VDJ: Visualization for T- and B-cell repertoires

This guide describes VDJ, a web-based tool for visualizing and interpreting the results of immunosequencing assays marketed by Adaptive Biotechnologies.

Adaptive offers a hosted version of the tool to its customers at https://vdjtool.azurewebsites.net. The code for VDJ is open-source, available on GitHub under MIT license. Details on modifying, building and hosting the application are available in the GitHub repository.

Note: while the tool currently supports only files generated by Adaptive Biotechnologies, it could trivially be updated to support alternative formats such as iRepretoire's pep.csv.

Contents

Login and Credentials

You'll be asked to log in before using any VDJ functionality. If you are using the Adaptive-hosted version of the tool, use the "Agate" credentials you received from Adaptive. Agate uses Microsoft Entra authentication, so your account name may be your own email or a new one with the domain "adaptiveagate.onmicrosoft.com."

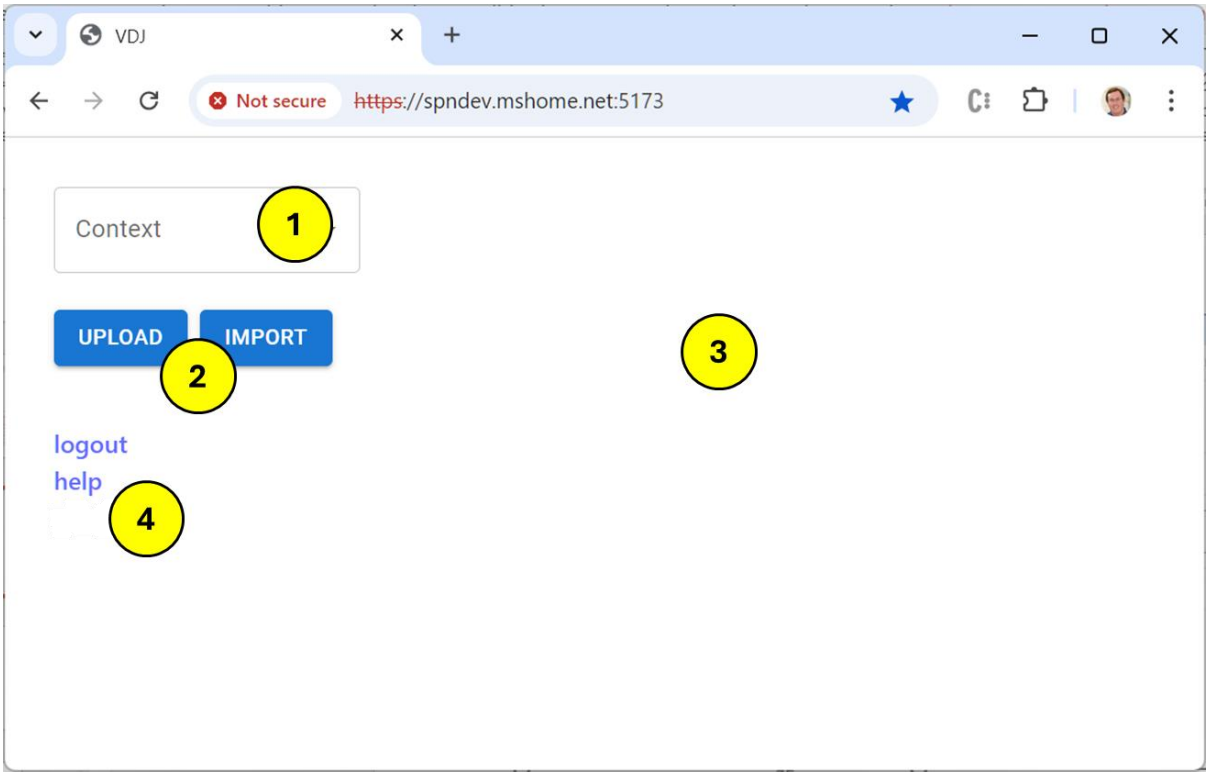
Note: If you are using another instance of VDJ (self-hosted or otherwise), your login may be a simple username and password, a corporate email, or it may defer to another third-party login system like GitHub, depending on how the administrator has configured the version of the application you're using.

Overview and Navigation

The main VDJ interface consists of a left-side navigation/action bar and a right-side content area. Related "repertoires" (samples) are grouped together into "contexts" --- a context may represent a patient, a study, or any group of samples that are analyzed together.

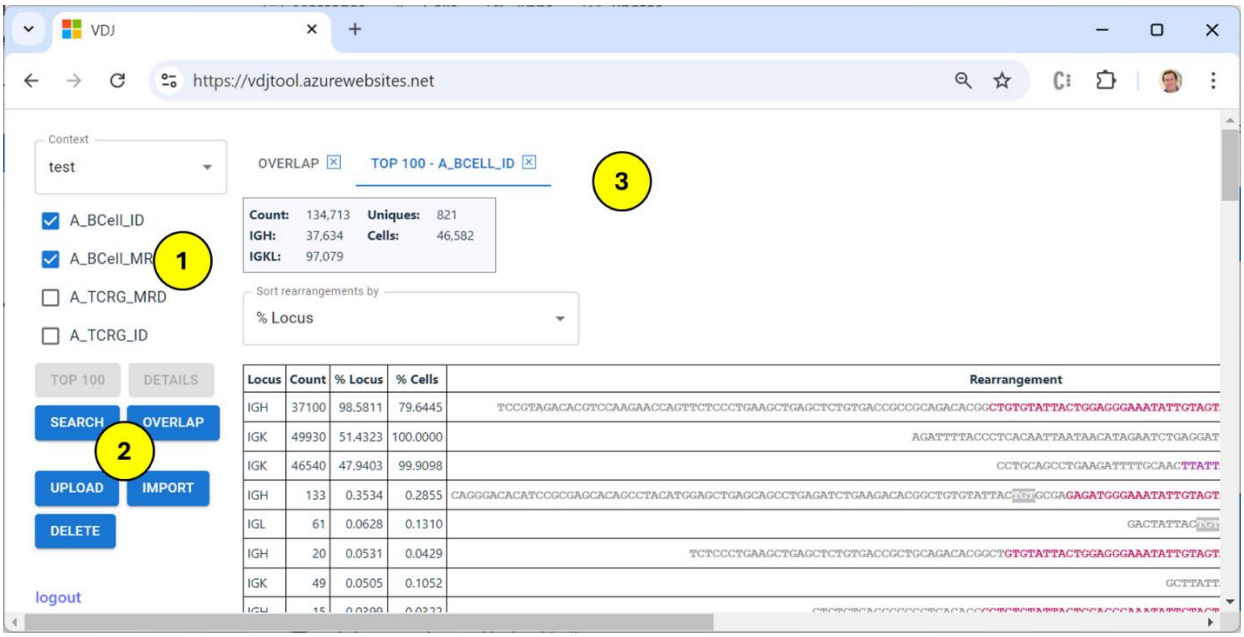
Note: by default each user will see their own collection of contexts and repertoires. If a group of users regularly collaborates on the same data, the VDJ administrator can assign those accounts to a "group" so that data is shared.

On initial login, no context is "active" and the screen will appear like this:



1. Click the context menu to see a list of available contexts and select one to work with. You can also type into this box to search for a context rather than scroll through them all.
2. The "Upload" and "Import" buttons are always available; this is how repertoires are added to your data set and are described in the "Managing Repertoires" section.
3. The right-hand side of the application starts out empty. This is where content will appear as you begin to work with repertoires.
4. The "logout" button terminates your working session; "help" will open the user manual in a new tab.

Once a context is selected using the dropdown, the repertoires in that context will be shown on the navigation bar and new action buttons will become available:



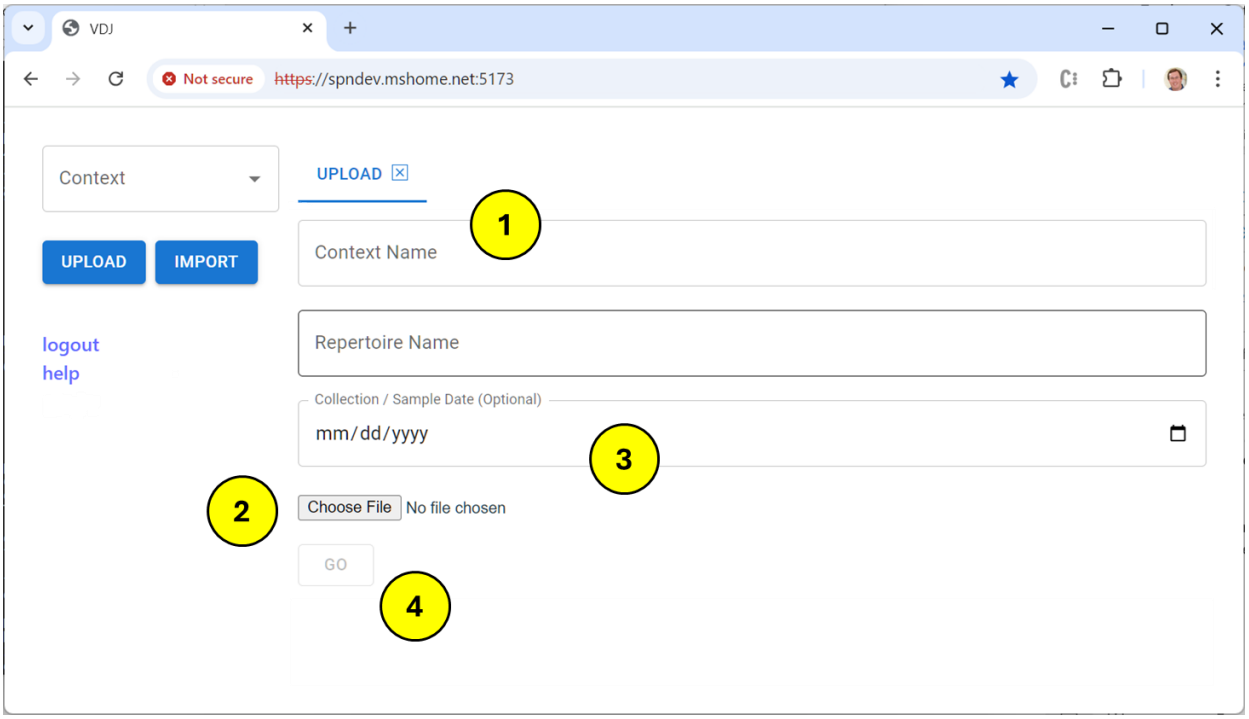
1. The repertoires in the context are shown in a list; use the checkboxes to select/unselect repertoires. Click on any repertoire and hit “control-a” to select all.
2. Buttons perform actions on the selected repertoires. They will be enabled/disabled based on the current selection --- e.g., “Top 100” can only be used when exactly one repertoire is selected, and “Overlap” requires at least two.
3. Actions open their results on the right side of the screen. Multiple result panes can be open at once; the “tabs” at the top can be used to switch between them. Use the “X” icon to close a tab when finished with it. When a new context is selected using the dropdown, all content tabs are cleared.

Managing Repertoires

Repertoires are added to each context either by uploading them from an Adaptive “TSV” file or importing them directly from Agate.

Upload

Adaptive “TSV” files can be exported from the immunoSEQ Analyzer, or you may receive copies from the client services team. Each TSV represents a single repertoire from a particular assay (e.g., B-Cell or TCRB) and can be added to a context using the “Upload” button.



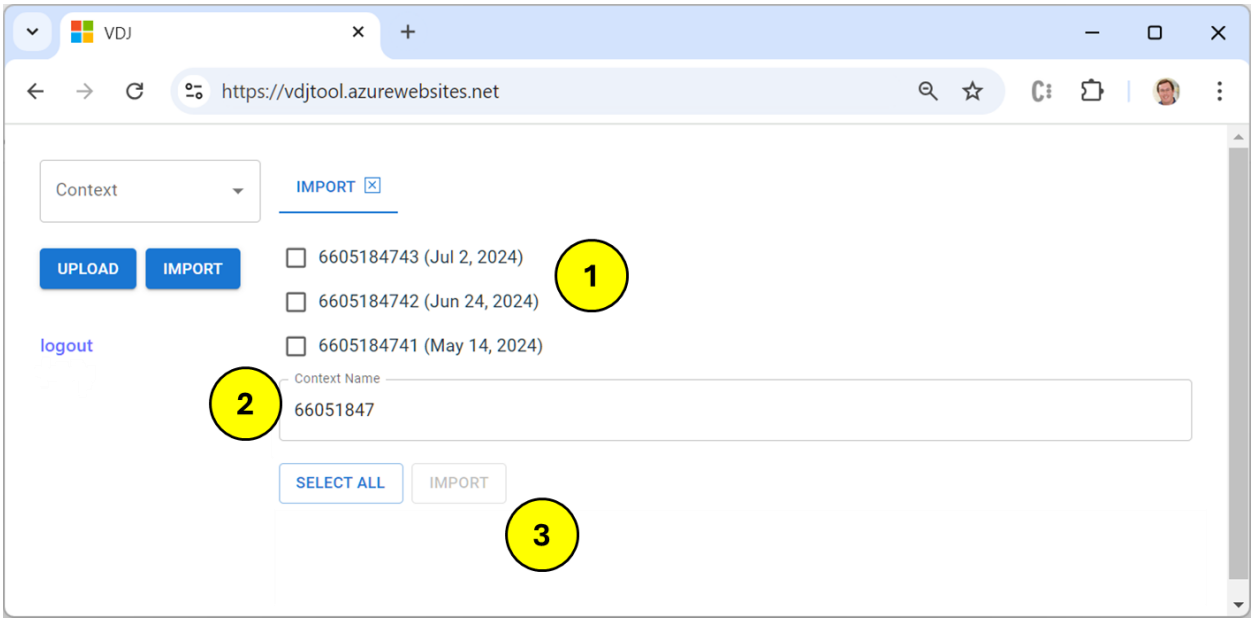
1. Enter the context in which the repertoire should be saved. This context does not have to already exist; new ones will be created automatically.
2. Choose a TSV file on your local machine to upload. These files may end in ".tsv" or they may be compressed (".tsv.gz" or ".tsv.zip"). Using a compressed file is recommended; doing so can dramatically reduce upload time. Choosing a file fills in the "Repertoire Name" field but you may update it if desired.
3. Optionally, provide a collection date for the sample. This value is used for display and sorting in some visualizations, but is not required.
4. Click the "Go" button to start the upload. After the upload is complete, close the tab using the "X" at the top of the content pane.

If this is the first repertoire in a new context, refresh your browser so it appears in the "context" dropdown. If the context already exists, the new repertoire will show up automatically.

Import

Note: If you are not using the Adaptive-hosted version of the tool, Agate import may or may not be available. If it is, you will be prompted to enter your Agate credentials, which will be different from the login and password used to log into VDJ.

Adaptive stores customer data in a cloud-based system called "Agate." Repertoires stored in Agate can be directly imported into VDJ. Start an import by clicking the "Import" button and entering a search string. The search is case-insensitive and searches with sample, project and order names. This query can take a bit of time to show results:



1. Check the boxes for samples to be imported. To select all listed samples, use the “Select All” button.
2. Verify the context for samples. VDJ attempts to pick a reasonable context name, but you can use whatever name works best for you.
3. Click “Import” to start the process; the resulting page will show progress.

Note: If you encounter an error during import, try logging out and logging back in again. This will refresh your access token and may resolve the problem.

If a selected repertoire already exists in the desired context, it will be skipped. This makes it easier to update data for a context over time (e.g., clonoSEQ results for a patient). Just search on the patient identifier, “select all” and then import --- new repertoires will be added and existing ones ignored.

As with Update, if this is the first repertoire in a new context, refresh your browser so it appears in the “context” dropdown. If the context already exists, new repertoires will show up automatically.

Delete

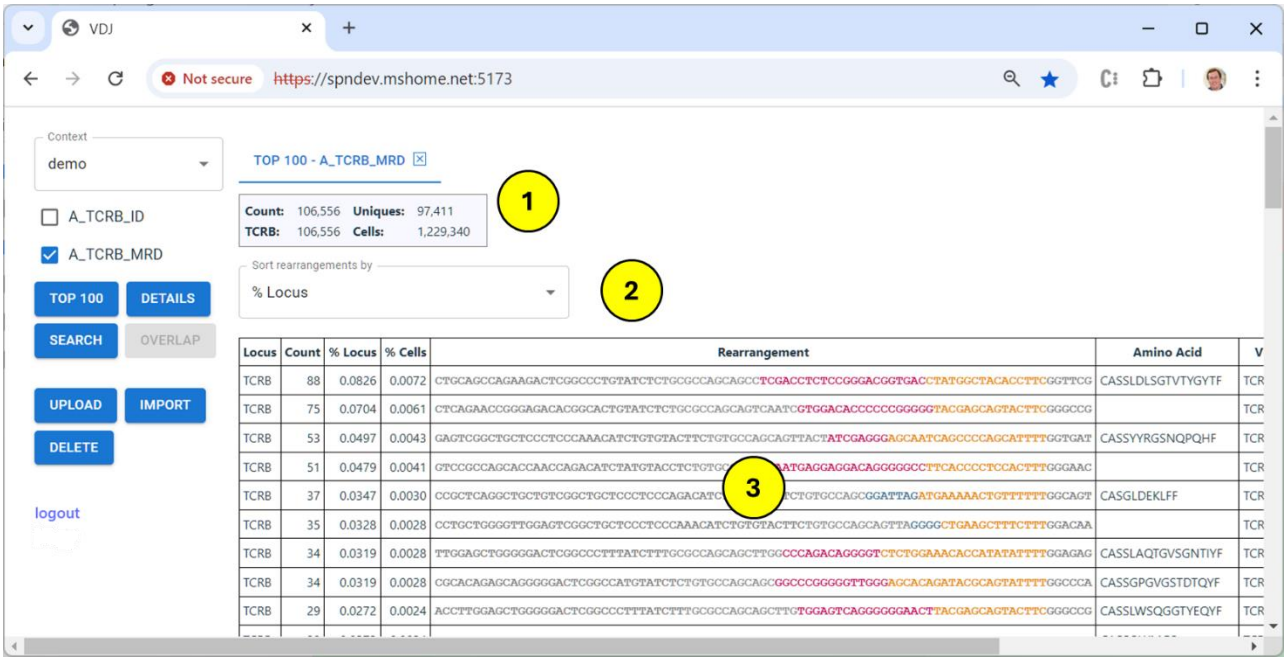
The “Delete” button can be used to remove repertoires present in the current context.

When the last repertoire is removed from a context, that context is automatically removed as well. You can recreate it at any time by Uploading or Importing repertoires using the original context name.

Actions

Top 100

This is the most common action for viewing details of a single repertoire. To make the feature performant, only the top 100 rearrangements sorted descending by the selected metric (% of Locus by default) are displayed.



1. The header shows aggregate values for the repertoire:
    a. Total sequence count
    b. Sequence counts grouped by locus group (TCRAD, TCRB, TCRG, IGH, IGKL)
    c. Total number of unique sequences
    d. Count of cells in the sample, or milliliters for cell-free repertoires.
2. The dropdown changes the sort metric:
    a. Simple Count
    b. Percentage of this sequence within cell count (or milliliters for cell-free)
    c. Percentage of this sequence within its locus group, by count
3. The results tale contains the following columns:
    a. Locus (TCRAD, TCRB, TCRG, IGH, DJ, IGK, IGL)
    b. Count as estimated by the assay
    c. % Locus
    d. % Cells (or Count/ML for cell-free)
    e. Nucleotide rearrangement sequence. Clicking this value will open up an analysis using IMGT V-Quest.
    f. Amino Acid, if the sequence is productive
    g. Identified V, D and J genes or alleles (depending on assay confidence)

h. Probability of sequence uniqueness, if available (log10)

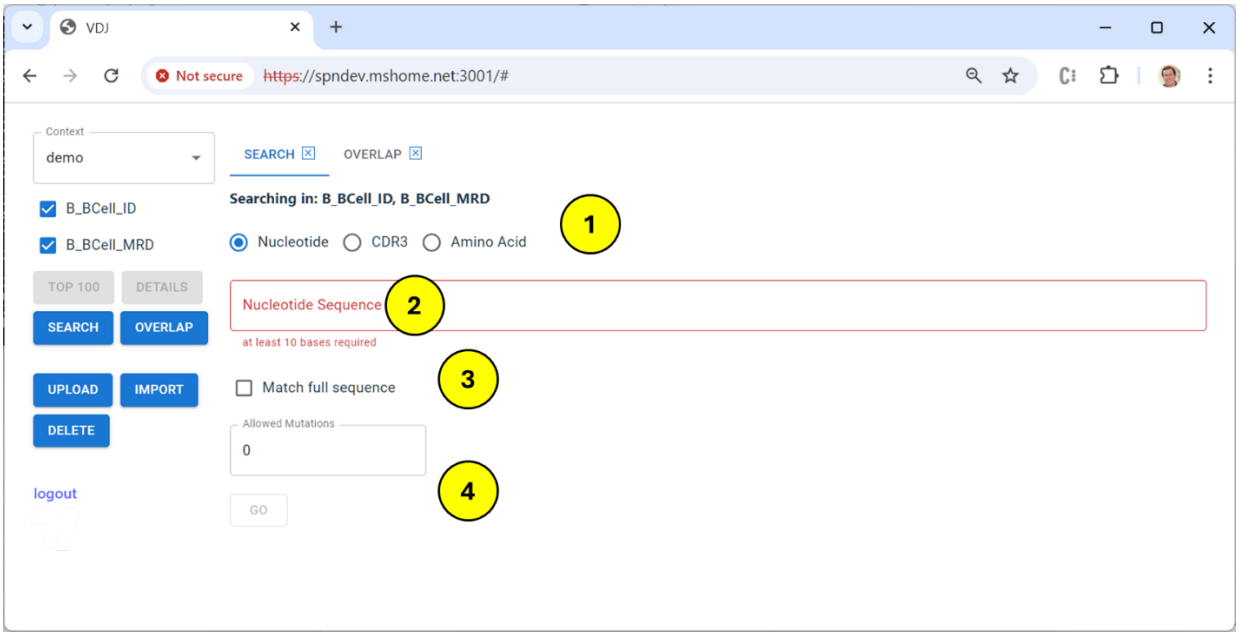
Note: The "rearrangement" column is colored to identify aligned regions: V, N1, D, N2 and J. However, this code is currently not particularly reliable --- an update is pending to make this a useful feature.

Details

The details action uses the same display format as Top 100, but returns all rearrangements in the repertoire. The rearrangements are unsorted and appear in the order they are present in the source TSV file. The "FORWARD" and "BACK" links page through the table.

Search

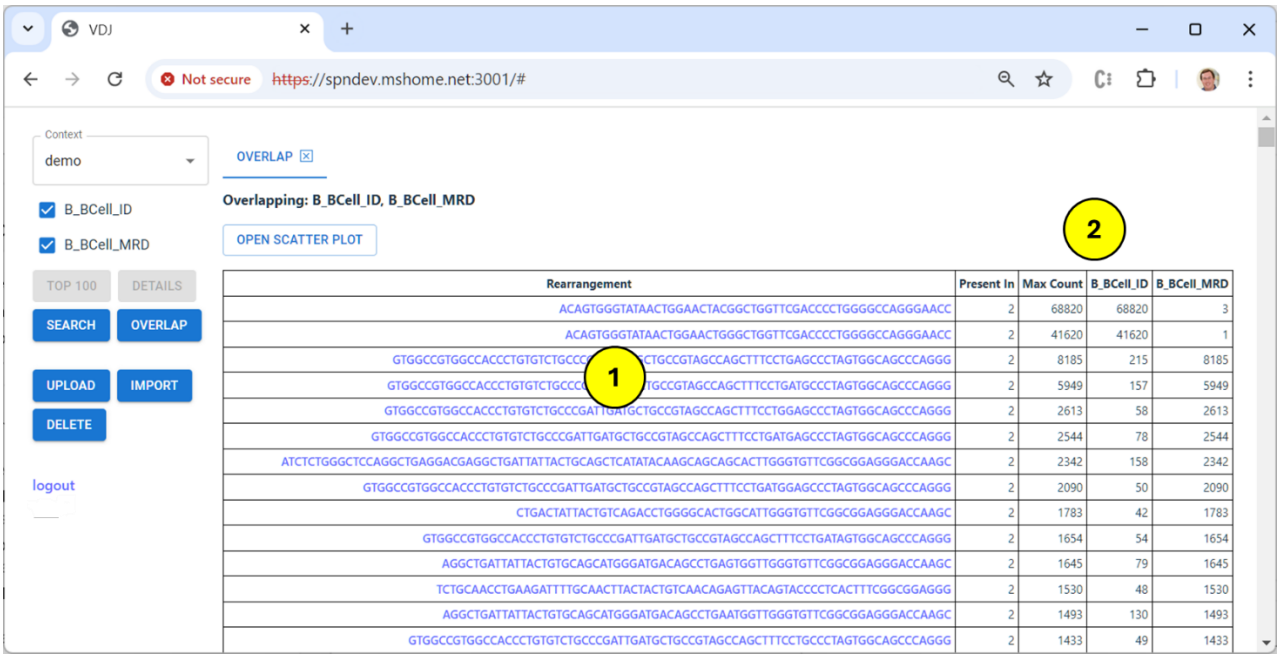
VDJ can search multiple samples for instances of a target nucleotide or amino acid sequence.



1. Select a target for the search (i.e., the "haystack"):
    a. Nucleotide: anywhere in the nucleotide sequence
    b. CDR3: the identified CDR3 region, a subset of the full sequence
    c. Amino Acid: the translated protein sequence for the CDR3 region
2. Enter the nucleotide or protein sequence to search for (i.e., the "needle")
3. If "Match full sequence" is checked, the needle and the haystack must match exactly, otherwise the needle may be any substring of the haystack. For example, the amino acid sequence "**CAEENWN**" will match the sequence "**CAEENWN**YGWFDPW" only if this box is UNchecked.
4. "Allowed Mutations" accepts a given number of mismatches (i.e., substitutions) between the needle and the haystack. For example, if this value is one, the needle "CAEEN**W**N" will match the haystack "CAEEN**A**N". This parameter works only for substitutions, not insertions or deletions.

Overlap

The Overlap action identifies commonality between two or more repertoires (currently set to a maximum of six). Comparison can be done by full nucleotide sequence, identified CDR3, or amino acid sequence.

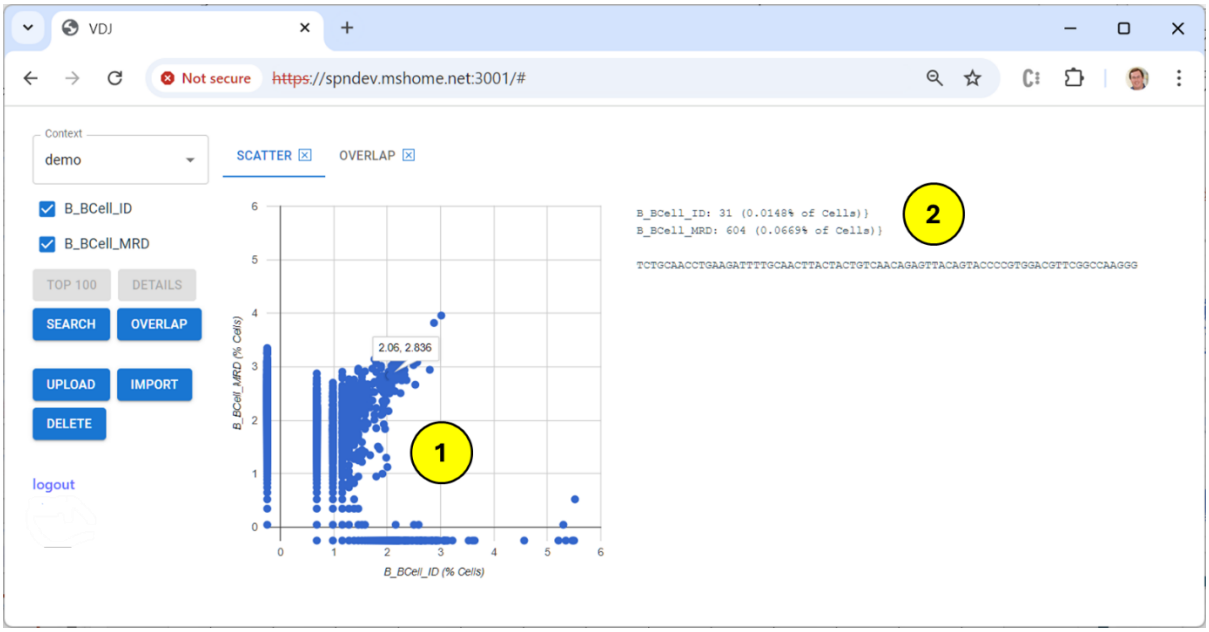


1. The resulting table shows the overlapping nucleotide or protein sequence. Note that in the case of "CDR3" matching this is a substring of the full nucleotide sequence. Clicking on a sequence will open a new "Search" tab showing detailed results.
2. Other columns in the table provide information on the overlap:
    a. "Present In" specifies the number of samples containing the overlap.
    b. "Max Count" is the maximum count found in any sample.
    c. Each repertoire is then represented in a column with its count value.

Only sequences that appear in at least two repertoires are returned in this table, which is sorted descending first by "Present In" and then by "Max Count."

Overlap Scatter Plot

When the Overlap action is run on exactly two repertoires, an additional option "Open Scatter Plot" will be present on the results pane. Clicking this button will open a comparison chart between the two repertoires using the same metric.



The chart uses a log10 scale to plot normalized values. This typically a percentage of cells, but can be count per milliliter for cell-free repertoires). Sequences with equivalent abundance in both repertoires will trend along the SW-NE diagonal line; those with differential abundance will fall in the top-left or bottom right quadrant.

Clicking on a data point will display the sequences it represents including exact count values.

Unlike the overlap results table, ALL sequences in both sequences are represented on the chart, including those that appear in one of the two repertoires. These values are plotted just below and to the left of the axis lines.

Export

Use this action to safe a repertoire file to your local computer:

1. “TSV” is exactly the file that was added to the VDJ tool by upload or import.
2. “FASTA” is a FASTA-compliant field with one entry per sequence in the repertoire. Two versions are available to assist in mapping output using this file:
    a. “by Row” sets the description for each sequence to a 0-based row number.
    b. “by Hash” sets the description for each sequence to the SHA256 value of the sequence itself.