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=xovmly11Bcs
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
Integrate tools to prioritize cancer missense mutations
Integrate tools to prioritize cancer missense mutations

Machine-learning of missense mutation impact

Missense mutation analysis and protein structure

Clustering patterns
Proximity to ligands and interfaces
Integrated user-friendly application with interactive results explorer
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

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
**Broaden scientific scope**

- Identify important genes and pathways for cohort-level analysis

<table>
<thead>
<tr>
<th>HUGO symbol</th>
<th>Number of variants</th>
<th>MAFP</th>
<th>LNA</th>
<th>Most common variant</th>
<th>VEST pathogenicity score pathogenicity</th>
<th>VEST pathogenicity score non-pathogenic</th>
<th>Driver genes</th>
<th>Occurrence in OMIM by primary site</th>
<th>Number of samples in study having the gene related</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP53</td>
<td>28</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>28</td>
</tr>
<tr>
<td>BGN</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>HIST1H2B</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>CNAC1C</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>GADD45A</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>EM6D2</td>
<td>4</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>TBP1</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>KAT1B</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>CHE</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>ANGEL2</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>MKL1</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>INK3</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>EP2</td>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>TPRP3</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>TFF1</td>
<td>3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

*Multiple criteria to sort genes by importance in a cohort*

*Find pathways enriched for important genes*
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
Broaden user base

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

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

2016

Users
6,516

Pageviews
19,561

3 months

PULLS
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