

SEQUENCHER™

Tutorial for Windows and Macintosh

Translated Variance Table

© 2007 Gene Codes Corporation

Gene Codes Corporation

T C A G E N E
A G T C O D E S

Gene Codes Corporation
775 Technology Drive, Ann Arbor, MI 48108 USA
1.800.497.4939 (USA) +1.734.769.7249 (elsewhere)
+1.734.769.7074 (fax)
www.genecodes.com info@genecodes.com

Translated Variance Table

Creating a Translated Variance Table from a Single Contig.....	3
Changing the Translation Range	6
Changing the exemplar sequence	7
Creating a Translated Variance Table from Multiple Contigs	7
Reviewing the data	8
Changing the Region of Interest	10
Sister tables.....	12
Conclusion	14

Translated Variance Table

The Translated Variance Table enables you to compare the differences between your exemplar (model or primary) sequence and other sequences you select. It is the ideal tool for scanning and evaluating your findings at the amino acid level. The Translated Variance Table directs your attention to the putative effects a base difference has on a codon and its associated amino acid. It saves you time and increases your accuracy in analyzing SNPs, mutations, heterozygotes or other significant coding changes.

You use the Translated Variance Table to compare differences among translations of sequences in the same contig. This type of Variance Table will be explored in the first part of the tutorial.

You use the Translated Variance Table to compare differences between the translations of consensus from multiple contigs that have all been assembled to a common Reference Sequence. This type of Variance Table will be explored in the second part of the tutorial.

In this tutorial you will learn how to use the Translated Variance Table and how to interpret its results. We will describe how to perform these functions using Sequencher's menu commands and button bars. You may also invoke commands using your right mouse button. The sample data provided in the tutorial project has already been trimmed and **Assembled by Name** to a Reference Sequence using the command **To Reference by Name**.

CREATING A TRANSLATED VARIANCE TABLE FROM A SINGLE CONTIG

The first table you are going to generate will show you what happens when you compare sequences within the same contig to an exemplar. Once you start working with your data you will need to decide which sequence will act as the exemplar and which sample sequences you want to compare to it.

- **Launch Sequencher to open an Untitled Project.**
- **Go to the File menu and select Import > Sequencher Project...**
- **Navigate to the Sample Data folder inside the Sequencher application folder.**
- **Select the Translated Variance Table project from the Sample Data folder and select Open.**

The Translated Variance Table project in this example contains 13 contigs. One contig contains exon data from 10 individuals. The other contigs contain multiple exon data from 12 individuals and will be used in the second half of this tutorial.

Sequencher gives you three choices for the exemplar sequence. You can choose the **Reference Sequence**, the **Consensus** sequence or the sequence at the top of your contig sequence list.

- **Choose Contig > Consensus Inclusively.**
- **Select the contig called Exon 11.**
- **Go to the Sequence menu and select Compare Translation To > Reference Sequence.**

The Translated Variance Table appears. Note that in some cases, if the names of the individual sequences are long, they appear truncated in the table view.

Description: 11 sequences compared to Reference Reference with 3 exons of contig Exon 11.
Comparison Range: Unfiltered
Base Positions: 14587..18495
Amino Acid Positions: 1..1303

Reference		Ref...	94	138	152	93	116	130	150	128	156	131	Total
17,458	AGC	AGC	AGC	AGC	AGG	AGC	AGS	AGS	AGS	AGS	AGG	AGG	7
958	S	S	S	S	R	S	?	?	?	?	R	R	
17,770	ACT	ACT	ACT	AMT	ACT	ACT	ACT	ACT	AMT	AMT	ACT	ACT	3
1,062	T	T	T	?	T	T	T	T	?	?	T	T	
17,785	TGG	TGG	TGA	TGA	TGA	TGR	TGA	TGA	TGA	TGA	TGA	TGA	10
1,067	W	W	.	.	.	?	
⚡	Total	0	1	2	2	1	2	2	3	3	2	2	20

In the table, the column order is the same as the original contig order of your sequences. To interpret the table you need to remember that each row represents a base position and each column represents a sample. The first row contains the names of the sequences. The first column shows the position of each variant found as a base position and a codon position. The cell in the equivalent position in the last column of the table indicates the total number of variants for that position. The first row, position 17,458 and codon 958, has 7 variants out of 11 samples the second has 3 and the third has 10. The last row of the table shows the total number of variants for each sample.

The total number of variants in the entire table is given in the cell at the bottom right hand corner of the table. In this table the total number of variants is 20.

You will notice pink headers in your Translated Variance Table. These indicate that the sequences in those columns do not span the entire length of the exemplar sequence. As you explore the tables you generate with your own data, you will notice that any time a sample sequence does not have a base for the equivalent position in the exemplar, the cell will have a pink cross in it.

As you browse through the table you will notice that some cells have lettering in black while others are in gray. Cells with black lettering contain variants. Cells with gray lettering are identical to the Reference Sequence. The difference in letter color helps you quickly focus on the variants as you scan the table.

- Scroll to the top of the table.
- For Sample Column 152, locate the variant at base position 17,458 which is equivalent to residue 958.

The sample sequence contains the codon AGG, as opposed to the Reference Sequence, which has an AGC codon at this position. This change substitutes an Arginine for the original Serine. This difference could prove to be interesting since Serine is a smaller polar amino acid while Arginine is a larger strongly basic amino acid.

Reference		Ref...	94	138	152	Total
17,458	AGC	AGC	AGC	AGC	AGG	7
958	S	S	S	S	R	

You can sort the table to see which sequence has the most variants.

- At the bottom left hand corner of the table is a Total button. Click on it once.

When you do this, the table will be rearranged. The sequences with the most variants will be on the left hand side. Those with the least will be on the right hand side.

Reference		128	150	131	156	130	116	152	138	93	Total
17,458 958	AGC S	AGS ?	AGS ?	AGG R	AGG R	AGS ?	AGS ?	AGG R	AGC S	AGC S	7
17,770 1,062	ACT T	AMT ?	AMT ?	ACT T	ACT T	ACT T	ACT T	ACT T	AMT ?	ACT T	3
17,785 1,067	TGG W	TGA .	TGR ?	10							
⚡	Total	3	3	2	2	2	2	2	2	1	20

Now suppose you wanted to see which sequence had the fewest variants.

- Click on the same **Total** button again.

The table will be rearranged again, only this time the sequences with the fewest variants will be on the left hand side.

Reference		Ref...	94	93	138	152	116	130	156	131	Total
17,458 958	AGC S	AGC S	AGC S	AGC S	AGC S	AGG R	AGS ?	AGS ?	AGG R	AGG R	7
17,770 1,062	ACT T	ACT T	ACT T	ACT T	AMT ?	ACT T	ACT T	ACT T	ACT T	ACT T	3
17,785 1,067	TGG W	TGG W	TGA .	TGR ?	TGA .	TGA .	TGA .	TGA .	TGA .	TGA .	10
⚡	Total	0	1	1	2	2	2	2	2	2	20

You can sort the rows of the table the same way clicking on the Total button in the top right hand corner of the table.

It is often the case that you have very long sequence names. You can resize the column widths all at once or individually.

- At the bottom left hand corner of the table is a resize button . Click on it once.

Reference		Reference with 3 exons			94	93	138	152	116	130	Total
17,458 958	AGC S	AGC S			AGC S	AGC S	AGC S	AGG R	AGS ?	AGS ?	7
17,770 1,062	ACT T	ACT T			ACT T	ACT T	AMT ?	ACT T	ACT T	ACT T	3
17,785 1,067	TGG W	TGG W			TGA .	TGR ?	TGA .	TGA .	TGA .	TGA .	10
⚡	Total	0			1	1	2	2	2	2	20

- Click on the resize button again.

Reference		...	94	93	Total	
17,458 958	AGC S	AGC S	AGC S	AGC S	AGC S	AGG R	AGS ?	AGS ?	AGG R	AGG R	AGS ?	AGS ?	7
17,770 1,062	ACT T	ACT T	ACT T	ACT T	AMT ?	ACT T	ACT T	ACT T	ACT T	ACT T	AMT ?	AMT ?	3
17,785 1,067	TGG W	TGG W	TGA .	TGR ?	TGA .	10							
⚡	Total	0	1	1	2	2	2	2	2	2	3	3	20

- Click on it one more time to return to the original state.

Notice how all the columns were resized each time you clicked the button. Use this button any time you want to check on the sequence name in the column headers or to squeeze more columns into the current table view.

- Move your cursor over the line between two column headers until the cursor changes shape.
- Click and hold down the mouse button.
- Move the mouse left or right while holding down the mouse button.

Notice how the column on the left hand side was resized as you moved the cursor.

Reference		...	94	93	...	152	Total	
17,458 958	AGC S	AGC S	AGC S	AGC S	AGC S	AGG R	AGS ?	AGS ?	AGG R	AGG R	AGS ?	AGS ?	7
17,770 1,062	ACT T	ACT T	ACT T	ACT T	AMT ?	ACT T	ACT T	ACT T	ACT T	ACT T	AMT ?	AMT ?	3
17,785 1,067	TGG W	TGG W	TGA .	TGR ?	TGA .	10							
+	Total	O	1	1	2	2	2	2	2	2	3	3	20

CHANGING THE TRANSLATION RANGE

The great power of the Translated Variance Table lies in its ability to use sequence features to restrict the range of comparison between the exemplar and sample sequences. In the first section of this tutorial the comparisons were performed along the entire length of the Reference Sequence. In the next section of this tutorial you will use a sequence feature to restrict the Comparison Range.

The Reference Sequence being used here, as the exemplar, is a stretch of genomic sequence from chromosome 22. It has a number of features, represented by different feature keys, marked on it. Each feature has several attributes, including a name, a location, and a feature key. You will use the CDS or coding sequence feature to restrict the **Translation Range** and direct your focus to variants that result in changes to the translation.

- Click on the **Translation Range** button in the **Translated Variance Table** button bar.
- Click the **Filter Translation by** radio button.
- Choose **CDS** from the **Feature Key:** drop down menu.
- Choose **HPS4a** from the window below the drop down menu.
- Click **OK**.

You will see that a new Translated Variance Table has been generated. This table is very different from the first table. There are now 5 rows of differences rather than the original unfiltered 3. The differences reported in this new filtered table accurately reflect the joined reading frames of the three exon features in the Reference.

Whenever you import a sequence from GenBank, it will include a feature table that Sequencher can read. You may have marked features on your exemplar yourself. In both cases, you can use the **Translation Range** function to restrict the comparison and focus on differences between your sequences in detail.

Notice that in some cells there is a question mark (?) instead of an amino acid residue. If you examine the codon in these cells you will see that one of the normal bases (A,G,C,T) has been replaced by an ambiguity code. For example in several samples at base position 17, 586 there is an M in the first position of the codon. M is the IUPAC code for A or C. The equivalent codons would be ACG or CCG. Later in the tutorial you will learn how to review this data and distinguish between true SNPs and base calling errors.

- **Close the Translated Variance Table.**

CHANGING THE EXEMPLAR SEQUENCE

Sequencher gives you three choices for your exemplar: the Reference Sequence, the consensus or the top sequence in your list. You can use any sequence as your exemplar simply by simply dragging it to the top of your list of sequences in the Contig Editor and choosing the **Compare Translation To > Top Sequence** command.

- Double Click on contig **Exon 11** to open the Contig Editor.
- Click on the **Bases** button.
- Maximize the Contig Editor.
- Locate the sequence **131** in the list at the bottom left hand side of the Contig Editor.
- Click on the name of this sequence.
- Note that the hand icon closes as you click.
- While holding the mouse button down drag the sequence name to the top of the list, above the Reference.
- **Select > Select All**, to select the names of all of the sequences.
- Go to the **Sequence** menu and select **Compare Translation To > Top Sequence**

The Translated Variance Table opens. Notice that the **Description** at the top of the table tells you which sequences you used and which sequence was the exemplar in this case.

Description: 11 sequences compared to sequence 131 of contig Exon 11.

Comparison Range: Unfiltered

Base Positions: 17288..17847

Amino Acid Positions: 1..186

Top Sequence		131	Ref...	94	138	152	93	116	130	Total
17,459 58	GGT G	GGT G	GCT A	GCT A	GCT A	GGT G	GCT A	GST ?	GST ?	8
17,585 100	CAC H	CAC H	CAC H	CMC ?	CMC ?	CAC H	CAC H	CAC H	CAC H	4
17,648 121	TCT S	TCT S	TCT S	TCT S	TYT ?	TCT S	TCT S	TCT S	TCT S	3
17,714 142	CAG G	CAG G	CAG G	CAG G	CMG ?	CAG G	CAG G	CAG G	CAG G	1
⬆	Total	0	2	2	5	0	2	1	1	21

You are now able to explore putative differences from the perspective of your new exemplar. The ability to change to a different exemplar is particularly useful in any type of comparative sequence analysis.

- Close the Translated Variance Table.
- Quit Sequencher

CREATING A TRANSLATED VARIANCE TABLE FROM MULTIPLE CONTIGS

In many instances you may want to work with data from different sample sources. Perhaps you are in a laboratory setting and are screening samples from different sources. Perhaps you are studying evolutionary changes among related species. You may be simply re-sequencing a cloned gene and want to make sure that no changes have occurred in a coding region. The commonality among these cases is that you have an established prototype sequence and are working with the consensus sequences from different contigs.

In the next section of the tutorial you will generate a **Translated Variance Table** using the **Compare Translation to Reference** command from the **Contig** menu. You will be working with an established an established prototype sequence and are working with the consensus sequences from different contigs.

- **Launch Sequencher.**
- **Go to the File menu and select Import > Sequencher Project...**
- **Navigate to the Sample Data folder, choose the Translated Variance Table project and select Open.**

The project contains 12 contigs from separate individuals. The auto-sequencing data comes from different human samples and spans three exons of a gene in chromosome 22.

When you import a Sequencher Project all of the imported are selected.

- **Use Apple + Click on a Mac or Ctrl + Click on a PC to deselect the Exon11 contig.**

Remember that the original sample sequences were assembled **To Reference by Name** so each contig already carries a copy of the Reference Sequence.

- **Go to the Contig menu.**
- **Choose Compare Translation to Reference.**

The Translated Variance Table appears in a new window. The column order is the same as the original contig order. Each column represents a contig. The contig name is contained in the column header. To interpret the table you need to remember that each row represents a consensus position but only displays this in black where there is a difference. The two left-most columns display the base position of the first base in the variant codon and the number of the variant amino acid. The right-most column displays the total number of variants for that position.

Notice the **Description** tells you how many consensus sequences are in the table and indicates the Reference Sequence to which they were compared. In most respects the Translated Variance Table has the identical descriptions, sorting and navigation functions as the Variance Table.

Description: 12 consensus sequences compared to Reference Reference with 3 exons.
Comparison Range: Unfiltered
Base Positions: 14587..18495
Amino Acid Positions: 1..1303

The information included in the table header provides information about the contents of the Translated Variance Table. The **Description** gives the total number of samples, sequences or contigs, included in the comparison and indicates which sequence is the exemplar. In this example, the exemplar sequence is called "Reference with 3 exons". The **Comparison Range** provides the status of the comparison range filter. The default setting for all Translated Variance Tables is Unfiltered. The table header also displays both the base and amino acid positions included in the comparison.

REVIEWING THE DATA

Now look at your data in detail. You will notice that each of the header cells is shaded in pink. This indicates that the sequence does not cover the full comparison range. In addition, some of the cells contain a pink X. This indicates that there is no base coverage for the sequence at this position.

FGTGGTTGGGACAGCTTTCTGGGCC**TG:G**CTAAGTTTTATGACAGATGGCACAGCAARTTCCTGATGGTTAGGTCA
GGC**C**AGTGCTAAGTTTTATGACAGATGGCACAGCAARTTC**C**TGATGGTTAGGTCA

70 | 15380 | 15390 | 15401..1 | 15410 | 15420 | 15430 | 15440
.....GGCC**AG**TGCTAAGTTTTATGACAGATGGCACAGCAARTTCCTGATGGTTAGGTCA

- Use the left arrow key to move through the table.

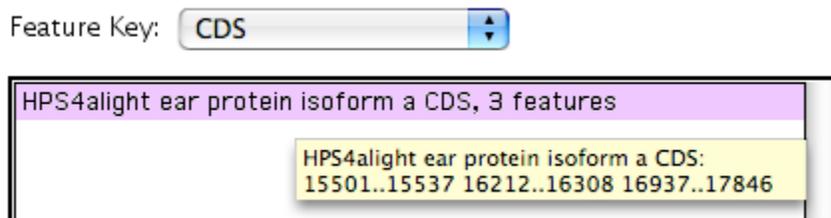
As you do this you will see the information in the Contig Editor and the Contig Chromatograms change. Remember each column represents the consensus for a single contig. As you move from cell to cell you are seeing the underlying data for each contig at that base position. Cells that contain a pink X will have no chromatogram data.

- Note the similarity in the data in each of the samples with coverage at base position 15,400.
- Close the Contig Editor and Contig Chromatogram windows.

CHANGING THE REGION OF INTEREST

When a GenBank sequence is imported into Sequencer for use as a Reference Sequence, its Feature Table is read and any joined features such as CDSs or Exons are noted. When you first construct a Translated Variance Table, the default comparison range is the full length of the exemplar sequence. In some instances this will be exactly what you want. However, the sequences used in this tutorial example are derived from genomic DNA and the Reference Sequence is marked up with the relevant coding regions. It would be more scientifically relevant to restrict the table to these regions.

- Click on the **Translation Range** button in the **Translated Variance Table** button bar.
- Click the **Filter Translation by** radio button.
- Choose **CDS** from the **Feature Key:** drop down menu.



The only CDS feature is selected in the Feature drop down box.

- Hover your mouse over **Feature HPS4 a light ear protein isoform a CDS, 3 features**, to trigger a tool tip that describes the feature and the location of the three coding regions.
- Click **OK**.

You will see that the Translated Variance Table has been refreshed. Note the changes in the header of the table. The **Comparison Range** has changed from Unfiltered to the name of the feature. Notice that **Base Positions:** now lists three discontinuous ranges, equivalent to the three CDS features in the Reference Sequence.

Description: 12 consensus sequences compared to Reference Reference with 3 exons.
Comparison Range: Filtered by HPS4alight ear protein isoform a CDS
Base Positions: 15501..15537 16212..16308 16937..17846
Amino Acid Positions: 1..348

The filtered Translated Variance Table reduces the original number of differences from 1928 to a total of 46 variants. Look at your data in detail. Find the base position with the most variants by looking at the right most column.

- Click on the **Total** button in the top right hand corner.

The table is now sorted so that the position with the most variants is in the top row. The total number of variants for this position (base position 15,516) is 12. In the majority of cases the predicted change is from Glutamic Acid to Glycine. Remember that in this table you are looking at consensus sequences.

- Select the top cell in sample 131 and double-click on it.

In the Translated Variance Table you have **GGG** representing the codon for Glycine. This has been substituted for the original Glutamic acid residue in the Reference Sequence. The acidic, polar residue has been replaced by a smaller, neutral, non-polar residue.

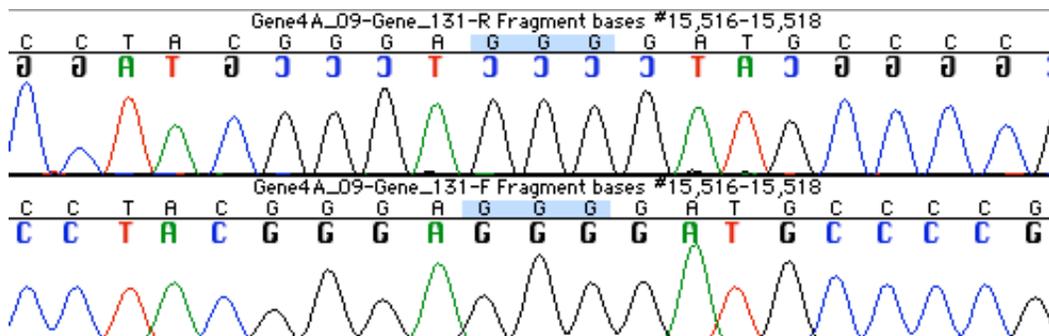
Looking at the Contig Editor you can see the Reference Sequence at the top of the editor window. The codon **GAG** (for Glutamic Acid) is highlighted in yellow, but the sample consensus sequence has **GGG** highlighted in yellow. Your selection extends from the first base of this codon down through the sequences providing coverage in the contig and through to the consensus sequence. You can see that the first "G" consensus call is supported by both sequences at this position in the contig and the Reference. In the second base of the codon the Reference has an "A", but the sample sequence has a "G". The background for both supporting bases is pale blue, indicating that these are high quality base calls.

```
CAGagactccctacgggagaggatgccccgcaggaacatgGTA
CAGAGACTCCCTACGGGAGGGGATGCCCCGCAGGAACATGGTA
CAGAGACTCCCTACGGGAGGGGATGCCCCGCAGGAACATGGTA
```

```
15500 15510 15520 15530
CAGAGACTCCCTACGGGAGGGGATGCCCCGCAGGAACATGGTA
```

The traces in the Contig Chromatograms window provide further support. You have two overlapping sequences. One is in the forward orientation and one is in the reverse orientation. Notice that the original base calls in the top chromatogram, displayed in color below the current base calls, are in mirror image. This is a visual cue indicating that the sequence is in the reverse orientation.

There are three clear peaks for the "GGG" codon at this position for both sequences. You have confirmed the variant.



In the reference sequence, the variant base is in red, distinguishing this position as a previously characterized variant. More details on the variant are available from the **Sequence > Feature Listing** or from the **Sequence > Edit Features...** commands.

- In the Contig Editor click on the name of the Reference Sequence.
- Choose **Sequence > Feature Listing**.
- Scroll to the **Variation** feature at position **15517**.

```
Name: variation
Key: variation
From: 15517 To: 15517 Strand: Complement Color: Red Style: Underline
complement(15517)
/gene="HPS4"
/replace="c"
/replace="t"
/db_xref="dbSNP:713998"
```

This Feature Listing indicates that you have found a new "G" variant. The previous variants at this position included only "C" and "T" substitutions.

- Close the **Feature Listing** window.

SISTER TABLES

Throughout this tutorial you have used the Contig Editor and the Contig Chromatograms command to support the findings you have made using the Translated Variance Table. You can simultaneously view the sister Variance Table alongside the Translated Variance Table, and you can generate Sister Tables with or without the editors available in Review mode.

- Go to the **Translated Variance Table** button bar.
- Click on the **Bases** button.

Arrange the windows to suit your viewing preference. If you would like the windows to be arranged this way every time you use the Sister Tables in Review mode, follow these steps.

- Go to the **Window** menu.
- Choose **Remember Window Layout > Double Variance Table Review**.

Now you can see four windows on your screen, the Variance Table, the Translated Variance Table, the Contig Editor and the Contig Chromatogram. There are several advantages to this layout. The Variance Table displays not only the individual bases but also shows any feature information. This is displayed as a color and text style. You control default styles through the User Preferences. You can also apply individual styles.

- Go to the Bases **Variance Table**. Its window is entitled **Compare Consensus to Reference**.
- Click on the **Comparison Range** button in the button bar.
- Check the **Filter Comparison By:** radio button.
- Choose the **Feature Key: CDS**.
- Click the **OK** button.
- Select the cell at base position 17,158 in sample 116.

This action opens the Contig Editor and the Contig Chromatogram for this sample and this position. The font of the base is lower case, underlined and green, indicating that the base is part of an exon, but the lack of red color indicates that this is the position of a new variant. With your selection still in the Variance Table, column 116 row 17, 158:

- Choose **Sequence > Mark Selection As Feature**.
- Check the **Feature Key: variation**.
- Change the **Feature Name: to variation new**.
- Change the **Feature Color: to cyan**.
- Click **OK**.

The style is immediately applied to the selected base and to the underlying sequence data in contig 116. You can see the name of the feature appear below the consensus line in the Contig Editor.



- Go to the **Translated Variance Table**. Its window is entitled **Compare Translation to Reference**.
- Locate the column header called 116. Click on the column header.
- Go to the **Edit** menu. Choose **Label > Label 2**.

This will change the text color of the header in both Variance Tables, giving you a visual cue about the status of the sample. You can set up names for the labels in the User Preferences.

Reference		116	Total	Reference		116	Total
15,517	A	<u>g</u>	12	15,516	GAG	GGG	12
16,277	C	<u>y</u>	5	6	E	G	12
17,061	G		1	16,277	CTC	YTC	5
17,158	A	<u>r</u>	1	35	L	?	5
17,229	G		1	17,061	GAT	GAT	1
17,353	G	<u>k</u>	1	87	D	D	1
17,419	C		1	17,157	CAC	CRC	1
17,460	C	<u>s</u>	7	119	H	?	1
17,586	A		4	17,229	GAA	GAA	1
17,649	C		3	143	E	E	1
17,715	A		1	17,352	AGC	AKC	1
17,771	C		3	184	S	?	1
17,787	G	<u>a</u>	10	17,418	TCT	TCT	
⚡	Total	6	50	⚡	Total	6	46

CONCLUSION

The Variance Table gives you a summary analysis of your data that focuses on differences. Not every base difference will cause a change in the amino acid. The Translated Variance Table lets you see the impact of a base change immediately and check the validity of any codon change you observe in the underlying sequence data. Working with the Translation Range allows you to focus on a feature of scientific interest such as a coding sequence (CDS), even when it is interrupted by regions of non-coding sequence.

The Translated Variance Table saves you time and allows you to double-check your work as you go along, making your efforts to track variants more efficient and relevant to your project.

You can find out more about the Variance Table and the Translated Variance Table in the manual and other tutorials in this series.