Single Cell Transcriptomics in Skeletal Muscle

Johanna GALVIS-LASCROUX

supervised by
Dr. Fabien LE GRAND
Institut Neuromyogène INSERM

M1 Master Bio-Informatique, UCBL

Mars-July 2020

Contents

Chapter 1

Introduction

1.1 Introduction

The muscle physiology and cell biology is the heart of research activity at Institute Neuro-Myogène. At Le Grand team, skeletal muscle regeneration and repair mechanisms are studied at the level of the single-cell, beyond the multi-nucleated muscle fibers, as its this micro-environment is composed of different types of cell populations with characteristics and complex interactions to be elucidated.

Because of its ability to differentiate into skeletal fibers, the Satellite cell (**MuSC**) has been the target of extensive research: by means of single cell RNA sequencing (scRNA-seq) technology on murine models it has been possible to describe the the transition of MuSCs from a quiescent state to an activated phase that conduces to proliferation or differentiation [4, 3, 8] and also interesting cellular sub-populations are emerging from these omics studies[6]. Moreover, throughout the repair process after injury induced by cardiotoxin, authors have also suggested the existence of different sub-types of Fibro-adipogenic progenitors (**FAPs**) mainly Pi16+ and Smoc2 sub-populations [3] whose expression profile has make possible to infer cell-fate decisions modeled in pseudo-time trajectories [4].

Hence, re-analyzing available data-sets in homeostasis, as well as during the repair process, would consolidate information about Satellite cells and FAPs and set a useful background to the next scRNAseq study by Le Grand team.

1.2 Objectives

The specific objectives of this work are to adapt and test tools to perform an adequate quality control, implement R packages and software to identify the cell populations (clusters), extract those genes exhibiting high variation among clusters and visualize changes in gene expression in different time-points after injury. As FAPs and MuSCs are of our interest, we will focus on these sub-populations to verify gene expression levels during skeletal muscle repair, run ligand-receptor analyses and visualize trajectories in pseudotime.

1.3 Single-Cell Samples Processing and cDNA sequencing

Before entering into expression data analysis itself, it is compulsory to start with a general description of the technical procedures common to all drop-based single cell transcriptomics projects.

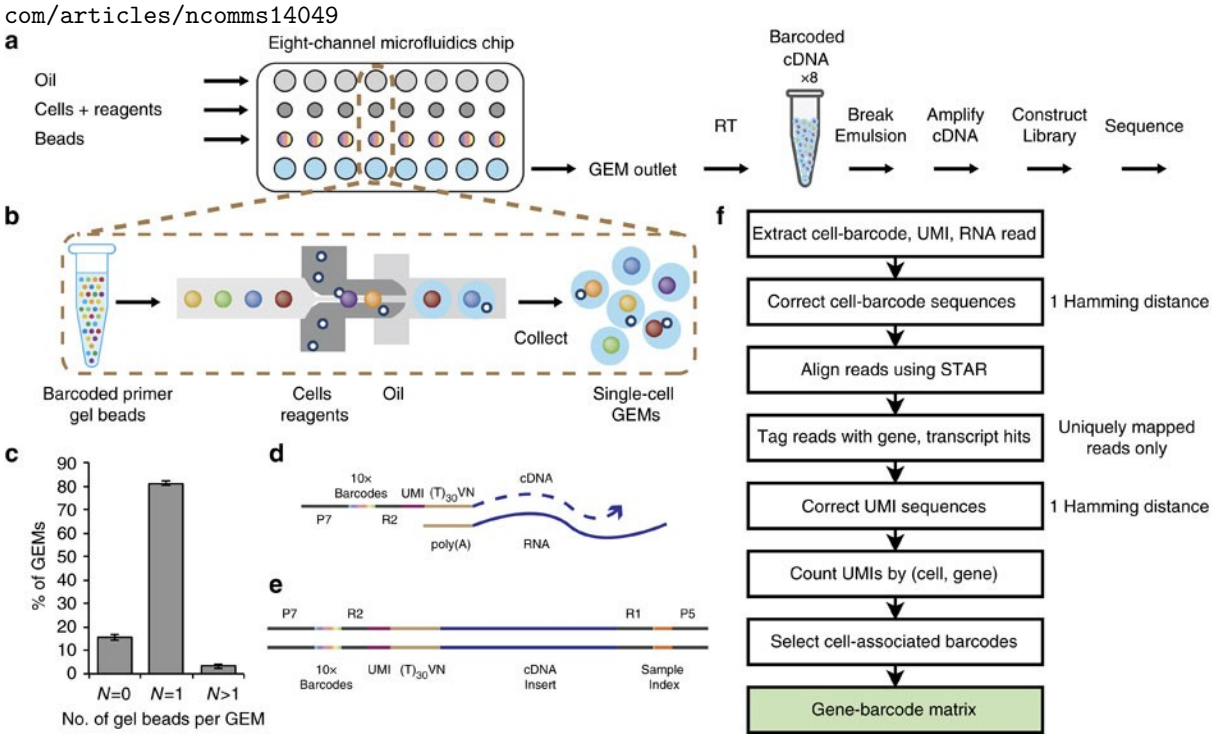
Isolated cells from skeletal muscle are sorted and subjected to the Chromium transcriptome 3' protocol (10X Genomics). The 10X protocol is often performed by a third party laboratory, Figure 1.1 resumes all these steps. Briefly, Gel Beads-in-Emulsion (GEM) technology allows to perform full cDNA sequencing of each cell inside independent droplets identified by unique barcoded beads (which are added to the beads during barcodes library preparation step). One GEM (or droplet) contains a cell and all the needed reagents, including unique molecular identifiers (UMIs), to perform RNA reverse transcription, fragmentation, **UMI linking**, amplification, library generation and sequencing. To note, UMI tags facilitate the merging of PCR duplicates, minimising amplification biases during analysis.

FASTQ files are generated by sequencing via Illumina technology. At this point, several options to perform **alignement and UMI counting** are available, for example the pseudoalignment approach

by Kallisto is an optimal choice when dealing with very big datasets (hundreds of barcodes) given its algorithmic efficiency (up to 7-fold less memory requirements) while yielding very close results to those by STAR when gross cell type identification is the goal. The most popular option is CellRanger pipeline, developed by 10X Genomics, which in turn includes task-specific pipelines:

- 'cellranger mkfastq' : demultiplexing, it is, assign reads to their corresponding barcodes.
- 'cellranger count' : alignement using the splice aware aligner STAR, UMI count by cell and gene, and filtering. and a count matrix with rows representing features/genes and columns representing the barcodes, hence gene expression values are given in the form of UMI counts per feature per cell.

Figure 1.1: Single-Cell Chromium 10X protocol with Gemcode technology, from `https://www.nature.com/articles/ncomms14049`



Raw count matrix is then ready to be subjected to Quality Control and downstream analysis

1.4 Quality Control: Pre-processing the count matrix

Whatever the pipeline/software for UMI counting is used in a given setting, it is compulsory to run routine strict quality control (QC) before any downstream analysis, to minimize artefactual findings that may conduce to misinterpretations.

Each single experimental batch must be subjected to QC. Note that QC on combined experimental batches (i.e. integrated objects or merged experiments) yields errors, especially during multiplets detection (more than one cell by drop). Accordingly to published reviews [7], the packages `scater` and `scran` are recommended for this purpose. These packages require the data-matrix to be transformed into a SingleCellExperiment object, so dedicated tools can be used to filter out empty barcodes, low-quality cells (based on UMI counts and ranked barcodes), or multiplets, that escaped upstream filters.

Moreover, with the help of other Bioconductor packages such as AnnotationDBI it is possible to detect contamination that eventually escaped previous filters, by verifying species origin (*Mus musculus* vs *Homo sapiens*) of every feature/gene in the matrix.

Full code to run a QC workflow following experts recommendations can be found at `https://github.com/LeGrand-Lab/QC_single_cell` , which includes most of code from Dr. L Modolo (`http://perso.ens-lyon.fr/laurent.modolo/scRNA/`) available under creative commons licence.

Chapter 2

Materials and Methods

2.1 scRNA-seq Data

Datasets from published papers are available at GEO (Gene Expression Omnibus) repository of the NCBI. **Raw count matrices** were obtained as seen in table 1.

Authors	Selected	GEO
Dell'Orso	Homeostasis	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3614992`
		`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3614993`
De Micheli	Homeostasis	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE143437`
Giordani	Homeostasis	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3520458`
		`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3520459`
Oprescu	Homeostasis and post-injury	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138826` ** **see 'Notes'

Table 2.1: Datasets

2.2 Software

The analysis were performed directly in R version 3.6.3, by using specific packages:

- scater v1.16.2, scran v1.16.0, DoubletFinder, and dependencies: for quality control.
- Seurat v3.1.5, and all dependencies : for clustering and differential gene expression inter-clusters.
- nichenetr (from saeyslab) : to perform ligand-receptor pairs analysis[2].
- ClusterProfiler v 3.16.0: for Gene Ontology enrichement analysis.
- Monocle v 3 alpha (2.99), and dependencies : for trajectory analysis.

The operative system architecture was a MacOs Mojave V 10.14.6 system, using a processor Intel Core i5 3GHz and 16 Go of memory (RAM). To allow correct installation of R packages and required compilers and dependencies, admin privileges are required, for instance to run 'make' and 'xcode-select –install" in the case of gcc (indispensable for Rcpp package). I also installed XQuartz for X11 server (formerly X Command Line Utilities) and homebrew.

2.3 Pipelines

2.3.1 QC

The "extended" QC version mentioned in section 1.2 of this document was replaced by an alternative version of QC , encapsulating steps into 'runPrep_scran', 'rundoublets_scran' and 'knee_plot' functions. (see `https://github.com/LeGrand-Lab/INMG_SingleCell/blob/master/scripts/functions_stock.R`) and using directly scater::isOutlier.

Specifically, the following were used:

scater::perCellQCMetrics yields a dataframe containing: counts sum (i.e. the library size) and number of detected features. Also contains a nested DataFrame of a subset (for example, sub-matrix of non-genomic features per barcode) if subset is specified. A threshold for the number of detected features (detection.limit) can be tuned, default is 0.

scater::addPerFeatureQC mean counts per feature, also allows to calculate for a negative control (passed as subset).

scran::computeSumFactors scaling normalization (implements deconvolution)

scater::isOutlier determine which values in a numeric vector are outliers based on the median absolute deviation (MAD). for inferior outliers preferably, as superiors are multiplets (these are marked otherwise).

scran::doubletCells on a log-transformed matrix, simulates thousands of doublets in the vicinity of each "cell", then determines doublets vs observed-cells densities for each barcode. The ratio simulated-doublets/observed-cells yields a score such that the greater the score, the most likely that "cell" is a multiplet. This is why must NOT to be done on combined batches (impossible multiplets will be reported!), same is valid for DoubletFinder.

DoubletFinder was runned independently of scater::doubletCells. As expected, scoring is different from one algorithm to another, so the values under the 0.95 percentile were considered as singlets on the hypothesis of a 5% multiplets formation. Barcodes found in the intersection of those classified as "Doublet" (or "doublets") by each method were marked to be excluded from the ulterior analysis.

2.3.2 Downstream Analyses

Seurat and Monocle tools were employed following respective official tutorials and vignettes. All related scripts are available at 'scripts' directory at INMG_SingleCell Github repo. Appendix A.A shows organization of the different pipelines and workflows, in diagrams.

Ligand-receptor and gene ontology scripts and results are found in `https://github.com/LeGrand-Lab/LigRecPairs_GO`.

Chapter 3

Results

3.1 Homeostasis (D0)

3.1.1 Re-Analysing Published Data in Homeostasis

Clustering and relevant markers identification were performed on each dataset (DellOrso, De Micheli, Giordani, Oprescu) in homeostasis (prefix 'D0' or 'Noninjury' on barcodes in the case of plain text data-matrices, or coherently named 10Xfolder/csvfiles). Nearly the same clusters were obtained across all four groups, but Giordani and Dell Orso showed a big proportion of immune cells, this being explained by the experimental design: popliteal lymph nodes were present as complete hind-limb was processed. Concordances in cell type clusters were better seen in integrated analysis Giordani + DellOrso + De Micheli

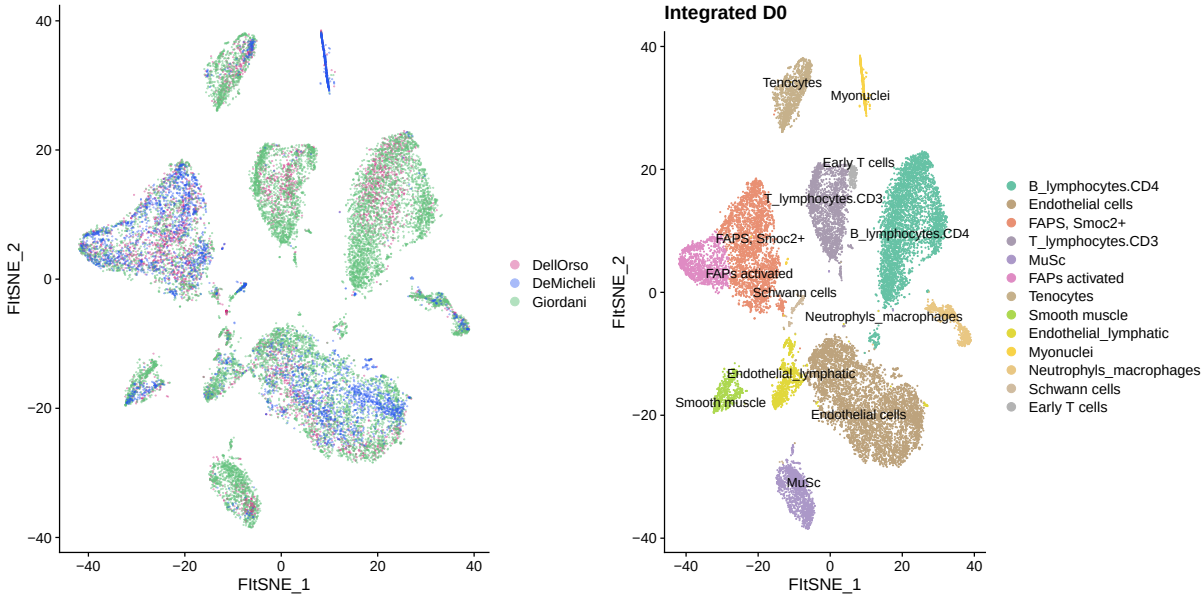


Figure 3.1: Cells co-clustering across three different data-sets

In a second time, Oprescu and De Micheli were analysed as an integrated Seurat object, as their experimental design was similar: exclusively tibialis anterior muscle was processed on both studies. Figure 3.2 shows key genes expression, which allows to distinguish the two FAPs sub-population of interest: Smoc2+ and Pi16+s. More detailed results are available at `https://github.com/LeGrand-Lab/LigRecPairs_GO/tree/master/results_DeOpNEWseu`
Using this integrated object, Gene Ontology enrichement and Ligand-Receptor pairs analyses were performed.

3.1.2 Detected ligands of relevance in Pi16 population

Computational predictions obtained by means of the nichenet model matrix on Oprescu+DeMicheli integrated seurat object in Homeostasis (D0) are shown in figure 3.3 .

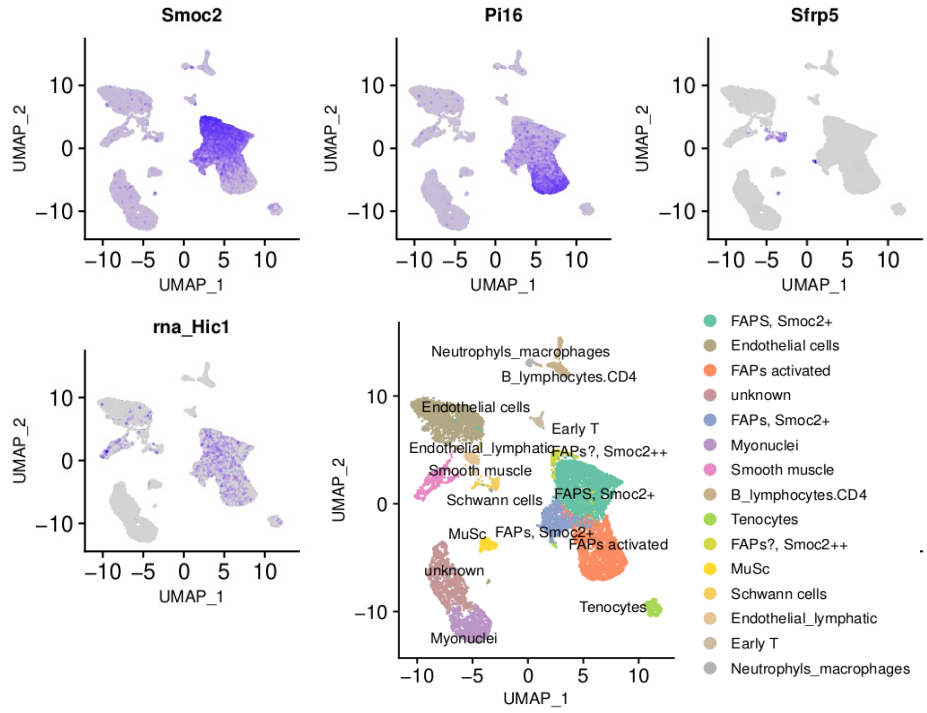


Figure 3.2: Oprescu and De Micheli integrated data. Note: "unknown" correspond to slow muscle fibers.

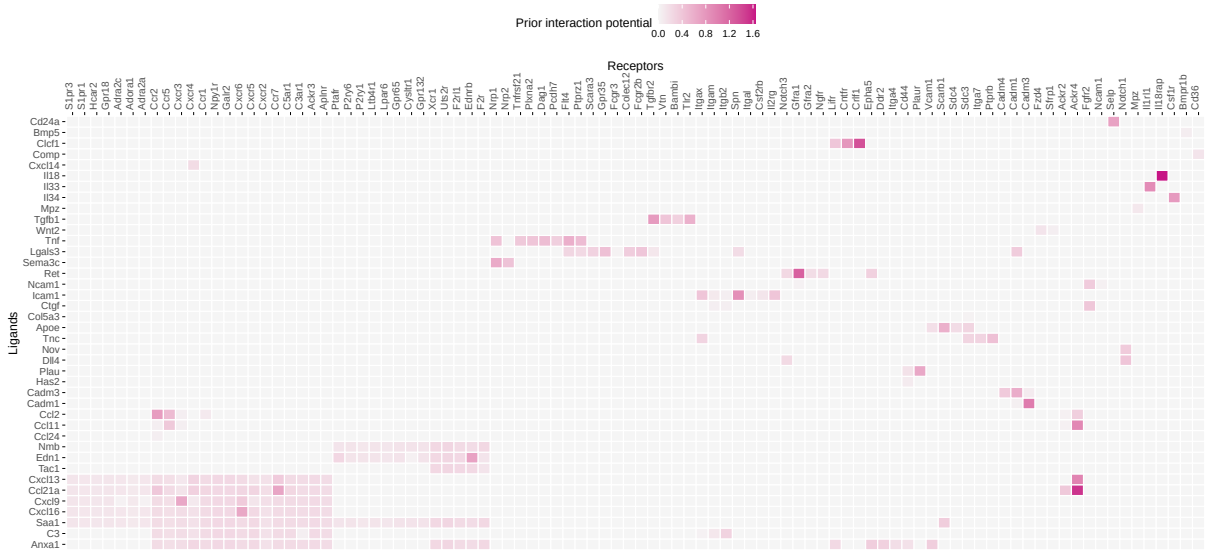


Figure 3.3: Activated FAPs ligands and their predicted targets in the local tissue.

Briefly, the nichenetr toolkit allows : a) to assign ligands to one of the posible sender cell types, being the sender cell the one that most strongly expresses this ligand (expression higher than average), or b) to assign ligands to the group of 'generally' expressed ligands, it is, ligands having expression less or equal the average expression (https://github.com/saeyslab/nichenetr/issues/5). The heatmap shows expressed ligands in FAPs Activated (Pi16+) population and their predicted targets (Receptors) in the local tissue. For each pair, a weight score is calculated based on the level of evidence that this interaction has in curated databases, so the higher the score, the more this interaction is supported. Note that as this result may be different depending on the databases update.

Clcf1 (cardiotrophin-like cytokine factor 1) and other cytokines/immune molecules exhibited high scores, however interesting potential ligands with more modest scores were also found, to cite, the complement component 3, and the endothelial related Edn1 and Semaphorine 3c.

3.1.3 Cross-talk between skeletal muscle resident cells

A graphical representation of the interactions between resident cells was build with the aid of nichenetr. Figure 4. synthetizes ligand-receptor pairs exhibiting the highest weight scores and suggests an active communication between Activated FAPs and Endothelial cells.

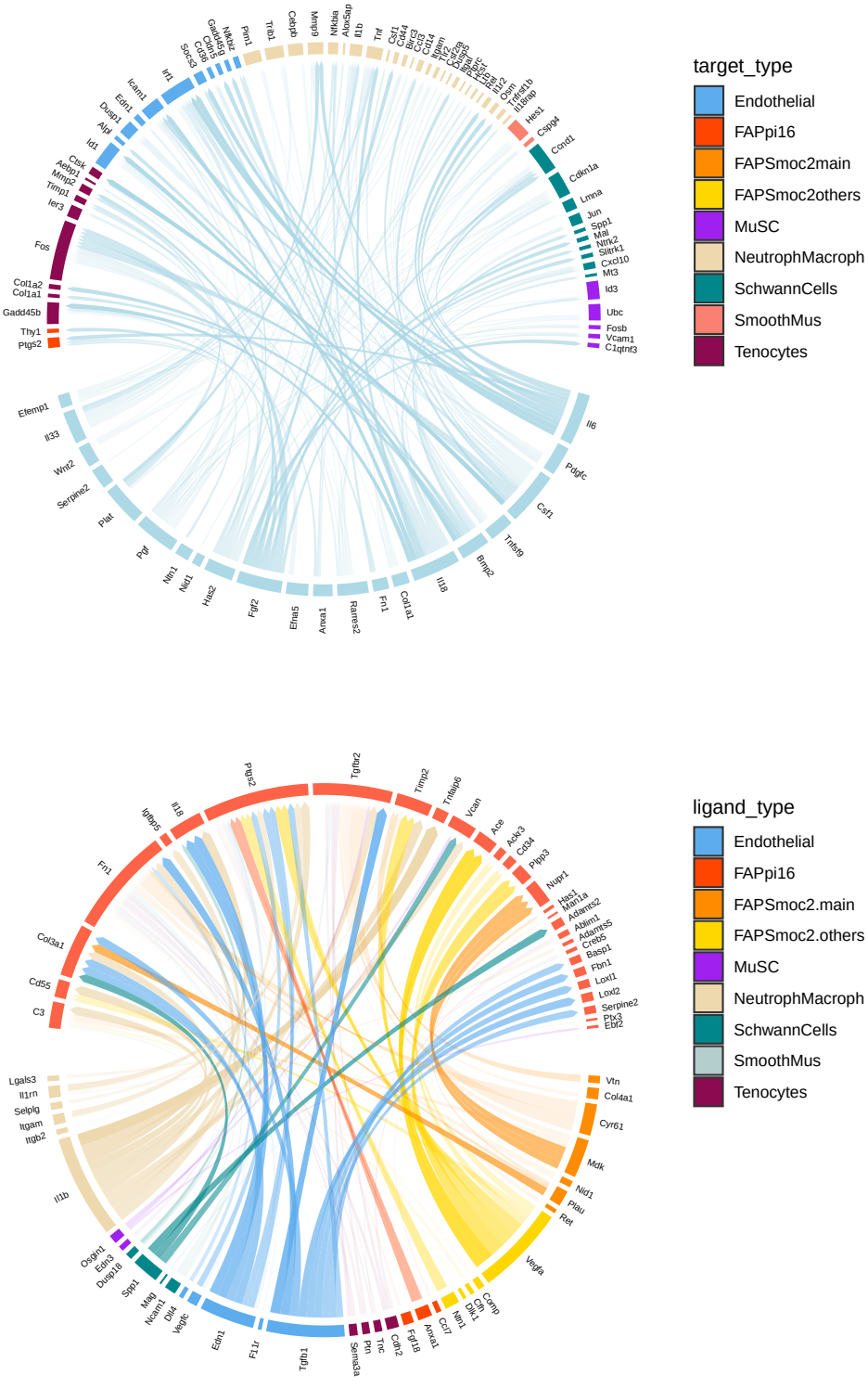


Figure 3.4: 'Circos' plot. Upper panel: FAPs Pi16+ (bottom, lightblue) ligands. Receptors colors as shown in legend (including P16+ itself). Lower panel: The inverse scenario.

3.1.4 GO enrichment

GO enrichement analysis performed by means of ClusterProfiler toolkit [11] on exclusively Smoc2+ or Pi16+ populations yielded biological functions expected to be associated with these cell types, i.e extracellular matrix organization. Interestingly, Pi16+ populations expression profile matched with processes related to cell differentiation and metabolism (see Supplementary Figures in Appendix A). Other results can be seen in GitHub repo 'LigRecPairs_GO/'.

3.2 After Injury

Oprescu full raw counts matrix was subjected to quality control (as explained in Methods), then generated tables could be integrated to a raw Seurat object. In this way adequate filtering was possible before subsequent clustering and Differential Expression (DE) analysis. Full resulting markers tables and high resolution figures are available at `https://github.com/LeGrand-Lab/INMG_SingleCell/tree/master/results/OprescuTimePoints`. Apart from the identification of cell populations (Figure 3.5), the level of expression of different genes of interest was visualized in violin-plots (see repo link) across the different timepoints and specifically across sub-populations. This allowed to confirm that Bmp4 gene expresses in MuSC renewing cells during acute injury, but in homeostasis and repaired muscle it is expressed almost exclusively by Activated(Asc) and Quiescent (Qsc) Satellite Cells.

3.2.1 Pseudotime analysis: proxy to cell fate determination

As seen in figure 3.5 (bottom) in the built trajectories of MuSC and myonuclei (SM = skeletal muscle, mature) it is possible to observe the quiescent (Qsc) and Activated (Asc) satellite cells at a pseudotime zero, that 'evolve' towards two different branches: one to fully mature myonuclei, the other to self renewing (MuSCrenew) satellite cells. See `https://github.com/LeGrand-Lab/INMG_SingleCell/blob/master/results/OprescuTimePoints/MUSC_markersGremMonoc.pdf` to observe genes expression behaviour across these pseudotemporal trajectories.
For FAPs subpopulations, a preliminar Monocle analysis was performed, showing equally a distribution of cells in branched trajectories that initiates with cells expressing markers Cxcl14 and Dpp4 towards activated FAPs or towards fibroblasts.

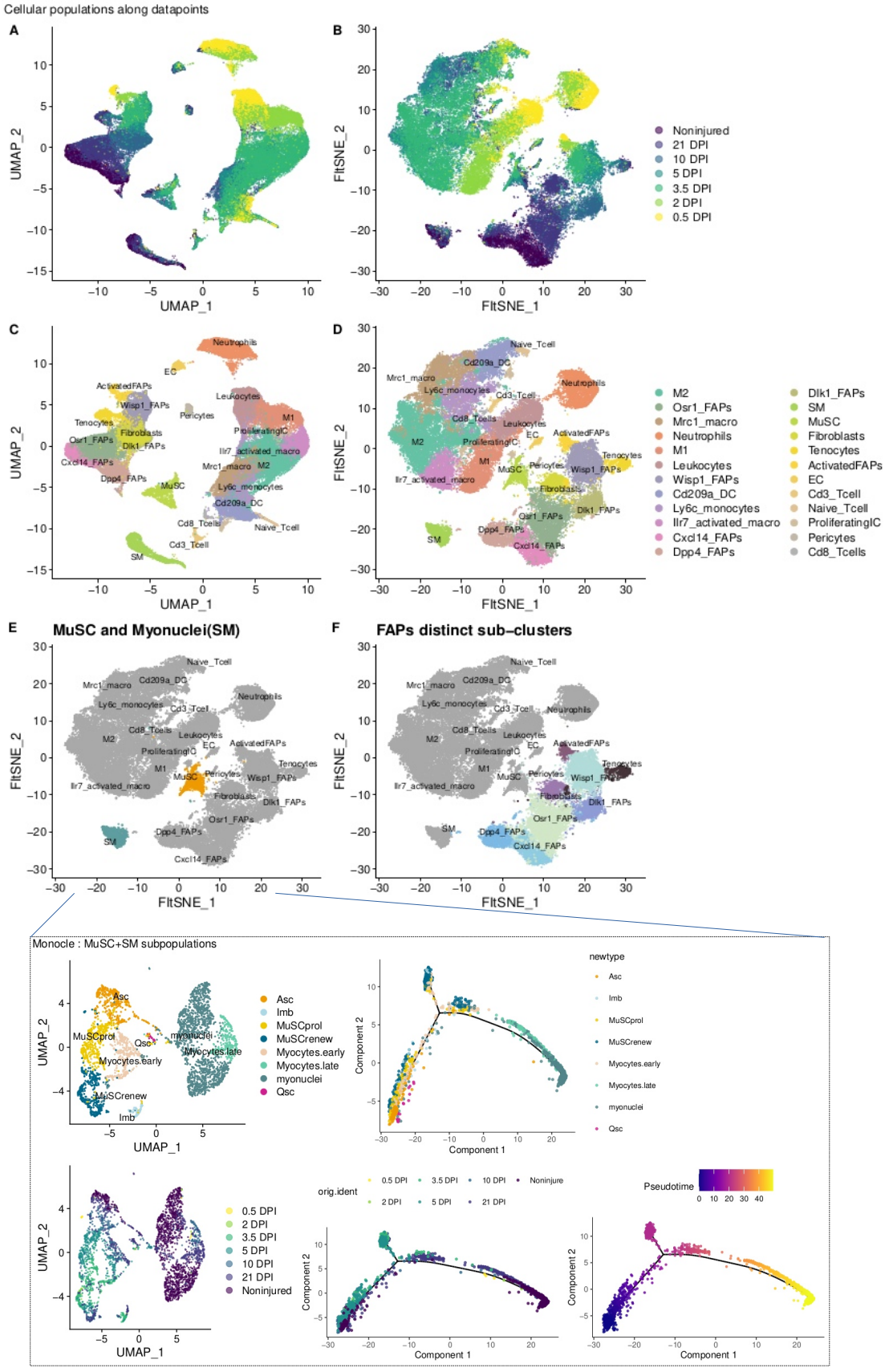


Figure 3.5: Cell Populations across injury datapoints. A. and C. correspond to UMAP algorithm for visual rendering, whereas B. and D. correspond to FItSNE. Note one advantage of FIt-SNE visualization given by a minor degree of cells (points) overlapping. E. and F. enphasize MuSC/SM and FAPs populations, respectively. Zoom on E. allows to identify subpopulations that where modeled into pseudotemporal trajectories.

Chapter 4

Discussion

4.1 Discussion

From the theoretical aspect of scRNAseq technology, one must to keep in mind that gene expression in each cell is a "bursty" process, it is, depending on stimulus and/or cell cycle stage, transcriptional activity in the nucleus triggers, and timing across the cells in a same tissue may vary. This cell-to-cell heterogeneity is overlooked in bulk RNA sequencing experiments, where a mean value of expression is obtained; therefore, single-cell whole cDNA sequencing appears as a promising approach to capture part of this heterogeneity, keeping in mind that it is still a static snapshot and finer molecular dynamics must be confirmed by functional analyses.

4.1.1 scRNAseq Data Pretreatment

In all scRNAseq experiments, a considerable investment in terms of time, laboratory resources and costs is made, so extracting trustful information from the data is a must. Careful data filtering is a first step to minimize subsequent artifactual findings and dangerous fake conclusions. A common practice is the exclusion of apparent superior and inferior outliers by establishing arbitrary cutoffs. Nevertheless, in the case of stem cells for example, low transcriptional activity traduces in low UMI counts which is still informative so there is interest on keeping these cells for pseudotemporal trajectories or other kind of constructs.

A bunch of tools are available for scRNAseq QC, related tutorials are to be taken as a general guide: avoid the use of standardized pipelines for all experiments. Pertinent recommendations can be found in the literature, I strongly recommend the paper of Malte Luecken and Fabian Theis [7]. DoubletFinder clearly states that you shouldn't run it on an aggregated dataset as it will produce false artificial doublets. I provide proof of this error here : `https://github.com/LeGrand-Lab/INMG_SingleCell/blob/master/qcdoubl/spli_FINDER/plotsComparisonsFINDER.pdf`, which resumes the results on an integrated dataset versus unified results obtained separately for each experimental batch ('splitted object').

4.1.2 Downstream Analyses

Different tools were applied, on publicly available scRNA-seq data, with the aim to obtain meaningful information about skeletal muscle micro-environment.

Clustering and differential expression analysis are automated by Seurat internal calling of dedicated algorithms (KNN, Louvain, and Wilcoxon rank sum test). For cell type imputation to each cluster, automated or semi-automated methods have been developed, methods that rely on databases such as the TabulaMuris. However not yet discovered/reported sub-populations may pose a challenge. Manual annotation (with the help of an experienced biologist) is often needed and even recommended when rare populations are expected.

Computational analysis of skeletal muscle in homeostasis suggest a relationship between FAPs and endothelial cells that merits further investigation. Sema3c was predicted to interact with Neuropilins 1 and 2. This interaction has not yet been proved experimentally in skeletal muscle, only a Sema3c

enrichement in activated FAPs on scRNA-seq analyses [9, 8] . In cancer cells and during developement, a "repulsive" role of Sema3c and other semaphorins has been reported: this guidance helps newly formed vessels to adopt a normal distribution [10]. In contraposition, semaphorin 7 is associated to pathological vessel distribution [5] as it seems to exhibit an "attractive" effect on newly formed vessels, favoring pathological vascular anastomoses.

During induced injury and subsequent repair timepoints, different patterns of expression were observed, as was previously reported. Nevertheless, the number of viable cells analysed at acute injury was by far inferior compared to the rest of timepoints (data not shown, but already published), which makes the observations less trustful. Despite this aspect, the general landscape of cell populations dynamics was coherent with the biological knowledge about the skeletal muscle repair process.

Notably, for some of the expressed genes associated to Gremlin signaling in MuSC/SC clusters, there is both spatial and temporal specificity. It was the case for example for Bmp4.

4.2 Conclusion

Predicting relevant markers is a very delicate task, that demands a deep understanding of the advantages and weaknesses of each employed tool. This analysis allowed to gain a further insight into skeletal muscle resident cell populations, and motivates the formulation of new biological questions.

4.3 Perspectives

For FAPs subpopulations, a preliminar Monocle analysis was performed, but as new correctly filtered seurat object is available, a confirmation would be ideal.

It is now an objective to test Semaphorin 3C by means of functional assays, to corroborate the validity of the predictions made about FAPs. On the other hand, a new scRNA-seq experiment is already in course, this time under ageing and pathological conditions. Apart from performing clustering, DE analysis and pseudotemporal trajectories, the analysis of RNA velocity can be interesting, but it demands big computational ressources. Finally **Regulatory gene network inference** approach by novel methods i.e the Python implementation SCENIC [1] can be a theme to develop in a future internship.

**This document is an overleaf project (in LaTeX). Members of LeGrand team are welcome to contribute, go to: https://www.overleaf.com/8589277959jqbrkddbkfnp

Chapter 5

Acknowledgments

To Dr. Fabien Le Grand for giving me the opportunity to work on this exciting subject and for his enormous help in understanding the biological meaning of computational representations. To all his team, which was very supportive.

Thanks also to to Dr L Modolo and Dr Gandrillon, professors at UCBL, for the rapid response to all my questions.

This work was made in the context of a voluntary internship, Master Bioinformatique Moléculaire: Méthodes et Analyses , Université Lyon1 , under direct supervision from Dr. LE GRAND.

Bibliography

[1] Sara Aibar et al. "SCENIC: Single-cell regulatory network inference and clustering". In: *Nature Methods* (2017). ISSN: 15487105. DOI: 10.1038/nmeth.4463.

[2] Robin Browaeys, Wouter Saelens, and Yvan Saeys. "NicheNet: modeling intercellular communication by linking ligands to target genes". In: *Nature Methods* 17.2 (2020), pp. 159–162. ISSN: 15487105. DOI: 10.1038/s41592-019-0667-5. URL: http://dx.doi.org/10.1038/s41592-019-0667-5.

[3] Andrea J. De Micheli et al. "Single-Cell Analysis of the Muscle Stem Cell Hierarchy Identifies Heterotypic Communication Signals Involved in Skeletal Muscle Regeneration". In: *Cell Reports* 30.10 (2020), 3583–3595.e5. ISSN: 22111247. DOI: 10.1016/j.celrep.2020.02.067. URL: https://doi.org/10.1016/j.celrep.2020.02.067.

[4] Stefania Dell'Orso et al. "Single cell analysis of adult mouse skeletal muscle stem cells in homeostatic and regenerative conditions". In: *Development (Cambridge)* 146.12 (2019). ISSN: 14779129. DOI: 10.1242/dev.174177.

[5] Mélanie Franco et al. "Semaphorin 7A: A novel marker of disease activity in Gaucher disease". In: *American Journal of Hematology* (2020). ISSN: 10968652. DOI: 10.1002/ajh.25744.

[6] Lorenzo Giordani et al. "High-Dimensional Single-Cell Cartography Reveals Novel Skeletal Muscle-Resident Cell Populations". In: *Molecular Cell* 74.3 (2019), 609–621.e6. ISSN: 10974164. DOI: 10.1016/j.molcel.2019.02.026.

[7] Malte D Luecken and Fabian J Theis. "Current best practices in single-cell RNA-seq analysis: a tutorial". In: *Molecular Systems Biology* 15.6 (2019). ISSN: 1744-4292. DOI: 10.15252/msb.20188746.

[8] Stephanie N. Oprescu et al. "Temporal Dynamics and Heterogeneity of Cell Populations during Skeletal Muscle Regeneration". In: *iScience* 23.4 (2020), p. 100993. ISSN: 25890042. DOI: 10.1016/j.isci.2020.100993. URL: https://doi.org/10.1016/j.isci.2020.100993.

[9] R. Wilder Scott et al. "Hic1 Defines Quiescent Mesenchymal Progenitor Subpopulations with Distinct Functions and Fates in Skeletal Muscle Regeneration". In: *Cell Stem Cell* 25.6 (2019), 797–813.e9. ISSN: 18759777. DOI: 10.1016/j.stem.2019.11.004. URL: https://doi.org/10.1016/j.stem.2019.11.004.

[10] Guido Serini and Luca Tamagnone. "Bad vessels beware! Semaphorins will sort you out!" In: *EMBO Molecular Medicine* (2015). ISSN: 1757-4676. DOI: 10.15252/emmm.201505551.

[11] Guangchuang Yu et al. "ClusterProfiler: An R package for comparing biological themes among gene clusters". In: *OMICS A Journal of Integrative Biology* (2012). ISSN: 15362310. DOI: 10.1089/omi.2011.0118.

Appendix A

Technical information

A.1 Notes

In the case of Oprescu dataset, do not use the "expression matrix" as it is already normalized, instead do:
`$ wget https://ftp.ncbi.nlm.nih.gov/geo/series/GSE138nnn/GSE1388ppl`
`/GSE138826%5Fregen%5Fdata%2Erds%2Egz`
which downloads an AnnotatedDataFrame object. Open it in R, extract and save the desired matrix as a .txt file.

A.2 Scripts Diagrams

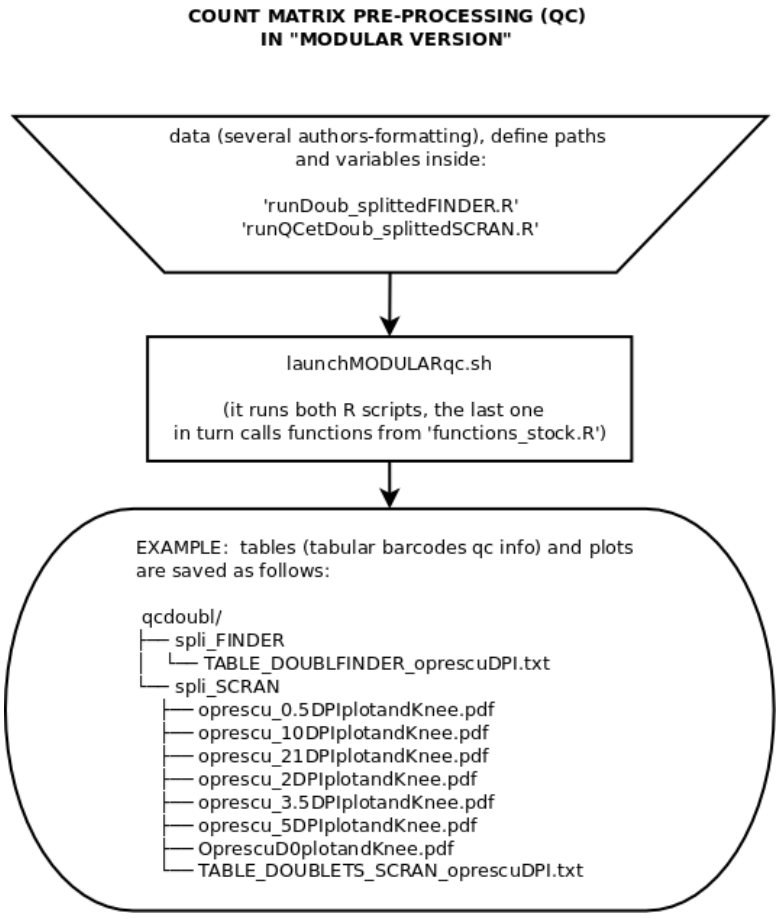


Figure A.1: .

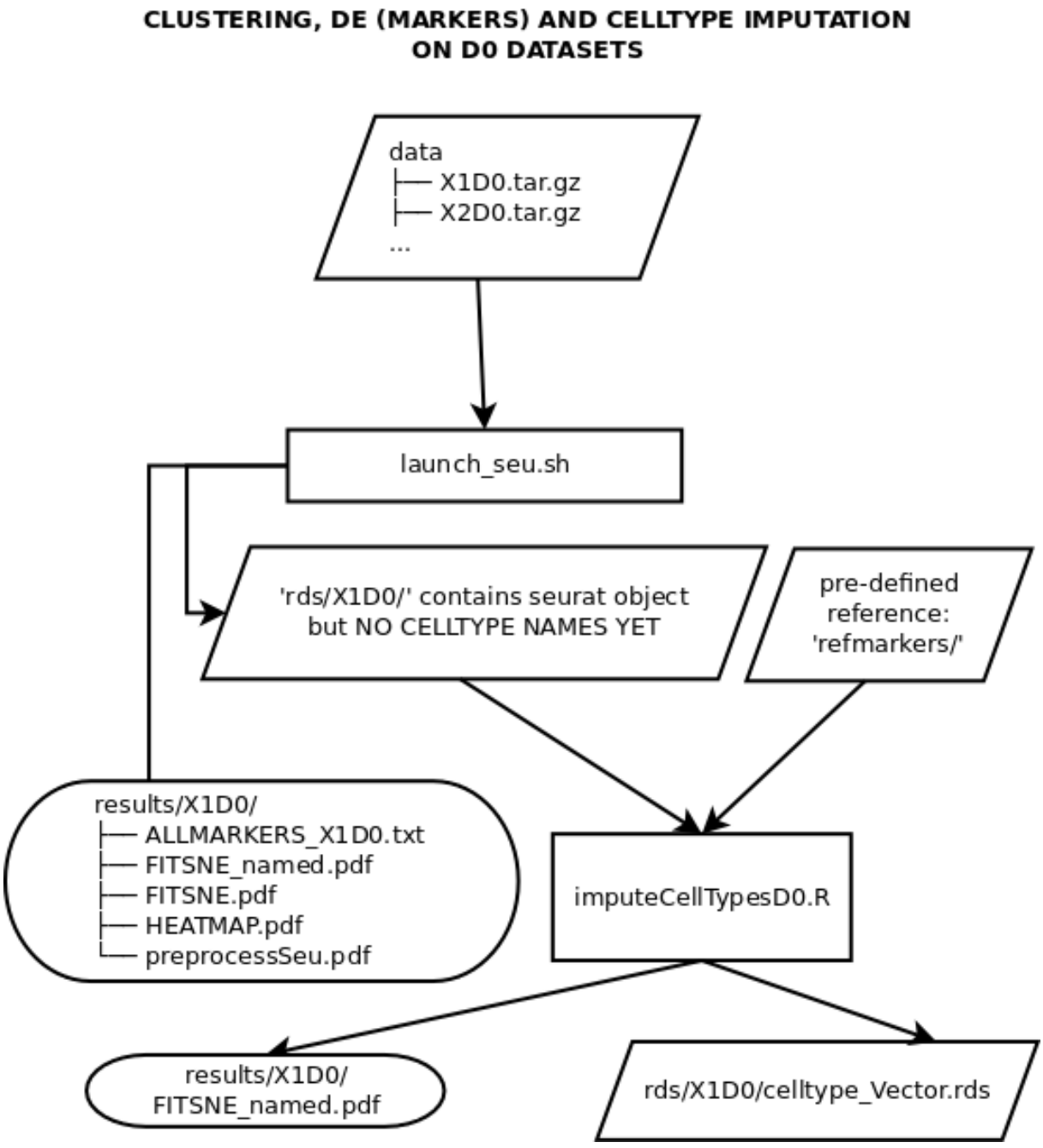

CLUSTERING, DE (MARKERS) AND CELLTYPE IMPUTATION
ON D0 DATASETS
data
├── X1D0.tar.gz
├── X2D0.tar.gz
...
launch_seu.sh
'rds/X1D0/' contains seurat object
but NO CELLTYPE NAMES YET
pre-defined
reference:
'refmarkers/'
results/X1D0/
├── ALLMARKERS_X1D0.txt
├── FITSNE_named.pdf
├── FITSNE.pdf
├── HEATMAP.pdf
└── preprocessSeu.pdf
imputeCellTypesD0.R
results/X1D0/
FITSNE_named.pdf
rds/X1D0/celltype_Vector.rds

Figure A.2: .

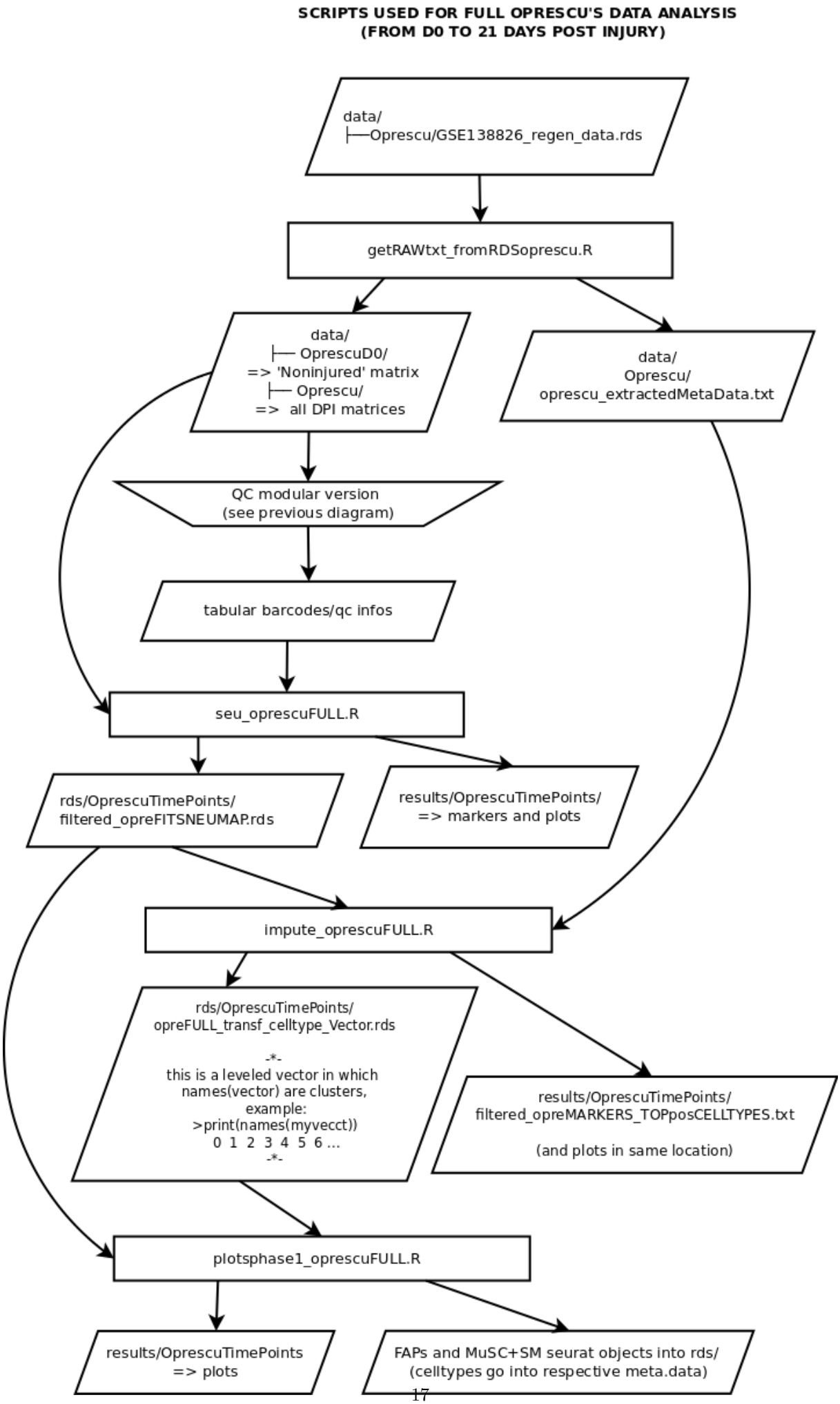

SCRIPTS USED FOR FULL OPRESCU'S DATA ANALYSIS
(FROM D0 TO 21 DAYS POST INJURY)
data/
├──Oprescu/GSE138826_regen_data.rds
getRAWtxt_fromRDSoprescu.R
data/
├── OprescuD0/
=> 'Noninjured' matrix
├── Oprescu/
=> all DPI matrices
data/
Oprescu/
oprescu_extractedMetaData.txt
QC modular version
(see previous diagram)
tabular barcodes/qc infos
seu_oprescuFULL.R
rds/OprescuTimePoints/
filtered_opreFITSNEUMAP.rds
results/OprescuTimePoints/
=> markers and plots
impute_oprescuFULL.R
rds/OprescuTimePoints/
opreFULL_transf_celltype_Vector.rds
-*-
this is a leveled vector in which
names(vector) are clusters,
example:
>print(names(myvecct))
0 1 2 3 4 5 6 ...
-*-
results/OprescuTimePoints/
filtered_opreMARKERS_TOPposCELLTYPES.txt
(and plots in same location)
plotsphase1_oprescuFULL.R
results/OprescuTimePoints
=> plots
FAPs and MuSC+SM seurat objects into rds/
(celltypes go into respective meta.data)

Figure A.3: .

Appendix B

Supplementary figures

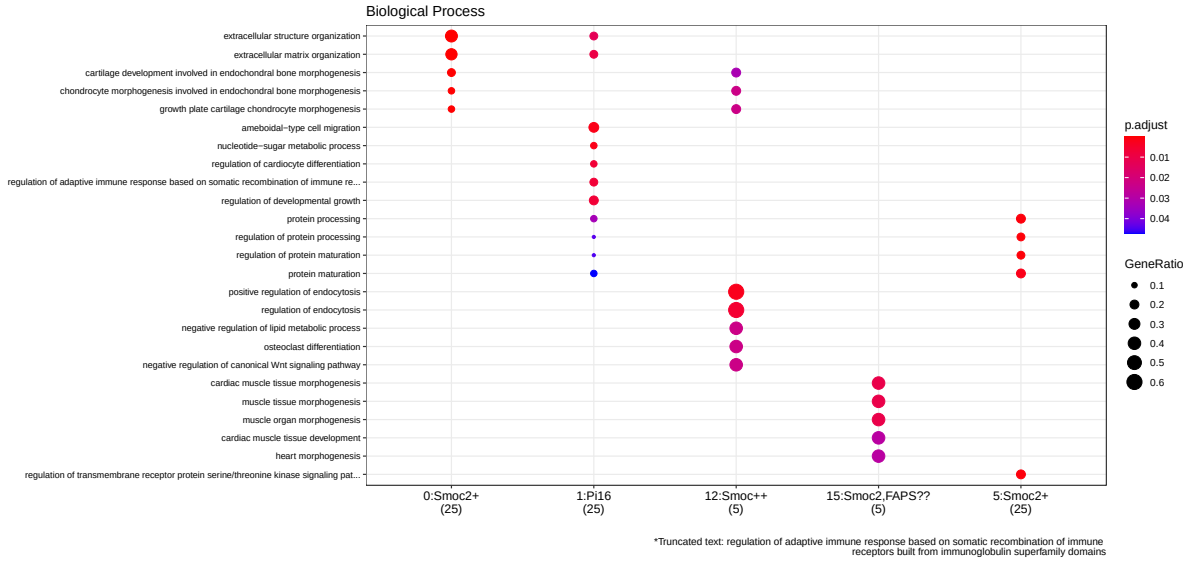


Figure B.1: GO on FAPs