A Nervous Breakdown

The Role of Caenorhabditis Elegans Peroxidasin PXN-2 in Axon Regeneration

- Research as the Path to 21st Century Skills
- Implementing Independent Student Research Programs for Teachers with a Lack of Research Experience
- Application of Bioinformatics in the Revelation of NSCLC Biomarkers and Potential Targeted Drug Therapies
- A Comparison of Drosophila CMI to Human MLL2/ALR Reveals Homologous Histone Binding and Recognition Preferences
- Analysis of Acceptance of Grammatical Changes in Standard American English Reveals Differences by Age and Education
Staff and Reviewers - October 2012

Academic Advisory Board
Cornelis Murre, PhD
Professor of Molecular Biology, UC San Diego

A.Ramachandra Rao, PhD
Emeritus Professor of Civil Engineering, Purdue University

Vanessa Svhla, PhD
Assistant Professor, Teacher Education
University of New Mexico

Nagesh Ragavendra, M.D.
Professor of Radiological Sciences, UCLA Medical Center

Editorial Board
Ravi Dinakar, PhD, Biology IC- Molecular Biology
Editor in Chief

Alicia Bicknell, PhD, Biology IC- Biochemistry
Post Doctoral Fellow, UMASS Medical School

Bernadette Roche, PhD
Chair, Dept. of Biology, Loyola University

Dahlia McGregor, PhD
President, New York Biology Teachers Association

Peter J. Cook, PhD
Research Fellow, Sloan Kettering Institute

Jennifer Stubbs, PhD
Scientist, Pathway Genomics

Judith A. Scheppler, PhD
Coordinator, Student Inquiry and Research, IMSA

Nancy C. Norton, PhD
Associate Professor, Chemistry and Biochemistry
University of Arizona

Claire Standen, PhD
Instructor in Molecular Medicine, UMASS Medical School

Diana Laboy-Rush, PhD
STEM Solutions Manager, Learning.com

Barbara Nicholson, PhD
Professor, Central Connecticut State University

Staff
Technology Director:
Scott Fenton, BS, Physics IC - Mechanics
Senior Infrastructure Engineer, MphasiS
CEO, Gnosis Tutoring

Legal Counsel:
Vijay Dinakar, Esq., Dinakar and Associates

Consultant:
Daniel Antonelli, Esq., Antonelli & Antonelli

Proof Editor:
Charissa Sorrentino, MA, English IC - Exposition
Teacher, Immaculata High School

October 2012 Issue Reviewers
Olga A. Timofeeva, PhD
Assistant Professor
Georgetown University, Lombardi Comprehensive Cancer Center

Jeffrey Field, PhD
Professor of Pharmacology
University of Pennsylvania

Randall J. Roper, PhD
Associate Professor of Biology
Indiana University-Purdue University Indianapolis

Chad Fertig, PhD
Assistant Professor of Physics
University of Georgia

Elizabeth A. Cowles, PhD
Professor and Assistant Chair, Biology
Eastern Connecticut State University

Pam Auburn, PhD, Lecturer at University of Houston

Aaron Alford, PhD
Executive Director of Science and Education
Paleo Quest

Rodolfo Hermans, PhD
Research Associate, London Centre for Nanotechnology

John Cuozzo, PhD
Director, Protein Biochemistry
X-Chem, Inc.

Ashton Powell, PhD
Instructor of Biology
North Carolina School of Science and Mathematics

Patrick C. Mathias, PhD
Illinois Medical Scholar, Department of Bioengineering
University of Illinois at Urbana-Champaign

Caroline K. Hutton, PhD
Drug Testing and Analysis Editorial Advisory Board Member
Sports Anti-Doping Science Consultant

Jonathan Wolf, M.A., Ed.M.
Adjunct Assistant Professor of Physics, Hofstra University

Veronica Gross, PhD
Lecturer at Curry College

Robert New, MA
Emerging Computational Linguistics

Kevin Knobbe, PhD IC - CERN
Professor of Physics, Estrella College

Reviewers from Editorial Board: Peter J. Cook, PhD, Jennifer Stubbs, PhD, Alicia Bicknell, PhD
Table of Contents

Editorials:

Research as the Path to 21st Century Skills, Mark Little, M.A., NBCT President-Elect, National Association of Biology Teachers, Broomfield High School: #1 Eagle Way, Broomfield, CO 80020, Donald P. French, PhD, President, National Association of Biology Teachers, Professor of Zoology, Oklahoma State University: 501 Life Sciences West, Stillwater, OK 74078; Implementing Independent Student Research Programs for Teachers with a Lack of Research Experience, Darci J. Harland, PhD, Assistant Director of Research at the Center for Mathematics, Science, and Technology, Illinois State University: Normal Ave & W Locust St, Normal, Illinois 61761, Author of the STEM Student Research Handbook

p. 1 - 5

In School Articles:

The Role of Caenorhabditis Elegans Peroxidasin PXN-2 in Axon Regeneration
Claudia Shin*, Christine Marshall-Walker2, and Joel Jacob2
Student1, Teacher2: Phillips Academy, Andover, MA 01810

p. 6 - 11

Internship Articles:

Application of Bioinformatics in the Revelation of NSCLC Biomarkers and Potential Targeted Drug Therapies
Mukund Subramanian1*, Jahangheer Shaik3, Rakesh Nagarajan3, and Theodore Waterson3
Student1, Teacher3: Ladue Horton Watkins High School, Saint Louis, MO 63124
Mentor2: Center for Biomedical Informatics, Department of Pathology and Immunology, Washington University, Saint Louis, MO 63124

p. 12 - 19

A Comparison of Drosophila CMI to Human MLL2/ALR Reveals Homologous Histone Binding and Recognition Preferences
Riva Trivedi1, Claudia Zraly2, and Andrew Dingwall2*
Student1: The Illinois Mathematics and Science Academy, 1500 Sullivan Road, Aurora, IL 60506
Mentor2: Loyola University Stritch School of Medicine, 2160 South 1st Avenue, Maywood, IL 60153

p. 20 - 24

Analysis of Acceptance of Grammatical Changes in Standard American English Reveals Differences by Age and Education
Sophie Lipson1*, Ileana Rios2, and Susan Behrens3
Student1, Teacher2: Trinity School, New York, NY 10024
Mentor3: Marymount Manhattan College: 221 East 71st Street, NY 10021 and Columbia Summer High School Program, New York, NY 10027

p. 25 - 27
Research as the Path to 21st Century Skills

Mark Little, M.A., NBCT
President-Elect, National Association of Biology Teachers
Broomfield High School: #1 Eagle Way, Broomfield, CO 80020

Donald P. French, PhD.
President, National Association of Biology Teachers
Professor of Zoology
Oklahoma State University: 501 Life Sciences West, Stillwater, OK 74078

Since 1938, the National Association of Biology Teachers (NABT) has promoted the adoption and dissemination of “best practices” in biology education, including authentic student research (see Watson 1957 for an early explicit example), at all levels. Today, more than a decade into the 21st century, we recognize the importance of biology education not only as a path to future employment in health care, biotechnology, and emerging research disciplines, but also for the nation’s citizenry who must make ethical, social, personal and professional decisions that affect themselves and society at large, both currently and in the future. As biology educators, we are transitioning from the National Science Education Standards to the Framework for K-12 Science Education (NRC 2012) and the Next Generation Science Standards, which will guide instruction at the pre-college level and prepare students for the revised college environment shaped by the guidelines outlined in reports like Vision and Change (AAAS 2010). While doing so, we see an increasing emphasis on treating scientific content and scientific practices as equal components of the well-designed instructional environment and on structuring education so it integrates these within and across disciplines, preparing students for careers where disciplinary boundaries are blurred and innovation is increasing.

What do these scientific practices include? They are the activities that biologists do to investigate the world around them: asking questions, creating and using models, experimenting, evaluating data and information, developing hypotheses (explanatory statements), thinking mathematically and computationally, arguing from evidence, and communicating (NRC 2012). Readers of this journal should clearly recognize these as the components involved in conducting research. It should also be clear that students will learn them best by engaging in these practices, reflecting on successes and failures, learning to judge their own competencies and receiving guidance from knowledgeable practitioners, proficient as researchers and mentors. By doing so, students should reach the goal of being able to perform science practices competently on their own in novel situations.

Should every student conduct research? Actually they do every day when they confront problems with unknown solutions, observations for which underlying causes are unknown, and propositions to evaluate. When the car doesn’t start or a MP3 player stops working, a student proposes an explanation or solution and tests it. Such simple examples require only a few rudimentary science skills and logical thinking, but they illustrate that science practices are directly imbedded in the lives of all students. As students take on adult responsibilities, it becomes pertinent to all of us that they can evaluate claims and data presented in organizational reports, marketing literature, political advertisements, jury trials, and ballots related to issues of climate change, teaching of evolution, stem cell research and use of animals or humans in research. These require not only content knowledge, but also an understanding of the nature of science and scientific research.

Does conducting research actually have the desired effects on students, e.g. improving their understanding of science or increasing the likelihood that they will obtain STEM degrees? A look at almost any of the multitude of studies at the high school and undergraduate level provides clear support, so a few examples should suffice here. Sadler et al. (2010) and Stake and Mares (2001) found positive effects on science career choices, including moving
away from choices of health professions toward the sciences as students became more aware of career options. Roberts and Wassursug (2009) examined a high school program’s effects on students over a period of decades and found that students were significantly more likely to be actively engaged in research as part of their careers if they conducted research as high school students. A review of the research literature from half a century (Sadler, 2010) found positive influences on students understanding of the nature of science though it may be limited without explicit instruction (Bell et al. 2003). Charney et al. (2007) found increases in both content knowledge, an area of concern for many skeptics, and the understanding of science. However, because the effects of research participation can vary with the support the mentor provides, how much the collaboration occurs in the research environment, how much the student is allowed to contribute to the direction and design of the research, the degree to which the student understands the importance of the research results, and the personal interest a student has in the research (Burgin et al. 2012), it is important for teachers and researchers who may be participating in guiding the students to consider their roles carefully and seek out professional development opportunities to refine their skills at mentoring, which means more than simply directing or advising (COSEPUP 1997).

Education is a community effort that requires shared and specialized contributions. Thus we urge all teachers, scientists, administrators, and parents to support efforts to involve students in research not only in every science class, but through opportunities to participate in science fairs and in research labs. More specifically, teachers need to develop curriculum and rethink their teaching practices to incorporate science practices into the classroom and to individualize instruction so each student can find a research topic of interest and all students can understand how science practices led to what we know. Scientists should accept high school students (and teachers) as researchers in their labs, recognizing that mentoring involves offering grade-appropriate guidance, the opportunity for students to develop ownership in their projects, and developing a relationship with a student. To excel, teachers and scientists should engage in professional development through educational and scientific organizations such as NABT, who are aimed at structuring and conducting high quality research experiences for students. They should also seek to establish networks between high schools, universities or businesses, and policy makers to promote and facilitate student research opportunities. NABT BioClubs were envisioned as incubators for such relationships. Administrators should strive to provide the resources needed for students to engage in laboratory experiences that model science including space, time, equipment, supplies, and professional development. Parents should encourage their children, and demand from educators and administrators that science is for all, not just some. Students should embrace the challenge, look for the opportunity, and perfect the skills and reasoning that scientific research offers.

Students should also encourage each other, just as scientists do, by celebrating successes, consoling friends after failures, offering suggestions, accepting critiques, and sharing opportunities. Some of the rewards of research are immediate, some delayed, but all prepare you with the skills and habits of mind needed for the research-based decision-making that will lead to success the 21st century.

References


Implementing Independent Student Research Programs for Teachers with a Lack of Research Experience

Darci J. Harland, PhD
Assistant Director of Research at the Center for Mathematics, Science, and Technology
Illinois State University: Normal Ave & W Locust St, Normal, Illinois 61761
Author of the STEM Student Research Handbook
dharlan@ilstu.edu

The importance of inquiry has long been promoted in the scientific education community as crucial for developing competent STEM professionals. However, like any buzzword, many classroom teachers are artificially attributing what they do in the classroom as inquiry.

In my article What Inquiry is NOT, I share how I incorrectly categorized my class as such, when in fact it was not…and how I was able to develop a curriculum that truly was inquiry-based for my classroom. While there are varying levels of inquiry, the best way to provide fully-engaged, student-centered inquiry is to allow students to experience the scientific method from beginning to end from deciding the problem to communicating the results.

Teachers and professors who implement student independent research programs know that students gain not only laboratory skills and advanced science content knowledge, but also intangibles such as a better understanding of the nature of science, increased confidence in their ability to perform science, and an ability to learn scientific facts in context of something about which they are passionate. Yet, even though the number of teachers who believe inquiry to be the best way for their students to learn is growing, most remain unable to overcome the hurdles of implementing these opportunities for their students.
Though each teacher has specific reasons that might prevent the implementation of independent student research in the classroom, the most common hurdles generally fall into the following categories: 1) Teachers are uncertain of their ability to teach research; 2) Teachers feel they lack the resources and support to provide quality experiences for students; 3) Teachers are unsure whether students have the laboratory and critical thinking skills to perform such intensive research; and 4) Teachers are accustomed to having full control of curricular content and outcomes. In this article, I will address the first of these hurdles.

Classroom teachers often lack science research lab experience, which then leads to a lack of confidence when it comes to implementing independent student research. The problem for some teachers begins in their preservice teacher education programs. While the science lab courses they take along side pre-med and future researchers, it likely that unless they’ve sough out research opportunities, they’ve never focused on a single topic for an extended period of time. And teachers who attend graduate school often do so in education programs, not in their area of science, where any original research is in education and pedagogy-based research, instead of using the scientific method to study a scientific topic. This lack of science research experience is problematic for two reasons. First, it creates doubt in teachers’ minds whether they have the skills and authority to guide students through their own scientific research projects—and the fear is well founded. Second, many career teachers have never written a scientific paper, and therefore may feel inadequate to guide students through this process.

Whether we want to admit it or not, those of us who are career educators have a different view of the scientific process than do career scientists. That’s not to say that it should keep us from guiding students through the process of long-term research projects, but transparency regarding our weakness and skewed career-teacher perspective is healthy, not only for our own sanity, but also for the learning of our students.

The best way to address this issue is to provide research experiences for teachers. I was involved with two programs that became instrumental in my ability and confidence to implement semester-long student research projects with high school students. The National Science Foundation (NSF) has “Research Experiences for Teachers” or RET grants that encourage laboratory scientists to develop partnerships with K-12 teachers which include research experience for the classroom teachers. And NSF’s Graduate Teaching Fellows in K-12 Education, (GK-12) was an NSF grant that funded projects that partnered scientists in graduate schools with classroom teachers. Instead of providing the teacher with research experience, it paired scientists with classroom teachers to develop curriculum that better mirrored how scientists conduct research. Any program or resource that bolsters teachers’ confidence and helps them to face their assumptions about science research is crucial for creating a change in today’s STEM classroom climate.

Not every teacher has the benefit of using an NSF program as a resource, so it is important that the teacher reach out to the scientific community directly. While teachers can be the sole expert contact for their students; students have a more intensive experience if paired with a research mentor. These mentors may be Principle Investigators or graduate students working in a university or industry lab. But mentors may also be online and work with a student from a distance to develop a strong research design. This relieves the teacher of having to be a content expert on every topic for each student. Instead, the mentor is the content expert, while the teacher provides a framework of the scientific method of which the student will need.

The lack of scientific research experience is also problematic because it means that most teachers have never written a scientific paper. Depending on the preservice teacher training, career science teachers may have turned in lab write-ups for their science courses, some of which may have been lengthy, but not the type of writing that would be turned into a scientific journal for publication. Teachers’ writing experience—like their research experience—is heavily centered on educational philosophy. While teachers know the parts of a scientific paper and have read their share of scientific journal articles, writing on an extensive experiment which they designed and implemented is not a skill that most teachers have. Even if a teacher is confident regarding their own scientific research skills, a teacher may be uncomfortable having to teach literacy skills to his or her students. Traditionally, English departments have been responsible for teaching students prewriting strategies, library research skills, documentation styles, and editing multiple drafts of large papers.
Similar to addressing the lack of research experience, teachers can still provide strong experiences for their students by acknowledging their own weaknesses. Science teachers must reframe how they view their own responsibilities. The Common Core Science Standards emphasize critical thinking and the ability to apply knowledge over memorizing facts. Even for students who do not become STEM professionals, being able to sift through mountains of content to find answers to questions and being able to write and communicate succinctly are skills that will benefit students no matter what they decide to do. Therefore, teachers should seek out resources to help them teach basic research skills. English teachers and librarians are invaluable here as may be the STEM Student Research Handbook, as it is written to help guide both students and teachers through the entire research process. Students should be expected to write papers—worthy of publication—in journals such as JESS. This provides students with an audience beyond the classroom, and an understanding of how science knowledge is acquired.

So the question boils down to, “Can a teacher with limited research experience teach and guide students in the research process?” I believe teachers, even with their lack of science laboratory experience, can and should provide students with rich research opportunities.

Teachers who conduct research with their students need to communicate with the science education community. Teachers must share their experiences with other teachers. We must encourage one another, provide tips, point out pitfalls, and share resources. This communication can be accomplished by submitting articles for publication (like NSTA publications) and by presenting at conferences. Only when teachers realize that they are not alone in their fears of teaching research will they be encouraged to try. If more teachers address their assumptions about how science is really done, reach out to the scientific community, and place a newfound emphasis on communicating scientific results, the more likely we are to get real change in science education.
IN SCHOOL ARTICLE

The Role of Caenorhabditis Elegans Peroxidasin PXN-2 in Axon Regeneration

Claudia Shin1*, Christie Marshall-Walker2, and Joel Jacob2
Student1, Teacher2: Phillips Academy, Andover, MA 01810
*Corresponding author: claudiashin@gmail.com

Abstract

We investigated the effect of peroxidasin pxn-2 on axon regeneration in C. elegans. Peroxidasins comprise a gene family with homologues in C. elegans, Drosophila, and Homo sapiens. Peroxidasins are believed to contribute to the extracellular matrix, but the function(s) of peroxidasin proteins are currently unknown. We used RNA interference to display a mutant-like phenotype through pxn-2 knockdown in wild type and unc-70 worms. The unc-70 worms carry a β-spectrin mutation rendering their axons unable to withstand mechanical strain, and hence they break in response to natural tensile forces incurred during movement. The axons of DD Gaba-ergic motor neurons of both wild type and unc-70 worm strains were tagged with green fluorescent protein (GFP), so the extent of regeneration in axonal commissures could be easily assessed. We found that the axon regeneration in the unc-70 worms that exhibited pxn-2 knockdown were significantly inhibited, with an average of 5.5 complete commissures compared with the average of 9.9 complete commissures found in the unc-70 worms with normal pxn-2 expression. With an average of 4.75 incomplete commissures and 4.23 missing commissures, the unc-70 worms affected by pxn-2 knockdown indicate that pxn-2 plays an important role in axon regeneration.

Introduction

One in fifty Americans suffers from some form of paralysis1. This amounts to over six million people with a permanent loss of muscle function and in some cases, a loss of sensory perception as well. The main cause of paralysis is spinal cord injuries, when a serious injury damages the spine in a way that severs the central nervous system axons. These broken axons are then unable to transmit information to and from the brain; therefore, any part of the body beyond the breakage cannot send or receive neural messages. There is currently no cure for paralysis, as damaged central axons in humans do not heal. However, some non-human species have the incredible ability to regenerate their central axons after they break2. Invertebrates, in particular, make excellent models for scientific study of axon regrowth since they can regenerate healthy and fully functional axons in a relatively short amount of time2. The challenge lies in identifying the genes that promote axon regeneration in invertebrates. If these genes can be isolated, and if they have human homologues, they would become potential therapeutic targets for the induction of axon regrowth in humans.

One promising gene family is the peroxidasin family, as recent research indicates that peroxidasin proteins play a role in axon regeneration3. Peroxidasins are relatively newly discovered genes with largely unknown functions, which make them a potential untapped resource for research. Furthermore, these genes are highly conserved across many species, including C. elegans, Drosophila, and, most importantly, Homo sapiens. Peroxidasins were first discovered in Drosophila in 1994 and characterized as peroxidase enzymes in the extracellular matrix (ECM)4. This finding was corroborated by further evidence of peroxidasins in the ECM of Drosophila myofibroblasts and fibrotic kidney cells5. Peroxidasins appear to reside in the ECM, but their function is unclear. Peroxidasins from Drosophila are known to be closely related to human peroxidase enzymes, so peroxidasins may catalyze hydrogen peroxide reactions such as radioiodination and oxidation in the same way that the glutathione, horseradish, and thyroid peroxidases do in humans6. This relation indicates that peroxidasin function may be conserved across species.

Potential roles for peroxidasins have been proposed, pointing to involvement in hydrogen peroxide reactions, consolidation and stabilization of the ECM, phagocytosis and apoptosis, and immune system defense4,5,6. However, these functions are still not definite, and certainly not exclusive. A recent study in C. elegans showed that peroxidasins promote embryonic development but inhibit axon regeneration1. The specific peroxidasin gene studied was pxn-2, which is homologous to the Drosophila peroxidasin and the human PXDN gene6. The same study showed that peroxidasins are located in the ECM and suggested that the substrate for peroxidasins may be type IV collagen in the basement membrane of C. elegans6,7. This exciting research highlights peroxidasins as a potential candidate gene for studies in search of novel therapies for spinal cord injuries.

In this experiment, peroxidasin pxn-2 was knocked down in C. elegans using RNAi by feeding to understand the role of pxn-2 in axon regeneration. A similar study indicated that pxn-2 should inhibit axon regrowth. In that study, laser axotomies were used to sever worms’ axons3. This current experiment utilizes RNAi sensitized unc-70 C. elegans with a β-spectrin mutation that removes the axons’ ability to withstand mechanical strain. The worms that carry this β-spectrin mutation break their axons naturally and more often than wild type worms. This genetic model for axon regeneration assays avoids the potentially confounding trauma of laser axotomy and should provide a better, more controlled environment to examine axon regeneration.

All experiments were conducted as post-embryonic developmental assays to avoid embryonic lethality and to focus on regeneration rather than normal axonal development. Wild type worms that possessed β-spectrin served as the control group, and the DD motoneuron axons in these worms were tagged with GFP. L1 stage wild type and unc-70 worms were placed on RNAi plates with the pxn-2 RNAi feeding strain to assess the effects of pxn-2 knockdown in worms with and without the β-spectrin mutation. In a separate experiment, unc-70 worms were fed either
pxn-2 RNAi feeding strain bacteria or control bacteria to assess the effects of pxn-2 knockdown in worms with the β-spectrin mutation.

These experiments demonstrated that the knockdown of pxn-2 inhibited axon regeneration in unc-70 worms but had virtually no effect on wild type worms. The difference in the numbers of complete and incomplete commissures between the unc-70 worms that fed on control versus pxn-2 bacteria was statistically significant, suggesting that pxn-2 is a key component of axon regeneration. The function of peroxidasins in the ECM is still unclear, but we propose three hypotheses that address potential functional roles for peroxidasins in axon regeneration.

**Materials and Methods**

**RNAi Experiments:** HT115 feeding strains containing L4440 vectors encoding targets to pxn-1 and pxn-2 were attained for this experiment (a kind gift from Dr. Andrew Chisholm, University of California, San Diego). To re-streak the bacteria onto fresh LB/AMP/TET plates and LB/AMP plates, a protocol was used from the Dolan DNA Learning Center in the Cold Spring Harbor Laboratory at Cold Spring Harbor, NY. Interestingly, the pxn-2 feeding strain grew on both types of plates, the L4440 empty vector grew only in LB/AMP, and the pxn-1 failed to grow. The experiment was conducted without the pxn-1 since that feeding strain could not be salvaged. RNAi worm plates were seeded with the pxn-2 feeding strain, and glycerol stocks were prepared for later study. For RNAi experiments, C. elegans were cultured on LB agar plates that contained 10 mg/mL sterile ampicillin and 1M IPTG (Isopropyl-B-D-thiogalactopyranoside), according to a standard protocol also from the Dolan DNA Learning Center (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

**C. elegans worms:** In this experiment, wild type and unc-70 RNAi sensitized C. elegans with a β-spectrin mutation were used (a kind gift from Dr. Marc Hammarlund at Yale University). All worms used in this experiment had axons of DD Gaba-ergic motor neurons tagged with green fluorescent protein (GFP). **Bleach protocol for unc-70 eggs:** To study post-embryonic development, unc-70 C. elegans eggs were picked rather than L4-stage worms. However, these paralyzed worms laid very few eggs, so instead, the plates were bleached to collect eggs, following a protocol from the Journal of Visualized Experiments (JoVE). The worms and eggs were washed off of the RNAi plate with sterile water and a bleach/NaOH solution composed of 2.5mL of bleach, 1 mL of 10N NaOH, and 6.5 mL of water was added. After the worms dissolved, the bleaching reaction was neutralized with 5 mL of M9 buffer. The remaining pellet of eggs was pipetted onto RNAi plates and allowed to hatch. However, when the RNAi plates were inspected later, no eggs were found. It seems that the unc-70 C. elegans were already so compromised by their β-spectrin mutation that the eggs were unable to withstand the bleach procedure. **Microscopy:** C. elegans were mounted onto agar padded slides and examined under an Axioplan Zeiss microscope. The agar padded slides contained 2% agarose and M9 buffer. Worms were examined using the GFP filter at 10x and 20x magnification and imaged using a Hamamatsu camera and Openlab software.

**Results**

In this experiment, we studied four groups of C. elegans: wild type worms feeding on control bacteria, wild type worms feeding on pxn-2 RNAi bacteria, unc-70 worms feeding on control bacteria, and unc-70 worms feeding on pxn-2 RNAi bacteria. To examine the role of pxn-2 in axon regeneration, we conducted two separate studies with these four worm groups. The first study compared wild type and unc-70 worms when both were exposed to pxn-2 RNAi bacteria, and the second study examined the effect of pxn-2 RNAi bacteria on unc-70 worms alone. Using the GFP filter on the microscope, complete and incomplete commissures were counted in L4 and adult stage C. elegans from each of these four worm groups. Figures 1 and 2 show examples of complete and incomplete DD commissures as seen through the GFP filter on a microscope. Figures 3, 4, and 5 are microscope images of the C. elegans in this experiment with GFP tagged commissures.

In the control group and in the wild type worms in pxn-2 bacteria, the median number of complete commissures was 15, and no incomplete commissures were observed. The wild type worms in control and pxn-2 bacteria had nearly identical numbers of commissures, with average numbers of 14.56 and 14.50 respectively, indicating that the pxn-2 bacteria had little to no effect on the worms. However, the unc-70 worms in control bacteria had an average of 9.90 complete commissures and 1.85 incomplete commissures, and the unc-70 worms in pxn-2 bacteria had an average of 5.55 complete commissures and 4.75 incomplete commissures; this suggests that the pxn-2 bacteria did have an impact on the unc-70 worms. Also, the unc-70 worms in control bacteria had an average of 11.75 complete and incomplete commissures, which is a lower average than the total commissures counted in the wild type worms. Assuming that C. elegans have similar total numbers of commissures, the unc-70 worms in control bacteria appear to be missing about 2.78 commissures, and the unc-70 worms in pxn-2 bacteria seem to be missing an average of 4.23 commissures. These missing commissures were most likely severed commissures that were unable to regenerate. The average numbers of counted commissures for the four worm groups are compared in Figure 6.

We then analyzed the data by making the box plots shown in Figures 7 and 8. The box plots display the range of the data: the middle 50% with gray boxes, called the Inner Quartile Range (IQR), and a median line in the middle of the IQR. The dots on the box plots represent outliers, which prevent a data set from being a “normal distribution”, which is a bell curve (an outlier is any data point greater than the 3rd quartile value plus 1.5 times the IQR or less than the 1st quartile value minus 1.5 times the IQR). The box plots found medians for the four worm groups to be 15 (wild type worms in control bacteria), 15 (wild type worms in pxn-2 bacteria), 10 (unc-70 worms in control bacteria), and 6 (unc-70 worms in pxn-2 bacteria), and the median coincided with the 3rd quartile value for wild type worms and unc-70 worms in pxn-2 bacteria. We also calculated standard deviations for the four worm groups and found medians for the four worm groups to be 1.47, 1.00, 2.07, and 1.05, respectively, meaning the data for the unc-70 worms in control bacteria varied the most. In order to test the data for statistical significance, we must assume that the data follows a normal distribution (independent variables, symmetric data, and no outliers). Thus, it should be noted that...
this initial condition was not met, meaning the statistical tests may not be entirely accurate.

However, it is rare that a real data set follows a normal distribution, so we proceeded to run an analysis of variance (ANOVA test) on the four worm groups. The null hypothesis for the ANOVA test was that populations of the four worm groups should all behave the same way. The ANOVA test found that the probability that the different behavior of the four worm groups was due to chance was less than 0.0001, or 1 in 10,000. Therefore, we reject the null hypothesis, and the statistical evidence indicates that the different behavior of the four worm groups was not due to chance. Although the ANOVA test does not specify which or how many of the four worm groups behave differently, the test allows us to compare multiple groups of worms simultaneously and prevents us from rejecting the null hypothesis on false authority. Now, specific worm groups can be directly compared with the 2-sample T test, and the box plots help determine which worm groups are the best choices for comparison.

We performed the 2-sample T test on the number of incomplete commissures in the unc-70 worm groups, and this data set is visualized by the box plot in Figure 8. The plot found medians of 2 and 5 and standard deviations of 1.39 and 1.25 for the unc-70 worms in control bacteria and pxn-2 bacteria, respectively. The null hypothesis for the 2-sample T test was again that all of the unc-70 worms should behave the same way. The 2-sample T test found that the probability was also less than 0.0001, but this time, we can assert with statistical evidence that these two groups behaved differently because of the pxn-2 bacteria, since this is the only variable that changed. The direct comparison made by the 2-sample T test allowed us to isolate the variable, the pxn-2 bacteria, and establish that the data was statistically significant. We can confidently reject the null hypothesis and state that the unc-70 worms behaved differently due to the pxn-2 bacteria, not chance. It should be noted that the null hypothesis of the 2-sample T test could not be rejected without first performing the ANOVA test.
and rejecting the ANOVA null hypothesis. By rejecting both hypotheses, we can assert that the behavior of the unc-70 worms in pxn-2 bacteria differs from that of the other worms through a changed variable rather than random chance. Through statistical analysis, we were able to convincingly demonstrate that the pxn-2 knockdown (changed variable) caused inhibited axon regeneration in the unc-70 worms in pxn-2 bacteria (different behavior).

In the unc-70 worms in pxn-2 bacteria, the majority of incomplete commissures stopped abruptly halfway between the ventral and dorsal nerve cords. Also, the unc-70 worms appeared to have more growth cones on their ventral nerve cord than the wild type worms. Lastly, on the worm plates that contained unc-70 worms in pxn-2 bacteria, very few adult stage worms were observed, indicating that the unc-70 worms in pxn-2 bacteria had trouble surviving to adulthood.

Discussion

The data from the unc-70 worms in pxn-2 bacteria strongly suggests that the presence of pxn-2 is important for axon regeneration, since the suppression of pxn-2 inhibited axon regeneration. As the function of peroxidasins is unknown, the cause of the inhibited axon regrowth is also unclear. In order for axon regeneration to occur, the environment in the ECM must be optimal. As pxn-2 is likely to interact with the axon regeneration environment, we will explore three hypotheses related to its potential functionality within the ECM.

Studies on the ECM of C. elegans have found that type IV collagen is responsible for the scaffolding of the basement membrane in C. elegans, which is a thin layer of the ECM that attaches the muscle to the epidermis. Another study shows that type XVIII collagen in the basement membrane of C. elegans affects axon guidance, indicating that basement membranes play a role in axonal guidance cues. Finally, type IV collagen is believed to be the substrate for peroxidasins, since the absence of peroxidasins caused muscle detachment in C. elegans. If peroxidasins are located in the basement membrane, the pxn-2 knockdown could have prevented the axons from integrating into the muscles by impeding the scaffolding process of the basement membrane. Without proper scaffolding, it would be extremely difficult for an axon to find its way from the ventral nerve cord to the dorsal nerve cord, especially if the guidance cues were not aligned with the basement membrane scaffold. Additionally, the muscle detachment explains the lack of adults on the worm plates containing unc-70 worms in pxn-2 bacteria, since the worms would not survive long without their muscles adhered to the epidermis by the basement membrane.

The next two hypotheses discuss axon guidance cues, specifically chemotaxis. In C. elegans, axons of GABAergic DD motor neurons are guided by secreted proteins called chemoattractants and chemorepellents. These axons are usually repelled away from the ventral nerve cord by chemorepellents and then attracted towards the dorsal nerve cord by chemoattractants. This way, the axon is pulled towards the higher concentration of chemoattractants, guiding the axon to the dorsal nerve cord. The unc-6 netrin is an example of an ECM protein that produces chemoattractants near the dorsal nerve cord. Since the qualitative data in this experiment indicates that the incomplete axons did make it away from the ventral nerve cord but did not travel all the way to the dorsal nerve cord, it seems that the chemoattractant gradient was disturbed. A functioning chemoattractant gradient has a high concentration of chemoattractants near the dorsal nerve cord that decreases in concentration and results in a low concentration of chemoattractants near the ventral nerve cord. This way, the axon is pulled towards the higher concentration of chemoattractants, guiding the axon to the dorsal nerve cord. The unc-6 netrin is an example of an ECM protein that produces chemoattractants and is located near the dorsal nerve cord. If the chemoattractant gradient is disrupted, the guidance cues are unclear, and axon regeneration fails. Since the incomplete commissures in the unc-70 worms in pxn-2 bacteria seemed to have trouble attracting to the dorsal nerve cord, pxn-2 knockdown may have disrupted the chemoattractant gradient.

The third hypothesis deals with another aspect of chemotaxis, the chemorepellent gradient. Through quantitative analysis, both groups of unc-70 worms appeared to be missing commissures, as their total average number of commissures did not match the total average number of commissures of the wild type worms. Through qualitative analysis, additional growth cones were observed near the ventral nerve cord in the unc-70 worms. Since growth cones are the foremost part of a regenerating axon, it is possible that the additional growth cones are the foremost parts of the missing commissures attempting to regrow. However, since few growth cones were observed in the middle of the unc-70 worms, it can be assumed that the growth cones were having trouble leaving the ventral nerve cord. If this is the case, the chemorepellent gradient may be disrupted as well, because it is the job of chemorepellents to guide regenerating axons away from the ventral nerve cord. The unc-5 netrin is an example of an ECM protein that produces chemoattractants and is located near the ventral nerve cord. The additional growth cones observed in the unc-70 worms may also be attributed to pxn-2 knockdown, which could...
have disturbed the chemorepellent gradient.

These hypotheses suggest three different functions of peroxidasins, but all three imply that peroxidasins help provide an optimal environment in the ECM for axon regeneration. Quantitative data from this experiment strongly indicates that pxn-2 is an essential element of successful axon regrowth. It is through qualitative data that we are able to make reasonable hypotheses as to the cause of the observed inhibition of axon regeneration. These hypotheses are more like shrewd conjectures than conclusive explanations, and as the box plots indicate, these results could always use more supporting data. However, we feel that the data provided is a fairly accurate representation of the populations and behaviors of the four worm groups.

In future studies, we would like to quantify some of the qualitative observations that these hypotheses were based upon. We focused on commissure counts as the most effective way to analyze the data, but growth cones could also be counted and distance of axons from the ventral nerve cord could be measured. This data may be able to clarify some of the unanswered questions in the three hypotheses from this experiment. For example, if type IV collagen is the substrate for pxn-2, then some of the effects of the pxn-2 knockdown should be unrelated to axon regeneration. If this is the case, it is odd that the wild type worms behaved almost identically, since the wild type worms should exhibit pxn-2 knockdown effects. Another question is whether the incomplete commissures were caused by a problem with the chemoattractant or chemorepellent gradient,

Figure 7. This box plot represents the data set of complete commissures. The gray boxes indicate the Inner Quartile Range (IQR), which is the middle 50% of the data, and the median lines are within the IQR. The medians are 15, 15, 10, and 6 for the wild type worms in control and pxn-2 bacteria and unc-70 worms in control and pxn-2 bacteria, respectively. The lines extending from the IQR show the range of data, not including outliers. The gray dots represent outliers (an outlier is any data point greater than the 3rd quartile value plus 1.5 times the IQR or less than the 1st quartile value minus 1.5 times the IQR).

Figure 8. This box plot represents the data set of incomplete commissures. The gray boxes indicate the Inner Quartile Range (IQR), which is the middle 50% of the data, and the median lines are within the IQR. The medians are 2 and 5 for the unc-70 worms in control and pxn-2 bacteria, respectively. The lines extending from the IQR show the range of data, not including outliers. The gray dots represent outliers (an outlier is any data point greater than the 3rd quartile value plus 1.5 times the IQR or less than the 1st quartile value minus 1.5 times the IQR).

or both.

Future studies of peroxidasins should explore the role of pxn-2 in other species with homologous genes, especially in model organisms like Drosophila. This experiment could be redesigned to focus on quantifying some of the qualitative data in C. elegans, but the importance of pxn-2 in axon regeneration has been demonstrated. Instead, a more effective research tactic would be to identify the function of peroxidasins in other creatures in order to corroborate or contradict the hypotheses from this experiment. Also, the location of peroxidasins in the basement membrane of the ECM is not yet certain, so manipulating peroxidasins in other species may help isolate the location and substrate of pxn-2 in C. elegans and other species.

To explain the results of this experiment, we hypothesized that the pxn-2 knockdown inhibited axon regeneration through its type IV collagen substrate, through the chemoattractant gradient, or through the chemorepellent gradient. We demonstrated that the effect of pxn-2 knockdown is statistically significant in inhibiting axon regrowth in C. elegans, indicating that pxn-2 plays an important role in axon regeneration. The eventual goal of this research is to identify homologous genes with Homo sapiens to try and induce axon regeneration of the central nervous system in humans. Since peroxidasins have homologous genes in humans, the results of this experiment may be significant in the search for a cure to paralysis. However, the function of the elusive peroxidasin protein family still remains a mystery to be solved by further research.
References


7. Chisholm, Andrew and Jennifer Gotenstein, UCSD. Personal communication.


Acknowledgments
The authors would like to thank Dr. Hammarlund at Yale University for his kind gift of C. elegans for this experiment. The authors would also like to thank Dr. Chisholm at the University of California, San Diego for his kind gift of HT115 feeding strains for this experiment.
Application of Bioinformatics in the Revelation of NSCLC Biomarkers and Potential Targeted Drug Therapies

Mukund Subramanian1*, Jahangheer Shaik2, Rakesh Nagarajan2, and Theodore Waterson1
Student1, Teacher2: Ladue Horton Watkins High School, Saint Louis, MO 63124
Mentor2: Center for Biomedical Informatics, Department of Pathology and Immunology, Washington University, Saint Louis, MO 63124
*Corresponding author: mookiesub@gmail.com

Abstract
Lung Cancer is the leading cause of cancer-related death in the United States; yet despite this fact, common misconceptions hinder progress towards a better understanding of this fatal carcinoma. It is divided into two crudely differentiated sub-types which are distinguished by the appearance of their respective cells under a microscope: Small-Cell (SCLC) and Non-Small-Cell Lung Carcinomas (NSCLC). In order to find possible genetic perturbations causal for or associated with these diseases, scientists habitually utilize the voluminous capacity of microarray technology, which harnesses the process by which single-stranded DNA, derived from mRNA of differentially expressed genes, hybridized to their complementary probes immobilized on a chip. This genome-wide data is provided to bioinformatics laboratories for analysis, the locus of this investigation. We started with expression microarray data of Adenocarcinoma, Large Cell Carcinoma and Bronchioalveolar Carcinoma (sub-types of NSCLC) tissue from 149 patients. After performing Quality Control Checks to ensure that the data was entirely free of experimental biases, such as the Batch Effect, we proceeded to conduct differential analysis testing on 120 samples from patients with Adenocarcinoma, Large-Cell Carcinoma and Bronchioalveolar Carcinoma: sub-types of NSCLC. Data from these sub-types were put through the Significance Analysis of Microarrays (SAM) Test, to obtain a list of genes that were more significantly expressed in a certain NSCLC sub-types in comparison to its counterparts. The False Discovery Rate (FDR) of significant genes was set to 5% overall or average local if needed. We also conducted a 0% FDRSAM Test in order to identify potential NSCLC Biomarkers. We plugged the filtered 5% FDR differentially-expressed gene set into the GRANITE tool which draws Gene Relational Networks and identifies enriched pathways in a set of genes. After identifying these pathways, the most enriched genes in these pathways were plugged into the DrugBank Database in order to isolate specific drugs that would stall a gene, inhibit a rogue pathway or potentially thwart the cancer itself. The resulting genes in the 0% FDR data set could be used in the diagnosis of lung cancer as potential biomarkers of the disease. Hopefully, these beneficial results will pave the way for potential personalized targeted drug therapies to Lung Carcinomas.

Introduction
Lung Cancer is the leading cause of cancer-related death in the United States; in fact the vicious carcinoma claimed the lives of a stunning 158,683 out of 203,536 lung cancer patients in 2007. The scale of lung cancer death is so colossal, that in 2009, a greater number of lung-cancer-diagnosed patients died than those who died of breast, colon, pancreas and prostate cancers combined. It is projected that, by 2015, the worldwide number of Lung Cancer deaths alone will amount to a staggering 1,676,000. The survival rate for patients with lung cancer is a mere 15% in 5 years: the lowest survival rate of any other cancer. Yet current methods used to “treat” lung carcinomas, specifically chemotherapy, have been largely fruitless in their attempts to extend the lives of lung-cancer diagnosed patients. A recent study demonstrated that chemotherapy only contributes to the 5-year survival of 2.3% of all cancer cases. In skin cancer research, another study showed that vemurafenib, a drug that specifically targets a gene mutation rather than an entire pathway (chemotherapy’s ineffective method), has proven to be 8 times more effective than chemotherapy in shrinking tumors in patients. It seems that the identification of targeted therapies such as this one will lead to a new dawn in treatment. However, in the United States, lung cancer receives just $1,200 of federal funding per death, while breast cancer receives more than $27,000 per death, followed by $14,000 for prostate cancer and $6,500 for colon cancer. This lack of funding leads to a subsequent deprivation of research, which has starved the carcinoma of potential targeted therapies. These stunning figures provided the motivation for us to pursue research in this cutting-edge field of oncology.

To begin research however, it was imperative that we truly understood Lung Cancer; contrary to common misinterpretations, lung cancer is actually a heterogeneous assortment of tumors of the lung, bronchus, alveoli and pleura. It is divided into two crudely differentiated sub-types known as Small-Cell and Non-Small-Cell Lung Carcinoma, indicative of what these respective cells look like under the scrutiny of a microscope. Of these Carcinomas, Non-Small Cell Lung Carcinomas, the focus of this study, represent the overwhelming majority of diagnoses at approximately 80.4%. Non-Small Cell Lung Carcinoma (NSCLC) is split into even more definitive sub-types: Adenocarcinoma, the most common form of all lung cancers, Squamous Cell Lung Carcinoma, Large-Cell Lung Carcinoma, and Bronchioalveolar Carcinoma, which, in this study, is considered to have its own explicit NSCLC Histology. These sub-types are widely accepted by the scientific community and provide a basis for oncologists to diagnose and treat with greater accuracy and specificity, due to differences in tumor appearance, location and formation.

The need for better analysis of tumors on a genome-wide
scale for specific genetic analysis has, as a result, instigated the increasingly habitual application of high-throughput screening assays by researchers, including DNA microarray technology. The voluminous capacity of this highly advanced technology enables it to analyze over 50,000 genes at once. The microarray is essentially a small chip with countless depressions upon it, each of which contains picomoles of various DNA sequences that encompass the entire human genome. Genes from samples are able to hybridize to their respective DNA counterpart sequences, permitting scientists to identify which genes are expressed more often in healthy patients and which genes are more active in cancer-diagnosed ones. Yet the true essence of microarray data is numbers. Each of these numbers defines levels of gene expression in cancer – the higher the number, the more the gene is expressed in cancer and vice-versa. Of course for a small data set, such tests could be easily done by hand. However the problem remains that one would have to painstakingly analyze 30,000 genes for each of the 149 patients in this study – no simple task by hand. This is where this project’s computer testing is absolutely imperative, for it effectively finds the genes that have the highest expression levels in a set of samples. After microarray testing, data is provided to bioinformatics laboratories for data analysis. 

This study, an application of advanced biomedical informatics, attempts to determine significant differentially-expressed genes in microarray data in order to ameliorate cancer diagnosis and prognosis of three of the four NSCLC sub-types: Adenocarcinoma, Large-Cell Carcinoma (LCC), and Bronchioalveolar Carcinoma (BAC).

**Materials and Methods**

To begin, we compiled microarray data from a wide variety of lung cancer patients, courtesy of the Washington University Department of Pathology and Immunology. Originally, we amassed a cohort of 149 lung cancer patients diagnosed with a variety of NSCLC Sub-Types and even further definitive classifications within these sub-types. This dataset was also a compilation of data from patients whose tests were conducted on two different Affymetrix Microarray Probe Platforms: Hu133+2 and Hu95AV2. Using an Integrated Genomics Suite called Partek\(^2\), we endeavored to ensure that we used only the most accurate samples in our analysis. Thus, we also put the dataset through several statistical quality control checks. We used a Principal Component Analysis (PCA) Plot at this stage in the quality control checking – a useful statistical technique used to find patterns in datasets of high dimension, which we planned to use to find patterns within the microarray data of the various NSCLC sub-types to find relative similarity but some difference amongst the data obtained from patients diagnosed with the aforementioned types of carcinomas. Subsequently, in order to identify the significantly differentially-expressed genes distinguishing each NSCLC histological category, we surmised that the Significance Analysis of Microarrays (SAM) Test\(^1\) would be the best option in terms of the determination of these gene sets. We proceeded to put the data obtained after the Quality Control Checks into an Excel 10 File in order to utilize SAM. We used a one-versus-all testing method in which we compared Adenocarcinoma to the other types; LCC to the other sub-types and so on. The purpose of this was to identify genes that were more significantly expressed in these specific carcinomas in relative comparison to their counterparts. We proceeded to conduct the SAM Test using a set number of 100 algorithms in each instance. We first, however, desired to set a specific False Discovery Rate (FDR) – in essence, a statistical measure which accounts for the fact that some of the genes identified by the test could be falsely identified (false positives) – that we wanted the SAM Test to output. So, based on a study done by Harvard, Stanford and UNC-Chapel-Hill researchers on Adenocarcinoma using SAM, we concluded that a 5% FDR would suffice to eliminate error, but still obtain a large list of cancer-causing genes. If less than 200 genes were selected, the conditions were iteratively softened until a dataset of the desired size was acquired. Three SAM Tests were conducted in this manner with an overall 5% false discovery rate. Although the Adenocarcinoma vs. All and Large Cell Lung Carcinoma vs. All SAM Tests yielded reasonable results, the BAC vs. All Test did not dispense as desirable an outcome. With an output of a mere 5 genes, we were forced to relax the restrictions on the SAM parameters to a 20.84% FDR. However, this was only the overall FDR, and the average of the obtained local FDRs for the significant genes in BAC samples was roughly 5.69%. The average local FDR took into account only the genes of interest, rather than allowing the underexpressed genes to sway the FDR to greater levels. Thus, using the local FDR in this instance is permissible and would yield better results. For the Adenocarcinoma and LCC Samples’ Tests the FDR was 5.68% and 5.34% respectively. The size of the resulting significant gene sets for Adenocarcinoma, LCC and BAC respectively were: 3589, 2821, and 204 genes. We also conducted three other SAM Tests in the same manner, but with a 0% FDR, in order to identify an explicit set of significant genes that would be instrumental in cancer diagnosis. In addition to identifying specific sets of genes unique to/enriched in each NSCLC sub-type, we utilized a novel application created by Jahangheer Shaik, one of the major contributors to this project, called the Gene RelAtional Network of InTeracting Elements (GRANITE)\(^3\), which assists Bioinformaticians and Systems Biologists in depicting Gene Relational Networks (GRNs); GRNs depict functional relationships among genes, and are extremely complex, involving multiple genes and gene products operating at multiple levels. GRANITE provides a user-friendly platform on which data from several different databases can be integrated onto one site. The GRANITE tool allows one to assess which biological pathways in the human body are enriched in a certain set of genes, as well as the transcription factors inhibiting certain genes in any given dataset. Before conducting the GRANITE Tests we determined a list of approximately 36 pathways that we believed were important in the process of Lung Cancer, after researching online at the four premier Pathway Databases: KEGG\(^4\), Reactome\(^5\), PANTHER\(^6\), and BioCarta. We conducted three different GRANITE tests by plugging in the enriched dataset obtained in the SAM Test for each respective NSCLC Histology. After conducting the tests we determined which pathways had the largest gene relational network and/or the largest number of genes for each NSCLC sub-class and determined the top three in each. If there were any ties in the number of GRNs in the pathways given, all of the tied pathways were listed. The final test we conducted was the
Results

Quality Control Check Results: When, at first, we viewed the PCA Plot of the data on Partek, as shown in Figure 1a, we were convinced that something was amiss. The data, despite the fact that all of it concerned Lung Cancer, was extremely definitive as it was split into two explicit groupings within the cohort of patients with the same disease. We then proceeded to label all of our data with proper probes mentioned earlier: Hu133+2 and Hu95AV2.

The results of this Probe Mapping are shown in Figure 1b. As is clearly visible the notorious Batch Effect is taking place in our data. The Batch Effect, one of the great drawbacks to microarray technology, occurs when microarray tests are conducted at different times or in different conditions, leading to a lack of comparability between the datasets. As our data contained the older Hu95AV2 and the newer Hu133+2 probes, the Batch Effect clearly grouped the two probe types together in a non-biologically-related, biased method. In order to rid the data of the batch effect, we normalized the data using the Quantile Normalization Process, a statistical technique for making two distributions identical in statistical properties, and obtained a Box-and-Whiskers Plot of the Data shown in Figure 1c. Satisfied that the normalization had gone smoothly, we then proceeded to get rid of the outliers.

However, the removal of the Batch Effect using Partek left no clear method by which to identify patterns in a PCA Plot, as shown in Figure 1d, we at last reasoned that we ought to remove the older Probe Type: the Hu95AV2 Probe and conduct our testing on the newer Hu133+2 Probe Type Data only. In the figure, it is impossible to find any subtle patterns which will help us slightly distinguish the various sample types. This “distinguishability” is an essential facet to any good dataset. Since there were only 25 samples conducted on the Hu95AV2 microarray probe-type, we chose to eliminate these probes to allow for comparability between the remaining samples without the Batch Effect or other errors, and since this older probe type is more prone to error anyway. After eliminating the Hu95AV2 Probe Data from our file and obtaining the PCA plot for the remaining Hu133+2 data, we eliminated the outliers and obtained the final PCA plot shown in Figure 1e. This excellent plot is not too distinct, yet it is explicit enough to be able to differentiate
between the sub-types, showing that our data is largely accurate and sufficient to test on. After completing Quality Control Checking there were a total of 120 microarray patient samples remaining (all Hu133+2); in total there were, 24 LCC samples, 13 BAC samples and 83 Adenocarcinoma samples. The successful completion of these cursory tests allowed the study to proceed using the final dataset from the Quality Control Checks.

**SAM Results:** After conducting the following 5% FDR SAM Tests – Adenocarcinoma vs. All, LCC vs. All and BAC vs. All — we obtained the plots shown in Figures 2a, 2b, and 2c in the same respective order. The significant genes obtained after the SAM Test were subsequently plugged in to the GRANITE tool in order to identify genes involved in cancerous pathways.

**GRANITE Results:** In Adenocarcinoma we discovered that the Janus Kinase-Signal Transducer and Activator of Transcription or Jak-STAT Signaling Pathway had the highest number of GRNs, followed by a three-way tie for second between Angiogenesis, mTOR Signaling, and Regulation of the Actin Cytoskeleton, and succeeded by Tight Junction in 3rd Place. However, in an odd turn of events the Regulation of the Actin Cytoskeleton Pathway possessed the largest number of genes, with Jak-STAT in tow and the remainder of the pathways following in the very same sequence. In Large-Cell Carcinoma, Jak-STAT was once again found as the pathway with the greatest amount of GRNs; however the remaining pathways did not at all resemble the results obtained in the Adenocarcinoma GRANITE test. After Jak-STAT, Wnt Signaling was found as the second largest pathway in terms of enriched genes followed by the EGF-Receptor (EGFR) Interaction Pathway in third. Lastly, for Bronchioalveolar Carcinoma, none of the pathways we ascertained to be significant in tumorigenesis and the progression of Lung Cancer were present, except one pathway: Wnt Signaling, which had quite an impressively large GRN for such a comparatively small dataset. Observers of this study must note that overly general and ambiguous pathways that could be linked to a plethora of different diseases were not used or analyzed as a part of this research. Two notable examples would be the MAP-Kinase (MAPK) Signaling Pathway and the well-known pathway calling for programmed Cell Death, Apoptosis.

The Results of GRANITE in terms of the enriched genes are displayed in the tables shown in Figures 3a, 3b and 3c, in the following respective order: BAC, Adenocarcinoma and LCC.

The GRANITE results of this study are consistent with some of the most highly-respected scientific literature currently in the field. We noticed that many of the genes involved in Adenocarcinoma were identified in both this study and others by Tongji University[1], and the University of Michigan[2]. Significant genes involved in LCC match many genes found in a study presented at the famous Seminar in Cell Developmental Biology[7]. Lastly, many significant genes corresponding to BAC in this...
study are consistent with those found in a North Shore University Hospital investigation\textsuperscript{25} and a University of California study\textsuperscript{18}. In addition, the aforementioned investigation conducted by Harvard University, Stanford University, and the University of North Carolina Chapel Hill\textsuperscript{7} also confirms several of our GRANITE-identified genes. This consistency with top literature clearly establishes the benefits of the methods utilized in our study and promises much.

Using the resulting Gene List obtained from the GRANITE Test, we proceeded to identify drugs which would halt the carcinogenic impacts of the significant genes involved on cancer-related pathways.

**DrugBank Results:** Overall, we obtained 21 drugs that inhibited some of the rogue pathways which aided lung tumor cells in tumorigenesis and tumor maintenance.

For Adenocarcinoma, we discovered that the following drugs would inhibit a certain mutant gene in their respective corresponding pathway: the drugs Adenovite, Adexphos and Azucaps all inhibited the PRKAA1 Gene involved in the mTOR Signaling Pathway, Sprycel impeded the STAT5B Gene involved in the Jak-STAT Signaling Pathway, Certican and Torisel each repressed the FRAP1 Gene in the mTOR Signaling Pathway, and lastly, Hydroxyfasudil suppressed the ROCK1 Gene in the Regulation of the Actin Cytoskeleton Pathway. These results are shown in Figure 4b.

For Large Cell Carcinoma, we found that the following drugs – Pegasys, Wellferon, Intron A, A-vaferon, Betaseron, Roferon A, Alferon and PEG-Intron – all inhibited the IFNAR2 Gene involved in the Jak-STAT Signaling Pathway, Actimmune impeded the IFNGR1 Gene involved in the Jak-STAT Signaling Pathway, Adagen repressed the GRB2 Gene in the Jak-STAT Signaling Pathway, Aerolone checked the PIK3R1 Gene involved in the Jak-STAT Signaling Pathway, Sprycel constrained the STAT5B Gene in the Jak-STAT Signaling Pathway and lastly, Eskalith suppressed the CTNNB1 Gene in the Wnt Signaling Pathway. These results are shown in Figure 4c.

Finally, for Bronchioalveolar Carcinoma, we ascertained the following drug as a proper inhibitor: the drug Velcade stalled the PSMB2 Gene in the Wnt Signaling Pathway. These results are shown in Figure 4a.
Discussion

According to the 2010 Report of the American Thoracic Society, early diagnosis plays a crucial role in prognosis in lung cancer and would be greatly supported by the discovery of a simple marker of lung cancer. Pinpointing genetic predisposition to lung cancer could augment the efficacy of tumor screening. This may be related to deviations within genes that are associated with constraining cell growth and other similar biological pathways involved in the potential creation and maintenance of lung tumors. Gene expression profiling experiments have indicated that biomarkers of this nature may, in the future, be valuable in the diagnosis of lung cancer.

For the 0% FDR SAM Test, we obtained a list of 135 Genes in Adenocarcinoma, and 45 Genes at the same FDR in LCC. However at the 0% FDR for BAC we only obtained one significant gene, so we suggest that those interested in this study use the original BAC 5% Average Local FDR Dataset in Diagnosis. This list of significant genes could be useful in the advancement and amelioration of early Lung Cancer identification, as these genes are the very same biomarkers which are imperative to the betterment of diagnosis. The clinical relevance of these genes could be promising to future oncologists, as they could identify whether a patient has a form of NSCLC in a faster and more reliable way, enhancing the chances of prevention in the early stages.

This study not only improves diagnosis, but lengthens prognosis as well. The drugs targeting the rogue pathways that we have identified through the GRANITE and DrugBank Tests could be future therapeutic substances taken by patients in order to ease the burden of the disease and potentially rid them of it. If a patient is diagnosed with Adenocarcinoma, for example he or she could be put on one of the identified drugs in the following list: Adenovite, Adephos, Azuscaps, Spryce, Certican, Torisel or Hydroxysufadil. These drugs would essentially inhibit certain genes in a rogue pathway and hopefully stop the pathway from aiding the cancerous tumor. For example, Azuscaps would stall the functioning of the Jak-STAT Signaling Pathway, a crucial participant in Non-Small-Cell Lung Cancer due to the fact that it is responsible for antiapoptosis.

Admittedly there could be flaws in the research. Our data did not include the Squamous Cell Carcinoma sub-type of NSCLC and it used only 100 algorithms in the SAM Test due to computer capabilities as opposed to the maximum of 2000 to ensure near-perfect accuracy. Or, perhaps, the Quality Control Checks did not go as planned. Whatever the issue, this study still proposes a promising idea for the future, and is accurate enough despite these setbacks.

In conclusion, the results of this investigation should be looked into by drug testers to determine the effectiveness of the genes we identified on lung-cancer-diagnosed patients. Hopefully this will aid doctors in making headway in the translation of the identified genes and drugs into proper cancer treatment.

Figure 2c. SAM Plot of BAC samples indicating a large number of samples underepressed in comparison to other sub-types and few overexpressed genes at the normal FDR. Although the desired 200 genes were found at an overall FDR of 20.84%, the average local FDR for the overexpressed genes were acceptable under the scrutiny of this study – 5.69%.

Figure 3a. BAC Results from the GRANITE Test, showing a large number of matches in the Wnt Signaling Pathway, a major pathway involved in lung cancer.

Figure 3b. Adenocarcinoma Results from the GRANITE Test showing gene correlation with numerous pathways involved in Lung Cancer according to major pathway databases.

Figure 3c. LCC Results from the GRANITE Test showing some similarity with Adenocarcinoma in the Jak-STAT Signaling Pathway, but impacting vastly different genes in comparison. These results also are consistent with many of the premier studies of this era in terms of Pathway correlation with Lung Cancer.

<table>
<thead>
<tr>
<th>Adenocarcinoma</th>
<th>Jak-STAT</th>
<th>Angiogenesis</th>
<th>mTOR</th>
<th>Aktin</th>
<th>Tight Junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIK3B</td>
<td>PLOCK1</td>
<td>FRA1P1</td>
<td>PIK3B</td>
<td>CLDN15</td>
<td></td>
</tr>
<tr>
<td>AKT2</td>
<td>PIK3B</td>
<td>AKT2</td>
<td>GSN</td>
<td>CLDN13</td>
<td></td>
</tr>
<tr>
<td>SRC4</td>
<td>AKT3</td>
<td>PIK3B</td>
<td>PIK3B</td>
<td>PIP2B6B</td>
<td></td>
</tr>
<tr>
<td>JAK3</td>
<td>PIK3C</td>
<td>PIK3A4</td>
<td>PIK3C</td>
<td>PIK3C</td>
<td></td>
</tr>
<tr>
<td>SOCS7</td>
<td>PIK3D</td>
<td>ULK2</td>
<td>PIK3C</td>
<td>PIP2B6B</td>
<td></td>
</tr>
<tr>
<td>LIP2</td>
<td>PIK3E</td>
<td>TSC2</td>
<td>PIK3C</td>
<td>PIP2B6B</td>
<td></td>
</tr>
<tr>
<td>AKT3</td>
<td>PIK3F</td>
<td>PIK3G4</td>
<td>PIK3C</td>
<td>PIP2B6B</td>
<td></td>
</tr>
<tr>
<td>STAT5B</td>
<td>PLOCK1</td>
<td>PIK3B</td>
<td>PIK3C</td>
<td>INAD1</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>BRS3</td>
<td>TAK1</td>
<td>PIK3C</td>
<td>INAD1</td>
<td></td>
</tr>
<tr>
<td>PIK4</td>
<td>AKT2</td>
<td>VAV2</td>
<td>PIK3C</td>
<td>INAD1</td>
<td></td>
</tr>
<tr>
<td>PIK5</td>
<td>PIK3C</td>
<td>PIK3G</td>
<td>AMHREF2</td>
<td>AMHREF2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Large Carcinoma</th>
<th>Cell</th>
<th>Jak-STAT</th>
<th>Wnt Signaling</th>
<th>EGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSMR</td>
<td>PMSC2</td>
<td>CBL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL6ST</td>
<td>PPSMB4</td>
<td>GRB2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JAK1</td>
<td>APC</td>
<td>PTPN11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL4R</td>
<td>BTRC</td>
<td>PAG1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNGR1</td>
<td>CTNNB1</td>
<td>KRAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFNAR2</td>
<td>SKP1</td>
<td>SHC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBL</td>
<td>PSMB7</td>
<td>PIK3R1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBL</td>
<td>CSNK1A1</td>
<td>SPRY1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTHR1</td>
<td>SPRY2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIAS1</td>
<td>STAT3</td>
<td>GSH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT5B</td>
<td>CCND3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bronchioalveolar Carcinoma</th>
<th>Wnt Signaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSMA4</td>
<td></td>
</tr>
<tr>
<td>PSMB2</td>
<td></td>
</tr>
<tr>
<td>PSMA1</td>
<td></td>
</tr>
<tr>
<td>PSMA2</td>
<td></td>
</tr>
<tr>
<td>PSMD12</td>
<td></td>
</tr>
<tr>
<td>PSMD14</td>
<td></td>
</tr>
<tr>
<td>PSMA6</td>
<td></td>
</tr>
</tbody>
</table>
References


identified by bioinformatics analysis of public expression data. Cancer Research, 67(15), 7431-7438. AMER ASSOC CANCER RESEARCH.


A Comparison of Drosophila CMI to Human MLL2/ALR Reveals Homologous Histone Binding and Recognition Preferences

Riva Trivedi¹, Claudia Zraly², and Andrew Dingwall*²

Student¹: The Illinois Mathematics and Science Academy, 1500 Sullivan Road, Aurora, IL 60506
Mentor²: Loyola University Stritch School of Medicine, 2160 South 1st Avenue, Maywood, IL 60153
*Corresponding author: adingwall@lumc.edu

Abstract
The MLL/ALR family consists of enormous multi-domain proteins found in large co-activator complexes (called COMPASS-like complexes) involved in nuclear receptor dependent gene transcription. These complexes activate gene transcription through binding and covalent modification of histone residues. Mutational loss of the human MLL/ALR genes has been implicated in developmental disorders and cancers. The Drosophila MLL/ALR homolog is a single representative split into two genes during evolution, known as cmi and trr, with each encoding for conserved portions essential for transcription regulation. To further define the functional relationship between cmi and trr, we used in vivo knock-down experiments using conditional shRNAi transgenes. We found depletion in cmi and trr gene levels greatly affected development of Drosophila, and resulted in lowered global histone lysine methylation. Different “loss of function” phenotypes were also noted, such as defects in wing vein development. To determine whether the fly and human MLL/ALR share similar histone binding properties, we cloned the human ALR/MLL2 PHD3 finger by PCR using reverse transcribed mRNA, performed protein expression analysis, and determined histone binding preference using whole genomic chromatin and histone peptide pulldown analyses. We found the human MLL/ALR genes and the Drosophila homologs share similar histone binding and recognition preferences.

Introduction
Cancers are caused by gene mutations followed by uncontrolled cell growth, with the rate of cell growth determining whether a cancer is benign or malignant/metastatic. Cancers are formed based on different environmental and genetic factors. This explains why even after countless years of research in effort to developing a “cure”, there is no single remedy that can be discovered to solve every problem. However, investigating the potential causes and effects that cause cancer to form can make it possible to learn more about the issue. This can be achieved by examining distinctive genetic effects in offspring based on mutations and modifications in their DNA. In this project, genes particularly known for forming cancerous diseases were studied.

The genetic material of a eukaryotic cell, known as DNA, is contained in the nucleus in structures called chromosomes. Because they are so long, chromosomal DNA strands must be compacted, twisted several times, and wrapped around histone octamers known as nucleosomes. These nucleosomes also twist into larger structures of chromosomes. The chromosomal DNA, along with proteins that package it in the nucleus of the cell, are often referred to as chromatin. Chromatin’s function is to stabilize and compact the DNA, prevent DNA damage, as well as regulate gene expression and DNA replication¹.

Histones have specific amino acid residues which can be covalently modified through acetylation, methylation, and phosphorylation, known as the “histone code”, which can lead to different effects in gene regulation². Specifically, the methylation of Histone H3 on lysine four (H3K4) can affect ‘epigenetic memory’. The epigenetic marks are placed onto the histone tail residues (e.g., trimethylated lysine 4 or H3K4me3) as a consequence of transcription and are sometimes heritable through successive cell generations³.

In humans, the paralogous ALR/MLL2 and HALR/MLL3 proteins are found as core components of COMPASS-like transcription coactivator complexes⁴. MLL2 and MLL3 are enormous (both are encoded by genes nearly 15,000 base pairs in length) and contain multiple functional domains, including a methyltransferase enzyme which puts methyl groups distinctively onto H3K4, nuclear receptor binding motifs and regions known as PHD (plant homeodomain) zinc fingers that are thought to mediate protein-protein interactions. The histone lysine methylation appears to generally correlate positively with cell growth and endurance and is associated with active gene transcription⁵. In addition to the enzymatic modification of histones, the MLL2 and MLL3 proteins also contain as many as 7 clustered PHD finger domains within the N-terminal portion. These PHD finger domains are each small ~60 amino acid portions of the protein that in some cases have been shown to recognize and bind to specific histone tail modifications. The single Drosophila MLL2/MLL3 counterpart is split into two proteins known as Cara Mitad (CMI) and Trithorax-related (TRR). The CMI protein is highly related to the N-terminal portion, while the TRR protein is related to the C-terminal portion of both MLL2 and MLL3⁶. CMI and TRR have been found in the same COMPASS-like complex in Drosophila. The CMI protein in Drosophila directly corresponds to the part of MLL2/MLL3 containing these PHD finger domains. In fact, the specific type 3 fingers (PHD3) in humans and Drosophila have a very close conservation of amino acid sequences⁷.

The Drosophila cmi and trr genes are essential for organismal development, with mutant phenotypes reminiscent of those associated with loss of hormone signaling⁸. It is already understood that the mammalian MLL2 and MLL3 genes both encode for H3K4 methyltransferases which function as epigenetic transcriptional activators during the fetal developmental period⁹. Recent studies have illustrated the effects of mutations in the
human MLL2 and MLL3 and murine MLL3 genes, with connections to dramatic changes in early development, cell survival, and diseases such as cancer. In the past several years, the genomes from patients with a variety of cancers have been sequenced to identify mutations correlated with disease. These studies have identified a significant number of somatic mutations in the MLL2 and MLL3 genes associated with non-Hodgkin lymphoma11. A similar study on patients with transitional cell carcinoma (the most common form of bladder cancer) also found a large number of mutations in chromatin remodeling genes, including MLL212. A rapidly growing number of publications have reported loss of MLL2 or MLL3 in association with childhood brain cancers (medulloblastomas), kidney, prostate and breast cancers, as well as pancreatic and gastric cancers. Additionally, other studies suggest Kabuki syndrome, a developmental disorder known to be linked to skeletal and cognitive defects, plus cardiac and immunological disorders, are also caused by mutations in MLL213,14. These studies have dramatically increased our understanding of how misregulation of histone modifications has caused great maladies that could lead to human cancer. Factors which moderate the additions, eliminations, and readings of modifications are causing this misregulation15.

The strong similarity between the human MLL2 and MLL3 and fruit fly CMI proteins is strong evidence that they are functional homologs. In order to examine this directly, we wished to test the hypothesis that the conserved PHDf3 domains would recognize and bind similar histone tail modifications. We first cloned the MLL2 PHDf3 region and then characterized the histone binding preference of both the MLL2 and CMI PHDf3 fingers. Due to its easily manipulative genetic tools, the fruit fly Drosophila melanogaster is a wonderful model system for investigating the effects of histone modifications during development. Using the Drosophila genetic system, we investigated the role of CMI in the regulation of epigenetic modifications. We individually depleted CMI and TRR in Drosophila using conditional expression of the Drosophila.

Results
As a first step in determining whether the human MLL2 and Drosophila CMI proteins are functionally related, we cloned and expressed a portion of the MLL2 protein that carried the conserved PHDf3 region. This PHD domain was predicted to recognize and bind to specifically modified histone residues based on restriction sites (BamHI and EcoRI) that allowed for cloning and manipulation. Amplification was performed using a proofreading polymerase to minimize the possibility of amplification errors. The PCR product was then cloned into pBlueScript II SK (Fermentas, Inc.) and sequenced. We created a bacterial fusion with the finger and glutathione-S-transferase (GST) in the vector pGEX4T (Amersham, Inc.) using the engineered restriction sites (Figure 1A). The GST:PHDf3 protein was estimated to be 32.3 kilodaltons (kD). In order to test for inducible expression of the fusion protein, bacteria containing the GST fusion plasmid were grown at 30°C, induced with IPTG, electrophoresed, and then run on 10% SDS PAGE protein gels for 75-90 minutes at 125 volts. Proteins were visualized by staining with coomassie blue. A 0-4 hour time course was run to measure induction (Figure 1B). In order to isolate the soluble protein fusion, proteins were extracted using the B-Per reagent (Pierce/Thermo-Fisher Scientific, Inc.) and soluble (Sol) and insoluble (Insol) protein was collected. The soluble GST:PHD fusion was purified away from other contaminating proteins using glutathione agarose resin. Soluble bacterial extract was added to Glutathione-agarose (Glu-agarose) resin (Sigma) and the GST:PHDf3 protein was eluted from the resin using free glutathione (Sigma). Varying amounts of the purified protein were examined by 10% SDS-PAGE gel analysis for 75-90 minutes at 125 volts (Figure 1C). Histone binding analysis. Native histones were prepared from either Drosophila or human cultured cells as described and incubated with similar amounts of the CMI7 or MLL2 PHDf3 fusion proteins. Bound histone was analyzed by immunoblot with αH3 antibodies (Abcam). A Western blot was run on the purified protein to illustrate it bound to Histone 3, ensuring its functionality (Figure 2 and Chauhan et al., 2012). We used modified histone peptide arrays to test binding preferences of both this PHDf3 as well as that of Drosophila17. Briefly, biotinylated histone peptides (Millipore, Inc.) were incubated with similar amounts of GST-CMI PHDf3.b, GST-CMI PHDf3.b (W680A-mutant version that is predicted to not bind histones), GST-ALR/MLL2 and GST-LID (control PHDf3). Bound GST fusions were analyzed by immunoblot with αGST antibodies (Abcam). Genetic studies in Drosophila. We used the ‘Driver-Responder’ system for gene-specific depletion frequently used in Drosophila melanogaster fruit flies, to address the role of certain genes in the regulation of epigenetic modifications16,17,18. Genetic combinations were made by using an engrailed GAL4 driver which activated a shRNAi (obtained from the Bloomington Drosophila Stock Center and the Vienna Drosophila Stock Center) to deplete certain genes. The driver serves as a transcription factor that “drives” the expression of the transgene carrying the short hairpin in a certain time and location within the developing fly. In our case, we depleted cmi and trr to observe the effects of losing either the PHD finger or the methyltransferase in the progeny of the Drosophila.

Materials and Methods
Cloning of the MLL2 PHDf3 region, bacterial expression and purification. The MLL2 PHDf3 region was cloned and expressed as a bacterial glutathione-S (GST) fusion protein. Due to the presence of an intron within the genomic DNA encoding for the PHDf3, we obtained complementary DNA (cDNA) from NIH293T human cell line. Specific Polymerase Chain Reaction (PCR) primers were used to amplify and isolate the specific finger region (amino acids 1503-1560). The primers were designed to include a portion of the PHD finger sequence with sequences containing unique

rationale to share analogous properties, or they could be unlike each other. We hypothesized the structures and mechanisms would be similar due to their highly conserved sequences.
The focus of this project was how the structures and mechanisms for recognition of specifically modified histones are similar and/or different between the Drosophila CMI and human MLL2/3 PHD3 fingers. We hypothesized both structures and mechanisms would share similar properties due to their highly conserved sequences. As both the CMI and the MLL2/ALR PHDf3 proteins bound to similarly modified H3K4, we can conclude they have comparable histone binding and recognition preferences. Additionally, by observing the vein patterning defects in the fruit flies, we could conclude that in vivo depletion of cmi and trr gene levels had a significantly negative effect on the development of Drosophila, and resulted in a reduction of global histone lysine methylation. Also, the “loss of function” phenotypes add to our conclusion that depleting these genes leads to developmental disorders.

We next examined the in vivo phenotypic consequences of depleting cmi, as well as its conserved methyltransferase partner trr. We found that conditional depletion of cmi and trr gene levels specifically in the developing wing tissue caused truncations in certain wing veins (Figure 3). Furthermore, several additional “loss of function” phenotypes were noted in the Drosophila, such as held-out wings and no flight capability (Table 1).

Figure 3. Depletion of CMI and TRR leads to specific vein pattern defects. A) Wild type Drosophila wing. Shown are the positions of the normal veins, including the longitudinal (L1-L5), posterior crossvein (pcv) and anterior crossvein (acv). The photos on the left were obtained at 63X magnification. Panels on the right at 100X views of the same wings. B) CMI depleted wing with an incomplete L5 vein (arrow in right panel). The arrow in the left panel represents the boundary between the anterior portion of the wing (top) and posterior portion (bottom). C) TRR depleted wing with an incomplete L5 vein, anterior cross wing, and posterior cross vein (arrows in right panels).

Discussion

The focus of this project was how the structures and mechanisms for recognition of specifically modified histones are similar and/or different between the Drosophila CMI and human MLL2/3 PHD3 fingers. We hypothesized both structures and mechanisms would share similar properties due to their highly conserved sequences. As both the CMI and the MLL2/ALR PHDf3 proteins bound to similarly modified H3K4, we can conclude they have comparable histone binding and recognition preferences. Additionally, by observing the vein patterning defects in the fruit flies, we could conclude that in vivo depletion of cmi and trr gene levels had a significantly negative effect on the development of Drosophila, and resulted in a reduction of global histone lysine methylation. Also, the “loss of function” phenotypes add to our conclusion that depleting these genes leads to developmental disorders.

As we know cmi and trr genetically interact in flies and depleting their gene products using RNAi leads to patterning defects and
a global reduction in the trimethylation of H3K4, we may infer cmi must be required for trr to function and activate its target genes. In other words, loss of the PHD finger affects the ability for the methyltransferase to function. This signifies rather than operating through independent parallel pathways to regulate gene expression, they seemingly function through the same hormone dependent pathway. Moreover, if the trimethylation of lysine 4 is normal when cmi is depleted, then cmi would not be required for trr to methylate target genes. However in this case, cmi function is necessary for trimethylation. This leads us to believe that CMI may be required to bring in trr function to the target genes. Therefore, the recognition of and binding to histones is essential in targeting specific genes. The methyltransferase is significant in histone regulation, and the PHD finger is evidently required for it to fulfill its function. Thus, mutations affecting these proteins could have harsh effects on growth and development if the correct genes are not targeted or are misregulated. This is supported by recent findings that several developmental disorders and cancers are associated with loss of MLL2 and MLL3.

Therefore, the data supports our hypothesis. It is important to note, since the CMI PHD3 finger is involved in recognition of certain histone tail peptides, and the MLL2 finger has the same recognition pattern, due to the high amino acid sequence conservation between the two, we can extrapolate information we examined from Drosophila developmental phenotypes and connect it to human diseases. Tests can be performed on Drosophila to further investigate the relationship of this PHD finger with the developmental maladies, potentially resulting in an expansion of our knowledge regarding growth disorders.

References


Acknowledgements

I would like to thank the IMSA Student Inquiry and Research Program for allowing me to pursue this opportunity. This work was supported by a grant from the National Science Foundation (MCB1122001).
Analysis of Acceptance of Grammatical Changes in Standard American English Reveals Differences by Age and Education

Sophie Lipson1*, Ileana Rios2, and Susan Behrens3
Student1, Teacher2: Trinity School, New York, NY 10024
Mentor3: Marymount Manhattan College: 221 East 71st Street, NY 10021 and Columbia Summer High School Program, New York, NY 10027
*Corresponding author: sophie.lipson13@trinityschoolnyc.org

Abstract
This study investigates the changing syntax of English and the acceptance of those changes by society. Participants were given a paragraph containing several grammatical mistakes, mistakes according to Standard American English, and were instructed to edit the paragraph. Results indicate that people under the age of 25 and people with less than a college education are more likely to miss grammatical errors. Additionally, certain mistakes were missed more often than others. These findings show that certain syntactic features of English are changing and have been accepted by younger generations and those with less education.

Introduction
Language is a tool of communication that continuously evolves: new words, phrases, and pronunciations emerge with increasing frequency. Despite the inevitability of variations, humans have a tendency to resist such changes. However, over time new language structures develop and become the new “normal.” This study determined people’s acceptance of prevalent grammatical features that, according to Standard American English, are incorrect. Standard American English is the dialect of English that is used in professional writing and is taught in American schools.

This study was partially inspired by the incorrect grammar frequently heard on reality TV. It is not uncommon to hear the word “good” where the speaker actually means “well.” This incorrect grammar is spoken by people of all status and is broadcast across America. It can be deduced that this replacement of adverbs with adjectives is becoming a part of “normal” English. This study was conducted in an effort to assess trends regarding this grammatical error.

Prominent linguists have recently noted grammatical changes in the English language. Richard Lederer1 expressed his perspective regarding language change during a segment called “Language Pet Peeves” on National Public Radio (NPR). One aspect of language he addressed was adjectives vs. adverbs. Lederer stated, “I agree that the adverb is under siege. I think it is endangered in things like ‘he speaks real good,’ which is a double violation.”

Another feature of language he discussed was relative pronouns. Lederer said that they are not so simple-the word “that” was used to refer to people in the historic past but today we would use the word “who” in reference to a person. These are two grammatical aspects of English that were investigated in the current study.

Materials and Methods
This study focused on the following seven grammatical features:

1. Singular-Plural Antecedent Agreement
2. First Person Singular Pronoun Cases
3. Adjectives vs. Adverbs
4. Pronoun Cases in a Prepositional Phrase
5. Subject-Verb Agreement
6. Third Person Singular Pronoun Cases
7. Relative Pronouns

A paragraph was written containing seven grammatically correct sentences and seven grammatically incorrect sentences. Each grammatically incorrect sentence had only one of the seven
mistakes. The grammatically incorrect sentences were dispersed randomly throughout the paragraph.

Paragraph Given To Participants:

Last Saturday I spent time with my friends and we went to the park. After a few hours, we parted because everyone had their own plans for the afternoon. Sally and myself went to the zoo where we saw a seal. The seal swam up to the zookeeper looking for food. The seal's baby quickly followed, and the large group swam towards the edge of the tank. Sally was very enthusiastic, but the zookeeper said she spoke too loud. We remained quiet for the remainder of the show. However, between you and I, I really wanted to see the giraffes. Luckily, they were close by and I later fed them. I was surprised that the legs of a giraffe was so tall. Overall, the day ended on a high note and we left feeling good. Sally was happy to have seen the seals, and I the giraffes. Her and I decided we would return soon. I know that people that live in cities don’t go to the zoo very often, but it is really worth the trip!

To determine if certain mistakes have infiltrated commonly used language, twenty people were instructed to edit the paragraph. Randomly selected participants were selected from the New York and Boston area.

Instructions were loosely scripted and given to participants verbally by the author and one surrogate. The instructions were to “edit” the passage, with no mention of the purpose. Participants were given as long as they wished, but not spoken to while reviewing. Participants had to review the passage immediately and in the same room as the author or surrogate.

To determine trends, the participants were categorized by the following self-reported background information: 1. Age, 2. Gender, and 3. Education Level (Table 1).

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategories</th>
<th>Number of Informants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15-24</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>25-44</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>45-64</td>
<td>3</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>11</td>
</tr>
<tr>
<td>Education Level</td>
<td>High School</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>College Degree</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Post-College</td>
<td>4</td>
</tr>
</tbody>
</table>

Results

In analyzing the edited paragraphs, the overall number of mistakes that were missed was marked, and the specific type of mistake that was missed was also marked. This data was then sorted and graphed by grammatical feature, age, and education level.

Some mistakes were missed more frequently than others. As Figure 1 reveals, singular-plural antecedent agreement was missed by 80% of participants, and the replacement of adverbs with adjectives was missed by 60% of participants. On the other hand, subject-verb agreement was only missed by 15% of participants. This suggests that some aspects of English are changing, while others are not. It appears that the usage of the word “their” as a singular gender-neutral pronoun is prevalent among all participants tested. Additionally, participants generally accept adjectives in the place of adverbs, suggesting that adverbs are being used less frequently. However, most participants notice subject-verb disagreement, which may indicate that people still expect verbs to be properly conjugated.

Younger people may be more open to grammatical change. Figure 2 shows that participants aged 15-24 missed 67% of grammatical errors, more than any other age group.

There is a strong trend concerning education level and acceptance of syntactical change. As shown in Figure 3, people with a high-school education missed 64% of mistakes, while those with a college degree missed 32%, and people with a Post-College Professional Degree missed only 14%. Gender did not impact acceptance of grammatical change, as shown in Table 2.

Statistical analysis reveals that the impacts of age and education level are significant at the 0.05 significance level. With a p-value of 0.00125, the difference between average mistakes missed by people ages 15-24 and average mistakes missed by people ages 25 or older is statistically significant. While there is an observable trend that higher education level correlates negatively with missed mistakes, the only statistically significant finding is the difference between a high school education level and a Post-College Professional Degree (p-value = 0.0022).

Table 2. Average Grammatical Mistakes Missed By Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>38</td>
</tr>
<tr>
<td>Female</td>
<td>44</td>
</tr>
</tbody>
</table>

Discussion

Results of this study suggest that there are significant trends regarding the acceptance of structural change in the English language. People ages 15-24 and people with a high school education missed more grammatical mistakes than other groups. There is a clear overlap between the two categories, as the 15-18 year old participants were in high school. Additionally, people over the age of 25 did not miss many mistakes, correlating with a higher incidence of college education. (9 of the 10 participants over the age of 25 had completed college.) This suggests that there is a relationship between the level of education and age concerning grammatical correctness: the higher the education level, the less mistakes missed. Perhaps this is because not only have people with a higher education level learned more grammar in the classroom, but also because they are expected to use correct grammar in higher academic work. There was a minimal difference between men and women, as was expected. This shows that gender does not greatly affect grammatical acceptance.

When the types of grammatical errors missed are examined, it appears that the distinction between adverbs and adjectives is shrinking. Additionally, the use of the word “their” as a singular gender-neutral pronoun has become commonplace. Both of these grammatical changes show a simplification of current syntactical structure, supporting George Kingsley Zipf’s Principle of Least Effort, in which he asserts that humans by nature take the path of least resistance; in this case simplifying grammatical structure.

One limitation of this study is that it was conducted at one point in time, as opposed to repeating it periodically with the
same participants. Without longitudinal data, one cannot know if the younger participants would test differently as they aged and obtained a higher level of education. A second limitation is that human psychology is complex. Confidence, affected by age and education level, can affect performance.

This study warrants further research regarding grammatical change in the English language. A large-scale study in different English-speaking populations would allow for a more thorough assessment of the acceptance of grammatical changes; one could better isolate variables such as age and education level.

References


Figure 1. Percentage of Grammatical Mistakes Missed Overall.

Figure 2. Average Grammatical Mistakes Missed by Age.

Figure 3. Average Grammatical Mistakes Missed by Education Level.
Science is a great game.
The playing field is the Universe itself.