# **Bioinformatics and Bioinformaticians**

#### What is it? Who are they? What do they do?

Johan Nylander

NBIS.se / NRM.se / SciLifeLab.se







# Introduction

In a time where we see transitions towards biology as a data science, and towards a 'life cycle' view of research data



Introduction



EDUCATION

# Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators

Lindsay Barone\*°, Jason Williams°, David Micklos°

Barone et al. 2018, PLoS Comput Biol

# Introduction



Fig 3. Unmet data analysis needs of National Science Foundation (NSF) Biological Sciences Directorate (BIO) principal investigators (PIs) (neuropet responding population) (218 < n < 510)

# Bioinformatics - What's in a Word?



"Scientist studying DNA sequences"

 "The application of information technology to the field of molecular biology" (wikipedia.org)

- "The application of information technology to the field of molecular biology" (wikipedia.org)
- Collection, archiving, organization, and interpretation of biological data

- "The application of information technology to the field of molecular biology" (wikipedia.org)
- Collection, archiving, organization, and interpretation of biological data
- Use of computers to address problems that are specific to biologists

- "The application of information technology to the field of molecular biology" (wikipedia.org)
- Collection, archiving, organization, and interpretation of biological data
- Use of computers to address problems that are specific to biologists
- Informational technology + computer science + mathematics and statistics

# A Bioinformatician is, then, Someone Who...

- > Applies information technology to the field of molecular biology
- Collects, archives, organizes, and interprets biological data
- Uses computers to address problems that are specific to biologists
- Uses informational technology + computer science + mathematics and statistics

Hi, my name is Johan, and I'm a Bioinformatician

Anyone else who wants to share?

(The following are quotes from job ads)

(The following are quotes from job ads)

Individual with a superior understanding of biology, computer science, and the latest trends in analyzing big data

(The following are quotes from job ads)

- Individual with a superior understanding of biology, computer science, and the latest trends in analyzing big data
- Have strong communication skills and ability to tell a story using numbers and data

(The following are quotes from job ads)

- Individual with a superior understanding of biology, computer science, and the latest trends in analyzing big data
- Have strong communication skills and ability to tell a story using numbers and data
- Ability to apply advanced statistical methods, machine learning, data science to molecular biology

- Strong background in programming and algorithms
- ▶ Fluent in one programming language (Python, C, C++ or Java) and familiarity with scripting languages (bash, Perl, R)

- Strong background in programming and algorithms
- Fluent in one programming language (Python, C, C++ or Java) and familiarity with scripting languages (bash, Perl, R)
- Expert knowledge of Unix operating system, and High-performance computing (HPC) systems

- Have hands-on experience with bioinformatics methods appropriate for NGS applications (targeted sequencing, RNASeq, CHiPSeq, and de novo genome assembly)
- Have hands-on experience with genome alignment, mapping, variant calling and annotation (e.g. BWA, Bowtie, STAR, GATK, samtools, bcftools)



Provide advanced bioinformatics analyses within research projects

> Provide advanced bioinformatics analyses within research projects

Development of tools and workflows for such analyses

Develop custom databases and web portals for managing raw and processed experimental data

- Develop custom databases and web portals for managing raw and processed experimental data
- Design and implement reusable bioinformatics analysis pipelines for processing next-generation sequencing, microarray, genomics, proteomics and chemogenomics data

- Develop custom databases and web portals for managing raw and processed experimental data
- Design and implement reusable bioinformatics analysis pipelines for processing next-generation sequencing, microarray, genomics, proteomics and chemogenomics data
- Develop novel algorithms and integrated data visualization applications when existing software packages are not available or are not adequate

Educating other scientists in bioinformatics through collaboration within supported projects, teaching at national courses, and through participating in various networks

- Educating other scientists in bioinformatics through collaboration within supported projects, teaching at national courses, and through participating in various networks
- Participating in the writing of scientific articles



Need to successfully deliver key insights to achieve greatness across our products

Need to successfully deliver key insights to achieve greatness across our products

Perform other duties as required

# One Real Example of Skills Requirements

The following profiles are expected:

- 1. Analysis of standard HTS data (variant calling, genome assembly, differential expression, ...),
- 2. Metagenomics
- 3. Metatranscriptomics
- 4. Single-cell analysis
- 5. Analysis of metabolomic data
- 6. Epigenomics and epitranscriptomics
- 7. Flux cytometry data analysis
- 8. Biological databases
- 9. Proteomics (Mass Spectrometry for Biology)
- 10. Omics data analysis and integration

- A Bachelor's degree in life sciences or computational sciences is required, although a MSc or PhD is preferred.
- Master's degree in Computer Science or Bioinformatics or a related computational discipline, or equivalent experience
- PhD or MSc in Bioinformatics, Computer Science, Computational Biology, Electrical Engineering/Signal processing or a related field with > 1 years' experience

A Bioinformatician is Not a Unicorn

# A biologist may have broad knowledge of biology, but will have expertise in just part

A Bioinformatician is Not a Unicorn

- A biologist may have broad knowledge of biology, but will have expertise in just part
- Similarly, a bioinformatician is not universal

- Organize other people's data
- Analyze other people's data
- Educate other people in the field of bioinformatics
- **Support** other people in analysis
- Develop tools for other people to use

Things that Bioinformaticians Do - and How Much of What?

Organize ~ 20%
 Analyze ~ 20%
 Educate ~ 20%
 Support ~ 30%
 Develop ~ 10%

# Things that Bioinformaticians Do - and How Much of What?

# 85% of the time a bioinformatician does "research"!



Chang 2015 Nature Communications

Career Opportunities and Salaries for Bioinformaticians?

- Need to offer career opportunities for bioinformatitians Can you?
- Academic suicide if not first or last author in a world based on counting publications
- Competetive salaries in academia?

Getting Bioinformatics Done - By Whom?

You?
Your students?
Your postdocs?
Your collaborators?
...
A company?

Getting Bioinformatics Done - By Whom?



# What Should I Learn?



# What Should I Learn?



# Be able to look at your data?

# Be able to look at your data?

- Locate your data
- Learn how to identify the "non-reproducible" parts of the data
- Transfer data
  - ...

\$ sshfs user-delivery01234@grus.uppmax.uu.se: ~/grus
\$ firefox ~/grus/P12345/00-Reports/\*multiqc\_report.html
\$ backup2nrm ~/grus/P12345



Other suggestions?

Do I Need to Learn Programming?

Do I Need to Learn Programming?

No, but Why not?

Some Words of Comfort: Can be Easy!

# Actually programming can be easy Doesn't require knowledge of computer science

# Some Words of Comfort: Just like a Protocol

```
# For every infile that ends in .fas:
for infile in *.fas
do
    alifile="${infile%.fas}.ali"
```

```
# 1. Do multiple sequence alignment
mafft --auto "$infile" > "$alifile"
```

```
# 2. Infer a phylogenetic tree
iqtree -s "$alifile" -nt AUTO -m TEST
```

#### done

# Some Words of Comfort: Just like a Protocol

You start with an existing one and tweak it



Some Words of Comfort: You know the Data!

You have a big advantage over the bioinformaticians:

- You generated the data
- You understand the biology
- You understand how things go wrong

# Learn How to Visualize Your Data

Great for learning about the data
Easy to pick out outliers and trends



# https://www.r-graph-gallery.com/







```
library(ggridges)
library(ggplot2)
ggplot(diamonds, aes(x = price, y = cut, fill = cut)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none")
```

Do I Need to Learn Programming?



Do I Need to Learn Programming?

# Well, up to you (your loss...)

# Have your student(s) learn programming!

- Include a programming course as part of their PhD
- Have your student(s) gather other students in "study circles"

Courses and Teach Yourself

# https://tess.elixir-europe.org/



