HHS Public Access

Author manuscript

*Softw Impacts*. Author manuscript; available in PMC 2022 March 14.

Published in final edited form as:

*Softw Impacts*. 2021 November ; 10: . doi:10.1016/j.simpa.2021.100142.

# Integrated Analysis Toolset for Defining and Tracking Alloreactive T-cell Clones After Human Solid Organ and Hematopoietic Stem Cell Transplantation

**Aleksandar Obradovic**[1,2], **Yufeng Shen, PhD**[2], **Megan Sykes, MD**[1,3,4], **Jianing Fu, PhD**[1,*]

[1]Columbia Center for Translational Immunology, Department of Medicine; Columbia University, New York, NY 10032, United States.

[2]Department of Systems Biology; Columbia University, New York, NY 10032, United States.

[3]Department of Surgery; Columbia University, New York, NY 10032, United States.

[4]Department of Microbiology & Immunology, Columbia University, New York, NY 10032, United States.

## Abstract

We have developed a suite of tools for integrated analysis of T-Cell-Receptor Sequencing data to define and track alloreactive T-cells in human transplant studies. This has enabled discovery of sequences and patterns of T-cell enrichment and deletion associated with clinical outcomes such as transplant rejection and tolerance. The codebase includes user-friendly default analyses with customizable parameters which greatly accelerate computational workflows and provide robust statistics comparing post-transplant specimens to pre-transplant baseline. It also includes helper functions for robust characterization of T-cell-repertoire diversity, sample-to-sample divergence, resolution of sample-of-origin ambiguity in pooled assays, and functions to output all sequences defined as alloreactive.

### Keywords

Solid Organ Transplantation; Hematopoietic Stem Cell Transplantation; TCR-Sequencing; alloreactive clones; immune tolerance

*Corresponding Author: Jianing Fu PhD, Columbia Center for Translational Immunology, Columbia University, 650 West 168th street, Black Building 1511, New York, 10032, Tel: 212.342.0237 jf2977@cumc.columbia.edu.

Authorship

Aleksandar Obradovic: Methodology, Software, Data curation, Visualization, Writing- Original draft preparation. Yufeng Shen: Conceptualization, Methodology, Software, Supervision, Writing- Reviewing and Editing. Megan Sykes: Conceptualization, Methodology, Supervision, Writing- Reviewing and Editing. Jianing Fu: Conceptualization, Investigation, Methodology, Software, Supervision, Writing- Original draft preparation, Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code metadata

| Nr | Code metadata description | Please fill in this column |
|---|---|---|
| C1 | Current code version | *v1* |
| C2 | Permanent link to code/repository used for this code version | https://github.com/Aleksobrad/TCR-analysis/blob/master/TCRanalysis.zip |
| C3 | Permanent link to reproducible capsule | https://codeocean.com/capsule/1539294/tree/v2 |
| C4 | Legal code license | *GPL-3.0* |
| C5 | Code versioning system used | *git* |
| C6 | Software code languages, tools and services used | *R* |
| C7 | Compilation requirements, operating environments and dependencies | *The code is stable up to the current R version 3.6.2* |
| C8 | If available, link to developer documentation/manual | *See* Table 1 |
| C9 | Support email for questions | azo2104@cumc.columbia.edu |

## Introduction

The purpose of this code package, implemented in the R programming language, is automated analysis of high-throughput T-Cell-Receptor Sequencing (TCR-Seq) data generated in the context of solid organ transplantation (SOT) or hematopoietic stem cell transplantation (HSCT) studies. In particular, where mixed lymphocyte reactions (MLRs) are performed from pre-transplant blood or lymphoid tissue samples, we are able to compare repertoires of baseline and antigen-stimulated T cells to define sets of host-versus-graft (HvG) or graft-versus-host (GvH) alloreactive clones as clones which increase (or decrease in the setting of tolerance) in frequency by a predefined frequency threshold following *in vivo* antigen exposure. Subsequently, we can track the identity and enrichment of these clones in follow-up biopsies and blood samples assessing different tissues or timepoints for the level of HvG or GvH immune response. This enables study of the dynamics of organ transplant rejection and tolerance, the onset of graft-versus-host disease (GVHD) after HSCT, as well as the development of novel sequencing-based biomarkers for early prediction of graft rejection and GVHD.

## Software Methodology

As input, the analysis pipeline assumes a data table of TCR clone counts aggregated such that each unique TCR clone is represented in a unique row of the data matrix, and each sample is represented as a column (see Figure 1). At minimum, columns are required for CD4 baseline unstimulated repertoire, CD4 antigen-stimulated repertoire, CD8 baseline unstimulated repertoire, CD8 antigen-stimulated repertoire, and at least one post-transplant blood or biopsy specimen (use biopsy as a demonstrative example in following text, figure and table). The data matrix format is provided as standard output from Adaptive Biotechnologies' ImmunoSeq platform, which is utilized by our lab and many other groups for TCR-Sequencing [1–8], and we therefore take this as input for downstream analysis.

Several major principles are applied to define biologically relevant alloreactive TCR clones, including 1) all sequences come from FACS-sorted CD4 and CD8 $CFSE^{low}$ populations to reflect their alloreactivity; 2) all sequences meet the ≥ 2–5 fold expansion criteria when comparing frequencies in $CFSE^{low}$ population with that in unstimulated populations to exclude dominant resting clones bleeding into the $CFSE^{low}$ sorted population likely mainly caused by sorting error, and to a lesser extent of bystander proliferation; 3) a minimal frequency cut-off for sequences in $CFSE^{low}$ population to avoid background noise; 4) resolve ambiguity between CD4 and CD8 T cells due to sorting errors and between donor and recipient sequences due to biological sharing and/or sorting imperfectness.

Recommended data pre-processing steps include setting rownames() of the data matrix as TCR-clone character strings (nucleotide or amino acid sequences), and defining separate CD4 and CD8 input objects for each comparison of biopsy versus baseline, as shown in Figure 1 and described in the code comments at https://github.com/Aleksobrad/TCR-analysis/blob/master/TCRanalysis.zip. The default run() function resolves ambiguities between sorted CD4 and CD8 T cells and defines alloreactive clones with an adjustable threshold of fold-expansion and minimum observed frequency, performing up to three statistical tests of alloreactive clone dynamics in pre-transplant baseline versus biopsy (Relative Expansion Analysis, Turnover Analysis, and Absolute Expansion Analysis, see Figure 1 and Table 1). To enable customizable workflows we also include a suite of helper functions to normalize clone counts to frequencies, resolve flow-sorting ambiguities, output a list of alloreactive clone TCR-sequences, and quantify repertoire diversity for any arbitrary vector of TCR clone counts. We also include functions for visualization of repertoire diversity by abundance plot as well as quantification of pairwise repertoire overlap in any number of samples. It is recommended to use abundance plot to visualize clonal frequency distribution of each sample before calculating diversity to get insights on whether the decreased diversity is due to a few dominant clones or overall clonal expansion as discussed in our previous study [3]. A full list of functions and their customizable parameters are detailed in Table 1.

## Impact Overview

Our platform combines functional assay (MLR) with high throughput TCR beta chain complementary determining region 3 (CDR3) DNA sequencing to define and track alloreactive clones across time and space after human transplantation [1–6], which serves as a great tool to study graft rejection after SOT [1–6] and GVHD after HSCT as discussed in published studies [9, 10] and ongoing work through personal communication with Dr. Ran Reshef at Columbia University (https://cdmrp.army.mil/search.aspx. Proposal Number: CA171008). We were able to identify biologically relevant TCR clones to correlate with patients' clinical outcomes. This method allows us to distinguish clonal deletion versus anergy, providing a deeper understanding of immune tolerance after transplantation.

Our R script is able to efficiently process large scale datasets and has been successfully applied for all kinds of alloreactivity studies after multiple types of SOT. The basic code workflow is intuitive and quickly picked up by non-computational research staff in the lab, while providing sufficient modularity to be easily customized by more experienced users.

It is worth noting that our R scripts' default parameters are tailored based on specific scientific questions and particular sequencing platforms but the suite of helper functions in this code package allows for easy customization and flexibility of workflows even for computationally inexperienced users. For instance, resolving ambiguities between donor and recipient TCR clones becomes important when studying bidirectional alloresponses after intestinal transplantation or lung transplantation, a problem easily fixed by pre-processing data with the resolveambiguous() function provided in the code pipeline. Of note, the frequency ratio used to resolve clonal ambiguity should be customized based on the characteristics of a user's particular dataset with respect to sample sizes and degree of sorting error. For greater sorting error or greater discrepancies in sample size, the ratio could be increased. In the extreme case, if the number of T-cells in one sample is several orders of magnitude smaller than the sample it is being compared to, the difference in minimum observable clone frequency between the two samples would far exceed our otherwise recommended cutoff ratio of 2–5 fold, and clones with reasonably high template counts would be defined as ambiguous and be overwritten by 0. In such cases we recommend resolution of clonal ambiguity by manual inspection of the data. The default run() function is user-friendly and enables very quick and powerful comparison of multiple samples across patients and across tissue sites in an automated way, reducing our previous workflow times from several days of manually manipulating excel spreadsheets to a few minutes of executing the code for each sample in a given experiment. Given the development of commercially available sequencing platforms, such as Adaptive Biotechnologies, which have gone through several generations of development, the data output changed from read counts to template counts. Therefore, we have adjusted the minimal frequency cut-off values based on power analysis reflecting the difference in technologies, taking advantage of the ability to adjust parameters for the run() and listAlloreactive() functions without having to re-write the code. For example, minimal frequency threshold to define alloreactive clone sequences from $CFSE^{low}$ samples is set at 0.001% when using read counts and 0.002% when using template counts, which serves to ensure 85% repeatability, as determined by power analysis.

## Application and Summary

Our method and R script [11] have been extensively used by us [1–6] and others [7, 8] in multiple types of human SOT studies, including intestine [2, 5, 6], liver [4], heart [7], kidney [8] and combined kidney and bone marrow transplants [1]. Our platform allows the integration of T cell clonotype, alloreactivity and RNA profiling on the single cell level to relate T cell functional profiles with clonotypes of individual T cells [6]. Additionally, by combining both pre- and post-transplant MLRs with single cell RNA+TCR sequencing, we are now able to run this pipeline on separate RNA-Seq phenotypes of TCR clones, defining multiple functional categories of HvG clones, including persistent, tolerant, *de novo* generated, acquired and missing HvG clones, to address questions of immune tolerance locally and systemically [12, 13]. Our unique platform and R script also help to develop treatment strategies to induce long persistent mixed chimerism after intestinal transplantation and induce tolerance. Based on our recent studies after human intestinal transplantation [5, 6], a pilot clinical trial to combine multivisceral transplantation with

donor $CD34^+$ cell infusion at the time of peak lymphohematopoietic GvH responses in circulation to promote long term multilineage chimerism is about to initiate at Columbia University Irving Medical Center. Furthermore, our method could help to develop alternative monitoring strategies to minimize or entirely avoid invasive and delayed diagnostic procedures after human kidney transplantation [8]. A list of scholarly publications enabled by this software is presented in the references [1–8], along with ongoing research projects [12–14].

## Acknowledgements

This work was supported in part by National Institutes of Health (NIH), National Institute of Allergy and Infectious Diseases (NIAID) grant #P01 AI106697 (to M. Sykes and Y. Shen), #UM1 AI109565 (to M. Sykes) and #R21 TR002279 (to M. Sykes) and by and American Society of Transplantation Research Network Directed Research Grant #gAST182D0MS (to M. Sykes). J. Fu was supported by a Congressionally Directed Medical Research Program (CDMRP) Discovery Award W81XWH-20-1-0159, funded by the Department of Defense (DoD), and a Nelson Faculty Development Award UR011630-01, funded by the Nelson Family Transplantation Innovation Award Program at Columbia University Irving Medical Center.

## References

1. Morris H, DeWolf S, Robins H, Sprangers B, LoCascio SA, Shonts BA, et al. Tracking donor-reactive T cells: Evidence for clonal deletion in tolerant kidney transplant patients. Sci Transl Med. 2015;7(272):272ra10. Epub 2015/01/30. doi: 10.1126/scitranslmed.3010760.
2. Zuber J, Shonts B, Lau SP, Obradovic A, Fu J, Yang S, et al. Bidirectional intragraft alloreactivity drives the repopulation of human intestinal allografts and correlates with clinical outcome. Science immunology. 2016;1(4):eaah3732. [PubMed: 28239678]
3. DeWolf S, Grinshpun B, Savage T, Lau SP, Obradovic A, Shonts B, et al. Quantifying size and diversity of the human T cell alloresponse. JCI Insight. 2018;3(15). Epub 2018/08/10. doi: 10.1172/jci.insight.121256.
4. Savage TM, Shonts BA, Lau S, Obradovic A, Robins H, Shaked A, et al. Deletion of donor-reactive T cell clones after human liver transplant. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2020;20(2):538–45. Epub 2019/09/12. doi: 10.1111/ajt.15592.
5. Fu J, Zuber J, Martinez M, Shonts B, Obradovic A, Wang H, et al. Human Intestinal Allografts Contain Functional Hematopoietic Stem and Progenitor Cells that Are Maintained by a Circulating Pool. Cell Stem Cell. 2019;24(2):227–39. doi: 10.1016/j.stem.2018.11.007. [PubMed: 30503142]
6. Fu J, Zuber J, Shonts B, Obradovic A, Wang Z, Frangaj K, et al. Lymphohematopoietic graft-versus-host responses promote mixed chimerism in patients receiving intestinal transplantation. J Clin Invest. 2021;131(8). Epub 2021/02/26. doi: 10.1172/jci141698.
7. Habal MV, Miller AMI, Rao S, Lin S, Obradovic A, Khosravi-Maharlooei M, et al. T cell repertoire analysis suggests a prominent bystander response in human cardiac allograft vasculopathy. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2021;21(4):1465–76. Epub 2020/10/07. doi: 10.1111/ajt.16333.
8. Aschauer C, Jelencsics K, Hu K, Heinzel A, Vetter J, Fraunhofer T, et al. Next generation sequencing based assessment of the alloreactive T cell receptor repertoire in kidney transplant patients during rejection: a prospective cohort study. BMC Nephrol. 2019;20(1):346. Epub 2019/09/04. doi: 10.1186/s12882-019-1541-5. [PubMed: 31477052]
9. Shah O, Tamaresis JS, Kenyon LJ, Xu L, Zheng P, Gupta P, et al. Analysis of the Whole CDR3 T Cell Receptor Repertoire after Hematopoietic Stem Cell Transplantation in 2 Clinical Cohorts. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2020;26(6):1050–70. Epub 2020/02/23. doi: 10.1016/j.bbmt.2020.01.020.

10. Bettens F, Calderin Sollet Z, Buhler S, Villard J. CD8+ T-Cell Repertoire in Human Leukocyte Antigen Class I-Mismatched Alloreactive Immune Response. Frontiers in Immunology. 2021;11(3440). doi: 10.3389/fimmu.2020.588741.

11. Obradovic A, Shen Y, Sykes M, Fu J. Integrated Analysis Toolset for Defining and Tracking Alloreactive T-cell Clones After Human Solid Organ and Hematopoietic Stem Cell Transplantation [Source Code]. 2021. doi: 10.24433/CO.8853858.v2.

12. Fu J, Wang Z, Martinez M, Frangaj K, Jones R, Guo XV, et al. Single cell immune profiling of human intestinal allografts reveals heterogeneity and alloreactivity of resident memory T cells in association with graft outcomes. Transplantation. 2020;104(S3):S72–S3. doi: 10.1097/01.tp.0000698616.71945.ca.

13. Fu J, Wang Z, Martinez M, Frangaj K, Jones R, Guo X, et al. O-40: Single-Cell Immune Profiling of Human Intestinal Allografts Reveals Differential Contributions of HvG T-cell Clones in Quiescent vs Chronically Rejecting Allografts. Transplantation. 2021;105(7S).

14. J W, M M, Z J, O A, J R, W E, et al. PE-2: Dynamic Reconstitution of Recipient Resident Memory T Cell Repertoire after Human Intestinal Transplantation. Transplantation. 2021;105(7S).

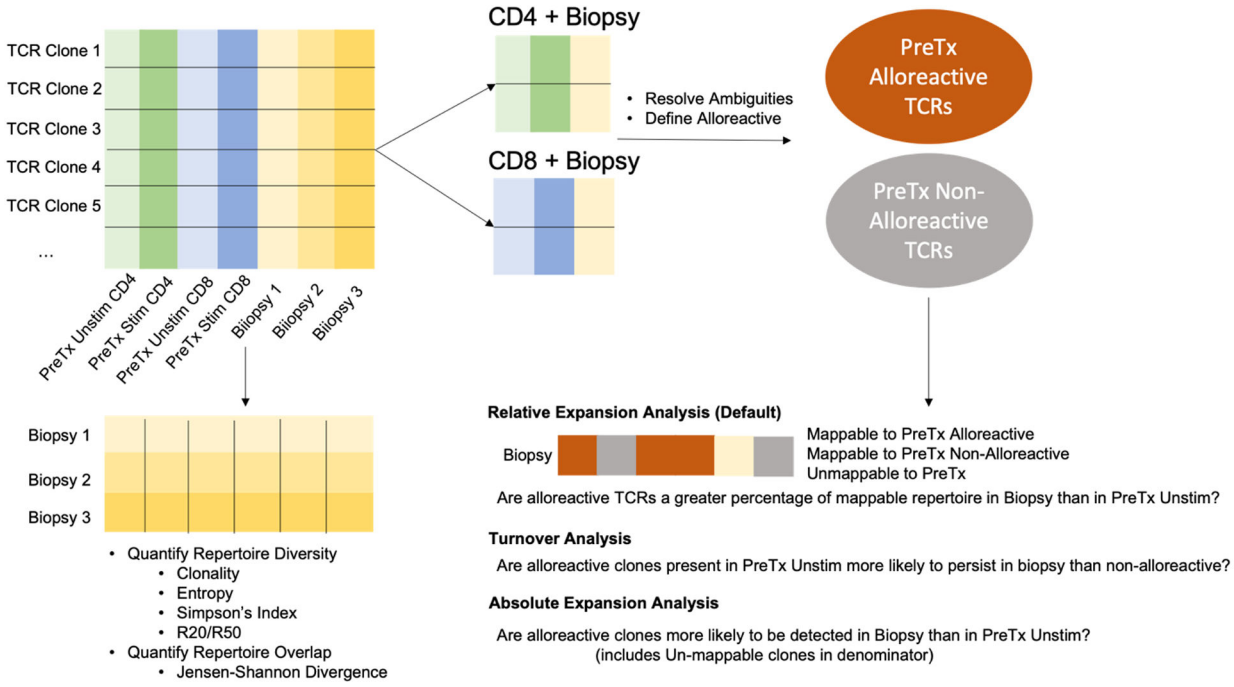


**Figure 1: Conceptual Analysis Workflow.**

PreTx: pre-transplant; Unstim: unstimulated; Stim: stimulated $CFSE^{low}$ population.

**Table 1:**

List of Functions and short description

<table>
<tr><th>Function</th><th>Description</th></tr>
<tr><td rowspan="4">run(cd4, cd8, fold=5, freq=0.00001, clonality=F, turnover =F, absoluteexpansion=F, ambiguityRatio=5, filename=”Default”)</td><td>By default, prints summary data tables to output file comparing pre-transplant baseline to a specified biopsy sample by total number of unique clones and numbers of mappable alloreactive vs non-alloreactive clones in each. Default analysis mode compares alloreactivity rate (by unique clone fraction and cumulative frequency) among mappable clones in biopsy vs baseline. This is done by applying Fisher’s exact test to the following contingency table
<table>
<tr><td>preTx alloreactive</td><td>total preTx – preTx alloreactive</td></tr>
<tr><td>biopsy alloreactive</td><td>Biopsy mappable as alloreactive or non-alloreactive – biopsy alloreactive</td></tr>
</table></td></tr>
<tr><td>Additional analyses included but not activated by default are turnover, absolute expansion, and clonality, see below.<br>cd4 (required): combined data frame of TCR clone counts with columns (in order) representing CD4 pre-transplant unstimulated, CD4 pre-transplant antigen-stimulated, and biopsy<br>cd8 (required): combined data frame of TCR clone counts with columns (in order) representing CD8 pre-transplant unstimulated, CD8 pre-transplant antigen-stimulated, and biopsy<br>fold: Minimum fold-expansion required from unstimulated to antigen-stimulated repertoire to define a TCR-clone as alloreactive. Set to 5 by default.<br>freq: Minimum frequency required to define a clone as alloreactive if detected in antigen-stimulated but not antigen-unstimulated repertoire (fold-expansion = Infinity in this case). Should represent fold * minimum detection threshold for the sequencing experiment.<br>clonality: When set to TRUE, outputs clonality of entire repertoire and of alloreactive clone subset in pre-transplant baseline and in biopsy.<br>turnover: When set to TRUE, performs statistical test of whether pre-transplant alloreactive clones are more likely to persist to biopsy timepoint than non-alloreactives. This is done by applying Fisher’s exact test to the following contingency table
<table>
<tr><td>nonAllo clones detected in preTx unstim and in biopsy</td><td>nonAllo clones detected in preTx unstim but not in biopsy</td></tr>
<tr><td>Allo clones detected in preTx unstim and in biopsy</td><td>Allo clones detected in preTx unstim but not in biopsy</td></tr>
</table></td></tr>
<tr><td>absoluteexpansion: When set to TRUE, performs statistical test of whether alloreactive clones are more likely to be detected in general in pre-transplant unstimulated sample or in biopsy. This is done by applying Fisher’s exact test to the following contingency table
<table>
<tr><td>Allo clones in preTx unstim</td><td>Total Allo clones defined from preTx stim – Allo clones in preTx unstim</td></tr>
<tr><td>Allo clones in biopsy</td><td>Total Allo clones defined from preTx stim – Allo clones in biopsy</td></tr>
</table></td></tr>
<tr><td>ambiguityRatio: ratio of clone frequencies used resolve CD4 vs CD8 sorting error. Clones present in both CD4 and CD8 with a frequency > ambiguityRatio higher in one population will be removed from the other. Clones with frequency difference less than ambiguityRatio will be removed from both.<br>filename: name of file to which output tables will be appended.</td></tr>
<tr><td>normalize(data)</td><td>Scales each column of input data matrix from raw TCR clone counts to normalized clone frequencies, summing to 1.<br>data: data frame of clone counts where rows are unique TCR sequenced and columns are samples</td></tr>
</table>

| Function | Description |
|---|---|
| cloneCal(x) | Computes Clonality (entropy divided by maximum possible entropy) for a given sample's TCR-repertoire. Ranges from 0 to 1 where higher scores indicate greater clonal expansion and lower repertoire diversity.<br>x: Vector of clone counts or normalized frequencies for a given sample |
| entropyCal(x) | Computes Entropy for a given sample's TCR-repertoire. Lowest possible score is 0, with upper bound dependent on number of unique clones.<br>x: Vector of clone counts or normalized frequencies for a given sample |
| simpsonCal(x) | Computes Simpson's Index for a given sample's TCR-repertoire. Sum of squares applied to clone frequencies, where higher scores indicate lower diversity and dominant clones are more heavily-weighted.<br>x: Vector of clone counts or normalized frequencies for a given sample |
| getR20(x,rval=0.2) | Computes R20 for a given sample's TCR-repertoire. This represents the minimum clone fraction required to capture 20% of overall clone frequency. R20 < 0.2 indicates non-uniform clone frequency, with extremely low values representing presence of a small number of highly dominant high-frequency clones.<br>x: Vector of clone counts or normalized frequencies for a given sample<br>rval: 0.2 by default can be adjusted to represent any overall clone frequency, enabling computation of R50, R10, and R5 scores, etc. |
| resolveambiguous(data,c1,c2,ratio=2) | For a given clone count data frame, returns a corrected-counts data frame removing counts of clones ambiguously present in two columns with should be mutually exclusive (e.g. CD4 and CD8) and removing contaminating counts resulting from flow-sorting error.<br>data: data frame of clone counts where rows are unique TCR sequenced and columns are samples<br>c1: column index of data to be disambiguated<br>c2: column index of data to be disambiguated<br>ratio: frequency fold-change ratio used to resolve ambiguous clones. Clones present in both c1 and c2 with a frequency > ratio higher in one population will be removed from the other. Clones with frequency difference less than ratio will be removed from both |
| listAlloreactive(cd4, cd8, fold=5, freq1=0.00001, ambiguityRatio=5) | Identifies alloreactive clones in the same way as the main run() function, and outputs CD4 and CD8 alloreactive clone TCR sequences as a list, indexed such that list item 1 is the set of all CD4 alloreactive clone sequences and list item 2 is the set of all CD8 alloreactive clone sequences.<br>cd4 (required): combined data frame of TCR clone counts with columns (in order) representing CD4 pre-transplant unstimulated and CD4 pre-transplant antigen-stimulated<br>cd8 (required): combined data frame of TCR clone counts with columns (in order) representing CD8 pre-transplant unstimulated and CD8 pre-transplant antigen-stimulated.<br>fold: Minimum fold-expansion required from unstimulated to antigen-stimulated repertoire to define a TCR-clone as alloreactive. Set to 5 by default.<br>freq: Minimum frequency required to define a clone as alloreactive if detected in antigen-stimulated but not antigen-unstimulated repertoire (fold-expansion = Infinity in this case). Should represent fold * minimum detection threshold for the sequencing experiment.<br>ambiguityRatio: ratio of clone frequencies used resolve CD4 vs CD8 sorting error. Clones present in both CD4 and CD8 with a frequency > ambiguityRatio higher in one population will be removed from the other. Clones with frequency difference less than ambiguityRatio will be removed from both. |
| abundancePlot(data) | For a given input of normalized counts matrices, creates a graphical abundance plot visualizing repertoire diversity, in a different color for each column of "data". Log-Log Plot of the number of TCR clones represented at each unique TCR frequency, where steeper slopes indicate more clonally diverse TCR repertoires.<br>data: data frame of normalized clone frequencies where rows are unique TCR sequenced and columns are samples. From a starting counts matrix, pre-process with normalize() |
| jensen_shannon(x,y) | Computes the Jensen-Shannon Divergence (JSD) of any two clone frequency vectors. This is a measure of repertoire similarity which accounts for the frequencies of shared clones (i.e. overlap in high-frequency clones is more heavily weighted than overlap in low-frequency clones). Computed as the sum of entropy(x) + entropy(y) divided by entropy of the summed vector x+y. Scales from 0 to 1 where 0 indicates identical repertoires with identical clone frequencies and 1 indicates no shared clones.<br>x: Vector of clone counts or normalized frequencies for a given sample<br>y: Vector of clone counts or normalized frequencies for a given sample |
| jsdThresholded(data) | For each pairwise combination of columns in data representing distinct TCR-repertoire samples, computes jensen_shannon divergence score. Outputs the resulting table. Repeats this process separately for all clones, the set of the top 5000 clones in each sample with the highest frequency, the set of the top 1000 clones, the set of the top 500 clones, and the set of the top 100 clones.<br>data: data frame of clone counts or frequencies where rows are unique TCR sequenced and columns are samples |