


IndexTools
A toolkit for accelerating
genomic analysis using NGS
index formats

 **John Didion**, Yih-Chii Hwang,
Steve Osazuwa, Aleksandra
Zalcman

INTRO

- Variant calling is embarrassingly parallel, can be accelerated by a “scatter-gather” approach.
- Optimal parallelization requires N data partitions of approximately equal size (N is multiple of # cores).
- BAM index files (.bai) contain a coarse-grained representation of read density across the genome.
- **IndexTools** leverages index files for ultra-fast genomic applications.

RESULTS

Compared to naïve parallelization, index-based partitioning of a 9.7 Gb Exome BAM file for GATK4 variant calling results in a **9%** speedup (mean 44 vs 48 minutes).

