

Click the orange “Initiate Analysis” button at the bottom of the page. To export the results, click the dark blue “Export” button to the upper left of the table, which will then download the file in a .zip format. See Section 3.6, “Viewing exported data.”

3.4.12 SAMPLE SUMMARY

Sample Summary is available for two or more samples of any locus. It returns a table with a per-sample summary of the types and quantities of sequences observed, as well as some overall sample statistics. Select your samples in the “Choose Samples” box and click the “Sample Summary” icon in the “Choose Analysis” box. You have three options: Show Max Frequency, which includes the maximum frequency for each sample, Gene Rearrangements (number of rearranged genomes present), and Filter DJ Sequences, which is by default checked to filter out DJ sequences, leaving only VDJ rearrangements. Click the orange “Initiate Analysis” button toward the bottom of the page.

Sample Summary: Multiple Samples				
Sample Summary				
Export				
Filters	Sequences	% of Total Sequences	Unique Sequences	% of Total Unique Sequences
Productive	16,315,773	85.625%	214,942	84.966%
Out of frame	2,502,185	13.132%	33,936	13.415%
Has stop	236,870	1.243%	4,097	1.62%
Total	19,054,828		252,975	

Individual Sample Summary												
Export												
Show 10 entries												
Sample Name	Total	Unique	Productive Total	Productive Uniques	Out of Frame Total	Out of Frame Uniques	Has Stop Total	Has Stop Uniques	Entropy	Clonality	Max Frequency	Gene Rearr.
TCRB_Control1	6,601,122	79,496	5,769,622	68,872	753,684	9,808	77,616	816	15.034514	0.06453084	1.082%	0
IGH_Memory_Control2	1,591,482	39,671	1,395,784	33,661	166,007	5,045	29,691	965	14.443698	0.039570477	0.347%	47,65
TCRB_Control2	9,647,898	95,622	8,078,094	78,690	1,459,722	14,467	110,082	1,465	14.29152	0.12225631	1.504%	0
IGH_Memory_Control1	1,214,326	38,186	1,072,073	32,719	122,772	4,616	19,481	851	14.412711	0.039014272	0.359%	45,03

Showing 1 to 4 of 4 entries First Previous 1 Next Last

You will see two tables. The top table is a summary of all of the combined samples. The bottom table is a summary of each individual sample. The all-samples summary shows the numbers and percentages of total and unique sequences for productive and non-productive clones. The individual sample summary shows more details about each sample including numbers of total and unique sequences that are both productive and non-productive, as well as entropy, clonality, maximum frequency, and gene rearrangements. To see all columns in the bottom table, you may need to place your cursor over the table and scroll to the right. You can export the tables via the dark blue “Export” button to the upper left of each table. The tables will export as a .tsv file.

3.4.13 SAMPLE EXPORT

Sample Export is available for any number of samples. Select the samples you want to export in the “Choose Samples” box and click the “Sample Export” icon under “Choose Analysis.” You may need to scroll down to view this option. For IGH samples, you can select whether to include DJ

sequences (by default these are filtered out and you will only view V(D)J sequences).

Click the orange “Initiate Analysis” button at the bottom of the page. The file will be downloaded as a zip file. It can be expanded into a folder that contains one .tsv file per sample. These tab-separated value files (.tsv files) can be opened using a program such as Microsoft Excel. You must set Excel to enable all files to be opened, not just “readable” files.

The column headers in the exported data are defined in the table below:

Column title	Definition
nucleotide	Full length nucleotide sequence
aminoAcid	CDR3 amino acid sequence
count	The count of sequence reads for the given nucleotide sequence
frequencyCount	The percentage of a particular nucleotide sequence in the total sample
cdr3Length	Total length of the CDR3 region
vMaxResolved	The highest degree of specificity obtained when identifying the V gene
vFamilyName	Name of the V family (if identified)
vGeneName	Name of the V gene if identified, otherwise 'unresolved'
vGeneAllele	Number of the V gene allele if identified, otherwise '1'
vFamilyTies	Name of the potential V families for a sequence if V family identity is unresolved
vGeneNameTies	Name of the potential V genes for a sequence if V gene identity is unresolved
vGeneAlleleTies	Name of the potential V gene alleles for a sequence if V allele identity is unresolved
dMaxResolved	The highest degree of specificity obtained when identifying the D gene
dFamilyName	Name of the D family (if identified)
dGeneName	Name of the D gene if identified, otherwise 'unresolved'
dGeneAllele	Number of the D gene allele if identified, otherwise '1' or 'unresolved' if D gene is unresolved
dFamilyTies	Name of the potential D families for a sequence if D family identity is unresolved
dGeneNameTies	Name of the potential D genes for a sequence if D gene identity is unresolved
dGeneAlleleTies	Name of the potential D gene alleles for a sequence if D allele identity is unresolved
jMaxResolved	The highest degree of specificity obtained when identifying the J gene
jFamilyName	Name of the J family (if identified)
jGeneName	Name of the J gene if identified, otherwise 'unresolved'
jGeneAllele	Number of the J gene allele if identified, otherwise '1'
jFamilyTies	Name of the potential J families for a sequence if J family identity is unresolved
jGeneNameTies	Name of the potential J genes for a sequence if J gene identity is unresolved

jGeneAlleleTies	Name of the potential J gene alleles for a sequence if J allele identity is unresolved
vDeletion	The number of deletions from the 3' end of the V segment
n1Insertion	The number of insertions in the N1 region
d5Deletion	The number of deletions from the 5' end of the D segment
d3Deletion	The number of deletions from the 3' end of the D segment
n2Insertion	The number of insertions in the N2 region
jDeletion	The number of deletions from the 5' end of the J segment
vIndex	Distance from the start of the sequence (designated 0) to the conserved cysteine residue in the V motif
n1Index	Distance from 0 to the start of the N1 region
dIndex	Distance from 0 to the start of the D region
n2Index	Distance from 0 to the start of the N2 region
jIndex	Distance from 0 to the start of the J region
estimated NumberGenomes	Estimated number of T/B cell genomes present in the sample
sequenceStatus	Whether the nucleotide sequence generates a functional amino acid sequence, or a sequence with a stop codon or out-of-frame sequence
cloneResolved	The gene segments used to construct the CDR3 (VDJ, VJ, DJ)
vOrphon	Not implemented at this time
dOrphon	Not implemented at this time
jOrphon	Not implemented at this time
vFunction	Not implemented at this time
dFunction	Not implemented at this time
jFunction	Not implemented at this time

3.4.14 EXPLORE EDIT DISTANCE

Explore Edit Distance is available for any single sample and visually represents the relationship between clones. To run this analysis, first select your sample in the “Choose Samples” box and then click the “Explore Edit Distance” icon in the “Choose Analysis” box. There are no configuration options for this analysis type, so simply click the orange “Initiate Analysis” button at the bottom of the page. The pull-down menus will give you the option to choose among particular V and J gene and CDR3 lengths and will show the number of unique clones fitting those criteria. These clones are represented by multicolored balls. The pull-down menu on the right will bin the clones based on the number of substitutions between the clones. This changes the representation by drawing a line between the multicolored balls.

3.4.15 SAMPLE OVERLAP

Sample Overlap is available for at least two to as many as 25 samples of any locus. It produces both an overlap metric as well as a Pearson's correlation (r^2) value for each possible pair-wise percent sharing between all pairs of samples. This is computed by averaging across the two ratios of shared reads over total reads for each sample. The correlation of frequencies for all shared clones is given by the r^2 .

	IGH_Memory_Control1	IGH_Memory_Control2	IGH_Naive_Control1	IGH_Naive_Control2
IGH_Memory_Control1		$r^2: 0.710$ $\alpha: 0.215$	$r^2: 0.000$ $\alpha: 0.004$	$r^2: 0.006$ $\alpha: 0.005$
IGH_Memory_Control2	$r^2: 0.710$ $\alpha: 0.215$		$r^2: 0.014$ $\alpha: 0.005$	$r^2: 0.011$ $\alpha: 0.006$
IGH_Naive_Control1	$r^2: 0.000$ $\alpha: 0.004$	$r^2: 0.014$ $\alpha: 0.005$		$r^2: 0.254$ $\alpha: 0.007$
IGH_Naive_Control2	$r^2: 0.006$ $\alpha: 0.005$	$r^2: 0.011$ $\alpha: 0.006$	$r^2: 0.254$ $\alpha: 0.007$	