APPLICANT BIOGRAPHICAL SKETCH

Use only for individual predoctoral and postdoctoral fellowships, dissertation research grants (R36),and Research Supplements to Promote Diversity in Health-Related Research (Admin Suppl). DO NOT EXCEED FIVE PAGES.

NAME OF APPLICANT: Jean Fan

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

POSITION TITLE: Graduate Student

EDUCATION/TRAINING *(Most applicants will begin with baccalaureate or other initial professional education, such as nursing. Include postdoctoral training and residency training if applicable. High school students should list their current institution and associated information. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	START DATE MM/YYYY	END DATE *(or expected end date)* MM/YYYY	FIELD OF STUDY
Johns Hopkins University, Baltimore, MD	BS	08/2009	05/2013	Biomedical Engineering, Applied Mathematics and Statistics
Harvard University, Boston, MA	PhD	06/2013		Bioinformatics and Integrative Genomics

A.   Personal Statement

My longterm research interests involve the development of a comprehensive understanding of key genetic, epigenetic, and other regulatory mechanisms driving cellular identity and heterogeneity within cellular groups, tissues, and organs. I am particularly interesting in heterogeneity in the context of cancer and how this heterogeneity shapes tumor progression, therapeutic resistance, and ultimately clinical impact.

My extensive scientific research experience has exposed me to both the wet lab, empirical approaches as well as the dry lab, computational approaches in tackling different sides of the same biological questions. As a high school student, I conducted wet lab research at the National Cancer Institute to identify putative oncogenes within the $8p_{11-12}$ amplicon driving breast cancer pathogenesis. As an undergraduate, I developed computational algorithms in Rachel Karchin's lab to predict the deleterious impact of mutations based on sequence conservation. I also assessed genetic variation and population structures on an organismal level in Shamil Sunyaev's lab. For my doctoral training, I will focus on developing statistical and computational methods for analyzing genomics, not on an organismal level, but on a single cell level. Since the start of my doctoral training, under the guidance and mentorship of Peter Kharchenko as well as my collaborators, I have developed software for inferring spatial localization of gene sets (Fan *et al*, manuscript in review) and pathway and gene set overdispersion analysis (Fan *et al*, manuscript in review) as well as applied my computational skills to analyze locally disordered methylation (Landau et al, Cell 2014) and clonal evolution in developing drug resistance (Burger et al, Cancer Discovery 2015) in chronic lymphocytic leukemia.

I hope to continue developing powerful statistical methods with user-friendly computation software in close collaboration with wet lab researchers and oncologists to enable more personalized cancer therapies in this era of precision medicine.

B.   Positions and Honors

ACTIVITY/ OCCUPATION	START DATE (mm/yy)	ENDING DATE (mm/yy)	FIELD	INSTITUTION/ COMPANY	SUPERVISOR/ EMPLOYER
Summer research intern	06/2008	09/2008	Cancer Biology	National Cancer Institute, National Institutes of Health	Paul Meltzer and Liang Ciao
Undergraduate research scientist	08/2009	05/2013	Bioinformatics, Evolutionary biology, Genomics	Institute for Computational Medicine, Johns Hopkins University	Rachel Karchin
Summer research intern	06/2012	08/2012	Population genetics	Harvard Medical School	Shamil Sunyaev
Teaching assistant	08/2012	12/2012	Mathematics	Johns Hopkins University	Donniell Fishkind
Teaching fellow	10/2014	10/2014	Genomics, Transcriptomics	Harvard Stem Cell Institute	Peter Kharchenko

Awards and Honors

- Siemens Competition in Math, Science and Technology SemiFinalist (2008)

- Intel Science Talent Search SemiFinalist (2009)

- DC-AMS Scholarship Winner (2009)

- Posse Scholarship Semi-Finalist (2009)

- Johns Hopkins University Dean's list (2009-2013)

- Provost's Undergraduate Research Award Winner (2012)

- National Science Foundation Graduate Research Fellowship Program awardee (2013)

- Department of Defense American Society for Engineering Education National Defense Science and Engineering Graduate Fellowship awardee (2013)

Memberships

- American Society of Human Genetics

- International Society for Computational Biology

- American Association for the Advancement of Science

C.   Contributions to Science (for predoctoral students and more advanced candidates only; high school students, undergraduates, and postbaccalaureates should skip this section)

A more complete understanding of chronic lymphocytic leukemia

Advancements in high-throughput sequencing technologies have uncovered tremendous genetic, epigenetic, and transcriptional heterogeneity in chronic lymphocytic leukemia (CLL) but its impact on clinical course is not well understood. I have established a close collaboration with the Wu lab at the Dana-Farber Cancer Institute, where I have focused on developing and applying bioinformatics methods for (1) assessing variability of single cell gene expression, (2) calling mutations from single cell qt-qPCR data, (3) differential expression and gene set enrichment tests for both bulk and single cell, RNA-sequencing and targeted qt-qPCR data. Our collaboration has led to many scientific findings that contribute to a more complete understanding of CLL:
- Landau DA, Clement K, Ziller MJ, Boyle P, **Fan J**, Gu H, Stevenson K, Sougnez C, Wang L, Li S, Kotliar D, Zhang W, Ghandi M, Garraway L, Fernandes SM, Livak KJ, Gabriel S, Gnirke A, Lander ES, Brown JR, Neuberg D, Kharchenko PV, Hacohen N, Getz G, Meissner A, and Wu CJ. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell 2014, Dec 8; 26(6):813-25
- Burger, JA, Landau DA, Taylor-Weiner A, Zhang H, Sarosiek K, Wang L, Stewart C, **Fan J**, Hoellenriegel H, Sivina M, Dubuc AM, Fraser C, Han Y, Livak K, Zou L, Wan Y, Konoplev SN, Sougnez C, Abruzzo LV, Carter CL, Keating MJ, Davids M, Wierda WG, Cibulskis K, Zenz T, Werner K, Kharchencko P, Cin PD, Neuberg D, Kantarjian H, Lander E, Gabriel S, O'Brien S, Letai A, Weitz D, Nowak MA, Getz G, and Wu CJ. Clonal evolution in patients with chronic lymphocytic leukemia developing resistance to BTK inhibition. Cancer Discovery (manuscript pending publication)
- Pleiotropic effects of splice variants generated by SF3B1 mutations in chronic lymphocytic leukemia (manuscript in preparation)

Statistical methods and software for analyses of single cell data

While heterogeneity within cellular systems has long been widely recognized, only recently have technological advances enabled measurements to be made on a single cell level. Applying traditional bulk analysis methods on single cells has met with varied degrees of success due to the high levels of technical as well as biological stochasticity and noise inherent in single cell measurements. Therefore, novel statistical methods are needed to identify and characterize heterogeneity in single cells. In the Kharchenko lab, I have focused on developing methods for analyzing single cell data, including differential expression analysis methods that takes into account sources of technical noise inherent to single cell RNA-seq data, clustering methods to identify pathways and gene sets that exhibit coordinated variability, and methods for spatial placement of cell subpopulations based on expression signatures. This work has led to the development of various statistical methods available as software for the scientific community:
- **Fan J**, Salathia N, Liu R, Kaeser G, Yung Y, Herman J, Kaper F, Fan JB, Zhang K, Chun J, and Kharchenko PV. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nature Methods (manuscript in review)

STEM Outreach

Women are underrepresented in science, technology, engineering, and math (STEM) fields. Improving the representation of women in STEM is pertinent to workplace diversity, gender equality, and American innovation. To help address this issue, I have been involved in a number of outreach efforts:
- I was the lead software engineer for the BioHazardz 3D video games (http://bioinfor.me/), which teach students the fundamentals of protein evolution through an intuitive and attractive gaming environment.
- I founded the 501(c)3 non-profit cuSTEMized (http://www.custemized.org/) and developed the website, software, graphics, and content to enable parents to generate and download free personalized motivational ebooks to help girls envision themselves in science, technology, engineering, and math.
- I am the co-chair of the Harvard Graduate Women in Science and Engineering student group. I manage and lead organization of networking, professional development, and mentoring events for women in natural sciences, social sciences, and engineering at Harvard University.

D. Scholastic Performance

YEAR	SCIENCE COURSE TITLE	GRADE	YEAR	OTHER COURSE TITLE	GRADE
	Johns Hopkins University			**Johns Hopkins University**	
2009	General Physics I	A	2009	Honors Multivariable Calculus	A
2009	General Physics Lab	A+	2009	BME Modeling and Design	A
2010	General Physics II	A-	2009	State and Family: Revised Classical Perspective	A+
2010	General Physics Lab II	A	2009	Philosophic Classics	A
2010	Organic Chemistry I	A-	2010	Linear Algebra	A-
2010	Molecules and Cells	A-	2010	Discrete Mathematics	B
2010	Independent Research	A	2010	BME Design Group	B+
2011	Statistical Mechanics/ Thermo	A-	2010	Differential Equations/Applications	A
2011	Independent Research	A	2010	Data Structures	B+
2012	Independent Study - BME	A	2010	Responsible Conduct of Research	Sat
2013	Senior Design Project I	A	2011	Intro to Probability	A-
2013	Senior Design Project II	A-	2011	Systems and Controls	B+
			2011	Models and Simulations	B+
	Harvard University		2011	Intro to Optimization	A
2013	Principles of Genetics	B+	2011	System Bioengineering I	B+
2013	Analysis of the Biological Literature	B+	2011	System Bioengineering Lab I	A
2013	Quantitative Genomics	B+	2011	Women, Family, & the Body Politic	A-
2014	Computational and Functional Genomics	A-	2012	Intro to Stochastic Processes	A
2014	Biological Macromolecules: Structure, Function and Pathways	A-	2012	Intro to Statistics	A
2014	Introduction to Biomedical Informatics I	A	2012	Systems Bioengineering II	A-
2014	Conduct of Science	Sat	2012	Systems Bioengineering Lab II	A
2015	Introduction to Biomedical Informatics II	A	2012	Business Law I	A
			2012	Computational Molecular Medicine	A-

YEAR	SCIENCE COURSE TITLE	GRADE	YEAR	OTHER COURSE TITLE	GRADE
			2012	Systems Bioengineering III	A
	GRE Quant	**163/170**	2012	Business Law II	B+
	GRE Verbal	**163/170**	2013	History of Modern Medicine	A
	GRE Writing	5.5/6	2013	Justice and Health	B+
			2013	Medical Informatics	A-
			2013	Machine Learning in Complex Domains	B-
			2013	Law and the Internet	B+
				Harvard University	
			2013	Statistical Inference	B-
			2014	Selected Topics in High Dimensional Analysis	Sat
			2014	Bayesian Methodology in Biostatistics	B+