Title	Prioritization of structural variants based on known biological information
Authors	*Brad Chapman*, Rory Kirchner, Miika Ahdesmaki, Justin Johnson, Shannan Ho Sui, Oliver Hofmann
Affiliations	Harvard Chan School Bioinformatics Core (http://hsphbio.ghost.io/), AstraZeneca Oncology (http://www.astrazeneca.com/Medicines/Oncology), Wolfson Wohl Cancer Research Centre (http://www.gla.ac.uk/researchinstitutes/cancersciences/ics/facilities/wwcrc/)
Contact	bchapman@hsph.harvard.edu
Availability	https://github.com/chapmanb/bcbio-nextgen
License	MIT

High-throughput human resequencing characterizes whole genome changes with the goal of linking variations to disease, drug responses or other phenotypes. The primary challenge following sensitive and precise variant detection is prioritizing the large number of results in the context of previously known biological information. This is especially problematic for samples that are not well explained by short variations like single nucleotide polymorphisms (SNPs) or small insertions and deletions. In these cases, structural variations such as larger insertions, deletions, rearrangments or copy number variations (CNVs) provide additional sources of causative variability. However, detecting structural variations from short reads is challenging, so biologists must search through a noisier dataset to find potentially relevant mutations for additional investigation.

We'll discuss an approach to help prioritize structural variations using pre-existing biological information. The approach is general and only reliant on inputs that link known mutations to genomic position, allowing incorporation of custom BED or VCF files into analyses. In this talk, we'll emphasize using public databases like COSMIC (https://cancer.sanger.ac.uk/cosmic), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) and CIViC (https://civic.genome.wustl.edu/) to evaluate cancer samples. We overlay known variants with existing annotations on genes, domains and other genome elements from Ensembl. Regions with pre-existing changes that match those found in BED files of structural variants are reported along with supporting information. We'll discuss practical examples of how this helps improve our ability to utilize structural changes in analysis of tumor variant calls.

The implementation is part of bcbio (https://github.com/chapmanb/bcbio-nextgen), which provides a configuration file and command line interface for running variant analysis on distributed machines. We have an open development community which contributed to our current cancer calling support (http://bcb.io/2015/03/05/cancerval/). We actively develop and support bcbio and hope to grow the community of users who both contribute and use it for answering biological questions.