

The role of tissue-specific antigens in different cancer entities

Analysis of CSN Genes

Supervisor: Dr. Maria Dinkelacker

Tutor: Nils Mechtel

Subject: Data Science SoSe 2021

Date of Presentation: 21.07.2021

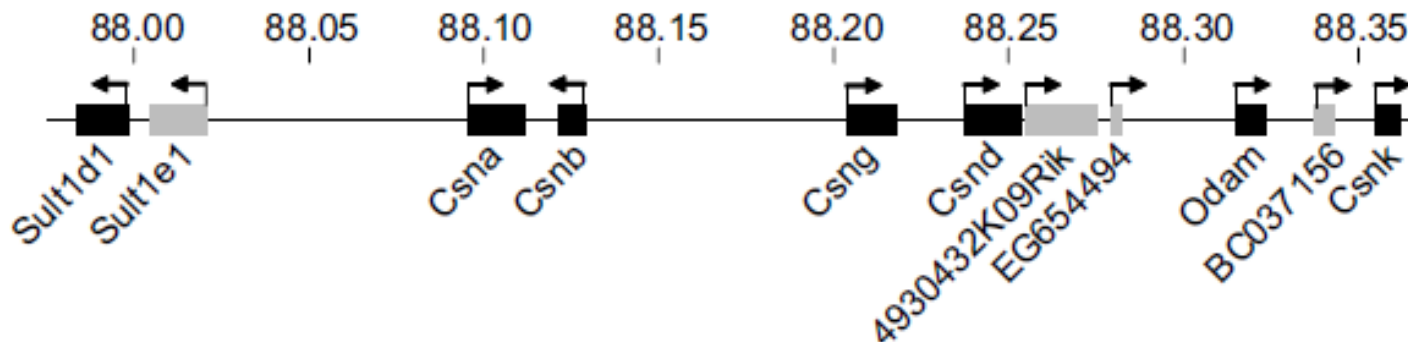
Participants: Anastassia Fink, Andreas Breuß, Yuan Sun

Biological Background

A thin vertical line is positioned to the right of the text. At the bottom of the slide, there is a solid horizontal bar in a dark orange color.

Recap: Casein Genes (*CSN*)

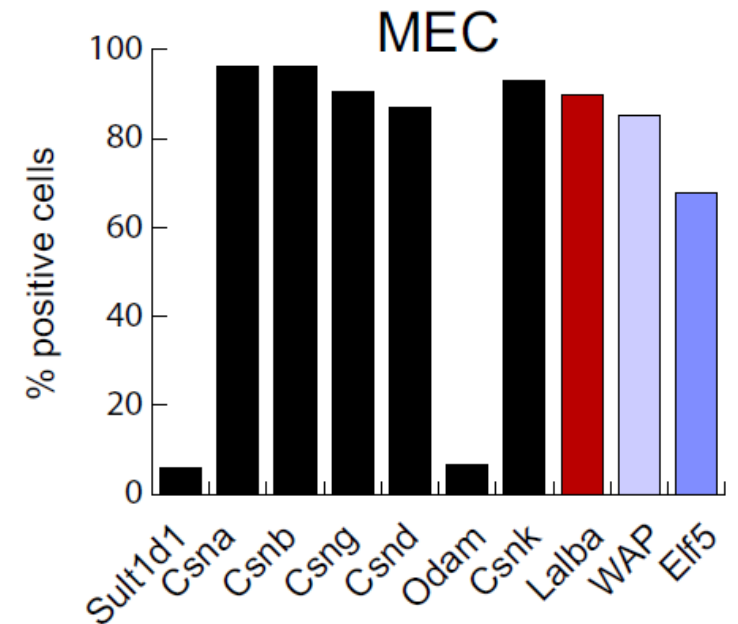
- Casein genes: *Csna*, *Csnb*, *Csng*, *Csnd*, *Csnk*
- Temporarily regulated during late pregnancy and postpartum lactation period (-> hormone dependent)
- TRA restricted to mammary gland epithelial cells (MEC)
- Location: casein gene region on chromosome 4 in humans (5 in mice)



Derbinski et al. 2008, Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism, Proc Natl Acad Sci U S A. 2008 Jan 15;105(2):657-62.

Recap: Correlation with other genes

- *Csna*, *Csnb*, *Csng*, *Csnd*, *Csnk*
-> Coexpressed in MEC
- Functionally related genes: lactalbumin- α (*Lalba*), whey acidic protein (*WAP*) and *Elf5*
-> These genes show correlated expression with *Csn* in MEC
- Colocalized genes: *Sult1d1* and *Odam*
-> These genes show no correlated expression with *Csn* in MEC



Derbinski et al. 2008, Promiscuous gene expression patterns in single medullary thymic epithelial cells argue for a stochastic mechanism, Proc Natl Acad Sci U S A. 2008 Jan 15;105(2):657-62.

Lung cancer dataset

- > Small cell lung cancer (SCLC)
- high-grade neuroendocrine tumor
- very bad prognosis, no targeted therapy is available
- initiating mutations: *TP53* and *RB1*

Breast cancer dataset

-> HER2 positive breast cancer

- overexpression of human epidermal growth factor receptor-2 (HER2)

-> Luminal A (LumA) and luminal B (LumB) breast cancer

- estrogen & progesterone receptors are expressed, additionally other hormone receptor-related genes

-> Triple negative breast cancer (TNBC)

- no expression of estrogen & progesterone receptors and no overexpression of HER2

- poor prognosis: high proliferation rate and genetic instability

Project Goals

A thin vertical line is positioned to the right of the text "Project Goals", extending from the top of the word "Project" down to the bottom of the word "Goals".A solid orange horizontal bar spans the entire width of the slide at the bottom.

Project goals

1. Investigation of the expression of *CSN* genes in cancer cells

- Are any *CSN* genes up-regulated in cancer?
- Could they be potentially good targets for targeted therapy of cancer?

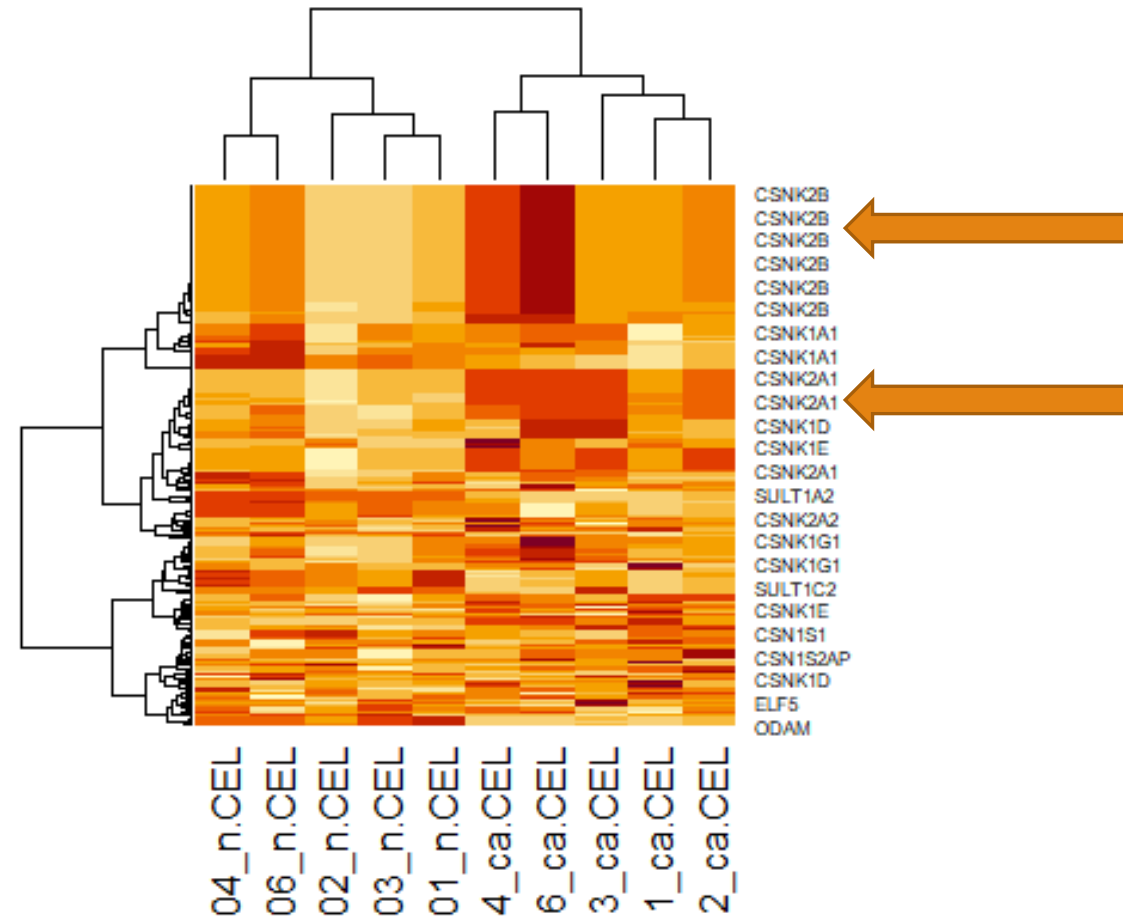
2. Investigation of the correlation of *CSN* with other genes

- Are functionally related, but not colocalized genes like *Lalba*, *WAP* and *Elf5* coexpressed with *CSN*?
- Are functionally unrelated, but colocalized genes like *Sult1d1* and *Odam* coexpressed with *CSN*?

Heatmap

A thin vertical line is positioned to the right of the word 'Heatmap'. At the bottom of the slide, there is a solid horizontal bar in a dark orange or brown color.

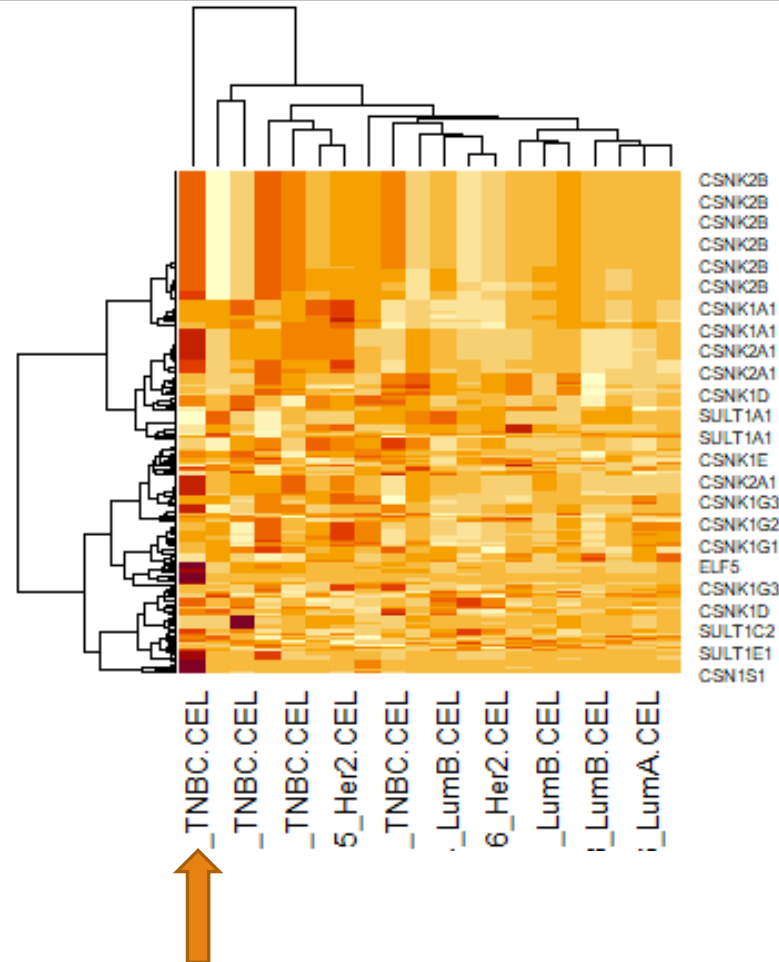
Heatmap from lung cancer data



CSNK2B and CSNK2A1
similar higher expression
value in cancerous group
than non-cancerous
group

Using t-test to prove it

Heatmap from breast cancer data



Find out whether breast cancer
TNBC are highly expressed in
all CSN genes

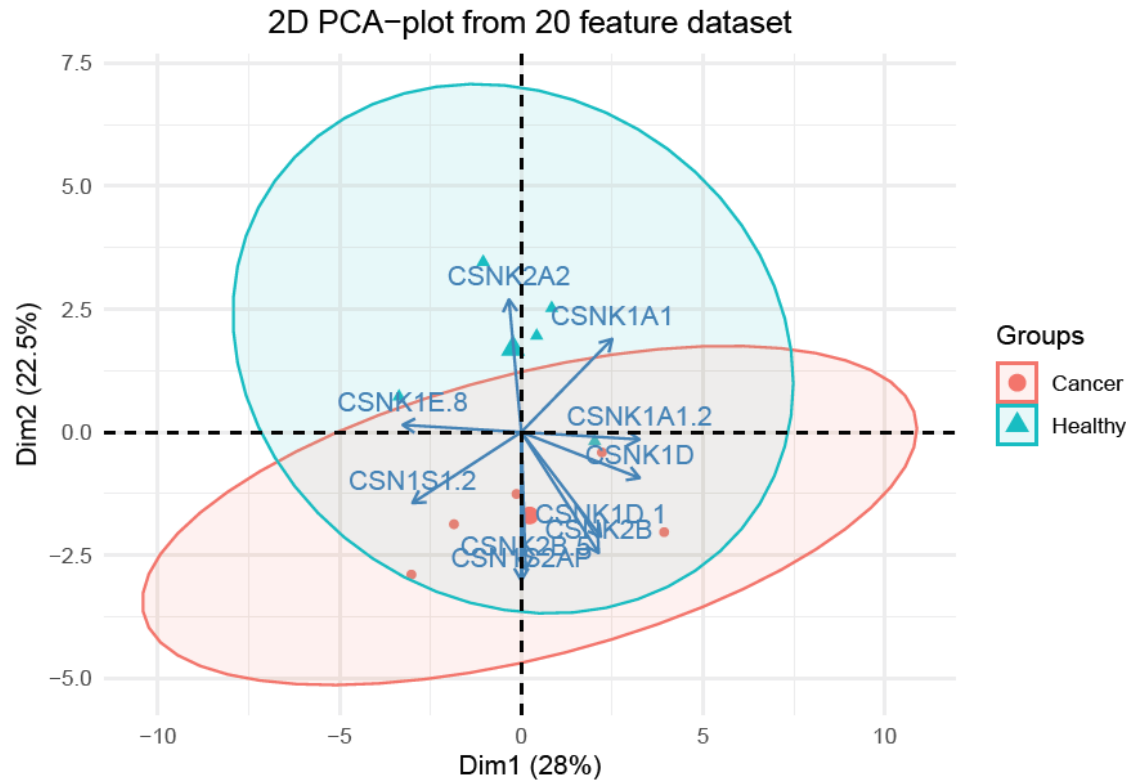


Using t-test to prove it

Principal Component Analysis



PCA: Lung cancer

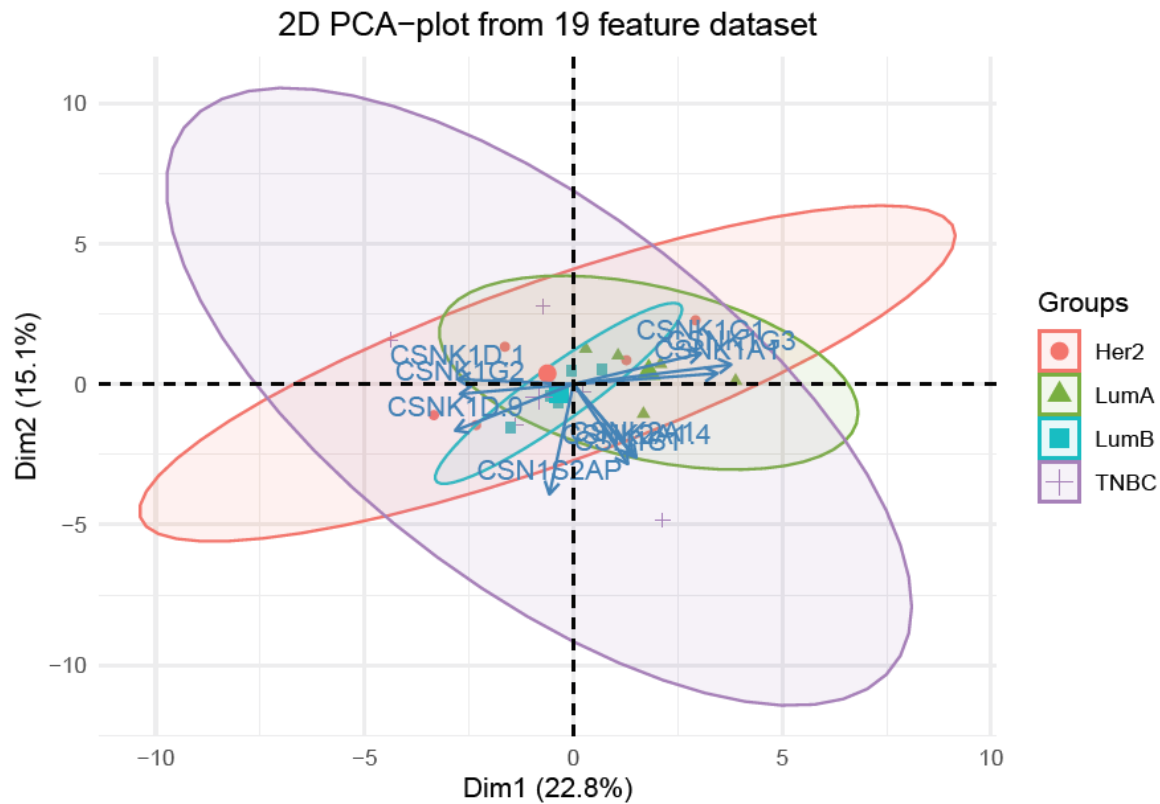


before the PCA: all transcripts with a correlation over 0,8 are filtered out
→ from 181 transcripts 20 remain

→ the groups of cancerous and healthy cells are overlapping, a separating trend is visible

→ possible differentiation of cancerous and healthy cells in a bigger dataset

PCA: Breast cancer



before the PCA: All transcripts with a correlation over 0,8 are filtered out

→ from 181 transcripts 19 remain

→ the groups are mostly overlapping, which means that all samples have similar characteristics

→ Interesting: LumA and LumB have less variation in the distribution their samples than TNBC or Her2

T-test

A vertical line is positioned to the right of the text 'T-test'. At the bottom of the slide, there is a solid horizontal bar in a dark orange color.

Results of t-test from lung cancer

	Gene of interest	P-value from t-test
→	CSNK2A1	3.928261e-38
→	CSN1S1	1.008727e-78
→	CSNK1G2	2.445394e-03
	CSNK1A1	7.133933e-01
	CSNK2A2	1.000000e+00
	CSNK1D	6.107149e-02
→	CSNK1E	8.955636e-03
	CSNK1G3	5.890529e-01
→	CSNK1G1	1.195431e-04
→	CSNK2B	9.661402e-14

6 genes out of 10 genes have higher expression value in cancerous group

Results of one way anova test from breast cancer

Gene of interest	P-value from one way anova test
CSNK2A1	0.18140659
CSN1S1	0.51918957
CSNK1G2	0.20523146
CSNK2B	0.08206299
CSNK1A1	0.68839922
CSNK2A2	0.10440108
CSNK1G1	0.05561539
CSNK1E	0.56146738
CSNK1D	0.25862612
CSNK1G3	0.24774170

All genes have higher p-value than 0.05 -> no significant difference under the four breast cancer type

Linear regression

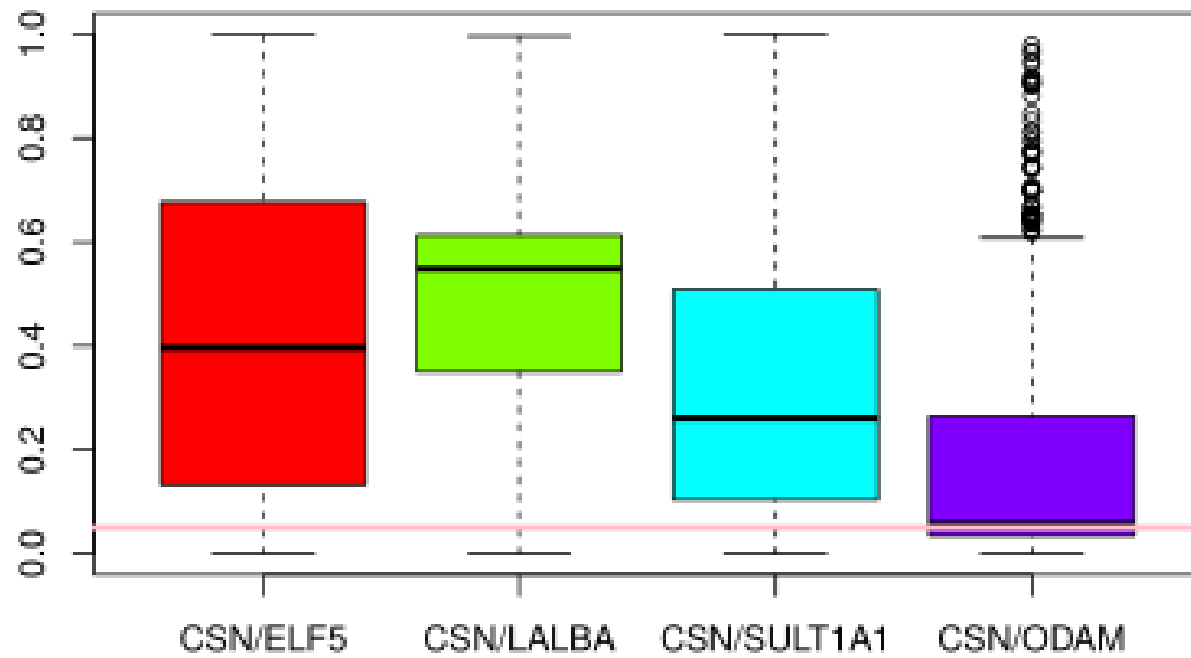
A thin vertical line is positioned to the right of the text. At the bottom of the slide, there is a solid horizontal bar in a dark orange color.

Linear regression

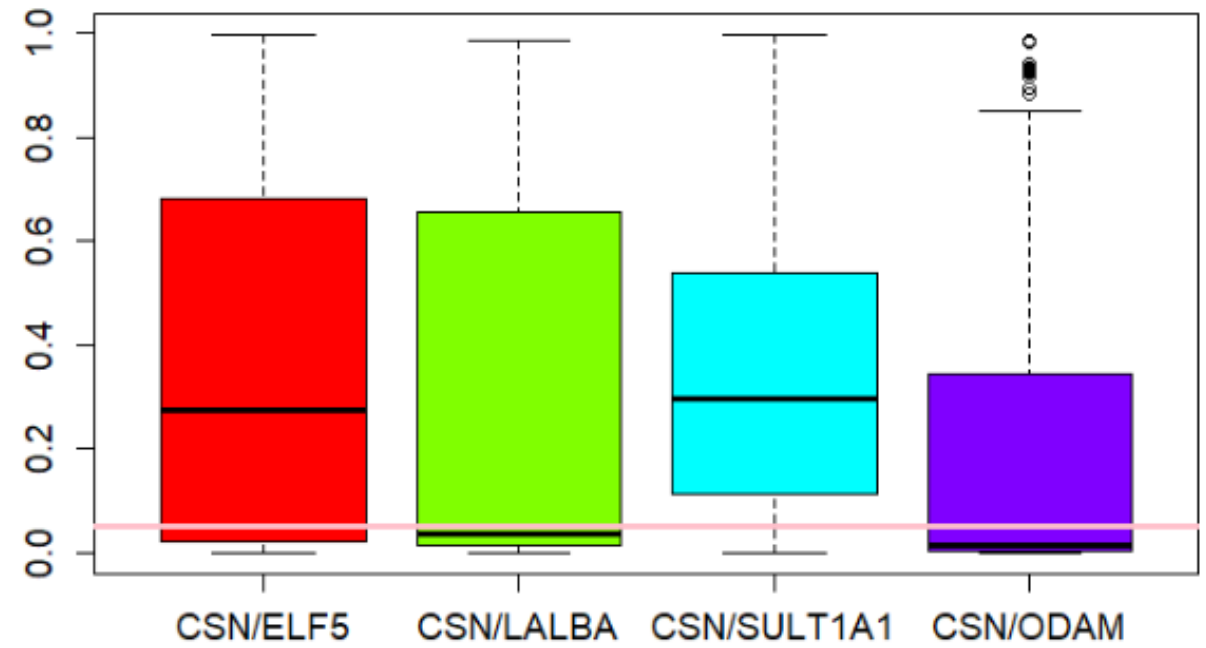
- Correlations between the casein genes and the genes Sult1A1, ODAM, ELF5 and ODAM
- Every CSN transcript was compared with one of the mentioned genes
- F-test was used to determine if the H0 thesis is valid or not / p-value

Linear regression

Lung cancer



Breast cancer



Discussion

A thin vertical line is positioned to the right of the word "Discussion". At the bottom of the slide, there is a solid orange horizontal bar.

Discussion

- Possible Obstacles to keep in mind
- Note quiet what we thought
- Cluster behaviour
- Expression values
- Correlations and localization

