Loupe™ V(D)J Browser Tutorial

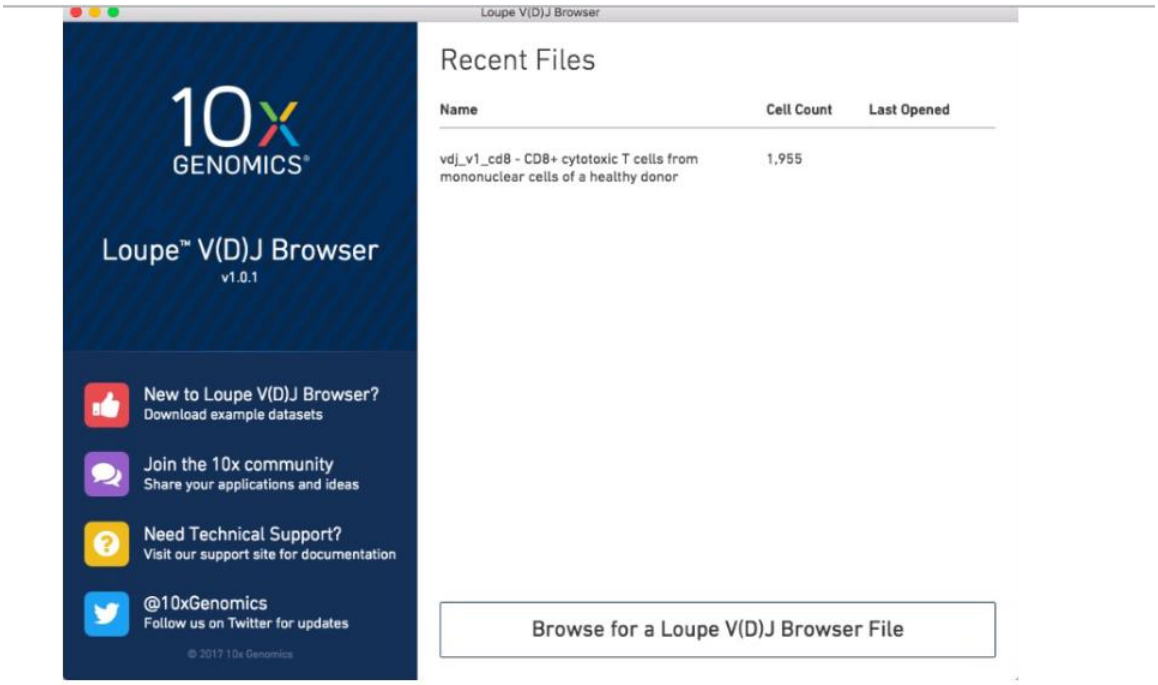
Here are the instructions for using Loupe V(D)J Browser. The example sample used for analysis is a CD8+ cytotoxic T-cell sample analyzed using the Cell Ranger pipeline. Loupe V(D)J Browser allows for in-depth analysis of the following:

1. Clonotype abundance analysis
2. V-J gene abundance analysis
3. CDR3 sequence analysis and mutation analysis
4. Validity of Cell Ranger's V(D)J spliced consensus sequences

1. Configuration
Installation package:

1. Windows: Loupe-VDJ-Browser-1.0.2.exe
2. Mac: Loupe-VDJ-Browser-1.0.2.dmg

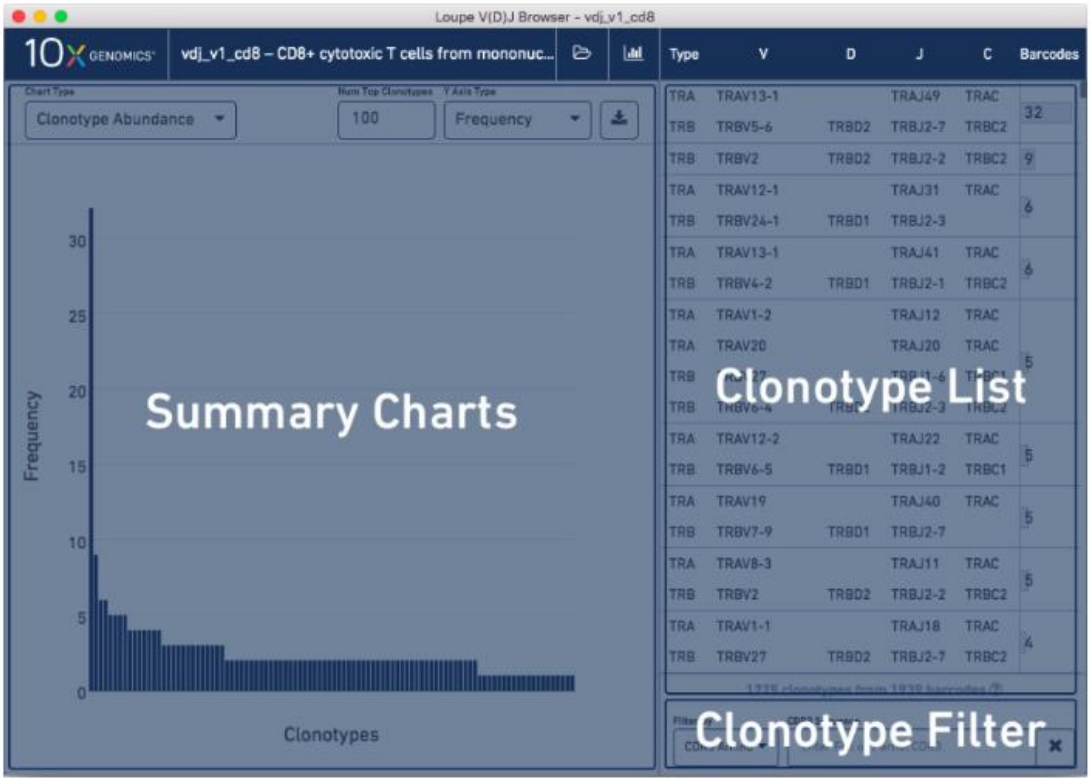
After installing, open the Loupe V(D)J Browser homepage. You can import files by clicking on 【Browse for a Loupe V(D)J Browser File】, as shown below:



Loupe™ V(D)J Browser: User Interface

After importing the data, you will enter the user interface.

2. Summary
   The Loupe V(D)J Browser interface consists primarily of two panels. The left panel is the content panel, and the right panel displays the clonotype list.



Content Panel: This panel is used for interaction and data display. Upon loading the dataset for the first time, the content panel displays the top 100 clonotypes based on their abundance distribution. Above the charts, there are controls for changing the chart type, specific chart options, and exporting to an image.

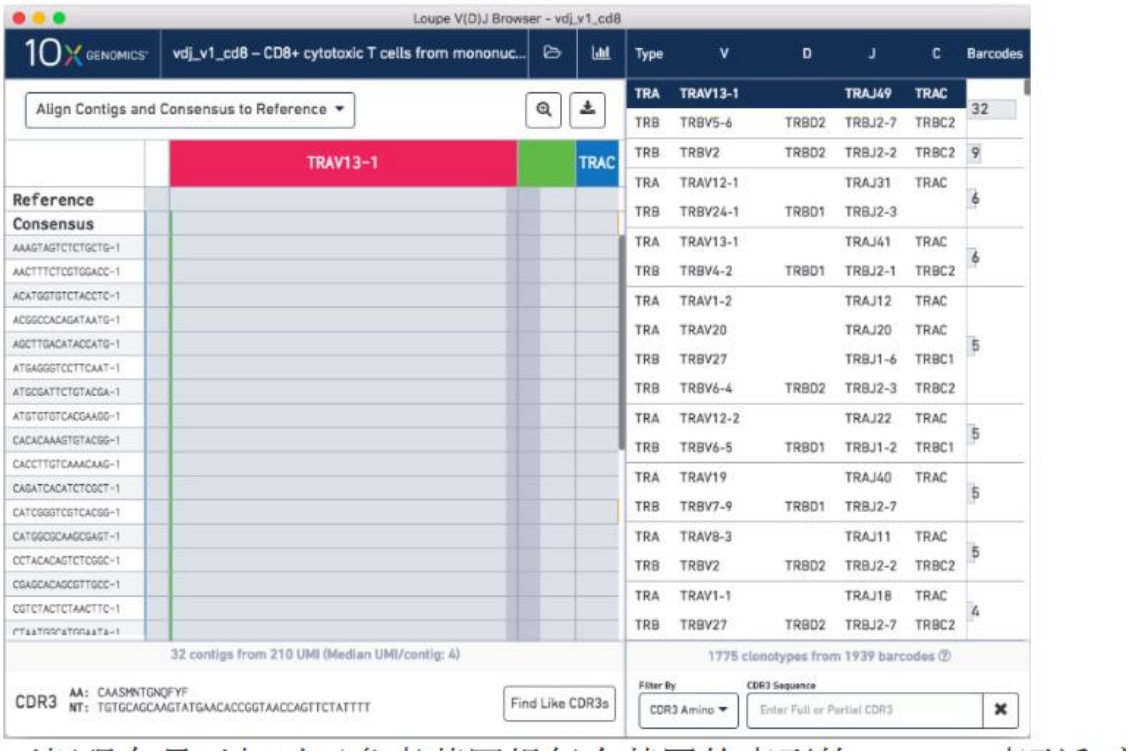
Clonotype List: Initially, it shows the list of all clonotypes in the CD8+ sample. Each clonotype contains one or more T cell receptor chains. From left to right in the list, the clonotype displays the chain type (α or β), V gene annotation, D gene annotation, J gene annotation, C gene annotation, and cell barcode count. In the rightmost column, there is a clonotype barcode count indicating the frequency of the clonotype in the sample. Hovering the mouse over the barcode count reveals the frequency of the clonotype.

Clonotype Filtering: Clonotypes can be filtered using the following options: 1. Gene annotation, 2. CDR3 amino acid sequence, 3. CDR3 nucleotide sequence, and 4. Barcode. We will explain how to use this filtering by listing the clonotypes in the next section.

3. Navigation Bar
Above the content panel, you can see the name of the currently opened file and two icons. Clicking the folder icon will take you back to the file selection window. Clicking the chart icon will return to the summary view.

4. Chain Information View

One of the main features provided by Loupe V(D)J Browser is the Chain Information View. It allows you to see how the contig sequences of each cell are spliced to form the α or β chain V(D)J spliced consensus sequences of a clonotype. You can view this information by clicking on any chain in the clonotype list on the right side. For example, if you click on the α chain (TRAV13-1/TRAJ49/TRAC) of the top clonotype, you should see the following content:



The content panel now displays the contig sequences and variants relative to the reference genome for each gene. By observing this, you can verify the contig sequence consistency within a clonotype and detect SNPs. The Chain View section will provide more detailed explanations of this view.

You can now return to the summary charts by clicking the chart icon at the top of the window. With the instructions provided above, you can now explore the clonotypes of the sample in more depth.

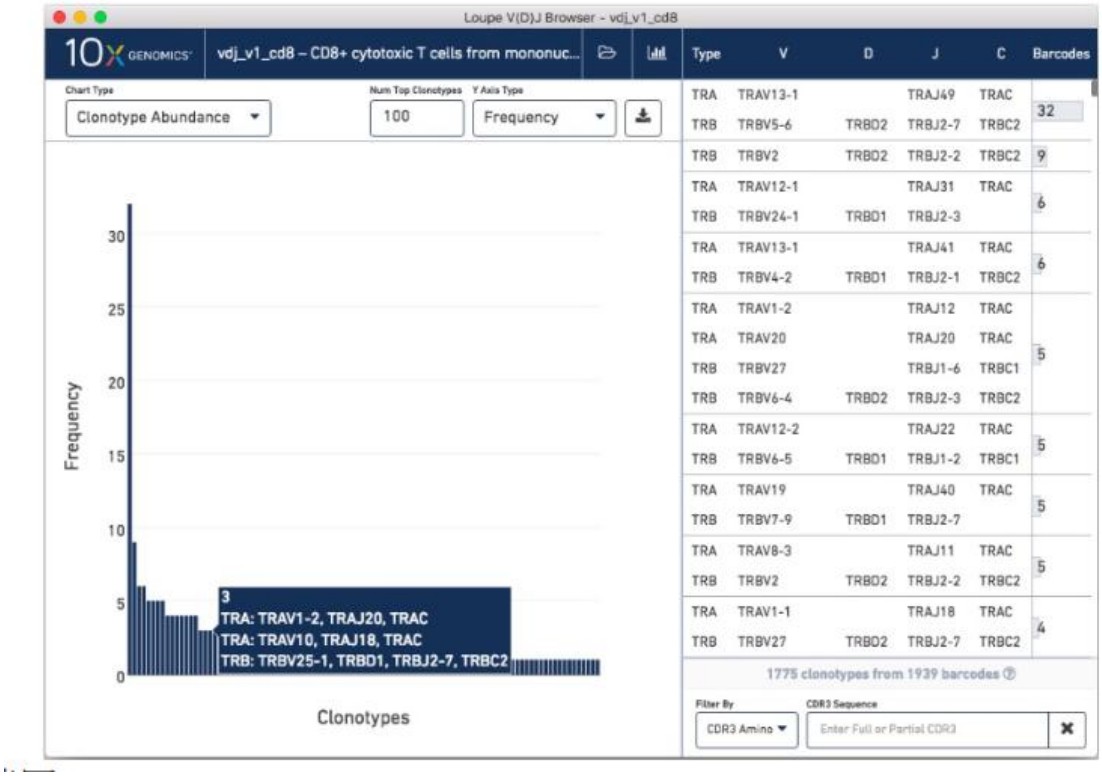
Loupe™ V(D)J Browser: Exploring Clonotypes

Loupe V(D)J Browser includes multiple tools to study the diversity of clone types in a sample. It allows for precise analysis of individual clones, genes, motifs, and even individual cell barcodes. The diversity of the data can be examined from a high level using the summary charts. Additionally, a combination of tools can be used to apply multiple criteria for clonotype filtering.

I. Summary Charts
There are four different charts that provide insights into the diversity of a cell sample:

1. Clonotype abundance: This chart displays the most abundant clonotypes. You can adjust the number of clonotypes shown by changing the value of 'Num Top Clonotypes.' The 'Y Axis Type' option allows you to display the clonotype (or gene) cell support count as either abundance or percentage. Hovering over a bar reveals specific information about the clonotype.
2. V gene usage: This chart shows the usage of V genes. It provides insights into the distribution and abundance of different V genes in the sample.
3. J gene usage: This chart shows the usage of J genes. It provides insights into the distribution and abundance of different J genes in the sample.
4. V-J gene heatmap: This heatmap displays the frequency of V-J gene combinations. It offers a visual representation of the diversity of V-J gene pairings.

Each chart has a set of options and the ability to export PNG images. Some charts allow for clonotype filtering. You can select the chart type by using the 'Chart Type' option in the top left corner. Here is an overview of each chart.



1. Clonotype abundance:
   The abundance chart displays the most abundant clonotypes. You can adjust the number of displayed clonotypes by changing the value of 'Num Top Clonotypes.' Additionally, the 'Y Axis Type' can be changed to represent the clonotype (or

gene) cell support count as either abundance or percentage. Hovering over a bar reveals specific information about the clonotype.

2. V/J gene usage:
   By selecting 'Chart Type,' you can choose V/J gene usage. If you select V gene usage, the chart in the content panel will change to a histogram showing the frequency of V genes. The Y-axis represents the number of cells expressing a specific V gene. You can observe that the TRVA1-2 V gene has the highest frequency, supported by 229 different cell barcodes.

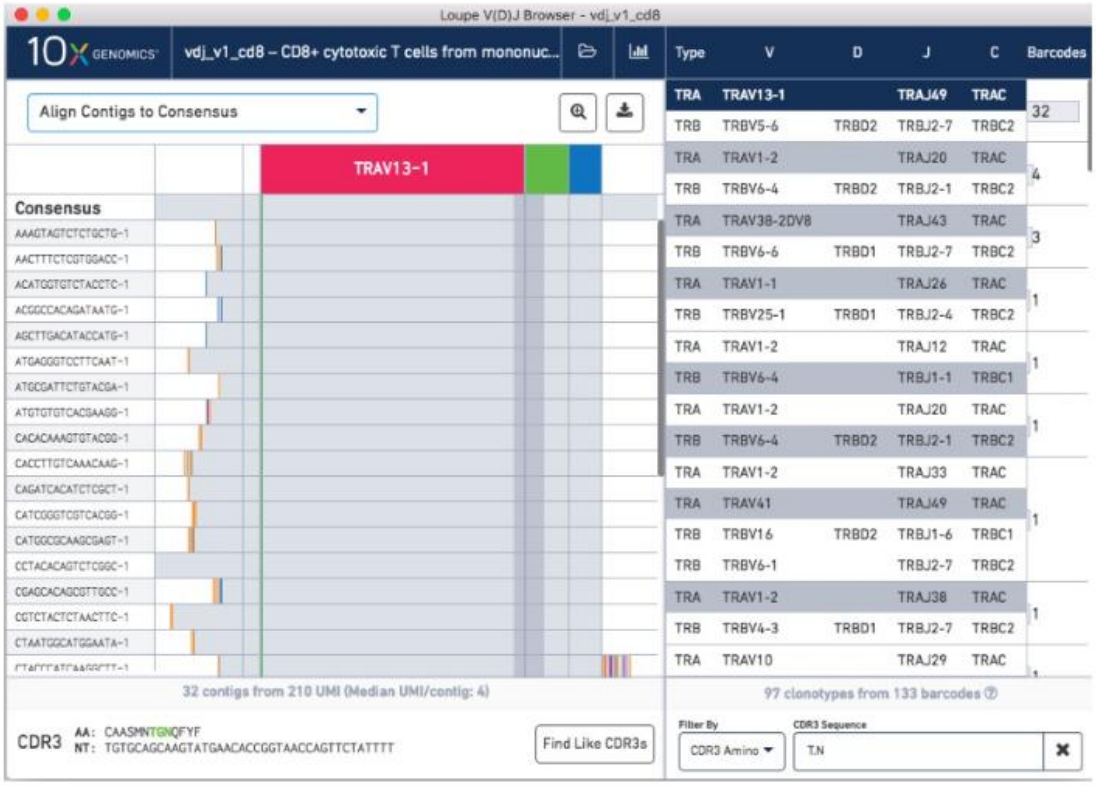
   You can limit the chart display to alpha or beta chains or display both by clicking the 'Chain Type' selection box above the chart. Additionally, hovering or clicking on individual bars will only show clonotypes that contain the specific gene annotation. If you click on TRVA1-2, the chart will display clonotypes containing the TRVA1-2 gene (the corresponding chain will be temporarily highlighted to emphasize the selection). The result of hovering is shown in the image below:

2. Filtering and Clonotype List

In addition to filtering through the V-J chart, you can also use the tools under Clonotype List to search for clonotypes using various methods. First, you can choose a filtering method from the 'Filter By' option. Then, enter a value in the input box to the right. The Clonotype List will display the filtered clonotype list and highlight the entered information. Each filtering method has some differences, as described below:

1. Gene annotation:
   When you select the 'Gene Name' filtering method, Loupe V(D)J Browser will automatically pop up a gene list above the input box for selection. You can click on genes to make selections, and multiple genes can be selected for filtering. The Clonotype List will only display clonotype chains that contain all the gene annotations entered in the input box. To remove genes, you can click the 'X' on the left side of each gene name or click the delete button on the right side of the input box to clear all selected genes.
2. CDR3 amino acid/nucleotide sequence:
   Clonotype filtering can be done based on CDR3 amino acid or nucleotide sequences. In both cases, the filtering will be based on partial matching of clonotype CDR3. This means that if any chain of a clonotype contains the specified amino acids or nucleotides in its CDR3, the clonotype will be retained in the displayed list. The '.' character can be used for fuzzy matching. For example, in CDR3 amino acid mode, entering "T.N" in the search will match clonotype chains containing TSN or TWN.

In addition, you can perform a quick search by entering a CDR3 nucleotide sequence to find clonotype chains. The corresponding clonotypes will be displayed in the Clonotype List on the right side.

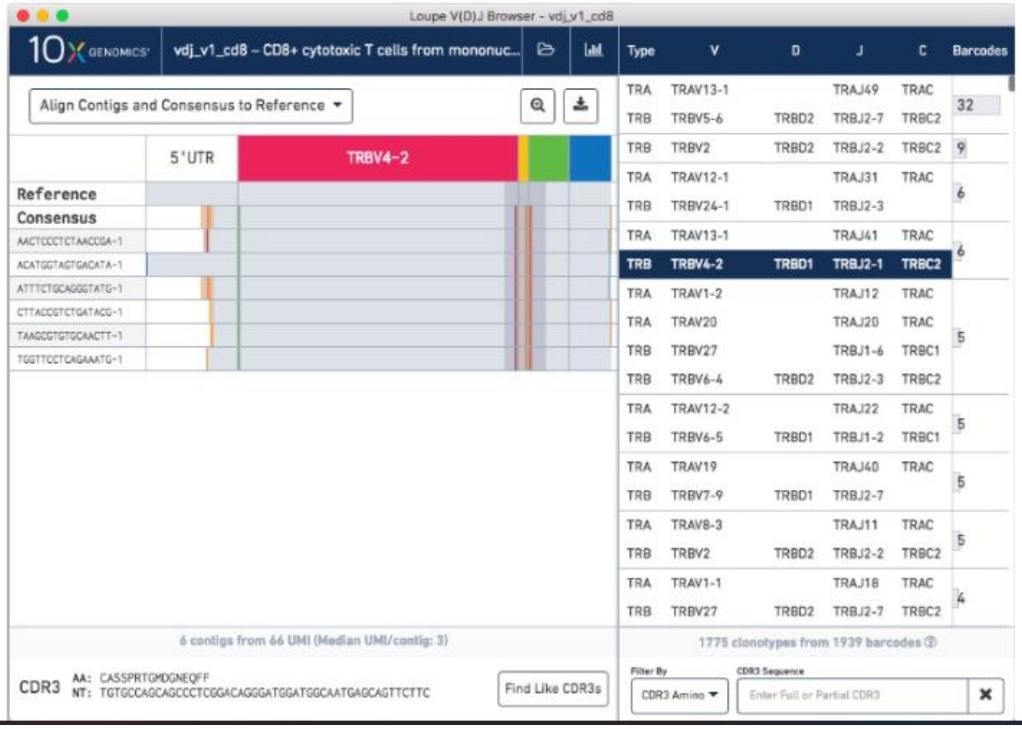


3. Barcode

If you want to search for specific cells in the analysis or filter out specific contigs for further study, you can use Cell barcode and contig ID for filtering. Both Cell barcode and contig ID searches will perform exact matches.

Loupe™ V(D)J Browser: Exploring Chains

The Chain View allows for a detailed display of clonotype chain sequences, both at the contig-by-contig level and the base-by-base level. By clicking on any chain in the Clonotype List, you can view the corresponding results. Here is an example of the displayed results for the fourth clonotype β chain (TRB: TRBV4-2: TRBD1: TRBJ2-1: TRBC2; CDR3 CASSPRTGMDNEQFF):



1. Summary
   When you first click on a clonotype chain, the sequences are displayed row by row at the cell barcode level. The barcode labels at the beginning of each row represent contig sequences assembled from the same cell barcode. Rows marked as 'Consensus' represent consensus sequences assembled from contigs, which represent a chain sequence within the clonotype. Above them is the reference sequence, which is constructed by concatenating sequences from various genes in the reference genome. Lastly, above the reference sequence and consensus sequence is the gene annotation.

As mentioned above, each row in the list represents a contig sequence. Gray regions indicate the alignment of contig sequences to the reference genome. If the 5' end of a row is white, it means that the contig sequence does not span the entire 5' UTR of the reference gene.

In addition, there are vertical lines on the contig sequences indicating various events, such as differences with the reference genome sequence or codons. Hovering over each vertical line will display its meaning. The following colors represent the meanings of the lines for reference:

• Green: Start Codon

• Red: Stop codon (not seen in this chain)

• Orange: Mismatch

• Purple: Deletion

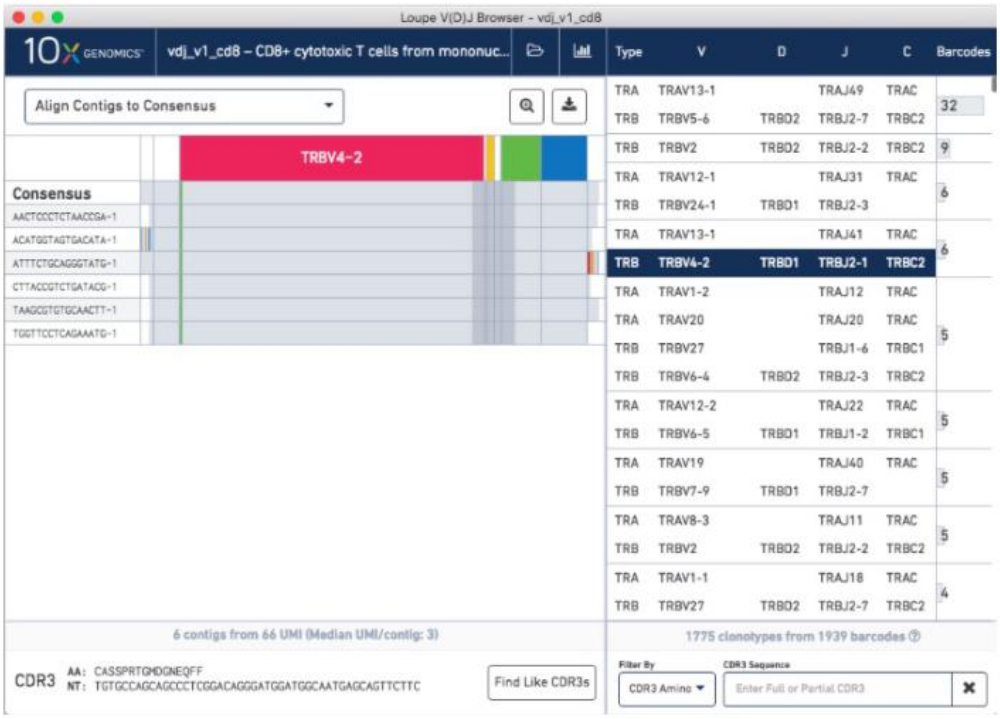
• Blue: Insertion

• Darker Gray Region: CDR3

Typically, the start codon appears at the very beginning of the V gene. However, stop codons do not appear because they are located at the 3' end of the C gene annotation, usually downstream of the primer amplification region.

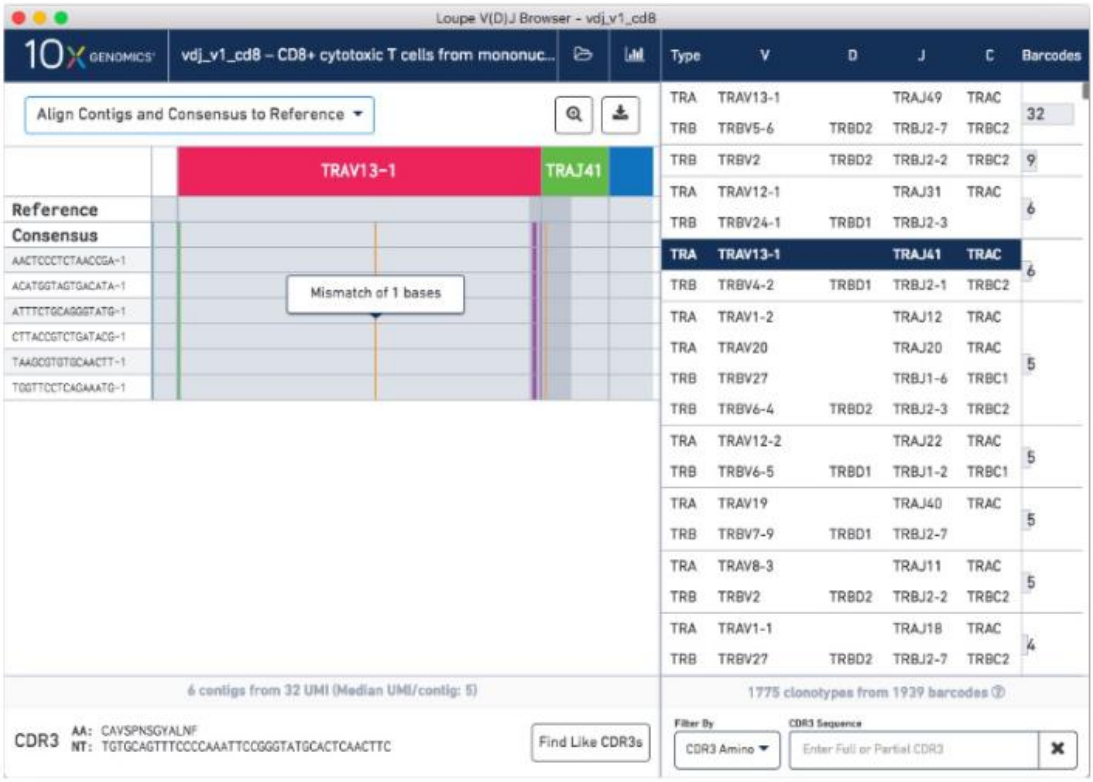
T cells are characterized not only by their V, D, and J genes, but also by the insertion (or deletion) of nucleotides at the gene junctions during V(D)J recombination, which further enhances the diversity of lymphocytes. Due to this additional variability, there are often more variations observed in the regions surrounding the gene junctions compared to the reference genome. However, the consensus sequence will remain consistent with the contig sequence!

When you need to further examine specific mutation information, you can select 'Align Contigs to Consensus' to align the Contig sequences with the Consensus sequence. When switching to this view, the display will be detached from the reference genome, showing regions outside of the annotations. These regions may contain mismatches and artifacts. In the 3' region (ATTTCTGCAGGGTATG-1), you can see some potential errors in stop codons.

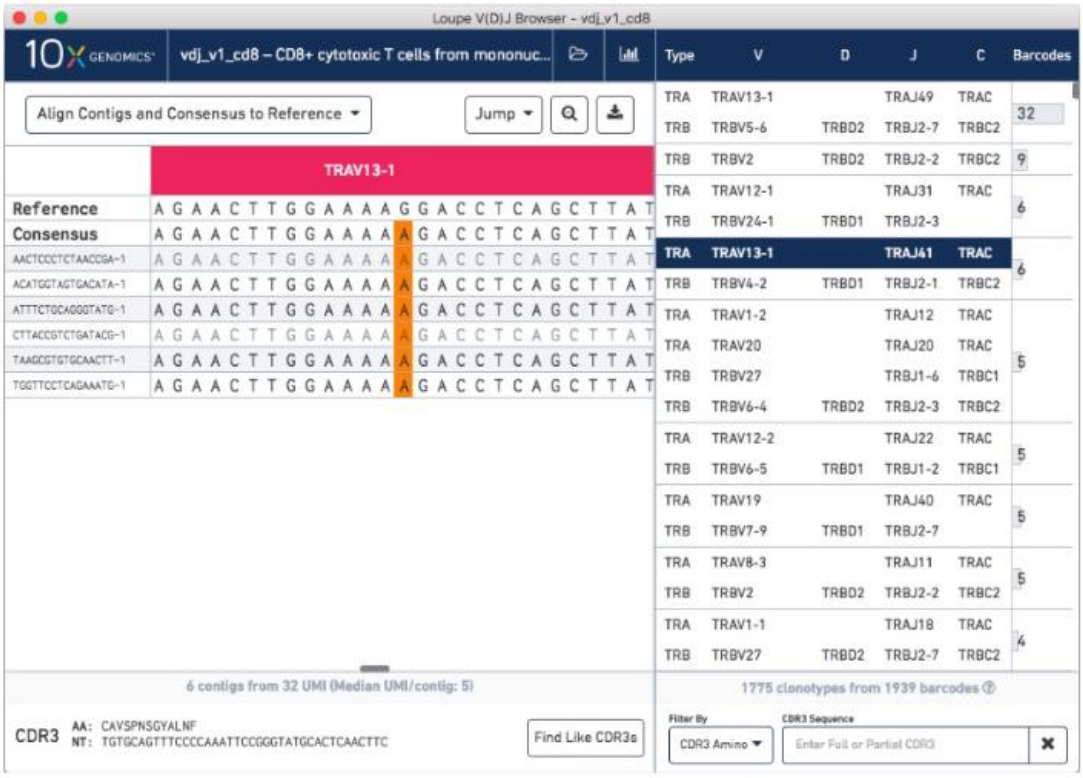


2. Sequence View

To investigate the chain of interest at the base level, you can click on the clonotype α chain, for example (TRA: TRAV13-1: TRAJ41: TRAC), and select the alignment mode in the top left corner as 'Align Contigs and Consensus to Reference' for visualization.



If you want to further examine the mismatches in the orange TRAV13-1 annotation, you can click on it to zoom in and view it at the sequence level. After zooming in, the feature colors remain the same, and orange still represents a mismatch. At this position, the reference genome is G, while the assembled contig is A. When the sequence is too long, you can drag horizontally to view it. Additionally, the 'Jump' option allows you to navigate directly to the selected gene segment.

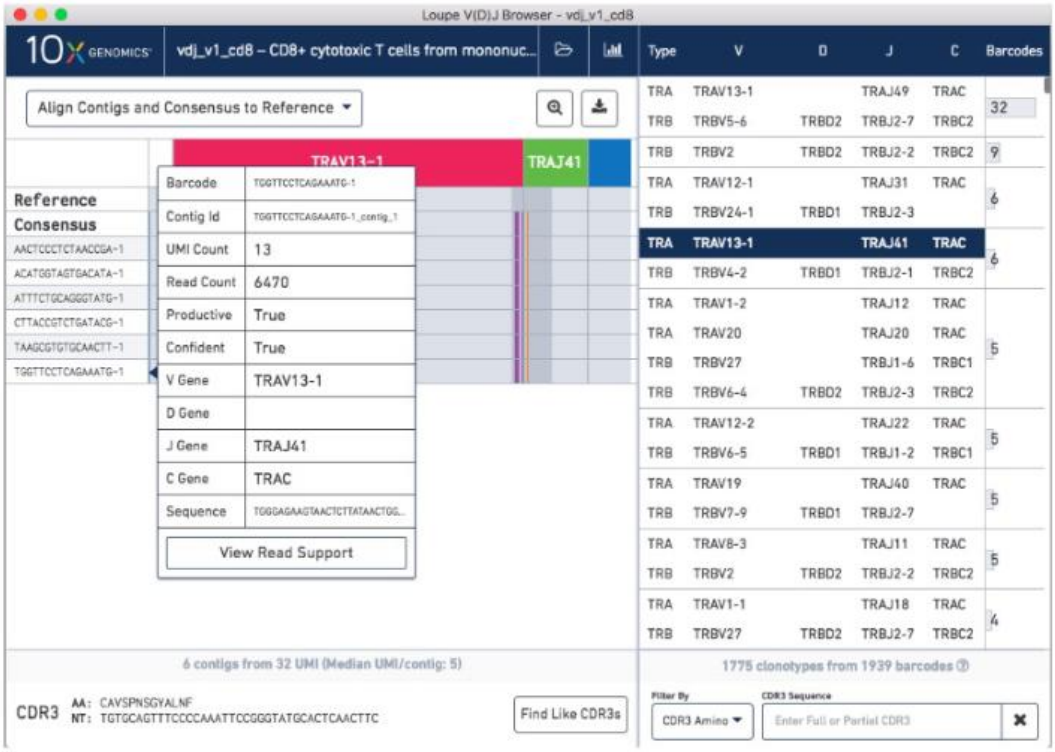


3. CDR3 Sequence, FASTA Output, Contig Data

1. CDR3 Sequence
At the bottom of the view, the consensus sequence of the CDR3 amino acid and nucleotide sequences is directly displayed. Clicking on the sequence allows you to copy it to the clipboard for further investigation.

2. FASTA Output
In the upper right corner of the view panel, you can see a download button. Clicking on it will download all sequence information (reference, consensus sequence, and contig sequences) into a FASTA file (nucleotide or amino acid format). The downloaded file can be used for downstream analysis with standard bioinformatics tools.

3. Contig Data
Hovering the mouse over any barcode label will pop up an additional data box. Clicking on these data allows you to copy them to the clipboard.



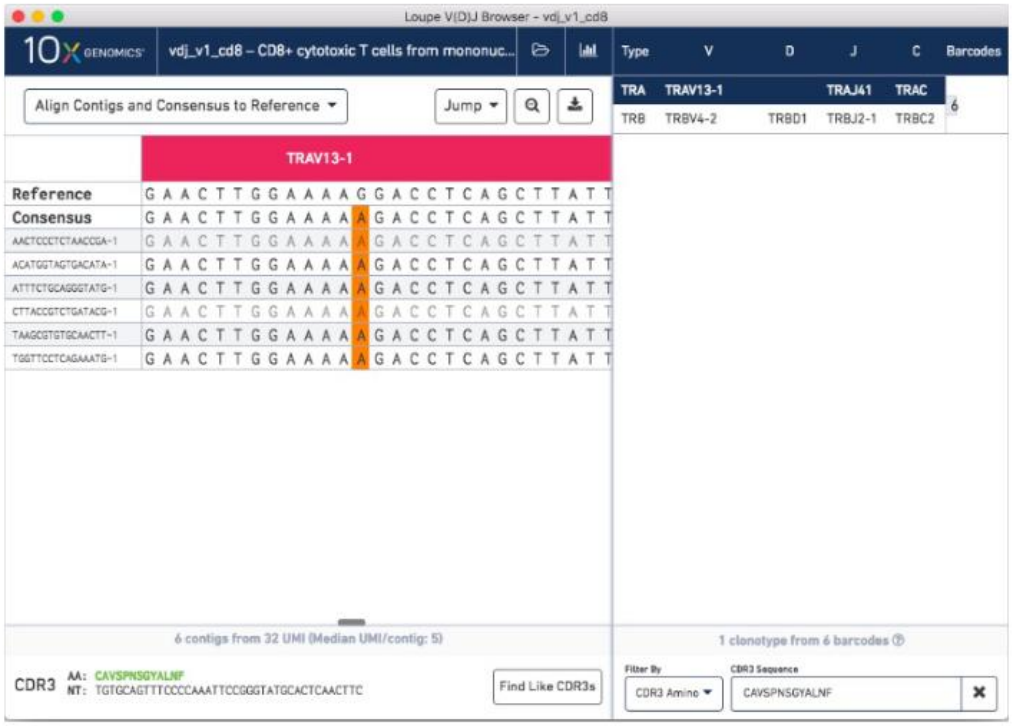
- UMI and read count per contig
- Whether the contig is productive and confident
- Contig sequence
- Cell barcode
- Contig Id

Loupe™ V(D)J Browser: Read Support

So far, you have been able to explore the consensus sequence and clonotype chain information at the contig-by-contig level. If you have a contig BAM file, you can further investigate the reads that support a particular contig sequence.

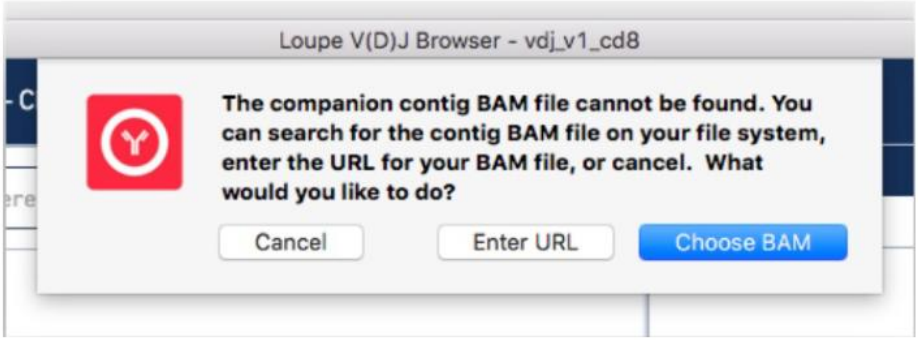
I. Configuration

To proceed with this investigation, first, you need to filter the Loupe V(D)J Browser to the following page (if you have any issues, you can refer to the previous instructions for operation, and make sure to select the yellow region for zooming in).



II. Authorization of .vloupe File for Reads Support

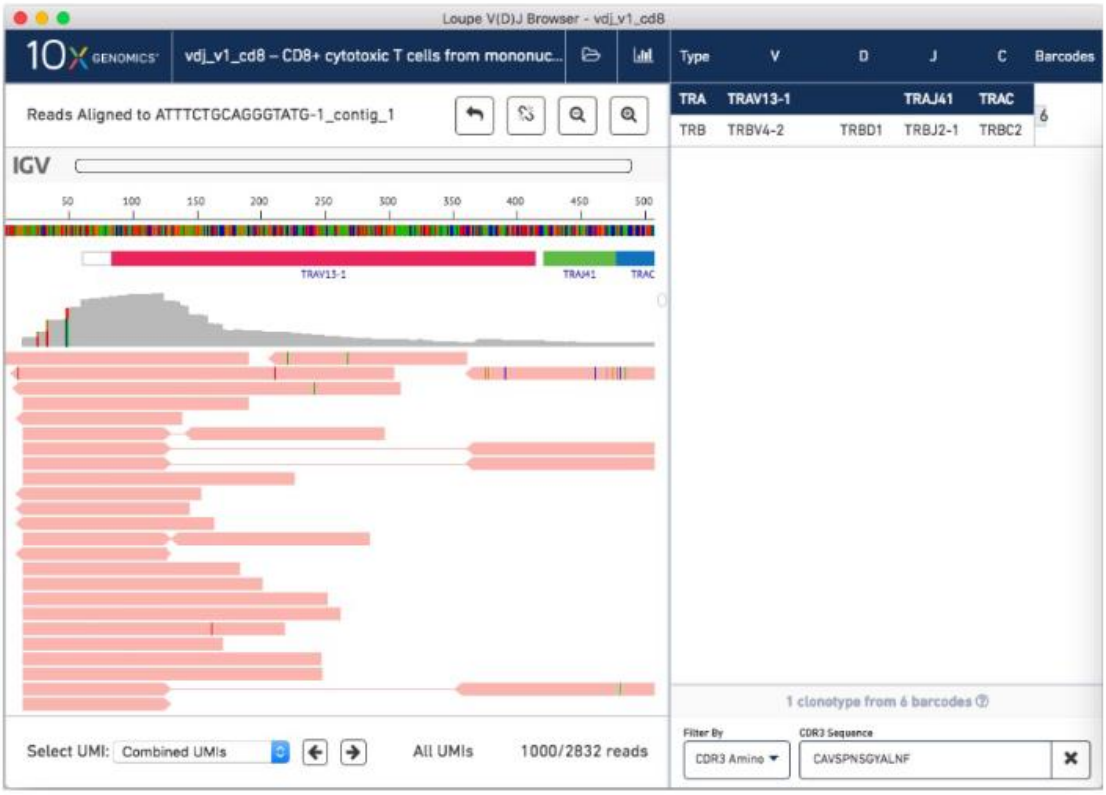
To view the reads supporting a contig, hover the mouse over the barcode on the left side of the chain. As shown in the image, hovering over the barcode "ATTTCTGCAGGGTATG-1" and clicking on "View Read Support." This will open a window. The popup window will appear as follows:



If the BAM file (all_contig.bam) is located in the same path as the .vloupe file, there will be no popup window. However, if they are in different paths, you will need to provide the path to the BAM file for reading. Additionally, you can provide a URL for accessing the BAM file. To provide a URL, click on the 'Enter URL' option and paste the URL into the input box.

III. IGV View

Now you can see the read support for the ATTTCTGCAGGGTATG-1_contig_1 sequence, as shown in the following image.



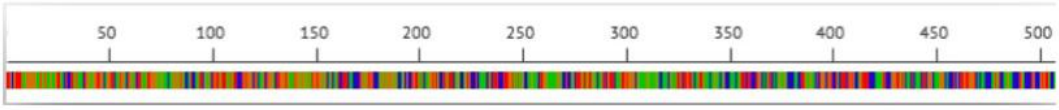
The specific reads are extracted from the BAM file and displayed using IGV.js (a lightweight version of Integrative Genome Viewer developed by the Broad Institute and University of California, San Diego).

1. IGV User Interface

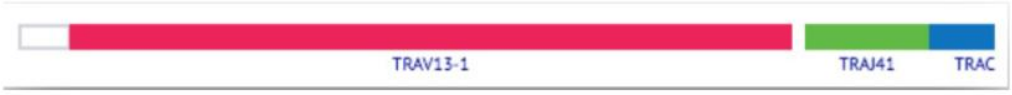
The IGV user interface consists of five sections:

1. The zoom bar: When zoomed in, this bar shows the current content's position in the larger contig region.

2. The ruler and base bar: The ruler displays position information related to the contig, and different bases are represented by colors. Green: A, Blue: C, Orange: G, Red: T. When zoomed in enough, colored bases can be displayed.

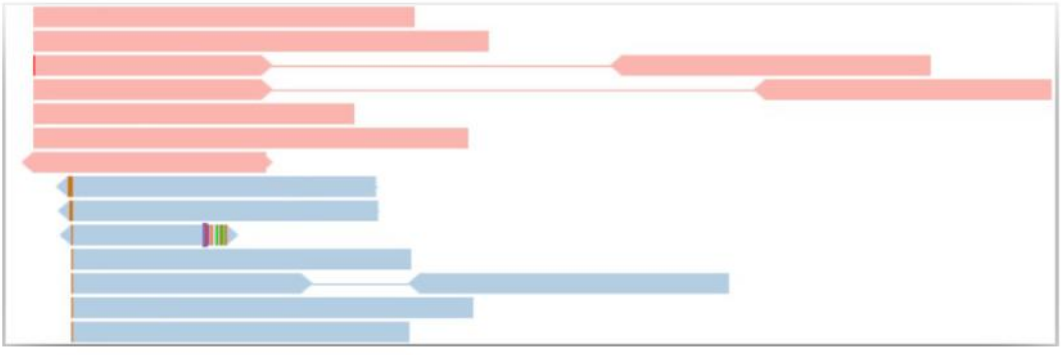


3. The annotation bar: The annotation bar indicates gene annotations on the contig sequence. When hovering over a gene, it displays the gene name and chain region information (V/D/J).



4. The coverage bar: This bar shows the abundance histogram for each position. Special positions are indicated by vertical color bars. Clicking on this bar displays information about that position.

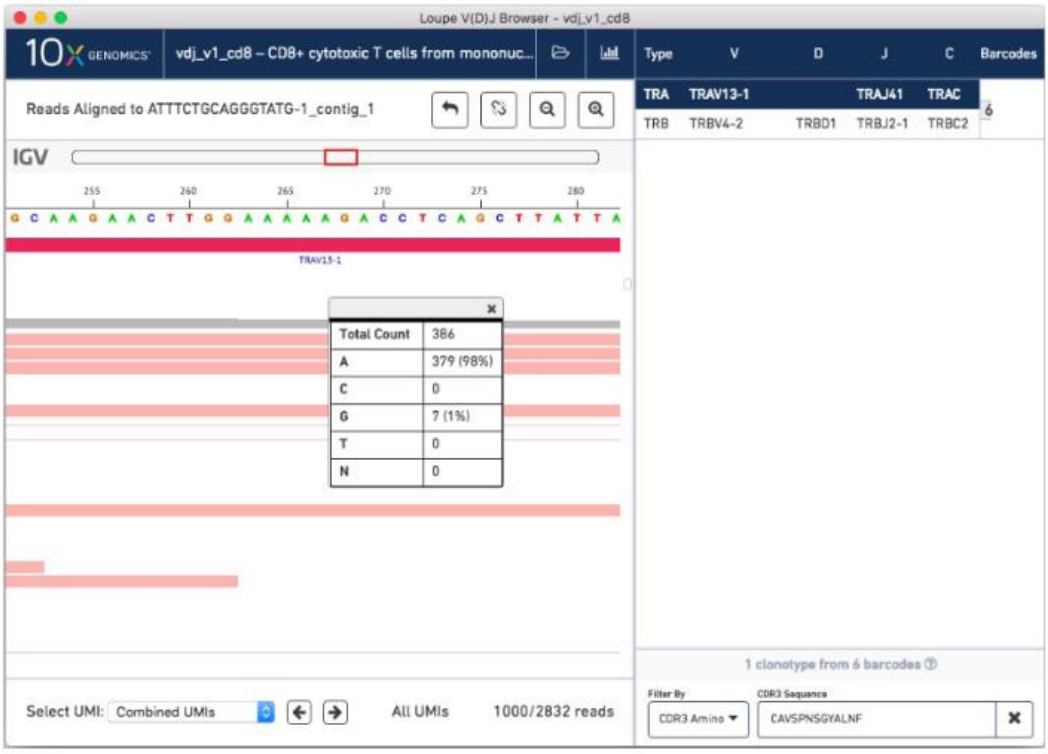
5. Reads bar: The remaining portion displays the reads themselves. Paired reads (R1/R2) are connected by lines, and different colors represent different UMIs. Hovering over a read displays detailed information such as the alignment position on the contig, read name, chain direction, and other BAM tag information. Under the IGV read support view, reads can be filtered and displayed using UMIs. The current browser performance can simultaneously display 1000 reads.



2. Confirming SNPs

You can double-click on the position of interest in the coverage bar to zoom in. By

clicking the 'Zoom In' button three times or a sufficient number of times to magnify the view, you can see the base information clearly. Clicking on the corresponding position in the coverage bar shows the cell support information for different bases at the SNP position. Based on the results, if 98% of the reads covering that base indicate base A instead of the reference genome's G, it confirms the presence of an SNP at that position. You can scroll through the Reads bar to view read information.



3. SNP Analysis

After confirming an SNP, you can further study the frequency of this variant occurring in the sample. In this example, the variant occurs in the TRAV13-1 gene. By selecting 'Gene Name' from the 'Filter Type' dropdown list under the clonotype list and typing TRAV13-1 in the input box, click on the clonotype chain related to TRAV13-1. Finally, you can obtain information about the chains in the sample where this variant occurs.