

Analyzer Export: Guide to the .tsv File Data

OVERVIEW

Data is exported from the immunoSEQ Analyzer as .tsv files. The following guide provides the column heading names and a data definition for each. Please note, there are some small column name differences between the .tsv file and the name displayed in Analyzer. These differences are noted in the Tables below.

Both sample- and rearrangement-level .tsv files can be exported. Sample-level files include a list of sample names and summary metrics for each sample; including total template counts, the locus sequenced, the counting method version of the assay, sample tags which describe the sample meta-data. Rearrangement-level files include each unique nucleotide sequence (rearrangement), the translated amino acid sequence, the template counts for each rearrangement, the estimated proportion or frequency of each rearrangement and other inferred metrics about each rearrangement.

TERMINOLOGY/DEFINITIONS

Some special definitions apply to the descriptions in this document, which are explained here (in alphabetical order):

Clonality

Clonality is an evenness index equal to $1 - \text{normalized Shannon's Entropy}$ (see below). Values for clonality range from 0 to 1, with values near 1 representing samples with one or a few rearrangements (monoclonal or oligoclonal samples) dominating the observed repertoire. Clonality values near 0 represent more polyclonal samples. For the TCRB assay, the median clonality of an adult T-cell repertoire in blood is about 0.075.

Entropy

Refers to Shannon's Entropy, a diversity metric commonly used in population biology. Samples with high entropy have a greater number of rearrangements and little dominance by a sub-set of rearrangements, while low entropy samples have fewer rearrangements with greater dominance by a subset. Entropy increases with the number of unique rearrangements, making this metric highly sensitive to small differences in sample size. Generally, only the normalized metric is used.

Index

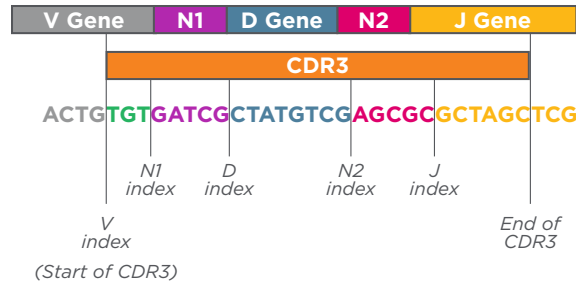
For TCRB samples, annotation delineates the location of gene regions within the rearranged VDJ segment. Indices are counted from the beginning of the reported V gene (first position = 0) the V index is the first V gene nucleotide of the CDR3. The N1 index is the first inserted nucleotide between the V gene and the D gene. The N2 index is the location of the first inserted nucleotide between the D gene and the J gene. The J index is the first nucleotide of the J gene.

Index definition continues on next page...

For Research Use Only. Not for use in diagnostic procedures.

In the .tsv file the following regions are distinguished (per the figure below):

- V (variable) gene
- Start of CDR3 in V gene
- N1 junction (V and D)
- D (diversity) gene
- N2 junction (D and J)
- J (joining) gene



Productive

Rearrangements and their corresponding templates and reads, are designated “productive” if the nucleotide sequence is capable of producing a functional peptide e.g. the sequences is in-frame relative to the conserved cysteine and phenylalanine, even if changes to the N-D-N region no longer code for these amino acids, and no stop codon is present.

Reads

Reads refer to the sequenced copies of a DNA template.*

Rearrangements

A rearrangement is a unique nucleotide sequence generated through V(D)J recombination and representative of a unique clonal lineage.

Somatic Hypermutation (SHM)

The process by which B cells introduce mutations to the variable regions of B Cell Receptor genes during affinity maturation.

Templates

Templates refer to the number of biological molecules put into the assay before PCR amplification and sequencing. Using the internal bias control of the immunoSEQ Assay, we can calculate the original number of templates of sample material sequenced.

***NOTE:** Because sequencing reads are a function of PCR efficiency and sequencer loading density, total counts do not correspond to a quantity of data. Current versions of the immunoSEQ Assay (counting method = v3 or higher) present templates for quantitating rearrangements. Reads were used in calculating frequency of a rearrangement in older versions (counting method = v1 and v2) of the immunoSEQ Assay.

SAMPLE-LEVEL .TSV FILE COLUMNS

To export sample data referring to a specific list of samples:

1. Select the sample checkboxes that you wish to export meta-data about and select ‘Analysis/Sample Overview’
2. Select the **Export** drop-down menu from the top right and select ‘Export view to file’ to download your .tsv file.

The exported file contains the following columns:

.tsv Column Tag	Analyzer Name	Description
sample_name	Sample Name	Sample Name is the user-defined label assigned to a sample.

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
total_templates	Total Templates (Sum)	Total Templates (Sum) is the sum of templates for all (productive and non-productive) rearrangements in the sample.
total_reads <i>(only v1 and v2 samples)</i>	Total Reads	Total Reads is the sum of post-amplification sequencing reads for all (productive and non-productive) rearrangements in the sample.
productive_templates	Total Productive Templates (Sum)	Total Productive Templates (Sum) is the sum of templates for all Productive Rearrangements in the sample.
total_productive_reads <i>(only v1 and v2 samples)</i>	Total Productive Reads	Total Productive Reads is the sum of post-amplification sequencing reads for all Productive Rearrangements in the sample.
fraction_productive	Fraction Productive	Fraction Productive is the fraction of the Total Templates which are productive. It is obtained by dividing Total Productive Templates by the number of Total Templates. For samples sequenced with the v1/v2 counting method this metric is calculated using Total Productive Reads and Total Reads.
total_rearrangements	Rearrangements	Total Rearrangements is the sum of unique (productive and non-productive) rearrangements in the sample.
productive_rearrangements	Productive Rearrangements	Productive Rearrangements is the sum of unique productive rearrangements in the sample.
productive_clonality	Productive Clonality	Productive Clonality is calculated using only the productive rearrangements in the sample.
max_productive_frequency	Max Productive Frequency	Max Productive Frequency is the highest Frequency value for a productive rearrangement in the sample.
Locus	Locus	Locus shows which chromosomal position was targeted for sequencing. In immunoSEQ Assays, two common examples of loci are TCRB and IGH.
sample_tags	Sample Tags	Sample Tags are metadata associated with the sample. Sample Tags are used to define sample groups for downstream analysis.
release_date	Release Date	Release Date is the data results were made available in the immunoSEQ Analyzer.
outofframe_rearrangements	Out of Frame Rearrangements	Out of Frame Rearrangements is the count of unique rearrangements in a sample where codons are out of frame with respect to the 5' cysteine and terminal phenylalanine.
stop_templates	Total Stop Templates (Sum)	Total Stop Templates (Sum) is the sum of all templates for all rearrangements, containing a stop codon.
dj_rearrangements <i>(IGH assay only)</i>	DJ Rearrangements	DJ Rearrangements is the number of IgH DJ rearrangements. Filtered by default but may be included by removing the project-level filter.
outofframe_templates	Total Out of Frame Templates (Sum)	Total Out of Frame Templates (Sum) is the count of all templates in a sample's Out of Frame Rearrangements.
stop_rearrangements	Stop Rearrangements	Stop Rearrangements is the count of unique rearrangements in a sample containing a stop codon.

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
dj_templates	Total DJ Templates (Sum)	Total DJ Templates (Sum) is the sum of all templates for all DJ rearrangements.
total_outofframe_reads	Total Out of Frame Reads	Total Out of Frame Reads is the sum of all reads for all Out of Frame Rearrangements.
total_stop_reads <i>(only v1 and v2 samples)</i>	Total Stop Reads	Total Stop Reads is the sum of all reads for all rearrangements containing a stop codon.
total_dj_reads <i>(only v1 and v2 samples)</i>	Total DJ Reads	Total DJ Reads is the count of all reads in a sample's DJ Rearrangements.
sample_clonality	Sample Clonality	Sample Clonality is calculated using all (productive and non-productive) rearrangements in the sample.
sample_entropy	Sample Entropy	Sample Entropy is Shannon's entropy calculated using all (productive and non-productive) rearrangements in the sample.
productive_entropy	Productive Entropy	Productive Entropy is the Shannon's entropy calculated using only the productive rearrangements in the sample.
max_frequency	Max Frequency	Max Frequency is the highest Frequency value of any rearrangement in the sample.
sample_cells <i>(available only for v4 samples)</i>	Nucleated Cells	Nucleated Cells is the total number of nucleated cells in the sample measured by internal controls.
total_t_cells <i>(available only for v4 samples)</i>	Total T Cells	Total T Cells shows the overall sum of T cells found within a sample.
fraction_productive_of_cells <i>(available only for v4 samples)</i>	Fraction T Cells of Nucleated Cells	Fraction T Cells of Nucleated Cells is the fraction of T (or B) cells out of total nucleated cells. This value is calculated by dividing the Productive Templates by the Nucleated Cells.
total_templates_agg <i>(available only for v4 samples)</i>	Total Templates (Aggregate)	Total Templates (Aggregate) is the total of all templates (productive and non-productive). It is an improved estimate over Total Templates, which can be biased upwards due to rounding errors.
sample_cells_mass_estimate	Nucleated Cells Estimate	Nucleated Cells Estimate is an estimate of the total number of nucleated cells, derived from the A260 absorbance measurements of the sample. This metric is less accurate than the v4 Nucleated Cells, which is based on internal controls.
fraction_productive_of_cells_mass_estimate	Fraction T/B Cells of Nucleated Cells Estimate	Fraction T/B Cells of Nucleated Cells Estimate is the estimated fraction of T (or B) cells out of total nucleated cells, based on the total nucleated cells estimated derived from the A260 absorbance. This value is calculated by dividing Productive Templates by Nucleated Cells Estimate.
sample_amount_ng (not reported if the concentration below 10ng/ul)	Sample Amount (ng)	Sample Amount (ng) is the mass of sample DNA (in nanograms) run in the assay, derived from the A260 absorbance.
Species	Species	Species is the target species (i.e. human or mouse) of the assay version used.

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
product_subtype	Product	Product is a designation of the assay depth at which the sample was run. If the immunoSEQ service was used, the following resolution labels may be reported: Survey, Deep, Ultra Deep, or Max Depth.
counting_method	Counting Method	Counting Method identifies which assay version was used to quantify the rearrangements. Current versions of the immunoSEQ Assay (counting method v3 or higher) display templates. Sequencing reads are provided for older assay versions (counting method v1 and v2, see Terminology/Definitions section above).
primer_set	Primer Set	Primer Set is a label that identifies the novel assay chemistry used to amplify immune cell templates.
order_name	Order Name	Order Name shows which order the sample is associated with.
kit_pool <i>(only samples run using the immunoSEQ Kit)</i>	Kit Pool	Kit Pool is the code corresponding to the ID and run number of a kit. If a sample has been run multiple times, only the most recently published results are displayed.
Sku	Test SKU	Test SKU (Stock Keeping Unit) is Adaptive's internal identification code for the type of immunosequencing assay performed.
test_name	Test Name	Test Name identifies which immunosequencing assay was performed on the sample. Each name is accompanied by a matching Test SKU code.
upload_date	(Not shown)	The date results were uploaded to the immunoSEQ Analyzer.

REARRANGEMENT-LEVEL .TSV COLUMNS

Rearrangement information is found by exporting individual sample data. Select the box(es) by your desired samples(s) on the immunoSEQ Analyzer "Project Overview" or "Sample Overview" then click on the "Export" button to download the associated .tsv file(s). The .tsv file contains the following columns:

.tsv Column Tag	Analyzer Name	Description
sample_name	Sample Name	Sample Name is the user-defined label assigned to a sample.
productive_frequency <i>(only for productive rearrangements)</i>	Productive Frequency	Productive Frequency is the frequency of the productive rearrangement among all Productive Rearrangements within the sample. This is calculated as the Templates number divided by the Total Productive Templates (for v1 and v2 samples, frequency is calculated using Reads and Total Productive Reads).
Frequency	Frequency	Frequency is the frequency of the rearrangement among all Rearrangements within the sample. This is calculated as the Templates number divided by the Total Templates (for v1 and v2 samples, frequency is calculated using Reads and Total Reads).

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
templates (<i>assay counting method v2 or higher</i>)	Templates	Templates is the number of inferred input molecules corresponding to a specific rearrangement.
Reads	Reads	Reads is the number of post-amplification sequencing reads for a specific rearrangement.
amino_acid (<i>productive rearrangements only</i>)	Amino Acid	Amino Acid is the translation of the nucleotide sequence of the CDR3 region to a peptide sequence.
Rearrangement	Rearrangement	Rearrangement is a unique nucleotide sequence detected by the assay.
frame_type	Frame Type	Frame Type is the status of the reading frame of the CDR3 sequence relative to the 5' canonical cysteine. Values are "In" (in-frame), "Out" (out-of-frame), or "Stop" (containing a stop codon).
cdr3_length	CDR3 Length	CDR3 Length is the total number of nucleotides contained in the CDR3 region of the rearrangement. This is measured from the first base of the codon for the 5' canonical cysteine in the V gene to the last base of the codon for the conserved Phenylalanine in the J gene.
v_resolved	V Resolved	V Resolved is the name most specific V gene family, gene, or allele identified during annotation for this rearrangement.
d_resolved	D Resolved	D Resolved is the name of the most specific D gene family, gene, or allele identified during annotation for this rearrangement.
j_resolved	J Resolved	J Resolved is the name of the most specific J gene family, gene, or allele identified during annotation for this rearrangement.
rearrangement_type	Rearrangement Type	Rearrangement Type is a list of the gene types annotated for a rearrangement (e.g. "VDJ" indicates the V, D and J gene regions were identified).
v_family	V Family	V Family is the name V family identified for this rearrangement.
v_gene	V Gene	V Gene is the name of the V gene identified for this rearrangement.
v_allele	V Allele	V Allele is the name of the V gene allele identified for this rearrangement.
d_family	D Family	D Family is the name of the D family identified for this rearrangement.
d_gene	D Gene	D Gene is the name of the D gene identified for this rearrangement.
d_allele	D Allele	D Allele is the name of the D gene allele identified for this rearrangement.
j_family	J Family	J Family is the name of the J family identified for this rearrangement.

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
j_gene	J Gene	J Gene is the name of the J gene identified for this rearrangement.
j_allele	J Allele	J Allele is the name of the J allele identified for this rearrangement.
v_deletions	V Deletions	V Deletions is the number of nucleotides deleted from the V gene during recombination.
d5_deletions	D5 Deletions	D5 Deletions is the number of nucleotides deleted from the 5' end of the D gene during recombination.
d3_deletions	D3 Deletions	D3 Deletions is the number of nucleotides deleted from the 3' end of the D gene during recombination.
j_deletions	J Deletions	J Deletions is the number of nucleotides deleted from the J gene during recombination.
n2_insertions	N2 Insertions	N2 Insertions is the number of nucleotides inserted in the N2 junction (between the J and D genes) during recombination.
n1_insertions	N1 Insertions	N1 Insertions is the number of nucleotides inserted in the N1 junction (between the V and D genes) during recombination.
v_index	V Index	V Index is the nucleotide position counting from the 5' end of the sequence beginning at "0" to the Cysteine at the start of the CDR3 in the V gene.
n1_index	N1 Index	N1 Index is the nucleotide position counting from the 5' end of the sequence beginning at "0" to the start of the N1 region (the junction of V and D).
n2_index	N2 Index	N2 Index is the nucleotide position counting from the 5' end of the sequence beginning at "0" to the start of the N2 region (the junction of D and J).
d_index	D Index	D Index is the nucleotide position counting from the 5' end of the sequence beginning at "0" to the start of the D gene.
j_index	J Index	J Index is the nucleotide position counting from the 5' end of the sequence beginning at "0" to the start of the J gene.
v_family_ties	V Family Ties	V Family Ties is a comma-separated list of possible families when there is insufficient sequence to identify a single V family.
v_gene_ties	V Gene Ties	V Gene Ties is a comma-separated list of possible genes when there is insufficient sequence to identify a single V gene.
v_allele_ties	V Allele Ties	V Alleles Ties is a comma-separated list of possible alleles when there is insufficient sequence to identify a single V alleles.
d_family_ties	D Family Ties	D Family Ties is a comma-separated list of possible families when there is insufficient sequence to identify a single D family.

For Research Use Only. Not for use in diagnostic procedures.

...table continued

.tsv Column Tag	Analyzer Name	Description
d_gene_ties	D Gene Ties	D Gene Ties is a comma-separated list of possible genes when there is insufficient sequence to identify a single D gene.
d_allele_ties	D Allele Ties	D Alleles Ties is a comma-separated list of possible alleles when there is insufficient sequence to identify a single D allele.
j_family_ties	J Family Ties	J Family Ties is a comma-separated list of possible families when there is insufficient sequence to identify a single J family.
j_gene_ties	J Gene Ties	J Gene Ties is a comma-separated list of possible genes when there is insufficient sequence to identify a single J gene.
j_allele_ties	J Allele Ties	J Alleles Ties is a comma-separated list of possible alleles when there is insufficient sequence to identify a single J allele.
v_shm_count <i>(IgH assay only)</i>	V SHM Count	V SHM Count is the total number of nucleotides in the V gene which were altered by SHM.
v_shm_indexes <i>(IgH assay only)</i>	V SHM Indexes	V SHM Indexes is a comma-separated list of nucleotide positions within the V gene sequence which were altered during somatic hypermutation.
Antibody	Flow Antibody	Flow Antibody is a standard flow cytometry antibody (derived from commercially available sources) that may identify the V gene of this rearrangement.

For Research Use Only. Not for use in diagnostic procedures.

IMMUNOSEQ, IMMUNOSEQ ANALYZER, ADAPTIVE BIOTECHNOLOGIES and associated designs are trademarks of Adaptive Biotechnologies.

Copyright © 2019 Adaptive Biotechnologies Corp. All rights reserved.
MRK-00342 | Page 8 of 8



adaptivebiotech.com