Constantin F. Aliferis, MD, Ph.D.
Narrator

Dominique A. Tobbell, Ph.D.
Interviewer

INSTITUTE FOR HEALTH INFORMATICS
HISTORY PROJECT

UNIVERSITY OF MINNESOTA

INSTITUTE FOR HEALTH INFORMATICS HISTORY PROJECT

In 2015, the Institute for Health Informatics (IHI) celebrates the 50th anniversary of health informatics at the University of Minnesota. Early institutional markers serve as the formal beginnings of the emergent discipline of health informatics at the University of Minnesota, designating the University of Minnesota as one of the first academic institutions to support and subsequently anchor the development of the new discipline. In 1965, the National Institute of Health (NIH) Division of Research Resources awarded the University of Minnesota's College of Medical Sciences a grant to establish a Biomedical Data Processing Unit at the University. Two years later, the Hill Family Foundation awarded a ten-year grant to Professor Eugene Ackerman to initiate a graduate research and training program in Biomedical Computing. In 1968, the College of Medical Sciences established the Division of Health Computer Sciences, which would serve as the administrative home for the NIH research resources grant, housed within the Department of Laboratory Medicine. The Division provided interdisciplinary training to pre-doctoral and post-doctoral students applying health computer sciences technology to health services research. In 1974, the University of Minnesota was awarded the prestigious National Library of Medicine Grant for Training in Health Computer Sciences, which formally established the Graduate Program in Health Informatics at the University of Minnesota. The Division and its institutional successor, the Institute for Health Informatics (created in 2006), received continuous training grants from the National Library of Medicine until 2009. For fifty years, the University of Minnesota has been one of the preeminent health informatics institutions in the United States.

The Institute for Health Informatics History Project captures, analyzes, and records the history of health informatics at the University of Minnesota. Through oral history interviews, the Project preserves the personal stories of faculty members and National Library of Medicine administrators who were involved in the early history of the field and have keen insights into the history of health informatics at the University of Minnesota.

Biographical Sketch

Constantin Aliferis received an MD degree from Athens University in Greece in 1990, and an MS in 1994 and PhD in 1998 in Intelligent Systems from the University of Pittsburgh. After completing his PhD, Aliferis returned to Greece to complete two years of mandatory military service. At the same time he served as a research associate in epidemiology at Athens University. In 2000, Randy Miller and William Stead recruited Aliferis to Vanderbilt University to serve as an assistant professor of biomedical informatics and founding director of Vanderbilt University's MS/PhD Program in Biomedical Informatics. While at Vanderbilt, Aliferis was founding director of the Discovery Systems Laboratory (2001-2008), faculty member at the Vanderbilt-Ingram Cancer Center (2000-2008), assistant professor of cancer biology (2005-2008), assistant professor of computer science (2007-2008), and assistant professor of biostatistics (2007-2008). In October 2008, Aliferis was recruited to New York University to serve as founding director of NYU's Center for Health Informatics and Bioinformatics, Director of the Biomedical Informatics Core of NYU's Clinical and Translational Science Institute, and associate professor of pathology (with tenure); positions he held until left NYU in June 2015. While at NYU, Aliferis also held the following positions: associate professor, computational biology program, Sackler Institute of Graduate Biomedical Studies (2009-2015); scientific director, Best Practices Integrative Informatics Consulting Core (2009-2015); scientific director, High Performance Computing Facility (2009-2015); founding director, NYU PhD Program in Biomedical Informatics (2011-2015); director, Biomedical Informatics Shared Resource of the Cancer Institute (2012-2015); and faculty, NYU Center for Data Science (2013-2015). In June 2015, Aliferis was recruited to the University of Minnesota. At the University of Minnesota, he holds the positions of professor and director of the Institute for Health Informatics, and the first University of Minnesota chief research informatics officer, heading the Clinical and Translational Institute Biomedical Informatics program. Aliferis will also build and lead a Big Data Analytics Unit for MHealth (collectively the Academic Health Center, Fairview Health Systems, and University of Minnesota Physicians) and carry the title of chief analytics officer.

Dr. Aliferis' research is focused on high dimensional modeling and analysis designed to transform biomedical data into novel actionable scientific knowledge. His three key areas of broad interest are (a) use of advanced informatics and analytics to accelerate and enhance the sophistication, volume, quality, and reproducibility of scientific research; (b) quality and cost improvements of healthcare using Big Data approaches; and (c) precision medicine.

Dr. Aliferis was elected as a fellow of the American College of Medical Informatics in 2007.

Interview Abstract

Constantin Aliferis begins by discussing his educational background, including his early interest in biomedical and health informatics. He describes the main focus of his research since graduate school, which has included machine learning and the analysis of complex and high-dimensional data sets; scientometrics and informatics retrieval; and model building, analysis, and knowledge discovery across a variety of disease domains. Aliferis goes on to briefly discuss his tenure at Vanderbilt University, followed by a more detailed discussion of his tenure at New York University. Next, Aliferis offers his definition of precision medicine. The remainder of the interview focuses on health informatics at the University of Minnesota. Aliferis describes his vision for the Institute for Health Informatics, reflects on the strong backing provided by the leadership of the University and the University's Academic Health Center to support this vision, and offers his perspective on the future of the field of biomedical and health informatics.

Interview with Doctor Constantin F. Aliferis

Interviewed by Dominique Tobbell, Oral Historian

Interviewed for the University of Minnesota
Health Informatics Project

Interviewed in Doctor Aliferis' Office
330 Diehl Hall, University of Minnesota Campus

Interviewed on June 8, 2015

Constantin Aliferis - CA
Dominique Tobbell - DT

DT: This is Dominique Tobbell. I'm here with Constantin Aliferis. I'm in Doctor Aliferis' office in 330 Diehl Hall. It is June 8, 2015.

To get us started, could you just tell me a little bit about your educational background?

CA: I was educated in medicine in Athens University School of Medicine. Once I completed my MD there, I decided to come to the United States to pursue postgraduate training in biomedical informatics. Specifically, I enrolled in the intelligent systems studies program at the University of Pittsburgh, which had one of the most well-known programs in medical artificial intelligence at the time. I did my master's and Ph.D. and a research postdoctoral fellowship specializing in artificial intelligence in medicine and machine learning.

DT: What led you to pursue your career in biomedical informatics?

CA: Initially, as a high school student, I was very good in a variety of subjects, so I could go and study a number of fields. My personal trend was towards physics and computation. I had a discussion with my parents who said that maybe from the standpoint of professional development and societal impact and value, I should take a close look at medicine as a profession. After I did that, I decided to go into medicine. Also, there were some personal family circumstances, persons in my immediate family that were terribly sick and ended up dying that really influenced my decision making: devoting my career to actually helping patients in any way that I could.

DT: What led you to continue your studies after you finished your medical degree and come to the United States?

CA: That's an interesting story. When I say it, most people are entertained quite a bit by how it's framed. When I entered medical school, I was just shocked by how unscientific medicine appeared to me. Mind you, that was the mid 1980s. Things like evidenced-based practice, standards of care guidelines, and other sets of quality oriented measures were not very widespread. The discipline and practice of evidence-based medicine came to wide recognition in the early and mid 1990s in the United States and elsewhere. So, at the time, things were very pre-scientific, if I can use this word. I started reading and I came across formal medical decision making theory, expert systems, and other data-driven approaches to making diagnosis and deciding optimal therapies. I figured that if one can program computers to make diagnosis and to make the right choice of treatment, then one has really understood deeply, on a scientific basis, medicine. That was my thinking, toward a more scientific understanding of medicine.

So I started working with a number of folks in Athens who were doing medical informatics and epidemiology and biostatistics. We published a number of papers. It was a great dress rehearsal, so to speak, for my subsequent career as a medical informaticist.

I understood, also, that unless I would come to the mecca of medical informatics, which has been the United States historically, I would never fulfill my full potential as biomedical informaticist.

DT: What, if you can tell me, has your research focused on?

CA: The research has focused on two levels. One is the methodological level. The other one is the application level. Methodologically, it's all about machine learning and the analysis of complex datasets and high-dimensional data sets. The specific areas are predictive modeling, causal modeling and discovery, both at the algorithm and the protocol and system levels. I focused quite a bit on what's known as Markov Blanket induction, which is a particular family of techniques for dealing with very, very large and high-dimensional datasets to figure out what's the smallest set of information that one can extract and be able to do diagnosis and prognosis and, also, to learn about the causal structure of the process that generates the data. I've done quite a bit of work on automated analysis systems known as auto modelers or intelligent front ends to analytics, also work in support vector machines, and Bayesian network methods for causal structure discovery, and for classification or regression. I've dealt with the issues of how do you model time and causality in order to provide decision support and so forth. I also looked at some fundamental issues of scientometrics and information retrieval as they apply in biomedicine.

Specifically, my students and I have worked on understanding what are the factors of determining citeability and impact, how one can construct computer systems that can scan the literature to find the highest quality papers, to extract their contents, to predict

their future citations, to scan the health web to find out what pages are providing validated and up-to-date medical advice to patients and their proxies as compared to websites that are pure quackery or out of date and dangerous basically for public health.

On the application level, I had the opportunity to collaborate with many colleagues from a variety of disciplines and to build models and do analysis and discovery in many disease domains, including various cancers, pneumonia, osteoarthritis, rheumatoid arthritis, infant sepsis, acute respiratory distress syndrome, psoriasis, respiratory infections, and, also, in a variety of application domains such as microbiomics, cancer therapeutics, risk modeling for patients. There are notable projects that come to mind. I don't want to take too much time, just discussing those in the context of this reply but I've been involved in many disciplines. That's also reflected through the fact that I had appointments in so many different disciplines throughout the years. Because the techniques that myself and my colleagues have been developing for analyzing data, modeling, and making discoveries are so widely applicable, I ended up engaging as a faculty with primary or joint appointments across many areas and I continue to work with many colleagues from many different disciplines.

DT: It really speaks to the interdisciplinarity of health informatics.

CA: Absolutely. Yes.

DT: It sounds like you've covered a whole range of areas and done so much research in so many areas.

CA: Yes, related to research in lung cancer, in osteoarthritis, in psoriasis, in diabetes, in stroke, in sepsis. It's endless. The range of applications is just endless.

DT: Am I correct in thinking that you—I think I got this from your bio on the website—work with "Big Data."

CA: Yes, back then, we were not calling it "Big Data." I could give you a little bit of historical context here. In the 1990s—I think this is an illustrative historical example—I remember as a grad student, I was involved in a project that was a collaboration between the University of Pittsburgh and CMU [Carnegie Mellon University]. There were four labs, two from the University of Pittsburgh and two labs from CMU. Very famous researchers, Bruce Buchanan, one of the founders of the field of biomedical informatics and computational biology, and [Gregory F.] Greg Cooper, the inventor of Bayesian networks and learning and many Bayesian network algorithms, were from the Pittsburgh side of things. From CMU, Tom [M.] Mitchell, who was for many years department chair of Machine Learning, one of the most successful departments of machine learning in the world, and the infamous [Peter] Spirets, [Clark] Glymor, [Richard] Scheines group for causal discovery, so really a first-rate collaboration. The goal was to contribute to the Pneumonia PORT [Patient Outcomes Research Team] project, which was a set of studies around creating risk models for community-wide pneumonia, dire outcomes, specifically to predict the risk for dying within one year of discharge. The Pneumonia PORT models

are being used very widely today in the United States and they're improving the quality of care for patients and the reduced in cost is just dramatic. It has been estimated back in the 1990s it was several billions back then when the cost of actually treating pneumonia was maybe ten times less than what it is today, so a very high impact type of project.

I'm saying this because among other things, I was doing causal graph discovery. I was doing some neural networks, but I was also acting as the head data curator and facilitator of data sharing and so forth for everybody in the project. The working dataset had 220 variables. I remember distinctly that the CMU folks coming from this highly distinguished machine learning environment did not have the tools to load a 220 variable dataset and find out variable number, 115, for example. They were using a word processor called Emacs. They had to scroll through counting columns until they could find the desired variables. That was twenty years ago, basically. Now, in our era, we're taking in high school seniors who are doing, maybe, their first internship in a scientific lab. They may do one week or two weeks of training and we give them data and tools to analyze gene expression array data with 50,000 variables and they are able to produce very accurate diagnosis and treatment recommendation models with the tools that we give them in a week or two. This gives you a sense of how things have dramatically improved in the ability to analyze really large and complicated datasets.

Today, we speak about "Big Data." Back then this term was not widely used. It's a very big discussion what "Big Data" is and exactly what makes it a really incredible field of activity and why I think NIH [National Institutes of Health] very, very wisely has chosen to define the field and support it. I'm not going to do it right now, unless you ask me for more details later.

DT: [chuckles]

CA: So, yes, I've been doing this for a very long time, as I said, twenty-five years now. But it was not called that back then.

DT: You said that after you finished your Ph.D., you went back to Greece for a couple of years.

CA: Yes. A consequence of being a Greek male is that you have to go to the army. You have to serve a mandatory military service. So I had to go there. At the same time, I was also a research scientist with the Department of Epidemiology and Biostatistics. I was very, very lucky because the chair of that department was a distinguished epidemiologist who happened to be, also, for many years, full professor and department chair of biostatistics at Harvard University holding the chairs in Athens and Harvard simultaneously, Dimitrios Trichopoulos, who, unfortunately, passed away last year. It was a big loss for all the students and all his collaborators, and science in general. He was a remarkable epidemiologist and statistician. My years in the Greek military did not really detract too much from my ability to do science and engage productively.

DT: That is fortunate.

Then, you moved to Vanderbilt University.

CA: Yes, in late May, early June of 2000. I was recruited by Vanderbilt University, by Randy Miller and Bill Stead specifically to who I am grateful for this. At the time, there was a division of biomedical informatics that very soon became a department. Now, it's one of the biggest departments in the world and, certainly, in the United States. It has, if I'm not mistaken, about sixty-plus core faculty. It exists in a very big informatics center and is very widely known about the quality and the scope of clinical informatics. In later years, it is also expanding very productively in translational bioinformatics.

What I was hired to do there was to address a number of goals. One was help them address both clinical informatics problems and bioinformatics problems using machine learning. Another one was that at the time, they had some trouble getting their Ph.D. training program off the ground internally, because it had been rejected internally—this is public knowledge what I describe here; it's not any private information—and, also, they had failed to get an NLM [National Library of Medicine] training grant at the time. They entrusted me with being the director of that training program, to try to get it through internally and then to use it to get the NLM training grant, which we were able to do. That was a very big success, and a great team effort. It ended up being one of the best programs in terms of the quality of its graduates and its broader recognition. Most, if not all, of the people who graduate from Vanderbilt now have become notable faculty being in academia and a few of them have become leaders in industry. So it's a really extraordinary program.

I also launched a lab, the Discovery Systems Laboratory, which was like a small division, in that sense, because it had also a few faculty in it and a number of students and staff. Our goal was to develop next generation algorithms for big data analytics, both on the causal side of things and the predictive modeling side of things. We were very, very lucky because we were able to do some breakthrough discoveries that were considered very important and at the same time very hard to achieve at the time.

That's my tenure at Vanderbilt in brief.

DT: How long were you there?

CA: Until 2008.

I moved then to NYU [New York University] where they hired me to be the founding director of the Center for Health Informatics and Bioinformatics. At NYU, I led the Center. I built a faculty base. The Center provided an academic home for informatics. We created informatics specific criteria for recruiting, appointing, reappointing, and tenuring faculty. I was also the leader of the informatics for the NYU CTSA [Clinical and Transitional Science Award] and the leader of informatics for the Cancer Institute. I supervised, from a scientific perspective, the high performance computing facility, which we built in-house. We also built a new training program in biomedical informatics. We

had the Ph.D. training program, which was giving also a master's on route to the Ph.D., as well as a very, very new terminal master's program, which was approved just before I left NYU. We also built an integrated consulting service that enabled hundreds of scientists to do frontier research. We also provided some secondary cores for other departments and/or large scale grant projects. We also had a portfolio of healthcare analytics that were more operationally oriented. We also launched seven research labs in total that were producing new methods in informatics. These ended up having quite a bit of impact actually, in terms of total footprint in terms of users of the methods produced by NYU informatics faculty. It was at 70,000 users in 120 countries, which is a very big number, certainly, I would say, one of the biggest in the country. This was the core activity of the Informatics Center at NYU.

DT: It's interesting that it's a center for both health informatics and bioinformatics. As I understand it, in a lot of places, those two are separate.

CA: Yes, there are two schools of thought about it. I think what the NIH really has emphasized over the years is the integrity of the field and the continuity of the spectrum of informatics. That's why we call the field biomedical informatics,to emphasize that there is this basic science end of it and, then, there is the translational part. There is the medical part. All of these have some continuity. That's when it becomes a very powerful as a field. If you try to isolate it and modularize it and artificially impose thresholds and boundaries, you cannot be as effective as a scientist or as an educator. Sometimes, artificially creating these boundaries has been a very historical consequence of how some units started and evolved and who was leading them, a political issue more than a scientific or technical one. At NYU, we sought, from the beginning, to preserve the continuity of the field and to build it as a continuum rather than in pieces. Of course, I would also argue that it makes perfect sense to have separate computational biology programs for the reason that computational biology is basically, in my mind at least, an offshoot of biology. It's truly an enabler for biology. But when you're dealing with the medical domain with patients with translational research, it's absolutely important to have the blending of the clinical informatics and the methodologically oriented and the translationally oriented informaticists all under the same roof. If we think about it, clinical informatics will include all of health informatics as well as the translational bioinformatics, because there is an explosively growing and fast integration between genomics, proteomics, all kinds of omics, and molecular medicine with clinical practice. The future will only become more and more integrated up to the point that it will be just completely seamless. So this cannot happen if you have separated entities.

DT: At Vanderbilt was bioinformatics integrated with clinical informatics?

CA: Yes. It was in the same realm. Absolutely. That was formally solidified with the recruitment of Dan Masys to lead the Department of Biomedical Informatics. Dan historically was the founding director of the NCBI [National Center for Biotechnology Information]. Yes, it was totally integrated.

DT: What led to your decision to join the University of Minnesota?

CA: We are very, very lucky as informaticists to live in this particular era because, with precision medicine on one hand and "Big Data" and healthcare analytics on the other hand, we're given the opportunity to have the biggest possible impact ever in the history of science and medicine and the health-related disciplines. This has created a very competitive environment where a lot of big investments have been made and many, many excellent universities are competing for talent. They're building units and they're controlling data resources and expanding them at a very rapid pace. At NYU, for a variety of reasons—some of them had to do with the historical trends of who was doing the leading research but, also, events that had to do with flooding twice in two years, two consecutive years, with more than a billion dollars worth of damages—the university found itself at a spot that it could not really compete in building the necessary infrastructure beyond what we did there in the first years. So a lot of folks started looking elsewhere at other opportunities and, like I said, because of the competiveness of the environment, it made sense.

One thing that challenged me personally was that the institution did not believe in personalized medicine as a strategic direction in its science strategy agenda and was not, also, making inroads in healthcare transformation using "Big Data" informatics and process. These are exactly the things that I love doing. Because I think that's my biggest opportunity to make the biggest possible contribution as a scientist and as an academic leader and administrator, I felt I should basically explore opportunity at places that had exactly the same vision.

So I'm very happy to say that here at the University of Minnesota, the vision of the leadership, which is precisely precision medicine and the needs of the research community for informatics to really enable them to do all the amazing projects that they want to do, but sometimes is difficult because they don't have all the necessary support in terms of high caliber informatics colleagues, creates for me an incredible opportunity to make a very, very big difference. Clearly, the University of Minnesota is one of *the* major institutions in health sciences and beyond…a huge history, a huge track record of great achievements. It's an academic environment of the highest caliber. At the same time, the vision of the leadership here is totally in the right direction in my opinion. The needs are huge, so we have the opportunity to come in and make a difference. I have to say that they also were very, very thoughtful in making an offer that would enable me and my team to really accomplish those things; although, as I said, the competition is stiff and there's a lot of players out there. They all want good informaticists. I was very lucky enough to also experience interest and offers from many places. I think I absolutely made the right choice by coming here and seeing what we can do here. I think it's going to be amazing, actually—and that's on the record.

DT: [laughter] It's there forever.

CA: Yes.

DT: Can you define precision medicine?

CA: Yes. Precision medicine is sometimes also called personalized medicine or individualized medicine. There are nuances in these things, but, at the end of the day, it means all of these activities: the scientific research, the technology and the organizational approaches and care workflows that are needed to deliver the right type of treatment to the right patient at the right time, not just on the basis of curing disease or alleviating symptoms, but also in terms of prevention in the context of acute disease and also preventing acute exigencies in chronic diseases. It has to do with producing a new generation of drugs that are highly rationally develop and they are targeting particular molecules, particular pathways because they reflect a deep mechanistic and causal scientific understanding of what is causing that particular disease that one is attempting to prevent or cure and, also, to understand how a particular individual is different than the whole population of individuals that are labeled under the same disease category with the express intent to maximize the benefits of treatment and minimize the risks of treatment, minimizing the toxicities, improving the outcomes, improving the patient's quality of life, making also a significant impact on the accessibility and the cost of care. All of these things are part of the same continuum of challenges and processes, the healthcare and biological process of disease in the whole of society, actually. They're all intertwined and connected with one another.

If you talk to ten different folks about what is personalized medicine to you, depending on their background, they will probably give ten different answers. There are many, many distinct ways to see threads under the same umbrella of personalized or precision medicine.

One of them has to do with molecular profiling, for example, for doing advanced diagnosis or prognostication about the outcome of the disease without treatment and post treatment. For example, there are some wonderful modalities for choosing whether breast cancer patients should get chemotherapy in addition to excising a tumor of early stage. A seminal example is that of the Oncotype Dx which is a molecular profile that is helping patients to avoid the unnecessary toxicities or unnecessary chemotherapy and receiving chemotherapy when there's a highly probability of future metastasis. That's improving survival. Also, they have great positive effects on the cost of care. There is a whole array of such modalities that are being developed and tested and actually reaching patients. We expect to see thousands of those in the next years. Molecular profiling is one way to do personalized medicine.

Another way is sequencing solid tumors. There have been both academic and commercial processes to this, but, basically, the idea is that if you have a tumor that does not respond to conventional chemotherapy by sequencing them and using the appropriate informatics models and methodologies and tools, one can identify mutations that are driving the dominant tumor clones in the solid tumor and, therefore, choosing from a menu of available targeted treatments, thereby converting a hopeless case, one that is guaranteed to lead to the demise of the patient, into possibly saving the patient. Unfortunately, with our limited range of treatments is a small ratio right now, but it's increasing and is going to get better and better and better.

Other ways to think about personalized medicine in the clinical realm is to risk-profile patients, by separating patients into high risk and low risk groups. One can make better admission decisions, better discharge decisions, better treatment decisions. All of this translates not just into improving the quality of the outcomes but also improving the cost of care, the investment, the financial sustainability of the healthcare systems, and so forth.

There are many additional ways. I won't go into all different types and pathways of personalized and precision medicine. Actually, on the white board behind me, you can see several of them drawn out because we were having this discussion with many colleagues the last few days.

[chuckles]

CA: This is just a huge space that leads to better treatments by looking at the individual characteristics of patients and developing a very, very deep understanding of disease on a molecular level.

DT: That's actually a fuller description. I know personalized medicine and I teach a little bit about it, but that's a much broader definition than I've heard and that goes to your point, that depending on who you talk to, you get a different definition.

CA: Yes, and I think the solution to this is just to accept that there are many, many different versions and all of it needs to be developed. They're all beneficial. If someone says that we want to develop a program in precision medicine or personalized medicine, that's as inclusive as you can imagine. It's not something that is exclusive. It is not like we do immunology or we don't do immunology. You can be an immunologist and have a wonderful program in personalized medicine or you can be a non-immunologist and have a wonderful program in personalized medicine. There are so just so many paths to the same ultimate goal for everyone regardless of their background or specific interests.

DT: I like this definition. Yes, it's a good one. I'll stick with this one.

[chuckles]

CA: Okay.

DT: What is your vision for the Institute for Health Informatics [IHI]?

CA: At the core of this vision are a number of elements.

Number one, we want to contribute to excellence in science. Once you achieve excellence in the science of informatics but also of science that's *enabled* by informatics, everything else follows. So the grants follow. The finances then follow. The growth

follows. The good career trajectories for the people involved follow and all of that. Excellence is very, very important.

The second one is educating the leaders of informatics of tomorrow.

The third one is educating in core informatics skills and knowledge everybody regardless of what is their background. Informatics is so universal right now that it's like statistics or mathematics. Every scientist needs to know core concepts and methods and tools in the field just to be conversant, just to understand the papers that he or she reads.

We want to make informatics into an engine for expanding the scope, the volume, the sophistication, and the impact of science across the [Academic] Health Center and the University. There are specific ways to do it. One major priority is to recruit a number of high-caliber, high-yield, and very collaboratively oriented informatics stars to come aboard and engage with the broader research community, providing for the research community the highest level of informatics, collaborative science, and expert consulting so that they can really push the frontier of what's possible in terms of research. We are working to create a one-stop shop integrative best practice oriented informatics consulting and collaborative science core that will provide informatics expert consulting and collaborative science across the spectrum of informatics to all engaged researchers.

So these are some of the very high priority targets: recruitments, integrated informatics services, best practices, toward the goal of excellence and impact for patients in society and, of course, the University as an agent of those changes for the benefit of everyone.

DT: With the teaching of core informatics skills to beyond the informatics folks, how do you see that taking place?

CA: In many ways. Clearly, there are so many educational tools one can deploy. Of course, mapping the different tools to the learning styles and learning needs of various trainees is part of setting up oneself for success. I can explain by example better than providing some general principle other than what I just said.

For example, in my prior life, we had a required course in bioinformatics for every Ph.D. student in the biological science program. So everybody will take that course so everybody would be conversant with the core techniques and core concepts.

Another way to do this is we had a number of seminars and workshops on topics that would be of broad usefulness.

A third one is to provide mentoring experiences. I'm talking now about teaching—if I understand your question correctly—informaticists and non-professionals in informatics. Provide short term mentoring, helping, for example, post docs in specific labs to come to terms with what are the tools and what are the core techniques that are relevant to that particular lab's long term research.

A fourth one is to develop customized training programs and career development programs for faculty, early career faculty, junior faculty, who are interested in getting a K [Career Development] award or something like this or more generally for pursuing a career that is to a large extent depending on having a solid understanding of informatics, not being informaticists but requiring informatics. For example, I have been a mentor to faculty who had an interest in modeling very complicated data using core techniques in psychiatry or other folks who had an interest in microbiomics. These kinds of faculty are not informaticists but they really need a deep understanding of a lot of informatics methods. The way we did that was we would develop mentored K award programs. We would develop a post doctoral fellowship. In some case, we will ask them to get a master's in informatics and so forth and, in some cases, not formal training but deep immersion in informatics with proper mentoring.

National and regional workshops, conferences, and tutorials are also helping a lot. Certificate programs, long distance training, the tools are endless. The question is what are the characteristics of them and how do you enable the recipient of this training to find the right training tool and opportunity for them?

DT: That sounds great.

Has the leadership of the Academic Health Center provided money so that you can recruit these informatics stars?

CA: Yes, they have; otherwise, I would not be here, actually.

DT: [chuckles]

CA: I think the leadership understands the importance of this, that people are the most important asset. They have provided resources for these but they, also, very wisely, understand that partnerships need to be built and collaborative links and bridges need to be built, by doing joint hires and basically driving recruitments according to the needs of the community rather just forming a build-it-and-they-will-come—strictly speaking—type of approach. You need to do some of that, because you need to really look into the future and be prepared for the future. You just cannot do this by being reactive all the time. You also have to drive some of your recruitments in the current strategy around existing needs and making sure that these needs are going to be fulfilled and met.

DT: Do you have a kind of timeline for how long this recruitment process…? Are you thinking in several years?

CA: Yes, the recruitment process will be ongoing because the needs are going to be ongoing and the new technologies and new opportunities will keep showing up if the past is any indication of the future. But I think we'll need to be very, very aggressive and to move fast and, to use an expression I use in my discussions lately, move in New York time…

DT: [chuckles]

CA: …making some core recruitments as fast as possible within the next half a year, so that we can really build some critical mass, and really showcase to the community what informatics can do for them, really build around the level and quality of those new collaborations and services, and, then, it will be a more measured and more deliberate and more relaxed approach to building the rest of the new faculty base. The world is spinning very fast. It's not going to stop for us and wait until we are ready,

DT: [laughter]

CA: Our competitors are moving very fast. They are building amazing organizations and shops. If we want to be competitive, we have to be aggressive, not reckless—I'm not thinking, obviously, of that—to utilize our personal influence and networks and connections and knowledge of where the stars are and what we will need to do in order to make it work so that we can recruit these kinds of folks.

DT: Where do you see as the relationship between the Institute and the CTSI [Clinical and Translational Science Institute] being?

CA: Technically speaking, there is a special informatics unit within the CTSI which is called the BMI, biomedical informatics program, which I lead under my capacity as chief research informatics officer. The IHI is kind of an organizationally separate entity but, in reality, all of this is just synergistic and symbiotic. The CTSI is a great ally in building informatics for the institution and informatics is a great ally for CTSI. My goal, fully aligned with Bruce Blazar (the CTSI Director), is to make sure that the CTSI grows and is extremely successful and becomes an engine for success and be a success for everybody. But I do not forget that at any time our measure of success will be the broad impact, not just the CTSI which is a major, major achievement and major undertaking but also to make sure that *through* the CTSI and through everything that we're doing informatics can really be an agent of progress and change for the benefit of everyone.

DT: Where do you see the future of the field of biomedical health informatics going?

CA: I think the future is just spectacularly amazing. I don't have the words actually to explain how promising the feeling seems to me. I think that the two major, major areas, aside from being an engine for productivity and quality of science, which I described at length previously, is enabling the vision of personalized medicine.

If we believe what all of these papers are saying and all of this prior research has shown and what the president of the United States has promised, personalized medicine, precision medicine will be the new human genome project, the new "race to the moon" on one hand, and all of this is just enabled and supported clinically by informatics, just cannot do it without high-ended informatics. That's very, very simply put. A lot of this is driven by the new assay technologies. All of them require informatics for execution

both at the assay level and at the level of hypothesis testing, hypothesis generation, interpretation of findings and pushing forward the science.

Then, on the other hand, we have all of the transformational healthcare, the "Big Data" approaches and, in general, data decision making and the reorganization of healthcare to achieve the goals of reduced costs, increased quality, improved access, and population wellness and health. There, informatics is just again, absolutely essential and critical.

So either way you look at it, either from the translational science and the personalized medicine or from the healthcare system and delivery of care perspectives, informatics is positioned to be really absolutely critical. As amazing as the past contributions in informatics have been, the new ones are going to be even bigger.

Let me very briefly mention some of these past contributions. I don't think that many people realize how important informatics has been. If you think about systems and methodologies like BLAST [Basic Local Alignment Search Tool] for doing sequence alignment, these are being used by every biologist on the planet daily almost. If you think about the shotgun sequencing that enabled the human genome project, if you think about PubMed and the other bibliographic databases that allow people to connect with scientific discoveries of their colleagues at a moment's notice, if you think about computational guidelines and other formal medical decision making theories in providing provider order entry systems, electronic health records, Bayesian networks, and other advanced technologies for decision making, there is a very big list of *monumental* contributions of informatics in science, especially the health sciences. So as big as these are and as important as they are, I think the future is even more promising.

DT: That's great.

My final question would be if you have any comments on what you think may be distinctive about the program here at Minnesota.

CA: In its current form or in its future?

DT: Either.

CA: Minnesota has a long history of informatics. I'm looking in front of me at a document that describes the celebration of fifty years of informatics at the University of Minnesota. Very few places have this institutional history and memory and deep understanding of the importance of informatics. That's a very, very important asset. Of course, there are notable figures in informatics in the United States that either were faculty here or are graduates of the training program and so forth. Another very important element is that we are having a number of good data access. We have a portfolio of data that constitutes a very rich data environment that can be put to good use for discovery and improving healthcare. We also have a range of training programs. We also have a range of people who are either active faculty or affiliate faculty and so forth,

Now, the challenge for the future will be to really focus more on the discovery side of things and to make sure that now we can pass the baton from access to data to major discoveries using this data and, also, to integrate more tightly with the research enterprise so that we can become participants and critical collaborators in the production of new scientific knowledge and to push forward projects that without the sophistication of informatics cannot be executed as of now, at least not easily.

So there is a rich tradition in its history and I think the future can be extraordinarily rewarding by emphasizing what I just mentioned.

DT: Great. Is there anything else that you would like to say for the record?

CA: I'll say one thing. I hope that two years from now, we will revisit this interview and we'll say, "Everything we said back then came to be." That will be certainly rewarding.

DT: Yes. Let's come back in two years. [chuckles]

CA: Okay.

DT: Thank you.

CA: I thank you very much.

[End of the Interview]

Transcribed by Beverly Hermes
Hermes Transcribing & Research Service
12617 Fairgreen Avenue, St. Paul, Minnesota, 55124
952-953-0730 bhermes1@aol.com