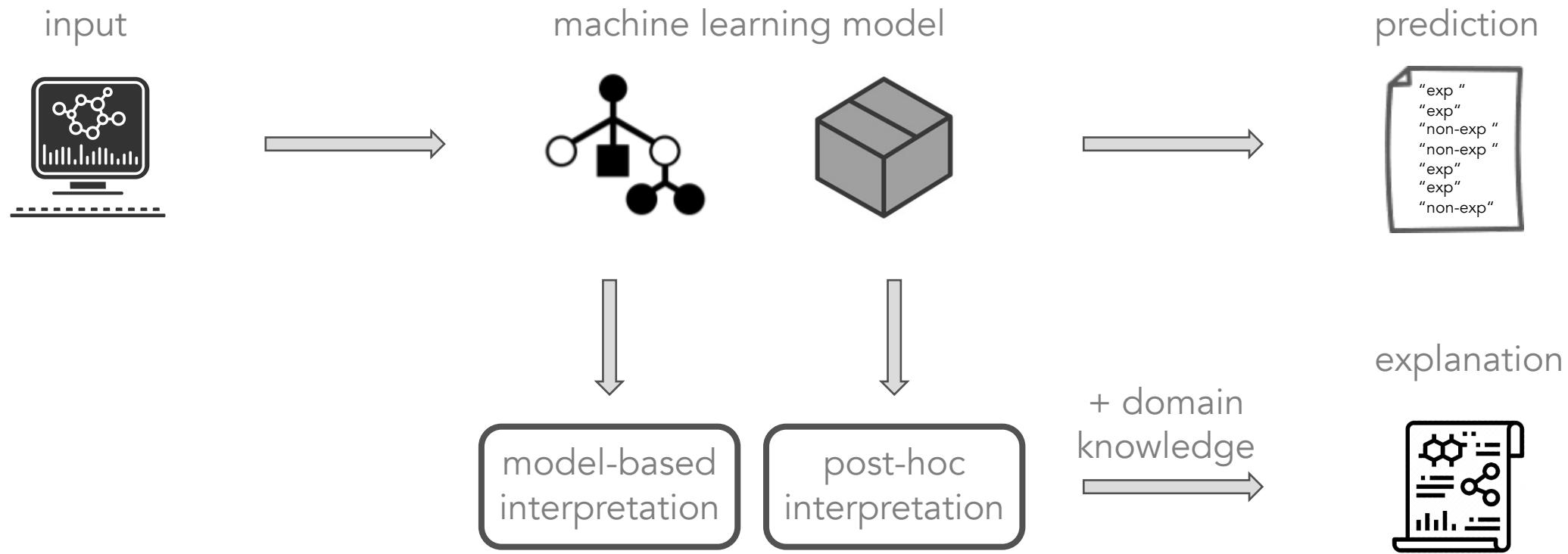


# XAI for Random Forest

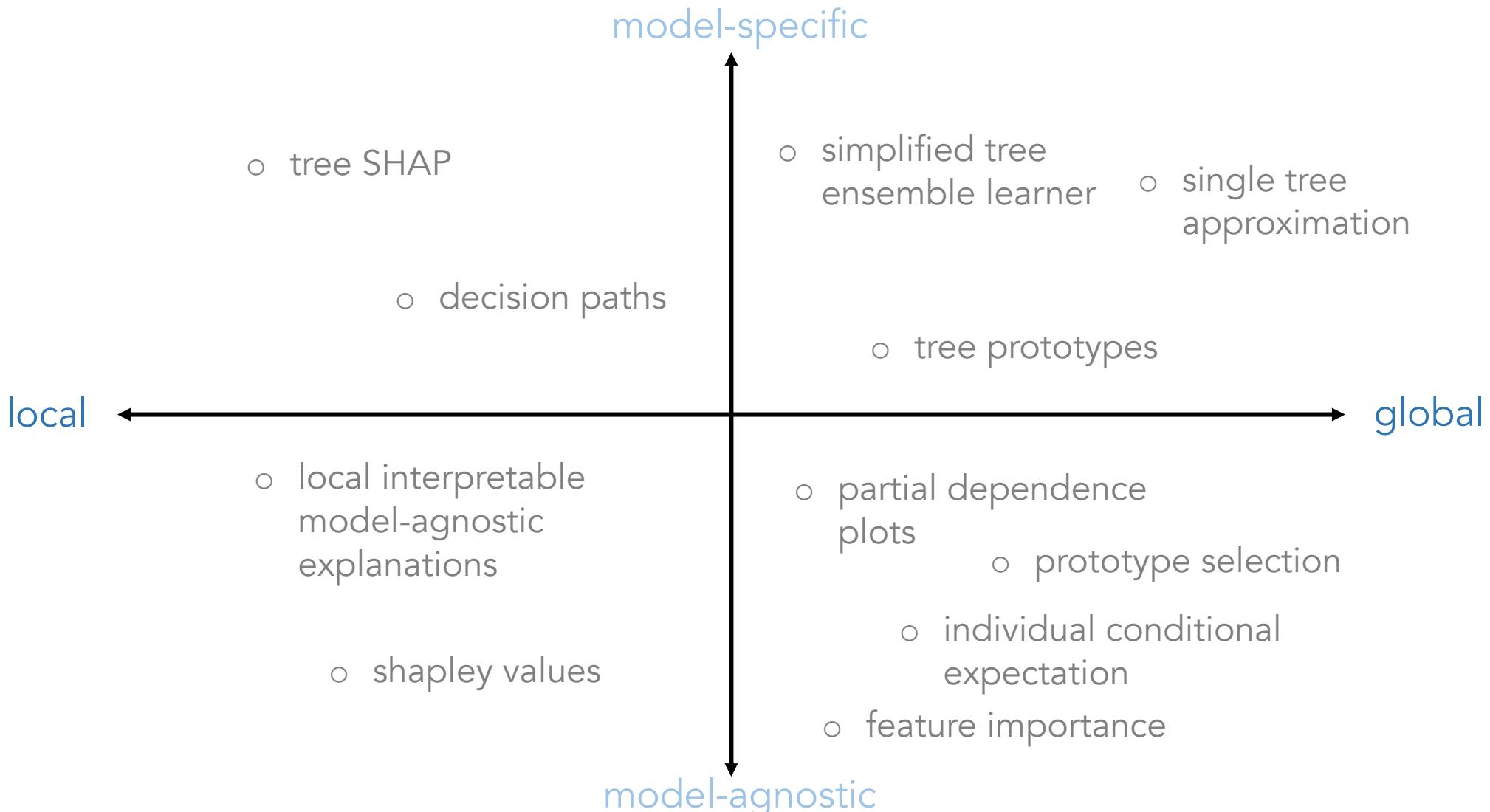


# How is interpretability defined?

---

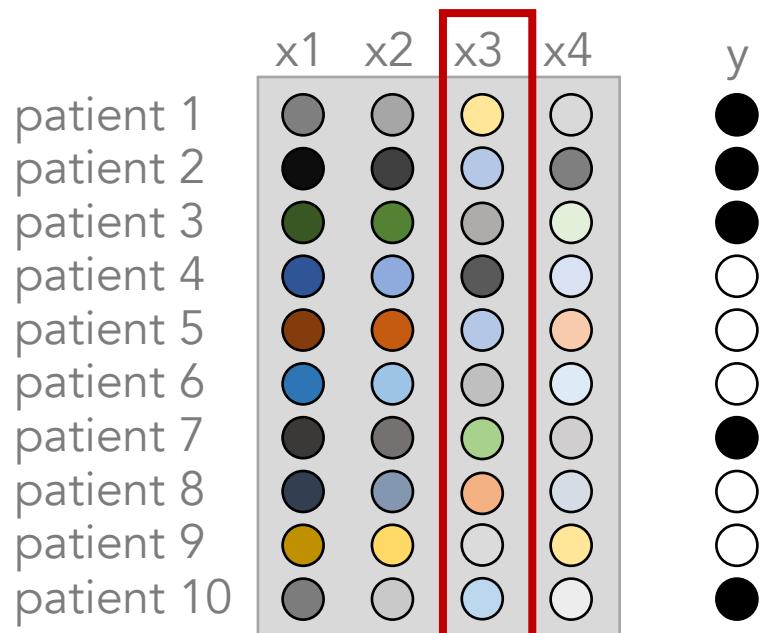


# Post-hoc interpretation methods



# Feature importance – a global model-agnostic method

idea: measure feature importance through feature permutation (Breiman, 2001)

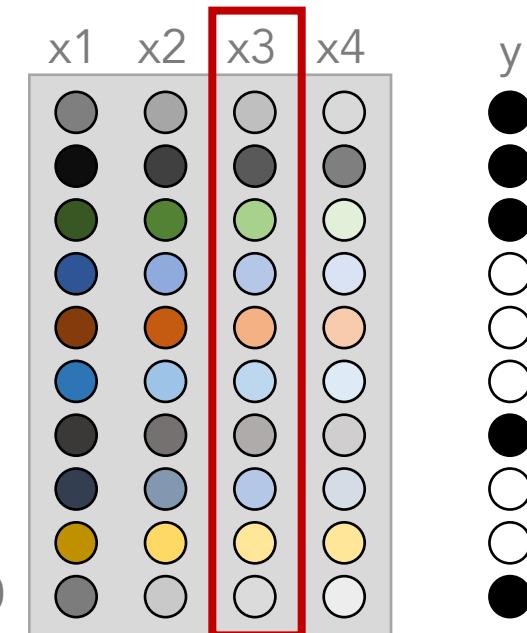


$$FI_{x_3}$$

$$e_{perm} = L(y, f(X_{perm}))$$

-

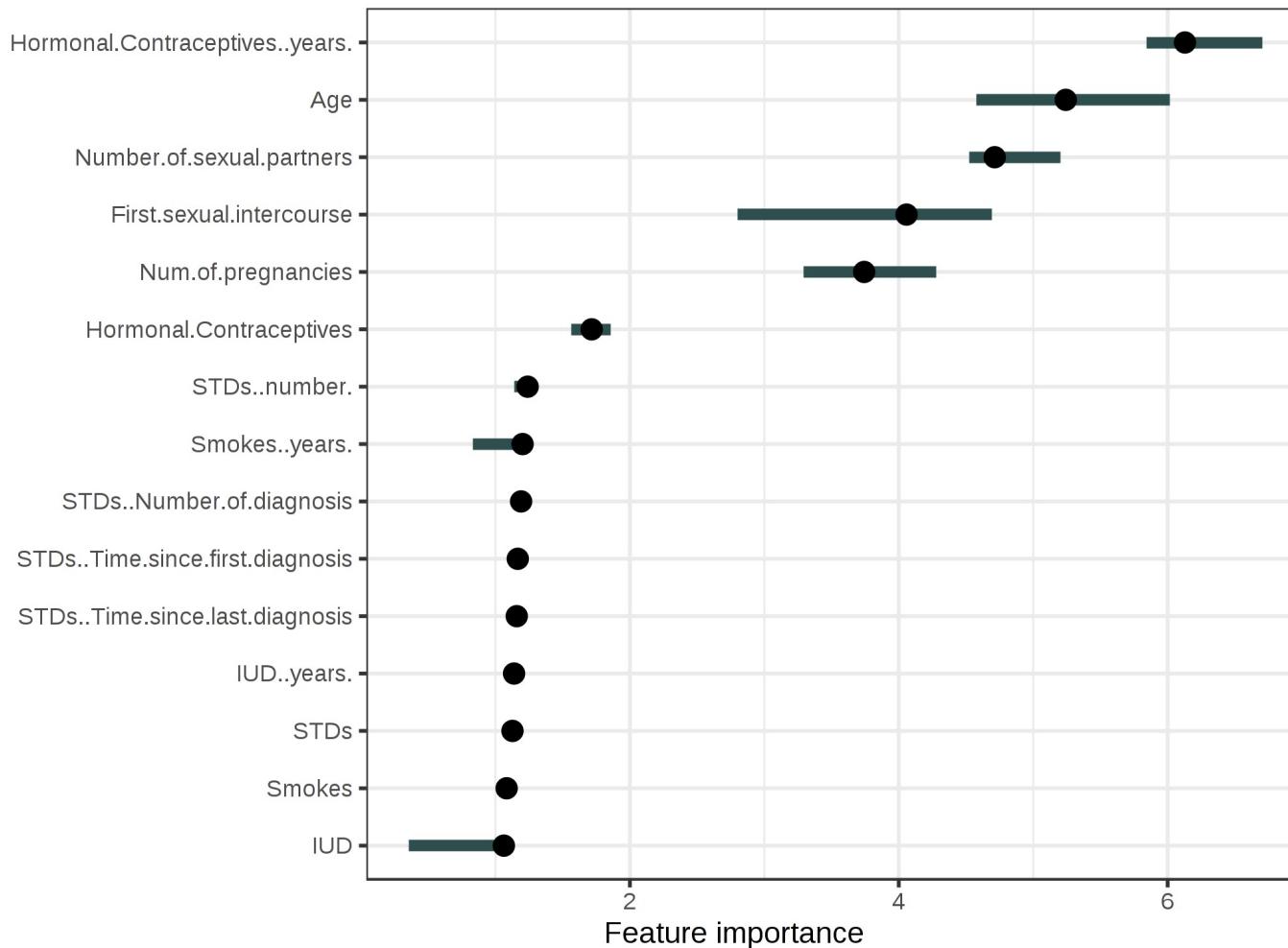
$$e_{orig} = L(y, f(X))$$



# Feature importance – a global model-agnostic method

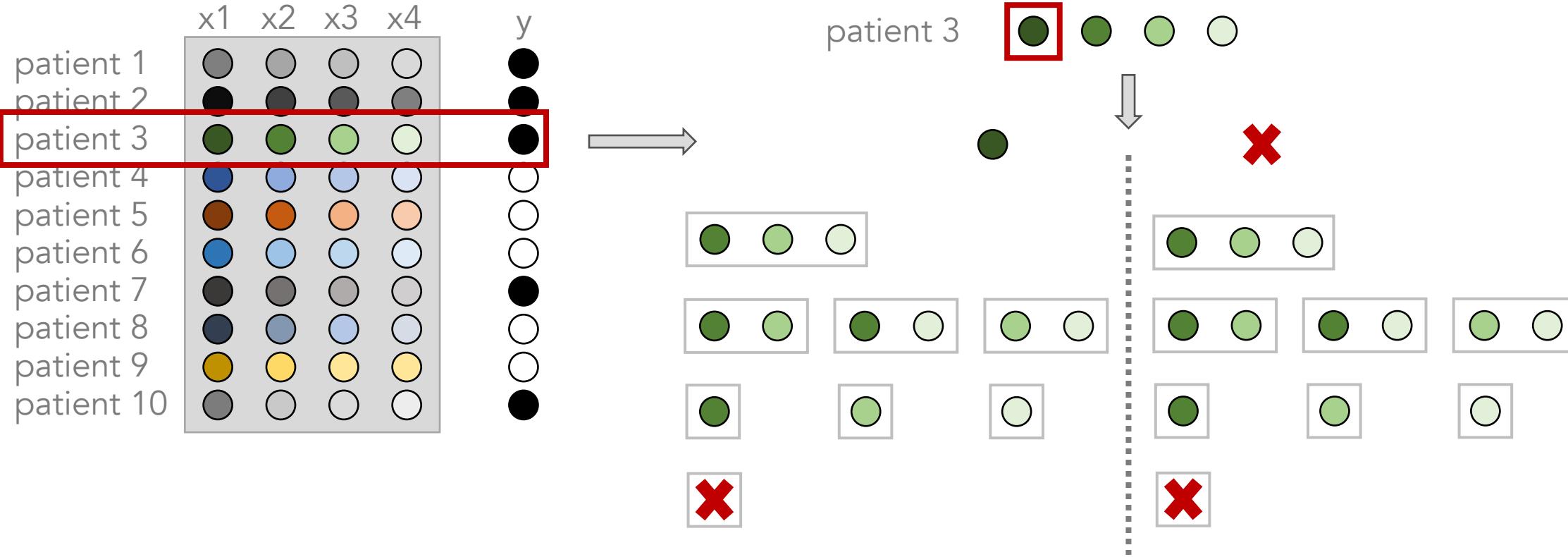
idea: measure feature importance through feature permutation (Breiman, 2001)

example for cervical cancer classification (Molnar, 2019):



# Shapley values – a local model-agnostic method

idea: compute feature contributions for single predictions (Shapley, 1953)

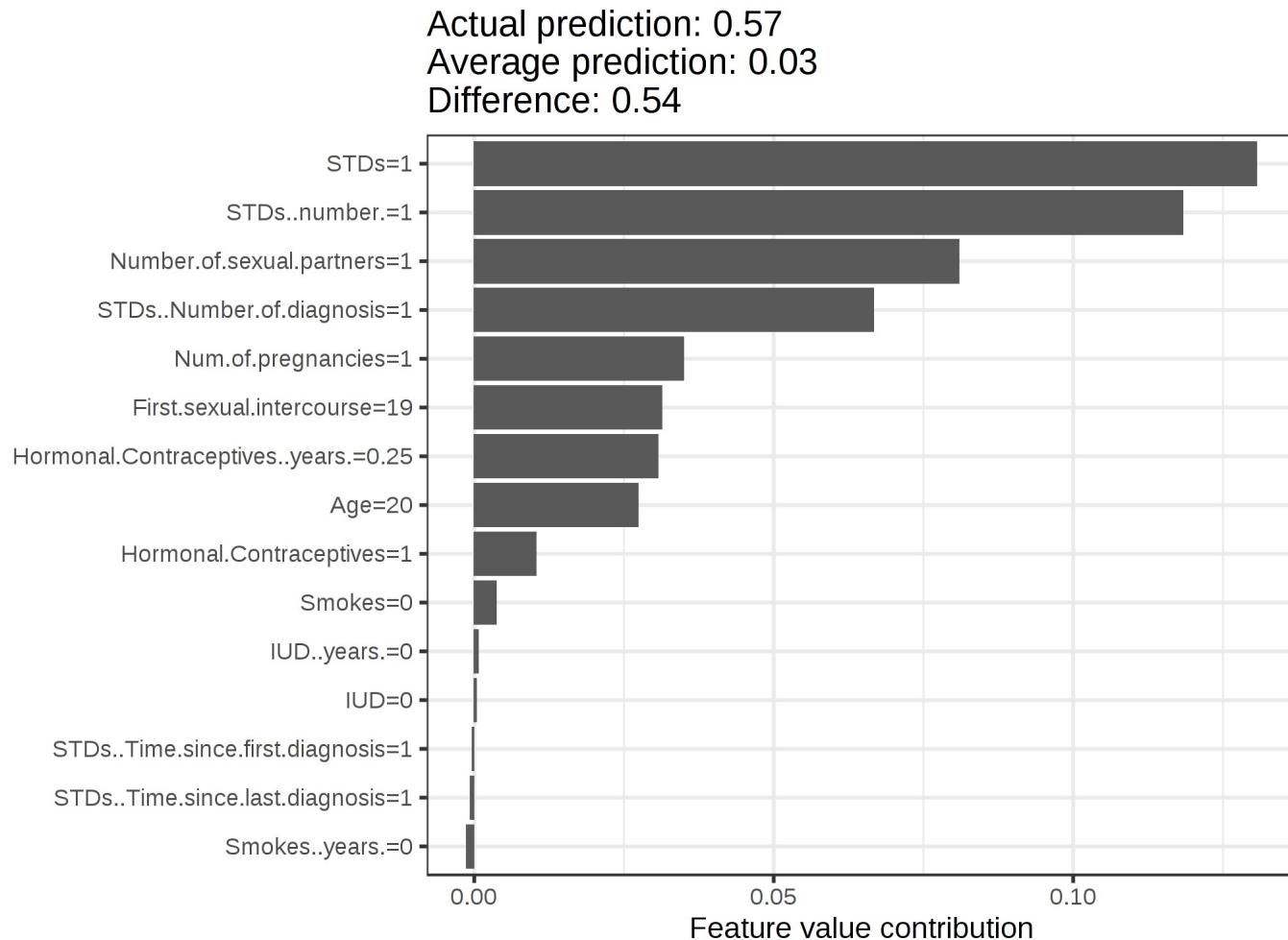


# Shapley values – a local model-agnostic method

---

idea: compute feature contributions for single predictions (Shapley, 1953)

example for cervical cancer classification (Molnar, 2019):

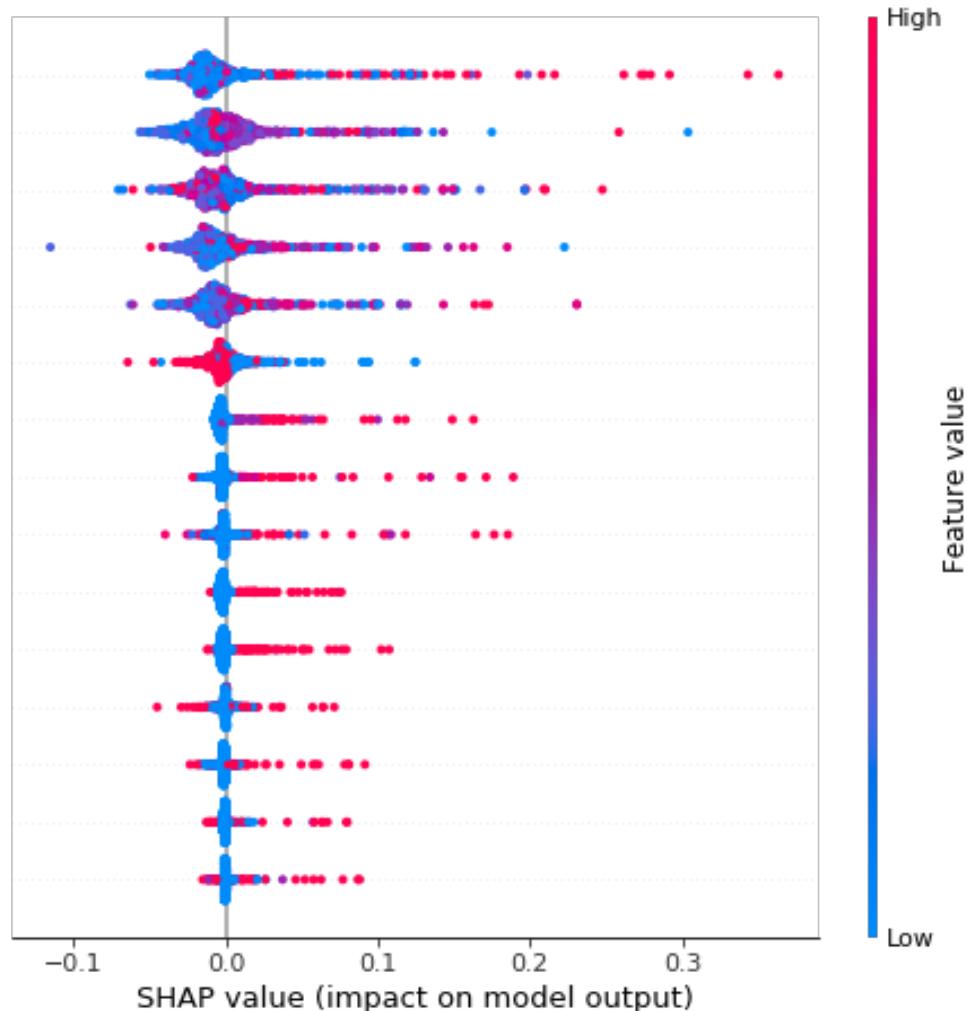


# SHapley Additive exPlanations (SHAP)

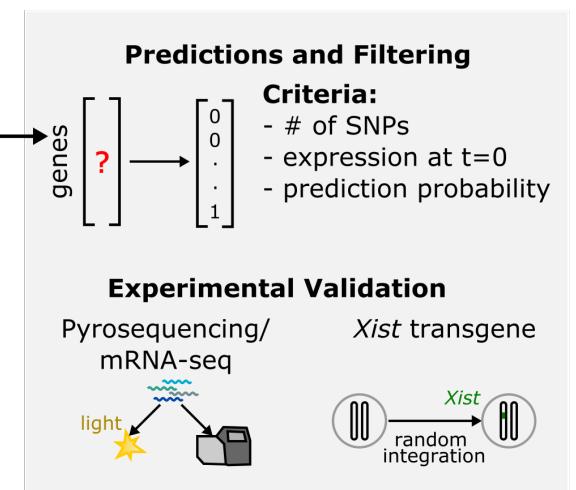
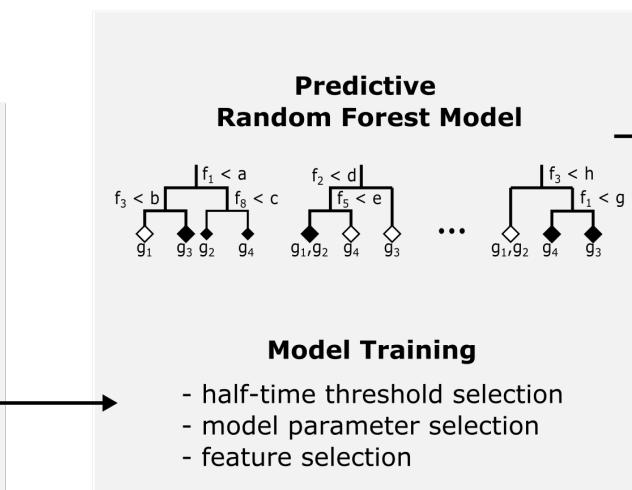
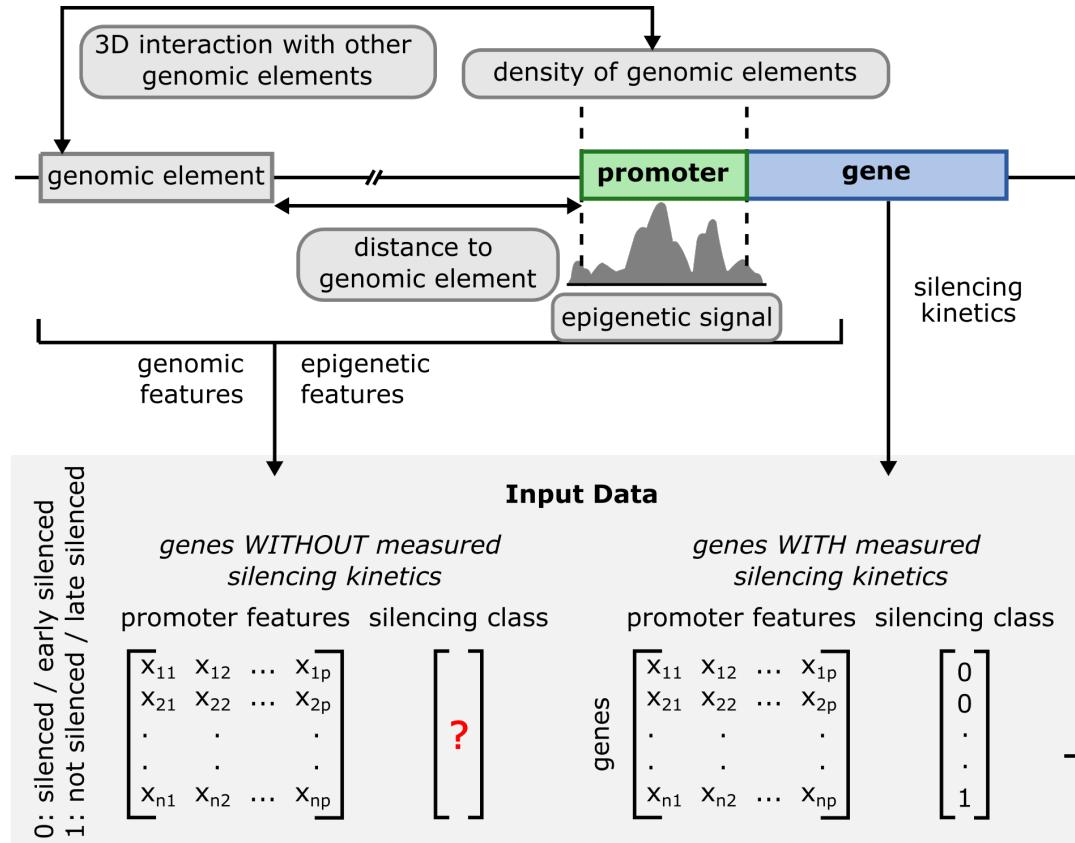
idea: alternative estimation method for shapley values to reduce computing time  
(Lundberg et al., 2017)

example for cervical cancer classification (Molnar, 2019):

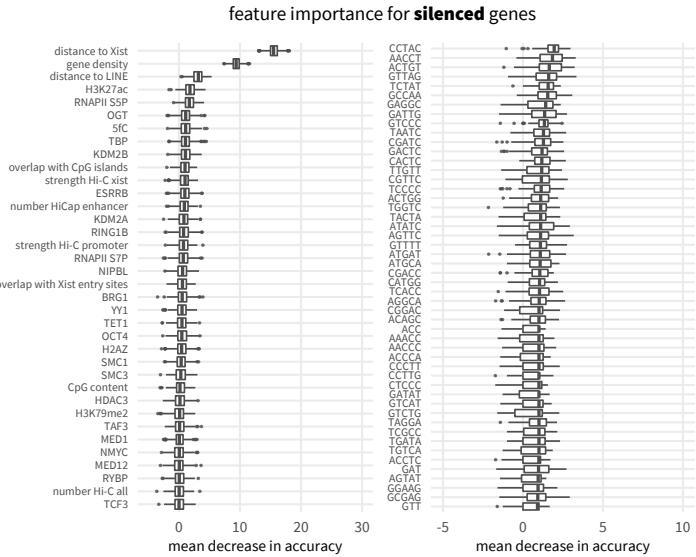
Hormonal.Contraceptives..years.  
First.sexual.intercourse  
Age  
Num.of.pregnancies  
Number.of.sexual.partners  
Hormonal.Contraceptives  
STDs..number.  
IUD..years.  
Smokes..years.  
STDs  
STDs..Number.of.diagnosis  
Smokes  
IUD  
STDs..Time.since.last.diagnosis  
STDs..Time.since.first.diagnosis



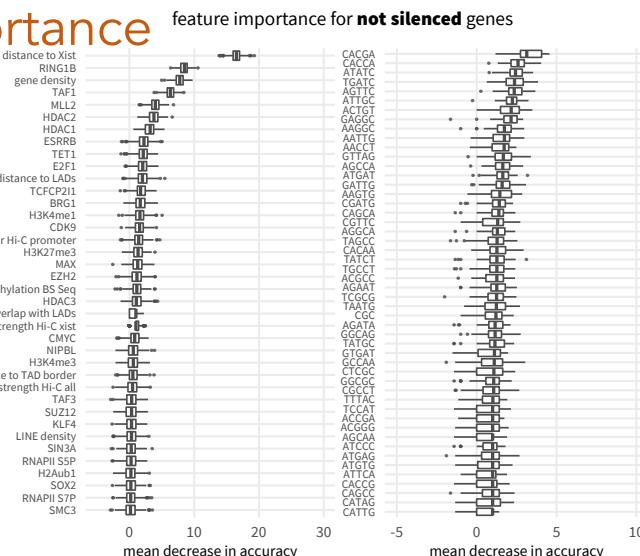
# Modelling approach



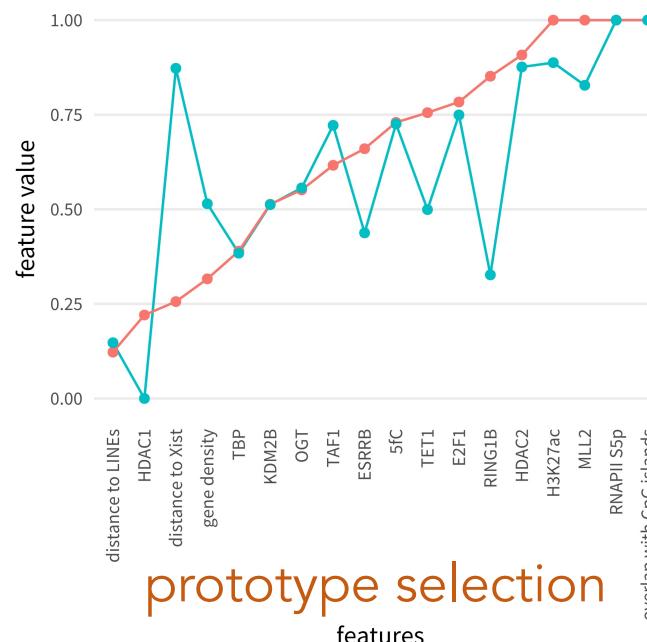
# Model Interpretation



# feature importance



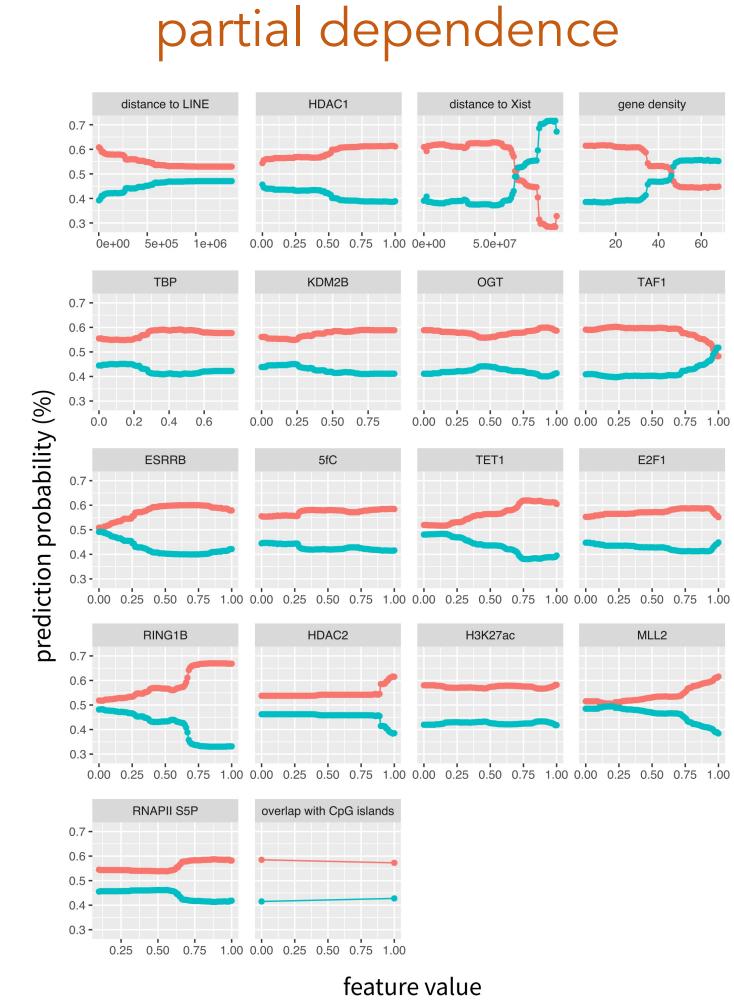
## XCI/escape model trained on epigenetic and genomic features



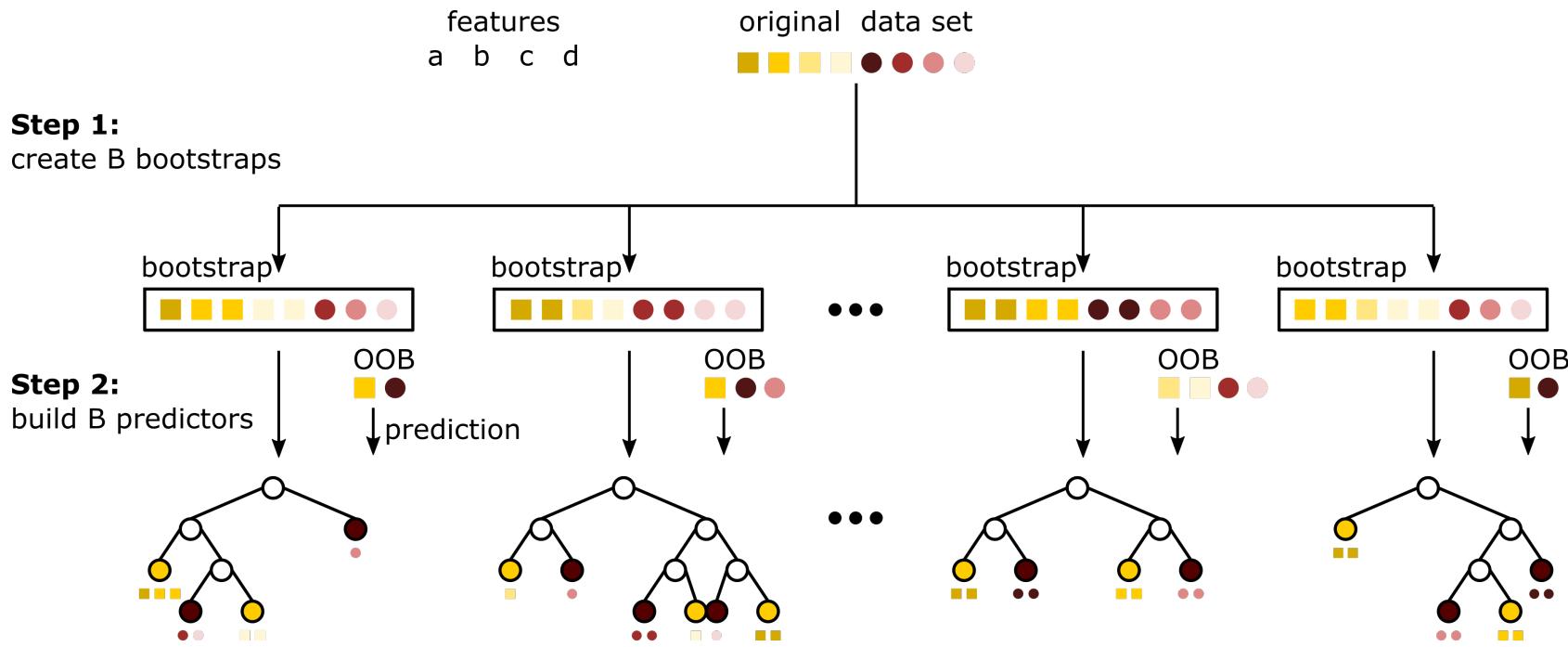
# prototype selection

features

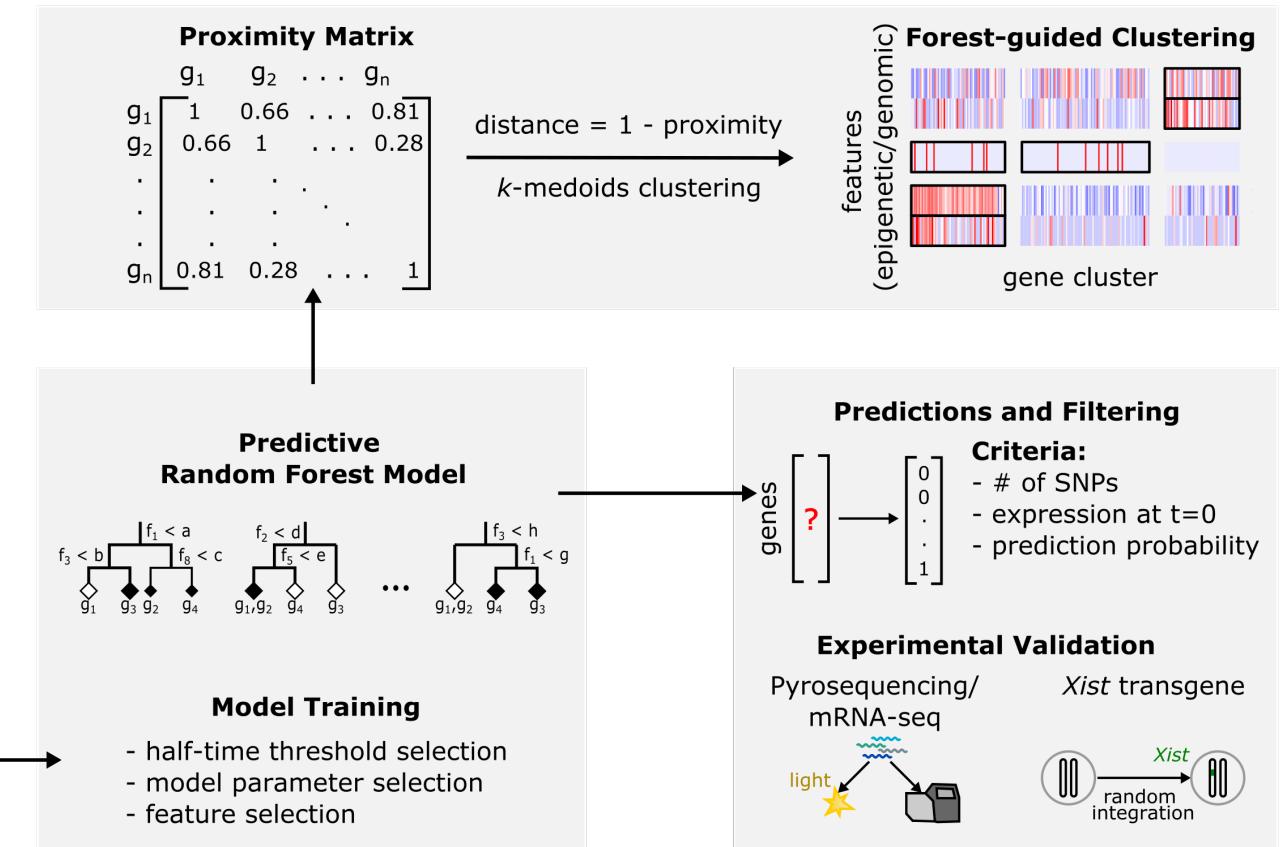
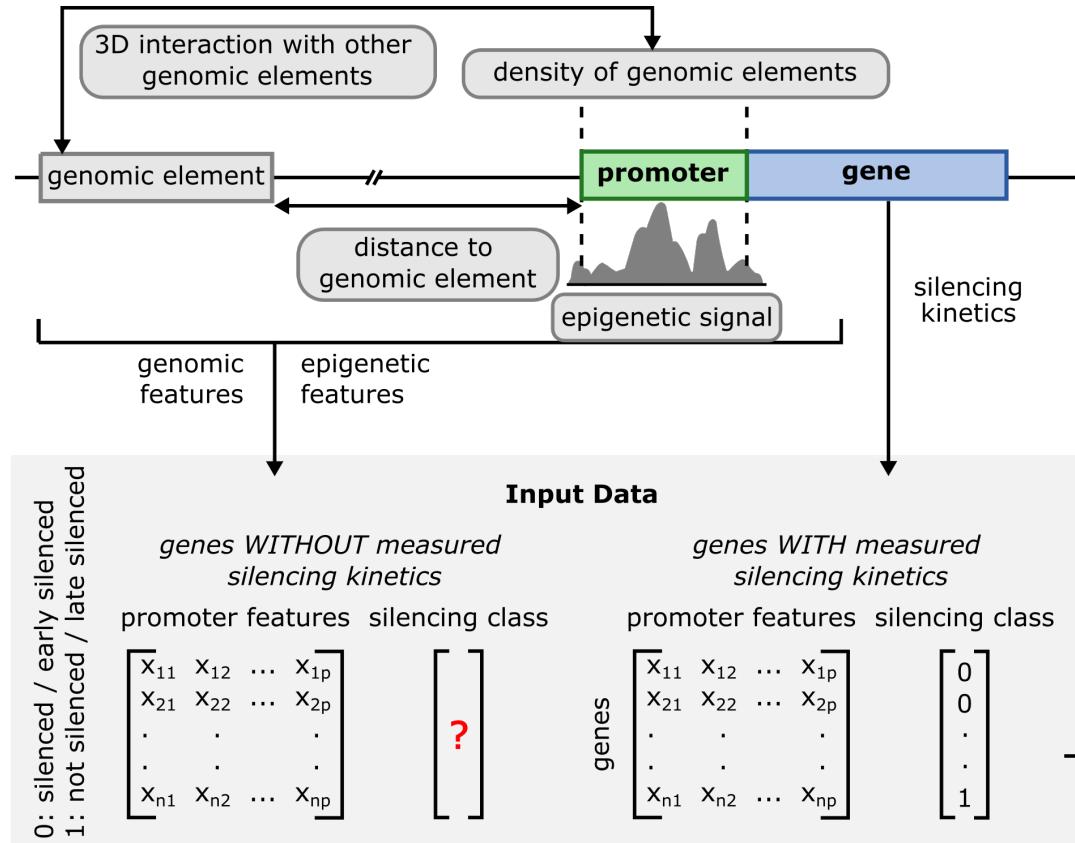
class    ● silenced genes  
          ● not silenced genes



# Random Forest Models



# Modelling approach



# Forest-guided clustering of XCI/escape model

