

---

# Thyroid specific antigens

— and how they affect thyroid cancer —

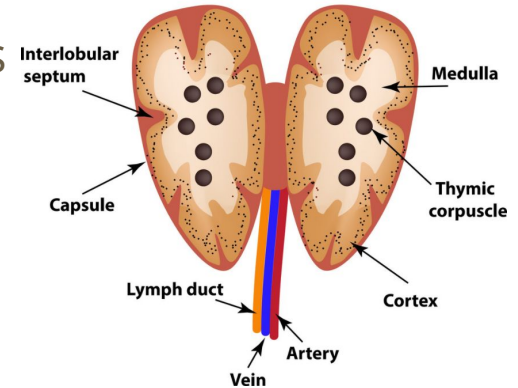
---

A data science project performed by Camila Vacas, Line Weiß and Carina Keßler  
Supervisor: Dr. Maria Dinkelacker, Dr. Carl Herrmann  
Tutor: Nils Mechtel

# **biological background**

# TRAs and thyroid cancer

- mTECs express TRAs
- Hashimoto's and Grave's diseases as most common autoimmune disorders
- Tumor Associated Antigens are expressed in thymus
- Radiation & thyroid cancer
  - sensitive to long-term effects
  - 80% papillary carcinoma
  - activation of MAPK signal cascade



structure of the thymus gland

<https://www.vectorstock.com/royalty-free-vector/structure-thymus-gland-infographics-vector-12904537>

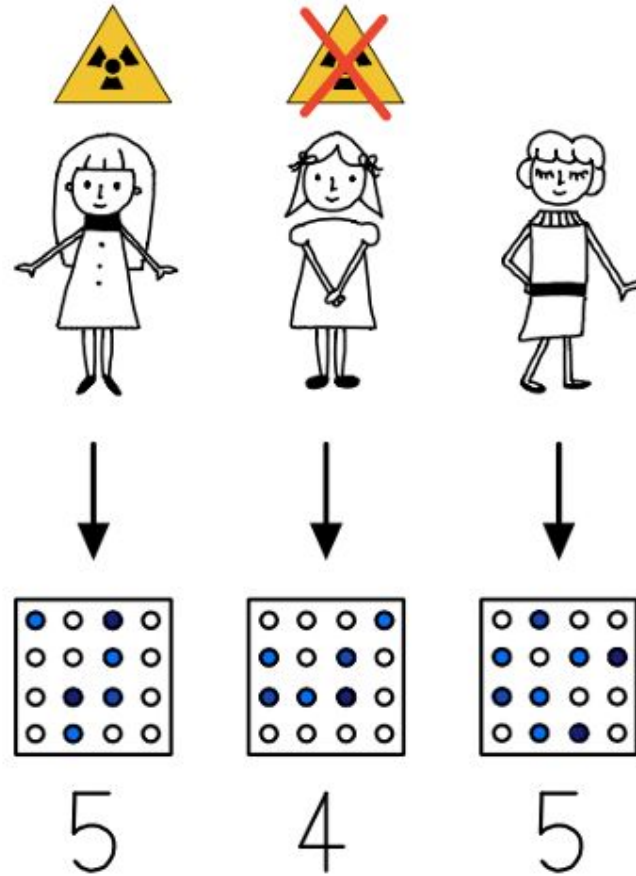
# our dataset

# the structure of our dataset

5 arrays of patients with papillary thyroid cancer (PTC), who were **exposed** to radiation

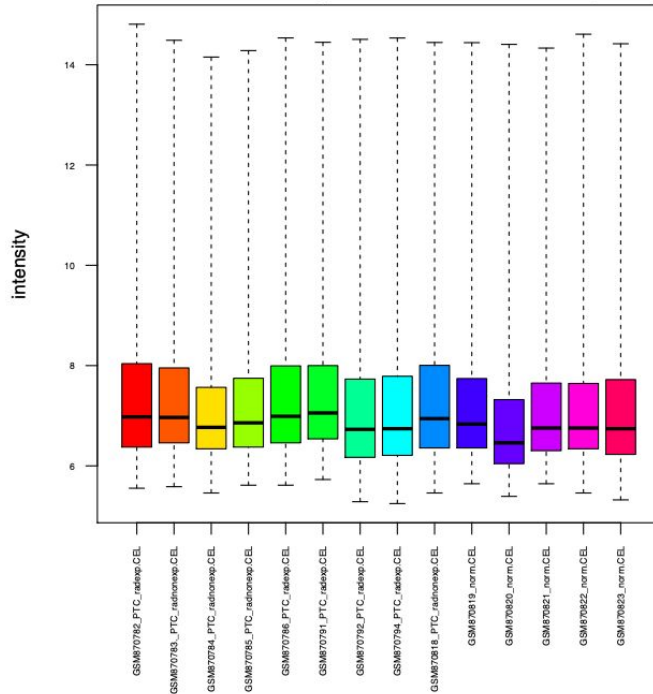
4 arrays of patients with PTC, who were **NOT exposed** to radiation

5 arrays of patients with **healthy** thyroid tissue

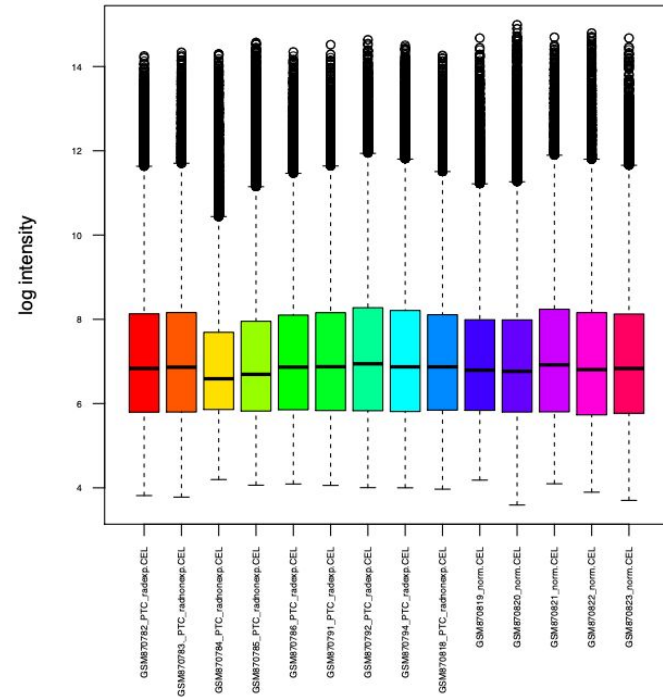


# quality control - boxplots

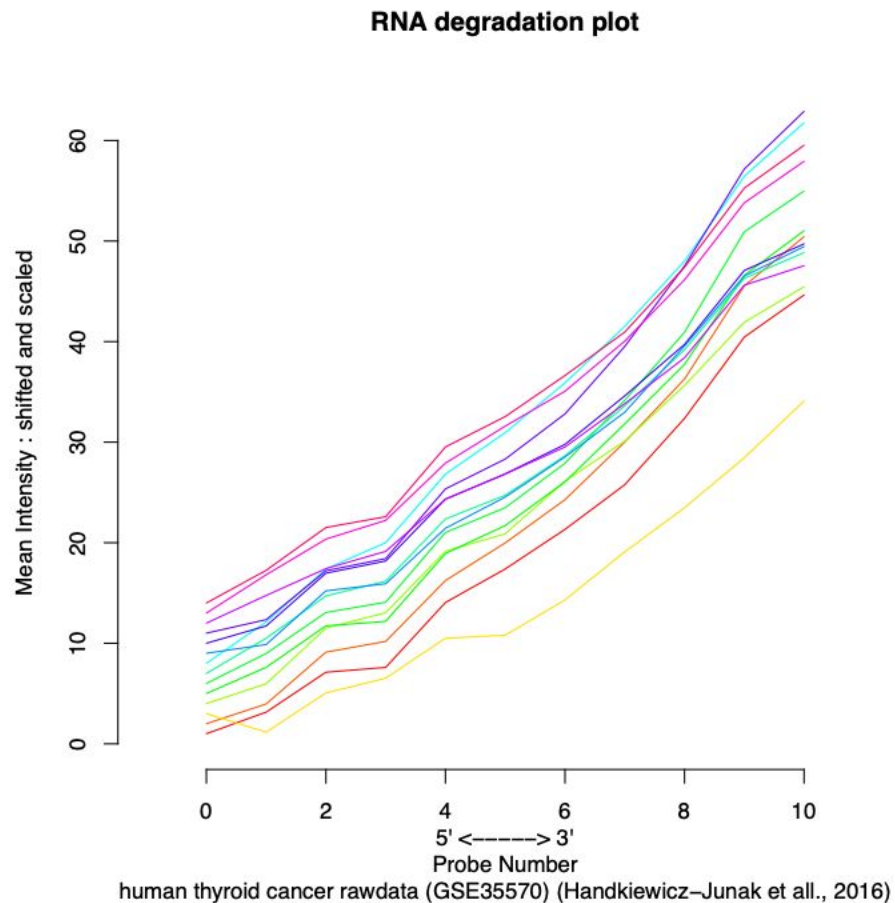
Gene expression in papillary thyroid cancer (GSE35570)  
before normalization (Handkiewicz-Junak et al., 2016)



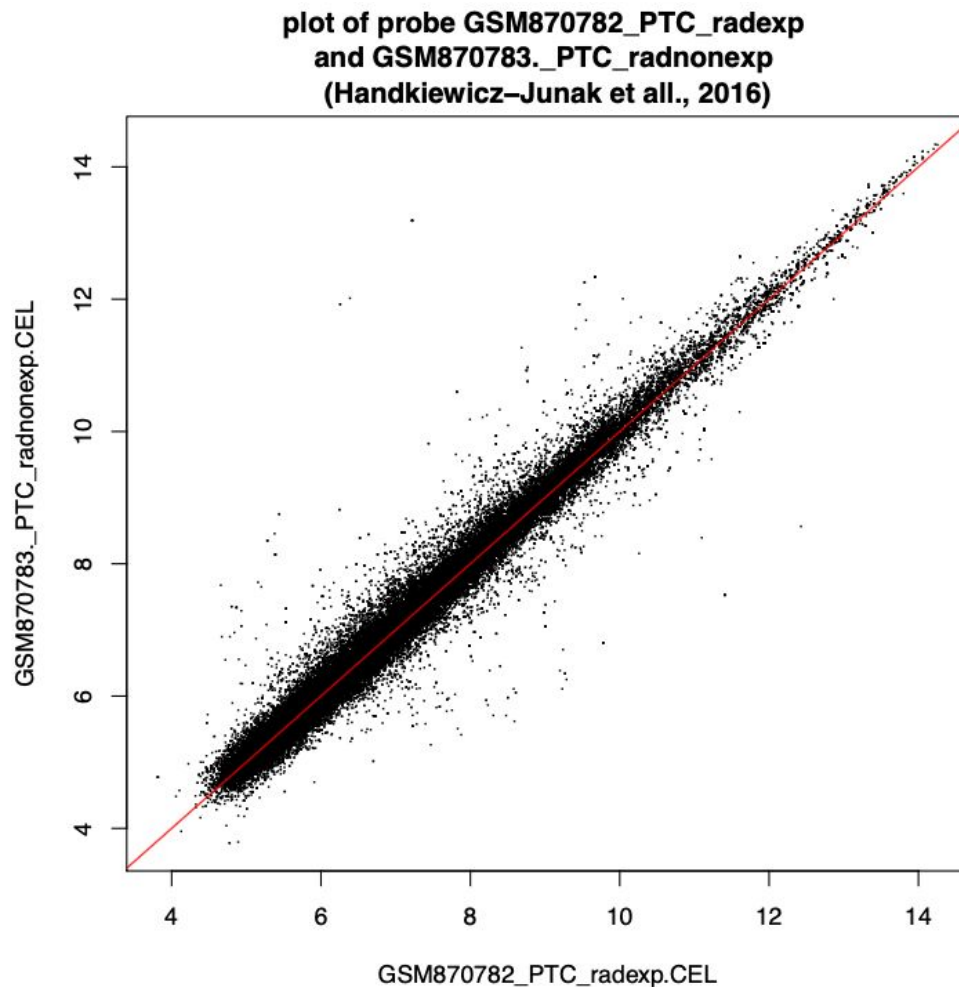
Gene expression in papillary thyroid cancer (GSE35570)  
after vsnrma normalization (Handkiewicz-Junak et al., 2016)



# quality control - RNA degradation plot

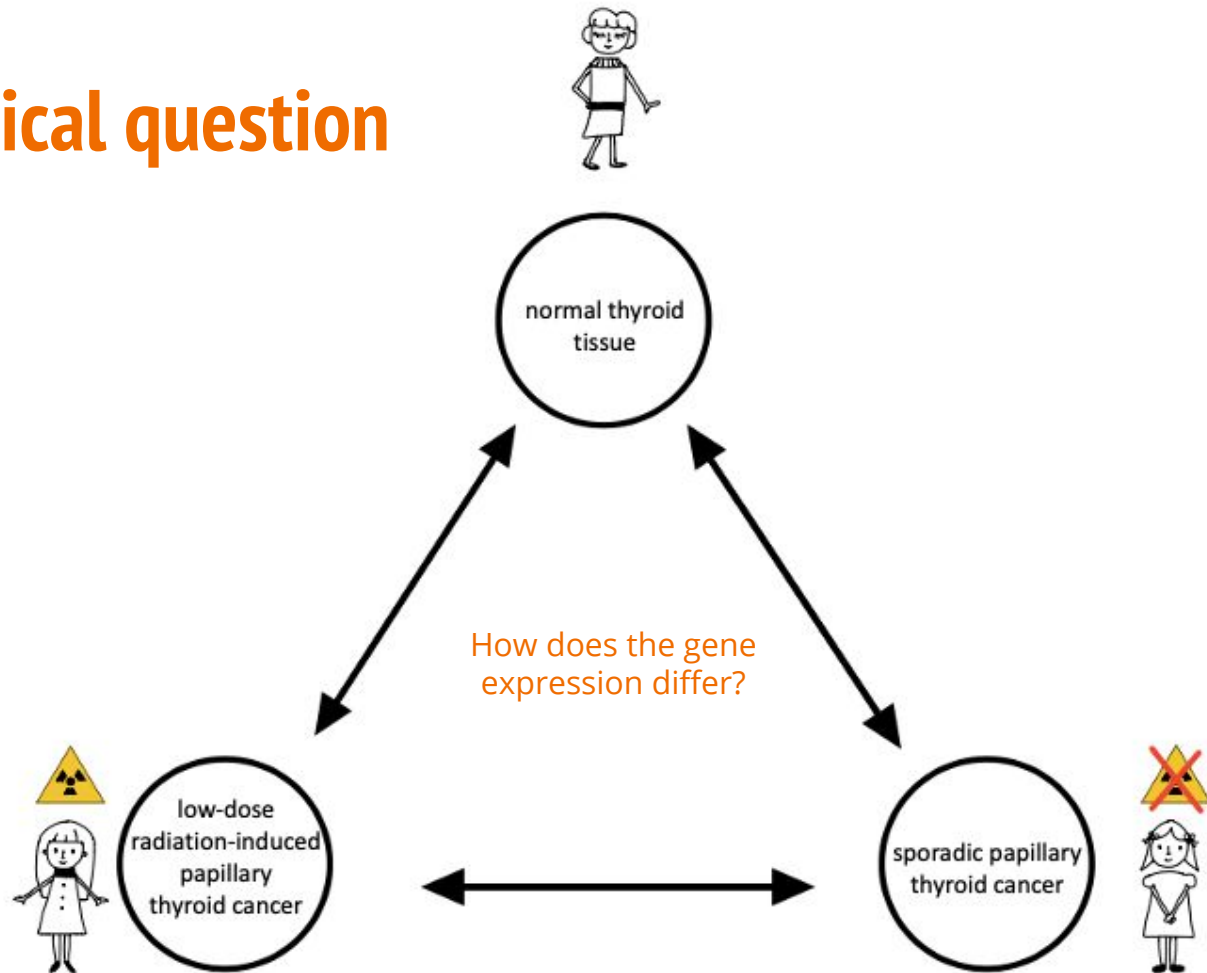


# quality control - scatterplots





# the biological question

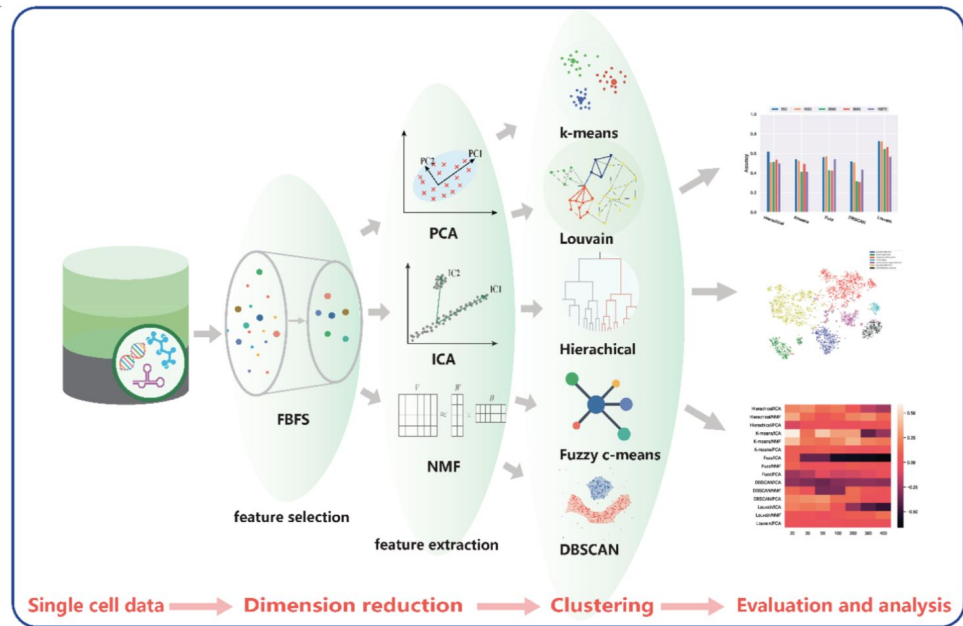


# objective

Identify which genes are upregulated or downregulated in radiation exposed samples compared to nonexposed samples and healthy samples for drug targets in cancer therapy

# General methods:

1. Data filtering and Cleanup
2. Descriptive statistics
3. Dimension reduction
  - PCA
  - Clustering
4. Differential expression analysis
  - One-way ANOVA
  - Post hoc test (Tukey)
5. Sorting of genes (Gene ontology)
6. Linear regression analysis and proportion test



Feng, C.; Liu, S.; Zhang, H.; Guan, R.; Li, D.; Zhou, F.; Liang, Y.; Feng, X. Dimension Reduction and Clustering Models for Single-Cell RNA Sequencing Data: A Comparative Study. *Int. J. Mol. Sci.* 2020, 21, 2181

# Clustering

Are genes significantly differentially expressed? Upregulated? Downregulated?  
Are there any genes co-expressed?

## Sorting of genes

Do genes differentially expressed can be clustered under a specific characteristic?  
Location in the cell? Function? Relation a specific signaling pathway? Do these genes have any relation to one another?

# Regression analysis

How well can the expression data of one gene be used to predict the expression of another?

# Future perspectives

Is the relation important for cancer? Diagnostic? Treatment? Cause of unregulated cell growth or an effect of this? (Further review bibliography and specific gene function)

# Timeline

- This timeline should be an approximation of how we expect the workload to be divided across the semester.
- Week 1, which has already passed, is also taken in consideration.

Week 1  
10.05

## Filtering and data cleanup

- Selecting samples of interest
- Cleaning data

Week 2  
17.05

## Descriptive Statistics

- Distribution of gene expression in sick and healthy patients.
- Breast Cancer Data Set analysis

Week 3  
24.05

## Dimension Reduction

- Clustering (k-means)
- Principal Component Analysis

Week 4  
31.05

## Expression analysis of different groups

- one-way ANOVA
- post hoc test
- (dendrogram)

Week 5  
7.06

## Expression and interpretation of different groups

- Genes downregulated or upregulated?
- Gene ontology - relationship between genes
- PCA with more chips

Week 6  
14.06

## Sorting of genes

- Function
- Location

Week 7 21.06	<b>Regression analysis</b>
	<ul style="list-style-type: none"> <li>• between certain characteristics of samples and their gene expression</li> <li>• Are there changes due to radiation?</li> </ul>
Week 8 28.06	<b>Linear regression analysis evaluation.</b>
	<ul style="list-style-type: none"> <li>• Evaluation of the results</li> <li>• F-Test</li> </ul>
Week 9 + 10 5.07	<b>Evaluation</b>
	<ul style="list-style-type: none"> <li>• Correcting mistakes</li> <li>• Summarizing project</li> <li>• Conclusions</li> <li>• RMarkdown</li> </ul>
date: : 21.07.21 Camila, Line and Carina	<b>Final Presentation</b>



**our results so far**

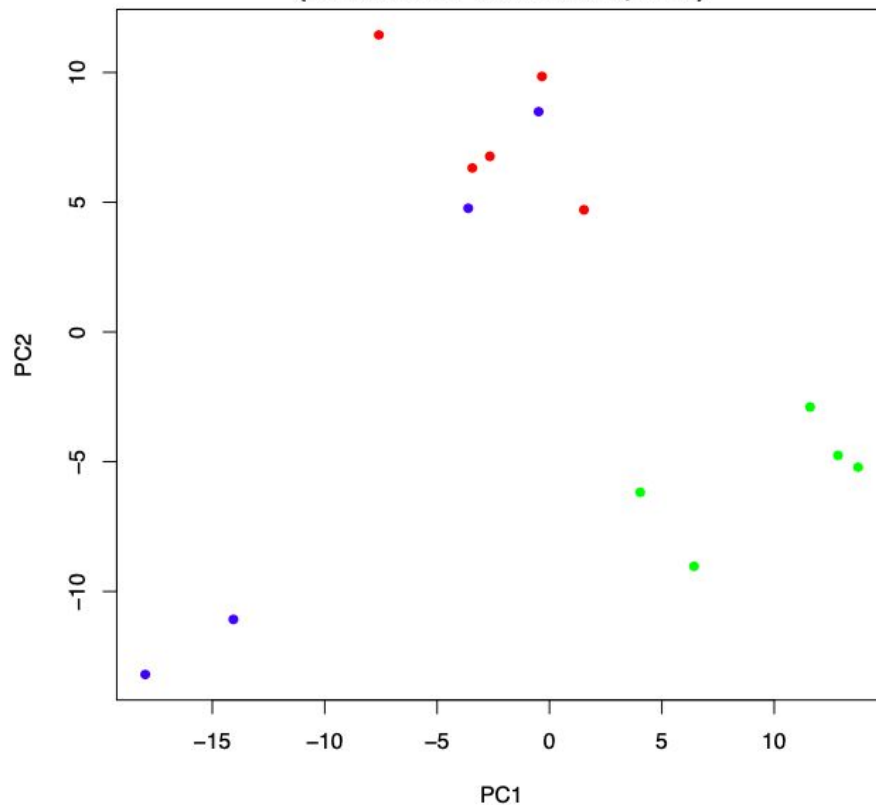
# PCA

Further questions:

Are those outliers of the  
not-exposed group?

→ adding for chips

PC1 vs. PC2 for radiation-exposed PTC (red),  
not-exposed PTC (blue) and healthy thyroid tissue (green)  
(Handkiewicz-Junak et al., 2016)

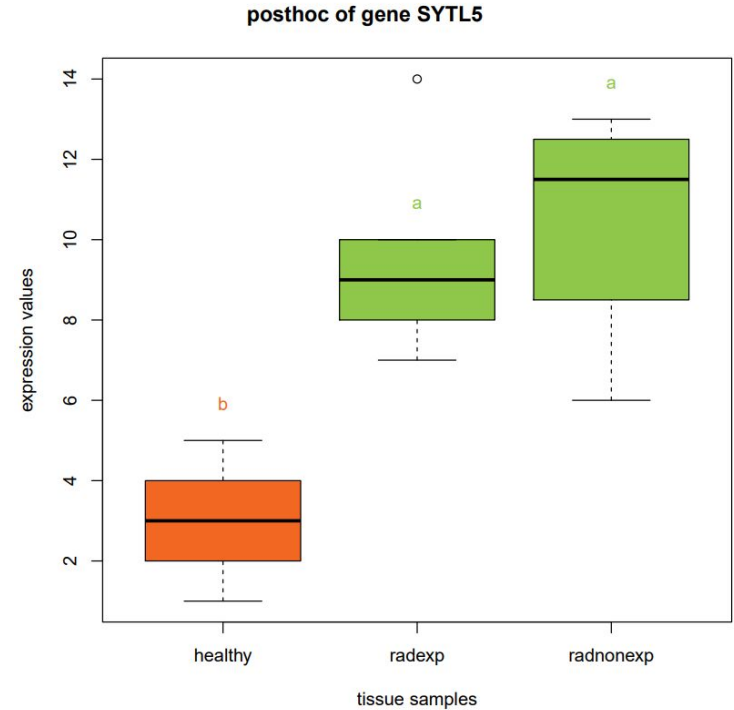


# ANOVA and post hoc test

=> SYTL5 is upregulated  
in thyroid cancer

Next step:

cluster genes in biological context



# Main literature

Dinkelacker 2007, "A database of genes that are expressed in a tissue-restricted manner to analyse promiscuous gene expression in medullary thymic epithelial cells." Diplomarbeit, Albert-Ludwigs-Universitaet, Freiburg, Germany, 2007.

Dinkelacker, 2019, PhD thesis, "Chromosomal clustering of tissue-restricted antigens", University of Heidelberg.