

FINAL PRESENTATION

**The role of tissue-specific antigens
in different cancer entities**

KALLIKREIN GENES

Anouk Dupé, Dustin Schilling, David Eckey, Maria Yemane

Supervisor: Dr. Maria Dinkelacker
Tutor: Nils Mechtel

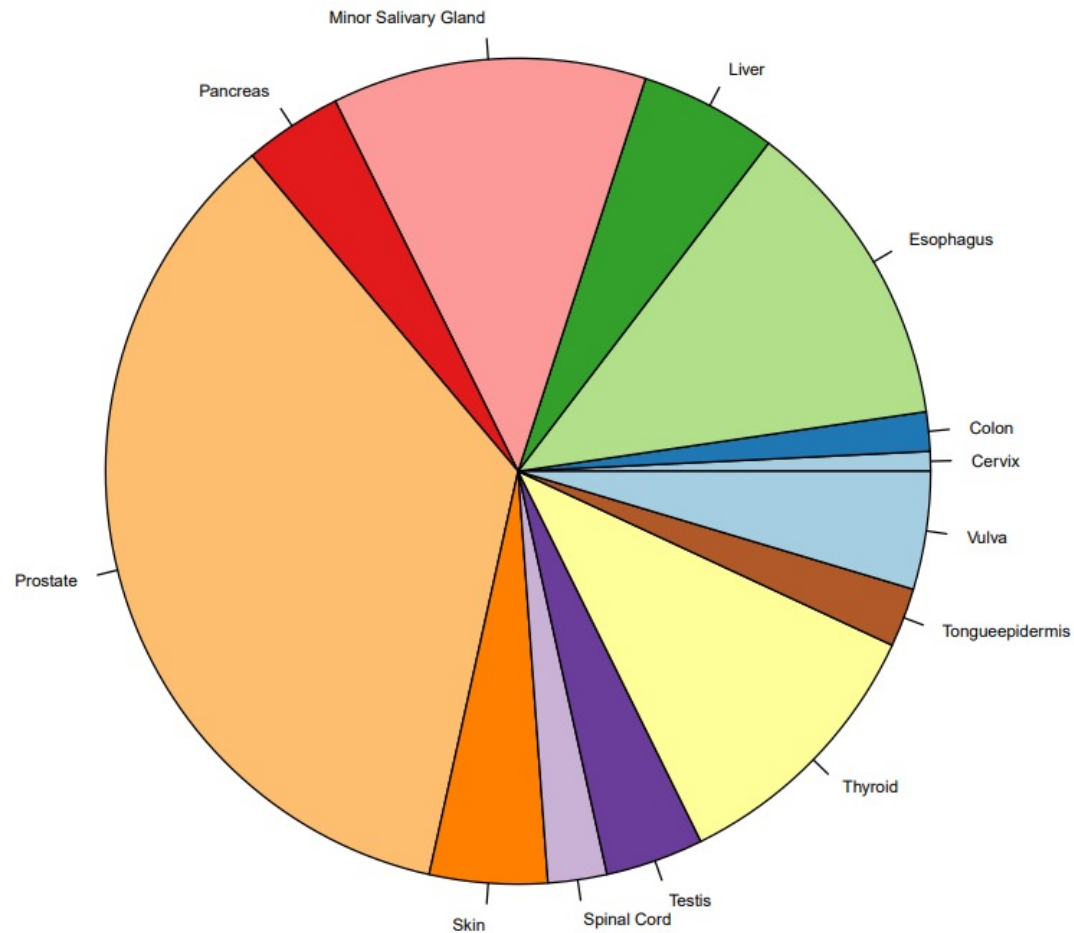
21.07.2021

Biological background and aim of the project

- 🧬 tissue restricted antigens (TRAs) show upregulation in cancer
- 🧬 TRAs can serve as biomarkers and/or potential drug targets
- 🧬 kallikrein genes (KLKs) are a group of serine proteases under steroid hormone regulation
 - involved in tissue remodelling
 - regulation blood pressure
 - skin desquamation
- 🧬 „Deviations in expression patterns of specific KLKs?“
 - potential in serving as biomarkers?

TRA distribution

Distribution of KLK-TRA genes over tissue



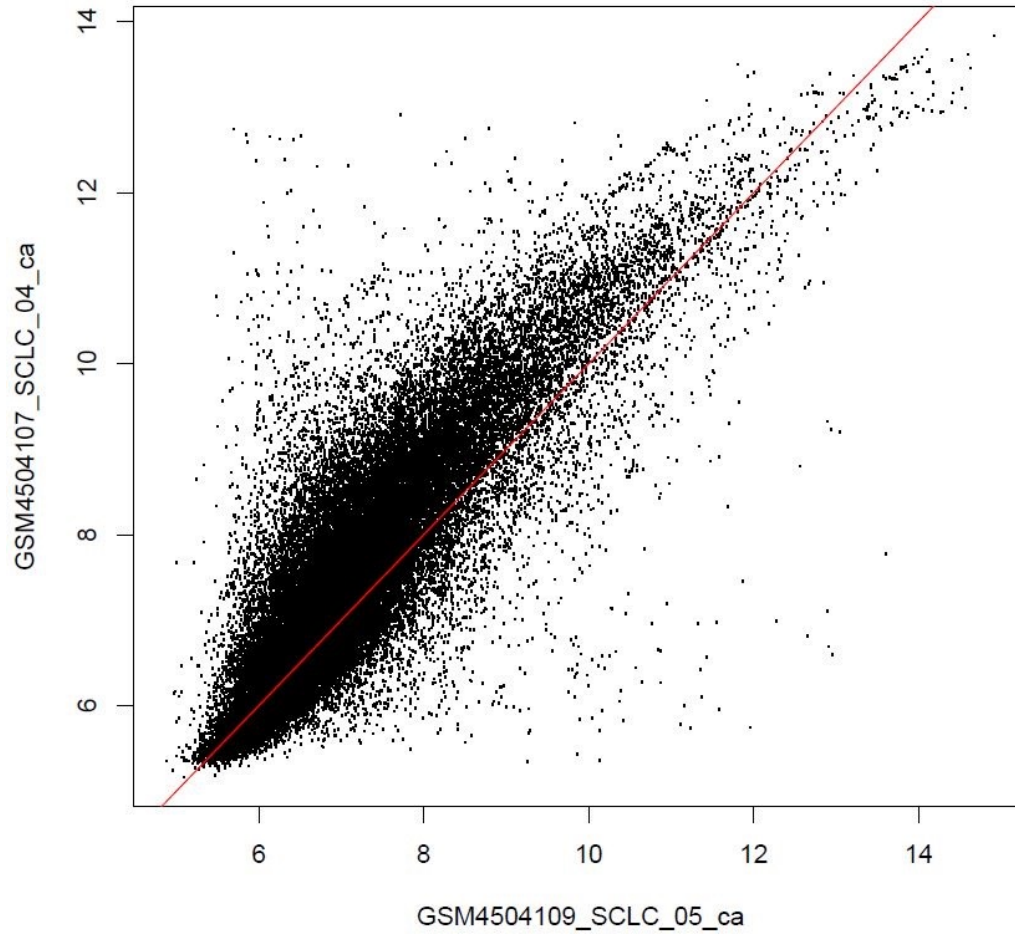
→ six TRA data sets

→ creation union KLK data set

→ removal identical entries

Quality control

scatterplot of probe GSM4504109_SCLC_05_ca and GSM4504107_SCLC_04_ca

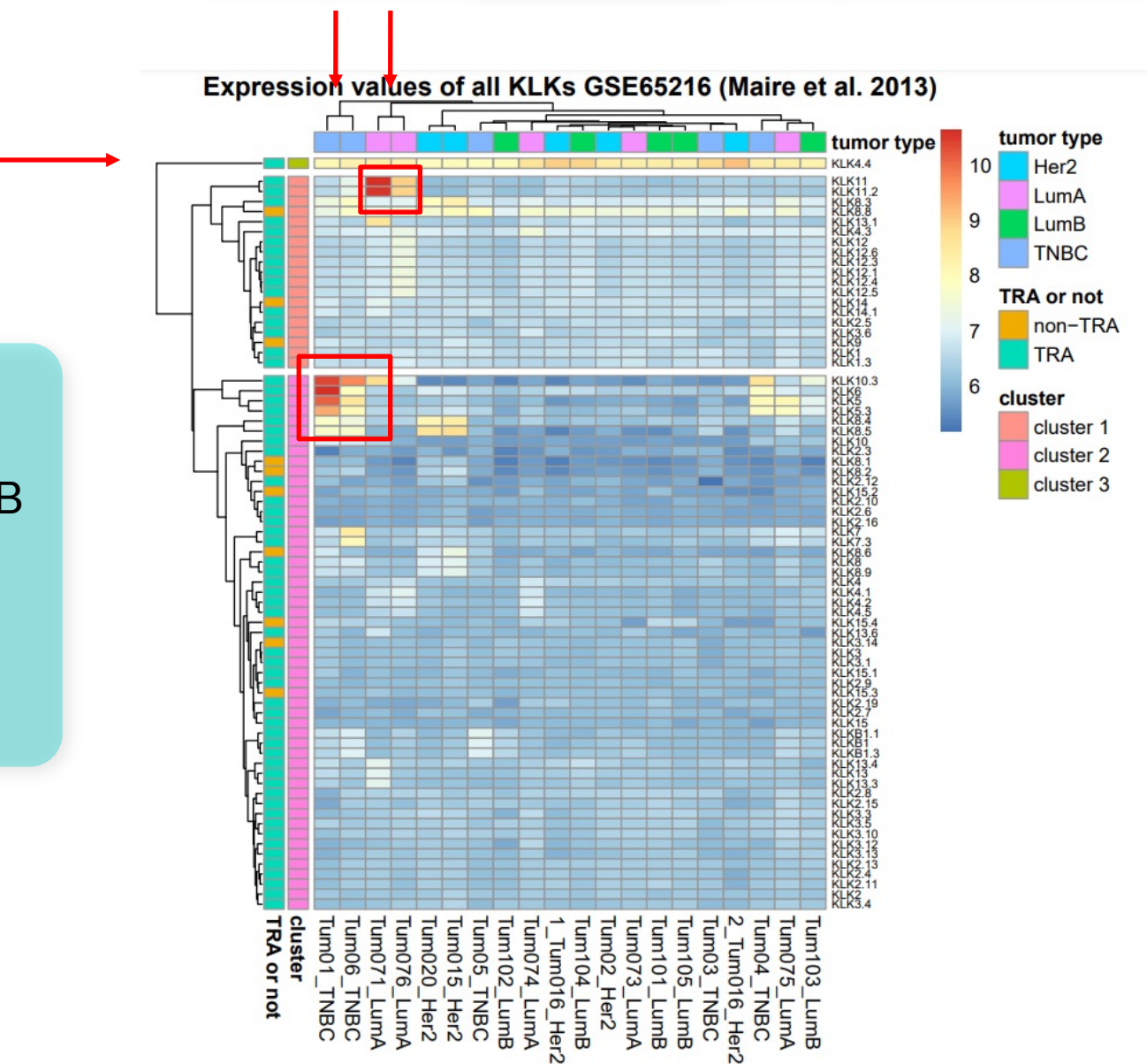


- scatterplot from lung cancer data set
- “banana form” detectable
- exchange of two chips

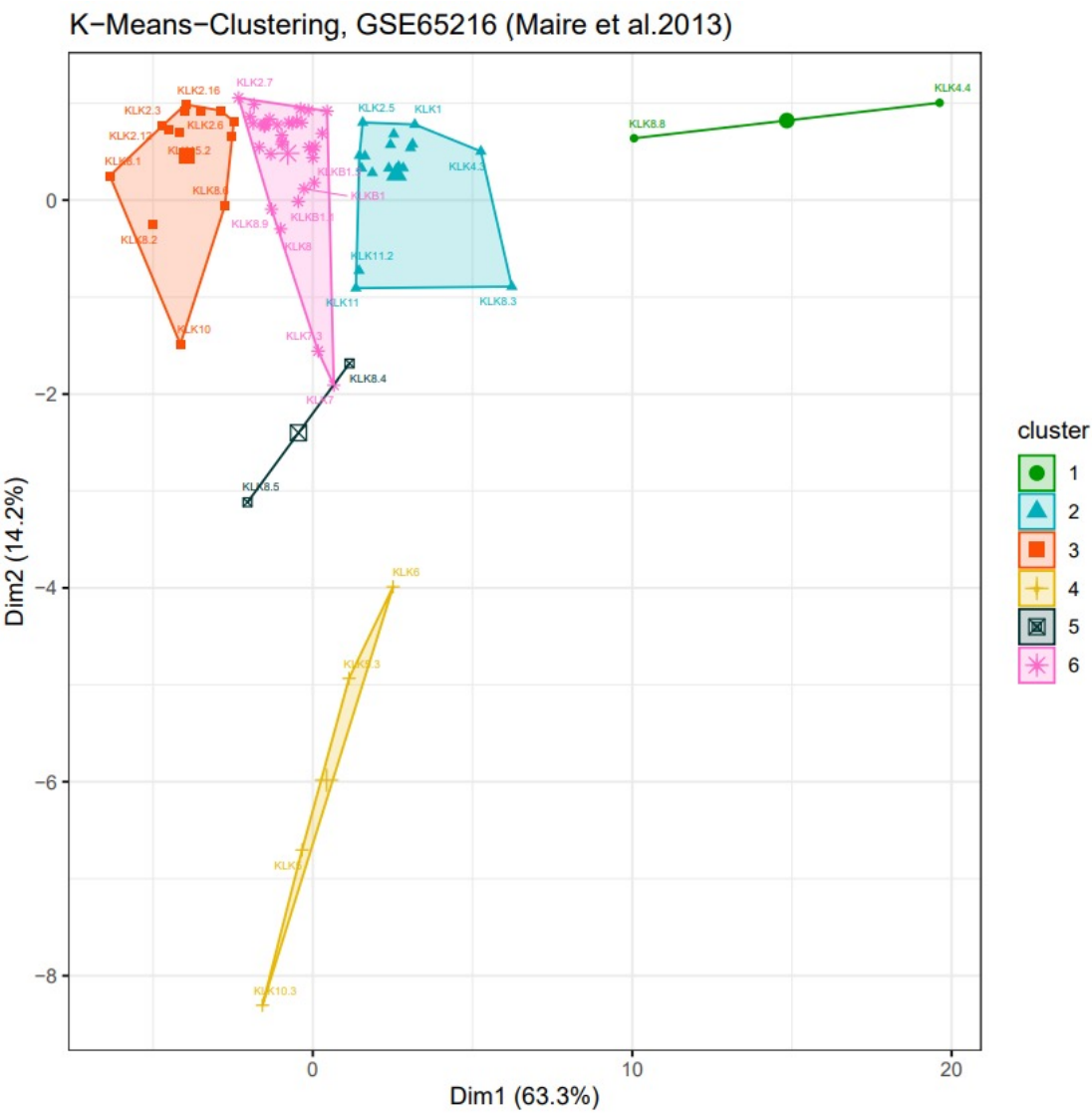
analogous QC presented by Dr. Maria Dinkelacker in “R Course Microarray Analysis”

Breast cancer - heatmap

- 20 samples
- four mutation: TNBC, Her2, LumA, LumB
- 73 KLK transcripts
- 63 TRAs



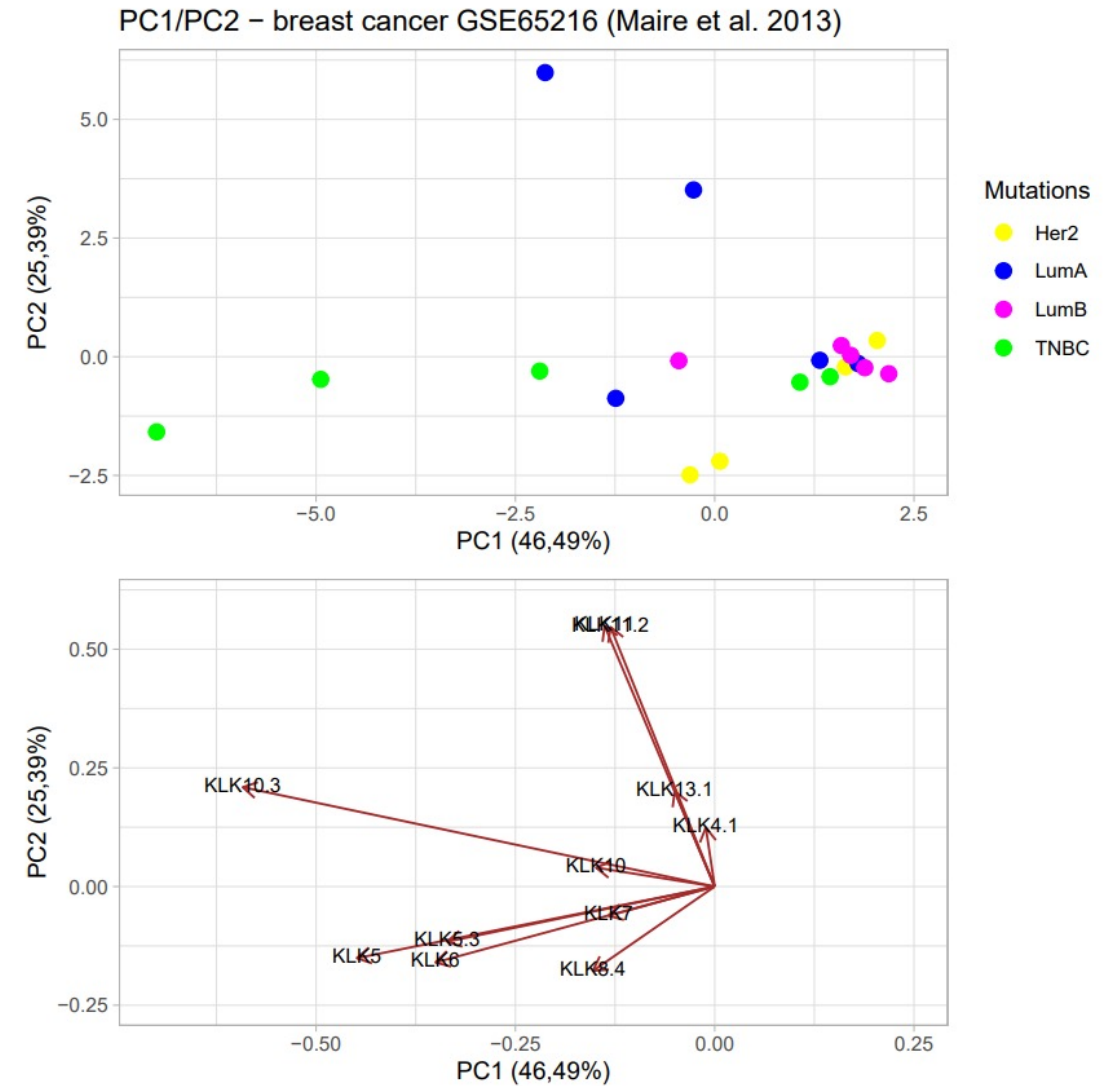
Breast cancer - k-means



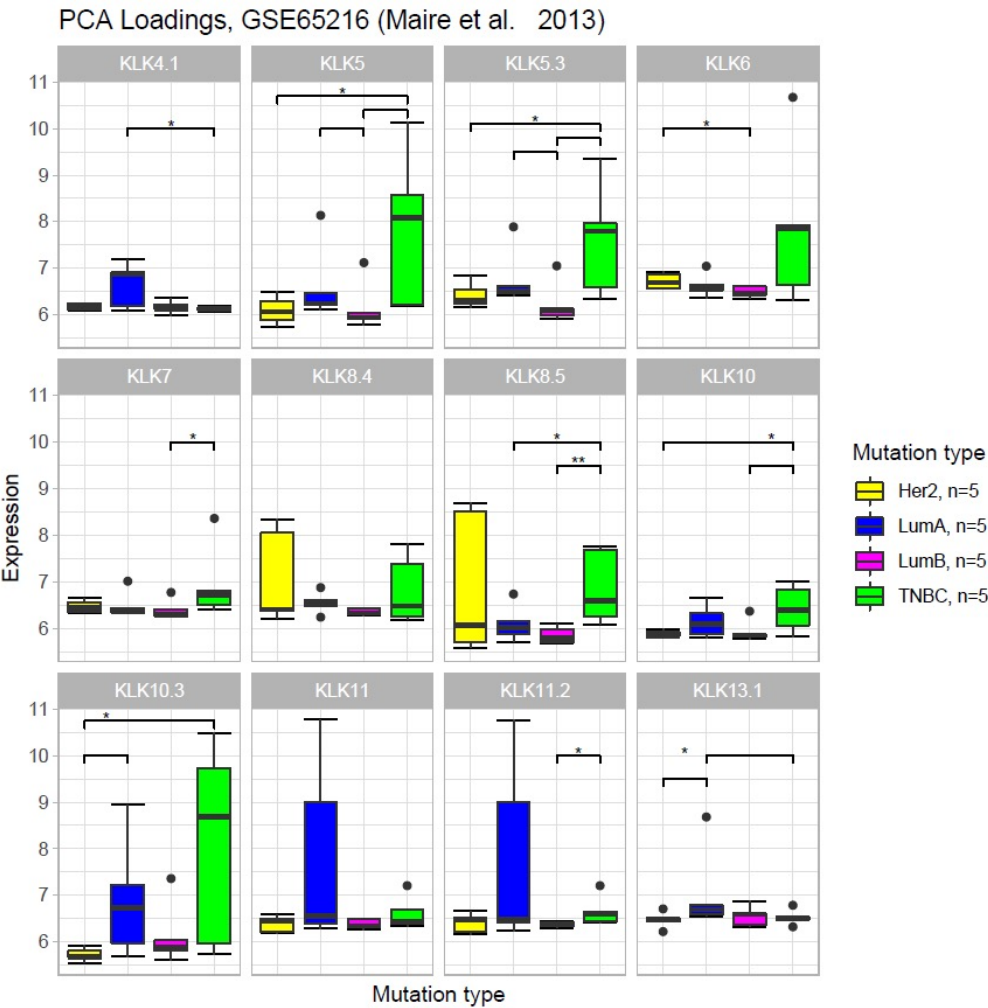
- cluster analysis for KLK genes
- two potentially interesting clusters
- cluster 1: KLK4.4 and KLK8.8
- cluster 4: KLKs 5, 5.3, 6 and 10.3

Breast cancer - PCA

- cumulative variance: 72%
- two LumA samples highly influenced by KLK11
- three TNBC samples influenced by KLK5,5.3,6
- clustered samples with no differentiation



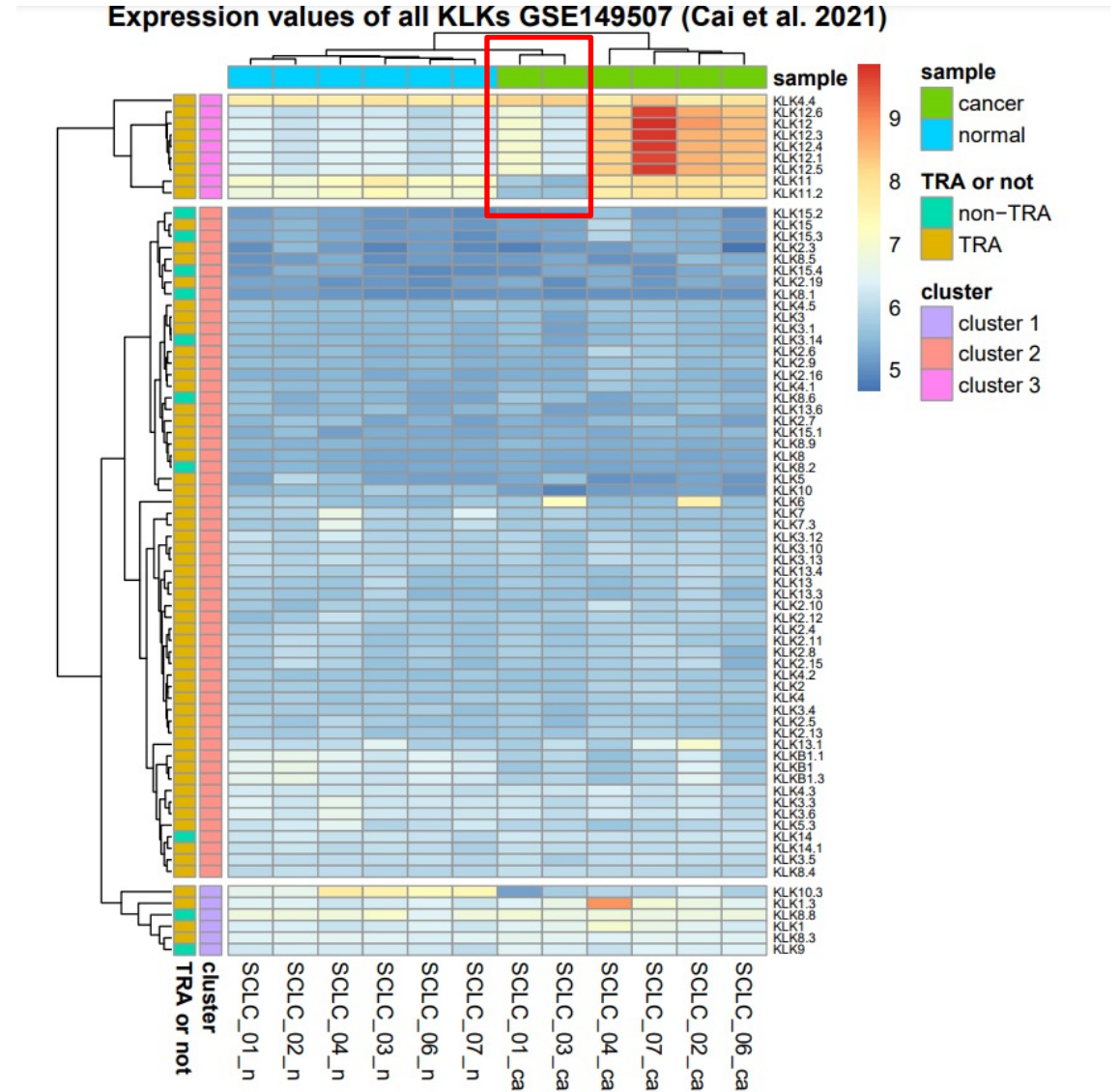
Breast cancer - hypothesis testing



- upper tail Wilcoxon-Mann-Whitney test
- LumA over-expression in KLK11 not significant
- TNBC over-expression in KLK5, KLK5.3 significant
- no specific expression pattern for Her2, LumA, LumB

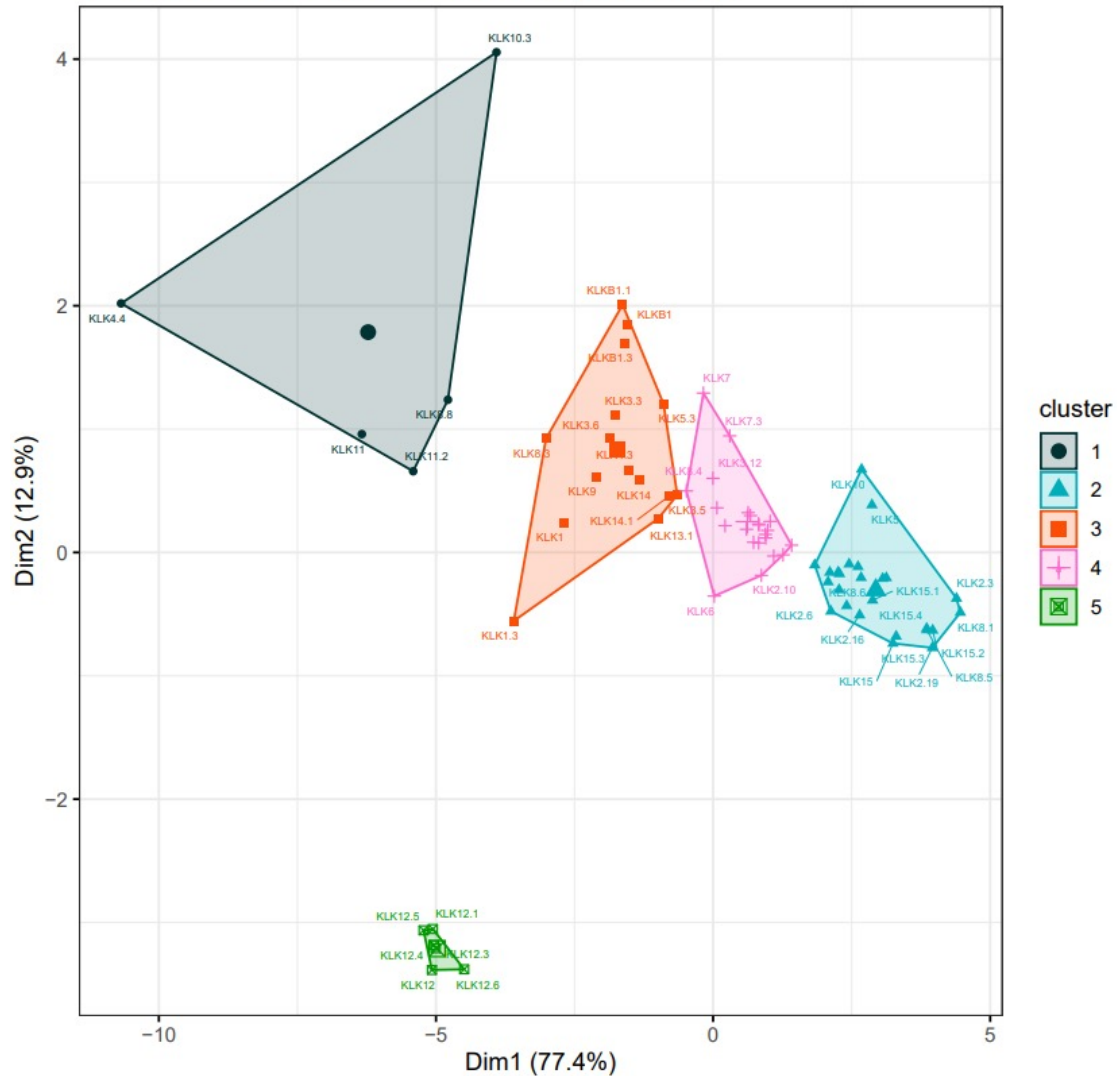
Lung cancer - heatmap

- twelve samples
- six patients with small cell lung cancer
- carcinoma and normal tissue samples



Lung cancer - k-means

K-Means-Clustering, GSE149507 (Cai et al. 2021)



→ cluster analysis for KLK genes

→ cluster 5 distinct

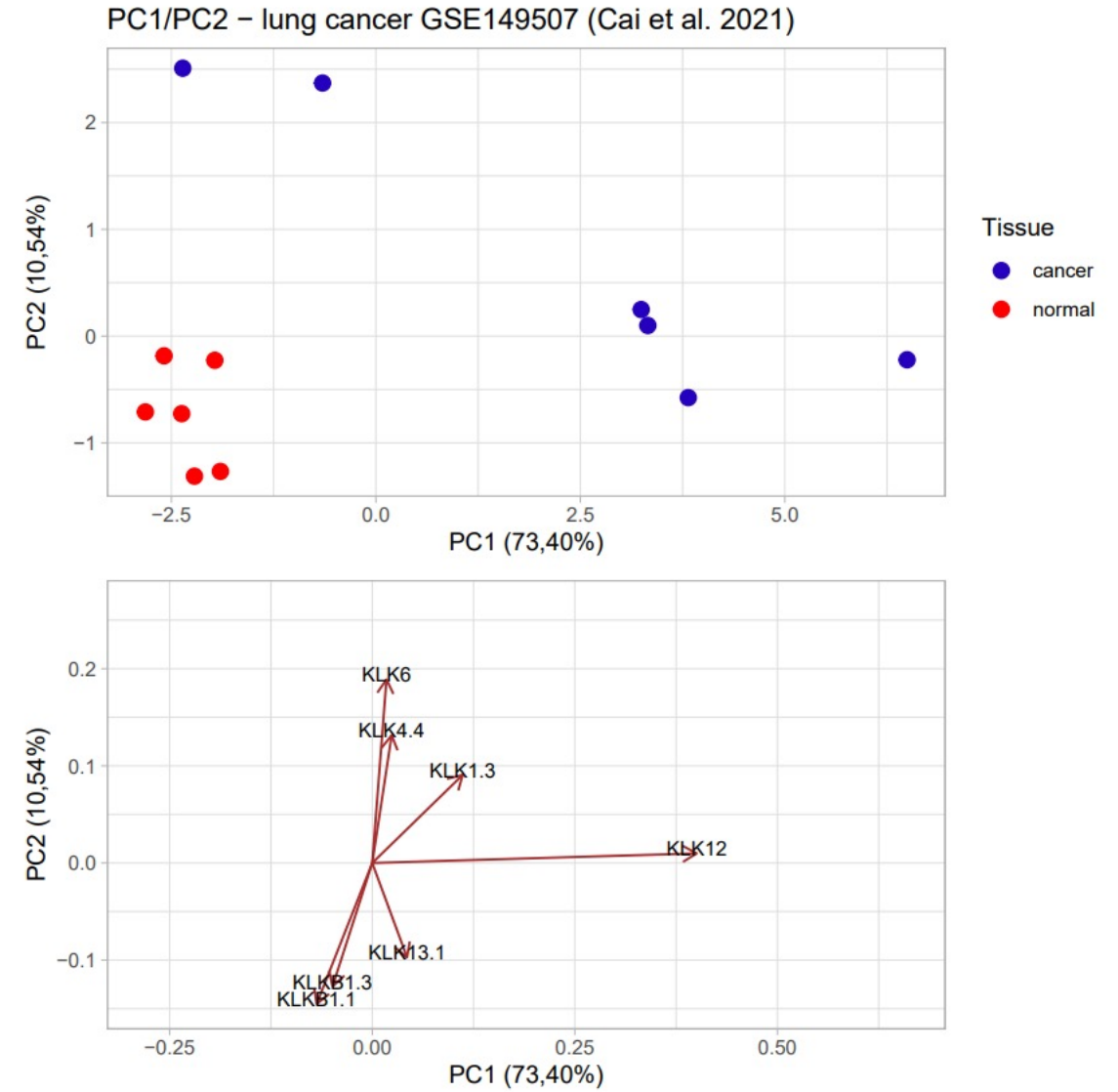
→ KLK12 & isoforms

→ cluster 1: KLKs 4.4, 10.3, 11 and 8.8

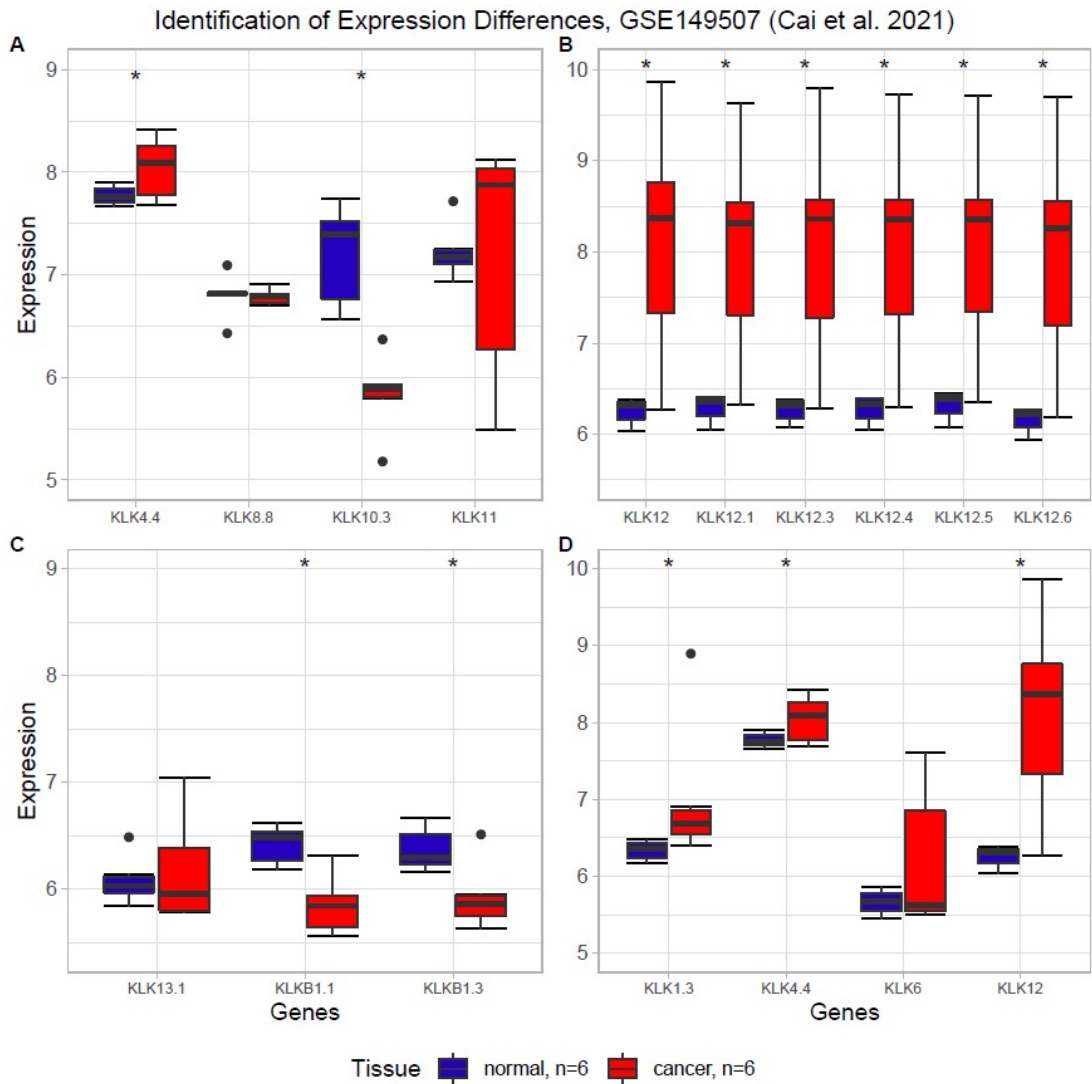
Lung cancer - PCA

→ cumulative variance: 85%

→ KLK12 marker for four microchips



Lung cancer - hypothesis testing



- Wilcoxon-signed rank test
- KLK10.3 upregulated in lung cancer → reduces proliferation
- KLK12 upregulated in lung → pro-angiogenic factor
- expression is significantly different in the tissues for most of the identified genes

Lung cancer - logistic regression analysis

Overfit model with KLKB1.1

```
Call:
glm(formula = tissue_type ~ KLKB1.1, family = binomial(), data = train.reg)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.930e-05 -2.100e-08  2.100e-08  2.100e-08  3.477e-05

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)      6872    5364632   0.001   0.999
KLKB1.1         -1037     809634  -0.001   0.999

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1.0585e+01  on 7  degrees of freedom
Residual deviance: 2.7531e-09  on 6  degrees of freedom
AIC: 4

Number of Fisher Scoring iterations: 25
```

→ main reason for overfit: low sample size

Regression model with KLK12

```
Call:
glm(formula = tissue_type ~ KLK12, family = binomial(), data = train.reg)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.00173  -0.61887   0.00364   0.11649   1.68489

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -35.542     35.715  -0.995   0.320
KLK12         5.257      5.372   0.979   0.328

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 10.5850  on 7  degrees of freedom
Residual deviance: 5.0733  on 6  degrees of freedom
AIC: 9.0733

Number of Fisher Scoring iterations: 8
```

Predictions

```
GSM4504107_SCLC_04_ca  GSM4504104_SCLC_02_n
                0.9994962                0.1825834

GSM4504106_SCLC_03_n  GSM4504114_SCLC_07_n
                0.3604154                0.2899213
```

→ p-value not significant

→ High uncertainty

Conclusion and outlook



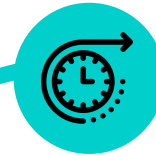
Results

- breast cancer: few overexpression
- lung cancer: different expression in carcinoma and healthy tissue



Note

- No breast cancer healthy tissue
- Small number of samples



Outlook

- Comparative value with healthy tissue
- More samples for analysis

Kallikrein genes are possible good biomarkers

Questions?

