

Latest news about Sarek



Formerly known as CAW

Maxime Garcia

 @gau

 @MaxUlysse

 <https://maxulysse.github.io/>

2018-02-20

SciLifeLab NGI / BarnTumörBanken






NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN


Karolinska
Institutet




Stockholms
Universitet


UPPSALA
UNIVERSITET



CAW

A screenshot of a web browser window displaying Google search results for the query "CAW". The browser interface includes a back/forward button, a refresh button, a home icon, a search bar with the URL "https://www.google.se/search", a download button, and a search bar with the word "Search". Below the search bar are the Google logo, a search button, and navigation links for "All", "Images", "Videos", "News", "Maps", and "More". A "Settings" and "Tools" link is also present. The main content area shows a search result for "Dictionary". The search term "caw" is entered in a search bar within the dictionary result. The definition for "caw" is provided, including its pronunciation (/kɔ:/), part of speech (noun), and verb definition ("utter a caw"). A note states "'rooks cawed in the dark trees'". At the bottom of the dictionary result, there is a link to "Translations, word origin, and more definitions". A feedback link is located at the bottom right of the search result.

What happens when you look for CAW

Sarek



James Frain as Sarek, "Star Trek: Discovery"

The real inspiration



Mount Skierfe in the Sarek National Park in Summer

How it should look today



Mount Skierfe in Winter

What is 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



What does Sarek do?



Sarek

What does Sarek do?



Sarek



Sarek
Germline

What does Sarek do?



Sarek



Sarek
Germline



Sarek
Somatic

What does Sarek do?



Sarek

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT
- Containers (portable, reproducible)

What does Sarek do?



- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
- SNPs, SNVs and indels
 - MuTect1, MuTect2, Strelka, and GATK HaplotypeCaller
- Structural variants
 - Manta
- Heterogeneity, ploidy and CNVs
 - ASCAT
- Containers (portable, reproducible)
 - Docker or Singularity

Where to use Sarek?

- Any POSIX compatible system

Where to use Sarek?



- Rackham

Where to use Sarek?



- Rackham
- Bianca

Where to use Sarek?



- Rackham
- Bianca
- Irma

Where to use Sarek?



- AWS Batch

Where to use Sarek?



Sarek with AWS Batch



Sarek with AWS Batch



- A single command line

```
nextflow run main.nf -profile awsbatch -w s3://caw-test-results/work \
--genome smallGRCh37 --sample s3://caw-test-data/tsv/tiny-s3.tsv \
--outDir s3://caw-test-results/Results
```

Sarek with AWS Batch



- A single command line

```
nextflow run main.nf -profile awsbatch -w s3://caw-test-results/work \
--genome smallGRCh37 --sample s3://caw-test-data/tsv/tiny-s3.tsv \
--outDir s3://caw-test-results/Results
```

🌐 <https://maxulysse.github.io/2017/11/16/Running-CAW-with-AWS-Batch/>

Going further



- Run a full size test sample

Going further



- Run a full size test sample
- Gather reports

Going further



- Run a full size test sample
- Gather reports
- Get a pricing

Going further



- Run a full size test sample
- Gather reports
- Get a pricing
- Wait for Amazon to finally come to Stockholm

The List of People Involved

Sebastian DiLorenzo	Markus Mayrhofer
Jesper Eisfeldt	Monica Nistèr
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

Where to find us?

- We are on the SciLifeLab Slack 
#cancer-pipeline

Where to find us?

- We are on the SciLifeLab Slack 
`#cancer-pipeline`
- We have a gitter channel
<https://gitter.im/SciLifeLab/Sarek>

Where to find us?

- We are on the SciLifeLab Slack 
`#cancer-pipeline`
- We have a gitter channel
<https://gitter.im/SciLifeLab/Sarek>
- Our code is hosted on Github 
<https://github.com/SciLifeLab/Sarek>

Any questions?

