Jibin John, Ph.D.

New York, USA •+1- 516-288-8006• johnjibinv@gmail.com; Google Scholar; PubMed ; LinkedIn

Summary

Accomplished bioinformatician with 10+ years of experience in genetics, genomics, and bioinformatics, specializing in the development and optimization of automated analysis pipelines for Next-Generation Sequencing (NGS), Genome-Wide Association Studies (GWAS), and multi-omics data integration. Proven expertise in large-scale statistical genetics, high-performance computing, and cloud-based workflows, with a strong track record in academic and industry collaborations. Skilled in Python, R, BASH, and proficient in leveraging HPC and GCP environments for scalable, reproducible research.

Technical Expertise

- **Programming & Scripting:** Python, R, BASH, UNIX/Linux
- **Computational Environments:** HPC, Docker, Singularity, Podman, Google Cloud Platform (GCP)
- **GWAS & Statistical Genetics:** PLINK, GCTA, regenie, EPACTS, RAREMETAL, GENESIS, IMPUTE2, SHAPEIT, LDSC, coloc, HyPrColoc, FUSION, SuSIE, FINEMAP, TwoSampleMR, MR-PRESSO
- **Sequencing Data Analysis:** BWA, Bowtie, minimap2, GATK, SAMtools, bcftools, Picard, IGV, Bedtools, DeepVariant, Clair3, Medaka
- **Variant Annotation:** VEP, ANNOVAR, SnpEff, ClinVar, COSMIC
- **RNA-Sequencing:** STAR, LeafCutter, DESeq2, edgeR, RUVSeq, ComBat-seq, StringTie, Ballgown, GSEA, ClusterProfiler
- **Single-Cell Analysis:** Seurat, Scanpy, Cell Ranger
- **Databases & Web Tools:** ClinVar, OMIM, Orphanet, DECIPHER, PharmGKB, DrugCentral, COSMIC, DisGeNET, ClinGen, GenCC, LOVD, dbGaP, TCGA, ENCODE, SRA, GEO, and others
- **Data Science & Visualization:** pandas, scikit-learn, matplotlib, seaborn, ggplot2, tidyverse, ComplexHeatmap
  Additional Expertise: Extensive experience with numerous other bioinformatics tools and databases; regularly evaluate and adopt new resources based on evolving project needs.

Professional Experience

Research Scientist | Feinstein Institutes for Medical Research, New York, USA | Feb 2022 – Present

- Conducted common variant genetic association analyses for cognitive traits (e.g., cognition, reaction time) using Regenie on UK Biobank datasets.
- Performed meta-analyses and pleiotropic analyses of public GWAS summary statistics across imaging-derived phenotypes (IDPs), psychiatric, and cognitive traits.
- Developed a Python-based automated pipeline for harmonizing and processing public GWAS summary statistics into VCF format.
- Estimated heritability and genetic correlations from GWAS summary data.
- Performed post-GWAS analyses including fine mapping, colocalization, gene mapping, gene-based and gene-set enrichment analysis, Mendelian Randomization (MR), SMR & HEIDI, and pleiotropic association testing using UK Biobank and deCODE datasets.
- Calculated polygenic risk scores (PRS) for first-episode psychiatric disorders and schizophrenia.
- Prioritized GWAS genes using multi-layer biological data with tools such as FLAMES, MAGMA, and PoPS.

- Integrated single-cell RNA-seq and ATAC-seq with GWAS data to identify disease-relevant cell types.
- Designed and maintained scalable, reproducible pipelines in Python and R for MR, gene prioritization, and meta-analysis, optimized for HPC and GCP environments.
- Mentored junior scientists and medical residents in statistical genetics and GWAS methodologies.
- Published two peer-reviewed manuscripts based on these research findings.
- Collaborated with interdisciplinary teams to translate genetic discoveries into biological and therapeutic insights.

Scientist – Bioinformatics | Semantic Web India Pvt. Ltd., Bengaluru, India | Apr 2022 – Dec 2022

- Involved in the developed a cloud-based SaaS platform for exome and genome sequencing data analysis, enabling identification of pathogenic variants in Mendelian disorders and cancers.
- Built an automated pipeline for Copy Number Variation (CNV) detection from exome data, improving analysis efficiency and accuracy.
- Designed an advanced variant filtering and prioritization system integrating patient clinical data with ACMG guidelines for germline variants and AMP/ASCO/CAP guidelines for somatic cancer variants.

Computational Laboratory Manager | CDFD, Hyderabad, India | May 2021 – Apr 2022

- Led daily operations of the computational biology lab, managing Next-Generation Sequencing (NGS) data analysis projects and ensuring smooth collaboration with wet-lab scientists for seamless workflows.
- Supervised, trained, and mentored a team of junior staff, fostering efficient facility operations and high-quality data analysis output.
- Conducted workshops on clinical exome sequencing and RNA-seq data analysis for clinicians, students, and faculty, expanding bioinformatics expertise across diverse audiences.
- Successfully executed NGS data projects for a wide range of clients, including clinicians, agricultural scientists, and microbiologists.
- Designed and implemented custom pipelines covering clinical exome analysis, target capture sequencing, de novo genome and transcriptome assembly, ChIP-Seq, 16S rRNA and shotgun metagenomics, bacterial genome analysis, RNA-seq, and COVID-19 sequencing.
- Pioneered an advanced Oxford Nanopore third-generation sequencing pipeline enabling sensitive detection of Single Nucleotide Variations (SNVs), Insertions/Deletions (Indels), and supporting COVID-19 and bacterial genome analyses.

Postdoctoral Researcher | McGill University, Montreal, Canada | 2018 – 2021

- Led genomic studies to identify common and rare variants (gene and pathway based) associated with psychiatric disorders by analyzing SNP microarray data and whole exome data

from large multi generational families (10 families with 230 samples) and unrelated individuals.
- Interpreted the biological relevance of associated loci using systems genomics approaches and conducted polygenic risk score analyses to understand genetic risk architecture.
- Performed both family-based and population-based rare and common variant association analyses.
- Managed and mentored a multidisciplinary team of graduate students, undergraduate researchers, and laboratory technicians, fostering a collaborative and productive research environment.

Postdoctoral Researcher | University of Delhi, New Delhi, India | 2017 – 2018

- Developed and maintained automated pipelines for whole-exome sequencing (WES) and RNA-sequencing data analysis.
- Identified genes and variants associated with psychiatric disorders, tardive dyskinesia, and strabismus through analysis of familial and sporadic WES datasets.
- Played a key role in RNA-seq data analysis of fibroblasts from rheumatoid arthritis patients.
- Trained junior researchers in WES and genotyping data analysis and result interpretation.
- Served as the laboratory's technical expert on NGS data analysis, providing support across multiple complex projects.
- Authored and published seven manuscripts, including several as first author.

Academic Health Science Fellow | University of Pittsburgh, Pittsburgh, PA, USA | 2012

- Awarded a prestigious training fellowship from the Fogarty International Center at the National Institutes of Health.
- Cultured and maintained induced pluripotent stem cells (iPSCs) and differentiated them into neurons.
- Performed quality control assays to ensure accuracy and validity of experimental results.

Education

- Ph.D., Genetics – University of Delhi, India (2017)
- M.Sc., Biotechnology – University of Madras, India (2010)
- B.Sc., Biotechnology – Bharathidasan University, India (2008)

Certifications and Honors

- **Professional Development Certificate** in Data Science and Machine Learning, McGill University, Canada
- **Healthy Brains for Healthy Lives (HBHL)** Initiative Postdoctoral Fellowship
- **Reviewer at** the Journal of Gerontology, Psychiatric Genetics and Imaging Neuroscience

Publications

A full list of publications can be found on Google Scholar and PubMed