BIOGRAPHICAL SKETCH

NAME: Malhotra, Ankit

eRA COMMONS USER NAME: ANKIT.MALHOTRA

POSITION TITLE: Associate Computational Scientist

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	Completion Date MM/YYYY	FIELD OF STUDY
Hans Raj College, University of Delhi, Delhi, India	B.S.	05/2003	Computer Science
University of Virginia, Charlottesville, VA	M.S.	01/2007	Computer Science
University of Virginia, Charlottesville, VA	Ph.D.	08/2010	Biochemistry

A. Personal Statement

The work described in the project would involve creating methods for discovering complex structural variation and INDELs from whole genome sequence datasets being generated by the various projects of the TOPMed program. The identified variants would be used in cross-program analyses to gain novel biological insight in human biology and disease mechanisms.

I have been working in the field of genetic variation for about eight years and have authored several important publications during my doctoral studies (in the lab of Dr. Anindya Dutta) as well as a postdoctoral researcher in the lab of Dr. Ira M Hall at the University of Virginia. My basic training (B.S. and M.S.) was in the field of Computer Science, and I earned my Ph.D. from the Department of Biochemistry at University of Virginia. This cross-disciplinary training has enabled me to bring sophisticated algorithms and methods from the field of computer science and apply them to important questions in biology. I was one of the first people to analyze data from high-throughput sequencing machines and developed important algorithms to discover and characterize structural variations in yeast and subsequently in human genomes. During my postdoctoral training, I worked on meta-analysis of complex structural variants from a large cohort of cancer datasets from the TCGA consortium and made very interesting and impactful observations. More recently, as part of the 1000 Genomes project, I have been involved in developing new methods and analysis to help generate the largest cohort of germline SVs from 2,504 individuals across 27 different populations. In total, I have authored/coauthored 18 peer reviewed publications on various aspects of genome analysis.

For this project, I will provide my expertise in the field of structural variation analysis. I will be involved in novel method development and analysis of sequencing datasets from the different TOPMed cohorts. As an Associate Computational Scientist at The Jackson Laboratory, I will lend my skills and experience in bioinformatics, cancer biology and genomic data analysis to this project.

B. Positions and Honors

2002-2003	Research Assistant, Dr. Harmeet Kaur, Hans Raj College, University of Delhi, Delhi, India
2003-2003	Research Assistant, Dr. P.R. Panda, Indian Institute of Technology, Delhi
2003-2010	Graduate Research Assistant, Dr. Anindya Dutta, University of Virginia, Charlottesville, VA
2010-2013	Research Associate, Dr. Ira M Hall, Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA
2013-Present	Associate Computational Scientist, Computational Sciences, The Jackson Laboratory for Genomic Medicine, Farmington, CT

C. Contribution to Science

My primary contributions are in field of cancer genomics, more specifically the study of structural variation and its impact on human health and disease. Over the last 8 years, I have contributed methods for discovering structural variation using next generation sequencing data. In 2013, we published the first meta-analysis of complex variations in multiple cancer genomes sequenced as part of the TCGA (The Cancer Genome Atlas) consortium. This study highlighted the importance of studying complex events, as they could be the initiating events in a patient's history. We also studied mechanisms that gave rise to such events and concluded that non-homologous repair of concurrently arising DNA double-strand breaks is the predominant mechanism underlying complex cancer genome rearrangements. I served as the primary investigator / analyst for all of these studies. More recently, as part of the Phase 3 of the 1000 Genomes project, I have also been involved in generating the largest set of known germline SVs from 2,504 individuals across 27 different world populations.

1. Shibata, Y., Malhotra, A. & Dutta, A. Detection of DNA fusion junctions for BCR-ABL translocations by Anchored ChromPET. *Genome Med* **2,** 70 (2010).
2. Malhotra, A. *et al.* Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms. *Genome Res.* **23,** 762–776 (2013).
3. Malhotra, A. *et al.* Ploidy-Seq: inferring mutational chronology by sequencing polyploid tumor subpopulations. *Genome Med* **7,** 6–6 (2015).
4. Sudmant, P et al. An integrated map of structural variation in 2,504 human genomes. *Nature 2015; 526(7571):75-81*

My earliest publications were as part of the pilot phase of the ENCyclopedia Of DNA Elements (ENCODE) project. I was part of a two-person team that worked on the analysis of the genomic data from the cell lines to produce a time of replication across the 1% of the genome. These resulted in the following publications:

1. ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
2. Karnani, N., Taylor, C., Malhotra, A. & Dutta, A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome Res.* **17,** 865–876 (2007).
3. Karnani, N., Taylor, C. M., Malhotra, A. & Dutta, A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol. Biol. Cell* **21,** 393–404 (2010).

Link to Publications:
(http://www.ncbi.nlm.nih.gov/sites/myncbi/ankit.malhotra.1/bibliography/48094753/public/?sort=date&direction=ascending)

D. Research Support

Current Research Support

U41 HG007497 Lee (PI) 08/01/13-06/30/17
NIH/NHGRI
An integrative analysis of structural variation for the 1000 Genomes Project
The major goal of this project is to develop and assess new methods for accurately identifying structural genomic variants in next generation DNA sequencing datasets.
Role: Co-Investigator, Computational Scientist

Completed Research Support

2011-2013: **Department of Defense Breast Cancer Postdoctoral Fellowship Award**, for proposal titled "*Role and mechanism of structural variation in progression of breast cancer*", awarded by Office of the Congressionally Directed Medical Research Programs (CDMRP), US Dept. of Defense
Role: PI