

Seven Bridges - Cancer Genomics Cloud

Tiffany Delhomme

Bioinfo meeting

10 oct 2017

NCI-supported Projects

Providing big data resources to accelerate cancer research

Hinkson et al.

Big Data in Cancer Research

TABLE 1 | Selected NCI-supported projects.

Project name	Lead institution(s)	Project URL
The Cancer Genome Atlas (TCGA)	National Cancer Institute National Human Genome Research Institute	cancergenome.nih.gov
Therapeutically Applicable Research to Generate Effective Treatments (TARGET)	NCI Office of Cancer Genomics	ocg.cancer.gov/programs/target
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	NCI Office of Cancer Clinical Proteomics Research	proteomics.cancer.gov/programs/cptac
Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) Network	Department of Defense Department of Veterans Affairs National Cancer Institute	proteomics.cancer.gov/programs/apollo-network
The Cancer Imaging Archive (TCIA)	University of Arkansas for Medical Sciences NCI Division of Cancer Treatment and Diagnosis	www.cancerimagingarchive.net
Genomic Data Commons (GDC)	NCI Center for Cancer Genomics	gdc.cancer.gov
Database of Genotypes and Phenotypes (dbGaP)	National Center for Biotechnology Information	www.ncbi.nlm.nih.gov/gap
NCI Cloud Resources	National Cancer Institute	cbit.cancer.gov/cloudresources
Broad FireCloud	Broad Institute	firecloud.org
Institute for Systems Biology Cancer Genomics Cloud (ISB-CGC)	Institute for Systems Biology	isb-cgc.org
Seven Bridges Cancer Genomics Cloud (SB-CGC)	Seven Bridges	www.cancergenomicscloud.org
NCI Cancer Research Data Commons	National Cancer Institute	cbit.cancer.gov/cancerdatacommons

NCI-supported projects annotated with lead institutions and URLs.

NCI-supported Projects

Providing big data resources to accelerate cancer research

Hinkson et al.

Big Data in Cancer Research

TABLE 1 | Selected NCI-supported projects.

Project name	Lead institution(s)	Project URL
The Cancer Genome Atlas (TCGA)	National Cancer Institute National Human Genome Research Institute	cancergenome.nih.gov
Therapeutically Applicable Research to Generate Effective Treatments (TARGET)	NCI Office of Cancer Genomics	ocg.cancer.gov/programs/target
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	NCI Office of Cancer Clinical Proteomics Research	proteomics.cancer.gov/programs/cptac
Applied Proteogenomics Organizational Learning and Outcomes (APOLLO) Network	Department of Defense Department of Veterans Affairs National Cancer Institute	proteomics.cancer.gov/programs/apollo-network
The Cancer Imaging Archive (TCIA)	University of Arkansas for Medical Sciences NCI Division of Cancer Treatment and Diagnosis	www.cancerimagingarchive.net
Genomic Data Commons (GDC)	NCI Center for Cancer Genomics	gdc.cancer.gov
Database of Genotypes and Phenotypes (dbGaP)	National Center for Biotechnology Information	www.ncbi.nlm.nih.gov/gap
NCI Cloud Resources	National Cancer Institute	cbit.cancer.gov/cloudresources
Broad FireCloud	Broad Institute	firecloud.org
Institute for Systems Biology Cancer Genomics Cloud (ISB-CGC)	Institute for Systems Biology	isb-cgc.org
Seven Bridges Cancer Genomics Cloud (SB-CGC)	Seven Bridges	www.cancergenomicscloud.org
NCI Cancer Research Data Commons	National Cancer Institute	cbit.cancer.gov/cancerdatacommons

NCI-supported projects annotated with lead institutions and URLs.

NCI Cancer Research Data Commons

A cloud-based network



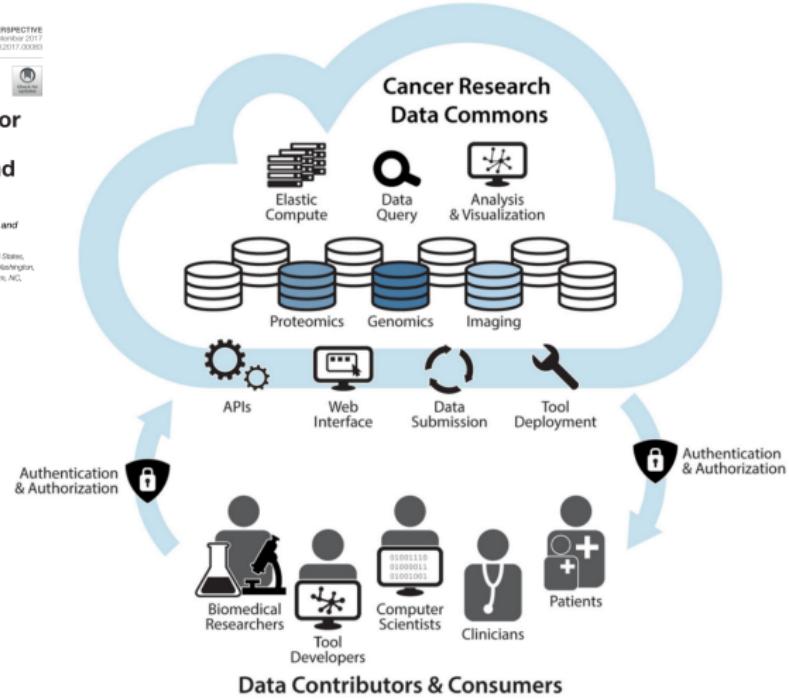
in Cell and Developmental Biology

PERSPECTIVE
published: 21 September 2017
doi: 10.3389/fcell.2017.00083

A Comprehensive Infrastructure for Big Data in Cancer Research: Accelerating Cancer Research and Precision Medicine

Izumi V. Hinikeri^{1,2}, Tanja M. Davidson¹, Juli D. Klemm¹, Anthony R. Kerlavage^{1,3} and Warren A. Kibbe^{1,2}

¹ Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD, United States; ² Science and Technology Policy Fellowship Program, American Association for the Advancement of Sciences, Washington, DC, United States; ³ Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, United States



International Agency for Research on Cancer

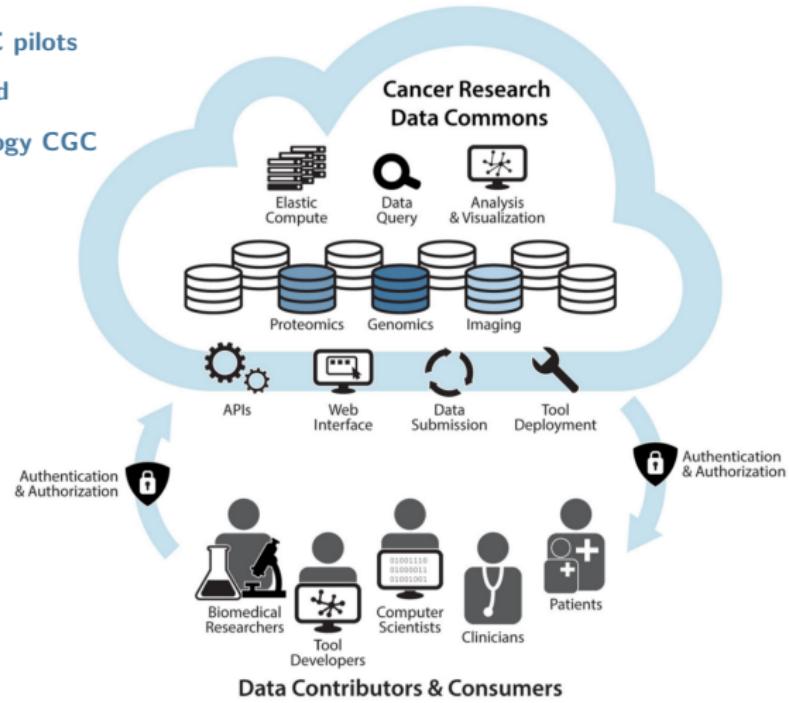


NCI Cancer Research Data Commons

A cloud-based network

3 NCI cloud resources : the CGC pilots

- Broad Institute's FireCloud
- Institute for Systems Biology CGC
- SevenBridges CGC

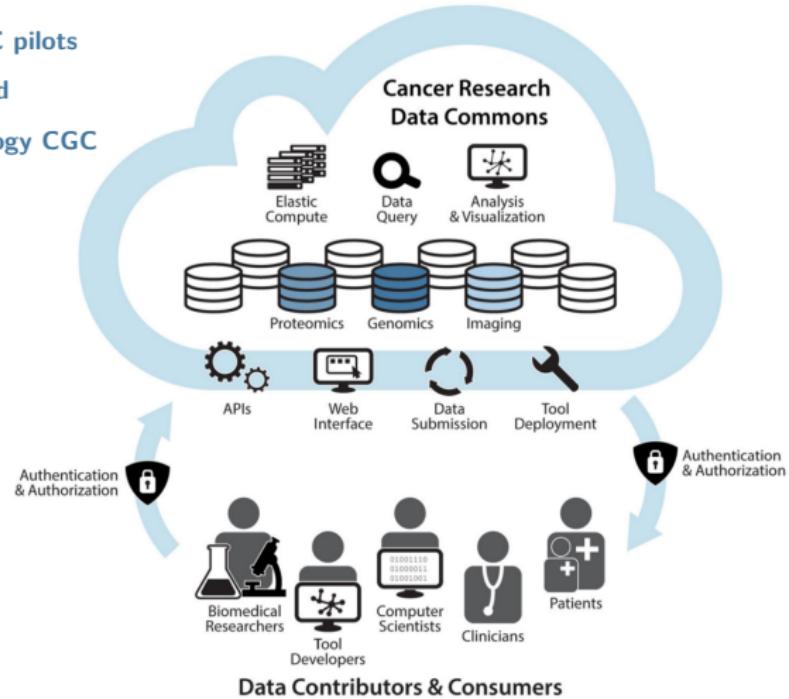


NCI Cancer Research Data Commons

A cloud-based network

3 NCI cloud resources : the CGC pilots

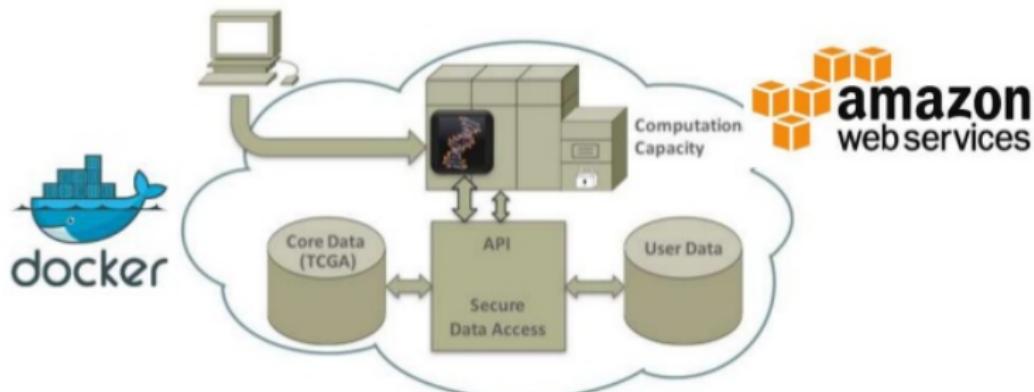
- Broad Institute's FireCloud
- Institute for Systems Biology CGC
- SevenBridges CGC



CGC cloud concept

Store and analyse in the same place

Cloud Pilots Concept: Co-located Compute & Data



Democratize access to NCI-generated genomic & related data
Provide cost-effective computational capacity for the cancer research community

NATIONAL CANCER INSTITUTE

Slide courtesy of Ishwar Chandramouliswaran, NCI

CGC available instances AWS instances pricing

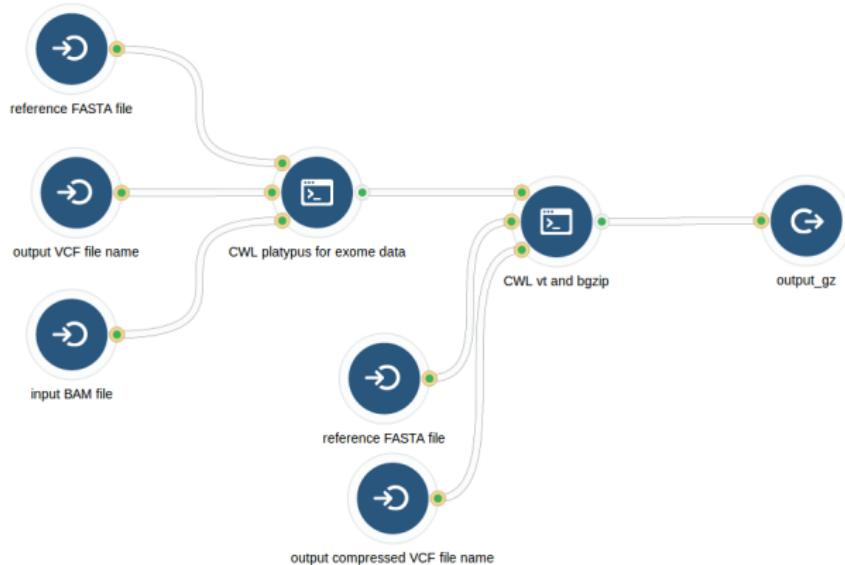
International Agency for Research on Cancer

World Health Organization

run analyses on the cloud

describe tools and workflows, get data, run tasks

- How to describe your app?
 - a **workflow** is composed of 1 or more **tools**, both are **apps**
 - a tool is described with: **inputs, outputs, parameters, requirements**



run analyses on the cloud

describe tools and workflows, get data, run tasks

- How to describe your app?

- a **workflow** is composed of 1 or more **tools**, both are **apps**
- a tool is described with: **inputs, outputs, parameters, requirements**

- How to get data to analyse?

- upload data
 - GUI , APIs, CancerGenomicsCloudUploader
- querying metadata
 - GUI, APIs

run analyses on the cloud

describe tools and workflows, get data, run tasks

- How to describe your app?

- a **workflow** is composed of 1 or more **tools**, both are **apps**
- a tool is described with: **inputs, outputs, parameters, requirements**

- How to get data to analyse?

- upload data
 - GUI , APIs, CancerGenomicsCloudUploader
- querying metadata
 - GUI, APIs

- How to run analyses?

- GUI, APIs (automatic way)
- task monitoring

International Agency for Research on Cancer



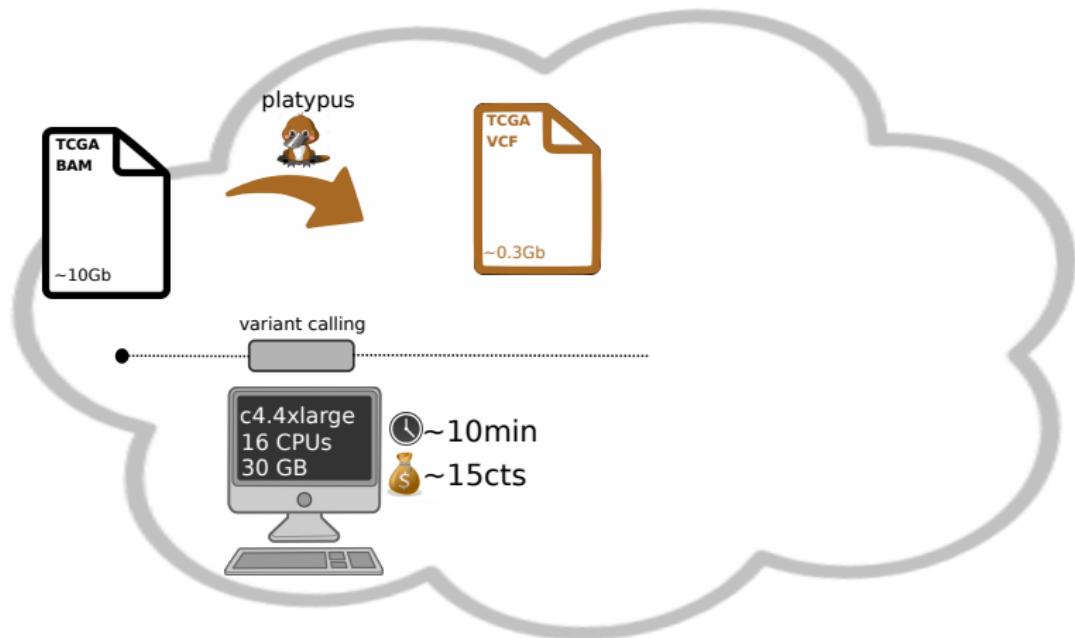
Application Programming Interface (API)

- sevenbridges APIs integrate the platform with other applications (such as R)
- automate procedures
 - uploading files
 - querying metadata
 - describe tools and workflows
 - executing analyses
- 3 API clients: R, Python, JAVA

sevenbridges R package

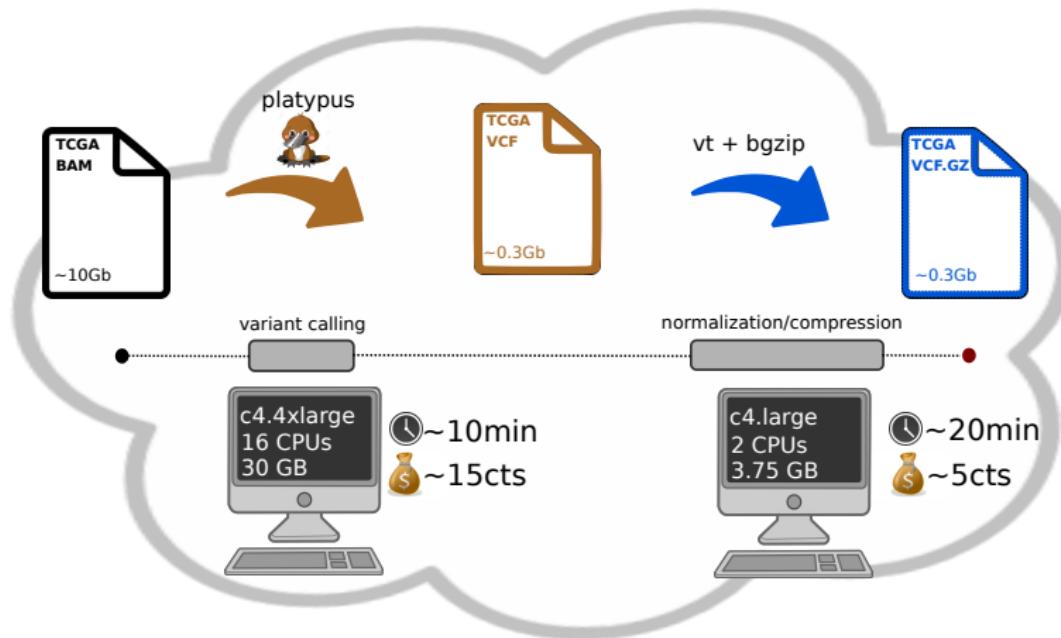
example of cloud analysis

TCGA germline variant calling with platypus



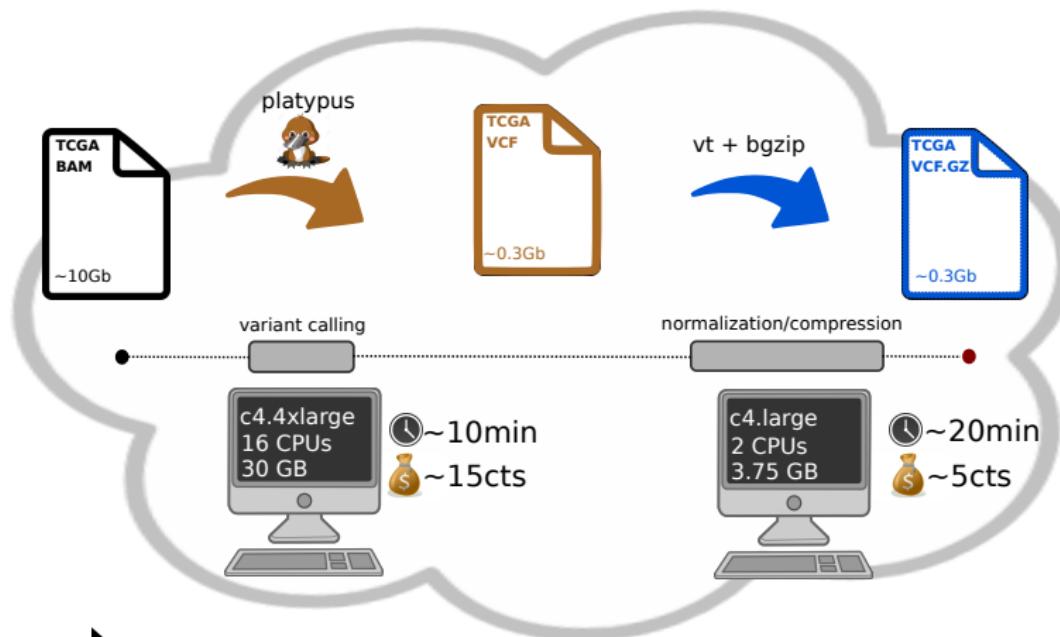
example of cloud analysis

TCGA germline variant calling with platypus



example of cloud analysis

TCGA germline variant calling with platypus



→ germline variant calling on lung TCGA (~1000 samples):
200 \$ (+ 6\$ storage/month)
5 hours

International Agency for Research on Cancer



Let's play

GUI demo