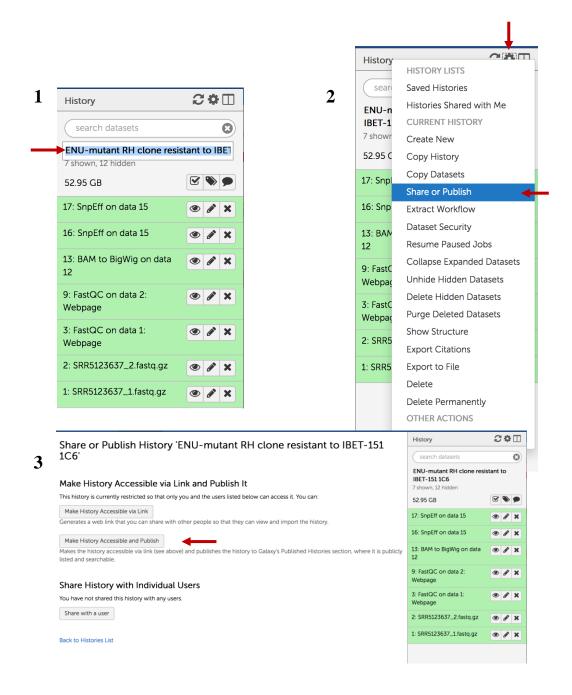
Analyzing Variant Call results using EuPathDB Galaxy, Part II

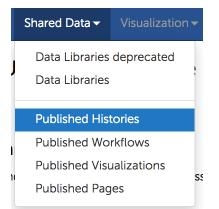
In this exercise, we will work in groups to examine the results from the SNP analysis workflow that we started yesterday. *The first step is to share your SNP workflow histories with the rest of the workshop participants:*

- 1. Give your workflow a meaningful name, eg. The sample or group name.
- 2. Click on the on the 'History options' link and select the 'share or Publish option'.
- 3. On the next page click on the 'Make History Accessible and Publish' link.



To import a shared history into your workspace follow these steps:

1. Select 'Published Histories' from the Shared data menu.

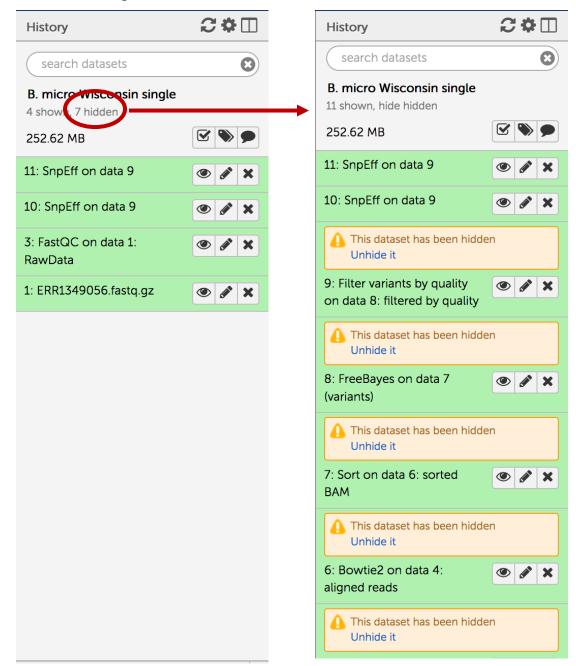


2. From the list of shared histories click on the one you want to import and on the next page select the 'Import' link in the upper right hand side.

| g 👌 globus Genomics | | | Shared Data 👻 | | |
|---|---|---------|---------------|--|----------------|
| Published Histories hb394 Group 1 results | | | | | Import history |
| Group 1 results 47.44 GB | | | | | |
| search datasets | | | | | 8 |
| Dataset | | Annotat | ion | | |
| 5: SRR1041268_1.fastq.gz | ۲ | > | | | |
| 6: SRR1041268_2.fastq.gz | ۲ | | | | |
| 7: SRR1041270_1.fastq.gz | ۲ | | | | |
| 8: SRR1041270_2.fastq.gz | ۲ | | | | |

Examining your results:

- 1. Click on the hidden files link in the history panel to reveal all workflow output files.
- 2. Examine the output files. What does the tool FASTQC do? What about Sickle?



3. The output of Sickle is used by a program called Bowtie2. What does this tool do? Bowtie generates a file called a BAM file. Whenever dealing with sequence alignment files you will likely hear of file formats called SAM or BAM. SAM stands for Sequence Alignment/Map format, and BAM is the binary version of a SAM file.

- 4. Many of the downstream analysis programs that use BAM files require a sorted BAM file. This allows access to reads to be done more efficiently.
- 5. The sorted BAM file is the input for a program called FreeBayes. This program is a Bayesian genetic variant detector designed to find small polymorphisms, specifically SNPs (single-nucleotide polymorphisms), indels (insertions and deletions), MNPs (multi-nucleotide polymorphisms), and complex events (composite insertion and substitution events) smaller than the length of a short-read sequencing alignment. The output for many variant callers is a file called a VCF file. VCF stands for variant interchange format.
- 6. Examine the VCF file in your results (click on the eye icon to view its contents). Detailed information about VCF file content is available here: <u>https://samtools.github.io/hts-specs/VCFv4.2.pdf</u>
- 7. What does tool SnpEFF do? SnpEff is a variant annotation and effect prediction tool. It annotates and predicts the effects of variants on genes (such as a mino acid changes).

Viewing VCF file results in a genome browser:

In order to view a VCF file in GBrowse, it first has to be converted to a format that GBrowse can understand like BigWig. To do this follow these steps:

- 1. Click on the edit attributes icon on the FreeBayes VCF output file.
- In the central window click on the 'Convert Format' tab.
- 3. Next select the 'Convert BED, GFF or VCF to BigWig' option and click on the 'Convert' link.



4. Notice a new step will appear in you history for the conversion step.

| Attributes Convert Format Dataty | pe Permissions |
|--|---|
| Convert to new format | |
| Convert VCF to BGZIP | |
| Convert VCF to VCF_BGZIP Convert Vcf to tabix | ntents of this dataset converted to a new format. |
| Convert BED, GFF, or VCF to BigWig | |

5. Once the conversion is done, you can export the BigWig file(s) to you EuPathDB dataset page

| | | | | | | | | | | Ŭ | | | | | | | - np | | | | |
|---------------|-------------------------------------|---------------------------|---------------------|---------------------|--|------------|-----------------------|--------------------|-----------------------|--------------------|--------|---------------------------|-----------------|----------------------|---------------|-----------|------------------|----------|--|----------|-----------------------------------|
| Overview | | 17_06_v3 | | | | | | | | | | | | | | | | | | | |
| | OM | 0.1 | | 0.2M | 0.3M | | 0.41 | 0.5M | 0.6 | | 0.7M | 0.8M | | 0.9M | 111 | | .1M | 1.2M | 1.3 | | 1.4M |
| Region | | | | | | | | | | | | | | | | | | | | | |
| | 360k | 370k | 380k | 390k | 400k | 410k | 420k | 430k | 440k | 450k | 460k | 470k | 480k | 490k | 500k | 510k | 520k | 530k | 540k | 550k | 560k |
| Details | | | | | | | TTL | | | | | | | | | | | | | | |
| | Pf 3D | 7_06_v3 | : 80 kb | P | | | 20 | kbp | | | | | | | | | | | | | |
| | 420k | | 43 | l + + + + | **** | 440k | **** | 45 | i0k | | 460k | | 4 | 70k | **** | 480k | **** | 4 | Pok | | 500k |
| | and the second second second second | property and property and | and an and a second | | and a second | La Cipline | desident and a second | and the local data | Company of the second | and a state of the | | description of the second | and good and it | in the second second | dentere ender | | the standard and | | den al a ser a | | |
| * 2 2 8 2 8 h | ttps_eupath | Idbworks | hop.glob | ousgenoi | nics.org_ | | | | | | | vig_plasm | nodb_bc | 5a8dd964 | 1414691_c | iata_gala | xy_9d443 | 2541559d | bc9d.bw | | 11.11 |
| * 8 8 8 2 8 h | ttps_eupath | idbworks | hop.glot | ousgenoi | nics.org_ | display_a | | on_9d4425 | | | | vig_plasm | nodb_bo | 15a8dd964 | 1414691_c | lata_gala | xy_9d44: | 2541559d | bc9d.bw | | [¹ ₀ . |
| | o.5 | anscripts | s (with U | TRs in gr | ay when a | | | | 111 | 111 | - Eğle | | | 111 | | 11 | | | | | [¹ ₀ .: |
| * 2 2 2 2 2 A | o.5 | | s (with U | TRs in gr | | | | Ē | 111 | 07_06103 | - Eğle | | | 5a8dd964 | | 11 | | | bc9d.bw | 611800.1 |] [¹ ₀ .: |
| | o.5 | anscripts | s (with U | TRs in gr 1C2) P | ay when a | available) | PF3D7_ | | PF3 | D7_06103 | - Eğle | PF3D7, | _0611000 | 111 | 7_0611400 | .1(SWIB) | PF3D7_ | | . PF3D7_0 | 611800.1 |] [¹ ₀ .: |

Filtering data in VCF files:

VCF files contain a lot of data about variants and their positions. SnpEff generates various analyses/summaries of VCF files (including GeneIDs that overlap variant positions). However, it is often necessary to filter VCF files further to obtain useful information for your specific question. For example, you may want to filter out SNP positions that have an impact on the coding sequence. One tool that can be used is called SnpSift Filter. This tool allows you to write complex expressions to filter a VCF file.

| 👩 🏷 globus Genomics | | Ar | nalyze Da | ata Workflow | | | | | | Using 902.1 GB |
|--|--|------------|-----------|----------------------------------|----------------------|------------------|--------------|---|---|------------------|
| Tools 1 | ##INFO= <id=n< td=""><td>UMALT,N</td><td>lumber=</td><td>1,Type=Integer,I</td><td>Description="Num</td><td>ber of unique</td><td>non-refere</td><td>nce alleles in called genotypes at this p</td><td>History</td><td>2 0 []</td></id=n<> | UMALT,N | lumber= | 1,Type=Integer,I | Description="Num | ber of unique | non-refere | nce alleles in called genotypes at this p | History | 2 0 [] |
| snpsift F | | | | | | | ique non-r | eference allele observations per sampl | e csearch datasets | 0 |
| and and a | ##INFO= <id=l< td=""><td></td><td></td><td></td><td>ription="allele leng</td><td></td><td></td><td></td><td>C</td><td></td></id=l<> | | | | ription="allele leng | | | | C | |
| VGS: Variant Detection | | | | Contraction Contraction Contract | | | | alternate alleles"> | Plasmodium Chloroqu | |
| VARSCAN TOOLS | and the second s | | | | | | | d reference alleles'> | 10 shown, 10 deleted, 9 hi | |
| SnpSift Filter Filter variants using | Interesting to the second s | | | | | | | alleles which are supported by proper | 100000 | 8 8 9 |
| arbitrary expressions | and the second s | | | | | | ed reference | ce alleles which are supported by prop | 28: SnpEff on data 26 | |
| SnpSift CaseControl Count samples | ##FORMAT= <ii< td=""><td>D=GT,Nu</td><td>mber=1,7</td><td>Type=String,Des</td><td>cription="Genotyp</td><td>ie'></td><td></td><td></td><td>28. Shipen on data 20</td><td>@ / X</td></ii<> | D=GT,Nu | mber=1,7 | Type=String,Des | cription="Genotyp | ie'> | | | 28. Shipen on data 20 | @ / X |
| are in 'case' and 'control' groups. | ##FORMAT= <ii< td=""><td>D=GQ,Nu</td><td>mber=1,</td><td>Type=Float,Desi</td><td>cription="Genotyp</td><td>e Quality, the</td><td>Phred-scal</td><td>ed marginal (or unconditional) probabi</td><td>27: SnpEff on data 26</td><td>@ / X</td></ii<> | D=GQ,Nu | mber=1, | Type=Float,Desi | cription="Genotyp | e Quality, the | Phred-scal | ed marginal (or unconditional) probabi | 27: SnpEff on data 26 | @ / X |
| SnpSift Annotate Annotate SNPs | ##FORMAT= !</td <td>D=GL,Nu</td> <td>mber=G.</td> <td>Type=Float,Desi</td> <td>cription='Genotyp</td> <td>e Likelihood, l</td> <td>og10-scale</td> <td>d likelihoods of the data given the call</td> <td>62.713 lines. 61 comme</td> <td>New York (199</td> | D=GL,Nu | mber=G. | Type=Float,Desi | cription='Genotyp | e Likelihood, l | og10-scale | d likelihoods of the data given the call | 62.713 lines. 61 comme | New York (199 |
| from dbSnp | ##FORMAT= <ii< td=""><td>D=DP,Nut</td><td>mber=1,7</td><td>lype=Integer,De</td><td>scription="Read D</td><td>epth"></td><td></td><td></td><td>format: vcf. database:</td><td></td></ii<> | D=DP,Nut | mber=1,7 | lype=Integer,De | scription="Read D | epth"> | | | format: vcf. database: | |
| | ##FORMAT= <i< td=""><td>D=RO,Nu</td><td>mber=1,</td><td>Type=Integer,De</td><td>escription="Referen</td><td>nce allele obse</td><td>ervation co</td><td>unt'></td><td>29_Pfalciparum3D7_C</td><td></td></i<> | D=RO,Nu | mber=1, | Type=Integer,De | escription="Referen | nce allele obse | ervation co | unt'> | 29_Pfalciparum3D7_C | |
| SnpSift Intervals Filter variants using intervals | ##FORMAT= <i< td=""><td>D=QR,Nu</td><td>mber=1,</td><td>Type=Integer,De</td><td>escription="Sum of</td><td>f quality of the</td><td>reference</td><td>observations"></td><td>BOCIM</td><td></td></i<> | D=QR,Nu | mber=1, | Type=Integer,De | escription="Sum of | f quality of the | reference | observations"> | BOCIM | |
| | ##FORMAT= <ii< td=""><td>D=AO,Nu</td><td>mber=A;</td><td>Type=Integer,De</td><td>escription="Alterna</td><td>ite allele obser</td><td>vation cou</td><td>nt"></td><td>BOCM</td><td></td></ii<> | D=AO,Nu | mber=A; | Type=Integer,De | escription="Alterna | ite allele obser | vation cou | nt"> | BOCM | |
| SNPSift concordance Calculate | ##FORMAT= <i< td=""><td>D=QA,Nu</td><td>mber=A</td><td>Type=Integer,De</td><td>escription="Sum o</td><td>f quality of the</td><td>alternate</td><td>observations"></td><td>display with IGV local</td><td></td></i<> | D=QA,Nu | mber=A | Type=Integer,De | escription="Sum o | f quality of the | alternate | observations"> | display with IGV local | |
| concordance between two VCF files | ##SnpEffVersio | n="4.11 (b | uild 2015 | 5-10-03), by Pat | olo Cingolani* | | | | 1. Chron | 2.Pos |
| orkflows | ##SnpEffCmd= | SnpEff - | ivcf -o v | cf -stats /scratc | h/galaxy/files/008 | /dataset_8077 | dat Plasma | DB-29_Pfalciparum3D7_Genome /sc | ##fileformat=VCFv4.1 | |
| All workflows | ##INFO= <id=a< td=""><td>NN,Num</td><td>ber=_Typ</td><td>e=String.Descri</td><td>ption="Functional</td><td>annotations: '</td><td>Allele Ann</td><td>otation Annotation_Impact Gene_N</td><td>##fileDate=20170616</td><td></td></id=a<> | NN,Num | ber=_Typ | e=String.Descri | ption="Functional | annotations: ' | Allele Ann | otation Annotation_Impact Gene_N | ##fileDate=20170616 | |
| All workitows | ##INFO= <id=l< td=""><td>OF,Numb</td><td>er=_Typ</td><td>e=String.Descrip</td><td>tion="Predicted lo</td><td>oss of function</td><td>effects for</td><td>this variant. Format: 'Gene_Name Ge</td><td>##source=freeBayes v</td><td>9.9.21-19-gc003c</td></id=l<> | OF,Numb | er=_Typ | e=String.Descrip | tion="Predicted lo | oss of function | effects for | this variant. Format: 'Gene_Name Ge | ##source=freeBayes v | 9.9.21-19-gc003c |
| | ##INFO= <id=n< td=""><td>MD,Num</td><td>ber=Ty</td><td>pe=String.Descr</td><td>iption="Predicted</td><td>nonsense med</td><td>fiated deca</td><td>y effects for this variant. Format: 'Gene</td><td>##reference=/mnt/gala</td><td>xyIndices2/genor</td></id=n<> | MD,Num | ber=Ty | pe=String.Descr | iption="Predicted | nonsense med | fiated deca | y effects for this variant. Format: 'Gene | ##reference=/mnt/gala | xyIndices2/genor |
| | #CHROM | POS | ID F | REF | ALT | QUAL | FILTER | INFO | ##phasing=none | |
| | Pf3D7_01_v3 | 30 | · 1 | ۱. | G | 106.836 | | AB=0;ABP=0;AC=0;AF=0;AN=2;AO= | ##commandline="freebo | yesbam local |
| | Pf3D7_01_v3 | 415 | . (| 3 | С | 100.39 | | AB=0;ABP=0;AC=0;AF=0;AN=2;AO= | an music state to the | ma Callela |
| | Pf3D7_01_v3 | 421 | | ATTA | ATTC | 93.0313 | | AB=0;ABP=0;AC=2;AF=1;AN=2;AO= | 26: Filter variants by qua on data 14: filtered by | lity 👁 🖋 🗙 |
| | Pf3D7_01_v3 | 466 | a 1 | E. | A | 211.522 | | AB=0;ABP=0;AC=2;AF=1;AN=2;AO= | | |
| | Pf3D7_01_v3 | 704 | a 1 | F: | С | 73.1939 | | AB=0;ABP=0;AC=2;AF=1;AN=2;AO= | 7 | |
| | Pf3D7_01_v3 | 709 | . (| 5 | С | 179.817 | | AB=0;ABP=0;AC=2;AF=1;AN=2;AO= | 20: Convert BED, GFF, o VCF to BigWig on data 1 | |
| | Pf3D7_01_v3 | 737 | . (| C | G | 52.3887 | | AB=0:ABP=0;AC=2;AF=1;AN=2;AO= | | |
| | Pf3D7_01_v3 | 781 | . (| STTA | CTTA | 69.6111 | | AB=0;ABP=0;AC=2;AF=1;AN=2;AO= | 2 14: FreeBayes on data 12 | 2 🕑 🖋 🗙 |
| | Pf3D7_01_v3 | 977 | . (| 3 | с | 169.554 | | AB=0:ABP=0:AC=2:AF=1:AN=2:AO= | g (variants) | |

The galaxy workflow we used already includes SnpSift Filter as the final step using the following expression:

(((ANN[*].IMPACT has 'HIGH') | (ANN[*].IMPACT has 'MODERATE')) & ((na FILTER) | (FILTER = 'PASS'))) - Examine the filtered VCF file. Notice that the GeneIDs are buried in the file but the file has some structure which means you can extract them either programmatically or using a program like Excel.

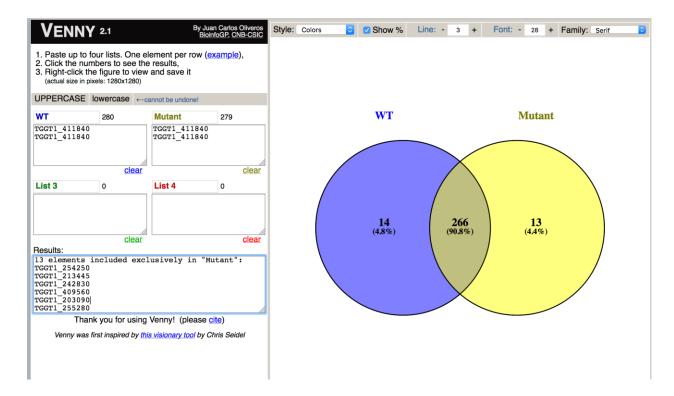


Here are some steps you can take to extract Gene IDs from two VCF files then compare them to identify genes that are in common or that distinguish the two files.

- 1. Download the SnpSift Filter output by clicking on the save icon
- 2. Open this file using excel and make sure you select tabs and | as column delimiters

| | Text Import Wizard - Step 1 of 3 | | | | | | | | | | | | Text Im | port Wizar | d - Step 2 | of 3 | | |
|---|---|--|--|--|---|--|---|---|--|---|--|---|---|---|--|---|--|-------------|
| he Text Wizard h | as detern | nined that v | our data i | is Delimite | ed. | | | | | Thie | ecreen le | te vou eet | the delim | iters your | data con | taine | | |
| this is correct, ch | | | | | | ribae unur | data | | | | | ta you set | une demin | iters your | uata com | cama. | | |
| this is conect, ci | IOUSE NEX | t, or choose | the Data i | spe that s | Jest desc | nues your | udid. | | | Delim | | | | | | | | |
| Delimited - C | | | | | | | | | | 💟 Ta | | | | | Treat co | nsecutive | delimiters | as one |
| Fixed width - Fi | elds are al | ligned in colu | umns with | spaces be | etween ea | ich field. | | | | Se | micolon | | | Te | ext qualifi | er: " | 0 | |
| art import at row | : 1 0 | File o | origin: M | acintosh | | | 0 | | | Co | omma | | | | | | | |
| | | | | | | | | | | | ace | | | | | | | |
| | | | | | | | | | | 🖸 Ot | her: (| | | | | | | |
| review of selecter | i data: | | | | | | | | | Dravie | nu of colo | cted data: | | | | | | |
| review of file /Use | | ownloads/G | alaxy29-fSr | npSift Filte | r on data | 271.vcf. | | | | FIEVA | ew of sele | cieu uata. | | | | | | |
| | ion officiants, p | i i i i i i i i i i i i i i i i i i i | alaxyes (or | apoint_r into | on_outo | Carlinon. | _ | | | | format-WCFv4 | .1 | | | | | | |
| Wfileformat=VCFv4.1 WfileDate=20170616 W#source=freeBayes v | 9. 21-19-oct | Billic1e | | | | | | | | ##sour | Date-2817861 | v0.9.21-19- | c003cle | | | | | |
| ##reference=/mnt/gal ##phasing=none | xyIndices2/g | genomes/PlasmoD | 8-29_Pfelcip | anun307_Geno | ne/seq/Plass | noDB-29_Pfal | ciporum307, | | | ##refe | rence-/mit/g | alaxyIndices | /genomes/Pla | smcDB-29_Pfa1 | ciparum307_ | Genome/seq/P | Losmo08-29_Pi | falciparumi |
| W/commandline_"freeb | ayesban la | ocalbam_0.bam - | -fasta-refer | ence /mnt/ga | laxyIndices | 2/genomes/Pl | asno08-29_F | | | stricom | andline-"fre | eboyes bon | locolban_0.t | on fasto-re tion-"Number | ference /m | t/galaxyIndi | ces2/genomes/ | Plasmob8- |
| WINFO- <id-op, number<="" td=""><td>1, Type-Inter</td><td>per, Description</td><td>-"Total read</td><td>depth ot th</td><td>e locus"></td><td>er herer in</td><td>mark and</td><td></td><td></td><td>and INFO</td><td> ID-OP_Nunt</td><td>er-1, Type-Ini</td><td>eger Descrip</td><td>tion="Total r</td><td>ead depth a</td><td>t the locus"</td><td>n Locus: hoses</td><td>in reads a</td></id-op,> | 1, Type-Inter | per, Description | -"Total read | depth ot th | e locus"> | er herer in | mark and | | | and INFO | ID-OP_Nunt | er-1, Type-Ini | eger Descrip | tion="Total r | ead depth a | t the locus" | n Locus: hoses | in reads a |
| PATH CONTRACTOR INVESTIGATION | The second second second | | | | - | | | | | | | | | Cancel | < Ba | _ | Next > | Finis |
| | | Car | ncel | < Back | Next | | Finish | | | | | | | Cancer | < De | JCK | Next > | Pins |
| | Paste | 0 • 2 | Page La Calibri (Body B I U | • • 12 | A | | = = % | iew • • •¶ • | | General | 3 | * [| iditional Fo | rmat Cell | • | nsert * Delete * | Σ · A Σ · Z Sort | |
| | A1 | • • • • • | talibri (Body B I <u>U</u> √ f _X I |) + 12 + + #fileformat | • A• | A• = | * | • • | | \$ • % | | Fo | matting as | Table Style | | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. .a |
| | A1 E | | Calibri (Body B I U ✓ fx I G |) * 12 ! * | × A ▲ ▲ × A ▲ × A ■ VCFv4.1 I | | E = 8 | | | |) •.0 | ▼ ↓ 0 P | ditional For matting as | rmat Cell Table Style | • | Delete * | ∑ A Z Z Sort | 7. .a |
| | A1 E 49 Senot | ¢ × | Calibri (Body B I U |) * 12 * • • * #fileformat H d marginal (or | VCFv4.1 | A - E | * | L otype*> | ÷÷ v M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. .a |
| | A1 E 49 Senot 50 Senot 51 "Rear | F type Quality, the type Ukelihood, I d Depth"> | alibri (Body B I U fx A G Phred-scaled log10-scaled I |) * 12 + • • • #fileformat H fmarginal (or ikelihoods of | VCFv4.1 | A - E | E S S S S S S S S S S S S S S S S S S S | L otype*> | ÷÷ v M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. .a |
| | A1 E 49 Senot 50 Senot 51 "Read 52 "Refe | F type Quality, the ype Ukelhood, I d Depth'> erence allele obs | alibri (Body B I U fx A G Phred-scaled log10-scaled I servation cour |) + 12 Hfileformat H marginal (or ikelihoods of nt"> | A | A - E | E S S S S S S S S S S S S S S S S S S S | L otype*> | ÷÷ v M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. .a |
| | A1 49 jenot 50 jenot 51 "Rea 52 "Refe 53 "Surr 54 -"Alte | F type Quality, the ype Ukelihood, I d Depth"serence allele obs no fiquality of the remate allele obs | alibri (Body B I U fx A G Phred-scaled log10-scaled I servation cours e reference of ervation cours |) + 12 H fileformat H d marginal (or ikelihoods of nt"> bservations"s it"> | A A | A - E | E S S S S S S S S S S S S S S S S S S S | L otype*> | ÷÷ v M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. .a |
| | A1 E 49 senot 50 senot 51 "Rea 52 "Refe 53 "Sun 54 -"Alte 55 "Sun | F type Quality, the ype Ukelihood, I d Depth"> erence allee obs n of quality of the | alibri (Body B I U fx A G Phred-scaled log10-scaled I servation cours e reference of ervation cours |) + 12 H fileformat H d marginal (or ikelihoods of nt"> bservations"s it"> | A A | A - E | E S S S S S S S S S S S S S S S S S S S | L otype*> | ÷÷ v M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 Senot 51 "Reas 52 "Refe 53 "Sum 54 "Alte 55 "Sun 56 ani" | F type Quality, the ype Ukelihood, I d Depth*> rence allele obs n of quality of the rmate allele obs n of quality of the | alibri (Body B I U G Phred-scaled log10-scaled I servation cours e reference of envation cours e alternate of |) * 12 Wfileformat H d marginal (or ikelihoods of ht*> bservations*> | • A • A • A • A • A • A • A • A • A • A | A J all probabilit en the called | E S S S S S S S S S S S S S S S S S S S | L otype"> possible geno | M M | \$ * % | 0 | P | Q | Table Style: | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 Senot 51 "Rea 52 "Refe 53 "Sur 54 -"Alte 55 -"Sur 56 ani" 57 Ses/0 58 p_ID | F Type Quality, the ype Ukelihood, I d Depth's enerce allele obs n of quality of th 08/dataset_807 Feature_Typ | alibri (Body B I U f ₃ A G Phred-scaled og10-scaled I servation cours e reference of ervation cours e alternate of 7.dat Plasmo(|) * 12 1 * 12 | A | A J all probabilit en the called | K y of the called gen | L otype"> possible genc | M M | N ed from the re | O eference and | P alternate alle | Q | R R ample ploidy | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Reas 52 "Refe 53 "Sum 54 "Alte 55 "Sun 56 ani" 57 ies/0 58 p_ID 59 zipts | F type Quality, the ype Ukelihood, I d Depth's erence allele obs n of quality of the rmate allele obs n of quality of the OB/dataset_B77; Feature T's | alibri (Body B I U f ₃ A G Phred-scaled og10-scaled I servation cours e reference of ervation cours e alternate of 7.dat Plasmo(|) * 12 1 * 12 | A | J all probabilit en the called | Ktyglaxy/files/00 | L otype"> possible genc | M stype generat | N ed from the re | O eference and | P alternate alle | Q. | R R ample ploidy | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Reas 52 "Refe 53 "Sum 54 "Alte 55 "Sun 56 ani" 57 ies/0 58 p_ID 59 zipts | F Type Quality, the type Quality, the type Quality of the of quality of the of quality of the OB/dstaset_807 Feature_Type Laffected "> | alibri (Body B I U f ₃ A G Phred-scaled og10-scaled I servation cours e reference of ervation cours e alternate of 7.dat Plasmo(|) * 12 1 * 12 | A | J all probabilit en the called | Ktyglaxy/files/00 | L otype"> possible genc | M stype generat | N ed from the re | O eference and | P alternate alle | Q. | R R ample ploidy | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Read 52 "Refa 53 "Sum 54 "Alte 55 "Sum 56 ani" 57 ies/0 58 a. 60 pripts 60 pripts 61 p. 208/c | F Type Quality, the type Quality, the type Quality of the of quality of the of quality of the OB/dstaset_807 Feature_Type Laffected "> | alibri (Body B I U f 2 f 2 Phred-scaled log10-scaled I ervation cours e reference of envation cours e elternate of 7.dat Plasmo(Feature_ID |) * 12 Hilleformat H marginal (or ikelihoods of tt"> bservations"> D8-29_Pfalcip Transcript_I | A A | J all probabilit enome /scrat HGVS.c | K K y of the called gen genotype for each HGV3.p | L otype"> possible genc | M stype generat | N ed from the re | O eference and | P alternate alle | Q. | R R ample ploidy | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Rea 52 "Ref 53 "Sur 54 'Alte 55 "Sur 56 an!" 57 ies/0 58 e_ID 59 cripts 61 o Cinj 62 308/c 63 | F | alibri (Body B I U fx I G Phred-scaled og10-scaled I revation cours e reference of ervation cours e alternate of 7.dat Plasmot Feature_ID t = /scratch/r |) * 12 * * * * * ##fileformat H amarginal jon ikelihoods of nt"> bservations" bservations"> bservations" | A A | J all probabilit enome /scrated HGVS.c | K K y of the called gen genotype for each HGV3.p | L otype"> possible genc | M stype generat | N ed from the re | O eference and | P alternate alle | Q. | R R ample ploidy | s S | ormat * | ∑ * A Z ¶ Z ¶ Fib | 7. 8 |
| | A1 49 Senot 50 Senot 51 "Rees 52 "Refe 53 "Sun 55 "Sun 55 seni" 57 Senot 58 s_ID 59 cripts 50 cripts 51 c Close 64 | Provide a second s | alibri (Body B I U fx A G Phred-scaled Ing10-scaled I ervation cours e alternate of reference of |) * 12 Wfileformat H marginal (or ikelinods of tt"> bservations"> tt"> bservations"> bservations"> tt"> bservations"> bservations"> tt"> bservations" bservations"> bservations" bservations"> bservations" bservati | A A | all probabilitien the called | K y of the called gene genotype for each HOYS p | L otype"> possible genc cDNA.pos / | M type generat 75.dat * (CDS.pos / C | N N ed from the re | O eference and | P alternate alle | Q. Q. les given the /ARNINGS / II | R R sample ploidy 4FO' "> | s S | telete * | ∑ * A | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Rea 52 "Ref 53 "Sur 54 'Alte 55 "Sur 56 an!" 57 ies/0 58 e_ID 59 cripts 61 o Cinj 62 308/c 63 | F | alibri (Body B I U fix A G Phred-scaled ing10-scaled Prode-scaled reference of envation coun e networkion coun e networkion e networ | I) * 12 III * III * Imarginal (or ikelihoods of nt"> bservations"> bservations"> bservations"> bservations"> D8-29_Pfalcip Transcript_I salasy/job_w INFO | A A | J all probabilit enome /scrat HGVS.c unknown a MODERNT | K K y of the called gen genotype for each HGV3.p | L otype"> possible geno cDNA.pos / PF307_010 | M https://www.eneratory.com/com/com/com/com/com/com/com/com/com/ | N ed from the re | O | P elternate alle ERRORS / V | Q es given the ARNINGS / II | R R ample ploidy | s s 1729/6492 | T T 1729/6492 | ∑ * A | 7. 8 |
| | A1 e 49 Senot 50 Senot 51 "Real 52 "Ref 55 "Sour 56 snl" 57 Jes/0 59 pipts 60 pipts 60 pipts 61 pipts 62 208/c 63 64 65 66 67 | Point Control of | alibri (Body B I U G Fhred-scaled og10-scaled pervation cours e reference of envation cours e reference of envation cours e alternate of restate (B) Fature (B) Future (B) Futur | i) v 12 iv i | The VCFv4.1 | A - J J J J J J J J J J J J J J | K y of the called gen genotype for each HOYS p P0/tmpAQDb8H* P107_0100100 HOYS p | L ctype"> possible genc cDNA.pos / PF3D7_010 PF3D7_010 | M http://www.commentationality.com/ http://wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww | N AA.pos / AA PF3D7_0100 PF3D7_0100 | Coding Coding Coding Coding | P alternate alle ERRORS / V 2-Jar | ARNINGS / II c.1729_173 c.172A-T c.420_425 | R ample ploidy FO' "> D p.Asp577Pro p.Ly5391Aan p.Thr147408 | 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 | T T T T T T T T T T T 773/6492 4420/6492 | ∑ * A | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Rea 52 "Ref" 53 "Sun 54 "Aite 55 "Sun 55 "Sun 55 "Sun 56 ani" 57 iesu 58 ani" 57 iesu 58 ani" 59 jipts 61 o Cințe 62 208/c 63 64 65 66 67 68 | Provide a second s | alibri (Body B I U G F A Internet Sector Internet Sector Pression Councers Feature_ID FutTER - - |) + 12 IIIII + IIIIIIIIIIIIIIIIIIIIIIIIIIIIII | The VCFv4.1 | A J J all probabilit enome /scrat HGVS.c tory/004/413 unknown a MODERATE a MODERATE a MODERATE | K y of the called gene genotype for each HOVS.p 20/tmpAQDb8H* PB307_0000100 FF307_0000100 | L type"> possible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 | M http://www.enerationality.com/ bt/pregenerationality.c | \$ * % N ed from the re (AA.pos / AA PF3D7_0100 PF3D7_0100 PF3D7_0100 | O eference and Distance O Coding O Coding O Coding O Coding O Coding O Coding | P P alternate alle ERRORS / V 2-iar 2-iar | Q es given the ARNINGS / II c.1729_173 c.1773A>T c.4432C-64 | R ample ploidy F() "> p.Asp577Pro p.Lys59IAsn p.Thr1474Gi p.Gin1478Gi | 5 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 | T 1729/6492 1773/6492 1773/6492 | ∑ * A | 7. 8 |
| | A1 e 49 Senot 50 ienot 51 "Rea 52 "Ref 53 "Sum 56 ani" 57 les/0 58 e_10 58 e_10 56 ao Cin 66 co 66 67 68 69 | F F F F F F F F F F F F F F | alibri (Body B I U fx A G Phred-scaled og10-scaled iservation count e reference of environ count e reference of environ count e enternate of z-dat Plasmot Feature_ID t e /scratch/n FILTER | Hfileformat Hfileformat H marginal (or kichoods of nt'> bservations'> bservations'> bservations'> bservations'> INFO AB=0;AB=0;AB=0; AB=0;AB=0;AB=0;AB=0;AB=0;AB=0;AB=0;AB=0; | A A | A · J all probabilite in the called HGVS.c tory/004/412 unknown a MODERATE a MODERATE a MODERATE a MODERATE a MODERATE | K y of the called gene genotype for each inf075.p //mpAQD88H* PB07.p00000 PB07.p00000 PB07.p00000 PB07.p00000 PB07.p00000 PB07.p00000 PB07.p00000 | L ctype"> possible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 | M N Nype generation 75.dat * (CDS.pos / C D transcript D transcript D transcript D transcript D transcript | S + % N ed from the re PF3D7_0100 PF3D7_0100 PF3D7_0100 | O O O O | P P alternate alle ERRORS / V 2-iar 2-iar | C.1729_173AFT c.4720_422_422 c.4430_442 c.4432C-6 c.4456C-A | R R FO' "> P.Asp577Pro p.Asp577Pro p.Jy591Asn p.Thr148446 p.Gin147866 p.Thr1474861 | s s 1729/6492 1773/6492 4432/6492 4436/6492 | T T T T T T T T T T T T T T T T T T T | ∑ * A | 7. 8 |
| | A1 49 Senot 50 ienot 51 "Reas 52 "Ref" 53 "Sum 54 "Aite 55 "Sum 55 "Sum 55 "Sum 56 ani" 57 ies/ 58 p.[D 59 origits 61 o Cinj 62 208/c 63 64 65 66 67 70 | F Type Quality, the ype Likelihood, I depth ³ F transcription F rence allele obso n of quality of th of quality of th rente allele obso n of quality of th rente allele obso n of quality of th Geldataset, 8076.da QUAL 165.615 59.2433 112.419 122.945 70.7189 20.1132 20.1132 | alibri (Body B I U G F A I Phed-scale log10-scaled I ervation count e afternate of 7.dat Plasmot Feature_ID t - e /scratch/I FILTER |) + 12 Hffieformat Hf ideformat H anginal (marginal (| A A | A j all probabilit enome /scrat HGVS.c tory/004/41 unknown MODERATE MODERATE MODERATE MODERATE MODERATE MODERATE | # # x # y of the called generative for each # KCh/galaxy/files/UO # HOYS.p # PRID7_100000 # # <t< td=""><td>L cotype"> pessible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010</td><td>M type generation 75.dat * (CDS.pos / C D transcript D transcript D transcript D transcript D transcript</td><td>\$ * % N ad from the re AA.pos / AA PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100</td><td>O Coding Coding</td><td>P P alternate alle ERRORS / V 2-iar 2-iar</td><td>Q Q es given the ARNINGS / II c.1729_173 c.1773A>T c.4420,442 c.4432C>G c.4455C>A</td><td>R R ample ploidy (FO' "> p.Asp577Pro p.Jy39JAan p.Thr1474651 p.Thr1474631</td><td>1729/6492 1773/6492 4432/6492 4432/6492</td><td>T 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7</td><td>∑ * A</td><td>7. 8</td></t<> | L cotype"> pessible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 | M type generation 75.dat * (CDS.pos / C D transcript D transcript D transcript D transcript D transcript | \$ * % N ad from the re AA.pos / AA PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 | O Coding | P P alternate alle ERRORS / V 2-iar 2-iar | Q Q es given the ARNINGS / II c.1729_173 c.1773A>T c.4420,442 c.4432C>G c.4455C>A | R R ample ploidy (FO' "> p.Asp577Pro p.Jy39JAan p.Thr1474651 p.Thr1474631 | 1729/6492 1773/6492 4432/6492 4432/6492 | T 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 | ∑ * A | 7. 8 |
| | A1 49 3eroto 50 ieroto 51 "Read 52 "Ref 53 "Sur 56 #"Alt 55 #"Sur 56 #S/0 58 #_D 50 zipts 60 zipts 61 208/c 63 64 65 66 67 68 69 70 71 | Frequencies and a second secon | alibri (Body B I U fr A G Phred-scaled envation course enference of reatrume_ID t e /scratch/r FILTER - - - |) + 12 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII | + A + A + A + A + A + A + A + A + A + A | A · J J all probabilit enome /scrat HGVS.c tory/004/413 MODERATE a MODERATE a MODERATE a MODERATE a MODERATE a MODERATE | K y of the called gene genotype for each HOT | L type"> possible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 | M https://www.internationality.com/ https://wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww | S + % N N AA.pos / AA PF3D7_0100 PF3D7_010 PF3D7_0100 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_0100 PF3D7_000 | O eference and Distance Coding | P alternate alle ERRORS / V 2-iar 2-iar 2-iar | Q Q les given the ARNINGS / II C.1729_173 C.1773A-T C.4420Cd C.4432Cd C.443Cd C.4432Cd | R R ample ploidy #O' "> 0_Asp577Pro p_Lys591Aan p_OIn1478G p_OIn147463Ly p_Lmr1439Ly p_Lap1552Ar p_Lap259_Aan | 1729/6492 5 5 5 5 5 5 5 5 5 5 5 5 | T T T T T T T T T T T T T T T T T T T | ∑ * A | 7. .a |
| | A1 49 Senot 50 ienot 51 "Reas 52 "Ref 53 "Sur 54 "Atte 55 "Sur 55 "Sur 56 is of 51 | Pype Quality, the pype Quality, the pype Quality, the pype Quality, the pype Quality of the pype Quality of the of quality of the renter allele observed of Quality, dthe perfected "> point" QUAL 168.615 59.2743 112.1419 112.945 70.7189 20.1132 149.70189 20.1132 2 | alibri (Body B I U G G Phred-scaled Phred-scaled Phred-scaled Phred-scaled G Phred-scaled G Phred-scaled G Phred-scaled G Phred-scaled G Phred-scaled C Phred-scaled Phred-scaled Phred-scaled Phred-scaled |) + 12 Hfileformal marginal (or ikelhoods of nt"> bservations"> tr> bservations"> bserva | A A | all probabilitien the called enome /scrat HGVS.c tory/004/413 unknoema MODERATE | ************************************ | L otype"> possible geno cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 | M type generative 75.dat * (CDS.pos / C D transcript D transcript D transcript D transcript D transcript D transcript D transcript D transcript | \$ * % N ed from the re PF3D7_0100 | O O O O | P alternate alle ERRORS / V 2-jar 2-jar 2-jar 2-jar 2-jar 2-jar | C.1729_173 C.1729_173 C.1773A-T C.4420_442 C.4455T>G C.4455T>G C.4733_473 C.44731_45 | R ample ploidy eFO' > p.Asp577Pro p.Ly591An p.Drh 17460 p.Gln147#GG p.Gln147#GG p.Gln147#GG p.Gln1575PAI p.Asp575PAI | * * * * * * * * * * * * * * * * * * * | Cormat * T T T T T T T T T | ∑ * A | 7. .a |
| | A1 49 3eroto 50 ieroto 51 "Read 52 "Ref 53 "Sur 56 #"Alt 55 #"Sur 56 #S/0 57 ies/0 58 #S/0 50 zipts 60 zipts 61 208/c 63 64 65 66 67 66 67 71 | Frequencies and a second secon | alibri (Body B I U fst A G Phred-scaled leavation course e alternate of 7.dat Plasmo(Fieture_iD t - e /scratch/i FiLTER |) + 12 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII | A A A A A A A A A A A A A A A A A A A | A J J all probabilition in the called HGVS.c tory/004/612 unknown a MODERATE a MODERATE | K y of the called gene genotype for each HOT | L type"> possible genc cDNA.pos / PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 PF3D7_010 | M http://www.internationality.com/ http://wwww.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://wwww.internationality.com/ http://wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww | S + % N ed from the m AA.pos / AA PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 | O eference and Distance Coding | P alternate alle ERRORS / V 2-lar 2-lar 2-lar 2-lar 2-lar | C.1729_173 C.1729_173 C.1773A-T C.4432C-6 C.4452C-6 C.4455C-A C.4455C-A C.4453C-6 C.4455C-6 C.4453C-6 C.4453C-6 C.4455C-6 C.4455C-6 C.4453C-6 C.4455C-6 C.445C-6 C | R R ample ploidy #O' "> 0_Asp577Pro p_Lys591Aan p_OIn1478G p_OIn147463Ly p_Lmr1439Ly p_Lap1552Ar p_Lap259_Aan | s s s s s s s s s s s s s s s s s s s | Delete * Format * T 1729/6492 1773/6492 4420/6492 4450/6492 4450/6492 4450/6492 4450/6492 4450/6492 4450/6492 4450/6492 4751/6492 5647/6492 | ∑ * A | 7. .a |
| | A1 E 49 3enoto 50 jenoto 51 "Reas 52 "Ref 53 "3un 56 ani" 55 ar, 56 ar, 56 ar, 50 jenoto 56 ar, 50 jenoto 56 ar, 50 jenoto 56 ar, 50 jenoto 50 jenoto 50 jenoto 51 "Reas 55 ar, 50 jenoto 50 jenoto 51 ar, 52 "Reas 60 jenoto 53 ar, 54 ar, 55 ar, 55 ar, 50 jenoto 55 ar, 50 jenoto 55 ar, 50 jenoto 50 jenoto | Frequencies and a second secon | alibri (Body B I U fr k G Fred-scaled Information course everation course eve |) + 12 Hilleformat H Haraginal (or ikelhoods of nt's bervations''s bervatis''s bervations' | A A A A A A A A A A A A A A A A A A A | A J J Jail probabilities in the called in the called in the called interv/OD4/431 MOCBATT MOCBATT MOCBATT MOCBATT MOCBATT MOCBATT MOCBATT MOCBATT | # # K # Y of the called generative for each # Information # Y of the called generative for each # Y of the called genera | L otype"> possible genc cDNA.pos/ cDNA.pos/ PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 PF307_010 | M http://www.internationality.com/ http://wwww.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://www.internationality.com/ http://wwww.internationality.com/ http://wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww | \$ + % N ed from the re PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 PF3D7_0100 | O O O O | P alternate alle ERRORS / V 2-lar 2-lar 2-lar 2-lar 2-lar 2-lar 2-lar 2-lar 2-lar 2-lar | ARNINGS / II c.1729_173 c.1773A-T c.4432C-6 c.4432C-6 c.4432C-6 c.4432-6 c.5473-6 c.5473-6 c.5472-6 c.547-6 c.447-6 c.547-6 | R R ample ploidy #FO' "> p.Asp577Pro p.Up591Am p.Jup591Am p.Thr1480H p.Am1852Ar | 1729/6492 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 | Cormat * T T T T T T T T T | ∑ * Z Z Z Z Sori/2163 Soyl/2163 Soyl/2163 1478/2163 1478/2163 158/2163 158/2163 | 7. .a |

- 3. Now you can look for Gene IDs of interest in the excel file. For example, if this is a known drug resistant line you can find the gene(s) that might be responsible for the resistance and see what kinds of SNPs are present.
- 4. If you are comparing a mutant and a wild type or two different strains you can extract gene IDs from both VCF files and use a website like http://bioinfogp.cnb.csic.es/tools/venny/



*Note that in the above steps you are ultimately comparing gene IDs – do you think you might be missing some important polymorphisms using this method? Of course, the answer is yes[©]

It is quite possible that a gene with a SNP in the WT and a SNP in the mutant that will be in the intersection of the two gene lists, contains different SNPs – you will miss this by doing the above steps. Below is a description of steps you can take to create a list of unique IDs for SNPs. This list of unique IDs can then be used in Venny.

- 1. Start with the same excel files that you opened in the above section.
- 2. To create a unique ID for SNPs we will combine information from multiple columns to create something that looks like this: chromosome:position:geneID
- 3. To do this you will use the concatenate function in Excel: =concatenate(cell#1,":",cell#2,":",cell#3)

Cell#1 = cell with chromosome number

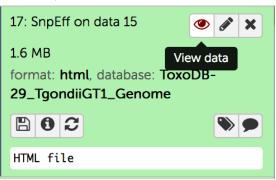
Cell#2 = cell with position Cell#3 = cell with GeneID

| · · · · | Calibri (| Body) | • 12 • | A- A- | | | × ∎ ∎ | => | Genera | al | Ŧ | ± * | 1. | 2. |
|--|--|---|--|--|---|--|---|---|--|--|-----------------------------|------------------------------------|---------|---------------|
| Paste 🧹 | B I | <u>U</u> • | | | | | •= | • • | \$ - | %) | 00, 0 • 0. | Conditional Formatting | | Cell tyles |
| NCATE 🗘 | × < 1 | x =CON | | E(A65.":".E | 65,":",K65 | | | | | | | | | |
| | • • J | 001 | | | t2], [text3], [te | wt41 Ftovt | 51 [toyt6] | | | | | | | _ |
| A | | | | | <u>121, [[ext3], [[i</u> | <u>ext4</u>], <u>[text</u> | 5], [lexi6], |) н | 1 | J | | К | B | s |
| ##SnpEffCmd= | | | | | 00/datasat 01/ | 7 dat Taual | DB 20 Trend | ICT1 Case | na /aastab | /aalauu /6laa /00 | D /datasa | 0105 dat " | | |
| | | | | | | | | | | HGVS.c | | | | |
| | | | | | | | Feature_ID | Transcript | | HGV3.C | HGVS. | , | | |
| ##INFO= <id=l0< td=""><td></td><td></td><td></td><td></td><td>_transcripts_af</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=l0<> | | | | | _transcripts_af | | | | | | | | | |
| ##INFO= <id=n< td=""><td></td><td></td><td></td><td></td><td>_transcripts_af</td><td>rected' "></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=n<> | | | | | _transcripts_af | rected' "> | | | | | | | | |
| ##SnpSiftVersio | | | | | | | | | | | | | | |
| ##SnpSiftCmd= | | | | | lataset_8106.d | at -e /scratc | h/galaxy/job_ | _working_di | rectory/004, | /4169/tmpBopc | fU" | | | |
| ##FILTER= <id=< td=""><td>SnpSift,D (AN</td><td>N[*].IMF (</td><td>FILTER = 'PA</td><td>.SS')))"></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></id=<> | SnpSift,D (AN | N[*].IMF (| FILTER = 'PA | .SS')))"> | | | | | | | | | | |
| #CHROM | POS | 10 | 0 | REF | ALT | QUAL | FILTER | INFO | FORMAT | unknown | | | | |
| TGGT1_chrla | I | 227230 | | A | С | 1156.55 | | AB=0;ABP= | 0 missense | va MODERATE | TGGT1 | 293300 | 65,":", | K65) |
| TGGT1_chrla | | 1340271 . | | G | с | 2387.77 | | AB=0;ABP= | 0; missense | va MODERATE | TGGT1 | 295040 | | |
| TGGT1 chrla | | 1396177 . | | A | с | 387.162 | | AB=0;ABP= | 0 missense | va MODERATE | TGGT1 | 295125 | | |
| TGGT1_chrlb | | 78769 . | | A | G | 1780.8 | | AB=0;ABP= | 0 missense | va MODERATE | TGGT1 | 207440 | | |
| TGGT1 chrlb | | 153771 . | | т | G | 1414.57 | | | | va MODERATE | | | | |
| TGGT1 chrlb | | 276348 | | т | G | 2066.14 | | | - | va MODERATE | | - | | |
| | | | | | | | | | | | | | | |
| TGGT1 chrlb | | 622140 | | G | C | 2335.06 | | $\Delta R = \Omega \cdot \Delta R P =$ | 0 missonso | Va MODERATE | TGGT1 | 208310 | | |
| TGGT1_chrlb | | 622140 . | | G | C T | 2335.06 | | | | va MODERATE | | | | |
| | | 622140 . 1446003 . 1446022 . | | G C G | C T T | 2335.06 60.6579 82.4046 | | AB=0;ABP= | 0; missense | _va MODERATE _va MODERATE _va MODERATE | TGGT1 | 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . | | C G | T T | 60.6579 | | AB=0;ABP= AB=0;ABP= | 0; missense | va MODERATE | TGGT1 | 209755B 209755B | Form. | at 🔻 🧳 |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . | | C G | T T | 60.6579 82.4046 | | AB=0;ABP= AB=0;ABP= | 0, missense 0, missense | va MODERATE | TGGT1 TGGT1 | 209755B 209755B | Form | at * 🧳 |
| TGGT1_chrlb TGGT1_chrlb | ■ 1 <u>⊍</u> ✓ fx | 1446003 . 1446022 . | D | C G | T T | 60.6579 82.4046 | | AB=0;ABP= AB=0;ABP= | 0, missense 0, missense | va MODERATE | TGGT1 TGGT1 | 209755B 209755B | Form: | at * 🧳 |
| TGGT1_chrlb TGGT1_chrlb C C C C C C C C C C C C C C C C C C C | ■ 1 ⊆ √ fx | C C 10-03, by F | D D | C G | T T F | 60.6579 82.4046 | • • • | AB=0;ABP= AB=0;ABP= + /0 | 0, missense 0, missense 2 .oc | va MODERATE va MODERATE | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C 1446003 . 1446022 . C 10-05, by F stats /scrat | D ratio cingorar ch/galaxy/file | C G E E so/008/datase | T T T F t_8107.dat Toxo | 60.6579 82.4046 = | HilGT1_Genome | AB=0;ABP= AB=0;ABP= P /0 | 0, missense 0, missense 1, .oc | va MODERATE | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auto Crigorar ch/galaxy/file ion_ Gene_N | C G E E so/008/datase lame Gene_II | T T F t_8107.dat Toxo Feature_Ty | 60.6579 82.4046 = | • • • | AB=0;ABP= AB=0;ABP= P /0 | 0, missense 0, missense 2 .oc | va MODERATE va MODERATE | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | ■ 1 ⊆ fx Eff -1vcf -0vcf umb Gene_ID | C C C C C C C C C C C C C C C C C C C | D auto Cingolary Ch/galaxy/file ion_Gene_N _of_Percent, | C G "E ts/008/datase ame Gene_II _of_transcrip | F t_8107.dat Toxo Feature_Ty ts_affected" > | 60.6579 82.4046 = | HilGT1_Genome | AB=0;ABP= AB=0;ABP= P /0 | 0, missense 0, missense 1, .oc | va MODERATE | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auto Cingolar ch/galaxy/file ion_ Gene_N _of_ Percent _of_ Percent | C G G E E S/008/datase ame Gene_II _of_transcrip _of_transcrip | T T F t_8107.dat Toxo Feature_Ty | 60.6579 82.4046 = | HilGT1_Genome | AB=0;ABP= AB=0;ABP= P /0 | 0, missense 0, missense 1, .oc | va MODERATE | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auto Cingolar ch/galaxy/file ion_Gene_N _of_Percent 03), by Pablo | C G s s/008/datase ame Gene_IC _of_transcrip Cingolani" | T T F t 8107.dat Toxo D Feature_Ty ts_affected' "> ts_affected' "> | 60.6579 82.4046 G DB-29_Tgonc F Feature_ID | H H Transcript_E | AB=0;ABP= AB=0;ABP= | 0; missense 0; missense J axy/files/008 HGVS.c | va MODERATE va MODERATE • • • • • • • • • • • • • • • • • • • | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auro cmgorar ch/galaxy/file ion_Gene_N _of_Percent, _of_Percent, _of_Percent, _of_percent, _of | C G s s/008/datase ame Gene_IC _of_transcrip Cingolani" | T T F t 8107.dat Toxo D Feature_Ty ts_affected' "> ts_affected' "> | 60.6579 82.4046 G DB-29_Tgonc F Feature_ID | H H Transcript_E | AB=0;ABP= AB=0;ABP= | 0; missense 0; missense J axy/files/008 HGVS.c | va MODERATE va MODERATE • • • • • • • • • • • • • • • • • • • | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auro cmgorar ch/galaxy/file ion_Gene_N _of_Percent, _of_Percent, _of_Percent, _of_percent, _of | C G s s/008/datase ame Gene_IC _of_transcrip Cingolani" | T T F t 8107.dat Toxo D Feature_Ty ts_affected' "> ts_affected' "> | 60.6579 82.4046 G DB-29_Tgonc F Feature_ID | H H Transcript_E | AB=0;ABP= AB=0;ABP= | 0; missense 0; missense J axy/files/008 HGVS.c | va MODERATE va MODERATE • • • • • • • • • • • • • • • • • • • | TGGT1 TGGT1 | 209755B 209755B | | |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . 10-05), uy r stats /scrat Number Number Number Id 2015-10- f /scratch/g Wf (FILTER ID | D auro cingoiar ch/galaxy/file ion_Gene_N _of_Percent, _of_Percent, _of_Percent, _alaxy/files/OP alaxy/files/OP | C G s/008/datass of_transcrip of_transcrip Cingolani" 18/dataset_8: | T T T F t_8107.dat Toxo D Feature_Ty ts_affected "> L06.dat -e /scrate | 60.6579 82.4046 B2.4046 G DB-29_Tgonc Feature_ID Filese Filese Filese | H H IllGT1_Genom Transcript_E working_dire | AB=0;ABP= AB=0;ABP= , a /0 , a | 0; missense 0; missense 0; missense 0; missense 0; daxy/files/008 HGVS.c 69/tmpBopq unknown | va MODERATE va MODERATE • • • • • • • • • • • • • • • • • • • | TGGT1 TGGT1 | 209755B 209755B | BS | |
| TGGT1_chrlb TGGT1_chrlb A SnpEffCmd=Snpl INFO=cID=ANN_N INFO=cID=ANN_N SnpSiftCmd="Snp FILTER=cID=SnpSi HROM | fx fx b fx fx b | C C C C C C C C C C C C C C C C C C C | D auto Cmgotasy/fiel ion_Gene_N _of_Percent, _of_Percent _of_Percent alaxy/files/0C = "PASS"))"> REF | C G s/008/datase lame Gene_II of_transcrip of_transcrip Cingolani* I8/dataset_8: ALT | T T T F st_8107.dat Toxo D Feature_Ty ts_affected "> ts_affected "> ts_affected "> ts_affected "> | 60.6579 82.4046 G G DB-29_Tgonc F Feature_ID Fh/galaxy/job FILTER 5. | H H IliGT1_Genom Transcript_E INFO AB=0;ABP=0; | AB=0;ABP= AB=0;ABP= , ap 70 , | 0; missense 0; missense 0; missense 1 0; missensensensensensensensensensensensensen | va MODERATE va MODERATE •••• Form K /dataset_8105.d HGVS.p | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | | C C C C C C C C C C C C C C C C C C C | D auto cmgorar ch/galaxy/file on_Gene_N _of_Percent _of_Percent _of_Percent _of_Percent _of_Percent _alaxy/files/00 _alaxy/fil | C G G s:s/008/datase ame Gene_II of_transcrip of_transcrip of_transcrip I28/dataset_8: ALT C | T T T T T T T T T T T T T T T T T T T | 60.6579 82.4046 = | H HIIGT1_Genomu Transcript_E INFO AB=0;ABP=0; AB=0;ABP=0; | AB=0;ABP= AB=0;ABP= I e /scratch/ga Rank ctory/004/41 FORMAT missense_va | 0, missense, 0, missense, 0, missense, 1, oc 1, | va MODERATE va MODERATE b • .0 Cond Form K /dataset_8105.d HGVS.p | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . 1446022 . 1446022 . 1446022 . 1446022 . 1446022 . 1446022 . 1446022 . 1446022 . 1446003 . 1446003 . 1446003 . 1446022 . 144602 . 144602 . 14602 | D auto cingorar ch/galaxy/file onGene_N ofPercent, 03), by Pablo alaxy/files/00 = 'PASS')))'> REF A G | C G s s/008/datase ame Gene_II of_transcrip of_transcrip Cingolani" 18/dataset_8: ALT C C | T T T t_ t8107.dat Tox0 Feature_Ty ts_affected "> ts_affected "> | 60.6579 82.4046 = | H H HIIGT1_Genomm Transcript_E INFO AB=0;ABP=0; AB=0;ABP=0; | AB=0;ABP= AB=0;ABP= , p /0 i e /scratch/ga Rank ctory/004/41 FORMAT missense_va missense_va | 0; missense; 0; missense; 0; missense; 1; axy/files/008; HGVS.c 69/tmpBopq unknown MODERATE MODERATE | va MODERATE va MODERATE Form k //dataset_8105.d HGVS.p U" TGGT1_293300 TGGT1_295040 | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . 144602 . 144602 . 14602 . 14600 . 14602 . 14600 . 146000 . | D raud Cingolary/file th/galaxy/file ion_Gene, N _of_Percent, _of_Percent, _of_Percent, _of_Percent, _of_Percent, _of_REF _A _G _A | C G S S/008/dataset ame Gene_IL of_transcrip of_transcrip of_transcrip Cingolani" 18/dataset_8: ALT C C C | T T T F F F F F F F F F F F F F C C C C | 60.6579 82.4046 = | H HIGT1_Genom Transcript_E INFO AB=0;ABP=0; AB=0;ABP=0; AB=0;ABP=0; | AB=0;ABP= AB=0;ABP= , , , , , , , , , , , , , , , , , , , | 0; missense; 0; missense; 0; missense; 1 axy/files/008 HGVS.c 69/tmpBopq unknown MODERATE MODERATE MODERATE | va MODERATE va MODERATE | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | | 1446003 . 1446022 . 144602 . 14 | auto Enigotar ch/galaxy/file on_Gene_N_ | C G G w s/OB/datase lame Gene_II of_transcrip cof_transcrip cof_transcrip Cingolani* B/dataset_8: ALT C C C G | T T T T t s 107.dat Toxcor t s affected "> t s affected "> t s a t s a t a t a t a t a t a t a t a | 60.6579 82.4046 g DB-29_Tgonc c FFeature_ID FILTER 5. 7. 2. 3. 7. | H H HIGT1_Genomi Transcript_E INFO AB=0;ABP=0; AB=0;ABP=0; AB=0;ABP=0; AB=0;ABP=0; AB=0;ABP=0; AB=0;ABP=0; | AB=0;ABP= AB=0;ABP= , 20 , 20 , 20 , 20 , 20 , 20 , 20 , 20 | 0; missense; 0; missense; 0; missense; 1; axy/files/008; HGVS.c 69/tmp8opq unknown MODERATE MODERATE MODERATE | va MODERATE va MODERATE 9 • 0 Form K K /dataset_8105.d HGVS.p U" TGGT1_29300 TGGT1_295125 TGGT1_295125 | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | Image: Second | 1446003 . 1446022 . 144602 . 144600 . 144602 . 144602 . 144600 . 1 | D auro Crigoara ch/galaxy/file of_ Percent _of_ Percent _alaxy/files/0C alaxy/files/0C alaxy/files/0C REF A G A A T T T G G | C G G s s s y S y OB/dataset amc Gene_II of_transcrip of_transcrip of_transcrip c of_transcrip Cingolani" DB/dataset 8: ALT C C C C G G G G G G G G G G G G C | T T T T t_ t_8107.dat Toxo D Feature_Ty ts_affected">> 006.dat = / scrats 006.dat = / scrats 0.006.dat = / scrats | 60.6579 82.4046 = | H HIGT1_Genom Transcript_E INFO AB=0;ABP=0;ABP=0; AB=0;ABP=0 | AB=0;ABP= AB=0;ABP= | 0 missense 0 missense 0 axy/files/008 HGVS.c 69/tmpBopq unknown MODERATE MODERATE MODERATE MODERATE | va MODERATE va MODERATE ************************************ | TGGT1 TGGT1 | 2097558 2097558 Table Styles | BS | |
| TGGT1_chrlb TGGT1_chrlb | P 1 ≤ fx fx b fx | 1446003 . 1446022 . 144602 . 144602 . 144602 . 144602 . 144602 . 144602 . 14602 . 14 | D auto cmgorar ch/galaxy/file of_gence.N _of_percent, _of_percent, _of_percent, _af | C G s s/008/datase ame Gene_IL of_transcrip of_transcrip Cf_transcrip | T T T T t t, \$107.dat Toxo D F Feature 7, b ts_affected "> ts_affected "> ts_affected "> ts_affected "> 106.dat e /scratt 1156.5 2387.7 387.16 1144.5 2066.1 | 60.6579 82.4046 = | H HilioT1_Genomm Transcript_E INFO AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, AB=0,ABP=0, | AB=0;ABP= AB=0;ABP= I e /scratch/ga Rank ctory/004/41 FORMAT missense_va missense_va missense_va missense_va | 0 missense 0 missense 1 J axy/files/008 HGV5.c 69/tmpBopq MODERATE MODERATE MODERATE MODERATE | va MODERATE va MODERATE ************************************ | TGGT1 TGGT1 atting as | 2097558 2097558 Table Styles | BS | |

- 4. You should get unique SNP IDs that look like this (for example): TGGT1_chrIb:1446003:TGGT1_209755B
- 5. Copy this function to the rest of the column to replicate the concatenate function.
- 6. Copy the these newly generated unique IDs into Venny and compare the mutant and wild type.

Examining SnpEff summary:

- Click on the view icon (eye) in the SnpEff output file that has the html format.



- This will open the html file right in galaxy where you can view it.
- The header contains a short summary and information about the run and it has several major components:
- Summary table that warns about possible genomic annotation errors or inconsistencies identified in the reference genome. If there are many, use caution interpreting results and examine associated gff files for any issues (ex. missing feature values in gff files, incomplete gene sequences, more than one stop codon per gene, etc.).
- 2. Summary statistics for variant types

Number variantss by type

| Туре | Total |
|----------|---------|
| SNP | 114,034 |
| MNP | 12,864 |
| INS | 6,907 |
| DEL | 7,304 |
| MIXED | 2,180 |
| INTERVAL | 0 |
| Total | 143,289 |

Here is an example of variant calls and what they mean in terms of nucleotide changes:

| Туре | What is means | Example |
|-------|----------------------------------|--------------------------------------|
| SNP | Single-Nucleotide Polymorphism | Reference = 'A', Sample = 'C' |
| Ins | Insertion | Reference = 'A', Sample = 'AGT' |
| Del | Deletion | Reference = 'AC', Sample = 'C' |
| MNP | Multiple-nucleotide polymorphism | Reference = 'ATA', Sample = 'GTC' |
| MIXED | Multiple-nucleotide and an InDel | Reference = 'ATA', Sample = 'GTCAGT' |

3. Statistics for the variant effects and impacts:

EXON

INTRON

NONE

INTERGENIC

TRANSCRIPT

UPSTREAM

SPLICE_SITE_ACCEPTOR

SPLICE_SITE_DONOR

SPLICE_SITE_REGION

| Type (alphabetical order) | Count | Percent |
|---------------------------|--------|---------|
| MISSENSE | 21,588 | 35.949% |
| NONSENSE | 131 | 0.218% |
| SILENT | 38,332 | 63.832% |
| | | |
| | | |
| | | |
| Type (alphabetical order) | Count | Percen |

67,505

74,749

1,064

1

5

4

176

333,432

12

8.451%

9.358%

0.133%

0.001%

0.001%

0.022%

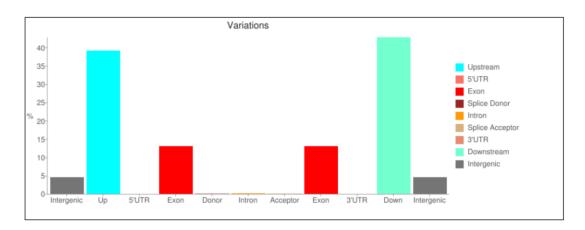
0.002%

41.741%

0%

Number of effects by functional class

Base changes summary. SnpEffhtml files provides a break down of SNPs across gene features:



The SNP workflow you are using is set up to generate certain files that will provide you with the information you can export and use further in your analysis (yellow stars).

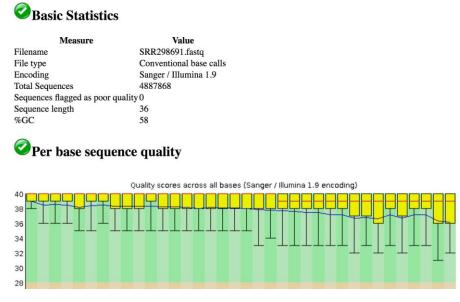
If you select certain options they will be shown in your history. If you do not select to display these files, you can view the output by clicking on displaying the hidden files from the history menu:

| History | €\$□ |
|---|-------------|
| search datasets | 8 |
| G2 C albicane SC5314 YP (paired) 10 shown, 22 <u>hidden</u> 19.39 GB | D and serum |

Now, lets take a look at the files generated by the workflow and steps that you can take to further evaluate them.

1. Examine sequence quality based on FastQC quality scores.

FastQC provides an easy-to-navigate visual representation sequencing data quality and distribution of nucleotides per read position.



2. Download vcf files and evaluate workflow results.

The vcf file generated by SnpEff contains information about SNPs and the genomic location.

| #CHROM | POS | ID | REF | ALT | QUAL | FILTER | INFO | FORMAT | unknown | | | |
|----------|---------|----|------|------|---------|--------|------------|------------|--------------|----------------|---------------|----------|
| CM001231 | 189057 | | AG | СТ | 787.449 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:143:0:0 | 143:5341:-20 | 7.887,-43.047 | 3,0 |
| CM001231 | 483825 | | G | A | 64.8756 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:4:0:0:4: | 146:-10.0999,- | -1.20412,0 | |
| CM001231 | 518226 | | G | С | 51.7908 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:8:0:0:7: | 276:-11.5007,- | 2.10721,0 | |
| CM001231 | 574021 | | С | G | 237.265 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:17:0:0:1 | 7:583:-39.079 | ,-5.11751,0 | |
| CM001231 | 609879 | | GAA | CAG | 55.2785 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:32:8:27 | 7:22:861:-18.1 | 711,-0.69473 | 5,0 |
| CM001231 | 1090073 | | G | т | 79.4156 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:8:2:75:6 | 5:238:-11.5539 | ,-1.36362,0 | |
| CM001231 | 1090104 | | Α | т | 70.961 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:6:0:0:6: | 220:-12.5146,- | -1.80618,0 | |
| CM001231 | 1153611 | | CCTC | GCTG | 111.123 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:8:5:188 | 3:97:-9.30616 | ,-6.1461,0 | |
| CM001231 | 1159150 | | СТ | GC | 126.126 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:31:0:0:1 | 9:741:-29.771 | 3,-5.71957,0 | |
| CM001231 | 1159438 | | с | G | 82.3312 | | AB=0;ABP=0 | GT:DP:RO:Q | 0/0:47:30:1 | 092:17:640:0,- | 9.53002,-3.5 | 0705 |
| CM001231 | 1159465 | | G | с | 249.656 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:126:47: | 1770:79:3013: | -53.8644,-25 | 2134,0 |
| CM001231 | 1159499 | | т | с | 124.95 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:143:32: | 1167:111:4248 | 8:-76.1575,-3 | 3.4865,0 |
| CM001231 | 1181576 | | CC | TG | 191.675 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:27:0:0:2 | 5:924:-41.744 | 8,-7.52575,0 | |
| CM001231 | 1293309 | | с | G | 51.22 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:2:0:0:2: | 78:-6.92763,-0 | .60206,0 | |
| CM001231 | 1323058 | | Π | GC | 71.3001 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:6:0:0:6: | 223:-12.5485,- | -1.80618,0 | |
| CM001231 | 1485397 | | Α | G | 3558.42 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:499:0:0 | :497:18671:-8 | 04.678,-149.6 | 12,0 |
| CM001231 | 1485429 | | G | A | 3783.33 | | AB=0;ABP=0 | GT:DP:RO:Q | 1/1:517:1:3 | 8:516:20010:- | 843.425,-151 | 978,0 |

Post-processing of SNP data is normally required to make sense of thousands of SNPs and to decide which ones have biological and functional importance. Data processing can help you to extract SNP distribution and parse associated data including GeneIDs, protein-coding annotations, and effects in sequence ontology terms such as missense or synonymous variants, stop codon gain, etc. and also link changes to the genome model.

Summary

| Genome | ToxoDB-29_TgondiiGT1_Genome |
|--|--|
| Date | 2017-06-17 05:56 |
| SnpEff version | SnpEff 4.11 (build 2015-10-03), by Pablo Cingolani |
| Command line arguments | <pre>SnpEff -i vcf -o vcf -stats /scratch/galaxy/files/008/dataset_8107.dat ToxoDB-29_TgondiiGT1_Genome /scratch/galaxy/files/008/dataset_8105.dat</pre> |
| Warnings | 3,941 |
| Errors | 0 |
| Number of lines (input file) | 8,411 |
| Number of variants (before filter) | 8,483 |
| Number of not variants (i.e. reference equals alternative) | 0 |
| Number of variants processed (i.e. after filter and non-variants) | 8,483 |
| Number of known variants (i.e. non-empty ID) | 0(0%) |
| Number of multi- allelic VCF entries (i.e. more than two alleles) | 72 |
| Number of effects | 14,149 |
| Genome total length | 63,945,332 |
| Genome effective | |