Sarek

SciLifeLab

Karolinska Institutet

GANATIONALCTAC
ATCAGENOMICSGT
INFRASTRUCTURE

NBIS

The Review

Maxime U. Garcia
maxulysse.github.io
@MaxUlysse
@gau

Barntumörbanken

What is Sarek?

http://sarek.scilifelab.se/

- Analysis germline and somatic workflow

What is Sarek?

http://sarek.scilifelab.se/

- Analysis germline and somatic workflow
- Whole genome or targeted sequencing

What is Sarek?

http://sarek.scilifelab.se/

- Analysis germline and somatic workflow
- Whole genome or targeted sequencing
- Developed with NGI and NBIS

NATIONAL GENOMICS INFRASTRUCTURE

NBIS

What is Sarek?

http://sarek.scilifelab.se/

- Analysis germline and somatic workflow
- Whole genome or targeted sequencing
- Developed with NGI and NBIS
- Support from The Swedish Childhood Tumor Biobank

Using

https://www.nextflow.io/

https://www.sylabs.io/singularity/

Using

https://www.nextflow.io/

Data-driven workflow language

https://www.sylabs.io/singularity/

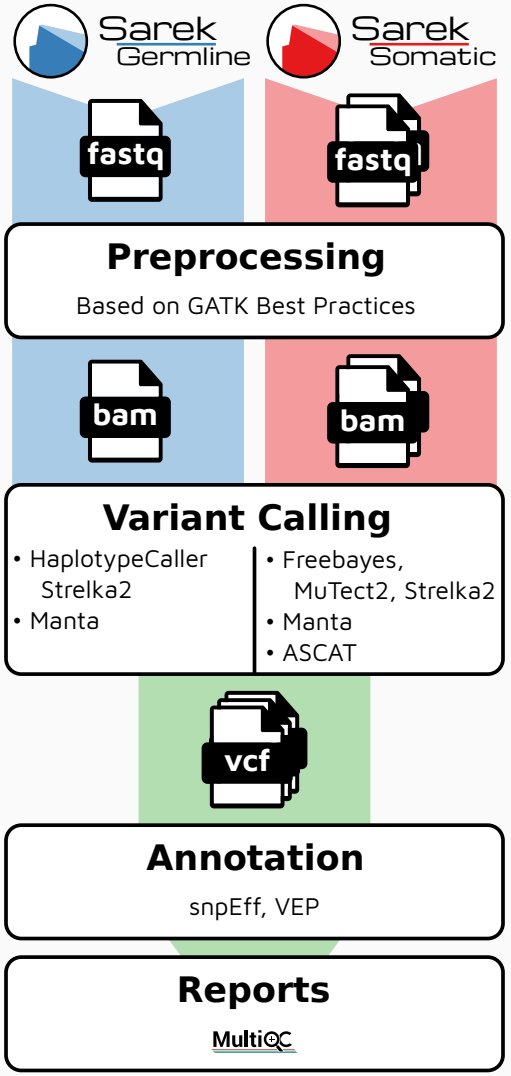
HPC specific container engine

Sarek exists in multiple flavors

Sarek exists in multiple flavors

Sarek exists in multiple flavors

Data and files workflow



AWS iGenomes

https://ewels.github.io/AWS-iGenomes/

- Human Grch37
- Human Grch38

AWS iGenomes

https://ewels.github.io/AWS-iGenomes/

- Human Grch37
- Human Grch38
- Non-Human

Preprocessing

https://software.broadinstitute.org/gatk/best-practices/

Based on GATK Best Practices (GATK 4.0)

Preprocessing

https://software.broadinstitute.org/gatk/best-practices/

Based on GATK Best Practices (GATK 4.0)

- Reads mapped to reference genome with `bwa`

Preprocessing

https://software.broadinstitute.org/gatk/best-practices/

Based on GATK Best Practices (GATK 4.0)

- Reads mapped to reference genome with `bwa`
- Duplicates marked with `picard MarkDuplicates`

Preprocessing

https://software.broadinstitute.org/gatk/best-practices/

Based on GATK Best Practices (GATK 4.0)

- Reads mapped to reference genome with `bwa`
- Duplicates marked with `picard MarkDuplicates`
- Recalibrate with `GATK BaseRecalibrator`

Variant Calling

- SNVs and small indels:

Variant Calling

- SNVs and small indels:
  - Freebayes 
  - HaplotypeCaller 
  - MuTect2 
  - Strelka2 

Variant Calling

- SNVs and small indels:
  - Freebayes 
  - HaplotypeCaller 
  - MuTect2 
  - Strelka2 
- Structural variants:

Variant Calling

- SNVs and small indels:
  - Freebayes 
  - HaplotypeCaller 
  - MuTect2 
  - Strelka2 
- Structural variants:
  - Manta 
- Sample heterogeneity, ploidy and CNVs:

Variant Calling

- SNVs and small indels:
  - Freebayes 
  - HaplotypeCaller 
  - MuTect2 
  - Strelka2 
- Structural variants:
  - Manta 
- Sample heterogeneity, ploidy and CNVs:
  - ASCAT 
  - Control-FREEC 

Annotation

- VEP and SnpEff
- 🗄 ClinVar, COSMIC, dbSNP, GENCODE, gnomAD, polyphen, sift, etc.

Prioritization

- First step towards clinical use

Prioritization

- First step towards clinical use
- Rank scores are computed for all variants
  - COSMIC, ClinVar, SweFreq and MSK-IMPACT (cancerhotspots.org)

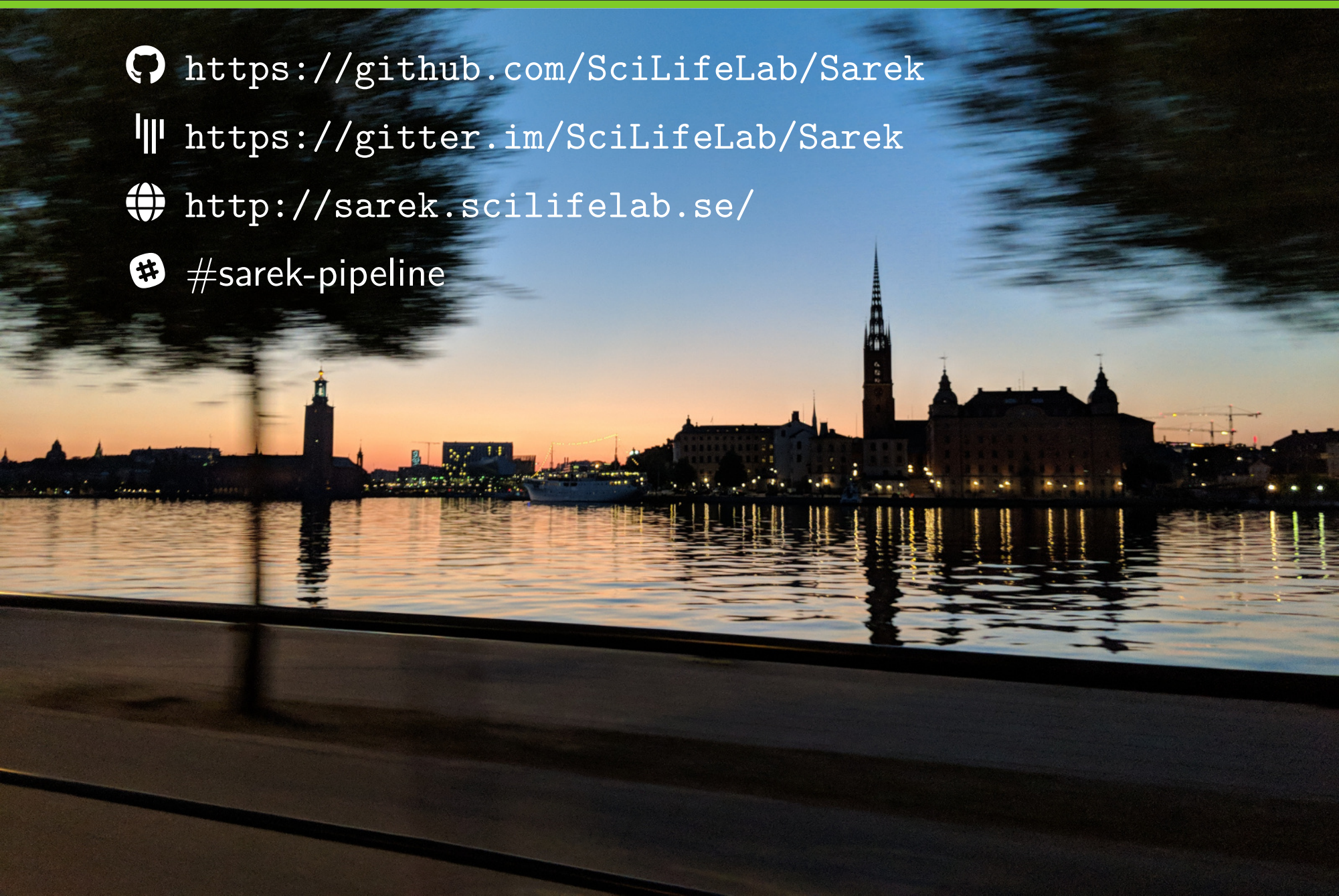
Prioritization

- First step towards clinical use
- Rank scores are computed for all variants
  - COSMIC, ClinVar, SweFreq and MSK-IMPACT (cancerhotspots.org)
- Findings are ranked in three tiers
  - 1^{st} tier: well known, high-impact variants
  - 2^{nd} tier: variants in known cancer-related genes
  - 3^{rd} tier: the remaining variants

Acknowledgments

Barntumörbanken	Elisa Basmaci	**NGI**	Johannes Alneberg	**NBIS**	Sebastian DiLorenzo
	Szilveszter Juhos		Anandashankar Anil		Malin Larsson
	Gustaf Ljungman		Franziska Bonath		Marcel Martin
	Monica Nistèr		Orlando Contreras-López		Markus Mayrhofer
	Gabriela Prochazka		Phil Ewels		Björn Nystedt
	Johanna Sandgren		Sofia Haglund		Markus Ringnér
	Teresita Díaz De Ståhl		Max Käller		Pall I Olason
	Katarzyna Zielinska-Chomej		Anna Konrad		Jonas Söderberg
			Pär Lundin		
Grupp Nistèr	Saad Alqahtani		Remi-Andre Olsen	**Clinical Genomics**	Kenny Billiau
	Min Guo		Senthilkumar Panneerselvam		Hassan Foroughi Asl
	Daniel Hägerstrand		Fanny Taborsak		Valtteri Wirta
	Anna Hedrén		Chuan Wang		
	Martin Proks			**Nextflow folks**	Paolo Di Tommaso
	Rong Yu				Sven Fillinger
	Jian Zhao	**Clinical Genetics**	Jesper Eisfeldt		Alexander Peltzer

Any questions?



https://github.com/SciLifeLab/Sarek

https://gitter.im/SciLifeLab/Sarek

http://sarek.scilifelab.se/

#sarek-pipeline