J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Editorial

Creating a pipeline of talent for informatics: STEM initiative for high school students in computer science, biology, and biomedical informatics

Joyeeta Dutta-Moscato[1], Vanathi Gopalakrishnan[1], Michael T. Lotze[2], Michael J. Becich[1]

[1]Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, [2]University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA

E-mail: *Joyeeta Dutta-Moscato - jod30@pitt.edu
*Corresponding author

Received: 02 Febrauary 14 Accepted: 03 Febrauary 14 Published: 28 March 14

This article may be cited as:
Dutta-Moscato J, Gopalakrishnan V, Lotze MT, Becich MJ. Creating a pipeline of talent for informatics: STEM initiative for high school students in computer science, biology, and biomedical informatics. J Pathol Inform 2014;5:12.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2014/5/1/12/129448

Abstract

This editorial provides insights into how informatics can attract highly trained students by involving them in science, technology, engineering, and math (STEM) training at the high school level and continuing to provide mentorship and research opportunities through the formative years of their education. Our central premise is that the trajectory necessary to be expert in the emergent fields in front of them requires acceleration at an early time point. Both pathology (and biomedical) informatics are new disciplines which would benefit from involvement by students at an early stage of their education. In 2009, Michael T Lotze MD, Kirsten Livesey (then a medical student, now a medical resident at University of Pittsburgh Medical Center (UPMC)), Richard Hersheberger, PhD (Currently, Dean at Roswell Park), and Megan Seippel, MS (the administrator) launched the University of Pittsburgh Cancer Institute (UPCI) Summer Academy to bring high school students for an 8 week summer academy focused on Cancer Biology. Initially, pathology and biomedical informatics were involved only in the classroom component of the UPCI Summer Academy. In 2011, due to popular interest, an informatics track called Computer Science, Biology and Biomedical Informatics (CoSBBI) was launched. CoSBBI currently acts as a feeder program for the undergraduate degree program in bioinformatics at the University of Pittsburgh, which is a joint degree offered by the Departments of Biology and Computer Science. We believe training in bioinformatics is the best foundation for students interested in future careers in pathology informatics or biomedical informatics. We describe our approach to the recruitment, training and research mentoring of high school students to create a pipeline of exceptionally well-trained applicants for both the disciplines of pathology informatics and biomedical informatics. We emphasize here how mentoring of high school students in pathology informatics and biomedical informatics will be critical to assuring their success as leaders in the era of big data and personalized medicine.

Key words: Bioinformatics, education, medical informatics, science, technology, engineering, and math education

Access this article online

Website: www.jpathinformatics.org

DOI: 10.4103/2153-3539.129448

Quick Response Code:

INTRODUCTION

Pathology is emerging from the Post-Genomics Era,[1] into the era of Personalized Medicine[2] and Big Data[3] and, as a result, the field will undergo a series of changes. We will redefine the way we engage in process sampling and analysis, requiring the use of interoperable imaging and molecular data. This will result in unprecedented need in information management, decision support and advanced analytics and herald the age of "computational pathology". Institutions with strong scientific leadership in biomedical informatics and computational and systems biology will be key to the role of Pathology Informatics emerging in computational pathology. Despite this very bright future for Pathology (and Biomedical) Informatics, getting the best and the brightest individuals to fill these critical positions is challenging. The need to introduce bioinformatics at the high school level is rapidly gaining recognition.[4] Therefore, to address this important need for informatics savvy trainees we created the Computer Science, Biology, and Biomedical Informatics (CoSBBI) Track in the Summer Academy in 2011. CoSBBI is our effort to begin "pipelining" the best and the brightest high school students to informatics through a science, technology, engineering, and math (STEM) oriented research academy as part of the University of Pittsburgh Cancer Institute (UPCI) International Academy (see http://www.upci.upmc.edu/summeracademy/).

There are many summer programs for high school students including those at Johns Hopkins University (http://cty.jhu.edu/summer/), Stanford (https://summerinstitutes.stanford.edu/), Northwestern University (http://osep.northwestern.edu/), the Pennsylvania Governor's School (see http://www.pgssalumni.org/about-pgss) as well as many others across the US. When we began the UPCI Summer Academy in 2009, it aimed to improve upon the existing summer programs for high school students by providing a mentored hands-on primary research experience, including the usual didactics associated with most other programs. This was challenging and required the enlistment of many faculty, post-docs, medical students, and laboratory personnel to provide concentrated mentorship. The major innovation in the UPCI Summer Academy was not only to do primary meritorious work with a mentor, but also have students present their work orally. This was done with faculty, staff, and families invited to be in the audience for the oral PowerPoint presentations. It also included a judged scientific poster session presented to the entire scientific community of the UPCI on the last day of the academy. The results of this effort have been remarkable and have changed the career aspirations of many of the students who have participated. In 2011, after 3 years of the Department of Biomedical Informatics' participation in the classroom component of the UPCI Summer Academy, we formed CoSBBI. Initially, CoSBBI was a partnership of the Departments of Biomedical Informatics (DBMI) and Computational and Systems Biology (CBS). Faculty from both departments provided mentored research experiences in computational biology, bioinformatics, biomedical informatics, and pathology informatics. It then 'fissioned', based on site changes, and gave rise to CoSBBI, more closely associated with the Department of Pathology and adjacent to the Hillman, and a new program oriented around drug discovery and systems biology for the past 3 years.

This editorial is a synopsis of that experience and encourages other programs to join us in this effort to create a pipeline of highest quality, research savvy trainees to our programs in pathology (and biomedical) informatics to ensure the successful integration of pathology (and computational pathologists) into the era of personalized medicine and big data.

HISTORY OF THE UPCI SUMMER ACADEMY

The UPCI Summer Academy was launched in 2009 and the Annual Reports since its inception can be downloaded at http://www.upci.upmc.edu/summeracademy/reports.cfm. These annual reports form the basis for the history of the UPCI International Academy (renamed in 2012) presented here (see also http://www.upci.upmc.edu/summeracademy/history.cfm).

Year 1-2009

The inaugural UPCI Summer Academy for Cancer Careers succeeded in its goals of encouraging students' interest in cancer careers, instilling knowledge of cancer biology and clinical care, and developing research and communication skills. The initial program received seed funding from philanthropic and UPCI sources, and was launched with five talented and motivated students: Four rising high school seniors recruited from among PA Governor School applicants and one recent high school graduate. The students attended a series of cancer biology lectures presented by a UPCI clinician/researcher, a biology professor, and a University of Pittsburgh medical student. They also attended presentations by clinicians and researchers from across UPCI disciplines focusing on clinical care, career options, and career preparation. Students were led on tours of a variety of clinical and research facilities at UPCI and University of Pittsburgh Medical Center (UPMC) Shadyside. Most of each student's day was spent conducting laboratory research in a UPCI lab. This inaugural year taught us much and formed the basis for the next 4 years of the academy.

Years 2 through 4-2010-2012

In 2010 (year 2) we expanded the participation

of underrepresented and disadvantaged/minority students by working closely with the Pittsburgh Public School system, the University of Pittsburgh's Office of Diversity, and national philanthropic programs, specifically the Jack Kent Cooke Foundation (JKCF). JKCF is a private, independent foundation established to help exceptionally promising scholars with modest family means reach their full potential through education. The academy also continued to receive support from local organizations such as the Pittsburgh Tissue Engineering Initiative and Bayer Material Sciences. Additionally, the academy was funded in early 2010 by a P30 CURE supplement from the National Cancer Institute. This grant enabled recruitment of under-represented minorities and under-represented students and provision of a $2,000 stipend and Robert Weinberg's "Biology of Cancer" textbook, allowing faculty mentors and laboratories to receive a $500 bench fee/supply stipend for their students' projects. In 2010, the academy established a field trip to the National Cancer Institute (NCI) as a regular part of the program, which now included 10 scholars. In 2011, the program was expanded to two additional sites—CoSBBI and the Magee Women's Research Institute's Women's Cancer Research Center (WCRC), and had 24 high school students as scholars. The academy continued the unique partnership with JKCF. In 2012, the academy hosted three out-of-state scholars under the JKCF's Young Scholars Program to spend the summer in Pittsburgh: The Hillman Cancer Center hosted two of the scholars, and the third spent the summer working in computational systems biology with the CoSBBI program. In addition, the Academy received funding from the Doris Duke Charitable Foundation (DDCF). Over these years, enrolment in the program grew from 10 in 2010 to 24 in 2011 and 28 in 2012.

Year 5 - Refocusing the UPCI summer academy to the UPCI international academy

With success of the Academy we underwent a dramatic expansion to include 56 student scholars. This included the expansion to two new areas: The Drug Discovery, Systems, and Computational Biology (DiSCoBio) and Tumor Immunology, both in the Oakland campus. The DiSCoBio component formed its own program, geographically separate from CoSBBI, due to the logistical challenges posed by the move of the Department of Biomedical Informatics from Oakland to the Shadyside campus. In addition, in 2013 we recruited our first international students to the program and renamed the UPCI Summer Academy, the UPCI International Academy. Students from the 2013 class of scholars were selected from 143 applicants, and included recruits from Hawaii, New Jersey, Virginia, Minnesota, Texas, Vermont, New York, North Carolina, Maryland, Germany, and Kazakhstan.

BACKGROUND ON COSBBI

The year 2013 was the third year of the CoSBBI Summer Academy. We enrolled a total of three (2011), seven (2012), and 11 (2013) high school students over the past three summers, and the number of mentors and number of students each mentor was willing to take on largely dictated this size. The 1st year was an experimental year, which turned out stellar students who went on to succeed in Intel Science Fair competitions regionally and nationally. We more than doubled the number of students the subsequent year, when CoSBBI was still co-located with the Department of CSB at the Oakland campus of the University of Pittsburgh. In 2013, CoSBBI was relocated to the Shadyside campus, and housed within the DBMI. CoSBBI students predominantly work on computational projects. We are able to take on younger students and have included seven sophomore high school students (rising juniors) thus far. We also partner with mentors from the Department of Pathology to provide experimental pathology projects focused on imaging informatics and next generation sequencing bioinformatics. In the class of 2013, we had five rising seniors and six rising juniors who developed and presented projects spanning a broad range of informatics topics including bioinformatics, computational biology, machine learning, image analysis, pharmacogenomics, and telemedicine.

A key component of the training that our students receive is from individual mentoring from outstanding scientists at the University of Pittsburgh's DBMI (see http://www.dbmi.pitt.edu/people) and Division of Pathology Informatics (see http://path.upmc.edu/cpi/). Research projects are usually prepared in advance by mentors when they are paired with the incoming high school student. Three publications from the mentor's laboratory or research area are sent to each trainee prior to the start of this 8-week academy. The scholars prepare to ask questions of the mentor regarding their specific project at the very start of their training. Some mentors also put two or three high school students onto the same project, providing them with adequate support for learning software programming and applying it to big datasets coming from genomic experiments. The role that mentors play is crucial to the success of the outcomes in terms of scholar satisfaction. Mentors often enlist postdoctoral and graduate students in their laboratories to provide additional support to enhance the learning experience of our CoSBBI scholars. On the final day of the academy, research mentors introduce their students to their colleagues, and the families of CoSBBI scholars are invited to attend. In the morning session, students describe their projects using an oral PowerPoint presentation. This is followed by a poster symposium at the Hillman Cancer Center in the afternoon. The

afternoon session is open to the entire UPCI community and the public. Beginning in 2013, we will encourage the faculty research mentors and CoSBBI scholars to publish their abstracts in the Journal of Pathology Informatics (JPI) as part of our commitment to the hard work and energy these students bring to our laboratories and research projects each year.

Graduates from our summer program continue to stay in touch with their research mentors over the subsequent year to obtain recommendations for college applications, and to continue their research projects towards publication. All UPCI summer academy scholars are invited to present their posters during a reunion each year in early October, at a prestigious Science conference organized by the University of Pittsburgh, which brings together cutting-edge science and technology in the region. This experience has enabled our scholars to learn about their own research questions in the context of broader, related, and integrated scientific and engineering applications. Our past scholars have been admitted into honors colleges at prestigious national universities. Some universities that our scholars have accepted include: Duke University, Harvard University, Massachusetts Institute of Technology (MIT), Penn State, Stanford University and University of Pennsylvania.

ROLE OF THE CLASSROOM COMPONENT

The classroom portion of CoSBBI was designed to provide a didactic introduction to biomedical informatics, promote an understanding of research, and expose students to career opportunities in this field. Towards this end, 4 weeks of daily lectures covered fundamental concepts and activities on information technology applied to biomedicine and health care. Each day was comprised of 1 instructional hour, led by doctoral, postdoctoral and medical students, followed by 1 hour of research presentation and discussion, led by faculty and industry guests. Lectures in the early weeks covered basics of molecular biology, bioinformatics tools, computational thinking, statistics, and data mining. In the absence of a standard undergraduate level textbook for our field, we selected the recently released compilation, Translational Bioinformatics (PLOS Computational Biology Collection, see syllabus). This online, open-access collection provided chapters crafted by leading experts in topics such as genomics, proteomics, Bayesian inference and decision modeling, and pharmacogenomics. These were complemented with lectures on human computer interaction and issues in technology incorporation for laboratory workflow, medication safety, and biosurveillance. The complete syllabus and links to PowerPoint presentations can be accessed at http://faculty.dbmi.pitt.edu/jod30/classes/cosbbi2013/. The classroom sessions were focused on concepts and application, with students encouraged to pursue deeper the skills relevant to their individual research project. Periodic sessions were held to discuss research progress, reading and presenting peer-reviewed papers, and writing an abstract.

Students in the CoSBBI program came with a background in advanced high school biology, but they varied in their mathematics and computational background, ranging from no programming experience to proficient in developing simple standalone applications. The students were administered a pre-test on the 1st day of class, consisting of multiple choice questions covering fundamentals from biology, genetics, protein interactions, computer science, bioinformatics, and biomedical informatics. The same test was again administered after the last day of class. Nine out of 11 students showed improvement in their scores, with a highest individual improvement rate of 83%. On an average, student performance increased by 26%.

IMPACT ON INFORMATICS - CREATING A PIPELINE

Biomedical Informatics training programs have existed since the 1970s[5] and our training program in Pittsburgh has been funded by the National Library Medicine since 1984 (see http://www.dbmi.pitt.edu/content/overview). Pathology informatics training has existed in the Division of the Department of Pathology at the University of Pittsburgh School of Medicine (UPSoM) since 1999 when we established the Center for Pathology Informatics (see http://path.upmc.edu/fellowship/informatics/index.htm). In both of these programs we recognize that we have had the opportunity to train excellent students. However, it is clear to us that the majority of recruits both in biomedical and pathology informatics come to our graduate training program with little relevant research experience and many decide on a career in informatics after training in other disciplines. At our Biomedical Informatics Training Program Retreat in 2012 we discussed this issue and found that only three of 113 people at the retreat had done relevant research in informatics during their high school and early college experiences. It turned out that those three people were the most widely cited authors and successful National Institute of Health (NIH) grant funded investigators in the group.

Spurred on by the successful launch of our high school student academy in 2011, we decided to establish a pipeline of the "best and the brightest" students and encourage them to pursue careers in informatics. The concept we sought to establish was to invite high school students to gain research experience and help them plan their college training to best prepare them for careers in informatics. We then established the following informatics directed "pipeline" plans which are in various stages of implementation today:

- Recruit the best high school students (as early

Table 1: Colleges and Universities Offering Undergraduate Degrees in Bioinformatics

College or University Name	Location	Program description URL
Baylor University	Waco, TX	http://www.ecs.baylor.edu/computer_science/index.php?id=29232
Brigham Young University	Provo, UT	http://saas.byu.edu/catalog/2013-2014ucat/departments/Biology/BioinformaticsMajor.php
Canisius College	Buffalo, NY	http://www.canisius.edu/bif/
City University of New York: Hunter College	New York, NY	http://www.hunter.cuny.edu/csci/for-students/the-computer-science-bioinformatics-concentration
City University of New York: New York City College of Technology	Brooklyn, NY	http://www.citytech.cuny.edu/academics/deptsites/biological/degrees.aspx
Claflin University	Orangeburg, SC	http://www.claflin.edu/academics/undergraduate-majors-minors
Clark University	Worcester, MA	http://www.clarku.edu/departments/mathcs/bioinformatics/index.cfm
College of Saint Rose	Albany, NY	http://strose.smartcatalogiq.com/en/2013-2015/Catalog/Programs-of-Study/Bioinformatics-BS
Gannon University	Erie, PA	http://www.gannon.edu/Academic-Offerings/Engineering-and-Business/Undergraduate/Bioinformatics/
George Mason University	Fairfax, VA	http://www.cs.gmu.edu/programs/undergraduate/acs/
George Washington University	Washington, DC	http://www.cs.gwu.edu/academics/undergraduate_programs/transfer/nvcc
Iowa State University	Ames, IA	http://bcbio.las.iastate.edu/
Loyola University Chicago	Chicago, IL	http://www.luc.edu/bioinformatics/academics_bs.shtml
Michigan Technological University	Houghton, MI	http://www.mtu.edu/admissions/programs/majors/bioinformatics/
Missouri Southern State University	Joplin, MO	http://www.mssu.edu/academics/programs/bioinformatics.php
New Jersey Institute of Technology	Newark, NJ	http://catalog.njit.edu/undergraduate/programs/bioinformatics.php
Pacific University	Forest Grove, OR	http://www.pacificu.edu/as/bioinformatics/
Portland State University and Oregon Health and Science University	Portland, OR	http://www.pdx.edu/computer-science/biomedical-informatics-program
Ramapo College of New Jersey	Mahwah, NJ	http://bioinformatics.ramapo.edu/bsbinf/
Rensselaer Polytechnic Institute	Troy, NY	https://www.rpi.edu/dept/bio/undergraduate/bsbioinfo.html
Rochester Institute of Technology	Rochester, NY	http://www.rit.edu/programs/bioinformatics-0
Saint Bonaventure University	St. Bonaventure, NY	http://www.sbu.edu/academics/schools/arts-and-sciences/departments-majors-minors/bioinformatics
St. Edward's University	Austin, TX	http://www.stedwards.edu/academics/bachelors/bioinformatics
St. Vincent College	Latrobe, PA	http://www.stvincent.edu/academics/bioinformatics/
Stevens Institute of Technology	Hoboken, NJ	http://www.stevens.edu/ses/ccbbme/undergrad
SUNY University at Buffalo	Buffalo, NY	http://undergrad-catalog.buffalo.edu/academicprograms/bioinfo.shtml
University of Alberta	Edmonton, AB	http://www.biology.ualberta.ca/programs/undergraduate/?Page=8825
University of California: Irvine	Irvine, CA	http://www.ics.uci.edu/~biomed/
University of California: Los Angeles	Los Angeles, CA	http://www.bioinformatics.ucla.edu/undergraduate/
University of California: Riverside	Riverside, CA	http://cnasstudent.ucr.edu/majors/biosci.html
University of California: San Diego	La Jolla, CA	http://bioinformatics.ucsd.edu/node/7
University of California: Santa Cruz	Santa Cruz, CA	https://bme.soe.ucsc.edu/bioinformatics
University of Maryland: Baltimore County	Baltimore, MD	http://www.cbcb.umd.edu/under-graduate-programs
University of Missouri - Kansas City	Kansas City, MO	http://www.umkc.edu/majormaps/
University of Nebraska - Omaha	Omaha, NE	http://bioinformatics.ist.unomaha.edu/undergraduate.php
University of Pennsylvania	Philadelphia, PA	http://www.upenn.edu/ben-penn/bioinfo.html#undergrad
University of Pittsburgh	Pittsburgh, PA	http://www.cs.pitt.edu/undergrad/bioinformatics/
University of St. Thomas	Houston, TX	http://www.stthom.edu/Academics/School_of_Arts_and_Sciences/Biology
University of Toronto	Toronto, ON	http://www.biochemistry.utoronto.ca/bcb/
University of Waterloo	Waterloo, ON	http://ugradcalendar.uwaterloo.ca/page/MATH-Computer-Science-Bioinformatics
Virginia Commonwealth University	Richmond, VA	http://www.vcu.edu/csbc/bioinformatics/bachelor/
Walsh University	North Canton, OH	http://www.walsh.edu/bioinformatics-degree
Wheaton College	Norton, MA	http://wheatoncollege.edu/bioinformatics/major/
Worcester Polytechnic Institute	Worcester, MA	https://www.wpi.edu/academics/bcb/ugrad-requirements.html

as sophomores) and provide mentored research experience via CoSBBI and establish a long-term relationship as their academic advisors
- Offer CoSBBI scholars paid research assistant positions in informatics in our laboratories as summer employees through high school and while in college
- Encourage students applying for college to consider programs which have undergraduate degrees in bioinformatics [Table 1] and encourage them to take coursework in math, statistics, computer science, biology, as well as the study of human disease pathobiology
- Emphasize to CoSBBI scholars the importance of publishing their work and to present at national pathology and informatics meetings (American Society for Clinical Pathology, American Medical Informatics Association, College of American Pathologists, and Pathology Informatics Summit). We also provide travel award scholarships
- Foster a "virtual" community of high school and college trainees interested in informatics trainees through the establishment of a not for profit with this goal as its core mission.

This proposed pipeline has many moving parts of which the most important is the recruitment and retention of high quality trainees. The UPCI International Academy has reached out to many organizations and is developing both national and international reach. However, there are many competing summer programs and reaching high school students is very difficult, compared to recruiting college and graduate students. Developing a 'viral' program via social media, coupled with word-of-mouth via successful CoSBBI scholars, is our current focus. Aggressive Pittsburgh School system outreach and open-house programs have been very successful in providing regional coverage, but national coverage remains a challenge. Nonetheless, we are making steady progress by increasing the visibility of our trainees through their participation in national meetings and science fair programs such as the Intel Science and Engineering Fair (ISEF). We have had three of our CoSBBI alumni (out of 21 to date) qualify for the International ISEF after winning regional science fairs.

To distinguish our high school scholars and the research experience from other programs we have decided this year to publish the abstracts of their summer work in the JPI. JPI is the publication sponsored by the Association for Pathology Informatics (http://www.pathologyinformatics.org). We have decided that the best way to attract other innovative students is to feature the work of CoSBBI scholars in the literature. We leave this decision to publish the abstract to the scholar's faculty mentor. In this inaugural year of this enhancement to our program 10 of the 11 scholars abstracts were published and are included below. In addition, two of our students have presented their work at our national meeting the Pathology Informatics Summit (http://www.pathologyinformatics.com). Discussions are currently underway to sponsor a high school scholar's session at the annual meeting of the American Medical Informatics Association Meeting.

CONCLUSION

This editorial describes a Pittsburgh-based effort to create a pipeline of training opportunities in informatics starting with high school and continuing through college. We aim to attract the best and brightest high school students nationally and train them for informatics as CoSBBI scholars. As part of the program, the high school CoSBBI scholars participate in a 4-week formal didactic session and a mentored research project, culminating in a formal presentation to the scientific community, as well as to their families. This 8-week experience closes with a competitive poster symposium. We have now chosen to publish abstracts of the CoSBBI scholars (see accompanying prologue by Dr. Vanathi Gopalakrishnan, CoSBBI course director). Lastly, we have developed a plan for this pipeline which is being fully implemented in Pittsburgh by 2015, including the potential establishment of a 501c3 not-for-profit organization to support this effort and create the 'virtual' community needed for its success. Three organizations are planning to implement CoSBBI-like informatics oriented STEM initiatives in their institutions. The Pittsburgh team will share all of the materials it has assembled to assist other organizations in order to increase the number of trainees interested in informatics as a career. Transformation through education is a team sport … game on!!!

ACKNOWLEDGMENTS

We appreciate the partnership we are forming with the Departments of Biology and Computer Science at the University of Pittsburgh and want to thank the Chairs of those Departments, Daniel Mosse, PhD and Paula Grabowski, PhD. We also want to thank Andrew King, BS and Peter Randall (soon to be BS) who are graduates and current trainees in the University of Pittsburgh Bioinformatics undergraduate degree program. Andrew is the first graduate of this program that we have accepted into our PhD program in Biomedical Informatics. Special thanks to administrative and project management support from Lucy Cafeo, Nancy Whelan, Albert Geskin, and Linda Mignogna. We want to particularly thank Megan Seippel, Joe Ayoob, and many other helpful staff of the UPCI International (formerly Summer) Academy who have kept us on track each year for this important effort. This CoSBBI track of the UPCI International Academy is supported by NIH grants R01 LM010950 from the National Library of Medicine, R01 GM100387 from the National Institute of General Medical

Sciences, Doris Duke Foundation, the National Cancer Institute CURE Program (3P30CA047904-22S1), and support from the Jack Kent Cook Foundation. In addition, this program would not be possible without the infrastructure and support teams from the DBMI NLM Training Program Grant in Biomedical Informatics (T15 LM007059), the University of Pittsburgh Cancer Institute (UPCI) Cancer Center Support Grant for the Cancer Bioinformatics Service (P30 CA47904), the Clinical and Translational Science Institute Biomedical Informatics Core (UL1 RR024153), as well as from funds and in-kind services from UPCI and DBMI, and the Departments of Surgery, Immunology, Computational and Systems Biology, Gynecology, and Pharmacology.

REFERENCES

1. Becich MJ. The role of the pathologist as tissue refiner and data miner: The impact of functional genomics on the modern pathology laboratory and the critical roles of pathology informatics and bioinformatics. Mol Diagn 2000;5:287-99.
2. Hamburg MA, Collins FS. The path to personalized medicine. N Engl J Med 2010;363:301-4.
3. Mattmann CA. Computing: A vision for data science. Nature 2013;493:473-5.
4. Machluf Y, Yarden A. Integrating bioinformatics into senior high school: Design principles and implications. Brief Bioinform 2013;14:648-60.
5. Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, *et al*. MIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. Methods Inf Med 2010;49:105-20.

Abstracts

Prologue to Abstracts

Vanathi Gopalakrishnan

Departments of Biomedical Informatics, Intelligent Systems and Computational Biology, Director, UPCI CoSBBI Summer Academy, University of Pittsburgh, United Stated of America

We present here the 2013 CoSBBI scholar abstracts which merit publication. Ten of our eleven summer 2013 scholars are represented in the 8 abstracts which follow. The decision to publish the work of our high school scholars was approved by the faculty of the Department of Biomedical Informatics and the Division of Pathology Informatics of the University of Pittsburgh School of Medicine. We feel it is quite telling that our faculty unanimously voted to approve this plan as we have been uniformly impressed by how these students have contributed to our laboratories in such a short period of time. Since high school students really do not know that some aspects of science are very hard to do, assigning them such a project can sometimes yield surprising scientific results. To cite a specific example, one mentor was told by laboratory members that an imaging analysis task was too time-consuming and that the publicly available images were not of great quality, and hence it would be hard to extract meaningful features from them. The same task was completed within a month by teaming together a CoSBBI high school student with another undergraduate summer trainee, which has yielded meaningful results that were subsequently published. We encourage the readers to make their own decision if the work presented below as abstract merits publication and provide feedback to the authors. We have made our decision and clearly feel that the continued training of these CoSBBI scholars is already changing our informatics community in a very positive way. The future of informatics is bright indeed!

CoSBBI 2013 Abstracts

Scoping Review of Telemedicine in Nursing Homes

Harishwer Balasubramani[1], Kayse L. Reitmeyer[2], Reza Sadeghian[2], Tanja Bekhuis[2*], Andrea M. Ketchum[3], Jill E. Foust[3], Steven M. Handler[2,4*]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, [3]Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, [4]Division of Geriatric Medicine, University of Pittsburgh, Pittsburgh, PA, *Co-Mentors.
E-mail: Steven Handler handler@pitt.edu

CONTEXT

It is anticipated that the population of US nursing home residents will double from 1.5 to 3 million by 2030. Inadequate access to appropriate and timely care poses a problem for a vast majority of these residents, as it can lead to an increase in unnecessary or inappropriate treatment, avoidable hospital admissions, and healthcare utilization.

TECHNOLOGY

Telemedicine is defined as the use of telecommunication and information technologies in order to provide clinical healthcare at a distance. This technology may be able to increase access to appropriate and timely care, resulting in improved processes of care, health-related outcomes, resource

utilization, or patient/provider satisfaction.

DESIGN

We conducted a scoping review of telemedicine in nursing homes using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology to assess the nature and extent of the research literature, as well as to identify research gaps. Two health sciences librarians wrote comparable queries for three electronic databases (PubMed, CINAHL, PsycINFO) and delivered a set of citations with titles, abstracts, and metadata as an EndNote version X6 library. Two reviewers independently screened the citations to identify relevant studies. We iteratively designed a data extraction form with guidance from a domain expert and a methodologist. We managed data in Excel, version 14.0, and recorded a variety of data including bibliographic information, study design, publication type, clinical setting, telemedicine category (store-and-forward, remote monitoring and real-time/interactive), resident and facility factors, and purpose of the consultation.

RESULTS

We retrieved 1,866 citations: PubMed N = 1,231; CINAHL N = 351; PsycINFO N = 284. After de-duplication, N = 1,760. Subsequent to screening citations, both reviewers agreed that 74 studies (4.2%) were relevant. Most of the included studies appear to be observational, report a variety of process measures and outcomes, and utilize all three telemedicine categories.

CONCLUSIONS

Preliminary evidence suggests that telemedicine can improve access to quality care, processes of care, health-related outcomes, resource utilization, and patient or provider satisfaction.

Distribution of Palindromes in the Human Genome

Sophia Cheng[1*], Ritwik Gupta[1*], Tonya Hammond[1*], Lavanya C. Viswanathan[2], and Madhavi K. Ganapathiraju[3]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology, and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2] Language Technologies Institute, Carnegie Mellon University, [3]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA.
E-mail: Madhavi Ganapathiraju Madhavi@pitt.edu
*These authors contributed equally to this work

CONTEXT

Palindromes are words that read the same when they are read forwards and backwards, like the word "racecar". In a DNA palindrome, a strand of nucleotides when read forwards is the complement of what it is when it is read backwards (e.g. 5'-CGATCG-3'). When a single strand of DNA is exposed, like during replication, palindromes can form into cruciform structures and hairpins by self-annealing. Short palindromes can possibly prevent degradation and help gene stability; whereas, long palindromes have potentially fatal implications such as mutations, instability of DNA, diseases, and cancer. We studied the prevalence of DNA palindromes in different genomic regions and also analyzed the distribution of palindromes in cancer genes.

DESIGN AND TECHNOLOGY

Human DNA palindromes were computed with the BLM Toolkit by Ganapathiraju *et al*. We computed the frequencies and lengths of palindromes in promoters, intergenic regions, exons, and introns of all chromosomes; we also studied their prevalence in cancer genes, and whether the palindromic regions are more prone to mutations observed in cancer patients. We downloaded the data from UCSC Genome Browser, COSMIC, and TCGA

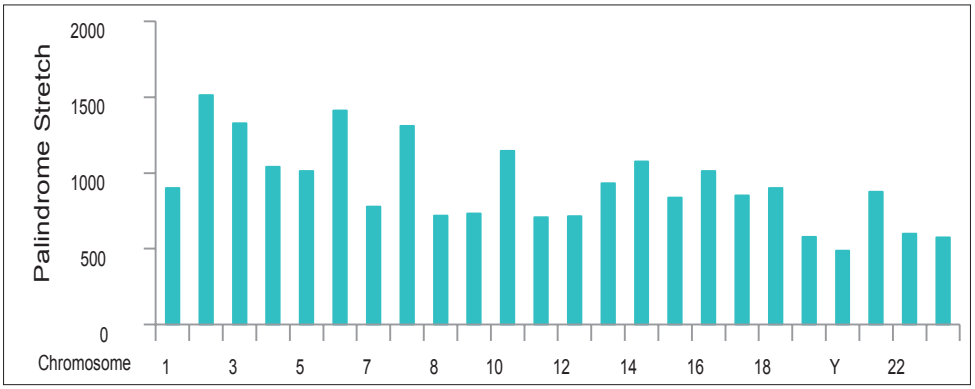

Figure 1: Length of longest palindromic stretch in each chromosome

Data Matrix, and wrote programs in Python, and interpreted the results. Long palindromes (>50 bp) were analyzed in the exons of non-cancer and cancer genes separately.

RESULTS

There are about 13,278 palindromic stretches (regions of contiguous palindromes) in exons and 423,292 palindromic stretches in introns per chromosome [Figure 1]. We found that chromosome Y has a much higher percentage of palindromes (53%) than the other chromosomes (35%). The distribution of long palindromes (>50 bp) in exons was found to be greater in cancer genes than non-cancer genes in chromosomes 3, 4, X, 9, 11, 12, 16, 19 and 21.

CONCLUSION

Human genome contains large number of palindromes, especially in chromosome Y. Furthermore, there is an underrepresentation of palindromes in exons. Furthermore, since the Y chromosome has a high percentage of palindromes, one can evaluate if palindromes are associated with sex-linked diseases. Prevalence of disease mutations and other genetic variants can be studied in relation to palindromes.

Interactive Visualization of FDA Pharmacogenomics Drug Labels

Meghana Ganapathiraju[1], Harry Hochheiser[2]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA.
E-mail: Harry Hochheiser harryh@pitt.edu

CONTEXT

Important pharmacogenomic information as presented in Structured Product Labels is difficult to read and interpret due to a lack of common organization. Time-pressed clinicians are unable to effectively use pharmacogenomic information in making clinical decisions. Our goal was to create a tool that would improve the comprehensibility and organization of pharmacogenomics information in structured product labels. These clearer presentations may help clinicians make more effective decisions, and avoid adverse drug reactions.

TECHNOLOGY AND DESIGN

Expert annotations describing pharmacogenomic information in Structured Product Labels formed the basis for our visualizations. The annotations contained the drug name and information on its class, primary associated gene, pharmacogenomic impact, and clinical recommendation. Using a Python program, annotations were converted into a JSON file used to create visualization in a webpage using the D3 JavaScript Library. Relevant pharmacogenomic information is organized into an interactive tree, allowing for collapsing and expanding groups to facilitate convenience and allow for a clearer view of the drug/gene of interest. Three hierarchies were created, which sort by drug class, drug name, or associated gene.

RESULTS

The visualization's three tree views help to enhance the interpretation of pharmacogenomic information. Each of the trees showcases a different perspective, as different users may need a different view. For example [Figure 2], a clinician looking for a substitute drug may use the view organized by class of drug, while a clinician looking to avoid drugs that interact with a gene may choose to group by gene.

CONCLUSIONS

Our tool covers only a small selection of drugs. To be of clinical importance, the visualization would have to be expanded to include all the drugs with relevant pharmacogenomic information, allowing clinicians to access all necessary information more easily. Only the tree structure is currently supported. Bringing different methods of visualization to the tool would help users from different perspectives to better analyze the data and make better clinical decisions. We hope to also conduct user tests and evaluate the effectiveness of visualization in simulated contexts.

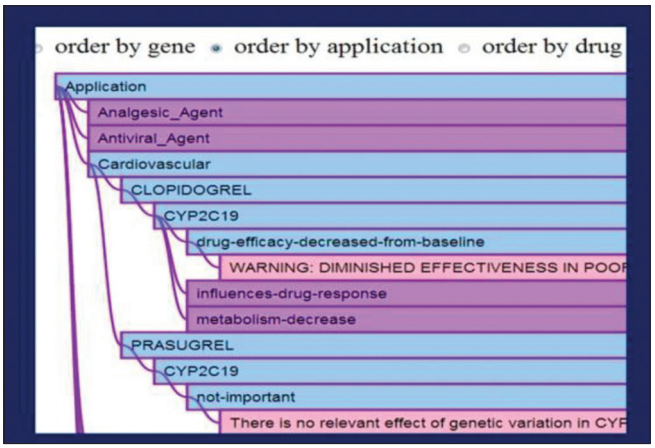

Figure 2: An example of view ordered by drug application

Evaluation of BRCA 1 and BRCA 2 Expression Profiles in a Large Series of Bladder Tumor Specimens: Correlation of Tumor Stage with Expression Profile

Huldah Kena[1], Anil Parwani[2]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh. PA.
E-mail: Anil Parwani parwaniav@upmc.edu

CONTEXT

Bladder cancer is a relatively common cancer with an estimated number of 72,570 new cases in 2013 in adults (54,610 men and 17,960 women) in the United States. A great number of bladder cancers are diagnosed at an early stage. The objective of this study was to evaluate the expression levels of BRCA1 and BRCA2 proteins in a large series of bladder carcinomas spanning four clinical stages to ascertain if these two genes may play a role in bladder neoplasia progression. **Technology:** Tissue microarrays and Aperio XT Slide scanner and software.

DESIGN

A total amount of 527 cases of bladder cancer were collected for this study of which 36 were stage T1, 81 were stage T2, 201 were stage T3 and 64 were stage T4. The tissue was processed for histology and paraffin blocks were used to produce tissue microarrays. Immunohistochemistry was used to stain the slides with antibodies against BRCA1 and BRCA2. The slides were digitized using a whole slide scanner (Aperio Technologies). Image analysis was performed using Aperio Cell Quant software. The data was analyzed and a subset of this data was correlated with a manual read of the stained slides by a board certified pathologist.

RESULTS

The average staining intensity of BRCA1 protein in bladder carcinomas was 116.3, Non Bladder was 81.7 and Normal Adjacent to Tumor was 116 (T1 = 117.7, T2 = 104.5, T3 = 120.7 and T4 = 122.3). The average staining intensity of BRCA2 protein in bladder carcinomas was 116.2, Non-Bladder tissue was 114.1 and normal adjacent to tumor was 123.3 (T1 = 119.2, T2 = 116.1, T3 110.9 and T4 118.6).

CONCLUSIONS

In conclusion, both tumor and normal adjacent to tumor specimens for both BRCA1 and BRCA2 had higher intensity as compared to non-bladder tissue. In addition, for BRCA1, the intensity was highest in higher stages, as compared to normal adjacent to tumor. For BRCA2, the reverse was true as the expression levels were lowest in the higher stages (T3 and T4) as compared to normal adjacent to tumor. Although not statistically significant, this data suggests that BRCA1 and BRCA2 may play a role in bladder tumor progression.

Automated Image Analysis for Immunohistochemical Evaluation of Protein Expression Levels to Assess Their Use as Biomarkers for Renal Cell Carcinoma

Sreejan Kumar[1], Anil Parwani[2]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA.
E-mail: Anil Parwani parwaniav@upmc.edu

CONTEXT

Renal Cell Carcinoma (RCC) is a common cancer and is expected to kill 13,680 people in 2013. Subtypes of RCC include clear cell, chromophobe, oncocytoma, and papillary carcinoma. Recent studies show these subtypes have unique amplifications and deletions in certain genes. Our goal was to evaluate antibodies expressed by four of these genes to determine their usefulness as biomarkers for RCC.

TECHNOLOGY

Slides of Tissue Microarrays were scanned using Aperio Scanscope XTslide scanner (Aperio Technologies, Vista, CA, USA) at ×20 magnification. Digital slides were analyzed using Aperio's annotation software (Imagescope v11.1.2.760) and its Positive Pixel Count Algorithm v9. The Pathologist used the Olympus light microscope model BX45TF (Olympus Corporations, Shinjuku, Tokyo, Japan).

DESIGN

Using immunohistochemistry, four antibodies were analyzed for their expression levels (ACR, ZNF860, MUC20, and MRC1) using TMAs comprised of specimens from patients with RCC. The slides were analyzed using image analysis software to obtain staining intensities. In addition, a pathologist reviewed a subset of the cases.

RESULTS

Chromophobe tumors had the greatest staining intensity followed by Oncocytoma in each protein. ZNF860 had the greatest staining intensity in clear cell and the lowest staining intensity in papillary RCC. Pathology review determined ZNF860 had weak nuclear staining in clear cell and strong cytoplasmic staining in papillary tumors. Also, ACR, MUC20, and MRC1 stained predominantly the distal renal tubules.

CONCLUSION

ACR, MUC20, and MRC1 were highly expressed in the distal renal tubules as well as both oncocytoma and chromophobe tumors. These antibodies may serve as biomarkers for both the distal renal tubules as well as tumors that originate therein like oncocytoma and chromophobe. Additionally, ZNF860 may be a biomarker for papillary carcinomas. Larger number of renal tumors will be further tested with these biomarkers to validate their efficacy as diagnostic biomarkers of renal neoplasms.

Machine Learning for Biomarker-based Classification of Alzheimer's Disease Progression

Amy McMillan[1], Shyam Visweswaran[2], Vanathi Gopalakrishnan[2]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Departments of Biomedical Informatics and Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA. E-mail: Vanathi Gopalakrishnan Vanathi@pitt.edu

CONTEXT

Patients with mild cognitive impairment (MCI) are at a significantly increased risk of developing Alzheimer's Disease (AD). While imaging and proteomic marker data exists for AD within the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, the accuracy of models obtained from such data using standard machine learning methods is lower than can be used for clinical testing. This research uses novel rule learning methods to assess whether there is improvement in classification performance over standard methods when imaging markers are combined with proteomic markers for distinguishing MCI to AD progression.

TECHNOLOGY

The Naïve Bayes algorithm was used from Weka (Waikato Environment for Knowledge Analysis), a collection of machine learning algorithms, while Bayesian Rule Learning was used from the PRoBE lab at University of Pittsburgh's Department of Biomedical Informatics.

DESIGN

Processed magnetic resonance imaging (MRI) data from ADNI was used. The data contained scores representing the spatial pattern of abnormalities for early recognition (SPARE) from a study of 212 patients. The SPARE values for distinguishing AD converters from the non-converters were input to the Naïve Bayes (NB) and Bayesian Rule learning (BRL) algorithms to learn classifiers evaluated over ten-fold cross validation. Proteomic biomarkers for 88 patients were also input to NB and BRL. The imaging and proteomic biomarkers were then integrated to create a subset of 38 patients and processed through the same classification methods using leave-one-out cross fold validation.

RESULTS

NB testing for SPARE scores had a classification accuracy of 52.35%, while BRL achieved 70.28%. For the proteomic data alone, NB produced a 52.83% accuracy rate, compared to BRL's 83% accuracy. When integrating the SPARE scores and the proteomic biomarkers, NB produced 74.63% accuracy, while BRL achieved significantly higher results (97.25%).

CONCLUSION

These findings suggest that not only is BRL more effective in classifying biomarkers of AD progression, but also that the integration of imaging and proteomic data decreases classification error compared to individual values. In the future, we would like to include more data from ADNI to increase our confidence in these results and to also explore new methods such as Markov Chains to model longitudinal data.

Comparison of Manual versus Computer-Assisted Proliferation Scoring in Tumors

Simran Parwani[1], Sara E. Monaco[2], Malini Srinivasan[2,3], Roger Day[4], Jon Duboy[2], Liron Pantanowitz[2,4]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA,[2] Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, [3]University

of Pittsburgh Cancer Institute, Pittsburgh, PA, [4] Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA. E-mail: Liron Pantanowitz pantanowitzl@upmc.edu

CONTEXT

Certain tumor types require an assessment of their proliferation index for diagnostic, therapeutic, and prognostic purposes. Ki-67, an immunohistochemical marker that stains proliferating nuclei in all active phases of the cell cycle, is used for this purpose. The objective of this study was to compare manual versus computer-assisted methods of quantifying the Ki-67 proliferation index in small biopsy material from a variety of tumors.

TECHNOLOGY

Aperio XT scanner and Aperio (Vista, CA, USA) nuclear image analysis algorithm (Version 9.1) was used to calculate the proliferation index.

DESIGN

We selected archival cytology cell block or core biopsy samples from 10 consecutive sarcomas, brain tumors, non-Hodgkin lymphomas, and neuroendocrine tumors that had a Ki-67 score reported. The manual score reported by the pathologist was recorded. A representative H and E and Ki-67 (MiB1 immunostain) glass slide were digitized using an Aperio XT scanner. Using whole slide images, 3-5 regions in each tumor with high Ki-67 proliferation (hot spots) were analyzed using the Aperio algorithm. The proportion of Ki-67 stained nuclei was calculated. Ki-67 was separately scored by two pathologists (expert consensus) using these hot spots and self-selected regions. Manual scores were compared to Aperio scores using R Studio statistical software.

RESULTS

The overall correlation between manual and image analysis scores was good for all tumor types. However, the algorithm failed to correctly calculate scores in three tumors. The presence of crushed nuclei caused the algorithm to underestimate the score. Large lymphocytes associated with tumor caused the algorithm to overestimate the proliferation index. These errors indicate why the Pearson correlation between the expert-selected and Aperio score was 0.868, but was 0.935 between the expert-selected and the reported score.

CONCLUSION

Employing image analysis to determine Ki-67 scores in tumors is challenging when dealing with small biopsy material, because tumor cells may be crushed or samples may contain many non-neoplastic proliferating lymphocytes. Pathologist involvement is recommended to avoid such artifacts when image analysis is applied to cytology and small biopsy samples. A computer algorithm for scoring Ki-67 using MATLAB is being explored to develop better image analysis methods for calculating the proliferation index in tumors.

Discovering Biomarkers for Cardiovascular Disease Using Rule Learning

Mara Staines[1], Lailonny Morris[3], Prahlad G. Menon[2,3], Joao Lima[4], Daniel C. Lee[5], Vanathi Gopalakrishnan[3]

[1]University of Pittsburgh Cancer Institute Computer Science, Biology and Biomedical Informatics (CoSBBI) Summer Academy, Pittsburgh, PA, [2]Sun Yat-sen University - Carnegie Mellon University Joint Institute of Engineering, Pittsburgh, PA, [3]Departments of Biomedical Informatics and Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, [4]Cardiovascular Imaging, The John Hopkins Hospital, Baltimore, MD, [5]Feinberg Cardiovascular Research Institute, Northwestern University, Evanston, IL. E-mail: Vanathi Gopalakrishnan Vanathi@pitt.edu

CONTEXT

Image-derived metrics of cardiovascular function require investigation into their potential use as biomarkers for cardiovascular disease. We also desired to test a new biomarker, reporting root mean square (RMS) error from average phase to phase regional left ventricular endocardial displacement (P2PD), computed on a patient specific basis. In this research, we investigate the classification of diseased and healthy cardiac subjects through rule learning applied to standard cardiac MRI (cMRI) metrics and the RMS-P2PD biomarker.

TECHNOLOGY

Short-axis cine cMRIs of 20 asymptomatic patients (MESA), and 25 symptomatic patients with coronary artery disease or left ventricle impairment (DETERMINE) were selected at random from the Cardiac Atlas Project database. To extract function metrics, the left ventricular endocardium and epicardium, as well as the right ventricular endocardium, were traced semi-automatically in each short-axis slice using Medviso Segment. Regional P2PD was established using an in-house shape analysis tool, by extracting left ventricular endocardial surface contours.

Table 1: Results after analyzing the subset of data with RMS-P2PD values

	No RMS/RMS		
	Accuracy	**Sensitivity**	**Specificity**
BRL	83.8/91.9	82.6/95.7	85.7/85.7
C4.5	89.1/97.3	82.6/100	100/92.8
JRip	94.6/94.6	91.3/95.7	100/93.0

RMS: Root mean square, BRL: Bayesian rule learning, P2PD: Phase to phase displacement

DESIGN

Supervised rule learning algorithms were utilized in this study, specifically, Bayesian Rule Learning, JRip, and C4.5. First, the entire study cohort was classified using only standard function metrics. Then, all three algorithms were run again on a subset of subjects for whom both function metrics and RMS values had been generated. The algorithms were run with and without RMS to determine whether RMS improved classification accuracy.

RESULTS

All rule learning methods successfully classified over 95 percent of patients over ten-fold cross validation. Additionally, all algorithms selected the same biomarkers to classify patients: left ventricular ejection fraction and end diastolic volume. The addition of RMS-P2PD as a marker consistently resulted in equal or better accuracy and improved sensitivity over cross-fold validation. In Table 1, bolded values denote equivalent or improved results through the addition of RMS-P2PD.

CONCLUSION

Analysis through rule learning algorithms indicates that standard cMRI metrics hold promise in the classification of cardiovascular disease. All three algorithms proved effective at classifying patients. RMS-P2PD shows merit as a cardiac function biomarker. Larger studies are needed to confirm these findings and generalize them for prospective diagnostic use.