



Transcriptomic Characterization of the Human Cell Cycle in Individual Unsynchronized Cells

Joakim Karlsson¹, Thomas Kroneis^{2,3}, Emma Jonasson²,
Erik Larsson^{1,†} and Anders Ståhlberg^{2,4}

1 - Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-40530 Gothenburg, Sweden

2 - Department of Pathology and Genetics, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-40530 Gothenburg, Sweden

3 - Institute of Cell Biology, Histology and Embryology, Medical University of Graz, Harrachgasse 21, 8010 Graz, Austria

4 - Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, SE-40530 Gothenburg, Sweden

Correspondence to Erik Larsson and Anders Ståhlberg: E. Larsson, Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-40530 Gothenburg, Sweden.

A. Ståhlberg, Department of Pathology and Genetics, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-40530 Gothenburg, Sweden.

erik.larsson@gu.se; anders.stahlberg@gu.se

<https://doi.org/10.1016/j.jmb.2017.10.011>

Edited by Dylan Taatjes

Abstract

The highly fine-tuned dynamics of cell cycle gene expression have been intensely studied for several decades. However, some previous observations may be difficult to fully decouple from artifacts induced by traditional cell synchronization procedures. In addition, bulk cell measurements may have disguised intricate details. Here, we address this by sorting and transcriptomic sequencing of single cells progressing through the cell cycle without prior synchronization. Genes and pathways with known cell cycle roles are confirmed, associated regulatory sequence motifs are determined, and we also establish ties between other biological processes and the unsynchronized cell cycle. Importantly, we find the G1 phase to be surprisingly heterogeneous, with transcriptionally distinct early and late time points. We additionally note that mRNAs accumulate to reach maximum total levels at mitosis and find that stable transcripts show reduced cell-to-cell variability, consistent with the transcriptional burst model of gene expression. Our study provides the first detailed transcriptional profiling of an unsynchronized human cell cycle.

© 2017 Published by Elsevier Ltd.

Introduction

The cell cycle demands that numerous processes are executed with high precision at exact time points. In cancer, these highly coordinated events are commonly dysregulated, resulting in uncontrolled cell proliferation and genomic instability [1]. This has motivated studies of global gene expression changes throughout the cell cycle, which have provided mechanistic insights into cell division and growth. A widely cited resource of periodic human cell cycle gene expression was compiled over a decade ago [2], based on the HeLa cell line. Similar experiments have also been performed in *Saccharomyces cerevisiae* [3,4]. While these studies have been highly informative, they have been limited by

the requirement for cells to be synchronized in order to ensure that they are representative of the cell cycle phase of interest, and have all been performed on bulk cell measurements. Cell synchronization is a potentially disruptive procedure involving, for example, treatment with thymidine, serum starvation, or centrifugal elutriation [5]. These methods are prone to artifacts [6] and the achieved synchrony may also be limited [5,7,8]. Detailed studies of cell cycle-associated gene expression patterns using methods that do not perturb cellular states are thus desired.

Recent years have seen a rapid development of methods for transcriptomic analysis at the level of single cells [9]. These methods are now being applied to a diverse range of topics, including tissue cell-type

composition [10], embryonic development [11], and lung epithelial cell differentiation [12]. This also offers an opportunity to study transcriptional changes at high resolution as single cells transition through the cell cycle, importantly without the need for synchronized cell populations. To date, however, gene expression tied to this crucial process has mostly been examined peripherally in studies primarily concerned with other biological phenomena [13–16]. For example, a recent transcriptomic study of cell-to-cell variability in mouse hematopoietic stem cells found that cell cycle effects were dominant and also proposed that old cells traverse through G1 faster than young cells [15].

Several methods have also been developed for the purpose of identifying cell cycle-correlated gene expression in single-cell data [17–19]. Buettner *et al.* [17] developed a method to regress out the effect of the cell cycle on gene expression measurements, facilitating the study of other biological processes. Scialdone *et al.* [19] built a computational method to predict the cell cycle states G1, S, or G2/M from single-cell transcriptome data. Leng *et al.* [18] developed an algorithm to identify oscillatory genes in single-cell data. While these studies all agree that cell cycle activity tends to be the dominant driver of gene expression variability in single-cell experiments, and that it is important to be able to identify components of expression attributable to the cell cycle, neither has provided a detailed account of the individual genes regulated in such a fashion.

Thus, so far, no systematic effort has been carried out to update the catalog of periodic gene expression in the human cell cycle established using synchronization techniques more than a decade ago, at single-cell resolution. The recent technological advances still hold the promise of providing new important insights into the complex nature of cell cycle expression dynamics. Here, we applied single-cell transcriptomics using a protocol that enables amplification and sequencing of complete transcripts [20], to investigate how gene expression patterns change during the course of the cell cycle in human cells. RNA sequencing was performed on 96 human myxoid liposarcoma cells that were first sorted according to cell cycle phase using flow cytometry. Time points were further resolved within each phase by the introduction of a pseudo-time metric. We then proceeded to examine gene expression dynamics throughout the cell cycle, taking into account both coding transcripts and long non-coding RNAs (lncRNAs).

Results

Transcriptome sequencing of G1, S, and G2/M phase cells

To identify cell cycle-regulated transcripts, we collected and sequenced individual cells from

different cell cycle phases (Fig. 1a). A human myxoid liposarcoma cell line (MLS 1765-92) was used, which has functional TP53 signaling and lacks other common cancer driver mutations [21]. We reasoned that these cells, despite being tumor derived, may offer a complementary perspective to previous studies performed on HeLa cells, which harbor major genomic rearrangements and accumulated mutations [22]. Other alternatives include primary cultures or embryonic stem cell lines, but this could potentially introduce additional gene expression heterogeneity owing to cells of different origins or differentiation states. Living cells were stained with a DNA binding dye that allowed cells with different DNA content to be collected without prior synchronization, which avoids biases from cell cycle arrest procedures [6]. We collected 96 cells from the G1, S, and G2/M phases (32 cells each; G2 phase cannot be separated from the M phase based on DNA content). The Smart-seq2 protocol [20] was applied for complete polyadenylated transcript profiling, using ERCC spike-in controls to allow determination of absolute transcript levels [23]. Two G1 cells were removed due to low RNA content for a final total of 94. Sequence reads were then aligned to the human hg19 reference genome, and gene expression was quantified relative to the GENCODE v17 annotation.

mRNAs accumulate during cell cycle progression

To study the effect of cell cycle phase on total polyadenylated transcript levels, we normalized the total number of gene-mapped read counts in each sample to the ERCC spike-in controls (see Materials and Methods). A comparison between G1, S, and G2/M revealed a gradual global transcriptome increase throughout the cell cycle (Fig. 1b). A similar pattern was observed on preamplified cDNA yields analyzed prior to sequencing library preparation (Fig. 1c). Thus, mRNAs accumulate to reach their peak total levels at mitosis.

A pseudo-timeline resolves expression dynamics within phases

To resolve dynamics of transcriptomic changes within each phase, we defined a pseudo-timeline, where cells were ordered based on similarity in gene expression with respect to the two neighboring cell cycle phases (Materials and Methods). Comparisons with known expression patterns of B-type cyclins, which peak during mitosis and maintain high expression in early G1, as well as CCNE2, elevated in late G1, and CCNA2, which reaches maximal levels in late G2/M [24], supported the validity of the pseudo-timeline (Fig. S1). The timeline also corresponded well with the sample relationships determined by

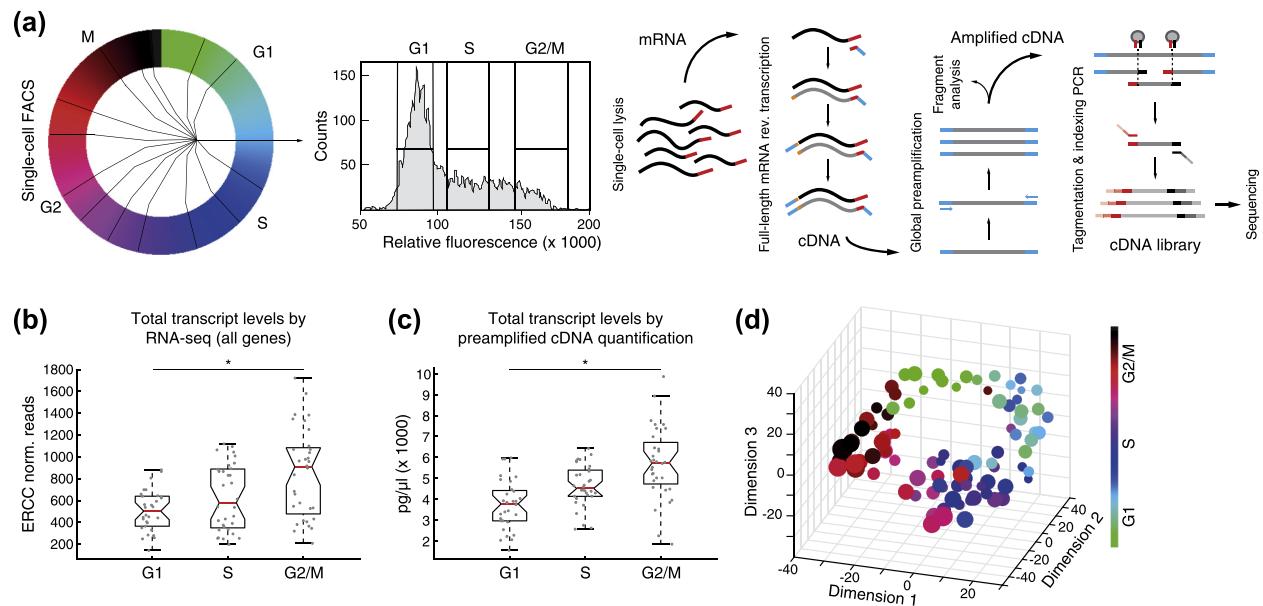


Fig. 1. Experimental workflow and observations on total transcript levels. (a) Single-cell library preparation using the Smart-seq2 protocol. Thirty-two cells were collected from G1, S, and G2/M, respectively, by fluorescence-activated cell sorting. Polyadenylated RNA molecules of lysed cells were full-length reverse transcribed, globally preamplified, fragmented, indexed, and sequenced. (b) Total poly(A) RNA transcript levels estimated by sequencing, as summed ERCC normalized reads in each cell. (c) Total poly(A) RNA levels determined by preamplified cDNA library quantification. For panels b–c, horizontal red lines indicate median; * indicates significant differences ($p < 0.05$, Kruskal–Wallis test). (d) t-SNE on coding genes expressed in >15 cells, based on \log_2 RPM. Colors represent pseudo-time derived cell cycle stages. Point sizes are proportional to total transcript levels. See also Fig. S1 and Table S6.

t-distributed stochastic neighbor embedding (t-SNE) [25], a dimensionality reduction technique reminiscent of principal component analysis (Fig. 1d).

Patterns of differential cell cycle expression

To discover genes with cell cycle-dependent expression, we tested for differences between phases using rank statistics (Materials and Methods). Tests were applied to all pairwise phase comparisons and resulted in a set of 472 genes, including 11 lncRNAs, at a false discovery rate (FDR) of <0.05 (Table S1, Fig. S2). Many genes with known cell cycle functions were confirmed, such as *CDK1*, *MCM2*, and *CCNF*. However, 269 genes (57%) were not annotated to the cell cycle in Gene Ontology (GO) [26] or Reactome [27], and did not have any periodic or cell cycle phenotype registered in Cyclebase 3.0 [24] (Table S2). Notably, 51 of these genes had homologs previously identified as cell cycle associated in a mouse hematopoietic stem cell line, supporting the validity of our findings [15]. To better understand the expression patterns of these genes, *k*-means clustering was performed, in which genes are assigned to one of *k* clusters in a way that maximizes the similarity to the respective cluster centroids [28]. Six clusters were found to provide a good description of the core trends (Fig. 2a). These co-expression groups were then

profiled further with respect to overrepresented GO terms, Reactome pathways and regulatory motifs, using MSigDB [30].

The first of the six co-expression groups (“Group 1,” Fig. 2a) described genes with low levels in G1, a peak in early S, and which then decreased until a second peak occurred in late G2/M. These genes mainly represented chromosome organization and DNA replication, but also transcriptional processes, exemplified by the terms “RNA splicing,” and “G1/S specific transcription” (detailed results in Table S3, summary terms in Fig. 2b). Other processes expressed according to this pattern were involved in protein localization and cellular stress responses. The latter may be indicative of either external stimuli pertaining to cell handling or potentially DNA replication stress, since some of these genes were also annotated to the GO terms “DNA metabolism” and “chromosome organization” (Table S3).

“Group 2” described genes whose expression progressively decreased from peak levels in early G1 to low levels in late G2/M. The dominating processes were related to translation, ribosome biogenesis, and post-translational control (Fig. 2a–b, Table S3). Out of 55 genes in this group, 10 encoded ribosomal proteins. RNA catabolic pathways were also highly represented, which, given elevated expression in G1, may act to degrade mRNAs that have accumulated during the previous cell cycle.

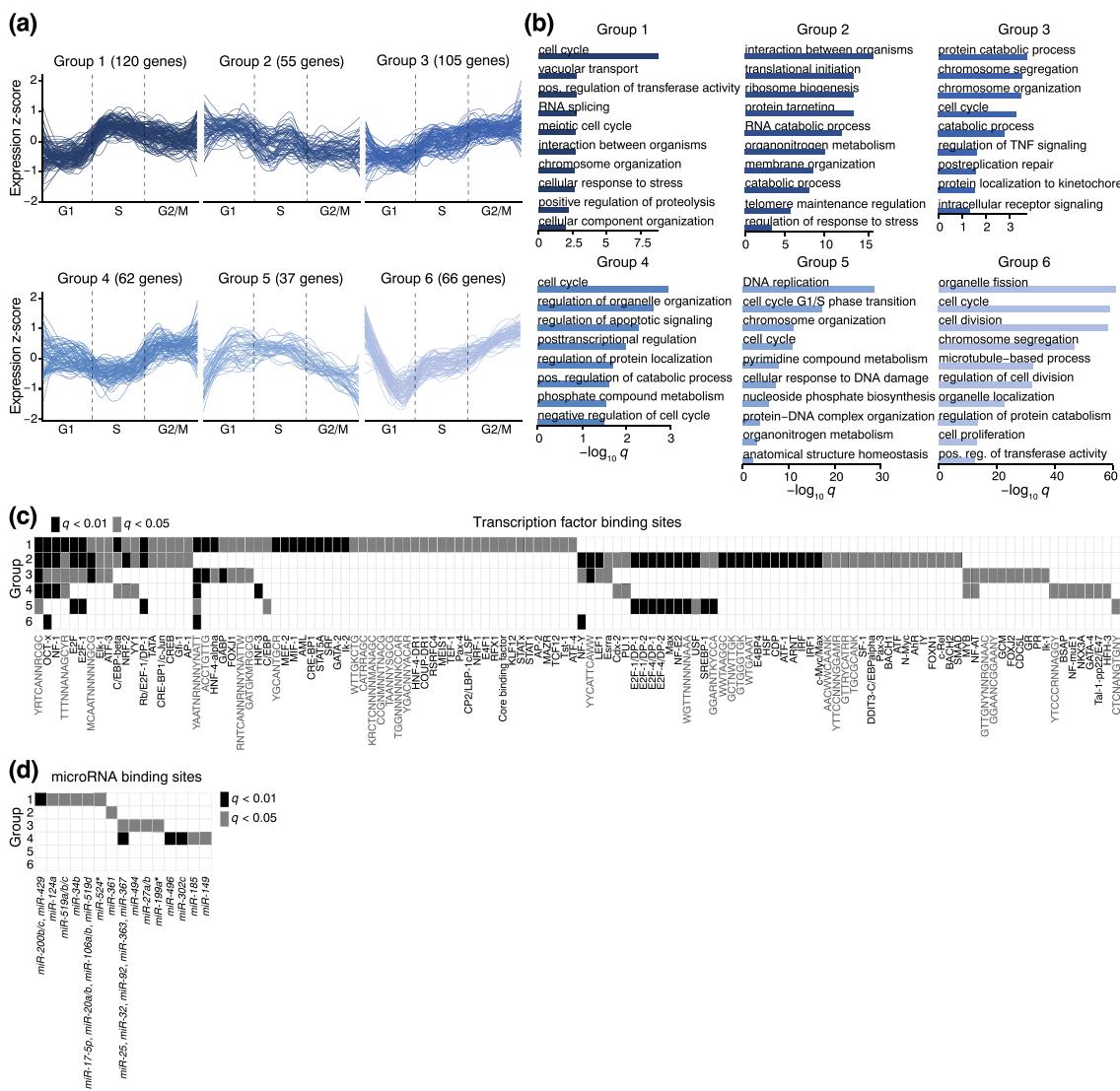


Fig. 2. Co-regulated cell cycle expression groups. (a) Expression profiles (cells ordered according to cell cycle pseudotime) of differentially expressed genes grouped with *k*-means clustering ($k = 6$), smoothed with LOESS regression (*R* function “*stat_smooth*,” *span* = 0.4) for increased clarity. (b) Enriched GO biological processes for the clusters in panel a. Similar terms were summarized using REVIGO [29] (default settings). Transcription factors (c) and miRNA binding motifs (d) associated with each cluster. Filled boxes indicate significant ($q < 0.05$, in gray, or $q < 0.01$, in black) enrichment of a given motif, based on MSigDB analyses. More detailed enrichment results, including Reactome pathways, are presented in Table S3. See also Fig. S2 and Table S1.

“Group 3” genes gradually increased in expression from G1 toward mitosis. These represented chromosome organization, post-replication repair, mitotic spindle assembly, and protein degradation (Fig. 2a–b, Table S3). Some of the protein degradation-associated genes, for instance, *BUB3*, *CDK1*, *UBC*, and *UBE2A*, were also annotated to GO terms such as “chromosome segregation” and “DNA repair” (Table S3). *UBC* and *UBE2A* encode ubiquitylation-associated proteins directly involved in protein destruction [31], while *BUB3* and *CDK1* are key cell cycle regulators that have more indirect ties to degradation

[32]. For instance, *BUB3* is implicated in the spindle checkpoint, which can prevent the anaphase-promoting complex from targeting other cell cycle regulators for degradation and thus allow further cell cycle progression [32].

“Group 4” represented genes with medium expression in G1, low levels in S, and peaks in early G2 and late M. Posttranscriptional and posttranslational control mechanisms, by means such as phosphorylation and ubiquitination-mediated proteasomal degradation, were among the most highly enriched processes (Fig. 2a–b, Table S3). Metabolic

pathways were also well represented, with mitochondrial inner membrane complexes and proton transporters enriched (Table S3).

"Group 5" genes followed a parabolic pattern that peaked in late G1 or early S, maintained high levels throughout S, and finally decreased toward mitosis. These genes, for instance, *CCNE2*, were involved in control mechanisms governing the G1/S transition, indicated by the Reactome pathways "cell cycle checkpoints" and "G1/S transition," but also nucleotide biosynthesis, DNA replication, and DNA repair (Fig. 2a–b, Table S3).

"Group 6" genes peaked during late mitosis/early G1. Enriched processes were associated with cytokinesis, exemplified by the terms "organelle fission," "microtubule based process," and "regulation of cell division" (Fig. 2a–b, Table S3). However, despite functions mainly associated with mitosis, these genes maintained high expression levels in early G1, consistent with observations in synchronized HeLa cells [2], and were also enriched for the Reactome pathway "G1/S transition," suggesting that these mitosis-associated genes may also have a role in G1 progression, rather than merely being passive byproducts of the previous cycle.

Transcription factor binding sites associated with cell cycle expression patterns

To investigate which transcription factors may regulate the identified expression patterns, enrichment analyses were performed using MSigDB transcription factor motif gene sets [30,33], which contain inferred target genes based on sequence patterns located in an interval of ± 2 kb from transcription start sites. Distinct combinations of enriched transcription factor motifs characterized the identified cell cycle-regulated groups (Fig. 2c, Table S3). These included sequence patterns for the E2F family proteins, known to regulate events at multiple stages of the cell cycle [34]. For instance, E2F1 activates, whereas E2F4 represses transcription in G1 [34]. E2F4 has additional activating roles [35] and also participates in cell cycle regulation in G2/M [36]. The established roles of these factors are consistent with the expression patterns of group 5 genes, which were elevated at the G1/S transition and declined throughout G2/M (Fig. 2a).

While most of the identified sequence elements were uniquely represented among genes of a single co-expression group, some were enriched in several. For instance, E2F, NF-1, Elk-1, ATF-3, YY1, C/EBP-beta, NRF-2, NF-Y, and octamer factor motifs (Fig. 2c, Table S3) were associated with three or more groups. Motifs of octamer factors, E2F, Elk-1, the ATF family, YY1, C/EBP, and NF-Y have also previously been associated with cell cycle gene expression in synchronized HeLa cells [37,38]. As the groups in which these broadly represented motifs are enriched have widely different profiles, this suggests that

they act combinatorially with other transcription factors. This has also been found to be the case for, among others, E2F, NF-Y, and CREB [37,38]. Notably, however, some of the most common motifs lacked known binding factors, the most prominent being the sequence patterns YRTCANNRCGC and YAATNRNN NYNATT, which were enriched in five out of six groups, respectively (Fig. 2c, Table S3). These may bind proteins with still unknown cell cycle involvement or uncharacterized complexes that include established factors.

MicroRNA binding sites associated with cell cycle expression patterns

Cell cycle control may also be mediated post-transcriptionally by microRNAs (miRNAs) binding to 3' UTRs of mRNAs. To discover such miRNAs, enrichment tests were performed using the MSigDB miRNA motif gene sets. Four of the co-expression groups, 1 to 4 (Fig. 2a), were significantly ($q < 0.05$) enriched for miRNA binding sites (Fig. 2d, Table S3). Several of the miRNAs that match these motifs are implicated in cell cycle control, including miR-20a/b, miR-25, miR-27a, miR-34b, miR-92, miR-106a/b, miR-124A, and miR-519 [39–45]. For instance, miR-27a may target genes involved in the G2/M checkpoint [39] and interfere with degradation of cyclin E [41]. miR-519, on the other hand, could reduce proliferation by inhibiting translation of the RNA-binding protein and translational regulator HuR [44], encoded by *ELAVL1*, which was also represented in the miR-519-associated group 1. Two other RNA binding proteins with miR-519 motifs, encoded by *SRSF2* and *PAPOLA*, were in the same group, suggesting that miR-519 may have additional influence over posttranscriptional mechanisms in a cell cycle context. Other enriched miRNAs lacked well-established cell cycle roles, although some evidence has been presented for miR-302c [46], miR-185 [47–49], miR-27b [50] and miR-494 [51,52].

Non-canonical cell cycle-associated genes

Some differentially expressed genes were not previously described as cell cycle regulated. These were in many cases part of processes not canonically considered in cell cycle progression (Table S4). Among those, a subset with high expression in G1 was associated with respiratory electron transport and the citric acid cycle, consisting of the genes *SDHB*, *SDHD*, *COX6C*, *COX7C*, *ATP5H* and *LDHB* (Fig. 3a). To investigate if these were cell cycle-associated beyond this MLS cell line, their expression levels were compared with public human embryonic stem cell (hESC) data [18]. Similar upregulation in G1 compared to S was observed also in hESCs, although the behavior in G2/M was

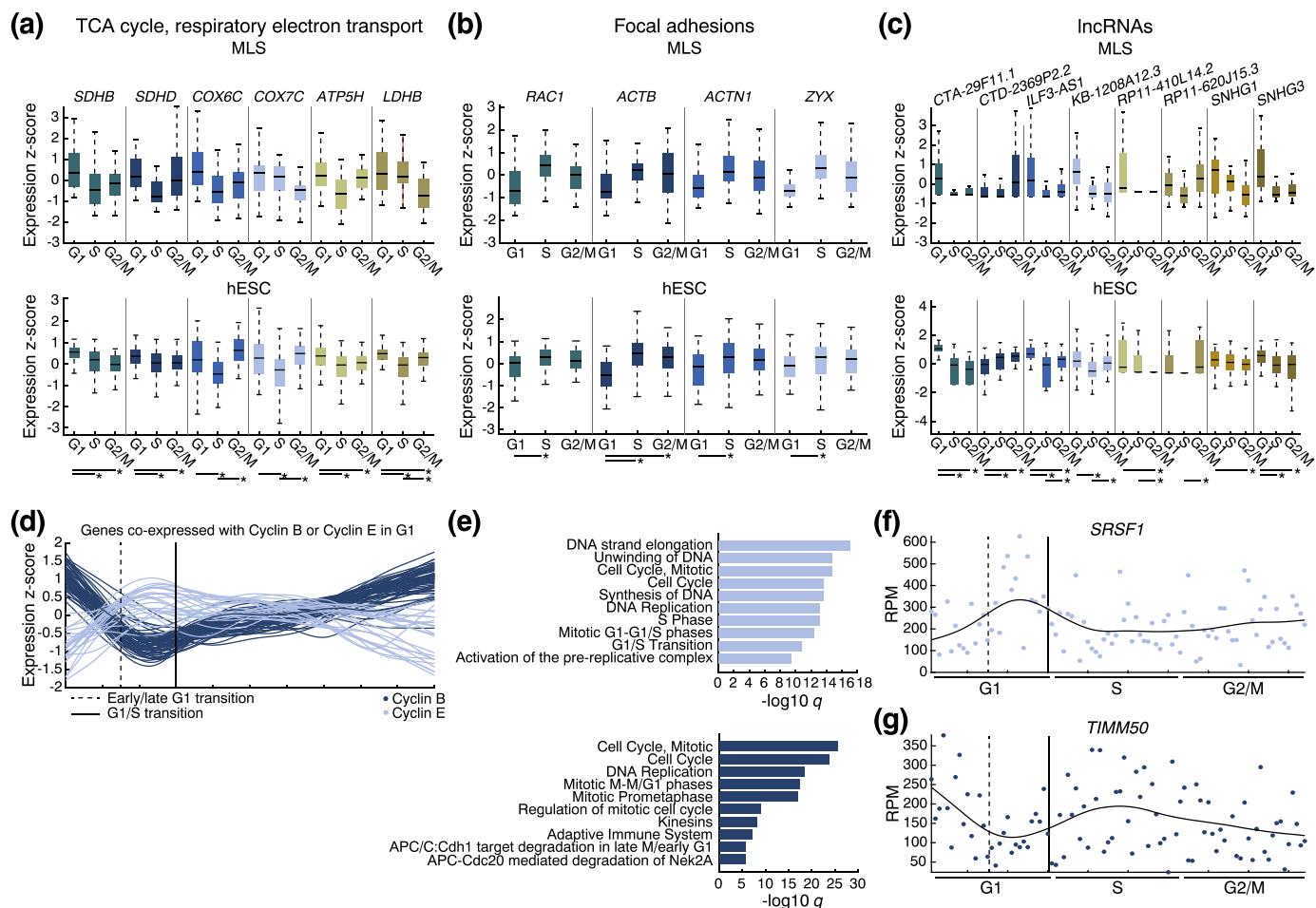


Fig. 3. Non-canonical cell cycle coupled genes and intra-G1 regulation. (a–c) Box plots of gene expression in different cell cycle phases for the indicated categories of genes, based on measurements in MLS cells and hESC [18]. * Significant differences ($p < 0.05$, two-tailed rank-sum test). (d) Trend fits of genes significantly co-expressed with either Cyclin B (blue, 59 genes) or Cyclin E (green, 31 genes) in G1. (e) Top 10 Reactome terms for genes co-expressed with Cyclin B (blue) or Cyclin E (green) in G1. (f–g) Expression profiles of *SRSF1* and *TIMM50*. See also Fig. S5 and Tables S4–S5 and S7.

in general different (Fig. 3a). Possibly, this upregulation of metabolic genes in G1 could be related to elevated energy requirements for protein and nucleotide biosynthesis as cells proceed to initiate DNA replication [53].

Among non-canonical processes highly expressed in S phase, on the other hand, were several associated with focal adhesions (Table S4). Genes related to this were *RAC1*, *ACTB*, *ACTN1*, and *ZYX* and, which also showed S phase upregulation compared to G1 in hESCs (Fig. 3b). Possibly, these illustrate crosstalk between the cell cycle and cell adhesion pathways, in which *RAC1* has a central role [54]. Non-canonical G2-associated genes were not enriched for any particular pathway, although they were associated with cell cycle-related transcription factors, such as SP1, MYB, NRF-1, Elk-1, GABP, and NF-Y (Table S4).

Genes that do not code for proteins have received less attention in a cell cycle context. Among the 11 lncRNAs we found to be significantly differentially expressed, 8 also displayed similar expression patterns in hESCs: *CTA-29F11.1*, *CTD-2369P2.2*, *ILF3-AS1*, *KB-1208A12.3*, *RP11-410L14.2*, *RP11-620J15.3*, *SNHG1*, and *SNHG3* (Fig. 3c), although with some minor differences, such as an increase in hESC G2/M expression for *KB-1208A12*. We conclude that the cell cycle affects, and is affected by, a broad range of cellular processes.

Early and late G1 characteristic genes

A substantial proportion of differentially expressed genes displayed large changes within G1 (Figs. 2a

and S2) and t-SNE analysis on all coding genes also showed that some G1 cells have transcriptomes similar to late M phase cells (Fig. 1d). We therefore concluded that two transcriptionally distinct G1 subpopulations exist, which we refer to as “early G1” and “late G1” cells, respectively, based on the pseudo-time ordering. We hypothesized that some genes would show cell cycle regulation only within G1, not detectable as differences in between phases. To investigate this in a manner unbiased by our visual classification of these subpopulations, we tested for significant co-expression with known marker genes that have patterns characteristic of exit from mitosis or entry into S phase. Specifically, within G1, *CCNB1* and *CCNB2* are known to be at their highest levels early in the phase, since they peak during mitosis, whereas *CCNE2* levels are higher at the S transition [24] (Fig. S1). Therefore, pairwise correlations between genes that met our thresholds for consideration (see Materials and Methods) and members of these two cyclin families were assessed. After correcting for multiple testing, 90 genes were discovered as significantly (FDR < 0.05) co-expressed with either (Fig. 3d, Table S5): 59 with cyclin B and 31 with cyclin E. Reassuringly, an overrepresentation analysis of Reactome pathways carried out using MSigDB [30] showed that genes coordinated with Cyclin B1 and B2 mRNAs were enriched for categories such as “Mitotic M-M/G1 phases,” while genes correlated with Cyclin E2 were characterized by terms such as “G1/S transition” (Fig. 3e).

Thirty of the detected genes only displayed significant cell cycle regulation within G1 and were not detected as differentially expressed in between

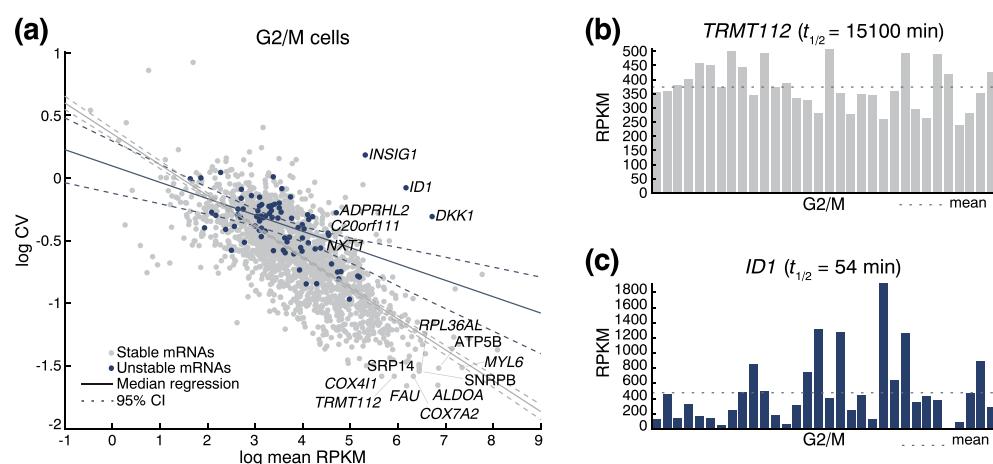


Fig. 4. The relationship between gene expression variability and mRNA decay rate. (a) Comparison of the coefficient of variation for stable (half-life greater of equal to 200 min) and unstable (half-life less than 200 min) mRNAs in G2/M cells (G1 and S results shown in Fig. S3). The coefficient of variation was calculated as the standard deviation divided by the mean of RPKM expression values for a given gene across all cells in the phase. Solid lines: median regression; dotted lines: 95% confidence intervals. Only genes expressed in over 90% of the cells were analyzed. Highlighted: highly expressed stable or unstable mRNAs. Expression of *TRMT112* (b; half-life: 15,100 min) and *ID1* (c; half-life: 54 min). See also Figs. S3–4.

phases. In addition, nine of these lacked previously documented cell cycle coupled expression. One of these was the splicing factor *SRSF1*, highly transcribed specifically in late G1, which has been proposed to regulate some cell cycle genes and to have oncogenic properties [55] (Fig. 3f). On the other hand, the expression of *TIMM50*, a mitochondrial inner membrane translocase, was reduced in late G1 (Fig. 3g).

Several transcription factor motifs were uniquely enriched in the early and late G1 sets, compared to those of other G1 regulated groups identified among genes differentially expressed in between phases (Tables S3 and S7). These included CREB, YRTCA NNRCGC, KRCTCNNNMANAGC, and CTGCAGY (early G1, compared to group 6) as well as NF-1 and GGAANCAGAANY (late G1, compared to group 5), which may thus potentially be involved in fine-tuning transcription within G1. In summary, cells in G1 are heterogeneous, and some genes show significant regulation only at distinct time points within this phase.

Stable mRNAs show reduced cell-to-cell variability

When studying cell population averages, the steady-state level of a given transcript is a function of its transcription and decay rates [56]. However, at the level of single cells, it is widely believed that transcription occurs in bursts [57,58]. One would therefore expect unstable (short-lived) transcripts to show not only lower levels, but also more cell-to-cell variability, as they will decay rapidly following a burst. On the other hand, stable transcripts should provide more buffering against this effect. This has, however, not been thoroughly investigated, and our single-cell data provided an opportunity to test this model.

We compared the variability of each gene across cells against mRNA decay rates previously determined in HeLa cells [59], defining short-lived mRNAs as those with half-lives less than 200 min. To avoid bias from cell cycle regulation, only data from one phase at a time were considered and both previously known cell cycle-regulated genes and those we detected as such here were excluded. Indeed, we observed that short-lived mRNAs displayed higher variability compared to more slowly degrading mRNAs, when expressed at a similar level (Fig. 4a shows G2/M phase; similar results were obtained in the other phases, Fig. S3A–B). This phenomenon is exemplified by *TRMT112* (half-life 15,100 min) and *ID1* (half-life 54 min), which have rather similar mean expression levels (477 and 373 RPKM, respectively) yet show marked differences in cell-to-cell variability (Fig. 4b–c, expression in G1 and S shown in Fig. S3C). To confirm that the HeLa mRNA decay rates were representative across cell lines, we also analyzed measurements from HepG2 cells and mouse fibroblasts. Consistent results were obtained when substituting for these decay rates

in the variability analysis (Fig. S3D–E). To investigate whether the steady-state mRNA levels of transcripts with different half-lives were affected by DNA staining and incubation time, we performed RT-qPCR measurements on representative genes with and without staining and incubation. No systematic bias was detected (Fig. S4). Our results support the transcriptional burst model of gene expression and suggest that low mRNA half-life may facilitate transient expression patterns across the cell cycle.

Discussion

Here, we describe how individual cell transcriptomes progress during the cell cycle, in cells that are undisturbed by synchronization protocols. We find that, as cells approach M phase, the total RNA amount increases, in agreement with a concurrent cell size increase [60]. While there is previous evidence that transcription is coordinated with cell size to maintain appropriate transcript concentrations, the exact relationship is complex and influenced by the cell cycle in a manner that varies between cell types and species [61–63]. Steady-state mRNA levels are influenced by multiple factors, including the frequency and duration of transcriptional bursts [57]. It has been observed that steady-state mRNA levels can lag transcription by as much as half a cell cycle for some genes [64]. An important determinant of these levels, besides the parameters of transcriptional pulsing, is mRNA decay. Here we find that more short-lived mRNAs show higher cell-to-cell variability in expression levels throughout the cell cycle. We speculate that processes requiring fast responses during the cell cycle may utilize rapid mRNA decay to allow better fine-tuning of temporal expression patterns.

By studying patterns of co-expression throughout the cell cycle, we gained insight into mechanisms that may be important at specific time points. An emergent theme is highly active translation in G1, as suggested by the elevated expression of genes involved in ribosome biogenesis and translational initiation, whereas we found that protein degradation genes are gradually upregulated toward G2/M. Proteomics studies have also noted decreased protein synthesis in G2/M [65,66]. Although it is possible that both processes cooperate to reduce total protein levels in relation to cell division, the degradation-associated genes may also target specific components and regulators of the mitotic machinery [67]. Genes involved in transcription and splicing, on the other hand, displayed a G1/S transition peak, suggesting increased transcriptional activity as the DNA replication program is initiated [34]. RNA catabolic processes, however, reached peak expression in G1, perhaps acting to degrade mRNAs that have accumulated throughout the previous cycle.

Our analysis of transcription factor binding motifs pointed to numerous factors that may have roles in coordinating these and other cell cycle processes. The E2F family is composed of core regulators participating in multiple and diverse patterns of gene co-expression. Others include NF-1, Elk-1, and the octamer motif binding proteins. These core regulators likely act combinatorially with other factors. For instance, many E2F binding sites also co-occur with NF-Y elements and YY1 motifs often co-occur with CREB [37]. The activity of the E2F family members is also modulated by co-factors such as DP-1 and Rb [68]. However, several motifs of unknown factors were also among the most prominently represented in our study, suggesting avenues for future research that may reveal additional key players in cell cycle control. Our results also supports the contribution of multiple miRNA families to cell cycle regulation, including several with less well established roles, such as miR-27b, miR-185, miR-302c, and mir-494.

By ordering cells on a pseudo-timeline, we could conclude that cells in early and late G1 phase showed marked differences in gene expression. A previously proposed model divides G1 into a “post-mitotic” and “pre-S” phase, separated by an internal G1 checkpoint, termed the restriction point [69]. This restriction point, in turn, has been suggested to represent a separate mechanism from that which then allows cells to enter S phase from G1-pre-S [69], although questions regarding the influence of serum starvation artifacts on these results have been raised [70]. However, our observations on transcriptionally distinct early and late G1 phases in an unsynchronized cell population, without additional major sources of biological variation, are consistent with such a separation. We additionally describe individual genes associated with this intra-G1 transition. Awareness of distinct G1 subpopulations may also be of relevance to other single-cell studies, where the removal of confounding elements attributable to cell cycle activity could be desired.

Moreover, we discovered a number of genes not traditionally considered cell cycle regulated, some of which were lncRNAs. Two of these, both with G1-specific expression, were small nucleolar RNA (snoRNA) host genes (*SNHG1* and *SNHG3*). SnoRNAs act as guides to mediate specific modifications on rRNAs as part of the formation of mature ribosomes [71]. This role is compatible with the theme of ribosome biogenesis-related gene expression in G1. *SNHG3*, as well as seven other lncRNAs, also had similar expression patterns in hESCs [18]. Other non-canonical cell cycle-regulated genes were involved in the citric acid cycle and the respiratory electron transport chain. These were highly expressed in G1, whereas genes involved in focal adhesions and actin cytoskeleton dynamics were upregulated in S phase. Previous studies have also found connections between metabolic gene expression and the cell cycle, although it may be difficult to exclude influence from

synchronization-induced stress responses [53,72]. Our study avoids synchronization artifacts. However, other factors of the experimental design may affect which cell cycle-regulated genes we detect. Although cell sorting is likely to have minimal impact [73], trypsinization, which dissociates cells to allow for single-cell sequencing, may influence cell adhesion and stress responsive genes [73]. The latter include ATF factors, which control both cell cycle and metabolic gene expression. Hence, although we observed similar G1/S-dependent regulation of these genes also in hESCs, there is a possibility for crosstalk between these pathways that may be influenced by external stresses. Further studies are needed to fully characterize the means by which cell cycle progression is dynamically tuned to the environment by transcription factors such as those of the ATF family and how the presence or absence of cell adhesion influences cell cycle regulation.

A prominent group of genes largely absent from our analysis were histones, known to be transcribed in a replication-dependent manner [74]. Most histone mRNAs are non-polyadenylated [75], and therefore, the Smart-seq2 protocol did not allow us to quantify these [20]. However, two histones, *H2AFX* and *HIST1H4C*, were detected as differentially expressed. This can be explained by the fact that *H2AFX* transcripts are bimorphic, that is, occasionally polyadenylated. Although *HIST1H4C* was found to be non-polyadenylated in HeLa and H9 hESCs [75], bimorphism is dependent on cell type and growth conditions [75,76], making it likely that also *HIST1H4C* is bimorphic.

Myxoid liposarcoma cells are likely to have some differences in cell cycle regulation to normal cells. This cell line is mainly driven by the FUS-DDIT3 fusion [21]. Here we detected enriched DDIT3 (a member of the C/EBP family) and C/EBP binding motifs among some cell cycle-regulated genes, suggesting means by which this fusion could have an influence. A group of genes potentially affected may be those involved in the unfolded protein response [77,78], which were highly expressed in G1 and present in a co-expression group enriched for DDIT3 motifs. However, this process could also be activated in response to the demands of ribosome biogenesis and protein production intimately coupled with G1 progression [78,79].

In conclusion, our study gives the first detailed picture of gene expression changes as individual cells progress through the human cell cycle while undisturbed by cell synchronization protocols. This allowed us to profile cell cycle-associated expression patterns of individual genes that were not artificially induced by such procedures. It also made it possible to describe important general characteristics, such as total RNA content as a function of cell cycle progression, an influence of mRNA decay on gene expression variability as well as cell-to-cell heterogeneity within phases.

Materials and Methods

Cell culture and single-cell sampling

A myxoid sarcoma cell line (MLS 1765-92) was cultured at 37 °C and 5% CO₂ in RPMI 1640 GlutaMax medium supplemented with 5% fetal bovine serum, 100 U/ml penicillin, and 100 µg/ml streptomycin. Cells were passaged with 0.25% trypsin containing 0.5 mM EDTA (all Thermo Fisher Scientific).

For single-cell sampling, 10⁶ trypsinized MLS 1765-92 cells/ml was resuspended in Hanks' balanced salt solution (Thermo Fisher Scientific) and stained for nuclear DNA using Vybrant DyeCycle Violet stain (Thermo Fisher Scientific) at 37 °C and a final concentration of 5 µM for 30 min. After staining, the cells were resuspended in Hanks' balanced salt solution and filtered via a 70-µm cell strainer (BD Biosciences) to remove cell aggregates. The cells were kept on ice until sorting.

Single cells were sorted according to cell cycle phases G1, S, and G2/M into 96-well plates (Thermo Fisher Scientific) containing 5 µl of 1 µg/µl BSA in 2.5% glycerol (Thermo Fisher Scientific) and 0.2% Triton X-100 (Sigma-Aldrich) in DNase/RNase-free water (Thermo Fisher Scientific) using a BD FACSAria II (Becton Dickinson) instrument and the FACSDiva software as described previously [80]. Collected cells were frozen on dry ice immediately after sorting and stored at -80 °C.

Single-cell transcriptome sequencing

Sequence-ready transcriptome libraries were generated with the Smart-seq2 protocol [20] with minor changes. In short, 1 µM adapter sequence-containing oligo-dT (5'-AAGCAGTGGTATCAACG CAGAGTACT₃₀VN-3', Sigma-Aldrich), 1 mM dNTP (Sigma-Aldrich), and ERCC spike-in controls [23] (corresponding to 1 µl of 1:2,500,000 diluted stock solution; Thermo Fisher Scientific) were incubated with direct lysed cells in a volume of 6.7 µl at 72 °C for 3 min. Reverse transcription (RT) was then performed in 15 µl by adding 1 × first-strand buffer [50 mM Tris-HCl (pH 8.3), 75 mM KCl, and 3 mM MgCl₂; Thermo Fisher Scientific], 1 M betaine (Sigma-Aldrich), 5 mM DTT, additional 10 mM MgCl₂ (both Thermo Fisher Scientific), 0.6 µM adapter sequence-containing template switching oligonucleotide (5'-AAGCAGTGG TATCAACGCAGAGTACATrGrG + G-3', where rG = riboguanosine and +G = locked nucleic acid modified guanosine, Eurogentec), 15 U RNaseOUT, and 150 U SuperScript II enzyme (both Thermo Fisher Scientific). RT was performed at 42 °C for 90 min and 70 °C for 15 min before being chilled to 4 °C. Final RT concentrations are shown. cDNA was stored at -20 °C until further use.

Preamplification of 7.5 µl cDNA was then performed in 50 µl, containing 1 × KAPA HiFi HotStart Ready Mix (KAPA Biosystems) and 60 nM adapter sequence-matching primer (5'-AAGCAGTGGTATCAACGCA-GAGT-3'; Sigma-Aldrich), as follows: denaturation at 98 °C for 3 min followed by 24 cycles of amplification (98 °C for 20 s, 67 °C for 15 s, and 72 °C for 6 min) and a final incubation at 72 °C for 5 min before being chilled to 4 °C. Preamplified cDNA was purified using the AMPure XP beads kit (Agencourt, Beckman Coulter), according to the Smart-seq2 protocol [20], with a sample to beads volume ratio of 0.8. Purified cDNA was eluted in 15 µl water and stored at -20 °C until further use. Concentration, fragment length distribution, and total transcriptome amount of preamplified libraries were analyzed by capillary gel electrophoresis using the High Sensitivity DNA Kit (Agilent) on a 2100 Bioanalyzer (Agilent).

Sequencing libraries were generated using the Nextera XT DNA Sample Preparation Kit and Nextera XT v2 Index Kit (both Illumina), following the manufacturer's recommendations with minor changes. In short, 0.1 ng of purified and preamplified cDNA was fragmented in 20 µl, containing 10 µl TD buffer and 5 µl ATM at 55 °C for 5 min. Fragmentation was stopped by adding 5 µl of NT buffer at room temperature incubating for another 5 min. Incorporation of index sequences was performed in 50 µl, containing 15 µl NPM PCR master mix solution, 5 µl i5 index primer, and 5 µl i7 index primer, as follows: heating to 72 °C for 3 min, denaturation at 95 °C for 30 s and 16 cycles of amplification (95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s), and a final extension step at 72 °C for 5 min before chilling to 10 °C. Indexed samples were purified using the AMPure XP beads, as described above, with a sample to beads volume ratio of 0.6. Mean fragment length and concentration of indexed single-cell libraries were assessed using High Sensitivity DNA Kit (Agilent) and dsDNA High Sensitivity Assay Kit (Qubit, Thermo Fisher Scientific), respectively. Single-cell libraries were diluted to 2 nM, pooled, and sequenced by the Genomics Core facility at the University of Gothenburg, Sweden, using Illumina NextSeq 500 to obtain 150-bp paired-end reads.

Data preprocessing

Reads were aligned to the hg19 reference, with ERCC spike-in sequences appended, using STAR version 2.4.1 [81], and splice junctions supplied from GENCODE [82] V17. Two cells with low RNA content were discarded. Gene expression was quantified using HTSeq [83] (options “-s no” and “-m intersection-strict”). RPM normalization (RPKM for some analyses) was preferred, since other methods may rely on assumptions that are violated in single-cell data [84]. Total transcript levels were estimated by the sum of raw gene reads divided by the sum of ERCC control reads. For comparisons with hESC data [18], reads obtained via

Gene Expression Omnibus entry PRJNA269761 were processed identically as above.

Differential and co-expression analysis

Differential expression analysis was performed using the Wilcoxon rank-sum test on RPM values of all genes that were either protein coding or lncRNAs and expressed in at least 10 samples. FDR correction was performed using Storey's procedure [85], with a cutoff of 0.05. The rank-sum test avoids the parametric assumptions of tools developed for bulk RNA-seq data, such as DESeq2 and edgeR, which neither account for dropout events due to cDNA amplification bias nor the stochastic nature of transcription in single cells.

To identify common patterns among genes that were differentially expressed between phases, k-means clustering was performed using the algorithm by Hartigan and Wong [28], using the R "kmeans" function with 1000 random starts, 500 max iterations and $k = 6$. Briefly, k-means aims to assign objects to the nearest "centroid" such that the within-cluster sum of squares is minimized. k determines the number of centroids, and thus clusters, that will be formed. The value of the parameter k was chosen by analyzing within-cluster sum of squares as a function of k , as well as by noting that the additional clusters obtained for higher values of k tended to be very similar to those obtained with $k = 6$. The clustering was performed based on expression values that had been normalized as $\log_2(\text{RPM} + 1)$ and then standardized such that each gene had mean 0 and standard deviation 1.

To determine early and late G1 characteristic genes, correlations with the representative markers *CCNB1*, *CCNB2*, and *CCNE2* were investigated. Given high similarities in expression of the B-type cyclins, their average was used as a less noisy substitute (*CCNE1* was only expressed in two cells and thus not considered for averaging with *CCNE2*). Only biologically variable genes, identified using functions in the scLVM R package [19] (default parameters), that were expressed in over 10 G1 cells were tested. FDRs of Pearson correlation p -values <0.05 were considered significant. Analyses of enriched categories and pathways among these genes were carried out with the MSigDB [30] online overrepresentation analysis tool, using gene sets of Reactome pathways [27], GO terms [26], transcription factor, and miRNA binding motifs [33]. Categories with q -values <0.05 were considered significant. For Fig. 2c, some transcription factor names were updated to current standards and some motifs implying the same transcription factor and were merged. The original motif names are presented in Table S3.

Identification of known cell cycle-regulated genes

To identify genes with previously known cell cycle-related function or expression, a comprehensive list

was compiled from the GO categories "cell cycle" (GO:0007049), "mitotic cell cycle" (GO:0000278), "cytokinesis" (GO:0000910), "regulation of cytokinesis" (GO:0032465), "regulation of cell separation after cytokinesis" (GO:0010590), "regulation of cell cycle" (GO:0051726), "cell separation after cytokinesis" (GO:0000920), "chromosome segregation" (GO:0007059), the Reactome category "cell cycle," and all genes included in the Cyclebase 3.0 downloadable files "human_periodic.tsv," "human_experiments_ptms.tsv," and "human_knowledge_ptms.tsv." Additional genes were added after automated NCBI Gene searches for summaries containing the phrase "cell cycle." The final list of 3235 genes is given in Table S2. To compare the genes identified as cell cycle regulated in this work with their corresponding homologs identified in a previous mouse hematopoietic stem cell population [15], gene identifiers were mapped between the two species based on information in the Mouse Genome Informatics homology database, specifically the downloadable table "HOM_MouseHumanSequence.rpt" [86].

Cell cycle pseudo-time derivation

To create a cyclic pseudo-timeline, each cell was assigned a sort order within its FACS determined phase. This order was based on relative similarity to the neighboring cell cycle phases. Given cells i , j , and k , which are members of G1, S, and G2 respectively, the relative similarity was estimated as the ratio $r_i = \max(\text{cor}(i,j))/\max(\text{cor}(i,k))$, where cor is the Pearson correlation between the gene expression values, as standardized $\log_2(\text{RPM} + 1)$, of each cell. For this purpose, a set of highly variable and expressed genes was used (heuristic thresholds: standard deviation above 1.5 and over 256 reads).

Another method for pseudo-time calculation was also evaluated: inferring the order entirely from the transcriptome based on a curve fit to the locations of the cells in two-dimensional t-SNE space [87] (Fig. S5). For this calculation, all coding genes expressed in more than 15 cells were included. To fit the curve, a k -segment-based algorithm implemented in MATLAB was used [88]. Overall, the order derived from this procedure agreed with well with our original method, as well as the FACS-defined labels. However, the order of cells in S phase appeared to be reversed in comparison to both our original method, as well as the FACS-defined labels and prior information about marker gene expression: six G2/M cells (according to the FACS labels) were assigned to the beginning of S and five G1 cells assigned to the end of S or beginning of G2. This resulted in an apparent second peak of *CCNE2* expression in late S, which does not agree with reference profiles of *CCNE2* expression from Cyclebase (Figs. S1A and S5B). Based on these factors, we decided to rely on the original correlation-based pseudo-time metric for the

main analyses. Indices of all cells based on both metrics are available in Table S6.

Variability and mRNA decay rate comparison

Human mRNA decay rates were obtained from publicly available data [59] and processed as previously described [89]. The variability of long (half-life greater or equal to 200 min) and short lived (half-life less than 200 min) transcripts was compared using median regression on the log coefficient of variation as a function of log mean RPKM, considering all non-cell cycle-regulated genes expressed in >90% of the cells that had half-life values assigned. Non-cell cycle-regulated genes were defined as those that were neither among the known cell cycle-regulated genes in Table S2 nor detected as regulated in our analyses. The number of genes passing these criteria for each phase was 1209 in G1, 1753 in S, and 1869 in G2/M. The coefficient of variation was calculated as the standard deviation divided by the mean of RPKM expression values for a given gene across all cells in the respective phases. Confidence intervals were estimated with 1000 bootstrap iterations, using an algorithm available at [http://mathworks.com/matlabcentral/fileexchange/32115-quantreg-x-y-tau-order-nboot-\(version 1.3\)](http://mathworks.com/matlabcentral/fileexchange/32115-quantreg-x-y-tau-order-nboot-(version 1.3)). To compare the variability of stable and unstable mRNAs based on HepG2 and mouse fibroblast half-lives, data were extracted from supplemental material of Yang *et al.* [90] and obtained from Gene Expression Omnibus entry GSE5324 [91], respectively. As for the HeLa data, non-cell cycle-regulated genes expressed in >90% of the cells and which had half-life values assigned were considered.

Assessment of the influence of DNA staining on selected genes

In order to assess the effect of DNA staining on the expression of mRNAs with different ranges of half-lives and corresponding to different biological processes, 10^6 trypsinized MLS 1765-92 cells/ml were resuspended in Hanks' balanced salt solution (Thermo Fisher Scientific) and stained for nuclear DNA using Vybrant DyeCycle Violet stain (Thermo Fisher Scientific), or water for control cells, at 37 °C and a final concentration of 5 µM for 30 min. Three parallel replicates were used. Cells were centrifuged and resuspended in 700 µl QIAzol lysis buffer (Qiagen). In addition, three replicates of cells lysed directly after trypsinization and washed with PBS were added as controls. RNA was extracted using miRNeasy mini kit (Qiagen) according to the manufacturer's instructions, including DNase treatment. Samples were eluted in 30 µl RNase-free water. RT was performed using GrandScript cDNA synthesis kit (TATAA Biocenter) according to the manufacturer's instructions. Briefly, RT was run in 10 µl containing 300 ng RNA, 1× TATAA GrandScript RT reaction mix, and 1× TATAA

GrandScript RT enzyme at 22 °C for 5 min, 42 °C for 30 min, and 85 °C for 5 min before being chilled to 4 °C. qPCR was performed using SYBR GrandMaster Mix (TATAA Biocenter) according to the manufacturer's instructions. Briefly, qPCR was run in 6 µl containing 2 µl cDNA, diluted 1:6, 1× TATAA SYBR GrandMaster Mix, and 400 nM of each forward and reverse primers. qPCR was run on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad) at 95 °C for 2 min followed by 40 cycles of incubation at 95 °C for 5 s, 60 °C for 20 s, and 70 °C for 20 s and melting curve analysis. Cq-values were determined by the second derivative maximum method using the CFX Manager Software version 3.1 (Bio-Rad). Data were analyzed using the GenEx software v6 (MultiD). Values were normalized to three reference genes (*CCNG1*, *RPL7*, and *CDT1*), chosen with the GenEx tool NormFinder, and \log_2 -transformed. Primer sequences are listed in Table S8.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2017.10.011>.

Accession numbers

RNA-seq data have been deposited in the ArrayExpress database at EMBL-EBI under accession number E-MTAB-6142.

Acknowledgments

This work was supported by the Assar Gabrielssons Research Foundation, BioCARE National Strategic Research Program, Johan Jansson Foundation for Cancer Research, Knut & Alice Wallenberg Foundation, Sahlgrenska Academy (ALF) at the University of Gothenburg, Swedish Cancer Society, Swedish Childhood Cancer Foundation, Swedish Foundation for Strategic Research, Swedish Research Council, Swedish Society for Medicine, VINNOVA, Wilhelm and Martina Lundgren Foundation for Scientific Research, Wallenberg Centre for Molecular and Translational Medicine, and Åke Wiberg's Foundation. Sequencing was performed by the Genomics Core Facility at the University of Gothenburg, Sweden. The authors declare no competing financial interests.

Author Contributions: Conceptualization: E.L. and A.S.; methodology: T.K., J.K., and A.S.; investigation: T.K.; formal analysis: J.K., E.J., and E.L.; validation: T.K. and E.J.; visualization: J.K. and T.K.; writing of the original draft: J.K. and T.K.; writing, review, and editing: J.K., T.K., E.J., E.L., and A.S.

Received 15 June 2017;

Received in revised form 5 October 2017;

Accepted 7 October 2017

Available online xxxx

Keywords:

cell cycle;
unsynchronized;
gene expression dynamics;
transcription factors;
single-cell analysis

†Lead contact.

Abbreviations used:

hESC, human embryonic stem cell; GO, Gene Ontology; lncRNA, long non-coding RNA; miRNA, microRNA; MLS, myxoid liposarcoma; t-SNE, t-distributed stochastic neighbor embedding.

References

- [1] M. Malumbres, M. Barbacid, Cell cycle, CDKs and cancer: a changing paradigm, *Nat. Rev. Cancer* 9 (2009) 153–166, <https://doi.org/10.1038/nrc2602>.
- [2] M.L. Whitfield, G. Sherlock, A.J. Saldanha, J.I. Murray, C.A. Ball, K.E. Alexander, J.C. Matese, C.M. Perou, M.M. Hurt, P.O. Brown, D. Botstein, Identification of genes periodically expressed in the human cell cycle and their expression in tumors, *Mol. Biol. Cell* 13 (2002) 1977–2000, <https://doi.org/10.1091/mbc.02-02-0030>.
- [3] R.J. Cho, M.J. Campbell, E.A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T.G. Wolfsberg, A.E. Gabrielian, D. Landsman, D.J. Lockhart, R.W. Davis, A genome-wide transcriptional analysis of the mitotic cell cycle, *Mol. Cell* 2 (1998) 65–73, [https://doi.org/10.1016/S1097-2765\(00\)80114-8](https://doi.org/10.1016/S1097-2765(00)80114-8).
- [4] M.V. Granovskiaia, L.J. Jensen, M.E. Ritchie, J. Toedling, Y. Ning, P. Bork, W. Huber, L.M. Steinmetz, High-resolution transcription atlas of the mitotic cell cycle in budding yeast, *Genome Biol.* 11 (2010) R24, <https://doi.org/10.1186/gb-2010-11-3-r24>.
- [5] P.K. Davis, A. Ho, S.F. Dowdy, Biological methods for cell-cycle synchronization of mammalian cells, *BioTechniques* 30 (2001) 1322–1331.
- [6] S. Cooper, K. Shedden, Microarray analysis of gene expression during the cell cycle, *Cell Chromosome* 2 (2003) 1, <https://doi.org/10.1186/1475-9268-2-1>.
- [7] M.H. Fox, R. a Read, J.S. Bedford, Comparison of synchronized Chinese hamster ovary cells obtained by mitotic shake-off, hydroxyurea, aphidicolin, or methotrexate, *Cytometry* 8 (1987) 315–320, <https://doi.org/10.1002/cyto.990080312>.
- [8] S. Cooper, Rethinking synchronization of mammalian cells for cell cycle analysis, *Cell. Mol. Life Sci.* 60 (2003) 1099–1106, <https://doi.org/10.1007/s00018-003-2253-2>.
- [9] A.A. Kolodziejczyk, J.K. Kim, V. Svensson, J.C. Marioni, S.A. Teichmann, The technology and biology of single-cell RNA sequencing, *Mol. Cell* 58 (2015) 610–620, <https://doi.org/10.1016/j.molcel.2015.04.005>.
- [10] D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, I. Amit, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types, *Science* 343 (2014) 776–779, <https://doi.org/10.1126/science.1247651>.
- [11] L. Yan, M. Yang, H. Guo, L. Yang, J. Wu, R.R. Li, P. Liu, Y. Lian, X. Zheng, J. Yan, J. Huang, M. Li, X. Wu, L. Wen, K. Lao, R.R. Li, J. Qiao, F. Tang, Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells, *Nat. Struct. Mol. Biol.* 20 (9) (2013) 1131, <https://doi.org/10.1038/nsmb.2660>.
- [12] B. Treutlein, D.G. Brownfield, A.R. Wu, N.F. Neff, G.L. Mantalas, F.H. Espinoza, T.J. Desai, M.A. Krasnow, S.R. Quake, Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq, *Nature* 509 (2014) 371–375, <https://doi.org/10.1038/nature13173>.
- [13] C. Trapnell, D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, L. Pachter, Differential analysis of gene regulation at transcript resolution with RNA-seq, *Nat. Biotechnol.* 31 (2013) 46–53, <https://doi.org/10.1038/nbt.2450>.
- [14] G.K. Marinov, B.A. Williams, K. McCue, G.P. Schroth, J. Gertz, R.M. Myers, B.J. Wold, From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing, *Genome Res.* 24 (2014) 496–510, <https://doi.org/10.1101/gr.161034.113>.
- [15] M.S. Kowalczyk, I. Tirosh, D. Heckl, T.N. Rao, A. Dixit, B.J. Haas, R.K. Schneider, A.J. Wagers, B.L. Ebert, A. Regev, Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells, *Genome Res.* 25 (2015) 1860–1872, <https://doi.org/10.1101/gr.192237.115>.
- [16] B. Reinius, J.E. Mold, D. Ramsköld, Q. Deng, P. Johnsson, J. Michaëlsson, J. Frisén, R. Sandberg, Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq, *Nat. Genet.* 48 (2016) 1430–1435, <https://doi.org/10.1038/ng.3678>.
- [17] F. Buettner, K.N. Natarajan, F.P. Casale, V. Proserpio, A. Scialdone, F.J. Theis, S.A. Teichmann, J.C. Marioni, O. Stegle, Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells, *Nat. Biotechnol.* 33 (2015) 155–160, <https://doi.org/10.1038/nbt.3102>.
- [18] N. Leng, L. Chu, C. Barry, Y. Li, J. Choi, X. Li, P. Jiang, R.M. Stewart, J.A. Thomson, C. Kendziora, Oscope identifies oscillatory genes in unsynchronized single-cell RNA-seq experiments, *Nat. Methods* 12 (2015) 947–950, <https://doi.org/10.1038/nmeth.3549>.
- [19] A. Scialdone, K.N. Natarajan, L.R. Saraiva, V. Proserpio, S.A. Teichmann, O. Stegle, J.C. Marioni, F. Buettner, Computational assignment of cell-cycle stage from single-cell transcriptome data, *Methods* 85 (2015) 54–61, <https://doi.org/10.1016/j.ymeth.2015.06.021>.
- [20] S. Picelli, Å.K. Björklund, O.R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells, *Nat. Methods* 10 (2013) 1096–1098, <https://doi.org/10.1038/nmeth.2639>.
- [21] A. Ståhlberg, C. Kåbjörn Gustafsson, K. Engström, C. Thomsen, S. Dolatabadi, E. Jonasson, C.Y. Li, D. Ruff, S.M. Chen, P. Åman, Normal and functional TP53 in genetically stable myxoid/round cell liposarcoma, *PLoS One* 9 (2014), e113110. <https://doi.org/10.1371/journal.pone.0113110>.
- [22] J.J.M. Landry, P.T. Pyl, T. Rausch, T. Zichner, M.M. Tekkedil, A.M. Stütz, A. Jauch, R.S. Aiyar, G. Pau, N. Delhomme, J. Gagneur, J.O. Korbel, W. Huber, L.M. Steinmetz, The genomic and transcriptomic landscape of a HeLa cell line, G3 (Bethesda). 3 (2013) 1213–1224, <https://doi.org/10.1534/g3.113.005777>.
- [23] L. Jiang, F. Schlesinger, C.A. Davis, Y. Zhang, R. Li, M. Salit, T.R. Gingras, B. Oliver, Synthetic spike-in standards for RNA-seq experiments, *Genome Res.* 21 (2011) 1543–1551, <https://doi.org/10.1101/gr.121095.111>.

- [24] A. Santos, R. Wernersson, L.J. Jensen, Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes, *Nucleic Acids Res.* 43 (2015) D1140–4, <https://doi.org/10.1093/nar/gku1092>.
- [25] L. Van Der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2605, <https://doi.org/10.1007/s10479-011-0841-3>.
- [26] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.* 25 (2000) 25–29, <https://doi.org/10.1038/75556>.
- [27] D. Croft, G. O'Kelly, G. Wu, R. Haw, M. Gillespie, L. Matthews, M. Caudy, P. Garapati, G. Gopinath, B. Jassal, S. Jupe, I. Kalatskaya, S. Mahajan, B. May, N. Ndegwa, E. Schmidt, V. Shamovsky, C. Yung, E. Birney, H. Hermjakob, P. D'Eustachio, L. Stein, Reactome: a database of reactions, pathways and biological processes, *Nucleic Acids Res.* 39 (2011) D691–7, <https://doi.org/10.1093/nar/gkq1018>.
- [28] J.A. Hartigan, M.A. Wong, Algorithm AS 136: a K-means clustering algorithm, *J. R. Stat. Soc. Ser. C* 28 (1979) 100–108, <https://doi.org/10.2307/2346830>.
- [29] F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of Gene Ontology terms, *PLoS One* 6 (2011), e21800. <https://doi.org/10.1371/journal.pone.0021800>.
- [30] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 15545–15550, <https://doi.org/10.1073/pnas.0506580102>.
- [31] D. Komander, M. Rape, The ubiquitin code, *Annu. Rev. Biochem.* 81 (2012) 203–229, <https://doi.org/10.1146/annurev-biochem-060310-170328>.
- [32] Z. Tang, H. Shu, D. Oncel, S. Chen, H. Yu, Phosphorylation of Cdc20 by Bub1 provides a catalytic mechanism for APC/C inhibition by the spindle checkpoint, *Mol. Cell* 16 (2004) 387–397, <https://doi.org/10.1016/j.molcel.2004.09.031>.
- [33] X. Xie, J. Lu, E.J. Kulbokas, T.R. Golub, V. Mootha, K. Lindblad-Toh, E.S. Lander, M. Kellis, Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals, *Nature* 434 (2005) 338–345, <https://doi.org/10.1038/nature03441>.
- [34] C. Bertoli, J.M. Skotheim, R.A.M. de Bruin, Control of cell cycle transcription during G1 and S phases, *Nat. Rev. Mol. Cell Biol.* 14 (2013) 518–528, <https://doi.org/10.1038/nrm3629>.
- [35] B.K. Lee, A.A. Bhinge, V.R. Iyer, Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis, *Nucleic Acids Res.* 39 (2011) 3558–3573, <https://doi.org/10.1093/nar/gkq1313>.
- [36] G.D. Grant, L. Brooks, X. Zhang, J.M. Mahoney, V. Martyanov, T.a. Wood, G. Sherlock, C. Cheng, M.L. Whitfield, Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors, *Mol. Biol. Cell* 24 (2013) 3634–3650, <https://doi.org/10.1091/mbc.E13-05-0264>.
- [37] Z. Zhu, J. Shendure, G.M. Church, Discovering functional transcription-factor combinations in the human cell cycle, *Genome Res.* 15 (2005) 848–855, <https://doi.org/10.1101/gr.339440>.
- [38] R. Elkon, C. Linhart, R. Sharan, R. Shamir, Y. Shiloh, Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells, *Genome Res.* 13 (2003) 773–780, <https://doi.org/10.1101/gr.947203>.
- [39] S.U. Mertens-Talcott, S. Chintharlapalli, X. Li, S. Safe, The oncogenic microRNA-27a targets genes that regulate specificity protein transcription factors and the G2-M checkpoint in MDA-MB-231 breast cancer cells, *Cancer Res.* 67 (2007) 11001–11011, <https://doi.org/10.1158/0008-5472.CAN-07-2416>.
- [40] I.K. Guttilla, B.A. White, Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cancer cells, *J. Biol. Chem.* 284 (2009) 23204–23216, <https://doi.org/10.1074/jbc.M109.031427>.
- [41] M. Lerner, J. Lundgren, S. Akhoondi, A. Jahn, H.F. Ng, F.A. Moqadam, J.A.F. Oude Vrielink, R. Agami, M.L. Den Boer, D. Grandér, O. Sangfelt, MiRNA-27a controls FBW7/hCDC4-dependent cyclin E degradation and cell cycle progression, *Cell Cycle* 10 (2011) 2172–2183, <https://doi.org/10.4161/cc.10.13.16248>.
- [42] M.J. Bueno, M. Malumbres, MicroRNAs and the cell cycle, *Biochim. Biophys. Acta* 1812 (2011) 592–601, <https://doi.org/10.1016/j.bbadi.2011.02.002>.
- [43] R.R. Chivukula, J.T. Mendell, Circular reasoning: microRNAs and cell-cycle control, *Trends Biochem. Sci.* 33 (2008) 474–481, <https://doi.org/10.1016/j.tibs.2008.06.008>.
- [44] K. Abdelmohsen, S. Srikantan, Y. Kuwano, M. Gorospe, miR-519 reduces cell proliferation by lowering RNA-binding protein HuR levels, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 20297–20302, <https://doi.org/10.1073/pnas.0809376106>.
- [45] J. Shao, J. Cao, Y. Liu, H. Mei, Y. Zhang, W. Xu, MicroRNA-519a promotes proliferation and inhibits apoptosis of hepatocellular carcinoma cells by targeting FOXF2, *FEBS Open Bio* 5 (2015) 893–899, <https://doi.org/10.1016/j.fob.2015.10.009>.
- [46] Y. Wang, S. Baskerville, A. Shenoy, J.E. Babiarz, L. Baehner, R. Blelloch, Embryonic stem cell-specific microRNAs regulate the G1–S transition and promote rapid proliferation, *Nat. Genet.* 40 (2008) 1478–1483, <https://doi.org/10.1038/ng.250>.
- [47] J.S. Imam, K. Buddavarapu, J.S. Lee-Chang, S. Ganapathy, C. Camosy, Y. Chen, M.K. Rao, MicroRNA-185 suppresses tumor growth and progression by targeting the Six1 oncogene in human cancers, *Oncogene* 29 (2010) 4971–4979, <https://doi.org/10.1038/onc.2010.233>.
- [48] Y. Takahashi, A.R.R. Forrest, E. Maeno, T. Hashimoto, C.O. Daub, J. Yasuda, MiR-107 and MiR-185 can induce cell cycle arrest in human non small cell lung cancer cell lines, *PLoS One* 4 (2009) <https://doi.org/10.1371/journal.pone.0006677>.
- [49] J.H. Yoon, Y.J. Choi, W.S. Choi, H. Ashktorab, D.T. Smoot, S.W. Nam, J.Y. Lee, W.S. Park, GKN1-miR-185-DNMT1 axis suppresses gastric carcinogenesis through regulation of epigenetic alteration and cell cycle, *Clin. Cancer Res.* 19 (2013) 4599–4610, <https://doi.org/10.1158/1078-0432.CCR-12-3675>.
- [50] J.-J. Lee, A. Drakaki, D. Iliopoulos, K. Struhl, MiR-27b targets PPAR γ to inhibit growth, tumor progression and the inflammatory response in neuroblastoma cells, *Oncogene* 31 (2012) 3818–3825, <https://doi.org/10.1038/onc.2011.543>.
- [51] A.V. Olaru, G. Ghiaur, S. Yamanaka, D. Luvsanjav, F. An, I. Popescu, S. Alexandrescu, S. Allen, T.M. Pawlik, M. Torbenson, C. Georgiades, L.R. Roberts, G.J. Gores, A. Ferguson-Smith, M.I. Almeida, G.A. Calin, E. Mezey, F.M. Selaru, MicroRNA down-regulated in human cholangiocarcinoma control cell cycle through multiple targets involved in the G1/S checkpoint, *Hepatology* 54 (2011) 2089–2098, <https://doi.org/10.1002/hep.24591>.

- [52] S. Yamanaka, N.R. Campbell, F. An, S.C. Kuo, J.J. Potter, E. Mezey, A. Maitra, F.M. Selaru, Coordinated effects of microRNA-494 induce G₂/M arrest in human cholangiocarcinoma, *Cell Cycle* 11 (2012) 2729–2738, <https://doi.org/10.4161/cc.21105>.
- [53] I.C. Lopez-Mejia, L. Fajas, Cell cycle regulation of mitochondrial function, *Curr. Opin. Cell Biol.* 33 (2015) 19–25, <https://doi.org/10.1016/j.celb.2014.10.006>.
- [54] P. Moreno-Layseca, C.H. Streuli, Signalling pathways linking integrins with cell cycle progression, *Matrix Biol.* 34 (2014) 144–153, <https://doi.org/10.1016/j.matbio.2013.10.011>.
- [55] O. Anczkukow, A.Z. Rosenberg, M. Akerman, S. Das, L. Zhan, R. Karni, S.K. Muthuswamy, A.R. Krainer, The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation, *Nat. Struct. Mol. Biol.* 19 (2012) 220–228, <https://doi.org/10.1038/nsmb.2207>.
- [56] M. Meryet-Figuere, B. Alaei-Mahabadi, M.M. Ali, S. Mitra, S. Subhash, G.K. Pandey, E. Larsson, C. Kanduri, Temporal separation of replication and transcription during S-phase progression, *Cell Cycle* 13 (2014) 3241–3248, <https://doi.org/10.4161/15384101.2014.953876>.
- [57] A. Raj, A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences, *Cell* 135 (2008) 216–226, <https://doi.org/10.1016/j.cell.2008.09.050>.
- [58] S. Tyagi, Tuning noise in gene expression, *Mol. Syst. Biol.* 11 (2015) 805, <https://doi.org/10.1525/msb.20156210>.
- [59] F. Iwamoto, M. Stadler, K. Chalupníková, E. Oakeley, Y. Nagamine, Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU, *Exp. Cell Res.* 314 (2008) 1378–1391, <https://doi.org/10.1016/j.yexcr.2008.01.006>.
- [60] A. Zetterberg, D. Killander, Quantitative cytophotometric and autoradiographic studies on the rate of protein synthesis during interphase in mouse fibroblasts in vitro, *Exp. Cell Res.* 40 (1965) 1–11.
- [61] S. Marguerat, J. Bähler, Coordinating genome expression with cell size, *Trends Genet.* 28 (2012) 560–565, <https://doi.org/10.1016/j.tig.2012.07.003>.
- [62] O. Padovan-Merhar, G.P. Nair, A.G. Biaesch, A. Mayer, S. Scarfone, S.W. Foley, A.R. Wu, L.S. Churchman, A. Singh, A. Raj, Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms, *Mol. Cell* 58 (2015) 339–352, <https://doi.org/10.1016/j.molcel.2015.03.005>.
- [63] A. Sanchez, I. Golding, Genetic determinants and cellular constraints in noisy gene expression, *Science* 342 (2013) 1188–1193, <https://doi.org/10.1126/science.1242975>.
- [64] Y. Liu, S. Chen, S. Wang, F. Soares, M. Fischer, F. Meng, Z. Du, C. Lin, C. Meyer, J.A. DeCaprio, M. Brown, X.S. Liu, H.H. He, Transcriptional landscape of the human cell cycle, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) 3473–3478, <https://doi.org/10.1073/pnas.1617636114>.
- [65] H. Fan, S. Penman, Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis, *J. Mol. Biol.* 50 (1970) 655–670, [https://doi.org/10.1016/0022-2836\(70\)90091-4](https://doi.org/10.1016/0022-2836(70)90091-4).
- [66] C.R. Stumpf, M.V. Moreno, A.B. Olshen, B.S. Taylor, D. Ruggero, The translational landscape of the mammalian cell cycle, *Mol. Cell* 52 (2013) 574–582, <https://doi.org/10.1016/j.molcel.2013.09.018>.
- [67] J.-M. Peters, The anaphase promoting complex/cyclosome: a machine designed to destroy, *Nat. Rev. Mol. Cell Biol.* 7 (2006) 644–656, <https://doi.org/10.1038/nrm1988>.
- [68] S. Sadasivam, J.A. DeCaprio, The DREAM complex: master coordinator of cell cycle-dependent gene expression, *Nat. Rev. Cancer* 13 (2013) 585–595, <https://doi.org/10.1038/nrc3556>.
- [69] A. Zetterberg, O. Larsson, K.G. Wiman, What is the restriction point? *Curr. Opin. Cell Biol.* 7 (1995) 835–842, [https://doi.org/10.1016/0955-0674\(95\)80067-0](https://doi.org/10.1016/0955-0674(95)80067-0).
- [70] S. Cooper, Reappraisal of serum starvation, the restriction point, G₀, and G₁ phase arrest points, *FASEB J.* 17 (2003) 333–340, <https://doi.org/10.1096/fj.02-0352rev>.
- [71] G. Donati, L. Montanaro, M. Derenzini, Ribosome biogenesis and control of cell proliferation: p53 is not alone, *Cancer Res.* 72 (2012) 1602–1607, <https://doi.org/10.1158/0008-5472.CAN-11-3992>.
- [72] J.R. Friedman, J. Nunnari, Mitochondrial form and function, *Nature* 505 (2014) 335–343, <https://doi.org/10.1038/nature12985>.
- [73] G.M. Richardson, J. Lannigan, I.G. Macara, Does FACS perturb gene expression? *Cytometry A* 87 (2015) 166–175, <https://doi.org/10.1002/cyto.a.22608>.
- [74] A.M.J. Rattray, B. Müller, The control of histone gene expression, *Biochem. Soc. Trans.* 40 (2012) 880–885, <https://doi.org/10.1042/BST20120065>.
- [75] L. Yang, M. Duff, B. Gravelley, G. Carmichael, L.-L. Chen, Genomewide characterization of non-polyadenylated RNAs, *Genome Biol.* 12 (2011) R16, <https://doi.org/10.1186/gb-2011-12-2-r16>.
- [76] V. Kari, O. Karpik, B. Tieg, M. Kriegs, E. Dikomey, H. Krebber, Y. Begus-Nahrmann, S.A. Johnsen, A. Subset, Of histone H2B genes produces polyadenylated mRNAs under a variety of cellular conditions, *PLoS One* 8 (2013) 1–12, <https://doi.org/10.1371/journal.pone.0063745>.
- [77] A. Jauhainen, C. Thomsen, L. Strömbom, P. Grundevik, C. Andersson, A. Danielsson, M.K. Andersson, O. Nerman, L. Rörkvist, A. Ståhlberg, P. Åman, Distinct cytoplasmic and nuclear functions of the stress induced protein DDIT3/CHOP/GADD153, *PLoS One* 7 (2012), e33208. <https://doi.org/10.1371/journal.pone.0033208>.
- [78] D.T. Rutkowski, R.S. Hegde, Regulation of basal cellular physiology by the homeostatic unfolded protein response, *J. Cell Biol.* 189 (2010) 783–794, <https://doi.org/10.1083/jcb.201003138>.
- [79] F. Zhang, R.B. Hamanaka, E. Bobrovnikova-Marjon, J.D. Gordan, M.S. Dai, H. Lu, M.C. Simon, J.A. Diehl, Ribosomal stress couples the unfolded protein response to p53-dependent cell cycle arrest, *J. Biol. Chem.* 281 (2006) 30036–30045, <https://doi.org/10.1074/jbc.M604674200>.
- [80] S. Dolatabadi, J. Candia, N. Akrap, C. Vannas, T. Tesan Tomic, W. Losert, G. Landberg, P. Åman, A. Ståhlberg, Cell cycle and cell size dependent gene expression reveals distinct subpopulations at single-cell level, *Front. Genet.* 8 (2017), 1. <https://doi.org/10.3389/fgene.2017.00001>.
- [81] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, *Bioinformatics* 29 (2013) 15–21, <https://doi.org/10.1093/bioinformatics/bts635>.
- [82] J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. Van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T.J.

- Hubbard, GENCODE: the reference human genome annotation for the ENCODE project, *Genome Res.* 22 (2012) 1760–1774, <https://doi.org/10.1101/gr.135350.111>.
- [83] S. Anders, P.T. Pyl, W. Huber, HTSeq—a python framework to work with high-throughput sequencing data, *Bioinformatics* 31 (2015) 166–169, <https://doi.org/10.1093/bioinformatics/btu638>.
- [84] A.T.L. Lun, K. Bach, J.C. Marioni, Pooling across cells to normalize single-cell RNA sequencing data with many zero counts, *Genome Biol.* 17 (2016), 75. <https://doi.org/10.1186/s13059-016-0947-7>.
- [85] J.D. Storey, A direct approach to false discovery rates, *J. R. Stat. Soc. Ser. B Stat Methodol.* 64 (2002) 479–498, <https://doi.org/10.1111/1467-9868.00346>.
- [86] J.A. Blake, J.T. Eppig, J.A. Kadin, J.E. Richardson, C.L. Smith, C.J. Bult, A. Anagnostopoulos, R.M. Baldarelli, J.S. Beal, S.M. Bello, O. Blodgett, N.E. Butler, L.E. Corbani, H. Dene, H.J. Drabkin, K.L. Forthofer, S.L. Giannatto, P. Hale, D.P. Hill, L. Hutchins, M. Knowlton, A. Lavertu, M. Law, J.R. Lewis, V. Lopez, D. Maghini, D. Perry, M. McAndrews, D. Miers, H. Montenko, L. Ni, H. Onda, J.M. Recla, D.J. Reed, B. Richards-Smith, D. Sitnikov, M. Tomczuk, L. Wilming, Y. Zhu, Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse, *Nucleic Acids Res.* 45 (2017) D723–D729, <https://doi.org/10.1093/nar/gkw1040>.
- [87] S. Petropoulos, D. Edsgård, B. Reinius, Q. Deng, S.P. Panula, S. Codeluppi, A. Plaza Reyes, S. Linnarsson, R. Sandberg, F. Lanner, Single-cell RNA-Seq reveals lineage and X chromosome dynamics in human preimplantation embryos, *Cell* 165 (2016) 1012–1026, <https://doi.org/10.1016/j.cell.2016.03.023>.
- [88] J.J. Verbeek, N. Vlassis, B. Kröse, A k-segments algorithm for finding principal curves, *Pattern Recogn. Lett.* 23 (2002) 1009–1017, [https://doi.org/10.1016/S0167-8655\(02\)00032-6](https://doi.org/10.1016/S0167-8655(02)00032-6).
- [89] E. Larsson, C. Sander, D. Marks, mRNA turnover rate limits siRNA and microRNA efficacy, *Mol. Syst. Biol.* 6 (2010), 433. <https://doi.org/10.1038/msb.2010.89>.
- [90] E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, J.E. Darnell, Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes, *Genome Res.* 13 (2003) 1863–1872, <https://doi.org/10.1101/gr.1272403>.
- [91] W.S. Lai, J.S. Parker, S.F. Grissom, D.J. Stumpo, P.J. Blackshear, Novel mRNA targets for tristetraprolin (TTP) identified by global analysis of stabilized transcripts in TTP-deficient fibroblasts, *Mol. Cell. Biol.* 26 (2006) 9196–9208, <https://doi.org/10.1128/MCB.00945-06>.