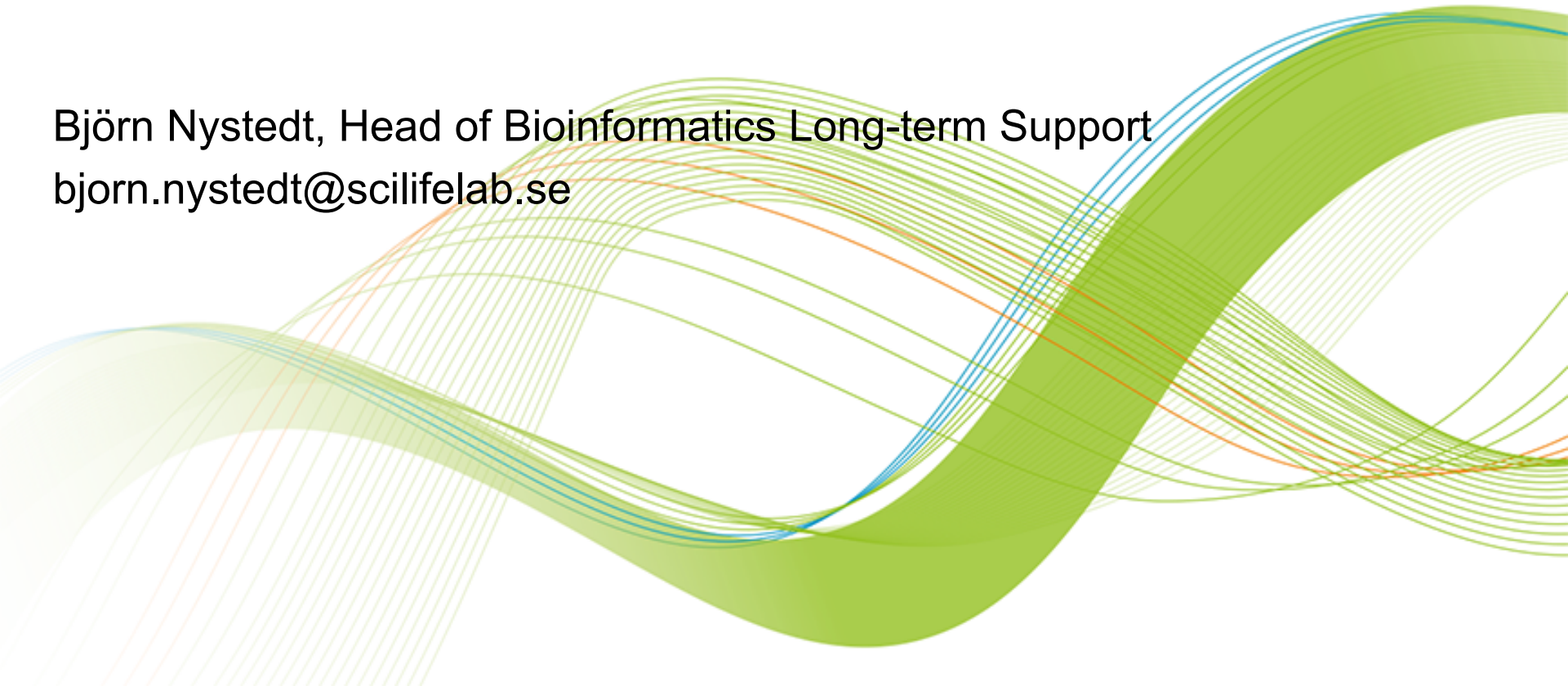
SciLifeLab

The National Bioinformatics Infrastructure Sweden (NBIS)

www.scilifelab.se/platforms/bioinformatics/

Björn Nystedt, Head of Bioinformatics Long-term Support
bjorn.nystedt@scilifelab.se



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 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 >70 FTE for custom-tailored project support, methods and systems development, data publishing, training
- **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 (LCBKI)
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

SciLifeLab platforms

SciLifeLab national service

National Genomics Infrastructure

Clinical Diagnostics

Functional Genomics

National Bioinformatics Infrastructure Sweden

Bengt Persson

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

VR

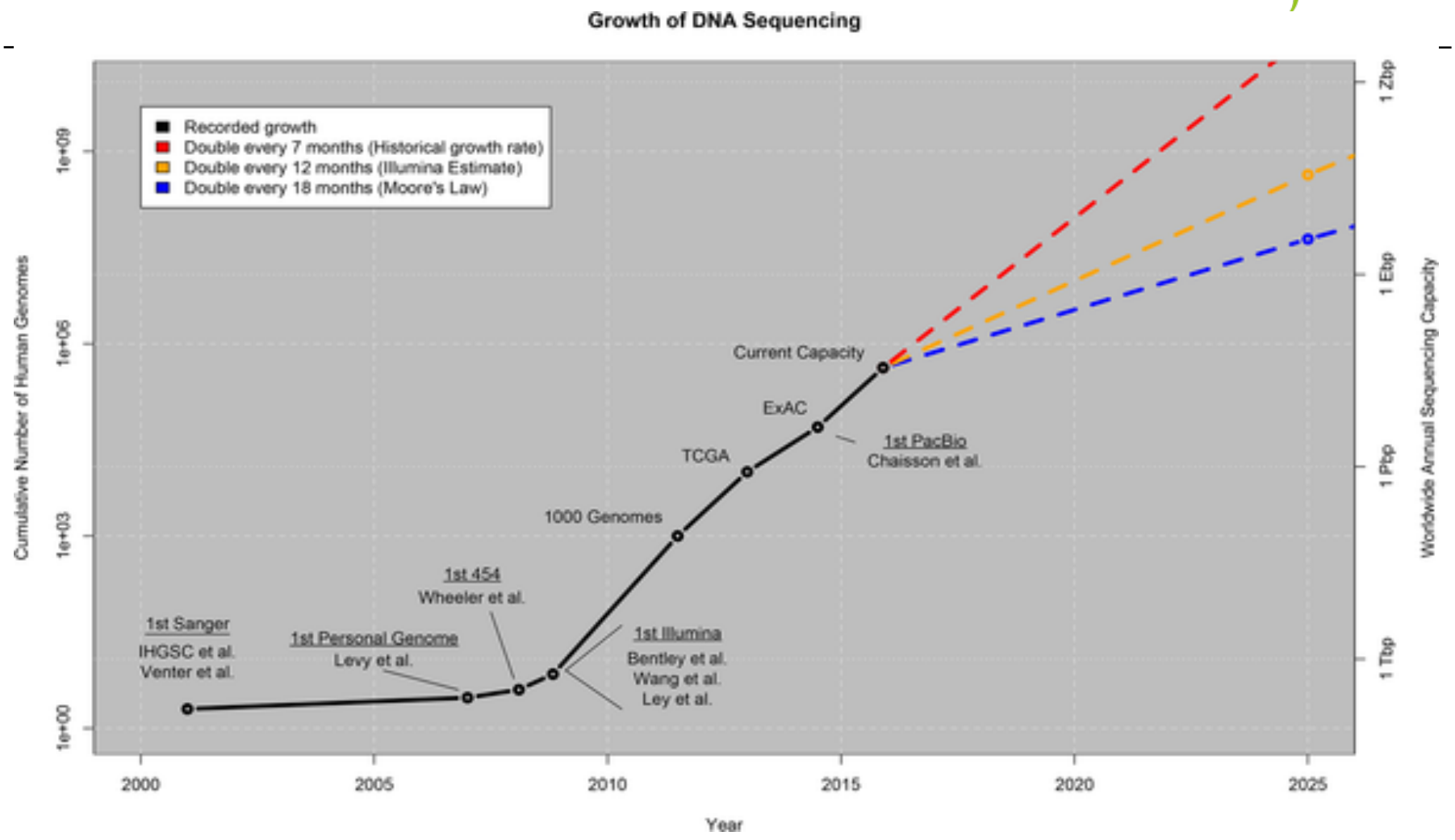
SNIC

Computer resources free for Swedish researchers

SciLifeLab

Why do we invest in a bioinformatics infrastructure?

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.

SciLifeLab

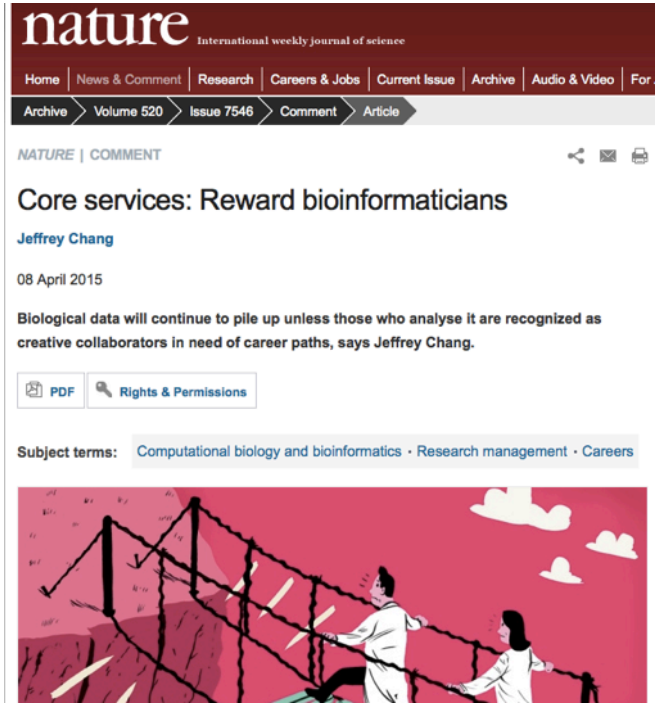
Data Phase	Astronomy	Twitter	YouTube	Genomics
Acquisition	25 zetta-bytes/year	0.5–15 billion tweets/year	500–900 million hours/year	1 zetta-bases/year
Storage	1 EB/year	1–17 PB/year	1–2 EB/year	2–40 EB/year
Analysis	In situ data reduction	Topic and sentiment mining	Limited requirements	Heterogeneous 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

PLOS | BIOLOGY

Bioinformatics know-how as infrastructure



"The scientific community has failed to craft attractive career paths for those who do the analyses it increasingly requires. Institutions and funding bodies must carve out a viable place for bioinformaticians who focus on collaborations, and reward them for their abilities to navigate the myriad demands of multidisciplinary projects."

http://www.nature.com/news/core-services-reward-bioinformaticians-1.17251

HEALTH AND AGING

ENVIRONMENT AND ECOLOGY

EVOLUTION AND ORIGIN OF LIFE

Support, tools and training

SciLifeLab

elixir

ISBE Infrastructure for Systems Biology Europe

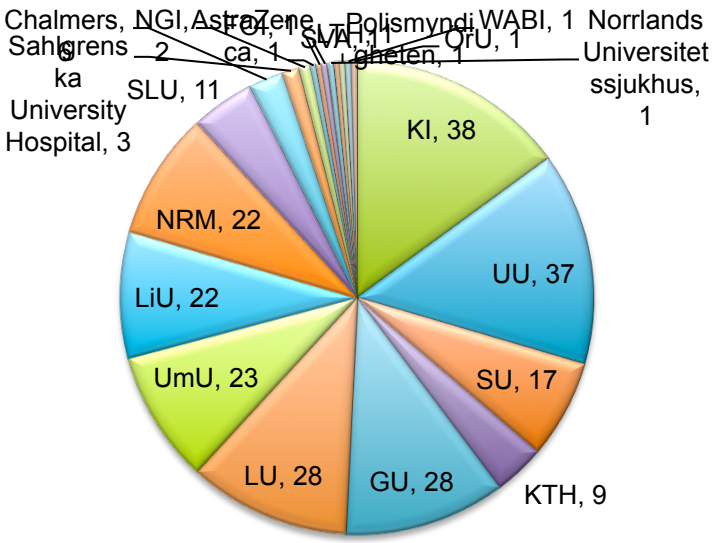

Support

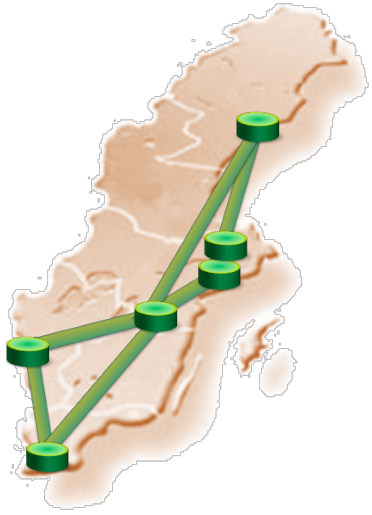
Custom-tailored support

New contact routes later 2016, stay tuned at www.scilifelab.se/platforms/bioinformatics/

- Study design consultation (free)
  support@bils.se
  + drop-in sessions every week @ all 6 sites
- Short-term support (≤40h, free)
  http://bils.se/resources/supportform/index.php
- Medium-term support (+40h, user fee)
  http://bils.se/resources/supportform/index.php
- Long-term support (500h, free, scientific evaluation)
  http://www.scilifelab.se/facilities/wabi/

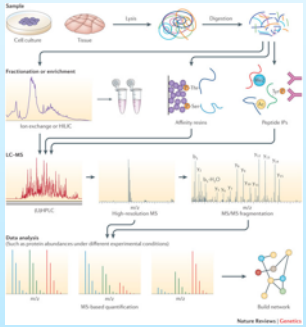
Next deadline for applications Feb 12!





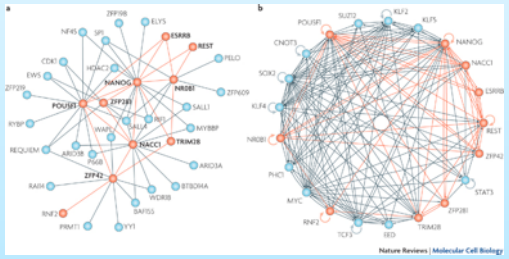
Potential increase in user fees later 2016 due to general infrastructure cut-down by VR

Short-term support



Proteomics

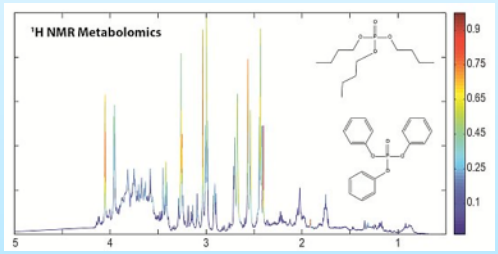
Genomics



Systems biology

400 projects/year!

Biostatistics



Metabolomics

Support decisions every 2nd week

Bioinformatics Long-term Support

Wallenberg Advanced Bioinformatics Infrastructure

www.scilifelab.se/facilities/wabi/

70% of funding

Tailored solutions – high impact

Applied bioinformatics: 500h free support/project

- Variant analyses
- Transcriptomics
- Single-cell analyses
- Epigenetics
- Metagenomics

Sweden's strongest unit for analyses of large-scale genomic data (24 FTE)

National committee reviews and selects projects based on scientific quality

Basic science!

Directors

Siv Andersson

Gunnar von Heijne

Managers

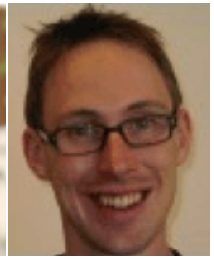
Björn Nystedt

Thomas Svensson

Bioinformatics Long-term Support


Johan Reimegård


Åsa Björklund


Pär Engström


Jakub Orzechowski Westholm

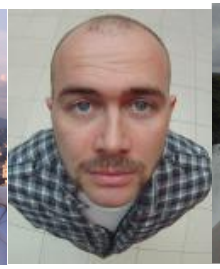

Estelle Proux-Wéra


Markus Ringnér


Sanela Kjellqvist


Björn Nystedt


Diana Ekman


Páll Ólason


Anna Johansson


Marcel Martin


Malin Larsson


Allison Churcher


Alvaro Martinez Barrio

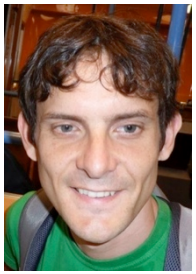

Per Unneberg


Leif Väremo


Nikolay Oskolkov


Lena Hansson

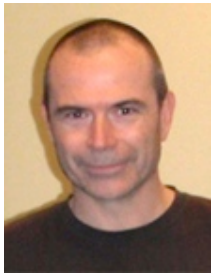

Sergiu Netotea


Mikael Huss


Bengt Sennblad


Rasmus Ågren


Stefania Giacomello


Thomas Svensson

Application procedure

- Open to all research groups in Sweden
- Applications 3 times every year (accept 5-10 projects per call)
- Requires hands-on involvement from the research group

National committee

- 500h effective time over ~6-18 calendar months
- Co-authors according to normal contribution criteria
- Staff 100% support (not driving own research)

www.scilifelab.se/facilities/wabi/

New!

Opening for a few projects in integrative omics as of Feb 12

Custom-tailored support
“Routinely unique”

Difficult to forsee/automate

Human health and disease (13)
5 Variant analyses (cohort, family, cell fate)
3 Epigenetics
2 RNA, method
1 Differential gene expression
1 Lipidomics
1 Integrative

(Medical) animal models (10)
4 single-cell RNA
2 Differential gene expression
2 Targeted
1 ChipSeq
1 miRNA

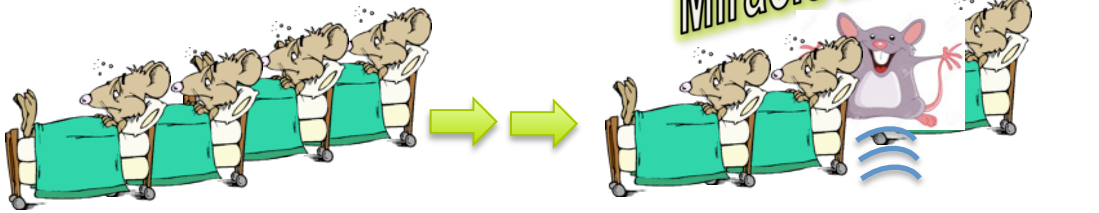
Ecology/Evolution (8)
3 Population genomics
2 De novo genome assembly/analyses
2 Phylogenomics/genome evolution
1 Epigenetics

Miracle mutation in rat model

SciLifeLab

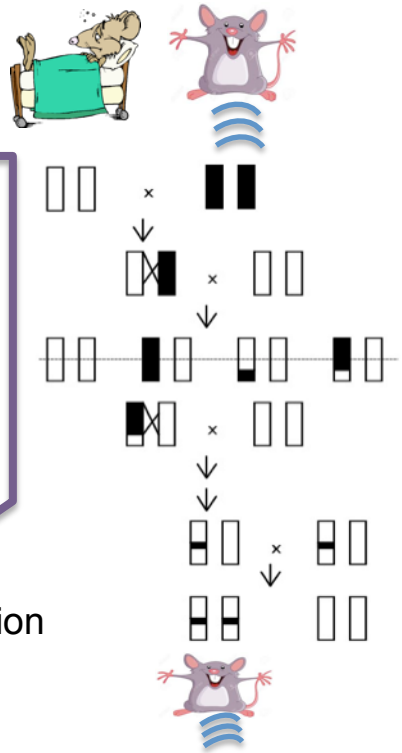
Diana Ekman

Disease model



Old and slow

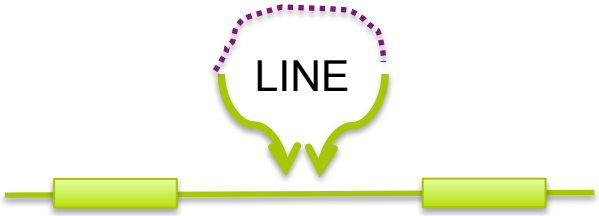
1Mb target region



New and fast

Global DNA and RNA sequencing

- 1 differentially expressed gene in region. But no SNPs.
- Manual inspection and local assembly of genomic reads.



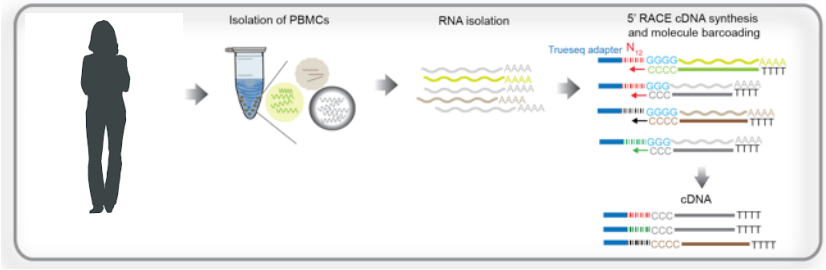
Complete protection by intronic LINE in unknown gene!

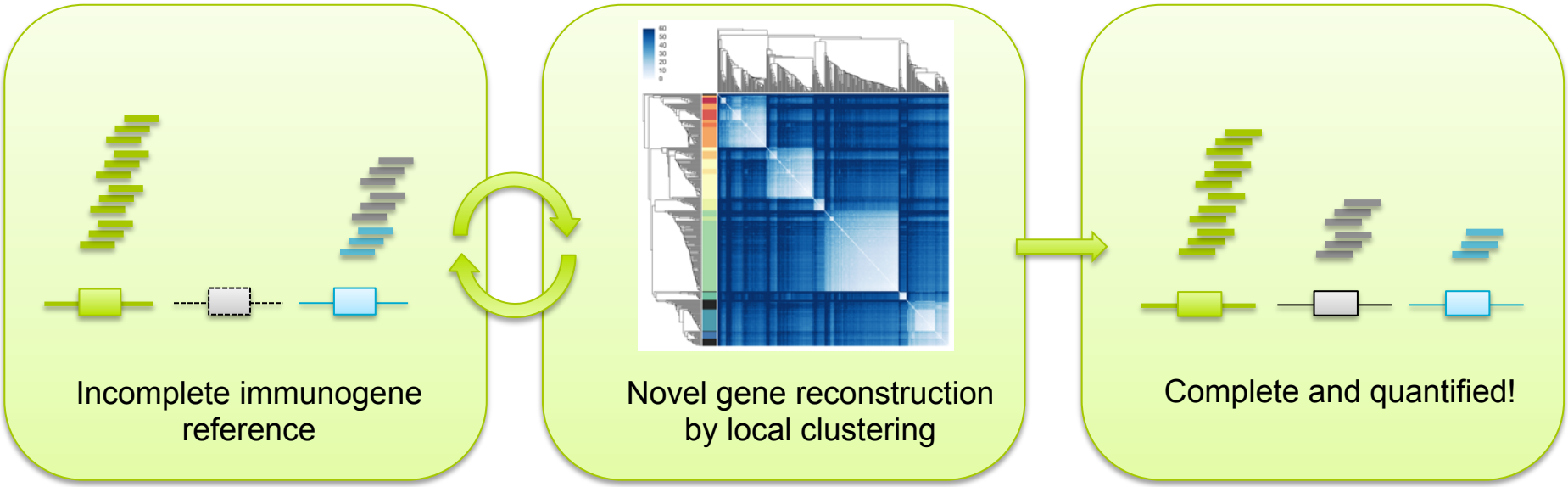
Ulrika Norin
Medical inflammation research

IgY-Pipe: Immunorepertoire profiling

SciLifeLab

Marcel Martin





IgY-Pipe

- Automatic V/D/J gene profiling
- Novel gene discovery works extremely well!
- Single-read tracing
- Any species (any region)
- Open Source release end of January 2016

Gunilla Karlsson Hedestam
Infection immunology

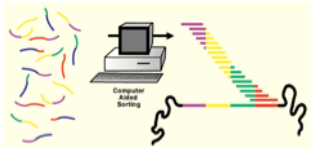
Speciation in action


Henrik Lantz

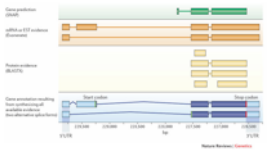

Per Unneberg

SciLifeLab

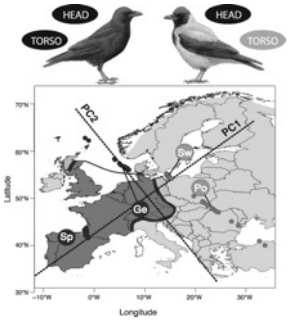
Genome assembly and annotation



1.2 Gbp

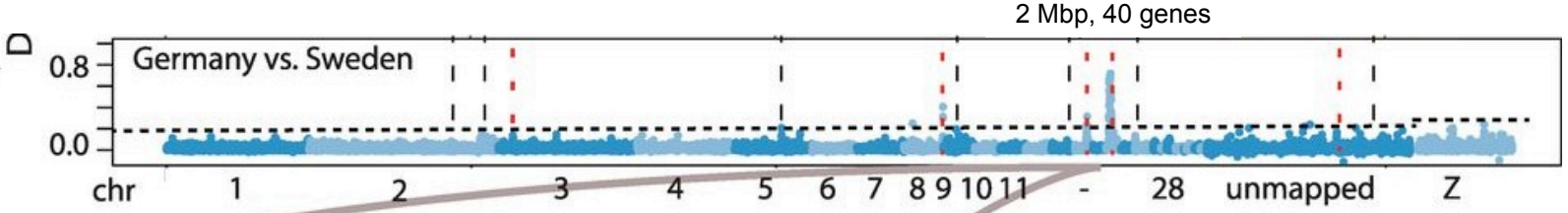

21,000 genes

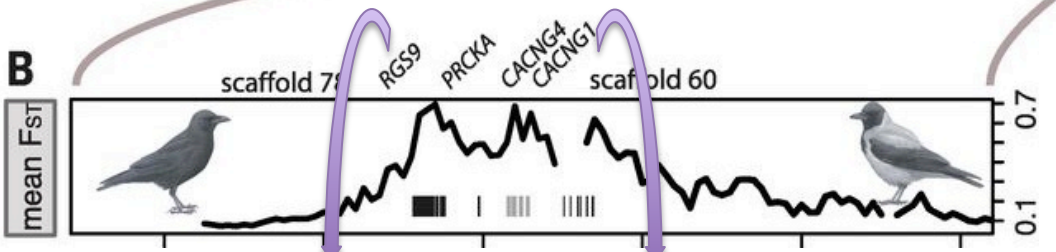
WGS re-sequencing



2+2 populations per species
60 individuals, 12X

Population contrasts





"Mating preferences and sexual selection alone can cause phenotypic and genotypic differentiation"

Affects visual perception

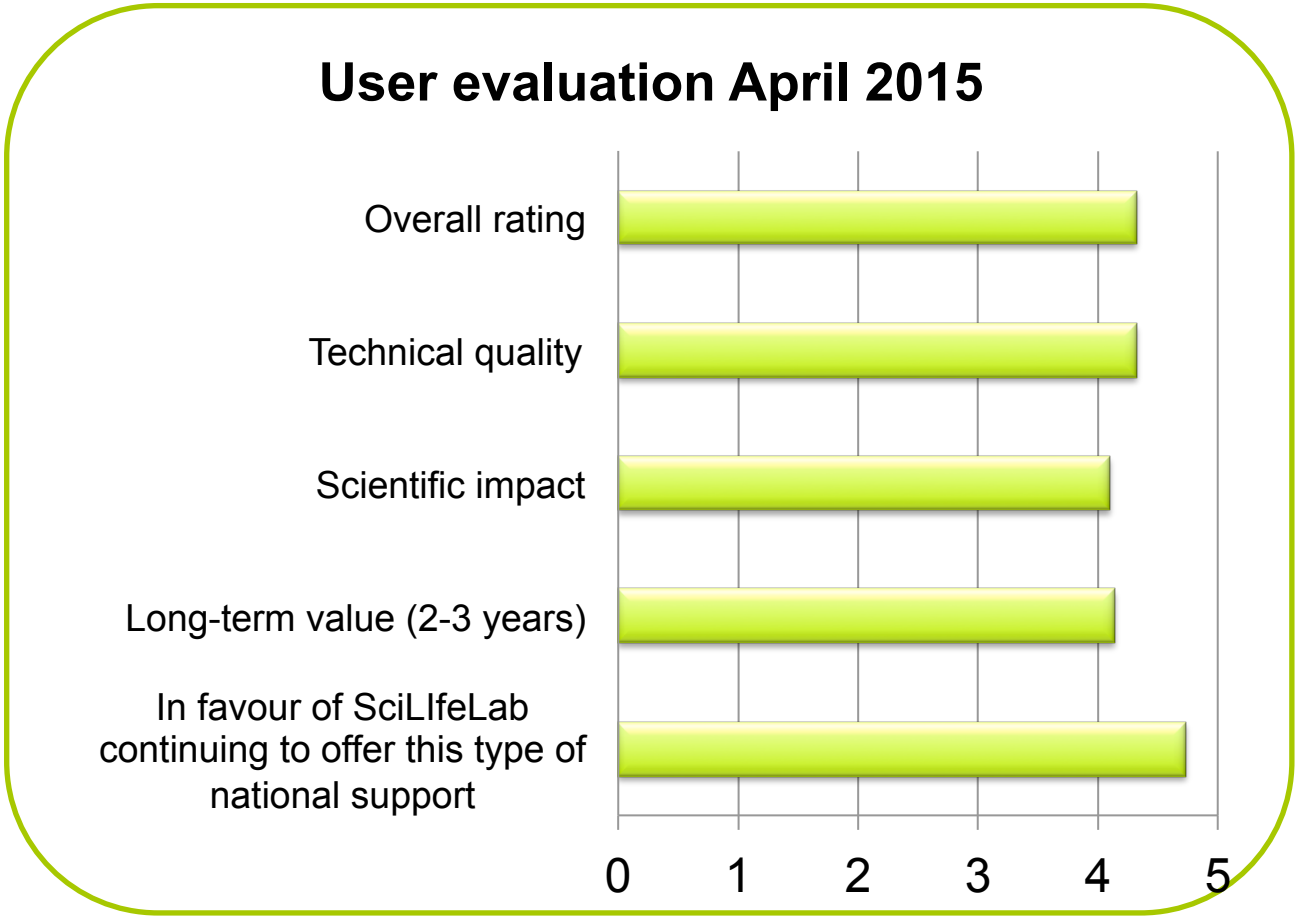
Transcription factor *MITF*

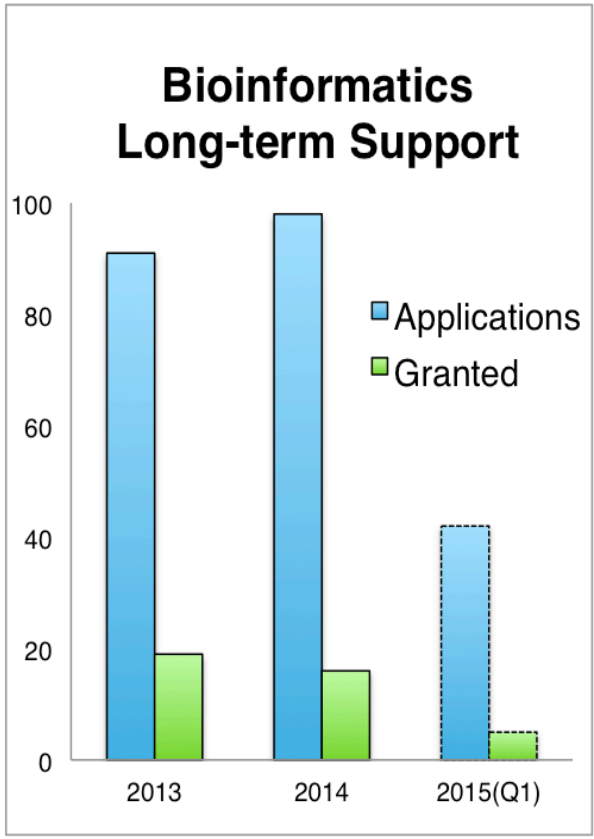
Melanogenesis pathway

Jochen Wolf
Evolutoinary biology

Poelstra et al. (2014)
Science 344:1410-1414

Happy users, high demand






Tools

Tools and resources in progress

SciLifeLab

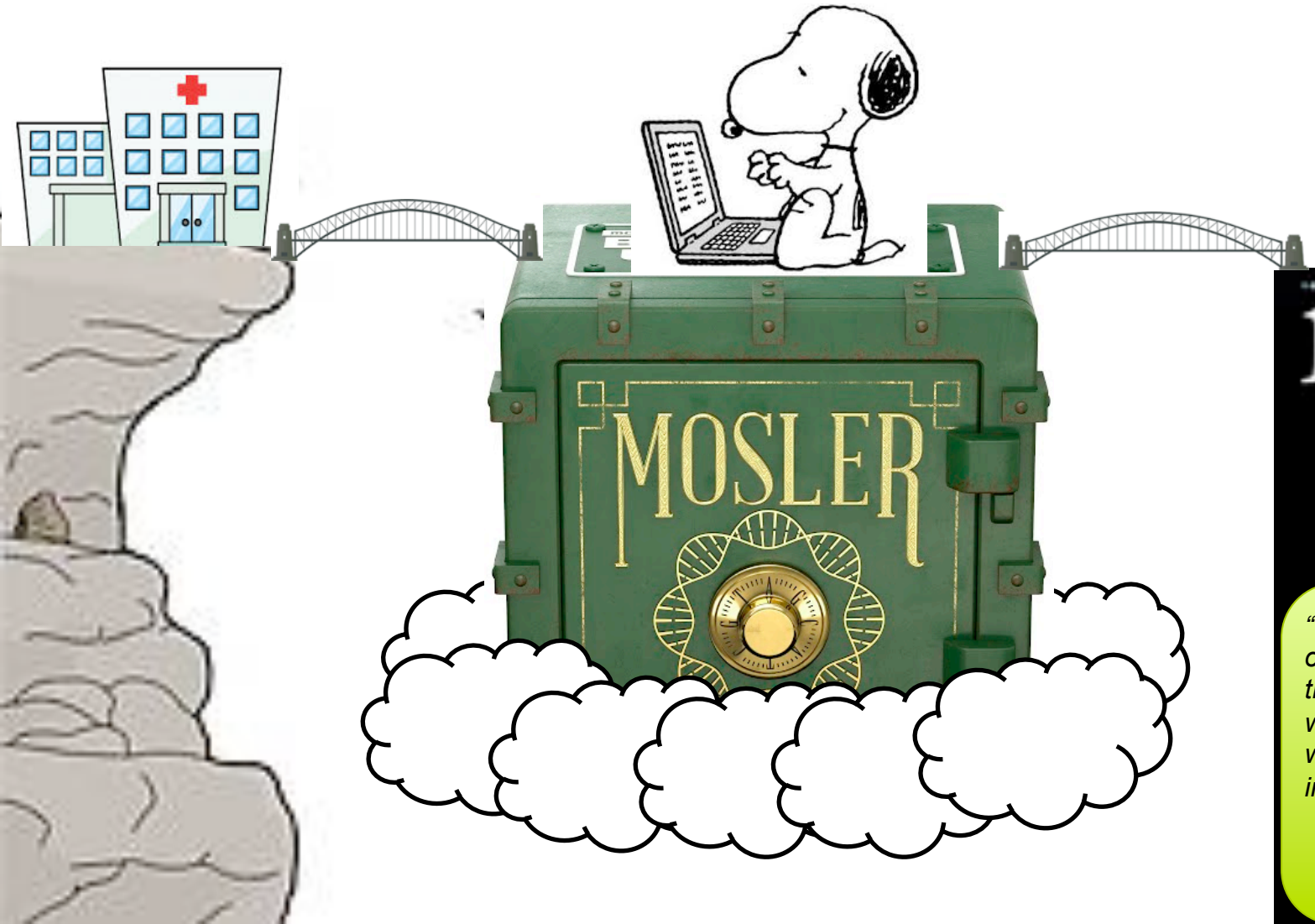
- Immuno gene repertiore profiling
- hg38-compatible GATK
- Haloplex variant calling pipeline
- ChIP-Seq pipeline
- Genomic phasing tool (long reads)
- Single-cell transcriptomics QC pipeline

- Snakemake workflow management system
- WGS structural variation pipeline
- WGS somatic variant calling pipeline
- …

High performance computing for sensitive personal data

SciLifeLab
Jonas Hagberg

From Personal Data Act to Publication



"We have had a sense of full security in using the Mosler system when doing research with sensitive personal information"

Tove Fall
Epidemiology

Genome assembly and annotation

SciLifeLab

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!


Training

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
			4.4
			4.7
	Oct 2014	20	4.5
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**

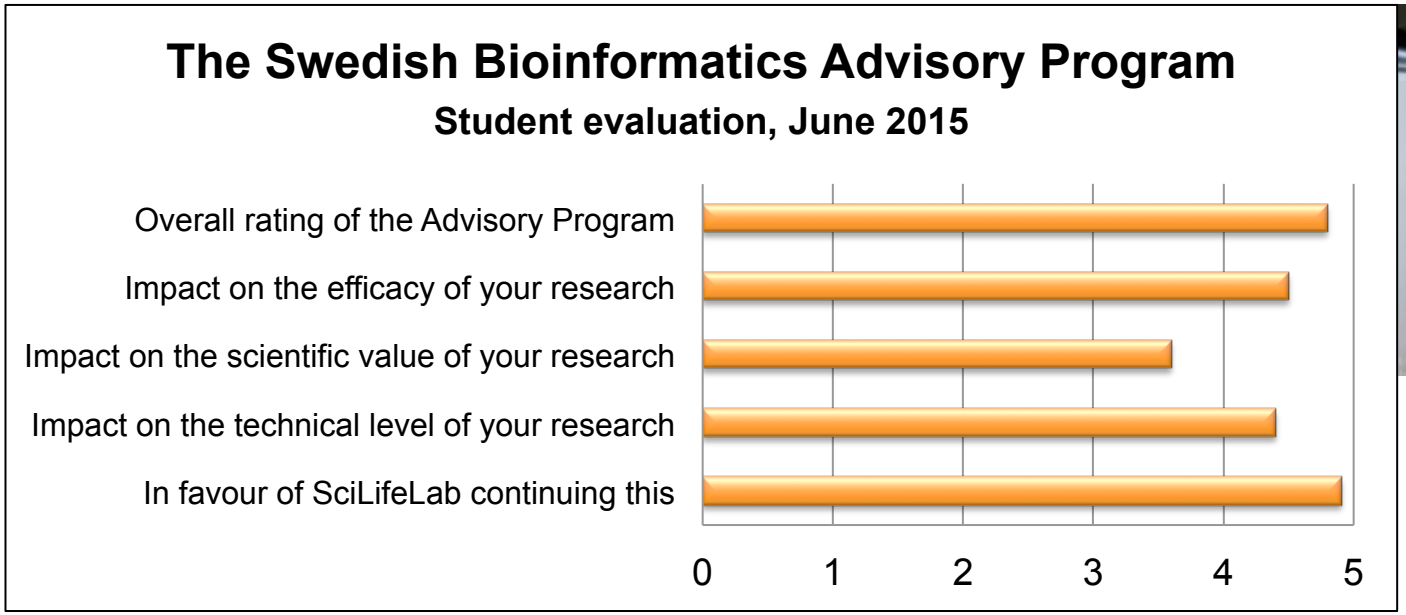
www.scilifelab.se/education/courses/

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/

Currently 27 PhD student enrolled

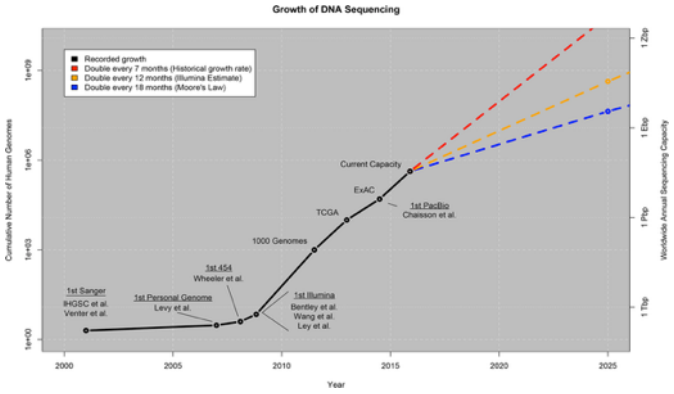


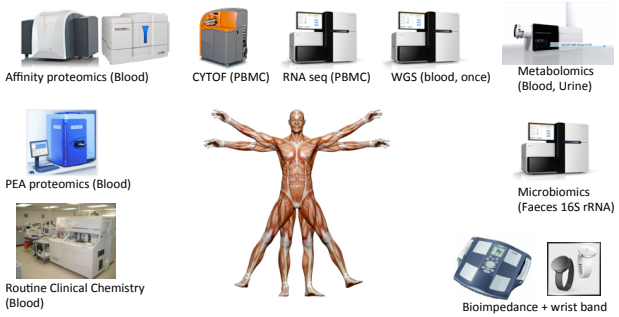
Looking ahead

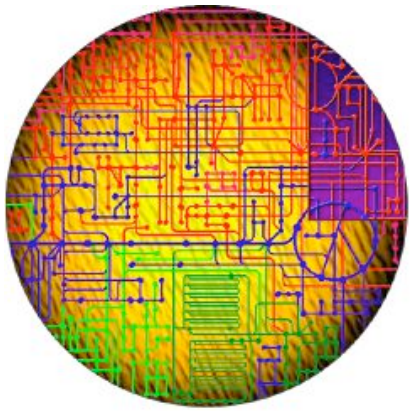
The future is bright ☺

..and integrated!
..and unbalanced!

SciLifeLab







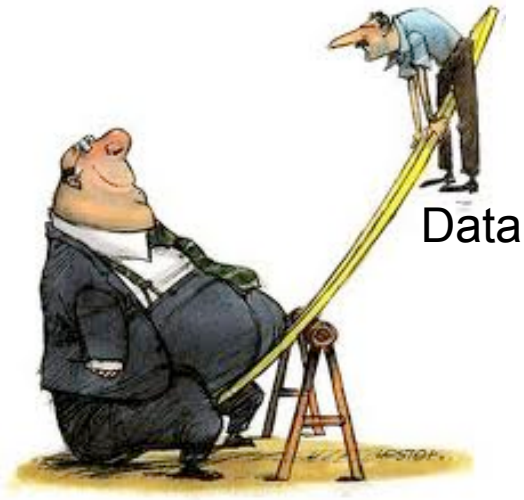
Volume

Integration

Systems/processes

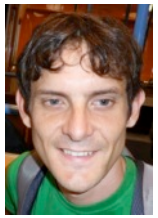
Strategic positioning

- Tools development
- Data management
- Integrative omics
- Systems Biology
- Medical genomics



Data scientists

Data



Mikael Huss
BigData/Integrative bioinformatics

We’re here for you!

SciLifeLab

Acknowledgements

Stockholm University
Uppsala University
Karolinska Institutet
The Royal Institute of Technology
Chalmers University of Technology
The University of Gothenburg
Linköpings University
Lund University
Umeå University
The Swedish Agricultural University