
Learning the signatures of structural variation in the genome

Victoria Popic

**Schmidt Fellow and Principal Investigator
Broad Institute of MIT and Harvard**

March 04, 2024

1:30PM-2:50PM

MSOB X303

Abstract:

Structural variants (SVs) are the exceptionally diverse set of large-scale genome rearrangements. SVs encompass mutations, such as deletions, inversions, duplications, translocations, and any complex combination thereof. Accounting for more base-pair differences across individuals than all other variant types combined, SVs are the greatest source of genetic diversity in the human genome and a key driver of disease. To date, SVs have been linked to a wide spectrum of disorders, including autism, lupus, Huntington's, Alzheimer's, and cancer. However, general SV discovery still remains an open problem. In large part, this is due to the intractability of accurately modeling the full range of SV signatures using manually-engineered heuristic approaches, which have been predominantly developed and employed to date.

In this talk, we will motivate the need for generalizable methods for SV discovery and introduce the deep learning method, Cue, for SV calling and genotyping. Cue casts SV discovery as a multi-class keypoint localization task in custom images designed to capture patterns of structural discordance derived from sequence alignments to a reference genome. We will show that Cue effectively learns signatures of multiple SV classes given inputs from different sequencing technologies, providing a unified framework for cross-platform analysis of structural variation.

Reading list:

Structural variant calling: the long and the short of it

Medhat Mahmoud, Nastassia Gobet, Diana Ivette Cruz-Dávalos, Ninon Mounier, Christophe Dessimoz, Fritz J. Sedlazeck

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1828-7>