

# SIDDHARTH S. TOMAR

## PERSONAL INFORMATION

*email* [siddhartht.94@gmail.com](mailto:siddhartht.94@gmail.com)  
*website* <https://www.scilen.se>  
*github* [siddharthst](#)  
*StackExchange* [Siddharth](#)  
*phone* (Swe) +46-702759580 · (Ind) +91-8894636427

## EDUCATION

*Master of Science* 2016-2018 KTH Royal Institute of Technology  
Master's programme in Molecular Techniques in Life Science, a joint program with Karolinska Institutet and Stockholm University  
Duration: 2 years (full time)  
Date of graduation: 31-May-2018  
Country: Sweden

*Bachelor of Technology* 2012-2016 Jaypee University of Information Technology  
GPA: 8.6 out of 10 · *Specialization in Bioinformatics*  
Duration: 4 years (full time)  
Date of graduation: 09-July-2016  
Country: India  
Additional electives: *Cryptography · High Performance Computing · Systems Biology*

## SKILLS

*Programming* C++/C (INCLUDING CUDA AND MPI), JAVA, PYTHON, PERL, R  
*Libraries* PANDAS, DASK, MATPLOTLIB, ALTAIR  
*Databases* MySQL  
*Machine learning* SCIKIT-LEARN, TENSORFLOW  
*Documentation* L<sup>A</sup>T<sub>E</sub>X, MARKDOWN  
*Version control* GIT (INCLUDING AUTOMATED TESTING AND CI), SNAKEMAKE  
*Workflows* Time series analysis, Weighted correlation network analysis

## OTHER

*Model organisms* HUMAN, MICE, *P. ABIES*  
*Workflows* Genome/Transcriptome assembly, Quantification, Differential expression, Novel gene discovery  
*Life Science* DNA isolation, RNA isolation, Plasmid and genomic DNA isolation, PCR, Sanger and Illumina sequencing

## COURSEWORK - KEYWORDS

*Computer science* Algorithm design and analysis, High performance computing, Cryptography, Object Oriented Programming, Relational Database Management Systems, Machine Learning - Support Vector Machines and Artificial Neural Networks

*Molecular biology* Systems biology, Genetics, translational research in cardiology, neurodegenerative disorders, homology search, multiple sequence alignment, phylogeny reconstruction, protein topology prediction, biophysical chemistry, molecular dynamics, comparative genomics, proteomics, mass spectrometry, antibody assays, short and long read sequencing, mircoarrays, bead based assays, droplet barcoding, single cell sequencing, ChIP sequencing

## RESEARCH EXPERIENCE

*Jun'18–present* KAROLINSKA INSTITUTET - Sweden

*SciLifeLab -  
Stockholm*

I am currently working as a research assistant in Science For Life laboratory, Stockholm. My current work builds upon my experience in my master's thesis and is more focused on the developmental dynamics of PIWI-interacting RNA (piRNA) in mice and their interplay with transposons. I mainly use Pandas for storing and operating on sequence alignment data, coupled with matplotlib and Altair for static and interactive visualisations. I have also repurposed algorithms and methods used in signal processing for the genomic annotation and discovery of novel elements in mice genome.

*Jun–Aug '17* KAROLINSKA INSTITUTET - Sweden

*SciLifeLab -  
Stockholm*

After being awarded the SciLifeLab Summer Fellowship, I joined Dr Claudia Kutters group at the Karolinska Institute/SciLifeLab. My work focused on annotating various RNA classes from mice liver at different developmental stages by using total RNA sequencing data. This project involved methods for discovering novel developmental stage-specific noncoding RNAs and correlation with other meta-resources to mark functionality. I mainly used Pandas along with Python REST interface to automatically acquire and create relational data links for my genes, thereby annotating them using multiple meta-resources in using a robust and reproducible framework. During this time I also explored the use of Python interactive visualisation library Bokeh for exploratory analysis. Reference: Hassan FOROUGHI ASL +46-702759580 · [hassan.foroughi@scilifelab.se](mailto:hassan.foroughi@scilifelab.se)

*Jun–Aug '15* INSTITUTE OF ONCOLOGY RESEARCH - Switzerland

*IOSI-IOR*

I worked under Dr Francesco Bertoni's group at Institute of Oncology Research, Bellinzona, Switzerland on performance profiling of short read aligners (using Nvidia CUDA framework) for testing scalability in GPGPU environment. I studied the impact of GPU based tools on the next generation sequencing analysis pipeline and included cost versus performance analysis based on scaling.

*Master's thesis*

*Micro-evolution of regulatory RNAs*

In this master thesis project, I investigated the functional contribution of regulatory RNA in speciation in closely related mice species. This involved studying the dynamics of their species-specific gene expression by utilising ChIP sequencing to define active promoters along with short and total RNA sequencing.  
Supervisor: Claudia KUTTER

*Bachelor's thesis*

*Improving the computation of positive selection using codon models*

In my bachelor's project, I was working on improving the computation of positive selection using codon models. In this project, I implemented the Branch-Site Model on GPU architecture, and test the efficiency gains which are potentially possible. I changed the underlying algorithm to reduce the redundancy and increase the sensitivity by optimising the subtree pruning. I mainly used (CUDA)C++ for this project. The primary challenge of this project was to implement the current model as it is, without any heuristics involved, and still archive exceptionally low runtimes.

## PROJECTS

*2017* KTH ROYAL INSTITUTE OF TECHNOLOGY

*Course Project*

Assembling reference transcriptome for *P. abies* using massively parallel and PacBio real time sequencing data. This project involved utilization of large amount of data (>700 million reads) and optimization techniques for transcriptome assembly.  
Reference: Lars ARVESTAD · +4687906436 · [lars.arvestad@scilifelab.se](mailto:lars.arvestad@scilifelab.se)

*2017* STOCKHOLM UNIVERSITY

*Course Project*

Constructing a SVM based predictor for classifying topological features of  $\beta$  barrels based on their sequence.

*2014* Jaypee University of Information Technology

*Course Project*

Identification of novel leads against dimeric interface(DBP) in *P. vivax* using computer-aided drug design. For this project, I investigated the structural dynamics of Duffy binding proteins with respect to the phylum Apicomplexa.

*2013* Jaypee University of Information Technology

*Course Project*

Created an application written in Java for searching repeat content in a given genome. Simple machine learning technique was used, applying the Hidden Markov Model. Lempel–Ziv–Welch algorithm was also used and modified according to requirements. Perl was used as a wrapper and as text acquisition/annotation search engine.

# POSTER PRESENTATION AND CONFERENCES

2018                      2<sup>nd</sup> Uppsala Transposon Symposium  
*Poster*                      Micro-evolution of regulatory RNA in *M. domesticus* and *M. castaneus*

## GRANTS

*Nvidia*                      2015 - Nvidia GPU Grant Program

## OTHER INFORMATION

*Scholarships*                      2017 · Science for Life Laboratory summer fellowship  
    2017 · Karolinska Institutet Foundation scholarship

*Languages*                      ENGLISH · Fluent (C2/ TOEFL iBT score 108)  
    HINDI     · Native  
    SWEDISH · Beginner

Last updated on May 28, 2019