UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Gabriel Baldissera

# ANALYSIS OF THE TRANSCRIPTIONAL AND TRANSLATIONAL PROFILE OF AGRP NEURONS

Porto Alegre
2019

Gabriel Baldissera

# ANALYSIS OF THE TRANSCRIPTIONAL AND TRANSLATIONAL PROFILE OF AGRP NEURONS

Undergraduate thesis presented to the Instituto de Biociências of the Universidade Federal do Rio Grande do Sul as the requirement to obtain a bachelor degree in Biological Sciences.

**Adviser:** Prof. MD. PhD Marcelo O. Dietrich
Comparative Medicine Department of Yale University and Visitor Professor of Departamento de Bioquímica – UFRGS
**Co-adviser:** MSc Delva P. Leão
PhD student of Departamento de Bioquímica – UFRGS

Porto Alegre
2019

**Evaluation comitee:**

Prof. MD. PhD Marcelo O. Dietrich (Adviser)
MSc Delva P. Leão (Co-Adviser)
PhD Fábio Klamt
PhD Kendi Nishino Miyamoto

*I dedicate this project to my cat Gordo for the 11 years of love, friendship and companionship we shared. Thank you.*

## ACKNOWLEDGMENTS

Education is the most powerful tool to shape the world into a better place. I thank all teachers and educators in my trajectory who have provided me the tools to learn how to learn.

We are the product of our life history and experiences. I look to all harms I faced and I recognize that, besides all the suffering, they made me the person I am. I thank the old me for surviving such difficulties with the tools I had available by that time.

We shape our environment and the environment shapes us. I thank all experiences that were proportioned by the biology undergraduate course and my university. I thank all the experience I had abroad. Such opportunities have transformed me into a better person.

We are a social animal. I thank all my friends for being an essential part of all the joy in my life.

## Summary

**ABSTRACT**

Agrp neurons are a small population of neurons in the arcuate nucleus of the hypothalamus known to drive feeding behavior. They are sensitive to hormones such as leptin, ghrelin, and insulin, and play a central role in the regulation of energy homeostasis. Even still, there is a need to understand how Agrp neurons work at the transcriptional and translational levels in scenarios of altered energy balance, e.g. deprivation of food for long periods of time. In this work, we investigated the transcriptome of Agrp neurons using public datasets of single-cell RNA-Seq and the translatome of Agrp neurons using the RiboTag strategy. The latter was coupled with RNA-Seq to obtain a read-out of Agrp neuron-specific mRNAs bounded to the ribosome upon food deprivation. We have identified pathways and biological processes enriched in both levels of regulation as well as shared genes and their function. Our findings suggest that Agrp neurons use distinct strategies to adapt to food deprivation and dynamically respond to a scenario of negative energy balance.

**Key-words**

Next-generation sequencing; Single-cell RNA-Seq; RNA-Seq; Agrp neurons; Food-deprivaion; Transcriptome; Translatome.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ARC | Arcuate nucleus of the hypothalamus |
| AGRP | Agouti-related peptide |
| Agrp | Agouti-related peptide neurons |
| CB | Cell barcode |
| cDNA | Complementary DNA |
| CPM | Counts per million |
| DE | Differential expression |
| ER | Endoplasmic Reticulum |
| FD | Food deprivation |
| FDR | False Discovery Rate |
| GEO | Gene Expression Omnibus |
| HA | Hemagglutinin |
| IP | Immunoprecipitated |
| LFC | Log Fold Change |
| MNN | Mutual nearest neighbors |
| mRNA | Messenger RNA |
| NGS | Next-generation sequencing |
| NPY | Neuropeptide Y |
| PCA | Principal component analysis |
| POMC | Pro-opiomelanocortin neurons |
| RNA-Seq | RNA sequencing |
| sc-RNA-Seq | Single-cell RNA sequencing |
| t-SNE | t-distributed stochastic neighbor embedding |
| UMI | Unique molecular identifiers |
| α-MSH | α-Melanocyte-stimulating hormone |

# INTRODUCTION

## 1. Agrp neurons and feeding behavior

The arcuate nucleus of the hypothalamus (ARC) is a region adjacent to the median eminence and the third ventricle, which integrates metabolic and physiological signals and is relevant for the control of energy homeostasis. Such integration can be achieved by particular attributes of this area, like the fenestrated blood-brain-barrier, which facilitates the contact with molecules circulating in the blood (Ciofi, 2011).

The ARC contains two neuronal populations that are particularly important: Pro-opiomelanocortin (POMC) neurons and agouti-related peptide (AGRP) neurons (Chronwall, 1985), herein Agrp neurons, which form the hypothalamic melanocortin system. The melanocortin system is a central pathway for the energy balance and regulation of feeding behavior. POMC neurons secret α-Melanocyte-stimulating hormone (α-MSH), which binds to melanocortin receptors (MC3R) in the Agrp and leads to decreased food intake. During a negative energy balance, e.g. periods of starvation, Agrp neurons are active and secrete AGRP and NPY. They partially block the binding and effect of α-MSH, leading to an increased food intake (Cansell et al., 2012). Additional studies showed that activation of Agrp neurons either by optogenetics (Aponte et al., 2011) or chemogenetics (Krashes et al., 2011) evokes voracious feeding within minutes. Likewise, acute ablation of Agrp neurons in adult animals is enough to stop feeding, leading to starvation (Gropp et al., 2005; Luquet et al., 2005).

Mechanisms for Agrp activation based on peripheral neuroendocrine hormones have been described as well. In a negative energy balance, neuroendocrine responses lead to an inhibition in the insulin secretion, to a reduction of the leptin levels and an increase in the ghrelin levels (Schwartz and Seeley, 1997). Agrp neurons have been found to express receptors for such hormones and therefore are responsive to their levels (Ashford et al., 1997; Havrankova et al., 1978; Van Den Top et al., 2004; Willesen et al., 1999). Leptin and insulin suppress Agrp neuronal

activity (Mizuno, 2004). Upon food deprivation (FD), levels of leptin and insulin are low, allowing Agrp activation and therefore, promotion of feeding behavior. On this context, high ghrelin levels also contribute to activation of Agrp neuronal activity and stimulation of AGRP and NPY secretion (Cowley et al., 2003). The reestablishment of the energy balance leads to an increase in leptin and insulin levels and decrease in ghrelin level, inhibiting again Agrp neuronal activity and, consequently, feeding behavior (Schwartz and Seeley, 1997).

## 2. Studying Agrp neurons in the genomics era

Many of the findings that placed Agrp neurons as key regulators of energy balance are being corroborated and further detailed due to advances in sequencing techniques, such as RNA sequencing (RNA-Seq) (Henry et al., 2015; Wang et al., 2009) and single-cell sequencing (Macosko et al., 2015; Ofengeim et al., 2017). Transcriptomic analysis of single or populations of neurons enabled the dissection of molecular mechanisms important for their specialized functions. For the first time, it was possible to access the expression profile of a neuron upon some intervention (diet, age or light cycle), with greater precision to estimate the expression of active genes in a given moment (Wang et al., 2009) or to analyze messenger RNA (mRNA) transcripts from thousands of individual cells simultaneously. These types of studies can reveal new cell-types and tissue diversity at the transcription level (Macosko et al., 2015).

Recently, the Drop-Seq strategy (a protocol widely used to analyze the transcriptome of single cells in a given tissue) (Macosko et al., 2015) was applied to the ARC, providing researchers with data on Agrp neuronal transcription at the single cell level, thus shedding light into the molecular dynamics of Agrp neurons upon food deprivation (Campbell et al., 2017). This study provided the first evidence on the existence of transcriptomic subtypes of Agrp neurons based on the presence of cells co-expressing both somatostatin (*Sst*) and *Agrp* gene markers, as well as cells expressing solely *Agrp* across the ARC (Campbell et al., 2017).

In addition to neuronal subtypes, molecular data from next-generation sequencing analysis also illuminated the pathways and genes involved in Agrp function upon food deprivation (Henry et al., 2015). However, there is still a need to further investigate the molecular machinery used by Agrp neurons upon food deprivation, not only at the transcriptional level but also at the translational level. It is known that post-transcriptional events regulate the activity of several cell types (Hossain et al., 2016). Several mechanisms of post-transcription regulation have been described, such as microRNA interference activity and polyadenylation (Schaefke et al., 2018). Thus, changes in gene expression at the transcriptional level do not necessarily imply changes at the translational level (Tebaldi et al., 2012).

## 3. Transcriptome and Translatome upon food deprivation

Changes in the environment and feeding conditions lead to intracellular signaling through energy sensors and transcription factors, altering the transcript expression profile of a given cell, i.e, the transcriptome (López-Maury et al., 2008). The expression profile is dynamic and important for the cell to functionally adapt to changes in the environment and physiological conditions (McGettigan, 2013). The transcripts, mRNAs, are key components in the translation process, where proteins are formed (McHale et al., 2013). From the entirety of mRNAs in a cell in a given condition, the ones associated with ribosomes for protein synthesis constitute the translatome (King and Gerber, 2016). A variety of methods have been developed to access the translatome of a cell or cell population. One of these methods is the ribosome affinity-purification. This method allows the study of gene expression of specific cell types without major contamination from surrounding cells. For that purpose, genetically modified animals expressing affinity-tagged ribosomal proteins that compose the large ribosomal subunit (60S) are used. After the tissue collection, these tagged ribosomes can be selected with specific antibodies or ligands and their associated mRNAs can be quantitatively measured with RNA-Seq (King and Gerber, 2016). The analysis of the transcriptome and translatome of specific cell types allows

the identification of gene expression patterns and the better understanding of important molecular pathways and biological processes.

## 4. NGS and the analysis of transcriptomic data

In the past decade, the classic Sanger sequencing has been substituted by newer sequencing technologies, called next-generation sequencing (NGS) (Heather and Chain, 2016). NGS comprises a set of strategies and methods that allows higher throughput of data. These new sequencing platforms allow the sequencing of a whole genome or transcriptome (Marguerat and Bähler, 2010) in an unprecedented time scale with reduced costs (Lister et al., 2009). Briefly, prior to sequencing, samples are submitted to library construction where the input DNA or cDNA is fragmented, attached to primers called adapters and amplified by polymerase chain reaction (PCR). The sequencing itself occurs in millions of parallels reactions that result in short length fragment called reads. The reads are then evaluated by quality, aligned to a reference genome, and submitted to downstream analysis. The throughput of data in the order of gigabases revolutionized bioinformatics and brought big data into biology (Metzker, 2010).

The study of RNA biology and gene expression in large scale, called RNA-Seq, has become one of the most popular NGS approaches to interrogate genes expressed in a given context (Mortazavi et al., 2008). Specifically, the development of differential expression (DE) tests allowed researches to evaluate the effect of different interventions, e.g. diet, in the gene expression profile having a profound impact on the way we understand biology (Huang et al., 2009; Zhang et al., 2014). The latest advances in NGS have also brought sequencing at the individual cell level to reality. Several single-cell sequencing protocols have been created to investigate the transcriptome of thousands of individual cells simultaneously (Tang et al., 2009). In combination, such technologies have revolutionized the exploration of cell diversity and gene expression dynamics (Ofengeim et al., 2017). This dissertation leverages theses powerful techniques to study Agrp neurons in the mouse brain.

## JUSTIFICATION

Despite the knowledge gathered on the intracellular functioning of Agrp neurons, there is a need to understand how post-transcriptional regulation impacts the transcriptome of these neurons upon food deprivation. Our hypothesis is that such regulation is important for distinct levels of regulation of Agrp neurons. Studying the translatome of these neurons can help us to better understand the activity of Agrp neurons during periods of food deprivation. Understanding the differences between the transcriptome and the translatome of Agrp neurons during such feeding conditions can also be valuable for the scientific community, having in mind that this feeding circuit plays a central in energy homeostasis and in obesity.

## AIMS

### 1. General Aim:

To analyze the transcriptome and the translatome of Agrp neurons upon food deprivation.

### 2. Specific Aims:

**2.1** To evaluate changes in genes expressed in the transcriptome and translatome of Agrp neurons during food deprivation.

**2.2** To characterize pathways relevant for Agrp neurons during food deprivation at the transcriptional and translational levels.

**2.3** To identify shared genes in the transcriptome and translatome of Agrp neurons during food deprivation.

## MATERIALS AND METHODS

### 1 Data obtainment

For this study, public single-cell RNA sequencing (sc-RNA-Seq) data were used to investigate the transcriptome of single Agrp neurons. Two sc-RNA-Seq datasets were chosen, both available on the NCBI Gene Expression Omnibus (GEO) database. We used data generated in our laboratory to characterize the translatome of Agrp neurons. A summary of all samples used in this study is available in the supplemental material (**Table S1**).

### 1.1 Transcriptome

The first public dataset was provided by Campbell (Campbell et al., 2017), identified on GEO by the tag GSE93374. Using Drop-Seq, this study analyzed 20,921 cells from the hypothalamic arcuate–median eminence complex (Arc-ME). To do so, 53 adult (4-12 weeks old) virgin male and female mice with the C57BL6/J genetic background were randomly assigned to the experimental groups. The experimental conditions used were: control (*ad libitum* chow-fed), 1-week low-fat diet (10%), 1-week high-fat diet (60%), food deprivation (24h), food deprivation (24h) + 2h re-feed. The groups were separated over 6 batches (one batch per day), the Arc-ME tissue was microdissected, pooled by experimental condition and dissociated with a papain-based protocol, having each batch been sequenced separately. Libraries were sequenced using the Illumina NextSeq500 platform. The read 1 was 20 bp long (bases 1–12 cell barcode (CB), bases 13–20 Unique Molecular Identifier (UMI)) and the read 2 (paired-end) was 60 bp long. We have downloaded the sequencing files in fastq format on the Single Cell Portal (https://portals.broadinstitute.org/single_cell) maintained by the Broad Institute.

The second dataset is comprised by the results of the investigation by Chen (Chen et al., 2017), GSE87544. Using Drop-Seq, they analyzed over 14,000 cells from the arcuate tissue. A total of 7 Young adult female (8 - 10 weeks) B6D2F1 mice (C57B6 female × DBA2 male) were used, 4 *ad libitum* fed control mice and 3 food-

deprived (24h) mice. All the experiments were performed in 5 batches, each animal was processed separately (cell dissociation, library preparation) and regarded as a biological replicate. In order to get the arcuate tissue, coronal sections from Bregma -0.22 to – 3.16 mm were cut and further sliced into 1 mm slices. Hypothalamic tissue was then dissected from each slice and dissociated with a papain-based protocol. Sequencing was performed in the Illumina Hiseq 2500 platform using the standard read construction for Drop-Seq (Macosko et al., 2015). We downloaded the sequencing data in the SRA format, available in the GEO database.

**1.2 Translatome**

In order to obtain the translatome of Agrp neurons, the RiboTag mouse was used (Sanz et al., 2009). This mouse line carries a *Rpl22* allele with a floxed wild-type C-terminal exon followed by an identical C-terminal exon with three copies of a hemagglutinin (HA) epitope inserted before the stop codon. When the RiboTag mouse is crossed to a cell-type-specific Cre recombinase expressing mouse, the first exon copy is excised and the ribosomal protein $RPL22^{HA}$ tagged with the HA epitope is incorporated into actively translating polyribosomes. Using the RiboTag mouse and an Agrp-Cre mouse, we can recover transcripts of Agrp neurons that are being translated by active polyribosomes.

A total of 13 samples were used in this study, with 6 being control samples of *ad libitum* diet and 7 food deprived samples (food deprivation for 16h). All animals were 40-45 days old. The animals were kept in boxes (48x26cm), in a room with stable temperature (22±1$^{o}$C) and 12h light/dark cycle. Briefly, the ARC was dissected and homogenized in 1 mL buffer solution followed by the immunoprecipitation protocol utilizing 2 mL of anti-HA antibodies and 4-hour incubation. Then, 200 μL of magnetic beads were added to the mixture for a 2-hour incubation. The mixture was washed 3 times in a high-salinity solution, using a magnetic plate. The bead-antibody relation was dismantled and the supernatant was collected for mRNA extraction utilizing the RNeasy Micro Plus kit (Qiagen). Before the immunoprecipitation protocol, a 25 μL aliquot of the whole ARC tissue was taken as an immunoprecipitation control and called “Input sample”. The immunoprecipitated samples were called “IP

samples". Both samples were sequenced in the Yale Center for Genome Analysis (YCGA) in the Illumina HiSeq 2500 platform, with a length of 75bp and 40 million reads per sample in a single-end design.

## 2 Transcriptome data analysis

### 2.1 Single-cell data pre-processing

#### 2.1.1 Fastq file extraction

We used the *fastq-dump* function from SRA Toolkit (available at https://www.ncbi.nlm.nih.gov/sra/docs/) to convert SRA to fastq files.

#### 2.1.2 Sequencing quality analysis

We used the FastQC software (Version 0.11.7) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess the sequencing quality data prior to any data processing.

#### 2.1.3 Read editing

In order to start the data processing, we generated a list of accepted cell barcodes (CB) using all barcodes available in the gene expression tables available as supplemental material for both studies. Next, we extracted the CBs+Unique Molecular Identifiers (UMI) from the read 1 files and added them to the read names of the read 2 files using the *Umi-tools extract* function (Version 0.5.5) (Smith et al., 2017). We also filtered out reads which CB did not match with the accepted CB list created in the previous step. To do so, we used the argument *--filter-cell-barcode* and provided the CB list to the *--whitelist* argument. The output of this step was a read 2 file with the actual gene sequence and the CB+UMI sequence at the read names. This file was ready to be mapped.

## 2.2 Single-cell data processing

### 2.2.1 Reference genome

We used the *Mus musculus* genome assembly GRCm38.p6. The reference FASTA and GTF files were obtained from Ensembl release 95, available in the repository (ftp://ftp.ensembl.org/pub/release-95/). The FASTA file contains the genome reference sequences without alternate contigs while the GTF file contains the annotated transcripts in the GTF format.

### 2.2.2 Genome indexing

For the purpose of optimizing the alignment processing, we built an indexed reference genome. To do so, we used the STAR tool *(*Version 2.6.0*)* (Dobin et al., 2013) with the *--runMode genomeGenerate* argument, setting *--sjdbOverhang* to 59.

### 2.2.3 Read alignment

We aligned the files using the STAR (Version 2.6.0) software using standard parameters. We did not allow for multi-mapping reads. The output of this step was a file in the BAM format (Li et al., 2009).

### 2.2.4 Alignment quality

We evaluated the quality of the read alignment as a checkpoint step, using the Piccard tool (Version 2.18.17) (http://broadinstitute.github.io/picard). We used the *CollectAlignmentSummaryMetrics* argument, which takes a BAM file input and produces a summary of alignment metrics detailing the quality of the read alignments.

### 2.2.5 Assignment of reads to genes

After aligning the reads to the genome, it was necessary to identify the gene identity of each read. We used the featuresCounts tool (Version 1.6.3) (Liao et al., 2014) from the Subread package to assign aligned reads to genes. We sorted the output BAM file and indexed it using Samtools *(*Version 1.6*)* (Li et al., 2009).

### 2.2.6 Estimative of gene expression

After all the previous steps, we were able to obtain the number of distinct UMIs. To do that, we used the Umi-tools count function (Smith et al., 2017). We set the arguments *--per-gene* and *--per-cell* to get the number of distinct UMIs in each gene and in each cell, respectively. We used the gene tag and assignment status tag provided by featureCounts to count the number of UMI assigned to each gene. The output of this step was a count table with the number of UMIs per gene (rows) in each cell (columns).

### 2.2.7 Additional quality controls

#### I. Cell quality control

After generating the UMI count lists for both datasets, we moved to the R platform, where we performed additional quality controls and further data analysis. We performed two quality controls at the cell level: the gene counts and library size. We filtered cells by the expressed gene count. As a general threshold, we removed cells with total gene count below 800. After the previous filtering, we tested the need to also filter cells by library size (total sum of raw counts across the genes). To do so, we assessed the presence of outliers: cells with log library-size with 3 median absolute deviations (MADs) below the respective library-size median values.

#### II. Cell-cycle effect correction

We assessed the impact of cell phases in our datasets using the Cyclone tool from the Scran package (Version 1.6.9) (Lun et al., 2016) in order to determine whether additional corrections were necessary to compensate for variability in gene expression due to distinct cell phases.

**III. Removal of low-abundance genes**

Low-abundance genes can be as problematic as zero counts since they are not reliable for statistical inferences and also increase the computational demand. Therefore, we removed genes with an average raw count value below a threshold of 0.07323569, which was the lowest average value for the *Agrp* gene (Chen dataset).

**IV. Normalization**

We normalized the data in order to correct for distinct capture efficiency and sequence depth between cells. We used the Scran (Lun et al., 2016) functions *Quickcluster*, *ComputeSumfactors* and *Normalize* to calculate the sum factors to each library and then normalize the data.

**V. Control for confounding factors: MNN correction**

In RNA-Seq data, part of the variability in gene expression can be due to technical factors. Such factors influencing the gene expression are called batch effects and their removal is an essential step in gene expression studies. Since we used two different datasets, each composed by independent biological samples in different diets, the batch correction was performed using both datasets through a method that allows us not only to correct for technical effects but also to make the distinct samples within each dataset comparable. For that purpose, the mnnCorrect function in the Scran package was used and, as a software requirement, the samples were subset to keep only the genes present in all of them. Such function uses the Mutual Nearest Neighbors (MNN) method (Haghverdi et al., 2018) to identify the

shared cell populations between two or more batches and to correct the expression of all cells.

### 2.3 Downstream analysis

#### 2.3.1 Clustering

In order to identify the groups of similar cells, specially cells from the same cells types across the samples in both datasets, we clustered the batch corrected expression values of all cells using the *buildSNNgraph* function of the Scran package (Lun et al., 2016). This method builds a shared nearest-neighbor graph using cells as nodes and the neighbors are identified based on Euclidean distances in their expression profiles.

#### 2.3.2 Dimensionality reduction

To visualize the data, we plotted a t-distributed stochastic neighbor embedding (t-SNE) plot on the MNN-corrected data. This is a technique suitable for reducing high-dimensional data for visualization in the low-dimensional space of two dimensions (VAN DER MAATEN et al., 2008). We colored the cells in the t-SNE according to the groups they were assigned to in the previous step.

#### 2.3.3 Agrp cells identification

In order to identify the Agrp neuron cells present in both datasets, we colored the cells in the t-SNEs according to their corrected expression values for the *Agrp* and *Npy* gene markers, respectively. We then identified the clusters that had MNN-corrected expression values > 0.05 for both gene markers. We selected those clusters based on their position in the t-SNE and on their cluster identity label. We subset the cells in those clusters, obtaining Agrp neuronal cells from both single-cell datasets.

The Campbell's dataset had more cells then Chen's dataset. Therefore, we expected its number of Agrp cells to be higher. We identified cells that had raw counts > 0 for *Agrp* and *Npy* in this dataset. We considered those cells as Agrp and they were merged with the ones subset from the clusters, resulting in a greater Agrp cell dataset. The totatality of Agrp cells identified were subset to keep only their shared genes and then the cells were used for the transcriptome obtainment and downstream analysis.

### 2.3.4 Batch correction for differential expression

In order to test the possibility of correcting for technical variability in the data to perform differential expression, we plotted a principal component analysis (PCA) to observe sample similarity. We observed the amount of variability explained by each variable utilizing the *plotExplanatoryVariables* function (Scater 1.10.1) (McCarthy et al., 2017). We also used the *svaseq* function (Sva 3.30.1) (Leek, 2014) to try to find possibly hidden variables accounting for batch effects.

### 2.3.5 Gene enrichment in food deprivation

We merged all Agrp cells identified previously in an unique SingleCellExperiment R object. We filtered this object in order to remove lowly expressed genes, keeping the ones that had at least 5 counts in at least 5 cells. We then transformed the raw counts into counts per million (CPM) using the *CPM* function from the SingleCellExperiment package (Version 1.4.1) (Lun and Risso, 2019). We calculated the average CPM gene expression in the Agrp cells by feeding condition and, based on it, we calculated the gene enrichment upon food deprivation by dividing the Food Deprived (FD) average count by the Fed average count.

$$FD\,Enrichment = mean(CPM)\,Food\,Deprived / mean(CPM)\,Fed$$

### 2.3.6 Transcriptome definition

We used the results of the gene enrichment described above in order to define the transcriptome. We applied a threshold to select genes that were impacted by food deprivation. We selected genes with enrichment value lower than 0.5 or higher than 1.105. This threshold allowed us to select genes that had a decrease of 50% in their count values upon food deprivation or had an increase equal/higher than the *Npy* marker enrichment which was 1.105. The resulting genes were considered the transcriptome of Agrp neurons upon food deprivation.

#### 2.3.7 Transcriptome functional characterization

We performed gene enrichment analysis utilizing the *Ingenuity Pathway Analysis* (IPA) software (Version v.01-14) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis) to obtain enriched pathways and the String tool (Version 11) (https://string-db.org/), without adding gene interactions, to obtain enriched gene ontology terms (Chen et al., 2013; Kuleshov et al., 2016).

## 3 Translatome data analysis

### 3.1 RiboTag data pre-processing

#### 3.1.1 Sequencing quality analysis

We analyzed the sequencing quality through the *FastQC* software, using the same considerations aforementioned for the single-cell data analysis (item 2.2).

### 3.2 RiboTag data processing

#### 3.2.1 Read alignment

We used the *STAR* software (Version 2.6.0) to perform gene alignment, using a genome index built with the same reference files used for the single-cell data and customized to the RNA-Seq read length (item 3.1).

### 3.2.2 Estimate gene expression

We used the *HTSeq* Software (Version 0.11.1) (Anders et al., 2015) to count the reads aligned to each gene and at this step, a count table was outputted.

### 3.2.3 Assessment of Agrp neuron ribosome immunoprecipitation

In order to verify the success of the Agrp ribosome immunoprecipitation, we analyzed the enrichment of the raw counts for the *Agrp* and *Npy* gene markers in the IP data in relation to the Input data. This analysis worked as a positive control. We analyzed the enrichment of marker genes for other cell types: Astrocyte, endothelial, microglia, neuron and oligodendrocyte. This analysis worked as negative control. The marker genes for such cell types were obtained from *PanglaoDB* database (Franzén et al., 2019).

$$Enrichment\ IP = raw\ counts\ IP / raw\ counts\ Input$$

## 3.3 Downstream analysis

### 3.3.1 Differential expression: Fed vs FD

The *HTSeq* output expression table was used as input in R, to perform gene differential expression using the *DESeq2* package (Version 3.8) (Love et al., 2014). Our design formula allowed us to compare food deprivation against fed condition.

$$Design = Seq.Round + Condition$$

### 3.3.2 Translatome definition

We used the results of the differential expression and identified the genes that were more impacted by the food deprivation based on the False Discovery Rate (FDR) and Log2 Fold Change (LFC). We selected only the genes that had an FDR < 0.05 and LFC > 0.5 or LFC < -0.5. The resulting genes were considered as the translatome of the Agrp neurons upon food deprivation.

### 3.3.3 Translatome functional characterization

We characterized the translatome using the same approach we used in the transcriptome: analyzing the enriched pathways and Gene Ontology terms via the *IPA* software and *String* tool, respectively.

## 4 Transcriptome vs translatome

We compared the transcriptome of food-deprived Agrp neurons against the translatome of food-deprived Agrp neurons in order to verify the gene similarity between both sets**.** We described each gene shared by the transcriptome and translatome using the *String* database. We also analyzed their related GO terms and pathways using *String* and *IPA* tools respectively.

# RESULTS

## 1 Transcriptome results

### 1.1 Single-cell data pre-processing results

#### 1.1.1 Sequencing quality analysis

The sequencing quality analyses demonstrated that both datasets have good quality and therefore no additional trimming was necessary, as we can see in Figure 1.

### 1.2 Single-cell data processing results

#### 1.2.1 Read alignment

The *STAR* read alignment results were satisfactory for both datasets as we can see in the amount of uniquely mapped reads in relation to the total amount of reads (**Figure 2**). We checked the alignment quality using the *Piccard* tool metrics and they were satisfactory for both datasets as well.

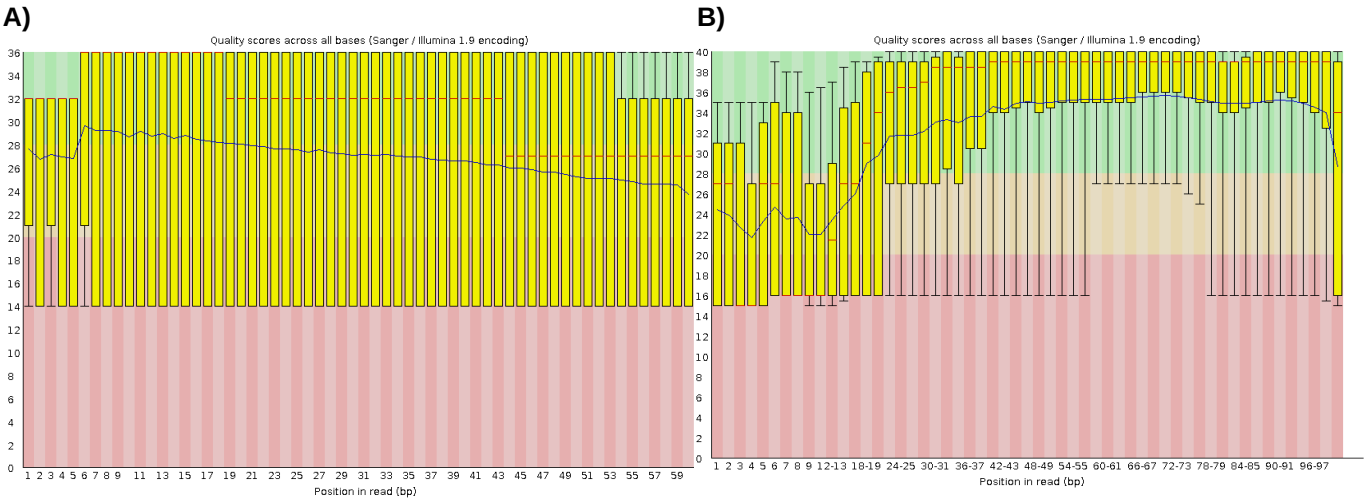
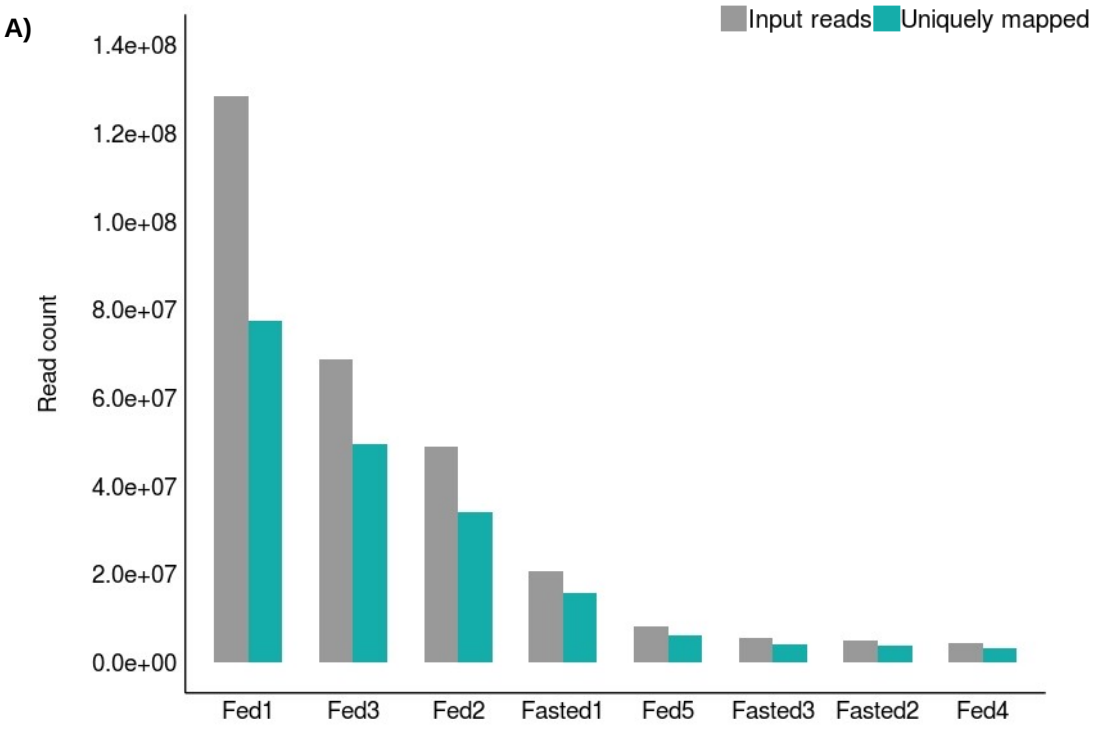



**Figure 1: A)** Per base sequence quality of the Fed 1 sample (Campbell dataset) **B)** Per base sequence quality of Fasted 4 sample (Chen dataset).



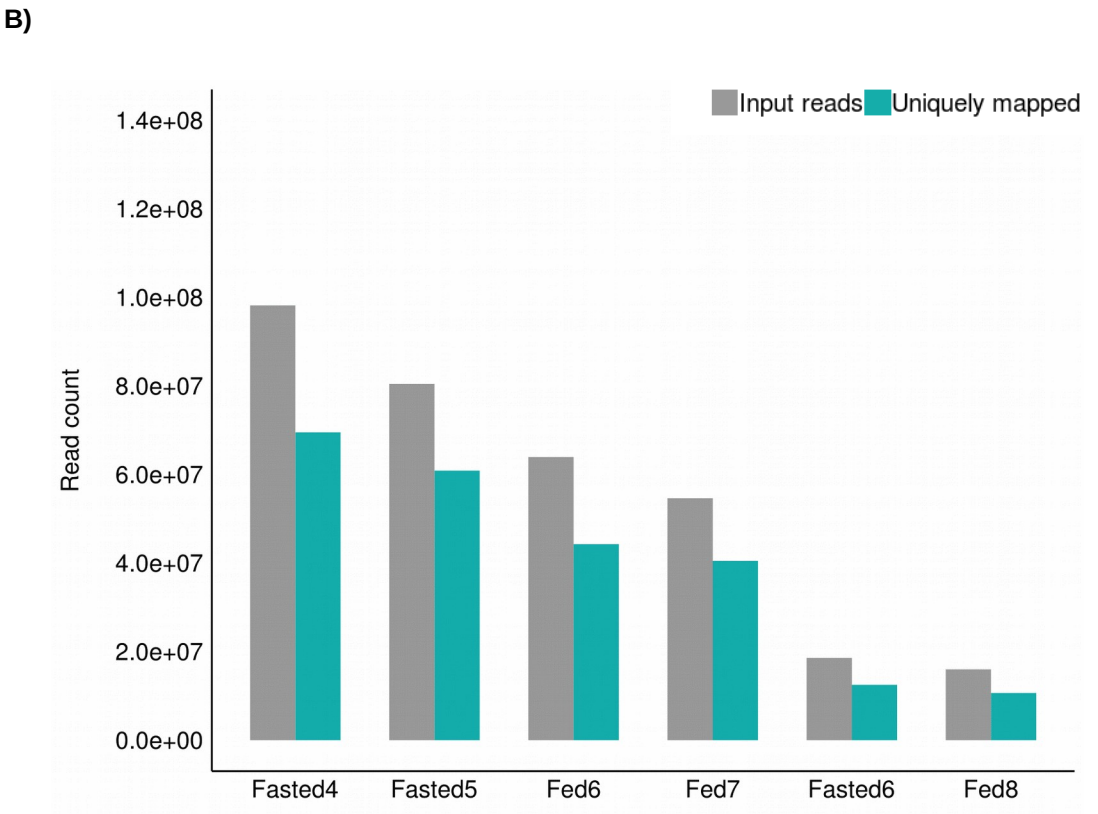
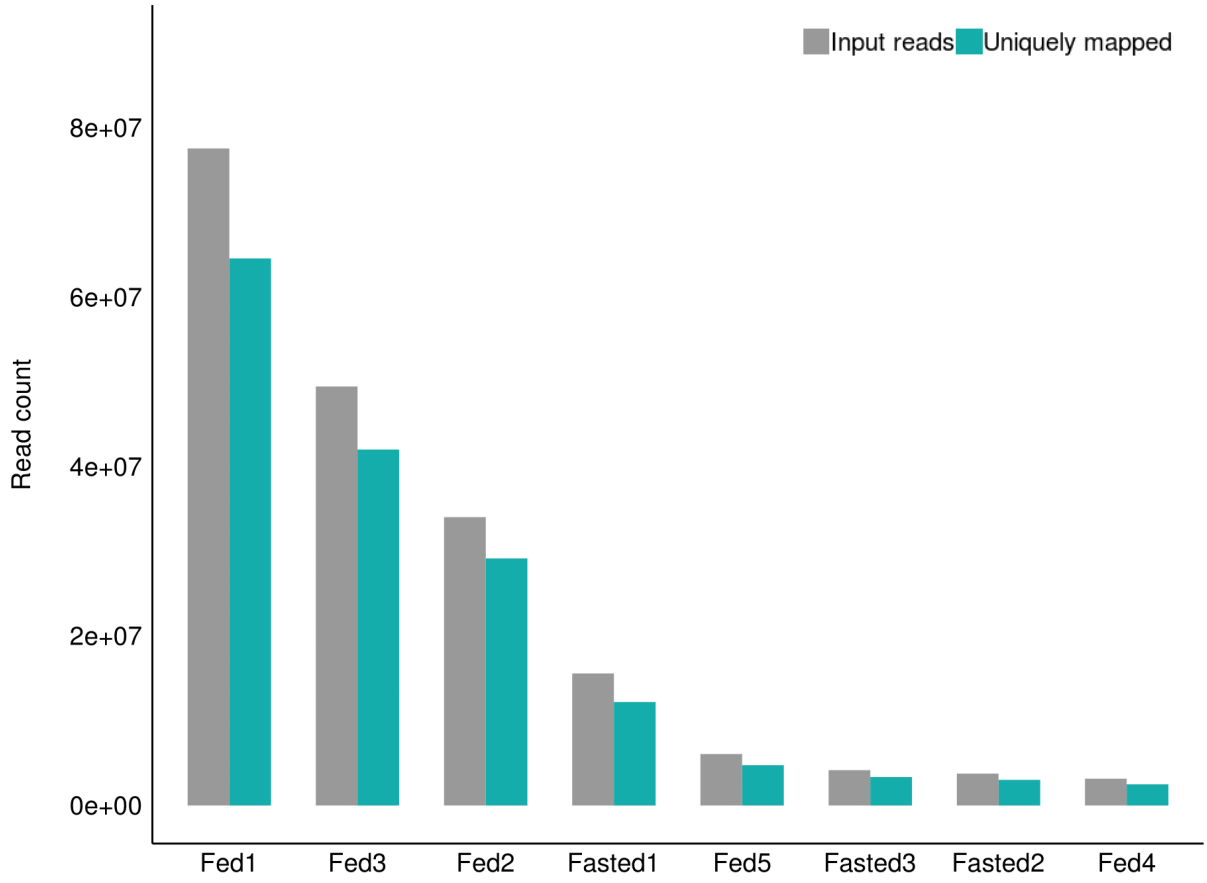




**Figure 2: A)** Histogram of raw read count used as input to *STAR* and uniquely mapped reads across the samples for the Campbell dataset. **B)** Histogram of raw read count used as input to *STAR* and uniquely mapped reads across the samples for the Chen dataset.

### 1.2.2 Assign reads to genes

We used the featureCounts tool to assign our mapped reads to genes. The samples had a satisfactory number of assigned reads (**Figure 3**). We then used these results for the following read-counting steps.

**A)**



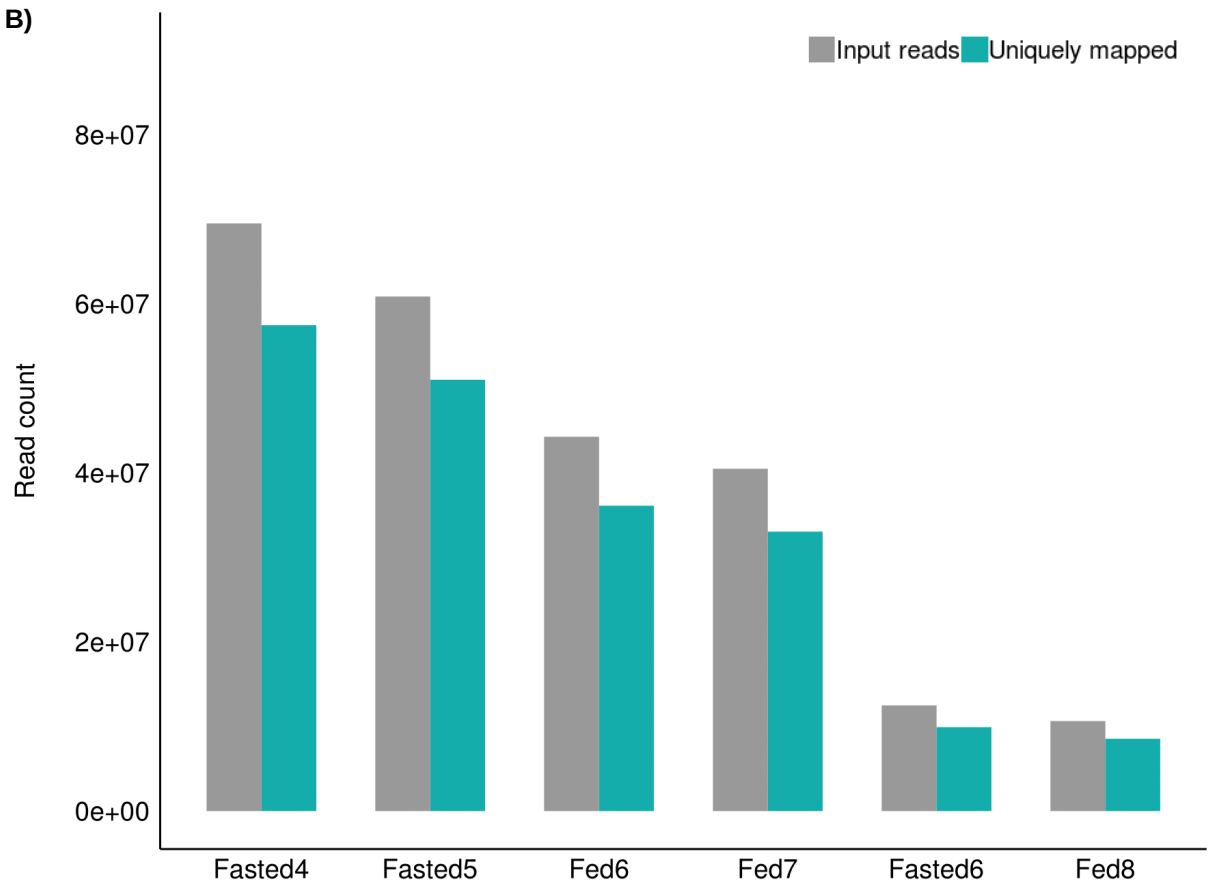



**Figure 3: A)** Histogram on the input and successfully assigned read count across the samples in the Campbell dataset. **B)** Histogram on the input and successfully assigned read count across the samples in the Chen dataset.

### 1.2.3 Estimate gene expression

We obtained the UMI count per-cell and per-gene using the *Umi-tools count* function. In this step, reads without a gene-assignment tag were not counted and duplicated reads were removed. Although the number of reads counted seem to be low (**Figure 4**), it was expected since duplicated reads are common in single-cell data under the Drop-Seq protocol. The raw expression tables per-gene and per-cell generated at this step were used for further quality controls.

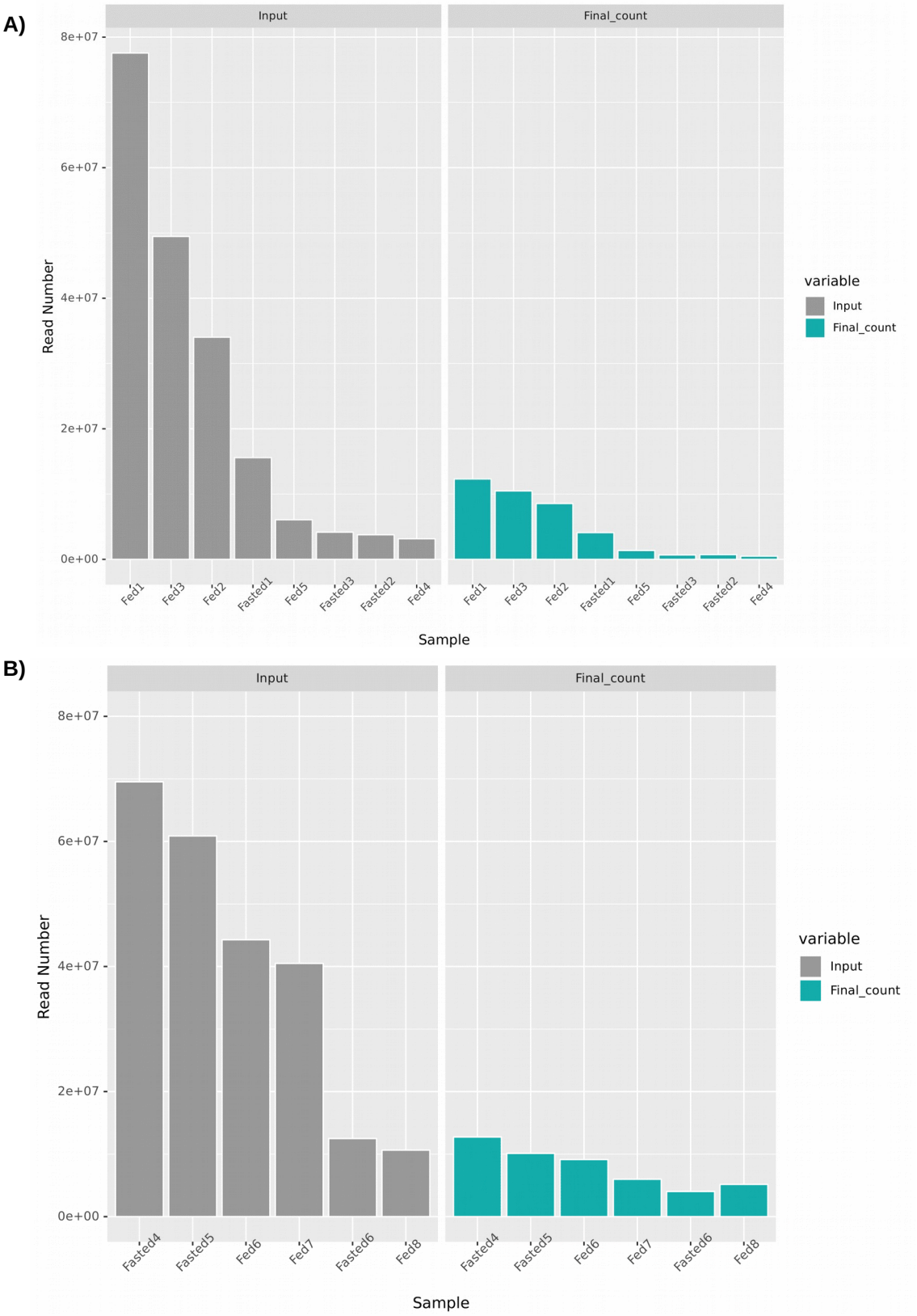


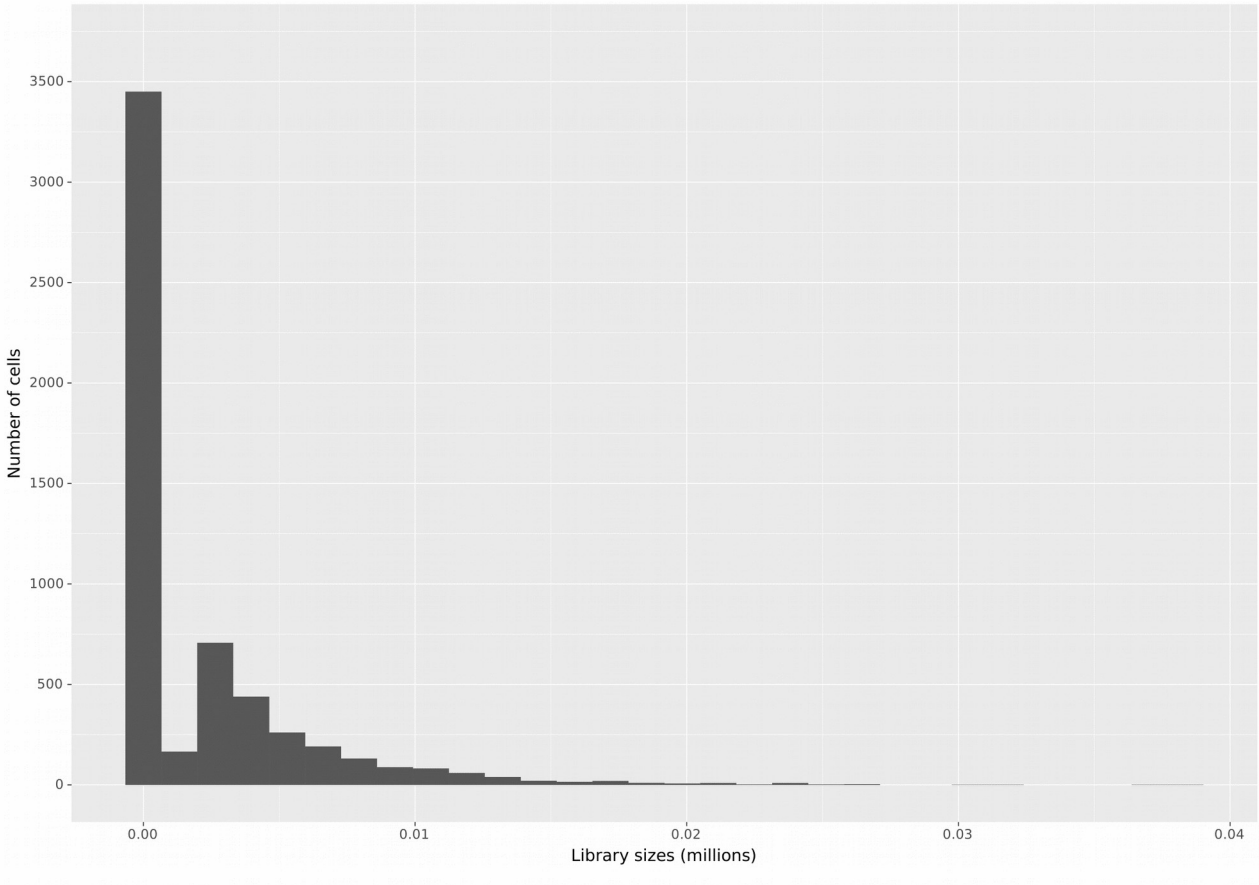

**Figure 4: A)** Histogram on the total number of reads inputed in Umi-tools count and the final number of counted reads across samples of the Campbell dataset. **B)** Histogram on the total number of reads inputed in Umi-tools count and the final number of counted reads across samples of the Chen dataset.

### 1.2.4 Additional quality controls

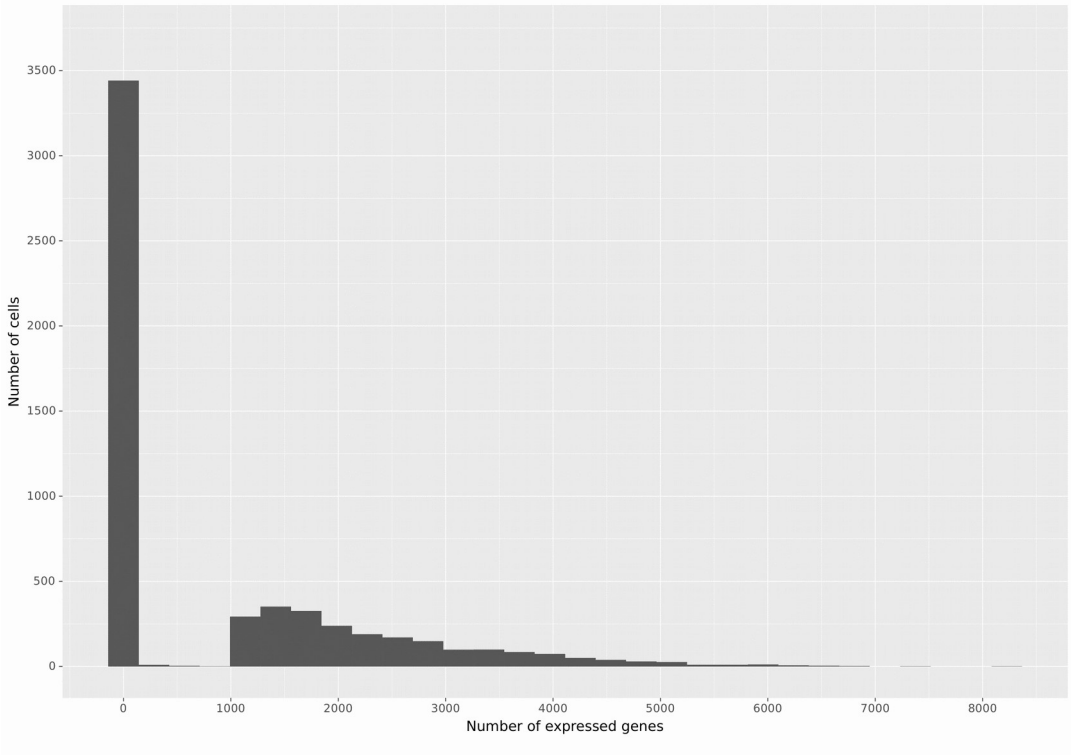
### A) Cell quality control

We first observed the distribution of the the values for the library-size and the number of expressed genes in our raw data. We can see in the histograms (**Figure 5)** the high presence of zero values, which indicates a high number of cells with zero gene count and a high number of cells with zero expressed genes. This type of cell is not informative and therefore need to be removed. This finding was consistent in all samples for both datasets.

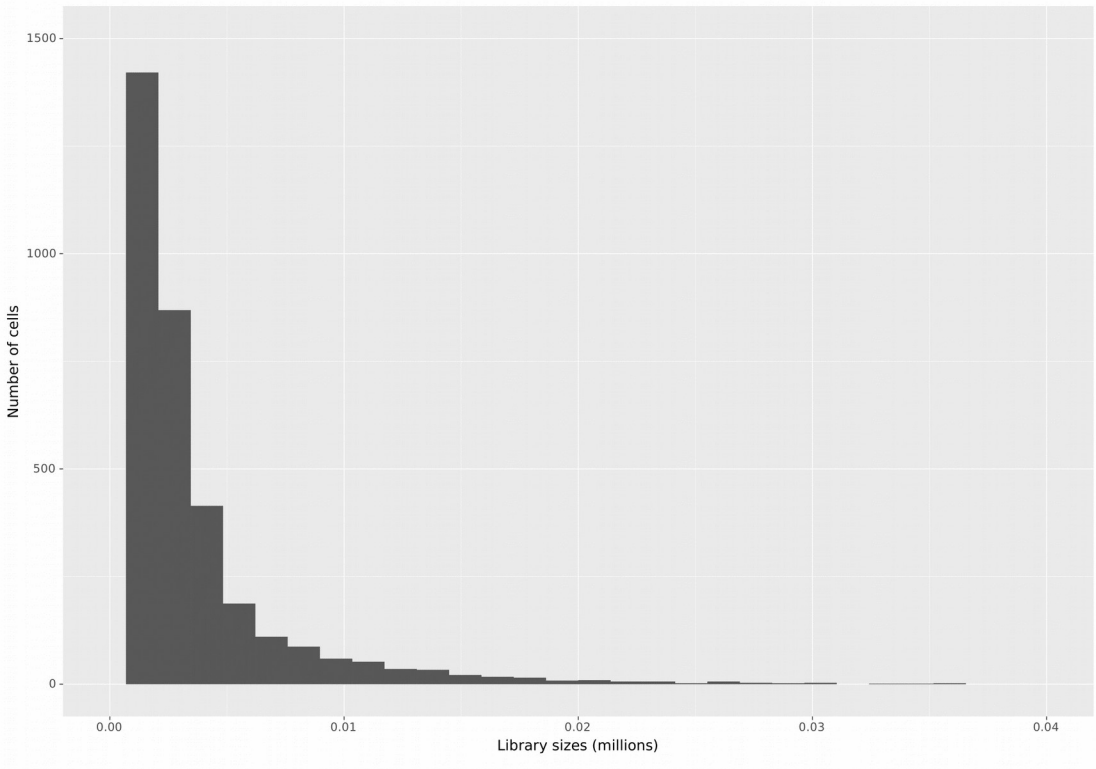
**A)**



**B)**


Number of cells
3500
3000
2500
2000
1500
1000
500
0
0
1000
2000
3000
4000
5000
6000
7000
8000
Number of expressed genes

**C)**


Number of cells
1500
1000
500
0
0.00
0.01
0.02
0.03
0.04
Library sizes (millions)

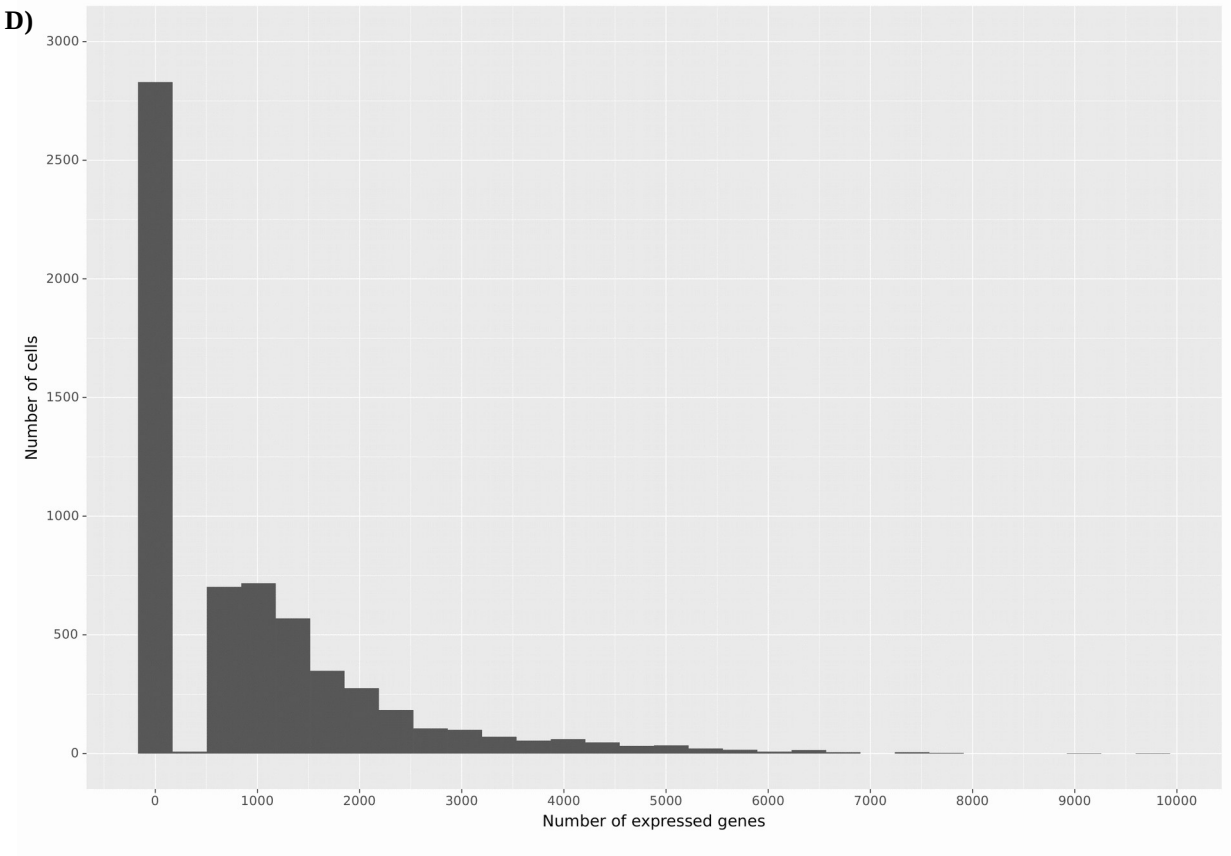
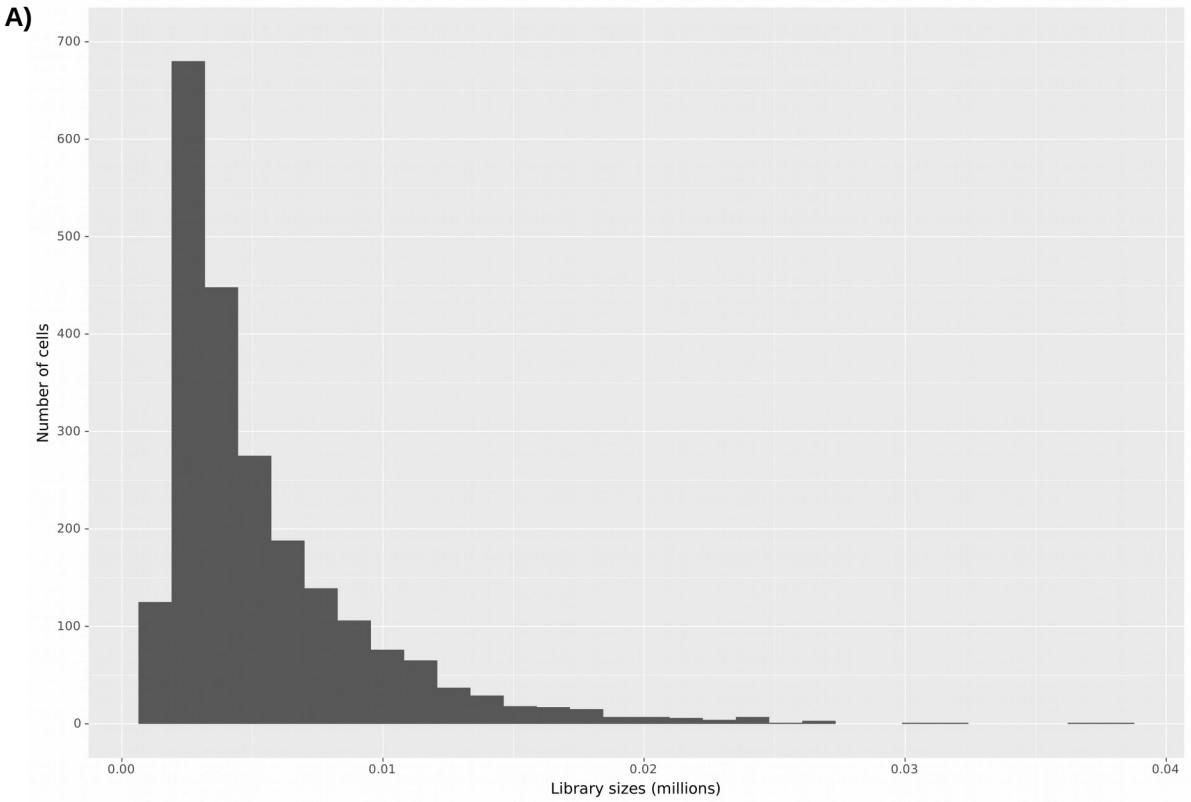
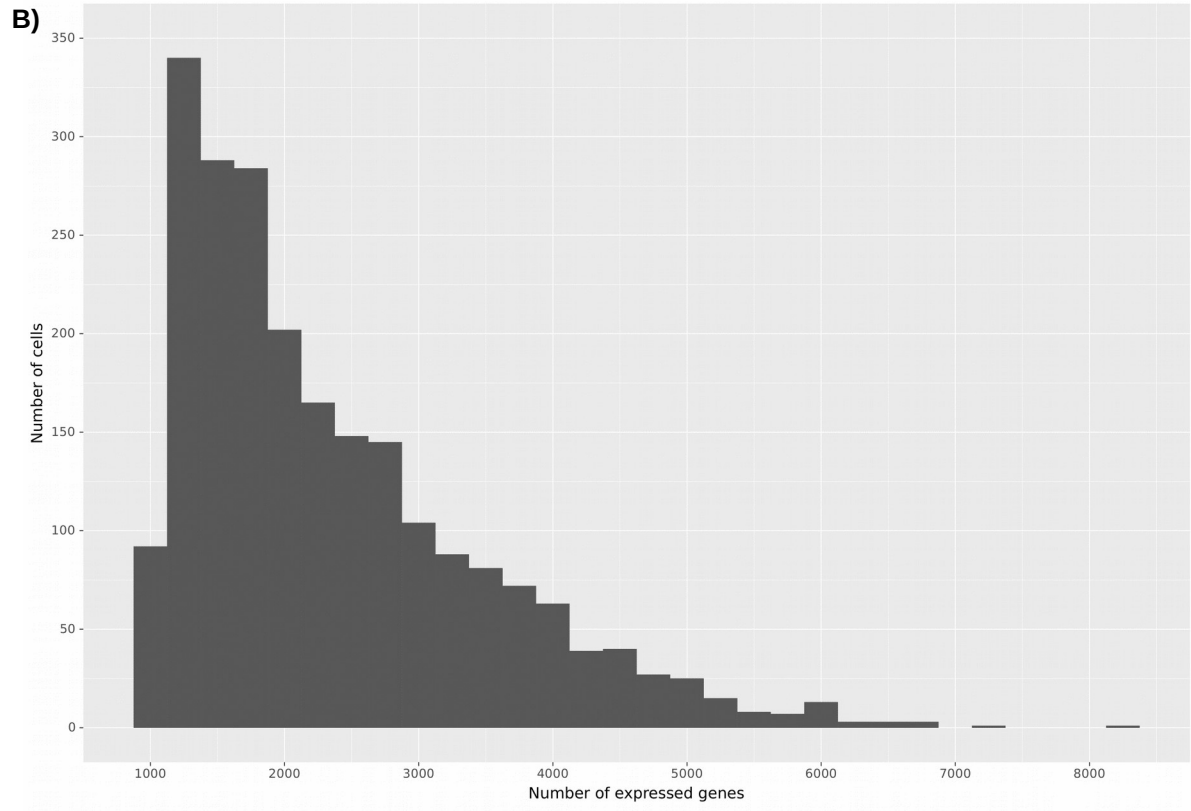



**Figure 5: A)** Histogram on the library-size for the Fed 1 sample (Campbell dataset) **B)** Histogram on the number of expressed genes for the Fed 1 sample (Campbell dataset) **C)** Histogram on the library-size for the Fasted 4 sample (Chen dataset) **D)** Histogram on the number of expressed genes for the Fasted 4 sample (Chen dataset).

We looked on the distribution of the same metrics and its improvement after applying a filter to remove cells with gene count below 800 (**Figure 6**). This result was consistent across all the samples in both datasets. The filter applied on the number of expressed genes was enough to improve the distribution of the library-size across the samples and therefore no specific filter was necessary.


A)
Number of cells
Library sizes (millions)
0
100
200
300
400
500
600
700
0.00
0.01
0.02
0.03
0.04


B)
Number of cells
Number of expressed genes
0
50
100
150
200
250
300
350
1000
2000
3000
4000
5000
6000
7000
8000

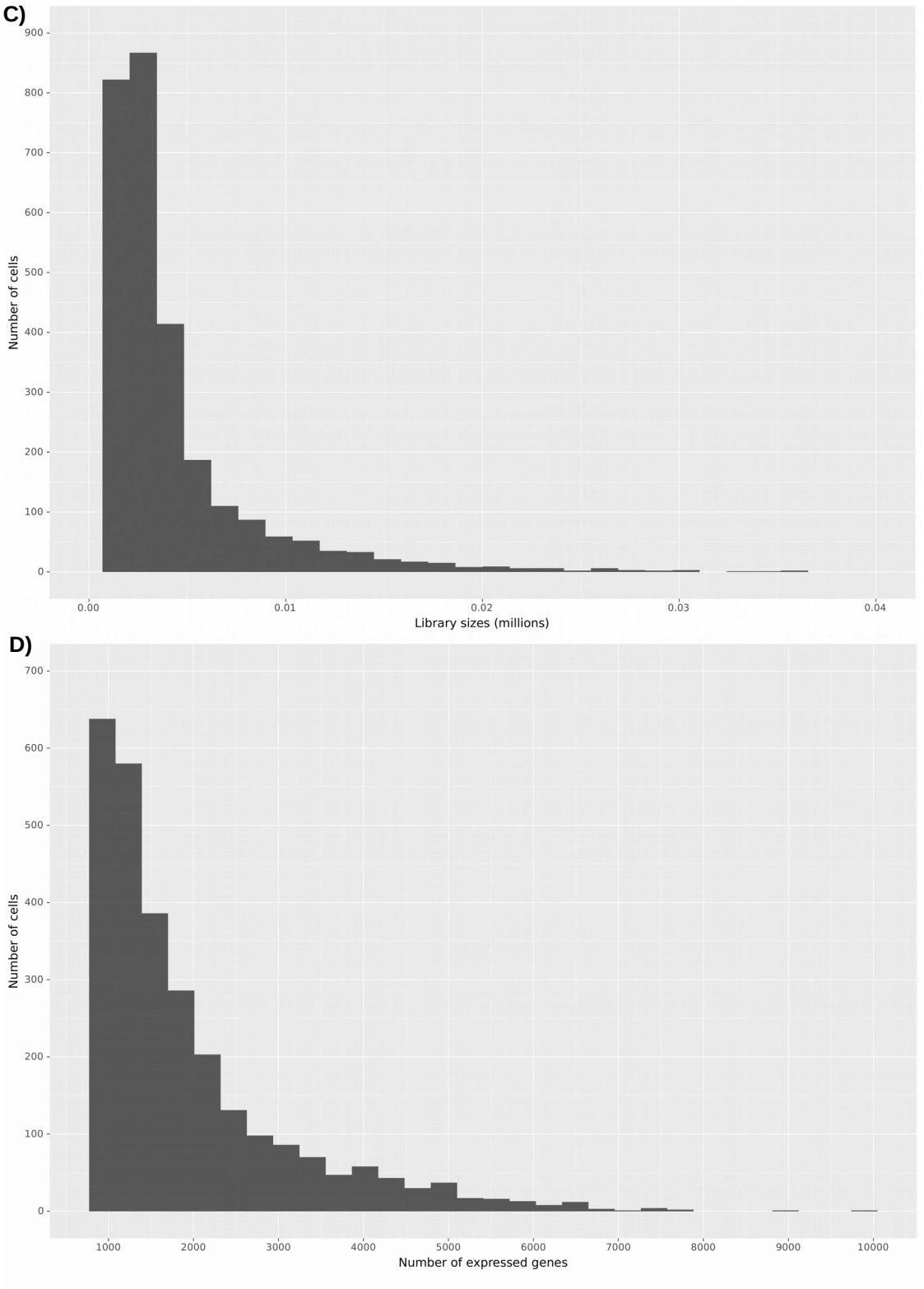



**Figure 6:** Plots after the filtering for number of expressed genes. **A)** Histogram on the library-size for the Fed 1 sample (Campbell) **B)** Histogram on the number of expressed genes for the Fed 1 sample (Campbell) **C)** Histogram on the library-size for the Fasted 4 sample (Chen) **D)** Histogram on the number of expressed genes for the Fasted 4 sample (Chen).

## B) Cell-cycle effect correction

The cell-cycle assignment in our analyzed samples indicated that most of the cells are in the G1 phase (**Figure 7**). It is important to have in mind that most of the tools currently available aim to classify the cells in one of the cell-cycle phases, which is not suitable for post-mitotic cells that do not belong in any of its phases. That could be the case of many neuron cell-types, where the classification would be irrelevant. Also, the cell-cycle effect would be relatively subtle compared to the differences between the cell types in our samples. Therefore, we decided not to perform any additional correction.

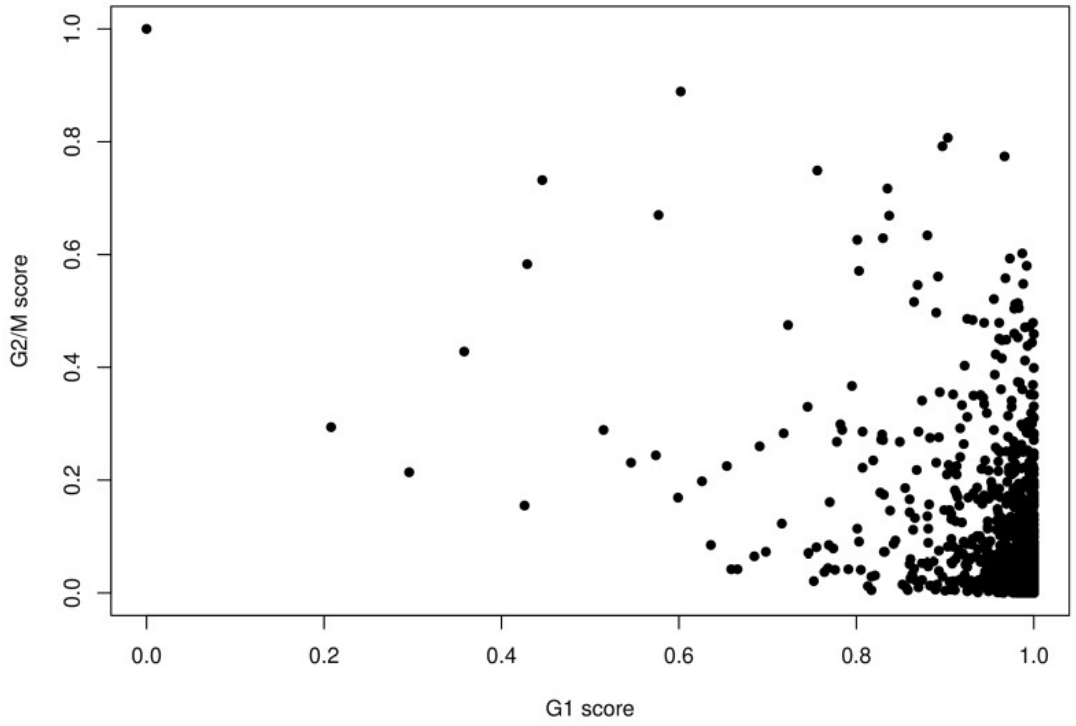
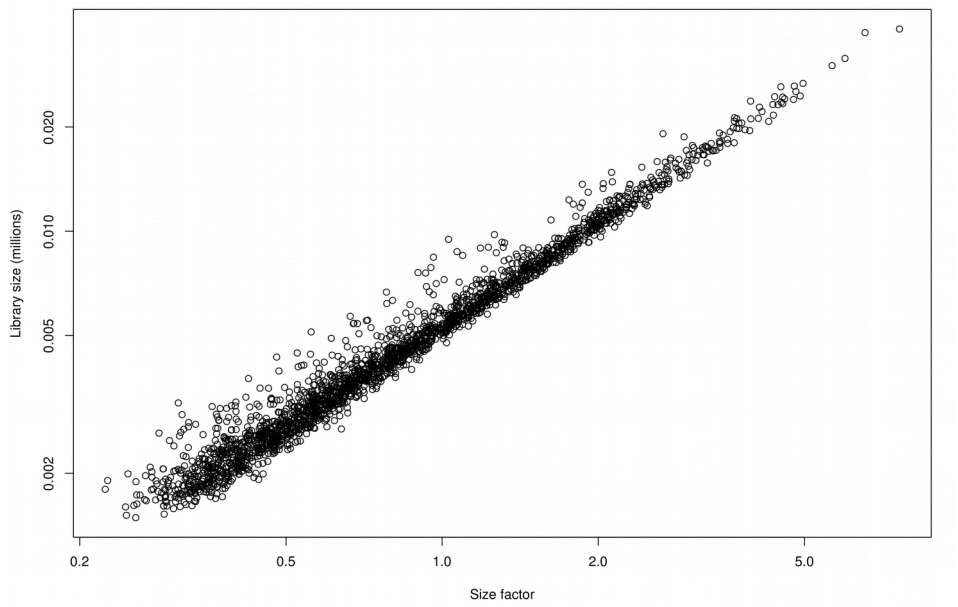



**Figure 7:** Cell-cycle assignment plot for the Fed 1 sample (Campbell) using the *Cyclone* tool. Each point represents a cell. Cells in the G1 phase have G1 score higher than 0.5 and G2/M score lower than the G1 one. The majority of the cells were assigned to the G1 phase.

## C) Normalization

In order to access the normalization quality, we plotted the relation of the library-size with the size-factors for each sample. Since we pool counts to improve the size-factor estimation, we would expect a tight relation of library-size and size-factor (**Figure 8**), which was found in all samples analyzed in both datasets. Therefore, the normalization processes succeed.

**A)**




**B)**

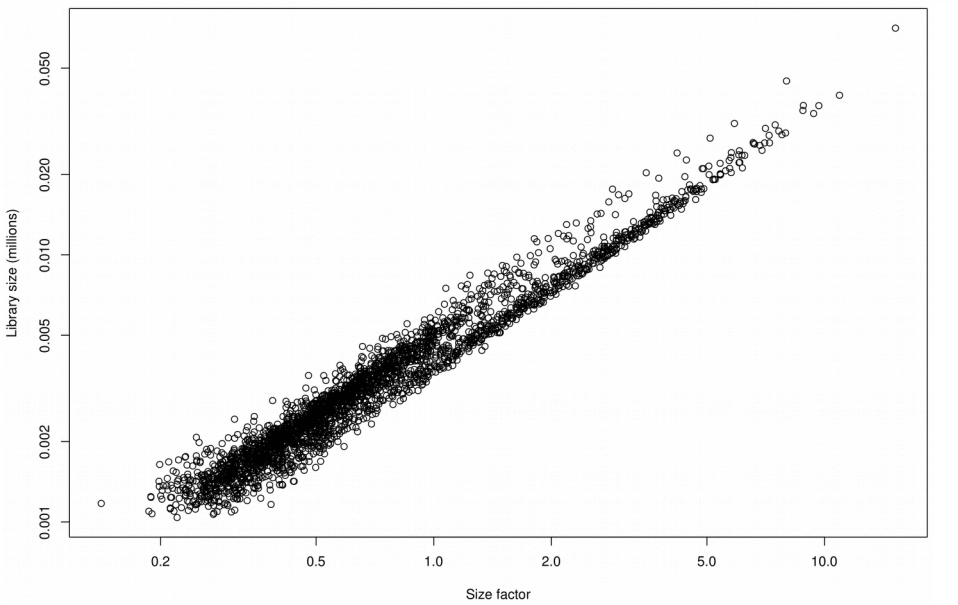



**Figure 8: A)** Correlation of the library-size and size factor for the Fed 1 sample (Campbell). **B)** Correlation of the library-size and size factor for the Fasted 4 sample (Chen).

**D) Correction via MNN**

The cells from both Chen and Campbell's dataset were used as input for the *MNNCorrect* function in order to perform the batch correction. The number of genes used was 6079, the ones present in all cells. The output of this step was the same number of cells and genes that were inputted but with corrected values accounting for technical variables.

### 1.3 Single-cell data downstream analysis results

#### 1.3.1 Clustering

We performed the clustering analysis in the log counts of the raw data and in the batch corrected data. In the raw data, 46 clusters were identified while in the batch corrected data 96 clusters were identified.

#### 1.3.2 Dimensionality reduction

In order to visualize the data, we plotted a t-SNE colored according to the identified clusters. The t-SNEs for the raw and batch-corrected data can be visualized in Figure 9 and Figure 10.

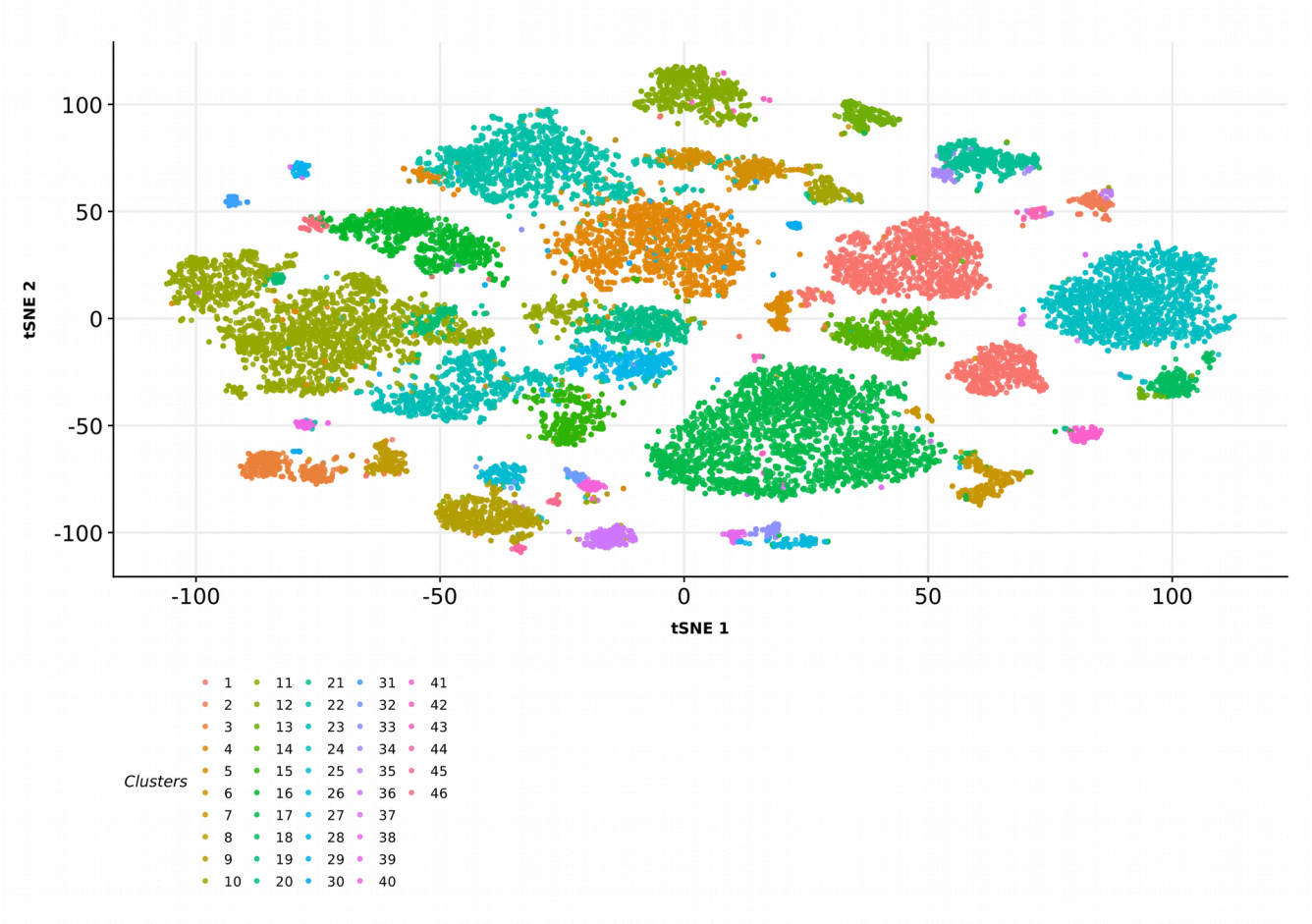



**Figure 9:** T-SNE plot of the normalized data, colored according to the clusters.

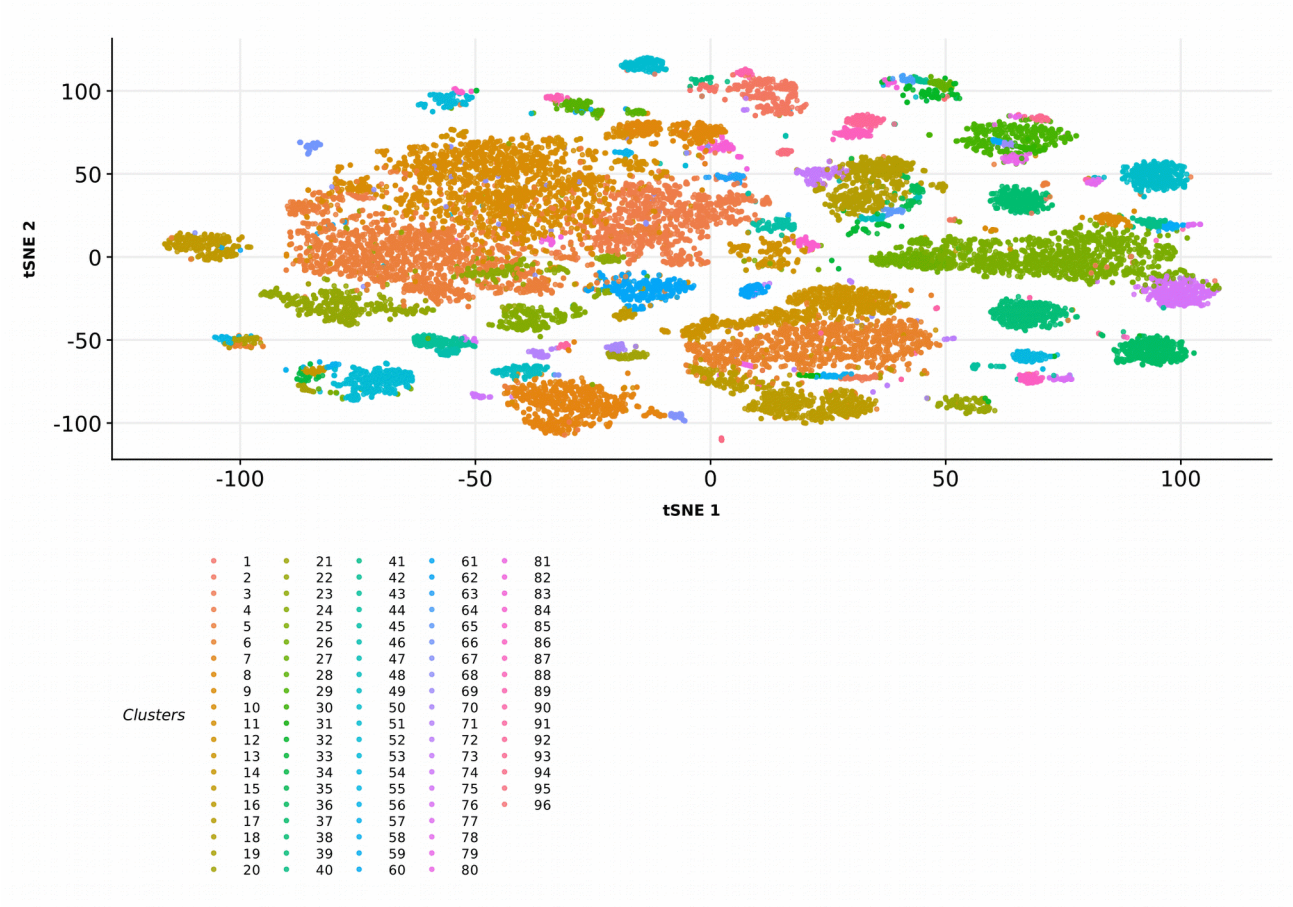



**Figure 10:** T-SNE plot of the MNN-corrected data, colored according to the clusters.

### 1.3.3 Agrp identification

In order to identify the Agrp cells, we plotted the t-SNE of the batch corrected data, coloring the cells according to corrected expression value for both *Agrp* and *Npy* genes. We visually identified 6 clusters that had corrected expression higher than 0.05 for both genes. We attributed a cell type label identifying all cells in such clusters as Agrp neurons (**Figure 11**) and they totalized 1000 cells. The identification of Agrp cells in the Campbell dataset based on the raw counts resulted in 1110 cells. The Agrp neurons identified with these two approaches resulted in 1191 cells.

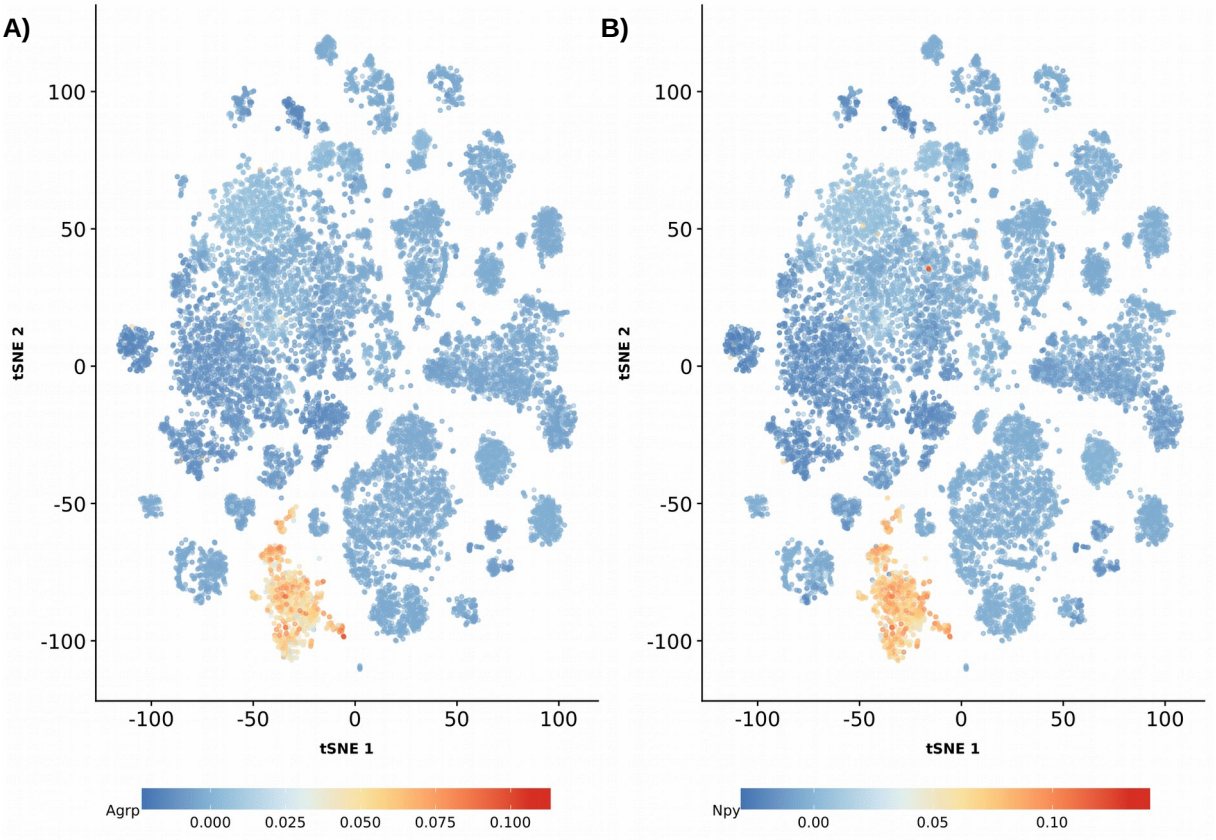



**Figure 11:** T-SNE plots on the MNN corrected data. The cells are colored according to the expression of a gene marker: **A)** *Agrp* **B)** *Npy*. The warmer the color, the higher the expression value of the given gene.

### 1.3.4 Batch correction for differential expression

The PCA on the data (**Figure S2**) showed that the origin of the dataset and the feeding condition were not responsible for the differences in the data. The results of the *plotExplanatoryVariables* function showed that the sub-batch variable, which describes how the batches were structured within each dataset, explains the highest amount of variation, even though it is considerably small (**Figure S3**). Observing such findings, we imagined that this variability could be caused by a hidden batch effect. However, the results of the svaseq function demonstrated that more than 20 variables could explain the variability in the data. Therefore, it was not possible to correct for the technical variation and perform differential expression. We performed gene enrichment as an alternative strategy.

### 1.3.5 Gene enrichment: food deprivation

The filtering for lowly expressed genes in all Agrp cells resulted in a total of 2591 genes. The distribution of the enrichment values for such genes can be visualized in the histogram in Figure 12. We can observe that the majority of the genes are distributed for enrichment values between 0.5 and 1.104988.

### 1.3.6 Transcriptome definition

The threshold applied to select the genes most impacted by food deprivation resulted in 1005 genes. Those genes were considered as the transcriptome of Agrp neurons upon food deprivation and were used for further downstream analyses.

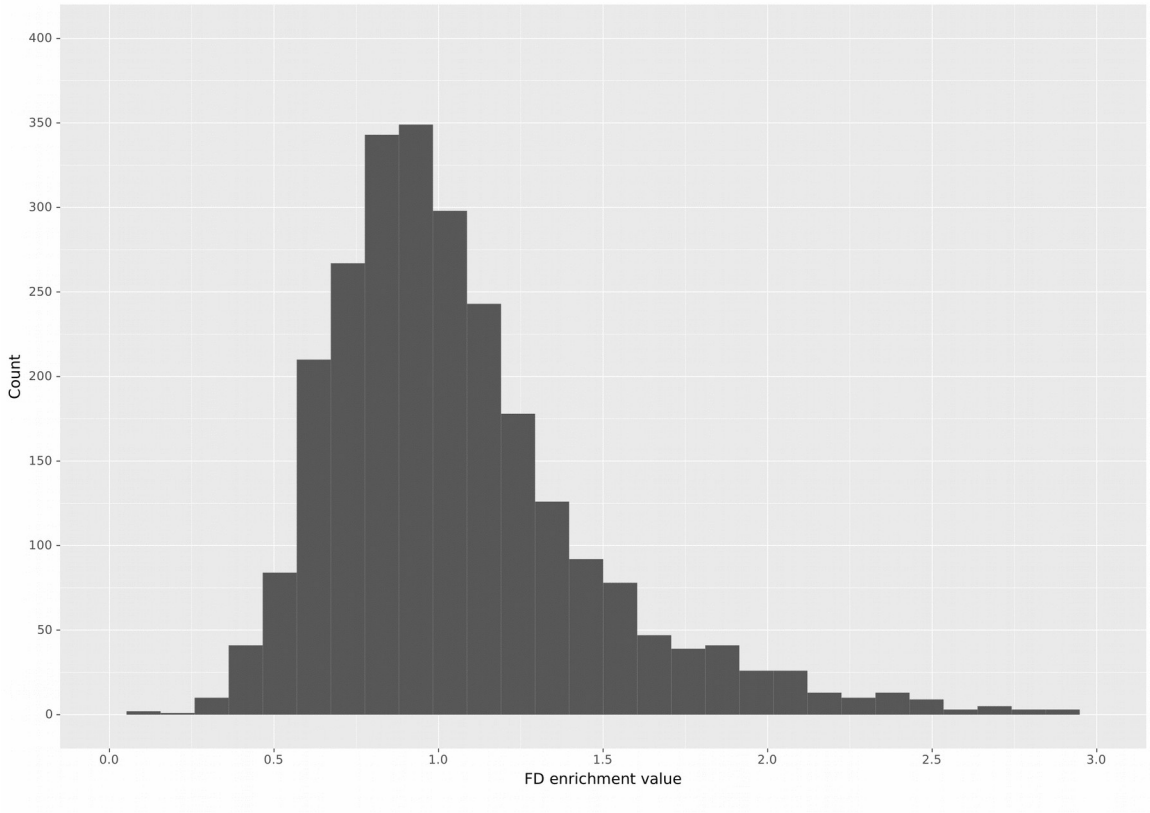



**Figure 12:** Histogram on the enrichment value upon FD in 2591 genes in the Agrp cells.

### 1.3.7 Transcriptome functional characterization

The gene ontology enrichment analysis in String associated the transcriptome of Agrp neurons to several process. We analyzed the ones which -log10(FDR) > 2, that is equivalent to a False Discovery Rate < 0.01. Then, we filtered all GO terms according to their hierarchy to obtain the most relevant terms. The most relevant biological processes associated to the transcriptome of Agrp neurons upon food deprivation can be seen in Figure 13. We observed a prevalence of processes associated to neuronal function, specially synapses and their regulation. We also observed processes associated to vesicle exocytosis and endoplasmic reticulum regulation.

For the pathway analysis, we also filtered the processes using a threshold of -log10(FDR) > 2. We noticed the presence of pathways related to synaptic signaling, synaptic regulation and protein metabolism; similar to what we have seen in the biological processes analysis. The pathways associated to the Agrp neuron transcriptome upon food deprivation can be visualized in the bar plot in Figure 14.

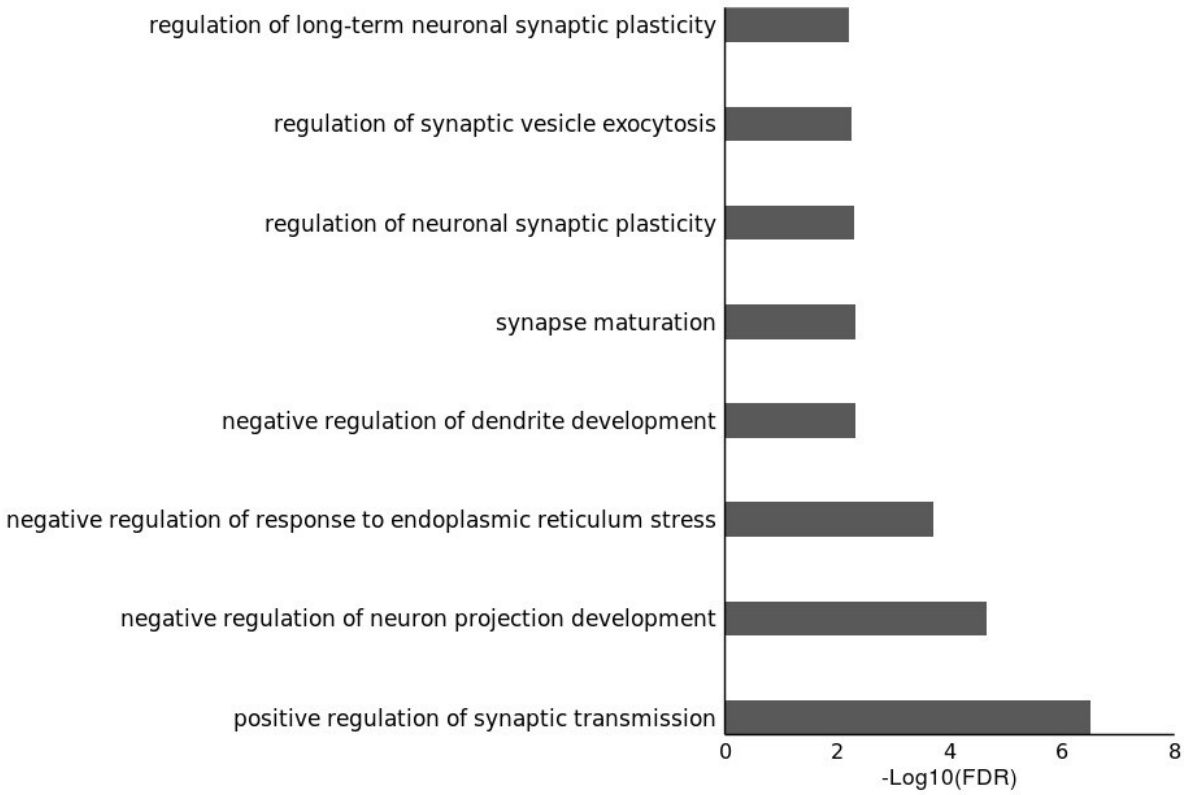



**Figure 13:** Bar plot on the main biological processes associated to the transcriptome in *String*.

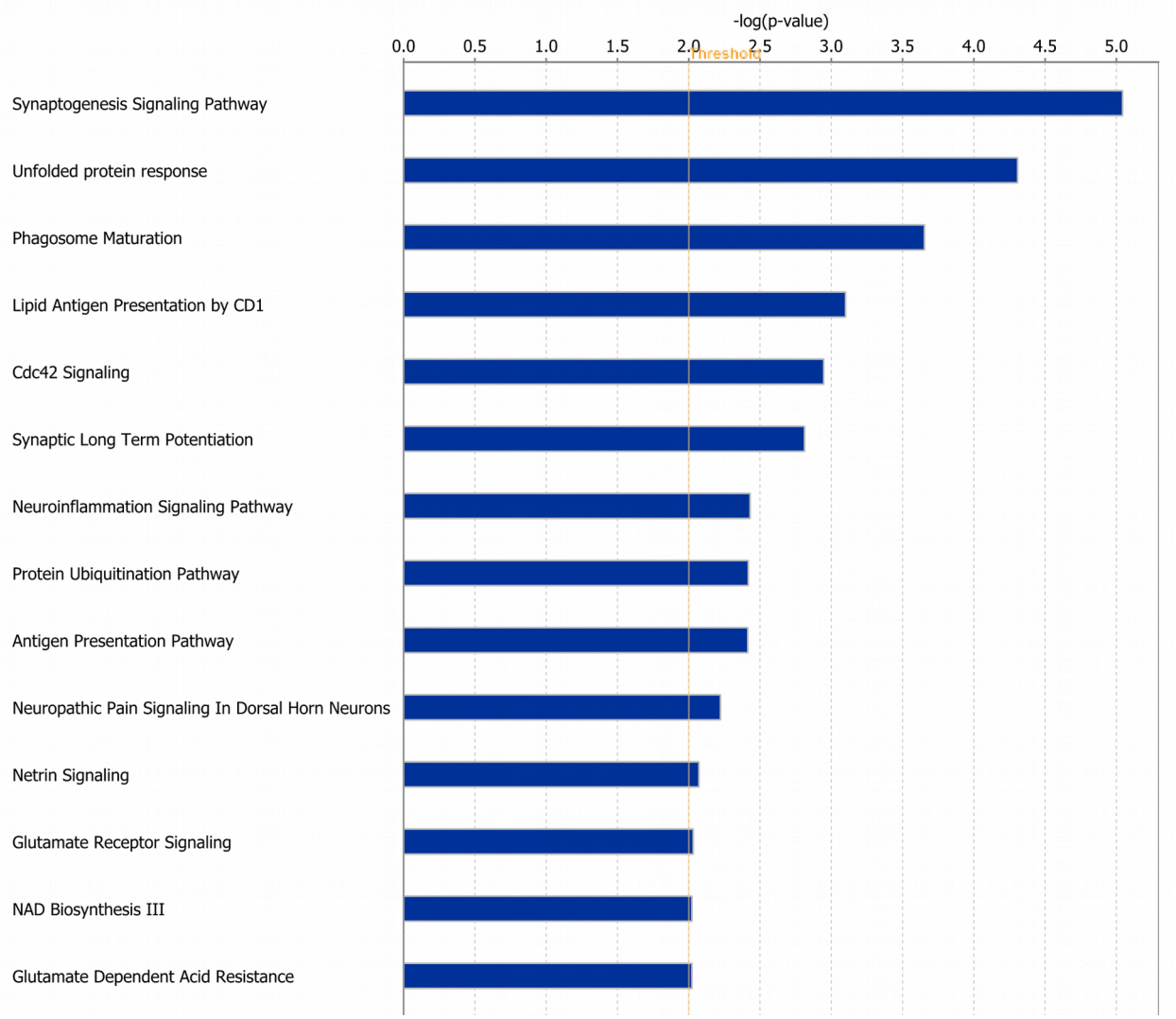



**Figure 14:** Bar plot on the pathways associated to the transcriptome upon food deprivation in *IPA*.

## 2 Translatome results

### 2.1 RiboTag data pre-processing results

#### 2.1.1 Sequencing quality analysis

The sequencing of the samples generated under the RiboTag technique presented a good quality, which can be seen in the per base sequence quality plot for one of the samples (**Figure 15**). Importantly, similar quality result was observed for all

samples used in this study. Since the metrics for sequencing quality indicated the sequencing was good enough in our samples, no base trimming was performed.

## 2.2 RiboTag data processing results

### 2.2.1 Alignment results

The alignment results were satisfactory for all the samples, with high proportions of uniquely mapped reads, as observed in Figure 16. It was valid for the input samples as well (**Figure S1**). The BAM file generated at this step was used for the counting step, which generated the expression tables used in the downstream analysis.

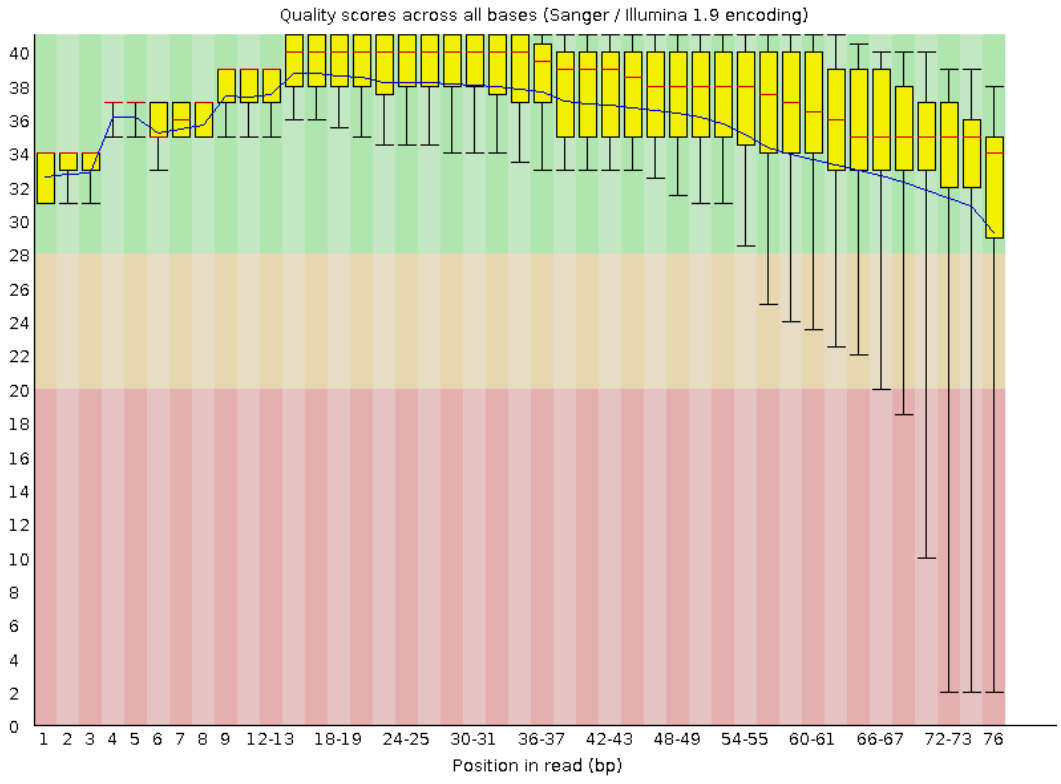



**Figure 15:** Per base sequence quality plot performed with *FastQC* for the Fed 4 sample.

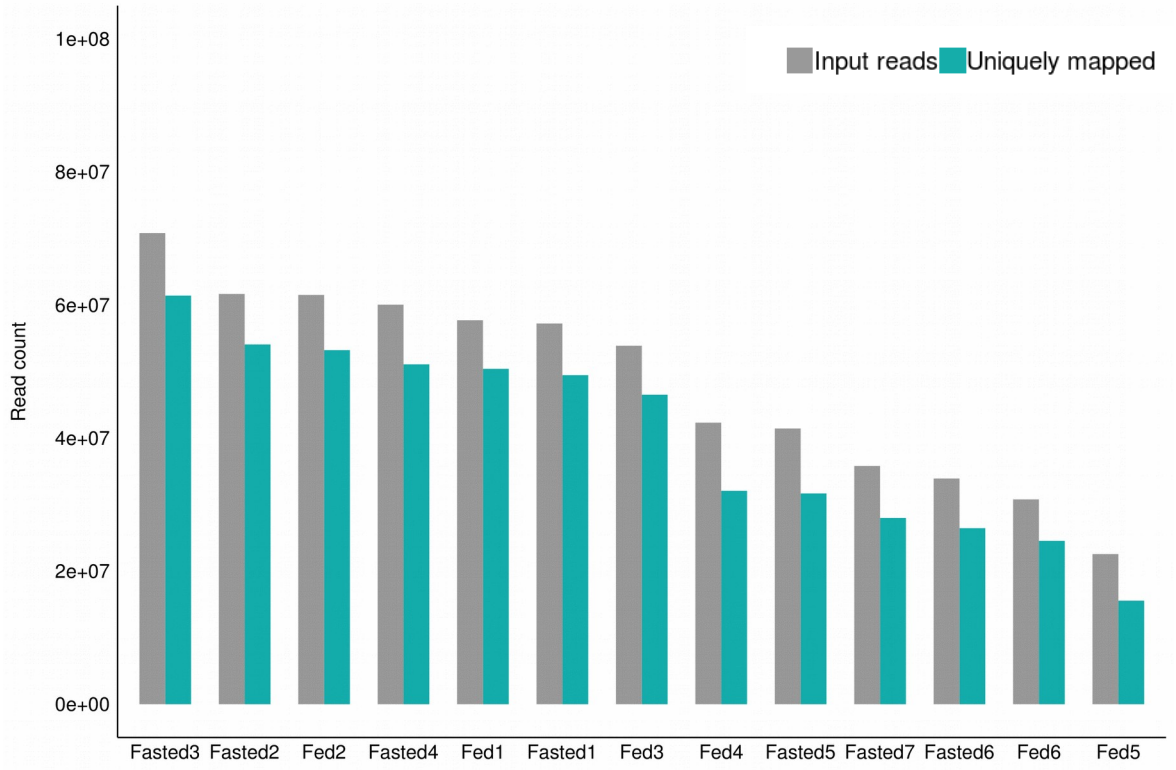



**Figure 16:** Histogram on the amount of input and uniquely mapped reads across all the IP samples generated under the RiboTag technique.

### 2.2.2 Assessment of Agrp neuron ribosome immunoprecipitation

We observed in our negative control that the enrichment values for markers of distinct cell types present in the ARC was very low. This shows that the immunoprecipitation protocol was able to deplete transcripts from cells other than Agrp neurons (**Figure 17**). The positive control demonstrated that we were able to select the Agrp neuronal transcripts, since there is an enrichment of more than 100-fold in *Agrp* and *Npy* markers in the IP data (**Figure 18**). Based on these control experiments, the data correspond to a highly enriched pool of actively translating transcripts from Agrp neurons.

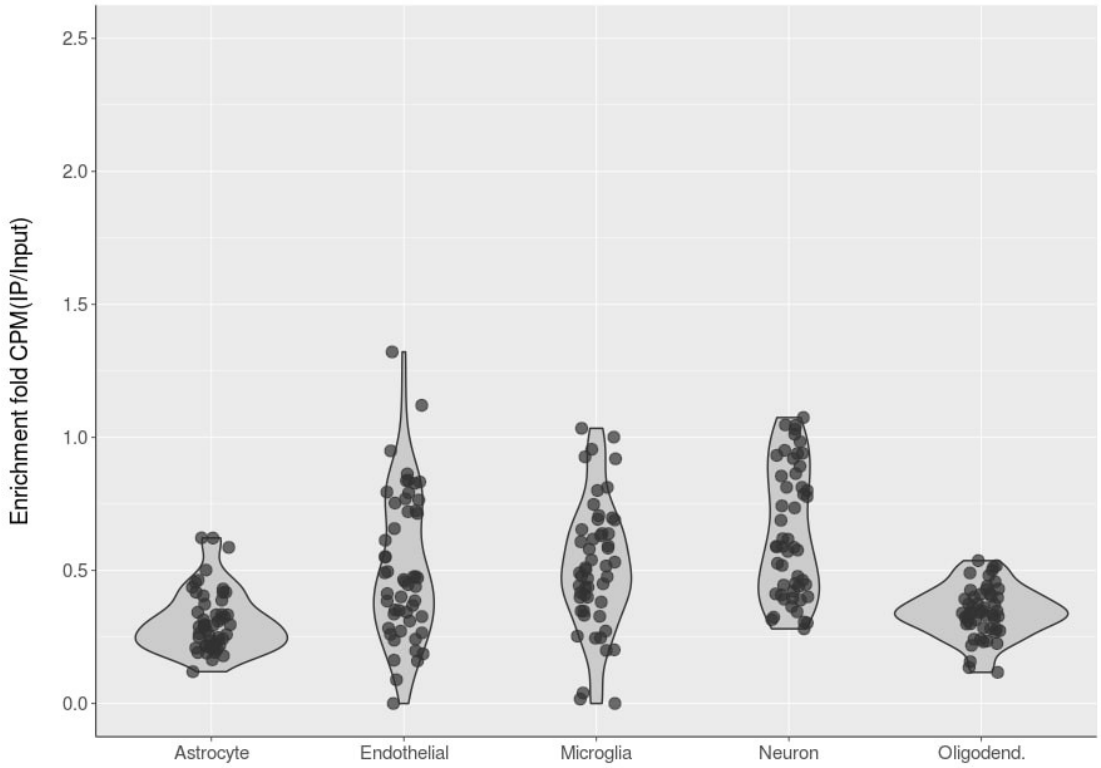



**Figure 17:** Negative control: Violin plot for the enrichment values of different cell-type gene markers.

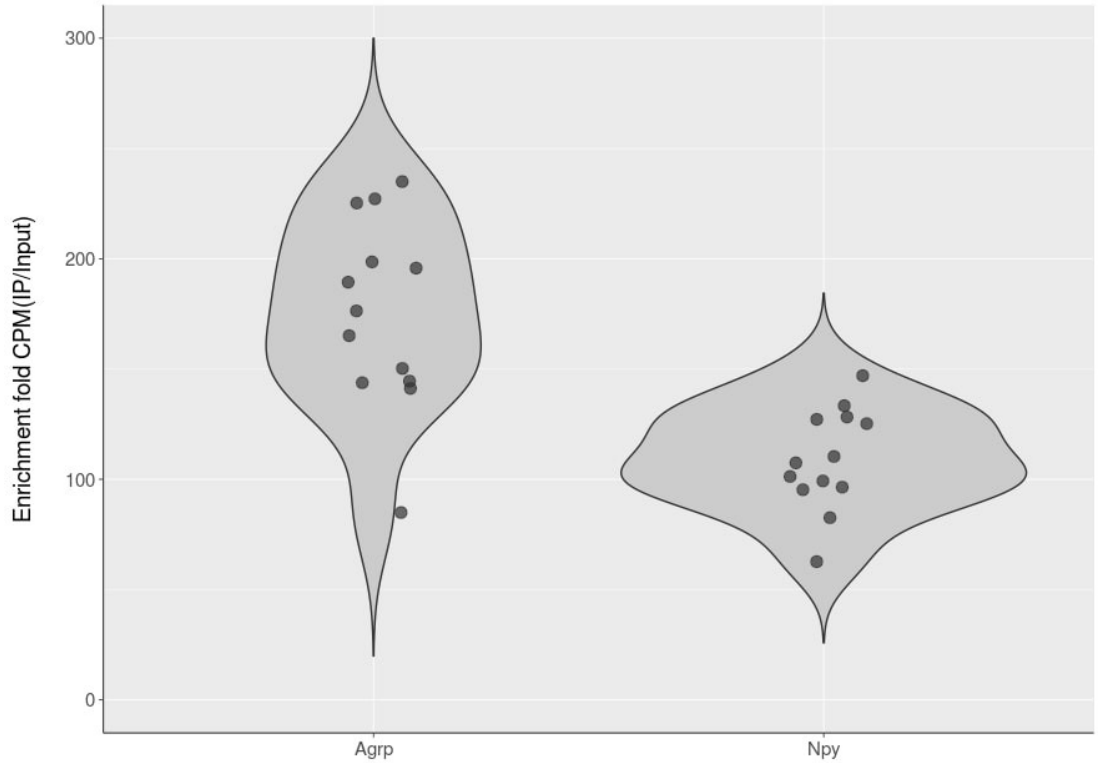



**Figure 18:** Positive control: Violin plot for the enrichment values of the Agrp and Npy gene markers.

## 2.3 RiboTag data downstream results

### 2.3.1 Differential expression: Fed vs FD

The expression data was successfully fitted to the generalized linear model used by the algorithm of *DESeq2*. This can be verified in the plot of the dispersion estimates in Figure 19.

The differential expression of Fed against Food Deprived generated a list of 1022 DE genes. We can see that the majority of the DE genes presented a log fold change varying from -4 to +4 (**Figure 20**).

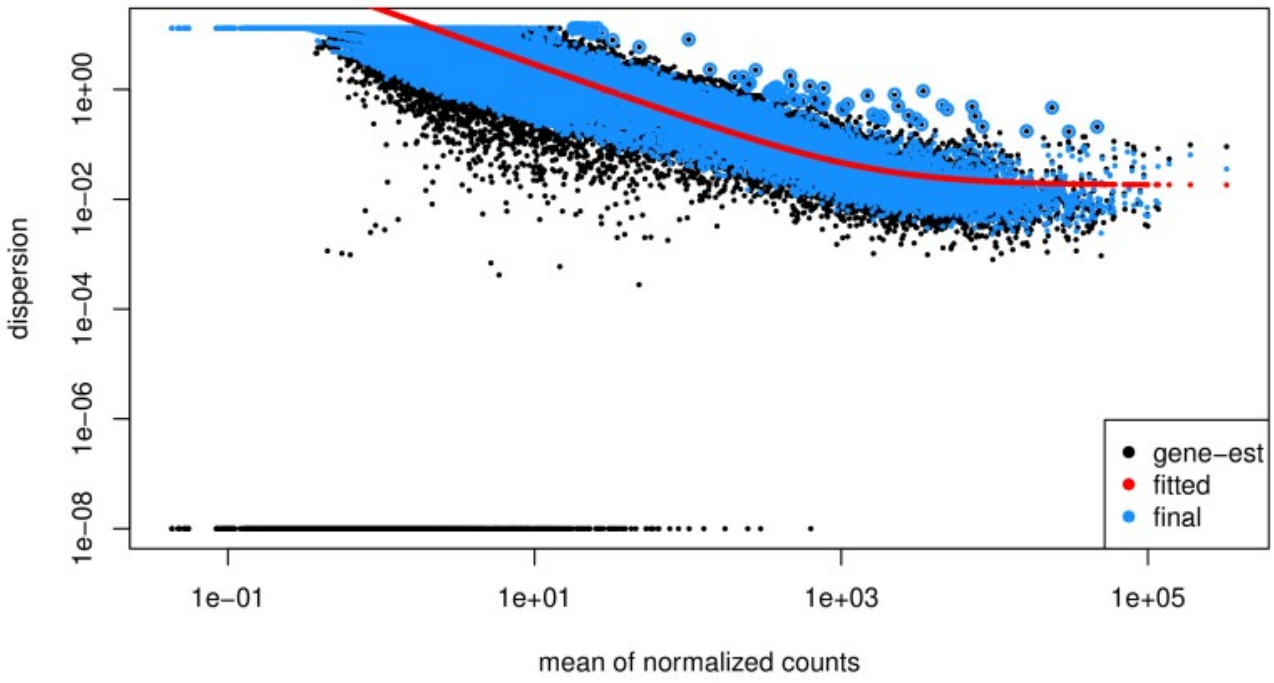



**Figure 19:** *DESeq2* dispersion plot indicating that the dispersion decreases according to the mean of normalized counts increase, as expected. We can also observe that the model fit trend, the red curve, suits the data.

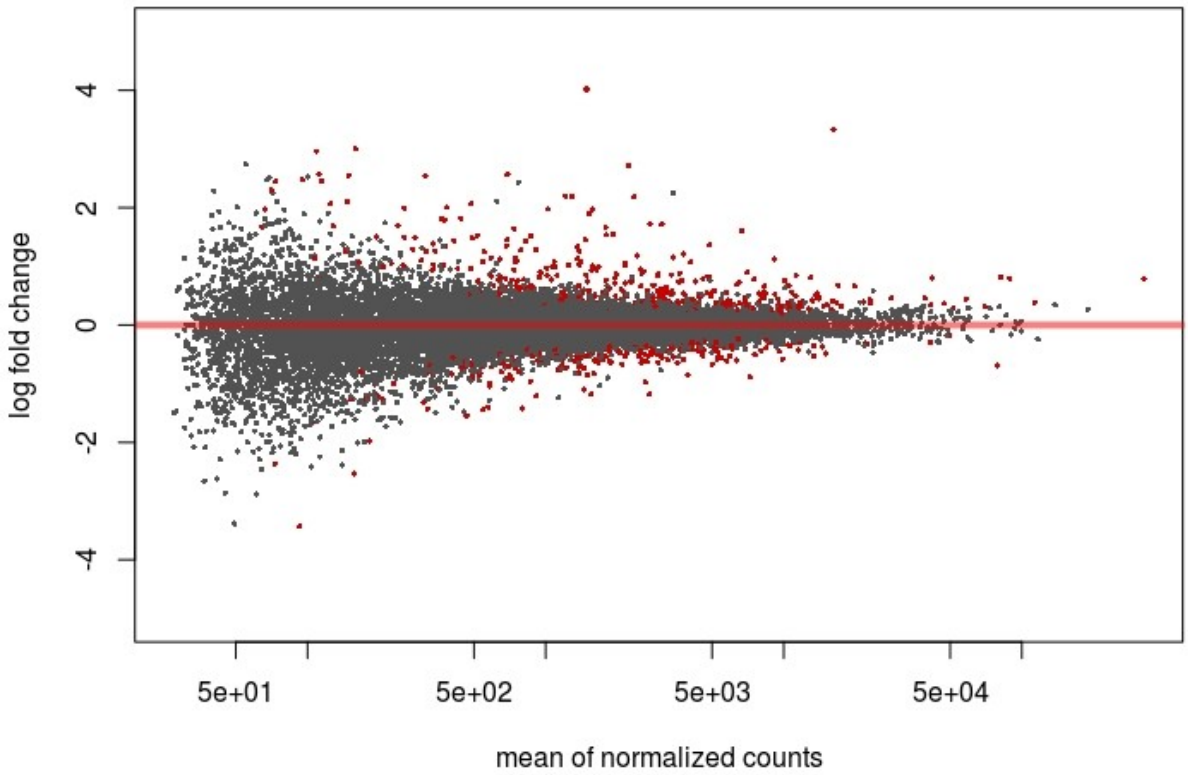



**Figure 20:** Mean of normalized counts vs log fold change. Each point stands for a gene, points in red represent significant DE genes (adjPvalue < 0.05).

### 2.3.2 Translatome definition

The filter applied on the DE list, based on FDR and LFC, resulted in a total of 529 genes. Such genes were considered the translatome of Agrp neurons upon food deprivation.

### 2.3.3 Translatome functional characterization

The gene ontology enrichment analysis in String associated the translatome upon food deprivation of Agrp neurons to several process. We filtered them using -log10(FDR) > 2 here as well. Then, we filtered all GO terms according to their hierarchy to obtain the most relevant terms. The most relevant biological processes associated to the translatome of Agrp neurons upon food deprivation can be seen in Figure 21.

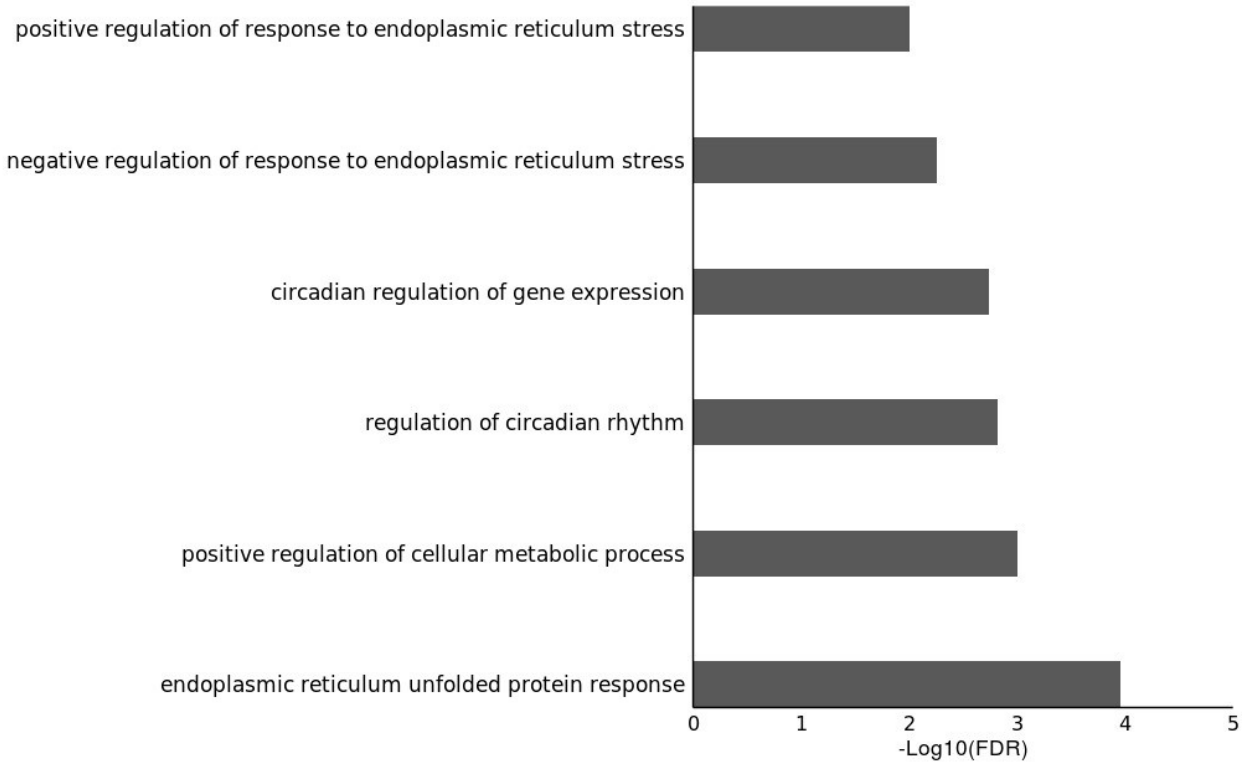



**Figure 21:** Bar plot on the main biological processes associated to the translatome in *String*.

We observed the presence of processes associated to gene expression regulation related to the biological clock and also processes associated to metabolism and endoplasmic reticulum protein folding. For the pathway analysis, we used the threshold of -log(FDR) = 2. We noticed the presence of pathways related to protein folding and gene regulation by the circadian clock, just like observed in the biological processes. We observed some general processes associated to metabolism and signaling. The pathways associated to the Agrp neuron translatome upon food deprivation can be visualized in the bar plot in Figure 22.

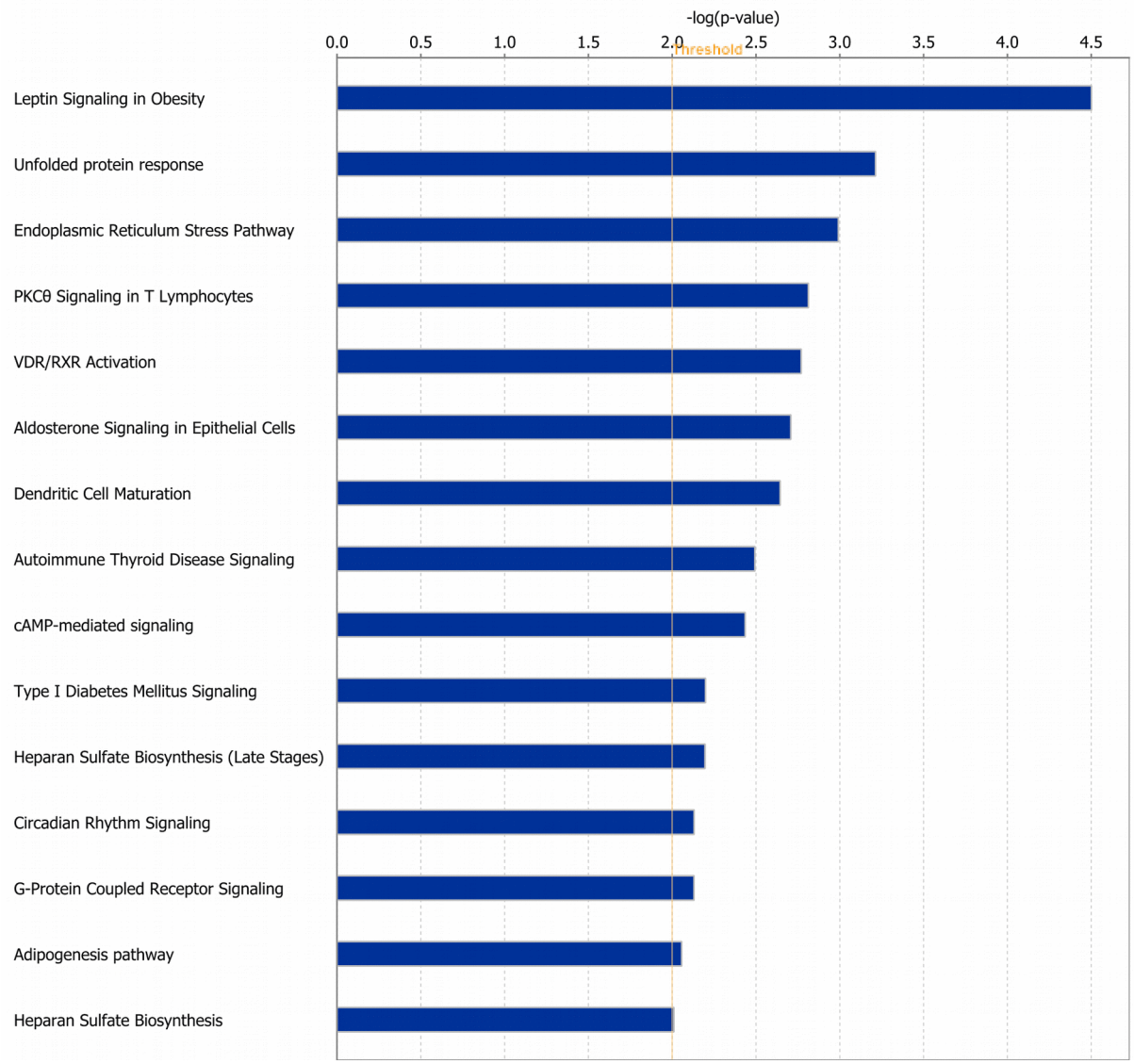



**Figure 22:** Bar plot on the pathways associated to the translatome upon food deprivation in *IPA*.

### 3 Transcriptome vs translatome

We compared the transcriptome and translatome of Agrp neurons via a Venn diagram (**Figure 23**) to see how many genes are shared by them. We observed that from the 1005 genes in the transcriptome, 946 are exclusive from it. In the translatome, in which we identified 529 genes, 470 genes are exclusive from it. 59 genes are shared between transcriptome and translatome, they are described in Table S2 according to the *String* database. We analyzed the shared genes using *String* in order to see the type of biological process they are associated to. We filtered them using -log10(FDR) > 2. Then, we filtered all GO terms according to their hierarchy to obtain the most relevant terms. The most relevant biological processes associated to the genes shared by the transcriptome and translatome of Agrp neurons upon food deprivation can be seen in Figure 24. We observed the presence of many processes associated to endoplasmic reticulum regulations, specially associated to cellular stress. We performed a pathway analysis in *IPA* using the threshold of -log(FDR) = 2. The results are available in Figure 25. We can note the presence of processes already seen in the biological processes: the ones associated to the endoplasmic reticulum and stress.

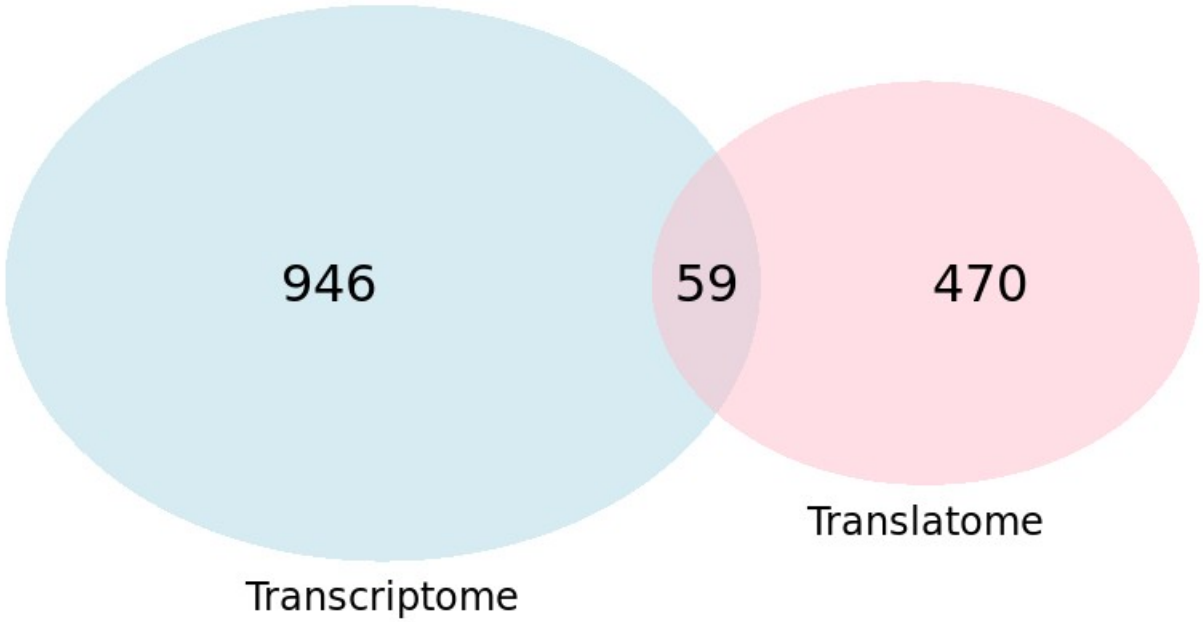



**Figure 23:** Venn diagram showing the differences in the genes that were found in the Transcriptome and Translatome.

We see the presence of leptin and other signaling pathways as well.

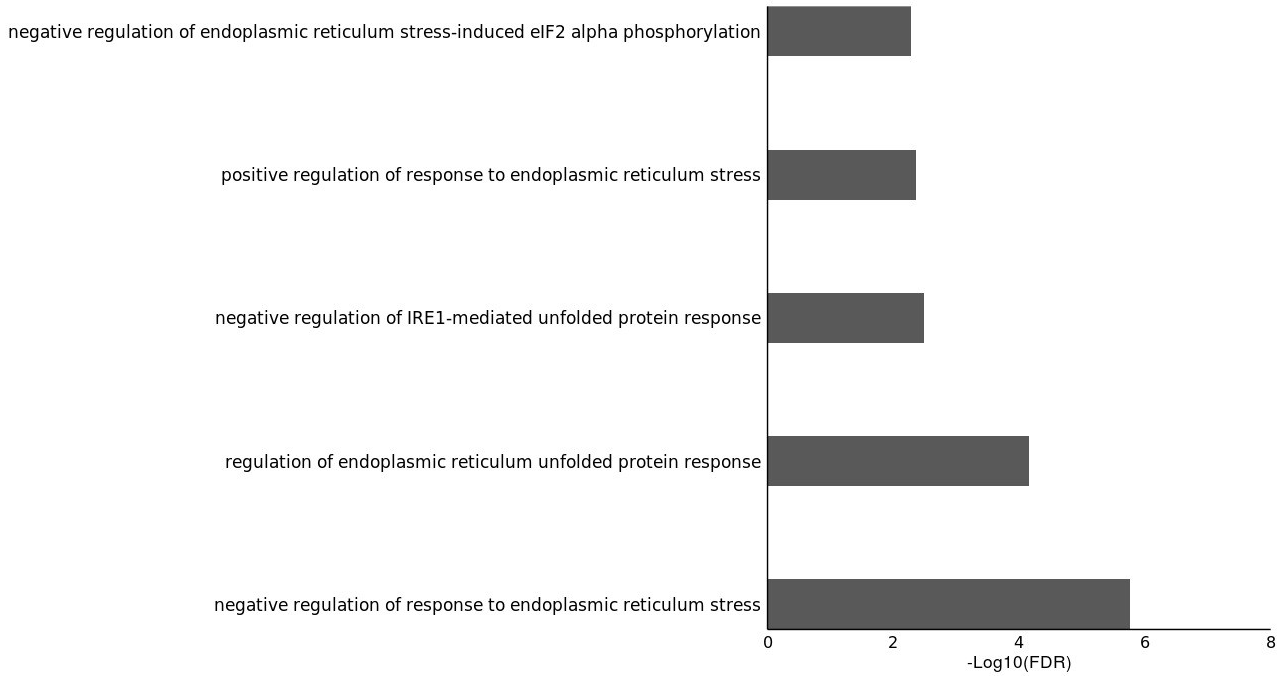



**Figure 24:** Bar plot on the main biological processes associated to the shared genes between transcriptome and translatome in *String*.

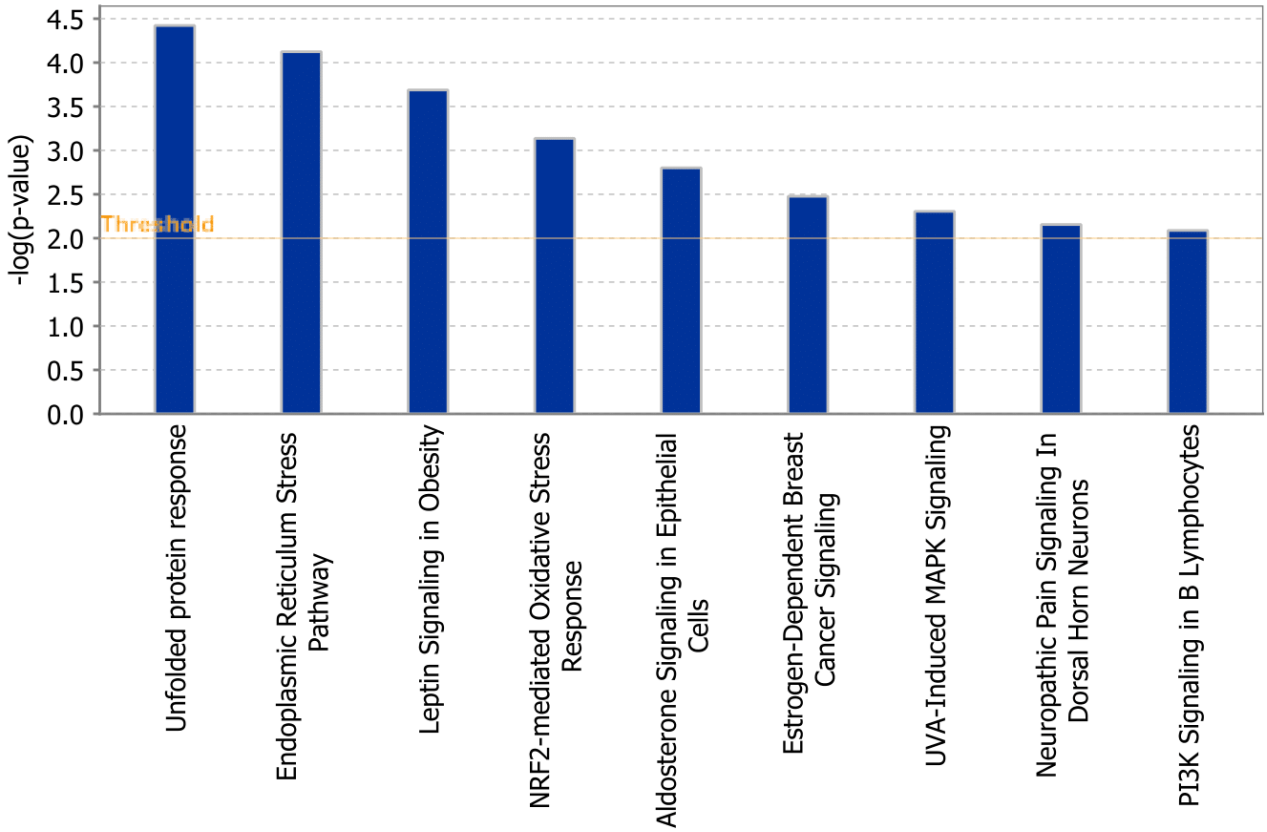



**Figure 25:** Bar plot on the pathways associated to the shared genes between transcriptome and translatome upon food deprivation in IPA.

## DISCUSSION

In this work, we have analyzed the transcriptional and translational profile of Agrp neurons utilizing both public datasets of sc-RNA-Seq and the RiboTag strategy coupled with RNA-Seq. We designed a processing pipeline suitable to analyze and identify genes and pathways functionally relevant for Agrp neurons in a scenario of food deprivation at the transcriptome and translatome levels. The results obtained after the data processing were enough to answer the general and the three specific goals we have initially proposed.

Our first aim (specific aim 2.1) was to evaluate changes in genes expressed in the transcriptome and translatome of Agrp neurons during food deprivation. For the transcriptome, we found a total of 1005 genes impacted by food deprivation (**Figure 12**). The majority of them changed within a magnitude of 0.5 to 1.5 with maximum values of increase around 3. For the translatome, we found a total of 529 genes (**Figure 20**) that changed their expression within the range of LFC -4 to 4, with majority ranging from -2 to 2.

Our second aim (specific aim 2.2) regarded the identification of pathways relevant upon food deprivation at the transcriptome and translatome levels. The pathways identified in the transcriptome (**Figure 13 and 14**) and in the translatome (**Figure 21 and 22**) revealed a slight distinction of both levels. In the transcriptome, the most relevant pathways were associated to synaptogenesis, neuroinflammation, long-term synaptic regulation and glutamate signaling. Synaptic plasticity in the hypothalamus (Frankfurt et al., 1990), and in the Agrp neurons specifically, has been described (Dietrich and Horvath, 2013). Literature findings indicate that a negative energy balance is capable of inducing an increase in the excitatory synaptic inputs to Agrp neurons (Yang et al., 2011) and an increase in the number of dendritic spines in Agrp neurons (Liu et al., 2012), demonstrating the importance of the glutamatergic signaling and synaptic plasticity induced by food deprivation in these neurons (Liu et al., 2012). These findings correlate with our transcriptome results where several processes for synapse regulation and glutamatergic signaling were observed. In the translatome, we observed a prevalence of pathways and biological processes associated to endoplasmic reticulum (ER) stress, leptin signaling and regulation of

circadian rhythms. Our translatome results correlated with findings in literature that have shown the relationship of food deprivation and neuroendocrine signaling with circadian clock regulation (Bechtold and Loudon, 2013); the connection of the circadian clock with energy levels via cAMP signaling (Lee and Kim, 2013) and, more recently, the association of circadian clock and Agrp neuron regulation of the feeding behavior (Cedernaes et al., 2019; Chen, 2019). Interestingly, there are evidence in the literature showing that stress (i.e. food deprivation) can lead to protein unfolding and ER stress, as well as a connection between both ER stress and leptin signaling pathways (Ramírez and Claret, 2015). In line with these findings, it has been demonstrated that ER stress in the hypothalamus of lean mice leads to leptin resistance and increase in the *Agrp/Npy* levels (Ozcan et al., 2009).

Our third aim (specific aim 2.3) was to identify possible genes shared by the transcriptome and translatome of Agrp neurons during food deprivation. We have described 59 genes (**Table S2**) and analyzed their biological roles. We observed that the majority of them relate to ER stress (protein unfolding response) and leptin signaling. Together with the findings described for aim 2.2, it is feasible to believe that food deprivation impacts the sensibility to leptin in Agrp neurons at both transcriptome and translatome levels, since protein unfolding process is identified in both instances and genes involved in leptin signaling were identified at the translatome and shared transcriptome-translatome levels.

Importantly, challenges in the data processing of sc-RNA-Seq datasets were found and we proposed alternative strategies to analyze the transcriptome of Agrp neurons at the single cell level. First, datasets created using the Drop-Seq protocol are ideal for identification of cell-type diversity at the cost of yield of mRNA captured per cell (Dal Molin and Di Camillo, 2018). Considering that both sc-RNA-Seq datasets used the ARC as input for tissue dissociation and the size of Agrp neuronal population in the ARC, we obtained a limited number of single Agrp neurons for the transcriptome definition. Nevertheless, we took advantage of the MNN method (Haghverdi et al., 2018) to identify Agrp neurons shared between the two analyzed datasets in order to maximize the number of single neurons analyzed in this study. Second, we were unable to perform a differential expression test in the isolated set of Agrp neurons in the transcriptome data. Upon characterization of the variables driving

the expression variability of our data (**Figure S1**), neither the origin of the dataset nor the feeding condition were sufficient to explain the differences observed between cells. In differential expression tests, the desired biological variable needs to drive the variability present in the dataset. When this is not the case, it is important to account for a possible technical variable (also called batch effect), known or unknown, that might be driving the variability of the dataset. Once the *origin dataset* variable was not sufficient to explain the technical variability of our data, we sought to investigate hidden batch effects with the *Scater* R package (**Figure S2**) and the *SVA* package. However, both methods were not able to identify a variable responsible for driving the observed batch effect and we decided to perform an enrichment-based strategy to identify genes affected by food deprivation in the transcriptome.

Altogether, our observations of the transcriptional and translational profile of Agrp neurons could indicate that these cells efficiently respond to a negative energy balance caused by food deprivation at both transcriptome and translatome levels. The transcriptome would lead to a response to the lack of food via an increase in the formation of synapse inputs, in a way to facilitate and sustain Agrp neuronal activation. On the other hand, the translatome would lead to direct changes at the protein level in genes known to regulate cellular stress and also in circadian clock regulation and its downstream genes, i.e. transcription factors associated to synaptic plasticity (Hannou et al., 2018). These distinct changes in Agrp neurons would be important to regulate the organism's response to food deprivation.

## CONCLUSIONS

We have successfully processed Drop-Seq and RiboTag data and characterized the transcriptome and translatome of Agrp neurons upon food deprivation. A clear separation of biological processes and pathways to each level (Transcription/Translation) was observed, whereas specific processes at both levels were also present. We observed the prevalence of synaptic plasticity changes associated to the transcriptome and the prevalence of circadian clock processes associated to the translatome. The genes shared by the transcriptome and translatome were associated to ER stress and leptin response. For future analysis, it would be important to solve the data variability issue in a way that DE analysis could be performed.

## REFERENCES


1. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
2. Aponte, Y., Atasoy, D., and Sternson, S.M. (2011). AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. Nat. Neurosci. *14*, 351–355.
3. Ashford, M.L.J., Spanswick, D., Smith, M.A., Groppi, V.E., and Logan, S.D. (1997). Leptin inhibits hypothalamic neurons by activation of ATP-sensitive potassium channels. Nature *390*, 521–525.
4. Bechtold, D.A., and Loudon, A.S.I. (2013). Hypothalamic clocks and rhythms in feeding behaviour. Trends Neurosci. *36*, 74–82.
5. Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. *20*, 484–496.
6. Cansell, C., Denis, R.G.P., Joly-Amado, A., Castel, J., and Luquet, S. (2012). Arcuate AgRP neurons and the regulation of energy balance. Front. Endocrinol. (Lausanne). *3*, 169.
7. Cedernaes, J., Huang, W., Ramsey, K.M., Waldeck, N., Cheng, L., Marcheva, B., Omura, C., Kobayashi, Y., Peek, C.B., Levine, D.C., et al. (2019). Transcriptional Basis for Rhythmic Control of Hunger and Metabolism within the AgRP Neuron. Cell Metab. *0*, 1–14.
8. Chen, Z. (2019). Temporal Control of Appetite by AgRP Clocks. Cell Metab. *29*, 1022–1023.
9. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.
10. Chen, R., Wu, X., Jiang, L., and Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. Cell Rep. *18*, 3227–3241.
11. Chronwall, B.M. (1985). Anatomy and physiology of the neuroendocrine arcuate nucleus. Peptides *6*, 1–11.
12. Ciofi, P. (2011). The arcuate nucleus as a circumventricular organ in the mouse. Neurosci. Lett. *487*, 187–190.

13. Cowley, M.A., Smith, R.G., Diano, S., Tschöp, M., Pronchuk, N., Grove, K.L., Strasburger, C.J., Bidlingmaier, M., Esterman, M., Heiman, M.L., et al. (2003). The distribution and mechanism of action of ghrelin in the CNS demonstrates a novel hypothalamic circuit regulating energy homeostasis. Neuron *37*, 649–661.

14. Dal Molin, A., and Di Camillo, B. (2018). How to design a single-cell RNA-sequencing experiment: pitfalls, challenges and perspectives. Brief. Bioinform. 1–11.

15. Dietrich, M.O., and Horvath, T.L. (2013). Hypothalamic control of energy balance: Insights into the role of synaptic plasticity. Trends Neurosci. *36*, 65–73.

16. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

17. Frankfurt, M., Gould, E., Woolley, C.S., and McEwen, B.S. (1990). Gonadal Steroids Modify Dendritic Spine Density in Ventromedial Hypothalamic Neurons: A Golgi Study in the Adult Rat. Neuroendocrinology *51*, 530–535.

18. Franzén, O., Gan, L.-M., and Björkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database *2019*, 1–9.

19. Gropp, E., Shanabrough, M., Borok, E., Xu, A.W., Janoschek, R., Buch, T., Plum, L., Balthasar, N., Hampel, B., Waisman, A., et al. (2005). Agouti-related peptide–expressing neurons are mandatory for feeding. Nat. Neurosci. *8*, 1289–1291.

20. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427.

21. Hannou, L., Roy, P.G., Ballester Roig, M.N., and Mongrain, V. (2018). Transcriptional control of synaptic components by the clock machinery. Eur. J. Neurosci. 0–2.

22. Havrankova, J., Roth, J., and Brownstein, M. (1978). Insulin receptors are widely distributed in the central nervous system of the rat. Nature *272*, 827–829.

23. Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics *107*, 1–8.

24. Henry, F.E., Sugino, K., Tozer, A., Branco, T., and Sternson, S.M. (2015). Cell type-specific transcriptomics of hypothalamic energy-sensing neuron responses to weight-loss. Elife *4*, 1–30.

25. Hossain, M.A., Claggett, J.M., Edwards, S.R., Shi, A., Pennebaker, S.L., Cheng, M.Y., Hasty, J., and Johnson, T.L. (2016). Posttranscriptional Regulation of Gcr1 Expression and Activity Is Crucial for Metabolic Adjustment in Response to Glucose Availability. Mol. Cell *62*, 346–358.

26. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

27. King, H.A., and Gerber, A.P. (2016). Translatome profiling: methods for genome-scale analysis of mRNA translation. Brief. Funct. Genomics *15*, 22–31.

28. Krashes, M.J., Koda, S., Ye, C., Rogan, S.C., Adams, A.C., Cusher, D.S., Maratos-Flier, E., Roth, B.L., and Lowell, B.B. (2011). Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. J. Clin. Invest. *121*, 1424–1428.

29. Kuleshov, M. V, Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-7.

30. Lee, Y., and Kim, E.K. (2013). AMP-activated protein kinase as a key molecular link between metabolism and clockwork. Exp. Mol. Med. *45*, e33-7.

31. Leek, J.T. (2014). svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. *42*, e161–e161.

32. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

33. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

34. Lister, R., Gregory, B.D., and Ecker, J.R. (2009). Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Curr. Opin. Plant Biol. *12*, 107–118.

35. Liu, T., Kong, D., Shah, B.P., Ye, C., Koda, S., Saunders, A., Ding, J.B., Yang, Z., Sabatini, B.L., and Lowell, B.B. (2012). Fasting Activation of AgRP Neurons Requires NMDA Receptors and Involves Spinogenesis and Increased Excitatory Tone. Neuron *73*, 511–522.

36. López-Maury, L., Marguerat, S., and Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat. Rev. Genet. *9*, 583–593.

37. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–34.

38. Lun, A., and Risso, D. (2019). SingleCellExperiment: S4 Classes for Single Cell Data.

39. Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research *5*, 2122.

40. Luquet, S., Perez, F.A., Hnasko, T.S., and Palmiter, R.D. (2005). NPY / AgRP Neurons Are Essential for Feeding in Adult Mice but Can Be Ablated in Neonates. Science (80-. ). *310*, 683–686.

41. VAN DER MAATEN, L., AND, and HINTON, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. 2579–2605.

42. Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets[Supplemental Information Highly. Cell *161*, 1202–1214.

43. Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. Cell. Mol. Life Sci. *67*, 569–579.

44. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

45. McGettigan, P.A. (2013). Transcriptomics in the RNA-seq era. Curr. Opin. Chem. Biol. *17*, 4–11.

46. McHale, C.M., Zhang, L., Thomas, R., and Smith, M.T. (2013). Analysis of the transcriptome in molecular epidemiology studies. Environ. Mol. Mutagen. *54*, 500–517.

47. Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31–46.

48. Mizuno, T.M. (2004). Hypothalamic Agouti-Related Protein Messenger Ribonucleic Acid Is Inhibited by Leptin and Stimulated by Fasting. Endocrinology *140*, 814–817.

49. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

50. Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., and Yuan, J. (2017). Single-Cell RNA Sequencing: Unraveling the Brain One Cell at a Time. Trends Mol. Med. *23*, 563–576.

51. Ozcan, L., Ergin, A.S., Lu, A., Chung, J., Sarkar, S., Nie, D., Myers, M.G., and Ozcan, U. (2009). Endoplasmic reticulum stress plays a central role in development of leptin resistance. Cell Metab. *9*, 35–51.

52. Ramírez, S., and Claret, M. (2015). Hypothalamic ER stress: A bridge between leptin resistance and obesity. FEBS Lett. *589*, 1678–1687.

53. Sanz, E., Yang, L., Su, T., Morris, D.R., McKnight, G.S., and Amieux, P.S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. Proc. Natl. Acad. Sci. *106*, 13939–13944.

54. Schaefke, B., Sun, W., Li, Y.-S., Fang, L., and Chen, W. (2018). The evolution of posttranscriptional regulation. Wiley Interdiscip. Rev. RNA *9*, e1485.

55. Schwartz, M.W., and Seeley, R.J. (1997). Neuroendocrine Responses to Starvation and Weight Loss. N. Engl. J. Med. *336*, 1802–1811.

56. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools : modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. 491–499.

57. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

58. Tebaldi, T., Re, A., Viero, G., Pegoretti, I., Passerini, A., Blanzieri, E., and Quattrone, A. (2012). Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. BMC Genomics *13*, 220.

59. Van Den Top, M., Lee, K., Whyment, A.D., Blanks, A.M., and Spanswick, D. (2004). Orexigen-sensitive NPY/AgRP pacemaker neurons in the hypothalamic arcuate nucleus. Nat. Neurosci. *7*, 493–494.

60. Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*, 57–63.

61. Willesen, M.G., Kristensen, P., and Rømer, J. (1999). Co-localization of growth hormone secretagogue receptor and NPY mRNA in the arcuate nucleus of the rat. Neuroendocrinology *70*, 306–316.

62. Yang, Y., Atasoy, D., Su, H.H., and Sternson, S.M. (2011). Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. Cell *146*, 992–1003.

63. Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., et al. (2014). A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. PLoS One *9*, e103207.

## SUPPLEMENTAL MATERIAL

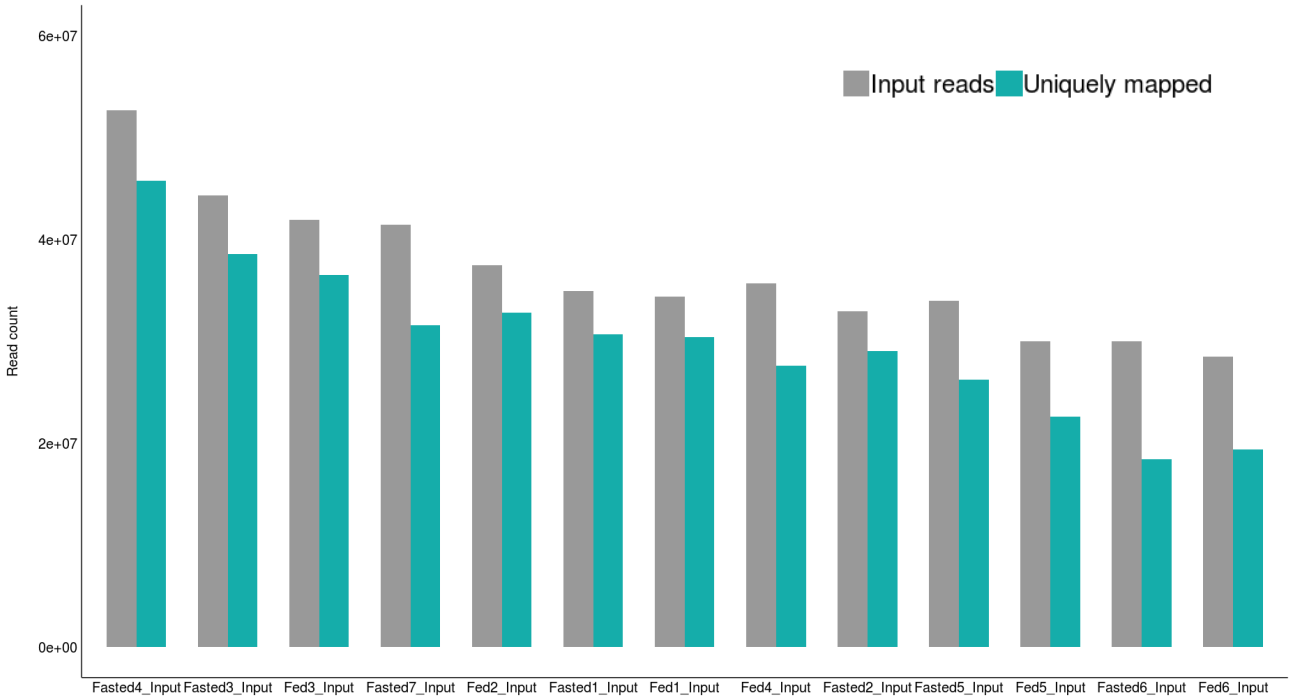



**Figure S1:** Histogram on the amount of input and uniquely mapped reads across all the Input samples generated under the RiboTag technique.

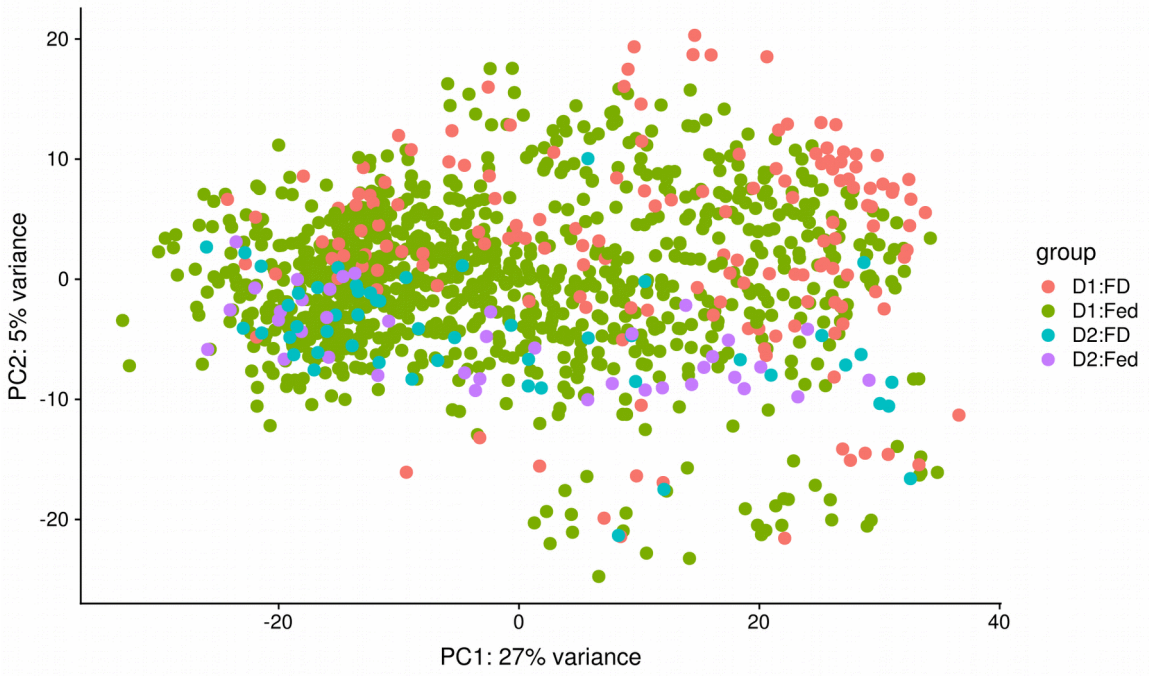



**Figure S2:** PCA on the raw data after variance stabilizing transformation showing that there is no clear separation of the cells according to their origin datasets or feeding conditions.

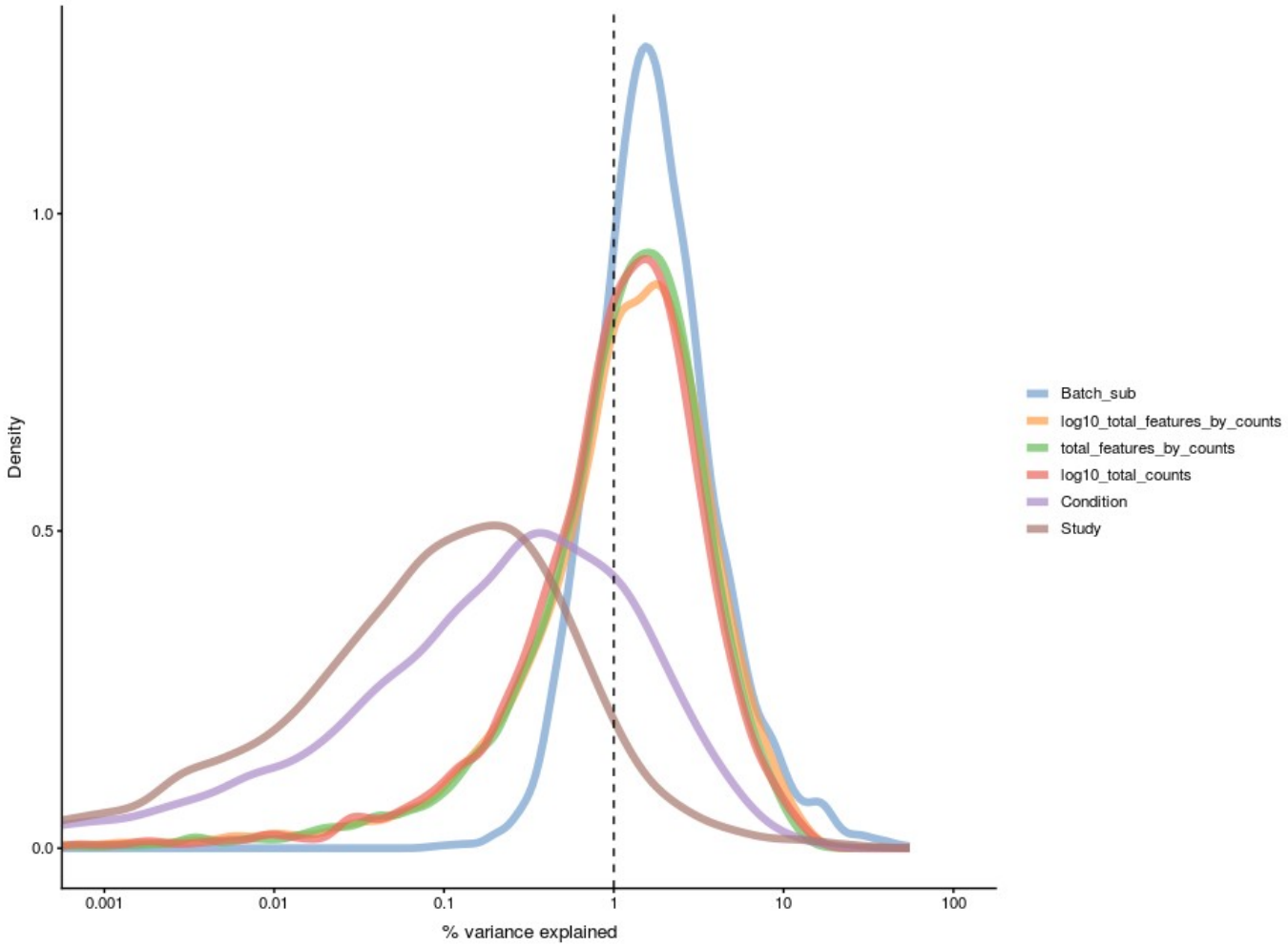



**Figure S3:** Plot showing the variables that explain the variability in the data, created with the *plotExplanatoryVariables* function (Scater 1.10.1) in R. The plot shows that although the Batch inside each dataset (Batch_sub) is the variable that explains most of the variance, a very small percentage of the variability is explained by the described variable. This means that there is a hidden batch effect we were not able to find.

**Table S1:** List of all samples used by dataset and protocol.

| Sample | Condition | Sample Accession Number (GSM) | Study Accession Number (GSE) | Dataset | RNA-Seq Protocol |
|---|---|---|---|---|---|
| Fed1 | Fed (*ad libtum*) | Not public | Not public | Dietrich Lab | RiboTag+ RNA-Seq |
| Fed2 | | | | | |
| Fed3 | | | | | |
| Fed4 | | | | | |
| Fed5 | | | | | |
| Fed6 | | | | | |
| Fed1_Input | | | | | |
| Fed2_Input | | | | | |
| Fed3_Input | | | | | |
| Fed4_Input | | | | | |
| Fed5_Input | | | | | |
| Fed6_Input | | | | | |
| Fasted1 | Food deprivation (16h) | | | | |
| Fasted2 | | | | | |
| Fasted3 | | | | | |
| Fasted4 | | | | | |
| Fasted5 | | | | | |
| Fasted6 | | | | | |
| Fasted1_Input | | | | | |
| Fasted2_Input | | | | | |
| Fasted3_Input | | | | | |
| Fasted4_Input | | | | | |
| Fasted5_Input | | | | | |
| Fasted6_Input | | | | | |
| Fed1 | Fed (*ad libtum*) | GSM2452122 | GSE93374 | Campbell | Single-Cell (Drop-Seq) |
| Fed2 | | GSM2452123 | | | |
| Fed3 | | GSM2452124 | | | |
| Fed4 | | GSM2452129 | | | |
| Fed5 | | GSM2452131 | | | |
| Fasted1 | Food deprivation (24h) | GSM2452127 | | | |
| Fasted2 | | GSM2452130 | | | |
| Fasted3 | | GSM2452132 | | | |
| Fed6 | Fed (*ad libtum*) | GSM2333583<br>GSM2333584<br>GSM2333585 | GSE87544 | Chen | Single-Cell (Drop-Seq) |
| Fed7 | | GSM2333586<br>GSM2333587 | | | |
| Fed8 | | GSM2333591<br>GSM2333592 | | | |
| Fasted4 | Food deprivation (24h) | GSM2333579<br>GSM2333580<br>GSM2333581<br>GSM2333582 | | | |
| Fasted5 | | GSM2333588<br>GSM2333589<br>GSM2333590 | | | |
| Fasted6 | | GSM2333593 | | | |

**Table S2:** Description based on the String database of the 59 genes shared by the transcriptome and translatome.

| Gene Name | Term Description |
|---|---|
| *Hlf* | Hepatic leukemia factor; Belongs to the bZIP family. PAR subfamily (295 aa) |
| *Agrp* | Agouti-related protein; Plays a role in weight homeostasis. Involved in the control of feeding behavior through the central melanocortin system. Acts as alpha melanocyte-stimulating hormone antagonist by inhibiting cAMP production mediated by stimulation of melanocortin receptors within the hypothalamus and adrenal gland. Has very low activity with MC5R. Is an inverse agonist for MC3R and MC4R being able to suppress their constitutive activity (By similarity). It promotes MC3R and MC4R endocytosis in an arrestin-dependent manner (By similarity) (131 aa) |
| *Dnajb9* | DnaJ homolog subfamily B member 9; Involved in endoplasmic reticulum-associated degradation (ERAD) of misfolded proteins. Acts as a co-chaperone with an Hsp70 protein (222 aa) |
| *N4bp2l1* | NEDD4 binding protein 2-like 1 (238 aa) |
| *Hist1h2bc* | Histone cluster 1, H2bc (126 aa) |
| *Btg2* | Protein BTG2; Anti-proliferative protein; the function is mediated by association with deadenylase subunits of the CCR4-NOT complex. Activates mRNA deadenylation in a CNOT6 and CNOT7-dependent manner. In vitro can inhibit deadenylase activity of CNOT7 and CNOT8. Involved in cell cycle regulation. Could be involved in the growth arrest and differentiation of the neuronal precursors. Modulates transcription regulation mediated by ESR1. Involved in mitochondrial depolarization and neurite outgrowth (By similarity); Belongs to the BTG family (158 aa) |
| *Slc35b1* | Solute carrier family 35 member B1; Probable sugar transporter; Belongs to the nucleotide-sugar transporter family. SLC35B subfamily (322 aa) |
| *Fos* | Proto-oncogene c-Fos; Nuclear phosphoprotein which forms a tight but non- covalently linked complex with the JUN/AP-1 transcription factor. On TGF-beta activation, forms a multimeric SMAD3/SMAD4/JUN/FOS complex, at the AP1/SMAD-binding site to regulate TGF-beta- mediated signaling (By similarity). Has a critical function in regulating the development of cells destined to form and maintain the skeleton. It is thought to have an important role in signal transduction, cell proliferation and differentiation. In growing cells, activates phospholipid synthesis, possibly by activating CDS1 an [...] (380 aa) |
| *Pcsk1* | Serine/threonine-protein kinase PLK2; Tumor suppressor serine/threonine-protein kinase involved in synaptic plasticity, centriole duplication and G1/S phase transition. Polo-like kinases act by binding and phosphorylating proteins are that already phosphorylated on a specific motif recognized by the POLO box domains. Phosphorylates CENPJ, NPM1, RAPGEF2, RASGRF1, SNCA, SIPA1L1 and SYNGAP1. Plays a key role in synaptic plasticity and memory by regulating the Ras and Rap protein signaling- required for overactivity-dependent spine remodeling by phosphorylating the Ras activator RASGRF1 an [...] (682 aa) |
| *Dnajc3* | DnaJ homolog subfamily C member 3; Involved in the unfolded protein response (UPR) during endoplasmic reticulum (ER) stress. Acts as a negative regulator of the EIF2AK4/ GCN2 kinase activity by preventing the phosphorylation of eIF-2-alpha at 'Ser-52' and hence attenuating general protein synthesis under ER stress, hypothermic and amino acid starving stress conditions. Co-chaperone of HSPA8/HSC70, it stimulates its ATPase activity. May inhibit both the autophosphorylation of EIF2AK2/PKR and the ability of EIF2AK2 to catalyze phosphorylation of the EIF2A. May inhibit EIF2AK3/PERK activit [...] (504 aa) |

| | |
|---|---|
| *Ifitm3* | Interferon-induced transmembrane protein 3; IFN-induced antiviral protein which inhibits the entry of viruses to the host cell cytoplasm, permitting endocytosis, but preventing subsequent viral fusion and release of viral contents into the cytosol. Active against multiple viruses, including influenza A virus, SARS coronavirus (SARS-CoV), Marburg virus (MARV) and Ebola virus (EBOV), Dengue virus (DNV), West Nile virus (WNV) and human immunodeficiency virus type 1 (HIV-1). Can inhibit- influenza virus hemagglutinin protein-mediated viral entry, MARV and EBOV GP1,2-mediated viral entry an [...] (137 aa) |
| *Tram1* | Translocating chain-associated membrane protein 1; Stimulatory or required for the translocation of secretory proteins across the ER membrane (374 aa) |
| *Ogfrl1* | Opioid growth factor receptor-like 1 (464 aa) |
| *Ptprn* | Receptor-type tyrosine-protein phosphatase-like N; Plays a role in vesicle-mediated secretory processes. Required for normal accumulation of secretory vesicles in hippocampus, pituitary and pancreatic islets. Required for the accumulation of normal levels of insulin-containing vesicles and preventing their degradation. Plays a role in insulin secretion in response to glucose stimuli. Required for normal accumulation of the neurotransmitters norepinephrine, dopamine and serotonin in the brain. In females, but not in males, required for normal accumulation and secretion of pituitary horm [...] (981 aa) |
| *Rgs4* | Regulator of G-protein signaling 4; Inhibits signal transduction by increasing the GTPase activity of G protein alpha subunits thereby driving them into their inactive GDP-bound form. Activity on G(z)-alpha is inhibited by phosphorylation of the G-protein. Activity on G(z)-alpha and G(i)-alpha-1 is inhibited by palmitoylation of the G-protein (By similarity) (205 aa) |
| *Hsd17b12* | Very-long-chain 3-oxoacyl-CoA reductase; Catalyzes the second of the four reactions of the long- chain fatty acids elongation cycle. This endoplasmic reticulum- bound enzymatic process, allows the addition of two carbons to the chain of long- and very long-chain fatty acids/VLCFAs per cycle. This enzyme has a 3-ketoacyl-CoA reductase activity, reducing 3- ketoacyl-CoA to 3-hydroxyacyl-CoA, within each cycle of fatty acid elongation. Thereby, it may participate in the production of VLCFAs of different chain lengths that are involved in multiple biological processes as precursors of memb [...] (312 aa) |
| *Pdia3* | Protein disulfide-isomerase A3; Protein disulfide isomerase associated 3 (505 aa) |
| *Npy* | Pro-neuropeptide Y; NPY is implicated in the control of feeding and in secretion of gonadotrophin-release hormone; Belongs to the NPY family (97 aa) |
| *Rpn1* | Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1; Essential subunit of the N-oligosaccharyl transferase (OST) complex which catalyzes the transfer of a high mannose oligosaccharide from a lipid-linked oligosaccharide donor to an asparagine residue within an Asn-X-Ser/Thr consensus motif in nascent polypeptide chains; Belongs to the OST1 family (608 aa) |
| *Nucb2* | Nucleobindin-2; Calcium-binding protein. May have a role in calcium homeostasis; Belongs to the nucleobindin family (420 aa) |
| *Pcsk1n* | ProSAAS; May function in the control of the neuroendocrine secretory pathway. Proposed be a specific endogenous inhibitor of PCSK1. ProSAAS and Big PEN-LEN, both containing the C-terminal inhibitory domain, but not the processed peptides reduce PCSK1 activity in the endoplasmic reticulum and Golgi. It reduces the activity of the 87 kDa form but not the autocatalytically derived 66 kDa form of PCSK1. Subsequent processing of proSAAS may eliminate the inhibition. Slows down convertase-mediated processing of proopiomelanocortin and proenkephalin. May control the intracellular |

| | |
|---|---|
| | timing of PC [...] (258 aa) |
| *Lrrn3* | Leucine rich repeat protein 3, neuronal (707 aa) |
| *Herc1* | HECT and RLD domain-containing E3 ubiquitin protein ligase family member 1; Hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1 (CHC1)-like domain (RLD) 1 (4859 aa) |
| *Ptplad1* | Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3; Catalyzes the third of the four reactions of the long- chain fatty acids elongation cycle. This endoplasmic reticulum-bound enzymatic process, allows the addition of two carbons to the chain of long- and very long-chain fatty acids/VLCFAs per cycle. This enzyme catalyzes the dehydration of the 3-hydroxyacyl-CoA intermediate into trans-2,3-enoyl-CoA, within each cycle of fatty acid elongation. Thereby, it participates in the production of VLCFAs of different chain lengths that are involved in multiple biological processes as precurs [...] (362 aa) |
| *Fam107a* | Family with sequence similarity 107, member A (144 aa) |
| *Ostc* | Oligosaccharyltransferase complex subunit OSTC; May act as substrate-specific component of the N- oligosaccharyl transferase (OST) complex which catalyzes the transfer of a high mannose oligosaccharide from a lipid-linked oligosaccharide donor to an asparagine residue within an Asn-X- Ser/Thr consensus motif in nascent polypeptide chains. May be involved in N-glycosylation of APP (amyloid-beta precursor protein). Can modulate gamma-secretase cleavage of APP by enhancing endoprotelysis of PSEN1; Belongs to the OSTC family (149 aa) |
| *Abca5* | ATP-binding cassette sub-family A member 5; May play a role in the processing of autolysosomes (1642 aa) |
| *Ppp1r15a* | Protein phosphatase 1 regulatory subunit 15A; Recruits the serine/threonine-protein phosphatase PP1 to dephosphorylate the translation initiation factor eIF-2A/EIF2S1, thereby reversing the shut-off of protein synthesis initiated by stress-inducible kinases and facilitating recovery of cells from stress. Down-regulates the TGF-beta signaling pathway by promoting dephosphorylation of TGFB1 by PP1. May promote apoptosis by inducing TP53 phosphorylation on 'Ser-15'. In case of infection with vesicular stomatitis virus (VSV), impairs viral replication; Belongs to the PPP1R15 family (657 aa) |
| *Pdia6* | Protein disulfide-isomerase A6; May function as a chaperone that inhibits aggregation of misfolded proteins. Negatively regulates the unfolded protein response (UPR) through binding to UPR sensors such as ERN1, which in turn inactivates ERN1 signaling (By similarity). May also regulate the UPR via the EIF2AK3 UPR sensor (By similarity). Plays a role in platelet aggregation and activation by agonists such as convulxin, collagen and thrombin (By similarity); Belongs to the protein disulfide isomerase family (445 aa) |
| *Xbp1* | X-box-binding protein 1; Functions as a transcription factor during endoplasmic reticulum stress by regulating the unfolded protein response (UPR). Required for cardiac myogenesis and hepatogenesis during embryonic development and the development of secretory tissues such as exocrine pancreas and salivary gland. Involved in differentiation of B lymphocytes to plasma cells and production of immunoglobulins. Modulates the cellular response to ER stress in a PIK3R-dependent manner. Binds to the cis-acting X box present in the promoter regions of major histocompatibility complex class II g [...] (267 aa) |
| *Pik3r1* | Phosphatidylinositol 3-kinase regulatory subunit alpha; Binds to activated (phosphorylated) protein-Tyr kinases, through its SH2 domain, and acts as an adapter, mediating the association of the p110 catalytic unit to the plasma membrane. Necessary for the insulin-stimulated increase in glucose uptake and glycogen synthesis in insulin-sensitive tissues. Plays an important role in signaling in response to FGFR1, FGFR2, |

| | |
|---|---|
| | FGFR3, FGFR4, KITLG/SCF, KIT, PDGFRA and PDGFRB. Likewise, plays a role in ITGB2 signaling (By similarity). Modulates the cellular response to ER stress by promoting nucle [...] (724 aa) |
| *Pam* | Peptidyl-glycine alpha-amidating monooxygenase; Bifunctional enzyme that catalyzes 2 sequential steps in C-terminal alpha-amidation of peptides. The monooxygenase part produces an unstable peptidyl(2-hydroxyglycine) intermediate that is dismutated to glyoxylate and the corresponding desglycine peptide amide by the lyase part. C-terminal amidation of peptides such as neuropeptides is essential for full biological activity (By similarity); In the N-terminal section; belongs to the copper type II ascorbate-dependent monooxygenase family (979 aa) |
| *Cd24a* | Signal transducer CD24; May have a pivotal role in cell differentiation of different cell types. May have a specific role in early thymocyte development. Signaling could be triggered by the binding of a lectin-like ligand to the CD24 carbohydrates, and transduced by the release of second messengers derived from the GPI-anchor. Modulates B-cell activation responses (By similarity). In association with SIGLEC10 may be involved in the selective suppression of the immune response to danger-associated molecular patterns (DAMPs) such as HMGB1, HSP70 and HSP90. Plays a role in the control of [...] (76 aa) |
| *Ier2* | Immediate early response gene 2 protein; DNA-binding protein that seems to act as a transcription factor (By similarity). Involved in the regulation of neuronal differentiation, acts upon JNK-signaling pathway activation and plays a role in neurite outgrowth in hippocampal cells (By similarity). May mediate with FIBP FGF-signaling in the establishment of laterality in the embryo (By similarity). Promotes cell motility, seems to stimulate tumor metastasis (By similarity) (221 aa) |
| *Lxn* | Latexin; Hardly reversible, non-competitive, and potent inhibitor of CPA1, CPA2 and CPA4 (By similarity). May play a role in inflammation; Belongs to the protease inhibitor I47 (latexin) family (222 aa) |
| *Scg2* | Secretogranin-2; Secretogranin-2 is a neuroendocrine secretory granule protein, which may be the precursor for other biologically active peptides (617 aa) |
| *Ezr* | Ezrin; Probably involved in connections of major cytoskeletal structures to the plasma membrane. In epithelial cells, required for the formation of microvilli and membrane ruffles on the apical pole. Along with PLEKHG6, required for normal macropinocytosis (By similarity) (586 aa) |
| *Junb* | Transcription factor jun-B; Transcription factor involved in regulating gene activity following the primary growth factor response. Binds to the DNA sequence 5'-TGA[CG]TCA-3'; Belongs to the bZIP family. Jun subfamily (344 aa) |
| *Cthrc1* | Collagen triple helix repeat-containing protein 1; May act as a negative regulator of collagen matrix deposition (245 aa) |
| *Erlec1* | Endoplasmic reticulum lectin 1; Probable lectin that binds selectively to improperly folded lumenal proteins. May function in endoplasmic reticulum quality control and endoplasmic reticulum-associated degradation (ERAD) of both non-glycosylated proteins and glycoproteins (By similarity) (483 aa) |
| *Herc2* | E3 ubiquitin-protein ligase HERC2; E3 ubiquitin-protein ligase that regulates ubiquitin-dependent retention of repair proteins on damaged chromosomes. Recruited to sites of DNA damage in response to ionizing radiation (IR) and facilitates the assembly of UBE2N and RNF8 promoting DNA damage-induced formation of 'Lys-63'-linked ubiquitin chains. Acts as a mediator of binding specificity between UBE2N and RNF8. Involved in the maintenance of RNF168 levels. E3 ubiquitin-protein ligase that promotes the ubiquitination and proteasomal degradation of XPA which influences the circadian oscill [...] (4836 aa) |

| | |
|---|---|
| *Sez6l* | Seizure 6-like protein; Candidate tumor suppressor gene. May contribute to specialized endoplasmic reticulum functions in neurons (963 aa) |
| *Gda* | Guanine deaminase; Catalyzes the hydrolytic deamination of guanine, producing xanthine and ammonia; Belongs to the metallo-dependent hydrolases superfamily. ATZ/TRZ family (454 aa) |
| *H13* | Minor histocompatibility antigen H13; Catalyzes intramembrane proteolysis of some signal peptides after they have been cleaved from a preprotein, resulting in the release of the fragment from the ER membrane into the cytoplasm. Required to generate lymphocyte cell surface (HLA-E) epitopes derived from MHC class I signal peptides. Involved in the intramembrane cleavage of the integral membrane protein PSEN1. Cleaves the integral membrane protein XBP1 isoform 1 in a DERL1/RNF139-dependent manner (By similarity). May play a role in graft rejection; Belongs to the peptidase A22B family (394 aa) |
| *Rasgrf2* | Ras-specific guanine nucleotide-releasing factor 2; Functions as a calcium-regulated nucleotide exchange factor activating both Ras and RAC1 through the exchange of bound GDP for GTP. Preferentially activates HRAS in vivo compared to RRAS based on their different types of prenylation. Functions in synaptic plasticity by contributing to the induction of long term potentiation (1188 aa) |
| *Rprm* | Protein reprimo; May be involved in the regulation of p53-dependent G2 arrest of the cell cycle. Seems to induce cell cycle arrest by inhibiting CDK1 activity and nuclear translocation of the CDC2 cyclin B1 complex; Belongs to the reprimo family (109 aa) |
| *Hspa5* | 78 kDa glucose-regulated protein; Plays a role in facilitating the assembly of multimeric protein complexes inside the endoplasmic reticulum. Involved in the correct folding of proteins and degradation of misfolded proteins via its interaction with DNAJC10, probably to facilitate the release of DNAJC10 from its substrate; Belongs to the heat shock protein 70 family (655 aa) |
| *Bmyc* | Protein B-Myc; Seems to act as an inhibitor of cellular proliferation (170 aa) |
| *Ufl1* | E3 UFM1-protein ligase 1; E3 protein ligase that mediates ufmylation, the covalent attachment of the ubiquitin-like modifier UFM1 to substrate proteins, a post-translational modification on lysine residues of proteins that may play a crucial role in a number of cellular processes. Mediates DDRGK1 ufmylation and may regulate the proteasomal degradation of DDRGK1 and CDK5RAP3 thereby modulating NF-kappa-B signaling. May also through TRIP4 ufmylation play a role in nuclear receptors-mediated transcription. May play a role in the unfolded protein response, mediating the ufmylation of multi [...] (793 aa) |
| *Nudt11* | Diphosphoinositol polyphosphate phosphohydrolase 3-alpha; Cleaves a beta-phosphate from the diphosphate groups in PP-InsP5 (diphosphoinositol pentakisphosphate), suggesting that it may play a role in signal transduction. Also able to catalyze the hydrolysis of dinucleoside oligophosphates, with Ap6A and Ap5A being the preferred substrates. The major reaction products are ADP and p4a from Ap6A and ADP and ATP from Ap5A. Also able to hydrolyze 5-phosphoribose 1-diphosphate; however, the relevance of such activity in vivo remains unclear; Belongs to the Nudix hydrolase family. DIPP subfamily (164 aa) |
| *Rab3b* | Ras-related protein Rab-3B; Protein transport. Probably involved in vesicular traffic (By similarity) (219 aa) |
| *Gem* | GTP-binding protein GEM; Could be a regulatory protein, possibly participating in receptor-mediated signal transduction at the plasma membrane. Has guanine nucleotide-binding activity but undetectable intrinsic GTPase activity; Belongs to the small GTPase superfamily. RGK family (295 aa) |
| *Dpp10* | Inactive dipeptidyl peptidase 10; Promotes cell surface expression of the potassium |

| | |
|---|---|
| | channel KCND2. Modulates the activity and gating characteristics of the potassium channel KCND2. Has no dipeptidyl aminopeptidase activity (Probable) (800 aa) |
| *Manf* | Mesencephalic astrocyte-derived neurotrophic factor; Selectively promotes the survival of dopaminergic neurons of the ventral mid-brain. Modulates GABAergic transmission to the dopaminergic neurons of the substantia nigra. Enhances spontaneous, as well as evoked, GABAergic inhibitory postsynaptic currents in dopaminergic neurons. Inhibits cell proliferation and endoplasmic reticulum (ER) stress-induced cell death (By similarity) (179 aa) |
| *Vgf* | VGF nerve growth factor inducible (617 aa) |
| *Fam46a* | Family with sequence similarity 46, member A |