Peer-graded Assignment: Course Project

1. Instructions
2. My submission
3. Discussions

Identification of polymoprphic sites in a family
Submitted on November 6, 2019
Shareable Link
Prompt

Use the text box below to submit a short write-up describing your results including the information requested above. (Maximum 300 words)

Quality of the given sequences were assesed using FastQC.
The sequences of father, mother and the child were mapped to human genome v19 using BWA-MEM v0.7.17.1.
Mapped files were merged using Merge Sam files and filtered
There are 3497 polymorphic sites are there in the data according to the VCF file (file attched) generated from FreeBayes v0.4.

1.Number of single nucleotide varients - 2963

2.Numberof insertions and deletions, insertions - 171, Deletions - 185

3.Number of multinucleotide varients - 34

4.Number of varients with multiple alternate alleles - 31

5. Genes which has largest number of polymorphic sites are
CACNA1H 42
USP7 40
ABAT 35
CLCN7 35
PKD1 32

Rubric

Use the space below to provide feedback about the write-up. Point out both strengths and areas for improvement. Providing substantive feedback is an important part of doing genomic data science in collaborative groups.

Luvena Ong
The description of what was done was very helpful!

Đỗ Ngọc Tuấn
you should use VCFFilter and Filter tools to take number of variants.

Has the student submitted a write-up?

- 1 point
  Yes
- 0 points
  No

Does the write-up accurately report the number of single nucleotide variants?

- 1 point
  Yes
- 0 points
  No

Does the write-up accurately report the number of insertion/deletion variants?

- 1 point
  Yes
- 0 points
  No

Does the write-up accurately report the number of multi-necleotide variants?

- 1 point
  Yes
- 0 points
  No

Does the write-up accurately report the number of variants with multiple alternate alleles?

- 1 point
  Yes
- 0 points
  No

Does the write-up accurately report the names of the 5 genes with the largest number of polymorphic sites?

- 1 point
  Yes
- 0 points
  No

Prompt

Upload your workflow file (.ga) by using the button below.

Galaxy-Workflow-Genomic_Data_science_with_Galaxy__Coursera_Assignment_.ga

Rubric

Use the space below to provide feedback about the workflow. Point out both strengths and areas for improvement. Providing substantive feedback is an important part of doing genomic data science in collaborative groups.

Luvena Ong
Pretty good though I see only 1 bwa-mem run..not 3..

Đỗ Ngọc Tuấn
You use too much FastQC, you should use FastQC before and after clean data if you want clean data

Has the student uploaded a .ga file?

- 1 point
  Yes
- 0 points
  No

Does the workflow file identify polymorphic sites in all three individuals?

- 1 point
  Yes
- 0 points
  No

Does the workflow file map the three sets of paired reads to the appropriate reference genome?

- 1 point
  Yes
- 0 points +0.5 pts because of a tie
  No

Did the student use a variant caller to identify sites that appear to have strong support for the presence of a polymorphism and call the genotype at that site for each sample?

- 1 point
  Yes
- 0 points
  No

Prompt

Upload your variation data file (.vcf) by using the button below.

Galaxy65-%5BFreeBayes_on_data_64_%28variants%29%5D.vcf
Rubric

Use the space below to provide feedback about the final VCF. Point out both strengths and areas for improvement. Providing substantive feedback is an important part of doing genomic data science in collaborative groups.

Luvena Ong
Very good!

Đỗ Ngọc Tuấn
good

Has the student uploaded a .vcf file?

- 1 point
  Yes
- 0 points
  No

Does the VCF file include only sites where the chance of a false positive call is 1 in 10,000 or better according to the VCF qual field?

- 1 point
  Yes
- 0 points
  No

Does the VCF file include a variant at position 2109995 (if hg38) or 2159996 (if hg19)?

- 1 point
  Yes
- 0 points
  No

Is the alternate allele at position 2109995 (if hg38) or 2159996 (if hg19) "A"?

- 1 point
  Yes
- 0 points
  No

Is the father's genotype called as homozygous for the alternate allele?

- 1 point
  Yes
- 0 points
  No

Is the daughter's genotype called as heterozygous for the alternate allele?

- 1 point
  Yes
- 0 points
  No

v