

---

# SciLifeLab Bioinformatics Platform

National Bioinformatics Infrastructure Sweden (NBIS)

***Björn Nystedt***

*Uppsala 01.12.2017*

SciLifeLab

## National service

The Swiss army knife for Swedish  
Life Science researchers

Local scientific  
center



Director: Olli Kallioniemi  
Co-director: Siv Andersson

### *Vision:*

*To be an internationally leading center that develops, uses and provides access to advanced technologies for molecular biosciences with focus on health and environment.*

2010: Strategic research initiative  
2013: National resource  
2015: New management/chairman

# SciLifeLab platforms

SciLifeLab national service

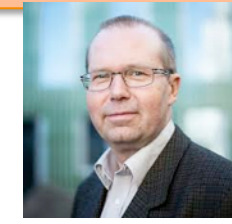
National Genomics Infrastructure

Diagnostics Development

Single-cell Biology

National Bioinformatics Infrastructure Sweden

Bengt Persson



VR

SNIC



Computer resources free for Swedish researchers



SciLifeLab Bioinformatics (NBIS) and Genomics (NGI) were both top-ranked in the recent VR evaluation (September 2017).

“NBIS is probably the largest genuinely national and fully established bioinformatics infrastructure in Europe.”

“[NBIS..] is crucial to the future competitiveness of Sweden in data-driven life sciences research, and is helping to keep Sweden in the European forefront in the area.”

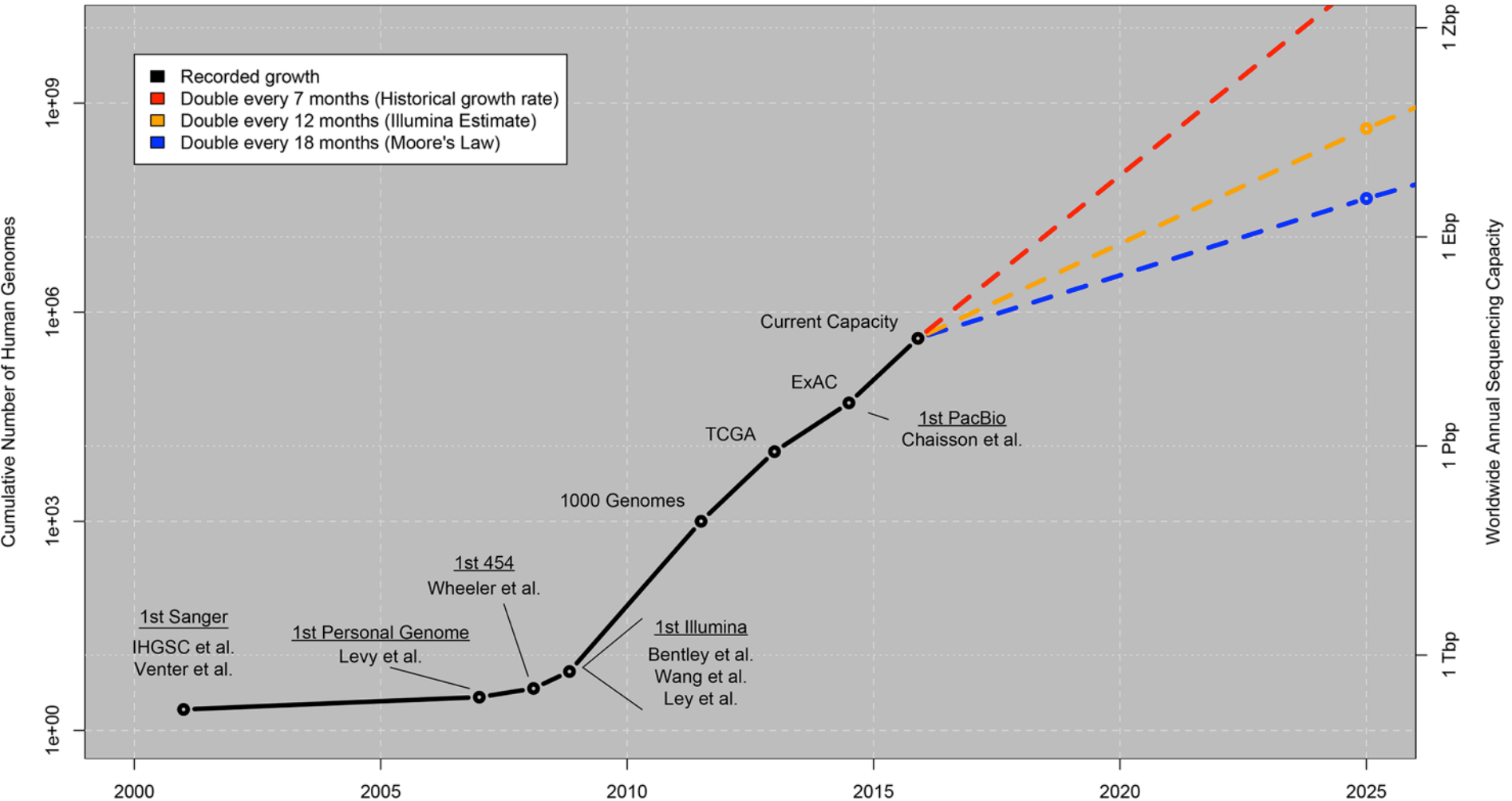


Scientific impact: 7/7 (“*Outstanding*”)

Overall score: 7/7 (“*Outstanding*”)

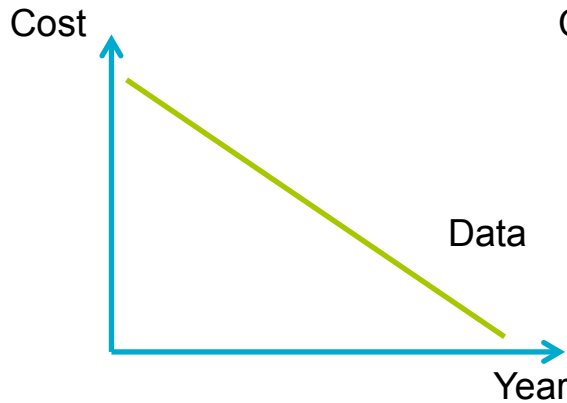
# Data growth

Growth of DNA Sequencing

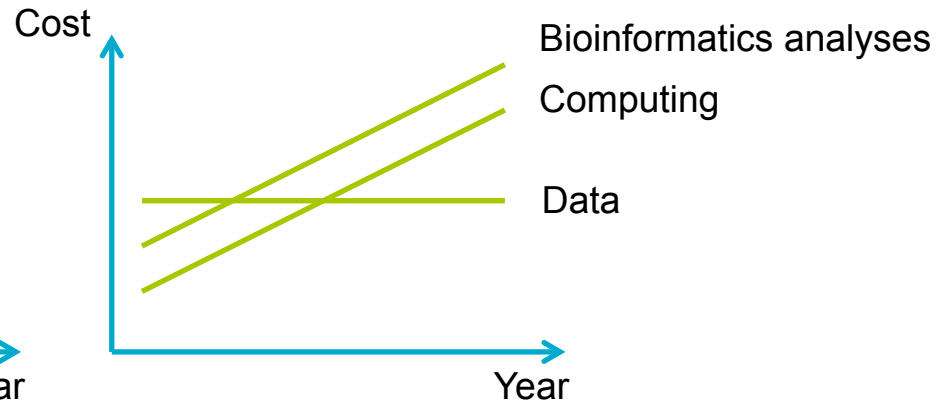


Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, et al. (2015) Big Data: Astronomical or Genomical?. PLoS Biol 13(7): e1002195. doi:10.1371/journal.pbio.1002195  
<http://127.0.0.1:8081/plosbiology/article?id=info:doi/10.1371/journal.pbio.1002195>

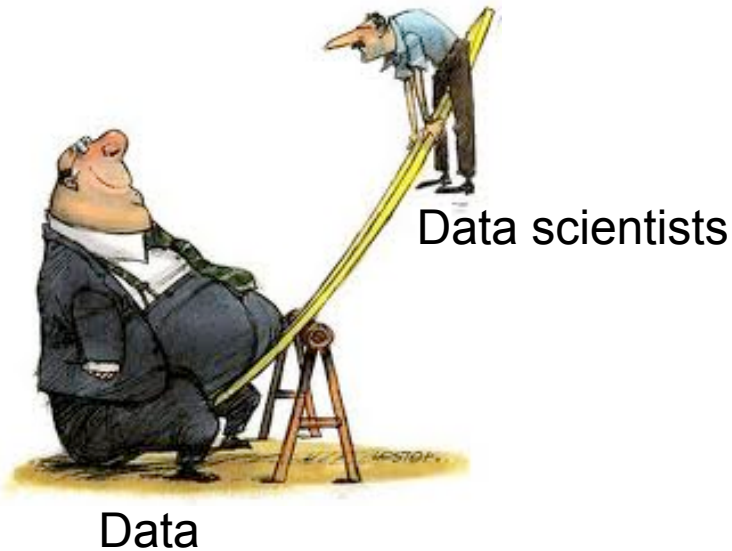




“Per base”



“Per project”



**Our role**

We want to help the Swedish Life Science community to **build knowledge** in large-scale data analysis, and to make bioinformatics **easily accessible** for all.

## NBIS activities

# Support, tools and training



Support



Tools



Training





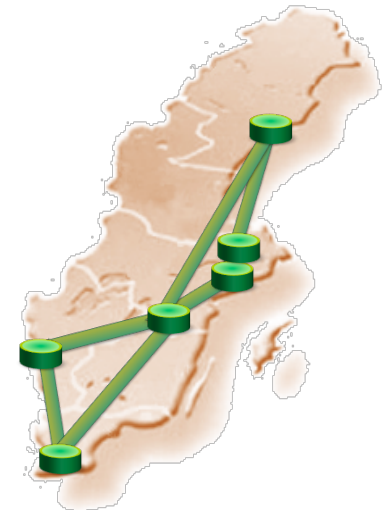


## Custom-tailored support

[www.nbis.se](http://www.nbis.se)

[www.scilifelab.se/platforms/bioinformatics/](http://www.scilifelab.se/platforms/bioinformatics/)

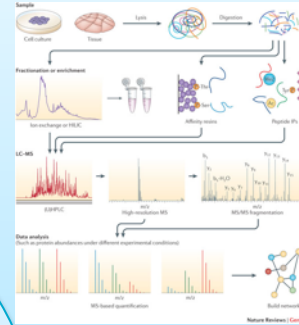
- Study design consultation (free)  
[www.nbis.se/support/supportform/index.php](http://www.nbis.se/support/supportform/index.php)  
+ drop-in sessions every week @ all 6 sites
- Support (User fee 800 kr/h)  
[www.nbis.se/support/supportform/index.php](http://www.nbis.se/support/supportform/index.php)
- Long-term support and systems biology  
(500h, free, scientific evaluation)  
[www.nbis.se/support/supportform/index.php?form=longterm](http://www.nbis.se/support/supportform/index.php?form=longterm)



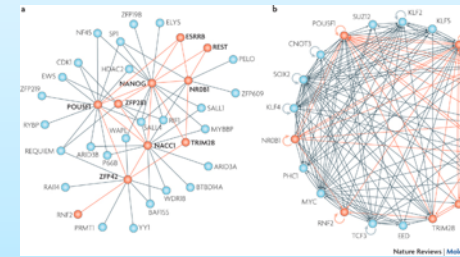
# Bioinformatics support



Genomics



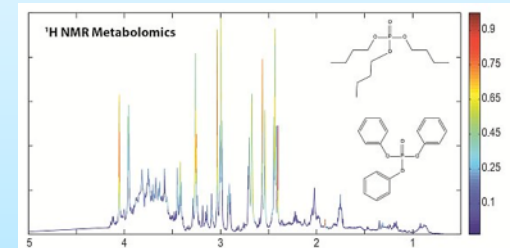
Proteomics



Systems biology



Biostatistics



Metabolomics

## 2 tracks!

- Fee-for-service (800kr/h)  
Rapid turnaround
- Scientific ranking (free)  
“Long-term Support”  
3 open calls/year

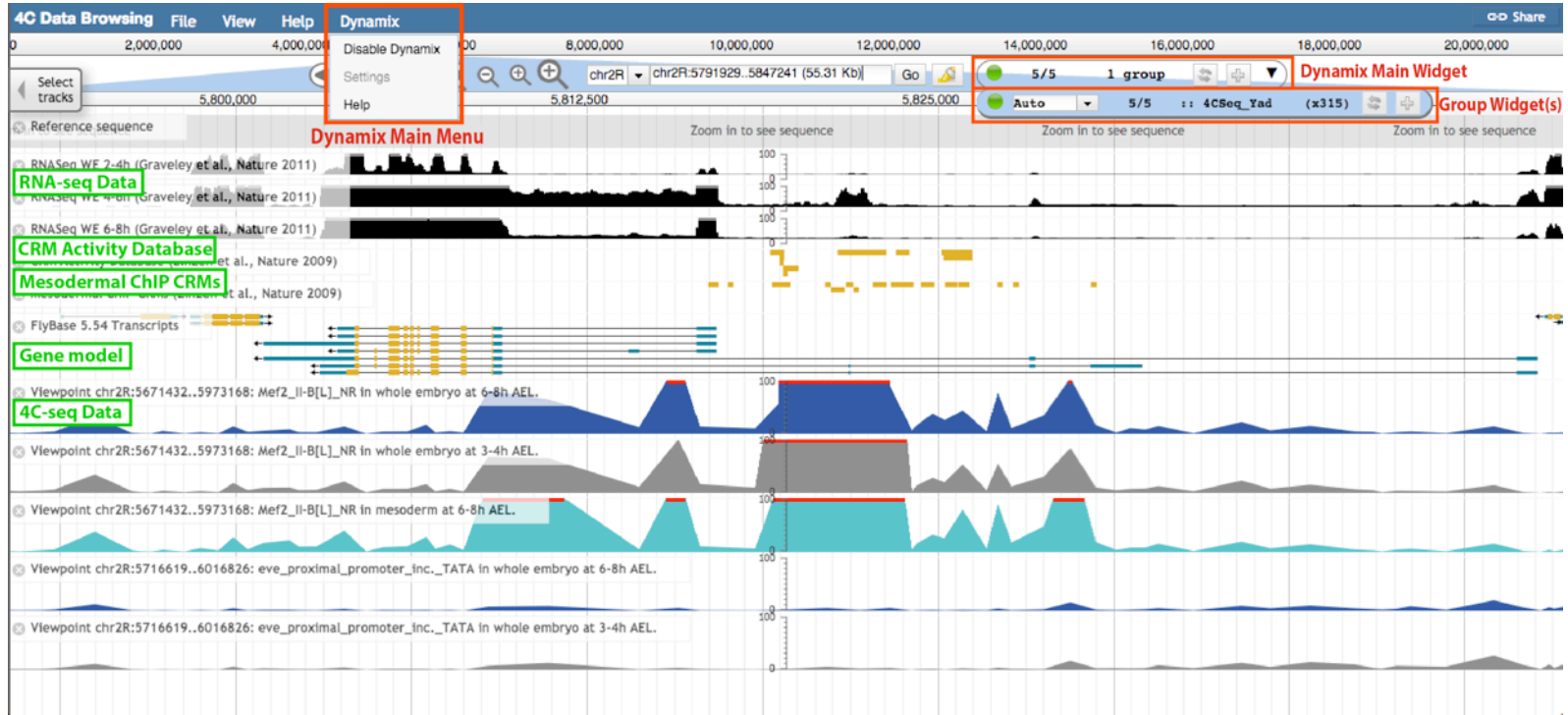
The screenshot shows a web browser window with the URL nbis.se. The page features a large header with the NBIS logo and the text "NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN". A navigation menu is visible at the top right, with the "Support" dropdown menu open. The menu items are: "General information", "Consultation", "Support", "Long-term Support", "Genome Assembly and Annotation", and "Apply here!". The "Apply here!" option is highlighted with a red circle. Below the header, there is a banner with the text "NBIS is a distributed national bioinformatics infrastructure, supporting life sciences in Sweden". At the bottom of the page, there are three green buttons with icons representing people, a gear, and a graduation cap.

Support ▾ Infrastructure ▾ Training ▾ News About ▾

- General information
- Consultation
- Support
- Long-term Support
- Genome Assembly and Annotation
- Apply here!

NBIS is a distributed national bioinformatics infrastructure, supporting life sciences in Sweden

nbis.se/support



- 10 - 20 projects per year
- Highly specialized staff and robust pipelines
- Tight user interaction
- Numerous manual and semi-manual QC steps
- Supports ENA submission
- Editable user interface

Cost effective with high quality!

# BigData/Integrative omics

## 4 FTE, joint effort by Long-term Support and Systems Biology

Projects apply in the regular Long-term Support calls

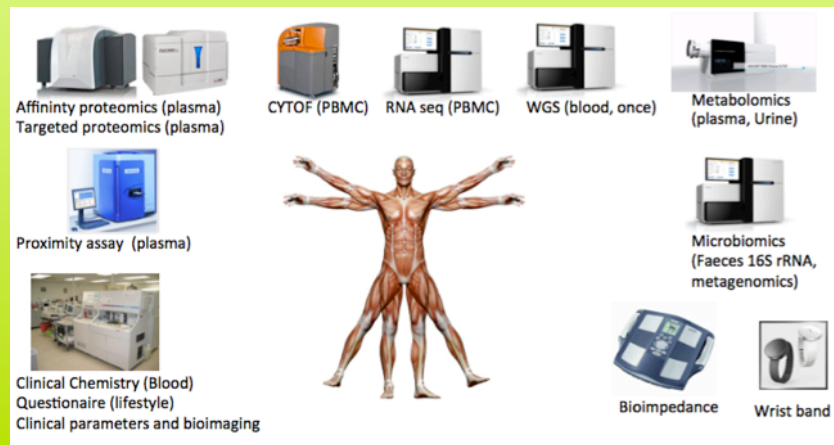
### Combine data from SciLifeLab platforms

- Building **tools** and **resources** for handling very large and/or complex biological data sets
- Typically performed in the context of longer support projects
- State-of-the-art analytical methods for integrating multi-modal biological data sets, eg
  - Machine learning/deep learning
  - Graph-based models
  - Genome-scale metabolic models

### Support track for integrative projects

First call Feb 2016; First few projects initiated

Involves extensive integration of data







## Compute and storage of sensitive data

- Local EGA
- ePouta integration pilot
- microMosler
- Pouta Blueprints
- web-servers with EGI cloud vo.NBIS.se

## WGS tools and resources

- SweGen 1000 genomes
- WGS somatic variant calling WF
- WGS structural variation WF

## Software maintenance

- MrBayes
- Structure prediction web services

## Assembly and annotation

- Falcon on Milou
- ENA submission help

## Other tools and resources

- Human Metabolic Atlas (HMA)
- Haloplex variant calling pipeline
- WhatsHap: Genomic phasing
- IgDiscover: Immunorepertoire

**NBIS** Support Infrastructure Training News Events About SciLifeLab

## Tools

The current tools under development and maintenance by NBIS staff are listed below. For more detailed information on the tools, follow the links. See also the [SciLifeLab Open Source website](#) for tools developed at other SciLifeLab platforms.

<b>CAW</b> URL: <a href="http://opensource.scilifelab.se/projects/caw/">http://opensource.scilifelab.se/projects/caw/</a>  CAW (Cancer Analysis Workflow) is a complete open source pipeline to detect somatic variants from WGS data.	Software type: Command-line tool, Workflow Platform: Linux Development status: Active Released: Yes Contact: <a href="mailto:maxime.garcia@scilifelab.se">maxime.garcia@scilifelab.se</a> Affiliation: scilifelab.se
<b>ChIP-seq QC pipeline</b> URL: <a href="https://github.com/ewels/clusterflow">https://github.com/ewels/clusterflow</a>  Pipeline for quality control of ChIP-seq data, implemented as a plug-in for the workflow manager ClusterFlow. See pipelines/chipseq_qc.config in the ClusterFlow repository on GitHub.	Software type: Workflow Platform: Linux Development status: Maintained Released: Yes Contact: <a href="mailto:jakub.weestholm@scilifelab.se">jakub.weestholm@scilifelab.se</a> Affiliation: nbis.se
<b>EMBLmyGFF3</b>	Software type: Command-line tool Platform: Linux, Mac

<http://nbis.se/infrastructure/tools/>  
<http://opensource.scilifelab.se/>

## SweGen Variant Frequency Database

- 950 twin registry + 50 Northern Sweden
- Deep coverage WGS (30X)
- ExAC browser interface
- Data Beacon
- Full SNP frequency table download



<https://swefreq.nbis.se/#/>  
1<sup>st</sup> release October 2016

Variant: 22:46615880 T / C

Filter Status: PASS  
dbSNP: rs1800234  
Allele Frequency: 0.0035  
Allele Count: 7 / 2000  
UCSC: 22-46615880-T-C  
ClinVar: [Click to search for variant in Clinvar](#)

Genotype Quality Metrics  
Site Quality Metrics

### Annotations

This variant falls on 7 transcripts in 1 genes:

missense  
• PPARA Transcripts

intron  
• PPARA - ENST00000434345

non coding transcript exon  
• PPARA - ENST00000493286

Note: This list may not include additional transcripts in the same gene that the variant does not overlap.

### Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
SweGen	7	2000	0	0.0035
Total	7	2000	0	0.0035

### Read Data

This interactive IGV.js visualization shows reads that went into calling this variant.

Note: Read data is not available for this variant.

chr22:46,615,730-46,616,030 301 bp

Funding: SciLifeLab  
Sequencing: NGI  
Variant calling: NGI  
QC: NBIS  
Data access interface: NBIS



# SciLifeLab course curriculum 2018-2020

Course	Location
Advanced Molecular Technology and Instrumentation for Proteome Analyses	Uppsala
Biophysical methods in drug discovery	Uppsala / Sthlm
Cellular profiling within the Human Protein Atlas/Spatial Proteomics	Stockholm
Chemical Proteomics	Stockholm
Cryo-EM sample preparation and data collection	Umeå
Cryo Electron Tomography and image processing	Stockholm
Intermediate level R for Bioinformatics. Summer school.	Gotland
Introduction to bioinformatics using NGS data (4 x / year)	UU, LiU, LU, UmU, GU
Opportunities for Affinity Proteomics	Stockholm
Single cell genomics - a practical and theoretical workshop	Uppsala
Single Particle Cryo-EM image processing	Stockholm

Date	Training
Application Nov/ Dec	The Swedish Bioinformatics Advisory Program (1-2 years)

## NBIS Courses

- Python Programming (2x / year)
- R programming (2x / year)
- RNA-seq (2x /year)
- Single cell RNA-seq analysis
- CHIP-seq data analysis
- Genome Annotation
- *De novo* Genome Assembly
- Metagenomics

<https://www.scilifelab.se/education/courses/>



## The Swedish Bioinformatics Advisory Program

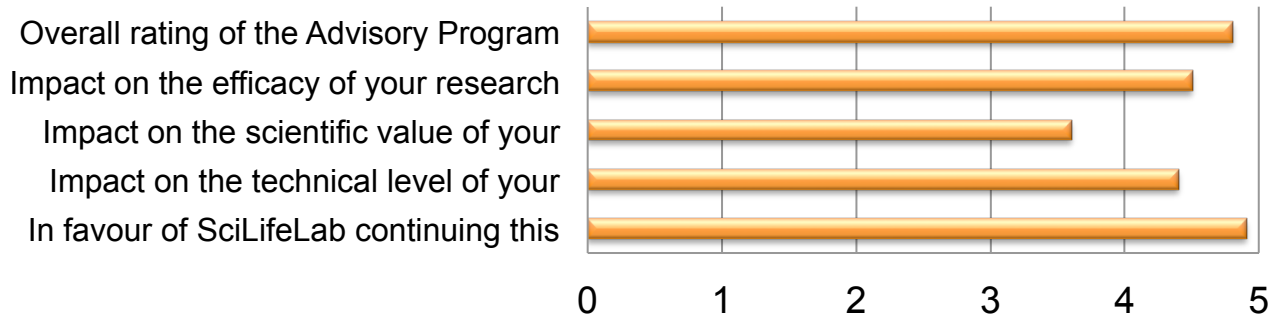
PhD students get a senior bioinformatician as a personal advisor during 2 years of their PhD. Monthly project meetings + two grand meetings per year to aid networking and knowledge transfer.

[www.scilifelab.se/education/mentorship/the-swedish-bioinformatics-advisory-program/](http://www.scilifelab.se/education/mentorship/the-swedish-bioinformatics-advisory-program/)

Last call (2017/2018): 111 applicants for 15 places  
**Deadline Dec 12, 2017!**

### The Swedish Bioinformatics Advisory Program

Student evaluation, June 2015



**We're here for you!**  
**[www.nbis.se](http://www.nbis.se)**

---

SciLifeLab



**NBIS**  
NATIONAL BIOINFORMATICS  
INFRASTRUCTURE SWEDEN