Synthetic Genomics Data Generation And Evaluation For The Use Case Of Benchmarking Somatic Variant Calling Algorithms

Styliani – Christina Fragkouli

PhD Candidate

sfragkoul@certh.gr
sfragkoul
@scfragkoul

National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITUTE OF APPLIED BIOSCIENCES
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Funded by the European Union

Horizon Europe Programme, Grant Agreement Number 101058573

MOTIVATION

- Variant calling plays an important in identifying genetic lesions.
- In the case of variants at low frequency (≤10%) identification becomes more challenging.
- The challenge that rises is the absence of a ground truth for reliable and consistent identification and benchmarking.

sfragkoul@certh.gr
sfragkoul
@scfragkoul

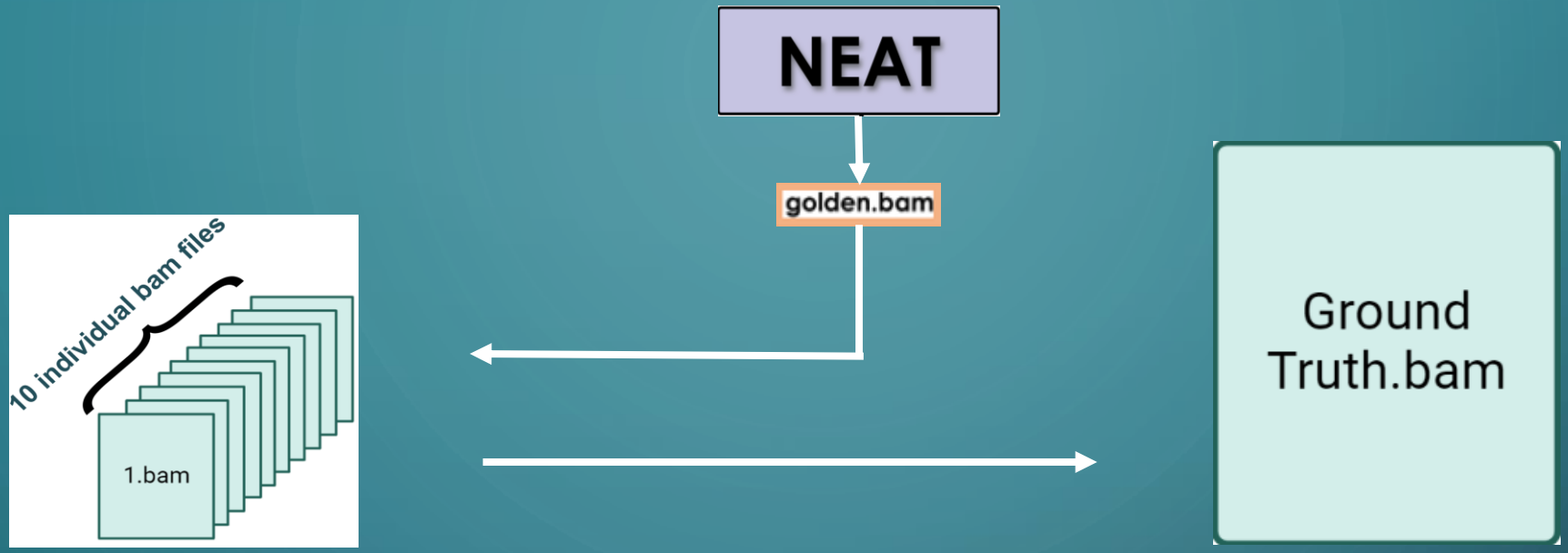
National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITOYTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573

SYNTHETIC «GOLD STANDARD» DATASET GENERATION



x500 coverage

"Gold Standard" Variants of 100% Allele Frequency

x5000 coverage

"Gold Standard" Variants of 10% Allele Frequency

sfragkoul@certh.gr
sfragkoul
@scfragkoul

National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITUTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RECEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573

COMPARING ALGORITHMS

In-silico generated dataset that contains «Ground Truth» SNIPs and INDELs

VS

Results from GATK somatic variant calling algorithm

gatk

Mutect2

sfragkoul@certh.gr
sfragkoul
@scfragkoul

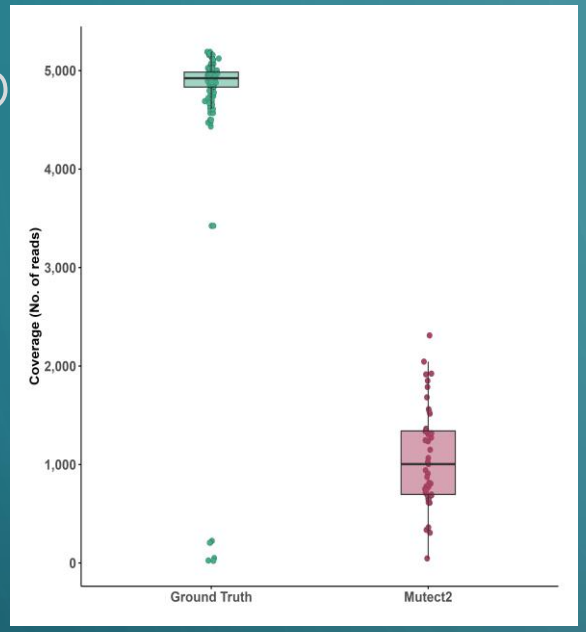
National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITUTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

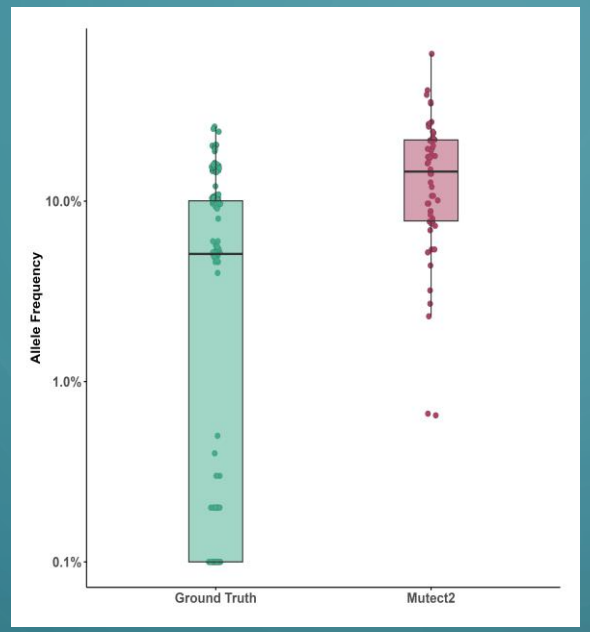
CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573

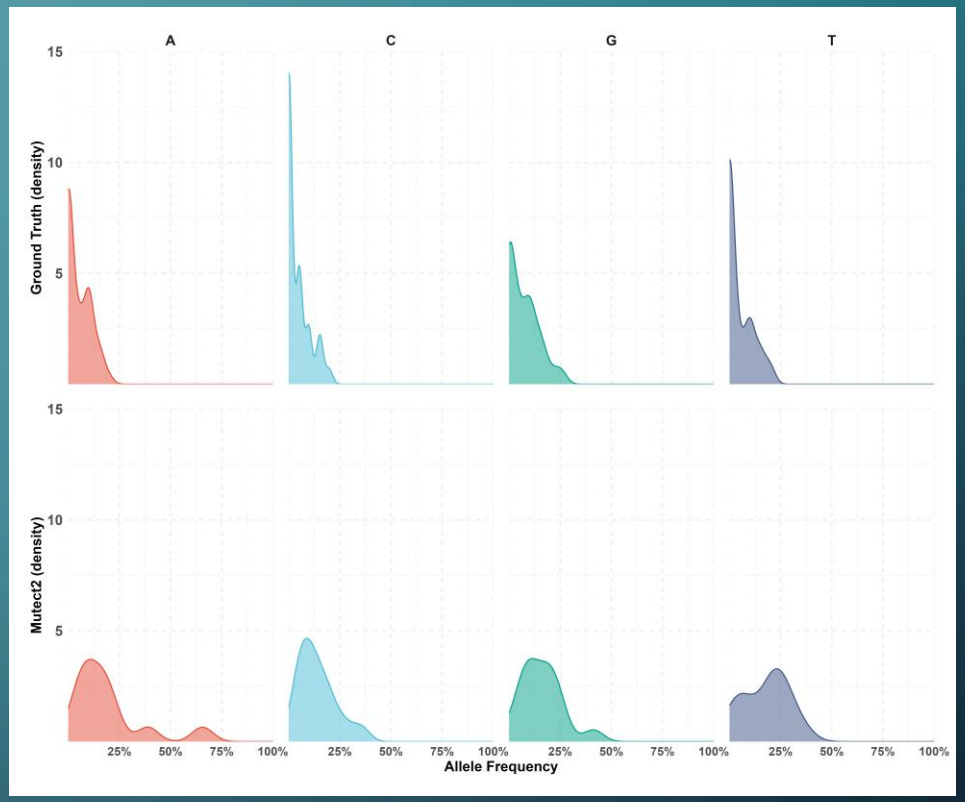
BENCHMARKING GATK-MUTECT2



Down-sampling of coverage of «Ground Truth» Variants



Differences in AF of «Ground Truth» Variants



Variance in AF Density plots of «Ground Truth» Variants per DNA Base

sfragkoul@certh.gr
sfragkoul
@scfragkoul

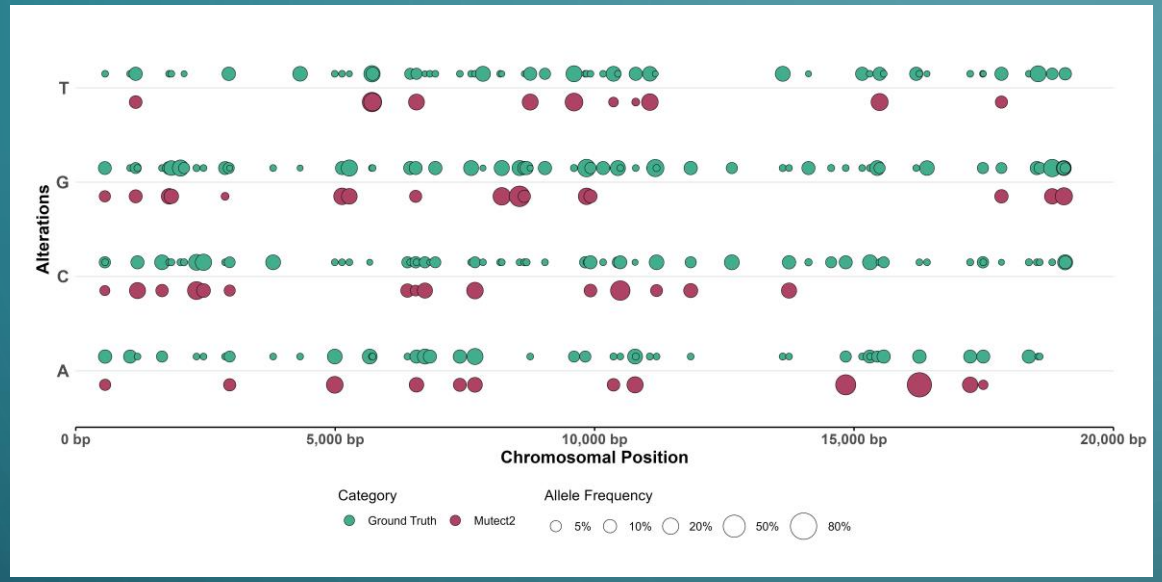
National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITOYTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

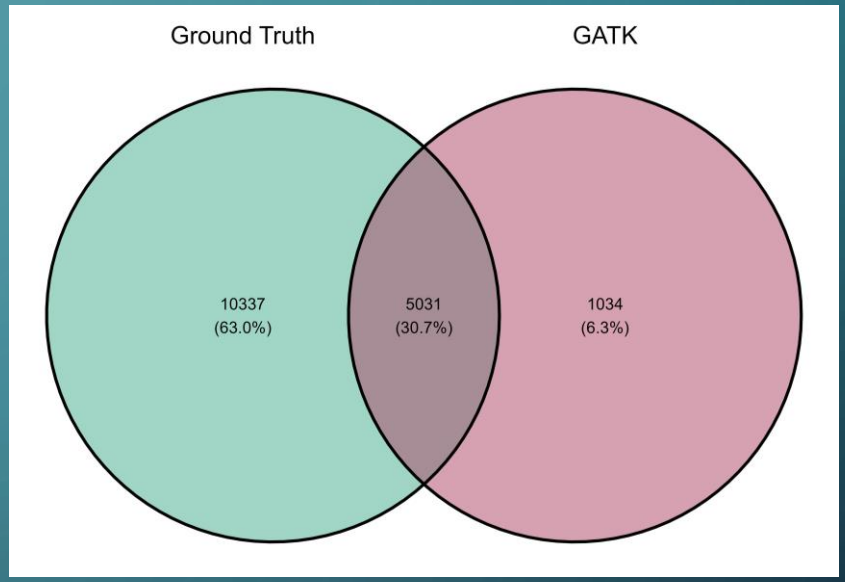
CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573

BENCHMARKING GATK-MUTECT2



Divergence in the identification of SNPs and their AF of «Ground Truth» Variants



Venn plot of the Overall Variants

sfragkoul@certh.gr
sfragkoul
@scfragkoul

National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITOYTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RECEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573

Highlights

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

To learn more about our work please visit

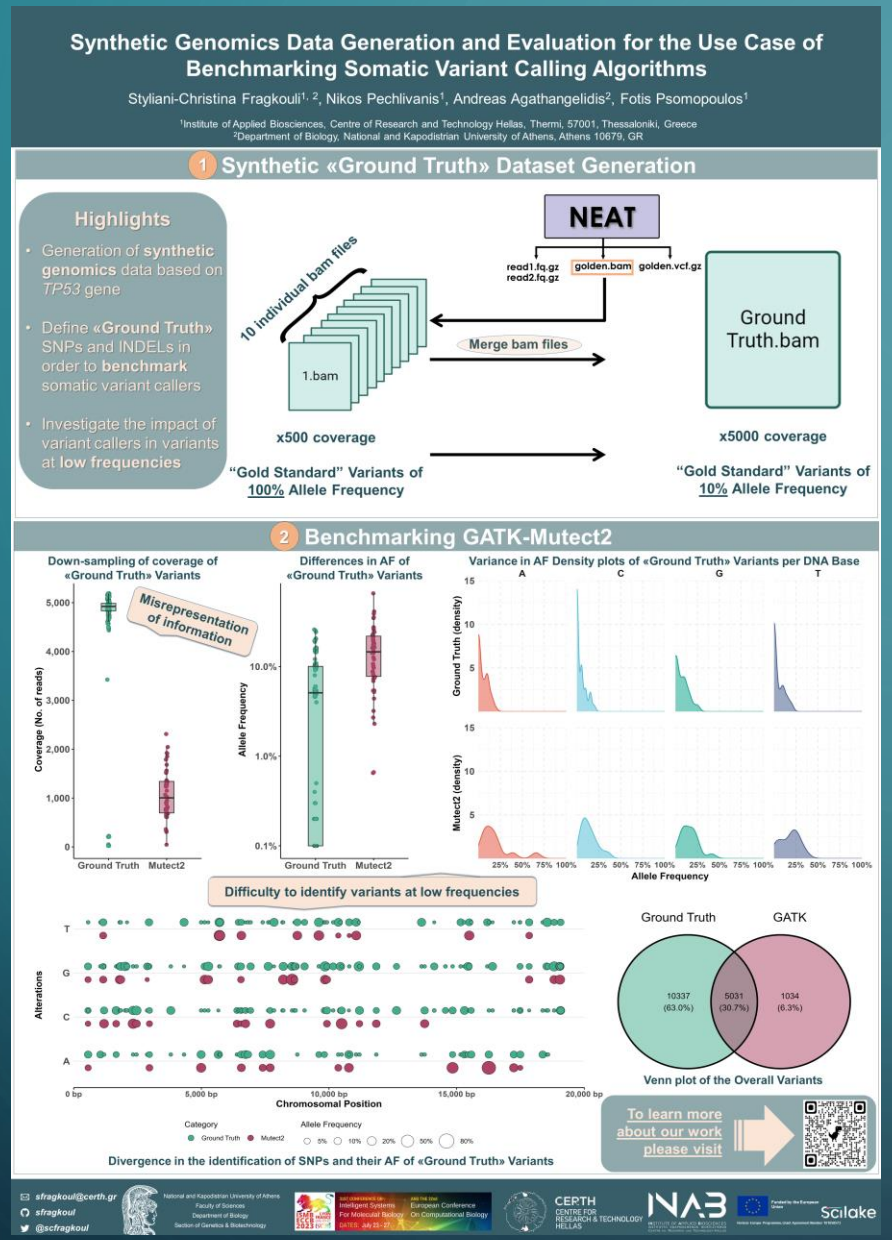
sfragkoul@certh.gr
sfragkoul
@scfragkoul

National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITOYTO EΦAPMOΣIMΩN BIOEΠIΣTHMΩN
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR
RESEARCH & TECHNOLOGY
HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573



Come say Hi :)

Monday 24th of July

Poster Session A

Track: BOSC

Poster: A-102

sfragkoul@certh.gr
sfragkoul
@scfragkoul

National and Kapodistrian University of Athens
Faculty of Sciences
Department of Biology
Section of Genetics & Biotechnology

INSTITUTE OF APPLIED BIOSCIENCES
INSTITOYTO EΦAPMOΣMENΩN BIOEΠIΣTHMΩN
CENTRE for RESEARCH and TECHNOLOGY-HELLAS

CERTH
CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

Funded by the European Union
Scilake
Horizon Europe Programme, Grant Agreement Number 101058573