An assessment of computational genotyping of Structural Variations for clinical diagnosis

Fritz Sedlazeck

Oct, 16, 2018

Scientific interests

Mapping/ Assembly reads

NextGenMap-LR
Sedlazeck et. al. (2018)

Falcon Unzip
Chin et.al. (2016)

NextGenMap
Sedlazeck et.al. (2013)

Detection of Variants

Clairvoyante
Lou et al. (in review)

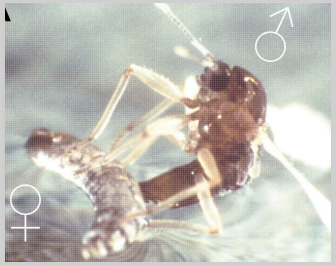
Sniffles
Sedlazeck et. al. (2018)

SURVIVOR
Jeffares et. al. (2017)

Benchmarking

Teaser
Smolka et.al. (2015)

Sequencing
Jünemann et.al. (2013)

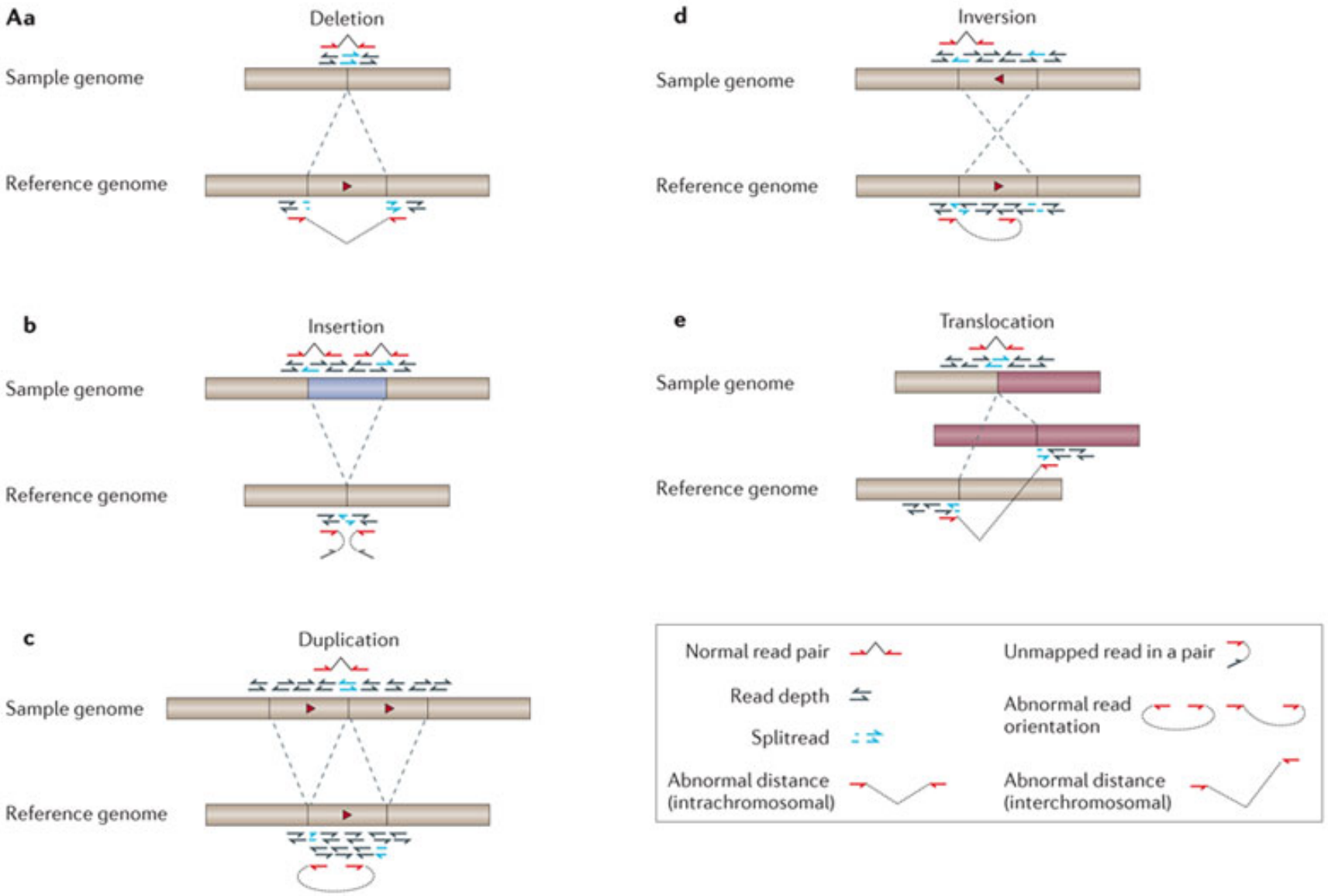


Applications

Model organisms:
-Cancer (SKBR3) (in preparation)
-miRNA editing (Vesely et.al. 2012)

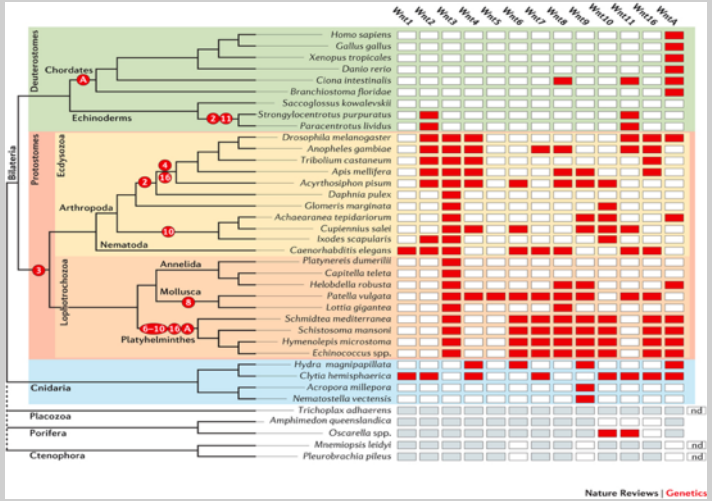
Non Model organisms:
-Cottus transposons (Dennenmoser et. al. 2017)
-Clunio (Kaiser et. al. 2016)
-Seabass (Vij et.al. 2016)
-Pineapple (Ming et.al. 2015)

How to detect Structural Variations

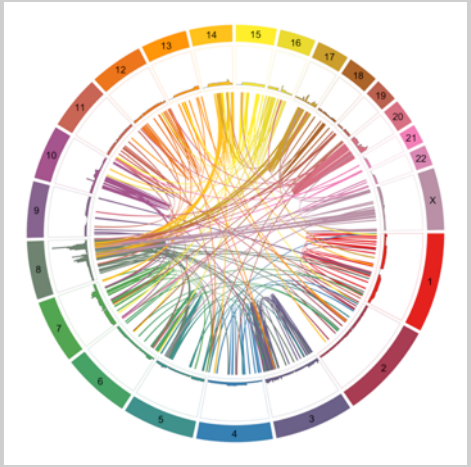


Structural Variations

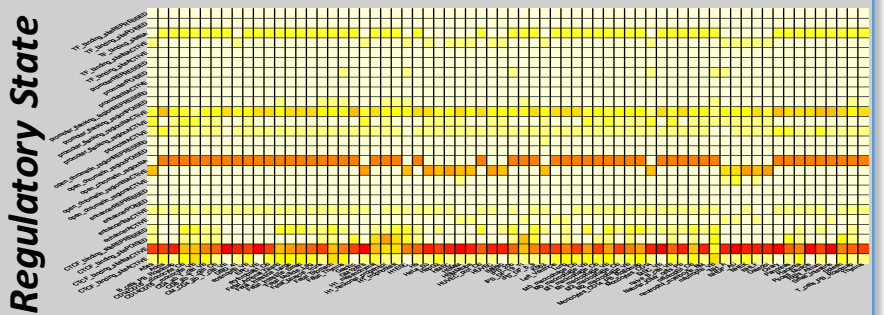
Evolution



Genomic Disorders



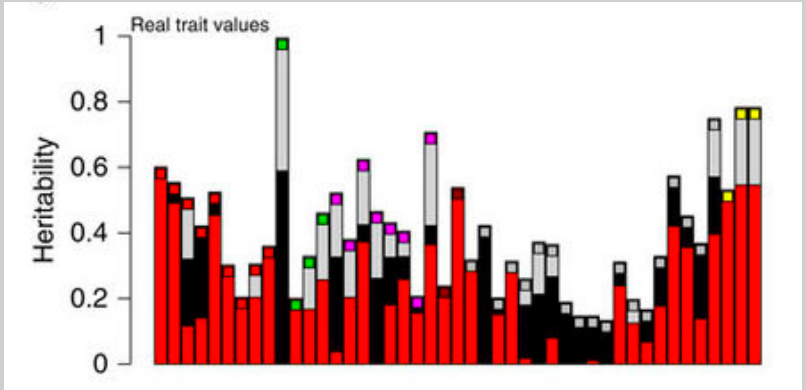
Impact on regulation



ENCODE

GTEx

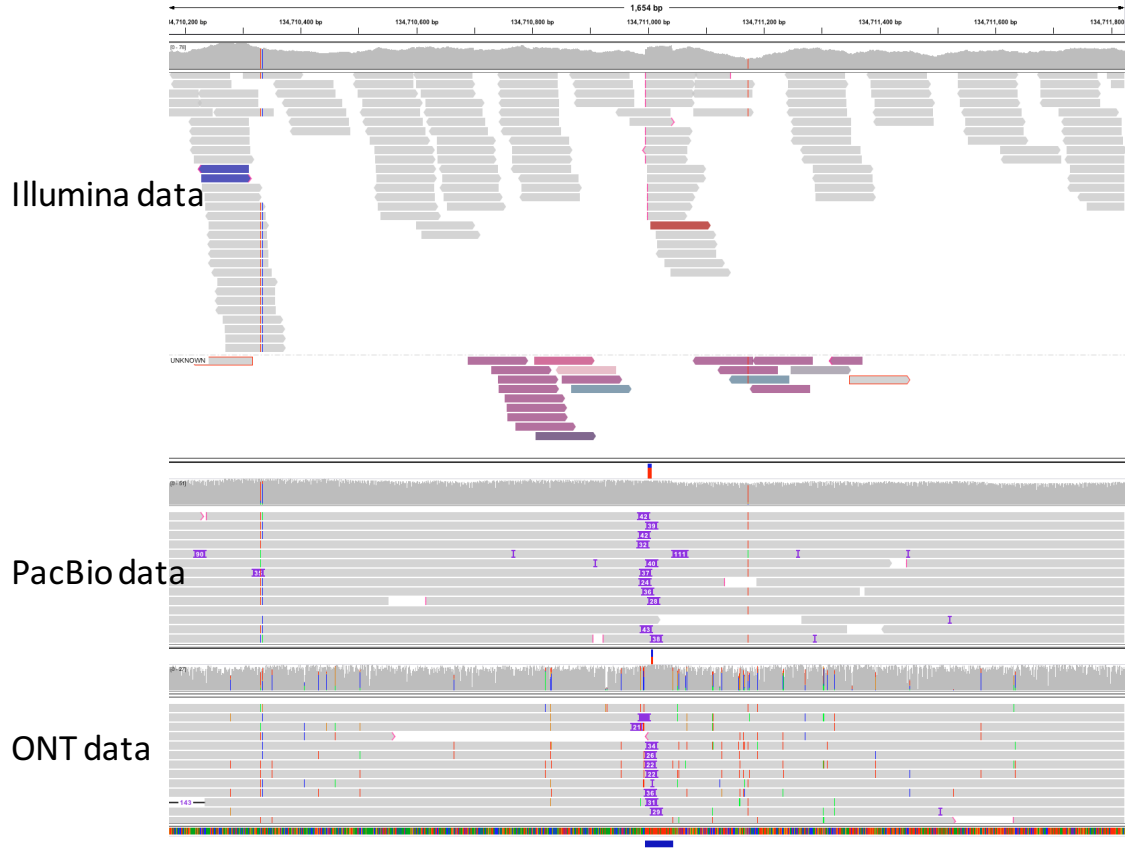
Impact on phenotypes



Remaining Challenges for SVs calling

1. Accuracy of the calls
    1. False positives
    2. False negatives

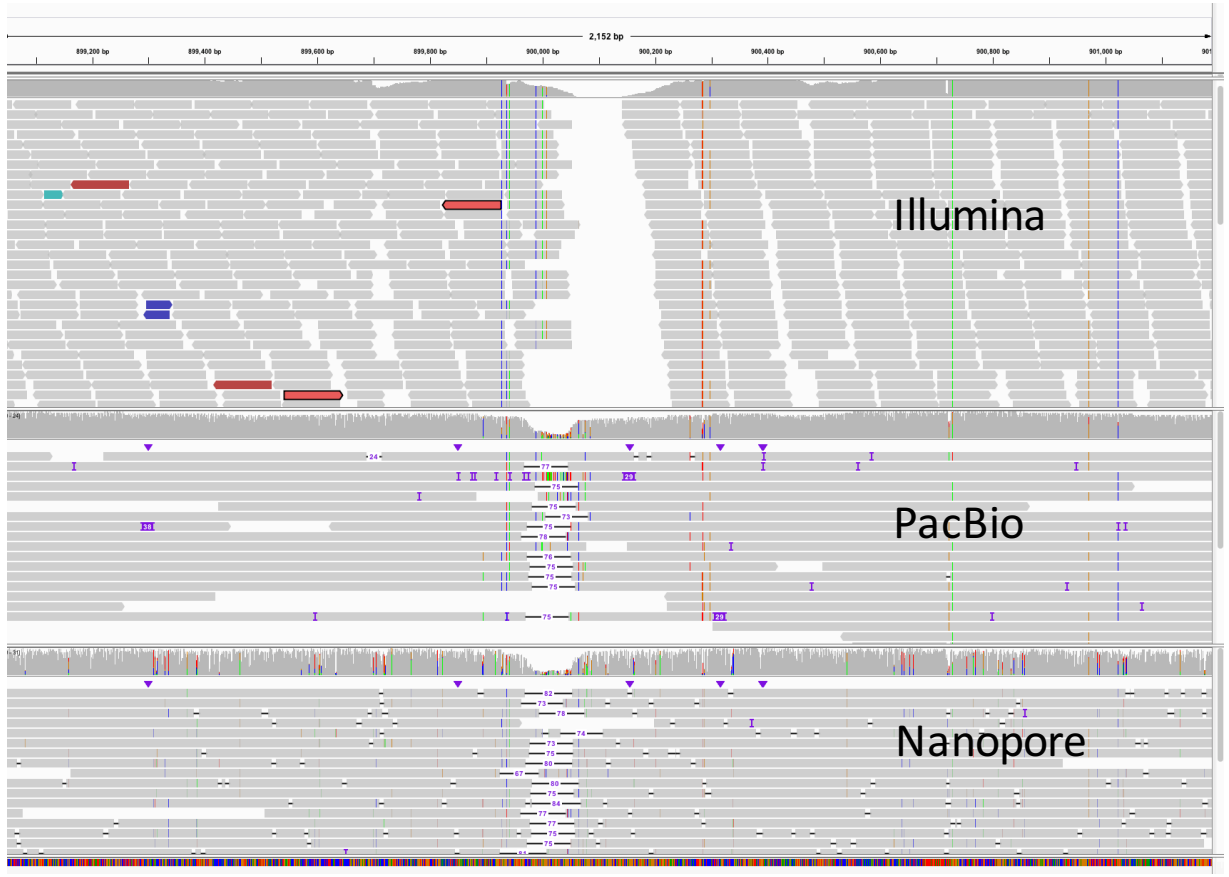
2. Functional interpretation?
    1. Population frequencies/ Curation



Remaining Challenges for SVs calling

1. Accuracy of the calls
    1. False positives
    2. False negatives

2. Functional interpretation?
    1. Population frequencies/ Curation



How to call SV in routine scans?

SV genotyping

- Advantages
  - Low/no false positives
  - False negatives ??
  - Focus on variants that are know to have an impact.

- Disadvantages
  - We cannot find novel SVs

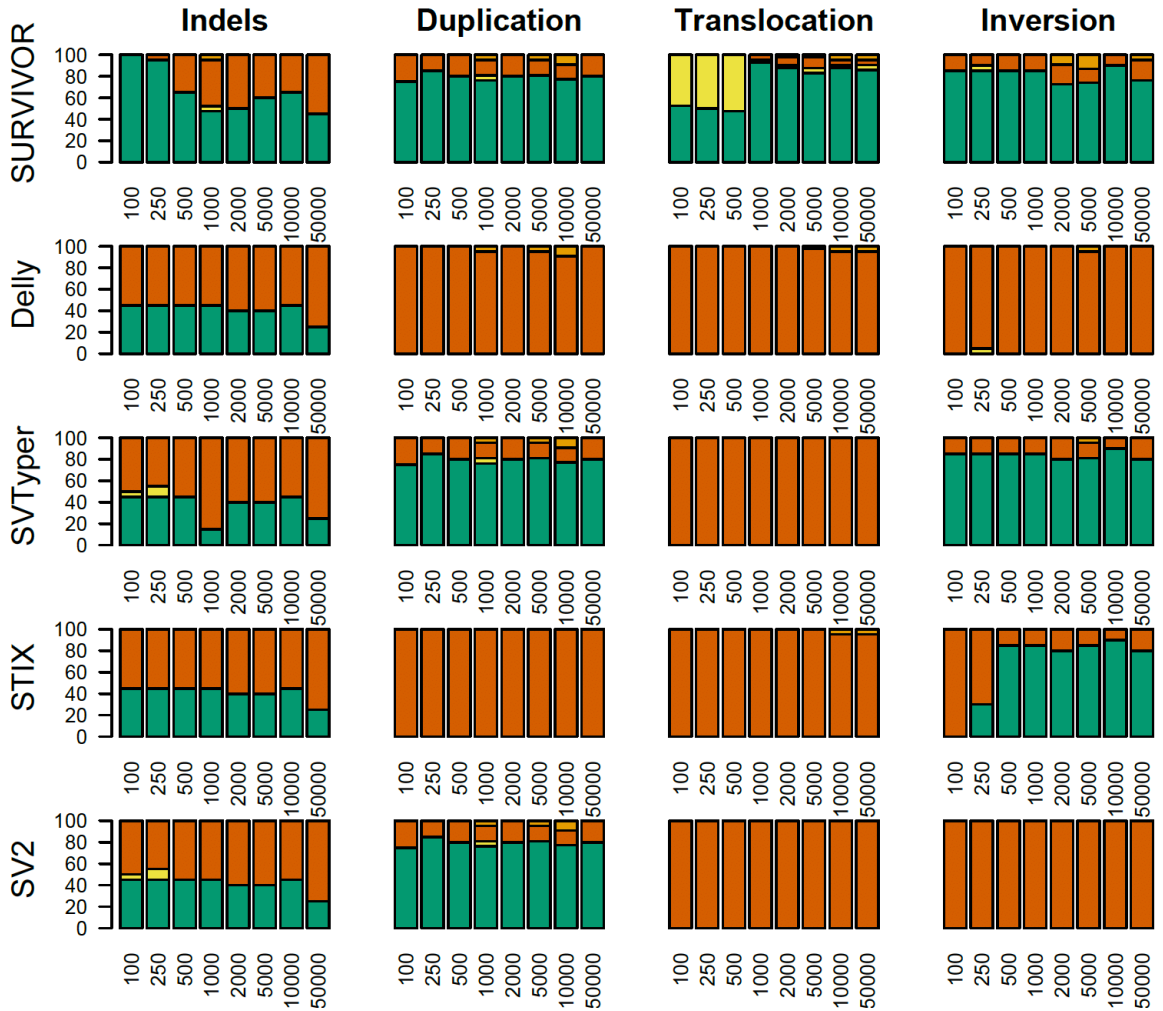
Varuna Chander

Approaches

- DELLY: SV caller that also supports genotyping
- STIX: SV genotyper
- SVTyper: SV genotyper
- SV2: SV genotyper

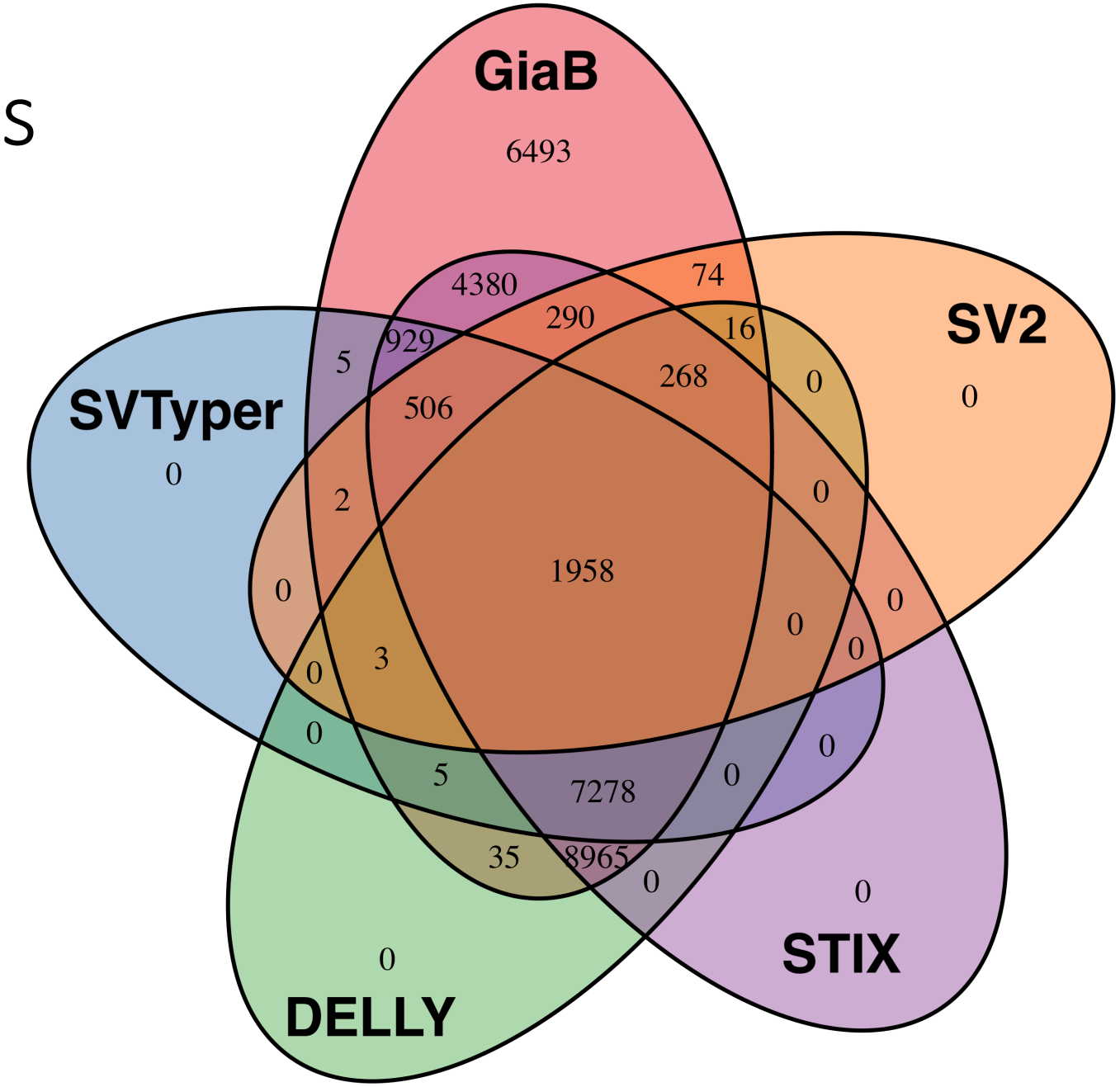
Simulated data

1. We simulated SVs of different types and sizes
2. Called SVs with Delly, Manta and Lumpy
3. Merged calls with SURVIVOR
4. Used the merges as input to the SV genotyper
5. Evaluated their results for SV that they support.



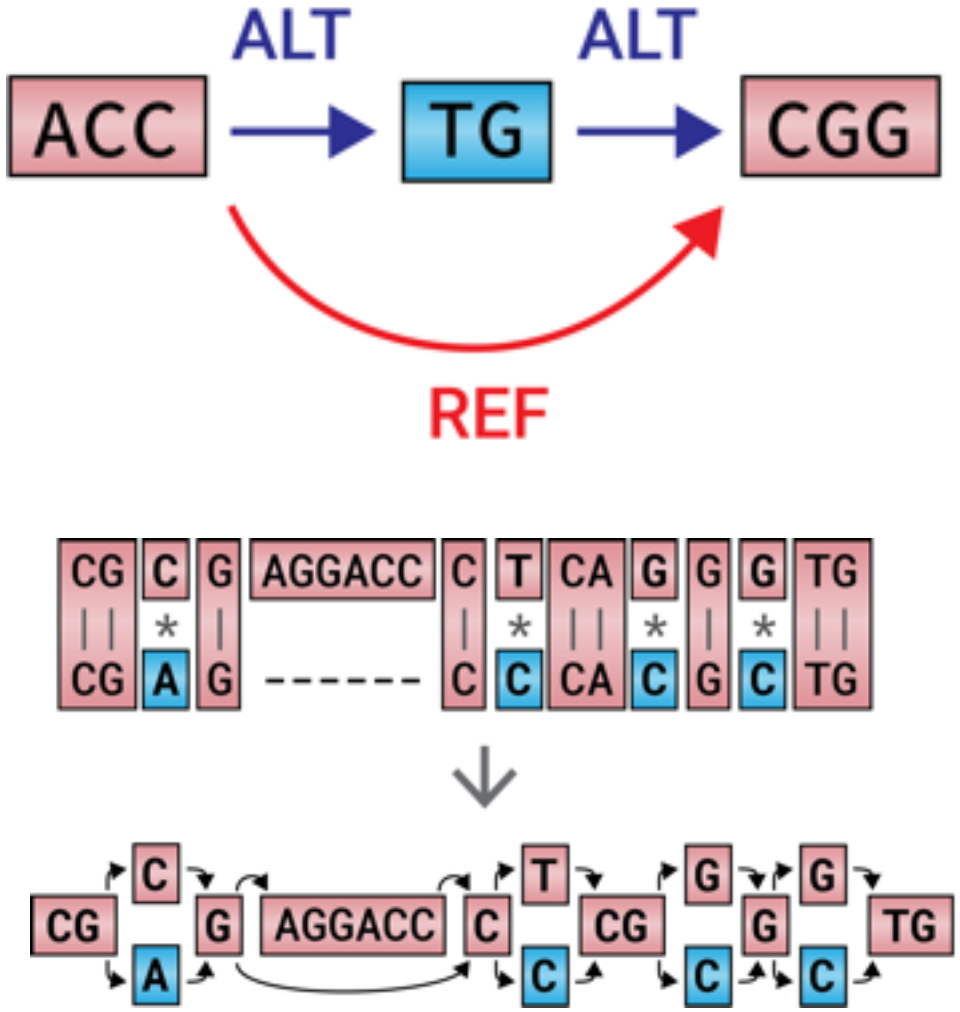
Giab v0.5.0 deletions

- Most of the genotyper only handle the DEL
- Constrain on the input format/field
- Lack of sensitivity



Paragraph

- Graph based SV genotyper
- GiaB all types:
  - Sensitivity: 82%
  - Precision: 99%
  - GT concordance: 80%
- Available:
  github.com/Illumina/paragraph



P Krusche et al (in preparation)

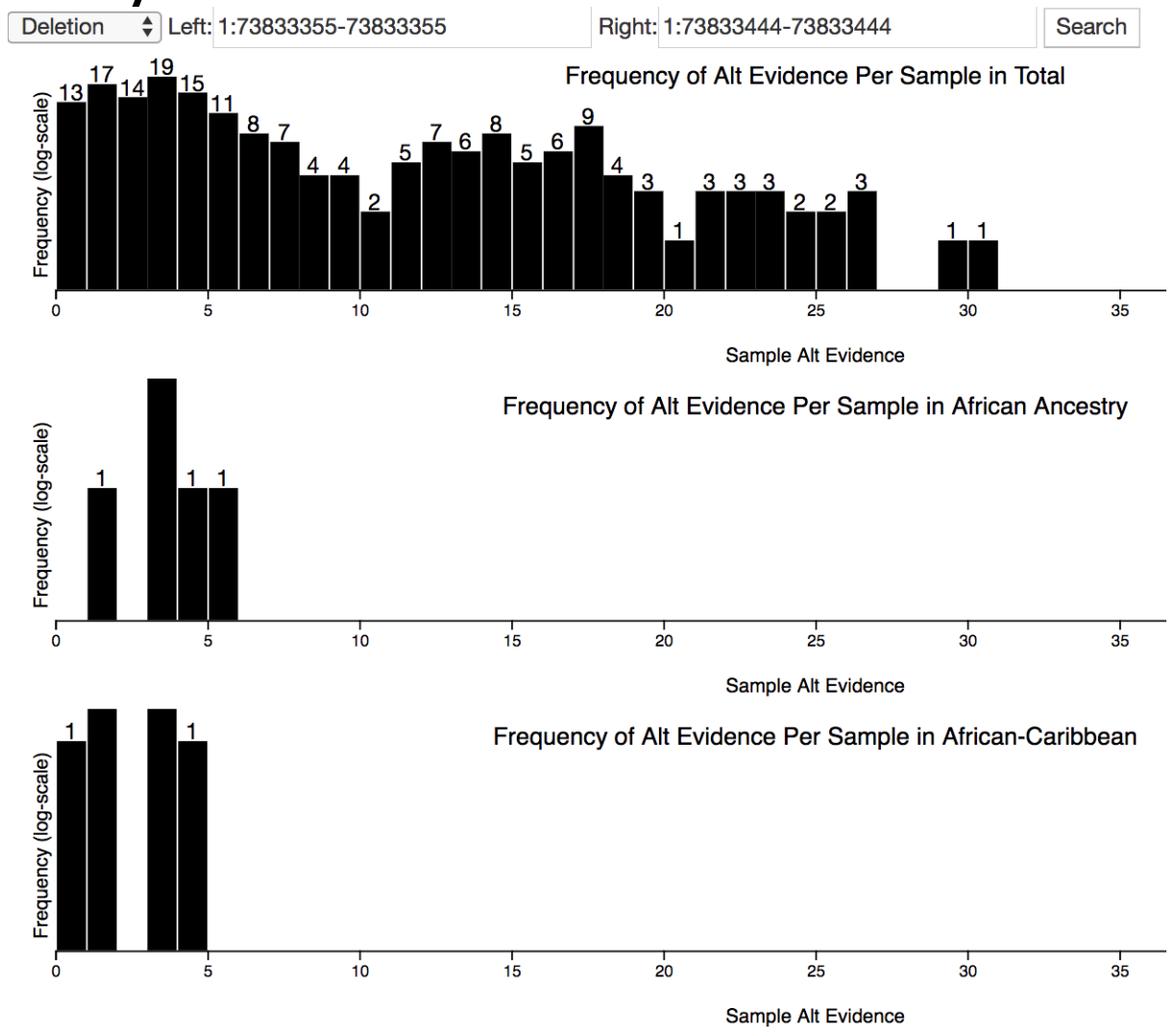
Remaining Challenges for SVs calling

1. Accuracy of the calls
    1. False positives
    2. False negatives

2. Functional interpretation?
    1. Population frequencies/ Curation

STIX: Population frequency

- Online framework to annotate your SVs with allele frequencies.
  - ~0.1 sec / SV
- Storing informative reads
  - (0.18% of BAM)
- Currently ~9000 samples
  - Multiple ethnicities



Layer et al. (in preparation)

Acknowledgments

Varuna Chander
William Salerno
Richard Gibbs

University of Colorado Boulder

Ryan Layer

illumina®

Peter Krusche
Sai Chen,
Mike Eberle

National Human Genome Research Institute