Powered by **GENOMAC HUB**

GENOMAC ONLINE RESEARCH SCHOLARSHIP ON GENOMICS AND BIOINFORMATICS

From Real-life Research Projects to Publishable Papers in 3 - 9 Months Without Previous Research Experience in Bioinformatics

*We envision inspiring and empowering life scientists to leverage **GENOMICS AND BIOINFORMATICS** to tackle critical challenges, drive innovation, and promote sustainable progress globally.*

Research Domain: Therapeutics R&D (Multi-Omics)
Research Focus: Anti-Cancer R&D
Research Topic: Discovery and Utilization of Biosynthetic Genes from Medicinal Sources in Peptide-Based Anti-Cancer Therapeutics through Multi-Omics and Machine Learning.
Research Aim: To systematically discover and leverage biosynthetic genes from medicinal sources, integrating them into the development of peptide-based anti-cancer therapeutics using a multi-omics and machine learning approach, with the goal of enhancing precision and efficacy in cancer treatment.
Research Objectives:

- Identify and characterize novel biosynthetic genes from diverse medicinal sources with potential anti-cancer properties through comprehensive multi-omics analysis, aiming to expand the repertoire of genetic elements available for the development of peptide-based anti-cancer therapeutics.
- Implement machine learning algorithms to analyze complex relationships within the identified biosynthetic gene data, predicting optimal combinations for the design and synthesis of peptide-based anti-cancer therapeutics, thereby advancing precision medicine in cancer treatment.

LEARNING OBJECTIVES

- **Understanding Biosynthetic Gene Discovery:** Gain a comprehensive understanding of the principles and methodologies involved in the discovery of biosynthetic genes from various medicinal sources with potential anti-cancer properties. This includes acquiring knowledge of advanced multi-omics techniques employed in the identification and characterization of these genes.
- **Mastering Multi-Omics Analysis:** Develop proficiency in utilizing multi-omics approaches, such as genomics, transcriptomics, and proteomics, to comprehensively analyze and interpret the molecular information obtained from medicinal sources.

- **Peptide Design for Anti-Cancer Activity:** Develop expertise in activity-based peptide design using specialized tools such as Anti-CP, focusing on the prediction and optimization of anti-cancer peptides with specific characteristics like length, terminus details, and activity scores.
- **In Silico Analysis and Docking Techniques:** Gain proficiency in silico analyses of designed peptides, including the prediction of phytochemical properties, structure, and function. Develop skills in molecular docking techniques, exploring various software tools for drug target identification, ligand selection, and evaluation of results.
- **Craft Research Papers for Publication:** Learn how to synthesize and present your findings coherently, culminating in the preparation of research papers suitable for publication, contributing to the broader understanding of the therapeutic research and development of anti-cancer peptides.

EXPECTATIONS WHILE UNDERTAKING THIS INTERNSHIP PROGRAM:

- **Knowledge of Genomics and Bioinformatics:** Develop a solid foundation in genomics and bioinformatics, including an understanding of key concepts, methodologies, and technologies used in the program
- **Proficiency in Data Analysis:** Gain proficiency in analyzing genomic data using bioinformatics tools and software. This includes skills in data preprocessing, quality control, data visualization, and statistical analysis.
- **Research Skills:** Acquire research skills necessary for conducting genomics and bioinformatics studies. This includes formulating research questions, designing experiments, collecting and analyzing data, and interpreting research findings.
- **Critical Thinking and Problem-Solving:** Develop critical thinking skills to analyze complex genomic and bioinformatics problems and propose creative solutions. You would be able to evaluate scientific literature, identify research gaps, and contribute to the advancement of knowledge in the field.
- **Computational Skills:** Gain proficiency in software and applications commonly used in bioinformatics, such as Geneious software, web servers etc. to analyze genomics data and interpret results
- **Communication Skills:** You would be able to effectively communicate your research findings and scientific concepts to both technical and non-technical audiences. This includes writing scientific reports, presenting research orally, and participating in scientific discussions and collaborations.
- **Collaboration and Teamwork:** Be able to develop skills in collaborating with peers and professionals in multidisciplinary research teams. This includes effective communication, teamwork, and the ability to contribute constructively to group projects.
- **Professional Development:** You would be able to develop a professional mindset, including skills in time management, organization, and project management. They should also be aware of current trends and advancements in genomics and bioinformatics, and actively seek opportunities for professional growth and development.
- **Publication and Dissemination:** Contribute to the scientific community by publishing their research findings in peer-reviewed journals

PROGRAM OUTLINE AND SCHEDULE

CLASSES	TOPICS/FOCUS	SCHEDULE & DELIVERABLES
General Classes	Overview of genomics, bioinformatics, and their applications in various fields	**WEEK 1**
	Understanding the central dogma of molecular biology	
	Introduction to genomics technologies and data generation	
	Data formats in Genomics and Bioinformatics (Practical)	
	Internet tools and Databases (Practical on data retrieval, Blast etc.)	
	Introduction to software tools and their installation, web servers, and pipeline tools (Practical), Basic Linux Command Line Interface	
	Genomics Data and its Analysis using cutting-edge tools (Practical DNA, RNA and Protein samples)	
Specialized Classes	Introduction to Therapeutic Peptide R&D **(Anti-Cancer)**	**WEEK 2**
	The experimental application of each of these in your field of study	
	Problem identification relative to the above area in the healthcare, industrial, and other life science research space	
	The use of critical thinking and problem-solving tools to design a hypothesis in solving identified problems	
PRACTICAL SESSIONS		
GENOMICS AND BIOINFORMATICS ANALYSIS		
PHASE ONE		
Data Collection and Preprocessing	**Raw Data Retrieval:** Gather diverse raw data samples, encompassing microbial metagenomics and other pertinent natural sources, to initiate comprehensive genomics and bioinformatics analyses crucial for the discovery and optimization of anti-cancer peptides.	
	Table 1: Construction of General Sequence Properties: via data table based on genome information which includes accession number, raw data size, sources, geographical regions platform, genome type, layout, file types, etc.	**Deliverable:** (Materials and Methods)
	Quality Control: Assess data quality, perform trimming, and filter out low-quality reads to ensure reliable results. **Genome Assembly:** Assemble the whole genome sequence of the diverse data samples recovered from different data sources.	**Deliverable:** (Results)
	Write Up 1 and 2: Reads Processing and Genome Assembly	
Comprehensive Genome Analysis	**Functional Genome Annotation:** Gene prediction, Protein features, Specialty features, Chromosomal properties, and Circus-view, among others.	**WEEK 3**
	Write Up 3: Functional Genome Annotation/Protein Identification	**Deliverable:** (Materials and Methods)

	Table 2: Construction of Chromosomal Genome Properties: CDS, Genes, RNA, Hypothetical Protein, Functional Protein, Go assignments, PGfam, Cripsr, etc.	**Deliverable:** (Results)
	PHASE TWO	**WEEK 4**
Genomics and Predictive Metabolite Analysis (Identification of Biomarkers)	**Prediction of Biosynthetic Genes/Secondary Metabolites Using Bagel4 or Antismash:** Identify potential biosynthetic genes and secondary metabolites in microbial, plant, or animal assemble sequence data relevant to targets • Utilize genomics and bioinformatics tools such as Bagel4 or Antismash for predictive analysis. • Apply these tools to diverse datasets to identify potential bioactive compounds that may serve as targets for anti-cancer • Conduct functional analysis to understand the potential roles and functions of the identified genes and metabolites in the context of anti-cancer peptide discovery	
	Write Up 4: Biosynthetic Gene Prediction or Prediction of Secondary Metabolites	**Deliverable:** (Materials and Methods)
	Post-Prediction Result Analysis and Documentation **Table 4:** Table Construction for Predicted Biosynthetic Genes/Secondary Metabolites that includes the following • Relevant features such as gene names, associated pathways, predicted metabolite structures, and any relevant scores or probabilities • Metadata details in the table, such as the source organism, sample type (microbial, plant, or animal), and specific datasets used **Figure 1:** Figure showing predicted biosynthetic genes/secondary metabolites with relevant figures.	**Deliverable:** (Results)
	PHASE THREE	**WEEK 5**
Prediction and Designing of Therapeutic Peptides with Machine Learning	**Machine Learning-Based Therapeutic Peptide Prediction and Design:** Each step in the machine learning-based therapeutic peptide prediction and design process contributes to the specificity, diversity, and customization of the designed peptides, fostering a more efficient and targeted drug development approach.	
	STAGE 1: AMP PREDICTION **CAMPR3 for AMP Collection:** CAMPR3 serves as a specialized machine-learning model for collecting AMPs from biosynthetic genes, enhancing the specificity of peptide selection for further analysis • **Prediction of One 99-Length AMP:** Predict a single 99-length AMP with the highest probability across all models • **Breakdown to Twenty 20-length AMPs:** Break down the prediction to identify twenty 20-length AMPs with the highest probability across the majority of models. **NOTE:** Predicting one 99-length AMP offers insights into the characteristics of longer peptides while breaking down twenty 20-length AMPs ensures diversity and practicality for experimental validation.	**Deliverable:** (Results)
	Write Up 5: Machine Learning-Based Therapeutic Peptide Prediction and Design	**Deliverable:** (Materials and Methods)

	Table 5: Containing the predicted AMPs for each category (99 and 20)	
	STAGE 2: ACTIVITY-BASED PEPTIDE RECONSTRUCTION AND FUNCTIONAL PREDICTION • Reconstructed predicted AMPs from CAMPR3, emphasizing anti-cancer activity, to generate novel peptides with enhanced therapeutic potential. • Prioritized parameters such as SVM scores, N & C terminus preferences, and other details during the redesign process.	**WEEK 6**
	Write Up 6: Activity-based Peptide Reconstruction and Functional Prediction	**Deliverable:** (Materials and Methods)
	Table 6: Contained the reconstructed peptides and parameters of the activities of each peptide	**Deliverable:** (Results)
	PHASE FOUR	
In Silico Characterization of Designed Therapeutic Peptides	**Phytochemical properties of designed therapeutic Peptides, in silico prediction of structure and function of the peptide.** • Hydrophobicity (numbers or ratio) • Number of G and p • Negative or Positive Net charge • Molecular formula • Boman Index • **Experimental Verification of peptide using BLAST** • Estimated half-life • **Predicted peptide toxicity** • Cell penetration and others • Instability Index, Aliphatic index, Grand average of hydropathicity (GRAVY)	**WEEK 7**
	Write Up: Phytochemical Properties and In Silico Prediction of Peptide Structure and Functional	**Deliverable:** (Materials and Methods)
	Visualization of all Peptide/Phytochemical Results (Tables and Figures for All or Most of the Phytochemical Properties)	**Deliverable:** (Results)
	PHASE FIVE	
Computational Drug Design/Molecular Docking	**In Silico Drug Design and Molecular Docking for Therapeutic Peptide Discovery** • Drug target identification and retrieval • Ligand selection of leads hits, and retrieval • Pharmacokinetics analysis and Lipinski rule of 5 • Exploring software used for docking, their peculiarities, similarities, and differences • Protein preparation and ligand preparation process • Exploring docking types based on protein structure: rigid-rigid, rigid-flexible, protein-ligand, protein-protein, protein-DNA • Docking analysis using different software like Pymol, Ds visualizer etc.	**WEEK 8 & 9**
	Write Up: Drug Design and Molecular Docking	**Deliverable:** (Materials and Methods)
	Evaluation and visualization of Results	**Deliverable:** (Results)

RESEARCH PROJECT OUTLINE FOR PUBLICATION	
Introduction	**BTW WEEK 6 -7**
Materials and Methods	**WEEK 9**
Results	**WEEK 10 and 11**
Discussion	**WEEK 12**
Conclusion	**WEEK 12**
References	**WEEK 13**
Certification and Recommendation Letter	
Follow up and Publication	