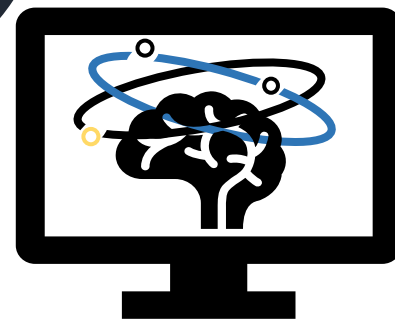
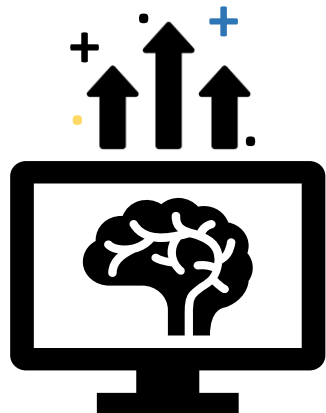


# Can active learning improve the performance of computational mass spectrometry?

Wei-Chieh Wang

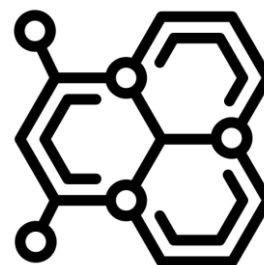
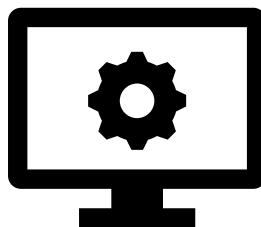
[Wei-chieh.wang@mmk.su.se](mailto:Wei-chieh.wang@mmk.su.se)



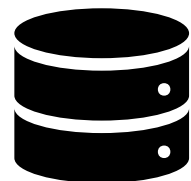


# Active learning workflow

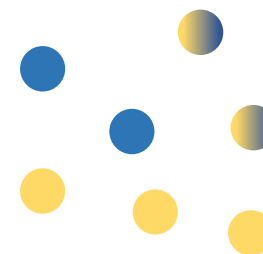
Train a model



Predict properties



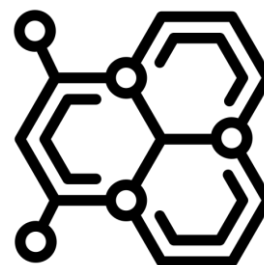
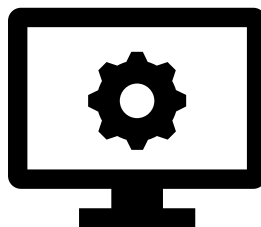
Collect training dataset



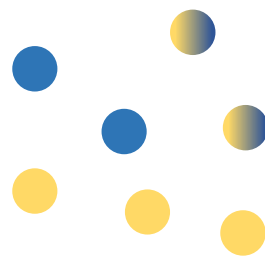
Obtain new data

# Active learning workflow

Train a model



Predict properties

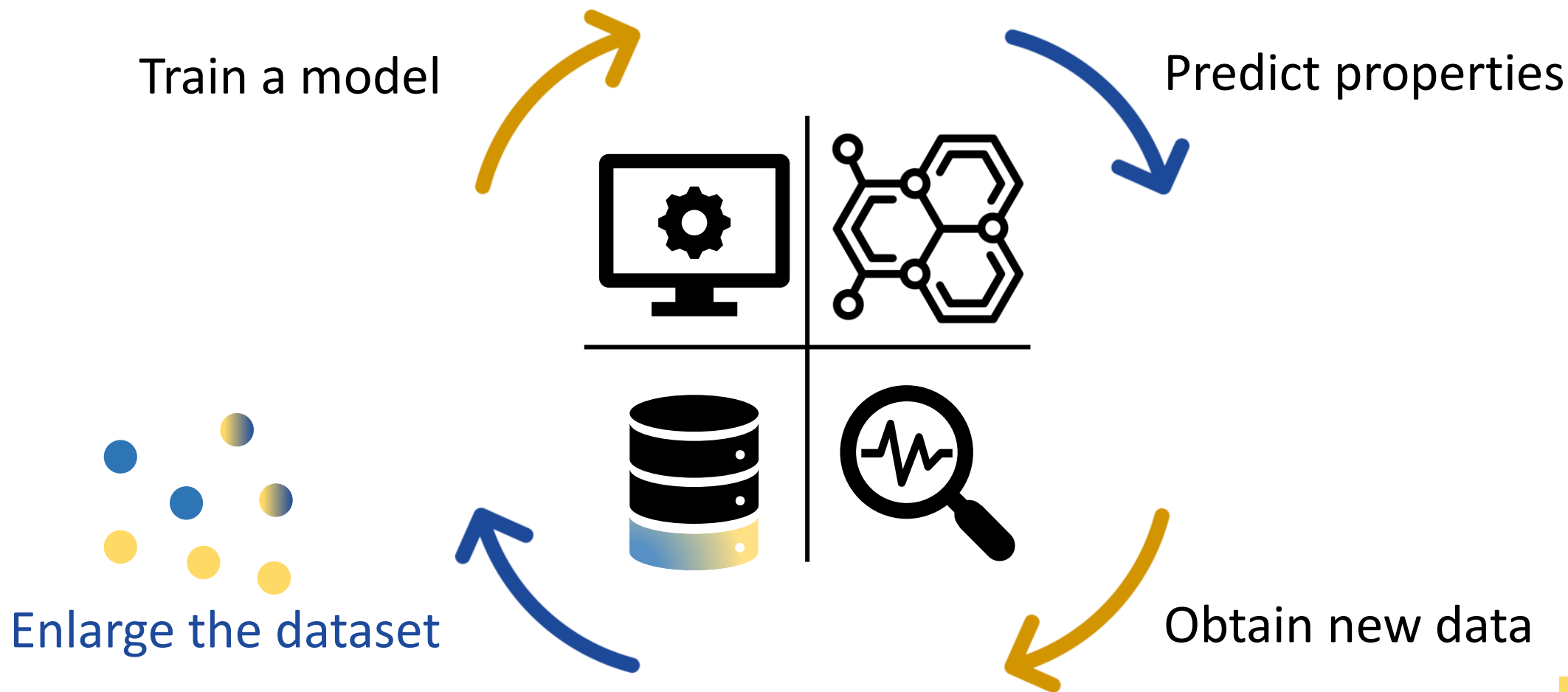


Enlarge the dataset



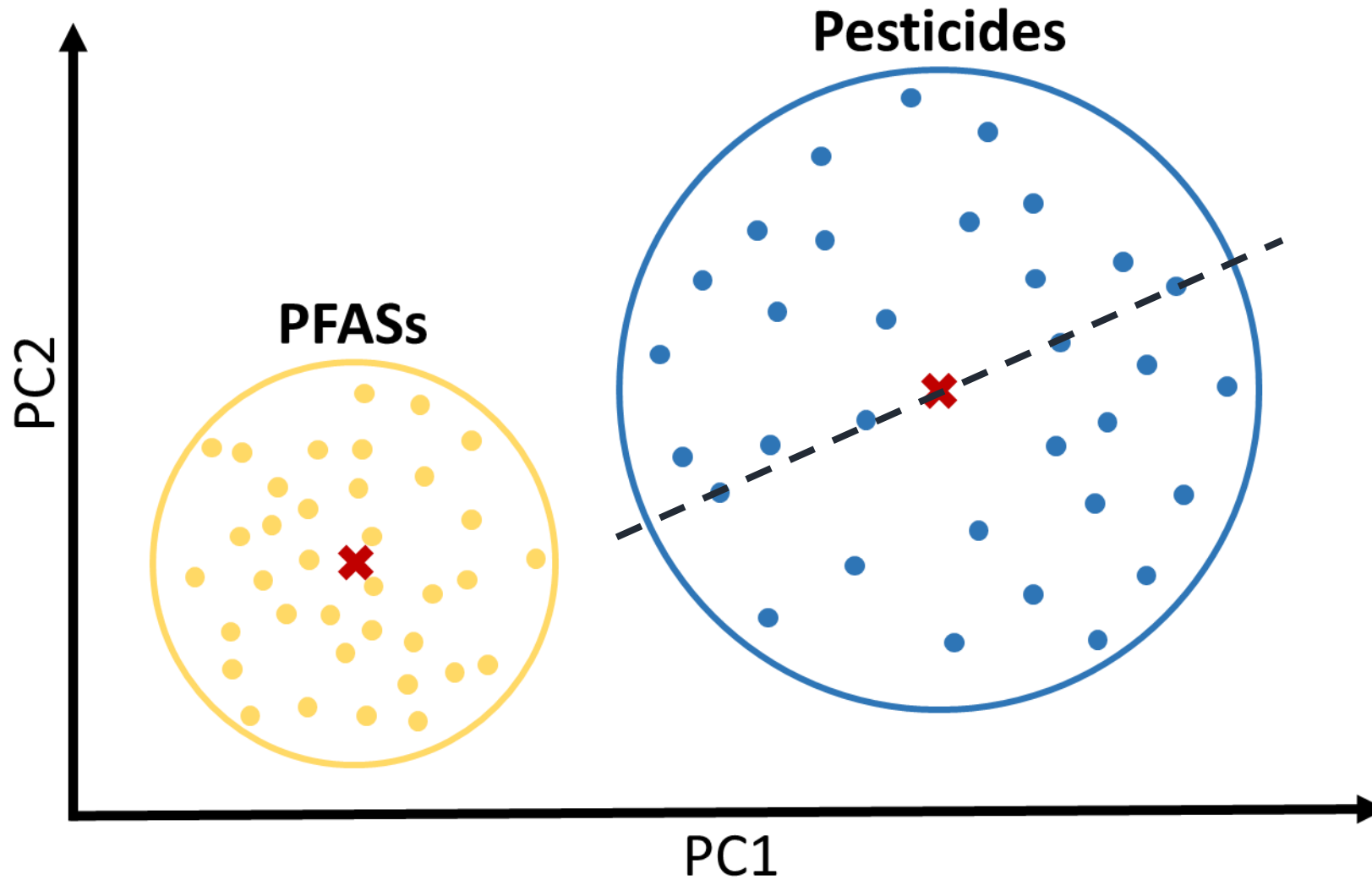
Obtain new data

# Active learning workflow

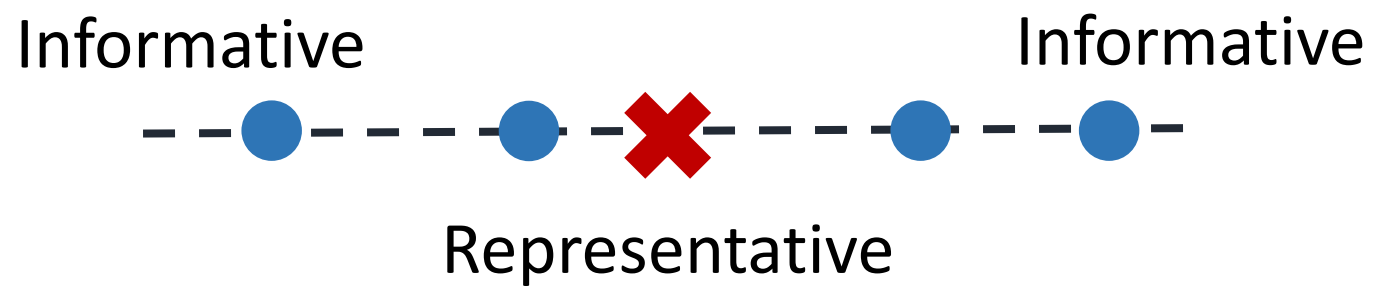
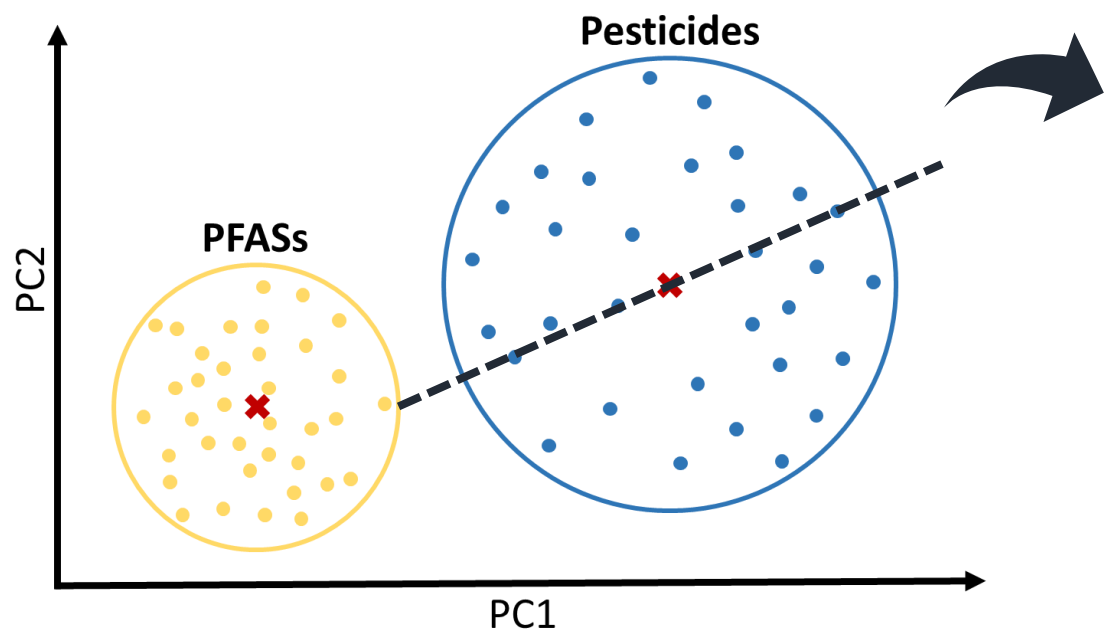




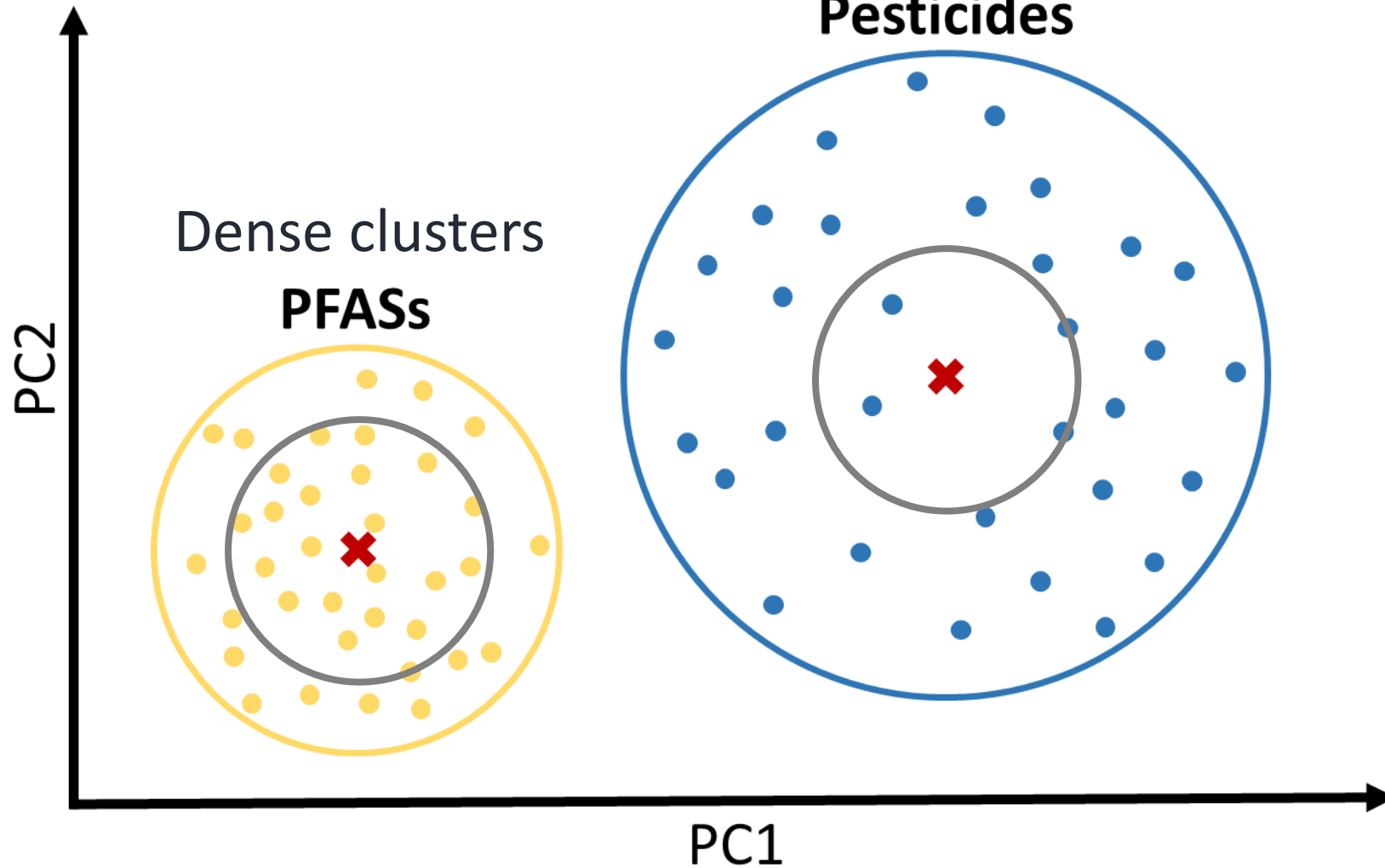
# Data properties



# Data properties



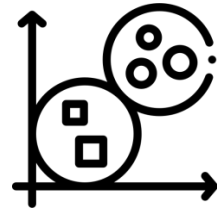
# Cluster properties





# Cluster and uncertainty

---



Cluster density

Dense clusters

$\text{Density} \leq \text{Mean}(\text{Density})$

 Representative

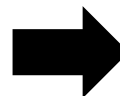
Sparse clusters

$\text{Density} > \text{Mean}(\text{Density})$

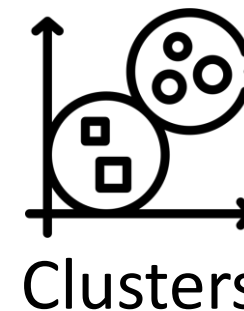
 Informative

Reducing the density threshold after each iteration

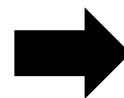
# Proposed active learning strategy



Cluster the compounds



Dense clusters



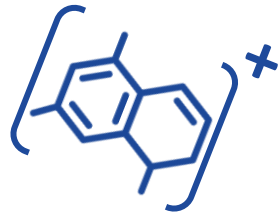
 Representative compounds

Sparse clusters

  
 Informative compounds

Reducing the density threshold after each iteration

# Case study



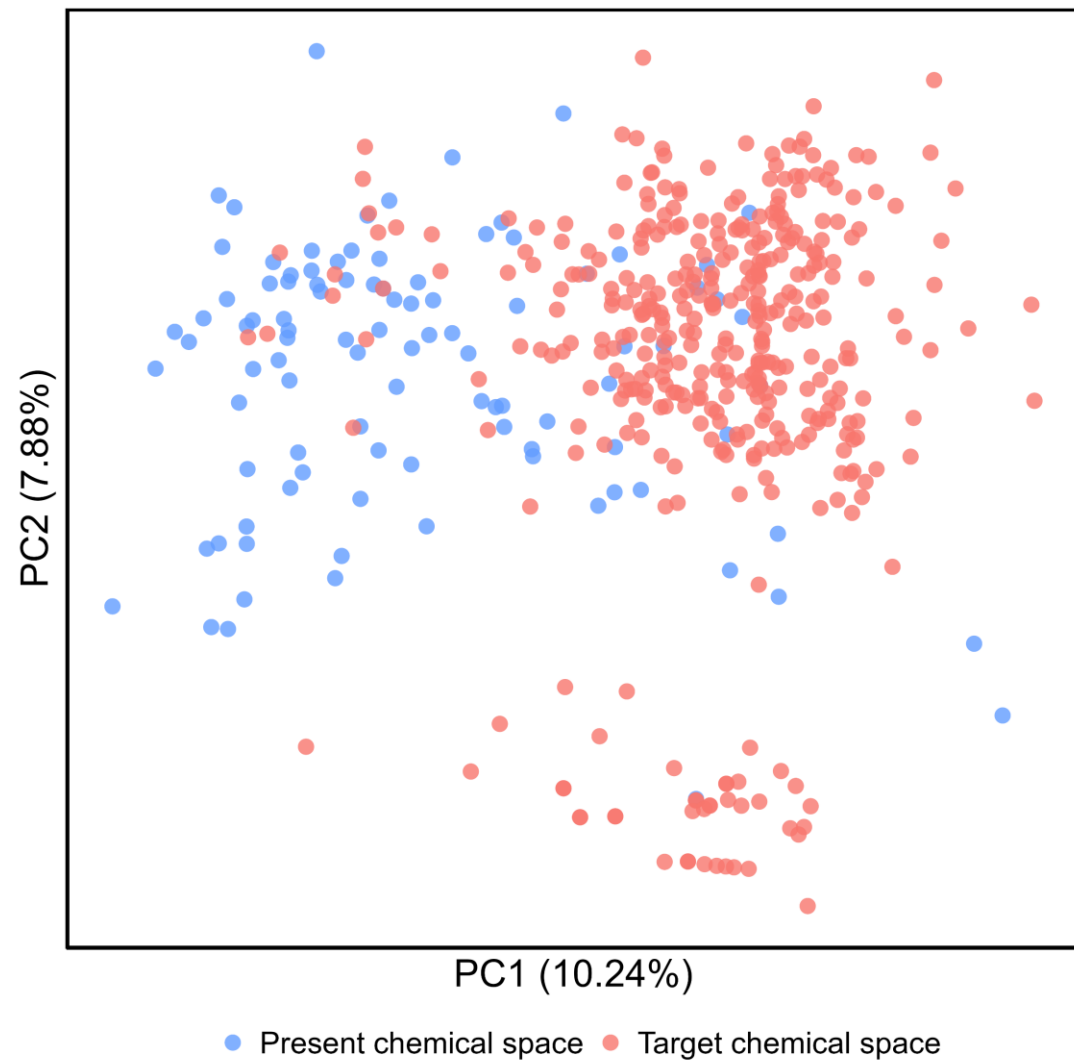
Ionization efficiency prediction model

100 known compounds, 400 new suspect compounds

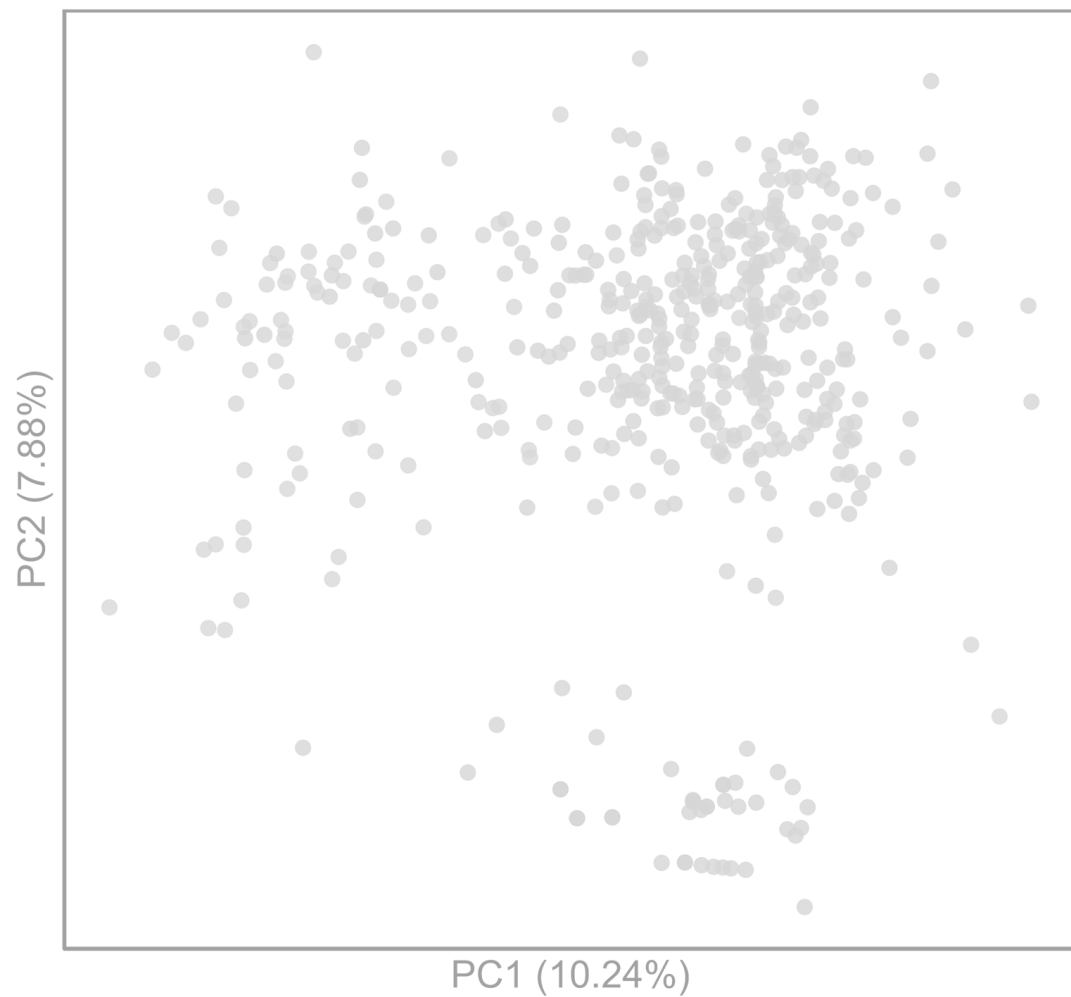
Obtain 15 new compounds in each iteration

Implement 8 iterations

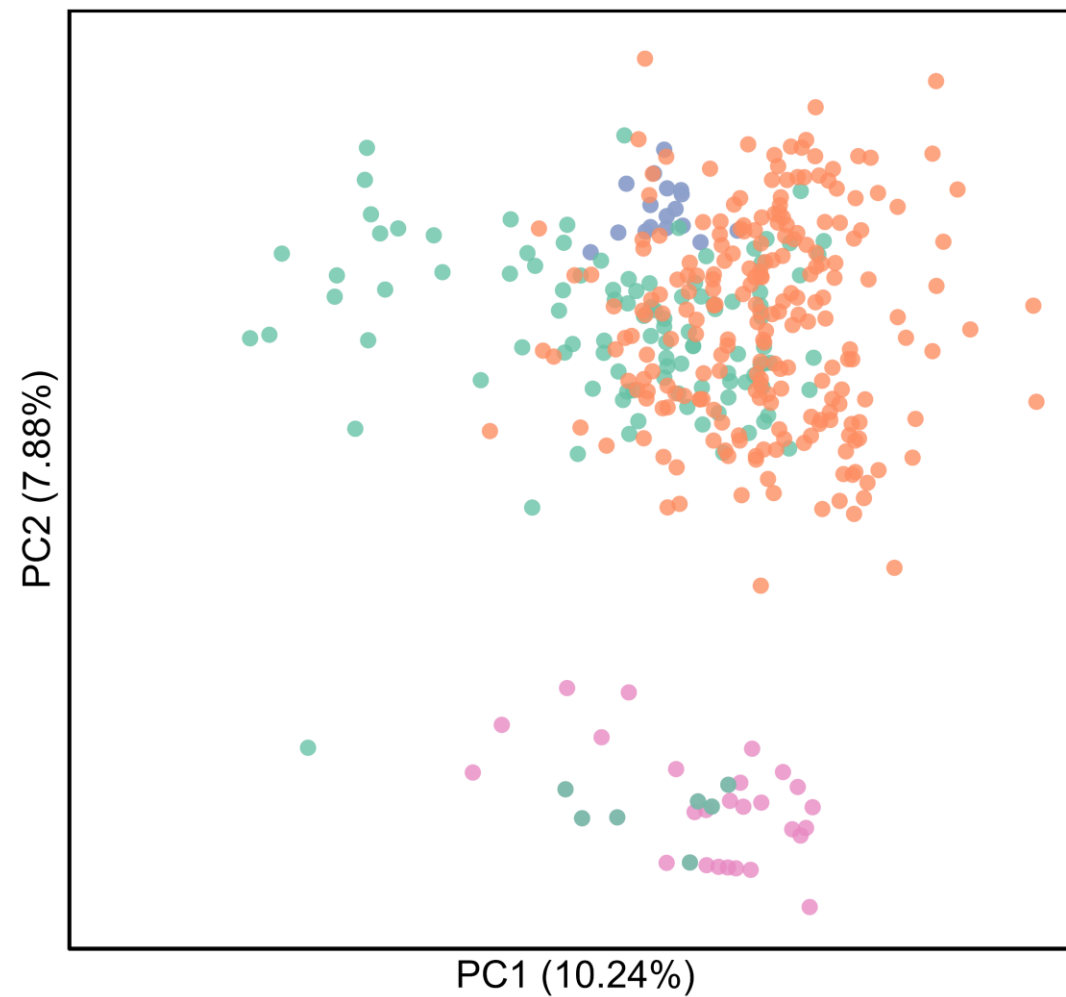
# Chemical space



# Chemical space

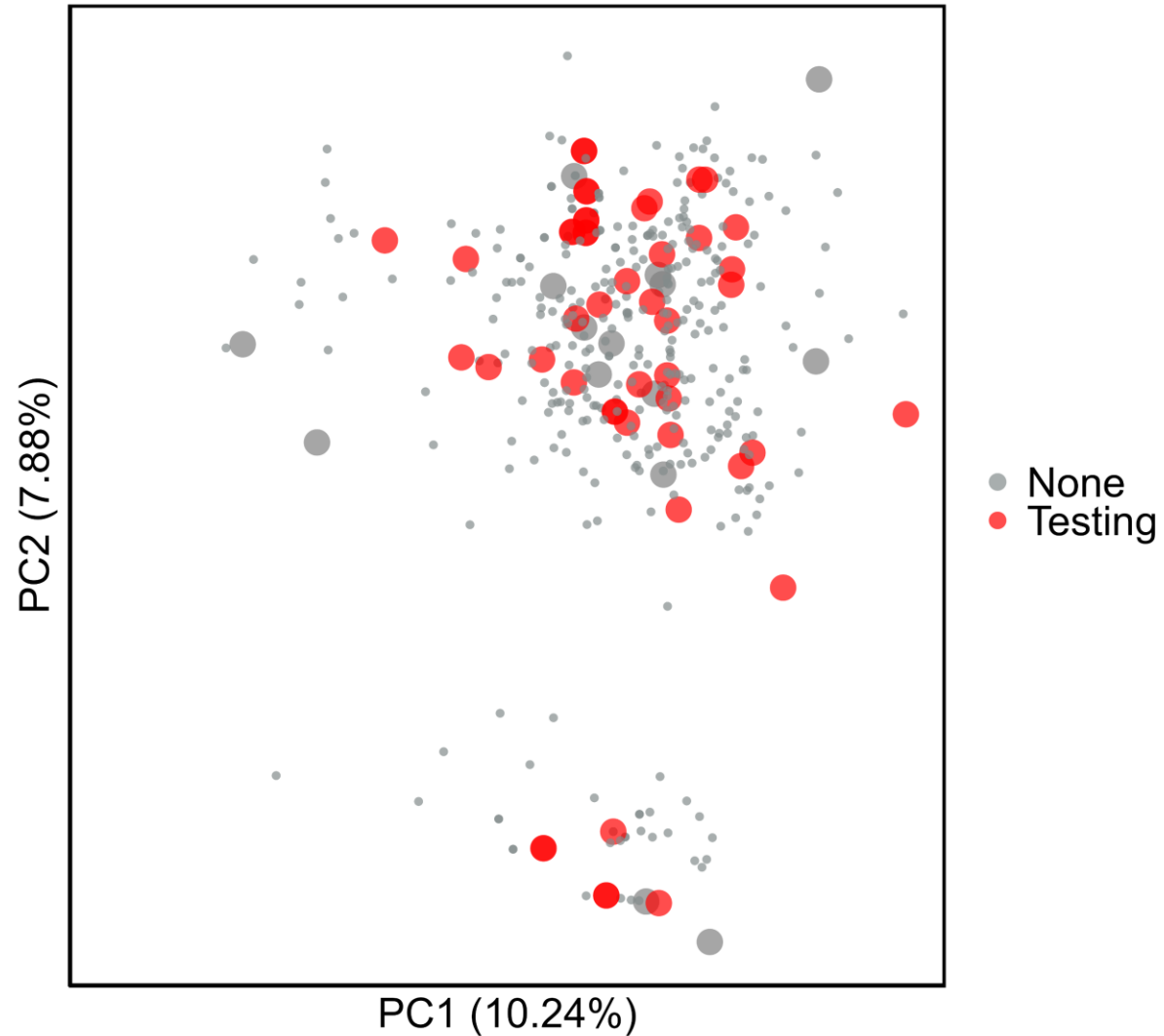


● Present chemical space ● Target chemical space

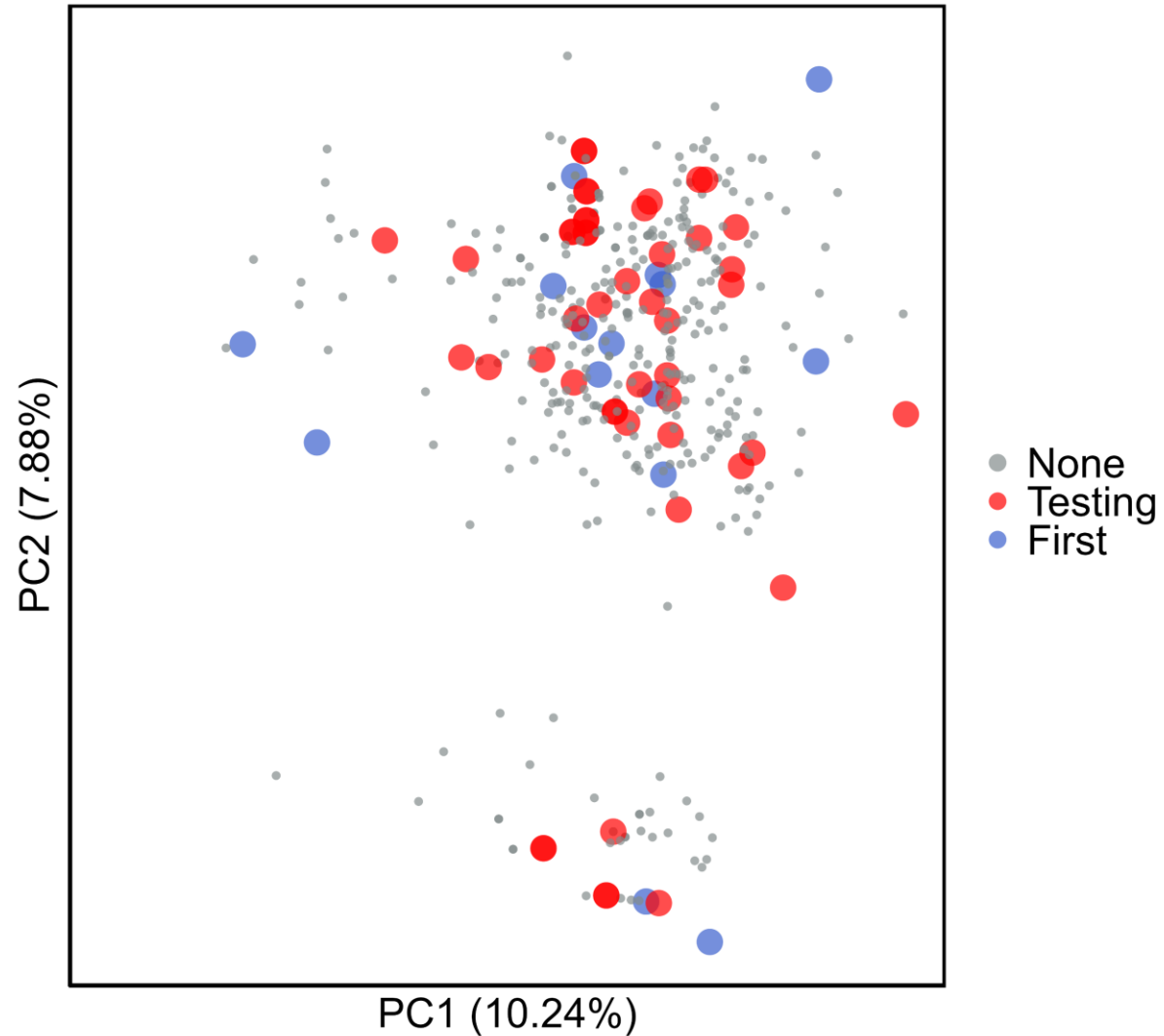


● Environmental contaminant ● Natural product ● PCB ● PFAS

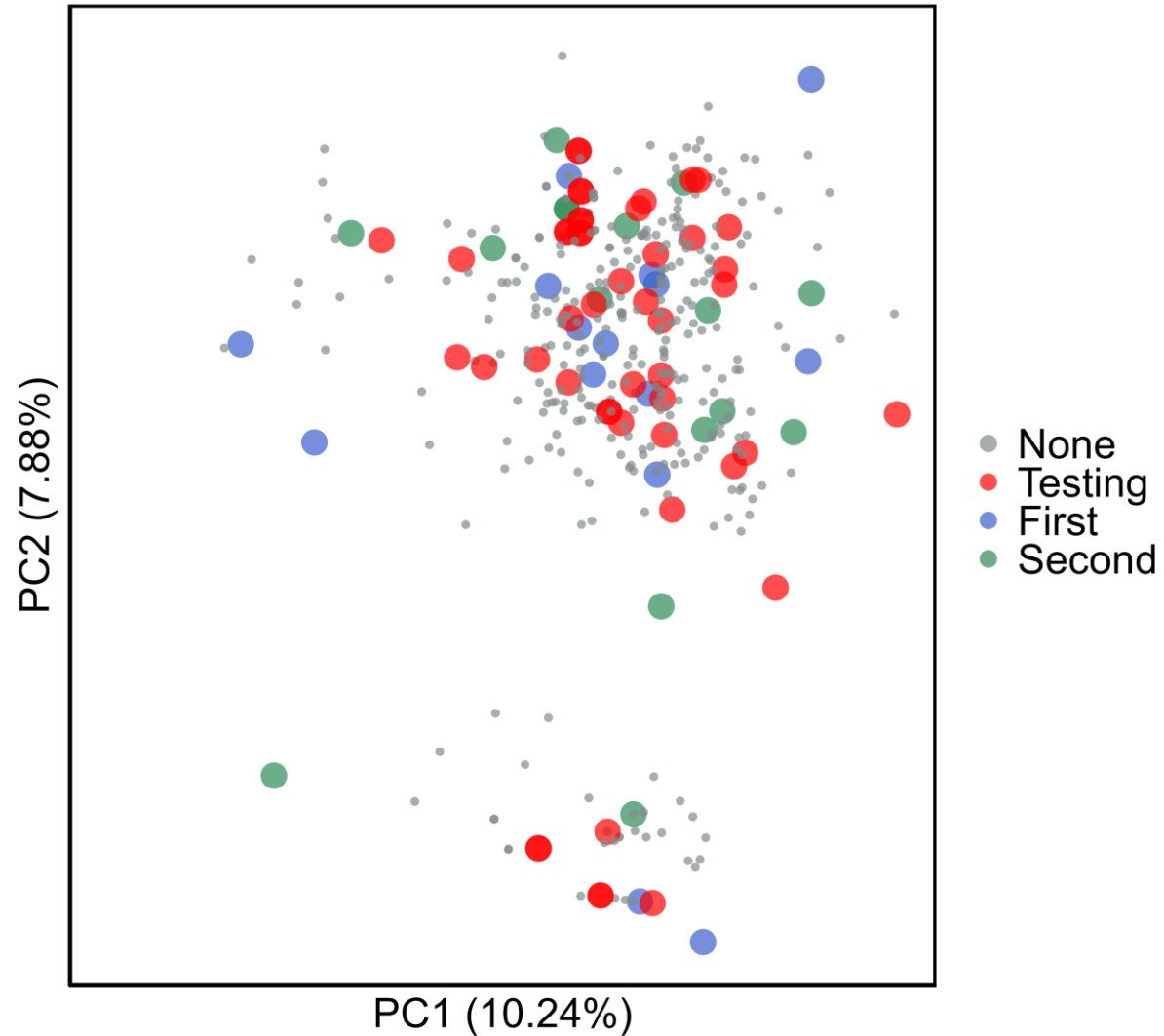
# Performance evaluation



# Performance evaluation

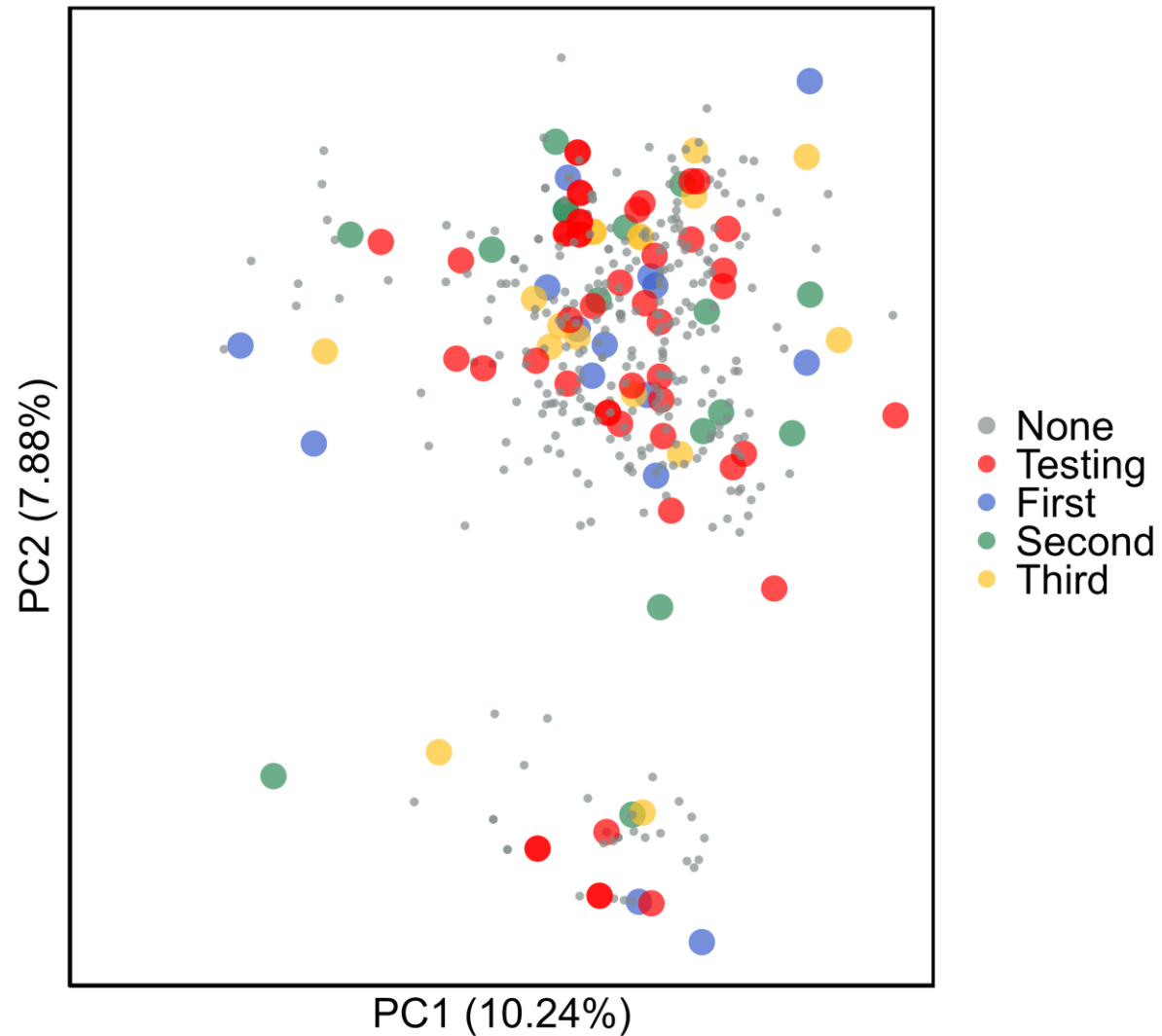


# Performance evaluation

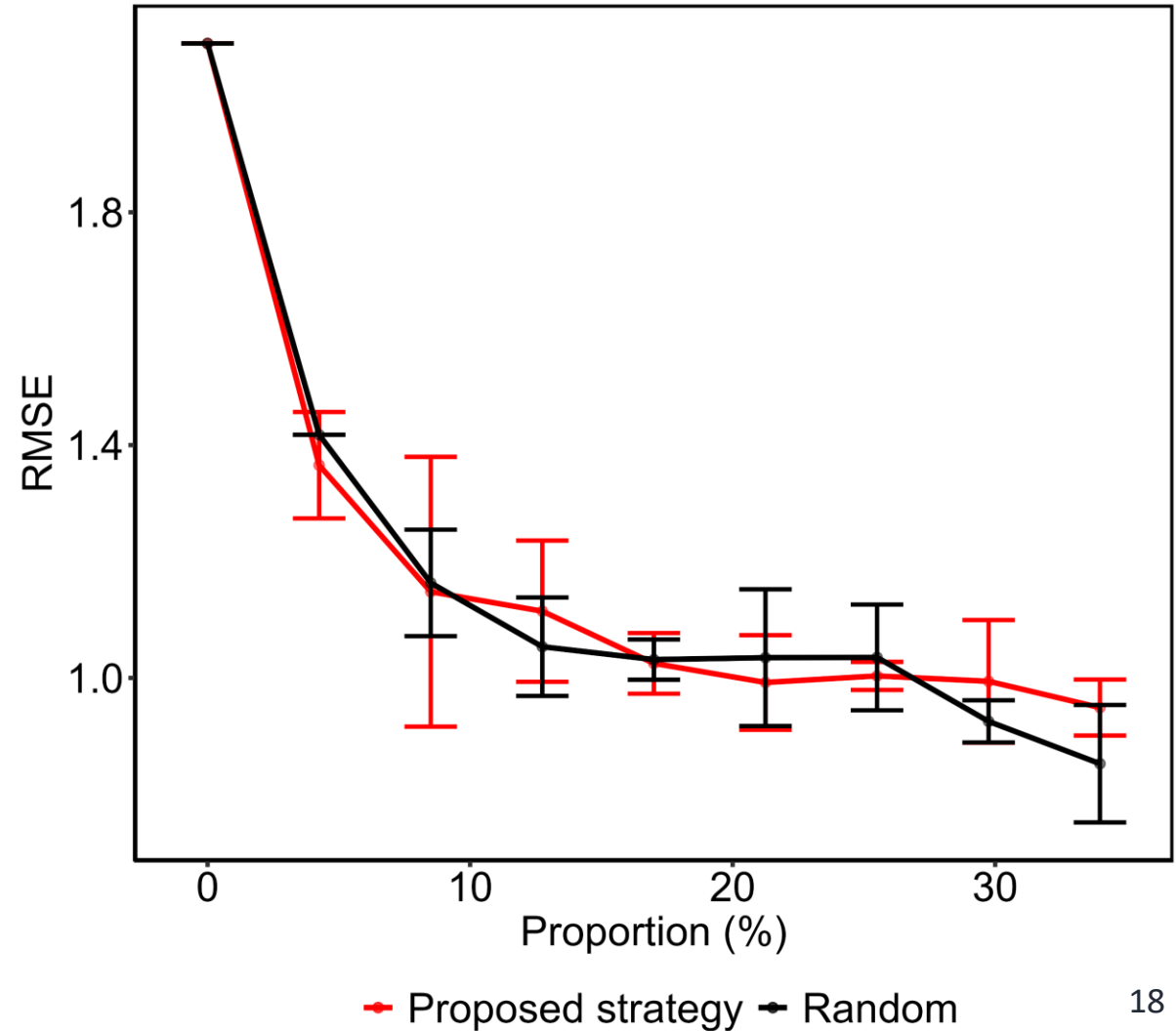
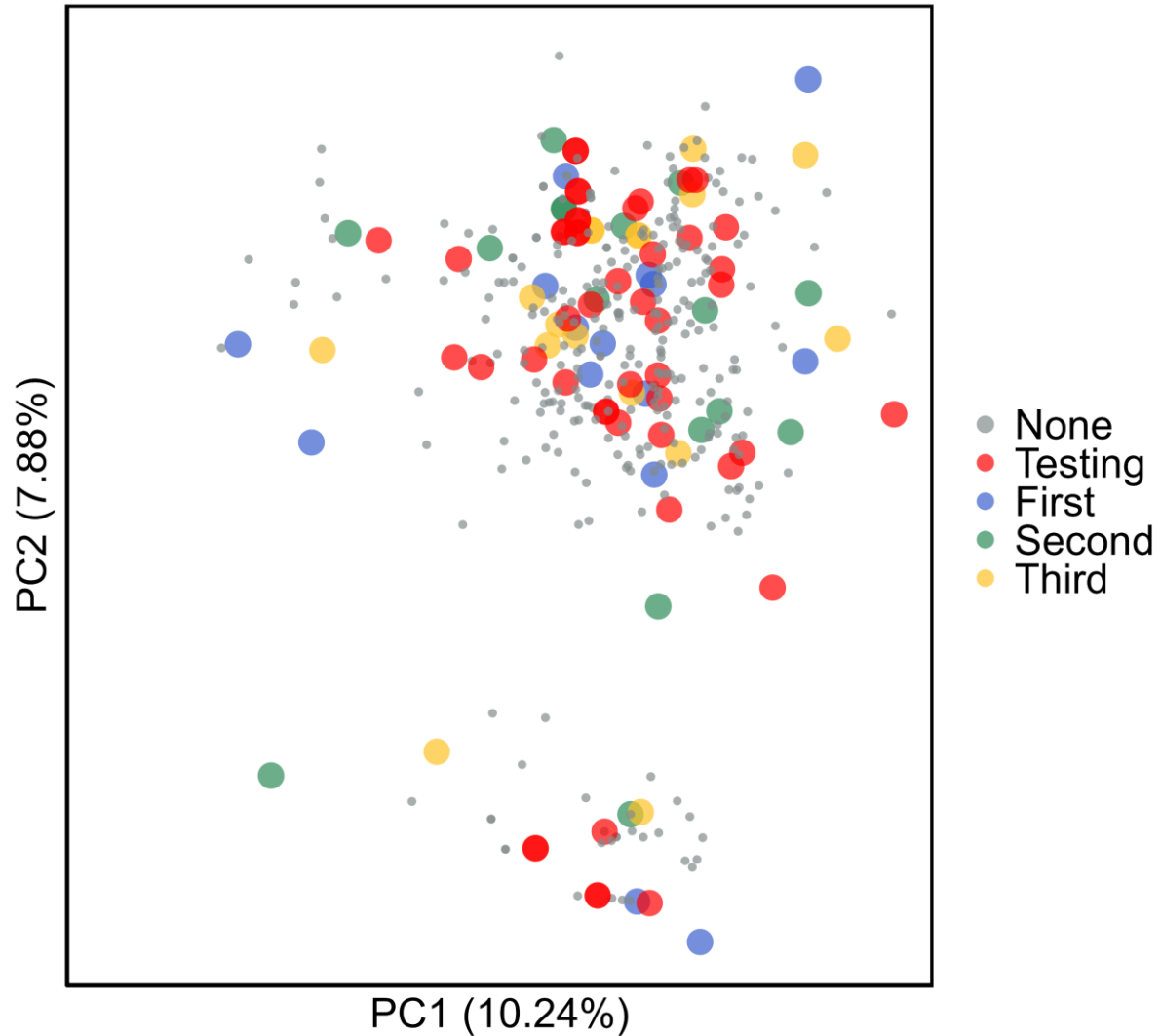




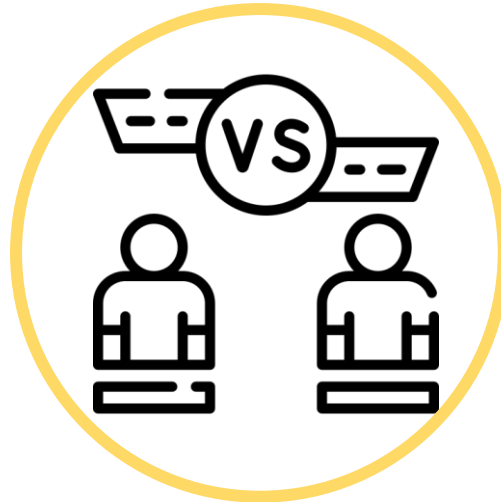
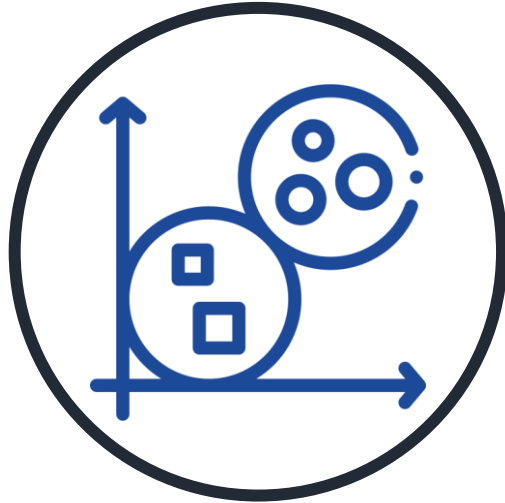
# Performance evaluation



# Performance evaluation



# Conclusions & Future works



# Thanks for listening.

## Let's make our ML models powerful by exploring new chemical space!

### Acknowledgment

Anneli Kruve, Meelis Kull

Helen Sepman, Caroline Huber, Nahid Amini,

Jaanus Liigand, Sara Khabazbashi

Kruvelab members

Wei-Chieh Wang

[wei-chieh.wang@mmk.su.se](mailto:wei-chieh.wang@mmk.su.se)



Stockholm  
University

The project is supported by the Swedish Research Council.

*Kruve lab*