DTU

DTU Health Technology
Bioinformatics

22126: Next Generation Sequencing Analysis
DTU - January 2024
Gabriel Renaud

Gabriel Renaud
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gabriel.reno@gmail.com

Who am I?

- PhD in Bioinformatics from Max Planck Institute for Evolutionary Anthropology in Leipzig
- Postdoc at KU
- Associate Professor at DTU in Dec. 2019
- Worked since 2006 with NGS
- slow response: gabre [at] dtu [dot] dk
- fast response: gabriel [dot] reno [at] gmail [dot] com

Who am I?

How to contact me?

- slow response: gabre [at] dtu [dot] dk
- medium response: gabriel [dot] reno [at] gmail [dot] com
- fastest response: Discord

Conflict of interest: none, I do not own any stocks or consulting for any sequencing company

Who am I?

What keeps me busy:

- Methods for NGS analysis
- Ancient DNA and modern samples
- Large sets of genotypes
- Pangenomes

Looking to do a special project/masters' project dealing with NGS, email me!

Who are we?

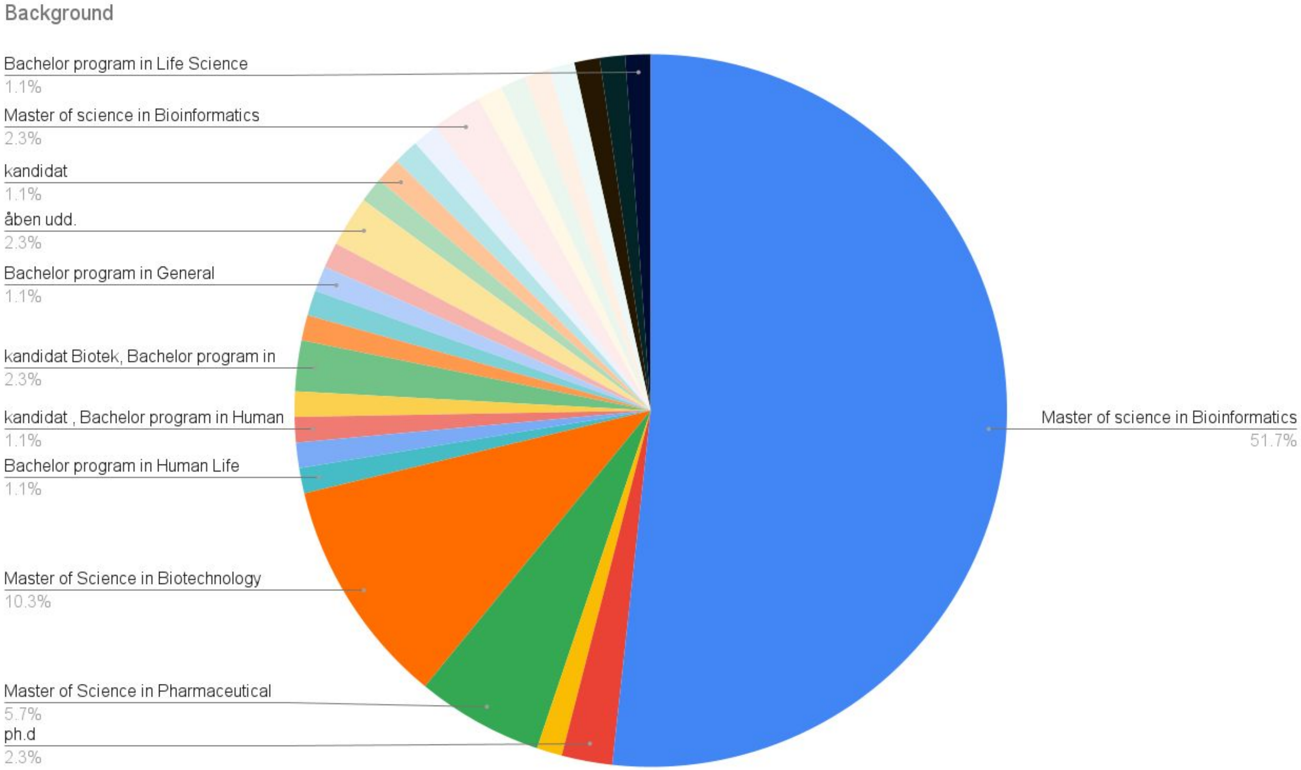
- Organizer:
  - Gabriel Renaud
  - Kristoffer Vitting-Seerup
  - Ole Lund
  - Frederikke Pedersen
  - Astrid Saksager
  - Grigorii Nos
  - DTU Bioinformatics
  - Peter Wad Sackett
- DTU Food
  - Pimlapas Leekitecharoenphon (Shinny)
- Copenhagen University:
  - Martin Sikora

Main teaching assistants

Astrid Saksager
Grigorii Nos

Who are you?

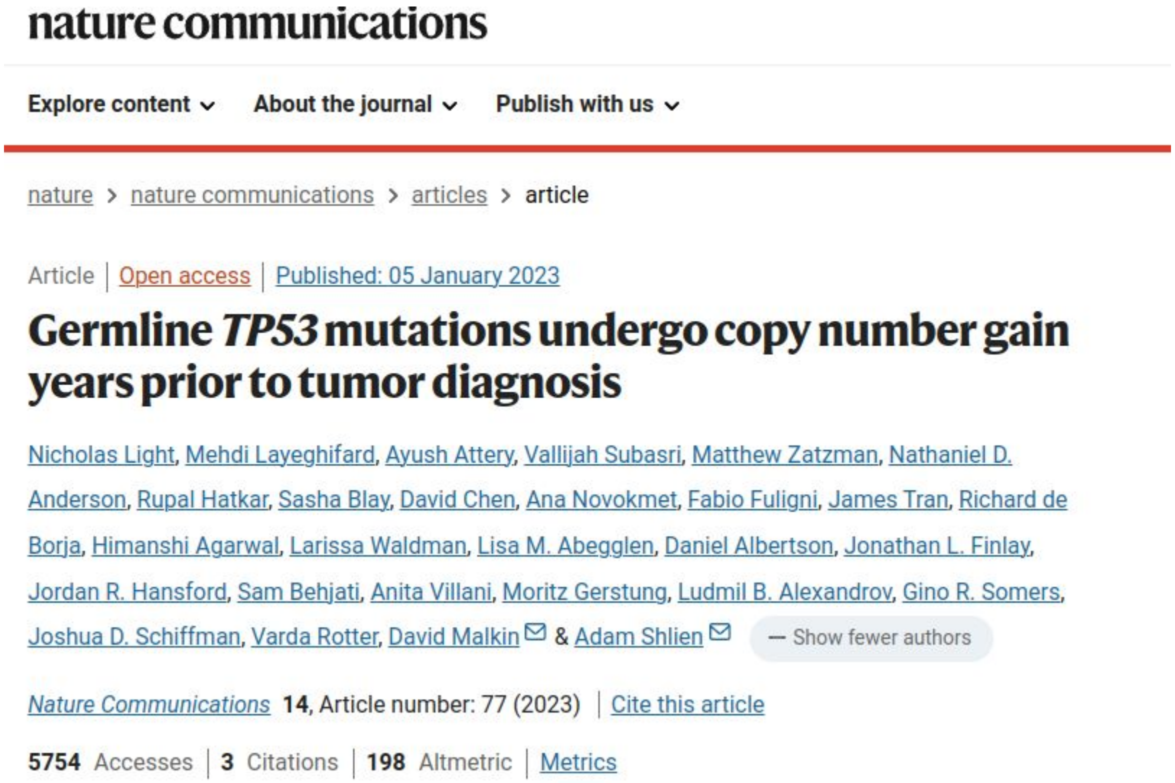
January 2024



Feedback

- My 5th time! 3rd time in person.
- We are still improving
- It is very difficult to keep up with new tech…
- NGS is very broad now, no one masters everything
- Please give us feedback !
  - – Please do the evaluation at DTU Inside

Why are we here?



nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

Article | Open access | Published: 05 January 2023

Germline *TP53* mutations undergo copy number gain years prior to tumor diagnosis

Nicholas Light, Mehdi Layeghifard, Ayush Attery, Vallijah Subasri, Matthew Zatzman, Nathaniel D. Anderson, Rupal Hatkar, Sasha Blay, David Chen, Ana Novokmet, Fabio Fuligni, James Tran, Richard de Borja, Himanshi Agarwal, Larissa Waldman, Lisa M. Abegglen, Daniel Albertson, Jonathan L. Finlay, Jordan R. Hansford, Sam Behjati, Anita Villani, Moritz Gerstung, Ludmil B. Alexandrov, Gino R. Somers, Joshua D. Schiffman, Varda Rotter, David Malkin ✉ & Adam Shlien ✉ — Show fewer authors

Nature Communications **14**, Article number: 77 (2023) | Cite this article

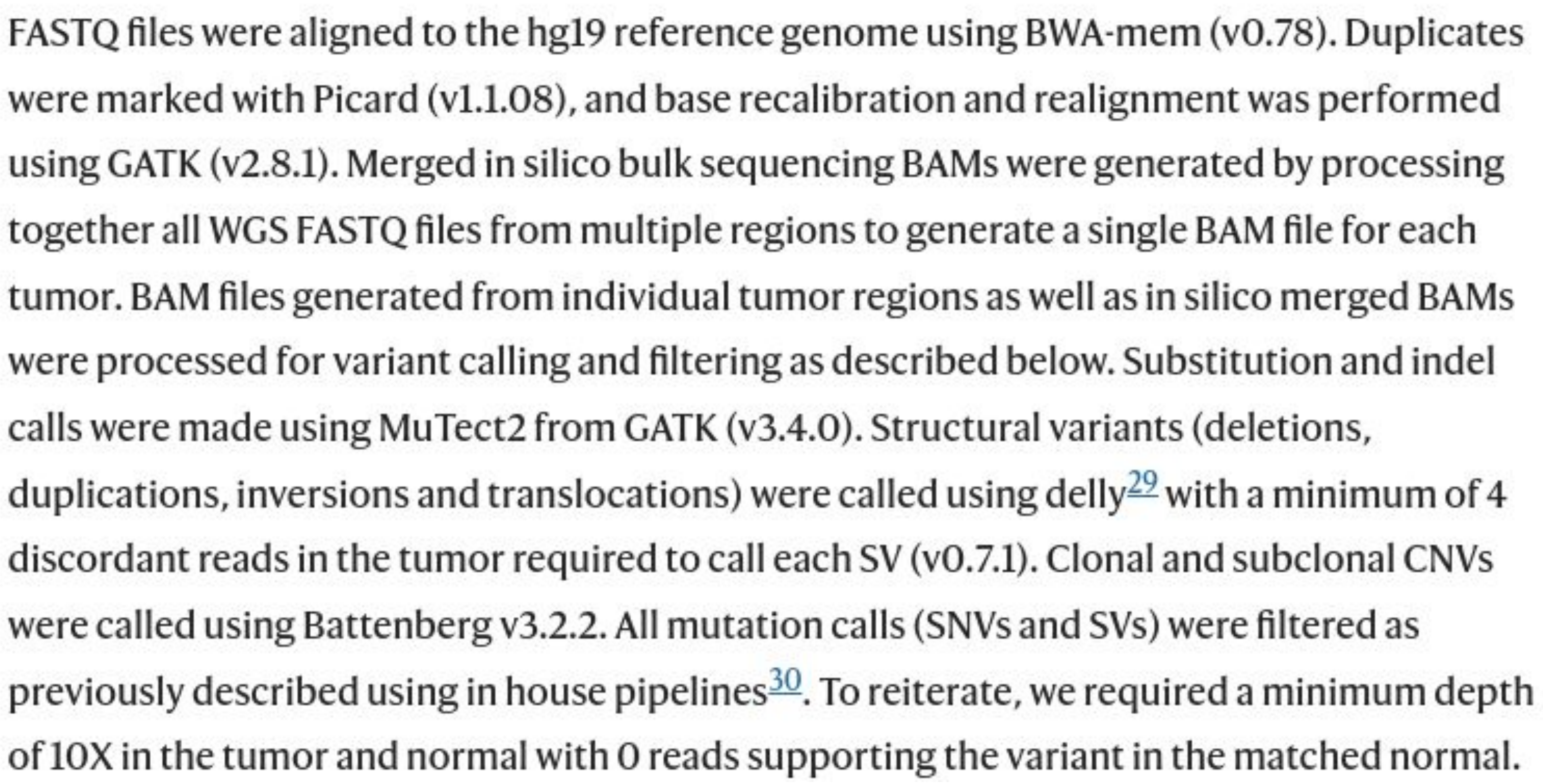
5754 Accesses | **3** Citations | **198** Altmetric | Metrics

Findings:

In people with Li-Fraumeni syndrome (a genetic condition that increases cancer risk), genetic mutations, occur early and are common in tumors, many years before cancer is diagnosed.

Published: 03 Jan 2023

Why are we here?



FASTQ files were aligned to the hg19 reference genome using BWA-mem (v0.78). Duplicates were marked with Picard (v1.1.08), and base recalibration and realignment was performed using GATK (v2.8.1). Merged in silico bulk sequencing BAMs were generated by processing together all WGS FASTQ files from multiple regions to generate a single BAM file for each tumor. BAM files generated from individual tumor regions as well as in silico merged BAMs were processed for variant calling and filtering as described below. Substitution and indel calls were made using MuTect2 from GATK (v3.4.0). Structural variants (deletions, duplications, inversions and translocations) were called using delly[29] with a minimum of 4 discordant reads in the tumor required to call each SV (v0.7.1). Clonal and subclonal CNVs were called using Battenberg v3.2.2. All mutation calls (SNVs and SVs) were filtered as previously described using in house pipelines[30]. To reiterate, we required a minimum depth of 10X in the tumor and normal with 0 reads supporting the variant in the matched normal.

Why are we here?

"Around 2 a.m. on Jan. 5, after working over 40 hours straight, Dr. Zhang and his team at the Shanghai Public Health Clinical Center sequenced the unknown virus on the NovaSeq™ 6000 System. They published its genome on **Jan. 10th 2020**."

https://www.illumina.com/company/news-center/blog/2020-in-genomics.html

Yong-Zhen Zhang

Why are we here?

"... Moderna's mRNA-1273, which reported a 94.5 percent efficacy rate on November 16, had been designed by **January 13th 2020**. This was just **two days** after the genetic sequence had been made public

…

It was completed **[...] more than a week before** the first confirmed coronavirus case in the United States."

https://nymag.com/intelligencer/2020/12/moderna-covid-19-vaccine-design.html

Yong-Zhen Zhang

Not a wet lab course...

...it’s a computational one

Tips

Tip: Do not memorize the name of the tools/procedure, they come and go

Tips

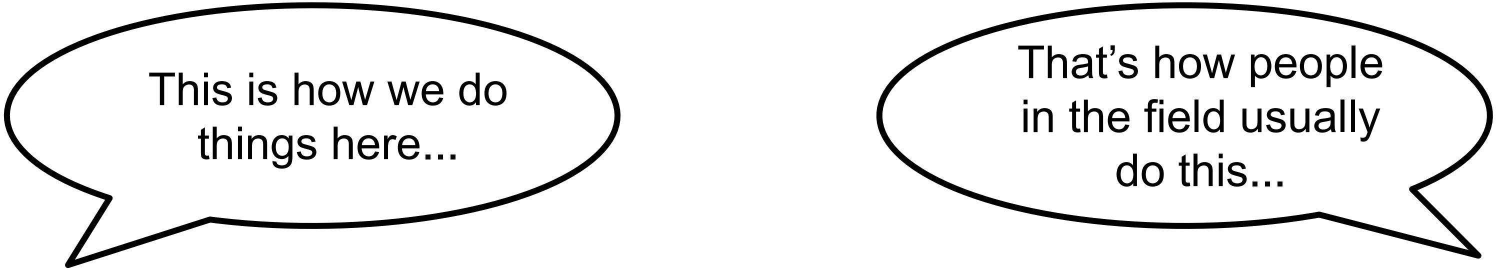
Tip: Understand the problem and how various tools work

Tips for NGS in general

- New tools or procedures get released all the time
- The best tool/format/pipeline in 2024 may not be the best in 2034
- Understand how they work, in which cases they perform well

Tips for NGS in general

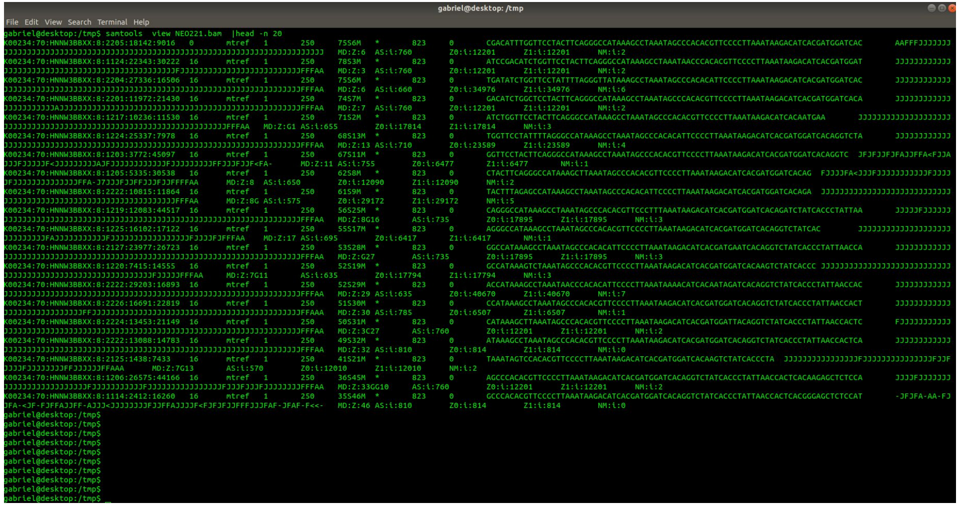
- Read benchmarking papers and reviews
- Beware of:



This is how we do things here...

That’s how people in the field usually do this...

The shell terminal



```
gabriel@desktop:/tmp$ samtools  view NEO221.bam  |head -n 20
K00234:70:HNNW3BBXX:8:2205:18142:9016   0       mtref   1       250     75S6M   *       823     0       CGACATTTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCAC       AAFFFJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ   MD:Z:6  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1124:22343:30222  16      mtref   1       250     78S3M   *       823     0       ATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAACCCACACGTTCCCCTTAAATAAGACATCACGATGGAT       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2204:27336:16506  16      mtref   1       250     75S6M   *       823     0       TGATATCTGGTTCCTATTTTAGGGTTATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCAC       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:6  AS:i:660        Z0:i:34976      Z1:i:34976      NM:i:6
K00234:70:HNNW3BBXX:8:2201:11972:21430  16      mtref   1       250     74S7M   *       823     0       GACATCTGGCTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACA       JJJJJJJJJJJ
JJJJJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:7  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1217:10236:11530  16      mtref   1       250     71S2M   *       823     0       ATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACAATGAA       JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G1 AS:i:655        Z0:i:17814      Z1:i:17814      NM:i:3
K00234:70:HNNW3BBXX:8:1224:25337:7978   16      mtref   1       250     68S13M  *       823     0       TGGTTCCTATTTTAGGGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:13 AS:i:710        Z0:i:23589      Z1:i:23589      NM:i:4
K00234:70:HNNW3BBXX:8:1203:3772:45097   16      mtref   1       250     67S11M  *       823     0       GGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTC  JFJFJJFJFAJJFFA<FJJA
JJJFJJJJJF<JJJJJJJJJAJFJJJJJJJJJJJJFJJJJJJJJJFFJJJFJJF<FA-      MD:Z:11 AS:i:755        Z0:i:6477       Z1:i:6477       NM:i:1
K00234:70:HNNW3BBXX:8:1205:5335:30538   16      mtref   1       250     62S8M   *       823     0       CTACTTCAGGGCCATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAG  FJJJJFA<JJJFJJJJJJJJJJJFJJJJ
JFJJJJJJJJJJJJJJFFA-J7JJJFJJFFJJJFJJFFFFAA      MD:Z:8  AS:i:650        Z0:i:12090      Z1:i:12090      NM:i:2
K00234:70:HNNW3BBXX:8:2222:10815:11864  16      mtref   1       250     61S9M   *       823     0       TACTTTAGAGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGA  JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA      MD:Z:8G AS:i:575        Z0:i:29172      Z1:i:29172      NM:i:5
K00234:70:HNNW3BBXX:8:1219:12083:44517  16      mtref   1       250     56S25M  *       823     0       CAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATGGATCACAGATCTATCACCCTATTAA       JJJJJFJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:8G16       AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1225:16102:17122  16      mtref   1       250     55S17M  *       823     0       AGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCAC        JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJFAJJJJJJJJJJFJJJJJJJJJJJJJJJJJJFJJJJFJFFFAA    MD:Z:17 AS:i:695        Z0:i:6417       Z1:i:6417       NM:i:1
K00234:70:HNNW3BBXX:8:2127:23977:26723  16      mtref   1       250     53S28M  *       823     0       GGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGAATCACAGGTCTATCACCCTATTAACCA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G27        AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1220:7415:14555   16      mtref   1       250     52S19M  *       823     0       GCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCC JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJFFFAA     MD:Z:7G11       AS:i:635        Z0:i:17794      Z1:i:17794      NM:i:3
K00234:70:HNNW3BBXX:8:2222:29203:16893  16      mtref   1       250     52S29M  *       823     0       ACCATAAAGCCTAAATAACCCACACATTCCCCTTAAATAAAACATCACAATAGATCACAGGTCTATCACCCTATTAACCAC       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:29 AS:i:635        Z0:i:40670      Z1:i:40670      NM:i:7
K00234:70:HNNW3BBXX:8:2226:16691:22819  16      mtref   1       250     51S30M  *       823     0       CCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACT       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFAAA   MD:Z:30 AS:i:785        Z0:i:6507       Z1:i:6507       NM:i:1
K00234:70:HNNW3BBXX:8:2224:13453:21149  16      mtref   1       250     50S31M  *       823     0       CATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATTACAGGTCTATCACCCTATTAACCACTC       FJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3C27       AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2222:13088:14783  16      mtref   1       250     49S32M  *       823     0       ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:32 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
K00234:70:HNNW3BBXX:8:2125:1438:7433    16      mtref   1       250     41S21M  *       823     0       TAAATAGTCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCCTA  JJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJF
JJJJFJJJJJJJJFFJJJJJJFFAAA      MD:Z:7G13       AS:i:570        Z0:i:12010      Z1:i:12010      NM:i:2
K00234:70:HNNW3BBXX:8:1206:26575:44166  16      mtref   1       250     36S45M  *       823     0       AGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACAAGAGCTCTCCA       JJJJFJJJJJJ
JJJJJJJJJJJJJJJJJFJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJFJJJFJJJFJJJJJJJJFFFAA   MD:Z:33GG10      AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1114:2412:16260   16      mtref   1       250     35S46M  *       823     0       GCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCAT       -JFJFA-AA-FJ
JFA-<JF-FJFFAJJFF-AJJJ<JJJJJJJJFJJFFAJJJJF<FJFJFJJFFFJJJFAF-JFAF-F<<-   MD:Z:46 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
```

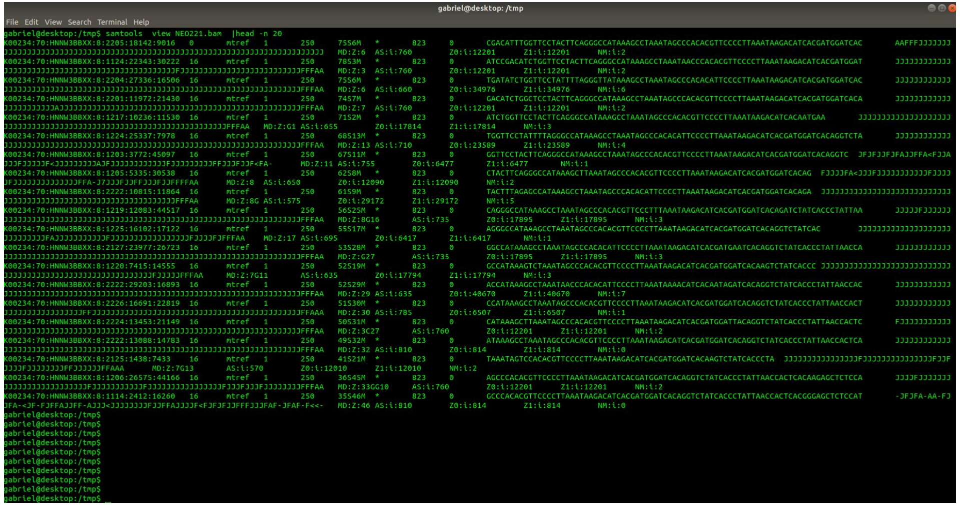
- Terminal allows users to interact with the computer using commands in the format:

```
command argument_1 argument_2
```

- Examples:

```
ls -al
pwd
```

The shell terminal



```
gabriel@desktop:/tmp$ samtools  view NEO221.bam  |head -n 20
K00234:70:HNNW3BBXX:8:2205:18142:9016   0       mtref   1       250     75S6M   *       823     0       CGACATTTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCAC       AAFFFJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ   MD:Z:6  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1124:22343:30222  16      mtref   1       250     78S3M   *       823     0       ATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAACCCACACGTTCCCCTTAAATAAGACATCACGATGGAT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2204:27336:16506  16      mtref   1       250     75S6M   *       823     0       TGATATCTGGTTCCTATTTTAGGGTTATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:6  AS:i:660        Z0:i:34976      Z1:i:34976      NM:i:6
K00234:70:HNNW3BBXX:8:2201:11972:21430  16      mtref   1       250     74S7M   *       823     0       GACATCTGGCTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACA       JJJJJJJJJJJJ
JJJJJJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:7  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1217:10236:11530  16      mtref   1       250     71S2M   *       823     0       ATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACAATGAA       JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G1 AS:i:655        Z0:i:17814      Z1:i:17814      NM:i:3
K00234:70:HNNW3BBXX:8:1224:25337:7978   16      mtref   1       250     68S13M  *       823     0       TGGTTCCTATTTTAGGGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:13 AS:i:710        Z0:i:23589      Z1:i:23589      NM:i:4
K00234:70:HNNW3BBXX:8:1203:3772:45097   16      mtref   1       250     67S11M  *       823     0       GGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTC  JFJFJJFJFAJJFFA<FJJA
JJJFJJJJJF<JJJJJJJJJAJFJJJJJJJJJJJJFJJJJJJJJJFFJJJFJJF<FA-      MD:Z:11 AS:i:755        Z0:i:6477       Z1:i:6477       NM:i:1
K00234:70:HNNW3BBXX:8:1205:5335:30538   16      mtref   1       250     62S8M   *       823     0       CTACTTCAGGGCCATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAG  FJJJJFA<JJJFJJJJJJJJJJJFJJJJ
JFJJJJJJJJJJJJJJFFA-J7JJJFJJFFJJJFJJFFFFAA      MD:Z:8  AS:i:650        Z0:i:12090      Z1:i:12090      NM:i:2
K00234:70:HNNW3BBXX:8:2222:10815:11864  16      mtref   1       250     61S9M   *       823     0       TACTTTAGAGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGA  JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA      MD:Z:8G AS:i:575        Z0:i:29172      Z1:i:29172      NM:i:5
K00234:70:HNNW3BBXX:8:1219:12083:44517  16      mtref   1       250     56S25M  *       823     0       CAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATGGATCACAGATCTATCACCCTATTAA       JJJJJFJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:8G16       AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1225:16102:17122  16      mtref   1       250     55S17M  *       823     0       AGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCAC        JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJFAJJJJJJJJJJJFJJJJJJJJJJJJJJJJJFJJJJFJFFFAA    MD:Z:17 AS:i:695        Z0:i:6417       Z1:i:6417       NM:i:1
K00234:70:HNNW3BBXX:8:2127:23977:26723  16      mtref   1       250     53S28M  *       823     0       GGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGAATCACAGGTCTATCACCCTATTAACCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G27        AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1220:7415:14555   16      mtref   1       250     52S19M  *       823     0       GCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCC JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJFFFAA     MD:Z:7G11       AS:i:635        Z0:i:17794      Z1:i:17794      NM:i:3
K00234:70:HNNW3BBXX:8:2222:29203:16893  16      mtref   1       250     52S29M  *       823     0       ACCATAAAGCCTAAATAACCCACACATTCCCCTTAAATAAAACATCACAATAGATCACAGGTCTATCACCCTATTAACCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:29 AS:i:635        Z0:i:40670      Z1:i:40670      NM:i:7
K00234:70:HNNW3BBXX:8:2226:16691:22819  16      mtref   1       250     51S30M  *       823     0       CCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFAAA   MD:Z:30 AS:i:785        Z0:i:6507       Z1:i:6507       NM:i:1
K00234:70:HNNW3BBXX:8:2224:13453:21149  16      mtref   1       250     50S31M  *       823     0       CATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATTACAGGTCTATCACCCTATTAACCACTC       FJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3C27       AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2222:13088:14783  16      mtref   1       250     49S32M  *       823     0       ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:32 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
K00234:70:HNNW3BBXX:8:2125:1438:7433    16      mtref   1       250     41S21M  *       823     0       TAAATAGTCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCCTA  JJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJF
JJJFJJJJJJJJFFJJJJJJJFFAAA      MD:Z:7G13       AS:i:570        Z0:i:12010      Z1:i:12010      NM:i:2
K00234:70:HNNW3BBXX:8:1206:26575:44166  16      mtref   1       250     36S45M  *       823     0       AGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACAAGAGCTCTCCA       JJJJFJJJJJJJ
JJJJJJJJJJJJJJJJJJFJJJJJJJJJJJFJJJJJJJJJJJJJJJJFJJJFJJJFJJJJJJJJFFFAA   MD:Z:33GG10     AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1114:2412:16260   16      mtref   1       250     35S46M  *       823     0       GCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCAT       -JFJFA-AA-FJ
JFA-<JF-FJFFAJJFF-AJJJ<JJJJJJJJFJJFFAJJJJF<FJFJFJJFFFJJJFAF-JFAF-F<<-   MD:Z:46 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
```

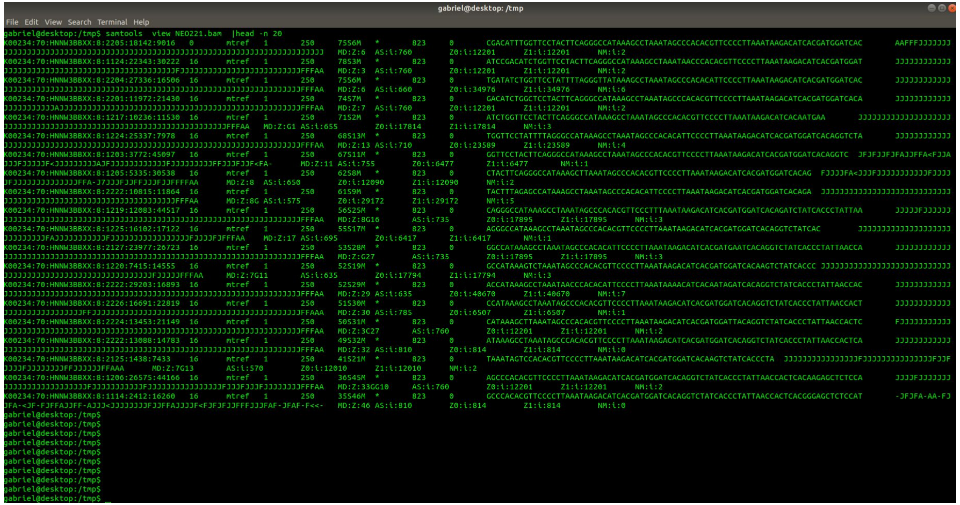
- Available on various platforms

Mac™OS

CentOS

Windows 10

Why the shell terminal

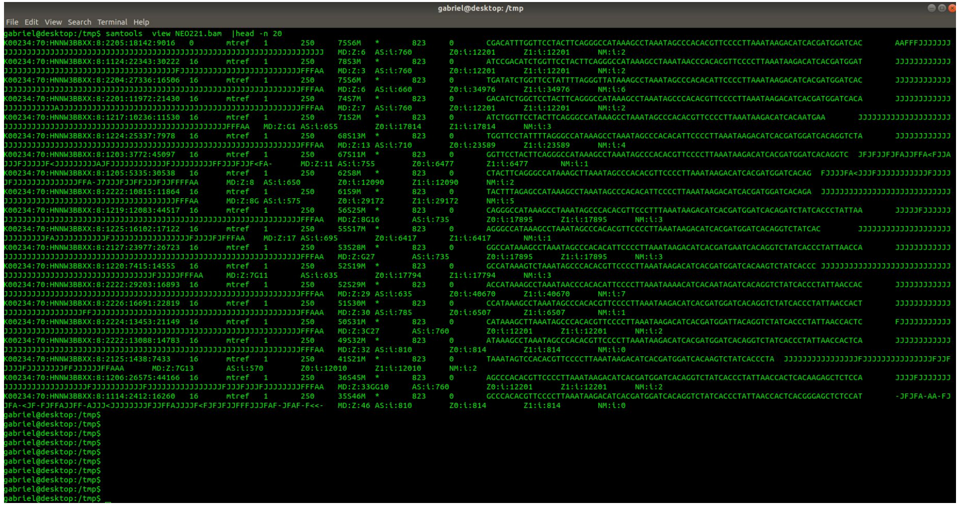


```
gabriel@desktop:/tmp$ samtools  view NEO221.bam  |head -n 20
K00234:70:HNNW3BBXX:8:2205:18142:9016   0       mtref   1       250     75S6M   *       823     0       CGACATTTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCAC       AAFFFJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ   MD:Z:6  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1124:22343:30222  16      mtref   1       250     78S3M   *       823     0       ATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAACCCACACGTTCCCCTTAAATAAGACATCACGATGGAT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2204:27336:16506  16      mtref   1       250     75S6M   *       823     0       TGATATCTGGTTCCTATTTTAGGGTTATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:6  AS:i:660        Z0:i:34976      Z1:i:34976      NM:i:6
K00234:70:HNNW3BBXX:8:2201:11972:21430  16      mtref   1       250     74S7M   *       823     0       GACATCTGGCTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACA       JJJJJJJJJJJJ
JJJJJJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:7  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1217:10236:11530  16      mtref   1       250     71S2M   *       823     0       ATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACAATGAA       JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G1 AS:i:655        Z0:i:17814      Z1:i:17814      NM:i:3
K00234:70:HNNW3BBXX:8:1224:25337:7978   16      mtref   1       250     68S13M  *       823     0       TGGTTCCTATTTTAGGGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:13 AS:i:710        Z0:i:23589      Z1:i:23589      NM:i:4
K00234:70:HNNW3BBXX:8:1203:3772:45097   16      mtref   1       250     67S11M  *       823     0       GGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTC  JFJFJJFJFAJJFFA<FJJA
JJJFJJJJJF<JJJJJJJJJAJFJJJJJJJJJJJJFJJJJJJJJJFFJJJFJJF<FA-      MD:Z:11 AS:i:755        Z0:i:6477       Z1:i:6477       NM:i:1
K00234:70:HNNW3BBXX:8:1205:5335:30538   16      mtref   1       250     62S8M   *       823     0       CTACTTCAGGGCCATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAG  FJJJJFA<JJJFJJJJJJJJJJFJJJJ
JFJJJJJJJJJJJJJJFFA-J7JJJFJJFFJJJFJJFFFFAA      MD:Z:8  AS:i:650        Z0:i:12090      Z1:i:12090      NM:i:2
K00234:70:HNNW3BBXX:8:2222:10815:11864  16      mtref   1       250     61S9M   *       823     0       TACTTTAGAGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGA  JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA      MD:Z:8G AS:i:575        Z0:i:29172      Z1:i:29172      NM:i:5
K00234:70:HNNW3BBXX:8:1219:12083:44517  16      mtref   1       250     56S25M  *       823     0       CAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATGGATCACAGATCTATCACCCTATTAA       JJJJJFJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:8G16       AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1225:16102:17122  16      mtref   1       250     55S17M  *       823     0       AGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCAC        JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJFAJJJJJJJJJJJJFJJJJJJJJJJJJJJJJFJJJJFJFFFAA    MD:Z:17 AS:i:695        Z0:i:6417       Z1:i:6417       NM:i:1
K00234:70:HNNW3BBXX:8:2127:23977:26723  16      mtref   1       250     53S28M  *       823     0       GGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGAATCACAGGTCTATCACCCTATTAACCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G27        AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1220:7415:14555   16      mtref   1       250     52S19M  *       823     0       GCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCC JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJFFFAA     MD:Z:7G11       AS:i:635        Z0:i:17794      Z1:i:17794      NM:i:3
K00234:70:HNNW3BBXX:8:2222:29203:16893  16      mtref   1       250     52S29M  *       823     0       ACCATAAAGCCTAAATAACCCACACATTCCCCTTAAATAAAACATCACAATAGATCACAGGTCTATCACCCTATTAACCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:29 AS:i:635        Z0:i:40670      Z1:i:40670      NM:i:7
K00234:70:HNNW3BBXX:8:2226:16691:22819  16      mtref   1       250     51S30M  *       823     0       CCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFAAA   MD:Z:30 AS:i:785        Z0:i:6507       Z1:i:6507       NM:i:1
K00234:70:HNNW3BBXX:8:2224:13453:21149  16      mtref   1       250     50S31M  *       823     0       CATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATTACAGGTCTATCACCCTATTAACCACTC       FJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3C27       AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2222:13088:14783  16      mtref   1       250     49S32M  *       823     0       ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:32 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
K00234:70:HNNW3BBXX:8:2125:1438:7433    16      mtref   1       250     41S21M  *       823     0       TAAATAGTCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCCTA  JJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJF
JJJJFJJJJJJJJFFJJJJJJFFAAA      MD:Z:7G13       AS:i:570        Z0:i:12010      Z1:i:12010      NM:i:2
K00234:70:HNNW3BBXX:8:1206:26575:44166  16      mtref   1       250     36S45M  *       823     0       AGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACAAGAGCTCTCCA       JJJJFJJJJJJJ
JJJJJJJJJJJJJJJJJJFJJJJJJJJJJJFJJJJJJJJJJJJJJJJFJJJFJJJFJJJJJJJJFFFAA   MD:Z:33GG10     AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1114:2412:16260   16      mtref   1       250     35S46M  *       823     0       GCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCAT       -JFJFA-AA-FJ
JFA-<JF-FJFFAJJFF-AJJJ<JJJJJJJJFJJFFAJJJJF<FJFJFJJFFFJJJFAF-JFAF-F<<-   MD:Z:46 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$ _
```

- Almost every tool for NGS analysis are command line only
- Generally more efficient/flexible, you can play around with the tools/data:
  - ex: put all text files with a specific string in a zipped archive

    a complete pain in a point-and-click windows environment, a breeze for the terminal

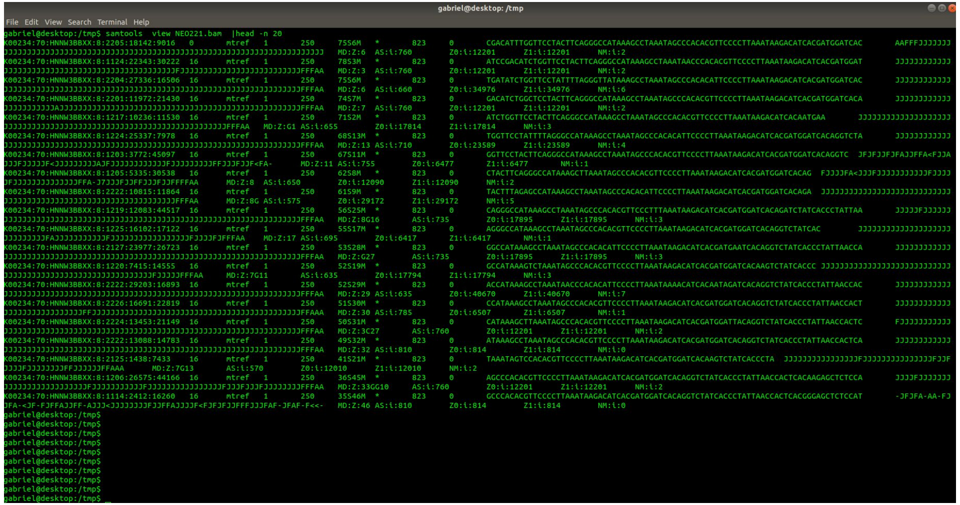
Why the shell terminal



```
gabriel@desktop:/tmp$ samtools  view NEO221.bam  |head -n 20
K00234:70:HNNW3BBXX:8:2205:18142:9016   0       mtref   1       250     75S6M   *       823     0       CGACATTTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCAC       AAFFFJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ   MD:Z:6  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1124:22343:30222  16      mtref   1       250     78S3M   *       823     0       ATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAACCCACACGTTCCCCTTAAATAAGACATCACGATGGAT       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2204:27336:16506  16      mtref   1       250     75S6M   *       823     0       TGATATCTGGTTCCTATTTTAGGGTTATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCAC       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:6  AS:i:660        Z0:i:34976      Z1:i:34976      NM:i:6
K00234:70:HNNW3BBXX:8:2201:11972:21430  16      mtref   1       250     74S7M   *       823     0       GACATCTGGCTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACA       JJJJJJJJJJJ
JJJJJJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:7  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1217:10236:11530  16      mtref   1       250     71S2M   *       823     0       ATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACAATGAA       JJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G1 AS:i:655        Z0:i:17814      Z1:i:17814      NM:i:3
K00234:70:HNNW3BBXX:8:1224:25337:7978   16      mtref   1       250     68S13M  *       823     0       TGGTTCCTATTTTAGGGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:13 AS:i:710        Z0:i:23589      Z1:i:23589      NM:i:4
K00234:70:HNNW3BBXX:8:1203:3772:45097   16      mtref   1       250     67S11M  *       823     0       GGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTC  JFJFJJFJFAJJFFA<FJJA
JJJFJJJJJF<JJJJJJJJJAJFJJJJJJJJJJJJJFJJJJJJJJJFFJJJFJJF<FA-     MD:Z:11 AS:i:755        Z0:i:6477       Z1:i:6477       NM:i:1
K00234:70:HNNW3BBXX:8:1205:5335:30538   16      mtref   1       250     62S8M   *       823     0       CTACTTCAGGGCCATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAG  FJJJJFA<JJJFJJJJJJJJJJJFJJJJ
JFJJJJJJJJJJJJJJJJFFA-J7JJJFJJFFJJJFJJFFFFAA      MD:Z:8  AS:i:650        Z0:i:12090      Z1:i:12090      NM:i:2
K00234:70:HNNW3BBXX:8:2222:10815:11864  16      mtref   1       250     61S9M   *       823     0       TACTTTAGAGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGA  JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA      MD:Z:8G AS:i:575        Z0:i:29172      Z1:i:29172      NM:i:5
K00234:70:HNNW3BBXX:8:1219:12083:44517  16      mtref   1       250     56S25M  *       823     0       CAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATGGATCACAGATCTATCACCCTATTAA       JJJJJFJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:8G16       AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1225:16102:17122  16      mtref   1       250     55S17M  *       823     0       AGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCAC        JJJJJJJJJJJJJJJJJJJ
JJJJJJJJJFAJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJFJJJJFJFFFAA    MD:Z:17 AS:i:695        Z0:i:6417       Z1:i:6417       NM:i:1
K00234:70:HNNW3BBXX:8:2127:23977:26723  16      mtref   1       250     53S28M  *       823     0       GGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGAATCACAGGTCTATCACCCTATTAACCA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G27        AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1220:7415:14555   16      mtref   1       250     52S19M  *       823     0       GCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCC JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJFFFAA      MD:Z:7G11       AS:i:635        Z0:i:17794      Z1:i:17794      NM:i:3
K00234:70:HNNW3BBXX:8:2222:29203:16893  16      mtref   1       250     52S29M  *       823     0       ACCATAAAGCCTAAATAACCCACACATTCCCCTTAAATAAAACATCACAATAGATCACAGGTCTATCACCCTATTAACCAC       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:29 AS:i:635        Z0:i:40670      Z1:i:40670      NM:i:7
K00234:70:HNNW3BBXX:8:2226:16691:22819  16      mtref   1       250     51S30M  *       823     0       CCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACT       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFAAA   MD:Z:30 AS:i:785        Z0:i:6507       Z1:i:6507       NM:i:1
K00234:70:HNNW3BBXX:8:2224:13453:21149  16      mtref   1       250     50S31M  *       823     0       CATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATTACAGGTCTATCACCCTATTAACCACTC       FJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3C27       AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2222:13088:14783  16      mtref   1       250     49S32M  *       823     0       ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCA       JJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:32 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
K00234:70:HNNW3BBXX:8:2125:1438:7433    16      mtref   1       250     41S21M  *       823     0       TAAATAGTCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCCTA  JJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJFJJF
JJJJFJJJJJJJJFFJJJJJJFFAAA      MD:Z:7G13       AS:i:570        Z0:i:12010      Z1:i:12010      NM:i:2
K00234:70:HNNW3BBXX:8:1206:26575:44166  16      mtref   1       250     36S45M  *       823     0       AGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACAAGAGCTCTCCA       JJJJFJJJJJJ
JJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJFJJJFJJJFJJJJJJJJFFFAA   MD:Z:33GG10     AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1114:2412:16260   16      mtref   1       250     35S46M  *       823     0       GCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCAT       -JFJFA-AA-FJ
JFA-<JF-FJFFAJJFF-AJJJ<JJJJJJJJFJJFFAJJJJF<FJFJFJJFFFJJJFAF-JFAF-F<<-   MD:Z:46 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
```

- They can be pipelined, i.e. analyzing 100 files in windowed mode is a pain ...
- Alternative approaches: Galaxy, CLC-workbench, Geneious

Why learn to use UNIX/Linux? (in general)

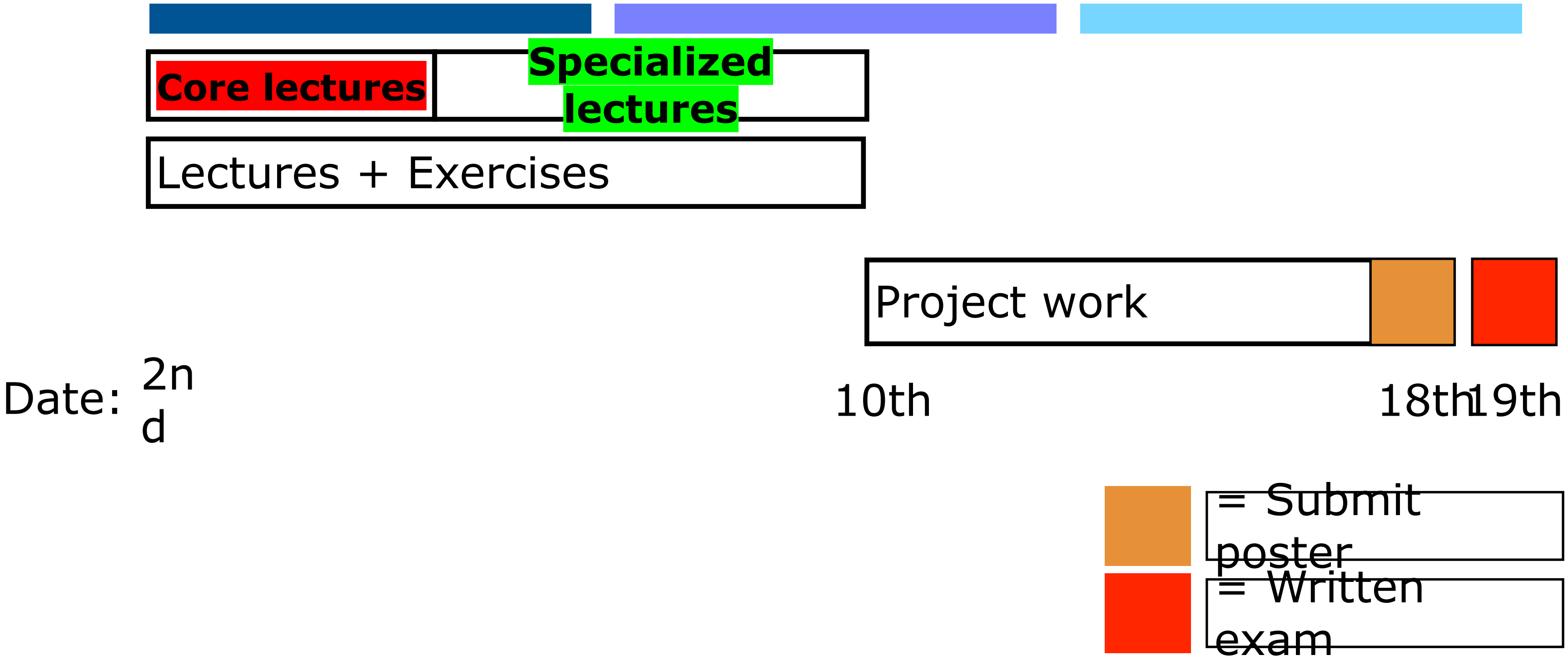


```
gabriel@desktop:/tmp$ samtools  view NEO221.bam  |head -n 20
K00234:70:HNNW3BBXX:8:2205:18142:9016   0       mtref   1       250     75S6M   *       823     0       CGACATTTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCAC       AAFFFJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ   MD:Z:6  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1124:22343:30222  16      mtref   1       250     78S3M   *       823     0       ATCCGACATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAACCCACACGTTCCCCTTAAATAAGACATCACGATGGAT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2204:27336:16506  16      mtref   1       250     75S6M   *       823     0       TGATATCTGGTTCCTATTTTAGGGTTATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:6  AS:i:660        Z0:i:34976      Z1:i:34976      NM:i:6
K00234:70:HNNW3BBXX:8:2201:11972:21430  16      mtref   1       250     74S7M   *       823     0       GACATCTGGCTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACA       JJJJJJJJJJJJ
JJJJJJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:7  AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1217:10236:11530  16      mtref   1       250     71S2M   *       823     0       ATCTGGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACAATGAA       JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G1 AS:i:655        Z0:i:17814      Z1:i:17814      NM:i:3
K00234:70:HNNW3BBXX:8:1224:25337:7978   16      mtref   1       250     68S13M  *       823     0       TGGTTCCTATTTTAGGGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:13 AS:i:710        Z0:i:23589      Z1:i:23589      NM:i:4
K00234:70:HNNW3BBXX:8:1203:3772:45097   16      mtref   1       250     67S11M  *       823     0       GGTTCCTACTTCAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTC  JFJFJJFJFAJJFFA<FJJA
JJJFJJJJJF<JJJJJJJJJAJFJJJJJJJJJJJJJFJJJJJJJJJFFJJJFJJF<FA-      MD:Z:11 AS:i:755        Z0:i:6477       Z1:i:6477       NM:i:1
K00234:70:HNNW3BBXX:8:1205:5335:30538   16      mtref   1       250     62S8M   *       823     0       CTACTTCAGGGCCATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAG  FJJJJFA<JJJFJJJJJJJJJJJFJJJJ
JFJJJJJJJJJJJJJJFFA-J7JJJFJJFFJJJFJJFFFFAA      MD:Z:8  AS:i:650        Z0:i:12090      Z1:i:12090      NM:i:2
K00234:70:HNNW3BBXX:8:2222:10815:11864  16      mtref   1       250     61S9M   *       823     0       TACTTTAGAGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGGATCACAGA  JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA      MD:Z:8G AS:i:575        Z0:i:29172      Z1:i:29172      NM:i:5
K00234:70:HNNW3BBXX:8:1219:12083:44517  16      mtref   1       250     56S25M  *       823     0       CAGGGCCATAAAGCCTAAATAGCCCACACGTTCCCTTTAAATAAGACATCACGATGGATCACAGATCTATCACCCTATTAA       JJJJJFJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:8G16       AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1225:16102:17122  16      mtref   1       250     55S17M  *       823     0       AGGGCCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCAC       JJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJFAJJJJJJJJJJJJFJJJJJJJJJJJJJJJJJFJJJJFJFFFAA    MD:Z:17 AS:i:695        Z0:i:6417       Z1:i:6417       NM:i:1
K00234:70:HNNW3BBXX:8:2127:23977:26723  16      mtref   1       250     53S28M  *       823     0       GGCCATAAAGCCTAAATAGCCCACACATTCCCCTTAAATAAGACATCACGATGAATCACAGGTCTATCACCCTATTAACCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:G27        AS:i:735        Z0:i:17895      Z1:i:17895      NM:i:3
K00234:70:HNNW3BBXX:8:1220:7415:14555   16      mtref   1       250     52S19M  *       823     0       GCCATAAAGTCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCC JJJJJJJJJJJJJJJJJJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFJJJJJFFFAA      MD:Z:7G11       AS:i:635        Z0:i:17794      Z1:i:17794      NM:i:3
K00234:70:HNNW3BBXX:8:2222:29203:16893  16      mtref   1       250     52S29M  *       823     0       ACCATAAAGCCTAAATAACCCACACATTCCCCTTAAATAAAACATCACAATAGATCACAGGTCTATCACCCTATTAACCAC       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:29 AS:i:635        Z0:i:40670      Z1:i:40670      NM:i:7
K00234:70:HNNW3BBXX:8:2226:16691:22819  16      mtref   1       250     51S30M  *       823     0       CCATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACT       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJFFJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFAAA   MD:Z:30 AS:i:785        Z0:i:6507       Z1:i:6507       NM:i:1
K00234:70:HNNW3BBXX:8:2224:13453:21149  16      mtref   1       250     50S31M  *       823     0       CATAAAGCTTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATTACAGGTCTATCACCCTATTAACCACTC       FJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:3C27       AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:2222:13088:14783  16      mtref   1       250     49S32M  *       823     0       ATAAAGCCTAAATAGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCA       JJJJJJJJJJJJ
JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJFFFAA   MD:Z:32 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
K00234:70:HNNW3BBXX:8:2125:1438:7433    16      mtref   1       250     41S21M  *       823     0       TAAATAGTCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAAGTCTATCACCCTA  JJJJJJJJJJJJJJJJFJJJJJJJJJJJJJJJFJJF
JJJJFJJJJJJJJJFFJJJJJJFFAAA      MD:Z:7G13       AS:i:570        Z0:i:12010      Z1:i:12010      NM:i:2
K00234:70:HNNW3BBXX:8:1206:26575:44166  16      mtref   1       250     36S45M  *       823     0       AGCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACAAGAGCTCTCCA       JJJJFJJJJJJJ
JJJJJJJJJJJJJJJJJJFJJJJJJJJJJJFJJJJJJJJJJJJJJJJFJJJFJJJFJJJJJJJJFFFAA   MD:Z:33GG10     AS:i:760        Z0:i:12201      Z1:i:12201      NM:i:2
K00234:70:HNNW3BBXX:8:1114:2412:16260   16      mtref   1       250     35S46M  *       823     0       GCCCACACGTTCCCCTTAAATAAGACATCACGATGGATCACAGGTCTATCACCCTATTAACCACTCACGGGAGCTCTCCAT       -JFJFA-AA-FJ
JFA-<JF-FJFFAJJFF-AJJJ<JJJJJJJJFJJFFAJJJJF<FJFJFJJFFFJJJFAF-JFAF-F<<-   MD:Z:46 AS:i:810        Z0:i:814        Z1:i:814        NM:i:0
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
gabriel@desktop:/tmp$
```

- Contains several little programs (sed, cut, grep, paste) that can be combined to make really powerful queries
- File descriptors and pipe can be used to spare you a lot of time/disk space
- Make/Snakemake/Nextflow can automate workflows
- Open source tools
- You can basically finish a PhD in computational bio. without knowing how to code

Course structure

- 3 weeks, 2 tracks



Core lectures

Specialized lectures

Lectures + Exercises

Project work

Date: 2nd 10th 18th 19th

🟧 = Submit poster

🟥 = Written exam

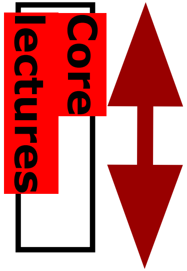
Course breakdown I

Core lectures

- Tuesday 2nd January
  - Introduction NGS technology
  - Unix and first look at data

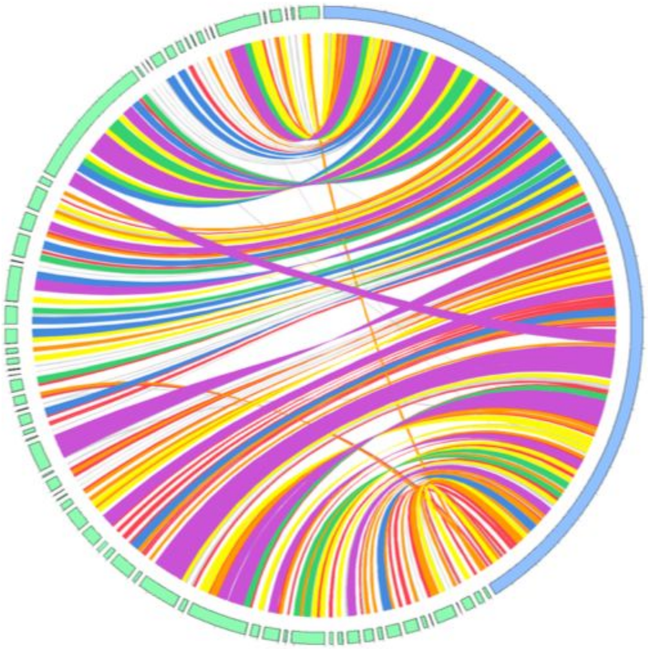
- Wednesday 3rd January
  - Data basics & preprocessing
  - Alignment
  - Reminder about Bayesian

Course breakdown II



- Thursday 4th January
  - Alignment processing
  - Genotyping
  - *de novo* assembly

- Friday 5th January
  - Ancient DNA
  - RNASeq



Course breakdown III

- Monday 8th January
  - Long read technologies
  - Recap test (after lunch)
  - Finish exercises

- Tuesday 9th January
  - Metagenomics

- Wednesday 10th January
  - Genomic Epidemiology
  - Group formation project

Course breakdown IV

- Thursday 11th January
  - Cancer-seq
  - Project work
  - Individual group talks
- Friday, 12th January
  - Project work
- Monday 13th - Thursday 17th
  - Project work
- Thursday 18th
  - Submit 1 page poster
- Friday 19th
  - Written Exam

Projects

- Try to analyze an empirical dataset and present results on poster
- 4-6 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
- Do **not** analyze very large datasets (time, resources)
- 2024: You have 2 days, a weekend and a week to finish

download+align data here!!

Points to remember

- **Understand** principles of the analysis
- The exercises will be useful for your projects and hopefully also later
- You don’t need to do all the exercises but the ones from the core lectures are important
- Have an exercise buddy and do them as a team, preferably on each individuals laptop so everyone gets to learn the command-line
- Please **just ask** questions at any time !

Points to remember

- You get the solutions for the exercises but **do not copy-paste!!**

- You will not get to copy-paste for the project

Cloud computing

- Pupil cluster
- We have 3 nodes

– pupil1	40 cores	252G RAM
– pupil2	24 cores	110G RAM
– pupil3	24 cores	94G RAM

- Be careful with disk space
- Limited computational power
- If you want software installed, ask me!

Poster

- Each group will create a poster
- The goal of the project is:
  - Do not memorize, **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 18th

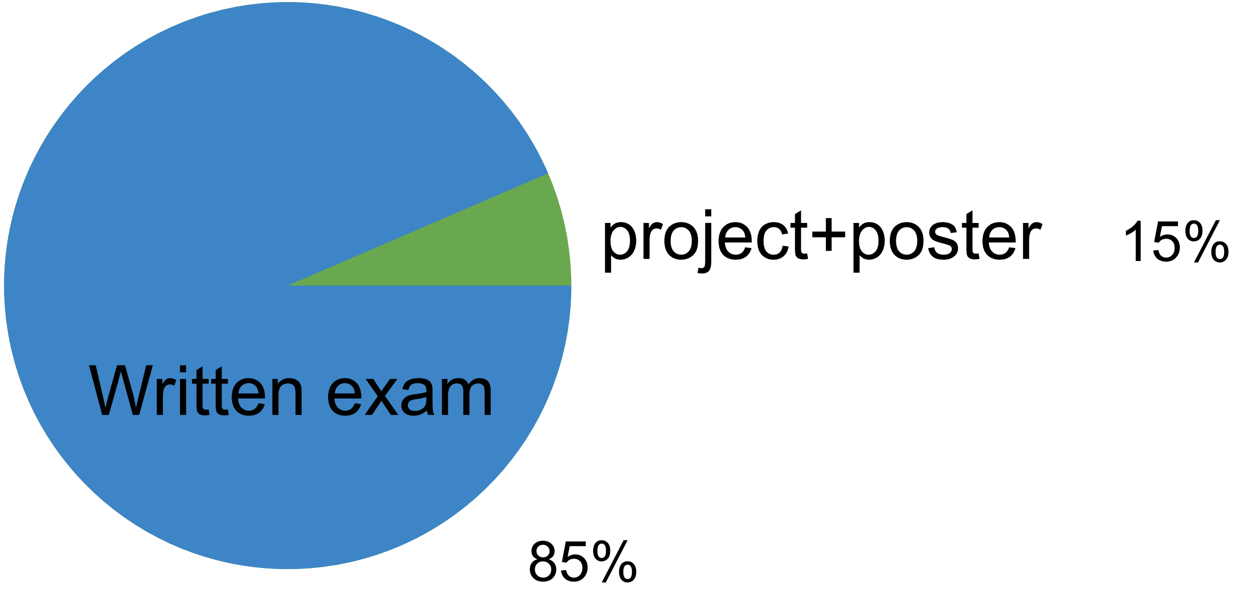
Written Exam

- **You cannot write the exam if you have not submitted the poster**
- Multiple choice exam
- Focuses on the core lectures
- Will have 1 very basic question per specialized lecture

Tips for this class

- Do not memorize definitions, **understand** concepts
- The core lectures are especially crucial
- The final exam is an oral one which will evaluate your understanding, not whether you can parroting definitions
- Do the exercises (esp. the first 3 days).
- Understand what you are doing:
  - inspect the input
  - inspect the output
  - play with parameters

Marking scheme



Disclaimer

- Sequencing technologies change very rapidly!
- We will dive into many areas and you will not learn to master everything
- However, we hope that the building blocks we provide will allow you to see new opportunities

Disclaimer

- We will talk about old techs, working with NGS means working with older datasets from previous studies

source: Dall-E

Be adventurous!

You do not have the ability to do anything destructive

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

Course webpage

- Course program, slides, handouts, exercises etc.

- http://teaching.healthtech.dtu.dk/22126

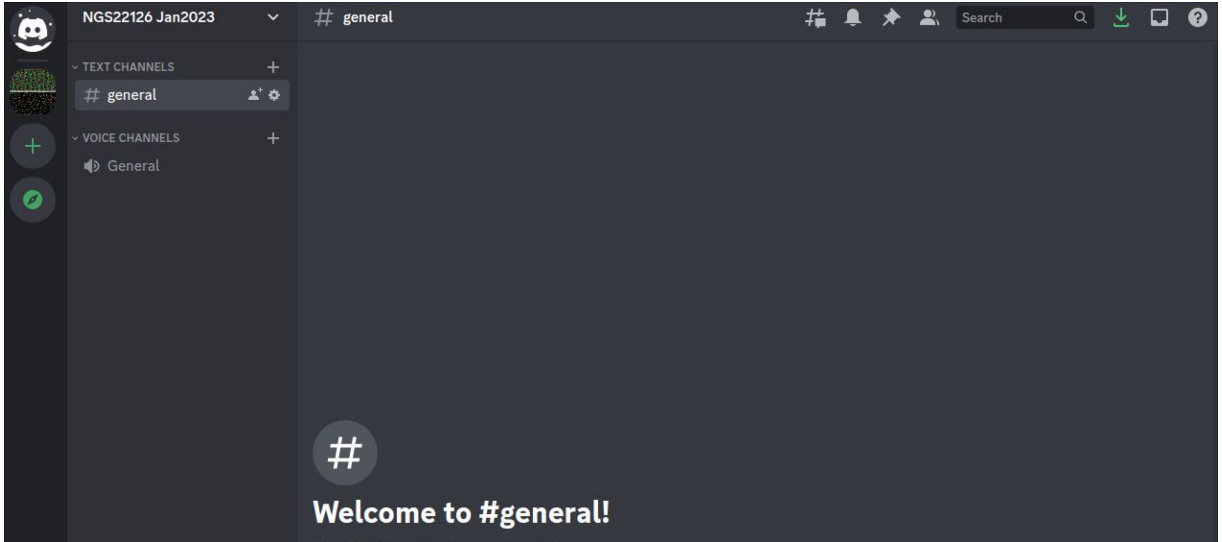
- We want the course page to be a repository for you!

Discord

- Chat with others during off-hours. Create channels!
- Please use your real name:

Jan Jansen ✓

n00b_owner_18 ✗



Welcome to #general!

Reading + wifi

- There are no textbooks for the course, it changes too rapidly
- Wireless networks
  - Use “dtu” and your dtu/campusnet login to get access to wireless
  - Eduroam

Pre-test

- Test your knowledge before we start

- Not used for grading or exam

- Used to understand where you are and what you need