

# Evaluation of Feb 28 benchmark vs v4.2.1 and verkko

Justin Zook, Nate Olson, and NIST team  
3/21/23

# Overall stats improved

- Benchmarking of verkko-dipcall vs TR benchmark within both benchmark regions
  - "TP-base": 3616,
  - "TP-comp": 3627,
  - "FP": 58,
  - "FN": 43,
  - "precision": 0.9842605156037992,
  - "recall": 0.9882481552336704,
  - "f1": 0.9862503046963221,
- Region summary
  - "TP": 2459,
  - "TN": 35894,
  - "FP": 53,
  - "FN": 33,
- Benchmarking of v4.2.1 small variants vs TR benchmark within both benchmark regions (excludes most SVs)
  - "TP-base": 1131,
  - "TP-comp": 1133,
  - "FP": 19,
  - "FN": 7,
  - "precision": 0.9835069444444444,
  - "recall": 0.9938488576449912,
  - "f1": 0.9886508560464089,
- Region summary:
  - "TP": 909,
  - "TN": 7261,
  - "FP": 17,
  - "FN": 6,

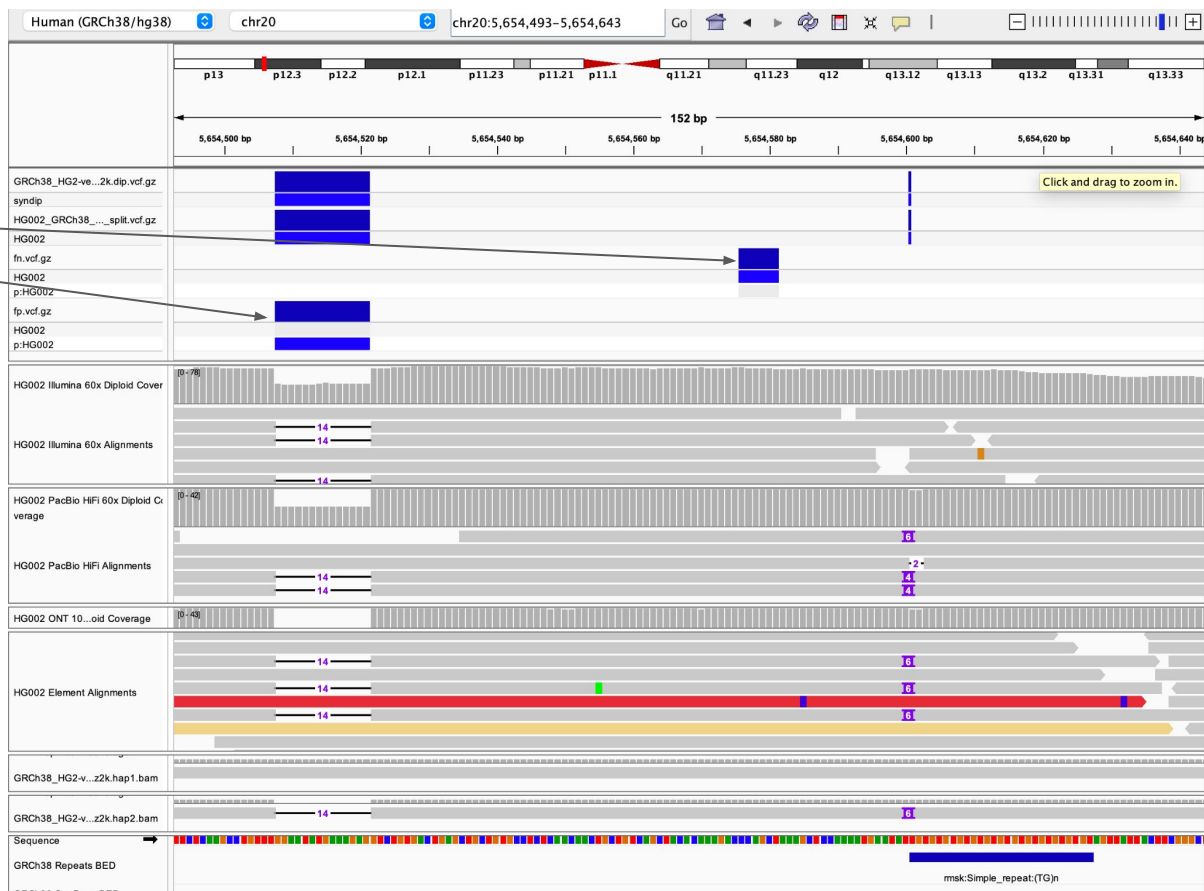
# v4 FP: TR benchmark correctly identifies error in v4 benchmark coming from noisy HiFi reads

v4 calls 6bp del instead of 4bp del



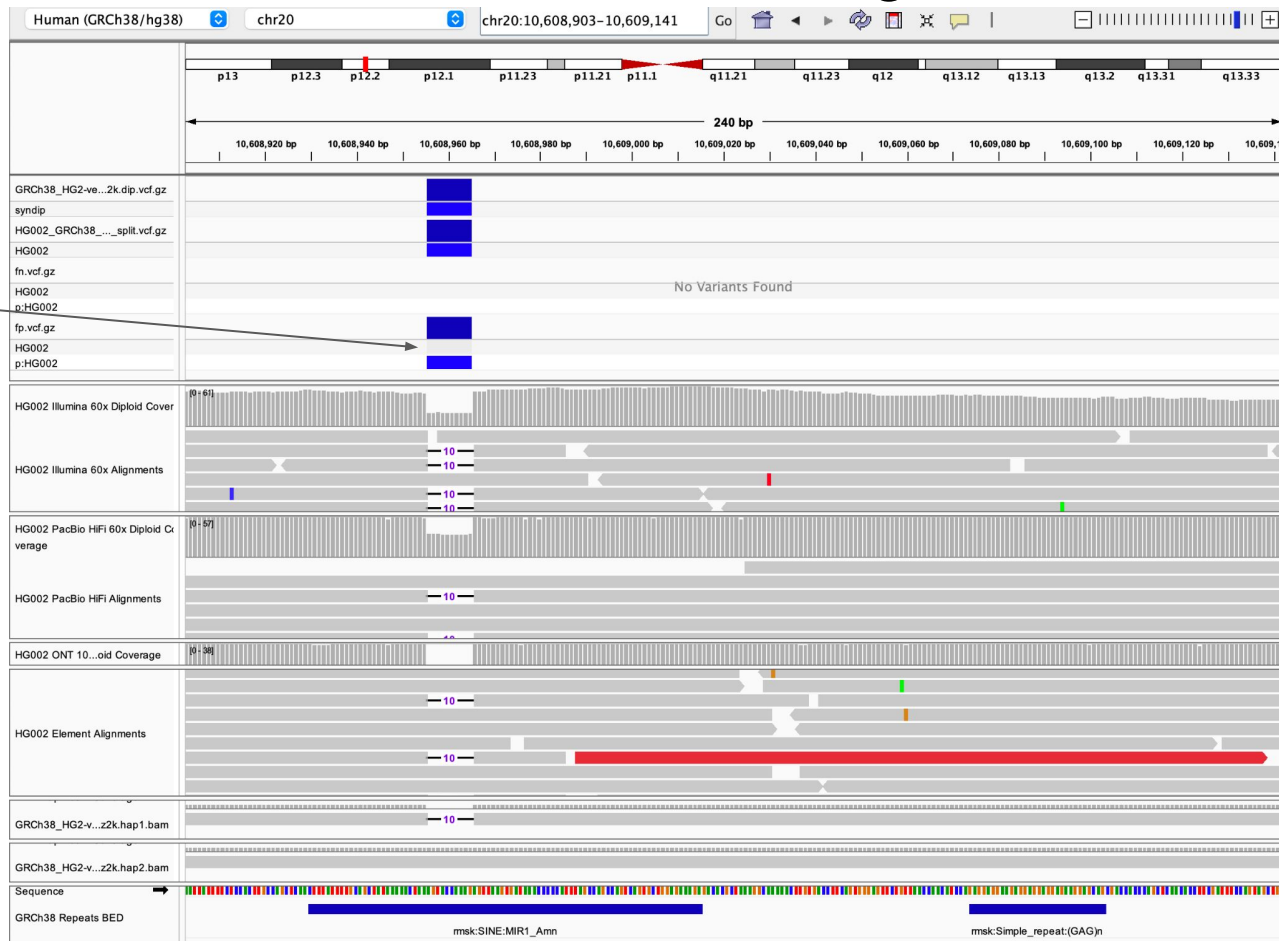
# v4 FP: TR benchmark miscalls del outside TR region

v4 calls correct  
del and TR  
benchmark  
calls incorrect  
del



# v4 FP: TR benchmark misses del outside TR region

v4 calls correct  
del





v4 FN: v4 called incorrect because of different representation and no phasing?

v4 calls  
unphased  
variants

