2017

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Columbia University Undergraduate Science and Engineering Research Showcase

Abstracts

Friday, October 13, 2017
12:00-1:30 p.m.
Roone Arledge Auditorium
Lerner Hall

Table of Contents

Abstracts

Raag Agrawal CC'20: Biology

ra2816@columbia.edu
Faculty Mentor: Harmen Bussemaker, professor of biological sciences and systems biology
Title: Inferring Gene and Transcription Factor Regulatory Mechanisms Through Differential Gene Expression Analysis

Abstract: As gene expression data for many different tissue types becomes widely available, it becomes possible to predict natural human genomic variation and infer gene expression differences between different populations. We use R, a statistical programming language, to create a model using Generalized Linear Models (GLMs) to predict natural human gene expression variation, and separate it from the effects of patient age, ethnicity, and other factors. We show that we can successfully predict natural variation within synthetic datasets, demonstrate possible additions to this model using transcription factor binding sites, and show future extensibility.

Mika Aly CC'20: Biology

ma3600@columbia.edu
Faculty Mentor: Laura Kaufman, professor of chemistry
Title: Molecular Weight Dependencies on Dynamic Heterogeneities in Polystyrene

Abstract: There is still a lot of uncertainty regarding the supercooled liquid regime in glass-forming systems. In order to fully study the specific dynamics of these glassy molecules, the Kaufman lab uses a home-built microscope in a wide-field configuration equipped with a vacuum cryostat for temperature control. This allows us to study the rotations of single molecules by measuring the rotational correlation times (Tfit and Tc) and the relaxation component (B) of each molecule. The lab has already conducted experiments that determine the dynamic heterogeneities with these molecules. Our experiment involves looking at the molecular weight dependencies of Polystyrene and the relative dynamics around their glass transition temperature (Tg+5K). Polystyrene with Mw = 168, 1364,118, 27.5, 4.8 kg/mol and pPDI dye was used. The average rotational correlation time, τc, was calculated from the fit values of τfit and β via $\tau c = (\tau fit/\beta)\cdot\Gamma(1/\beta)$, where Γ is the gamma function. All data analysis was performed using IDL software. We still hope to see dynamic heterogeneity even as we lower in molecular weight.

Adam Block CC'19: Mathematics

abb2190@columbia.edu
Faculty Mentor: Daniel Litt, postdoctoral research fellow and postdoctoral research scientist in the Department of Mathematics
Title: Resolving Symmetric Powers by Schur Functors

Abstract: Recently, under favorable hypotheses, Daniel Litt proved a bound on the Frobenius amplitude of symmetric powers of vector bundles in positive characteristic. We attempt to extend this result by resolving symmetric powers of polynomial representations by Schur functors. Because Schur functors vanish relative to the dimension of the vector space, such a resolution would be an important step in proving the above result.

Viggo Blomquist CC'21: Neuroscience and Behavior

vb2446@columbia.edu
Faculty Mentor: Christopher Ahern, associate professor of molecular physiology and biophysics at the University of Iowa Carver College of Medicine
Title: A Novel Suppression Technique of Premature Termination Codons and its Application to Congenital Long QT Syndrome 2

Abstract: The human ether-à-go-go related gene (hERG) codes for the pore-forming subunit of the voltage dependent potassium ion channel in cardiac cells known as Kv11.1. Potassium channels are voltage gated protein channels within the plasma membrane that allow potassium to move down the electrochemical gradient. The channel's central role is to initiate the repolarization of the cardiac action potential. Premature termination codons in hERG result in a hereditary and potentially deadly disease called Long QT Syndrome 2 (LQTS2). The purpose of this investigation was to test if novel codon edited human V10 tryptophan tRNAs would be able to recognize the premature termination codon within the mRNA strand and add tryptophan to the growing polypeptide chain, preventing premature termination and rescue of Kv11.1. In order to prove successful suppression, Western blotting, patch-clamp technique, and immunofluorescence microscopy were conducted on HEK293 cells transfected with DNA plasmids containing a gene for codon edited V10 Tryptophan tRNA and the mutated hERG gene. I determined that codon edited V10 Tryptophan tRNAs were capable of suppressing the W1001TGA premature stop in hERG, which causes LQTS2. Successful suppression of premature stops by using such tRNAs could lead to major medical advances in gene therapy for not only LQTS2, but also for other diseases caused by premature termination codons.

Sharon Chen SEAS'19: Computer Science

syc2138@columbia.edu
Faculty Mentor: Julia Hirschberg, the Percy K. and Vida L. W. Hudson Professor of Computer Science and chair of the Department of Computer Science
Title: Scraping Code-Switching Data for Modern Language Tools

Abstract: Code-switching (CS) is the fluent alternation between languages during communication. CS is a prevalent phenomenon in multilingual communities, but little effort has been put into developing natural language tools that can handle it. Machine Learning (ML) algorithms are often used in Natural Language Processing (NLP), but they often require large amounts of data, which are currently unavailable for CS. In this research project, we designed and built a system to efficiently find online Spanish-English CS conversational data. Our system makes use of a web-scraper and a crawler using the

Python Reddit API Wrapper (PRAW) for scraping Reddit comments, and a scraper for Reddit Live Threads.

We implemented several CS detection methods into our system to collect CS data and filter out monolingual data. We used lists of anchors, which are words unique to a specific language, and taggers from the Natural Language Toolkit (NLTK), in our methods. Our four methods detected a sentence as CS if it (1) has both a Spanish anchor and an English anchor; (2) has a Spanish anchor and one English anchor, where anchors are not named entities; (3) is detected to be in a language and has an anchor of the other language; or (4) is detected to be in a language and has a non-named-entity anchor of the other language. We tested these methods on labelled Twitter tweets and found that their accuracies were 38.8%, 39.0%, 50.8%, and 51.0%, with high rates of false negatives. We reduced the false negative rate to 64.5% and increased the accuracy to 52.0% by combining the methods to create a classifier that labeled a sentence as CS if any of the four methods labeled it as CS. The f-scores were increased to 50.1% for the CS class and 53.8% for the non-CS class.

Overall, we collected more than a million sentences from Reddit, for a total of 5,022 Spanish-English CS sentences. This new corpus can be used to further study the phenomenon of CS in informal conversations.

Joshua Choe CC'20: Biology

joshua.choe@columbia.edu
Faculty Mentor: Carol L. Prives, the Da Costa Professor of Biology
Title: The oligomeric status of tumor suppressor p53 affects its stability and resistance to chemotherapy

Abstract: Tumor suppressor p53, mutated in about 50% of all cancers, plays a crucial role in DNA damage repair, cell cycle regulation, metabolism, programmed cell death, and other important cellular processes. p53 exists in three oligomeric states: monomer, dimer, and tetramer. Under stress, p53 is found mostly in its tetrameric structure, which is considered the active form. Tetramer formation is essential for DNA binding, transcriptional regulation and posttranslational modification. While the tetrameric form of p53 is very well-studied, the functions of dimeric p53 is far from being fully understood, even though the dimeric form is considered to be the most abundant form in basal conditions. Moreover, patients with Li-Fraumeni syndrome (LFS) are highly cancer prone and harbor germ line mutations in p53. A considerable frequency (19%) of Li-Fraumeni syndrome mutations occurs within the p53 oligomerization domain (OD). Thus, unraveling the changes caused by altered oligomeric states of p53 remains essential for both basic research and clinical studies.

In order to study the dimeric form of p53 in a more physiologically relevant and endogenous setting, dimeric mutations within the OD (E343K, A347D) were generated using CRISPR in an osteosarcoma cell line (U2OS). Following confirmation that these CRISPR mutants do indeed form dimers, dimeric p53 was found to be better degradable by its negative regulator Mdm2, which targets both p53 and itself for degradation in the proteasome. Furthermore, dimeric p53 appears to exhibit partially reduced

transcriptional activity. Interestingly, dimer mutants appear to be less sensitive to the DNA damage inducing chemotherapeutic drug etoposide than cells with wild-type p53, suggesting that dimeric p53 may be less able to induce cell death upon DNA damage. Thus, further work needs to be done investigating the mechanisms underlying this reduced sensitivity which may ultimately allow for more targeted therapeutic approaches for patients harboring p53 mutations in the OD.

Dylan Cooper CC'18 Neuroscience and Behavior

djc2198@columbia.edu
Faculty Mentor: Rae Silver, the Helene L. and Mark N. Kaplan Professor of Natural and Physical Sciences, Barnard College
Title: Morphological Investigation of the Hippocampus in the Context of Neuropsychiatric Diseases

Abstract: Autism Spectrum Disorders (ASD) and Schizophrenia (SCZ) are cognitive, neuropsychiatric disorders with individually complicated neurobiological architectures but common behavioral pathologies, such as impaired social cognition. Accumulating evidence reveals a strong genetic association within these pathologies, and the advent of genetically modified mouse models allows for an in-depth analysis of their underlying mechanisms. Our research aims to determine if the CNTNAP2 and Nrg1 loci, linked to ASD and SCZ in humans, respectively, result in an impaired network connectivity of area CA2, a region in the hippocampus that has recently been identified as crucial for the encoding of social memory. We employ electrophysiological and immunohistochemical techniques to determine whether these mouse models display: (a) a loss of PV+ interneurons in area CA2 (which has been observed in brain tissue from individuals with schizophrenia or bipolar disorder), and (b) altered intrinsic excitability or synaptic circuitry of CA2 pyramidal neurons (which was observed in a separate mouse model of schizophrenia). Understanding the cellular and molecular changes that take place in this brain region may shed light into the etiology of these disorders and may uncover novel targets for future drug development.

Anastasia Dmitrienko CC'21: Statistics; Computer Science

ad3473@columbia.edu
Faculty Mentor: Richard Zink, JMP principal research statistician, SAS Institute
Title: Risk-based Monitoring Rules for Binomial and Poisson Outcomes in Clinical Trials with Software Implementation in JMP

Abstract: Risk-based monitoring (RBM) strategies have been broadly applied in clinical trials to detect sites with unusual characteristics. Monitoring rules may be defined using clinically-based criteria, such as the risk thresholds suggested by TransCelerate Biopharma. Alternatively, monitoring rules may utilize control limits that define an expected amount of variability for the outcome of interest. This poster focuses on developing control limits for binomial and Poisson outcomes. The control limits are based on asymptotic and exact confidence intervals for the corresponding parameters. Factors that affect the expected amount of variability such as patient's time on study are considered when computing the control limits. The methods developed in this poster are applied to define rules for monitoring binary

outcomes such as patient discontinuation and count outcomes such as the number of adverse events per patient in several real clinical trials. RBM rules for binomial and Poisson outcomes are conveniently summarized using funnel plots with asymptotic and exact control limits defined using the methods described above. These RBM strategies are implemented using JMP software.

Costin Dobrin CC'20: Sustainable Development; Economics-Mathematics

costin.dobrin@columbia.edu
Faculty Mentor: Jason Smerdon, the Lamont Research Professor in the Lamont-Doherty Earth Observatory
Title: Tree-Ring Constraints for the Climatic Conditions of the 14th Century European Great Famine

Abstract: The Great Famine was a period of overall agricultural failure in 14th century Western Europe. In the past, the hydro-climatic history of Europe's climate has been predominantly characterized by documentary observations in chronicles and journals. The recently published Old World Drought Atlas (OWDA) allows comparisons between these subjective documentary observations with quantitative estimates from tree-ring archives. According to historical sources, the Great Famine was caused by consecutive years of increased rainfall and cold, in contrast to other agricultural hardships that have been tied to drought conditions. Documentary evidence roughly characterizes the beginning and end of the famine, its spatial expansion across Europe, and its overall evolution. Using the OWDA, the event can now be more accurately characterized in both time and space. Analyzing the data allows a more accurate understanding of the magnitude and the effects of the Great Famine event in the context of the before-and-after conditions. Also, the Great Famine can now be compared and ranked among other hydro-climatic events.

Ailis Dooner CC'19: Neuroscience and Behavior

acd2174@columbia.edu
Faculty Mentor: Francisco Martin, Group Leader for Molecular Physiology of Behavior, Cajal Institute of the Spanish National Research Council
Title: Circadian Rhythms in Drosophila

Abstract: In Drosophila melanogaster larvae, the prothoracicotropic hormone (PTTH) has been previously shown to affect the animals' reaction to light: namely, via interaction with the Bolwig's organ and torso receptors in class IV neurons (peripheral) (Yamanaka et al 2013). However, PTTH's job in circadian rhythmicity regulation in adult flies is not totally grasped. In one portion of this project, the role of PTTH in modulating circadian activity in the adult fly brain was assessed. Different crosses of flies with distinctive expression patterns were established and maintained, and the male offspring were used in Trikinetics experiments. The results of the project were inconclusive, and further investigation will be required. In another portion of the project, confocal microscopy was performed to analyze expression patterns in larval fly brains.

Nicholas Herrera CC'18: Computer Science

nh2473@columbia.edu
Faculty Mentor: Rocco Servedio, professor of computer science
Title: Toward Lower Bounds on Constant-Depth Frege Systems with Modular Connectives

Abstract: The field of proof complexity aims to understand the minimum length of a proofs of theorems in logically complete proof systems, most notably Frege systems whose proofs are sequences of propositional formulas. A major open problem in proof complexity is showing that there exist theorems without Frege proofs of length polynomial in the length of the theorems' statements.

We consider the subsystem of Frege commonly called $AC^0[p]$-Frege whose proofs are restricted to being constant-depth formulas with mod p connectives by analogy with the complexity class $AC^0[p]$. Despite lower bounds for both AC^0-Frege and $AC^0[p]$ being known for nearly 30 years, non-trivial unconditional proof size lower bounds for $AC^0[p]$-Frege have remained elusive. Recent work in the area of algebraic proof complexity — particularly the recently proposed ideal proof system (IPS) — provides connections between our problem and algebraic complexity theory. However, there appear to be several difficulties in applying existing tools for dealing with algebraic problems to our problem. In particular, the existing methods for proving lower bounds on subsystems of IPS are not well developed and work only for IPS over fields of characteristics either zero or very large rather than p, the case linked to $AC^0[p]$-Frege. Extending existing techniques to prove strong enough lower bounds on IPS to prove lower bounds appears at least as hard as extending these techniques to prove stronger circuit lower bounds than are currently known. This suggests that further work must be done on more basic problems in algebraic complexity theory before we can prove lower bounds in the less understood setting of proof complexity.

Albino Folcarelli GS'20: Biochemistry

agf2128@columbia.edu
Faculty Mentor: Dr. Steven Kernie, professor of pediatrics in neurology and director of the Division of Pediatric Critical Care Medicine
Title: The Effects of Social Isolation Followed by Enrichment on Neurogenesis in the Dentate Gyrus of the Hippocampus

Abstract: Social isolation has been shown to facilitate the accumulation of neural stem cells in adult mice. Despite this finding it is unknown how an accumulation of neural stem cells affects their downstream conversion into neurons in the dentate gyrus of the adult hippocampus. Using the mouse as a model system we test the hypothesis that social isolation followed by enrichment will lead to increased neurogenesis in the adult hippocampus compared to enrichment alone. Although our results show trends in favor of the above hypothesis in terms of overall cell proliferation and neurogenesis in the dentate gyrus, they failed to produce statistical significance. We believe that this may have been caused by the small sample size of our study, and that our findings warrant continued investigation. We intend to increase our sample size and continue with the study in order to conclusively confirm or reject our hypothesis.

Khrystofor Khokhlov CC'19: Chemistry

kk3106@columbia.edu

Faculty Mentor: Professor Colin Nuckolls, the Sheldon and Dorothy Buckler Professor of Material Science in the Department of Chemistry

Title: Toward an Efficient Synthesis of Helical Graphene Nanoribbons: Methoxylated PDIs as Novel and Versatile Building Blocks

Abstract: Organic photovoltaics (OPVs) have potential to become a widespread, low-cost energy source. However, although fullerene derivatives are used as electron acceptors in most of existing OPVs, their poor visible light absorption and complicated energy tuning limit power conversion efficiency (PCE). This has fueled new research to explore alternative electron acceptors that could overcome limitations of fullerenes while retaining sufficient solubility and appropriate energy level for popular donor materials (e.g. PTB7). One class of molecules that fulfill all these requirements is perylene diimides (PDIs), with PDI tetramer showing PCE surpassing that of fullerenes. However, currently regiopure 1,7 and 1,6-dibromo PDIs, two crucial building blocks for synthesis of PDI-based scaffolds, can only be obtained via HPLC separation, with largest loading of ~100 mg. Furthermore, presence of two reactive bromines in PDI molecule results in oligomerization and polymerization in Pd-catalyzed couplings; this dramatically complicates their synthetic application. Therefore, we set out to develop a synthetic pathway that would allow us to circumvent this bottleneck in supply of regiopure dibrominated PDIs. Fortunately, bromination in the bay area of PDI aromatic scaffold can by rendered regioselective by introduction of a directing group, similarly to bromination in benzene core. Methoxy group in the PDI bay region renders both first and second stages of bromination regioselective and is easily converted to triflate in two steps. This allowed us to develop a novel, scalable synthetic route towards PDI derivatives: molecules synthetically identical to 1,7 and 1,6 dibromo PDIs were obtained via bromination of 1-methoxy PDI. These molecules can be made on gram-scale in regiopure form while using only standard procedures for separation. Synthetic equivalence of developed molecules and dibromo PDIs is demonstrated via synthesis of a series of PDI oligomers, functionalized by triflate group in regioselective fashion.

Young Joon Kim CC'20: Biology

yk2611@columbia.edu

Faculty Mentor: Dr. Omar Abdel-Wahab, assistant member in the Human Oncology and Pathogenesis Program and attending physician on the Leukemia Service in the Department of Medicine, Memorial Sloan Kettering Cancer Center

Title: Synthetic Lethal and Convergent Biological Effects of Spliceosomal Gene Mutations

Abstract: Mutations affecting RNA splicing factors SF3B1, SRSF2, and U2AF1 constitute the most common class of genetic alterations in myeloid malignancies including myelodysplastic syndromes (MDS), chronic myelomonocytic leukemia (CMML) and acute myeloid leukemia (AML). These mutations exhibited strict heterozygosity that is functionally distinct from loss of function. Notably, splicing factor mutations are almost always mutually exclusive with each other.

To understand the functional basis for this mutual exclusivity, we generated mice co-expressing Srsf2P95H and Sf3b1K700E mutations in the hematopoietic stem and progenitor cells (HSPC) and assess their ability to contribute to hematopoiesis in vivo. Double-mutant HSPCs showed severe defects multi-lineage reconstitution relative to single-mutant or control HSPCs. Transcriptomic analyses revealed that gene expression changes caused by each mutation alone were largely distinct while double-mutant cells showed additional downregulation of key regulators of HSPCs including c-Mpl, Itga2b and Pbx1.

Interestingly, mutations in both splicing factors contributed to hyper-activation of NF-kB signaling via mis-splicing of signaling intermediates. SF3B1K700E mutation promoted mis-splicing of MAP3K7, which encodes a kinase that mediates NF-kB signaling. Similarly, SRSF2P95H mutation generated a novel truncated isoform of caspase-8 (CASP8 TR), which also induced upregulation of NF-kB signaling. These observations suggest that spliceosomal gene mutations have shared downstream effect and converge on the innate immune signaling pathway.

In summary, this study provides a biological explanation for the mutual exclusivity of splicing factor mutations in leukemia. These findings also open up new therapeutic opportunities as inhibition of MAP3K7 function in SF3B1-mutant cells or NF-kB signaling in splicing mutant cells may be effective against spliceosomal mutant cancers.

Suppawat Kongthong CC'18: Biology

s.kongthong@columbia.edu
Faculty Mentor: Daniel Mucida, associate professor of immunology, virology and microbiology, The Rockefeller University
Title: Dissecting the Role of the Commensal Microbiota in Intestinal T Cell Fate

Abstract: The gut mucosa is constantly exposed to a large number of various antigens – both harmless and pathogenic. As a result, maintaining a balance between pro- and anti-inflammatory responses is of key importance. Two subpopulations of intestinal lymphocytes, peripheral regulatory T cells (pTregs) and intraepithelial CD4+ T cells (CD4-IELs), play a role in suppressing immune responses and maintaining intestinal homeostasis. Conversion and expansion of these two subpopulations can be affected by the intestinal microbiota through antigen recognition via the T cell receptor (TCR). To investigate how microbiota influence lymphocyte fate and the TCR repertoire, we performed a single-cell analysis of the TCRαβ pairs on the Illumina MiSeq platform on intestinal lymphocytes from inducible fate-mapping reporter mice that enabled us to single-cell sort pTregs, Tregs that converted into CD4-IELs, and CD4-IELs. Our preliminary data show the TCR repertoire of pTregs and CD4-IELs in the intraepithelial compartment is restricted to a few expanded clones whereas pTregs in the mesenteric lymph nodes are as diverse (polyclonal) as expected. Our data suggest the intraepithelial pTregs and CD4-IELs are being activated by a restricted set of antigens. To screen for commensal bacteria that these TCRs recognize, we used an in vitro antigen presentation assay consisted of NFAT-GFP T cell hybridromas reconstituted with selected TCRs. When the hybridomas become activated, they express GFP under NFAT control. The

next step is to find which bacterial antigens these expanded TCRs recognize. This can shed light on how immunity against commensals is achieved while keeping intestinal immune homeostasis.

Ashley Koo CC'19: Computer Science

hk2827@columbia.edu

Faculty Mentor: Smaranda Muresan, research scientist in the Department of Computer Science and adjunct associate professor of computer science

Title: Argumentative Discourse Study: Exploring the Semantic Types of Argumentation in an Online Discussion Forum

Abstract: Columbia's Center for Learning Computer Systems aims to understand which arguments are persuasive and why. With the subreddit ChangeMyView (CMV) discussion posts as the dataset, the ultimate goal is to automate the detection and generation of persuasive arguments with Artificial Intelligence (AI). Motivation: My own project focuses on discourse study rather than persuasion. Provided that certain discussion topics seem more 'emotional' or 'logical' in nature, the original post may be expected to advance a similar type of argument. Since the title is treated as the main claim of the post, the ensuing claims and premises of the post substantiate the main claim and thus have a nontrivial relation with the title. This relation between the title (which summarizes the post) and the discourse pattern is investigated by considering the semantic types of claims and premises in the original post. Method: 38 CMV threads are annotated in terms of two semantic types for titles (interpretation and evaluation), three for premises (ethos, pathos, and logos) and three for claims (interpretation, rational evaluation, emotional evaluation). A frequency distribution table groups these premise types and the claim types together as dependent variables (y-axis) and the title types as independent variables (x-axis). Chi-squared test is used to compute the statistical significance between the two proportions. A separate test is run for each subgroup (premise and claim). Results & Discussion: Due likely to the small dataset, the difference in premise type counts between the title pair is statistically insignificant, with p-value of 0.4608 (N.B.: premises tend to support the immediately relevant claim rather than the larger main claim). There is almost a two-fold difference in the claim types counts for the title pairs (p-value = 0.0334). As expected, for evaluative titles, 'emotionally evaluative' claims appear 32% of the time while for interpretative titles it is 16%. Similarly, interpretative claims appear twice more frequently for interpretative titles than for evaluative titles. The findings suggest that given a larger dataset, and a more rigorous taxonomy to cover title types and discussion topics, we may be able to make substantive inferences about the semantic model of a given argumentative discourse.

Avik Laha CC'19: Physics; Computer Science

al3510@columbia.edu

Faculty Mentor: Greg Bryan, professor of astronomy

Title: Supernova Feedback in Cosmological Simulations

Abstract: We study various implementations of supernova feedback in the Enzo simulation for reproduction of realistic galaxy populations. In particular, we wish to reproduce the radial density

distributions of the dark matter in dwarf galaxies, which recent observations suggest is both less centrally located and more varied than was previously expected. In large-scale simulations, where galactic evolution is best studied, naive implementations of supernova feedback do not produce good results due to poor resolution of individual supernova events. We thus explore novel schemes, which have not yet been used in large-scale Enzo simulations, in an attempt to systematically quantify the efficacy of these schemes. Ultimately, this should indicate whether supernova feedback can indeed explain the deviation of observed distribution in radial density in dwarf galaxies from earlier expectations.

Hyunwook Lee CC'18: Biochemistry

hl2863@columbia.edu
Faculty Mentor: Thomas Ludwig, associate professor in the Department of Biology and Genetics, The Ohio State University College of Medicine
Title: Mechanisms and Phenotypes of Acquired Drug Resistance in PALB2-Deficient Tumor Cells

Abstract: It is well known that mutations in the mammalian genes encoding for homologous recombination proteins, such as BRCA1, BRCA2, and PALB2, can lead a higher risk of breast and ovarian cancer. BRCA1, BRCA2, and PALB2 are implicated in error-free repair of double strand breaks by homologous recombination. Mutations in either of any of these genes can thus cause cells to be hypersensitive to interstrand DNA crosslinking agents, such as cisplatin, mitomycin C, and poly(ADP-ribose) polymerase (PARP) inhibitors. These chemotherapeutic drugs can take advantage of the increased genetic vulnerability of cancer cells and prevent rapid cell division.

However, there are many instances where patients undergoing this type of therapy relapse and recurring tumors show resistance to these aforementioned chemotherapeutic agents. Some cases of resistance are believed to be caused by secondary BRCA1/2 mutations, which seem to restore the repair functions that BRCA1/2 serve. Still, there are other mechanisms that lead to drug resistance that are not well understood and those involving PALB2 are not well studied.

To begin investigating the mechanisms that give rise to this drug resistance, we exposed PALB2-deficient tumor cells to mitomycin C and olaparib, an FDA-approved PARP inhibitor, over a three week period to look for any emerging colonies that became resistant to the interstrand DNA crosslinking agents. Through immunofluorescence, resistant cells from these colonies were found to have elevated levels of RAD51, indicating that the resistant cells were actively performing DNA repair through homologous recombination. In addition, there were significantly fewer chromosome aberrations and exchanges in these resistant cells through karyotype analysis.

These results are extremely important in light of its clinical ramifications and understanding cancer caused by PALB2 mutations as well as BRCA1/2 mutations. The main form of chemotherapy for BRCA1-related triple-negative breast cancer is through these interstrand DNA crosslinking agents and PARP inhibitors. Understanding the nature and mechanism by which these tumor cells acquire drug resistance

can allow us to identify new targets for drug development to suppress the restoration of repair mechanisms.

Kai-Zhan Lee SEAS'19: Computer Science

kl2792@columbia.edu

Faculty Mentor: Julia Hirschberg, the Percy K. and Vida L. W. Hudson Professor of Computer Science and chair of the Department of Computer Science

Title: Data Filtering for Intelligibility in Deep Learning TTS Voices

Abstract: We train deep learning text-to-speech models using the University of Edinburgh's neural network generator, Merlin. We use the Macrophone corpus, a collection of 200,000 transcribed utterances (text and audio) recorded over telephone from 5,000 American speakers. Prior work with the Macrophone corpus has used HTS (Hidden-Markov-Model-based Speech Synthesis System) to synthesize speech. Because deep learning techniques have been steadily developing and improving, the objective of this work is to test the efficacy of the University of Edinburgh's deep learning speech synthesis model, Merlin relative to HTS, as well as to further develop training data selection methods for speech models. To discover the best clusters of data for training with Merlin, we merge utterances by speaker into single audio files and extracted standard acoustic features for each speaker. We experiment with various methods for calculating heuristic values for each utterance to select 2-, 4-, and 10-hour clusters of the best training data. We evaluate voices by computing word error rate (WER) between the IBM Watson automatic speech recognition API's predictions on synthesized audio.

The baseline model, trained on a random subset of 10 hours of utterances, achieved WER of 81.8%, and the best model generated using speakers specifically selected by feature extraction and combination into single heuristic values achieved WER of 42.9%, representing a 47.6% improvement from baseline. The best 2-hour model achieved a WER of 76.6%, representing a 44.3% improvement from the best 2-hour HTS model trained using the same dataset information. However, there was significant variance between voices trained on the same data; one voice trained on the same data had a WER that varied by 18.1%.

The most useful features for intelligibility were IBM Watson ASR word error rate, voiced-to-total frame ratio within the utterance, mean energy, and F0 mean average slope. Future work may include crowd-sourcing trained voices for human naturalness and intelligibility evaluation, more robust voice training to account for model initialization variance, and clustering by utterance instead of by speaker.

Emily Li CC'20

el2857@columbia.edu

Faculty Mentor: Julia Hirschberg, the Percy K. and Vida L. W. Hudson Professor of Computer Science and chair of the Department of Computer Science

Title: Identifying Audio Features that Characterize TTS Corpora to Improve Naturalness in Voice Synthesis

Abstract: The expansion of speech technology in the past few years has resulted in the development of intelligible and natural-sounding voices for high-resource languages (HRLs) such as Mandarin, English, and Japanese. These languages have extensive, high-quality Text-to-Speech (TTS) corpora collected from professional speakers. However, there are over 6000 languages spoken around the world which do not have these resources. Thus, our goal is to develop methods to synthesize similar, high-quality voices from "found" data, which refers to audio that is recorded for non-TTS purposes such as broadcast news and telephone speech. We are working on characterizing corpora created specifically for TTS in terms of pitch, energy, speaking rate, and articulation in order to filter "found" data for utterances that share similar characteristics to TTS corpora. We collect statistics for both TTS corpora and non-TTS corpora by finding mean and standard deviation of these features for all the utterances for a particular speaker. Thus far, our results are consistent with the assumption that TTS speakers generally have lower standard deviations across the aforementioned features than non-TTS speakers do. The next steps might be to compare the distributions by gender to identify the characteristics of the group that produces the most natural-sounding voices. We aim to differentiate different groups via clustering and other statistical models. It may also be fruitful to examine other aspects of these acoustic features, such as the rate of energy change, to see if there are additional ways groups can be characterized. Moreover, we are looking to develop a metric to rank speaker quality based on how natural voices trained on their data are. Ultimately, we hope to identify high quality speakers and the characteristics of their voices in order to inform the synthesis of voices from "found" data.

Noah Miller CC'19: Physics; Mathematics

nmm2166@columbia.edu
Faculty Mentor: Frederik Denef, professor of physics
Title: dS/CFT Correspondance

Abstract: The conjectured dS/CFT correspondence, unlike the AdS/CFT correspondance, is very poorly understood and might not even exist. This project explores the prospects of the holographic principle in de Sitter space.

Richard Nederlander CC'18: Astrophysics

rhn2105@columbia.edu
Faculty Mentor: Professor Keivan Stassun, professor of physics and astronomy, Vanderbilt University
Title: Radiation Characterization of a Silicon Avalanche Photodiode

Abstract: The Fisk/VU CubeSat Program's goal is to launch multiple small, low-cost satellites to an asteroid for data collection of the elemental abundance on the asteroid's surface. As cosmic radiation bombards the asteroid, the process of cosmic spallation breaks apart atoms on its surface (desired signal), and potentially the satellite itself (undesired signal). Not only do we want to differentiate between the two different signals, but we also want to characterize the element given the detection of the charged particles' emissions. To diminish as much noise as possible, we characterized the radiation

hardness of a Silicon Avalanche Photodiode to begin constructing a satellite that is resistant to radiation effects in space. Inside of our prototype satellite is the scintillator crystal known as strontium iodide doped with europium (SrI2(Eu)). In response to a gamma-ray hitting the crystal, it scintillates and creates many optical photons. Electrical pulses of photons were collected using a Silicon Avalanche Photodiode (Si-APD), which is directly beneath the crystal, and processed in Python. We read those pulses using an oscilloscope, which are then plotted in Python. These pulses, also known as Single Event Effects (SEE) measure Si-APD's sensitive to radiation. Specifically, 15 µCi of cesium-137 and 8.8 µCi of americium-241 were used to measure the Si-APD's radiation hardness in two different configurations and two different environments at both Fisk University and Vanderbilt University. We discuss the results of the radiation effects on the Si-APD to differentiate between noise influenced by increased levels of radiation and charged particle signals. Proposed future work will be to determine whether to continue using the Si-APD or test other types of APD's for better efficiency, as well as to filter the data to discern the type of charged particles being detected (i.e. protons, electrons, alpha particles, etc). We also plan to launch the satellite on a high-altitude balloon to analyze radiation levels at 35 kilometers above the Earth's surface, as well as verify the satellite's durability in those conditions

Aunoy Poddar CC'19: Biology

ap3441@columbia.edu
Faculty Mentors: Alice Heicklen, senior lecturer in the discipline of biological sciences
Jae Woo Lee, senior lecturer in the discipline of computer science
Title: Using ferroptosis to selectively eliminate multidrug resistant cancer cells

Abstract: Ferroptosis is an iron dependent, non-apoptotic form of programmed cell death dependent on the depletion of an essential tripeptide antioxidant, glutathione (GSH). To identify regulators of GSH metabolism, and perhaps better understand the regulation of ferroptosis, we conducted a fluorescence activated cell sorting (FACS)-based haploid genetic screen. Unexpectedly, we discovered the ATP-Binding Cassette Subfamily C Member 1 gene ABCC1 to negatively regulate the level of GSH inside the cell. ABCC1 encodes the ATP-dependent drug pump multidrug resistance protein 1 (MRP1), known to be overexpressed in multidrug resistant cancer phenotypes. We investigated a wide panel of cancer cell lines to observe the modulation of MRP1 and its relationship with GSH. Next, we proceeded to analyze the effects of genetic knockdown of MRP1 and lentiviral overexpression. We discovered that knockdown of MRP1 confers resistance, while overexpression of MRP1 sensitizes cells, to ferroptotic cell death. Finally, we screened 261 bioactive compounds to identify novel collaterally sensitive agents to target MRP1 overexpression, and found 10 compounds to selectively kill multidrug resistant cancer cells.

Manasa Prahlad CC'20: Biochemistry

mp3331@columbia.edu
Faculty Mentor: Lawrence Allen Chasin, the William R. Kenan, Jr. Professor of Biological Sciences
Title: Treating Duchenne Muscular Dystrophy by Altering the Splicing of pre-mRNA Transcripts Using Small Molecules

Abstract: Duchenne muscular dystrophy (DMD) is a genetic condition that is caused by a nonsense mutation in the gene for the dystrophin protein. It is possible to induce the skipping of certain exons in the pre-mRNA transcript. While the type of mutation that results in DMD can occur at multiple locations in the dystrophin gene, skipping Exon 51 can cause an in-frame shift that would in some cases mitigate the effect of the original mutation to produce a shorter, but still functional form of dystrophin. It is known from prior research that small molecules can bind to RNA and modify their secondary structure and therefore their function. We believe that such small molecules exist that can also bind to specific sites in pre-mRNA, and therefore influence splicing at these sites.

To test this, we are developing a mammalian cell model using Chinese hamster ovary cells. We constructed a plasmid containing Exon 51 from the human dystrophin gene and the puromycin resistance gene. The plasmid was transfected into the cells, and after incubation, their mRNA was extracted to verify that the transfection was successful and that the transfected DNA was being stably expressed. Next steps include testing the strength of the splicing sites at Exon 51 by using splicing inhibitors. Once the model is complete, it can be used to screen tens of thousands of molecules for efficacy in promoting exon skipping. After identifying several effective molecules, we plan to focus on improving these molecules' efficacy by creating and testing their derivatives.

Samantha Rhoads CC'18: Psychology; Sociology

skr2150@columbia.edu
Faculty Mentor: Valerie Purdie-Vaughns, associate professor of psychology and special advisor to the executive vice president for Arts and Sciences
Title: Gender in Negotiations: An Investigation of Network Centrality

Abstract: An analysis of paired negotiations between 84 MBA students revealed that men ranked higher in overall negotiation performance than did women, as measured by David's scores ($p<.01$). Moreover, a social network analysis showed that students' network centrality predicted their subsequent negotiation performance ranking using a variety of centrality measures (degree: $p<.01$; eigenvector: $p<.01$;betweenness: $p=.09$; and closeness: $p=.07$). Since past results show that women tend to occupy more peripheral positions in networks in male-dominated contexts (Lazega, 2001)—in that women are less likely to have reciprocated ties than are men—we hypothesized that social network position would explain the relationship between gender and negotiation performance. However, gender failed to predict network position—for neither betweenness ($p=.91$), closeness ($p=.44$), nor eigenvector centrality ($p=.31$), and only marginally for degree centrality ($p=.08$), removing the possibility of mediation. Gender and network centrality thus seem to be independent predictors of negotiation performance.

Daniela Riedlova CC'20: Neuroscience and Behavior

dr2888@columbia.edu
Faculty Mentor: Sarah Hansen, lecturer in the discipline of chemistry
Title: Metavisualization in Chemistry Learning

Abstract: Using eye-tracking and visual feedback to see whether viewing patterns change and/or student becomes more confident in their knowledge of chemistry. Experiment used was a precipitation experiment with different animations following it.

Jaewook Ryu CC'19: Biochemistry

jr3576@columbia.edu
Faculty Mentor: Ruben Gonzalez, professor of chemistry
Title: Sortase Based Strategy for Site-specific Fluorophore Labelling of Multi-component Biological Complexes

Abstract: Single-molecule techniques, such as single molecule Fluorescence Resonance Energy Transfer (smFRET), have provided profound mechanistic insights into biological systems which were unrecognizable by conventional ensemble-scale approaches. However, the application of smFRET is limited due to the difficulty of preparing site-specific, fluorophore labelled biomolecular components of multi-component biological complexes without compromising their native biochemical activities. To address this critical bottleneck, we propose a sortase-based labelling strategy of multi-component biological complexes coupled with multiplex genome engineering. Multiplex genome engineering enables the genomic insertion of sortase recognition sequence into a protein component, which, after expression, can be correctly incorporated and assembled in vivo. After purification of the multi-component complex, the sortase recognition sequence on this protein can be specifically labelled with fluorophore-conjugated peptides with compatible sequences by sortase activity. We demonstrate the feasibility of this approach by site-specific Cyanine 3 dye labelling of Escherichia coli ribosomal protein S19 by Streptococcus pyogenes sortase A, followed by smFRET analysis of translational initiation dynamics. Since this strategy is widely applicable in other multi-component biological systems, even those of other species, it will widen the scope of smFRET analyses, thereby providing unprecedented insights into how biological systems function.

Arjun Srivatsa CC'20

ass2186@columbia.edu
Faculty Mentor: Songtao Jia, associate professor of biological sciences
Title: The effect of Telomere Tethering in Fission Yeast on Homolog Pairing

Abstract: During Prophase I in Meiosis, homologous chromosomes (homologs) pair and undergo recombination. This process is essential in most eukaryotic organisms, and contributes to proper chromosomal segregation during meiosis. Additionally, recombination is a significant source of genetic variation. The process by which homologs pair, known as synapsis, involves the processes of movement and searching in the nucleus. Studying this pairing process is crucial because abnormalities in this process have been linked to infertility, miscarriages, and developmental disorders in humans. To study this pairing process, we use the fission yeast, Schizosaccharomyces pombe, as a model organism. In S. Pombe, homolog pairing is facilitated by a structure called the "bouquet". The bouquet consists of telomeres of chromosomes tethered to a structure called the Spindle Pole Body (SPB) on the nuclear

envelope. The nuclear envelope then undergoes movement, which is crucial for appropriate pairing. In our experiment, we sought to examine whether the tethering of chromosomes to the Spindle Pole Body (SPB) was crucial to appropriate pairing, and hence spore viability. To test this, we used a yeast strain with a telomere repeat and mated this with two different yeast strains; the first yeast strain was a healthy wild type strain and the second strain was another telomere repeat strand. If our hypothesis were correct, we would expect the pairing of the telomere repeat strain and wild type strain to have a low viability since the chromosomal pairs are misaligned with the telomere repeat strain having inserted base pairs. Additionally, we would expect the cross between the two telomere repeat strains to have a higher viability since both of these strains had additional base pairs inserted, their corresponding chromosomes would match better on the bouquet. Indeed, we saw this trend; the cross between the two telomere repeat strands had a significantly higher spore survival rate than the cross between the telomere repeat strand and wild type strain. In the second part of the experiment we sought to determine whether the cause of the increased spore viability was actually the difference in tethering between the telomeres and SPB, and not some extraneous factor. To do this, we made two new strains of yeast, each tagged with Green Fluorescent Protein (GFP). The first strain had a GFP marker inserted at the site of the telomere repeat. The other strain had GFP inserted on a different chromosome from the telomere repeat; this chromosome was unmodified. If our hypothesis is correct, we would expect to observe defects in the division of the chromosome with the telomere repeat and observe proper division of the unmodified chromosome. We have yet to perform this part of the experiment.

Maya Talukdar CC'20: Computer Science; Biology

mt3134@columbia.edu
Faculty Mentor: Harmen Bussemaker, professor of biological sciences and systems biology
Title: Enhanced Biological Annotation of Single-Cell RNA-Sequencing Datasets via Singular Value Decomposition

Abstract: The recent advent of single-cell RNA-sequencing (scRNA-Seq) technology has revolutionized the scientific study of cellular heterogeneity and rare cell subtypes. However, the inherent stochasticity of background gene expression as well as the large and multivariable nature of these gene expression datasets have presented complex new computational and statistical questions of how to best facilitate the identification of previously unknown and scientifically valid cell subtypes while also controlling for the significant amounts of noise found in such datasets. This is particularly complicated by the widespread use of dimensionality reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE) and principal components analysis (PCA) that contribute to a "cluster first, interpret second" mentality that forces researchers to try and assign biological meaning to groups of cells clustered together via purely statistical means. Furthermore, though PCA is mathematically based in the singular value decomposition (SVD), a direct analysis of a gene expression dataset's SVD is often eschewed in the biological analysis of sCRNA-seq data. Here, I utilize the singular value decomposition, both uniquely and in its relation to PCA, to simultaneously reduce the dimensions of such scRNA-seq gene expression datasets and to biologically annotate these otherwise abstract reduced dimensions in order to better understand the in vivo significance of cell clusters first identified through PCA.

Michelle Vancura CC'19: Biochemistry

mv2433@columbia.edu
Faculty Mentor: Laura Kaufman, professor of chemistry
Title: Gelatin-Methacrylate (Gel-MA) Hydrogel as a Model to Study Breast Cancer Cell Invasion

Abstract: Breast cancer is the most common cancer in women worldwide, with fatalities primarily resulting from metastasis to other vital organs. To become metastatic, cells must invade the dense extracellular matrix, which is composed of collagen-I, surrounding the tumor. Invasion of breast cancer in vivo is often studied in collagen-I matrices; however, the mechanical stiffness of collagen-I matrices is typically ~10 Pa, which is below that of both healthy and cancerous breast tissue. Furthermore, the ability to vary the stiffness of collagen gels without also changing pore size, a potential co-variable that can dominate breast cancer cell invasive speed and mode, is limited. Gelatin-methacrylate (Gel-MA) is an alternative biopolymer matrix that may not have the same complications as collagen-I matrices. Initial work has shown that the full range of stiffness, from 10 Pa to 100 kPa, can be achieved with limited influence on gel pore size. Here, we show optimization of Gel-MA synthesis and gel preparation and characterization of gel properties for future use in breast cancer cell invasion studies.

Justin Whitehouse CC'19: Mathematics; Computer Science

jaw2228@columbia.edu
Faculty Mentor: Kyle Chard, research specialist at the Computation Institute, University of Chicago
Title: Rich Context Search: Using Machine Learning Techniques to Explore Research Publications

Abstract: While there is an ever increasing number of academic publications available to both public and private entities, it can be difficult to find papers related to a given topic of interest or dataset. In general, the most reliable universal methods for discovering relevant research publications consist of searching key words and phrases along with using filters in various databases. Some specific journals do have taxonomies to classify papers into subjects, but these taxonomies are generated by professionals, and thus require large investments of time and effort to generate. We propose a system that, given a set of documents, uses a variety of unsupervised and supervised learning techniques to tag documents and datasets with key terms (such as location and keywords), relate documents to relevant datasets, and generate a hierarchical taxonomy consisting of wiki-labels that captures the diversity of topics found within the documents. We tested our proposed structure on the Journal of Policy Analysis and Management (JPAM), Journal of Public Administration Research and Theory (JPART), and the Public Administration Review (PAR), and created a search site around the automatic tagging processes. On the site, we show the tags alongside the text of the document, and allow users to mark the tags a relevant or irrelevant, and even propose new tags for supervised learning methods. Furthermore, users can upload relevant documents of their choosing with tags they apply themselves to improve the learning process. While we are still awaiting more results from interactions with the site, we are observing that our unsupervised tags and classification results are similar in quality to those of other journals that use manual tagging and classification. This process, overall, provides a novel and useful way for exploring large amounts of data within a given field.

James Xu CC'21: Computer Science

jrx2000@columbia.edu

Faculty Mentor: Alexander Gorbach, chief of the Infrared Imaging and Thermometry Unit within the Biomedical Engineering and Physical Science Shared Resource Program at NIH, National Institute of Biomedical Imaging and Bioengineering

Title: Using Mobile Phones to Monitor Nailfold Capillary Density in Humans

Abstract: Nailfold capillaroscopy plays an important role in the diagnosis and prognosis of rheumatic diseases, such as myositis, psoriasis, and systematic sclerosis. However, most capillaroscopic devices are still traditional optical microscopes, which are not easily portable. For this project, both hardware and software for a portable capillaroscope were designed and built. For hardware, a case and lens holder were 3D-printed. The iPhone 7 Plus was chosen to be the mobile phone for use because it can capture high-resolution images. For software, an app that runs on the iOS operating system was developed. The computer vision library OpenCV 3.2 was integrated into the app to allow it to recognize the capillaries in an image. Then, a smoothed z-score algorithm was used to filter the pixel intensity data that was acquired from the image. Finally, the library iOS-Charts helped visualize the data by plotting an intensity profile for analysis. A calibration tool was also incorporated to enhance the z-score algorithm and account for the range of lighting that could be present between images. When attached to a mobile phone, the current prototype of the device can monitor nailfold capillaries and quantify capillary density, one of the most commonly measured capillary parameters. Analysis of capillary density usually requires importing data to a computer. However, through this project, it was determined semi-automatic data processing on the mobile phone could provide a more efficient alternative.

Katherine Xu SEAS'20: Chemical Engineering

kex2000@columbia.edu

Faculty Mentor: Corey McMillan, research assistant professor of neurology, University of Pennsylvania Perelman School of Medicine

Title: Individualized Cortical Thickness Heatmaps for Improved Clinical Care in Neurodegenerative Disorders

Abstract: Neurodegenerative disorders are characterized by gradual atrophy in the brain, resulting in a decline in behavior, motor skills, and/or language. The disease progression often includes social impairments and difficulty with executive functioning, which necessitates capable caregivers. As part of treatment, patients are routinely given MRIs to assess brain condition. In clinic visits, it is often difficult to communicate to patients and their caregivers the results from MRI images due to their 3D nature and subtle differences to an untrained eye. Our objective was to develop a method to automatically render standardized images that would better communicate to patients their MRI results. We also aimed to offer caregivers and patients a better understanding of their disease progression and some explanation for symptoms. We developed a computer program that utilizes the open source visualization tool, Connectome Workbench, to display MRI values on a color scale. Using T1-weighted MRI scans from the Penn Frontotemporal Degeneration Center, cortical thickness measurements were calculated using

Advanced Normalization Tools (ANTs). To provide patients and clinicians with understandable results, we calculated z-scores of patients' cortical thickness values against healthy controls for every voxel in the MRI. Cortical thickness is an important biomarker that correlates with gray matter health and the number of neurons in the brain. The patient's z-scores, on a color scale were imposed onto a brain render such that significant brain atrophy could be easily visualized. Various disease phenotypes were tested and the resulting renders were visually powerful in identifying the characteristic areas of atrophy. Furthermore, the z-score heat maps could display disease progression and help evaluate prognosis. We hope to implement these individualized heatmaps for clinic use to improve MRI assessment for clinicians and understanding of the disorder for the patient and caregiver.

Helen Zhang CC'20: Biology

helen.zhang@columbia.edu
Faculty Mentor: Charles Zuker, professor of biochemistry and molecular biophysics and of neuroscience
Title: The Molecular Mechanisms of the Salt Taste Behavior of Salt-Deprived Animals

Abstract: The mammalian taste system senses five basic taste qualities: sweet, bitter, sour, salty, and umami. Sweet and umami are innately attractive, while bitter and sour are innately repulsive. Salt, however, can be either attractive or repulsive depending on the concentration of salt and the physiological condition of the taster. Interestingly, salt-deprived animals are more attracted to salt solutions than control animals. This study aimed to elucidate the molecular mechanisms that produce this behavior. Some hormones, such as the endothelins (ETs), are intimately involved in the regulation of salt homeostasis. The activation of endothelin receptor type B by ET-1 has been shown to inhibit the epithelial sodium channel (ENaC) in the kidney collecting duct. Since ENaC channels expressed by taste receptor cells (TRCs) in the tongue mediate salt taste, we hypothesized that the endothelins (ETs) inhibit ENaC activity in the tongue as it does in the kidney to control salt consumption. To test this hypothesis, we first examined the expression levels of the ETs in the tongues of salt-deprived and control animals using immunohistochemistry. We then recorded the sodium tastant-induced action potentials in the chorda tympani nerve, which innervate and carry taste information from the TRCs to the geniculate ganglion neurons, of control animals and animals treated with ETs. This study will contribute to our understanding of the molecular mechanisms underlying animal behavior after salt deprivation.

Catherine Zhang CC'18: Biology; Environmental Science

crz2111@columbia.edu
Faculty Mentor: Alberto Ciccia, assistant professor of genetics and development
Title: Role of SMARCAL1 in Protection of Stalled Replication Forks in BRCA1 Mutated IOSE

Abstract: BRCA1 is a tumor suppressor gene, essential in DNA repair, homologous recombination, and cell cycle control. BRCA1 protects DNA replication forks arrested upon replication stress from nuclease degradation. The mechanism by which fork degradation results in genomic instability such as chromosomal aberrations is not understood. An assay to study the formation of anaphase bridges during mitosis was developed to investigate this source of genomic instability. The Ciccia Lab has

recently identified SMARCAL1 and its related protein ZRANB3 as key DNA translocases that protect and allow restart of stalled replication forks. Depletion of SMARCAL1 rescues protection of stalled forks in BRCA1-deficient breast cancer cells. However, the mechanisms by which BRCA1 heterozygous mutations predispose to ovarian cancer are currently unknown. It was found that SMARCLA1 depletion could similarly rescue protection of stalled forks in patient ovarian cancer cells carrying BRCA1 heterozygous mutations.