scRNA-seq Analysis Workshop
VDJ session

@ Kirby Institute, UNSW Sydney

August 31, 2023

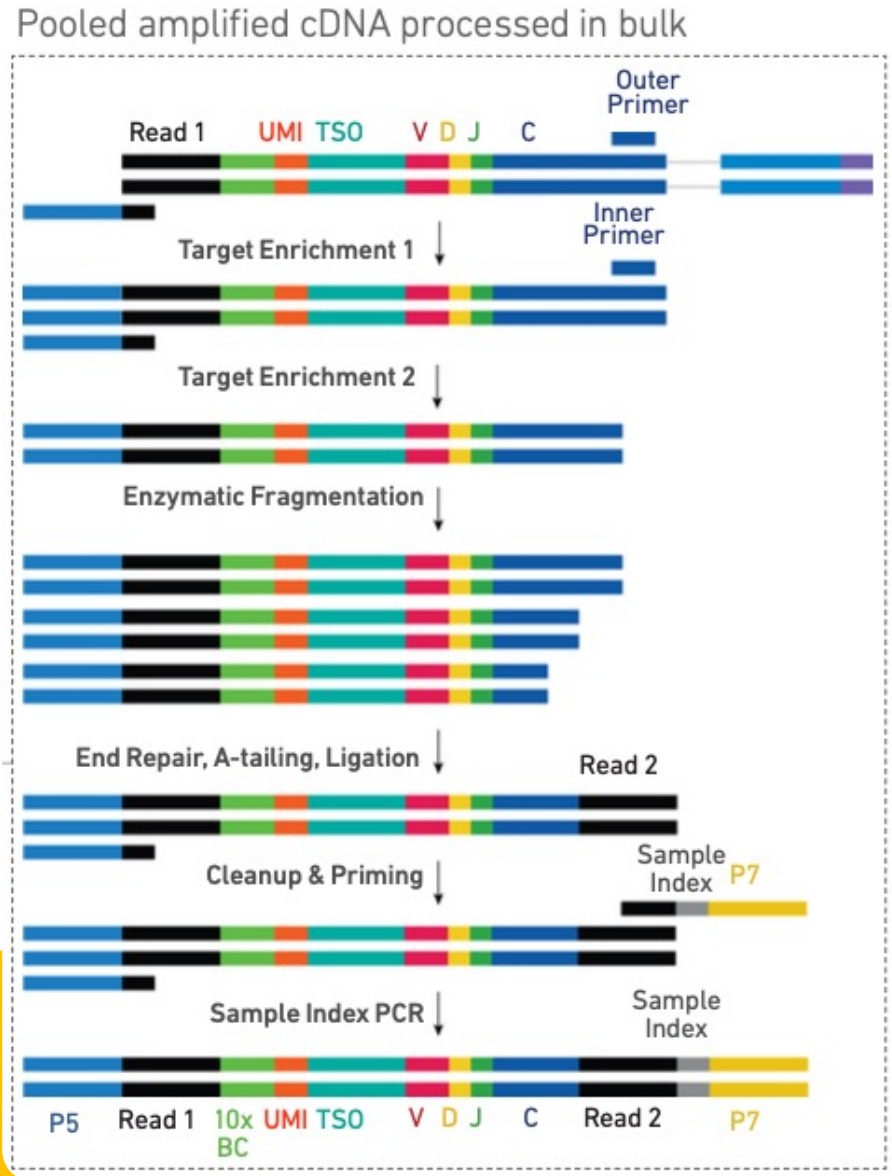
Dr Katherine Jackson

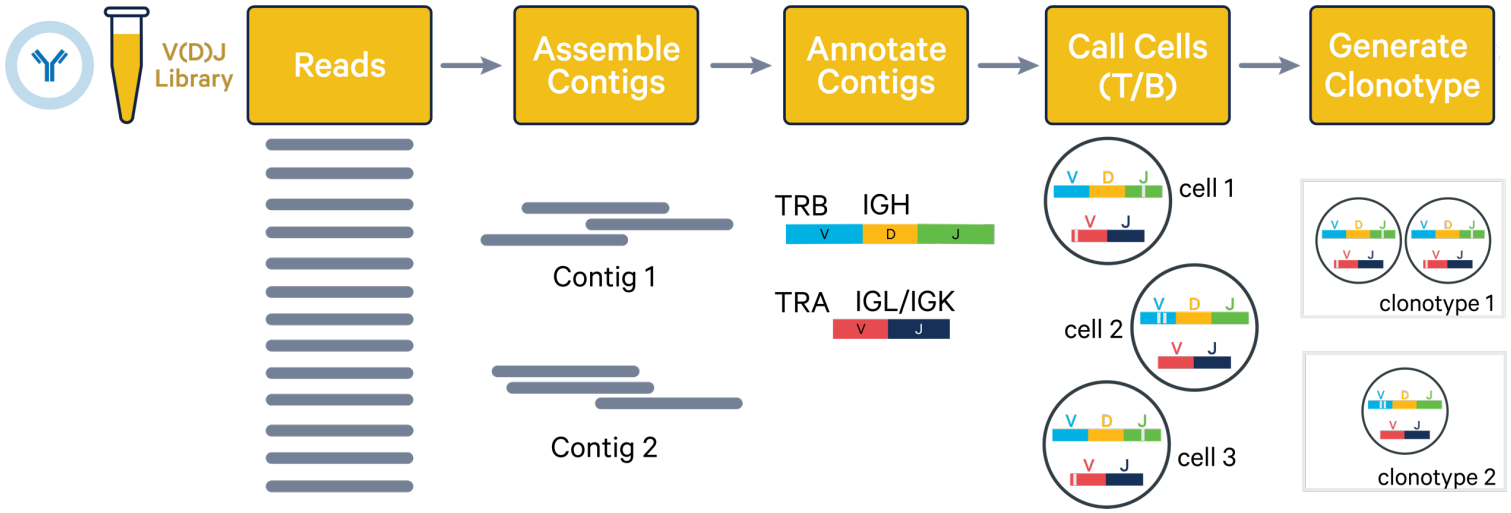
Garvan Institute of Medical Research

scVDJ-seq session outline

- Overview of 10x scVDJ-seq
- Post-processing outputs from `cellranger vdj` or `cellranger multi`
- Working with 10x VDJ data in R
  - Integrating scRNA-seq and scVDJ-seq in Seurat
  - Annotating T cells with epitope specificity
  - Using VDJ features for DGE in Seurat

Generating VDJ amplicons with 10x 5` VDJ





https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/algorithms/overview

10x Document CG000207 Rev G

Generating VDJ amplicons with 10x 5` VDJ

Running vdj

After determining your input arguments and options, run `cellranger vdj`:

```
$ cd /home/jdoe/runs
$ cellranger vdj --id=sample345 \
                 --reference=/opt/refdata-cellranger-vdj-GRCh38-alts-ensembl-7.1.0 \
                 --fastqs=/home/jdoe/runs/HAWT7ADXX/outs/fastq_path \
                 --sample=mysample \
                 --localcores=8 \
                 --localmem=64
```

https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/using/vdj

Generating VDJ amplicons with 10x 5` VDJ

```
outs
├── airr_rearrangement.tsv
├── all_contig_annotations.bed
├── all_contig_annotations.csv
├── all_contig_annotations.json
├── all_contig.bam
├── all_contig.bam.bai
├── all_contig.fasta
├── all_contig.fasta.fai
├── all_contig.fastq
├── cell_barcodes.json
├── clonotypes.csv
├── concat_ref.bam
├── concat_ref.bam.bai
├── concat_ref.fasta
├── concat_ref.fasta.fai
├── consensus_annotations.csv
├── consensus.bam
├── consensus.bam.bai
├── consensus.fasta
├── consensus.fasta.fai
├── filtered_contig_annotations.csv
├── filtered_contig.fasta
├── filtered_contig.fastq
├── metrics_summary.csv
├── vdj_contig_info.pb
├── vdj_reference
├── vloupe.vloupe
└── web_summary.html
```

```
barcode,is_cell,contig_id,high_confidence,length,chain,v_gene,d_gene,j_gene,c_gene,full_length,productive,cdr3,cdr3_nt
AAACCTGTCATATCGG-1,true,AAACCTGTCATATCGG-1_contig_1,true,556,IGK,IGKV1-8,,IGKJ4,IGKC,true,true,CQQYDELPVTF,TGTCAACAATA
AAACCTGTCCGTTGTC-1,true,AAACCTGTCCGTTGTC-1_contig_1,true,551,IGK,IGKV1-8,,IGKJ1,IGKC,true,true,CQQYYSYPRTF,TGTCAACAGTA
AAACCTGTCCGTTGTC-1,true,AAACCTGTCCGTTGTC-1_contig_2,true,565,IGH,IGHV1-69D,IGHD3-22,IGHJ3,IGHM,true,true,CATTYYYDSSGYY
AAACCTGTCGAGAACG-1,true,AAACCTGTCGAGAACG-1_contig_1,true,642,IGL,IGLV5-45,,IGLJ3,IGLC3,true,true,CMIWHSSAWVV,TGTATGATT
AAACCTGTCGAGAACG-1,true,AAACCTGTCGAGAACG-1_contig_2,true,550,IGH,IGHV1-2,,IGHJ3,IGHM,true,true,CAREIEGDGVFEIW,TGTGCGAC
AAACCTGTCTTGAGAC-1,true,AAACCTGTCTTGAGAC-1_contig_1,true,551,IGK,IGKV1D-8,,IGKJ2,IGKC,true,true,CQQYYSFPYTF,TGTCAACAGT
AAACCTGTCTTGAGAC-1,true,AAACCTGTCTTGAGAC-1_contig_2,true,557,IGH,IGHV5-51,,IGHJ3,IGHM,true,true,CARHIRGNRFGNDAFDIW,TGT
AAACGGGAGCGACGTA-1,true,AAACGGGAGCGACGTA-1_contig_1,true,534,IGH,IGHV4-59,,IGHJ3,IGHM,true,true,CARVGYRAAAGTDAFDIW,TGT
AAACGGGAGCGACGTA-1,true,AAACGGGAGCGACGTA-1_contig_2,true,633,IGL,IGLV3-19,,IGLJ2,IGLC2,true,true,CNSRDSSGNHVVF,TGTAACT
AAACGGGCACTGTTAG-1,true,AAACGGGCACTGTTAG-1_contig_1,true,633,IGL,IGLV3-21,,IGLJ2,IGLC2,true,true,CQVWDSSSDHVVF,TGTCAGG
AAACGGGCACTGTTAG-1,true,AAACGGGCACTGTTAG-1_contig_2,true,537,IGH,IGHV4-39,,IGHJ3,IGHM,true,true,CARRLITMIEGGAFDIW,TGTG
AAAGATGAGGATGCGT-1,true,AAAGATGAGGATGCGT-1_contig_1,true,578,IGH,IGHV3-33,,IGHJ4,IGHM,true,true,CAKVMIEHPSNRGHFDYW,TGT
AAAGATGAGGATGCGT-1,true,AAAGATGAGGATGCGT-1_contig_2,true,568,IGK,IGKV2-28,,IGKJ2,IGKC,true,true,CMQALQTPYTF,TGCATGCAAG
AAAGATGGTACTTCTT-1,true,AAAGATGGTACTTCTT-1_contig_1,true,566,IGH,IGHV2-5,IGHD3-22,IGHJ3,IGHM,true,true,CVNNKGNYYDSSRYG
AAAGATGGTACTTCTT-1,true,AAAGATGGTACTTCTT-1_contig_2,true,551,IGK,IGKV1-5,,IGKJ3,IGKC,true,true,CQQYNSYSQTF,TGCCAACAGTA
AAAGATGGTCGAATCT-1,true,AAAGATGGTCGAATCT-1_contig_1,true,685,IGL,IGLV6-57,,IGLJ2,IGLC2,true,true,CQSYDSSNVVF,TGTCAGTCT
AAAGATGGTCGAATCT-1,true,AAAGATGGTCGAATCT-1_contig_2,true,590,IGH,IGHV3-15,,IGHJ4,IGHM,true,true,CTTDDEKRPYSGSYLPFDYW,T
AAAGATGGTGAGGGAG-1,true,AAAGATGGTGAGGGAG-1_contig_1,true,551,IGH,IGHV4-61,,IGHJ6,IGHM,true,true,CARETTPVVTASTYYYYYGMDV
AAAGATGGTGAGGGAG-1,true,AAAGATGGTGAGGGAG-1_contig_2,true,644,IGL,IGLV2-14,,IGLJ3,IGLC3,true,true,CSSYTSSSTWVF,TGCAGCTG
AAAGTAGCAGATCCAT-1,true,AAAGTAGCAGATCCAT-1_contig_1,true,555,IGK,IGKV1-27,,IGKJ3,IGKC,true,true,CQKYNSAPFTF,TGTCAAAAGT
AAAGTAGCAGATCCAT-1,true,AAAGTAGCAGATCCAT-1_contig_2,true,577,IGH,IGHV1-2,,IGHJ6,IGHM,true,true,CARGGRVSVAVYWDYYYYGMDVV
AAAGTAGGTACAAGTA-1,true,AAAGTAGGTACAAGTA-1_contig_1,true,551,IGK,IGKV1-5,,IGKJ1,IGKC,true,true,CQQYNSYSWTF,TGCCAACAGTA
AAAGTAGGTACAAGTA-1,true,AAAGTAGGTACAAGTA-1_contig_2,true,549,IGH,IGHV2-5,IGHD3-9,IGHJ3,IGHM,true,true,CAHSAQYYDILTGYYK
```

```
>AAACCTGTCATATCGG-1_contig_1
TGGGGAGGAGTCAGTCCCAACCAGGACACGGCCTGGACATGAGGGTCCCTGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCTCAGGTGCCAG
ACAATTATGTAAATTGGTATCAGCAGAAACCAGGGAAAGCCCCTAAACTCCTGATCTACGATGCATTGAATTTAGAAATAGGGGTCCCATCAAG
ACGAACTTCCCGTCACTTTCGGCGGAGGGACCAATGTGGAAATGAGACGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGA
>AAACCTGTCCGTTGTC-1_contig_1
AGGAGTCAGACCCTGTCAGGACACAGCATAGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTGCTCTGGCTCCCAGGTGCCAGATGTG
TATTTAGCCTGGTATCAGCAAAAACCAGGGAAAGCCCCTAAGCTCCTGATCTATGCTGCATCCACTTTGCAAAGTGGGGTCCCATCAAGGTTCA
TACCCTCGGACGTTCGGCCAAGGGACCAAGGTGGAAATCAAACGAACTGTGGCTGCACCATCTGTCTTCATCTTCCCGCCATCTGATGAGCAGT
>AAACCTGTCCGTTGTC-1_contig_2
ATCACATAACAACCACATTCCTCCTCTAAAGAAGCCCCTGGGAGCACAGCTCATCACCATGGACTGGACCTGGAGGTTCCTCTTTGTGGTGGCA
TTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGAGGGATCATCCCTATCTTT
GAGATCTGAGGACACGGCCGTGTATTACTGTGCGACTACGTATTACTATGATAGTAGTGGTTATTACCAGAATGATGCTTTTGATATCTGGGGC
AACCCTTTTCCCCCTCGTCTCCTGTGAGAATTCCCCGTCGGATACGAGCAGCGTG
>AAACCTGTCGAGAACG-1_contig_1
ACTGTGGGGGTAAGAGGTTGTGTCCACCATGGCCTGGACTCCTCTCCTCCTCCTGTTCCTCTCTCACTGCACAGGTTCCCTCTCGCAGGCTGTG
GATATATTGGTACCAGCGGAAGCCAGGGAGTCCTCCCCAGTATCTCCTGAGGTACAAATCAGACTCAGATAAGCAGCAGGGCTCTGGAGTCCCC
TATGATTTGGCACAGCAGCGCTTGGGTGGTCGGCGGAGGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTC
GTGTCTCATAAGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCC
```

https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/output/overview

barcode	AAACCTGTCCGTTGTC-1	AAACCTGTCCGTTGTC-1
is_cell	TRUE	TRUE
contig_id	AAACCTGTCCGTTGTC-1_contig_1	AAACCTGTCCGTTGTC-1_contig_2
high_confidence	TRUE	TRUE
length	551	565
chain	IGK	IGH
v_gene	IGKV1-8	IGHV1-69D
d_gene		IGHD3-22
j_gene	IGKJ1	IGHJ3
c_gene	IGKC	IGHM
full_length	TRUE	TRUE
productive	TRUE	TRUE
cdr3	CQQYYSYPRTF	CATTYYYDSSGYYQNDAFDIW
cdr3_nt	TGTCAACAGTATTATAGTTACCCTCGGACGTTC	TGTGCGACTACGTATTACTATGATAGTAGTGGTTATTACCAGAATGATGCTTTTGATATCTGG
reads	5679	4161
umis	43	51
raw_clonotype_id	clonotype701	clonotype701
raw_consensus_id	clonotype701_consensus_2	clonotype701_consensus_1

No alleles for gene assignments, no way to determine germline identity for SHM for B cells

Post-processing 10x VDJ data

https://kjlj.github.io/scRNA-seq_VDJ/

IgBLAST post-processing of 10x contigs

`Cell Ranger` output does not include allele level gene assignment and doesn't provide information about identity to the germline gene that is needed for calculating somatic hypermutation (SHM) levels for the B cell datasets. To obtain this extra level of detail the contigs from 10x VDJ are re-aligned against the IMGT reference directory using IgBLAST. To get tab-delimited AIRR-C format for the VDJ alignments we will use the stand-alone version of `IgBLAST` rather than the web-based version.

In the interest of time, we will not undertake the `IgBLAST` post-processing during this session.

Rather we will start with datasets that have already been processed.

The steps for generating the datasets that we will be working with are documented:

1. Setting up IgBLAST and human reference databases
2. Obtaining VDJ datasets from 10x genomics
3. Running IgBLAST
4. Combining IgBLAST and Cell Ranger VDJ data and summarising by each cell barcode
5. Clone clustering for B cells

The results of running the above steps are two tab-delimited files; one for the T cells and one for B cells:

- B cell VDJ data with B cell clones defined: pbmc-tumour_Ig_cellranger_igblast_per-barcode_clns.tsv
- T cell VDJ data with clonotypes defined: pbmc-tumour_TR_cellranger_igblast_per-barcode.tsv

These are the two files that we will use for integrating VDJ data with scRNA-seq data.

Do I really have to post-process the datasets?

If working with T cell VDJ data only, then probably not, but for B cell VDJ data it is probably worth the effort, but you don't *have* to!

Ig contigs

Running `IgBLAST` on Ig:

```
#for the PBMCs
## navigate to the directory where the 10x data was downloaded
cd ~/data/scRNA-seq_workshop/pbmcs/

#IgBLAST requires the IGDATA environmental variable to be set
export IGDATA=~/data/apps/ncbi-igblast-1.21.0/
#can check what IGDATA is get to with
#echo $IGDATA

##running IgBLAST against the Ig references
##output to tab-delimited AIRR-C format
~/data/apps/ncbi-igblast-1.21.0/bin/igblastn \
    -germline_db_V ~/data/apps/ncbi-igblast-1.21.0/references/imgt_ig_v_human.fa \
    -germline_db_D ~/data/apps/ncbi-igblast-1.21.0/references/imgt_ig_d_human.fa \
    -germline_db_J ~/data/apps/ncbi-igblast-1.21.0/references/imgt_ig_j_human.fa \
    -c_region_db ~/data/apps/ncbi-igblast-1.21.0/references/ncbi_human_c_genes \
    -auxiliary_data ~/data/apps/ncbi-igblast-1.19.0/optional_file/human_gl.aux \
    -domain_system imgt -ig_seqtype Ig -organism human \
    -outfmt 19 \
    -query sc5p_v2_hs_PBMC_10k_b_filtered_contig.fasta \
    -out sc5p_v2_hs_PBMC_10k_b_filtered_contig.ig.igblast.tsv
```

TR contigs

Running `IgBLAST` on TR contigs:

```
#for the PBMCs
## navigate to directory for the PBMC data
cd ~/data/scRNA-seq_workshop/pbmcs/

##set IGDATA
export IGDATA=~/data/apps/ncbi-igblast-1.21.0/

##output to tab-delimited AIRR-C format
~/data/apps/ncbi-igblast-1.21.0/bin/igblastn \
    -germline_db_V ~/data/apps/ncbi-igblast-1.21.0/references/imgt_tr_v_human.fa \
    -germline_db_D ~/data/apps/ncbi-igblast-1.21.0/references/imgt_tr_d_human.fa \
    -germline_db_J ~/data/apps/ncbi-igblast-1.21.0/references/imgt_tr_j_human.fa \
    -c_region_db ~/data/apps/ncbi-igblast-1.21.0/references/imgt_tr_c_human.fa \
    -auxiliary_data ~/data/apps/ncbi-igblast-1.19.0/optional_file/human_gl.aux \
    -domain_system imgt -ig_seqtype TCR -organism human \
    -outfmt 19 \
    -query sc5p_v2_hs_PBMC_10k_t_filtered_contig.fasta \
    -out sc5p_v2_hs_PBMC_10k_t_filtered_contig.tr.igblast.tsv
```

Combine Cell Ranger & IgBLAST data for each contig:
https://kjlj.github.io/scRNA-seq_VDJ/docs/joining_cellranger_igblast.html

Build B cell clonal lineages:
https://kjlj.github.io/scRNA-seq_VDJ/docs/building_b_cell_clones.html

Working with 10x VDJ data in R

https://kjlj.github.io/scRNA-seq_VDJ/

Integrating scRNA-seq and VDJ

This session with focus on utilising post-processed VDJ data from 10x VDJ for integration with scRNA-seq.

We will build off the earlier sessions that generated the Seurat objects with the Azimuth annotation and the cell clustering. The Seurat objects are available from the Dropbox at pbmc and tumour, but have been pre-loaded into RStudio in Posit.cloud.

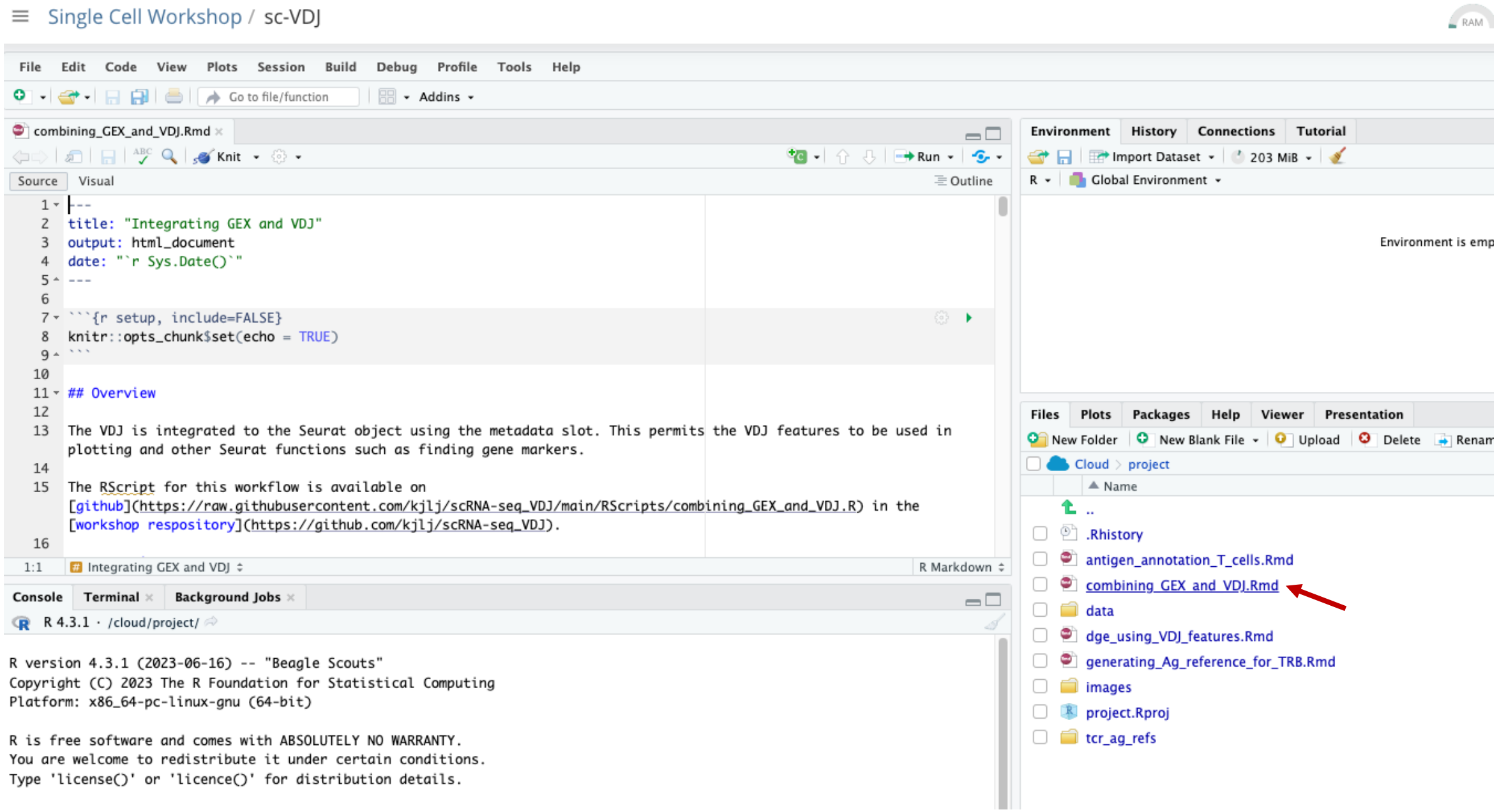
Topics:

- Adding VDJ to scRNA-seq including displaying VDJ features on UMAPs.
- Antigen annotation for T cells
    - generating reference datatset
    - annotating single cell clontoypes
- DGE between clones/clonotypes

Integrating scVDJ-seq with scRNA-seq

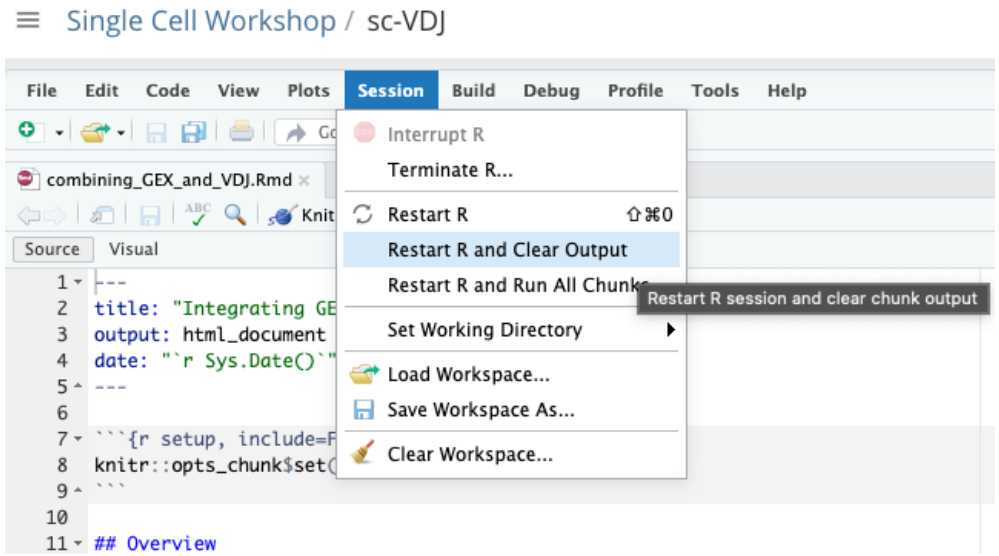
- Log into Posit Cloud
- Select ‘Kirby Institute, UNSW, Sydney’ as Organization
- Open the ‘single cell workshop’ space
- Start the ‘sc-VDJ’ assignment
- Open ‘combining_GEX_and_VDJ.Rmd’ in RStudio on Posit Cloud.

https://kjlj.github.io/scRNA-seq_VDJ/docs/combining_GEX_and_VDJ.html



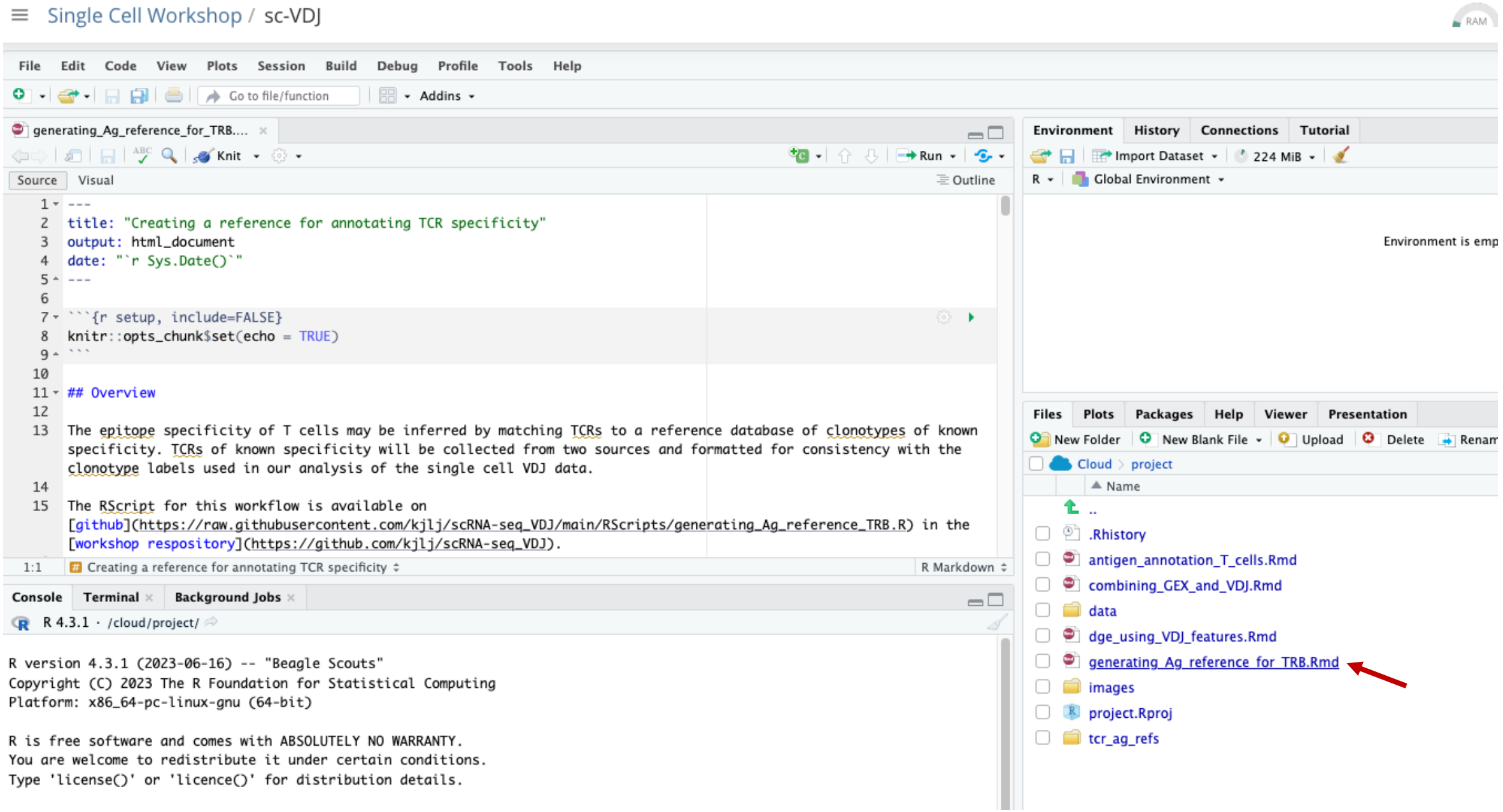
Annotating T cell epitope specificity

- Restart R session and clear all outputs

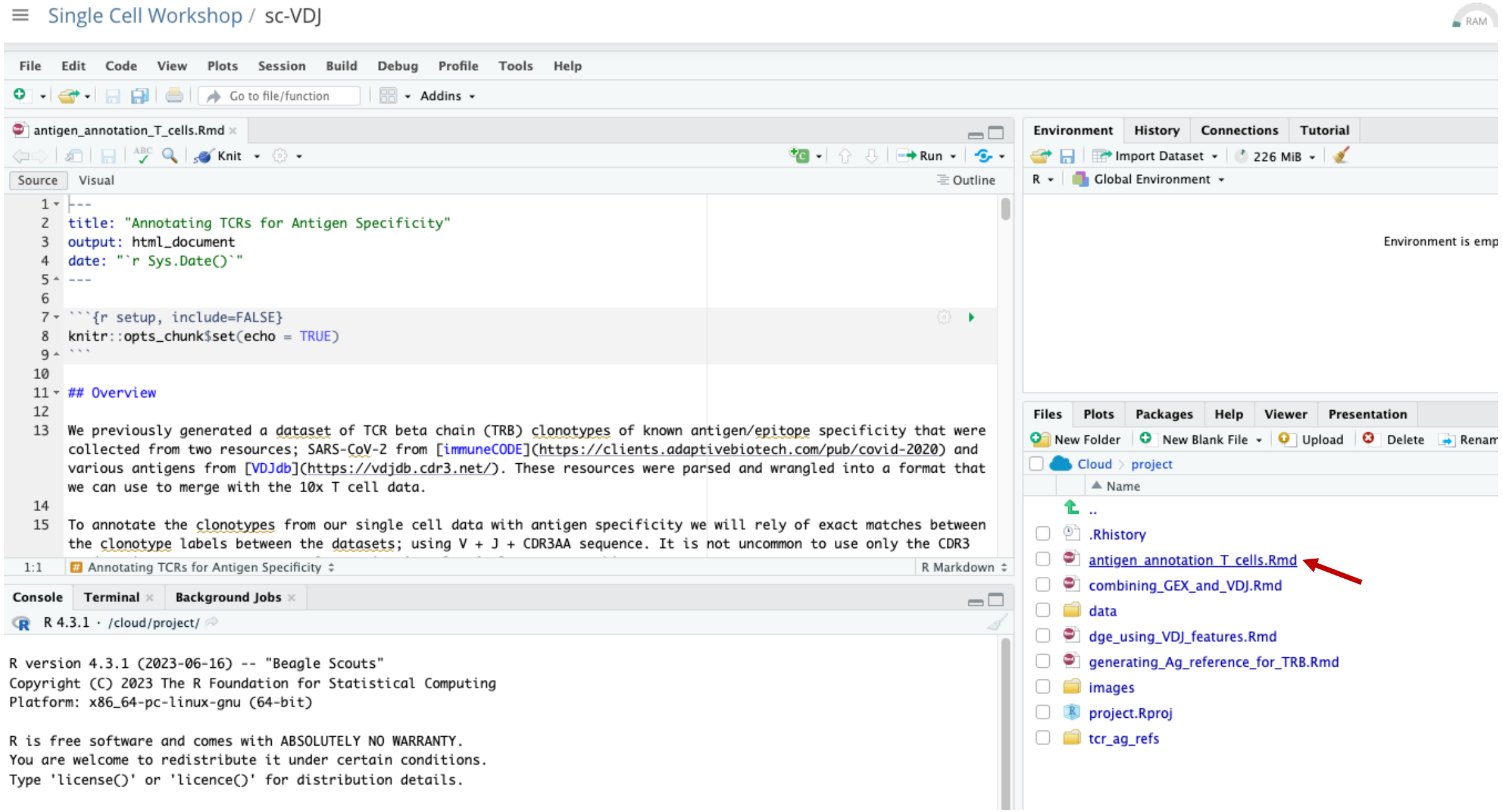


- Open ‘generating_Ag_reference_for_TRB.Rmd’

https://kjlj.github.io/scRNA-seq_VDJ/docs/generating_Ag_reference_for_TRB.html

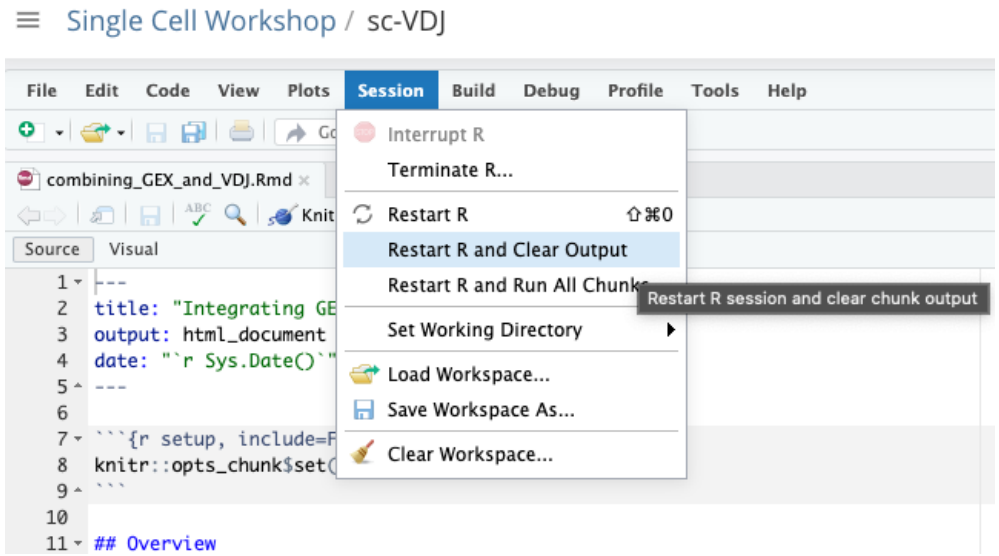


https://kjlj.github.io/scRNA-seq_VDJ/docs/antigen_annotation_T_cells.html



DGE with VDJ features

- Restart R session and clear all outputs



- Open ‘dge_using_VDJ_features.Rmd’

https://kjlj.github.io/scRNA-seq_VDJ/docs/dge_using_VDJ_features.html

Single Cell Workshop / sc-VDJ

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

dge_using_VDJ_features.Rmd

Knit | Run | Source | Visual | Outline

```
1  ---
2  title: "DGE using VDJ features"
3  output: html_document
4  date: "`r Sys.Date()`"
5  ---
6
7  ```{r setup, include=FALSE}
8  knitr::opts_chunk$set(echo = TRUE)
9  ```
10
11 ## Overview
12
13 Combing the metadata from the VDJ assay with the `FindMarkers()` function from `Seurat` permits differential gene
   expression (DGE) comparisons using the VDJ features such as expanded vs. unexpanded clones, comparing
   clones/clonotypes to each other, or comparing groups of cells with particular shared VDJ features like V isotype or
   germline gene usage.
14
15 The RScript for this workflow is available on
   [github](https://raw.githubusercontent.com/kjlj/scRNA-seq_VDJ/main/RScripts/dge_using_VDJ_features.R) in the
```

1:1 | DGE using VDJ features | R Markdown

Console | Terminal | Background Jobs

R 4.3.1 · /cloud/project/

```
R version 4.3.1 (2023-06-16) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
```

Environment | History | Connections | Tutorial

Import Dataset | 226 MiB

R | Global Environment

Environment is emp

Files | Plots | Packages | Help | Viewer | Presentation

New Folder | New Blank File | Upload | Delete | Renam

Cloud > project

Name

- ..
- .Rhistory
- antigen_annotation_T_cells.Rmd
- combining_GEX_and_VDJ.Rmd
- data
- dge_using_VDJ_features.Rmd
- generating_Ag_reference_for_TRB.Rmd
- images
- project.Rproj
- tcr_ag_refs

Additional resources

https://kjlj.github.io/scRNA-seq_VDJ/

Thank you

Resources will be available from the Github repository:

- https://github.com/kjlj/scRNA-seq_VDJ
- https://kjlj.github.io/scRNA-seq_VDJ/

Today's session, like a light chain CDR3, was short! 🤓

If you would like a deeper dive into VDJ repertoire analysis please let us know in the feedback so that we can plan for future sessions and update to the resources on the Github.