

V(D)J Analysis

- [Background](#)
 - [Variable \(V\), Diversity \(D\), and Joining \(J\) Recombination Analysis](#)
 - [Terminology](#)
 - [Understanding Clone Composition](#)
- [Import Data](#)
- [Analyzing the Single Cell V\(D\)J Node](#)
 - [Summarize V\(D\)J](#)
 - [Clonotype Frequency Plot](#)
- [Tips for Figure Making](#)
- [References](#)

Background

Variable (V), Diversity (D), and Joining (J) Recombination Analysis

V(D)J recombination occurs in lymphocytes when T and B cells assemble variable (V), diversity (D), and joining (J) gene segments, contributing to the generation of receptors that recognize and respond to perturbations. V(D)J recombination produces clones of unique T cell receptor (TCR) chains or B cell receptor (BCR) chains giving rise to the diverse repertoire of T and B cell populations which are imperative to adaptive immune system function¹. The frequency of generated clones can be measured and explored, giving researchers a powerful view into variation, expansion, and diversity within the biological system. You can import filtered **Contig Annotation CSV files**² from the 10x Genomics Cell Ranger V(D)J or multi pipeline³. If there is matching gene expression data, it can also be imported and analyzed within the same project. We recommend uploading the filtered feature barcode matrices as either the **Hierarchical Data Format (H5 or HDF5)**⁴ or **Market Exchange Format (MEX)**⁵.

Terminology

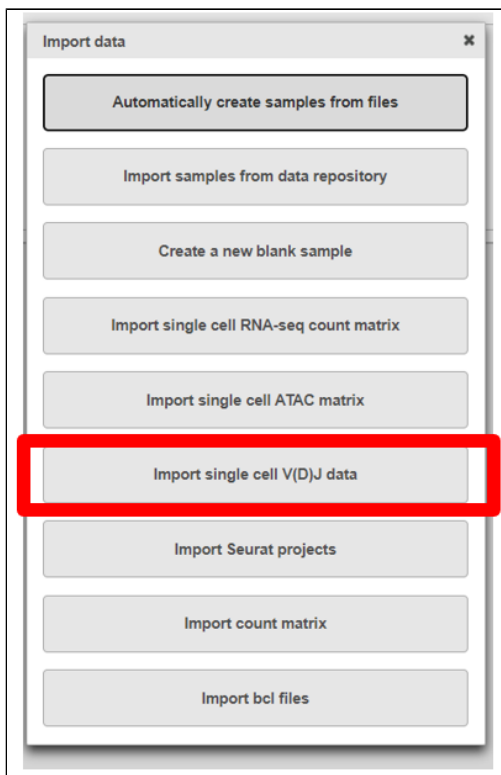
- *UMI* (Unique Molecular Identifier): random 10 bp nucleotide sequence that distinguishes which reads came from the same transcript.
- *Barcode*: the unique identifier in each droplet that usually contains reads from a single cell.
- *Contig*: assembled sequence of bases⁶.
- *Complementarity-Determining Region*: CDR1, CDR2, and CDR3 are important in antigen binding of a T or B cell receptor.
- *CDR3* (Complementarity-Determining Region 3): CDR3 spans the V(D)J junction. There is one CDR3 nucleotide sequence for each V(D)J contig.
- *Clonotype* (clone): cells derived from a common ancestor during clonopoesis which have a particular composition. The cells in a clonotype can have a different number of chains or different CDR3 regions but still be considered a single clone (CDR3 is a highly variable region used for binding; an example of different CDR3 regions would be from affinity maturation which can occur in memory B cells).

Understanding Clone Composition

Multiple cells can have the same clonotype and each clonotype can have multiple makeups. Each clonotype contains one or more chains (TRA and TRB for T cells and IGH, IGK and IGL for B cells), the highest scoring V, D, and J gene segments, and CDR3 nucleotide sequence. T cells have a TRA and TRB chain with V, D, J, and C regions. In B cells, IGH is the heavy chain that has a V, D, and J region while IGK and IGL are the light chains with a V and J region. The Immunoglobulins have two identical heavy chains and two identical light chains. B cell isotypes are antigenic determinants that characterize the classes and subclasses of heavy chains and types and subtypes of light chains; the constant region (C gene) produced by the B cell changes but the V regions and specificity do not. Constant regions do not participate in antigen recognition, instead, C regions interact to mediate biological function; so isotypes have a different function but can bind the same antigen.

Import Data


- **Create** a new project to upload your data. Ensure that you have transferred the filtered contig_annotations.csv file(s)² for each sample from either the *cellranger vdj*⁷ or *cellranger multi*⁸ pipeline to the server. If you have matching gene expression data, import the filtered feature barcode matrices as well in H5⁴ or MEX⁵ format from the *cellranger multi* pipeline for each sample.
- Click **Import**, then select **Import single cell V(D)J data**




- Upload either the filtered contig_annotations.csv file alone or, if you have matching gene expression data, with the filtered count matrix per

sample and give each sample a name. To add a sample use the  Action. The example below is using the default settings and there is one sample with two files, one for V(D)J, and one for gene expression. Click **Finish**.


Home > VDJ > Import single cell V(D)J data


Sample name	Contigs	Annotations	Files	Browse files	Action
Melanoma 7K 10X Genomii	4399	29	7k_melanoma_TBNK_connect_7k_melanoma_TBNK_connect_count_sample_feature_7k_melanoma_TBNK_connect_7k_melanoma_TBNK_connect_vdj_b_filtered_contig_ai	<input type="button" value="Browse"/>	

Feature annotation


Use annotation file  ☐

Counts format

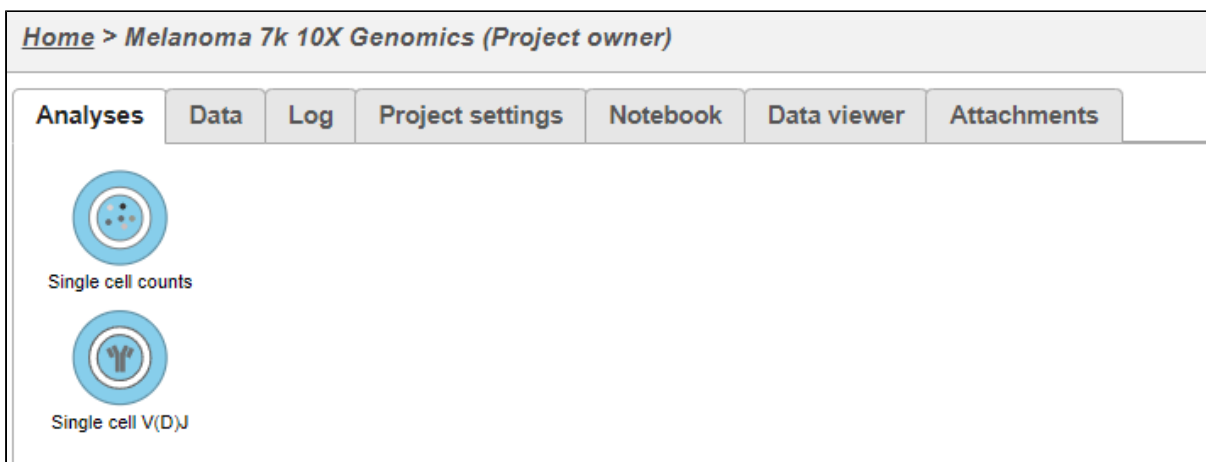
Raw counts  ☒

Report features without counts  ☒

Feature deduplication

Deduplication method  ☒ Mean ☐ Maximum ☐ Sum

- This results in two starting nodes, one for single cell counts and one for single cell V(D)J as shown below. Note that once subsequent tasks are performed on a node, no more data can be imported into this project. The single cell counts node can be processed as usual; for help related to this please see the tutorial for [Analyzing Single Cell RNA-Seq Data](#).



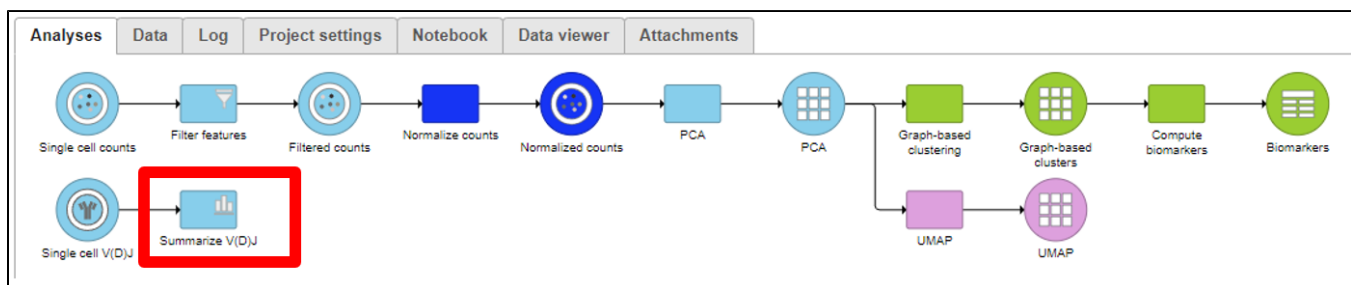
Analyzing the Single Cell V(D)J Node

Summarize V(D)J

- Under *QA/QC* tasks is the **Summarize V(D)J** task which will summarize the V(D)J contents by Sample name, # Cells, Barcode count, Clonotypes, Variable genes, Diversity genes, Joining genes, and Constant genes.



- Double-click on the completed task to view the contents which can be downloaded.



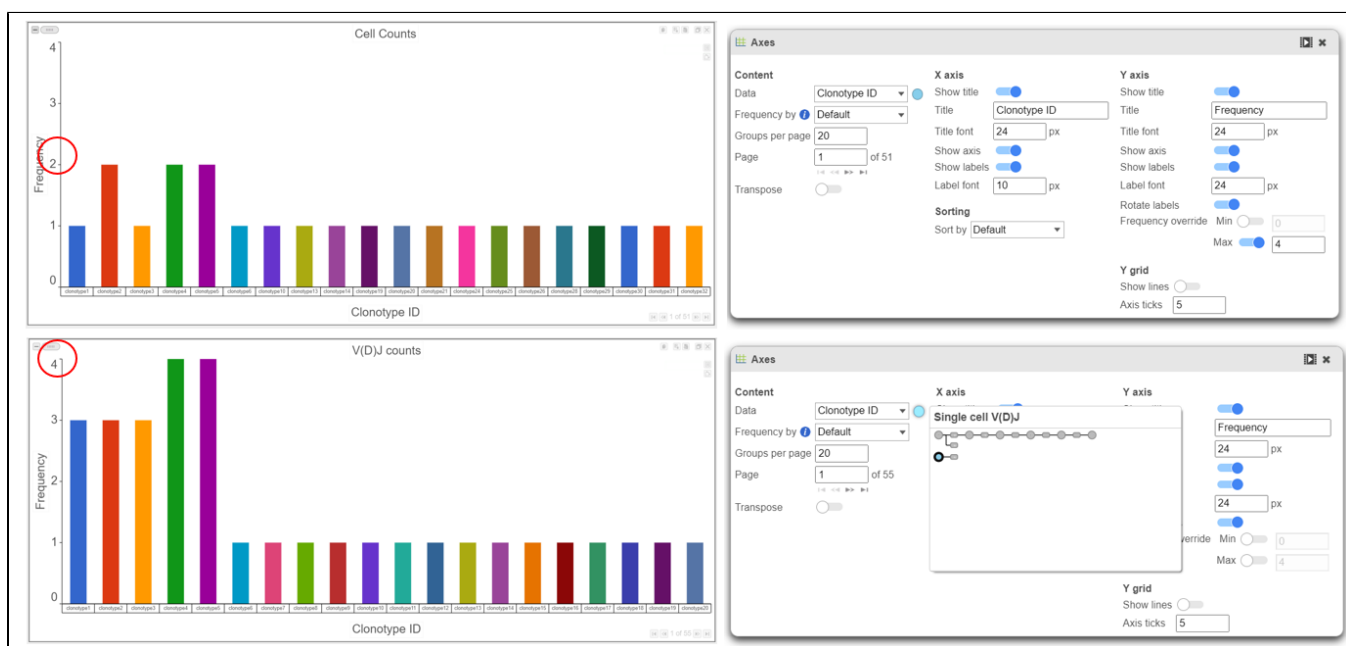
Clonotype Frequency Plot

Under the *Exploratory Analysis* task is the **Clonotype Frequency Plot** task which will summarize the V(D)J node into plots of interest in the Data Viewer. The same or different comparisons can be made in the Data Viewer (see *Tips for Figure Making* below). These may include determining the T cell receptor and B cell receptor chains that makeup clonotypes in the samples, quantifying the clone diversity by frequency, comparing the immune repertoire between samples, and visualizing clones and gene expression data together on scatterplots like a UMAP.

- The example below shows the results from the **Clonotype Frequency Plot** task which is accessed by choosing to perform this task from the *Single cell V(D)J node* and will automatically open in the data viewer for modification. In this case, the barcode frequency is the number of clonotypes per cell because the barcode usually represents a single cell, so there are two cells that have clonotype5. Clonotype5 is made of two compositions (a frequency of four for clonotype5 from the V(D)J node) with a Chain, V gene, D gene, and J gene as seen below the bars and by hovering.

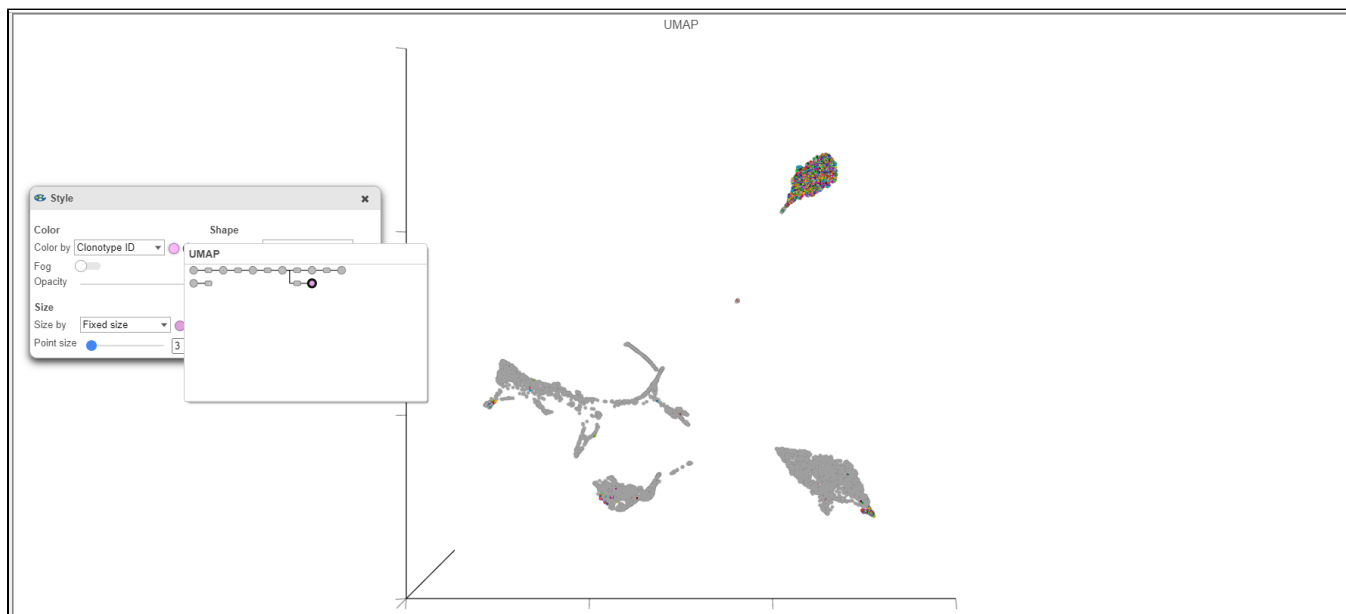


- Plotting Clonotype ID frequency from the Cell counts and VDJ counts nodes highlights the difference between the two nodes. In this example, the top plot is the number of cells per clonotype and the bottom plot is the number of V(D)J clonotypes present. Note that Cell Ranger does not always call the barcode a cell and this can affect frequencies when making comparisons between cell frequency per clonotype and barcode frequency per clonotype. An example of this would be clonotype1 when comparing the figure above and below.



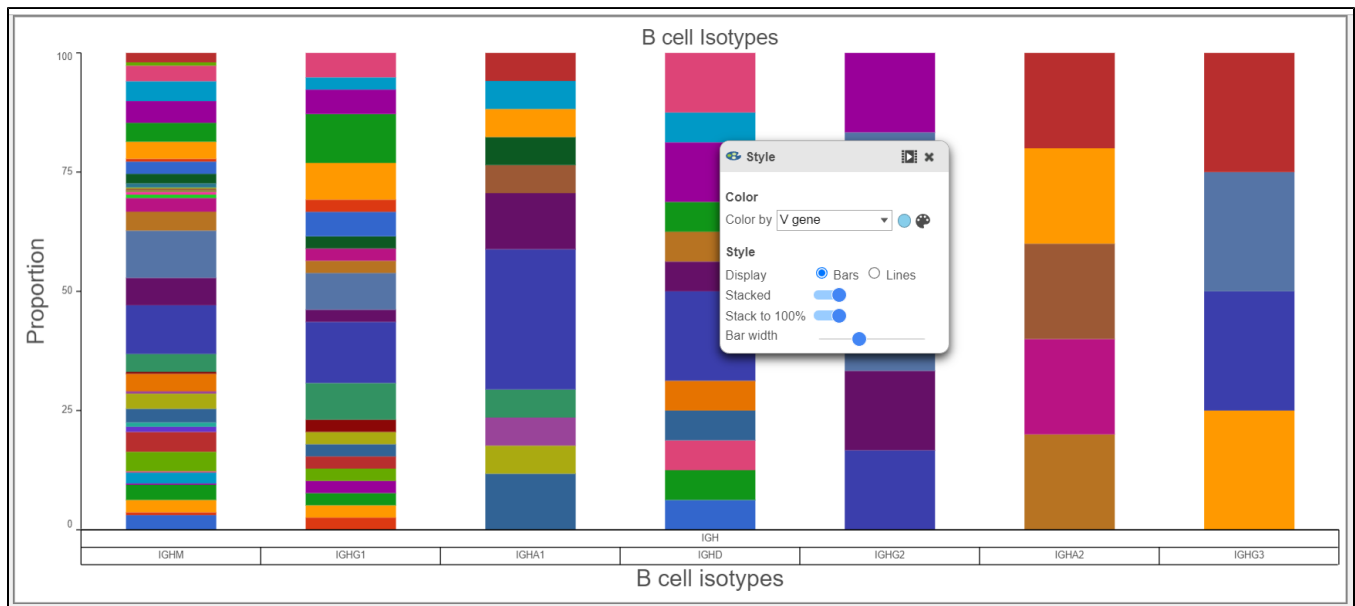
Tips for Figure Making

- When overlaying the Clonotype ID on plots from the Single cell counts analysis pipeline (such as the **3D Scatter plot** example below), make sure to use the Clonotype ID from the Single cell counts node and not the VDJ node.

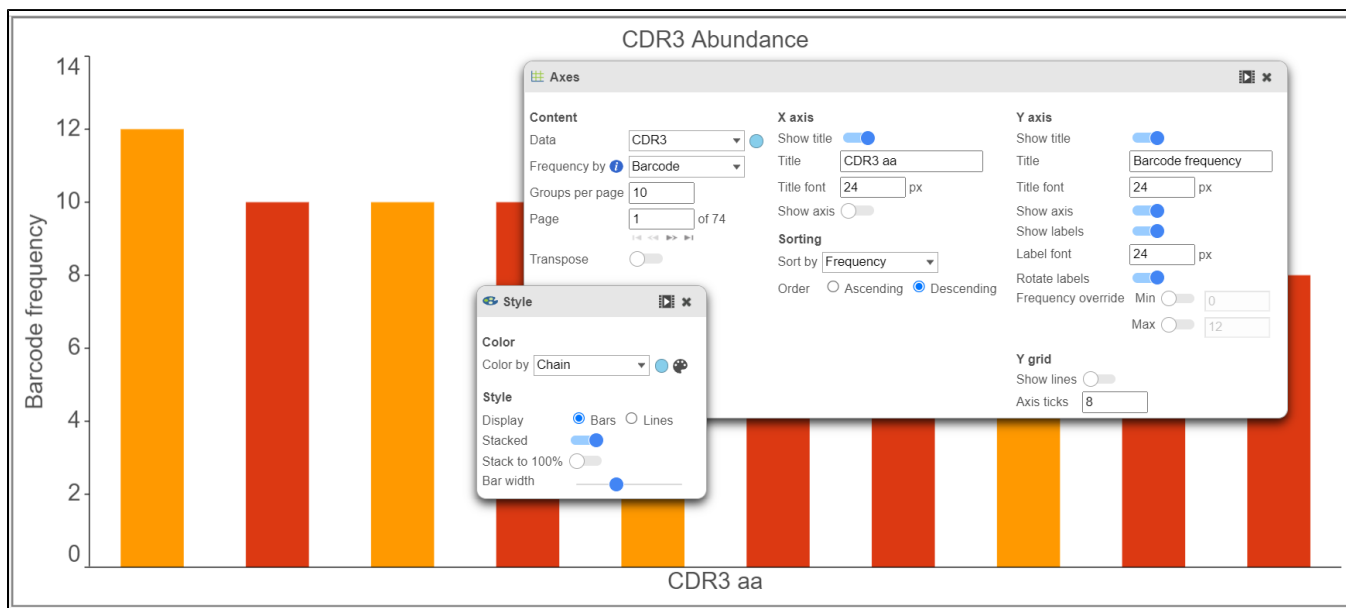
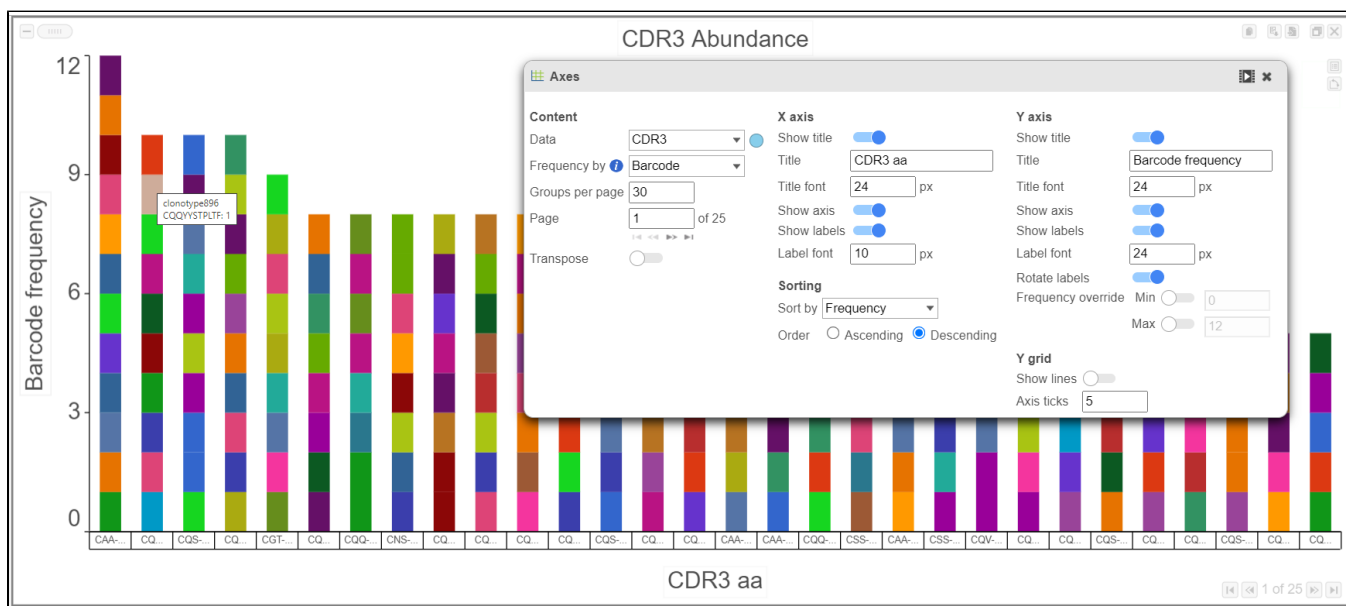


- B cell isotypes are defined by the Chain and C gene. In the examples below, the Chain and C gene are plotted by Barcode frequency using a **Bar chart**. On the top, no selection and filtering have been performed. On the second plot, the data has been selected and filtered by the heavy chain. Certain criteria can be focused on by using **Select & filter**. The third plot is stacked to 100% and colored by the V gene because the V regions do not change specificity during isotype switching.

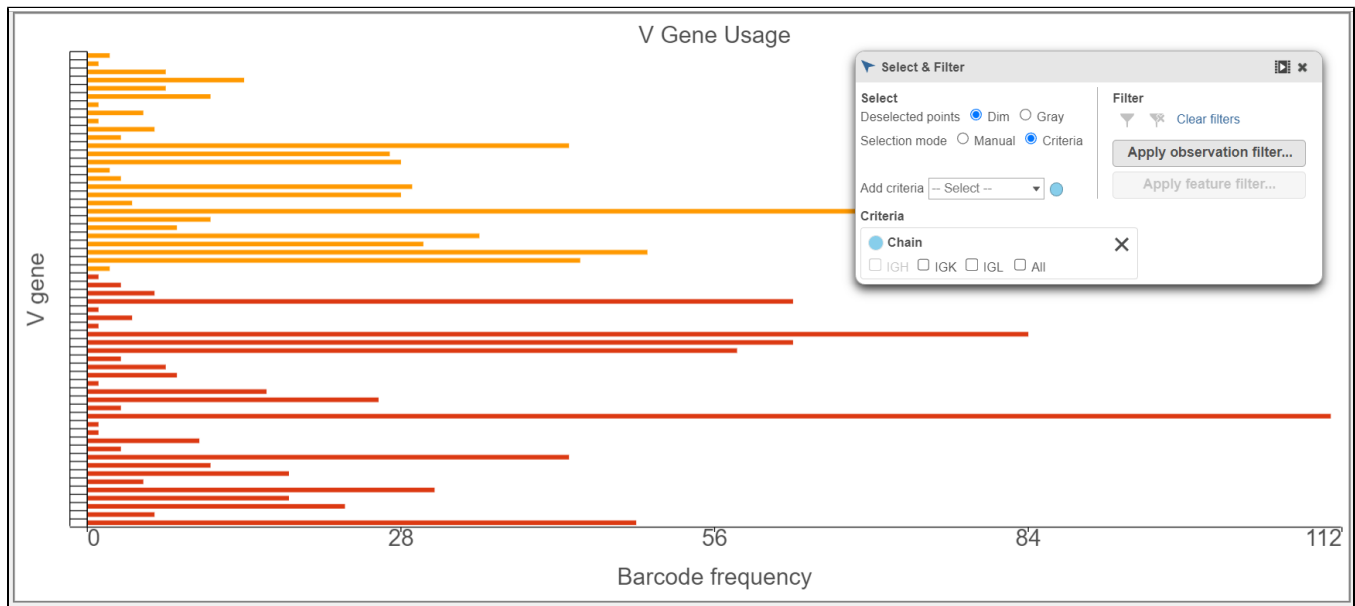
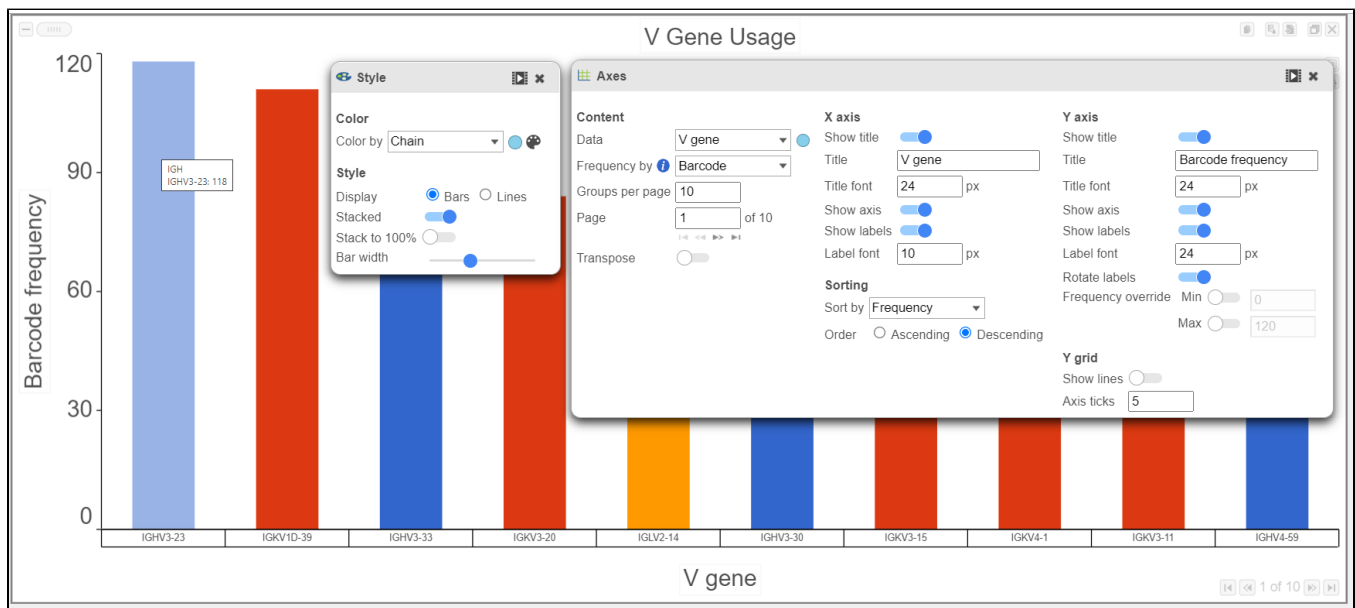




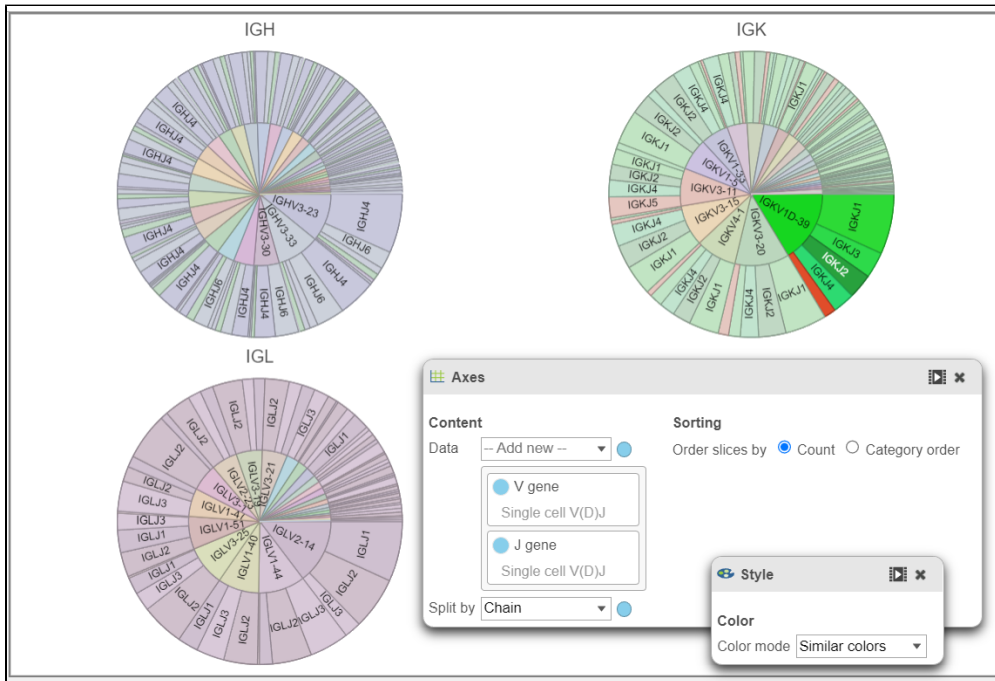
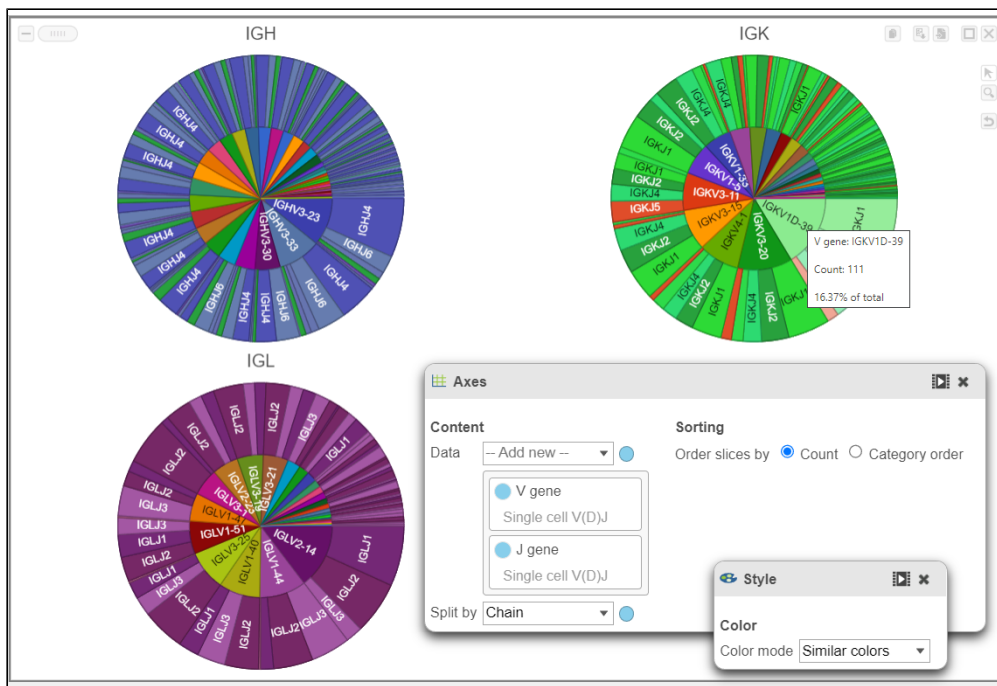
- Below in the top chart, CDR3 abundance is plotted by barcode frequency and colored by Clonotype ID. In the bottom chart, the plot is colored by Chain and other modifications, such as axis ticks and the number of groups per page. Note that the predicted CDR3 amino acid sequence is plotted here. The predicted CDR3 nucleotide sequence (cdr3_nt), as well as information for other Complementarity-Determining Regions, is also available.



- Gene usage plots for the V, D, and J genes can be plotted in many ways, as seen in the V gene examples below. In the top plot, the barcode frequency for V genes is sorted by frequency in descending order and colored by Chain. The transposed plot below shows all of the groups sorted by ascending value and the heavy chain has been excluded.



- In the example below, a **Pie chart** split by Chain is used to plot the V gene and J gene. This is another way to explore and visualize frequency.



References

1. Tonegawa, S. Somatic generation of antibody diversity. *Nature* 302,575–581 (1983). <https://doi.org/10.1038/302575a0>
2. <https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/output/annotation#contig-annotation>
3. <https://support.10xgenomics.com/single-cell-vdj/software/overview/welcome>
4. https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/7.0/advanced/h5_matrices
5. <https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/7.0/output/matrices>
6. <https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/algorithms/annotation#productive>
7. <https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/using/vdj>
8. <https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/using/multi>