OMICS**LOGIC**
BIOINFORMATICS EDUCATION

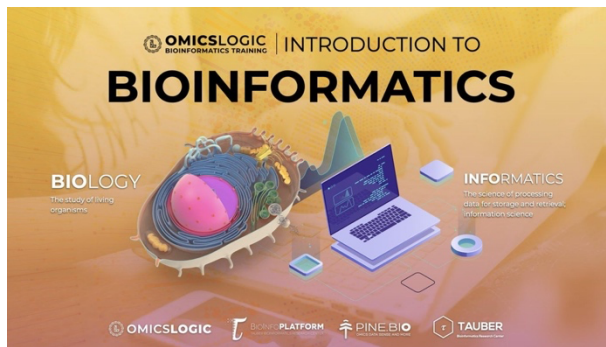|	Omics Logic Bioinformatics	Bioinformatics for Infectious Diseases	SARS-COV2: Genomics	Data Science for Biomedical Data	Research Fellowship
Program Name	**Omics Logic Bioinformatics**	**Bioinformatics for Infectious Diseases**	**SARS-COV2: Genomics**	**Data Science for Biomedical Data**	**Research Fellowship**
Type	**Asynchronous**	**Blended**	**Blended**	**Blended**	**Mentor-guided**
Length	**3 months**	**3 months**	**1 months**	**1 months**	**3 or 6 months**
Live sessions	None	1 per week	2 per week	2 per week	1 per week
Number of Live Group Sessions**	**0**	**12**	**10**	**10**	**12**
Online Chat Assistance	Yes	Yes	Yes	Yes	Yes
Expert Q&A**	NA	Yes	Yes	Yes	Yes
One-one-one support	NA	Yes	Yes	Yes	Yes
Bioinformatics Project	NA	Yes	NA	NA	Yes
Access to Educational Platform	asynchronous* online courses	asynchronous* online courses	asynchronous* online courses	asynchronous* online courses	asynchronous* online courses
Access to Code playground	Yes	Yes	Yes	Yes	Yes
Access to Server	Educational	Educational	Educational	Educational	Educational
Certificate	Program Certificate	Program Certificate	Program Certificate	Program Certificate	Program Certificate + Project

Full Program details (Schedule & Topics)

OmicsLogic Bioinformatics - 3 months asynchronous project-based learning in bioinformatics

A self-paced introduction to bioinformatics, or the intersection of biology and data. Includes a beginner's guide to key terminology and history of computational biology, such as the genomic revolution, next-generation sequencing (NGS) and big data. Provides examples of the application of bioinformatics in every area of life science, including research, biomedical, biotechnology, and agrobiology.

Level: Beginner (Undergraduate)

This program is best suited for students interested to learn about various -omics technologies and assess the significance and applications of computational analysis approaches. Program access includes unlimited and self-guided asynchronous online course, project, and tutorials with basic course certification

- **Flexible schedule:** access to asynchronous coursework for self-guided study
- **Accessibility:** An introduction to bioinformatics for beginners (Bytes and Molecules, Introduction to Bioinformatics, Introduction to Genomics and Metagenomics)
- **Depth:** Overview of 6 key -omics areas: 1) Bioinformatics, 2) Genomics, 3) Transcriptomics, 4) Epigenomics, 5) Metagenomics and 6) Data Science
- **Technical Skills:** Curated project datasets and tutorials for hands-on analysis using big data analysis tools, as well as programming in R and Python
- **Application:** Self-paced research study examples on how to apply omics data analysis in oncology, neuroscience, infectious diseases, and agriculture

ONCOLOGY

Omics data is used in oncology to understand disease onset, risks and progression. In this specialization track, you will learn about how to find cell line, animal model and patient-level data from various cancers and understand how modern oncology is leveraging molecular precision to provide a more precise and personalized treatment for patients.

INFECTIOUS DISEASES

With increasings risks of global pandemics and chronic infectious diseases on the rise, it is important to learn about viral genomics, pathogen drug resistance and vaccine design. This track will allow you to master genomic data analysis, multiple sequence alignment and understand how to study the processes that cause pathogens to infect, replicate and spread.

NEUROSCIENCE

Tremendous efforts are dedicated to the study of the human body, and yet, scientists are puzzled by how our brain works and breaks down. From basic biology to treatment of neurodegenerative diseases, neuroscience is a frontier of research & discovery. In this track, you will learn about the current issues from a data perspective, analyzing genomic, transcriptomic and epigenomic data.

AGRICULTURE

Food security is under threat of a changing climate, pollution and growing demands around the globe. In the study of modern agriculture, molecular data is being used to understand how certain crops can be grown more efficiently, providing the world with a more reliable and healthier future. Here we describe experiments with model organisms and analyze data from crops.

*In this program you will be able to learn about the application of bioinformatics in various areas of research and then how can you start doing it yourself that is to apply NGS applications for the identification of germline and somatic pathogenic variants, measurement of gene expression, detection of methylation patterns on DNA and even studies of microbial communities on human skin, in the gut, lungs, and other organs. This online program is designed for everyone including students who don't have a background in Bioinformatics. The hands-on learning experience is enabled by a cloud-based T-BioInfo analytical platform that allows for the processing and analytics of BIG data on various **NGS data types, Mass-Spectroscopy, Structural Biology, and Machine Learning**. This 3-month self-paced program student tracking of learning progress for automatic grading, access to online tutorials and challenge-based coding practice in R and Python.*
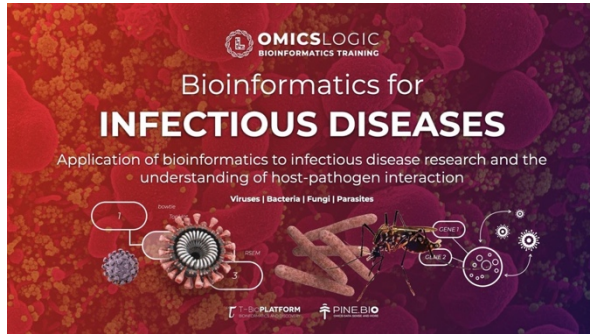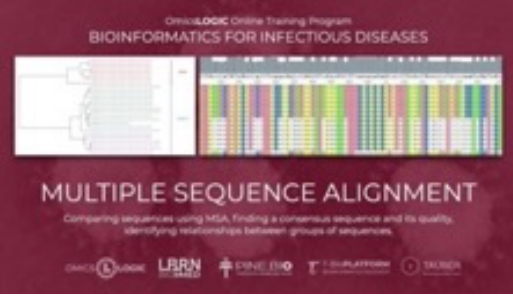
Bioinformatics for Infectious Diseases: 3 months trainer-guided program on the applications of bioinformatics to study human diseases caused by viral, bacterial and parasite pathogens

This program is dedicated to the applications of bioinformatics to the study of pathogen genomics and host responses. We will explore genomic diversity of pathogens in epidemics, study viral zoonotic spillover, apply evolutionary analysis to understand adaptation and explore examples of viral and bacterial disease development. Participants will have the chance to learn about genomics and apply their understanding to public-domain data. As a result, every participant will learn a) to understand relationships between genomes, strains and haplotypes, b) to find differences in genomic sequences, c) to interpret the functional consequences of identified variants, and d) to study host responses connecting pathogen variation with disease and immunity. This program is an opportunity to gain hands-on experience with curated datasets from the public domain as well as guidance and support of bioinformaticians with experience.

This program is recommended for **intermediate to advanced** students, as well as faculty interested in genomics and bioinformatics with some background in microbiology, virology, or molecular biology. Topics covered:

- Analysis of genomic data (consensus and sub-consensus analysis of pathogen genomes)
- Relationship between sequence, structure, and function – (UCSF Chimera to visualize conservation, chemical properties, and functional domains)
- Find predicted B-cell and T-cell epitopes and repertoire using online tools
- Analysis of transcriptomic data (RNA-Seq) for Immunology: host response to infection at the host, organ, tissue, and single-cell level

Sessions	Topics
	Next-generation sequencing: how to find viral genomes in the host transcriptome • Overview of NGS: reads, sequences, file formats • Alignment, annotation, and separation of non-mapped reads for downstream analysis • Alignment to databases of viral genomes with manual validation
	Multiple Sequence Alignment (MSA) and Phylogeny – Reconstructing a Phylogenetic Tree from Alignment • Comparing sequences (Multiple Sequence Alignment) • Finding a consensus sequence from raw sequencing data (reads) • Identifying relationships between sequences (phylogeny, conservation)

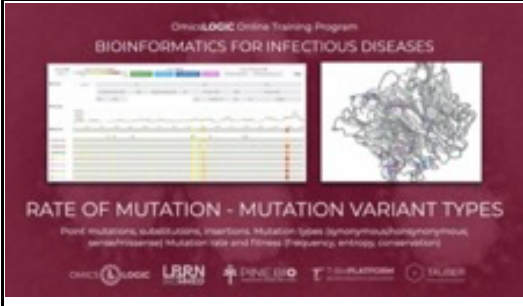|	
	Hands-on session on preparing and running your pipeline in the cloud (using server.t-bio.info) ● Multiple sequence alignment of viral genomes and building a phylogenetic tree ● Finding full genome sequences and preparing FASTA files ● Selecting appropriate genomic sequences ● Preparing a full pipeline of MSA and Phylogeny
	Q&A and DISCUSSION of pipeline results and outputs ● Workflows: what to do if we have FASTA/FASTQ files? ● Which databases to use: Detection of viral genomes by mapping on databases ● Interpretation of Phylogenetic Analysis: Evolutionary relationships between genomes, evolutionary time
	From Infection to Pandemic: using genomics to study viral adaptation and find relationships between sequence and phenotype ● Hosts and origins ● Transmission ● Cell entry and tissue tropism
	Symptom severity: genomic variation associated with disease progression and outcomes ● Viral proteins ● Replication ● Immune evasion
	Hands-on project example and discussion EXAMPLE: the origin of human infection with MERS, SARS, and SARS-2 pandemics
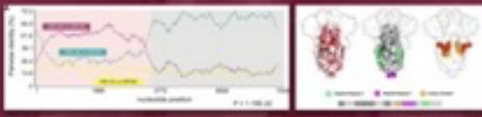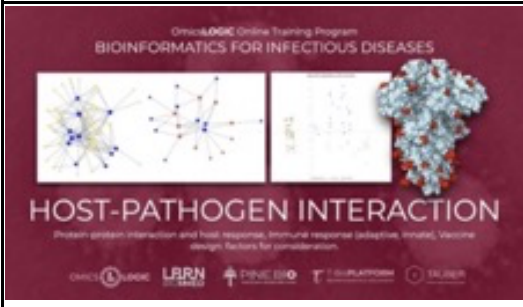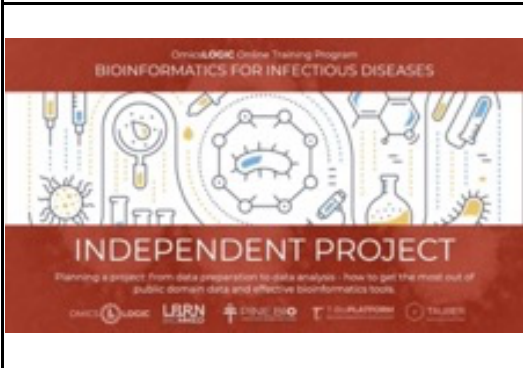
	Rate of Mutation - mutation variant types, low frequency mutations and amino-acid changes ● Point mutations, substitutions, insertions ● Mutation types (synonymous/nonsynonymous; sense/missense) ● Mutation rate and fitness (frequency, entropy, conservation)
	Mutation Annotation & Significance for analysis evaluation and reporting ● Codon/amino-acid and chemical properties ● Location on genome and protein function ● Protein-protein interaction ● B and T-cell epitope prediction
	Host-pathogen interaction: transcriptomic analysis of host response to infection and treatment ● Protein-protein interaction and host response ● Immune responses (adaptive, innate) ● Organism, tissue and single-cell resolution of host gene expression
	Final Review & Project proposal ● Planning your project proposal, ● How to present your scientific data and hypothesis ● Audio & Video Presentation, ● Case Studies & Publications, ● Datasets.

Upon completion of all the requirements, participants will receive a certificate of completion from the Louisiana Biomedical Research Network.
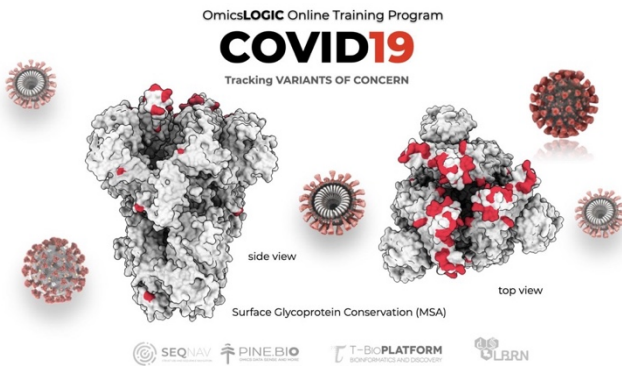
COVID-19 Genomics: 1 month program on the genomics and epidemiology of SARS-CoV2 pandemic

In this program, we will learn about the way the SARS-COV-2 pandemic has transformed our appreciation of genomics and bioinformatics. Participants will learn about genomic data analysis tools that can be used to identify specific viral strains, understand multiple sequence alignment, phylogenetic analysis, and the significance of mutations in the context of viral protein structure and function. We will further discuss the viral genome of SARS-CoV-2 (pathogen causing Covid-19) structure: genes, sub-genomic DNA fragments, proteins, and the virology of the disease. We will discuss how the data and the various analysis tools can help in the characterization of viral genomes, compare, and distinguish between viral strains, identify the impact of mutations on the functionality of viral proteins, and discuss the emerging challenges with the tracking of Variants of Concern (VOCs) reported in media.

This program is geared towards a **beginner level**. Topics:

- Using online repositories like NCBI and GISAID to find and download data
- Use graphical interfaces for phylogenetic tree reconstruction and molecular evolution
- The application of coding to genomic data analysis and variant detection
- Tracking and reporting variants of concern reported in the news

LBRN SARS COV-2 : COVID 19 Genomics & VOC- 2021 (1-Month)		
Sessions	**Topics**	**Dates**
	Using NGS data to find a new pathogen: Bioinformatics pipelines and processing steps to structure genomic data using BowTie, HiSat and STAR, annotation of identified sequences Using the NCBI virus reference database.	08 June 2021
	Extracting reads that did not map to the host genome from FASTQ files after alignment, visualizing how they align to specific viral genomes from a database of viral genomes. Understanding genomic variation in short reads.	10 June 2021

|		
COVID-19 GENOMICS **USING GENOMIC REFERENCE DATABASES OF VIRAL SEQUENCES** *Detection of viral genomes in host samples (Part 2)*	Extracting reads that did not map to the host genome from FASTQ files after alignment, visualizing how they align to specific viral genomes from a database of viral genomes. Understanding genomic variation in short reads.	15 June 2021
COVID-19 GENOMICS **NCBI & GISAID: HIGH QUALITY COVID-19 GENOMIC DATA** *Finding and preparing files for analysis on NCBI Virus*	Public resources where SARS-COV-2 data is made available. Types of databases, access control and utilization. Finding the right genomic sequences, using NCBI alignment to check for quality and preparing data for analysis.	17 June 2021
COVID-19 GENOMICS **HANDS-ON SESSION: PREPARING AND RUNNING YOUR PIPELINE**	Bioinformatics – a step-by-step overview of a pipeline that is used to align sequences, translate trinucleotide segments into amino acids and use the alignment for phylogenetic analysis using BEAST.	22 June 2021
COVID-19 GENOMICS **TRENDS BEHIND THE EMERGENCE OF SARS-COV-2 VARIANTS (VARIANTS OF CONCERN, VOC)**	Trends behind the emergence of SARS-COV-2 variants (variants of concern, VOC) Nature of emerging variants - why they are causing concern and what might be the driving forces behind the observed clades or substrings of this virus as it continues to spread around the world.	24 June 2021
COVID-19 GENOMICS **GENOMIC SURVEILLANCE: DATA COLLECTION, SEQUENCING AND BIOINFORMATICS ANALYSIS**	Hands-on training to understand the logic behind analysis of variants of concern and develop a workflow for variant reporting using curated data from clinical and environmental sequencing.	29 June 2021
COVID-19 GENOMICS **DOWNSTREAM ANALYSIS OF GENOMIC DATA: EVOLUTIONARY STUDIES/ PHYLOGENY, ORIGIN TRACING, REPORTING**	Studying evolutionary analysis of viral genomes – comparison of variable regions and identification of genomic variation. Mutations, conservation, and viral evolution. Modeling relationships between sequences based on probability (phylogeny, evolution and conservation)	01 July 2021

	Hands-on session to load files from an NGS pipeline into R and use custom scripting to produce a report on variants identified as consensus and sub-consensus levels for detection of major and minor variants	06 July 2021
	Future expectations as this type of data is likely to grow in abundance when many countries around the world are investing billions of dollars in sequencing, variant detection, vaccination and other interventions for SARS-COV2	08 July 2021

Upon completion of all the requirements, participants will receive a certificate of completion from the Louisiana Biomedical Research Network.
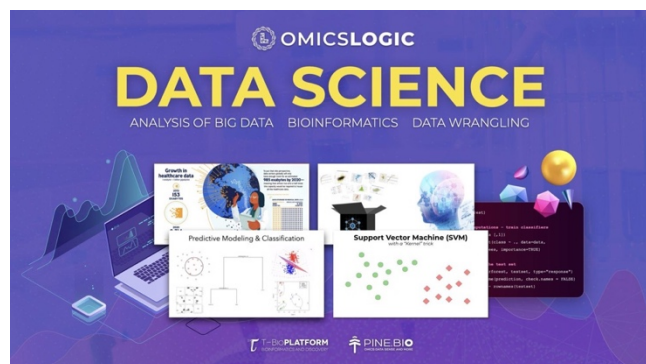
Data Science for Biomedical Data: 1 month program

The rapid growth of high-throughput data, including -omics technologies, gave rise to significant demand for data science skills and experience with bioinformatics methods of analysis. This online training program is designed for beginners and students interested in data-driven research questions. The program will include aspects of data science, such as data wrangling, visualization, statistical analysis, and machine learning. The methods will be reviewed in the context of biomedical and other scientific problems.

- Big data challenges, HPC and cloud computing
- NGS: omics data types and use cases
- Processing data using computational pipelines
- Introduction to Programming: R and Python
  - 1. Data visualization
  - 2. Statistical Analysis
  - 3. Machine Learning
  - 4. Annotation using databases

Summer Training in Biomedical Data Science (31 May - 02 July 2021)		
Sessions	**Topics**	**Date**

#DATASCIENCE OMICS2021 — OMICS: Introduction to data types and properties	**Session 1: Omics: Introduction to data types and properties** ● Overview of commonly used "omics" data ● NGS, Mass-Spec, phenotypic data (genomics, transcriptomics, metagenomics) ● Phenotypes: clinical, imaging, metadata (research, clinical, biotech, pharma) ● The need for preparation of raw data for analysis	31 May 2021
#DATASCIENCE OMICS2021 — Big Data Challenges and Opportunities	**Session 2: Big Data Challenges and Opportunities (conceptual and computational)** ● Availability and variability of data ● Unprecedented Detail and volume ● Data heterogeneity, complexity, and noise ● Need for structure and reproducibility	4 June 2021
#DATASCIENCE OMICS2021 — Cleaning Loading and Processing data	**Session 3: Cleaning, loading and processing data (Logical steps and a practice)** ● Analysis logic: from raw reads to a table of expression (RNA-seq example) ● Common sources of unwanted technical variation ● pre-processing steps, filtering and cleaning the table of expression ● Loading processed data for analysis	7 June 2021
#DATASCIENCE OMICS2021 — Exploratory data analysis: summary and visualization	**Session 4: Exploratory data analysis: data summary and effective visualization** ● Summary statistics (histogram, boxplot, a scatterplot of 2 samples compared to each other, Excel "summary statistics" operation) ● Visualization of practice data - compare the regular and ln scale of gene expression and discuss distribution and log-normal distribution ● Missing data and data errors (remove 0s, filter anything below 2 in ln scale in R) ● Summary statistics in R	11 June 2021
#DATASCIENCE OMICS2021 — Hands-on: handling large and complex datasets	**Session 5: Hands-on: handling large and complex data (data properties, statistical summary, and preparation for analysis)** ● Learn how to make statistical representations of the data and how to address missing or data errors. ● How do you compare the same gene from different samples in the same condition and how to compare all genes between different conditions	14 June 2021

	● How to find a sample that has poor quality reads or lots of missing gene, low expression (outliers) ● 3 types of data: gene expression (continuous), clinical (categorical) and drug response (LD50 - continuous, but different variance, fewer features)	
	Session 6 - Introduction to Machine Learning (ML) and Artificial Intelligence (AI) ● Hypothesis testing 101: compare conditions and find the p-value ● Data-driven discovery: discover groups or conditions ● Process of inference for a machine versus human. (What and how machines learn.)	18 June 2021
	Session 7 - Unsupervised Machine Learning: dimensionality reduction and clustering ● Finding patterns in the data and methods of data mining. ● PCA, k-means, h-clustering (run example on T-Bio and then open the script in R and modify it)	21 June 2021
	Session 8 - Supervised Machine Learning: classification and feature selection ● Conceptual Introduction: Known sample data is used to train the computer to use these patterns to correlate to unknown data. ● Binary decision trees, random forest, then LDA, then swLDA.	25 June 2021
	Session 9 - Model accuracy and validation ● Technical accuracy (ROC curve) ● Logical or biological relevance (compare feature selection with PCA by subtype or clinical phenotype) ● Trained Model validation: Learning how a model used to analyze data is accurate and valid across multiple datasets. ● Hands-on example: cross-validation, Leave 1 out analysis	28 June 2021

	Session 10 - ML in production ● The interaction between artificial intelligence and human ● Differences between ML and AI ● In what ways can AI support human research and decision-making ● Training and research extractions are applied in new ways.	02 July 2021

Research Fellowship: 3- or 6-months program with expert guidance on the development of an independent research project

The research fellowship program is designed to provide support and guidance on development of an independent research project within a given timeframe (3-6 months). During this time, participants meet on a weekly basis to share their progress and get feedback from the program coordinators. They can also request one-on-one sessions with a mentor on a weekly basis or meet in smaller focus groups. The program is structured around access to various training sessions, weekly progress updates and reviews or focused group meetings on topics of special interest with an expert mentor.

Recommended level: depending on student objective or interest, this program can be suited for anyone interested to learn and develop a project that can be used for a poster presentation or as a draft for a publication using in-silico analysis methods. The program is conducted in these stages:

Stage 1: Introduction and Orientation

Introduction to bioinformatics theory, methodology and applications:

- cloud tools for big data processing
- Introductory coursework
- NGS data types: Genomics, Transcriptomics, Metagenomics, Epigenomics

Stage 2: Evaluation and Assessment

Independent project topic selection through literature review:

- Research Proposal (Project Title, Project Description, Reference Publication & Datasets),
- Experimental design: Pipelines & Algorithms, Data Processing, and normalization.

Stage 3: Practicum

Practical coursework and exemplar projects in cancer, neuroscience, agriculture, biotechnology and infectious diseases:

- Processing, analysis, and integration methods
- Reproducibility of research findings
- Extracting meaningful insights from large datasets
- Controlled and "real-world" experiments and data sources

Stage 4: Comprehension and Analysis

Exploratory analysis of a focused set of publicly available datasets to identify major trends and patterns for investigation:

- Overview of project meta-data, practice of analytical tests
- Processing of data for analysis
- Design of a statistical metric for inference
- Analysis review and internal presentation

Stage 5: Investigation

In-depth review and practice of analytical methods, best data science practices and algorithms applied to data type of choice

- Exploratory data analysis
- Data-driven research hypothesis
- Statistical testing and inference
- Unsupervised and supervised Machine Learning

Stage 6: Writing, Editing, and preparing the Project for submission

Preparing the project report for an external audience:

- Background, significance, methodology and results
- Discussion of project findings and limitations
- Comparison to other research work

The Research Fellowship program is designed for those interested in the intersection of data and biology for a research project. Most research fellows come from a life science background and learn to use bioinformatics tools in the first month of the program. Program length can be 3-6 months and follows a well-defined structure with mentor guidance completely done online via ZOOM in individual and group settings.

<u>**Outcomes**</u>

Outcome 1: Understanding of challenges associated with bioinformatics

Even if the research fellow does not complete the program with a project, they will get to experience what a research activity is like, how it is different from training and what are the challenges with independent research.

Outcome 2: Training and Skill Development

Another possible outcome is to get familiar with analytical methods and approaches, including some level of hands-on experience. As a result of completing such training, students get a certificate of participation and course completion that reflects the skills they have learned and mastered.

Outcome 3: Independent or Group Research Project

A typical outcome is to get a project completed independently or as a group and present it to a group of fellows in our sessions. This can also be adapted to a science fair, conference poster or a class presentation.

Outcome 4: Poster presentation or Research Publication

For a competitive project, the next step is to submit the manuscript to a journal. That still does not guarantee acceptance - for example, many manuscripts take years to revise, edit and complete before passing peer-review. However, there are many journals that accept research projects developed by novices, high school students and undergraduates.

Registration is on a rolling schedule.

For any additional questions, you can view previous research project submissions and outcomes from this program: https://learn.omicslogic.com/blog/post/research-fellowship-with-pine-biotech-independent-research-projects-using-bioinformatics