MILRD Virtual Training Projects

Research Staff · Postdocs · Graduate Students

VTP OVERVIEW

Variant Calling + COVID-19
Collaboration with Dr. Adriana Heguy, NYU Medical Center

Aim

Characterize nucleotide and amino acid variants from SARS-CoV-2+ patient samples

Learning Goals

Discussion Topics	*Bioinformatics Tasks + Methods*
• Next Generation Sequencing methods and data structures • Principles and Methods resequencing analysis and variant detection • Principles and methods of data QC, alignment, variant calling	• Data QC • Human read removal • Alignment • Variant Calling • Variant Filtering • Visualization • NT and AA variant characterization • Evaluation in GisAID

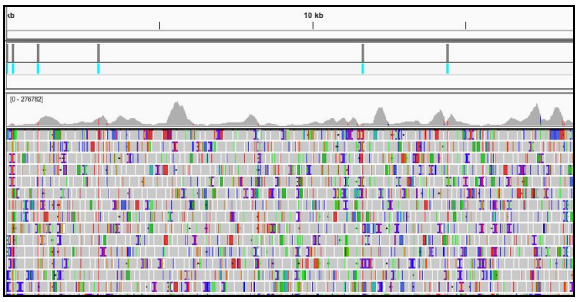
Suggested Preparation

Linux/Unix command-line fundamentals (optional)

Summary

The Covid-19 Pandemic—caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)—has infected over 6.9 million people and claimed over 400,000 lives in the past six months. This pandemic underscores the pressing need to interrogate the cellular, molecular, and physiological aspects of SARS-CoV-2 to help expand diagnostic and clinical treatments for Covid-19.

In this VTP—created with the help of Adriana Heguy's Research Group at NYU Medical Center—you'll analyze your own assigned dataset with help from your mentor and in collaboration with your cohort.



Throughout the week, participants will characterize clinical SARS-Cov2+ genomic data. In the Linux terminal, you will perform genomic data quality control, genome alignment, variant calling and visualization on SARS-Cov2+ samples.

Source Data

Maurano et al. *Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City Region.* Genome Research 2020. Preprint on MedRxiv. New York Times coverage.

MILRD Virtual Training Projects

Research Staff · Postdocs · Graduate Students

Schedule

Total Effort: ~10 hours

MILRD Provides

- Unlimited support from expert mentors
- Access to all required high-performance cloud-compute resources (AWS), analysis tools and software
- Access to all source data required to complete your project
- Optional Pre-VTP Preparation

Participants Provide

- A computer running Windows or MacOS
- Google Chrome, Safari, Firefox, or Edge
- A stable Internet connection

Sign Up

Review VTP dates and enrollment instructions on our Enrollment & Contact page.

```
VCFv4.2
PASS,Description="All filters passed">
sion=1.9+htslib-1.9
mand=mpileup --redo-BAQ --adjust-MQ 50 --gap-frac 0.05 --max-depth 10000 --max-idepth 200000 --output-type u -f /home/ubuntu/databases/NC_045512

ile:///home/ubuntu/databases/NC_045512_2.fasta
NC_045512.2,length=29903>
escription="Represents allele(s) other than observed.">
DEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
V,Number=1,Type=Integer,Description="Maximum number of reads supporting an indel">
F,Number=1,Type=Float,Description="Maximum fraction of reads supporting an indel">
,Number=1,Type=Integer,Description="Raw read depth">
B,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version="3">
B,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
B,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
B,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
SB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
B,Number=1,Type=Float,Description="Segregation based metric.">
0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
GT,Number=1,Type=String,Description="Genotype">
GQ,Number=1,Type=Integer,Description="Phred-scaled Genotype Quality">
B,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
B,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
,Number=1,Type=Integer,Description="Average mapping quality">
lVersion=1.9+htslib-1.9
lCommand=call --ploidy 1 --keep-alts --multiallelic-caller -f GQ -o BS06813A_calls.vcf; Date=Fri Sep  4 02:31:56 2020
   ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  BS06813A_sorted.bam
   24      .       G       .       29.5864 .       DP=51;MQ0F=0;AN=0;DP4=0,0,0,0;MQ=.      GT      .
   25      .       T       .       29.5864 .       DP=56;MQ0F=0;AN=0;DP4=0,0,0,0;MQ=.      GT      .
   26      .       A       .       29.5864 .       DP=57;MQ0F=0;AN=0;DP4=0,0,0,0;MQ=.      GT      .
   27      .       A       .       89      .       DP=58;MQ0F=0;AN=1;DP4=5,0,0,0;MQ=23     GT      0
   28      .       C       .       126     .       DP=58;MQ0F=0;AN=1;DP4=55,0,0,0;MQ=23    GT      0
   29      .       A       .       91      .       DP=59;MQ0F=0;AN=1;DP4=4,0,0,0;MQ=23     GT      0
   30      .       A       .       81      .       DP=60;MQ0F=0.0166667;AN=1;DP4=4,0,0,0;MQ=17     GT      0
   31      .       A       .       33.5884 .       DP=60;MQ0F=0.0166667;AN=1;DP4=1,0,0,0;MQ=0      GT      0
   32      .       C       .       98      .       DP=96;MQ0F=0.0104167;AN=1;DP4=48,0,0,0;MQ=23    GT      0
   33      .       C       .       127     .       DP=96;MQ0F=0.0104167;AN=1;DP4=53,0,0,0;MQ=23    GT      0
   34      .       A       .       101     .       DP=98;MQ0F=0.0102041;AN=1;DP4=6,0,0,0;MQ=22     GT      0
   35      .       A       T       5.64754 .       DP=98;VDB=0.0221621;SGB=-0.511536;RPB=0;MQB=0.666667;BQB=1;MQ0F=0.0102041;AC=0;AN=1;DP4=2,0,

   36      .       C       .       91      .       DP=99;MQ0F=0;AN=1;DP4=4,0,0,0;MQ=28     GT      0
   36      .       C       CT      148     .       INDEL;IDV=88;IMF=0.888889;DP=99;VDB=5.03075e-33;SGB=-0.693143;MQ0F=0.010101;AC=1;AN=1;DP4=61,
1:175,0:127
   37      .       C       .       42.5869 .       DP=99;MQ0F=0;AN=1;DP4=1,0,0,0;MQ=24     GT      0
   37      .       CA      CTA,CATCTTA     153     .       INDEL;IDV=9;IMF=0.0909091;DP=99;VDB=1.6822e-31;SGB=-0.693143;MQ0F=0.010101;AC=1,0;AN=
       1:180,0,148:127
   38      .       A       .       137     .       DP=99;MQ0F=0;AN=1;DP4=58,0,0,0;MQ=24    GT      0
   39      .       A       .       184     .       DP=99;MQ0F=0;AN=1;DP4=95,0,0,0;MQ=31    GT      0
```

info@milrd.org