31 Mar 2025

Sarek_Mic

Advanced Molecular Detection
Southeast Region Bioinformatics

This resource was made possible through funding provided under the Epidemiology and Laboratory Capacity for Prevention and Control of Emerging Infectious Diseases (ELC) Cooperative Agreement (CK24-0002), Project D: Advanced Molecular Detection to the Florida Department of Health. The conclusions, findings, and opinions expressed by authors do not necessarily reflect the official position of the U.S. Department of Health and Human Services, the Public Health Service, or the Centers for Disease Control and Prevention.

Updates

- Complete BRR Needs Assessment if you haven't yet – the more responses, the better!
- BRR/BTL Training - Delayed this year after changing from ABiL to UGA as our tutorial provider. Please update your ELC Milestones to notify them of this. We maintain contact with SaGE/OAMD team on these changes as well!

Overview

Purpose

- Analyze microorganisms such as Candida auris, SARS-CoV-2, TB, and more with references which result in SNP calling and annotation that enhances for genomic research.

Usage

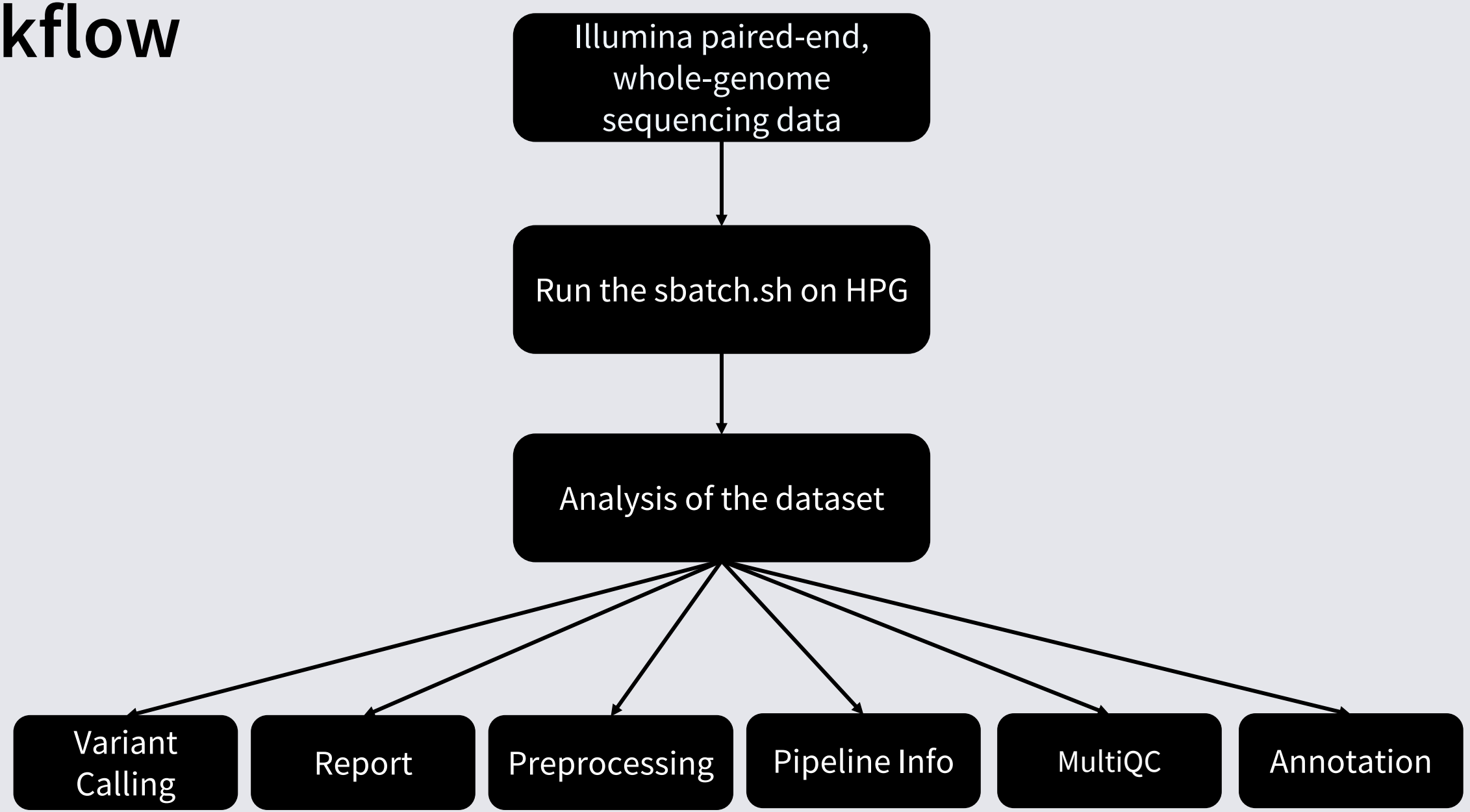
- To support public health and researchers by providing detailed reports and analyses of the data which enables insights into drug resistance monitoring, genomic research, outbreak surveillance and epidemiological studies.

Dependencies

- Nextflow
- nf-core

Advanced Molecular Detection
Southeast Region Bioinformatics

Workflow



Advanced Molecular Detection
Southeast Region Bioinformatics

Application

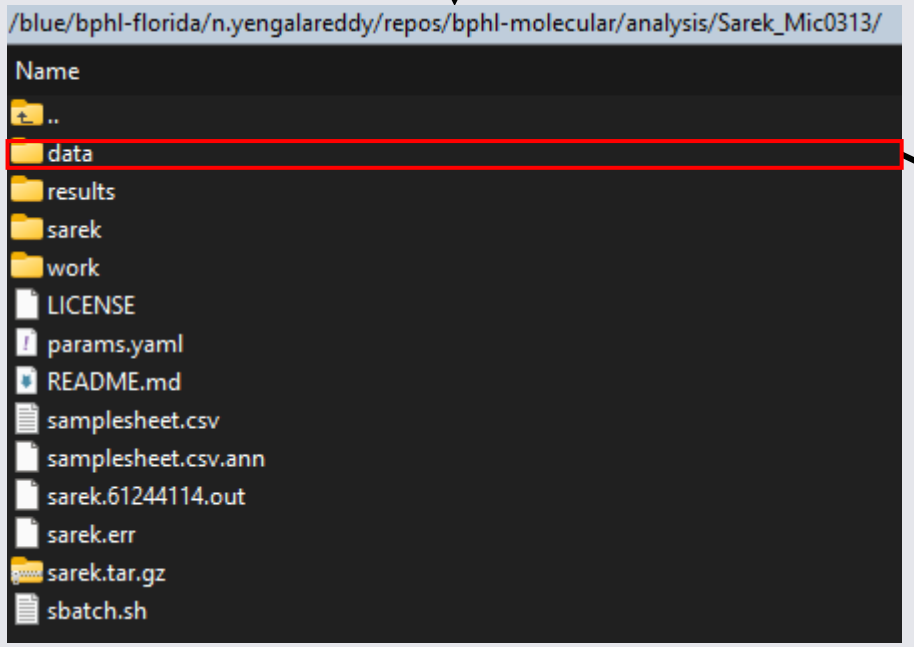
Objective

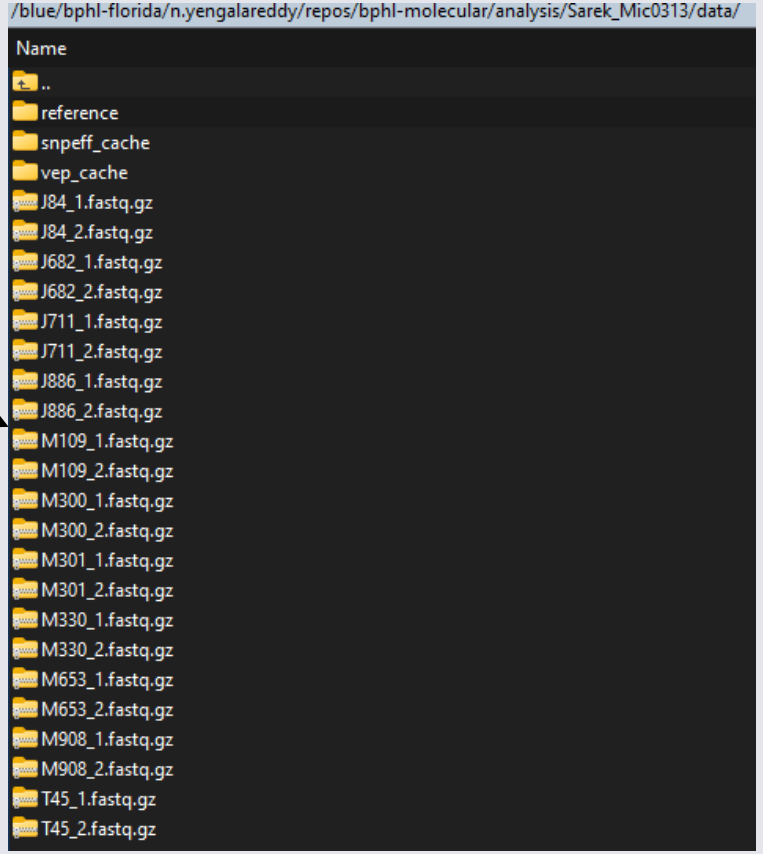
Use SARS-CoV-19 dataset and analyze using Sarek-Mic Pipeline

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
cd /blue/bphl-<state>/<user>/repos/bphl-molecular/
git clone https://github.com/BPHL-Molecular/Sarek_Mic
mkdir analysis/
cd analysis/
cp /blue/bphl-<state>/<user>/repos/bphl-molecular/ Sarek_MIc/
copy .fastq to data directory
```





Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

nano params.yaml

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/

Name

- ..
- data
- results
- sarek
- work
- LICENSE
- params.yaml
- README.md
- samplesheet.csv
- samplesheet.csv.ann
- sarek.61244114.out
- sarek.err
- sarek.tar.gz
- sbatch.sh

GNU nano 2.9.8 params.yaml

```
input: './samplesheet.csv'
outdir: './results/'
#genome: 'null'
tools: 'mpileup,haplotypecaller,freebayes,strelka,snpeff,vep'
skip_tools: 'baserecalibrator,baserecalibrator_report,dnascope_filter'
#step: 'annotate'
igenomes_ignore: 'true'
snpeff_cache: '/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/data/snpeff_cache'
snpeff_db: 'MN908947.3'
vep_cache: '/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/data/vep_cache'
vep_species: 'sars_cov_2'
vep_genome: 'ASM985889v3'
vep_cache_version: '101'

fasta: '/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/data/reference/covid/nCoV-2019.reference.fasta'
fasta_fai: '/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/data/reference/covid/nCoV-2019.reference.fasta.fai'
only_paired_variant_calling: 'true'
```

Application Cont.

nano sbatch.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/

Name

- ..
- data
- results
- sarek
- work
- LICENSE
- params.yaml
- README.md
- samplesheet.csv
- samplesheet.csv.ann
- sarek.61244114.out
- sarek.err
- sarek.tar.gz
- sbatch.sh

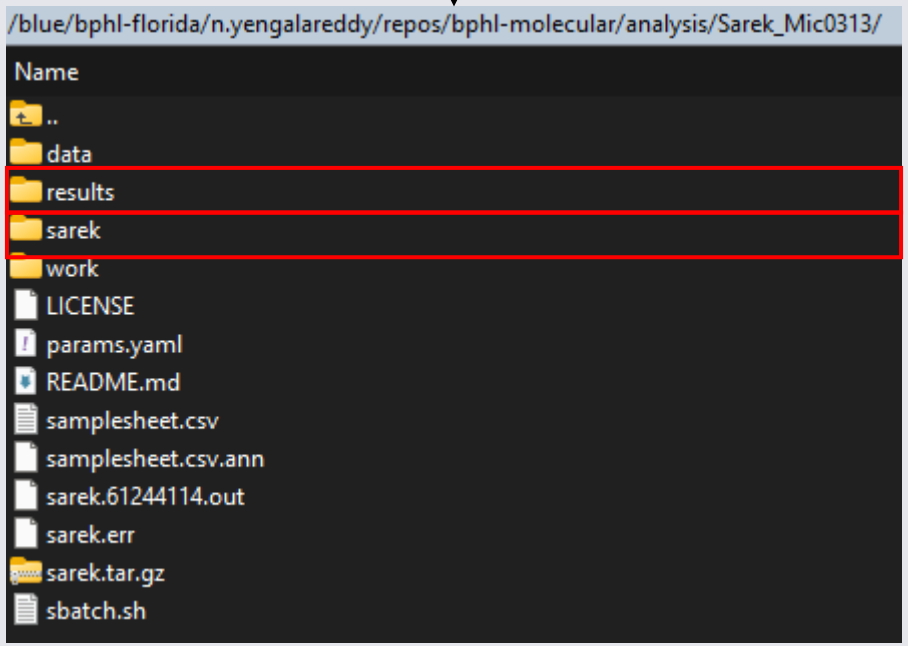
GNU nano 2.9.8 sbatch.sh

```
#!/usr/bin/bash
#SBATCH --account=bphl-umbrella
#SBATCH --qos=bphl-umbrella
#SBATCH --job-name=sarek
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=20
#SBATCH --mem=50gb
#SBATCH --time=48:00:00
#SBATCH --output=sarek.%j.out
#SBATCH --error=sarek.err
#SBATCH --mail-user=nikhil.yengala@flhealth.gov
#SBATCH --mail-type=FAIL,END

module load nextflow

APPTAINER_CACHEDIR=./
export APPTAINER_CACHEDIR

#nextflow run nf-core/sarek -profile singularity --input ./samplesheet.csv --outdir ./output --tools 'haplotypecaller,snpeff'
#GRCh38
#nextflow run ./sarek -profile singularity --input ./samplesheet.csv --outdir ./output --tools 'freebayes,snpeff'

#hg38 ucsc from igenomes
#nextflow run ./sarek -profile singularity --input samplesheet.csv --outdir ./ --genome hg38 --tools 'freebayes,snpeff' --skip_tools baserecal

# custom genomes
#nextflow run ./sarek -profile singularity --input samplesheet.csv --outdir ./ --tools 'freebayes' --genome null --igenomes_ignore --fasta ./r
#nextflow run ./sarek -profile hs37d5,singularity --input samplesheet.csv --outdir ./ --tools 'freebayes,snpeff' --skip_tools baserecalibrator
#nextflow run ./sarek -profile singularity --input samplesheet.csv --outdir ./Candida --tools 'freebayes,snpeff' --genome null --igenomes_igno
```

Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

```
tar xvzf sarek.tar.gz
activate nf-core env
sbatch ./sbatch.sh
```



Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/

Name
- ..
- data
- results
- sarek
- work
- LICENSE
- params.yaml
- README.md
- samplesheet.csv
- samplesheet.csv.ann
- sarek.61244114.out
- sarek.err
- sarek.tar.gz
- sbatch.sh

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/results/

Name
- ..
- annotation
- build
- csv
- multiqc
- pipeline_info
- preprocessing
- reference
- reports
- variant_calling

/.../n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/results/annotation/

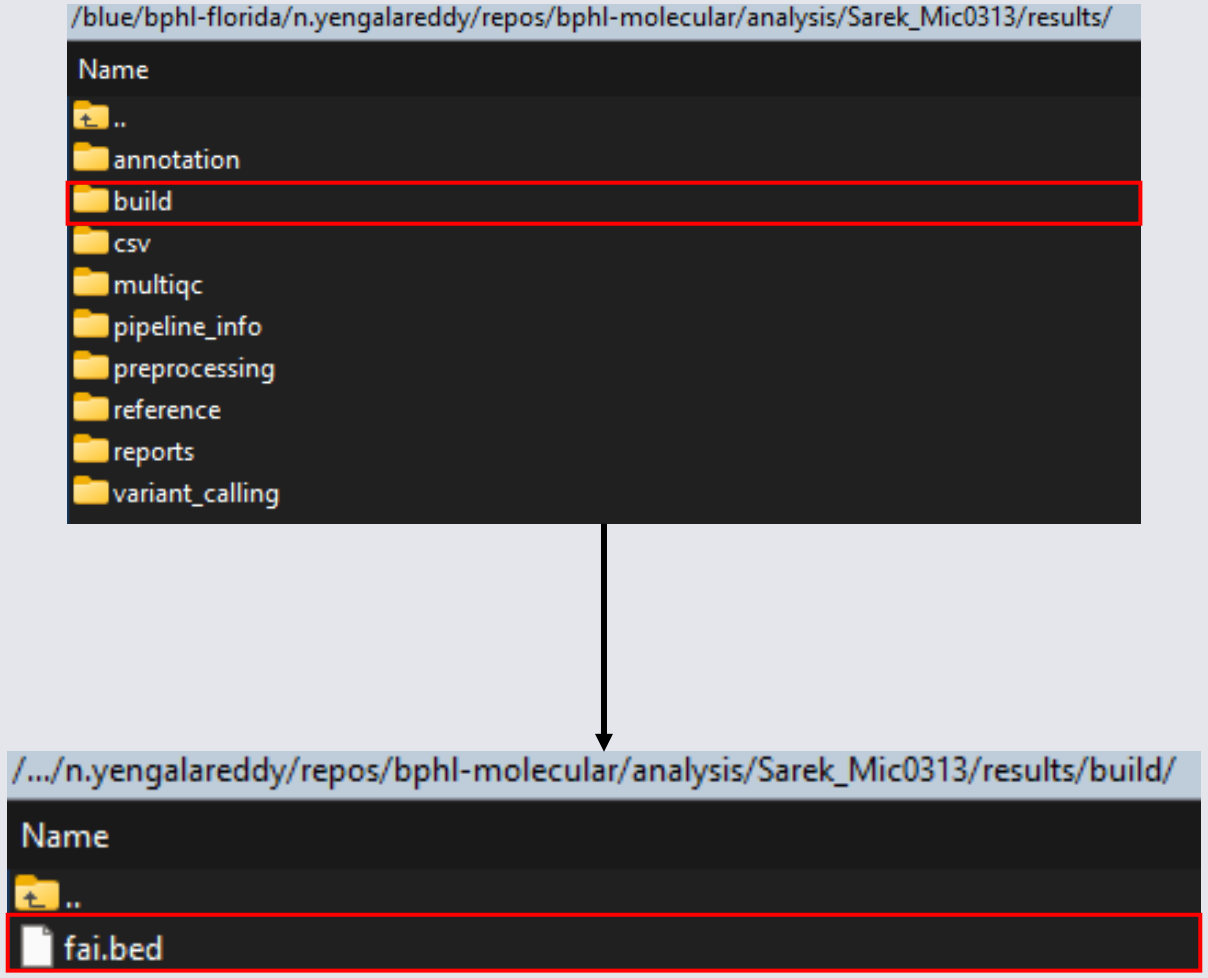
Name
- ..
- bcftools
- freebayes
- strelka

/.../repos/bphl-molecular/analysis/Sarek_Mic0313/results/annotation/bcftools/

Name
- ..
- J84
- J682
- J711
- J886
- M109
- M300
- M301
- M330
- M653
- M908
- T45

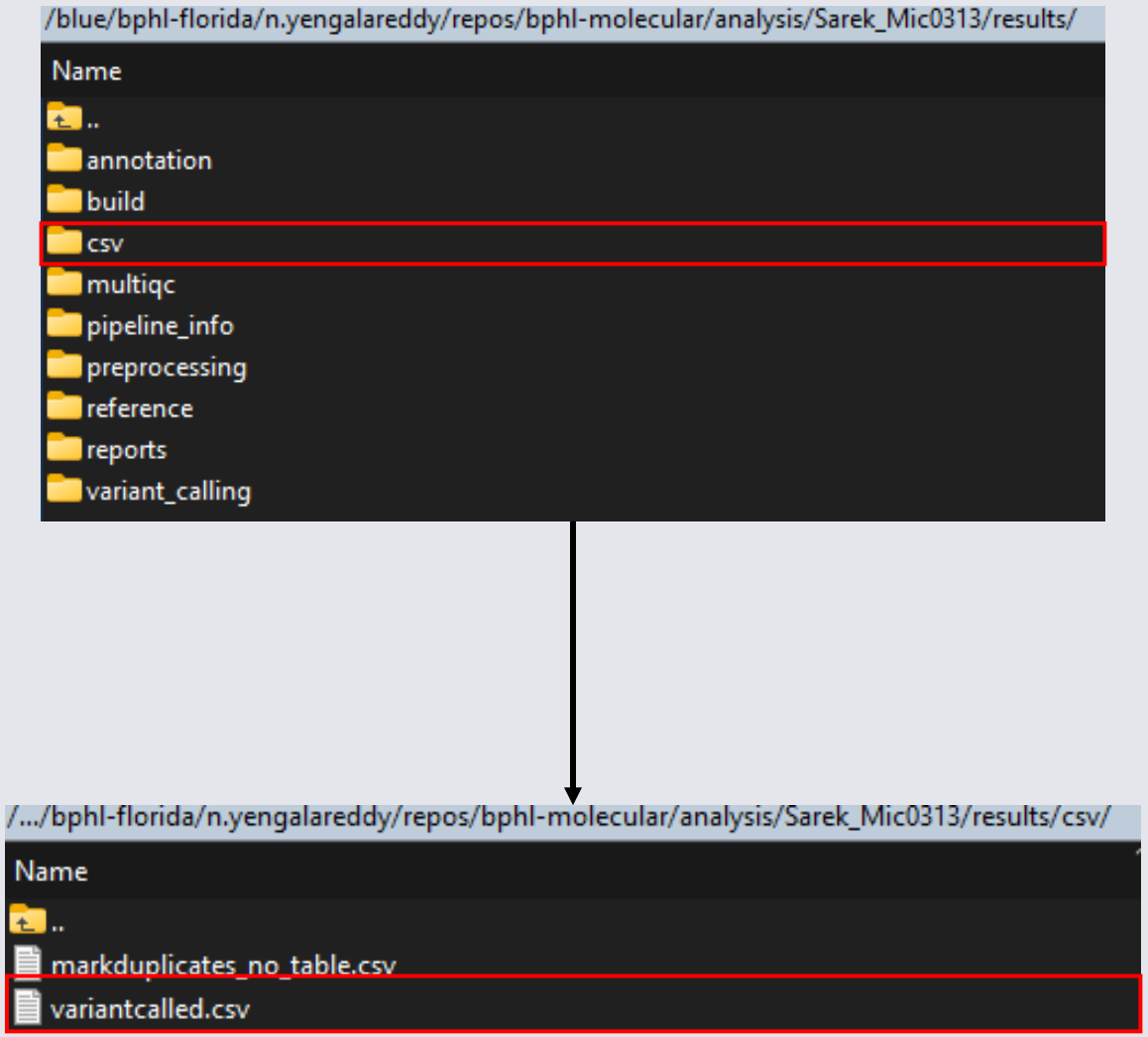
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

/blue/bphl-florida/n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/results/

Name

- ..
- annotation
- build
- csv
- multiqc
- pipeline_info
- preprocessing
- reference
- reports
- variant_calling

/.../n.yengalareddy/repos/bphl-molecular/analysis/Sarek_Mic0313/results/multiqc/

Name

- ..
- multiqc_data
- multiqc_plots
- multiqc_report.html

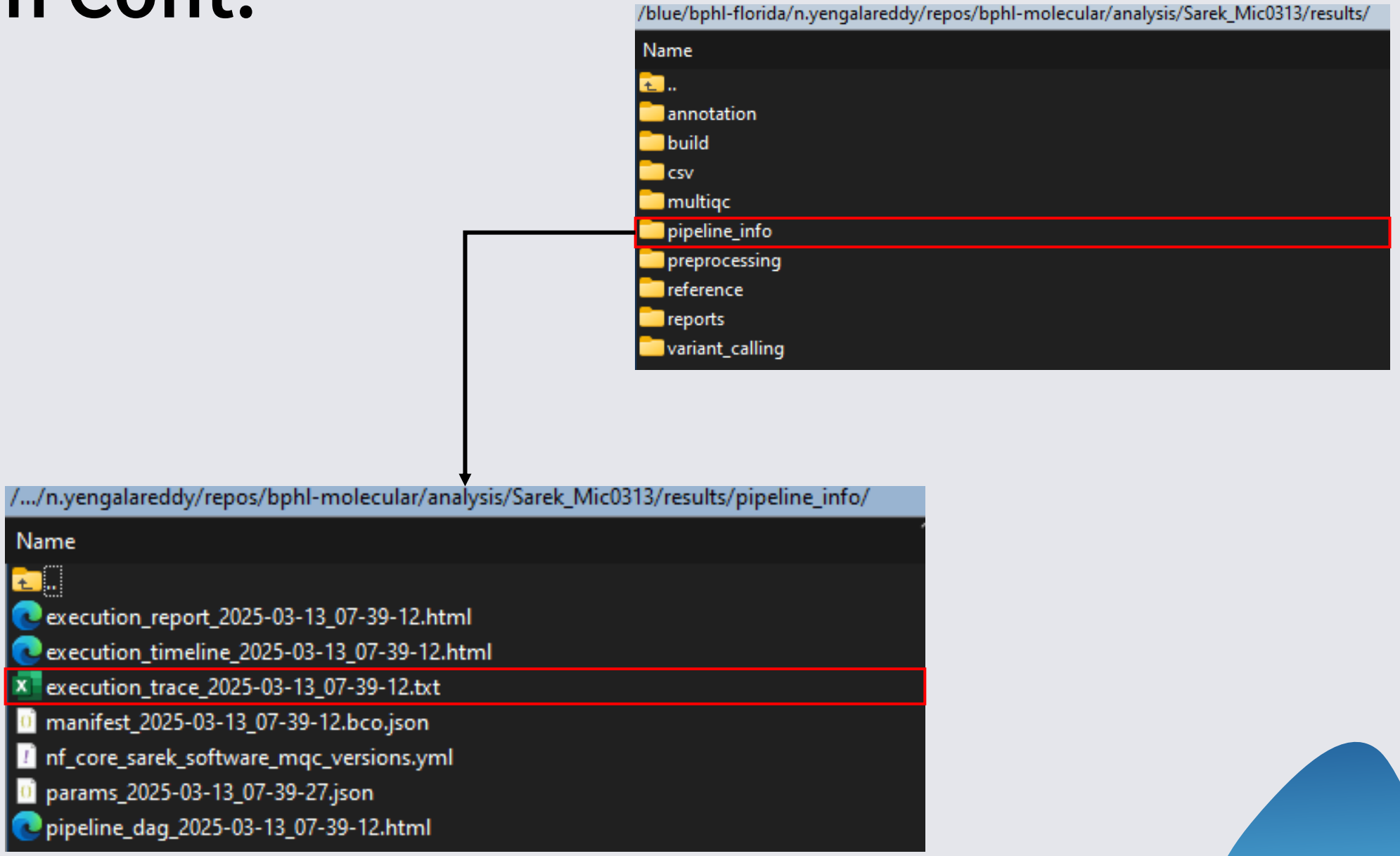
/.../repos/bphl-molecular/analysis/Sarek_Mic0313/results/multiqc/multiqc_data/

Name

- ..
- mosdepth_cov_dist.txt
- mosdepth_cumcov_dist.txt
- mosdepth_perchrom.txt
- multiqc.log
- multiqc_bcftools_stats.txt
- multiqc_citations.txt
- multiqc_data.json
- multiqc_fastp.txt
- multiqc_fastqc.txt
- multiqc_general_stats.txt
- multiqc_picard_dups.txt
- multiqc_samtools_stats.txt
- multiqc_snpeff.txt
- multiqc_software_versions.txt
- multiqc_sources.txt
- picard_histogram.txt
- picard_histogram_1.txt
- picard_histogram_2.txt
- vcftools_tstv_by_count.txt
- vcftools_tstv_by_qual.txt
- vep.txt

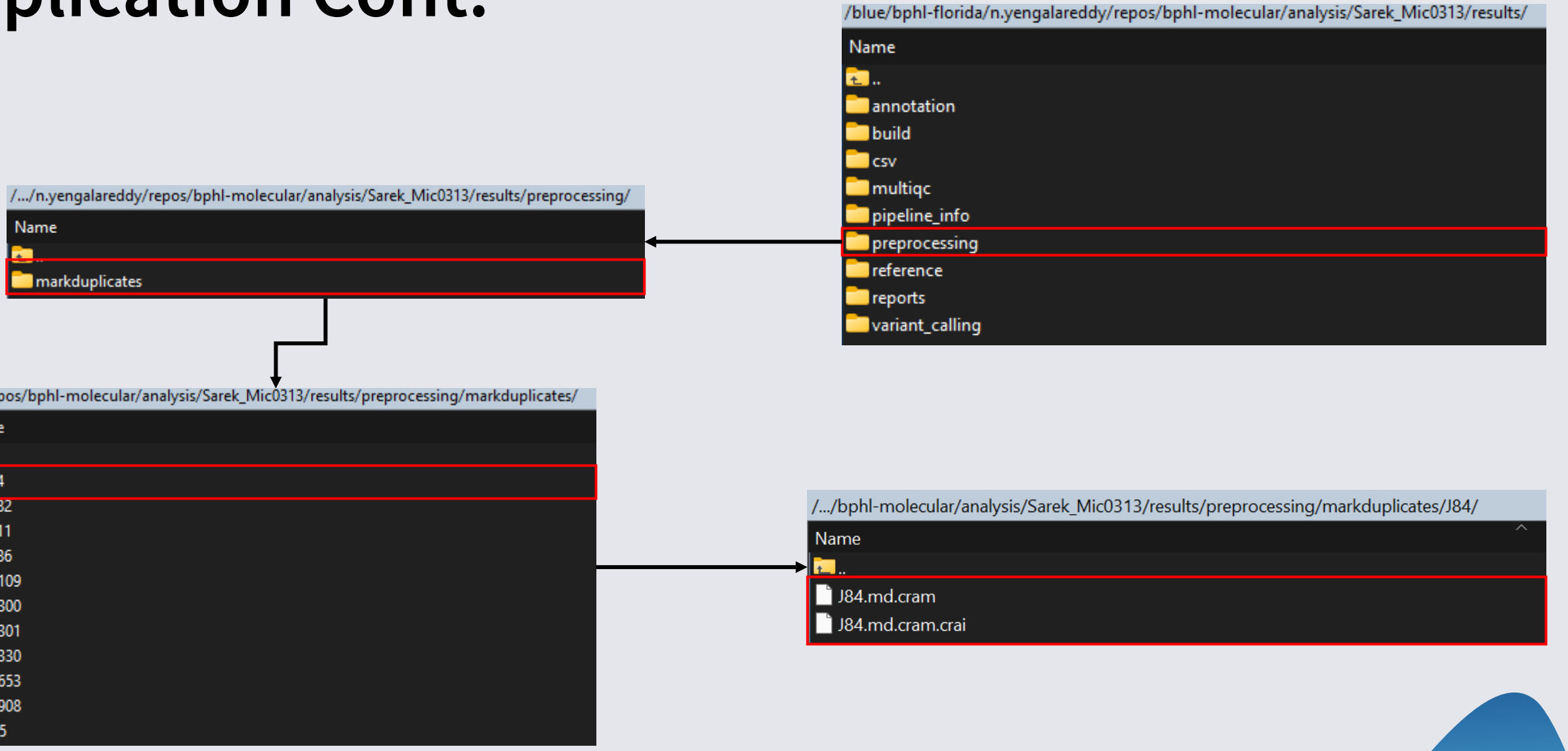
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



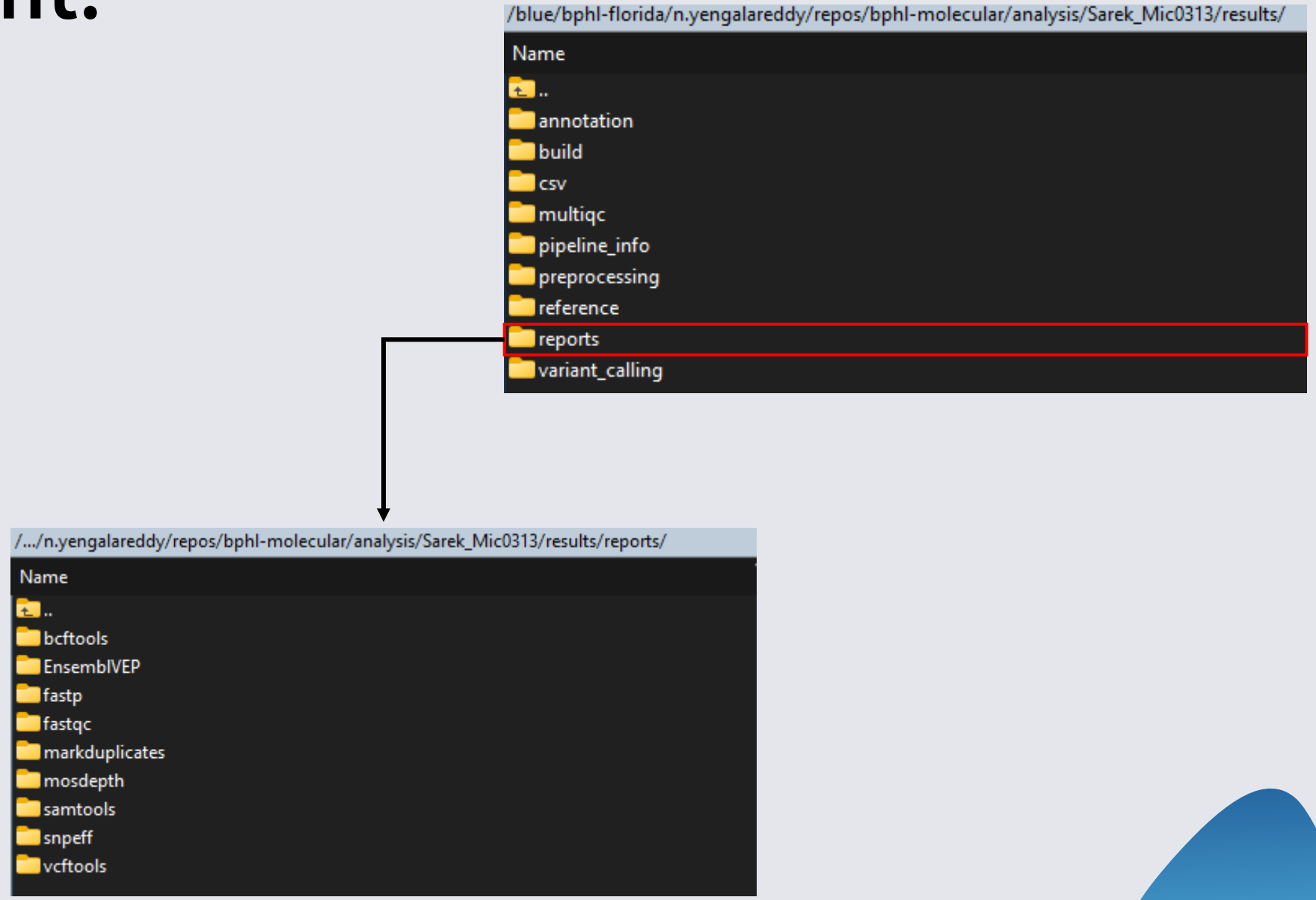
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



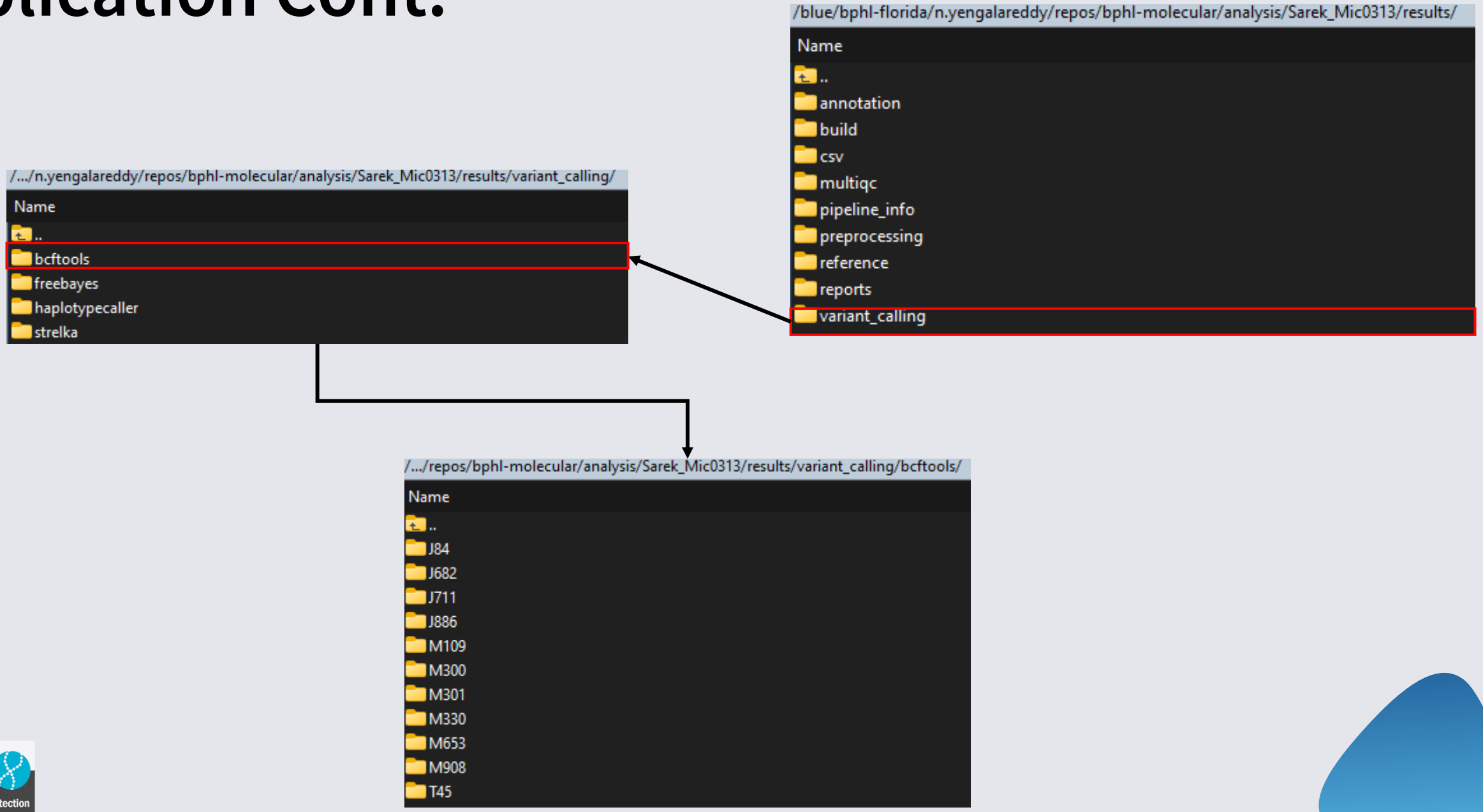
Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.



Advanced Molecular Detection
Southeast Region Bioinformatics

Application Cont.

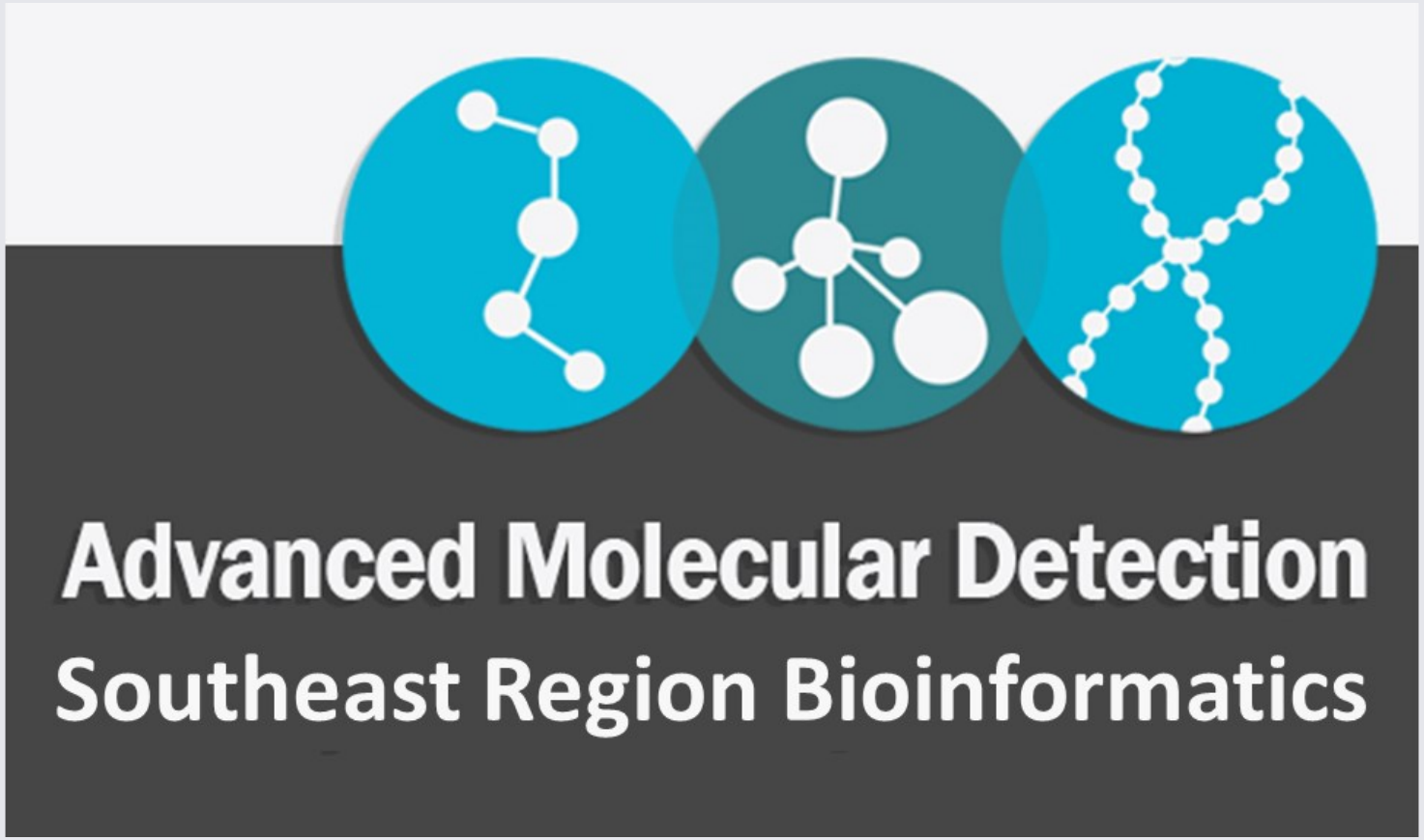


Advanced Molecular Detection
Southeast Region Bioinformatics

Conclusion

- Fundamentals of Sarek_Mic
- Installation and setup of Sarek_Mic in HPG
- Successfully executed job query for Sarek_Mic
- Generated output files

Advanced Molecular Detection
Southeast Region Bioinformatics



Questions?

bphl-sebioinformatics@flhealth.gov

Molly Mitchell, PhD
Bioinformatics Supervisor
Molly.Mitchell@flhealth.gov

Nikhil Reddy, MS
Bioinformatician
Nikhil.Yengala@flhealth.gov

Sam Bernhoft, MPH
Bioinformatician
Samantha.bernhoft@flhealth.gov