single cell RNAseq

We prepared all our scRNAseq libraries using the technology PIPseq and thus we also always use their proprietary software PIPSEEKER for the basic analysis, followed by seurat and other tools.

At the beginning, we used to use the pipseeker v2.1.4 but this is no longer compatible with some of the features present in the libraries prepared with their updated experimental protocols (and as of 1/2025 the company was acquired by Illumina so there might be more changes coming in the future). Note that the experimental protocol evolved quite a bit and so did the pipeline. Hence, we now use the docker version fluent-pipseeker\_3.3.0.sif that is retro-compatible, but for each run, you have to specify which chemistry version was used (notes in the script) as depending on the version, there are differences on how the reads are assambled. The docker image is located here, along with the running script 01\_pipseeker\_docker.sh:

/share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/01\_pipseeker\_docker.sh

Here is a backup of mock tutorial data for testing the pipseeker pipeline if needed

/share/lab\_avram/HPC\_Cluster/user/tomas/scRNAseq\_tutorial-data.tar.gz

Note that pipseeker also requires references to be in a specific format so we just used whatever they provided. For human and mouse it is:

GRCm39 (GENCODE vM29 2022.04, Ensembl 106)

GRCh38.p13 (GENCODE v40 2022.04, Ensembl 106)

They also provide a reference file that is combination of human+mouse (pipseeker-gex-reference-GRCh38-and-GRCm39-2022.04) which can be used for mixed experiments that we also performed. All these references are here:

/share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/references

As of 3-6-2025, there are 4 sets of scRNAseq data that we performed:

1) The most important data is here, focusing on transferred antigen-specific CD8 T cells isolated from ovarian (OT1) and melanoma (PMEL) tumors. This will be included in the upcoming Bcl11b-CD8 paper. The basic folder contains the basic analysis by pipseeker pipeline (focus on the analysis by pipseeker v3.3.0, such as fodler: results\_v3.3.0\_PIP3M8OTK2\_no\_annot). Each sample folder also contains results from TRUST4 analysis to identify TCR transcripts and the summarizing trust4 analysis is saved in tcr\_trust folder, all in:

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023

Then, we used the matrix files generated by pipseeker and placed them as input into the BERLIN folder for further processing. In the BERLIN folder, there is Output folder that contains all the scRNAseq results that we will use in the paper. Mostly focus on the folder ovarT\_integrated\_noRefUsed which contains integrated data from four ovarian samples - two WT and two KO samples from Bcl11bF/F GzmbCre mice (melanoma mouse data is also there as melanT\_integrated\_noRefUsed but Dorina now does not plan to include it).

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/BERLIN/2-Single\_Cell\_RNAseq\_Pipeline/Output/ovarT\_integrated\_noRefUsed

This is the script used for the analysis of both ovarian and melanoma data is here (just comment or uncomment the ovarian vs melanoma section at the beginning and the rest should re-create the plots in the output folders):

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/BERLIN/12\_berlin\_v5.0.1\_combined.R

The most important Rds file for further processing is this one for ovarian data:

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/BERLIN/2-Single\_Cell\_RNAseq\_Pipeline/Output/ovarT\_integrated\_noRefUsed/03\_reclustered\_ovarT\_integrated\_noRefUsed\_seurat\_obj.Rds

And this one for melanoma:

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/BERLIN/2-Single\_Cell\_RNAseq\_Pipeline/Output/melanT\_integrated\_noRefUsed/03\_reclustered\_melanT\_integrated\_noRefUsed\_seurat\_obj.Rds

In the "reclustered" folder you can find all kinds of plots.

**FOR RNA velocity analysis**, we needed to prepare a few more files. These commands are also saved in 01\_pipseeker\_docker.sh. Specifically:

sample="PIP3M8OTW2"

cd /share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/wt\_ovarian\_mouse\_tumor\_cd8\_ot1\_rep2

# first run the full pipseeker run to analyze the sample

singularity run /share/lab\_avram/HPC\_Cluster/user/tomas/pipseeker-docker/fluent-pipseeker\_3.3.0.sif full --chemistry v4 --fastq $sample --star-index-path /share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/references/pipseeker-gex-reference-GRCm39-2022.04 --output-path results\_v3.3.0\_"$sample"\_no\_annot --save-svg --dpi 400 --random-seed 42 --resume-last-run --skip-version-check

# then run the barcode version of pipseeker

singularity run /share/lab\_avram/HPC\_Cluster/user/tomas/pipseeker-docker/fluent-pipseeker\_3.3.0.sif barcode --chemistry v4 --fastq $sample --star-index-path /share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/references/pipseeker-gex-reference-GRCm39-2022.04 --output-path results\_v3.3.0\_"$sample"\_no\_annot --random-seed 42 --sorted-bam --resume-last-run --skip-version-check

# append the sample barcode to beginning in bam

python /share/lab\_avram/HPC\_Cluster/user/tomas/bin/bamtagregex.py starsolo\_sorted.bam starsolo\_sorted\_renamedCB\_"$sample".bam --tag CB --pattern ^ --replace "${sample}\_"

# index and sort the bam file

samtools index starsolo\_sorted\_renamedCB\_"$sample".bam

samtools sort -t CB -O BAM -o cellsorted\_starsolo\_sorted\_renamedCB\_"$sample".bam starsolo\_sorted\_renamedCB\_"$sample".bam

samtools index cellsorted\_starsolo\_sorted\_renamedCB\_"$sample".bam

# then convert this modified bam file into loom file that is used as an input for scVelo or UNITVELO. This is done by velocyte tool that is run from a docker image, as described in this script:

/share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/03\_shortcake.sh

Specifically:

cd /share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/wt\_ovarian\_mouse\_tumor\_cd8\_ot1\_rep2/results\_v3.3.0\_PIP3M8OTW2\_no\_annot

singularity exec /share/lab\_avram/HPC\_Cluster/user/tomas/shortcake-docker/shortcake\_full.3.0.0.sif run\_env.sh shortcake\_default velocyto run -b barcodes.filt.FULL\_BARCODE.OTW2\_10-15-24.txt -o velocyto\_wt\_ovarian\_mouse\_tumor\_cd8\_ot1\_rep2\_7-23-24 starsolo\_sorted\_renamedCB\_OTW2.bam ../../gencode.vM29.primary\_assembly.annotation.gtf

# separately, in R, prepare a loom file from the processed integrated data - this will contain also all the metadata so it can be put all together - the loom part preparation is part of the main script here:

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/BERLIN/12\_berlin\_v5.0.1\_combined.R

Then to perform the RNA velocity analysis, follow the steps saved in this python notebook:

/share/lab\_avram/HPC\_Cluster/user/tomas/10\_PIPseq2-4\_scRNAseq\_CUTRUN\_CD8\_melanoma\_ovarian\_mouse\_10-19-2023/scvelo\_7-23-24/unitvelo\_10-17-24\_WT\_KO\_separately.ipynb

In the same folder, you can also find the input files and results.

2) Next there is another set of data, this is the most complex data containing some samples that are only mouse - all immune cells from recipient mouse, and others which are combined mouse+human TILs, where mouse are again the recipient immune cells from ovarian tumors transferred with CD8 T cells and the human are melanoma TILs from Shari Pilon-Thomas lab

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024

On an example of combined mouse+human sample:

nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen\_huTILs\_p84\_P12HM4OC4WN

We see that there are three folders:

results\_v3.3.0\_P12HM4OC4WN\_no\_annot

results\_v3.3.0\_P12HM4OC4WN\_no\_annot\_human\_mapping

results\_v3.3.0\_P12HM4OC4WN\_no\_annot\_mouse\_mapping

The first folder is output from pipseeker pipeline v3.3 analysis using the combined mouse+human reference file. Make sure to select also the right chemistry in the pipseeker script as shown in the script

/share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/01\_pipseeker\_docker.sh

This output is then used to take the raw\_matrix (matrix, feature, barcode) files and use them as an input for the R script. I used this R script to first preprocess the human+mouse data:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/1\_berlin\_v5.0.1\_mixed\_humanmouse.R

This is also as part of the 3\_berlin\_v5.0.1\_human.R script but one has to change the organism at the beginning of the script and comment out some sections at the beginning of the script.

from there, we have an output in the form of:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/1\_mixed\_human\_mouse\_samples/Output/ nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen\_huTILs\_p84/Single\_Cell\_RNAseq\_Output/raw\_counts\_sub\_withNAs\_nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen\_huTILs\_p84.txt

Next we use this script:

sbatch /share/lab\_avram/HPC\_Cluster/user/tomas/bin/scRNAseq\_pipeline\_pipseeker\_v3.3.0/02\_convert\_barcodes\_meltData.sh

This will further modify the raw\_counts files which are also located here where this analysis takes place, ultimately producing barnyard\_metrics\_nt\_recip\_dnmt3\_tbetcre\_huTILs\_p41.txt and some figures:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/1\_mixed\_human\_mouse\_samples/Output/extract\_barnyard\_plots\_and\_species\_barcodes

The resulting barnyard\_metrics\_nt\_recip\_dnmt3\_tbetcre\_huTILs\_p41.txt are then used as an input for the separate processing of mouse and human data, such as you can see here:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/2\_recipients\_mouse/Input/pipseeker\_v3.3/nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen

For the separate processing of mouse and human data, we need to the other folder that were created by the basic pipseeker pipeline, i.e. the folders where we specifically mapped the reads to either mouse or human genome, such as this folder for mouse data:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen\_huTILs\_p84\_P12HM4OC4WN/results\_v3.3.0\_P12HM4OC4WN\_no\_annot\_mouse\_mapping

From this folder, we need to use again the raw matrix files (matrix, feature, barcode) and use them as an input for mouse-specific analysis, where the script will utilize the information on the mouse-specific barcodes from barnyard\_metrics\_nt\_recip\_bcl11b\_cd4ercre\_noTamoxifen\_huTILs\_p84.txt which should also be in the input directory. Then to analyze the mouse data, just run this script:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/2\_berlin\_v5.0.1\_mouse.R

This is where the output is then located for mouse recipient data:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/2\_recipients\_mouse/Output

For human data, analogically use barcodes/feature/matrix files from the human-mapped folder results\_v3.3.0\_P12HM4OC4WN\_no\_annot\_human\_mapping and put them in the input folder along with the corresponding barnyard\_metrics file, such as here:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_human\_TILs/Input/pipseeker\_v3.3/huTILs\_p19

Then run this human script (which is the ultimate script that could also be run for the pre-processing or for the mouse data if the organism and a few more lines are correctly commented at the beginning of the script:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_berlin\_v5.0.1\_human.R

The output for human TIL data can then be found here:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_human\_TILs/Output/hg\_melanT\_integrated\_noRefUsed\_noReg/reclustered

the final human dataset that we may use in the paper would be this one - it is CD8 T cells isolated based on clustering…. /share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_human\_TILs/Output/hg\_melanT\_integrated\_noRefUsed\_noReg/reclustered/01\_cd8\_cluser-based\_seu\_reclusthg\_melanT\_integrated\_noRefUsed\_noReg\_030325.Rds

This dataset contains all the TILs from all three patients we have right now that are integrated together

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_human\_TILs/Output/hg\_melanT\_integrated\_noRefUsed\_noReg/01\_hg\_melanT\_integrated\_noRefUsed\_noReg\_seurat\_obj.Rds

And there are also Rds files for individual patients stored here, in their respective folders:

/share/lab\_avram/HPC\_Cluster/user/tomas/15\_PIPseq8-13\_10-18-2024/BERLIN/2\_recipients\_mouse\_human\_TILs\_PIPseq8-13\_10-18-2024/3\_human\_TILs/Output/huTILs\_p41/Single\_Cell\_RNAseq\_Output/huTILs\_p41.Rds

3) Next, this is from peritoneal CD8 T cells from mouse ovarian mouse model - NOT NEEDED NOW, as we focus on intra-tumor T cells. This resides here:

/share/lab\_avram/HPC\_Cluster/user/tomas/09\_pipseq1\_6-30-23

Analysis is in the BERLIN folder in there, and interactive way to check the data is here in the shiny app:

https://biostools.moffitt.org/4476125/09\_pipseq1\_6-30-23/

4) CD4 T cells from ovarian model - NOT WORKING WELL

This dataset is of very poor quality as explained in the email to Dorina from 2/6/2024. I included this data and text of the email in the BERLIN folder (in ppt), which also contains the analyzed data. The basic processing of the data is in the parental directory here:

/share/lab\_avram/HPC\_Cluster/user/tomas/11\_PIPseq6\_ovarian\_CD4\_OT2\_01-11-2024/

This is also available in the shiny app:

https://biostools.moffitt.org/4476125/11\_PIPseq6\_ovarian\_CD4\_OT2\_01-11-2024