

# Interpretable classification of molecular measurements via pathway-induced multiple kernel learning

—  
Matteo Manica

# Roadmap

Molecular data classification

Pathway-Induced Multiple Kernel Learning (PIMKL)

PIMKL benchmarking

PIMKL application

# Roadmap

Molecular data classification

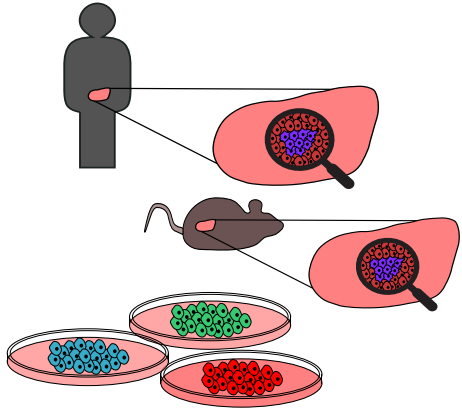
Pathway-Induced Multiple Kernel Learning (PIMKL)

PIMKL benchmarking

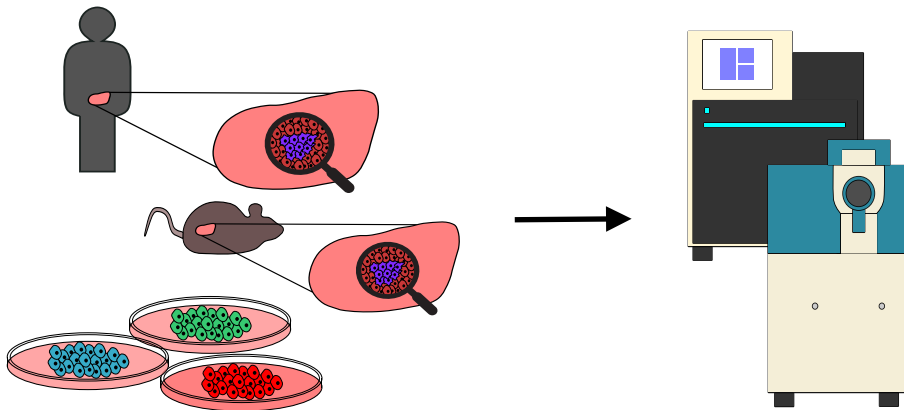
PIMKL application

# Molecular data classification

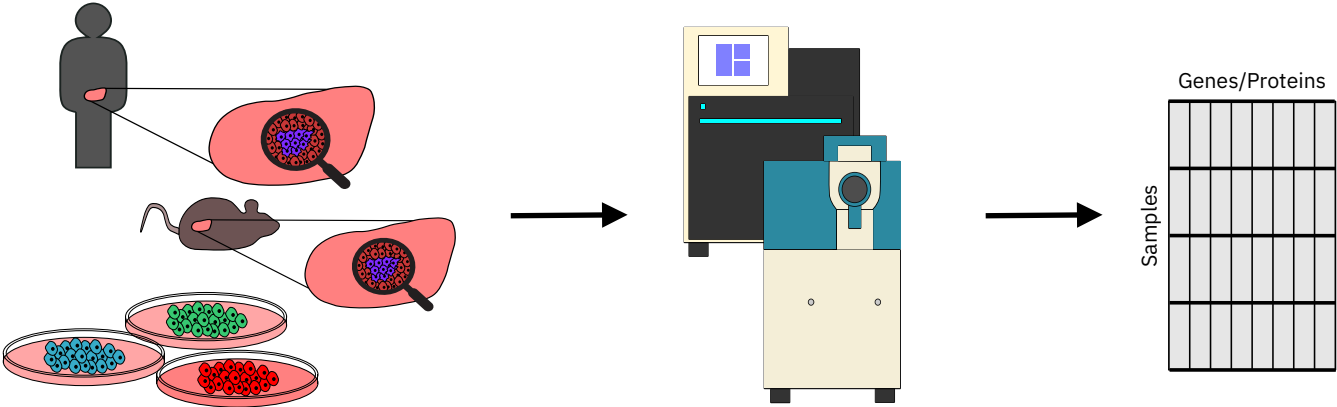
# Molecular data classification



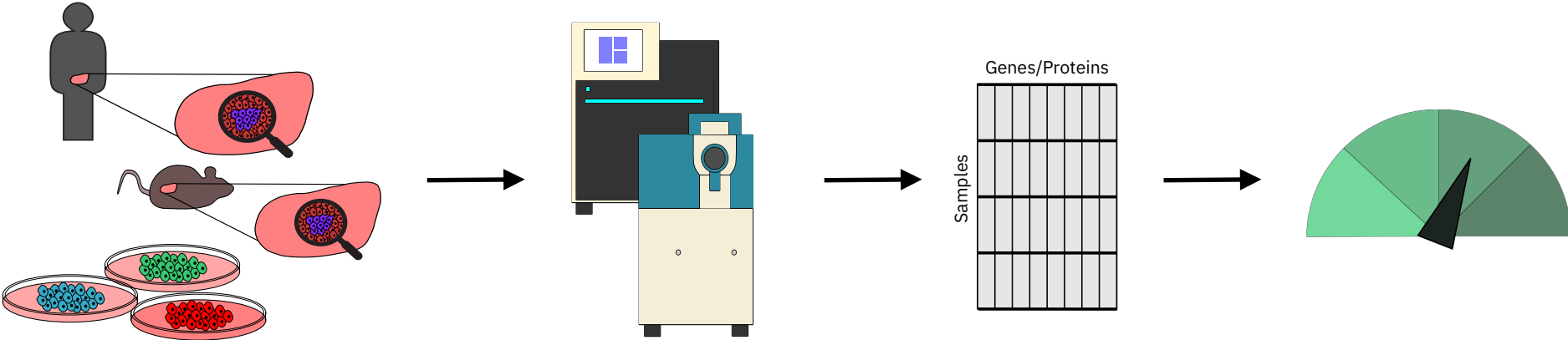
# Molecular data classification



# Molecular data classification



# Molecular data classification

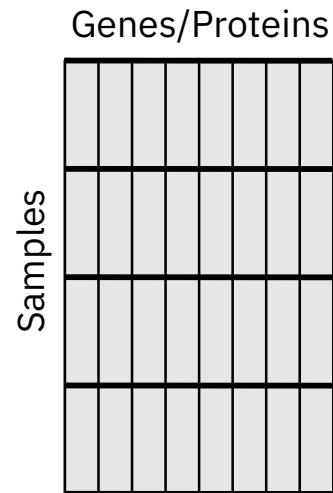




# Molecular data classification - challenges

Experiments costs and noise

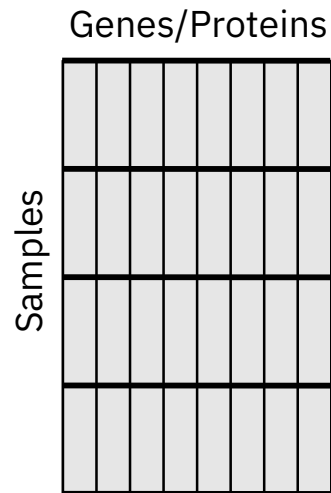
Scarce sample availability in high throughput experiments



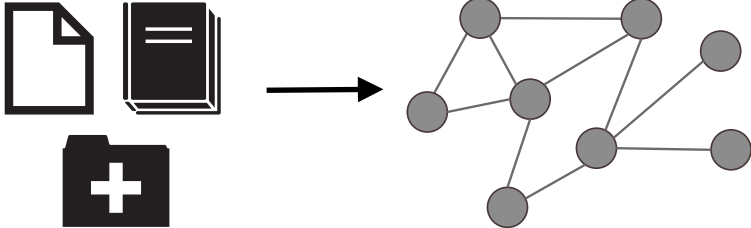
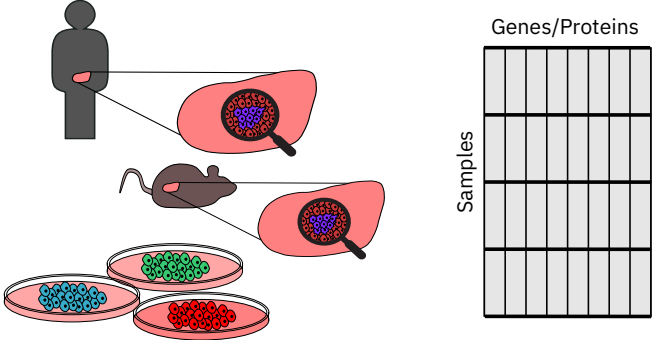
# Molecular data classification - challenges

Experiments costs and noise

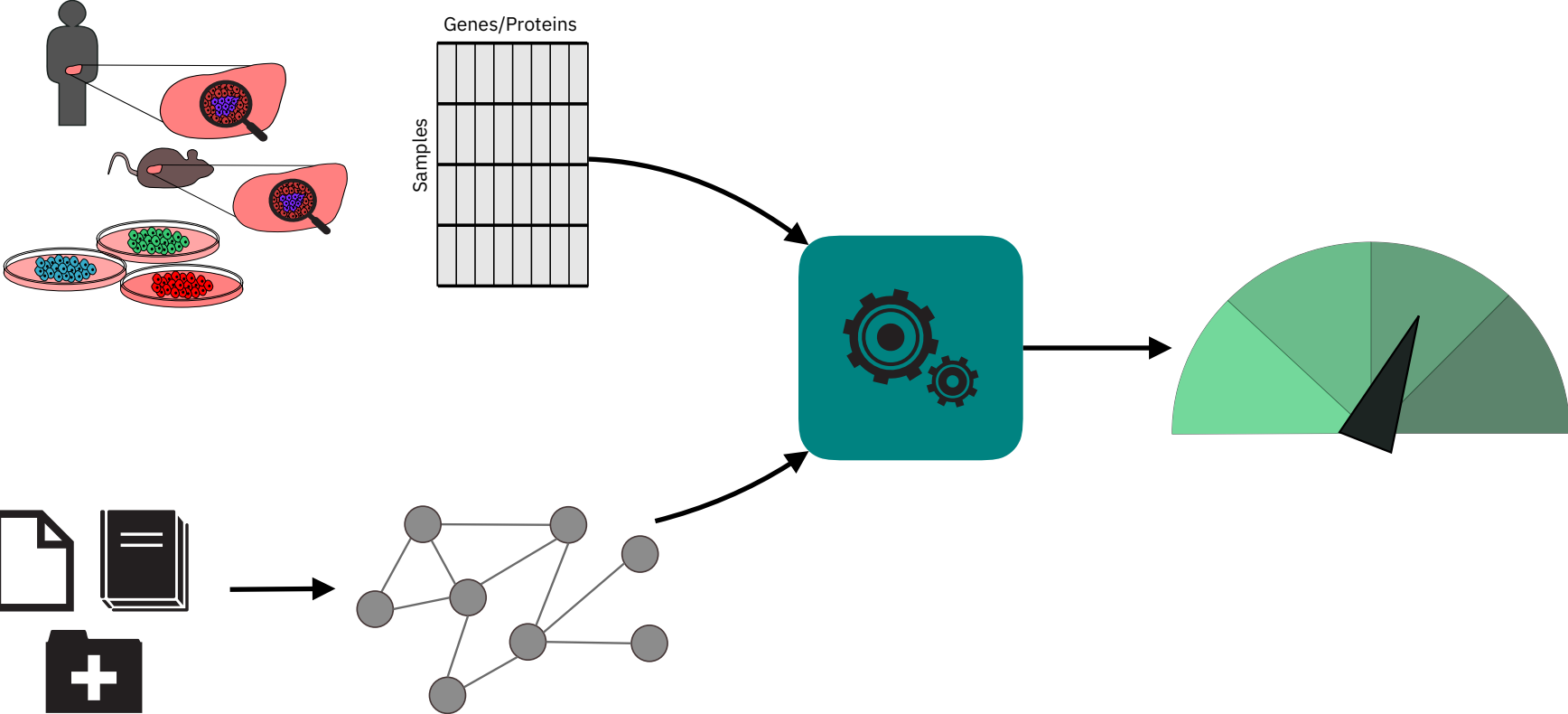
Scarce sample availability in high throughput experiments



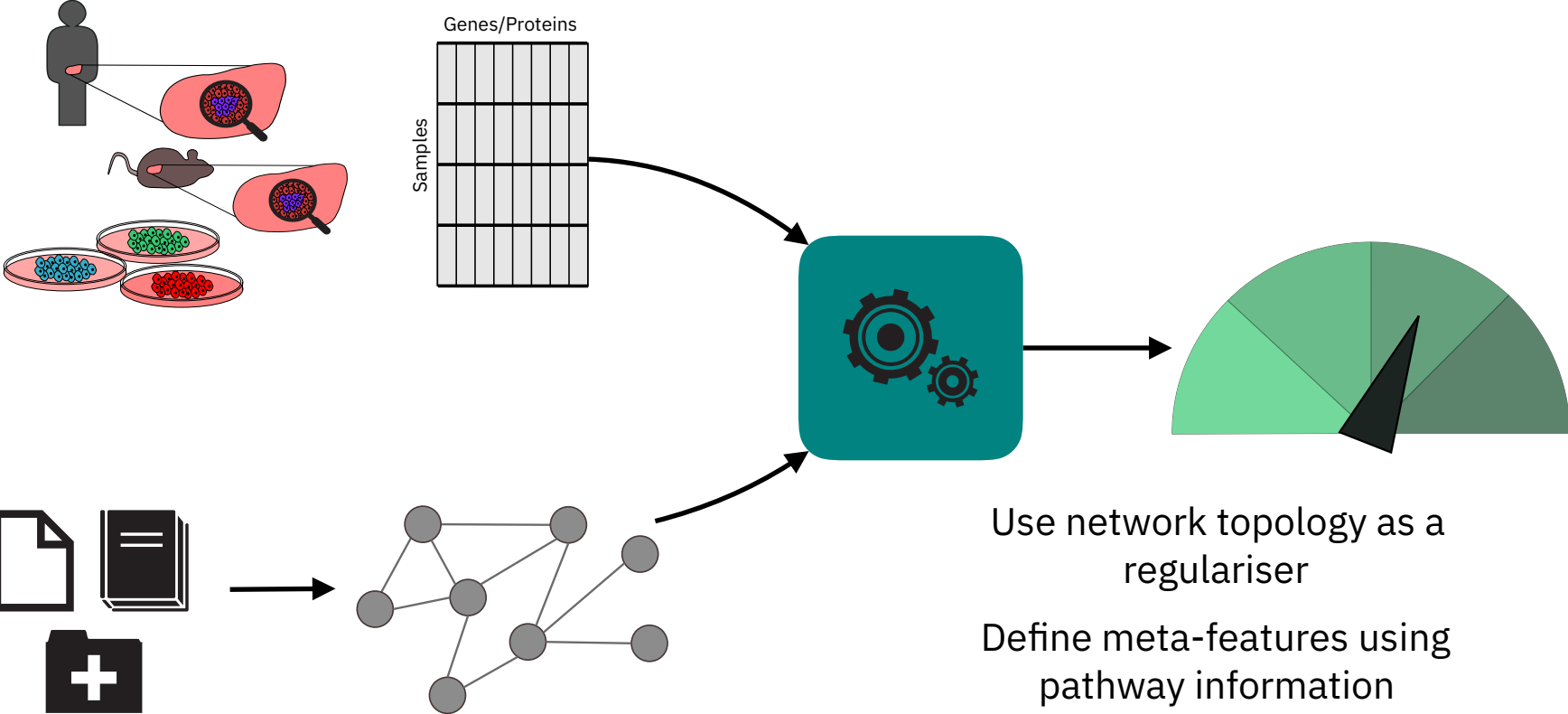
# Molecular data classification - exploit prior knowledge



# Molecular data classification - exploit prior knowledge



# Molecular data classification - exploit prior knowledge



L. Chen et al., *BMC Systems Biology*, 2011  
Z. Guo, *BMC Bioinformatics*, 2005

# Roadmap

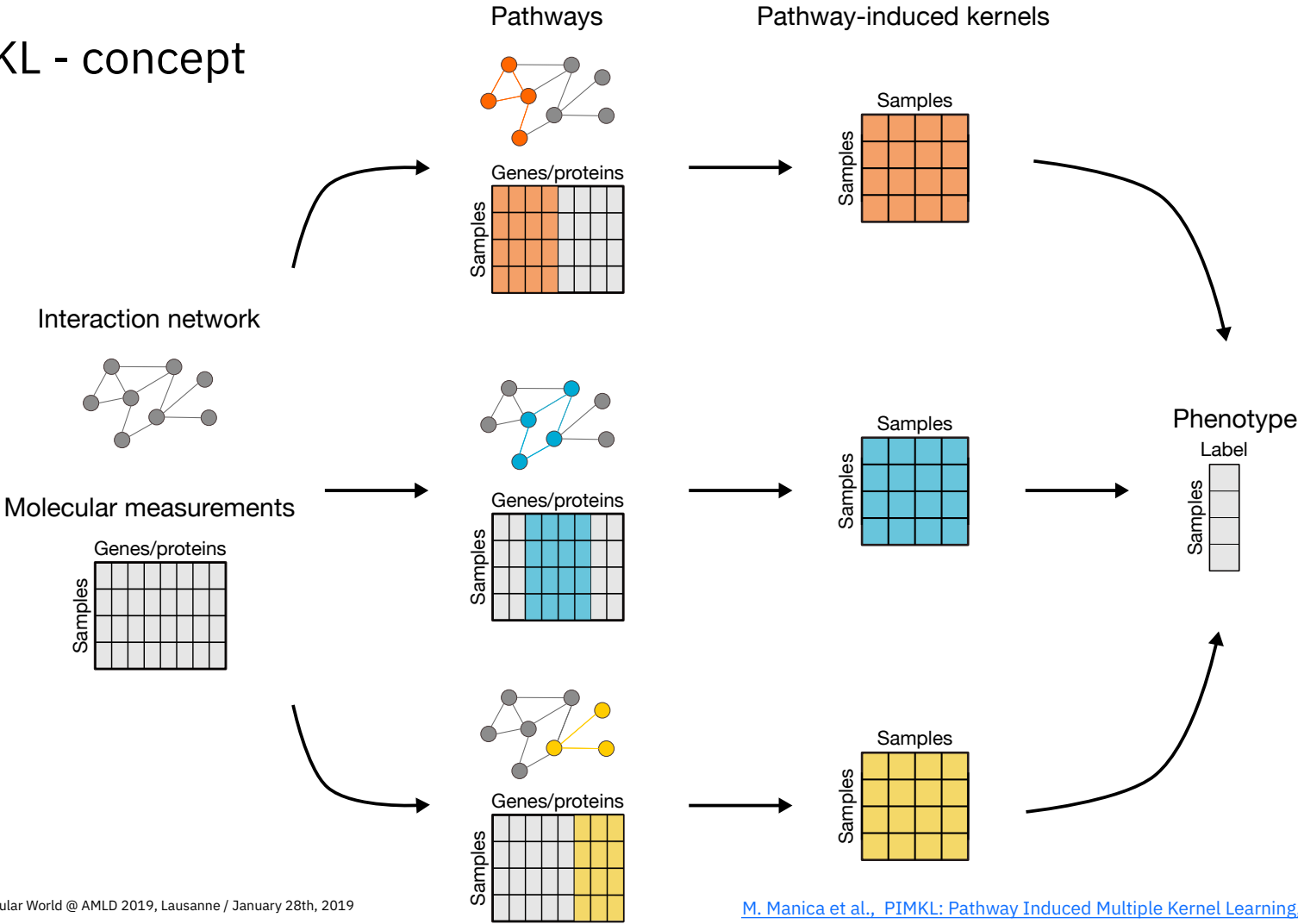
Molecular data classification

Pathway-Induced Multiple Kernel Learning (PIMKL)

PIMKL benchmarking

PIMKL application

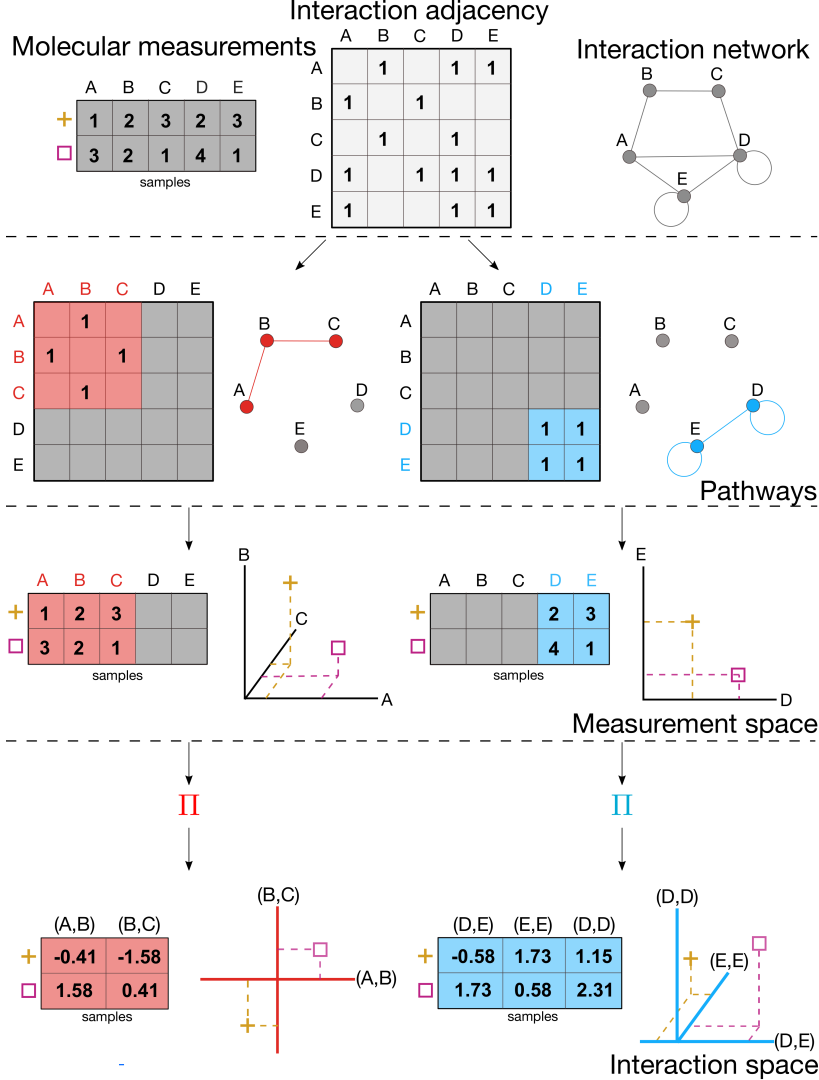
# PIMKL - concept



# PIMKL - pathway induction

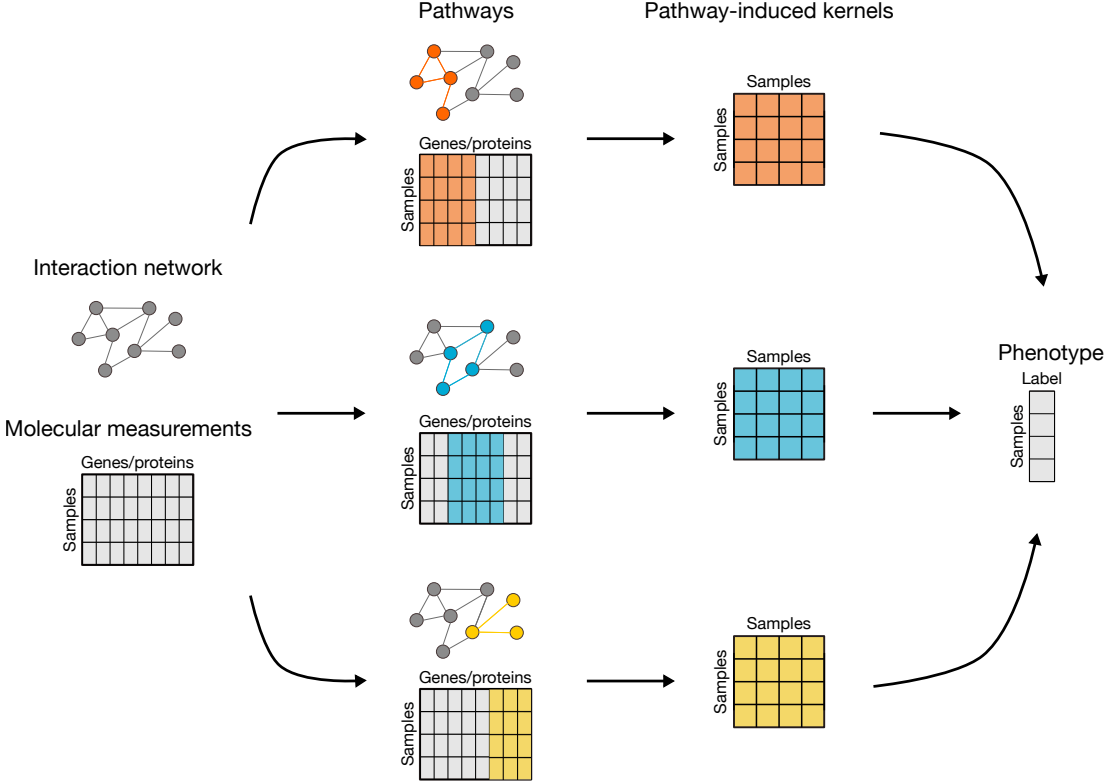
$$\begin{cases} k_{L_\omega}(x, y) = x^T L_\omega y = x^T S_\omega S_\omega^T y = \Pi(x)^T \Pi(y) \\ S_\omega = D^{-\frac{1}{2}} S W^{\frac{1}{2}} \end{cases}$$

Pathway induction map the data from the gene space to the interaction space





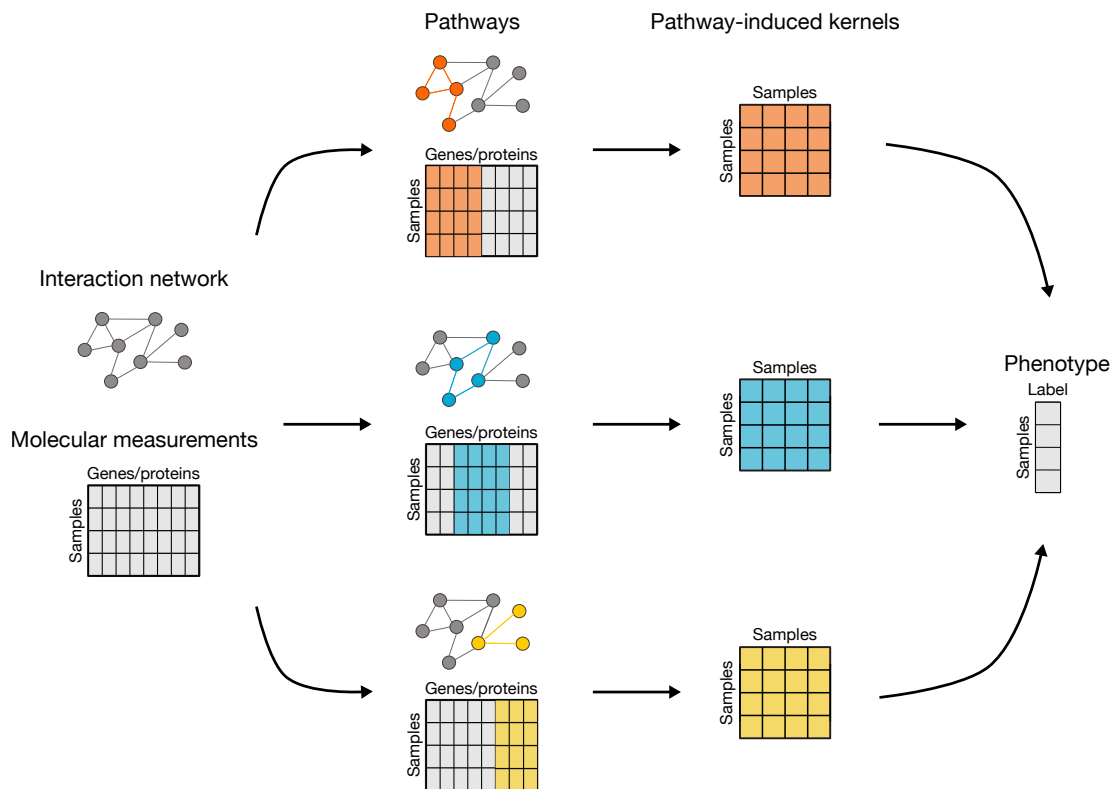
# PIMKL - multiple kernel learning



Weighted combination of pathway-induced kernels to optimize phenotype prediction

$$\begin{cases} K_{ij}^p = k_{L^p}(x_i, x_j) \\ K = \sum_{p=1}^P w_p K^p, \quad w_p \geq 0 \end{cases}$$

# PIMKL - multiple kernel learning



Weighted combination of pathway-induced kernels to optimize phenotype prediction

$$\begin{cases} K_{ij}^p = k_{L_o}^p(x_i, x_j) \\ K = \sum_{p=1}^P w_p K^p, \quad w_p \geq 0 \end{cases}$$

Included using the kernel trick in an iterative optimization procedure

$$\begin{aligned} & \max_{\alpha_i \geq 0, w_p \geq 0} \sum_i \alpha_i - \frac{1}{2} \sum_p \sum_i \sum_j \alpha_i \alpha_j y_i y_j w_p K_{ij}^p \\ & \text{subject to} \quad 0 \leq \alpha_i \leq C \\ & \quad \quad \quad 0 \leq w_p \leq C' \\ & \quad \quad \quad \sum_i \alpha_i y_i = 0 \end{aligned}$$

# Roadmap

Molecular data classification

Pathway-Induced Multiple Kernel Learning (PIMKL)

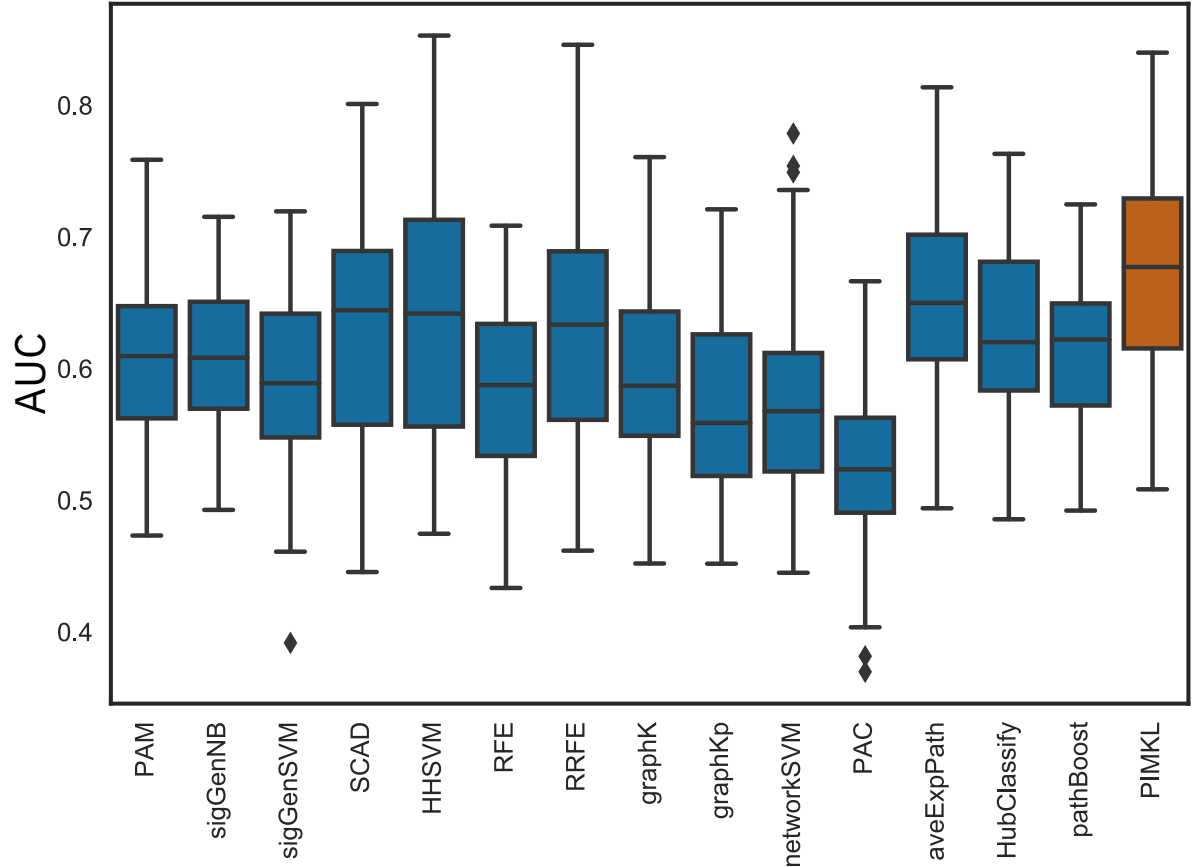
**PIMKL benchmarking**

PIMKL application

# PIMKL benchmarking

Benchmark against other prior knowledge informed methods on multiple breast cancer cohorts

PIMKL significantly improves prediction of tumor relapse



Y. Cun et al. *BMC bioinformatics*, 2012

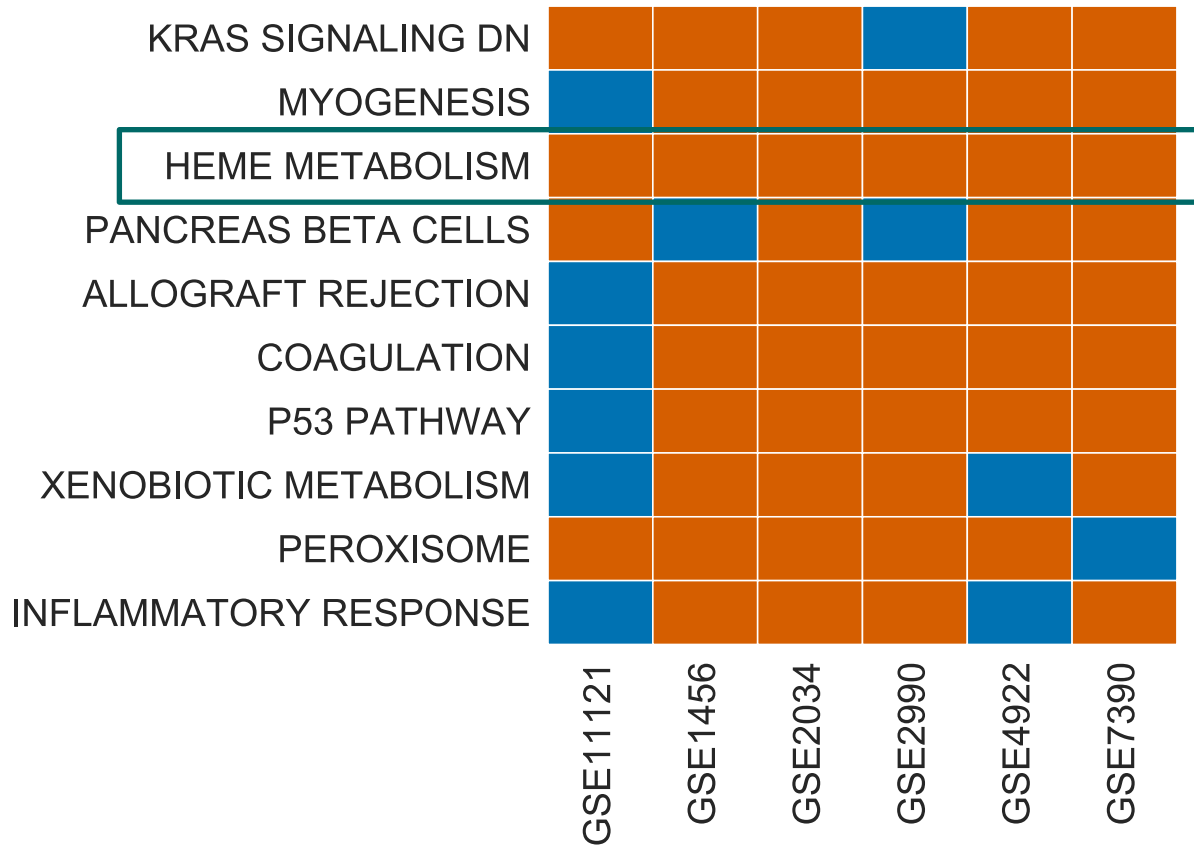
A. Liberzon et al. *Cell systems*, 2015

M. Kanehisa et al. *Nucleic Acids Research*, 2016

E. G. Cerami et al. *Nucleic Acids Research*, 2011

# PIMKL benchmarking

PIMKL detects stable signatures across cohorts



Y. Cun et al. *BMC bioinformatics*, 2012

A. Liberzon et al. *Cell systems*, 2015

M. Kanehisa et al. *Nucleic Acids Research*, 2016

E. G. Cerami et al. *Nucleic Acids Research*, 2011

# Roadmap

Molecular data classification

Pathway-Induced Multiple Kernel Learning (PIMKL)

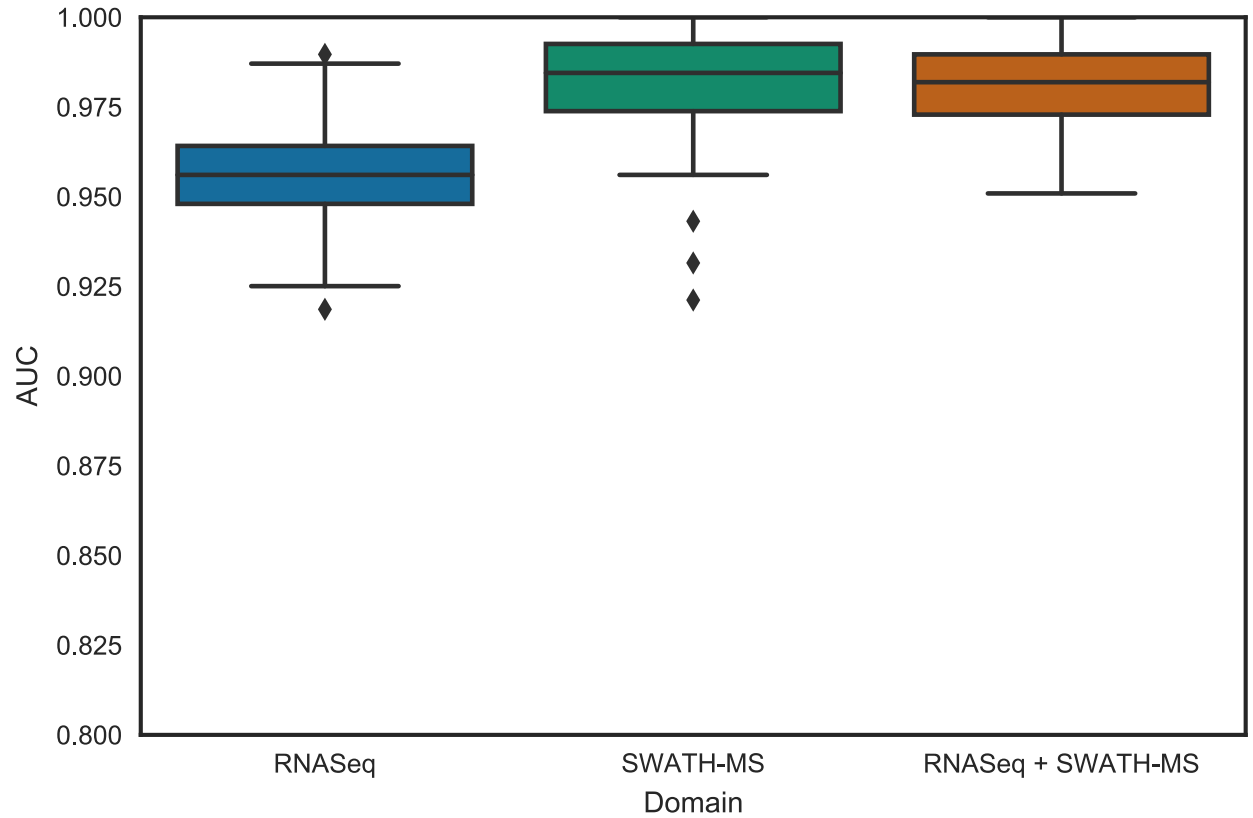
PIMKL benchmarking

PIMKL application

# PIMKL application - prostate cancer

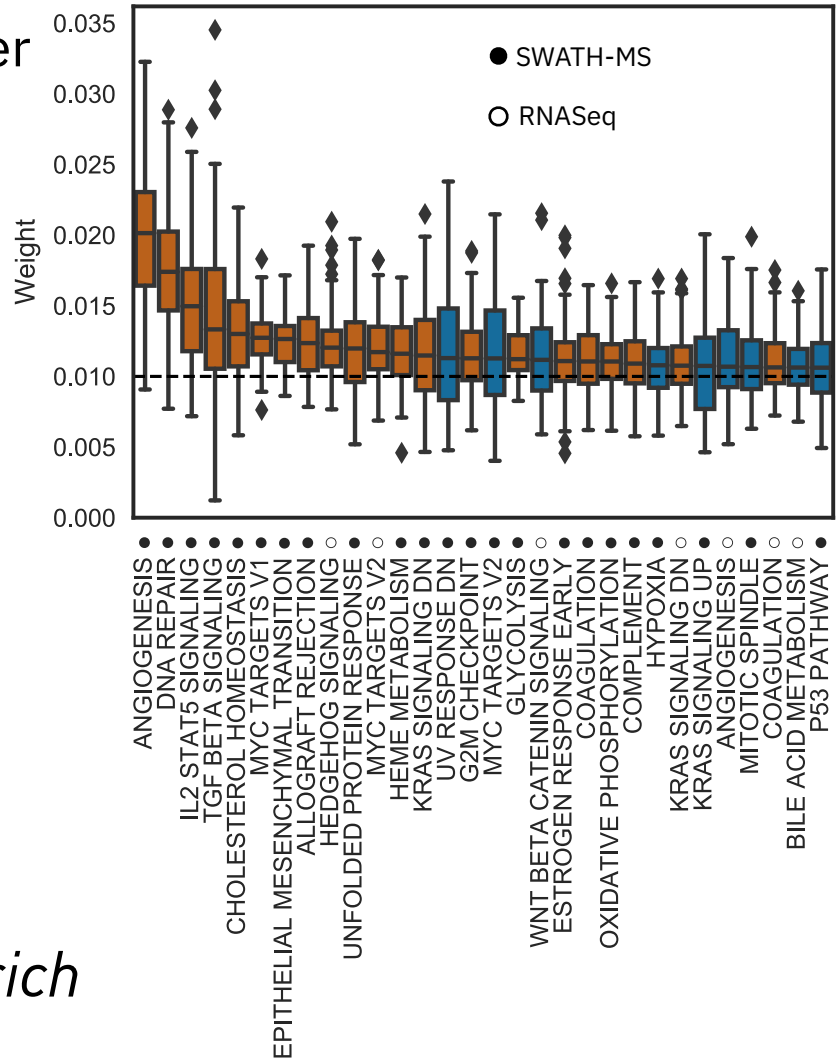
PIMKL detects tumor samples accurately

PIMKL integrates multiple omics seamlessly



# PIMKL application - prostate cancer

PIMKL identifies relevant pathways for each data type

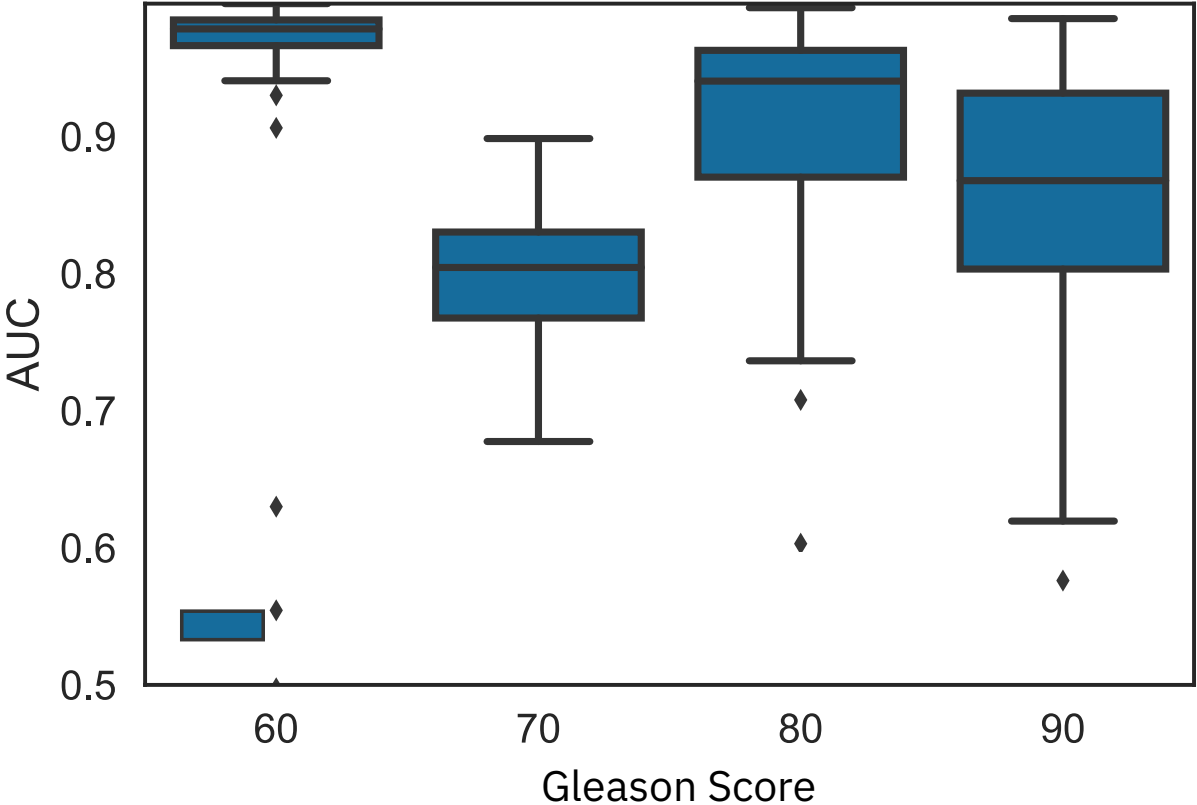




# PIMKL application - prostate cancer

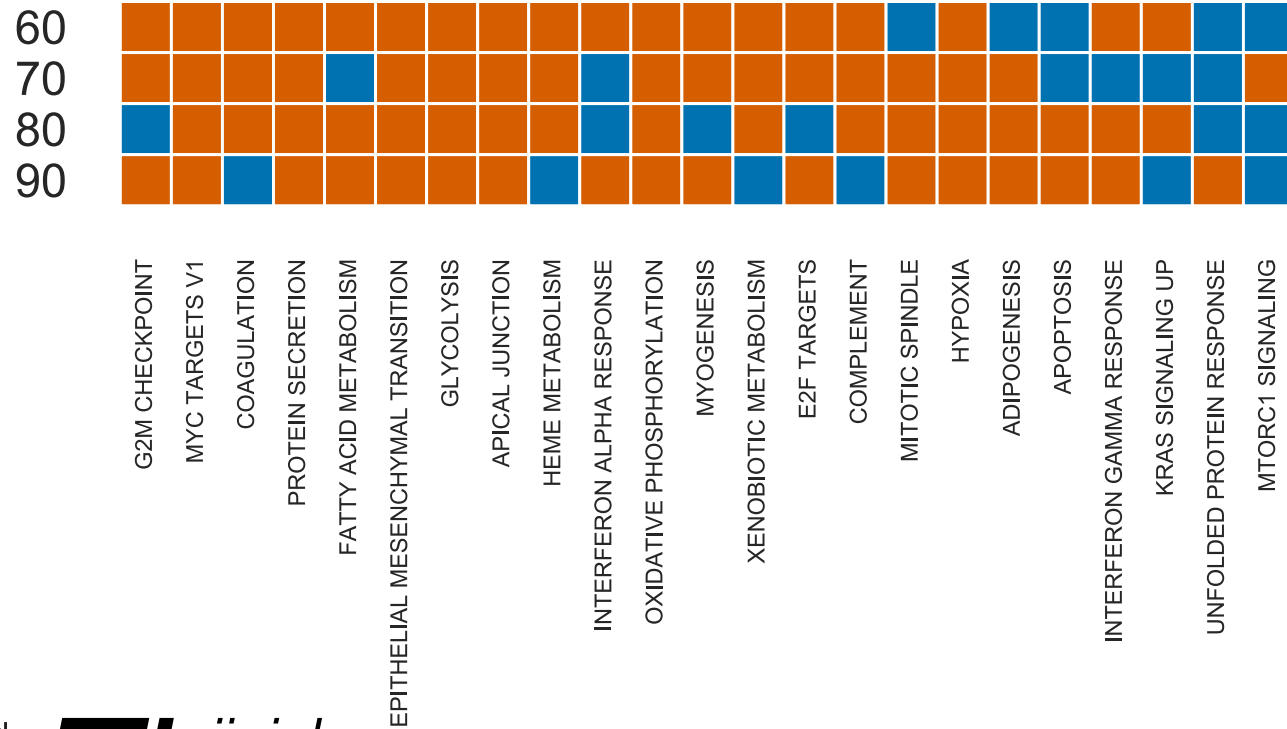
PIMKL stratifies different disease grades

PIMKL reaches high performance by training on 30 samples per class



# PIMKL application - prostate cancer

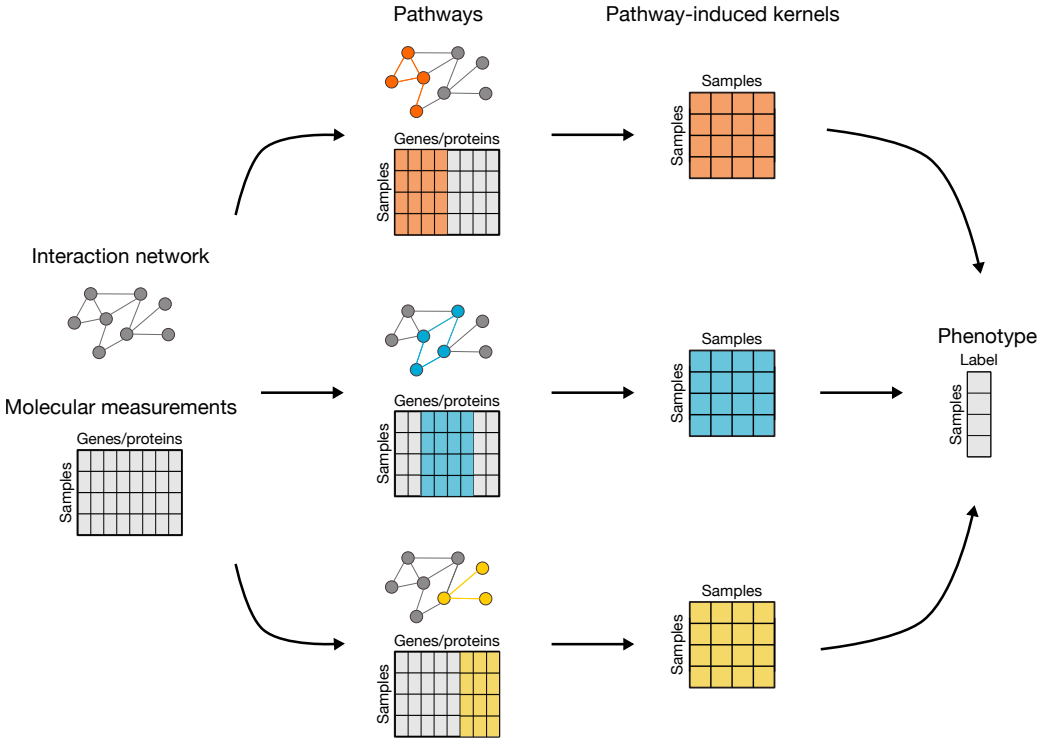
PIMKL helps to find the differences in the active pathways for each grade



# PIMKL - pathway-induced multiple kernel learning



Available on IBM Cloud



<https://ibm.biz/pimkl-aas>

# Thanks for your attention

