Title
Exploring the roles of Drosophila *IK2* gene in nervous system function

Author
Olamide Foluso Adefioye*
Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord. Many genetic mutations have been identified in ALS patients, including the mutations found in the gene *TBK*1, it is not yet well understood how these mutations contribute to the pathogenesis of ALS. In this study, we utilized Drosophila transgenic lines with lowered expression of the *IK2* gene, the fly version of the human *TBK1* gene, to investigate the phenotypic and molecular changes of the affected flies. We checked the effect of the toxicity caused by lowered expression of *IK2* on the eye phenotype. We carried out a few behavioral analysis assays which include the climbing and crawling assays on the affected lines. In addition, we conducted out RNA analysis on the affected lines and examined their longevity via a viability experiment. These research data demonstrated that *IK2* gene is very important for Drosophila nervous system function, in the end they will help establish the fly model of ALS to explore the potential mechanism of *TBK1* related ALS.

Title
Analysis of single nucleotide variants on patients with leukemia

Authors
Amanda Bataycan,[1] Jon Mohl,[1,2,3] and Ming-Ying Leung[1,2,3]
[1]Computational Science Program, [2]Department of Mathematical Sciences, and [3]Border Biomedical Research Center, The University of Texas at El Paso, El Paso, TX

Abstract
The overall goal of the study is to identify and analyze the single nucleotide genomic variants found in patients with leukemia and conduct a comparison of the mutational profiles between patients with acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The leading cause of any patient diagnosed with cancer is from either an environmental or hereditary genetic mutation in the DNA. Understanding single nucleotide mutations on a genomic level by examining their locations and identifying their effects on the gene can help provide insights for potential gene therapies. The methodology of this study begins by transforming the data from variant call format files, extracted from the Genomic Data Commons Portal, into a more reader-friendly csv file. Throughout our study, the GRCh38 (Genome Reference Consortium Human Build 38) is used as the reference genome along with 24 refFlat files, one for each chromosome, to assist with genomic region identification. Each individual refFlat file contains the names, locations, as well as start and stop positions of the coding sequences for all the genes found on its corresponding chromosome. Mainly, Python programming is used to organize the data, but R is used for the statistical analysis. Additionally, we will look beyond the mutational data to include patient demographical and clinical details such as gender, age, race, and the leukemia subtype using the French-American-British classification system for interpreting their correlations with patients' overall survival.

Title
Genomics-assisted breeding in agriculture

Authors
Ratna Jyothi Bolem[1,2] and Utkarsh Singh[1]
[1]School of Agriculture Sciences, Baba Farid Institute of Technology, India
[2]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
The global population is expected to reach nine billion before 2050, which means that there is a need for more food to satisfy this increasing demand. The development of modern agriculture has thus been a major driver in the growth of biotechnology as a means to increase crop yield and/or quality. However, traditional breeding programs require many crossing and selection trials and a great number of resources and are thus often not feasible in developing countries. Therefore, genomics-assisted breeding (GAB) represents a strategy to overcome these limitations. With the development of sequencing technologies, the amount of genomic information that can be gained from a single crop has reached a level that is now possible to be harnessed in agriculture. The integration of genomics in the development of new plant varieties can be traced back to the seventies, but only a few successful examples are reported in the literature today. In this presentation, the potential of GAB in the development of improved crop varieties is described. The challenges that need to be overcome to achieve the full potential of GAB are discussed, and possible approaches to overcome these challenges are outlined.

Title
Review Presentation: "Use of metatranscriptomics in microbiome research"

Presenter
Robert Diaz, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
This poster is a review of the paper "Use of Metatranscriptomics in Microbiome Research" by Bashiardes *et al.* (2016) published in *Bioinformatics and Biology Insights.* The abstract, as originally published by the authors, is as follows:

"The human intestinal microbiome is a microbial ecosystem that expresses as many as 100 times more genes than the human host, thereby constituting an important component of the human *holobiome*, which contributes to multiple health and disease processes. As most commensal species are difficult or impossible to culture, genomic characterization of microbiome composition and function, under various environmental conditions, comprises a central tool in understanding its roles in health and disease. The first decade of microbiome research was mainly characterized by usage of DNA sequencing based 16S rDNA and shotgun metagenome sequencing, allowing for the elucidation of microbial composition and genome structure. Technological advances in RNA-seq have recently provided us with an ability to gain insight into the genes that are actively expressed in complex bacterial communities, enabling the elucidation of the functional changes that dictate the microbiome functions at given contexts, its interactions with the host, and functional alterations that accompany the conversion of a *healthy* microbiome toward a disease driving configuration. Here, we highlight some of the key metatranscriptomics strategies that are implemented to determine microbiota gene expression and its regulation and discuss the advantages and potential challenges associated with these approaches."

Title

Screening of maize and maize products in Cape Coast markets for genetically modified products

Authors

Lucky K. Gbeda[1,2] and Emmanuel P. Otwe[1]

[1]Department of Molecular Biology and Biotechnology, University of Cape Coast, Ghana

[2]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract

Genetically modified organisms (GMOs) are organisms generated using genetic engineering techniques. GMOs are produced purposefully to increase agricultural yield, reduce pesticides and herbicides use, improve nutritional quality of foods, and prevent disease through edible vaccines. However, GMOs come with biosafety concerns and perceived risks. Hence, it is necessary to screen for genetic modification in major food crops like maize and maize products. The research identified several improved maize varieties in Cape Coast markets, Ghana. Flinders Technology Associates (FTA) paper and cetyltrimethylammonium bromide (CTAB) methods were used to extract DNA from the improved maize varieties ("abontem," "honampa," "obatampa," "omankwa," "sanzal-sima," and "panner53") for the purpose of transgene amplification using polymerase chain reaction (PCR). PCR was used to amplify the transgenic regions for the cauliflower mosaic virus (CaMV) 35S promoter and nopaline synthase (NOS) terminator of expected band size of 199 bp and 127 bp respectively. The PCR method of GMO detection provided significant evidence that no trace of genetic modification was found in maize varieties present in the Cape Coast markets, Ghana.

<u>Title</u>
Review Presentation: "An informatics approach to distinguish RNA modifications in nanopore direct RNA sequencing"

<u>Presenter</u>
Luis Antonio Gracia Mazuca, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

<u>Abstract</u>
This poster is a review of the paper "An informatics approach to distinguish RNA modifications in nanopore direct RNA sequencing" by Ramasamy *et al*. (2022) published in *Genomics*. The abstract, as originally published by the authors, is as follows:

"Modifications in RNA can influence their structure, function, and stability and play essential roles in gene expression and regulation. Methods to detect RNA modifications rely on biophysical techniques such as chromatography or mass spectrometry, which are low throughput, or on high throughput short-read sequencing techniques based on selectively reactive chemical probes. Recent studies have utilized nanopore-based fourth-generation sequencing methods to detect modifications by directly sequencing RNA in its native state. However, these approaches are based on modification-associated mismatch errors that are liable to be confounded by SNPs. Also, there is a need to generate matched knockout controls for reference, which is laborious. In this work, we introduce an internal comparison strategy termed "IndoC," where features such as 'trace' and 'current signal intensity' of potentially modified sites are compared to similar sequence contexts on the same RNA molecule within the sample, alleviating the need for matched knockout controls. We first show that in an IVT model, 'trace' is able to distinguish between artificially generated SNPs and true pseudouridine (Ψ) modifications, both of which display highly similar mismatch profiles. We then apply IndoC on yeast and human ribosomal RNA to demonstrate that previously reported Ψ sites show marked changes in their trace and signal intensity profiles compared with their unmodified counterparts in the same dataset. Finally, we perform direct RNA sequencing of RNA containing Ψ intact with a chemical probe adduct (N-cyclohexyl-N′ -β-(4-methylmorpholinium) ethylcarbodiimide [CMC]) and show that CMC reactivity also induces changes in trace and signal intensity distributions in a Ψ specific manner, allowing their separation from high mismatch sites that display SNP-like behavior"

Title
Review Presentation: "ISOGlyP: de novo prediction of isoform-specific mucin-type O-glycosylation"

Presenter
Luisa Gracia Mazuca, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
This poster is a review of the paper "ISOGlyP: de novo prediction of isoform-specific mucin-type O-glycosylation" by Mohl *et al.* (2021) published in *Glycobiology*. The abstract, as originally published by the authors, is as follows:

"Mucin-type O-glycosylation is one of the most common posttranslational modifications of proteins. The abnormal expression of various polypeptide GalNAc-transferases (GalNAc-Ts) which initiate and define sites of O-glycosylation are linked to many cancers and other diseases. Current O-glycosyation prediction programs utilize O-glycoproteomics data obtained without regard to the transferase isoform (s) responsible for the glycosylation. With 20 different GalNAc-Ts in humans, having an ability to predict and interpret O-glycosylation sites in terms of specific GalNAc-T isoforms is invaluable.

To fill this gap, ISOGlyP (Isoform-Specific O-Glycosylation Prediction) has been developed. Using position-specific enhancement values generated based on GalNAc-T isoform-specific amino acid preferences, ISOGlyP predicts the propensity that a site would be glycosylated by a specific transferase. ISOGlyP gave an overall prediction accuracy of 70% against in vivo data, which is comparable to that of the NetOGlyc4.0 predictor. Additionally, ISOGlyP can identify the known effects of long- and short-range prior glycosylation and can generate potential peptide sequences selectively glycosylated by specific isoforms.

ISOGlyP is freely available for use at ISOGlyP.utep.edu. The code is also available on GitHub (https://github.com/jonmohl/ISOGlyP)."

<u>Title</u>
Progression-free survival of patients with prostate cancer under radiation therapy

<u>Authors</u>
Kelvin Ofori-Minta[1] and Ming-Ying Leung[1,2]
[1]Data Science Program, Department of Mathematical Sciences and [2]Border Biomedical Research Center, The University of Texas at El Paso, El Paso, TX.

<u>Abstract</u>
Progression free survival (PFS), defined as the time from random assignment in a clinical trial to disease progression or death by any other cause, has seen increased interest in its adoption as a substitute for the overall survival (OS) endpoint, partly due to the ability of PFS to monitor situations where OS improvements by follow-up treatments were marginal. These numerous upvotes for PFS endpoint argue that a survival benefit is experienced when tumor progression is stunted, hence an indication of tumor stabilization. Using the clinical data freely accessible via cBioPortal for cancer genomics, we study the effect of radiation therapy, a treatment using high doses of radiation to damage cancer cells, on the growth of tumor in patients at different cancer stages. Employing the non-parametric Kaplan-Meier estimator to analyze the survival trends of the data cohorts, we found out several indicative effects of some medical variables. Considering cohorts with different cancer staging, we realized radiation therapy has a reduced but insignificant effect on PFS. Moreover, patients with reported neoplasm experienced significant reduction in PFS than those without, and neoplasm patients experienced an insignificant PFS advantage over patients without neoplasm when exposed to radiation therapy. This analysis was collectively done for patients at all cancer stages but more separate studies on the PFS of patients at different stages will be conducted. We further seek to examine the validity of the Cox proportional-hazards assumption for our set of PFS predictor variables.

Title
Review Presentation: "Cancer omics in Africa: Present and prospects"

Presenter
Francis Owusu-Dampare, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
This poster is a review of the paper "Cancer Omics in Africa: Present and Prospects" by El Jaddaoui *et al*. (2020) published in *Frontiers in Oncology*. The abstract, as originally published by the authors, is as follows:

"During the last century, cancer biology has been arguably one of the most investigated research fields. To gain deeper insight into cancer mechanisms, scientists have been attempting to integrate multi omics data in cancer research. Cancer genomics, transcriptomics, metabolomics, proteomics, and metagenomics are the main multi omics strategies used currently in the diagnosis, prognosis, treatment, and biomarker discovery in cancer. In this review, we describe the use of different multi omics strategies in cancer research in the African continent and discuss the main challenges facing the implementation of these approaches in African countries such as the lack of training programs in bioinformatics in general and omics strategies in particular and suggest paths to address deficiencies. As a way forward, we advocate for the establishment of an "African Cancer Genomics Consortium" to promote intracontinental collaborative projects and enhance engagement in research activities that address indigenous aspects for cancer precision medicine."

Title
Review Presentation: "GMSECT: Genome-wide massive sequence exhaustive comparison tool for structural and copy number variations"

Presenter
Salvador Alexis Rodarte Villa, Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Abstract
This poster is a review of the paper "GMSECT: Genome-Wide Massive Sequence Exhaustive Comparison Tool for Structural and Copy Number Variations" by Singh (2021) published in the *Journal of Proteomics & Bioinformatics*. The abstract, as originally published by the authors is as follows:

"GMSECT is a parallel robust 'Application Interface' that efficiently handles the large genomic sequences for rapid and efficient processing. It is a 'message passing interface' based parallel computing 'Tool' that can be operated on a cluster for 'Massive Sequences Exhaustive Comparison', to identify matches such as the structural variants. The GMSECT algorithm can be implemented using other parallel application programming interfaces as well such as Posix-threads or can even be implemented in a serial submission fashion. There is complete flexibility to the choice of comparison tool that can be deployed and with the optional parameters as of the choice of comparison tools to suit the speed, sensitivity and specificity of pair wise alignment. The algorithm is simple and robust, and can be applied to compare multiple genomes, chromosomes or large sequences, of different individuals for personalized genome comparison and works good for homologous as well as distant species. The tool can even be applied to smaller genomes like the microbial genome such as the *Escherichia coli* or algae such as *Chlamydomonas reihardtti* or yeast Saccharomyces cerevisiae to quickly conduct comparisons, and thus finds its application to the pharmaceuticals and microbial product-based firms for research and development. The application interface can efficiently and rapidly compare massive sequences to detect for the presence of numerous types of DNA variation existing in the genome ranging from Single Nucleotide Polymorphism (SNPs) to larger structural alterations, such as Copy-Number Variants (CNVs) and inversions. The new algorithm has been tested for comparing the chromosome 21 of Celera's R27c compilation with all the 48 chromosomes of Celera's R27c compilation and with all the 48 chromosomes of the human Build 35 reference sequence, which took just 2 Hours and 10 minutes using the pair wise BLAST algorithm choice and with 110 processors each with 2.2 GHz capacity and 2 GB memory. GMSECT facilitates rapid scanning and interpretation in personalized sequencing project. The application interface with the above resources and alignment choice is estimated to do exhaustive comparison of the human genome with itself in just 2.35 days. An exhaustive comparison of an individual's genome with a reference genome would comprise of a two 'self-genome' comparison and a 'non-self-genome' comparison which is estimated to take about 9.4 days with the above resources. With the advent of personalized genome sequencing project, it would be desirable to compare 100s of individual's genome with a reference genome. This would involve a 'non-self-genome' and a 'self-genome' comparison for each genome and would take around 7 days for each individual's genome using GMSECT and the above-mentioned resources."

Title
Characterization of FOXA1 gene in breast, liver, and prostate cancer

Authors
Dayo Shittu,[1] Enrique Ramos,[2] and Shrikanth Gadad[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX
[2]Department of Molecular and Translational Medicine, Texas Tech University Health Sciences Center, El Paso, TX

Abstract
FOXA1 is a member of FOX gene family of transcription factors which have been found to be imperative in multiple stages of mammalian life, from early development to metabolism and homeostasis in the adult. It is expressed in multiple tissues and can bind to the promoters of several genes related with cell signaling regulation and the cell cycle. The goal of the study is to do a comprehensive in-silico characterization of FOXA1 gene to help understand its roles in different cancers. Specifically, we aim to identify which of five FOXA1 transcript variants are more relevant for which cancer types, and to understand how the gene behaves in different cell lines and different cancer types. This is done by processing RNA-seq data from the Genotype-Tissue Expression (GTEx) project, The Cancer Genome Atlas (TCGA), the Ensembl Genome Database, and the UCSC genome browser. As FOXA1 is a pioneer factor that determines the expression of other genes and plays a role in differential regulation of gene expression among different isoforms of FOXA1, we will compare expression levels of FOXA1 in normal samples to those in cancer samples. The different transcripts of FOXA1 gene and how they are regulated in different cancer types will be investigated. Using a set of bioinformatics tools in Bioconductor (e.g., recount3, biomaRt), as well as custom scripts in the R programming language, we will determine how different alternate splicing occurs, how the exons align to form the different FOXA1 transcripts, and which transcripts are involved in a specific cancer types. These will help us gain insights into cancer progression in breast, liver, and prostate tissues where FOXA1 is known to be highly expressed.

Title
Safety and efficacy of COVID-19 and influenza vaccines

Authors
Diana Vielma[1] and Mingtao Zeng[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX
[2]Department of Molecular and Translational Medicine, Texas Tech University Health Sciences Center, El Paso, TX

Abstract
The objective of this study is to evaluate the safety and efficacy of the COVID-19 and influenza vaccines. Investigating these vaccines ability to prevent disease is important in the protection of the population because it leads to a reduction in hospitalizations, deaths, birth defects and harmful reactions. Safety is a priority throughout the vaccine approval lifecycle and is continuously monitored post-approval. The efficacy of vaccines evaluates its effectiveness to prevent disease and is measured by comparing a vaccinated group against an unvaccinated group using a ratio calculated with the percentages of people who contracted the disease. Vaccine Adverse Event Reporting System (VAERS) is a reporting system that allows individuals who have received any type of vaccines to report side effects experienced ranging from minimal to fatal. The VAERS database contains a collection of CSV files organized by calendar years starting at 1990. For this analysis, the CSV files from 2021 to 2022 will be incorporated to analyze efficacy of the COVID-19 and influenza vaccines.

Title
Computational genomics and computer-aided drug design for organic compounds against mutant p53 cancer gene

Authors
Kanishk Yadav,[1,2*] Meena Bohra,[2] and Vinod Kumar Gupta[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX
[2]Rapture Biotechnology, Inc., Noida, India

Abstract
The p53 gene is one of the essential genes in our body that prevents cancer from happening. However, when there is a mutation in that gene due to reasons such as ionizing radiation, chemical exposure (like methyl orange), etc., it can lead to p53 gene becoming cancerous. Our goal was to study organic compounds (like drugs) in experimental and investigational stages that would help us in stopping such cancer by binding to the mutant p53 and inactivating or inhibiting its cancerous activity. For this we used many different types of bioinformatics software such as BLAST, Clustal Omega, ORF, Protein DataBank, UniProt, STRING, KEGG Pathway, DrugBank, PubChem, etc. Some of our major jobs were done using PyRx and Biovia Discovery Studio. There were interest findings in the study such as all the organic compounds had different binding energy and two of the promising compounds for deactivating cancerous mutant p53 gene were "AZD-1775" and "Gantespib."

*Mr. Kanishk Yadav performed this research project for his bachelor's dissertation while pursuing bachelor's degree at Jaipur National University, India, under the supervision by Dr. Vinod Kumar Gupta with the assistance of his research associate Ms. Meena Bohra for completing and working towards the dissertation.