



Gestational Age prediction using CpG clusters in regularized regression

ANTS: Alena van Bömmel, Tushar Patel, Steve Hoffmann

40th Winter Seminar, Bled – 11th February, 2025







Reverse Engineering Assessment and Methods (DREAM) - are an open science, collaborative challenge framework - to spur innovative solutions.

Examples from some of the previous challenges include:

- Prediction of drug sensitivity in cell lines using genomic data.
- Prediction of patient outcomes in metastatic prostate cancer using clinical trial data.



Develop more capable and accurate placental clocks to predict gestational age based on DNA methylation measured by the Illumina methylation arrays.













Selecting top gestational aging correlated CpGs

















PCA on selected-clustered features









An elastic-net regression model for GA prediction



Actual GA (weeks)

An elastic-net regression model for GA prediction



Actual GA (weeks)

Important CpGs across models and their biology





SC1_v1.0 : On all 450k data

SC1_v1.1 : On PCA outlier removed 450k data

Wang, S.S et al. 2023

Enrichment of common CpGs associated genes





Based on UCSC_ref_gene_name associated with CpGs from EPIC annotation.

Model comparison for CV performance







R

Top 3 performers

Congratulations to our top-performing teams!

submitterid	createdOn =	dockerrepositoryname =	evaluationid	RMSE (wks) =
@herdiants	9/3/2024 4:02 AM	docker.synapse.org/ syn6240 🕂	Placental Clock SC1	1.245
ANTS	8/30/2024 5:25 PM	docker.synapse.org/ <u>syn6184</u> 🕂	Placental Clock SC1	1.603
С 5-ВВК	9/1/2024 6:24 AM	docker.synapse.org/ syn6196 🕂	Placental Clock SC1	1.683





The DREAM Challenges PRESENT THIS CERTIFICATE OF EXCELLENCE TO

Tushar Patel, Alena van Bömmel and Steve Hoffmann

Fritz Lipmann Institute



Gustavo Stolovitzky DREAM Challenges Emeritus Chair







- Filtering and clustering CpG approach captures the gestational aging comprehensively
 - defining important role of feature selection and engineering in ML approaches.
- A simple multivariate elastic-net regression works better in this case compared to more complex non-linear models such as Gaussian Process.
- The selected CpGs indicate to known changing immune related processes in placenta highlighting the biological explainability of our modeling approach.

Acknowledgements





- The placental tissue has an overall unique-lower levels of genome-wide CpG methylation. [Mayne, Benjamin T et al. 2019]
- An overall progressive increase in average methylation from first to third trimester. [Novakovic, Boris et al. 2011]



1) Average methylation of all samples for first, second and third trimester for ~26k probes 2) Unsupervised clustering based on 954 CpG probes with $\Delta\beta$ > 0.2 between First and Third trimester





ChAMP pipeline v2.8.6

Y Tian, T Morris, A Webster, Z Yang, S Beck, A Feber, & A Teschendorff (2024)

✓ Filter out problematic CpG probes ~106k

✓ Perform normalization for type 1 and type 2 illumna probes - BMIQ

× Attempt for batch-effect correction - Combat

Finding gestational aging signature



22





Science Immunology

An immune clock of human pregnancy

 NIMA AGHAEEPOUR, EDWARD A. GANIO

 , DAVID MCILWAIN, AMY S. TSAI
 , MARTHA TINGLE
 , SOFIE VAN GASSEN
 , DYANI K. GAUDILLIERE
 , QUENTIN BACA,

 LESLIE MCNEIL
 , L..., AND BRICE GAUDILLIERE

Authors Info & Affiliations
SCIENCE IMMUNOLOGY

 1 Sep 2017

Aghaeepour et al. used mass cytometry to examine the precise timing of these pregnancy-induced changes in immune function and regulation.

• *Example*: STAT5ab featured in immune clock & JAK3 from our findings.



An immunological "clock" of pregnancy. Credit: Carla Schaffer / AAAS



GP-age: chronological age prediction from blood DNA methylation using few sites and cohort-based algorithm



Cell Reports Methods

Article

Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm

> Miri Varshavsky, Gil Harari, Benjamin Glaser, Yuval Dor, Ruth Shemer, Tommy Kaplan



Benchmarking

1) model type

- 2) feature selection method
- 3) number of input features to the model



met2 : sd_variance met3 : k-means