

NBIS

The national bioinformatics infrastructure Sweden

www.scilifelab.se/platforms/bioinformatics/

Björn Nystedt, bjorn.nystedt@scilifelab.se

SciLifeLab

SciLifeLab

National service

Local scientific center

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.



Director (July 2015)
Olli Kallioniemi

Co-director
Kerstin Lindblad-Toh

2010: Strategic research initiative
2013: National resource

www.scilifelab.se

Main sites (Stockholm/Uppsala)



Uppsala (Biomedical center)



Stockholm (KI campus)

9 national service platforms
200+ research groups

SciLifeLab provides state-of-the art services

- **NGI (One of the largest sequencing centers in Europe)**
X-Ten, HiSeq, MiSeq, PacBio, IonTorrent, MinIon, Optical mapping
- **Clinical Diagnostics**
Sequencing and other omics for new clinical applications
- **Bioinformatics**
Approaching 100 FTE for custom-tailored project support, methods and systems development, data publishing
- **Functional Genomics**
Single-cell transcriptomics, genomics, and proteomics
- ...

National facilities

Affinity Proteomics

Biobank Profiling
Cell Profiling
Fluorescence Tissue Profiling
Mass Cytometry
PLA Proteomics
Protein and Peptide Arrays
Tissue Profiling

Bioimaging

Advanced Light Microscopy
Fluorescence Correlation Spectroscopy

Bioinformatics

Bioinformatics Compute and Storage (UPPNEX)
Bioinformatics Long-term Support (WABI)
Bioinformatics Short-term Support and Infrastructure (BILS)

Chemical Biology Consortium Sweden

Laboratories for Chemical Biology Umeå (LCBU)
The Laboratories for Chemical Biology at Karolinska Institutet (LCBK)
Uppsala Drug Optimization and Pharmaceutical Profiling (UDOPP)

Clinical Diagnostics

Clinical Biomarkers
Clinical Genomics
Clinical Sequencing

Drug Discovery and Development

ADME (Absorption Distribution, Metabolism Excretion) of Therapeutics (UDOPP)
Biochemical and Cellular Screening
Biophysical Screening and Characterization
Human Antibody Therapeutics
In Vitro and Systems Pharmacology
Medicinal Chemistry – Hit2Lead
Medicinal Chemistry – Lead Identification
Protein Expression and Characterization

Functional Genomics

Eukaryotic Single Cell Genomics
Karolinska High Throughput Center (KHTC)
Microbial Single Cell Genomics
Single Cell Proteomics

National Genomics Infrastructure

NGI Stockholm (Genomics Applications)
NGI Stockholm (Genomics Production)
NGI Uppsala (SNP&SEQ Technology Platform)
NGI Uppsala (Uppsala Genome Center)

Structural Biology

Protein Science Facility

The national bioinformatics infrastructure Sweden (NBIS) constitutes the Bioinformatics platform at SciLifeLab

NBIS will be official as of 2016. Most of the activities are already fully operational, and NBIS is the continuation and expansion of present activities, unifying all national bioinformatics support into a single infrastructure.

SciLifeLab platforms

SciLifeLab

National
Genomics
Infrastructure

Joakim Lundeberg
Ann-Christine Syvänen
Ulf Gyllensten



National
Bioinformatics
Infrastructure
Sweden

Bengt Persson



Clinical
Diagnostics

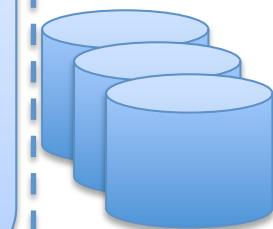
Lars Engstrand



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

VR

SNIC



Computer
resources
free for
Swedish
researchers

Data producing platforms
(eg sequencing)

BIOINFORMATICS
PLATFORM

Development
Lab Bioinfo

Project support and
tools development

Data

Platform
accounts

Supercomputer center

Customer project
accounts

DELIVERY

PREDEFINED ANALYSES

CUSTOM-TAILORED ANALYSES

The Bioinformatics Platform 2016

Funding

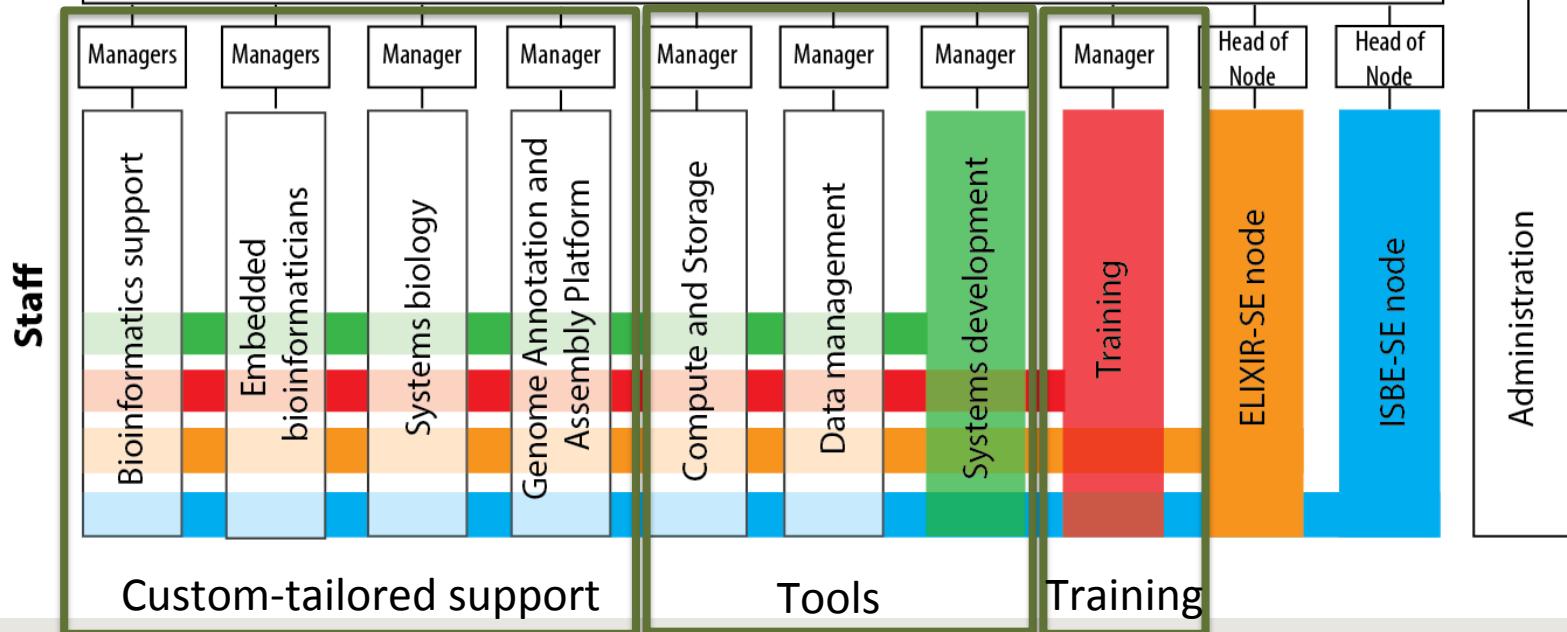
- The Research Council
- SciLifeLab
- KAW foundation
- Host universities

Recently applied at the Research Council as continued national infrastructure 2016-2023. Decision late 2015.



Reference Group

Management (Director + Managers)



Why bioinformatics infrastructure?

A continuous technical scale-up will provide an unprecedented amount of heterogeneous omics data

- Support, Tools, Training

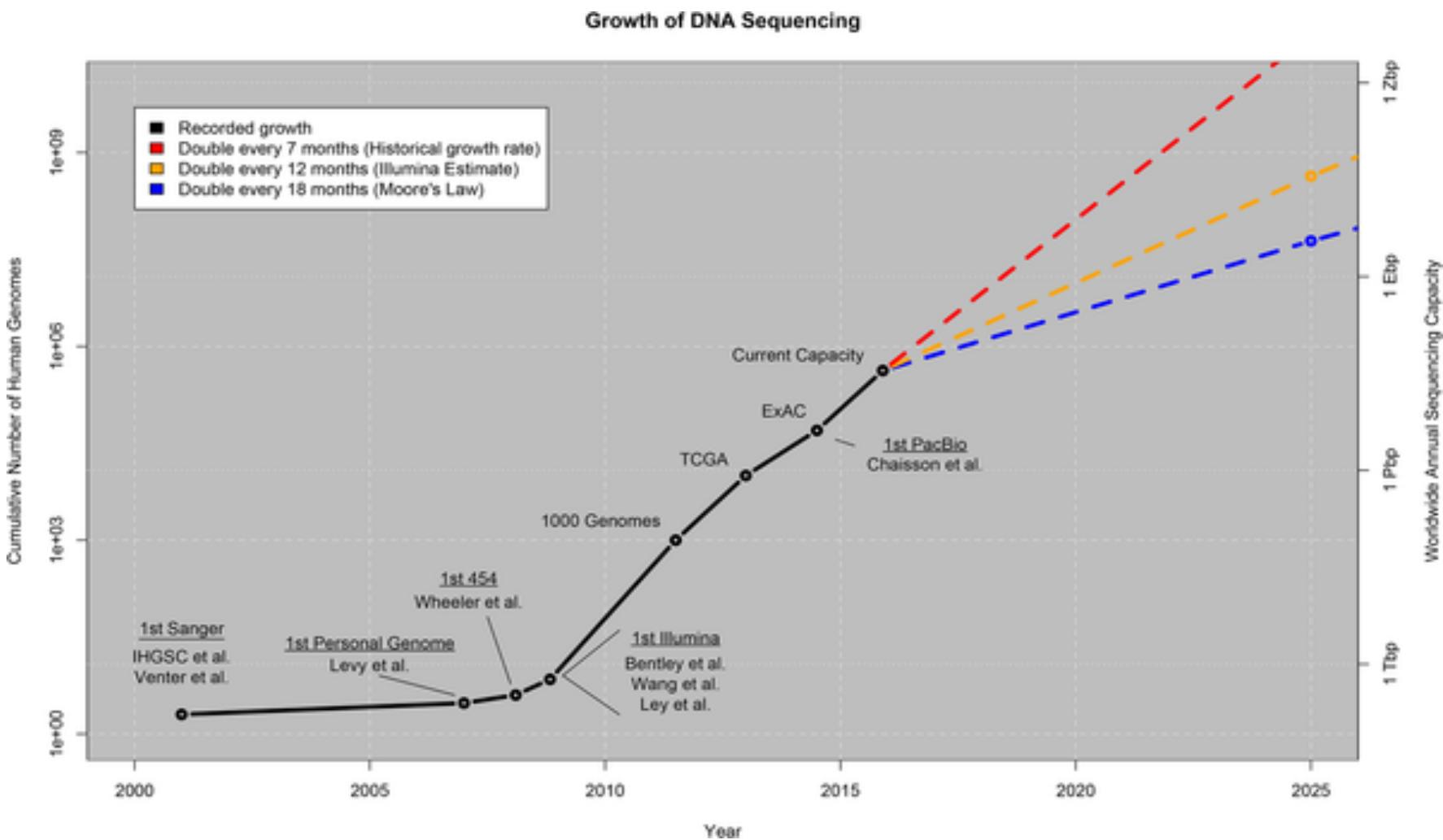
System-level analyses in biomedical research will transform life science

- Strategic positioning in systems biology

Large-scale omics is will make a major leap into translational research and diagnostics

- Method adaptation and expert advice

Fig 1. 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>

Table 1. Four domains of Big Data in 2025.

Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous data and analysis
	Real-time processing	Metadata analysis		Variant calling, ~2 trillion central processing unit (CPU) hours
	Massive volumes			All-pairs genome alignments, ~10,000 trillion CPU hours
Distribution	Dedicated lines from antennae to server (600 TB/s)	Small units of distribution	Major component of modern user's bandwidth (10 MB/s)	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

doi:10.1371/journal.pbio.1002195.t001

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>

Short-term Support and Infrastructure

BILS – Bioinformatics Infrastructure for Life Sciences

www.scilifelab.se/facilities/bils/

- BILS is a distributed national research infrastructure supported by the Swedish Research Council (VR)
 - Nodes at each of the 6 large universities
 - Coordination with other bioinformatics activities
 - Training activities, annual symposium/user meeting
 - BILS provides
 - Support/Consultancy (“human eCloud” for life sciences)
80 hours for free, additional hours for a subsidised fee
 - Databases and software tools
Computing and storage in collaboration with SNIC
 - BILS coordinates the Swedish node in the European infrastructure for biological information ELIXIR
 - Annual budget ~3.5 MEUR (2014)
 - ~2.2 MEUR from Swedish Research Council
 - ~1.3 MEUR from SciLifeLab and participating universities
- | | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Director</p>  <p>Bengt Persson</p> | <p>Technical coordinator</p>  <p>Mikael Borg</p> | <p>Proteomics coordinator</p>  <p>Fredrik Levander</p> |
| <p>Genomics coordinators</p>   <p>Magnus Alm-Rosenblad Henrik Lantz</p> | <p>Syst. development coordinator</p>  <p>Jonas Hagberg</p> | <p>Training coordinator</p>  <p>Sara Light</p> |



SciLifeLab



UNIVERSITY OF
GOTHENBURG



SLU



SciLifeLab



Long-term Support

Wallenberg Advanced Bioinformatics Infrastructure
www.scilifelab.se/facilities/wabi/

Tailored solutions – high impact

Applied bioinformatics: 500h free support/project

- Variant analyses in health and disease
- Transcriptomics
- Novel genomes
- Metagenomics

Swedens strongest unit for analyses of large-scale genomic data (13 + 10 FTE)

National committee reviews and selects projects based on scientific quality

Ongoing expansion with staff in Lund, Gothenburg, Linköping and Umeå.

Directors



Siv Andersson Gunnar von Heijne

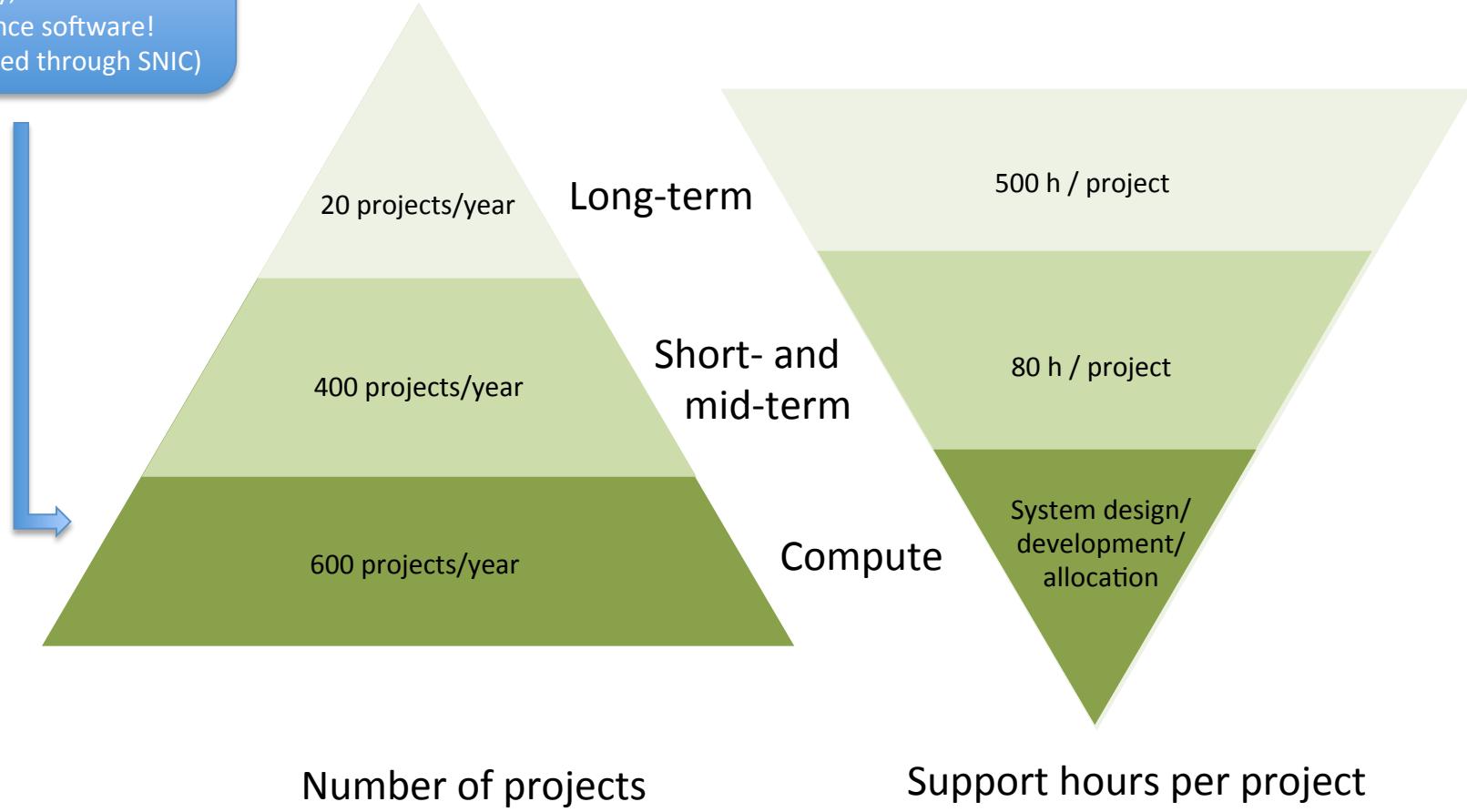
Managers



Björn Nystedt Thomas Svensson

The bioinformatics platform

A centralized computer resource for the entire country, with 250+ life science software!
(Organized through SNIC)



User access to the Bioinformatics Platform

- Compute and storage (free; majority of hardware and system administration belongs to SNIC)

<http://www.uppmax.uu.se/>

- Study design consultation (free)

support@bils.se

- Short-term support (≤ 80 h; free; first come first serve)

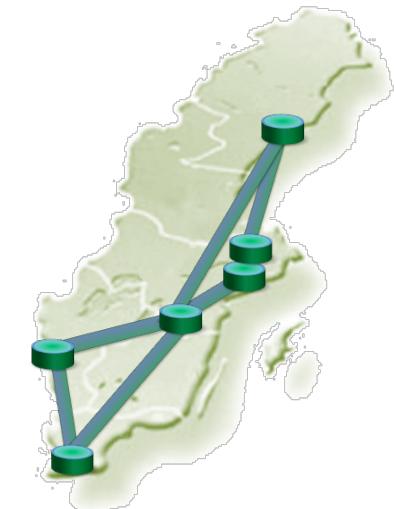
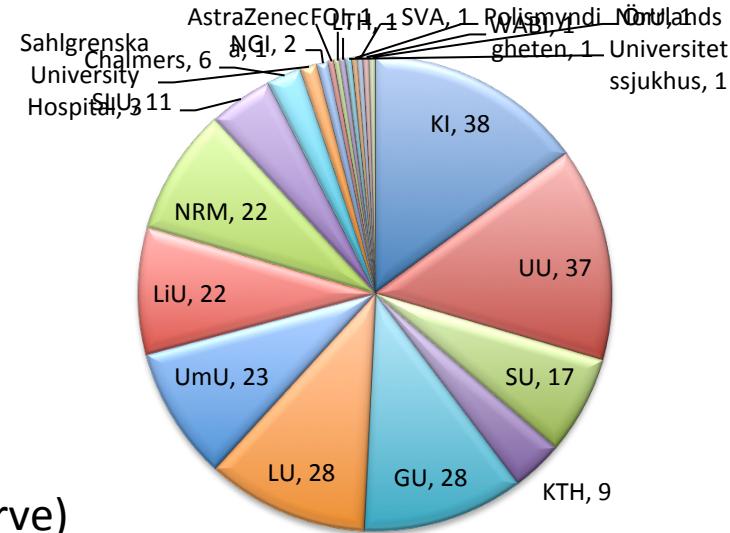
- Medium-term support (> 80 h; user fee; first come first serve)

<http://bils.se/resources/supportform/index.php>

- Long-term support (500h; free: scientific evaluation)

<http://www.scilifelab.se/facilities/wabi/>

Next deadline for applications Oct 30!



Expert teams

Assembly/annotation of novel genomes

support@bils.se

(2 + 2 people, running)

Human WGS ToolBox

Method implementation, community building

<https://wabi-wiki.scilifelab.se/display/SHGATG/>

(2+ people, running)

BigData/Integrative bioinformatics

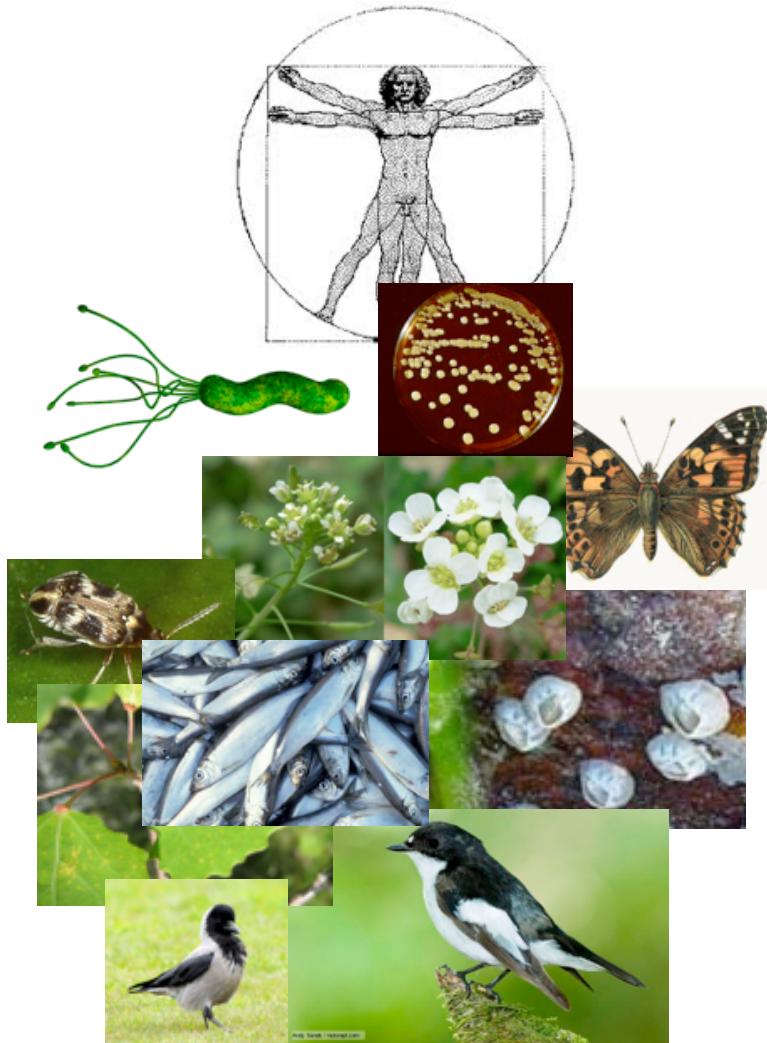
Method development, project support

<http://www.scilifelab.se/facilities/wabi/>

(4 people, hiring now, part of Long-term Support)

“Routinely unique”

Difficult to foresee/automate



- 7 Variant analyses in medical research
- 4 Differential gene expression
- 3 single-cell RNAseq
- 2 Epigenetics
- 1 miRNA
- 1 Allele-specific expression + methylSeq
- 1 Immunorepertoire
- 1 Integrative
- 20**

- 5 Variant calling/Population genomics
- 4 De novo genome assembly/analyses
- 3 Gene expression
- 2 RNAseq method development
- 1 RNAseq + proteomics
- 1 Metagenomics
- 1 Non-coding RNA
- 17**

Human

Animals/plants

Animals/plants

Animals/Plants

Universal

Fungus

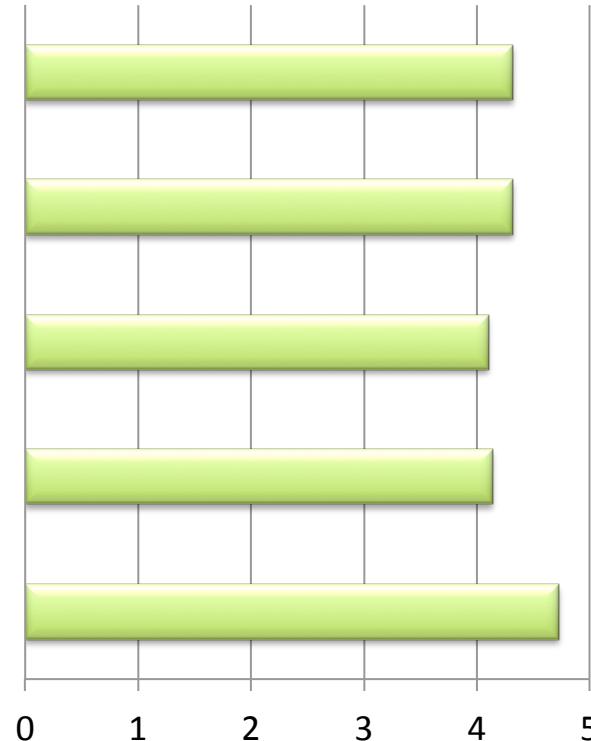
Human microbiota

Bacteria

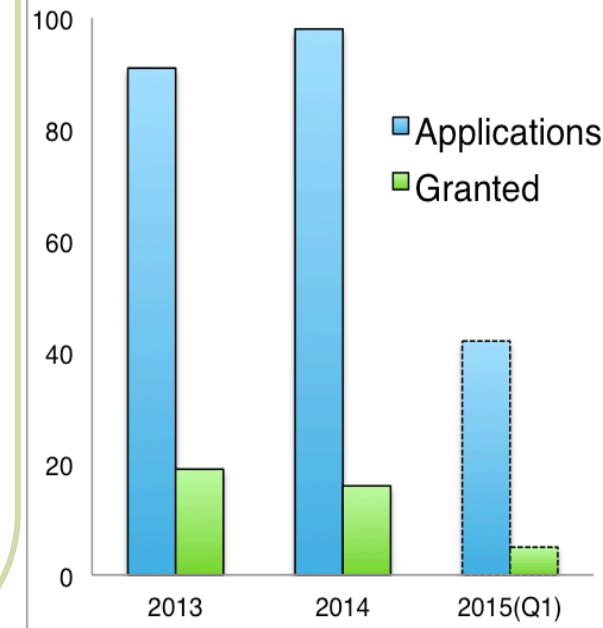
Happy users, high demand

User evaluation April 2015

Overall rating
Technical quality
Scientific impact
Long-term value (2-3 years)
In favour of SciLifeLab continuing
to offer this type of national
support



Bioinformatics
Long-term Support



Scientific highlights 2014/2015

Methods

- Boekel J, Chilton JM, Cooke IR, Horvatovich PL, Jagtap PD, Kall L, Lehtio J, Lukasse P, Moerland PD, Griffin TJ (2015) Multi-omic data analysis using Galaxy. *Nat. Biotechnol.*, 33:137–139.
- Branca RM, Orre LM, Johansson HJ, Granholm V, Huss M, Perez-Bercoff A., Forshed J, Kall L, Lehtio J (2014) HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods*, 11, 59–62.
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C (2014) Binning metagenomic contigs by coverage and composition. *Nat. Methods*, 11:1144–1146.

Evolution and biodiversity

- Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JB (2014) The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, 344:1410-1414
- Martinez Barrio A*, Lamichhaney S*, Fan G*, Pettersson M, Rafati N, Zhang H, Dainat J, Ekman D, Höppner D, Jern P, Martin M, Nystedt B, Liu X, Chen W, Liang X, Shi C, Fu Y, Ma K, Zhan X, Feng C, Gustafson U, Rubin C-J, Blass M, Casini M, Folkvord A, Laikre L, Ryman N, Ming-Yuen Lee S, Xu X, Andersson L (2015) The Atlantic herring genome – the genetic architecture underlying ecological adaptation (*Nature, in review*)
- Steige KA, Reimegård J, Koenig D, Scofield DG, Slotte T (2015) Analysis of allele-specific expression reveals *cis*-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella* (2015) (*Mol. Biol. Evo.*, submitted)

Medical research

- Alkasalias T, Flaberg E, Kashuba V, Alexeyenko A, Pavlova T, Savchenko A, Szekely L, Klein G, Guven H (2014) Inhibition of tumor cell proliferation and motility by fibroblasts is both contact and soluble factor dependent. *Proc. Natl. Acad. Sci. U.S.A.*, 111:17188–17193.
- Lindqvist CM, Nordlund J, Ekman D, Johansson A, Moghadam BT, Raine A, Overnäs E, Dahlberg J, Wahlberg P, Henriksson N, Abrahamsson J, Frost BM, Grandér D, Heyman M, Larsson R, Palle J, Söderhäll S, Forestier E, Lönnherholm G, Syvänen AC, Berglund EC (2015) The Mutational Landscape in Pediatric Acute Lymphoblastic Leukemia Deciphered by Whole Genome Sequencing. *Human Mutation* 36:118–128
- Einarsdottir B, Bagge RO, Bhadury J, Jespersen H, Mattsson J, Nilsson L, Truvé K, López MD, Naredi P, Nilsson O, Stierner U, Ny L, Nilsson J (2014). Melanoma patient-derived xenografts accurately model the disease and develop fast enough to guide treatment decisions. *Oncotarget*, 5:9609–9618.
- Lindholm ME, Huss M, Solnestam BW, Kjellqvist S, Lundeberg J, Sundberg CJ (2014) The human skeletal muscle transcriptome: sex differences, alternative splicing, and tissue homogeneity assessed with RNA sequencing. *FASEB J.* 28:4571-4581
- Browall S, Norman M, Tångrot J, Galanis I, Sjöström K, Dagerhamn J, Hellberg C, Pathak A, Spadafina T, Sandgren A, Bättig P, Franzén O, Andersson B, Örtqvist Å, Normark S, Henriques-Normark B (2014) Intralclonal Variations Among Streptococcus pneumoniae Isolates Influence the Likelihood of Invasive Disease in Children. *Journal of Infectious Diseases*, 209: 377–388.

SciLifeLab Bioinformatics Courses

Course	Date	Participants	Evaluation score (max 5)
Introduction to bioinformatics using NGS data	April 2013	24	4.6
	Nov 2013	24	4.3
	March 2014	24	4.5
	April 2014	24	3.8
	Sept 2014	24	4.1
	Nov 2014	24	4.3
Perl programming for biological sciences	May 2013	20	4.4
	Oct 2014	20	4.4
Genome Assembly	Nov 2013	20	4.1
	Nov 2014	20	4.4
Human Genetic Variation	June 2013	15	4.5
	Sept 2013	20	3.9
RNAseq	June 2013	15	4.1
	Sept 2013	20	4.2
	Oct 2014	20	4.3
RNAseq and proteomics	June 2014	20	4.1
Metagenomics	Nov 2014	20	4.2
TOTAL 2013 + 2014		394	4.3

www.scilifelab.se/education/courses/

We're here for you!

