysis were measured on Elecsys (Roche Diagnostics) at HELIOS Kliniken GmbH. sFlt1/PlGF ratios were calculated[4].

The healthy (n = 5) and PE (n = 5) primary trophoblast cells were isolated from human placenta samples obtained from HELIOS Klinikum in Berlin. Human placenta sampling was approved by the Regional Committee of the Medical Faculty of Charité Berlin. All placentas were processed within 2 h after delivery. Control trophoblast cells were isolated from placentae of uncomplicated pregnancies collected following Caesarean section at term. PE

was defined by hypertension (SBP ≥140 mmHg or DBP ≥90 mmHg) and proteinuria (≥0.3 g in a 24-hour urine specimen) according to the American College of Obstetricians and Gynecologists (ACOG) classification.

Cluster analysis of microarray data

We employed R platform (https://cran.r-project.org/) and various Bioconductor packages (https://bioconductor.org/biocLite.R) to analyze the human placenta and decidua microarray data. We extracted significance level of normalized expression values against the background corresponding to each probe IDs using "*lumi*", a R Bioconductor package where the variance-stabilizing transformation (VST) was applied to deal with sample replicates and robust spline normalization (RSN) for normalization[7]. Probes with p-value < 0.05 were further transformed to $\log_2$ scale and IDs were annotated as gene names from "*illuminaHumanV3.db*" from Bioconductor annotation data package containing 47,324 probes. Consequently, we generated expression matrix of 38,382 significant probes in this study. Expression values of a gene with multiple probes were assigned by their mean, resulting in 26,886 unique genes. In order to achieve consistency between arrays, $\log_2$ transformed expression values underwent through quantile normalization using "normalizeQuantile" function built-in "*limma*" package. Quality of processed data was further assured by PCA visualization[8]. Each gene value was further assigned as their relative abundance value which is the ratio of expression value of gene in each sample and the mean value of expression for corresponding gene across the samples. The resulting relative expression matrix was subjected to unsupervised hierarchical clustering which blindly classifies all samples based on their transcriptome pattern without the prior knowledge of their disease status. Hierarchical clustered dendrograms were generated using "average" linkage method (mean distance between an observation in one cluster and an observation in the other cluster) on the dissimilarity matrix. Pairwise distances were calculated using Euclidian distance method on dissimilarity matrix obtained by pairwise

Spearman correlation. P-value threshold for correlation test of matrix was kept up to 0.01. In order to examine statistical reliability of clustering we performed bootstrapping (1000 replicates) on unbiased hierarchically clustered dendrogram of distance matrix by "average" method using pvclust[9]. In present study, fold change of differential gene expression (DEG) between samples on the $\log_2$ scale was analyzed using linear and e-Bayesian model algorithms from "*limma*" R Bioconductor package[10]. P-value of DEGs between the three PE clusters (n < 8) vs clustered Controls (n = 16) was obtained from empirical Bayesian built-in function from "*limma"*. We did not adjust the p-values by using multiple hypothesis tests in order to preserve the family-wise error-rate (FWER) that were supportive for further linking the genes with clinical phenotypes. Pairwise differential gene expression is shown on the heatmap displaying $\log_2$ fold change values, which were detected in all three PE clusters against controls at a significant level p-value < 0.05. Venn diagram was used to represent the distribution of differentially expressed gene in the clusters using "*venny*", an online tool (http://bioinfogp.cnb.csic.es/tools/venny/).

Furthermore, imprinted genes were extracted from the differentially expressed genes to compare their expression in the identified preeclamptic clusters. We analyzed differential expression of maternally (MEGs) and paternally (PEGs) expressed genes in all PE placenta samples. The list of imprinted genes was analyzed according to the Geneimprint website that gives known and predicted human imprinted genes (http://www.geneimprint.com/site/home/), and recently published list of imprinted genes in human placenta[11]. We defined imprinting as genes showing statistically significant allelic expression of < 20% mean expression ratio between maternal and paternal allele as a paternally expressed genes, and those showing > 80% as maternally expressed genes according to the allelic expression in the study[11]. This resulted in a total of 257 MEGs and 150 PEGs (Supplementary Table 3). For analysis of imprinting, we considered also M/P allelic expression ratio (M[%] – P[%]) following values:

60-40, 70-30, as well as 90-10. Wilcoxon test was subjected to analyze mean $\log_2$ fold change of differentially expressed MEGs and PEGs in PE.

Pathway analysis of deregulated genes

Canonical pathways and biological function of the differentially expressed genes identified in microarray data sets were investigated using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). The Ingenuity Knowledge Base (www.ingenuity.com) was used as reference gene sets (pathways). Overrepresentation of biological pathways was assessed by Fisher's exact test and corrected for multiple testing by the Benjamini-Hochberg procedure. The ratio (overlap) is calculated as a number of genes from the dataset that map to the pathway divided by the number of total genes included into the pathway.

Clinical data analysis

Clinical data and biomarker parameters of the preeclamptic patients subjected for microarray were analyzed to assess the difference in clinical outcome in each preeclamptic cluster. We leveraged our clinical and transcriptome datasets to infer similarity between clinical phenotypes, patients (including or excluding controls) and moreover, genes contributing to each clinical phenotype by our pairwise correlational analysis. First, we created the matrix of phenotypical values for each individual subject considering only those that had continuous variables (non-binary values of clinical phenotypes) and in addition, we removed those that had missing data points (<75% of samples). It resulted in 33 phenotypes (in black in Suppl. Mat, Table S1. Patient`s clinical data, page) for 46 samples (24 Patients and 22 Controls). Missing data points (>75% of samples) in resulting matrix was assigned as mean value (since we further considered values relative to mean for comparison) of respective clinical variable. Finally, we transformed data points relative to mean of their respective clinical variables. We

then constructed dendrograms via bootstrapping (10 replicates) based hierarchical clustering using ranked correlation and complete linkage method on relative values from the clinical variables as stated above. Height of dendrograms represents the Euclidian distance of dissimilarity matrix; numbers in red and blue indicate au and bp values from bootstrapping. We used similar strategy to perform sample's (Patients with or without Controls) and clinical phenotype's clustering on the basis of variable clinical data. In addition, when analyzing only Patients data we added four more variables (in blue in Suppl. Mat, Table S1. Patient`s clinical data) because this data was available for Patients but not for Controls.

In order to make our analysis consistent and to find association between genes and clinical phenotype, expression data was also assigned as relative to mean values and resulting two dataframes (clinical and expression data) that were merged on the basis of their sample IDs. Pairwise ranked correlation was calculated for each clinical phenotype with each gene across the dataframe; to obtain the genes associated with any of given phenotypes correlational threshold was kept to 60 (any pair doesn't fit to the criterion was given white color in the heat map displaying correlogram, whereas red denotes positive and black as negative associations). As a novelty, we show the list of genes corresponding/associate with the clinical phenotype in the supplement (Suppl. Mat. Figure S2B).

mRNA isolation, RT-PCR

Total RNA was isolated from tissues (homogenized by ceramic beads) and cells using QIAzol lysis reagent and Qiagen RNeasy mini kit (including on-column DNAase I digestion) (Qiagen) according to the manufacturer's protocol. RNA quality and concentration was measured by NanoDrop-1000 spectrophotometer (PeqLab). RNA was reverse transcribed into cDNA using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and detected by real-time polymerase chain reaction (PCR) on ABI 7500 Fast Sequence Detection

System (Applied Biosystems). Data was analyzed by 7500 Fast System Software (Applied Biosystems). Primers and probes were designed with PrimerExpress 3.0 (Applied Biosystems), synthetized by Biotez (Germany) and are listed in Table S2. The expression of all of the genes was normalized to 18S expression.

Genomic DNA extraction

Frozen placenta tissues were pulverized with liquid nitrogen. DNA from placentas was extracted using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. DNA was stored at −20°C.

Loss of imprinting assay

Loss of imprinting assay (LOI) was carried for DLX5 gene. Detail of assay design has been already described[12]. To test differential allelic mRNA expression a single nucleotide polymorphism (SNP) for DLX5 (rs73708843) was used. gDNA from 97 preeclamptic and controls placental samples were genotyped for heterozygous SNP using standard sequencing protocol. Next cDNAs from heterozygous placental samples were amplified with gene-specific primers bracketing the readout polymorphism using FastStar Essential DNA Green Master (Roche). The reaction was carried on 20 ng cDNA templates with 0.2mM primers in a final volume of 20 µl. Cycling conditions were: 95.0°C for 10 min, followed by 35 cycles of 95.0°C for 30 sec, 70.0°C for 30 sec and 72.0°C for 30 sec. Quantitative allele-specific PCR (qAS-PCR) was detected using allele-specific primers. These primers were optimized for allelic discrimination using heterozygote genomic DNA amplified on the LightCycler480™ (Roche). The reaction was carried with FastStar Essential DNA Green Master (Roche) on 20 ng gDNA template, 0.2mM primers and final volume 20 µl. Cycling conditions were: 95.0°C for 10 min, followed by 45 cycles of 95.0°C for 30 sec, 70.0°C for 30 sec and 72.0°C for 30 sec.

PCR cDNA amplicons from heterozygous placental samples for the selected readout SNPs were diluted 1:40-1:100 accordingly to the abundance of the mRNA and amplified on the LightCycler480™ (Roche) in the reaction containing: FastStar Essential DNA Green Master Mix (Roche), 0.2mM primers, diluted DNA template; final volume 20 µl. Cycling conditions were: 95.0°C for 10 min, followed by 45 cycles of 95.0°C for 30 sec, 70.0°C for 30 sec and 72.0°C for 30 sec. The primers are listed in Table S2. LOI was measured as an expression of the silenced allele, which was calculated as:

$LOI = 2^{-|\Delta Cp|}$

where the $-|\Delta Cp|$ corresponds to the absolute difference between allele-specific Cp values on cDNA level corrected for the specificity of the allele-specific PCR[12].

To validate allelic expression of DLX5 in human placenta we subjected cDNA of the placenta samples showing no LOI for standard sequencing analysis.

DLX5 CpG methylation analysis

We reanalysed methylation and gene expression data of a PE dataset[13]. We explored "*methylumi*", "*lumi*", "*minfi*" for methylation analysis. "*minfi*" gave highest level of Pearson's correlation within the groups of patient and controls (data is available). We used Bioconductor package "*minfi*"[14] for methylation data analysis and "*lumi*" for gene expression data analysis (as described previously). We calculated $\log_2$ fold change of differential gene expression level in PE patients compared to controls using linear modelling algorithm from Bioconductor package "*limma*" with multiple testing correction by applying Benjamini and Hochberg's (BH) false discovery rate (FDR) analysis.

We preprocessed and normalized raw data by performing background correction against controls. Furthermore, we constructed data frame of normalized data by assigning the Beta (β) value for each probe across the samples.

We fetched the probe IDs of CpG methylation data frame, falling in the genomic window of DLX5 locus in addition with that of 10KB upstream region from the normalized data frame of 20 PE patients and 20 controls. We calculated the differential methylation regions (for the probes, which had ≥ 50% of methylation in either of the cases). We calculated $\log_2$ fold change of methylation level in PE patients compared to controls using linear modelling algorithm from Bioconductor package "*limma*".

Secondly, we created data frame of DLX5 expression and methylation level across 8 controls and 8 PE patients. We then assigned relative expression and methylation values to each probe across the samples. We calculated pairwise correlation of relative expression of DLX5 with relative methylation of various methylation sites using "*rcorr"* function from *"Hmisc"* package using Spearman's test. For each pair, we obtained p-values using "*cor.test*" and FDR was calculated using Benjamini and Hochberg's (BH) test for multiple testing using *p.adjust* function.

Single cell RNA sequencing data analysis

We reanalyzed single cell RNA sequencing data from two studies for human embryogenesis[15] and for mouse embryogenesis[16].

Reads were mapped to the human genome (hg19) and mouse genome (mm10) using STAR[17] with our defined settings i.e. *--alignIntronMin 20 --alignIntronMax 1000000 --chimSegmentMin 15 --chimJunctionOverhangMin 15 --outFilterMultimapNmax 20*, and only uniquely mapped reads were considered. We obtained counts using featureCounts[18] at gene level with RefSeq annotations. Gene expression levels were calculated at Transcript Per Million (TPM) from counts over the entire gene (defined as any transcript locating between TSS and TES). This we did using our *in-house* R script (available on request). We chose samples expressing more than 5000 genes. We subsequently selected for further analysis those genes that express ($\log_2$ TPM > 2) in at least 5 of these samples. This resulted in 1285

single cells expressing 15501 genes for human samples, whereas for mouse we recovered 259 single cells expressing 15181 genes (Note: DLX5 was not recovered in mouse as it didn`t pass the threshold). We adjusted the batch effects arising from single cell sequencing using COMBAT[19] from SVA[20] package in R.

Each gene was represented by an across sample vector of z-scores. The resulting z vectors for all the genes were clustered by Pearson's correlation method. Single-cell RNA-seq data was analyzed using "Seurat" R package[21]. In order to obtain significant clusters of cell population at E4 and E5 stage, we first excluded all genes that were very lowly expressed on average (log(TPM+1)<2) and identified the set of genes that was most variable across the dataset by z-score cutoffs, as described in[22]. We next reduced the dimensionality of our dataset using PCA. As previously described in[22], we ran PCA using the '*prcomp*' function in R using most variable genes as an input. We did 200 random samplings to find significant genes, each time randomly permute 1% of genes. This gave us p-value for each gene in each PC, based on how likely the gene/PC score would have been observed by chance. Next, we utilized a modified randomization approach ('jack straw'), a built-in function in "Seurat" package to identify 'statistically significant' principal components in the datasets. 'Significant' PCs will have a strong enrichment of genes with low p-values. We found PC 1-9 as statistically significant. We then chosen significant genes for PC1-9, allowed each PC to contribute a max of 200 genes (as described in "Seurat" package). When we run tSNE using our 9 significant PCs as an input (spectral tSNE), we got distinct point clouds. Finally, we calculated density cluster of the tSNE map (G.use parameter is the density parameter for the clustering – we set it to 7). This yielded distinct clusters from scRNA-seq datasets. We used recently developed single-cell visualization approach based on the nonlinear dimensionality reduction t-distribution stochastic neighborhood embedding (t-SNE) machine learning algorithm for clustering the transcriptome of human and mouse single cell transcriptomes. We used t-SNE instead PCA since unlike PCA, t-SNE does not map the points between the two contrasting eigen vectors

(Principle components) i.e. higher and lower dimension spaces; t-SNE maps the points in the way that points that are close in the higher dimensional space will be close (with high probability) in the low-dimensionality embedding. It converts similarities between data points to joint probabilities and tries to minimize the divergence between joint probabilities of low-dimensional embedding and high-dimensional data. Differentially expressed genes between single cell clusters were calculated using SCDE algorithms[23].

We ran t-SNE on mouse and human cells and defined clusters of cell populations to identify genes expressing exclusively in those clusters. For human, we redefined clusters of cell populations from 353 single cells only from E4-E5 stages where inner cell mass (ICM) and trophectoderm (TE) split from Morulae. For human, we identified three major clusters containing approximately 90% of total cells and three minor clusters containing < 10% of total cells from E4 and E5 stage of human embryonic development, whereas for mouse we found seven clusters (data not shown). We defined the cell type for each cluster according to the known markers[24, 25]. We identified human TE defining cells and early and late mouse TE cells.

For cynomolgus monkeys (*Macaca fascicularis*), we obtained processed data frame from GEO datasets GSE74767[26] as RPKM and converted it to TPM. We grouped the cells as their pre-implantation embryonic days from E6-E9 and extracted the level of DLX5 expression for each stage. Furthermore, to analyze cell type specific expression pattern of DLX5 at the pre-implantation stage, we analyzed the expression level of DLX5 in pre-implantation hypoblast, ICM, Epiblast, early and late trophectoderm.

For imprinting expression analysis during human embryogenesis, we obtained raw fastq files from GEO database under accession number GSE36552[27]. We calculated the expression on $\log_2$ TPM as above. The data was further scaled to z-score for each gene in order to demonstrate the relative expression. We removed any gene that had < 1 TPM expression level in all the analyzed samples. We used list of imprinted genes as described above and extracted

from the data frame. Hierarchical clustering analysis of imprinted genes expressed during early (oocyte, zygote, 2-cell and 4-cell stage), mid (8-stage, Morulae) embryogenesis and late blastocyst was subjected separately for maternally (179/257) and paternally (86/150) expressed genes.

RNA sequencing of human trophoblasts

Total RNA was isolated from five control and five early-onset PE human trophoblasts using Trizol lysis reagent and Direct-zolTM RNA MiniPrep kit including DNase I on-column digestion (Zymo Research) according to the manufacturer`s protocol. Concentration of RNA was quantified on NanoDrop Spectrophotometer ND-1000 and the quality of RNA was analysed using Agilent RNA 6000 Nano Kit on Agilent 2100 Bioanalyzer machine. Library for RNA sequencing was prepared from 550 ng of RNA using Illumina TruSeq Stranded mRNA LT Set A kit (cat. no. RS-122-2101) according to TruSeq Stranded mRNA Sample Prep LS Protocol. Samples were indexed with sample-specific indices which allow for the pooling and sequencing of all libraries in two pools of five samples. Expression profiling of trophoblast transcriptomes by high throughput sequencing were performed in BIMSB Genomics Platform of Max Delbrück Center for Molecular Medicine (Berlin, Germany) on Illumina HiSeq2000 sequencing platform. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using PE TruSeq Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, sequencing was performed on an Illumina HiSeq 2000 platform as 100 bp first strand specific paired-end reads.

Sample-specific barcoded sequencing reads were de-multiplexed from multiplexed flow cells. The resulting BCL files were converted to FASTQ format files using CASAVA 1.8.2. The quality of the raw sequence reads was determined with the FastQC[28]. Reads with quality score < 30 were removed. We also truncated 2nt from the end of sequencing reads, since their

average quality score was not same as the rest of nucleotides. This resulted in at least 70 million reads per sample. Next, reads were mapped over the reference genome (Human hg19/GRCh37) and transcriptome model (hg19.refseq.gtf), downloaded from USCS tables (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/) using TopHat v2.0.8, samtools 0.1.17.0 and Bowite 2.0.5.0 applying parameters as: *"tophat2 -p 8 –r 150--mate-std-dev 140 --library-type fr-firststrand"*. On average 75% of total reads were uniquely mapped on the annotated gene models, approximately 10% of total reads are uniquely mapped on repeated fraction of genome (data not shown). Transcript assembly for each individual sample was conducted using Cufflinks v2.08 measured as FPKM. For calculation of differentially expressed genes (DEGs) we calculated Counts Per Million (CPM) using featureCounts[18] and algorithms from "*DESeq2*" which performed quantization and statistical inference of systematic changes between conditions, as compared to within-condition variability[29]. The package "*DESeq2*" provides methods to test for differential expression by use of negative binomial generalized linear models; the estimates of dispersion and logarithmic fold changes incorporate data-driven prior distributions. In addition, for the outlier's samples (trophoblast sample PE1) DEGs were calculated using a single-replicate model. The read counts were calculated with *featureCounts* from *subread* package (http://subread.sourceforge.net/), FPKM was calculated using *bamutils* (http://ngsutils.org/modules/bamutils/count/). Next, *Random Variable1 (Var1) = n.l.x* formula was used, where *x* (Random Variable 2) is the expression level of this gene, *n* is reflecting the sequencing depth and *l* is the gene length. The two random variables were used in the published model of GFOLD algorithm, which calculate the normalization constant and variance to extract fold changes from unreplicated RNA-seq data [30].

Distinct loci of repeat elements i.e. obtained from repbase libraries (http://www.girinst.org/repbase/). We counted one alignment per uniquely mapped reads over the distinct loci and calculated counts per million (CPM) normalized to library size. In order

to decipher the expression-set for each repeat elements family, we used both uniquely and non-uniquely mapped reads over the distinct families and calculated expression in similar fashion. Non-unique reads were counted considering only one alignment per read for the each repeat elements family that is mapped exclusively within the respective family. We used Kolmogorov-Smirnov test (KS) on the distribution of RNA-seq read counts per million (log scale) over distinct loci of L1 and SVA elements. Observed p-values from KS test were further processed for multiple testing using Benjamini and Hochberg's (BH) test to calculate FDR.

Immunofluorescent staining

Human control placenta sections were deparaffinized in xylene and followed by rehydration through graded ethanol. Antigen was retrieved by heat-induced epitope retrieval (HIER) method in 0,01 M Citrate buffer pH 7. Non-specific antibody binding was blocked with Ultra V Block (Lab Vision). Double immunostaining of rabbit anti-DLX5 antibody (Sigma-Aldrich) (diluted 1:50 in Antibody Diluent (Dako)) and mouse anti-Cytokeratin-7 (1:1000, Abcam) was applied for 45 minutes. The secondary antibody goat anti-rabbit Alexa 555 (1:200) and goat anti-mouse Alexa 488 (1:200) was incubated for 30 minutes. After extensive washes with phosphate-buffered saline with 0.05% Tween-20 (PBS/T) cell nuclei were stained with DAPI (diluted 1:2000 in PSB) and sections were mounted with ProLong Gold Antifade Mounting Medium (Life Technologies).

Immunohistochemistry staining

Human control and PE placenta sections were deparaffinized and dehydrated. Antigen was retrieved by heat-induced epitope retrieval (HIER) method in 0.01 M Citrate buffer pH 6. Endogenous peroxidase activity was blocked with Hydrogen peroxide (Lab Vision/Thermo scientific) for 10 minutes. Tris-buffered saline with 0.05 % Tween-20 (TBS/T) was used in all

washing steps. Non-specific immunoglobulin binding was blocked with Ultra V Block (Lab Vision). Rabbit anti-DLX5 antibody (Sigma-Aldrich) was diluted (1:200) in Antibody Diluent (Dako) and applied for 45 minutes. The secondary antibody HRP Polymer (ready-to-use; Lab Vision) was incubated for 20 minutes. Peroxidase activity was detected with aminoethylcarbazole (AEC) chromogen (ready-to-use; Lab Vision). All sections were counterstained with hemalaun and mounted with Kaiser's Glyceringelatine (Merck).

Constructs generation

Human DLX5 coding sequence (sequence ID: refNM_005221.5) was amplified by PCR from placental cDNA with primers containing restriction sites for cloning into pCAGGS-HA tag vector. PCR reaction was carried with Phusion High-Fidelity DNA Polymerase (New England BioLabs) on 50ng of cDNA template with 0.2 mM dNTPs, 0.5 μM primers, 3% DMSO in a final volume of 20 μl. Cycling conditions were: 98.0°C for 30s, followed by 35 cycles of 98.0°C for 10 sec, 56.0°C for 30 sec and 72.0°C for 1.5 min and final extension of 10 min at 72.0°C. The PCR product of ~870bp was extracted from 1% agarose gel, digested with EcoRV and NotI restriction enzymes and cloned in pCAGGS-HA vector into the EcoRV/NotI restriction site. It resulted in generation of N-terminal HA-tagged DLX5 expressing construct (pCAGGS-HA-N-DLX5). To generate a construct for stable transgene overexpression, the CAGGS-HA-DLX5 cassette was cut out with SspI restriction enzyme from pCAGGS-HA-N-DLX5 and cloned into EcoRV site, generated by site directed mutagenesis into pT2B-puro *Sleeping Beauty* vector carrying the SV40-puro cassette for puromycin selection.

Cell culture and stable DLX5 overexpression and microarray

SGHPL-4 cells were a kind gift from Judith E. Cartwright (St George's University of London, London, United Kingdom). This cell line is derived from primary human first trimester extravillous trophoblasts (EVT), transfected with the early region of SV40 and show similar

invasive capabilities to primary EVTs and retain features of normal EVTs [31]. SGHPL-4 cells were cultivated in HAM's F10 (Biochrom) media containing 10% (v/v) FCS, 2mM glutamine and 1% (v/v) penicillin/streptomycin (P/S) in 37°C, 5% CO_2.

SGHPL-4 cells were electroporated with 10:1 ratio of vector carrying DLX5 overexpression cassette and plasmid containing *Sleeping Beauty* transposase using Neon Transfection System (Life Technologies). Two days post-transfection puromycin selection was carried for two weeks. DLX5 overexpression was confirmed by RT-qPCR and western blot.

Transcriptome analysis of DLX5-overexpressing SGHPL-4 cells was performed on Illumina HumanHT-12_V4_0_R2 BeadChip according to manufacturer's protocol. Total RNA was isolated from SGHPL-4 cells overexpressing human DLX5 protein using Direct-zol™ RNA MiniPrep kit including DNase I on-column digestion (Zymo Research) according to the manufacturer`s protocol. Six independent transfections as biological replicates were used for the study. Expression values against the background corresponding to each probe IDs were analyzed using "*lumi*" R Bioconductor package. Probes with p-value < 0.05 were further transformed onto $\log_2$ scale. Probe IDs were annotated as gene names from "*illuminaHumanv4.db*" from Bioconductor annotation data package. Expression values of multiple probes for one gene were assigned by their mean. We then adjusted the upper quantile variation, which resulted in normalized data matrix of genes from samples generated in present study. Fold change of differentially expressed genes between wild type and OE-DLX5 samples on the $\log_2$ scale were analyzed using linear and e-Bayesian model algorithms from "*limma*" R Bioconductor package with multiple testing correction by applying Benjamini and Hochberg's (BH) false discovery rate (FDR) analysis on the threshold of 0.05. DEGs are presented on *MA plot* where each point represents the $\log_2$ fold change of genes (y-axis) plotted against the $\log_2$ average expression (x-axis) and the significant level of p-value < 0.05 are colored as red for upregulated genes and blue for downregulated genes.

Intersection of preeclamptic microarray and DLX5 microarray

Quantile normalized datasets from the two microarrays described above were used in the analysis. In order to generate a matrix of expression level for unique genes in each sample, two datasets from different platforms were merged by their unique gene names in total for 58 samples. The batch effect arising from two different platforms was corrected by normalizing surrogate variances from "Combat" package from R Bioconductor[19]. The corrected batch effect was confirmed by Principal Component Analysis (PCA). In order to remove the uneven biases, we considered equal number of patients and controls or equal number of OE-DLX5 and wild type samples prior to intersecting datasets for further analysis. Each gene expression value was further assigned as their relative abundance value, which is the ratio of expression level of a gene in each sample and mean expression values of the gene across the samples. We considered comparison of 6 wild type and 6 OE-DLX5 SGHPL-4 transcriptomes with each of the three preeclamptic clusters and equal number of controls which were randomly chosen. The resulting expression matrix was subjected to hierarchical clustering (Spearman rank correlation, average linkage). P-value threshold for correlation test for matrix was kept up to 0.01 where the hierarchically clustered dendrogram displays gene expression patterns relative to the mean of all the replicated samples included in this analysis. In order to examine statistical reliability of clustering we performed bootstrapping (1000 replicates) on unbiased hierarchically clustered dendrogram of distance matrix using Ward method.

Furthermore, to find a significant impact of DLX5 on PE we clustered the transcriptomes of wild type and OE-DLX5 SGHPL-4 cells with preeclamptic placenta samples that displayed low and high DLX5 expression level, respectively. Six OE-DLX5 and six wild type transcriptomes were analyzed with the equal number of patients expressing higher level of DLX5 and lower level of DLX5. This resulted in two groups of 12 samples: OE-DLX5 and wild type cells, and PE DLX5-high and PE DLX5-low. For each gene, the expression was calculated as relative to the average value of gene expression across the 12

samples for each group. Then both dataframes were merged together to form a single matrix. The batch effect was removed as defined previously. In the next step, genes showing significant variation (we applied Poisson's distribution model on the data, genes lying out of curve were considered as genes showing higher variance in the matrix of 24 samples) were retrieved, which resulted in total of > 3000 genes. Furthermore, we removed 52 genes, which were showing more variation within the group than between the groups. Finally, the heatmap and clusters were constructed as defined previously.

To find a significant correlation gene networks we applied Weighted Gene Co-expression Network Analysis (WGCNA), which is used to identify clusters (modules) of highly correlated genes[32]. We employed this algorithm to find the set of genes co-expressed in our total merged 58 sample's transcriptome. Total microarray datasets of 58 samples including wild type and OE-DLX5 cells and placenta tissues from individuals with and without PE, provides higher level of variability thus giving stronger power to this study to generate a significantly enriched co-expression networks and moreover, to check whether DLX5 is part of them. WGCNA gave us significant co-expression networks as various modules with ~ 3000 genes. We then used DLX5 as a probe to find out the networks associated with DLX5 from WGCNA output. The metamodel derived from Spearman's correlation threshold of positive (60%) and negative (55%) gave the probable significant networks. We kept this threshold to extract all genes, which were correlated at provided correlation thresholds with DLX5 from extracted networks. Finally, pairwise ranked correlation matrix of identified genes was generated.

Western blot

Equal numbers of cells were lysed with 3X Sample Buffer (cooked in 95^{o}C for 7 min) and denaturized proteins were separated on SDS-PAGE gel. Proteins were then transferred to PDVF membrane using semi-dry blotting. Unspecific binding was blocked with 2.5% w/v

non-fat dry milk in TBS-T. Primary antibodies: rabbit anti-DLX5 (1:1000, Sigma-Aldrich) and mouse anti-β-actin (1:1000, Dianova) were incubated overnight at 4°C. Secondary antibody goat anti-rabbit (1:5000, Thermo Scientific) and goat anti-mouse (1:5000, Thermo Scientific) against the host of primary antibody was used conjugated with horseradish peroxidase and membranes were developed with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare). Membranes were then stripped using mild stripping buffer, blocked again and incubated with loading control antibody.

Immunostaining of DLX5 in SGHPL-4 cells

SGHPL-4 wild type and OE-DLX5 cells were seeded in 8-wells chamber slide and incubated in 10% (v/v) FCS Ham´s F10 medium, 1% AA and incubated O/N at 37°C, 5% CO_2. Cells were washed with PBS and 500 µl of ice-cold methanol was added. After 10 min of incubation on ice, methanol was removed. Next, cells were washed 3 times with 500 µl of PBS-T (0.3% TritonX-100). Cells were blocked by incubation with 500 µl of normal donkey serum (10% NDS in PBS) for 30 min at RT. Rabbit anti-DLX5 antibody (Sigma-Aldrich) was diluted (1:50) in PBS with 1% FCS and applied for O/N incubation at 4°C with gentle shaking. After washing, secondary antibody donkey CyTM3 IgG anti-rabbit (1:200; Jackson ImmunoResearch) was added and cells were incubated for 2h at RT in dark. After extensive washes with PBS cell nuclei were stained with DAPI solution (Vectashield with DAPI) for 10 min at RT in dark, washed with PBS and mounted with AquaPoly/Mount solution (Polysciences).

Time-lapse microscopy: proliferation and apoptosis

Proliferation and apoptosis of SGHPL-4 cells was observed over time using an Olympus IX70 inverted microscope equipped with a Hamamatsu C4742-95 digital camera. SGHPL-4 wild type and DLX5-overexpressing cells were preincubated O/N in 0.5% (v/v) FCS in HAM´s

F10 medium (Biochrom) containing 1% P/S. Next control (no treatment), TNFa-treated (30ng/ml) wild type and DLX5-overexpressing cells were followed over 48 h using time-lapse microscopy in HAM´s F10 medium with 0.5% (v/v) FCS, 1% P/S. The microscope and stage were enclosed within a heated (37°C) chamber (Solent Scientific, UK) and cells were cultured in 5% CO_2 in air. Images were captured every 15 min for 48 hours and analyzed using Image Pro Plus software (Media Cybernetics, USA). For analysis, 40 cells in one field of view for each condition were randomly chosen at the beginning of the time-lapse sequence and the distance moved by each cell was recorded. Cell proliferation was assessed as a cumulative number of dividing cells over the 48 h of incubation. Apoptosis was scored according to the time at which clear apoptotic cell morphology was observed during 48 h of incubation (characterized as a membrane blebbing, cytoplasmic shrinkage, nuclear condensation, a phase bright appearance and the formation of blisters)[33].

Cell count using High Throughput Samples (HTS)

SGHPL-4 wild type and DLX5-overexpressing cells were seeded in 12-well plate in 10% (v/v) FCS HAM´s F10 medium (Biochrom), 1% AA and allowed to attach for 3h in 37^{o}C, 5% CO_2. Next control (no treatment) and TNFa-treated (TNFa treatment at conc. 30ng/ml) cells were incubated for 48h in HAM´s F10 medium containing 0.5% (v/v) FCS and 1% Antibiotic-Antimycotic (AA) (Gibco). Cells were trypsinized and collected in a total volume of 100 µl of cell suspension in a 96-well plate and counted using high throughput sampler (HTS) device on BD FACSCanto II system using BD FACSDiva software. Data was analyzed using FlowJo software and is presented as absolute cell number in in SGHPL-4 cell gate.

MTT viability assay

SGHPL-4 cells were seeded in 96-well plate in 100µl of HAM´s F10 medium (Biochrom) with 10% (v/v) FCS, 1% AA and incubated for 48h in 37^{o}C, 5% CO_2. Medium was removed

and replaced with 80µl fresh medium and 20µl of a yellow tetrazole MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) solution (4.15 mg/ml) (Sigma) was added. After 3 h of incubation in 37°C, 5% CO_2 medium was removed and 50µl of DMSO was added to dissolve the insoluble purple formazan. The absorbance was measured at 570 nm by a spectrophotometer.

Mitochondrial assay

To determine the effect of DLX5 overexpression on the mitochondrial respiratory capacity XF Cell Mito Stress Test Kit and XF-24 Extracellular Flux Analyzer (Seahorse Biosciences) was used. Prior to day of mitochondrial assay, sensor cartridge with 1 ml/well XF24 calibrant solution on one utility plate was pre-incubated in a 37°C non-CO_2 incubator overnight. On the day of the assay, SGHPL-4 cells were seeded in the Mito Stress XF24 culture plate in 100µl HAM`s F10 medium with 10% (v/v) FCS and 1% AA and allowed to attach for 4 h (after 1h additional 150µl of medium was added). XF basal medium was supplemented with 6.1 mM D-glucose, 1 mM Na-pyruvate and 0.5% FCS, pH was adjusted to 7.4 and the complete XF medium was sterile filtrated. Next mitochondrial drugs were diluted in the complete XF medium at concentration as follow: 0.75 µM oligomycin, 1 µM FCCP, 1 µM antimycin A + 0.1 µM rotenone and injected into port A, B and C in the pre-incubated sensor cartridge, respectively, and loaded into XF24 Extracellular Flux Analyzer instrument for calibration. In the meantime HAM`s F10 medium in SGHPL-4 cells was replaced with 500 µl of complete XF medium and cells were incubated in a 37°C non-CO_2 incubator for equilibration. After calibration was completed, the utility plate was replaced with pre-incubated cell plate and the mitochondrial activity was measured. Five consecutive measurements were obtained under basal conditions and after the sequential compounds injection as follow:

Command	Time (min)
Calibrate	Fixed
Equilibrate	Fixed

Loop Start Mix Wait Measure Loop End	5x 2 2 3
Inject from port A Loop Start Mix Wait Measure Loop End	 5x 2 2 3
Inject from port B Loop Start Mix Wait Measure Loop End	 5x 2 2 3
Inject from port C Loop Start Mix Wait Measure Loop End	 5x 2 2 3
Program End	

After the run was completed, cells were collected and counted using high throughput sampler (HTS) on BD FACSCanto II system using BD FACSDiva software for normalization. Data was analyzed using FlowJo software. The mitochondrial assay was run in four independent replicates.

ROS production

Wild-type and OE-DLX5 SGHPL-4 cells were seeded in 24-well plate in 1 ml of Ham´s F10 medium with 10% (v/v) FCS, 1% AA and incubated for 24h in 37oC, 5% CO2. Cells were labelled with 0.05mM DCFH-DA for 45 min. After incubation, cells were washed 3 times with PBS, trypsinized and GFP signal was analyzed on FACSCalibur system using CellQuest Software. Data is presented as a mean fluorescent intensity (MFI) of the GFP signal.

ER stress

The degree of ER stress is directly proportional to the degree of oxidative stress. We applied the *in vitro* model of oxidative cellular stress that was developed by Yung et al. using newly derived BeWo cells, BeWo-NG, which grow at physiological concentration of glucose[34]. In short, different degrees of oxidative cellular stress were induced. For the mild and severe oxidative stress, BeWo cells were incubated in serum-free medium (modified DMEM/F12, containing 5.5 mM glucose) under repetitive 6 h cyclic conditions between 5% and 20% O_2 (5/20 H/R) or 1% and 20% O_2 (1/20 H/R), respectively (ExVivo, BioSpherix, USA) for 24 h. Constant 20% O_2 was used as a normoxic control. The severity of ER stress ranks from low to high as follows: 5% O_2 < 1% O_2 < 5%/20% O_2 HR < 1%/20% O_2 HR. After 24h cells were lysed and the DLX5 protein expression was checked by Western blot. To quantify the DLX5 expression in different conditions, ImageJ software was used to measure the intensity of the band.

BeWo cell syncytialization

To study the role of DLX5 in syncytialization, we cultured wild type BeWo cells in 6-well plates in DMEM/F12, 10% (v/v) FCS, 1% AA and incubated for 24h in 37oC, 5% CO_2. Next, we induced syncytium formation with 100μM Forskolin (Sigma). Cells were lysed after 24h and 48h of incubation using with 3X Sample Buffer (cooked in 95^{o}C for 7 min) and denaturized proteins were separated on SDS-PAGE gel. Proteins were then transferred to PDVF membrane using semi-dry blotting. Unspecific binding was blocked with 2.5% w/v non-fat dry milk in TBS-T. Primary antibodies: rabbit anti-DLX5 (1:1000, Sigma-Aldrich), rabbit anti-gCG (1:1000 Abcam), mouse E-cadherin (1:1000, BD Transduction) and mouse anti-β-actin (1:1000, Dianova) were incubated overnight at 4^{o}C. Secondary antibody goat anti-rabbit (1:5000, Thermo Scientific) and goat anti-mouse (1:5000, Thermo Scientific) against the host of primary antibody was used conjugated with horseradish peroxidase and

membranes were developed with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare).

Statistics

Data is presented as either mean ± SEM (for normally distributed data) or median with inter-quartile range (for non-normally distributed data). Normality was assessed by Kolmogorov-Smirnov tests. Data sets were compared using the unpaired t test, nonparametric Mann–Whitney test, one-way ANOVA or Kruskal-Wallis test as appropriate. Post hoc testing included Bonferroni and Dunn´s tests for multigroup comparisons. Statistical correlation was performed using nonparametric Spearman´s rank correlation. Techniques for each analysis are specified in the Figure legends. Two-tailed testing with normal-based 95% confidence interval was performed for each analysis, and $p < 0.05$ was considered statistically significant.

Supplementary Tables

Table S1. Patient`s clinical data (provided as an Excel file).

Clinical data of patients subjected for clustering analysis. Table presents clinical and biomarker parameters, n = 36 (in blue are additional parameters used for analysis of PE patients). Table represents the mean ± SEM value of each parameter in Control and PE patients. PE patients were subdivided into EO-PE, LO-PE and PE+IUGR and the p-value was calculated for each group against Controls either using 2-way ANOVA for normally distributed data (Bonferroni multiple comparison test) or Kruskal-Wallis (Dunn`s multiple comparison test) for not normally distributed data.

Table S2. Ingenuity Pathway Analysis of DEGs in three PE clusters (provided as an Excel file).

Table S3. Imprinted gene expression analysis in PE (provided as an Excel file).

Table S4. Primers and probes

	Gene		Primer	Sequence
RT-qPCR	18S		Forward	5'-ACATCCAAGGAAGGCAGCAG-3'
			Reverse	5'-TTTTCGTCACTACCTCCCCG-3'
			Probe	5'-FAM-CGCGCAAATTACCCACTCCCGAC-TAMRA-3'
	hDLX5		Forward	5'-TCCGATCCGGCGACTTC-3'
			Reverse	5'-CCTGAGACGGATGGTGCATA-3'
			Probe	5'-FAM-AAGCTCCGTTCCAGACGTCCGCA-TAMRA-3'
LOI	Genotyping	hDLX5amp (+)	Forward	5`-AGACTGGGAGTCGTGAAGT-3`
			Reverse	5`-ATTGTCCCCAGACATCACAGA-3`
	Amplicon	LOIhDLX5amp	Forward	5`-ACAGATAGACTAAGACCCCTCCCCAC-3`
			Reverse	5`-CGGGGTGGATCTGGTTCTATTGGC-3`
	qASPCR	qAShDLX5-c	Forward	5`-CCACAACAAGCAAAAACACACACAAGC-3`
		qAShDLX5-t	Forward	5`- CCACAACAAGCAAAAACACACACAAGT-3`
		LOIhDLX5amp	Reverse	5`-CGGGGTGGATCTGGTTCTATTGGC-3`

DLX5 cloning	**hDLX5_EcoRV**	Forward	5´-atcgatatcATGACAGGAGTGTTTGACAG-3´
	hDLX5_NotI	Reverse	5´-atcgcggccgcCTAATAGAGTGTCCCGGAG-3`

Table S5. CpG methylation at the DLX5 locus in control and PE placenta samples (provided as an Excel file).

Table S6. Ingenuity Pathway Analysis of DEGs upon DLX5 overexpression (provided as an Excel file).

Table S7. Imprinted gene expression analysis during human embryogenesis (provided as an Excel file).

Supplementary Figures:

Fig. S1. Cluster analysis of microarray gene expression data.

(A) Clustering analysis based on Euclidean distances of microarray data identified three PE groups in placenta: PE_P1 (blue), PE_P2 (yellow), PE_P3 (green). (Control placenta n = 22, PE placenta n = 24). **(B)** Clustering analysis of decidual gene expression (Control placenta n = 22, PE placenta n = 24).

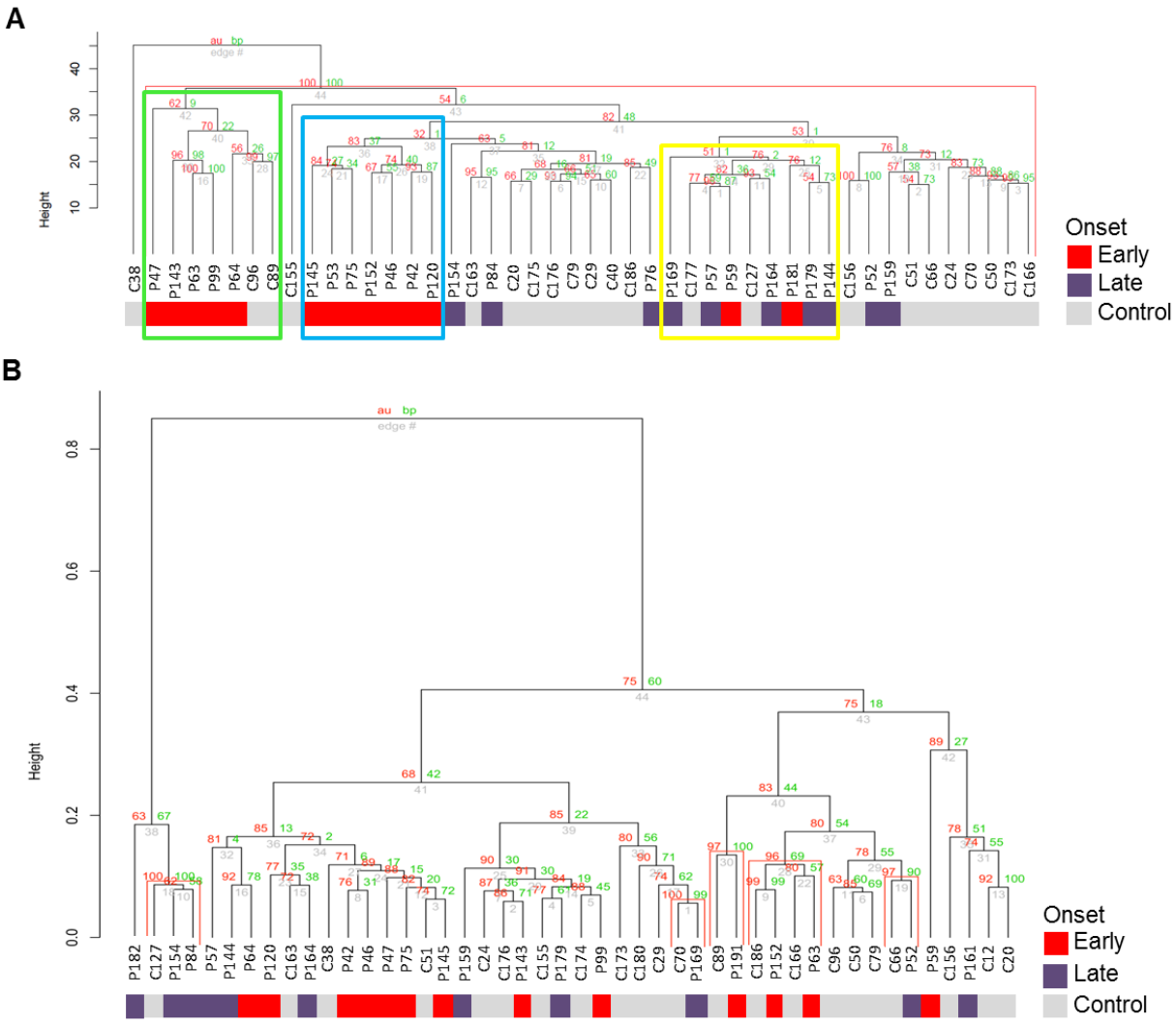
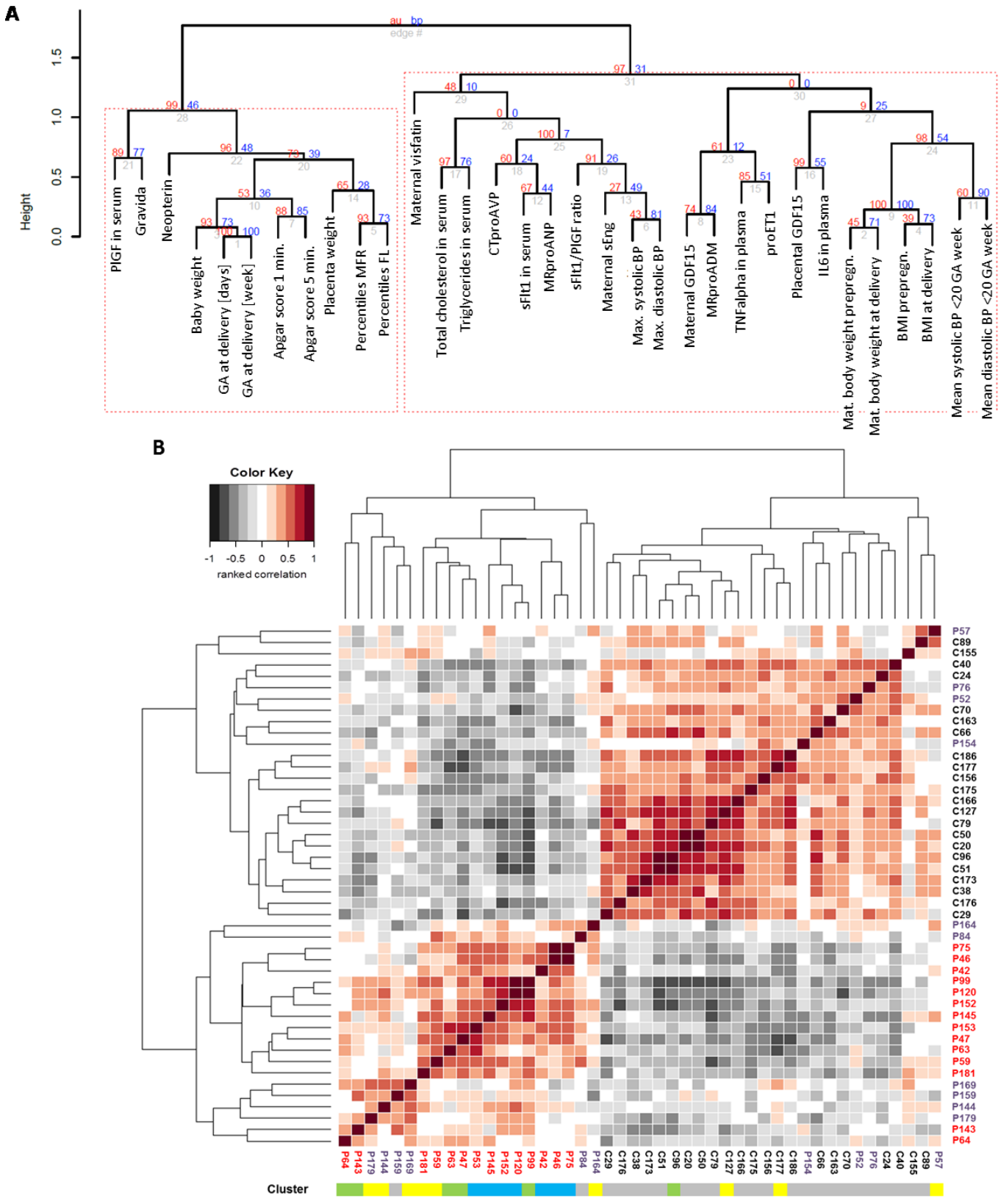


Fig. S2. Clustering analysis of clinical data.

(A) Hierarchical cluster dendrogram calculated using ranked correlation and complete linkage method on relative values from the clinical variables (n = 33). Height of dendrograms represents the Euclidian distance. Numbers in red and blue indicate au and bp values from bootstrapping. **(B)** Heatmap representing ranked correlation between samples based on the clinical data for both Controls and Patients. Most of the Patients and Controls were separated. Only four LO-PE patients were grouped with Controls. Color code as for gene expression clusters: PE_P1 – blue, PE_P2 – yellow, PE_P3 – green. **(C)** Genes associated with the clinical phenotype. Pairwise ranked correlation was calculated for each clinical phenotype with each gene across the samples. Shown pairs are those that passed the correlational threshold of |0.6|.


A
au bp
edge #
Height
0.0
0.5
1.0
1.5
PlGF in serum
Gravida
Neopterin
Baby weight
GA at delivery [days]
GA at delivery [week]
Apgar score 1 min.
Apgar score 5 min.
Placenta weight
Percentiles MFR
Percentiles FL
Maternal visfatin
Total cholesterol in serum
Triglycerides in serum
CTproAVP
sFlt1 in serum
MRproANP
sFlt1/PlGF ratio
Maternal sEng
Max. systolic BP
Max. diastolic BP
Maternal GDF15
MRproADM
TNFalpha in plasma
proET1
Placental GDF15
IL6 in plasma
Mat. body weight prepregn.
Mat. body weight at delivery
BMI prepregn.
BMI at delivery
Mean systolic BP <20 GA week
Mean diastolic BP <20 GA week
B
Color Key
-1 -0.5 0 0.5 1
ranked correlation
Cluster

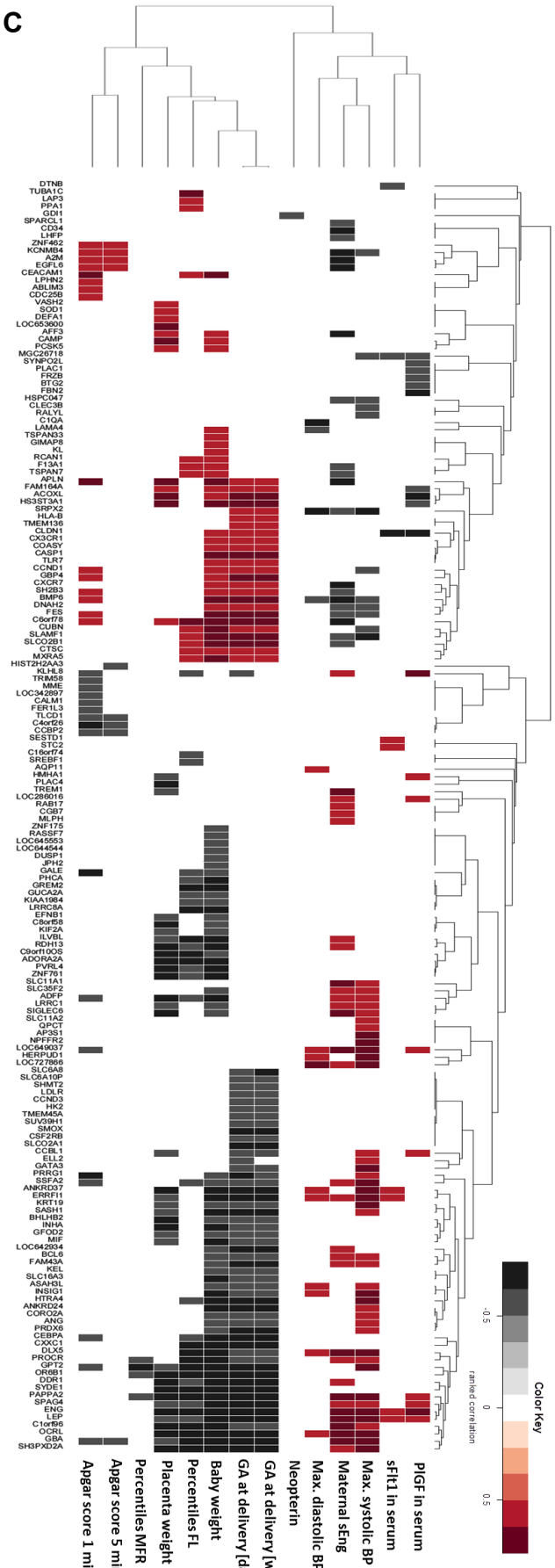

C
PlGF in serum
sFlt1 in serum
Max. systolic BP
Maternal sEng
Max. diastolic BP
Neopterin
GA at delivery [week]
GA at delivery [days]
Baby weight
Percentiles.FL
Placenta weight
Percentiles.MFR
Apgar score 5 min.
Apgar score 1 min.
Color Key
ranked correlation
-0.5
0
0.5

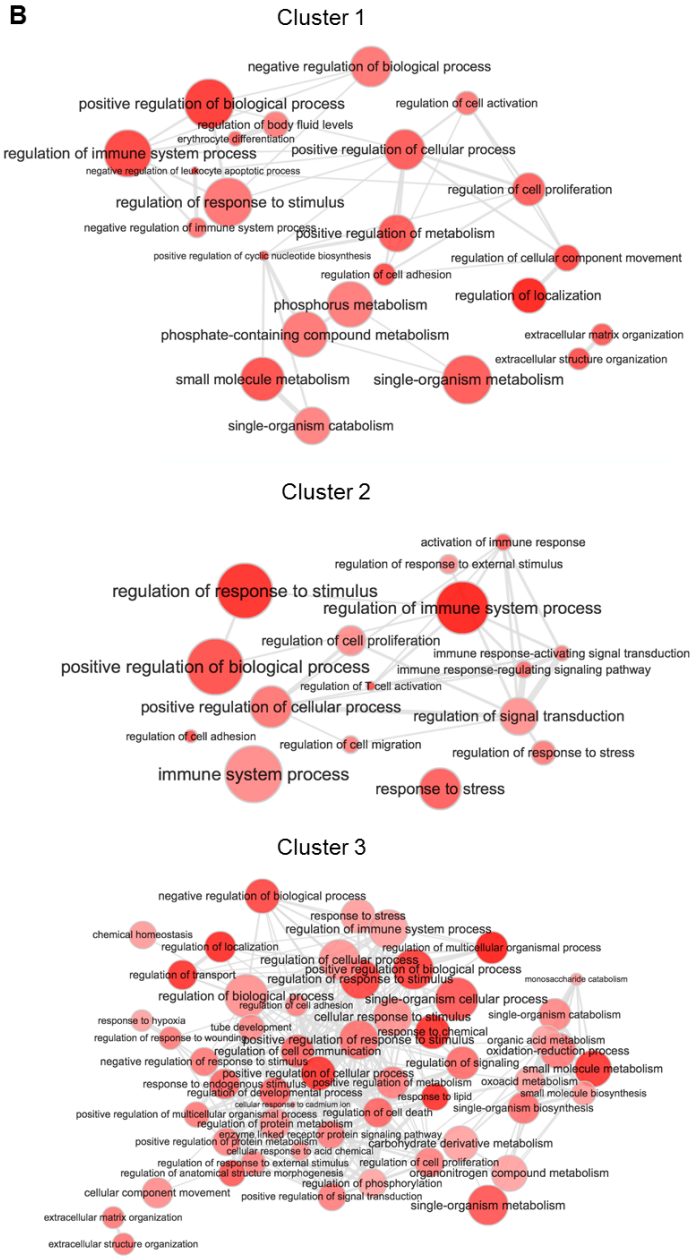
Fig. S3. Molecular characteristics of the three types of PE.

(A) Canonical pathway summary using Ingenuity Pathways Analysis (IPA). The top 5 canonical pathways overrepresented in each cluster of PE placenta. **(B)** Gene ontology enrichment analysis using GOrilla. Presented are gene ontology terms overrepresented in the PE clusters. The size of the circle and intensity of color depends on the enrichment and p-value, respectively.

A

Ingenuity Canonical Pathways	-log(B-H p-value)	Ratio
Cluster 1		
ILK Signaling	3.56	0.285
Hepatic Fibrosis / Hepatic Stellate Cell Activation	3.56	0.279
Axonal Guidance Signaling	2.56	0.222
LXR/RXR Activation	2.33	0.289
p38 MAPK Signaling	2.33	0.291
Cluster 2		
IL-8 Signaling	3.33	0.23
Axonal Guidance Signaling	3.33	0.182
Ephrin Receptor Signaling	3.33	0.23
Breast Cancer Regulation by Stathmin1	3.17	0.22
Androgen Signaling	2.91	0.252
Cluster 3		
NRF2-mediated Oxidative Stress Response	2.51	0.306
G-Protein Coupled Receptor Signaling	2.08	0.27
Axonal Guidance Signaling	2.08	0.245
Gαi Signaling	2.08	0.317
Gαq Signaling	2.08	0.299

B



The analysis revealed that Cluster 1 is enriched in genes involved in ILK signaling, hepatic fibrosis and hepatic stellate cell activation, LXR/RXR activation and p38 MAPK signaling. Among the biological processes regulation of localization and locomotion, negative regulation of leukocyte apoptotic process, regulation of

immune system and regulation of cell motility and cell adhesion were the most overrepresented terms. Cluster 2 is enriched in immune system process and response, regulation of response to stimulus, cell adhesion and T cell activation. Among the top dysregulated pathways, IL-8 signaling, ephrin receptor signaling, androgen signaling and pathway of breast cancer regulation by stathmin-1 are the most significant. Genes identified in Cluster 3 featured dysregulated oxidative stress responses and endothelial dysfunction as the identified enriched genes resided in the NFR2-mediated oxidative stress response and G-protein coupled receptor signaling, as well as Gαi signaling and Gαq signaling. According to the gene ontology analysis the most overrepresented terms are: regulation of multicellular organismal process, regulation of localization and secretion and transport, and response to chemical and lipid. Axonal guidance signaling was the top pathway overrepresented in all three PE clusters.

Fig. S4. DLX5 expression *in vivo*.

(A) Microarray validation of DLX5 overexpression using qPCR. PE samples were subdivided into PE+IUGR and PE without IUGR. Data is presented as median ± IQR (Control 0.765 ± 0.6175-0.9125, n = 26; PE 1.06 ± 0.78-1.25, n = 19; PE+IUGR 1.14 ± 0.88-1.79, n = 7; *p < 0.05, Kruskal-Wallis test, Dunn's multiple comparisons test). **(B)** qPCR validation of DLX5 overexpression in the samples subjected for the microarray analysis according to clusters. Data presented as mean ± SEM (Control Cluster 1 and 2: 0.962 ± 0.045, n = 16; PE Cluster 1: 1.434 ± 0.156, n = 7; PE Cluster 2: 1398 ± 0.186, n = 7; PE Cluster 3: 1.328 ± 0.089, n = 5; *p < 0.05; ANOVA, Bonferroni's multiple comparisons test). **(C)** Western blot analysis of DLX5 expression in Control (n = 10) and PE (n = 10) placenta tissues. Actin was used as a loading control. **(D)** Western blot quantification of DLX5 expression in placenta tissues using ImageJ software. Relative intensity: DLX5/Actin (p=0.0414; Unpaired t-test). **(E)** Dependence between DLX5 expression in the placenta and gestational age in Control and PE samples (Control, n = 28; PE, n = 50, Spearman rank correlation). **(F)** DLX5 expression across different tissues (maternal muscle and fat tissue, decidua, placenta, primary human trophoblasts and fetal macrophages). Data presented as a mean ± SEM; TBs – trophoblasts. **(E)** DLX5 mRNA expression in the decidua of the second cohort shows that DLX5 transcript is increased in late PE. Data shown as mean ± SEM (Control: 1.008 ± 0.115, n = 36; EO-PE+IUGR: 1.338 ± 0.22, n = 8; LO-PE+IUGR: 1.092 ± 0.2, n = 5; EO-PE: 0.916 ± 0.13, n = 13; LO-PE: 1.677 ± 0.28, n = 15) (*p < 0.05; ANOVA, Bonferroni's multiple comparisons test).

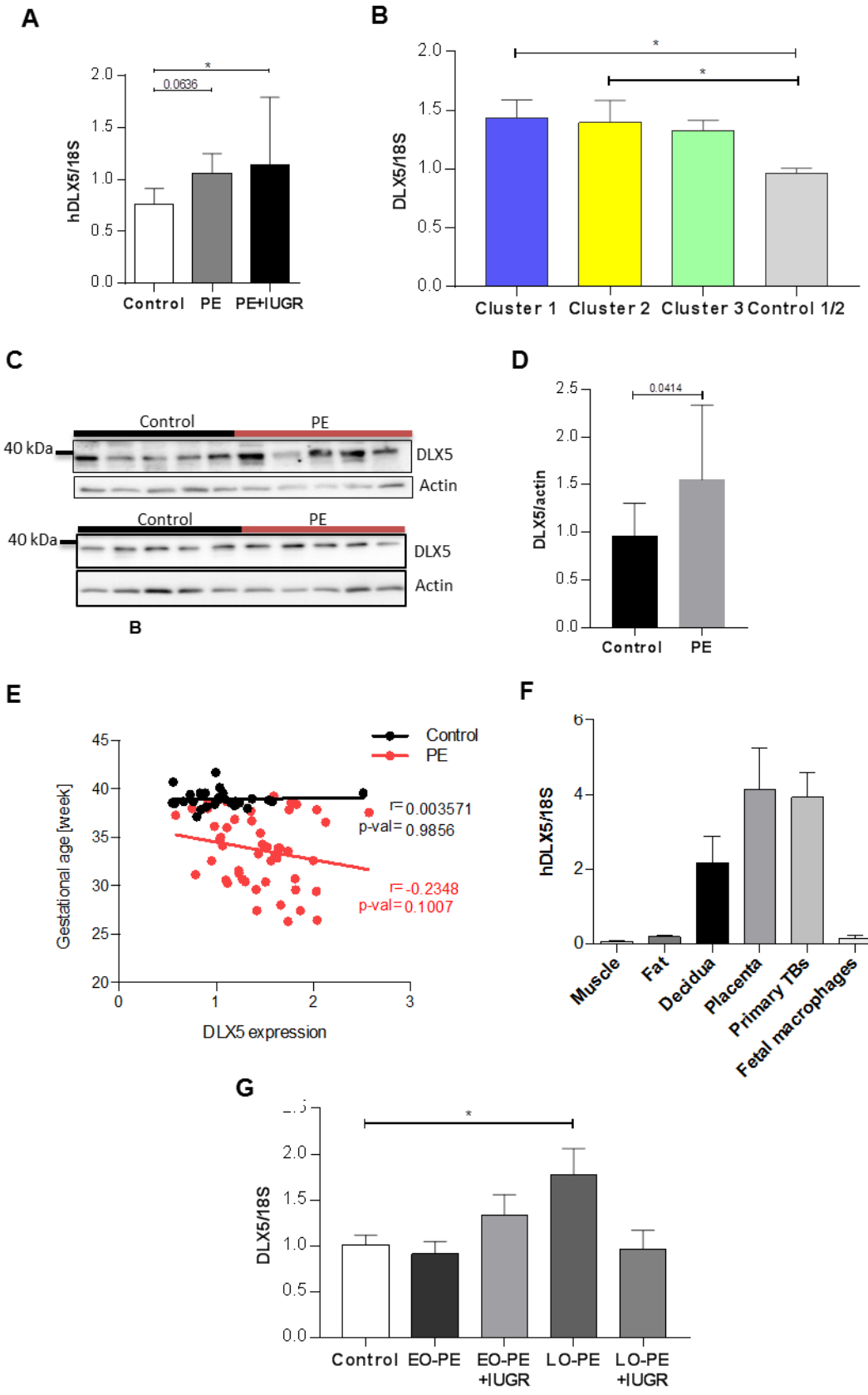

A
hDLX5/18S
2.0
1.5
1.0
0.5
0.0
0.0636
*
Control
PE
PE+IUGR
B
DLX5/18S
2.0
1.5
1.0
0.5
0.0
*
*
Cluster 1
Cluster 2
Cluster 3
Control 1/2
C
Control
PE
40 kDa
DLX5
Actin
Control
PE
40 kDa
DLX5
Actin
B
D
DLX5/actin
2.5
2.0
1.5
1.0
0.5
0.0
0.0414
Control
PE
E
Gestational age [week]
45
40
35
30
25
20
0
1
2
3
DLX5 expression
Control
PE
r=0.003571
p-val=0.9856
r=-0.2348
p-val=0.1007
F
hDLX5/18S
6
4
2
0
Muscle
Fat
Decidua
Placenta
Primary TBs
Fetal macrophages
G
DLX5/18S
2.0
1.5
1.0
0.5
0.0
*
Control
EO-PE
EO-PE +IUGR
LO-PE
LO-PE +IUGR

Fig. S5. Overexpression of human DLX5 protein *in vitro*.

(A) Using *Sleeping Beauty* transposon system human DLX5 gene was stably overexpressed in SGHPL-4 cells. **(B)** DLX5 overexpression was confirmed on mRNA by qPCR (n=6, **p=0.0044, Unpaired t-test) and **(C)** protein level by western blot (n=3). **(D)** Immunofluorescence staining of wild-type cells and OE-DLX5 in SGHPL-4 cells. Predominant nuclear localization of DLX5 is observed.

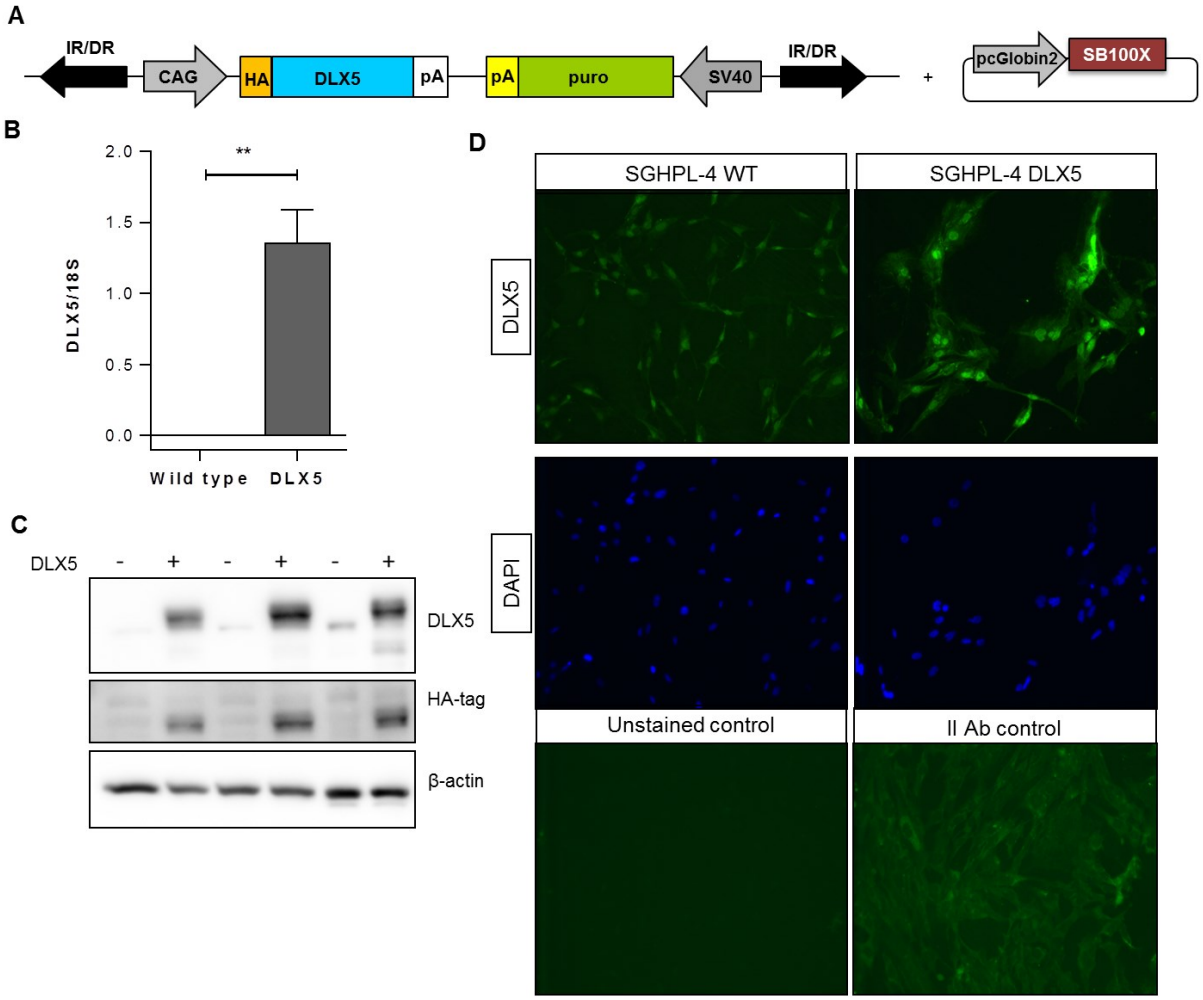
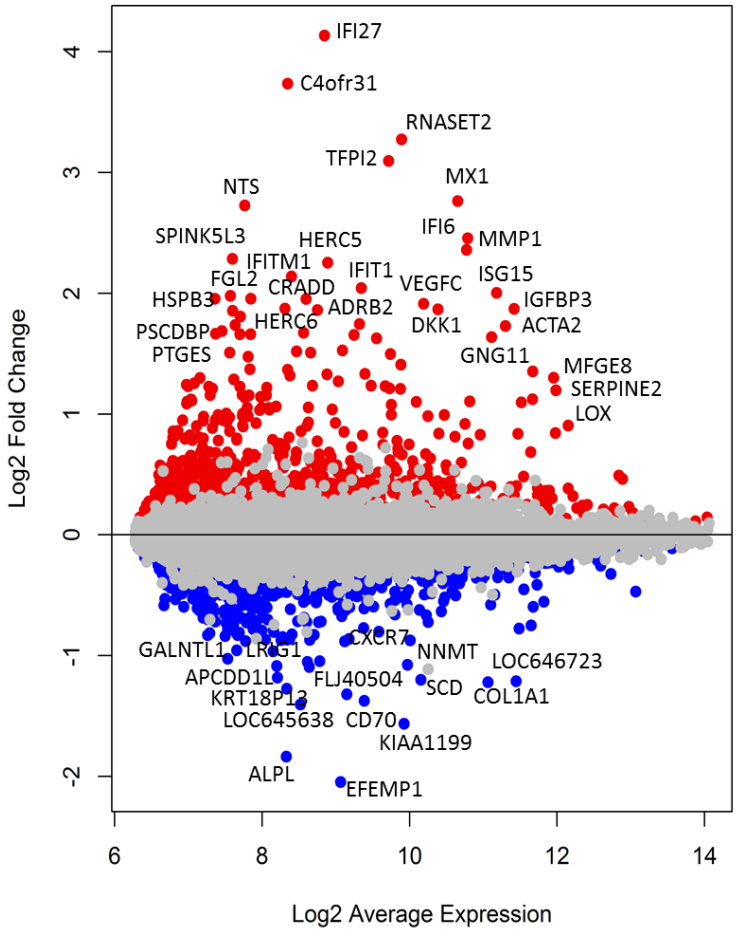


Fig. S6. Transcriptome analysis of SGHPL-4 cells overexpressing human DLX5 protein. **(A)** Fold change of DEGs on the $\log_2$ scale in OE-DLX5 trophoblasts at a significant level of p-value < 0.05; each point represents the $\log_2$ fold change of genes (y-axis) plotted against the $\log_2$ average expression (x-axis); 1897 genes were upregulated (red), 1753 genes were downregulated (blue); n = 6. **(B)** Molecular and cellular functions and physiological activities enriched for genes whose expression is altered in OE-DLX5 cells. Presented are the top five scoring hits in these categories using IPA, which includes the significance scores (p-values) and the number of genes included in each class.

A



B

Name	P-value	Number of molecules
Molecular and cellular functions		
Cellular growth and proliferation	$7.25 \times 10^{-5} - 6.52 \times 10^{-26}$	885
Cell death and survival	$5.39 \times 10^{-5} - 3.28 \times 10^{-19}$	833
Cellular movement	$9.14 \times 10^{-5} - 8.94 \times 10^{-19}$	546
Cellular development	$1.00 \times 10^{-4} - 1.40 \times 10^{-16}$	844
Gene expression	$9.20 \times 10^{-6} - 1.27 \times 10^{-10}$	554
Physiological system development and function		
Organismal survival	$9.48 \times 10^{-6} - 6.72 \times 10^{-17}$	640
Cardiovascular system development and function	$9.30 \times 10^{-5} - 3.14 \times 10^{-13}$	380
Organismal development	$9.75 \times 10^{-5} - 3.14 \times 10^{-13}$	846
Tissue development	$1.00 \times 10^{-4} - 5.25 \times 10^{-12}$	722
Embryonic development	$9.75 \times 10^{-5} - 3.61 \times 10^{-11}$	534

Fig. S7. Intersection of PE transcriptomes with the DLX5^{high} transcriptome. **(A)** DLX5 log_2 relative expression across healthy (blue) and PE (red) placenta samples subjected to the microarray. The placenta samples from patients displaying high (Patients DLX5_high) and low (Patients DLX5_low) DLX5 expression level are indicated. **(B)** Top 6 DLX5-high (Patient_high) and 6 DLX5-low (Patient_low) expressing PE placenta samples were further clustered with WT and OE-DLX5 (DLX5^{high}) SGHLP-4 cells. Heatmap represents Z-score of gene expression across samples (Spearman correlation and distances between observations were calculated using Euclidian distances and average linkage).

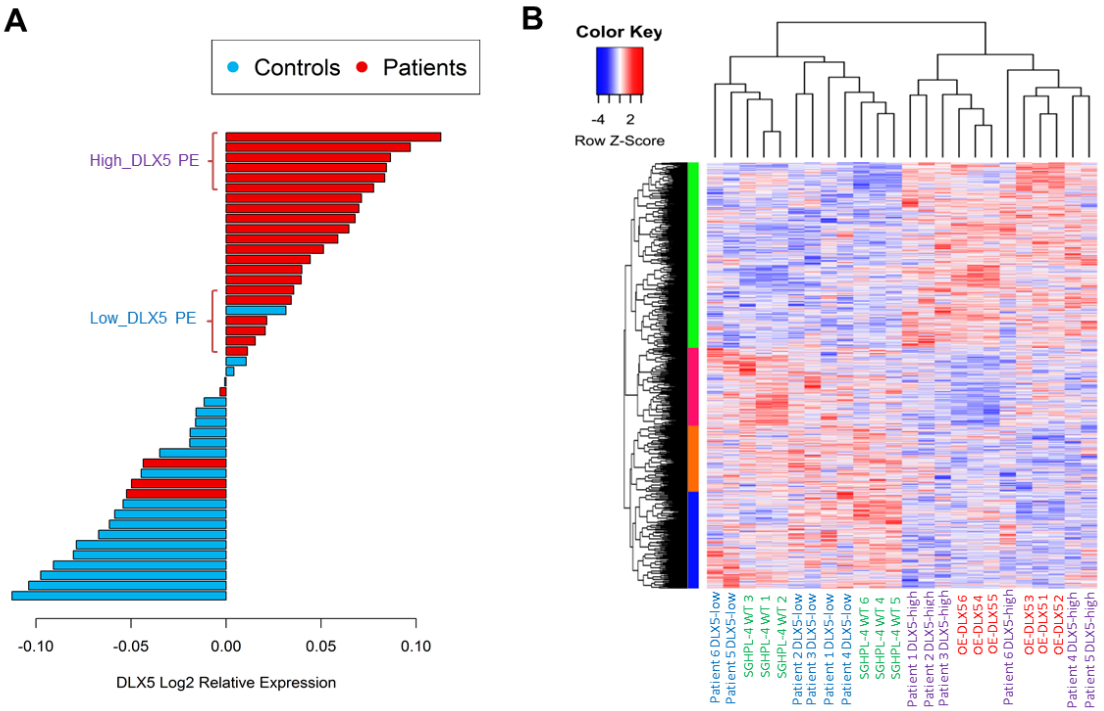


Fig. S8. Effect of DLX5 overexpression on SGHPL-4 cell apoptosis.

(A) DLX5 overexpression slightly increased cell apoptosis as measured by scoring cells undergoing apoptosis over 48h of incubation (not significant, 2-Way ANOVA, Bonferroni's multiple comparisons test). **(B)** After 48h of incubation cell apoptosis is significantly increased after TNFα treatment (conc. 30ng/ml) (DLX5 median 25, IQR: 23.13-31.88 vs. wild type median 12.50, IQR: 6.875-17.50) (*p<0.05; Kruskal-Wallis test, Dunn's multiple comparisons test) but not in untreated cell (DLX5 median 18.75, IQR: 13.75-25 vs. wild-type median 8.75, IQR: 4.375-15.63).

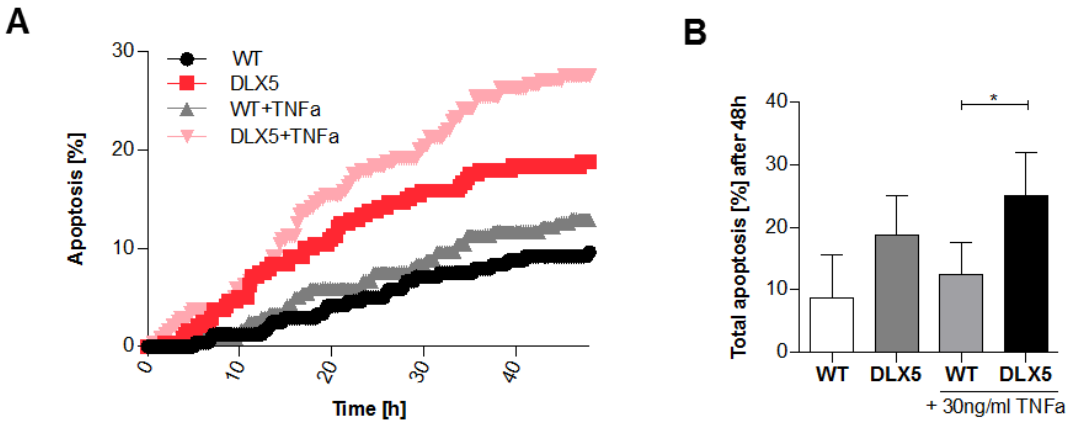


Fig. S9. DLX5 overexpression is associated with ER stress.

(A) DLX5 protein expression level upon induction of ER stress in BeWo cells. 20N – normoxia, 5/20 HR and 1/20 HR - cyclic condition of 6 hour of incubation in 5%/20% O_2 and 1%/20% O_2. **(B)** The transcriptome analysis of OE-DLX5 cells revealed alternation in unfolded protein response pathway (UPR). Shown are upregulated (red) and downregulated (green) UPR genes (at p-value < 0.05). **(C)** Venn diagram representing differentially expressed genes from the UPR pathway in all PE placenta samples and OE-DLX5 cells. 8 dysregulated genes are shared between PE and OE-DLX5 cells. **(D)** Western blot representing expression level of DLX5, E-cadherin, hCG in wild type BeWo cell line treated with 100uM Forskolin to induce syncytialization, either for 24h or 48h. Actin was used as a loading control. DLX5 expression increased upon Forskolin treatment in a time dependent manner.

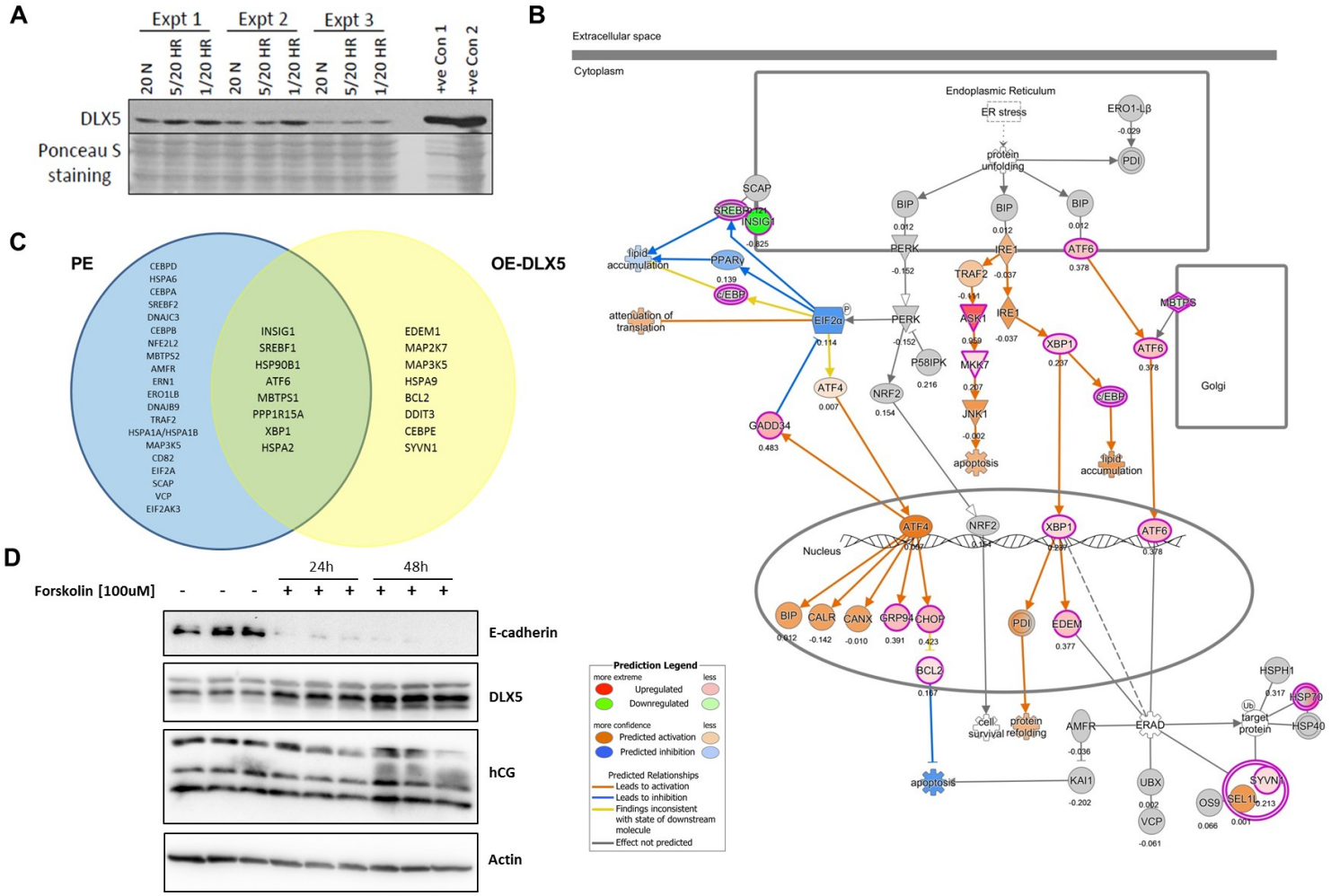
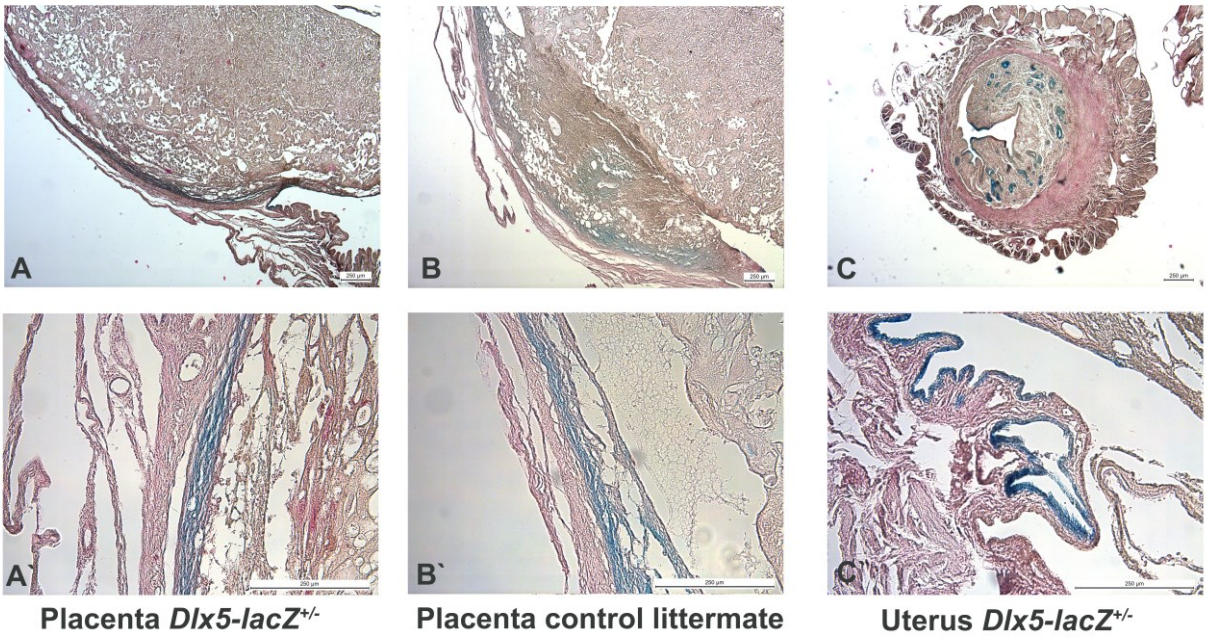


Fig. S10. Dlx5 expression pattern in mouse placenta.

Dlx5-lacZ$^{+/-}$ (A and A`) and control placenta (B and B`) sections were stained for *lacZ* expression. Uterus section of the *Dlx5-lacZ*$^{+/-}$ was used as a positive control. In the placenta weak LacZ staining in the Dlx5-LacZ+/-on the external muscular layer of the placenta, but as similarly, a weak positive staining was also present in controls, we did not consider it. Representative staining are shown.



Supplemental References:

1. Herse F, Lamarca B, Hubel CA, Kaartokallio T, Lokki AI, Ekholm E, Laivuori H, Gauster M, Huppertz B, Sugulle M, Ryan MJ, Novotny S, Brewer J, Park JK, Kacik M, Hoyer J, Verlohren S, Wallukat G, Rothe M, Luft FC, Muller DN, Schunck WH, Staff AC, Dechend R. Cytochrome p450 subfamily 2j polypeptide 2 expression and circulating epoxyeicosatrienoic metabolites in preeclampsia. *Circulation*. 2012;126:2990-2999
2. Sugulle M, Herse F, Seiler M, Dechend R, Staff AC. Cardiovascular risk markers in pregnancies complicated by diabetes mellitus or preeclampsia. *Pregnancy Hypertens*. 2012;2:403-410
3. Braekke K, Holthe MR, Harsem NK, Fagerhol MK, Staff AC. Calprotectin, a marker of inflammation, is elevated in the maternal but not in the fetal circulation in preeclampsia. *American journal of obstetrics and gynecology*. 2005;193:227-233
4. Przybyl L, Haase N, Golic M, Rugor J, Solano ME, Arck PC, Gauster M, Huppertz B, Emontzpohl C, Stoppe C, Bernhagen J, Leng L, Bucala R, Schulz H, Heuser A, Weedon-Fekjaer MS, Johnsen GM, Peetz D, Luft FC, Staff AC, Muller DN, Dechend R, Herse F. Cd74-downregulation of placental macrophage-trophoblastic interactions in preeclampsia. *Circulation research*. 2016;119:55-68
5. Burton GJ, Sebire NJ, Myatt L, Tannetta D, Wang YL, Sadovsky Y, Staff AC, Redman CW. Optimising sample collection for placental research. *Placenta*. 2014;35:9-22
6. Staff AC, Ranheim T, Khoury J, Henriksen T. Increased contents of phospholipids, cholesterol, and lipid peroxides in decidua basalis in women with preeclampsia. *American journal of obstetrics and gynecology*. 1999;180:587-592
7. Du P, Kibbe WA, Lin SM. Lumi: A pipeline for processing illumina microarray. *Bioinformatics (Oxford, England)*. 2008;24:1547-1548
8. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*. 2015;43:e47
9. Suzuki R, Shimodaira H. Pvclust: An r package for assessing the uncertainty in hierarchical clustering. *Bioinformatics (Oxford, England)*. 2006;22:1540-1542
10. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3
11. Hamada H, Okae H, Toh H, Chiba H, Hiura H, Shirane K, Sato T, Suyama M, Yaegashi N, Sasaki H, Arima T. Allele-specific methylome and transcriptome analysis reveals widespread imprinting in the human placenta. *The American Journal of Human Genetics*. 2016;99:1045-1058
12. Lambertini L, Diplas AI, Lee MJ, Sperling R, Chen J, Wetmur J. A sensitive functional assay reveals frequent loss of genomic imprinting in human placenta. *Epigenetics*. 2008;3:261-269
13. Blair JD, Yuen RK, Lim BK, McFadden DE, von Dadelszen P, Robinson WP. Widespread DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. *Molecular human reproduction*. 2013;19:697-708
14. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. *Bioinformatics (Oxford, England)*. 2014;30:1363-1369

15. Petropoulos S, Edsgard D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R, Lanner F. Single-cell rna-seq reveals lineage and x chromosome dynamics in human preimplantation embryos. *Cell*. 2016;165:1012-1026
16. Deng Q, Ramsköld D, Reinius B, Sandberg R. Single-cell rna-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science (New York, N.Y.)*. 2014;343:193-196
17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. Star: Ultrafast universal rna-seq aligner. *Bioinformatics (Oxford, England)*. 2013;29:15-21
18. Liao Y, Smyth GK, Shi W. Featurecounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*. 2014;30:923-930
19. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics (Oxford, England)*. 2007;8:118-127
20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England)*. 2012;28:882-883
21. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*. 2015;33:495-502
22. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161:1202-1214
23. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nature methods*. 2014;11:740-742
24. Blakeley P, Fogarty NME, del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK. Defining the three cell lineages of the human blastocyst by single-cell rna-seq. *Development*. 2015;142:3151-3165
25. Biase FH, Cao X, Zhong S. Cell fate inclination within 2-cell and 4-cell mouse embryos revealed by single-cell rna sequencing. *Genome research*. 2014;24:1787-1796
26. Nakamura T, Okamoto I, Sasaki K, Yabuta Y, Iwatani C, Tsuchiya H, Seita Y, Nakamura S, Yamamoto T, Saitou M. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*. 2016;537:57-62
27. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Li R, Qiao J, Tang F. Single-cell rna-seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology*. 2013;20:1131-1139
28. Andrews S. Fastqc a quality control tool for high throughput sequence data. *Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc*. 2010
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*. 2014;15:550
30. Feng J, Meyer CA, Wang Q, Liu JS, Shirley Liu X, Zhang Y. Gfold: A generalized fold change for ranking differentially expressed genes from rna-seq data. *Bioinformatics (Oxford, England)*. 2012;28:2782-2788
31. Cartwright JE, Holden DP, Whitley GS. Hepatocyte growth factor regulates human trophoblast motility and invasion: A role for nitric oxide. *British journal of pharmacology*. 1999;128:181-189
32. Langfelder P, Horvath S. Wgcna: An r package for weighted correlation network analysis. *BMC bioinformatics*. 2008;9:559

33. Dash PR, Cartwright JE, Whitley GS. Nitric oxide inhibits polyamine-induced apoptosis in the human extravillous trophoblast cell line sghpl-4. *Hum Reprod*. 2003;18:959-968
34. Yung HW, Atkinson D, Campion-Smith T, Olovsson M, Charnock-Jones DS, Burton GJ. Differential activation of placental unfolded protein response pathways implies heterogeneity in causation of early- and late-onset pre-eclampsia. *The Journal of pathology*. 2014;234:262-276