Shankaracharya, Ph. D.
Principal Bioinformatics Scientist II
SeQure Dx, A MaxCyte Company, Waltham, MA.
Website: https://www.linkedin.com/in/dr-shankaracharya-7308268/
E. mail: shankaracharya2222@gmail.com
Phone: 713-497-6813
(US Permanent Resident)

PROFESSIONAL SUMMARY

- Leading Bioinformatics scientist and Mentor with extensive experience in **developing methods and computational pipeline** for analyzing Genomics, Transcriptomics and other next generation NGS Datasets and communicating the results effectively.
- Authored more than 35 publications in high impact journals like *Nature Biotechnology*, *Nature Genetics*, *Nature Communications*, *Brain*, *Neuron*, *Cancer Discovery*, *Human Gene Therapy*, *Movement Disorder* and others.
- Developed industry-leading software packages for CRISPR/cas9/cas12 off-target nomination, annotation, and confirmation analysis as well as various variant calling (SNPs, Indels and large translocation).
- Experience applying supervised, unsupervised machine learning, and deep learning methods to get the biological insight of the data.
- Experience in hiring best talent in Bioinformatics to make high performance team and mentor them.

RESEARCH SUMMARY

- More than 12 years of experience (post Ph.D.) in mentoring, developing, maintaining, and customizing NGS pipeline (based on the project need) on High Performance Cluster and Cloud Computing environment as well as local server to reduce the time and cost of analysis.
- Industry standard project planning experience with Azure DevOps environment including scrum planning, Containerization, code control with GitHub (CI/CD) and running the applications on batch (AWS) as well as automating the pipeline using Next flow and/or Airflow.
- Production level Industry standard *state-of-the-art solutions* for Genomics, Transcriptomics and CRISPR-Cas based assay using various NGS based analysis methods, tools and pipeline.
- Comprehensive experience in identifying intermediate and high-risk cancer and neurological disease susceptibility loci from analysis of rare genetic variation using whole genome and whole exome DNA sequence datasets.
- Experience in designing, maintaining, and improving bulk and single cell RNA sequence data analysis pipeline. Differential gene expression analysis, splice site and cryptic splice site detection with RNA-Seq datasets.

PROFESSIONAL EXPERIENCE

SeQure Dx, Waltham, Massachusetts **(Feb 2022 – Sept 2025)**
Principal Bioinformatics Scientist

- Lead the development of tool to access the relative risk of predicting CRISPR-Cas9 guide-RNA performance.
- Lead the development of customized pipeline (upstream and downstream) for different experimental gene editing nomination (Digenome-seq, ONE-seq, GUIDE-seq) assay to nominate the off-target sites.
- Lead the development of customized confirmation assays pipeline for SNPs and small Indels (rhAMP-seq) and long translocations (Long read Illumina sequencing) to validate the detected cut-sites.
- Lead the development of detection of long inter-chromosomal, intra-chromosomal and deletion-based translocations due to CRISPR-Cas activity (SAFER-Detection).
- Lead the development of customized annotation pipeline for different nomination libraries and performing the data science to get insight of off-target effect in an individual/population.
- Lead the development of Primer design pipeline considering different filters for amplicon sequence assay.
- Implemented pipeline for amplicon sequencing NGS data analysis for off-target confirmation analysis.
- Leading the development of various other computational analysis pipeline and industry level data management for genomics.
- Lead the effort of making tool for gRNA Risk assessment for CRISPR-CAS guides to help select the best guide (with least off-target effect and high on-target efficiency and sensitivity).
- Provided one stop *state-of-the-art solutions* for Bioinformatics related topic about development of new assay strategy/pipeline and data analysis issues.
- Help developing high performance team by hiring the best possible candidate in the field and mentor them to get the best of them.

Prevention Genetics LLC, Marshfield, Wisconsin **(Jan 2021 – Feb 2022)**
Senior Bioinformatician

- Lead the Development of production level Structural Variant analysis pipeline for Whole Genome Sequence analysis including computational detection and validation of SVs.
- Lead the development of Mitochondrial DNA analysis pipeline in industry setting.
- Developed production level repeat expansion detection pipeline for Whole Genome Sequence data analysis.
- Lead the development of various other analysis pipeline and industry level data management for genomics in genetic testing space.

UMASS Medical School, Worcester, MA **(Oct 2017- Jan 2021)**
Bioinformatician III (June 2019 – Jan 2021):

- Lead the development of Rare Variant Analysis pipeline for detection of SNPs and small Indels, Copy Number Variations (CNVs) using very large cohort (>70,000 samples) neurodegenerative disease case-control Whole Genome and Whole Exome sequence dataset.
- Collaborated with scientist to develop Biomarker strategies using different NGS Datasets.
- Developed Structural variants and repeat expansion detection pipeline with WGS dataset.
- Developed automated workflow on google cloud's Terra platform for WGS NGS data processing and variant calling.
- Applied Bioinformatics approaches on various in-house and collaborative projects towards causative gene finding for neurological genetic disorder using state of art methods with DNASeq and RNASeq data analysis towards development of actionable therapeutics.
- Developed single cell RNA (scRNA) and single nucleus RNA (snRNA) analysis pipeline for 10x data.
- Managed high end Unix RAID servers and storage to manage Petabyte size datasets along with switching the analysis in cloud environment.
- Published studies in high impact journal like Nature Genetics, Cell, Nature Communications etc.

Postdoctoral Research Associate (October 2017- June 2019):

- Actively participated in the discovery of KIF5A gene causing ALS.
- Participated actively in updating SQL based ALS Variant Server (http://als.umassmed.edu/) with KIF5A gene finding.
- Involved in the discovery of Parkinson's Disease, ALS and other neurodegenerative disease gene with whole exome and whole genome large scale case control cohort datasets.

MD Anderson Cancer Center, Houston, TX, USA **(Feb 2013 - Oct 2017)**
Postdoctoral fellow:

- Designed Disease risk prediction ML based modeling tool PREdiction by SUpervised Learning Toolkit (PRESULT) to simplify the development, validation, and optimization of machine learning (ML) risk prediction models using various epidemiological and genetic datasets. The tool produces portable models and program objects that can be distributed and used for validation or prediction on new datasets. (http://www.hufflab.org/software/presult/),
- Actively participated for designing and conducting experiments for the development and publication of Pedigree Variant Annotation, Analysis, and Search Tool (pVAAST). pVAAST is a Software tool for identifying genetic variants that directly influence disease risk in families (http://www.hufflab.org/software/pvaast/)
- Participated for generating the results and publication of second version ERSA 2.0. ERSA (Estimation of Recent Shared Ancestry) estimates recent shared ancestry between pairs of individuals based on the number and lengths of chromosomal segments that they share identically-by-descent through common ancestors (IBD segments). (http://www.hufflab.org/software/ersa/)
- Tested Genetic Anticipation and calculated Mutation rate estimate in Lynch syndrome using Whole Genome DNA sequence data. Performed mapping, variant analysis of 8 quartets and quintets Lynch syndrome families with 3 trio control families. We applied general purpose and pedigree aware variant callers for variant calling.
- Trained graduate students and research assistants in various Bioinformatics techniques.

Birla Institute of Technology, Mesra, Ranchi, India. **(Dec 2006 - Dec 2013)**
Associate Lecturer:

- Designed tool for early detection of Diabetes in Indian and American Pima Indian population using Machine Learning (Mixture of Experts) tool.
- Contributed as Mentor in the development of Java based GUI (Swift Modeler) for homology modeling of protein. (https://github.com/shankaracharya)
- Taught various Bioinformatics, programming and statistical courses to M.Sc. Bioinformatics and B.E Biotechnology classes.

EDUCATION

- **Ph.D. – Technology (Bioinformatics),** Birla Institute of Technology, Mesra, Ranchi, India. (2009- 2012)

Thesis: Development of an intelligent system for early detection of diabetes risk in Indian population and evaluation of its performance.

- **APGD (Bioinformatics)** – Jawaharlal Nehru University, New Delhi, India (2005- 2006)
- **M.Sc. (Biotechnology)** – Awadesh Pratap Singh University, Rewa, Madhya Pradesh, India (2001- 2003)
- **B.Sc. (Biotechnology)** – Ranchi University, Ranchi, Jharkhand, India (1997- 2001)

Technical Skills

- **Bioinformatics:**
  Gene editing: Gene editing target nomination, annotation and analysis, bulk primer designing, analysis pipeline development for Nomination assay using ONESeq, DigenomeSeq, GuideSeq. Analysis pipeline development for Confirmation assay using amplicon sequences, Hybrid capture technology (using Sure Select, IDT etc.), rhAMP-seq, and large Structural variants detection (SAFER-Detection), gRNA risk assesment.
  Variants calling: GATK, BWA, SAMtools, Picard tools, VAAST, pVAAST, DeNovoGear, Polymutt, FamSeq, Variant annotation and analysis, Pipeline development of NGS Analysis, Expertise in NGS Data analysis for whole genome and whole Exome datasets.
  WGS Structural variant pipeline: Structural variant calling using CNVnator, Lumpy, Manta, Delly etc. SV genotyping using SVTyper
  Transcriptomic Data Analysis: scRNA and snRNA – 10x pipeline (Cellranger mkfastq, count, agr, Loupe etc.)
  Bulk RNA Seq: DE Analysis – STAR analysis, DESeq2, Kallisto, Sleuth.
  Biomarker Development using NGS: ENCODE, TCGA, SRA, COSMIC, GEO, VEP etc.
- **High performance computing environment**: PBS and LSF HPC clusters, big servers, job automation for mapping and variants analysis. Managing Unix based RAID servers and storage, Google cloud computing.
- **Containerization**: building and using docker based applications.
- **Cloud Computing**: AWS Batch, AWS EC2, Google Cloud, Terra pipeline for WGS variant calling analysis, Google Colab
- **Machine Learning**: Disease risk prediction modeling for Cancer and Diabetes datasets, Classification, regression, clustering, Statistical Methods: regression models, hypothesis testing and Confidence Intervals.
- **Programming Languages:** Bash script, MySQL (Local and server-based experience), Stored Procedure uses, Python (scikit-learn, numpy, scipy, pandas), Matlab, R.
- **Operating System**: Linux, Mac OSx and Windows 10.

SELECTED RECENT PUBLICATIONS: (Total publication 35)

- Hop *et al*., (2024) Systematic rare variant analyses identify RAB32 as a susceptibility gene for familial Parkinson's disease., ***Nature Genetics***. 56:1371–1376
- Fumes *et* al., (2024) Expression of ALS-PFN1 impairs vesicular degradation in iPSC-derived microglia. ***Nature Communications***. Vol 15, Article number: 2497
- Glass *et al*., (2022) Ataxin-2 intermediate expansions in ALS: Prevalence and implications for disease progression, **Brain**, 2022 May 6; awac167. doi: 10.1093/brain/awac167.
- Baron *et al*., (2022) ALS-associated KIF5A mutations abolish autoinhibition resulting in a toxic gain of function, **Cell Rep**. 39(1):110598; doi: 10.1016/j.celrep.2022.110598.
- Straniero *et al*., (2022) Role of Lysosomal Gene Variants in Modulating GBA-Associated Parkinson's Disease Risk, **Mov Disord**. 37(6):1202-1210; doi: 10.1002/mds.28987.
- Dewan *et al*., (2021): Pathogenic Huntingtin repeat expansions in patients with frontotemporal dementia and amyotrophic lateral sclerosis. **Neuron.** 109(3):448-460. e4
- Batista *et al.,* (2020) *Ly6a* Differential Expression in Blood-Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B., **Hum Gene Ther**., 31(1-2):90-102
- Dobson-Stone *et al*., (2020): *CYLD* is a causative gene for frontotemporal dementia – amyotrophic lateral sclerosis. ***Brain***, 143(3): 783–799 (2020); https://doi.org/10.1093/brain/awaa039
- Nicolas *et al*., (2018): Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. ***Neuron***. 97(6):1268-1283.e6. doi: 10.1016/j.neuron.2018.02.027.
- Hu H. *et al.,* (2014): A unified test of linkage analysis and rare-variant association for robust analysis of sequenced pedigrees. ***Nature Biotechnology***. **32**:663–669 (2014); doi: 10.1038/nbt.2895
- Park DJ, *et al.,* (2014): Rare mutations in *RINT1* predispose carriers to breast and Lynch Syndrome-spectrum cancers. ***Cancer Discovery***. 4(7): 804–815; doi: 10.1158/2159-8290.CD-14-0212
- Li H, Glusman G, Hu H, **Shankaracharya** *et al.* (2014): Relationship Estimation from Whole-Genome Sequence Data**.** ***PLoS Genet****,* 10(1): e1004144.