

Saket Choudhary

CONTACT INFORMATION	1076W 30 th Street Apartment #13 Los Angeles California 90007	<i>Email:</i> skchoudh@usc.edu <i>Homepage:</i> http://saket-choudhary.me [Google Scholar] [CrossValidated] [Github]
EDUCATION	University Of Southern California[USC] , Los Angeles, USA <i>PhD Student</i> , Computational Biology and Bioinformatics [2014 - Ongoing] Adviser: Prof. Anton Valouev Project: Tools for Motif conservation Analysis http://moca.usc.edu [Poster] [Abstract] Indian Institute of Technology Bombay[IITB] , Mumbai India <i>Bachelor of Technology, Master of Technology</i> , Chemical Engineering [2009 – 2014] Masters Thesis: Pattern Recognition in Clinical Data	
RESEARCH INTERESTS	<ul style="list-style-type: none">- Machine Learning applications in Biology- Computational methods to study transcription regulation- Single-cell omics- Mathematical Statistics	
HONORS AND AWARDS	<ul style="list-style-type: none">- Provost Fellowship, awarded to outstanding incoming PhD students at USC [2014]- Gandhian Young Technological Innovation Award by Indian Institute of Management Ahmedabad, for designing a low cost water impurity detection device [2013]- Institute Technical Special Mention for 3 consecutive years at IITB [2010-12]- Undergraduate Research Award, for developing ‘Scilab on Cloud’ [2012]- Kishor Vagnyanik Protsahan Yojana(KVPY) Fellowship by Department of Science and Technology (DST), Government of India [2007]- Homi Bhabha Young Scientists’ Gold Medal [2005]	
OLYMPIADS	<ul style="list-style-type: none">- Top 6 to be selected for Indian National Mathematics Olympiad(INMO), selection level exam for International Mathematical Olympiad(IMO) [2008]- Top 30 in Regional Mathematics Olympiad(RMO) [2009]- Top 250 in Indian National Physics Olympiad (INPhO) [2009]- Top 300 in Indian National Astronomy Olympiad (INAO) [2009]	
PUBLICATIONS /PREPRINTS	<ol style="list-style-type: none">1. Choudhary, Saket and Anton Valouev. <i>MoCA: Tool for motif conservation analysis</i> (In preparation)2. Choudhary, Saket, Leyla Garcia, Andrew Nightingale, and Maria-Jesus Martin. <i>BioJS-HGV Viewer: Genetic Variation Visualizer</i>. bioRxiv (2015): 032573. [Preprint]3. Rahman, Syed Asad, Gilliean Torrance, Lorenzo Baldacci, Sergio M. Cuesta, Nimish Gopal, Saket Choudhary, John May, Gemma L. Holliday, Christoph Steinbeck and Janet M Thornton. <i>Reaction Decoder: Extracting Chemical Features from Chemical Reactions</i> (In Review)4. Syed, Parvez, Shabarni Gupta, Saket Choudhary, Narendra Goud Pandala, Apurva Atak, Annie Richharia, Heng Zhu et al. <i>Autoantibody Profiling of Glioma Serum Samples to Identify Biomarkers Using Human Proteome Arrays</i> Scientific reports 5 (2015). [Online]	

5. Yachdav, Guy, Tatyana Goldberg, Sebastian Wilzbach, David Dao, Iris Shih, **Saket Choudhary**, Steve Crouch et al. *Anatomy of BioJS, an open source community for the life sciences*. eLife 4 (2015): e07009. [Preprint]
6. **Choudhary, Saket**, and Santosh B. Noronha. *GalDrive: Pipeline for comparative identification of driver mutations using the Galaxy framework*. bioRxiv (2014): 010538. [Preprint]
7. **Choudhary, Saket**, Vishnu Raj, K. Sanmugasundaram, Gyan Singh Patel, and Kannan Moudgalya. *Scilab on Cloud and Textbook Companion Project: A Web 2.0 Service for Open Source Education*. In 2013 International Conference on Cloud Computing and Big Data. [Online]
8. Gatline, Pradip, Swati Gatline, Sushanth Poojary, **Saket Choudhary**, and Santosh Noronha. *Development of piezo-electric sensor based noninvasive low cost Arterial Pulse Analyzer*. In Biomedical Engineering International Conference (BMEiCON), 2013 6th, pp. 1-4. IEEE, 2013. [Online]
9. Dilip Save, Yogesh, R. Rakhi, N. D. Shambhulingayya, Amit Srivastava, Manas Ranjan Das, **Saket Choudhary**, and Kannan M. Moudgalya. *Oscad: An open source EDA tool for circuit design, simulation, analysis and PCB design*. In Electronics, Circuits, and Systems (ICECS), 2013 IEEE 20th International Conference on, pp. 851-854. IEEE, 2013. [Online]

RESEARCH EXPERIENCE

Tools for Motif Conservation Analysis, PhD Project

May 2015 - Ongoing

Guide: Prof. Anton Valouev

Dept. of Preventive Medicine, Keck School of Medicine, USC

Motifs predicted by motif discovery tools can often not be the ‘true motifs’ and can have significant p-value(or E-values) for even ‘false motifs’. We hypothesized that a ‘true motif’ should exhibit high evolutionary conservation scores. MoCA makes use of PhyloP and Gerp scores to assess the conservation profile of motif bases.

We used MoCA to analyze ENCODE Chip-Seq datasets and found that the ‘true motifs’(ones which have been validated experimentally) do exhibit high conservation scores and that these are statistically significant when compared to the scores of flanking regions or randomly sampled regions.

MoCA is available as a web-service at <http://moca.usc.edu>

Source: <https://github.com/saketkc/moca>

Improvements to Amordad and Methpipe, Rotation Project

Aug 2014 - Feb 2015

Guide: Prof. Andrew Smith

Computational Biology & Bioinformatics, USC

My first rotation project was on improving [methpipe](#), a pipeline for analyzing bisulfite sequencing and an internal tool to manage [methbase](#). I also worked on analyzing bisulfite sequencing data.

In my second part of the rotation I worked on [Amordad](#), an alignment-free approach to search similar metagenomic sequences. I implemented an optimized version of the server backend and added unit tests.

Pattern Recognition in Clinical Data, Masters Thesis

Apr 2013 - Jul 2013

Guide: Prof. Santosh Noronha

Dept. of Chemical Engineering, IIT Bombay

Awarded Outstanding Thesis Award

Multiple methods exist for determining oncogenic ‘driver’ mutations. These tools often have non overlapping predictions and input format is tool specific.

We developed a Galaxy based toolbox to run such prediction tools in parallel with a standard input format. The end results are presented as an intuitive heatmap indicating mutations which are predicted to be drivers by a majority of the tools. The results are summarized in the following preprint: ["GalDrive: Pipeline for comparative identification of driver mutations using the Galaxy framework"](#)

In a separate project, we analyzed proteomics data from Glioblastoma patients and predicted a smaller set of marker genes. This resulted in the following publication: ["Autoantibody Profiling of Glioma Serum Samples to Identify Biomarkers Using Human Proteome Arrays"](#).

Automated Mining of Reaction Patterns May 2012 - Jul 2012, Jan 2014 - Mar 2014
Guide: Dr. Syed Asad Rahman Dr. Dame Janet Thornton Lab, EMBL-EBI, Cambridge(UK)

EC-BLAST is a novel tool to compare enzymes and map reactions. We used clustering based approaches to highlight misclassified enzymes in the established enzyme classification system(EC).

We developed a web-service that facilitated automated job submissions to back end clusters at EBI that led to significant reduction in job runtime.

Next Generation Sequencing, Supervised Learning Project Jul 2012 - Dec 2012
Guide: Prof. Santosh Noronha Dept. of Chemical Engineering, IIT Bombay & ACTREC

We developed automated pipelines using Python to analyze whole genome sequence data of cancer tumors. As part of the project, I also contributed BWA and samtools wrappers to Biopython, a Python based open source library for bioinformatics.

Scilab On Cloud May 2012 - Jul 2012
Guide: Prof. Kannan Moudagalya Dept. of Chemical Engineering, IIT Bombay

Scilab is an open source software for numerical computation and is primarily command line/GUI based. We developed a back-end that allowed running Scilab through browser much like the modern day IPython notebooks. This enabled accessing Scilab remotely, even on low configuration devices.

Presented at: [IEEE Conference Cloud Computing and Big Data \(CloudCom-Asia\), 2013](#)

PROFESSIONAL EXPERIENCE

Google Summer of Code 2015 | Mixed Effect Models for *statsmodels* May 2015 - Jul 2015
Student Contract Developer

- *statsmodels* is a Python based library for statistical modeling
- Implemented IPython based notebooks illustrating varied applications of Mixed Effects Models
- Implemented likelihood ratio tests
- Progress Report: <http://statsmodels-mlm-gsoc2015.blogspot.com>

Google Summer of Code | BioJavascript Jul 2014 - Sep 2014
Student Contract Developer

- BioJavascript is an open source library to facilitate biological data visualization
- Developed 'Human Genetic Variation Viewer', a d3.js based component to visualize genetic variations from SNP databases
- Preprint: [BioJS-HGV Viewer: Genetic Variation Visualizer](#)

Google Summer of Code | Galaxy Project
Student Contract Developer

Jul 2013 - Sep 2013

- Galaxy Project is an open source web-based platform used for reproducible bioinformatics analysis
- Implemented 'nested workflows' that allows users to run a workflow inside a workflow, obviating the need to replicate steps
- Added 'edit on the go' functionality to edit default parameters before runtime
- Progress Report: <http://galaxy-gsoc2013.blogspot.com>

Google Summer of Code | Connexions Project
Student Contract Developer

Jul 2012 - Sep 2012

- Developed a Python module to allow embedding slide-shows in online notebooks
- Created functionality to add user defined quiz as an additional achievement
- Progress Report: <http://oerpub.github.io/oerpub.rhaptoslabs.slideimporter/>

OTHER PROJECTS

Image Analysis of Tuberculosis Samples

Jan 2013 - Apr 2013

Supervised Learning Project, Collaborator: Hinduja Hospital, Mumbai

- Used image processing algorithms to detect probable cases of TB from sputum images
- Developed a user friendly GUI to aid histopathologists thus reducing the overall delay in analysis

Pratham, Student Satellite Program

May 2010 - Oct 2010

India's First Students' Satellite Team, IIT Bombay

- Executed hardware testing of the On-board Computer system
- Implemented signal processing pipeline for communications subsystem

**TEACHING
EXPERIENCE**

- Teaching Assistant, Computer Programming and Utilization Fall 2011
- Teaching Assistant, Artificial Intelligence in Process Engineering Fall 2013

**POSITIONS OF
RESPONSIBILITY**

Web Manager, UG Academic Council

Jul 2012 - Apr 2013

- Initiated a number of web portals, thus improving online accessibility of academic resources
- Awarded **Institute Organizational Award**

TechniC, Core Group Member

Jul 2010 - Apr 2011

- Organized institute wide technical events; mentored students

**STANDARDIZED
TEST SCORES**

- GRE: Quantitative: 170/170 Verbal: 153/170 Analytical Writing: 3.5/6
- TOEFL: Reading: 29/30 Listening: 28/30 Speaking: 24/30 Writing: 28/30 Total: 109/120

**RELEVANT
COURSEWORK AT
USC**

- Machine Learning
- Numerical Analysis
- Biostatistics
- Methods of Statistical Inference
- Mathematical Statistics
- Analysis of Algorithms
- Molecular Biology
- Introduction to Computational Biology
- Seminar in Statistical Consulting
- Applied Probability