



# Venkatesh Kamaraj

 [venkatk89.github.io](https://venkatk89.github.io)  
 [github.com/venkatk89](https://github.com/venkatk89)  
 [vengatesh.vengatesh89@gmail.com](mailto:vengatesh.vengatesh89@gmail.com)

## RESEARCH EXPERIENCE

### Project Scientist

Jun 2022 – Until now

### Junior Research Fellow

Jan 2021 – May 2022

Centre for Integrative Biology and Systems Medicine, IIT Madras

Chennai, India

*Advisors: Dr. Himanshu Sinha, Dr. Karthik Raman, Dr. Manikandan Narayanan*

### Key Projects

- **Analysis of human genomes with genome graphs**
  - \* Designed scalable computational workflows with multiple bioinformatics tools for WGS studies
  - \* Developed novel methods for genome graph annotation to identify regions of significance
  - \* Performed detailed examinations to study the structural complexities in the genome graphs
- **GenomeIndia: Unveiling unique variants in Indian (sub) populations**
  - \* Analyzed the WGS of individuals from India's diverse populace using genome graphs
  - \* Used the dynamic nature of the genome graph to capture prevalent and rare variants in Indians
  - \* Studied the consequences of genetic variations within and between Indian sub-populations
- **Approachability of genomics and visualization of variants**
  - \* Ideated and developed SCI-VCF, a cross-platform *application* for genomic data analysis
  - \* Designed an intuitive GUI for the tool to aid users irrespective of their programming expertise
  - \* The software helps users summarise, compare, and visualize genetic variants from VCF files
- **Polygenic Risk Scores for common complex diseases**
  - \* Computed individual-level genetic risk scores for type-2 diabetes in an Indian cohort
  - \* Assessed the effect scores from homogeneous and trans-ethnic GWA Studies as base datasets
  - \* Employed machine learning techniques to enhance the predictive power of the calculated scores

## EDUCATION

### Bachelor of Technology in Engineering Physics

Aug 2015 – May 2019

Indian Institute of Technology, Madras | CGPA: 7.18/10

Chennai, India

## PUBLICATIONS

- **Venkatesh Kamaraj** and Himanshu Sinha "SCI-VCF: A cross-platform GUI solution to Summarise, Compare, Inspect, and Visualise the Variant Call Format" | bioRxiv 2023 

## PRESENTATIONS

### GenomeIndia: Cataloguing the genetic variations in Indians

Aug 2023

Poster presentation at the RBCDSAI AI/ML conclave on healthcare

IITM

### A comprehensive analysis of human genomes with genome graphs

July 2023

Poster presentation at Network Biology Day by The Institute of Mathematical Sciences

IMSc

### Sequence graph representations of yeast and human genomes

April 2021

Talk at the Data Science workshop organized by the GenomeIndia consortium

Online

## CONFERENCES AND TRAINING

---

<b>Data Science-driven solutions to improve maternal and child health</b>	Feb 2023
Reviewed the development of pregnancy dating models and challenges in clinical translation	
<b>Microbiomes in Environment, Space, and Human Health</b>	Nov 2022
Explored the core competencies and the recent advancements in microbial omics research	
<b>Clinical Genomics to Systems Medicine: Transforming Healthcare</b>	Feb 2022
Focused on the novel multi-omics analysis methods for understanding complex human diseases	
<b>Data Scientist Career track with R – DataCamp</b>	Sep 2018
Gained in-depth expertise in the multifarious aspects of data science to interpret complex data	

## WORK EXPERIENCE

---

<b>Data Science Consultant</b>	Sep 2019 – Feb 2020
Indus Insights and Analytical Services Gurgaon, India	
<ul style="list-style-type: none"><li>• Developed recommendation engine for the biggest airlines in the USA, using cutting-edge ML</li><li>• Leveraged cloud computing to train and evaluate deep learning models on over 100 GB of data</li><li>• Audited the marketing models of a US-based small business lender by performing detail-oriented examinations on the SAS-based models to ensure robustness during deployment</li></ul>	
<b>Data Science Intern</b>	May 2018 – Jul 2018
Blitzkrieg Retail Private Limited Chennai, India	
<ul style="list-style-type: none"><li>• Developed a recommendation engine for the online pharmacy store by mining association rules</li><li>• Created a proprietary image processing application for the company from scratch, using OpenCV</li><li>• Built an end-to-end ETL pipeline on Firebase and MSSQL to develop the company's dashboard</li></ul>	

## OTHER WORKS

---

<b>DNA Sonification Tool</b>	Oct 2020
<ul style="list-style-type: none"><li>• Incorporated music theory and developed an auditory display <i>tool</i> to sonify genomic sequences</li><li>• Enhanced the musicality of the tool's output while maintaining its overall analytical capabilities</li></ul>	
<b>DengAI – Disease Spread Prediction</b>	Jul 2020
<ul style="list-style-type: none"><li>• Designed ensemble methods to forecast the spread of dengue using TSA and statistical modeling</li><li>• Trained the models on Collab GPUs with data collected over a decade in South American cities</li></ul>	
<b>Climate Data Analysis – Department of Chemical Engineering, IITM</b>	Aug 2019
<ul style="list-style-type: none"><li>• Analyzed the data collected across India and validated key insights through hypothesis testing</li><li>• Created temporal plots and geospatial heat maps for radiation and air quality-related parameters</li></ul>	
<b>Fraudulent Transaction Prediction – Department of Computer Science, IITM</b>	Apr 2018
<ul style="list-style-type: none"><li>• Conceptualized and designed an online fraud prediction algorithm using random forest classifiers</li><li>• Won the first position in the Exebit Data Science Challenge, 2018</li></ul>	

For more details, visit  [portfolio website](#)

## SKILLS

---

**Bioinformatics:** Genome analysis, Single-cell transcriptomics, Statistical methodology, High performance computing, Data visualization, Machine learning, Deep learning, Software development  
**Programming:** R, Python, SQL, C, HTML, Shell scripting, Snakemake, Docker, Git