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: Purvesh Khatri

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

POSITION TITLE: Assistant Professor, Biomedical Informatics Research Center, Institute for Infection, Transplant and Immunity, Stanford University

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)*	Completion Date MM/YYYY	FIELD OF STUDY
BVM Engineering College, India	B.Eng.	09/98	Electronics Engineering
Wayne State University	M.S	04/06	Computer Science
Wayne State University	Ph.D.	04/06	Computer Science
Wayne State University	Postdoc	08/06	Computer Science
Stanford University	Postdoc	06/08	Systems Medicine/Transplant

A. Personal Statement

Dr. Khatri is an Assistant Professor at the Institute for Immunity, Transplantation, and Infection and the Center for Biomedical Informatics Research, Department of Medicine at Stanford University. Dr. Khatri has more than 15 years of professional experience in the areas of bioinformatics, computational biology, and translational medicine. He actively collaborates with many investigators on the Stanford campus, and at other institutes, with a goal to disseminate and implemented newly-invented diagnostic markers and therapeutic targets. He develops methods for the integration and analysis of high throughput genomics and proteomics data. He is well known for work on the development of ontological and pathway analysis of high throughput molecular data, and leveraging publicly available data for integrated, multi-cohort analyses for identification of diagnostic and therapeutic biomarkers. Dr. Khatri developed the first tool, Onto-Express, for analysis of microarray data using Gene Ontology. He expanded his work in ontological analysis to develop a suite of web-based open access tools, Onto-Tools. Currently, more than 15,000 people around the world are registered as Onto-Tools users. His recent work is focused on developing computational methods for integrated, multi-cohort analysis of publically available data to increase the sample size as well as better account for heterogeneity observed in real world patient population. Using these methods, he has integrated data sets from multiple centers consisting of distinct patient cohorts with different biological and technical confounders (i) to identify highly specific and sensitive biomarkers for acute rejection across all transplanted organs, cancers (pancreatic cancer, small cell and non-small cell lung cancer), and infectious diseases (sepsis, respiratory viral infections, tuberculosis) (ii) to suggest repositioning of FDA-approved drugs for treating transplant patients, and (iii) to identify novel gene involved in non small cell lung cancer and pancreatic cancer carcinogenesis that may be a potential drug target.

B. Positions and Honors

Positions and Employment

2001-2006 Research Assistant, Wayne State University, Department of Computer Science, Detroit, MI
2006-2008 Postdoctoral Scholar, Wayne State University, Department of Computer Science, Detroit, MI
2008-2010 Postdoctoral Scholar, Stanford University, Center for Biomedical Research, Stanford, CA
2010-2013 Research Associate, Stanford University, Division of Systems Medicine, Stanford, CA

2013-2014 Acting Assistant Professor, Stanford University, Department of Medicine, Stanford, CA
2014-Present Assistant Professor, Stanford University, Department of Medicine, Stanford, CA

Academic and Professional Honors

2010 "Young Investigator Award" at American Transplant Congress 2010 for Meta-analysis of Solid Organ Transplant Data Sets Identifies Differentially Expressed microRNAs common in Heart, Kidney and Liver Allografts.

2005 "Fast Breaking Paper" award in the field of Computer Science (Bioinformatics) for Khatri et. al. Bioinformatics 2005 Sep; 21(18):3587-3595 by ISI Thomson-Scientific Essential Science Indicator. According to ISI Thomson, these papers comprise the top 1% of papers in each field and each year (http://esi-topics.com/fbp/fbp-october 2006.html).

C. Contribution to Science

1. Developing novel methods for ontological analysis for interpretation of high throughput molecular data. In the late 1990s, following the advent of DNA microarrays, it was evident that these technologies posed a challenge of understanding underlying biology from the large amount of data generated by them. Dr. Khatri developed the first tool, called Onto-Express, for ontological analysis of transcriptomic data using the Gene Ontology annotations for identifying significant biological processes in a condition under study. The approach he proposed has been very successful, and a large number of tools similar to Onto-Express have been developed in the 10 years following its release. Following wide adoption of Onto-Express, he expanded his work to develop a suite of ontology-based analytical tools, called Onto-Tools, which has more than 15,000 registered users worldwide.

a. Purvesh Khatri, Sorin Draghici, G. Charles Ostermeier, Stephen A Krawetz. Profiling gene expression using Onto-Express. **Genomics** 2002, 79(2): 266-270. (PMID: 11829497)
b. G. Charles Ostermeier, David J. Dix, David Miller, Purvesh Khatri, and Stephen A. Krawetz. *Spermatozoal RNA profiles of normal fertile men.* **Lancet**, 2002 Sep; 360 (9335): 772-777.
c. Sorin Draghici, Purvesh Khatri, Rui P. Martins, G. Charles Ostermeier, and Stephen A. Krawetz. *Global functional profiling of gene expression.* **Genomics**, 2003 Feb; 81(2): 98-104.
d. Purvesh Khatri and Sorin Draghici. *Ontological analysis of gene expression data: current tools, limitations, and open problems.* **Bioinformatics**, 2005 Sep: 21(18): 3587-3595.

2. Developing novel methods for pathway analysis of high throughput molecular data. As ontological analysis approaches similar to Onto-Express were widely adopted, it became increasingly clear that these methods did not leverage the knowledge embedded in pathway knowledgebase such as KEGG, Reactome, BioCarta, etc. to their full potential. For instance, these approaches did not account for regulatory interactions (activation or inhibition) between genes in different pathways, and did not consider the type of genes (e.g., ligand, vs. receptor vs. transcriptional factor). Dr. Khatri developed novel pathway analysis methods, Pathway-Express and SPIA, which account for pathway topology to take advantage of the information embedded in biological pathways.

a. Sorin Draghici, Purvesh Khatri, Adi Laurentiu Tarca, Kashyap Amin, Arina Done, Calin Voichita, Constantin Georgescu and Roberto Romero. *A systems biology approach for pathway level analysis.* **Genome Research**, 2007 Oct; 17(10): 1537-1545.
b. Adi Laurentiu Tarca, Sorin Draghici, Purvesh Khatri, Sonia S. Hassan, Pooja Mittal, Jung-sun Kim, Chong Jai Kim, Juan Pedro Kusanovic, and Roberto Romero. *A novel signaling pathway impact analysis*. **Bioinformatics**, 2009; 25(1): 75-82.
c. Purvesh Khatri, Marina Sirota, and Atul J. Butte. *10 years of Functional Pathway Analysis: Current Approaches and Unaddressed Challenges.* **PLoS Computational Biology** 8(2): e1002375, 2012.

3. Developing novel methods to leverage heterogeneity present in public data through integrated multi-cohort analysis. A typical biological experiment is a controlled experiment, in which all samples are obtained from the same tissue, have been treated similarly, and are profiled using the same technology. Although tremendously useful and successful, a limitation of this approach is that before the results can be translated in to a clinical practice, they are required to be validated in multiple, independent cohorts because the results of a controlled experiment could still be influenced by an unknown biological or technological confounding factor. Explosive growth in the amount of data available in recent years provides a unique opportunity to address this challenge quickly and inexpensively. Integration of publicly available experimental data from multiple independent laboratories, studying the same phenotype or disease under similar or dramatically different

conditions, into a single analysis allows nearly-comprehensive representation of heterogeneity of the phenotype being studied. These data sets are generated by independent groups that follow (slightly) different experimental protocols, and use different technologies (e.g., oligonucleotide vs. cDNA microarrays), they represent various technological confounding factors in the data. Furthermore, different strains of an organism are used in *in vitro* experiments, or samples are collected from different countries in the case of human studies, and such genetic variation represents the biologic confounding factor. However, presence of these biological and technological confounding factors in different datasets also present challenges in their integration in a single analysis. Dr. Khatri has developed a novel framework for performing integrated, multi-cohort analysis of these diverse public data to identify robust signatures of disease phenotypes that are observed across multiple datasets, and are not affected by various confounding factors present in individual datasets. Using this framework, Dr. Khatri and colleagues analyzed more than 1,100 transplant biopsy samples from four organs (heart, lung, liver, kidney) in 13 cohorts from 11 hospitals in 6 countries profiled using 4 different types of microarrays to identify a 11-gene set that (1) significantly correlated with extent of graft injury, (2) diagnose stable and acute rejection patients with high specificity and sensitivity (AUC >0.8), (3) could diagnose subclinical injury to allograft 18 months earlier than clinical diagnosis (AUC =0.88). Using these 11 genes, we predicted two FDA-approved drugs that could be repurposed to treat transplant patients, and validated these predictions using mouse model of cardiac transplant and electronic medical records of more than 2,500 renal transplants patients followed for up to 10 years. Dr. Khatri and colleagues have also used this approach to identify tricyclic antidepressants that could be repurposed to treat small cell lung cancer patients.

a. Purvesh Khatri, Silke Roedder, Naoyuki Kimura, Katrien De Vusser, Alexander A. Morgan, Yongquan Gong, Michael P. Fischbein, Robert C. Robbins, Maarten Naesens, Atul J. Butte, and Minnie M. Sarwal. *A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation.* **Journal of Experimental Medicine**, 2013. 210(11):2205-2221.
b. L. Li*, P. Khatri*, T. K. Sigdel*, T. Trana, L. Ying, M. J. Vitalone, A. Chen, S. Hsieh, H. Dai, M. Zhang, M. Naesens, V. Zarkhin, P. Sansanwala, R. Chen, M. Mindrinos, W. Xiao, M. Benfield, R. B. Ettenger, V. Dharnidharka, R. Mathias, A. Portale, R. McDonald, W. Harmon, D. Kershaw, V. M. Vehaskari, E. Kamil, H. J. Baluarte, B. Warady, R. Davis, A. J. Butte, O. Salvatierra and M. M. Sarwal. *A Peripheral Blood Diagnostic Test for Acute Rejection in Renal Transplantation.* **American Journal of Transplantation**, 2012. 12(10): 2710-2718.
c. Ron Chen*, Purvesh Khatri*, Pawel K. Mazur, Melanie Polin, Yanyan Zheng, Dedeepya Vaka, Chuong D. Hoang, Joseph Shrager, Yue Xu, Silvestre Vicent, Atul J. Butte, and E. Alejandro Sweet-Cordero. *A Meta-analysis of Lung Cancer Gene Expression Identifies PTK7 as a Survival Gene in Lung Adenocarcinoma*. **Cancer Research**, 2014. 74(10): 2892-2902.
d. Pawel K. Mazur*, Nicolas Reynoird*, Purvesh Khatri, Pascal W. T. C. Jansen, Alex W. Wilkinson, Shichong Liu, Olena Barbash, Glenn S. Van Aller, Michael Huddleston, Dashyant Dhanak, Peter J. Tummino, Ryan G. Kruger, Benjamin A. Garcia, Atul J. Butte, Michiel Vermeulen, Julien Sage, and Or Gozani. *SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer.* **Nature**, 2014. 510(7504): 283-287.

4. Identification of diagnostic biomarkers for infectious diseases using public data. Sepsis is a whole-body inflammation syndrome set off when the immune system wildly overreacts to the presence of infectious pathogens. It is the leading cause of hospital deaths in the United States, accounting for nearly half of the total number, and is tied to the early deaths of at least 750,000 Americans each year. Its estimated annual cost to the health-care system exceeds $24 billion. It is critical for clinicians to diagnose sepsis accurately and quickly, because the risk of death from this condition increases with every passing hour it goes untreated. However, there are no rapid, definitive diagnostic blood tests for sepsis. Using our multi-cohort analysis framework, we analyzed 27 independent cohorts composed of more than 2,900 blood samples to identify a “recovery signature” in trauma patients as they recover during their stay in hospitals, which confounds majority of the sepsis studies. Further, we showed that accounting for this recovery signature identifies a robust signature of sepsis diagnosis, which allows diagnosis 2-to-5 days prior to clinical diagnosis of sepsis. This work is an example of using large amounts of publicly available data to identify transcriptional signatures capable of distinguishing different types of inflammation that are clinically usable.
Similarly, we have analyzed 27 independent cohorts from 19 data sets consisting of 3,819 samples that were collected in 7 countries, representing infections from 7 viruses and 4 bacteria in whole blood, PBMC and epithelial cells. Using these data, we have identified a common host transcriptional signature across different

respiratory viral infections that can distinguish individuals with viral infections from healthy controls and those with bacterial infections. We have also identified an influenza-specific host response signature that (i.) can distinguish individuals infected with influenza from those with either bacterial or other respiratory viral infections; (ii.) is both a diagnostic and prognostic indicator in influenza-pneumonia patients and influenza challenge studies; (iii.) can discriminate symptomatic from asymptomatic subjects, and identify symptomatic subjects prior to symptom onset in challenge studies; and (iv.) is predictive of response to influenza vaccines.

a. Timothy E. Sweeney, Aaditya Shidham, Hector R. Wong, Purvesh Khatri. *A comprehensive time-course–based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set.* **Science Translational Medicine** 2015 7 (287), 287ra71-287ra71.
b. Marta Andres-Terre, Helen M McGuire, Yannick Pouliot, Erika Bongen, Timothy E Sweeney, Cristina M Tato, Purvesh Khatri. Transcriptional signatures of viral infection across multiple respiratory viruses derived from integrated, multi-cohort analysis. **Immunity** 2015 43(6):1199-1211.