22126: Next Generation Sequencing Analysis
DTU - January 2026
Mick Westbury

Mick Westbury
Associate Professor
Section of Bioinformatics
Technical University of Denmark
micwe@dtu.dk

Who am I?

- Undergrad + honours – University of Otago, New Zealand
- PhD in Evolutionary biology – University of Potsdam, Germany
- Postdoc – University of Copenhagen
- Assistant professor – University of Copenhagen
- Associate professor (since Aug. 2025) – DTU

Research interests

- Rise and fall of biodiversity
- Conservation
- Improve study of non-model organisms
- Cross-disciplinary approaches

Key research themes

- **Palaeogenomics (ancient DNA)**:
  - Species extinction, adaptation
- **Conservation Genomics**
  - Genomic erosion/rescue
- **Methodological Innovation**:
  - Developing pipelines for low-quality DNA

Study systems

Extinct species

Endangered species

Who are we?

- **DTU**
  - Mick Westbury (organiser)
  - Mads Hartmann (TA)
  - Amanda Gammelby Qvesel (TA)
  - Kristoffer Vitting-Seerup
  - Rasmus Lykke Marvig

- **University of Copenhagen**
  - Martin Sikora
  - Juan Rodríguez
  - David Duchene
  - Luke Holman

Feedback

- My 1st time teaching this course
- Working on improving
- NGS is very broad now, no one masters everything
- Give us feedback
  - Please do the evaluation at DTU Inside

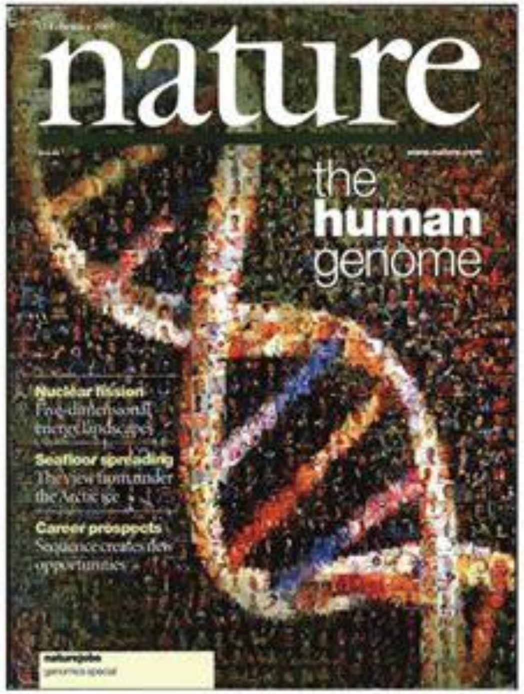
Why we (next generation) sequence?

Why we sequence

- DNA encodes all genetic information
- Differences in DNA explain variation between
  - Individuals
  - Species
  - Diseases
- Sequencing is how we read this information.

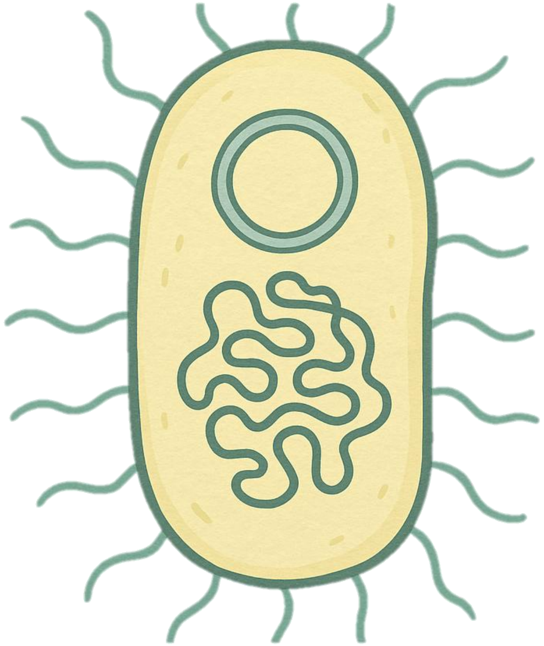
Human genome project (HGP)

- Pre-next generation sequencing
  - Sanger sequencing + BAC clones
- Started 1990
- First draft in 2001
- Completed 2004.
- Cost: ~5.5 billion USD (adjusted)
- Involved 20 research centres
  - 6 countries



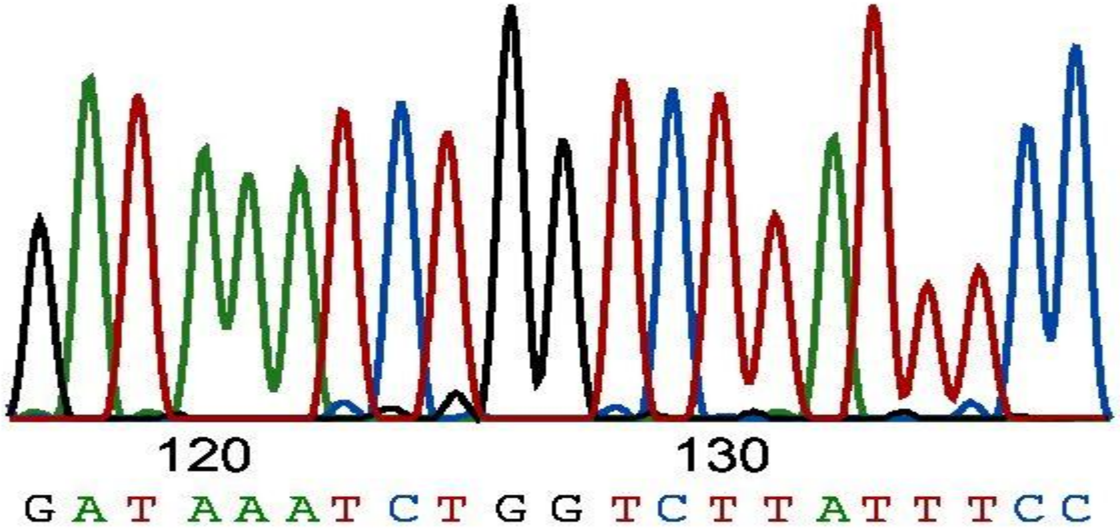
HGP – long DNA problem

- Human genome is ~3.2 billion bases long
- Challenge: How to break DNA, sequence the pieces, and reconstruct the original genome?
- BAC (Bacterial Artificial chromosomes)
- Break into long fragments
  - Break long into smaller
  - Insert to bacteria and clone
  - Sequence plasmids



HGP – long DNA problem

- Technology of the time ~700-1,000bp
  - Sanger sequencing
  - 3.2 million single sequencing reactions = $$$



HGP - shotgun sequencing

- Developed at scale by Celera Genomics during the HGP
- Randomly fragmenting DNA into small pieces and
- Assembling them with overlaps.
- First demonstration of high-throughput
- Shotgun sequencing introduced
  - Massive parallelization
  - Volume over accuracy
  - Ideas that directly led to NGS

First NGS genome

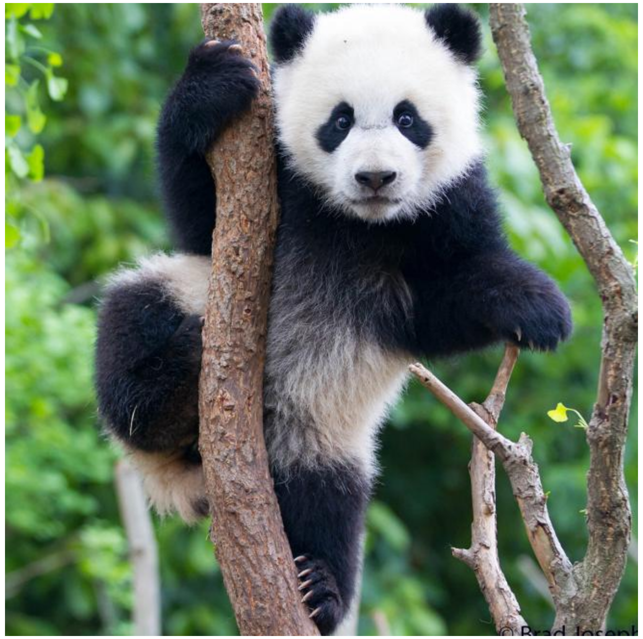
Article | Open access | Published: 13 December 2009

The sequence and *de novo* assembly of the giant panda genome

Ruiqiang Li, Wei Fan, Geng Tian, Hongmei Zhu, Lin He, Jing Cai, Quanfei Huang, Qingle Cai, Bo Li, Yinqi Bai, Zhihe Zhang, Yaping Zhang, Wen Wang, Jun Li, Fuwen Wei, Heng Li, Min Jian, Jianwen Li, Zhaolei Zhang, Rasmus Nielsen, Dawei Li, Wanjun Gu, Zhentao Yang, Zhaoling Xuan, … Jun Wang ✉

Nature **463**, 311–317 (2010) | Cite this article

93k Accesses | **1038** Citations | **175** Altmetric | Metrics

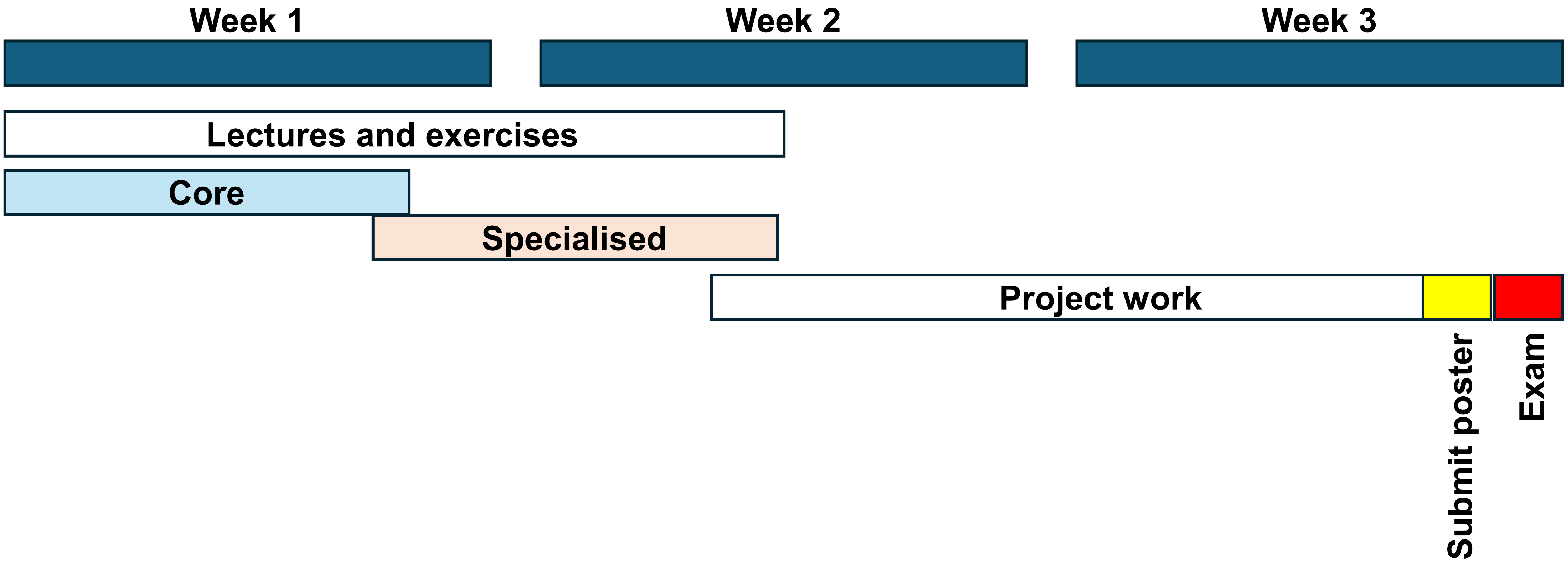


Why sequencing matters today

- Rapid outbreak response
- Cancer profiling and precision oncology
- Rare disease diagnostics
- Prenatal cell-free DNA screening
- Microbiome and ecosystem studies
- Evolution, population genomics, phylogenomics

Course information

Course overview



Upcoming course topics

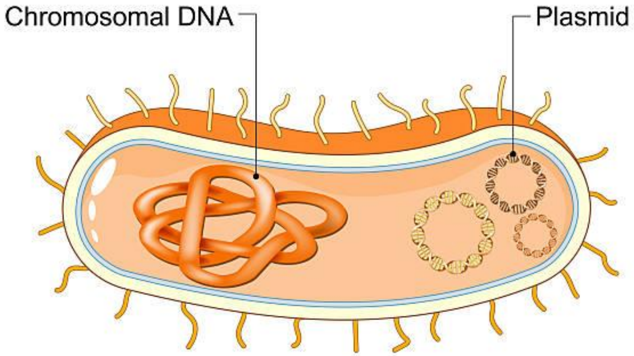
- **Day 1 (core)**
  - Basics of NGS
  - Unix introduction
  - First look at data

- **Day 2 (core)**
  - Data basics and preprocessing
  - Alignments

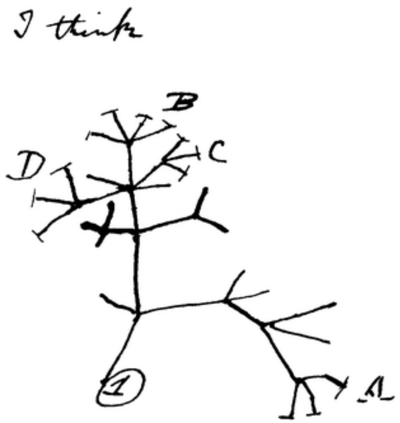
Upcoming course topics

- **Day 5 (specialised)**
  - 3D genomics
  - Ancient DNA

- **Day 6 (specialised)**
  - Microbial genomics



- **Day 7 (specialised)**
  - Phylogenomics



Upcoming course topics

- **Day 8**
  - Metabarcoding **(specialised)**
  - Project introduction

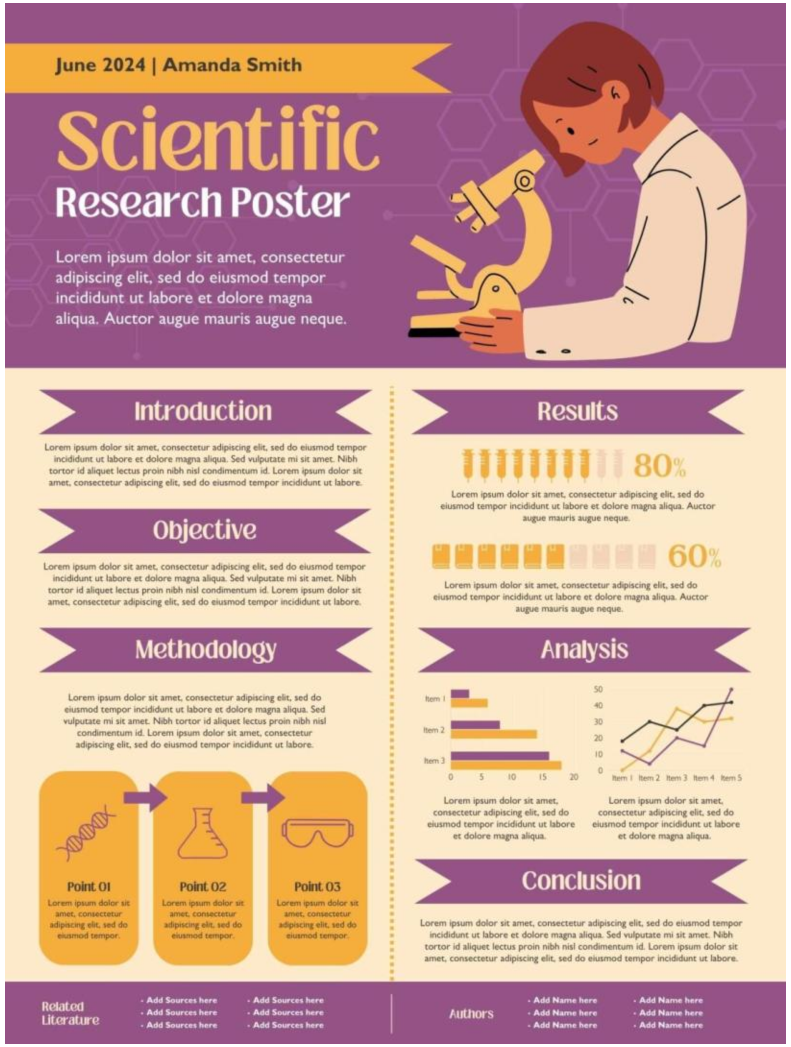
- **Days 9-13**
  - Project work

Upcoming course topics

- **Day 14**
  - Project hand-in
- **Day 15**
  - Written exam

Projects

- Analyse an empirical dataset
- Present results as a poster
- 5-6 per group
- You can find a dataset on SRA/ENA
- Empirical data
  - Recreate a previous study
  - Use your own data if group agrees
- Do **not** analyse very large datasets
  - Time + resources = limited



June 2024 | Amanda Smith

Scientific
Research Poster

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Auctor augue mauris augue neque.

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed vulputate mi sit amet. Nibh tortor id aliquet lectus proin nibh nisl condimentum id. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

Objective

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed vulputate mi sit amet. Nibh tortor id aliquet lectus proin nibh nisl condimentum id. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

Methodology

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed vulputate mi sit amet. Nibh tortor id aliquet lectus proin nibh nisl condimentum id. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

Point 01
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor.

Point 02
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor.

Point 03
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor.

Results

80%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Auctor augue mauris augue neque.

60%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Auctor augue mauris augue neque.

Analysis

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Sed vulputate mi sit amet. Nibh tortor id aliquet lectus proin nibh nisl condimentum id. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

Related Literature
- Add Sources here
- Add Sources here
- Add Sources here
- Add Sources here
- Add Sources here
- Add Sources here

Authors
- Add Name here
- Add Name here
- Add Name here
- Add Name here
- Add Name here
- Add Name here

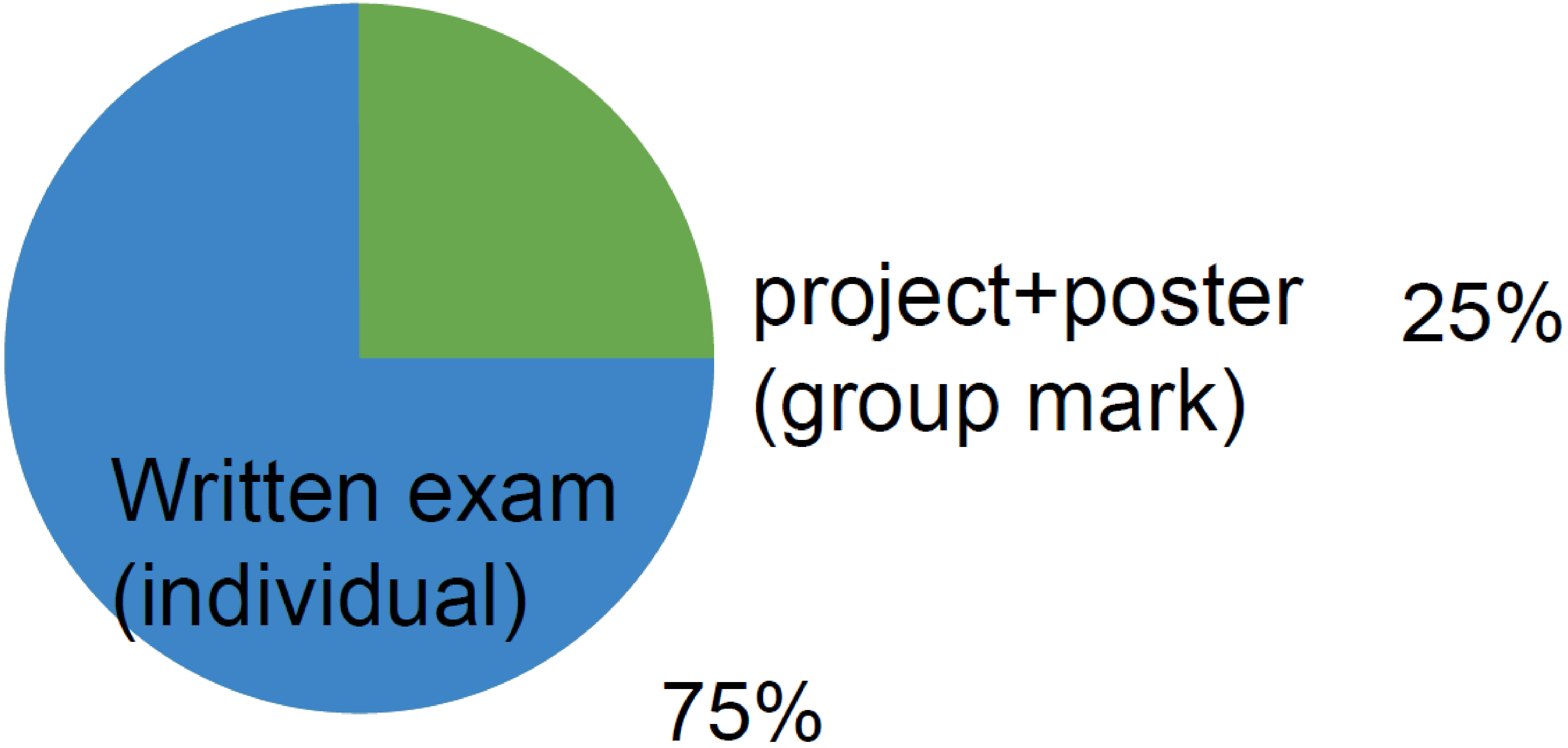
Projects

- Each project group will create a poster
- The goal of the project is:
  - Do not memorise
  - Understand what you are doing during the project
  - Understand the concepts taught in class
  - Learn NGS from firsthand experience
- Please send the PDF **before noon on Thursday the 22nd January**

Written Exam

- You cannot write the exam if you have not submitted the poster
- Multiple choice exam
- Focuses on the core lectures
- One basic question per specialised lecture
- Euroteq = oral exam

Marking scheme



Points to remember

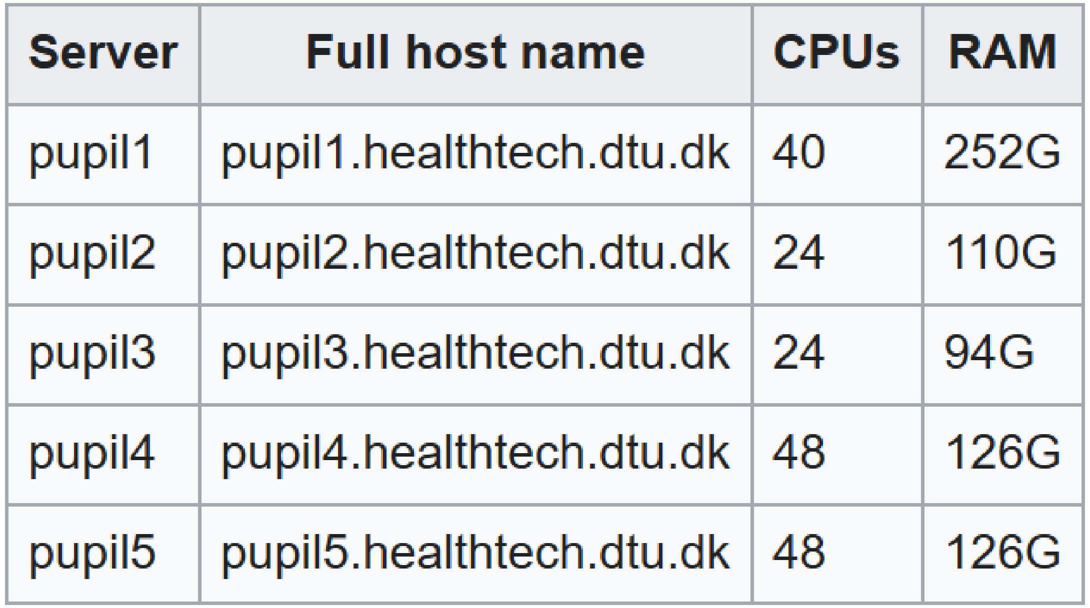
- Understand principles of the analysis
- The exercises will be useful for your projects and hopefully later
- You don’t need to do all the exercises
  - However, the core lectures are important

Points to remember

- Have an exercise buddy and do them as a team
    - On each individuals’ laptop
    - Everyone learns the command-line
- Ask questions at any time!
- You get the solutions for the exercises but **do not copy-paste!!**
- You will not get to copy-paste for the project
- Read the **error message**

Cloud/HPC computing

- Pupil cluster
  - 5 nodes
- Be careful with disk space
- Limited computational power
- Need software installed?
  - Ask me!



Server	Full host name	CPUs	RAM
pupil1	pupil1.healthtech.dtu.dk	40	252G
pupil2	pupil2.healthtech.dtu.dk	24	110G
pupil3	pupil3.healthtech.dtu.dk	24	94G
pupil4	pupil4.healthtech.dtu.dk	48	126G
pupil5	pupil5.healthtech.dtu.dk	48	126G

Course webpage

- Course program, slides, handouts, exercises etc.

- https://teaching.healthtech.dtu.dk/22126/index.php/Program_2026

- The course page is a repository for you

Discord

- Chat with others during off-hours. Create channels!
- Please use your real name
- https://discord.gg/mgFJxM3W

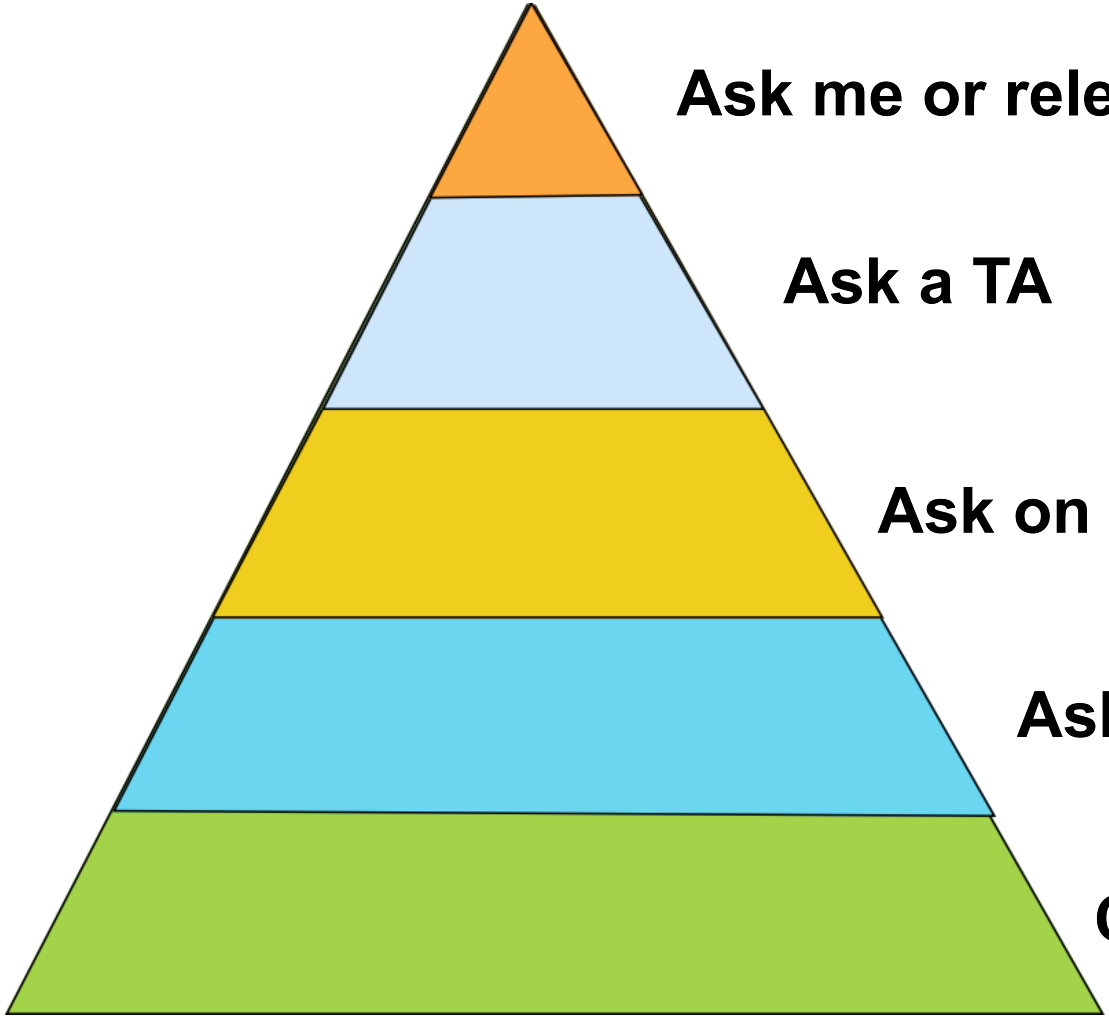
Tips

- Do not memorise definitions
  - understand concepts
- The core lectures are especially crucial
- Do not memorise software names
- Understand what you are doing
  - inspect the input
  - inspect the output
  - play with parameters
- Read the error messages

Learning objectives

- Are you able to:
  - work in group and delegate tasks?
  - set realistic objectives?
  - use the command line?
  - understand the strength and weakness of each tool?
  - explain key steps in a critical manner?

Any Questions?



Ask me or relevant lecturer

Ask a TA

Ask on discord/your neighbour

Ask ChatGPT/google or equivalent

Check the error message