BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Russell Greiner PhD	Professor, Department of Computing Science, University of Alberta
eRA COMMONS USER NAME (credential, e.g., agency login)	

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
California Institute of Technology, Pasadena, CA	BSc	05/76	Mathematics & Computer Science
Stanford University, Palo Alto, CA	MSc	05/79	Computer Science
Stanford University, Palo Alto, CA	Ph.D.	05/85	Computer Science

A. Personal Statement

There are now a great many medical datasets, whose individual records each describe some specific patient, providing (say) histological, glucose level, genomic, proteomic and metabolomic information for that patient, as well as some relevant outcome information -- eg, the specific diagnosis, or the efficacy of some treatment, or survival time. The standard bio-statistical analysis seeks univariate correlations, which identify the individual "features" (genes, proteins, metabolites ...) that are most correlated with the outcome. While this is extremely useful when designing the next experiment to run (eg, which gene to knock out?), it is not, in-and-of-itself, sufficient for *personalized treatment* -- i.e., it does not provide the survival time for Ms Smith, nor identify which treatment will work best, for her. For this task, it is important to produce a "predictive model"; that is, a set of rules that can use *all* of the information about Ms Smith, to determine the treatment that is best for her. Fortunately, there are tools for this exact task, many produced by the field of Machine Learning, which is my area of expertise.

Over the last decade, I have applied this technology (both standard approaches, and novel extensions) to a wide variety of fascinating medical applications, in collaboration with many talented medical researchers. These projects have led to a number of publications -- some that analyze existing datasets **[1,4,5,7,8,9,11,14,15,18,23,28,29]**, others that provide novel approaches to better understand data [**2**], including some in medical imaging **[9,15,21,C1,C4,C5]**, and yet others that provide a completely novel way to predict the survival time for an individual patient -- think of a Kaplan-Meier plot that is specific for Ms Smith, based not just on the site and stage of her cancer, but also on her age, gender, histology, and every other factor known about her **[C2]**.

Among other tasks, this project seeks ways to improve clinic outcomes, by determining which treatment will work best for each individual patient. This task fits perfectly into the machine learning framework: given such characteristics of historical patients – including the treatment administered, and the outcome – produce a classifier that can accurately predict the effectiveness of each specific treatment on a new patient, towards identifying the one that is most effective. This same idea can be extended to identify which sequence of tests and treatments (that is, patient care pathway) is most appropriate for the patient population, which could be honed based on the information collected about each patient. This body of tools could help provide the prompt and accurate diagnoses, and personalized treatments.

As evidence of my international reputation in machine learning in general, note that I have served on the editorial boards of essentially every prominent journal in Machine Learning, as Editor-in-Chief on a major journal in the larger field of Artificial Intelligence, and as Program Chair, then General Chair, in the field's most prominent fully-refereed conference. I was also the founding scientific director of one of the world's most prominent centres in Machine Learning (AICML). In summary, I have a demonstrated record of successful and productive research projects in this general area of applying machine learning to medical

problems, and my expertise and experience have prepared me to be an essential participant of this exciting project.

B. Positions and Honors

07/1998 – Present Full Professor, University of Alberta, Canada: Dept Computing Science

- Best papers at: Cdn AI (2010); IJCAI (2005); UM (2005)
- Killam Fellowship (2007); McCalla Professorship (2005)
- Fellow of AAAI (2007)
- ASTech Award (AICML)
- CoFounder of ChenomX Inc (2001) – a metabolomics software and service company.
- Public website [C2]: Patient-Specific Survival Predictions: http://pssp.srv.ualberta.ca/

C. Selected Peer-reviewed MEDICAL/BIOLOGICAL Publications [from 61 journals; 152 top tier conferences)[1]]

[1] M. Bastani, L. Vos, et al. "A Machine Learned Classifier that uses Gene Expression Data to Accurately Predict Estrogen Receptor Status". *PLoS One*, June 2013.

[2] S. Ravanbakhsh, M. Gajewski, et al. "Determination of the optimal tubulin isotype target as a method for the development of individualized cancer chemotherapy*". Theor. Biology and Medical Modelling*, 2013.

[3] C. Stretch, S. Khan, et al. "Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature". *PLoS One*, April 2013.

[4] M. Hajiloo, et al. "ETHNOPRED: a Novel Machine Learning Method for Accurate Continental and Sub-continental Ancestry Identification and Population Stratification Correction". *BMC Bioinformatics*, 14:61, 2013.

[5] M. Hajiloo, B. Damavandi, et al. "Breast Cancer Prediction Using Genome Wide Single Nucleotide Polymorphism Data ". *BMC Bioinformatics*, January 2013.

[6] D. Wishart, T. Jewison, et al. "HMDB 3.0 - The Human Metabolome Database in 2013". *Nucleic Acids Research*, November 2012.

[7] G. Sidhu, et al. "Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD". *Frontiers in Systems Neuroscience*, 6, October 2012.

[8] M. Brown, et al. "ADHD-200 Global Competition: Diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements ". *Frontiers in Systems Neuroscience*, 6, Sept 2012.

[9] C Stretch, T Eastman, et al. Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach. *J Nutrition*, 2012.

[10] B. Saha, N. Ray, et al. Quick detection of brain tumors and edemas: A bounding box method using symmetry. *Computerized Medical Imaging and Graphics*, 2011.

[11] D. Moulavi, M. Hajiloo, et al. Combining gene expression and interaction network data to improve kidney lesion score prediction. *Inter'l J Bioinformatics Research and Applications*, 2011.

[12] B. Sehrawat, M. Sridharan, et al. "Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility". *Human Genetics*, March 2011.

[13] N. Psychogios, D. Hau, et al. The Human Serum Metabolome. *PLoS One*, February 2011.

[14] R. Eisner, J. Xia, et al. Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. *Metabolomics*, July 2010.

[15] N. Asgarian, et al. Learning to Predict Relapse in Invasive Ductal Carcinomas based on the Subcellular Localization of Junctional Proteins. *Breast Cancer Research and Treatment*, 121(2), May 2010.

[16] A. Kerhet, C. Small, et al. Application of Machine Learning Methodology for PET-Based Definition of Lung Cancer. *Current Oncology*, 17(1), July 2009.

[17] S. Damaraju, B. Sehrawat, et al. Candidate and Whole-Genome SNP Association Studies of Late Radiation Toxicity in Prostate Cancer Patients. *Radiation Research*, Sept 2008.

[18] A. Fyshe, Y. Liu, D. Szafron, R. Greiner, P. Lu. Improving Subcellular Localization Prediction using Text Classification and the Gene Ontology. *Bioinformatics*, August 2008.

[1] This list includes several conference papers. All were full-paper reviewed, and appear in archival venues, which serve as the primary means for disseminating results in computing science. See also http://tinyurl.com/RG-papers for a more comprehensive list of publications.

[19] D. Wishart, M. Lewis, et al. The Human Cerebrospinal Fluid Metabolome. *Journal of Chromatography B,* August 2008.
[20] C. Slupsky, K. Rankin, et al. Investigations of the Effects of Gender, Diurnal Variation and Age in Human Urinary Metabolomic Profiles. *Analytical Chemistry*, August 2007.
[21] D. Wishart, D. Tzur, et al. HMDB: the Human Metabolome Database. *Nucleic Acids Research*, 35, pp D521 -- D526, January 2007.
[22] M. Morris, R. Greiner, J. Sander, A. Murtha, M. Schmidt. Learning a Classification-based Glioma Growth Model Using MRI Data. *J of Computers*, 1(7), pp 21-31, November 2006.
[23] L. Pireddu, D. Szafron, P. Lu, R. Greiner. The Path-A metabolic pathway prediction web server. *Nucleic Acids Research*, 34, pp W714 -- W719, July 2006.
[24] S. Damaraju, D. Murray, et al. Association of DNA Repair and Steroid Metabolism Gene Polymorphisms with Clinical Late Toxicity in Patients Treated with Conformal Radiotherapy for Prostate Cancer. *Clinical Cancer Research*, 12(8), April 2006.
[25] G. Van Domselaar, P. Stothard, et al. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research*, 33, pp W455-W459, July 2005.
[26] P. Lu, D. Szafron, et al. PA-GOSUB: A searchable Database of Model Organism Protein Sequences with Their Predicted GO Molecular Function and Subcellular Localization. *Nucleic Acids Research*, 3, pp D147--D153, January 2005.
[27] D. Szafron, P. Lu, et al. Proteome Analyst: Custom Predictions with Explanations in a Web-based Tool for High-Throughput Proteome Annotations. *Nucleic Acids Research*, 32, pp W365-W371, July 2004.
[28] J. Listgarten, S. Damaraju, et al. Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms. *Clinical Cancer Research*, April 2004.
[29] Z. Lu, D. Szafron, et al. Predicting Subcellular Localization of Proteins using Machine-Learned Classifiers. *Bioinformatics*, 20(4), pp 547--556, March 2004.
[30] D. Wishart, L. Querengesser, et al. Medical Resonance Diagnostics --- A New Technology for High Throughput Clinical Diagnostics. *Journal of Clinical Chemistry*, October 2001

[C1] I. Diaz, P. Boulanger, et al. "An Automatic Brain Tumor Segmentation Tool". IEEE Engineering in Medicine and Biology Society, July 2013.
[C2] C. Yu, R. Greiner, H. Lin, and V. Baracos. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In NIPS(*), 2011.
[C3] S. Ravanbakhsh, B. Poczos, R. Greiner. A Cross-Entropy Method that Optimizes Partially Decomposable Problems: A New Way to Interpret NMR Spectra. AAAI (*), July 2010.
[C4] A. Farhangfar, R. Greiner, C. Szepesvari. Learning to Segment from a Few Well-Selected Training Images. *ICML(*)*, June 2009.
[C5] M. Schmidt, I. Levner, et al. Segmenting Brain Tumors using Alignment-Based Features. *International Conference on Machine Learning and Applications*, December 2005.

D. Research Support - Current:

1. Sponsor: Pfizer Psychiatry Research Award Treatment
   Title: Prediction of Individual Treatment Response Based on Brain Changes in the Early Phase of Antidepressant
   Type: Operating
   Total Award: $99,822; 2013-2015
   Investigators: R Ramasubbu + R Greiner, M Brown
2. Sponsor: Canadian Institutes of Health Research (CIHR)
   Title: Diagnostics and Prognostics for Alzheimer's Disease and Related Dementias: Machine Learning Analysis of Existing Neuroimaging Datasets
   Type: Catalyst Grant for Secondary Analysis of Neuroimaging Databases
   Total Award: $99,900; 2013-2014
   Investigators: R Greiner and S Dursun
3. Sponsor: Terry Fox Research Institute
   Title: The Virtual Biopsy Project: Non-Invasive Molecular Diagnosis in Glioblastoma
   Type: Operating
   Total Award: $100,000; Jan 20122 – December 2012
   Investigators: R. Mitchell, A. Murtha; G. Cairncross; R. Greiner, T. Magliocco, J. Sander

4. Sponsor: Alberta Innovates Health Solutions (AIHS)
   Title: The Alberta Food Metabolome Proposal-Comprehensive micronutrient characterization of Alberta-grown foods
   Type: Operating
   Total Award: $480,000 / year; December 2011 – November 2012
   Investigators: PI: D Wishart, CoPI: R Greiner
5. Sponsor: Alberta Health Services
   Title: Alberta Innovates Centre for Machine Learning-Alberta Health Services Collaboration
   Type: Operating
   Total Award: $250,000 / year; December 2009 – March 2013
   Investigators: R Greiner
6. Sponsor: NSERC Discovery Grant
   Title: Learning for Bio- and Medical-Informatics
   Type: Operating
   Total Award: $42,000 / year; April 2007 – March 31, 2014
   Investigators: R. Greiner
7. Sponsor: Alberta Innovates Technology Futures
   Title: Alberta Innovates Centre for Machine Learning
   Type: Operating
   Total Award: $2,000,000 / year; April 2009 – March 31, 2014
   Investigators: R. Greiner
8. Sponsor: Terry Forx Research Institute
   Title: Improved Assignment to Best Available Therapy for Myelodysplasia/Acute Myeloid Leukemia
   Type: Operating
   Total Award: $1,000,000 / year; July 2012 – June 2013
   Investigators: S. Couban (PI) and R. Greiner, et al.

D. Research Support - Past

- Screening for colorectal cancer and its precursors: a novel partnership; NSERC Engage; Jun12-Dec12; $24,721
- Alberta Innovates Centre for Machine Learning; Apr11-Mar12; $500,000/year; Alberta Innovates Technology Futures.
- Quantitation of Ki67 in ER positive Breast Cancers; Jan -Dec 11; $25,000; University Hospital Foundation (UHF) Medical Research Competition; J Hugh (PI) + R Chibbar, R Greiner.
- Development and evaluation of genomic selection methods for improving economically important cattle traits Agriculture Funding Consortium; Jul 08 – Jun 11; $156,666/yearZ Wang, P Stothard + S Moore; G Plastow; G Lin; R Greiner.
- Creation of the Alberta Transplant Applied Genomics Centre (ATAGC); Jan07–Jan10: $1,688,136/year; CFI; P Halloran [PI] and others.
- Creation of the Alberta Transplant Applied Genomics Centre (ATAGC); Jan07–Jan10; $1,378,312/year; Alberta Science and Research Investment Program (ASRIP); P Halloran [PI] and others.
- Genome-Wide Single Nucleotide Polymorphism Based Association Studies in Metastatic Breast Cancer; ACBRI [Operating]; Apr07–Mar10; $383,788/year; S Damaraju [PI], J Tuszynski, J Mackey, C Cass, R Lai, R Berendt, R Greiner.
- Identification and Validation of Pathways Associated with Failure of Standard Adjuvant Therapy in Early Stage Breast Cancer Duration: Apr07–Mar09; $245,457/year; Alberta Cancer Board [Operating Grant]; J. Mackey [PI], R. Lai, C. Cass, K. Graham, R. Greiner.