Translation on Demand as a post-transcriptional regulation mechanism of ES cell differentiation

Master thesis for acquisition of the degree of Master of Science (M.Sc.) in Computational Biology of the faculty of mathematics and natural sciences at the University of Cologne

submitted by Elias Schwall

June 21, 2025

Reviewers:

Prof. Dr. Andreas Beyer

Prof. Dr. Achim Tresch

Supervised by:

Luke Brandwood

Declaration of originality

I hereby declare that this thesis is my own work and that no other sources and tools than stated were used.

Cologne, June 21, 2025 Elias Schwall

Acknowledgement

Table of contents

Abstract

Introduction

Mouse embryonic stem cells

Mouse embryonic stem cells (mESCs) are defined by their self-renewal capabilities and pluripotency, allowing them to differentiate into any cell type of the three germ layers. While this differentiation process is limited in vivo, pluripotent stem cells (PSCs) cultivated in vitro can self-renew indefinitely under specific culture conditions (Nichols and Smith 2009). One such condition is the 2i+LIF (two inhibitors with leukemia inhibitory factor) medium, which maintains the naive pluripotent state of ESCs, closely mimicking the inner cell mass of pre-implantation mouse embryos.

The continuum of pluripotency is described by a progression from the naive state to a primed state, characterized by the priming for differentiation (Nichols and Smith 2009). Naive mESCs serve as a model for the inner cell mass of the pre-implantation embryo, exhibiting minimal lineage commitment. In contrast, epiblast stem cells (EpiSCs), which are isolated from later developmental stages, correspond to the post-implantation epiblast and are primed for differentiation (Boroviak et al. 2014).

During the transition from the naive to the primed states, sequential changes occur at multiple molecular levels, encompassing alterations in signaling pathways, epigenetic markers, and gene expression (Dejosez and Zwaka 2012). Specifically, early modifications in the phosphoproteome serve as indicators of these state transitions, initiating a cascade of changes that ultimately impact the transcriptome and proteome (Needham et al. 2019). This progression underscores the dynamic nature of pluripotency and highlights the intricate molecular mechanisms governing cell fate decisions in early development.

It is worth noting that early mESCs development differs from human embryonic stem cells (hESCs). While the molecular mechanisms controlling developmental processes are evolutionarily conserved, the speed at which these processes occur can vary significantly between species. For instance, the differentiation of motor neurons from embryonic stem cells is more than twice as fast in mice compared to humans (Rayon et al. 2020).

Regardless of the species, however, early developing embryonic stem cells need to undergo rapid changes in their gene expression profiles to ensure proper development and

differentiation. Mechanisms such as ToD, can allow for swift adjustments in protein levels to aid these rapid changes. This ensures that cells can quickly respond to developmental cues and environmental changes, maintaining the necessary pace of development. Such mechanisms are crucial for adapting to the specific temporal requirements, ensuring that embryonic stem cells can effectively progress through the stages of development.

2i Condition for maintaining naive pluripotency

Since serum contains undefined factors that can induce differentiation, the 2i medium was developed to provide a defined, chemically controlled environment that preserves mESCs in the naive pluripotent state (Ying et al. 2008). The 2i medium relies on the synergy between two inhibitors, PD0325901 and CHIR99021, to sustain mESCs in an undifferentiated state. PD0325901, a MEK inhibitor, blocks the MEK/ERK signaling pathway, suppressing differentiation signals from FGF and ERK that would otherwise prompt mESCs to exit pluripotency (see Figure 1). Meanwhile, CHIR99021, a GSK3α/β inhibitor, stabilizes β-catenin and activates Wnt/β-catenin signaling, promoting the expression of self-renewal genes and reinforcing pluripotency (Ying et al. 2008). Together, these inhibitors create a controlled environment that preserves mESCs in a ground state resembling the undifferentiated pre-implantation epiblast (Ying et al. 2008). Alongside leukemia inhibitory factor (LIF), which activates the JAK-STAT3 pathway to further inhibit differentiation, the 2i medium maintains mESCs in a naive pluripotent state, enabling them to retain full differentiation potential without lineage bias (Ying et al. 2008).

The removal of the 2i medium triggers the transition from the naive pluripotent state by reactivating the MEK/ERK and GSK3 pathways, allowing ERK signaling to promote differentiation and reducing Wnt signaling to enable lineage specification. This shift initiates a cascade of biological changes, including the downregulation of naive markers (such as Nanog and Esrrb) and upregulation of formative markers (such as Otx2), along with epigenetic remodeling, altered chromatin accessibility, and a metabolic shift from glycolysis to oxidative phosphorylation (see Figure 1) (Wray et al. 2011).

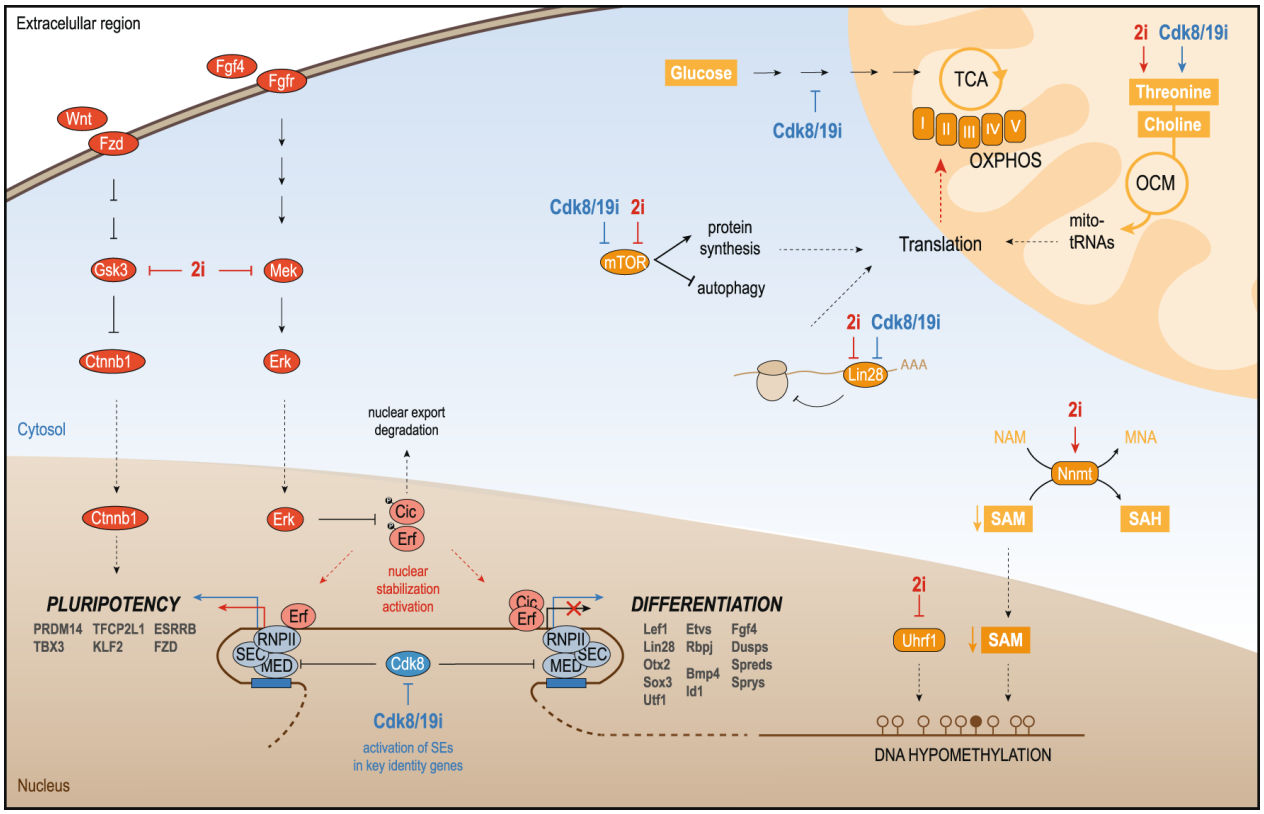

Figure 1: **2i mechanism of stabilizing naive pluripotency**
The inhibition of Mek/Erk through 2i-dependent mechanisms prevents the phosphorylation of Erf and Cic, resulting in the stabilization of these factors within the nucleus. These transcriptional effectors may play a role in activating genes associated with naïve pluripotency while repressing those involved in primed/formative states. Cdk8/19i influences the phosphorylation of proteins that are part of the elongation and transcription machinery, including the super elongation complex (SEC), RNA polymerase II (RNPII), and Mediator (MED). This could lead to the rapid activation of genes regulated by super-enhancers, which include factors that stabilize the naïve state as well as certain factors from the primed/formative states that are also upregulated in response to Cdk8/19i. The activation of the naïve circuitry through 2i and Cdk8/19i may support the enhanced mitochondrial oxidative phosphorylation (OXPHOS) capacity of these cells. The degradation of Lin28a might alleviate its repressive effects on target mRNAs, leading to the upregulation of mitochondrial proteins among others. The balance between autophagy and protein synthesis may sustain the translation of these mitochondrial components. Additionally, the re-wiring of mitochondrial sources for one-carbon units may further support the translation of mitochondrial-encoded proteins. Furthermore, the increased availability of S-adenosylmethionine (SAM) levels, due to NNMT upregulation, combined with the impairment of DNA methylation caused by Uhrf1 degradation, contributes to the characteristic hypomethylated DNA state typical of 2i conditions (Martinez-Val et al. 2021).

ERK and mTOR signalling pathway and its effects on translation

ERK signaling influences translation primarily by modulating the activity of eukaryotic translation initiation factor 4E (eIF4E) (Lavoie, Gagnon, and Therrien 2020). ERK activates MAPK-Interacting Kinases (MNK)1 and MNK2 kinases, which phosphorylate eIF4E, enhancing its binding to the mRNA cap structure and promoting the formation of the eIF4F complex (see Figure 2). This complex, consisting of eIF4E, eIF4G, and eIF4A, is essential for recruiting the ribosome to mRNA, scanning for the start codon, and initiating protein synthesis. ERK signaling can selectively enhance the translation of specific mRNAs, particularly those with complex 5′ UTRs, leading to the production of proteins involved in cell growth and survival. Additionally, ERK interacts with other pathways, such as mTOR, to coordinate translation regulation in response to various signals, thus optimizing cellular responses to environmental changes.

The mTOR pathway complements the role of ERK in translation regulation. It comprises two distinct complexes: mTORC1 and mTORC2. mTORC1 is primarily responsible for regulating protein synthesis, autophagy, and metabolism, while mTORC2 is involved in cell survival and cytoskeletal organization. Activation of mTORC1 occurs through nutrient availability, with amino acids serving as key activators, and through growth factors that activate the PI3K-AKT pathway (see Figure 2) (Ma and Blenis 2009). mTORC1 influences translation by phosphorylating 4E-BP1, which releases eIF4E to initiate cap-dependent translation (see Figure 2) (Ma and Blenis 2009). Furthermore, mTORC1 activates S6K1, which phosphorylates ribosomal protein S6, enhancing the translation of ribosomal proteins and elongation factors. These actions collectively increase the global protein synthesis capacity of the cell (see Figure 2) (Ma and Blenis 2009).

The ERK signaling pathway

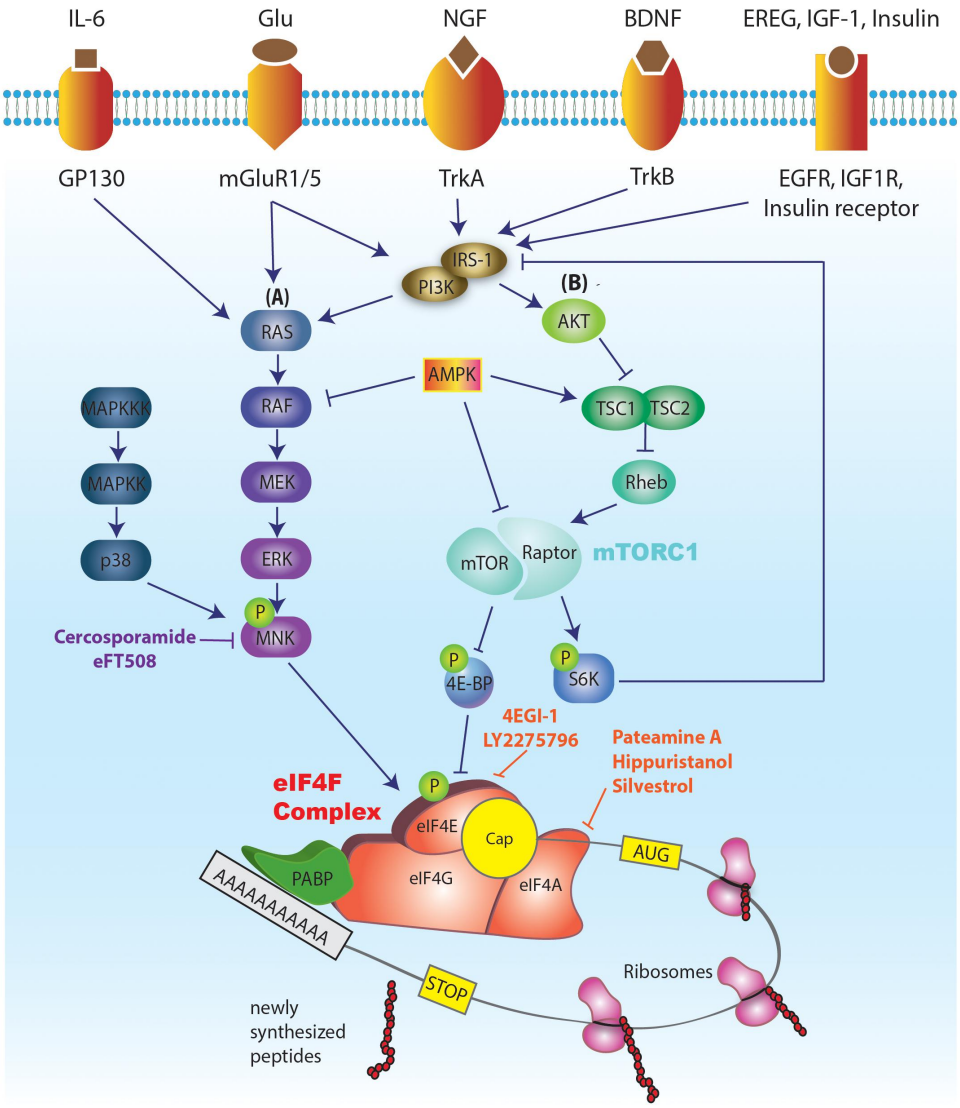

Figure 2: **ERK and mTORs effects on eIF4E regulation**
The ability of eIF4E to bind to the mRNA cap makes it a key regulator of translation. A crucial step in translation initiation involves eIF4E binding to the mRNA cap, facilitating the formation of the eIF4F complex on the cap structure (comprising a 7mGp linked to the first nucleotide). The eIF4F complex includes eIF4E, eIF4G (a scaffolding protein), and eIF4A (a helicase). The successful assembly of the eIF4F complex on the mRNA cap promotes the recruitment of the pre-initiation complex (PIC), followed by scanning of the 5′ UTR to locate the start codon AUG and the joining of the 60S ribosomal subunit, marking the completion of translation initiation. eIF4E acts as a downstream effector of both the mTORC1 pathway (via 4E-BP-dependent repression) and the ERK pathway (via eIF4E phosphorylation by MNK 1/2). The activities of mTORC1 and ERK signaling pathways are influenced by various external signals and internal cues (Uttam et al. 2018).

Additional mechanisms influencing translation

Naturally, there are many other described mechanisms that can influence translation. RNA-binding proteins (RBPs) can bind mRNA through modular RNA-binding domains, which primary sequence of amino acids dictates the specificity of the RBP (Lunde, Moore, and Varani 2007). By binding the mRNA, these proteins can influence the turnover rate of their specific mRNA, stability of the mRNA as well as form complexes with other proteins to recruit the ribonucleoproteins (RNPs) of the translation machinery (Glisovic et al. 2008).
Another prominent example of post-transcriptional regulation is the interaction of mircoRNAs (miRNAs) with mRNAs. miRNAs have been shown to bind targeted sequences at the 3' untranslated region (UTR) for blocking translation, deadenylation of the mRNA or for decapping (Huntzinger and Izaurralde 2011). In mESCs specifically, families of miRNAs have been shown to maintain pluripotency and suppressing genes that drive differentiation (Melton, Judson, and Blelloch 2010). The fact that, ERK signalling globally downregulates miRNAs, further highlights the role of miRNAs to maintain pluripotency and ERKs to exit the naive state (Sun et al. 2016).
Lastly epigenetic regulation, although indirectly, can influence translation. In mESCs, the incorporation of specific histone variants, such as H3.3, has been shown to modulate the chromatin environment. Regulatory complexes can modify chromatin to control gene expression during development by influencing the accessibility and stability of specific genomic regions (Banaszynski et al. 2013).

Translation on demand

We define ToD as an increase in protein abundance due to enhanced translation, occurring without any changes in mRNA levels of the corresponding gene. In other words, ToD is a regulatory mechanism where translation efficiency is selectively increased for transcripts of specific genes. This upregulation allows for rapid adjustments in protein levels, supporting cellular responses that need immediate protein synthesis without requiring new mRNA transcription. Such translation driven increases could be particularly useful in situations where quick adaptations are necessary, as it enables cells to respond to changes in the

environment or developmental cues by rapidly elevating the protein output from existing mRNA pools.

In vivo, there are already described mechanisms that facilitate forms of selective, rapid translation of mRNAs. For example, neurons need local translation at active synapses for quick cellular responses (Holt and Schuman 2013). In early embryos, maternal mRNA is stored and activated after fertilization, enabling the embryo to rapidly synthesize proteins necessary for early development without new mRNA synthesis (De Leon, Johnson, and Bachvarova 1983). In stress response, rapid and selective synthesis of stress proteins, such as heat shock proteins (HSPs), is facilitated by ToD, allowing cells to quickly adapt and protect themselves under stress conditions without increasing mRNA levels (Holcik and Sonenberg 2005).

Aim of this study

In this project, we aim to identify ToD candidates from a multi-omic dataset, which includes a time series of bulk RNA sequencing and proteomic data (Yang et al. 2019). These multiple omic layers enable us to explore how protein abundance responds to changes in mRNA levels and to identify genes that may be regulated through ToD mechanisms. Next to finding genes that are subject to ToD regulation, we want to illustrate why these specific genes are prioritized and to uncover the underlying mechanisms of this regulation. This involves exploring the biological roles and functional significance of these genes in cellular processes. Additionally, we are interested in understanding the gene regulatory network and proteome changes that occur as mESCs exit pluripotency, as these regulatory networks are most likely to be regulated by ToD.

By analyzing the temporal dynamics within these data layers, we can distinguish cases where protein levels increase independently of mRNA changes, a hallmark of ToD regulation. This approach provides a comprehensive view of the post-transcriptional mechanisms that contribute to protein abundance, offering insights into gene-specific regulatory processes. Understanding the more nuanced relationship between RNA and protein expression, as well as the timing of regulation, will provide valuable insights into the funda-

mental principles of molecular biology and enhance our comprehension of cellular regulatory networks as well as early development.
Furthermore, this study should serve as a framework to detect ToD candidates in by leveraging the use of transciptome and proteom data together. We also aim to validate our findings and approaches with two other datasets to check the validity of our methods and see if we obtain similar results. This validation will help ensure that the mechanisms identified are robust and applicable across different biological contexts.

Methods and tools

Experimental Approach and Workflow

To research ToD, we exploit the fact that protein translation is largely inhibited under 2i conditions. By sampling mouse embryonic stem cells under 2i conditions and then periodically after 2i removal, we can take snapshots of the early changes in the transcriptome and proteome (see Figure 3). Upon 2i removal, we trigger the transition from the naive pluripotent state to the primed formative state. This allows us to detect changes in the translational behavior of genes aiding this transition.
We connect transcriptomic and proteomic dynamics by comparing their Log Fold Change (LFC) between time points. This way, we can compare qualitative trends for transcript and protein abundance rather than comparing them by count values. After identifying genes that are likely ToD-regulated, we assess their biological function and place them in the context of early development (see Figure 3).

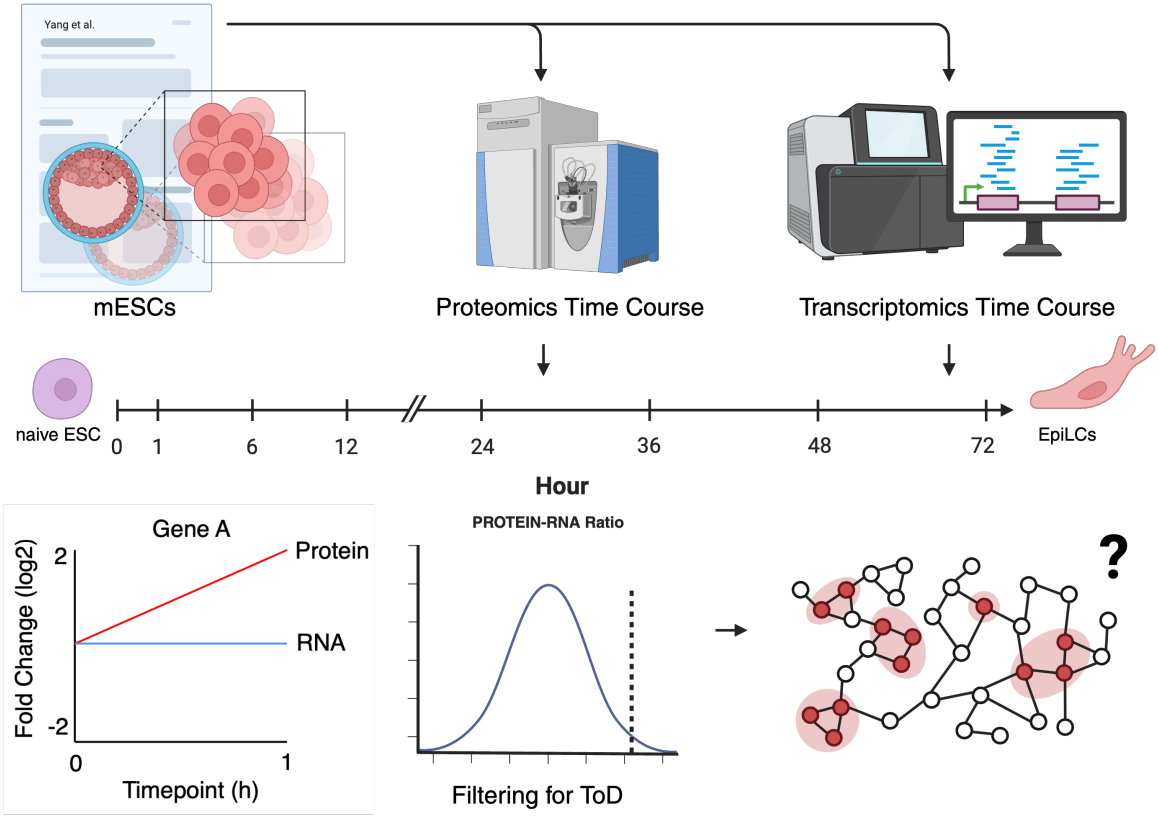

Figure 3: **Experimental Approach and Workflow** Exemplary workflow for the Yang dataset. Proteomic and transcriptomic time courses were used to examine LFC changes at different time intervals. Filtering criteria were applied to determine if a specific gene is likely to be ToD regulated. ToD candidates were tested for functional enrichment and other annotations methods to determine their role in early development.

Data overview and processing

The data for this project is comprised of three datasets with our main focus being the Yang Data, due to its higher temporal resolution in early time points and more consistent time course in the proteomic data (see Table 1). This dataset was sourced from Supplementary Table S2 (temporal proteomic data) and Supplementary Table S3 (temporal transcriptomic data) in the publication Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency (Yang et al. 2019).

The temporal proteomic data is in the form of Log2 Label-Free Quantification (LFQ) intensities and included the following time points: 0min, 30min, 1h, 6h, 12h, 24h, 36h, 48h,

and 72h. Each time point had four biological replicates, which were also pooled. The temporal transcriptomic data consisted of non-normalized transcript counts and included the following time points: 0min, 1h, 6h, 12h, 24h, 36h, 48h, and 72h. Each time point had two biological replicates, which were averaged. Since the transcriptomic dataset did not include a 30min time point, this time point was removed from the proteomic data for further analysis (see Table 1).

To increase our accuracy in calling meaningful ToD candidate genes, we realigned the transciptomic data with the GRCm38 reference genome using STAR (Mudge et al. 2025; Dobin et al. 2013).

The Leeb data has a temporal resolution from 0 to 32 hours with 4 hour intervals. Each time point has four biological replicates in the proteomic time course and no replicates in the transcriptomic time course (see Table 1). Gaussian Process Regression (GPR) was used to reduce noise in the transcriptomic time course data.

The Leeb MC data includes time course data for 0h, 4h and 8h for the purpose of calling ToD candidates. Additionally, it includes a time course with the same temporal resolution, where 2i conditions were kept while a medium change was still applied to the cells. Each time point has four biological replicates in the proteomic time course and three replicates in the transcriptomic time course (see Table 1).

Table 1: Datasets

Dataset	Temporal Resolution (h)	Replicates Transciptomics	Replicates Proteomics
Yang Data	0, 1, 6, 12, 24, 36, 48, 72	2	4
Leeb Data	0, 4, 8, 12, 24, 32	1	4
Leeb MC Data (Medium Change)	0, 4, 8	3	4

Transcriptomic Data Normalization

For the Yang dataset as well as the Leeb MC dataset we used DESeq2 to normalize the transciptomic data (Love, Huber, and Anders 2014). DESeq2's normalization procedure operates on the assumption that the majority of genes exhibit no differential expression. To normalize gene counts, DESeq2 uses the median of ratios method. This involves dividing the count of each gene in a sample by its geometric mean across all samples, resulting in a ratio (see Equation 1; Equation 2). The geometric mean serves as a measure of central tendency, representing the average gene count across samples, and provides a standard for comparing individual sample counts. The size factor for a given sample is then determined by calculating the median of these ratios (see Equation 3). This method is robust, accounting for sequencing depth and RNA composition, and effectively compensates for sample specific biases.

$$\text{Geometric Mean}_g = \left(\prod_{s=1}^{n} \text{Count}_{gs} \right)^{1/n} \tag{1}$$

- Geometric Mean_g : The geometric mean of counts for gene g
- Count_{gs} : The count of gene g in sample s
- n : The total number of samples

$$\text{Ratio}_{gs} = \frac{\text{Count}_{gs}}{\text{Geometric Mean}_g} \tag{2}$$

$$\text{Size Factor}_s = \text{median}\left(\text{Ratio}_{gs} \mid \text{for all } g\right) \tag{3}$$

Calculation of Fold Changes and Ratios

LFC values between consecutive time points for proteomic and transcriptomic time courses were calculated with Equation 4.

$$\text{logFC}_{\text{between time points}} = \text{logFC}_2\left(\frac{\text{Expression}_{\text{time point}_{n+1}}}{\text{Expression}_{\text{time point}_n}}\right) \tag{4}$$

LFC values against the first time point (0h) for proteomic and transcriptomic time courses were calculated with Equation 5.

$$\text{logFC}_{\text{against time point 1}} = \text{logFC}_2\left(\frac{\text{Expression}_{\text{time point}_n}}{\text{Expression}_{\text{time point}_1}}\right) \tag{5}$$

We calculated a Protein-to-RNA-LFC-Ratio (LFC-Ratio) to highlight genes with greater changes in protein than RNA, identifying possible ToD candidates (see Equation 6). A positive Protein-to-RNA FC Ratio indicates that protein expression has increased more than RNA expression, suggesting potential translational upregulation through ToD.

$$\text{LFC-Ratio} = \text{logFC}_{\text{between time points protein}} - \text{logFC}_{\text{between time points RNA}} \tag{6}$$

Gene Overrepresentation Analysis with clusterProfiler

We performed Gene Overrepresentation Analysis (ORA) on potential ToD candidates at each time point using the R package clusterProfiler (Yu et al. 2012). This analysis aimed to identify overrepresented biological functions, molecular activities, or cellular localizations among ToD candidates compared to a background reference. Using a hypergeometric model, clusterProfiler evaluates the enrichment of gene set terms within each set of ToD candidates, generating p-values to quantify significance. To control for false discovery rates (FDR), we applied Benjamini-Hochberg correction to the resulting p-values (Benjamini and Hochberg 1995). This approach highlighted time-specific biological processes potentially regulated at the translational level among ToD candidates. The corresponding complete data set (genes both present in the proteomic and transcriptomic data) was used as a background (or universe) for the analysis.

Gene Set Enrichment Analysis

Motif enrichment

Software

Results

Detecting Translation on Demand Candidates

A hallmark of ToD regulation is a rapid increase in translation, while the rate of transcription stays stable. As a filtering criteria of ToD regulation, we selected only genes who showed RNA LFCs which stayed inside a stable range for each time interval. The stringency of this range depends on the biological question we want so answer with a specific list of ToD candidates. For example, for gene set enrichment it might make sense to broaden the range we define as stable to detect functional trends in a certain time interval which would not be detectable otherwise. If not stated otherwise we used a stable LFC RNA range from minus one to one. The next filtering step to determine if a gene might be subject to ToD regulation at a specific time interval was to filter for a LFC-Ratio of at least one to ensure a substantial increase in protein levels while maintaining a relatively stable RNA LFC, which filters for robust protein level changes independent of RNA fluctuations (see Figure 4 A). Again, this LFC-Ratio threshold depends on the biological context and needs to be adapted for the dataset in question. If not stated otherwise, we used a LFC-Ratio threshold of one. This filtering step results in an initial ToD candidate list generated with the Yang Data of 113 ToD candidates for the first time interval (0h to 1h), 103 for the second (1h to 6h) and 78 for the third (6h to 12h) etc. (see Figure 4 B). Corrected for time, the number of ToD candidates are 103, 17 and 13 for the first three time intervals respectivly (see Figure 4 B). This trend continues in later time intervals. The fact, that we detect more ToD candidates per unit of time in early time intervals might be due to the relevance of ToD regulation during these early develompental time intervals, because rapid protein systhesis of specific transcipts is needed at that time to initiate priming for differentiation.

To make the ToD candidate calling less subject to noise, we implemented a third filtering step. After a gene passes the LFC-Ratio threshold at a given time interval, we look at its protein LFC trajectory normalized for the first time point (0h) and see if the subsequent LFC value after the threshold passing is within a certain range defined by the time interval where it passed the LFC-Ratio threshold. For example, if a gene passed the LFC-Ratio threshold at the 0h to 1h time interval, we look in the protein LFC trajectory of that gene for 1h vs 0h and the next LFC value at 6h vs 0h. The LFC value at the 6h vs 0h time interval is only allowed to go down by 30% compared to the previous value at 1h vs 0h to eliminate genes that showed higher levels of noise in their protein LFC trajectory. After applying this filtering method, the number of ToD candidates for the first three time intervals were 61, 52, and 34 respectively (see Figure 4 C). The trend of having more ToD candidates when correcting for time in the early time intervals remained evident after applying the filtering step.

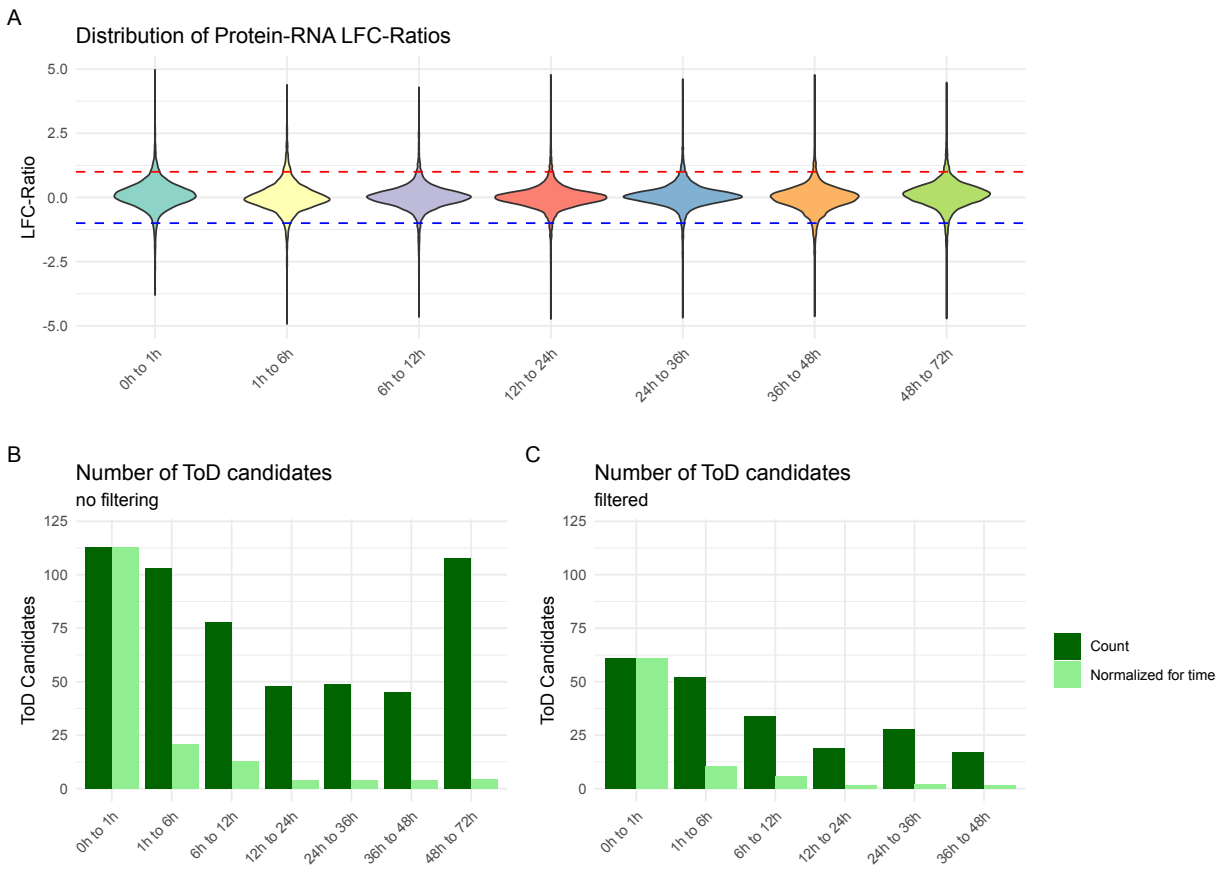

Figure 4: **Filtering for ToD candidates in the Yang Data**
(A) Distribution of LFC-Ratios at different time intervals. The dashed red line (LFC-Ratio of 1) shows the LFC-Ratio threshold a gene needs to pass. The dashed blue line (LFC-Ratio of -1) shows the opposite behaviore where we see a decreased protein expression.
(B) The barplot shows the number of ToD candidates at each time interval which passed the LFC-Ratio of one. The dark green bars show the absolute number of ToD candidates at a given time interval. The light green bars show the number of ToD candidates corrected for time.
(C) The barplot shows the number of ToD candidates at each time interval which passed the LFC-Ratio and which passed the third filtering step. The dark green bars show the absolute number of ToD candidates at a given time interval. The light green bars show the number of ToD candidates corrected for time.

Early development is shaped by ToD candidates

Upon examining the filtered ToD candidates across various temporal intervals, it is evident that these candidates include numerous genes crucial for early developmental processes. Specifically, during the initial interval from 0 to 1 hour, a significant proportion of the

ToD candidates are associated with developmental functions. Notably, genes such as Agap1, Dnajc14, Emc7, Rab27a, Vti1a and Napb play roles in protein trafficking and vesicular transport. This observation suggests that protein trafficking is likely subjected to substantial demand during this initial period, coinciding with the onset of protein translation following the removal of the 2i condition. The increased load on protein trafficking mechanisms may reflect the cellular necessity to rapidly establish the proteomic landscape required for subsequent developmental stages.

Not suprisingly then, genes asscotiated with cytosceletal regulation like Cdc42bpa, Cep95, Plekhh1 and Stil are represented among ToD candidates in the first time interval.

Another group of genes, such as Celf2, Rex2, and Rsrp1, are involved in post-transcriptional regulation. Specifically, Celf2 encodes an RNA-binding protein that plays a critical role in regulating several post-transcriptional events. It is involved in pre-mRNA alternative splicing, mRNA translation, and stability. Celf2 mediates exon inclusion and/or exclusion in pre-mRNA, which is subject to tissue specific and developmentally regulated alternative splicing (Ladd, Charlet-B., and Cooper 2001). Lasty, another group of genes like Axin1, Csnk1g2, Jag1, Rap2c and Rgs19 contributes to regulatory networks that govern developmental processes.

ToD candidats contributing to the mentioned processes can also be found in subsequent time intervals, like the splicing factor Srsf11 or developmental regulators like Epha2 in time interval 1h to 6h among others (see Table 2).

These findings illustrate that the ToD candidates identified in the Yang dataset play a crucial role in facilitating the transition from the naive pluripotent state to the primed formative state. By contributing to essential processes such as protein trafficking, cytoskeletal regulation, post-transcriptional modification, and developmental signaling, these genes help orchestrate the complex molecular changes necessary for early development.

Gene over-representation analysis reveals time dependent pathway regulation of developmental terms by ToD candidates

Gene ontology (GO) as well as Hallmark ORA provided further validation of these observations. In the 0h to 1h time interval, we identified a nuanced overrepresentation of Notch signaling as well as Wnt/β-catenin signaling (see Figure 5 A). The prominence of these pathways during the initial time interval suggests that they are among the first to respond to the removal of the 2i conditions. This response likely initiates a cascade of molecular events that prepare the cells for subsequent differentiation, highlighting their importance in the early phases of transitioning from a naive to a primed state. Both pathways are well known for their involvement in embryonic development (Zhou et al. 2022; Liu et al. 2022). Additionally, in the 1h to 6h time interval ToD candidates associated with TNFA signaling and TGF beta signaling where significantly over-represented (see Figure 5 B). During embryogenesis and organogenesis, TNFA signaling plays a crucial role in mediating inflammatory responses and regulating apoptosis, which can influence tissue remodeling and development (You et al. 2021). TNFA's involvement in these processes underscores its importance in maintaining cellular homeostasis and promoting the appropriate differentiation pathways necessary for proper tissue formation. Similarly, TGF-beta signaling is pivotal in embryonic development, as it regulates cell proliferation, differentiation, and migration. TGF-beta signaling is essential for the establishment of the body plan and the development of various tissues and organs (Deng et al. 2024). The overrepresentation of TGF-beta signaling in this interval suggests that ToD candidates may be actively contributing to the cellular transitions required for mESCs to exit the naive state and progress towards lineage specification.

Next to these signalling Hallmark terms, we saw a significant over-representation of GO biological processes associated with development in that time interval (see Figure 5 C). Tissue specific developmental terms such as tube morphogenesis, vasculature development, epithelial cell proliferation, skeletal system development, ossification and response to growth factor show the progression toward priming for lineage commitment. This priming requires sophisticated and rapid reprogramming of the epigenetic, gene

expression, and protein landscapes of ESCs. ToD might be a central mechanism enabling cells to make these necessary rapid changes.

Lastly, ToD candidates of the 6h to 12h time interval displayed a significant over-representation in GO molecular functions, specifically in calcium ion binding and calcium dependent protein binding as well as a nuanced over-representation in calcium dependent phospholipid binding (see Figure 5 D). Calcium-dependent signaling is a central and conserved mechanism essential for development by orchestrating key processes such as cell proliferation, differentiation, and migration. It plays a crucial role in various stages of embryogenesis, including the establishment of body plans, organogenesis, and tissue morphogenesis. The ability of calcium to act as a versatile second messenger allows it to integrate and decode diverse extracellular signals, thereby regulating the dynamic cellular environments necessary for proper development (Stewart and Davis 2019).

These findings underscore the critical role that ToD candidate genes play in early development. The proteins encoded by these genes are essential for facilitating the transition from a naive to a formative primed state. This aligns with our initial hypothesis that ToD may serve as a cellular mechanism that enables swift alterations in the proteomic state of stem cells during development. By integrating cellular signaling cues, ToD effectively regulates gene networks to ensure a proper developmental transition.

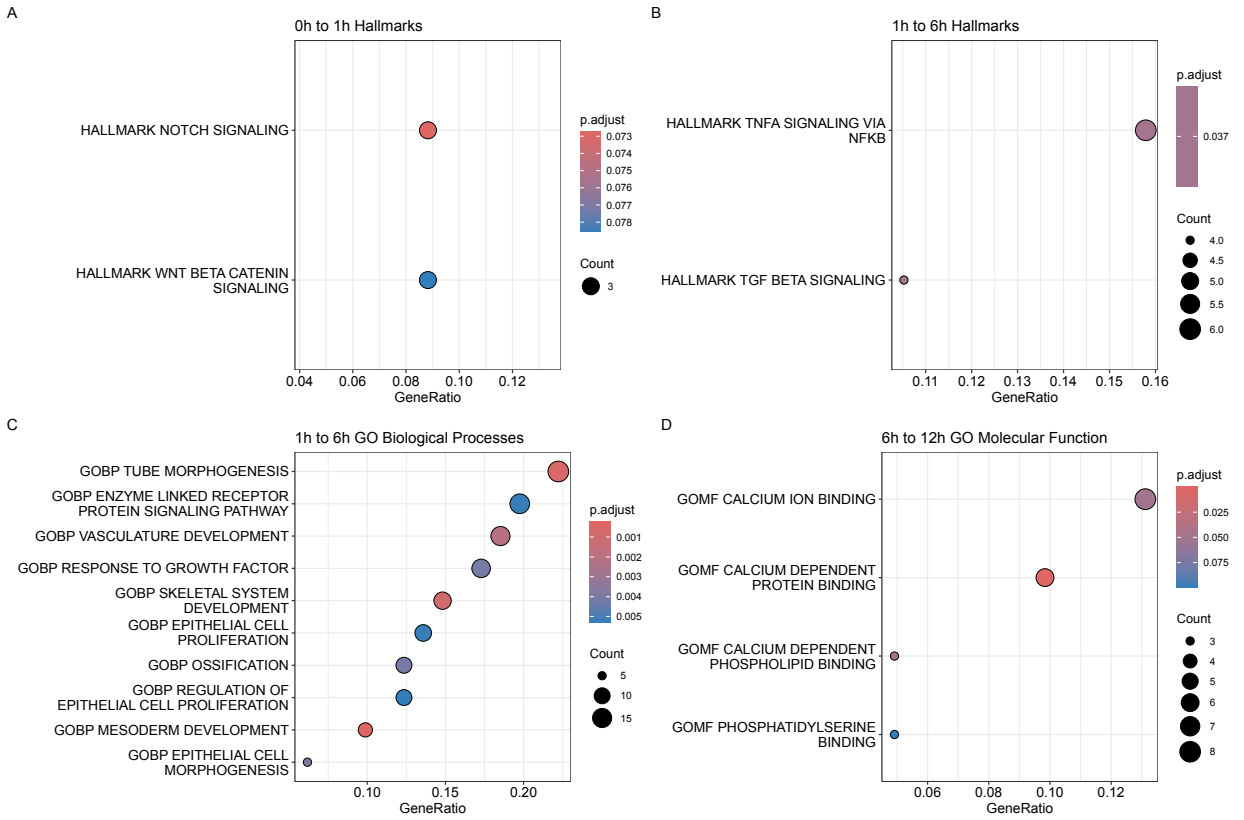

Figure 5: **Gene overrepresentation analysis of early ToD candidates in Yang Data**
(A) Hallmark term over-representation analysis at the 0h to 1h time interval. Adjusted p-value threshold of 0.1. Stable transcript range for ToD call of -2 to 2.
(B) Hallmark term over-representation analysis at the 1h to 6h time interval. Adjusted p-value threshold of 0.05. Stable transcript range for ToD call of -2 to 2.
(C) GO term over-representation analysis for biological processes (BP) at the 1h to 6h time interval. Adjusted p-value threshold of 0.05. Stable transcript range for ToD call of -2 to 2.
(D) GO term over-representation analysis for molecular function (MF) at the 6h to 12h time interval. Adjusted p-value threshold of 0.05. Stable transcript range for ToD call of -2 to 2.
The color gradient represents the adjusted p-value, with red indicating a lower adjusted p-value and blue indicating a higher adjusted p-value. The size of each dot corresponds to the number of genes present in the respective gene set from the dataset.

ToD candidates regulate gene expression as transcription factors

ToD is influenced by RNA binding proteins

GSEA reveals

Discussion

Conclusion

References

Banaszynski, Laura A., Duancheng Wen, Scott Dewell, Sarah J. Whitcomb, Mingyan Lin, Nichole Diaz, Simon J. Elsässer, et al. 2013. "Hira-Dependent Histone H3.3 Deposition Facilitates PRC2 Recruitment at Developmental Loci in ES Cells." *Cell* 155 (1): 107–20. https://doi.org/10.1016/j.cell.2013.08.061.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300. https://www.jstor.org/stable/2346101.

Boroviak, Thorsten, Remco Loos, Paul Bertone, Austin Smith, and Jennifer Nichols. 2014. "The Ability of Inner-Cell-Mass Cells to Self-Renew as Embryonic Stem Cells Is Acquired Following Epiblast Specification." *Nature Cell Biology* 16 (6): 513–25. https://doi.org/10.1038/ncb2965.

De Leon, Victor, Andrew Johnson, and Rosemary Bachvarova. 1983. "Half-Lives and Relative Amounts of Stored and Polysomal Ribosomes and Poly(A)+ RNA in Mouse Oocytes." *Developmental Biology* 98 (2): 400–408. https://doi.org/10.1016/0012-1606(83)90369-X.

Dejosez, Marion, and Thomas P. Zwaka. 2012. "Pluripotency and Nuclear Reprogramming." *Annual Review of Biochemistry* 81 (Volume 81, 2012): 737–65. https://doi.org/10.1146/annurev-biochem-052709-104948.

Deng, Ziqin, Tao Fan, Chu Xiao, He Tian, Yujia Zheng, Chunxiang Li, and Jie He. 2024. "TGF-β Signaling in Health, Disease and Therapeutics." *Signal Transduction and Targeted Therapy* 9 (1): 61. https://doi.org/10.1038/s41392-024-01764-w.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-seq Aligner." *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Glisovic, Tina, Jennifer L. Bachorik, Jeongsik Yong, and Gideon Dreyfuss. 2008. "RNA-binding Proteins and Post-Transcriptional Gene Regulation." *FEBS Letters* 582 (14):

1977–86. https://doi.org/10.1016/j.febslet.2008.03.004.

Holcik, Martin, and Nahum Sonenberg. 2005. “Translational Control in Stress and Apoptosis.” *Nature Reviews Molecular Cell Biology* 6 (4): 318–27. https://doi.org/10.1038/nrm1618.

Holt, Christine E., and Erin M. Schuman. 2013. “The Central Dogma Decentralized: New Perspectives on RNA Function and Local Translation in Neurons.” *Neuron* 80 (3): 648–57. https://doi.org/10.1016/j.neuron.2013.10.036.

Huntzinger, Eric, and Elisa Izaurralde. 2011. “Gene Silencing by microRNAs: Contributions of Translational Repression and mRNA Decay.” *Nature Reviews Genetics* 12 (2): 99–110. https://doi.org/10.1038/nrg2936.

Ladd, Andrea N., Nicolas Charlet-B., and Thomas A. Cooper. 2001. “The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing.” *Molecular and Cellular Biology* 21 (4): 1285–96. https://doi.org/10.1128/MCB.21.4.1285-1296.2001.

Lavoie, Hugo, Jessica Gagnon, and Marc Therrien. 2020. “ERK Signalling: A Master Regulator of Cell Behaviour, Life and Fate.” *Nature Reviews Molecular Cell Biology* 21 (10): 607–32. https://doi.org/10.1038/s41580-020-0255-7.

Liu, Jiaqi, Qing Xiao, Jiani Xiao, Chenxi Niu, Yuanyuan Li, Xiaojun Zhang, Zhengwei Zhou, Guang Shu, and Gang Yin. 2022. “Wnt/β-Catenin Signalling: Function, Biological Mechanisms, and Therapeutic Opportunities.” *Signal Transduction and Targeted Therapy* 7 (1): 3. https://doi.org/10.1038/s41392-021-00762-6.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Lunde, Bradley M., Claire Moore, and Gabriele Varani. 2007. “RNA-binding Proteins: Modular Design for Efficient Function.” *Nature Reviews Molecular Cell Biology* 8 (6): 479–90. https://doi.org/10.1038/nrm2178.

Ma, Xiaoju Max, and John Blenis. 2009. “Molecular Mechanisms of mTOR-mediated Translational Control.” *Nature Reviews Molecular Cell Biology* 10 (5): 307–18. https://doi.org/10.1038/nrm2672.

Martinez-Val, Ana, Cian J. Lynch, Isabel Calvo, Pilar Ximénez-Embún, Fernando Garcia, Eduardo Zarzuela, Manuel Serrano, and Javier Munoz. 2021. "Dissection of Two Routes to Naïve Pluripotency Using Different Kinase Inhibitors." *Nature Communications* 12 (1): 1863. https://doi.org/10.1038/s41467-021-22181-5.

Melton, Collin, Robert L. Judson, and Robert Blelloch. 2010. "Opposing microRNA Families Regulate Self-Renewal in Mouse Embryonic Stem Cells." *Nature* 463 (7281): 621–26. https://doi.org/10.1038/nature08725.

Mudge, Jonathan M, Sílvia Carbonell-Sala, Mark Diekhans, Jose Gonzalez Martinez, Toby Hunt, Irwin Jungreis, Jane E Loveland, et al. 2025. "GENCODE 2025: Reference Gene Annotation for Human and Mouse." *Nucleic Acids Research* 53 (D1): D966–75. https://doi.org/10.1093/nar/gkae1078.

Needham, Elise J., Benjamin L. Parker, Timur Burykin, David E. James, and Sean J. Humphrey. 2019. "Illuminating the Dark Phosphoproteome." *Science Signaling* 12 (565): eaau8645. https://doi.org/10.1126/scisignal.aau8645.

Nichols, Jennifer, and Austin Smith. 2009. "Naive and Primed Pluripotent States." *Cell Stem Cell* 4 (6): 487–92. https://doi.org/10.1016/j.stem.2009.05.015.

Rayon, Teresa, Despina Stamataki, Ruben Perez-Carrasco, Lorena Garcia-Perez, Christopher Barrington, Manuela Melchionda, Katherine Exelby, et al. 2020. "Species-Specific Developmental Timing Is Associated with Differences in Protein Stability in Mouse and Human." *Science (New York, N.Y.)* 369 (6510): eaba7667. https://doi.org/10.1126/science.aba7667.

Stewart, Teneale A., and Felicity M. Davis. 2019. "An Element for Development: Calcium Signaling in Mammalian Reproduction and Development." *Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research*, 15th European Symposium on Calcium, 1866 (7): 1230–38. https://doi.org/10.1016/j.bbamcr.2019.02.016.

Sun, Hui-Lung, Ri Cui, JianKang Zhou, Kun-yu Teng, Yung-Hsuan Hsiao, Kotaro Nakanishi, Matteo Fassan, et al. 2016. "ERK Activation Globally Downregulates miRNAs Through Phosphorylating Exportin-5." *Cancer Cell* 30 (5): 723–36. https://doi.org/10.1016/j.ccell.2016.10.001.

Uttam, Sonali, Calvin Wong, Theodore J. Price, and Arkady Khoutorsky. 2018. "eIF4E-

Dependent Translational Control: A Central Mechanism for Regulation of Pain Plasticity." *Frontiers in Genetics* 9 (October). https://doi.org/10.3389/fgene.2018.00470.

Wray, Jason, Tüzer Kalkan, Sandra Gomez-Lopez, Dominik Eckardt, Andrew Cook, Rolf Kemler, and Austin Smith. 2011. "Inhibition of Glycogen Synthase Kinase-3 Alleviates Tcf3 Repression of the Pluripotency Network and Increases Embryonic Stem Cell Resistance to Differentiation." *Nature Cell Biology* 13 (7): 838–45. https://doi.org/10.1038/ncb2267.

Yang, Pengyi, Sean J. Humphrey, Senthilkumar Cinghu, Rajneesh Pathania, Andrew J. Oldfield, Dhirendra Kumar, Dinuka Perera, et al. 2019. "Multi-Omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency." *Cell Systems* 8 (5): 427–445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

Ying, Qi-Long, Jason Wray, Jennifer Nichols, Laura Batlle-Morera, Bradley Doble, James Woodgett, Philip Cohen, and Austin Smith. 2008. "The Ground State of Embryonic Stem Cell Self-Renewal." *Nature* 453 (7194): 519–23. https://doi.org/10.1038/nature06968.

You, Kai, Hui Gu, Zhengwei Yuan, and Xuewen Xu. 2021. "Tumor Necrosis Factor Alpha Signaling and Organogenesis." *Frontiers in Cell and Developmental Biology* 9 (July): 727075. https://doi.org/10.3389/fcell.2021.727075.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. "clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters." *Omics: A Journal of Integrative Biology* 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.

Zhou, Binghan, Wanling Lin, Yaling Long, Yunkai Yang, Huan Zhang, Kongming Wu, and Qian Chu. 2022. "Notch Signaling Pathway: Architecture, Disease, and Therapeutics." *Signal Transduction and Targeted Therapy* 7 (1): 95. https://doi.org/10.1038/s41392-022-00934-y.

Appendices

Supplementary figures

Supplementary tables

Table 2: ToD candidate genes from the Yang Data

Time Interval	ToD candidate genes
0h to 1h	Agap1, AI987944, Arhgap27, Axin1, Bbs5, Bbs9, Bex4, Btbd9, Cacna1a, Cbx6, Ccdc91, Ccnf, Cdc42bpa, Celf2, Cep95, Cgn, Csnk1g2, Csrnp2, Dcbld1, Dnajc14, Ebpl, Emc7, Enox2, Epha1, Fbxo9, Gba2, Gm5141, Gstp2, Hps4, Jag1, Napb, Nudt22, Odc1, Otud5, Plch1, Plekhh1, Rab27a, Rap2c, Rassf1, Rex2, Rgs19, Rhbdd3, Rnf187, Rsrp1, Slc38a2, Smim20, Sowahb, Stil, Taf11, Tex15, Tm2d3, Tmc6, Tmub2, Tnks, Tpst2, Vangl1, Vti1a, Zcchc2, Zfp850, Zik1, Zmym6
1h to 6h	Agtrap, Alg3, Arhgap30, Avpi1, Bmpr1a, Capn3, Ccdc112, Cdh1, Coro2b, Ctdsp2, Dip2c, Epha2, Fam168b, Fat1, Fzr1, G2e3, Hsf2, Ifrd1, Irs2, Itgb3, Jak1, Kifc5b, Lgals8, Lin28b, Maml1, Mapkapk2, Micall1, Nfe2l2, Nptx2, Nr2c1, Pbx1, Pde8a, Pecam1, Rimklb, Rnf187, Rps6ka6, Slc19a2, Slc9a8, Sptb, Sqstm1, Srsf11, Ss18l2, Sulf1, Suv39h1, Syde2, Tex15, Tmem106b, Traip, Trim37, Trmt44, Zfand5, Zfp84
6h to 12h	Adck2, Akap11, Bbx, Bcl2l11, Bex4, Birc2, Bnc2, Cenpj, Cers5, Cox7a2, Dzip1l, Eif1b, Enox2, Gpc4, Htra1, Kctd9, Klhdc10, Lnx2, Mpped2, Myo1d, Ormdl3, Parp9, Pnp, Rab40c, Rapgef2, Rbm44, Rims2, S100a10, Scd1, Sdc1, Slc25a31, Tmem17, Ttll4, Ube2q2

Time Interval	ToD candidate genes
12h to 24h	Alox15, Arhgef25, Bbc3, Camkv, Dapk1, Dgkd, Mapk6, Mief1, Mthfsd, Nuak2, Phf6, Pop5, Rab27a, Sh3kbp1, Slc35b1, Spop, Tktl2, Trim23, Zfp644
24h to 36h	Actn3, Asnsd1, Capn3, Chchd7, Cnn2, Dgkd, Enpp3, Fam83b, Fbxw17, Foxp1, Gng3, Gstm4, Hmga1, Hmox1, Lnx2, Mrpl52, Nr2c1, Otub2, Peli1, Plce1, Pop5, Rsl24d1, Siva1, Spta1, Srgap1, Sumo3, Syne1, Zfp644
36h to 48h	B4galt1, Clip2, Cthrc1, Fam222b, Flnc, Flywch1, Glt1d1, Irs2, Mlxip, Mutyh, Phykpl, Ptpn13, Rbm47, Rragd, Tmem245, Trio, Wwp1

Code and Data availability

Abbreviations

Table 3: List of abbreviations

Abbreviation	Definition
ToD	Translation on demand
mESCs	Mouse embryonic stem cells
PSCs	Pluripotent stem cells
EpiSCs	Epiblast stem cells
hESCs	Human embryonic stem cells
LIF	Leukemia inhibitory factor
HSPs	Heat shock proteins
eIF4E	Eukaryotic translation initiation factor 4E
MNKs	MAPK-Interacting Kinases
RBPs	RNA-binding proteins
RNPs	Ribonucleoproteins
miRNAs	mircoRNAs
UTR	Untranslated region
LFC	Log Fold Change
GPR	Gaussian Process Regression
LFC-Ratio	Protein-to-RNA-LFC-Ratio
GO	Gene ontology
BP	Biological processes
MF	Molecular function