

Дмитрий Сергеевич Буг [05.04 15:30]:

https://drive.google.com/file/d/1V4jX3-FdLk-aQyHQFWeT2O1iCOY2OUJ/view?usp=share_link

https://drive.google.com/file/d/1EGZU_urVnPatuTRfvvublEibcVjZLvW7/view?usp=share_link

Дмитрий Сергеевич Буг [05.04 15:31]:

Необходимо провести контроль качества, выравнивание на референс (любой Homo sapiens), провести дальнейшую предподготовку, коллинг вариантов и выяснение их клинической значимости для пациентов (воспользуйтесь онлайн-версией VEP). В качестве ответа укажите по одному варианту в формате HGVS для каждой из исходных проб и пришлите VCF файлы на почту преподавателя.

Дополнительная информация: проводилось таргетное секвенирование генов BRCA1 и BRCA2, при подготовке библиотеки использовалось ампликоновое обогащение, ищите индел (bcftools для этого подходит плохо).

Дмитрий Сергеевич Буг [05.04 15:31]:

Файлы индексации тут:

https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/

```
bwa mem -t 24 nastyslav@nastyslav-In-555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур  
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi  
c.fna N3040_S2_L001_R1_001.fastq.gz N3040_S2_L001_R2_001.fastq.gz |  
samtools view -Sb - > N3040_S2_L001.bam
```

```
bwa mem -t 24 nastyslav@nastyslav-In-555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур  
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi  
c.fna N8058_S14_L001_R1_001.fastq.gz N8058_S14_L001_R2_001.fastq.gz |  
samtools view -Sb - > N8058_S14_L001.bam
```

```
gatk AddOrReplaceReadGroups -I N3040_S2_L001.bam -O  
N3040_S2_L001_rg.bam -RGID 1 -RGLB lib1 -RGPL ILLUMINA -RGPU unit1 -  
RGSM test_sample1
```

```
k AddOrReplaceReadGroups |  
-I N8058_S14_L001.bam |  
-O N8058_S14_L001_rg.bam |  
-RGID 2 |
```

-RGLB lib1 |
-RGPL ILLUMINA |
-RGPU unit1 |
-RGSM test_sample2

gatk SortSam -I N3040_S2_L001_rg.bam -O N3040_S2_L001_rg_sorted.bam -SO coordinate

gatk SortSam -I N8058_S14_L001_rg.bam -O N8058_S14_L001_rg_sorted.bam -SO coordinate

samtools index N3040_S2_L001_rg_sorted.bam

samtools index N8058_S14_L001_rg_sorted.bam

gatk MarkDuplicates -I N3040_S2_L001_rg_sorted.bam -O N3040_S2_L001_rg_sorted_dm.bam -M N3040_S2_L001_dm_metrics.txt

gatk MarkDuplicates -I N8058_S14_L001_rg_sorted.bam -O N8058_S14_L001_rg_sorted_dm.bam -M N8058_S14_L001_dm_metrics.txt

gatk BaseRecalibrator |

-R nastyslav@nastyslav-In-555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomic.fna |

-I N3040_S2_L001_rg_sorted_dm.bam |

--known-sites GRCh38_latest_clinvar.vcf |

-O N3040_S2_L001_recal.table

gatk BaseRecalibrator |

-R nastyslav@nastyslav-In-555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomic.fna |

-I N8058_S14_L001_rg_sorted_dm.bam |

--known-sites GRCh38_latest_clinvar.vcf |

-O N8058_S14_L001_recal.table

gatk ApplyBQSR |

-R nastyslav@nastyslav-In-

*555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna |*

-I N3040_S2_L001_rg_sorted_dm.bam |

-bqsr N3040_S2_L001_recal.table |

-O N3040_S2_L001_bqsr.bam

gatk ApplyBQSR |

-R nastyslav@nastyslav-In-

*555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna |*

-I N8058_S14_L001_rg_sorted_dm.bam |

-bqsr N8058_S14_L001_recal.table |

-O N8058_S14_L001_bqsr.bam

gatk HaplotypeCaller -R nastyslav@nastyslav-In-

*555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna -I N3040_S2_L001_bqsr.bam -O*

output/1/em_N3040_S2_L001_variants.g.vcf --emit-ref-confidence GVCF

gatk HaplotypeCaller -R nastyslav@nastyslav-In-

*555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna -I N8058_S14_L001_bqsr.bam -O*

output/1/em_N8058_S14_L001_variants.g.vcf --emit-ref-confidence GVCF

gatk GenotypeGVCFs |

-R nastyslav@nastyslav-In-

555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур

*a/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna |*

-V em_N3040_S2_L001_variants.g.vcf |

-O N3040_S2_L001_final_variants.vcf

gatk GenotypeGVCFs |

-R nastyslav@nastyslav-In-

*555:/media/nastyslav/F29E58A99E5867DD/Users/nasty/Desktop/Магистратур
а/биоинформатика/Зачет/genome/GCF_000001405.40_GRCh38.p14_genomi
c.fna |*

-V em_N8058_S14_L001_variants.g.vcf |

-O N8058_S14_L001_final_variants.vcf