

U24: Informatics tools for cancer research

ITCR Annual PI Meeting

University of California Santa Cruz Genomics Institute

Santa Cruz, California

May 31, 2017

Rachel Karchin, Ph.D.

Department of Biomedical Engineering

Institute of Computational Medicine

Department of Oncology

Johns Hopkins University

Michael Ryan, Ph.D.

In Silico Solutions

Overview

- What is CRAVAT?
- Brief news and updates
- Outreach efforts and usage stats
- Visualization tools and mini-demo about seeing protein mutations in 3D
- How you can incorporate us into your tools

CRAVAT is a cancer mutation analysis portal

The screenshot displays the CRAVAT 4.3 web interface. At the top, the CRAVAT logo is shown with the version number 4.3 and a 'Follow' button. A navigation bar includes links for Help, Release notes, How to Cite, Contact, My Jobs, My Profile, and Log-out. The main content area is divided into three panels: 1 Input, 2 Analysis, and 3 Results. The Input panel has checkboxes for VCF and CRAVAT format examples, a text area for variants, and a file upload section with a 'Choose File' button and a 'Clear' button. The Analysis panel shows a 'Choose analysis program' section with checkboxes for VEST, CHASM, GeneCard and PubMed annotation, and SnpGet. The Results panel has a text input for the email address to send the report to, with 'karchin@jhu.edu' entered, and a checkbox for 'Include text reports'. A large 'SUBMIT' button is at the bottom right. A footer contains copyright information for the Karchin lab at Johns Hopkins University, funding information, and links to Disclaimer, Privacy Policy, and Licensing.

CRAVAT 4.3

Follow

Welcome, karchin@jhu.edu

Help | Release notes | How to Cite | Contact | My Jobs | My Profile > | Log-out

1 Input

☐ Check for VCF-format input example.
☐ Check for CRAVAT-format input example.

Enter variants below: ?

or upload a variant file: ?

Choose File pam03-somatic-all.vcf Clear

☐ hg18 ?

2 Analysis

Choose analysis program

☒ VEST ?
☐ CHASM ?
☒ GeneCard and PubMed annotation ?
☐ SnpGet ?

3 Results

Send the analysis report ? to:

karchin@jhu.edu

☐ Include text reports ?

SUBMIT

What will I get?

© 2016 Karchin lab, Institute for Computational Medicine, The Johns Hopkins University | Funding: 5U01CA180956-03 | Disclaimer | Privacy Policy | Licensing

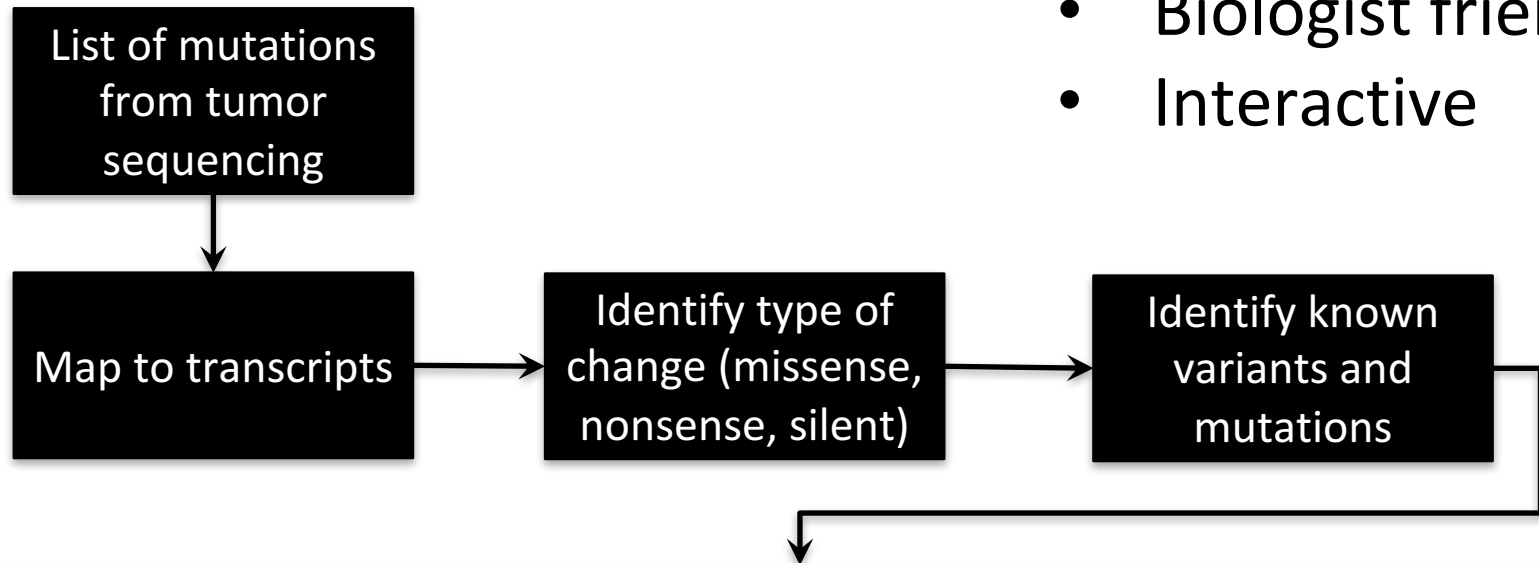
<http://cravat.u>

s

<https://hub.docker.com/r/karchinlab/cravatmupit/>

CRAVAT in a nutshell

- Biologist friendly
- Interactive



Analysis

Predict driver vs.
random
mutations

Predict functional
impact of
mutations

Visualize
mutations

Find important
mutated genes
and pathways

News and updates

- Two major CRAVAT releases (7/2016, 11/2016)
- Underlying databases entirely refreshed/rebuilt.
- Support for HGVS input and output
- Full recoding of MuPIT protein 3D visualization tool to use latest WebGL technology
- Updated CHASM and VEST classifiers.
- Coming soon: new driver gene and mutation classifiers, next CRAVAT release for GRCh38 genome assembly

Outreach and usage



647
PULLS

Publications

[Proc Natl Acad Sci U S A](#). 2016 Dec 13;113(50):14330-14335. Epub 2016 Nov 22.

Evaluating the evaluation of cancer driver genes.

[Tokheim CJ](#)^{1,2}, [Papadopoulos N](#)^{3,4}, [Kinzler KW](#)^{3,4}, [Vogelstein B](#)^{5,4}, [Karchin R](#)^{6,2,7}.

[Cancer Res](#). 2016 Jul 1;76(13):3719-31. doi: 10.1158/0008-5472.CAN-15-3190. Epub 2016 Apr 28.

Exome-Scale Discovery of Hotspot Mutation Regions in Human Cancer Using 3D Protein Structure.

[Tokheim C](#)¹, [Bhattacharya R](#)¹, [Niknafs N](#)¹, [Gygax DM](#)², [Kim R](#)², [Ryan M](#)², [Masica DL](#)¹, [Karchin R](#)³.

Cancer Res. 2017 in press

CRAVAT 4: Cancer-Related Analysis of Variants Toolkit.

Masica DL, Douville C, Tokheim C, Bhattacharya R, Kim R, Moad K, Ryan M, Karchin R.

Poster

- GRC Human Genetic Variation and Disease 2016
- ASHG 2016

Tutorial

- Invited workshop ASHG 2017
- NCI Bioinformatics Training & Education Program 2016 and 2017

Talk

- NGS'16 Genome Annotation (Barcelona, Spain)
- Technical University of Munich
- ETH Zurich
- Canceromatics III (Madrid, Spain)
- University of Pennsylvania
- Indiana University School of Medicine
- NYU Genome Center

Public webportal direct usage

Total jobs processed

2015 June-2016 May 5674

2016 June-2017 May 29851

Total mutations processed

2015 June-2016 May 340709750

2016 June-2017 May 399534327

Total unique users submitting jobs

2015 June-2016 May 518

2016 June-2017 May 497

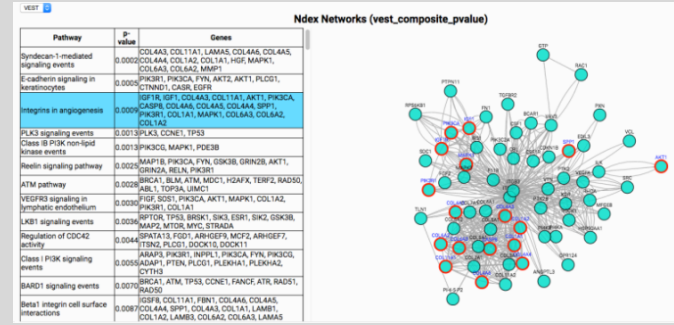
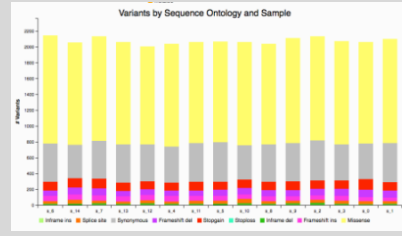
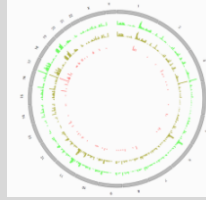
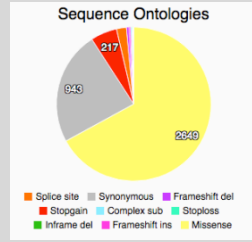
Public webportal web services calls

2015 June-2016 May 514

2016 June-2017 May 2446731

Dynamic interactive tables and visualization widgets

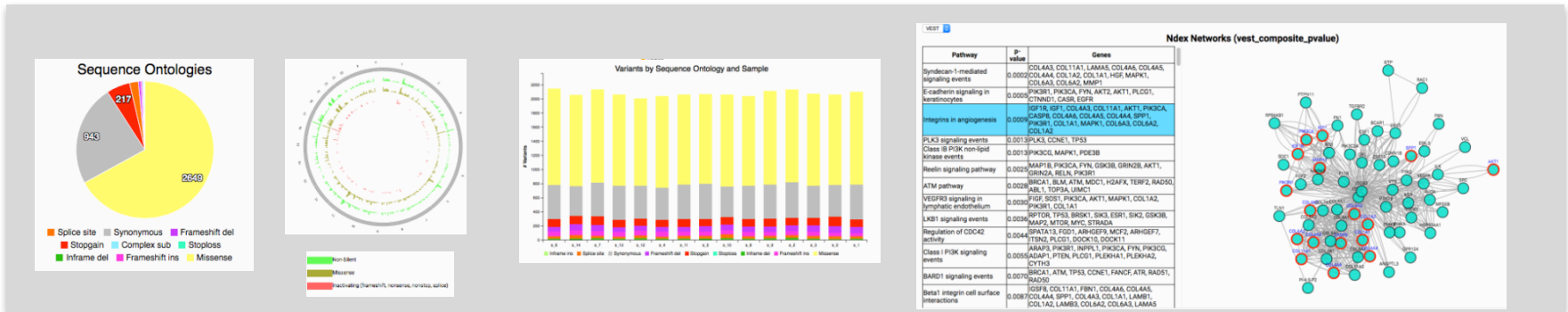
Study summary widgets



interactive tables and visualization

Dynamic interactive tables and visualization widgets

Study summary widgets



interactive tables and visualization

Summary Gene Variant Noncoding Error

Job Info: eprice_20170202_062906

Filter: CHASM p-value, VEST p-value

Columns: Gene Info, Variant Impact, VEST, CHASM, Disease Assoc, Structure, Study

Gene: PRKCB

Crystal structure of a C2 domain of a protein kinase C alpha (PRKCA) from Homo sapiens at 1.90 A resolution (Structure ID: 4DNL)

Protein Mutations Annotations

Chains

Style: Cartoon Color By: Secondary Structure Opacity: 1

Protein Kinase C Alpha Type

Small Molecules

Style: Stick Color By: Molecule (custom)

MSE NA

Show labels on clicked atoms Clear Labels

Capture Image Size: Small

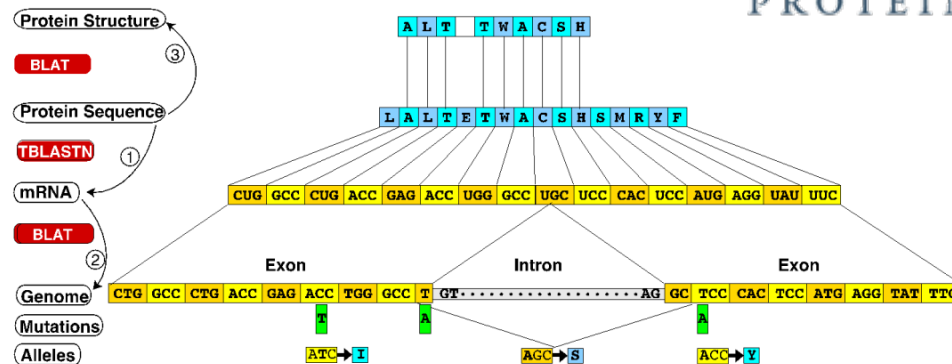
© 2016 Karchin lab | Funding: U24CA204817-01 | To cite: PMID: 23793516 | Credits | Visualization uses 3dMol.js

CGC Inheritance: None
CGC Somatic Tumors: None
CGC Germline Tumors: None
COSMIC gene hits: 260

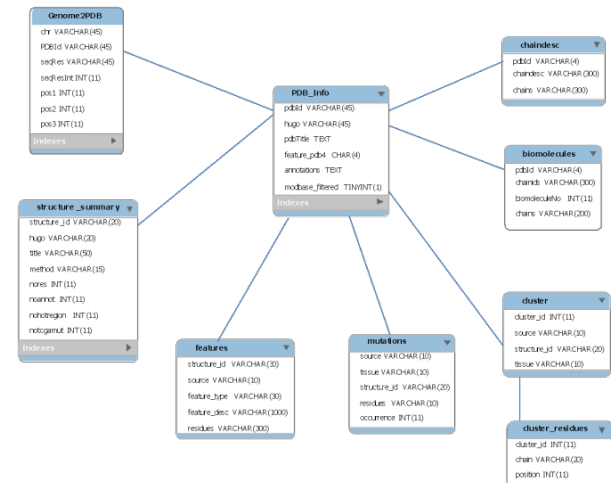
Infrastructure for MuPIT 3D mutation visualization




HotMAPS



42,268 PDB structures
99,929 Homology models
15,697 Human genes



Mini-demo of MuPIT



The screenshot displays the MuPIT web application interface. On the left, a sidebar contains the MuPIT logo, social media icons, and a search icon. The main header area shows the gene name "Gene: BRAF" and the structure description "Crystal structure of the BRAF:MEK1 complex (Structure ID: 4MNE)". A protein structure is visible in the background. A top navigation bar includes tabs for "Protein", "Mutations", and "Annotations". Below these, a "Chains" section shows the protein name "Serine/threonine-protein Kinase B-raf" and options for style, color, and opacity. A large, light blue-bordered "Query" dialog box is centered on the screen, featuring a search icon and a red arrow pointing to it. The dialog box contains two input methods: "Enter genomic positions (GRCh37)" with a "Load example" button, and "Upload a file." with a "Choose File" button. The example text in the first input field is "chr10 123298184" and "chr10 123325191". At the bottom of the dialog are "Submit", "Clear", and "Cancel" buttons.

Gene: BRAF
Crystal structure of the BRAF:MEK1 complex (Structure ID: 4MNE)

Protein Mutations Annotations

Chains

Style: Cartoon Color By: Secondary Structure Opacity: 1

Serine/threonine-protein Kinase B-raf

Query

Enter genomic positions (GRCh37) or Upload a file.

OR Load example

chr10 123298184
chr10 123325191

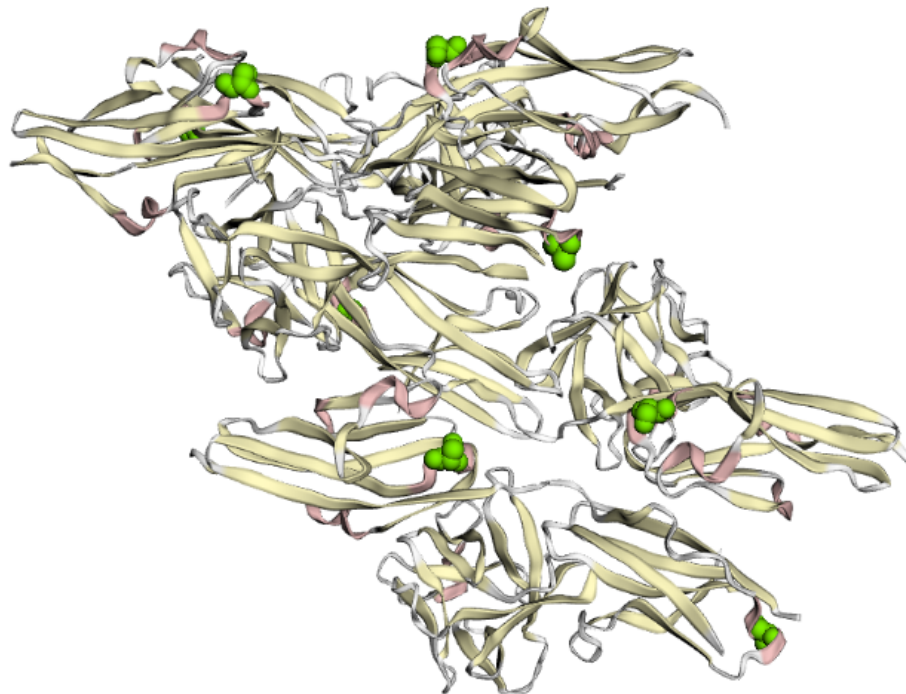
Choose File No file chosen

Submit Clear Cancel



Gene: FGFR2

CRYSTAL STRUCTURE OF FGF2 IN COMPLEX WITH THE EXTRACELLULAR LIGAND BINDING DOMAIN OF FGF RECEPTOR 2 (FGFR2) (Structure ID: 1EV2)



Protein Mutations Annotations

Chains

Style: Color By: Opacity:

Small Molecules

Style: Color By:

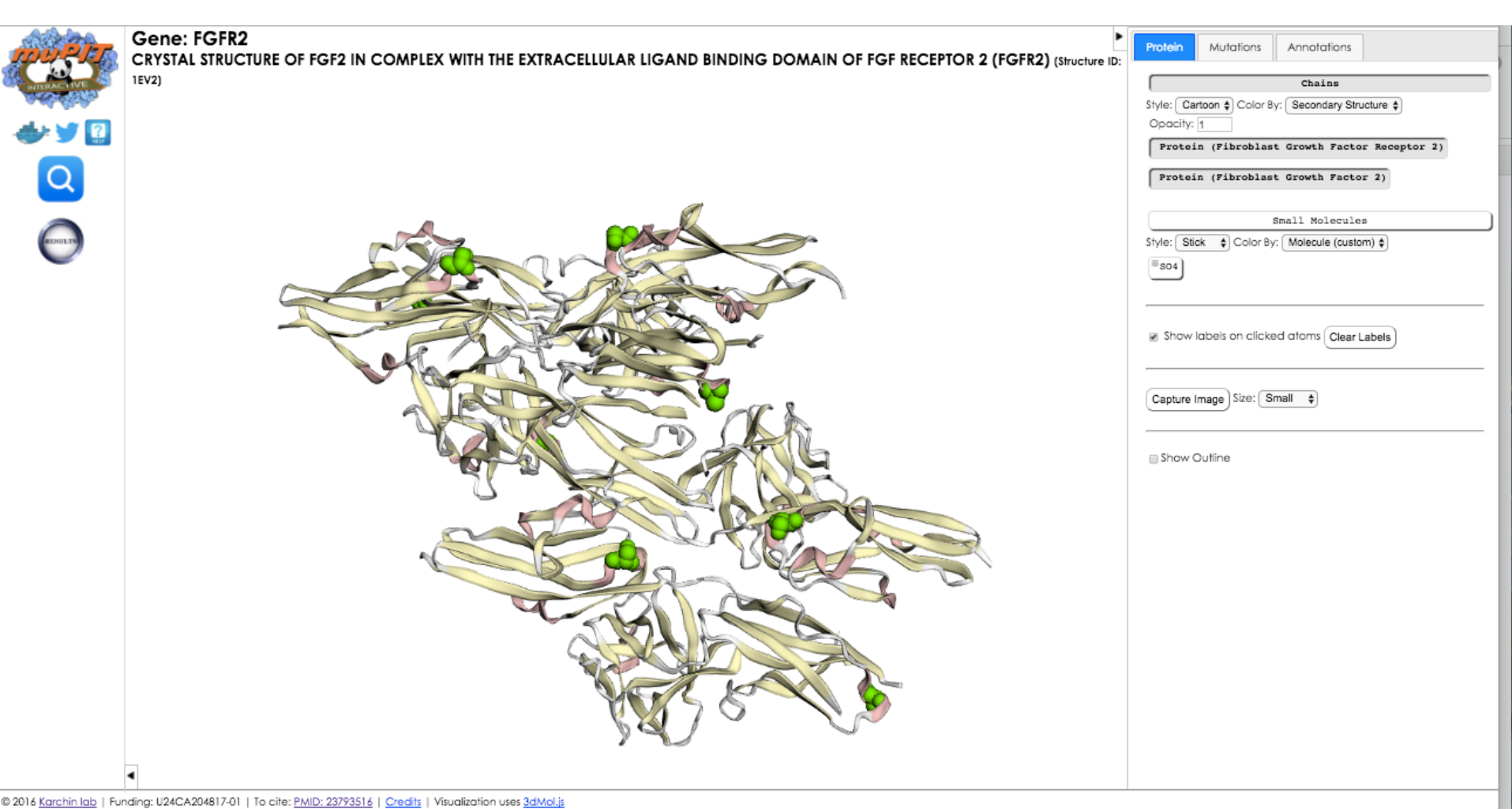
☒ Show labels on clicked atoms

Size:

☐ Show Outline

© 2016 Karchin lab | Funding: U24CA204817-01 | To cite: PMID: 23793516 | Credits | Visualization uses 3dMol.js

- Default structure for FGFR2 (Fibroblast Growth Factor Receptor) in complex with EC domain of FGF
- MuPIT displays the BioMolecule – here it's an octamer.
- Protein tab is selected enabling control over protein



- Default structure for FGFR2 (Fibroblast Growth Factor Receptor) in complex with EC domain of FGF
- MuPIT displays the BioMolecule – here it's an octamer.
- Mutations submitted by user are shown as green spheres

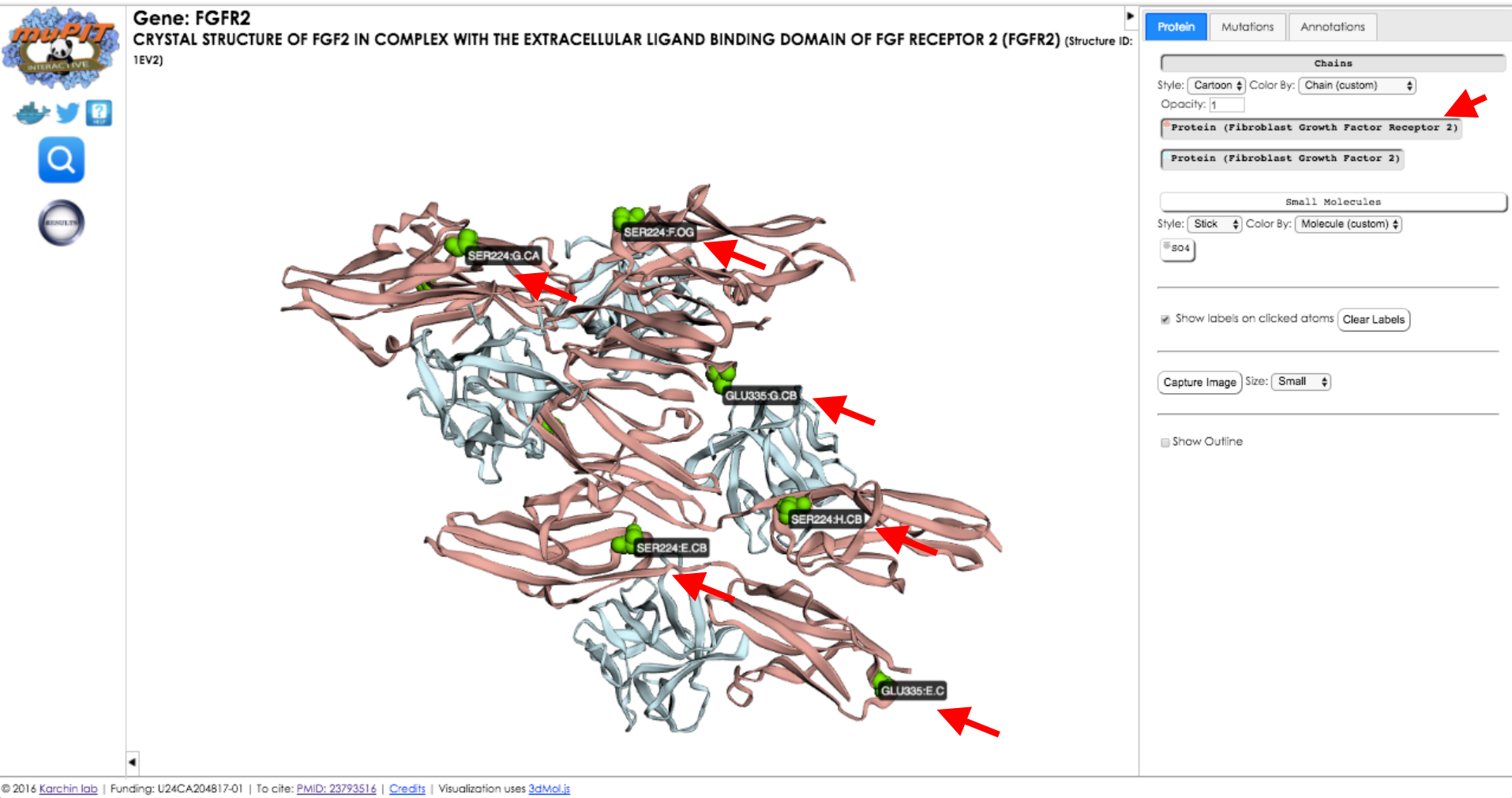


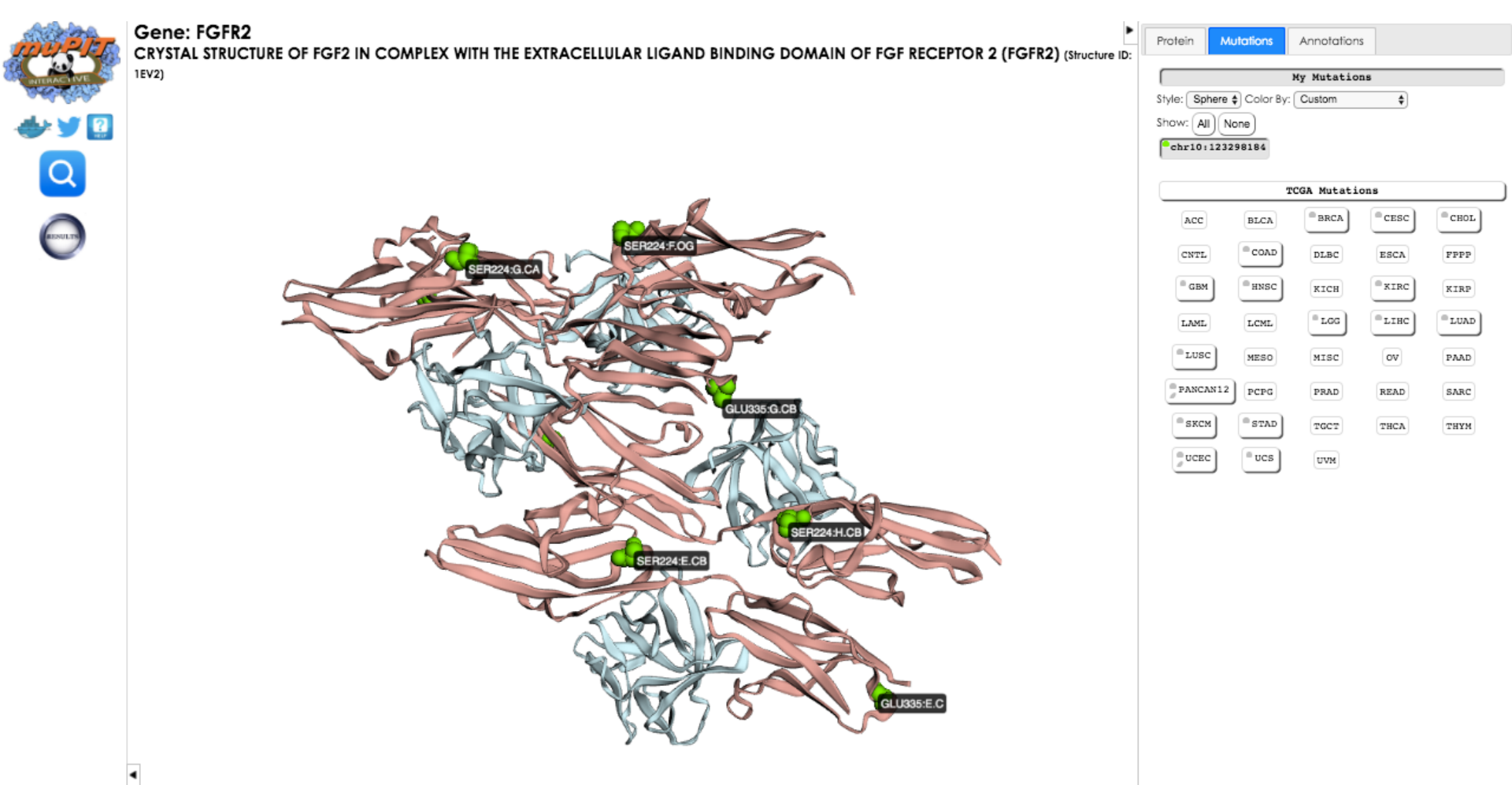
CRYSTAL STRUCTURE OF FGF2 IN COMPLEX WITH THE EXTRACELLULAR LIGAND BINDING DOMAIN OF FGF RECEPTOR 2 (FGFR2) (Structure ID: 1EV2)

FGFR2 29 structures						
⌵	LIGAND-BINDING PORTION OF FIBROBLAST GROWTH FACTOR RECEPTOR 2 IN COMPLEX WITH FGF1				X-ray	1djs
	Structure residues: 682	Query mutations: 1	Annotations: 2	TCGA mutations: 68	TCGA hot regions: 1	Mutations in TCGA hot regions: 0
⌵	Crystal structure of a ternary FGF1-FGFR2-heparin complex				X-ray	1e0o
	Structure residues: 1322	Query mutations: 1	Annotations: 3	TCGA mutations: 130	TCGA hot regions: 2	Mutations in TCGA hot regions: 0
⌵	CRYSTAL STRUCTURE OF FGF2 IN COMPLEX WITH THE EXTRACELLULAR LIGAND BINDING DOMAIN OF FGF RECEPTOR 2 (FGFR2)				X-ray	1ev2
	Structure residues: 2588	Query mutations: 1	Annotations: 12	TCGA mutations: 246	TCGA hot regions: 4	Mutations in TCGA hot regions: 0
⌵	CRYSTAL STRUCTURE OF SER252TRP APERT MUTANT FGF RECEPTOR 2 (FGFR2) IN COMPLEX WITH FGF2				X-ray	1ii4
	Structure residues: 656	Query mutations: 1	Annotations: 3	TCGA mutations: 58	TCGA hot regions: 0	Mutations in TCGA hot regions: 0
⌵	CRYSTAL STRUCTURE OF PRO253ARG APERT MUTANT FGF RECEPTOR 2 (FGFR2) IN COMPLEX WITH FGF2				X-ray	1iii
	Structure residues: 656	Query mutations: 1	Annotations: 3	TCGA mutations: 59	TCGA hot regions: 2	Mutations in TCGA hot regions: 0
⌵	Crystal Structure Analysis of the FGF10-FGFR2b Complex				X-ray	1nun
	Structure residues: 679	Query mutations: 1	Annotations: 0	TCGA mutations: 69	TCGA hot regions: 0	Mutations in TCGA hot regions: 0

Chains

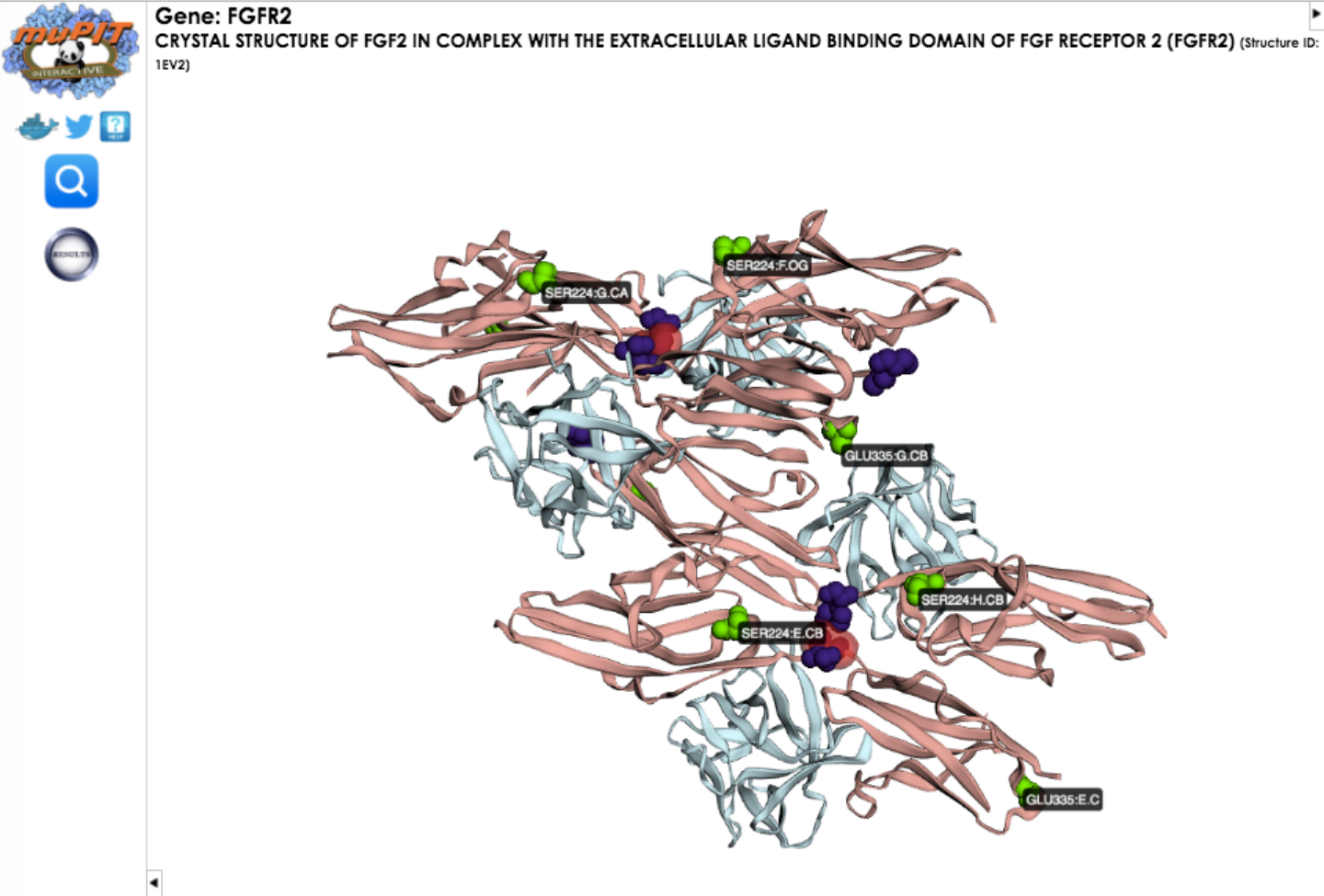
- Browse and select other structures
- See quickly which structures have the most annotations.





© 2016 [Karchin lab](#) | Funding: U24CA204817-01 | To cite: [PMID: 23793516](#) | [Credits](#) | Visualization uses [3dMol.js](#)

- Visualization of mutations from published studies controlled in the mutations tab
- 38 TCGA cancer types currently available
- HotMAPs 3D mutation hotspots available, calculated from TCGA data.



Protein Mutations Annotations

My Mutations

Style: Sphere Color By: Custom

Show: All None

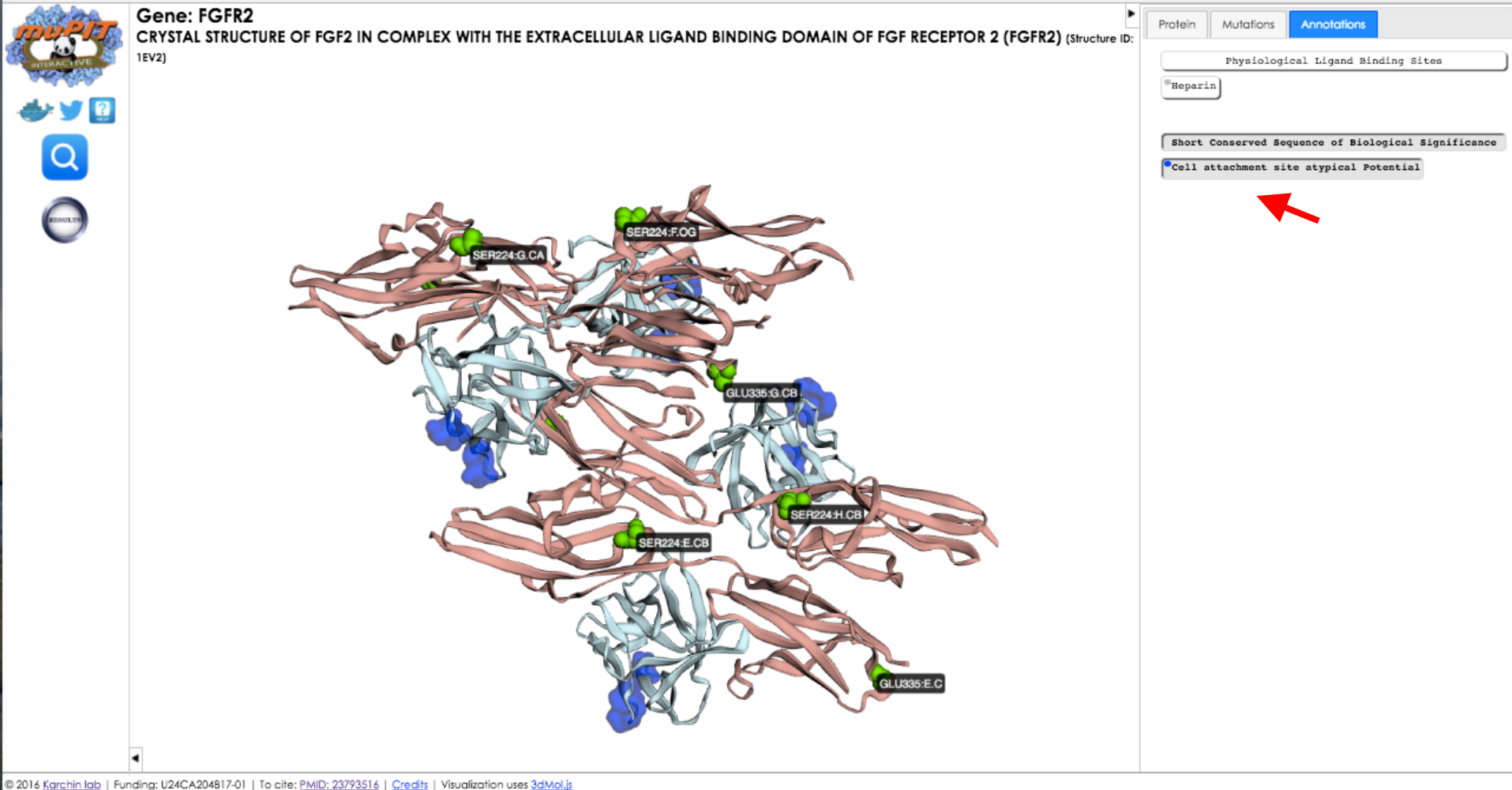
chr10:123298184

TCGA Mutations

ACC	BLCA	BRCA	CEC	CHOL
CNTL	COAD	DLBC	ESCA	FPPP
GBM	HNSC	KICH	KIRC	KIRP
LAML	LCML	LOG	LHIC	LUAD
LUSC	MESO	MISC	OV	PAAD
PANCAN12	PCPG	PRAD	READ	SARC
SKCM	STAD	TGCT	THCA	THYM
UCEC	UCS	UVM		

A red arrow points to the UCEC button in the TCGA Mutations table.

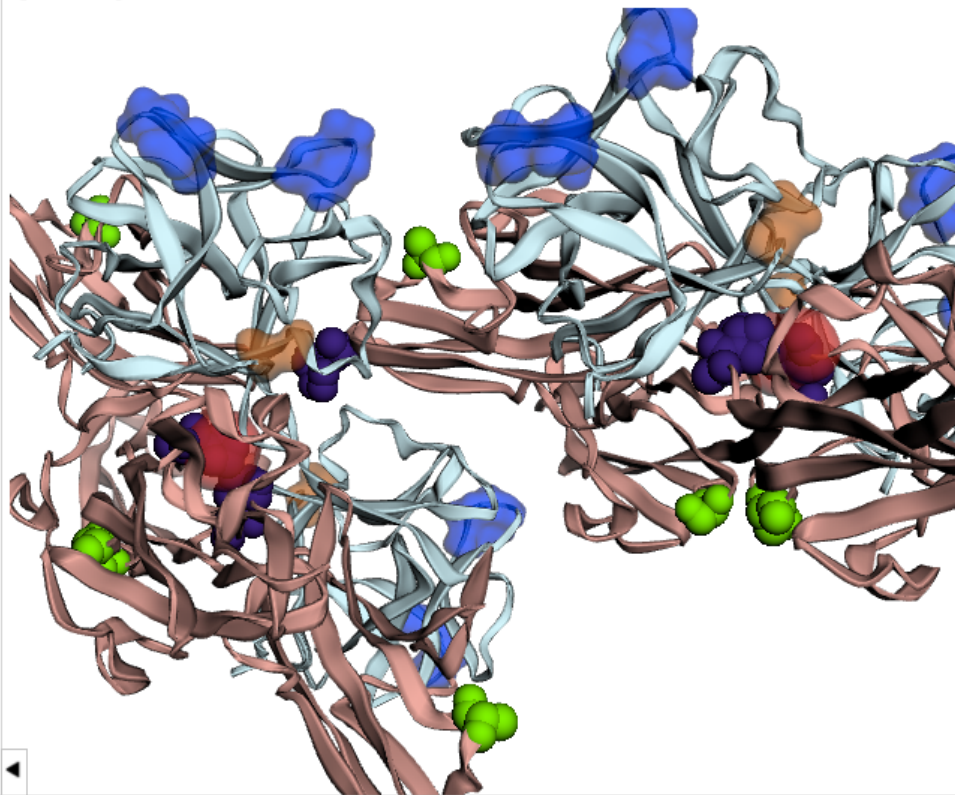
- TCGA UCEC mutations and HotMAPS 3D hot spot



- Visualization of annotated features is controlled in the Annotations tab
- Example: a region annotated as a cell attachment site for FGF is highlighted



Gene: FGFR2
**CRYSTAL STRUCTURE OF FGF2 IN COMPLEX WITH THE
EXTRACELLULAR LIGAND BINDING DOMAIN OF FGF RECEPTOR 2
(FGFR2)** (Structure ID: 1EV2)



Protein Mutations Annotations

Chains

Style: **Cartoon** Color By: **Chain (custom)** Opacity: 1

Protein (Fibroblast Growth Factor Receptor 2)


Protein (Fibroblast Growth Factor 2)

Small Molecules

Style: **Stick** Color By: **Molecule (custom)**

S04

☒ Show labels on clicked atoms **Clear Labels**

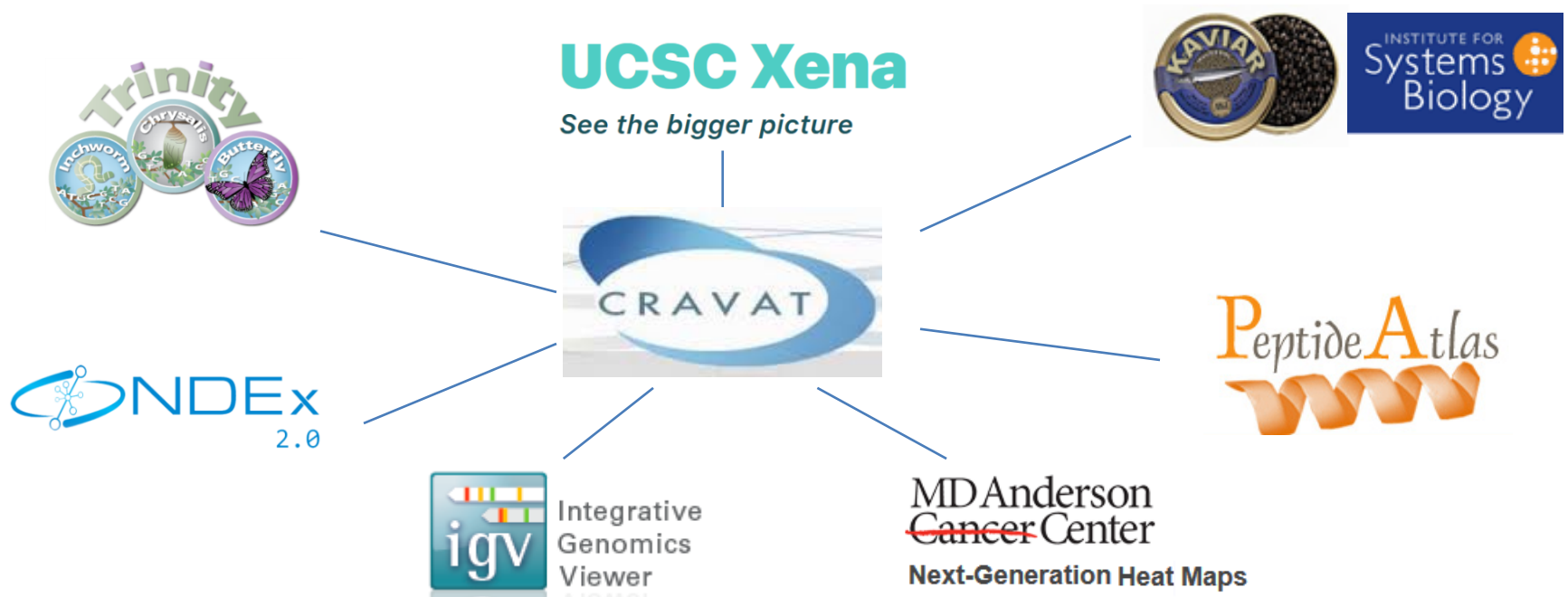
Capture Image Size: **Small** 

☐ Show Outline

- Easy export for publication-quality figures!

CRAVAT / MuPIT Interfaces

- Interoperability is a goal of the ITCR program. (Based on pleas from research community)
- Interconnected tools speed research, enhance user experience, and enable deep knowledge mining.
- Connected tools increase user base for all of the tools.
- Current CRAVAT/MuPIT Integrated Tools:



Incorporating MuPIT into your tool

- Link out to us with structured urls to get our default viewer for
 - **Single Genomic Location to 3D Structures Mapping**
 - **Multiple Genomic Locations to 3D Structures Mapping**
 - **Uniprot Protein Coordinates to 3D Structures Mapping**
 - **Gene to 3D Structures Mapping**
 - **Direct Viewing of a 3D Structure**

Incorporating MuPIT into your tool

- Link out to us with structured urls to get our default viewer for
 - **Single Genomic Location to 3D Structures Mapping**



Incorporating MuPIT into your tool

- Display MuPIT in an iframe within your own portal

Summary Gene Variant Noncoding Error

Job Info: eprice_20170202_062906

PRKCB

Has a mutation in a

Most severe

Open Close X

CHASM cancer p-value (sensitivity)

6 0

6 0

5 0

Gene: PRKCA
Crystal structure of a C2 domain of a protein kinase C alpha (PRKCA) from Homo sapiens at 1.90 A resolution (Structure ID: 4DNL)

muPIT INTERACTIVE

Gene Info+
Variant Impact+
VEST+
CHASM+
Disease Association+
Structure+
Study+

Q

RESULTS

Chains

Style: Cartoon Color By: Secondary Structure Opacity: 1

Protein Kinase C Alpha Type

Small Molecules

Style: Stick Color By: Molecule (custom)

MSE NA

☒ Show labels on clicked atoms Clear Labels

Capture Image Size: Small

© 2016 Karchin Lab | Funding: U24CA204817-01 | To cite: PMID: 23793516 | Credits | Visualization uses 3dMol.js

CGC Inheritance: None
CGC Somatic Tumors: None
CGC Germline Tumors: None
COSMIC gene hits: 260

xx x ↔ x x

Incorporating MuPIT into your tool

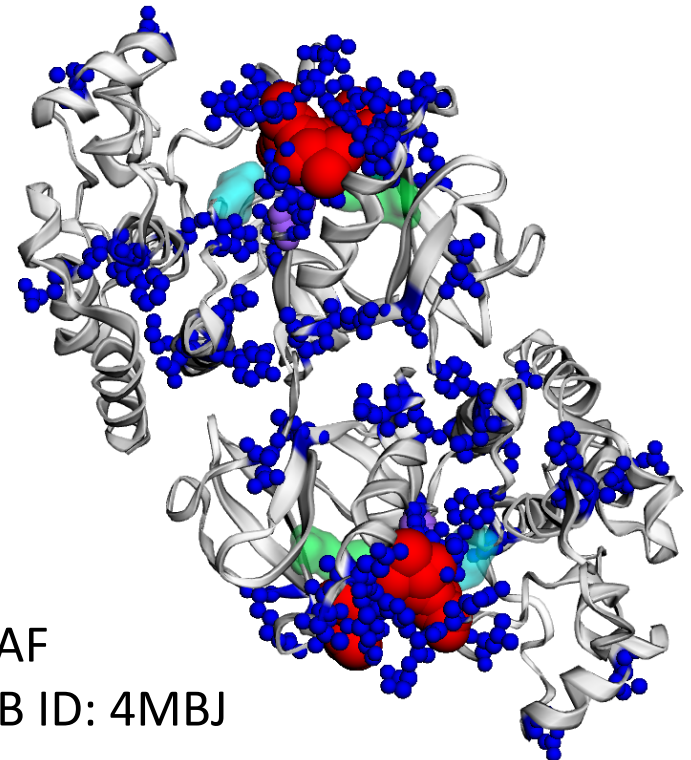
- Link out to us with structured urls to get specially configured visualizations



Dr. Jing Zhu



UCSC Xena
Browser



BRAF
PDB ID: 4MBJ

Small molecules
mutation
recurrence

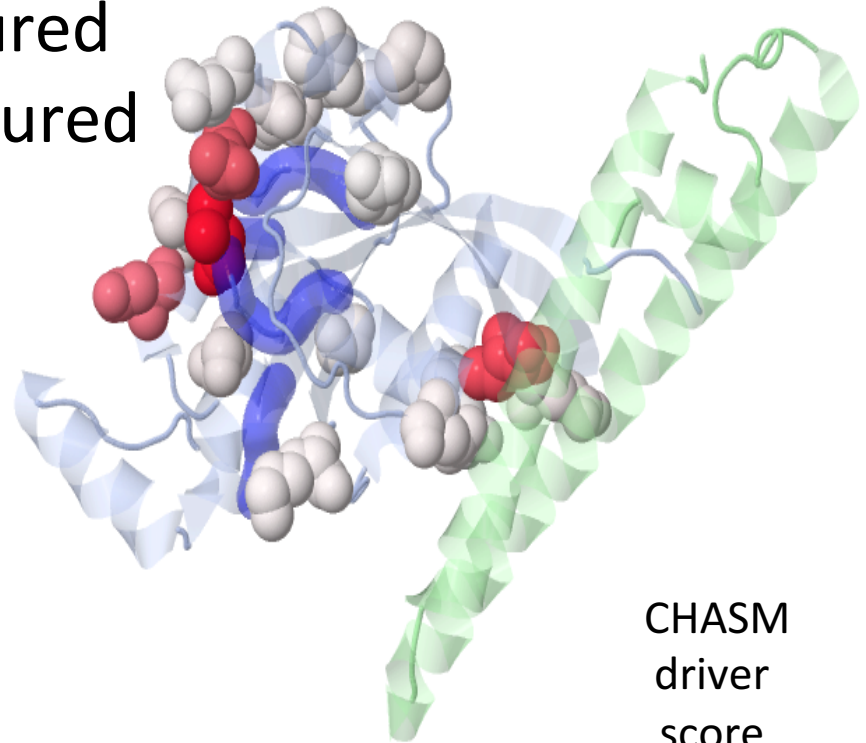
TCGA PanCan Mutations

Incorporating MuPIT into your tool

- Link out to us with structured urls to get specially configured visualizations



Dr. Jing Zhu



CHASM
driver
score

■ q value 0
□ q value 0.5

RHOA GTPase and effector
domain of PKN/PRK1
PDB ID: 1CXZ



UCSC Xena
Browser

TCGA PanCan Mutations

MuPIT HTTP Integration

Link to MuPIT 3D Visualization with URL:

For a mutation, just pass its genomic position (one or more):

http://mupit.icm.jhu.edu/MuPIT_Interactive/?gm=chr17:7577510

Just link to a gene:

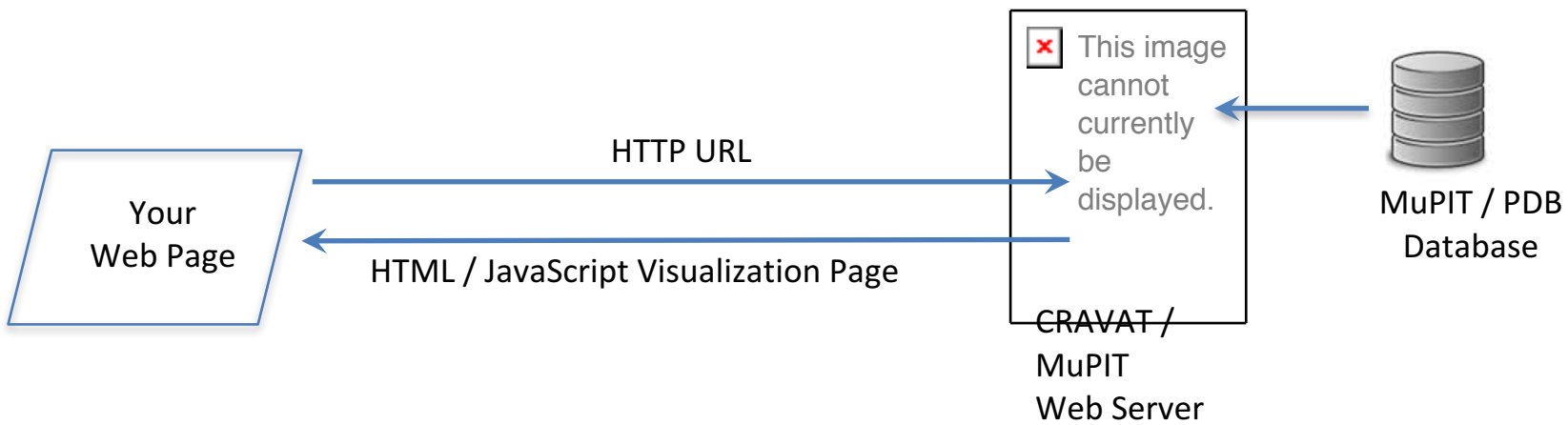
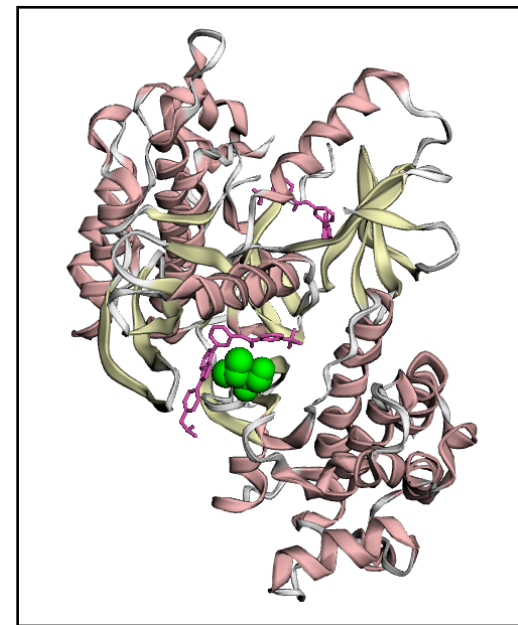
http://mupit.icm.jhu.edu/MuPIT_Interactive?gene=TP53

Link using a protein position:

http://mupit.icm.jhu.edu/MuPIT_Interactive/?gm=P14317:235&protquery=y

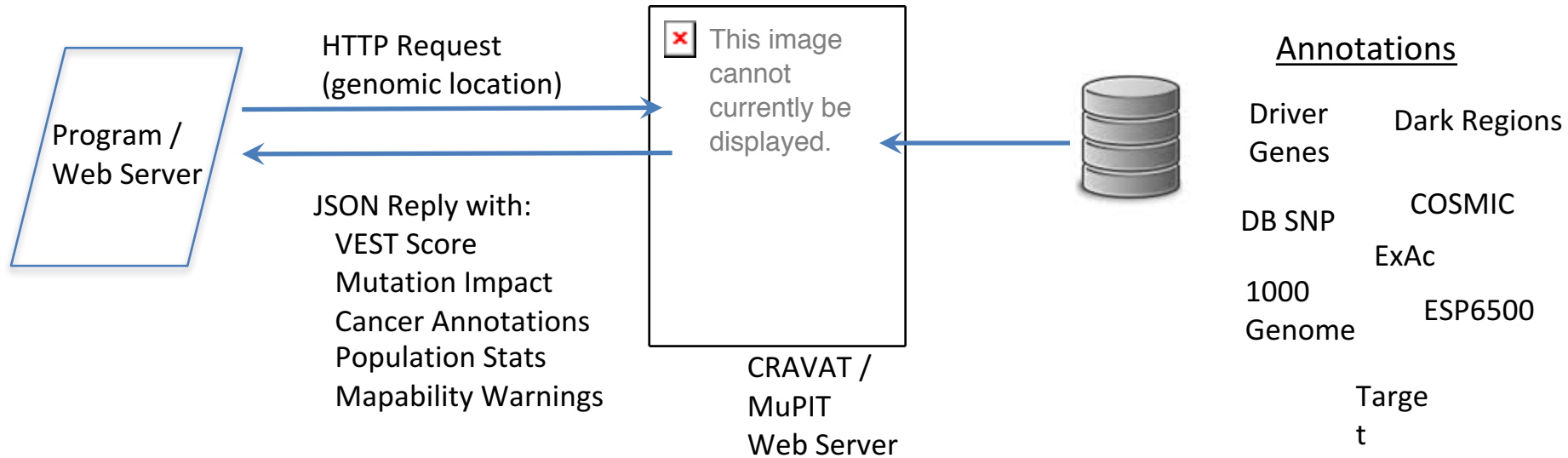
To check that a structure exists:

http://mupit.icm.jhu.edu/MuPIT_Interactive/rest/showstructure/check?pos=chr1:69094



CRAVAT Web Service

A web service call provides fast access to single mutation annotation to external systems.



CRAVAT Web Service call

(substitution):

http://www.cravat.us/CRAVAT/rest/service/query?mutation=chr22_30421786_+A

CRAVAT Web Service call

(indel):

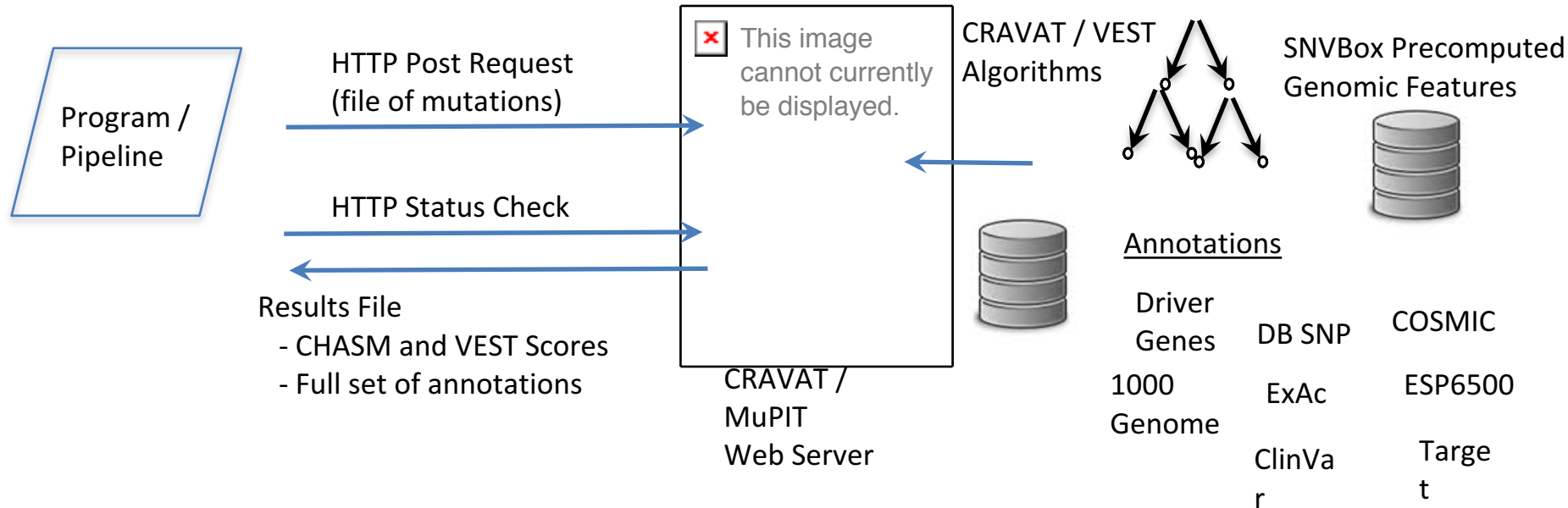
http://www.cravat.us/CRAVAT/rest/service/query?mutation=chr22_30421786_+_TAC

JSON Reply:

```
{ "Chromosome": "chr22", "Position": "30421786", "Strand": "+", "Reference base(s)": "-", "Alternate base(s)": "TAC", "HUGO symbol": "MTMR3", "Sequence ontology transcript": "NM_021090.3", "Sequence ontology protein change": "1198", "Sequence ontology": "II", "Sequence ontology all transcripts": "NM_153051.2:1161(II), NM_153050.2:1170(II), ENST00000323630:1062(II), ENST000003351488:1161(II), ENST00000333027:1170(II), ENST00000406629:1170(II), ENST00000401950:1198(II)", "ExAC total allele frequency": "0", "ExAC allele frequency (African/African American)": "0", "ExAC allele frequency (Latino)": "0", "ExAC allele frequency (East Asian)": "0", "ExAC allele frequency (Finnish)": "0", "ExAC allele frequency (Non-Finnish European)": "0", "ExAC allele frequency (Other)": "0", "ExAC allele frequency (South Asian)": "0", "1000 Genomes allele frequency": "0", "ESP6500 allele frequency (European American)": "0", "ESP6500 allele frequency (African American)": "0", "Transcript in COSMIC": "", "Protein sequence change in COSMIC": "", "Occurrences in COSMIC [exact nucleotide change]": "", "Occurrences in COSMIC by primary sites [exact nucleotide change]": "", "Mappability Warning": "", "Driver Genes": "", "TARGET": "", "dbSNP": "", "MuPIT Link": "" }
```

CRAVAT RESTful Service

Method for automated submission of large sets of mutations



Any programming language (Python example)

```
# Job submission via POST with an input file
import requests
r=requests.post('http://dev.cravat.us/rest/service/submit',
                files={'inputfile':open('input_file/vcf_input.txt')},
                data={'email':'rkim@insilico.us.com','analyses':''})

# Checking job status
import requests
r=requests.get('http://www.cravat.us/rest/service/status',
               params={'jobid':'rkim_20160315_103245'})

r.text # contains the job status as a string. Includes URL to get results when done.
```

CRAVAT Mutation Browser

In our upcoming release, link to interactive mutation detail web page:

http://cravat.us/CRAVAT/variant.html?variant=chr22_40814500_-

A G

Show Summary

Variant Info

UID: None

Gene: [MKL1](#)

Chromosome: chr22

Position: 40814500

Ref base(s): A

Alt base(s): G

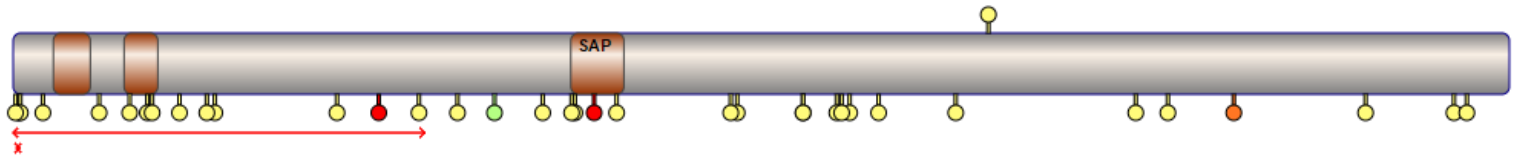
dbSNP: [rs878756](#)

Sample: None

samples: None

Protein Diagram

My Mutations TCGA Mutations:



MuPIT

Mutation Call Quality

Mappability: None

QUAL: N/A

Filter: N/A

Zygosity: N/A

Variant reads: N/A

Sample reads: N/A

Disease Association

Driver genes: None

TARGET: None

PubMed gene hits: None

ClinVar Clin. Sig.: None

ClinVar Diseases: None

ClinVar Dis. Ref.: None

COSMIC variant hits: [1](#)

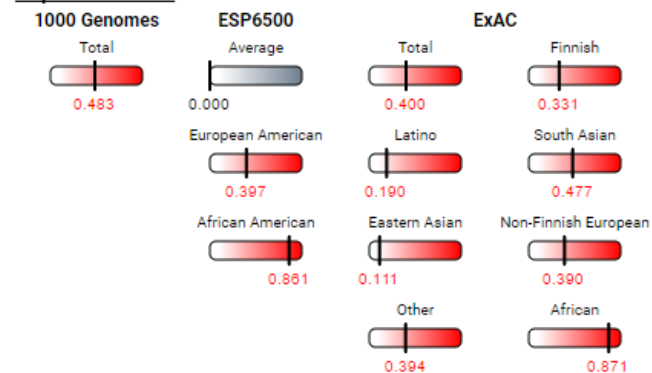
Tissue # samples

stomach 1

Variant Impact & Pathogenicity & Cancer Driver

Transcript	Sequence Ontology	Protein Change	VEST		CHASM	
			Score	p-value	Score	p-value
NM_020831.3	Missense	S648G	0.990	0.005	0.414	0.332
ENST00000396617	Missense	S648G	0.414	0.113	0.260	0.780
ENST00000402042	Missense	S598G	0.987	0.006	0.414	0.332
ENST00000407029	Missense	S648G	0.990	0.005	0.360	0.479
ENST00000355630	Missense	S648G	0.986	0.006	0.414	0.332

Population Stats



Platform Integration

By integrating with common platforms – we become open to users of the platform.



CRAVAT tool in Galaxy
Toolshed

The screenshot shows the Galaxy web interface with a table of genomic data. The table has columns for ID, Chromosome, Position, Strand, Reference base(s), Alternate base(s), Sample, 1000 Genomes allele frequency, Driver Genes, and ESP6500 allele frequency. The data includes various genomic regions and their associated frequencies.

ID	Chromosome	Position	Strand	Reference base(s)	Alternate base(s)	Sample	1000 Genomes allele frequency	Driver Genes	ESP6500 allele frequency
rs6054257_A	chr20	14370	+	G	A	A	0		
rs6054257_G	chr20	14370	+	G	A	A	0		
VAR21_A	chr20	17330	+	T	A	A	0		
rs6040355-1_N	chr20	1110696	+	A	G	N	0		
rs6040355-2_N	chr20	1110696	+	A	T	N	0		
rs6040355-1_A	chr20	1110696	+	A	G	A	0		
rs6040355-2_A	chr20	1110696	+	A	T	A	0		
rs6040355-2_G	chr20	1110696	+	A	T	0	0		
microRNA1-1_N	chr20	1234568	+	TC	-	N	0		
microRNA1-2_A	chr20	1234570	+	T	A	Unknown	0		
microRNA1-1_G	chr20	1234568	+	TC	-	0	0		
TR1_A	chr22	30421786	+	A	T	A	0.0077875400893390178680419921875		0.0226962994
TR1_G	chr22	30421786	+	A	T	0	0.0077875400893390178680419921875		0.0226962994
VAR26_N	chr22	29446079	+	A	G	N	0		0.001040040049
VAR26_G	chr22	29446079	+	A	G	0	0		0.001040040049
TR3_A	chr22	40814500	+	T	C	A	0.483227014541625975825		0.86064
VAR28_N	chr22	40815256	+	C	T	N	0.0279552992433309555037109375		0.083409
VAR28_G	chr22	40815256	+	C	T	A	0.0279552992433309555037109375		0.083409
VAR29-1_N	chr12	123466293	+	GAGAGAGA	-	N	0		0.132282
VAR29-1_A	chr12	123466293	+	GAGAGAGA	-	A	0		0.132282
VAR29-2_A	chr12	123466293	+	GAGAGAGA	-	A	0		0.132282
VAR29-1_G	chr12	123466293	+	GAGAGAGA	-	0	0		0.132282
VAR29-3_G	chr12	123466293	+	GAA	-	0	0		0.132282



CRAVAT in Docker
(multi-container integration)

The screenshot shows the Docker Hub page for the repository "karchinlab/cravatmupit". It includes a search bar, repository details, a short description, a full description, and installation instructions for Docker and CRAVAT/MuPIT.

Short Description

Dockerized CRAVAT and MuPIT

Full Description

Note: The docker version of CRAVAT requires a large reference database so 200 GB of disk space will be required.

How to Install Docker and CRAVAT/MuPIT Container

Please scroll to read the installation instructions for your OS.

> Windows Installation

Installing Docker

To install Docker, please follow the guide at <https://www.docker.com/docker-toolbox>.

After installing Docker, please allocate at least 2 Gb of memory to the Docker VM. By default <https://hub.docker.com/r/karchinlab/cravatmupit/> will require using Oracle's VirtualBox (installed with Docker). In

Docker Pull Command

```
docker pull karchinlab/cravatmupit
```

Owner

karchinlab

Thank you!

Karchin Lab

Collin Tokheim
David Masica
Christopher Douville
Noushin Niknafs
Violeta Beleva-Guthrie
Rohit Bhattachaya
Ashok Sivakumar

In Silico Solutions

Michael Ryan
Rick Kim
Derek Gyga
Kyle Moad

ITCR

Jerry Li
Julie Klemm



Funding: NCI U24CA204817