Cancer Genomics Cloud

- The **Cancer Genomics Cloud** - CGC is a scalable cloud-based environment for analyzing genomics data
- **Securely** store, analyze, and share data amongst team members
- Collaborate more **efficiently**
- Main GOAL → Help researchers to get to results **faster** in **reproducible** and **reliable** manner
Single Variant run - 07-16-18 15:39:04

Executed on July 16, 2018 08:41 by milan_kovacevic

Spot Instances enabled | Price: $0.11 | Refund | View refunds | Duration: 11 minutes

App: Single Variant - Revision: 0

App Settings

null_model.R

null_model.R

covars

sex

Population

outcome

Outputs

assoc_combined

assoc_single_chr1.RData
assoc_single_chr2.RData
assoc_single_chr3.RData
assoc_single_chr4.RData
assoc_single_chr5.RData

... and 17 more files

assoc_plots_output

assoc_single_plots_manh.png
assoc_single_plots_qq.png

and 17 more files
<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Modified by</th>
<th>Modified on</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>GRM</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>KING Robust</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>King Robust and PC-Air and PC-Relate unpacked</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:27</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>Id_pruning</td>
<td>CommandLineTool</td>
<td>milan_kovacevic</td>
<td>July 18, 2018 23:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>PC Air</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>PC Air and PC Relate</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:29</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>PC Relate</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:21</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>Single Variant</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>Sliding Window</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:21</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>vcf2gds</td>
<td>CommandLineTool</td>
<td>milan_kovacevic</td>
<td>July 18, 2018 23:20</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>VCF to GDS</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:19</td>
<td>⚡️ Run</td>
</tr>
<tr>
<td>VCF to GDS and GRM</td>
<td>Workflow</td>
<td>milan_kovacevic</td>
<td>July 16, 2018 07:30</td>
<td>⚡️ Run</td>
</tr>
</tbody>
</table>
Add apps to "TOPMed Analysis Pipeline"
Rabix Composer

Try out Rabix Composer, our new desktop editor. Create CWL 1.0 apps which are portable and which enable enhanced functionalities for workflow development. Enjoy an improved UI with better debugging during development. Learn more.

Download (macOS) Need it for Windows or Linux?

Workflow

Workflows are chains of interconnected tools.

Create a Workflow

Tool

Tools are programs for processing data on the Platform.

Create a Tool
<table>
<thead>
<tr>
<th>File name</th>
<th>Experimental strategy</th>
<th>Path</th>
<th>Created on</th>
<th>Type</th>
<th>Size</th>
<th>Disease type</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1KG_phase3_subset_chr1.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>120.7 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr10.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>128.6 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr11.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>125.6 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr12.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>124.3 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr13.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>127.8 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr14.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>117.3 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr15.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>120.8 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr16.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>119.2 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr17.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td>114.0 KB</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1KG_phase3_subset_chr18.vcf.gz</td>
<td>-</td>
<td>Files</td>
<td>July 16, 2018 07:41</td>
<td>VCF.GZ</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1KG_phase3_subset.gds

1.4 MB (1,423,738 bytes) Uploaded on July 16, 2018 08:37 (Pacific Daylight Time), Modified on July 18, 2018 07:19 (Pacific Daylight Time)

File
- Experimental strategy
- Library ID
- Platform
- Platform unit ID
- File segment number
- Quality scale
- Paired-end
- Reference genome

General
- Investigation

Case
- Case ID
- Case UUID

Case/Demographic
- Gender
- Age

Other files uploaded with this one
- 1KG_phase3_subset_chr18.vcf.gz
- 1KG_phase3_subset_chr6.vcf.gz
- 1KG_phase3_subset_chr17.vcf.gz
- 1KG_phase3_subset_chr20.vcf.gz
- 1KG_phase3_subset_chr8.vcf.gz
- 1KG_phase3_subset_chr9.vcf.gz
- 1KG_phase3_subset_chr2.vcf.gz
- 1KG_phase3_subset_chr19.vcf.gz

and 86 more Preview files (94)
### Add files to “TOPMed Analysis Pipeline”

<table>
<thead>
<tr>
<th>File name</th>
<th>Size</th>
<th>Path</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ucac_Bisulfite_Genome_bowtie2.tar.gz</td>
<td>8.4 GB</td>
<td>Files</td>
<td>TAR.GZ</td>
</tr>
<tr>
<td>GRCh37_To_Hg19.over.chain.gz</td>
<td>1.8 KB</td>
<td>Files</td>
<td>CHAIN.GZ</td>
</tr>
<tr>
<td>hg19_To_GRCh37.over.chain.gz</td>
<td>1.7 KB</td>
<td>Files</td>
<td>CHAIN.GZ</td>
</tr>
<tr>
<td>hg38_To_Hg19.over.chain.gz</td>
<td>1.2 MB</td>
<td>Files</td>
<td>CHAIN.GZ</td>
</tr>
<tr>
<td>hg19_To_Hg38.over.chain.gz</td>
<td>222.4 KB</td>
<td>Files</td>
<td>CHAIN.GZ</td>
</tr>
<tr>
<td>ucac.hg19.msisensor_microsats_list.txt</td>
<td>1.4 GB</td>
<td>Files</td>
<td>TXT</td>
</tr>
<tr>
<td>hg19.IobSTR_v3.0.2.ref.tar</td>
<td>4.6 GB</td>
<td>Files</td>
<td>TAR</td>
</tr>
<tr>
<td>ucac_hg19_Broad_variant_lift.gtf</td>
<td>132.6 MB</td>
<td>Files</td>
<td>GTF</td>
</tr>
<tr>
<td>exome_targets.hg19.sorted.bed</td>
<td>749.1 KB</td>
<td>Files</td>
<td>BED</td>
</tr>
<tr>
<td>GSM1629193_hg19.ERRCC.refFlat</td>
<td>31.9 MB</td>
<td>Files</td>
<td>REFFLAT</td>
</tr>
<tr>
<td>GSM1629193_hg19.ERRCC.fasta</td>
<td>3.0 GB</td>
<td>Files</td>
<td>FASTA</td>
</tr>
</tbody>
</table>
Type, paste or drag and drop your links in the field below.

*How do I use this tool?*

ftp://john:mypassword@superseq.com/results/NA18507

or

Browse for a file with links

Import
Command Line Uploader
Upload your private data using fast and secure upload client, taking advantage of parallelization where possible.
Download
Learn more

Desktop Uploader
Upload your private data from a personal computer or hard drive to any of your existing projects.
Download (macOS)
Need it for Windows or Linux?
Learn more

SBFS (Seven Bridges File System) Beta
Mount your projects and use files locally.
Important note: before installing SBFS, please install FUSE for macOS.
To install SBFS, run the following command:

curl https://static.abgenomics.com/sbfs/install.sh -sSf | sudo sh
Or download the executable file for macOS.
Need it for Linux? (not available for Windows)
Learn more

Upload files via the API
Upload files using the Seven Bridges Python library.

```
files = [
    '/foo/bar/baz.bam
'/foo/bar/qux.fastq'
]
for file in files:
    api.files.upload(project='my-project',
                   path=file)
```
Learn more
Volumes

No volumes connected

Connect a volume

Select your cloud storage provider:

- Amazon Web Services
- Google Cloud Platform

Next
Welcome to your new project!

Projects are the core building blocks of the CGC Platform. Each project corresponds to a distinct scientific investigation, serving as a container for its data, analysis pipelines, and results. Projects are shared only by designated project members.

Within your project, you can:
- Start exploring public datasets straight away
- Install your tools on the CGC and create workflows
- Upload your own private data and analyze it along with public datasets
- Collaborate securely with other researchers

Please record the details of your project here, such as its aims, experimental context, and any other ideas that you’d like to share with your project members. Remember that details of each pipeline execution you run on the CGC are logged on the task page. This notepad is just for your own notes.

You can also use markdown here to add formatting to your notes.

Good luck with your research! If you get stuck, take a look at the Knowledge Center.

The Seven Bridges CGC Team
Welcome to your new project!

Projects are the core building blocks of the CGC, scientific investigation, serving as a container for shared only by designated project members.

Within your project, you can:
- Start exploring public datasets straight away
- Install your tools on the CGC and create workflows
- Upload your own private data and analyze it against other researchers
- Collaborate securely with other researchers

Please record the details of your project here, such as ideas you’d like to share with your project members and execution you run on the CGC.

You can also use markdown here to add formatting.

Good luck with your research! If you get stuck, talk to The Seven Bridges CGC Team.
Single Variant run - 07-16-18 15:39:04

Executed on July 16, 2018 08:41 by milan_kovacevic
Spot Instances enabled | Price: $0.11 | Duration: 11 minutes

App: Single Variant - Revision: 0

### Inputs

- **Variant include files**
  - variant.include_chr1.RData
  - variant.include_chr10.RData
  - variant.include_chr14.RData
  - variant.include_chr16.RData
  - variant.include_chr2.RData
  - ...and 17 more files

- **conditional_variant_file**
  - No files selected

- **gds_file_for_null_model**
  - No files selected

- **gds_files**
  - 1KG_phase3_subset_chr9.gds
  - 1KG_phase3_subset_chr6.gds
  - 1KG_phase3_subset_chr10.gds
  - 1KG_phase3_subset_chr13.gds

### App Settings

- **null_model.R** (`/null_model_r`)
  - **covars**
    - sex
    - Population
    - outcome
  - outcome

### Outputs

- **assoc_combined**
  - assoc_single_chr1.RData
  - assoc_single_chr2.RData
  - assoc_single_chr3.RData
  - assoc_single_chr4.RData
  - assoc_single_chr5.RData
  - ...and 17 more files

- **assoc_plots_output**
  - assoc_single_plots_manh.png
  - assoc_single_plots_qq.png
null_model_r
assoc_single_r
define_segments_r
abg_flatten
cmd.log

Rscript /usr/local/analysis_pipeline/R/assoc_single.R --segment 137 assoc_single.config
Explore genomics data

Understand complex genomics data with interactive analysis tools.

**Genome Browser**
Visualize alignments, SNV/Indels, annotation tracks, check coverage and mismatch, assess alignments and variants

**Data Cruncher**
Analyze and explore data using Jupyter notebooks

**Variant Browser BETA**
Filter and interpret your annotated data
Variant type per chromosome

Show the count of different variant types per chromosome. When normalized, it is displayed in SVs per base pair.

```python
In [12]:
normalize = True
removemitochondrial = True
show(variant_type_per_chromosome(vcf, normalize=normalize,
                                      karyotype=karyotype, removemitochondrial=removemitochondria
                                      ))
```
<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>Ref</th>
<th>Alt</th>
<th>Gene name</th>
<th>Zygosity</th>
<th>Features count</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>764625</td>
<td>ACTG</td>
<td>A</td>
<td>AC090044.1</td>
<td>Heterozygous</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1363304</td>
<td>CT</td>
<td>C</td>
<td>CNTN6</td>
<td>Heterozygous</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3217128</td>
<td>AT</td>
<td>A</td>
<td>CRBN</td>
<td>Homozygous</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>4940288</td>
<td>AG</td>
<td>A</td>
<td>BHLHE40-AS1</td>
<td>Homozygous</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>7172236</td>
<td>AA</td>
<td>A</td>
<td>GRM7</td>
<td>Heterozygous</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>7600819</td>
<td>CT</td>
<td>C</td>
<td></td>
<td>Homozygous</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>9519363</td>
<td>AT</td>
<td>A</td>
<td>SETD5</td>
<td>Heterozygous</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>9743398</td>
<td>TG</td>
<td>T</td>
<td></td>
<td>Homozygous</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>10075900</td>
<td>TG</td>
<td>T</td>
<td></td>
<td>Heterozygous</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>10083526</td>
<td>CT</td>
<td>C</td>
<td></td>
<td>Homozygous</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>10105661</td>
<td>CT</td>
<td>C</td>
<td>FANCD2</td>
<td>Heterozygous</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>10142749</td>
<td>TA</td>
<td>T</td>
<td></td>
<td>Heterozygous</td>
<td>10</td>
</tr>
</tbody>
</table>
Single Variant run - 07-16-18 15:39:04

Executed on July 16, 2018 08:41 by milan_kovacevic

Spot Instances enabled | Price: $0.11 Refund | View refunds | Duration: 11 minutes

App: Single Variant - Revision: 0

App Settings

null_model.R (null_model.r)

covars

sex

Population

outcome

Outputs

assoc_combined

assoc_single_chr1.RData
assoc_single_chr2.RData
assoc_single_chr3.RData
assoc_single_chr4.RData
assoc_single_chr5.RData

...and 17 more files

assoc_plots_output

![](assoc_single_plots_manh.png)
![](assoc_single_plots_qq.png)

1KG_phase3_subset_chr9.gds
1KG_phase3_subset_chr6.gds
1KG_phase3_subset_chr10.gds
1KG_phase3_subset_chr13.gds
Single Variant

Copy of Single Variant (Latest revision), by milan_kovacevic on July 16, 2018 07:20

Description

No description.

Basic Information

- CWL Version: v1.0
- Contributors: milan_kovacevic
- App Id: milan_kovacevic/topmed/single-variant

Workflow steps

- null_model.R
- define_segments.R
- assoc_single.R
- SBG Prepare Segments
- assoc_combine.R
- SBG Group Segments
- SBG FlattenLists
- assoc_plots.R

Ports

Inputs  App Settings  Outputs
```json
"class": "Workflow",
"cwVersion": "v.1.0",
"id": "omnioric/gtopmed-development/single-variant/10",
"label": "Single Variant",
"snamespaces": {
  "sbp": "https://sevenbridges.com"
},
"inputs": [
  {
    "id": "phenotype_file",
    "sbp:fileTypes": "RDATA",
    "ctype": "File",
    "sbpy": -450.1188659669769,
    "sbpx": -971.215951367188
  },
  {
    "id": "conditional_variant_file",
    "sbp:fileTypes": "RDATA",
    "ctype": "File",
    "sbpy": 209.274169921875,
    "sbpx": 969.2352876171875
  },
  {
    "id": "gds_file_for_null_model",
    "sbp:fileTypes": "GDS",
    "ctype": "File",
    "sbpy": 75,
    "sbpx": -970.89697265625
  },
  {
    "id": "pmd_file",
    "sbp:fileTypes": "RDATA, GDS",
    "ctype": "File",
    "sbpy": -64.44434301578125,
    "sbpx": -962.235462966308
  },
  {
    "id": "pca_file",
    "sbp:fileTypes": "RDATA",
    "ctype": "File",
    "sbpy": -255.3557831845703125,
    "sbpx": -984.889538542188
  },
  {
    "id": "pocrelate_file",
    "sbp:fileTypes": "GDS",
    "ctype": "File",
    "sbpy": -324.85238126953125,
    "sbpx": -220.4960899047852
  }
}
```
DOCKER IMAGE

Docker Repository
uwgac/topmed-master:latest

BASE COMMAND

Rscript /usr/local/analysis_pipeline/R/vcf2gds.R vcf2gds.config

ARGUMENTS

Set arguments such as parameters or options to hard code them for every execution of a tool. For instance, set the output file name as an argument instead of an input port to use a fixed output file name.

Add an Argument

Learn More

INPUT PORTS

<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>vcf_file</td>
<td>File</td>
<td>unbound</td>
</tr>
<tr>
<td>gds_file_name</td>
<td>string</td>
<td>unbound</td>
</tr>
</tbody>
</table>

Rscript /usr/local/analysis_pipeline/R/vcf2gds.R vcf2gds.config
Links

- CGC
- CGC Docs
- Rabix Composer
- Seven Bridges
Questions?