

Machine learning approaches to predict preterm birth by non-invasive means

Samantha L Wilson, PhD

Postdoc Fellow | Princess Margaret Cancer Centre | Hoffman Lab

R Ladies Toronto, June 18th 2019



R Packages

plyr

dplyr

knitr

limma

ggplot2

GridExtra

boot

reshape2

glm

lm

MLR

carrot

rmarkdown

RSamtools

stringr

dendextend

impute

pvclust

GenomicFeatures

sigclust2

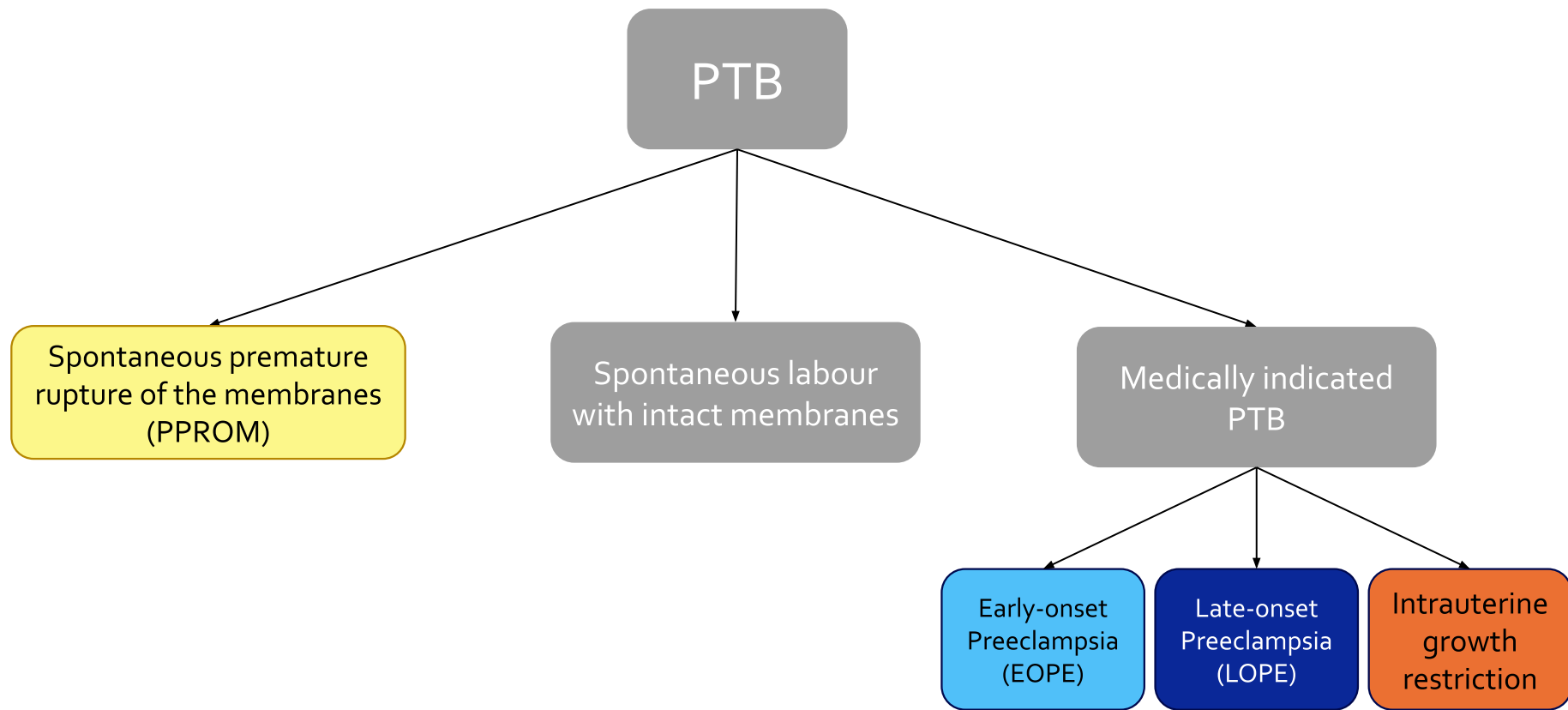
WGCNA

ggjoy

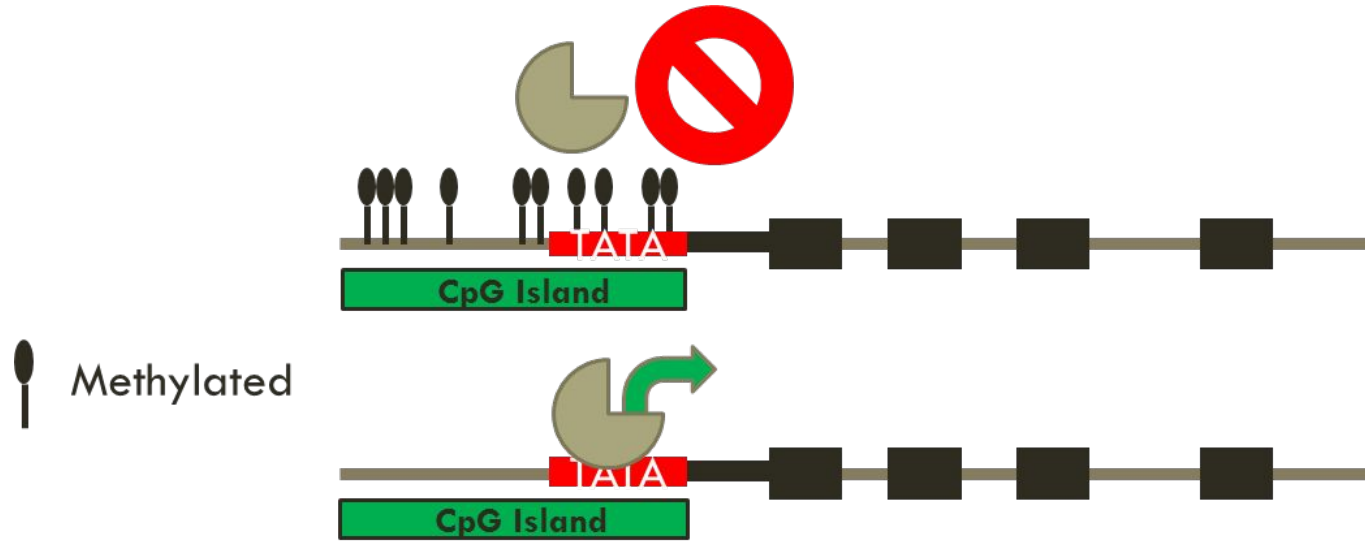
lumi

methylumi

What is preterm birth and why study it?

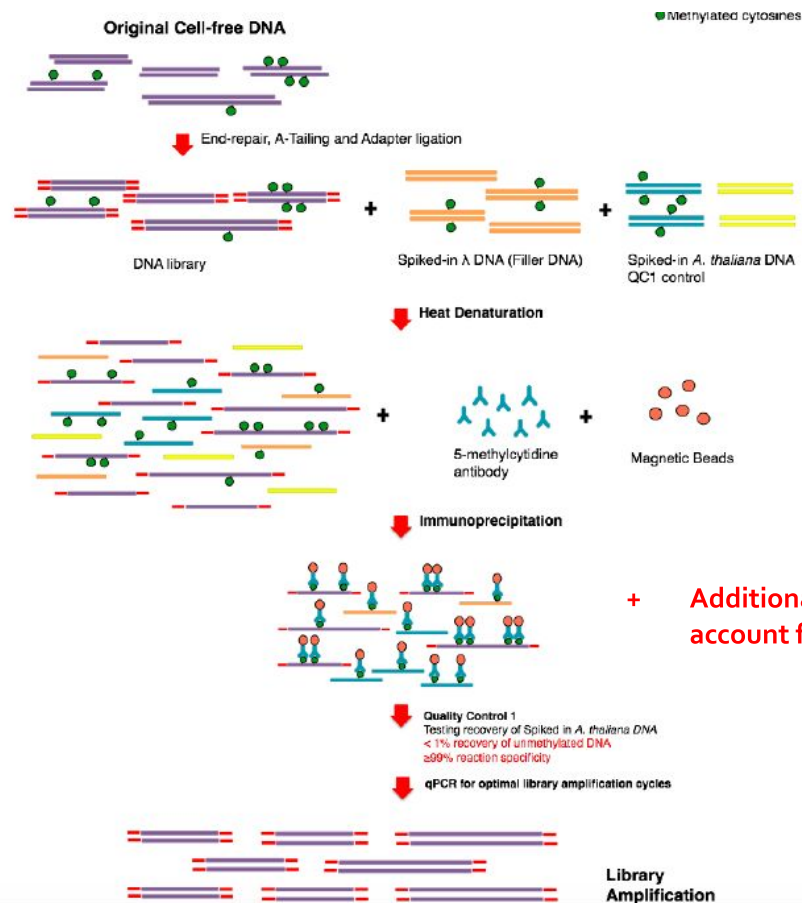


Hypothesis: Different subtypes of PTB are distinct entities which will be associated with distinct cfp-DNA signatures which we can use to build a predictive model of PTB.



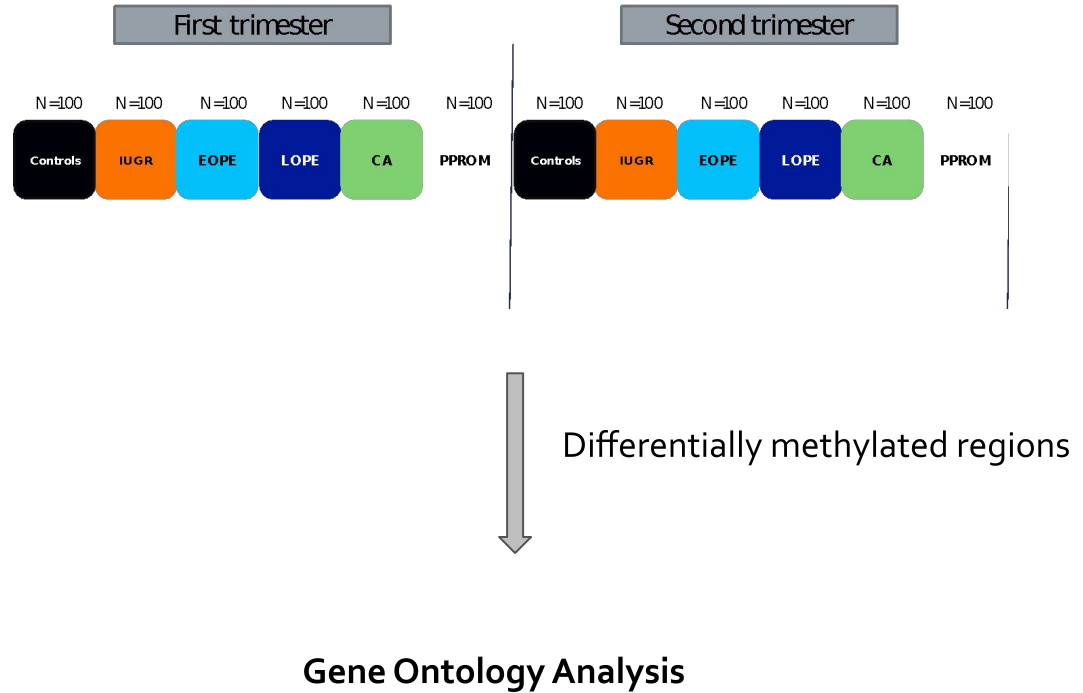
DNA methylation

Measuring DNAm: cfMeDIP



Shen *et al.* Nature. 2019.

Aim 2: Identify differential methylation in preterm birth subtypes



Aim 3: Build a predictive model of preterm birth

Group	Sample <i>n</i> =67 per group	Clinical features		DNAm data per 300bp window (Adjusted allele-specific methylation)				
		Fetal sex	Maternal Chronic hypertension	Greater Allele <i>chr1.90300.90600</i> <i>chr1.90601.90900</i>		Lesser Allele <i>chr1.90300.90600</i> <i>chr1.90601.90900</i>		
A.	Controls	1	M	-	0.77	0.81	0.22	0.17
	
		67	F	+	0.64	0.84	0.15	0.33
LOPE		68	M	+	0.98	0.77	NA	0.27
	
		134	F	+	0.59	0.92	0.04	0.14
EOPE		135	F	-	0.64	0.81	0.17	0.34
	
		201	F	+	0.66	0.84	0.33	NA
IUGR		202	M	-	0.77	0.77	0.34	0.22
	
		268	M	-	0.92	0.64	NA	0.15
CA		269	M	-	0.77	0.98	0.27	NA
	
		335	F	-	0.64	0.59	0.14	0.04
SPL		336	M	-	0.81	0.64	NA	0.17
	
		402	M	+	0.84	0.66	0.14	0.33

B. Training data

Numerous predictive models built on training data

Multiple-fold cross validation

Tuning data

Numerous predictive models tested for utility

C. Testing data

n=33 per group, same data type and feature. Withheld from building predictive model. Best model is tested on validation model for utility