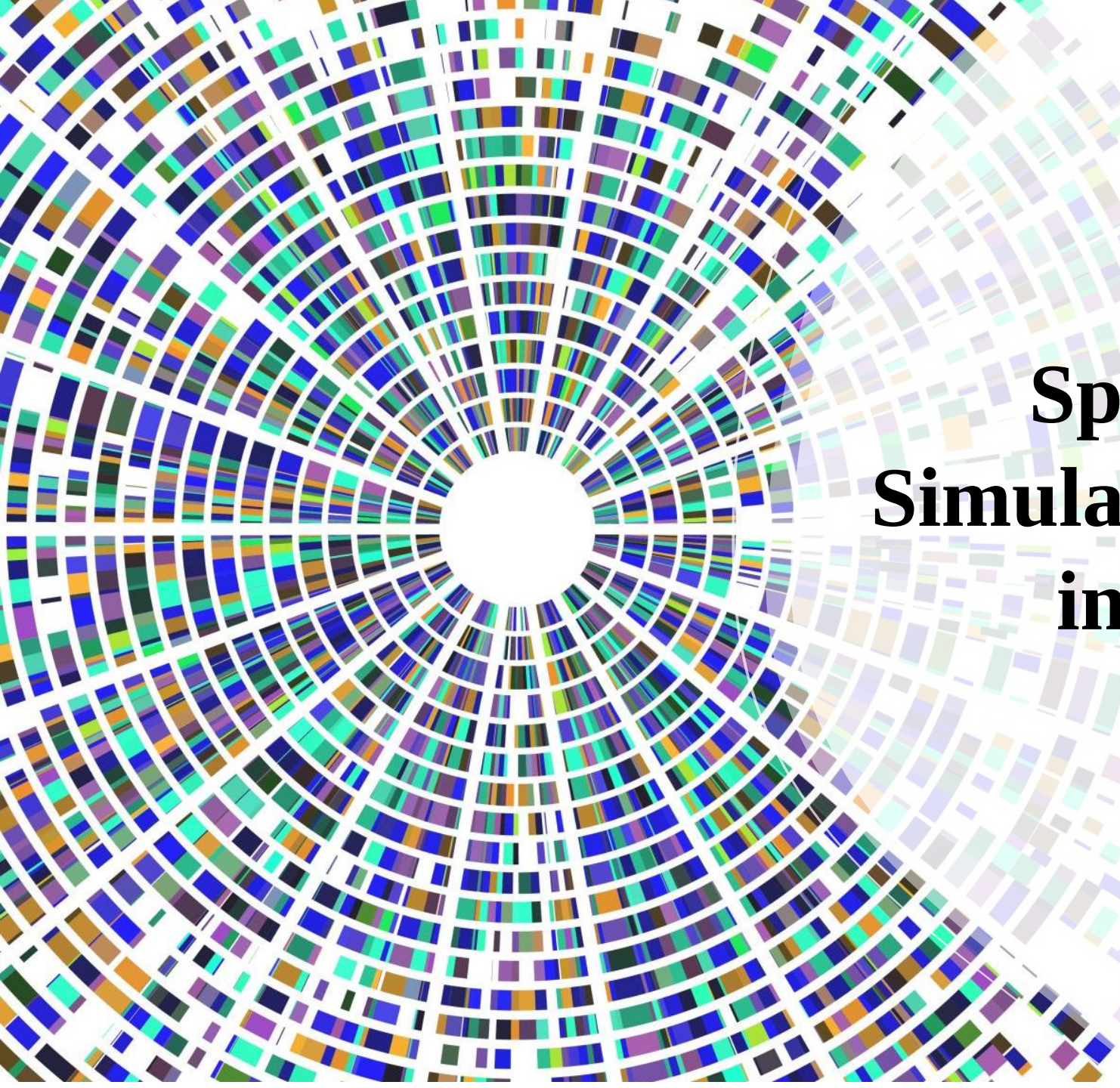


# Group 2 :
# SpikeVar & TykeVar
# Simulation of Mosaic Variants
# in Sequencing Data

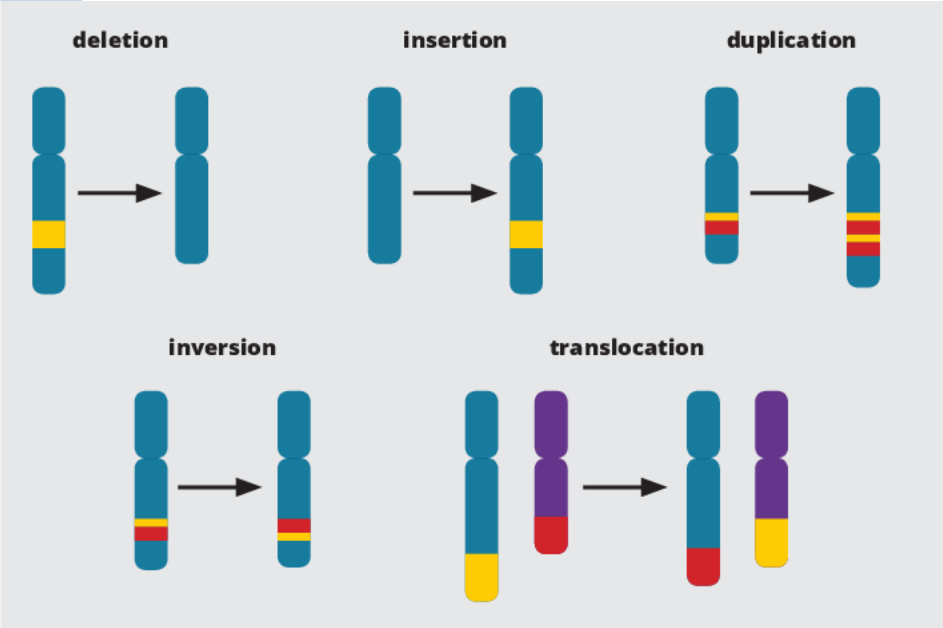

**Farhang Jaryani,**

**Fritz Sedlazeck, Lab**

**Human Genome Sequencing Center**

**Baylor college of Medicine**

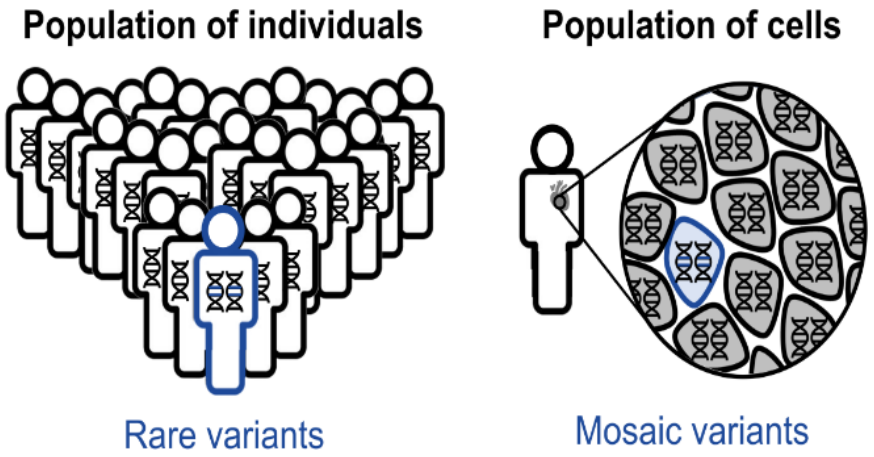
# Structural Variants




| Reference | 1 | 2 | 3 | | | |
|---|---|---|---|---|---|---|
| Insertion | 1 | 2 | 5 | 3 | | |
| Deletion | 1 | 3 | | | | |
| Inversion | 1 | 3 | 2 | | | |
| Copy Number Variation | 1 | 1 | 1 | 1 | 2 | 3 |
| Tandem Duplication | 1 | 1 | 2 | 3 | | |
| Dispersed Duplication | 1 | 2 | 1 | 3 | | |
| Mobile Element Insertion | 1 | 2 | Mobile Element | 3 | | |
| Translocation | 1 | | | | | |
| | 10 | 11 | 12 | 2 | 3 | |


HUNTING STRUCTURAL VARIANTS , Luke Hickey
https://api.semanticscholar.org/CorpusID:204805198

# Mosaic Variants




## Mosaic Variants:

- ✓ **Potentially associated with disease**
- ✓ **Mixed in sequence data**
- ✓ **Requires mosaic variant callers to detect**

THE OHIO STATE UNIVERSITY
COLLEGE OF MEDICINE

Home Profiles Research units **Research output** Projects

### Disease-associated mosaic variation in clinical exome sequencing: a two-year pediatric tertiary care experience

Cecelia R. Miller, Kristy Lee, Ruthann B. Pfau, Shalini C. Reshmi, Donald J. Corsmeier, Sayaka Hashimoto, Ashita Dave-Wala, Vijayakumar Jayaraman, Daniel Koboldt, Theodora Matthews, Danielle Mouhlas, Maggie Stein, Aimee McKinney, Tom Grossman, Benjamin J. Kelly, Peter White, Vincent Magrini, Richard K. Wilson, Elaine R. Mardis, Catherine E. Cottrell

Comprehensive Cancer Center, Comprehensive Cancer Center - Innate Immunity, Pathology, Pediatrics, Center for Clinical and Translational Science, Comprehensive Cancer Center - Molecular Biology and Cancer Genetics, Comprehensive Cancer Center - Experimental Therapeutics

*Research output: Contribution to journal › Article › peer-review*

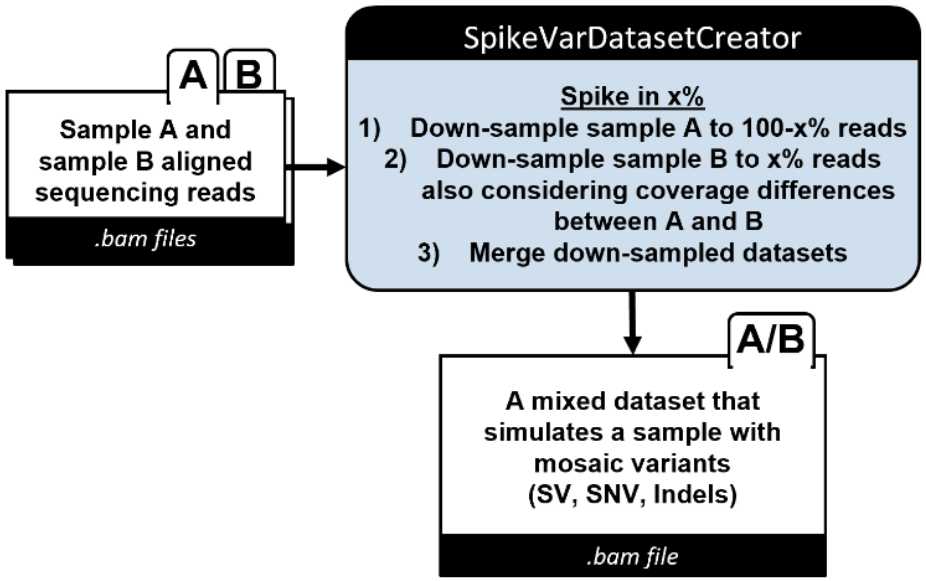
Disease-associated mosaic variation in clinical exome sequencing: a two-year pediatric tertiary care experience. Cold Spring Harb Mol Case Stud. 2020;6(3). Epub 20200612. doi: 10.1101/mcs.a005231. PubMed PMID: 32371413; PMCID: PMC7304353.

# *Spike in Known Exogenous Variants*

## ❖ *Introducing User-Defined Spiking in Sequencing Data*

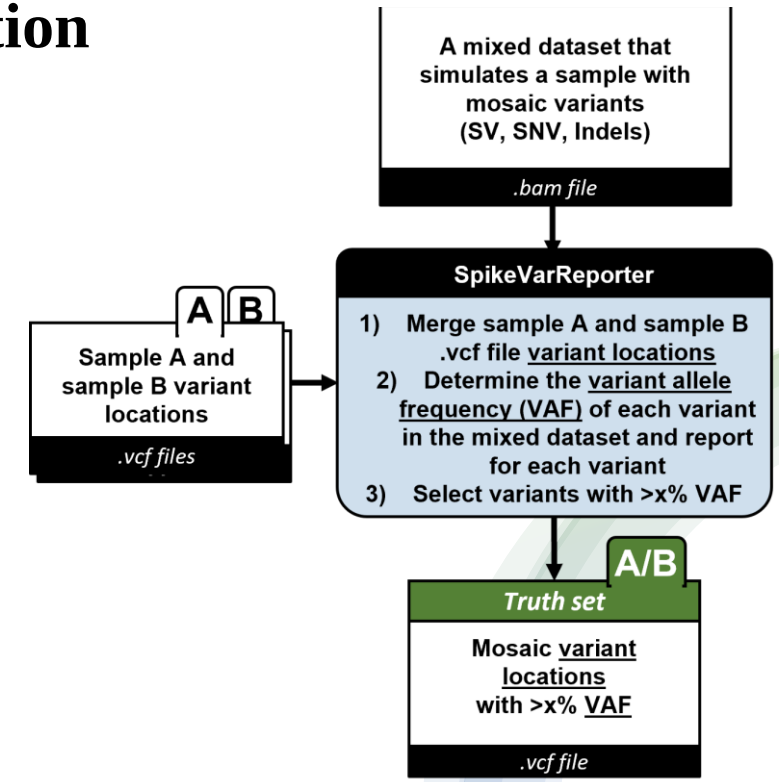
**1) SpikeVar DatabaseCreator –**

**Generate Spiked-in Dataset**

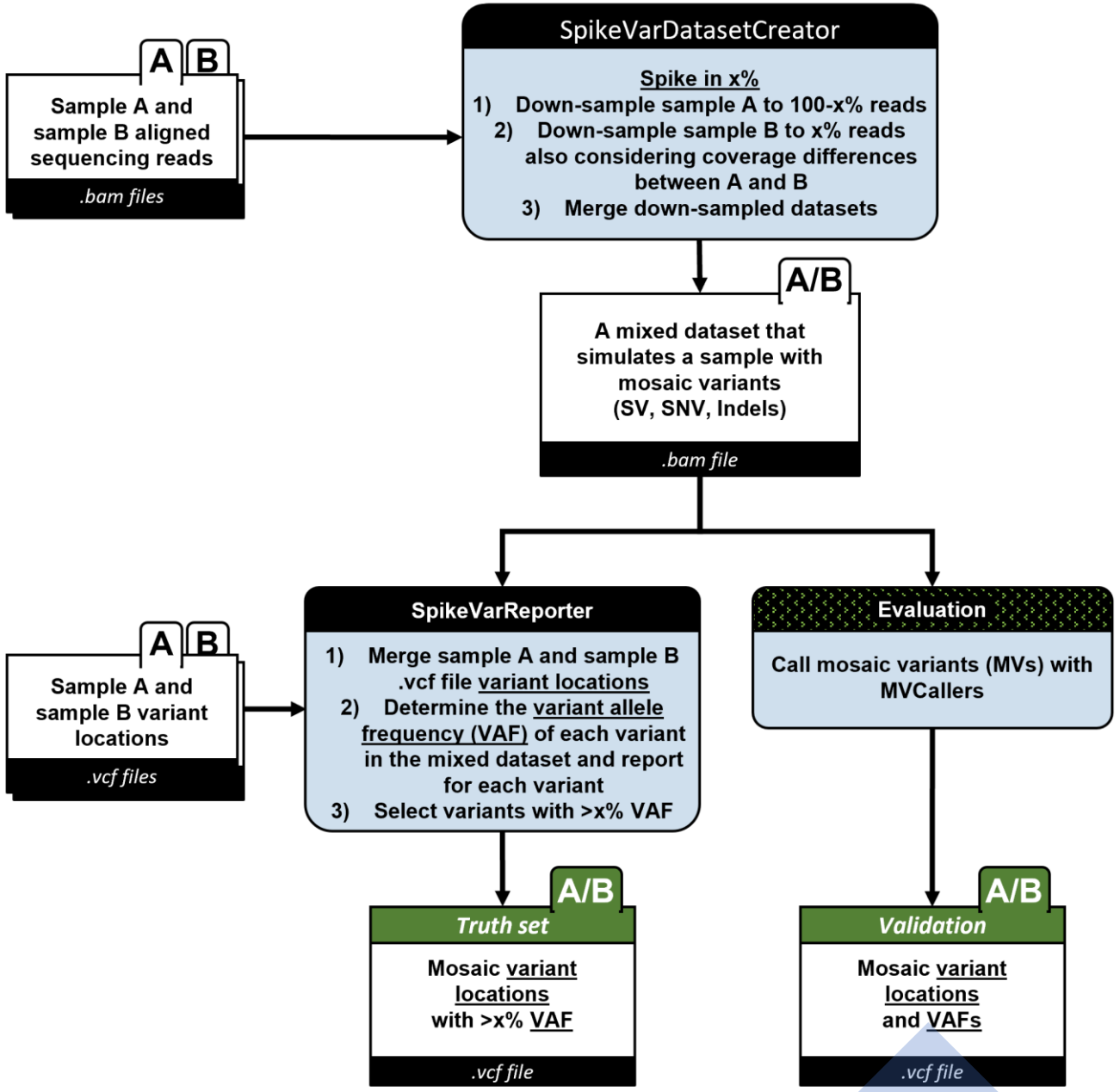



**2) SpikeVarReporter –**

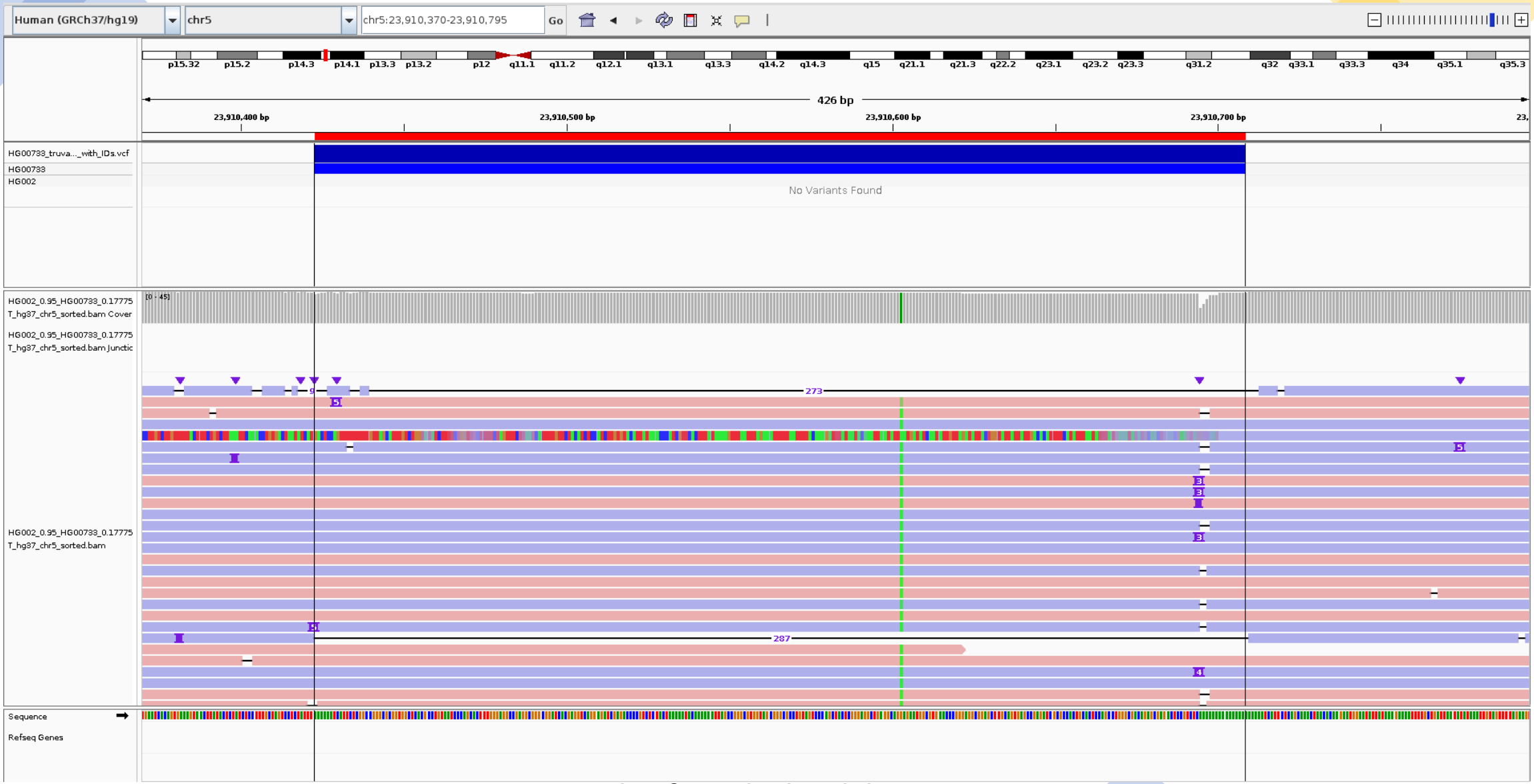
**Filter Reads After Variant Allele Frequency Recalculation**



# *Sp*ike *i*n *K*nown *E*xogenous *Var*iants

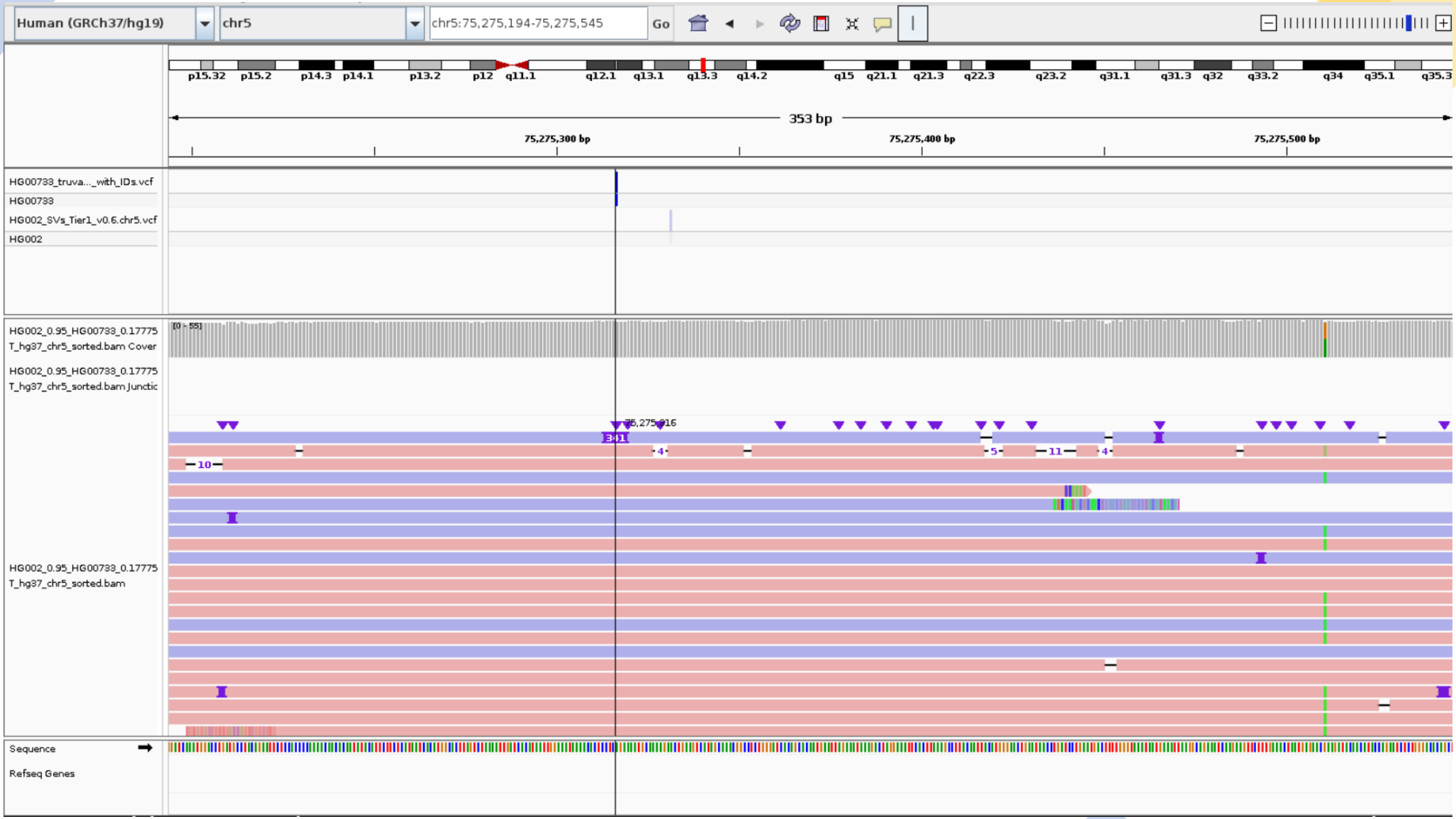


# *Spike in Known Exogenous Variants*




**Example of a spiked in deletion.**

# *Sp*ike *i*n *K*nown *E*xogenous *Var*iants



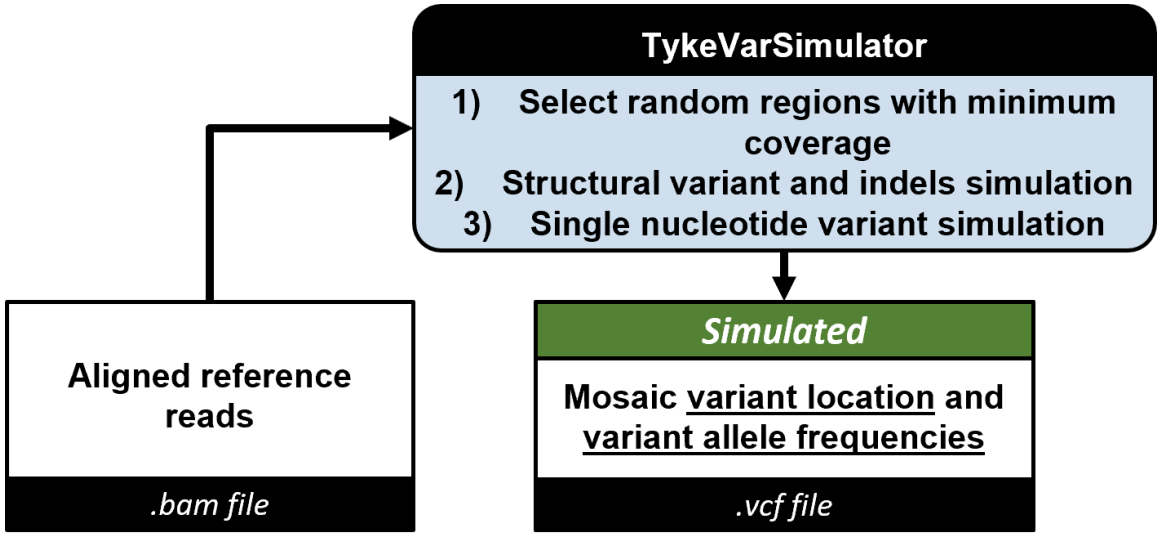

**Example of a spiked in insertion.**

# *T*rack in *Y*our *K*ey *E*ndogenous *Var*iants

- creates a list of random mutations
- modifies a fraction of existing reads to match the user-defined MV frequency

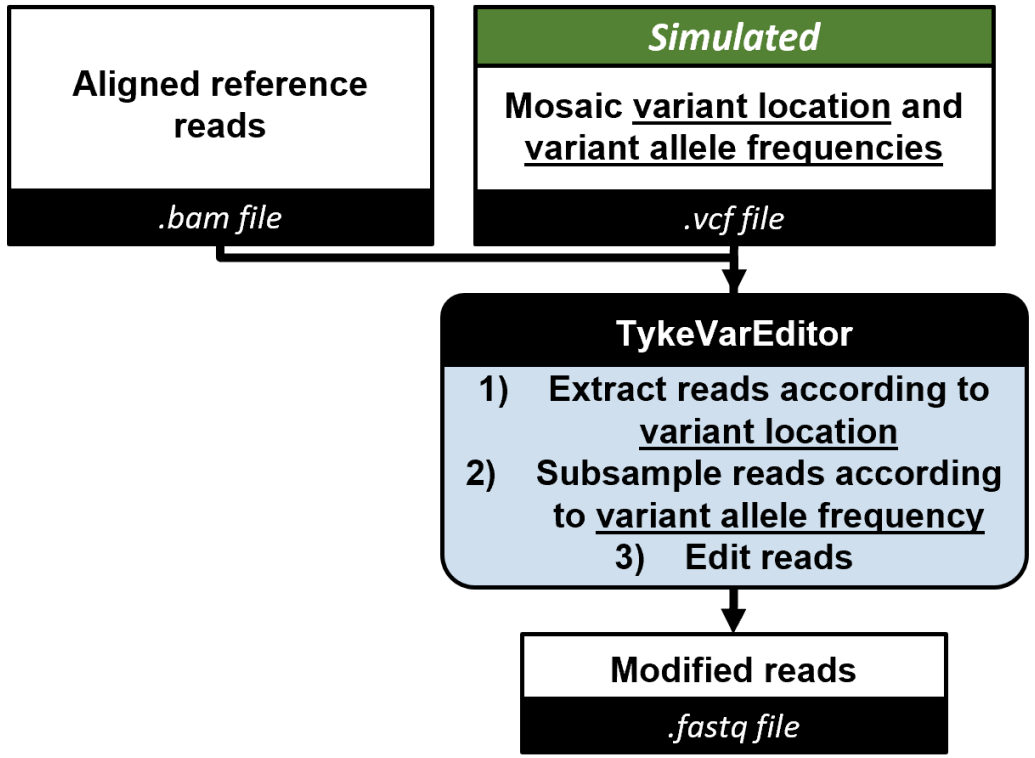
## 1) TykeVarSimulator – Generate Simulated VCF

```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=AD,Number=R,Type=Integer,Description="Read depth for each allele">
##INFO=<ID=AF,Number=1,Type=Float,Description="Allele Frequency">
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT
chr22   32405780        .       A       C       1500    PASS    AF=0.06 GT:AD   0/0:55:4
chr22   27945528        .       G       C       1500    PASS    AF=0.21 GT:AD   0/0:34:10
chr22   32267021        .       T       A       1500    PASS    AF=0.23 GT:AD   0/0:33:11
chr22   27935277        .       T       C       1500    PASS    AF=0.11 GT:AD   0/0:46:6
chr22   30860204        .       G       C       1500    PASS    AF=0.09 GT:AD   0/0:51:6
chr22   46593675        .       C       T       1500    PASS    AF=0.14 GT:AD   0/0:48:8
```

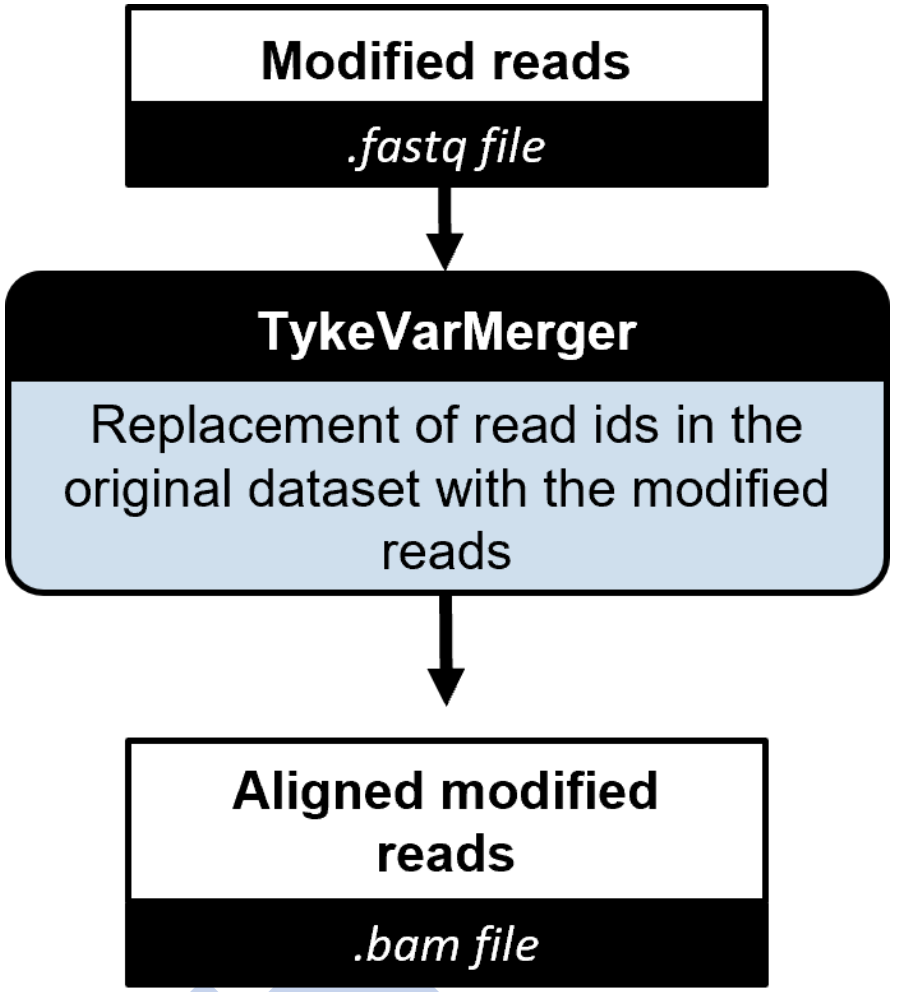


# *T*rack in *Y*our *K*ey *E*ndogenous *Var*iants

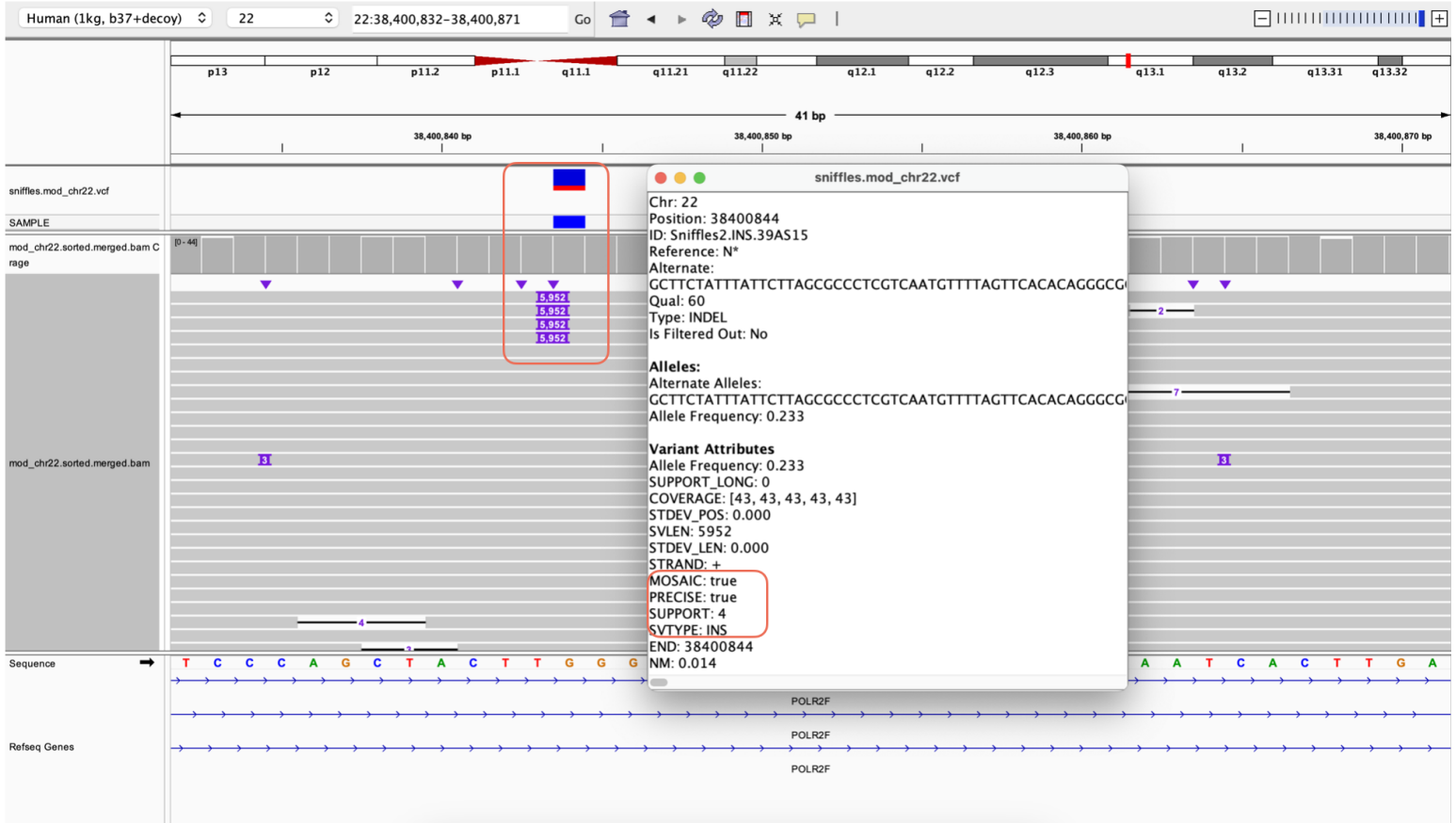
## 2) TykeVarEditor – Generate Edited Reads Based on Simulated VCF

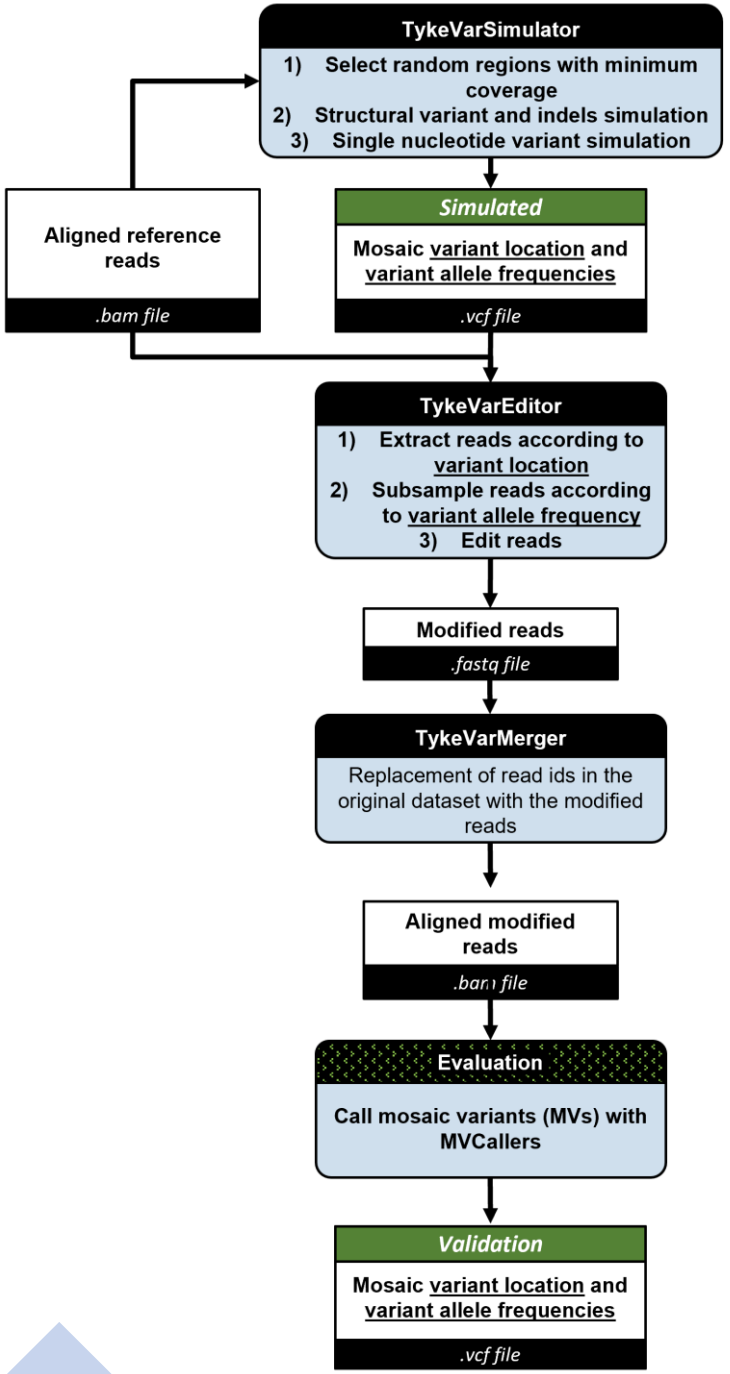



## 3) TykeVarMerger – Re-Align Modified Reads and Merge Them



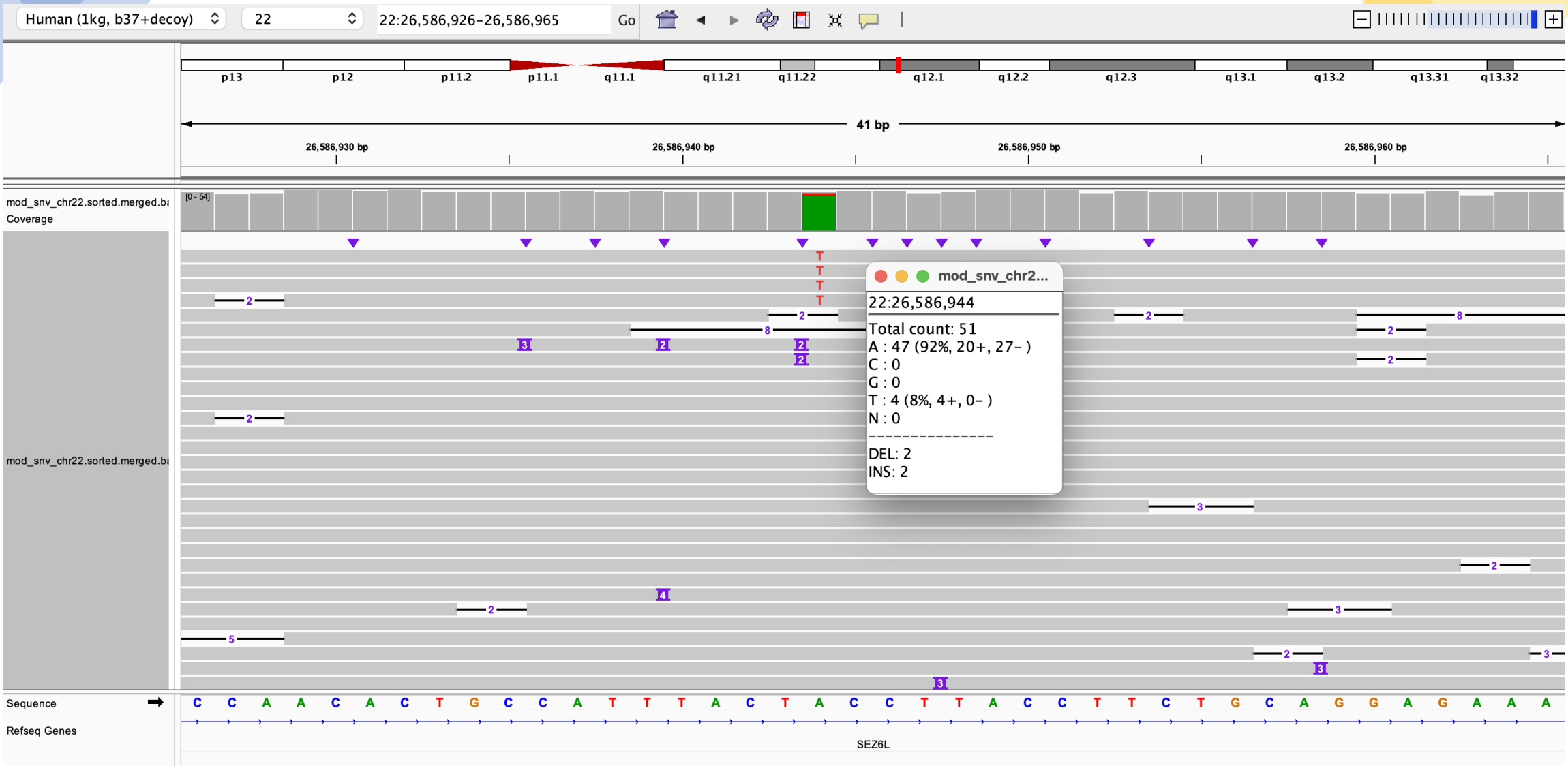
# TykeVar – Creation of Sequencing Data With a Subset of Modified Reads




**A mosaic SV insertion.**

# *T*rack in *Y*our *K*ey *E*ndogenous *Var*iants




**A mosaic SNP.**

# Acknowledgement


## Sedlazeck lab

- Fritz J Sedlazeck
- Luis Paulin
- Xinchang Zheng
- Medhat Mahmoud
- Sairam Behera
- Daniel Paiva Agustinho
- Adam C English
- Michal Izydorczyk
- Yilei Fu