BioBytes

NEWSLETTER

Aug 2019

001

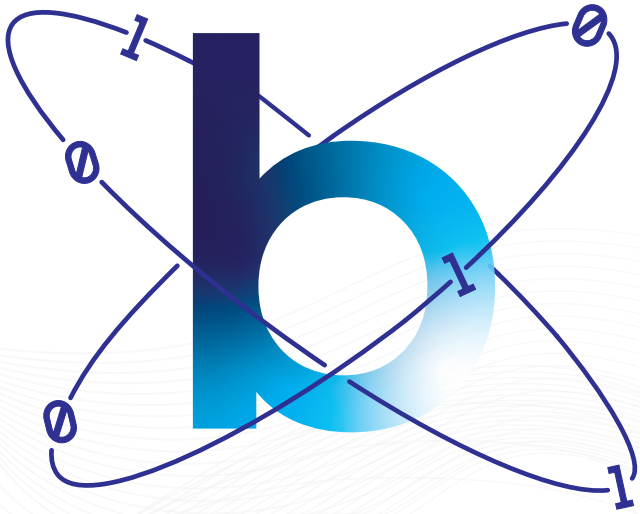
INTRODUCTION TO CB

AND MORE ABOUT BIOBYTES.

Throughout the course of humanity, we humans have been constantly evolving our machines and tools to make our lives easier. Having gained total dominance over the silicon empire, in the past 30 years alone, we have managed to achieve innumerable feats with the advancements in computing technologies. We have looked to nature for inspiration and in recent years there has been a surge in the development of tools to understand the natural phenomena and to simulate them. Understanding the deep and intricate natural design has become relatively easier with the computers simulating billions of natural processes in seconds to produce an outcome which, for us, would have taken years to produce. Computational algorithms have been coded which encompass most of the biological world. This field of study, as an umbrella term, is referred to as computational biology. More formally -

Computational Biology is the development of models and algorithms to understand biological systems. Experimental Biology produces a vast amount of data and so arose a need to process this data using computational tools and thus computational biology is becoming an essential field because of the possibilities it gives rise to in fields like healthcare and drug discovery. Contrary to popular belief, one does not come across names like “Myrmekiaphila neilyoungi” on a daily basis and nor does one have to cut open a frog to find the cause of a disease, instead data analysis and modelling are a computational biologist’s subsistence.

We at BioBytes aim to showcase the research being undertaken in the field of computational biology and help students regardless of their knowledge of biology take part in some of them. On top of that, we would be interacting with prominent personalities in this field, organizing sessions on the various techniques being used and have long challenges which would require you to solve real-world problems by developing state of the art algorithms. We will also initiate team projects soon exclusive to the members of BioBytes where we would be mentored by professors of the CB department to be a part of their research projects. These projects would potentially be published in research papers or could even be further incubated into startups.



BioBytes

Research

1. Microsoft Research

At Microsoft's research labs around the world, computer scientists, programmers, engineers and other experts are trying to crack some of the computer industry's toughest problems. A subset of those scientists, engineers and programmers are trying to use computer science to solve one of the most complex and deadly challenges humans face: Cancer. Although the individual projects vary widely, Microsoft's philosophy toward solving cancer focuses on two basic approaches

One approach is rooted in the idea that cancer and other biological processes are information processing systems. Using that approach the tools that are used to model and reason about computational processes – such as programming languages, compilers and model checkers – are used to model and reason about biological processes.

The other approach is more data-driven. It's based on the idea that researchers can apply techniques such as machine learning to the plethora of biological data that has suddenly become available, and use those sophisticated analysis tools to better understand and treat cancer.

"The collaboration between biologists and computer scientists is actually the key to making this work," said Jeannette M. Wing.

"We've reached the point where we are drowning in information. We can measure so much, and because we can, we do," said Jasmin Fisher, who is a senior researcher. "How do you take that information and turn that into knowledge? That's a different story. There's a huge leap here between information and data, and knowledge and understanding."

"We can use methods that we've developed for programming computers to program biology, and then unlock even more applications and even better treatments," said Andrew Phillips.

Of course, none of these tools will help fight cancer and save lives unless they are accessible and understandable to biologists, oncologists and other cancer researchers. And researchers at Microsoft are taking great pains to make these tools accessible and easy to understand.

One approach Fisher and her team are taking is called Bio Model Analyzer, or BMA for short. It's a

Microsoft Research

cloud-based tool that allows biologists to model how cells interact and communicate with each other, and the connections they make.

Here's one-way BMA might work: Let's say a patient has a rare and often fatal form of brain cancer. Using BMA, clinicians could enter all the biological information about that patient into the system. Then, they could use the system to run all sorts of experiments, comparing the cancer patient's information with that of a healthy patient, for example, or simulating how the patient's system might respond to various medications.

The ability to run these types of experiments "in silico" – or using computers – instead of with pen and paper or test tube and beaker also allows the researchers to quickly test many more possibilities. That, in turn, is giving them a better understanding of how cancers develop, evolve and interact with the rest of the body.

A system such as BMA can allow us to try out all sorts of ideas, which makes it more likely they will hit on the correct ones – and less likely they'll miss the dark horse candidates.

2. Nvidia Clara

Medical imaging developers around the world are discovering numerous ways to use AI to automate workflows, make instruments run faster and improve image quality, in addition to assisting doctors in detecting and diagnosing disease.

NVIDIA's Clara is an open computing platform that enables developers to build and deploy medical imaging applications into computing environments to create intelligent instruments and automated healthcare workflows.

Augmenting radiology with artificial intelligence and deep learning is reinventing the way we visualize medical images. A Magnetic Resonance Imaging (MRI) scan, is an irreplaceable tool for diagnosing many medical conditions. It is time-consuming and there are millions of them around the world that would take decades to upgrade. Normally in an X-ray scan, the body attenuates the signal, the workaround this problem is to simply increase the intensity of X-rays (which obviously is harmful). The regular method that CT machines have been using is called filtered back projection. CT machines collect the back-projection and reconstruct the image (the image has low fidelity). The left-most image shows what is called the iterative reconstruction, this method, one beam at a time calculates, estimates and corrects. The creation of images through this algorithm has been accelerated many folds through the use of GPUs. It's essentially the Ray-tracing of medical imaging. The brilliant thing it achieves is that it helps to reduce to dosage tremendously when compared to filtered back-projection method used in CT machines while increasing the resolution and fidelity of the reconstruction.

So essentially what you get is a higher fidelity image and at the same time reducing the dosage of x-rays (by a factor of 6). Another way to put this in perspective is that we can finally use the CT machines for children.

Once we obtain this image, we can now apply Clara AI to it, this neural network is trained to detect organs volumetrically. We now achieve the image in the middle (the organs are marked and can be seen distinctly. Finally, by applying ray-tracing, we obtain a very clear and precise 3D model (rightmost image).

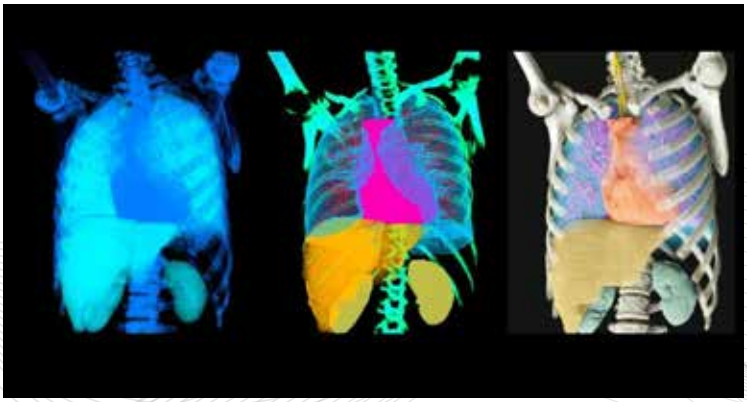


Image source : https://bit.ly/2Zzi8HT

3. Verily

Owned by Alphabet, Verily is a life sciences company previously known as Google Life Sciences. They develop tools to analyse health data to help in timely decision-making and effective interventions. By researching ways to predict and prevent disease onset and progression, they aim to transform the way healthcare is delivered.

Onduo is a virtual diabetes program by Verily that provides the tools, coaching and access to specialty doctors that you need to take control of your diabetes and learn what works for you. Onduo collects data from members directly, from members' wearable and other connected devices, and from the healthcare system. The data is used to understand your health and health risk factors, track progress, measure outcomes, and deliver meaningful insights.

verily

4. Deepmind

Owned by Alphabet, DeepMind seeks to solve real-world problems through the use of AI. One such field where DeepMind has managed to achieve excellence in is the healthcare domain.

DeepMind

In many clinical specialities, there is a relative shortage of this expertise to provide timely diagnosis and referral. For example, eye care professionals use optical coherence tomography (OCT) scans to help diagnose eye conditions. These 3D images provide a detailed map of the back of the eye, but they are hard to read and need expert analysis to interpret. And the worst part is that for the initial assessment of many of the sight-threatening diseases, OCT scans are indispensable. Artificial intelligence (AI) provides a promising solution for such medical image interpretation and triage. The system that has been developed seeks to address this shortage of skilled personnel. Not only can it automatically detect the features of eye diseases in seconds, but it can also prioritise patients most in need of urgent care by recommending whether they should be referred for treatment. This instant triaging process should drastically cut down the time elapsed between the scan and treatment, helping sufferers of diabetic eye disease and age-related macular degeneration avoid sight loss.The early results show that the system could handle a wide variety of patients found in routine clinical practice. In the long term, we hope this will help doctors quickly prioritise patients who need urgent treatment – which could ultimately save sight

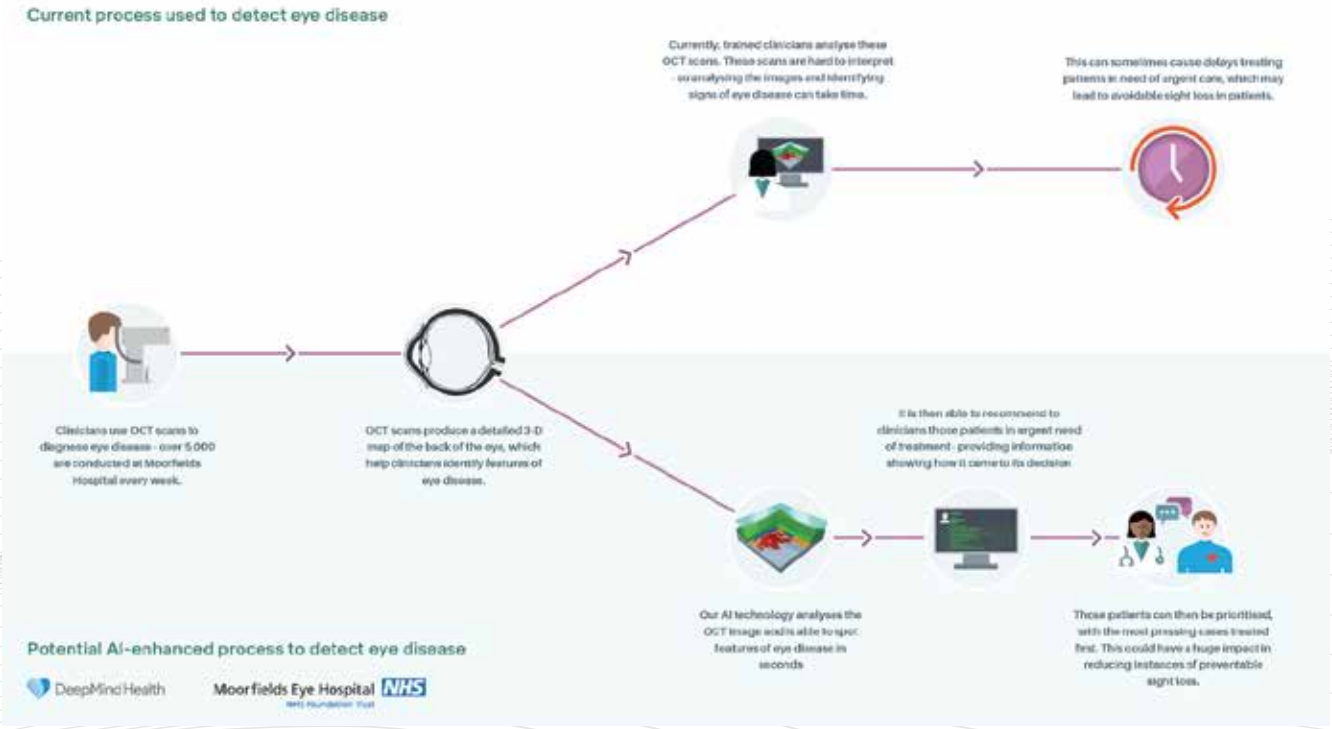
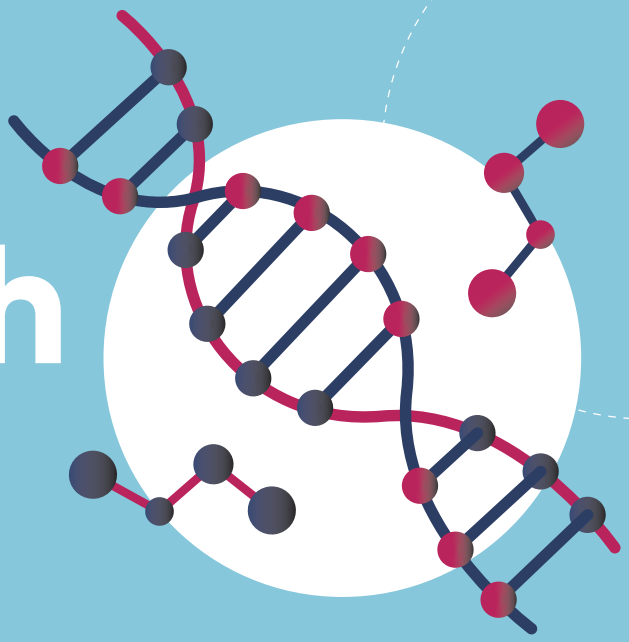


Image source : https://bit.ly/30P8gHr

Research Paper



Discovery of rare cells from voluminous single cell expression data

By Aashi Jindal, Prashant Gupta, Jayadeva & Debarka Sengupta

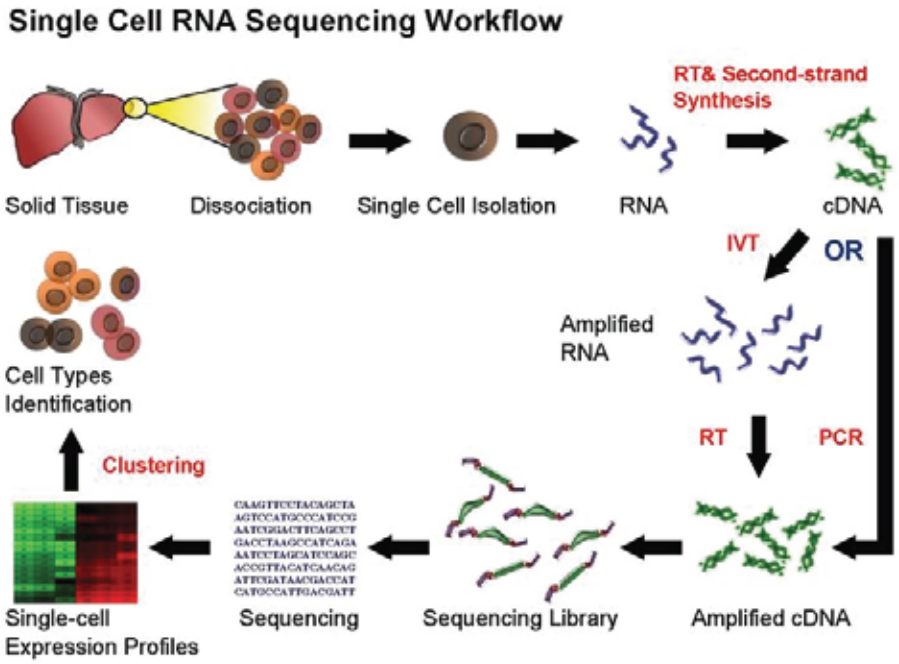
Continued progress in technology over the past years has made transcriptome analysis of individual cells a reality. (The study of all the RNA molecules within a cell is known as transcriptome analysis. Many studies of the transcriptome focus on messenger (m)RNA molecules only, which reflect the genes that are being actively expressed (as protein products) in a cell or tissue at a given time or in a given situation).

ScRNA-seq analysis is used to detect the rare cells from a cluster of cells. To understand why detecting rare cells is important, it is imperative to understand the importance of rare cell types. Rare cell types include circulating tumor cells, cancer stem cells, circulating endothelial cells, endothelial progenitor cells, antigen-specific T cells, invariant natural killer T cells, etc. Despite low abundance, rare cell populations play an important role in many biological processes. For example, determining the pathogenesis of cancer, mediating immune responses, angiogenesis in cancer and other diseases, etc. Stem cells have an ability to replace damaged cells, and to treat diseases like Parkinson's, diabetes, heart diseases, etc. Circulating tumor cells offer unprecedented insights into the metastatic process with real-time leads for clinical management.

The workflow involves isolating single cells from a cluster of cells obtained from a tissue and collecting the expression profiles through a series of steps involving amplification and sequencing. This process is tedious and time-consuming. Quite fortunately, recent discovery of the droplet-based single-cell transcriptomics has enabled the parallel profiling of tens of thousands of single cells, at a significantly reduced per-cell cost.

The analysis of this single cell expression profiles data does the task of categorizing cells into clusters and involves computation. (to understand, see image 1 on next page)

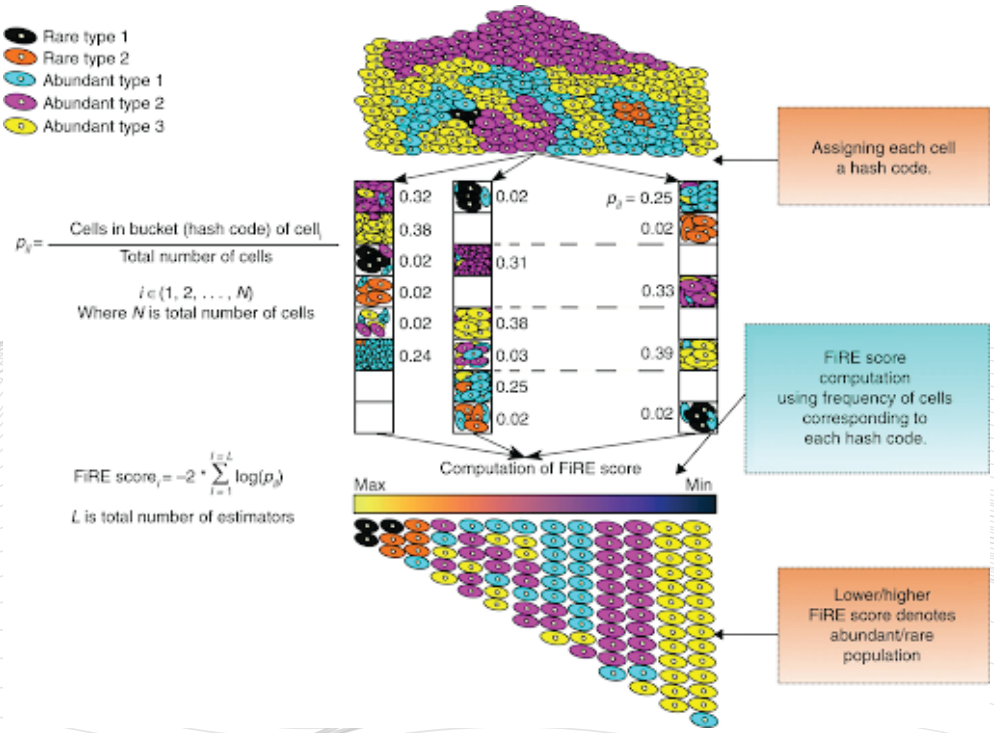
The techniques that we have used till now are based on unsupervised learning (clustering). Prominent among these are rare cell-type identification (RaceID) and GiniClust, both use clustering as an intermediate process. The shortcoming with clustering techniques is that it is often dependent



on a number of sensitive parameters and works inefficiently as density varies across data points. Resolution of group identities is also a major problem, therefore what is required is clustering in multiple stages (every stage refines the categorization process).

To avoid clustering, Researchers from IIITD under Dr. Debarka Sengupta designed Finder of Rare Entities (FiRE), a conspicuously fast algorithm to identify the rare-cell types. The algorithm is based on sketching technique. Sketches are compact data structures that can be used to estimate the properties of the original data in building large scale search engines and data analysis systems. The expression-profile data is very feature rich and is inherently high dimensional. Sketching is a powerful technique for low-dimensional encoding of a large volume of data points.

FiRE works by randomly projecting high dimensional data points to low dimensional hash-codes while preserving the l1 distance between the two points (l1 distance is the sum of absolute difference between the measures in all dimensions of two points). A hash code can be imagined as a bucket that tends to contain data points which are close by in the concerned higher-dimensional space. Cells which have the same hash-code share the bucket with cells of its type. The cell originating from a large cluster shares its bucket with many other cells, whereas a rare cell shares its bucket with only a few. FiRE uses the populousness of a bucket as a marker to classify rare cell types. Each type is assigned a FiRE score which denotes the abundant/rare population. FiRE assigns a continuous score to each cell, such that outlier cells and cells originating from the minor cell populations are assigned higher values in comparison to cells representing major subpopulations. A continuous score gives users the freedom to decide the degree of the rareness of the cells, to be further investigated. For better understanding, see this:



FiRE is fast, accurate and cheap. For example, FiRE took ~31 s to analyze a scRNA-seq dataset containing ~68 k expression profiles (for reference this would have taken >78h for RaceID). Such unrivaled speed, combined with the ability to pinpoint the truly rare expression profiles, makes the algorithm future proof.

FiRE is scalable and fast
FiRE can recover artificially planted rare cells
FiRE is sensitive to cell type identity

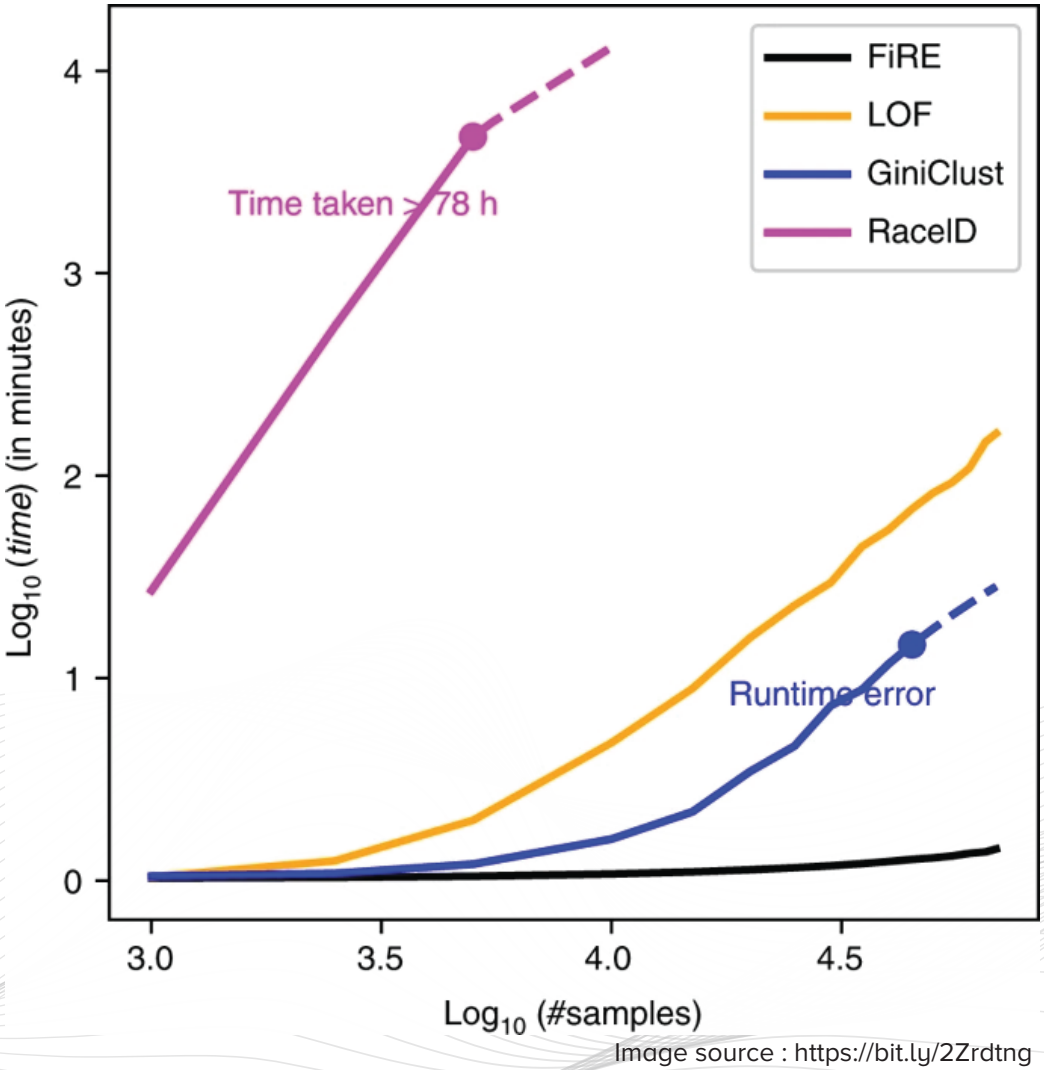


Image source : https://bit.ly/2Zrdtng

PhD Students joining in 2019

Ridam Pal

B.Tech CSE from Techno India University, Kolkata, WB.

Prageesha Chawla

M.Sc. Math with CS from Jamia Millia Islamia

Passed-out students

Shiju Sisobhan S.

• Currently a Postdoctoral Researcher at Northwestern University, Chicago.
• Specialization: Computational modeling of circadian Dynamics.
• M.Tech ECE at College of Engineering, Chengannur, Kerala.
• B.Tech ECE at College of Engineering, Kottarakkara.

List of Publications in 2019

- S. Kumar, J. Thakur, K. Yadav, M. Mitra, S. Pal, A. Ray, S. Gupta et al. "Cholic Acid-Derived Amphiphile can Combat Gram-Positive Bacteria-mediated Infections via Disintegration of Lipid Clusters." ACS Biomaterials Science & Engineering (2019).

- C. Goswami, S. Poonia, L. Kumar, D. Sengupta. "Staging System to Predict the Risk of Relapse in Multiple Myeloma Patients Undergoing Autologous Stem Cell Transplantation". Frontiers in Oncology, 12 July 2019.

- R. Kumar, S. Patiyal, V. Kumar , G. Nagpal, G.P.S Raghava. "In Silico Analysis of Gene Expression Change Associated with Copy Number of Enhancers in Pancreatic Adenocarcinoma". International Journal of Molecular Sciences 20(14), 3582. (2019).

- S. Ahmad, M. Gromiha, G.P.S. Raghava, C. Schönbach, S. Ranganathan. "APBioNet's annual International Conference on Bioinformatics (InCoB) returns to India in 2018.". BMC Genomics. 2019 Apr 18;19(Suppl 9):266.

- D. Kaur, S. Patiyal, N. Sharma, S. S. Usmani, G.P.S Raghava. "PRRDB 2.0: a comprehensive database of pattern-recognition receptors and their ligands". Database, Volume 2019.

- A. Nagori, L. S. Dhingra, A. Bhatnagar, R. Lodha, T. Sethi. "Predicting Hemodynamic Shock from Thermal Images using Machine Learning". Scientific Reports 9, Article number: 91 (2019)

- P. Agrawal, S. Patiyal, R. Kumar, V. Kumar, H. Singh, P. K. Raghav, G. P. S. Raghava. "ccPDB 2.0: an updated version of datasets created and compiled from Protein Data Bank." Database, Volume 2019.

- A. Mongia, D. Sengupta, A. Majumdar. "McImpute: Matrix Completion Based Imputation for Single Cell RNA-seq Data." Frontiers in Genetics.

- V. Ravindran, J. C. Nacher, T. Akutsu, M. Ishitsuka, A. Osadcenco, V. Sunitha, G. Bagler, J. Schwartz, D. L. Robertson. "Network controllability analysis of intracellular signalling reveals viruses are actively controlling molecular systems." Scientific Reports 9, Article number: 2066 (2019)

- P. Agrawal, H. Singh, H. K. Srivastava, S. Singh, G. Kishore, G. P. S. Raghava. "Benchmarking of different molecular docking methods for protein-peptide docking." BMC Bioinformatics2019
  S. S. Usmani, P. Agrawal, M. Sehgal, P. K. Patel, G. P S Raghava "ImmunoSPdb: an archive of immunosuppressive peptides." Database, Volume 2019

- P. Agrawal, S. Kumar, A. Singh, G. P. S. Raghava, I. K. Singh. "NeuroPIpred: a tool to predict, design and scan insect neuropeptides." Scientific Reports 9

- R. Kumar, G. Nagpal, V. Kumar, S. S. Usmani, P. Agrawal, G. P. S. Raghava. "HumCFS: a database of fragile sites in human chromosomes." BMC Genomics 2019.

- R. Tuwani, S. Wadhwa, G. Bagler. "BitterSweet: Building machine learning models for predicting the bitter and sweet taste of small molecules."Scientific Reports 9, Article number: 7155 (2019).

ON AIR WITH BioBytes

Dr. Debarka Sengupta

Dr. Debarka Sengupta is an Assistant Professor of the Departments of Computational Biology and Computer Science at IIIT-Delhi. Debarka did his doctoral research at the Machine Intelligence Unit of Indian Statistical Institute. After graduation in 2013, he pursued his postdoctoral research at the Genome Institute of Singapore where he got exposed to the then-emerging field of single-cell genomics. His group pioneered single-cell research in India and published several breakthrough findings, including the discovery of a rare subtype of pars tuberalis lineage in mouse brain (Jindal et al., Nature Communications, 2018). His current research focuses on early cancer detection using liquid biopsy techniques and functional interpretation of human variants. Debarka leads the data science research at Circle of Life Healthcare Pvt. Ltd., a health analytics company. He is a recipient of the prestigious INSPIRE Faculty Award. He serves on the editorial boards of PLOS One and Scientific Reports (A Nature group journal).

Get the full interview on Biobytes' Youtube channel

Support us by subscribing to our channel using the QR code.

SCAN ME

A production of :

BioBytes

The Computational Biology Club of IIIT Delhi

INDRAPRASTHA INSTITUTE of INFORMATION TECHNOLOGY DELHI