THE UCLA Undergraduate Science Journal

VOLUME 36 | SPRING 2023

Established 1986.
The UCLA Undergraduate Science Journal is an officially recognized student organization at UCLA.
The UCLA Undergraduate Science Journal is entirely managed and produced by students.
Its contributions are the original work of UCLA undergraduates.

FRONT COVER ILLUSTRATION BY FATIMAH ISMAIL
INSIDE COVER IMAGE COURTESY OF WILLY ARISKY

2022-2023 USJ STAFF

Editor-in-Chief & Managing Editor of Review
Aishwarya Atmakuri

Asst. Managing Editors of Review
Caroline Sha
Cole Woulbroun
Sophia Isabel S Pulido
Rachel Ma
Rebekah Jin
Jordan Lin

Editor-in-Chief & Managing Editor of Editorial
Emily Lin

Asst. Managing Editors of Editorial
Caitlyn Brahim
Zihui (Hailey) Li
Sohan Talluri
Martin Sevcik

Editor-in-Chief & Managing Editor of Layout
Elise Tran

Asst. Managing Editors of Layout
Dashrit Pandher
Chelsea Lai

Review Board
Aadil Rehan
Agamroop Kaur
Aki Takahashi
Caden Chow
David Mastro
Eliana Bohn
Emma Villinski
Ethan Hung
Haley Xiaohe Zhang
Jessica Goeij
Malvika Iyer
Marissa Li
Melody Jiang
Miki Matsuoka
Natalie Tsubamoto
Nathan Joshua
Nathan Ou
Nathaniel Chen
Nyah Zhang
Oliver Wang
Raphael Low
Roy Zhao
Ryan Wong
Wasila Sun
Yuelai Feng

Editorial Board
Andrew Wang
Aniket Das
Bethany Woo
Brendan Sam
Brynn Beatty
Charlotte Chen
Chloe Nelson-Torakawa
Evelynn Chen
Isabel Angres
Kyle Nguyen-Ngo
Lily Zello
Maanasi Dhiraj Narayan
Nhi Pham
Ruthy Shin
Timothy Liu
Ying yin (Lina) Zhu

Layout Board
Ahmad Ismail
Christine Zhang
Christy Ma
Crystal Woo
Truman Ma

Faculty Advisor
Jorge A. Avila, Ph.D.

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

A Letter from the Editors

Dear Reader,

We are incredibly proud to present to you the 36th Edition of the UCLA Undergraduate Science Journal. We are honored to be a part of this scientific collaboration and we are confident you will see these values exemplified in this year's volume of the USJ. The entirety of this issue is the direct result of undergraduate student effort, from the outstanding manuscripts to the peer review, editing, and layout of the journal. We are truly impressed with the caliber of research conducted by this year's undergraduate authors, as well as the tireless diligence and dedication of the staff members in our Review, Editorial, and Layout Boards. This volume of the USJ includes both original research and review articles, ranging from studying hummingbird foraging behavior to modeling electron density. The articles featured in this journal may only offer a small glimpse into the breadth of research that undergraduates undertake at UCLA, but they are a true representation of the ingenuity and diversity of our students.

Our first year in-person since 2020, the USJ staff has seen many changes. We have witnessed changes in leadership structure, board expansion, and a new faculty advisor. Historically, the USJ has had two Editors-in-Chief (EICs) who oversee the USJ logistics and ensure all boards are on track to publish the complete journal on time and three Managing Editors (MEs) who each oversee one of the three boards of the USJ. However, this past year, the leadership structure of the USJ staff has been unconventional. In lieu of the traditional two EICs, our three MEs of the Review, Editorial, and Layout boards have stepped up and filled both the EIC and ME positions, without previous experience in either role. We also got a new faculty advisor, Dr. Jorge Avila, who has helped us navigate uncharted territory.

Despite these changes in the USJ leadership, we did not let it dissuade us from our organization's goals. Using the new structure to our advantage, we were all able to learn together, support one another, and increase communication between the boards, resulting in a more cohesive USJ community and what we think is our best issue yet. In addition, Dr. Avila has been extremely supportive and encouraging of our new ideas to improve USJ, including completely reworking our previous timeline. We are so excited to be publishing the USJ during the school year in time for UCLA's Undergraduate Research Week, far earlier than it has been published in previous years.

This journal would not be possible without the labor and love of our entire team. We would like to especially commend the passion and dedication of all of our USJ staff members who worked with us and made the journal what it was, as we navigated our new dual-responsibilities as EICs and MEs. A huge thank you goes to our assistant managing editors for the amazing work and long hours you have all dedicated to the USJ. Of course, we would like to express our most heartfelt gratitude to our faculty advisor, Dr. Jorge Avila, who supported us throughout the entire publication process and without whom the USJ would not exist. Additionally, we sincerely thank Dr. Tama Hasson, director of the Undergraduate Research Center – Sciences (URC-Sciences) who provides pivotal support to publish our journal, the UCLA Clinical and Translational Science Institute for their generous continued support of our annual issue, and Alec Hernandez, Program Coordinator of the URC-Sciences. Finally, we would like to thank Colornet Press for helping us produce this beautiful journal.

We hope you are inspired while reading about the science featured in this journal, in fields that touch many aspects of our daily lives. This journal stands as a testament to the sheer courage and determination of our UCLA journal staff and undergraduate authors—we could not be more honored to share it with you.

Warm regards,
Aishwarya Atmakuri, Emily Lin, and Elise Tran

TABLE OF CONTENTS

How Loss of Social Status May Affect the Brain

Aadil Rehan

We all know how it feels to be embarrassed in public. Your face grows hot; you might start sweating or shaking. Recent findings from Fan et al. have determined that something similar in mice – sudden loss of social status – also affects neural circuits. When two mice are placed in a tube in which only one can pass, one "dominant" mouse will forcibly pass the other to get to the other side. When researchers rigged the competition by suddenly blocking the dominant mouse's exit (thereby "losing" to the subordinate mouse), it resulted in increased stress and decreased reward pathway activation in the formerly-dominant mouse – a pattern closely resembling depression in humans. But all is not lost for these mice, as stimulating a particular neural circuit restores activity of the reward pathway and social status. These mice could be valuable test subjects for medications and behavioral treatments for patients with depression.

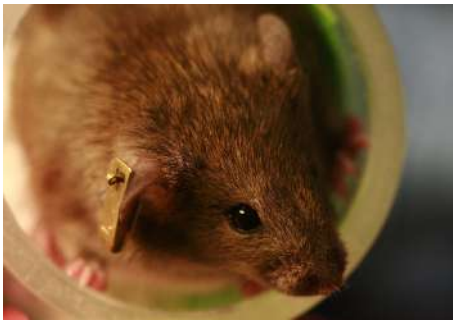


Image: Mice are commonly used as model organisms for a wide variety of laboratory experiments.
Reference: Z. Fan, et al., Neural mechanism underlying depressive-like state associated with social status loss. Cell. 186, 560-576 (2023). doi: 10.1016/j.cell.2022.12.033.
Image Source: "Lab mouse mg 3308" by Wikipedia user Rama. Licensed under CC BY-SA 2.0 FR. https://commons.wikimedia.org/wiki/File:Lab_mouse_mg_3308.jpg

Grasslands, More Than a Place for a Picnic

Brendan Sam

Covering more than a third of the Earth's landmass, grasslands provide forage for livestock and wildlife, sequester carbon from the atmosphere, and support the growing population of people worldwide. As grasslands are sensitive to rising temperatures and changing precipitation patterns associated with global climate change, Havrilla and colleagues developed species distribution models (SDM) of C3 (cool-season) and C4 (warm-season) grass species to predict the long-term impacts of climate change on grasslands in the western USA. The study found that C3 abundance declined in regions experiencing temperature increases while C4 species were projected to increase across the western USA, especially in higher-latitude regions. Factors such as temperature, precipitation, and soil characteristics were also found to influence grass species, particularly in restricting C4 species in low-altitude, arid locations. The researchers intend to use integrated SDMs for predicting future climate changes to make long-term resource management strategies for grassland conservation, restoration, and predictability.

Image: View of the Santa Monica Mountains taken at Point Mugu State Park in the Western United States.
Reference: C. A. Havrilla et al., Divergent Climate impacts on C3 versus C4 grasses imply widespread 21st century shifts in grassland functional composition. Diversity and Distributions. 29, 373-394 (2023). doi: 10.1111/ddi.13669
Image Source: Photo taken by Brendan Sam

A Fun Guide to Fungi Communication: Using Electrical Signals as Words

Caroline Sha

While changes in electrical potential are traditionally associated with human nervous system communication, a recent study has found evidence that suggests fungi also utilize electrical potential spikes to communicate. They do so through an underground network of branching filaments which may relay important information like injury sites and food sources. Similar to human words, these spikes differ in size, duration, and incidence. To determine whether these variations could indicate communication, the study author used an algorithm to analyze the linguistics of these patterns. The program found that there was an average of 4.7 spikes per cluster, comparable to the average English word length of 4.8. They also found that fungi have up to 50 unique groups of spikes. Though the author cautions against hastily making direct connections to human speech, this study provides insight into whether mushrooms can actually "talk" to each other—a concept that is still relatively unknown.

Image: Fungi (such as the Yellow-orange Fly Agaric pictured) may communicate with each other via electrical signals sent by underground filaments.
Reference: Andrew Adamatzky, Language of fungi derived from their electrical spiking activity. R. Soc. Open Sci. 9, 211962 (2022).
Image Source: "Fearless Fungi" by FLICKR user Distant Hill Gardens and Nature Trail. Licensed under CC BY-NC-SA 2.0. https://www.flickr.com/photos/59898141@N06/8196158070.

AI Identifies Novel Antimicrobials From Human Gut Microbiome Data

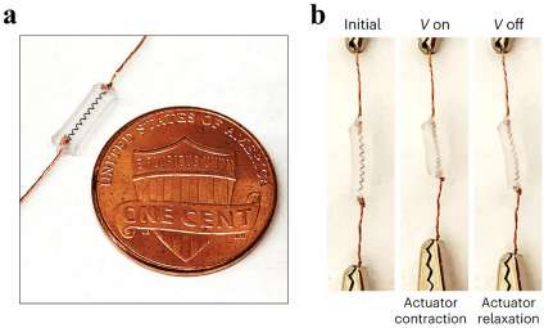
Sohan Talluri

Multidrug-resistant bacteria, or bacteria that cannot be killed by a broad range of antibiotics, cause an estimated 5 million deaths per year worldwide. Recently, researchers at the Chinese Academy of Sciences have brought artificial intelligence (AI) into the battle against these pathogens. Utilizing human gut microbiome data, the AI model identified over two thousand candidate antimicrobial peptides, from which researchers synthesized the 216 most promising sequences. The vast majority of these candidates were novel with less than 40% similarity to previously characterized peptides, and 84% of the synthesized peptides displayed antibacterial activity in vitro. Furthermore, in a mouse model of multidrug-resistant bacterial pneumonia, the three top candidates lowered bacterial load tenfold and facilitated significantly faster recovery, demonstrating the safety and efficacy of these novel peptides. Though clinical trials are necessary before patient application, AI could become a rapid, high-throughput drug discovery tool in our ever-evolving fight against antibiotic resistance.

Image: Multidrug-resistant Klebsiella pneumoniae bacteria used in mouse experiments.
Reference: Y. Ma et al., Identification of antimicrobial peptides from the human gut microbiome using deep learning. Nat. Biotechnol. 40, 921–931 (2022). doi:10.1038/s41587-022-01226-0
Image Source: Klebsiella pneumoniae Bacteria, by National Institute of Allergy and Infectious Diseases, NIH. https://www.flickr.com/photos/nihgov/24463154650. Available under CC BY-NC 2.0.

Combining Materials and Medicine: Creating Unique Biomedical Devices

Emma Villinski



Researchers at Harvard University have developed a biomedical device that employs mechanotransduction as a method to regenerate and rehabilitate muscle in patients that experience muscle atrophy. Mechanotransduction is a process that converts physical signals into changes in cell function. Stimulating muscle through stretching and compression has the potential to alleviate muscle atrophy; however, this had not been previously explored beyond massage therapy. The biomedical device, named MAGENTA, employs a shape memory alloy (SMA) to facilitate contraction. When a voltage is run through the alloy, it contracts into its original shape, a tightly wound spring. Without a voltage, it expands and stretches out. MAGENTA is attached directly to muscle tissue by a hydrogel-based adhesive, so that the muscle expands or contracts with the device. MAGENTA shows that applying materials with unique properties to medicine can create innovative biomedical devices and advance opportunities for treatment worldwide.

Image: (a) an example of MAGENTA, (b) contraction and relaxation of MAGENTA when voltage is applied.
Reference: Nam, S. et al. Active tissue adhesive activates mechanosensors and prevents muscle atrophy. Nat. Mater. 22, 249–259 (2023).doi:10.1038/s41563-022-01396-x
Image Source: Nam, S. et al. Active tissue adhesive activates mechanosensors and prevents muscle atrophy. Figure 1. Used with permission from author.

Why the Salton Sea Is Releasing Toxic Dust into Southern California

Oliver Wang

The Salton Sea, a large inland lake in Southern California formed from the diversion of the Colorado River for agriculture, has been rapidly drying up in the last few decades. This has led to toxic chemicals in the lakebed becoming exposed, turning into airborne dust and threatening the health of the surrounding land, wildlife, and residents. In October 2022, researchers from the University of California, Riverside identified the cause of this change as a decrease in water inflow from the Colorado River. The identification of this cause, as opposed to other proposed culprits such as increased temperatures or shifts in agricultural practices, helps environmental authorities pinpoint measures to preserve the ecological balance. The researchers suggest that if the Salton Sea—currently considered an independent system—was included in the Colorado River watershed, more resources could be allocated to restoring the health of this ecosystem.

Image: The Salton Sea's waters, a haven for migratory birds, are at risk of becoming depleted.
Reference: J.S. Acero Triana, H. Ajami, Identifying Major Hydrologic Change Drivers in a Highly Managed Transboundary Endorheic Basin: Integrating Hydro-Ecological Models and Time Series Data Mining Techniques. Water Resour. Res. 58, e2022WR032281 (2022). doi:10.1029/2022WR032281
Image Source: Photo taken by Sylvia Zhang

"DishBrain"— Neurons in a Dish Learned How to Play Games

Yuelai Feng

Imagine a group of neurons in a dish playing a video game—yes, you read that right. Recently, scientists from Cortical Labs in Australia have successfully developed DishBrain, creating a living neural network in a dish that can learn and play the classic 1972 Atari video game Pong. By growing cortical neurons on a grid of electrodes, the researchers trained the network of cells to perform tasks by stimulating them with electrophysiological signals and providing feedback based on their electrical activity. In just a few minutes, the neurons became proficient at the game, demonstrating their ability to learn and adapt. DishBrain has brought new insights into today's understanding of the human brain and intelligence, and it may offer exciting implications for the development of new drugs and better AI. The ability of neurons to adapt and learn, even in an artificial setting, highlights the incredible potential of biological systems.

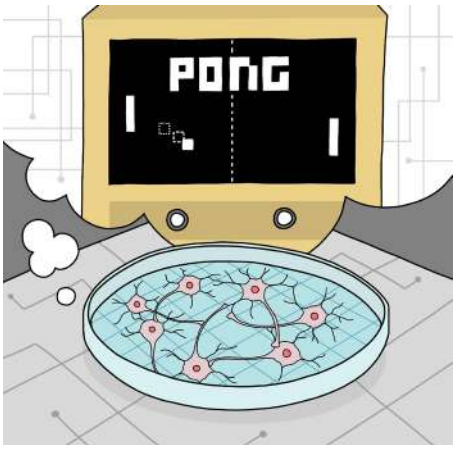
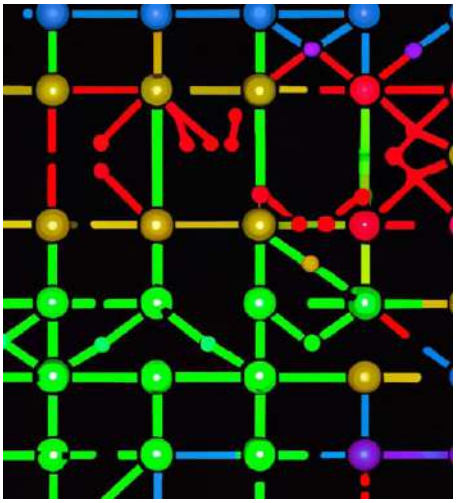


Image: Scientists from Cortical Labs developed DishBrain, which successfully trained a dish of neurons how to play the arcade game Pong.
Reference: B. J. Kagan et al., In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. Neuron. 110, 3952–3969 (2022), doi:10.1016/j.neuron.2022.09.001.
Image Source: Digital Illustration by Yuelai Feng

Surface Code for Suppressing Error Probability in Quantum Computers

Raphael Low



A large hurdle in realizing effective quantum computers is the strict demand on reducing error rates. A technical paradox seemingly arises where additional qubits are required to provide error correction, but introducing more qubits simultaneously increases the error rates of the setup. Google's Quantum AI has found a new way of converting physical qubits (e.g. superconducting qubits) to logical qubits (i.e. theoretical qubits) via surface codes, such that the error correction improves exponentially while the error probability increases at a linear rate when more physical qubits are added to the system. This breakthrough allows for scaling quantum systems while keeping the logical error density roughly constant with network size, with a caveat that the total noise of the system stays within a threshold. This prompts the possibility of large-scale effective quantum computers, which could realize algorithms such as Shor's or Grover's algorithm for cybersecurity and unstructured search respectively.

Image: AI-driven artistic interpretation of a quantum circuit.
Reference: Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. Nature 614, 676–681 (2023). doi:10.1038/s41586-022-05434-1
Image Source: Open-source creation by NightCafe AI using the DALL-E 2 Algorithm, text prompt "network of 5 hexagonal shaped qubits highlighted in pink shaded outline" and "Nightcafe" Preset style, with no seed image and 2 iterative generations.

AI Advances the Human Microbiome Project through Machine Learning Algorithms

Chloe Nelson

Often referred to as the second human genome project due to its importance in understanding health, the human microbiome consists of all the microorganisms within the body. While many are beneficial, some cause infectious diseases and influence cancer risk and immune health. Scientists are using machine learning to study the interactions between human and microbial proteins. Sequence-based algorithms, like Convolutional Neural Networks (CNN), analyze amino acid sequences to predict protein-protein interactions (PPI), while location-preserving algorithms, like Rotation Forest (RoF), use component analysis theory to analyze protein structure at the atomic level. Sequence-based algorithms have already been used to study infectious diseases, identifying 1,326 human proteins that the SARS CoV-2 virus targets when spreading infection. While these algorithms are limited by current knowledge of the microbiome, machine learning can help scientists better understand PPI and biological processes, aiding in the study of infectious diseases and the development of microbiome-based medications.

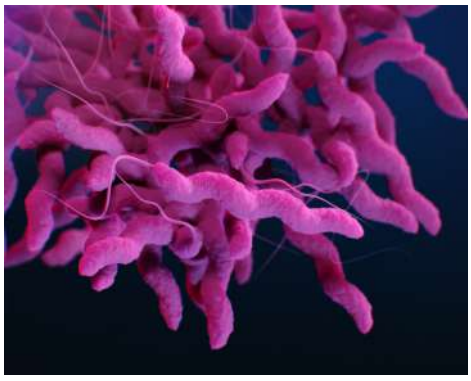


Image: Illustration of Campylobacter, drug-resistant bacteria that can colonize the intestinal microbiome.
Reference: H. Lim, Artificial intelligence approaches to human-microbiome protein-protein interactions. Curr Opin Struct Biol. 73, 102328 (2022). doi:10.1016/j.sbi.2022.102328
Image Source: Available under public domain. Medical illustration presented in the Centers for Disease Control and Prevention (CDC) publication entitled, Antibiotic Resistance Threats in the United States, 2019 (AR Threats Report). https://www.cdc.gov/media/subtopic/images.htm

How Life May Have Originated From Space Rocks

Ryan Wong

Scientists have long tried to explain the mysterious origins of life ever since the Miller-Urey experiment, which demonstrated that amino acids, the fundamental building blocks of life, could be synthesized under early-Earth conditions. Recent research from Japan proposes an alternative hypothesis: meteorites. The researchers suggest that gamma rays emitted from the radioactive decay of 26Al in star-forming regions could have provided the energy necessary to induce a chemical reaction that converted organic and nitrogenous compounds found in meteorites, such as ammonia or methanol, into basic amino acids, including alanine. It just so happens that carbonaceous chondrites—the oldest of the universe's asteroids, rich in water, carbon, nitrogen, and oxygen—provide the perfect reaction site. In fact, samples of carbonaceous chondrites brought back from space already contain some of these macromolecules produced by the researchers, including amino acids, lending evidence to their hypothesis. So, perhaps life did originate from outer space.

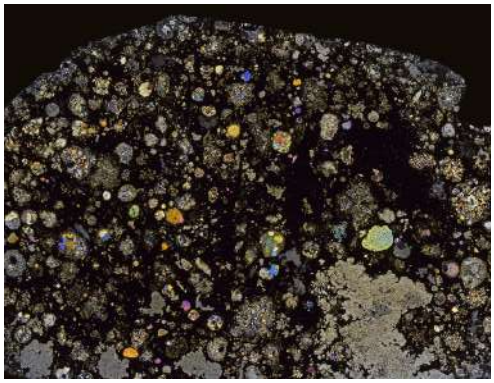
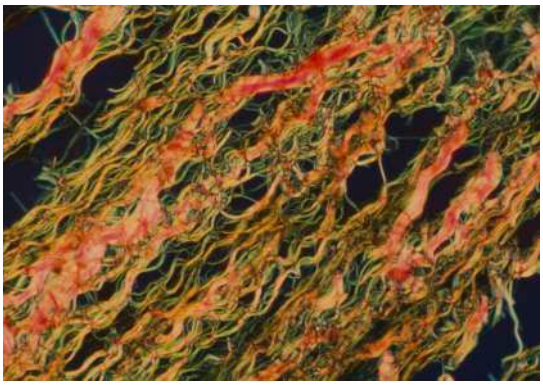


Image: Thin Section from Allende, the largest carbonaceous chondrite to impact Earth
Reference: Y. Kebukawa, S. Asano, A. Tani, I. Yoda, K. Kobayashi, Gamma-ray-induced amino acid formation in aqueous small bodies in the early solar system. ACS Central Science. 8, 1664–1671 (2022), doi:10.1021/acscentsci.2c00588.
Image Source: Photograph by Peter Marmet. Used with permission. http://www.thinsections.ch/ts/HOME.html

Collagen Expression in the Efficiency of Dermal Wound Healing Processes

Lina Zhu



The healing process of a wound involves cells protecting against infection, lowering inflammation, refilling the wound, and generating collagen used to provide structure. However, excess collagen build-up can lead to thicker, raised, and less elastic scars surrounding the skin. Researchers at the University of California, Los Angeles recognized that the timing of wounds within the circadian rhythm affects the behavior of the healing cells, leading to differences in the rates of wound healing. A recent study tested five FDA-approved compounds that suppress the Neuronal PAS domain 2 (Npas2) circadian gene used to regulate the body's natural rhythm. The compounds were applied to scratched samples of adult human dermal fibroblasts. Two of the five compounds successfully modulated cell migration and prevented excess collagen synthesis, thus, speeding up the healing process. This work serves as a foundation for future investigations involving collagen homeostasis and accelerated wound healing.

Image: Collagen fibers of varying thickness in scar tissue
Reference: A. Clements et al., In vitro assessment of Neuronal PAS domain 2 mitigating compounds for scarless wound healing. Front Med. 9, 1-12 (2023). doi:10.3389/fmed.2022.1014763
Image Source: "Scar collagen" by Wikimedia Commons user PW31. Licensed under CC BY 4.0. https://creativecommons.org/licenses/by/4.0

Smartphones or Social Connections? Can't Have the Best of Both Worlds

Dashrit Pandher

"It's because you're always on your phone." Every smartphone user has heard this when complaining about feeling unwell. Researchers examined the correlation between screen time and psychological well-being based on social connectedness. The displacement hypothesis states that smartphone use replaces socially-rewarding activities. The interference hypothesis states that smartphone use during socialization makes social activities unsatisfying. Data collected showed individuals reported a poor emotional state when socialization was low but screen time was high, indicating a positive relationship between emotional state and socialization. Individuals who used their smartphones while engaging in social interactions found them unrewarding, affirming the interference hypothesis. Ultimately, a "vicious" cycle was found. Increased screen time led to poor social connectedness and psychological well-being, which led people to seek socialization through their smartphones and further increase their screen time. Smartphones are revolutionary but pervasive devices, so the impacts they have on users' psychological and social well-being must be further studied.

Image: A collage of individuals with many having their faces blocked by their smartphone.
Reference: C. Anderl, M. K. Hofer, & F. S. Chen, Directly-measured smartphone screen time predicts well-being and feelings of social connectedness. J. Soc. Pers. Relat. 40, (2023). doi:10.1177/02654075231158300
Image Source: Created by Dashrit Pandher

Do Spiders Dream? Signs of REM Sleep in Jumping Spiders

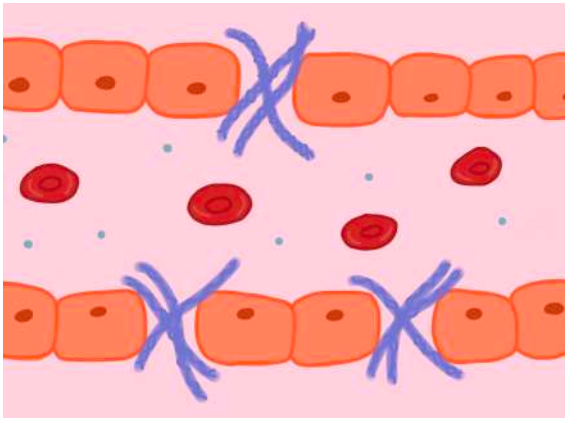
Nathan Joshua

Researchers at Harvard University recently discovered that spiders may be capable of dreaming in their sleep. While monitoring the sleep of E. arcuata, a species of jumping spiders, they found that the arachnids perform certain body movements like twitching and leg curling along with repeated intervals of retinal movements. This suggests that they experience a state similar to rapid-eye movement (REM) sleep, a stage of sleep that is associated with dreaming in humans. Although definitive proof has not been established, scientists hypothesize that eye movements during REM sleep are connected to visual scenes in dreams. This hints at the possibility that jumping spiders experience visual dreams, and raises an even bigger question - what purpose do visual dreams in animals serve?

Image: Male Jumping spider - *E. arcuata*
Reference: Rößler, D. C., Kim, K., De Agrò, M., Jordan, A., Galizia, C. G., & Shamble, P. S. (2022). Regularly occurring bouts of retinal movements suggest an REM sleep–like state in jumping spiders. Proceedings of the National Academy of Sciences, 119(33), e2204754119. doi:10.1073/pnas.2204754119
Image Source: Lukas Jonaitis - https://www.flickr.com/photos/38628972@N05/4932577872/. Licensed under the Creative Commons Attribution 2.0 Generic License

New Biomaterial Can Be Intravascularly Infused to Treat Heart Attack Damage

Ruthy Shin



Extracellular matrices (ECM) from decellularized tissues have great potential in tissue engineering. However, current ECM biomaterials exist only as surgical patches and localized injections. Researchers have developed a new, less invasive formulation of ECM that can be intravascularly infused to target leaky vasculature that result from injury such as acute myocardial infarction (MI), commonly known as heart attack. The infusible ECM (iECM) had optical properties nearly identical to saline and was hemocompatible. iECM was delivered intravascularly to preclinical rat MI models, in which it was localized specifically to damaged tissue. Minimal material was seen in the neighboring myocardium and no material was observed in the remote myocardium. Further studies demonstrated that iECM works by filling the gap between endothelial cells in areas of damaged tissue to minimize vascular leakage. If applied to humans, iECM could be administered immediately following a heart attack to heal damaged tissue and minimize permanent scarring.

Image: Artistic rendition of iECM fibers (purple) binding in endothelial cells of a leaky vessel.
Reference: M.T. Spang et al., Intravascularly infused extracellular matrix as a biomaterial for targeting and treating inflamed tissues. Nat. Biomed. Eng. 7, 94-109 (2023). doi:10.1038/s41551-022-00964-5
Image Source: Illustration by Ruthy Shin.

Unlocking an Empathetic Code: Advancing Computer Recognition of Emotions

Brynn Beatty

Your computer cannot tell when you are having a bad day, but with new research into improving a machine's understanding of its users' emotions, this could soon change. Rather than studying non-physiological signals such as a subject's expressions and movements, researchers at Toyama Prefectural University in Japan investigated various data processing techniques to more accurately identify the emotions of both individuals and groups using electroencephalography (EEG). Specifically, the researchers built a convolutional neural network model utilizing data from 32 EEG electrodes as participants watched 40 videos with different emotional components. Their proposed model achieved an accuracy of 85% for group data, displaying a marked improvement in recognizing the emotions of multiple users at once. Eliminating the barriers between human-computer interaction will enable scientists to delve into new branches of neuromarketing research and explore how computers understand why you act the way you do.

Image: New advancements in electroencephalography analysis methods help computers identify human emotions.
Reference: Cui, G., Li, X. & Touyama, H. Emotion recognition based on group phase locking value using convolutional neural network. Sci Rep 13, 3769 (2023). doi:0.1038/s41598-023-30458-6
Image Source: Illustration by Brynn Beatty.

The Use of miRNA Therapies for Cancer Treatment

Fady Samuel Michael[1], John J. Colicelli[2]

[1]Department of Molecular, Cell Developmental Biology, University of California, Los Angeles. [2]Department of Biological Chemistry, David Geffen School of Medicine .

ABSTRACT

MicroRNAs (miRNAs) are small, single-stranded biomolecules that bind complementary to messenger RNA to regulate gene expression by preventing translation. miRNAs have various implications in cancer research due to their ability to attenuate growth pathways, and it has been observed that dysregulation of endogenous miRNA is correlated to aggressive cancer growth both *in vivo* and *in vitro* across multiple cell lines. Since cancer typically involves overexpression of growth pathways or underexpression of tumor-suppressor pathways, miRNAs that complement these respective transcripts work to regulate the activity of these pathways. This review postulates the use of exogenous miR-34 therapy to mitigate the expression of overactive pathways in both preclinical and clinical models of cancer, and examines the use of miRNA therapies as a novel way to control overactive growth pathways. The therapeutic viability of exogenous miRNA is examined in both murine and human subjects, and its viability as a tool for oncologists is explored.

INTRODUCTION

miRNA and its Involvement In Cancer Progression

Cancer, as a disease, is unique due to its genotypic and phenotypic variability from patient to patient. Novel therapies are continuously being developed to personalize cancer treatment, and one recent avenue being explored focuses on microRNAs. MicroRNAs (miRNAs) are known to influence genetic regulation by acting as post-transcriptional silencers on a molecular level. There is empirical support that miRNA may be over- or under-expressed in various forms of cancer (*1*). High amounts of miR-195 have been observed in breast cancer patients (*2*) and an increase in circulating plasma miR-195-5p, miR-199a-3p, miR-320a, and miR-374a-5p has been recorded in osteosarcoma patients (*3*). Additionally, there was a significant decrease in these miRNAs following surgical resection, indicating that the expression of these miRNAs are influenced by cancer.

A specific subtype of miRNA known as miR-34 has been extensively studied for its role in modulating cancer progression. The miR-34 family, first identified in *Caenorhabditis elegans* in 2007, encodes a single miRNA which is evolutionarily conserved among several invertebrates. In mammals, three homologous transcripts exist: miR-34a, miR-34b, and miR-34c (*4*). The miR-34a gene maps to band 36, sub-band 22 on the short arm of chromosome 1 (1p36.22) in humans, and deletions of this 1p36 region are frequently observed in a variety of human cancers including neuroblastoma, glioma, breast cancer, non-small cell lung cancer, colorectal cancer, and melanoma (*5*). miR-34 has also been documented as a master regulator of tumor suppression by Chang *et al*. (*6*). The group found that miR-34a expression is induced by the tumor suppressor p53 following DNA damage in a colon cancer cell line (HCT116). The phenotypic effects of miR-34a expression were also modeled by transfecting miR-34 into p53 proficient and deficient HCT116 cell lines and observing that miR-34 transfection induced apoptosis regardless of the presence of p53, albeit at a reduced amount in the p53 deficient than proficient model. This study illustrates the anti-oncogenic effects of exogenous miR-34 delivery *in vitro* despite the lack of major tumor suppressors that regulate endogenous miRNA levels.

Given that overexpressed miRNA levels have been observed to stabilize following tumor resection in breast and osteosarcoma cancer patients and that exogenous expression of miR-34 in HCT116 cell lines reduced cancer growth regardless of the presence of p53, it is evident that miRNAs are a critical component in modulating pathways that could contribute to cancer progression. It is possible that increasing the amount of miR-34 in individuals with cancers shown to be deficient in this particular miRNA can attenuate phenotypes such as metastasis, size, and vascularity without the need for surgery. Therapies based on miR-34 may not entirely eradicate an individual's cancer, but may instead slow cancer progression.

miRNAs function as gene regulators by binding to complementary messenger RNAs (mRNAs) and halting their translation into protein via various degradation mechanisms. There exist a variety of non-canonical mechanisms for miRNA production; however, only canonical mechanisms are discussed in this review (Figure 1). In animals, miRNA recognizes its target mRNA by binding to complementary sequences in the 3' untranslated region (UTR). The first step of miRNA production begins with RNA polymerase II binding to genomic DNA at an intronic, intergenic, or polycistronic loci to transcribe a Primary miRNA (pri-miRNA) structure with a stem-loop hairpin motif that contains both a 5' 7-MG cap and Poly-A tail (*7, 8*). It has been demonstrated that RNA polymerase III can also be recruited for this process (*9, 10*).

Following pri-miRNA creation, precursor miRNA (pre-miRNA) is created through the microprocessor complex (MPC), which contains both Drosha and the DiGeorge Critical Region 8 (DGCR8). Specifically, Drosha binds to the basal "UG" motif and the stem region located on the pri-mRNA hairpin. DGCR8 binds to the double-stranded

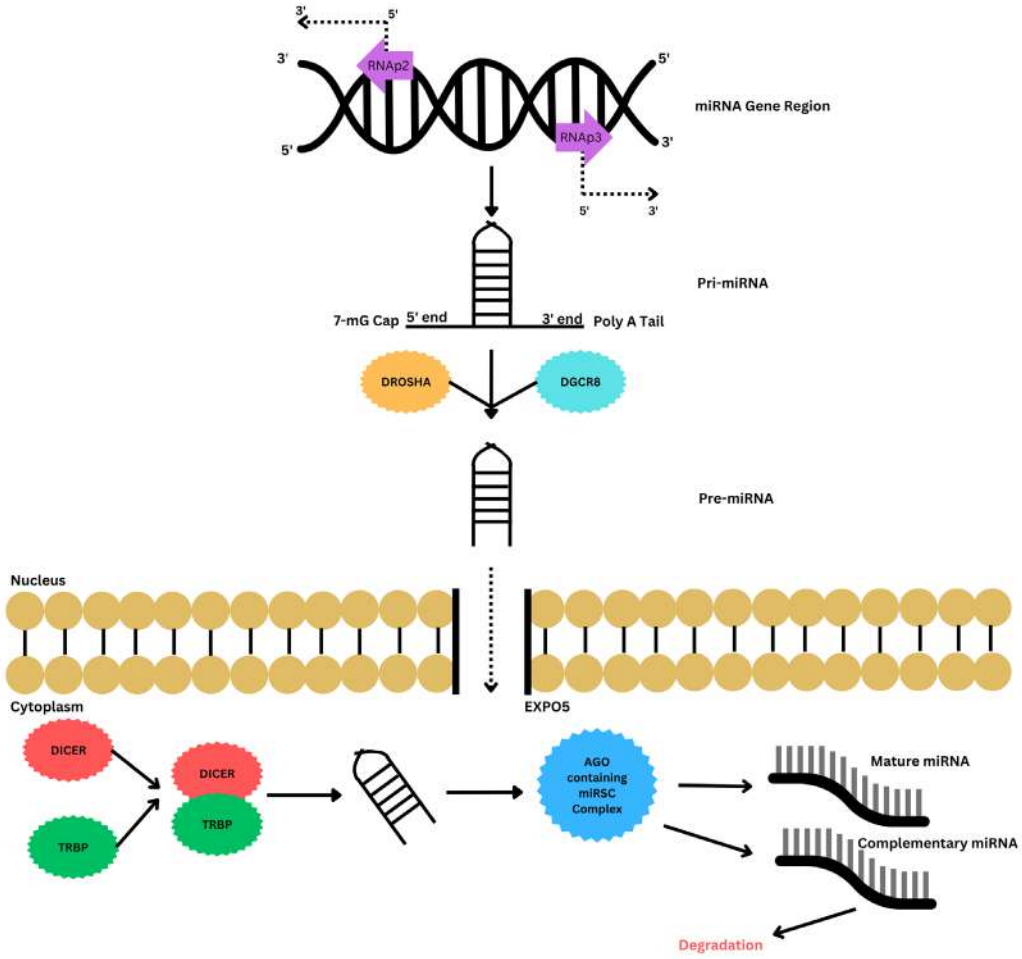


Figure 1: Canonical miRNA Synthesis and Production.
The production of microRNA (miRNA), a small, single-stranded biomolecule that binds complementary to messenger RNA (mRNA) to regulate gene expression by preventing translation, is facilitated by RNA polymerases II and III. After RNA polymerase recognizes the respective gene element that the cell intends on regulating via miRNA, miRNA is transcribed into an initial primary miRNA (pri-miRNA), which is structurally identified by its stem-loop hairpin motif. Pri-miRNA passes through the microprocessor complex (MPC) which contains DiGeorge Critical Region 8 (DGCR8) to yield pre-miRNA. Following translocation to the nucleus via Ran GTPase Exportin-5 (EXPO5), pre-miRNA is modified by the Dicer-trans-activating response RNA-binding protein (TRBP) complex to generate mature miRNA. Mature miRNA is further modified via an Argonaute protein (AGO)-containing miRNA-induced silencing complex (miRISC) duplex to yield the final, mature miRNA strand. miRNA regulation is facilitated by tumor suppressors, such as Tumor Protein 53 (TP53). Mutations arising in tumor suppressor proteins may conversely affect the synthesis of miRNA. This figure illustrates the complex process of miRNA synthesis and emphasizes the potential for miRNA administration therapies when synthesis is dysregulated by the incidence of cancer. Adapted from Ahmed *et al.* (*34*).

RNA-binding domain of pri-miRNA by complementation with the "UGU" motif (*11*). Pre-miRNA is then translocated to the cytoplasm via a Ran GTPase Exportin-5 complex (EXP5/XP05). This relocated pre-miRNA is then modified by the Dicer-trans-activating response RNA-binding protein (TRBP) complex to generate mature miRNA.

Following the creation of mature miRNA via TRBP, it enters into an Argonaute protein (AGO)-containing miRNA-induced silencing complex (miRISC) duplex. The pre-miRNA unwinds its duplex within this complex to generate a mature miRNA strand that is biochemically active, and a complementary miRNA strand that is typically degraded (*12, 13*).

This review investigates the viability of exogenous miRNA delivery systems and whether they can be employed in a safe and effective manner to treat patients with downregulated miR-34 observed in different forms of cancer. The regulatory mechanisms of miR-34 on growth pathways involving various forms of cancer are explored. Given the advancements in biotechnology within the last decade, miRNA therapy and treatment may provide a new avenue that minimizes the use of surgery and chemotherapy as the primary means of controlling cancer growth.

miR-34A Exhibits Anti-Oncogenic Activity Through Inhibiting Growth-Signaling Pathways and is Reduced in Cancer

The miR-34 family of miRNAs is directly induced by p53 (*14*). P53-regulated miRNAs mediate tumor suppression and cellular stress responses by regulating several key intracellular processes, such as cell cycle progression, migration, epithelial–mesenchymal transition, stemness, metabolism, differentiation, and cell

survival (*15*). miR-34a has been shown to directly target several oncogenic mRNAs and inhibit their functions, including BCL2, SIRT1, FRA1, E2F, MET, Notch1, Notch2, CDK4/6, VEGF, ARAF, PIK3R2, cyclin D3, cyclin E2, and PLK1 (*16*). The overactivity of these proteins drives cells into extensive growth, and miR-34 has an inhibitory effect on these pathways by localizing to the 3' UTR of their respective mRNA sequence and promoting degradation (*16*).

Given that miR-34 is able to regulate growth pathways and hyperactivity in these pathways results in cancer progression, it follows that several variations of cancer express reduced levels of miR-34. Parallel to Chang *et al*'s findings, Achari et al. found that ectopic expression of the miR-34 variant miR-34c induced G2/M growth arrest in a breast cancer derived cell line (*17*). miR-34a and miR-34c share the same seed sequence, defined as the first 2-8 nucleotides starting from the 5' end counting towards the 3' end. Therefore, these miRNAs target the same genes. Despite this, it has been hypothesized by Kim *et al*. that the two miRNAs vary in their phenotypic effects across different cancerous cell lines, so deviations between various cell lines should be considered (*18*). Nevertheless, the general principle that miR-34 administration reduces cancerous phenotypes by attenuating growth pathways remains consistent. This principle was strengthened by Lodygin *et al.*, who found that transfection of miR-34a led to an increase in senescence among human prostate and pancreatic carcinoma samples that exhibited reduced miR-34a expression following DNA damage (*19*). These three studies demonstrate that miR-34 administration may have potential therapeutic effects, making this form of miRNA a viable candidate for examining the benefits of exogenous miRNA delivery in models with reduced levels of miR-34.

Before examining the potential of miR-34 therapeutics *in vivo*, the molecular mechanisms which cause a decrease in miRNA expression should be considered. Notably, Lodygin *et al*. concluded from their findings that miR-34a expression was silenced in several types of cancer due to irregular CpG methylation of its promoter (*19*). This finding is consistent with prior research by Vogt *et al.*, which examined over 110 different cancer cell lines of diverse origin and found that over 75% of colorectal cancers (CRC) experience irregular methylation in miR-34a or miR-34b/c (20). Cytosine methylation of these promoters blocks RNA polymerase from accessing the gene to create miRNA. Overcoming DNA methylation in individuals may be difficult and further emphasizes the need for miR-34 based exogenous therapeutics.

To provide a succinct examination of miR-34 regulation *in vivo*, CRC-specific studies are referenced. For instance, Jiang *et al*. created an APC deficient mouse model ($Apc^{Min/+}$) to promote CRC formation (*21*). The group then created various miR-34 family knockouts and found that miR-34a/b/c reduction increased the number of intestinal stem cells compared to mice that did not have miR-34 knocked out. This reduction in miR-34a/b/c led to larger CRC tumor formation and a more severe cancerous phenotype compared to the control group. Additionally, tumor number and risk of cell death were significantly increased among the experimental group that expressed low levels of miR-34a/b/c. $Apc^{Min/+}$ mice deficient for either the miR-34a or the miR-34b/c allele did not show a statistically significant change in lifespan. This study demonstrates that miR-34 may contribute to minimizing the severity of cancer if reduced levels are brought back to baseline in cancers experiencing a reduction in miR-34.

Specific targets of miR-34, such as VEGFR-2 mRNA, allow insight into explaining the observed reduction in cancer severity when miR-34 levels are brought to baseline (Figure 2). Vascular Endothelial Growth Factor (VEGF) is a growth signal responsible for modulating various cellular kinases and signal transducers that promote endothelial cell viability. Activation of VEGF is responsible for proliferation, survival, migration, vascular permeability, lymphangiogenesis, vasculogenesis, and angiogenesis. As angiogenesis results in increased nutrient and blood supply, overexpression of this pathway has been consistently linked with cancer growth and severity. miR-34 targets the 3' UTR of the post-transcriptionally modified VEGF receptor (VEGFR-2) sequence (*22*). This regulatory mechanism works to attenuate hyperactive growth pathways (Figure 2) by degrading the mRNA that codes for the most upstream signal transducer within the pathway, the receptor itself. Furthermore, Zhang *et al*. found that miR-34 overexpression in CRC cell lines leads to the underexpression of phosphorylated-FAK (p397Y-FAK) (*23*). FAK is an important protein found within the VEGF pathway and phosphorylation of the 397 tyrosine residue constitutes its activation. FAK activation is responsible for transducing signals for focal adhesion and proliferation within the cell. It was also found that underexpression of miR-34 led to an overexpression of p397Y-FAK, which promoted oncogenic characteristics such as increased cell migration and invasive behavior of the cell lines examined. Since miR-34 levels were low in the experimental group, VEGF protein was not degraded by miR-34. ELISA, Western blot, and RT-PCR results verified that cells treated with miR-34 exhibited significantly reduced levels of VEGF and p397Y-FAK. The amount of p397Y-FAK was rescued by extra VEGF treatment, verifying that miR-34 directly regulates the VEGF receptor and could possibly regulate various other growth pathways in a similar manner (*24*).

The APC deficient studies by Jiang *et al*. illustrate that a lack of miR-34a/b/c in CRC induced mice increases cancer severity. The studies by Zhang *et al*. show that low levels of miR-34 increase cancerous characteristics by promoting overactivity of the VEGF pathway via p397Y-FAK. From these studies, it can be concluded that low levels of miR-34 may play a prominent role in cancer progression and development. Other studies have examined the pathophysiology of mice surrounding exogenous miR-34 treatment, and several therapies that have attempted to restore miR-34 expression as a means to alleviate cancer growth and progression are explored in the following sections.

The Potential for Exogenous miR-34 Administration in Cancer Patients with Reduced Levels of miR-34

miR-34 directly targets and attenuates various growth pathways by degrading mRNA prior to its translation into protein. The regulatory mechanisms of miR-34 itself may provide anti-oncogenic benefits against overactive growth pathways. In particular, cytosine methylation has been discussed as either a permanent repressive epigenetic modification or one that is subject to the environment (*26*). Nevertheless, it is difficult to restore the function of a hypermethylated miR-34 gene sequence through demethylation. Furthermore, miR-34 is regulated by tumor suppressor proteins. If a mutated tumor suppressor protein reduces the level of miR-34,

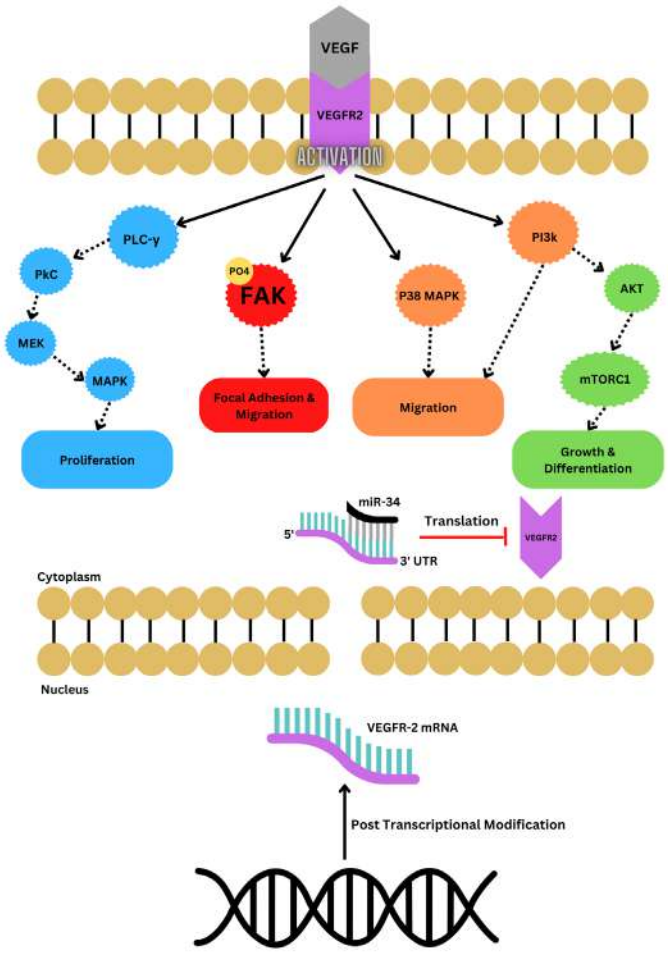


Figure 2: Methods of attenuating hyperactive growth pathways through the usage of exogenous miR-34 administration.
miR-34 binds to the 3' untranslated region (UTR) of Vascular Endothelial Growth Factor Receptor-2 mRNA (VEGFR-2 mRNA), resulting in mRNA degradation. Degrading the mRNA coding for the receptor leads to a decrease in Vascular Endothelial Growth Factor (VEGF) receptor density, which decreases tumor vascularization and growth. The findings from Zhang *et al*. demonstrate the potential for miRNA therapies in modulating other growth pathways, including the VEGF pathway seen in this figure, in a translationally-dependent manner (*23*). Several of the secondary messengers seen in this figure are also interlinked by other growth pathways, such as the Mammalian Target of Rapamycin 1 (mTORC1). mTORC1 has been shown to target the eIF4A and p70S6K1 proteins, both of which are responsible for promoting translation of proteins responsible for cell growth (25). miRNA modulates cell growth as a whole via mitigating the secondary response of other messengers interlaced within numerous growth pathways. The ability of miRNAs to target multiple pathways simultaneously can result in attenuation of overactive growth pathways in cancerous cells.

it may be challenging to rescue the mutant function of the tumor suppressor protein in order to restore miR-34 levels, as these mutations involve permanent allele differences within DNA. Given the murine studies by Jiang *et al.*, it is hypothesized that exogenous delivery of miR-34 may circumvent the aforementioned challenges and reduce the severity of cancer when miR-34 expression is sufficiently low.

The following section examines the murine research studying exogenous miR-34 delivery, and how this research has informed the investigation of similar pilot drugs in humans. The biggest challenge facing miR-34 delivery in humans is ensuring that the miRNA is able to effectively target cancerous cells while minimizing interaction with non-cancerous cells to avoid disrupting natural physiological functions. Additionally, the introduction of a foreign vehicle to deliver the drug may cause an immune response that is harmful to cancer patients who may already experience a hyperactive immune system that is attempting to eliminate cancer cells.

miR-34 Administration as a Potential Therapeutic-Murine Research. In the murine setting, Pramanik *et al.* delivered an exogenous miR-34 technology via a synthesized 100 nm lipid-based nanoparticle. His group used the MIA PaCa-2 cell line to induce a pancreatic cancer xenograft in mice, followed by systemic tail-vein administration of the drug three times each week for three weeks. Not only was tumor growth inhibited, but there was also a significant decrease in proliferation and an increase in apoptosis of cancerous tissue, indicating that the drug minimized cancer progression (*27*). Additionally, blood, liver, and kidney tests showed no significant toxicity after treatment with the drug. The Pramanik group showed that exogenous miR-34 delivery is safe and effective in reducing the progression of pancreatic cancer in mice.

Using a slightly different vehicle from the one employed by Pramanik *et al*., Wiggins *et al*. found that a Neutral Lipid Emulsion (NLE)-based nanoparticle delivery system carrying exogenous miR-34 was able to reduce cancerous effects in non-small-cell lung cancer (NSCLC) xenografted mice that exhibited reduced levels of miR-34 within their tumors (*28*). They also found that there was no significant increase in inflammatory cytokines, liver, or kidney enzymes and that this drug was able to effectively localize to the tumor. The Wiggins group showed that systemic delivery of miR-34 was not only able to localize to the tumor without disrupting normal physiology, but that it also reduced the progression of NSCLC xenograft mouse models that specifically experienced a reduction in miR-34.

To verify that the administration of miR-34 was safe and effective in cancerous mouse models, Trang *et al.* (*29*) replicated several methods from the Wiggins *et al.* study. The purpose of this study was to examine the effects of miR-34 and let-7 exogenous miRNA treatments in alleviating NSCLC burden. Similar to miR-34's anti-oncogenic effects, let-7 is a miRNA that negatively regulates several oncogenes that are frequently mutated in cancer. Let-7's targets include proteins responsible for cell growth and viability: RAS, MYC, and HMGA2 (30). In the Trang *et al.* study, a constitutively active KRAS G12D mutant was created using Cre recombinase. It was found that 100% of the mice treated with Cre recombinase developed NSCLC along with extensive diffuse hyperplasia and adenomas. The mice treated with exogenous miR-34a displayed a 60% reduction in tumor area compared to mice treated with a negative control. The mice treated with let-7 also displayed significantly decreased tumor burden. This proof of concept study demonstrated that cancer severity was reduced by administration of two different miRNA drugs in a NSCLC mouse model induced from a KRAS G12D mutation.

Taken together, these three studies demonstrate that exogenous administration of lipid-based nanovectors containing miRNA is safe and effective in mice models. The studies also extend the question of possible complications that may occur in human subjects if similar drugs were to transition to clinical trials. While mice may be an effective model to examine the pharmacokinetics and pharmacodynamics of drugs in mammals, they are not perfect. However, if data collected by clinical research groups corroborate the results obtained from these murine models, miRNA delivery technologies may have potential in the clinic.

Challenges Transitioning to the Clinic & mRNA Clinical Trials. The variability that exists between mice and human genomes must be considered when evaluating the transition from murine to human research. As such, introducing miRNA therapies into the clinic may be met with several barriers. These biological barriers include preventing systemic inflammation, ensuring the drug is properly metabolized with minimal toxicity, and selective targeting of cancer cells. In particular, minimizing inflammation from exogenous lipid-based drug delivery should be considered when transitioning to human trials. In a meta-analysis of fourteen studies, it was found that long chain triglycerides (LCT), medium chain triglycerides (MCT), and fish oil (FO) did not produce a significant immunological response in humans upon intravenous injection (*31*). The participants included in this study experienced a range of ailments including cancer, sepsis, malnourishment, or an undefined critical illness prior to receiving an injection. The lipid vectors developed by the Pramanik group, Wiggins group, and Trang group contained structurally similar lipid molecules that produced no immunological response when injected intravenously in mice. Given the data from the meta-analysis and the Pramanik et al., Wiggins et al., and Trang et al. studies, similar miRNA vehicles may be viable for further testing in humans, beginning with a phase one clinical trial in individuals with treatment-resistant cancer that exhibit reduced levels of miR-34.

MRX34 is a liposomal miR-34 drug that entered phase one clinical testing in 2020 with the intent of evaluating the safety of miR-34 exogenous delivery in treatment-resistant patients aged 18 and over (*32*). Patients with a variety of cancer types (Table 1) were included in the study. However, MRX34 clinical testing was discontinued due to lack of efficacy and continuing disease progression. Additionally, multiple adverse effects were observed as a result of treatment (Table 2). MRX34 treatment was the first clinical trial examining miRNA therapies in humans, but it was unable to pass into phase two testing due to lack of efficacy and associated toxicity.

Total Patients	n=85
Age	60
Sex: Male, n(%)	62 (73)
Race: Asian/Caucasian/Black/Other, %	49/33/2/16
Cancer Type	**n (%)**
Hepatocellular Carcinoma	36 (42)
Melanoma	9 (11)
Renal Cell Carcinoma	8 (9)
Lung	8 (9)
Gastrointestinal Stromal Tumor	6 (7)
Neurondocrine	6 (7)
Other	12 (14)
MRX34 Cycles Delivered, Median (Range)	2 (1-16)

Table 1: Patient demographics and cancer subtypes.
Patients from the United States of America and the Republic of Korea were enrolled into 3+3 dose-escalation cohorts. MRX34 was initially given intravenously twice every week (BIW) for 3 weeks of a 4-week cycle. During the final portion of the study, treatment was modified to daily injections for 5 days along with dexamethasone (DEX) pre-medication twice daily (BID) for 7 days in week 1, followed by 2 weeks of rest in 3-week cycles (QDx5 schedule). The QDx5 dosage scheduling was implemented following immune mediated toxicity and adverse events (AE) caused by BIW scheduling (32).

While the murine studies examining miR-34 administration performed by Pramanik *et al*., Wiggins *et al*., and Trang *et al*. are promising, the efficacy of this treatment in humans remains unclear. This discrepancy highlights that murine miRNA research may not be directly translatable to humans. Furthermore, sepsis, hypoxia, cytokine release syndrome, and hepatic failure of numerous patients involved in the MRX34 trial suggested a pattern of immune-mediated toxicity. One potential cause of the poor MRX34 treatment outcomes, most notably in terms of toxicity, may have been the vehicle used to administer the drug. The researchers used a liposome, which confers an anionic charge under physiological pH. There is evidence that anionic phospholipids serve as modulators of the coagulation cascade and that anionic liposomes induce

aggregation of platelets in guinea pigs, and blood cell aggregation in mice (*33*). Furthermore, the experimental design did not screen participants to ensure they had reduced miR-34 levels prior to starting treatment, which may have reduced the internal validity of the study. Although phase one studies focus on safety rather than efficacy, if blood serum testing or tumor biopsies were performed to ensure that prior miR-34 levels were reduced in these individuals, the treatment outcomes may have been improved by ensuring that the treatment was targeting a deficiency in miR-34. Overall, if MRX34 therapy had been targeted to more effectively localize to the tumor and mitigate sepsis-induced toxicity, the outcome of the trial may have been different.

Side Effects		**Laboratory Anomalies**	
Fever	72%	Lymphocytopenia	44%
Chills	53%	Thrombocytopenia	29%
Fatigue	51%	Neutropenia	21%
Back/Neck Pain	36%	Elevated AST	11%
Nausea	36%	Elevated ALT	11%
Dyspnoea	25%	Hyperglycemia	17%
		Hyponatremia	17%

Table 2: Physical side effects and laboratory anomalies following MRX34 administration.
85 total patients were treated with at least one dose of MRX34. The final treatment regimen included daily injections for 5 days along with dexamethasone (DEX) pre-medication twice daily for 7 days in week 1, followed by 2 weeks of rest in 3-week cycles (QDx5 schedule). MRX34 drug delivery was discontinued due to lack of clinical efficacy and excessive adverse effects (32).scheduling (32).

CONCLUSION

The significance of miRNA therapies as a treatment for cancer progression is supported by several studies across *in vitro* settings and murine research. The *in vitro* and murine studies explored open avenues for continued miRNA research and successfully developed tools to target various cancers. However, miR-34 treatment in the clinic raises serious concerns, and clinical researchers must learn from the failed MRX34 trial to effectively engineer new miRNA therapeutics. The termination of the phase one MRX34 trial was in part due to lack of clinical efficacy and excessive adverse effects among the patient cohort. Alternative outcomes may have been observed if the methodology of the study was modified to specifically target miR-34 deficient patients, rather than a cohort of individuals with various cancers (*32*). Additionally, modifying the mechanism of MRX34 drug delivery may have reduced adverse effects and toxicity. Given that a steroid-based anti-inflammatory drug (DEX) was provided alongside MRX34 administration, immune system hyperactivity may have been a moderating variable contributing to decreased efficacy and excessive side effects. This review highlights the need for increased research and engineering to discover and curate viable miRNA therapies for patients who have treatment-resistant tumors that can be targeted by specific miRNAs, along with effective vehicles to administer these novel therapies. miRNA treatment holds promise for the field of clinical oncology, but currently remains unsafe for human patients and requires further development before successful implementation.

REFERENCES

1. S. Filipów, L. Łaczmański, Blood Circulating miRNAs as Cancer Biomarkers for Diagnosis and Surgical Treatment Response. *Front Genet*. **10**, 169 (2019). doi: 10.3892/ol.2017.6028
2. H. M. Heneghan *et al.*, Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Ann Surg*. (2010). doi: 10.1097/SLA.0b013e3181cc939f.
3. F. Lian, Y. Cui, C. Zhou, K. Gao, L. Wu, Identification of a plasma four-microRNA panel as potential noninvasive biomarker for osteosarcoma. *PLoS One*. **10**, (2015). doi: 10.1371/journal.pone.0121499
4. M. Agostini, R. A. Knight, miR-34: from bench to bedside. *Oncotarget*. **5**, 872-881 (2014). doi: 10.18632/oncotarget.1825
5. K. O. Henrich, M. Schwab, F. Westerman, 1p36 tumor suppression--a matter of dosage? *Cancer Res*. **72**, 6079-6088 (2012). doi: 10.1158/0008-5472.CAN-12-2230
6. T. C. Chang, *et al*., Transactivation of miR-34a by p53 Broadly Influences Gene Expression and Promotes Apoptosis. *Mol Cell*. **26**, 745-752 (2007). doi: 10.1016/j.molcel.2007.05.010
7. E. Huntzinger, E. Izaurralde, Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet*. **12**, 99-110 (2011). doi: 10.1038/nrg2936
8. Y. Lee, *et al.*, MicroRNA genes are transcribed by RNA polymerase II. EMBO J. **23**, 4051-4060 (2004). doi: 10.1038/sj.emboj.7600385
9. S. Jiang, W. Yan, Current View of microRNA Processing. *Signal Transduct Insights*. **5**, (2016). doi:10.4137/STI.S12317
10. G. Borchert, W. Lanier, B. Davidson, RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol*. **13**, 1097-1101 (2006). doi: 10.1038/nsmb1167
11. A. N. Packer, Y. Xing, S. Q. Harper, L. Jones, B. L. Davidson, The bifunctional microrna mir-9/mir-9* regulates rest and CoREST and is downregulated in Huntington's disease. *J Neurosci.* **28**, 14341-14346 (2008). doi: 10.1523/JNEUROSCI.2390-08.2008
12. V. C. Auyeung, I. Ulitsky, S. E. McGeary, D. P. Bartel, Beyond Secondary Structure: Primary-Sequence Determinants License Pri-miRNA Hairpins for Processing, *Cell*. **152**, 844-858 (2013). doi: 10.1016/j.cell.2013.01.031
13. G. Tang, siRNA and miRNA: an insight into RISCs. *Trends Biochem Sci*. **30**, 106-114 (2005). doi: 10.1016/j.tibs.2004.12.007
14. H. Hermeking, MicroRNAs in the p53 network: micromanagement of tumor suppression. *Nat Rev Cancer*. **12**, 613-626 (2012). doi: 10.1038/nrc3318
15. L. He, *et al.*, A microRNA component of the p53 tumor suppressor network. *Nature*. **447**, 1130-1134 (2007). doi: 10.1038/nature05939
16. B. Wang, *et al*., miR-34A directly targets tRNAiMet precursors and affects cellular proliferation, cell cycle, and apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7392-7397 (2018). doi: 10.1073/pnas.1703029115
17. C. Achari, S. Winslow, Y. Ceder, C. Larsson, Expression of miR-34c induces G2/M cell cycle arrest in breast cancer cells. *BMC Cancer*. **14**, (2014). doi: 10.1186/1471-2407-14-538
18. J. S. Kim, *et al.*, MiR-34a and MiR-34b/c have distinct effects on the suppression of lung adenocarcinomas. *Exp Mol Med*. **58**, 1-10 (2019). doi: 10.1038/s12276-018-0203-1
19. D. Lodygin, *et al.*, Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. *Cell Cycle*. **7**, 2591-2600 (2008). doi: 10.4161/cc.7.16.6533
20. M. Vogt, *et al.*, Frequent concomitant inactivation of miR-34a and miR-34b/c by CpG methylation in colorectal, pancreatic, mammary, ovarian, urothelial, and renal cell carcinomas and soft tissue sarcomas. *Virchows Arch*. **458**, 313-322 (2011). doi: 10.1007/s00428-010-1030-5
21. L. Jiang, H. Hermeking, miR-34a and miR-34b/c Suppress Intestinal Tumorigenesis. *Cancer Res.* **7**, 2746–2758 (2017). doi: 10.1158/0008-5472.CAN-16-2183
22. P. Carmeliet, VEGF as a Key Mediator of Angiogenesis in Cancer. *Oncology.* **69**, 4-10 (2005). doi: 10.1159/000088478
23. D. Zhang, J. Zhou, M. Dong, Dysregulation of MicroRNA-34A Expression in Colorectal Cancer Inhibits the Phosphorylation of FAK Via VEGF. *Dig Dis Sci.* **59**, 958–967 (2014). doi: 10.1007/s10620-013-2983-4
24. M. Kowanetz, N. Ferrara, Vascular Endothelial Growth Factor Signaling Pathways: Therapeutic Perspective. *Clin Cancer Res*. **12**, 5018-5022 (2006). doi: 10.1158/1078-0432.CCR-06-1520
25. R. A. Saxton, D. M. Sabatini, mTOR Signaling in Growth, Metabolism, and Disease. *Cell*. **169**, 361-371 (2017). doi: 10.1016/j.cell.2017.02.004

26. B. Jin, Y. Li, K. D. Robertson, DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy?. *Genes Cancer.* **2**, 607-617 (2011). doi: 10.1177/1947601910393957
27. D. Pramanik, *et al.*, Restitution of tumor suppressor microRNAs using a systemic nanovector inhibits pancreatic cancer growth in mice. *Mol Cancer Ther*. **10**, 1470-1480 (2011). doi: 10.1158/1535-7163.MCT-11-0152
28. J. F. Wiggins, *et al.*, Development of a lung cancer therapeutic based on the tumor suppressor microRNA-34. *Cancer Res*. **70**, 5923–5930 (2010). doi: 10.1158/0008-5472.CAN-10-0655
29. P. Trang, *et al.*, Systemic Delivery of Tumor Suppressor microRNA Mimics Using a Neutral Lipid Emulsion Inhibits Lung Tumors in Mice. *Mol Ther*. **19**, 1116-1122 (2011). doi: 10.1038/mt.2011.48
30. S. M. Johnson, *et al.*, RAS is regulated by the let-7 microRNA family. *Cell.* **120**, 635-647 (2005). doi: 10.1016/j.cell.2005.01.014
31. M. Wirtitsch, *et al.*, Effect of different lipid emulsions on the immunological function in humans: a systematic review with meta-analysis. *Clin Nutr.* **26**, 302-313 (2007). doi: 10.1016/j.clnu.2007.02.001
32. D. S. Hong, *et al.*, Phase 1 study of MRX34, a liposomal miR-34a mimic, in patients with advanced solid tumours. *Br J Cancer*. **122**, 1630-1637 (2020). doi: 10.1038/s41416-020-0802-1
33. C. T. Inglut, *et al.*, Immunological and Toxicological Considerations for the Design of Liposomes. *Nanomaterials*. **10**, (2020). doi: 10.3390/nano10020190
34. F. Ahmed, *et al.*, Modification of miRNA Expression through plant extracts and compounds against breast cancer: Mechanism and translational significance. *Phytomedicine*. **68**, 153168 (2020). doi: 10.1016/j.phymed.2020.153168

Radiological Features Associated With Lung Adenocarcinoma and Aggressiveness

Joshua J. Perez[1], William Hsu[2]

[1]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles. [2]Department of Radiological Sciences, David Geffen School of Medicine.

ABSTRACT

Lung adenocarcinoma (LUAD) is the most common form of lung cancer, displaying highly heterogeneous behavior that hinders therapeutic response, which results in low survival rates. While morphologic patterns observed during histopathological analysis have been shown to correlate with specific subtypes of LUAD and aggressive cancers, radiologist-interpreted features from computed tomography (CT) scans are also highly useful in disease characterization and prognostication. This review surveys current scientific literature on prognostic, radiologic, and histopathologic features in early-stage LUAD (before the tumor has metastasized to other tissues). These surveyed studies were reviewed and filtered using a set of predefined inclusion/exclusion criteria, and 15 papers were further analyzed in greater detail for features significantly correlated to LUAD. While nodule size and consistency are found to be common predictors of LUAD, pleural invasion, solid component size, and intratumoral density have also been identified as prognostic factors of aggressive LUAD. With the increased adoption of low-dose CT lung cancer screening, quantitative and semantic characterization of nodule features could provide useful information to earlier identify patients at risk of aggressive disease. Understanding various aggressive features will allow physicians to not only better identify and diagnose unique forms of LUAD, but also create more effective treatments for aggressive early-stage LUAD.

INTRODUCTION

Lung cancer is the second most common cancer type, with nonsmall cell lung cancer (NSCLC) representing 80% of all lung cancers (*1*). Lung adenocarcinoma (LUAD), which is the most common of the three subtypes of NSCLC, represents 40% of NSCLC and develops in the glandular tissue of the lung (*2*). Depending on the stage of LUAD, mortality rates vary. When caught at an early stage (e.g., stage IA/IB, where cancer has not spread into the lymph nodes), the 5-year survival rate is >50%. However, when caught at later stages, survival drops to 20% (*3*). In order to understand the basis of this review, aggressive LUAD is characterized by disease that develops and spreads quickly, often resulting in metastatic disease, disease recurrence, or poor survival (less than 5 years). Patients with aggressive tumors can be candidates for more intensive treatment, such as adjuvant radiation therapy post-surgery to reduce the risk of metastasis or recurrence. Identifying LUAD at earlier stages is critical to reducing lung cancer-related mortality.

Routine low-dose computed tomography (CT) is used to screen high-risk individuals (older age, current or former smoker, or family history of cancer) for lung cancer (*4*). CT provides detailed anatomical *in-vivo* images of the lungs and was found to have >90% sensitivity in detecting lung cancer in the National Lung Screening Trial (*5*). Scans are acquired at routine intervals, allowing radiologists to detect the presence of suspicious nodules (solid components) or changes in the appearance of these nodules over time (indicators of malignancy). These nodules can contain varying areas of increased density in the lung, visualized on CT as ground-glass opacity (GGO), which is a measure of consistency within the lungs (*6*). GGOs can consist of either a solid component, part-solid nodule (the presence of a solid component within a lesion and GGO) or pure nodule (containing no solid components, only GGO) (Figure 1, Panel A). Guidelines such as the lung imaging reporting and data system (Lung-RADS) utilize a combination of tumor size, consistency, and size changes in follow-up scans to assess the risk of nodule malignancy (*7*). For high-risk nodules, diagnostic workup in the form of diagnostic CT imaging, positron emission tomography (PET) imaging, and/or tissue sampling (e.g., CT-guided biopsy) may be recommended. Beyond the features used in Lung-RADS, additional semantic descriptors, such as the roundness of the nodule and border definition (well or ill-defined border), have also been reported as potential predictors of malignancy (*8*).

This review surveys the current literature on quantitative and semantic imaging features and their relationship with lung cancer biology. Radiologic features can diagnose aggressive cancers earlier and inform of need for adjuvant therapy. The review is organized around the association between radiologic features with 1) LUAD diagnosis and 2) whether the identified cancer is aggressive.

IMAGING FEATURES ASSOCIATED WITH LUNG ADENOCARCINOMA

Prior literature review has shown that nodule size (maximum diameter and volume), morphology (shape), and opacity (consistency of nodule) are predictors of malignancy risk (*9*). For example, a nodule with larger diameter, irregular shape, or

spiculations (a lump of tissue with spikes or points on the tissues) would indicate a higher malignancy risk. Pathology is also an important diagnosis factor, as one study revealed an association between histopathological types (e.g., acinar, lepidic, solid, and papillary) and CT imaging results based on the mass of nodules of 148 patients in South Korea (*10*). It is imperative to consider all types of features present when diagnosing LUAD.

Various studies discuss a correlation between different types of GGOs and LUAD (Table 1). From a study of 2,010 patients at a single institution, Fu et al. showed that the presence of GGO is a strong indication and predictive marker of patients with invasive pathological stage I NSCLC (*11*). Based on analysis of thin-section CT scans acquired from 1,697 patients at the same institution, Suzuki et al. defined six types of LUAD consistency: pure, semi-consolidation, halo, mixed, solid nodule with GGO, and solid nodule without GGO (*12*). Pure and semi-consolidated GGOs are homogeneous and do not contain a solid component, while halo and mixed GGOs are heterogeneous and have a solid component of less than 50% of its diameter. Solid nodules contain a solid component of more than 50% of its diameter, and can contain GGO or lack GGO (Figure 1, Panel A). The two solid types are associated with invasive LUAD and lymphatic node metastases, while the others are considered early-stage LUAD with minimally invasive pathologic features. Solid-patterned tumors, therefore, justify more aggressive and attentive treatment. Similarly, Katsumata et al. found that tumors with more solid components had less predominantly lepidic (growth of tumor-like cells without vascular invasion) LUAD characteristics, meaning they require careful diagnosis for treatments (*6*). The different types of GGOs appear to have unique characteristics distinguishing LUAD, whereby these subtypes must be differentiated to be correctly related to gene markers.

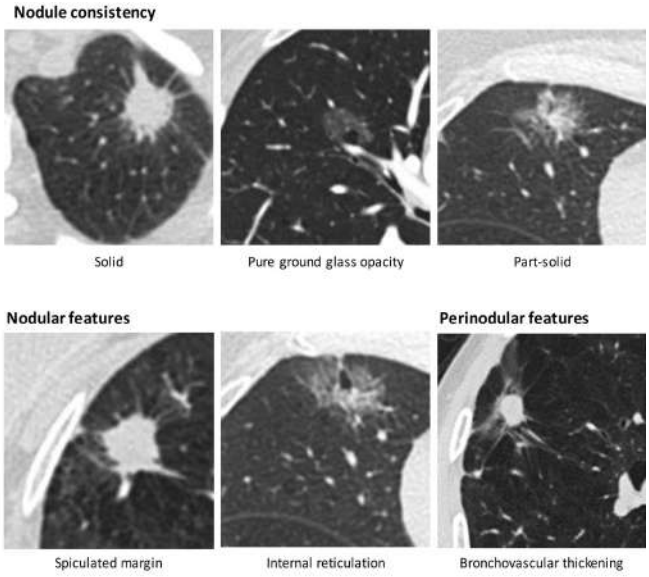


Figure 1: Examples of CT features.
Panel A (top row): nodule consistency (solid, pure GGO, part-solid), which is associated with varying risk of lung cancer. Panel B (bottom row): nodular features (spiculated margin, internal reticulation), which describe the internal architecture of the nodule, and perinodular features (bronchovascular thickening), which describe characteristics of the region surrounding the nodule.

Author Name	Cohort Size	Relevant Lesion Types	Imaging Characteristics	Molecular Characteristics	Takeaway Point
Hong et al.	250	Papillary Lepidic Acinar Solid	Size Shape Border Enhancement GGO	EGFR	Absence of GGO predictor of negative EGFR.
Katsumata et al.	1228	Lepidic	GGO Solid components Size	N/A	GGO proportion not related to characteristics and survival; presence on CT shows relationship
Lee et al.	148	Acinar Lepidic Solid Papillary	Mass Appearance Volume Diameter	EGFR KRAS	Pathological appearance/score related to DFS and mass of nodule
Suzuki et al.	1,697	N/A	Solid GGO Node metastases	N/A	Solid GGO related to invasive LUAD
Yin et al.	487	Micropapillary Solid	HGGNs PSNs GGO	EGFR	SSNs with solid components are more likely in mediastinal window and have poor differentiation.

Table 1: Summary of Imaging Features Associated with LUAD.
Common features found include lesion size, presence of GGO, EGFR gene marker, and lepidic histopathological subtype.

Gene markers also play an important role in determining if a patient is likely to develop GGO in their tumors. In one study, Hong et al. found higher rates of mutation in EGFR (with exons 19 and 21) in women and never-smokers, a trait correlated to tumors with GGO consistency (*12*). The most common histopathological subtypes in the group of EGFR patients were papillary, lepidic, acinar, and solid. Additionally, another study associated patients without an EGFR mutation with an absence of GGO (*13*). Focusing on GGOs, Yin et al. found that part-solid GGOs were more likely to have micropapillary subtypes and higher rates of genetic mutations (*14*). These studies reveal the effect of the EGFR mutation on GGO presence, which is similar to other features that can be variably assessed.

It is important to note the type of analysis used in these studies. GGOs are interpreted either through qualitative analysis describing imaging features, or quantitative image analysis characterizing GGOs and objectively assessing their appearance using intensity and texture features. For example, Zhu *et al.* analyzed intra-nodular and peri-nodular radiomic features, including determining GGO thickness and confirming the presence of Ki-67, a protein that is an indicator of cell proliferation. Ki-67's expression level is assessed by the Ki-67 labeling index (LI), where patients were sorted into either high Ki-67 expression or low Ki-67 expression (*15*). This study also found that the following two features were linked as risk markers for Ki-67: a spiculated margin (one of the main features distinguishing malignant pulmonary nodules) and the maximum 2D diameter of the nodule (Figure 1, Panel B). Another study using examining radiomic features to differentiate invasive carcinomas focused on voxel-level changes. This feature uses a 'voxel count', whereby each 3D pixel is counted and determines the size of a nodule, whereby smaller nodules contain fewer voxels. In this study, the voxel helped clarify distinguishing types of LUAD, along with the combination tumor surface area (*16*). Measures of quantitative analysis can help more accurately point out defining features related to GGO and LUAD.

Altogether, better understanding of radiologic and genomic features is key to obtaining better patient prognoses. Of the 15 papers reviewed between Tables 1 and 2, 11 discuss the relationship between solid or part-solid nodules and invasive cancers. A prior review examined the role of EGFR and KRAS mutations in LUAD and their association with the duration and intensity of smoking (*17*). The aforementioned discussion of gene markers demonstrated that EGFR were related to GGO nodules with a solid component. While 14 of these 15 papers identified EGFR mutation status as associated with GGO appearance, Katsumata et al. did not find a statistical difference in their cohort (*6*). To give some examples of papers determining this association as significant, Cohen et al. found EGFR associated with GGO in a retrospective study, and Han et al. in Korea correlated imaging presentation of the tumor to EGFR mutational status, either negative or positive (*18*). Additionally, Han et al. found that tumors with EGFR showed a higher proportion of GGO than tumors without mutations (*19*). Other factors relating to LUAD development and invasiveness in these papers included consolidation to tumor ratio and tumor size greater than 20mm. Lepidic and papillary subtypes of LUAD were reported to have the lowest recurrence rates. A broad spectrum of radiologic features is correlated with specific gene mutations, which cause poorer prognosis for LUAD.

IMAGING FEATURES ASSOCIATED WITH AGGRESSIVENESS

To determine the aggressiveness of specific LUAD cases, it is important to consider specific imaging characteristics (Table 2). GGO is particularly instrumental in predicting invasiveness. As discussed by Mimae et al., 23% of the pure GGOs analyzed contained an invasive component, or a portion of a tumor contributing to aggressiveness (*20*). To determine if the lesion was likely to be recurrent (reappear later in treatment), solid tumor size was analyzed, whereby a solid component less than or equal to 6mm was considered indolent. Similarly, Fu et al. found that the absence of GGO was a strong independent risk factor for predicting recurrence-free survival in stage I lesions (*11*). Their study identifies stage I lesions as having the highest risk of recurrence, where the absence of GGO or solid subtypes displays these the highest recurrence rates. In another manner, Kitazawa et al. discuss part-solid nodules less than 5 mm in size that have invasive potential. They found that quantitative assessment using CT attenuation values of GGOs distinguished preinvasive lesions or minimally invasive adenocarcinoma, apart from invasive LUAD (*21*). These assessments are key to determining aggressiveness, where tumor size must be considered.

In addition to tumor consistency, larger tumor size is associated with a higher recurrence rate. In one study, Kuroda et al. looked at tumors with sizes greater than 3 cm, where 324 patients from 4 different hospital institutions had poor prognosis. Mediastinal diameter, calculated using the mediastinal window volume, was associated with a higher risk of lymph node involvement (*22*). Also, Suzuki et al. found that size distinguished noninvasive LUAD. For example, a tumor with less than 2.0 cm and less than 0.25 consolidation was likely to be indolent (*23*). Similarly, Mimae et al. found that solid tumor size, whole tumor size, and pleural invasion (the penetration of beyond the visceral pleural area of the lungs) predict indolent lung tumors (*20*). As demonstrated by these studies, considering tumor size is necessary for evaluating aggressiveness in LUAD alongside other unique features.

Additional characteristics that predict aggressiveness are tumor shape and intratumoral density (Figure 1, Panel B). Looking at both of these characteristics, Grove et al. observed how tumor shape complexity and intratumoral density variation were linked to the prognosis of LUAD. This study revealed that these features were descriptive and correlated with imaging characteristics and patient survival in LUAD. Radiomic features extracted from imaging were used to describe tumor shape complexity and intratumoral density. While the study found that both of these features were associated with overall survival in their development cohort, it could not show statistical significance of the intratumor density feature itself when using a second cohort (*24*). Focusing on intratumoral density, Choi et al. found a strong association of intratumoral density with pathologic complexity, or the presence of varied heterogeneous pathologic components. This study proposed that radiomic variables such as density were predictors of heterogeneity in the whole tumor (*25*). Put together, these studies reveal that irregular tumor shape and abnormal densities can contribute to poorer prognosis. Thus, considering tumor shape and intratumoral density will help better predict aggressive nodules.

Author Name	Cohort Size	Relevant Lesion Types	Imaging Characteristics	Molecular Characteristics	Takeaway Point
Choi et al.	89	N/A	Intratumoral density	N/A	Tumoral density is predictive of pathologic complexity and can help predict intratumoral heterogeneity.
Deng et al.	403	N/A	Tumor size Pleural indentation Spiculation	Differentiation	Visceral pleural invasion is correlated with LUAD and unfavorable prognosis in early-stage patients
Dou et al.	200	N/A	Radiomic analysis of peritumoral regions	N/A	Parenchymal tissue is an important indicator of more invasive types of LUAD, and multiple programs should be run to detect these predictors.
Fu et al.	2,010	Acinar Micropapillary Lepidic Papillary Solid	GGO Size	EGFR KRAS	Presence of GGO is a strong indication and prediction of patients with invasive stage 1 pathological NSCLC
Grove et al.	61; 47	N/A	Convexity Entropy ratio Shape	N/A	Tumor shape complexity with a higher convexity is a predictor of poor prognosis and survival.
Kitazawa et al.	96	Lepidic	Size Maximum diameter GGO	N/A	3D CT attenuation values distinguish pre-invasive lesions/minimally invasive adenocarcinoma from invasive ADC
Mimae et al.	85	Lepidic Papillary	Solid Component Size GGO	N/A	Predicting invasive types of LUAD can be based on solid component size and pure GGOs
Suzuki et al.	545	N/A	Size Consolidation Invasive GGO	N/A	Aligned pathologic and radiologic prognoses have good treatment rates; misaligned prognoses are undertreated or overtreated
Yotsukura et al.	1032	Non-lepidic Lepidic Papillary Acinar	Tumor size Invasive component size	EGFR	Pure lepidic adenocarcinoma with CAF have excellent prognosis; non-lepidic growth patterns have a worse prognosis
Zhang et al.	762	Acinar Papillary Solid	Air bronchogram Maximum tumor diameter Maximum solid component diameter CTR	EGFR ROSI ALK	Indicators of STAS will help better predict patients who will have higher recurrence rates

Table 2: Summary of the relationship between imaging features and aggressive disease.
Common features found include presence of GGO, lesion size, containing a solid component, and papillary histopathological subtype.

Other factors, including STAS (spread through air spaces) status and VPI (visceral pleural invasion), should also be considered when diagnosing aggressive LUAD. Examining predictive value of features, Dou et al. focused on peritumoral features (Figure 1, Panel B). Using data from 200 NSCLC patients, this study used radiomic characterization of the parenchyma (a portion of lungs involved in gas exchange) and the tumor nodule. The radiomic signature was found to have a higher prognostic performance (a better 5-year survival rate) than similar models in other studies. Moreover, STAS has been noted as a risk factor for stage IA LUAD recurrence (*26*). Additionally, Zhang *et al.* analyzed 762 individuals at a single institution and found that air bronchogram (a malignant pattern of air-filled bronchi), maximum tumor diameter, and maximum solid component diameter had high sensitivity (95%) but poor specificity (30%) in predicting aggressive disease. Furthermore, STAS was more frequent in acinar, micropapillary, and solid subtypes. VPI, which is the tumor penetration of the layer behind the visceral pleura, can indicate cancer cells crossing sections to invade a particular lobe of the lung (*27*). While Zhang et al. focused on air bronchogram and STAS with VPI, Deng et al. observed that NSCLCs with VPI tended to have more significant rates of pleural indentation and spiculation (*28*). VPI could be an early predictor of whether NSCLC would be poorly differentiated and have more invasive solid components. Taking into account these imaging features is important to correctly and effectively determine a patient's outcome.

DISCUSSION

This literature review summarizes quantitative and qualitative features recognized by radiologists and pathologists to better understand when to diagnose and treat early-stage LUAD patients. Imaging features include consistency (e.g., solid vs. GGO), tumor size, and uneven shape or borders of tumors (*18*). However, these features cannot fully replace the need for tissue sampling due to uncertainty when assessing imaging features. Overall, more effective treatments are necessary to increase survival rates. Additionally, being able to characterize the aggressiveness of a tumor is critical to determine the treatment plan and its resulting effectiveness. While some nodules may be less invasive, others must be treated immediately and efficiently to prevent cancer metastasis.

Opportunities for newer image analysis techniques to characterize LUAD and its aggressiveness are likewise developing. One promising direction in this review is the area of quantitative image analysis, including radiomics. Also known as radiogenomics, this is known as the correlation between radiomic features and molecular characteristics. Although radiogenomics studies may ultimately help guide treatment decisions, they still require further investigation and validation over multiple cohorts. Furthermore, deep learning models and radiomic analysis can identify malignancies and determine their aggressiveness, providing the best course alongside histopathological analysis to predict patient treatment outcomes.

REFERENCES

1. A. Mogi, H. Kuwano. TP53 mutations in nonsmall cell lung cancer. *J. Biomed. Biotechnology.* **2011**, 583929 (2011). doi: 10.1155/2011/583929.
2. N. Kim et al., Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020). doi: 10.1038/s41467-020-16164-1.
3. B.Y. Wang et al., The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. *J. Cancer Res. Clin. Oncol.* **146**, 43–52 (2020). doi: 10.1007/s00432-019-03079-8.
4. T.V. Bilfinger, D. Albano, M. Perwaiz, R. Keresztes, B. Nemesure. Survival Outcomes Among Lung Cancer Patients Treated Using a Multidisciplinary Team Approach. *Clin. Lung Cancer.* **19**, 346–51 (2018). doi: 10.1016/j.cllc.2018.01.006.
5. The National Lung Screening Trial Research Team. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *N. Engl. J. Med.* **368**, 1980–91 (2013). doi.org: 10.1056/NEJMoa1209120.
6. S. Katsumata et al., Pathological features and prognostic implications of ground-glass opacity components on computed tomography for clinical stage I lung adenocarcinoma. *Surg. Today.* **51**, 1188–202 (2021). doi: 10.1007/s00595-021-02235-3.
7. P.F. Pinsky et al., Performance of Lung-RADS in the National Lung Screening Trial: A Retrospective Assessment. *Ann. Intern. Med.* **162**, 485–91 (2015). doi:10.7326/M14-2086.
8. Q. Li et al., Comparison Between Radiological Semantic Features and Lung-RADS in Predicting Malignancy of Screen-Detected Lung Nodules in the National Lung Screening Trial. *Clin. Lung Cancer.* **19**, 148-156 (2018). doi: 10.1016/j.cllc.2017.10.002.
9. M.M. Wahidi, J.A Govert, R.K. Goudar, M.K. Gould, D.C. McCrory. Evidence for the Treatment of Patients With Pulmonary Nodules: When Is It Lung Cancer? *Chest*. **132**, 94S-107S (2007). doi: 10.1378/chest.07-1352.
10. H.Y. Lee et al., Solitary Pulmonary Nodular Lung Adenocarcinoma: Correlation of Histopathologic Scoring and Patient Survival with Imaging Biomarkers. *Radiology*. **264**, 884–93 (2012). doi: 10.1148/radiol.12111793.
11. F. Fu et al., Distinct Prognostic Factors in Patients with Stage I Non–Small Cell Lung Cancer with Radiologic Part-Solid or Solid Lesions. *J. Thorac. Oncol.* **14**, 2133–42 (2019). doi: 10.1016/j.jtho.2019.08.002.
12. K. Suzuki, M. Kusumoto, S. Watanabe, R. Tsuchiya, H. Asamura. Radiologic classification of small adenocarcinoma of the lung: radiologic-pathologic correlation and its prognostic impact. *Ann. Thorac. Surg.* **81**, 413–9 (2006). doi: 10.1016/j.athoracsur.2005.07.058.
13. S.J. Hong et al., Radiogenomic correlation in lung adenocarcinoma with epidermal growth factor receptor mutations: Imaging features and histological subtypes. *Eur. Radiol.* **26**, 3660–8 (2016). doi: 10.1007/s00330-015-4196-z.
14. J. Yin et al., Solid Components in the Mediastinal Window of Computed Tomography Define a Distinct Subtype of Subsolid Nodules in Clinical Stage I Lung Cancers. *Clin. Lung Cancer.* **22**, 324–31 (2021). doi: 10.1016/j.cllc.2021.02.015.
15. M. Zhu et al., Predicting Ki-67 labeling index level in early-stage lung adenocarcinomas manifesting as ground-glass opacity nodules using intra-nodular and peri-nodular radiomic features. *Cancer Med.* (2022). doi: 10.1002/cam4.4719.
16. W. Li et al., Radiomic analysis of pulmonary ground-glass opacity nodules for distinction of preinvasive lesions, invasive pulmonary adenocarcinoma and minimally invasive adenocarcinoma based on quantitative texture analysis of CT. *Chin J. Cancer Res. Chung-Kuo Yen Cheng Yen Chiu.* **30**, 415–24 (2018). doi: 10.21147/j.issn.1000-9604.2018.04.04.
17. S. Dogan et al., Molecular Epidemiology of EGFR and KRAS Mutations in 3,026 Lung Adenocarcinomas: Higher Susceptibility of Women to Smoking-Related KRAS -Mutant Cancers. *Clin. Cancer Res.* **18**, 6169–77 (2012). doi: 10.1158/1078-0432.CCR-11-3265.
18. J.G. Cohen et al., Lung adenocarcinomas: correlation of computed tomography and pathology findings. *Diagn. Interv. Imaging.* **97**, 955–63 (2016). doi: 10.1016/j.diii.2016.06.021.
19. X. Han et al., Value of CT features for predicting EGFR mutations and ALK positivity in patients with lung adenocarcinoma. *Sci. Rep.* **11**, 5679 (2021). doi: 10.1038/s41598-021-83646-7.
20. T. Mimae et al., What are the radiologic findings predictive of indolent lung adenocarcinoma? *Jpn J. Clin. Oncol.* **45**, 367–72 (2015). doi: 10.1093/jjco/hyv005.
21. S. Kitazawa et al., Three-dimensional mean CT attenuation value of pure and part-solid ground-glass lung nodules may predict invasiveness in early adenocarcinoma. *Clin. Radiol.* **74**, 944–9 (2019). doi: 10.1016/j.crad.2019.09.130.

22. H. Kuroda et al., Prognostic significance of combined radiologic imaging modalities for prognosis of clinical IA adenocarcinomas. *Oncotarget*. **9**, 10745–53 (2017). doi: 10.18632/oncotarget.23395.
23. K. Suzuki et al., A prospective radiological study of thin-section computed tomography to predict pathological noninvasiveness in peripheral clinical IA lung cancer (Japan Clinical Oncology Group 0201). *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer.* **6**, 751–6 (2011). doi: 10.1097/JTO.0b013e31821038ab.
24. O. Grove et al., Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PloS One.* **10**, e0118261 (2015). doi: 10.1371/journal.pone.0118261.
25. E.R. Choi et al., Quantitative image variables reflect the intratumoral pathologic heterogeneity of lung adenocarcinoma. *Oncotarget*. **7**, 67302–13 (2016). doi: 10.18632/oncotarget.11693.
26. T.H. Dou, T.P. Coroller, J.J.M. van Griethuysen, R.H. Mak, H.J.W.L. Aerts. Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PloS One.* **13**, e0206108 (2018). doi: 10.1371/journal.pone.0206108.
27. Z. Zhang et al., Predictive value of radiological features on spread through air space in stage cIA lung adenocarcinoma. *J. Thorac. Dis.* **12**, 6494–504 (2020). doi:10.21037/jtd-20-1820.
28. H.Y. Deng et al., Novel biologic factors correlated to visceral pleural invasion in early-stage non-small cell lung cancer less than 3 cm. *J. Thorac. Dis.* **10**, 2357–64 (2018). doi: 10.21037/jtd.2018.03.185.

Maintenance of Hemiplegic Migraines and Alternative Treatments

Maggie T. Han[1]

[1]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles.

ABSTRACT

Migraines are a disabling condition that disrupt the lives of those who are diagnosed. Hemiplegic migraines are a rare and particularly disruptive kind of migraine. Symptoms include unilateral weakness accompanied by nausea, dizziness, and ataxia. Current available drugs involve abortive and preventative medications with no standardized protocol for treatment options other than addressing symptoms. This paper provides a comprehensive review on the latest alternative treatments for hemiplegic migraines, such as triptans, marijuana, optical treatments, Botox, and peripheral neurostimulation. Lifestyle factors such as nutrition, sleep, and exercise, which can contribute to migraines, are also discussed in terms of maintenance to reduce the frequency and severity of migraines. Due to the complex pathophysiology of hemiplegic migraines, future research should be focused on upstream factors and alternative treatments to treat patients who may be contraindicated for standard pharmaceutical drugs. Promising treatments include Botox, marijuana, and nutraceuticals. To ensure the efficacy of these treatments, clinical studies should be conducted with larger sample sizes, follow-ups, and appropriate control groups. To integrate alternative treatments more efficiently into a standardized protocol for treatment, studies should also be conducted to see how novel treatments complement currently prescribed medications.

INTRODUCTION

Migraines are a type of neurovascular dysfunction disease accompanied by photophobia (fear of light), phonophobia (fear of sound), and nausea or vomiting. Migraines also have a significant impact on society as a whole. They are the sixth most disabling condition in the world in terms of years lost due to disability and affect 15-20% of people living in the United States (*1*). The National Health Interview Survey found that approximately 1.4 billion dollars are lost annually due to disability and decreased quality of life, with direct medical costs approximating 1 billion dollars in the US (*2*).

Hemiplegic migraines are a particularly disabling and rare type of migraine. They are preceded by an aura, which induces intense symptoms or sensations such as flashing lights. Long-lasting auras can lead to hospitalization. Hemiplegic migraines are accompanied by hemiplegia (unilateral weakness) and various other symptoms such as nausea, confusion, dizziness, and aphasia (*3*). Although the global prevalence of hemiplegic migraines is unknown, a study of the Danish population found a prevalence of 0.01% (*4*).

Currently there are few studies on hemiplegic migraines and no well-established protocols for treatment. Diagnosis relies on clinical assessments with genetic confirmation, and current pharmacologic treatments for hemiplegic migraines rely on evidence from individual reports due to a lack of controlled trials (*3*). In addition to pharmacologic treatments, risk factors and triggers of migraines can be mitigated to reduce frequency of migraines.

This review provides a comprehensive overview of hemiplegic migraine treatment, with an emphasis on non-pharmaceutical treatment to supplement the current standard of care (Table 1). A variety of treatments will be discussed in terms of their efficacies, benefits, and disadvantages. Lifestyle modifications such as sleep behaviors, physical exercise, and dietary changes will also be discussed, since they can also reduce the frequency of migraines. Unfortunately, there are few studies researching how promising upcoming treatments can complement risk factor control and lifestyle modifications. Overall, a variety of factors can be investigated to create a more standardized protocol for treating patients, contributing to the better understanding of the pathophysiology of migraines.

RISK FACTORS AND GENETICS

The risk of developing hemiplegic migraines depends on genetics. Hemiplegic migraines are described as familial if there is a family history of first-degree relatives with the same condition and sporadic if there is no previous family history. The inheritance pattern of familial hemiplgeic migraine (FHM) is autosomal dominant with reduced penetrance. Penetrance is defined as the percentage of individuals with a particular mutation or genotype who exhibit clinical signs or phenotype of the associated disorder or genotype. While the pathophysiology of sporadic hemiplegic migraines is unclear, certain genes have been implicated in familial hemiplegic migraines (*6*). Those genes are the CACNA1A gene, ATP1A2 gene, and the SCN1A

gene, each of which corresponds with a different form of FHM and is associated with calcium channels and nerve cell signaling (Figure 1). These genes increase the likelihood of spreading cortical depression associated with auras and possibly even migraines themselves. Cortical depression is a neurophysiological phenomenon characterized by abrupt changes in intracellular ion gradients and sustained depolarization of neurons. Cortical depression leads to a loss of electrical activity, changes in synaptic architecture, and altered vascular responses (*7*). Functional studies reveal that these genes may also be involved in neuronal hyperexcitability that plays a pivotal role in the pathogenesis of hemiplegic migraines (*8*). CACNA1A encodes a subunit of P/Q type calcium channels and is associated with ataxia. ATP1A2 encodes a catalytic subunit of a sodium/potassium ATPase and is theorized to cause depolarization of neurons followed by activation of the trigeminovascular system, which has been associated with the headache aspect of migraines (*9*). Lastly, SCN1A encodes a transmembrane alpha subunit of a brain sodium channel that has been theorized to lead to neuronal excitability. Current hypotheses propose that neuronal excitability makes the brain increasingly more sensitive to perturbations until cortical depression is achieved (*10*). It is also possible to exhibit the symptoms of hemiplegic migraines even when no mutations are present in any of these genes.

These genes are also occasionally seen in patients with sporadic hemiplegic migraine, which accounts for 25% of hemiplegic migraines. Another gene linked with sporadic hemiplegic migraines is SLC1A3, which encodes the glutamate transporter EAAT1. Mutations in this gene can cause neuronal excitability (*8*). For sporadic hemiplegic migraines, the migraines can also arise from mutagens or from chance.

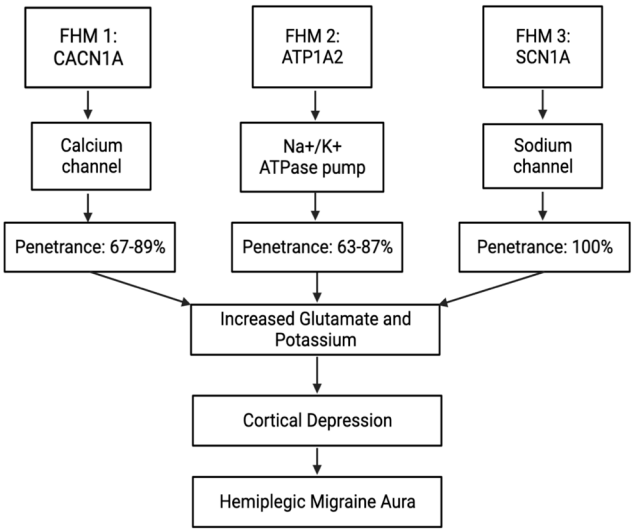


Figure 1: Proposed Mechanism of Hemiplegic Migraine Aura.
CACN1A (Calcium Voltage-Gated Channel Subunit Alpha 1A), ATP1A2 (ATPase Na+/K+ Transporting Subunit Alpha 2), and SCN1A (Sodium Voltage-Gated Channel Alpha Subunit 1) are the three genes that cause the three types of familial hemiplegic migraines. These genes are involved with the regulation of ions such as calcium, potassium, and sodium. Increased glutamate and potassium lead to neuronal excitability, and this in turn can lead to cortical depression which has been hypothesized to be involved with hemiplegic migraine auras. Other than SCN1A, the genes show reduced penetrance which refers to the percentage of individuals having a particular mutation or genotype who exhibit clinical signs or phenotype of the associated disorder or genotype

Non-genetic risk factors can be defined as the circumstances that trigger migraines. Patients who are unable to avoid these triggers in daily life are therefore at higher risk for migraines than those who can. Non-modifiable risk factors include age, sex, hormonal changes, and other socioeconomic variables. Migraines are three times more prevalent in women compared to men. This trend may be related to hormonal changes as migraines have been linked to menstrual cycles and pregnancy and often improve after menopause (*11*). Migraines tend to develop during adolescence with a peak in frequency during the late teens and 20s then again around 50 years of age (*12*).

Although the pathophysiology of migraines is not fully understood, it is hypothesized that migraines are triggered by changes in the brainstem modulating afferent signals and interactions with the trigeminal nerve. Current research is also evaluating the role of neurochemicals such as serotonin or calcitonin gene-related peptides (CGRP) and their influence on pain regulation (*13*).

	Treatment	Mechanism	Administration
Pharmaceutical	Verapamil	Reduces vasoconstriction and blocks calcium influx	Intravenous and oral administration
	Ketamine	NMDA glutamate receptors antagonism	Intranasal administration
	Triptans	Agonism of 5-HT receptors which respond to 5-HT involved in serotonin	Oral or subcutaneous
	Naloxone	Opiate-agonism Possible role of endorphins	Intravenous
	Acetazolamide	Unknown Possibly due to improved calcium channel functioning	Oral 250-500 mg twice a day
Alternative	Visual Treatment	Reduce photosensitive cell stimulation or green light exposure	Optical filters or Green Light Emitting Diode (GLED)
	Peripheral Neurostimulation	Electrical pulses to stimulate certain areas of the brain	Implantation or single pulses
	Botox	Inhibit exocytosis of neurochemicals, proinflammatory molecules and excitatory neuropeptides	31-39 injections of Botox for roughly 12 weeks
	Marijuana	Blocking cannabinoid receptors and targeting inflammation	Oral

Table 1: Pharmacological and Alternative Treatments of Migraines.
A summary of current pharmaceutical and alternative treatments for migraines with their mechanisms and method of administration. Pharmaceutical drugs are FDA-approved and the current standard of care, but triptans are currently contraindicated for hemiplegic migraines. Additionally, the prescription of these pharmaceutical drugs is not standardized and often results in side effects that can not be tolerated by patients. When paired with pharmaceutical treatments, alternative treatments can effectively mitigate the severity of migraine symptoms and decrease the frequency of migraines.

LATEST RESEARCH AND TREATMENTS

Due to their rarity, the mechanisms of hemiplegic migraines are not as well-understood as other conditions. Current treatments for hemiplegic migraines primarily focus on pharmaceutical methods of managing migraines. However, research has shifted from a focus on drugs to a focus on creative, natural, and preventative solutions. As the use of drugs is not a comprehensive treatment and is ineffective for some patients, more holistic solutions need to be researched.

Current standard of care

Standard of care for migraines usually involves abortive and preventative medications. Abortive medications focus on stopping a migraine that is already happening, while preventative medications focus on preventing migraines from happening in the future. Common drugs used to treat migraines are Verapamil, Ketamine, Naloxone, and Acetazolamide. Other methods include preventative medications such as tricyclic antidepressants (amitriptyline), anti-seizure medications (topiramate and valproic acid), and the avoidance of other medications like beta-blockers (propranolol) (*3*). For long-term treatments, abortive drugs include paracetamol and nonsteroidal anti-inflammatory drugs like ketoprofen and aspirin (*14*). Hemiplegic migraines are treated with similar drugs, and there is no specific standardized protocol to follow when treating these types of migraines. Doctors treat hemiplegic migraine patients on a case-by-case basis and frequently (if applicable) recommend that patients identify triggers and avoid them. Common triggers of migraines include hormonal changes in women, alcohol, caffeine, stress, sensory stimuli, sleep changes, physical exercise, weather changes, medications, and foods or food additives (*15,16*). Severe migraine attacks may lead to hospitalization, and diagnosis and treatment of this condition is difficult due to symptoms resembling other serious neurological diseases such as stroke (*17*).

Side effects and contraindications

Although current medications can be effective for the general population, sometimes there are side effects that may lead to individuals discontinuing treatment. In a study of 2444 patients, two-thirds of the patients reported delaying or avoiding medication due to worries of adverse side effects (*18*). One possible side effect is Medication Overuse Headache (MOH), which occurs when patients with migraines overuse their medication and, in turn, increase the intensity and frequency of their migraines (*19*). Another common side effect is liver and kidney toxicity (*20*). Common contraindications for migraine medications include cardiovascular and cerebrovascular disease, ischemic stroke, Reynaud's disease, pregnancy, lactation, hypertension, renal failure, and liver failure. For children, however, few clinical studies have been conducted for the efficacy of medications due to high placebo responses (*21*). For those who suffer from adverse effects or have conditions that interfere with them taking certain medications, non-pharmacological and alternative treatments may be a more effective choice.

Triptans

Triptans are commonly used to treat migraines other than hemiplegic migraines. Current research suggests that the triptan mechanism of action involves vasoconstriction, the narrowing of blood vessels. This could cut off blood flow to some tissues, causing heart attacks or ischemic strokes (when brain tissue does not get enough blood or nutrients) (*22*). Vascular contributions may play a causative component in migraines, and this communication between blood vessels and neurons could reveal the pathophysiology of migraines (*23*). For hemiplegic migraines, however, the use of triptans is contraindicated. Due to the concern that the vasoactive properties of triptans can increase the risk of stroke, researchers exclude those diagnosed with hemiplegic migraines from most migraine studies involving triptans (*8*). However, a study with 76 patients diagnosed with hemiplegic migraines demonstrated that the use of a triptan did not result in patients experiencing an ischemic stroke or heart attack. The study suggested that triptans were safe to use for hemiplegic migraines; however, caution is still needed as the retrospective setting of the study is not optimal (*22*). Another study with 40 patients showed that the use of triptan resulted in side effects. Although patients experienced headaches, flushing, dizziness, and shortness of breath, they did not experience strokes (*24*). These studies highlight the potential of triptans as an abortive medication to reduce frequency for hemiplegic migraines and underline the need for well-designed clinical studies with larger sample sizes as multiple studies, including the previously mentioned ones, stated limited theoretical statistical power. In addition to patient studies, more in vitro studies should be done to determine if triptans directly increase the risk of vascular events as there is currently no sufficient evidence to suggest this (*26*).

Anti-CGRP treatments

Calcitonin gene-related peptide (CGRP) is a neuropeptide involved in migraine physiology as it causes neurogenic inflammation through vasodilation, and can lead to hypersensitivity of the trigeminovascular system (*26*). With the understanding of the mechanism for the CGRP, researchers developed treatments against CGRP such as Lasmiditan, Ubrogepant, Rimegepant, and CGRP monoclonal antibodies (mAbs) as a method of migraine prevention (*27*). Benefits of anti-CGRP monoclonal antibodies include minimal drug interactions, fewer side effects, and a high affinity and selectivity for CGRP. Experts have stated that anti-CGRP mAbs have advantages such as ease of use, quick onset of action, and persistent efficacy (*28*). However, anti-CGRP mAbs should be avoided during pregnancy (*29*). Another anti-CGRP treatment that is a promising alternative to triptans is Lasmiditan, which is the first FDA-approved drug of the ditan category. Studies have revealed that Lasmiditan does not have any vasoconstrictive effects; thus, it can circumvent the risk of ischemic stroke that triptans have. The precise mechanism of action for this medication is not currently known. Lasmiditan has passed through phase 2 and phase 3 clinical trials showing efficacy, but also adverse side effects impacting the central nervous system, such as driving impairment (*30*). Although the short-term effects of anti-CGRP treatments have been well-established (*28,31*), more research needs to be done on the long-term effects and the exact mechanisms of action.

Visual treatment

As symptoms of photophobia are prevalent in migraine patients, researchers are investigating possible visual treatments. One potential treatment is to reduce the stimulation of photosensitive retinal ganglion cells by using an optical notch filter integrated into glasses lenses. A study with 37 patients showed that this filter reduced the Headache Impact Test (HIT-6) score for 23 patients, suggesting that these filters are a promising treatment. Another method of treatment is phototherapy for pain management. Green Light Emitting Diode (GLED) exposure significantly reduced headache pain intensity for patients with chronic headaches (*33*). GLED exposure functions by flashing light into the eyes of the user, which reduces migraine frequency and even stops migraines as they occur. The exact mechanism of how it accomplishes these results is not

clear, but it is hypothesized that the light enforces a regular brain alpha rhythm (*34*). Alpha rhythms are sinusoidal-like oscillations measured using an electroencephalogram, and irregular alpha rhythms can serve as a predictive measure for possible migraines during asymptomatic periods (*35*). The exact relationship between alpha rhythms and migraines is not currently known, but both are associated with each other. By treating the photophobic aspect of migraine symptoms, hemiplegic migraines could potentially be reduced while relieving pain. Although the results of these studies are promising, most studies have been conducted on rats. More clinical studies involving humans need to be conducted to fully understand how the visual treatment is related to migraine and pain physiology. Barriers to implementation of these treatments include creating inexpensive small portable devices that are user-friendly.

Peripheral neurostimulation

Peripheral neurostimulation (PNS) is a method of treating migraines using electrical pulses that stimulate certain areas of the brain and nervous system that might be related to migraine formation. One study showed that PNS leads to a decrease in Migraine Disability Assessment (MIDAS) scores for patients, with an average improvement of 88.7%. Patients also reported a decrease in the number of days with migraines from an average of 75.6 days to 37.5 days (*36*).

Occipital neurostimulation (ONS) and supraorbital neurostimulation (SNS) are forms of implanted neurostimulation. Combined occipital and supraorbital neurostimulation, both types of PNS, offer significant potential to decrease migraine frequency. According to case reports, patients who had both ONS and SNS achieved greater reduction in the frequency of migraines and severity of hemiplegia than if they had only one implant (*37*). Additionally, a 24-month study using ONS and epicranial temporal neurostimulation showed a reduction of migraine frequency by 50% (*38*). Reports show that the combination of treatments targeting different areas of the brain have been particularly helpful for migraines with auras, such as hemiplegic migraines. If patients do not want an implant, other options include occipital single-pulse transcranial magnetic stimulation (TMS) and transcutaneous supraorbital stimulation. TMS has been theorized to help migraine patients by modulating central pain areas and normalizing abnormalities of cortical responsivity between migraines (*39*). Thus, peripheral neurostimulation could provide another alternative to the standard pharmacological and prophylactic treatment for hemiplegic migraines. However, before peripheral neurostimulation can be fully implemented, clinical trials must be randomized and sham-controlled. Side effects such as lightheadedness, sinusitis, and drowsiness should also be taken into consideration (*39*).

PROHYLACTIC TREATMENT

Botox

Botox, although commonly used as a cosmetic treatment, can provide significant benefits for migraine patients by reducing muscle contractions (*40*). Possible mechanisms for botulinum toxin type-A (BoNT-A) include inhibiting the release of neurochemicals, proteins, proinflammatory cells, neurotransmitters, and excitatory neuropeptides of the motor, sensory, and nervous systems. It is theorized that the injection of BoNT-A reduces the input of pain signals from the trigeminal and cervical nerves that can lead to the hyperexcitability of neurons involved in migraine physiology (*41*). Each treatment consists of 31 to 39 injections around the head and neck and lasts for roughly 12 weeks (*42*). One study found that 9 of the 11 patients injected with Botox reported reduction of frequency and severity of headaches and aura after 3 months with few adverse effects (*43*). Currently, Botox is FDA-approved for those with chronic migraines, a condition for individuals older than 18 who have migraines at least 15 days a month (*44*). Additionally, a systematic review of 24 articles shows that BoNT-A is a low-cost option for the treatment of various kinds of migraines (*42*). Although there is significant research on the effect of Botox for other types of migraines, there are currently no studies focused specifically on hemiplegic migraines. As a result, the mechanism of how Botox alleviates hemiplegic migraine aura is currently unknown. However, it is theorized that Botox reduces CGRP and trigeminal sensory input, leading to changes in the cortical excitability threshold (*43*). To determine the validity of this mechanism, in-vitro studies should be conducted to observe if knocking down CGRP produces an effect in migraine models.

Marijuana

Marijuana or medical cannabis (MC) is a topic of interest because it may reduce the pain from migraines and ease symptoms such as nausea after 6 months of use by blocking cannabinoid receptors involved in nociceptive pain and targeting inflammatory components of migraines (*45*). Cannabidiol (CBD), a type of cannabinoid, is of particular interest because it does not cause intoxication like other cannabinoids. While medical marijuana seems to be a recent development, it has been used as a remedy for headaches throughout the nineteenth and twentieth centuries (*46*). Medical marijuana was shown to decrease the frequency of migraines by 40% (*47*). One review states that cannabis has a wide variety of effects including anticonvulsive, analgesic, antiemetic, and anti-inflammatory effects that make it promising as a migraine treatment (*48*). Although there are limited well-designed experimental studies to assess the effectiveness and safety of MC, a review of studies reveals that there are enough anecdotal results of analgesic effects from using marijuana or cannabidiol oil to warrant clinical trials (*48*). This speaks to the potential of marijuana as a treatment for migraines, including hemiplegic migraines. However, research into the use of marijuana in clinical practice is limited by politics. As of 2021, 36 states had legalized cannabis, but it still remains a Schedule 1 substance under the Federal Controlled Substance Act (*49*), resulting in limited research. Well-designed clinical trials are needed to better understand the efficacy and effects of cannabis long-term and on hemiplegic migraines specifically.

NON-PHARMACOLOGICAL INTERVENTIONS

Nutraceuticals

As with many other conditions and diseases, nutrition can influence the course of migraines. As discussed above, alcohol and caffeine are known diet-related triggers of migraines (*50*). Diet interventions such as low-fat diets and elimination diets are also related to a decrease in the frequency of migraines. Other common food triggers include chocolate, dairy products

such as aged cheese, food preservatives with nitrates and nitrites, monosodium glutamate (MSG), and artificial sweeteners such as aspartame (*51*). In addition to these common food triggers, there are also food triggers specific to individuals. It has been found that avoiding these food triggers can reduce migraines in 20-50% of patients. Patients can also benefit from nutrition supplements such as magnesium, riboflavin, and coenzyme Q10. A study found that taking these supplements reduced the average number of days per month with migraines from 6.2 to 4.4 and decreased scores on a maximal pain intensity scale for migraines (*52*). Other studies have shown that migraines may be reduced by focusing on the gut microbiota and gut-brain axis through consumption of fiber and a low glycemic-index diet, supplemented with vitamin D (*53*). Nutrition has been shown to be a promising alternative to pharmaceuticals due to its limited side effects, especially for pregnant women who are contraindicated for most medications. However more studies need to be conducted to determine the effects of nutrition and specific diets on the severity and duration of migraines (*54*).

Sleep

Maintaining a healthy sleep schedule can also reduce the frequency of migraines (*55*). The exact relationship between sleep and migraines is not clearly understood, but there is a clear correlation as studies have discovered that migraines are associated with sleep disorders, especially insomnia (*56*). One area of study is rapid eye movement (REM) sleep and how neurotransmitters released during sleep can lead to the development of migraines. A common pathophysiology between migraines and sleep problems may exist as parts of the brainstem-cortical networks, which are factors in the migraine pathway with signaling molecules such as serotonin and dopamine (*57*). A study at University of North Carolina, Chapel Hill demonstrated that a behavioral sleep intervention addressing harmful sleep habits associated with increased insomnia led to a reduction in headache frequency and headache index among those who had episodic migraines. Habits that were most effective in preventing migraines were sleeping 8 hours consistently, no electronics in bed, visualization techniques to shorten time to sleep onset, eating no later than four hours before bedtime, and eliminating naps (*58*). Helpful future studies would include identifying which behavior modifications are most useful in decreasing migraine frequencies and how to incorporate those modifications into a standardized protocol for treatment.

Exercise

There is a correlation between low levels of exercise and high frequencies of migraines (*59*). Exercise suppresses inflammatory modulators such as cytokines and cortisol as well as microvascular health, which may be implicated in cortical spreading depression and therefore impact the severity of migraines (*60*). Aerobic exercise has been investigated as a migraine-preventative treatment and has been found to produce notable improvements in migraine symptoms (*61*). Studies have shown that a rigorous aerobic regimen yields a statistically significant reduction in migraine frequency, intensity, and duration (*60*). Regular exercise has also been theorized to alter an individual's migraine triggering threshold. However, there are limitations on the type of physical activity as high-intensity exercise with insufficient warmup can also be a trigger (*59*). More research is needed with a focus on how frequency and intensity of exercise can influence migraines, and future studies could focus on determining how much exercise can be tolerated by migraine patients, how to standardize aerobic exercise, and the possible effects of aerobic exercise (*61*).

DISCUSSION

Abortive treatment, which involves stopping the migraine, and treatments that ease migraine symptoms are practical options for individuals with migraines. However, preventative and alternative treatments could be more conducive for patient health. For many, migraine drugs can have strong adverse side effects which can impact their ability to continue with medication. According to a survey, 67% of those who use prescription migraine medications delayed use out of concern for adverse side effects (*62*). Frequent use of medication can also lead to conditions such as medication overuse headache. Although non-drug treatments may take longer for the onset of benefits, their effects could be broader and more durable with fewer side effects (*63*).

Many factors that lead to migraines cannot be changed, such as genetics, age, and sex. Therefore, more flexible factors such as nutrition and sleep are the key to reducing frequency and severity of hemiplegic migraines for all patients. The exact pathophysiology of hemiplegic migraines is currently unknown, which presents a challenge for preventing these types of migraines, but research on more alternative treatments may help researchers deduce the proper mechanism and treatment plan. For example, studies involving Botox resulted in the creation of theories on the role of nerves and neurotransmitters in migraines and deepened the understanding of neurological symptoms associated with hemiplegic migraines. Thus, current research should focus more on the benefits of alternative treatments and nonpharmaceutical prevention. While there are studies comparing the effects of combining multiple pharmaceutical medications, there is a lack of research in the efficacy of combining common medications with other preventative treatments. Current studies on alternative treatments also have methodological limits, such as small sample sizes, little randomization, and a lack of appropriate control groups (*64*). Future studies with physician-confirmed diagnoses of participant hemiplegic migraines and follow-ups can address research gaps and study limitations. Medical professionals developing a protocol for treating patients with hemiplegic migraines should focus on upstream factors as well as symptoms. An example of a possible protocol that focuses on lifestyle modifications such as sleep, exercise, and diet is the SEEDS (Sleep, Exercise, Eat, Diary, Stress) mnemonic suggested by the Cleveland Clinic (*65*). With a holistic view towards research and treatment plans, patients suffering from hemiplegic migraines can more effectively manage their symptoms and improve their overall quality of life.

REFERENCES

1. P. Golden L. Peters, Migraine Overview and Summary of Current and Emerging Treatment Options. *Suppl Featur Publ.* **25** (2019) (available at https://www.ajmc.com/view/migraine-overview-and-summary--of-current-and-emerging-treatment-options).
2. X. H. Hu, L. E. Markson, R. B. Lipton, W. F. Stewart, M. L. Berger, Burden of migraine in the United States: disability and economic costs. *Arch Intern Med* **159**, 813–818 (1999). doi:10.1001/archinte.159.8.813
3. V. Di Stefano et al., Diagnostic and therapeutic aspects of hemiplegic migraine. *J Neurol Neurosurg Psychiatry*. **91**, 764–771 (2020). doi:10.1111/

head.12482
4. L. Lykke Thomsen et al., An epidemiological survey of hemiplegic migraine. *Cephalalgia Int J Headache.* **22**, 361–375 (2002). doi:10.1046/j.1468-2982.2002.00371.x.
5. J. C. Jen, "Familial Hemiplegic Migraine" in GeneReviews®, M. P. Adam, D. B. Everman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, A. Amemiya, Eds. (University of Washington, Seattle, Seattle (WA), 1993; http://www.ncbi.nlm.nih.gov/books/NBK1388/).
6. L. Thomsen, J. Olesen, Sporadic Hemiplegic Migraine. *Cephalalgia*. **24**, 1016–1023 (2004). doi:10.1111/j.1468-2982.2004.00788.x
7. O. Cozzolino, M. Marchese, F. Trovato, E. Pracucci, G. M. Ratto, M. G. Buzzi, F. Sicca, F. M. Santorelli, Understanding Spreading Depression from Headache to Sudden Unexpected Death. *Front Neurol*. **9** (2018) (available at https://www.frontiersin.org/articles/10.3389/fneur.2018.00019). doi:10.3389/fneur.2018.00019
8. M. B. Russell, A. Ducros, Sporadic and familial hemiplegic migraine: pathophysiological mechanisms, clinical characteristics, diagnosis, and management. *Lancet Neurol.* **10**, 457–470 (2011). doi:10.1016/S1474-4422(11)70048-5
9. S. M. Gritz, R. A. Radcliffe, Genetic effects of ATP1A2 in familial hemiplegic migraine type II and animal models. *Hum Genomics.* **7**, 8 (2013). doi:10.1186/1479-7364-7-8
10. M. Scheffer, A. van den Berg, M. D. Ferrari, Migraine strikes as neuronal excitability reaches a tipping point. *PloS One*. **8**, e72514 (2013). doi:10.1371/journal.pone.0072514
11. B. L. Peterlin, S. Gupta, T. N. Ward, A. MacGregor, Sex Matters: Evaluating Sex and Gender in Migraine and Headache Research. *Headache.* **51**, 839–842 (2011). doi:10.1111/j.1526-4610.2011.01900.x
12. T. W. Victor, X. Hu, J. C. Campbell, D. C. Buse, R. B. Lipton, Migraine prevalence by age and sex in the United States: a life-span study. *Cephalalgia Int J Headache*. **30**, 1065–1072 (2010). doi:10.1177/0333102409355601
13. M. Aggarwal, V. Puri, S. Puri, Serotonin and CGRP in Migraine. *Ann Neurosci.* **19**, 88–94 (2012). doi:10.5214/ans.0972.7531.12190210
14. J. J. Y. Ong, M. De Felice, Migraine Treatment: Current Acute Medications and Their Potential Mechanisms of Action. *Neurotherapeutics*. **15**, 274–290 (2018). doi:10.1007/s13311-017-0592-1
15. A. W. Hauge, M. Kirchmann, J. Olesen, Trigger factors in migraine with aura. *Cephalalgia Int J Headache*. **30**, 346–353 (2010). doi:10.1111/j.1468-2982.2009.01930.x
16. J. M. Hansen, A. W. Hauge, M. Ashina, J. Olesen, Trigger factors for familial hemiplegic migraine. *Cephalalgia Int J Headache.* **31**, 1274–1281 (2011). doi:10.1177/0333102411415878
17. N. Pelzer et al., Clinical spectrum of hemiplegic migraine and chances of finding a pathogenic mutation. *Neurology*. **90**, e575–e582 (2018). doi:10.1212/WNL.0000000000004966
18. R. M. Gallagher, R. Kunkel, Migraine medication attributes important for patient compliance: concerns about side effects may delay treatment. *Headache*. **43**, 36–43 (2003). doi:10.1046/j.1526-4610.2003.03006.x
19. M. A. Fischer, A. Jan, "Medication-overuse Headache" in StatPearls (StatPearls Publishing, Treasure Island (FL), 2022; http://www.ncbi.nlm.nih.gov/books/NBK538150/).
20. A. González-Hernández, B. A. Marichal-Cancino, A. MaassenVanDenBrink, C. M. Villalón, Side effects associated with current and prospective antimigraine pharmacotherapies. *Expert Opin. Drug Metab. Toxicol.* **14**, 25–41 (2018). doi:10.1080/17425255.2018.1416097
21. S. Evers et al., EFNS guideline on the drug treatment of migraine – revised report of an EFNS task force. *Eur J Neurol.* **16**, 968–981 (2009). doi:10.1111/j.1468-1331.2009.02748.x
22. V. Artto et al., Treatment of hemiplegic migraine with triptans. *Eur J Neurol.* **14**, 1053–1056 (2007). doi:10.1111/j.1468-1331.2007.01900.x
23. B. N. Mason, A. F. Russo, Vascular Contributions to Migraine: Time to Revisit? *Front Cell Neurosci.* **12**, 233 (2018). doi:10.3389/fncel.2018.00233
24. G. Singh, J. Sherman, G. Zinkhan, Are Triptans safe in Basilar and Hemiplegic Migraine? A retrospective chart review (S23.002). *Neurology*. **84** (2015) (available at https://n.neurology.org/content/84/14_Supplement/S23.002).
25. H.-C. Diener, The Risks or Lack Thereof of Migraine Treatments in Vascular Disease. *Headache*. **60**, 649–653 (2020). doi:10.1111/head.13749
26. S. Iyengar, K. W. Johnson, M. H. Ossipov, S. K. Aurora, CGRP and the Trigeminal System in Migraine. *Headache*. **59**, 659–681 (2019). doi:10.1111/head.13529
27. D. Szkutnik-Fiedler, Pharmacokinetics, Pharmacodynamics and Drug-Drug Interactions of New Anti-Migraine Drugs-Lasmiditan, Gepants, and Calcitonin-Gene-Related Peptide (CGRP) Receptor Monoclonal Antibodies. *Pharmaceutics*. **12**, 1180 (2020). doi:10.3390/pharmaceutics12121180
28. F. Cohen, H. Yuan, E. M. G. DePoy, S. D. Silberstein, The Arrival of Anti-CGRP Monoclonal Antibodies in Migraine. *Neurother J Am Soc Exp Neurother.* **19**, 922–930 (2022). doi:10.1007/s13311-022-01230-x
29. A. Negro, P. Martelletti, Patient selection for migraine preventive treatment with anti-CGRP(r) monoclonal antibodies. *Expert Rev Neurother.* **19**, 769–776 (2019). doi:10.1080/14737175.2019.1621749
30. J. K. Beauchene, T. L. Levien, Lasmiditan: Acute Migraine Treatment Without Vasoconstriction. A Review. *J Pharm Technol JPT Off Publ Assoc Pharm Tech.* **37**, 244–253 (2021). doi:10.1177/87551225211024630
31. T. de Vries, C. M. Villalón, A. MaassenVanDenBrink, Pharmacological treatment of migraine: CGRP and 5-HT beyond the triptans. *Pharmacol Ther.* **211**, 107528 (2020). doi:10.1016/j.pharmthera.2020.107528
32. R. N. Hoggan et al., Thin-film optical notch filter spectacle coatings for the treatment of migraine and photophobia. *J Clin Neurosci.* **28**, 71–76 (2016). doi:10.1016/j.jocn.2015.09.024
33. K. Cheng, L. F. Martin, H. Calligaro, A. Patwardhan, M. M. Ibrahim, Case Report: Green Light Exposure Relieves Chronic Headache Pain in a Colorblind Patient. *Clin. Med. Insights Case Rep.* **15**, 11795476221125164 (2022). doi:10.1177/11795476221125164
34. M. M. Ibrahim et al., Long-lasting antinociceptive effects of green light in acute and chronic pain in rats. *Pain.* **158**, 347–360 (2017). doi:10.1097/j.pain.0000000000000767
35. T. Nyrke, P. Kangasniemi, H. Lang, Alpha rhythm in classical migraine (migraine with aura): abnormalities in the headache-free interval. *Cephalalgia Int J Headache.* **10**, 177–181 (1990). doi:10.1046/j.1468-2982.1990.1004177.x
36. C. A. Popeney, K. M. Aló, Peripheral Neurostimulation for the Treatment of Chronic, Disabling Transformed Migraine. *Headache J Head Face Pain.* **43**, 369–375 (2003). doi:10.3390/biomedicines9111588
37. K. L. Reed, K. R. Will, F. Conidi, R. Bulger, Concordant Occipital and Supraorbital Neurostimulation Therapy for Hemiplegic Migraine; Initial Experience; A Case Series. *Neuromodulation Technol Neural Interface.* **18**, 297–304 (2015). doi:10.1111/ner.12267
38. K. K. Deshpande, K. L. Wininger, Feasibility of combined epicranial temporal and occipital neurostimulation: treatment of a challenging case of headache. *Pain Physician.* **14**, 37–44 (2011). doi:10.1111/ner.12267
39. J. Schoenen, B. Roberta, D. Magis, G. Coppola, Noninvasive neurostimulation methods for migraine therapy: The available evidence. *Cephalalgia Int J Headache.* **36**, 1170–1180 (2016). doi:10.1177/0333102416636022
40. W. J. Binder, M. F. Brin, A. Blitzer, L. D. Schoenrock, J. M. Pogoda, Botulinum toxin type A (BOTOX) for treatment of migraine headaches: an open-label study. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg.* **123**, 669–676 (2000). doi:10.1067/mhn.2000.110960
41. R. Burstein, A. M. Blumenfeld, S. D. Silberstein, A. Manack Adams, M. F. Brin, Mechanism of Action of OnabotulinumtoxinA in Chronic Migraine: A Narrative Review. *Headache*. **60**, 1259–1272 (2020). doi:10.1111/head.13849
42. N. Shaterian et al., Botox (OnabotulinumtoxinA) for Treatment of Migraine Symptoms: A Systematic Review. *Pain Res Manag.* 2022, 3284446 (2022). doi:10.1155/2022/3284446
43. T. Y. Chen, I. Garza, D. W. Dodick, C. E. Robertson, The Effect of OnabotulinumtoxinA on Aura Frequency and Severity in Patients With Hemiplegic Migraine: Case Series of 11 Patients. *Headache*. **58**, 973–985 (2018). doi:10.1111/head.13317
44. C. P. Herd et al., Botulinum toxins for the prevention of migraine in adults. *Cochrane Database Syst Rev.* 2018, CD011616 (2018). doi:10.1002/14651858.CD011616.pub2
45. B. O. Okusanya, B. E. Lott, J. Ehiri, J. McClelland, C. Rosales, Medical Cannabis for the Treatment of Migraine in Adults: A Review of the Evidence. *Front Neurol.* **13**, 871187 (2022). doi:10.3389/fneur.2022.871187
46. E. Russo, Cannabis for migraine treatment: the once and future prescription? An historical and scientific review. *Pain*. **76**, 3–8 (1998). doi:10.1016/s0304-3959(98)00033-5
47. D. N. Rhyne, S. L. Anderson, M. Gedde, L. M. Borgelt, Effects of Medical Marijuana on Migraine Headache Frequency in an Adult Population. *Pharmacotherapy.* **36**, 505–510 (2016). doi:10.1002/phar.1673
48. P. Leimuranta, L. Khiroug, R. Giniatullin, Emerging Role of (Endo)Cannabinoids in Migraine. *Front Pharmacol.* **9**, 420 (2018). doi:10.3389/fphar.2018.00420

49. A. I. Perlman et al., Medical Cannabis State and Federal Regulations: Implications for United States Health Care Entities. *Mayo Clin Proc.* **96**, 2671–2681 (2021). doi:10.1016/j.mayocp.2021.05.005
50. N. A. Hindiyeh et al., The Role of Diet and Nutrition in Migraine Triggers and Treatment: A Systematic Literature Review. *Headache*. **60**, 1300–1316 (2020). doi:10.1111/head.13836
51. Z. Zaeem, L. Zhou, E. Dilli, Headaches: a Review of the Role of Dietary Factors. *Curr Neurol Neurosci Rep*. **16**, 101 (2016). doi:10.1007/s11910-016-0702-1
52. C. Gaul, H.-C. Diener, U. Danesch, Migravent® Study Group, Improvement of migraine symptoms with a proprietary supplement containing riboflavin, magnesium and Q10: a randomized, placebo-controlled, double-blind, multicenter trial. *J Headache Pain.* **16**, 516 (2015). doi:10.1186/s10194-015-0516-6
53. P. Gazerani et al., A randomized, double-blinded, placebo-controlled, parallel trial of vitamin D3 supplementation in adult patients with migraine. *Curr Med Res Opin.* **35**, 715–723 (2019). doi:10.1080/03007995.2018.1519503
54. K. Kaur et al., The Efficacy of Herbal Supplements and Nutraceuticals for Prevention of Migraine: Can They Help? *Cureus*. **13**, e14868. doi:10.7759/cureus.14868
55. C. Tiseo, A. Vacca, A. Felbush, T. Filimonova, A. Gai, T. Glazyrina, I. A. Hubalek, Y. Marchenko, L. H. Overeem, S. Piroso, A. Tkachev, P. Martelletti, S. Sacco, Migraine and sleep disorders: a systematic review. *J Headache Pain.* **21**, 126 (2020). doi:10.1186/s10194-020-01192-5
56. S. J. Kim, K.-T. Han, S.-Y. Jang, K.-B. Yoo, S. J. Kim, The Association between Migraine and Types of Sleep Disorder. *Int J Environ Res Public Health.* **15**, 2648 (2018). doi:10.3390/ijerph15122648
57. A. Vgontzas, J. M. Pavlović, Sleep Disorders and Migraine: Review of Literature and Potential Pathophysiology Mechanisms. *Headache*. **58**, 1030–1039 (2018). doi:10.1111/head.13358
58. A. H. Calhoun, S. Ford, Behavioral sleep modification may revert transformed migraine to episodic migraine. *Headache*. **47**, 1178–1183 (2007). doi:10.1111/j.1526-4610.2007.00780.x
59. F. M. Amin et al., The association between migraine and physical exercise. *J. Headache Pain.* **19**, 83 (2018). doi:10.1186/s10194-018-0902-y
60. M. Barber, A. Pace, Exercise and Migraine Prevention: a Review of the Literature. *Curr Pain Headache Rep.* **24**, 39 (2020). doi:10.1007/s11916-020-00868-6
61. T.-J. Song, M. K. Chu, Exercise in Treatment of Migraine Including Chronic Migraine. *Curr. Pain Headache Rep*. **25**, 14 (2021). doi:10.1007/s11916-020-00929-w
62. R. M. Gallagher, R. Kunkel, Migraine medication attributes important for patient compliance: concerns about side effects may delay treatment. *Headache*. **43**, 36–43 (2003). doi:10.1046/j.1526-4610.2003.03006.x
63. R. E. Wells, E. Loder, Mind/Body and behavioral treatments: the evidence and approach. *Headache*. **52** Suppl 2, 70–75 (2012). doi:10.1111/j.1526-4610.2012.02238.x
64. R. E. Wells, J. Beuthin, L. Granetzke, Complementary and Integrative Medicine for Episodic Migraine: An Update of Evidence from the Last 3 Years. *Curr Pain Headache Rep.* **23**, 10 (2019). doi:10.1007/s11916-019-0750-8
65. J. Robblee, A. J. Starling, SEEDS for success: Lifestyle management in migraine. *Cleve Clin J Med.* **86**, 741–749 (2019). doi:10.3949/ccjm.86a.19009

Characterizing Structural Changes in Cytochrome *C* Protein Induced by ATP

Sahana S. Shah[1], Joseph A. Loo[2]

[1]Department of Biochemistry and Biophysics, Oregon State University. [2]Department of Chemistry and Biochemistry, University of California, Los Angeles.

ABSTRACT

Adenosine 5'-triphosphate (ATP) plays a key role in storing and transferring energy in cells, allowing important biochemical reactions to occur in living things. It has been shown that ATP can effectively stabilize protein structure and inhibit aggregation. When this occurs, ATP is hydrolyzed to adenosine diphosphate (ADP), releasing energy which the protein utilizes to change shape or make the enzyme catalytically active, therefore altering its function. Lysozyme, ubiquitin, and malate dehydrogenase were three proteins found to contain ATP-susceptible regions characterized by high flexibility and an abundance of charged residues. ATP cluster formation mediated by cations was found to suppress the structural fluctuations of malate dehydrogenase and promote thermal stability, helping to maintain the structure and overall function of this enzyme within the citric acid cycle and central metabolic pathway. Another protein that interacts with ATP is cytochrome **c** (cyt*c*), a soluble heme protein that plays a key role in the electron transport chain of mitochondria. Previously, mechanisms of cyt*c* dimerization have been discussed, which can alter the effectiveness of cyt*c* as an electron carrier, but none have discussed the potential role of ATP-facilitated dimerization. To better understand this phenomenon, native top-down mass spectrometry and Fourier Transform Ion Cyclotron Resonance were employed as tools to probe structural changes of bovine, equine, and pigeon cyt*c* in the presence of ATP. It was found that the interaction of cyt*c* with ATP occurred across all three species and was detected in the presence of a cyt*c* dimer in solution. However, ATP was not found to induce cyt*c* dimerization. This study has the potential to explore whether cyt*c* dimerization, which could alter cyt*c* functionality in the electron transport chain, functions within the complex network of human metabolic pathways.

INTRODUCTION

Adenosine 5'-triphosphate (ATP) is an important energy molecule for crucial biological processes, such as protein synthesis, muscle contraction, and the transport of molecules across cell membranes. The function of many proteins is highly dependent on their interaction with ligands, with ATP being an important ligand that often functions as a coenzyme (*1*). Based on previous studies, ATP has been shown to exhibit additional physiological roles by acting as an effective cosolvent that stabilizes protein structure, inhibiting liquid-liquid phase separation, and enhancing protein solubility (*2*).

Additionally, an ATP concentration of only ~100 μM is needed for most enzymes that utilize ATP, while the physiological concentration is 10 to 100-fold higher, between 1 and 10 mM. Recent biochemical and molecular dynamics simulations have shown that ATP can effectively stabilize protein structure and inhibit aggregation at a concentration less than 10 mM, which is significantly lower than cosolvent concentrations used in standard mechanisms and suggests high efficiency in the mechanism of ATP-induced protein stabilization (*2*). Based on the wide range of ATP functionality, it is important to examine the interactions between proteins of interest and ATP to study how this affects the function of the protein in biological pathways.

In a recent study, the interactions of ATP with lysozyme, ubiquitin, and malate dehydrogenase were studied (Figure 1). Malate dehydrogenase was found to contain ATP susceptible regions characterized by high flexibility and an abundance of charged residues. Through molecular dynamics simulations, it was shown that ATP binding promoted the thermal stability of proteins by forming cation mediated clusters around the protein, binding to surface regions with high flexibility and high degrees of hydration. Given that the working temperature of the mitochondria is several degrees higher than atmospheric temperature, the role of ATP in enhancing thermal stability is crucial. The high concentration of ATP in the mitochondria contributes to maintaining the structure and function of malate dehydrogenase, allowing the protein to play a proper role in regulating the citric acid cycle (*2*).

Another protein of interest that interacts with ATP is cytochrome *c* (cyt*c*), a soluble heme protein that plays a key role in the electron transpo3t chain of mitochondria to produce ATP and is an initiator of cellular apoptosis. It has been proposed that the role of ATP interaction with cyt*c*, which occurs when the phosphorylation potential of the system is high under physi-

ological conditions, can decrease electron flow in the mitochondrial electron transport chain of cellular respiration (*3*). An enhanced understanding of how ATP binds and affects components of the central metabolic pathway is needed to explore the unique mechanism in which ATP affects protein structure.

To better characterize protein structural changes, the Loo Lab at UCLA monitored oligomerization state alterations in cytc in the absence and presence of ATP in solution. The self-association of proteins to form higher-order oligomers, particularly dimers, is a very common phenomenon that is a key factor in the regulation of proteins (*4*). Previous work has discussed ways that cytc can potentially dimerize, but the potential role of ATP-induced dimerization remains poorly understood. A recent study discussed an alternate conformer of cytc called a domain swapped dimer, in which two proteins exchange an identical element to form a dimer, was shown to control the activation of apoptosis by acting as a pH-inducible switch and oxygenating cardiolipin, a component of the inner mitochondrial membrane (*5*). By integrating knowledge about the role of ATP in protein stabilization and structural conformations of cytc, it was hypothesized that ATP could interact with cytc and promote dimerization of this protein, potentially inhibiting the function of cytc as an electron carrier (*6*).

To better understand this phenomenon, mass spectrometry was employed to probe structural changes of cytc in the presence of ATP. Native mass spectrometry (MS) and native top-down MS are primarily used for structural biology studies of macromolecular complexes and can be used to measure protein molecular weight, identify post-translational modifications, and to help determine the stoichiometry of ligand binding and ligand binding sites (*7*).

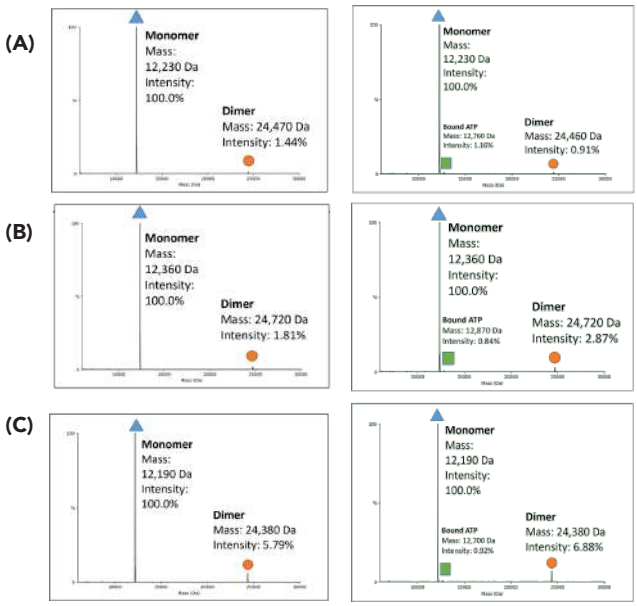


Figure 1: MS1 Spectra shows that ATP does not induce cytc dimerization.
Left- and right-side spectra show monomeric/dimeric states and intensities for each cytc species (absence of ATP and presence of ATP, 1:5 cytc: ATP ratio). Data acquired using Time of Flight (TOF) Mass Spectrometer and deconvoluted using UniDec (Marty *et al.*, 2015). Peak Detection Threshold of 0.005. **(A)** Bovine cytc **(B)** Equine cytc **(C)** Pigeon cytc

This study attempted to elucidate structural changes of bovine, equine, and pigeon cytc that were potentially induced by the presence of ATP in solution. This study demonstrates the possible biological relevance of cytc dimerization, which is not fully understood, and better characterizes the interaction of ATP with cytc by employing MS methods. Given that the synthesis of ATP requires the function of cytc in the electron transport chain, it is possible that ATP-induced dimerization of cytc could cause a buildup of ATP in the cell, therefore inhibiting electron transport chain function and altering the effectiveness of cytc as an electron carrier in this process (*6*).

METHODS

Materials for native top-down MS experiments

Cytc proteins from bovine heart, cytc from equine heart, and cytc from pigeon breast muscle were purchased from Sigma-Aldrich (St. Louis, MO, USA) and used without further purification. Proteins were chosen based on their commercial availability and polypeptide sequences, dictated by species of origin. LC-MS grade water was obtained from Fisher Chemical (Hampton, NH, USA). ATP·Mg^{2+} was obtained from Sigma-Aldrich and a 200 μM stock solution was prepared.

Sample preparation for native top down MS1 experiments

Cytc from bovine heart, cytc from equine heart, and cytc from pigeon breast muscle were dissolved in LC-MS grade water at a concentration of 1 mg/mL. These solutions were desalted with 10K Amicon Ultra centrifugal filter units (Millipore Corp, Billerica, MA, USA) and buffer exchanged into 20 mM ammonium acetate. The samples were diluted to 10 μM for analysis on the Waters Synapt G2-Si High-Definition Mass Spectrometer (HDMS quadrupole time-of-flight instrument) (Milford, MA, USA). Cytc and ATP·Mg2+ were prepared at a 1:1, 1:2, and 1:5 ratio of molar concentration (cytc/ATP·Mg2+ = 1:1, 1:2, 1:5). This was done to observe if the abundance of the cytc dimer would change in response to varying the relative ATP concentration. The 1:2 and 1:5 ratios of molar concentration were further diluted to 5 μM for experimentation (cytc/ATP·Mg2+ = 5:10 μM, 5:25 μM). For analysis on the Bruker solariX 15-Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) (Billerica, MA), samples were buffer exchanged into 20 mM ammonium acetate, diluted to 5 μM and prepared at a 1:10 ratio of molar concentration (cytc/ATP·Mg^{2+} = 1:10).

Mass spectrometry for native top-down MS experiments

Protein samples were loaded into in-house pulled glass capillaries coated with platinum (Harvard Apparatus, Holliston, MA) and analyzed with nano electrospray ionization (nanoESI) by applying a voltage between 1.3 and 1.5 kV using the Waters Synapt G2-Si HDMS, with a m/z range from 100-5000. On this instrument, a sampling cone value of 20-30 V, a source offset of 10-20 V, a temperature of 130-150°C, and a Nano Flow gas value of 0-0.2 bar were applied. Between 120-300 scans were acquired and averaged. Data were deconvoluted and analyzed using the Unidec software program (*8*). These parameter values were chosen based on standard instrument practice and fine-tuned to optimize ionization of cytc, therefore producing the most stable spectrum possible.

Sample preparation for denaturing MS experiments using FTICR-MS

For a top-down MS experiment utilizing the Bruker solariX 15-Tesla FTICR-MS, each species of cyt*c* (1 mg/mL) was dissolved in a denaturing solution of 50% H_2O, 50% acetonitrile (ACN), and 0.1% acetic acid (CH_3COOH, pH~3).

Mass spectrometry for native and denaturing top-down MS experiments using FTICR-MS

Additional experiments were conducted utilizing the Bruker solariX 15-Tesla FTICR-MS. Protein samples were loaded into in-house pulled glass capillaries coated with platinum and electrosprayed by applying a voltage between 0.7-0.9 kV on the ESI capillary. MS and MS/MS spectra were collected for denatured bovine, equine, and pigeon cyt*c*. For each spectrum, up to 150 scans were obtained and averaged. All data were deconvoluted using the SNAP 2.0 algorithm from the Bruker Daltonics Data Analysis Software. For the MS/MS spectra, fragments were generated, verified manually, and matched using the ProSight Lite program with a fragmentation ion matching tolerance of 15 ppm. ProSight Lite is a freely available software tool to aid the identification of MS fragments that correlate to protein sequence, confirming manual assignments of the fragments.

RESULTS

Native top-down mass spectrometry confirms cyt*c* dimer formation

Cyt*c* dimerization was detected in solution by comparing native MS spectra of bovine, equine, and pigeon cyt*c* in the absence of ATP. By applying native MS using nanoESI, it is possible to gain valuable insight about the properties of the protein, including exact mass, the presence of post-translational modifications, and binding. The 7+ charge state was the most abundant ion of cyt*c* for all three species (Figure 3). Across the three species, the presence of a small amount of the cyt***c*** dimer was detected (Figure 1).

Presence of ATP in solution does not induce cyt*c* dimerization

As the concentration of ATP increased in relation to cyt*c*, there was no consistent or significant change in cyt***c*** dimer intensity, which never exceeded 10% across all three species (Figure 1, 2). In comparison to bovine and equine cyt*c*, the pigeon cyt*c* dimer was consistently higher in intensity across all cyt*c*: ATP ratios (Figure 1, 2). FTICR-MS was also utilized to analyze a 1:10 ratio of cyt*c*: ATP. The charge state distribution is consistent between species, and the dimeric states also lack intensity across all three species (Figure 3).

FTICR-MS confirms differences in amino acid sequences across species

By utilizing FTICR-MS, collisionally activated dissociation (CAD) to generate MS/MS fragmentation spectra was applied to all three species. Several B and Y product ions resulting from the proteins fragmenting by CAD could be matched to the corresponding polypeptide sequence. The denatured MS/MS spectra helps to confirm differences between the cyt*c* species, based on the amino acid sequences and unique fragmentation patterns observed (Figure 4, 5).

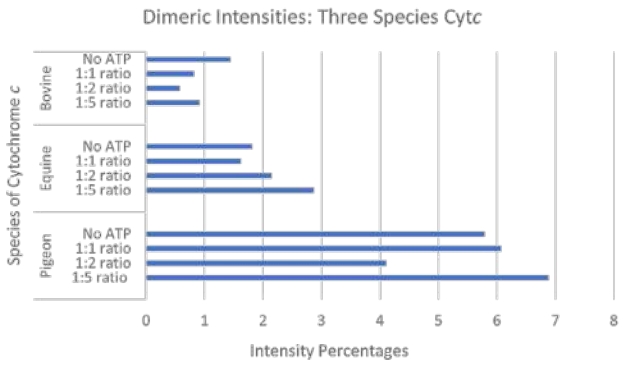


Figure 2: ATP does not induce cyt*c* dimerization.
Bar chart denoting dimeric intensities as percentages in the absence and presence of ATP for bovine, equine, and pigeon cyt*c*. As shown by the data, there is no correlation between the low cyt*c* dimeric intensity percentages and the amount of ATP in solution. This supports that ATP does not seem to induce cyt*c* dimerization.

DISCUSSION

The aim of this research was to study cyt*c* in the presence of different concentrations of ATP and determine any changes in the dimerization relative abundance of the protein. Based on the native MS spectra of cyt*c* in the absence and presence of ATP across three different species, it was concluded that cyt*c* and ATP associate across all three species and cyt*c* dimers form in solution. However, ATP does not seem to specifically induce cyt*c* dimerization. This is supported by the low intensity of cyt*c* dimer and the lack of correlation in the rise or fall of these intensities as the amount of ATP in solution increases. Each species exhibits slightly different monomeric and dimeric mass values based on differences in amino acid sequences. Ultimately, with the obtained results, it can be concluded that ATP lacks a role in inducing cyt*c* dimerization.

MS/MS (MS2) spectra were acquired to confirm the differences in sequences between each species. B ions are the fragment ions when the charge is retained on the N-terminus, while Y ions are the fragments when the charge is retained on the C-terminus (*9*). The differences in amino acid sequences can best be confirmed by directly comparing the fragmentation marks (denoted in blue) within the sequence images (Figure 5). Overall, the presence of N-terminal acetylation is consistently reflected in each species, in addition to two cysteines at the same amino acid residues that covalently link the heme group. In CAD, peptide bond cleavage is governed by the mobility of a proton added during ionization (*10*). The proton affinities of each peptide's side chain determine the most stable locations for available protons, influencing at which amide bonds the fragmentation will occur. For example, cleavages that are located C terminal to acidic residues dominate the spectra for peptides with a localized proton while cleavages that are N terminal to proline dominate the spectra for peptides with a mobile proton (*11*). Overall, this experiment helps to confirm the differences in cyt*c* species on a residue level and shows where fragmentation occurs in the sequence.

In this study, cyt*c* dimers were observed at low intensity levels and are different from dimers described in the literature. Cyt*c* dimerization is a phenomenon that has been previously observed

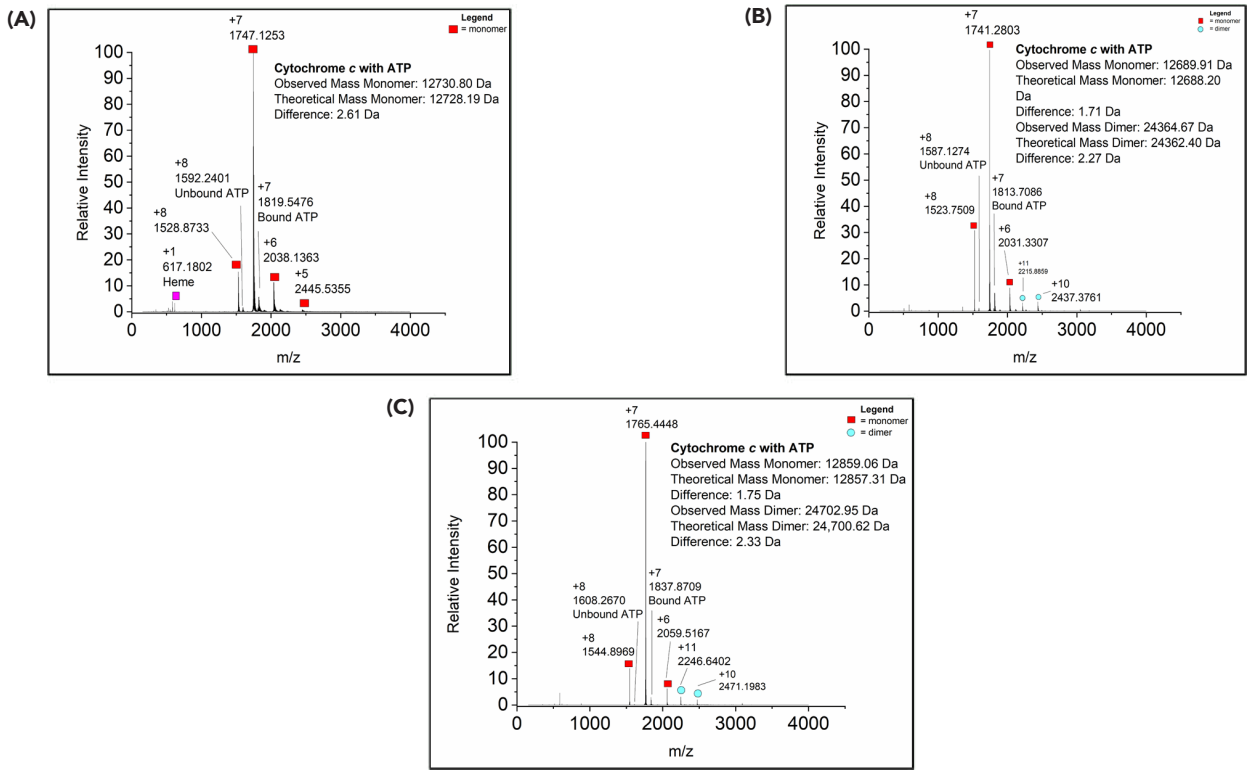


Figure 3: FTICR MS1 spectra supports that ATP does not induce cytc dimerization.
MS1 spectra for native cytc species in presence of ATP (1:10 ratio). Monomeric and dimeric state shown by red and blue marker, respectively. Heme group represents pink marker. Data acquired using FTICR-MS. **(A)** Bovine cytc- Only the monomer was detected in solution; ATP does not induce cytc dimerization. **(B)** Equine cytc- Dimer lacks intensity near 100%; ATP does not induce cytc dimerization. **(C)** Pigeon cytc- Dimer lacks intensity near 100%; ATP does not induce cytc dimerization.

(*5*). Cytc has been shown to oxidize cardiolipin (CL) and initiate the intrinsic apoptotic pathway. A domain-swapped dimer (DSD) of cytc has previously been reported and was proposed to have evolved into a pH-inducible switch to control the activation of apoptosis near pH 7.0. It has been proposed that cytc DSD has a set of properties that provide tighter regulation of the intrinsic pathway of apoptosis, and it has been discussed that different species of cytc can influence dimer stability. For example, human DSDs have been shown to be more kinetically stable than horse and yeast DSDs. Domain-swapped protein dimers have become common and are believed to have biological significance by creating access to the heme group and increasing peroxidase activity of the cytc in relation to the monomer (*5*). The homodimer formation detected in this study appears to be different from the formation of DSDs discussed in the previous study. This suggests the presence of a cytc dimer with a different structure and signifies a possible new biological function. From the current work, it seems that ATP does not induce cytc dimerization. More studies will be necessary to elucidate the potential biological role of the cytc dimer detected in this work.

It is important to note that the presence of the charge distribution representing a cytc dimer differed slightly in the data collected with a quadrupole time-of-flight instrument. Comparing the native spectra acquired using FTICR-MS in a 1:10 protein:ATP ratio, it can be observed that across the three species, there is a slight difference in the presence of the monomeric and dimeric charge distributions. As an example, pigeon cytochrome *c* contains a peak at *m/z* 2437 that is likely to be both the 10+ dimer and 5+ monomer due to the presence of a 11+ dimer peak and 6+ monomer peak (Figure 3C). A dimeric charge state (Eq. 1) is reduced to the same ratio for a monomeric charge state (Eq. 2). This means that a mixture of the monomeric and dimeric states exists in solution. It is possible that the FTICR-MS used for the present work is too *harsh*, or too energetic for weakly-bound cytc dimers to survive the gas-phase measurement. However, it is known that other weakly bound protein complexes can be measured using the FTICR-MS. More testing of the differences between the two instrument platforms is necessary to address this question.

$$\frac{m}{z} = \frac{2M}{+10} \tag{1}$$

$$\frac{m}{z} = \frac{M}{+5} \tag{2}$$

In this work, it is important to recognize the inherent discrepancies that exist between different mass spectrometers, based on variations in resolution, sensitivity, and accuracy. This study, which utilized two different mass spectrometers, reflects that ATP does not seem to induce cytc dimerization on a scale or uti-

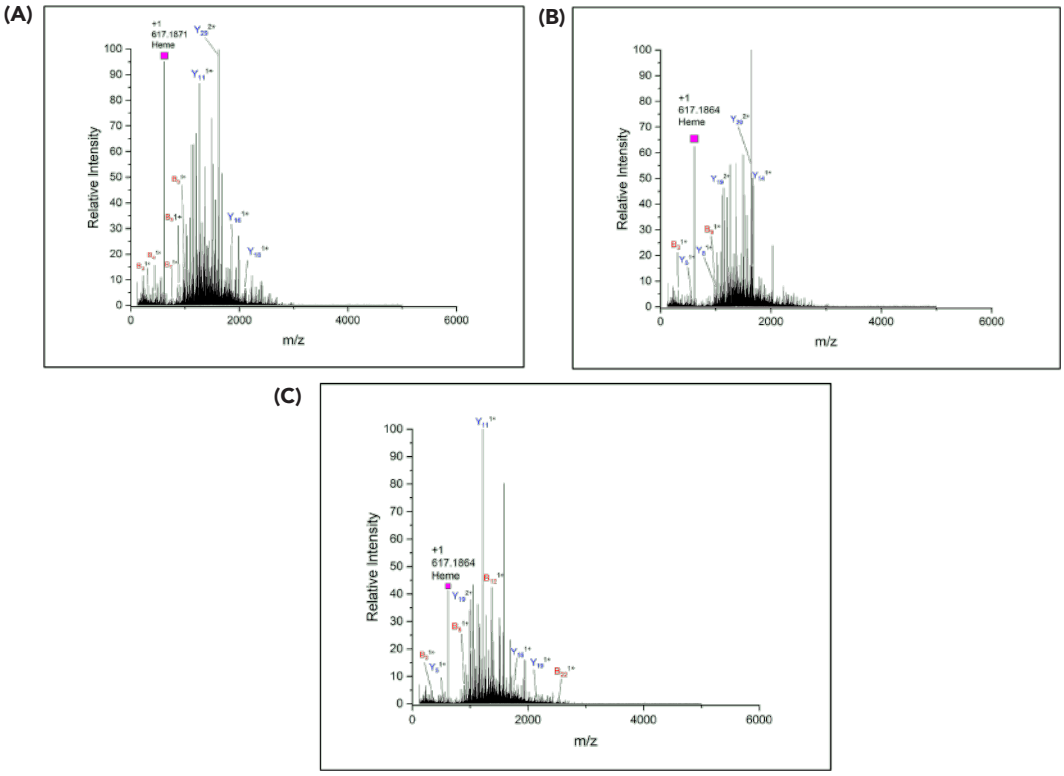


Figure 4: FTICR MS2 spectra with B and Y ion assignments show various fragments for cytc sequences.
MS2 spectra shown for three species of denatured cytc with selected B and Y ions shown with fragment position (numerical subscript). Heme group represents pink marker. Data acquired using FTICR-MS with CAD conditions. **(A)** Bovine cytc (Isolated 1359 *m/z*, 37 V) denotes B/Y ions assignments with unique fragment positions, showing an individualized sequence. **(B)** Equine cytc (Isolated 1374 *m/z*, 35 V) denotes B/Y ions assignments with unique fragment positions, showing an individualized sequence. **(C)** Pigeon cytc (Isolated 1355 *m/z*, 35 V) denotes B/Y ions assignments with unique fragment positions, showing an individualized sequence.

lizing a pattern that is easily observable. Native MS has evolved to acquire multi-layer structural information for proteins, including composition, stoichiometry, and topology, but also sequence, post-translational modifications, and ligand binding sites. In this study, the Waters Synapt G2-Si platform is a time of flight (TOF) mass spectrometer that allows macromolecular assemblies to be analyzed with fast-scanning speeds and no theoretical upper mass limit (*7*). However, compared to FTICR instruments, which are known for their ultrahigh mass-resolving power and accuracy, the resolving power of TOF analyzers is lower.

To address restricted resolving power of TOF instruments, current developments in methodology have worked to extend ion transport distance, but the capability of TOF to determine accurate molecular weights is still somewhat limited by salt and buffer adducts. To improve the analysis of macromolecular assemblies, the applied magnetic field induction of the instrument has been refined and increased over time (*7*). Additionally, although FTICR exhibits higher ion sensitivity, analysis utilizing this technique is limited by space-charge effect, which causes frequency shifts due to Coulombic interactions between ions in the analyzer cell. To combat this issue, it is sometimes necessary to reduce the number of ions in the cell (*9*). It is important to address the limitations of the instruments in the context of this study, as different intensities in the dimeric state peak between experiments and between instruments have been observed.

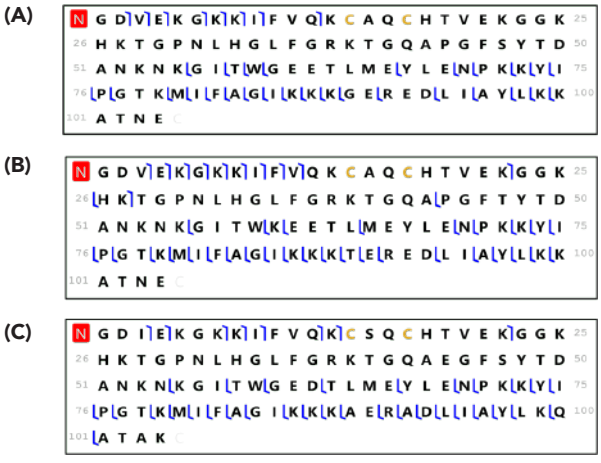


Figure 5: Sequence maps show amino acid differences in cytc species.
Sequence maps denote amino acid sequences for cytc species. N-terminal acetylation is denoted in red and there are two cysteines that covalently link the heme group, shown in gold. **(A)** Bovine cytc shares high similarity with amino acid sequence for the equine species, yet contains different fragmentation patterns (blue labels), confirming uniqueness of bovine cytc. **(B)** Equine cytc shares high similarity with amino acid sequence for the bovine species, yet contains different fragmentation patterns (blue labels), confirming uniqueness of equine cytc. **(C)** Pigeon cytc sequence is overall more distinct from bovine and equine with a distinct fragmentation pattern (blue labels), confirming uniqueness of pigeon cytc.

It is also crucial to note the differences in charge state distribution based on different solution conditions. In this study, mass spectra were acquired under native and denaturing solution conditions. In denaturing conditions, there are higher and more charge states due to the protein unfolding event, reflecting signals that are lower intensity and appear at a lower *m/z* (Figure 4). Under native conditions, there are lower and less charge states, causing the signals to have higher intensity and appear at a higher *m/z*. Within each condition, the charge state distributions are generally consistent between species, reflecting the similarity between the three cyt*c* species tested. This reflects that cyt*c* is a highly conserved protein regardless of experimental conditions applied. The spectra obtained under native and denaturing conditions demonstrate that ATP does not seem to induce cyt*c* dimerization, the central question of the study.

In general, protein dimerization is a key factor in the regulation of a wide variety of proteins. Self-association between proteins can help to minimize genome size while maintaining complex formation. In addition, dimerization has been shown to improve stability and control overactive site accessibility, while producing new sites for allosteric regulation. It has also been shown that the mechanism of dimerization seems to be specific to the protein involved and that dimer interfaces have evolved to optimize association (*4*). It is critical to investigate the mechanisms of ligand binding, allosteric modulation, and dimerization in the central metabolic pathway, leading to a better understanding of how this process is regulated and further insight into the creation of continuous energy for cellular function.

For next steps, it could be beneficial to investigate the binding of cyt*c* oxidase to ATP. It has been previously shown that cyt*c* oxidase acts as a site of regulation of oxidative phosphorylation through feedback inhibition by ATP, and an ATP binding site has been characterized on cyt*c* in different species such as horse and yeast. Additionally, previous work has shown that cyt*c* oxidase has a dimeric structure. However, in this study, it had also been shown that inhibition experiments using free ligands could not distinguish between ATP binding of cyt*c* and cyt*c* oxidase (*12*). Thus, the role of ATP binding should be investigated further to delineate between the interactions with these two proteins. As another potential substrate, ADP could be bound to cyt*c* to observe if the intensity of dimerization changes. Cyt*c* could also be studied in the presence of cardiolipin, and native MS could be used to confirm dimerization, since it is already known that cyt*c* dimerizes in the presence of this lipid. In addition, a biochemical assay could be performed to better characterize the interaction of ATP with potential binding partners, including cyt*c* and proteins such as malate dehydrogenase, a key enzyme within the central metabolic pathway. Finally, if the intensity of the ATP induced dimer is optimized, chemical crosslinking could be applied to stabilize the dimer, and bottom-up MS could be performed to identify the structural regions of the binding interface on an amino acid level.

Further studies can contribute to a better understanding of the function of cyt*c* and the role of cyt*c* dimerization. Overall, characterizing cytochrome *c* dimerization and continuing to investigate the role of ATP in this mechanism remains imperative to better understand the role of cyt*c* in the key pathway of oxidative phosphorylation.

ACKNOWLEDGEMENTS

This work was supported by the Amgen Scholars Program (to S.S.) and by the National Institutes of Health (NIH; grant R01GM103479 to J.L.). We acknowledge the contributions and help from the Loo Lab members, including Jessie Le, Carter Lantz, Eileen Olivares, and Merin Rixen.

REFERENCES

1. J. S. Chauhan, N. K. Mishra, G. P. Raghava, Identification of ATP Binding Residues of a Protein from Its Primary Sequence. *BMC Bioinformatics*. **10,** 434 (2009). doi:10.1186/1471-2105-10-434
2. X. Ou et al., ATP Can Efficiently Stabilize Protein Through a Unique Mechanism. *J Am Chem Soc*. **1**, 1766–1777 (2021). doi:10.1021/jacsau.1c00316
3. D. B. Craig, C. J. A. Wallace, ATP Binding to Cytochrome c Diminishes Electron Flow in the Mitochondrial Respiratory Pathway. *Protein Sci*. **2**, 966–976 (1993). doi:10.1002/pro.5560020610
4. N. J. Marianayagam, M. Sunde, J. M. Matthews, The Power of Two: Protein Dimerization in Biology. *Trends Biochem Sci*. **29**, 618–625 (2004). doi:10.1016/j.tibs.2004.09.006
5. H. B. B. Steele, M. M. Elmer-Dixon, J. T. Rogan, J. B. A. Ross, B. E. Bowler, The Human Cytochrome c Domain-Swapped Dimer Facilitates Tight Regulation of Intrinsic Apoptosis. *Biochemistry-US*. **59**, 2055–2068 (2020). doi:10.1021/acs.biochem.0c00326
6. T. Shur-Perek, Y. Avi-Dor, The Effect of Cytochrome c and Its 'Dimer' on Electron Transfer and Energy Transformation. *Biochemical Journal*. **126**, 709-716 (1972). doi:10.1042/bj1260709
7. R. Liu, S. Xia, H. Li, Native Top-Down Mass Spectrometry for Higher-Order Structural Characterization of Proteins and Complexes. *Mass Spectrom Rev.* (2022). doi:10.1002/mas.21793
8. M.T. Marty *et al.*, Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal Chem*. **87**. 4370-4376 (2015). doi:10.1021/acs.analchem.5b00140
9. H. Steen, M. Mann, The ABC's (and XYZ's) of Peptide Sequencing. *Nat Rev Mol Cell Bio*. **5**, 699–711 (2004). doi:10.1038/nrm1468
10. J. T. Adamson, K. Hakansson, Chapter 14 - Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Lectin Analysis. In Lectins, C. L. Nilsson, ed. (Amsterdam: Elsevier Science B.V.), pp. 343–371 (2007). doi:10.1016/B978-044453077-6/50015-6
11. C. A. Sherwood ***et al.***, Correlation Between Y-Type Ions Observed in Ion Trap and Triple Quadrupole Mass Spectrometers. *J Proteome Res*. **8**, 4243–4251 (2009). doi:10.1021/pr900298b
12. R. Bisson, G. Schiavo, C. Montecucco, ATP Induces Conformational Changes in Mitochondrial Cytochrome c Oxidase. Effect on the Cytochrome c Binding Site. *J Biol Chem*. **262**, 5992–5998 (1987). doi:10.1016/S0021-9258(18)45527-5

Transformer Phonotactics via IPA G2P Fine-Tuning

Mario Peng Lee[1]

[1]Stanford University Online.

ABSTRACT

Phonotactics is a core component in the field of linguistics, yet its role in natural language processing remains relatively unexplored. While traditional Grapheme to Phoneme (G2P) models are constructed by linguist-defined phonotactics, deep learning requires real-world natural language examples and no defined rulesets, allowing undefined rule learning. This is more representative modeling of human knowledge acquisition and offers a greater opportunity to provide unprecedented insights to the field of phonotactics. In this study, different sizes of the pretrained Transformer language model T5 were fine-tuned on G2P tasks and tested on qualitative and quantitative benchmarks. The experiments showcase G2P fine-tuning on transformers as an easy approach to creating phonotactic models. Moreover, the results demonstrate that all T5 models acquire phonotactic rules such as /z/ devoicing and pass the Wug test. Ultimately, findings suggest that language models acquire phonotactic knowledge in a similar manner, and they provide future research directions for human phonotactic modeling through deep learning, along with two Sequence-to-Sequence G2P datasets to encourage future G2P deep learning research.

INTRODUCTION

Phonotactics

The particular rules that govern a language's phonemic composition is the language's phonotactics. Linguists revealed that a native speaker of a language can easily classify unprecedented words into either 'possible' or 'not possible' in their language. For example, most native English speakers can classify the word 'brick' as 'English' and 'bnick' as 'non-English' (*1*).

One influential experiment on English phonotactics acquisition was the Wug Test (*2*), which explored the acquisition of morphological and phonological knowledge in children. In the experiment, an imaginary creature is presented and taught to be 'a wug'. Right after, two wugs would be presented, and the child is expected to say 'two wugs'. The correct answer would be pronouncing the 's' in 'wugs' as the voiced sound /z/, as in bugs instead of a voiceless /s/. The 'wug' is a made-up word to evaluate children's inference abilities while controlling for memorization. This is similar to zero-shot evaluation in machine learning.

Grapheme to phoneme

Grapheme to Phoneme (G2P) is a process that involves mapping written language (graphemes) to its pronunciation (phonemes) (Figure 1). A speaker's phonotactic knowledge is directly influenced by their lexical knowledge (*3*). Therefore, G2P is a direct way to study the phonetic understanding of a grapheme-based machine. G2P is challenging because a single grapheme can correspond to multiple phonemes (*4*), creating a complex set of rules that cannot be encapsulated by simple linear mapping algorithms. For example, the word 'mercedes' (Figure 2) has 3 different pronunciations for the same letter 'e'. Therefore, G2P is often used by linguists to study the fundamental understandings a speaker has over the language.

Traditional computational linguistics has studied language modeling using sets of defined constraints (*5*). These types of models had suboptimal performance on unprecedented tasks (*6*). On the other hand, deep learning AI has been proven to allow better generalization, more similar to human knowledge acquisition (*7*). Aided by the advancement of AI, linguists discovered new insights about human language modeling (*8*). This evolution is now being exhibited in the field of G2P (*9*).

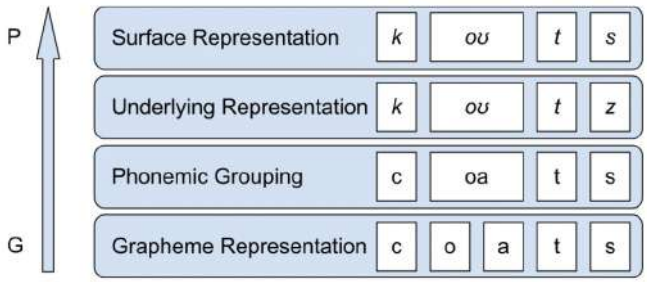


Figure 1: Pictorial representation of the Grapheme to Phoneme process bottom-up.
The word 'coats' is first grouped into phonemes; in this case, the diphthong 'oa' is grouped together. Then, these phonemic clusters are directly transposed to their respective International Phonetic Alphabet representations. Finally, to obtain the surface representation, phonotactic rules of the language are applied, in this case the /z/ to [s] devoicing.

The Transformer architecture

Natural Language Processing (NLP) went through a rapid development in the past decades (*10,11*). A Language Model (LM) can be conceptualized as a probability distribution over words or sentences.

This simple concept can be applied to text generation (*12*), machine translation (*13*), image recognition (*14*), text-to-image generation (*15*), and many more domains. This demonstrates that rich knowledge is being encoded in these probabilities, encouraging researchers to decode and interpret these knowledge representations (*7*). Phonotactics remains relatively outdated as compared to semantics and syntax in NLP despite being a core component in the field of linguistics.

The last major advancement in NLP has been the Transformer architecture, which introduced attention heads (*2*), a powerful technology that propelled LM performance in both syntactic and semantic knowledge (*16*). The attention mechanism effectively determines the importance of each part of the sentence, allowing the Transformer to perform context-sensitive processing to solve complex NLP tasks such as machine translation. For the purposes of this paper, three different model architectures were considered.

The Generative Pretrained Transformer

Generative Pretrained Transformer 3 (GPT) (*3*) is a Transformer model specialized on text generation and is known as one of the most versatile general-purpose LMs. However, it is suboptimal for G2P tasks since it is optimized for text generation rather than translation. Its architecture only takes into account context on the left, which will miss subtle linguistic information.

Bidirectional Encoder Representations from Transformers

In contrast, Bidirectional Encoder Representations from Transformers (BERT) (*17*) considers the context of words both before and after each word, lending improved representation learning for translation tasks.

Text-to-Text Transfer Transformer

Text-to-Text Transfer Transformer (T5), however, is perfectly suited for G2P. It was designed for translation tasks and devised to be fine-tuned to domain-specific applications. This involves defining the specific source language and target language for translation (*13*).

Due to its appropriateness for G2P, Rezackova *et al.* (*18*) fine-tuned T5 on English and Czech G2P tasks, yielding SOTA G2P results in uniword and homograph tasks on the 'sampa' alphabet. Later on, Zhu *et al.* (*19*) performed G2P tests on ByT5 for increased G2P performance. However, neither analyzed the phonological behavior of the LMs they developed. It could be expected that G2P tasks on Transformer LMs should yield significantly better results as well as new knowledge about phonetics as this evolution is already present in language modeling (*18,20*). Although an LM's language understanding seems unrelated to human psycholinguistics, it is undeniable that deep learning is becoming increasingly more similar to human inference learning (*21*). By examining real-world examples rather than relying on pre-defined language rules by linguists, deep learning may provide novel insights into phonotactics through example-based learning. This approach corresponds with the idea that a language is best studied by examining the language itself (*4,7*).

This paper

As G2P tasks can shed light on phonotactics and language development, and as deep learning models can serve as proxies for human reasoning (*21*), there exists an opportunity to further explore the relationship between G2P and NLP. This has yet to be fully realized (*22*) and represents a gap in NLP research that deserves attention.

In this paper, experiments were first conducted with transfer learning on pretrained Transformer models with simple G2P uniword tasks. This established the benchmark and motivated further fine-tuning. Then, data augmentation techniques were applied to the original data. Two English to International Phonetic Alphabet (IPA) labeled datasets were developed, ipa_dict_seq2seq and OpenWebIPA. These datasets can be accessed online at https://huggingface.co/datasets/mariopeng/openIPAseq2seq and https://huggingface.co/datasets/mariopeng/openwebIPA. Finally, the fine-tuned model was tested using quantitative metrics, such as accuracy, and qualitative metrics, such as the Wug Test (*2*) and loanword generalization.

The experiment shows that the model acquired advanced phonotactic knowledge despite few training epochs. Findings conclude that the larger the dataset, the more phonotactic patterns are acquired, similar to how a human's vocabulary correlates to their morphemic and phonotactic understanding of a language (*3*). Mainly, this paper establishes a foundation for future research on the understanding of human phonetic knowledge through deep learning technology.

METHODS

G2P is algorithmically similar to machine translation where a word or sequence of words is mapped to its respective pronunciation, in this case denoted by the IPA. The lack of English to IPA datasets motivated the introduction of new datasets in addition to existing data for this work.

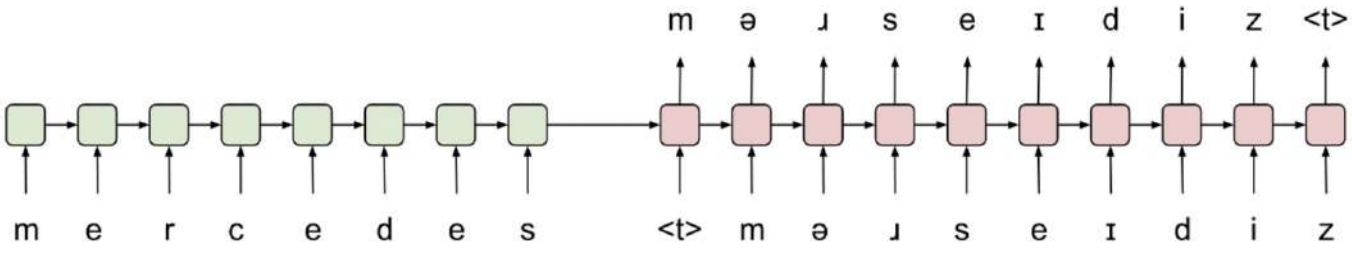


Figure 2: Overview of the Grapheme to Phoneme (G2P) process in an encoder-decoder transformer architecture such as Text-to-Text Transfer Transformer (T5). The single grapheme /e/ can be 'mapped' to 3 different phonemes depending on the surrounding context, supporting that G2P is a non-trivial process that could benefit from subtle rules learning such as those acquired by deep learning AI.

Single word dictionary dataset

The ipa_dict_en_us dataset was used for the main G2P task. In essence, it is a dictionary dataset of 12527 entries with words as keys and pronunciation transcribed in IPA as values (*23*). (Figure 3). The dataset was randomly shuffled with a constant seed and split using the 90/10 rule into a training set of 113328 entries and an evaluation set of 12599 entries with an average of 7.49 characters per word and 8.05 characters per label. Taking words as prompts and IPA transcriptions as labels, this dataset was ideal for the models to learn primitive G2P translation. Although ipa_dict offers a variety of languages, only English was used to better understand monolingual behavior before advancing into bilingual or multilingual in future work.

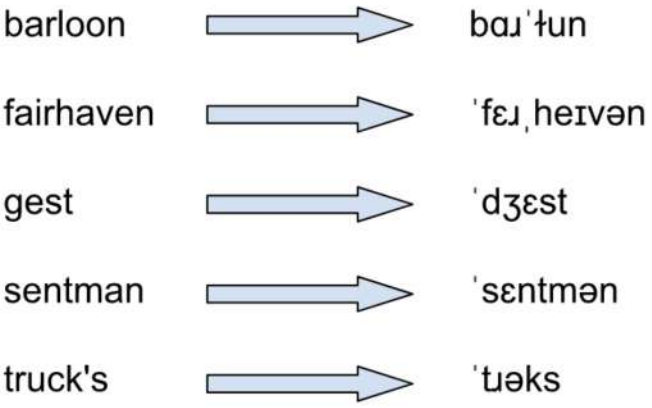


Figure 3: Random sample from the ipa_dict en_US , powered mostly by the Carnegie Mellon University Pronunciation Dictionary.
On the left, an array of training data is provided, on the right, their correct labels for G2P translation. As training data, the left side would be the features and the right side the labels.

Sequence to sequence dataset

The potential performance increase through prompt enrichment (*12,24*) motivated the creation of a sequence to sequence (Seq2Seq) dataset. For Seq2Seq fine-tuning, a dataset that contains sequences of plain text in English with their respective sequences of IPA as labels was needed. Two approaches were taken for preprocessing two new datasets (Figure 4). The first dataset was created by concatenating 15 to 30 random words from the ipa_dict dataset; this created nonsense sentences, but correctly allowed the model to perform multi-word English to IPA translation. To craft the second data set, the python module eng-to-ipa3, a module based on the CMU pronunciation dictionary (*23*), was used on a subset of the OpenWebText corpus, a collection of 8013769 passages from the web. This process resulted in 300,000 entries of internet text with IPA labels. Words not in the dictionary parser module were deleted by a linear algorithm.

Model

For all tasks, T5 encoder-decoder architecture was utilized to convert English text to IPA (*13*). To explore the simulated dynamics of language acquisition, three different sizes of T5 were evaluated: t5-small with 60M parameters and 6 layers, t5-base with 220M parameters and 12 layers, and t5-large with 770M parameters and 24 layers, all pretrained on the same corpus (Table 1). The PyTorch implementation of T5 from the Hugging Face Transformers library was used to simplify the training process. Since the base models of T5 do not contain IPA symbols in their vocabulary, all special IPA characters contained in the labels of the datasets were added to the tokenizer and the model dimensions were adjusted accordingly (Figure 5).

Evaluation

Two evaluation methods, a quantitative analysis on G2P performance and a qualitative linguistic analysis on unprecedented prompts, were implemented to provide a thorough understanding of the cross-model comparisons and the model's understanding of the language.

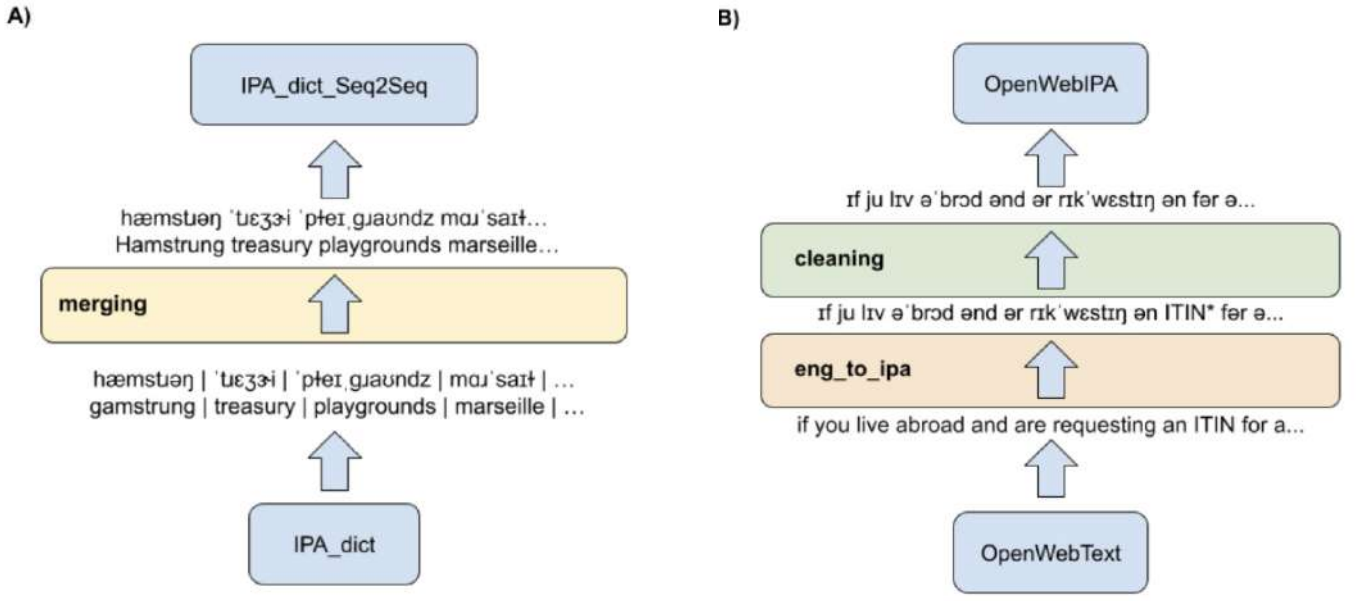


Figure 4: Data preprocessing pipelines for the Sequence-to-Sequence datasets along with an example entry from each dataset.
(A) demonstrates the concatenation and translation process in which an array of words are translated and then concatenated to form pseudo-sentences.
(B) depicts IPA translation followed by the 'out-of-dictionary' cleaning process which removes unprecedented words.

Model Training Specifications				
Parameter	t5-small	t5-small-s2s	t5-base	t5-large
Layer Norm Epsilon	1e-06	1e-06	1e-06	1e-06
Feed Forward Projection	ReLU	ReLU	ReLU	ReLU
Learning Rate	5e-5	5e-5	5e-5	5e-5
Optimizer	AdamW	AdamW	AdamW	AdamW
Drop Out Rate	0.1	0.1	0.1	0.1
Batch Size	256	256	128	64
Number of Epochs	100	5	60	10

Table 1: Details of the different Text-to-text Transfer Transformer (T5) model configurations for fine-tuning on Grapheme to Phoneme tasks.
Across models, most of the hyperparameters remained the same. Due to computational limitations, the Batch Size and number of training epochs were reduced as the model size increased.

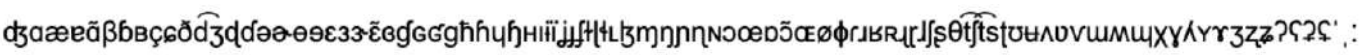


Figure 5: Special characters introduced in the datasets.
Since the original model did not include most of the IPA characters in its tokenizer, these were the symbols used to update the tokenizer to avoid unknown token errors.

Quantitative performance

The standard metric Character Error Rate (CER) was used to evaluate the accuracy of the model's predictions. Concretely, CER is the sum of the number of substitutions, deletions, and insertions divided by the sum of the number of substitutions, deletions, and correct characters from the label. The denominator is equivalent to the number of characters in the reference.

$$CER = \frac{S+D+I}{S+D+C} = \frac{S+D+I}{N} \quad (1)$$

CER was chosen over Word Error Rate, which evaluates the correct insertion and deletion of words to evaluate a sentence's correctness. This choice is due to the level of high precision necessary for the scope of this research. Generalizing a partially correct word into 'correct' or 'incorrect' will not accurately capture the nuances of allophones (Figure 2).

CER was evaluated on G2P tasks across 3 different sizes of T5 fine-tuned on the ipa_dict en_us dataset to give a broad perspective on how parameter size relates to phonotactic learning.

Qualitative performance

To test the model's acquisition of certain hypothesized phonotactics, a series of single word prompts were tested and the generated outputs were evaluated. All the prompt words for this part were ensured to not be in the dataset to avoid pattern matching or memorization.

The Wug Test (*2*) was conducted on the models by feeding the following five words: "wuks", "wugs", "musips", "absapps", and "appsabs". These made-up words explore phonotactic understanding of the possible allophones of the /z/ phoneme. Imaginary minimal pairs were used for control; for example, in "wuks" and "wugs", the only difference is the character 'k' and 'g', which only differ in their voicing feature. The prompts "absapps" and "appsabs" both encapsulate the same consonant clusters but differ in ordering, which changes the phonotactics. Therefore, "appsabs" should exhibit different surface representations from "absapps".

Additionally, to further test phonotactic understanding, a collection of 10 loan words from 6 different languages and 3 commonly mispronounced English words were compiled and tested on each model (Table 2). Each word was carefully selected to test a subset of the English phonotactics. Since loan words originate from other languages, they might produce interesting results which can either differ or align with native English speakers' productions. For example, 'Illinois' is an Irenwa word that does not follow English phonotactics yet most native speakers can produce correctly. Both accordance or dissonance of the results would reveal the extent to which the model understands phonotactics. These qualitative tests were evaluated with the 'native speaker judgements', a metric commonly used by linguists to evaluate the general feasibility of speech production.

RESULTS

Model size improves performance

When given a single word, t5-large, the highest-performing model, had a CER of 44.77% (Table 3). When given a sequence of words as opposed to a single word, performance dropped significantly for all models except t5-small-s2s, which was only trained on Seq2Seq tasks. A discrepancy was found in the relation between result pairs. In t5-small's and t5-base's results, despite a 6% difference in uniword performance, there is negligible difference in Seq2Seq performance. Furthermore, t5-large has approximately a

7% performance difference with t5-small on uniword tasks, yet their difference on Seq2Seq tasks is only 0.17%.

Phonotactics acquired

All models passed the Wug Test by correctly articulating the plural form of 'wug' and 'wuk' as expected by the original experiment. t5-small failed at 'musips' while t5-base and t5-large both passed. All three models produced 'absapps' and 'appsabs' correctly with a minor yet admissible difference in t5-small's response for 'appsabs' (Table 4).

The models differed greatly in loan word responses. For 'tacos', 'wednesday', and 'fresco', despite acceptable differences, all models demonstrated admissible native English speaker's pronunciations; for most of the words: 'brigade', 'coyote', 'cocoa', 'zucchini', 'Illinois', 'chow mein', and 'receipt', a linear progression from small to large models can be observed—the larger the model, the closer the resemblance to native speech. For 'tsunami' none of the models produced an admissible result; for 'gif', t5-small produced the most admissible output; and finally for 'mischievous' only t5-base produced the wrong phoneme (Table 2).

Single Shot Loan Word Prompts			
prompt	t5-small	t5-base	t5-large
tacos	tækoʊz	tɑkoʊz	tɑkoʊz
brigade	bɹeɪgd	bɹɪgəd	bɹaɪgeɪd
coyote	kaɪət	kɔɪoʊt	koʊjoʊt
fresco	fɹɛskoʊ	fɹɛskoʊ	fɹɛskoʊ
zucchini	kəntʃiz	zətʃɛni	sətʃuni
Illinois	ɪləkənz	ɪləsəns	ɪɹnaʊ
chow mein	khaʊmaɪn	tʃaʊmaɪn	tʃaʊmaɪn
tsunami	taɪmnɑs	taɪndɹɑmən	təmsuoʊ
cocoa	kəkoʊdə	kəkɑmə	koʊkoʊə
gif	gɪf	dʒaɪf	dʒaɪf
mischievous	mɪstʃivəs	mɪʃivəs	mɪstʃivəs
wednesday	wɛnzdeɪ	wɛnzdeɪ	wɛndzdeɪ
receipt	ɹisɪpʃən	ɹisɛpʃən	ɹisit

Table 2: 10 Loan Word Grapheme to Phoneme prompts, borrowed from French, Spanish, Italian, Irenwa, Chinese, and Japanese, and several commonly mispronounced English words were included.
Correctness is color coded. Yellow is "slightly incorrect,"Red is "severely incorrect," and uncolored is "admissible."

DISCUSSION

Phonotactics is a core component of linguistics, yet it remains relatively outdated in terms of its language modeling analysis. Computational tools for phonetics studies have followed NLP's lead in focusing on model semantics and syntax analysis; this study takes the next step by investigating phonetic studies in the large pretrained Transformers paradigm. This exploratory investigation builds on recent works and highlights a research gap that might be promising for human phonetic understanding. Further investigation is necessary, and this paper demonstrates that deep learning G2P is a promising field with significant potential for expansion.

Analysis

In this study, the model was significantly outperformed by traditional computational models. Considering SOTA is 10.9% CER achieved by joint n-gram models, this model's 44.77% CER indicates performance is poor (*26*). However, this paper's results are not representative of the maximum potential performance of Transformers on G2P tasks. With only 10 fine-tuning epochs, the larger model outperformed the smaller model which trained for 100 epochs, indicating that training becomes more efficient as model size increases, in accordance with scaling laws. Further training is likely to yield logarithmically better results (*20*). This observation highlights the efficiency of using fine-tuning as a short-cut for high-performing G2P models as opposed to training them from scratch. It is worth noting, however, that direct comparisons of this paper's results cannot be made with other deep learning models, as prior deep learning G2P SOTA has focused on Word Error Rate (WER) which is inappropriate for uniword tasks (*6,18,27*) .

G2P performance dropped considerably for sequences of words, likely due to T5's heavy positional embeddings. If future work investigates the attention heads and layers, it could be hypothesized that mostly those layers or blocks governing the first word in a sequence were effectively trained during the uniword G2P fine-tuning while the other layers or blocks were not. The Uniword-Seq2Seq pair discrepancy seems to align with the interpretation that the Seq2Seq models tend to predict the first word correctly and proceed to fail on subsequent words. This seems to be in accordance with the fact that each entry in the dataset has on average 22.5 words per sequence. Assuming the first word of each of those sequences is exactly right, the average CER should be around 95%. However, since no model has 0% error rate, it is expected that 95% is the maximum possible performance if only the first word is properly translated.

The Wug Test provided useful information on phonotactic acquisition in language models. The results suggest that the plural morpheme 's' is correctly voiced whenever preceded by a voiced phoneme. However, the fact that this behavior was correct on 'appsabs' yet not on 'absapps' across all models suggests that positional information of a character is affecting the model's phonotactic understanding (Table 4). Supported by the 'mercedes' example (Figure 2), it can be inferred that deep learning is constructing a phonotactical explanation with allophonic representation rather than linearly mapping graphemes to phonemes.

CER performance (%)		
Model	Uniword	Seq2Seq
t5-small	51.89	92.25
t5-small-s2s	116.13	91.83
t5-base	46.08	92.27
t5-large	44.77	92.08

Table 3: Character Error Rate (CER) in percentages.
As observed in Equation 1, CER can exceed 1.0 or 100% when the number of extra incorrect insertions are larger than the number of correct characters. The uniword dataset evaluates the model on single word prediction while the Seq2Seq dataset evaluates the model on strings of 15 to 30 words. The best performance was achieved by t5-large with 44.77% CER on the Uniword dataset. Performance increased with model size for Uniword, yet remained relatively uniform for Seq2Seq.

The loan word experiments showcase that larger models seem to acquire phonotactic rules at a faster rate than smaller models. Additionally, they are able to better generalize this knowledge to better predict words such as 'chow mein' and 'coyote' while smaller models struggle to do so. An interesting prompt is 'cocoa'; since English phonotactics disallow the /oa/ diphthong, the models applied a voiced consonant insertion in their answers between /o/ and /a/, generating [kəkoʊdə]. This suggests a partial understanding of restricted diphthongs in the language, proving that specific phonotactic rules can be probed by prompt engineering. However, t5-large's correct production could potentially imply that it learns to generalize better, as a true human speaker would, by breaking some smaller rules to apply broader rules. Over-generalization is exhibited when a speaker knows enough rules to apply them but not enough to capture exceptions. This is usually seen in children (*1*), especially when they encounter irregular verbs. Over-generalization from the model could be in support of the claim that the model is acquiring rules in a way similar to human speakers, making them viable subjects for linguistics studies.

Another evidence of acquired phonotactics is seen upon prompting the model with 'tsunami'. None of the models produced the /ts/ consonant cluster which is not allowed by English phonotactics. This is expected since there are no words starting with 'Ts' in the training data. One could expect that a native English speaker unexposed to this word would also have trouble pronouncing it correctly. Furthermore, models seem to be able to correctly replicate Spanish and Italian words such as 'tacos' and 'fresco' with high similarity as native speakers would. Lastly, for the prompt 'receipt', only t5-large was close to the correct answer, showing great phonotactic knowledge. Overall, all the examples (Table 2) exhibit similarities to human phonetic reasoning, and further analysis with a better trained model should be conducted for more accurate results (*1,4*).

Limitations

Due to the size of large pretrained models, training was resource intensive. As a result, there may not have been enough training on the models. Additionally, the lack of monolingual open-source IPA labeled data required considerable effort to create new datasets for this specific G2P task. Although Seq2Seq IPA labeled datasets were generated, they were not utilized. Furthermore, the datasets presented do not include phonotactic interactions across words, a phenomenon which exists in natural language production. For example, "green bean" is pronounced as "greembean" even though "green" by itself would not be pronounced with an "m" at the end; this is called place assimilation. In terms of validity, only uniword behavior was studied thoroughly, and behavior may change drastically on sequences of words or with Seq2Seq English to IPA machines. By using uniword datasets, heteronyms, words with same spelling but different pronunciation, were completely neglected in these experiments. A richer dataset could resolve these issues. The qualitative analysis of the models is weak due to the absence of a more formal metric. A stronger metric could be attained by conducting a survey on native speakers or trying to quantitatively evaluate the productions with CER. Finally, multilingual data augmentation could have been implemented theoretically without damaging the quality of monolingual phonetic understanding.

Future work

This research raises numerous questions that should be answered in future work. Three ideas that were out of scope could build on top of this work. First, a phonetic model could be conjoined with a simple IPA to Acoustic dictionary, resulting in a cheaper text to speech with performance dependent on English to IPA CER. Second, BERT could be tested on G2P and its performance could be compared to T5. Third, a model could be trained on the Seq2Seq datasets to theoretically yield better G2P performance.

In addition, bilingual or multilingual studies of this same project could yield results that align with the hypotheses of this paper and present new behaviors that are not present in monolingual models. This could be done by implementing multilingual data augmentation, which theoretically would instrumentally yield better performance on G2P tasks.

Wug Test Input-Output Pairs

Model	wuks	wugs	musips	absapps	appsabs
t5-small	[wəks]	[wəgz]	[məsəps]	[æbsəps]	[æpsəbz]
t5-base	[wəks]	[wəgz]	[mjuzɪps]	[æbsəps]	[æpsæbz]
t5-large	[wəks]	[wəgz]	[mjuzɪps]	[æbsəps]	[æpsæbz]

Table 4: The Wug Test results.
All models correctly employed the s-to-z Grapheme to Phoneme transformation when "s" is followed by a voiced consonant in "wugs." This can be contrasted to the control prompt "wuks", in which "s" is transcribed as /s/. t5-small failed to do s-to-z in the "musips" test case. Curiously, models did not generalize s-to-z for the substring "bs" in "absapps" while the opposite was true for "appsabs".

Another possible research direction to explore is the creation of a large pretrained language model purely on phonological domains, either G2P or phoneme to phoneme tasks. This model's knowledge representation should be compared to traditional Text to Text models, as there is potentially promising interpretability research on these two different paradigms using tools such as those presented by Tenney *et al.* (*29*). A further motivating factor is potential improved performance, as phonemic representations seem to capture more information than graphemic representations (*24*). Finally, few-shot prompting as introduced by Brown *et al.* (*12*) and a generative phonemic model could reveal novel behavior about generalized knowledge. There is likely to be growing interest in interdisciplinary interpretability research along with cognitive science and psycholinguistics to understand both human and deep learning reasoning and data representation.

In terms of IPA datasets, a SOTA G2P model could create a large English to IPA dataset for further phonetic analysis. Related to this, self-supervised learning (*30*) should be explored to counteract the lack of IPA labeled data for G2P training. Masked training as seen in BERT is another alternative solution (*17*). Machine generated data would resolve the issue of deficient datasets to train inter-word phonetic interactions or heteronym labeling, since unexpected interactions may occur as in the case of /e/ being mapped to multiple phonemes (Figure 2).

In conclusion, this exploratory investigation highlights the potential of large pretrained Transformers in the field of G2P, demonstrating their ability to acquire and generalize phonotactic knowledge similar to human speakers, with promising avenues for future research.

AUTHORSHIP STATEMENTS

Author worked individually, with no external collaborators for this project.

ACKNOWLEDGEMENTS

I want to thank the Fall 2022 XCS224U staff, particularly my Course Facilitator, Raul, and Professor Potts for this learning opportunity.

REFERENCES

1. N. Chomsky, M. Halle, *The Sound Pattern of English*. (Harper & Row Publishers, 1968).
2. J. B. Gleason, The Wug Test, in *Encyclopedia of Language Development, SAGE Publications*, 2014. pp. 687–688.
3. S. A. Frisch, M. R. Brea-Spahn. *Lab Phonol.* **1**, 345–360 (2010).
4. N. Chomsky, *Language and Mind* (Cambridge University Press, ed. 3, 2006), pp. xviii, 190.
5. M. Bisani, H. Ney, Joint-sequence models for grapheme-to-phoneme conversion. Speech *Commun*. **50**, 434–451, (2008).
6. A. Deri, K. Knight, in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, K. Erk, Ed. (Association for Computational Linguistics, 2016). pp. 399–408.
7. H. Sajjad *et al.*, in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* M. Carpuat, Ed. (Association for Computational Linguistics, 2022). pp. 3082–3101.
8. C. Mayer, M. Nelson, in *Proceedings of the Society for Computation in Linguistics 2020*, A. Ettinger, Ed. (Association for Computational Linguistics, 2020). pp. 291–301.
9. A. Sherstinsky, Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. Physica D. 404, (2020). *doi*:10.1016/j.physd.2019.132306.
10. Z. S. Harris, Distributional Structure. *WORD*. **10**, 146-162 (1954). *doi*:10.1080/00437956.1954.11659520Z.
11. A. Vaswani *et al.*, Attention Is All You Need. *arXiv.* (2017). *doi*:10.48550/ARXIV.1706.03762.
12. T. B. Brown *et al.*, Language Models are Few-Shot Learners. *arXiv.* (2005). *doi*:10.48550/ARXIV.2005.14165.
13. C. Raffel *et al.*, Exploring the Limits of Transfer Learning with a Unified Text-to Text Transformer. *arXiv. (2020)*. *doi*:1910.10683.5.
14. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition. *arXiv*. (2015). *doi*:1512.03385.
15. A. Ramesh *et al.*, Zero-Shot Text-to-Image Generation. *arXiv*. (2021). *doi*:2102.12092.
16. A. Chowdhery *et al.*, PaLM: Scaling Language Modeling with Pathways. *arXiv*, (2022). *doi*:10.48550/ARXIV.2204.02311.
17. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv*, (2019).
18. M. ˇRez´aˇckov´a, J. ˇSvec, and D. Tihelka, in Proc Interspeech 2021.
19. J. Zhu, C. Zhang, D. Jurgens, *ByT5 model for massively multilingual grapheme-to-phoneme conversion*, *arXiv*, 2022.
20. J. Kaplan *et al.*, *Scaling Laws for Neural Language Models*, *arXiv*, 2020.
21. L. P. Argyle *et al.*, Out of One, Many: Using Language Models to Simulate Human Samples, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, *arXiv*, 2022. pp. 819–862.
22. K. Rao, F. Peng, H. Sak, F. Beaufays, G2P Using LSTM RNN, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, ISSN, 2015. pp. 4225–4229.
23. CMU, The Carnegie Mellon University Pronouncing Dictionary, *2015*.
24. H. Zhang, H. Zhan, Y. Zhang, X. Yu, Y. Lin, Revisiting IPA-based Cross-lingual Text to-speech, *arXiv*, 2021.
25. A. Gokaslan, V. Cohen, OpenWebText Corpus, *2019*.
26. L. Galescu, J. F. Allen, in *Proc. 7th International Conference on Spoken Language Processing, ICSLP, 2002*. pp. 109– 112.
27. M. N. Sundararaman, A. Kumar, J. Vepa, Phoneme-BERT: Joint Language Modelling of Phoneme Sequence and ASR Transcript, *arXiv*, 2021.
28. C. Zhu, K. An, H. Zheng, Z. Ou, Multilingual and crosslingual speech recognition using phonological-vector based phone embeddings, *arXiv*, 2022.
29. I. Tenney *et al.*, The Language Interpretability Tool: Extensible, Interactive Visualizations and Analysis for NLP Models, *arXiv*, 2020.
30. G. E. Hinton, S. Osindero, Y.-W. Teh, A Fast Learning Algorithm for Deep Belief Nets, *Neural Computation 18*, 2006. pp 1527–1554.

Effects of Nectar Toxicity and Flower Color on Foraging of *Calypte anna*

Rosie Fitzsimmons[1], Madeline Wiygul[2], Yuerong Xiao[2]

[1]University of California, Davis. [2]University of California, Los Angeles.

ABSTRACT

The coevolution between plants and pollinators has driven great diversity and reproductive success in flowering plants. In particular, plants have evolved with hummingbirds to create a distinct subset of long, tubular flowers, generating great diversity amongst flowering plants. Within this hummingbird-flower dynamic, the foraging behavior of hummingbirds is affected by many variables, such as color cues and toxicity level in the nectar. This study used artificial flowers to mimic *Nicotiana glauca* (*N. glauca*) and *Epilobium canum* (*E. canum*) to examine the effect of flower color and nectar type on the behavior of *Calypte anna* (*C. anna*) by looking at *C. anna*'s flower preference and duration at each flower. Flower color preference, toxic nectar tolerance, and spatial preference were the variables used to analyze *C. anna* behavior. The experiment aimed to control for spatial preference by randomizing locations of flower color and nectar type. The hummingbirds did not show a preference for red versus yellow flower color and nectar toxicity had no effect on the length of time hummingbirds spent at the flower, but potential spatial preference independent of color and nectar type was suggested. The relationship between *N. glauca*, *E. canum*, and *C. anna* sheds light on the complexities of plant-pollinator interactions and motivates future study on plant-pollinator coevolution amidst climate change. Hummingbird-driven evolution has been a generator of biodiversity for millions of years, so understanding the relationship between plants and hummingbirds is essential for maintaining biodiversity in light of anticipated mass global extinction.

INTRODUCTION

Coevolution between plants and pollinators has driven distinct morphological changes on both ends, particularly in specialized pollinating systems. Pollinators are major drivers of natural selection in flowering plants *(1)*. Flowering plants reproduce by fertilizing a plant's pistil (female flower part) with their pollen (male flower part) either sexually or asexually via self-pollination. Both processes typically require a pollinator to transport the pollen to the pistil. One method flowering plants employ to attract pollinators is to produce sugary nectar inside the flower *(2)*. Pollination syndrome is the process of flowering plants changing their morphology, phenology, and physiology to better match their pollinator counterpart and exclude outside competitors *(2)*. For example, flowering plants can alter flower shape, color, nectar type, nectar amount and more, which drives changes in pollinators' size, behavior, and morphology *(3)*. This coevolutionary relationship has led to specialization between plants and pollinators and heightened reproductive success. As a result, plant-pollinator specialization has become widespread across flowering plants and is responsible for the unique diversity seen in flowering plants today *(4)*.

One prevalent example of plant-pollinator specialization is hummingbirds and hummingbird-specialist plants *(2)*. There are many key characteristics that plants exhibit under the influence of hummingbird-driven pollination syndrome, such as tubular flower shape and lack of a landing platform (Figure 1). Hummingbird-specialists have developed a corolla, a tube-like structure composed of long, thin petals that protects reproductive organs inside the flower *(5)*. This corolla structure coupled with a lack of a landing platform make it difficult for most pollinators to access the flower's nectar. Hummingbirds, however, have long, thin beaks and tongues and the unique ability to hover while flying that allow them to pollinate such flowers and obtain a largely untouched nectar reward *(6,7,8)*. Uncovering further specialization between hummingbirds and flowers beyond the flower-beak relationship could shed light on the importance of traits like flower nectar type or color in plant-pollinator coevolution.

Unlike most other pollinators, hummingbirds perceive red coloration, causing hummingbird-specialist plants to contain red pigment in the flower petals or corollas *(9)*. Because red flowers are not frequented by other nectar-consumers, hummingbirds have learned to connect red flowers with higher nectar volumes and demonstrate a hierarchical preference for red coloration *(10)*. However, other research has found that when nectar rewards are equal, hummingbirds drink from flowers regardless of color and depend on spatial memory to locate flowers that were not emptied previously *(6)*. Overall, the impact of color cues on hummingbirds' preference when spatial cues are removed is not well understood.

Although plentiful nectar increases the frequency of hummingbird visits and reproductive success, producing nectar is energy-expensive and limits other flower functions *(11, 12)*. To increase the efficiency of nectar production and maximize reproductive success hummingbird-specialist plants have developed many strategies. The strategy this study focuses on is the production of toxic nectar,

which is theorized to cause hummingbirds to spend less time at one plant but frequent more plants with longer distances in between *(13)*. Another theory is that toxic nectar is used by hummingbird-specialists to reduce the volume of nectar consumed per hummingbird visit and to increase resistance to herbivory *(14, 15)*.

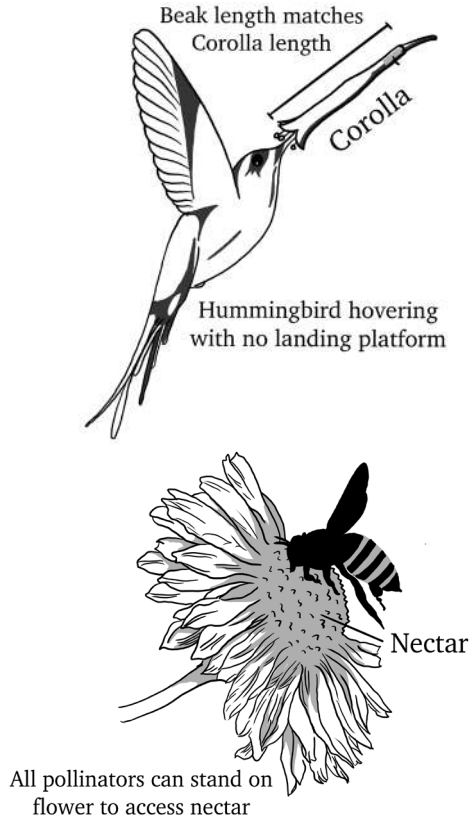


Figure 1: Illustration of a hummingbird and a hummingbird specialized flower, with emphasis on the tubular structure of the flower and the long, thin structure of the hummingbird's beak compared to other pollinators.
The illustration depicts the tubular structure of the flower, with the nectar sitting at the end of the long corolla. The pollinators and the flower next to one another show how nectar is difficult to reach by anything but the hummingbird's long, thin beak. The complementary structures demonstrate the co-evolution of flowers and hummingbirds. Other pollinators are unlikely to reach the nectar at the bottom of the specialized flower, as depicted in the illustration, which leaves an enticingly large nectar reward for hummingbirds. Illustration by Chelsea Lai.

Nicotiana glauca (*N. glauca*) is a hummingbird-specialist originating from South America that has invaded California and now relies on California-native hummingbirds for reproduction *(16)*. *N. glauca* has yellow, tubular flowers, toxic nectar, and no landing platform — distinguishing characteristics for hummingbird-specialist plants *(17)*. *N. glauca* produces a pyridine alkaloid, anabasine toxin. Sharing the same range as *N. glauca* across California, *Epilobium canum* (*E. canum*) is a California native hummingbird-specialist plant that produces red, tubular flowers. In contrast to *N. glauca*, *E. canum* produces a sucrose-based nectar with amino acids that provides essential nutrients to hummingbirds *(19)*. While *N. glauca* and *E. canum* differ in color and nectar toxicity, they both rely on hummingbird pollination and have found reproductive success in their shared environment. These dynamics shed light into the complex relationship between hummingbirds and flowering plants, and prompt exploration of why both species appear equally successful in their environment. While previous studies have examined hummingbird flower color preference and toxic nectar tolerance, potential confounding variables, such as spatial cues, have never been considered. This study explores the effects of flower color and nectar type specifically in the absence of spatial cues.

Calypte anna (*C. anna*) is a native hummingbird species that lives in Southern California chaparral and relies heavily on the nectar of *E. canum (12)*. *N. glauca* was introduced to California a little over 100 years ago and has spread to disturbed soils and riparian areas throughout the Southern California chaparral in which *C. anna* resides *(16)*. Thus, *C. anna* has been living among nonnative *N. glauca* plants for a few decades, but it is unknown when the hummingbird began feeding on *N. glauca* nectar. Conversely, *C. anna* has likely been feeding from the native *E. canum* for hundreds of years in Southern California *(12)*.

The null hypothesis states that hummingbirds do not favor flower color nor nectar type *(6)*. The study investigates the effects of flower color and nectar type on C. anna preferences when choosing a flower and the length of flower visits. It was hypothesized that *(1)* because red flowers are associated with high nectar rewards, hummingbirds would have an initial preference for red flowers, *(2)* hummingbird visits would decrease in duration as toxicity increases, and *(3)* hummingbirds would not show spatial preference amid the randomized flower color and nectar type.

METHODS

Site characterization

The study was conducted at Sedgwick Reserve from November 1, 2021 through November 3, 2021. Sedgwick Reserve is located in Santa Ynez Valley in Santa Barbara County, California and consists mostly of oak savanna containing a diverse array of flowering plants, shrubs, and trees. Two types of hummingbird-specialist plants that grow in Sedgwick Reserve were chosen as flower models of the study: the red, non-toxic, native *E. canum* and the yellow, toxic, invasive *N. glauca*. *C. anna* are still very active in November because of the abundant food sources *(20)*. The garden plot was selected because there are naturally flowering E. canum where regular *C. anna* visits were observed. This decision made it more likely that the hummingbirds would locate the experimental plot within the three day observation period. The plot, however, was set up in a location that does not directly border any *E. canum* plants (Figure 2).

Artificial flower construction

Although mirroring natural conditions is preferable in most experimental designs, using real flowers would have made controlling nectar concentration, nectar type, and spatial randomization impossible. For example, putting the nontoxic nectar in real *N. glauca* flowers would have likely led to a seepage of toxic compounds into the nectar from the *N. glauca* flower petal. Using artificially constructed flowers allowed control over potential confounding variables. Artificial flower construction is common practice, especially for hummingbird and flower experiments, because of hummingbird's willingness to drink from artificial nectar sources *(21)*. Nearly identical artificial flower construction methods were used in Fenster *et al.*'s 2006 paper that tested the association of nectar

and floral traits in hummingbird attraction *(21)*. This experiment used 15 yellow and 15 red artificial flowers with a 30 cm wooden dowel for the stem, a piece of a 5-mL plastic pipette to hold the nectar, and colored duct tape for the petals (Figure 2). Red and yellow were chosen to mimic the coloration in *E. canum* and *N. glauca* respectively; pipette tubes were used to mimic the tubular shape of the hummingbird-specialist flowers. Each petal was made of red or yellow duct tape that was cut into a 2 cm wide circle with a 1.3 cm diameter hole in the middle where the tip of the pipette rested. The bottom of the pipette was glued to the dowel tip at a 30 degree angle and then wrapped in the colored duct tape for durability and waterproofing.

Artificial nectar creation

To make artificial nectars with traits of each flower species, the flowering part of *E. canum* and *N. glauca* plants was wrapped in fiberglass mesh at randomly chosen locations around Sedgewick Reserve at least 0.5 km apart. The flowers with their untouched nectar were collected the next morning. There were three different nectar types created: sugar water, sugar water mixed with *E. canum*, and sugar water mixed with *N. glauca*. The sugar water was made with a 1:5 volume ratio of Great Value granulated white table sugar to distilled water, which has been demonstrated as an ideal ratio for hummingbird feeders by Waser et al. (2018) *(22)*. To create the *E. canum* sugar water solution, 410 mL of sugar water was brought to a rolling boil, removed from the heat source, mixed with six *E. canum* flowers, and stirred for 10 minutes before removing the flowers. The concentration of *N. glauca* nectar in the *N. glauca* sugar water and nectar solution gradually increased each day for the three days of the experiment. The lowest concentration used on day one was made with three *N. glauca* flowers and used the same process as the *E. canum* treatment. On the second day, the process was repeated using four *N. glauca* flowers and again on day three with ten *N. glauca* flowers. The nectar was stored in three sealable jars and transferred from the jars into the flowers using 5-mL pipettes. The nectar was made each morning, and replenished throughout the day to keep the flower nectar levels full at all times.

Field observations

The flowers received a random nectar type, with five yellow flowers and five red flowers receiving each nectar type. The 5-mL pipette tube was filled to half as a standardized volume. A 0.37 m × 0.81 m flower plot was set up with three columns and ten rows (Figure 2). Each flower was inserted 5 cm into the ground, standing about 30 cm above the ground. The location of each flower was randomized within the plot every morning before the start of the observation period. The observation started at 07:00 and ended around 18:00 when the sun set each day. At 13:00 on day two and day three, the *N. glauca* nectar was replaced with the *N. glauca* nectar of the next higher concentration. Besides the replacement of *N. glauca* nectar, all of the flowers were refilled with their corresponding nectars at 07:00 and 13:00.

Observations were collected every two hours for 11-12 hours across three days. These recordings were done with a handheld stopwatch and recorded on a paper with a premade table. During each hummingbird visit, the hummingbird's gender, order of flowers the hummingbird drank from, and the length of time spent at each flower were recorded. If a hummingbird momentarily lifted its head then continued to drink from the same flower, it was to be considered one continuous drinking period. If a hummingbird left the plot for more than five seconds but came back, it was counted as a new visit to the plot. A premade, hand drawn garden plot was made in addition so the exact flower visited could quickly be marked while still recording the time.

Figure 2: Picture of the artificial flower set up in the garden.
30 artificial tubular flowers in two colors and three nectar types set up in a 3×10 plot. 15 yellow and 15 red artificial flowers were constructed to mimic the coloration of *E. canum* and *N. glauca*, respectively. Five of each colored flowers were filled with one of the three nectar types: sugar water mixed with *E. canum*, sugar water mixed with *N. glauca*, and sugar water as control. The location of the flowers in the 3×10 plot was randomized every day.

Statistical testing

Statistical analyses were conducted using JMP Statistical Software version 16.1.0 and RStudio version 1.3.1073. A two-tailed t-test was used to look at how flower color affected the number of hummingbird visits. To analyze the effects of different nectar types and the increasing concentration of *N. glauca*, two one-way ANOVA tests were conducted with the length of hummingbird visits as the dependent variable. It was not possible to compare color preference with the different concentrations of *N. glauca* nectar due to insufficient sample size, because each toxicity level was only tested for one day. In RStudio, the package "ggplot" was used to generate a heat map of hummingbird visits on flowers during the three-day observation period.

RESULTS

In total, hummingbirds drank from artificial flowers 223 times across 106 visits. All of the hummingbirds observed were *C. anna* females, and no more than two hummingbirds were observed at the same time.

Among all flowers that hummingbirds drank from over the three days, there was no significant relationship between flower color and hummingbirds' initial preference ($N = 223$, $t = -1.41$, $P = 0.16$; Figure 3). Hummingbirds also did not show a significant preference for red or yellow coloration when choosing the first flower they drank from during one visit alone ($N = 69$, chi-square$=1.23$, $P=0.266$).

The length of time hummingbirds spent at each flower was not significantly affected by nectar type ($N = 223$, $F = 1.46$, $P = 0.24$; Figure 4). Looking at just the artificial flowers with *N. glauca* nectar, it was found that increasing the concentration of *N. glauca* in the nectar did not have a measurable impact on the length of visit at each flower ($N = 146$, $F = 2.93$, $P = 0.06$; Figure 5).

The heat map of hummingbird visits showed that hummingbirds visited the last five flowers on the third row most often despite the flower color and nectar type being randomized in different locations each day ($P < 0.001$, Figure 6).

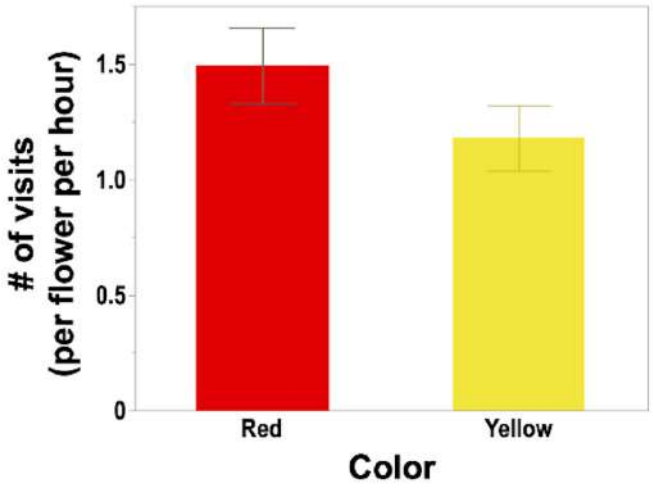


Figure 3: Graph of color of flower and number of hummingbird visits per flower per hour.
T-test of the effects of flower color on the number of visits by *C. anna*. A t-test was conducted on JMP version 16 to show that over the three days, flower color had no effect on the number of visits by *C. anna* per flower per hour (N = 223, t = -1.41, P = 0.16). The x-axis showed two different colors, and the y-axis is the mean number of vi`sits by *C. anna* per flower per hour over the three days. The *E. canum* bar showed the number of visits to red flowers, and the *N. glauca* bar showed the number of visits to yellow flowers.

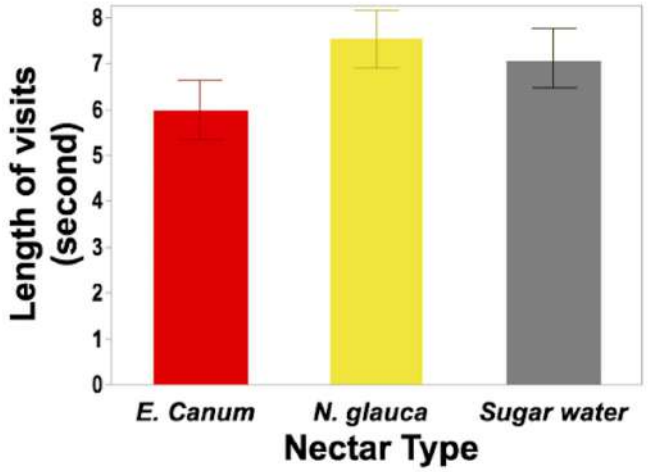


Figure 4: Graph of nectar type and length of hummingbird visits in seconds.
One-way ANOVA was conducted on JMP version 16 to test the effects of nectar type on the length of visits by *C. anna*. The tests showed that over the three days, *C. anna* did not have a preference to stay longer at any specific nectar type (N = 223, F = 1.46, P = 0.24). The *E. canum*, *N. glauca* and Sugar Water bars refer to flowers with *E. canum* nectar solution, *N. glauca* nectar solution, and sugar water solution respectively. The x-axis showed three different nectar types, and the y-axis is the mean length of visits by *C. anna* over the three days.

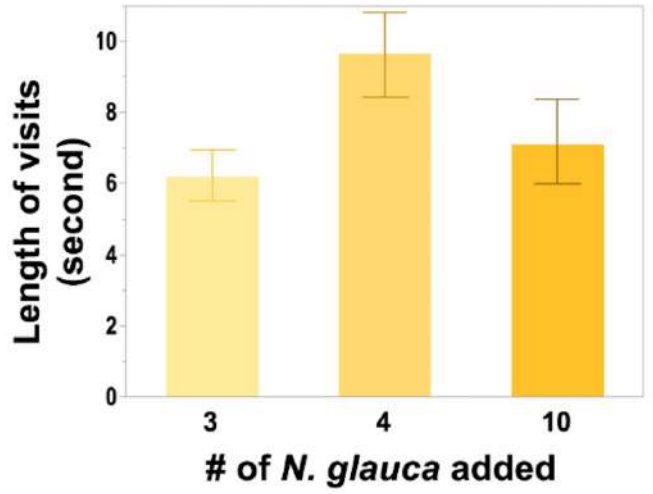


Figure 5: Graph of percent toxic nectar added to sugar solution and number of hummingbird visits in seconds.
One-way ANOVA was conducted on JMP version 16 to test the effects of increasing *N. glauca* concentration on the length of visits by *C. anna*. The test showed that as the *N. glauca* concentration in the nectar solution increased, the average length of visits by *C. anna* was not impacted (N = 146, F = 2.93, P = 0.06). The x-axis shows the increasing number of *N. glauca* flowers added to make the nectar solution, and the y-axis is the average length of visits by *C. anna*.

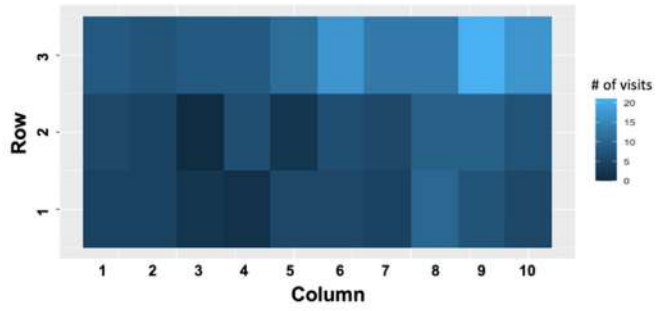


Figure 6: A heat map that replicates the flower plot, with darkest colors representing where most time was spent by hummingbirds.
Package "ggplot" was used in RStudio version 1.3.1073. The heat map suggests hummingbirds have potential spatial preference. The possible spatial preference in this plot would indicate *C. anna* preference visit on the third row from column 5 to column 10.

DISCUSSION

When looking at hummingbirds' preference on which flower to visit, no significant difference was found to support the hypothesis that *C. anna* has an initial preference for red flowers. This lack of color preference was similar to the finding of a previous study, which suggested that although red flowers were sometimes preferred, hummingbirds learn to visit different colored flowers in nearly equal frequencies if nectar rewards are constant regardless of coloration *(7)*. If it holds true that nectar reward is a more important factor than flower color, one possible explanation of the result could be the effect of the same nectar volume used in every flower. In the study's design, the pipette tubes were always filled to 2.5 mL as a standard volume. This volume of nectar was more than what a hummingbird would receive from an actual flower in an equal amount of time, thus the normal depletion of nectar by other nectar-consumers was not simulated *(7)*. As a result, *C. anna* might have visited any flower knowing that there would always be sufficient nectar rewards. It is yet to be determined whether hummingbirds prefer red flowers when nectar rewards in flowers are kept at a much lower amount, which is closer to a natural situation.

Although flower colors had no impact on the preference of *C. anna* in the study, it is possible that the flowers were chosen based on their locations. The observed spatial preference aligned with other studies that suggested hummingbirds memorize spatial cues such as local landmarks rather than visual cues such as flower color *(6)*, and further confirmed that the spatial cues were still used to inform hummingbirds' foraging preference even when location of flowers were randomized every day (Figure 6). Specifically, the possible location of visits in the study plot was close to an average-height flowering tree and a low fence. It was observationally noted that the hummingbirds rested on the tree and fence in between some visits; insects attracted by the flowering tree were also consumed by the hummingbirds. The most visited flowers were on the corner of the plot, closest to the fence and the tree. Other factors, such as convenience and insects as food sources, might also impact the foraging preference of hummingbirds. A comprehensive study that crosses flower characteristics, such as color or nectar type, with environmental factors, such as time of year or food availability, should be conducted to explore why hummingbirds visit their food sources in different frequencies.

When looking at the length of time *C. anna* spent at each flower, it was found that neither the nectar type nor the toxic nectar concentration affected the number of visits or length of visits. This pattern contradicts the hypothesis that *C. anna* would have longer visits at flowers with sugar water and with *E. canum* nectar and that increasing toxin concentration would decrease overall visits by *C. anna*. A similar study in Palestine was performed, where *N. glauca* was pollinated by native sunbirds. Sunbirds' size, behavior, diets, and long, thin beaks are similar to hummingbirds and the two species are often compared *(19)*. This sunbird study found that the pollinators chose the nectar with higher toxicity levels less often and showed reduced gut transit time after ingestion of the toxic nectar *(19)*. Because of this, it is unexpected to find that increasing toxicity levels did not influence the preference of *C. anna* like it did to sunbirds.

One explanation for this lack of response to toxic nectar could be the variability in the nectar toxicity of the flowers collected. Nectar toxicity has been found to not only vary between individual flowers on the same plant, but also differ in concentration from the floral or leaf tissues within the plant *(13, 14)*. For example, an analysis of *N. glauca* nectar toxicity found that the average concentrations of alkaloids in *N. glauca* nectar were 0.5 ppm nicotine and 5.0 ppm anabasine, but higher concentrations were likely to be found in other plant tissues *(23, 24)*. Other compounds within the floral tissues could have also been introduced into the solution that may attract hummingbirds and cancel out the deterrent effect of the alkaloids. Because the experiment used the entire *N. glauca* flower to make the nectar solution, it is difficult to determine whether the solution contained enough toxicity to deter pollinators as compared to other studies. There is a category of compounds that contribute to nectar flavor and aroma of the flower known as terpenes; when paired with nectar naturally containing nicotine and anabasine, odor emitting monoterpenes were found to increase hummingbird nectaring time compared to nectaring time on flowers without the odor emitting compounds *(14)*. Due to the constraints of the project, measurements of the exact concentration of alkaloids, specifically nicotine and anabasine, or other floral compounds present in the *N. glauca* nectar solution were not recorded. Quantitative information about toxicity level would allow for a more precise future experiment aimed at understanding hummingbird, particularly *C. anna*'s, ability to tolerate these toxins.

Potential coevolution between *N. glauca* and *C. anna* could also account for *C. anna*'s ability to tolerate the nectar. Different pollination methods in *N. glauca* have been observed in different regions of the world; for example, in parts of Western Europe, *N. glauca* relies on self-pollination more than pollinators *(25)*. As a result, these *N. glauca* are not as toxic as their counterparts in South America. Although little research has been done on toxicity levels of *N. glauca* in California, native pollinators could have imposed selective pressure on the non-native plant to reduce its toxicity. The pollination syndrome in *N. glauca* in California might explain why *C. anna* is able to tolerate the nectar solution. Alternatively, it is possible that California native *C. anna* has developed toxin tolerance in response to the natural selection imposed by *N. glauca* nectar.

Lastly, the insignificant effect of nectar type or nectar concentration on hummingbird foraging behavior could support the theory that nectar toxicity may instead be a residual effect of the chemical defenses the plant will deploy in response to leaf herbivory *(13, 14)*. Other studies have revealed that leaf herbivory induces an increase in the nectar alkaloids that deter pollinators *(24, 26)*. It was not observed whether the *N. glauca* plants used to make the nectar had been subject to damage or leaf herbivory. Future research could perform a similar experiment but damage the leaves of the *N. glauca* plant prior to collecting flower samples. This would determine whether the interaction between leaf herbivory and nectar alkaloids has an effect on hummingbird foraging behavior. Studying nectar reward as a variable in hummingbird flower preference could be an important future direction, as it is still unclear what the strongest motivating factor is in hummingbird flower choice.

The experimental study highlighted that *C. anna* do not have a preference in floral color and nectar toxicity does not affect the length of time hummingbirds will spend at a given flower. Instead, features within their environment could have a larger influence on their foraging behavior. The finding sheds light on the complexities of plant-pollinator interactions and the limited knowledge of these systems. Studies on hummingbirds' foraging behaviors will help to inform the direction and scope of future hummingbird-plant coevolution imposed by the selective pressure from the hummingbirds on specific plant characteristics. Important future directions

include variations in toxic nectar creation and analysis, quality and quantity of nectar reward, consideration of other environmental variables such as nectar robbers, and time of year or duration of experiment. The work done in this experiment could easily be used in studies for different flower species and could be used to guide further studies on hummingbird behavior. Understanding interactions between flowering plants and pollinators is becoming more important amid the ongoing global extinction that has led to major losses in biodiversity across most species *(9)*. Biodiversity loss has been declared by the United Nations as a global emergency alongside climate change *(17)*. As such, gaining a more comprehensive understanding of pollination syndrome in flowering plants, which has created some of the most diverse species in history, is an essential step towards future conservation plans. The relationship between plants and pollinators, and the specific variables that drive each species' behaviors, can inform conservation experts as to how biodiversity can be both protected and restored.

ACKNOWLEDGEMENTS

This work was performed at the University of California's Sedgwick Reserve doi: https://doi.org/10.21973/N3C08R. We want to thank Dr. Tim Miller, Dr. Renske Kirchholtes, Tanner Dulay and the CEC Fall 2021 cohort. The support we received from them helped a simple curiosity grow into an entire research project. We want to thank Kate McCurdy for allowing us to set up our experiment plot in her yard, and for maintaining the beautiful and accommodating field station at Sedgwick Reserve. We finally thank the female hummingbirds that lived near Kate's yard for always being hungry.

REFERENCES

1. A.C. Lewis, Learning and the evolution of resources: pollinators and flower morphology. *Insect Learning*. 219-242 (1993). doi:10.1007/978-1-4615-2814-2_9.
2. C.I. Rodrígues-Flores, J.F. Ornelas, S. Wethington, M. del C. Arizmendi, Are hummingbirds generalists or specialists? Using network analysis to explore the mechanisms influencing their interaction with nectar resources. *PLOS ONE*. **14**, e0211855 (2019). doi:10.1371/journal.pone.0211855.
3. J. Fründ, K. E. Linsenmair, N. Blüthgen, Pollinator diversity and specialization in relation to flower diversity. *Oikos*. **119**, 1581-1590 (2010). doi:10.1111/j.1600-0706.2010.18450.x (4) Barrett, S. C. H. 2008. Major Evolutionary Transitions in Flowering Plant Reproduction: An Overview. International Journal of Plant Sciences, 169(1), 1–5. doi:10.1086/522511
4. S. H. Barrett, Major Evolutionary Transitions in Flowering Plant Reproduction: An Overview. *International Journal of Plant Sciences*. **169**, 1-5 (2008). doi:10.1086/522511
5. W.A. Guzman, P. Wilson, Hummingbirds at artificial flowers made to resemble ornithophilous versus melittophily. *Journal of Pollination Ecology*. **8**, 67-68 (2012). doi:10.26786/1920-7603(2012)9.
6. C. C. Missagia, M. A. Alves, Does beak size predict the pollination performance of hummingbirds at long and tubular flowers? A case study of a Neotropical spiral ginger. *J. Zool*. **305(1)**, 1-7 (2018). doi:10.1111/jzo.12539.
7. Melendez-Ackerman, P. Rico, Hummingbird behavior and mechanisms of selection on flower color in Ipomopsis. *Ecology (Durham)*. **78**, 2532-2541 (1997). doi:10.2307/2265912.
8. S. T. Hurly, S. D. Healy, Memory for flowers in rufous hummingbirds: location or local visual cues?. *Animal Behavior*. **51**, 1149-1157 (1996).
9. A. L. Laine, Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *Journal of Experimental Botany*. **60(11)**, 2957-2970 (2009). doi:10.1093/jxb/erp168.
10. R. K. Suarez, C. L. Gass, Hummingbird foraging and the relation between bioenergetics and behavior. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*. **133**, 335-343 (2002). doi:10.1016/s1095-6433(02)00165-4.
11. F. Soteras, M. Moré, A. C. Ibañez, M. dR. Iglesias, A. A. Cocucci, Range overlap between sword-billed hummingbird and its guild of long-flowered species: an approach to the study of coevolutionary mosaic. *PLOS*. **13**, e0209742 (2018). doi:10.1371/journal.pone.0209742.
12. O. P. Pearson, The daily energy requirements of a wild Anna hummingbird. *The Condor*. **56**, 317-322 (1954).
13. D. Kessler, S. Bhattacharya, C. Diezel, E. Rothe, K. Gase, et al., Unpredictability of nectar nicotine promotes outcrossing by hummingbirds in *Nicotiana attenuata*. *The Plant Journal: Cell and Molecular Biology*. **71**, 529-538 (2012). doi:10.1111/j.1365-313x.2012.05008.x.
14. D. Kessler, I. T. Baldwin, Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata* and function of floral secondary metabolites. *The Plant Journal*. **49**, 840-854 (2007). doi:10.1111/j.1365-313x.2006.02995.x.
15. J. Ollerton, S. Watts, S. Connerty, J. Lock, L. Parker, et al., Pollination ecology of the invasive tree tobacco *Nicotiana glauca*: comparisons across native and nonnative ranges. *Journal of Pollination Ecology*. **9**, 85-95 (2012). doi:10.26786/1920-7603(2012)12.
16. E. A. Issaly, A. N. Sérsic, A. Pauw, A. A. Cocucci, A. Traveset, et al., Reproductive ecology of the bird-pollinated *Nicotiana glauca* across native and introduced ranges with contrasting pollination environments. *Biological Invasions*. **22**, 485-498 (2020). doi:10.1007%2Fs10530-019-02104-8.
17. H. Tadmor-Melamed, S. Markman, A. Arieli, M. Distal, M. Wink, Limited availability of Palestine sunbirds *Nectarinia osea* to cope with pyridine alkaloids in nectar of tree tobacco *Nicotiana glauca*. *Functional Ecology*. **18**, 844-850 (2004). doi:10.3389%2Ffmicb.2021.639808.
18. J. L. Castorena, J. C. Garriott, F. E. Barnhardt, R. F. Shaw, A fatal poisoning from *Nicotiana glauca*. *Journal of Toxicology: Clinical Toxicology*. **25**, 429-435 (1987). doi:10.3109/15563658708992646.
19. D. L. Boose, Sources of variation in floral nectar production rate in *Epilobium canum* (*Onagraceae*): implications for natural selection. *Oecologia*. **110**, 493-500 (1997). doi:10.1007/s004420050185.
20. E. I. Greig, M. Wood, D. N. Bonter, Winter range expansion of a hummingbird is associated with urbanization and supplementary feeding. *Proceedings of the Royal Society B: Biological Sciences*. **284**, 20170256 (2017). doi:10.1098/rspb.2017.0256.
21. C. B. Fenster, G. Cheely, M. R. Dudash, R. J. Reynolds, Nectar reward and advertisement in hummingbird-pollinated *Silene virginica* (*Caryophyllaceae*). *American Journal of Botany*. **93**, 1800-1807 (2006). doi:0.3732/ajb.93.12.1800.
22. N. M. Waser, P. J. CaraDonna, M. V. Price, Atypical flowers can be as profitable as typical hummingbird flowers. *The American Naturalist*. **192**, 644-653 (2018). doi.:10.1086/699836.
23. R. L. Kaczorowski, A. Koplovich, F. Sport, M. Wink, S. Markman, Immediate effects of nectar robbing by Palestine sunbird (*Nectarinia osea*) on nectar alkaloid concentrations in tree tobacco (Nicotiana glauca). *Journal of Chemical Ecology*. **40**, 325-330 (2014). doi:10.1007/s10886-014-0411-7.

24. D. Roe, Biodiversity loss – more than an environmental emergency. *The Lancet Planetary Health*. **3(7)**, e287-e289 (2019). doi:10.1016/S2542-5196(19)30113-5.
25. S. K. Schueller, Island-mainland difference in *Nicotiana glauca* (Solanaceae) corolla length: a product of pollinator mediated selection?. *Evol Ecol*. **21**, 81-98 (2007). doi:10.1007/s10682-006-9125-9.
26. L. S. Adler, M. Wink, M. Distl, A. J. Lentz, Leaf herbivory and nutrients increase nectar alkaloids. *Ecology Letters*. **9**, 960-967 (2006). doi:10.1111/j.1461-0248.2006.00944.x.

Metabolic Regulation of Cell Identity and Therapy Response in Prostate Cancer

Aishwarya Atmakuri[1], Andrew S. Goldstein[1,2]

[1]Department of Molecular, Cell, and Developmental Biology, Univeristy of California, Los Angeles. [2]Department of Urology, Broad Stem Cell Research Center, Jonsson Comprehensive Cancer Center, and David Geffen School of Medicine, University of California, Los Angeles.

ABSTRACT

Prostate cancer growth is driven by the androgen receptor (AR) signaling axis. Androgen deprivation therapy (ADT) is standard treatment for prostate cancer, but the majority of ADT-treated patients develop resistance, which often involves the loss of luminal lineage identity and AR-independent growth. Modulating metabolism can regulate cell fate in many tissues. Therefore, it is important to understand the mechanisms underlying how altered metabolism affects lineage identity and response to AR blockade, a type of treatment for prostate cancer involving blocking AR signaling. Inhibition of mitochondrial pyruvate carrier, an essential metabolic enzyme, with small molecule UK5099 in mouse basal-derived nonmalignant organoids blocked luminal differentiation and UK5099-treated organoids retained a basal phenotype. AR expression decreased substantially in vehicle-treated organoids, but had a more modest decrease in UK5099-treated organoids in castrated conditions. The organoid size assay showed that in certain conditions, UK5099-treatment can decrease sensitivity to castration. An investigation into differentiation and metabolic pathways affected by UK5099-treatment revealed differential activation of nuclear factor kappa B, changes in β-catenin nuclear translocation, and increased levels of citrate synthase, alpha-ketoglutarate dehydrogenase, and malate dehydrogenase 2 in the nucleus. This demonstrated that UK5099 affects both differentiation and metabolism. When investigating specific mechanisms of UK5099 action, it was found that c-Myc signaling, lactate supplementation, and hypoxic conditions are able to phenocopy some of the effects of UK5099-treatment. Collectively, these results show the complexity in metabolic regulation of different cellular pathways and its effects on lineage identity and response to AR blockade. Modulating metabolism may potentially be a novel method of driving luminal differentiation to re-sensitize tumors to therapy.

INTRODUCTION

Prostate cancer and therapy resistance

The prostate secretes fluid that nourishes and transports sperm, making it an important component of the male reproductive system. Prostate cancer is the result of malignant, uncontrollable proliferation of a cell in the prostate. Globally, over 1.25 million men are diagnosed with prostate cancer, and over 350,000 men die from prostate cancer annually *(1)*. Prostate cancer growth is driven by the androgen receptor (AR), which is a transcription factor *(2,3)*. High serum levels of prostate specific antigen (PSA), a downstream AR target, have been used to screen for prostate cancer and indicate that increased AR activity is associated with prostate cancer *(3,4)*. AR becomes activated and translocates to the nucleus after binding to dihydrotestosterone (DHT), the more potent intracellular reduced form of testosterone *(2,3)*. AR is important for prostate cancer progression and migration *(5,6)*. Androgen deprivation therapy (ADT) is the first line of therapy against prostate cancer and targets AR (Figure 1A) *(7)*. ADT refers to either surgical or medical castration in order to decrease the production of androgens, and it is especially useful in cases of advanced prostate cancer and high-risk localized disease *(7)*.

However, ADT can select for phenotypes that lead to therapy resistance with AR-independent mechanisms of growth *(8)*. The majority of ADT-treated patients develop castration resistant prostate cancer (CRPC) after 2-3 years *(9)*. Several factors underlying CRPC have been identified, including ligand-independent AR activation, increased levels of AR expression, or mutations that make AR insensitive to anti-androgen treatment *(2,3)*. CRPC is traditionally treated with Enzalutamide, an androgen receptor pathway inhibitor (Figure 1B) *(10,11)*. While initially effective, most cancers develop resistance to Enzalutamide (Figure 1C). Analysis of Enzalutamide-resistant tumors shows that nonresponders possessed low AR transcriptional activity and a stem-like program, indicating AR-independent growth and proliferation *(12)*.

One mechanism behind this resistance is a change in lineage identity primarily characterized by a loss of luminal features and the development of basal and neuroendocrine features *(13,14)*. Basal and luminal cells are both epithelial cells in the prostate; basal cells sit on the basement membrane and contribute to the structural integrity of the prostate, while luminal cells produce secretory proteins important for prostate function *(15,16)*. Luminal cells can be identified using cytokeratin 8 (K8) and cytokeratin 18 (K18) and are highly AR-dependent for growth *(17,18)*. Basal cells can be identified using cytokeratin 5 (K5) and p63 and are mostly AR-independent *(18)*. Luminal cells are fully differentiated with limited progenitor capacity, while basal cells are the main source of prostate stem cells *(19)*. The plasticity of epithelial cell lineage

identity has been implicated in ADT resistance (19). Therapies targeting the AR signaling axis are effective against cells with a luminal identity, and the transition to a basal and/or neuroendocrine phenotype has been associated with therapy resistance (Figure 1D) *(13,20–23)*. Studying the mechanisms that regulate lineage identity can lead to new approaches to reduce therapy resistance and improve therapeutic outcomes for patients with prostate cancer.

Metabolism and lineage identity

Treatment-resistant prostate cancers are increasingly associated with a loss of luminal phenotype *(13)*. The Goldstein Laboratory has previously shown that altering metabolism through mitochondrial pyruvate carrier (MPC) inhibition through UK5099 treatment blocks luminal differentiation and promotes a more basal phenotype with increased glycolytic metabolism (Figure 1E). UK5099 is a small molecule that inhibits the MPC, which is responsible for transporting pyruvate into the mitochondria for the tricarboxylic acid (TCA) cycle. UK5099-treatment allows for the controlled modulation of metabolism, leading to increased glycolysis. UK5099-treatment in Lymph Node Carcinoma of the Prostate (LNCaP) prostate cancer cells has shown to increase glycolysis, decrease oxidative phosphorylation (OXPHOS), create a more stem-like phenotype, and increase resistance to chemotherapy (24). Other studies have found that metabolism can regulate cell fate in many tissues and also affect tumor formation *(25–28)*. The interplay between metabolism and lineage identity in the prostate should therefore be further elucidated to understand its role in therapy response and resistance.

Potential pathways

It was hypothesized that the change in lineage identity from modulating metabolism could be due to alterations in expression of key developmental pathways in the prostate, epigenetic changes in cell lineage identity from translocation of TCA cycle enzymes, c-Myc, and lactate metabolism *(8,29–32)*. Changes in these pathways have been associated with differentiation, lineage identity, epigenetics, zygotic gene expression, and metabolism *(30,33–35)*.

Experimental approach

This study examines how MPC inhibition alters sensitivity to AR blockade to understand how regulation of lineage identity through metabolism can affect response to therapy. Metabolic and differentiation pathways that are involved in UK5099 action will also be investigated. It is expected that more glycolytic conditions will lead to decreased response to ADT and that key regulatory developmental pathways, translocation of TCA cycle enzymes, c-Myc, and lactate metabolism in the prostate are involved in lineage identity transitions. Uncovering the role of metabolism and lineage identity in prostate cancer may provide strategies to direct luminal differentiation and drive resensitization to therapy.

METHODS

Organoid culture and collection

Primary mouse-derived prostate cells were collected from wild type mice. Prostate was dissected and then processed into single cells using dissociation media (Roswell Park Memorial Institute (RPMI) base media with 10% fetal bovine serum (FBS), 1% penicillin and streptomycin (P/S), collagenase, dispase, DNase, and Rho-associated, coiled-coil containing protein kinase (ROCK) inhibitor (RI) added). Cells were centrifuged, washed in phosphate buffered saline (PBS), and incubated in trypsin. Cells were syringed, filtered, centrifuged, and resuspended in dissociation media. Cell sorting antibodies and 4',6-diamidino-2-phenylindole (DAPI) were used to prepare cells for fluorescence-activated cell sorting. Basal cells were isolated using epithelial cellular adhesion molecule+ (EpCAM+) and cluster of differentiation $49f^{high}$ ($CD49f^{high}$) cells.

Single knock out (SKO) organoids were collected from phosphatase and tensin homolog (PTEN) floxed mice that had been infected with Cre. SKO organoids were used to conduct experiments in a cancer model, rather than a wild type (WT), benign model.

Primary mouse-derived basal cells were plated into Matrigel in organoid culture at 2500 cells/well on 24-well tissue culture plates with Poly(2-hydroxyethyl methacrylate) (poly-HEMA) coating with 350 µL to 1000 µL of fresh media based on organoid size. Media was changed every 48 hours. Control mouse organoid media consisted 34.5 mL advanced Dulbecco's Modified Eagle Medium (DMEM) F12, 5 mL R-spondin conditioned media, 1 mL B27, 500 µL Glutamax, 125 µL NAC, 50 µL Normocin, 50 µL Noggin, 50 µL DHT, 5 µL epidermal growth factor (EGF), and 5 µL A-8301. Castrated media was the same as control mouse organoid media, except with no DHT. Both UK5099 and Enzalutamide were added to media at 10 µM with vehicle control dimethyl sulfoxide (DMSO). Vehicle and UK5099 pretreatment began when cells were plated; Enzalutamide and castration treatment started 96 hours after plating.

Organoids were harvested using dispase containing media (1 mg dispase per mL advanced DMEM F12 + RI 1:1000). Media was removed, Matrigel was blasted with the harvesting media, and collected cells were incubated in harvesting media for 30 minutes to 1 hour. Cells were centrifuged, resuspended in PBS, and centrifuged again before PBS was removed. Cell pellets were flash frozen in dry ice and methanol and stored at -80˚C.

16D cell culture and collection

16D prostate cancer cells, established by Bishop *et al.*, were cultured in poly-L coated 10 cm tissue culture plates with RPMI media with 25 mM L-glutamine, 25 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES), 1% P/S, and 10% FBS *(36)*. Cells were passaged when confluent, every 2-4 days. Treatment with 10 µM Enzalutamide, MYCi361, 10074-G5, or 10-50 mM GSK 2837808A (GSK) was started at the same time as seeding the cells.

Cells were passaged by first removing media and then washing with PBS. Cells were incubated in trypsin in 37˚C for 5 minutes then quenched with media. Cell volumes were collected based on the passaging ratio, which ranged from 1:2 to 1:6, centrifuged at 1200 RPM for 5 minutes, resuspended into 10 mL of media, and plated into poly-L coated 10 cm tissue culture plates.

16D cells were collected by first aspirating media and washing with PBS. 850 µL – 1 mL of radio-immunoprecipitation assay (RIPA) buffer was added to each confluent plate and wells were scraped using a cell scraper. Collected cells were transferred into an Eppendorf tube, sonicated in ice water, and incubated on ice for 30 minutes. Samples were syringed 6 times with a 25-gauge needle and centrifuged at 17,000 g at 4˚C for 5 minutes.

Drugs

The different drug treatments were used to modulate metabolism, serve as an API, and block pathways of interest (Table 1). Each drug treatment was paired with appropriate vehicle treatments.

Drug	Purpose	Concentration Used	Potency	Manufacturer	Catalog Number
UK5099	MPC inhibitor	10 µM	49 µM (K_i)	Sigma-Aldrich	PZ0160-5MG
Enzalutamide	AR pathway inhibitor	10 µM	36 nM (IC_{50})	Selleck Chemicals	S1250
MYCi361	c-Myc inhibitor	0.5 µM	3.2 µM (K_d)	Selleck Chemicals	S8905
10074-G5	c-Myc inhibitor	0.5 µM, 1 µM, 5 µM	2.8 µM (K_d)	Selleck Chemicals	S8426
GSK 2837808A	LDHA and LDHB inhibitor	10 µM, 50 µM	LDHA: 2.6 nM (IC_{50}), LDHB: 43 nM (IC_{50})	Tocris	5189
DMSO	Vehicle			Thermo Fisher	BP231-100

Table 1: Drug treatments used in cell culture.
UK5099 and Enzalutamide were used to modulate metabolism and inhibit androgen receptor (AR) signaling, respectively in organoid and 2D cell culture models. MYCi361, 10074-G5, and GSK 2837808A were used to inhibit c-Myc and LDHA/B, respectively. Dimethyl sulfoxide (DMSO) is used as a vehicular control.

Western blot

Western blots were performed to visualize the abundance of different proteins of interest to identify changes in expression under certain conditions. Cell pellets were resuspended in 40 – 300 µL of RIPA buffer containing protease inhibitors and phosphatase inhibitors depending on pellet size. Pellets were sonicated and then placed on ice for 30 minutes. Pellets were stored at -20˚C.

A bicinchoninic acid (BCA) assay was used to find the concentration of protein in each sample. The assay generated a standard curve based on absorbance, which was used to calculate protein concentration of each sample and ensure even loading.

Each lane was loaded with 10 µg protein, loading buffer, reducing agent, and water to ensure even volumes. Master mixes for each sample were boiled at 70˚C for 10 minutes and then cooled on ice for 5 minutes. Samples were loaded onto NuPAGE™ 4 to 12% Bis-Tris gels and ran with running buffer (40 mL 20x MOPS NuPAGE SDS running buffer and 760 mL deionized (DI) H2O, inner chamber had added 500 µL NuPAGE antioxidant). Blue2 Plus protein ladder was used to identify protein location. Gels ran at 200 V for 50 minutes.

Gels were transferred onto a polyvinylidene fluoride (PVDF) membrane in a Wattmann paper sandwich after full submersion in transfer buffer (15 mL 20x NuPAGE transfer buffer, 0.3 mL NuPAGE antioxidant, 30 mL methanol, 255 mL DI H2O). Wet transfer ran at 30 V for 1 hour. Membrane was removed and allowed to dry after transfer.

SYPRO Ruby Stain was used to ensure even loading and the absence of air bubbles in regions of interest. The dried membrane was floated in fixation solution (5 mL methanol, 3.5 mL acetic acid, 41.5 mL water) for 15 minutes. Membrane was washed with water 4 times for 5 minutes and then floated face down in SYPRO Ruby for 15 minutes. Membrane was subsequently washed 4 more times for 1 minute and air dried.

Membrane was then trimmed to allow for multiple proteins to be probed for. Methanol-activated membrane was blocked in 5% PBS Tween (PBST) milk (Non-fat Dry Milk, LabScientific) for 1 hour. Primary antibodies were used for protein detection (Table 2). Membrane was incubated in primary antibody overnight at 4˚C.

Membrane was washed in PBST 3 times for 5 minutes. Membrane was then placed in 5% PBST milk with secondary antibody for 1 hour. Membrane was washed with PBST 3 times for 5 minutes for fluorescent detection and 6 times for 5 minutes for chemiluminescent detection.

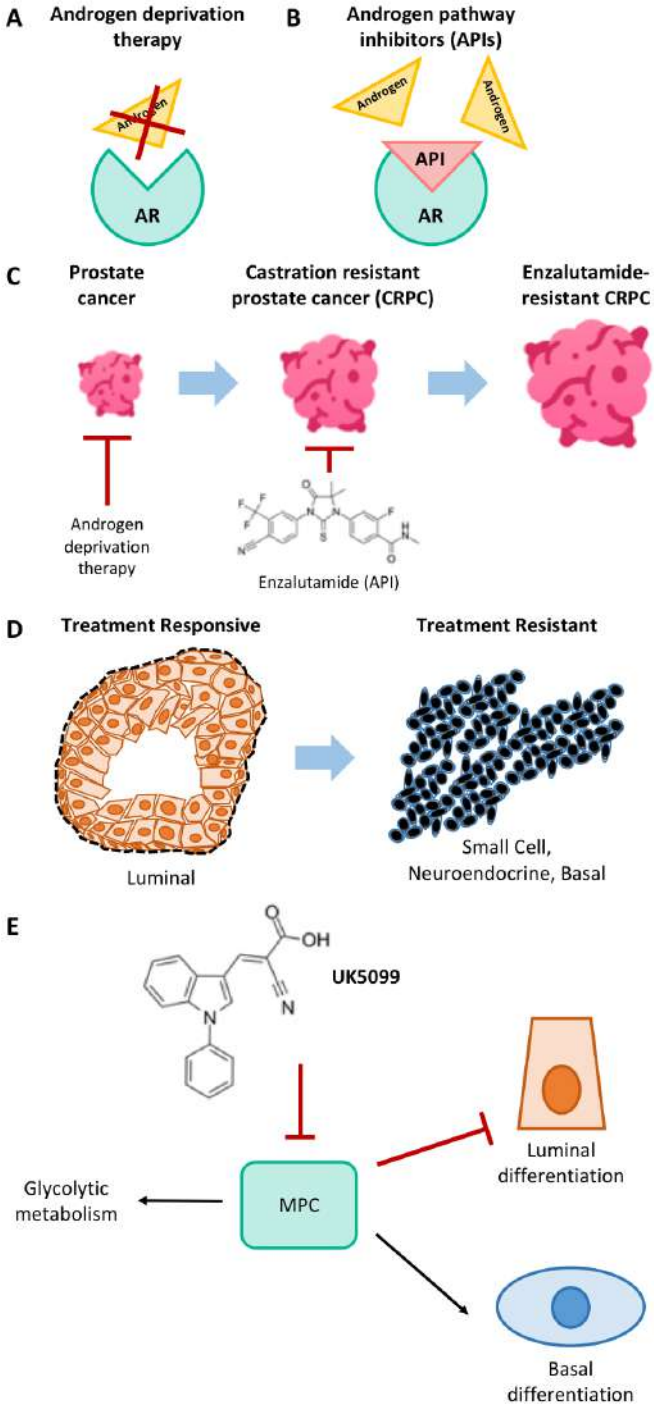


Figure 1: Schematic of prostate cancer progression and treatments.
(A) Androgen deprivation therapy (ADT) functions by medically or surgically decreasing levels of androgens. **(B)** Androgen pathway inhibitors (APIs) work by physically blocking androgen receptor (AR) binding and signaling. **(C)** Prostate cancer treated with androgen deprivation therapy can progress into castration-resistant prostate cancer (CRPC), which is treated with an API known as Enzalutamide. Most cases of CRPC progress to Enzalutamide-resistant CRPC. **(D)** Treatment responsive prostate cancer has a luminal phenotype, while therapy resistance is accompanied by a change in lineage identity to small cell, neuroendocrine, and basal phenotypes. **(E)** UK5099 inhibits the mitochondrial pyruvate complex (MPC), which results in increased glycolytic metabolism, blocking of luminal differentiation, and retention of a basal phenotype.

Fluorescent and chemiluminescent secondary antibodies were used (Table 3). Fluorescent detection used Alexa Fluor® 647 conjugated antibodies and the iBright 1500 imaging system. Chemiluminescent antibodies were read using horseradish peroxidase substrates.

Antibody	Dilution	Reactivity	Detection	Manufacturer	Catalog Number
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	1:1000	Rabbit	Fluorescent	Thermo Fisher	A21244
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	1:1000	Mouse	Fluorescent	Thermo Fisher	A21235
Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP	1:10000	Rabbit	Chemiluminescent	Thermo Fisher	31463
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP	1:10000	Mouse	Chemiluminescent	Thermo Fisher	31430

Table 3: Secondary antibodies.
Secondary antibodies were used during Western blot procedures to quantify proteins of interest by binding to primary antibodies. Secondary antibodies were designed for either fluorescent or chemiluminescent detection.

Organoid size quantification

Organoids were cultured in vehicle or 10 µM UK5099 media for 4 days pre-treatment followed by 6 days of treatment with control, castrated (-DHT), or 10 µM Enzalutamide. At the end of the treatment period, 25 representative organoids were randomly chosen and imaged on a phase contrast microscope. The diameter of each organoid was measured by digitally drawing a circle around the organoids. The diameter of the resultant circle was used to calculate the diameter of the organoids using a conversion factor determined by the scale bar. Organoids were individually quantified and normalized to the control mean. Statistical analysis was performed using a t-test with Welch's correction.

Subcellular fractionation

ThermoFisher Scientific Subcellular Protein Fractionation Kit was used for Cultured Cells (number 78840). The kit enables stepwise separation of cytoplasmic, membrane, nuclear soluble, chromatin-bound, and cytoskeletal protein extracts from whole cell lysate. First, cells were incubated in cytoplasm extraction buffer (CEB) for 10 minutes, centrifuged at 500 g for 5 minutes, and the supernatant-containing cytoplasmic extract was transferred to a separate tube. Cell pellets were washed and centrifuged with the same amount of PBST as CEB. Cells were then incubated in membrane extraction buffer (MEB) with protease inhibitors, vortexed, incubated for 10 minutes, centrifuged at 3000 g for 5 minutes, and supernatant containing membrane extract was transferred to a separate tube. Cell pellets were washed and centrifuged with the same amount of PBST. Cells were then incubated in nuclear extraction buffer (NEB) with protease inhibitors, vortexed, incubated for 30 minutes, and centrifuged at 500 g for 5 minutes. Supernatant containing soluble nuclear extract was transferred to a separate tube. Cell pellets were washed and centrifuged with the same amount of PBST. Cells were then incubated in NEB with protease inhibitors, CaCl2, and micrococcal nuclease, vortexed, incubated for 15 minutes, and centrifuged at 500 g for 5 minutes. Supernatant containing nuclear chromatin-bound extract was transferred to a separate tube.

Organoid transduction

Transduction was performed to overexpress or downregulate certain genes in organoids. Cells from organoids were resuspended in a small volume (~100 µL) of mouse organoid media containing polybrene and virus (~25 µL), incubated at 37˚C for 30 minutes, spun for 90 minutes at 1800 RPM, then plated using Matrigel.

16D cell transduction

Transduction was performed to overexpress or downregulate certain genes in 16D prostate cancer cells. 16D cells were transduced at 50% confluency. An 8 µM polybrene media master mix was created and split into 10 mL aliquots. The virus with the gene to be transduced was added to each aliquot. The original media was replaced with the virus media and left to incubate for 72 hours. Virus vectors were cloned in the Goldstein Laboratory.

Hypoxia

Hypoxic conditions were used to study increased anaerobic and lactate metabolism. Cells were incubated in a 1% oxygen incubator to induce hypoxic conditions. Hypoxic conditions were validated through measuring hypoxia-inducible factor 1-alpha (HIF1α).

RESULTS

Inhibiting mitochondrial pyruvate uptake blocks luminal differentiation

The Goldstein Laboratory has previously shown that mitochondrial pyruvate carrier inhibition using UK5099 blocks luminal differentiation and promotes a more basal phenotype. Given that prostate cancer growth is driven by AR and that existing therapies target the AR signaling axis, the effect of UK5099 on AR expression was investigated, which validated that change in luminal identity in basal-derived organoids. AR expression did not change between vehicle and UK5099 treatment. The luminal marker K8, however, did decrease in UK5099-treated organoids. Vehicle-treated basal-derived organoids differentiate from basal to luminal over the course of a week in organoid culture. Modulating metabolism through UK5099 treatment inhibited the basal to luminal differentiation, as indicated by low K8 expression (Figure 2A).

In a more detailed experiment looking at the effects of vehicle vs. UK5099-treatment in the presence of different modes of androgen blockade, Enzalutamide or chemical castration (-DHT), lineage identity was measured through luminal marker K8, basal marker K5, and basal marker p63Δn (Figure 2B, C). There was higher expression of K8 in vehicle-treated organoids compared to UK5099 illustrating that basal-derived organoids differentiate into a more luminal phenotype. UK5099 inhibits this lineage transition. In both vehicle and UK5099-treated organoids, castration decreased K8 expression. Enzalutamide increased K8 expression in UK5099-treated organoids. This showed that lineage identity is sensitive to modulations in metabolism and AR blockade. K5 expression was higher in UK5099-treated organoids, providing further evidence that UK5099-treated organoids retain their basal identity (Figure 2B). p63Δn was higher in control compared to castration in both groups and higher in UK5099-treated organoids compared to vehicle (Figure 2C). This indicated that UK5099-treated organoids have a greater basal phenotype and that castration decreases epithelial identity.

Blocking mitochondrial pyruvate uptake affects AR expression

Given the effect of UK5099-treatment on lineage identity, the effects of UK5099 treatment on sensitivity to androgen pathway inhibitors was investigated. UK5099-treatment was studied in combination with two different forms of androgen blockade: chemical castration or Enzalutamide.

Antibody	Protein Function	Size	Species	Dilution	Detection	Company	Catalog Number	Reactivity
Actin	Loading control	42 kDa	Mouse	1:15000	Fluorescent	Invitrogen	MA1-140	Human, mouse
AR		95 kDa	Rabbit	1:1000	Chemiluminescent	LS Bio	LS-C331862	Human, mouse
AR		95 kDa	Rabbit	1:2000	Fluorescent	Cell Signaling	5153S	Human
AR		98 kDa	Rabbit	1:1000	Chemiluminescent	Abcam	ab133273	Human, mouse
β-catenin	WNT/β-catenin pathway	86 kDa	Mouse	1:5000	Chemiluminescent	Proteintech	66379-1-Ig	Human, mouse
Citrate synthase	TCA cycle enzyme	47 kDa	Rabbit	1:1000	Fluorescent	Cell Signaling	14309S	Human, mouse
Cleaved caspase-3	Apoptosis marker	18 kDa, 15 kDa	Rabbit	1:1000	Chemiluminescent	Cell Signaling	9661L	Human, mouse
phospho-ERK1/2	Active EGFR pathway marker	42kDa	Rabbit	1:250	Chemiluminescent	Cell Signaling	4370	Human, mouse
ERK1/2	EGFR pathway	42 kDa	Mouse	1:1000	Fluorescent	Cell Signaling	4696	Human, mouse
HIF1a	Hypoxia marker	93 kDa	Rabbit	1:500	Chemiluminescent	GeneTex	GTX127309	Human, mouse
HK2	Glycolysis marker	90 kDa	Rabbit	1:1000	Chemiluminescent	Cell Signaling	28675	Human, mouse
Total histone H3	Nuclear chromatin bound loading control	14 kDa	Rabbit	1:1000	Fluorescent	Cell Signaling	9717S	Human, mouse
K5	Basal marker	55 kDa	Rabbit	1:3000	Fluorescent	Biolegend	905504	Human, mouse
K8	Luminal marker	52 kDa	Mouse	1:1000	Fluorescent (human), Chemiluminescent (mouse)	Biolegend	904804	Human, mouse
Lamin A	Nuclear soluble control	62 kDa	Rabbit	1:1000	Chemiluminescent	Abcam	Ab26300	Human, mouse
LDHA	Lactate metabolism	35 kDa	Mouse	1:1000	Fluorescent	Millipore Sigma	MABC150	Human
MCT1	Membrane control	54 kDa	Rabbit	1:1000	Fluorescent	Abcam	ab85021	Human
phospho-Myc	Active c-Myc	62 kDa	Rabbit	1:1000	Chemiluminescent	Cell Signaling	46650S	Human, mouse
Myc-c		52 kDa	Rabbit	1:1000	Chemiluminescent	Abcam	ab32072	Human, mouse
Nse	Neuroendocrine marker	45 kDa	Mouse	1:5000	Fluorescent	Proteintech	66150-1-Ig	Human, mouse
NUP62	Nuclear soluble control	62 kDa	Rabbit	1:1000	Fluorescent	Proteintech	13916-1-AP	Human, mouse
OGDH	TCA cycle enzyme	98 kDa	Rabbit	1:5000	Fluorescent	Proteintech	15212-1-AP	Human, mouse
p63△n	Basal marker	62 kDa	Rabbit	1:1000	Chemiluminescent	Biolegend	619002	Human, mouse
p65	NF-κB pathway	60 kDa	Rabbit	1:1000	Chemiluminescent	Cell Signaling	8242	Human, mouse
pAKT	Active EGFR pathway marker	52 kDa	Rabbit	1:1000	Chemiluminescent	Cell Signaling	3787	Human, mouse
PCNA	Proliferation marker	32 kDa	Mouse	1:1000	Fluorescent	Thermo Fisher	13-3900	Human, mouse
PDHA1	TCA cycle enzyme	43 kDa	Rabbit	1:1000	Fluorescent	Proteintech	18068-1-AP	Human, mouse
phospho-p65	Active NF-κB pathway marker	60 kDa	Rabbit	1:250	Chemiluminescent	Cell Signaling	3033	Human, mouse
PSA	AR target gene	33 kDa, 28 kDa, 24 kDa	Rabbit	1:1000	Fluorescent	Cell Signaling	5877	Human
Tubulin	Cytoplasm control	49 kDa	Mouse	1:500	Fluorescent	DSHB	12G10	Human, mouse
VDAC1	Membrane control	35 kDa	Mouse	1:1000	Fluorescent	Abcam	ab14734	Human, mouse
Vinculin	Loading control	105 kDa	Rabbit	1:8000	Fluorescent	Abcam	Ab129002	Human, mouse

Table 2: Primary antibodies.
Primary antibodies were used during Western blot procedures to quantify proteins of interest. Primary antibodies bound to specific proteins on a membrane.

AR expression was highest in vehicle control and Enzalutamide-treated groups. In vehicle treated organoids, AR expression was lower in castration than with Enzalutamide treatment. In UK5099 treated organoids, AR expression was lower with Enzalutamide than with castration (Figure 2B).

When evaluating proliferation and apoptosis, there were no large differences between samples in proliferating cell nuclear antigen (PCNA) or cleaved caspase 3, but there appeared to be slightly higher cleaved caspase 3 in both castration treatments, depicting greater cell death (Figure 2C).

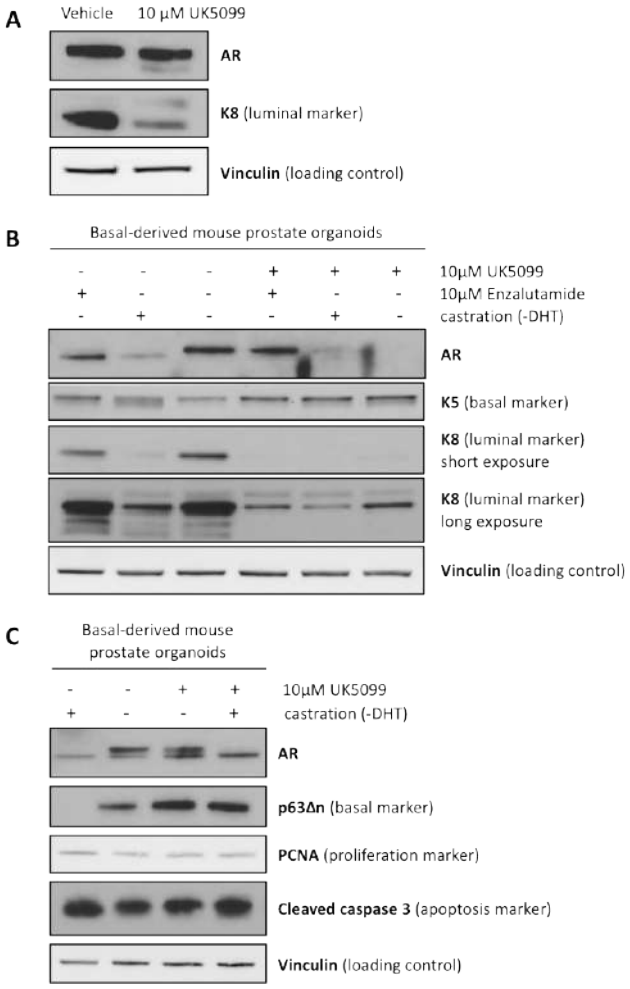


Figure 2: Lineage identity and growth response to androgen blockade and increased glycolysis.
(A) Western blot analysis of AR, K8 (luminal marker), and vinculin (loading control) was performed in vehicle vs. 10 μM UK5099-treated basal-derived primary mouse prostate organoids. UK5099-treatment showed decreased luminal identity. **(B)** Western blot analysis of basal marker cytokeratin 5 (K5) and luminal marker cytokeratin 8 (K8) with vehicle vs. 10 μM UK5099 pre-treatment was followed by 10 μM Enzalutamide or chemical castration (-Dihydrotestosterone (DHT)) in basal-derived primary mouse prostate organoids. UK5099-treatment altered responses to castration and Enzalutamide. **(C)** Western blot analysis of AR, p63Δn (basal marker), proliferating cell nuclear antigen (PCNA) (proliferation marker), and cleaved caspase 3 (apoptosis marker) with vehicle vs. 10 μM UK5099 pre-treatment was followed by chemical castration (-DHT) in basal-derived primary mouse prostate organoids. Basal identity increased with UK5099-treatment and remained high when treated with both UK5099 and castrated. Castration slightly increased apoptosis marker levels, but did not greatly affect proliferation markers.

Overall, treatment with UK5099 led to alterations in AR expression after treatment with different forms of androgen blockade.

Androgen blockade affects organoid growth

An organoid growth size quantification assay was developed to understand what regulates castration resistance in 3D *ex vivo* basal-derived organoid culture. Organoid growth was initially measured after 6 days of castration treatment (Figure 3A). Results show that castration significantly decreased relative size by 16% in vehicle-treated organoids ($p < 0.05$), yet appeared to increase relative size in UK5099-treated organoids (Figure 3B). Enzalutamide treatment had a similar but less pronounced effect (Figure 3B). Differences between vehicle, castration, and Enzalutamide treatment were greater in vehicle-treated organoids.

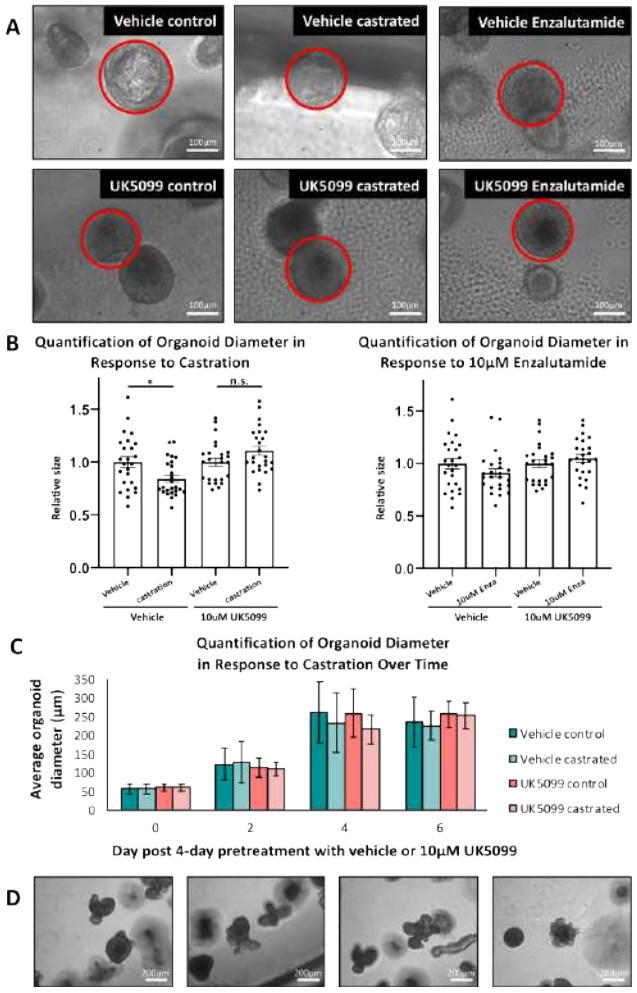


Figure 3: Androgen blockade affects organoid size.
(A) Representative images of basal-derived primary mouse prostate organoids with vehicle vs. 10 μM UK5099 pre-treatment followed by control, 10 μM Enzalutamide, or chemical castration (-DHT). **(B)** Quantification of relative mean organoid diameter based on 25 representative samples comparing castrated and 10 μM Enzalutamide-treated organoids. Vehicle castrated organoids were significantly smaller than vehicle control organoids ($p < 0.05$). **(C)** Time series of average quantified organoid diameter of basal-derived organoids with 4-day 10 μM UK5099 pre-treatment and 6-day castration. Trends in organoids size changed over time. **(D)** Day 4 UK5099-treated organoid blebbed phenotype.

The experiment was followed with a time series that measured organoid diameter every two days post castration onset (Figure 3C). On day 0, the mean vehicle and UK5099-treated organoid sizes were roughly the same (57 µm and 61 µm, respectively) with UK5099-treated organoids being slightly larger. On day 2, vehicle-treated organoids were on average 14 µm larger than UK5099-treated organoids, with minimal differences between castrated and control. On day 4, there was the greatest mean difference (35 µm) between control and castrated organoid size, but less differences between vehicle and UK5099-treated organoids. Many UK5099 organoids had a blebbed phenotype compared to vehicle-treated organoids on day 4 (Figure 3D). On day 6, the relative sizes of the different groups matched the previous trend seen in the initial experiment but were less drastic. UK5099-treated organoids were larger than vehicle-treated organoids, but there were smaller differences between control and castrated organoids. This showed that the effect of modulating metabolism on castrated organoids varies with the duration of castration. Overall, the data showed that modulating metabolism and castration can alter organoid growth. AR blockade affected organoid size and had differential effects with altered metabolism.

Blocking mitochondrial pyruvate uptake decreases TNF/NF-κB signaling and increases Wnt/ β-catenin signaling

After identifying that metabolism can affect lineage identity and response to ADT, potential mechanisms linking UK5099-treatment and blocking MPC with lineage identity were investigated, attempting to identify manipulations that are able to phenocopy UK5099-treatment.

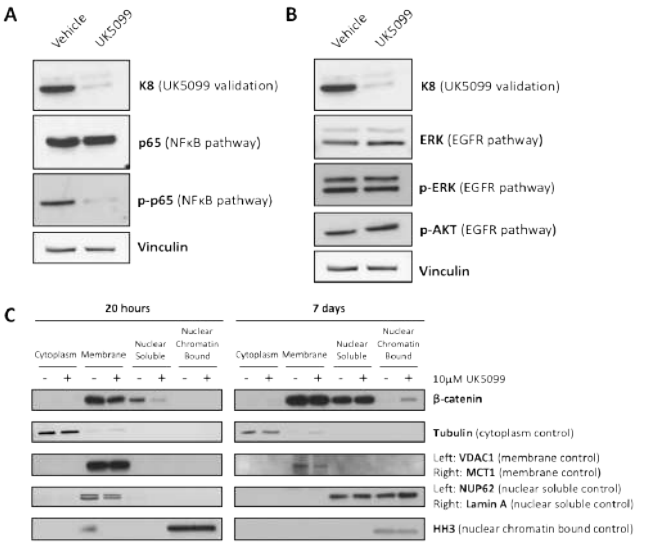


Figure 4: Protein abundance of pathways that regulate differentiation in the prostate.
(A) NF-B pathway in vehicle vs. UK5099-treated organoids after 7 days of treatment. UK5099-treatment decreased the amount of active NF-κB. **(B)** EGFR pathway vehicle vs. UK5099-treated organoids after 7 days of treatment. UK5099-treatment did not greatly affect activation and abundance of downstream EGFR signaling proteins. **(C)** Subcellular fractionation of Wnt/β-catenin pathway vehicle vs. UK5099-treated organoids in 20 hours (left) vs. 7 days (right) to assess β-catenin translocation. UK5099-treatment had less β-catenin translocation after 20 hours, but more after 7 days.

Expression of proteins associated with pathways that affect luminal differentiation in the prostate was evaluated in vehicle vs. UK5099-treated organoids. The tumor necrosis factor (TNF)/nuclear factor kappa B (NF-κB) pathway has been shown to suppress luminal differentiation, while Notch, Wnt/β-catenin, epidermal growth factor receptor (EGFR) and Kruppel-like factor 5 (KLF5) have been shown to promote luminal differentiation *(29,37,38)*.

In vehicle vs. UK5099-treated organoids, expression of p-p65 indicating activation of the NF-κB pathway decreased with UK5099-treatment along with decreased luminal expression shown by K8 (Figure 4A).

In vehicle vs. UK5099-treated organoids, expression of EGFR pathway indicators phosphorylated extracellular signal-regulated kinase (p-ERK) and phosphorylated protein kinase B (p-AKT) were relatively similar, indicating that the pathway likely does not play a major role in lineage identity shifts caused by UK5099 (Figure 4B).

A subcellular fractionation was used to visualize β-catenin translocation from the membrane to the nucleus where it acts as a transcription factor when activated. In vehicle vs. UK5099-treated organoids, there was less translocation of β-catenin to the nucleus at the 20-hour time point, but greater translocation at the 7-day timepoint (Figure 4C).

Blocking mitochondrial pyruvate uptake increases nuclear translocation of citrate synthase, OGDH, and MDH2

The translocation of TCA cycle enzymes into the nucleus has been shown to be associated with altered gene expression, epigenetic changes, and zygotic gene activation (Figure 5A) *(30,33–35)*. The subcellular location of several TCA cycle enzymes was observed with vehicle vs. UK5099-treatment. UK5099-treatment led to increased nuclear translocation of citrate synthase, alpha-ketoglutarate dehydrogenase (OGDH), malate dehydrogenase 2 (MDH2) (Figure 5B). Pyruvate dehydrogenase A (PDHA), an enzyme normally found in the mitochondrial membrane, also translocated to the nucleus with both vehicle and UK5099 treatment (Figure 5B).

The role of c-Myc in UK5099 action remains unclear

c-Myc is a major oncogene, and may be involved in the change in cell identity with UK5099-treatment, based on previous data from the literature and the Goldstein Laboratory *(39)*. When transfected with *c-Myc* to generate an overexpression (OE) in WT primary mouse prostate basal-derived organoids, the *c-Myc* OE did not phenocopy the inhibition of luminal differentiation as seen with UK5099-treatment as measured by K8 protein abundance through Western blot. Specifically, the *c-Myc* OE did not lead to the decrease in K8 that is seen with UK5099-treatment (Figure 6A, B).

c-Myc inhibitor MYCi361 was used in WT primary mouse prostate basal-derived organoids to see if c-Myc inhibition could rescue the UK5099 phenotype, indicated by detection of luminal marker K8 and absence of basal marker p63Δn. The results showed that MYCi361 alone has similar K8 and p63Δn readouts as with vehicle-treatment. In combination with UK5099, however, there was a loss of both basal and luminal epithelial identity indicated by the lack of K8 and p63Δn. However, it is unclear whether the inhibitor is working as there is no change in c-Myc or phosphorylated c-Myc levels (Figure 6C).

A second c-Myc inhibitor 10074-G5 was used to test the rescue of the UK5099 phenotype in a cancer context using single knock out (SKO) mouse organoids (missing PTEN). The inhibitor only appeared to block c-Myc at the 1 µM concentration in vehicle organoids. At the 1 µM concentration, there was a decrease

in both luminal and basal identity, marked by less K8 and p63Δn. UK5099-treated organoids have greater c-Myc abundance, and decreasing K8 levels with increasing c-Myc inhibition suggest a dose-dependent response (Figure 6D).

Increasing lactate abundance may affect response to modulation of metabolism

UK5099-treatment results in increased lactate production since there is greater pyruvate availability for conversion into lactate. GSK inhibits lactate dehydrogenase (LDH), the enzyme that generates lactate. Preventing lactate formation resulted in a slight increase in neuron-specific enolase (NSE) at 50 μM GSK, increase in PSA at 10 μM GSK, and decrease in c-Myc at 50 μM GSK in 16D human prostate cancer cell lines (Figure 7A). K8 levels remained unchanged (Figure 7A). While preventing lactate buildup through LDH inhibition may alter cell and tumor identity, decreased lactate buildup alone is insufficient in changing epithelial identity markers. It should be noted that the on-target effects of GSK still need to be validated.

In hypoxia, cell metabolism shifts towards anaerobic respiration, leading to increased amounts of lactate through lactic acid fermentation. If the mechanism behind UK5099 relies on increasing lactate, hypoxic conditions should phenocopy the effects of UK5099-treatment, lactate supplementation, and lactate carrier inhibition. Hypoxia and the effect on lactate were validated through increased abundance of LDHA in hypoxia samples in basal-derived primary mouse prostate cancer organoids (Figure 7B). Hypoxia also led to decreased K8 expression, which decreased further

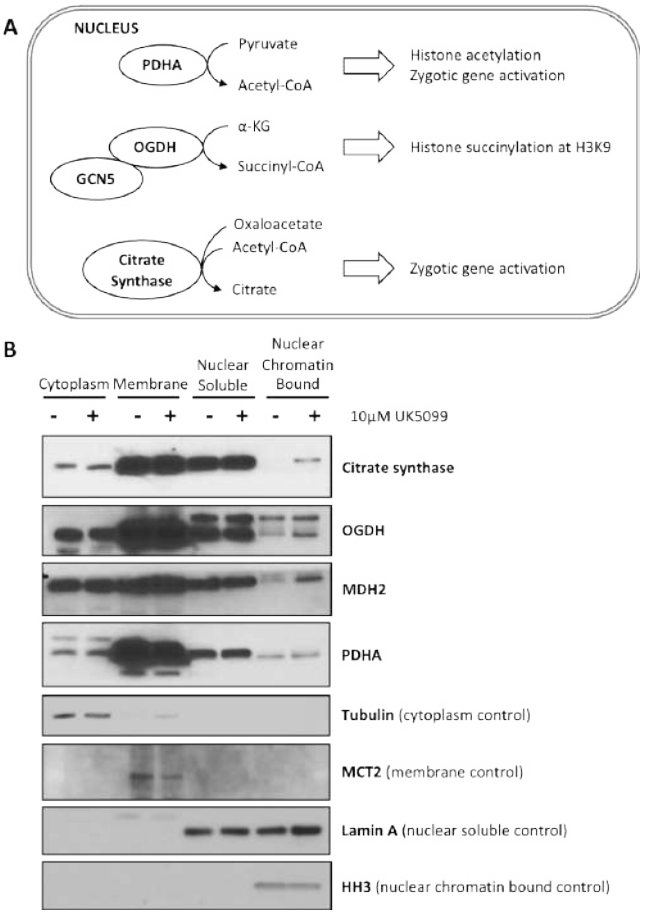


Figure 5: Protein expression and localization of citric acid cycle enzymes. **(A)** Nuclear functions of TCA cycle enzymes PDHA, OGDH, and citrate synthase and metabolites, which can regulate histones and gene expression. **(B)** Localization and expression of citrate synthase, OGDH, MDH2, and PDHA in vehicle vs. UK5099-treated organoids.

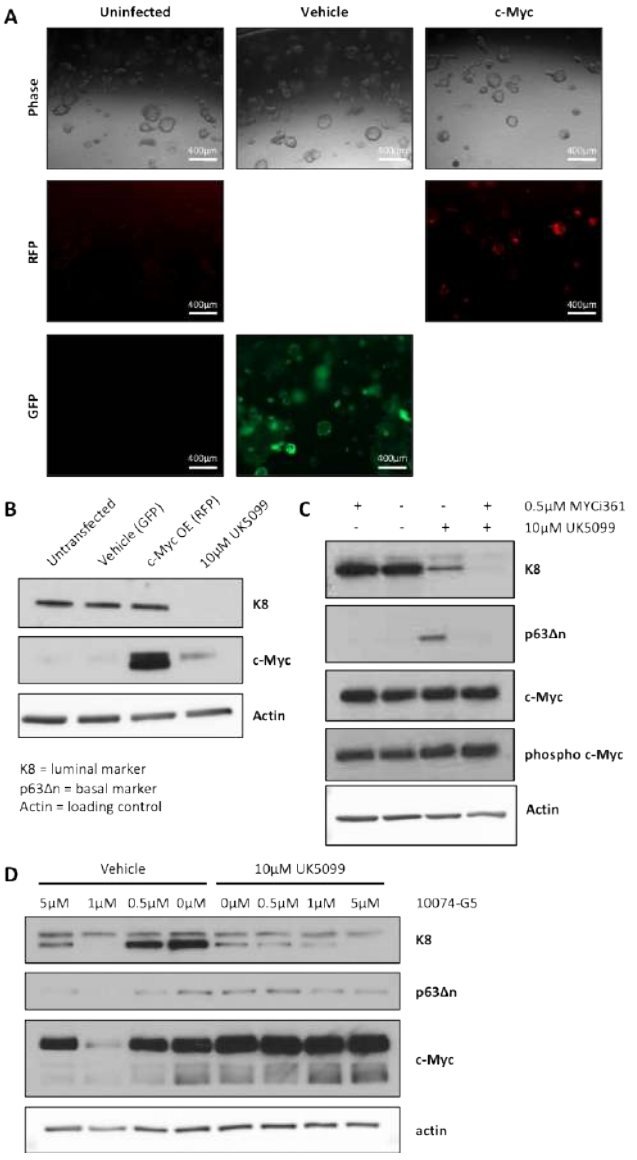


Figure 6: Examining c-Myc in the modulation of metabolism. **(A)** Fluorescent and phase images of uninfected, vehicle (GFP), and c-Myc (RFP) show successful infection. **(B)** Western blot of K8 and c-Myc to test if c-Myc overexpression phenocopies UK5099 treatment. c-Myc overexpression did not affect luminal marker expression, but UK5099-treatment increased c-Myc expression. **(C)** Western blot of K8 and p63Δn (basal marker) to test if c-Myc inhibition using MYCi361 rescues the UK5099 phenotype. Treatment with MYCi361 and UK5099 decreased luminal marker expression the most, but also decreased basal marker expression. **(D)** Western blot of K8, p36Δn, and c-Myc in single knock out (SKO) organoids (PTEN knockout). Treatment with 10074-G5 decreased markers of epithelial identity.

with UK5099-treatment. Interestingly, abundance of basal markers p63Δn and K5 also decreased in hypoxia but were partially rescued with UK5099-treatment. It remains unclear whether there was a decrease in luminal identity or epithelial identity as a whole. Both PCNA and c-Myc abundance decreased in hypoxic conditions, indicating decreased cell proliferation, but UK5099-treatment was partially able to rescue c-Myc (Figure 7B). This was also supported by the slightly smaller organoid size with hypoxia treatment (Figure 7C). Overall, hypoxia and the corresponding increase in lactate affected epithelial identity in some ways similar to UK5099-treatment.

DISCUSSION

This project sought to understand the effect of and mechanism behind modulating metabolism on lineage identity and AR blockade in basal-derived organoids. Clinically, therapy resistance is often accompanied by a loss of luminal identity and a gain of basal and neuroendocrine features *(14,39)*. It is essential to understand the mechanisms of this transition to learn how to drive luminal differentiation, increase sensitivity to therapy, and decrease resistance.

Since UK5099 has been shown to inhibit luminal differentiation in basal-derived organoids, it was originally hypothesized that they would be less sensitive to AR blockade given their basal characteristics. Several experimental findings support this hypothesis. K8, K5, and p63Δn expression shows that vehicle-treated basal-derived organoids are able to differentiate into luminal cells, but UK5099 blocks this transition and retains the basal phenotype. These findings are similar to the observations in the study by Mahendralingam et al. where glycolysis was enriched in basal cells and oxidative phosphorylation was enriched in luminal cells in mammary epithelial cells in the context of breast cancer *(27)*. Altered lineage identity may be related to the effects of castration on AR expression. AR expression in castrated organoids decreased far more in vehicle-treated organoids than UK5099-treated organoids, showing that AR expression in UK5099-treated castrated organoids is less dependent on external androgens.

The organoid size quantification assay, developed to understand what regulates castration experiments, was helpful in determining the role of AR blockade and altered metabolism. The assay suggests that UK5099-treated organoids are less sensitive to AR blockade. This can be seen where vehicle-treated organoids significantly decrease in size in castrated conditions, while UK5099-treated organoids actually appear to increase in size. However, these results are inconsistent over time, potentially indicating that the time after castration can affect relative differences in organoid size. The validity of the results of the time series could have been influenced by organoid shape and size. The blebbed phenotype of UK5099-treated organoids made it difficult to draw an accurate circle to determine organoid size on day 4. Additionally, many smaller organoids that previously could not be seen nor measured were able to be visualized on day 4, leading to greater variability in organoid diameter. In terms of AR expression, while there are no large differences between control vehicle and UK5099-treated organoids, the results are most distinct with AR blockade through castration or Enzalutamide. Collectively, the results show that altered metabolism can affect response to AR blockade.

These findings are in line with the clinical research study published by Tewari *et al.* Tewari *et al.* performed RNAseq on patients who were exceptional responders and non-responders to androgen

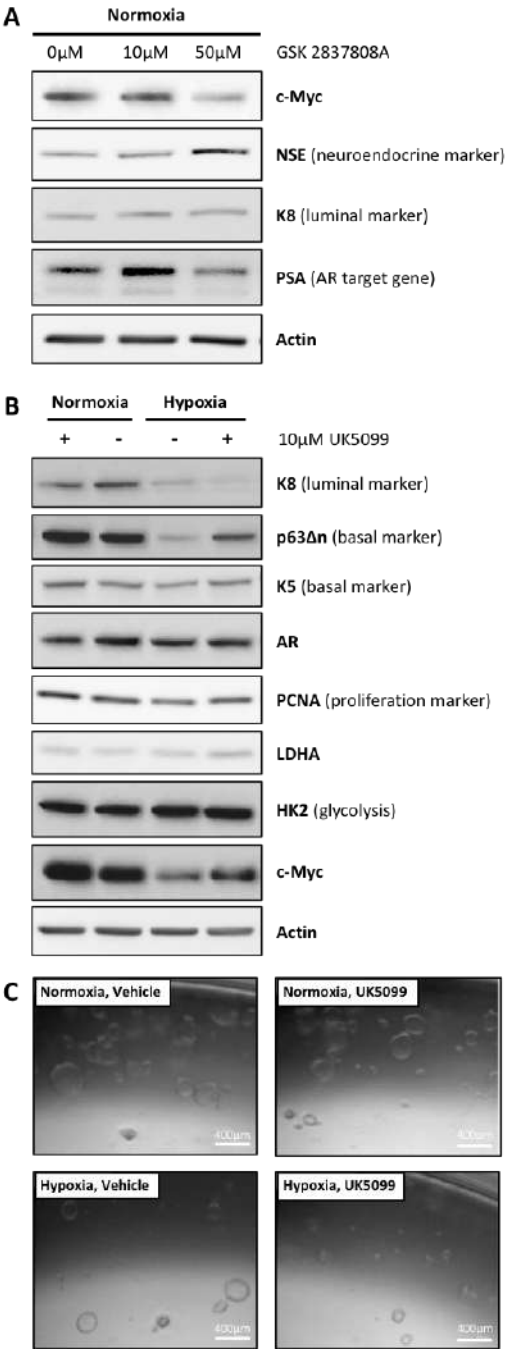


Figure 7: Examining the role of lactate and hypoxia in the modulation of metabolism.
(A) Western blot of c-Myc, neuron-specific enolase (NSE), K8, prostate-specific antigen (PSA), and Actin in 16D human prostate cancer cells treated with GSK 2837808A (GSK). GSK-treatment led to decreased c-Myc, increased NSE, and decreased PSA expression. **(B)** Western blot of K8 (luminal marker), p63Δn and K5 (basal markers), AR, PCNA (proliferation marker), lactate dehydrogenase A (LDHA), hexokinase 2 (HK2), and c-Myc with hypoxia and UK5099-treatment. Hypoxia and UK5099-treatment together decreased luminal marker expression, increased basal marker expression relative to hypoxia alone, and increased c-Myc expression relative to hypoxia alone. **(C)** Representative images of basal-derived primary mouse prostate organoids with hypoxia starting day 4 and 10 μM UK5099 1 day before harvesting. Hypoxic organoids are slightly smaller than normoxic organoids.

pathway inhibitor treatment *(40)*. Analysis by the Goldstein Laboratory showed that patients who were exceptional responders had a higher luminal signature and a lower basal signature than non-responders. Exceptional responders also had higher MPC1 and 2 expression than non-responders. These results in human prostate cancer correlate to this study's findings in mouse-derived nonmalignant basal organoids that MPC inhibition and a basal-phenotype decrease sensitivity to AR blockade.

Several developmental pathways have been shown to regulate differentiation in the prostate. The TNF/NF-κB pathway has been shown to suppress luminal differentiation while Notch, Wnt/β-catenin, EGFR, and KLF5 have been shown to promote luminal differentiation *(29,37)*. Altered expression of the TNF/NF-κB pathway and the Wnt/β-catenin signaling pathway suggests potential involvement in lineage identity shifts associated with UK5099-treatment. The Wnt/β-catenin pathway initially had higher expression in vehicle organoids at the 20 hour time point, but then had the opposite trend in the longer time point. It is possible that the pathways changed expression according to the literature soon after UK5099 treatment, but the observed time point allowed organoids to overcompensate for the change and pathways rebounded past normal expression levels. Although TNF/NF-κB and Wnt/β-catenin are altered with UK5099-treatment, it is unclear whether they are the mechanism of action in metabolic-induced lineage identity changes.

UK5099-treatment could result in the translocation of TCA cycle enzymes to the nucleus because they are no longer needed in the mitochondria. In the nucleus, TCA cycle enzymes have been shown to be associated with epigenetic changes that could affect cell lineage identity, chromatin dynamics, and zygotic gene expression *(30,34)*, providing potential insight on UK5099 action. PDHA generates acetyl-CoA, which promotes histone acetylation and gene expression *(33)*. OGDH and its product succinyl-CoA can induce histone succinylation at H3K9, which also promotes gene expression *(34,35)*. OGDH's substrate α-KG is important for demethylation of DNA and histones *(34)*. Several TCA cycle enzymes including citrate synthase, aconitase, isocitrate dehydrogenase, and PDH can provide metabolic fuels needed for zygotic gene activation *(30)*. Increased translocation of OGDH, MDH2, and citrate synthase with UK5099-treatment suggests that UK5099 may induce epigenetic changes that alter gene expression. The extra-mitochondrial localization of PDHA as observed with both vehicle and UK5099-treatment is a phenotype found in cancer cell lines and shown to be important for the activation of zygotic genes *(30)*, showing a potential connection between metabolic enzymes and differentiation.

c-Myc is a major oncogene and driver of prostate cancer growth *(31)*. Myc is an AR antagonist, specifically in castration resistant tumors *(38,41)*, so it is critical to understand its role in metabolism. Previous data has shown that MPC inhibition leads to increased c-Myc expression and activity. The Goldstein Laboratory has shown that c-Myc may be involved in lineage identity as c-Myc overexpression in long-term Enzalutamide-treated cells results in decreased luminal identity. Data from the Cancer Genome Atlas demonstrates a negative correlation between the Myc signature and luminal marker expression, suggesting that it could be involved in lineage identity. Experiments on c-Myc and lactate have shown that they are involved in cell lineage identity, but alone are not able to fully phenocopy the effects of UK5099. Given the effect of c-Myc treatment on altering epithelial identity, there is potential for synergistic effects with UK5099-treatment. Further studies should be done on elucidating the specific role of the c-Myc and lactate metabolism with UK5099-treatment.

It has also been observed that treatment with UK5099 increases intracellular lactate and glycolysis. Cancer cells preferentially utilize glycolytic metabolism, as shown in the Warburg effect. The upregulation of MCT transporters has been well-characterized *(32)*, suggesting potential involvement in the UK5099-mediated change in lineage identity. Exploration into the role of lactate in modulating the lineage identity transition through inhibition and hypoxia experiments have shown potential effects on epithelial identity. Hypoxia appears to be able to modulate the response to UK5099, suggesting involvement and further research. Additional experiments in the Goldstein Laboratory have shown that lactate supplementation can phenocopy the effects of UK5099-treatment. Given the role of lactate metabolism in cancer and observations in UK5099-treated organoids, lactate metabolism may influence lineage identity in the prostate and consequently affect response to therapy.

CONCLUSION

These experiments have underscored the complex interactions between metabolism, developmental pathways, epigenetic regulation from TCA cycle enzyme translocation, key oncogenes, lineage identity and plasticity, and response to therapy. The overarching effects of altering metabolism interact with several key cellular pathways. The isolation of a single pathway responsible for the effects of metabolism on lineage identity and therapy response is not representative of these intricate networks. The results show that a single pathway alone is unable to phenocopy the effects of UK5099 and prove to be its mechanism of action. Another challenge comes from the difficulty of rescue experiments, given the overarching effects of altering metabolism. It is difficult to separate the effect of metabolism on cell survival and lineage identity, which highlights the need for better model systems to exclusively study lineage identity. This research, however, elucidated the connection between alterations in metabolism, lineage identity, and sensitivity to therapy, while investigating the unknown mechanisms of UK5099 on lineage plasticity.

Further research should continue to elucidate the pathways involved in metabolism and lineage identity. Additional research should take place in a greater variety of *in vitro*, *ex vivo*, and *in vivo* model systems and investigate how altering different pathways affects sensitivity to AR-blockade based therapies. Experiments can also focus on testing the opposite effects of UK5099 by increasing OXPHOS and studying sensitivity to AR blockade. Understanding the mechanisms of UK5099-treatment may offer insight on driving luminal differentiation in the clinic to resensitize patients to therapies. Translating such findings may provide promising new strategies to harness the power of metabolism to identify novel methods of overcoming therapy resistance in prostate cancer.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Andrew Goldstein and Jenna Giafaglione for their guidance and mentorship of this project. The author would also like to thank all the members of the Goldstein Laboratory for their support. This project was funded in part by the Undergraduate Research Fellows Program.

REFERENCES

1. P. Rawla, Epidemiology of Prostate Cancer. *World J Oncol*. Metabolic regulation of cell identity and therapy response in prostate cancer. **10**, 63–89 (2019). doi:10.14740/wjon1191
2. K. Fujita, N. Nonomura, Role of Androgen Receptor in Prostate Cancer: A Review. *World J Mens Health*. **37**, 288–295 (2019). doi:10.5534/wjmh.180040
3. M. E. Tan, J. Li, H. E. Xu, K. Melcher, E. Yong, Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol Sin*. **36**, 3–23 (2015). doi:10.1038/aps.2014.18
4. R. M. Hoffman, Screening for Prostate Cancer. *N Engl J Med*. **365**, 2013–2019 (2011). doi:10.1056/NEJMcp1103642
5. G. Castoria et al., Androgen-induced cell migration: role of androgen receptor/filamin A association. *PLoS One*. **6**, e17218 (2011). doi:10.1371/journal.pone.0017218
6. C. A. Heinlein, C. Chang, Androgen receptor in prostate cancer. *Endocr Rev*. **25**, 276–308 (2004). doi:10.1210/er.2002-0032
7. N. Sharifi, J. L. Gulley, W. L. Dahut, Androgen Deprivation Therapy for Prostate Cancer. *JAMA*. **294**, 238–244 (2005). doi:10.1001/jama.294.2.238
8. M. Che *et al.*, Opposing transcriptional programs of KLF5 and AR emerge during therapy for advanced prostate cancer. *Nat Commun*. **12**, 6377 (2021). doi:10.1038/s41467-021-26612-1
9. T. Chandrasekar, J. C. Yang, A. C. Gao, C. P. Evans, Mechanisms of resistance in castration-resistant prostate cancer (CRPC). *Transl Androl Urol*. **4**, 365–380 (2015). doi:10.3978/j.issn.2223-4683.2015.05.02
10. M. Bishr, F. Saad, Overview of the latest treatments for castration-resistant prostate cancer. *Nat Rev Urol*. **10**, 522–528 (2013). doi:10.1038/nrurol.2013.137
11. H. I. Scher *et al.*, AFFIRM Investigators, Increased survival with enzalutamide in prostate cancer after chemotherapy. *N Engl J Med*. **367**, 1187–1197 (2012). doi:10.1056/NEJMoa1207506
12. J. J. Alumkal *et al.*, Transcriptional profiling identifies an androgen receptor activity-low, stemness program associated with enzalutamide resistance. *PNAS*. **117**, 12315–12323 (2020). doi:10.1073/pnas.1922207117
13. H. Beltran et al., The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. *Clin Cancer Res*. **25**, 6916–6924 (2019). doi:10.1158/1078-0432.CCR-19-1423
14. R. Tiwari *et al.*, Androgen deprivation upregulates SPINK1 expression and potentiates cellular plasticity in prostate cancer. *Nat Commun*. **11**, 384 (2020). doi:10.1038/s41467-019-14184-0
15. J. W. Park *et al.*, Prostate epithelial cell of origin determines cancer differentiation state in an organoid transformation assay. *Proc Natl Acad Sci U S A.* **113**, 4482–4487 (2016). doi:10.1073/pnas.1603645113
16. T. Kurita, R. T. Medina, A. A. Mills, G. R. Cunha, Role of p63 and basal cells in the prostate. *Development*. **131**, 4955–4964 (2004). doi:10.1242/dev.01384
17. Q. Xie *et al.*, Dissecting cell-type-specific roles of androgen receptor in prostate homeostasis and regeneration through lineage tracing. *Nat Commun*. **8**, 14284 (2017). doi:10.1038/ncomms14284
18. L. Xin, Cells of origin for cancer: an updated view from prostate cancer. *Oncogene*. **32**, 3655–3663 (2013). doi:10.1038/onc.2012.541
19. D. Zhang, S. Zhao, X. Li, J. S. Kirk, D. G. Tang, Prostate Luminal Progenitor Cells in Development and Cancer. *Trends Cancer*. **4**, 769–783 (2018). doi:10.1016/j.trecan.2018.09.003
20. S. Y. Ku *et al.*, Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*. **355**, 78–83 (2017). doi:10.1016/j.cmet.2019.01.016
21. A. Berger *et al.*, N-Myc–mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *J Clin Invest*. **129**, 3924–3940 (2019). doi:10.1172/JCI127961
22. P. Mu *et al.,* SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science*. **355**, 84–88 (2017). doi:10.1126/science.aah4307
23. A. Davies, A. Zoubeidi, L. A. Selth, The epigenetic and transcriptional landscape of neuroendocrine prostate cancer. *Endocr Relat Cancer*. **27**, R35–R50 (2020). doi:10.1530/ERC-19-0420
24. Y. Zhong *et al.*, Application of mitochondrial pyruvate carrier blocker UK5099 creates metabolic reprogram and greater stem-like properties in LnCap prostate cancer cells in vitro. *Oncotarget*. **6**, 37758–37769 (2015). doi:10.18632/oncotarget.5386
25. M. J. Rodríguez-Colman *et al.*, Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature*. **543**, 424–427 (2017). doi:10.1038/nature21673
26. Y. Yu *et al.*, Glutamine metabolism regulates proliferation and lineage allocation in skeletal stem cells. *Cell Metab*. **29**, 966-978.e4 (2019). doi:10.1016/j.cmet.2019.01.016
27. M. J. Mahendralingam *et al.*, Mammary epithelial cells have lineage-rooted metabolic identities. *Nat Metab*. **3**, 665–681 (2021). doi:10.1038/s42255-021-00388-6
28. S. C. Baksh *et al.*, Extracellular serine controls epidermal stem cell fate and tumour initiation. *Nat Cell Bio*. **22**, 779–790 (2020). doi:10.1038/s41556-020-0525-9
29. A. Centonze *et al.*, Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature*. **584**, 608–613 (2020). doi:10.1038/s41586-020-2632-y
30. R. Nagaraj *et al.*, Nuclear Localization of Mitochondrial TCA Cycle Enzymes as a Critical Step in Mammalian Zygotic Genome Activation. *Cell*. **168**, 210-223.e11 (2017). doi:10.1016/j.cell.2016.12.026
31. C. M. Koh *et al.*, MYC and Prostate Cancer. *Genes Cancer*. **1**, 617–628 (2010). doi:10.1177/1947601910379132
32. N. Pértega-Gomes, F. Baltazar, Lactate Transporters in the Context of Prostate Cancer Metabolism: What Do We Know? *Int J Mol Sci*. **15**, 18333–18348 (2014). doi:10.3390/ijms151018333
33. G. Sutendra, E. D. Michelakis, Pyruvate dehydrogenase kinase as a novel therapeutic target in oncology. *Front Oncol*. **3**, 38 (2013). doi:10.3389/fonc.2013.00038
34. X. Liu *et al.*, The existence of a nonclassical TCA cycle in the nucleus that wires the metabolic-epigenetic circuitry. *Sig Transduct Target Ther*. **6**, 1–18 (2021). doi:10.1038/s41392-021-00774-2
35. Y. Wang *et al.*, KAT2A coupled with the α-KGDH complex acts as a histone H3 succinyltransferase. *Nature*. **552**, 273–277 (2017). doi:10.1038/nature25003
36. J. L. Bishop *et al.*, The Master Neural Transcription Factor BRN2 Is an Androgen Receptor–Suppressed Driver of Neuroendocrine Differentiation in Prostate Cancer. *Cancer Discov*. **7**, 54–71 (2017). doi:10.1158/2159-8290.CD-15-1263
37. B. Zhang *et al.*, Klf5 acetylation regulates luminal differentiation of basal progenitors in prostate development and regeneration. *Nat Commun*. **11**, 997 (2020). doi:10.1038/s41467-020-14737-8
38. S. B. Frank, C. K. Miranti, Disruption of Prostate Epithelial Differentiation Pathways and Prostate Cancer Development. *Front Oncol*. **3**, 273 (2013). doi:10.3389/fonc.2013.00273
39. D. Tuerff, T. Sissung, W. D. Figg, Cellular identity crisis: Antiandrogen resistance by lineage plasticity. *Cancer Biol Ther*. **18**, 841–842 (2017). doi:10.1080/15384047.2017.1323599
40. A. K. Tewari *et al.*, Molecular features of exceptional response to neoadjuvant anti-androgen therapy in high-risk localized prostate cancer. *Cell Reports*. **36**, 109665 (2021). doi:10.1016/j.celrep.2021.109665
41. X. Qiu *et al.*, MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. *Nat Commun*. **13**, 2559 (2022). doi:10.1038/s41467-022-30257-z

Analytical and Chebyshev Filtering to Calculate Electron Densities Using the Tight Binding Model

Justin Purnomo[1], Daniel Neuhauser[1]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles.

ABSTRACT

Determining the chemical properties of a generic quantum system often involves the exact diagonalization of a convoluted Hamiltonian matrix in order to find the given system's energy eigenvalues and eigenvectors. However, the matrix diagonalization process is computationally costly, as it scales to the third power in terms of necessary operations. To circumvent matrix diagonalization, this paper presents two previously developed stochastic methods—analytical and Chebyshev filtering—that are able to estimate a given system's electron density in a more efficient manner. In order to better compare the respective computational costs of these filtering methods, the tight binding model was selected as the quantum system due to its low complexity. Application of these filtering methods on the aforementioned model has yet to be explored; however, it serves as an ideal candidate for determining electronic properties of large-scale systems. The tight binding model describes the system as real-space Hamiltonian matrices expressed on a manageable number of parameters, leading to substantially lower computational costs than using other models. By comparing the approximate densities obtained by these filtering methods to the analytical density obtained by matrix diagonalization, the accuracy of the methods in discerning physical properties of a given system can be gauged. This paper concludes that both analytical and Chebyshev filtering were effective in approximating the system's analytical density, with errors of less than 10 percent. Though analytical filtering demonstrated higher accuracy, Chebyshev filtering was more efficient since it avoided matrix diagonalization altogether. Furthermore, the error for both filtering methods decreased as the number of stochastic orbitals used increased.

INTRODUCTION

The density matrix, an $M \times M$ matrix that describes the quantum state of a physical system, is used to reveal microscopic details of molecules and materials *(1)*. Essentially, the density matrix operator is an alternate representation of the quantum state for which the more well-known wavefunction is often used. Whereas wave functions can only represent pure states, density matrices can also represent mixed states, which are states that consist of statistical ensembles of different quantum states *(2)*. Thus, they are invaluable tools in areas of quantum mechanics such as quantum statistical mechanics and quantum information that deal with mixed states.

The density matrix operator is able to yield the probabilities of the outcomes of any measurement performed upon this system. This is done by determining the matrix's eigenvectors and eigenvalues. However, extracting these eigenvectors and eigenvalues requires matrix diagonalization, which entails complex numerical calculations with computational costs of $O(M^3)$. With the rise of quantum computing in areas such as theoretical chemistry and physics, it is important to mitigate the aforementioned computational costs in memory and execution time. Achieving this generates unprecedented advantages, particularly in the areas of optimization and simulation *(3)*.

Density functional theory (DFT) uses matrix diagonalization to calculate the electronic properties of many-body systems with functionals of the electron density. Its rise in popularity can be attributed to the useful balance between accuracy and computational cost that these approximate functionals provide *(4)*. DFT provides a way to approach a complex interacting problem by mapping it to a simple non-interacting problem. As a result, more realistic systems such as molecules can be better studied giving DFT immense predictive power and potential for applications *(5)*.

Although useful, DFT's most prominent issue is its cubic scaling, which can be computationally costly given large systems. It is important to determine both how long a simulation is likely to take before beginning it and how much computational resources the simulation will necessitate. If the simulation requires unreasonable timescales or computer resources, then researchers would need to design a different study. At the forefront of tackling this problem is stochastic density-functional theory (sDFT), a tool that uses stochastic orbitals and a low-pass energy filter to investigate the same electronic structure of chemical systems as DFT at a lower computational cost *(6)*. By generating stochastic orbitals, which are random linear combinations of occupied and unoccupied orbitals, the stochastic density of chemical systems can be calculated without having to perform matrix diagonalization. The findings of Chen et al. demonstrate that sDFT leads to linear scaling, which is significantly less costly than the $O(M^3)$ computational costs mentioned previously *(7)*. Obtaining the system's electronic density in this way requires filtering of the system's unoccupied orbitals. Hence, performing both analytical and Chebyshev filtering on the

tight binding Hamiltonian model can calculate its stochastic density. Obtaining this density is essential for obtaining other physical properties about the system such as the potential and kinetic energies.

The tight binding model is a way to imitate a chain of atoms (*8*, Figure 1). This model associates a wavefunction for the entire system by a technique called the linear combination of atomic orbitals—specifically, a linear combination of the system's individual atomic wavefunctions. Hence, the tight binding Hamiltonian matrix must be adjusted to take into account the potential created by extra sites. Diagonalizing this matrix then yields the information needed to obtain the system's analytical density, which will be the standard to which the stochastic density will be compared.

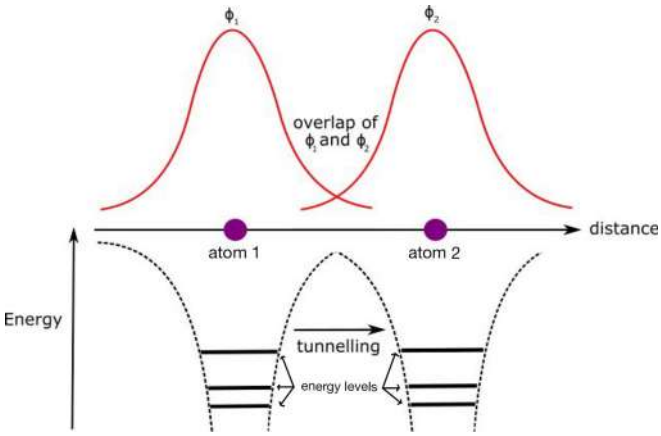


Figure 1: Visual representation of tight binding model.
The tight binding model for two atom sites in a chain. The wavefunctions for each site overlap slightly, and this overlap becomes larger when the distance between the atom sites is decreased. For greatest accuracy, the wavefunctions on each site should act as an isolated atom model. The small overlap should only reach into the neighboring site, as shown in the figure.

METHODS

Analytical density

The tight binding Hamiltonian matrix was adjusted to account for the potential created by extra sites. Specifically, this matrix was an $n \times n$ matrix with alternating off-diagonal entries. Additionally, its top left and bottom right entries were non-zero. This adjustment ensured the matrix was a periodic tight binding Hamiltonian matrix, so as to give a band gap, which is the energy range in which no electronic states can exist. Diagonalizing this matrix yielded its occupied eigenvectors ψ, allowing the analytical density $n(r)$ to be calculated by

$$n(r) = 2\sum_{i=1}^{\mathrm{N_{occ}}} |\psi_i(r)|^2 \tag{1}$$

where $\mathrm{N_{occ}}$ is the number of occupied orbitals.

Analytical filtering

The stochastic density was first calculated by analytical filtering, which uses the projection operator to project the stochastic orbitals χ onto the occupied eigenvectors ψ_j of the tight binding Hamiltonian matrix H in order to obtain the filtered orbitals ξ. These filtered orbitals were then used to calculate the stochastic density. The stochastic density was also calculated by Chebyshev filtering, which circumvents the need of possessing the matrix's eigenstates by expressing the projection operator in terms of Chebyshev polynomials (*9*). Analytical filtering uses the projection operator P_{occ}

$$P_{occ} = \sum_{j=1}^{Nocc} |\psi_j\rangle \langle\psi_j| \tag{2}$$

to project the occupied eigenvectors ψ_j of the tight binding Hamiltonian matrix H onto the stochastic orbitals χ in order to obtain the filtered orbitals ξ (*10*).

$$P_{occ} |\chi\rangle = |\xi\rangle \tag{3}$$

The system's stochastic density $n^s(r)$ was then found with the following equation.

$$n^s(r) = \frac{2}{\mathrm{N_{sto}}} \sum_{i=1}^{\mathrm{N_{sto}}} |\xi_i(r)|^2 \tag{4}$$

Though the occupied eigenvectors are not explicitly used in the stochastic density formula, they are still implicitly used by projection. Thus, analytical filtering assumes that the original matrix's eigenvectors have already been obtained through matrix diagonalization. However, analytical filtering is important since it serves as a debugging tool for Chebyshev filtering, which circumvents matrix diagonalization altogether. The formulation of this filtering method and the subsequent algorithm implemented in this method is what is wholly original in this experiment.

Chebyshev filtering

Chebyshev filtering represents the projection operator that analytical filtering uses as the error function

$$f(x) = \frac{1}{2} erfc(\beta(x - \mu)) \tag{5}$$

Doing so allows further Chebyshev expansion on this error function in terms of Chebyshev polynomials. However, the Chebyshev expansion is only reliable for $x \in [-1,1]$, meaning that the Hamiltonian matrix must be scaled so that all of its eigenvalues are between -1 and 1. Scaling the Hamiltonian matrix requires an algorithm to numerically determine the maximum and minimum eigenvalues of a matrix, which will be used to find the important values needed in the scaled Hamiltonian formula. The aforementioned algorithm applies the Hamiltonian n times to a wavefunction (which does not have to be an eigenstate of H) where n is a large number.

$$H^n |\psi\rangle = H^n \sum_i a_i |\phi_i\rangle = \sum_i a_i E_i^n |\phi_i\rangle \tag{6}$$

Defining C_n as the following quantity

$$C_n = \frac{\langle\psi|H^n|\psi\rangle}{\langle\psi|H^{n-1}|\psi\rangle} = \frac{\sum_i a_i E_i^n}{\sum_i a_i E_i^{n-1}} \tag{7}$$

and the largest magnitude, E_1, which can be E_{max} or E_{min}, it can be seen that as n approaches infinity, the eigenvalues with magnitudes not $|E_1|$ become much smaller than $|E_1|^n$ and that

$$\lim_{n\to\infty} C_n = |E_1| \tag{8}$$

Numerically, $|E_1|$ can be approximated

$$B = \lim_{n \to N} C_n \tag{9}$$

In this paper, N was 5,000. To find the next highest magnitude eigenvalue $|E_2|$, a modified Hamiltonian was defined as where the estimate of E_1, is subtracted off.

$$H_{mod} = H - B \tag{10}$$

Here, the estimate of $|E_2|$, B' was equal to $|E_{max}\text{-}E_{min}|$. This eigenvalue was found in the identical manner as B except using the modified Hamiltonian in Eq. 3. By defining $B''=B+B'$, B'' could be E_{max} or E_{min}, where E_{max} and E_{min} were the larger and smaller quantities of B and B'' respectively. B and B' were then calculated by using a for/do loop. To accommodate the large N value of 5,000 mentioned previously, $\langle \psi | H^n | \psi \rangle$ was normalized after every iteration in order to prevent the eigenvalue magnitudes from increasing exponentially. Defining E_3 as equal to E_2+E_1, E_{max} and E_{min} are the larger and smaller values between E_1 and E_3, respectively.

h_{avg} is defined as

$$h_{avg} = \frac{E_{max} + E_{min}}{2} \tag{11}$$

and dh as

$$dh = \frac{E_{max} - E_{min}}{2} \tag{12}$$

The scaled Hamiltonian matrix was then generated with the following formula:

$$H_{scaled} = \frac{H - h_{avg} I}{dh} \tag{13}$$

Checking the accuracy of the scaled Hamiltonian involved verifying that all of its eigenvalues were between -1 and 1, which was done by a matrix diagonalization subroutine in Fortran.

The function used to approximate the projections of analytical filtering is of the form

$$\theta(\mu - H) = \sqrt{erfc(\beta(\mu - H))} \tag{14}$$

where $\mu=\frac{1}{2}(\epsilon_H+\epsilon_L)$ is the chemical potential of the system and β is chosen such that the function completely decays in the band gap. Here ϵ_H and ϵ_L represent the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), respectively. Note that they are equal to the N_{occ}th and $(N_{occ}+1)$th eigenvalues of the Hamiltonian matrix, respectively.

Denoting the HOMO and LUMO of the scaled Hamiltonian matrix as $\bar{\epsilon}_H$ and $\bar{\epsilon}_L$, the following two formulas were used to obtain these values.

$$\bar{\epsilon}_H = \frac{\epsilon_H - h_{avg}}{dh} \tag{15}$$

$$\bar{\epsilon}_L = \frac{\epsilon_L - h_{avg}}{dh} \tag{16}$$

We then found the scaled chemical potential $\bar{\mu}$ with the following formula.

$$\bar{\mu} = \frac{\bar{\epsilon}_H + \bar{\epsilon}_L}{2} \tag{17}$$

The last parameter to be discerned was $\bar{\beta}$. Similar to β, this was found by determining the value at which the scaled function completely decays in the band gap. This culminated in the following function that would be used in the proceeding steps:

$$\hat{\theta}(\bar{\mu} - \bar{H}) = \sqrt{erfc(\bar{\beta}(\bar{\mu} - \bar{H}))} \tag{18}$$

Recall that the analytical filter, the projection of the occupied states of the Hamiltonian onto χ, is $\hat{\theta}_{ana} \approx \sum_{i=1}^{N_{sto}} |\psi_i\rangle\langle\psi_i|$. The desired Chebyshev filter $\hat{\theta}_{ana} \approx \sum_{n=0}^{npoly} a_n T_n \bar{H}$ approximated $\hat{\theta}_{ana}$ to filter out the unoccupied contributions in the stochastic orbitals χ. Note that *npoly* amount of Chebyshev polynomials T_n are not needed due to the recursion relation of Chebyshev polynomials with $n \geq 2$:

$$T_{n+1}(\bar{H}) = 2\bar{H}T_n(\bar{H}) - T_{n-1}(\bar{H}) \tag{19}$$

Hence, the three original vectors are the only ones needed, namely $T_0(\bar{H})=1$, $T_1(\bar{H})=\bar{H}$, and $T_2(\bar{H})=2\bar{H}\ T_1(\bar{H})\text{-}T_0(\bar{H})=2\bar{H}^2\text{-}1$. Denoting $T_0(\bar{H})$, $T_1(\bar{H})$, and $T_2(\bar{H})$ as the temporary vectors $|V_n\rangle$, $|V_{n-1}\rangle$, and $|V_{n-2}\rangle$, respectively, a do loop can be used to redefine the vectors after each iteration, allowing the calculation of the filtered orbital ξ.

μ Algorithm

Discerning the exact chemical potential μ requires an algorithm. Since we used stochastic orbitals, μ skewed closer either to the system's LUMO or HOMO. The aforementioned algorithm generates a set of stochastic orbitals, and uses a singular vector for calculating the electron density. Then it searches for a value μ such that the following equation holds true.

$$\mathrm{N_e} = 2 \sum_{\mathrm{n=0}}^{\mathrm{npoly}} \mathrm{a_n}(\mu, \beta) \mathrm{R_n}(\bar{\mathrm{H}}) \tag{20}$$

where N_e is the number of electrons, *npoly* is the number of Chebyshev polynomials, a_n are the Chebyshev coefficients, R_n are the residues, and $\bar{H}$ is the scaled Hamiltonian matrix.

Taking advantage of the recursion relation of the Chebyshev polynomials T_n mentioned earlier, the residues were calculated. This requires only the three vectors to begin with, $T_0(\bar{H})=1$, $T_1(\bar{H})=\bar{H}$, and $T_2(\bar{H})=2\bar{H}\ T_1(\bar{H})\text{-}T_0(\bar{H})=2\bar{H}^2\text{-}1$. The vectors are redefined after each iteration in order to obtain all Chebyshev polynomials. Then using

$$\mathrm{R_n}(\bar{\mathrm{H}}) = \frac{1}{\mathrm{N_{sto}}} \sum_{\mathrm{i=1}}^{\mathrm{N_{sto}}} \langle \chi_\mathrm{i} | \mathrm{T_n}(\bar{\mathrm{H}}) | \chi_\mathrm{i} \rangle \tag{21}$$

where χ_i represents a random vector, the residues were calculated. Note that $T_n(\bar{H})|\chi_i\rangle$ for $i=0,2$ were calculated and these results were called ϕ_i. As mentioned before, since ϕ_i is redefined after each iteration, the entire residue array R_n could be calculated.

Once the residues had been calculated, the number of electrons were then found using the following formula.

$$\mathrm{N_{calc}} = \sum_{\mathrm{i=1}}^{\mathrm{npoly}} \mathrm{a_n}(\bar{\mu}, \beta)\mathrm{R_n} \tag{22}$$

Since R_n is known, the next step was to solve for the coefficients a_n. This can be done in two ways: either through Fourier coefficients, obtained by calling a Fast Fourier Transform subroutine, or through Chebyshev coefficients.

Next, N_{calc} was solved for 500 different $\bar{\mu}$ values, by using $\sum a*R$ for each iteration. Varying $\bar{\mu}$ would vary how many unoccupied orbitals got filtered. Hence, a higher $\bar{\mu}$ value would mean that there are more electrons in the system (Figure 2).

The plateau between -0.2 and 0.2 can be explained by the fact that the actual $\bar{\mu}$ value is between the HOMO and the LUMO. There exists a band gap between the HOMO and the LUMO, which is the energy range in which no electronic states can exist. The next step was to find the exact $\bar{\mu}$ value by using binary search.

Binary search, or binary chop, is a search algorithm that locates the position of a target value in a sorted array. It does this by comparing the target value to the middle element of the array. If they are unequal, the half in which the target value cannot lie is eliminated. The search then continues with the remaining half, again comparing the target value to the middle element. This process repeats itself until the target value is found. If the search ends with the remaining half being empty, the target is not in the array.

In this case, the target value $\bar{\mu}$ is a number such that Equation 9 holds true. The initial middle point Es_{mid} will be defined to be the midpoint of the highest and lowest $\bar{\mu}$ values, Es_{max} and Es_{min}. By using a series of if statements, the halves of the array in which the target value cannot exist can be eliminated. This was done by redefining Es_{max} and Es_{min} after each iteration. If $N_{calc}>n$, then Es_{max} is defined to be Es_{mid}. If $N_{calc}<n$, then Es_{min} is defined to be Es_{mid}. Eventually, $\bar{\mu}$ will equal Es_{mid} and then the target value can be used $\bar{\mu}$ for Chebyshev filtering, which was already completed. Note that in the actual code, N_{calc} is assumed to be equal to n once their absolute difference $|N_{calc}-n|$ is less than 1*E*-10.

Upon performing the binary search algorithm, the target value $\bar{\mu}$ is obtained.

RESULTS

Analytical vs. Chebyshev filtering accuracy

This section compares the difference between the analytical density $n(r)$ and the stochastic densities $n^a(r)$ and $n^c(r)$ obtained from both of the methods applied on the tight binding model. $n^a(r)$ and $n^c(r)$ represent the stochastic density obtained from analytical and Chebyshev filtering, respectively. Table 1 displays the stochastic density error as a function of the number N_{sto} of stochastic orbitals used. The second column contains the results for analytical filtering, while the third column contains the results for Chebyshev filtering. Note that for each Chebyshev filtering data point, the same number of Chebyshev polynomials (i.e. - 1000) was used in the Chebyshev expansion to ensure consistency. Figure 3 displays the same data as a graph, showing the error for both filtering methods decreasing as a function of the number of stochastic orbitals used. The stochastic error was defined to be the maximum difference between the stochastic density and the analytical density. The error is measured by a single point basis to ensure that other factors (i.e. - N_{sto}) that might affect the accuracy of the results are considered.

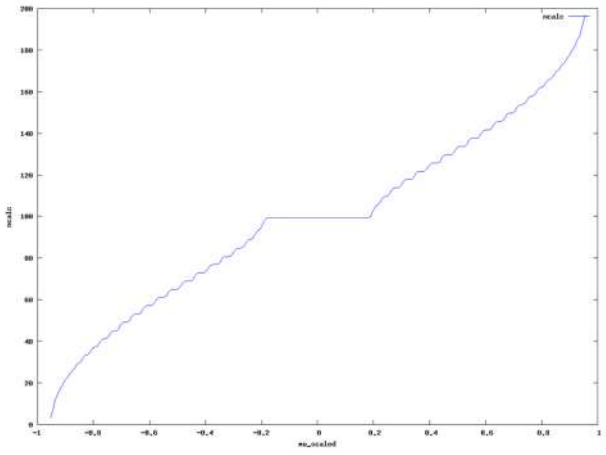


Figure 2: Graph of $\bar{\mu}$ vs. *ncalc*.
The number of electrons in the chemical system, ncalc, as a function of the scaled chemical potential $\bar{\mu}$ In accordance with the error function used that approximates the projection operator, a higher $\bar{\mu}$ value would mean that there are more electrons in the system, as less orbitals would be filtered out. The plateau between -0.2 and 0.2 is explained by the fact that the actual $\bar{\mu}$ value is between the HOMO and the LUMO. There exists a band gap between the HOMO and the LUMO, which is the energy range in which no electronic states can exist.

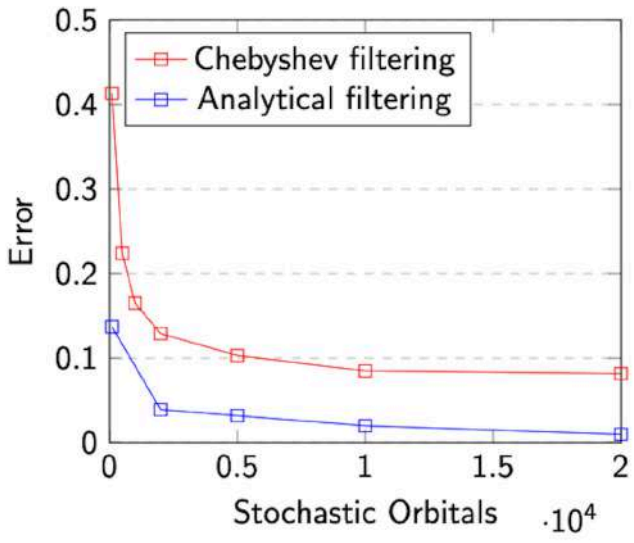


Figure 3: Stochastic density error dependence on number of stochastic orbitals for analytical and Chebyshev filtering.
The error of Chebyshev filtering and analytical filtering, as functions of the number of stochastic orbitals used. Chebyshev filtering yields greater errors than analytical filtering, indicating that Chebyshev filtering is less accurate than analytical filtering. Additionally, the errors for both filtering methods decrease as the number of stochastic orbitals used increases, indicating that greater accuracy can be achieved with greater stochastic samplings.

Calculating global error utilizing some sort of averaging or integration would otherwise disregard such influential factors.

Analytical vs. Chebyshev Filtering Efficiency

In addition to the accuracy of analytical and Chebyshev filtering, their efficiencies are also compared. Table 2 compares the computer run time in seconds of both filtering methods as a function of N_{sto}. This was done by subtracting the end time t_{end} from the start time t_{start}.

DISCUSSION

Approximation methods such as projections and polynomial expansions—the analytical and Chebyshev filtering—address the problem of computational costs hindering a complete understanding of a chemical system's physical properties. The error differences between the analytical and stochastic densities of a given chemical system measure the accuracy and efficiency of these filtering methods.

According to the results, both filtering methods were effective in approximating the system's analytical density, as the error for both the analytical and Chebyshev filtering results were below 10% once enough stochastic orbitals were used. Additionally, the stochastic density obtained from analytical filtering was more accurate than the stochastic density obtained from Chebyshev filtering. The error for analytical filtering was below 5% when using just 2,000 stochastic orbitals, while it took 20,000 stochastic orbitals for the error to reach 8% for Chebyshev filtering. These findings make intuitive sense, as analytical filtering used the exact eigenvectors acquired from matrix diagonalization. When taking this into consideration, it is clear that analytical filtering and the full analytical method demonstrate equivalent computational efficiencies. An 8% error is still acceptable, as it falls within the 10% error benchmark considered to show significant results. Note that the physical consequences of error entail influencing calculations of other properties of the system. Since these properties are calculated by using the density, an error in the density can lead to an error in the aforementioned calculated properties. Furthermore, the results show that the error for both filtering methods decreased as the number of stochastic orbitals increased. Generating more stochastic orbitals means more projections, thus corroborating the higher degree of accuracy.

N_{sto}	maxval$\lvert n^a(r) - n(r)\rvert$	maxval$\lvert n^c(r) - n(r)\rvert$
100	0.137	0.413
500	0.0976	0.224
1000	0.0631	0.165
2000	0.0391	0.129
5000	0.03194	0.103
10000	0.0199	0.0849
20000	0.00987	0.0816

Table 1: Stochastic Orbitals vs. Filtering Error.
The absolute difference between the analytical density and the stochastic density obtained from analytical filtering (column 2) and Chebyshev filtering (column 3) as a function of the number of stochastic orbitals N_{sto} used. The fact that the absolute difference decreases as N_{sto} increases demonstrates that the accuracy of the stochastic density increases as more stochastic orbitals are used.

In regard to the efficiency of these filtering methods, the time required to compute the stochastic densities was roughly two orders of magnitude less for Chebyshev filtering than analytical filtering. Even when using 20,000 stochastic orbitals in Chebyshev filtering, the computation time was less than when using 100 stochastic orbitals in analytical filtering. This finding is in accordance with the fact that analytical filtering first requires matrix diagonalization to obtain the system's occupied eigenvectors. This efficiency is advantageous when trying to extract data from more realistic systems with more than one electron, such as molecules. By observing the density value that the Chebyshev approach converges to for such systems, the results of the analytical approach can be predicted.

These findings demonstrate the computational power of sDFT. By circumventing matrix diagonalization, computation run time shortens while maintaining a high degree of accuracy. This provides a potential solution to the problem discussed earlier regarding the unreasonable timescales of computer resources impeding simulations needed for research studies. Future work may entail changing other parameters such as the number of Chebyshev polynomials used to approximate the error function. Additionally, work can be done to identify the exact number of stochastic orbitals needed such that the error value and efficiency are both optimized Additionally, other forms of the Hamiltonian such as the harmonic oscillator and rigid rotor models may be considered in addition to the tight binding model that was used in this research project.

N_{sto}	Analytical $t_{end} - t_{start}$	Chebyshev $t_{end} - t_{start}$
100	4.01E-3	2.31E-5
500	4.52E-3	2.60E-5
1000	5.04E-3	3.91E-5
2000	6.10E-3	5.01E-5
5000	8.01E-3	6.52E-5
10000	1.60E-2	1.83E-4
20000	3.22E-2	3.67E-4

Table 2: Stochastic Orbitals vs. Filtering Efficiency.
The computation time to calculate the stochastic density obtained from analytical filtering (column 2) and Chebyshev filtering (column 3), as a function of the number of stochastic orbitals N_{sto} used. The computation time required for Chebyshev filtering is lower than that of analytical filtering, indicating a higher level of efficiency. Additionally, the computation time for both filtering methods decreased as the number of stochastic orbitals used increased, indicating that greater stochastic samplings lead to lower efficiency.

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. Daniel Neuhauser, for his instructional guidance and knowledge in overseeing this research project. I am also grateful to my graduate student mentor Minh Nguyen for his invaluable patience and feedback.

Lastly, this endeavor would not have been made possible without the generous support of the University of California Leadership Excellence through Advanced Degrees program, who was responsible for funding my research.

REFERENCES

1. F. A. Gonźalez, A. Gallego, S. Toledo-Cortés, V. Vargas-Calderón, Learning with density matrices and random features. *Quantum Machine Intelligence*. **4**, pp.2 (2022). doi:10.1007/s42484-022-00079-9.
2. J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley Publishing Company, Boston, ed. 1, 1967).
3. Langione, Matt, et al. "Where Will Quantum Computers Create Value-and When?" *BCG Global*, BCG Global, 15 Dec. 2021, https://www.bcg.com/publications/2019/quantum-computers-create-value-when.
4. K. Burke, *The ABC of DFT* (Department of Chemistry, University of California, Irvine, CA, 2007), pp. 15–19.
5. T. van Mourik, M. Bühl, M.-P. Gaigeot, Density functional theory across chemistry, physics and biology. *Philos T Roy Soc A*, **372**, pp. 1-3 (2014). doi:10.1098/rsta.2012.0488.
6. M. D. Fabian *et al.*, Stochastic density functional theory. *WIREs Computational Molecular Science*. **9**, pp.2 (2018). doi:10.1002/wcms.1412.
7. M. Chen, R. Baer, D. Neuhauser, E. Rabani, Stochastic density functional theory: Real- and energy-space fragmentation for noise reduction. *J Chem Phys*. **154**, 204108 (2021). doi: 10.1063/5.0044163
8. J. C. Slater, G. F. Koster, Simplified LCAO Method for the Periodic Potential Problem. *Phys Rev*. **94**, 1498–1524 (1954). doi:10.1103/PhysRev.94.1498
9. J. Mason, D. C. Handscomb, *Chebyshev Polynomials* (Chapman and Hall/CRC Press, London, 2002).
10. A. Szabo, N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, Inc., Mineola, NY, 1996).

Recommendation Algorithms to Increase Network Fairness

Naisha Agarwal[1]

[1]Department of Computer Science, University of California, Los Angeles.

ABSTRACT

This paper introduces a novel problem of improving fairness in networks as defined by access to influencers. Most real-world networks, especially within social media, have a small fraction of highly-influential nodes. Access to these influential nodes leads to better future opportunities. However, recommendation systems used in social media often maximize the total number of edges to increase engagement and advertising revenue. This creates bias and often leads to inequitable access to influencers among the nodes, thereby creating unequal access to future opportunities. This paper provides a principled formulation of this problem using concepts in graph algorithms. In particular, we first introduced a novel scale-free measure to quantify this fairness in networks. A novel recommendation algorithm is then introduced that can seamlessly work with existing recommendation algorithms to increase this fairness in networks. The algorithm introduced in this paper is inspired by the seminal work of Watts-Strogatz in the context of decentralized search. As such, the approach is to introduce random edges in a controlled fashion to create more weak ties in a network and reduce the overall distance of nodes from the influencer set. Through extensive simulations on two real-world network data sets and comparing seven different algorithms, the fairness of the graph is shown to increase monotonically as the amount of controlled randomness introduced increases. This current approach assumes the graph is connected; future pursuits include generalization of the proposed method to disconnected graphs. This work provides a foundation to improve machine learning algorithms used in networks, particularly social media, using foundational concepts in graph theory.

INTRODUCTION

Networks are ubiquitous and significantly influence daily lives. Oftentimes these networks have a highly influential small set of nodes, and having access to these nodes can be quite advantageous. For instance, on a social media site like LinkedIn, access to decision makers at large corporations can increase the likelihood of obtaining jobs, sales, and other opportunities. Similarly, on Twitter and Facebook, access to influencers can help businesses market messages more easily and extensively (*1*).

To ensure a level playing field for all network nodes, nodes in a graph must have equal access to influencers. However, in many real world applications, new edges (connections) are formed via recommendation systems that are based on network attributes like triangle closings (the number of mutual friends) and node characteristics like demographics, industry, school, or interests (*2*). While these systems are great at optimizing the number of edges that are formed in the network over time, they create unintentional biases in the system where some nodes have an unfair advantage of gaining more access to opportunities and utilities through influencers. The algorithms that power these recommendation systems further exacerbate the situation due to the attributes they rely on.

The main objective of this paper is to show that it is possible to modify machine learning algorithms used in social media to balance two competing objectives: profit for the social media companies and fair access to influencers across all nodes. Under these modified algorithms, people will not be penalized based on who they know and will instead be rewarded based on what they know. This will result in a more equitable society despite the increasing influence of social media and digitization.

This paper's technical approach is inspired by the seminal work of Watts and Strogatz (W-S). In their paper, W-S demonstrated that introducing randomness can lead to more connections in a social network and reduce the diameter of the network (*3*). While W-S inspired the idea of adding randomness in a controlled fashion, this paper's approach applied this idea to node recommendation algorithms to provide more equitable access to influencers. A systematic study of introducing this randomness into node recommendation algorithms in the context of networks to increase fairness has not been studied before.

There have been several solutions proposed to measure fairness in networks, such as the Atkinson Index and using various other types of algorithms in networks (*4*,*5*). However, these algorithms provide variance in utility associated with a node across the entire network. In this paper, a scale-free fairness measure is calculated using the influencer set and is improved using novel node recommendation algorithms.

The efficacy of the approach is proved through extensive simulations on two real-world datasets comparing seven different

algorithms. It was shown that adding controlled randomness results in fairness increasing monotonically with increasing randomness. This paper will motivate further research that can leverage well established mathematical models in networks and graphs to improve machine learning algorithms used for graphs in social media.

METHODS

Graphs used in this paper represent social networks. A given graph is assumed to have nodes with the same attributes (ex. skill level, age, gender, location) and edges that represent the connections among those nodes. In other words, the likelihood of any pair of nodes connecting is exactly the same in terms of the attribute properties of the different nodes. The difference in the likelihood of any pair of nodes connecting comes from graph structure and the differences in the number of mutual edges. The graph is also assumed to be connected, hence a path always exists. Disconnected graphs are discussed in future work.

2.1 - Finding an appropriate measure for fairness

There are three main components to finding an appropriate measure for fairness:

1) finding the influencers in the graph using graph theoretic measures;
2) finding the fairness measure of the overall graph; and
3) normalizing the graph fairness measure

These methods are summarized and explained in further detail below.

2.1.1 - Finding influencers

When measuring the influence of a node on the network, betweenness centrality was used as the measure of centrality. Betweenness centrality measures the proportion of times a node acts as a bridge along the shortest path between two other nodes (*6*). The equation for betweenness centrality for a node v is

$$c(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma st(v)}{\sigma st} \quad (1)$$

σ_{st} is the total number of shortest paths from a source node, s, to a target node, t, and $\sigma_{st}(v)$ is the number of σ_{st} that pass through v (*7*).

Influencers were found by calculating the betweenness centrality of every node in the graph and taking the top k percent of nodes as the influencers.

Algorithm 1 Determining Influencers

Require: G(graph), k(top percent of nodes as per betweeness centrality that will be considered as influencers)

Ensure: array of the top k percent of nodes that are considered influencers. Each element has 2 components: the node number and its corresponding betweeness centrality.

2.1.2 - Finding the fairness measure

Fairness is defined to be the minimum shortest path to the influencer set of an existing path. To compute the fairness of the entire network, the fairness of a single node is first found, and then an average is taken of each node (described in further detail below) to compute the fairness of the entire network. Using Dijkstra's algorithm (*8*), the fairness of a node can be measured using

$$\text{fairness}(v) = \min\{s \text{ in } S\} \text{ dist}(v, s) \quad (2)$$

where S is the influencer set, s is the source node in S, and v is a node in the graph. If no path exists from a node to the influencer set, individual node fairness is not defined. In this paper, it was assumed the graph is connected, hence there is always a path.

Algorithm 2 Calculating Fairness of a Node

Require: G(graph), inf(influencer array), n(node for fairness required)
Ensure: fairness value of the node
min is initialized to a large number;
while end of inf is not reached **do**
  if there is path between n and element in inf **then**
    get shortest path distance d between n and element in inf in graph G
    if d is less than min **then**
      min = d
    else
      return false
    end if
  end if
  go to next element in inf
end while

2.1.2.2 - Getting fairness of total graph

The fairness of the graph is defined to be the average of the top *t percent* of node fairness values. This can be modeled by the equation:

$$\text{fairness}(G) = \frac{1}{\text{top t\% } \{v \text{ in V}\} \text{ fairness}(v)} \quad (3)$$

where G is the graph in question, v is a node in the graph, and V is the set of vertices in the graph not including influencers. As the denominator becomes smaller, more nodes have better access to influencers and the fairness of the graph increases.

Algorithm 3 Calculating Fairness of the Graph

Require: G(graph), inf(influencer array), t(percent of nodes sorted in descending order of node fairness that should be used to calculate the fairness of the graph)
Ensure: single integer value indicating the total fairness of the graph
min is initialized to a large number;
while all nodes in graph G have not been visited **do**
  get fairness of node using fairness measure function
  add to final array arr of all node fairnesses
end while
sort array in descending order
calculate top number of nodes that should be used to calculate total fairness using top percent
return reciprocal of average of node fairnesses of t percent of nodes

2.1.3 - Calculating the fairness index

Since graphs have different numbers of nodes and edges, it is important to normalize the fairness values to ensure it is scale-free and easily comparable across different graphs. To normalize the fairness measure,the fairness of a random graph with the same number of nodes and edges as the observed graph was computed. A random graph was used to obtain a scale free fairness index that makes it possible to compare fairness after adjusting for the varying number of edges and nodes in the graph. The fairness index is defined as:

$$\text{fairnessindex} = \frac{\text{fairnesss(G)}}{\text{fairness(random graph)}} \quad (4)$$

The Erdos-Renyi graph was used to produce random graphs. The Erdos-Renyi graph randomly connects nodes with every edge included in the graph with selected probability p.

The probability for generating a graph with n nodes and m edges is modeled by

$$p^m(1-p)^{\binom{n}{2}-m} \tag{5}$$

where the likelihood of adding more edges to the given graph increases as probability p increases (*9*).

The Barabasi-Albert was another method used to produce random graphs. This graph begins with an initial connected network of n nodes, where new nodes are added to existing nodes with probability p_i. p_i is the probability that a new node is connected to node i. This is modeled through the equation

$$p_i = \frac{k_i}{\sum_j k_j} \tag{6}$$

where k_i is the degree of node i and the baseline sum is made over all pre-existing nodes j (*10*).

When generating Erdos-Renyi and Barabasi-Albert graphs, different numbers of random graphs (e.g. 5, 10, and 100) were tested and fairness values were averaged for each graph. The fairness values typically converged and stabilized around 10 random graphs and above. There was an average of 10 random graphs when calculating the fairness index.

Algorithm 4 Calculating the Fairness Index of the Graph

Require: G(graph), n(total number of nodes in graph), m(total number of edges), k(top percent of nodes that will be selected as influencers), s (random seed), t (top percentage of nodes that are considered for fairness calculation
Ensure: single integer value representing the fairness index
  p = (2 * m)/(n* (n-1))
  create Erdos-Renyi/Barabasi-Albert graph l with parameters n,p,s
  find influencer array with parameters l, k
  find fairness of graph l with parameters l,inf, t using total graph fairness
  fairness index = fairness of graph G/fairness of graph l
  return fairness index

2.2 - Recommendation algorithms to increase the fairness index in a network

The main idea of the proposed recommendation algorithm was motivated by a similar idea in the Watts-Strogatz model in the context of the small-world experiment and decentralized search (*3*). The high-level mathematical model behind this approach states for a given node n, with some probability P, the proposed algorithm recommends a node based on the number of mutual friends between n and the candidate node. The larger the number of mutual friends, the more likely it is to recommend the node. With probability $Q = (1 - P)$, the algorithm adds some randomness and creates a weak tie by forming a new edge with a randomly selected target node. This experiment introduced a modification where the weak tie is selected via an importance sampling algorithm that is proportional to some function based on degree and distance to the influencer set of the candidate node. This algorithm is illustrated below in pseudocode:

2.2.1 - Details

For a given node n, consider the array $M[n] = \{(m, i(m), w(i(m))\}$. Here, m denotes the target candidate nodes that n is not yet connected to. For each node m, $i(m)$ denotes the number of mutual friends (edges) between n and m, and $w(i(m))$ denotes the weight of m. The influencer set is not included in $M[n]$ to avoid trivial solutions.

Algorithm 5 Recommendation Algorithm

Require: N(number of visits), a, b, P(probability of recommending using mutual friends), inf (influencer array), G(graph)
Ensure: Recommends users proportional to degree and distance to the influencer set of the candidate node
  gnew = G, nvisits = 0
  while nvisits is less than N **do**
    select random node n
    generate M(n)
    generate R(n)
    if probability P **then**
      sample an edge from M
    else probability Q
      sample an edge from R
    end if
    gnew = gnew U new edge
    nvisits++
  end while

Algorithm 6 generate M(n)

Require: i is the number of mutual friends
  get nodes that n is not connected to (excluding influencers)
  compute mutual friends with n for each node
compute weights for each node using e^i

Algorithm 7 generate R(n)

Require: i is the number of mutual friends
  get nodes that n is not connected to (excluding influencers)
  compute degree and the minimum distance to an influencer of each node
sample with weights equal to D^a/d^b

2.2.2 - Calculate w(i(m)) values

In order to give people with more mutual friends a higher probability of connecting, a weight function for each node was calculated. The weight $w(i(m))$ should be an increasing function of $i(m)$ because the more mutual friends there are between n and m, the more likely they are to connect. Therefore, they should be recommended with a higher probability. One popular w function used is the sigmoid function $q(i) = 1/(1 + e^{a-i})$ where a is a constant. Since $q(0) = 1/(1 + e^a)$ is a constant, the weight function is

$$w(i) = \frac{q(i)}{q(0)} = \frac{(e^{-a} + 1)}{(e^{-a} + e^{-i})} \approx e^i \tag{7}$$

The approximation was made assuming a was very large due to the probability of two nodes connecting with zero mutual friends being very small. Hence, $w(i) = e^i$.

Consider another array for a given node n as $R[n] = \{m, D(m), d(m), wr(a, b)\}$. Here, $D(m)$ is the degree of m, $d(m)$ is the minimum distance of m to the influencer set, $wr(a,b)$ is the weight function written as $wr(a,b) = D(m)^a/d(m)^b$, and a and b are non-negative constants. Recall m refers to the target candidate nodes that n is not yet connected to. The reasoning behind choosing this weight function is described later below.

2.2.3 - Algorithm details

For probability P, the algorithm recommends a node for a given visit node n by sampling a node from M with weights proportional to $w(i)$'s. Probability $Q = 1 - P$ recommends a node by sampling from R with weights proportional to $wr(a,b)$. Note that $wr(0,0) = 1$. Just like in the Watts-Strogatz model, for probability P, the new algorithm recommends a node that is close to n's neighborhood, while for probability Q, the algorithm selects a node at random (assuming $a = b = 0$). These random nodes shorten the path to the influencer set and create more fairness.

For simplicity and efficiency in the simulations, the algorithm always connects to the recommended node. In practice, this does

not happen and recommendations have different connection probabilities that are not constant. For instance, the connection probability to a node with zero mutual friends is typically lower than connecting to a node with ten mutual friends. Since the goal of this paper is only to study the impact of adding diversity to fairness, this assumption can be made without loss of generality. In the end, what matters is the budget in terms of adding nodes at random and the fairness index it produces.

Similarly to the Watts-Strogatz model, diversity can be introduced in a controlled manner. Two intuitions were used. First, sampling higher degree nodes with higher weights is helpful since they are likely to create more paths to the influencer set. Additionally, if two nodes are given that have the same high degree, the one that is closer to the influencer set is more likely to increase fairness.

However, it is not clear what is better: a very high degree node far away from the influencer set (e.g., a node with a degree of 100 that is a distance of 3 from an influencer) or a less high degree node closer to the influencer set (e.g., a node with a degree of 50 that is a distance 1 from an influencer). Simulations tested multiple a and b values, which will be described in further detail in the experiments section (Table 1).

To get more intuition into the algorithm, the distance between $w(i)$ (the mutual friends algorithm) and $wr(a,b)$ was computed using the Kullback-Leibler distance function. The distance is given by

$$\mathrm{KL}(w, \mathrm{wr}(a = 0, b = 0)) = A\sum_{k=1}^{K} f(k)e^{k} + C \tag{8}$$

where k denotes the number of mutual friends, K is the maximum value of mutual friends there are in the data, $f(k)$ denotes the frequency of target nodes that have k mutual friends with the source node, and A and C are constants. We assume $a=b=0$ for comparing the distance with a completely random distribution as the baseline. As can be seen above, the distance from the random distribution increases when the source node has a higher concentration of target nodes with a large number of mutual friends. These are the scenarios where the randomization is breaking the "rich gets richer" characteristics of the usual recommendation algorithm and creating more fairness.

2.2.4 - Deriving the Kullback-Leibler equation

For a given source node n, let m subscript be the target node the source node can connect to and let P and Q be the probability distributions for weights w and $wr(a=b=0)$, respectively. Therefore, $P = Aw$ and $Q=B(wr)$, where A and B are normalizing constants.

a value	b value	representation
0	0	completely random
1	0	proportional to degree
1	1	proportional to D/d
1/2	1/2	proportional to $\sqrt{D/d}$
0	2	proportional to $1/d^2$
2	2	proportional to $(D/d)^2$
1	2	proportional to D/d^2

Table 1: Comparing different values of a and b being tested.
This table details the different values of a and b that are being tested in our overall D^a/d^b recommendation algorithm, and that are graphed in Figure 3.

$$\mathrm{KL}(P, Q) = \sum_{m} P(m)\log\left(\frac{P(m)}{Q(m)}\right) \tag{9}$$

Since $w(m)=e^{i(m)}$ and $wr(a=b=0) = 1$, where $i(m)$ is the number of mutual friends between n and target node m, the expression simplifies to:

$$\mathrm{KL}(P, Q) = A\sum_{m} i(m)e^{i(m)} + C, \text{ where C is a constant} \tag{10}$$

Let k be the number of mutual friends and $f(k)$ denote the frequency of target nodes with k mutual friends with the source node n. Doing a frequency tabulation of $i(m)e^{i(m)}$over all target nodes, the expression becomes

$$\mathrm{KL}(P, Q) = A\sum_{k=1} f(k)ke^{k} + C \tag{11}$$

2.3 - Experiments

Simulations were conducted on multiple datasets to show the efficacy of the proposed methods: a Facebook social circles graph and a LastFM graph (Figure 1). Both data sets had different network structures and provided complementary insights to the methodology in this paper. For each data set, the D^a/d^b algorithm was run for seven pairs of a and b (Table 1).

The Barabasi-Albert random graph was used to calculate the fairness index of nodes (Figure 2).

Additionally, the influencers of the LastFM graph were manually changed to show the robustness of this algorithm (Figure 3). The new influencers had very low betweenness centralities, representative of situations where influencers who don't have a big social media presence are still influential in the real world.

RESULTS

3.1 - Random graph

It was found that the Barabasi-Albert random graph converged faster than the Erdos-Renyi graph, and provided a better normalizing value for calculating the fairness index (Figure 3). The fairness index of nodes were calculated using the Barabasi-Albert graph.

3.2 - Facebook data

It was observed that degree (a =1, b = 0) was the best algorithm for the Facebook graph as it consistently produced the highest data points. D/d^2 and $(D/d)^2$ were close seconds (Figure 3).

Figure 1: Visualization of Facebook and LastFM networks.
Nodes and edges of the graph were visualized using the NetworkX draw feature, which draws the graph using Matplotlib. Nodes (dots) on the graph represent a person in a social media network, and edges (lines) represent a connection in the network. The graph on the left representing the Facebook social network shows several dense communities loosely connected by weak ties. The graph on the right representing the LastFM network shows one central community with some nodes on the periphery.

Each of the algorithms had the same value for $Q=0$ due to the locations' algorithms being based on mutual friends and not randomness. The fairness index increased when there was a higher chance of randomness in the network, indicated by the decreased P and increased Q.

3.3 - LASTFM data

It was observed that $(D/d)^2$ resulted in the highest fairness values for the LastFM data (Figure 3).

3.4 - Studying the impact of changing influencers on the fairness index

The algorithm was robust to a change in influencers by showing that $(D/d)^2$ resulted in the highest fairness values. (Figure 3).

DISCUSSION

As the amount of randomness, Q, in the network increased, the overall fairness in the network in terms of access to influencers increased. This is the most critical and basic observation, and it shows the efficacy of the approach. Based on this experiment, it appeared that introducing randomness as some polynomial function of D and d makes a difference relative to complete randomness. $(D/d)^2$ seemed to be the best recommendation algorithm to increase fairness in the network; although it is specific to the data-set at hand. This result is significant and important in order to find algorithms that can achieve high fairness with lower randomization. More randomization would lead to more loss in number of edges and hence engagement and revenue. Additionally, the algorithm found is robust to small changes in the quantity and set of influencers. The overall increase in fairness with increasing randomness deteriorated but the monotone increasing pattern remained intact. This is again significant because the definition of influencers in the real-world could be defined in various ways. For instance, it could include a famous personality who is known to be an influencer but does not necessarily engage in a social network, and hence, they could have low betweenness centrality. This result shows networks can add such nodes in the influencer set based on relevant domain knowledge and still obtain fairness using the algorithm proposed.

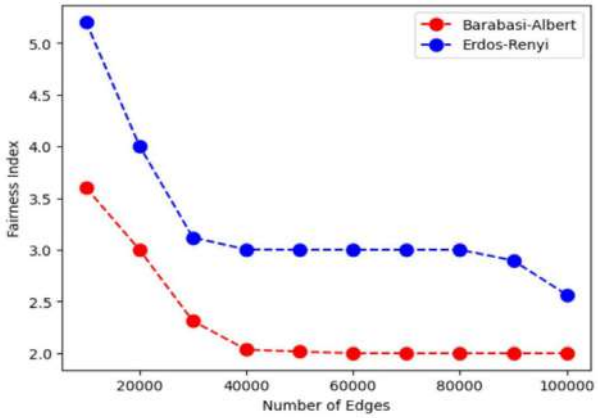


Figure 2: Fairness convergence for random graphs for different amounts of edges.
This graph compares the Erdos-Renyi and Barabasi-Albert random graphs. The x-axis indicates the different number of edges and the y-axis represents the fairness index of these graphs. The random graphs' fairness indices begin to converge to a singular value as the number of edges increases. The Barabasi-Albert random graph converges faster than the Erdos-Renyi graph; while the Erdos-Renyi graph starts to converge at 50,000 edges and then changes values at 100,000 edges, the Barabasi-Albert graph begins to converge at 40,000 edges and stays at the same fairness value through the 100,000 edges. Therefore, the Barabasi-Albert random graph serves as a better normalizing value for calculating the fairness index.

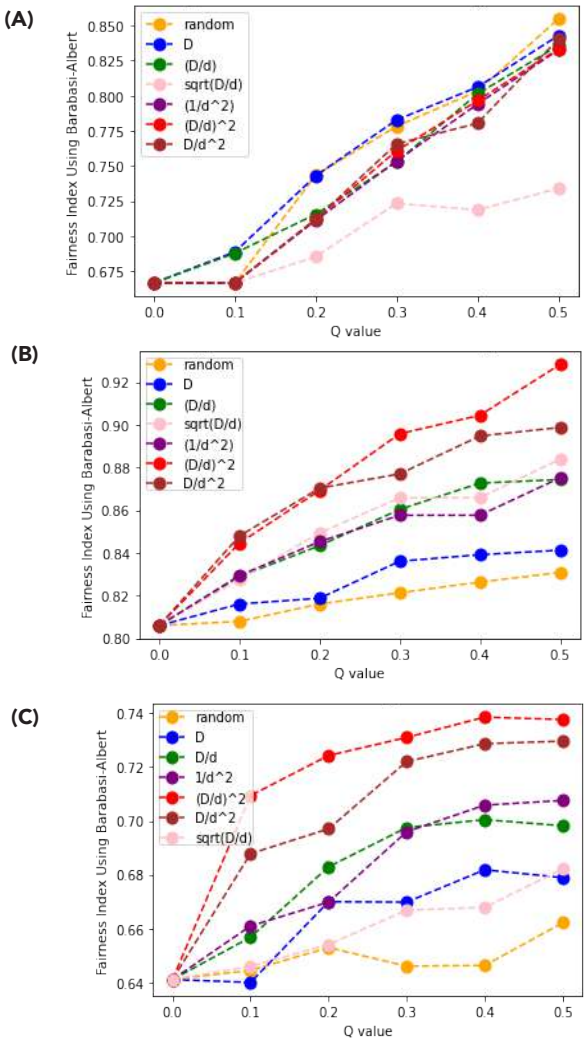


Figure 3: Changes in fairness with different recommendation algorithms for different social networks
The x-axis in graphs **(A)**, **(B)**, and **(C)** is the probability Q that the recommendation algorithm recommends a node based on randomness. The y-axis is the fairness index, where a higher value indicates more fairness in the graph. Each graph compares numerous recommendation algorithms of the form D^a/d^b, with various values of a and b plotted. Graph **(A)** compares recommendation algorithms in the Facebook social network. The blue line with the degree algorithm (a = 1, b = 0) consistently appears to be the highest line of the graph, indicating that it achieves the greatest degree of fairness. Graph **(B)** compares recommendation algorithms in the LastFM social network. The red line, or the $(D/d)^2$ algorithm (a = 2, b = 2), consistently earns the highest fairness index, indicating it produces the greatest fairness. Graph **(C)** also compares different recommendation algorithms for the LastFM social network, except with a different choice in influencers (low betweenness centralities). It appears that $(D/d)^2$ still produces the highest fairness indexes, proving the robustness of the algorithms with a change in influencers.

CONCLUSION

A novel problem of measuring network fairness in terms of access to influencers was introduced in this paper. Novel methods to measure this fairness and increase this fairness with recommendation algorithms were proposed. The efficacy of the approach was tested through extensive simulations on two real world data-sets comparing several variations of the algorithm. First, experiments conducted proved the efficacy of the proposed approach: fairness tends to increase monotonically with increasing randomness. This is important as increased randomness will also monotonically increase the number of edges lost compared to the canonical recommendation algorithm. Hence, the outlined approach provides a clear mechanism to incorporate fairness in social media through a simple randomization mechanism that can be easily incorporated into existing systems. This study also showed that introducing randomness via importance sampling through a polynomial function of D and d can yield better fairness with lower randomness. Other functions based on node attributes can improve this even further. Machine learning applications can be explored in future work. Finally, this study showed that the monotone relationship is intact with changes in the influencer set. This is a critical finding, as influencers may not be necessarily determined solely based on the graph theoretic measure we introduced. Some domain knowledge may also be important. For instance, a famous personality who is not very active on social media could still be considered as an influencer.

Previous work studied bias and fairness in supervised machine learning algorithms that address fairness by modifying the training data and the supervised algorithms (*5*). While this work is similar in motivation, it is highly specific to networks. Additionally, there is very recent work being done on dyadic preferences in a graph, where the authors of the paper address potential bias in user recommendation (*11*). However, their work is specific to ensuring demographic parity with respect to node attributes in the network. This work addresses a more specific problem of ensuring fairness with respect to access to influencers. While it is more specific, it does require new methods to measure fairness and modify the recommendation algorithms. The area of studying fairness in networks is relatively new and hence the literature in this area is sparse.

The methods discussed only apply to connected graphs as the fairness measure would not be defined if the graph is disconnected. Future investigations include studying generalizations to disconnected graphs. Additionally, since the simulations in the experiment were based on small data-sets, it encourages future studies of behavior with very large graphs. This will require large scale distributed computing with graphs that we plan to explore. The activeness of a user can also be taken into account when running simulations. Currently, it is assumed all nodes are equally likely to visit the network in the simulations. It is also assumed that the number of nodes in a graph remains constant. In reality, the number of nodes visiting a network depends on how active a user is, and nodes in a graph are not constant due to users constantly joining and leaving.

ACKNOWLEDGEMENTS

We would like to thank MIT Professor Julian Shun for his guidance throughout this research.

REFERENCES

1. Thorpe, Jessica. How Influencers Can Help Your Business in Times of High Inflation. *Fast Company.* (2022).
2. How We Analyzed Twitter Social Media Networks. *PewResearchCenter*, Social Media Research Foundation. (2014).
3. D. Watts, S. Strogatz, Collective dynamics of 'small-world' networks. *Nature*. **393**, 440–442 (1998). doi: 10.1038.30918
4. G. Saint-Jacques, A. Sepehri, , I. Perisic, N. Li, Fairness through Experimentation: Inequality in A/B Testing as an approach to responsible design. *arXiv*. (2020). doi: 2002.05819
5. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A., A Survey on Bias and Fairness in Machine Learning. *arXiv*: 1908.09635 (2019, September 17).
6. Brandes, U, A Faster Algorithm for Betweenness Centrality. *J Math Sociol*, **25(2)**, 163-177 (2001).
7. Freeman, L, A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40(1)**, 35-41 (1977). doi: 10.2307/3033543.
8. Dijkstra, E. W, "A Note on Two Problems in Connexion with Graphs". *Numer Math*. **1(1)**, 269–271 (1959). doi: 10.1007/BF01386390
9. Erdős, P.; Rényi, A, "On Random Graphs. I". *Publ Math-Debrecen*. 6: 290–297 (1959).
10. Barabási, A.-L., Albert, R, Emergence of scaling in random networks. *Science*, *286*(5439), 509-512 (1999). doi: 10.1126/science.286.5439.509
11. Li, Pezhou., Wang, Yifei., Zhao, Han., et.al. On Dyadic Fairness: Exploring and Mitigating Bias in Graph Connections. *International Conference on Learning Representations*. (2021).

The Effect of Guiding-Center Approximations on Energetic Particle Transport in Tokamaks

Nathaniel Chen[1], Garrett Wright[2], Phillip Bonofiglo[2], Mario Podesta[2]

[1]Department of Physics and Astronomy, University of California, Los Angeles. [2]Princeton Plasma Physics Laboratory.

ABSTRACT

Tokamaks are toroidal plasma confinement devices considered to be among the best candidates for fusion energy. Plasma modeling can inform tokamak designs and sensor placement. A key assumption often made in order to improve computational speed is to use the guiding center approach, where gyromotion is ignored. However, gyromotion may become non-negligible when particles are near the outboard walls of a tokamak or when particles have high kinetic energy. To better understand the implications of particle gyromotion in these conditions, both the guiding center and full orbit effects of energetic particles in the National Spherical Torus Experiment – Upgrade are simulated with ORBIT-GPU, a particle pushing code specialized in quickly solving plasma transport inside of a tokamak. Results show that the approximate particle losses given by this approximation are significant, and that future burning plasma transport simulations may need to account for gyromotion in order to give better results when designing modern tokamaks.

INTRODUCTION

Nuclear fusion is often perceived after as the holy grail of clean energy — a process that uses very common materials found on Earth to generate an enormous amount of power with negligible waste and carbon emission. Fusion occurs most often in plasmas, a state of matter in which atoms are ionized through a combination of high temperatures and low number densities. Plasmas can be thought of as a large ensemble of electromagnetic particles. In this sense, Maxwell's equations, which describe the electromagnetic interactions of fields and particles, and the fluid equations, which describe the collective motion of particles, can both provide a strong description of the ongoing state of plasma (*5*). In certain cases, such as with hydrogen isotopes like deuterium and/or tritium, fusion can produce much more energy than the amount put in. Since these particles can exceed 100 million kelvins, they must be thermally isolated from the surrounding environment to maintain an ongoing reaction known as burning plasma (*13*).

One technique used to enable burning plasma conditions is magnetic confinement fusion (MCF), which contains plasma inside a magnetic chamber. While there are many ways to magnetically confine plasma, a well-established MCF candidate is the tokamak. Tokamaks, a transliteration of the Russian "toroidal chamber with axial magnetic field," are doughnut-shaped devices that confine plasmas within a vacuum chamber using three main sets of magnets: a central solenoid which keeps plasma flowing around the tokamak, and a toroidal and poloidal field coil system which shapes the plasma to keep it from touching the outboard walls (Figure 1) (*4*). When modeling the process that occurs inside a tokamak, Maxwell's equations and the fluid equations can be combined with two-dimensional (2D) axial symmetry to form the Grad-Shafranov equation:

$$\$\Delta^*\psi = -\mu_0 R^2 \frac{\mathrm{d}p}{\mathrm{d}\psi} - F\frac{\mathrm{d}F}{\mathrm{d}\psi}\$ \quad (1)$$

where the term $F = RB_\phi$ is a surface quantity related to the net poloidal current in the plasma, $R\$, \$\phi\$, \$Z$ are the cylindrical coordinates (Z is not explicitly shown in the equation), and ψ is the stream function related to the flux, or flow of the plasma. Δ^* represents an elliptic operator, a kind of generalization of the Laplace operator (*5*).

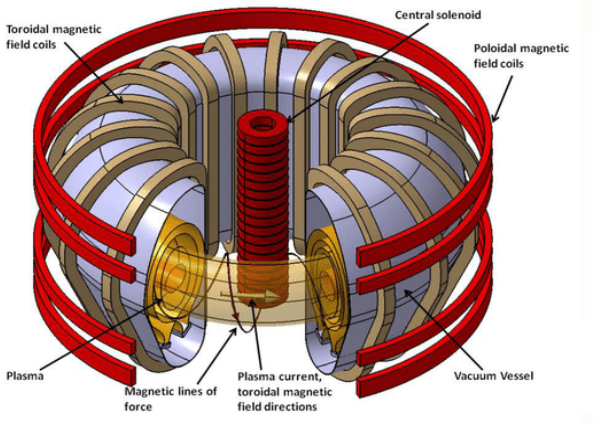


Figure 1: Schematic of a tokamak fusion reactor.
Plasma (in orange) is contained within the torus. Each plasma layer within the tokamak is known as a magnetic flux surface. The last closed flux surface (LCSF) is the layer closest to the walls of the tokamak. This figure was taken from Figure 1 of reference 4 and used under CC BY 4.0 (*4*).

However, there are fundamental challenges to tokamaks, such as possible plasma instabilities and islands that trap plasma (*10*). To better understand these effects, plasmas can be numerically solved. This is difficult given that plasmas operate on a wide range of timescales – simply using the smallest timescale evolution would lead to long computation times and large deviations over time due to numerical errors. Hence, the challenge for these types of simulations is to provide a computationally inexpensive description of plasma behavior without compromising accuracy (*1*). One way to do this is to calculate the plasma in multiple steps: first getting the equilibrium/stability profile, which describes the slow evolution of fields in a plasma, and then calculating the plasma transport, which describes the fast evolution and distribution of particles in a plasma (*2*).

Traditionally, guiding center (GC) transport solvers specialize in quickly calculating the motion of plasma by ignoring particle gyromotion. Gyromotion can be described as the fast precession of a particle around a magnetic field line, characterized by a Larmor radius and cyclotron frequency. Cyclotron frequency is often large while the Larmor radius is small (*9*). In order to save computation time, gyromotion can often be ignored and the particle can be assumed to be at the GC as opposed to at the full orbit (FO) position, which includes the particle's gyromotion path (Figure 2). A possible issue with this assumption in simulation may come from energetic particles (EPs). As burning plasma fusion conditions are becoming more of a reality due to advances in magnets and infrastructure, EPs may begin to play a considerable role in correctly characterizing fusion. Compared to thermal particles, which are on the order of 1-20 keV, EPs are suprathermal, meaning they can vary from 100 keV to the MeV range. An example can be found in deuterium tritium (DT) reactions, which can produce energetic 3.5 MeV alpha particles. Although EPs make up roughly 1-10% of plasma, they provide the bulk of plasma heating and are responsible for maintaining fusion conditions in plasma. In this sense, they exert a large influence over plasma properties such as temperature, density profiles, turbulence, and equilibrium (*6*).

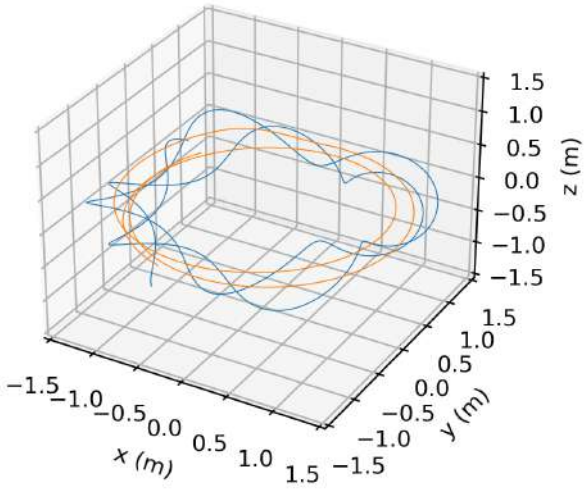


Figure 2: Simulated particle trajectory within a tokamak. Guiding center (orange) is the direct and simplified computation of the particle, which ignores gyromotion. Full orbit (blue) is the full motion of the particle - the sum of the phase angle and Larmor radius to the guiding center.

However, since EPs are extremely hot, they may exhibit different behavior than other plasmas. For example, their Larmor radius may be very large and thus non-negligible.

MATERIALS AND METHODS

Orbit-GPU and Supplementary Codes

ORBIT-GPU is a CUDA-compatible graphics processing unit (GPU) specialized in solving Hamiltonian EP motion with GC principals. The equations it uses were originally derived by White and Chance and directly provide the solution to particle coordinates in a tokamak (Equation A) (*12*).

Equation A

$$\dot{\theta} = \frac{\rho_\parallel B^2}{D}(1-\rho_\parallel g') + \frac{g}{D}\left[(\mu+\rho_\parallel^2 B)\frac{\partial B}{\partial \psi_p} + \frac{\partial \Phi}{\partial \psi_p}\right]$$

$$\dot{\psi}_p = -\frac{g}{D}\left[(\mu+\rho_\parallel^2 B)\frac{\partial B}{\partial \theta} + \frac{\partial \Phi}{\partial \theta}\right]$$

$$\dot{\rho}_\parallel = -\frac{(1-\rho_\parallel g')}{D}\left[(\mu+\rho_\parallel^2 B)\frac{\partial B}{\partial \theta} + \frac{\partial \Phi}{\partial \theta}\right]$$

$$\dot{\zeta} = \frac{\rho_\parallel B^2}{D}(q+\rho_\parallel I'_{\psi_p}) - \frac{I}{D}\left[(\mu+\rho_\parallel^2 B)\frac{\partial B}{\partial \psi_p} + \frac{\partial \Phi}{\partial \psi_p}\right]$$

The coordinates (ψ_p, θ, ζ) are for a given helical symmetric system where ψ_p is the poloidal flux, $\rho_\parallel$ is a normalized gyroradius parallel to the magnetic field, Φ is the electric potential, g is force (F), and μ is the normalized magnetic moment. This code approximates particles as their GC and solves for EP transport in response to a static, perturbed magnetic field. It is good for analyzing many types of mode-induced particle loss and particle transport. ORBIT-GPU can accurately approximate the movement of each EP independent of each other, essentially rendering it a Monte Carlo simulator.

ORBIT-GPU encodes the resonant wave-particle interactions via the fast ion constants of motion with the ORBIT-kick variation (*8*). A kick-model describes EP diffusion in response to a perturbed magnetic field (*7*). The fast ion constants of motion include energy (E), canonical toroidal momentum (P_ζ), and magnetic moment (μ). ORBIT assumes conservation of magnetic moment for low frequency activity ($\omega << \omega_{ci}$ where ω_{ci} is the ion cyclotron frequency) and the GC approximation. FO effects are included by simply adding a particle's gyrophase to its GC position and tracking the gyrophase.

The equilibrium files used in ORBIT-GPU are from the code TRANSP, which is widely used for solving the magnetic equilibrium of plasma among other features (*3*). However, it can not solve for perturbations within equilibrium. Instead, this is solved with a separate code, the kinetic post-processor NOVA-k.

Simulation Parameters

Equilibrium data and fast ion distributions from National Spherical Tokamak Experiment (NSTX), a tokamak at Princeton Plasma Physics Laboratory (PPPL), run 128741 are passed from TRANSP into ORBIT-GPU for the main particle pushing. Simulations are run on the Princeton Traverse Cluster typically using 1-2 IBM POWER9 nodes (with 4 Nvidia V100 GPUs per node).

Parameters represent typical fusion environments and environments within NSTX. Particles are set to ones involved in common fusion paradigms, including deuterium, tritium, alpha particles, and protons. Neutrons are not included because they are neutral particles, and thus have no magnetic field interactions. Electrons are not included because their dynamics occur at a much shorter time scale due to their low mass and high velocity. The perturbation type used is a neoclassical tearing mode, which is a radial structure for a mode taken from an analytic theory. Perturbations are approximated to 2 mode levels. The background magnetic field is 4.85 kG. The simulation environment runs for 500 ms.

Validation Metrics

A simple way of validating this code is to compare the end particle distributions of simulations with FO and GC effects. Most importantly, this can be used to analyze particle confinement. The last closed flux surface (LCSF) is the topological boundary that contains information about the magnetic field. In the case of NSTX runs, it also coincides with the outboard walls of a tokamak which is practically where the ionized particles will undergo much more complicated edge physics. Seeing how many particles fall outside the LCSF with FO effects can give an estimate of how well the GC approximation works for the core plasma.

RESULTS

From the increasing number of particles outside the LCFS, the GC approximations seemed to greatly decrease in accuracy as the energy and mass of the particles increased. For example, triton particles with energies ranging from 100-1010 keV are 3.64 times more likely to be found outside the LCFS than triton particles with 10-100 keV, demonstrating how approximations decrease with energy. Triton particles with energies ranging from 100-100 keV are 1.04 times more likely to be found outside the LCFS than lower-mass deuteron particles with 10-100 keV, demonstrating how approximations decrease with mass. The various particles, energy ranges, and percentage of particles that fall outside the LCFS are represented in Table 1. From it, the closest-modeled approximations include particles from NBIs and helium-3 fusion, with roughly 8-13% of particles outside the LCFS given FO effects. Examples include the 10-100 keV deuteron and triton orbits (Figure 3). These two distributions have less than 10% deviation between GC and FO, with 8.332% and 8.646% of FO particles respectively falling outside the surface. Particles in the 100-1000 keV range display a varying amount of accuracy (Figure 4). Helion is 13.8% while triton jumps to 31.4%. The rest of these simulations were less accurate by an order of 10, with roughly 30% particles outside the LCFS, such as the proton and alpha particle distributions (Figure 5). From Table 1, this can range from 13.772% outside for helium-3 to 31.404% outside for triton. In addition, a visual analysis showed that these effects were most pronounced at the outboard edge near the walls.

Table 1

Emin (keV)	Emax (keV)	Particle	% Outside
10	100	D	8.332
10	100	T	8.646
100	820	$He3$	13.772
100	1010	T	31.404
100	3020	p	31.273
100	3520	$alpha$	33.320

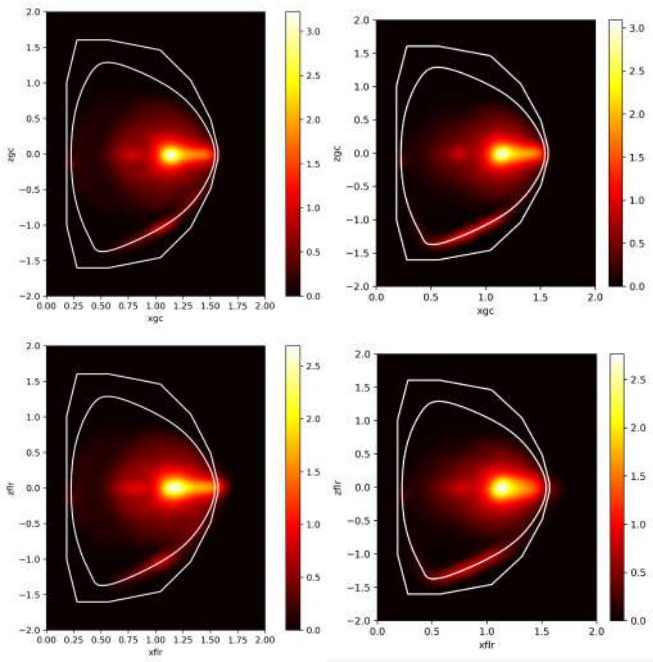


Figure 3: 1-100 keV cross-sectional guiding center density (top) and full orbit (bottom).
Guiding center and full orbit distributions are closely related, with around 8% full orbit particles outside the last closed-flux surface (LCFS) written in Table 1. Color is probability density function and scale is in meters.

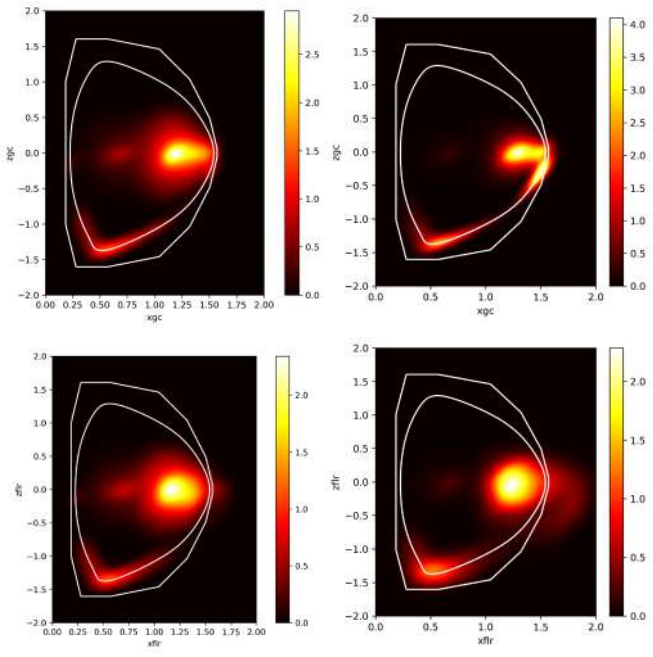


Figure 4: 0.1-1MeV cross-sectional guiding center density (top) and full orbit (bottom).
Distributions differ more than in Figure 3, with around 13% full orbit particles outside the last closed-flux surface (LCFS) written in I. Color is probability density function and scale is in meters.

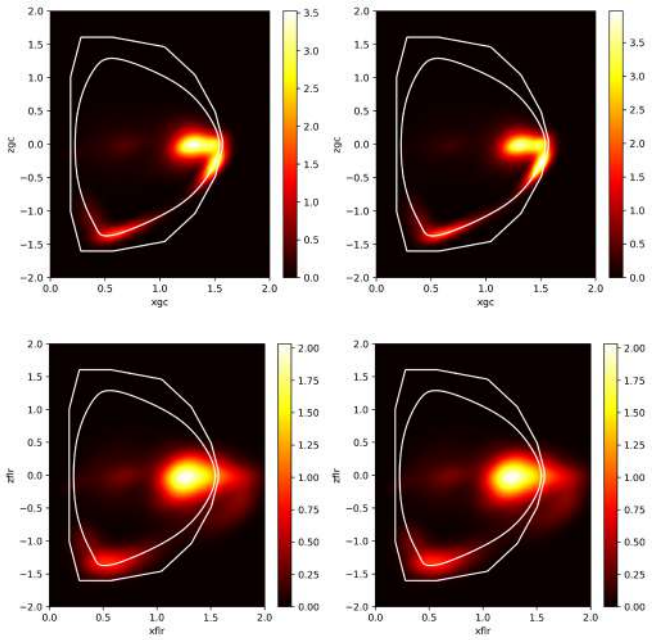


Figure 5: 0.1-3MeV cross-sectional guiding center density (top) and full orbit (bottom).
Distributions widely differ, with around 30% full orbit particles outside the last closed-flux surface (LCFS) written in I. Color is probability density function and scale is in meters.

DISCUSSION

An initial test of particles that fall outside the LCFS given FO effects shows that the deviation between GC and FO grows much larger, especially for particles in the MeV range. Therefore, while the guiding center approximation is more than 90% accurate for particles in the keV range, which is typical of NSTX injections, it greatly drops to around 60% accuracy once particles reach burning plasma thresholds. This implies that future experiments which aim to target burning plasma conditions, including ITER and SPARC, may be inaccurately represented by these kinds of calculations, especially in the edge regions of a tokamak.

It is important to note that the number of particles outside only counts as particles outside the LCFS at a given time when using GC approximation. Given their large radii, particles that circle outside and back into the flux regions are still counted as inside the LCFS. Furthermore, roughly half of the particles which stop at the LCFS will have a phase angle into the tokamak that counts them as inside the chamber. Therefore, this loss criterion can be massively improved within the ORBIT-GPU code by directly analyzing particle trajectory and extending ORBIT-GPU to the vacuum region (*11*). Since we observe a change in losses by more than 30% when including FO effects, it is safe to assume that the GC assumptions fail for these ion species.

Future improvements to the simulation can be used to better characterize this effect. For example, we can visualize fast ion interaction with the mode as well as the mode itself in order to show the radial mode structure and fast ion resonance. This feature can be further adapted to mark particles when they reach a limiting zone, such as the outboard wall. In addition, a more accurate GC approximation is yet to be implemented, such as calculating the vacuum plasma transition outside the LCFS and tracking every particle in their respective orbits while keeping the code efficient.

An analysis of particle transport using the GC code ORBIT-GPU demonstrates that while the GC approximation is a good cost-effective method of simulating transport for low energy EPs, it becomes much less accurate for fusion condition EPs. These can inform diagnostics, design, and control of a tokamak. With magnetic confinement burning plasma experiments on the horizon, finding accurate methods to model these processes is critical for research to progress.

ACKNOWLEDGEMENTS

This work was made possible by funding from the Department of Energy for the Summer Undergraduate Laboratory Internship (SULI) program. This work is supported by the US DOE Contract No. DE-AC02-09CH11466.

REFERENCES

1. C. F. Clauser, R. Farengo, H. E. Ferrari, Focus: A full-orbit CUDA solver for particle simulations in magnetized plasmas. *Comput Phys Commun.* **234**, 126–136 (2019), doi:10.1016/j.cpc.2018.07.018.
2. F. M. Poli, Integrated Tokamak Modeling: When physics informs engineering and research planning. *Phys Plasmas.* **25**, 055602 (2018). doi:10.1063/1.5021489
3. J. Breslau et al., TRANSP. (2018), doi:10.11578/dc.20180627.4.
4. J. Cizek et al., Cold sprayed tungsten armor for tokamak first wall. *Coatings*. **9**, 836 (2019). doi:10.3390/coatings9120836
5. J. P. Freidberg, (2014). Ideal MHD (Cambridge: Cambridge University Press).
6. K. G. McClements, E. D. Fredrickson, Energetic particles in spherical tokamak plasmas. *Plasma Phys Contr F.* **59**, 053001 (2017). doi:10.1088/1361-6587/aa626e
7. M. Podestà et al., Development of a reduced model for energetic particle transport by Sawteeth in Tokamaks. *Plasma Phys Contr F.* **64**, 025002 (2021). doi:10.1088/1361-6587/ac37fe
8. M. Podestà, M. Gorelenkova, R. B. White, A reduced fast ion transport model for the Tokamak Transport Code Transp. *Plasma Phys Contr F.* **56**, 055003 (2014). doi:10.1088/0741-3335/56/5/055003
9. N. A. Crocker et al., Novel internal measurements of ion cyclotron frequency range fast-ion driven modes. *Nucl Fusion.* **62**, 026023 (2022). doi.org/10.1088/1741-4326/ac3d6a
10. N. N. Gorelenkov et al., Beam ion driven instabilities in the National Spherical Tokamak Experiment. *Phys Plasmas.* **11**, 2586–2593 (2004), doi:10.1063/1.1689667.
11. P. J. Bonofiglo et al., Numerical Studies on saturated kink and sawtooth induced fast ion transport in Jet Iter-like Plasmas. *Nucl Fusion.* **62**, 112002 (2022). doi:10.1088/1741-4326/ac888c
12. R. B. White, M. S. Chance, Hamiltonian guiding center drift orbit calculation for plasmas of arbitrary cross section. *Phys Fluids.* **27**, 2455 (1984). doi:10.1063/1.864527
13. S. E. Wurzel, S. C. Hsu, Progress toward fusion energy breakeven and gain as measured against the Lawson Criterion. *arXiv.* (2022). doi:10.1063/5.0083990.

2022-2023

REVIEW BOARD

(From left to right) Top row: Oliver Wang, Aki Takahashi, Cole Woulbroun, Jordan Lin, Nathaniel Chen, and David Mastro. Second row: Nathan Ou, Miki Matsuoka, Emma Villinski, Ethan Hung, Agamroop Kaur, Aadil Rehan, and Nathan Joshua. Third row: Third row: Marissa Li, Ryan Wong, Nyah Zhang, Malvika Iyer, Eliana Bohn, Wasila Sun, Yuelai Feng, and Raphael Low. Bottom row: Melody Jiang, Aishwarya Atmakuri, Caroline Sha, Sophia Isabel S Pulido, Rebekah Jin, Jessica Goeij, and Caden Chow. Not pictured: Rachel Ma, Natalie Tsubamoto, Haley Xiaohe Zhang, and Roy Zhao.

LAYOUT BOARD

(From left to right) Top row: Truman Ma, Ahmad Ismail, Dashrit Pandher, and Christy Ma. Bottom row: Crystal Woo, Elise Tran, Christine Zhang, and Chelsea Lai.

EDITORIAL BOARD

(From left to right) Top row: Andrew Wang, Martin Sevcik, Kyle Nguyen-Ngo, Aniket Das, and Emily Lin. Middle row: Bethany Woo, Brendan Sam, Sohan Talluri, Lily Zello, Brynn Beatty, and Timothy Liu. Bottom row: Ruthy Shin, Nhi Pham, Chloe Nelson-Torakawa, Ying Yin (Lina) Zhu, Isabel Angres, and Maanasi Dhiraj Narayan. Not pictured: Caitlyn Brahim, Zihui (Hailey) Li, Charlotte Chen, and Evelynn Chen.

USJ STAFF

The UCLA Undergraduate Science Journal is a peer-reviewed publication registered with the Library of Congress, featuring top-quality research performed by undergraduates in all STEM fields, including life sciences, engineering, statistics, physics, mathematics, computer science, and more. We aim to provide a multidisciplinary platform that allows students to publish their research and engage with the larger research community at UCLA and also involves students with the peer review process. Submitted manuscripts undergo a rigorous review and publication process run completely by undergraduates.

USJ accepts manuscript submissions and staff applications annually in the fall.

Correspondence should be addressed to Dr. Jorge A. Avila, Life Sciences 2121, 621 Charles E. Young Dr. S. Los Angeles, CA 90095 or the UCLA Undergraduate Science Journal at usj@ucla.edu.

We acknowledge the generous support of the Undergraduate Research Center – Sciences and of the UCLA Clinical and Translational Science Institute (CTSI) grant UL1TR001881 to publish this journal.

UCLA Undergraduate Education Research Centers – Sciences

UCLA CTSI Grant #UL1TR001881