

# Sleep signature project

Goals:

- I) Demonstrate sleep state is encoded in gene expression matrix
- II) Find subset of genes encoding this state (~feature selection)

# 1) Decode state from gene expression

## Data generation:

Monitor animals with ethoscope platform > sample @ diff ZT and cond > SCS > analysis with Kris Davie's scanpy pipeline (documented on his notebook) > .h5ad import

## Proposal:

Train MLP with different architectures on the same datasets:

6 different clusters: y, ab, abp, AST, EG, SG

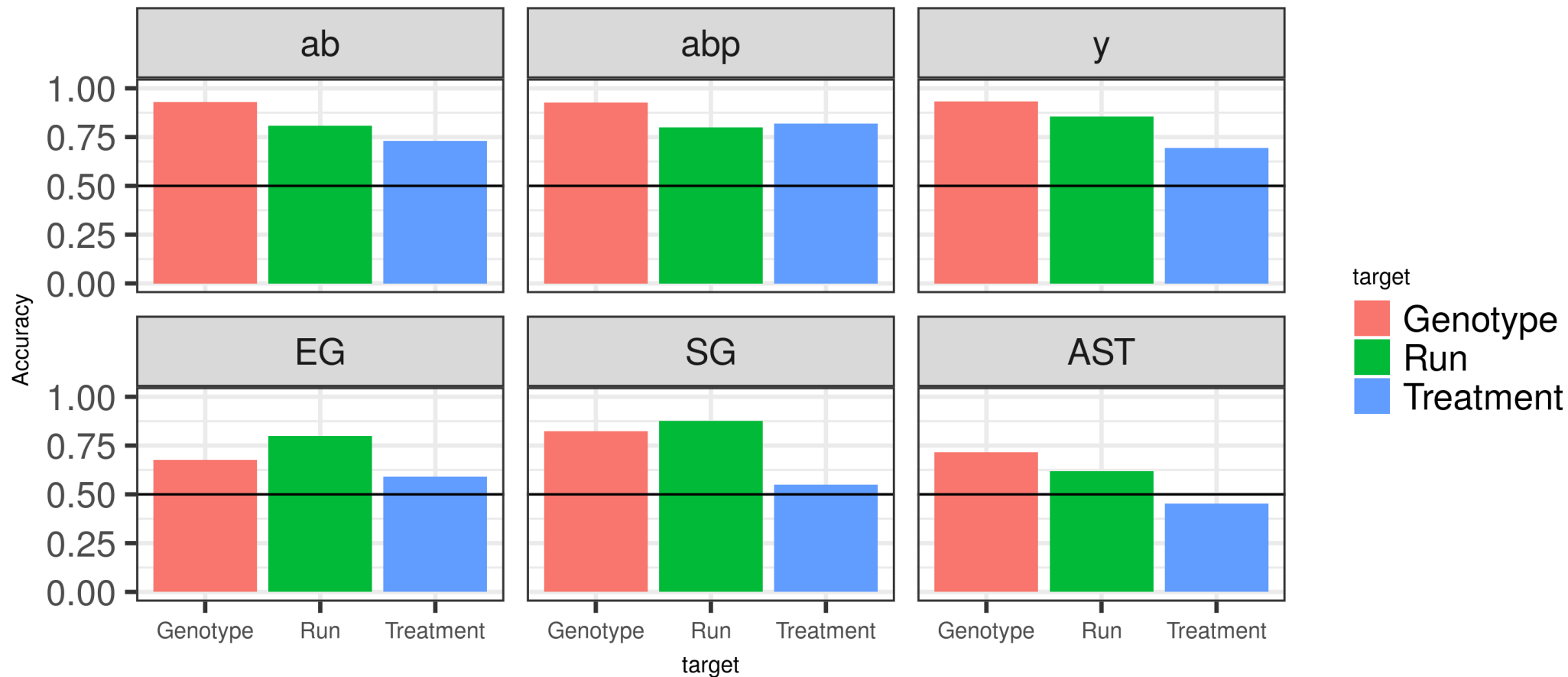
Consider:

- \* 1 timepoint (ZT20) → remove circadian effects
- \* 2 conditions (sleep vs sleep deprivation) → simplicity
- \* 2 DGRP lines (303 and 88) → minimise DGRP effect

# 1) Decode state from gene expression

ARCH:200-100-10

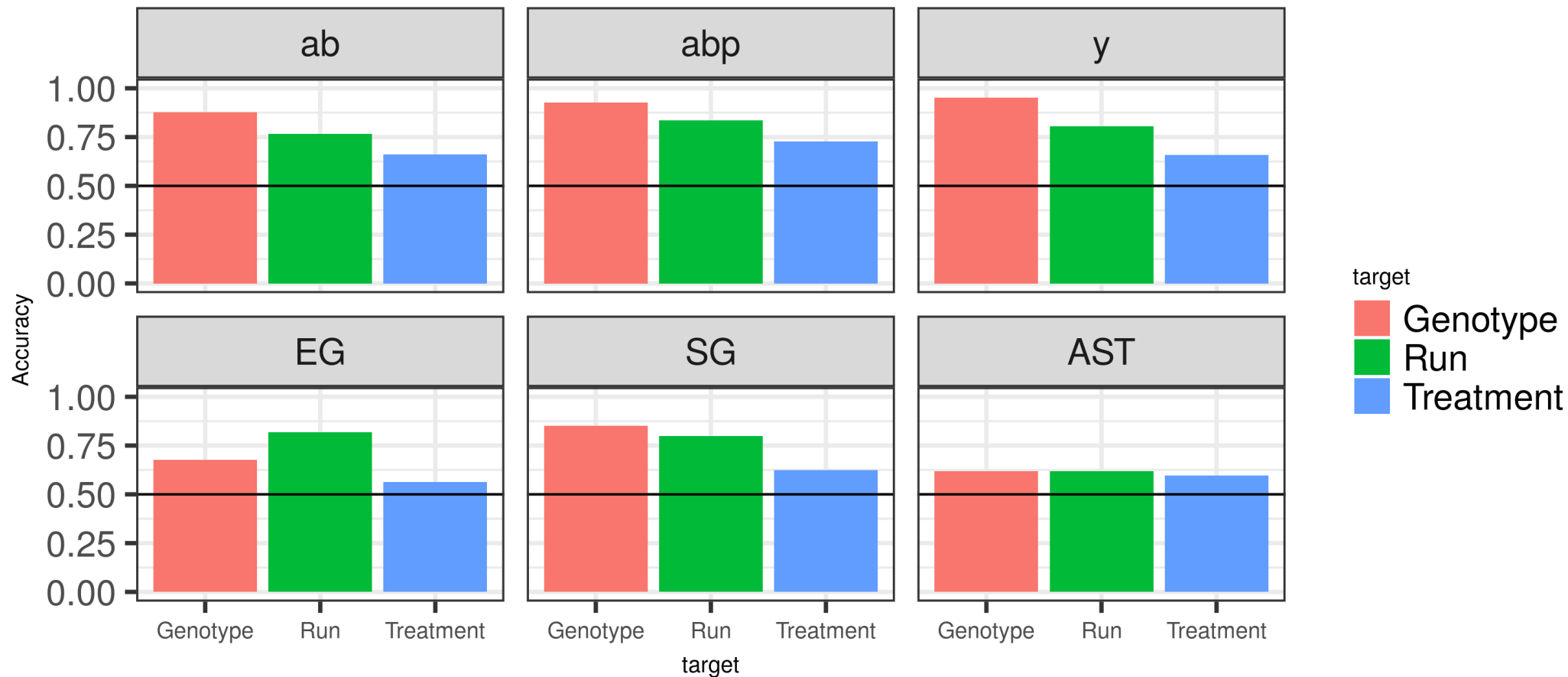
sklearn.neural\_network.MLPClassifier defaults



# 1) Decode state from gene expression

ARCH:50-25-5

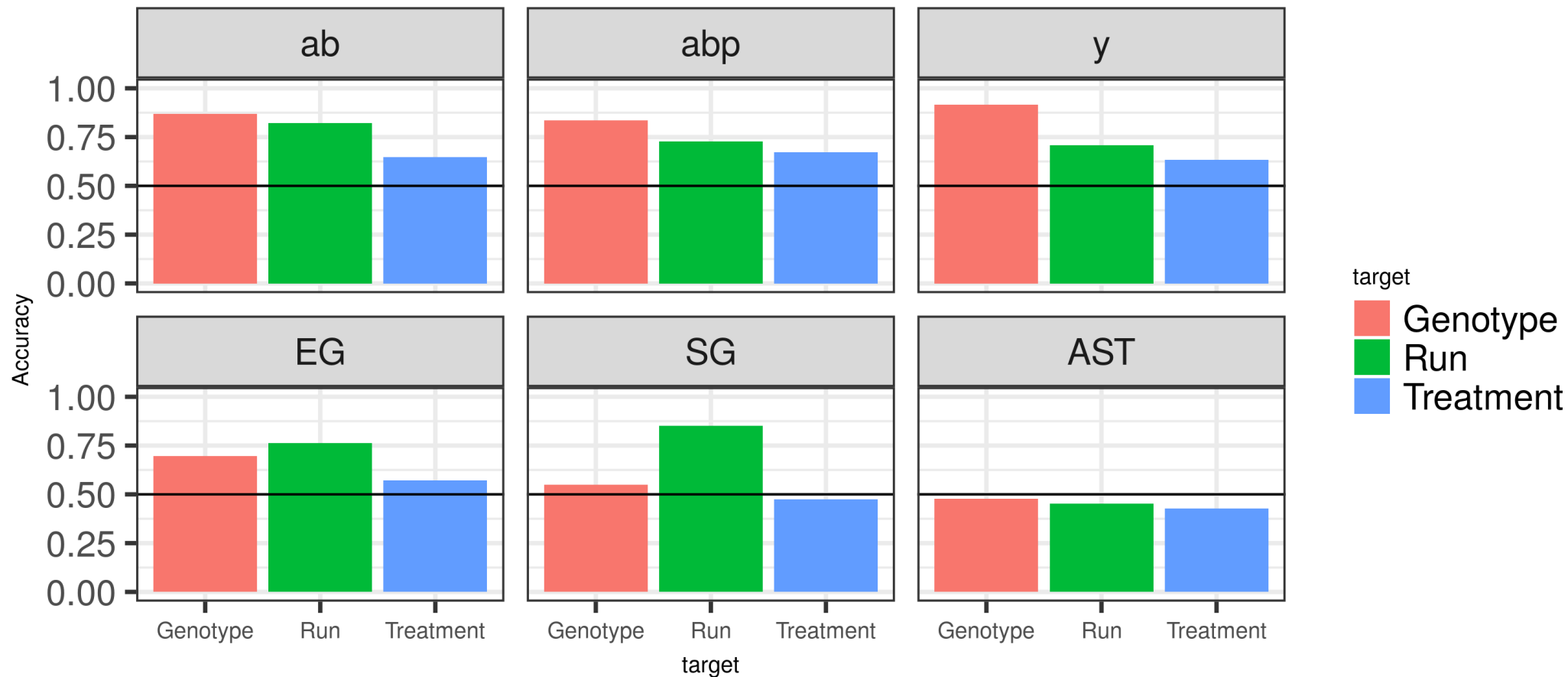
sklearn.neural\_network.MLPClassifier defaults



# 1) Decode state from gene expression

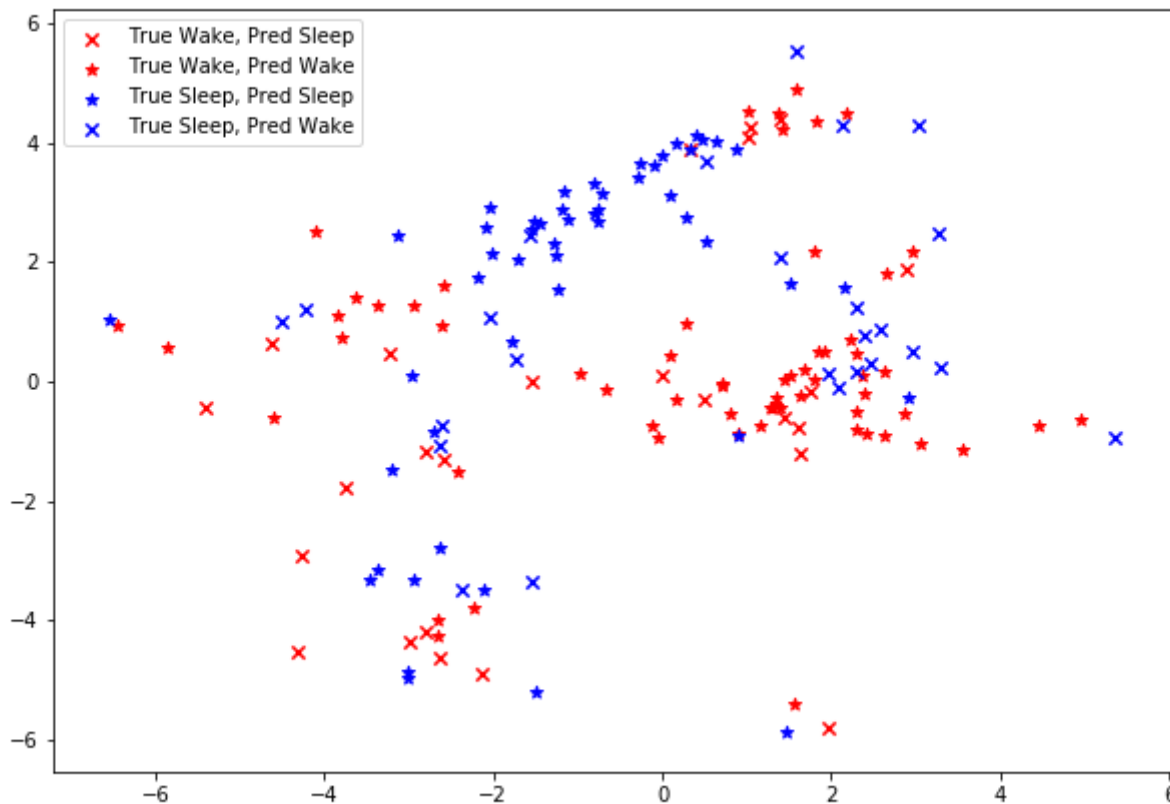
ARCH:3-2-1

sklearn.neural\_network.MLPClassifier defaults



# 1) Decode state from gene expression

1320 genes, y KCs, 69.5 % accuracy



# Conclusions

- \* State partially predictable
- \* Predictable on some clusters, not others (!)
- \* Technical effect very predictable with simple model

# Conclusions

- \* Remove technical components?
  - SVD, ...?
- \* Is state still predictable after that?
  - Evidence of state encoding
- \* Better models? Autoencoder, etc



## 2) Find subset of genes

- \* DEA: Wilcoxon rank test

- (already done by Joana)

- \* Feature selection of genes? Which genes support state prediction?

- *in silico* gene knockdown