WELCOME

27626 & 27826: Next Generation Sequencing Analysis
DTU - June 2017
Josef K Vogt

Josef Korbinian Vogt
Assistant Professor
DTU Bioinformatics
Technical University of Denmark
josef@cbs.dtu.dk

27626 - Next Generation Sequencing Analysis
Who are we?

- **Organizers:**
  - Josef Korbinian Vogt (Me, main teacher 2017)
  - Simon Rasmussen (Course responsible)
  - Thomas Sicheritz-Ponten

- **CBS-teachers:**
  - Peter Wad Sackett
  - Kosai Al-Nakeeb
  - Henrike Zschach
  - Jose Armenteros
  - Jakob Nissen

- **Lasse Følkersen**
- **Jose Izarzugaze (Txema)**
- **Marcin Krzystanek**
- **Aron Eklund**

- **DTU Food**
  - Frank Aarestrup

- **University of Copenhagen**
  - Kristian Hanghøj
  - (Mikkel Schubert)
Who am I?

• Msc in Biotechnology from DTU

• PhD in bioinformatics

• Assistant Professor - Metagenome and Microbiome Analysis course (3 years), NGS analysis (first time)

• Metagenomics, Human Microbiome, Machine Learning
Who is Simon?

- Msc in Chemistry/Biotechnology from DTU
- Research scientist SSI - TB vaccine development
- PhD in bioinformatics
- Associate Professor - NGS analysis (6 years)
  - Human genetics, Metagenomics, Bacterial evolution
Who is Thomas?

- Co-organizer
- Professor in Metagenomics
- Metagenomics group leader
Teaching Assistants

Kosai Al-Nakeeb  
PhD student  
Metagenomics

Henrike Zschach  
PhD student  
Immunoinformatics and Machine Learning

Jakob Nissen  
PhD student  
Metagenomics

Jose Armenteros  
PhD student  
Disease Intelligence and Molecular Evolution
Who are you?

- According to Campusnet
- Bsc students: 5
- Msc students: 22
- PhD students: 17
- Open education: 3
Feedback

• Fifth time we are running the course
• We are still improving - quite some changes from last year

• Please give us feedback!
  • Please do the evaluation at DTU inside (Campusnet)
Learning objectives

Cooperate in groups

Strength / Weaknesses

Analytical & reflective

Applications

Formulate/perform a project

NGS

Explain key steps

Theoretical principles

Unix command line
Why shell terminal?

• Almost all tools for NGS analysis are command line only

• Generally more efficient/flexible, you can play around with the tools/data

• They can be pipelined, ie. analyzing 100 files in windowed mode is a pain ...

• Alternative approaches: Galaxy, CLC-workbench
Course structure

• 3 weeks, 2 tracks:

Lectures + Exercises + Pres.

Project work

June: 1. 13. 22.

= Poster exam in front of CBS
Course program

- course outline: http://www.cbs.dtu.dk/courses/27626/
- Breaks are included in the program
- Coffee is provided by CBS:
  - cart in front of CBS door
Course breakdown

- **1st June**
  - Introduction NGS tech.
  - Tech talk groups
  - Unix and first look at data

- **2nd June**
  - Tech talk presentations
  - Data basics & preprocessing
Course breakdown

• 6th June
  • Alignment
  • Alignment processing
  • Variant calling

• 7th June
  • De novo assembly
  • De novo metagenomics
Course breakdown

- **8th June**
  - Quantitative Metagenomics
  - RNAseq
- **9th June**
  - Test + recap
  - Cancer seq
- **12th June**
  - Genomic Epidemiology
  - Ancient DNA
Course breakdown

- **13th June**
  - Project group formation
- **14th June**
  - Short project presentations
  - Project work
- **15th - 21st June**
  - Project work
- **22nd June**
  - Poster Exam
Some points

- Learn principles of the analysis
- The exercises will be useful for your projects and later
- Team up 2 and 2 for exercises (or do them yourself but discuss with neighbour)
- Please just ask questions at any time!
Cloud computing

• For the first time moved course to Cloud!

• Danish National Supercomputer for Life Science (Computerome) located at DTU Risø

• 16048 cores, 92 Tb RAM an 3Pb storage

• We have 2 nodes

• 28 cores, 128 Gb RAM
Projects

• Try analyse a “real” dataset and present results on poster

• 4-5 pr. group

• You can find a dataset on SRA/ENA

• You can use your own data if everyone in the group agrees and it can be presented on a poster

• Don’t analyse too large datasets (time, resources)
Projects

• Teachers and TA will be available to help you with your projects

• ‘Office hours’ during project period: 1pm - 3pm

• Use Piazza as a platform to communicate with your peers, TAs and teachers:
  • Use it as discussion platform to benefit from your collective knowledge
  • Simon will follow remotely if any problems arise
  • You will be granted access to the course platform today
Exam

• Each group will create a poster

• You can print posters at the DTU library for 20-30 kr

• Each group will present the poster for the examiners

• Then each individual in the group will one-by-one be asked questions on the learning objectives and your project (5-10min).
Disclaimer

- Sequencing technology changes very rapidly!
- We will dive into many areas - you will not master everything

but

- There are many opportunities - hopefully you will learn to see them
Be adventurous!

You do not have the ability to do anything destructive

Unless you physically destroy our computers!

The worst that can happen is that you lose your own data

Source: Angus
Course web-page

• Course program, Slides, Handouts, Exercises

• http://www.cbs.dtu.dk/courses/27626/index.php

• We want the course page to be a repository for you!
Reading + wifi

• There are no text-book for the course

• There are papers uploaded to campusnet that you can read for more information

• Wireless network
  • Use “dtu” and your dtu/campusnet login to get access to wireless
  • Alternative wifi: “You can haz wifi”
Pre-test

• Test your knowledge before we start

• Not used for grading or exam

• Used to understand where you are