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
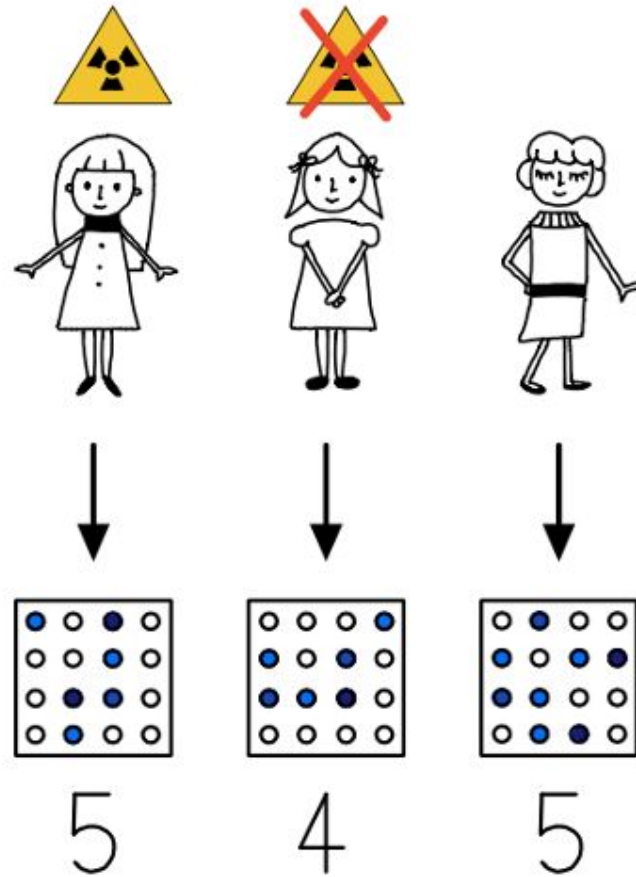 - activation of MAPK signal cascade
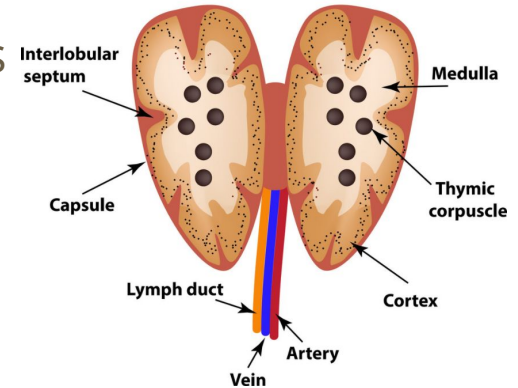
structure of the thymus gland

our dataset

the structure of our dataset

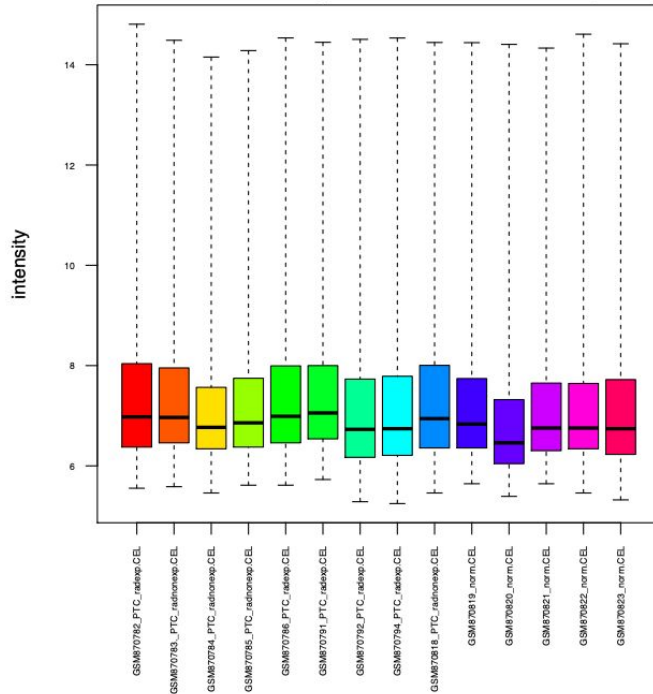5 arrays of patients with papillary thyroid cancer (PTC), who were exposed to radiation

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

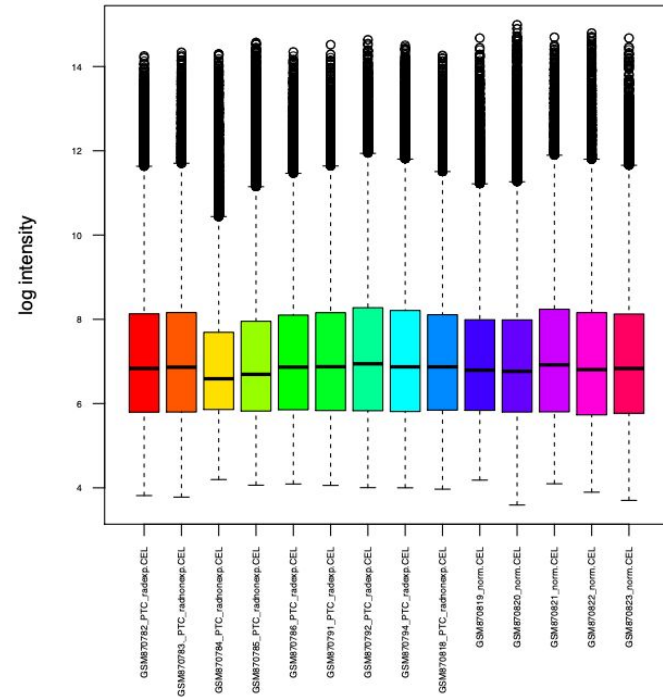5 arrays of patients with healthy thyroid tissue

quality control - boxplots

Gene expression in pappillary thyroid cancer (GSE35570) before normalization (Handkiewicz–Junak et all., 2016)

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

quality control - RNA degradation plot

RNA degradation plot

Mean Intensity : shifted and scaled

5' <-----> 3'
Probe Number

human thyroid cancer rawdata (GSE35570) (Handkiewicz–Junak et all., 2016)
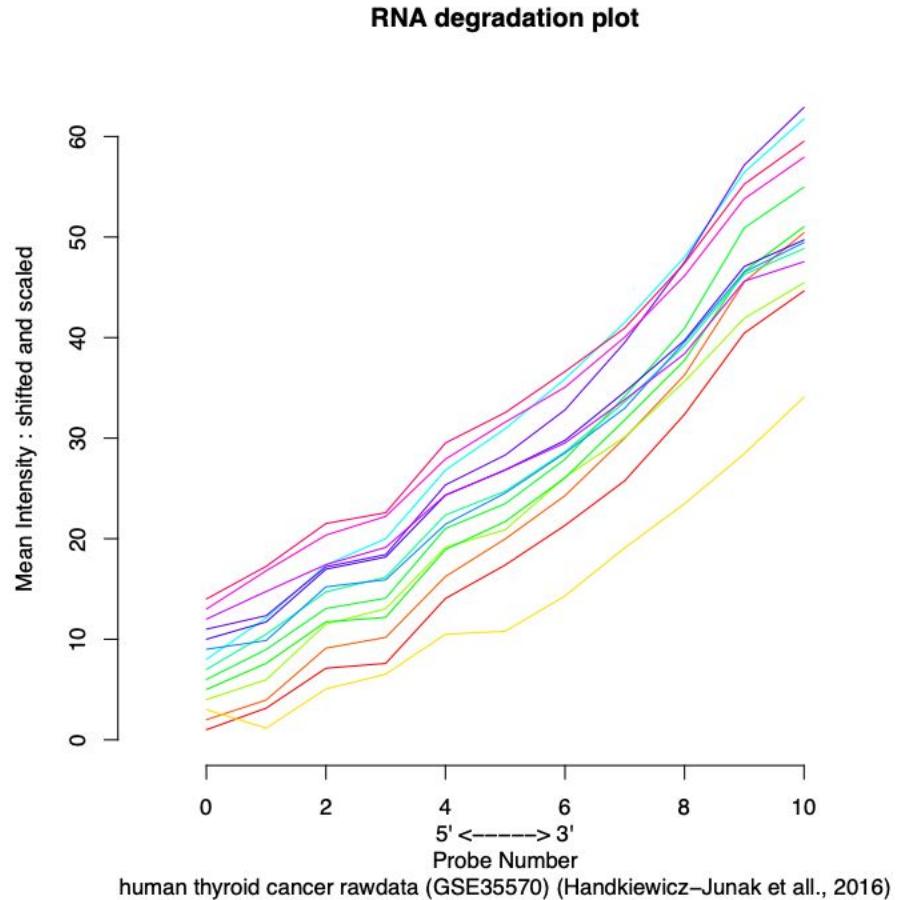
quality control - scatterplots

plot of probe GSM870782_PTC_radexp
and GSM870783._PTC_radnonexp
(Handkiewicz–Junak et all., 2016)
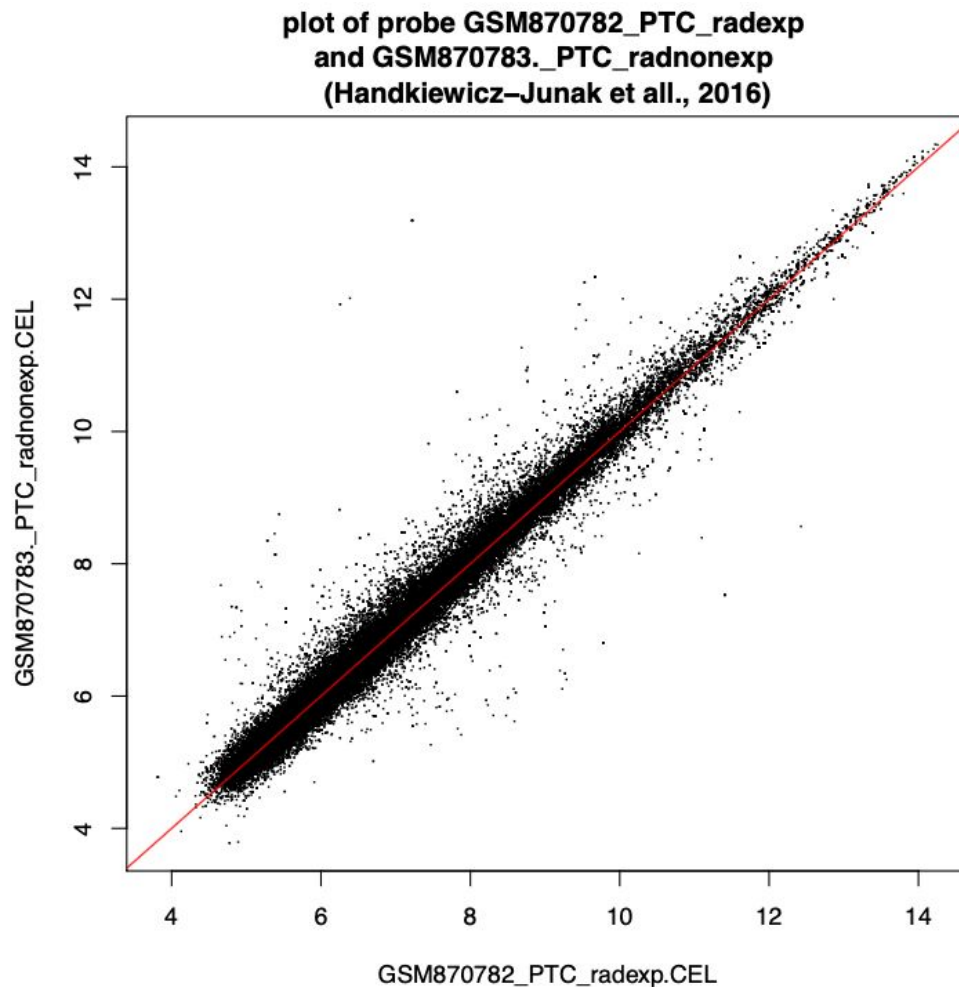
the biological question

normal thyroid tissue

low-dose radiation-induced papillary thyroid cancer

How does the gene expression differ?

sporadic papillary thyroid cancer
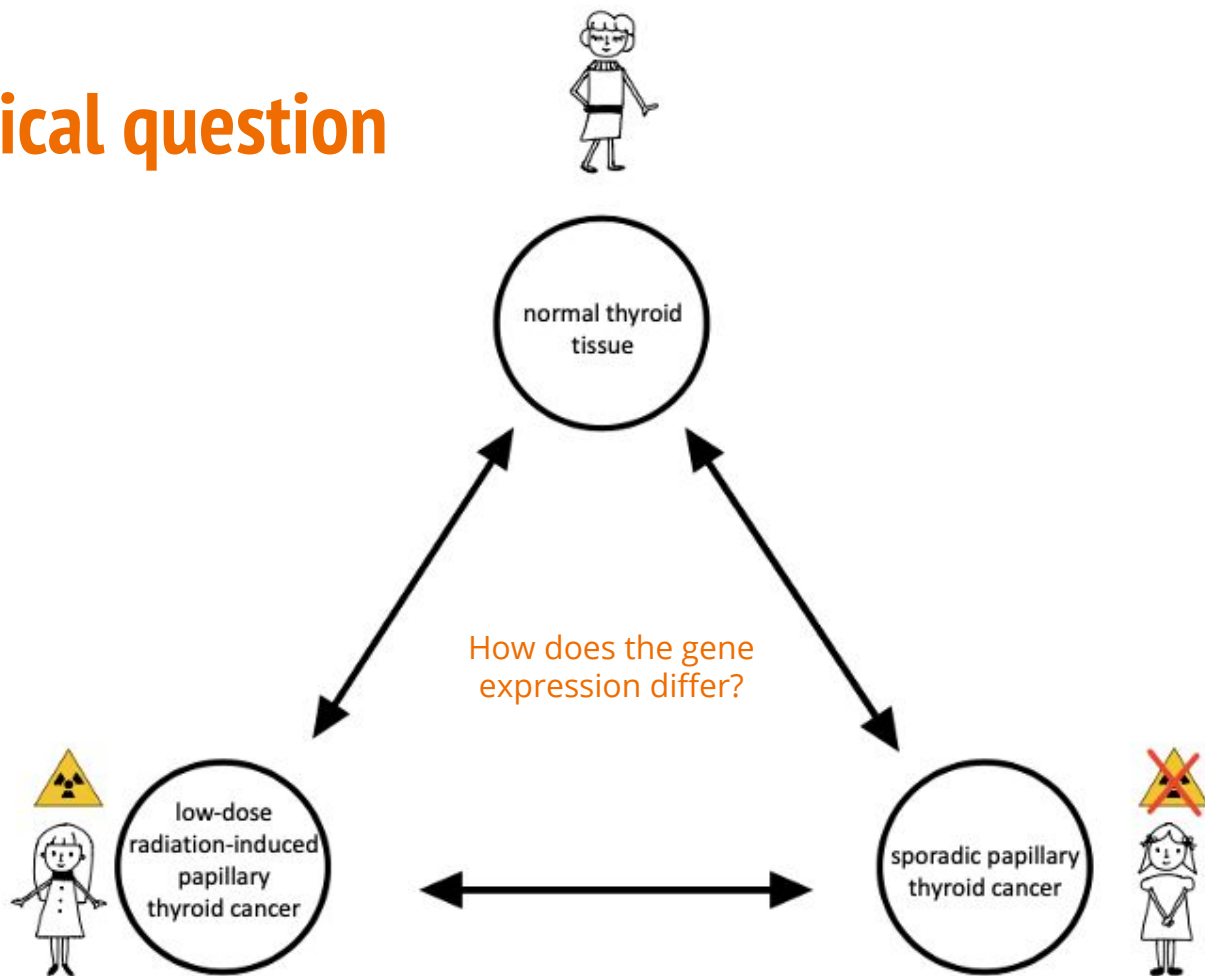
Objective

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:

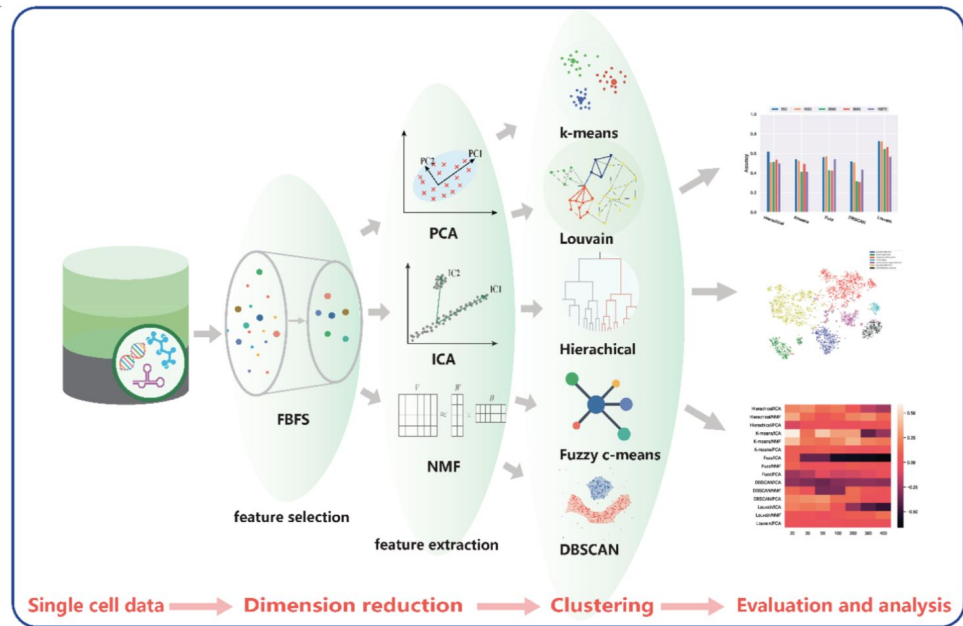1. Data filtering and Cleanup
2. Descriptive statistics
3. Dimension reduction
   - PCA
   - Clustering
4. Differential expression analysis
   - one-way ANOVA
   - pairwise T-test
5. Sorting of genes (Gene ontology)
6. Linear regression analysis and proportion test

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
Camila + Line
(R-Markdown: Carina)

Filtering and data cleanup

- Selecting samples of interest
- Cleaning data

Week 2
Carina + Camila
(R-Markdown: Line)

Descriptive Statistics

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

Week 3
Carina + Line
(R-Markdown: Camila)

Dimension Reduction

- Clustering (k-means)
- Co-expression analysis

Week 4
Camila + Carina
(R-Markdown: Line)

Expression analysis of different groups

- one-way ANOVA
- post hoc test
- dendrogram and heatmap

Week 5
Camila + Line
(R-Markdown: Carina)

Expression and interpretation of different groups

- Genes downregulated or upregulated?
- Gene ontology - relationship between genes

Week 6
Line + Carina
(R-Markdown: Camila)

Sorting of genes

- Function
- Location

Week 7
Carina + Camila
(R-Markdown: Line)

Regression analysis

- between certain characteristics of samples and their gene expression
- Are there changes do to radiation?

Week 8
Camila + Line
(R-Markdown: Carina)

Linear regression analysis evaluation.

- Evaluation of the results
- F-Test

Week 9
Carina + Line
(R-Markdown: Camila)

Evaluation

- Correcting Mistakes
- Summarizing Projekt
- Conclusions

Week 10
Camila, Line and
Carina

Final Presentation