SciLifeLab Bioinformatics Platform

National Bioinformatics Infrastructure Sweden (NBIS)

Björn Nystedt
www.nbis.se

RNA course

Uppsala 13.03.2017

SciLifeLab

SciLifeLab

SciLifeLab

National service
The Swiss army knife for Swedish Life Science researchers

Local scientific center

Director: Olli Kallioniemi
Co-director: Lena Claesson-Welsh

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

National Genomics Infrastructure

Next Generation Diagnostics

Single-cell omics

National Bioinformatics Infrastructure Sweden

Bengt Persson

Merge of BILS, WABI and more; complete 2016. National, distributed

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

VR

SNIC

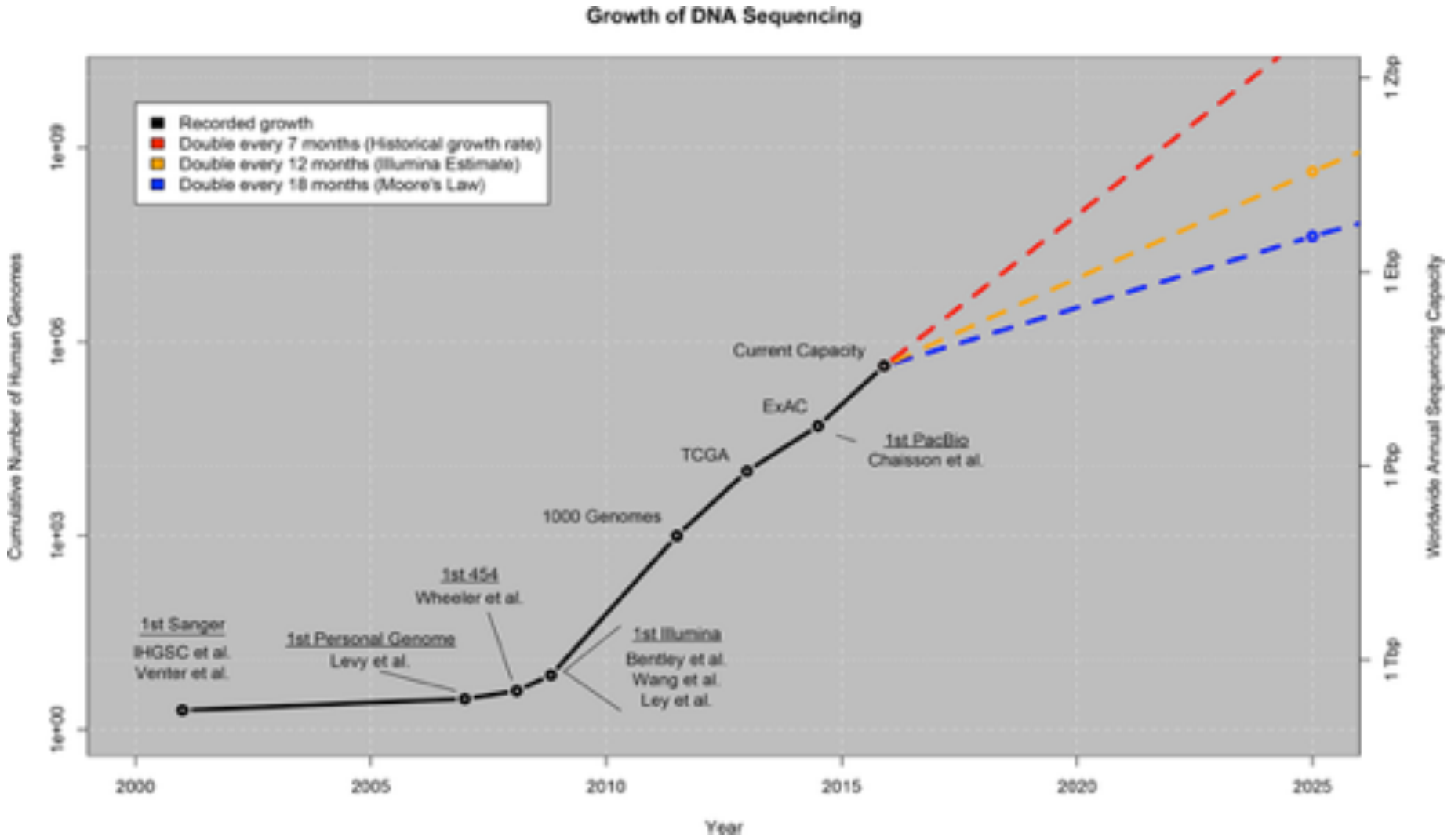
Computer resources free for Swedish researchers

NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

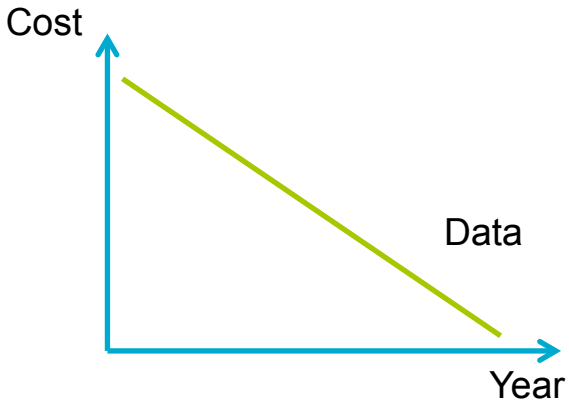
SciLifeLab

Bioinformatics as infrastructure

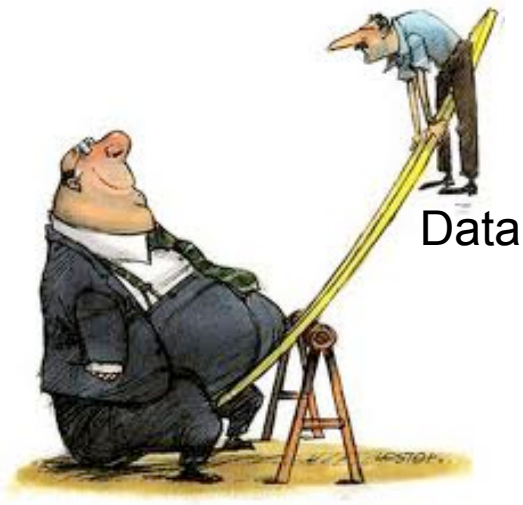
Data growth


Growth of DNA Sequencing
Recorded growth
Double every 7 months (Historical growth rate)
Double every 12 months (Illumina Estimate)
Double every 18 months (Moore's Law)
Current Capacity
ExAC
1st PacBio
Chaisson et al.
TCGA
1000 Genomes
1st 454
Wheeler et al.
1st Sanger
IHGSC et al.
Venter et al.
1st Personal Genome
Levy et al.
1st Illumina
Bentley et al.
Wang et al.
Ley et al.
Cumulative Number of Human Genomes
1e+00
1e+03
1e+06
1e+09
Worldwide Annual Sequencing Capacity
1 Tbp
1 Pbp
1 Ebp
1 Zbp
2000
2005
2010
2015
2020
2025
Year

Production is cheap, analysis is not



"Per base"

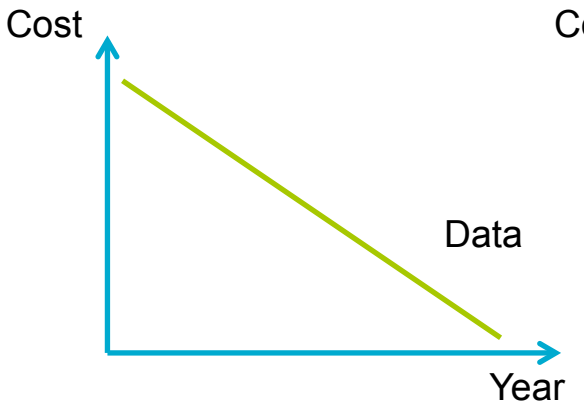


Data scientists

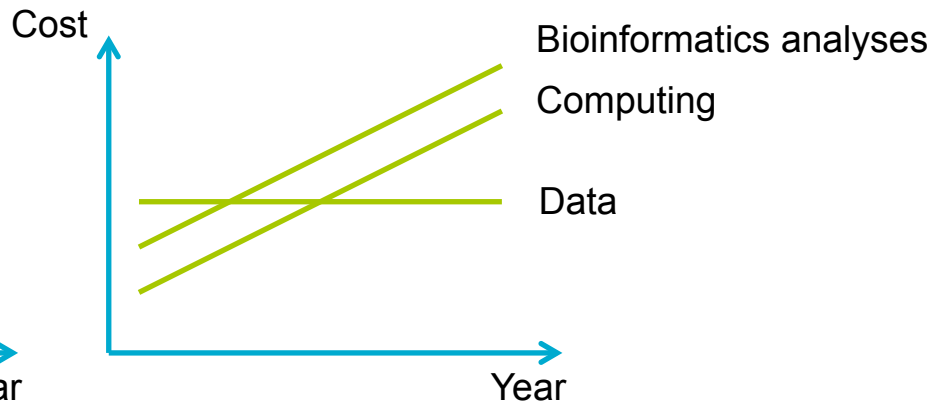
Data

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.

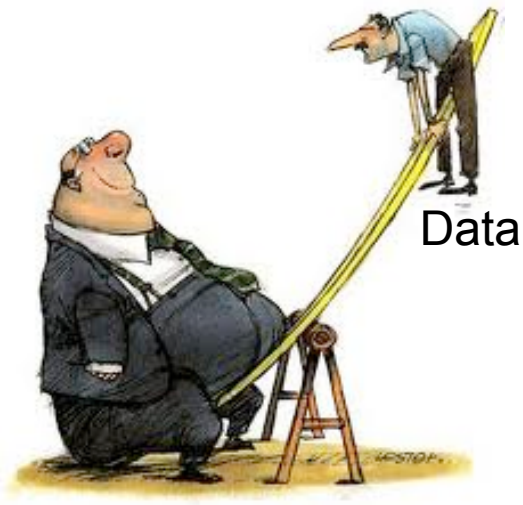
Production is cheap, analysis is not



“Per base”



“Per project”



Data scientists

Data

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

elixir

ISBE Infrastructure for Systems Biology Europe

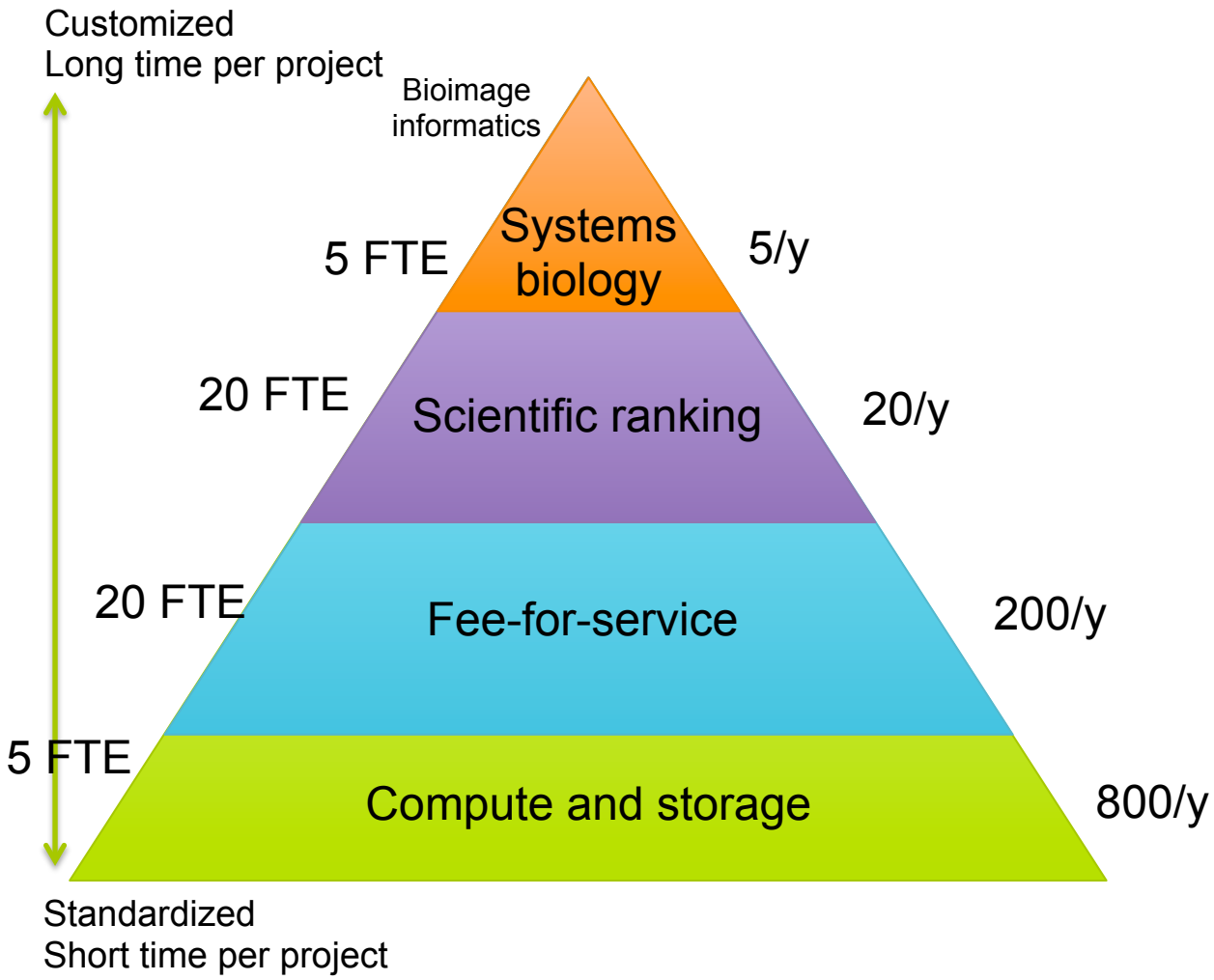
NBIS
NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

4 facilities, ~60 FTEs

- **Support and Infrastructure**
  Wide competence in bioinformatics, Assembly/Annotation, SysDev
- **Long-term support (WABI)**
  Large collaborative projects selected by scientific ranking
- **Systems biology**
  Network analyses and Integrative bioinformatics
- **Compute and storage**
  Computational and storage resources for bioinformatics, especially next-generation sequencing

NBIS



Custom-tailored support

Training

Systems development

Data management

User benefits

Design of compute, storage, archiving

Study design, Consultation, Grant applications

Compute resource allocation

Support

Tools

Training

Data submission, Reproducibility

Support

NBIS NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

SciLifeLab

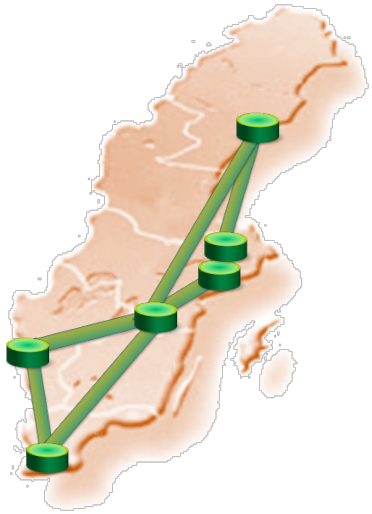
Custom-tailored support

www.scilifelab.se/platforms/bioinformatics/
www.nbis.se

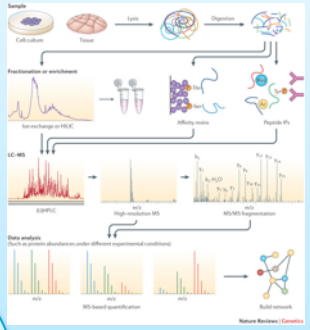
- Study design consultation (free)
  www.nbis.se/support/supportform/index.php
  + drop-in sessions every week @ 6 sites

- Short- and Medium-term support (User fee 800 kr/h)
  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



Bioinformatics support



Proteomics

Genomics

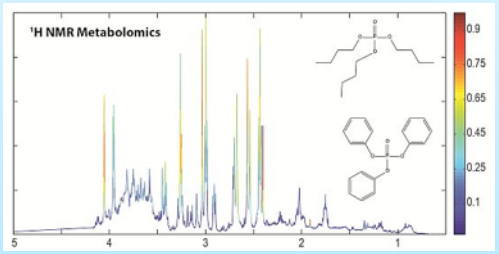
Systems biology

2 tracks!

- Fee-for-service (800kr/h)
  Rapid turnaround

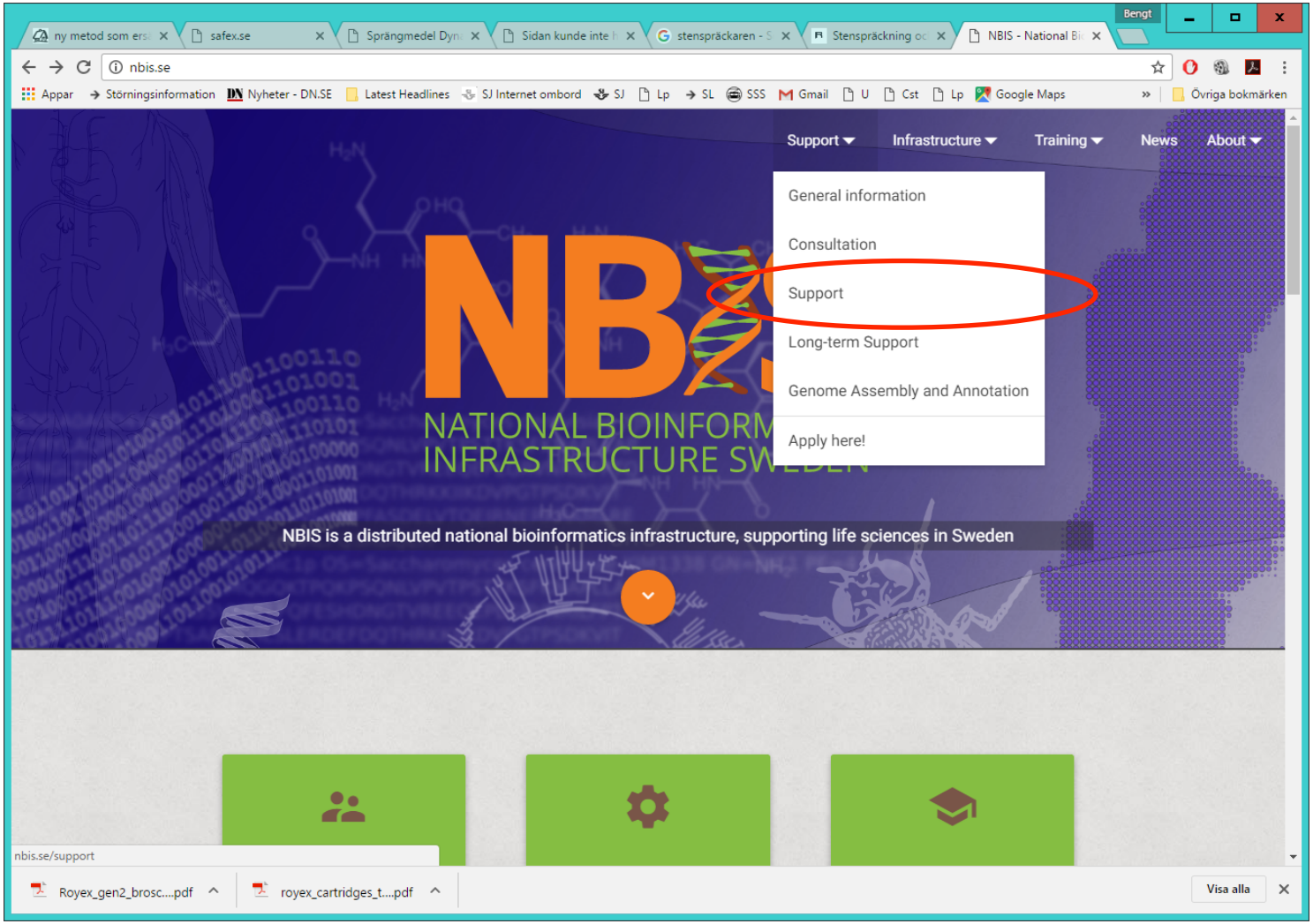
- Scientific ranking (free)
  "Long-term Support"
  3 rounds/year

Biostatistics

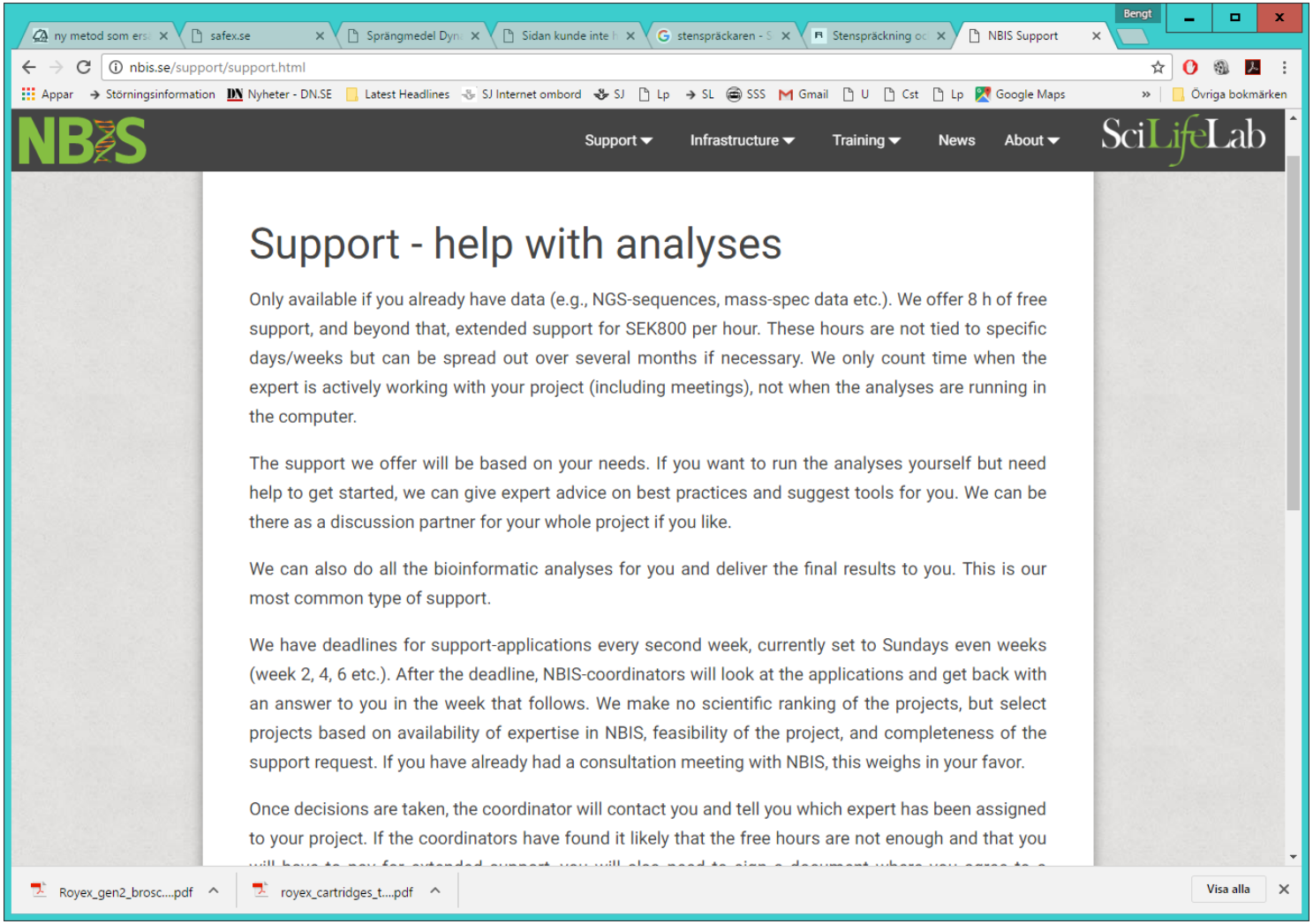


Metabolomics

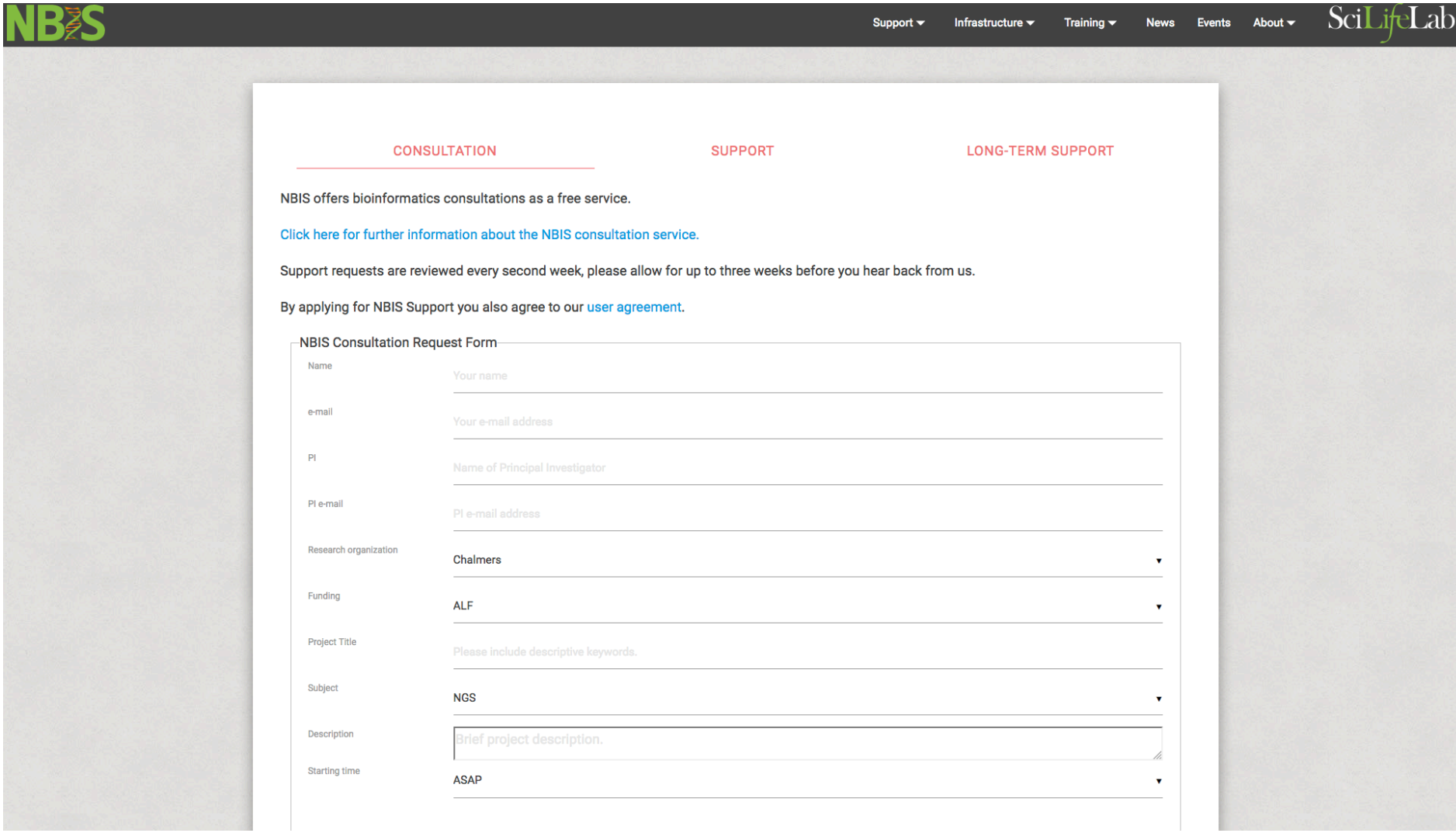
How to get support nbis.se



Support information

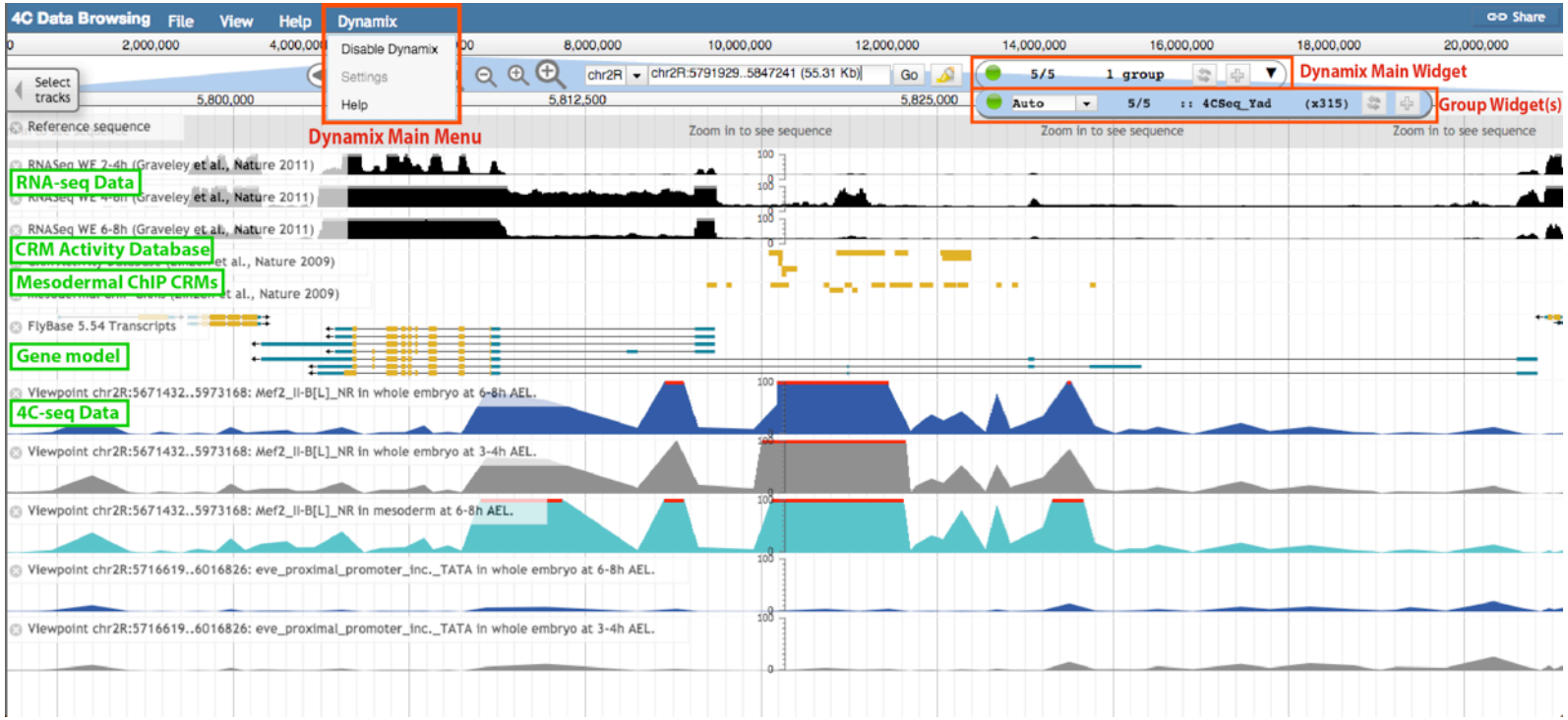


Support form



Genome assembly and annotation

Henrik Lantz



- 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

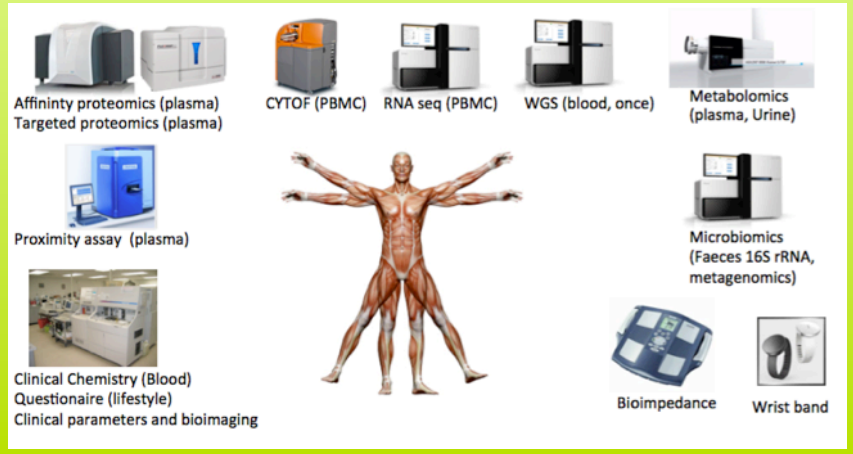
4 FTE, joint effort by Long-term Support and Systems Biology
Projects apply in the regular Long-term Support calls

Combine data from SciLifeLab platforms

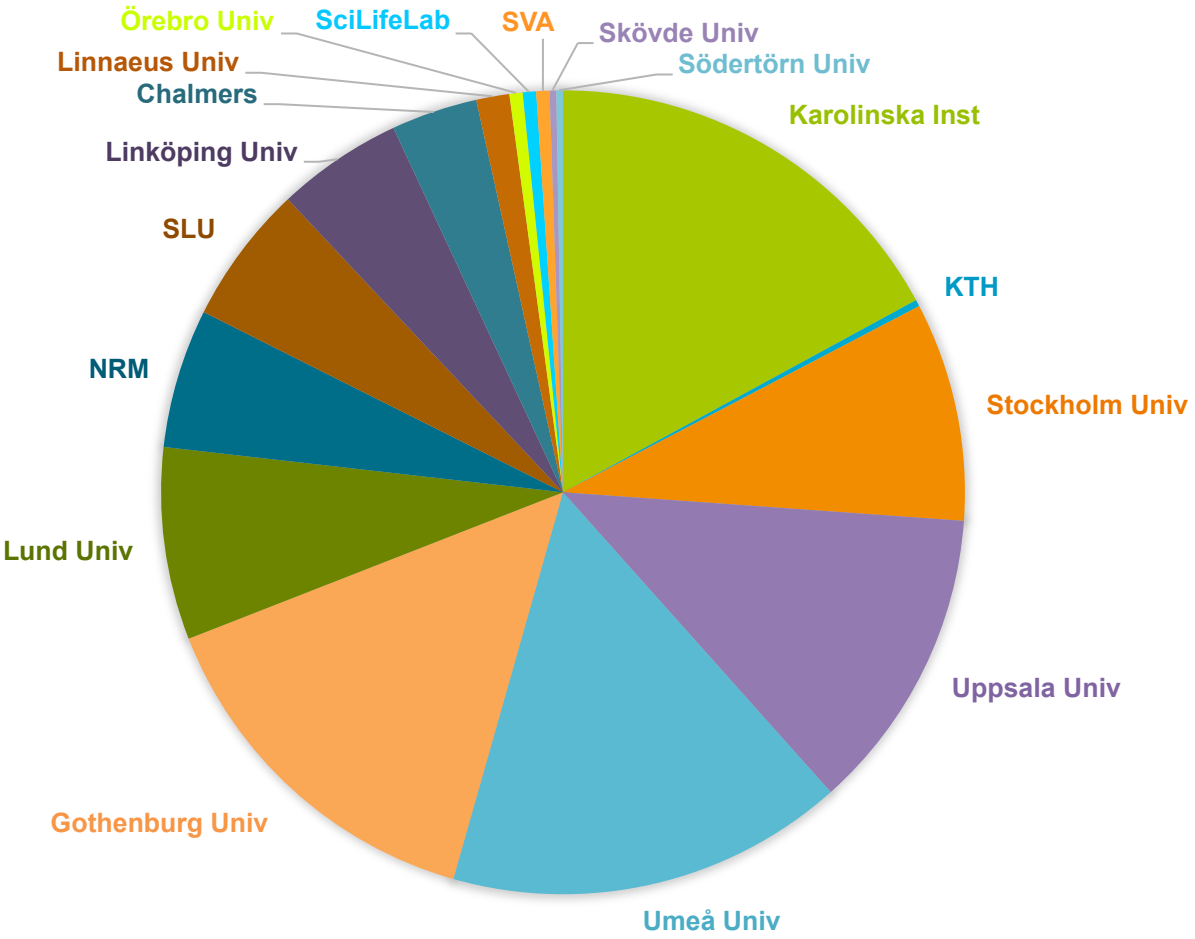
- Building **tools** and **resources** for handling very large and/or complex biological data sets
- Typically performed in the context of longer support projects
- State-of-the-art analytical methods for integrating multi-modal biological data sets, eg
  - Machine learning/deep learning
  - Graph-based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

Support track for integrative projects

First call Feb 2016; First few projects initiated
Involves extensive integration of data



Geographical Distribution of Projects 2015



Tools

Tools and infrastructure

https://docs.google.com/spreadsheets/d/1PrehKn2eb0ymfaFtCfvbLrOSKtpTL3qLcWZ2YwoXOIU/edit#gid=0

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

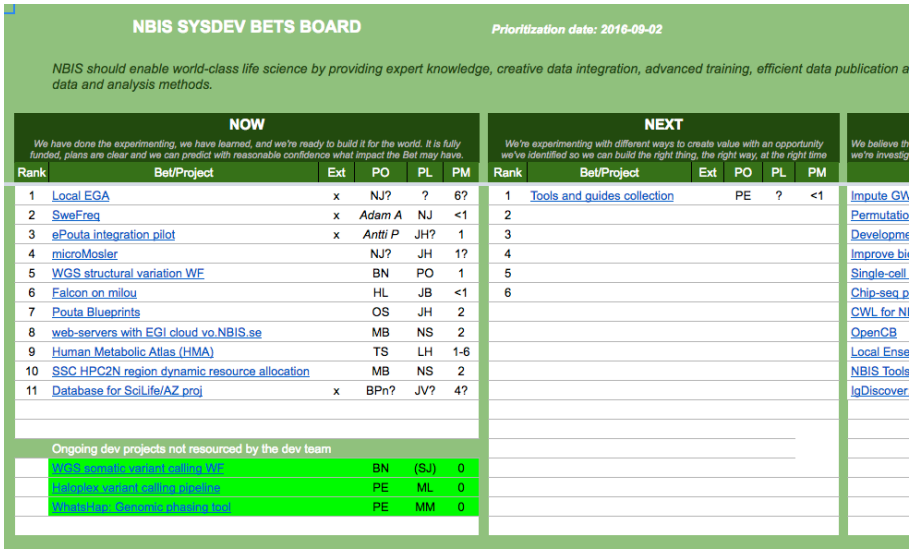
- MrBayes
- Structure prediction web services

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 SYSDEV BETS BOARD *Prioritization date: 2016-09-02*

NBIS should enable world-class life science by providing expert knowledge, creative data integration, advanced training, efficient data publication a data and analysis methods.

NOW — *We have done the experimenting, we have learned, and we're ready to build it for the world. It is fully funded, plans are clear and we can predict with reasonable confidence what impact the Bet may have.*

Rank	Bet/Project	Ext	PO	PL	PM
1	Local EGA	x	NJ?	?	6?
2	SweFreq	x	*Adam A*	NJ	<1
3	ePouta integration pilot	x	*Antti P*	JH?	1
4	microMosler		NJ?	JH	1?
5	WGS structural variation WF		BN	PO	1
6	Falcon on milou		HL	JB	<1
7	Pouta Blueprints		OS	JH	2
8	web-servers with EGI cloud vo.NBIS.se		MB	NS	2
9	Human Metabolic Atlas (HMA)		TS	LH	1-6
10	SSC HPC2N region dynamic resource allocation		MB	NS	2
11	Database for SciLife/AZ proj	x	BPn?	JV?	4?
	Ongoing dev projects not resourced by the dev team				
	WGS somatic variant calling WF		BN	(SJ)	0
	Haloplex variant calling pipeline		PE	ML	0
	WhatsHap: Genomic phasing tool		PE	MM	0

NEXT — *We're experimenting with different ways to create value with an opportunity we've identified so we can build the right thing, the right way, at the right time*

Rank	Bet/Project	Ext	PO	PL	PM
1	Tools and guides collection		PE	?	<1
2					
3					
4					
5					
6					

Open prioritization and background descriptions

Tools and development projects needs to be much more visible! Work in progress…

SweGen: 1000 Swedish genomes

SciLifeLab

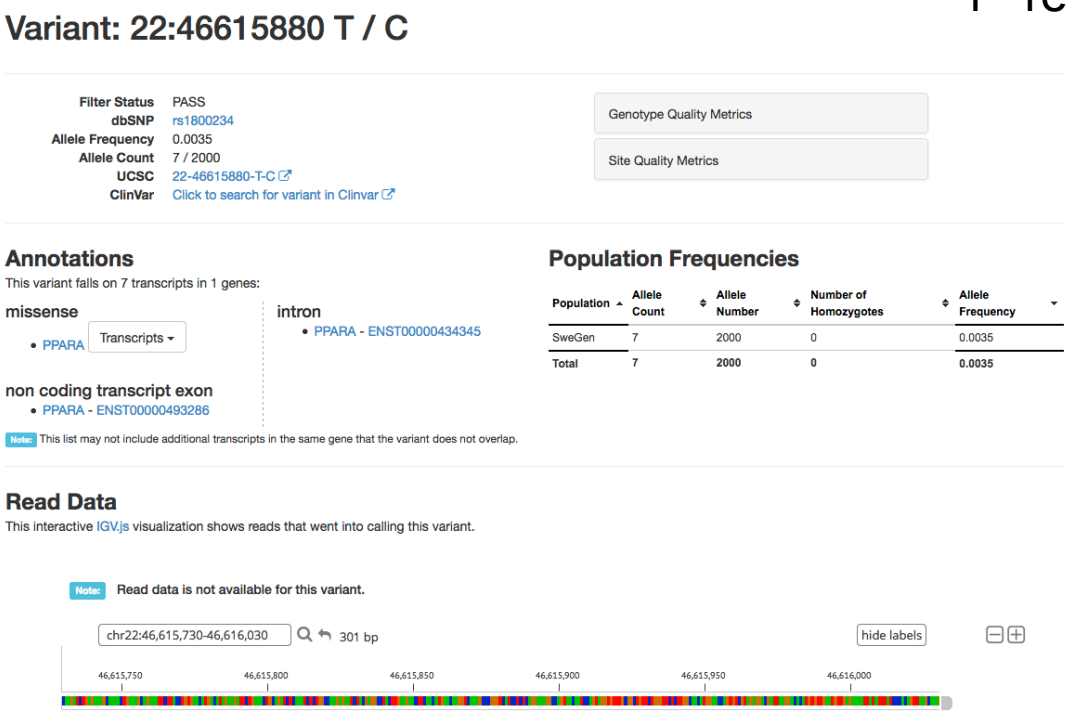
SweGen Variant Frequency Database

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

SweGen

https://swefreq.nbis.se/#/

1st release October 2016!



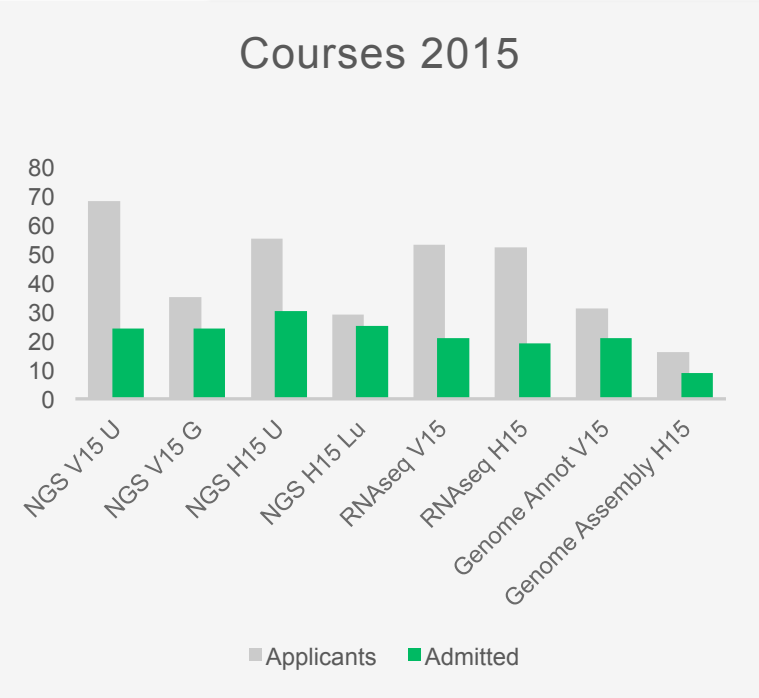
Funding: SciLifeLab
Sequencing: NGI
Variant calling: NGI
QC: NBIS
Data access interface: NBIS

Training

Outreach & Training

- Bioinformatics Drop-In
  - Weekly at all sites – initial consultations
- 20-odd courses every year
  - Introduction to Bioinformatics using NGS
  - Introduction to Linux
  - Perl programming
  - Introduction to genome annotation
  - Introduction to multivariate analysis
  - RNA-seq
  - Advanced workshop on NGS data analysis
  - Advanced functional genomics
  - Advanced bioinformatics
- Additional local activities
- Bioinformatics Advisory programme
  - Mentorship in bioinformatics

Gender balance:
54% female / 46% male



From spring 2017, we plan to double our training efforts to match the increased demands from the scientific community

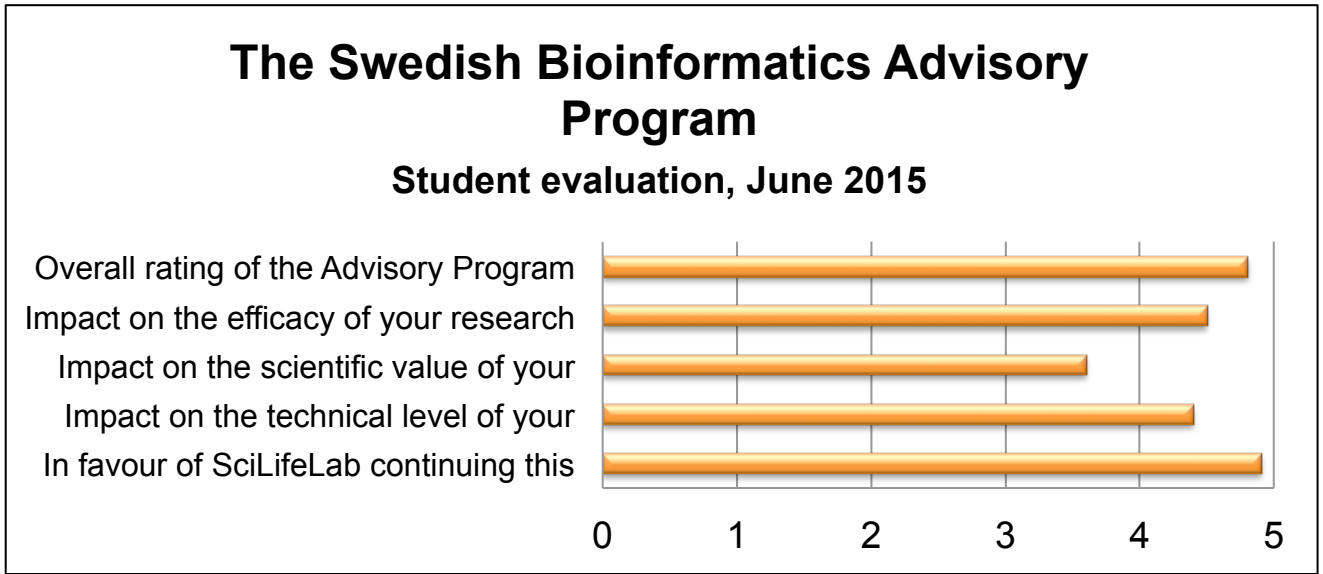
www.scilifelab.se/education/courses/
www.nbis.se/training/events.html

Teaching and mentoring

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/

Recent call (2017/2018): 111 applicants for 15 places (!)



Elixir

Why ELIXIR?

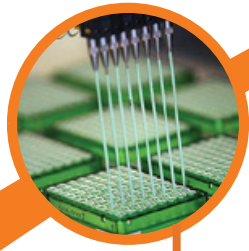
- Creating a robust infrastructure for biological information is a **bigger** task than any individual organisation or nation can take on alone
- These are issues of such complexity that no single institution or country can tackle alone

- Biology has by far the largest research community:
  - ~3 million life science researchers in Europe
  - >7 million web hits a day at EMBL-EBI alone

ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data

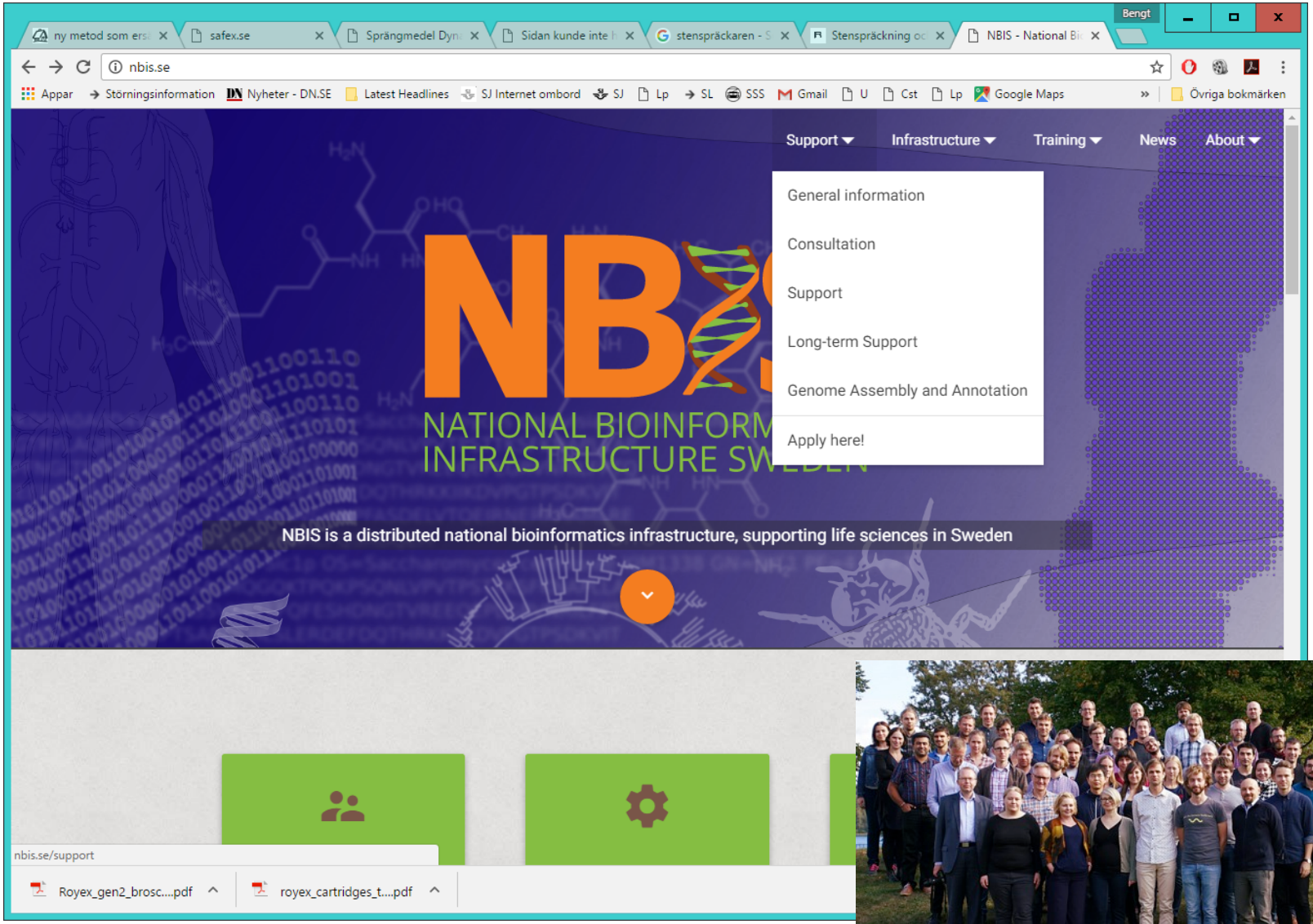
medicine

agriculture



bioindustries

environment

ELIXIR underpins life science research – across academia and industry

We're here for you!
nbis.se