

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



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 download icon, a star icon, a search bar containing "Search", and a URL bar showing "https://www.google.se/search". Below the search bar are tabs for "All", "Images", "Videos", "News", "Maps", and "More", with "All" being the selected tab. To the right are "Settings" and "Tools" buttons. A vertical scroll bar is visible on the right side of the main content area.

The main content area shows search results for "CAW". At the top, it says "About 64 400 000 results (0,55 seconds)". Below this, there is a "Dictionary" section with a search bar containing "caw" and a search icon. The word "caw" is defined as:

caw
/kɔ:/ ⓘ
noun
1. the harsh cry of a rook, crow, or similar bird.
verb
1. utter a caw.
"rooks cawed in the dark trees"

At the bottom of the dictionary card, there is a "Translations, word origin, and more definitions" link with a downward arrow icon. A "Feedback" link is located at the bottom right of the card.

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?



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?

