

VDJ Explorer How-To

Background

Variable (V), diversity (D), and joining (J) regions of lymphocyte immune cell receptor proteins are capable of undergoing recombination, which produces a set of unique alpha and beta chain pairs (aka clonotypes), the sum totality of which is sometimes called the repertoire of T and B cell populations. Measurements of clonotype diversity give researchers a nuanced and powerful view into the expansion of subpopulations of these cell types. Particular T cell and B cell receptors (TCR / BCR), and the diversity of these epitopes are vital to the proper function of the immune system, and can be indicators of changes in response to system perturbation.(1)

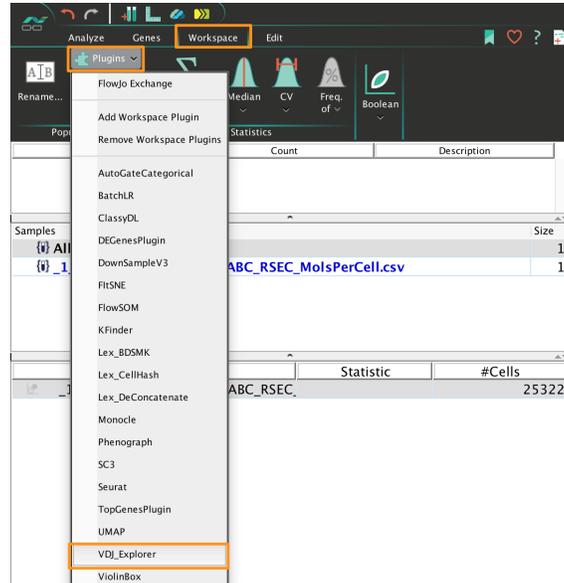
Immune repertoire analysis of single cells is now made possible by advances in single cell RNA sequencing. This is the process by which researchers will characterize T-Cell and B-Cell receptor diversity within a sample of using next generation sequencing techniques from one of a variety of platforms.(2)

However, the task of parsing and illustrating the information from V(D)J recombination can be quite complicated, due to the ‘many to many’ mapping of these relationships. Data required for this task include both the single-cell RNA-sequencing expression matrix for a sample, and the corresponding meta information on V(D)J identification, in CSV format.

Here we detail the workflow in which clonotypes are elucidated using the **VDJ Explorer** plugin.

Install

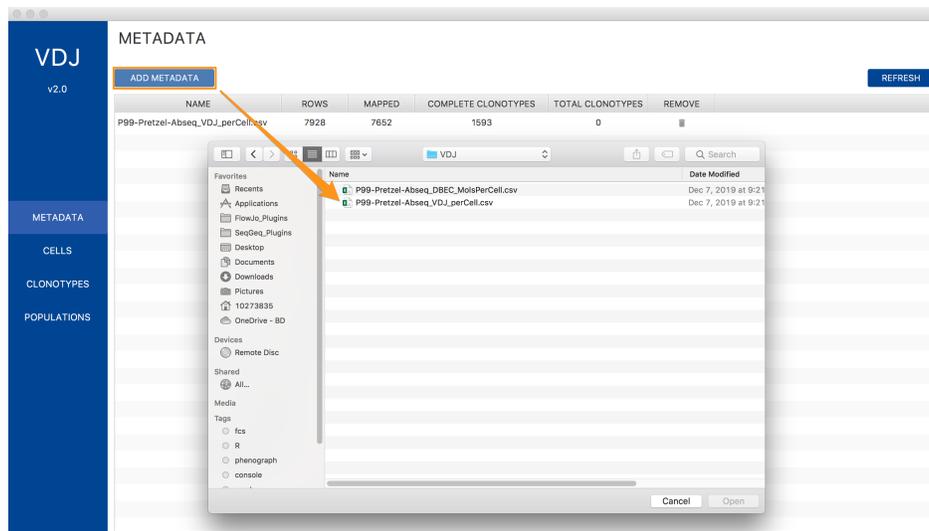
The VDJ Explorer plugin for SeqGeq has been coded in Java, and therefore does not require any R connection or dependencies, thus you can simply download the plugin JAR file, and place that into your SeqGeq plugins folder. Restarting SeqGeq should illustrate that plugin within the workspace Plugins dropdown list:



Use

Samples Section

Once the plugin has been run on an appropriately sequenced gene expression matrix (GEX) file, it will require a researcher connect the file to their V(D)J CSV annotations file, either named “all_contig_annotations.csv” or “VDJ_perCell.csv”. This is accomplished by clicking on the gene expression matrix file within SeqGeq, and opening the VDJ Explorer, where you will be prompted to select the meta-info CSV file(s):



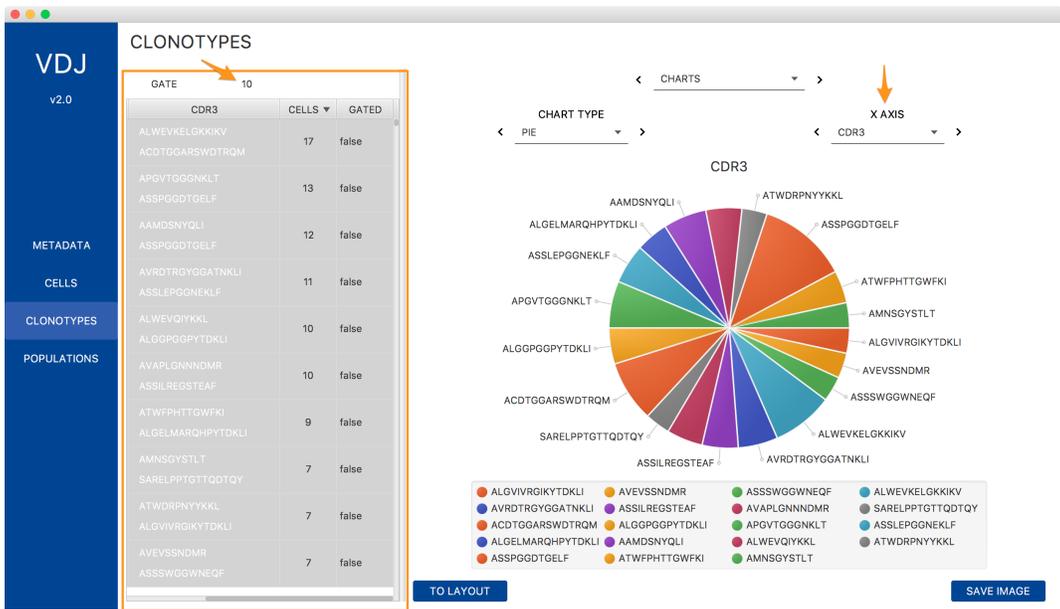
Repertoire

The 'Cells' tab of the VDJ Explorer gives researchers the ability to filter their Metadata file for cells corresponding to particular clonotypes of interest. This is achieved in part by applying filters to the Metadata columns displayed by right clicking on the column header of interest to apply a filter.

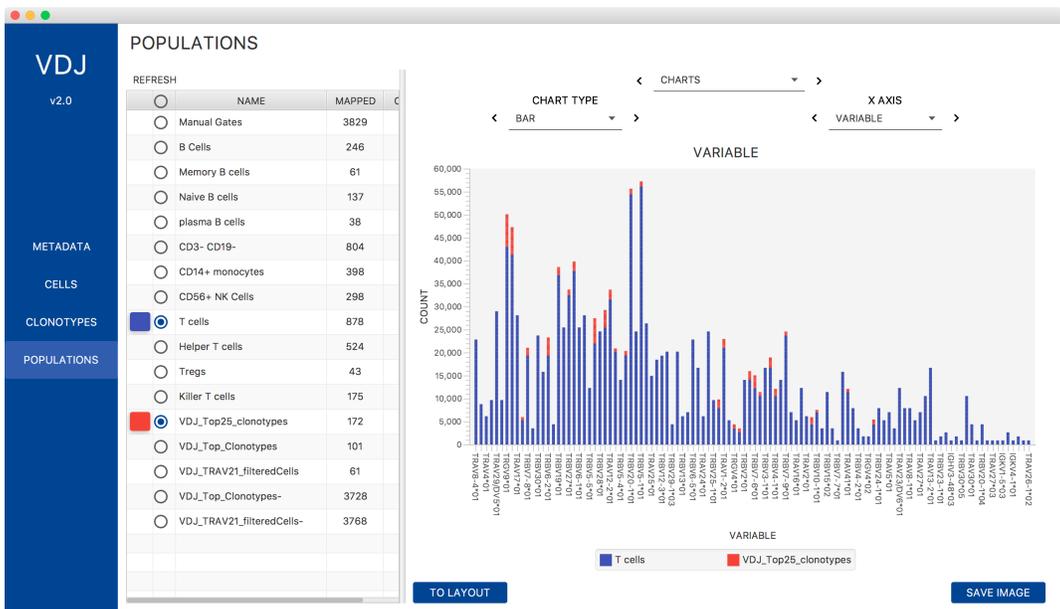
After applying the filter, you can select all of the cells in the list by pressing cmd+A (Ctrl+A on Windows) then right click and choose "Population from Selection" to create a new population in the workspace. You can then rename the population in the workspace appropriately.

Owned	TYPE	CHAIN	READS	UMIS	VARIABLE
true	TCR	Alpha	4027.0	18.0	TRAV21*01
true	TCR	Alpha	1320.0	3.0	TRAV21*01
true	TCR	Alpha	4624.0	14.0	TRAV21*01
true	TCR	Alpha	1389.0	7.0	TRAV21*01
true	TCR	Alpha	5908.0	14.0	TRAV21*01
true	TCR	Alpha	1774.0	9.0	TRAV21*01
true	TCR	Alpha	1281.0	7.0	TRAV21*01
true	TCR	Alpha	8103.0	25.0	TRAV21*01
true	TCR	Alpha	342.0	4.0	TRAV21*01
true	TCR	Alpha	4683.0	12.0	TRAV21*01
true	TCR	Alpha	2812.0	13.0	TRAV21*01
true	TCR	Alpha	2643.0	8.0	TRAV21*01
true	TCR	Alpha	6346.0	15.0	TRAV21*01
true	TCR	Alpha	4482.0	9.0	TRAV21*01
true	TCR	Alpha	2207.0	6.0	TRAV21*01
true	TCR	Alpha	7269.0	21.0	TRAV21*01
true	TCR	Alpha	3225.0	8.0	TRAV21*01
true	TCR	Alpha	812.0	4.0	TRAV21*01
true	TCR	Alpha	1333.0	10.0	TRAV21*01
true	TCR	Alpha	2129.0	5.0	TRAV21*01
true	TCR	Alpha	151.0	6.0	TRAV21*01
true	TCR	Alpha	1471.0	8.0	TRAV21*01
true	TCR	Alpha	5792.0	12.0	TRAV21*01
true	TCR	Alpha	96.0	1.0	TRAV21*01

Comparisons can be accessed in the Clonotypes and Populations section. Selecting Clonotypes on the left will allow for comparisons to be shown by selecting the Charts option at the top and selecting the desired Chart Type and X Axis options. Here we are comparing the top ten clonotypes selected from the left, and their third complementarity-determining regions (CDR3).



Selecting the Populations tab will allow users to perform comparisons between populations based on their V(D)J information. You can then select the Charts option and X Axis comparator to view different plot types. Here we are comparing the T cell population vs the Top25_clonotype population and their variable gene usage across the populations.



The figures themselves can be exported from the plugin as PNG figures, or directly to the Layout Editor by clicking on the corresponding button within the main window of VDJ Explorer.

Downstream

Artifacts generated from the plugin, such as the most frequently occurring clonotypes can be used for further in depth analysis throughout SeqGeq's platforms. While in the Clonotypes tab, sorting by the number of cells will allow the top most clonotypes to be selected and the resulting cells exported as a new population by using 'Cmd+A' to select all the rows of the table, then right-clicking to export as a population.

The screenshot displays the SeqGeq VDJ interface. On the left, a sidebar contains navigation options: VDJ v2.0, METADATA, CELLS, CLONOTYPES (selected), and POPULATIONS. The main window is titled 'CLONOTYPES' and shows a table with columns: GATE, CDR3, CELLS, and GATED. An orange arrow points to the 'CELLS' column. Below this table is a larger table with columns: ID, TYPE, CHAIN, READS, UMIS, VARIABLE, DIVERSITY, and JOINING. An orange arrow points to a 'Population From Selection' button within this table. At the bottom, there are 'TO LAYOUT' and 'SAVE IMAGE' buttons.

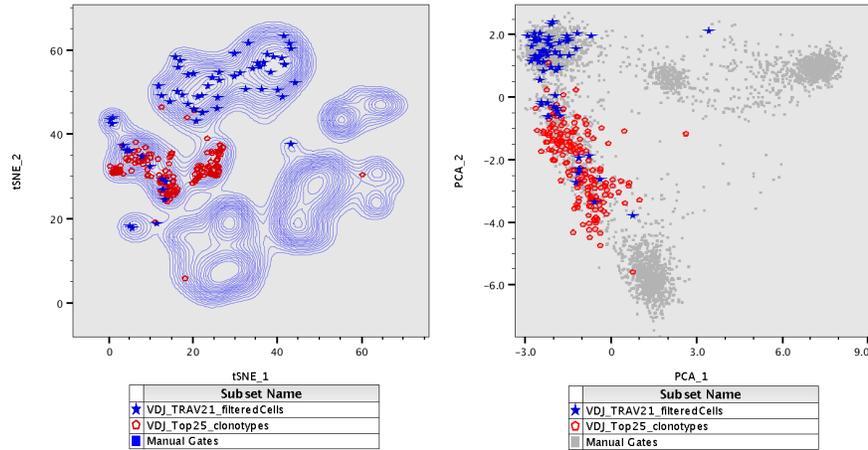
Differential Expression and Geneset Enrichment Analysis

As with any population in SeqGeq, we can begin to ask what the transcriptome is doing within clonotypes of interest using the Volcano Plotting tool to analyze differentially expressed genesets there, and follow that with Geneset Enrichment analyses:

The screenshot shows three windows from the SeqGeq interface. On the left is a Volcano Plot with 'log2(q-value/Manual Gates/VDJ_Top_Clonotype)' on the y-axis and 'log2(fold_change/Manual Gates/VDJ_Top_Clonotype)' on the x-axis. A vertical line is drawn at x=1, and a horizontal dashed line is at y=0.05. A region of points above the horizontal line and to the right of the vertical line is shaded orange and labeled 'UP in Top Clonotype'. In the center is the 'Gene Set Inspector' window, showing 'Gene Count: 16' and 'Gene Set Type: ANALYTIC'. It lists parameters for 'fold_change/Manual Gates/VDJ_Top_Clonotype' and 'q-value/Manual Gates/VDJ_Top_Clonotype'. On the right is the 'Gene Set Library' window, showing a list of genesets with columns for Name, Count, and Description. An orange arrow points from the 'UP in Top Clonotype' label in the Volcano Plot to the 'Gene Set Inspector' window, and another orange arrow points from the 'Gene Set Inspector' window to the 'Gene Set Library' window.

Clonotypes Within Clusters

Clonotypes detected by V(D)J sequencing can be compared with unbiased clustering coming from other platforms in SeqGeq, and visualized in dimensionally reduced spaces:



References

1. F. Alt, et al. "VDJ recombination." Immunology Today 13.8. (1992)
2. M. De Simone, et. al. "Single Cell TCR Sequencing: techniques and future challenges." Frontiers in Immunology 9. (2018)
3. J. Lin. "Divergence measures based on the Shannon entropy." IEEE Transactions on Information Theory 37.1. (1991)