

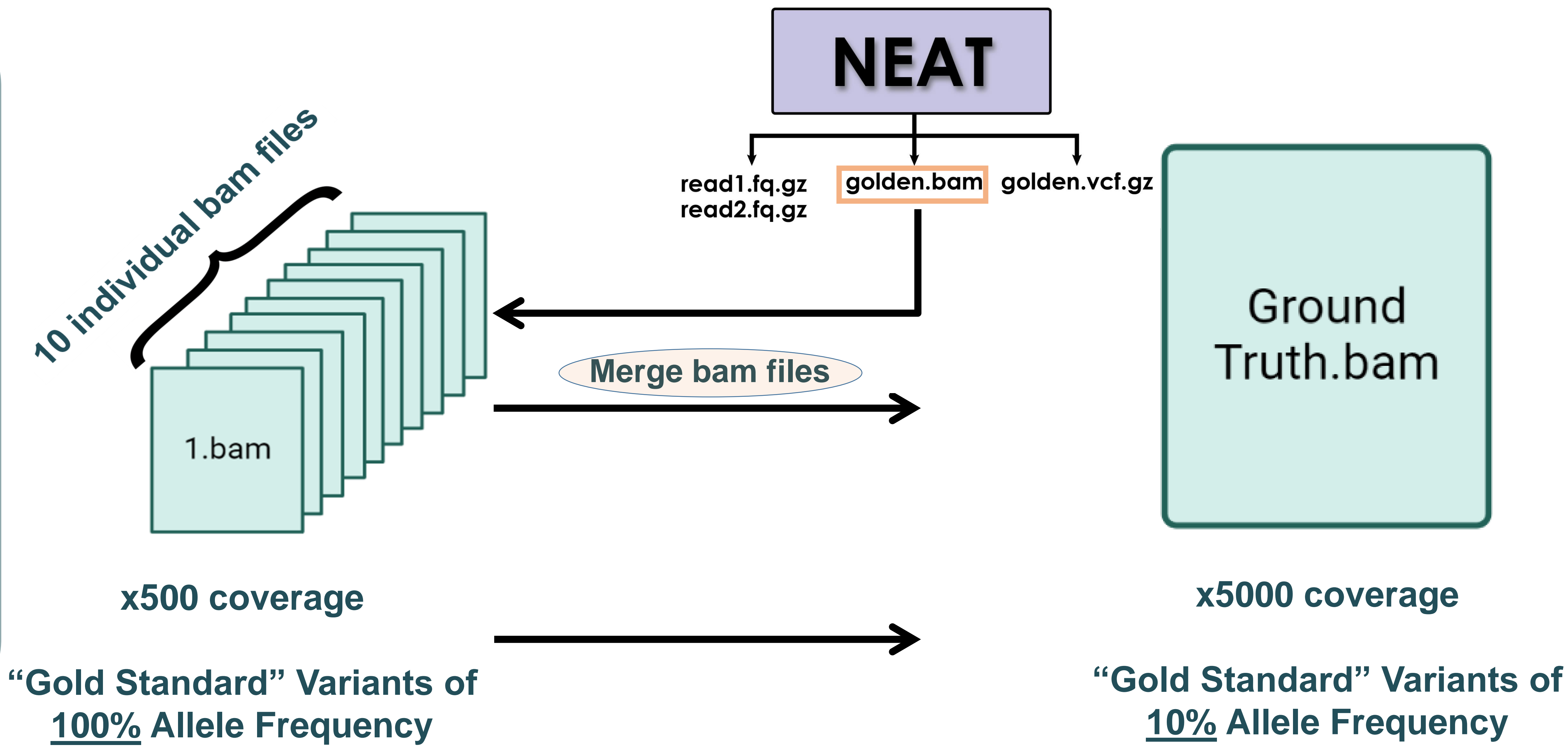
# Synthetic Genomics Data Generation and Evaluation for the Use Case of Benchmarking Somatic Variant Calling Algorithms

Fragkouli Styliani-Christina<sup>1, 2</sup>, Pechlivanis Nikos<sup>1</sup>, Agathangelidis Andreas<sup>2</sup>, Psomopoulos Fotis<sup>1</sup>

<sup>1</sup>Institute of Applied Biosciences, Centre of Research and Technology Hellas, Thessaloniki, Greece  
<sup>2</sup>Department of Biology, National and Kapodistrian University of Athens, Athens 10679, GR

## 1 Synthetic «Gold Standard» Dataset Generation

- Highlights
- Generation of **synthetic genomics** data based on *TP53* gene
  - Define «**Ground Truth**» SNIPs and INDELs in order to **benchmark** somatic variant callers
  - Investigate the impact of variant callers in variants at **low frequencies**



## 2 Benchmarking GATK-Mutect2

