Cancer Genome COLLABORATORY

Cloud computing for collaborative research

Vincent Ferretti, PhD

Chicago

June 7th, 2016

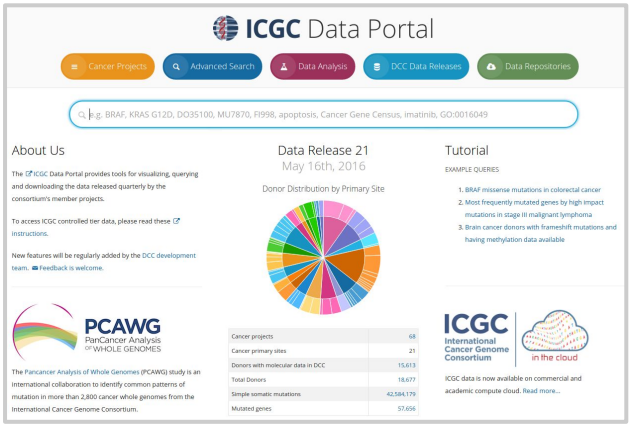
International Cancer Genome Consortium

Background

- Goal is to sequence 25,000 tumor genomes from 50 tumor types and subtypes by end of 2018

- Release 21 (May 2016)
  - 15,613 donors with molecular data
  - 68 cancer projects including 24 from TCGA
  - > 4,000 WGS; 7,406 tumor WGX, 8,766 tumor RNA-Seq

- ICGC Data Portal at *dcc.icgc.org*
  - Search, visualisation & analysis tools
  - Data download
    - Mutation calls, CNV, Meth., Expr.

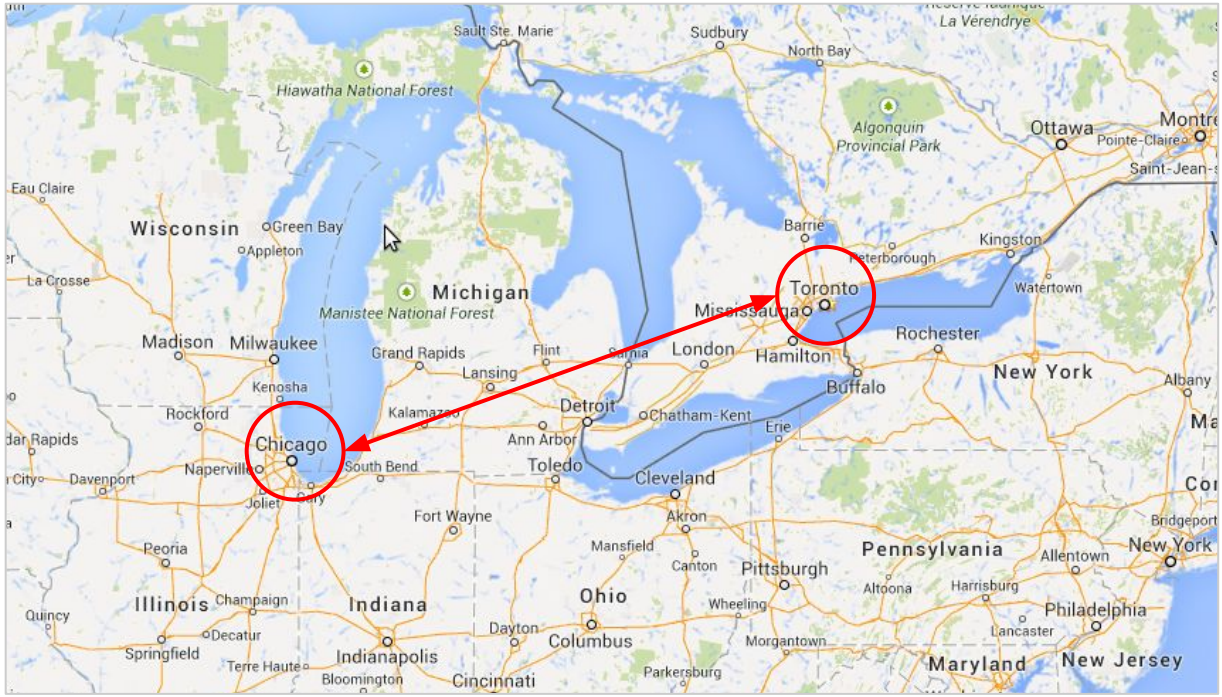


- Raw data stored at EGA, cgHub & GDC

The Cancer Genome Collaboratory

- A new academic research **compute** cloud resource hosting the ICGC *raw sequencing* data

- Two Physical Data Centres
  - **SciNet** - Compute Canada, **Toronto** -> ICGC but TCGA
  - **BioNimbus** Protected Data Cloud (PDC), **Chicago** -> TCGA



Presentation Outline

- Hardware infrastructure at Scinet-Toronto
- Overview of the software architecture
- Our “FAIR” solution for ICGC data
  - The ICGC Data Portal indexing and searching tools
  - A Universal data download for better ICGC clouds/repositories interoperability
- Further work

Infrastructure (SciNet-OICR)

Monitoring platform

ZABBIX

High performance networking

20 Gbps intra-rack
240 Gbps (6x40) inter-rack

OpenStack Controllers

3 controllers (for quorum)

OpenStack VMs

4 compute servers per node

R1	1 Node
R2	4 Nodes
R3	4 Nodes
R4	3 Nodes
R5	3 Nodes
R6	3 Nodes

Ceph Storage

144 to 288 TB per node

R1	7 Nodes
R2	8 Nodes
R3	6 Nodes
R4	8 Nodes
R5	8 Nodes
R6	8 Nodes

Next 2 Purchases

	Currently	Sept 2016	March 2017
Storage (PB)	3.26	4.1	6.7
Compute Cores*	1152	2592	2592
Compute RAM (TB)	8.9	17.9	17.9

** 2 cores per server are reserved for infrastructure operation*

OpenStack - Cloud System

Six different virtual machine “flavours” users can launch

Name	Logical CPU	RAM (GB)	Local Storage (GB)	Max # of instances (current)	Max # of Instances (March 2017)
c1.micro	1	8	162	1080	2448
c1.small	2	16	325	540	1224
c1.medium	4	32	650	252 + 36 small	576 + 72 small
c1.large	8	56	1,300	108 + 36 medium + 36 small	252 + 72 small + 72 medium
c1.xlarge	15	125	2,600	72	144 + 36 large
c1.xxlarge	30	250	5,200	36	72 + 36 large

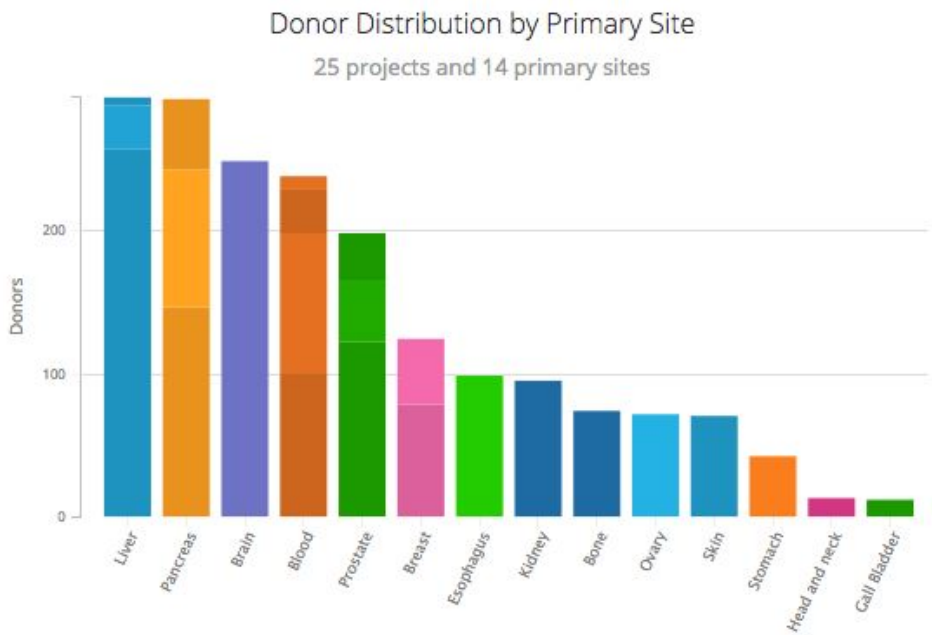
Data Available (as of June 1st)

PCAWG latest data release

- 42,549 files, 478 TB
- 25 projects, 14 primary sites
- 1907 ICGC donors
- 3940 WGS BAMs
- 4 variant sets (DKFZ, Sanger, Broad, MUSE)

Coming soon

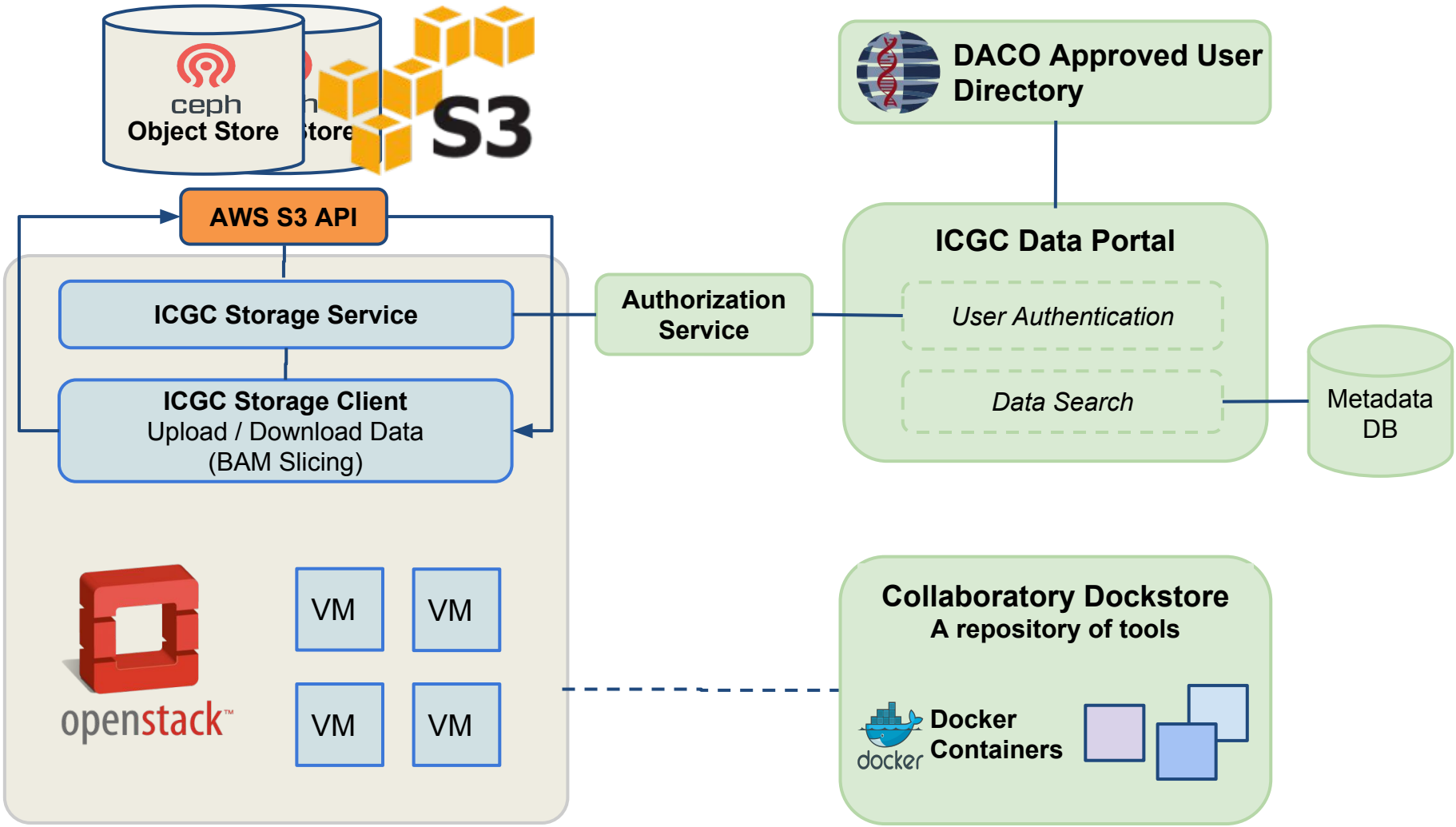
- PCAWG Merged variant calls
- PCAWG mini BAMs
- PCAWG RNA-Seq BAMs



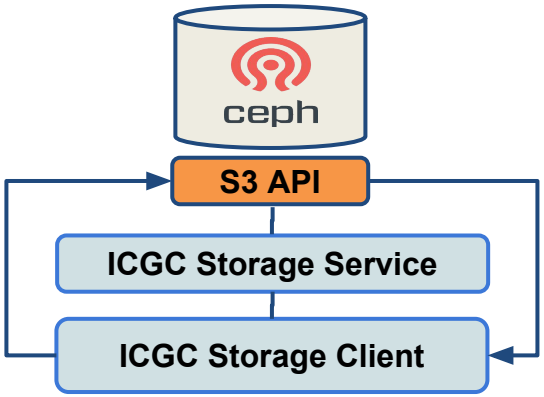
1,907 Donors **42,549** Files **477.81 TB**

Data Type	# Donors	# Files	Format	Size
SGV	1,907	6,101	VCF	344.22 GB
StGV	1,907	4,066	VCF	5.32 GB
Aligned Reads	1,907	3,940	BAM	477.35 TB
Simple Somatic Mutations	1,907	14,233	VCF	123.77 GB
Copy Number Somatic Mutations	1,907	4,066	VCF	85.26 MB
Structural Somatic Mutations	1,907	10,143	VCF	668.49 MB

System Overview



The *Storage* Module



- A client-server application for both uploading and downloading data

- Core features
  - Support for encrypted and authorized transfers
  - High-throughput: multi-part parallel upload/download
  - Resumable
  - MD5 checksum validation

- Download-specific features
  - Support for BAM slicing
  - Support for Filesystem in Userspace (FUSE)

ICGC-Storage-Client

Download manifest data (text files or manifest IDs)

```
%: icgc-storage-client download --manifest 4jdyyqs099ew22
--output-dir data --output-layout bundle
```

Download BAM slices (individual files or manifest)

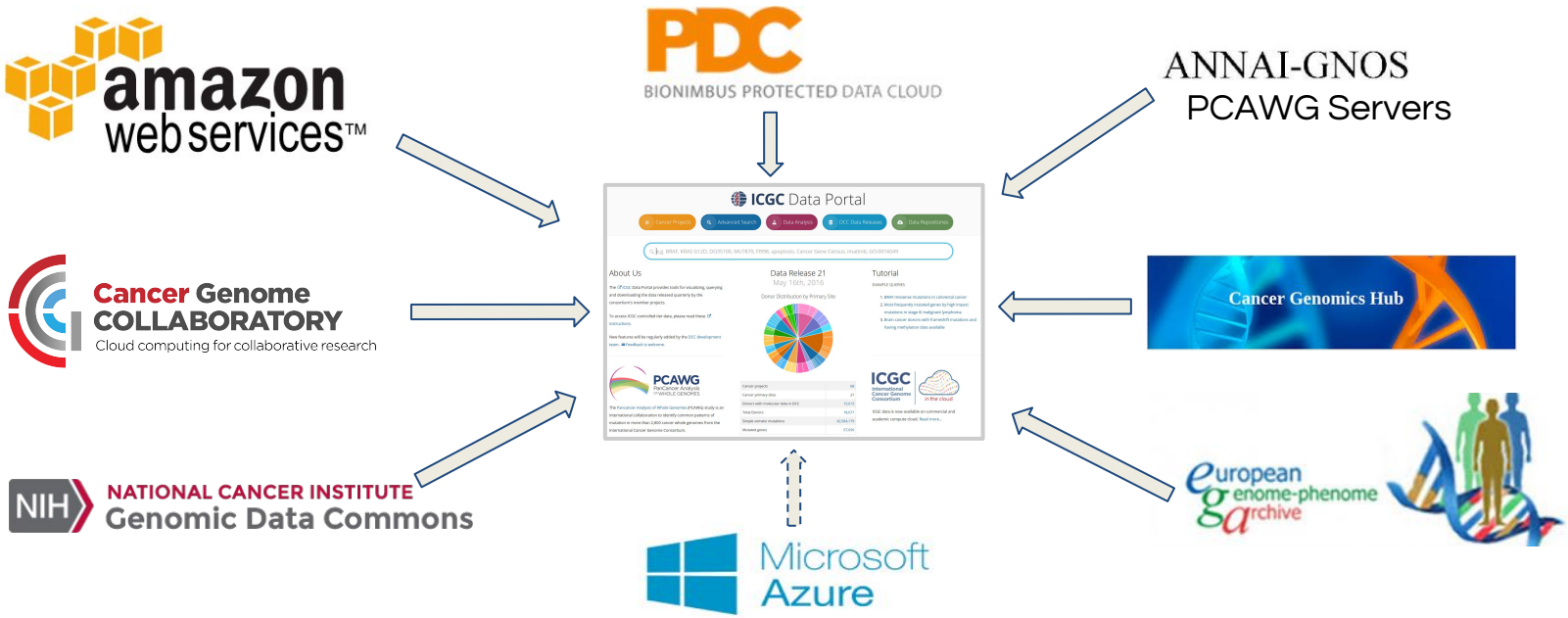
```
%: icgc-storage-client view --object-id ea17647-17f6-5ae0
--query 12:25357723-25403870
```

Mounting a manifest (FUSE)

```
%: icgc-storage-client mount --manifest 4jdyyqs099ew22
--mount-point /tmp/ --cache-metadata
%: ls /tmp
%: samtool view /tmp/<fileName.bam> 1:10000-20000
```

Data Searching

- ICGC data hosted on compute clouds/repos worldwide are nightly linked to ICGC donors, aggregated and indexed
- The ICGC Data Portal provides intuitive web interfaces for searching donors and associated raw data



ICGC Data Portal

Quick Search

DATA REPOSITORIES

Files | Donors

Share | Repository IS Collaboratory - Toronto

Download manifests | 7,172 Files | 1,190 Donors | 313.00 TB

SAVE DONOR SET

Showing **1** - **25** of **7,172** files

- File
  - Enter File ID, Name or Object ID
- Repository
  - Collaboratory - ... 7,172
  - AWS - Virginia 7,309
- Experimental Strategy
  - WGS 7,172
- Only Files in Study
  - PCAWG 7,172
- File Format
  - VCF 4,704
  - BAM 2,468
- Analysis Software
  - Sanger variant call p... 4,704
  - BWA MEM 2,468

Showing **1** - **25** of **7,172** files

	File ID	Donor	Repository	Project	Study	Data Type	Strategy	Format	Size
☐	FI9981	DO46390	PCAWG - Chicago (ICGC), Collaboratory, AWS - Virginia	OV-AU	PCAWG	SSM	WGS	VCF	47.19 MB
☐	FI998	DO52556	PCAWG - Barcelona, Collaboratory, AWS - Virginia	BRCA-UK	PCAWG	CNSM	WGS	VCF	4.19 KB
☐	FI9979	DO46390	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	OV-AU	PCAWG	StSM	WGS	VCF	48.95 KB
☐	FI9977	DO46390	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	OV-AU	PCAWG	CNSM	WGS	VCF	4.16 KB
☐	FI9975	DO46390	PCAWG - Chicago (ICGC), Collaboratory, AWS - Virginia	OV-AU	PCAWG	SSM	WGS	VCF	2.66 MB
☐	FI9916	DO47108	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	174.53 GB
☐	FI9915	DO47108	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	229.06 GB
☐	FI9883	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	SSM	WGS	VCF	58.19 MB
☐	FI9881	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	StSM	WGS	VCF	20.91 KB
☐	FI9879	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	CNSM	WGS	VCF	2.33 KB
☐	FI9877	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	SSM	WGS	VCF	7.09 MB
☐	FI9876	DO49113	PCAWG - Barcelona, PCAWG - Chicago (ICGC), AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	188.71 GB
☐	FI9875	DO49113	PCAWG - Barcelona, PCAWG - Chicago (ICGC), AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	80.17 GB
☐	FI9858	DO46933	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	153.97 GB
☐	FI9857	DO46933	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	128.29 GB

ICGC Data Portal

Quick Search

DATA REPOSITORIES

Files | Donors

Share | Repository IS Collaboratory - Toronto

Donor

e.g. DO45299, SA501608

Upload Donor Set

Project

LIRI-JP	1,595
PACA-CA	814
PRAD-CA	658
PAEN-AU	626
PACA-AU	582

12 more

Primary Site

Pancreas	2,022
Liver	1,709
Prostate	911
Bone	452
Kidney	444

6 more

Specimen Type

Primary tumour - so...	5,181
Normal - blood deri...	873
Primary tumour - bl...	273
Normal - solid tissue	235
Metastatic tumour - ...	153

10 more

Only Donors in Study

PCAWG 7,172

Download manifests | 7,172 Files | 1,190 Donors | 313.00 TB

SAVE DONOR SET

Showing **1** - **25** of **7,172** files

File ID	Donor	Repository	Project	Study	Data Type	Strategy	Format	Size
FI9981	DO46390	PCAWG - Chicago (ICGC), Collaboratory, AWS - Virginia	OV-AU	PCAWG	SSM	WGS	VCF	47.19 MB
FI998	DO52556	PCAWG - Barcelona, Collaboratory, AWS - Virginia	BRCA-UK	PCAWG	CNSM	WGS	VCF	4.19 KB
FI9979	DO46390	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	OV-AU	PCAWG	StSM	WGS	VCF	48.95 KB
FI9977	DO46390	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	OV-AU	PCAWG	CNSM	WGS	VCF	4.16 KB
FI9975	DO46390	PCAWG - Chicago (ICGC), Collaboratory, AWS - Virginia	OV-AU	PCAWG	SSM	WGS	VCF	2.66 MB
FI9974	DO46390	PCAWG - Barcelona, PCAWG - Chicago (ICGC), PCAWG - London, AWS - Virginia	OV-AU	PCAWG	Aligned Reads	WGS	BAM	124.84 GB
FI9973	DO46390	PCAWG - Barcelona, PCAWG - Chicago (ICGC), PCAWG - London, AWS - Virginia	OV-AU	PCAWG	Aligned Reads	WGS	BAM	94.48 GB
FI996	DO52556	PCAWG - Barcelona, Collaboratory, AWS - Virginia	BRCA-UK	PCAWG	SSM	WGS	VCF	5.92 MB
FI995	DO52556	PCAWG - Barcelona, AWS - Virginia	BRCA-UK	PCAWG	Aligned Reads	WGS	BAM	124.22 GB
FI994	DO52556	PCAWG - Barcelona, AWS - Virginia	BRCA-UK	PCAWG	Aligned Reads	WGS	BAM	98.96 GB
FI9916	DO47108	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	174.53 GB
FI9915	DO47108	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	229.06 GB
FI9883	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	SSM	WGS	VCF	58.19 MB
FI9881	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	StSM	WGS	VCF	20.91 KB
FI9879	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	CNSM	WGS	VCF	2.33 KB
FI9877	DO49113	PCAWG - Chicago (ICGC), AWS - Virginia, Collaboratory	PACA-AU	PCAWG	SSM	WGS	VCF	7.09 MB
FI9876	DO49113	PCAWG - Barcelona, PCAWG - Chicago (ICGC), AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	188.71 GB
FI9875	DO49113	PCAWG - Barcelona, PCAWG - Chicago (ICGC), AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	80.17 GB
FI9858	DO46933	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	153.97 GB
FI9857	DO46933	PCAWG - London, PCAWG - Barcelona, AWS - Virginia	RECA-EU	PCAWG	Aligned Reads	WGS	BAM	128.29 GB

ICGC Data Portal

Quick Search

ADVANCED SEARCH

Donors | Genes | Mutations

Share

Gene IS KRAS AND Functional Impact IS High

Mutation

e.g. MU123

Consequence Type

- Frameshift 2
- Missense 28
- Stop Gained 1

Functional Impact

- ✓ High 31
- Low 40
- Unknown 373

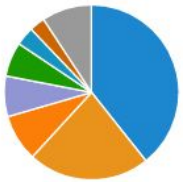
Type

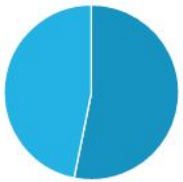
- Substitution 28
- Deletion 2
- MSub 1

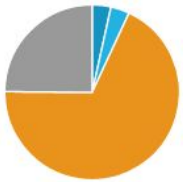
Platform

Analysis Type

Donors 173 | Genes 1 | Mutations 31







Show More

SAVE DONOR SET | DOWNLOAD DONOR DATA | VIEW IN DATA REPOSITORIES

Donors

Showing **1** - **10** of **173** donors

ID	Project	Site	Gender	Age	Stage	Survival (days)	Available Data Types											# Mutations	# Genes
							SSM	CNSM	StSM	SGV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN		
DO36874	READ-US	Colorectal	Female	62			✓	✓	--	--	✓	--	--	✓	✓	✓	--	1	1
DO32827	PACA-AU	Pancreas	Female	58		1,778	✓	--	--	--	✓	--	✓	--	--	--	--	1	1

ICGC Data Portal

Quick Search

DATA REPOSITORIES

Files | Donors

Share | Repository IS Collaboratory - Toronto AND Donor IN (Input donor set)

File

✓ Input donor set

Enter File ID, Name or Object

Repository

✓ Collaboratory -...	502
AWS - Virginia	439
PCAWG - Barcelona	50
PCAWG - Chicago (...	286
PCAWG - Chicago (...	114
PCAWG - Heidelbe...	144

6 more

Data Type

SSM	168
StSM	118
SGV	72
Aligned Reads	48
CNSM	48

1 more

Experimental Strategy

WGS	502

Download manifests | **502** Files | **24** Donors | **6.24 TB**

SAVE DONOR SET

Showing **1** - **25** of **502** files

	File ID	D	Repository	Project	Study	Data Type	Strategy	Format	Size
	FI9367	DO33392	PCAWG - Barcelona, PCAWG - London, Collaboratory - Toronto, AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	177.76 GB
	FI9366	DO33392	PCAWG - Barcelona, PCAWG - London, AWS - Virginia, Collaboratory - Toronto	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	79.81 GB
	FI5311	DO51505	PCAWG - London, PCAWG - Barcelona, PCAWG - Tokyo, Collaboratory - Toronto, AWS - Virginia	PACA-CA	PCAWG	Aligned Reads	WGS	BAM	121.82 GB
	FI5310	DO51505	PCAWG - London, PCAWG - Barcelona, PCAWG - Tokyo, Collaboratory - Toronto, AWS - Virginia	PACA-CA	PCAWG	Aligned Reads	WGS	BAM	101.12 GB
	FI5206	DO33544	PCAWG - Barcelona, PCAWG - London, AWS - Virginia, Collaboratory - Toronto	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	207.03 GB
	FI5205	DO33544	PCAWG - Barcelona, PCAWG - London, Collaboratory - Toronto, AWS - Virginia	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	136.85 GB
	FI49625	DO33208	PCAWG - Barcelona, PCAWG - London, AWS - Virginia, Collaboratory - Toronto	PACA-AU	PCAWG	Aligned Reads	WGS	BAM	201.24 GB
	FI49624	DO33208	PCAWG - Barcelona, PCAWG - London, AWS - Virginia,	PACA-AU	PCAWG	Aligned	WGS	BAM	133.71 GB

ICGC Data Portal

Quick Search


```
▼ <ResultSet date="2015-11-25 21:20:04" id="229969040">
    <Query> analysis_id:afeccc96-1be9-46dc-8f02-90a8af641f59 </Query>
    <Hits> 1 </Hits>
  ▼ <Result id="1">
      <analysis_id> afeccc96-1be9-46dc-8f02-90a8af641f59 </analysis_id>
      <state> live </state>
      <reason />
      <last_modified> 2014-11-10T18:56:52Z </last_modified>
      <upload_date> 2014-10-31T20:31:49Z </upload_date>
      <published_date> 2014-11-01T00:37:02Z </published_date>
      <center_name> DKFZ </center_name>
      <study> dkfz_pancancer </study>
      <aliquot_id> bd1ce02b-bfaf-4cde-aaac-06097a12e248 </aliquot_id>
    ▶ <files>
      <sample_accession />
      <dcc_project_code> PBCA-DE </dcc_project_code>
      <dcc_specimen_type> Primary tumour - solid tissue </dcc_specimen_type>
      <participant_id> ICGC_MB153 </participant_id>
      <specimen_id> ICGC_MB153 </specimen_id>
      <sample_id> ICGC_MB153 </sample_id>
      <use_cntl> 1dc17d0f-c906-46e0-b600-a158520de3e4 </use_cntl>
      <analyte_code />
      <library_strategy> WGS </library_strategy>
      <platform> ILLUMINA </platform>
      <refassem_short_name> GRCh37 </refassem_short_name>
    ▶ <analysis_xml>
    ▶ <experiment_xml>
    ▶ <run_xml>
    ▼ <analysis_detail_uri>
        https://gtrepo-dkfz.annailabs.com/cghub/metadata/analysisDetail/afeccc96-1be9-46dc-8f02-90a8af641f59
      </analysis_detail_uri>
```


		Actions ?
		Manifest Metadata
	GNOS	Manifest Metadata
	GNOS	Manifest Metadata
	S3	Manifest Metadata

BAM iobio

Real time BAM inspection

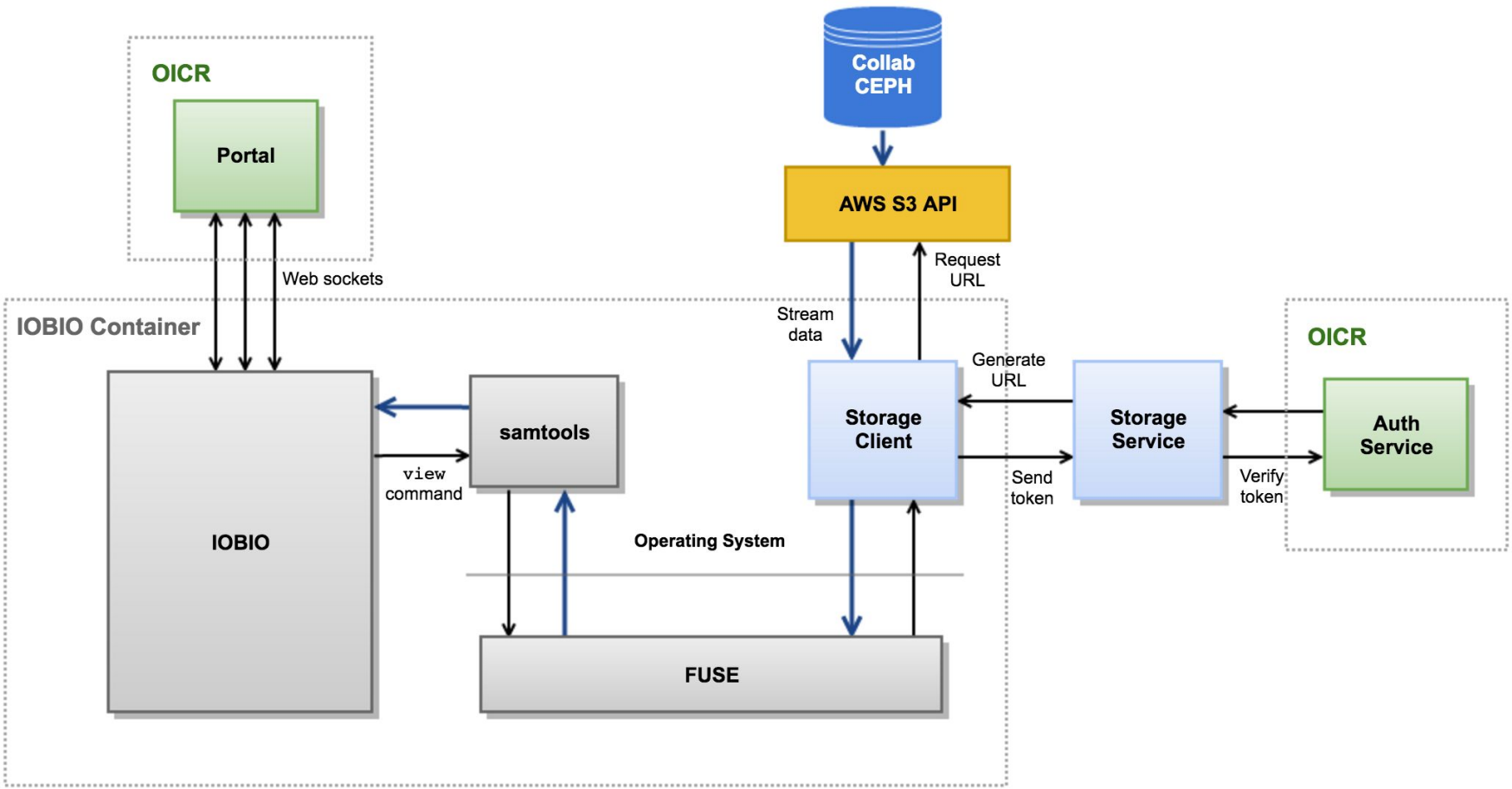
Developed by Gabor Marth's lab

Streaming reads sampled from AWS and Collab

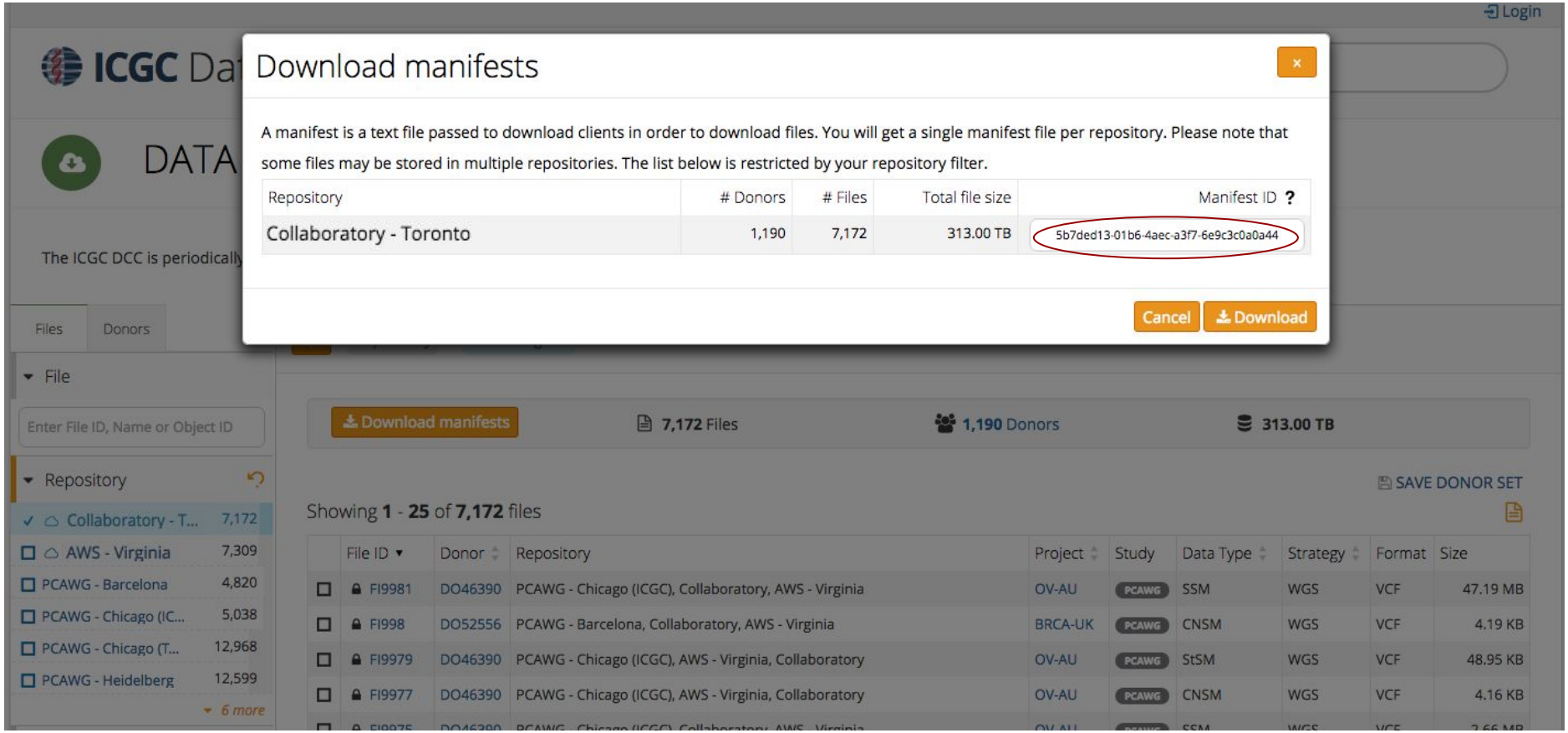
[BAM LiveDemo]

[VCF LiveDemo]

Iobio - An application of the ICGC-storage's Fuse Feature



Downloading/Saving Manifests




```
%: icgc-storage-client download --manifest
5b7ded13-01b6-4aec-a3f7-6e9c30a0a44 --output-dir data
--output-layout bundle
```

Downloading from multiple clouds/repos

Download manifests

A manifest is a text file passed to download clients in order to download files. You will get a single manifest file per repository. Please note that some files may be stored in multiple repositories. The list below is restricted by your repository filter.

Repository	# Donors	# Files	Total file size	Manifest ID ?
GDC - Chicago	10,660	76,334	126.57 TB	Not Applicable
Collaboratory - Toronto	1,907	42,549	477.81 TB	1a6ae681-1e5f-4d44-b297-d09403fa720d
AWS - Virginia	1,394	31,204	365.38 TB	eae74f02-899d-40a9-8bd9-a6609a51d055
EGA - Hinxton	6	12	1.40 TB	Not Applicable
PDC - Chicago	1	1	140.30 GB	Not Applicable

Cancel Download

Tar archives that contains a manifest file for each selected repository are generated

Lack of Interoperability among Clouds/Repos

- Cloud-specific manifest file formats and software for accessing data

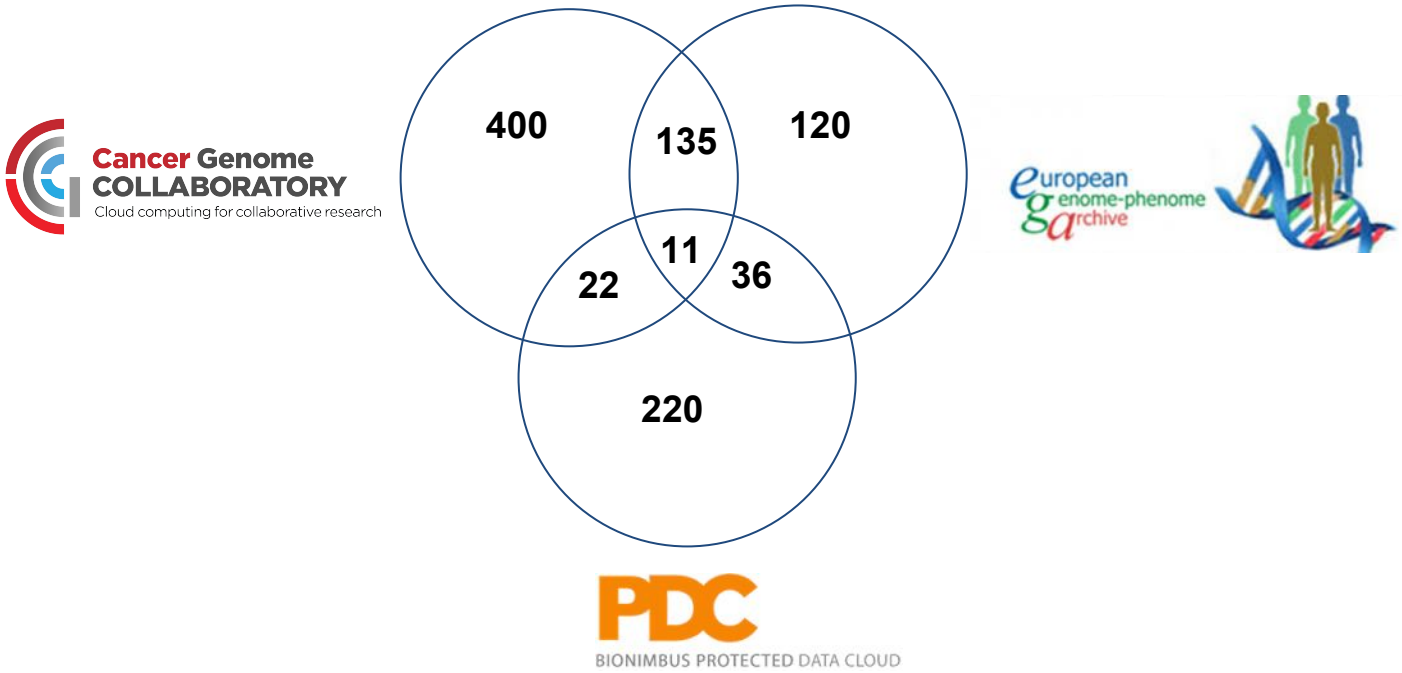
Cloud/Repository	Manifest format	Download Software
Annai-GNOS	XML	gtDownload, JVM
GDC	TSV but soon YAML	GDC Data Transfer Tool
Collab, AWS	TSV or saved manifests with IDs	ICGC-Storage-Client
EGA	No support for manifests, ICGC generates a shell script with file IDs injected	JVM, EGA Download client
PDC	No support for manifests, ICGC generates a shell script with file IDs injected	AWS client

- Users need to install and learn how to use several applications with different manifest formats

Redundant Downloads

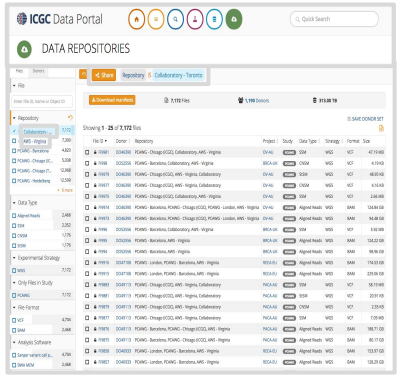
- Potentially, users will download multiple times the files that are stored in two or more clouds/repositories

# of queried files per repository

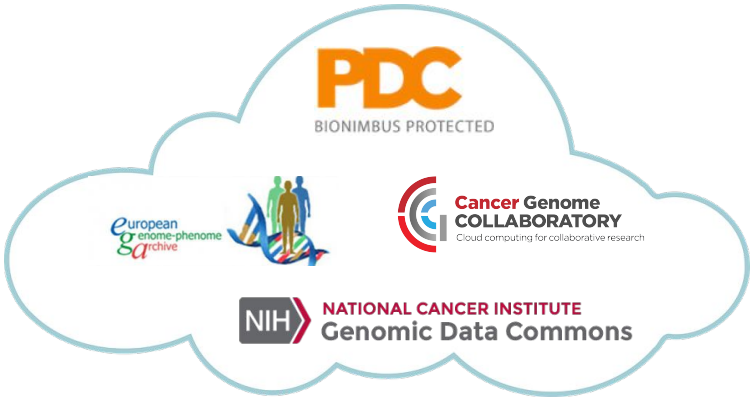


A Universal Download Client for ICGC

- ICGC-get
  - A single and easy-to-use tool for downloading ICGC data residing on multiple clouds / repositories
  - Works with EGA, PDC, AWS, Collaboratory, GNOS, GDC
  - Makes ICGC clouds/repos interoperable and seamless
  - Deep integration with the ICGC Portal for optimal usability
  - Released by end of June



II - Get a manifest_ID



I - Search Data

III - ICGC-get download <manifest_ID>

ICGC-get : Convenient deployment

- Single Docker container hosted on DockerHub
- Easy to install and update

Docker Pull Command

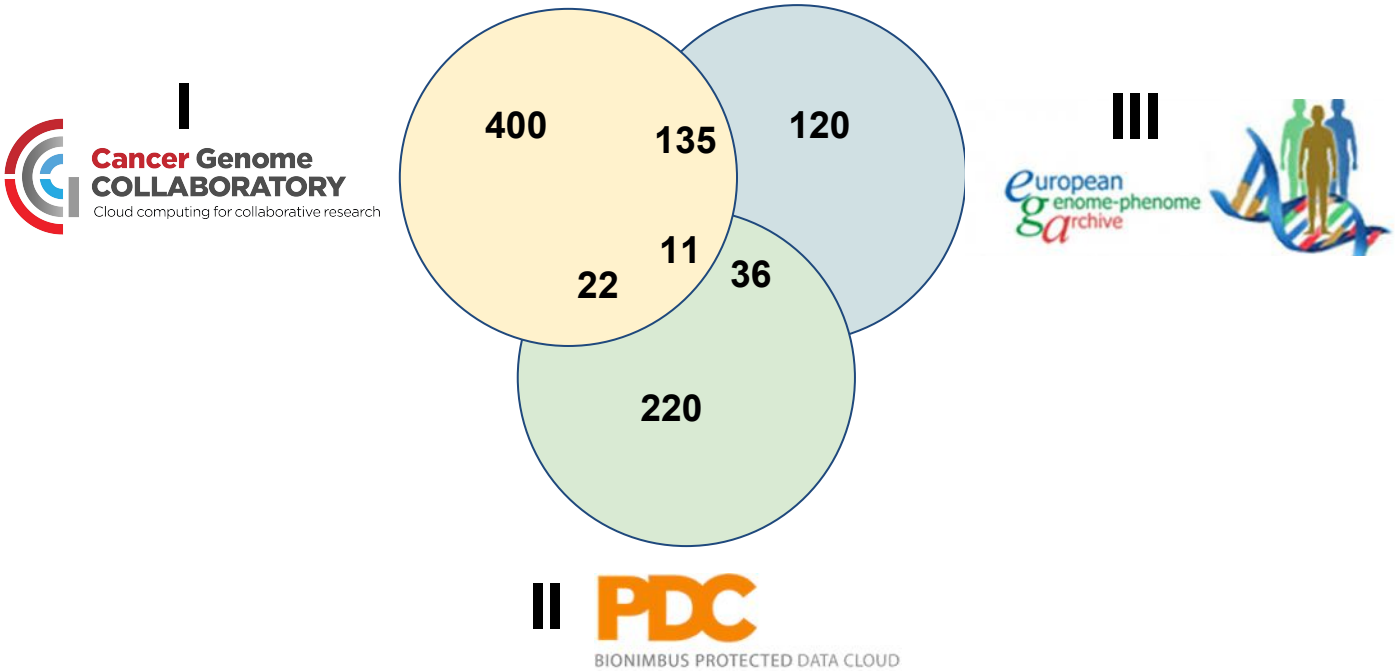
```
docker pull icgc/icgc-get
```

- Combines all required clients and their dependencies

gtDownload, JVM
GDC Data Transfer Tool
ICGC-Storage-Client
JVM, EGA Download client
AWS client

ICGC-get : Some Features

- File-less manifests for easy context switching and sharing
- Single file ID scheme independent of repo (e.g. FI1234)
- Preview mode to verify large downloads before proceeding
- Download a particular file copy based on repos preference:
  - Access, geographical proximity, reliability of cloud, etc.
- Configurable cloud/repo precedence



ICGC-get client

Inspect the version of each client

```
%: icgc-get version
AWS CLI Version: 1.10.34
GDC Client Version v0.7
EGA Client Version: 2.2.2
Gtdownload Release 3.8.7
ICGC Storage Client Version: 1.0.13
ICGC-Get Version: 0.5
```

Download a single file from a particular repository

```
%: icgc-get download FI378424 --repo collaboratory
```

Download multiple files using a manifest with repo priorities

```
%: icgc-get download --manifest 123e4567-e89b-12d3-426655
--repo pdc --repo collaboratory
```

Cloud/repo-specific Authentication (still)

Token Manager

OVERVIEW

Personal access tokens function like ordinary OAuth access tokens, similar to those offered by GitHub.com. They can be used instead of a password to access ICGC resources. Tokens allow you to associate *scopes* which limit access to that needed for the target environment. From this page, you can create your own personal API tokens for use in scripts and on the command line. This feature is required when using the Storage Client or programatically downloading controlled access data from the Data Portal.

Please note that access tokens are associated with your user credentials so you must never share your personal tokens with anyone.

GENERATE NEW TOKENS

Select the desired set of scopes below and click "Generate". You may also enter a description to remind yourself what the token is for. Your new token will be shown in the next section. Note that there is currently a limit of one token per unique set of scopes.

Select Required Scopes

- [ ] portal.download — *Allows secure downloads from the Data Portal*
- [x] collab.download — *Allows secure downloads from the Collaboratory cloud*
- [ ] aws.download — *Allows secure downloads from AWS S3*

Enter a Description

What is this token for?

Generate

MANAGE ACTIVE TOKENS

The following are your **1** active token(s). You may revoke a token if it is no longer needed or you believe it has been compromised.

Token Id	Description	Scopes	Expires in	
dfa0f7ae-7f7a-4554-944b-96fd0f894384		collab.download	213.05 days	

Comprehensive user guide on docs.icgc.org

ICGC DCC Docs

Home | Portal | Cloud | Submission | Dictionary | Software

CLOUD GUIDE

Overview

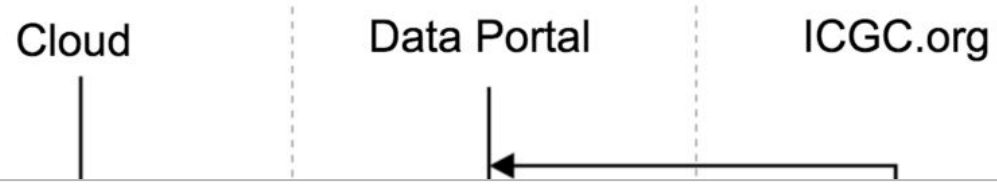
This user guide describes the steps to securely explore and analyze ICGC data stored in Amazon (*AWS*) or Collaboratory (*OpenStack*) cloud environments. For more information about ICGC cloud initiatives, please see ICGC in the Cloud.

Please see Terms for a glossary of terms used in this guide.

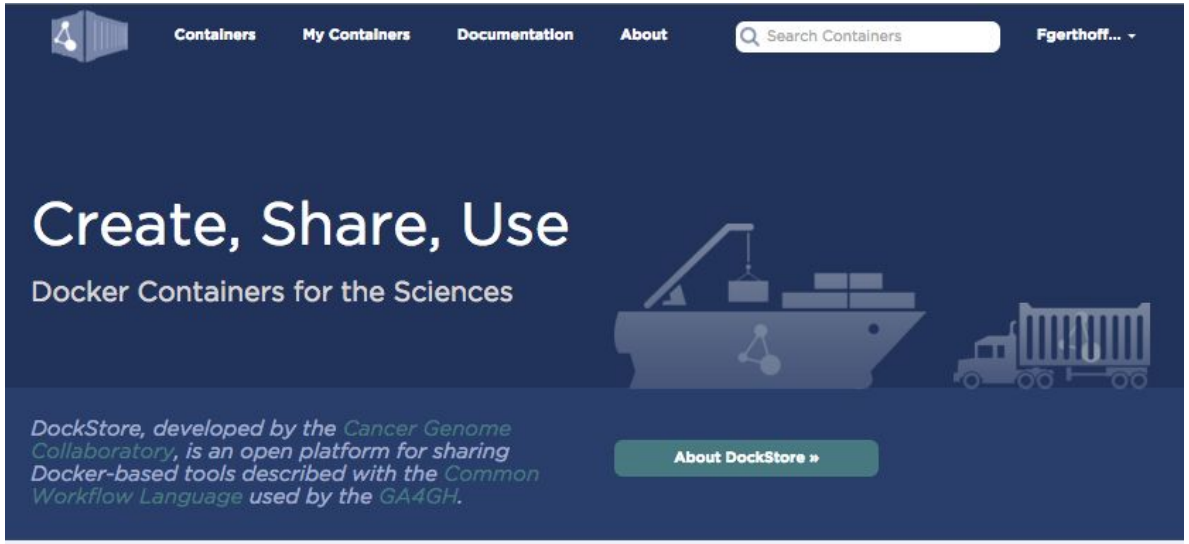
Process

The figure below illustrates the overall process and systems involved:

1. Authorization Apply for DACO *Cloud Access* if not already approved Upon approval, login to the *Data Portal* Generate an *Access Token* for cloud download
2. Compute Prerequistes Provision a *Compute Instance* in the target cloud
3. Installation Download and install the ICGC *Storage Client*
4. Configuration Configure the Storage Client to use the generated Access Token
5. File Search Identify files of interest using the Data Portal
6. Storage Client Usage Download or view data with the provided Storage Client or via an external tool



The Collaboratory Dockstore



www.dockstore.org

- An open platform for sharing docker-based tools
- Containers described with Common Workflow Language (CWL) and/or Workflow Description Language (WDL)
- API for programmatic access
- Developed in close collaboration with GA4GH

What's next for the Collaboratory

- Complete integration of BioNimbus PDC and Collaboratory software infrastructures for better interoperability

- Continue to upload ICGC data on PDC, AWS and Collab

- A new data submission system
  - Clinical data, metadata and raw sequencing data
  - Data dictionary driven
  - Based on our experience with the ICGC and the GDC submission systems
  - To be used by the ICGCmed (200,000 donors)

- Implement
  - the GA4GH Search API for reads and variants
  - a cost-recovery system based on usage

Acknowledgment to ICGC/Collab team

- Andy Yang
- Bob Tiernay
- Christina Yung
- Denis Yuen
- Dusan Andric
- Francois Gerthoffert
- George Mihaiescu
- Junjun Zhang
- Lincoln Stein
- Linda Xiang
- Nodirjon Fayzullaev
- Phuong-My Do
- Solomon Shorser
- Vincent Ferretti
- Vitalii Slobodianyk

Funders

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

NSERC CRSNG

GenomeCanada

CIHR IRSC
Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

Ontario

GenomeBritishColumbia

OntarioGenomicsInstitute

GenomeQuébec