

# Sarek and nf-core for Reproducibility

## SPHN Workflow Interoperability Workshop

SciLifeLab



- 👤 Maxime U. Garcia
- 🐦 @gau
- ⌚ @MaxUlysse
- 🌐 maxulyssse.github.io

Science for Life Laboratory



# SciLifeLab

 <https://scilifelab.se/>

SciLifeLab is a national centre for molecular biosciences  
with focus on health and environmental research



# SciLifeLab

 <https://scilifelab.se/>

## Infrastructure Services

Genomics

Proteomics

Metabolomics

Single Cell Biology

Bioimaging and Molecular  
Structure

Chemical Biology and Genome  
Engineering

Drug Discovery

Diagnostics

Bioinformatics



🌐 <https://ngisweden.scilifelab.se/>

Our mission is to offer a **state-of-the-art infrastructure** for massively parallel DNA sequencing and SNP genotyping, available to researchers all over Sweden.

We provide **guidelines and support** for sample collection, study design, protocol selection and bioinformatics analysis.

# National Genomics Infrastructure



🌐 <https://ngisweden.scilifelab.se/>





🌐 <https://ngisweden.scilifelab.se/>

- Stockholm
- Uppsala

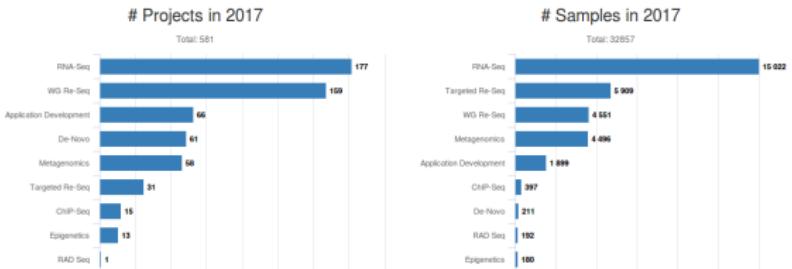


🌐 <https://ngisweden.scilifelab.se/>

- Genomics Production - Genomics Applications Development
- SNP&Seq - Uppsala Genome Center



🌐 <https://ngisweden.scilifelab.se/>



# National Bioinformatics Infrastructure Sweden



🌐 <https://www.nbis.se/>



- Distributed infrastructure
- Swedish node in ELIXIR

# Sarek



James Frain as Sarek, "Star Trek: Discovery"

# What is Sarek?

---



Sarek

🌐 <http://opensource.scilifelab.se/projects/sarek/>

# What is Sarek?

---



🌐 <http://opensource.scilifelab.se/projects/sarek/>

- Nextflow pipeline

# What is Sarek?



🌐 <http://opensource.scilifelab.se/projects/sarek/>

- Nextflow pipeline
- Developed at NGI



# What is Sarek?



🌐 <http://opensource.scilifelab.se/projects/sarek/>

- Nextflow pipeline
- Developed at NGI
- In collaboration with NBIS



# What is Sarek?



🌐 <http://opensource.scilifelab.se/projects/sarek/>

- Nextflow pipeline
- Developed at NGI
- In collaboration with NBIS
- Support of The Swedish Pediatric Tumor Biobank



Sarek is now in two flavors

---



Sarek

# Sarek is now in two flavors

---



Sarek

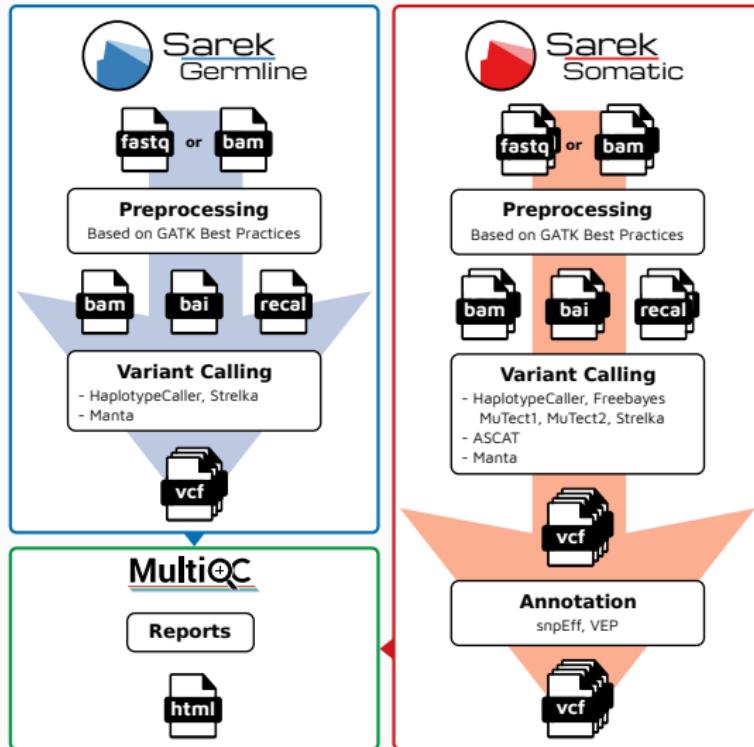


Sarek  
Germline



Sarek  
Somatic

# What does Sarek do?



# What does Sarek do?

---



- WGS analysis (Tumor/Normal pair or Germline)
  - FASTQ(s) to annotated VCF(s)
  - Multiple entry points

# What does Sarek do?

---



- WGS analysis (Tumor/Normal pair or Germline)
  - FASTQ(s) to annotated VCF(s)
  - Multiple entry points
- Handles both GRCh37, GRCh38 and custom genomes

# What does Sarek do?

---



- WGS analysis (Tumor/Normal pair or Germline)
  - FASTQ(s) to annotated VCF(s)
  - Multiple entry points
- Handles both GRCh37, GRCh38 and custom genomes
- Uses containers (Reproducibility, Portability)
  - Docker or Singularity

## Where do I use Sarek?

---

- My own Computer

## Where do I use Sarek?

---

- My own Computer
- UPPMAX - Irma - NGI Production Cluster
- UPPMAX - Rackham - Cluster
- UPPMAX - Bianca - Secure Cluster

## Where do I use Sarek?

---

- My own Computer
- UPPMAX - Irma - NGI Production Cluster
- UPPMAX - Rackham - Cluster
- UPPMAX - Bianca - Secure Cluster
- Travis CI

## Where do I use Sarek?

---

- My own Computer
- UPPMAX - Irma - NGI Production Cluster
- UPPMAX - Rackham - Cluster
- UPPMAX - Bianca - Secure Cluster
- Travis CI
- AWS

## Where can Sarek be used?

---

- Any POSIX compatible system

# Interoperability

---

- Reproducibility

# Interoperability

---

- Reproducibility
- Portability

# Interoperability

---

- Easy Reproducibility
- Easy Portability

# NGI pipelines



**NGI**-RNAseq



**Sarek**



**NGI**-MethylSeq



**NGI**-smRNASeq



**NGI**-ChIPseq



**NGI**-ExoSeq



**NGI**-RNAfusion



**NGI**-NeutronStar

SciLifeLab

NGI stockholm

NouGAT (*de-novo*)

# Who is using NGI pipelines?

## Participating Institutes

NGI-RNAseq is now used by a number of core sequencing and bioinformatics facilities. Some of these are listed below. If you use this pipeline too, please let us know in an issue and we will add you to the list.

|   |   |   |
|---|---|---|
|  | National Genomics Infrastructure (NGI),<br>Sweden | <a href="https://ngisweden.scilifelab.se/">https://ngisweden.scilifelab.se/</a>                         |
|  | Quantitative Biology Center (QBiC), Germany       | <a href="https://portal.qbic.uni-tuebingen.de/portal/">https://portal.qbic.uni-tuebingen.de/portal/</a> |

# NGI-RNASeq

 <https://github.com/SciLifeLab/NGI-RNAseq>

Vladimir Kiselev  
@wikiselev

Following

Just had a error-free run of [@tallphil](#)'s RNA-Seq pipeline using [@PaoloDiTomaso](#)'s [@nextflowio](#) on our LSF cluster! Started yesterday, came back today, all done! Thanks a lot guys for your great work!

The screenshot shows a Nextflow workflow report. At the top, it says "flow workflow report" and "allen". Below that, a green bar indicates "Run completed successfully". The main area displays a chart titled "% Requested CPU Used" with several colored bars representing different processes. Below the chart, a process graph shows nodes connected by arrows, with labels like "SampleSheet", "FastQC", "Trim", "Map", "GATK", "VCF", "LocusZoom", and "LocusZoom". A legend at the bottom identifies the colors for each process: blue for SampleSheet, orange for FastQC, red for Trim, purple for Map, green for GATK, yellow for VCF, and pink for LocusZoom.

11:18 AM - 22 Feb 2018

4 Retweets 24 Likes

4 24



🌐 <https://nf-core.github.io/>

A community effort to collect a curated set of Nextflow analysis pipelines

- GitHub organisation to collect pipelines in one place
- No institute-specific branding
- Strict set of guideline requirements
- Automated testing for code style and function



🌐 <https://nf-core.github.io/>

All Nextflow pipelines must have:

- MIT licence
- Software bundled using Containers
- Continuous Integration testing
- Stable release tags
- Common pipeline structure and usage
- Excellent documentation

# NGI pipelines



**NGI**-RNAseq



**Sarek**



**NGI**-MethylSeq



**NGI**-smRNASeq



**NGI**-ChIPseq



**NGI**-ExoSeq



**NGI**-RNAfusion



**NGI**-NeutronStar

SciLifeLab

NGI stockholm

NouGAT (*de-novo*)

## A first pipeline transferred



Phil Ewels @ewels

We have a pipeline!

# **nf-core/ methylseq**



<https://github.com/nf-core/methylseq>

# The List of People Involved

---



|                     |                        |
|---------------------|------------------------|
| Sebastian DiLorenzo | Markus Mayrhofer       |
| Jesper Eisfeldt     | Monica Nistér          |
| Phil Ewels          | Björn Nystedt          |
| Maxime Garcia       | Pall Olason            |
| Szilveszter Juhos   | Markus Ringnér         |
| Max Käller          | Pelin Sahlén           |
| Malin Larsson       | Johanna Sandgren       |
| Marcel Martin       | Teresita Díaz De Ståhl |

# The List of People Involved

---



- The National Genomics Infrastructure (NGI), SciLifeLab, Stockholm, Sweden
- The Quantitative Biology Center (QBiC), Universität Tübingen, Germany
- The Genomics Institute of Singapore (GIS), A\*STAR, Singapore

# The List of People Involved

---



|                  |                            |
|------------------|----------------------------|
| Chih Chuan       | Denis Moreno               |
| Paolo Di Tommaso | Remi-Andre Olsen           |
| Phil Ewels       | Senthilkumar Panneerselvam |
| Sven Filinger    | Alexander Peltzer          |
| Maxime Garcia    | Chuan Wang                 |
| Rickard Hammarén | Andreas Wilm               |

## Get involved!

- We have gitter channels
  -  <https://gitter.im/SciLifeLab/Sarek>
  -  <https://gitter.im/nf-core/Lobby>

## Get involved!

- We have gitter channels
  -  <https://gitter.im/SciLifeLab/Sarek>
  -  <https://gitter.im/nf-core/Lobby>
- Our code is hosted on Github
  -  <https://github.com/SciLifeLab/Sarek>
  -  <https://github.com/nf-core>

# Any questions?

