

## **ScanIndel: a novel and versatile indel caller for next generation sequencing in cancer research and clinical applications**

**Rendong Yang**<sup>1</sup>, Andrew C. Nelson<sup>2</sup>, Christine Henzler<sup>1</sup>, Bharat Thyagarajan<sup>2</sup> and Kevin A.T. Silverstein<sup>1</sup>

<sup>1</sup> Supercomputing Institute for Advanced Computational Research, University of Minnesota, Minneapolis, MN 55455.

<sup>2</sup> Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455.

**Background:** Comprehensive identification of insertions/deletions (indels) across the full size spectrum from next generation sequencing (NGS) is challenging due to the relatively short read length inherent in the technology. Different indel calling methods exist but are limited in detection to specific sizes with varying accuracy and resolution. Comparing with coding point mutations, copy number alterations and translocations that have been explored and defined in the prostate cancer genome, indels have received far less attention comparing with. We believe an important reason is the limitation of current bioinformatics methods for efficiently discovering indels from short reads provided in NGS.

**Methods:** We present ScanIndel, an integrated framework for detecting indels with multiple heuristics including gapped alignment, split reads and *de novo* assembly. Using simulation data, we demonstrate ScanIndel's superior sensitivity and specificity relative to several state-of-the-art indel callers across various coverage levels and indel sizes.

**Results:** ScanIndel yields higher predictive accuracy with lower computational cost compared to existing tools for both amplicon-based targeted resequencing data from tumor specimens and high coverage whole-genome sequencing data from the human NIST standard NA12878. Thus we anticipate ScanIndel will improve indel analysis in both clinical and research settings. ScanIndel is implemented in Python, and is freely available for academic use at <https://github.com/cauyrd/ScanIndel>

**Conclusions:** We present ScanIndel as a robust method for detecting indels from targeted amplicon-based to WGS data. In particular, ScanIndel reliably detects medium-size indels and has comparable performance with existing methods for detecting very large indels. ScanIndel is capable of detecting indels across the full size spectrum with base-pair resolution. We anticipate ScanIndel will enable identification and elucidation of clinically actionable indels that are currently difficult to characterize.

**Conflict of Interest:** No conflicts.

**Funding Acknowledgements:** RY is supported by PCF Young Investigator Award (YIA 2015)