# NBIS Workshop

# **RNAseq Data Analysis**

Umeå, November 14-16, 2018

# National Bioinformatics Infrastructure Sweden (NBIS)

## SciLifeLab Bioinformatics Platform

***Jeanette Tångrot***

*National Bioinformatics Infrastructure Sweden (NBIS)*

*Computational Life Science Cluster (CLiC)*

*Umeå University*

# What is SciLifeLab?

Government: One of the three main research infrastructures in Sweden, along with MAX-IV and ESS

Solna
Uppsala
Umeå
Linköping
Lund
Gothenburg

# SciLifeLab

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

**www.scilifelab.se**

# SciLifeLab platforms

SciLifeLab national service

VR

National Genomics Infrastructure

Metabolomics

Single-cell Biology

National Bioinformatics Infrastructure Sweden

Bengt Persson

SNIC

Computer resources free for Swedish researchers

e.g. NGI

Bioinformatics platform

Samples

Data generation

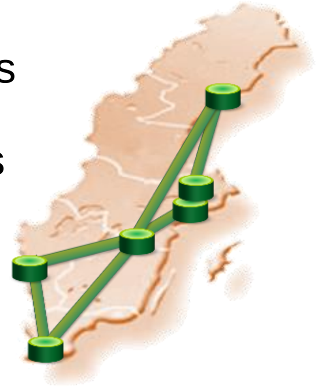
Standard analyses

Project-specific analyses

Results

Data to be stored and published

# NBIS – the Bioinformatics Platform



Distributed research infrastructure with nodes at the 6 large university towns

The Swedish node in ELIXIR, the European infrastructure for bioinformatics

Staff: ~75 full-time equivalents

Services:

- **Support**
- **Infrastructure / Tools**
- **Training**

Funding from:

# Two tracks for support!

If data is **not** available: **Consultation**

Free, always available. No data will be handled by NBIS.

If data is available: **Support**

**Short Term Support - for a fee**

- Always open for applications
- Quick turn-over
- Work usually starts within 2 weeks from signed agreement
- Most projects 30-60 hours, stretching over 1-3 months
- 800 kr/h, time logged per hour used

**Long Term Support - free**

- Application rounds, 3 times/year
- Scientific committee
- Projects chosen for suppor scientific excellence
- 500 h of sup over 1-2 years
- Hands-on invol om research group mandatory

Next deadline:
January 2019

+ drop-in sessions every week @ all 6 sites

# Two tracks for support!

If data is **not** available: **Consultation**

Free, always available. No data will be handled by NBIS.

If data is available: **Support**

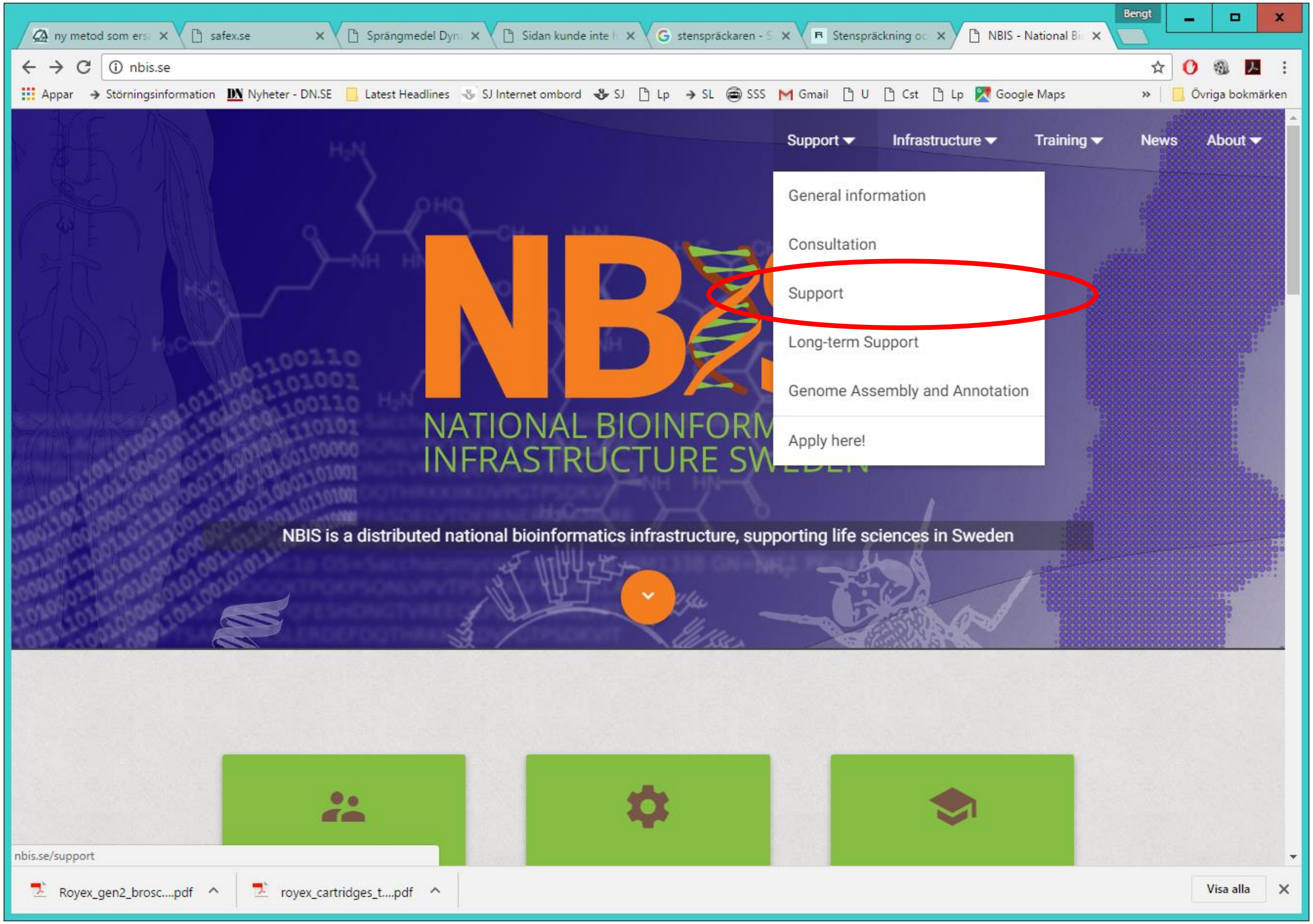
**Short Term Support - for a fee**
- Always open for applications
- Quick turn-over
- Work usually starts within 2 weeks from signed agreement
- Most projects 30-60 hours, stretching over 1-3 months
- 800 kr/h, time logged per hour used

**Long Term Support - free**
- Application rounds, 3 times/year
- Scientific committee
- Projects chosen for support based on scientific excellence
- 500 h of sup rt, usu ng over 1-2 years
- Hand group m

+ drop-in sessions every week

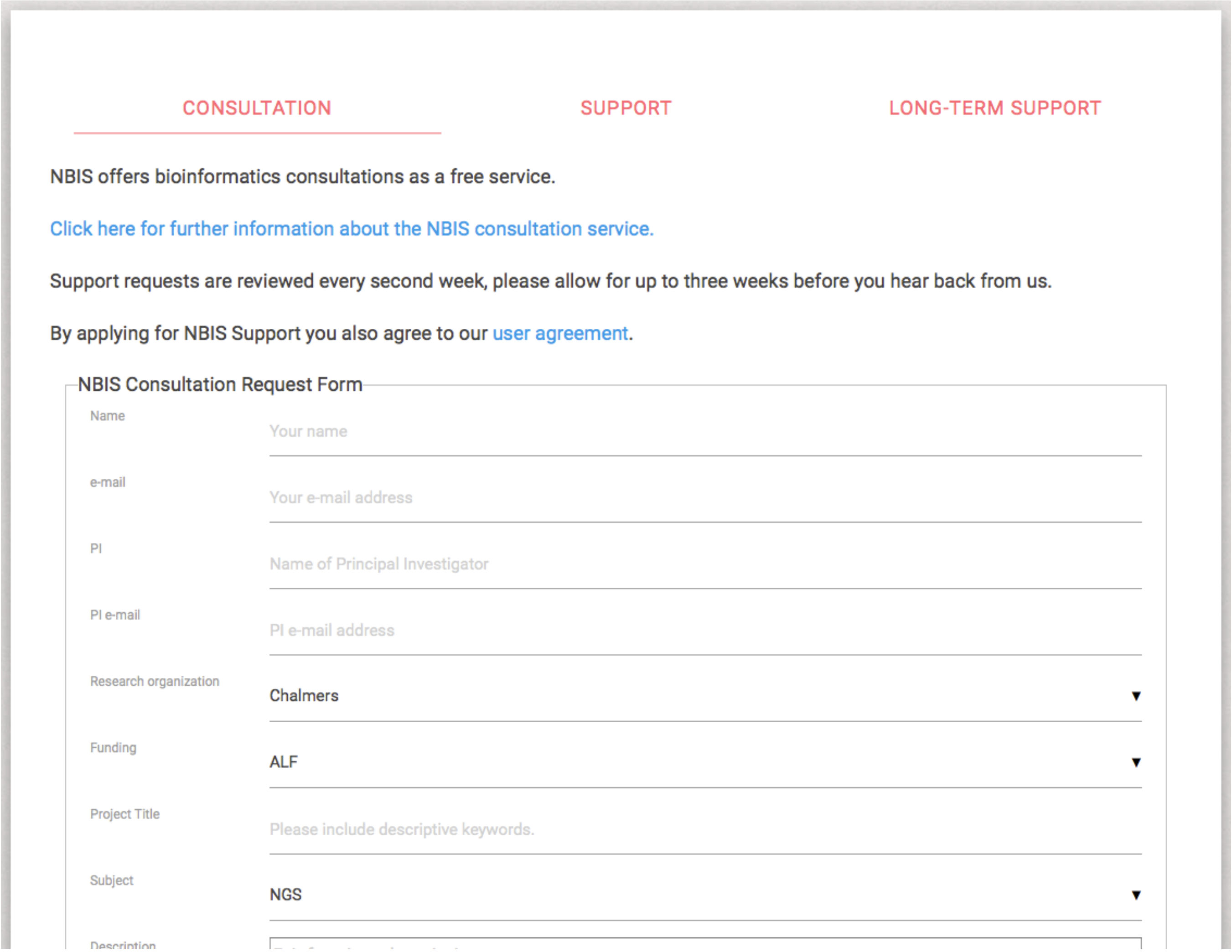
Thursdays 10 am
Even weeks: common ground
Odd weeks: KBC cafeteria

# How to get support
# nbis.se



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

# Support form



NBIS offers bioinformatics consultations as a free service.

Click here for further information about the NBIS consultation service.

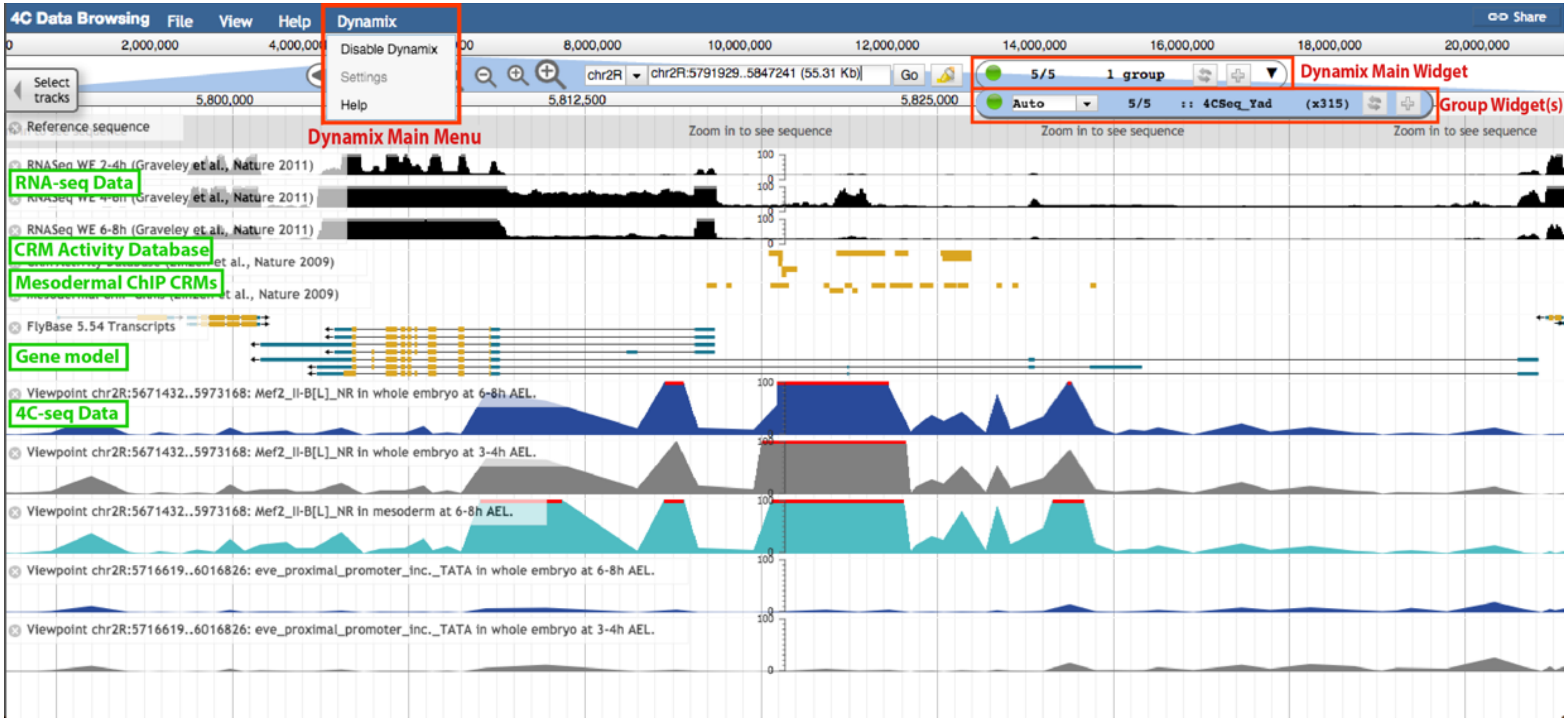
Support requests are reviewed every second week, please allow for up to three weeks before you hear back from us.

By applying for NBIS Support you also agree to our user agreement.

NBIS Consultation Request Form

| | |
|---|---|
| Name | Your name |
| e-mail | Your e-mail address |
| PI | Name of Principal Investigator |
| PI e-mail | PI e-mail address |
| Research organization | Chalmers ▼ |
| Funding | ALF ▼ |
| Project Title | Please include descriptive keywords. |
| Subject | NGS ▼ |
| Description | |

# Focus area:
# Genome assembly and annotation



- Highly specialized staff and robust pipelines
- Tight user interaction
- Numerous manual and semi-manual QC steps
- Supports ENA submission

*Cost effective with high quality!*

# Tools and infrastructure

## 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

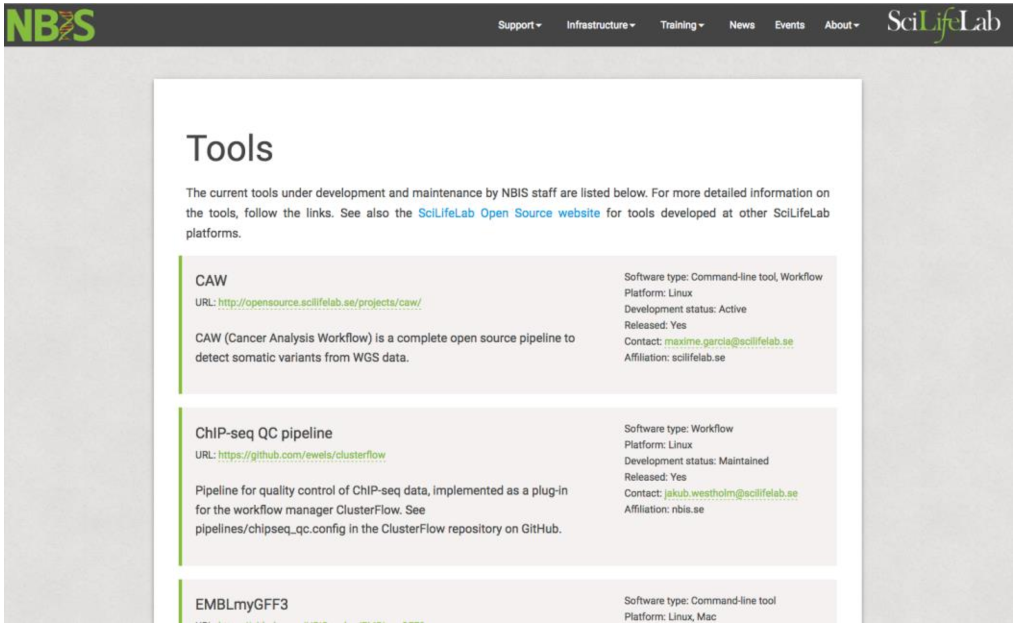
- MrBayes
- Structure prediction web services

## Assembly and annotation

- Falcon on Uppmax
- ENA submission help

## Other tools and resources

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



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.

CAW

URL: http://opensource.scilifelab.se/projects/caw/

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: maxime.garcia@scilifelab.se
Affiliation: scilifelab.se

ChIP-seq QC pipeline

URL: https://github.com/ewels/clusterflow

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: jakub.westholm@scilifelab.se
Affiliation: nbis.se

EMBLmyGFF3

Software type: Command-line tool
Platform: Linux, Mac

http://nbis.se/infrastructure/tools/

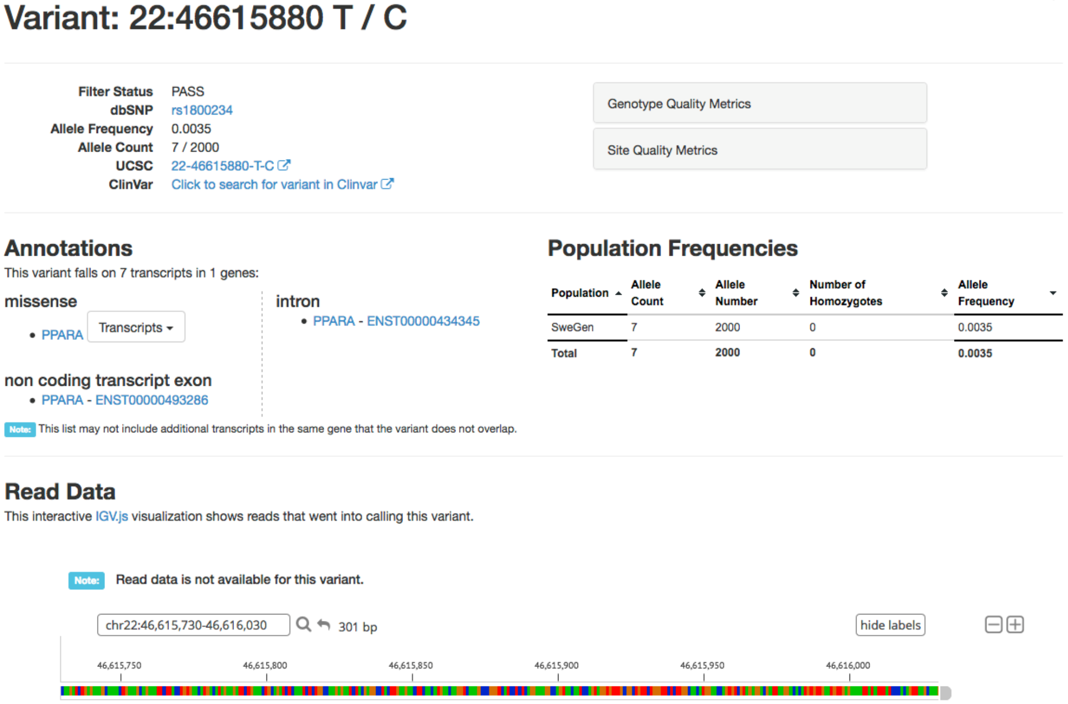
# SweGen: 1000 Swedish genomes

**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/#/

1st 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

### Annotations

This variant falls on 7 transcripts in 1 genes:

missense
- PPARA

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.

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

NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

# National PhD level bioinformatics courses

SciLifeLab

Joint efforts by SciLifeLab staff and researchers (+ more)

**Introduction to omics data analysis**
4 x Introduction to Bioinformatics using NGS data, 5 days (Manfred Grabherr)
1 x Omics data analysis, 10 days (Erik Fredlund, Lukas Orre)

**Advanced topical courses**
2 x RNAseq, 3 days (Johan Reimegård)
1 scRNA, 2 days (Åsa Björklund)
1 x Genome assembly, 2 days (Henrik Lantz)
1 x Genome annotation, 2 days (Henrik Lantz)
1 x Metagenomics, 2 days (Anders Andersson, Thijs Ettema)
1 x Proteomics, 4 days (Fredrik Levander)

**Programming and reproducible research**
1 x Reproducible research, 2 days (Rasmus Ågren, Leif Wigge)
2 x Python, 2 days (Frédéric Haziza, Sergiu Netotea)
2 x Perl, 5 days (Matt Webster)
1 x R Summer School, 10 days: RaukR (Marcin Kierczak)
1 x R programming Foundations for Life Scientists, 5 days (Marcin Kierczak)
1 x Unix/Linux Tutorial for beginners, 3 days (Mihaela Martis)

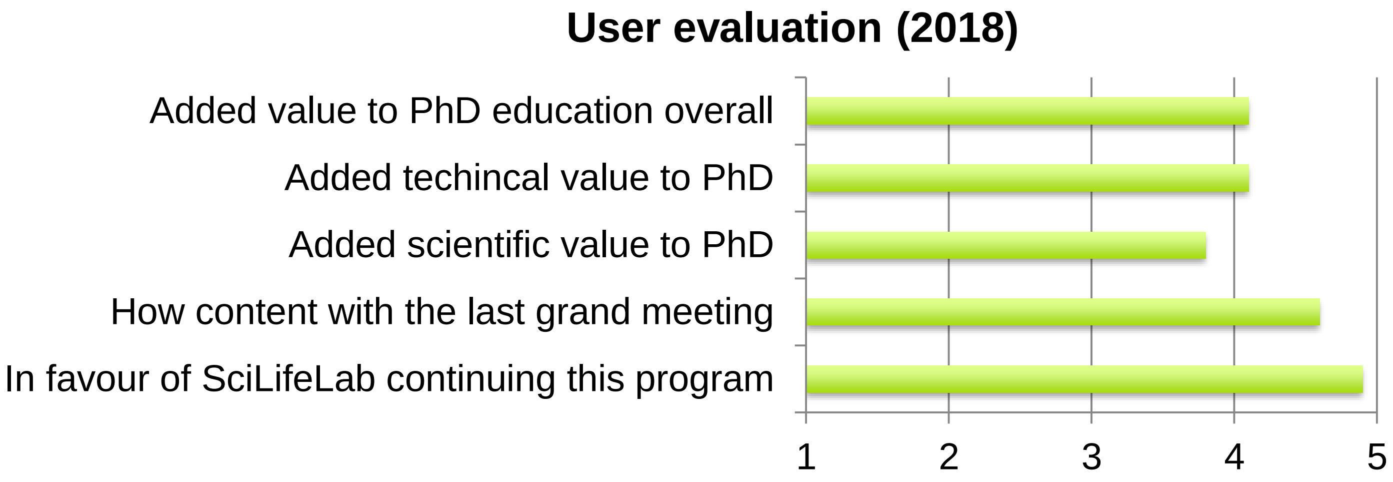
500+ students/year
Very positive course evaluations
Teacher-dense
Good for networking

New 2018!
R Summer School,
Visby

[www.scilifelab.se/education/courses/](http://www.scilifelab.se/education/courses/)
[www.nbis.se/training/events.html](http://www.nbis.se/training/events.html)

# Bioinformatics Advisory Program

- PhD students get a senior bioinformatician as advisor for 2 years
- Monthly project meetings with advisor
- Two grand meetings per year for networking and knowledge exchange



Next call planned to open end of 2018

www.scilifelab.se/education/mentorship/

# NBIS coordinates the Swedish node in ELIXIR – the European infrastructure for biological information

Data infrastructure for Europe's
life-science research
20 members + 1 observer

Five platforms

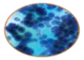
- *Data*
- *Interoperability*
- *Tools*
- *Compute*
- *Training*

Four user communities

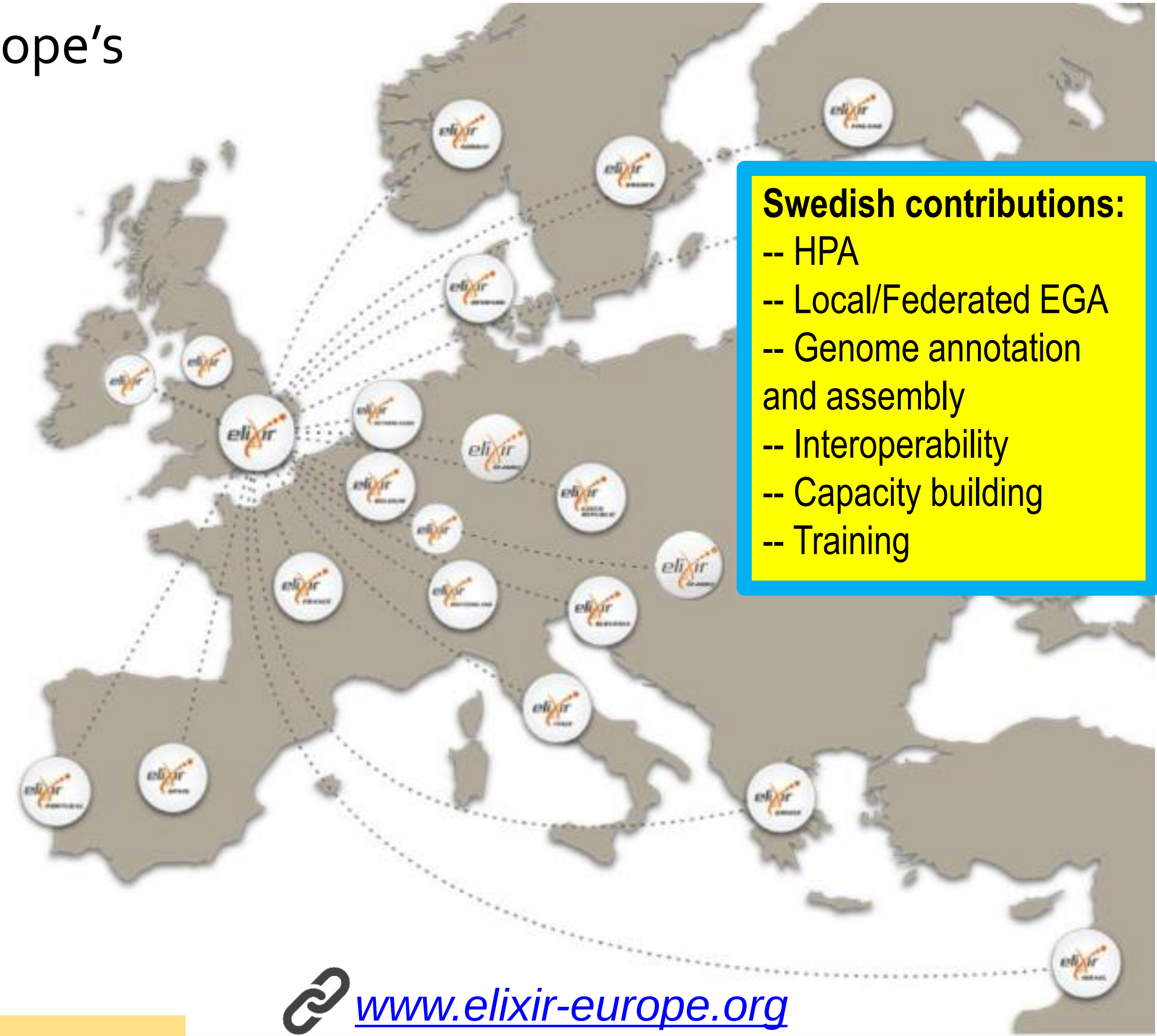
- *Marine metagenomics*

- *Crop and forest plants*

- *Human data*



- *Rare diseases*



**Swedish contributions:**
-- HPA
-- Local/Federated EGA
-- Genome annotation and assembly
-- Interoperability
-- Capacity building
-- Training

At the global level interactions with:
RDA, NIH BD2K, GA4GH

[www.elixir-europe.org](http://www.elixir-europe.org)

[@ELIXIREurope]()

# We're here for you!

# www.nbis.se

NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

# Practical information for the workshop

**Attendance sheet**

- Please sign morning and afternoon

**Fika and lunches together every day**

- 10, 12:00, 15
- Fika outside the lecture room

**Course dinner**

- Tue 15/11, at 18:00, TC
- Sign up latest today before lunch
- Included: pre-ordered menu, 1 drink