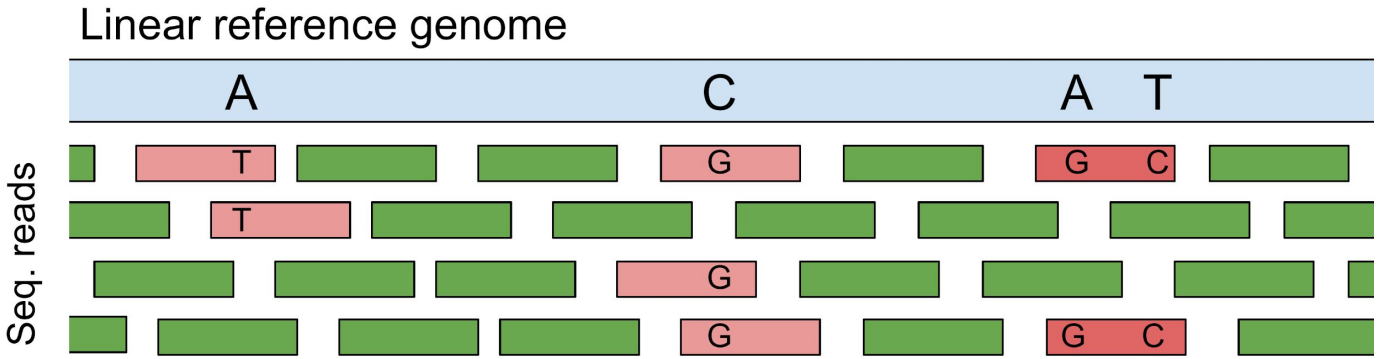
Genotyping structural variants in pangenome graphs using the vg toolkit

Jean Monlong

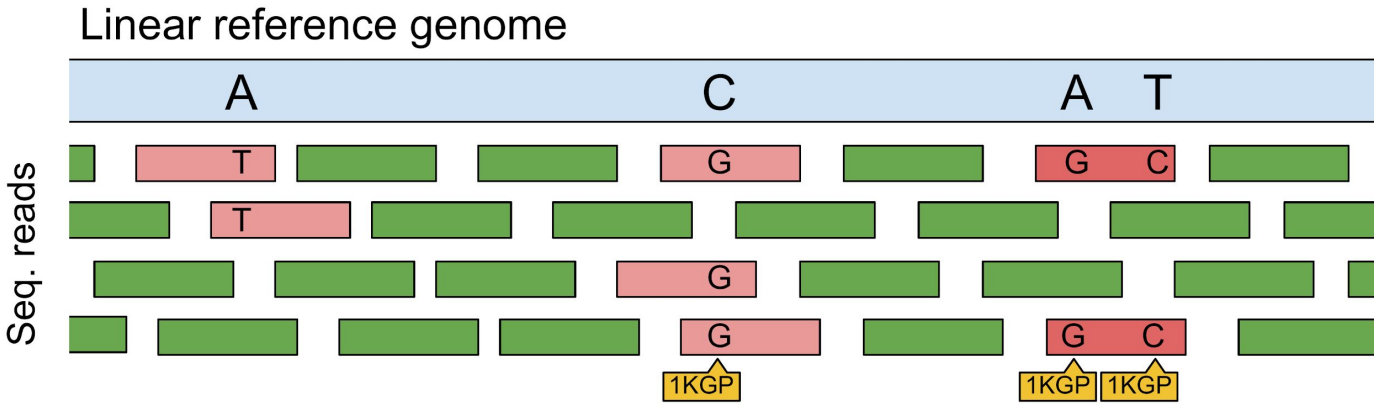
NCCB 2019

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute

Re-sequencing approach: short reads mapped to reference genome

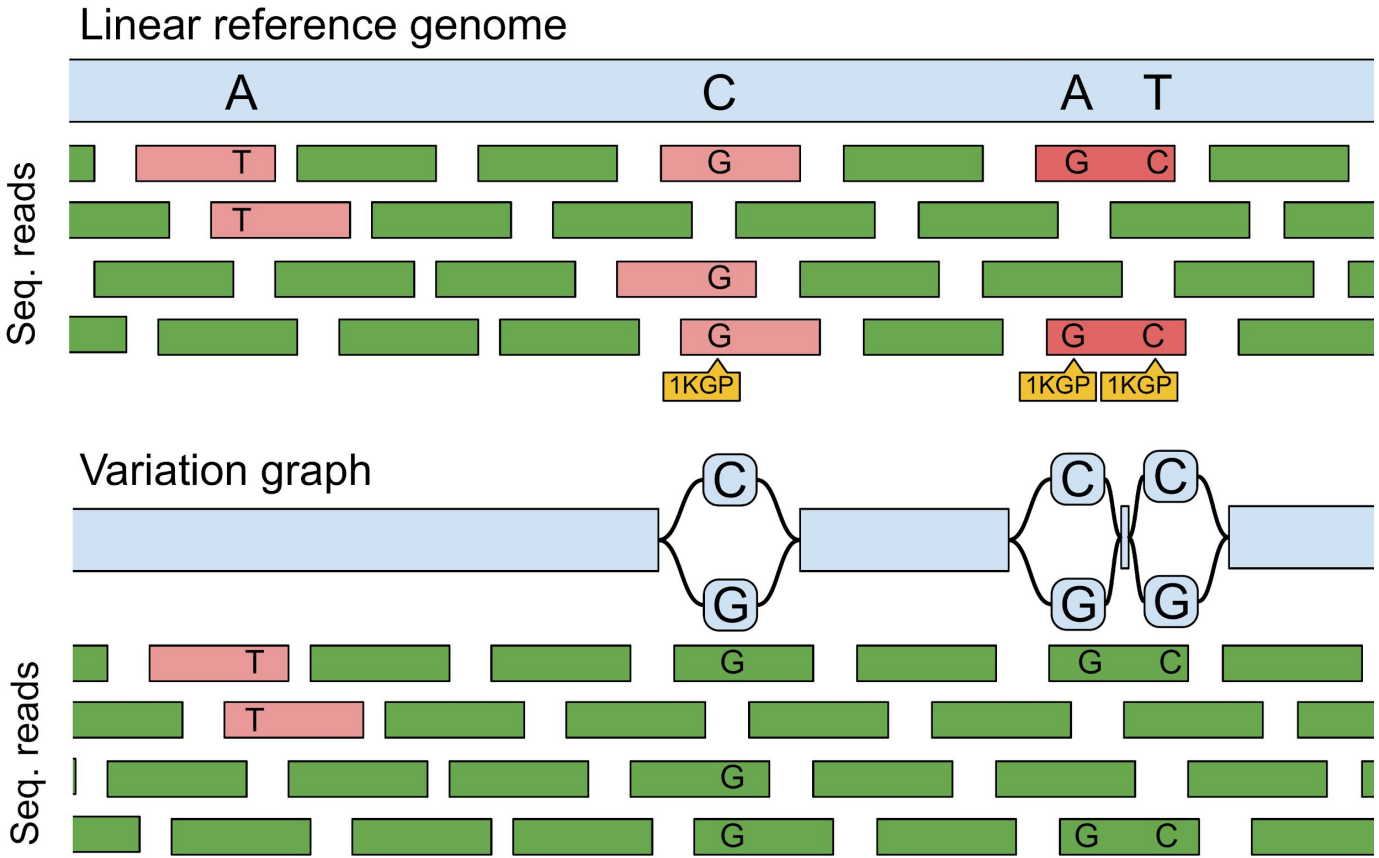


Re-sequencing approach: short reads mapped to reference genome



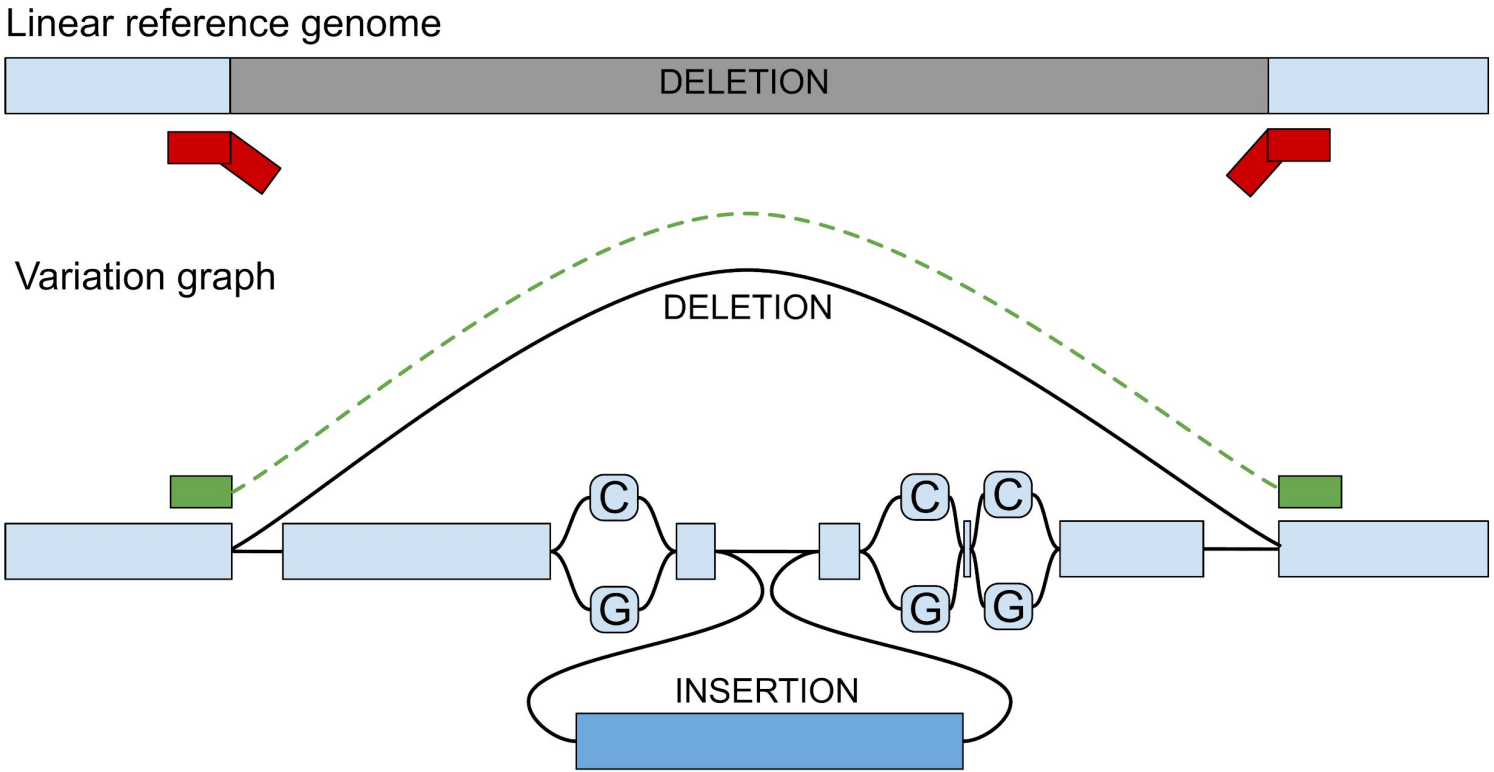
Sequencing reads map better on variation graphs.

(Garrison et al. Nat. Biotech. 2018)



Structural Variants (SVs)

Genomic variants >50bp. E.g. insertions, deletions, inversions.



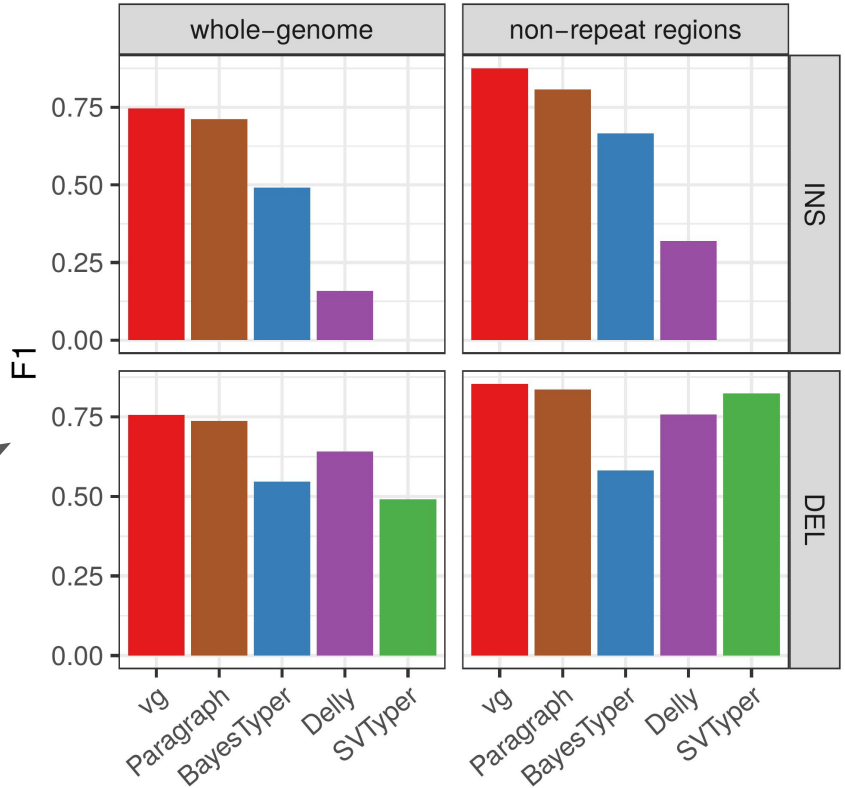
High-quality SV catalogs from long-read sequencing studies
(HGSVC 2019, GIAB 2019, SVPOP 2019).

Can we genotype these SVs from short-read data?

High-quality SV catalogs from long-read sequencing studies
(HGSVC 2019, GIAB 2019, SVPOP 2019).

Can we genotype these SVs from short-read data?

Yes, and better than traditional approaches.



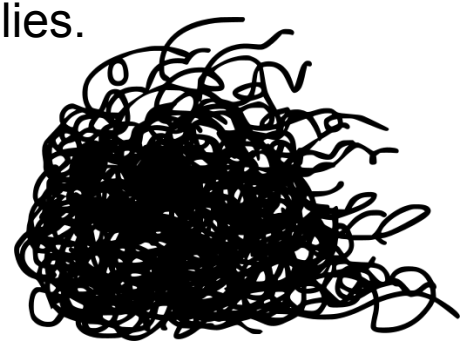
Graph-based SV genotypers: **vg**, Paragraph, BayesTyper

Traditional SV genotypers: Delly, SVTyper

Hickey et al. bioRxiv 2019

Work in progress

Starting from whole-genome alignment of de novo assemblies.



- No need to “catalog” SVs relative to a reference.
- High-quality assemblies from nanopore seq.
  (see Ryan Lorig-Roach’s poster)

Genotyping large cohorts of individuals.

- Get frequency estimates.
- Fine-tune SV breakpoints.
- Disease association studies.

NHLBI TOPMed

Acknowledgements

Glenn Hickey
David Heller
Jonas Sibbesen
Jouni Siren
Jordan Eizenga
Eric Dawson
Erik Garrison
Adam Novak
Benedict Paten

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics Institute

vg team
https://github.com/vgteam/vg

Check out Jonas' poster on spliced variation graphs for RNA-seq analysis.