Doctoral Dissertation and Other Research Experience:

Summer High School Research Intern **June 2008 - Oct 2008**
Mentored by Paul Meltzer and Liang Ciao at the National Cancer Institute, National Institutes of Health
Project: Delineating the Role of BRF2 in Breast Cancer Pathogenesis
The $8p_{11-12}$ amplicon is a region of genetic amplification present in 10-15% breast cancers and is associated with poor prognosis. While putative driving oncogenes have been proposed, genes within this amplicon had yet to be definitively implicated in cancer growth, survival, or pathogenesis. I evaluated the role of *BRF2*, one gene located within this amplicon, in breast cancer growth and development. I performed immunoblot analysis and used previous CGH and expression array studies to establish a strong correlation between *BRF2* gene amplification and BRF2 protein overexpression, a trait consistent with an oncogenic role. I employed lentiviral-mediated gene transfer to deliver *BRF2* shRNA into breast cancer cells with $8p_{11-12}$ amplification and subsequently established longterm stable cell lines with shRNA constructs targeting *BRF2*, leading to marked reduction of BRF2. Using clonogenic assays, I determined that BRF2 inhibition resulted in impeded growth and proliferation rates as well as increased cell death. My findings suggest that *BRF2* is a relevant oncogene in the $8p_{11-12}$ amplicon and may play a role in breast cancer growth and pathogenesis.

- Fan J, Yu Y, Meltzer PS, Cao L. Delineating the Role of BRF2 in Breast Cancer Pathogenesis. HURJ 2011. 14, 53-55

Undergraduate Research Scientist **Aug 2009 - May 2013**
Mentored by Rachel Karchin at the Institute for Computational Medicine, Johns Hopkins University
Project: Computational Assessment of the Utility of Limiting Orthologous Sequence Depth in Mutation Impact Prediction Performance
To predict for the function impact of mutations, current computational models often use sets of orthologous sequences, which are presumed to originate from a common ancestor such that their differences can be attributed to mutation and selective pressures. However, the extent to which these orthologous sequences have been subjected to the same selective pressures and subsequently the validity of using overly distant orthologous sequences remains unknown. I devised a SVM classifier approach as well as implemented published approaches such as SIFT and PolyPhen2 to assess the utility of limiting orthologous sequence depth in mutation impact prediction performance in 33 Mendelian disease-related genes. I developed the feature scores used by the SVM classifier to capture information concerning the physiochemical differences between reference and variant amino acid residues as well as the evolutionary conservation of amino acid residues up to a certain phylogenetic distance depth limit. I measured the overall performance of predictions using standard protocols for statistical learning including calculation of ROC and AUC. My results suggested an orthologous sequence depth limit at the divergence point between vertebrates and invertebrates that may improve mutation impact prediction performance.

- Fan J, Karchin R. Computational Assessment of the Utility of Limiting Orthologous Sequence Depth in Mutation Impact Prediction Performance. International Congress of Human Genetics/American Society of Human Genetics Conference, Montreal, 2011 (Poster), the BME Undergraduate Research Day, Johns Hopkins University, 2012 (Poster), and Provost's Undergraduate Research Poster Session, Johns Hopkins University, 2012 (Poster)

Additional projects: Estimating the Phylogenetic Distance Between Target Organisms, Missense Mutation Trends in PIK3CA, Investigation of Pseudogenes as Potential Confounders of Mutation Function Prediction, Critical Assessment of Genome Interpretation - The Personal Genomes Challenge

Summer Undergraduate Research Intern **June 2012 – August 2012**
Mentored by Shamil Sunyaev for the Harvard-MIT HST and i2b2 BIG Program
Project: Detecting Synergistic Epistasis in Humans
The prevalence of sexual reproduction, despite its inherent two-fold cost disadvantage, suggests that sexual reproduction must confer some compensatory evolutionary advantage. The deterministic mutation hypothesis for the evolution of sex posits that such an evolutionary advantage may be achieved contingent on synergistic epistasis, whereby accumulations of deleterious mutations lead to larger decreases in relative fitness. We devised a theoretical test using variance-mean ratios of mutations accumulated since the out-of-Africa migration to detect synergistic epistasis in humans. I applied this test to Genome of the Netherlands (GoNL) data and compared various functional classes of mutations, hypothesizing that variance will be depleted for deleterious mutations but not for benign or neutral mutations. I devised and conducted statistical tests including nonparametric bootstrap, ANOVA, and principal component analysis to assess the significance of results and performed quality control tests to assess for potential batch and flow-cell effects. While detection of synergistic epistasis in humans remains inconclusive, my results did suggest segregation in variance-mean ratios between benign and damaging mutations.

Rotation Student **June 2013 - Sept 2013**
Mentored by Peter Kharchenko at the Center for Biomedical Informatics, Harvard University
Project: Transcriptional heterogeneity in mouse neural progenitor cells
Recent technological advances have revealed tremendous transcriptional heterogeneity among single cells. But how this transcriptional heterogeneity plays a role in cell behavior, fate, and function is still not well understood. We developed PAGODA to resolve multiple, potentially overlapping aspects of transcriptional heterogeneity by identifying known pathways or novel gene sets that show significant excess of coordinated variability among the measured cells. We demonstrate that PAGODA effectively recovers the subpopulations and their corresponding functional characteristics in a variety of single-cell samples, and use it to characterize transcriptional diversity of neuronal progenitors in the developing mouse cortex. Specifically, I contributed to the development of various clustering approaches to identify de novo gene sets that exhibit coordinated variability across cells and ultimately cluster cells into putative subpopulations. Integrating data from the Allen Brain atlas, I also developed an R package to spatially location cells based on their gene expression signatures. Our work resulted in the development of software that can be readily applied to diverse single cell RNA-seq datasets to assess transcriptional heterogeneity.

- 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 pending publication)

Rotation Student **Sept 2013 – Nov 2013**
Mentored by Nir Hacohen and Catherine Wu at the Broad Institute
Project: Locally disordered methylation in chronic lymphocytic leukemia
Intratumoral heterogeneity plays a critical role in tumor evolution. How DNA methylation contributes to this heterogeneity is not well understood. We performed genome-scale bisulfite sequencing of 104 primary chronic lymphocytic leukemias (CLLs) in bulk. We found that, compared with normal B cell samples, CLLs consistently displayed higher intrasample variability of DNA methylation patterns across the genome. I helped perform transcriptome analysis of single CLL cells revealed that methylation disorder was linked to low-level expression.

- 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

Additional projects: Machine learning algorithm for inferring VDJ-recombination from SNP6-array data

Graduate Student **December 2013 – Present**
Mentored by Peter Kharchenko at the Center for Biomedical Informatics, Harvard University
Project: Precise dissection of genetic and transcriptional heterogeneity in chronic lymphocytic leukemia by single cell analysis
Intratumoral genetic heterogeneity is the basis of tumor cell plasticity. To more accurately dissect this heterogeneity, detect subclones, define phylogenetic relationships, and to directly uncover genotype-phenotype relationships, we developed a versatile approach based on qPCR for simultaneous targeted mutation and gene expression detection from single cells. I developed the computational methods for mutation calling from raw florescence readouts. We have applied and will continue to apply this method to study chronic lymphocytic leukemia, revealing distinct genetic subclones at different stages of CLL progression.

- 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)

Additional projects: Transcriptional heterogeneity and similarity in ferret and human neural progenitor cells, Genetic and transcriptional heterogeneity in multiple myeloma, Progenitor origins in chronic lymphocytic leukemia, Mouse model for SF3B1 mutation