

Summarize Cohort Mutations


Variant information is stored on a per sample basis, but it can be informative to view variants in the context of recurrent variants identified within the project's sample cohort to identify both the frequency of variants and the samples that share a particular variant. The *Summarize cohort mutations* task can be invoked from any *Variants* or *Annotated variants* data node to generate a report of shared variants identified from detection against a reference sequence or among paired samples.

Summarize cohort mutations dialog

The *Summarize cohort mutations* task, user needs to specify *Minimum coverage for genotype calls*. In general, it is likely that if a variant is not called in a sample at a particular locus then the sample has a homozygous reference genotype. Yet this may not always be the case as factors such as insufficient depth or low quality bases at that locus may lead to an inability of the variant caller to identify any genotype at that locus. As such, setting a minimum coverage will make the assumption that the sample contains a homozygous reference genotype if the depth requirement is met. This is done for the purpose of generating genotype calls for all samples (even reference homozygotes) at all variant loci within the project.

For paired variant caller reprot, if *Merge pairs* check button is unselected, pairs will be analyzed separately. If it is selected, all samples will be analyzed together.

Cohort mutation summary report

The **Cohort mutation summary report** provides a row in the table for all variant sites, either SNVs or INDELs, identified in the project (Figure 2). Hovering over a column header will provide a brief description of the column data. Columns presented in the table include the following information: *View* provides a link to [Chromosome View](#) by selecting the chromosome icon , *Chr* represents chromosome from the reference assembly, *Position* represents the base position in the chromosome, *Mutation type* is the category of variant (Substitution for SNVs and Insertion or Deletion for INDELs), *Reference allele* is the base(s) in the reference assembly sequence, *Case genotypes* are the genotypes of the samples with a variant at the locus, *Variant frequency* represents the frequency of the variant site in the sample cohort, *Sample count* is the fraction of samples in the cohort with the variant, and *Samples* are the names of the samples that contain the variant. The *Summarize cohort mutations* task is not available for variants detected by [LoFreq](#) as no genotypes are produced from the caller. If variant detection was performed on paired samples in [Samtools](#) (Figure 3), the *Genotype* column will be replaced with four columns: *GT* *Change* presents the possible change in zygosity between cases and controls at the variant locus, *Control Genotypes* are the genotypes of the designated control samples in the pairs, and *Case Genotypes* are the genotypes of the cases in the pairs. Additional columns can be added to the *Cohort mutation summary report* table by selecting *Optional columns*. The optional columns are dependent upon the information present in the underlying vcf file and include variant and sample metrics from variant detection and information from the annotation. Hovering over a term in the list will provide a brief description of the data contained in that column. *Optional columns* can also be used to exclude default columns in the table.









Optional columns									
	View	Chr	Position	MutationType	Reference allele	Case genotypes	Variant frequency	Sample count	Samples
		Search...	Search...	Search...	Search...	Search...	Search...	Search...	Search...
1		10	23319553	Substitution	A	AG	.50	1/2	2
2		10	23319559	Substitution	A	AG	.50	1/2	2
3		10	76790198	Deletion	CCA	REF/ALT	.50	1/2	1
4		11	19213968	Insertion	A	REF/ALT	.50	1/2	2
5		11	19214005	Insertion	T	REF/ALT	.50	1/2	2
6		11	58378422	Substitution	G	GA	.50	1/2	1
7		12	97073612	Insertion	T	REF/ALT	.50	1/2	1
8		12	97073619	Deletion	TC	REF/ALT	.50	1/2	1

Figure 1. Example of the Cohort mutations summary table (truncated) for variants detected against a reference sequence

◇ GT Change	◇ Control genotypes	◇ Case genotypes	◇ Control/reference match
Search...	Search...	Search...	Search...
HET->HET	TG	TG	no
HET->HOM	AG	GG	no
HET->HOM	TC	CC	no
HET->HET	CG	CG	no
HET->HOM	TC	CC	no
HET->HOM	GA	AA	no

Figure 2. Additional columns added to the Cohort mutation summary report for variants detected by Samtools paired analysis

Below each data column header in the *Cohort mutation summary report*, the *Search...* section allows for filtering of the table (Figures 2, 3 and 4). The search can be useful for limiting the list of variants to those of interest when large numbers of variants are present in the table. For columns with numbers, exact values or ranges using either ">" or "<" can be utilized in the search. For columns with letters or words, and exact string of characters must be entered in order to obtain a match. In the case of table cells with multiple entries, there must be an exact match between the query and 1 entry to retain the table row.

If the *Summarize cohort mutations* task is performed upon an *Annotated variants* data node, additional information can be presented in the *Cohort mutation summary report* table. Annotation with known variants from a variant database will add the *ID* column, which provides the identifier of the variant within the database (Figure 4). Annotation with gene/features will allow for the following columns to be added using *Options* above the top right corner of the table (Figure 5): *Type* provides the detailed category of variant relative to a gene model (3-prime UTR, 3-prime UTR indel, 5-prime UTR, 5-prime UTR indel, Intron, Intron indel, missense, Nonsense, Non-coding RNA, Non-coding RNA indel, Promoter, and Synonymous), *Gene Symbol* provides the HUGO gene name, *Transcript* provides the annotation model transcript identifier, *Strand* provides the stranding of the gene, *Gene section* provides the location of the variant in the gene model (Exon, Intron, Promoter), *Nt change* provides the location and base change of the variant for the gene model, and *AA change* provides amino acid changes produced by coding variants. It may be the case that a cell has multiple entries due to the overlap of information in the annotation model. For example, a gene may have multiple transcripts and these will show up as multiple entries within a cell. Annotation with [SnpEff](#) and [VEP](#) will add additional columns to the table, some of which may be redundant with other annotations.

Optional columns										
	View	Chr <div>Search...</div>	Position <div>Search...</div>	ID <div>Search...</div>	MutationType <div>Search...</div>	Reference allele <div>Search...</div>	Case genotypes <div>Search...</div>	Variant frequency <div>Search...</div>	Sample count <div>Search...</div>	Samples <div>Search...</div>
1		10	23319553	.	Substitution	A	AG	.50	1/2	2
2		10	23319559	.	Substitution	A	AG	.50	1/2	2
3		10	76790198	.	Deletion	CCA	REF/ALT	.50	1/2	1
4		11	19213968	.	Insertion	A	REF/ALT	.50	1/2	2
5		11	19214005	.	Insertion	T	REF/ALT	.50	1/2	2
6		11	58378422	rs79135825	Substitution	G	GA	.50	1/2	1
7		12	97073612	.	Insertion	T	REF/ALT	.50	1/2	1
8		12	97073619	.	Deletion	TC	REF/ALT	.50	1/2	1

Figure 3. Addition of the ID column in the Cohort mutation summary report when variants are annotated with a known database

Optional columns

☒ Case genotypes

☒ Variant frequency

☒ Sample count

☒ Samples

RefSeq Transcripts - 2016-05-02 Columns

☐ Type

☐ Gene symbol

☐ Transcript

☐ Strand

☐ Gene section

☐ Nt change

☐ AA change

samtools Columns

☐ 1000 Genome phase 3

☐ AC

☐ AC1

☐ AF1

☐ AN

☐ ASP

☐ ASS

◊ MutationType

Search...

◊ Reference allele

Search...

◊ Case genotypes

Search...

◊ Variant frequency

Search...

◊ Sample count

Search...

◊ Samples

Search...

Substitution

A

AG

.50

1/2

2

Substitution

A

AG

.50

1/2

2

Deletion

CCA

REF/ALT

.50

1/2

1

Insertion

A

REF/ALT

.50

1/2

2

Insertion

T

REF/ALT

.50

1/2

2

Substitution

G

GA

.50

1/2

1

Insertion

T

REF/ALT

.50

1/2

1

Deletion

TC

REF/ALT

.50

1/2

1

Deletion

GC

REF/ALT

.50

1/2

2

Insertion

T

REF/ALT

.50

1/2

1

Insertion

A

REF/ALT

.50

1/2

2

Deletion

AG

REF/ALT

.50

1/2

2

Substitution

A

CC

.50

1/2

1

Deletion

CCA

REF/ALT

.50

1/2

2

Deletion

GCT

ALT/ALT

.50

1/2

2

Insertion

A

REF/ALT

.50

1/2

1

Substitution

C

CT

.50

1/2

1

Deletion

GA

REF/ALT

.50

1/2

1

Substitution

C

CT,TT

1.00

2/2

1, 2

Substitution

G

GA

.50

1/2

1

19

17

19713740

rs150122

20

17

39728050

rs8070680

At any point, information in the *Cohort mutation summary report* table can be saved in text or vcf format by selecting *Download* at the bottom right corner of the table. If the table is exported in text format, the visible table will be appended with additional columns for all samples in the project. These columns specify the genotype call for each variant locus in the project. In instances where no variant was detected within a sample, the value specified by *Minimum coverage for genotype calls* in the task dialog will be used to call either a homozygous reference genotype if above the specified threshold or no genotype if below the specified threshold.

If you need additional assistance, please visit [our support page](#) to submit a help ticket or find phone numbers for regional support.



Your Rating:  Results:  39 rates