

CRAVAT 4.2: informatics tools for high-throughput analysis of exome variants

Collin Tokheim^{1,*}, Christopher Douville^{1,*}, David L. Masica^{1,*}, Rick Kim^{3,*}, Rohit Bhattacharya², Noushin Niknafs¹, Derek Gygax³, Kyle Moad³, Michael Ryan³ and Rachel Karchin^{1,4}

¹Department of Biomedical Engineering ²Department of Computer Science, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA

³In Silico Solutions, Fairfax, VA, USA ⁴Department of Oncology, Johns Hopkins Medical Institutions, Baltimore MD, USA. *Authors contributed equally.

The diagram illustrates the CRAVAT 4.2 architecture. It starts with a user interface for '1 Input' (VCF or CRAVAT-format), leading to '2 Analysis' (selecting VEST, CHASM, or SnvGet programs) and then '3 Results' (emailing reports). A central 'Jobs' section manages tasks, which can be downloaded as 'Text' or 'Excel' files. An 'Explore' interface provides a graphical results browser with various filters and visualizations.

Cancer-Related Analysis of VARIants Toolkit

- Comprehensive and easy-to-use web interface for assessing and prioritizing genes and variants important in cancer.
- Scores and annotates all coding exonic non-silent mutation types, including missense, in-frame and frame-shift insertions and deletions, splice variants and stop mutations.
- Analyze one variant or tens of millions, fast results are quickly emailed to users in spreadsheet and machine-friendly format.
- Create an account to store your projects online.
- Create an account to access the new graphical interactive results browser.

Important new features

- Scoring for all non-silent mutation types.
- Cancer mutation hotspot regions identified by new HotMAPS algorithm.
- Docker version, useful for protected data, allows analysis on local computers or in the cloud.
- Interactive results viewer includes: lollipop diagrams, population MAF widget, dynamic sorting and columns, and pathway enrichment viewer.

Broadened reach

- CRAVAT is now available in GALAXY.
- Integration with NDEx, Xena, IGV, and Trinity.
- Follow on twitter @CravatMupitTool.

Future development

- New study overview summary page coming soon!
- Machine learning prediction of cancer driver genes from user's data
- Annotation for non-coding variants.
- Availability on NCI cloud pilot platforms.
- Support for genome assembly hg38.

Publications this year

Tokheim C. Exome-scale discovery of hotspot mutation regions. *Cancer Research* (2016)
Douville C. Assessing the pathogenicity of indels. *Human Mutation* (2015)

CRAVAT usage past two years

A world map shows CRAVAT usage across continents, with a legend indicating session counts (1 to 1,390).

Users	6,516
Pageviews	19,561
karchinlab/cravatmupit	public 3 months 376 PULLS

HotMAPS algorithm

A 3D ribbon model of a protein structure with colored regions representing mutation hotspots.

Name	Color	Chain IDs
F-BOX/WD_REPEAT_PROTEIN_7	B	
S-PHASE_KINASE-ASSOCIATED_PROTEIN_1A	A	
CYCLINE_C-TERMINAL_DEGRON	C	

TCGA Mutations

	TCGA Mutations	TCGA 3D Mutation Hot Regions		
Tissue	Color	# Mutated Residues in Structure	Color	Number of TCGA Hot Regions
BLCA	■	■	■	■
BRCA	■	■	■	■
CESC	■	■	■	■
COAD	■	■	■	■

New summary page widgets

Sequence Ontologies

A pie chart showing the distribution of sequence ontologies: 217 Splice site, 943 Synonymous, 2649 Missense, 1 Stopgain, 1 Complex sub, 1 Inframe del, 1 Frameshift ins, 1 Stoploss.

CRAVAT integration

A diagram showing CRAVAT integration with various platforms: Xena, docker, NDEx, Galaxy, Trinity, and igv.