



2020 LBRN Virtual Summer Program

(for undergraduate, graduate students, faculty and staff of LBRN Institutions)

The Louisiana Biomedical Research Network (LBRN) Summer Program is Supported by the Louisiana Board of Regents and NIH:NIGMS P20GM103424

Institution									
Program	Omics Logic Basics	Bioinformatics for Infectious Diseases	SARS-CoV2: Genomic Data Analysis	Information Visualization	Computational -Aided Drug Discovery (CADD) of Anti-Viral Therapeutics for COVID-19	Bacteriophage Investigations – in Silico Bacteriophage Annotation Project	Computer Aided Recognition (CAR) System	Quantum Dots Imaging Project	
Length (months)	3	2	1	2	2	2	2	2	
Number of Students	50	15	15	12	24	15	3	3	
Instructor	LBRN / PINE.BIO	LBRN / PINE.BIO	LBRN / PINE.BIO	Dr. Marjan Trutschl	Dr. Elahe Mahdavian	Dr. Ann Findley / Dr. Chris Gissendanner	Dr. Omer Soysal	Dr. Patrick Moyer	
Certificate	Course Certificate	Program Certificate	Program Certificate	LBRN Certificate of Completion					
				Invitation to present at either the LBRN annual meeting or Annual Bioinformatics Conference					

LBRN Summer Program Options
I - Pine Biotech

2020 Summer Bioinformatics Programs Details:

1. OmicsLogic Online 3-months

This program is best suited for students interested to learn about various -omics technologies and how bioinformatics is used in biotechnology, healthcare, agriculture and basic research. Program access provides access to all the asynchronous* online courses (basic course certificates only):

1. **Introduction to bioinformatics (1 course)**
2. **Genomics (3 courses)**
3. **Transcriptomics (4 courses)**
4. **Metagenomics (3 courses)**
5. **Epigenomics (4 courses)**
6. **Biomedical Data Science (3 courses)**

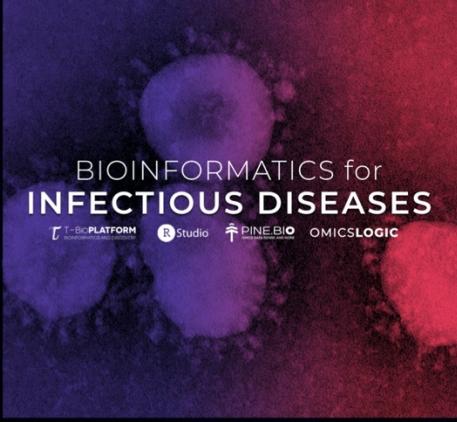
Program Length: 3 months, certification: individual course

Each course includes quizzes and some courses include practical assignments that can be completed before program end to receive certificates of completion. A passing quiz grade is required for completion. Assignments are checked for pass/no pass and can be resubmitted after feedback.

Level of difficulty: intermediate, great for a flexible schedule and independent learning.

Bioinformatics for Infectious Diseases

Bioinformatics for Infectious Diseases



BIOINFORMATICS for INFECTION DISEASES

T-BioPLATFORM R Studio PINE.BIO OMICSLOGIC

T-BioPLATFORM PINE.BIO TAUBER

BIOINFORMATICS AND DISCOVERY OMICS DATA SENSE AND MORE BIOINFORMATICS RESEARCH CENTER

Understand Viruses and Infectious Diseases:

- Zoonotic Transmission
- Viral and Bacterial infection
- Replication and Diversity

Learn about Analysis Approaches to

- Detection of Viral genomes in NGS Data
- Find differences between strains and haplotypes
- Vaccine Design, Antivirals and Antibiotics
- Epidemiological Trends and Drug resistance

Curated Projects and Datasets

- Recent epidemics like Ebola, Zika, Dengue
- Corona viruses like SARS, MERS
- Mycobacterium: Tuberculosis and Leprosy

This program is dedicated to the study of viral diversity and its role in epidemic infectious diseases that keep re-emerging, including zoonotic spillover, transmission between humans and the process of viral and bacterial disease development. Participants will get a chance to learn about bioinformatics and analyze genomic data by applying various analysis approaches to study viral genomes. As a result, you will learn to understand relationships between viral strains and haplotypes, find differences in sequence data and see the implications for drug and vaccine design. This program will provide opportunities to practice analyzing data to gain hands-on experience with curated datasets from public domain collections, guided by experts with bioinformatics experience and knowledge about virology.

The topics we will cover include:

- Finding genomic data from epidemic outbreaks and research projects (i.e. genomic sequences and NGS data)
- phylogenetics and Multiple Sequence Alignment (MSA)
- Downstream analysis of genomic data (differential mutations, data mining and association with phenotype)
- GWAS studies for viral and bacterial genomes
- Variation mapping on protein structures

Bioinformatics for Infectious Diseases is an online training program designed for biologists, clinicians and students that are interested in virology and immunology and would like to learn about the use of bioinformatics and big data for infectious disease research, diagnostics as well as drug and vaccine development.

In this program, you will:

1. Understand Viral Diversity and its role in epidemic Infectious Diseases that keep re-emerging, including zoonotic transmission, transmission between humans and the process of viral and bacterial disease development
2. Learn about Bioinformatics Analysis Approaches to study viral genomes and understand relationships between viral strains and haplotypes, finding differences in sequence data and seeing the implications for drug and vaccine design
3. Gain hands-on experience by analyzing Curated Datasets from public domain collections, guided by experts with bioinformatics experience and knowledge about virology.

Program Topics:

- Finding genomic data from epidemic outbreaks and research projects (i.e. genomic sequences and NGS data)
- phylogenetics and Multiple Sequence Alignment (MSA)
- Downstream analysis of genomic data (differential mutations, data mining and association with phenotype)
- GWAS studies for viral and bacterial genomes
- Variation mapping on protein structures

Program Syllabus:

Part1: INTRODUCTION:

Introduction to bioinformatics basics and data sources:

- 1. Next-generation sequencing: viral genomes in host transcriptome**
 - Overview of NGS: reads, sequences, file formats
 - alignment, annotation and non-mapped reads
 - Alignment to databases of viral genomes
- 2. Multiple Sequence Alignment and Phylogeny**
 - Comparing sequences (Multiple Sequence Alignment)
 - Finding a consensus sequence
 - Identifying relationships between sequences (phylogeny, conservation)
- 3. Hands-on session, preparing and running your pipeline:**
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
- 4. Q&A and DISCUSSION of pipeline results:**
 - 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

Part 2: Genomics and Virology:**How can bioinformatics be used to study pandemics?**

5. From Infection to Pandemic: viral adaptation

- Hosts and origins
- Transmission
- Cell entry and tissue tropism

6. Symptom severity: Disease progression and outcomes

- Viral proteins
- Replication
- Immune evasion

7. Hands-on project discussion

EXAMPLE: the origin of human infection with MERS, SARS, and SARS-2 pandemics

8. Rate of Mutation - mutation variant types

- Point mutations, substitutions, insertions
- Mutation types (synonymous/h nonsynonymous; sense/missense)
- Mutation rate and fitness (frequency, entropy, conservation)

9. Mutation Annotation & Significance for analysis

- Codon/amino-acid and chemical properties
- Location on genome and protein function

10. Host-pathogen interaction

- Protein-protein interaction and host response
- Immune response (adaptive, innate)
- Vaccine design: factors for consideration

CONCLUSION and OUTCOMES:

In this program, we will learn about important principles of bioinformatics in application to virology, including:

Use of bioinformatics in virology

- Methods of analysis
- Databases and references
- Raw data types and repositories

Important factors for vaccine design

- Exposed parts of viral proteins
- Vaccine types: protein-based, virus-based
- Novel approaches to rapid-response vaccines: interference

Important factors for antivirals

- Prevention of cell entry
- Inhibition of replication
- Toxicity, specificity
- Solubility, permeability

Case studies we will utilize in this program:

- Coronaviruses and the recent COVID-19 epidemic
- EBOLA outbreaks over the last decade: emerging diseases

- Flu and other respiratory disorders: challenges with vaccines and antivirals
- Tuberculosis: bacterial chronic diseases and antibiotic resistance

Bioinformatics methods:

Alignment, annotation, characterization of mutations, GWAS, phylogenetic analysis, RNA-seq

SARS-COV-2: Genomic Data Analysis

Using genomics to understand the COVID-19 pandemic (high school students and undergraduate college students interested in research as well as “citizen scientists”)



The landing page features a purple background with a central graphic of the SARS-CoV-2 virus structure. To the left, large yellow text reads "STAY@ HOME& LEARN ABOUT THE VIRUS". Below this is a description of the program's goals. To the right, a section titled "INTRODUCTORY COURSE" is titled "ANALYSIS OF SARS-COV-2 GENOMIC DATA". It includes four numbered steps with icons: 1. Learn where to find appropriate genomic data for analysis (magnifying glass icon), 2. Analyze big data using intuitive cloud based tools (monitor icon), 3. Read COVID-19 publications with understanding (document icon), and 4. Understand key principles of vaccine and drug design (test tube icon). Logos for BioInfoPlatform, PINE BIO, and TAUBER are at the bottom.

COVID19

OMICSLOGIC
BIOINFORMATICS TRAINING

STAY@ HOME& LEARN
ABOUT THE VIRUS

In this program, you will learn about genomic data, where to find it and how to analyze it. By analyzing publicly available data you will learn about the COVID-19 pandemic origin, evolution, cell entry and replication.

JOIN TODAY!

© Pine Biotech 2020. All rights reserved.

INTRODUCTORY COURSE

ANALYSIS OF SARS-COV-2 GENOMIC DATA

- 1 Learn where to find appropriate genomic data for analysis
- 2 Analyze big data using intuitive cloud based tools
- 3 Read COVID-19 publications with understanding
- 4 Understand key principles of vaccine and drug design

BIOINFOPLATFORM PINE BIO TAUBER

Program Topics:

The 2019–20 coronavirus pandemic has had far-reaching consequences lethal for many and disruptive to almost everyone around the world. Many of us are now interested to contribute our efforts to study, research and understand how we **COVID-19** can be stopped and what can be done to prevent such pandemics in the future. The Pine Biotech team along with our collaborators at the Tauber Bioinformatics Research Center, Louisiana Biomedical Research Network and Georgetown University Medical Center has been actively working to provide computational resources and training materials to scientists, researchers and students all over the globe who want to learn more about this virus. As a citizen scientist, you too can be part of this global initiative and contribute by understanding what data is available and how it can be studied using **bioinformatics**.

In this program, you will learn about the novel coronavirus genomics and understand how genomic data analysis tools can help identify specific viral strains, understand how they differ from previously studied viruses by using such tools as multiple sequence alignment, phylogenetic analysis and characterization of mutations in the context of viral protein structure and function. We will apply these analytical methods to publicly available SARS-CoV2 genomic data, learning about the virus emergence, spread and disease pathology. As a result, you will have the chance to perform your own analysis and develop a research question related to this pandemic.

This online program includes: Instructor support via online interactive sessions, tracking of student progress via online course and pipeline activity, access to online educational resources and analysis tools for self-paced learning, access to bioinformatics experts for guidance and problem-solving as well as educational use of analytical pipelines. SARS-COV-2: Genomic Data Analysis program is supported by our cloud-based tools that enables all program participants to analyze complex genomic data generated during this pandemic without having to invest in high-end computational infrastructure.

Registration Details: Online Coursework and 10 ZOOM sessions (total of 15 hours), Duration: 30 days training, followed by an optional research license for independent research.

Program Outcomes:

- A. Learn where to find appropriate genomic data for analysis
- B. Analyze publicly available data using intuitive cloud-based tools
- C. Learn to read research publications on this topic with understanding
- D. Understand key principles of vaccine & antiviral drug design

SARS-COV2 Genomics Program Syllabus

1. What is Next-generation sequencing?

Learn about Next Generation Sequencing techniques used to detect viral genomes in clinical samples and generate genomic data about pathogens and infected cells: short reads, genomic sequences, file formats and laboratory techniques for data preparation.

Associated online course/resource: Introduction to Metagenomics, Metagenomics 1

2. Looking for viral genomes in a sample (part 1)

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.

Associated online course/resource: Spodoptera frugiperda and the contamination of biologics production (<https://edu.t-bio.info/projects/spodoptera-frugiperda-fall-armyworm/>)

3. Identifying a novel virus genome

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.

Associated online course/resource: Introduction to Genomics

4. Multiple Sequence Alignment and Phylogeny

How closely related are different viruses? Comparing genomic sequences (Multiple Sequence Alignment) and finding a consensus sequence (ShoRAH) in FASTQ files or between multiple FASTA files available on NCBI.

Associated online course/resource: Introduction to Metagenomics

5. Finding reliable genomic data on COVID-19 on NCBI

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.

Associated online course/resource: NCBI Virus Database

6. Phylogenetic Analysis and Probable Origins of SARS-COV2

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)

Associated online course/resource: Genomics 1

7. Hands-on session: using found data to prepare and run a bioinformatics 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.

8. Q&A and DISCUSSION of pipeline results

Different Workflows: what to do if we have FASTA or FASTQ files? Which databases to use: Detection of viral genomes by mapping on databases. Interpretation of Phylogenetic Analysis: Evolutionary relationships between genomes, evolutionary time and specific changes of interest.

9. Discussion about Program topics and literature review

Utilization of NGS genomics to study viral diseases and understand emergence and evolution of pandemic outbreaks. Zoonotic spillover events and cross-species viral adaptation. Analysis of viral genomic data to explore potential SARS-COV2 origins, and transmission by finding and analyzing public domain data.

10. Bioinformatics in vaccine research and antiviral drug design

In conclusion, we will review how recent examples from the current and previous outbreaks of viral infections (i.e. Ebola 2014-2016) reveal the significance of widely accessible and accurate bioinformatics tools as well as the significance of bioinformatics skills for data-driven life science research enabled by high throughput data and rapid molecular sequencing. The direct outcome of evolutionary studies, functional annotation and publicly accessible genomic data enables the translation of research findings into vaccine and antiviral drug applications.

LBRN Summer Program Options
II - LSU-S

Dr. Marjan Trutschl

Information Visualization (CSC464/ CSC664 LSUS summer online course)

Systems administrators, biomedical researchers, banks, fraud and [cyber]security investigators as well as other entities possess vast quantities of high-dimensional data (a.k.a., big data), which hide fundamental relations that need to be exploited to their advantage. This explains the demand for big data analysts, which is at an all-time high!*

The ability to view and explore graphical representations of such data and being able to say 'Ah Ha! That is interesting!' is the focus of Information Visualization.

The focus of this hands-on online course is examining the issues covering a wide range of visual displays, from classic to novel. Students will learn about visualizations and visualization packages (such as Tableau, 3D.js, Circos, Google Charts, etc.) and how to apply them to various data.

Understanding the principles of Information Visualization will enable students to evaluate existing and create new visualizations and visualization systems, with the goal of effectively analyzing the data from various sources.

Upon completion of this course, the students will have sufficient knowledge to prove their technical skills with a Tableau Specialist Exam and will be on their way towards passing the Tableau Certified Associate Exam, the leading data visualization tool in business intelligence and data visualization software industry.

How many students could you accommodate

12

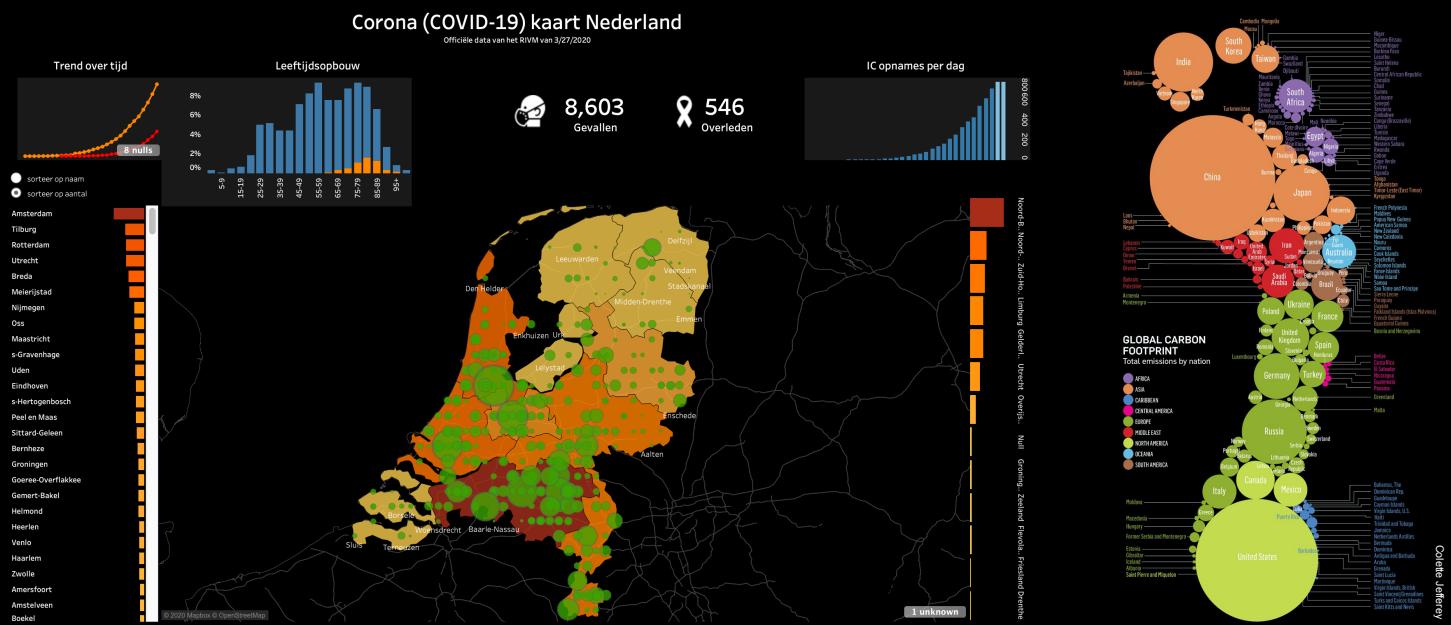
How will this be delivered - what requirements would there be for students to participate

- Priority will be given to
 - graduate students
 - students who will be seniors and juniors during the Fall 2020 semester
 - students with knowledge of
 - programing
 - statistics
 - Linux or command-line
- Students need Internet access and a computer with the following minimum specifications to run:
 - Mac
 - iMac/MacBook computers 2009 or newer
 - macOS High Sierra 10.13, macOS Mojave 10.14 and macOS Catalina 10.15
 - 1.5 GB minimum free disk space
 - Windows
 - Microsoft Windows 7 or newer (x64)
 - Microsoft Server 2008 R2 or newer
 - 2 GB memory
 - 1.5 GB minimum free disk space

Projected start would be June 8 (with students application and selection done before then)

The summer online course at LSUS runs from June 1 to July 31, 2020. If students join after the semester begins they will have a bit extra work the first week of participation.

Students certificates - would you be able to "grade" them and determine if they are successful (and how)
Yes, they can earn a certificate by completing the [minimum certificate] assigned work with 80% or better.



Information Visualization

CSC464 / CSC664

Systems administrators, biomedical researchers, banks, fraud and [cyber]security investigators as well as other entities possess vast quantities of high-dimensional data (a.k.a., big data), which hide fundamental relations that need to be exploited to their advantage. This explains the demand for big data analysts, which is at an all-time high!*

The ability to view and explore graphical representations of such data and being able to say '*Ah Ha! That is interesting!*' is the focus of Information Visualization.

The focus of this hands-on online course is examining the issues covering a wide range of visual displays, from classic to novel. Students will learn about visualizations and visualization packages (such as 3D.js, Circos, Google Charts, Tableau, etc.) and how to apply them to various data.

Understanding the principles of Information Visualization will enable students to evaluate existing and create new visualizations and visualization systems, with the goal of effectively analyzing the data from various sources.

Dr. Elahe Mahdavian

Title: A project-based course in computational and virtual medicinal chemistry

Instructor: Dr. Elahe Mahdavian, emahdavi@lsus.edu

Learning Outcomes: The students will recognize the availability and the ease of retrieval of a substantial amount of annotated chemical and structural biology data and how these are leveraged in inquiry-based research projects. Web-accessible databases and cheminformatics tools facilitate linking of multi-layered and cross-referenced data into an effective framework for early-stage drug design, discovery, and optimization. The students will also gain a practical knowledge of how the structure of drug molecules and biological protein receptors are represented to model and understand their similarities, physiochemical properties, and their interactions that lead to biological/therapeutic responses. Subsequently, they will use the data to create hypotheses that can be tested and validated both virtually and experimentally.

Prerequisite: One year of freshman chemistry and freshman biology courses; scripting skill is not required

Maximum number of students: 24

Assessment methods: Hands-on exercises, case studies, and a final report in a poster format. Students will be encouraged to present their posters in the 2021 annual LBRN meeting (in person or virtual!)

Description:

Computational-Aided Drug Discovery (CADD) of Anti-Viral Therapeutics for COVID-19

The novel coronavirus, SARS-CoV-2 is the causative agent for COVID-19, and it has created an unprecedented global viral pandemic with a significant negatively disruptive impact on human health and the economy. Due to the lack of approved drugs for COVID-19, there is an urgent unmet medical need for new and effective therapeutics against this deadly virus. This course provides a practical overview of how modern computational and cheminformatics tools are now used as virtual shortcuts in making drug discovery pipeline more efficient in time and cost and less risky. Concepts and technical skills related to CADD, including virtual **high-throughput drug screening**, ligand- and structure-based drug library design, and molecular docking will be discussed. This course will enable students to integrate various public domain and proprietary computational tools in the context of an anti-viral drug library design and optimization process (a virtual drug to a first-in-human drug) for COVID-19. This includes:

- I. **Molecular Fingerprints and Similarity Searching.** We will use a three-pronged metric framework to expand the drug library for chemotype diversity based on a promising antiviral drug, arbidol, these include: **Chemical** (2D/3D structure, electrophores, spectrophores); **Biological** (binding affinity, bioactivity), and **Pharmacophore** descriptors.
Tools and Databases: Swiss *Bioinformatics Institute* (SIB), SEA, UniProt, Novartis *HTP-FP*, and *Open Babel*
- II. **Chemical Structure Optimization.** We will use standard protocols to prepare the energy-minimized conformational ensembles for the drugs and the respective protein targets (SPIKE and SPIKE-ACE2 complex).
Tools: *Spartan* and *YASARA*
- III. **Virtual Molecular Docking.** We will include practical predictive application of ligand binding sites in targets, binding modes in drug-protein complexes, and scoring functions for drug binding affinity to their respective targets.
Tools and Databases: *MOE-CCG*, *AutoDock Vina*, *BIND*, *Binding Interface Database*
- IV. **Selection of Drug Candidates for Experimental Validation.** We will use a two-pronged metric framework to select four final drug candidates for phenotypic viral cell-based assays. The first approach is ligand-based, which includes the application of the lead-triage concept to filter out drugs with undesirable physiochemical properties and those that are over-promiscuous using ADMET, PAINS, Lipinski's ROF filters. The second approach is structure-based, the application of dynamic molecular modeling and simulations to rank drugs for their affinities towards the protein targets of interest (SPIKE & SPIKE-ACE-2) and the corresponding stability of those complexes.

LBRN Summer Program Options
III - ULM

**Bacteriophage Investigations – *in Silico*
A Bacteriophage Annotation Project
Spring, 2020**

Instructors: Dr. Ann Findley and Dr. Chris Gissendanner

Phone: Findley: 342-1817; Gissendanner: 342-3314

e-mail: Findley: afindley@ulm.edu; Gissendanner: gissendanner@ulm.edu

Course Objectives: Congratulations! You have been selected to participate in a unique and challenging project that will provide you an opportunity to make a real and important contribution to the scientific community. The Howard Hughes Medical Institute Science Education Alliance is a national, collaborative project consisting of twelve institutions of higher education. The purpose of this project, and this course, is to introduce undergraduate biology students to real, discovery-based science. The aim of this program is to discover and characterize a new Actinobacter phage (viruses that infect bacteria of the various Actinobacter species). The data generated will be incorporated with the data of other participating institutions. The discoveries made in this lab will increase our knowledge of bacteriophage diversity, genetics, and evolution. It is data that will be published and disseminated to other scientists world-wide. Therefore, you will not just be a student; you will also be a scientist. The University of Louisiana at Monroe is a member of Cohort 1 of the HHMI-SEA PHAGES Program and during the past 10 years has worked with over 250 ULM undergraduates to successfully isolate over 150 novel phages. To date, 25 ULM-annotated phages have appeared in GenBank and several of these novel isolates have been featured in peer-reviewed publications.

Course Description: This project is traditionally delivered as a two-semester course sequence. In the first semester wet-lab portion, the focus is on the isolation of new Actinobacter bacteriophage from the environment. Students generate a pure, high-titer culture of these phages and characterize their structure using electron microscopy. In addition, they isolate and purify the genomic DNA of these bacteriophages and submit this DNA for sequencing to the University of Pittsburgh. The second portion of this project involves the characterization and annotation of the genomic DNA.

Course Reading Materials: You will be provided with an electronically-accessible laboratory manual and additional reading materials.

Student Expectations: As mentioned above, you will be both a student and a scientist in this course. Therefore, you will be expected to quickly master techniques and the relevant quantitative skills. Since you are generating scientific data that will ultimately be published, you will be required to properly maintain a laboratory notebook and you will be held to the same ethical and professional standards required of all practicing scientists. The instructors reserve the right to remove any students, at any time, that fail to meet these standards.

Instructional Methods and Activities: Labs will incorporate both Zoom lectures and computer lab exercises. Students will complete annotation exercises either in groups or individually.

Evaluation and Assessment: Assessment will be based on the following criteria:

Conduct (safety, attitude, participation, professionalism, collegiality, and ethics) - 20%

Lab notebook (gene product annotation notes) and postings to DNA Master - 30%

Mastery of annotation scheme and associated topics (survey responses; reflection essay submission) – 30%

Assignments/Quizzes/Moodle Postings - 20%

Minimum Requirements for Enrollment in this Project:

- Personal computer and high-speed internet access
- Basic computer skills – can download/upload files, ability to perform web-based searches
- Basic knowledge of molecular biology and genetics
- Sophomore, junior, or senior undergraduate status
- Past participation in an independent research project or the SEA PHAGES program is a plus but not required

Proposed schedule:

Week of June 15: General overview; *in silico* resource guide; DNA sequencing video

Week of June 22: Introduction to genome organization and annotation; DNA Master

Week of June 29: DNA Master – Gordonia phage - Lamberg

Week of July 6: DNA Master – Gordonia phage - Lamberg

Week of July 13: DNA Master – Gordonia phage – Lamberg; annotation review/submission
DNA Master – Gordonia phage - Dexdert

Week of July 20: DNA Master – Gordonia phage – Dexdert;
Gattaca – view movie; *Gattaca* survey

Week of July 27: DNA Master – Gordonia phage – Dexdert; annotation review/submission

LBRN Summer Program Options
IV - SELU

Dr. Omer Soysal

Abstract and Specific Aims

The main goal of this proposal is to design a prototype Computer Aided Recognition (CAR) system for recognition of lung tumors utilizing the proposed hierarchical deep fusion-learning scheme. The PI will employ the proposed prototype system to obtain extensive preliminary results to be nationally competitive as well as utilize the fund to build a Biomedical Informatics Research Lab in the department of Computer Science.

In this proposal, the PI aims to explore a hierarchical deep fusion network that combines multi-channel and multi-perspective decisions. In addition, the PI will investigate performance of the proposed network integrated with association rule mining based decision making at the fully connected layer.

The idea proposed in this research is innovative such that the model learns how to make fusion of different predictions in contrast to using a voting scheme or simple descriptive approach. The idea of utilizing association rules proposed in this research is also a novel approach in object recognition.

As professional development objectives, the PI aims to build a Biomedical Informatics Research Lab where he will advise undergraduate and graduate students at Southeastern Louisiana University (Southeastern). As an adjunct faculty and a former non-tenure track faculty at Louisiana State University (LSU), the PI has a strong relationship with LSU in terms of academic activities such as advising graduate students, teaching graduate level courses, and collaboration in research. Through this proposal, the PI plans to extend his biomedical research collaboration at LSU and in other local institutions such as Pennington Biomedical Research Center and Mary Bird Perkins Cancer Center. Funding of this proposal will greatly support cross-university and cross-disciplinary research of the PI.

Background and Preliminary Results

Rationale:

The PI Dr. Soysal is an Assistant Professor in the Department of Computer Science at Southeastern Louisiana University as well as an adjunct faculty in the School of Electrical Engineering and Computer Science at Louisiana State University. He has advised 14 graduate level theses and supervised several projects in his former employment at LSU while working as a non-tenure track Research Assistant Professor. He has managed the development of several software applications. In 2009, he proposed a novel hierarchical fusion-learning scheme for recognition of lung tumors; he developed a prototype Computed Aided Recognition application using MATLAB. After appointment as a Research Assistant Professor, he has extended his research on hierarchical learning and lung tumor detection. Currently, his focus is integrating deep learning into his hierarchical learning scheme for concept mining of 3D/4D spatio/temporal data.

The PI has been working on building a Biomedical Informatics Research Lab in the Department of Computer Science. In this lab, he will conduct research in biomedical informatics working closely with undergraduate and graduate students to prepare them for the work force. In this regard, the PI will develop a curriculum that teaches students how to conduct scientific research and translate the outcome of the research into technology.

A student research assistant who has good programming skills will be hired to assist in research, implementation of the proposed algorithms, and preparation of data for experiments. This project will greatly facilitate the PI's ability as an independent researcher who can manage a research project successfully. As a result, the PI will be more competitive nationally.

The PI foresees three barriers to be competitive at the national level:

- 1) Publication in the field of lung cancer recognition: It is important to collect more publications in the field to be a strong candidate at the national level.
- 2) Extensive preliminary results: The PI needs to conduct more experiments using a larger set of data to strengthen his proposed approach for lung cancer detection.
- 3) Independent researcher: The PI has supervised several projects in his previous employment and currently working on his startup project. If funded, this proposal will make the PI more competitive at nationwide funding opportunities.

Significance and Potential Impact: Although lung cancer is the second most commonly diagnosed cancer in both men and women, it is the leading cancer type that causes mortality in both men and women [1]. Lung nodule detection is a very challenging task. The research team in [2] explored the effect of the low-dose CT scans in cancer mortality. Utilizing either low-dose CT or chest radiography, they screened around 53K high lung cancer risk patients three times a year between August 2002 and April 2004. The results of their study show that there is a 20% reduction in mortality of the patients who were screened by low-dose CT scan. Even though the CT scan helps to reduce the mortality rate, the radiologists' decision may differ significantly in identification of the lung nodules from the CT scans. As an example, [3] shared the results of two radiologists' examinations over 25 CT scans; the results show that one of the radiologists detected 20 nodules, whereas the other radiologist detected 63 nodules from the same CT scans.

A Computer Aided Recognition (CAR) system increases the performance of the nodule detection substantially. The study conducted by [4] showed that the CAR system significantly reduced the number of false positives. The research in [5] that studied the effect of a CAR system in detection of small nodules shared the results of six radiologists' examinations over 52 CT scans with/without a CAR system. The results show that the CAR system improves a radiologist's performance considerably. In [6], the performance of the commercial software Lung-CAD VB10A and Siemens AG Healthcare were compared with the performance of two independent readers for detecting the pulmonary nodules in NELSON dataset. The study showed that sensitivity of CAR was 96.7% with a 3.7 FPs/scan and sensitivity of double reader was 78.3% with 0.5 FPs/scan. Therefore, the CAR system with a higher nodule detection rate can be a highly useful tool for radiologists to decrease the number of missed nodules, particularly, the small nodules in their early stages.

Innovation: The outcome of this proposed research will contribute to the increase of our knowledge about the most recent learning approach, deep learning, and its application to lung nodule detection. In object recognition, the idea of learning-based fusion is a new approach in contrast to using a voting scheme. The proposed idea is also novel in applying the approach for the lung tumor recognition. The proposed learning scheme hierarchically improves the decision about the class of an input 3D volume. In addition, association rule mining (ARM) based decision making will be integrated to the proposed hierarchical deep fusion network to improve recognition

performance. The proposed fusion approach and integration of ARM is a novel approach in the field.

Preliminary Results:

The PI's earlier preliminary study [7] shows that the performance of a hierarchical fusion-learning decision engine improves at each successive layer as seen in Figure 1. In this proposal, the PI aims to extend his earlier idea utilizing a deep learning scheme. The PI recently proposed a deep learning method that integrates a well-known data mining (DM) technique, Association Rule Mining, into a deep convolutional neural network (DCNN) at the final fully connected layer [8]. The PI used this DCNN+DM learning machine for prediction of action units (AU), which is a very challenging multi-label classification problem. The preliminary results show that this new approach is promising in recognition of facial expressions. Table 1 summarizes performance of DCNN integrated data mining techniques for action unit classification. In this proposal, the PI will explore the performance of such content representation for lung nodule detection within the deep fusion-learning schema.

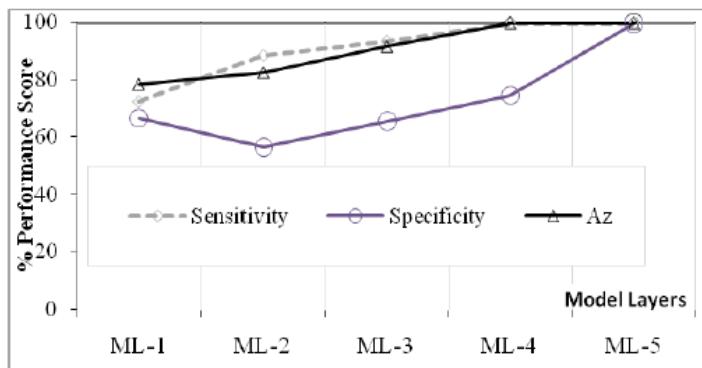


Figure 1 Classification performance through hierarchical fusion-learning

Table 1 Classification performance for AUs

Method	Accuracy	Specificity	Sensitivity
DCNN	63.77%	49.53%	47.53%
DCNN+Causality Max	49.53%	52.09%	45.00%
DCNN+Causality Count	47.53%	49.21%	47.67%
Decision Tree	53.61%	47.31%	58.22%
DCNN+CBA	69.70%	47.36%	64.88%

Dr. Patrick Moyer

Patrick Moyer, Department of Chemistry and Physics, Southeastern Louisiana University

LBRN projects – proposed Summer 2020



VISION and MISSION of the SPM Laboratory: To become an internationally recognized scanning photon and probe imaging laboratory. Our mission is to develop state-of the-art high spatial and temporal resolution imaging modalities and image analysis techniques. Part of our mission is to become a resource for collaborators in the state of Louisiana in order to provide high resolution and novel imaging methods and analysis techniques.

BACKGROUND: Optical microscopy and scanning probe microscopy (SPM) are among the most widely utilized techniques in scientific and engineering studies. We are developing our laboratory at Southeastern Louisiana University (SLU), the Southeastern Photon/Probe Microscopy (SPM) lab, to become an international recognized laboratory for the development of state-of-the-art imaging techniques and image analysis methods.

The SPM lab will consist of two components: (1) an image acquisition and instrumentation facility comprising laser scanning confocal microscopy and scanning probe microscopy, or atomic force microscope (AFM), and (2) an image analysis component. The imaging instrumentation will be home-built by students and capable of single molecule and single photon fluorescence sensitivity, nanometer localization, and dynamic fluorescence measurements with picosecond temporal resolution. LabView will be the interface for instrument control.

In terms of specific applications, we will address nanoscale scientific problems of significance including electronic dynamics of nanoparticles and the physics of plasmons and plasmon-enhanced photon emission from nanoparticles. We will also be a resource for fluorescence imaging for biologist and chemists. Our instrumentation will have high applicability to such fields as molecular imaging and cellular imaging.

Figure 1 shows a schematic diagram of the optical portion of the lab facility that will be developed in the fall of 2020. This is not part of this summers' project but it is provided to give context for the proposed projects this summer.

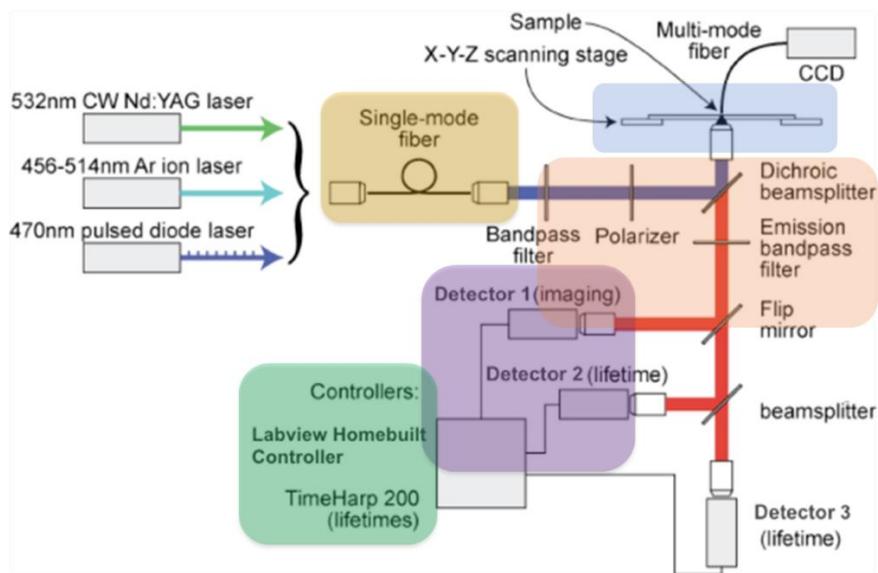


Figure 1. Schematic diagram of the laser scanning confocal microscopy instrumentation.

We have focused our previous nanoscale optical research on studying the electronic properties of quantum dots, myosin V motor protein dynamics, and the study of intracellular viscosity using fluorescence anisotropy.

Quantum dots offer advantages over conventional organic bioimaging markers in the following aspects:

- Narrower emission band providing multiple images to be acquired simultaneously with multiple labels.
- Increased photostability (absence or prolonging of photobleaching).
- Broader excitation spectrum allowing for single source excitation.
- Single molecule sensitivity via increased quantum efficiency (yield).
- Single and two photon emission control allowing for the application of well known non-linear imaging mechanisms and lifetime imaging methods.

Applications include drug delivery, theranostic agents, single molecule sensors of local biological environments/activities, and real-time *in vivo* deep tissue imaging agents. See, for example, the following references of biological applications of quantum dots:

- Nanodots Featuring Efficient FRET for Real-Time Monitoring of Drug Delivery and Two-Photon Imaging, *Adv. Mater.* 25(45) (2013) 6569-6574.
- H. Tada, H. Higuchi, T.M. Wanatabe, N. Ohuchi, In vivo real-time tracking of single quantum dots conjugated with monoclonal anti-HER2 antibody in tumors of mice, *Cancer Res.* 67(3) (2007) 1138-1144.

- Q.L. de Chermont, C. Chaneac, J. Seguin, F. Pelle, S. Maitrejean, J.P. Jolivet, D. Gourier, M. Bessodes, D. Scherman, Nanoprobes with near-infrared persistent luminescence for in vivo imaging, Proc. Natl. Acad. Sci. U.S.A. 104(22) (2007) 9266-9271.

Below are the specific GOALS AND DELIVERABLES for this summer's work (Summer 2020). Project #1 will be projects worked on by Southeastern-sponsored students. Project #2 and project #3 will be LBRN-sponsored student projects.

1. LabView control interface (work supported by Southeastern)

Figure 2 shows a version of the control interface that we (PJ Moyer lab) developed at UNC Charlotte. Two Southeastern undergraduate physics majors have begun work to develop this interface this summer. The interface provides the following features:

- x and y analog output voltages to control the piezoelectric scanner
- tip control and feedback loop for scanned probe microscopy tip control
- multiple simultaneous image acquisition channels (multi-colored fluorescence imaging), sample topography, etc.
- zoom capabilities
- stop and park over a specific point for spectroscopic analysis

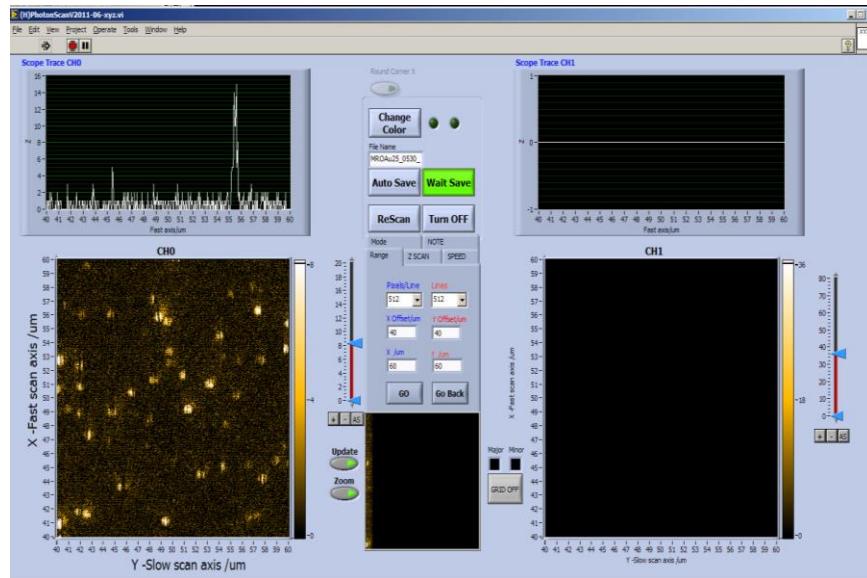


Figure 2. Interface that will be developed by Southeastern students.

DELIVERABLE – A functional interface and data acquisition program for an optical and scanning probe microscope. The interface would include feedback control for the AFM.

2. A stand-alone image deconvolution program (work supported by LBRN)

All imaging methods have a ‘blurring’ mechanism or some fundamental technical limitation to how finely the detail of a sample can be imaged. Optically, this is determined by diffraction effects which are limited by the wavelength of light being used. For SPM, the limitation is the physical size of the tip. Either way, the imaging mechanism has what is known as a point spread function (psf). The final image is a convolution of the sample itself and the point spread function. This deliverable will take the final image and deconvolve that image with the point spread function to ‘recover’ information that may have been lost in the imaging process. This program can be used on any imaging data that an external researcher would send us provided we could estimate the point spread function of their imaging method.

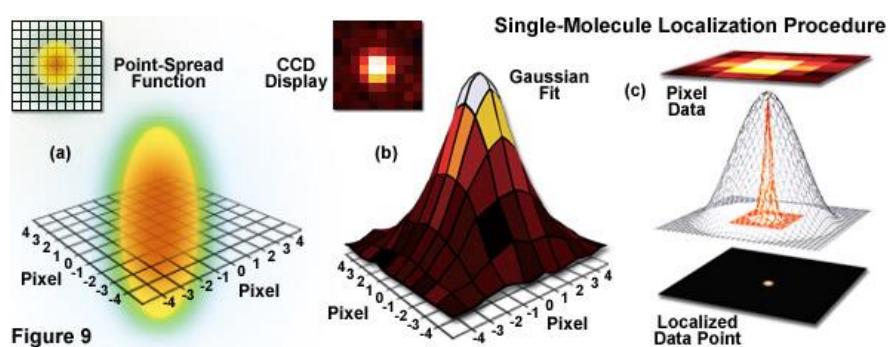


Figure 3. Depiction of the concept of a point spread function (reference <http://zeiss-campus.magnet.fsu.edu/articles/superresolution/introduction.html>)

Figure 4 shows an example of a deconvolution program we wrote and applied in our lab at UNC Charlotte. In this experiment, we used a Bessel function as the PSF. In this project, we will use a Gaussian function. The Gaussian is determined by the image of single molecule or a single quantum dot.

The molecules shown in figure 4 are localized to a position of less than 10 nm using this method. This feature has applicability to future projects via the localization of single biomolecules and tracking them on the nm scale. For example, dynamic processing myosin motor protein molecules along an actin filament can be studied under various physiological conditions. Figure 5 shows an example of previous data we acquired by myosin motor proteins (labeled with green fluorescent markers) on actin filaments (red fluorescence image). This provides an example of how this instrument and deconvolution program can study biological processes on the nanometer scale.

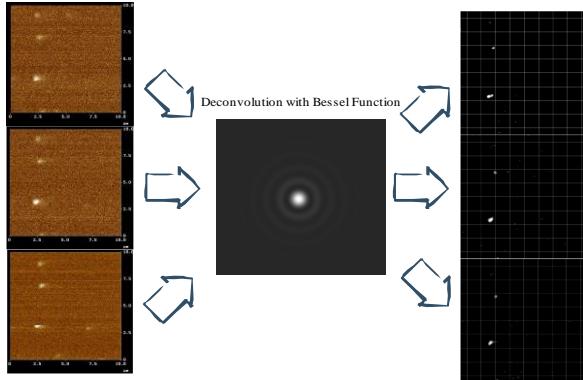


Figure 4. Demonstration of single molecule imaging and deconvolution to yield better than 4 nm localization. Images on left are raw data. Images on right result from deconvolution of raw image with a Bessel function (shown in center) to yield localization of molecules with 4 nm accuracy.

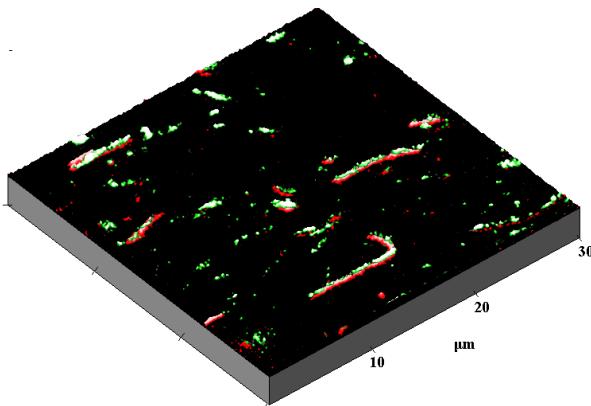


Figure 5. Myosin V motor protein imaging (green) overlapped with actin filaments (red).

DELIVERABLE – Stand-alone image deconvolution program for any image and the corresponding point spread function (PSF) for that particular imaging system.

3. **Two photon imaging using fluorescence lifetime imaging microscopy (FLIM) of single quantum dots**

Two photon imaging of conventional organic fluorescent molecules is a technique used commonly in bio-imaging to effectively narrow the point spread function via the nonlinear nature of the interaction of the excitation laser with the sample. The PSF is narrowed because it is the result of the square of the linear PSF. Multiplying a Gaussian function by itself significantly narrows the instrument PSF and improves the spatial resolution.

This particular project is a novel method proposed by PJM by utilizing the two-photon excitation/emission of quantum dots and the familiar FLIM method. This method has the potential to add high spatial and temporal resolution to bioimaging with the benefits of quantum dots over that of organic fluorescent molecules.

This work is based on work that was previously published by the group of PJM. It takes advantage of two-photon emission from quantum dots in the presence of plasmonic nanostructures. The work does not require plasmon coupling but it makes it much more efficient.

Figure 6 demonstrates two photon excitation/emission of single quantum dots {This work has been published in *NanoLetters*, Sharonda Johnson LeBlanc, Mason McClanahan, Marcus Jones, and Patrick J. Moyer, “Enhancement of multiphoton emission from single CdSe quantum dots coupled to gold films,” *NanoLetters*, **13**, 1662-1669 (2013)}. Figure 6 demonstrates the significant multiphoton emission using the Hanbury-Brown Twiss (HBT) geometry utilized in photon correlation spectroscopy.

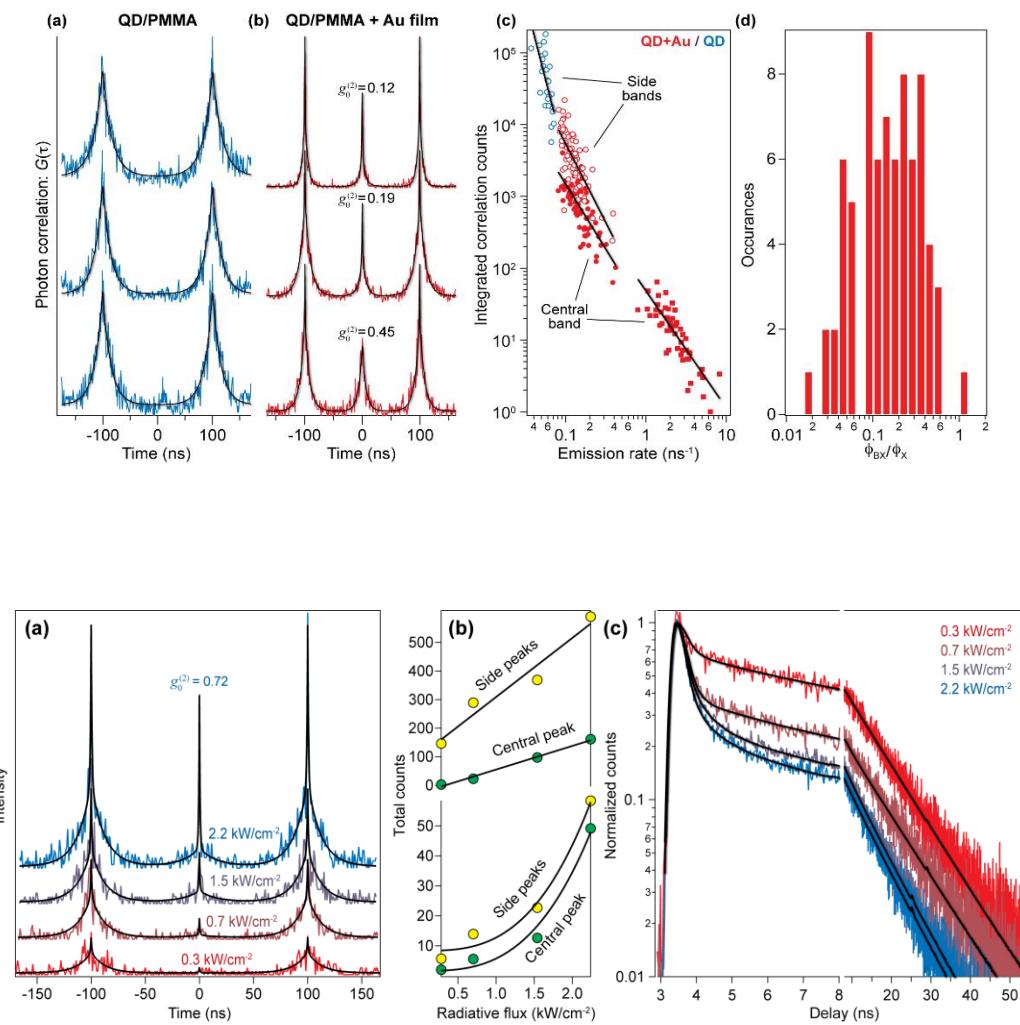


Figure 6. Two photon emission experiments as published in *NanoLetters* by the PI (Moyer). This is one calibration that will be performed during the set up stage for confirming proper instrument performance.

Figure 6 demonstrates (1) multiphoton emission of a single quantum dot (note the central peak of the photon correlation spectra), (2) the fact that we can separate out fast vs. slow lifetime components, and (3) the nonlinear nature of the fast component.

These characteristics provide an excellent opportunity for a novel high spatial and temporal bioimaging mechanism using single quantum dots.

This project will involve the simulation of the type of data presented in figure 6 as a function of xy image position, and a comparison of two images: (1) the spatial mapping of the linear single photon slow fluorescence component and (2) the spatial mapping of the nonlinear two-photon nonlinear fast fluorescence component.

We will also work on a patent study and consider the submission of a provisional patent based on this work.

DELIVERABLE – A full study and comparison of one-photon vs. two-photon imaging of single quantum dots, including applicability to biological systems. Also, we will provide a feasibility study for a patent using this technique.