Tobias Frisch – [frisch@imada.sdu.dk](mailto:frisch@imada.sdu.dk)

Average graph

find most varied average genes/highest variance

HiSeq - Variance of zero genes

Do they explain a PAM50 label well?

judge or dont remove

Zero means unmeasureable, not explicitly errors

If its only present in under 5 (or only present in a specific group) samples remove it

Remove by low variance?

Clean more explicitly

Judge better (informed decisions)

Outlier patients

Lower dimentionality for this ^

UMAP

T-SNE

PCA

For next week:

Feature selection

If so, how

Clean data better

Present plan/next steps

Find better classification (interpretable) vs PAM50