How to V(D)J Explorer

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)

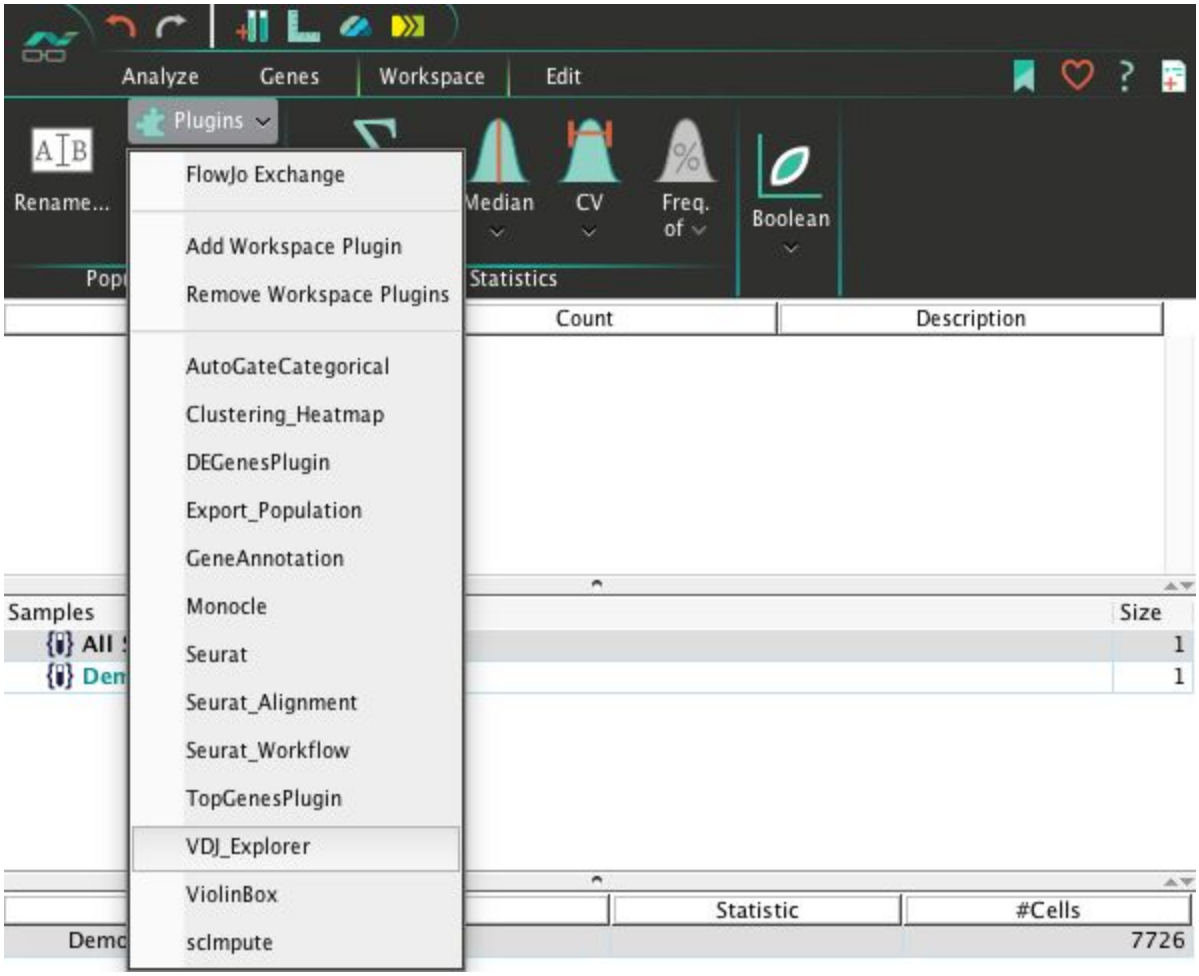
Immune repertoire analysis of single cells is now made possible by advances in single cell RNA sequencing. This is the process by which researchers will characterize T-Cell and B-Cell receptor diversity within a sample of using next generation sequencing techniques from one of a variety of platforms.(2)

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 **V(D)J Explorer** plugin.

Install

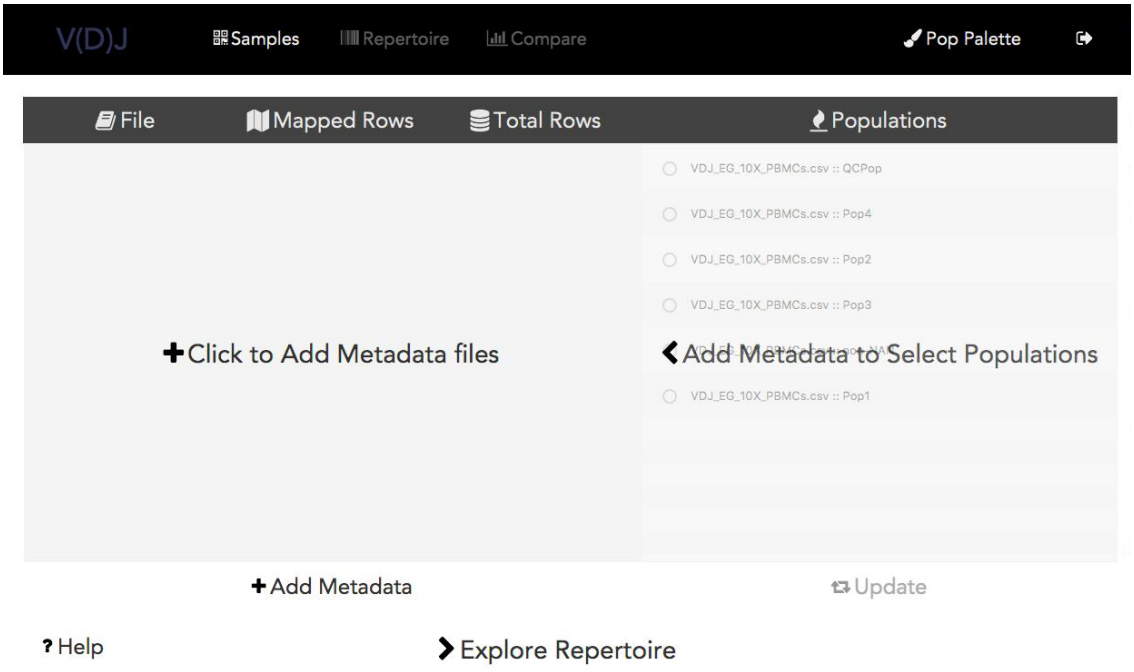
The V(D)J Explorer plugin for SeqGeq has been coded in JavaFX, and therefore does not require any R connection or dependencies, thus you can simply download the plugin JAR file, and place that into your SeqGeq plugins folder. Restarting SeqGeq should illustrate that plugin within the workspace:



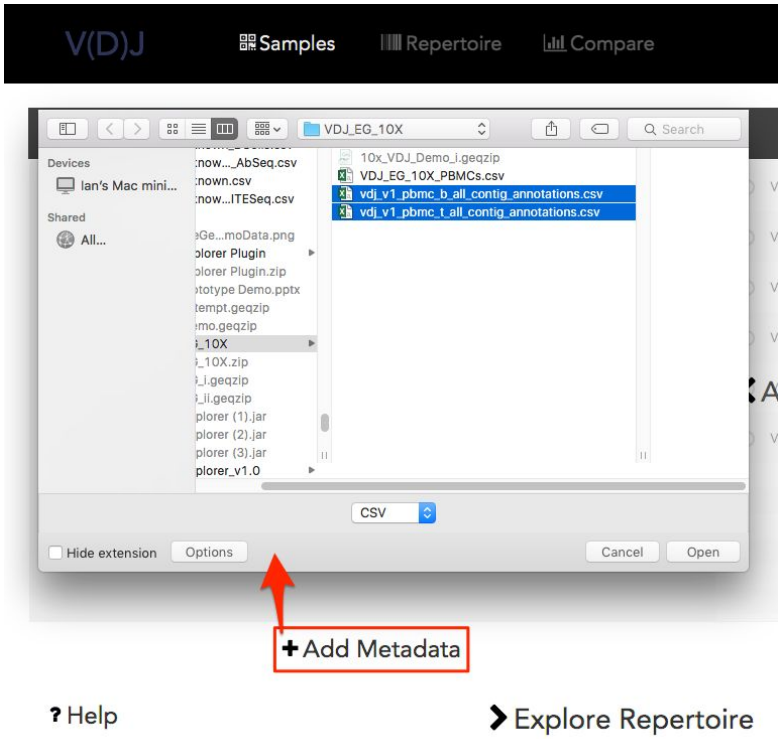
Use

Samples Section

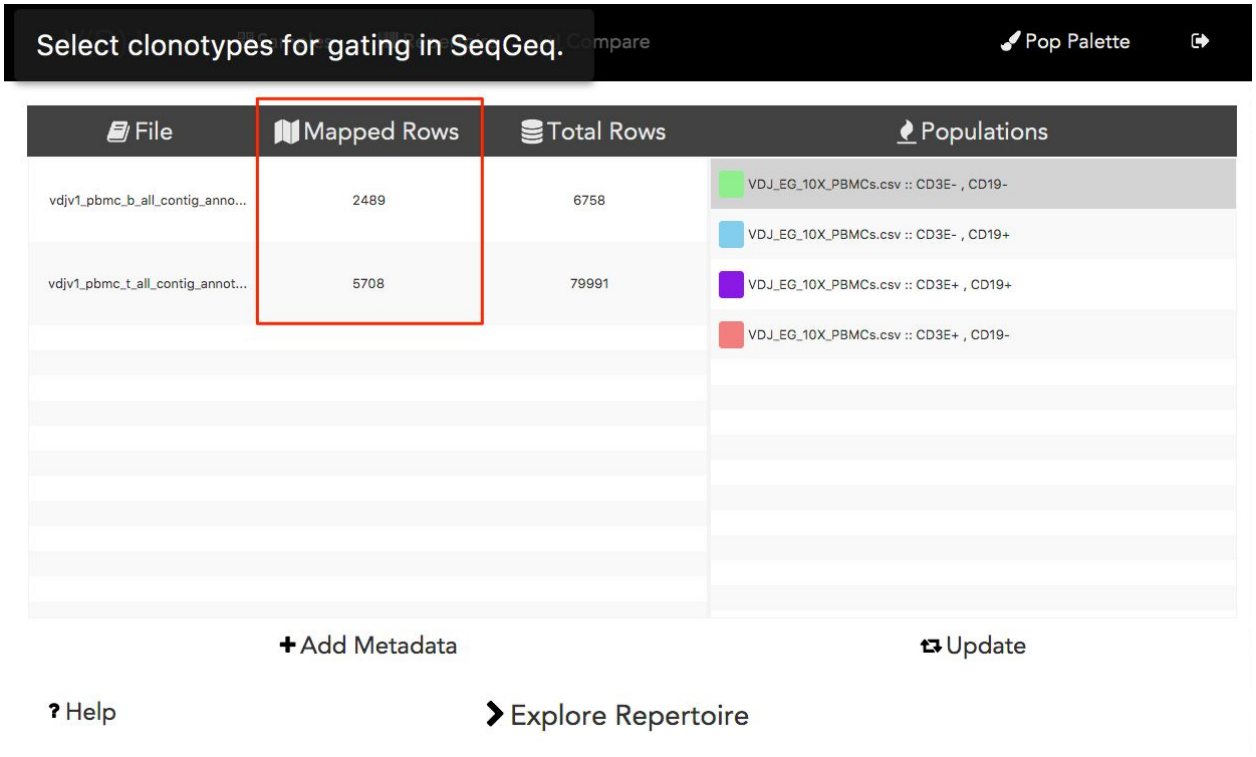
Once the plugin has been run on an appropriately sequenced gene expression matrix (GEX) file, it will require a researcher connect the file to their V(D)J CSV meta info file, usually named “all_contig_annotations.csv”. This is accomplished by clicking on the GEX file within SeqGeq, and opening the VDJ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



Click on “Add Metadata” within the resulting plugin dialog to choose the TCR and/or BCR meta-info CSV file(s) :



Selecting populations of interest for further comparison will indicate the number of total rows within the Metadata CSV mapped to the population(s) selected there:



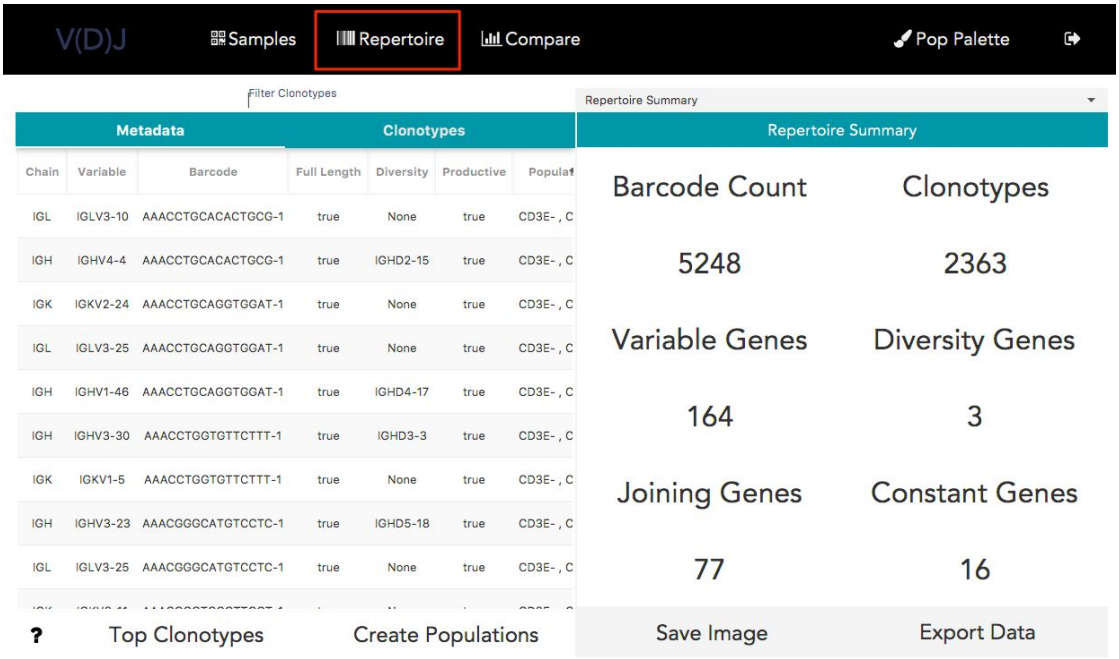
Note: Rows in the metadata do not directly correspond to a particular number of cells because each barcode (associated with a given cell) can appear multiple times within the V(D)J sequencing. This is due to the nature of V(D)J sequencing, wherein many T and B cells will generate many different chains.

Double-clicking on a population will deselect that node.

Repertoire

The next section of the V(D)J Explorer gives researchers the ability to filter their metadata file for cells corresponding to particular clonotypes of interest. This is achieved in part by stacked-bar charts which compare populations selected in the previous window.

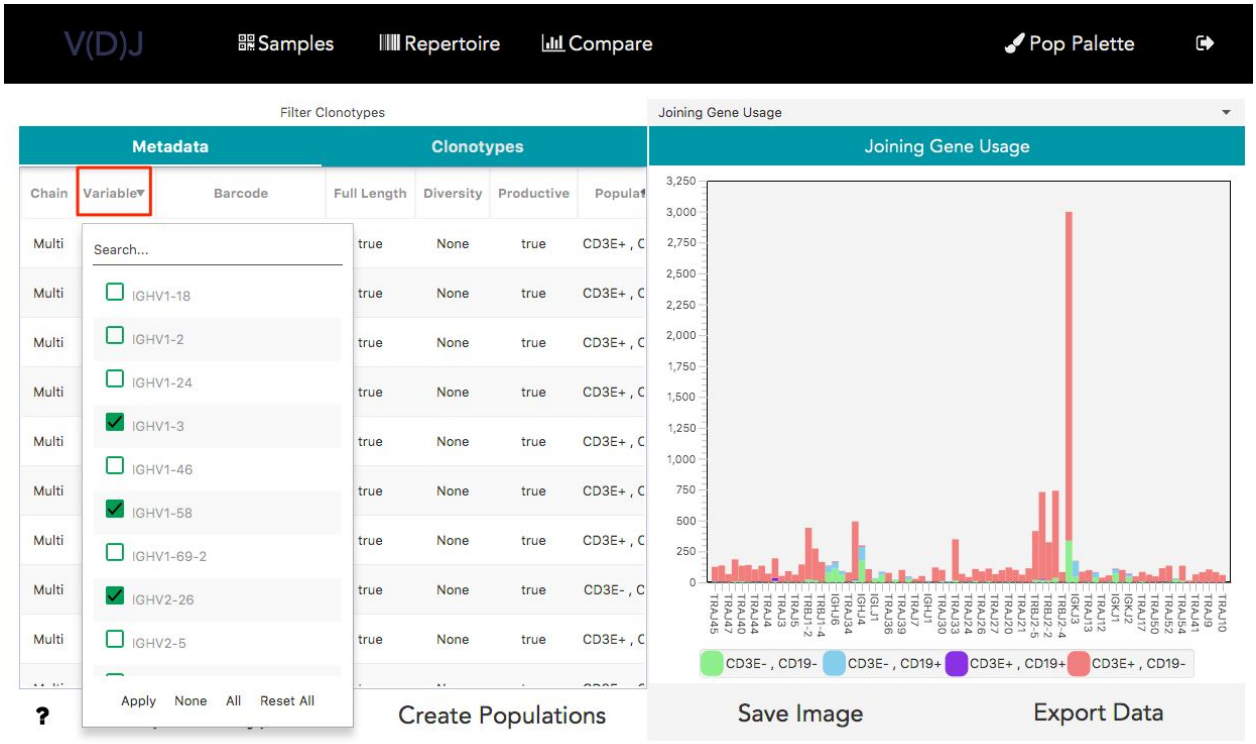
Initially you'll be presented with some summary information on the populations selected:



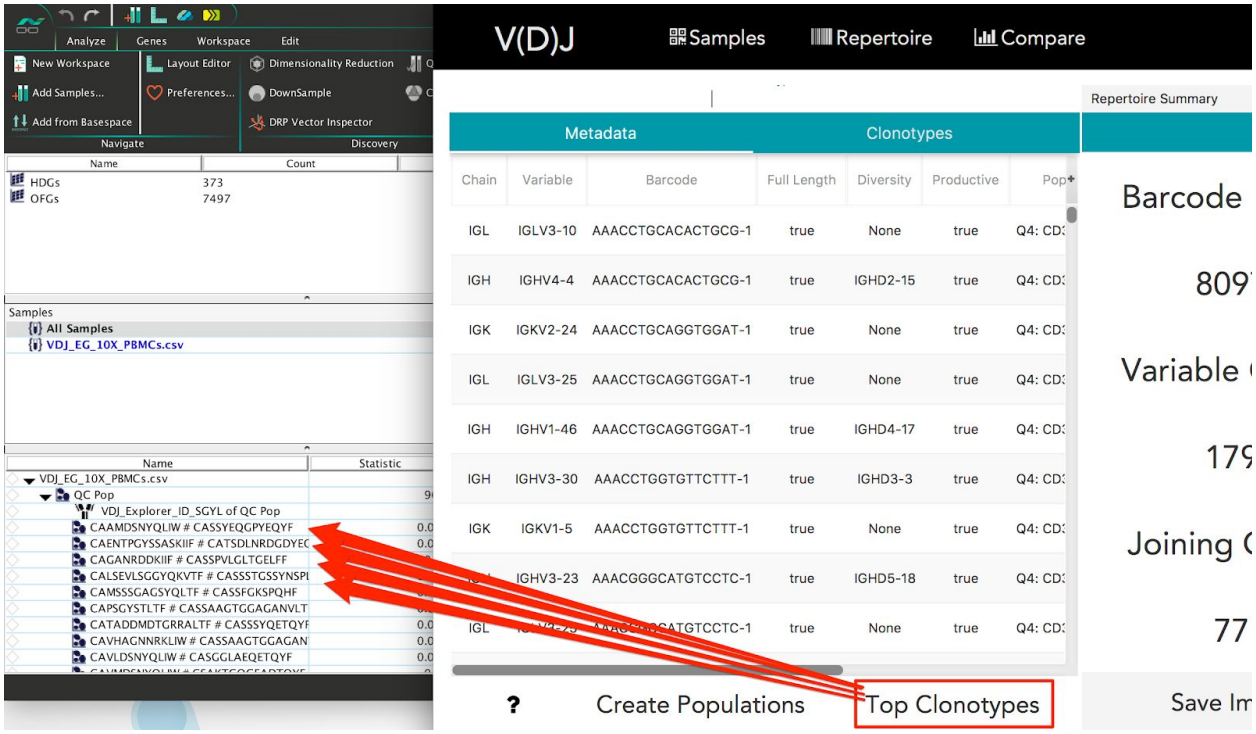
Choosing any of the numbers within that summary sections will illustrate the stacked-bar chart associated with the populations being compared. For example, this will allow you to view variable gene usage across the populations compared:

Other comparisons can be accessed at the top of the right hand pane in this Repertoire section.

Selecting populations within the stacked bar charts will filter the meta information on the left. Similarly you can right click on the column headers within the meta-information to filter there directly:

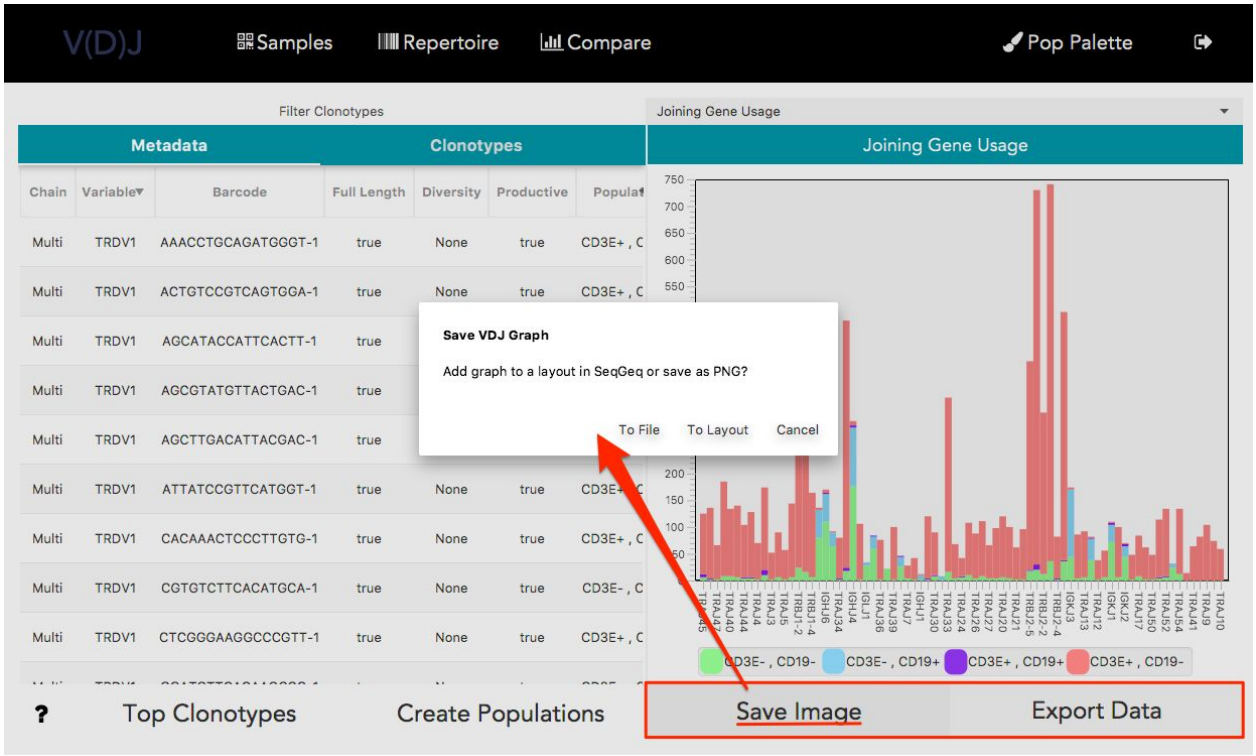


Selecting a set of rows from the metadata section of the plugin and clicking “Create Populations” will create corresponding populations within the workspace. Researchers can also create populations corresponding the top ten most frequent clonotypes using this plugin:



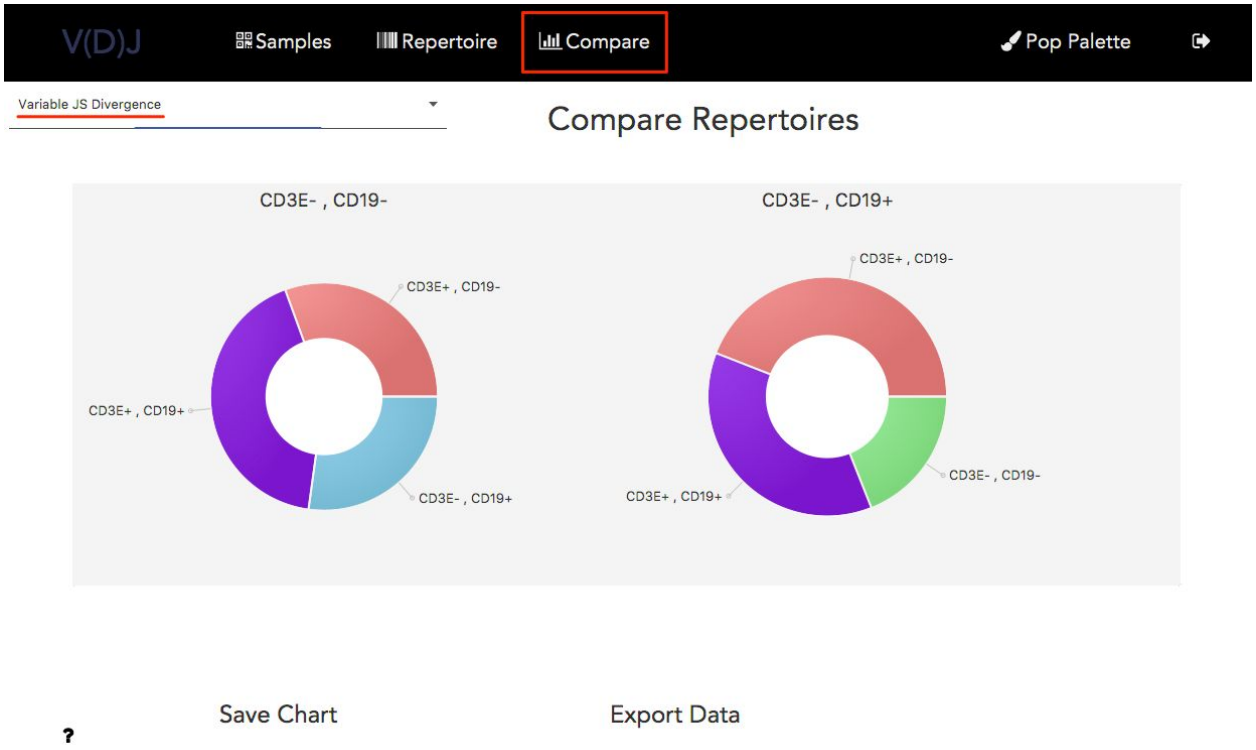
Note: The name of colonotypes corresponds to the Complementarity-Determining Region amino acid sequence of each chain.

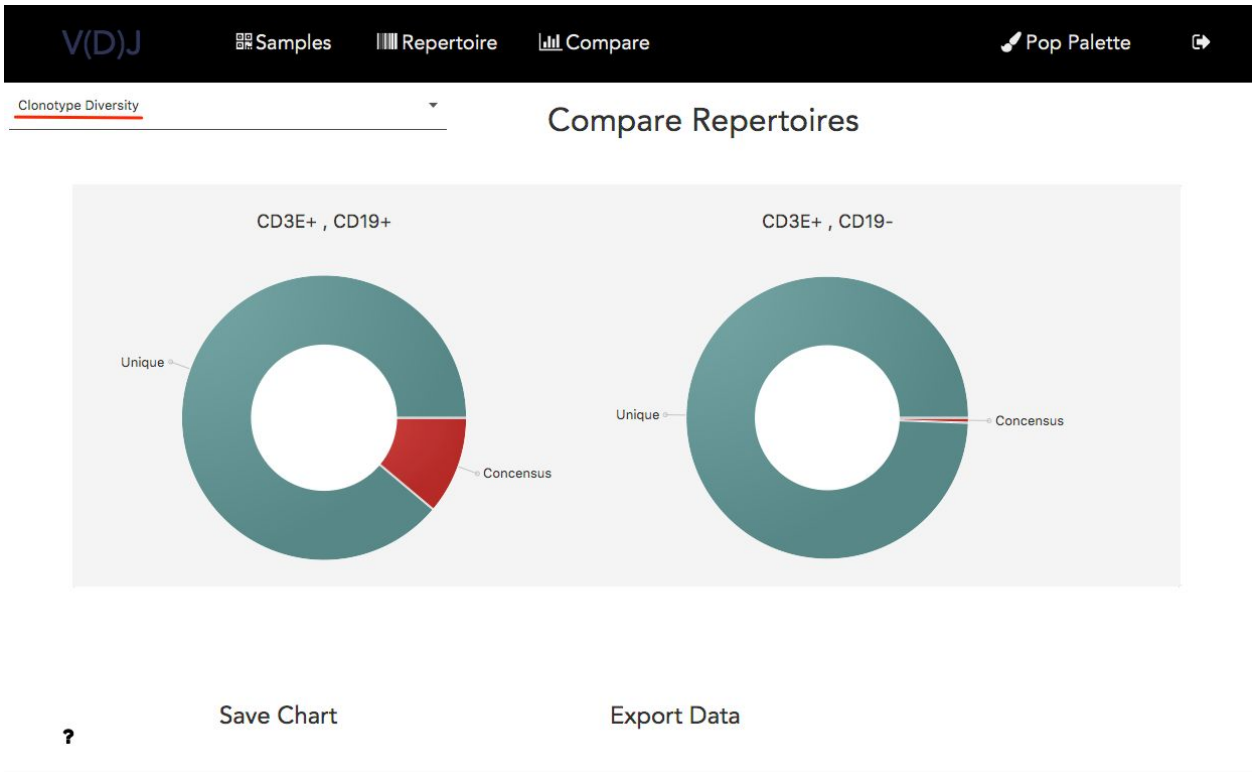
Information from tables and figures themselves can be exported from the plugin (as CSV information, PNG figures, or directly to the Layout Editor):



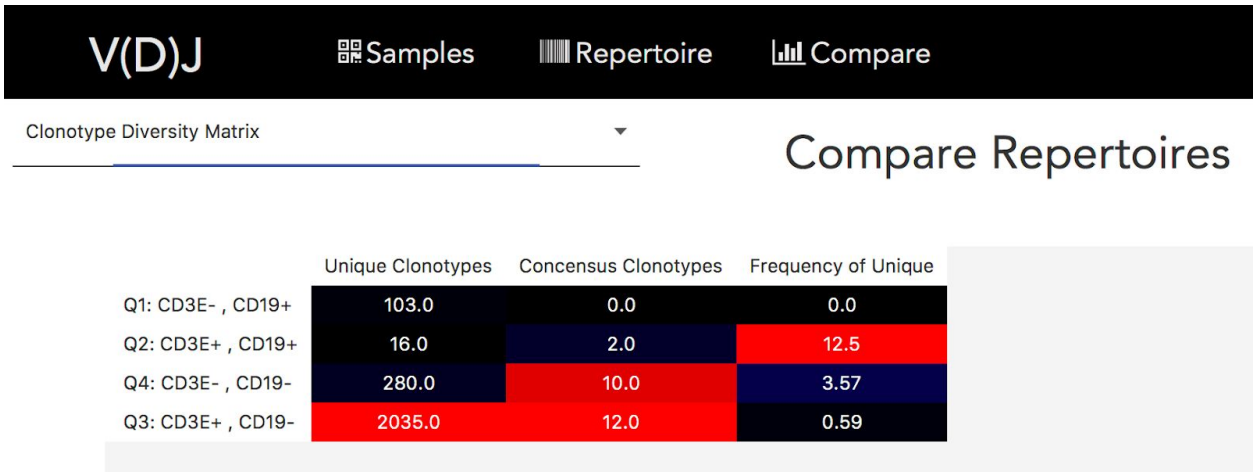
Compare

The final section of the V(D)J Explorer illustrates broader comparisons between populations. Comparisons available there are Jensen-Shannon Divergence (aka “Information Radius” or “Shannon Entropy Index“)(3) for Variable and Divergent regions, and Clonotype Diversity:





Note: These plots are also exportable to the Layouts in SeqGeq. And can be viewed as a heat-mapped table of values as well as a in pie-chart format:

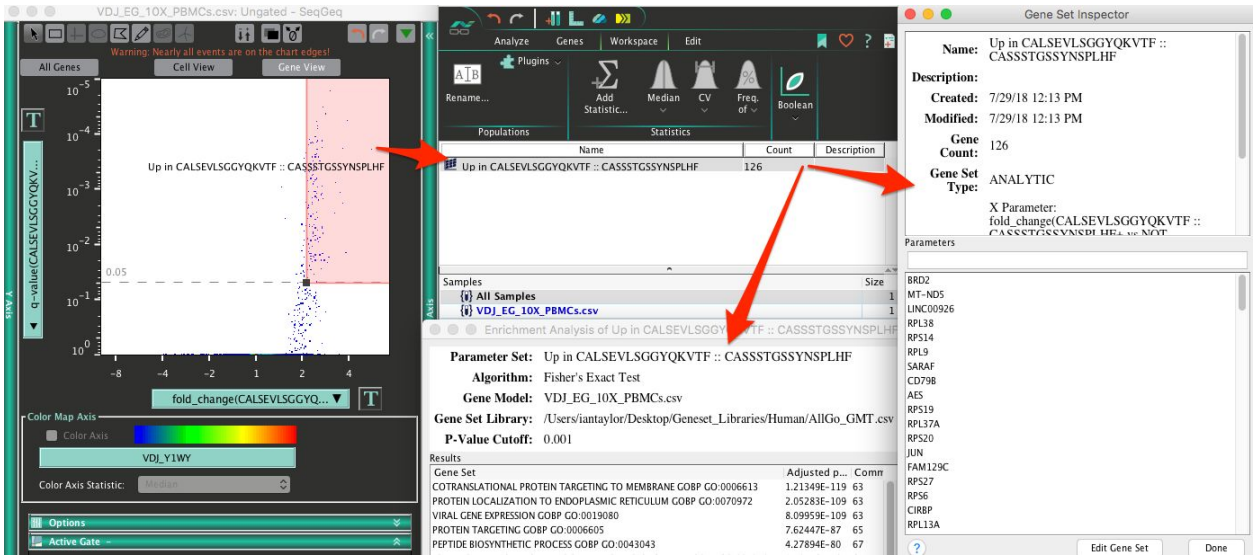


Downstream

Artifacts generated from the plugin, such as the most frequently occuring clonotypes can be used for further in depth analysis throughout SeqGeq's platforms.

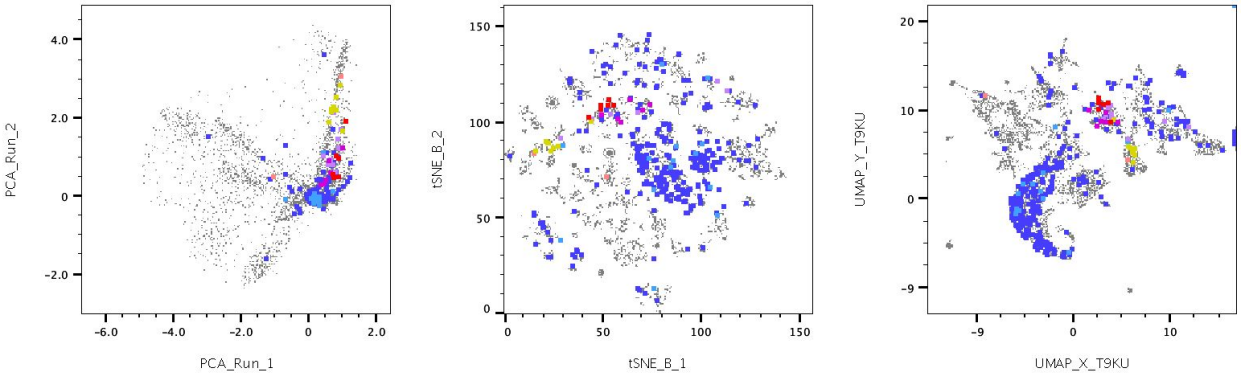
Differential Expression and Geneset Enrichment Analysis

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:

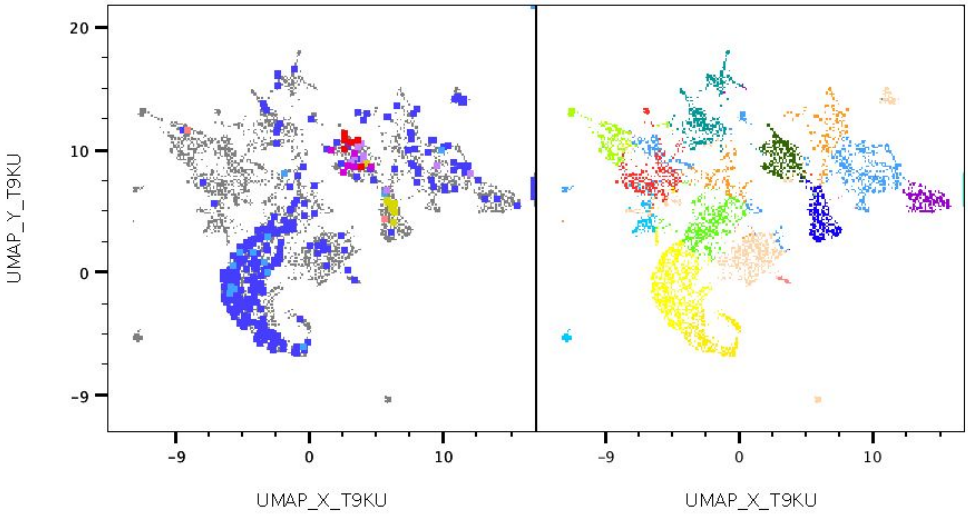


Clonotypes Within Clusters

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



	Subset Name
	CAYNNNDMRF :: CSVRTGKGTEAFF
	CAVRDSNYQLIW :: CATSREVAGGPSTDTQYF
	CAPSGYSTLTF :: CASSAAGTGGAGANVLTF
	CAVHAGNNRKLIW :: CASSAAGTGGAGANVLTF
	CAGANRDDKIIF :: CASSPVLGLTGELFF
	CALSEVLSGGYQKVTF :: CASSSTGSSYNSPLHF
	CAVMDSNYQLIW :: CSAKTGQGEADTQYF



	Subset Name
	CAYNNNDMRF :: CSVRTGKGTEAFF
	CAVRDSNYQLIW :: CATSREVAGGPSTDTQYF
	CAPSGYSTLTF :: CASSAAGTGGAGANVLTF
	CAVHAGNNRKLIW :: CASSAAGTGGAGANVLTF
	CAGANRDDKIIF :: CASSPVLGLTGELFF
	CALSEVLSGGYQKVTF :: CASSSTGSSYNSPLHF
	CAVMDSNYQLIW :: CSAKTGQGEADTQYF

	Subset Name
	PhenoGraph_Cluster_15
	PhenoGraph_Cluster_14
	PhenoGraph_Cluster_13
	PhenoGraph_Cluster_12
	PhenoGraph_Cluster_11
	PhenoGraph_Cluster_10
	PhenoGraph_Cluster_9
	PhenoGraph_Cluster_8
	PhenoGraph_Cluster_7
	PhenoGraph_Cluster_6
	PhenoGraph_Cluster_5
	PhenoGraph_Cluster_4
	PhenoGraph_Cluster_3
	PhenoGraph_Cluster_2
	PhenoGraph_Cluster_1

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)