IBSE Annual Report

Jan 2022 - Dec 2023

https://ibse.iitm.ac.in

OBJECTIVES

The Centre for Integrative Biology and Systems mEdicine (IBSE) will establish state-of-the-art research cores and will continue to work to maintain, improve, and promote the interactions, collaborations, and Internationalisation by expanding current partnerships through increased faculty mobility into and out of IBSE. As an interdisciplinary group dedicated to pioneering innovative approaches and algorithms that integrate multi-dimensional data across scales to understand, predict and manipulate complex biological system, IBSE will conduct data-driven disease and clinical research

The key research activities in the IBSE are – (i) Development of algorithms using machine learning, data and graph mining to study clinical data to understand and predict clinical outcomes and generate disease risk models; (ii) Biological network analysis, which includes the prediction of molecular function in cells and populations, and environmental and evolution-dependent modulation of networks.

RESEARCH IMPACT

IBSE has conducted some cutting-edge research work which was published in reputed international journals. The research can be broadly classified into the development of a computational tool for bio-production, computational tools for cancer research, computational tools for basic research, anti-cancer treatment studies and design principles for adaptation. Here is a summary of the research work.

Focus Areas:

➤ Metabolic Modelling, Flux Analysis, Metabolic Engineering
➤ Systems Theory, Systems Identification, Control Theory
➤ Machine Learning, Graph Mining, Data & Text Mining, Reinforcement Learning,
➤ Network Biology, Regulatory Networks, Protein Folding, Tissue- specific Networks, Community Detection
➤ Multi-omics data analysis for medical research

IBSE FACULTY

HIMANSHU SINHA, Biotechnology
KARTHIK RAMAN, Biotechnology
NIRAV BHATT, Biotechnology
MEIYAPPAN LAKSHMANAN, Biotechnology
RAVINDRAN BALARAMAN, Computer Science
MANIKANDAN NARAYANAN, Computer Science
RAGHUNATHAN RENGASWAMY, Chemical Engineering

RESEARCH HIGHLIGHTS

:

COMPUTATIONAL TOOLS FOR MEDICAL RESEARCH

Personalised identification of cancer-causing genes

With the availability of multi-omic data, we use mutation, expression, copy number variations and network information to build multi-omic machine learning (ML) models. Our tool, PIVOT, is the first ML-based approach to classify driver genes as Tumour suppressors (TSGs) and oncogenes (OGs). We also identify rare driver genes altered in as few as a single sample and validate using literature (DOI)

COWAVE: A Labelled COVID-19 Wave Dataset for Building Predictive Models

We have collated COVID-19 case data from around the world and provided various definitions for waves. Using these definitions to define labels, we create a labelled dataset, which can be used while building supervised learning classifiers. We also use a simple eXtreme Gradient Boosting (XGBoost) model to provide a minimum standard for future classifiers trained on this dataset and demonstrate the utility of our dataset for the prediction of (future) waves. This dataset will be a valuable resource for epidemiologists and others interested in the early prediction of future waves. The datasets are available here.

METABOLIC MODELS

Elucidating genotype-phenotype or variant-to-function relationships remains a pivotal challenge in genetics. For quantitative traits, causal SNPs act either additively or epistatically, resulting in complex interactions which are difficult to dissect molecularly. We have developed gene co-expression networks and genome-scale metabolic models for all combinations of causal SNPs of yeast sporulation efficiency, a quantitative trait, to determine how genetic interactions drive phenotypic variation. Genome-scale differential flux analysis was used to identify differentially activated metabolic reactions across multiple metabolic pathways in each SNP-specific model. Our study presents a modelling framework that has the potential to unravel specific causal metabolic pathways and reactions regulated by SNPs and their combinations, thus providing insights into GWAS variants of metabolic diseases (DOI)

BIOMOLECULE PRODUCTION

By reconstructing metabolic models, we can predict plant metabolic behavior, facilitating the selection of suitable approaches and saving time, cost, and energy, over traditional hit and trial experimental approaches. In this study, we reconstructed a genome-scale metabolic model for *N. nimmoniana* and curated it using experimentally obtained biochemical data. We also used in silico tools to identify and rank suitable enzyme targets for overexpression and knockout to maximize camptothecin production. One of the top-ranked overexpression targets, strictosidine synthase, was chosen to generate metabolically engineered cell lines of *N. nimmoniana* using *Agrobacterium tumefaciens*-mediated transformation. The transformed cell line showed a 5-fold increase in camptothecin production. (DOI)

MICROBIOME RESEARCH

PATTERNING IN MICROBIAL CONSORTIA
Microbial consortia exhibit spatial patterning across diverse environments. Since probing the self-organization of natural microbial communities is limited by their inherent complexity, synthetic models have emerged as attractive alternatives. We have developed novel frameworks of bacterial communication and explore the emergent spatiotemporal organization of microbes, especially in quorum-sensing consortia. Overall, we demonstrate the potential of spatial organization as a tunable parameter in synthetic biology by introducing a communication paradigm based on the location and strength of coupling of microbial strains (DOI).

PRE-TERM BIRTH PREDICTION
We present a crowdsourcing approach to predict: (a) preterm or (b) early preterm birth from 9 publicly available vaginal microbiome studies representing 3,578 samples from 1,268 pregnant individuals, aggregated from raw sequences via an open-source tool, MaLiAmPi. We validated the crowdsourced models on novel datasets representing 331 samples from 148 pregnant individuals (DOI).

DEEP-SEA MICROBIOMES
Using microbiomes from the Guaymas Basin hydrothermal system in the Pacific Ocean to shed more light on the key species in these communities and their interactions, we have built metabolic models from metagenomically assembled genomes (MAGs) and inferred possible metabolic exchanges and horizontal gene transfer (HGT) events within the community. Ten microbes, including eight bacteria and two archaea, were identified as key contributors, by donating metabolites that cannot be produced by any other community member. Over 200 horizontal gene transfer events have also been predicted in the community (DOI).

MICROBIOME ON ISS
In order to study the interactions between the various microbes aboard the ISS and how they shape the microbiome, we apply a computational approach to predict possible metabolic interactions in the ISS microbiome and shed further light on its organization. (DOI)

EVOLUTION OF MICROBIAL INTERACTIONS
The evolution of two-member microbial communities *in silico* has been simulated using genome-scale metabolic models of organisms from the human gut, to study how evolutionary forces can shape the interactions between organisms. Depending upon the starting interaction, the metabolic capabilities of the organisms, and other evolutionary constraints, the community evolves to yield fitness benefits. Our approach of evolving microbial communities *in silico* provides an exciting glimpse of the dynamics of microbial interactions and offers several avenues for future investigations (DOI).

SCI-VCF
As genomics advances swiftly and its applications extend to diverse fields, bioinformatics tools must enable researchers and clinicians to work with genomic data irrespective of their programming expertise. We introduce SCI-VCF, a comprehensive toolkit for summarizing, comparing, and visualizing genetic variants from the variant call format. With an intuitive GUI, SCI-VCF aims to bridge the approachability gap in genomics that arises from the predominantly existing command-line utilities (DOI).

PANERA: A NOVEL FRAMEWORK FOR MICROBIAL COMMUNITY MODELLING
Panera, represents a potent and effective framework for metabolic modelling, enabling robust predictions of the metabolic phenotypes of microbial communities by building pan-genus metabolic models. The models are used to explore the comprehensive metabolic abilities of a genus (even in the absence of accurate species level mapping), shedding light on metabolic commonalities between microbial groups. It can be applied in characterizing microbial community models using 16S data. Our hybrid community models, which combine both GSMMs and pan-genus metabolic models, exhibit a 10% reduction in prediction error, with error rates diminishing as community size increases. We tested the performance of the models from our approach by analyzing their ability to capture the functionality of the entire genus and individual species within a genus. This advancement has the potential to greatly impact the field of microbiome research, offering new insights into the metabolic dynamics of diverse microbial ecosystems (DOI).

Review Articles

P. Sengupta, S. K. M. Sivabalan, A. Mahesh, I. Palanikumar, D. K. Kuppa Baskaran, and K. Raman, "Big Data for a Small World: A Review on Databases and Resources for Studying Microbiomes," Journal of the Indian Institute of Science, 2023.

COMPUTATIONAL TOOLS FOR BASIC RESEARCH

BIG DATA FOR A SMALL WORLD

We have catalogued databases with system-specific microbiomes, discussing publicly available common databases as well as specialized databases for a range of microbiomes. If the new datasets generated in the future could maintain at least biome-specific annotation, then researchers could use those contemporary tools for relevant and bias-free analysis of complex metagenomics data (DOI).

PHENOTYPIC PLASTICITY

Phenotypic plasticity aids survival in novel environments. Herein, we review studies from yeast that allow us to start uncovering the genetic architecture of phenotypic plasticity. Genetic variants and their interactions impact the phenotype in different environments, and distinct environments modulate the impact of genetic variants and their interactions on the phenotype. Because of this, certain hidden genetic variation is expressed in specific genetic and environmental backgrounds. A better understanding of the genetic mechanisms of phenotypic plasticity will help to determine short- and long-term responses to selection and how wide variation in disease manifestation occurs in human populations (DOI).

GUT MICROBES AND ANTI-CANCER TREATMENT:

Certain gut microbes such as *Enterococcus* spp. are known to have anticancer potential and therefore it is essential to study the effect of anti-cancer treatments on such microbes. IBSE studied the effect of silver nanoparticles that are promising candidates for cancer treatment on the *Enterococcus* spp. and found that the silver nanoparticles treated *Enterococcus* cells were successful in decreasing the viability of colorectal cancer cells. The study also observed the production of folate in *Enterococcus* cells treated with silver nanoparticles. While the previous studies show that folate helps in the progression of colorectal cancer cells, this new study by IBSE researchers signals that folate may instead help in killing colorectal cancer cells (DOI).

MULTI-LAYER TISSUE COMMUNICATION

To systematically identify inter-tissue mediators, we present a novel computational approach MultiCens (Multilayer/Multi-tissue network Centrality measures). Unlike single-layer network methods, MultiCens can distinguish within- vs. across-layer connectivity to quantify the "influence" of any gene in a tissue on a query set of genes of interest in another tissue. MultiCens enjoys theoretical guarantees on convergence and decomposability, and performs well on synthetic benchmarks. On human multi-tissue datasets, MultiCens predicts known and novel genes linked to hormones. MultiCens further reveals shifts in gene network architecture among four brain regions in Alzheimer's disease. MultiCens-prioritized hypotheses from these two diverse applications, and potential future ones like "Multi-tissue-expanded Gene Ontology" analysis, can enable whole-body yet molecular-level systems investigations in humans (DOI)

RESEARCH PROJECTS

GARBH-Ini - modelling of preterm birth risk

In collaboration with our clinical partners, the Translational Health Science and Technology Institute (THSTI), we are analyzing the GARBH-Ini cohort, a hospital-based observational pregnancy cohort to study multidimensional correlates of preterm birth in India. GARBH-Ini is one of the five "Atal Jai Anusandhan Biotech Missions – UNaTI" launched by the Government of India. We are building regression and machine learning models using anthropometric, clinical, and obstetric characteristics of the mothers to estimate gestational age and predict preterm birth. These models, when validated and deployed will be instrumental reducing maternal and child mortality in India.

A second round of validation of the developed models has been started at pan-India level, sponsored by BMGF.

GenomeIndia: Cataloguing the genetic variations in Indians

The aim of the 'GenomeIndia: Cataloguing the Genetic Variation in Indians' project is to do whole genome sequencing (WGS) for 10,000 healthy individuals of the country's diverse populace. This project will generate a unique dataset which will catalogue the unique variations of representative population groups across the country to create a catalogue of genomic variants for Indians. The project will help in cataloguing the common as well as low-frequency genetic variations in the Indian population, along with a comprehensive understanding of the copy number variations specific to our population and thus facilitate future large-scale genome-wide association studies for researchers across India.

Indo-European Consortium for Next Generation Influenza Vaccine (INCENTIVE)

The INCENTIVE consortium addresses the global health and the economic challenge posed by influenza infections. INCENTIVE's strategic goals are to provide seminal knowledge on the underlying mechanisms of poor responsiveness to influenza vaccines in vulnerable individuals and advance the development of two next-generation universal influenza vaccines. As a data science partner in INCENTIVE, we provide the state-of-the-art technology and methodology for vaccine research.

Multi-Omics of Mother and Infant (MOMI) - Predicting preterm birth outcomes from a combination of clinical and multi-omic molecular markers

Neonatal morbidity and mortality due to complications of prematurity can be controlled by the early detection of Pre-term Birth (PTB). Identification of molecular markers with respect to PTB remains crucial to minimise PTB-related complications. Thus, ML-based multi-omics data analysis will be helpful in identifying the key molecular markers that play a vital role in PTB and, therefore, targeted as novel prognostic or therapeutic candidates. In this project, we primarily focus on analysing the multi-omics data specific to subjects with preterm and term births corresponding to multiple cohorts comprising low and middle-income countries. Our data consisted of proteomics, metabolomics and lipidomics measurements that included 240 samples (120 each for preterm and term births).

EVENTS

SYNTHETIC BIOLOGY WORKSHOP
2 - 17 March 2023
Prof. Victor de Lorenzo, CSIC Spain

This hands-on course addressed the double-sided ability of SynBio (for understanding and for doing) along with its potential not only to transform the health & agricultural sectors but also moving the current oil-based industry into a sustainable bio-based economy. The Program was complemented with a daily journal Club and presentations of students' projects.

RBCDSAI-IBSE SEMINAR
3 March 2023
Dr. Suresh Sudarsan, DTU

For engineering a strain with a fermenterphile phenotype, a quantitative understanding of the metabolic and transcriptional regulatory network response under production conditions are essential. The demand for an improved strain design requires an "end-in-mind" approach with suitable tools such as small-scale scale-down bioreactors to mimic the spatial heterogeneity of large-scale industrial bioreactors and assess the response of production strains to the constant changing conditions.
This lecture addressed the challenges of a production strain in comparison to the parent strain, state-of-the-art methods/tools to identify & resolve biomanufacturing challenges, an integrated big data analytical approach to accelerate the strain/process development cycle

NGS TUTORIAL
5 April 2023
Prof. Vijay Tiwari, QUB, UK

This was a half-day hands-on intensive tutorial session on NGS data analysis. There were around 30 participants, mainly students and researchers from IBSE. The workshop covered techniques In RNA Sequencing and its applications in personalized cancer medicine at single-cell level. The participants also interacted with him extensively, seeking his inputs for their own research too. This was followed by a talk titled "Data-driven precision medicine" by him.

TRANSCRIPTOMICS WORKSHOP
13 July 2023
Prof. Vijay Tiwari, QUB, UK

This was another half-day intensive practical online workshop on Single-Cell transcriptomics. Participants were taught about the techniques used and given real-life research examples.

IBSE SEMINAR
17 April 2023
Karthikeyani Chellappa, Brown University
This talk was about the use isotope tracing and metabolomics to study the implications of host-microbiome interactions in aging and age-related diseases.

Organized a one-day event on 22nd February 2023, for stakeholder meeting of GARBH-Ini (Interdisciplinary Group for Advanced Research on BirtH outcomes - DBT INdia Initiative), at IIT Madras, Chennai, along with PIMS, Puducherry and THSTI, Faridabad.

Co-organized **AI-ML Conclave in Healthcare** with RBCDSAI, IITM on August 17, 2023

CSI- IITM SYMPOSIUM
3 - 4 October 2023
To foster collaborative research, this 2-day in-person event was an internal symposium featuring research talks by faculty from ITIM and CSI, NUS, Singapore. Researchers and faculty from 5 labs in CSI and 6 labs in IITM presented their work, along with scientists from Cancer Institute, Chennai.

IBSE TALK
21 September 2023
Dr. Krishnan. J, Imperial College, London
Titled **"**Structured systems investigations of information processing in biological networks", he discussed three different studies involving structured systems analyses which transcend an individual context, but can then be used to connect with individual contexts. The first part of the talk focussed on multisite substrate modification by enzymes. The second part of the talk involved an analysis of biphasic responses in biochemical enzyme/substrate networks. The third part of the talk focusses on the effect of spatial localization in biochemical systems.

IBSE SEMINAR
16 October 2023
Prof. Ganesh Bagler, IIIT D
The application of data-driven strategies for investigating gastronomic questions has created an all-new paradigm for studying food and cooking. Computational Gastronomy asks food-centric questions to seek answers via the structured compilation of data and their analysis. Making food computable will enable data-driven innovations and transform the food landscape, helping achieve better public health, nutrition, and a sustainable food ecosystem.

COLLOQUIA

11TH IBSE COLLOQUIUM
27 September 2023
Dr. Karthik Anantharaman, University of Wisconsin- Madison
This was an interesting talk on "(Re)defining the roles of viruses in microbiomes." The talk dwelt upon interactions in microbial communities using sulfur transformations as a model – virus-microbe, microbe-microbe, and microbial "metabolic handoffs." It focused on virus-microbe interactions involving "auxiliary metabolic genes.

12TH IBSE COLLOQUIUM
19 October 2023
Prof. Rouso Juho, Aalto University, Finland
This talk's focus was on Machine Learning for Small Molecules. Small molecules have a pivotal role in many applications in life sciences, including biomedicine and drug discovery, environmental sciences, and biotechnology. At the same time, the emergence of large open datasets is fueling the development of new machine learning technologies. Prof. Juho Rousu will discuss a generic machine learning task of predicting the a compatibility score F(x,y) of a pair of objects x and y, or in general a set of multiple interacting objects. Machine learning tasks such as predicting structured output, link prediction in networks as well as multi-variate association analysis in paired datasets falls under this umbrella.

IN THE NEWS

- IIT Madras researchers metabolically engineer plant cells to boost production of anti-cancer drug Camptothecin

- Assisting physicians with AI-based diagnostics, New Indian Express

- IIT Madras Robert Bosch Centre For Data Science & AI Organises AI, ML Conclave

- IIT Madras develop 'MultiCens' approach for Human Inter-Organ Communication

- MutBLESS: A tool to identify cancer-prone sites

- சென்னை ஜஐடி ஆராய்ச்சியாளர்கள் கண்டறிந்துள்ள புதிய தொழினுட்பம்

❖ **Developed courses for IITM Online BSc in Data Science and AI degree**
❖ **Prof. Ravindran, B. was awarded the Fellowship of the Indian National Academy of Engineering**
❖ **Ms. Purnima Kovuri received the International Immersion Experience Award from IIT Madras**
❖ **IBSE Travel Grants awarded to Ms. Purnima Kovuri and Mr. Srijith Sasikumar to Queen's University, Belfast and DTU, Denmark respectively**
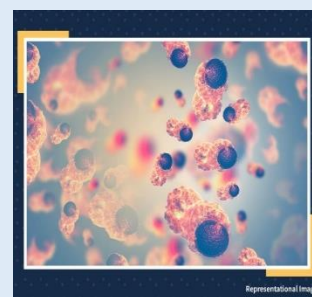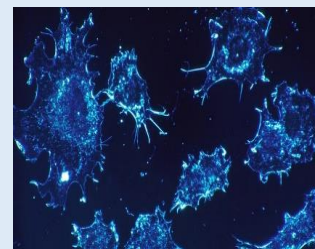
IBSE Personnel

- ○ **We heartily welcome Prof. Meiyappan Lakshmanan** as IBSE Faculty. He had worked with A-STAR Singapore and now utilizes a combination of computational systems biology and machine learning tools, as well as molecular biology/cell biology/"-omics" based experimental techniques to deeply characterize the molecular processes of mammalian and microbial cells. He also aims to develop computational tools to analyze large-scale multi-omics data and build mechanism based computational models.

- ○ **IBSE Post- baccalaureate Fellows:**
    - Deepikka Senthilkumar
    - Barathi L
    - Tanisha Malpani
    - Vijay Yuvaram Singh

- ○ **IBSE Post-doctoral Fellows:**
    - ○ Shree Chaitranjali Yadla

OUR NEXT AREA OF FOCUS

1. *A multicentric external validation of gestational age estimation models developed using data from the GARBH-Ini cohort*

   It is crucial to accurately estimate gestational age (GA) during pregnancy for effective antenatal care and epidemiological purposes. Traditionally, GA has been estimated using the last menstrual period (LMP), but this method is prone to errors due to various factors. Ultrasound (US) imaging has emerged as a safe and accurate method for GA estimation during pregnancy, especially in the later stages when fetal biometry parameters other than crown-rump length (CRL) are used. However, the current biometry-based methods may not be reliable in low and middle-income countries (LMICs) like India, where the prevalence of small for gestational age fetuses is high. To address this issue, the GARBH-Ini cohort study was initiated in India in 2015 to acquire fundamental knowledge of adverse pregnancy outcomes and identify sustainable solutions to reduce associated mortality. The study has enrolled and followed up with over 10,000 pregnant women and collected ultrasound images and videos throughout their pregnancy. The aim was to use this data to develop more accurate methods for GA estimation, particularly in the later stages of pregnancy, using biometry and non-biometry-based information from US images. We have developed two GA estimation models (Garbini-GA1, Garbini-GA2 (based on US biometry) for predicting GA. We tested these models on small external validation datasets. Before these tools can be deployed in clinical practice, we must validate them using pan-India datasets.

2. *Center of Excellence in AI for Health*

   Recently Govt of India announced a Call for Proposal (CFP) to establish Centres of Excellence (CoE) in Artificial Intelligence in three sectors: Agriculture, Sustainable Cities, and Health. One such Center shall be put up at IIT Madras. The first phase will attract RS 2 Cr funding to build towards a more robust application for funding to set up a Rs 330 Cr Centre for five years.

   IBSE, along with RBCDSAI and CeRAI, the School of DSAI and the Dept of MST, IIT Madras, are putting together a strong proposal for building a world-class hub for developing and deploying AI solutions to address critical healthcare challenges in India. We have established collaboration with several technical education institutes, hospitals, central and state government agencies, NGOs, and industry and startup companies. Building on the success of IBSE in the maternal and child health (MCH) domain, the key focus of this proposal is to develop AI tools to improve maternal neonate outcomes under the theme "AI tools for the first 1000 days of life".

PICTURES

IBSE is grateful to [Dr. Prakash Arunachalam](#), Lead Data Scientist, BNY Mellon, USA, for his generous contribution supporting the IBSE and funding the various other IBSE activities through the Office of International and Alumni Relations, IIT Madras.

The initial funding for IBSE has helps us to recruit talented personnel and raise further funds (~ Rs. 855 lakhs) through various projects sponsored by government agencies and private companies.