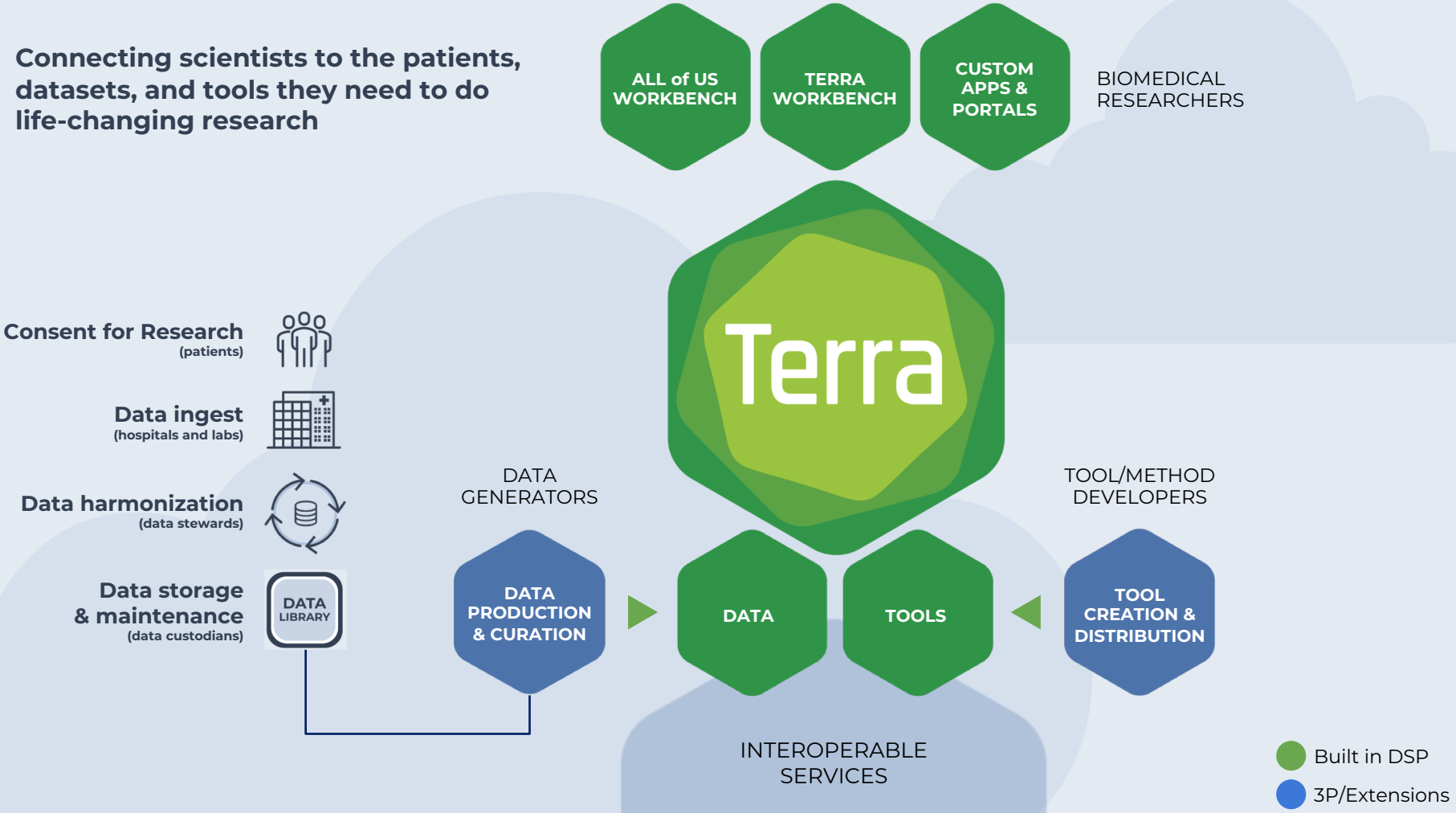
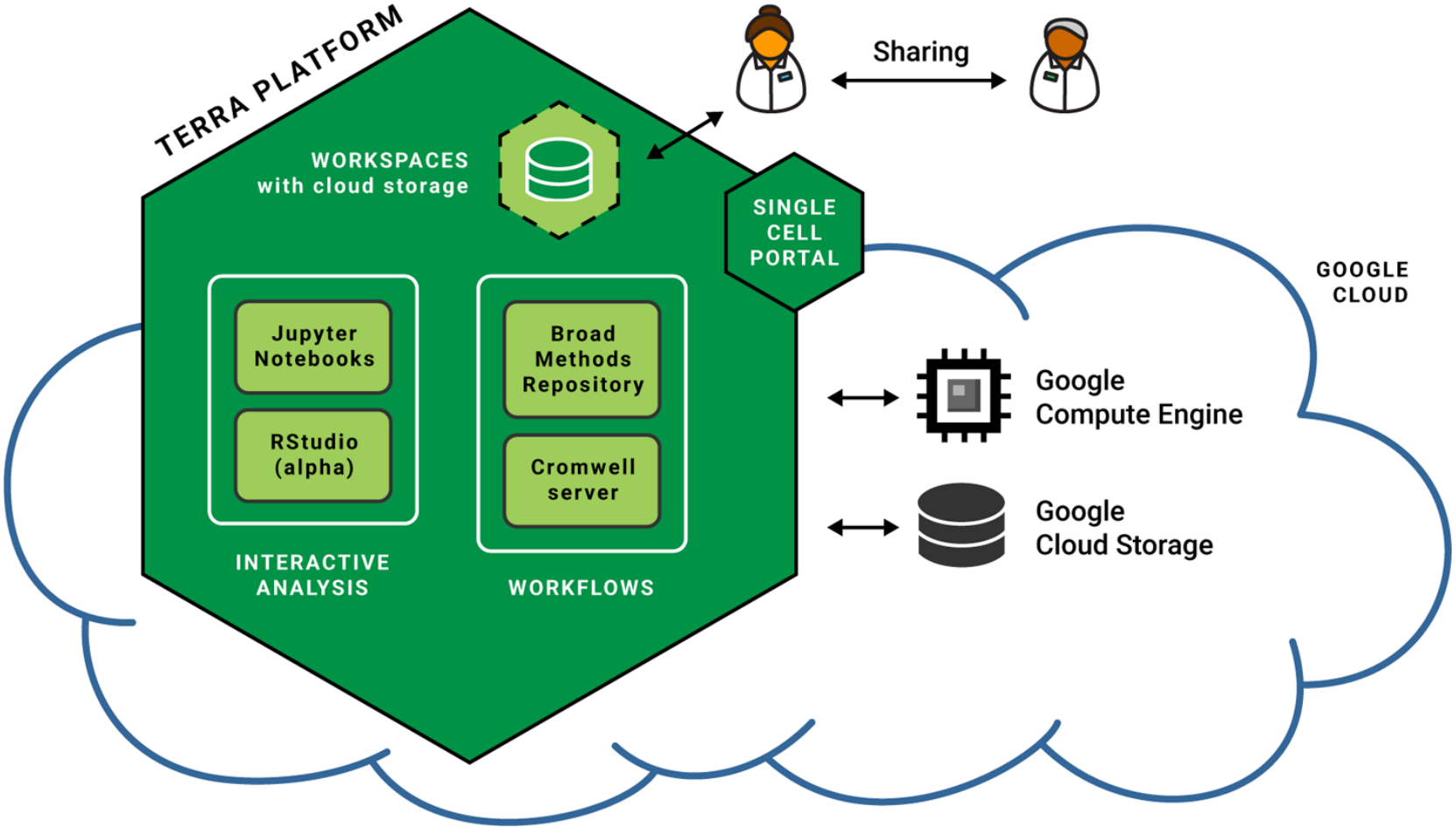
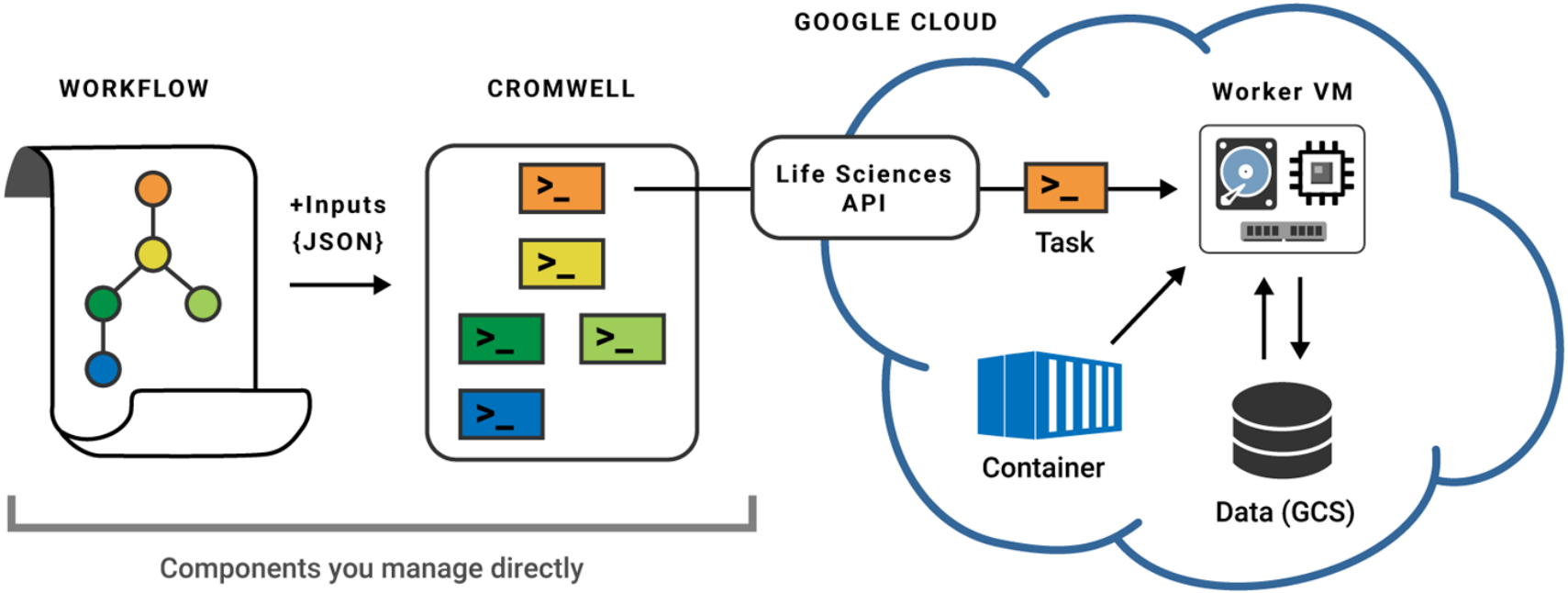
Connecting scientists to the patients, datasets, and tools they need to do life-changing research

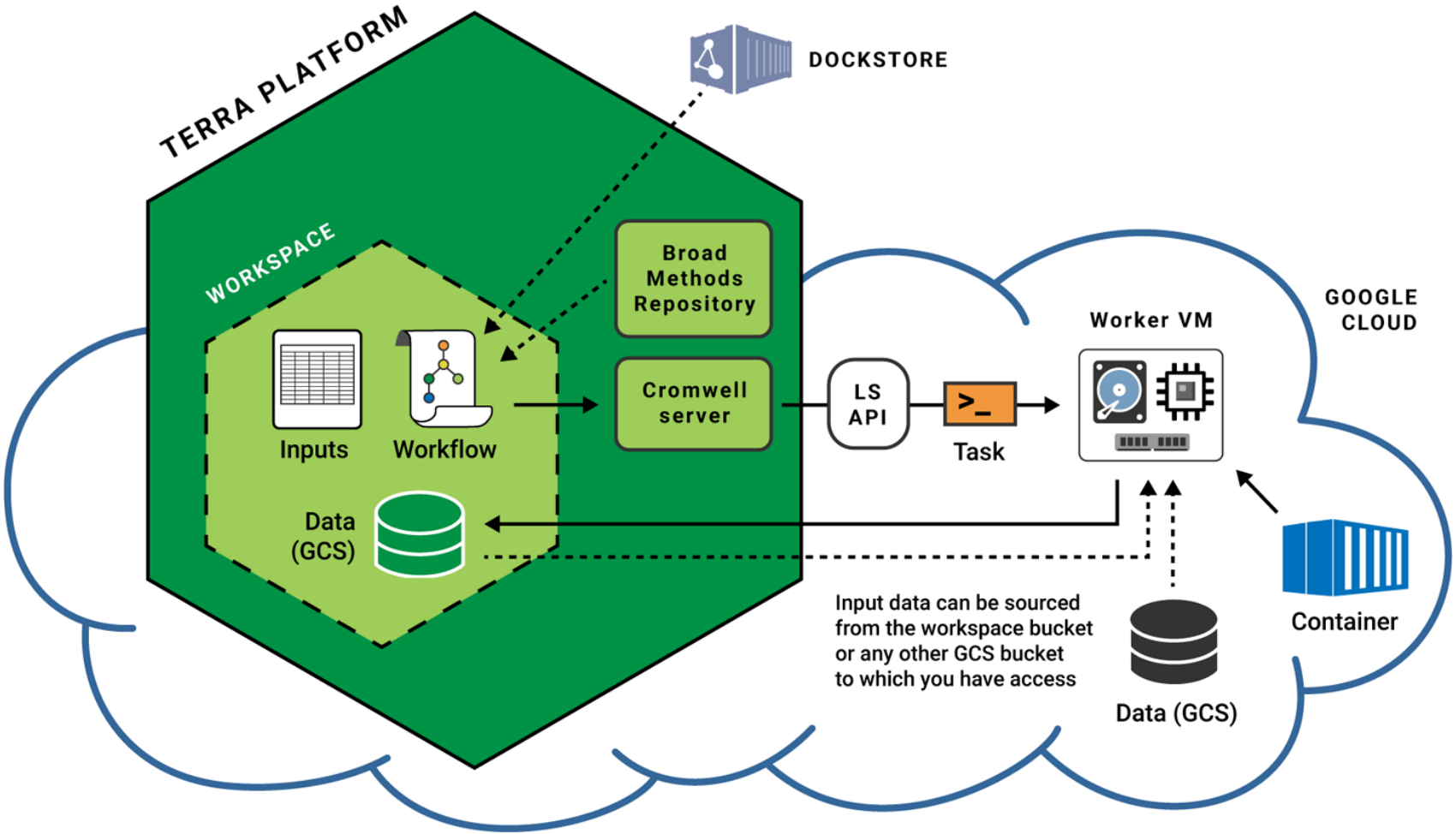


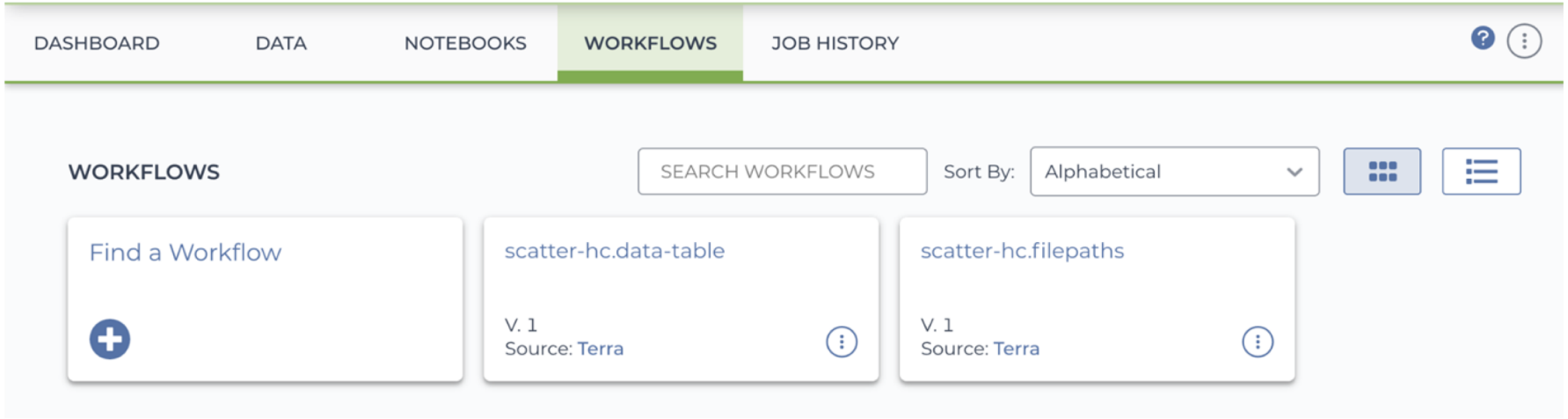

TERRA PLATFORM
WORKSPACES
with cloud storage
Sharing
SINGLE
CELL
PORTAL
GOOGLE
CLOUD
Jupyter
Notebooks
RStudio
(alpha)
INTERACTIVE
ANALYSIS
Broad
Methods
Repository
Cromwell
server
WORKFLOWS
Google
Compute Engine
Google
Cloud Storage

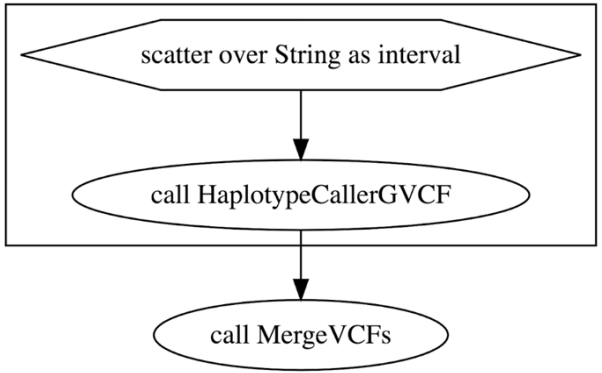
RUNNING WORKFLOWS

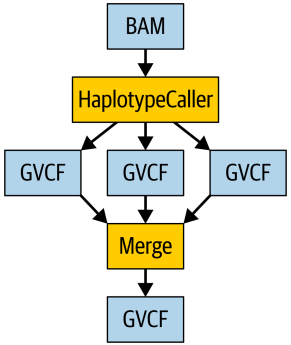


Cromwell dispatching workflows to Google Cloud


TERRA PLATFORM
DOCKSTORE
WORKSPACE
Inputs
Workflow
Data
(GCS)
Broad
Methods
Repository
Cromwell
server
LS
API
Task
Worker VM
GOOGLE
CLOUD
Container
Input data can be sourced
from the workspace bucket
or any other GCS bucket
to which you have access
Data (GCS)


DASHBOARD
DATA
NOTEBOOKS
WORKFLOWS
JOB HISTORY
WORKFLOWS
SEARCH WORKFLOWS
Sort By:
Alphabetical
Find a Workflow
scatter-hc.data-table
V. 1
Source: Terra
scatter-hc.filepaths
V. 1
Source: Terra


scatter over String as interval
call HaplotypeCallerGVCF
call MergeVCFs


BAM
HaplotypeCaller
GVCF
GVCF
GVCF
Merge
GVCF

SCRIPT • • INPUTS • • OUTPUTS • • RUN ANALYSIS

```
## This workflow runs the HaplotypeCaller tool from GATK4 in GVCF mode
## on a single sample in BAM format. The execution of the HaplotypeCaller
## tool is parallelized using an intervals list file. The per-interval
## output GVCF files are then merged to produce a single GVCF file for
## the sample, which can then be used by the joint-discovery workflow
## according to the GATK Best Practices for germline short variant
## discovery.

version 1.0

workflow ScatterHaplotypeCallerGVCF {

    input {
        File input_bam
        File input_bam_index
        File intervals_list
    }

    String output_basename = basename(input_bam, ".bam")

    Array[String] calling_intervals = read_lines(intervals_list)

    scatter(interval in calling_intervals) {
```

SCRIPT • • INPUTS • • OUTPUTS • • RUN ANALYSIS

Download json | Drag or click to upload json SEARCH INPUTS

Task name	Variable	Type	Attribute
HaplotypeCallerGVCF	docker_image	String	"broadinstitute/gatk:4.1.3.0"
HaplotypeCallerGVCF	java_opt	String	"-Xmx8G"
HaplotypeCallerGVCF	ref_dict	File	"gs://genomics-on-the-cloud/book-bundle-v0/data/germline/ref/ref.dict"
HaplotypeCallerGVCF	ref_fasta	File	"gs://genomics-on-the-cloud/book-bundle-v0/data/germline/ref/ref.fasta"

Direct file paths

SCRIPT • • INPUTS • • OUTPUTS • • RUN ANALYSIS

Download json | Drag or click to upload json SEARCH INPUTS

Task name	Variable	Type	Attribute
HaplotypeCallerGVCF	docker_image	String	workspace.gatk_docker
HaplotypeCallerGVCF	java_opt	String	"-Xmx8G"
HaplotypeCallerGVCF	ref_dict	File	workspace.ref_dict
HaplotypeCallerGVCF	ref_fasta	File	workspace.ref_fasta
HaplotypeCallerGVCF	ref_index	File	workspace.ref_fasta_index
MergeVCFs	docker_image	String	workspace.gatk_docker
MergeVCFs	java_opt	String	"-Xmx8G"
ScatterHaplotypeCallerGVCF	input_bam	File	this.input_bam
ScatterHaplotypeCallerGVCF	input_bam_index	File	this.input_bam_index

References to data tables

this.input_bam

workspace.gatk_docker

“Workspace data”

workspace.gatk_docker

Key	Value
gatk_docker	broadinstitute/gatk:4.1.3.0
intervals_list_full	snippet-intervals-full.list
intervals_list_min	snippet-intervals-min.list
ref_dict	ref.dict
ref_fasta	ref.fasta
ref_fasta_index	ref.fasta.fai

Input data

this.input_bam

☐	book_samples_id	input_bam	input_bam_index
☐	father	father.bam	father.bai
☐	mother	mother.bam	mother.bai
☐	son	son.bam	son.bai

SCRIPT • • INPUTS • • OUTPUTS • • RUN ANALYSIS

Output files will be saved to

Files / *submission unique ID* / ScatterHaplotypeCallerGVCF / *workflow unique ID*

References to outputs will be written to

Tables / book_samples

Fill in the attributes below to add or update columns in your data table

Download json | Drag or click to upload json

SEARCH OUTPUTS

Task name	Variable	Type	**Attribute** | Use defaults
ScatterHaplotypeCallerGVCF	output_gvcf	File	this.output_gvcf

TABLES

book_samples (3)

book_samples_set (1)

REFERENCE DATA

OTHER DATA

Workspace Data

Files

DOWNLOAD TABLE TSV | COPY TO CLIPBOARD | OPEN WITH...

Search

book_samples_id	input_bam	input_bam_index	output_gvcf
father	father.bam	father.bai	father.merged.g.vcf
mother	mother.bam	mother.bai	mother.merged.g.vcf
son	son.bam	son.bai	son.merged.g.vcf

DASHBOARD DATA NOTEBOOKS WORKFLOWS JOB HISTORY

Back to list

Workflow Statuses
Submitted: 2

Workflow Configuration
dsp-comms-dev/scatter-hc.data-table

Data Entity
scatter-hc-data-table_2019-12-16T22-45-48
book_samples_set

Submitted by
genomics.book@gmail.com
Dec 16, 2019, 5:46 PM

Submission ID
adbaa401-af54-4c58-be37-bf552e69c77a

Total Run Cost
N/A

Call Caching
Enabled

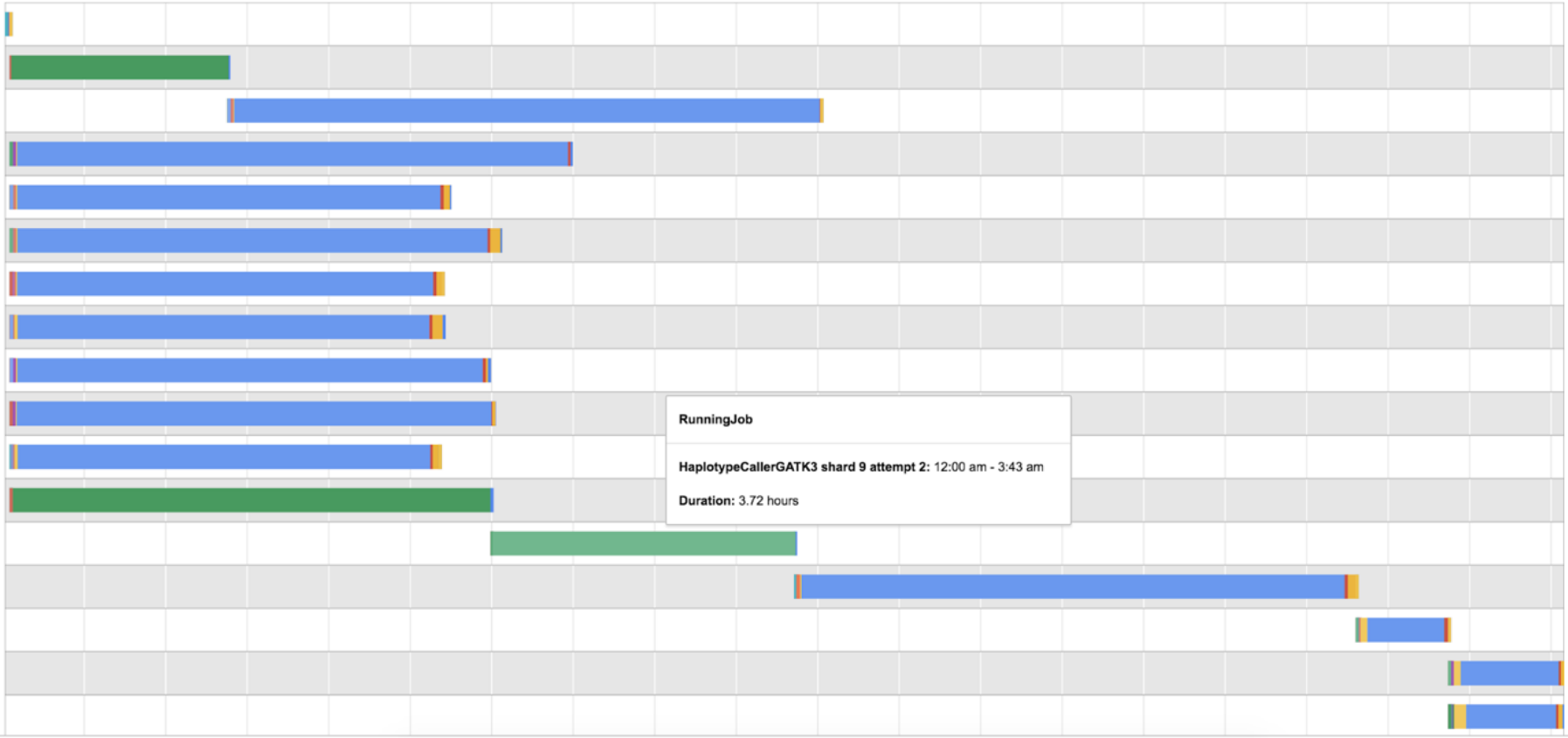
Search | Completion status | Queue Status

	Data Entity	Last Changed	Status	Run Cost	Messages	Workflow ID
View	father (book_samples)	Dec 16, 2019, 5:46 PM	Submitted	N/A		5a8ed577-1593-45f6-bc32-0c494a3dd638
View	mother (book_samples)	Dec 16, 2019, 5:46 PM	Submitted	N/A		b52bf87e-0b15-42fd-9f5d-b34706c5a243

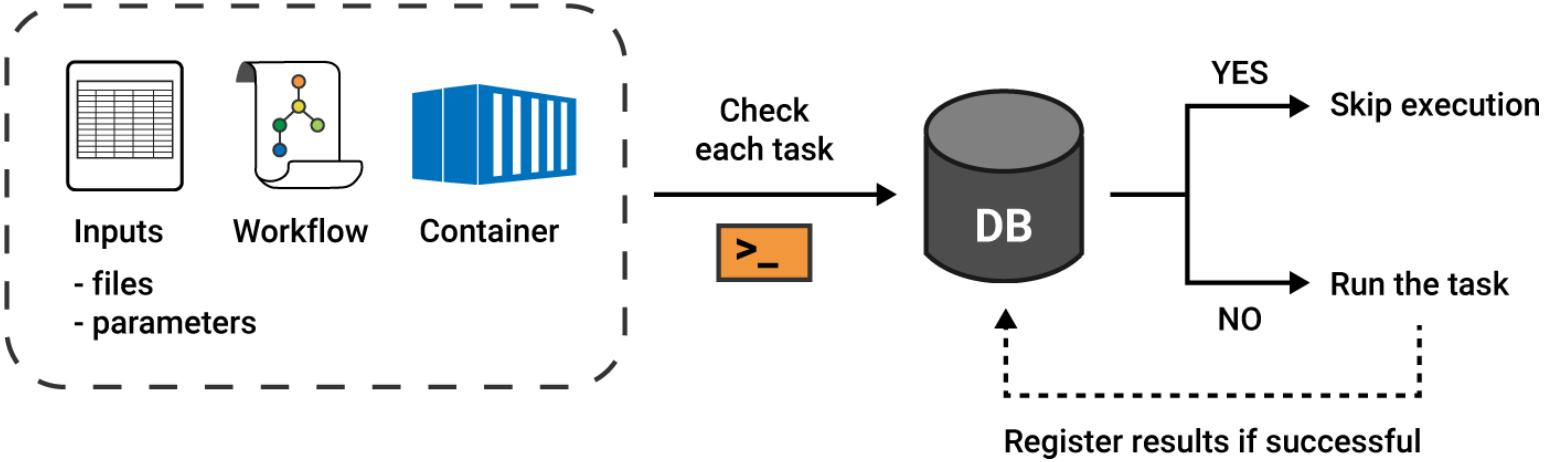
LIST VIEW INPUTS OUTPUTS LABELS TIMING DIAGRAM

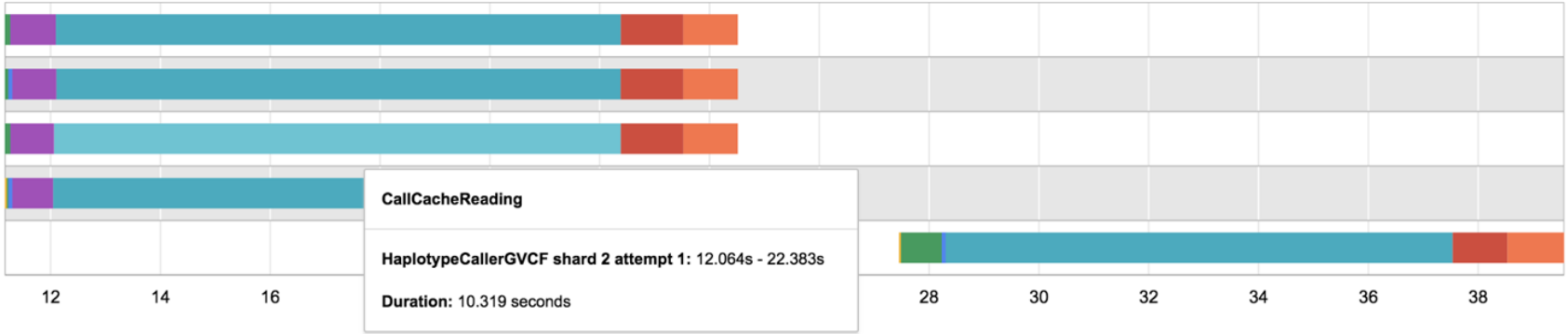
Task Name	Status	Start	Duration	Inputs	Outputs	Links	Attempts
HaplotypeCallerGVCF		Today, 7:13 PM	0h 4m				
MergeVCFs		Today, 7:18 PM	0h 5m				1

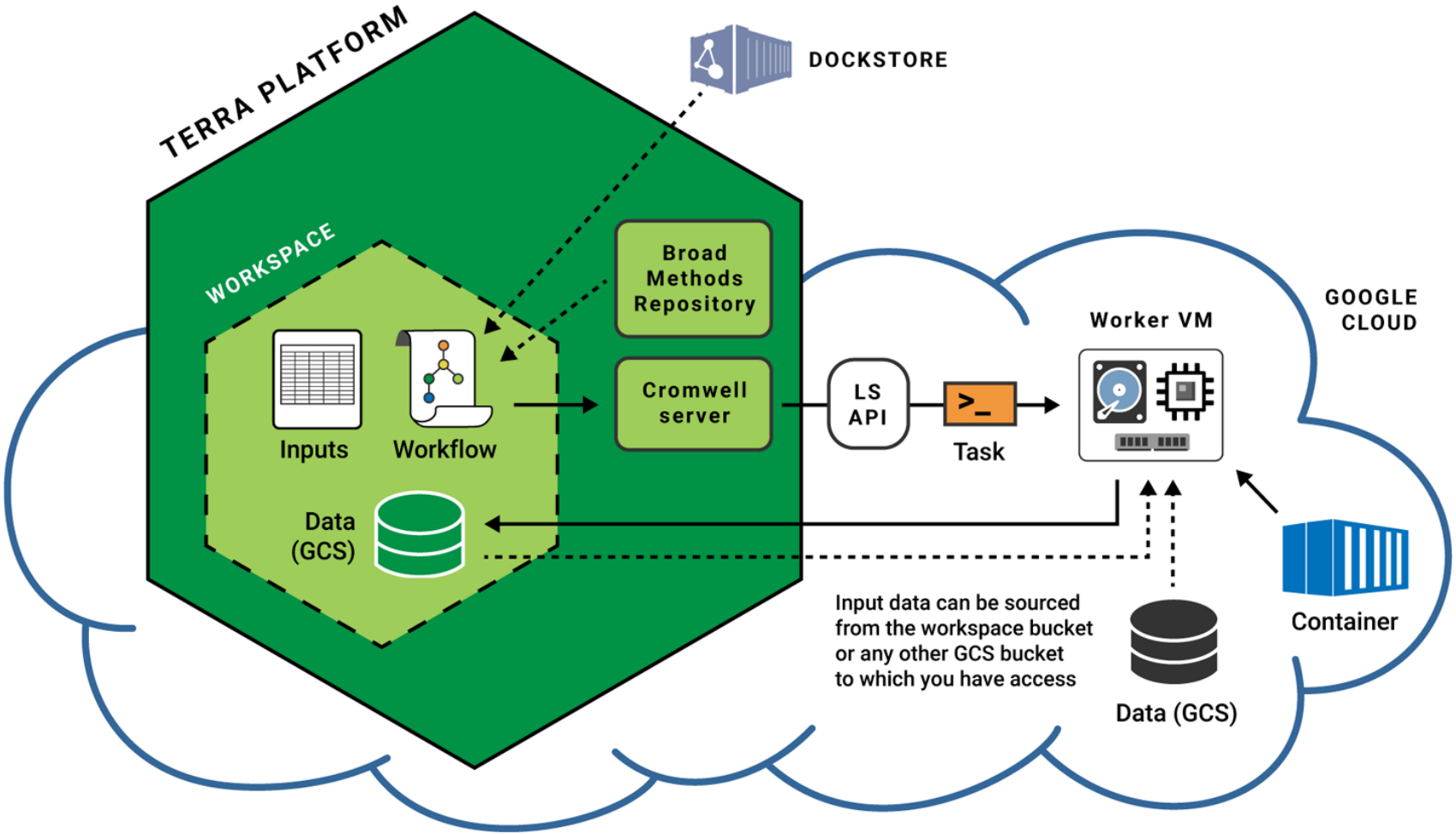
LIST VIEW INPUTS OUTPUTS **TIMING DIAGRAM**



Has this combination been run before?

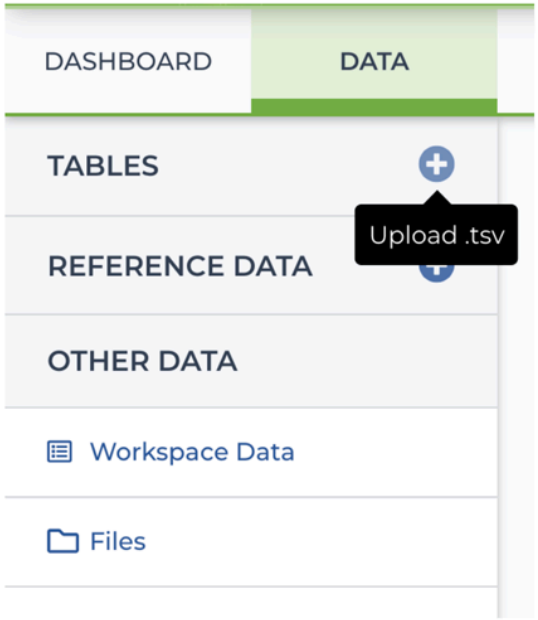




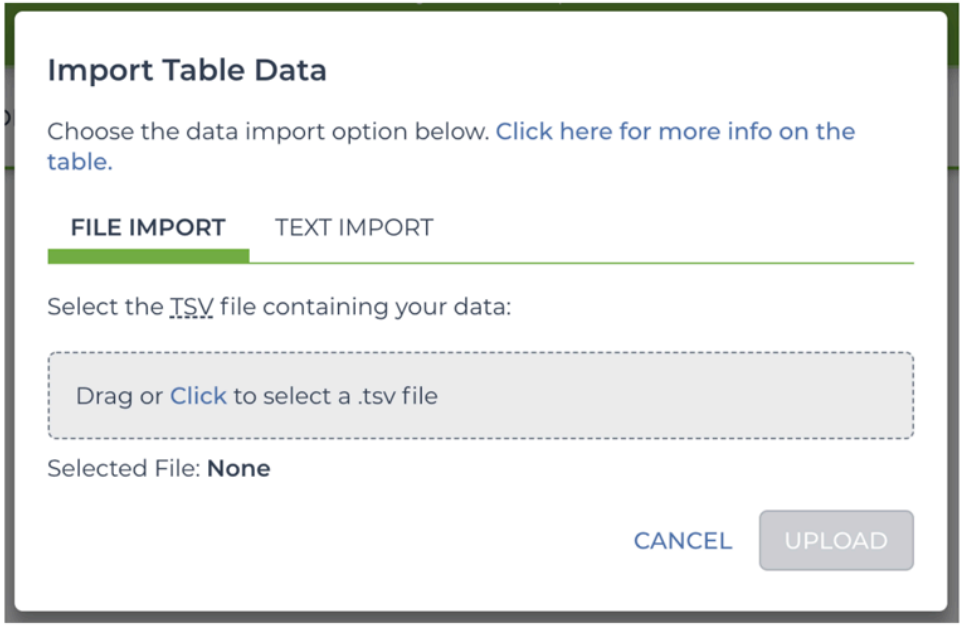

TERRA PLATFORM
DOCKSTORE
WORKSPACE
Inputs
Workflow
Data
(GCS)
Broad
Methods
Repository
Cromwell
server
LS
API
Task
Worker VM
GOOGLE
CLOUD
Container
Input data can be sourced
from the workspace bucket
or any other GCS bucket
to which you have access
Data (GCS)

GETTING DATA

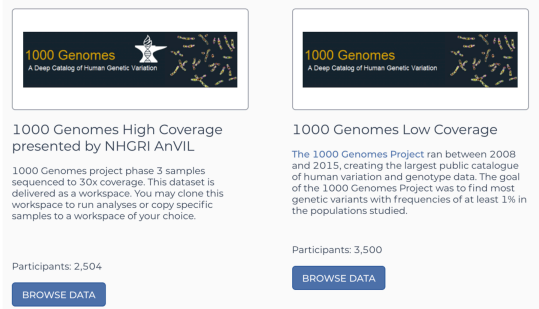
A.

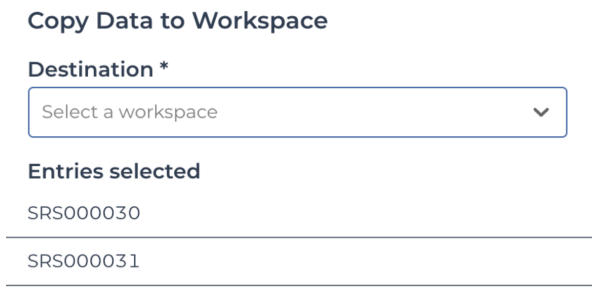


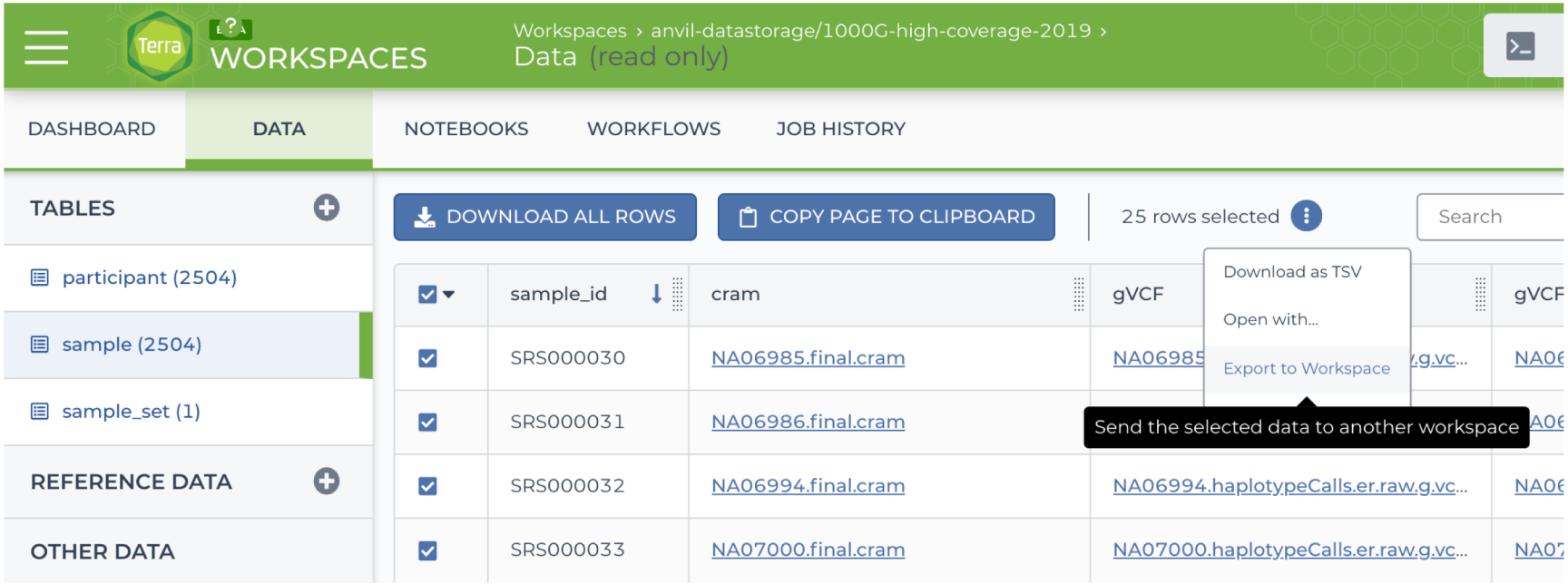
B.



	A	B	C	D	E	F
1	entity:book_samples_id	input_bam	input_bam_index			
2	father	gs://genomics-on-the-cloud/	gs://genomics-on-the-cloud/book-bundle-v0/data/germline/bams/father.bai			
3	mother	gs://genomics-on-the-cloud/	gs://genomics-on-the-cloud/book-bundle-v0/data/germline/bams/mother.bai			
4	son	gs://genomics-on-the-cloud/	gs://genomics-on-the-cloud/book-bundle-v0/data/germline/bams/son.bai			
5						


1000 Genomes
A Deep Catalog of Human Genetic Variation
1000 Genomes High Coverage presented by NHGRI AnVIL
1000 Genomes project phase 3 samples sequenced to 30x coverage. This dataset is delivered as a workspace. You may clone this workspace to run analyses or copy specific samples to a workspace of your choice.
Participants: 2,504
BROWSE DATA
1000 Genomes
A Deep Catalog of Human Genetic Variation
1000 Genomes Low Coverage
The 1000 Genomes Project ran between 2008 and 2015, creating the largest public catalogue of human variation and genotype data. The goal of the 1000 Genomes Project was to find most genetic variants with frequencies of at least 1% in the populations studied.
Participants: 3,500
BROWSE DATA


Copy Data to Workspace
Destination *
Select a workspace
Entries selected
SRS000030
SRS000031


Terra
WORKSPACES
Workspaces › anvil-datastorage/1000G-high-coverage-2019 ›
Data (read only)
DASHBOARD
DATA
NOTEBOOKS
WORKFLOWS
JOB HISTORY
TABLES
participant (2504)
sample (2504)
sample_set (1)
REFERENCE DATA
OTHER DATA
DOWNLOAD ALL ROWS
COPY PAGE TO CLIPBOARD
25 rows selected
Search
Download as TSV
Open with...
Export to Workspace
Send the selected data to another workspace
sample_id
cram
gVCF
SRS000030
NA06985.final.cram
SRS000031
NA06986.final.cram
SRS000032
NA06994.final.cram
NA06994.haplotypeCalls.er.raw.g.vc...
SRS000033
NA07000.final.cram
NA07000.haplotypeCalls.er.raw.g.vc...

broad-gdr-encode.appspot.com

ENCODE

6 Donors | 201 Files

EXPORT

Reference Genome Assembly=GRCh38 ×

Assay Type=ATAC-seq ×

Send to Terra

Donor Accession

Search...

- [ ] ENCDO000AAB
- [ ] ENCDO000AAC
- [ ] ENCDO000AAD
- [ ] ENCDO000AAE
- [ ] ENCDO000AAF
- [ ] ENCDO000AAG

Age

0 — 150

Age Units

Search...

- [ ] day
- [ ] month
- [ ] week
- [ ] year

Health Status

Search...

- [ ] abdominal sarcoma
- [ ] acute promyelocytic leukemia
- [ ] acute T cell leukemia
- [ ] Adenocarcinoma of cecum
- [ ] apparently healthy
- [ ] B cell lymphoma

Sex

Search...

- [ ] female
- [ ] male
- [ ] mixed

Assay Type

Search...

- [x] ATAC-seq
- [ ] ChIA-PET
- [ ] ChIP-seq
- [ ] DNase-seq
- [ ] Hi-C
- [ ] microRNA-seq

Terra BETA IMPORT DATA

Importing Data

From: broad-gdr-encode.appspot.com

The dataset(s) you just chose to import to Terra will be made available to you within a workspace of your choice where you can then perform analysis.

Destination Workspace

Choose the option below that best suits your needs.

Start with an existing workspace
Select one of your workspaces

Start with a new workspace
Set up an empty workspace that you will configure for analysis

Terra BETA WORKSPACES

Workspaces › dsp-comms-dev/TEST-ANVIL-1234 ›
Data

Notebook Runtime
RUNNING ($0.76 hr)

DASHBOARD | DATA | NOTEBOOKS | WORKFLOWS | JOB HISTORY

TABLES

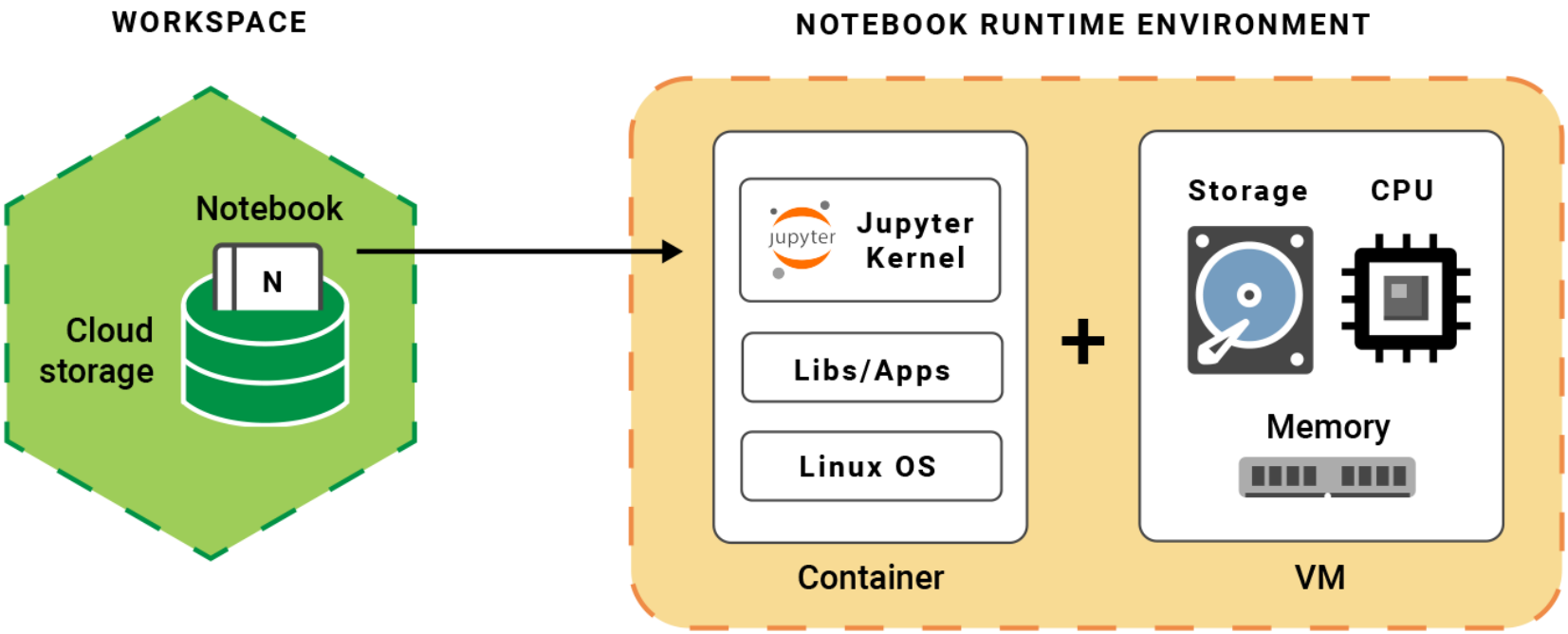
- donor (6)
- donor_set (1)
- file (201)
- file_set (1)

DOWNLOAD ALL ROWS | COPY PAGE TO CLIPBOARD | 0 rows selected | Search

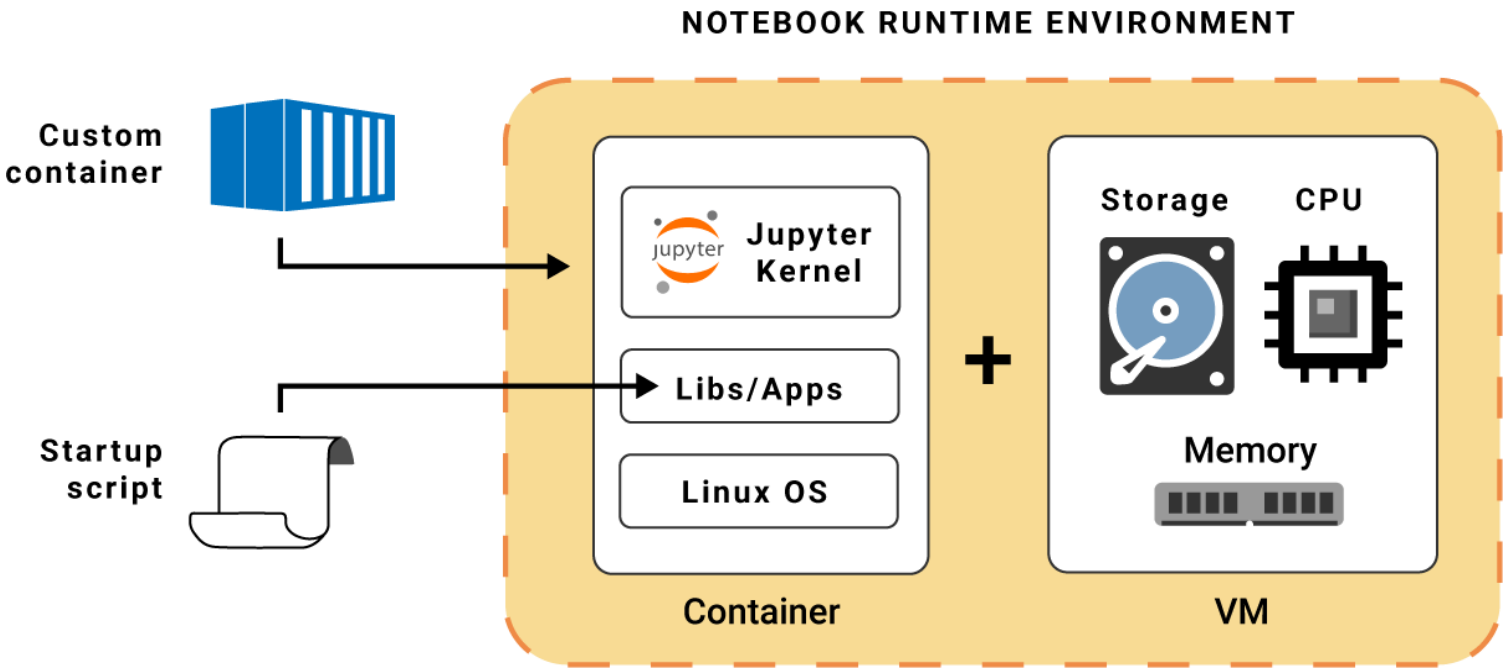
file_id	assay_type	biosamples	biosample_term_id
ENCFF013HFS	ATAC-seq	1 item	EFO:0001086
ENCFF015FKA	ATAC-seq	1 item	UBERON:0001150
ENCFF020COS	ATAC-seq	1 item	EFO:0001086

INTERACTIVE ANALYSIS

(Jupyter Notebooks)



Interactive work is done in a runtime environment



Options for customizing the runtime environment

RUNTIME CONFIGURATION

Choose a Terra pre-installed runtime environment (e.g. programming languages + packages) or choose a custom environment

PRE-INSTALLED ENVIRONMENT | CUSTOM ENVIRONMENT

Environment: Default (Python 3.6.8, R 3.5.2, Hail 0.2.11)

What's installed on this environment?

Updated: Aug 25, 2019
Version: FINAL

COMPUTE POWER

Select from one of the compute runtime profiles or define your own

Profile: Default (Moderate) computer power

CPUs: 4 | Memory (GB): 15 | Disk size (GB): 50

Cost: $0.19 per hour

CANCEL | CREATE

INSTALLED PACKAGES

Default (Python 3.6.8, R 3.5.2, Hail 0.2.11)

Updated: Aug 25, 2019
Version: FINAL

Installed packages Python

- Python
- R
- Tools

Package	Version
python	3.6.8
hail	0.2
wrapt	1.11.2
absl-py	0.7.1
arrow	0.14.5

COMPUTE POWER

Select from one of the runtime profiles or define your own

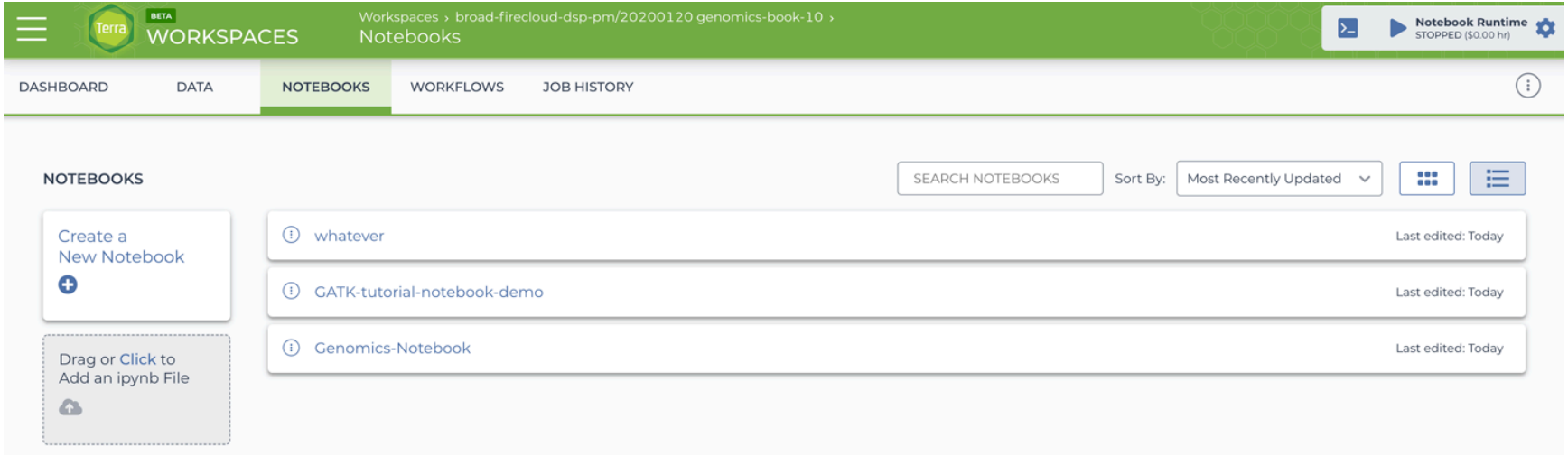
Profile: Custom

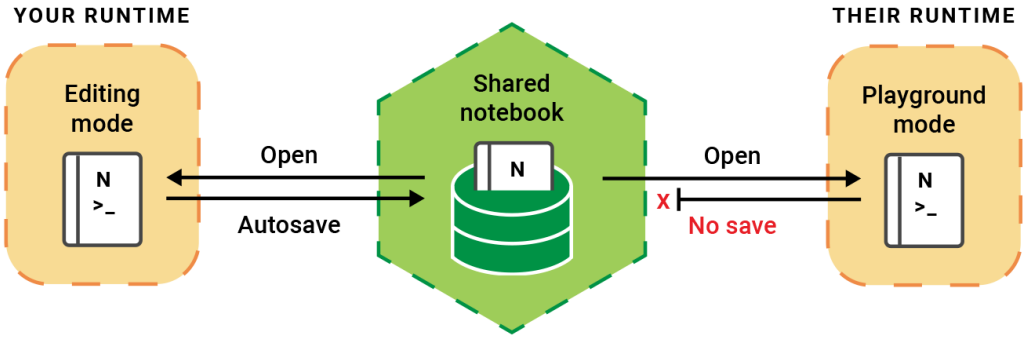
CPUs: 4 | Memory (GB): 15 | Disk size (GB): 50

Startup script: gs://genomics-on-the-cloud/book-bundle-v0/code/notebooks/insta

☐ Configure as Spark cluster

COST: $0.19 per hour


Terra
BETA
WORKSPACES
Workspaces › broad-firecloud-dsp-pm/20200120 genomics-book-10 ›
Notebooks
Notebook Runtime
STOPPED ($0.00 hr)
DASHBOARD
DATA
NOTEBOOKS
WORKFLOWS
JOB HISTORY
NOTEBOOKS
SEARCH NOTEBOOKS
Sort By:
Most Recently Updated
Create a New Notebook
Drag or Click to Add an ipynb File
whatever
Last edited: Today
GATK-tutorial-notebook-demo
Last edited: Today
Genomics-Notebook
Last edited: Today


YOUR RUNTIME
THEIR RUNTIME
Editing mode
N
>_
Open
Autosave
Shared notebook
N
Open
X
No save
Playground mode
N
>_

Terra WORKSPACES

Workspaces › dsp-comms-dev/genomics-book-10 › notebooks › Genomics-Notebook.ipynb

Notebook Runtime RUNNING ($0.76 hr)

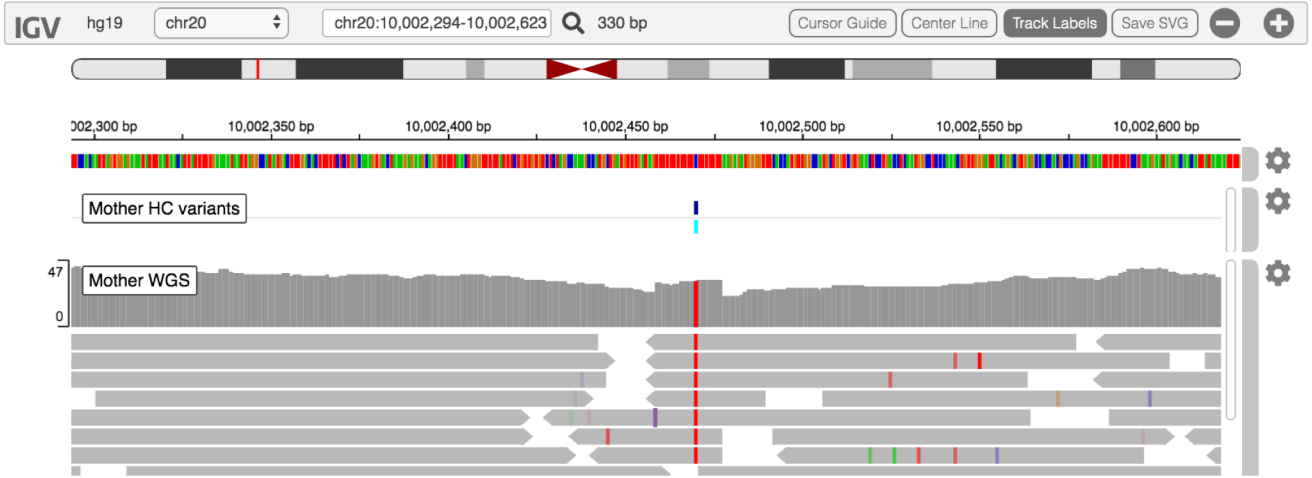
jupyter Genomics-Notebook (autosaved)

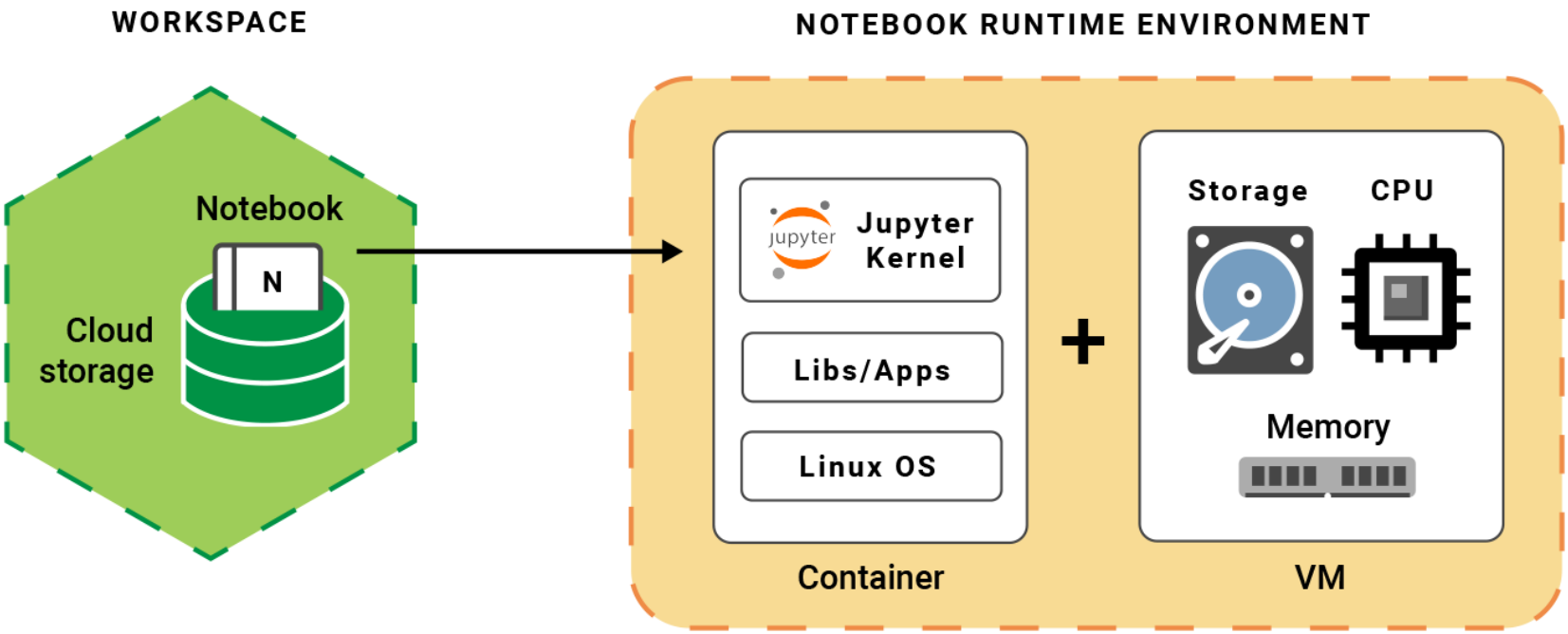
File Edit View Insert Cell Kernel Navigate Widgets Help | Not Trusted | Edit Mode | Python 3

Run | Markdown

In [21]:
```
# Create an interactive IGV browser with genome reference and coordinates specified.
IGV_InspectCalls = igv.Browser(
    {"genome": "hg19",
     "locus": "chr20:10,002,294-10,002,623"
    }
)

# Tell Python to display it below
IGV_InspectCalls.show()
```





Coming next: other apps for interactive analysis

Connecting scientists to the patients, datasets, and tools they need to do life-changing research