Running CAW with AWS Batch

Hopefully my last talk about CAW ever...

Maxime Garcia

@gau
@MaxUlysse
https://maxulysse.github.io/

2018-01-31

SciLifeLab NGI / BarnTumörBanken
Definition

Go to a Google search for the word "CAW".

About 64,400,000 results (0.55 seconds)

Dictionary

caw

caw

/ko:/

noun
1. the harsh cry of a rook, crow, or similar bird.

verb
1. utter a caw.
   "rooks cawed in the dark trees"

Translations, word origin, and more definitions
What is CAW?

http://opensource.scilifelab.se/projects/caw/

- Nextflow pipeline
What is CAW?

- Nextflow pipeline
- Developed at NGI

http://opensource.scilifelab.se/projects/caw/
What is CAW?

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

http://opensource.scilifelab.se/projects/caw/
What is CAW?

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

http://opensource.scilifelab.se/projects/caw/
What does CAW do?

- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GA4TK best practices for processing FASTQ files
- SNPs, SNVs and indels
- MuTect1, MuTect2, Strelka, and GA4TK HaplotypeCaller
- Structural variants
- Manta
- Heterogeneity, ploidy and CNVs
- ASCAT
- Containers (portable, reproducible)
  - Docker or Singularity
What does CAW do?

- WGS analysis (Tumor/Normal pair or Germline)
What does CAW do?

- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
What does CAW do?

- WGS analysis (Tumor/Normal pair or Germline)
- Handles both GRCh37 and GRCh38
- Based on GATK best practices for processing FASTQ files
What does CAW 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 CAW 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 HaplotyeCaller
- Structural variants
  - Manta
- Heterogeneity, ploidy and CNVs
  - ASCAT
- Containers (portable, reproducible)
  - Docker or Singularity
What does CAW 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 CAW 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 CAW 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 HaplotyeCaller
- Structural variants
  - Manta
- Heterogeneity, ploidy and CNVs
What does CAW 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 HaplotyeCaller
- Structural variants
  - Manta
- Heterogeneity, ploidy and CNVs
  - ASCAT
What does CAW 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 CAW 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 HaplotyeCaller
- Structural variants
  - Manta
- Heterogeneity, ploidy and CNVs
  - ASCAT
- Containers (portable, reproducible)
  - Docker or Singularity
Where to use CAW?

- Any POSIX compatible system
Where to use CAW?

- Rackham
Where to use CAW?

- Rackham
- Bianca
Where to use CAW?

- Rackham
- Bianca
- Irma
Where to use CAW?

- Rackham
- Bianca
- Irma
Where to use CAW?

- AWS Batch
Where to use CAW?

OH, REALLY?

PLEASE, TELL ME MORE
CAW 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

CAW with AWS Batch

- A single command line

```bash
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
```
CAW 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

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
- We have a gitter channel
  https://gitter.im/SciLifeLab/CAW
- Our code is hosted on Github
  https://github.com/SciLifeLab/CAW
Where to find us?

- We are on the SciLifeLab Slack #
  #cancer-pipeline
- We have a gitter channel !
  https://gitter.im/SciLifeLab/CAW
Where to find us?

- We are on the SciLifeLab Slack
  
  #cancer-pipeline

- We have a gitter channel

  https://gitter.im/SciLifeLab/CAW

- Our code is hosted on Github

  https://github.com/SciLifeLab/CAW
Any questions?