



Civil and  
Environmental  
Engineering

# High-accuracy mutation calling for bacterial isolates using deep learning

Herui Liao, Arolyn Conwill, Tami Lieberman



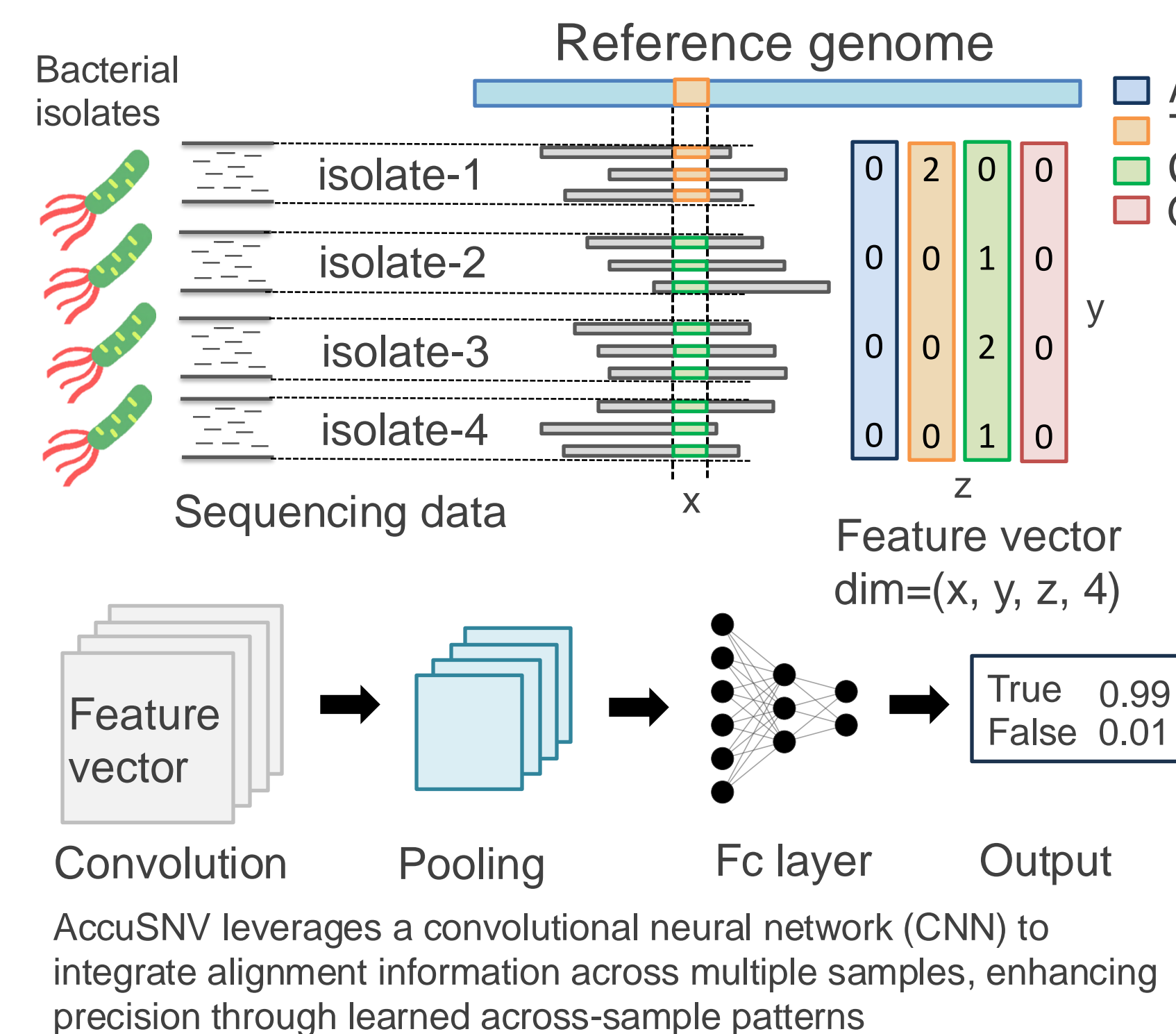
## Problem

Tracking single nucleotide variants (SNVs) - single-letter changes in DNA sequences - is key to studying bacterial evolution, disease outbreaks, and antibiotic resistance.

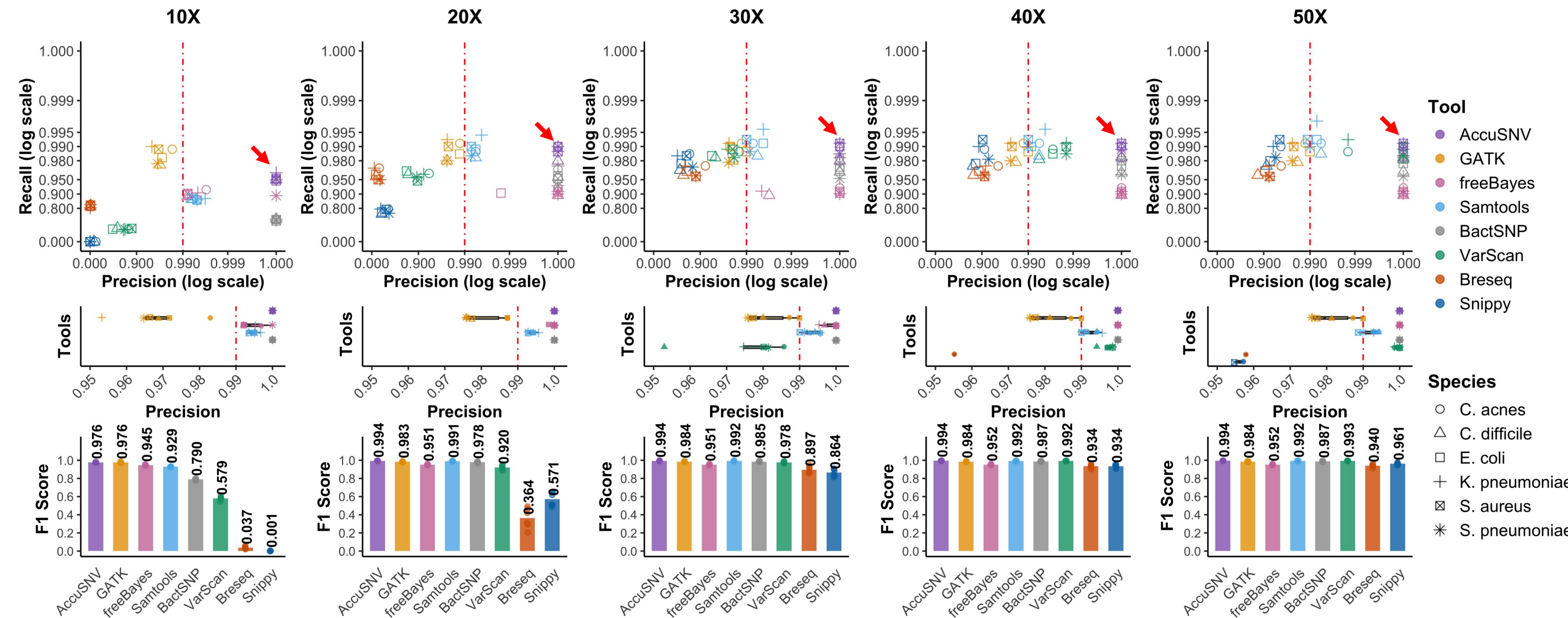
## The challenge

Unlike human genomes, which are 99.9% identical between individuals, bacterial genomes vary widely across strains. This variation leads to more mismatches between sequencing data and the reference genome, challenging the accuracy of existing mutation calling tools. Solutions that maintain high precision while adapting to diverse bacterial genomes are needed.

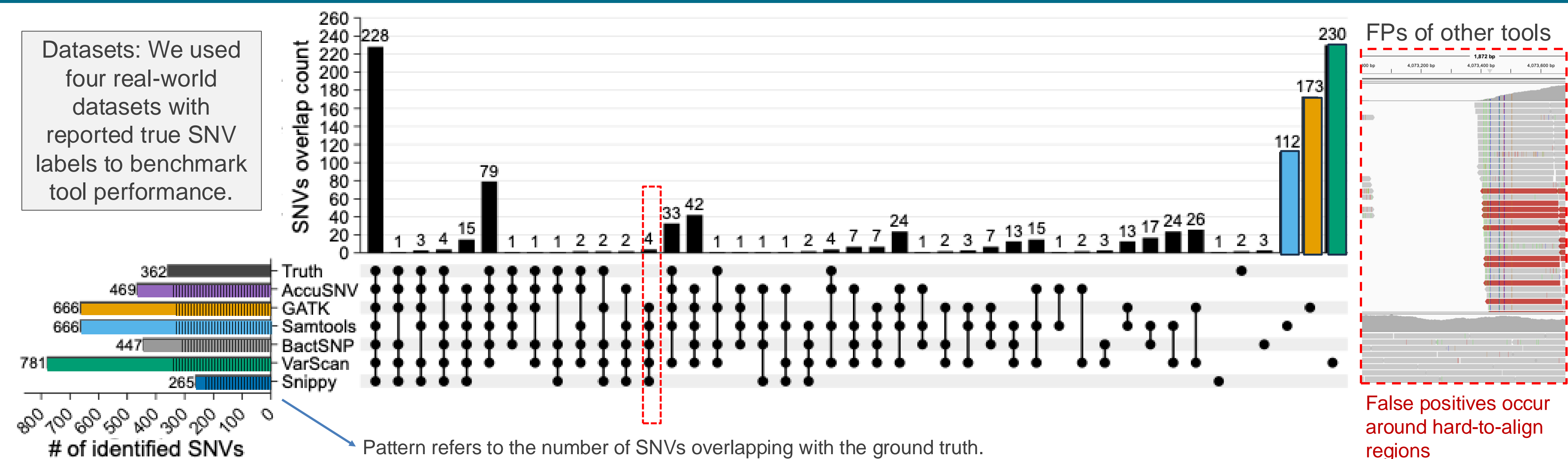
## A deep learning framework that leverages across-sample comparisons



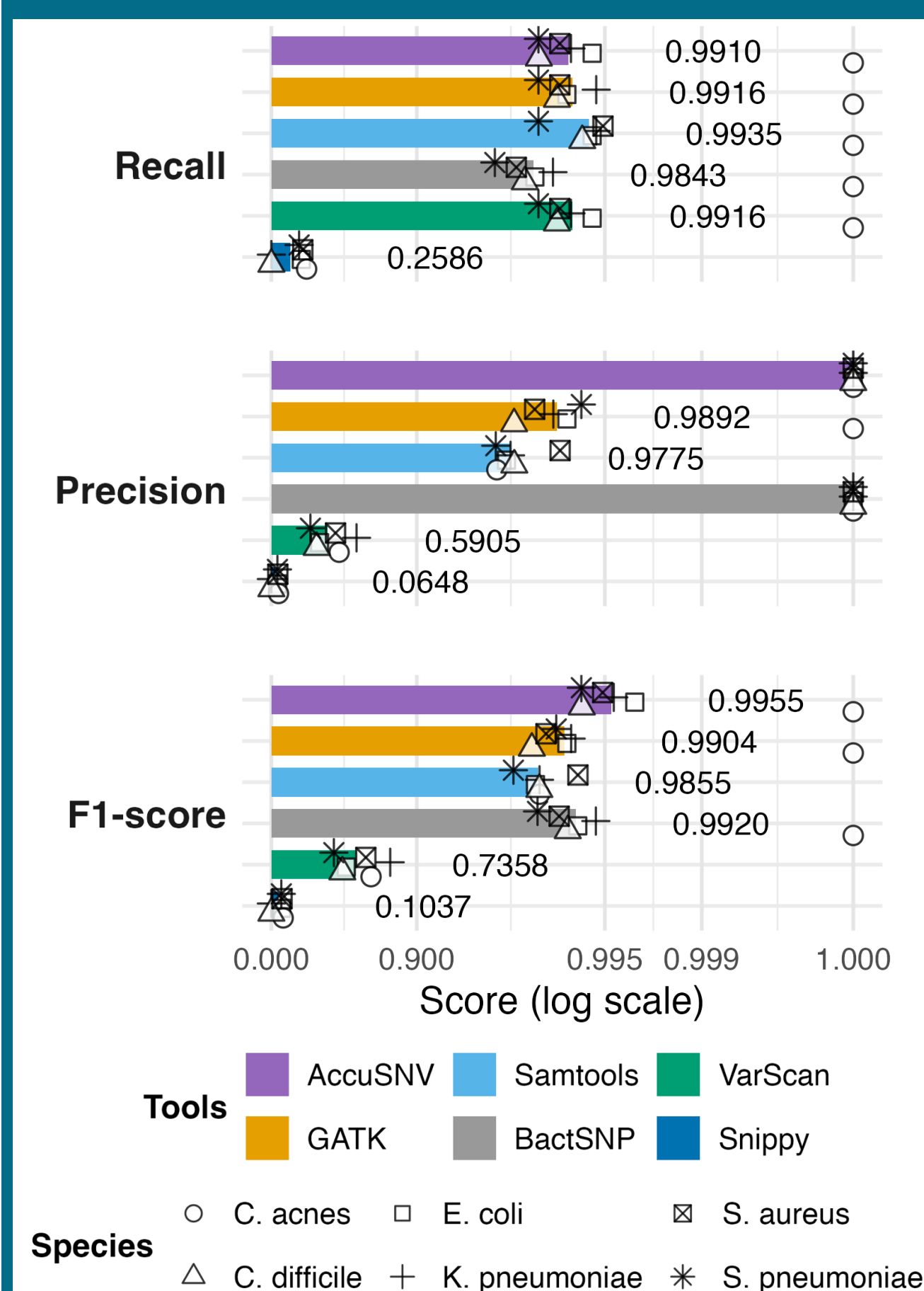
## AccuSNV shows high precision across sequencing depths



## AccuSNV achieves high accuracy on real data (test data)



## AccuSNV achieves high accuracy on large and noisy simulated datasets



Each simulated dataset contains 100 strains with sequencing depths ranging from 15X to 70X. To mimic real-world conditions, the number of mutations per strain was simulated using msprime with a mutation rate of 5e-10.

## Conclusions

- AccuSNV shows higher precision and accuracy than other tools across a range of conditions

- AccuSNV provides comprehensive output, making high-precision SNV-based analysis accessible to all users.

Demo report