Cancer Survival Rate Analysis: Cytoskeletal Protein Stop and Frame Shift (fs) Codons vs. Missense Codons

CBIO001

Computational Biology and Bioinformatics

Jessie Orrin Strickland

Wildwood High School, Wildwood, FL

Cancer is something we hear in our everyday lives, it takes friends and loved ones and there seems like there is no help for them, especially when the treatment does not work. Through this project, data was gathered to determine if there should be different cancer treatments for different types of cancer mutations. The different types of mutations inside specific cancer cells were compared. Specific genes were used to compare the mutations; the four genes used were Mucin 4, Mucin 16, Fat 4, and Collagen type XI alpha 1. Five cancers were selected to use in the experiment, the five cancers were Lung Adenocarcinoma, Breast Invasive Carcinoma, Skin cutaneous melanoma, Colorectal Adenocarcinoma, and Prostate Adenocarcinoma. All of this experiment was done on Cbioportal.org and Excel. By comparing the two different types of mutations inside a cancer cell using Cbioportal.org using graphs and p-values, it showed that the effects of stop and frameshift mutations are much worse than the effects of missense mutations on the cancers survival rate. Several factors cause these results. The main factor is that frameshift and stop mutations leave the cytoskeleton in worse condition than missense does. The experiment showed that in the Breast Invasive Carcinoma has a statistical difference in the different types of mutations, meaning that it supported my hypothesis. The other cancers neither proved nor disproved my hypothesis.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Utilization of Artificial Intelligence Assisted Brain-Computer Interface to Allow Patients with Motor Impairments or Paralysis to Regain a Range of Mobility

Cindy Mo, Freddie Hao Liang

President Theodore Roosevelt High School, Honolulu, HI

CBIO002T

Computational Biology and Bioinformatics

The purpose of the project was to create a brain-computer interface headset prototype capable of allowing patients with motor impairments and paralysis to regain a range of mobility utilizing the Applied Program Interface, Keras artificial intelligence. According to the 2017 Disability Statistics Compendium, 6.7% of the American population were diagnosed with ambulatory disabilities. These motor disabilities prevent patients from performing daily activities and living a normal, independent life. Electroencephalography (EEG) electrodes were programmed to collect and record the electrical activity of the brain to a Raspberry Pi which transmits the data to be displayed on OpenBCI Graphical User Interface (GUI) which shows the data in graphs and sequences. Softwares were used to display the data collected from each sensor and converted the EEG data into Fast Fourier Transform (FFT) data which were used to train a 1D convolutional neural network. The AI processed data and also determined which thoughts were correlated with certain sequences. Python was used to program the prototype of a Spinal Cord Stimulator (SCS) to output voltages at different locations to simulate how Spinal Cord Stimulation would work inside a patient with ambulatory disabilities. EEG data were collected and commands were issued out to the SCS to stimulate a certain part of a patient's body which then allows a patient to move; therefore, we were able to conclude that it is possible to utilize EEG to control the bodies of disabled patients with motor impairments.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

[✓ human participants] potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:07 PM*

Correlation between Protein Type and Fatality Rates in Zaire Ebolavirus

CBIO003

Computational Biology and Bioinformatics

Ellen Rose Jones

Haines City IB East, Lake Wales, FL

The student researcher tested for correlation between protein type and fatality rate in Zaire ebolavirus, hypothesizing that there is a significant difference in the fatality rates due to protein types, with the null hypothesis being no significant difference. Records were pulled from the Virus and Pathogen Resource Database for Zaire ebolavirus and sorted by protein type and by UniProtKB Accession. Fatality rates were gathered from the WHO and the CDC and then matched based on the year of collection. Protein types with the highest fatality rates were polymerase (58%), glycoprotein 1, 2 (53%), and small secreted glycoprotein (53%). An Analysis of Variance (ANOVA) was performed upon the data. The student researcher determined that the protein types and associated fatality rates supported the alternative hypothesis that presence of protein type does impact fatality rate in the instance of Zaire ebolavirus. This was concluded because the p-value, 3. 05 x 104_71, is less than the alpha, 0. 05. The f input, 24. 8, is also greater than the f critical input, 1. 63. Due to the difference in the p-value and alpha, the null hypothesis fails to be accepted and the alternative hypothesis is supported beyond the realm of chance.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):		YES	✘	NO
3. This project is a continuation of previous research (Form 7):		YES	✘	NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):		YES	✘	NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	✘	YES		NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	✘	YES		NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

MedBrain: A Novel Deep Learning Tool for Holistic Patient Data Interpretation and Clinical Event Prediction

CBIO004

Computational Biology and Bioinformatics

Suraj Anand

Palos Verdes Peninsula High School, Rolling Hills Estates, CA

Approximately 55,000 patients are treated daily in ICUs within the U.S., most of whom are extremely sick and require rapid clinical decision-making. However, copious patient data engenders information overload, leading to slower care and poorer outcomes. Current computer-aided diagnostic solutions for decision-making support are used minimally, either unable to capture complex patient data or lacking interpretability. MedBrain is an interpretable deep learning python framework created to improve clinician decision-making by predicting significant clinical events and identifying relevant patient variables based on charts. Time-series patient data from Beth Israel Deaconess Hospital (2001-2012) was manually preprocessed via anomaly detection, imputation, and quality control. A Skip-gram and Bidirectional Transformer were created to encode textual diagnoses and notes into numeric embeddings. With these data, multiple Long Short-Term Memory Recurrent Networks with Attention were constructed to predict future likelihood of the critical short-term events of mortality, sepsis, respiratory failure, acute kidney injury, vancomycin admin, and myocardial infarction (MI). These achieve an Area under the Receiver Operating Characteristic (AUROC) of 0.94, 0.88, 0.86, 0.85, 0.83, and 0.80, respectively. Attention maps identify relevant patient variables, like troponin as a predictor of MI. Moreover, spatial analysis of embeddings enables physicians to discern possible overlooked diagnoses associated with the patient's history. This innovative prediction model harnesses the power of deep learning while providing unprecedented transparency via attention mechanisms, distilling vast patient data to show relevant variables. MedBrain could reduce cost and improve patient outcomes.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:08 PM*

Analyzing the Mechanics of Pollutant Degradation

CBIO005

Computational Biology and Bioinformatics

Melannie Paulette Nimocks

College Park High School, The Woodlands, TX

As pollution levels rise, the relationship between pollutant degrading proteins across various organisms is vital to create a synthetic microbe that mitigates substance toxicity. To understand this relationship, the programming language R was used to write a program that analyzed trends among proteins involved in toxic substance breakdown. Due to similarity of function across the proteins of various organisms, it was hypothesized that a strong relationship exists between the proteins. In order to analyze how pollutant consumption is related across organisms, a variety of methods from the Bio3D package under the computer language R were employed on eight specific proteins. Once these methods were called, it was found that sets 3, 4, 11, 12, 13, 14, and 17 contained cores, or areas of highly similar protein make-up. Sets 3, 11, and 12 exhibited particularly strong similarities. The normal mode analysis procedure quantified the structural variations between each protein of each set and organized it in a visual plot. These results support the patterns from the core analysis and indicate a level of similarity in both amino acid sequence and structure among various sets of proteins, specifically, between proteins that degrade plastics, toxic chemicals and heavy metals, and hydrocarbons. However, as it is just a visual depiction, further analysis is being explored. These similarities signal the potential to create a synthetic protein that could attack multiple pollutants, specifically, hydrocarbons and plastics or heavy metals, herbicides, and hydrocarbons at once. As the proteins have little variance in structure, a small genetic change that adds the functionality of the other effective proteins is plausible.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:08 PM*

Persistent Homology and Analysis of Histological Data

CBIO006T

Computational Biology and Bioinformatics

Georgii Kadantsev, Aleksandr Sinitsyn

School 564, St. Petersburg, Russian Federation

Colorectal cancer is the second highest cause of cancer occurrence and death in men and women in the United States combined. A specialist usually confirms their diagnosis by a careful microscopical examination of a tissue sample. As a thorough analysis like this can be quite difficult when time is of the essence, close attention has been given to developments towards a computer-based diagnosis system in recent years. In a computer-aided examination of a tissue sample its digital copy is usually divided into many small images called patches. Each patch is then analysed individually. In this project we study these histology images using methods from topological data analysis (TDA). TDA is a modern approach to data analysis which aims to extract certain topological features from data. The primary characteristic of a patch for us is persistent entropy of an image, which is extracted from its 0-th persistent homology. It can be viewed as a certain numerical measure for the chaos present in an image described in a language of TDA. The goal of this project is to show that the concept of persistent entropy can be helpful in diagnosis of colorectal cancer. We have developed a fast and original algorithm for computing 0-th persistent homology and persistent entropy. We have analysed a big dataset of healthy tissue images and tumor images and observed a significant difference between the entropy of two patch classes. These findings can become pivotal in a new computer-based system for colorectal cancer diagnosis.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　potentially hazardous biological agents

vertebrate animals　microorganisms　rDNA　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES ✘ NO

3. This project is a continuation of previous research (Form 7): YES ✘ NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES ✘ NO

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: ✘ YES NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ✘ YES NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:08 PM*

Exploring the Mechanisms of Major Depression and Antidepressant Response Using Gene and miRNA Expression

CBIO007

Computational Biology and Bioinformatics

Ethan Joseph Dunsworth

Wentzville Holt High School, Wentzville, MO

Major depressive disorder (MDD) is a debilitating illness, and antidepressant treatment often yields inadequate response rates. It is not currently understood why antidepressant response rates remain so low, partially due to the incomplete knowledge of how MDD itself functions in the brain. This study attempts to further elucidate the biological mechanisms that define MDD and antidepressant response. To do this, the miRNA expression of MDD patients who did and did not respond to treatment were compared to uncover patterns in differential expression. Analysis showed that miRNAs 146a-5p, 146b-5p, miR-24-3p, and 425-3p all are significantly down-regulated in responders following treatment, have stronger down-regulation in responders, and are all likely involved in biological processes which reduce depressive symptoms—findings corroborated by one of the studies whose data this project uses. Cross-referencing multiple databases that predict the target genes of miRNA suggests that antidepressant response correlates with a disruption in axon guidance, MAPK signalling, and TCR-signalling pathways. To explore the presence of these miRNA in external validation sets, the gene expression of postmortem brain tissues were analyzed to find differential gene expression between control subjects and subjects diagnosed with MDD. The results of the genetic analysis supported the miRNA findings and suggest that the biological processes of MDD are sexually dimorphic—again consistent with the findings of the original study's authors.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:09 PM*

Predicting Epileptic Seizures Using Discrete Wavelet Transform and Machine Learning

CBIO008

Computational Biology and Bioinformatics

Daye Kwon

Arkansas School for Mathematics, Sciences and the Arts, Hot Springs, AR

Epilepsy is one of the most common neurological disorders worldwide, and the unpredictable nature of epileptic seizures makes life challenging for many patients. The objective of this study was to develop generalized and patient-specific machine learning models that are able to predict seizures by identifying the preictal state. A five level discrete wavelet transform with the Daubechies 4 wavelet was applied to preictal and interictal EEG segments. Energy, Shannon entropy, and absolute mean were extracted from each subband as features. Variance Decomposition Proportions were analyzed to detect multicollinearity and remove redundant features. Four machine learning algorithms- Decision Tree, K-Nearest Neighbors, Logistic Regression, and Support Vector Machine- were trained using 5-fold cross validation and tested on the corresponding validation sets. The decision tree algorithm had the best performance as a generalized predictor, achieving a specificity of 95.2%, a sensitivity of 84.8% and an accuracy of 88.4%. The best-performing patient-specific models all achieved a sensitivity of at least 75% and specificity of at least 80%. In a clinical context, the algorithm would be able to detect the onset of a preictal state within 25 seconds with a 99.97% probability. If implemented in wearable devices, such machine learning based seizure predictors can greatly improve the quality of life for people with epilepsy. Further research in this field includes exploring different types of features, applying different machine learning algorithms, and using data from a larger number of patients.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants | potentially hazardous biological agents

vertebrate animals | microorganisms | rDNA | tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:09 PM*

A Biologically-inspired, Biomarker-driven, Rapid Early Warning System for Epileptic Onset Prediction and Seizure Detection Using Machine Learning

Pratik Vangal

Sunset High School, Portland, OR

CBIO009

Computational Biology and Bioinformatics

Epilepsy is a chronic brain disorder impacting over 65 million people globally. Each year, >100000 patients die from Sudden Unexpected Death in Epilepsy (SUDEP), many from fatal falls. A reliable seizure forecasting, and early warning system can help patients stay safe. This work presents real-time algorithmic methods for automated seizure detection by performing rapid feature extraction using CHB-MIT's scalp electroencephalogram (EEG) epilepsy database. A new time-frequency domain Discrete Wavelet Transform analysis enables near 100% seizure detection accuracy. Predicting seizures before they occur is a challenging research problem! By analyzing over 200 hours of chronic epileptic physiological data, three unique biomarker pre-seizure patterns were identified and utilized to develop a novel machine learning epilepsy prediction framework. In 16 of 23 patients, EEG data analysis shows distinct bursts of high-frequency oscillations (60-100 Hz range) preceding a seizure. A second biomarker was identified by examining fluctuations in electrocardiogram (ECG) data, called heart rate variability. Stress can precipitate seizures. By periodically monitoring variations in Cortisol, the stress hormone in the human body - elevated cortisol levels can be correlated to seizure onsets. The predictive feature vectors extracted from all three biomarkers are used to train supervised machine learning (ML) classifiers. The final trained ML model can successfully predict seizures 1-22 minutes prior to clinical onset with 91% classification accuracy. All three proposed biomarkers allow for non-invasive patient monitoring. A low-cost (<US$10) open-source electronics platform shows promise for a wearable "epilepsy alert device" to improve emergency response times and help save lives worldwide.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES ☒ NO

3. This project is a continuation of previous research (Form 7): ☒ YES NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES ☒ NO

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: ☒ YES NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ☒ YES NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:09 PM*

Identifying Tablets Using Neural Networks

CBIO010

Computational Biology and Bioinformatics

Isha Narang

Ardrey Kell High School, Charlotte, NC

According to the FDA, approximately 1.3 million people are injured due to medicine errors annually in the United States. A large percentage of these people are the elderly and people with multiple medical conditions. So, the purpose of my project was to utilize data analytics and machine learning methods to provide a simple way to identify tablets, thus reducing medicine errors.

First, I took pictures of different tablets and applied various filters to them in WEKA. Then, I ran the Decision Trees algorithm on features generated by each filter and selected the Auto Color Correlogram filter because its features resulted in the highest classification accuracy of 88.75%. The accuracy of this filter with Neural Networks (NN) was 97.5%. With this evidence of NN being a good classifier, I ran the algorithm available at Teachable Machine on my dataset to generate an NN model using TensorFlow Lite. I imported this model and embedded it into an Android app, which I named 'Tablet Identifier'. I downloaded this app onto a virtual phone, and connected a webcam to my desktop in order to test the app. It correctly identified tablets with 99% accuracy. When published, more medications can be added for numerous medical conditions. Thus, anyone with a mobile phone can download the app, and identify any tablet when their phone's camera is pointed towards it.

As a result, people will be able to use this user friendly app at home before taking their medication, which will lower the number of medicine errors.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants	potentially hazardous biological agents		
vertebrate animals	microorganisms	rDNA	tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	YES	[✓ NO]
3. This project is a continuation of previous research (Form 7):	YES	[✓ NO]
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES	[✓ NO]
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	[✓ YES]	NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	[✓ YES]	NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:10 PM*

Understanding Schizophrenia: Changing the Game with the Oral Microbiome

CBIO011T

Computational Biology and Bioinformatics

Madeline Kate Hace, Eliana Mae Zubert

Chanhassen High School, Chanhassen, MN

This experiment sought to determine if individuals with schizophrenia have observable differences in the genetics of their oral microbiomes compared to healthy individuals. This was done by investigating the significance of oral microbiota phyla and metabolic pathways. The research was performed using genetically sequenced data samples of the human oral microbiomes of six schizophrenics and six healthy individuals from the Integrated Microbial Genomes and Microbiomes (IMG/M) database. To ascertain if distinctions were present between the two oral microbiomes, the data was compared using principal coordinates analysis (PCoA) models, based on phyla composition and protein pathways. Differentiations in the phylogenetic distribution of the samples metagenomes were shown graphically. Additionally, variations between metabolic pathways were identified using Kyoto Encyclopedia of Genes and Genomes (KEGG) models. The PCoA models illustrated a clear division between the samples using spatial separation. The phylogenetic distribution of metagenomes models showed healthy samples had increased amounts of bacteroidetes bacteria, while schizophrenic samples had more firmicutes bacteria. Dissimilarities in phyla between the two oral microbiomes illustrated schizophrenia's affect on phylum diversity. Through analysis of metabolic pathways, schizophrenic samples showed higher concentrations of the metabolic transporters urea and glutamine/glutamate. As a result of the differences between these various analyses, there is a prevalent and observable difference between the oral microbiomes of schizophrenic and healthy individuals. In the future, this research could be developed further regarding how the oral microbiome can be utilized during the diagnosis and treatment of Schizophrenia.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	☒ YES	NO
3. This project is a continuation of previous research (Form 7):	YES	☒ NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES	☒ NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	☒ YES	NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	☒ YES	NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:10 PM*

Auditory Attention Decoding Approach to Cocktail Party Problem Using Deep Learning

CBIO012T

Computational Biology and Bioinformatics

Govardhan Thirumurthy Poondi, John Rho

Plano West Senior High School, Plano, TX

[Problem] For individuals with hearing loss, distinguishing between various voices in noisy environments is difficult. Current hearing aids are unable to identify and attenuate background voices due to software limitations and a single-microphone design.

[Approach] The current audiological approach to this issue (the cocktail party problem) is auditory attention decoding, a process that harnesses the natural ability of the brain to identify and attenuate background noise. Non-invasive EEG data produced a predicted audio file through an existing state-of-the-art AAD framework, incorporating deep learning. A Deep Attractor Network was then employed to separate this file into its component voices through the use of attractor points. To compare the attended audio file and its component voices, a novel methodology implementing a Fast Fourier Transform deconstructed these files into their constituent frequencies. A correlation analysis then determined the component voice with the highest match to the attended audio file.

[Testing] The degree of correlation between the component voices and the predicted audio file was measured by a Perceptual Evaluation of Speech Quality index and Mean Opinion Score index using a Tensorflow program. Both indices demonstrated correlation at ~90%, indicating excellent audio quality and comprehension from both the human and computer perspective. Thus, the integrated AAD approach developed is suitable for EEG-based speech separation and attenuation.

[Applications] Since the algorithm functions in real-time and trains on unclean speech sources, it is ideal for implementation into hearing aids. The developed AAD technology also shows potential in military environments with heavy background noise and existing smart home devices.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES ☒ NO

3. This project is a continuation of previous research (Form 7): YES ☒ NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES ☒ NO

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: ☒ YES NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ☒ YES NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:10 PM*

Determining Gene Interactions in Congenital Heart Disease for Development of a Comprehensive Fetal Cardiac Diagnostics Platform

CBIO013

Computational Biology and Bioinformatics

Krithik Ramesh

Cherry Creek High School, Greenwood Village, CO

Cardiovascular malformations are the most common type of birth defects. Congential HeartDisease occurs in approximately 1% of all births globally and has a 48.1% infant mortality rate. Giventhat there is significant ambiguity with identification of which biomarkers and genetic factors that areassociated with CHD, there is significant need for a comprehensive and longitudinal understanding of CHD.Iinvestigated the gene interactions in congenital heart disease by using a generative tensorial reinforcementlearning network (GENTRL) to map the active kinase trends and molecular structural trends to see activationpatterns in amniotic fluid. This system was able to identify 132 novel gene interaction pathways. Based onthe genetic analysis trends I developed a conditional generative adversarial network that is able to predictthe morphological deformation and develop a 3D model. The reconstruction accuracy was evaluated at86.32%±5.84% as evaluated by the dice similarity coefficient. Based on the 3D reconstruction a Gaussianapproximation metric was used to create pseudo ECG data with 94.6% accuracy. the data suggests that thecombined genetic and morphological metric serves as a viable early-detection and diagnostic tool for CHD.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:11 PM*

Finding the Most Influential Factors in the Healing Process of Diabetic Foot Ulcers Using Parameter Space Geometry

CBIO014

Computational Biology and Bioinformatics

Gloria Huang

The Carol Martin Gatton Academy of Mathematics and Science in Kentucky, Bowling Green, KY

The treatment of chronic wounds has long been a challenge to wound care professionals and presents a substantial economic burden to healthcare systems globally. Over $50 billion is spent on the treatment of chronic wounds each year, with the annual cost rising as chronic wounds are becoming more prevalent and difficult to treat. To combat this issue, a mathematical model describing the interactions between matrix metalloproteinases (MMPs), their regulators (TIMPs), fibroblasts, and the extracellular matrix (ECM), which is the primary measure for the healing response in the wound, was developed and analyzed to find the most influential factors, or parameters, in the healing process of diabetic foot ulcers. Using the differential equation model with de-identified patient data, the three-dimensional geometry of parameter space was visualized for all combinations of the twelve parameters in the model to more precisely see how these parameters affect the biological system. Knowledge of the identifiability of parameters can, in turn, streamline treatment by allowing us to individualize treatment for each patient. This approach plots two parameters against the sum of squared errors to generate a three-dimensional graph. By analyzing the minimum of the graph, we can conclude if a parameter is able to be uniquely determined, or identified. The identifiability of a parameter signifies its importance in the healing response. This research shows that the regulators of MMPs (TIMPs) are the most influential parameters in a wound-healing model. With this knowledge we can better illuminate the unpredictable nature of wound healing.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants | potentially hazardous biological agents

vertebrate animals | microorganisms | rDNA | tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):		YES	✘	NO
3. This project is a continuation of previous research (Form 7):		YES	✘	NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):		YES	✘	NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	✘	YES		NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	✘	YES		NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Identification of Novel LEAD Compounds for Solute Transporter Protein Member 5 in Sensorineural Hearing Loss

CBIO015

Computational Biology and Bioinformatics

Saaim Ali Khan

Cranbrook Kingswood School, Bloomfield Hills, MI

2 statistically likely, novel drug candidates for solute transporter protein member 5, SLC5, were identified in this study. Dysfunction of a roundworm orthologue of SLC5, sulp-5, was linked to sensorineural hearing loss (SNHL) in 2015. The goal of this project was to examine how this phenomenon in roundworms translates to humans and identify molecular drug candidates for SLC5 in an effort to advance SNHL treatability. Homology modeling with loop refinement and energy minimization was used to generate a competent 3D model of SLC5. The model was then subjected to rigorous topological shape analysis to identify its primary functional binding pocket. Upon detection of this binding pocket, 2 genetic variants were found directly on it, emphasizing the need for drug candidates specific to this functional surface on SLC5. With this knowledge in hand, a virtual protein-ligand docking screen of 8.1M drug candidates was conducted on SLC5 using a self-assembled 16 node HPC cluster. Through careful evaluation of docking results, ethyl 3-methyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidine-5-carboxylate(L4) and 1-(4-(azanyl)butyl)-3-((methyl-azanyl)methyl)-1-azetidine(L1) were culled as the likely drug candidates for mitigation of SLC5 dysfunction. These predictions were further substantiated by a self-assembled probabilistic bootstrap computation model that yielded p-values of 4.02E-5 and 1.19E-18 respectively. Not only were these candidates identified with reputable statistical confidence, but the main functional binding pocket of SLC5 was discovered with precise residual accuracy. Furthermore, the accessibility of the binding pocket to genetic variance was pinpointed. With these discoveries, the future of molecular therapeutics for SNHL is brought appreciably closer to reality.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

An Enhanced Early Detection Model of Dengue Fever Outbreaks Using SEIR Infectious Disease Epidemiological Compartments, Generalized Linear Regression, and Statistical Computing

CBIO016

Computational Biology and Bioinformatics

Tarun Kumar Martheswaran

The Waterford School, Sandy, UT

Dengue Fever is the most rapidly spreading mosquito-borne disease, with more than 25,000 deaths annually. To date, no vaccine has been developed for Dengue Fever due to the existence of four virus serotypes. This project aims to innovate a novel approach to detect outbreaks of this fatal disease. My primary motivation to continue this research was the “2019 Dengue Epidemic,” in which there was an unexpected surge in case count in several Latin American and Southeast Asian countries. I used the SEIR model to track the dynamics and transmission of the disease between humans and vectors (mosquitoes). Using outbreak data collected from Singapore’s Health Database, I then created a linear regressional relationship between variables and the Density of Infectious Mosquitoes. The independent climatic variables were Temperature, Rainfall, and Humidity. Each variable was tested for predictive significance at different amounts of delay or “lag time.” With the ability to predict the Density of Infectious parameter in the SEIR model, I was then able to conduct week by week simulations in R. The model was then tested on five actual Singapore Dengue Outbreaks from years 2013-2019. Statistical testing using cross-correlation showed a significant similarity between the test data and actual Dengue outbreaks. These were a 2013 value of 0.827, a 2014 value of 0.678, 2015 value of 0.813, 2017 value of 0.968, and 2019 value of 0.806. In addition, the model was applied to 2019 outbreaks Honduras and Cambodia with successful results of 0.935 for Cambodia and 0.899 for Honduras. With this reliable early detection framework, health organizations gain lead time to implement intervention strategies to eradicate mosquito populations, and resultingly, large outbreaks of Dengue Fever.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants
potentially hazardous biological agents
vertebrate animals
microorganisms
rDNA
tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): [✓ YES] NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:12 PM*

Abstract Text Mining to Create an Exhaustive Disease-disease Correlation Database

CBIO017

Computational Biology and Bioinformatics

Suchir Misra

Jericho High School, Jericho, NY

Disease-disease correlations lend to improved treatment modalities, however, current disease-disease correlation databases fail to include poorly-studied diseases. Craniosynostosis, despite being the second most common craniofacial abnormality, is one such disease typically overlooked for disease-disease correlations. Text mining, specifically abstract text mining, may serve as a reliable process to establish disease-disease correlations.

In this study, a computational approach was designed to create a disease-disease correlation database using solely abstract text mining. Python programs were written to collect abstract IDs for all genetic (i.e. terms related to genetics) papers (N = 2,056,144) to identify disease-disease associations. Abstracts were used to extract gene names. Disease-disease correlations were determined via gene overlaps.

The top ten disease-disease connections overall and shared drugs were previously elucidated in literature, validating the effectiveness of the proposed algorithm and the power to improve disease treatment modalities through use of it. Of the top ten disease-disease connections for craniosynostosis, four were newly elucidated, illustrating the power of the database to find novel correlations for lesser studied diseases.

This study created the most comprehensive disease-disease database available and the first to be created using solely abstract text mining. Future iterations of the database will include real-time search functions and automatic updates to the data collection. Physicians and researchers will both be able to use the database to design disease treatments for both rare and common diseases that lack viable treatment options today.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	✗ YES	NO
3. This project is a continuation of previous research (Form 7):	✗ YES	NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES	✗ NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	✗ YES	NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	✗ YES	NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

A Smart Heart Monitoring System with ECG Analysis Based on Deep Transfer Learning

CBIO018

Computational Biology and Bioinformatics

Jason D. Park

Baton Rouge Magnet High School, Baton Rouge, LA

In 2015, 13,000 Louisiana residents died of heart disease. Due to expensive cost of ECG analysis, underrepresented community people cannot get medical supports. There are challenges on medical datasets because it is more expensive to collect and label medical datasets than typical image datasets due to cost and patient privacy issues. Moreover, datasets collected from new medical devices are different and cannot be used for previous analysis algorithms. Therefore, it is challengeable to apply deep learning algorithms to different datasets, such as hospital grade 12-lead ECG data and cheap mobile single-lead ECG data. This project proposed a software based on a noble deep transfer learning algorithm to analyze mobile single-lead ECG signals based on the knowledge acquired from 12-lead ECG datasets. In order to train the deep CNN model, the project used a public ECG dataset produced by MIT and Boston's Beth Israel Hospital (BIH). To transfer knowledge from 12-lead data to mobile ECG data, the project has two-step training: 1st step pre-training with a 12-lead public ECG dataset (MIT-BIH) and 2nd fine-tune training with a single-lead ECG dataset (Alivecor). The experiment results showed that the developed approach's accuracy results, such as precision and recall, are better than the accuracy results of Cardiologists. Furthermore, the experiment results showed that the accuracy results are higher than those of the previous works (e.g., Stanford University). In conclusion, the project showed that the proposed Smart Heart Monitoring System with ECG Analysis based on Deep Transfer Learning has a potential to help under-represented community with low cost and high accuracy of prediction.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

[✓ human participants] potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

The Development of Neural Network Inversions on Synthetic MRI Data Masks to Accurately Estimate Brain Tissue Stiffness Indicative of Alzheimer's Dementia

CBIO019

Computational Biology and Bioinformatics

Andrew Cao

Century High School, Rochester, MN

Abnormally stiff brain tissue is associated with neural damage and can indicate Alzheimer’s dementia, other dementias, and normal pressure hydrocephalus. Currently, early Alzheimer’s diagnosis is conducted with inadequate, qualitative mental interviews that are not applicable to all patients. Magnetic Resonance Elastography (MRE)—a subsection of MRI that identifies tissue stiffness—creates a numerical measurement that can indicate forms of dementia and increase chances of early diagnosis. However, the traditional MRE method of direct inversion (DI) tested on in vivo MRIs yielded a low correlation of 0.49 and is biased around the brain’s edges. Thus, training neural network inversions (NNIs) on MRIs could increase stiffness estimation accuracy and generalization. Training NNIs on artificial data could also produce higher accuracy and generalization than previously described models. The dataset contained 5 regions per 85 peoples’ in vivo MRI masks, the workable 3D representations of the brain. Seven main NNIs were trained through an Inception-like convolutional architecture: no-masks (NM), all-masks (AM), one-person’s masks (OPM), one-region’s masks (ORM), randomized-masks (RandM), and artificial-masks (ArtifM). The models were tested on three datasets: AM, OPM, and ORM. The coefficients of determination were found to be: NM-0.42, AM-0.99, OPM-0.98, ORM-0.99, RandM-0.95, ArtifM-0.83. In summary, the NNI significantly estimated brain tissue stiffness more accurately than DI, while increasing generalization and decreasing bias, and proves to be a potentially accurate, efficient MRE technique for early Alzheimer’s dementia diagnosis. The NNIs in this study have real-world implications on decreasing time and manpower spent, improving early diagnosis of Alzheimer’s dementia.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:13 PM*

Multiparameter Optimization of the Extracellular Matrix: A Novel Approach to Controlling Cancer

CBIO020

Computational Biology and Bioinformatics

Antara Pal

Acton Boxborough Regional High School, Acton, MA

Cancer originates from genetic and non-genetic alterations which induce not only abnormal cell proliferation, but also abnormal migration. Local invasion involves motile cancer cells advancing through the extracellular matrix (ECM) by secreting Matrix Metalloproteinase (MMP) and contributes to over 90% of cancer mortalities. Thus, it is critical to develop novel approaches to limiting local cancer invasion by studying cancer cell-ECM interactions. ECM microenvironments differ in cell-matrix adhesions and fiber rigidities. To understand the roles these ECM parameters play in local invasion, the Cellular Potts Model (CPM) was used to simulate local invasion through parallel linear ECM configurations of varying cell-matrix adhesions and fiber rigidities. ECM fibers with strong cell-matrix attachments were found to generate cell pseudopodia, which aid in increasing local invasion rate, while weaker attachments prevent the cells from forming protrusive regions, limiting invasion. ECM arrays with rigid fibers elongate the cell body, allowing the cells to form cell protrusions, resulting in increased cancer invasiveness. Conversely, soft fibers stimulate cell rounding, which is associated with limited migration. Combining the weakest cell-ECM attachments and softest ECM fibers decreased local invasion by over 90%. To the best of my knowledge, this is the first time multi-parameter optimization has been applied to ECM parameters. Understanding the interactions between the ECM and cancer cells may provide insight into novel therapeutic approaches to prevent cancer migration and improve survival.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	YES [✓ NO]
3. This project is a continuation of previous research (Form 7):	[✓ YES] NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES [✓ NO]
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	[✓ YES] NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	[✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:13 PM*

Improving Hazard Characterization in Bacterial Pathogens: Predicting Efficiency of Antibiotics Using Machine Learning

CBIO021

Computational Biology and Bioinformatics

Munira AlBerahim

AlTarbyah AlIslamiya Schools, Riyadh, Riyadh, Saudi Arabia

New species of bacteria are being discovered on a daily basis. Fast detection and risk characterization are vital for taking proper action. Current techniques and methods used for microbial risk assessment in clinical, environmental, and food samples rely upon conventional clinical microbiology monitoring approaches that are laborious, time-consuming, expensive, and suffer from a number of considerable drawbacks. Thus, the primary objective of this project is to improve hazard characterization in bacterial pathogens by predicting the efficiency of antibiotic concentrations using machine learning, specifically artificial neural networks (ANN). Algorithmically, opensource data on bacterial pathogens and related clinical literature were utilized to extract features that represent the relationship between drug exposure conditions and their outcome response (i.e. MIC of Penicillin on E. coli). The ANN predictors were developed using Python libraries (TensorFlow and scikit-learn). The predictors were trained using backpropagation algorithm and then validated based on annotated bacterial pathogens through k-fold cross validation. The predictors were then applied to currently unclassified novel bacterial pathogens and identified their antibiotic response with an accuracy score greater than 80%. This study proposes a futuristic assistive laboratory product, that is affordable and efficient in predicting hazardous bacterial life forms in laboratories, with the potential of advancing the field of biotechnology in various industries; such as, clinical research, drug discovery, diagnostics, and human health care worldwide.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

A Pilot Benchmarking Study of Deep Neural Network Performance on Low Magnification Pathology ROIs

CBIO022T

Computational Biology and Bioinformatics

Christopher Pondoc, Alexander Plisov

Frederick High School, Frederick, MD

Deep Neural Networks (DNNs) have successfully demonstrated superior overall performance in many image classification and recognition tasks on H&E histology images. Reported studies typically utilize high quality (20x or 40x) Whole Slide Images (WSIs) in order to deliver optimal performance. However, it remains uncertain how well DNNs can perform on lower quality region-of-interest (ROI) histology images in real life scenarios. The NCI Patient Derived Models Repository (PDMR) database hosts a catalog of low magnification (4x) ROIs of tissue histology images across a total of 60 cancer models, providing an ideal test case for evaluating DNNs performance in real life scenarios. This study used 5 pre-trained models to benchmark the NCI PDMR database ROIs on a selected set of popular DNN classifiers. Overall, on the binary carcinoma vs. sarcoma classification test, the downsizing models have reached 89.57% top-1 accuracy on 4X ROIs and the patch-based models have reached 84.18% top-1 accuracy on 4x ROIs. On the multiclass carcinoma classification test, the downsizing models reached 72.06% top-2 accuracy on 4X ROIs and the patch-based models reached 78.07% top-2 accuracy on 4x ROIs. With such accuracies, the goal is to utilize the DNNs to perform crucial tele-pathological tasks in underdeveloped countries and rural areas, enabling scientists to take a cell phone picture and feed that image into a battery powered small computer for a quick screening on the field.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification 5/16/2020 6:07:14 PM

Activity-by-Contact Model to Predict Enhancer-Gene Connections: A Tool to Increase Our Understanding of Cancer

Lillian Kay Petersen, Maximilian Eaton Corliss

Los Alamos High School, Los Alamos, NM

CBIO023T

Computational Biology and Bioinformatics

Gene expression is regulated by proteins known as transcription factors, which bind to specific DNA sequences called enhancers. Enhancers activate nearby genes, but there is still a limited understanding of which genes they regulate. We created an Activity-by-Contact (ABC) model to predict enhancer-gene connections based on the three-dimensional structure of the genome. Predicting enhancer-gene connections is important because it can identify mutant transcription factors causing the up-regulation of oncogenes in cancer patients. First, we conducted a validation in the K562 cell line and found that the ABC model predicted enhancer-gene connections significantly better than the previous method of using linear distance. Next, the model was applied to study 24 B-Cell Leukemia patients. The samples were first grouped into subtypes by comparing principal component analysis of their gene expression data to 2,000 previously identified samples. Differential enhancers, differential genes, and those with high ABC scores to each other were identified within each subtype. In these cases, the differential enhancer likely regulates the differential oncogene. We were able to identify specific enhancers that regulate known leukemia oncogenes such as FOXO4 and HUWE1. This can allow for the development of novel drugs to target these mutant transcription factors and thereby treat the cancer. This model builds a better understanding of the mechanisms of gene regulation and supports the theory that genes are regulated by enhancer activity and enhancer-promoter contact frequency. The ABC model has the ability to illuminate pathways of oncogene activation, identify mutant transcription factors, and lead to the development of new drugs for targeted treatment of cancer.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:14 PM*

Functional Genetic Biomarkers of Alzheimer's Disease and Gene Expression from Peripheral Blood

CBIO024T

Computational Biology and Bioinformatics

Amish Sethi, Andrew Warren Ni

Pine-Richland High School, Gibsonia, PA

Detecting Alzheimer’s Disease (AD) at the earliest possible stage is key in advancing AD prevention and treatment but is challenged by confounders with the normal aging processes in addition to other neurodegenerative diseases. Recent genome-wide association studies (GWAS) have identified associated alleles, but it has been difficult to transition from non-coding genetic variants to underlying mechanisms of AD. We sought to reveal functional genetic variants and diagnostic biomarkers underlying AD using machine learning techniques. We first developed a Random Forest (RF) classifier using blood microarray gene expression data from 744 participants in Alzheimer’s Disease Neuroimaging Initiative cohort. After initial feature selection, 5-fold cross-validation of the 100-gene RF classifier achieved an accuracy of 98.4%. The high accuracy of the RF classifier supports the possibility of a powerful and minimally invasive tool for screening of AD. Then, unsupervised clustering was used to identify relationships among differentially expressed genes (DEGs) the RF selected. Results suggest downregulation of global sulfatase and oxidoreductase activities in AD through mutations in SUMF1 and SMOX respectively. Finally, we used Greedy Fast Causal Inference (GFCI) to find potential causes of AD within DEGs. In the causal graph, HLA-DPB1 emerges as the largest node. HLA-DPB1 is downregulated and indirectly causes AD, validated by its mechanisms in the immune system which lead to increased neuron death and the progression of neurodegenerative disorders through its role in T-cell receptors and antibody/antigen production. This study further advances understanding of molecular mechanisms underlying AD and provides potential gene targets for further experimentation.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:14 PM*

Genes that Make You Go HMM: Errors in Protein Sequences Propagated by Heuristic Conveniences Fixed with Probabilistic Models

Qijia Liu

Westview High School, Portland, OR

CBIO025

Computational Biology and Bioinformatics

The sequences of most proteins are inferred from the longest open reading frame (ORF) found in their parent mRNA. While the rule is simple and convenient, it does not model the underlying biology, such as composition of sequences and the Kozak consensus. As genomes and their gene parts are in immense quantity and lay the foundation of modern biology, even a small percentage of potential misannotations is critical. To improve the accuracy of gene annotation, hidden Markov models of mRNA structure are built to identify the most probable protein sequences. Results show that approximately 4.132% of the proteins do not follow the longest ORF rule and may be misannotated; in addition, for 39.7% of all the differences in gene annotation, the hidden Markov model predicted a longer ORF than the longest ORF model due to outdated data. Further examinations of a subset of highly conserved proteins corroborate this interpretation.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	[✓ YES] NO
3. This project is a continuation of previous research (Form 7):	YES [✓ NO]
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES [✓ NO]
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	[✓ YES] NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	[✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

PANCREAS.AI: Novel Deep Learning-based Screening Towards Precision Applications for Pancreatic Cancer

CBIO026

Computational Biology and Bioinformatics

Rishab Jain

Westview High School, Portland, OR

Pancreatic cancer is a devastating and incurable disease with a 5-year survival rate of just 10%. Due to genetic alterations known as mutations, certain cancer-inhibiting and cell vitality genes are affected, allowing cancer to rapidly metastasize across the body. Today, three main treatments exist for pancreatic cancer: radiotherapy, chemotherapy, and immunotherapy. Precision medicine aims to tailor treatments to patients based on their genetic mutations; this yields better survival-rates than conventional methods. However, for pancreatic cancer, current diagnostic tools can lose up to one month of valuable treatment time in turn-around procedures. My research aims to solve the problem of lost treatment time and provides a highly confident artificial-intelligence tool enabling precision medicine treatments for pancreatic cancer. By classifying genetic mutations of patients, this research enables doctors to use targeted treatments potentially improving patient survivability by up to 13%. My research utilizes deep learning to accurately predict the genetic mutations of patients, based on the tissue from their biopsy. Using over 450 pancreatic cancer biopsy images, radiomics allowed me to quantify imaging features such as density and texture from cancer tissue. I utilized the MATLAB environment to train a “custom build” of the Inception-v3 deep learning network on the dataset. The network was able to successfully predict five pancreatic cancer mutations including KRAS, TP53, and CDKN2A. This research will allow oncologists to recommend targeted therapies yielding a higher probability of success -- while saving up to 30 days of turn-over time. Future work will explore feature-based prediction of progression for pancreatic cancer.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:15 PM*

A Real-Time miRNA-based Machine Learning Approach for Precision Cancer Therapeutics

CBIO027

Computational Biology and Bioinformatics

Darsh S. Mandera

Jesuit High School, Portland, OR

Artificial intelligence has been used to identify cancer types with high accuracy, but despite this, we see increased death rates in cancer. Even though cancer is a complex and extremely heterogeneous condition, the current practice of treating cancer is a one-size-fits-all approach that is expensive, time-consuming, causes the patients to suffer, and worse, prescribed cancer drugs are ineffective 75% of the time.

Advances in cancer genomics and pharmacogenomics have generated a wealth of omics data. Recent research shows that microRNAs (miRNAs), which serve to regulate gene expression, become dysregulated in cancer cells and tissues. This has led many to investigate the use of miRNAs as biomarkers for cancer detection and targeted therapy prediction. Machine learning can be used to deliver precision cancer therapeutics based on patients' genomic profiles. The solution is a machine learning platform that analyzes pharmacogenomic and miRNA data of various cancer types and predicts targeted drug efficacy with high accuracy. In this research, models were built using miRNA and drug response data of real human patients - rather than cancer cell lines and humanized mice models - from the Cancer Genome Atlas, a publicly available data repository. The machine learning algorithms tested in this research were OneVsRestClassifier, K-NearestNeighborsClassifier, AdaBoostClassifier, and DecisionTreeClassifier. An ensemble learning method combining multiple weak learners, OneVsRest was able to predict drug efficacy with the highest accuracy of 74.1%. The results show that the approach is superior to current state-of-the-art research that predicts drug efficacy based on miRNA data with an accuracy of 67%.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:15 PM*

Mining Common Structural Features of Aggregating Proteins in Neurodegenerative Diseases

CBIO028

Computational Biology and Bioinformatics

Shuwen Zhang

Vivian Webb School, Claremont, CA

I set out to find the similarities between different aggregating proteins' amino acid sequences and 3D conformations in neurodegenerative diseases (NDDs) using computational methods. Using the found similarities, I then pinpointed drug targets between some proteins, which may become useful references for future NDD therapeutics. Because the amino acid sequences and the crystallography structures of most NDD-related proteins are publicly available on Protein Database (PDB), I used the derived PDB files and FASTA sequence files to conduct several computational analyses. First, I used tools such as BLAST and Topmatch to align the amino acid sequences of the proteins to see how similar the sequences are. Then, I employed algorithms such as FATCAT and Dali to search for similarities between their 3D structures. From the most similar pairs of aggregating proteins, I then used icn3d to hone in on specific similar sequences between Transportin-1 and Tau, the most similar protein pair. Though I found no significant similarities between the amino acid sequences of aggregating proteins, there were six significant structural alignments found during my research. Of these six pairs, three (Transportin-1 and Tau, Transportin-1 and APP, and APP and Tau) had a significant structural similarity. I then used the obtained binding sites for Trn1 and Tau to determine respective drug targets since there are no drugs that target both simultaneously. Possible future targets to be tested for Tau are S-Adenosylmethionine (DB00118) and Sirolimus (DB00877). A potential drug that could be repurposed to target Trn1 is Acetic acid (DB03166).

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

DeepChem: Using Generative Adversarial Networks to Develop Novel Inhibitors of Carcinomas

CBIO029T

Computational Biology and Bioinformatics

Harsh Deep, Krishay Mukhija

The Harker School, San Jose, CA

In our project, we sought to use machine learning to generate new lead molecules and expedite drug discovery. Lead molecules are molecules with an affinity for inhibiting the target protein, and identifying suitable leads is the first stage in drug discovery. Our approach for generating lead molecules differed from current methodologies in 2 areas. First, instead of only training our network on drugs that affected the target protein, we found drugs that targeted proteins that were structurally similar to the target protein. This helped create a more diverse dataset and helped augment the dataset for proteins that currently do not have many drugs that can inhibit them. Secondly, whereas many previous machine learning approaches to drug discovery utilized Recurrent Neural Networks, we used Generative Adversarial Networks(GANs) which provide the benefit of having 2 competing neural networks that continuously learn. Over time, the generator becomes more adept at producing realistic drugs and the discriminator becomes better at discerning which drugs are real versus generated. The effect of this is that the generated molecules more closely resemble the style of input molecules while also containing slight variations. Overall, our generated inhibitors had an average docking score of -7.9 which was better than that of current lead molecules which sat at -7.1, a more negative docking score is preferred; however, our molecules were still below the average docking scores of current carcinoma drugs which had an average docking score of -9.3.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Modeling Distant Metastasis-Free Survival: Applications to Hazard Prediction and Pairwise Gene Interaction Discovery

CBIO030

Computational Biology and Bioinformatics

Russell Yang

The Harker School, San Jose, CA

Models of metastasis-free survival based on microarray gene expression can be powerful tools for clinicians and have broad applications in risk quantification, genetic research, and palliative care. This project focused on developing various survival models for distant metastasis-free survival and exploring some of the applications of those models in genetic and clinical settings.

Several linear and nonlinear dimensionality reduction methods were employed to transform a set of 20,000+ genes to a smaller dataset. Various survival models were considered, including the traditional semiparametric Cox Proportional Hazards Model, nonparametric ensemble-based methods, and a Cox Proportional Hazards Deep Neural Network. Each combination of dimensionality reduction and survival model was evaluated using a nonparametric bootstrapping approach (B=100). Concordance indices were used to assess model prediction ability. The assumptions of the Cox Proportional Hazards Model were rigorously tested using Schoenfeld, Deviance, and Martingale residuals.

Next, an algorithm was devised to identify potential pairwise epistatic gene interactions between MYC (a proto-oncogene) and other genes in the dataset. A Kolmogorov-Smirnov test was performed, showing a highly statistically significant difference in gene z-score density upon the addition of MYC to the model. Lastly, preranked gene set enrichment analysis was performed on significant model-identified genes (with Hallmark and C1 positional sets as references) to find pathways and their associated normalized enrichment scores.

The models developed in this study can be used to estimate the prognostic effect for any gene of interest. Perhaps more importantly, the models can also be used for survival prediction in a palliative setting.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants
potentially hazardous biological agents
vertebrate animals
microorganisms
rDNA
tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Gene Embedding: A Novel Computational Hybrid Approach to Somatic Mutation-Based Primary Cancer Type Identification and Biomarker Discovery

Sidra Yang Xu

The Harker School, San Jose, CA

CBIO031

Computational Biology and Bioinformatics

Cancer is a leading cause of death worldwide. Because the organ and cell type that generated the tumor determine a patient's response to therapies, quick and accurate identification of the primary site is critical for guiding effective treatment. Yet there is no rapid and effective approach available to aid early screening.

Somatic point mutation-based cancer typing has the potential of differentiating similar tumors and delivering accurate results, but researchers face limited accuracy due to the challenge of modeling complex gene interactions. To address this issue, a novel approach is implemented in this study: to harness the power of gene expression data and to include it in a somatic mutation-based cancer identification model through embeddings. An embedding is a mapping of complex categorical variables to vectors of continuous numbers with the capability of uncovering relationships between these variables that are otherwise regarded as independent entities. By introducing portable gene expression embeddings, the model can harvest information in both somatic mutation and gene expression without requiring the latter from patients, often not available in clinical settings.

Results in this work show that when a gene expression embedding extracted from all cancer-related genes in TCGA databases is applied, the model provides a prediction accuracy of 76% on 12 tumor classes, an improvement of more than 15% compared with previous studies without embeddings. My research is the first to demonstrate the success of a hybrid genetic model. Furthermore, feature ranking analyses reveal a number of genetic markers specific to each cancer type, which could be studied for a better understanding of each cancer and utilized as therapeutic targets in the future.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:17 PM*

Real-time Correlation of Electrode Stimuli onto a Movement Model

CBIO032

Computational Biology and Bioinformatics

Adriel Torresola Merced

Centro Residencial de Oportunidades Educativas de Mayagüez, Mayagüez, Puerto Rico, Puerto Rico

When the conscious mind declares a movement in the body, the central nervous system carries a signal to the muscles by way of neurons. This signal causes the muscles to contract in the form that is desired. In electromyography, this signal is measured and recorded for analysis. However, this data can fluctuate greatly therefore complicating typical analysis. In this research project, the investigator intends to utilize machine learning algorithms to successfully identify movements being conducted by the subject. The software utilizes a database of measured nerve stimulus and its respective movement as a training to identify movements. The motion data is collected utilizing the Kinovea software to track points in the arm which can be utilized to calculate characteristics such as distance, angle, and velocity. For the electromyography readings, the investigator utilized the Backyard Brains Heart and Brain Spikershield. The software references the database to be able to accurately predict movement utilizing the measured muscle voltage. To verify the validity of the predictions, the researcher compared the machine generated movement predictions, with the actual recorded movements. In comparing both sets, the predicted information was found to be significantly similar, therefore accepting the hypothesis. Because surface electromyography possesses minimal risk and a cheaper cost, utilizing programs such as these to facilitate a direct nerve interface could allow for more accessible and cost-effective prosthetics for disabled patients. For this, the project could be expanded, and given more training data for the software to analyze and improve the predictions.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

[✓ human participants] potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): [✓ YES] NO

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED

REGENERON ISEF

May 10-15, 2020

Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:17 PM*

Implementing a Machine Learning Model in an Effective and Low-Cost Parkinson's Disease Diagnosis Algorithm

CBIO033T

Computational Biology and Bioinformatics

Sanjita Pamidimukkala, Dennis L. Chan

Dougherty Valley High School, San Ramon, CA

Parkinson’s Disease (PD) is a disorder that inhibits the dopaminergic neurons in the brain. Currently, there is no accurate non-clinical way to diagnose PD, forcing medical professionals to resort to subjectivity. Doctors utilize interviews, motor tests, and neurological examination using infrastructure such as DaTscan technology. This places a heavy financial burden on patients and costs the USFG over 25 billion dollars. Even if diagnosis is accurate, it is subject to months of delays and 60% of the time it comes after substantial neuron loss. This research aimed to develop an effective, low cost, and quick way to objectively diagnose PD. The UCI Machine Learning Repository was used to obtain data collected from Archimedean Spiral Tests. Two machine learning classifiers were tested: Deep Neural Network and Random Forest. The algorithm implemented a DNN with 2 hidden layers with 10 nodes in each. Additionally, a RF classifier was also evaluated. RF is a collection of decision trees that each make its own prediction and will collectively sway towards a class prediction. Four attributes were used to prepare data to be trained: x-position, y-position, time stamp, and test ID. Mean speed, skew of speed, radial velocity, and coefficient of variance of speed were extracted from the attributes using mathematical formulas. Each model was trained with the data and evaluated. The DNN had an average accuracy of 90.8% and the RF had an 85.2% average. Although RF had less spread, the DNN was classified as the best model because DNNs are stochastic and use randomness to their advantage. Overall, this algorithm proved to be over 15% more accurate than current rudimentary diagnosis methods. It reduces delays by months, reduces costs by thousands, and increases accessibility.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:17 PM*

Survival Analysis of Pediatric Neuroblastoma Patients Using Gene Expression Data: Identification of Novel Neural Differentiation Related Biomarkers

CBIO034

Computational Biology and Bioinformatics

Ankita Anand Menon

Nikola Tesla STEM High School, Redmond, WA

Of the over 100 types of cancers that together claim 2,000 lives in the United States daily, neuroblastoma is the most common in children under the age of two. This project seeks to identify novel prognostic biomarkers for neuroblastoma that can be used to develop faster/more efficient therapies and detection models, with a focus on the NEK2 gene and genes involved in neural differentiation (especially the Wnt and Notch signaling pathways). The gene expression and clinical data of 249 pediatric neuroblastoma patients was obtained from the TARGET database, and univariate Cox hazard regressions (a statistical survival analysis) were applied to each of the 23,000 genes to determine whether higher gene expression correlates with lower event-free survival time (prognosis). After applying univariate regression models, genes with a hazard ratio (comparison between probability of an event occurring based on a set variable such as gene expression) above 1.0 and p-value below 0.05 were researched for relevancy to the project purpose and novelty to neuroblastoma, and then individually put into multivariate Cox regression models (Cox regression factoring covariates such as age and gender) to determine whether the genes would still be significant prognostic markers. The hypothesis that NEK2 would be a prognostic biomarker was supported (with a multivariate hazard ratio of 1.368 and p-value of 0.0353), and additionally 16 other novel neural-differentiation related biomarkers were identified, which not only gives new insight into the pathways that neuroblastoma arises from, but also paves the way for novel, noninvasive gene therapies that minimize the chance of recurrence and rapid, preemptive cancer detection models to detect the cancer before metastasis.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	YES [✓ NO]
3. This project is a continuation of previous research (Form 7):	YES [✓ NO]
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES [✓ NO]
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	[✓ YES] NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	[✓ YES] NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:18 PM*

Identification of Novel Antimicrobial Peptides with Designed Activity through a QSAR Based Machine Learning Model

CBIO035

Computational Biology and Bioinformatics

Fiona Lin Luo

Monta Vista High School, Cupertino, CA

Antimicrobial peptides (AMPs) are a growing field of therapeutics with lower risks of antimicrobial resistance compared to conventional antibiotics. However, experimentally testing peptides is costly and inefficient, making a preliminary in-silico screening more efficient. Our goal is to design a machine-learning program to predict the activity of AMPs, and to determine novel AMP drug leads.

We present a novel model structure for predicting AMP activity by combining a QSAR (quantitative structure-activity relationship) model and an LSTM (long short-term memory) model. We built a QSAR neural network to predict probability of peptide activity using 29 calculated physicochemical descriptors as a representation of each sequence. We then used a generative LSTM network to sample 10,000 promising de novo sequences and validated the samples' activity with the QSAR model. We chose samples predicted to have over 99% probability of activity and ran them through a protein secondary structure prediction server (CABS-fold) to analyze structure. Finally, we applied our program to anti-HIV sequences to generate novel anti-HIV drug leads.

Our QSAR model achieved an accuracy of 92.60% on AMP prediction and 81.67% on anti-HIV prediction. We determined 69 novel anti-HIV drug leads and 707 antimicrobial leads, largely predicted to be alpha-helices, that can be further tested in vitro.

Overall, our project improves on prior research in several ways. Our model is flexible and can be further applied to AMPs of designed activities (e.g. antiviral, antifungal, antimalarial), and we provide the specific example of anti-HIV peptides. To our best knowledge, we are the first machine learning program to predict anti-HIV activity, and the first QSAR and LSTM combined model which analyzes AMPs.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	YES	☒ NO
3. This project is a continuation of previous research (Form 7):	YES	☒ NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES	☒ NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	☒ YES	NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	☒ YES	NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:18 PM*

Getting Cryptic with Bioinformatics!

CBIO036

Computational Biology and Bioinformatics

Katheryne Lochart

Broadneck High School, Annapolis, MD

This investigation researches if using amino acids instead of nitrogenous bases in bacterial data storage systems protects against the corruption of the stored data due to mutations.
The hypothesis was the storage system that utilizes amino acids as the foundation instead of nucleotides (currently used system) will result in more accurate RNA and the least amount of errors due to mutation. Python was used to conduct a simulation of a randomized codex assigned to a message. The message was decoded through the nitrogenous based system and an amino acid codex. Analysis of the data shows that the amino acid system was more efficient in maintaining the message's integrity than the nitrogenous base system. The nitrogenous base system consistently showed the corruption of the message by allowing the mutation to directly impact the decoded message. In later trials, where mutation rates of 7 and 14 mutations per strand were applied, the nitrogenous base results yielded all of the trials being corrupted. The amino acid-based system tells a different story, for even as mutations increased, amino acid systems were consistently lower than the binary results (e. g. Binary System with 3 and 7 mutations=534 failures while Amino Acid System with 3 and 7 mutations=499, Binary System 5 mutations= 35 while Amino Acid System 5 mutations=26). The amino acid systems consistently yielded close to or zero corruptions, but the nitrogenous base system had many corruptions after decoding, therefore proving that this new system may lead to further data storage innovation.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants　　potentially hazardous biological agents

vertebrate animals　　microorganisms　　rDNA　　tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): YES [✓ NO]

3. This project is a continuation of previous research (Form 7): YES [✓ NO]

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES [✓ NO]

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: [✓ YES] NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. [✓ YES] NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:18 PM*

GADDNET: A Platform for Connecting Researchers via the Genes and Diseases They Studied and Will Study

CBIO037

Computational Biology and Bioinformatics

Emma Yang

Brearley School, New York, NY

Authors, Genes, Diseases, and Drugs in a Network (GADDNET) is a web-based platform that connects researchers and institutions via the genes and diseases they study. GADDNET can help researchers to explore understudied genes and enhance opportunities for collaborations. The platform integrates multiple datasets and presents these datasets to the user as a network centered on a researcher, a gene, a drug, or a disease. The nodes in this network are connected based on published genes and proteins, diseases, and drugs, as well as predicted relevant genes. GADDNET has the potential to dramatically accelerate the progress of drug and target discovery.

The goal of the project is to create a web-based and mobile app that dynamically creates networks linking data centered around a specific researcher, a gene, a drug, or a disease. For example, the app uses the researcher's name to look up the genes, the drugs, and the diseases they have published. These genes, drugs and diseases are found in PubMed abstracts. The genes, drugs, and diseases found are used to identify other researchers who have published research about the same genes, drugs, and diseases. The marquee feature of the platform is that the app also provides genes and drugs predicted to be similar in function to the genes and drugs the researcher has published. These predicted genes and drugs are found based on gene-gene and drug-drug similarity matrices constructed from several different resources. These connections may help researchers identify understudied genes that could become key drug targets and drivers of disease mechanisms, as well as identify opportunities for drug repurposing.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): ✘ YES NO

3. This project is a continuation of previous research (Form 7): YES ✘ NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES ✘ NO

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only: ✘ YES NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ✘ YES NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Determining the Most Cost-Efficient Indicators of Alzheimer's Disease Using Neural Networking

CBIO038

Brea Avery Swartwood

Mililani High School, Mililani, HI

Computational Biology and Bioinformatics

Comparison of methods used to diagnose Alzheimer’s Disease (AD) via imaging, mental, and genetic data remains an open area of investigation. In this investigation, it was hypothesized that after designing and creating a full feed-forward neural network that accurately classifies cognitively normal (CN), mildly cognitively impaired (MCI), or Alzheimer’s Disease (AD) patients, the top 8 indicators of AD will include ApoE 4 and structural volumetric measures. A key public AD Neuroimaging Initiative dataset containing 68 potential indicators for 14,037 patients was cleaned, one-hot encoded, and normalized to 29 potential key indicators and 2,646 patients. A full feed-forward network was carefully designed, which employed a categorical cross-entropy loss function and Adam Stochastic Gradient Descent Optimizer to determine the most accurate indicators of AD. BigML Analysis and Principal Component Analysis were conducted to isolate top indicators of AD, and more than 1,000 trials were run using the created neural network to find the greatest accuracy that could be obtained using 8 indicators. A cost analysis was also conducted to assess the practicality and accessibility of these top indicators. The results partially supported the hypothesis, suggesting that ApoE E4, Clinical Dementia Rating Scale Sum of Boxes, the 11 and 13 item versions of ADAS-Cog, ADASQ4, the Mini-Mental State Examination, hippocampal volumetric measurements, and intracranial volume adjustment measurements most accurately indicate AD. Based on the cost analysis, getting a Mini-Mental State Examination ($79.35) and ApoE 4 testing ($99-$199) are the most cost-efficient indicators of AD.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants potentially hazardous biological agents

vertebrate animals microorganisms rDNA tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	✘ YES	NO
3. This project is a continuation of previous research (Form 7):	YES	✘ NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):	YES	✘ NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only:	✘ YES	NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	✘ YES	NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

An Optimized Molecular Docking Protocol Targeting Musashi RNA-Binding Proteins for Cancer Drug Design

CBIO039

Computational Biology and Bioinformatics

Khushi Kohli

Olathe North High School, Olathe, KS

Acute Myeloid Leukemia (AML) is a deadly blood cancer in which abnormal blast cells crowd out healthy blood cells in bone marrow. Although AML is the most common adult leukemia, its survival rate remains just 27.4%. The Musashi 2 (MSI2) protein, which regulates stem cell division, is necessary for AML progression and a promising drug target for AML targeted therapy. But due to its high flexibility as an RNA-binding protein (RBP), this “undruggable” protein lacks well-defined binding pockets. This presents significant challenges to conventional virtual screening that uses rigid-body docking calculations to evaluate binding affinities of MSI2 inhibitors, rendering RBP inhibitor design incredibly difficult. Because RBPs undergo extensive conformational changes upon ligand binding, it was hypothesized that incorporating protein flexibility in docking calculations will improve accuracy. In this study, rigid-body and flexible docking protocols were evaluated with known MSI2 inhibitors using its NMR and X-Ray structures. Ten actives and 1453 decoys of MSI2 were collected for retrospective docking using AutoDock. Docking performance was assessed through binding geometry, Enrichment Factor, and area under Receiver Operating Characteristic curve (AUC) analysis. Enabling flexibility of target site residues, using optimized rigid-body docking parameters and the NMR structure, significantly improved binding site specificity but diminished the accuracy of the docking scoring function. These findings suggest reasonable incorporation of receptor flexibility is essential for accurate docking of RBPs, including MSI2. The optimized docking protocol for this challenging class of drug targets is expected to advance novel drug lead discovery for a broad spectrum of diseases.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C): ✘ YES   NO

3. This project is a continuation of previous research (Form 7): YES ✘ NO

4. My display board includes non-published photographs/visual depictions of humans (other than myself): YES ✘ NO

5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year’s work only: ✘ YES   NO

6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work. ✘ YES   NO

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:19 PM*

Predicting Recurrence in Triple-Negative Breast Cancer Patients through Analysis of the Tumor-Immune Microenvironment

CBIO040

Computational Biology and Bioinformatics

Aalok Nital Patwa

Archbishop Mitty High School, San Jose, CA

My research identifies novel ways to predict recurrence in triple-negative breast cancer patients. I used multiplexed ion beam imaging to analyze cell prevalence and protein expression in the tumor-immune microenvironment. My results show that the organization of cells and biomarkers effectively predicts recurrence in TNBC and can help doctors pursue strategic therapies, saving more lives.

1. In this research project, the student directly handled, manipulated, or interacted with (check ALL that apply):

human participants   potentially hazardous biological agents

vertebrate animals   microorganisms   rDNA   tissue

2. I/we worked or used equipment in a regulated research institution or industrial setting (Form 1C):	✘	YES		NO
3. This project is a continuation of previous research (Form 7):		YES	✘	NO
4. My display board includes non-published photographs/visual depictions of humans (other than myself):		YES	✘	NO
5. This abstract describes only procedures performed by me/us, reflects my/our own independent research, and represents one year's work only:	✘	YES		NO
6. I/we hereby certify that the abstract and responses to the above statements are correct and properly reflect my/our own work.	✘	YES		NO

The stamp or embossed seal attests that this project is in compliance with all federal and state laws and regulations and that all appropriate reviews and approvals have been obtained including the final clearance by the Scientific Review Committee.

SCIENTIFIC REVIEW COMMITTEE APPROVED
REGENERON ISEF
May 10-15, 2020
Anaheim, California

Official Intel ISEF 2020 Abstract and Certification *5/16/2020 6:07:20 PM*