Science Translational Medicine
AAAS

stm.sciencemag.org/cgi/content/full/13/592/eabd9898/DC1

Supplementary Materials for

Integrated trajectories of the maternal metabolome, proteome, and immunome predict labor onset

Ina A. Stelzer, Mohammad S. Ghaemi, Xiaoyuan Han, Kazuo Ando, Julien J. Hédou, Dorien Feyaerts, Laura S. Peterson, Kristen K. Rumer, Eileen S. Tsai, Edward A. Ganio, Dyani K. Gaudillière, Amy S. Tsai, Benjamin Choisy, Lea P. Gaigne, Franck Verdonk, Danielle Jacobsen, Sonia Gavasso, Gavin M. Traber, Mathew Ellenberger, Natalie Stanley, Martin Becker, Anthony Culos, Ramin Fallahzadeh, Ronald J. Wong, Gary L. Darmstadt, Maurice L. Druzin, Virginia D. Winn, Ronald S. Gibbs, Xuefeng B. Ling, Karl Sylvester, Brendan Carvalho, Michael P. Snyder, Gary M. Shaw, David K. Stevenson, Kévin Contrepois, Martin S. Angst, Nima Aghaeepour, Brice Gaudillière*

*Corresponding author. Email: gbrice@stanford.edu

Published 5 May 2021, *Sci. Transl. Med.* **13**, eabd9898 (2021)
DOI: 10.1126/scitranslmed.abd9898

This PDF file includes:

Materials and Methods

Mass Cytometry from whole blood

Sample barcoding and minimization of experimental batch effect

To minimize the effect of experimental variability on mass cytometry measurements between serially collected samples, samples corresponding to the entire time series collected from one woman were processed, barcoded, pooled, stained and run simultaneously. To minimize the effect of variability between study participants, sample sets of two women were run per day and the run was completed within consecutive days, while carefully controlling for consistent tuning parameters of the mass cytometry instrument (Helios CyTOF, Fluidigm Inc., South San Francisco, CA).

Antibody staining and mass cytometry

The mass cytometry antibody panel included 28 antibodies that were used for phenotyping of immune cell subsets and 11 antibodies for the functional characterization of immune cell responses (**table S1**). Antibodies were either obtained preconjugated (Fluidigm, Inc.) or were purchased as purified, carrierfree (no BSA, gelatin) versions, which were then conjugated in-house with trivalent metal isotopes utilizing the MaxPAR antibody conjugation kit (Fluidigm, Inc.). After incubation with Fc block (Biolegend), pooled barcoded cells were stained with surface antibodies, then permeabilized with methanol and stained with intracellular antibodies. All antibodies used in the analysis were titrated and validated on samples that were processed identically to the samples used in the study. Barcoded and antibody-stained cells were analyzed on the mass cytometer.

Identification of immune cell subsets

The mass cytometry data was normalized using Normalizer v0.1 MATLAB Compiler Runtime (MathWorks) (*89*). Files were then de-barcoded with a single-cell MATLAB debarcoding tool (*90*). Manual gating was performed using CellEngine (https://immuneatlas.org/#/) (Primity Bio, Fremont, CA). The following cell types were included in the analysis: Granulocytes, B cells, Natural Killer cells ($CD3^-CD7^+$), $CD56^{bright}CD16^-$NK, $CD56^{dim}CD16^+$NK ($CD69^-$ and $CD69^+$), TCRγδ T cells, $CD4^+$ T cells, $CD4T_{naive}$ ($CD45RA^+CD45RO^-$), $CD62L^+CD4T_{naive}$, $CD4T_{effector\ (eff)}$ ($CD45RA^+CD62L^-$), $CD4T_{memory\ (mem)}$ ($CD45RA^-CD45RO^+$), $CD69^+CD4T_{mem}$, $CD4T_{central\ memory\ (cm)}$ ($CD62L^+CD45RO^+$), $CCR5^+CCR2^+CD4T_{cm}$, $CD4T_{effector\ memory\ (em)}$ ($CD62L^-CD45RO^+$), $CCR5^+CCR2^+CD4T_{em}$, $CD25^+$ $FoxP3^+CD4^+$T cells (T_{reg}), $CD4^+$Tbet$^+$T cells (T_h1), $CD8^+$ T cells, $CD8T_{naive}$ ($CD45RA^+CD45RO^-$), $CD62L^+CD8T_{naive}$, $CD8T_{eff}$ ($CD45RA^+CD62L^-$), $CD8T_{mem}$ ($CD45RA^-CD45RO^+$), $CD69^+CD8T_{mem}$, $CD8T_{cm}$ ($CD62L^+CD45RO^+$), $CCR5^+CCR2^+CD8T_{cm}$, $CD8T_{em}$ ($CD62L^-CD45RO^+$), $CCR5^+CCR2^+CD8T_{em}$, NKT cells ($CD56^+CD3^+$), $CD14^+CD16^-$ classical monocytes (cMCs), $CD14^-CD16^+$ non-classical MCs (ncMCs), $CD14^+CD16^+$ intermediate MCs (intMCs), $CCR2^+$cMC, $CCR2^+$intMC, $CCR2^-$ncMC, $CD14^+CD11b^+$HLA-DR^{lo} myeloid-derived suppressor cells (MDSC), $CD14^-CD16^-$HLA-DR^+ dendritic cells (DC), myeloid DC ($CD11c^+$ mDC), and plasmacytoid dendritic cells ($CD123^+$ pDC).

Proteomics from plasma

Blood was collected into EDTA tubes, kept on ice, and centrifuged (1500 x g, 20 min) at 4℃ within 60 min. Separated plasma was stored at ⁻80°C until further processing. The 200 µL plasma samples were analyzed by the Genome Technology Access Center (St. Louis, MO) using a highly multiplexed, aptamer-based platform capturing 1317 proteins (SomaLogic, Inc., Boulder, CO) (*27, 91*). The assay quantifies proteins over a wide dynamic range (> 8 log) using chemically modified aptamers with slow off-rate kinetics (SOMAmer reagents). Each SOMAmer reagent is a unique, high-affinity, single-strand DNA endowed with functional groups mimicking amino acid side chains. In brief, samples were incubated on 96-well plates with a mixture of SOMAmer reagents. Two sequential bead-based immobilization and washing steps were used to eliminate nonspecifically-bound proteins, unbound proteins, and unbound SOMAmer reagents from protein target-bound reagents. After eluting SOMAmer reagents from the target proteins, the fluorescently-labeled reagents were quantified on an Agilent hybridization array (Agilent Technologies, Santa Clara, CA). Data were normalized in 4 specific steps and according to assay data quality control procedures defined in the good laboratory practice quality system of SomaLogic, Inc. Normalization steps control for signal intensity biases introduced by differential hybridization efficiencies and the overall brightness of plates, collection protocol artifacts, and batch effects between different plates.

Untargeted metabolomics from plasma by liquid chromatography mass spectrometry (LC-MS)

Sample preparation and data acquisition

Plasma samples were thawed on ice, prepared and analyzed randomly as previously described (*27*). Briefly, metabolites were extracted using 1:1:1 acetone:acetonitrile:methanol, evaporated to dryness under nitrogen and reconstituted in 1:1 methanol:water before analysis. Metabolic extracts were analyzed using a broad-spectrum platform comprising two chromatographic systems (HILIC and RPLC) and two ionization modes (positive and negative). Data were acquired on a Q Exactive HF mass spectrometer for HILIC and a Q Exactive mass spectrometer for RPLC (Thermo Scientific, San Jose, CA, USA). Both instruments were equipped with a HESI-II probe and operated in full MS scan mode. MS/MS data were acquired on quality control samples (QC) consisting of an equimolar mixture of all samples in the study. HILIC experiments were performed using a ZIC-HILIC column 2.1 x 100 mm, 3.5 µm, 200Å (cat# 1504470001, Millipore, Burlington, MA, USA) and mobile phase solvents consisting of 10-mM ammonium acetate in 50/50 acetonitrile/water (A) and 10-mM ammonium acetate in 95/5 acetonitrile/water (B). RPLC experiments were performed using a Zorbax SBaq column 2.1 x 50 mm, 1.7 µm, 100Å (cat# 827700-914, Agilent Technologies, Santa Clara, CA) and mobile phase solvents consisting of 0.06% acetic acid in water (A) and 0.06% acetic acid in methanol (B). Data quality was ensured by (i) injecting 6 and 12 pooled samples to equilibrate the LC-MS system prior to run the sequence for RPLC and HILIC, respectively, (ii) injecting a pool sample every 10 injections to control for signal deviation with time, and (iii) checking mass accuracy, retention time and peak shape of internal standards in each sample.

Data processing

Data from each mode were independently processed using Progenesis QI software (v2.3, Nonlinear Dynamics, Durham, NC). Metabolic features from blanks and that did not show sufficient linearity upon dilution in QC samples ($r < 0.6$) were discarded. Only metabolic features present in >2/3 of the samples were kept for further analysis. Inter- and intra-batch variations were corrected using the LOESS (locally estimated scatterplot smoothing Local Regression) normalization method on QC injected repetitively along the batches (span = 0.75). Data were acquired in five and three batches for HILIC and RPLC modes, respectively. Missing

values were imputed by drawing from a random distribution of low values in the corresponding sample. Data from each mode were merged and resulted in a dataset containing 3,529 metabolic features that was used for downstream analysis. Metabolic features of interest were tentatively identified by matching fragmentation spectra and retention time to analytical-grade standards when possible or matching experimental MS/MS to fragmentation spectra in publicly available databases. 12 of the 24 metabolomic most informative model features were successfully annotated with metabolite identifiers derived from public data bases and subsequently visualized. In individual cases, metabolite features were additionally verified by comparing their peaks to commercially available metabolite standards. Three metabolites with elemental composition $C_{21}H_{30}O_3$ (331.2264_8.4, 331.2264_8.1, 331.2265_8.9) were identified as isomers of 17-Hydroxyprogesterone, which correlated with 17-Hydroxyprogesterone, indicating similar biological functions and/or belonging to similar pathways. Similarly, metabolite 361.2017_7.1 ($C_{21}H_{30}O_5$) was highly correlated with the peak of the standard metabolite for cortisol and identified as its isomer.

Statistical analyses

Cross-validation

An underlying assumption of the LASSO algorithm is statistical independence between all observations. In this analysis, although participants are independent, the samples collected on different days throughout the 100 days before the day of labor corresponding to the same subject are not. To address this, a leave-one-subject-out cross-validation (LOOCV) strategy was designed. In this setting, a model is trained on all available samples from all subjects but one. This procedure is repeated for each subject and a model is trained excluding it from the training. The remaining sample is used for testing. The reported results are exclusively based on the blinded subject. For stacked generalization (SG), a two-layer cross-validation strategy was implemented where the inner layer selects the best values of λ. Then, the outer layer tests the models on the blinded subjects. A similar strategy was used for the SG step. Cross-validation folds were carefully synchronized between the individual models from each of the omics. Features whose median across all LOOCV iterations have a non-zero coefficient were reported in the set of most informative features for the prediction task. To assess the relative importance of each feature to the model, features within each individual omic data set were ranked based on the model contribution index, calculated from (-*log10*(p-value)*abs(model coefficient)).

Model validation

Using the results for the cross-validated models on each omic as well as for the SG model combining the omics on the training cohort, we validated the results for the model by predicting the TL of each new sample of the test cohort (N=10 patients, n=27 samples). Metabolomic, proteomic and mass cytometry data for the test cohort was generated using the same procedure as for the training cohort and the features were selected by the individual LASSO to obtain the predictions. The final model coefficients are the medians of the LASSO coefficient across each fold of the cross-validation procedure. Finally, we applied the same preprocessing on the predictions obtained from each omic and computed the final SG model predictions using the same method as described above. In the metabolomic panel, one feature of the training cohort was not detected in the test cohort (331.2264_8.1). In order to compute the predictions, we simply did not include this term in the SG regression equation. Proteomic measurements were unavailable for 6 of the 27 test samples. For these 6 samples, we did not include the proteomic prediction term in the SG regression equation.

Correlation network

All features in each individual omic dataset were visualized using graph structures. Each biological feature was denoted by a node. The graph was visualized using the t-SNE algorithm applied to the complete correlation matrix. For visualization purposes, only the top correlations among features were selected manually and are represented by edges.

Pattern fitting

A classification method was designed to identify function patterns in the features studied. The method was first to separate features with a linear behavior from features with a quadratic behavior in relation to time to labor and then determine if the second derivative of the quadratic fits was positive (acceleration) or negative (deceleration).

The first step of this classification method compared two linear regression fits for each feature X_i: one using the feature X_i and the other using the feature X_i and its square, X_i^2. Both fits were compared using Akaike information criterion (AIC), and the model with the lower AIC value was selected. The AIC values goodness of fit, but penalizes the number of parameters in the models. In this case, if the squared feature, X_i^2, did not sufficiently increase the goodness of fit, the feature was considered linear. Then the feature is classified as accelerating or decelerating based on the coefficients of the model fitted. The fits chosen were associated with p-values computed from the F-statistic. The p-value (< 0.05) were used to determine the relevance of the fit chosen and discard the fits with poor association with either a linear or quadratic model.

Interactome analysis

The interactome was described by the Spearman correlation coefficients between features from different omics. From the correlation matrix of all features, we filtered different thresholds to visualize the connections between the different omics. The intensity of a link between each omic was computed from this filtered correlation matrix: we counted the number of correlations passing the threshold for features from one omic with features from the other omic and we normalized by the total number of possible interactions between both -omics. In order to control for the FDR, we applied a decoy-to-target method generating one random feature using random sampling with replacement from each real feature in our multi-omic dataset. The generated "decoy" dataset of randomized features was then used to estimate the correlations passing different FDR thresholds. All correlations are controlled at FDR < 0.05.

Confounder analysis

Linear regression analysis was used as a statistical model to examine the association between multiple covariates available including the model cross-validated values and the TL outcome variable. This model can be employed as a multiple linear regression to see through confounding and isolate the relationship of interest. Using this method on the training dataset we identified the confounding effects of the covariate variables and asserted the validity of the models and robustness to confounders.

Bootstrap analysis and comparison of ranking

For each omics dataset, we performed a bootstrap analysis where we repeat a random sampling with replacement procedure on the dataset and train a cross-validated model. At each iteration, we keep the non-zero coefficients selected by the LASSO model on the bootstrapped dataset and we repeat the procedure 1000 times. We report the frequency of selection of the features as well as their median coefficient in all the bootstraps. To assess the relative importance of each feature

to the model, we ranked features in each omic dataset based on their frequency of selection. This allowed us to compare the importance of the feature between the complete and term-only model to assess the robustness of the top features to the preterm samples.

Pathway enrichment analysis

Pathway enrichment was performed on the top proteomics and metabolomics features using the Fisher's test (*29*) and hypergeometric test (*30*), respectively. In a first analysis, all 45 selected features from each modality were included in the pathway analysis. To further examine the possibility of multiple correlations of interacting features across omics data contributing to different pathways, the top hits from the multivariate model were visualized using a correlation network. The nodes were divided into two major clusters and were similarly analyzed for pathway enrichment.

A

Time to labor (TL) prediction model	RMSE (days)		
	term	preterm	validation
Original model	17.1	23.7	17.4
Term-only model	15.2	27.3	NA

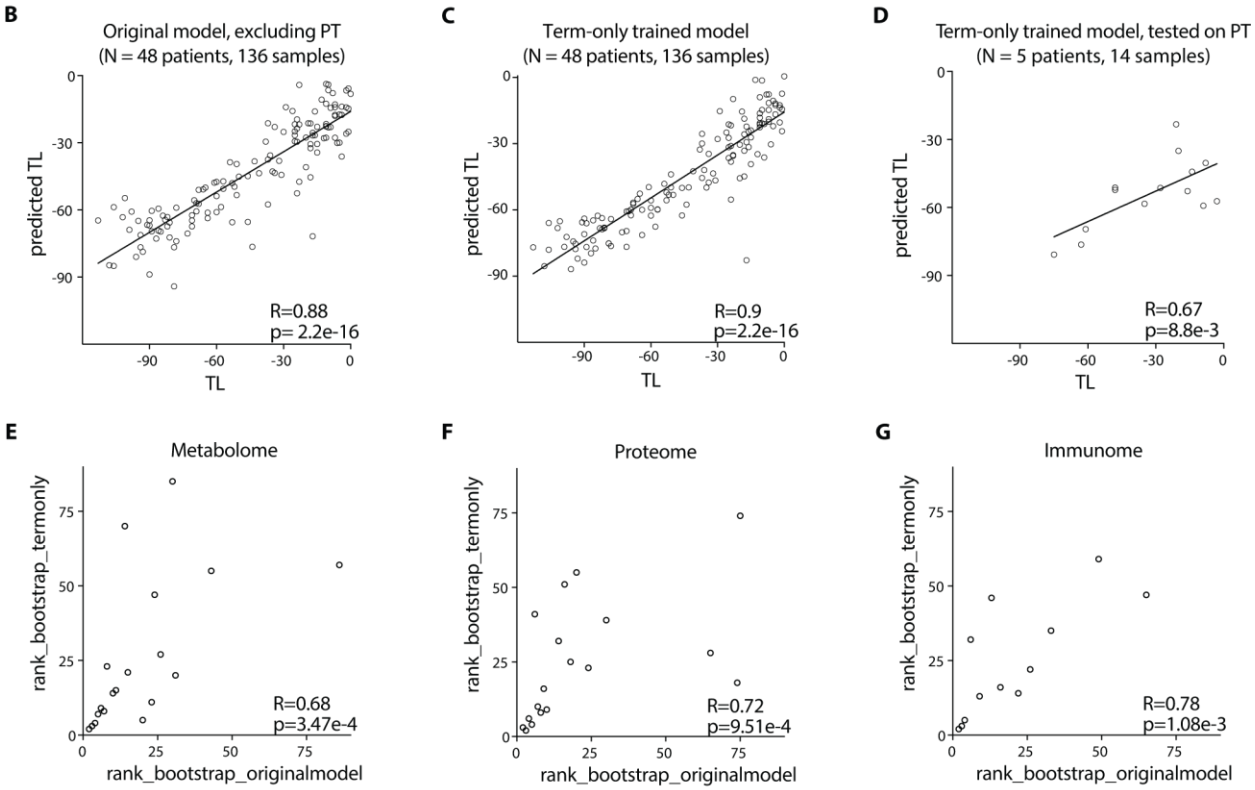

Fig. S1. Analyses of the subcohort of patients with preterm (PT) labor. (**A**) Prediction accuracy (root mean square error (RMSE) in days) of models predicting the TL (original model trained on N = 48 term and N = 5 preterm birth; term-only model trained on N = 48 term birth). (**B**) Regression of predicted vs. true TL for original model plotting term birth data only (N= 48). (**C**) Regression of predicted vs. true TL for model trained in term birth data only (N = 48). **(D**) Regression of predicted vs. true TL for a model trained in term birth data only and tested in cohort of women with preterm birth (N = 5). (**E-G**) Comparison of feature-ranking by bootstrap analyses (original model and term-only model) shows significant correlations between ranks of most informative features, indicating that the original model captured biological parameters independent of studied gestational lengths. Related to Fig. 3.

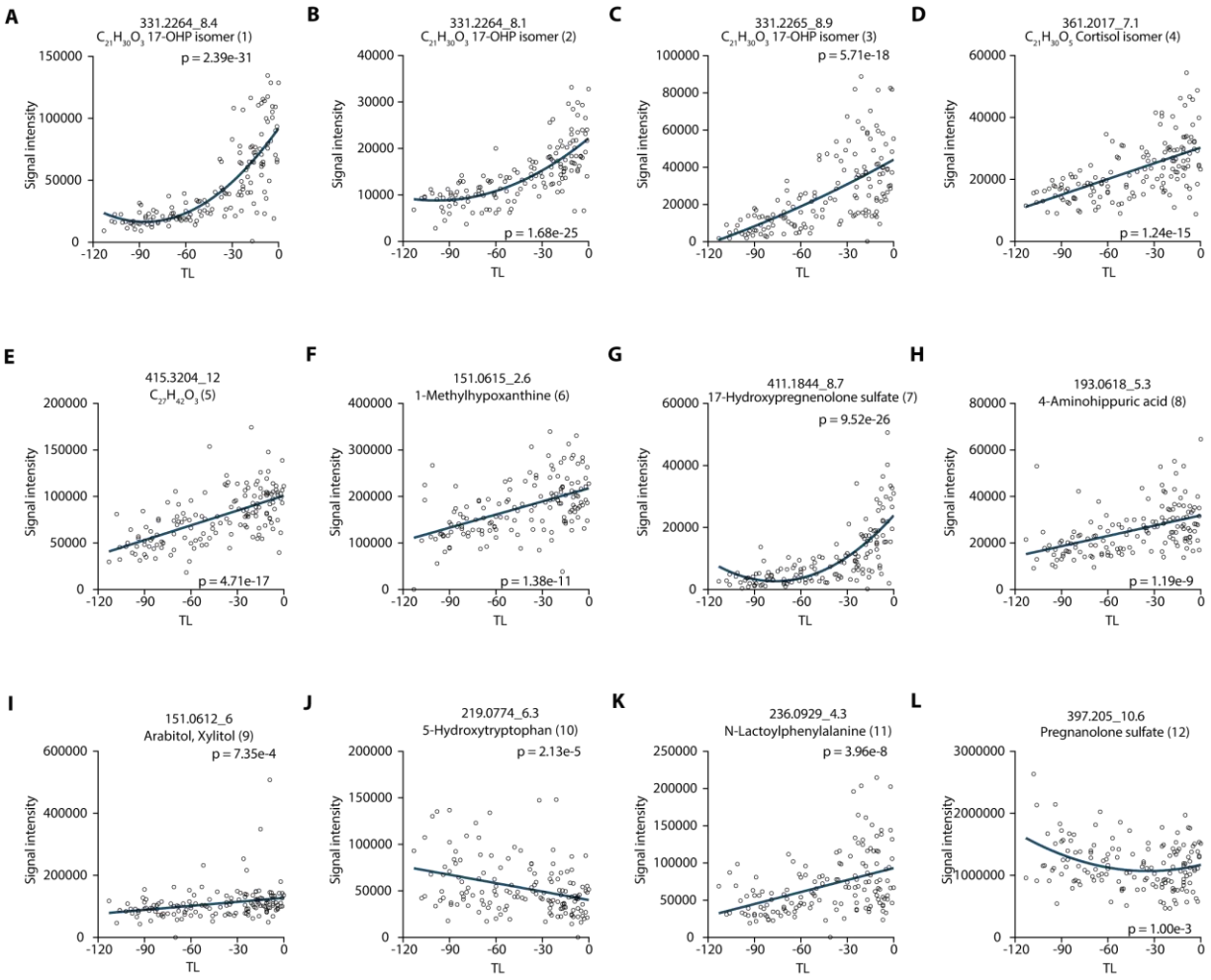

Fig. S2. Metabolic features most informative for the integrated prediction model (*N* = 53 patients, *n* = 150 samples, training cohort). Features are ranked by model index. (**A**) $C_{21}H_{30}O_3$ 17-OHP isomer, (**B**) $C_{21}H_{30}O_3$ 17-OHP isomer, (**C**) $C_{21}H_{30}O_3$ 17-OHP isomer, (**D**) $C_{21}H_{30}O_5$ Cortisol isomer, (**E**) $C_{27}H_{42}O_3$, (**F**) 1-Methylhypoxanthine, (**G**) 17-OH pregnenolone sulfate, (**H**) 4-Aminohippuric acid, (**I**) Arabitol, Xylitol, (**J**) 5-Hydroxytryptophan, (**K**) N-Lactoylphenylalanine, (**L**) Pregnanolone sulfate. Lines represent linear/quadratic curves based on goodness of fit of a pattern fitting model (Akaike information criterion (AIC)); p-value associated with F-statistic for comparison of fits (see table S4). See also table S3. Related to Fig. 4.

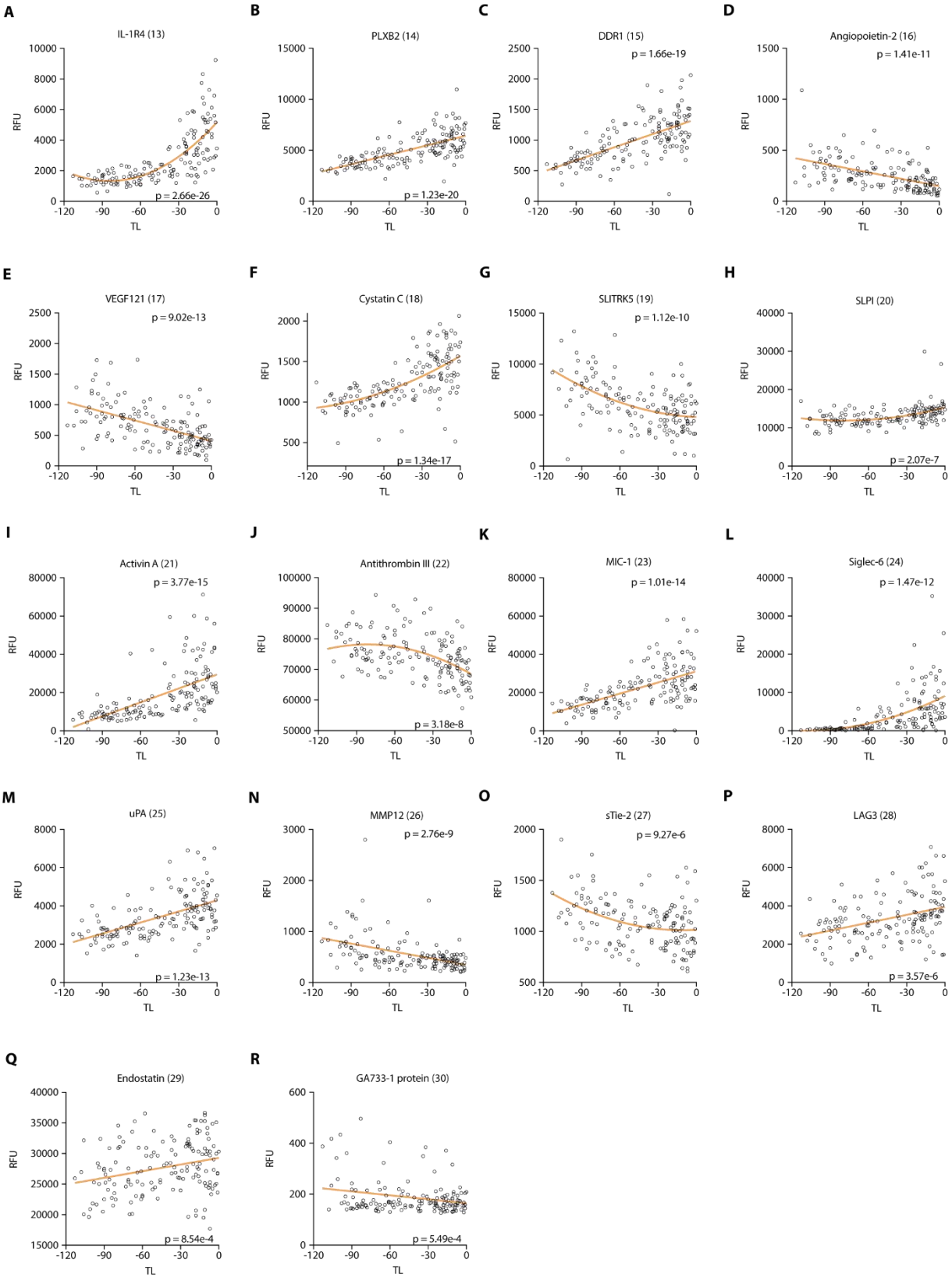

Fig. S3. Proteomic features most informative for the integrated prediction model (*N* = 53 patients, *n* = 150 samples, training cohort). Features are ranked by model index. (**A**) IL-1R4, (**B**) Plexin-B2 (PLXB2), (**C**) Discoidin

domain receptor 1 (DDR1), (**D**) Angiopoietin-2, (**E**) Vascular endothelial growth factor 121 (VEGF121), (**F**) Cystatin C, (**G**) SLIT and NTRK-like protein 5 (SLTRK5), (**H**) Secretory Leukocyte Peptidase Inhibitor (SLPI), (**I**) Activin A, (**J**) Antithrombin III, (**K**) Macrophage inhibitory cytokine-1 (MIC-1), (**L**) Siglec-6, (**M**) urokinase-type plasminogen activator (uPA), (**N**) Matrix metalloproteinase (MMP) 12, (**O**) Soluble tunica interna endothelial cell kinase (sTie)-2, (**P**) LAG3, (**Q**) Endostatin, (**R**) GA733-1 protein. Lines represent linear/quadratic curves based on goodness of fit of a pattern fitting model (Akaike information criterion (AIC)); p-value associated with F-statistic for comparison of fits (see table S4). RFU = Relative Fluorescence Unit. See also table S3. Related to Fig. 4.

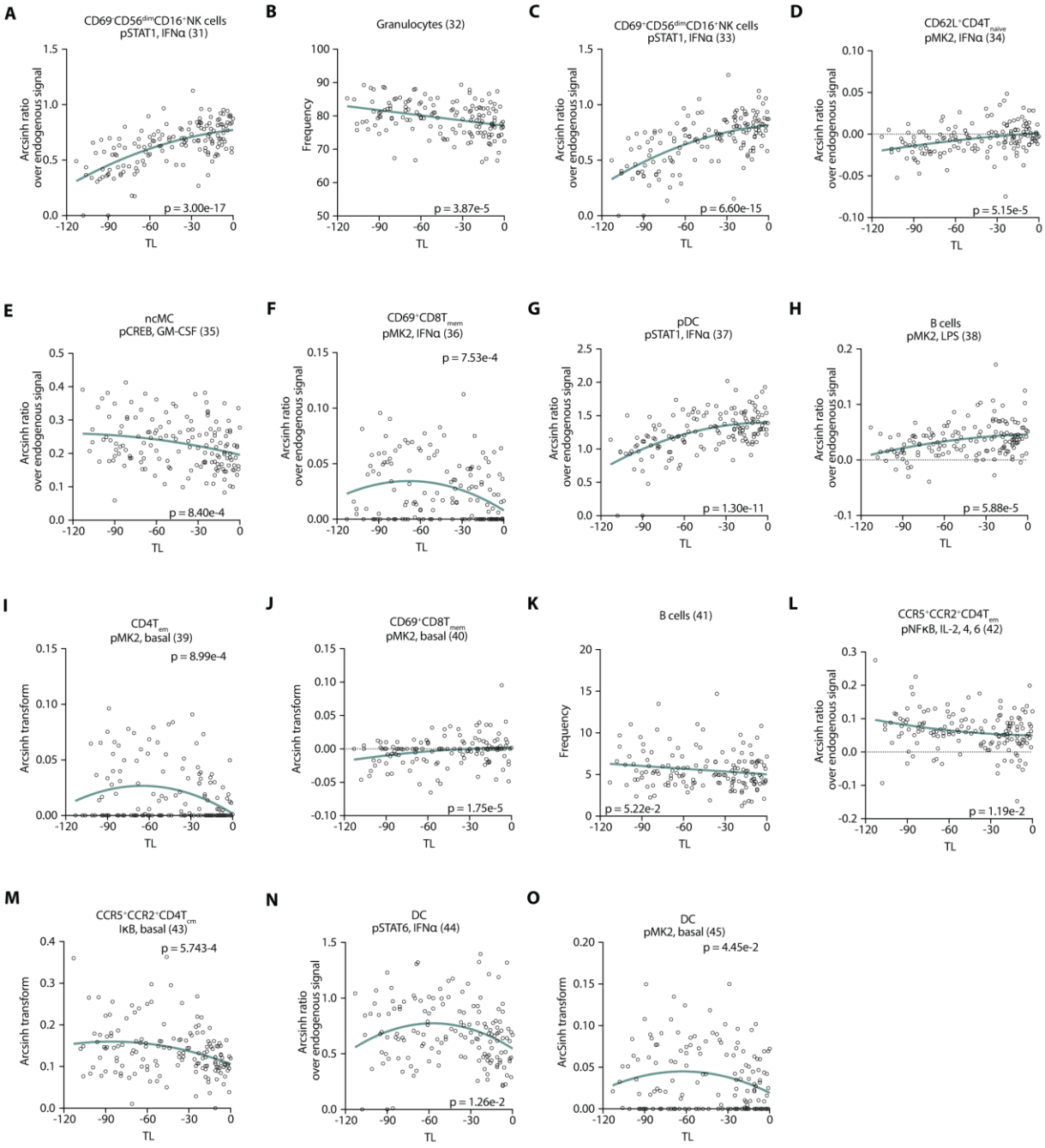

Fig. S4. Immune features most informative for the integrated prediction model (*N* = 53 patients, *n* = 150 samples, training cohort). Features are ranked by model index. (**A**) CD69^{-}CD56^{dim}CD16^{+}NK, pSTAT1, IFNα, (**B**) Granulocytes (freq), (**C**) CD69^{+}CD56^{dim}CD16^{+}NK, pSTAT1, IFNα, (**D**) CD62L^{+}CD4T$_{naive}$, pMK2, IFNα, (**E**) ncMC, pCREB, GM-CSF, (**F**) CD69^{+}CD8T$_{mem}$, pMK2, basal, (**G**) pDC, pSTAT1, IFNα, (**H**) B cells, pMK2, LPS, (**I**) CD4T$_{em}$, pMK2, basal, (**J**) CD69^{+}CD8T$_{mem}$, pMK2, IFNα, (**K**) B cells (freq), (**L**) CCR5^{+}CCR^{+}CD4T$_{em}$, pNFκB, IL-2,4,6, (**M**) CCR^{+}CCR2^{+}CD4T$_{cm}$, IκB, basal, (**N**) DC, pSTAT6, IFNα, (**O**) DC, pMK2, basal. Lines represent linear/quadratic curves based on goodness of fit of a pattern fitting model (Akaike information criterion (AIC)); p-value associated with F-statistic for comparison of fits (see table S4). See also table S3. Related to Fig. 4.

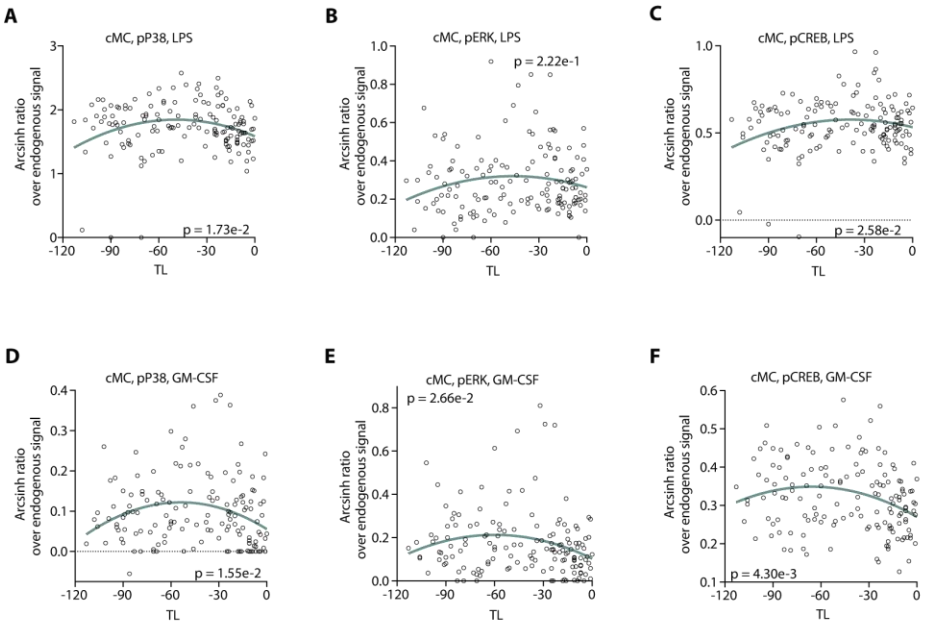

Fig. S5. Innate immune responsiveness decelerates in the prelabor phase (*N* = 53 patients, *n* = 150 samples, training cohort). Classical monocytes (cMCs) in response to lipopolysaccharide (LPS) (**A**–**C**) and GM-CSF (**D**–**F**) show a decrease in MyD88-signaling responses (pP38 (**A**, **D**), pERK1/2 (**B**, **E**), and pCREB (**C**, **F**)) with approaching labor. Lines represent linear/quadratic curves based on goodness of fit of a pattern fitting model (Akaike information criterion (AIC)); p-value associated with F-statistic for comparison of fits. Related to Fig. 4.

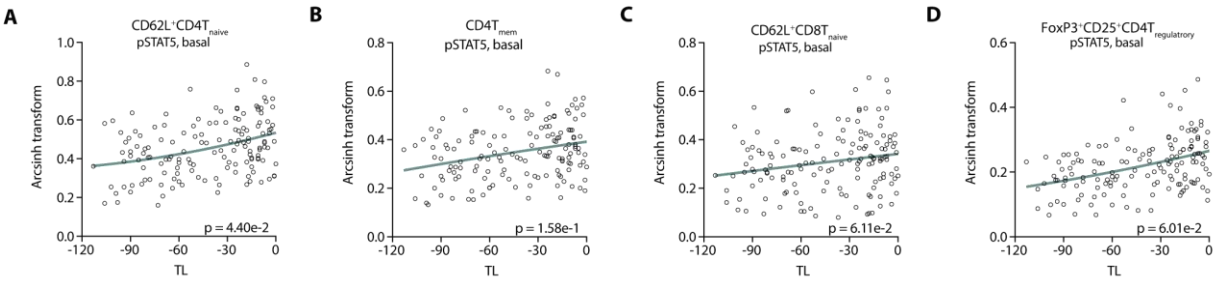

Fig. S6. Basal adaptive immune activity in the prelabor phase (*N* = 53 patients, *n* = 150 samples, training cohort). Phosphorylation of STAT5 (pSTAT5) in naïve (**A**) and memory (**B**) $CD4^+$ T cell subsets, naïve $CD8^+$ T cells (**C**) and $FoxP3^+CD25^+$ regulatory $CD4^+$ T cells (**D**), top informative features for the prediction of gestational age throughout pregnancy (Aghaeepour et al. 2017 (*12*)), increases with approaching labor, but is not informative for the prediction of time to labor (TL). Lines represent linear/quadratic curves based on goodness of fit of a pattern fitting model (Akaike information criterion (AIC)); p-value associated with F-statistic for comparison of fits. Related to Fig. 4.

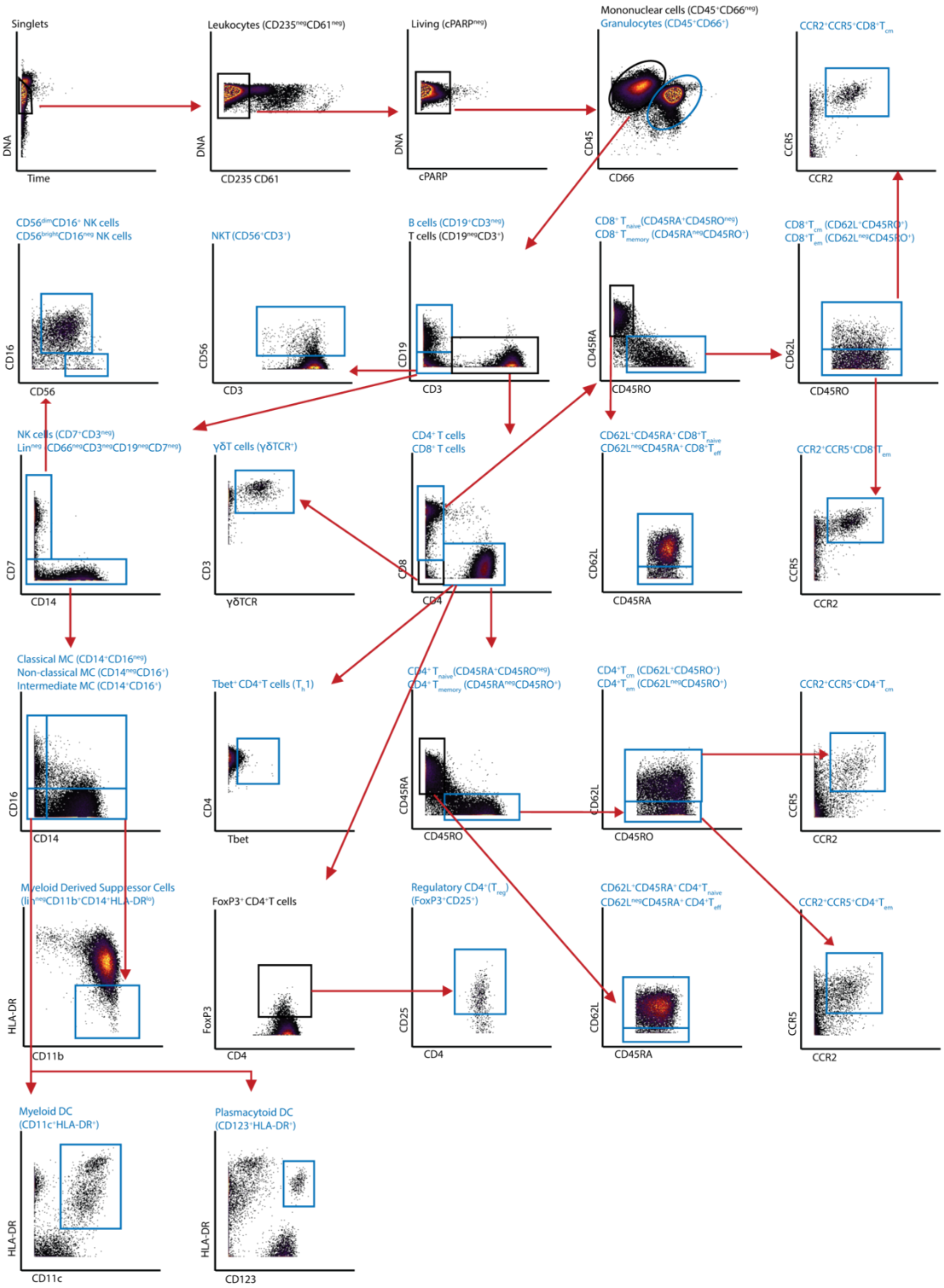

Fig. S7. Gating strategy for mass cytometry analyses. Live, non-erythroid cell populations (blue subpopulations) were used for analysis.

Table S1. Mass cytometry antibody panel.

Antibody	Manufacturer	Atomic Symbol	Atomic Mass	Clone	Comment	Catalogue Number	RRID
Barcode 1	Trace Sciences	Pd	102		Barcode		
Barcode 2	Trace Sciences	Pd	104		Barcode		
Barcode 3	Trace Sciences	Pd	105		Barcode		
Barcode 4	Trace Sciences	Pd	106		Barcode		
Barcode 5	Trace Sciences	Pd	108		Barcode		
Barcode 6	Trace Sciences	Pd	110		Barcode		
CD235	BioLegend	In	113	HIR2	Phenotype	306615	AB_2562825
CD61	BD	In	113	VI-PL2	Phenotype	555752	AB_396093
CD45	Biolegend	In	115	HI30	Phenotype	304045	AB_2562821
CD66	BD	La	139	B1.1/CD66	Phenotype	551354	AB_394166
CD7	BD	Pr	141	M-T701	Phenotype	555359	AB_395762
CD19	BioLegend	Nd	142	HIB19	Phenotype	302247	AB_2562815
CD45RA	BioLegend	Nd	143	HI100	Phenotype	304143	AB_2562822
CD11b	BioLegend	Nd	144	ICRF44	Phenotype	301337	AB_2562811
CD4	BioLegend	Nd	145	RPA-T4	Phenotype	300541	AB_2562809
CD8a	BioLegend	Nd	146	RPA-T8	Phenotype	301053	AB_2562810
CD11c	BioLegend	Sm	147	Bu15	Phenotype	337221	AB_2562834
CD123	BioLegend	Nd	148	6H6	Phenotype	306027	AB_2562823
pCREB (pS133)	CST	Sm	149	87G3	Function	9198	AB_2561044
pSTAT5 (Tyr694)	CST	Nd	150	C11C5	Function	9359	AB_823649
pP38 (pT180/pY182)	BD	Eu	151	36/p38	Function	612289	AB_399606
TCRγδ	BD	Sm	152	B1	Phenotype	555715	AB_396059
pSTAT1 (pY701)	BD	Eu	153	14/P-STAT1	Function	612133	AB_399504
pSTAT3 (Tyr705)	CST	Sm	154	M9C6	Function	4113	AB_2198588
pS6 (pS235/pS236)	CST	Gd	155	D57.2.2E	Function	4858	AB_916156
IκB	CST	Gd	156	L35A5	Function	4814S	
CD69	BD	Gd	157	FN50	Phenotype	555529	AB_395914
CD33	BioLegend	Gd	158	WM53	Phenotype	303419	AB_2562818
pMPK2 (Thr334)	CST	Tb	159	27B7	Function	3007	AB_490936
Tbet	Thermo Fisher	Gd	160	4B10	Phenotype	14582580	AB_763635
cPARP	BD	Dy	161	F21-852	Function	552596	AB_394437
FoxP3	Thermo Fisher	Dy	162	PCH101	Phenotype	14477682	AB_467554
CD45RO	Fluidigm	Dy	164	UCHL1	Phenotype	3164007B	AB_2811092
CD16	BioLegend	Ho	165	3G8	Phenotype	302051	AB_2562814
pNFκB (pS529)	BD	Er	166	k108951250	Function	558393	AB_647284
pERK1/2 (pT202/pY204)	CST	Yb	167	D13.14.4E	Function	4370	AB_2315112
pSTAT6 (Tyr641)	BioLegend	Er	168	A15137E	Function	686002	AB_2616820
CD25	BioLegend	Tm	169	M-A251	Phenotype	356102	AB_2561752
CD3	BioLegend	Er	170	UCHT1	Phenotype	300402	AB_314056
CCR5	Fluidigm	Yb	171	NP6G4	Phenotype	3171017a	
CD62L	Thermo Scientific	Yb	172	DREG.200	Phenotype	BMS1015	AB_10596353
CCR2	BioLegend	Yb	173	K036C2	Phenotype	357202	AB_2561851
HLA-DR	BioLegend	Yb	174	L243	Phenotype	307651	AB_2562826
CD14	BioLegend	Yb	175	M5E2	Phenotype	301843	AB_2562813
CD56	BD	Yb	176	NCAM16.2	Phenotype	559043	AB_397180
DNA1/2	Fluidigm	Ir	191/192		DNA		

Table S2. Confounder analysis. None of the potentially confounding variables significantly influences the prediction of the TL in the original training model (see Methods). Related to Fig. 3. *"Betamethasone treatment" is co-linear with variable "Preterm delivery (<37wks)"

Variable	Estimate	Std.Error	t	value	Pr(>|t|)
Prediction	1.308304	0.059853	21.859	<2e-16	***
Age (years)	0.44338	0.510009	0.869	0.3867	
BMI 1st trimester (kg/m2)	2.056674	1.194251	1.722	0.0881	
BMI 3rd trimester (kg/m2)	-0.406995	1.004374	-0.405	0.6862	
GA at delivery, all (wks)	-2.477985	2.186258	-1.133	0.2597	
Preterm delivery (<37 wks)*	-0.056306	11.72105	-0.005	0.9962	
Gravidity	-2.657588	1.82363	-1.457	0.1481	
Parity (% nulliparous)	0.762594	3.100135	0.246	0.8062	
Infant sex	-4.016345	3.494093	-1.149	0.2531	
Birthweight (g)	-0.00641	0.004302	-1.49	0.1393	
Race East Asian	88.571079	89.228857	0.993	0.3232	
Race South Asian	81.119883	91.381792	0.888	0.3768	
Race Southeast Asian	87.743665	89.408413	0.981	0.3287	
Race Native Hawaiian/Pacific Islander	76.950741	92.525918	0.832	0.4075	
Race Middle Eastern	68.949986	90.949048	0.758	0.4501	
Race White	85.241423	90.572585	0.941	0.3489	
Race Two or More Races	130.86968	87.843884	1.49	0.1394	
Race Other	87.38296	92.766224	0.942	0.3484	
Gestational diabetes	3.709263	6.795246	0.546	0.5864	
Gestational hypertension	-7.140725	5.651122	-1.264	0.2093	
History of preterm birth	2.650519	6.589623	0.402	0.6884	
Preeclampsia	-0.09008	7.124061	-0.013	0.9899	
Progesterone treatment	-6.987755	9.992922	-0.699	0.486	

Table S3. Forty-five most informative features of the integrated multiomic labor prediction model in the training (blue) and test (gray) cohort. Features in the training cohort were ranked based on the model index (calculated from: -*log10*(pval)*abs(model coef)). Spearman correlation coefficients and associated *p*-values were calculated for the association of the individual features with TL in both cohorts. Related to Fig. 3, 4.

No.	Modality	Feature name	Model index	Training Spearman Correlation Coefficient (R)	Training *p*-value (N = 53 patients)	Test Spearman Correlation Coefficient (R)	Test *p*-value (N = 10 patients)
1	metabolome	331.2264_8.4 ($C_{21}H_{30}O_3$17-OHP isomer)	200.12	0.82	3.94E-37	0.79	9.19E-07
2	metabolome	331.2264_8.1 ($C_{21}H_{30}O_3$17-OHP isomer)	161.21	0.76	1.47E-29	Not detected	Not applicable
3	metabolome	331.2265_8.9 ($C_{21}H_{30}O_3$17-OHP isomer)	47.61	0.69	5.98E-23	0.53	4.74E-03
4	metabolome	361.2017_7.1 ($C_{21}H_{30}O_5$ Cortisol isomer)	27.84	0.62	5.29E-17	0.39	4.35E-02
5	metabolome	415.3204_12 ($C_{27}H_{42}O_3$)	16.43	0.63	8.63E-18	0.41	2.05E-03
6	metabolome	151.0615_2.6 (1-Methylhypoxanthine)	8.71	0.53	3.95E-12	0.45	3.42E-02
7	metabolome	411.1844_8.7 (17-OH pregnenolone sulfate)	7.73	0.75	2.10E-28	0.57	1.79E-02
8	metabolome	193.0618_5.3 (4-Aminohippuric acid)	5.11	0.52	1.40E-11	0.21	3.02E-01
9	metabolome	151.0612_6 (Arabitol, Xylitol)	3.01	0.34	2.69E-05	0.62	2.29E-01
10	metabolome	219.0774_6.3 (5-Hydroxytryptophan)	2.42	-0.33	3.57E-05	-0.24	1.67E-01
11	metabolome	236.0929_4.3 (N-Lactoylphenylalanine)	2.17	0.43	4.96E-08	0.27	9.52E-01
12	metabolome	397.205_10.6 (Pregnanolone sulfate)	0.24	-0.15	7.39E-02	-0.012	5.89E-04
13	proteome	IL-1R4	191.68	0.78	7.98E-32	0.83	3.12E-03
14	proteome	Plexin-B2 (PLXB2)	89.89	0.71	3.68E-24	0.12	5.37E-01
15	proteome	Discoidin domain receptor 1 (DDR1)	68.74	0.68	2.84E-21	0.44	4.87E-01
16	proteome	Angiopoietin-2	52.63	-0.58	1.16E-14	-0.61	6.13E-02
17	proteome	Vascular Endothelial Growth Factor 121	45.08	-0.56	1.39E-13	-0.55	1.43E-01
18	proteome	Cystatin C	38.77	0.67	3.98E-21	0.26	5.55E-01
19	proteome	SLIT and NTRK-like protein 5 (SLTRK5)	31.36	-0.49	2.83E-10	-0.26	5.37E-01
20	proteome	Secr. Leukocyte Peptidase Inhibitor (SLPI)	24.90	0.47	1.73E-09	0.31	2.27E-01
21	proteome	Activin A	19.23	0.71	2.89E-24	0.34	9.15E-01
22	proteome	Antithrombin III	10.27	-0.45	6.11E-09	-0.46	7.33E-02
23	proteome	Macrophage inhibitory cytokine-1 (MIC-1)	9.96	0.61	1.79E-16	0.62	1.15E-02
24	proteome	Siglec-6	8.60	0.73	4.43E-26	0.85	4.49E-03
25	proteome	urokinase-type Plasminogen Activator (uPA)	8.09	0.59	2.44E-15	0.39	2.42E-01

Table S3 cont'd

No.	Modality	Feature name	Model index	Training Spearman Correlation Coefficient (R)	Training *p*-value (N = 53 patients)	Test Spearman Correlation Coefficient (R)	Test *p*-value (N = 10 patients)
26	proteome	Matrix metalloproteinase (MMP) 12	5.87	-0.51	2.67E-11	-0.69	9.25E-03
27	proteome	Soluble tunica interna endothelial cell kinase (sTie)-2	4.91	-0.36	8.33E-06	-0.49	6.80E-02
28	proteome	LAG3	3.24	0.39	6.59E-07	0.09	4.68E-01
29	proteome	Endostatin	1.73	0.24	3.22E-03	0.05	8.43E-01
30	proteome	GA733-1 protein	0.27	-0.12	1.52E-01	0.19	9.38E-01
31	immunome	$CD69^{-}CD56^{dim}CD16^{+}$NK, pSTAT1, IFNα	161.23	0.61	1.31E-16	0.78	1.56E-06
32	immunome	Granulocytes (freq)	18.20	-0.34	1.74E-05	-0.31	1.12E-01
33	immunome	$CD69^{+}CD56^{dim}CD16^{+}$NK, pSTAT1, IFNα	10.18	0.58	5.96E-15	0.61	7.07E-04
34	immunome	$CD62L^{+}CD4T_{naive}$, pMK2, IFNα	8.92	0.35	1.39E-05	0.42	3.01E-02
35	immunome	ncMC, pCREB, GM-CSF	8.11	-0.28	5.98E-04	-0.14	4.94E-01
36	immunome	$CD69^{+}CD8T_{mem}$, pMK2, basal	5.90	-0.27	1.90E-02	-0.32	8.77E-02
37	immunome	pDC, pSTAT1, IFNα	5.05	0.51	3.95E-11	0.65	2.78E-04
38	immunome	B cells, pMK2, LPS	2.96	0.31	1.38E-04	-0.01	9.54E-01
39	immunome	$CD4T_{em}$, pMK2, basal	1.69	-0.19	1.85E-02	-0.23	2.44E-01
40	immunome	$CD69^{+}CD8T_{mem}$, pMK2, IFNα	1.28	0.19	7.57E-04	0.34	1.09E-01
41	immunome	B cells (freq)	1.11	-0.14	9.30E-02	-0.15	4.62E-01
42	immunome	$CCR5^{+}CCR2^{+}CD4T_{em}$, pNFκB, IL-2, Il-4, IL-6	1.10	-0.19	2.26E-02	0.12	5.68E-01
43	immunome	$CCR5^{+}CCR2^{+}CD4T_{cm}$, IκB, basal	0.97	-0.28	5.87E-04	-0.42	2.75E-02
44	immunome	DC, pSTAT6, IFNα	0.80	-0.18	2.43E-02	-0.20	3.08E-01
45	immunome	DC, pMK2, basal	0.36	-0.13	1.14E-01	-0.24	2.25E-01

Table S4. Goodness of fit of a pattern-fitting model [Akaike information criterion (AIC)] for the 45 most informative features of the integrated multiomic labor prediction model in the training cohort. The fits chosen were associated with p-values computed from the F-statistic to determine their relevance. Model index calculated from: *-log10*(pval)*abs(model coef). See also table S3. Related to Fig. 4.

No.	Modality	Feature name	Model index	AIC degree 1	AIC degree 2	p-value	Class	Pattern
1	metabolome	331.2264_8.4 ($C_{21}H_{30}O_3$17-OHP isomer)	200.12	3425.85	3402.78	2.39e-31	quadratic	acceleration/ increase
2	metabolome	331.2264_8.1 ($C_{21}H_{30}O_3$17-OHP isomer)	161.21	2940.09	2930.63	1.68E-25	quadratic	acceleration/ increase
3	metabolome	331.2265_8.9 ($C_{21}H_{30}O_3$17-OHP isomer)	47.61	3326.09	3327.88	5.71E-18	linear	constant
4	metabolome	361.2017_7.1 ($C_{21}H_{30}O_5$ Cortisol isomer)	27.84	3103.48	3103.91	1.24E-15	linear	constant
5	metabolome	415.3204_12 ($C_{27}H_{42}O_3$)	16.43	3428.04	3429.92	4.71E-17	linear	constant
6	metabolome	151.0615_2.6 (1-Methylhypoxanthine)	8.71	3678.81	3680.80	1.38E-11	linear	constant
7	metabolome	411.1844_8.7 (17-OH pregnenolone sulfate)	7.73	3088.91	3059.97	9.52E-26	quadratic	acceleration/ increase
8	metabolome	193.0618_5.3 (4-Aminohippuric acid)	5.11	3158.65	3160.09	1.19E-09	linear	constant
9	metabolome	151.0612_6 (Arabitol, Xylitol)	3.01	3669.95	3671.74	7.35E-04	linear	constant
10	metabolome	219.0774_6.3 (5-Hydroxytryptophan)	2.42	3490.65	3492.62	2.13E-05	linear	constant
11	metabolome	236.0929_4.3 (N-Lactoylphenylalanine)	2.17	3580.68	3582.19	3.96E-08	linear	constant
12	metabolome	397.205_10.6 (Pregnanolone sulfate)	0.24	4275.53	4271.66	0.001	quadratic	deceleration/ decrease
13	proteome	IL-1R4	191.68	2567.88	2548.48	2.66E-26	quadratic	acceleration/ increase
14	proteome	Plexin-B2 (PLXB2)	89.89	2537.32	2539.01	1.23E-20	linear	constant
15	proteome	Discoidin domain receptor 1 (DDR1)	68.74	2108.89	2110.75	1.66E-19	linear	constant
16	proteome	Angiopoietin-2	52.63	1890.58	1892.08	1.41E-11	linear	constant
17	proteome	Vascular Endothelial Growth Factor 121	45.08	2121.34	2122.95	9.02E-13	linear	constant
18	proteome	Cystatin C	38.77	2074.798	2074.796	1.34E-17	quadratic	acceleration/ increase
19	proteome	SLIT and NTRK-like protein 5 (SLTRK5)	31.36	2713.88	2713.27	1.12E-10	quadratic	deceleration/ decrease
20	proteome	Secr. Leukocyte Peptidase Inhibitor (SLPI)	24.90	2765.13	2761.58	2.07E-07	quadratic	acceleration/ increase
21	proteome	Activin A	19.23	3218.71	3218.79	3.77E-15	linear	constant
22	proteome	Antithrombin III	10.27	3059.82	3056.07	3.18E-08	quadratic	acceleration/ decrease
23	proteome	Macrophage inhibitory cytokine-1 (MIC-1)	9.96	3157.13	3158.62	1.01E-14	linear	constant
24	proteome	Siglec-6	8.60	2952.64	2951.56	1.47E-12	quadratic	acceleration/ increase
25	proteome	urokinase-type Plasminogen Activator (uPA)	8.09	2480.50	2482.13	1.23E-13	linear	constant

Table S4 cont'd

No.	Modality	Feature name	Model index	AIC degree 1	AIC degree 2	p-value	Class	Pattern
26	proteome	Matrix metalloproteinase (MMP) 12	5.87	2128.60	2129.72	2.76E-09	linear	constant
27	proteome	Soluble tunica interna endothelial cell kinase (sTie)-2	4.91	2052.00	2051.45	9.27E-06	quadratic	deceleration/ decrease
28	proteome	LAG3	3.24	2536.34	2537.59	3.57E-06	linear	constant
29	proteome	Endostatin	1.73	2932.02	2932.58	8.54E-04	linear	constant
30	proteome	GA733-1 protein	0.27	1678.15	1674.65	5.49E-04	quadratic	deceleration/ decrease
31	immunome	$CD69^{-}CD56^{dim}CD16^{+}$NK, pSTAT1, IFNα	161.23	136.88	-137.58	3.00E-17	quadratic	deceleration/ increase
32	immunome	Granulocytes (freq)	18.20	901.11	902.06	3.87E-05	linear	constant
33	immunome	$CD69^{+}CD56^{dim}CD16^{+}$NK, pSTAT1, IFNα	10.18	93.57	-94.15	6.60E-15	quadratic	deceleration/ increase
34	immunome	$CD62L^{+}CD4T_{naive}$, pMK2, IFNα	8.92	796.26	-794.41	5.15E-05	quadratic	deceleration/ increase
35	immunome	ncMC, pCREB, GM-CSF	8.11	364.62	-362.96	8.40E-04	quadratic	acceleration/ decrease
36	immunome	$CD69^{+}CD8T_{mem}$, pMK2, basal	5.90	662.37	-665.48	7.53E-04	quadratic	acceleration/ decrease
37	immunome	pDC, pSTAT1, IFNα	5.05	26.62	24.68	1.30E-11	quadratic	deceleration/ increase
38	immunome	B cells, pPK2, LPS	2.96	638.73	-637.35	5.88E-05	quadratic	deceleration/ increase
39	immunome	$CD4T_{em}$, pMK2, basal	1.69	680.21	-684.44	8.99E-04	quadratic	acceleration/ decrease
40	immunome	$CD69^{+}CD8T_{mem}$, pMK2, IFNα	1.28	810.97	-809.10	1.75E-05	quadratic	deceleration/ increase
41	immunome	B cells (freq)	1.11	670.77	672.74	5.22E-02	linear	constant
42	immunome	$CCR5^{+}CCR2^{+}CD4T_{em}$, pNFκB, IL-2,4,6	1.10	427.82	-426.13	1.19E-02	quadratic	deceleration/ decrease
43	immunome	$CCR5^{+}CCR2^{+}CD4T_{cm}$, IκB, basal	0.97	418.89	-418.56	5.74E-04	quadratic	acceleration/ decrease
44	immunome	DC, pSTAT6, IFNα	0.80	45.62	40.80	1.26E-02	quadratic	acceleration/ decrease
45	immunome	DC, pMK2, basal	0.36	554.01	-555.32	4.45E-02	quadratic	acceleration/ decrease