OVERVIEW

This repository contains the complete computational pipeline and analysis code from the master's thesis " Development of a pipeline for spatial transcriptome analysis of samples from Hodgkin's lymphoma patients" (2026, Justus Liebig University Giessen).

The pipeline enables automated, reproducible analysis of 10x Genomics Visium low-density spatial transcriptomics data from FFPE tissue samples, with applications to classical Hodgkin lymphoma (cHL) and beyond.

CONTENTS

- Bash script for automated preprocessing with Space Ranger (v3.1.3)

- JupyterLab notebooks (`.ipynb`) for quality control, batch correction (Harmony, scVI), and downstream analysis

- PDF reports documenting all analysis steps and results

- Helper script (Python) for data organization (`collect_output.py`)

- CSV templates (`SampleSheet.csv`, `Aggregation.csv`) for metadata input

- Test dataset analysis for pipeline validation using mouse spleen samples (GEO: GSE254652)

- HTML tables with QC metrics, correlation matrices, and differential expression results

- Figures and visualizations from all analyses

DATA ANALYSIS WORKFLOW

1. Automated sample processing (Space Ranger `count` & `aggr`)

2. Quality control across 13 classical Hodgkin lymphoma samples

3. Integration attempts using Harmony and scVI

4. Marker gene expression analysis

5. HRS spot characterization

6. Pipeline validation using test dataset (mouse spleen)

TECHNICAL REQUIREMENTS

- Operating System: Ubuntu 24.04.2 LTS (or compatible Linux distribution)

- Space Ranger: v3.1.3

  - Download: https://www.10xgenomics.com/support/software/space-ranger/downloads/previous-versions

- Python: 3.12.10 with Scanpy (v1.10.4), Squidpy (v1.6.5), harmonypy (v0.0.10), scVI (v1.3.1.post1)

  - Installation:

    Create conda environment:

    - `conda create -n <env_name> python=3.12.10`
    - `conda activate <env_name>`

    Install packages:

    - `conda install scanpy=1.10.4`
    - `conda install conda-forge::squidpy=1.6.5`
    - `conda install -c conda-forge scvi-tools=1.3.1.post1`
    - `conda install bioconda::harmonypy==0.0.10`

- Reference files (10x Genomics):

- Transcriptome: `refdata-gex-GRCh38-2020-A`
- Probe set: `Visium_Human_Transcriptome_Probe_Set_v2.0_GRCh38-2020-A`

  Download: https://www.10xgenomics.com/support/software/space-ranger/downloads

  Note: Select "Download 2020-A references" for transcriptome

- Reference files for mouse spleen validation (10x Genomics):

- Transcriptome: `refdata-gex-mm10-2020-A`
- Probe set: `Visium_Mouse_Transcriptome_Probe_Set_v1.0_mm10-2020-A.csv`

  Download: https://www.10xgenomics.com/support/software/space-ranger/downloads

  Note: Select "Download 2020-A references" for transcriptome

FILE STRUCTURE

Main Analysis Scripts (`classical_Hodgkin_lymphoma_samples/`)

- `script.sh`: Main automation script for Space Ranger processing
- `collect_output.py`: Organizes Space Ranger outputs into `sample_info.csv`
- `sample_info.csv`: Structured sample metadata (generated by `collect_output.py`)
- `SampleSheet.csv`: Input metadata for Space Ranger `count`
- `Aggregation.csv`: Input metadata for Space Ranger `aggr`
- `Quality_Control.ipynb`: Quality control across all samples
- `Integration_All_Samples.ipynb`: Integration with Harmony/scVI
- `Integration_Best_Samples.ipynb`: Integration of high-quality samples
- `Markergenes_Best_Samples.ipynb`: Marker gene expression analysis
- `HRS_Cells_K001927-19.ipynb`: HRS spot characterization
- PDF reports for all notebooks
- `tables/`: HTML tables with QC metrics, correlation matrices, and differential expression results (DEG only for K001927-19)
- `figures/`: Generated plots and visualizations

Validation Dataset (`test/`)

- Pipeline validation using mouse spleen samples (GEO: GSE254652)
- Quality control of test samples
- Adapted marker gene analysis for mouse tissue
- Integration workflow validation
- `tables/`: HTML tables with QC metrics, and correlation matrices
- `figures/`: Generated plots and visualizations

USAGE

Prerequisites:

- Install Space Ranger v3.1.3 (https://www.10xgenomics.com/support/software/space-ranger/downloads/previous-versions)

- Download reference transcriptome (`refdata-gex-GRCh38-2020-A`) and probe set

- Install conda environment and required Python packages (see TECHNICAL REQUIREMENTS)

Step-by-step workflow:

1. Prepare input files:

   - Create `SampleSheet.csv` with sample metadata (`SAMPLE_ID, FASTQ_PATH, IMAGE_PATH, CYTAIMAGE_PATH, SLIDE_ID, AREA, TRANSCRIPTOME, PROBE`)

   - Create `Aggregation.csv` with spaceranger aggr parameters (`LIBRARY_ID, MOLECULE_H5, CLOUPE_FILE, SPATIAL_FOLDER`)

2. Configure Space Ranger path:

   - Update Space Ranger installation path in `script.sh`

   - Update base path in `collect_output.py`

3. Run automated preprocessing:

```
bash script.sh
```

   (Executes `spaceranger count` for all samples + `spaceranger aggr`)

4. Organize outputs:

```
python collect_output.py
```

   (Generates `sample_info.csv` with structured metadata)

5. Configure JupyterLab paths:

   - Update base paths in all JupyterLab notebooks to your data directory

6. Execute downstream analyses sequentially:

   a. `Quality_Control.ipynb` - QC metrics and sample comparison

b. `Integration_All_Samples.ipynb` - Batch correction attempts (Harmony, scVI) on all samples, no filtering
c. `Integration_Best_Samples.ipynb` - Integration of high-quality samples (K001927-19, K43-19), with QC filtering
d. `Markergenes_Best_Samples.ipynb` - Cell type marker expression analysis
e. `HRS_Cells_K001927-19.ipynb` - HRS spot characterization and differential gene expression

7. (Optional) Validate pipeline using test dataset:

- Navigate to `test/` directory

- Follow same workflow with mouse spleen samples

ACKNOWLEDGMENTS

This work was conducted as part of a Master's thesis at Justus Liebig University Giessen (2026).

Supervision:

Prof. Dr. Marek Bartkuhn (Institute of Biochemistry)

David Glaser, M.Sc. (Institute of Biochemistry)

Experimental Work:

Dr. Rredhi (Institute for Pathology, AG Prof. Dr. Bräuninger)

- Sample preparation, H&E staining, Visium workflow, sequencing, and manual HRS spot annotation

Study Context:

Patient samples from EuroNet-PHL-C2 study