University of North Carolina at Chapel Hill Certificate Program Application Form

Please use this application form as a guide for your Certificate Program proposal.

Name of Proposed Certificate Program: Biomedical Big Data to Knowledge (BD2K)

Sponsoring Academic Unit: Biostatistics (primary contact unit) and Mathematics

Administering Unit, if different: <u>Biostatistics (primary administrative unit)</u> and <u>Mathematics</u>

Primary Contact Name: Michael R. Kosorok (primary contact) and M. Gregory Forest

Address and CB #: <u>3101 McGavran-Greenberg Hall, CB 7420</u>

Phone Number: <u>919-966-8107</u> Email: <u>kosorok@unc.edu</u> (primary) and <u>forest@unc.edu</u>

First Term the Certificate Would be Offered: Spring 2017

- **1.** Describe the Certificate Program and provide a statement of educational objectives. This certificate program is designed to prepare primarily doctoral students in biomedical and quantitative sciences for interdisciplinary data science research that incorporates and integrates four domain areas: biomedical research, biostatistics/statistics, computer science/informatics, and mathematical modeling. Students in Master Degree programs will be individually considered. The successful student will be able to work in teams and contribute to team-based research that uses data, including Big Data, to solve challenging biomedical research questions. To accomplish this, students will be successful in their domain area for their doctoral degree, learn fundamental data science techniques, become acquainted with each of the four domain areas, and experience collaborating in multidisciplinary teams. This certificate is connected closely to the Big Data to Knowledge Training Program grant (T32 CA201159 and T32 LM012420) led by principal investigators Greg Forest (Department of Mathematics) and Michael Kosorok (Department of Biostatistics), both of whom will also jointly lead the certificate program. In addition to required coursework, students will participate in monthly BD2K seminar meetings and will collaborate in one or more interdisciplinary research teams leading to the submission of at least one manuscript or proceedings paper where the student is a co-author and/or at least one dissertation chapter will involve a significant, interdisciplinary biomedical data science component.
- 2. Include a statement about the need for such a Certificate Program, and specifically why there is a need to offer this Certificate at UNC-Chapel Hill. Is the Certificate offered at other universities or community colleges?

The National Institutes of Health has initiated a significant national program for training biomedical researchers in Big Data (the BD2K initiative). The goal is to train biomedical and quantitative researchers to collaborate in multi-domain, interdisciplinary teams to address

Completed and approved applications should be submitted to: Stephanie Schmitt, Associate Dean for Academics, <u>sschmitt@email.unc.edu</u>

and potentially solve important public health problems and biomedical challenges by learning the basics of Big Data science and being able to work across four disciplinary areas: biomedical science, biostatistics/statistics, computer science/informatics, and mathematical modeling/applied mathematics. Our BD2K training program was one of the first three training programs in the US to be funded by this new mechanism in May 2015 (the other two were UCLA and Columbia). This training program meets a hitherto unmet need. The purpose of the certificate is to recognize those doctoral students at UNC who have successfully completed Big Data to Knowledge training, including both funded and unfunded trainees. As far as we are aware there is no other systematic effort to train in the unique cross-disciplinary manner done in this training program, either at UNC or anywhere else in North Carolina or the US. We also have the enthusiastic support across the UNC campus of over twenty-five departments spanning four Schools (Information and Library Science, Medicine, Pharmacy, Public Health), the College of Arts & Sciences, and the Bioinformatics and Computational Biology PhD program, all with the commitment that many of their students—as well as their disciplines—will benefit from this training program and provide a tangible enhancement to their PhD as well as their future career.

3. Describe the demographics of the target student population for the Certificate Program. Double click each box that applies and describe the intended audience.

Undergraduate Students
Graduate Students
Professional Students: _____
Degree-seeking, Matriculated Students
Non-Degree-Seeking Students

4. Why is the Certificate Program necessary beyond offering the program as a minor, supporting area, or specialization/concentration/track?

The certificate program provides a truly trans-disciplinary experience in Big Data and should not be tied to a particular degree since there are many disciplines that must fully participate in and contribute to this program. An inclusive, campus-wide certificate program is the most suitable format for this. A UNC Graduate School BD2K in Biomedicine Certificate will enhance the PhD of programs across campus spanning diverse disciplines, and endow a valuable credential to recipients for a spectrum of jobs.

5. Provide specific courses and other requirements for the Certificate Program. Separate listings of courses may be included with the proposal. The certificate consists of course and module requirements, seminar and program activity participation, and an interdisciplinary biomedical data science research project:

Coursework: Each student will be required to take at least 9 credit hours of the following kinds of courses. One 3 credit hour course on biomedical data science within their degree program discipline; one 3 credit hour graduate-level course that is in a different domain than their degree field; and 3 one credit hour modules under the auspices of the BD2K training program involving a biomedical research area and at least two of the remaining three domain areas. The first two required courses must be approved by the BD2K in Biomedicine Program Directors, currently Dr. Forest and Dr. Kosorok. We attach a list of potential courses, but it is difficult to fully enumerate all possible courses since so many departments

are involved in the BD2K training program. We have attached several examples from several departments. We also include the one credit hour modules from Spring 2016 as well as planned modules in Spring 2017 as exemplars of current and future modules.

Seminar Participation: Students are required to regularly attend for at least two semesters the BD2K seminar co-led by Drs. Forest and Kosorok.

Research Project: Students are required to submit as a co-author one research manuscript or proceedings for peer review and/or include one dissertation chapter that represents a significant, interdisciplinary biomedical data science research project involving both biomedical research and at least one of the three quantitative domains (biostatistics/statistics, computer science/informatics, applied mathematics). These can arise from laboratory research, summer internships, or other relevant, collaborative research projects.

6. Provide a statement on the relationship of the Certificate Program to degree programs within the unit(s). To what extent will requirements for the Certificate overlap with requirements for bachelor's, master's or doctoral degrees? Confirm how course credit transfer policies will be applied to students.

The goal of the BD2K in Biomedicine Certificate is to *enhance* primary doctoral degrees across campus, with at most one course overlap toward the primary PhD (the first coursework requirement in the above list). The modules and BD2K seminar requirements are new contributions to the graduate curriculum across campus, whereas the multi-domain research project is a constraint specific to the BD2K training program.

7. Will the Certificate Program be offered jointly with another university? If yes, describe the relationship with the joint unit.

No.

8. Will the Certificate Program be offered on campus, as a distance education program, or a combination? Describe any distance education components in detail.

The program will be offered on campus with no planned distance education components.

9. Describe the admissions criteria and process in detail. Differentiate between processes for degree-seeking students and non-degree-seeking students, where applicable. Include information about residency for tuition purposes as needed.

All UNC graduate students enrolled / admitted in one of the participating doctoral programs (over 25 thus far) are eligible to participate in this Big Data to Knowledge training program funded by NIH (T32 CA201159 and T32 ML012420) and to seek the BD2K in Biomedicine Certificate. Student applications are submitted to and reviewed by Drs. Kosorok and Forest. Applications for full funding by the BD2K program are submitted each Spring and reviewed by Drs., Kosorok and Forest and a BD2K training committee (the committee currently consists of Drs. Penny Gordon-Larsen from Nutrition, David Gotz from the School of Information and Library Science, Weili Lin from Radiology and BRIC, Marc Niethammer from Computer Science, and Stephen Marron from the Department of Statistics and Operations Research) on the basis of their interest in Big Data and their background and interests in Biomedical Science, Biostatistics/Statistics, Computer Science/Informatics, and

Applied Mathematics, as well as their overall academic excellence. Students are admitted independently of funding by the BD2K program provided they have sufficient strength and interest in at least three out of the four given domains. Each Spring we select a diverse group of 6 or 7 funded trainees, as resources permit, spanning all domains and participating PhD programs, while leveraging additional support from the BD2K mentors for all other student funding. All student applications include a statement of purpose for pursuing the BD2K Certificate and how participation in the BD2K program is aligned with and will enhance their plan of study. Funded trainees must be US permanent residents (as per the NIH requirement). All students in the training program are eligible to participate in the certificate program. All students seeking the BD2K Certificate are required to sign a statement of understanding of the extra requirements beyond their PhD requirements.

The departments and doctoral programs participating in the training program include: Applied Physical Sciences, Biochemistry and Biophysics, Bioinformatics and Computational Biology, Biology, Biomedical Engineering, Biostatistics, Chemistry, Computer Science, Environmental Sciences and Engineering, Epidemiology, Genetics, Genetics and Molecular Biology, Information and Library Sciences, Mathematics, Medicine, Microbiology and Immunology, Neurology, Nutrition, Pediatrics, Pharmacology, Divisions within the Eshelman School of Pharmacy, Psychology and Neurosciences, Physics and Astronomy, Radiology, School of Dentistry, School of Information and Library Science, Sociology, Statistics and Operations Research, Surgery, and Toxicology.

10. Provide a three-year, semester-by-semester projection of enrollments and course offerings.

We anticipate that at steady-state new students will primarily enroll in the Fall of their first or second year, and will complete all required coursework and requirements by their third or fourth year. Thus far, students have enrolled at all stages since the program started May 2015, including entering students in the 2016 class who arrived early to participate in the Summer program, through to advanced students with 1 year remaining. Thus far, there are over 50 students across campus participating, most of whom aspire to complete requirements for the proposed BD2K Certificate. We conservatively project that we will have at least 10 new students enrolled each year. The modules and seminar will continue to be offered for the foreseeable future.

11. Provide a three-year projection of the Certificate Program's financial plan. Include the impact on campus resources, such as classrooms and instructional faculty/personnel. Also include plans for tuition and billing, if separate from standard rates across schools.

The funding for the funded trainees will continue to come from grants T32 CA201159 and T32 LM012420 (with Drs. Kosorok and Forest as co-PIs) through Spring 2020. The time provided by Drs. Kosorok and Forest as well as other mentors and module instructors we anticipate will continue to be funded indirectly as part of their usual faculty funding, since they view this training program and certificate to be consistent with their faculty mission. Some additional funding for team building activities is being provided by distinguished professorship funds provided to Drs. Kosorok and Forest, and the Department of Mathematics and the College of Arts and Sciences has allocated Phillips Hall 365, the former Math-Physics Help Center, as the physical location for the BD2K Graduate Training Program as the graduate training component of a pan-university research institute led by Dr. Forest.

12. List all faculty members who will be responsible for planning and participating in the Certificate Program. Programs are encouraged to provide advising for students through the identification of one faculty member as the director of the Certificate.

Drs. Kosorok and Forest will be co-directors of this program, with Dr. Kosorok serving as the main contact. There is a training committee, current members listed earlier, that serves the dual role as an advisory board. We plan to formalize the Advisory Board, including members external to UNC, in the Spring of 2017. We attach a list of mentors who have requested to be part of the BD2K in Biomedicine training program.

Describe the evaluation plans for the Certificate Program.

We will review success of the program through informal feedback from mentors, students, and instructors, as well as certificate completion rates, publications of peer-reviewed journal papers and conference proceedings, and successful placement of graduates and biomedical Big Data jobs upon graduation. There is a very intensive review process proscribed by NIH for T32 training grants that we will also follow.

Once a year, we will request annual reports on progress from each of the students enrolled, review with the BD2K T32 training committee (consisting of Drs. Forest and Kosorok along with five to six other faculty across the departments and schools involved in the program). We will then provide feedback on progress to the students, meet with them, and help them strategize as needed to maintain good progress. We will also reach out annually to each of the departments and programs involved and ask for their feedback on what is working and what is not working. Finally, we will also track first time job placements to help evaluate whether the program is preparing students for the right professional opportunities.

13. Appropriate letters of support should be included with the proposal. All units sponsoring and participating in a Certificate Program should approve the proposal and provide support letters, including letters from units supporting the Certificate through resources (e.g., faculty time, course slots). Approval letters from the home school should accompany the proposal submitted to the Graduate School for final approval.

Please see attached.



THE UNIVERSITY of North Carolina at Chapel Hill

DEPARTMENT OF SOCIOLOGY

CAMPUS BOX 3210 155 HAMILTON HALL CHAPEL HILL, NC 27599-3210 T 919.962.1007 F 919.962.7568 http://sociology.unc.edu

17 August 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Imital

Kenneth Andrews Professor and Chair, Department of Sociology



DEPARTMENT OF BIOLOGY COKER HALL CAMPUS BOX 3280 CHAPEL HILL, NC 27599-3280 T 919.962.2077 F 919.962.1625 biology.unc.edu

August 16, 2016

Michael R. Kosorok, Ph.D. W.R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences

University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Sincerely yours,

Victoria L Bautch, PhD Beverly Long Chapin Distinguished Professor and Chair of Biology



DEPARTMENT OF BIOSTATISTICS

CAMPUS BOX 7420 Chapel Hill, NC 27599-7420 T 919.966.7250 F 919.966.3804 www.sph.unc.edu/bios

October 4, 2016

JIANWEN CAI, PHD Interim Chair

AMY H. HERRING, ScD Interim Vice Chair

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

januen Car

Jainwen Cai, PhD Cary C. Boshamer Distinguished Professor and Interim Chair of Biostatistics



OFFICE OF THE DEAN

205 SOUTH BUILDING CAMPUS BOX 3100 CHAPEL HILL, NC 27599-3100

T 919.962.1165 F 919.962.2408 college.unc.edu

August 25, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Sincerely,

hris Clemens Senior Associate Dean for Natural Science

Jaroslav Folda Distinguished Professor of Physics and Astronomy





Donna B. Gilleskie, PhD Professor donna gilleskie@unc.edu http://www.unc.edu/~dgill/

August 20, 2016

To Drs. Michael Kosorok and Greg Forest:

We are writing this letter to express our support for and admiration of the NIH-funded Big Data to Knowledge (BD2K) in Biomedicine graduate training program you have implemented at the University of North Carolina at Chapel Hill. I am a full professor in the Department of Economics at UNC-CH, and my research is in the area of health economics. I am currently working with Dr. Maria Ferris from the UNC School of Medicine to study successful transition of chronically ill adolescents from pediatric care providers to adult care providers. One of my Ph.D. graduate students in Economics, Yi Zhong, is working as a research assistant and co-author with us on several studies using longitudinal data on patient transition readiness measured since 2006. Recently we obtained access to medical record data spanning the patient's initial contact with UNC (as early as 1984) to present. Needless to say, we presented Yi with the tremendous task of cleaning these data and making them usable for our longitudinal transition study.

We heard about the training program, and thought it was worth exploring given the size and complexity of our data. We had no idea how much our research assistant would benefit nor what could be accomplished in one summer (and beyond). The BD2K program encouraged trainees to work in groups on a project of interest to them. Our project attracted four additional graduate students pursuing degrees in Epidemiology, Biostatistics, Rehabilitation Counseling, and Statistics and Operations Research.

The funded multi-disciplinary graduate students worked initially to clean the detailed medical record data to create a unique, useable, longitudinal history of a patient's health. The raw data included millions of observations on dated patient level encounters (scheduled physician visits, outpatient visits, inpatient stays, ER visits), diagnosis conditions, vitals and lab results on important markers of disease progression, and prescriptions and refills, as well as charge information at the procedure level. Cleaning these data was accomplished four (or more) times faster given the coordination within the group and the contributions from different perspectives.

The diverse backgrounds of the students have complemented our initial desires for obtaining these data. We intend to develop analysis models and interpret this new information in conjunction with measured transition readiness analyses as well as to better understand the disease process. In fact, the funding provided by your program has allowed our group to advance the science of a newer field in health, the "Life course of pediatric-onset chronic conditions." The understanding afforded by the breadth of these data will

guide professionals of multiple disciplines to identify more effective ways to improve patient outcomes.

We are grateful for the opportunities and training your BD2K program has provided.

Donna B. Hillissie

Donna Gilleskie Professor Department of Economics UNC-Chapel Hill

milie

Maria Ferris, MD, MPH, PhD Associate Professor School of Medicine UNC-Chapel Hill



THE UNIVERSITY of North Carolina at Chapel Hill

DEPARTMENT OF PSYCHOLOGY AND NEUROSCIENCE

DAVIE HALL CAMPUS BOX 3270 CHAPEL HILL, NC 27599-3270

T 919.843.5467 F 919.962.2537 psychology.unc.edu

DONALD T. LYSLE, Ph.D. KENAN PROFESSOR AND CHAIR

August 24, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Dr. Kosorok and Dr. Forest,

The Department of Psychology and Neuroscience is enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine.

We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. Dr. Kathleen Gates, Assistant Professor of Quantitative Psychology, has further agreed to serve as a mentor for students involved in the BD2K in Biomedicine training program. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Jonald T. tayle

Donald T. Lysle, Ph.D., Chair Kenan Distinguished Professor



DEPARTMENT OF ENDODONTICS

1098 FIRST DENTAL BUILDING CAMPUS BOX 7450 CHAPEL HILL, NC 27599-7450 T 919-537-3403 F 919-537-3408 www.dentistry.unc.edu

August 16, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

I am very supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. I would be very interested in collaborating with you on your successful BD2K T32 training program that would benefit many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

I also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. I look forward to working with you on this important certificate and training program.

Ashraf F. Fouad, DDS, MS Freedland Distinguished Professor and Chair



THE UNIVERSITY of North Carolina at Chapel Hill

DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY

125 MASON FARM ROAD 6TH FLOOR MARSICO HALL CAMPUS BOX 7290 CHAPEL HILL, NC 27599-7290

T 919.966.9580 F 919.962.8103 goldman@med.unc.edu

August 21, 2016

WILLIAM E. GOLDMAN, PH. D. *Professor and Chair of Microbiology and Immunology*

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

This letter is to confirm my Department's enthusiastic support of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

William E. Goldman

William E. Goldman, Ph.D. Professor and Chair of Microbiology and Immunology



SCHOOL OF MEDICINE

Campus Box 7310 T 919.962.8400 Chapel Hill, NC 27599 F 919.966.9863

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

August 17th, 2016

On behalf of the Curriculum in Toxicology (CiT) and as a mentor to several students, I am enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC, including the CiT. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

The CiT also feels that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Sincerely,

Ilona Jaspers, PhD Director; Curriculum in Toxicology Professor; Departments of Pediatrics, Microbiology & Immunology, and Env. Sci. & Engineering Deputy Director; Center for Env. Med. Asthma & Lung Biology University of North Carolina at Chapel Hill





August 16, 2016

DEPARTMENT OF PEDIATRICS

KEVIN J. KELLY, MD Interim Chair, Department of Pediatrics Interim Physician in Chief, N.C. Children's Hospital

VICE CHAIRS:

TERRY NOAH, MD Senior Vice Chair

ALLEN HABIF, CMPE Administration

IAN BUCHANAN, MD Clinical Integration

BENNY JOYNER, MD Hospital Inpatient Services

MICHAEL STEINER, MD Outreach & Business Development

TONI DARVILLE, MD Research

DAVID PEDEN, MD, MS Translational Research

JULIE BYERLEY, MD Education

STUART GOLD, MD Patient Engagement & Diversity Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Sincerely,

fllls nos

Kevin J. Kelly, MD Professor and Chair of Pediatrics Physician-in-Chief, NC Children's Hospital

260 MACNIDER HALL CAMPUS BOX #7220 CHAPEL HILL, NC 27599

Telephone: (919) 966-1505 Facsimile: (919) 966-7299



DEPARTMENT OF SURGERY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

DEPARTMENT of SURGERY 4041 BURNETT WOMACK CAMPUS BOX 7050 CHAPEL HILL, NC 27599-7050

T 919 966-4320 F 919 966-6009 melina_kibbe@med.unc.edu

MELINA R. KIBBE, MD, FACS, FAHA Zack D. Owens Professor and Chair

August 17, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Melina R. Kibbe, MD Zack D. Owens Professor and Chair Department of Surgery Editor-in-Chief, JAMA Surgery



100 MANNING HALL 216 LENOIR DRIVE, CB 3360 CHAPEL HILL, NC 27599-3360 T 919.962.8363 F. 919.962.8071 gary@ils.unc.edu

GARY MARCHIONINI, Ph D. Dean and Cary C. Boshamer Distinguished Professor

August 16, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

I am enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. SILS is already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S. Your proposed program is well-conceived and I believe will attract top students from doctoral programs at UNC and also be helpful in recruiting talent to doctoral programs across the campus.

I believe that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. The SILS faculty and I look forward to working with you on this important certificate and training program. Sincerely,

infundi ...

Gary Marchionini Dean and Cary C. Boshamer Professor



Matthew A. Mauro, MD FACR FSIR FAHA Chairman, Department of Radiology and CEO, UNC Faculty Physicians Ernest H. Wood Distinguished Professor of Radiology Professor of Surgery

August 22, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Man Mauro

Matthew A. Mauro MD FACR FSIR FAHA Ernest H. Wood Distinguished Professor & Chair Department of Radiology Professor of Surgery CEO, UNC Faculty Physicians

GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH SCHOOL OF MEDICINE

DEPARTMENT OF NUTRITION

CAMPUS BOX 7461

CHAPEL HILL, NC 27599-7461 T 919-966-7218 F 919-966-7215

ELIZABETH J MAYER-DAVIS, PhD Chair of the Department of Nutrition Cary C. Boshamer Distinguished Professor

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

August 23, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

We are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Elitchem 9 Mayor - Dais

Elizabeth J Mayer-Davis, PhD Chair of the Department of Nutrition Cary C. Boshamer Distinguished Professor of Nutrition Gillings School of Global Public Health School of Medicine



DEPARTMENT OF EPIDEMIOLOGY F 919.966.2089 MCGAVRAN-GREENBERG HALL CAMPUS BOX 7435 CHAPEL HILL, NC 27599-7435

August 17, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

On behalf of the Department of Epidemiology I enthusiastically support your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. Our department faculty are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Andrew F. Olshan, Ph.D. Barbara S. Hulka Distinguished Professor and Chair



THE UNIVERSITY of North Carolina at Chapel Hill

DEPARTMENT OF APPLIED PHYSICAL SCIENCES

244 CHAPMAN HALL CAMPUS BOX 3216 CHAPEL HILL, NC 27599-3216 T 919.843.9334 F 919.843.8412 apsc.unc.edu

August 16, 2016

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

Dear Michael and Greg,

Applied Physical Sciences enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. APS faculty members are already collaborating with you on your successful BD2K T32 training program. The proposed certificate is essential for students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers in North Carolina.

The proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Sincerely,

Very truly yours,

Edward T. Samulski, Chair Applied Physical Sciences



JEFF SEKELSKY PROFESSOR OF BIOLOGY AND GENETICS

CB #7100, 303 FORDHAM HALL CHAPEL HILL, NC 27599 sekelsky@unc.edu

PHONE: (919) 843-9400 FAX: (919) 962-4574 http://sekelsky.bio.unc.edu

Michael R. Kosorok, Ph.D. W. R. Kenan, Jr. Distinguished Professor Departments of Biostatistics and Statistics & Operations Research University of North Carolina at Chapel Hill

M. Gregory Forest, Ph.D. Graham Dahlstrom Distinguished Professor Departments of Mathematics, Biomedical Engineering, and Applied Physical Sciences University of North Carolina at Chapel Hill

October 4, 2016

Dear Michael and Greg,

We in the Curriculum in Genetics and Molecular Biology are enthusiastically supportive of your proposed graduate certificate that recognizes student training and achievement in interdisciplinary Big Data to Knowledge (BD2K) in Biomedicine. We are already collaborating with you on your successful BD2K T32 training program that is currently benefitting many individual students, research endeavors, and educational programs at UNC. The proposed certificate will further strengthen momentum in UNC's aspirations to build educational programs enabling students to prepare for careers addressing emerging challenges in Big Data Biomedical Science and to facilitate a new, data-driven generation of biomedical researchers and practitioners in North Carolina and the U.S.

We also feel that the proposed training will not pose a significant additional burden to our students, faculty or staff, but will blend efficiently with our current degree requirements while simultaneously enhancing the BD2K aspects of our training. We look forward to working with you on this important certificate and training program.

Professor of Biology and Genetics Director, Curriculum in Genetics and Molecular Biology

Required Courses Description

General Concepts

We view BD2K education as spanning and integrating across four disciplinary domains: (1) Biomedicine (includes health and human biology related disciplines), (2) Statistics (includes Biostatistics), (3) Computer Science (includes Informatics), and (4) Mathematics (includes Applied Mathematics). Any doctoral program at UNC with significant coverage of one or more of these domains and who is interested in using data science in a significant way to move biomedical research forward is welcome to be included. The BD2K program is not a doctoral program on its own but supplements the education experience of other doctoral programs having interest in furthering BD2K research. One of the challenges of listing all of the eligible courses for the BD2K certificate is that there are a large number in many departments and programs across the university, and it would difficult to list all of the eligible ones. The main approach we plan to use is to approve the courses each student wishes to use on a case by case basis and provide general guiding principles and some examples.

Within-Program Electives

Each student is required to take a 3 (or more) credit hour doctoral elective within their own doctoral program which is BD2K-intensive and clearly relevant for BD2K research. This course cannot be a required core doctoral course and must include graded assignments and/or projects. Usually this will be a course at the 700 level but alternative course numbers can be approved on a case by case basis by the BD2K program leaders. Below are some examples in Biostatistics and HPM for illustration:

BIOS 735 Statistical Computing - Basic Principles and Applications

- BIOS 752 Design and Analysis of Clinical Trials
- BIOS 764 Advanced Survey Sampling Methods
- BIOS 765 Models and Methodology in Categorical Data
- BIOS 767 Longitudinal Data Analysis (not acceptable since it is required)
- BIOS 772 Statistical Analysis of MRI Images
- BIOS 775 Statistical Methods in Diagnostic Medicine
- BIOS 776 Casual Inference in Biomedical Research
- BIOS 780 Theory and Methods for Survival Analysis (not acceptable since it is required)
- BIOS 781 Statistical Methods in Human Genetics
- BIOS 782 Statistical Methods in Genetic Association Studies
- HPM 760 Healthcare Quality and Information Management.
- HPM 770 Operations Research for Healthcare Systems.
- HPM 771 Introduction to Regression Models for Health Services Research.
- HPM 772 Techniques for the Economic Evaluation of Health Care.

HPM 779 Advanced Analytics and Operations Research.

HPM 785 Advanced Decision Modeling.

Out-of-Program Electives

Each student is required to take a 3 (or more) credit hour graduate elective in another graduate program which is in a different domain than their own doctoral program and normally which could count as a supporting program elective. Usually this will be a course at the 600 or 700 level but alternative course numbers can be approved on a case by case basis by the BD2K program leaders. Since these are out-of-program electives, degree required courses (in the out-of-program degree) can be included. For example, the two Biostatistics courses listed above as not acceptable are in fact acceptable for students not in a Biostatistics or Statistics PhD program. Some examples of 600 level courses from Biostatistics and HPM which would work include the following (many other courses in these and other departments and programs would also work):

BIOS 664 Sample Survey Methodology.

BIOS 665 Analysis of Categorical Data.

HPM 625 Diagnosis and Design of Multilevel Intelligence for a Smart Health System.

Modules

Each student is required to take 3 1-credit BD2K modules approved by the program leaders. These courses must all have a biomedical application and must utilize at least 2 of the quantitative domains (Statistics, Computer Science and/or Mathematics). These are multidisciplinary modules which teach students BD2K skills and prepare students to work on BD2K research projects in the research area of the module, including working in labs. The format is a research seminar, including lectures, in-class discussion, and assignments. Usually these courses are led by faculty in one of the BD2K or related disciplines and includes instructors from 2 or more of the BD2K domain areas. These modules vary from year to year, depending on faculty availability. Below are the five modules taught during Spring 2016:

Cancer Genomics and Class Discovery. MATH 891-004; February 15 – March 21, Monday 5:00-7:00pm. The objectives of this course are to: (1) Gain an understanding of the genomic data types being collected on human tumors; Learn about the basics of DNA sequence analysis coming from NGS platforms. (2) Learn about pattern discovery tools including hierarchical clustering and biclustering; Understand the challenges of integrating heterogeneous data types. (3) See real world examples of complex data integration for class discovery that can be applied to any disease type for which multiple technologies are used. Instructors: Chuck Perou (lead), Katie Hoadley, Joel Parker (Genetics, Pathology and Laboratory Medicine, Lineberger CCC), Steve Marron and Andrew Nobel (Statistics and Operations Research), Tim Elston (Pharmacology, Bioinformatics and Computational Biology) and Greg Forest (Applied Math, BME, APS).

Chromosome Organization and Dynamics in Living Cells. MATH 891-003, BIOL 890-002; Start and end dates, meeting time TBD. The objectives of this course are to: (1) Develop mathematical models of an entire eukaryotic genome in a living cell; (2) Extract quantitative data from the models and experimental systems to tune model parameters; (3) Discuss the power of mathematical modeling in guiding intuition at the nano to micron scales; (4) Discuss future avenues such as network analysis and fluid dynamics to

explore pathways of information flow in a living cell. Instructors: Kerry Bloom (lead-Biology), Greg Forest (Applied Math, BME, APS), David Adalsteinsson (Applied Math).

Microrheology and Transport – Particle Tracking in Biological Fluids. MATH 891-001, PHYS 893-001; Start and end dates, meeting time TBD. Topics covered include: (1) Transport properties of biological fluids; (2) Particle tracking experiments; (3) Fluorescence microscopy; (4) Particle tracking analysis; (5) Inferences from tracking data. Instructors: Sam Lai (lead-Eshelman School of Pharmacy), Rich Superfine (Physics & Astronomy), David Hill (P&A, Marsico Lung Institute), Greg Forest (Applied Math, BME, APS).

Neuromechanics. MATH 891-002, BIOL 890-001; January 13 – February 10, Wednesday 4:30pm – 7:00 pm. The course will begin with a review of basic neurobiology and muscular biology. We will then discuss traditional mathematical models in neuromechanics, including the Hodgkin Huxley equations, Hill muscle model, and integrate and fire models. We will then develop integrative models for connecting motor neurons, neuromuscular junctions, muscle activation, muscle contraction, resulting limb movement, and sensory feedback. The results of these models will be compared to experimental data. Instructors: Laura Miller (lead-Biology and Applied Math), Katie Newhall (Applied Math), seeking experimentalist in UNC SoM.

Predictive Analysis for High-Dimensional Data Sets. COMP 790-201; January 12 – February 11, T/TH 9:30-10:45 AM. We will cover multiple techniques for each of four problems: sparse regression, classification, clustering, and dimensionality reduction. The only graded assignments in the class will be four short projects that apply the techniques presented in class to real data. Our focus in the course will be on understanding the basic idea behind the methods and using existing implementations, so it should be a gentle introduction even for those with no background in statistics or machine learning. The projects themselves are drawn from a wide range of data types and biomedical applications, including audio data from patients with Parkinson's disease, image data from tumor biopsies, and gene expression data from cancer cells. Instructors: Joshua Welch and Robert Corty (both funded BD2K trainees), Jans Prins (faculty lead-Computer Science), and Dirk Dittmer (Microbiology and Immunology).