

# Nf-core ?????? ??????? ?? (scrnaseq)

📄 : <https://catalog.us-east-1.prod.workshops.aws/workshops/76d4a4ff-fe6f-436a-a1c2-f7ce44bc5d17/en-US>

📄 📄 📄 📄 📄 . 📄 📄 📄 .

2025 📄 6 📄 6 📄 scrnaseq v4.0.0 📄 📄 . (nextflow 📄 📄 nf-scheam 📄 📄 )

📄 📄 📄 📄 2.7.1 📄 📄 .

## ????? ??

📄 📄 📄 **Bash** 📄 📄

```
cd ~

export yourbucket="your-bucket-name"
export your_account_id="your-account-id"
export region="your-region"
export workflow_name="your-workflow-name"
export omics_role_name="your_omics_rolename"

# if not exist the bucket, let's create.
#aws s3 mb $yourbucket
```

nf-core repository???? ?????? ??

```
git clone https://github.com/nf-core/scrnaseq --branch 2.7.1 --single-branch
```

# Docker Image Manifest? ??

```
cp ~/amazon-ecr-helper-for-aws-healthomics/lib/lambda/parse-image-uri/public_registry_properties.json namespace
```

inspect\_nf.py [ ] [ ][ ][ ][ ][ ] .

```
python3 amazon-omics-tutorials/utlils/scripts/inspect_nf.py \
--output-manifest-file scrnaseq_271_docker_images_manifest.json \
-n namespace.config \
--output-config-file omics.config \
--region $region \
~/scrnaseq/
```

```

[ ] [ ] [ ] scrnaseq_271_docker_images_manifest.json [ ] omics.config [ ] .

```

scrnaseq\_271\_docker\_images\_manifest.json       :

```
[ec2-user@ip-172-31-22-158 ~]$ cat scrnaseq_271-docker_images_manifest.json
{
  "manifest": [
    "quay.io/biocontainers/bioconductor-alevinqc:1.12.1--r41h9f5acd7_0",
    "quay.io/biocontainers/fastqc:0.12.1--hdfd78af_0",
    "quay.io/biocontainers/gffread:0.12.1--h8b12597_0",
    "quay.io/biocontainers/gffread:0.12.7--hd03093a_1",
    "quay.io/biocontainers/kb-python:0.28.2--pyhdfd78af_2",
    "quay.io/biocontainers/mulled-v2-1fa26d1ce03c295fe2fdcf85831a92fbcdb7e8c2:ded3841da0194af2701c780e9b3d653a85d27489-0",
    "quay.io/biocontainers/multiqc:1.23--pyhdfd78af_0",
    "quay.io/biocontainers/p7zip:16.02",
    "quay.io/biocontainers/python:3.9--1",
    "quay.io/biocontainers/scanpy:1.7.2--pyhdfd78af_0",
    "quay.io/biocontainers/simpleaf:0.10.0--h9f5acd7_1",
    "quay.io/biocontainers/star:2.7.10b--h9ee0642_0",
    "quay.io/nf-core/cellranger-arc:2.0.2",
    "quay.io/nf-core/cellranger:8.0.0",
    "quay.io/nf-core/seurat:4.3.0",
    "quay.io/nf-core/ubuntu:22.04",
    "quay.io/nf-core/universc:1.2.5.1",
    "quay.io/quay.io/biocontainers/bioconductor-dropletutils:1.18.0--r42hf17093f_1"
  ]
}
```

???? ??? (into Amazon ECR)

```
aws stepfunctions start-execution \  
--state-machine-arn arn:aws:states:$region:$your_account_id:stateMachine:omx-container-puller \  
--input file://scrnaseq_271_docker_images_manifest.json
```

Step Functions

State machines

Activities

Developer resources

Online learning workshop

Local Development

Data flow simulator

Feature spotlight

Documentation

Join our feedback panel

Updated extension for VS Code

With the latest AWS Toolkit extension for Visual Studio Code, you can edit and test workflows graphically with Workflow Studio. [Learn more](#).

Download the AWS Toolkit

State machines (2)

View execution counts

View details

Edit

Copy to new

Delete

Search for state machines

Any type

Name	Type	Creation date	Status
<a href="#">omx-container-builder</a>	Standard	Apr 15, 2025, 15:44:36 (UTC+09:00)	Active
<a href="#">omx-container-puller</a>	Standard	Apr 15, 2025, 14:55:08 (UTC+09:00)	Active

Executions

Monitoring

Logging

Definition

Aliases

Versions

Tags

Executions (5)

View details

Stop execution

Redrive

Start execution

Filter executions by property or value

Filter by status

Last 15 months

5 matches

1

Name	Status	Start Time (local)	End Time (local)	Duration
<a href="#">465b6f92-b88a-4277-8465-8f45bbd8755c</a>	Succeeded	Aug 20, 2024, 15:31:10	Aug 20, 2024, 15:32:25	00:01:15.020
<a href="#">14ad44ec-264a-49ca-bb2e-04bde507ac7e</a>	Succeeded	Aug 19, 2024, 21:18:42	Aug 19, 2024, 21:21:41	00:02:59.200
<a href="#">6bb79dab-d511-41a0-8ba1-db9e905b8fb0</a>	Succeeded	Jun 21, 2024, 21:09:49	Jun 21, 2024, 21:15:16	00:05:26.830
<a href="#">edc05886-0499-4b95-baa4-4ca995d97e2f</a>	Succeeded	Jun 17, 2024, 13:19:44	Jun 17, 2024, 13:20:45	00:01:00.930
<a href="#">83eda741-4152-4fa9-a5ae-5ed29ba0941c</a>	Succeeded	Jun 14, 2024, 18:30:55	Jun 14, 2024, 18:31:54	00:00:59.320

nf-core project ?? ?????

```
mv omics.config scrnaseq/conf
```

```
echo "includeConfig 'conf/omics.config'" >> scrnaseq/nextflow.config
```

AWS HealthOmics ?????? ???

## 1. AWS HealthOmics

parameter-description.json

```
cat << EOF > parameter-description.json
{
  "input": {"description": "Samplesheet with sample locations.",
    "optional": false},
  "protocol": {"description": "10X Protocol used: 10XV1, 10XV2, 10XV3",
    "optional": false},
  "aligner": {"description": "choice of aligner: alevin, star, kallisto",
    "optional": false},
  "barcode_whitelist": {"description": "Optional whitelist if 10X protocol is not used.",
    "optional": true},
  "gtf": {"description": "S3 path to GTF file",
    "optional": false},
  "fasta": {"description": "S3 path to FASTA file",
    "optional": false},
  "skip_emptydrops": {"description": "module does not work on small dataset",
    "optional": true}
}
EOF
```

## 2. Workflow Setup

```
zip -r scrnaseq-workflow.zip scrnaseq -x "*/\.*" "*/\.*/*"
```

```
aws s3 cp scrnaseq-workflow.zip s3://${yourbucket}/workshop/scrnaseq-workflow.zip
```

```
aws omics create-workflow \
  --name ${workflow_name} \
  --definition-uri s3://${yourbucket}/workshop/scrnaseq-workflow.zip \
  --parameter-template file://parameter-description.json \
  --engine NEXTFLOW
```

## 3. Workflow ID Retrieval

```
workflow_id=$(aws omics list-workflows --name ${workflow_name} --query 'items[0].id' --output text)
echo $workflow_id
```

????? ??????

?? ??? ?? ??

???? ??

parameter-description.json 1 111 11 111 11 1111 input.json 111 11 1111 . 11  
11111 1111 11 S3 11 11 1111 1111 .

111 111 111 1111 11111 . (111)

111 11 11

11 1111 111 11 1111 11 1 1111 .

```
wget https://github.com/nf-core/test-datasets/raw/scrnaseq/samplesheet-2-0.csv
wget https://github.com/nf-core/test-datasets/raw/scrnaseq/reference/GRCm38.p6.genome.chr19.fa
wget https://github.com/nf-core/test-datasets/raw/scrnaseq/reference/gencode.vM19.annotation.chr19.gtf
```

11 11 11111 1111 1 1111 111 1111 1 1111 .

```
aws s3 sync . s3://omics-output-us-east-1-462922227709/workflow_migration_workshop/nfcore-scrnaseq-
v4.0.0/
```

sample sheet 111

```
cat << EOF > samplesheet-2-0.csv
sample,fastq_1,fastq_2,expected_cells,seq_center
Sample_X,s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/Sample_X_S1_L001_R1_001.fastq.gz,s3://aws-genomics-static-
${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/Sample_X_S1_L001_R2_001.fastq.gz,5000,"Broad Institute"
Sample_Y,s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/Sample_Y_S1_L001_R1_001.fastq.gz,s3://aws-genomics-static-
${region}/workflow_migration_workshop/nfcore-scrnaseq-v2.3.0/Sample_Y_S1_L001_R2_001.fastq.gz,5000,"CRG
Barcelona"
Sample_Y,s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
```

```
v2.3.0/Sample_Y_S1_L002_R1_001.fastq.gz,s3://aws-genomics-static-
${region}/workflow_migration_workshop/nfcore-scrnaseq-v2.3.0/Sample_Y_S1_L002_R2_001.fastq.gz,5000,"CRG
Barcelona"
EOF
```

## ## **samplesheet** ## s3 ##

```
aws s3 cp samplesheet-2-0.csv s3://${yourbucket}/nfcore-scrnaseq/samplesheet-2-0.csv
```

## **json** ## (## **sample sheet**## ## ## **input**## ## ## )

```
cat << EOF > input.json
{
  "input": "s3://${yourbucket}/nfcore-scrnaseq/samplesheet-2-0.csv",
  "protocol": "10XV2",
  "aligner": "star",
  "fasta": "s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/GRCm38.p6.genome.chr19.fa",
  "gtf": "s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/gencode.vM19.annotation.chr19.gtf",
  "skip_emptydrops": true
}
EOF
```

?? ??? ??

(## : ##)

## ## ##

```
aws s3 sync s3://ngi-igenomes/test-data/scrnaseq/ s3://${yourbucket}/test-data/scrnaseq/ --exclude "*" --include
"pbmc8k_S1_L00*"
```

**sample sheet** ##

```
cat << EOF > samplesheet_2.0_full.csv
sample,fastq_1,fastq_2,expected_cells
```

```
pbmc8k,s3://${yourbucket}/test-data/scrnaseq/pbmc8k_S1_L007_R1_001.fastq.gz,s3://${yourbucket}/test-
data/scrnaseq/pbmc8k_S1_L007_R2_001.fastq.gz,10000
pbmc8k,s3://${yourbucket}/test-data/scrnaseq/pbmc8k_S1_L008_R1_001.fastq.gz,s3://${yourbucket}/test-
data/scrnaseq/pbmc8k_S1_L008_R2_001.fastq.gz,10000
EOF
```

❏❏    ❏❏    **samplesheet**❏    **s3**❏    ❏❏

```
aws s3 cp samplesheet_2.0_full.csv s3://${yourbucket}/nfcore-scrnaseq/samplesheet_2.0_full.csv
```

❏❏    **json**    ❏❏❏    (❏    **sample sheet**❏❏❏    ❏❏    ❏❏❏    **input**❏    ❏❏❏    ❏❏❏❏❏    )

```
cat << EOF > input_full.json
{
  "input": "s3://${yourbucket}/nfcore-scrnaseq/samplesheet_2.0_full.csv",
  "protocol": "10XV2",
  "aligner": "star",
  "fasta": "s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/GRCm38.p6.genome.chr19.fa",
  "gtf": "s3://aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-
v2.3.0/gencode.vM19.annotation.chr19.gtf",
  "skip_emptydrops": true
}
EOF
```

## Policy ??

### Prepare IAM service role to run AWS HealthOmics workflow

your-bucket-name, your-account-id, your-region❏    ❏❏    ❏❏    ❏❏❏    ❏❏    ❏❏❏❏    ❏❏❏❏❏    .

omics\_workflow\_policy.json    ❏❏❏

```
cat << EOF > omics_workflow_policy.json
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
```

```

        "s3:GetObject"
    ],
    "Resource": [
        "arn:aws:s3:::${yourbucket}/*",
        "arn:aws:s3:::aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-v2.3.0/*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "s3:ListBucket"
    ],
    "Resource": [
        "arn:aws:s3:::${yourbucket}",
        "arn:aws:s3:::aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-v2.3.0",
        "arn:aws:s3:::aws-genomics-static-${region}/workflow_migration_workshop/nfcore-scrnaseq-v2.3.0/*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "s3:PutObject"
    ],
    "Resource": [
        "arn:aws:s3:::${yourbucket}/*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "logs:DescribeLogStreams",
        "logs:CreateLogStream",
        "logs:PutLogEvents"
    ],
    "Resource": [
        "arn:aws:logs:${region}:${your_account_id}:log-group:/aws/omics/WorkflowLog:log-stream:*"
    ]
},
{
    "Effect": "Allow",

```



```

    "Action": [
        "logs:CreateLogGroup"
    ],
    "Resource": [
        "arn:aws:logs:${region}:${your_account_id}:log-group:/aws/omics/WorkflowLog:*"
    ]
},
{
    "Effect": "Allow",
    "Action": [
        "ecr:BatchGetImage",
        "ecr:GetDownloadUrlForLayer",
        "ecr:BatchCheckLayerAvailability"
    ],
    "Resource": [
        "arn:aws:ecr:${region}:${your_account_id}:repository/*"
    ]
}
]
}
EOF

```

```
echo "omics_workflow_policy.json" > omics_workflow_policy.json
```

```
trust_policy.json
```

```

cat << EOF > trust_policy.json
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "omics.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "${your_account_id}"
                }
            }
        }
    ]
}
EOF

```

```

        "ArnLike": {
            "aws:SourceArn": "arn:aws:omics:${region}:${your_account_id}:run/*"
        }
    }
}
]
}
EOF

```

```
echo "trust_policy.json" | cat /dev/null
```

## IAM Role ??

```
aws iam create-role --role-name ${omics_role_name} --assume-role-policy-document file://trust_policy.json
```

Policy document

```
aws iam put-role-policy --role-name ${omics_role_name} --policy-name OmicsWorkflowV1 --policy-document
file://omics_workflow_policy.json
```

## ????? ??

```

input.json, input_full.json

```

```

aws omics start-run \
  --name scrnaseq_workshop_test_run_1 \
  --role-arn arn:aws:iam::${your_account_id}:role/${omics_role_name} \
  --workflow-id ${workflow_id} \
  --parameters file://input.json \
  --output-uri s3://${yourbucket}/output/

```

Revision #31

Created 22 July 2024 11:45:23 by Hyunmin Kim

Updated 6 June 2025 14:47:36 by Hyunmin Kim