

Department of Health and Human Services Public Health Services Grant Application <i>Do not exceed character length restrictions indicated.</i>		LEAVE BLANK—FOR PHS USE ONLY.		
		Type	Activity	Number
		Review Group		Formerly
		Council/Board (Month, Year)		Date Received
1. TITLE OF PROJECT (<i>Do not exceed 81 characters, including spaces and punctuation.</i>) Novel Translational Informatics Software: an Application to Kidney Cancer				
2. RESPONSE TO SPECIFIC REQUEST FOR APPLICATIONS OR PROGRAM ANNOUNCEMENT OR SOLICITATION <input checked="" type="checkbox"/> NO <input type="checkbox"/> YES <i>(If "Yes," state number and title)</i>				
Number: Title: Mentored Research and Career Development (KL2) Program				
3. PROGRAM DIRECTOR/PRINCIPAL INVESTIGATOR				
3a. NAME (Last, first, middle)		3b. DEGREE(S)	3h. eRA Commons User Name	
Bokov, Alex		PhD MS BS	bokova	
3c. POSITION TITLE Instructor		3d. MAILING ADDRESS (<i>Street, city, state, zip code</i>) 7703 Floyd Curl Drive San Antonio, TX 78228		
3e. DEPARTMENT, SERVICE, LABORATORY, OR EQUIVALENT Department of Epidemiology and Biostatistics				
3f. MAJOR SUBDIVISION Clinical Informatics Research Division				
3g. TELEPHONE AND FAX (<i>Area code, number and extension</i>)		E-MAIL ADDRESS:		
TEL: 210 723-9814 FAX:		bokov@uthscsa.edu		
4. HUMAN SUBJECTS RESEARCH		4a. Research Exempt	If "Yes," Exemption No.	
<input type="checkbox"/> No <input checked="" type="checkbox"/> Yes		<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes		
4b. Federal-Wide Assurance No. FWA00005928		4c. Clinical Trial	4d. NIH-defined Phase III Clinical Trial	
		<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No <input type="checkbox"/> Yes	
5. VERTEBRATE ANIMALS <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes		5a. Animal Welfare Assurance No.		
6. DATES OF PROPOSED PERIOD OF SUPPORT (<i>month, day, year—MM/DD/YY</i>)		7. COSTS REQUESTED FOR INITIAL BUDGET PERIOD		8. COSTS REQUESTED FOR PROPOSED PERIOD OF SUPPORT
From	Through	7a. Direct Costs (\$)	7b. Total Costs (\$)	8a. Direct Costs (\$)
7/1/2018	6/30/2020	83,500	98,710	167,000
9. APPLICANT ORGANIZATION		10. TYPE OF ORGANIZATION		
Name The Univ. of TX Health Science Center at San Antonio		Public: → <input type="checkbox"/> Federal <input checked="" type="checkbox"/> State <input type="checkbox"/> Local		
Address 7703 Floyd Curl Drive San Antonio, TX 78229-3900		Private: → <input type="checkbox"/> Private Nonprofit		
		For-profit: → <input type="checkbox"/> General <input type="checkbox"/> Small Business		
		<input type="checkbox"/> Woman-owned <input type="checkbox"/> Socially and Economically Disadvantaged		
11. ENTITY IDENTIFICATION NUMBER 1-74-1586031A3				
DUNS NO. 80-077-2162		Cong. District	TX-021	
12. ADMINISTRATIVE OFFICIAL TO BE NOTIFIED IF AWARD IS MADE		13. OFFICIAL SIGNING FOR APPLICANT ORGANIZATION		
Name Rebecca G. Vega		Name	Chris G. Green, CPA	
Title Manager, Finance & Administration		Title	Director, Office of Sponsored Programs	
Address Mail Code 7933 7703 Floyd Curl Drive San Antonio, TX 78229-3900		Address	Mail Code 7828 7703 Floyd Curl Drive San Antonio, TX 78229-3900	
Tel: (210) 567-0836 FAX: (210) 567-0921		Tel: (210) 567-2340	FAX: (210) 567-8107	
E-Mail: vegarg@uthscsa.edu		E-Mail: grants@uthscsa.edu		
14. APPLICANT ORGANIZATION CERTIFICATION AND ACCEPTANCE: I certify that the statements herein are true, complete and accurate to the best of my knowledge, and accept the obligation to comply with Public Health Services terms and conditions if a grant is awarded as a result of this application. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties.		SIGNATURE OF OFFICIAL NAMED IN 13. <i>(In ink. "Per" signature not acceptable.)</i>		DATE

PROJECT SUMMARY (See instructions):

US-born Latinos may be at increased risk for kidney cancer according to a recently published analysis of death certificates in Texas and California. If this disparity can be confirmed using an independent data source, it will be important to understand how social risk factors, lifestyle, and genetic predisposition might interact to create it. I will test the disparity hypothesis and a secondary hypothesis that metabolic syndrome mediates poor outcomes via an inverse propensity treatment weighted Cox survival model and data from an Integrating Informatics from Bench to Bedside (i2b2) [9] data warehouse. Our i2b2 has electronic medical records (EMR) linked with socioeconomic Census data and detailed case histories from a cancer registry. As a post-doc and then as an instructor, I helped set up this data repository to advance population health and health services research on our campus. An obstacle that remains is the lack of a good way to download big datasets from i2b2 in an analyzable form. Over the last few years I have been writing an informatics app to overcome this obstacle. To insure that my software is responsive to the needs of clinician scientists and relevant to patient outcomes, I need to gain an inside perspective on clinical care and how clinicians use EMR systems. With my local mentors, among them the Chair of Urology, I chose a sequence of Medical School courses and other training opportunities that will help me become a successful independent researcher in medical informatics as well as a valuable collaborator to clinician scientists.

To take my informatics skills to the next level, I recruited a co-author of i2b2 and co-PI of the CTSA/ACT network to be my external mentor. He will supervise my programming efforts and my externship at Massachusetts General Hospital (MGH) in Boston.

In Aim-1 of my study, I will complete my data extraction app (DataFinisher) [10] and use it to acquire de-identified data for at least a thousand kidney cancer patients. In Aim-2 of my study, I will test the existence of the proposed ethnic disparity and identify possible mediating variables. In Aim-3 I will test the generalizability both of my results and of my informatics software at MGH during my Externship.

RFI FVANC (See instructions):

The clinical relevance of this work will be to determine, amidst as yet sparse published evidence, whether a disparity in kidney cancer outcomes affects two *regional* Hispanic populations (San Antonio and Boston). In the process, information will be generated to help plan follow-up studies (effect sizes, covariance structures, candidate covariates). The informatics relevance of this work will be to demonstrate feasibility and utility of my software for supporting multi-site studies in a flexible, convenient, and efficient manner.

PROJECT/PERFORMANCE SITE(S) (if additional space is needed, use Project/Performance Site Format Page)

Project/Performance Site Primary Location

Organizational Name:	The Univ. of TX Health Science Center at San Antonio		
DUNS:	80-077-2162		
Street 1:	7703 Floyd Curl Drive	Street 2:	
City:	San Antonio	County:	Bexar
Province:		Country:	USA
Project/Performance Site Congressional Districts:	TX-021		

Additional Project/Performance Site Location

Organizational Name:	Massachusetts General Hospital		
DUNS:	07-313-0411		
Street 1:	55 Fruit Street	Street 2:	
City:	Boston	County:	Suffolk
Province:		Country:	USA
Project/Performance Site Congressional Districts:	MA-008		

SENIOR/KEY PERSONNEL. See instructions. *Use continuation pages as needed* to provide the required information in the format shown below. Start with Program Director(s)/Principal Investigator(s). List all other senior/key personnel in alphabetical order, last name first.

Name	eRA Commons User Name	Organization	Role on Project
Amelie Ramirez	aramirez	UTHSCSA	Mentor/Chair
Ronald Rodriguez	rrodrig	UTHSCSA	Mentor
Joel Michalek	michalekj	UTHSCSA	Mentor
Shawn Murphy	smurphy	Harvard/MGH	Mentor

OTHER SIGNIFICANT CONTRIBUTORS

Name	Organization	Role on Project
TBA	DEB or Urology	Staff Database Programmer

Human Embryonic Stem Cells No Yes

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list:
https://grants.nih.gov/stem_cells/registry/current.htm. Use continuation pages as needed.

If a specific line cannot be referenced at this time, include a statement that one from the Registry will be used.

Cell Line

BUDGET FOR ENTIRE PROPOSED PROJECT PERIOD
DIRECT COSTS ONLY

BUDGET CATEGORY TOTALS	INITIAL BUDGET PERIOD <i>(from Form Page 4)</i>	2nd ADDITIONAL YEAR OF SUPPORT REQUESTED	3rd ADDITIONAL YEAR OF SUPPORT REQUESTED	4th ADDITIONAL YEAR OF SUPPORT REQUESTED	5th ADDITIONAL YEAR OF SUPPORT REQUESTED
PERSONNEL: <i>Salary and fringe benefits. Applicant organization only.</i>	58,500	58,500			
CONSULTANT COSTS					
EQUIPMENT					
SUPPLIES					
TRAVEL	2,000	2,000			
INPATIENT CARE COSTS					
OUTPATIENT CARE COSTS					
ALTERATIONS AND RENOVATIONS					
OTHER EXPENSES	23,000	23,000			
DIRECT CONSORTIUM/CONTRACTUAL COSTS					
SUBTOTAL DIRECT COSTS <i>(Sum = Item 8a, Face Page)</i>	83,500	83,500			
F&A CONSORTIUM/CONTRACTUAL COSTS					
TOTAL DIRECT COSTS	83,500	83,500			
TOTAL DIRECT COSTS FOR ENTIRE PROPOSED PROJECT PERIOD				\$ 167,000	

JUSTIFICATION. Follow the budget justification instructions exactly. Use continuation pages as needed.

The \$53,000 covers 75% of my salary; \$2000 will cover yearly travel to the CTSA/ACT conference in Washington DC. The remaining 23,000 will be spent on effort coverage for staff programmers: assuming a base salary of \$59,000 with F&A 30%, this will cover 3.6 months per year. The programmer's tasks will include: implementing changes to the i2b2 ETL process to address the specific needs of my kidney cancer disparity study; assisting with the testing/documentation/refactoring of DataFinisher code; assisting with DataFinisher web-app (see Aim-1); answering questions from MGH informatics team and testing/integrating code patches contributed by them and other teams in the ACT network. At this stage in the development of research data warehouses it is par for the course that funded research projects spend some of their budgets creating/improving quality and functionality in their domain of research interest thus gradually improving the curation of the system. As a member of CIRD this is what I have told collaborators using our services and therefore I am holding myself to the same standard in this grant proposal.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Bokov, Alex

eRA COMMONS USER NAME (credential, e.g., agency login): bokova

POSITION TITLE: Instructor

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
University of Michigan, Ann Arbor, MI	BS	2001	Cellular and Molecular Biology
University of Texas Health Science Center at San Antonio, San Antonio, TX	PHD	2008	Physiology
University of Texas at San Antonio, San Antonio, TX	MS	2014	Applied Statistics
University of Texas Health Science Center at San Antonio, San Antonio, TX	Postdoctoral Fellow	2014	Biostatistics, Data Science

A. Personal Statement

I am an informatician with graduate training in statistics and physiology. I am interested in using electronic medical records to better understand cancer and improve care decisions. In so doing I wish to make it easier for other researchers to use the analytic methods and software I develop for their own research.

My PhD training was in Physiology at UTHSCSA. I worked at the Audie Murphy VA Hospital's GRECC center and followed a traditional basic research track, working with animal models of longevity and stress resistance. As I progressed through the program I became increasingly concerned about how under-utilized modern statistical methods are in pre-clinical research. By the time I defended my dissertation, "The Role of Somatotropic and Estrogen Signaling in Longevity and Resistance to Oxidative Stress", I was passionately interested in data analysis and had taught myself the R statistical programming language. I took a Post-Doctoral position at the Barshop Institute for Aging and Longevity Studies in the Program Improvement Core of the Nathan Shock Center under the supervision of Dr. Jon Gelfond. Here I honed my programming, database, and data analysis skills through constant and daily use, providing analysis, study design, grant preparation efforts to multiple projects. Concurrently, I enrolled in an M.S. program in Applied Statistics at UTSA which I completed in 2014.

At the end of 2013 I accepted a position at the newly formed Clinical Informatics Research Division (CIRD) of the Department of Epidemiology and Biostatistics. In 2015 I was promoted to faculty. At CIRD, I helped stand up UTHSCSA's first production i2b2 data warehouse. I used my understanding of CIRD's operations together with my research training to develop and obtain IRB approval for a repository protocol that removes many administrative obstacles to feasibility counts and de-identified data requests while maintaining exceptionally high standards of patient privacy.

With all the pieces needed for a first-rate translational informatics program in place, I wish to demonstrate the utility of this system for research by using it to determine whether Hispanic patients are impacted by a disparity in the progression of kidney cancer and what factors mediate this disparity. In the course of this study I will complete my work on an informatics app that will streamline and automate data acquisition from i2b2.

- a. **Bokov A**, Wilson D, Espinoza SE, Shireman PK. Side by side comparison of Rockwood Index with Risk Analysis Index as predictors of surgical complications at a Texas safety net hospital serving a large Hispanic population. Presented at the International Conference on Frailty & Sarcopenia Research 2018, Miami Beach, FL, March 2018
- b. Connolly Dan, **Bokov A**, Waitman Russ, Tirado-Ramos A. Using Python and Paver to Control a Large Medical Informatics ETL Process. PyData 2014, Silicon Valley; 2014 May 30; Menlo Park, CA, USA.

B. Positions and Honors

Positions and Employment

2015 - pres Instructor, University of Texas Health Science Center at San Antonio, Department of Epidemiology and Biostatistics, Clinical Informatics Research Division, San Antonio, TX

Other Experience and Professional Memberships

2011 - pres Peer Reviewer, PLoS One
2012 - 2012 Peer Reviewer, Aging Cell
2015 - pres Member, Committee on Graduate Studies, Masters of Science in Clinical Investigation, University of Texas Health Science Center at San Antonio
2015 - pres Member, San Antonio Pepper Center Leadership Advisory Committee
2015 - 2017 Reviewer, Education in Computational Science, International Conference on Computational Science
2016 - pres Masters Dissertation Committee, University of Texas Health Science Center at San Antonio (two current students, one graduated)
2016 - pres Member, Institute for the Integration of Medicine and Science, Leadership Advisory Committee
2017 - pres Peer Reviewer, American Medical Informatics Association
2017 Grant Internal Reviewer, Greater Plains Collaborative

Honors

2009 Paul F. Glenn Award: For meritorious research in the area of biomedical gerontology by a post-doctoral fellow, American Aging Association, American Foundation for Aging Research

C. Contribution to Science

1. When my department's Clinical Informatics Research Division (CIRD) was first created, I used my knowledge of Linux, the Python programming language, and Oracle to start adapting an extract/transform/load (ETL) process to our local environment in order to deploy an i2b2 (Integrating Informatics from Bench to Bedside). I interviewed our programmers and taught them what I learned when they came on board. I am the informatics lead for the GPC demonstration project Family Weight and Health Survey (FWHS). This study combines a patient survey with extraction of EMR data for the same patients across 10 sites in multiple cities running a mix of Cerner and Epic EMRs, but having in common the i2b2 data warehouse platform. Recruitment and data collection from this project are now complete, and the first of many manuscripts is in preparation. While working on FWHS and other GPC projects I confronted two serious gaps: first the need for simple, general-purpose data extraction from i2b2 without writing new SQL code each time, and second, the need to automate the process of characterizing cohorts. It was hard finding anybody with enough cross-domain context in both statistics and database programming to understand the problem. So, I wrote two apps: DataFinisher and ChinoType. DataFinisher converts an i2b2 star-schema into a simple CSV table that can be read by almost any statistical software--SAS, R, Stata, Prism, etc. DataFinisher does more than dynamically generate SQL it also has heuristics for processing data elements differently depending on what they are. These rules are configurable, to allow for different semantics at each site. In the next revision, DataFinisher will have greatly simplified syntax for specifying data-handling rules and a simple way to preserve metadata while still retaining the universal compatibility of the CSV format. ChinoType is a plugin for the i2b2 web-app, allowing a researcher to ranked all the codes in the EMR from most over-represented to least in the patient cohort they selected. ChinoType is like enrichment analysis, except instead of gene-lists it applies to electronic health records. Both my apps are open source, freely available, and vendor-agnostic because they rely on i2b2's star schema. Informatics is transforming clinical and translational science in the same way that genomics and proteomics transformed cellular and molecular biology: rapid, seemingly effortless testing of thousands of hypotheses in parallel on massive retrospective datasets spanning a national network of data warehouses built on open standards and non-proprietary technology. But each one of these data warehouses needs to be kick-started by someone with the necessary cross-domain skills and willingness to spend several years of their academic career on the slow, complicated, and often invisible build-out of the prerequisite technical and organizational infrastructure. At UTHSCSA that individual was me.

- a. **Bokov A**, Manuel L S, Tirado-Ramos A, Gelfond JA, Pletcher S D. Biologically relevant simulations for validating risk models under small-sample conditions Proceedings of the IEEE Symposium on Computers and Communication 2017 Jul:290-295.
 - b. **Bokov A**, Bos A, Manuel L S, Tirado-Ramos A, Kittrell P, Jackson CE, Olin G P. Using Prevalence Patterns to Discover Un-mapped Flowsheet Data in an Electronic Health Record Data Warehouse Computer-Based Medical Systems 2017 Jun:324-327.
 - c. **Bokov A**, Bos A, Manuel L S, Tirado-Ramos A, Kittrell P, Olin G P, Jackson CE. Exhaustively Characterizing a Patient Cohort by Prevalence of EMR Facts: a Generalized, Vendor-Agnostic Method for Quality Control and Research AMIA Annual Symposium Proceedings 2017 Nov
 - d. **Bokov A**, Tirado-Ramos A, Bos, Angela B., Chen, Catherine, Manuel, Laura S.. Denormalize and Delimit: How not to Make Data Extraction for Analysis More Complex than Necessary Procedia Computer Science 2016 May;80:1033-1041..
2. Prior to my current work with data-mining health record informatics, I did the data analysis for a number of high-throughput "omics" and systems biology studies some of which are represented by these references.
 - a. Fok WC, **Bokov A**, Gelfond J, Yu Z, Zhang Y, Doderer M, Chen Y, Javors M, Wood WH 3rd, Zhang Y, Becker KG, Richardson A, Pérez VI. Combined treatment of rapamycin and dietary restriction has a larger effect on the transcriptome and metabolome of liver. Aging Cell. 2014 Apr;13(2):311-9. PubMed PMID: [24304444](#); PubMed Central PMCID: [PMC3989927](#).
 - b. Fok WC, Chen Y, **Bokov A**, Zhang Y, Salmon AB, Diaz V, Javors M, Wood WH 3rd, Zhang Y, Becker KG, Pérez VI, Richardson A. Mice fed rapamycin have an increase in lifespan associated with major changes in the liver transcriptome. PLoS One. 2014;9(1):e83988. PubMed PMID: [24409289](#); PubMed Central PMCID: [PMC3883653](#).
 - c. Butler JA, Mishur RJ, **Bokov AF**, Hakala KW, Weintraub ST, Rea SL. Profiling the anaerobic response of *C. elegans* using GC-MS. PLoS One. 2012;7(9):e46140. PubMed PMID: [23029411](#); PubMed Central PMCID: [PMC3459875](#).
 - d. Pérez VI, Pierce A, de Waal EM, Ward WF, **Bokov A**, Chaudhuri A, Richardson A. Detection and quantification of protein disulfides in biological tissues a fluorescence-based proteomic approach. Methods Enzymol. 2010;473:161-77. PubMed PMID: [20513477](#).
 3. As a Postdoctoral fellow I played a key role in analyzing data, helping the other researchers interpret the results, and writing the results up for publication. Among many other lines of research at the Barshop Institute, I worked on the early animal studies that identified rapamycin as a possible agent for extending lifespan and healthspan in mammals. I have another rapamycin manuscript in preparation with Zelton Sharp.
 - a. Zhang Y, **Bokov A**, Gelfond J, Soto V, Ikeno Y, Hubbard G, Diaz V, Sloane L, Maslin K, Treaster S, Réndon S, van Remmen H, Ward W, Javors M, Richardson A, Austad SN, Fischer K. Rapamycin extends life and health in C57BL/6 mice. J Gerontol A Biol Sci Med Sci. 2014 Feb;69(2):119-30. PubMed PMID: [23682161](#); PubMed Central PMCID: [PMC4038246](#).
 - b. Livi CB, Hardman RL, Christy BA, Dodds SG, Jones D, Williams C, Strong R, **Bokov A**, Javors MA, Ikeno Y, Hubbard G, Hasty P, Sharp ZD. Rapamycin extends life span of Rb1+/- mice by inhibiting neuroendocrine tumors. Aging (Albany NY). 2013 Feb;5(2):100-10. PubMed PMID: [23454836](#); PubMed Central PMCID: [PMC3616197](#).
 - c. Bhattacharya A, **Bokov A**, Muller FL, Jernigan AL, Maslin K, Diaz V, Richardson A, Van Remmen H. Dietary restriction but not rapamycin extends disease onset and survival of the H46R/H48Q mouse model of ALS. Neurobiol Aging. 2012 Aug;33(8):1829-32. PubMed PMID: [21763036](#).
 - d. Tabatabai-Mir H, Sataranatarajan K, Lee HJ, **Bokov AF**, Fernandez E, Diaz V, Choudhury GG, Richardson A, Kasinath BS. Rapamycin selectively alters serum chemistry in diabetic mice. Pathobiol Aging Age Relat Dis. 2012;2PubMed PMID: [22953036](#); PubMed Central PMCID: [PMC3417581](#).
 4. Discovered that the life-extension conferred by the *Igf1r^{+/+}* genotype in mice is conditional upon background strain or husbandry conditions. In female C57Bl/6 mice that experience a normal healthy lifespan for their

species, I found that the *Igf1r^{+/−}* genotype extends lifespan by just 5% and not the 33% originally reported in a short-lived strain. However, female *Igf1r^{+/−}* mice and *Prop-1* dwarf mice of both sexes (which are confirmed by multiple investigators to have extended lifespans) are both resistant to the paraquat and diquat models of oxidative stress. Thus I demonstrated that oxidative stress resistance is not tightly coupled to long life in mice. I went on to investigate the sexual dimorphism of oxidative stress resistance in C57Bl/6 and found that ovariectomy diminished paraquat resistance in females but resistance could be rescued by estradiol supplementation. Conversely, orchidectomy was protective against paraquat in males, and estradiol had no additional protective effect.

- a. **Bokov AF**, Garg N, Ikeno Y, Thakur S, Musi N, DeFronzo RA, Zhang N, Erickson RC, Gelfond J, Hubbard GB, Adamo ML, Richardson A. Does reduced IGF-1R signaling in *Igf1r^{+/−}* mice alter aging?. *PLoS One*. 2011;6(11):e26891. PubMed PMID: [22132081](#); PubMed Central PMCID: [PMC3223158](#).
- b. **Bokov AF**, Lindsey ML, Khodr C, Sabia MR, Richardson A. Long-lived Ames dwarf mice are resistant to chemical stressors. *J Gerontol A Biol Sci Med Sci*. 2009 Aug;64(8):819-27. PubMed PMID: [19414510](#); PubMed Central PMCID: [PMC2981464](#).
- c. **Bokov AF**, Ko D, Richardson A. The effect of gonadectomy and estradiol on sensitivity to oxidative stress. *Endocr Res*. 2009;34(1-2):43-58. PubMed PMID: [19557590](#); PubMed Central PMCID: [PMC2750774](#).
- d. **Bokov A**, Chaudhuri A, Richardson A. The role of oxidative damage and stress in aging. *Mech Ageing Dev*. 2004 Oct-Nov;125(10-11):811-26. PubMed PMID: [15541775](#).

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

Pilot Grant, San Antonio Claude Pepper Older Americans Independence Center
Shireman (PI)

05/01/17-05/01/18

Effectiveness of frailty screening in predicting post-operative morbidity and mortality

Role: Co-Investigator

CDRN-1501-26643, Patient Centered Outcomes Research Institute

Waitman (PI)

09/01/15-09/01/18

Greater Plains Collaborative Clinical Data Research Network

Broadening the network partnerships across clinical data research networks and five patient-powered research networks to create enhanced data resources to support observational/interventional research, adding record linkage and duplication analysis capabilities and deploying free text note de-identification and processing to support advanced large scale computable phenotyping.

Role: Faculty

1P30AG044271-01, NIH

Musi (PI)

06/01/15-04/01/20

San Antonio Claude D. Pepper Older Americans Independence Center Medicine

Role: Faculty

Completed Research Support

CDRN-1306-04631, Patient Centered Outcomes Research Institute

Waitman (PI)

03/01/14-09/01/15

Greater Plains Collaborative Clinical Data Research Network

The goal of this grant is to establish the Greater Plains Collaborative which will develop a large clinical data research network of electronic health record information to facilitate the conduct of clinical trials and comparative effectiveness research investigations.

Role: KP

RC2AG036613, NIH

Richardson (PI)

09/01/09-08/01/11

Can Rapamycin Retard Age-Related Diseases?

The TC2 grant is a unique, large, multi-investigator study designed to answer the following questions: (1) Does rapamycin retard/reduce age-related diseases [e.g., cancer, neurodegeneration, atherosclerosis, and nephropathy]? (2) Do mice given rapamycin maintain a better/longer healthspan (e.g., sensitivity of mice to infectious agents, autoimmunity, and biological function)? (3) Does rapamycin improve pathways that impact healthspan (e.g., autophagy and inflammation)?

Role: Post-Doctoral Scholar

2P30AG013319, NIH

Richardson (PI)

07/01/05-06/01/10

Nathan Shock Center of Excellence Program Enrichment Core

To provide administrative and analytic support to the core.

Role: PDC

5T32AG021890, NIH

Austad (PI)

09/01/06-09/01/08

Training Grant on the Biology of Aging

Role: GR

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Ramirez, Amelie G.

eRA COMMONS USER NAME (credential, e.g., agency login): aramirez

POSITION TITLE: Professor and Interim Chair, Department of Epidemiology and Biostatistics, and Director of the Institute for Health Promotion Research, University of Texas Health San Antonio

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	Completion Date MM/YYYY	FIELD OF STUDY
University of Texas, Houston, TX	B.S.	05/73	Psychology
University of Texas Health Science Center, School of Public Health, Houston	M.P.H.	05/77	Health Services Administration
University of Texas Health Science Center, School of Public Health, Houston	Dr.P.H.	05/92	Health Promotion

A. Personal Statement

I have a **proven track record of scientific research and innovation**, local, national, and international leadership, and pioneering health communication networks in chronic disease control and prevention that qualify me to serve as PI for the proposed R01 Research Project on HCC. I have more than 30 years of experience designing, implementing, directing, and evaluating **more than 100 large-scale collaborative research studies**, randomized controlled trials, and large research networks focused on human and organizational communication targeting positive behavior changes in cancer, genetic testing, chronic disease, healthy lifestyles, and obesity prevention among underserved minority populations, especially Latinos, in a variety of settings from the clinic to the community to the nation. All my prior projects utilized proven scientific theories and methodologies, incorporated collaborations with other researchers, and effectively administered project logistics. These projects led to unique health communication models and interventions that have identified racial/ethnic cancer and chronic disease disparities and risk factors and contributed to behavior change, such as increasing cancer screening rates, improving access to health care, and improving healthy lifestyles to reduce cancer and obesity risk. I also have **developed and managed national networks with multiple sites and objectives**, including experimental interventions, health policy research, and dissemination. For example, the National Cancer Institute funded my national Latino cancer network *Redes En Acción* from 2000-2017 to fuel research, education, and awareness advancements with leaders in five U.S. regions. I have also **led several liver cancer studies** that have identified persistent increases in hepatocellular carcinoma (HCC) rates in South Texas and among Latinos, as well as studies examining HCC risk factors. I also have **personally trained and/or mentored more than 200 Latino** undergraduates, doctoral students, and early-career faculty members. I bring all of this experience to my leadership role for the proposed R01 project to identify and study the interaction of HCC risk factors in South Texas, a region with disproportionately high HCC rates.

1. Ramirez AG. [The dire need for cancer health disparities research](#). *Health Educ Res*. 2013 Oct;28(5):745-7. doi: 10.1093/her/cyt089. PubMed PMID: 24038473.
2. Ramirez AG, Thompson IM, Vela L. [The South Texas health status review: A Health Disparities Roadmap](#). New York: Springer Science + Business Media; 2013.

B. Positions and Honors**Positions and Employment**

1983 – 1988 Assistant Professor, Department of Community Medicine, Baylor College of Medicine, Houston, TX

1983 – 1988	Assistant Professor, Experimental Medicine, Department of Medicine, Baylor College of Medicine, Houston, TX
1985 - 1985	Director, Cardiovascular Health Promotion Section, Debakey Heart Center, Houston, TX
1985 – 1992	Co-Principal Investigator, <i>A Su Salud</i> Project, Center for Health Promotion Research and Development, The University of Texas Health Science Center at Houston, Houston, TX
1990 – 1993	Principal Investigator, <i>MIRAME</i> : Substance Abuse Prevention Video Series for Hispanic Adolescents Project, South Texas Health Research Center, University of Texas (UT) Health San Antonio (formerly University of Texas Health Science Center at San Antonio), TX
1990 – 1995	Co-Principal Investigator, <i>Salud</i> : Mexican-American Participation in Cancer Prevention Project, South Texas Health Research Center, UT Health San Antonio, TX
1992 – 1997	Principal Investigator, <i>En Acción</i> : National Hispanic Leadership Initiative on Cancer Project, South Texas Health Research Center, University of Texas Health San Antonio, TX
1993 – 1997	Associate Professor, Department of Family Practice, UT Health San Antonio, TX
1993 – 1997	Co-Program Leader, Cancer Prevention and Population Studies, Cancer Therapy and Research Center (formerly the San Antonio Cancer Institute), an NCI-Designated Cancer Center, UT Health San Antonio, TX
1995 – 1997	Director, South Texas Health Research Center, University of Texas Health San Antonio, TX
1997 – 9/2006	Associate Director for Cancer Prevention and Control, Cancer Therapy and Research Center Research Foundation, San Antonio, TX
1997 – 9/2006	Associate Professor, Department of Medicine, Baylor College of Medicine, Houston, TX
1997 – 9/2006	Deputy Director, Chronic Disease Prevention and Control Research Center, Baylor College of Medicine, Houston, TX
2000 – 2007	Clinical Associate Professor, Department of Pediatrics, UT Health San Antonio, TX
2004 – 9/2006	Professor, Department of Medicine, Baylor College of Medicine, Houston, TX
2005 – 9/2006	Director, Office of Cancer Health Disparities Research, The Dan L Duncan Cancer Center, Baylor College of Medicine, Houston, TX
10/06 – 8/2012	Co-Program Leader for Cancer Prevention and Control, Cancer Therapy and Research Center Research Foundation, San Antonio, TX
10/06 – Present	Dielmann Chair, Health Disparities Research and Community Outreach, UT Health San Antonio, TX
10/06 – Present	Professor, UT Health San Antonio, Dept of Epidemiology and Biostatistics, San Antonio, TX
10/06 – Present	Founding Director, Institute for Health Promotion Research, UT Health San Antonio, TX
03/07 – Present	Voelcker Chair, Cancer Healthcare Disparities and Outreach, The Cancer Therapy and Research Center at UT Health San Antonio, TX
09/12 – Present	Associate Director for Cancer Prevention and Health Disparities, The Cancer Therapy and Research Center at UT Health San Antonio, TX
10/14 – Present	Chair Ad Interim, Dept of Epidemiology and Biostatistics, UT Health San Antonio, TX

Other Experience and Professional Memberships

1998-2002	Member, National Heart, Lung, and Blood Advisory Council, National Institutes of Health, Bethesda, MD
1999-2004	Member, National Cancer Policy Advisory Board, National Cancer Institute, Bethesda, MD
1999-2004	Member, National Cancer Advisory Board, National Cancer Institute, Bethesda, MD
1999-2004	Member, Behavior Change Expert Panel, Office of National Drug Control Policy, Wash, DC
2000-2005	Member, Tobacco Counter-Marketing Manual Review Committee, Department of Health & Human Services, Center for Disease Prevention and Control (CDC), Atlanta, GA
2000-2005	Member, Review Committee, Office on Smoking and Health, Atlanta, GA
2005-Present	Member, Board of Directors, Lance Armstrong Foundation, Austin, TX
2012-Present	Treasurer, Board of Directors, Lance Armstrong Foundation, Austin, TX
2007-2010	Chair, NBCCEDP Federal Advisory Board, CDC, Atlanta, GA
2007-Present	Member, Scientific Advisory Board, Susan G. Komen for the Cure, Dallas, TX
2007-2011	Member, Scientific Advisory Board, Avon Foundation Breast Cancer Crusade, New York, NY
2007-Present	Member, Institute of Medicine, National Academies, Washington, DC
2012-Present	Membership Chair, Institute of Medicine, National Academies, Section 9
2010-Present	Member, Institute of Medicine's Roundtable on the Promotion of Health Equity and the Elimination of Health Disparities, Washington, DC
2010-Present	Member, Texas Academy of Medicine, Engineering, and Science

Selected Honors

1. Dr. LaSalle D. Leffall, Jr. Cancer Prevention Award in recognition of efforts to promote cancer awareness among minorities and the economically disadvantaged, UTMD Anderson Cancer Center, 1991.
2. Sarah Mazelis Award in recognition of outstanding practice in the field of health education, the American Public Health Association (APHA), Public Health Education and Health Promotion Section, 1996.
3. Humanitarian Award, National American Cancer Society, 2003.
4. Cancer Prevention Laurel for Dedication to Community Programs, Cancer Research and Prevention Foundation, 2004.
5. IABC-San Antonio Bronze Quill award for the publication *Nuestras Historias*, 2005.
6. Professor of Survivorship, Susan G. Komen For the Cure, 2007.
7. Elected to the Institute of Medicine of the National Academies, 2007.
8. White House "Champion of Change," 2011.
9. Making a Difference Award, Latina Contra Cancer, 2014.
10. Everett M. Rogers Award, APHA, Public Health Education and Health Promotion Section, 2014.

C. Contributions to Science

I pioneered the **Diffusion Acceleration Model**, based on the Social Cognitive and Diffusion of Innovation theories, which uses dual-link communications to accelerate adoption of health practices that are not widely present in a population by presenting examples of role models who have already acquired the practice, synchronically using mass media and interpersonal communications (what we call "behavioral journalism."). I led studies using this model to accelerate behavioral and policy changes among Latinos for many health issues, including improving cancer screening rates and reducing tobacco use in South Texas.

- a. Ramirez AG, McAlister A. Mass media campaign—A Su Salud. *Prev Med* 1988 Sep;17(5):608-621.
- b. Ramirez AG, Chalela P, Suarez L & Gallion KJ. A Su Salud *En Acción*: Replicating a model to increase utilization of cancer screening among low income Latinas. *J Health Dispar Res Pract*. 2011; 5, (1), 65-79.
- c. Ramirez A, Villarreal R, McAlister A, Gallion KJ, Suarez L, Gomez P. Media and community campaign effects on adult tobacco use in Texas. *J Health Commun*. 2004 Mar;9(2):95-109.
- d. Ramirez AG, Villarreal R, McAlister A, Gallion K, Suarez L, Gomez P. Advancing the role of participatory communication in the diffusion of cancer screening among Hispanics. *J Health Commun* 1999 Jan;4(1):31-36.

Before the 1990s, Latinos were considered homogeneously in studies. As PI of a national multi-site project funded by NCI from 1992-1999, I initiated the **first comprehensive assessment of cancer risk factors among the major Latino populations**: Mexican Americans in San Diego, San Antonio and Brownsville, Texas; Cubans in Miami; Puerto Ricans in New York City; and Central and South Americans in San Francisco. The project built a cohort of over 18,000 cases and was first to identify levels of risk, adherence to cancer screening recommendations, and healthy behaviors among different Latino population groups.

- a. Ramirez AG, McAlister AL, Villarreal R, Suarez L, Talavera GA. [Prevention and control in diverse Hispanic populations: A national initiative for research and action](#). *Cancer*. 1988 January; 83(8 Suppl):1825-1829.
- b. Ramirez AG, Suarez L, Laufman L, Barroso C, Chalela P. [Hispanic women's breast and cervical cancer knowledge, attitudes, and screening behaviors](#). *Am J Health Promot*. 2000 May-Jun; 14(5):292-300. PubMed PMID: 11009855.
- c. Pérez-Stable EJ, Ramirez A, Villareal R, Talavera GA, Trapido E, Suarez L, Marti J, McAlister A. [Cigarette smoking behavior among US Latino men and women from different countries of origin](#). *Am J Public Health*. 2001 Sep; 91(9):1424-30. PubMed PMID: 11527775; PubMed Central PMCID: PMC1446798.
- d. Ramirez AG, Talavera GA, Villarreal R, Suarez L, McAlister A, Trapido E, Pérez-Stable E, Marti J. [Breast cancer screening in regional Hispanic populations](#). *Health Educ Res*. 2000 Oct; 15(5):559-68. PubMed PMID: 11184215.

Building on this multisite framework, I **developed a national research network** (*Redes En Acción*; NCI U01/U54 from 2000-2017) for research, training, and awareness to reduce Latino cancer health disparities. My network leadership helped identify issues with tobacco use, late-stage diagnosis, low screening rates, and low clinical trial and genetic testing participation among Latinos. The network, headquartered under my direction in San Antonio with regional sites in San Francisco, San Diego, New York City, Miami, and later also in Chicago, has conducted and published 300+ research projects; surpassed \$232 million in leveraged

funding; conducted 2,400 community events; earned awards for its PSAs, magazines, newsletters, and *fotonovelas*, and funded 26 junior faculty who received \$1 million in funding for Latino cancer research.

- a. Ramirez AG, Talavera GA, Marti J, Penedo FJ, Medrano MA, Giachello AL, Pérez-Stable EJ. [Redes En Acción. Increasing Hispanic participation in cancer research, training, and awareness.](#) *Cancer*. 2006 Oct 15; 107(8 Suppl):2023-33. PubMed PMID: 16958026.
- b. Ramirez AG, Gallion KJ, Suarez L, Giachello AL, Marti JR, Medrano MA, Pérez-Stable EJ, Talavera GA, Trapido EJ. [A national agenda for Latino cancer prevention and control.](#) *Cancer*. 2005 Jun 1; 103(11):2209-15. PubMed PMID: 15822119.
- c. Ramirez AG, Suarez L, Chalela P, Talavera GA, Marti J, Trapido EJ, Villarreal R, Pérez-Stable EJ. [Cancer risk factors among men of diverse Hispanic or Latino origins.](#) *Prev Med*. 2004 Aug; 39(2):263-9. PubMed PMID: 15226034.

I led a study that identified **South Texas Latinos as having the highest rate of liver cancer in the nation**. The study found that liver cancer incidence rates in South Texas were 3.1 higher in men and 4 times higher in women than their non-Latino White counterparts. South Texas Latinos had even higher rates. Due to the many potential risk factors for liver cancer and hepatocellular carcinoma (HCC), I also led a study that found Latinos with liver cancer had much higher levels of aflatoxins than those without liver cancer. Aflatoxins are cancer-causing chemicals produced by mold that can contaminate improperly stored foods. We analyzed aflatoxin exposure in 42 liver cancer cases and 42 non-cases from clinics in San Antonio, Texas (two-third Latino) and found that liver cancer cases had 6 times higher odds of having detectable levels of aflatoxins in their blood, compared to non-cases. The study was the first to link liver cancer with aflatoxin exposure among Latinos.

- a. Ramirez AG, Muñoz E, Long Parma D, Michalek JE, Holden AEC, Phillips TD, Pollock BH. Lifestyle and clinical correlates of hepatocellular carcinoma in South Texas: a matched case-control study. *Clin Gastroenterol Hepatol*. 2017 Mar 23; PubMed PMID: 28344065.
- b. Ramirez AG, Munoz E, Holden AE, Adeigbe RT & Suarez L. Incidence of hepatocellular carcinoma in Texas Latinos, 1995-2010: an update. *PLoS One*, 2014. 9(6):e99365.

Only about 5% of Latinos participate in NCI cancer clinical trials. My studies **identified cultural and other barriers and enablers** to Latino clinical trial participation among patients and physicians. This led to a current Komen study that hypothesizes that a technologically oriented educational intervention can increase Latinas' self-efficacy to make informed decisions about breast cancer clinical trials (results not yet published). I also published a Clinical Trials Outreach for Latinos replication manual (<http://bit.ly/1NrpbnK>).

- c. Ramirez AG, Wildes K, Talavera G, Nápoles-Springer A, Gallion K, Pérez-Stable EJ. [Clinical trials attitudes and practices of Latino physicians.](#) *Contemp Clin Trials*. 2008 Jul; 29(4):482-92. Epub 2007 Nov 21. PubMed PMID: 18155966; PubMed Central PMCID: PMC2587358.
- d. Chalela P, Suarez L, Muñoz E, Gallion KJ, Pollock BH, Weitman SD, Karnad A, Ramirez AG. [Promoting factors and barriers to participation in early phase clinical trials: Patients perspectives.](#) *J Community Med Health Educ*. 2014 Apr 24; 4(281):1000281. PubMed PMID: 25077043; PubMed Central PMCID: PMC4112537.
- e. Ramirez AG, Miller AR, Gallion K, San Miguel de Majors S, Chalela P, García Arámburo S. [Testing three different cancer genetics registry recruitment methods with Hispanic cancer patients and their family members previously registered in local cancer registries in Texas.](#) *Community Genet*. 2008; 11(4):215-23. doi: 10.1159/000116882. Epub 2008 Apr 14. PubMed PMID: 18417969.

Given Latinos' heavy use of mobile Internet, texting, and social media, I **led studies using the latest mobile technologies**—mobile phone apps, text messaging programs, and social media innovations, etc.—to improve health behaviors among Latinos. I recently led the design and creation of *Quitxt*, a bilingual tobacco-cessation service for young Latino adults using mobile-phone text messages and mobile media; the service yielded a 20% quit rate among enrollees at seven-month follow-up. I also am testing an interactive mobile app to improve adherence to endocrine hormonal therapy among breast cancer patients (Susan G. Komen; results not yet published) and a text-messaging program to promote health in hard-to-reach groups (San Antonio Life Sciences Institute; results not yet published). I also designed a digital content curation strategy for health information for Latino audiences via blogs and social media.

- a. Ramirez AG, Chalela P, Akopian D, Munoz E, Gallion KJ, Despres C, Morales J, Escobar R, & McAlister AL. Text & mobile media smoking cessation service for young adults in south Texas: Operation & costs effectiveness estimation. *Health Promotion Practice*, 2017 [Epub ahead of print].
- b. Ramirez AG, Baldwin S, Adeigbe RT, Aguilar RP, Gallion KJ, Despres C. *SaludToday*: Curating Latino health information for a new generation. *J Commun Healthc*, Vol. 9, Iss. 1, 2016.

D. Research Support

Ongoing Research Support

- National Cancer Institute 1 R25 CA134301-01A2 Ramirez (PI) 08/01/2010 – 07/31/2020
Éxito! Latino Cancer Research Leadership Training
This project trains Latino master's-level students and professionals to pursue a doctoral degree and cancer research. Over 20% of program alum have applied for and/or enrolled in doctoral programs.
Role: Principal Investigator (PI)
- Robert Wood Johnson Foundation RWJF-75213 Ramirez (PI) 02/15/2018 – 08/14/2018
Salud America! The RWJF Research Network to Prevent Obesity Among Latino Children
The reauthorization of *Salud* activates its 100,000-person network to create healthy changes to reduce Latino childhood obesity. SA! will monitor online network activities with robust online and social media analytical tools and conduct periodic surveys/polls and other assessments to determine the extent to which SA! educational materials and network activities are positively influencing levels of self and collective efficacy of network members. The analysis will inform future efforts to promote the Robert Wood Johnson Foundation's Health Kids, Healthy Weight initiative and reduce Latino childhood obesity. Role: Principal Investigator (PI)
- Cancer Prevention & Research Institute of Texas PP170099 Ramirez (PI) 03/01/2018 – 02/28/2021
SMS Cessation Service for Young Adult Smokers in South Texas
This award is a continuation of PP140176. The project aims to overcome accessibility and barriers by using culturally and linguistically appropriate mobile channels and messages to recruit smokers to an evidence-based SMS cessation service for South Texans. Role: Principal Investigator (PI)
- Susan G. Komen Foundation (National) SAB160005 Ramirez (PI) 04/01/2016 – 03/31/2019
Improving Adherence to Endocrine Hormonal Therapy (EHT) among Breast Cancer Patients
This study is enrolling 120 breast cancer patients for an evidence-based intervention with 1) a bilingual, culturally tailored, personalized, interactive mobile app; and 2) support from a patient navigator; vs. usual care.
Role: Principal Investigator (PI)

Completed Research Support

- National Cancer Institute Ramirez (PI) 09/01/2010 – 08/31/2017
Redes En Acción: The National Latino Cancer Research Network
This project was a comprehensive set of outreach, research and training activities carried out nationally.
Role: Principal Investigator (PI)
- Susan G. Komen Foundation (National) Ramirez (PI) 04/01/2008 – 03/31/2015
Increasing Access of Latinas into Breast Cancer Clinical Trials
This project was part of a multi-year program testing the efficacy of family-based and provider-based strategies to increase accrual of Latinas into breast cancer clinical trials.
Role: Principal Investigator (PI)
- Robert Wood Johnson Foundation RWJF-74301 Ramirez (PI) 02/15/2017 – 02/14/2018
Salud America! The RWJF Research Network to Prevent Obesity Among Latino Children
The reauthorization of *Salud* activates its 100,000-person network to create healthy changes to reduce Latino childhood obesity. Since 2007, *Salud* has created new evidence, researchers, tools, and communications.
Role: Principal Investigator (PI)
- Cancer Prevention & Research Institute of Texas PP140176 Ramirez (PI) 09/01/2014 – 08/31/2017
SMS Cessation Service for Young Adult Smokers in South Texas
This project aims to overcome accessibility and barriers by using culturally and linguistically appropriate mobile channels and messages to recruit smokers to an evidence-based SMS cessation service for South Texans.
Role: Principal Investigator (PI)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES

NAME: Ronald Rodriguez, M.D., Ph.D.

eRA COMMONS USER NAME (credential, e.g., agency login): rrodrig

POSITION TITLE: Interim Dean, School of Medicine; Professor & Chair, Department of Urology

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	MM/YY	FIELD OF STUDY
Massachusetts Institute of Technology, Cambridge, MA	B.S.	1984	Life Sciences
Baylor College of Medicine, Houston, TX	Ph.D.	1990	Cell Biology
Baylor College of Medicine, Houston, TX	M.D.	1992	Medicine
Johns Hopkins School of Medicine, Baltimore, MD	Residency	1994	Surgery
Johns Hopkins School of Medicine, Baltimore, MD	Residency	1998	Urology

A. Personal Statement

I balance a fulltime clinical schedule while running a translational research lab, overseeing the departmental operations for the Department of Urology and teaching students and residents. I spent 8 years as the residency program director for the department of Urology at Johns Hopkins University, and have a clear record of leadership in resident education. I also have a longstanding history of scientific innovation. I have extensive expertise in treating prostate, kidney and bladder cancer patients. As well as being a surgeon I am a trained molecular biologist and have pioneered the first recombinant oncolytic adenovirus targeted for a specific cancer, and have been a leader in adenoviral gene therapy for 15 years. I developed the largest series of cryoablation for renal masses in the country, and have made major contributions in molecular endocrinology, molecular imaging, outcomes from cancer research, and translational research. While at Hopkins I have been involved in a number of 1st in human clinical trials. As a chairman, my vision for the department has been to have a strong basic science foundation which interacts and supports the clinical enterprise. In addition, as a native of San Antonio, I have a strong interest in outreach to the community and serving the minority population so prevalent in this part of the country.

1. Rodriguez R, Schuur ER, Lim HY, Henderson GA, Simons JW, Henderson DR. Prostate attenuated replication competent adenovirus (ARCA) CN706: a selective cytotoxic for prostate-specific antigen-positive prostate cancer cells. *Cancer Res.* 1997 Jul 1;57(13):2559-63. PMID: 9205053
2. Neuman BP, Eifler JB, Castanares M, Chowdhury WH, Chen Y, Mease RC, Ma R, Mukherjee A, Lupold SE, Pomper MG, Rodriguez R. Real-time, near-infrared fluorescence imaging with an optimized dye/light source/camera combination for surgical guidance of prostate cancer. *Clin Cancer Res.* 2015 Feb 15;21(4):771-80. PMID: 25501577
3. Hötö N, Yang S, Aiyetan P, Kumar B, Hu Y, Clark D, Eroglu AU, Shah P, Johnson T, Chowdery WH, Zhang H, Rodriguez R. Overexpression of Exportin-5 Overrides the Inhibitory Effect of miRNAs Regulation Control and Stabilize Proteins via Posttranslational Modifications in Prostate Cancer. *Neoplasia.* 2017 Oct;19(10):817-829. Epub 2017 Sep 4.
4. Shanmugasundaram K, Nayak BK, Friedrichs WE, Kaushik D, Rodriguez R, Block K. NOX4 functions as a mitochondrial energetic sensor coupling cancer metabolic reprogramming to drug resistance. *Nat Commun.* 2017 Oct 19;8(1):997. PMID: 29051480

B. Positions and Honors**Positions and Employment**

1999-2006 Assistant Professor, Brady Urological Institute, Johns Hopkins, Baltimore, MD
 2006-2012 Associate Professor, Brady Urological Institute, Johns Hopkins, Baltimore, MD
 2017-2018 Interim Dean, School of Medicine,
 2013-Present Professor & Chair, UTHSCSA, Department of Urology, San Antonio, TX

Other Experience and Professional Memberships

1995-present Member, American Urological Association
 1999-present Member, American Society of Gene Therapy
 1999-present Member, American Association for Cancer Research

2005-2012 Residency Program Director & Residency Director, Urology, Johns Hopkins University
 2007-2013 Associate Editor, Urology, Basic Sciences Section

Honors and Awards

1998	Robert Wood Johnson Minority Faculty Development Award
2000	Pfizer Scholar in Urology Award
2003	CaPCure Research Award
2005	Honorary Professor, Shandong Provincial Hospital, Dept of Urology
2010	Healthnetwork Service Excellence Award Recipient (one of ten nationwide)

C. Contributions to Science

1. I have investigated the role of Histone Deacetylase Inhibitors in treating cancer, particularly the utility of Valproic acid (VPA). The toxicity profile of VPA is very well understood as it has been in clinical use for treating epilepsy over 5 decades and it also is a very inexpensive drug. We have demonstrated the effectiveness of VPA in reducing invasiveness and progression of bladder cancer, as well as its utility in treating prostate cancer in pre-clinical models. In a recently terminated Phase II clinical trial we have collected data that suggest VPA could be used to delay prostate cancer patients with biochemical recurrence from going onto Androgen Deprivation Therapy.
 - 1.. Shabbeer S, Kortenhorst MS, Kachhap S, Galloway N, **Rodriguez R**, Carducci MA. Multiple Molecular pathways explain the anti-proliferative effect of valproic acid on prostate cancer cells in vitro and in vivo. *Prostate*. 2007 Jul 1;67(10):1099-110. PMID: 17477369
 2. Sidana A, Wang M, Shabbeer S, Chowdhury WH, Netto G, Lupold SE, Carducci M, **Rodriguez R**. Mechanism of growth inhibition of prostate cancer xenografts by valproic acid. *J Biomed Biotechnol*. 2012;2012:180363. PMID: 23093837
2. For the better part of my scientific career I have focused on developing & understanding the biology of oncolytic gene therapy vectors. I pioneered the development of tissue specific oncolytic viruses and was part of the 1st group to take a recombinant oncolytic adenovirus into clinical trial. My group was the 1st to describe the attenuating effects of Androgen Receptor (AR) in AR dependent recombinant oncolytic adenoviruses, and we also described methods to circumvent the problem. We have studied the role of p21 in viral replication, and also developed recombinant viruses that work in conjunction with radiation and anti-Androgen therapy. We have also developed a system to generate and screen peptides on the viral capsid to facilitate the development of tissue specific gene delivery via adenovirus.
 1. Höti N, Chowdhury WH, Mustafa S, Ribas J, Castanares M, Johnson T, Liu M, Lupold SE, **Rodriguez R**. Armoring CRAdS with p21/Waf-1 shRNAs: the next generation of oncolytic adenoviruses. *Cancer Gene Ther*. 2010 Aug;17(8):585-97. PMID: 20448671 Free PMC Article
 2. Liu C, Zhang Y, Liu MM, Zhou H, Chowdhury W, Lupold SE, Deweese TL, **Rodriguez R**. Evaluation of continuous low dose rate versus acute single high dose rate radiation combined with oncolytic viral therapy for prostate cancer. *Int J Radiat Biol*. 2010 Mar;86(3):220-9. PMID: 20201650
3. I developed the largest series of cryoablation for renal masses in the country, and a large part of my clinical practice deals with diseases of the kidney. We have optimized the technique of percutaneous cryoablation of malignant tissue under CT guidance. I have also studied the pathophysiology of kidney stones.
 1. **Rodriguez R**, Cizman Z, Hong K, Koliatsos A, Georgiades C. Prospective analysis of the safety and efficacy of percutaneous cryoablation for pt1NxMx biopsy-proven renal cell carcinoma. *Cardiovasc Intervent Radiol*. 2011 Jun;34(3):573-8. PMID: 20628879
 2. Sidana A, Aggarwal P, Feng Z, Georgiades CS, Trock BJ, **Rodriguez R**. Complications of renal cryoablation: a single center experience. *J Urol*. 2010 Jul;184(1):42-7. PMID: 20478601
4. I have also made major contributions in molecular imaging. We have developed a Near Infra Red fluorescent imaging system which can revolutionize how future prostate cancer surgeries are performed.
 1. Neuman BP, Eifler JB, Castanares M, Chowdhury WH, Chen Y, Mease RC, Ma R, Mukherjee A, Lupold SE, Pomper MG, Rodriguez R. Real-time, near-infrared fluorescence imaging with an optimized dye/light source/camera combination for surgical guidance of prostate cancer. *Clin Cancer Res*. 2015 Feb 15;21(4):771-80. PMID: 25501577
 2. Castanares MA, Mukherjee A, Chowdhury WH, Liu M, Chen Y, Mease RC, Wang Y, Rodriguez R, Lupold SE, Pomper MG. Evaluation of prostate-specific membrane antigen as an imaging reporter. *J Nucl Med*. 2014 May;55(5):805-11. PMID: 24700883

Complete List of Published Work in MyBibliography:

https://www.ncbi.nlm.nih.gov/sites/myncbi/1J_ohscyWuPAd/collections/47927472/public/

BIOGRAPHICAL SKETCH

NAME: Michalek, Joel

eRA COMMONS USER NAME: MICHALEKJ

POSITION TITLE: Professor

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	Completion Date MM/YYYY	FIELD OF STUDY
Wayne State University, Detroit, MI	BS	1966	Mathematics
Wayne State University, Detroit, MI	MS	1968	Mathematics
Wayne State University, Detroit, MI	PhD	1973	Mathematical Statistics

A. Personal Statement

With regard to your KL2 Scholar application, I will provide mentoring with regard to statistical methods, power and sample size calculations, manuscript-writing, and responses to journal reviewers as needed. I have a broad background in biostatistics pertaining to theory and methods, preclinical and clinical trials, and epidemiology. Since 2006, in my current position, I have written protocols and grants, analyzed data, and co-authored manuscripts arising from clinical studies in surgery, emergency medicine, cancer, and pediatrics, clinical trials pertaining to medical devices and therapeutic medications, and an epidemiological study. From 1992 to 2005 I was Principal Investigator of the Air Force Health Study, a 20-year prospective epidemiological study of veterans who sprayed Agent Orange and other herbicides in Vietnam; while in that position, I authored or co-authored papers on statistical methods, pharmacokinetics, survival analysis, and epidemiology and presented results to the National Academy of Sciences and the United States Congress. My research interests include applications of signal processing in clinical medicine and genomics. I am currently active at the Cancer Therapy and Research Center (CTRC) as a member of the Protocol Review Committee; additionally I provide statistical guidance and co-authorship on CTRC investigator initiated grants and protocols, and as a research Biostatistician for the Department of Emergency Medicine. Four co-authored papers relevant to cancer research are listed. My complete bibliography is available at http://www.experts.scival.com/uthscsa/orgaDetail.asp?o_id=20&oe_id=1.

1. Landgren O, Shim Y, **Michalek JE**, Costello R, Burton D, Ketchum N, Calvo K, Caporaso N, Raveche E, Middleton D, Marti G, Vogt R. Agent Orange exposure and monoclonal gammopathy of undetermined significance. An Operation Ranch Hand veteran cohort study. *JAMA Oncology* doi:10.1001/jamaoncol.2015.2938.
2. Abraham J, Nunez-Alvarez Y, Hettmer S, Carrio E, Chen HI, Nishijo K, Huang ET, Prapjapati SI, Walker RL, Davis S, Rebeles J, Wiebush H, McCleish AT, Hampton ST, Bjornson CR, Brack AS, Wagers AJ, Rando TA, Capecchi MR, Marini FC, Ehler BR, Zarzabal LA, Goros MW, **Michalek JE**, Meltzer PS, Langenau DM, LeGallo RD, Mansool A, Chen Y, Suelves M, Rubin BP, Keller C. Lineage of origin in rhabdomyosarcoma informs pharmacological response *Genes Dev* 2014 Jul;25(14):1578-1591
3. Mahalingam D, Malik L, Beeram M, Rodon J, Sankhala K, Mita AC, Benjamin D, Ketchum N, **Michalek JE**, Tolcher A, Wright J, Sarantopoulos J. Phase II Study Evaluating the Efficacy, Safety, and Pharmacodynamic Correlative Study of Dual Antiangiogenic Inhibition Using Bevacizumab in Combination with Sorafenib in Patients with Advanced Malignant Melanoma *Cancer Chemother Pharmacol* 2014 May
4. Ha CS, **Michalek JE**, Elledge R, Kelly KR, Ganapathy S, Su H, Jenkins CA, Argiris A, Swords R, Eng TY, Karnad A, Crownover RL, Swanson GP, Goros M, Pollock BH, Yuan ZM. A p53-Based strategy to reduce hematological toxicity of chemotherapy: A proof of principle study. *Mol Oncol*. 2015 Sep 18. pii: S1574-7891(15)00163-5. doi: 10.1016/j.molonc.2015.09.004.

B. Positions and Honors

Positions and Employment

1973-1976	Assistant Professor, Mathematics, Syracuse University, Syracuse, NY
1976-1977	Mathematical Statistician, Consultant, Transportation Systems Center, Cambridge, MA
1977-2005	Principal Investigator, Government, Air Force Research Laboratory, Brooks City-Base, TX
1993-pres	Adjunct Professor, Management Science and Statistics, UT at San Antonio, TX
2001-pres	Adjunct Professor, UT School of Public Health, Houston, TX
2005	Clinical Professor, Epidemiology and Biostatistics, Medical School, UT Health Science Center at San Antonio, TX
2005-pres	Professor, Epidemiology and Biostatistics, Medical School, UT Health Science Center at San Antonio, TX

Honors and Professional Memberships

1993	CDC & ATSDR Statistics Award, for statistical excellence demonstrated by the publication of "Effects of Measurement Error on Estimating Biological Half-Life" Journal of Exposure Analysis and Environmental Epidemiology 1993; 2:463-476
1997	CDC and ATSDR Statistical Science Award, Paper Finalist in recognition of "Pharmacokinetics of TCDD in Veterans of Operation Ranch Hand: 10-Year Follow-up." Journal of Toxicology and Environmental Health
2016	Fellow of the American Statistical Association

C. Contribution to Science

1. Three contributions are worth mentioning, a proof of the mathematical equivalence of two apparently different scoring methods for linear rank procedures in survival analysis, a series of papers pertaining to endocrine disruption in Vietnam veterans occupationally exposed to Agent Orange and its TCDD contaminant, and, motivated by the Air Force Health Study, a correction for bias in statistical estimation of the elimination rate in a first order pharmacokinetic model. Our scoring paper [1] was written in 1981 at a time when new methods for the analysis of censored survival data were becoming available, in particular those pertaining to score statistics for testing the null hypothesis of no treatment effect under a variety of survival models and independent and arbitrary right censoring. Our paper clarified the relation between two scoring methods and helped move the field forward. Closely related papers [2] and [3] applied the resulting theory to matched survival studies and numerical routines for score calculation.

- a. Mehrotra KG, **Michalek JE** and Mihalko D. A relationship between two forms of linear rank procedures for censored data. *Biometrika* 1982;69:674-676.
- b. **Michalek JE** and Mihalko D. Linear rank procedures for matched data. *Biometrics* 1984;40:487-491.
- c. Mehrotra KG, **Michalek JE**, Mihalko D, White T. Score computation for linear rank procedures. *Journal of Statistical Computation and Simulation* 1983;16:201-211.

2. My work in endocrine disruption and exposure to Agent Orange and TCDD is summarized in a sequence of papers describing results from the Air Force Health Study, also known as the Ranch Hand Study which began in 1980 and ended in 2006. The study and others were reviewed and summarized by the Institute of Medicine in a series of text books entitled *Veterans and Agent Orange* and were considered by the Department of Veterans Affairs in determining health policy relevant to Vietnam veterans and their families. The first of these [1] displayed a trend of increased risk of type 2 diabetes with TCDD body burden and subsequent papers described the finding in relation to the NIOSH study of chemical workers exposed to TCDD, showing conflicting results, [2], a lack of variation in the TCDD elimination rate with diabetes [3], insulin sensitivity and TCDD in a glucose clamp study, and a pattern of increased risk after adjustment for days of spraying and calendar period of service [4]. This work was done while I was Principal Investigator of the study and during that period I presented results and responded to critique at the National Academy of Sciences and at hearings before House and Senate subcommittees of the United States Congress.

- a. Henriksen GL, Ketchum NS, **Michalek JE**, Swaby JA. Serum dioxin and diabetes mellitus in veterans of Operation Ranch Hand. *Epidemiology* 1997 May;8(3):252-258.
- b. Steenland K, Calvert G, Ketchum N and **Michalek JE**. Dioxin and diabetes mellitus: an analysis of the combined NIOSH and Ranch Hand data. *Occupational and Environmental Medicine* 2001 Oct;58:641-648.
- c. **Michalek JE**, Ketchum NS, Tripathi RC. Diabetes mellitus and 2,3,7,8-tetrachlorodibenzo-p-dioxin elimination in veterans of Operation Ranch Hand. *J Toxicol Environ Health A* 2003 Feb;66(3):211-221.

- d. Kern PA, Said S, Jackson WG, **Michalek JE**. Insulin sensitivity following agent orange exposure in Vietnam veterans with high blood levels of 2,3,7,8-tetrachlorodibenzo-p-dioxin. *J Clin Endocrinol Metab* 2004 Sep;89(9):4665-4672.
3. With the application of the serum TCDD measurement to the Air Force Health Study in 1987, interest focused on characterizing whole body elimination of TCDD in the study population. First attempts employed a first order pharmacokinetic model with parameters estimated based repeated measurements made in a subset of exposed veterans, the subset being those veterans who had a baseline TCDD body burden greater than 10 ppt. It was immediately obvious that the ordinary least squares estimate would be biased due to truncation. We examined this bias and mathematically derived a correction algorithm such that off the shelf software, such as SAS proc mixed, would produce unbiased estimates and summarized the algorithm in 1998 [1], subsequently applied it to an expanded database [2], and generalized it to maximum likelihood estimation [3]. Near the end of the study, we were motivated to replace the first order model with a physiologically based pharmacokinetic model based on an analysis of data from the Seveso Study [4], for which the issue of bias due to truncation was not addressed [4].
- Michalek JE**, Tripathi RC, Kulkarni PM, Gupta PL and Selvavel K. Correction for bias introduced by truncation in pharmacokinetic studies of environmental contaminants. *Environmetrics* 1998;9:165-174.
 - Michalek JE**, Tripathi RC. Pharmacokinetics of TCDD in veterans of Operation Ranch Hand: 15-year follow-up. *J Toxicol Environ Health A* 1999 Jul;57(6):369-378.
 - Mehrotra KG, Kulkarni PM, Tripathi RC, and **Michalek JE**. Maximum likelihood estimation for longitudinal data with truncated observations. *Statistics in Medicine* 2000;19:2975-2988.
 - Michalek JE**, Pirkle JL, Needham LL, Patterson Jr DG and Caudill SP. Pharmacokinetics of 2,3,7,8-tetrachlorodibenzo-p-dioxin in Seveso adults and veterans of Operation Ranch Hand. *J Expo Anal Environ Epidemiol*. 2002 Jan-Feb;12(1):44-53.

Complete List of Published Work in MyBibliography:

<https://www.ncbi.nlm.nih.gov/sites/myncbi/joel.michalek.1/bibliography/52450974/public/?sort=date&direction=ascending>

D. Research Support

Ongoing Research Support

CAP Project 5 (344-19) (Peterson)

06/01/16 – 5/31/2018

Audie Murphy Mem Vets Hospital

The overarching goals of the STRONG STAR Consortium to Alleviate PTSD (STRONG STAR-CAP) are to improve the psychological and physical health and well-being of Service Members (active duty, National Guard, and Reserve) and Veterans who have deployed in support of OEF/OIF/OND by developing and evaluating the most effective diagnostic, prognostic, preventive, treatment, and rehabilitative strategies for combat-related PTSD and comorbid conditions. The STRONG STAR Consortium to Alleviate PTSD will support projects funded by VA-ORD and DoD. This mobility assignment is funded by the Veterans Administration to support this initiative.

Role: Co-Investigator

P30 CA054174-23 (Gelfond)

08/01/09 – 07/31/19

NIH/NCI

This is a P30 Cancer Center Support Grant at the University of Texas Health Science Center at San Antonio. This grant provides funding for the conduct of cancer clinical trials, prevention studies and assistance with statistical design and analysis of translational cancer research projects. Specifically, this is for support of the biostatistics and medical informatics core which has the following aims: provide flexible, cost-efficient services to CTRC basic, clinical and translational researchers; provides services to aid in the administration of the CTRC including monitoring and reporting functions for compliance with local and Federal regulatory requirements, quality assurance review, preparation of reports and analyses for administrative meetings; to provide informatics components for the Protocol Review Committee, the Data Safety Monitoring Committee and to provide statistical assistance with the development and conduct of clinical protocol-based research.

Role: Co-Director for Biostatistics in the Bio and Bioinformatics Shared Resource

1R01 AG052697-01A1 (Espinoza)

06/01/17 – 04/30/18

NIH/NIA

Metformin for Preventing Frailty in High Risk Older

Grant Detail: The research described in this proposal will test a pharmacological intervention that could potentially be used to delay/reduce the development of frailty in older adults. In this initial study we will enroll glucose intolerant subjects, a population which encompasses approximately 1/3rd of older adults and is most likely to benefit from metformin. Because of the enormous costs associated with frailty (both personal and economic), an intervention that prevents or delays frailty, even in a sub-population of older adults, would have a major positive impact in our society. To our knowledge, this research will be the first to study a potential intervention targeted toward central pathogenic processes (inflammation and insulin resistance) involved in the etiology of frailty in humans.

Role: Consultant

5 R01 FD004400-04/FDA (Brenner)

07/01/17 – 06/30/18

CTRC

Phase 2 Study of Th-302 for the Treatment of Glioblastoma

Grant Detail: One of the most exciting and promising aspects of modern cancer drug development is the potential to personalize treatments by developing drugs that inhibit specific molecular targets. While this potential has come to fruition somewhat, overall this has been limited [37]. Success stories of personalized cancer treatments include erlotinib for anti-epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) inhibitors for non small-cell lung cancer with modest improvements in survival, anti-human epidermal growth factor receptor 2 (HER2) for breast cancer, and more recently BRAF inhibitors for patients with melanoma whose tumors carried the V600E BRAF mutation. The key to developing such targeted therapies lies in identifying responsive patient populations and tumor characteristics. Due to the molecular heterogeneity of most tumors, however, this has proven extremely challenging. In this application, we take a novel approach in targeting the metabolic state of the tumor, an approach that has not yet been significantly described in the literature. Just as novel is the pairing of an imaging agent and a therapeutic agent whose uptake mechanisms are identical to identify those patients with the highest probability of response. Finally, we integrate the least invasive and potentially simplest means of identifying these patients, through the use of emerging biomarkers of hypoxia.

Role: Consultant

1 R01 AG057431-01/NIH (Salmon)

09/15/17 – 05/31/18

NIH/NIA

Primary Fibroblast Resiliency as a Predictor of Health and Lifespan in Mice

There has been considerable effort to understand the relationship between cellular resiliency (largely defined in these studies as stress resistance) and longevity. However, prior studies have largely focused on this relationship within populations of animals; that is, addressing whether modifying longevity has a similar effect on stress resistance (or vice versa). This study expands on these prior studies with 3 innovations: 1) we will demonstrate for the first time whether patterns of resiliency in the cells of individual mice may be predictive of the lifespan or cause of death of an individual mouse. 2) Because we can establish cell lines throughout the life of the animal, we will determine whether these cell resiliency patterns change with age and determine at what age may be the most predictive marker. 3) We will discover whether interventions known to affect aging alter these patterns of cellular resiliency as a means to adapt this assay as a screen for potential novel interventions to extend longevity.

Role: Co-investigator

5 UL1 TR001120-05/BIOSTATISTIC (Clark) 05/01/15 – 04/30/18

NIH-National Center for Adv Translational Science

ISG: Inst for Integration of Med & Sci – Biostatistics

The objective of the Biomedical Informatics Program (BMIP) is to transform our clinical and translational research enterprise by providing access to and promoting the use of an interconnected set of data resources and informatics tools. We will re-engineer our informatics infrastructure in order to optimize the development of innovative methods to solve increasingly complex multidisciplinary research problems and to integrate genotypic, phenotypic, clinical, and public health data sources. In doing so, the BMIP will provide a valuable resource that promotes the overall goals of the Institute for Integration of Medicine and Science (IIMS).

Role: Co-investigator

NTI-NTRR15-11/W81XWH-15-2-0089 (Michalek) 12/01/15-09/29/18

NIH/NTI

A National Coordinating Center for Trauma Research/The PROspective Observational Vascular Injury Trial
The objective of this grant is to establish a prospective, multicenter, observational study through the AAST
Multicenter Trials Committee. This study, referred to as PROspective Observational Vascular Injury Trial or
PROOVIT, aims to capture key elements of vascular trauma presentation, diagnosis, management, outcomes
from leading trauma institutions in the US.

Role: Consultant

W81XWH-17-1-0560 (Jiang) 09/30/17-09/29/21

DoD

Development of Hemichannel-Targeting Antibody Therapies for Breast Cancer Bone Metastasis

Breast Cancer Breakthrough Award Level 3

The major objective of this grant is to develop a therapeutic antibody that targets connexin hemichannel and
can be used to treat breast cancer bone metastasis.

Role: Co-Investigator

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Shawn Norman Murphy, MD, Ph.D.

eRA COMMONS USER NAME (credential, e.g., agency login): smurphy

POSITION TITLE: Professor of Neurology, Harvard Medical School; Chief Research Information Officer, Partners HealthCare

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	Completion Date MM/YYYY	FIELD OF STUDY
University of Notre Dame, Notre Dame, IN	BS	1983	Chemistry
University of Chicago, Chicago, IL	PhD	1989	Pharmacology
Pritzker School of Medicine, Univ of Chicago, IL	MD	1991	Medicine
Beth Israel Hospital, Boston, MA	(Intern)	1992	Medicine
Massachusetts General Hospital, Boston, MA	(Resident)	1995	Neurology
Massachusetts General Hospital, Boston, MA	(Fellow)	1996	Epilepsy
Massachusetts General Hospital, Boston, MA	(Fellow)	1997	Bioinformatics

A. Personal Statement

As Chief Research Information Officer, I direct the team of Research IS and Computing at Partners HealthCare. This leverages the extensive work done in the past as Director of the Research Patient Data Registry (RPDR) for Partners HealthCare System, Inc., beginning in 1999, connecting human studies investigators to vast amounts of patient clinical data. The RPDR is a large data warehouse, with 6.7 million patients and 2.5 billion rows of clinical data, serving clinical information to 1200 researchers each year. I have unlocked billions of records for research both nationally and internationally by creating online, self-serve tools that connect researchers with the clinical-care data from entire enterprises as the open source project, Informatics for Integrating Biology and the Bedside (i2b2). This has allowed me to have a large impact on clinical research that makes use of data from medical records and their accompanying artifacts such as patient genomic and imaging data. I have worked to make the raw data from medical records useful by inventing new methods that establish high quality phenotypic data from these records, and utilized the resulting high-quality data in many original research studies. This work is with diverse teams of medical scientists performing electronic health record (EHR)-based analysis of patient phenotypes and their association with genotypes, biomarkers, and physiologic characteristics. The high throughout phenotyping and sample collection enabled by the i2b2 team, as well as many other methods they developed, has resulted in over 70 publications in the fields of depression, bipolar disease, rheumatoid disease, inflammatory bowel disease, and multiple sclerosis. I developed the ability to extend the i2b2 population analytics directly to the provider with a Strategic Health IT Advanced Research Projects (SHARP) grant from the Office of the National Coordinator, where i2b2 is extended into a national EHR infrastructure such that the integration it provides can extent into “Apps” that can run in a hospital EHR system.

Wattanasin, N., Porter, A., Ubaha, S., Mendis, M., Phillips, L., Mandel, J., Ramoni, R., Mandl, K., Kohane, I., **Murphy, S.N.** Apps to display patient data, making SMART available in the i2b2 platform. AMIA Annu Symp Proc. 2012; 2012:960-9. Epub 2012 Nov 3

B. Positions and Honors

Thesis: Modulation of Intracellular Calcium Homeostasis by Excitatory Amino Acids in Mouse Central Neurons (1989)

Board Certification: The American Board of Psychiatry and Neurology, Specialty in Neurology (1997)

Licensure: Massachusetts Board of Registration in Medicine, issue date 5/24/1995

Positions:

1997-	Attending in Neurology, Associate Neurologist, Massachusetts General Hospital, Boston, MA
1997-2005	Instructor in Neurology, Harvard Medical School, Boston, MA
2001-2013	Director, Research Patient Data Registry, Partners HealthCare, Boston, MA
2005-	Associate Director, Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA
2005-2011	Assistant Professor of Neurology, Harvard Medical School, Boston, MA
2007-2013	Medical Director, Partners HealthCare Research Computing, Boston MA
2011-2017	Associate Professor of Neurology, Harvard Medical School, Boston, MA
2013-	Chief Research Information Officer, Partners HealthCare, Boston, MA
2014-	Associate Professor of Biomedical Informatics, Harvard Medical School, Boston, MA
2017-	Professor of Neurology, Harvard Medical School, Boston, MA

MAJOR COMMITTEE ASSIGNMENTS

2011-2012 Advisory Committee to Director of NIH: Data and Informatics Section Co-Chair
2016- Board of Scientific Counselors, Lister Hill National Center for Biomedical Communications

PROFESSIONAL SOCIETIES

2007- Fellow of the American College of Medical Informatics (ACMI)

C. Contribution to Science

I have led the development of systems to integrate patient EHR data with genotypes, biomarkers, and physiologic characteristics that have been implemented extensively nationally and internationally. I achieved my first success with the invention, establishment, and continued development of the Research Patient Data Registry (RPDR) originally released in 1999. The RPDR serves as a data warehouse that integrates clinical, administrative, and research data from many data sources for the primary purpose of supporting research. Over 5,000 researchers have used the RPDR over the past 15 years, accessing it with a Web-based query tool on the private PHS Intranet. Authorized users may query against RPDR data for aggregate totals, and with proper IRB approval may extract detailed clinical data. The system delivers over 2,500 sets of detailed EHR data for research every year, supporting over 550 teams of clinical researchers, and providing critical data for \$94-\$136 million of NIH funded studies at PHS (Nalichowski et. al. AMIA Annu Symp Proc. 2006; 1044). Security and confidentiality are an integral part of the project, and we control and audit the distribution of patient data within the policies of the Institutional Review Board. Beginning in 2004, as PI of the software development cores of the Informatics for Integrating Biology and the Bedside (i2b2) NIH grant, I made the technology developed in the RPDR generally available throughout the country. The technology is now widely accepted and installed in over 120 hospitals nationally and internationally.

Murphy, S.N., Morgan, M., Chueh, H., Barnett, G.O. Optimizing Research Data Warehouse Design through Past COSTAR Query Analysis. Journal of the American Medical Informatics Association, Symposium Supplement 1999; 892-896.

Murphy, S.N., Chueh, H (2002). A Security Architecture for Query Tools Used to Access Large Biomedical Databases. Journal of the American Medical Informatics Association, Symposium Supplement 2002, pages 552-556. PMID 12463885

Murphy, S.N., Weber, G., Mendis, M., Gainer, V.S., Churchill, S., Kohane, I.S. Serving the Enterprise and Beyond with Informatics for Integrating Biology and the Bedside (i2b2). Journal of the American Medical Informatics Association, 2010 March 1; 17(2): 124-130. PMID: 20190053

Beginning in 2009 I led the creation of patient cohort analysis and discovery networks across the United States, leveraging the fact that i2b2 had already been widely accepted and installed in over 60 hospitals nationally, including four of the Harvard hospitals and five hospitals of the University of California system. We were able to tie the ability to query many of the installations together using the Shared Health Informatics Network (SHRINE) which can be used to query multiple sites simultaneously in a distributed fashion under one infrastructure. We demonstrated successful ways that such networks could be created from mixtures of i2b2 and other population tools in a demonstration for the Office of the National Coordinator as part of Query Health. I became Co-PI of one of eleven Clinical Data Research Networks for the Patient-Centered Outcomes Research Institute in 2014. Our network has 10 Hospitals that have cross-query capability and can support distributed clinical trials with i2b2 software. I also became Co-PI of a Big Data to Knowledge Center of Excellence in 2014 where we extend the i2b2 software to perform distributed computing upon big-data resources.

- Weber, G.M., **Murphy, S.N.**, McMurry, A.J., Macfadden, D., Nigrin, D.J., Churchill, S., Kohane, I.S. The Shared Health Information Network (SHRINE): A prototype federated query tool for clinical data repositories. *Journal of American Medical Informatics Association* Sep-Oct 2009; 16(5):624-630. PMID: 19567788
- Klann, J.G., Buck, M.D., Brown, J., Hadley, M., Elmore, R., Weber, G.M., **Murphy, S.N.** Query Health: standards-based cross-platform population health surveillance. *J Am Med Inform Assoc.* 2014 Jul; 21(4):650-6. PMID: 24699371.
- Mandl, K.D., Kohane, I.S., McFadden, D., Weber, G.M., Natter, M., Mandel, J., Schneeweiss, S., Weiler, S., Klann, J.G., Bickel, J., Adams, W.G., Ge, Y., Zhou, X., Perkins, J., Marsolo, K., Bernstam, E., Showalter, J., Quarshie, A., Ofili, E., Hripcsak, G., **Murphy, S.N.** Scalable Collaborative Infrastructure for a Learning Healthcare System (SCILHS): Architecture. *J Am Med Inform Assoc.* 2014 Jul; 21(4):615-20. PMID: 24821734.

I oversee the entirety of Research IS and Computing at Partners HealthCare as Director of this Department with over 45 full time personnel. Part of this work is to engage with diverse teams of medical scientists to perform research as part of a team using health record (EHR) based analysis of patient profiles/phenotypes and their association with genotypes, biomarkers, and physiologic characteristics. This leverages the extensive work I have done in the past as part of the science arm of i2b2 where overall we have produced over 50 publications in the fields of depression, bipolar disease, rheumatoid disease, inflammatory bowel disease, and multiple sclerosis. This capability is of considerable interest to Pharma and other industries and, as Director, I engage in many activities that help sponsor research using industry funding. I served as a Principal Investigator for Partners HealthCare in the Observational Medical Outcomes Partnership, through which we applied the latest algorithms and technology to investigate adverse drug events. We have invented a novel method to collect biological samples from the clinical pathology workflow of the hospitals, enabling a high-throughput method of coupling genomic analysis and the phenotypes from the EHR. This has resulted in many i2b2 team publications in the field of rheumatoid arthritis, major depression, inflammatory bowel disease, and multiple sclerosis unifying EMR and genomic data.

Murphy, SN, Churchill, S., Bry, L., Chueh, H., Weiss, S., Lazarus, R., Zeng, Q., Dubey, A., Gainer, V., Mendis, M., Glaser, J., Kohane, I.; Instrumenting the Health Care Enterprise for Discovery Research in the Genomic Era. *Genome Research* 2009; September 19(9): 1675-81. PMCID: PMC2752136

Brownstein, J.S., **Murphy, S.N.**, Goldfine, A.B., Grant, R.W., Sordo, M., Gainer, V., Colecchi, J.A., Dubey, A., Nathan, D.M., Glaser, J.P., Kohane, I.S. (2010). Rapid Identification of Myocardial Infarction Risk Associated With Diabetes Medications Using Electronic Medical Records. *Diabetes Care* 33: 526-531. PMID: 20009093

O'Dushlaine, C., Ripke, S., Ruderfer, D.M., Hamilton, S.P., Fava, M., Iosifescu, D.V., Kohane, I.S., Churchill, S.E., Castro, V.M., Clements, C.C., Blumenthal, S.R., **Murphy, S.N.**, Smoller, J.W., Perlis, R.H. Rare copy number variation in treatment-resistant major depressive disorder. *Bio Psychiatry*. 2014 Oct 1;76(7):536-41. Doi: 10.1016/j.biopsych.2013.10.028. PMID: 24529801.

Castro, V.M., Apperson, W.K., Gainer, V.S., Ananthakrishnan, A.N., Goodson, A.P., Wang, T.D., Herrick, C.D., **Murphy, S.N.** Evaluation of matched control algorithms in EHR-based phenotyping studies: a case study of inflammatory bowel disease comorbidities. *J Biomed Inform.* 2014 Dec;52:105-11. Doi:10.1016/j.jbi.2014.08.012. Epub 2014 Sep 6. PMID: 25196084.

Images are among the best phenotype information that we work with in medicine, and I have led projects to utilize high-throughput images for clinical research that have allowed the high density of information that can be acquired in images to be used in studying patient phenotypes. This was shown in my early work that was done at a cellular lever where it was shown that calcium movement resulted from glutamate activation in Hippocampal neuronal cultures. Images were added to a phenotyping workflow within the Bioinformatics Research Network, and that led to allowing imaging studies to be acquired in a high-throughput fashion, where through a CTSA Supplement and later an RO1, we led the development of medical imaging software that allows the research self-serve access to a Picture Archiving and Communication System (PACS). This was applied to the creation of normative brain magnetic resonance imaging of children from 0 to 6 years old that could then be used to form the baseline for comparison of abnormal imaging such as we acquired for children with hypoxic ischemic injury.

- Murphy, S.N.**, Miller, R.J. A unique glutamate receptor regulates Ca²⁺ mobilization in hippocampal neurons. Proceedings of the National Academy of Sciences, USA 1988; 85:8737-8741.
- Murphy, S.N.**, Mendis, M.E., Grethe, J.S., Gollub, R., Kennedy, D., Rosen, B.R.; A Web Portal that Enables Collaborative Use of Advanced Medical Image Processing and Informatics Tools through the Biomedical Informatics Research Network (BIRN), AMIA, Fall Symposium 2006: 579-83. PMID: 17238407
- Murphy, S.N.**, Herrick, C., Wang, Y., Wang, T.D., Sack, D., Andriole, K.P., Wei, J., Reynolds, N., Plesniak, W., Pieper, S., Gollub, R.L. High Throughput Tools to Access Images from Clinical Archives for Research. J Digit Imaging 2015 Apr;28(2):194-204 PMID: 25316195.

Complete List of Published Work in Harvard Catalyst “Profiles”:

<https://connects.catalyst.harvard.edu/Profiles/display/Person/66554>

D. Research Support

Ongoing Research Support

- 3 U01 HG008685-03S1 (Murphy/Weiss/Smoller/Karlson) 09/2015-05/2019
NIH
eMERGE Phase III Clinical Center at Partners HealthCare
The goal of this project is to develop and implement phenotype algorithms for detection and classification of diseases by mining many data sources.
Role: Multi-Principal Investigator
- 5 U54 HG007963-04 (Murphy/Kohane) 09/2014 - 08/2018
NIH
Patient-Centered Information Commons
The goal of this project is enable the use of big data for research on patients.
Role: Multi-Principal Investigator
- CDRN-1306-04608 (Murphy/Mandl/Schneeweiss) 04/2014-10/2018
PCORI
Scalable Collaborative Infrastructure for a Learning Healthcare System Phase II
The goal of this project is to create a clinical data research network to enable the selection of patient cohorts for clinical effectiveness research.
Role: Multi-Principal Investigator
- 5 R01 HG009174-02 (Murphy/Macrae/Aronson/Rehm) 09/2016-08/2020
NIH
Developing i2b2 into a Health Innovation Platform for Clinical Decision Support in the Genomics Era
The goal of this project is to build a decision support system based upon “SMART” Apps that use i2b2 as the population analytics repository of phenotypic and genotypic information.
Role: Multi-Principal Investigator

3 OT2 OD024612-01S1 (Murphy/Smoller/Weiss/Karlson/O'Connor) NIH A New England Enrollment Center for the PMI Cohort Program The goal of this project is to recruit patients for a national biobank cohort. Role: Multi-Principal Investigator	09/2016-02/2018
5 R00 LM011575-04 (Wagholarikar) NIH A Framework to Enhance Decision Support by Invoking NLP The goal of this project is to develop a framework for reminder systems that can encompass broader areas of practice, due to their capability to utilize free-text in addition to structured EHR data. Role: Co-Investigator	02/2016-01/2019
1 R01 LM012594-01 (Lin) NIH Improving comparative effectiveness research through electronic health records continuity cohorts Produce generalizable algorithms to identify high-validity continuity cohorts in a given EHR. Role: Co-Investigator	09/2016-08/2020
5 UL1TR001102-05 (Nadler) NIH Harvard Clinical and Translational Science Center Provide enriched resources to educate and develop the next generation of researchers trained in the complexities of translating research discoveries into clinical trials and ultimately into practice. Role: Co-Investigator	09/2013-04/2018
5 UL1TR001857-02 (Reis) NIH University of Pittsburgh Clinical and Translational Science Institute – ACT Supplement Enable the regulatory-compliant identification of patients who are eligible for clinical studies. Role: Site PI	07/2016–05/2021
<u>Completed Research Support</u>	
NIH 5 R01 EB014947-03 (Murphy/Gollub/Grant) NIH mi2b2 Enabled Pediatric Radiological Decision Support The goal of this project is to extend the recently developed Medical Imaging Informatics Bench to Bedside (mi2b2) Workbench software to make medical images collected during routine clinical care available to clinical translational investigators. Role: Multi-Principal Investigator	08/2012-10/2016
OBS-1505-30683 (Weiler) PCORI PCORNet Bariatric Surgery Study to determine the outcome of bariatric surgery in obese patients. Role: Site PI	02/2016-01/2018
OBS-1505-30699 (Weiler) PCORI PCORNet Obesity Observational Study: Short and Long-Term Effects of Antibiotics on Childhood Growth Determine the effects of antibiotics in children and whether they lead to obesity. Role: Site PI 1/17/18	02/2016-01/2018

OTHER SUPPORT – BOKOV, A. F.**ACTIVE**

CDRN-1306-04631 Phase II (Tirado-Ramos – PI) 09/05/15 – 09/04/18 2.40 Calendar

Patient-Centered Outcomes Research Institute \$414,410

The goal of this grant is to establish the Great Plains Consortium which will develop a large clinical data research network of electronic health record information to facilitate the conduct of clinical trials and comparative effectiveness research investigations.

Role: Co-Investigator

1 P30 AG044271-01A1 (Gelfond – PI) 05/01/15 – 04/30/20 0.84 Calendar

NIH/NIA \$396,171

San Antonio Claude D. Pepper Older Americans Independence Center

The purpose of this program is to test interventions in non-human primates and humans that are hypothesized to slow aging and prevent or delay age-related diseases.

Role: Co-Investigator

PENDING

1UL1 TR002362 (Clark, Hargreaves, Shireman, Dougherty - MPI) 05/01/18 to 04/30/22 2.4 Calendar

NIH/NCATS (CTSA) ~\$3,700,000/year

Institute for Integration of Medicine & Science: A Partnership to Improve Health (CTSA)

The major goal of this project is to achieve optimal integration of clinical and translational research, education, training, and career development across all UTHSCSA schools and among our partner organizations in the South Texas region. IIMS will focus existing and newly developing resources and intellectual capital on advancing the discipline of clinical and translational research for the improvement of human health. There is no support for Dr. Bokov's research from this CTSA grant.

Role: Co-Investigator

OVERLAP

None

If awarded the KL2, I will reduce my effort on other projects, so that my total funded research effort does not exceed 100%.

For New and Renewal Applications – DO NOT SUBMIT UNLESS REQUESTED**PHS 398 OTHER SUPPORT**

Provide active and pending support for all senior/key personnel. **Other Support** includes all financial resources, whether Federal, non-Federal, commercial or institutional, available in direct support of an individual's research endeavors, including but not limited to research grants, cooperative agreements, contracts, and/or institutional awards. Training awards, prizes, or gifts do not need to be included.

There is no "form page" for other support. Information on other support should be provided in the *format* shown below.

For instructions and information pertaining to the use of and policy for other support, see [NIH Grants Policy Statement, Section 2.5.1: Just-in-Time Procedures](#). Neither the application under consideration nor the current PHS award for this project should be listed as Other Support.

Effort devoted to projects must be measured using person months. Indicate calendar, academic, and/or summer months associated with each project.

Format**NAME OF INDIVIDUAL****ACTIVE/PENDING**

Project Number (Principal Investigator) Source Title of Project (<i>or Subproject</i>)	Dates of Approved/Proposed Project Annual Direct Costs	Person Months (Cal/Academic/ Summer)
The major goals of this project are...		

OVERLAP (*summarized for each individual*)

Ramirez, Amelie G.**ACTIVE**

R25 CA134301-01A2 Ramirez (PI) National Cancer Institute <i>Éxito! Latino Cancer Research Leadership Training</i>	08/01/2010 -07/31/2020 \$264,710	1.56 calendar
---	-------------------------------------	---------------

This project trains Latino master's-level students and professionals to pursue a doctoral degree and cancer research. Over 20% of program alum have applied for and/or enrolled in doctoral programs.

Role: Principal Investigator (PI)

RWJF-75213 Ramirez (PI) Robert Wood Johnson Foundation <i>Salud America! The RWJF Research Network to Prevent Obesity Among Latino Children</i>	02/15/2018 – 08/31/2018 \$500,000	3.0 calendar
---	--------------------------------------	--------------

Salud America! The RWJF Research Network to Prevent Obesity Among Latino Children

The reauthorization of *Salud* activates its 100,000-person network to create healthy changes to reduce Latino childhood obesity. SA! will monitor online network activities with robust online and social media analytical tools and conduct periodic surveys/polls and other assessments to determine the extent to which SA! educational materials and network activities are positively influencing levels of self and collective efficacy of network members. The analysis will inform future efforts to promote the Robert Wood Johnson Foundation's Health Kids, Healthy Weight initiative and reduce Latino childhood obesity.

Role: Principal Investigator (PI)

PP170099 Ramirez (PI) Cancer Prevention & Research Institute of Texas (CPRIT) SMS Cessation Service for Young Adult Smokers in South Texas	03/01/2018 – 02/28/2021 \$1,302,641	1.2 calendar
--	--	--------------

This award is a continuation of PP140176. The project aims to overcome accessibility and barriers by using culturally and linguistically appropriate mobile channels and messages to recruit smokers to an evidence-based SMS cessation service for South Texans.

Role: Principal Investigator (PI)

SAB160005 Ramirez (PI) 04/01/2016–03/31/2019 .24 calendar
Susan G. Komen Foundation (National) \$200,000
Improving Adherence to Endocrine Hormonal Therapy (EHT) among Breast Cancer Patients

This study is enrolling 120 breast cancer patients for an evidence-based intervention with: 1) a bilingual, culturally tailored, personalized, interactive mobile app; and 2) support from a patient navigator; vs. usual care.
Role: Principal Investigator (PI)

1 R13 MD012457-01 (Ramirez (PI) 09/19/2017 – 08/31/2018 .12 calendar
NIH \$49,935
Advancing the Science of Cancer in Latinos

The conference, titled “Advancing the Science of Cancer in Latinos” will be held from February 21-23, 2018 in San Antonio, Texas. The conference will be hosted by the Cancer Therapy and Research Center and the Institute for Health Promotion Research at UT Health San Antonio, and will be the first ever international conference focused specifically on the science of cancer health disparities among Latino communities.
Role: Principal Investigator (PI)

RWJF-74949 Ramirez (PI) 10/01/2017 – 05/31/2018 .12 calendar
Robert Wood Johnson Foundation \$179,200
Salud America! Sustainability Assessment Project

Salud America! (Client) is a national program of the Robert Wood Johnson Foundation (RWJF), managed and directed by the Institute for Health Promotion Research at UT Health San Antonio. Focused on addressing the important issue of Latino childhood obesity, Salud America! has become a leading researcher and advocate, capable of disseminating its message through a national network developed with key leaders in local communities. Through its partnership with RWJF, the Salud America! program has flourished, but it is now time to determine if Salud America! is programmatically and economically viable beyond the support of RWJF.

As Salud America! looks towards its future, it seeks to assess the viability of its model, and, in doing so, develop a sustainable business model to guide its efforts. Salud America! will subcontract with La Piana Consulting to formulate a new organizational strategy and create a business plan that will enable it to accomplish these goals.

PENDING

None.

OVERLAP

None.

OTHER SUPPORT – Rodriguez, Ronald

ACTIVE

None

PENDING

None

OVERLAP

None

exceed 100%.

For New and Renewal Applications (PHS 398) – DO NOT SUBMIT UNLESS REQUESTED**PHS 398 OTHER SUPPORT**

Provide active and pending support for all senior/key personnel. **Other Support includes all financial resources, whether Federal, non-Federal, commercial or institutional, available in direct support of an individual's research endeavors, including but not limited to research grants, cooperative agreements, contracts, and/or institutional awards.** Training awards, prizes, or gifts do not need to be included.

There is no "form page" for other support. Information on other support should be provided in the *format* shown below, using continuation pages as necessary. The sample below is intended to provide guidance regarding the type and extent of information requested.

For instructions and information pertaining to the use of and policy for other support, see Other Support in the Supplemental Instructions, Part III, Policies, Assurances, Definitions, and Other Information.

Effort devoted to projects must be measured using person months. Indicate calendar, academic, and/or summer months associated with each project.

Format

**NAME OF INDIVIDUAL
ACTIVE/PENDING**

Project Number (Principal Investigator) Source Title of Project (<i>or Subproject</i>)	Dates of Approved/Proposed Project Annual Direct Costs	Person Months (Cal/Academic/ Summer)
The major goals of this project are... <u>OVERLAP</u> (<i>summarized for each individual</i>)		

MICHALEK, J.E.**ACTIVE**

CAP Project 5 (344-19) (Peterson) 06/01/16 – 5/31/2018 1.2 CM
Audie Murphy Mem Vets Hospital \$21,047

The overarching goals of the STRONG STAR Consortium to Alleviate PTSD (STRONG STAR-CAP) are to improve the psychological and physical health and well-being of Service Members (active duty, National Guard, and Reserve) and Veterans who have deployed in support of OEF/OIF/OND by developing and evaluating the most effective diagnostic, prognostic, preventive, treatment, and rehabilitative strategies for combat-related PTSD and comorbid conditions. The STRONG STAR Consortium to Alleviate PTSD will support projects funded by VA-ORD and DoD. This mobility assignment is funded by the Veterans Administration to support this initiative.

Role: Co-Investigator

P30 CA054174-23 (Gelfond) 08/01/09 – 07/31/18 2.4 CM
NIH/NCI \$737,916

This is a P30 Cancer Center Support Grant at the University of Texas Health Science Center at San Antonio. This grant provides funding for the conduct of cancer clinical trials, prevention studies and assistance with statistical design and analysis of translational cancer research projects. Specifically, this is for support of the biostatistics and medical informatics core which has the following aims: provide flexible, cost-efficient services to CTRC basic, clinical and translational researchers; provides services to aid in the administration of the CTRC including monitoring and reporting functions for compliance with local and Federal regulatory requirements, quality assurance review, preparation of reports and analyses for administrative meetings; to provide informatics components for the Protocol Review Committee, the Data Safety Monitoring Committee and to provide statistical assistance with the development and conduct of clinical protocol-based research.

Role: Co-Director for Biostatistics in the Bio and Bioinformatics Shared Resource

1R01 AG052697-01A1 (Espinoza) NIH/NIA	06/01/17 – 04/30/18 \$377,643	1.2 CM
Metformin for Preventing Frailty in High Risk Older		
Grant Detail: The research described in this proposal will test a pharmacological intervention that could potentially be used to delay/reduce the development of frailty in older adults. In this initial study we will enroll glucose intolerant subjects, a population which encompasses approximately 1/3rd of older adults and is most likely to benefit from metformin. Because of the enormous costs associated with frailty (both personal and economic), an intervention that prevents or delays frailty, even in a sub-population of older adults, would have a major positive impact in our society. To our knowledge, this research will be the first to study a potential intervention targeted toward central pathogenic processes (inflammation and insulin resistance) involved in the etiology of frailty in humans.		
Role: Consultant		
5 R01 FD004400-04/FDA (Brenner) CTRC	07/01/17 – 06/30/18 \$300,024	1.8 CM
Phase 2 Study of Th-302 for the Treatment of Glioblastoma		
Grant Detail: One of the most exciting and promising aspects of modern cancer drug development is the potential to personalize treatments by developing drugs that inhibit specific molecular targets. While this potential has come to fruition somewhat, overall this has been limited [37]. Success stories of personalized cancer treatments include erlotinib for anti–epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK) inhibitors for non small-cell lung cancer with modest improvements in survival, anti–human epidermal growth factor receptor 2 (HER2) for breast cancer, and more recently BRAF inhibitors for patients with melanoma whose tumors carried the V600E BRAF mutation. The key to developing such targeted therapies lies in identifying responsive patient populations and tumor characteristics. Due to the molecular heterogeneity of most tumors, however, this has proven extremely challenging. In this application, we take a novel approach in targeting the metabolic state of the tumor, an approach that has not yet been significantly described in the literature. Just as novel is the pairing of an imaging agent and a therapeutic agent whose uptake mechanisms are identical to identify those patients with the highest probability of response. Finally, we integrate the least invasive and potentially simplest means of identifying these patients, through the use of emerging biomarkers of hypoxia.		
Role: Consultant		
1 R01 AG057431-01/NIH (Salmon) NIH/NIA	09/15/17 – 05/31/18 \$309,396	0.96 CM
Primary Fibroblast Resiliency as a Predictor of Health and Lifespan in Mice		
There has been considerable effort to understand the relationship between cellular resiliency (largely defined in these studies as stress resistance) and longevity. However, prior studies have largely focused on this relationship within populations of animals; that is, addressing whether modifying longevity has a similar effect on stress resistance (or vice versa). This study expands on these prior studies with 3 innovations: 1) we will demonstrate for the first time whether patterns of resiliency in the cells of individual mice may be predictive of the lifespan or cause of death of an individual mouse. 2) Because we can establish cell lines throughout the life of the animal, we will determine whether these cell resiliency patterns change with age and determine at what age may be the most predictive marker. 3) We will discover whether interventions known to affect aging alter these patterns of cellular resiliency as a means to adapt this assay as a screen for potential novel interventions to extend longevity.		
Role: Co-investigator		
5 UL1 TR001120-05/BIOSTATISTIC (Clark)	05/01/15 – 04/30/18 NIH-National Center for Adv Translational Science \$184,969 ISG: Inst for Integration of Med & Sci – Biostatistics	1.8 CM
The objective of the Biomedical Informatics Program (BMIP) is to transform our clinical and translational research enterprise by providing access to and promoting the use of an interconnected set of data resources and informatics tools. We will re-engineer our informatics infrastructure in order to optimize the development of innovative methods to solve increasingly complex multidisciplinary research problems and to integrate		

genotypic, phenotypic, clinical, and public health data sources. In doing so, the BMIP will provide a valuable resource that promotes the overall goals of the Institute for Integration of Medicine and Science (IIMS).

Role: Co-investigator

NTI-NTRR15-11/W81XWH-15-2-0089 (Michalek) 12/01/2015-11/30/2017 0.3 CM
NIH/NTI \$31,688

A National Coordinating Center for Trauma Research/The PROspective Observational Vascular Injury Trial
The objective of this grant is to establish a prospective, multicenter, observational study through the AAST Multicenter Trials Committee. This study, referred to as PROspective Observational Vascular Injury Trial or PROOVIT, aims to capture key elements of vascular trauma presentation, diagnosis, management, outcomes from leading trauma institutions in the US.

Role: Consultant

1 R01 DK103841-02/NIH (DeFronzo) 09/01/2015-08/31/2018 0.6 CM
NIH-Diabetes/Digestive/Kidney Diseases \$968,789

Durability of Early Combination Therapy vs Conventional

The hypothesis is that the initiation of combination therapy with a glucagon-like peptide-1 analogue (liraglutide) plus a thiazolidinedione (pioglitazone) in type 2 diabetic subjects poorly controlled blood glucose levels on metformin plus sulfonylurea (SU) will achieve superior and more durable glycemic control with fewer side effects compared to basal-bolus insulin. Accordingly we will compare the efficacy and safety of initiating combination therapy with liraglutide plus pioglitazone versus basal (glargine) plus premeal insulin in type 2 diabetic subjects who are suboptimally controlled ($HbA1c > 7.0\%$) with metformin plus SU on the achievement of optimal glycemic control ($HbA1C < 6.5\%$), the amelioration of whole body (muscle) insulin resistance and improvement in beta cell function.

Role: Consultant

5 R01 DK107680-02/NIH (DeFronzo) 07/15/2016-06/30/2018 0.6 CM
NIH-Diabetes/Digestive/Kidney Diseases \$1,038,846

Sglt2 Inhibition and Stimulation of Endogenous Glucose

Role: Consultant

Grant Detail: We and others have demonstrated that glucosuria produced by inhibition of the renal sodium-glucose transporter-2 (SGLT2) lowers the fasting plasma glucose (FPG) concentration but causes a “paradoxical” increase in endogenous glucose production (EGP). Although the increase in EGP can be viewed as a compensatory mechanism that opposes urinary glucose loss and prevents the development of hypoglycemia in NGT individuals, in T2DM individuals it occurs while the plasma glucose concentration is still in the hyperglycemic range. Moreover, it offsets by ~50% the urinary glucose loss produced by SGLT2 inhibitors and attenuates their glucose lowering ability. The increase in EGP following SGLT2 inhibition is associated with an increase in plasma glucagon concentration and decrease in plasma insulin concentration. Because renal glucose production is stated to be unresponsive to an increase in the plasma glucagon concentration, it is likely that the liver contributes, at least in part, to the increase in EGP [which is triggered by glucosuria]. However, an increase in glucose production by the kidney cannot be excluded. We hypothesize that there is a previously unrecognized, [completely novel] “reno-hepatic” interaction that participates in the regulation of plasma glucose concentration.

Funding Agency: DoD

Title Development of Hemichannel-Targeting Antibody Therapies for Breast Cancer Bone Metastasis (PI: Jiang)

Period: 9/2017 - 08/2021 0.6 CM

Role: Co-Investigator

Total Costs: \$3,241.607

Grant Detail: The biggest challenge facing breast cancer patients is cancer recurrence and metastasis. Bone metastases occur frequently in patients with advanced breast cancers and typically render the disease incurable. Currently the treatment option is very limited. Therefore, there is a pressing need to identify novel

targets and develop new, specific therapies with improved therapeutic efficacies but less toxicity. Our team recently established connexin (Cx) 43 hemichannel as a de novo drug target for breast cancer bone metastasis. More importantly, we have developed a monoclonal, humanized antibody against Cx43 and have shown that activation of the hemichannels by this antibody inhibits breast cancer cell migration/invasion in vitro and bone metastasis in vivo with minimal toxicity. We propose to assess and optimize the therapeutic value of this novel, humanized antibody that represents a first-in-class therapy for breast cancer bone metastases which remains an unmet medical need, and we will conduct preclinical studies on the optimized antibody therapy in the treatment of breast cancer bone metastasis.

PENDING

Funding Agency: CPRIT

Title Repurposing Anti-Depressant Imipramine for Treating Breast Cancers (PI: Kaklamani)

Period: 03/2018 - 02/2022 0.6 CM

Role: Co-Investigator

Total Costs: \$1,472,948

Grant Detail: The PD-1/PD-L1 pathway is a target for FDA approved immunotherapy against several types of cancers. Interestingly, imipramine is known to inhibit the expression of the IL-10 and to elevate IL-12, which activates the innate immunity via natural killer (NK) cells and adaptive immunity via cytotoxic T-lymphocytes. Based on these observations, we hypothesize that imipramine is an effective drug for treating TNBC; that it could enhance PARPi response by targeting DNA damage response, and could improve immunosurveillance and anti-PD-1/PD-L1 immunotherapy response. To test these hypothesis, we propose the following aims: In aim 1, we will perform a four cohort trial with imipramine as monotherapy or in combination with TSR-042 (an anti PD-1 antibody) or PARPi niraparib. In aim2, we will validate that imipramine acts by targeting the FOXM1 and associated signaling-dependent DNA damage response and by targeting the PD-1/PD-L1 immunomodulatory axis. The proposed study will set the stage for a new paradigm of treating TNBCs using Imipramine therapeutics.

Funding Agency: CPRIT

Title Targeting Fatty Acid Synthase in Lipogenic Metastatic Breast Cancer (PI: Brenner)

Period: 03/2018 - 02/2022 1.2 CM

Role: Co-Investigator

Total Costs: \$1,204,876

Grant Detail: The proposed studies combine a novel hypothesis – that omega-3 fatty acids can prevent AI recurrence through inhibition of COX-2 activity, while targeting a specific patient population, the obese postmenopausal woman. Importantly, while the combination therapy of celecoxib with AI therapy has already been explored in two prior studies, including the GINECO Phase III and CAAN neoadjuvant trials, we are the first to show that resistance to AIs is encountered in part due to prostaglandin impact on local aromatase expression and that this was associated with the obese state and not in patients of normal habitus. There have been no prior studies of COX2 inhibitors with AIs to overcome resistance in the obese population, and prior studies were inadequately powered to detect this in subgroup analysis. These studies will be the first to evaluate clinically an association between omega-3 fatty acid supplementation, ER α activity and tumor prognostic markers. These studies have the potential to significantly move the field forward and have an immediate impact both on our understanding of the biology of obesity-related breast cancer as well as treatment interventions to improve outcome by preventing relapse and disease progression.

Funding Agency: NIH

Title Institute for Integration of Medicine & Science: A partnership to Improve Health (PI: Clark)

Period: 05/2018 - 04/2023 1.2 CM

Role: Co-Investigator

Total Costs: \$5,600,000

Grant Detail: The objective of the Biomedical Informatics Program (BMIP) is to transform our clinical and translational research enterprise by providing access to and promoting the use of an interconnected set of data resources and informatics tools. We will re-engineer our informatics infrastructure in order to optimize the development of innovative methods to solve increasingly complex multidisciplinary research problems and to integrate genotypic, phenotypic, clinical, and public health data sources.

Role: Co-Investigator

Funding Agency: DoD

Title Boosting Antitumor Immunity for Triple Negative Breast Cancer trough ER-beta (PI: Li)

Period: 10/2017 - 09/2020 0.6 CM

Role: Co-Investigator

Total Costs: \$829,137

Grant Detail: The Principal Investigators (PIs) of this application propose to test the hypothesis that rallying estrogen receptor (ER)-beta antitumor activity using clinically safe ER-beta agonists will improve clinical responses in patients with triple-negative breast cancer (TNBC) and further that ER-beta-selective agonists will improve the clinical efficacy of existing anticancer immunotherapies and make them effective to treat TNBC. The project's specific aims are (1) elucidate the mechanism(s) by which ER-beta signaling boosts antitumor immunity in TNBC and (2) galvanize ER-beta signaling to boost the antitumor efficacy of TNBC immunotherapy using validated mouse models, including humanized mice challenged with autologous patient-derived xenografts (PDX).

Funding Agency: NIH

Title Study of Targeted Agents with Radiochemotherapy in Pediatric Tumor Xenografts Malignancy (PI: Houghton)

Period: 04/2018 - 03/2023 0.24 CM

Role: Co-Investigator

Total Costs: \$1,143,750

Grant Detail: We hypothesize that a strategy to protect the genomic integrity of normal cells will reduce the incidence of therapy-related malignancy. We have developed a novel strategy to protect the normal cells from DNA double strand breaks (DSBs) in-vivo and to keep the proliferating normal human lymphocytes from losing telomere length after radiation therapy in-vitro. In this application, we propose to investigate the mechanisms behind these findings to ultimately come up with a new strategy such as small molecules to reduce the incidence of therapy-related malignancy.

OVERLAP

If pending projects are awarded, effort will be adjusted accordingly to prevent any over commitment.

For New and Renewal Applications – DO NOT SUBMIT UNLESS REQUESTED**PHS 398 OTHER SUPPORT**

Murphy, Shawn
ACTIVE

3 U01 HG008685-03S1 (Murphy/Weiss/Smoller/Karlson) NIH eMERGE Phase III Clinical Center at Partners HealthCare The goal of this project is to develop and implement phenotype algorithms for detection and classification of diseases by mining many data sources.	9/2015 – 5/2019	1
5 U54 HG007963-04 (Murphy/Kohane) NIH Patient-Centered Information Commons The goal of this project is enable the use of big data for research on patients.	09/2014 – 08/2018	1
CDRN-1306-04608 (Murphy/Mandl/Schneeweiss) PCORI Scalable Collaborative Infrastructure for a Learning Healthcare System Phase II The goal of this project is to create a clinical data research network to enable the selection of patient cohorts for clinical effectiveness research.	04/2014-10/2018	1
5 R01 HG009174-02 (Murphy/Macrae/Aronson/Rehm) NIH Developing i2b2 into a Health Innovation Platform for Clinical Decision Support in the Genomics Era The goal of this project is to build a decision support system based upon “SMART” Apps that use i2b2 as the population analytics repository of phenotypic and genotypic information.	09/2016-08/2020	1
3 OT2 OD024612-01S1 (Murphy/Smoller/Weiss/Karlson/O'Connor) NIH A New England Enrollment Center for the PMI Cohort Program The goal of this project is to recruit patients for a national biobank cohort.	09/2016-02/2018	1
5 R00 LM011575-04 (Wagholarikar) NIH A Framework to Enhance Decision Support by Invoking NLP The goal of this project is to develop a framework for reminder systems that can encompass broader areas of practice, due to their capability to utilize free-text in addition to structured EHR data.	02/2016-01/2019	1
1 R01 LM012594-01 (Lin) NIH Improving comparative effectiveness research through electronic health records continuity cohorts Produce generalizable algorithms to identify high-validity continuity cohorts in a given EHR.	09/2016-08/2020	1

5 UL1TR001102-05 (Nadler) NIH Harvard Clinical and Translational Science Center Provide enriched resources to educate and develop the next generation of researchers trained in the complexities of translating research discoveries into clinical trials and ultimately into practice.	09/2013-04/2018	1
5 UL1TR001857-02 (Reis) NIH University of Pittsburgh Clinical and Translational Science Institute – ACT Supplement Enable the regulatory-compliant identification of patients who are eligible for clinical studies.	07/2016–05/2021	1

PENDING

None.

OVERLAP

None.

CANDIDATE'S BACKGROUND

I graduated from the University of Michigan with a BS in Cellular and Molecular Biology in 2000. In 2008 I defended my dissertation, "The Role of Somatotropic and Estrogen Signaling in Longevity and Resistance to Oxidative Stress" and earned my PhD in Physiology at the UT Health Science Center Graduate School of Biomedical Sciences [1]–[3]. During my doctoral training I noticed that a disproportionate share of research effort goes into data acquisition compared to analysis and interpretation.

I went back to earn an MS in Applied Statistics at UT San Antonio so that I could develop tools which would make it easier to analyze data in a valid, transparent, and reproducible manner. Statistical methods often find uses in multiple domains, and I wanted to do research that more directly impacts patients and communities. There are limits to the applicability of animal models, and no good analogues to complex external influences like socioeconomic disparities. The next step in my journey from basic to translational research was in late 2013 when Dr. Tirado-Ramos hired me into the newly created Clinical Informatics Research Division (CIRD) at the Department of Epidemiology and Biostatistics (DEB). Soon after I was promoted to Instructor.

Funded by a PCORI subcontract, we deployed an Integrating Informatics from Bench to Bedside (i2b2) data warehouse which we loaded with de-identified patient data from the UT Med electronic medical record (EMR) system. We supported mostly early T1 clinical trials and T3 health services and comparative effectiveness projects. Through our partnership with TriNetX (which I facilitated by helping navigate the review and approval process) we became a link to T2 pharma-sponsored trials, and our recent entry into the NCATS Accrual for Clinical Trials (ACT) network may multiply our clinical trial participation several-fold.

I worked on projects spanning many different topics and specialties including functional status in ALS patients [4], [5], stochastic mortality models [6], recruitment for a clinical trial of Metformin's ability to delay age-related frailty, identifying high-risk surgery patients [Bokov et al., in preparation], and a large multi-site study of obesity and attitudes toward participation in clinical research [Davis et al., in preparation]. But the overarching theme tying together my roles in these projects was finding innovative ways to use EMR data for translational research. A major bottleneck is the time and effort needed to organize data for analysis. I wrote software called DataFinisher to fill this gap which I will describe in my Research Strategy (below) where it plays a central role.

My time at CIRD so far has been a crash-course not only in medical informatics but also IRB protocols, multi-site grants, planning and budgeting projects, and learning to build trust and consensus between institutional stakeholders. Now I need a structured, comprehensive mentored translational research program to consolidate my on-the-job experience and add to it an organized introduction to how EMR data is created and used in the clinic and how population health data is used to inform efforts at prevention and intervention. Furthermore I need to learn design principles for informatics systems in general and advanced topics in i2b2 architecture in particular. Our campus has a shortage of experienced research mentors in medical informatics, especially after some recent faculty departures. To bridge this gap I asked Dr. Shawn Murphy at Harvard Medical School, co-author of i2b2, to join my mentoring team and he agreed. He will supervise my software efforts and host me for a 4-8 week externship at Partners Healthcare in Massachusetts.

Collaborations

Statistics and informatics are inherently collaborative fields. One must understand not only the data, but the perspective of one's collaborators to be able to devise effective ways to achieve their research, quality, and patient-care goals. Of my 37 peer-reviewed publications, roughly half have been collaborations with other research teams. I am completing the first paper reporting the results of a large PCORI-funded multi-site patient survey with co-authors at Kansas University Medical Center and University of Wisconsin-Madison, helping a local pathologist who won a multi-site pilot grant launch his project, and in March I completed my contribution to an in-preparation manuscript by a lead author at Zhejiang University School of Medicine in Hangzhou, China-- he contacted me on ResearchGate because of my expertise in the R statistical language.

I advocate a collaborative funding model for CIRD wherever practical-- i.e. using existing capabilities for preliminary data and budgeting new features into a grant. Some level of fee-for-service turned out to be unavoidable, but I helped prepare at least 20 different grants, plus a blitz of small pilot applications co-sponsored by IIMS and the Greater Plains Collaborative (GPC) wherein I helped 10 local PIs submit letters of intent in a one-month period, 4 of them going on to submit proposals. Not everything I worked on got funded

but I helped secure a significant portion of CIRD's coverage.

Activity	Fa 2018	Sp 2019	Fa 2019	Sp 2020
Didactic Training - 20%				
TSCI 6015 Topics in Cancer Prevention	1hr/wk			
TSCI 6065 Health Services Research		2hr/wk		
TSCI 6103 Selected Topics in Advance Research Ethics			1hr/wk	
TSCI 6102 Practicum in IRB Procedures		1hr/wk		
TSCI 5070 Responsible Conduct of Research	2hr/wk			
TSCI 6001 Introduction to Translational Science	1hr/wk			
MEDI 4153 Informatics and Advanced Evidence-Based Medicine		1hr/wk		
INTD 4104 Improving Patient Outcomes	1hr/wk			
INTD 5030 Introduction To Patient Care			5hr/wk	
ELEC 5004 Surgical Oncology Service	1hr/wk			
EPIC & Sunrise training	1 wk			
Seminars, Workshop, Meetings: Local - 10%				
Grand Rounds - Urology	1hr/wk	1hr/wk	1hr/wk	1hr/wk
Urology Tumor Board	2hr/2wk	2hr/2wk	2hr/2wk	2hr/2wk
Population Science and Prevention Program Monthly Meeting	1hr/mo	1hr/mo	1hr/mo	1hr/mo
Rising Leaders Workshop	1 day			
Advancing the Science of Cancer in Latinos	3d/yr		3d/yr	
GWNI/GrantSeekers	2d/mo	2d/mo	2d/mo	2d/mo
Health Services Research (events at UHS & VA)	1d/wk	1d/wk	1d/wk	1d/wk
Meetings with mentors (separate or concurrent, as needed)	1-4hr/wk	1-4hr/wk	1-4hr/wk	1-4hr/wk
Seminars, Workshop, Meetings: KL2 - 7%				
Entering Mentoring (together with Dr. Michalek)	8hr			
SOAR Seminars	1hr/wk	1hr/wk	1hr/wk	1hr/wk
SPARK Seminars	1hr/wk	1hr/wk	1hr/wk	1hr/wk
Managing Career Development Seminars	2hr/mo	2hr/mo	2hr/mo	2hr/mo
Scholar/Director Meetings	2hr/mo	2hr/mo	2hr/mo	2hr/mo
IIMs Frontiers of Translational Science Research Day		1dy/yr	1dy/yr	
Association for Clinical and Translational Science Meeting		4dy/yr	4dy/yr	
Seminars, Workshop, Meetings: External - 5%				
American Medical Informatics Association	3dy/yr		3dy/yr	
American Society of Clinical Oncology		5d/yr		5d/yr
American Association for Cancer Research - Health Disparities	4d/yr		4d/yr	
American Society for Preventative Oncology		3d/yr		3d/yr
Externship - 5%				
MGH with Dr. Murphy			4-8 wk	
Research Plan - 53% (see research strategy section for detailed timeline)				

Table 1 Training and career development activities

Potential To Develop Into A Successful Independent Investigator

For an early-career researcher I have a good amount of grantsmanship experience, collaborations, and peer-reviewed publications. Since becoming faculty, I averaged 2-3 peer reviewed publications per year. I been first-author on three of my last four. I also have experience with IRB protocols and non human-subjects declarations (for de-identified data). I am an active member of the medical informatics community having presented at two international IEEE conferences and one national AMIA conference. By now I know 6 different programming languages, four dialects of SQL, and have a working knowledge of medical coding systems-- ICD9/10, CPT, and LOINC. These might seem like technical skills rather than academic ones, but there is no substitute for mastery of operational details when developing realistic research plans, budgets, and timelines.

Another strength I have as a researcher is that I know first-hand some of the important gaps in the use of

medical informatics for T3 and T4 studies. My research plan emanates directly from obstacles I encountered and observed my clinician scientist colleagues encounter.

On the other hand, I still do not know what a surgeon's routine interaction with the EMR system looks like nor how treatment decisions are made for kidney cancer. I don't know the typical sequence of events between intake and discharge, nor how each of those gets documented and billed. Knowing these things will make me better able to spot errors in the data as well as recognize simplifying principles that could streamline data analysis, and this is why clinical mentorship under the KL2 grant will be crucial.

Commitment To KL2 And Further Plans

I will commit at least 75% of full-time professional effort to the KL2 program and related career development activities and I have the full support of my department chair in this.

CAREER GOALS AND OBJECTIVES

My long-term goal is to make retrieval of patient data for analysis fast, unobtrusive, and robust so that researchers spend most of their effort answering questions rather than asking them, with my own models of health disparities serving as an example. For this I need to achieve within the next 5 years my medium-term goals of advancing to the academic rank of professor, building my reputation in translational medical informatics, and leading a research team to assist with code maintenance, testing, dissemination, and implementation of innovative features while. This in turn requires grant support and training by a multidisciplinary mentoring team like the one I brought together here. I will use this KL2 to demonstrate the clinical and translational usefulness of my software and statistical model and in the process prepare my first grant as lead PI. A possible candidate is the National Library of Medicine K01 PAR-16-204. I will use that to build a track record toward my first R01 grant as lead PI. During the KL2 period I will replicate my study two patient populations-- South-Central Texas (Aim-2) and Massachusetts (Aim-3). But for this to change clinical practice, it will be necessary to show that it generalizes across a representative sample of health systems, in a larger multi-site study later on. If participating sites are CTSA/ACT members and already have i2b2, many of the costs, like design of data collection forms and hiring personnel to fill them can be avoided.

CANDIDATE'S PLAN FOR CAREER DEVELOPMENT/TRAINING ACTIVITIES DURING AWARD PERIOD, INCLUDING ANY PLANNED EXTERNSHIPS

In order to achieve my career objectives, I have identified a set of training needs which I list below and then discuss in greater detail.

- Gain a detailed understanding of current best practices in preventing and treating kidney cancer and what the important evidence gaps are in these fields.
- Learn advanced concepts in writing open-source informatics software and disseminating it to researchers.
- Receive structured mentorship in continuing to grow my professional skills as a translational scientist: leadership, grantsmanship, scientific communication, and organization.
- Build, strengthen, and extend collaborative relationships with other translational scientists including clinicians, population health researchers, and other informaticians.

Aim-3 of my research plan is built around a 4-8 week externship at MGH, where Dr. Murphy practices. I will work with Dr. Murphy's informatics team to deploy my DataFinisher app on their i2b2.

Professional Responsibilities Beyond KL2 Activities

I anticipate outside my 75% KL2 time assisting Dr. Ramirez, who is also my Department Chair, in helping IIMS and other stakeholders utilize CIRD services in an impactful and sustainable manner in coordination with other DEB services. I will be glad to participate in a search committee to recruit additional faculty for CIRD, which it needs. Also, I have been named our site's representative to the NCATS Informatics Domain Task Force which has a monthly webinar and a yearly breakout session at the AMIA summit. All of these activities are by definition relevant to translational informatics and will help further my goals of improving leadership skills and collaboration-building.

MENTOR STATEMENT

Dr. Bokov has worked out with Dr. Rodriguez and Dr. Ramirez a thorough curriculum for deepening his understanding of cancer prevention and the current standards of care for kidney cancer. Through the MEDI, INTD, and ELEC courses listed in Table 1, Dr. Bokov will get some of the same training that medical students do, learning what information physicians look for when interviewing patients, how they document it in EMR systems, and how they use this information to make decisions. Dr. Bokov will shadow surgeons in the department of Urology. He has already started attending meetings of the Urology tumor board and of the Population Science and Prevention Program at the Mays Cancer Center. We expect him to be ready to submit his first KL2-related poster for the September 2018 deadline of ASPO 2019 Annual Meeting and then present a further update of his progress at the 2019 Advancing the Science of Cancer in Latinos conference. To grow his leadership skills, Dr. Bokov has signed up for the September Rising Leaders workshop offered by our University. He will use some of this non-KL2 time to submit the remainder of his in-preparation manuscripts during the last quarter of 2018. The consensus of the mentoring team is that Dr. Bokov will increase his current publication rate of 2-3 peer-reviewed articles per year to 4 with at least half of them first author, with additional abstracts and presentations.

We are aware of the shortage of mid-level and senior informatics faculty at UT Health. This should emphasize the need for career advancement of junior faculty like Dr. Bokov rather than be an obstacle to it. Luckily, Dr. Bokov was able to bring in informatics co-mentorship from outside in the way of Dr. Shawn Murphy MD from Harvard Medical School. Dr. Murphy is a co-author of i2b2 and during the first year will guide Dr. Bokov's programming efforts remotely, helping Dr. Bokov master advanced methods in informatics programming, best practices in disseminating software, i2b2 design goals and philosophy, as well as the professional skills of planning, budgeting, and promoting an informatics software project. Dr. Murphy and Dr. Bokov are both accustomed to working with remote-collaboration tools of their field including Atlassian, GitHub, online chat, and videoconferencing. Dr. Bokov has also linked with Dr. Murphy via the American Medical Informatics Association mentorship program. In early 2019, after DataFinisher is feature-complete as described in Aim 1 of the Research Strategy, Dr. Bokov will do an 8 week externship at Massachusetts General Hospital (MGH) and Harvard Medical school. Dr. Bokov will work alongside Dr. Murphy's informatics team to deploy DataFinisher first in a test environment and then on an i2b2 instance containing de-identified records equivalent to the CIRD i2b2. Under Dr. Murphy's supervision, Dr. Bokov will submit a DataFinisher presentation to the AMIA 2019 Informatics Summit (due date: August 2018) and use the feedback from the summit to submit an extended paper to the Journal of the American Medical Informatics Association.

Dr. Michalek is Dr. Bokov's mentor of record for Promotion & Tenure, and will work closely with him on his promotion packet. Dr. Michalek and Dr. Bokov will both enroll in Entering Mentoring, when it is next offered in July. Dr. Bokov is enrolling because this will develop his professional leadership and organization skills. In addition to weekly meetings with the mentoring team, Dr. Michalek and Dr. Bokov already see each other on a daily basis because they have adjacent offices and Dr. Michalek is also a collaborator of Dr. Rodriguez and Dr. Ramirez. Of the senior faculty in Dr. Bokov's home department Dr. Michalek especially understands the challenges and opportunities of i2b2 and "big data" because of his own expertise with the National Inpatient Sample, a large de-identified data source which Dr. Michalek is using for his research into national cancer trends, and provides mentorship for Dr. Bokov in advanced statistical methods.

In summary, Dr. Bokov will meet with each of his four mentors at least one hour per week either as a group or individually depending on his needs at the time, with Dr. Murphy joining remotely. Dr. Bokov's presentations at National conferences (AMIA, ASCO, ASPO, AACR, ACTS), IDTF service, externship at Harvard Medical School, active participation in local events (Grand Rounds, Tumor Board, ASPP, and ASCL), and the KL2 structured activities will all serve his career development goal of building and strengthening a network of collaborators that he can leverage in future proposals for multi-site informatics studies which are the next step after the initial two-site study he will do during this KL2. Dr. Bokov will receive continued responsible conduct of research training via coursework (TSCI 5017, 5070, and 6102); from Grant Writing for New Investigators; at KL2 activities including SOAR, SPARK, and director meetings; and informally from each of his mentors.

Dr. Bokov's data extraction software is a portable product of his research to take with him wherever his career leads. His research protocol will make provisions to let him keep any de-identified data he obtains here during

his KL2 so that he can continue to publish using that if he moves to a new institution.

At this stage in the evolution of our data warehouse, most data analysis projects end up requiring staff time to make adjustments to the data warehouse or the data extraction process. This is a normal part of translation and results in permanent enhancements that benefit all subsequent researchers. Dr. Bokov is capable of making far more efficient use of the local i2b2 team's effort than any other researcher on campus (and the express goal of his Aim 1 is to streamline data acquisition even more). But if his KL2 research budget does not cover the necessary testing, troubleshooting, and other services needed by his project, the DEB will gladly donate additional staff effort as needed.

Dr. Bokov teaches TSCI 5050, Introduction to Data Science and co-teaches TSCI 5076, Introduction to Informatics. In order to free his time up for research and career development, he will be exempted from teaching these classes during the KL2 period. During his non-KL2 time will continue to advise Dr. Ramirez on CIRD-related topics and then whomever she appoints to be the new Chief of CIRD. Finally Dr. Bokov is our representative to the NCATS Informatics Domain Task Force, whose yearly in-person meeting is concurrent with the yearly AMIA summit that Dr. Bokov already plans to attend as part of his career development plan. These responsibilities will each contribute to Dr. Bokov's goal of gaining leadership and organizational experience while and networking with future collaborators.

Whatever the outcome of this KL2 proposal, we are helping Dr. Bokov prepare his promotion packet to Assistant Professor. At that rank he will be a candidate for Chief of the Clinical Informatics Research Division if that position is still open by the time he completes his KL2. Independent of that, he will use this KL2 to prepare a grant proposal for an expanded multi-site study which we are confident will be his next step in a successful career as an independent researcher. All four members of Dr. Bokov's mentoring team are senior faculty experienced in mentoring junior faculty. Dr. Ramirez has mentored over 200 individuals at least 20 of which were junior faculty. Dr. Michalek sits on both the DEB and the School of Medicine Promotion and Tenure committees and is mentoring two Assistant Professors. Dr. Rodriguez has mentored 6 students, 9 postdoctoral trainees, and over a dozen other individuals with doctoral degrees under various arrangements. Dr. Murphy has mentored 12 individuals, 6 of them junior faculty.

SPECIFIC AIMS

Among US-born Texans of Hispanic ancestry (7.3 million, 27% of the State's population), annual age-adjusted mortality rates kidney cancer are 1.5-fold and 1.4-fold those of non-Hispanic whites for males and females respectively [7]. Consistently with this, in my preliminary analysis (Figure 1 in Research Plan) I found an accelerated progression to metastasis among Hispanic patients at UT Health. Understanding how socioeconomic status (SES), lifestyle, interaction with the healthcare system, metabolic syndrome, and family history each contribute to this disparity will help design appropriate cancer prevention strategies, improve clinical care pathways, and target them where they will have the most impact.

I will use an inverse propensity of treatment weighted (IPTW) [8] survival model to rank the importance of several groups of possible mediators of disparity in kidney cancer progression among Hispanic patients compared to non-Hispanic whites. I will obtain the data from i2b2 [9], an open-source data warehouse used by sites in the CTSA ACT network including ours. Developed by my co-mentor Dr. Shawn Murphy at Harvard under an NIH grant, i2b2 is used to great advantage for clinical trial recruitment but wider use in population health and health services research is impeded by lack of a simple data export path for statistical analysis. I wrote a prototype app called DataFinisher [10] which bridges this gap. I will use the protected time from this KL2 to finish this DataFinisher under the guidance of Dr. Murphy (Aim-1) then extract and analyze local data (Aim-2), and then disseminate DataFinisher to Massachusetts General Hospital (MGH) where I will replicate the data collection and analysis to demonstrate the technical feasibility of using DataFinisher for multi-site data extraction and to determine the extent to which South Central Texas results generalize to a population where Hispanic patients are primarily of Caribbean rather than Mexican descent.

To guide me in my cross-disciplinary informatics software and health-services project I recruited a mentoring team of nationally recognized experts consisting of: Dr. Shawn Murphy MD (medical informatics), Dr. Amelie Ramirez DPH (disparities and cancer prevention), Dr. Ronald Rodriguez MD (surgical oncology), and Dr. Joel Michalek PhD (retrospective analysis of large data sets). They will guide me in obtaining the training and experience I will need to carry out the following Specific Aims:

- To complete my work on open source software for data-extraction from the i2b2 data warehouse.
- To use data extracted with the novel app completed in Aim-1 to test the primary hypothesis that Hispanic kidney cancer patients have an increased risk of progression to metastasis and the secondary hypothesis that a maternal history of diabetes and cancer mediates this disparity by way of metabolic syndrome.
- To deploy my software and replicate my analysis at MGH to determine applicability of the findings to a population where the majority of Hispanic patients are of Caribbean descent and in the process demonstrate feasibility of a larger multi-site study for a future grant.

At the conclusion of this study I will have a de-identified dataset representing kidney cancer patients in three health systems (UHS, UTMed, MGH) spanning the years 2013 to 2019 along with an analysis pipeline which can be used to follow this patient cohort prospectively over multiple data updates in the respective i2b2 systems. I will use the publications and collaborations developed during this KL2 to obtain funding for a continuation of this study and expansion to additional i2b2 sites in the ACT, GPC, or ARCH networks.

SIGNIFICANCE

Since 2015, the population-adjusted incidence of kidney cancer has been increasing at 1-1.3% per year[11]–[13] with 65,340 new cases expected in 2018 [14]. Mortality rates are about 1 in 3 but among US-born Texans of Hispanic ancestry (7.3 million, 27% of the State's population), annual age-adjusted mortality rates kidney cancer are 1.4 to 1.5-fold those of non-Hispanic Whites [7]. My own preliminary analysis of UT Health patient records shows that risk of progression to metastasis is worse among Latinos (Figure 1).

There is evidence that glycolytic switch [15] is a key event in the pathogenesis of kidney cancer [16]— cells of the renal epithelium shift away from oxidative phosphorylation in the mitochondria toward glycolysis in the cytoplasm. This metabolic pathway is less efficient but can produce ATP and glycolytic intermediates rapidly to support rapid proliferation and resistance to chemotherapeutic agents. The glycolytic switch is triggered by increased reactive oxygen species (ROS) by way of HIF-1a. The ROS are produced by NADPH oxidases NOX4 and NOX1 in the mitochondria. These in turn are believed to be inappropriately stimulated by chronic low-grade inflammation associated with metabolic syndrome. This is where the molecular mechanisms link up to the organismal and environmental variables that I will be analyzing in order to understand what it is about about being Hispanic and living in South Central Texas that mediates the outcome disparities we and others [ref] have observed.

Dr. Rodriguez is starting a gene sequencing effort using his department's biorepository. My Aim-1 will deliver the ability to extract any needed set of variables from i2b2 for linkage to that gene expression data and my Aim-2 will suggest which variables should be linked. Biosample studies often over-represent patients of European origin [17]. My propensity scores from Aim-2 will help improve future studies by informing better sampling and weighting strategies. The major mediators identified by the model will help design and target better prevention programs and care pathways, contributing to a comprehensive model of kidney cancer that encompasses exposures at the cellular, organismal, clinical, and community scales.

EMR data can provide sample-sizes for T3 patient outcomes studies and T4 population health studies far beyond what would be feasible to accrue by any other means. The i2b2 data warehouse [9] addresses the problems of deidentification and terminology alignment and can gather records from multiple silos under a common data model. Together with a related system called Shared Health Research Information Network (SHRINE) [18], i2b2 forms the backbone of the NCATS Accrual to Clinical Trials (ACT) network. Every CTSA site that meets ACT's membership requirements has an i2b2 instance, including our own.

Use of i2b2 to get non-aggregated visit-level data for analysis is less common than for eligibility counts and other high-level descriptive results. To get the data from an i2b2 star-schema into one unified table requires a structured query language (SQL) statement with at least as many self-join clauses as there are variables. As the number of variables becomes large, such queries become complicated to write. But the obstacle to automated SQL generation has been that different variables need to be handled in different ways. One example are laboratory results with numeric values versus those with coded values (such as color, or normal/abnormal). The same variable may need different representations depending on the goals of the researcher— some studies only need to know whether a patient is a current smoker, while others need to know whether they are a smoker at the time of the visit, or the specific type of tobacco product they consume. I am working on an i2b2 enhancement, DataFinisher, which represents a compromise between generalizability and customizability [10]. DataFinisher analyzes the properties of each variable, including cardinality and missingness and assigns the most appropriate transformation, using a list of rules that can be customized by the local informatics team. My prototype is already available to UT Health researchers with the appropriate IRB

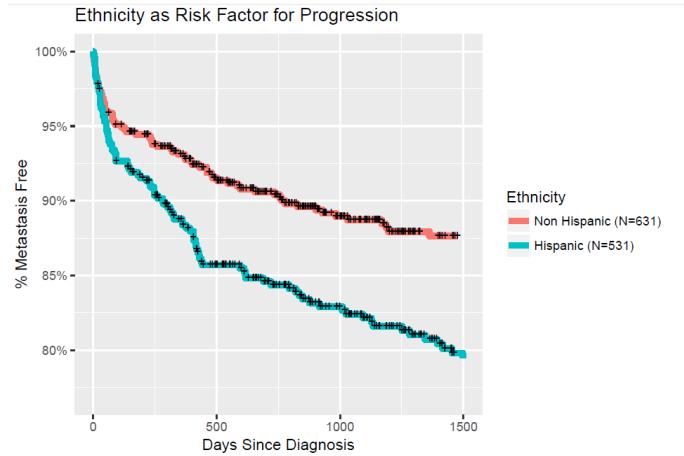


Figure 1 time until metastatic progression, Hispanic vs Non-Hispanic

authorization. The final result is a CSV file that can be read directly into Excel, SAS, R, and almost any analysis software but a significant amount of work remains on the rules and on preparing it for dissemination to other i2b2 sites.

By saving translational scientists the complex and error-prone chore of manually post-processing i2b2 data and offering the flexibility to amend many aspects of the dataset without re-running the query, DataFinisher can democratize T4 research and encourage wider utilization of i2b2 in health outcomes and population health studies by translational scientists who are not programmers or informaticians. This, in turn, will create additional returns on the investment that ACT sites have made in deploying i2b2.

I will be the first researcher to use DataFinisher to develop a causal model for kidney cancer based on EMR and public data sources, then test it out at a second site. This will set the stage for a subsequent study for validating this model at a larger sample of US health systems. Such a multi-site study would answer the question of whether the kidney cancer disparity we are investigating affects Hispanic persons in general or specifically those of Mexican descent-- but even the current two-site design can yield preliminary data for this question.

Clinical trials are the focus of ACT, and these too would be enhanced by streamlined data extraction. Alongside patient demographics and contact information, researchers could, for example, be provided with each patient's most recent labs, vitals, principal diagnoses, or any other supported data elements they need. Researchers could reference these columns as a cross-check during eligibility screening or use them as covariates later, when analyzing their results.

INNOVATION

For most cancers Hispanic persons have lower incidence and mortality, but for kidney cancer there is conflicting evidence. Analysis of data from the Texas Cancer Registry and the National Inpatient Sample [Michalek et al. in preparation] fails to produce evidence of disparity, while Pinheiro et al. [7] and my own preliminary analysis of UT Medicine patients suggest worse outcomes for Hispanic patients (Figure 1). Because of the size of the UHS catchment area and the fact that it is a safety-net hospital, it is a more representative sample of the Bexar County population than UT Medicine alone and can provide substantive evidence for or against Pinheiro et al. [7] at least at the regional level. Aim-3 then goes beyond the regional level, replicating the analysis at MGH whose Hispanic population is primarily of Caribbean rather than Mexican descent. If no disparity is found, there will still be useful outcomes from the MGH arm of this study and I discuss them in the Pitfalls and Alternative Approaches section, below.

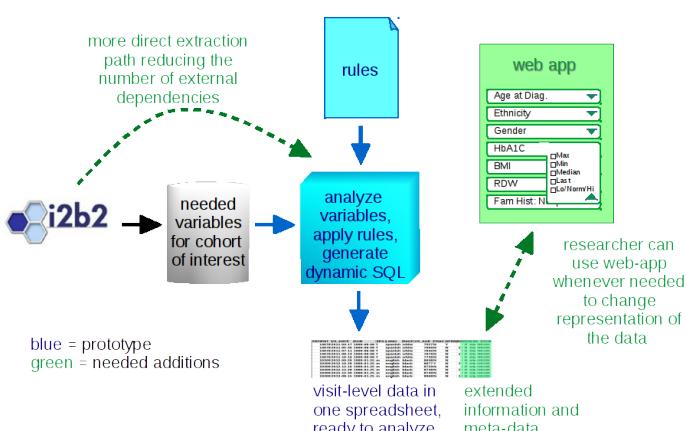


Figure 2 existing and planned features for DataFinisher

The informatics aspect of this study is innovative in that it could overturn the prevailing expectation that collection and preparation of secondary data must be a slow, labor-intensive processes and that the needs of T4 retrospective studies vary too much for automation to be feasible. My statistical training and experience with past collaborations lead me to the opposite conclusion. There is an infinite diversity of possible research plans but underlying them is a more constrained set of analysis strategies funneling into an even smaller number of expected tabular structures. Most of these, in turn, are various special cases of just one general schema similar to the "tidy data" paradigm of Hadley Wickham [19] but simpler because it only needs to encompass studies using EMR data. Briefly, each patient is represented by several rows of data, one for each follow-up period (visit, day, week, month, etc depending on the study); each data element has its own column; and patient-level values which do not change (e.g. demographics) are repeated for each row belonging to that patient. This covers survival analysis, mixed-effect models for longitudinal data, generalized linear models, linear regression, and can be easily aggregated to produce contingency tables. Embodying this approach as a tangible piece of software will help overcome the perception by many clinician scientists that EMR data is

impractical to use beyond cohort selection and feasibility counts.

I will not be the first to use causal analysis and IPTW to identify mediators of health disparities [20] but I will be the first to apply this powerful method to kidney cancer. I am proud to be an early adopter of study pre-registration, an emerging best practice to promote transparency and combat publication bias. Prior to commencing data collection and analysis I will register this study with the Center for Open Science [21].

APPROACH

Aim 1: To complete my work on open source software for data-extraction from the i2b2 data warehouse.

Initial data source on which DataFinisher will operate: The i2b2 data warehouse from which DataFinisher will pull data has de-identified electronic health records from over 379,000 adult patients going back to 2007. In 2017, the CIRD repository protocol was amended to authorize making merged data from UT Health and UHS available for research use. The initial testing is now complete and the dataset goes live this June (2018). The number of unique patients will increase to 1.2 million. My inclusion criteria are adult patients diagnosed with kidney cancer (ICD10 C64 or ICD9 189.0) followed by nephrectomy (i.e. non-surgical cases are excluded). Table 1 shows eligible patients used for my preliminary data, (i.e. not yet including UHS). For the upcoming study I will update this query to specifically select patients with renal clear cell carcinoma (RCC, which is 75% of cancer cases) but if a sufficient sample sizes are available, I may do additional analysis on papillary (10%) carcinoma. Also, I will exclude the tiny numbers of patients with HIV or with metastatic tumors *prior* to their first RCC diagnosis. The former, to avoid skewing the CCI and the latter, to insure that the index cancer is in fact the primary tumor as well as to avoid skewing the CCI (in which metastatic cancer and AIDS are the two most heavily weighted components).

In addition to EMR fields (diagnoses, lab results, vital signs, medications, procedures, and demographics), our i2b2 contains death dates from Social Security records, and the contents of the local North American Association of Central Cancer Registries (NAACCR). At the time of preliminary analysis, NAACCR records were not linked to the EMR, but after the June 2018 data refresh it will be possible to match NAACCR data with EMR data for the same patients. Two other features scheduled to become available at this time are median household incomes and educational attainment from the 2016 American Community Survey (ACS) at the block-group level of precision. They will be imported into i2b2 using code I wrote earlier this year. The variables for my study will require each of these sources, and will be described in the methods for Aim 2 below.

Software development: DataFinisher is written in Python. The code will be maintained in a public GitHub repository in strict isolation from real data (even though it is de-identified) enforced by configuration settings and git commit hooks. No visit-level data will be published but the i2b2 queries producing the data will be contributed to PheKB [22] to facilitate re-use at other sites. DataFinisher documentation will generated from code-comments using Sphinx [23], a best practice among Python developers to insure documentation remains synchronized with code.

I have had a prototype of DataFinisher installed on our local i2b2 since 2015, updated as time permitted. I have three proximal goals for completing this app: streamline the variable-handling rules, retain extended data, and facilitate dissemination.

Currently, the variable-handling rules DataFinisher uses for choosing how to represent each variable are themselves complex and difficult to modify so my first goal is to generalize and simplify them.

My second goal is to implement a key new feature-- retaining additional information about each variable (including modifiers, units, and value flags). This information will be stored in specially encoded columns

	UT Health	Diagnosed	Metastatic
Total	379481	973	294
Gender			
Female	215,394 (56.8%)	399 (41%)	94 (32%)
Male	164,087 (43.2%)	574 (59%)	200 (68%)
Vital Status			
Deceased	8,177 (2.2%)	87 (8.9%)	68 (23.1%)
Living	371,304 (97.8%)	886 (91.1%)	226 (76.9%)
Ethnicity			
Hispanic	92,499 (24.4%)	392 (40.3%)	140 (47.6%)
Non-Hispanic	198,861 (52.4%)	536 (55.1%)	148 (50.3%)
Unknown/Other	88,121 (23.2%)	45 (4.6%)	6 (2%)
Age			
18-34	85,177 (22.4%)	23 (2.4%)	6 (2%)
35-44	55,622 (14.7%)	63 (6.5%)	16 (5.4%)
45-54	60,735 (16%)	166 (17.1%)	43 (14.6%)
55-64	71,031 (18.7%)	311 (32%)	98 (33.3%)
65-74	59,372 (15.6%)	268 (27.5%)	87 (29.6%)
75-84	30,094 (7.9%)	110 (11.3%)	36 (12.2%)
=>85	17,450 (4.6%)	32 (3.3%)	8 (2.7%)

Table 2 UTHealth cohort from pilot data

alongside ordinary numeric or categoric data. Analysis software will still treat a file produced by DataFinisher as a table of data (with some additional columns of text that can be easily skipped). Yet, uploading the file back into a DataFinisher web-app which I will also develop will give the researcher a menu-driven interface to modify their variables whenever they wish, even specifying multiple derived columns for the same variable (e.g. one column with only maternal family history codes and another with only paternal ones).

My third goal for DataFinisher is making it easy to disseminate to other sites-- required for my long-term goal of leading and facilitating multi-site studies leveraging ACT and PCORI networks. I will refactor and document the installation process in accordance to i2b2 plugin deployment standards, with guidance from Dr. Murphy. As part of this, I will reduce the dependencies on external Python libraries and decouple DataFinisher from a GPC-developed app called DataBuilder [24] on which it currently relies for database connections. In Aim-3 I will describe field-testing DataFinisher at MGH.

Aim 2: To use data extracted with the novel app completed in Alm-1 to test the primary hypothesis that Hispanic kidney cancer patients have an increased risk of progression to metastasis and the secondary hypothesis that a maternal history of diabetes and cancer mediates this disparity by way of metabolic syndrome.

Hypothesized causal structure: In Figure 2 is an directed acyclic graph (DAG) with green arrows indicating hypothesized causal chains connected to Hispanic ethnicity and black arrows indicating independent effects. The dashed arrow is the direct correlation of Hispanic ethnicity with cancer progression not explained by other variables. The thick arrows represent the secondary hypothesis.

Variables: Except where indicated otherwise all variables will be from the UT Health and UHS EMR systems by way of i2b2. The main exposure will be Hispanic ethnicity. Sex and age at diagnosis will be the primary covariates. The main response variable will be time from initial diagnosis to either the first diagnosis of a secondary tumor or last tumor-free follow-up with an accompanying censoring indicator. Socioeconomic (SES) variables will include: median household income [ACS], fraction of adults completing high school [ACS], country of birth, preferred language, and type of insurance. Lifestyle variables will include smoking, alcohol, and use of anti-inflammatory drugs. Family history (Fam Hist) of neoplasia and diabetes (distinguished by maternal and paternal) will be used as a proxy for genetic predisposition. Variables related to metabolic syndrome (Metab) will include: hemoglobin A1c levels, diagnosis of diabetes, HDL, VLDL, systolic and diastolic blood pressure, and BMI. Overall comorbidity (Comorb) will be represented by the Charlson Comorbidity Index [25] calculated from problem-list codes as per Quan et al. [26]. Proxies for access to and seeking of care (Care) will be: number of visits per year, number of lab tests and imaging orders per visit, time spent with provider per visit, time from diagnosis to surgery, enrollment in adjuvant trials, and stage at presentation (NAACCR).

Statistical analysis:

Each patient in the dataset will be randomly assigned to a development, validation, or test subset. The development subset will be used to fit the propensity scores that will be described below, as an initial screen of the correlation structure predicted by the causal diagram, and for fitting the Cox proportional hazard models also described below. The validation set will be used to determine if the results obtained from the development data are repeatable. After all models and other analysis decisions are finalized, the test subset will be used to obtain significance levels and parameter estimates for publication.

To test the primary hypothesis, a Cox proportional hazard model modified to allow multiple follow-up visits per patient [27] will be fit using Sex, Age at presentation, Charlson comorbidity index, and Hispanic ethnicity as the

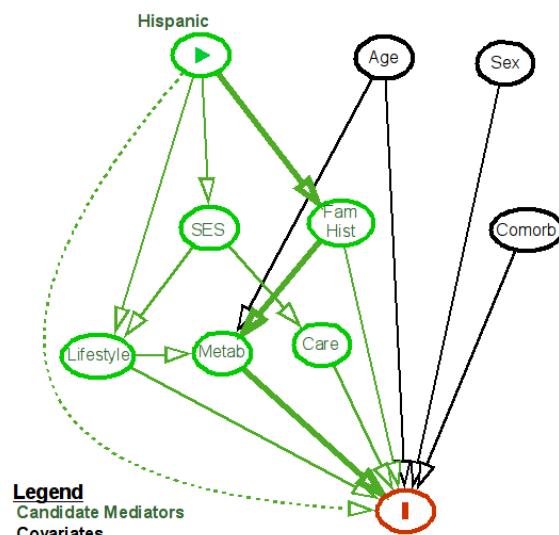


Figure 3 hypothesized causal structure

predictors. The response will be days elapsed from initial diagnosis to secondary tumor or last recorded visit.

If the primary hypothesis is confirmed (Hispanic variable is significant), the secondary hypothesis will be tested in stages. An ordinal variable representing maternal family history (2 = diabetes *and* kidney cancer, 1 = diabetes *or* kidney cancer, 0 = neither) will be added to the model. The hypothesis predicts that the effect of the Hispanic variable will diminish and the maternal history variable will have a significant effect. A paternal history variable will likewise be added. The hypothesis predicts that the maternal variable will remain significant. A propensity score will be constructed from the SES variables (see above) for Hispanic ethnicity. The inverse of this propensity score, the Inverse Probability of Treatment Weighting (IPTW) [8], will be used to adjust the Cox model. This will be repeated, adding variables for Lifestyle, Care, and Metabolic Syndrome to the propensity score. The hypothesis predicts that the maternal history variable will remain significant until the inclusion of Metabolic Syndrome variables in the IPTW. After that point the prediction is that the maternal history variable's effect will diminish.

If the primary hypothesis is not confirmed, the above secondary hypothesis will be modified to treat family history rather than Hispanic ethnicity as the main exposure.

Interpretation: As more variables are added, the IPTW adjusts for more and more differences between Hispanic patients and the non-Hispanic white (NHW) comparator group, so the metastasis risk attributable to Hispanic ethnicity diminishes. The fraction by which it diminishes is interpreted as the mediating effect for that group of variables. Beyond testing the causal hypothesis, the behavior of the other main effects as the IPTW term is updated will provide valuable exploratory data as a byproduct, for improved understanding of the other causal paths shown in Figure 2.

Power analysis: Time to progression will be of central importance in Specific Aim 2. Our work with existing data has suggested that time to progression (Figure 1), varies significantly with ethnicity (Hispanic SH(4.1)=0.79, hazard rate=0.057, Non-Hispanic SN(4.1)=0.87, hazard rate=0.034), where SH(t) and SN(t) are the progression free survival distribution functions for Hispanic and Non-Hispanic patients respectively and time is measured in years. Assuming proportional hazards, 1 year of accrual, and a total time of 5 years, and no losses to follow-up, the to attain 80% power for testing $H_0: SH(t)=SN(t)$ versus $H_1: SH(t)\neq SN(t)$ with $\alpha=0.05$ the proposed study will require only $N=624$ patients (equal number of Hispanic and non-Hispanic patients) [PASS Version 15, NCSS Kaysville UT 2017]. The sample sizes attainable with the merged UHS/UTHealth dataset readily surpass this number.

Reproducibility and transparency: I will follow RECORD guidelines [29] in formally documenting my research strategy and will pre-register the protocol with the Open Science Framework, a practice designed to combat publication bias [21]. RMarkdown and Pandoc will be used to generate the final manuscript submissions to insure that tables, figures, and values embedded in manuscripts are automatically synchronized with the analysis results. Analysis scripts will be written using the R language [30] and versioned (separately from the data) in GitHub to insure that an audit trail exists of every modification made. The data-files will not be versioned but their MD5 hashes will be recorded whenever analysis is run, and visible in the resulting report documents.

Aim 3: To deploy my software and replicate my analysis at MGH to determine applicability of the findings to a population where the majority of Hispanic patients are of Caribbean descent and in the process demonstrate feasibility of a larger multi-site study for a future grant.

MGH: MGH, located in Boston, MA and operated by Partners Healthcare, is the teaching partner of Harvard Medical School. MGH has 1.5 million ambulatory visits and 51,000 inpatient stays per year. According to Dr. Murphy, there are a total of 18,500 kidney cancer cases in the MGH EMR. 19% of Boston's population is Hispanic. Of Hispanic Bostonians the most common ancestries are Puerto Rican (43%) and Dominican (17%), with Mexican descent comprising less than 7%.

Informatics software dissemination: If linkage to census data is not already in place in MGH, I will share with them the code we use for doing so at our site. Likewise I can share our NAACCR code if needed. I will work with Dr. Murphy and his informatics team to test and deploy DataFinisher at MGH.

Statistical analysis and interpretation: I will replicate my query at MGH, and use DataFinisher to extract a de-

identified dataset. For terms included in the ACT ontology (e.g. diagnoses) the ACT SHRINE mappings will be used to assist with replicating the query. Nevertheless, this step will likely reveal differences in how the data is coded and organized between the two sites, and I will need to do several more revisions of the DataFinisher rule file before a final dataset can be extracted. This dataset will have exactly the same structure as the one in Aim-2, but different values and a different number of results. The analysis will be as described for Aim-2 and using the same scripts but substituting in the results from MGH rather than UT Health/UHS permitting a side-by-side comparison of the sites.

Possible Pitfalls and Alternatives

Failure to confirm the primary hypothesis in Aim-2 despite a large sample size and adjustment for all relevant covariates would mean that, together Dr. Michalek's finding of no disparity in data from the National Inpatient Sample and from the Texas Cancer Registry, we would have publishable evidence contrary to Pinheiro et al. The secondary hypothesis can fail at any of the iterative steps described in Aim-2, and such failures can be informative. Briefly, part of the causal chain can be supported by the results, and the point where results are no longer as predicted is where the hypothesis needs to be updated and tested in future studies. As mentioned in the Aim-2 section, the behavior of other terms in the model can give clues about what changes are needed.

My secondary hypothesis privileges one particular set of causal relationships over many possible others. This starting model is informed by the literature about kidney cancer being a disease of energy metabolism [15], [16] and the expert opinion of my mentors about a patient's journey through the health system. I will check my results with sensitivity analysis, varying the order in which variables are added to the IPTW.

I will then perform exploratory analysis to update the causal paths for use in the confirmatory analyses of future studies. In addition what I described above, I will generate a pairwise correlation matrix of all individual data elements to eliminate non-significant relationships, shorten the list of candidate predictors, and find errors in my prior assumptions about which data elements are closely related. For example, some variables in the "Care" group are mostly influenced by provider behavior while others by the patient... if a strong clustering structure is observed, this will provide an empirical basis to separate them into two or more groups. I will also run a backwards elimination process (using Akaike information criterion) [31] to find an optimal set of main effects and interactions for predicting metastatic progression. Both exploratory analyses will be repeated at MGH and conserved features will be identified as being robust against regional differences.

Milestones/Deliverables*	Fa 2018	Sp 2019	Fa 2019	Sp 2020
Complete DataFinisher	X			
Submit Aim-1 poster to AMIA Summit	X			
Submit Aim-1 manuscript to JAMIA or JSS		X		
Pre-register research plan on Center for Open Science				
Extract and prepare San Antonio data for analysis		X		
Data analysis, San Antonio		X		
Submit Aim-2 poster to AACR, ASCP or ASCO		X		
Submit Aim-2 manuscript to Cancer Research			X	
Externship and deployment of DataFinisher at MGH			X	
Extract and prepare MGH data for analysis, do data analysis			X	
Submit Aim-3 poster to AIMA Summit			X	
Submit Aim-3 manuscript to JAMIA				X
Prepare Federal K-grant	X	X	X	X
Submit Federal K-grant			X	X

* I anticipate additional publications from this project, as well as from pre-KL2 work that I am about to submit.

Table 3 timeline for research plan and main publications. Grey 'X' represent resubmission, if necessary.

TRAINING IN THE RESPONSIBLE CONDUCT OF RESEARCH

Formal Training

I have regularly taken UT Health online training in Conflict of Interest, HIPAA Privacy Training Level 2, and General Compliance Awareness Training. I have completed CITI training in Human Research/Biomedical Research in 2015 with a refresher course in 2018. As a KL2 scholar I will continue my training by taking TSCI 5070, Responsible Conduct of Research; TSCI 6102, Practicum in IRB Procedures; and TSCI 6103, Selected Topics in Advanced Research Ethics. TSCI 5070 is a two-credit interdisciplinary course at the end of which students will be able to: (1) delineate a history of hallmark abuses of humans enrolled in clinical research, (2) describe the evolution of national and international codes and regulations guiding inclusion of human subjects in clinical investigations, (3) list the elements of informed consent and describe procedures and precautions for enrolling special populations into clinical investigation, (4) write a consent form in understandable language, (5) recognize different forms of scientific misconduct, (6) describe the role and processes of a peer review board to judge violations in research ethics, (7) develop strategies for self-assessment and validation of scientific objectivity in one's own research, and (8) recognize the ethical responsibilities and consequences of whistle blowing. TSCI 6102 is a one-credit in-depth introduction to IRB taught through a combination of readings, monthly attendance at selected IRB meetings, and discussions with faculty. I will also participate in Spotlight on Research Integrity, a monthly workshop covering current topics in responsible conduct of research. TSCI 6103 is a one-credit course where students prepare literature reviews on a topic in research ethics.

Informal Training

I will receive training in the responsible conduct of research through weekly meeting with my mentors each of which has extensive training and experience in the responsible conduct of research. The bi-weekly workshop Grant Writing for New Investigators will also provide me with training in all RCR topics as they pertain to grant application guidelines and development. I will also take programmatic RCR training offered by the KL2 program in the form of Scholars Optimizing Achievement in Research (1 hour, monthly), the Scholars Preparing Aims for R and K awards Peer-Mentoring Seminar (1.5 hours, bi-weekly), as well as monthly one-on-one meetings with the KL2 directors.

INSTITUTIONAL ENVIRONMENT

Institute for Health Promotion Research (IHPR)

IHPR investigates the causes of and solutions to the unequal impact of cancer, chronic disease and obesity among Latinos in San Antonio, South Texas and the nation. Areas of Expertise include: Latino health disparity research/training; Health promotion/communication; Cancer control research from primary prevention to survivorship; Healthy lifestyle promotion; Prevention of tobacco use, obesity, diabetes.

Masters of Science in Clinical Investigation Program

Conducted through the Graduate School of Biomedical Sciences at the UT Health Science Center, the MS-CITS degree program offers coursework and mentored research for degree-seeking and non-degree seeking students. The courses from this curriculum that I will take include: TSCI 6015 Topics in Cancer Prevention; TSCI 6065 Health Services Research; TSCI 6102 Practicum in IRB Procedures; TSCI 5070 Responsible Conduct of Research; and TSCI 6001 Introduction to Translational Science

Long School of Medicine

The Long School of Medicine, ranked one of the top three in the United States for Hispanic students by Hispanic Business magazine, has a strong and supportive faculty and numerous opportunities for building clinical and research skills. Our medical research institutes and nationally recognized cancer treatment programs combine education and research to provide some of the country's most innovative care. I will be taking the following courses here: MEDI 4153 Informatics and Advanced Evidence-Based Medicine; INTD 4104 Improving Patient Outcomes; INTD 5030 Introduction To Patient Care; ELEC 5004 Surgical Oncology Service

UHS

UHS is a nationally recognized academic medical center, network of outpatient clinics strategically located in at-risk communities, a Level I trauma center, and operates a Federally Qualified Health Center (CommuniCare). UHS is the largest Safety Net Hospital in South Texas and treats a predominately Hispanic population. Many patients are seen in UT Health clinics before and after surgical procedures at UHS.

UT Health Faculty Practice

UT Health Physicians (formerly called UT Medicine) features more than 700 physicians and health care providers offering advanced services and technologies for you and your family's needs with a patient population of over 480,000 managed via the Epic EMR system.

Clinical Informatics Research Division

The IIMS Informatics Core (B) provides EHR and billing data from our faculty practice plan (UT Health) and from University Hospital System (UHS) linked into one coherent dataset in an i2b2 data warehouse managed by the Clinical Informatics Research Division (CIRD). UT Health has been available since 2015 and UHS scheduled for production release to local researchers for June of this year, bringing the total number of unique patients from 480,000 to 1.2 million. The CIRD i2b2 data warehouse also contains mortality data from the Social Security Death Master File (SSDMF), detailed cancer reports from the North American Association of Central Cancer Registries (NAACCR) and as of June 2018 will also contain income and educational attainment data from the 2016 American Community Survey 5-year Summary linked to patients based on their addresses.

Massachusetts General Hospital

MGH: MGH, located in Boston, MA and operated by Partners Healthcare, is the teaching partner of Harvard Medical School. MGH has 1.5 million ambulatory visits and 51,000 inpatient stays per year. According to Dr. Murphy, there are a total of 18,500 kidney cancer cases in the MGH EMR. 19% of Boston's population is Hispanic. Of Hispanic Bostonians the most common ancestries are Puerto Rican (43%) and Dominican (17%), with Mexican descent comprising less than 7%.

STATEMENT OF HOW THE RESEARCH IS TRANSLATIONAL

This research falls into the T3 and T4 stages of translational research: translation to practice and translation to population health. This is translation to practice because Aims 2 and 3 involve quality improvement-- the causal network in Figure 2 of the Research Strategy funnels down into four specific and intervenable aspects of a patient's interaction with the health system at and after presentation, all of which have been selected on the basis of construct validity. Aims 1 and 3 are about dissemination of non-commercial software that will make electronic health records more accessible to researchers, as well as a demonstration project of this software.

The overall project is outcomes research that reaches all the way from the clinic into society at large where some of the underlying mediators of disparity may reside, using detailed socioeconomic data from the census linked to electronic health records.

PROTECTION OF HUMAN SUBJECTS

The proposed study merges identified data with additional sources as proxy measures for social risk factors. This research falls under the non-exempt Human Subjects Research category.

1. Risks to Human Subjects

1a. Human Subjects Involvement: Characteristics and Design: Inclusion criteria for this study are adult patients with a diagnosis of kidney cancer and who underwent nephrectomy at UT Health/UHS, or at Massachusetts General Hospital. Both sites have i2b2 data warehouses, where visit data (such as labs, diagnoses, procedures, medications, and vital signs) are stored separately from any identifying information. This enables research use of de-identified electronic health records.

There will be at least 900 subjects in Aim 1/2 of this study and at least as many additional subjects in Aim 3, the Massachusetts General Hospital arm of the study. Potential risks are limited to inadvertent release of Protected Health Information (PHI). Studies will be implemented only after final Institutional Review Board (IRB) and other appropriate regulatory committee review and approval using the SMART IRB platform.

1b. Study Procedures, Materials and Potential Risks: All analysis will be done on de-identified data. The software development activities do not require access to identified data. Actual handling of identified data will be limited to a) chart reviews to validate accuracy and b) if necessary, linkage of i2b2 patient records to sources such as US Census data or the local NAACCR cancer registry. Identifying patient data will never be transmitted between institutions and a key strategic purpose of this project is to demonstrate that a sophisticated multi-site retrospective study can be carried out without any need to transmit identifiers in the first place. Both institutions are CTSA sites with rigorous data security systems and protocols in place.

2. Adequacy of Protection Against Risks

2a. Informed Consent and Assent: This study will not actively recruit patients or use informed consent. This will be a data-only study with waived consent.

2b. Protections Against Risk: Both i2b2 teams already have established procedures in place to store identifiers on secure servers behind the firewalls of their respective with access granted only to members of the site informatics teams. Access to identifiers pursuant to this project would be limited to quality control/validation and linkage to supplementary data sources.

2c. Vulnerable Subjects: Patients in i2b2 could be prisoners, mentally disabled, have dementia or other conditions that can be classified as vulnerable but they are not being explicitly selected for these qualities. This is a data-only study; access to vulnerable subjects records will be performed with the same care as other study subjects.

3. Potential Benefits of Research to Human Subjects and Others

The researcher will not directly interact with any of the patients in the study, only patient data. Patients in the study will not directly benefit other than the knowledge generated from the proposed studies.

4. Importance of Knowledge to be Gained

While mortality rates for most cancers are declining, those for kidney cancer are increasing, and Hispanic patients are disproportionately impacted. This study will measure how much various candidate mediators of this disparity actually contribute to progression to metastasis among patients who were diagnosed with kidney cancer. This information can help improve the design and targeting of future prevention, screening, physician training, and health system navigation efforts. It will also help understand (in Aim 3) whether this disparity primarily affects Hispanic patients of Mexican heritage, or also Hispanic patients of Caribbean heritage. Finally, it will provide a more focused set of clinical covariates to use in future clinical trials and high-throughput analysis of biospecimens.

BIBLIOGRAPHY AND REFERENCES CITED

- [1] A. F. Bokov, D. Ko, and A. Richardson, "The Effect Of Gonadectomy And Estradiol On Sensitivity To Oxidative Stress," *Endocr. Res.*, vol. 34, no. 1–2, pp. 43–58, 2009.
- [2] A. Bokov, H. Y. Liang, W. B. Qi, H. VanRemmen, W. Ward, and A. Richardson, "Diquat induced oxidative modifications track or precede cell death in mouse liver," *FREE Radic. Biol. Med.*, vol. 37, no. Suppl. 1, p. S105, 2004.
- [3] A. F. Bokov *et al.*, "Does Reduced IGF-1R Signaling in Igf1r^{+/−} Mice Alter Aging?," *PLoS ONE*, vol. 6, no. 11, p. e26891, Nov. 2011.
- [4] J. Bokov, Olin, Bos, Kittrell, Tirado-Ramos, "Using Prevalence Patterns to Discover Un-Mapped Flowsheet Data in an Electronic Health Record Data Warehouseigital," presented at the Computer-Based Medical Systems, Thessaloniki, Greece, 2017.
- [5] A. F. Bokov *et al.*, "Exhaustively Characterizing a Patient Cohort by Prevalence of EMR Facts: a Generalized, Vendor-Agnostic Method for Quality Control and Research," in *AMIA Annual Symposium Proceedings*, 2017, vol. 2017.
- [6] A. F. Bokov, L. S. Manuel, A. Tirado-Ramos, J. A. Gelfond, and S. D. Pletcher, "Biologically relevant simulations for validating risk models under small-sample conditions," in *IEEE Symposium on Computers and Communication*, Heraklion, Greece, 2017, pp. 290–295.
- [7] P. S. Pinheiro *et al.*, "High cancer mortality for US-born Latinos: evidence from California and Texas," *BMC Cancer*, vol. 17, no. 1, Dec. 2017.
- [8] P. C. Austin, "The performance of different propensity score methods for estimating marginal hazard ratios," *Stat. Med.*, vol. 32, no. 16, pp. 2837–2849, Jul. 2013.
- [9] S. Murphy *et al.*, "Instrumenting the health care enterprise for discovery research in the genomic era," *Genome Res.*, vol. 19, no. 9, pp. 1675–1681, 2009.
- [10] A. Bokov, L. Manuel, C. Cheng, A. Bos, and A. Tirado-Ramos, "Denormalize and Delimit: How not to Make Data Extraction for Analysis More Complex than Necessary," *Procedia Comput. Sci.*, vol. 80, pp. 1033–1041, 2016.
- [11] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015: Cancer Statistics, 2015," *CA. Cancer J. Clin.*, vol. 65, no. 1, pp. 5–29, Jan. 2015.
- [12] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016: Cancer Statistics, 2016," *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, Jan. 2016.
- [13] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA. Cancer J. Clin.*, vol. 67, no. 1, pp. 7–30, Jan. 2017.
- [14] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018: Cancer Statistics, 2018," *CA. Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.
- [15] L. Yu, X. Chen, X. Sun, L. Wang, and S. Chen, "The Glycolytic Switch in Tumors: How Many Players Are Involved?," *J. Cancer*, vol. 8, no. 17, pp. 3430–3440, 2017.
- [16] K. Shanmugasundaram and K. Block, "Renal Carcinogenesis, Tumor Heterogeneity, and Reactive Oxygen Species: Tactics Evolved," *Antioxid. Redox Signal.*, vol. 25, no. 12, pp. 685–701, Oct. 2016.
- [17] A. Q. Haddad and V. Margulis, "Tumour and patient factors in renal cell carcinoma—towards personalized therapy," *Nat. Rev. Urol.*, vol. 12, no. 5, pp. 253–262, May 2015.
- [18] G. M. Weber *et al.*, "The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories," *J. Am. Med. Inform. Assoc.*, vol. 16, no. 5, pp. 624–630, Sep. 2009.
- [19] H. Wickham, "Tidy Data," *J. Stat. Softw.*, vol. 59, no. 10, 2014.
- [20] A. F. Beck, B. Huang, K. A. Auger, P. H. Ryan, C. Chen, and R. S. Kahn, "Explaining Racial Disparities in Child Asthma Readmission Using a Causal Inference Approach," *JAMA Pediatr.*, vol. 170, no. 7, p. 695, Jul. 2016.
- [21] University of Virginia & Center for Open Science and B. A. Nosek, "Opening Science," in *Open: The Philosophy and Practices that are Revolutionizing Education and Science*, Kwantlen Polytechnic University, CA, R. S. Jhangiani, R. Biswas-Diener, and Noba Project, Eds. Ubiquity Press, 2017, pp. 89–99.

- [22] J. C. Kirby *et al.*, "PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability," *J. Am. Med. Inform. Assoc.*, vol. 23, no. 6, pp. 1046–1052, Nov. 2016.
- [23] K. Rother, "Documentation," in *Pro Python Best Practices*, Berkeley, CA: Apress, 2017, pp. 245–259.
- [24] B. Adagarla *et al.*, "SEINE: Methods for Electronic Data Capture and Integrated Data Repository Synthesis with Patient Registry Use Cases," 2015.
- [25] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, "A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation," *J. Chronic Dis.*, vol. 40, no. 5, pp. 373–383, Jan. 1987.
- [26] H. Quan *et al.*, "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data," *Med. Care*, vol. 43, no. 11, pp. 1130–1139, Nov. 2005.
- [27] P. K. Andersen and R. D. Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study," *Ann. Stat.*, vol. 10, no. 4, pp. 1100–1120, Dec. 1982.
- [28] D. A. Schoenfeld, "Sample-Size Formula for the Proportional-Hazards Regression Model," *Biometrics*, vol. 39, no. 2, p. 499, Jun. 1983.
- [29] S. G. Nicholls *et al.*, "The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement: Methods for Arriving at Consensus and Developing Reporting Guidelines," *PLOS ONE*, vol. 10, no. 5, p. e0125620, May 2015.
- [30] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017.
- [31] W. N. Venables, *Modern applied statistics with S*, 4th ed. New York: Springer, 2002.

Diversity Questionnaire (Required by NIH)

What is your sex/gender? Female Male

What is your ethnic background?

Hispanic or Latino? Yes No

Not Hispanic or Latino? Yes No

Unknown or Not Reported Yes No

What is your racial background?

American Indian or Alaska Native

Native Hawaiian or other Pacific Islander

Asian

Black or African American

White (non-Hispanic)

More Than One Race

Unknown or Not Reported

Do you have a disability? Yes No Do Not Wish to Provide

Are you from a disadvantaged background?

Yes No Do Not Wish to Provide

Individuals from disadvantaged backgrounds are defined as:

1. Individuals who come from a family with an annual income below established low-income thresholds. These thresholds are based on family size, published by the U.S. Bureau of the Census; adjusted annually for changes in the Consumer Price Index; and adjusted by the Secretary for use in all health professions programs. The Secretary periodically publishes these income levels at <http://aspe.hhs.gov/poverty/index.shtml>. For individuals from low-income backgrounds, the institution must be able to demonstrate that such candidates (a) have qualified for Federal disadvantaged assistance; or (b) have received any of the following student loans: Health Professional Student Loans (HPSL), Loans for Disadvantaged Student Program; or (c) have received scholarships from the U.S. Department of Health and Human Services under the Scholarship for Individuals with Exceptional Financial Need.

2. Individuals who come from a social, cultural, or educational environment such as that found in certain rural or inner-city environments that have demonstrably and recently directly inhibited the individual from obtaining the knowledge, skills, and abilities necessary to develop and participate in a research career. Recruitment and retention plans related to a disadvantaged background are most applicable to high school and perhaps undergraduate candidates, but would be more difficult to justify for individuals beyond that level of achievement.



Amelie G. Ramirez, MPH, DrPH

Professor and Chair *ad Interim*

Department of Epidemiology & Biostatistics

May 6, 2018

Dear colleagues,

On behalf of the Department of Epidemiology and Biostatistics (DEB) at UT Health San Antonio, I am very pleased to support the application of Dr. Alex Bokov for the Mentored Research and Career Development (KL2) Program of the Institute for the Integration of Medicine and Science. Our institution and our department are making a strong push to rapidly grow the number of faculty capable of leveraging electronic health records in their research, and Dr. Bokov's career development and research plan is directly responsive to this need.

Dr. Bokov has made tremendous contributions in creating the Clinical Informatics Research Division (CIRD) at DEB. But because of the rare and in-demand skill-set required for this domain, it has been difficult to find experienced informatics mentors for Dr. Bokov, especially after the recent departure of CIRD leader Dr. Alfredo Tirado-Ramos from our institution. To fill this gap, Dr. Bokov used his networking skills to "import" needed specialized mentorship from Harvard School of Medicine by engaging Dr. Shawn Murphy as a co-mentor. Dr. Murphy co-wrote the i2b2 data warehouse on which the CTSA Accrual to Clinical Trials network relies. Dr. Murphy will supervise Dr. Bokov's continued work on an informatics app intended to speed up and simplify the process of obtaining retrospective data from i2b2, and will organize an externship at Massachusetts General Hospital.

Equally important is the thorough plan Dr. Bokov and this mentoring team developed for obtaining training in clinical documentation and decision-making. Dr. Bokov will be taking several courses from the medical student curriculum and shadowing co-mentor and collaborator Dr. Ronald Rodriguez in Urology. Dr Bokov has chosen an important and under-researched health disparity: the high prevalence of kidney cancer in South-Central Texas, especially among Hispanic patients. During his KL2 training Dr. Bokov will use his informatics software to analyze medical histories from a large sample of kidney cancer patients in order to identify possible causes of this disparity then replicate this analysis at Massachusetts General Hospital (to see, among other things, the extent to which his findings generalize to Hispanic patients of Caribbean heritage).

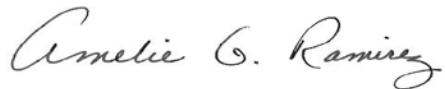
The DEB is fully committed to supporting Dr. Bokov's transition to an independent investigator. Even before Dr. Bokov started preparing his proposal, Dr. Joel Michalek and I started helping him prepare his packet for consideration for promotion to Research Assistant Professor. I guarantee 75% protected time for Dr. Bokov to pursue research and career development plus a third year of commitment of 75% protected time if he has not yet been awarded an individual K or R grant by the time the KL2 funding cycle has ended.

Dr. Bokov will also be given in-kind support by way of staff effort on a as needed to bases with software development and data warehouse enhancements for his KL2 research. Furthermore, Dr. Bokov will continue to receive the usual departmental support provided to faculty including office space, a computer, software, grant preparation, and administrative assistance.

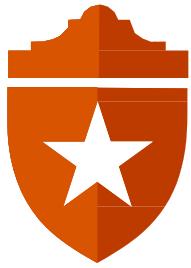
Again, I offer my highest support for Dr. Alex Bokov for the Mentored Research and Career Development (KL2) Program of the Institute for the Integration of Medicine and Science.

If I can provide any further information regarding my support, please do not hesitate to contact me at 210-562-6500 or ramirezag@uthscsa.edu.

Sincerely,

A handwritten signature in black ink that reads "Amelie G. Ramirez". The signature is fluid and cursive, with "Amelie" and "G." on the first line and "Ramirez" on the second line.

Amelie G. Ramirez, DrPH
Professor and Chair *ad Interim*
Department of Epidemiology & Biostatistics
Director, Institute for Health Promotion Research



UT Health San Antonio

Urology

Department of Urology

May 17, 2018

Administration

Ronald R. Rodriguez, MD, PhD
Chair
Stephen R. Kraus, MD, FACS
Vice Chair

Urologic Oncology

Robert S. Svatek, MD
Division Chief
Joseph W. Basler, MD, PhD
Javier Hernandez, MD
Dharam Kaushik, MD
Michael A. Liss, MD
Ronald Rodriguez, MD, PhD

Neurourology/ Urodynamic Voiding Dysfunction

Stephen R. Kraus, MD
Division Chief

Endourology, Laparoscopy Stone Disease

Timothy Tseng, MD

General Urology and Reconstructive

Thomas A. Rozanski, MD

Research

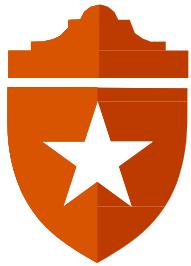
Robin J. Leach, PhD
Division Chief
Donna P. Ankerst, PhD
Dean J. Bacich, PhD
Rita Ghosh, PhD
Denise O'Keefe, PhD
A. Pratap Kumar, PhD
Susan S. Padalecki, PhD
Teresa Johnson-Pais, PhD

KL2 Application, CTSA Program UT Health San Antonio

Dear Dr. Dougherty, Dr. Tsevat, and Reviewers:

As Professor and Chair of the Department of Urology, I am writing to offer my strongest possible personal recommendation and support for Dr. Alex Bokov in his application for the Mentored Research Career Development (KL2) Program in Clinical and Translational Science. He has developed an innovative yet thorough proposal that leverages electronic medical records from the i2b2 data warehouse to better understand risk factors for progression to metastasis in kidney cancer patients. I reviewed and advised him on his research and career development plans and I look forward to being his mentor. It is not typical for a KL2 scholar to propose a multi-site study but this particular scholar can make it happen because of his thorough understanding of the technology involved, his relationship with Dr. Murphy at Harvard Medical School, his careful attention to project scope, and his experience as a post-doc and junior faculty working with PCORI and more recently ACT. I am particularly impressed by how Dr. Bokov's research plan is designed to produce publishable results at regular intervals over the course of the KL2 funding period and to yield useful results whichever way any individual sub-hypothesis turns out. Along the way, he takes every opportunity to generate preliminary data and evidence of feasibility for future grant proposals. Dr. Bokov already has access to Epic and Sunrise data through the CIRD repository protocol, which will cover the software development work he proposes in Aim-1. I am PI on several research protocols to which I can add him if necessary for Aim-2 and will sponsor him to be credentialed to the Sunrise EMR front-end if necessary. For Aim-3, he will likely need to use the SMART IRB process but should already be familiar with that from his PCORI work. All proposed work is low-risk and not likely to encounter regulatory obstacles.

When I started practicing at UT Health, I was surprised by how much more common advanced stage renal cancers are here compared to Johns Hopkins, and they seemed to be especially common in Latinos. I started collaborating with Dr. Ramirez and Dr. Michalek from the Department of Epidemiology and Biostatistics (DEB) to understand why. They introduced me to Dr. Bokov as the resident expert on getting data from an EMR system into a format that other researchers can use. They were right-- I have never met somebody so passionately committed to building collaboration and understanding between researchers in the clinical and quantitative fields.



UT Health
San Antonio

To become even more effective as a translational informatician Dr. Bokov needs, and is actively seeking, front-line hands-on training in every aspect of clinical practice that, in his words, "a Ph.D. is legally allowed to do". I helped him identify courses he can take alongside medical students where he will learn to collect patient medical histories, perform chart reviews, understand standards of care in cancer, and use EMR systems under real-world clinical conditions. Part of his training will include shadowing surgeons in my department. He has already started regular attendance at our GU tumor board. Over the course of the KL2 I will meet with him weekly, and he will continue to attend GU tumor board, GU research meetings, the Population Science and Prevention Program at CTRC, and Urology Grand Rounds.

I have mentored dozens of residents and fellows. From time to time an analytically-inclined one of them branches out into computer programming and data science... but this is the first time I see a young translational scientist already trained in data analysis go the other direction, to learn more about my profession. That is something we should support. Dr. Bokov is a valuable asset to our institution and I want do everything possible to promote and retain this talented early-career translational scientist so that he can achieve his full potential and help establish a reputation of excellence in health informatics for our campus.

Sincerely,

A handwritten signature in black ink, appearing to read "Ron Rodriguez".

Ron Rodriguez, MD, PhD
Professor and Chairman, Department of Urology
Henry B. and Edna Smith Dielmann Memorial Professor of Urologic Science,
UT Health San Antonio Joseph R and Teresa L Long School of Medicine
7703 Floyd Curl Drive, Mail Code 7845
San Antonio, TX 78229-3900
Office: 210-567-5643
Fax: 210-567-6868



Department of Epidemiology & Biostatistics

02/10/18

Dear IIMS KL2 selection committee:

It is my extreme privilege and pleasure to write the strongest possible letter of support for Dr. Alex Bokov regarding the Mentored Research Career Development (KL2) Program In Clinical and Translational Science and to serve as his mentor.

I am a Fellow of the American Statistical Association and tenured Professor in the Department of Epidemiology and Biostatistics, Dr. Bokov's home department. I am a member of the Protocol Review Committee Co-director of the Biostatistics Shared Resource at the Mays Cancer Center, and I have an extensive track record of peer-reviewed publications in oncology. Together with Dr. Shawn Murphy, Dr. Amelie Ramirez, and Dr. Ronald Rodriguez I will mentor Dr. Bokov.

I have known Dr. Bokov since at least 2013 when he was still a post-doc. One of our first interactions was when he showed me a web-app he wrote that would allow users to perform and visualize multi-variable regression without needing to know R or SAS. This was before he was hired at CIRD and his skills have progressed even further by now. I am absolutely certain that he can complete the software development aims he sets out in his Research Plan. In my opinion, Dr. Bokov will benefit from external mentorship from Dr. Murphy by learning advanced topics in interoperability and dissemination for his software as well as guidance on career advancement opportunities specific to medical informatics. As for the statistical approach in his Aim-2, to be replicated at Massachusetts General Hospital in Aim-3, I advised him on its design and can vouch for its soundness and validity. A strength of his research plan is the thought he put into insuring that he will obtain interpretable results to inform his future research regardless of whether or not his starting hypotheses are confirmed.

My roles will include guidance on advanced statistical topics, grantsmanship, and navigating the Promotion and Tenure (P&T) process at our institution. My P&T mentoring of Dr. Bokov's as well as my collaborations with Dr. Rodriguez and Dr. Ramirez position me for these roles. My office is adjacent to Dr. Bokov's, we already meet almost daily, and intend to continue doing so. In July, we will both enroll in the Entering Mentoring program. Among the funding opportunities Dr. Bokov's is targeting for this next grant is the National Library of Medicine K01 career development award in Biomedical Informatics and Data Science (PAR-16-204). He will start working on his proposal early, updating it with new publications and results as they become available. As someone with a long track record of successful grantsmanship I have valuable experience to share with him.

Dr. Bokov's research is complementary to my collaboration with Dr. Rodriguez and Dr. Ramirez but distinct from it, and not only in the area of writing informatics software. My approach is from the macro level-- clinical trials and large public datasets such as SEER and the National Inpatient Sample. Dr. Bokov's approach is from the health-system level-- the Epic and Sunrise EMR systems via i2b2. Together we can learn how representative national and regional trends are of the actual patient population we serve. I see many opportunities for co-authorship in addition to the first-author publication pipeline Dr. Bokov has planned out in the table near the end of his Research Plan.



Department of Epidemiology & Biostatistics

We all hope that in a few years informatics infrastructure will mature to the point that any researcher will be able to carry out the sort of project that Dr. Bokov is proposing. But today, the only junior faculty on our campus who can credibly do such a study is Dr. Bokov and this is exactly what he needs to launch a successful career as an independent translational scientist in the rapidly growing field of medical informatics. I hope we use this opportunity to give him the protected time, structured mentorship, and clinical training that he needs.

Sincerely,

A handwritten signature in black ink that reads "Joel E. Michalek".

Joel E. Michalek, PhD FASA
Professor



Academic Programs
399 Revolution Drive
Suite 725
Somerville, MA 02145
Tel: 857-282-3769

Shawn N. Murphy, MD, Ph.D.
*Chief Research Information Officer
Partners HealthCare
Professor of Neurology
Harvard Medical School*

May 17, 2018

Dear Colleagues at UT Health:

I am writing in enthusiastic support of Dr. Alex Bokov in his application for the Mentored Research Career Development (KL2) Program In Clinical and Translational Science and to affirm my commitment be his mentor.

In 1999, at Massachusetts General Hospital, I developed a data warehouse system called Research Patient Data Registry (RPDR). In 2004 I got an NIH grant to disseminate this system to other institutions. Today this system is known as Integrating Informatics from Bench to Bedside (i2b2) and it is deployed at over 120 hospitals all over the world, including all sites in the ACT network.

I first met Dr. Bokov online in 2013 when he posted a question on the "i2b2 Install Help" Google group that I run. The next year we met in person at the i2b2 Academic Users' Group annual meeting and again several more times, most recently at the 2017 American Medical Informatics Symposium in Washington DC where we were both on an oral presentation panel about Enhanced Cohort Identification and Retrieval. Dr. Bokov was presenting one of his two i2b2-leveraging software projects, and I am pleased to see that in his KL2 application he is seeking protected time to complete his other one-- DataFinisher. He and I have been talking about this app for a while now-- it automates the process of joining relational data from an i2b2 datamart into a single output table. This could give data warehouse users more flexibility and save informatics teams a significant amount of honest broker effort.

As one of i2b2's authors and a Co-I of the ACT network it is my opinion that Dr. Bokov's DataFinisher software will be useful to translational researchers, is novel yet feasible, and strongly aligned with the goals of CTSA both in its functionality and in the transparent, open-source development approach Dr. Bokov is taking.

Dr. Bokov clearly has a natural talent for data analysis and software development. He can become a highly successful and productive medical informatician but he needs structured mentorship from a more experienced colleague. I am very happy to provide this mentorship, both through his KL2 training and through AMIA.

Informaticians are accustomed to remote collaboration-- Dr. Bokov and I use chat rooms, teleconferences, and forums all the time in our work with ACT and PCORNet. I can assure you that our geographic distance will not be an obstacle to a strong mentoring relationship. I anticipate us meeting on a biweekly basis and setting aside time to catch up in person during AMIA and i2b2 events. Furthermore, at some point during the internship I will invite Dr. Bokov to do a 4-8 week externship with me and my team at Massachusetts General Hospital. During this time Dr. Bokov will deploy the DataFinisher software that he will have completed under my direction and

replicate the kidney cancer study he will have done in San Antonio under the direction of his local mentors. We have data elements in our i2b2 that are equivalent to the income, education, and tumor stage variables Dr. Bokov describes in his Aim-2. We have a well-organized data governance process in place and I do not anticipate any problems getting Dr. Bokov the necessary approvals.

If informatics is an important part of translational science, then the career development of informaticians should be important to CTSA sites, and you are fortunate to have on your campus a young informatician of exceptional potential.

Sincerely,

A handwritten signature in black ink, appearing to read "Shawn N. Murphy".

Shawn N. Murphy, MD, Ph.D.



UT Health
San Antonio

Institute for Integration
of Medicine & Science

Mentorship Agreement

[1] **Goals** (what you hope to achieve as a result of this relationship; e.g., gain perspective relative to skills necessary for success in academia, explore new career opportunities/alternatives, obtain knowledge of organizational culture, networking, leadership skill development, etc.):

- Gain a detailed understanding of current best practice in preventing and treating kidney cancer and what the important evidence gaps are in these fields. [Rodriguez, Ramirez, Michalek]
- Gain better understanding of the behavioral and social determinant impact on kidney cancer.
- Learn advanced concepts in writing open-source informatics software and disseminating it to researchers. [Murphy]
- Get structured mentorship in continuing to grow my professional skills as a translational scientist: leadership, grantsmanship, scientific communication, and organization. [All]
- Build, strengthen, and extend collaborative relationships with other translational scientists including clinicians, population health researchers, and other informaticians. [All]

[2] **Steps to achieving goals** as stated above (e.g., meeting regularly, manuscripts/grants, collaborating on research projects, steps to achieving independence, etc.): mentor meetings, coursework and other activities as described in the table in the Career Development Plan, collaboration on manuscripts and presentations, notifying mentee of relevant networking and funding opportunities.

[3] **Meeting frequency:** One hour per week on average. Mentee will come prepared with an agenda to discuss areas of importance.

[4] **Confidentiality:** Any sensitive issues that we discuss will be held in the strictest of confidence.

[5] Plan for **evaluating relationship effectiveness** (e.g., bi-annual review of mentorship meeting minutes, goals, and outcomes/accomplishments): de-brief after KL2 director's meeting, reviewing yearly faculty career development plan. Mentee will regularly post online in a publicly readable form the hours he worked each day and what they were spent doing. Mentee's progress can also be tracked via GitHub commit logs.

[6] **Relationship termination clause:** In the event that either party finds the mentoring relationship unproductive and requests that it be terminated, we agree to honor that individual's decision without question or blame.

[7] **Duration:** This mentorship relationship will continue as long as both parties feel comfortable with its productivity or until the conclusion of the KL2 funding period or the optional 1-year DEB-covered grace period if needed for Dr. Bokov to secure a Federal K or R grant.

Mentor's Signature

Anneke G. Ramirez

Mentee's Signature

Bokov

Date

5/18/2018



Mentorship Agreement

[1] **Goals** (what you hope to achieve as a result of this relationship; e.g., gain perspective relative to skills necessary for success in academia, explore new career opportunities/alternatives, obtain knowledge of organizational culture, networking, leadership skill development, etc.):

- Gain a detailed understanding of current best practice in preventing and treating kidney cancer and what the important evidence gaps are in these fields. [Rodriguez, Ramirez, Michalek]
- Learn advanced concepts in writing open-source informatics software and disseminating it to researchers. [Murphy]
- Get structured mentorship in continuing to grow my professional skills as a translational scientist: leadership, grantsmanship, scientific communication, and organization. [All]
- Build, strengthen, and extend collaborative relationships with other translational scientists including clinicians, population health researchers, and other informaticians. [All]

[2] **Steps to achieving goals** as stated above (e.g., meeting regularly, manuscripts/grants, collaborating on research projects, steps to achieving independence, etc.): mentor meetings, coursework and other activities as described in the table in the Career Development Plan, collaboration on manuscripts and presentations, notifying mentee of relevant networking and funding opportunities.

[3] **Meeting frequency:** one hour per week on average, in private or group meetings discussing data, manuscript revisions, strategies and career development.

[4] **Confidentiality:** Nothing at this time.

[5] Plan for **evaluating relationship effectiveness** (e.g., bi-annual review of mentorship meeting minutes, goals, and outcomes/accomplishments): de-brief after KL2 director's meeting, reviewing yearly faculty career development plan. Mentee will regularly post online in a publicly readable form the hours he worked each day and what they were spent doing. Mentee's progress can also be tracked via GitHub commit logs.

[6] **Relationship termination clause:** In the event that either party finds the mentoring relationship unproductive and requests that it be terminated, we agree to honor that individual's decision without question or blame.

[7] **Duration:** This mentorship relationship will continue as long as both parties feel comfortable with its productivity or until the conclusion of the KL2 funding period or the optional 1-year DEB-covered grace period if needed for Dr. Bokov to secure a Federal K or R grant.

Mentor's Signature



Mentee's Signature



Date

5/16/18



Mentorship Agreement

[1] **Goals** (what you hope to achieve as a result of this relationship; e.g., gain perspective relative to skills necessary for success in academia, explore new career opportunities/alternatives, obtain knowledge of organizational culture, networking, leadership skill development, etc.):

- Gain a detailed understanding of current best practice in preventing and treating kidney cancer and what the important evidence gaps are in these fields. [Rodriguez, Ramirez, Michalek]
- Learn advanced concepts in writing open-source informatics software and disseminating it to researchers. [Murphy]
- Get structured mentorship in continuing to grow my professional skills as a translational scientist: leadership, grantsmanship, scientific communication, and organization. [All]
- Build, strengthen, and extend collaborative relationships with other translational scientists including clinicians, population health researchers, and other informaticians. [All]

[2] **Steps to achieving goals** as stated above (e.g., meeting regularly, manuscripts/grants, collaborating on research projects, steps to achieving independence, etc.): mentor meetings, coursework and other activities as described in the table in the Career Development Plan, collaboration on manuscripts and presentations, notifying mentee of relevant networking and funding opportunities.

[3] **Meeting frequency** : one hour per week on average.

[4] **Confidentiality:** Any sensitive issues that we discuss will be held in the strictest of confidence. No issues pertaining to this research are sensitive.

[5] Plan for **evaluating relationship effectiveness** (e.g., bi-annual review of mentorship meeting minutes, goals, and outcomes/accomplishments): de-brief after KL2 director's meeting, reviewing yearly faculty career development plan. Mentee will regularly post online in a publicly readable form the hours he worked each day and what they were spent doing. Mentee's progress can also be tracked via GitHub commit logs.

[6] **Relationship termination clause:** In the event that either party finds the mentoring relationship unproductive and requests that it be terminated, we agree to honor that individual's decision without question or blame.

[7] **Duration:** This mentorship relationship will continue as long as both parties feel comfortable with its productivity or until the conclusion of the KL2 funding period or the optional 1-year DEB-covered grace period if needed for Dr. Bokov to secure a Federal K or R grant.

Mentor's Signature

José E. Michalek

Mentee's Signature

Bokov

Date

17 May 2018



Mentorship Agreement

[1] **Goals** (what you hope to achieve as a result of this relationship; e.g., gain perspective relative to skills necessary for success in academia, explore new career opportunities/alternatives, obtain knowledge of organizational culture, networking, leadership skill development, etc.):

- Gain a detailed understanding of current best practice in preventing and treating kidney cancer and what the important evidence gaps are in these fields. [Rodriguez, Ramirez, Michalek]
- Learn advanced concepts in writing open-source informatics software and disseminating it to researchers. [Murphy]
- Get structured mentorship in continuing to grow my professional skills as a translational scientist: leadership, grantsmanship, scientific communication, and organization. [All]
- Build, strengthen, and extend collaborative relationships with other translational scientists including clinicians, population health researchers, and other informaticians. [All]

[2] **Steps to achieving goals** as stated above (e.g., meeting regularly, manuscripts/grants, collaborating on research projects, steps to achieving independence, etc.): mentor meetings, coursework and other activities as described in the table in the Career Development Plan, collaboration on manuscripts and presentations, notifying mentee of relevant networking and funding opportunities.

[3] **Meeting frequency** : one hour every two weeks on average.

[4] **Confidentiality:** Any sensitive issues that we discuss will be held in the strictest of confidence. Issues that are off limits for discussion include: Nothing at this time.

[5] Plan for **evaluating relationship effectiveness** (e.g., bi-annual review of mentorship meeting minutes, goals, and outcomes/accomplishments): de-brief after KL2 director's meeting, reviewing yearly faculty career development plan. Mentee will regularly post online in a publicly readable form the hours he worked each day and what they were spent doing. Mentee's progress can also be tracked via GitHub commit logs.

[6] **Relationship termination clause:** In the event that either party finds the mentoring relationship unproductive and requests that it be terminated, we agree to honor that individual's decision without question or blame.

[7] **Duration:** This mentorship relationship will continue as long as both parties feel comfortable with its productivity or until the conclusion of the KL2 funding period or the optional 1-year DEB-covered grace period if needed for Dr. Bokov to secure a Federal K or R grant.

Mentor's Signature

A handwritten signature in blue ink, appearing to read "Bokov".

Mentee's Signature

A handwritten signature in blue ink, appearing to read "Bokov".

Date

A handwritten date in blue ink, reading "17 May 2018".



Individual development plan

The purpose of this template is to assist you in documenting information relevant to your career trajectory including short and long-term goals, barriers, accomplishments, and plans for acquiring the skills necessary to achieve your career goals. This template is aligned with the Annual Faculty Evaluation and will allow you to easily cut and paste information from one document to another. The document may be useful for self-reflection or for meetings with mentors or your Career Development Committee. This is a template meant to be a working document, to be utilized at the discretion of the end user, and to be altered in order to meet individual needs.

Name and degree	Alex Bokov, Ph.D.
Year of initial faculty appointment	2015
Current academic rank	Instructor

Goals for the current year, progress toward meeting goals, barriers encountered/foreseen, mitigation strategies. Goals may be created in the categories of research and scholarly activities, teaching and mentoring activities, clinical activities, service and citizenship activities, and leadership and management activities.

Goals	Progress	Barriers encountered/foreseen	Mitigation strategies
Complete DataFinisher	Modest	time	submitting KL2, limiting new projects, investing in user training
Complete in preparation manuscripts	Almost done	time	setting aside dedicated time
Start working on NLM K01	Underway	time, clinical domain knowledge	submitting KL2, collaborating with Dr. Rodriguez
Prepare i2b2 for study: NAACCR, SES, vital status	Almost done	competing priorities for CIRD staff time	do most of the programming myself



2. Additional significant achievements.

Develop leadership, initiative, and collaboration skills by participating in the Informatics Domain Task Force, ACT. Build a good working relationship with Urology faculty.

3. Distribution of effort (Should total 100%)

Research and scholarly activities	Teaching and mentoring activities	Clinical activities	Service and citizenship activities	Leadership and management activities
30%	10%	0%	40%	20%

4. Long-term career goals (3–5+ years)

Goals	Competencies/skills/knowledge needed (Areas you need to develop)
Become Assistant Professor	grantsmanship; scientific communication; thorough understanding of clinical practice; multidisciplinary collaborations
Develop DataFinisher into a federated data extraction system capable of running on a SHRINE network	advanced informatics skills, reputation for excellence in open source informatics software;
Become chief of an informatics division and lead PI of a multi-site EMR-based study of kidney cancer predictors	leadership; organizational skills; multidisciplinary collaborations; thorough understanding of clinical practice particularly urological oncology and surgery



**5. Development plan for acquiring competencies/skills/knowledge needed
(Areas you need to develop in order to achieve your long-term goals)**

Specific competencies/skills/knowledge needed	Action steps for acquiring	Involvement of manager, mentors, etc.	Target dates/incremental milestones	Outcomes (successes or failures)
grantsmanship	continue submitting and contributing to grants; go to GWNI	mentoring team, KL2 leadership, Dr. McManus	have next target grant chosen by December	specific aims and draft of research plan
understanding clinical practice	attend Grand Rounds, Tumor Board, PSPP	Dr. Rodriguez	start analysis of training subset of the data in 2018, ahead of schedule	Enough early results to submit poster for September 2018 ASPO deadline
advanced informatics skills, reputation for excellence in open source informatics software;	start collaboration with Dr. Murphy, finish DataFinisher	Dr. Murphy	finish DataFinisher by December	production release of DataFinisher
leadership; organizational skills	assist with smooth CIRD transition, continue to be active in ACT; enroll in leadership course	Dr. Ramirez, colleagues at other CTSA informatics departments	training program launched in advance of local ACT SHRINE launch	