

cBioPortal

for Cancer Genomics

Webinar 5: API & R Client

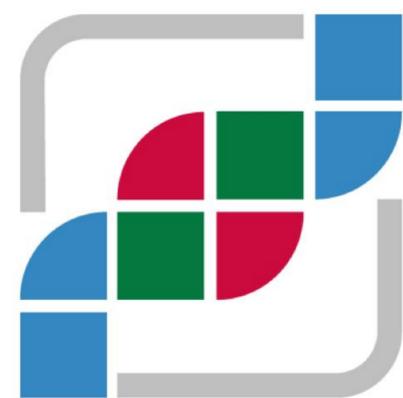
May 28, 2020

Webinar will begin momentarily.



Memorial Sloan Kettering
Cancer Center™





cBioPortal

for Cancer Genomics

Webinar 5: API & R Client

May 28, 2020



Memorial Sloan Kettering
Cancer Center™



Webinar Schedule

- April 30: Introduction to cBioPortal
- May 7: Mutation Details & Patient View
- May 14: OQL & Expression
- May 21: Group Comparison
- May 28: API & R Client

Webinar Schedule

- April 30: Introduction to cBioPortal
- May 7: Mutation Details & Patient View
- May 14: OQL & Expression
- May 21: Group Comparison
- May 28: API & R Client **(The last webinar for now 🥲)**

Acknowledgements



Nikolaus Schultz
Jianjiong Gao
Benjamin Gross
S. Onur Sumer
Yichao Sun
Hongxin Zhang
Adam Abeshouse
Ritika Kundra
Ino de Bruijn
Robert Sheridan
Angelica Ochoa
Aaron Lisman
Manda Wilson
Avery Wang
Jing Su
Ramyasree Madupuri
Gaofei Zhao
Xiang Li
Christoph Kreitzer



Ethan Cerami
Chris Sander
Tali Mazor
Luke Sikina
Pieter Lukasse
Priti Kumari
Augustin Luna
James Lindsay



The Children's Hospital of Philadelphia

Adam Resnick
Allison Heath
Karthik Kalletla
John Maris



Marcel Ramos
Levi Waldron



Fedde Schaeffer
Oleguer Plantalech
Pim van Nierop
Paul van Dijk
Sander Rodenburg
Sjoerd van Hagen
Kees van Bochove



Trevor Pugh
Kelsey Zhu



Olivier Elemento
Alexandros Sigaras

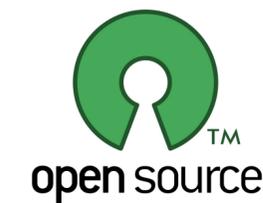


Ugur Dogrusoz
Ziya Erkoc

Alumni

B. Arman Aksoy
Istemi Bahceci
Caitlin Byrne
Hsiao-Wei Chen
Ersin Ciftci
Fred Criscuolo
Leonard Dervishi
Gideon Dresdner
Andy Duffile
Catherine Del Vecchio Fitz
Arthur Goldberg
Zachary Heins
Michael Heuer
Anders Jacobsen
Ngoc Nguyen

Erik Larsson
Dong Li
Tamba Monroe
Peter Kok
Irina Pulyakhina
Pichai Raman
M. Furkan Sahin
Kaan Sancak
Sander de Ridder
Sander Tan
Jiaojiao Wang
Stuart Watt
James Xu
Dionne Zaal
Riza Nugraha



Funding: Present & past



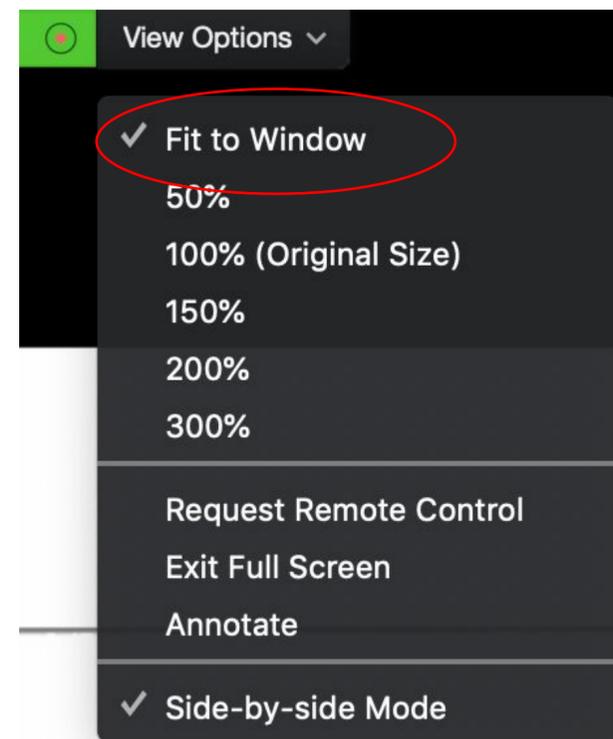
Email: cbioportal@googlegroups.com

 [@cBioPortal](https://twitter.com/cBioPortal)

 github.com/cBioPortal

Questions?

- Please ask questions using the Q&A feature
- We will try to answer some live, some directly
- The recording of all webinars will be posted on <https://cbioportal.org/tutorials/>
- If you still have questions after the webinar, please ask them via our Google Group: cbioportal@googlegroups.com



- Platform for **exploratory and interactive visualization, analysis and download** of large-scale cancer genomics data sets
- **Open source** software jointly developed by Dana-Farber Cancer Institute, Memorial Sloan Kettering Cancer Center, Princess Margaret Cancer Centre, Children’s Hospital of Philadelphia, and The Hyve
- **Public website** (cBioPortal.org) with public data (TCGA, ICGC, published sequencing studies)
 - Private instances are installed at academic and commercial institutions world-wide
 - You can make OncoPrints and Lollipop plots with your own data (“Visualize Your Data” page)

- Platform for **exploratory and interactive visualization, analysis and download** of large-scale cancer genomics data sets
- **Open source** software jointly developed by Dana-Farber Cancer Institute, Memorial Sloan Kettering Cancer Center, Princess Margaret Cancer Centre, Children's Hospital of Philadelphia, and The Hyve
- **Public website** (cBioPortal.org) with public data (TCGA, ICGC, published sequencing studies)
 - Private instances are installed at academic and commercial institutions world-wide
 - You can make OncoPrints and Lollipop plots with your own data (“Visualize Your Data” page)
 - **Today:** make your own visualizations using public data

What you **WILL** learn in this webinar

1. What is a REST API and why use one?
2. How to use the R package cBioPortalData to pull data from the API

What you **WILL NOT** learn in this webinar

- Basics of R
- tidyverse
- Bioconductor (MultiAssayExperiment, SummarizedExperiment, etc)

Please see these other excellent resources:

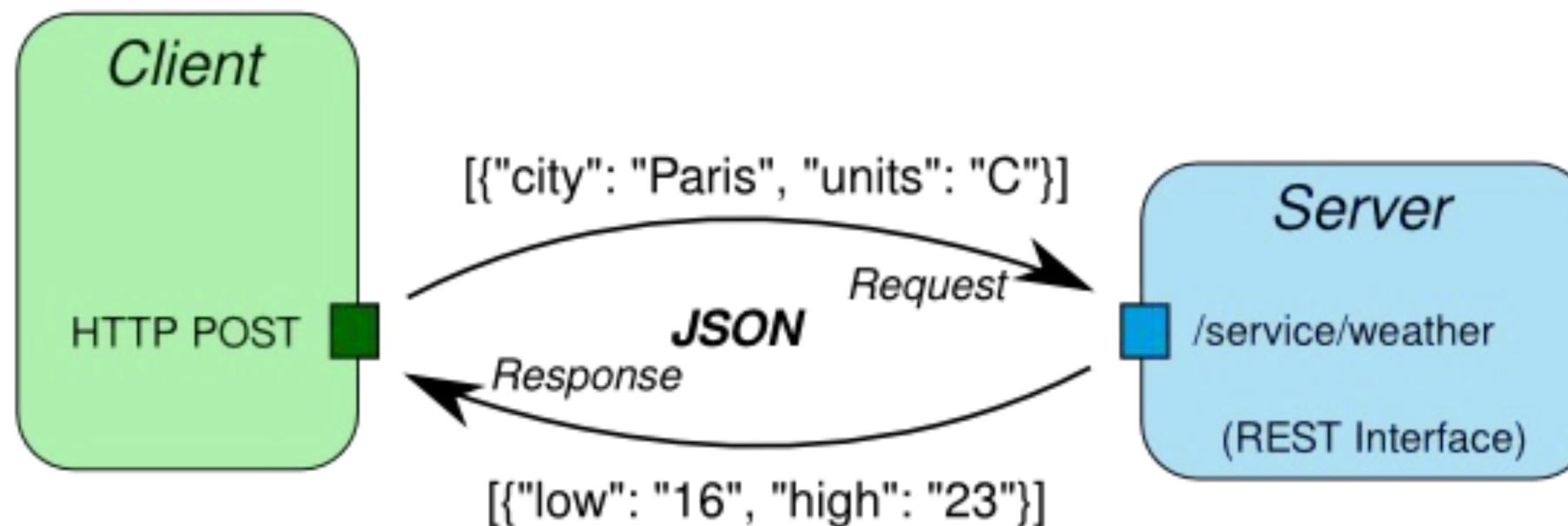
Bioconductor: <https://bioconductor.github.io/BiocWorkshops/>

MultiAssayExperiment: <https://waldronlab.io/MultiAssayExperiment/>

Tidyverse: <https://www.tidyverse.org/learn/>

What is a REST API?

- **API** (**A**pplication **P**rogramming **I**nterface) is a set of rules by which an application can interact with another
- **A REST API** uses **RE**presentational **S**tate **T**ransfer (REST), an architectural style and approach to communications often used in web services
- Client and server communicate with **HTTP** requests and send data in **JSON** format



REST APIs are used everywhere

In Chrome open Developer Tools, Open the network tab and go to e.g. [google.com](https://www.google.com)

The image shows a Chrome browser window with the Google logo. The Chrome Developer Tools are open, with the Network tab selected. The 'XHR' filter is active, and a request to `search?q&cp=0&client=.../complete` is highlighted. The 'Preview' tab is open, showing the JSON response. The response is a list of search suggestions, including 'what is a rest api', 'rest api', 'what is a rest api for dummies', 'import transform visualize', 'web api', 'what is rest api', and 'tidyverse tutorial'. The 'REQUEST' and 'RESPONSE' labels are overlaid in red on the image.

REQUEST

RESPONSE

```
)]}' [,...]  
  0: [{"json", 35, [362, 39],...}, {"what is a rest api", 35, [362, 39], {...}}, {"rest api", 35, [362, 39],...},...]  
    0: [{"json", 35, [362, 39],...}  
    1: [{"what is a rest api", 35, [362, 39], {...}}  
    2: [{"rest api", 35, [362, 39],...}  
    3: [{"what is a rest api for dummies", 35, [362, 39], {...}}  
    4: [{"import transform visualize", 35, [362, 39], {...}}  
    5: [{"web api", 35, [362, 39],...}  
    6: [{"what is rest api", 35, [362, 39], {...}}  
    7: [{"tidyverse tutorial", 35, [362, 39], {...}}]
```

cBioPortal.org's REST API

<https://cbioportal.org/api>

 **swagger** Select a spec

cBioPortal web API [Beta] 1.0 (beta)

[Base URL: www.cbioportal.org/api]
<https://www.cbioportal.org/api/api-docs>

A web service for supplying JSON formatted data to cBioPortal clients. Please note that this API is currently in beta and subject to change.

[cBioPortal - Website](#)
[Send email to cBioPortal](#)
[License](#)

A. Cancer Types ∨

GET	<code>/cancer-types</code>	Get all cancer types
GET	<code>/cancer-types/{cancerTypeId}</code>	Get a cancer type

Why use cBioPortal.org's REST API?

- **Flexible:** fetched data can be further analyzed in any way
- **Structured data:** relatively little work to transform
- **Slice data:** get only what you need
- **Curated data:** many well curated studies available
- **Automated processes** (no manual steps involved with downloading)
- **Changes little** over time (few breaking changes)
- **Reproducible**

rerun against updated data or e.g. private data in a local instance
reproduce exact data pulling that happens when visiting the website

Why use cBioPortal.org's REST API?

- Example use cases:
 - a. Reproducible analysis for a **paper**
 - b. **Custom dashboard** (e.g. daily new cases counts for a specific type of cancer)
 - c. Some other **custom visualization/analysis** not yet in cBioPortal
 - d. **Performance heavy analyses** e.g. analysing all genes

Etc...

When to use cBioPortal.org's REST API?

- Several ways to **download data**:

- a. **Manually** using download option in plots on cBioPortal.org (TXT, TSV)

- For a one time quick thing

- b. **Datahub** (<https://github.com/cBioPortal/datahub>)

- Get complete data dump (TSV/TXT)

- Analyze really large datasets (e.g. >100K samples, all genes)

- Semi-Structured (TSV/TXT): Transforming data is more work

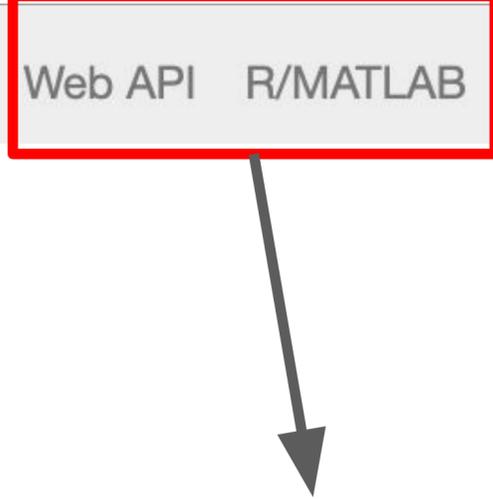
- Note: cBioPortalData R package provides an interface for this called "cBioDataPack"

- c. **REST API**

- All other cases

- Note: fine to pull large datasets as well with cBioPortalData b/c it does caching

How to use cBioPortal.org's REST API?



<https://docs.cbioportal.org>

Clients available in many languages including Python, JavaScript, R

How to use cBioPortal.org's REST API in R?

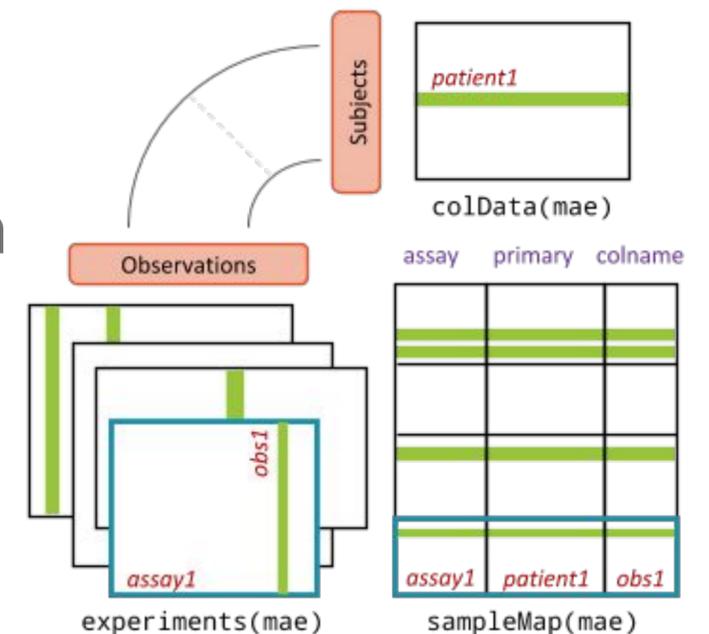
There are two R clients:

1. cgdsr (will be deprecated)

- Uses an old version of our web API (cbioportal.org/webservice.do)

2. cBioPortalData

- Access all API endpoints (cbioportal.org/api)
- MultiAssayExperiment interface
 - Convenient popular data structure
 - Easily combine multiple assays e.g. mutations, expression, methylation
 - Select columns by patient's clinical/pathological data
 - Flexible API including harmonized subsetting and reshaping
- Provides complete datasets, not just subsets of genes
- Automatic local caching, thanks to BiocFileCache



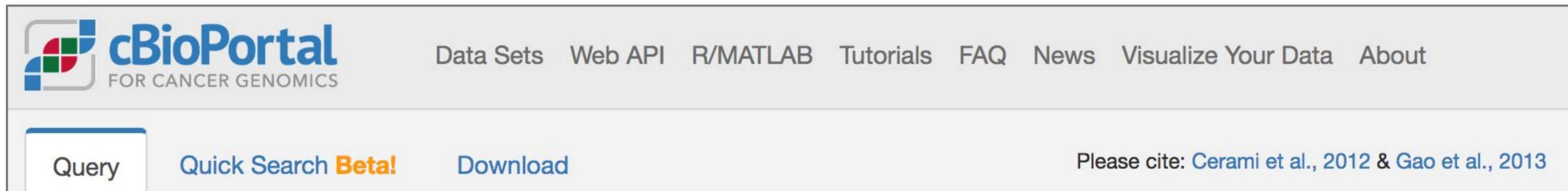
More info: <https://waldronlab.io/cBioPortalData/>

R coding time!

<https://github.com/cBioPortal/2020-cbioportal-r-workshop>

Using cBioPortal for publication

- **Please use cBioPortal in your publications!**
- Cite the studies that generated the data you are using if applicable
- Cite cBioPortalData manuscript (in review) if applicable
- Cite cBioPortal



The screenshot shows the top navigation bar of the cBioPortal website. On the left is the cBioPortal logo with the text "FOR CANCER GENOMICS". To the right of the logo are several navigation links: "Data Sets", "Web API", "R/MATLAB", "Tutorials", "FAQ", "News", "Visualize Your Data", and "About". Below the navigation bar is a search bar with the text "Query" inside. To the right of the search bar are two buttons: "Quick Search Beta!" and "Download". On the far right of the search bar area, there is a citation notice: "Please cite: Cerami et al., 2012 & Gao et al., 2013".

Questions?

- The recording of all webinars will be posted on <https://cbioportal.org/tutorials/>
- If you still have questions, please email: cbioportal@googlegroups.com
-



Thank you for joining us!