Utrecht University

Hubrecht Institute
Developmental Biology and Stem Cell Research

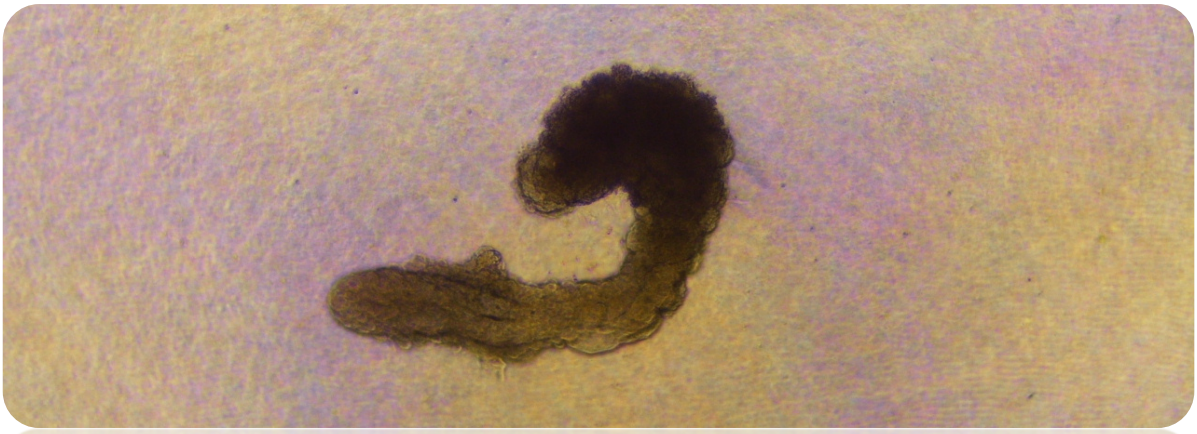
MSc. Cancer, Stem Cells and Developmental Biology
Graduate School of Life Sciences
Utrecht University

Developmental Biology and Stem Cell Research
Hubrecht Institute
KNAW

MAJOR RESEARCH INTERNSHIP

Translational gene regulation in mouse gastruloids

EIRINI DASKALAKI
Student number: 0181633



Supervisor: Professor Alexander van Oudenaarden
Daily Supervisor: Dr. Michael VanInsberghe
Hubrecht Institute
Utrecht

January 2023

Layman's summary

A cell is defined as the smallest unit of life. All multicellular organisms start as a single cell that undergoes multiple changes and develops into a finely structured complex organism. Therefore, it is only natural that the study of the biological processes taking place in a single cell should be studied thoroughly to better understand how all living beings and the world around us is made. To better understand how life is built, further investigating the changes that take place after fertilization of that single cell takes place is important and this is what the study of embryo development does.

Embryo development is a very complicated process and must be strictly controlled to ensure that all processes will be executed properly and at the right timing. Studies in the field of molecular biology that focus on how gene expression is regulated have shown that there are multiple different levels to achieve that. The end goal of those regulation mechanisms is to control all the biological processes that take place in a cell and are necessary for the proper development of an organism.

As it is described in the central dogma of biology, information is encoded in DNA in genes, which is then passed on to the form of RNA through the process of transcription and at the final stage RNA is translated into protein by the ribosome. So far, it was believed that gene expression is regulated mostly at the level of transcription, so most studies focused on that. There have been indications from multiple experiments that control also takes place at the level of translation. The aim of this project is thus to further explore that field during embryo development.

Having a proper model system for the study of every biological question is important. In this study we chose to use a novel system that has been recently developed and that is the gastruloid system. It has been developed from mouse stem cells, which have been guided through multiple protein factors and mechanical stimuli in the lab, to imitate the processes that drive the early stages of embryo development in the uterus. It is known that developmental processes that happen during embryogenesis are highly conserved across closely related species, meaning that they're very similar. Through the study of mouse gastruloid development we can get insight to the developmental processes that take place during mammal development.

Novel methods have been developed that allow us to quantify transcription and translation levels at a single cell resolution. These methods are called single cell VASA-seq and Ribo-seq and were used in this project to quantify and compare transcription and translation levels in single cells derived from the mouse gastruloids. The quantification and comparison are done by computationally analyzing the sequences of RNA and RNA-ribosomal complexes identified in our samples. During the analysis the different cell types that are present in our samples are identified and grouped together. The levels of RNA and active translation in those cells for specific genes are compared. Based on the central dogma of biology we expect that protein levels change according to how RNA levels do. In cases where protein levels do not correspond to the levels of RNA in the cells, we consider that regulation takes place at the level of translation and causes these differences.

We identified different groups of genes for which that was the case. Therefore, we concluded that there are groups of genes that are active during embryonic development in mice and their expression is controlled by changes that take place during translation. We concluded that those genes are mostly involved in processes that control cell growth and the development of the different cell types, as it is expected during embryogenesis.

Abstract

Mechanisms to regulate gene expression have been studied thoroughly throughout the years mostly on the transcriptional level. There have been examples though where the presence of translational control regulated important biological processes. In this project we investigated whether translational control plays an important role during early embryonic development. For that we used mouse embryonic stems cells and 120-hour mouse gastruloids that develop somites. The experiments were performed on a single cell level and two genome-wide sequencing methods were used to quantify transcription (scVASA-seq) and translation (scRibo-seq). With scVASA-seq, all types of RNA molecules were sequenced. With scRibo-seq, ribosomal footprints were created that capture actively translating ribosomes on mRNA molecules. From the analysis, cells from both techniques were integrated computationally, and eight different clusters of cell types were identified. The cell-type identities of these clusters were annotated based on marker gene expression. Changes in Translational Efficiency (ΔTE) between clusters were calculated and compared. We identified 86 genes that undergo translational control that further clustered into three groups. An enrichment analysis for biological processes of the genes in these groups provided further information about the biological procedures these genes are involved in. This work conducted the first genome-wide measurements of translation and translation efficiency during mouse gastrulation, providing insight into the role post-transcriptional regulation plays in this important developmental process.

Introduction

The three main processes that occur during embryonic development, which are cleavage, gastrulation, and organogenesis, are highly conserved across species. Gastrulation takes place during the early stages of development, during which the three germ layers are created through cell migrations. During gastrulation, reptiles, avians, and mammals develop three germ layers: the mesoderm, endoderm, and ectoderm. These different germ layers are located in different parts of the embryo and will give rise to different cell lineages while they further differentiate. At that stage the embryo develops a body plan and transitions from a two-dimensional surface to a three-dimensional multilayered structure called the gastrula, with dorsal/ventral and cranial/caudal axes and left/right symmetry [1,2]. After gastrulation is complete, cardiogenesis, neurulation, and somitogenesis start taking place in the that order. During somitogenesis, somites will start forming that will lead to the formation of cartilage, tendons, dermis and the skeletal muscles of the embryo [3–5].

In vitro models of early development are a powerful tool to study these developmental processes. Gastruloids are three dimensional aggregates of embryonic stem cells that produce an embryo-like organization and behave similar to embryos. It has been demonstrated that with the right stimuli during culturing, gastruloids can develop somites, cardiac structures, and even brain-like tissue [6], which are important for the body plan/axis establishment [7] and simulate in vivo embryotic processes on a transcriptional level [8]. Their ability to recapitulate these developmental processes, combined with their relative ease of manipulation, makes gastruloids an attractive model for investigating the gene regulatory networks that drive gastrulation and study mammalian development in vitro. Moreover, it has been shown that it is possible to make gastruloids from different species. The mouse gastruloid system has been studied more thoroughly and was the first one to be established [9]. Human gastruloids have also been developed from human ES cells [10].

Genome-wide single-cell technologies have been used to map RNA expression, gene regulatory networks, and epigenetic changes, and to identify the different cell types involved

in early mammalian development [11]. However, there is evidence in similar systems that translational regulation may also play an important role. For example, in mice, disrupting the translation of certain HOX genes results is the appearance of an extra set of ribs, amongst other developmental defects. This study showed that the Ribosomal Protein L38 (Rpl38) affects the 80S-mRNA complex formation on some HOX gene mRNAs. When mutations are induced to the Rpl38 gene, the translation of these HOX gene proteins is disrupted, while global protein synthesis levels do not change [12]. Another example of translation playing a major role in gene expression regulation is the period before and after the maternal-to-zygotic transition (MZT) in zygotes. The zygote's transcriptional machinery is mostly inactive, as protein production is not dependent on zygote RNA levels, but on proteins and RNA molecules that are passed on by the mother. It has previously been shown that during that stage of embryo development translational control plays a role for gene expression regulation. Specifically, changes in mRNA poly-(A) tail length were found to cause translational up or down regulation [13].

So far, genome-wide measurements of translational regulation during embryonic development have been restricted to systems where enough material can be obtained. Ribosome profiling measures mRNA molecules that are actively being translated [14–16]. In this technique, mRNA molecules are digested at the positions where they are not bound with ribosomes, generating ribosome-protected footprints (RPFs). Sequencing, mapping, and counting these footprints quantifies the instantaneous translation levels for each expressed gene. The rate of translation per mRNA provides a measure of the translation efficiency (TE) for that gene. Thus, by comparing the levels of translation to separate matched measurements of mRNA abundance between different conditions, it is possible to identify and quantify changes in translation efficiencies. However, due to the large sample input requirements demanded by ribosome profiling [15],the application of these measurements to early development has been restricted to systems where enough material can be obtained.

Recently developed technologies for measuring translation and transcription at the single-cell level provide an opportunity to measure translational regulation at a higher resolution than previously possible. The development of single cell Ribo-seq has made quantification of translation possible at a single cell level, making the acquisition of enough sample material no longer an issue [17]. VASA-seq allows us to quantify all RNA molecules present in a cell [18]. The advantage of this technique is that it captures all different types of transcripts regardless of whether or not they are polyadenylated. Compared to standard scRNA-seq, VASA-seq provides a more complete measurement of RNA abundance as it is more sensitive and captures non-polyadenylated transcripts such as histone genes. Additionally, as changing to poly-A tail length is a mechanism used to regulate translation efficiency [13], using VASA-seq removes this potential source of technical bias.

In this project we explored the role of post-transcriptional regulation in early embryonic development by applying single-cell RNA sequencing (VASA-seq) and single-cell ribosome profiling (scRibo-seq) first on mouse stem cells and later on cells isolated from 120 h mouse gastruloids. By computationally integrating the measurements of transcript abundance and translation levels, marker genes from prior publications were used to annotate the different cell types identified in the data. After that, genes were identified whose translational output could not be solely explained by the transcript abundances by comparing transcription and translation levels and their changes between clusters. Finally, the biological role of the genes that were differentially translated was further explored.

Results

Method overview

The experimental workflow consisted of making gastruloids (Fig. 1A) dissociating them to a single-cell suspension, FACS sorting the single cells into wells of 384-well plates and using these single cells to perform scRibo-seq (Fig. 1B) and VASA-seq (Fig. 1C). The end product of these techniques is single-cell libraries. The final part of this project is the sequencing of these libraries. The data retrieved are later analyzed to cluster cells into groups of closely related cell types and examine changes in translation and transcription for those to identify possible cases of post-transcriptional regulation (Fig. 1D).

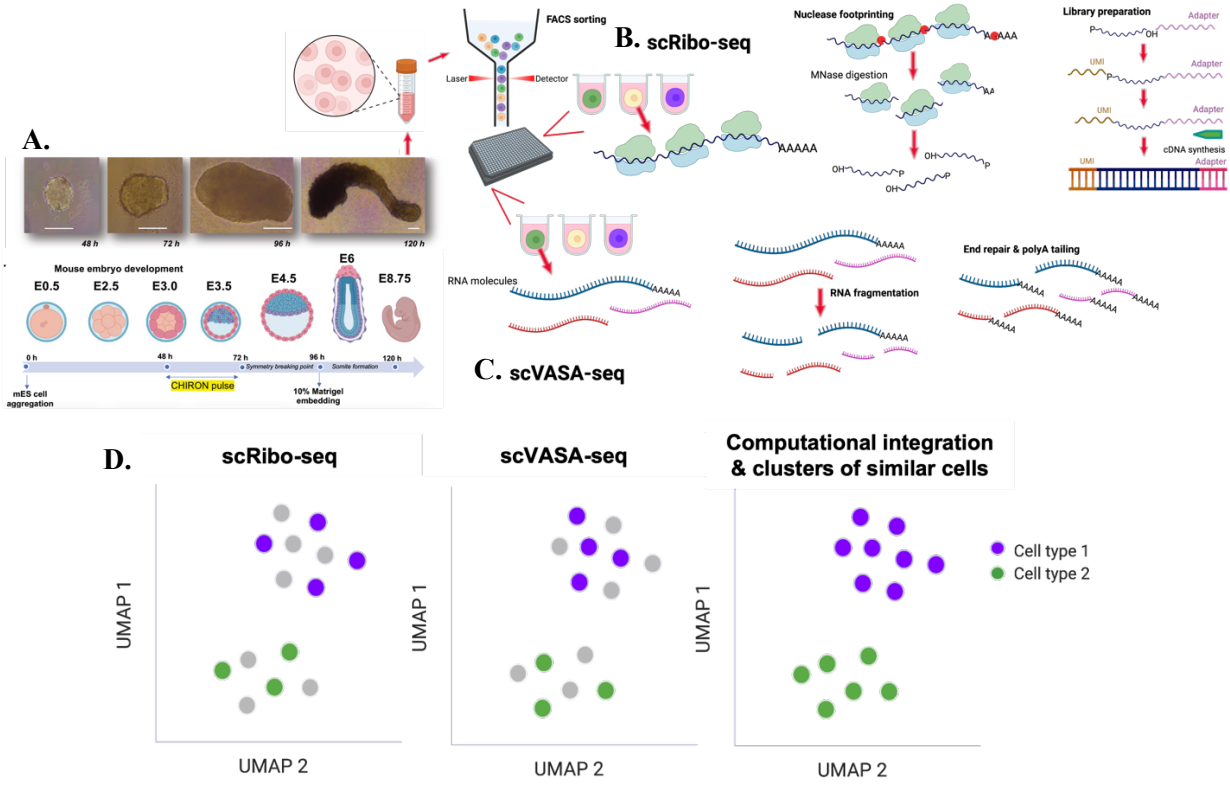


Figure 1. Methods Outline| A. Mouse gastruloids: A brief outline for making 120 h gastruloids that resemble E8.5 mouse embryos. When embedded in 10 % Matrigel they develop somite-like structures. The gastruloids are collected and dissociated to a single-cell suspension that is later used for FACS sorting., **B. scRibo-seq:** Used to quantify translation. mRNA molecules that are actively being translated by ribosomes are digested with MNase to create ribosome-protected footprints, which are used to construct the final DNA library that will be sequenced., **C. scVASA-seq:** Used to quantify the total RNA molecules present in a cell. All RNA molecules are fragmented, and a poly-A tail is added to them. These molecules are used to construct the final DNA library that will be sequenced., **D. scRibo-seq and scVASA-seq translation efficiency analysis:** Each dot represents a single cell. The single-cell gene expression and translation matrices obtained using each method are computationally integrated. This integration is used to group cells into clusters that contain closely related cell types. These cell groupings are used to quantify the ΔTE of genes across clusters.

Library QC metrics

The initial scRibo-seq and scVASA-seq experiments were performed on mouse stem cells. The purpose of that was to explore the extent of translational regulation in pluripotent cells. Besides that, these experiments allowed us to ensure that the techniques can be executed properly in a simpler system before moving on to a more complex one, the gastruloid system.

In total for the final analysis one batch of mouse stem cells was generated which is referred to as mES1 and three batches of mouse gastruloids (mGAS1, mGAS2 and mGAS3). For the mouse stem cell experiments the quality filtering resulted in the selection of 649 cells from the scVASA-seq experiment and 320 cells from the scRibo-seq experiments. For the gastruloids, quality filtering resulted in the selection of 1250 cells from the scRibo-seq experiments and 2637 cells from the scVASA-seq experiments in total for the mouse gastruloid experiments.

In order to verify that the methods worked we compared metrics to those that have been previously reported in published papers. Firstly, we compared the log10 number of unique coding-sequence mapped reads per cell (Fig. 2B) and the log10 number of protein-coding genes detected per cell (Fig. 2C) for both Ribo-seq and VASA-seq. We see that the read distributions are very similar for the scRibo-seq with the ones presented in the original paper [17] and for the scVASA-seq as well[18]. For our experiments, the average number of unique protein coding reads 9.960 ± 275 for RPF and 19.799 ± 323 for VASA and average number of protein coding genes 2.763 ± 46 for RPF and 5.141 ± 30 for VASA.

Moreover, for each experimental batch a heatmap was made that shows the proportion of total protein-coding reads that align at that specific position using the scRibo-seq data (Fig. 2A). The left region is -40 nt to +60 nt aligned around the start codon, the middle region is 100 to 200 nt aligned from the start codon, and the right region is -60 to +40 aligned at the stop codon. The plot quantifies the number of reads whose 5' ends align at each location in the metagene regions. After comparing the heatmap of our data to the one published in the original paper [17] we see that the expected 3-base periodicity of the ribosome can be detected in all batches. Moreover, we can also see the high presence of the ribosome at the start and end codon.

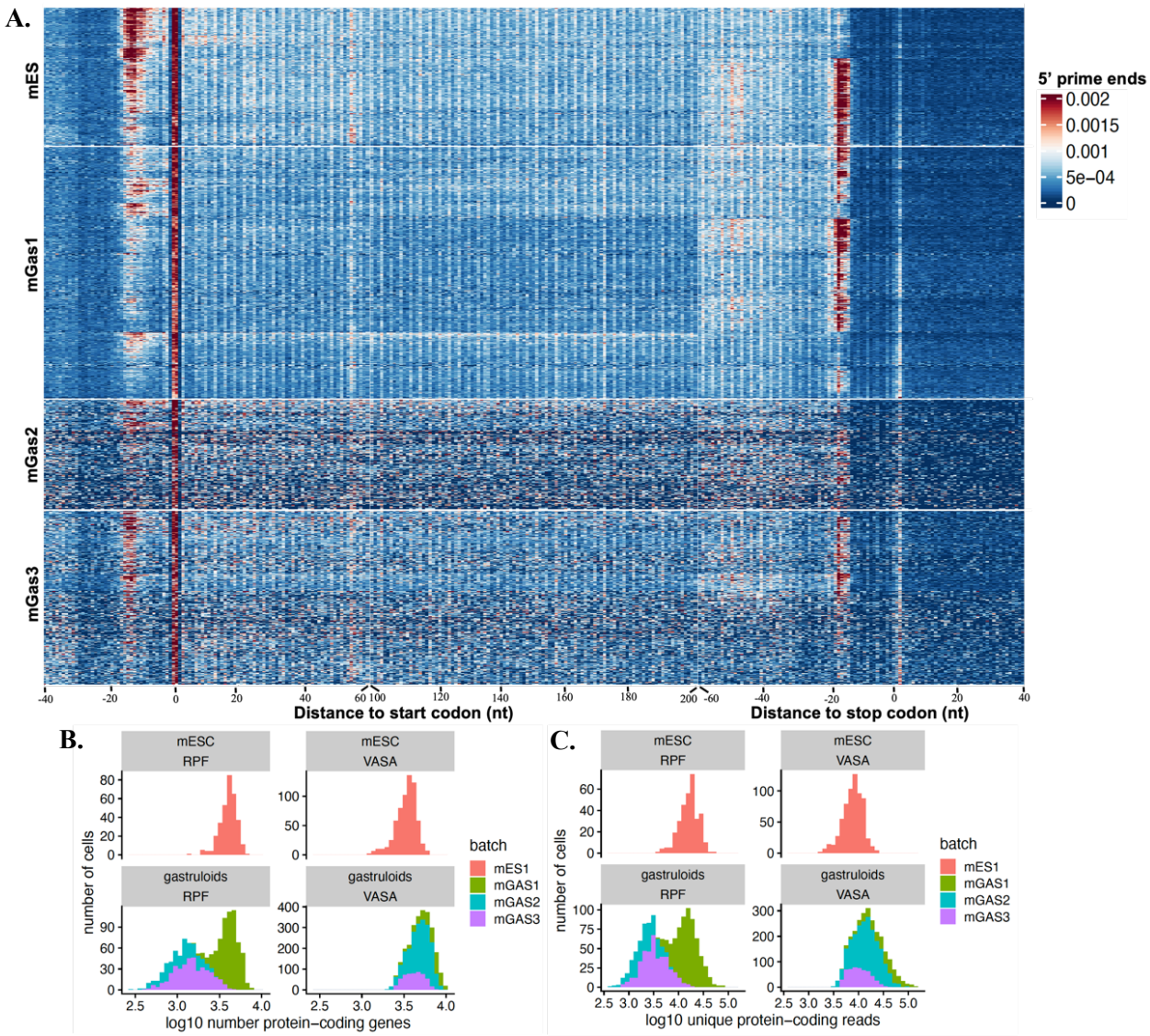


Figure 2. Library metrics for scRibo-seq and VASA-seq| A. 5' end read heatmap: Heatmap of the proportion of total protein-coding reads that align to -40 nt to +60 nt around the start codon, 100 to 200 nt from the start codon, and -60 to +40 nt to the stop codon for the mouse ES and gastruloid Ribo-seq experiments. The different plate colors correspond to the different plates that were processed in each batch., **B. Number of protein coding genes per cell:** Distribution of the $\log_{10}$ number of protein-coding genes detected per cell for the mouse stem cell and gastruloid experiments, for all batches and methods., C. **Number of unique protein coding reads:** Distribution of the $\log_{10}$ number of unique coding-sequence mapped reads per cell for the mouse stem cell and gastruloid experiments, for all batches and methods.

Translation regulation in unperturbed mESCs

Having verified the quality of the scRibo- and VASA-seq libraries, we next aimed to determine the extent of translational regulation in the unperturbed mESCs. After the integration of the mouse stem cell (mESC) data a UMAP was created with 5 different clusters (Fig. 3A), each representing a different cell type. By identifying the marker genes of each cluster, a heatmap was made that consists of genes that are significantly differentially expressed between cells in each cluster (Fig. 3B). With this information it was possible to annotate the clusters to cell types. The mES cells cluster mainly based on cell cycle stage differences. For example, we can see that in cluster 1 mostly histone genes like *H4c4* and *H3c6* are expressed, which is characteristic of cells in S phase. Additionally, we identified a cluster of neural cells, cluster 5, as the gene *Prtg* which is associated with neurogenesis and a gene mainly expressed in neural cells like *Nes* were expressed in that cluster.

By using the methodology explained (Methods, Computational Analysis) the ΔTE of multiple genes was calculated. Three genes were identified that indicate that their expression is regulated on a translational level (Fig. 3C). This conclusion can be drawn through two UMAPs and a box plot for each gene (Fig. 3C). The two UMAPs were colored by the normalized gene expression for each technique. Two of the three genes that were identified are related to pluripotency (*Dppa5a, Nanog*) and the third is a histone gene (*H2ax*). We see that the translation of the pluripotency associated gene *Dppa5a* is repressed in cluster 5. The same observation can be made for *Nanog*, a stem cell transcription factor, which appears repressed both in transcription and further in translation in cluster 5. Biologically the down regulation of the expression of those genes can be explained, as it is observed in a cluster of differentiating cells, that are gradually losing their pluripotency status. On the other hand, we see the translation of a gene that is highly involved in the processes taking place during the S phase of the cell cycle being up regulated. The above changes in gene behaviour reflect the type of biological changes that the cells go through.

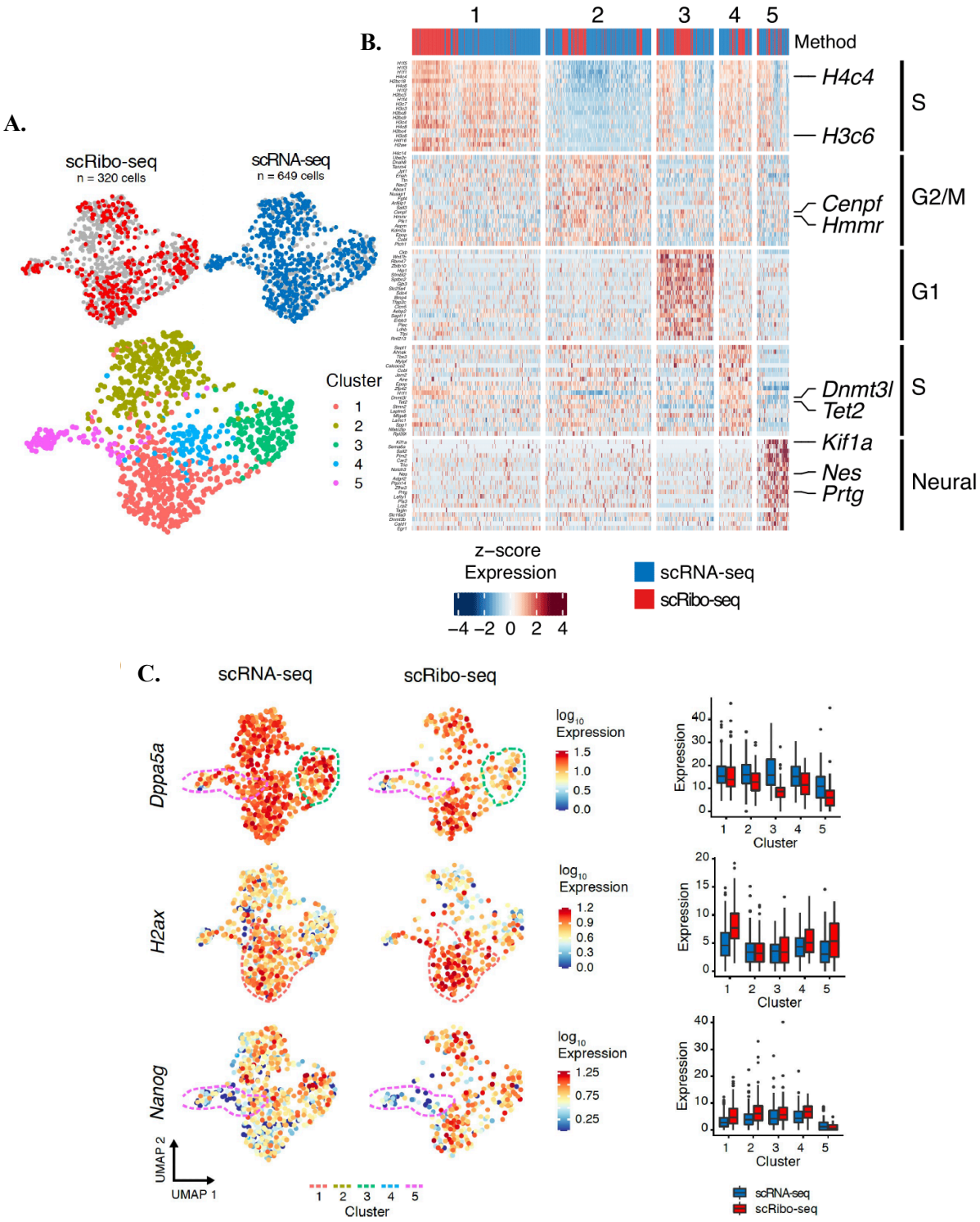


Figure 3. mES data analysis| A. UMAP generated after data integration: Top UMAPs: The data integrate well since there is a good representation of cells across the UMAP space from both techniques. Bottom UMAP: Five different clusters are identified., **B. Marker gene Heatmap:** The marker genes for each cluster are identified. Clusters 1-4 represent cell in different stages of the cell cycle. Cluster 5 represents neural cells. The gene names can be seen in Table 3 in the Supplementary Information section, **C. Genes with possible translational control:** Translation of *Dppa5a* developmental pluripotency associated gene is repressed in clusters 5 and 3. Translation of *H2ax* histone is increased in cluster 1, which represents a population of cells in the S phase of the cell cycle. Translation of *Nanog* is repressed in cluster 5, which represents a population of differentiating neural cells.

Translational regulation in mouse gastruloids

Having seen signs of differentiation-related regulation of translation in mESCs, we next aimed to extend these observations in a more complicated differentiation system, the mouse gastruloids. After integrating data from each method from each of the three experimental batches, a UMAP was created with 8 different clusters (Fig. 5A). The identities of the clusters were annotated by checking the expression of previously described marker genes based on scRNA-seq data of 120 hour gastruloids[8]. Simultaneous expression of specific combinations of marker genes are linked to the presence of specific cell types (Table 2).

After identifying the marker genes of each cluster, a heatmap was made that consists of genes that are significantly differentially expressed between cells in each cluster (Fig. 4B), to further investigate the cells present in the different clusters. The results of these annotations can be seen in Figure 5A. An example of how the annotations were done follows for clusters 0,1,2 which represent somite cells on different stages of their cell cycle (Figure 4A & 4B). In cluster 0 we see the expression of different histone genes like *H1f1*, in cluster 2 we see the expression of the gene *Cenpf*, which is mostly expressed in genes during the transition between the G2 to the M phase of the cell cycle. At the same time the marker genes for the presence of differentiated somites (*Meox2, Pax3*) are expressed almost exclusively in cells of these three clusters (Supplementary Figure 2).

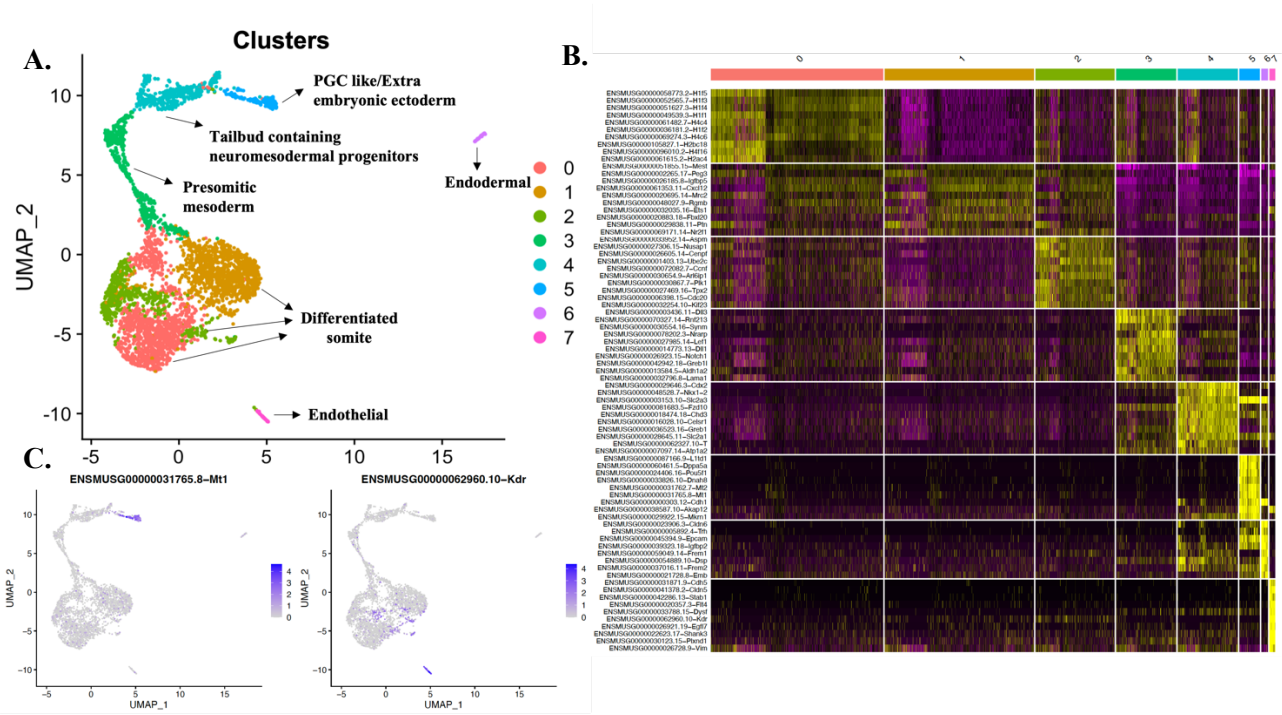


Figure 4. mGAS data analysis| A. mGAS UMAP annotation: Based on the expression of various marker genes, cell clusters were annotated to different cell types. The markers genes used were for cluster 0,1,2: *Meox2, Pax3*, cluster 3: *Tbx6, Hes7*, cluster 4: *T (Brachyury), Fgf8*, cluster 5: *Pouu5f1, Mt1*, cluster 7: *Etv2, Kdr*., **B. mGAS Heatmap:** Visualization of the top 10 genes expressed in each cluster. The gene names can be seen in Table 4 in the Supplementary Information section, **C. Feature Plots:** Examples of plots used to visualize the expression of marker genes on the different clusters of the UMAP.

The same statistical model that was used for the analysis of the mES cells was used to calculate the ΔTE of the genes in the gastruloid data. Some examples of genes that were found to undergo translational control of gene expression in gastruloids are the genes *Atp1a1*, *Eif4g1* and *Prtg* (Fig. 5A). The changes between transcription and translation are visualized with scatter plots. In each scatter plot, we see the mean normalized expression for each cluster for Ribo seq on the y axis and for VASA seq on the x axis. For the case of *Atp1a1*, we

see that while some clusters of cells (clusters 0,2,4,7) have the same levels of RNA, there are different levels of translation (cluster 7). This is also the case for *Prtg*, where clusters 5 and 7 are translated at different amounts for the same levels of RNA. In the case of *Eif4g1*, same levels of translation (clusters 0,1 and 2,6) correspond to different levels of RNA for two groups of clusters. These differences and the fact that transcription and translation change in a different manner compared to how they change in other clusters, show a very strong indication that protein production is regulated at a post transcriptional level, since changes in RNA production do not cause a proportional change to protein production for the same gene.

The genes that showed a significant difference in their ΔTE change were plotted in a heatmap. To identify underlying biological processes that might be similar between the genes that belonged in each group, the genes showing translational regulation were clustered. Based on how similar the changes where, they were grouped in 3 groups of genes. The genes seemed to change in patterns. For most of them and most of the clusters' translation seemed to be up regulated. Only in the case of cluster 7, which corresponds to endothelial cells, translation was highly down regulated in most cases. Only exception was the genes that belonged to group 3, which we know are involved in the establishment of the cardiovascular system. In cluster 7, cells labelled as haemato-endothelial progenitors were present, therefore up regulation of translation can be explained biologically.

The genes from these groups were used as input for a GO term analysis for biological processes (Fig. 5B). The results of the GO term enrichment analysis showed that genes that belong to Group 1 are mostly involved in epigenetic changes, chromatin/nucleosome assembly. The genes from Group 2 are involved in processes regarding cell growth and development, but also neural cell fate establishment. Moreover, the genes from Group 3 mainly participate in processes relevant to organ and muscle development and the establishment of the cardiovascular system.

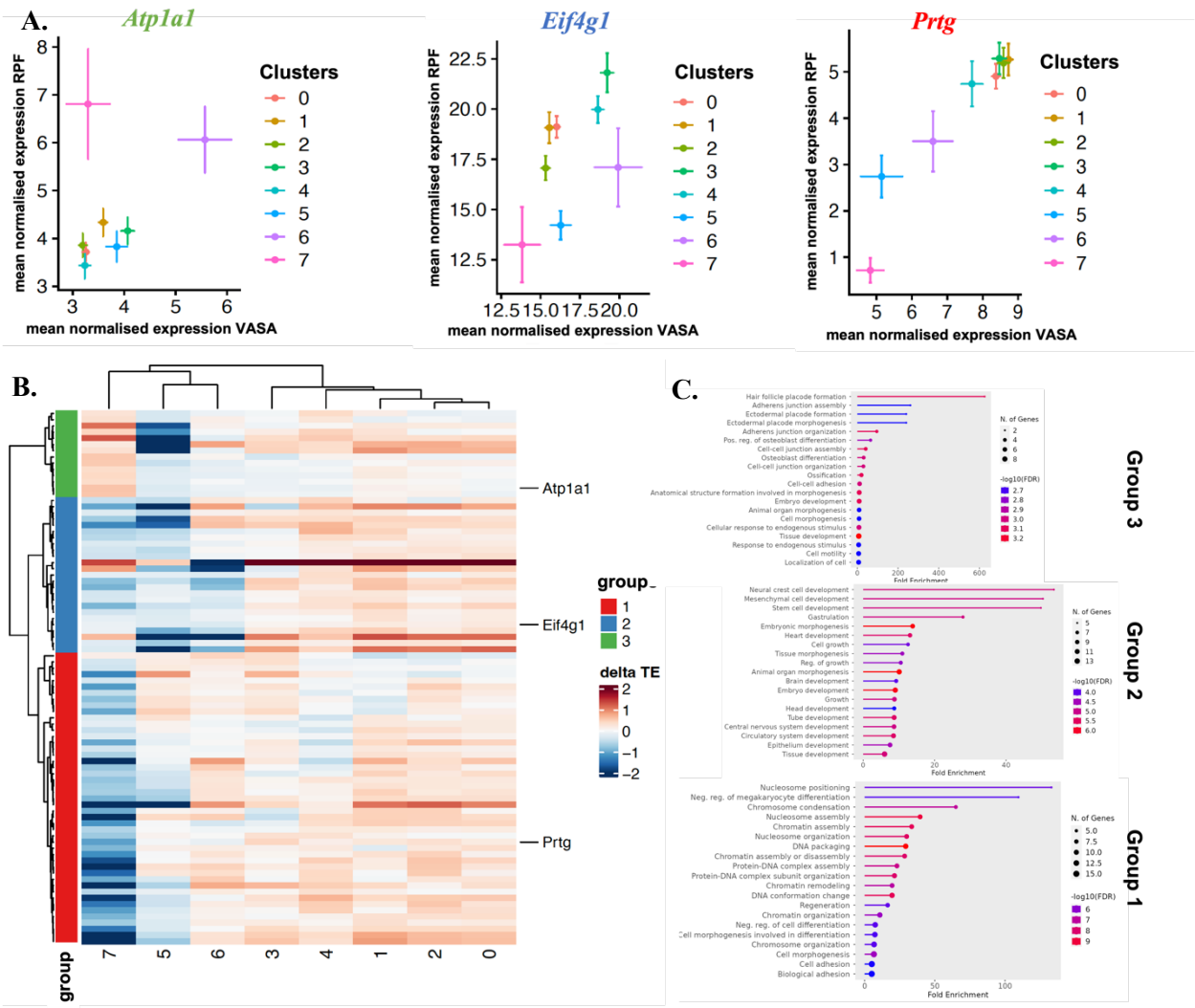


Figure 5. Translation efficiency analysis in mouse gastruloids | A. Examples of genes that show translational control of gene expression. The scatter plots show the mean normalized expression for VASA-seq and Ribo-seq for each gene.**, B. Heatmap of significantly different genes based on their ΔTE change:** Three groups of genes are identified., **C. GO term enrichment analysis:** The enrichment analysis was performed to investigate the biological processes in which the groups of genes are involved. From top to bottom the clusters presented are 3, 1 and 2. Shown in the same order as they are shown on the heatmap.

Discussion

For this project, translational regulation was investigated as a mechanism for the control of gene expression. All experiments were performed in systems derived from mice and the main focus was to study cases where this type of regulation might take place during embryo development. For this purpose, mouse gastruloids were used to imitate the signaling and cell to cell interactions that take place during the gastrulation stage of mouse embryo development. All the experiments were performed at a single-cell level and the goal of the techniques used was to compare changes in RNA and translation levels in similar cells.

Single cell Ribo-seq and VASA-seq were both performed for the first time on mouse stem cells and gastruloids. During the analysis, the data of the two methods integrated well, although batch effects could be detected (data not included). We saw that it is possible to remove them with the Canonical Correlation Analysis performed by Seurat and adding a term to the linear model.

After that, the expression of known marker genes was investigated and we confirmed that the gastruloids developed during this project express important genes that are shown to be expressed in previously published work with gastruloids[8]. This allowed us to validate the presence of various cell types and germ cell layers in our data and measure and compare transcription and translation in cells that belong to the same cell type.

The data analysis showed that it is possible to detect cases of translational control in gastruloids and stem cells with the experimental pipeline that was described in this report. We investigated the presence of translational control, identified as genes whose changes in the levels of translation did not align with the changes in their RNA levels.

We verified that the linear model can identify genes that are translationally regulated. Some examples of those genes were discussed, and in the gastruloids, the genes that were found to show a significant change in translational efficiency clustered into three groups. More specifically, from the analysis of the translational efficiency changes in the different cell types that were annotated, it is observed that there is a very high translational repression present in endothelial cells (cluster 7) and the primordial germ cell like cells (cluster 5) in 120 h gastruloids. For the other cell clusters, we can see an up regulation in translational efficiency for the majority of genes that show significant changes in their ΔTE. This can indicate that translational control plays a major role in the regulation of gene expression. Many of the genes that were translationally regulated showed an increase in TE in the early stages of their cell fate commitments (e.g., comparing clusters 0-4, 6 to 5, 7). These changes were independent of the cell-cycle stage.

The genes that showed evidence of translational regulation between gastruloid cell types clustered into three groups. The genes from each of these groups were used as input for a GO term enrichment analysis for biological processes to further investigate the type of biological processes they are involved in. We saw that the processes which these groups of genes represent are important processes that drive development, like cell growth and morphogenesis, DNA accessibility, and organ development. The main observation from this analysis was that we can detect distinct behaviors in translational regulation between the different cell types and also the translationally regulated gene modules identified agree with the biological changes that are occurring.

A challenge that we faced during this project was that good wet lab results from the library size and concentration measurements could not always guarantee that a lot of cells would pass the data analysis QC control step. Moreover, another challenge was that the statistical model returned a lot of false positive translationally regulated genes. Although various filters were applied to minimize this (p-value and fold change thresholds were used), false positive genes kept on being presented as results. Manual inspection of the results revealed that these hits were mostly due to the presence of outlier cells. Additionally, genes that were expressed only in a small number of cells and showed very extreme changes in their ΔTE. A solution to this issue would be to increase the quality of the analysed data. Having high numbers of cells that pass the QC steps would help minimize the outlier effect in the ΔTE step.

In general, we conclude that large datasets can be analyzed with the data analysis pipeline that is used and batch effects can be removed so that safe conclusion can be drawn. This experimental set-up is providing the chance to perform a very high-resolution analysis of translational gene expression regulation and thus it is providing a very useful tool to the

scientific community to further expand the knowledge of a field that hasn't been explored much.

Further experiments that can be done regarding this system would be the repetition of these experiments on a higher scale to identify a larger number of genes that show translational control. After that is done, it would be interesting to perform a motif analysis on those genes to identify possible similarities in their RNA sequence that may explain their similar behavior and maybe uncover a mechanism that allows interaction with the translational machinery or RNA binding proteins. On the already existing data obtained from the experiments done for this project, different types of computational analyses can be done. For example, studying how different types of RNAs besides mRNA are involved in translation, as VASA-seq also allows the detection of these molecules. Finally, different types of experiments can be performed to study how different treatments on the gastruloids can affect translation. An example can be the development of different mutant gastruloids with knockouts or knock-ins of various genes and study how that can affect translation and development.

We know that development is a very complex and dynamic process that requires very strict control of gene expression. Many diseases that are the result of poor gene expression regulation during embryonic development are solid proof of that. The presence therefore of an additional mechanism that would regulate gene expression on the post-transcriptional level, would give the opportunity to the developing organism to use an extra layer of control to fine tune these very complicated processes, and ensure that development will be completed successfully. Therefore, the presence of translational control could play a very important role during development and therefore it should be further investigated.

Methods

mES cell culture

Mouse embryonic stem cell line LfngT2AVenus15 [8] was used, with the selection cassette removed. The mES cells were thawed and cultured in 6 well plate wells coated with MEFs. They were kept in culture with mES media (GMEM, sodium pyruvate, non-essential amino acids, GlutaMAX, penicillin-streptomycin, β-mercaptoethanol, 10 % FBS) supplemented with ESLIF (1000 x dilution from original stock [10 million U/mL]), to maintain pluripotency in a humidified incubator (5 % CO_2, 37 ºC). After a MEF depletion step the mouse ES cells were collected for FACs sorting.

Gastruloid culture

The protocol that was followed for the generation of the gastruloids has been described in [8]. The cells were stored in liquid nitrogen and each aliquot was thawed, two days prior to the aggregation step of the gastruloid protocol, into a well of a 6 well plate, pre-coated with 0.1 % gelatin. The cells were maintained in standard conditions in mES media (GMEM, sodium pyruvate, non-essential amino acids, GlutaMAX, penicillin-streptomycin, β-mercaptoethanol, 10 % FBS) supplemented with ESLIF (1000 x dilution from original stock [10 million U/mL]), to maintain pluripotency in a humidified incubator (5 % CO_2, 37 ºC). The cells were used without passaging them to make gastruloids within two days after the day they were thawed. The stem cells that are two days in culture were detached and dissociated with prewarmed TrypLE (Thermo Fisher), which was deactivated with the addition of mES media. The cells were collected by centrifuging them and the pellet was resuspended in N2B27 media (NDiff27 media, Takara), to make a single-cell suspension. Each gastruloid was grown in a well of U-bottomed low-adhesion 96 well plate. In each well ~ 300 cells were transferred. After 48 hours of culture, 150 μL fresh N2B27 media was added in each well that

contained Chi99021 (Chi) in a final concentration of 3 µM in two steps of 75 µL with force. After, 72 hours of culture the media was replaced with 150 µL of N2B27 media. At 96 hours, the gastruloids were embedded in 10 % Matrigel (growth factor reduced, Corning) containing N2B27 media. At 120 hours the gastruloids were collected and dissociated.

Gastruloid dissociation

Only the gastruloids that had successfully elongated and developed somites in one body axis were selected. The Matrigel was removed from each well by adding Cell Recovery Solution (Corning) and leaving the plate on ice. Washes with PBS0 followed and dissociation was done by adding prewarmed trypsin-EDTA (0.5 % with phenol red, Thermo Fisher). Trypsin was quenched by adding ice cold sorting buffer (PBS0 + 10 % FBS) and the single-cell suspension was kept on ice.

FACS sorting

Polystyrene FACS tubes are used (FALCON) with the lid of polypropylene FACS tubes (FALCON). The tubes were precoated with 100 % FBS. The cell suspension was added to each tube through strainer caps. Sorting buffer was added until the final volume was 1 mL. DAPI was added at 1000x dilution from the original stock [10 µg/µL]. Tubes were kept on ice.

DAPI negative single cells were selected with a BD FACSJazz Cell Sorter (BD Biosciences). An example gating strategy is shown in Supplementary figure 3. A single cell was sorted to each well of a 384 well hard-shell plate (Bio-Rad) and the plates were snap-frozen and stored at -80 ºC until processed.

Single-cell Ribo-seq

Library construction was done by performing **1) cell lysis**, **2) ribosome footprint generation**, **3) RNA library preparation**, **4) plate pooling and purification**. The reagents were dispensed in the wells of the 384-well plates by either Nanodrop II (Innovadyne Technologies) or Mosquito (TTP Labtech). After every dispensing step, the plates were spun at 2000 ×g for 2 min. The protocol followed is the same as the one described in the original paper [17].

1) Cell lysis: Before sorting, plates were filled with 5 µL of light mineral oil (Sigma-Aldrich) and 50 nL of lysis brew.

Lysis brew	per single cell [nL]
1M Tris pH 7.5	1,1
1M MgCl2	0,8
100 mM CaCl2	2,8
5 M NaCl	1,7
10% Triton X-100	5,5
RNase IN plus (40 U/µL)	2,8
Cyclohexamide [10 mg/mL]	0,6
H2O	39,9

2) Ribosome footprint generation: In this step MNase digestion took place. MNase digestion brew was added in each well and plates were incubated at 37 ºC for 30 min.

MNase Digestion brew	per single cell [nL]
H2O	47,4
1/10 MNase (10.000 U/µL)	2,6

To stop the reaction, a proteinase K containing brew was added in each well and plates were incubated at 37 ºC for 30 min, 55 ºC for 10 min and then held at 4 ºC.

Proteinase K brew	**per single cell [nL]**
Thermolabile Prot K (120 U/mL)	7,8
0.5M EGTA	6,2
0.5M EDTA	1,7
5M GuSCN	7
H2O	27,4

3) RNA library preparation: Firstly, the end repair reaction was done. The reagents added per well can be seen in the table that follows and plates were incubated at 37 ºC for 1 h and then held at 4 ºC.

End Repair brew	**per single cell [nL]**
10× T4 RNL2 buffer	20,5
1M MgCl2	0,8
UTP [100 mM]	2
T4 PNK (10 U/ μL)	4,4
RNase IN Plus (40 U/μL)	1
H2O	21,2

Next, the addition of the 3’ prime ligation brew followed, and plates were incubated at 4 ºC for 18 h.

3' Ligation brew	**per single cell [nL]**
PEG 8000 (50%)	186,9
3' Adapter (OMV 630 20 μM)	13
10× T4 RNL2 buffer	26,4
10% Tween-20	2,6
RNase IN Plus (40 U/μL)	6,6
T4 Rnl2tr KQ (200 U/ μL)	28,1

The reverse transcription primer brew was added for the cDNA synthesis and plates were incubated at 65°C for 1 min, 37°C for 2 min, 25°C for 2 min, and held at 4°C.

RT primer brew	**per single cell [nL]**
OMV 572 (100 μM)	2,6
100 mM ATP	6,7
10% Tween	0,5
H2O	40,2

After that the 5’ ligation brew was prepared and added for the addition of the 5’ adapters and the plates were incubated at 37 °C for 2 h and then held at 4 °C.

5' Adapter brew	**per single cell [nL]**
OMV632 (20 μM Denatured)	3,9
10× T4 RNL2 buffer	15,6
10% Tween-20	1,6
T4 RNA Ligase (10 U/μL)	38,9
PEG 8000 (50%)	95,7

Next, the Reverse Transcription brew was added for the cDNA synthesis, and the plates were incubated at 50 °C for 1 h, then 85 °C for 5 min and held at 4 °C.

Reverse Transcription brew	per single cell [nL]
5× RT Buffer	288,9
10 mM dNTP mix	96,3
10% Tween 20	14,4
H2O	298,5
RNase IN Plus (40 U/µL)	36,1
Maxima H Minus Enzyme (200 U/µL)	36,1

In this step each well contains a single-cell library. In each unique cell and plate sequencing primers were added along with the PCR brew. The plates were incubated at 98 °C for 30 s and 10 cycles of 98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s followed. A final incubation of 72 °C for 5 min was done and the plates were held at 4 °C and they were either kept there or frozen at −20 °C until pooling.

PCR Brew	per single cell [nL]
10% Tween-20	48,2
dH2O	463,5
2× Q5 Hot Start Mix	2407,7
Adapter Plate [20 µM]	150,5
RPI [10 µM]	301

4) Plate pooling and purification: For the pooling process each plate was put in a VBLOK200 reservoir (Click Bio) and, by centrifuging for 2 min at 2000 ×g, its contents (the libraries) were collected, and then transferred to a tube. The two phases, light oil and aqueous one, were separated by centrifugation and the aqueous phase (contains the libraries) was concentrated to ~ 500 µL with n-butanol.

The libraries were then **purified** by using AMPure XP beads (Beckman Coulter) that were diluted 6× in bead binding buffer (20% PEG-8000, 2.5 M sodium chloride) and added to each collection of libraries at a 2.1:1 ratio. The purified product was resuspended in 20 µL low TE buffer (LoTET, 3 mM Tris-HCl pH 8.0, 0.2 mM EDTA pH 8.0, 0.1% Tween-20).

Half of each sample was run in a 10 cm 7% acrylamide gel (75 mL) at 300 V for ~ 3 h and the band that corresponds to 180 bp was excised for each sample. Each band was crushed and soaked overnight at 4 °C in a tube containing 5:1 LoTET:7.5 M ammonium acetate.

The gel was filtered, and the 500 µL of each library were precipitated in 1380 EtOH µL, 50 µL 3 M Sodium Acetate and 3 µL linear acrylamide overnight at -20°C. EtOH was removed, and the library is resuspended in low TE buffer. The libraries were quantified with Qubit and the size of the excised bands was checked with a high sensitivity dsDNA bioanalyzer chip. The expected size was ~ 185 bp.

Single-cell VASA-seq

Library construction was done by performing **1) cell lysis and RNA fragmentation, 2) cDNA preparation, 3) plate pooling and purification, 4) IVT, purification and library construction.** The reagents were dispensed in the wells of the 384 well plates by either Nanodrop II (Innovadyne Technologies) or Mosquito (TTP Labtech). After every dispensing step, the plates were spun at 2000 ×g for 2 min. The protocol followed is the same as the one named VASA-plate in the original paper [18].

1) Cell lysis and RNA fragmentation: Before sorting the cells on the plates, 5 μL of light mineral oil was dispensed (Sigma-Aldrich) in each well of the 384-well plates and 50 nL of Cel-seq2/Sort-seq 0.25 μM primer.
In each well the lysis and fragmentation brew were added, and the plates were incubated at 25 °C for 1 h, 55 °C for 10 min and held at 4 °C. Then an incubation followed at 85 °C for 3 min to fragment the cells and plates were held at 4 °C.

Lysis and fragmentation brews	**per single cell [nL]**
5x First Strand buffer	34
Thermolabile Prot K (120 U/mL)	10
IGEPAL 2% (diluted form 10% in H_20)	5
Spike in 1:2500	1

2) cDNA preparation: End repair & poly(A)-tailing brew was added to each well of the plate. The plates were incubated at 37 °C for 1 h and held at 4 °C.

End repair and poly(A)-tailing brew	**per single cell [nL]**
H_2O	13
5x First Strand buffer	6
DTT (0.1M)	10
ATP (0.1mM, diluted from 10mM in 50mM Tris)	3,75
Poly (A) Polymerase (10x diluted in 50mM Tris) (5 U/μL)	7,5
T4 PNK (10 U/ μL)	5
RNaseOUT (40U/μl)	5

After that, the reverse transcription brew was added in each well and the plates were incubated at 50 °C for 1 h and held at 4 °C.

Reverse transcription brew	**per single cell [nL]**
H_20	36
10mM dNTP mix	10
Superscript III Reverse Transcriptase (200 U/μL)	4

The second strand synthesis brew was then added. The plates were incubated at 16 °C for 2 h and held at 4 °C. Enzyme inactivation followed by incubating the plates at 85 °C for 20 min and then they were held at 4 °C.

Second strand synthesis brew	**per single cell [nL]**
H_20	390
5x Second Strand buffer	125
10mM dNTP mix	12,5
E.coli DNA polymerase I (10 U/μL)	17,5
RNAseH (2U/ μL)	5

3) Plate pooling and purification: Plates were pooled as it was described in the scRibo-seq protocol and the libraries that corresponded to the plates were transferred in tubes. The libraries were then purified by using AMPure XP beads that were diluted 8× in bead binding buffer and added to each collection of libraries at a 1:1 ratio. The purified product was eluted in 8 μL 0.25% Tween-20.

4) IVT, purification and library construction: Invitrogen's IVT kit was used for in vitro transcription and the samples were incubated at 37 ºC for 16 h and held at 4 ºC. After that Exo-sap was added to all samples, and they were incubated at 37 ºC for 15 min and held at 4 ºC.

IVT Reagents	Volume added per sample [µL]
MegaScript T7 buffer	2
MegaScript T7 enzyme	2
MegaScript ATP	2
MegaScript CTP	2
MegaScript GTP	2
MegaScript UTP	2

Then, the libraries were **purified** by adding undiluted AMPure XP beads to each collection of libraries at a 2:1 ratio. The purified RNA was eluted in 10 µL 0.25% Tween-20. The concentration of the purified RNA was diluted 10x and measured with Qubit. For the **rRNA depletion** the concentration of the samples should be 100 ng/µL and the volume 6 µL. In each sample 2 µL of rRNA depletion probes (25 µM) and 2 µL of Hybridization buffer (pH 7.5, 500 mM Tris-HCl, 1M NaCl) were added. The samples were incubated at 95 ºC for 2 min and then the temperature was decreased to 45 ºC at a rate of 0.1 ºC/s.

After that, a mix of 2 µL Thermostable RNaseH and 8 µL of RNaseH buffer (pH 7.5, 125 mM Tris-HCl, 250 mM NaCl, 50 mM $MgCl_2$) was prepared for each sample and prewarmed at 45 ºC before being added in each sample. The samples were incubated at 45 ºC for 30 min and held at 4 ºC. In the last step of the rRNA depletion, 4 µL of RQ DNase I (1 U/µL), 21 µL of dH2O and 5 µL of $CaCl_2$ (10 mM) were added in each sample and they were incubated at 37 ºC for 30 min and held at 4 ºC.

The libraries were **purified** again with undiluted AMPure XP beads to each collection of libraries at a 1.6:1 ratio. Purified RNA was eluted in 6 µL 0.25 % Tween-20.

Adapter ligation followed and 1 µL of adapter OMV630_miRNA4_3App (20 µM) was added to each sample and incubated at 70 ºC for 2 min and held at 4 ºC. Then, the rest of the adapter brew was added in each sample, and they were incubated at 25 ºC for 1 h and held at 4 ºC.

Adapter Ligation brew	Volume added per sample [µL]
10x T4 RNA Ligase Reaction Buffer	1
T4 RNA Ligase 2, truncated (200 U/µL)	1
RNaseOUT (40U/µl)	1
H_2O	1

For the **second cDNA synthesis** 1 µL of dNTPs (10 mM) and 2 µL of RTP oligonucleotide (20 µM) were added in each sample and incubated at 65 ºC for 5 min and held at 4 ºC. After that the second cDNA synthesis brew was added in each sample and they were incubated at 50 ºC for 1 h, 70 ºC for 15 min and held at 4 ºC.

Second cDNA synthesis brew	Volume added per sample [µL]
5x First Strand Buffer	4
DTT (0.1M)	1
RNaseOUT (40 U/µl)	1
H_2O	1

RNA degradation followed and 1 µL of RNaseA was added to each sample and it was incubated at 37 ºC for 30 min and held at 4 ºC. Two **purification** steps followed with undiluted AMPure XP beads at a 1:1 ratio. The purified DNA was eluted in 20 µL 0.25% Tween-20 each time.

The last reaction performed was the **PCR amplification** for which PCR amplification brew was added to each sample and a PCR reaction took place. The amplification programme included an initial heat denaturation at 98°C for 30 sec, 7-8 cycles for 10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C, and a final extension step at 72°C for 10 min and the samples were held at 4°C.

PCR amplification brew	**Volume added per sample [µL]**
2x NEBNext PCR mix	22
RNA PCR Primer (RP1)	1.8
Index (one per sample)	2
H_2O	10

Lastly, two **purification** steps followed with undiluted AMPure XP beads for each sample. The first one at a 0.8:1 ratio and the purified DNA was eluted in 25 µL low TE buffer and the second and final one at a 1.25:1 ratio and the purified DNA was eluted in 10 µL low TE buffer. The libraries were quantified with Qubit and the size of the libraries was checked with a high sensitivity dsDNA bioanalyzer chip.

Sequencing

scRibo-seq

Libraries were sequenced using a NextSeq 2000 (Illumina) with 72 cycles for R1. Run type P3: 88-cycle: 72 cycles for read 1, 6 cycles for the i7 index read, and 10 cycles for the i5 index read.

scVASA-seq

The VASA samples were sequenced on a NextSeq 2000 (Illumina). Run type P3 100-cycle: 26 cycles for R1 and 106 cycles for R2, 6 cycles for i7 index read.

Computational analysis

In total, one batch of mES cells and three batches of gastruloids were processed and sequenced. Sample demultiplexing and read trimming, alignment, deduplication, and counting were done as described in the original paper [17]. From the sequencing reads that were retrieved from each method for each cell, a count matrix was created containing the number of counts per gene per single cell. Next, quality control (QC) steps were performed to filter the cells. Only the ones that were within set thresholds were used for further analysis. Different types of thresholds were applied for the scRibo-seq and the scVASA-seq data.
For Ribo-seq, two thresholds were used to select cells that did not fail during library preparation and that had most of their reads aligning to coding sequences. Specifically, the log10 of the total unique protein-coding reads per cell was plotted against the fraction of these reads that align to coding sequences (CDS). For the mESCs, cells were selected for further analysis if they had at least 4000 protein-coding reads and 70 % aligned to CDSs; for the gastruloids, cells were selected if they had at least 4000 (batch mGAS1) or 500 (batches mGAS2 and mGAS3) protein-coding reads, with at least 65 % aligning to CDSs (Supplementary Figure 1A.).
For the VASA-seq, the cells that were selected were those that had the highest number of counts for the log10 of the number of genes and cells (Supplementary Figure 1B.); these

exact thresholds were modified to adjust for different sequencing depths per plate but ranged from 2.000 to 4.000.

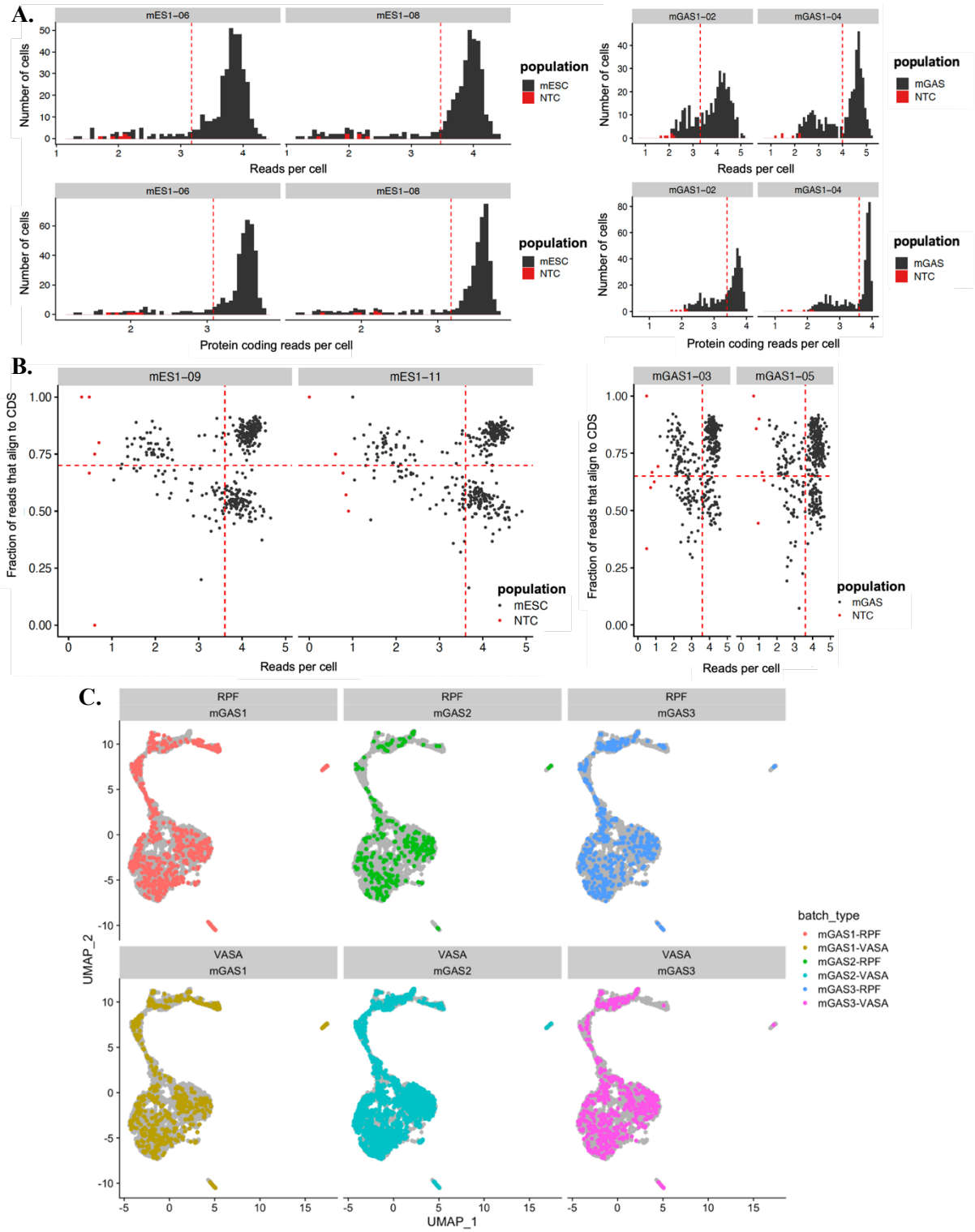
After that, the data from both methods were normalized and integrated to remove batch and method effects and further analysis of the counts was done using the R (version 2022.07.2) package Seurat (version 4.1.1). All statistics and plotting were done using ggplot2 (version 3.3.6), dplyr (version 1.0.9), tidyr (version 1.2.0), DESeq2 (version 1.34.0) packages. All analysis was performed in Rstudio (version 2022.07.2).

Specifically, the data were normalized, merged and batch effects were removed after the QC controls for both methods were complete. This allowed the computational data integration in a Seurat object. The Seurat package used the highly variable genes across all groups in the data to make clusters that contained similar cells with shared biological features. Seurat performed a Canonical Correlation Analysis (CCA) [19] and identified common varying genes, which were considered anchors and were used as a reference to perform the integration, dimensionality reduction, and clustering.
The result of that was displayed on a UMAP, where each cell is represented with a dot and the relative distance between cells across the UMAP space revealed how similar they are. At this step of the analysis, the success of the integration could be assessed by checking the representation of all batches and methods on the UMAP space. It can be considered successful if there is an equal representation of all batches and methods across the UMAP. As can be seen in Supplementary Figure 1C, integration was done successfully, with the cells measured from each batch using each method mixing in the UMAP representation. The comparison between transcription and translation would be possible across clusters, since each cluster contains similar cells coming from all the experiments performed (Supplementary Figure 1C.).

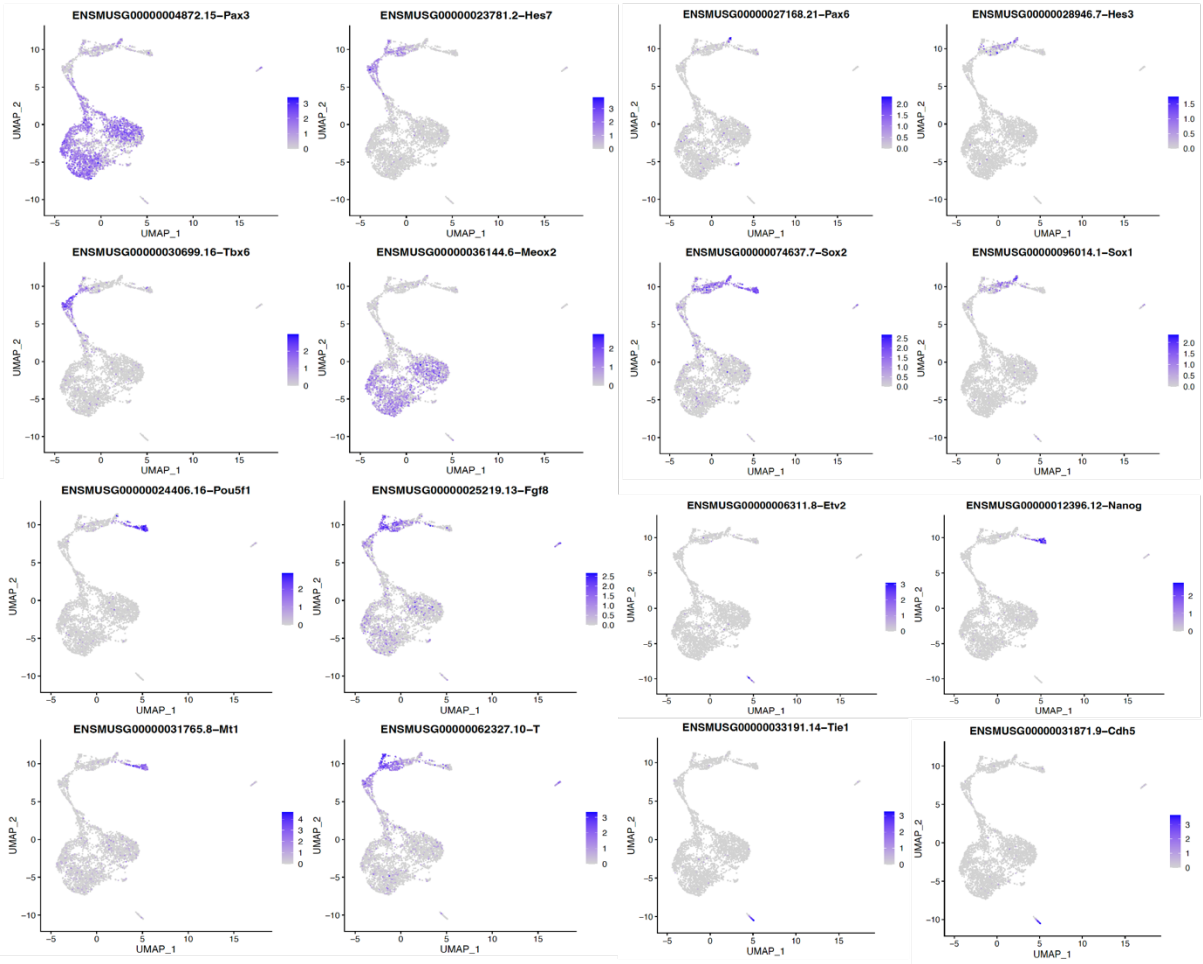
A generalized linear model was used to measure changes in translation between cell clusters, taking into account potential batch effects, technical differences between scRibo-seq and VASA-seq, and differences in RNA abundance between cell clusters. Specifically, this was parameterized as modelling the counts for each gene as a function of the batch, method, cell cluster, and an interaction term between method and cell cluster. For genes that are only transcriptionally regulated, the levels of translation would be expected to change in proportion to the amount of RNA. Changes to the translational regulation of a gene between clusters, however, would present as a deviation from this relationship. Such behaviour is captured in the interaction term between the method and cell cluster. Thus, a likelihood ratio test was used to test the full model above to the reduced model not containing the interaction term. The model and testing were implemented in DESeq2. For the mESC data, genes were selected if they had an adjusted p-value less than 0.1; for the gastruloids, genes were selected if they had an absolute fold change greater than 0.35 and less than 8 and an adjusted p-value less than 0.05.

The biological processes in which the genes that showed indication of translational control were further investigated by doing a GO term enrichment analysis for Biological Processes using ShinyGO (version 0.76.3).

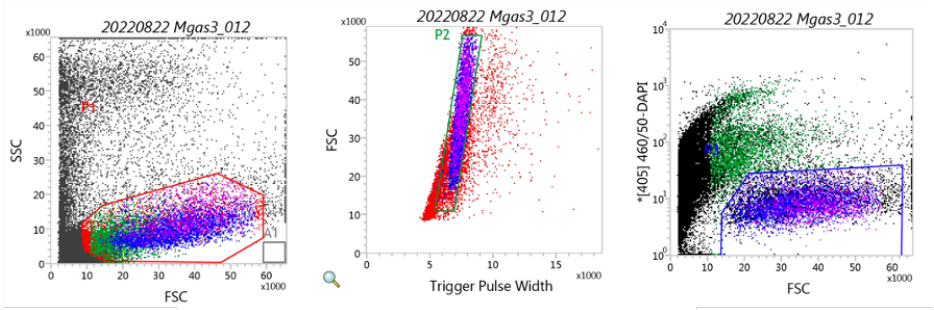
Supplementary figures



Supplementary Figure 1. Quality control of scRibo-seq and VASA-seq| A. Example QC histograms for VASA-seq: 1) mouse ES (mES) experimental data. Each plot represents the cell population of a plate., 2) mouse gastruloids (mGAS) experimental data. Each plot represents the total cell population of a plate. The red dashed lines represent the thresholds set for the cell selection., **B. Example QC plots for Ribo-seq:** 1) mouse ES (mES) experimental data. Each plot represents the cell population of a plate, and each dot represents a single cell., 2) mouse gastruloids (mGAS) experimental data. Each plot represents the cell population of a plate, and each dot represents a single cell. The red dashed lines represent the thresholds set for the cell selection., **C. Integrated scRibo-seq and scVASA-seq:** Each color represents a different batch for the gastruloid experiments, and each dot represents a cell. In each UMAP we see the cells that come from each measurement type for each experimental replicate



Supplementary Figure 2. Marker gene expression visualised in Feature plots. |In these plots the expression of marker genes for different cell types is shown for all the clusters identified. These plots were used to annotate cell types for each cluster. Table 2 shows in detail the annotation between the expression of groups of genes and the cell types they correspond to.



Populations: *20220822 Mgas3_012*

Populations	Even...	% Total	% Parent
All Events	27,940	100.00%	####
A1	0	0.00%	0.00%
Accudrop	27,940	100.00%	100.00%
P1	10,398	37.22%	37.22%
P2	7,757	27.76%	74.60%
P3	4,741	16.97%	61.12%
P4	45	0.16%	0.95%
P5	3,452	12.36%	72.81%
P6	1,088	3.89%	22.95%

Supplementary Figure 3. Gating strategies for FACS sorting:| An example of how gating was done during the FACS sorting to select viable single cells for the experiments.

Supplementary Information

Three batches of mouse gastruloids were generated named mGAS1, mGAS2 and mGAS3 and one batch of mouse stem cells named mES. Different numbers of plates were processed for each batch and each experimental technique (Table 1).

Table 1.| Number of 384 well plates that were processed per batch and per experiment

	mES	mGAS1	mGAS2	mGAS3
scRibo-seq	2	3	8	10
scVASA-seq	2	2	8	3

Table 2.| Marker gene expressed in the data and their corresponding cell types

Marker gene expression	Corresponding cell type
Etv2	haemato-endothelial progenitors
Kdr	haemato-endothelial progenitors & endothelium
Cdh5, Tie1	endothelium
Ephx2, Mt1, Utf1, Pou5f1	primordial-germ-cell-like or extra-embryonic ectoderm
Col4a1, Epcam, Sox17	endoderm
Hand2, Gata6	heart
Meox2, Pax3	differentiated somite
Hes7, Tbx6	presomitic mesoderm
Wnt3a, Fgf17, Fgf8, Cyp26a1, Nkx1-2, T	tail bud containing neuromesodermal progenitors
Pax6, Sox1, Hes3, Sox2	differentiated neural cells

Table 3.| Genes shown in the clusters of the heatmap Figure 3B.

Clusters	Genes
Cluster 1	*H1f5, H1f3, H1f1, H4c4, H2bc18, H4c6, H1f2, H2bc3, H1f4, H3c7, H3c3, H2bc8, H2bc9, H3c4, H4c8, H2bc4, H3c6, H4f16, H2aw*
Cluster 2	*H4c14, Ube2c, Dnah8, Tenm4, Jpt1, Enah, Ttn, Nav2, Abca1, Nusap1, Fgf4, Arl6ip1, Sall3, Cenpf, Hmmr, Plk1, Aspm, Kdm2a, Epop, Cobl, Ptch1*
Cluster 3	*Ckb, Wnt7b, Rbm47, Zbtb10, Hip1, Sfmbt2, Sptbn2, Gjb3, Slc25a4, Sdc4, Bmp4, Tfap2c, Clcn5, Aebp2, Sept11, Erbb3, Plec, Ldhb, Tfpi, Rnf213*
Cluster 4	*Sept1, Ahnak, Tbx3, Mylpf, Calcoco2, Cobl, Jam2, Aire, Epop, Zfp42, H1f1, Dnmt3l, Tet2, Stmn2, Laptm5, Mfge8, Lamc1, Spp1, Nfatc2ip, Rpl39l*
Cluster 5	*Kif1a, Sema6a, Sall2, Pim2, Car2, Trio, Notch3, Nes, Adgrl2, Ptpn14, Zfhx3, Prtg, Lefty1, Pls3, Lrp2, Tagln, Slc16a3, Dnmt3b, Cald1, Egr1*

Table 4. | Genes shown in the clusters of the heatmap Figure 5B.

Clusters	Genes
Cluster 0	*H1f5, H2ac4, H4f16, H2bc18, H4c6, H1f2, H4c4, H1f1, H1f4, H1f3, H2ac4*
Cluster 1	*Mest, Nr2f, Ptn, Fbxl20, Ets1, Rgmb, Mrc2, Cxcl12, Igfbp5, Peg3*
Cluster 2	*Aspm, Kif23, Cdc20, Tpx2, Plk1, Arl6ip1, Ccnf, Ube2c, Cenpf, Nusap1*
Cluster 3	*Dll3, Lama1, Aldh1a2, Greb1l, Notch1, Dll1, Lef1, Nrarp, Synm, Rnf213*

Cluster 4	*Cdx2, Atp1a2, T, Slc2a1, Greb1, Celsr1, Chd3, Fzd10, Slc2a3, Nkx1−2*
Cluster 5	*L1td1, Mkrn1, Akap12, Cdh1, Mt1, Mt2, Dnah8, Pou5f1, Dppa5a*
Cluster 6	*Cldn6, Emb, Frem2, Dsp, Frem1, Igfbp2, Epcam, Trh*
Cluster 7	*Cdh5, Vim, Plxnd1, Shank3, Egfl7, Kdr, Dysf, Flt4, Stab1, Cldn5*

Acknowledgements

Firstly, I would like to thank Alexander for giving me the opportunity to do my major research internship in his lab. I had the opportunity to learn new experimental techniques and work in a very professional space with scientists that excel in their field. The skills that I acquired during the time I spent in the AvO lab and the things I've learned have helped me evolve as a scientist, become better and well prepared for my future steps in the field. My experience here made me realise that deciding to study abroad was indeed the right choice.

Secondly, I would like to thank my daily supervisor, Mike. He taught me everything I learnt in this lab and was always patient and eager to help me solve any problems I encountered. I had to learn a lot of things from scratch for this project and he guided me through everything and always answered my questions, which were a lot. His feedback and advice helped me improve and become a better scientist and motivated me to always strive for the best possible outcome.

I would also like to thank Jerry, for introducing me to all cell culture techniques and giving me feedback multiple times and also Nune for teaching me the gastruloid protocol. Also, I would like to thank Max for his valuable advice and feedback for multiple topics, science related and not. Also, I would like to thank Marloes and Vincent for sharing with me a lot of tips and advice for various experiments and topics.

All in all, I would like to thank everyone in the lab for creating a pleasant working environment and making me feel welcome. I also want to thank Josi and Annemiek for always making sure that everything works properly and Renier for sorting all the cells for the experiments. A big shoutout to the people in the "Dutch" office for always making it fun to work in and for all the interesting conversations we had. And of course, I would like to thank anyone that ever brought any kind of dessert or snack to the lab, the extra calories and glucose always gave a pleasant "kick" and made working easier.

I would also like to thank my boyfriend Deniz for all his emotional support and help. He is always there for me when I need him and gives me strength to chase my dreams and never quit. He was always my audience when practising my lab meeting presentations and although he is not a biologist, he grew fond of gastruloids and thinks they're pretty cool.

Last but not least, I also want to thank my family for supporting me all these years. Without them I wouldn't be able to be here in the first place and their love has always been valuable to me.

References

1. Nishioka N, Yamamoto S, Kiyonari H, et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev*. 2008;125(3-4):270-283. doi:10.1016/J.MOD.2007.11.002
2. Yagi R, Kohn MJ, Karavanova I, et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development*. 2007;134(21):3827-3836. doi:10.1242/DEV.010223
3. Keynes RJ, Stern CD. Mechanisms of vertebrate segmentation. *Development*. 1988;103(3):413-429. doi:10.1242/dev.103.3.413
4. Ramaswami LS. The development of the avian vertebral column. 2000:271-280.
5. White PH, Farkas DR, McFadden EE, Chapman DL. Defective somite patterning in mouse embryos with reduced levels of Tbx6. *Development*. 2003;130(8):1681-1690. doi:10.1242/dev.00367
6. Girgin MU, Broguiere N, Mattolini L, Lutolf MP. Gastruloids generated without exogenous Wnt activation develop anterior neural tissues. *Stem Cell Reports*. 2021;16(5):1143-1155. doi:10.1016/j.stemcr.2021.03.017
7. van den Brink SC, van Oudenaarden A. 3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation. *Trends Cell Biol*. 2021;31(9):747-759. doi:10.1016/J.TCB.2021.06.007
8. van den Brink SC, Alemany A, van Batenburg V, et al. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*. 2020;582(December 2018). doi:10.1038/s41586-020-2024-3
9. Beccari L, Moris N, Girgin M, et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*. 2018;562(7726):272-276. doi:10.1038/s41586-018-0578-0
10. Arias AM, Marikawa Y, Moris N. Gastruloids: Pluripotent stem cell models of mammalian gastrulation and embryo engineering. *Dev Biol*. 2022;488(April):35-46. doi:10.1016/j.ydbio.2022.05.002
11. Alberio R. Transcriptional and epigenetic control of cell fate decisions in early embryos. doi:10.1071/RD17403
12. Kondrashov N, Pusic A, Stumpf CR, et al. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*. 2011;145(3):383-397. doi:10.1016/J.CELL.2011.03.028
13. Xiong Z, Xu K, Lin Z, et al. Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol*. 2022;24(6):968-980. doi:10.1038/s41556-022-00928-6
14. Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (80-)*. 2009;324(5924):218-223. doi:10.1126/science.1168978
15. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol*. 2015;16(11):651-664. doi:10.1038/nrm4069
16. McGlincy NJ, Ingolia NT. Transcriptome-wide measurement of translation by ribosome profiling. *Methods*. 2017;126(2017):112-129. doi:10.1016/j.ymeth.2017.05.028
17. VanInsberghe M, van den Berg J, Andersson-Rolf A, Clevers H, van Oudenaarden A. Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. *Nature*. 2021;597(7877):561-565. doi:10.1038/s41586-021-03887-4
18. Salmen F, De Jonghe J, Kaminski TS, et al. High-throughput total RNA sequencing in single cells using VASA-seq. *Nat Biotechnol*. 2022;40(December). doi:10.1038/s41587-022-01361-8
19. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol Vol*. 2015;33. doi:10.1038/nbt.3192