

## **BIOINFORMATICS PIPELINE FOR DETECTION OF IMMUNOGENIC CANCER MUTATIONS BY HIGH THROUGHPUT MRNA SEQUENCING**

Jorge Duitama<sup>1</sup>, Ion Mandoiu<sup>1</sup>, and Pramod K. Srivastava<sup>2</sup>

<sup>1</sup>Computer Science & Engineering Department, University of Connecticut

<sup>2</sup>Center for Immunotherapy of Cancer and Infectious Diseases,  
University of Connecticut Health Center

Immunotherapy is a promising cancer treatment approach that relies on awakening the immune system to the presence of antigens associated with tumor cells. The success of this approach depends on the ability to reliably detect immunogenic cancer mutations, the vast majority of which are expected to be tumor-specific.

We present a bioinformatics pipeline for detection of immunogenic cancer mutations from high throughput mRNA sequencing data. To improve the success rate and accuracy of read mapping, our pipeline maps mRNA reads against both reference genome and consensus coding sequences (CCDS) and combines mapping results. To improve SNP calling, we implemented a model that chooses the genotype with highest posterior probability based on counts and base quality scores. Validation against known SNP genotypes of a Hapmap individual shows that the combined mapping strategy yields improved SNP calling accuracy compared to performing genome or CCDS mapping alone. 96 immunogenic mutations predicted by our pipeline from approximately 11 million mRNA reads generated from a mouse cancer tumor cell line using a Illumina platform are currently under experimental validation.