

# COMBAT TB Explorer, a TB data exploration

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum



## Introduction

Current data repositories only present pre-computed data and do not provide the computational toolkit for biomedical researchers to analyze their own data. We have created the COMBAT TB Explorer, a Galaxy-based environment for annotating and exploring Mycobacterium tuberculosis sequence data. COMBAT TB Explorer currently combines a genomic variant calling pipeline with a web based tool for exploring the relationship between variants and known annotation and allows the user to perform geneset enrichment analysis.

## Methods & Results

- We implemented a bacterial genomic variant calling workflow in Galaxy, based on previous SANBI work on analysing mutations in drug resistance strains of M. tuberculosis.
- Variant calls were combined with a reference database of M. tuberculosis genome annotation to produce a new Galaxy datatype.
- z\_mplemented a COMBAT TB Explorer, a Galaxy Interactive Environment. This allows novel data to be explored both in the context of the M. tuberculosis genome (visualised in the JBrowse genome browser) and with reference to known annotation.

- Trimming
- Novo Align
- Novo Sort
- Mark Duplicates
- Realigner Target Creator
- Indel Realigner
- Depth of Coverage
- Snpeff

History

COMBAT-TB Pipeline

168: MPileup on data 1 and data 55 (log)

167: MPileup on data 1 and data 55

166: NOVO ALIGN on data 165 and data 164

165: Trim Galore on data 2 and data 3: trimmed reads pair 2

164: Trim Galore on data 2 and data 3: trimmed reads pair 1

163: Trim Galore on data 3: trimmed reads

162: Trim Galore on data 2: trimmed reads

161: QualMap: Multi BamQC PDF output

65: Haplotype Caller on data 1 and data 55 (VCF)

56: Indel Realigner on data 53, data 1, and data 51 (log)

55: Indel Realigner on data 53, data 1, and data 51 (bam)

54: Realigner Target Creator on data 1 and data 51 (log)

53: Realigner Target Creator on data 1 and data 51 (GATK interval)

52: Collect Alignment Summary Metrics on data 1 and data 51: Summary stats

51: MarkDuplicates on data 48:

50: MarkDuplicates BAM output

Galaxy / My own Galaxy browser

Tools

COMBAT-TB Pipeline

168: MPileup on data 1 and data 55 (log)

167: MPileup on data 1 and data 55

166: NOVO ALIGN on data 165 and data 164

165: Trim Galore on data 2 and data 3: trimmed reads pair 2

164: Trim Galore on data 2 and data 3: trimmed reads pair 1

163: Trim Galore on data 3: trimmed reads

162: Trim Galore on data 2: trimmed reads

161: QualMap: Multi BamQC PDF output

65: Haplotype Caller on data 1 and data 55 (VCF)

56: Indel Realigner on data 53, data 1, and data 51 (log)

55: Indel Realigner on data 53, data 1, and data 51 (bam)

54: Realigner Target Creator on data 1 and data 51 (log)

53: Realigner Target Creator on data 1 and data 51 (GATK interval)

52: Collect Alignment Summary Metrics on data 1 and data 51: Summary stats

51: MarkDuplicates on data 48:

50: MarkDuplicates BAM output

Tools

COMBAT-TB Pipeline

168: MPileup on data 1 and data 55 (log)

167: MPileup on data 1 and data 55

166: NOVO ALIGN on data 165 and data 164

165: Trim Galore on data 2 and data 3: trimmed reads pair 2

164: Trim Galore on data 2 and data 3: trimmed reads pair 1

163: Trim Galore on data 3: trimmed reads

162: Trim Galore on data 2: trimmed reads

161: QualMap: Multi BamQC PDF output

65: Haplotype Caller on data 1 and data 55 (VCF)

56: Indel Realigner on data 53, data 1, and data 51 (log)

55: Indel Realigner on data 53, data 1, and data 51 (bam)

54: Realigner Target Creator on data 1 and data 51 (log)

53: Realigner Target Creator on data 1 and data 51 (GATK interval)

52: Collect Alignment Summary Metrics on data 1 and data 51: Summary stats

51: MarkDuplicates on data 48:

50: MarkDuplicates BAM output

COMBAT-TB

geneset enrichment analysis

GO Term ID	GO Term Name	Raw p-value	Corrected p-value
GO:0045454	cell redox homeostasis	0.0016287817582708746	0.016075395698978433
GO:0004803	transposase activity	0.003152393480534542	0.016075395698978433
GO:0006313	transposition, DNA-mediated	0.003314455982926706	0.016075395698978433
GO:0006739	NADP metabolic process	0.004465387694160676	0.016075395698978433
GO:0003957	NAD(P)+ transhydrogenase (B-specific) activity	0.004465387694160676	0.016075395698978433
GO:0016655	oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	0.008913042420826633	0.022919251939268484

History

CTB

6.9 MB

6: Build\_ctb\_gene\_tool on data 2

Neo4 database (multiple files)

Neo4 database: database 2

uploaded gtf3 file

display with IGV (local)

1: snpeff

snpeff version 3

## Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.