ITCR Annual Meeting

Informatics Tools for High-throughput Analysis of Cancer Mutations

Karchin Lab
Departments of Biomedical Engineering and Oncology
Institute of Computational Medicine
Johns Hopkins University

Link to video demo:

https://www.youtube.com/watch?v=xovmIy11Bcs

Goals for U01 funded in 2012

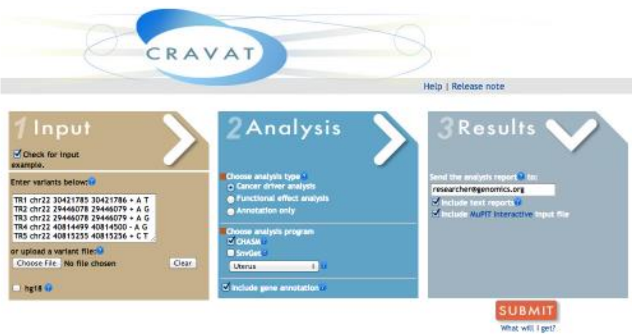
1. Integrate tools we developed to prioritize cancer missense mutations
    a. Create single user-friendly application that provides analysis of large-scale data
    b. Make it accessible to research scientists who are not bioinformatics experts

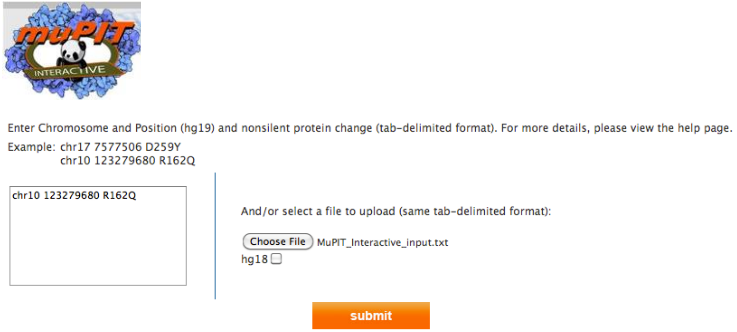
Goals for U01 funded in 2012

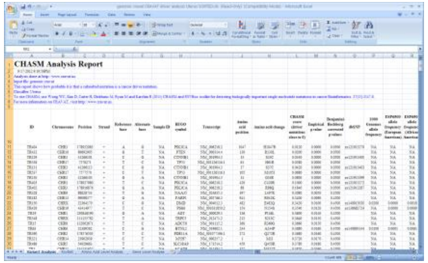
2. Broaden the tools scientifically
    a. Handle small mutations in cancer exome beyond missense mutations
    b. Identify important genes and pathways
    c. Enable cohort-level analysis

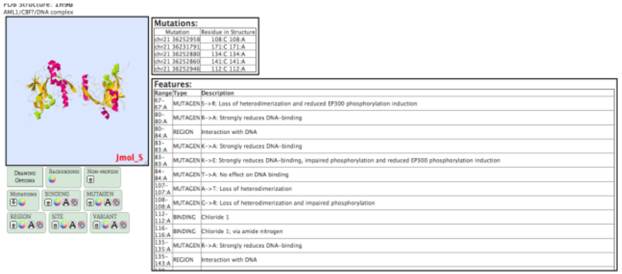
2012

Integrate tools to prioritize cancer missense mutations









2012

Integrate tools to prioritize cancer missense mutations

Machine-learning of missense mutation impact

Cancer-Specific High-Throughput Annotation of Somatic Mutations: Computational Prediction of Driver Missense Mutations

Hannah Carter[1], Sining Chen[2,3], Leyla Isik[1], Svitlana Tyekucheva[3], Victor E. Velculescu[4], Kenneth W. Kinzler[4], Bert Vogelstein[4], and Rachel Karchin[1]

Identifying Mendelian disease genes with the Variant Effect Scoring Tool

Hannah Carter, Christopher Douville, Peter D Stenson, David N Cooper and Rachel Karchin

RESEARCH ARTICLE

Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses

Siân Jones[1,*], Xiaosong Zhang[1,*], D. Williams Parsons[1,2,*], Jimmy Cheng-Ho Lin[1,*], Rebecca J. Leary[1,*], Philipp Angenendt[1,*], Parminder Mankoo[3], Hannah Carter[3], Hirohiko Kamiyama[4], Antonio Jimeno[1], Seung-Mo Hong[4], Baojin Fu[4], Ming-Tseh Lin[4], Eric S. Calhoun[1], Mihoko Kamiyama[4], Kimberly Walter[4], Tatiana Nikolskaya[5], Yuri Nikolsky[6], James Hartigan[7], Douglas R. Smith[7], Manuel Hidalgo[1], Steven D. Leach[1,8], Alison P. Klein[1,4], Elizabeth M. Jaffee[1,4], Michael Goggins[1,4], Anirban Maitra[1,4], Christine Iacobuzio-Donahue[1,4], James R. Eshleman[1,4], Scott E. Kern[1,4], Ralph H. Hruban[1,4], Rachel Karchin[3], Nickolas Papadopoulos[1], Giovanni Parmigiani[1,9], Bert Vogelstein[1,†], Victor E. Velculescu[1,†], Kenneth W. Kinzler[1,†]

RESEARCH ARTICLE

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

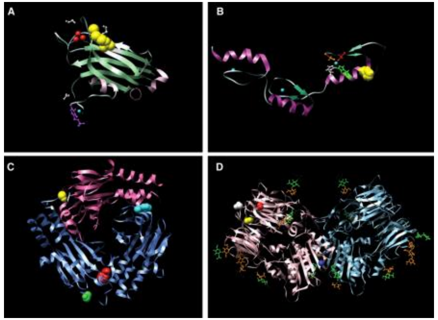
D. Williams Parsons[1,2,*], Siân Jones[1,*], Xiaosong Zhang[1,*], Jimmy Cheng-Ho Lin[1,*], Rebecca J. Leary[1,*], Philipp Angenendt[1,*], Parminder Mankoo[3], Hannah Carter[3], I-Mei Siu[4], Gary L. Gallia[4], Alessandro Olivi[4], Roger McLendon[5], B. Ahmed Rasheed[5], Stephen Keir[5], Tatiana Nikolskaya[6], Yuri Nikolsky[7], Dana A. Busam[8], Hanna Tekleab[8], Luis A. Diaz Jr.[1], James Hartigan[9], Doug R. Smith[9], Robert L. Strausberg[8], Suely Kazue Nagahashi Marie[10], Sueli Mieko Oba Shinjo[10], Hai Yan[5], Gregory J. Riggins[4], Darell D. Bigner[5], Rachel Karchin[3], Nick Papadopoulos[1], Giovanni Parmigiani[1], Bert Vogelstein[1,†], Victor E. Velculescu[1,†], Kenneth W. Kinzler[1,†]

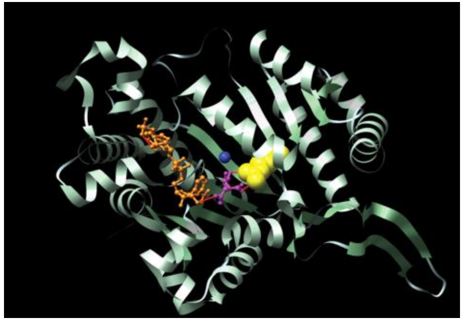
Missense mutation analysis and protein structure

MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures

Noushin Niknafs, Dewey Kim, RyangGuk Kim, Mark Diekhans, Michael Ryan, Peter D. Stenson, David N. Cooper, Rachel Karchin

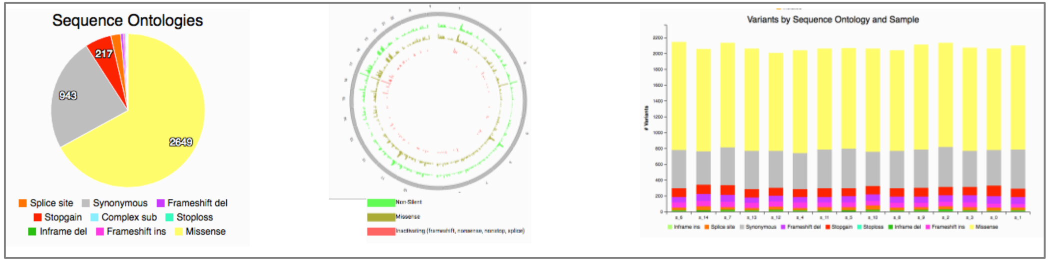
Clustering patterns
Proximity to ligands and interfaces

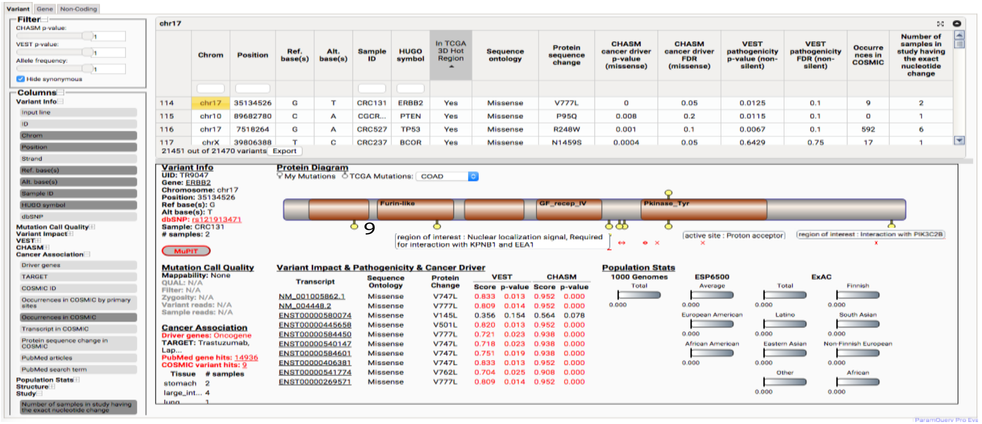


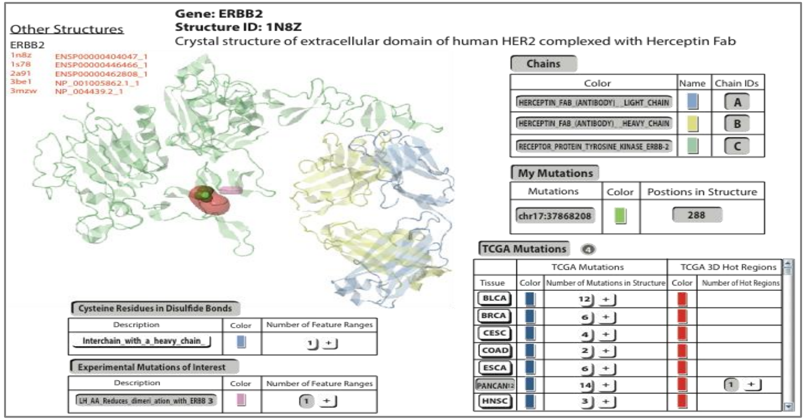


2016

Integrated user-friendly application with interactive results explorer







2016

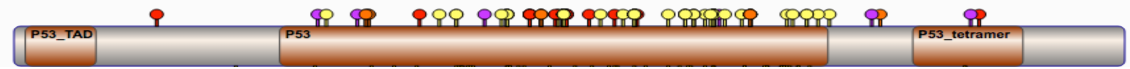
Broaden scientific scope

- Annotation/scoring of all small non-silent mutation types
  - Machine learning classifiers for specific mutation consequence types
  - Integrated P-values support a unified prioritization

Chromosome	Position	Strand	Reference base(s)	Alternate base(s)	Sample ID	HUGO symbol	Sequence ontology	Sequence ontology protein sequence change	VEST pathogenicity p-value (non-silent)	VEST pathogenicity FDR (non-silent)
chr12	2795367	+	C	T	SS6004357	CACNA1(	SG	R1989*	0.0007	0.2
chr4	3156068	+	C	T	SS6004357	HTT	SG	R1183*	0.0015	0.2
chr7	151921149	+	G	A	SS6004357	MLL3	SG	R1092*	0.0020	0.2
chr12	83251120	+	C	G	SS6004356	TMTC2	MS	R139G	0.0057	0.2
chr2	129026227	+	G	T	SS6004357	HS6ST1	MS	R249S	0.0058	0.2
chr2	129026227	+	G	T	SS6004356	HS6ST1	MS	R249S	0.0058	0.2
chr17	19284136	+	G	C	SS6004356	MAPK7	MS	R205P	0.0068	0.2
chr5	169135251	+	C	T	SS6004357	DOCK2	MS	R488C	0.0073	0.2
chr11	108192148	+	G	A	SS6004357	ATM	SS	_2191_	0.0075	0.2
chr4	134072602	+	G	C	SS6004357	PCDH10	MS	R436P	0.0078	0.2
chr11	116629090	+	T	C	SS6004357	BUD13	MS	D465G	0.0080	0.2
chr11	116629090	+	T	C	SS6004356	BUD13	MS	D465G	0.0080	0.2
chr5	140052204	+	C	A	SS6004357	DND1	MS	G144C	0.0083	0.2
chr17	14110451	+	A	C	SS6004356	COX10	MS	H418P	0.0083	0.2
chr5	7706895	+	G	A	SS6004357	ADCY2	MS	R383H	0.0088	0.2
chr1	233497916	+	C	T	SS6004357	KIAA180	MS	R477W	0.0089	0.2
chr1	233497916	+	C	T	SS6004356	KIAA180	MS	R477W	0.0089	0.2

Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-Indel)

Christopher Douville, David L. Masica, Peter D. Stenson, David N. Cooper, Derek M. Gygax, Rick Kim, Michael Ryan, Rachel Karchin



2016

Broaden scientific scope

- Identify important genes and pathways for cohort-level analysis

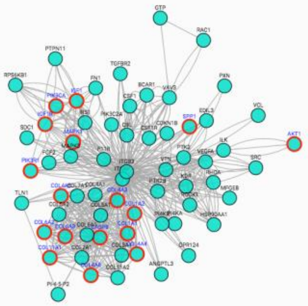
	HUGO symbol	Number of variants	MuPIT Link	Most severe sequence ontology	VEST pathogenicity composite p-value (non-silent)	VEST pathogenicity FDR (non-silent)	Driver genes	Occurrences in COSMIC by primary sites	Number of samples in study having the gene mutated
1	TP53	28	Yes	Frameshft del	0	0.05	TSG	large_intestine(3317);vulva(122);pleura(21);peritoneum...	28
2	EGFR	6	Yes	Missense	0	0.05	Oncogene	large_intestine(270);pleura(5);peritoneum(11);endome...	6
3	HIST1H3B	3	Yes	Missense	0	0.05	Oncogene	upper_aerodigestive_tract(1);cervix(2);large_intestine(...	3
4	CACNA1E	4	Yes	Missense	0	0.05		large_intestine(187);pleura(1);peritoneum(1);endometr...	4
5	PIK3CA	31	Yes	Missense	0	0.05	Oncogene	large_intestine(1792);vulva(7);pleura(1);peritoneum(2)...	31
6	DNAH2	4	Yes	Missense	0.0001	0.05		large_intestine(228);peritoneum(1);endometrium(103)...	4
7	TRIP12	3	Yes	Missense	0.0001	0.05		upper_aerodigestive_tract(4);cervix(6);large_intestine(...	3
8	MAP1B	3		Missense	0.0002	0.05		large_intestine(149);peritoneum(1);endometrium(57);l...	3
9	GNE	3	Yes	Missense	0.0002	0.05		upper_aerodigestive_tract(2);cervix(3);large_intestine(...	3
10	RNASE4	2	Yes	Missense	0.0002	0.05		large_intestine(5);stomach(5);central_nervous_system...	2
11	MLL	3		Stopgain	0.0002	0.05			3
12	INSR	3	Yes	Missense	0.0003	0.1		large_intestine(72);endometrium(27);lung(33);skin(53)...	3
13	EPC1	2	Yes	Missense	0.0004	0.1		upper_aerodigestive_tract(1);cervix(3);large_intestine(...	2
14	TRPM3	3	Yes	Missense	0.0004	0.1		large_intestine(119);endometrium(48);lung(55);skin(7...	3
15	ITPR2	3	Yes	Splice site	0.0004	0.1		large_intestine(135);endometrium(46);lung(63);skin(5...	3
16	TUBD1	3	Yes	Missense	0.0004	0.1		thyroid(1);cervix(4);large_intestine(23);stomach(7);cen...	3

Multiple criteria to sort genes by importance in a cohort

VEST

Ndex Networks (vest_composite_pvalue)

Pathway	p-value	Genes
Syndecan-1-mediated signaling events	0.0002	COL4A3, COL11A1, LAMA5, COL4A6, COL4A5, COL4A4, COL1A2, COL1A1, HGF, MAPK1, COL6A3, COL6A2, MMP1
E-cadherin signaling in keratinocytes	0.0005	PIK3R1, PIK3CA, FYN, AKT2, AKT1, PLCG1, CTNND1, CASR, EGFR
Integrins in angiogenesis	0.0009	IGF1R, IGF1, COL4A3, COL11A1, AKT1, PIK3CA, CASP8, COL4A6, COL4A5, COL4A4, SPP1, PIK3R1, COL1A1, MAPK1, COL6A3, COL6A2, COL1A2
PLK3 signaling events	0.0013	PLK3, CCNE1, TP53
Class IB PI3K non-lipid kinase events	0.0013	PIK3CG, MAPK1, PDE3B
Reelin signaling pathway	0.0025	MAP1B, PIK3CA, FYN, GSK3B, GRIN2B, AKT1, GRIN2A, RELN, PIK3R1
ATM pathway	0.0028	BRCA1, BLM, ATM, MDC1, H2AFX, TERF2, RAD50, ABL1, TOP3A, UIMC1
VEGFR3 signaling in lymphatic endothelium	0.0030	FIGF, SOS1, PIK3CA, AKT1, MAPK1, COL1A2, PIK3R1, COL1A1
LKB1 signaling events	0.0036	RPTOR, TP53, BRSK1, SIK3, ESR1, SIK2, GSK3B, MAP2, MTOR, MYC, STRADA
Regulation of CDC42 activity	0.0044	SPATA13, FGD1, ARHGEF9, MCF2, ARHGEF7, ITSN2, PLCG1, DOCK10, DOCK11
Class I PI3K signaling events	0.0055	ARAP3, PIK3R1, INPPL1, PIK3CA, FYN, PIK3CG, ADAP1, PTEN, PLCG1, PLEKHA1, PLEKHA2, CYTH3
BARD1 signaling events	0.0070	BRCA1, ATM, TP53, CCNE1, FANCF, ATR, RAD51, RAD50
Beta1 integrin cell surface interactions	0.0087	IGSF8, COL11A1, FBN1, COL4A6, COL4A5, COL4A4, SPP1, COL4A3, COL1A1, LAMB1, COL1A2, LAMB3, COL6A2, COL6A3, LAMA5

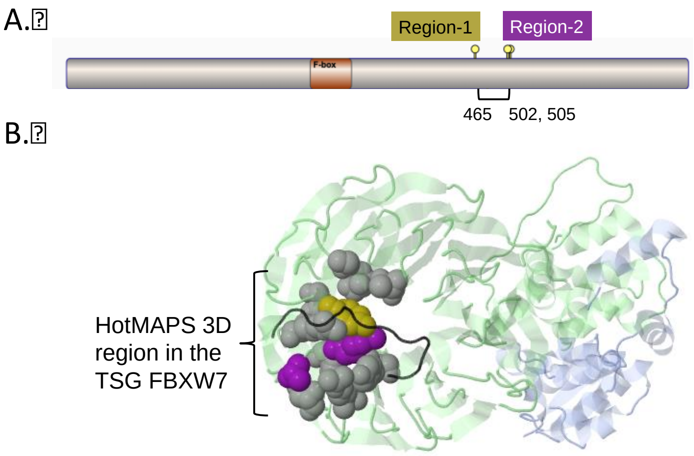


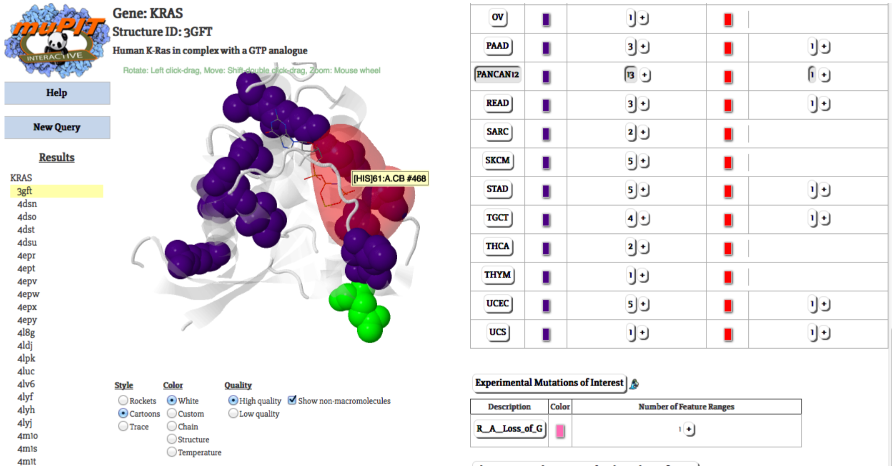
Find pathways enriched for important genes

2016

Broaden scientific scope

- HotMAPS algorithm and MuPIT viewing of 3D hotspot mutation regions in cancer cohorts
  - Positional clustering of somatic mutations is a signal of positive selection
  - Clustering in 3D detects hotspots that are missed in 1D





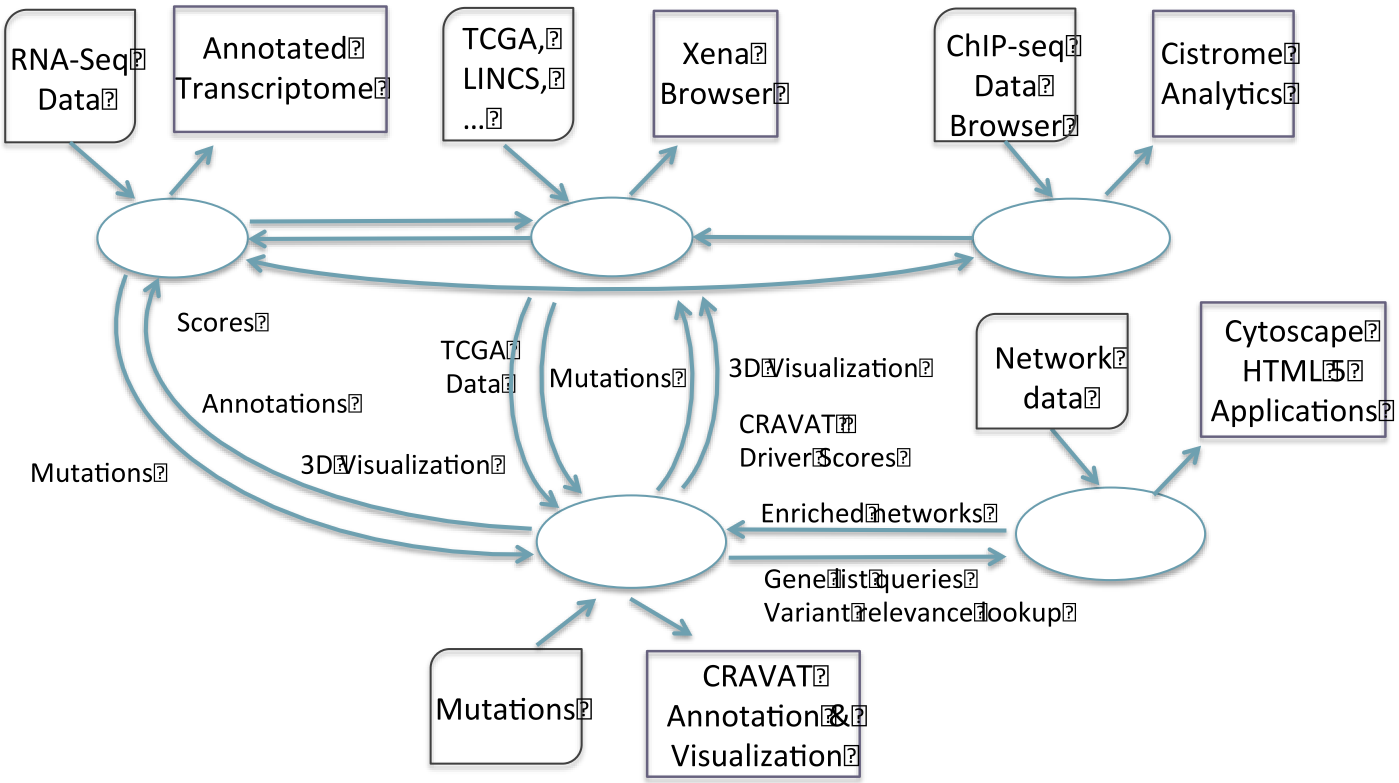
Cancer Res. 2016 Apr 28. pii: canres.3190.2015. [Epub ahead of print]

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 MC[3], Masica D[1], Karchin R[4].

2016

ITCR project collaborations



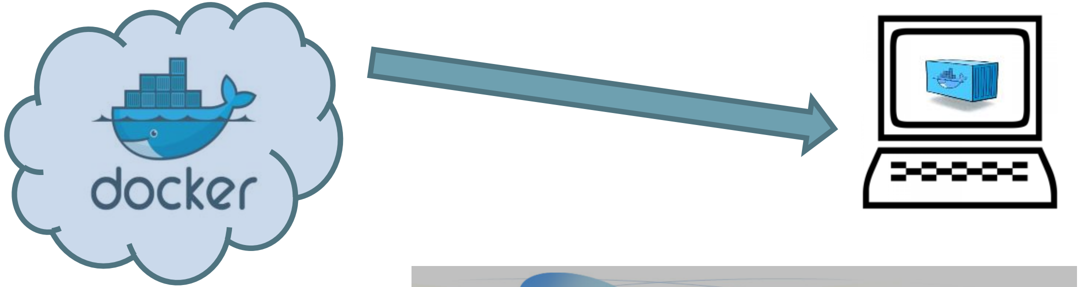
Administrative supplements

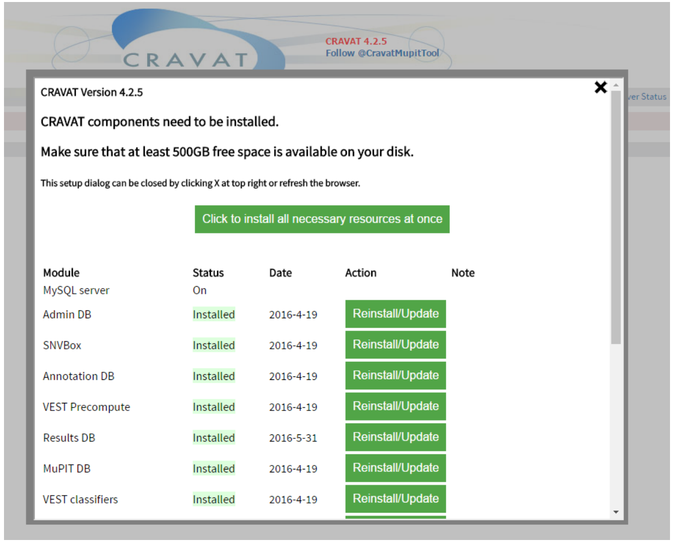
2016

Broaden user base

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

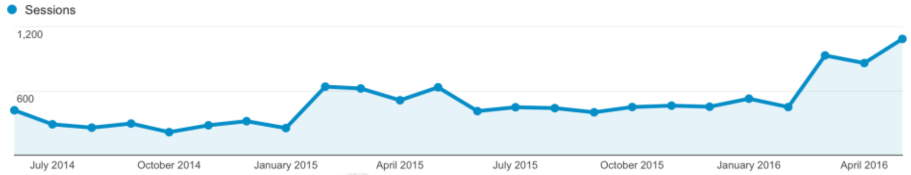
- Galaxy tools
- Docker containers
  - Run locally
  - Run in cloud
  - Handle protected data

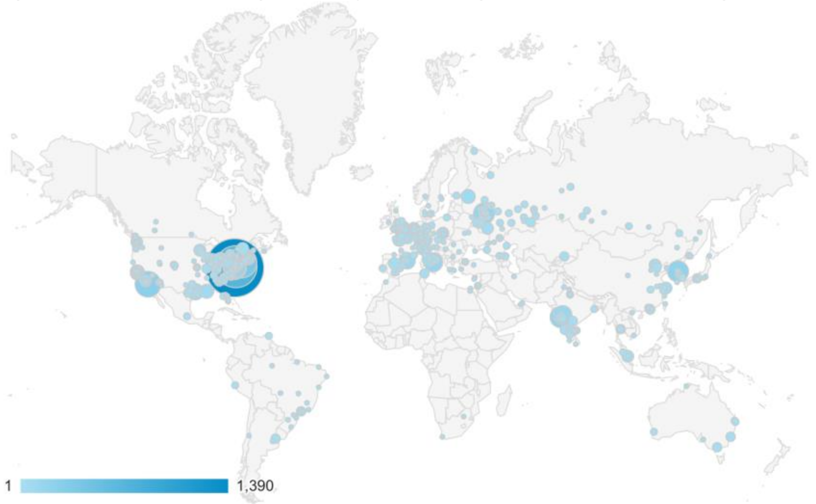




2016

Usage





Users

6,516

Pageviews

19,561

karchinlab/cravatmupit

public

376

PULLS

3 months

U24 Aims

- New features to make the tools more broadly useful
  - Add new mutation, gene and pathway scoring methods and annotations
  - Expand to non-coding mutations
  - Customized modular analysis and installation
- Maximize interoperability and interactions with other tools
  - NCI cloud pilot projects
  - Additional Galaxy tools
- Keep system up-to-date, user support and outreach
  - Rebuild underlying databases for hg38
  - Increase presence on genomics-focused social media