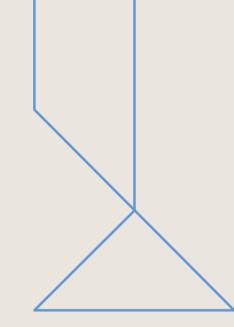


# BB1000 Programming in Python Lab 2: Data analysis

Josefine H. Andersen





## Introduction to Lab 2



## Background

The Human Protein Atlas:<sup>1</sup>

- A program initiated with the aim to map all the human proteins in cells, tissues, and organs
- Knowledge resource with open access data
- 12 sections on particular aspects of the genome-wide analysis of the human proteins



#### The pathology section:<sup>2</sup>

- Explores the gene expression profiles of human cancers
- mRNA and protein expression data from 17 different forms of human cancer
- Correlation between mRNA expression and cancer patient survival



## The project

Replicate the results from the pathology section

- Survival analysis: (Un)favorable prognostic genes
- Specificity of a gene in cancer types



#### You will need to:

- Have a general understanding of what gene expression is
- Become familiar with the content in the Pathology section
- Understand what Kaplan-Meier survival estimators are
- Understand what log rank tests are



## The purpose

#### Practical knowledge:

- Python for data analysis
- Statistical analysis

#### Fundamental knowledge:

- The process behind conclusions
- Critical approach to conclusions
- How to understand, limit, and carry out a (small) project



## The process

**Today**: A warm-up to Pandas and the HPA

Before next lab: Self study

Next lab: Independent work (data will be provided)

Pro tip: work in pairs!



#### Links and references

The Human Protein Atlas (HPA): https://www.proteinatlas.org/

The Cancer Genome Atlas (TCGA): https://www.cancer.gov/ccg/research/genome-sequencing/tcga

Uhlen et al., A pathology atlas of the human cancer transcriptome. Science 357 (2017). DOI: 10.1126/science.aan2507

Weinstein, J., Collisson, E. et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45, (2013). DOI: 10.1038/ng.2764

