

An Indian-Australian research partnership

Project Title:

Application of Artificial Intelligence in human genomics

Project Number

IMURA0885



Monash Main Supervisor

(Name, Email Id, Phone)

Dr Sonika Tyagi, sonika.tyagi@monash.edu

Monash Co-supervisor(s)

(Name, Email Id, Phone)

Monash Head of

Dept/Centre (Name,Email)

Prof Moira O'Bryan, moira.obryan@monash.edu

Monash Department:

School of Biological Sciences, Faculty of Science

Monash ADRT

(Name,Email)

Peter Betts

IITB Main Supervisor

(Name, Email Id, Phone)

Dr Mithun Mitra, mithun@phy.iitb.ac.in

IITB Co-supervisor(s)

(Name, Email Id, Phone)

IITB Head of Dept

(Name, Email, Phone)

Dr. K. G. Suresh, hod@phy.iitb

IITB Department:

Department of Physics

Research Clusters:

Research Themes:

**Highlight which of the Academy's
CLUSTERS this project will address?**

(Please nominate JUST one. For more information, see
www.iitbmonash.org)

**Highlight which of the Academy's Theme(s) this
project will address?**

(Feel free to nominate more than one. For more information, see
www.iitbmonash.org)

1	Material Science/Engineering (including Nano, Metallurgy)	1	Advanced computational engineering, simulation and manufacture
2	Energy, Green Chem, Chemistry, Catalysis, Reaction Eng	2	Infrastructure Engineering
3	Math, CFD, Modelling, Manufacturing	3	Clean Energy
4	CSE, IT, Optimisation, Data, Sensors, Systems, Signal Processing, Control	4	Water
5	Earth Sciences and Civil Engineering (Geo, Water, Climate)	5	Nanotechnology
6	Bio, Stem Cells, Bio Chem, Pharma, Food	6	Biotechnology and Stem Cell Research
7	Semi-Conductors, Optics, Photonics, Networks, Telecomm, Power Eng	7	Humanities and social sciences
8	HSS, Design, Management	8	Design

The research problem

Define the problem

Single nucleotide polymorphisms called SNPs (snips), are the most common type of genetic variation among people. SNPs occur once in every 300 nucleotides on average in human genome. There are roughly 10 million SNPs in the human genome, and most commonly, these variations are found in the DNA between genes. In addition to SNPs, DNA-rearrangements, gene silencing, and satellites are also known to cause diseases. High throughput technologies are now able to generate genome scale data at DNA, RNA, protein and metabolite level. Integrating these various types of biological and health data is required to design personalized medicine and therapies. The field of Artificial Intelligence (AI) and machine learning (ML) holds promises to develop models to integrate complex, multi model and high dimensional data from biological experiments and health records to improve diagnosis, treatment and overall healthcare quality.

Project aims

Define the aims of the project

- Catalogue of genes, SNPs, and DNA elements associated with high probability of diseases (literature review, and database construction).
- Developing specific DNA-marker database(s) with special emphasis on diseases relevant to Indian population: Global and Indian population SNP map, and their comparison.
- GWAS to estimate disease risk: Development and application of genome-based strategies for the early detection, diagnosis, and treatment of disease.
- Development of robust bioinformatics pipeline to study genes and DNA elements utilising 'multi-omics' information in large scale genomic data analysis settings.
- Integration of genomic data with health records and other possible information such as from IoT devices using analytics tools based on machine learning
- Use of AI and ML in early diagnosis of diseases, through advanced data processing (viz. image data analysis), connecting genotype to phenotype, predicting regulatory function, and classifying mutation types.

Expected outcomes

Highlight the expected outcomes of the project

- Novel biomarker discovery in human disease.
- Machine Learning based framework for modeling disease of interest.
- New predictive disease models suitable for personalized medicine.

How will the project address the Goals of the above Themes?

Describe how the project will address the goals of one or more of the 6 Themes listed above.

The project will combine the knowledge of Artificial Intelligence/Machine Learning and human genomics and integrate them to discover novel biomarkers of disease.

Capabilities and Degrees Required

List the ideal set of capabilities that a student should have for this project. Feel free to be as specific or as general as you like. These capabilities will be input into the online application form and students who opt for this project will be required to show that they can demonstrate these capabilities.

- Candidates with a degree in Computer Science or Bioinformatics will be preferred.
- Demonstrated advanced knowledge of computer programming and data exploration, analysis and visualization : preferred
- Expertise in computer science and engineering with experience in distributed systems and networking; software engineering: preferred
- Knowledge of human genetics and disease: preferred
- Knowledge of genomics and current high throughput genomic technology: preferred
- Competencies in computational biology and data analysis algorithms: preferred
- Prior exposure to machine learning frameworks (such as tensorflow, keras, PyTorch) is desirable

Potential Collaborators

Please visit the IITB website www.iitb.ac.in OR Monash Website www.monash.edu to highlight some potential collaborators that would be best suited for the area of research you are intending to float.