



Article

Unsupervised Machine Learning Reveals Temporal Components of Gene Expression in HeLa Cells Following Release from Cell Cycle Arrest

Tom Maimon, Yaron Trink, Jacob Goldberger and Tomer Kalisky * 

Faculty of Engineering, Bar-Ilan Institute of Nanotechnology and Advanced Materials (BINA), Bar-Ilan University, Ramat Gan 5290002, Israel; tm.maimon@gmail.com (T.M.); trinkya@biu.ac.il (Y.T.); jacob.goldberger@biu.ac.il (J.G.)

* Correspondence: tomer.kalisky@biu.ac.il; Tel.: +972-3-738-4656

Abstract

Gene expression measurements of tissues, tumors, or cell lines taken over multiple time points are valuable for describing dynamic biological phenomena such as the response to growth factors. However, such phenomena typically involve multiple biological processes occurring in parallel, making it difficult to identify and discern their respective contributions at any time point. Here, we demonstrate the use of unsupervised machine learning to deconvolve a series of time-dependent gene expression measurements into its underlying temporal components. We first downloaded publicly available RNAseq data obtained from synchronized HeLa cells at consecutive time points following release from cell cycle arrest. Then, we used Fourier analysis and Topic modeling to reveal three underlying components and their relative contributions at each time point. We identified two temporal components with oscillatory behavior, corresponding to the G1-S and G2-M phases of the cell cycle, and a third component with a transient expression pattern, associated with the immediate early response gene program, regulation of cell proliferation, and cervical cancer. This study demonstrates the use of unsupervised machine learning to identify hidden temporal components in biological systems, with potential applications to early detection and monitoring of diseases and recovery processes.

Keywords: cell cycle; Fourier transform; topic modeling; machine learning; HeLa cells



Academic Editor: Tae-Hong Kang

Received: 10 June 2025

Revised: 19 August 2025

Accepted: 4 September 2025

Published: 28 September 2025

Citation: Maimon, T.; Trink, Y.;

Goldberger, J.; Kalisky, T.

Unsupervised Machine Learning

Reveals Temporal Components of

Gene Expression in HeLa Cells

Following Release from Cell Cycle

Arrest. *Int. J. Mol. Sci.* **2025**, *26*, 9491.

[https://doi.org/10.3390/](https://doi.org/10.3390/ijms26199491)

[ijms26199491](https://doi.org/10.3390/ijms26199491)

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)).

1. Introduction

Complex biological systems such as tissues and tumors contain millions of cells. These cells dynamically change their transcriptional state over time, as multiple gene programs—coordinated sets of genes that work together to perform specific biological tasks—are activated and repressed through a complex network of interactions. Dynamic changes in transcriptional cell states underlie fundamental biological processes, such as embryo development, tissue regeneration, and the onset and progression of cancer. A common approach to characterize cell state dynamics is to perform RNA sequencing at multiple time points, either at the ‘bulk’ or single-cell level, yielding a series of gene expression profiles that represent the underlying cell states, and that usually form continuous trajectories in latent space. However, these expression profiles are typically a combination of multiple time-dependent components, where each component is associated with distinct biological processes and gene programs. A major challenge is to “deconvolve” these temporal components. This involves delineating each component, identifying its associated gene

programs, and inferring its relative contribution to the overall gene expression profiles at each time point.

Here, we demonstrate the use of unsupervised machine learning to reveal the temporal components in a simple model biological system. We first downloaded a dataset of bulk RNAseq measurements performed by Dominguez et al. on synchronized HeLa cells at 14 consecutive time points—approximately two cell cycles—following release from cell cycle arrest [1,2]. Then, we performed Fourier analysis and found that most genes have a periodicity of either one or two cycles over time. Next, we used topic modeling, an unsupervised machine learning technique, and found that the series of gene expression profiles can be represented as a mixture of three components, two of which are periodic over time and correspond to the G1-S and G2-M phases of the cell cycle, and a third topic with a transient temporal behavior, that is associated with the immediate early response gene program, regulation of cell proliferation, and cervical cancer. This study demonstrates the potential of machine learning algorithms to deconvolve hidden temporal components in complex biological systems, with potential applications for early disease detection, as well as monitoring disease progression and recovery processes.

2. Results

2.1. RNA Velocity Analysis of Periodically Expressed Genes Reveals a Time Lag Between Spliced and Un-Spliced mRNA

We downloaded a published dataset of RNAseq measurements that were performed by Dominguez et al. [1,2]. In that study, HeLa cells were first synchronized by a double thymidine block. Then, following release from cell cycle arrest, RNA was collected and sequenced at 14 consecutive time points which corresponds to approximately two cell cycles (Figure 1A, Table S1). The authors also identified a set of 67 “core” cell cycle related genes with periodic behavior, and categorized them as “G1-S related” or “G2-M related” according to the stage at which their expression is maximal (Table S1).

We started by manually inspecting the expression levels of spliced and un-spliced mRNA of “core” cell cycle genes such as *CCNE2* and *UNG* (G1-S related) and *MKI67* and *TOP2A* (G2-M related, Figure 1B) [3]. As expected, we observed that the yet-unspliced mRNA typically precedes the spliced mRNA by a single sampling interval, which corresponds to a time interval of approximately 1.5–3 h. We confirmed this by performing a cross-correlation between the spliced and un-spliced mRNA levels averaged over the 67 “core” cell cycle genes (Figure 1C).

2.2. Fourier Analysis Identifies Sets of Genes with Potentially Transient and Oscillatory Behaviors over Time

To test the periodicity of every gene in the transcriptome, we performed Fourier analysis. For each gene we plotted its “dominant frequency”, that is, the frequency with the highest spectral density, versus its “dominant frequency score”, which measures the degree to which the score of this frequency is higher than the scores of the other frequencies (see Section 4 and Figure 2A). We found that the first and second dominant frequencies—that correspond to either one or two whole cycles—contain genes with scores that are higher than those within other dominant frequencies (Figure 2A). Moreover, we observed that genes with dominant frequencies corresponding to three cycles and higher have scores comparable to those derived from randomized datasets, which were generated by randomly shuffling the order of counts for each gene (Figure S1). This suggests that most of the information in our dataset is contained in the expression levels of genes with a periodicity of either one or two cycles, indicating potential transient or oscillatory behavior, respectively.

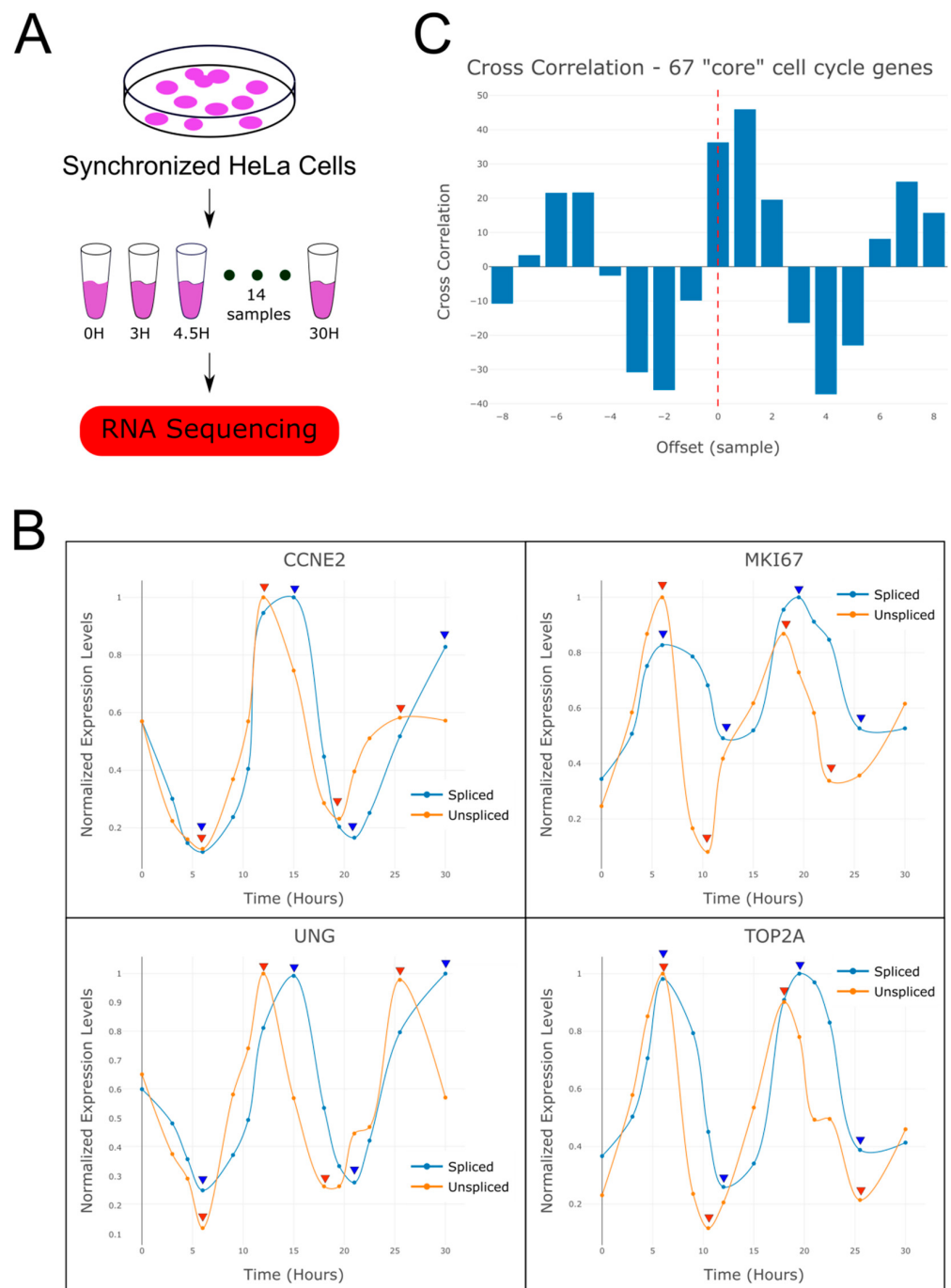


Figure 1. RNA velocity analysis of periodically expressed genes reveals a time lag between spliced and un-spliced mRNA. **(A)** Shown is a sketch of the experiment performed by Dominguez et al. [1,2] from which the RNAseq dataset was produced. HeLa cells were sampled at 14 consecutive time points following release from cell cycle arrest and their RNA was sequenced. **(B)** Shown are expression levels over time of spliced and unspliced mRNA for selected genes that are known to be over-expressed at the G1-S phases (*CCNE2*, *UNG*) or the G2-M phases (*MKI67*, *TOP2A*) of the cell cycle. These indicate that the cells completed approximately two cell divisions during the experiment. It can be seen that the yet-unspliced mRNA precedes the spliced mRNA. To assist visibility, expression levels were normalized to their maximum and spline curves were plotted through the data points. Arrowheads mark the maxima and minima of spliced (blue) and unspliced (red) mRNA levels. **(C)** A cross-correlation plot between the spliced and un-spliced mRNA levels averaged over the 67 “core” cell cycle genes (i.e., periodically expressed genes identified by Dominguez et al. [1,2]) shows a significant time lag between the spliced and yet un-spliced mRNA.

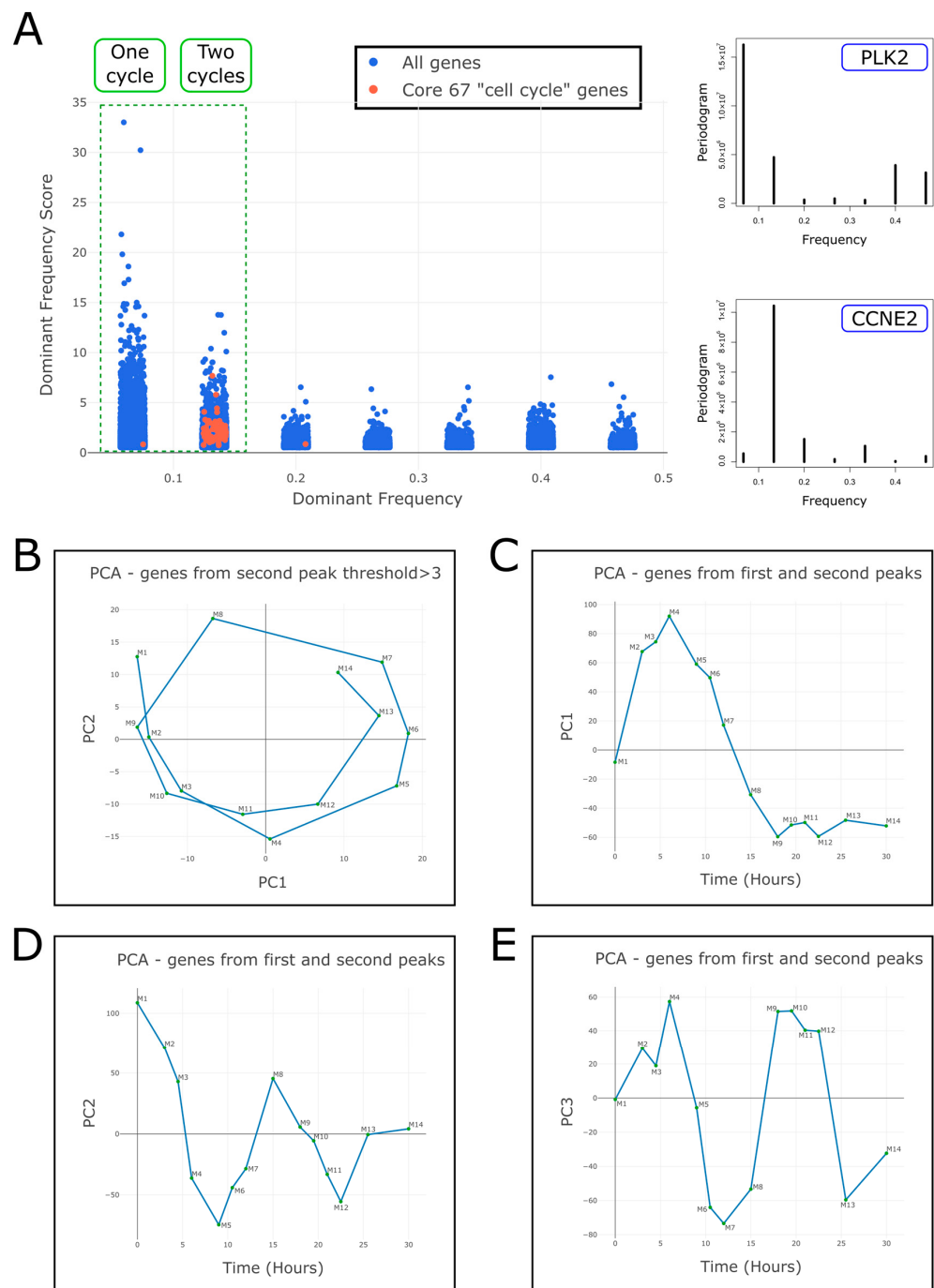


Figure 2. Fourier analysis identifies sets of genes with potentially transient and oscillatory behaviors over time. **(A)** For each gene, a periodogram was calculated (see specific examples for *PLK2* and *CCNE2* in the right panels) and its “dominant frequency” and “dominant frequency score” were plotted (left panel, see Section 4). It can be seen that almost all of the “core” cell cycle genes (65 out of 67 red dots) have a dominant frequency corresponding to two whole cycles. Likewise, the first and second dominant frequencies contain genes with scores that are higher than those within other dominant frequencies, whereas genes in the 3rd dominant frequency and onwards have scores comparable to those of randomly shuffled datasets (Figure S1). This indicates that the genes within the first and second dominant frequencies contain most of the periodic information in our dataset. **(B)** When we select genes from the second dominant frequency with scores > 3 and perform PCA, the samples form a circular trajectory in latent space that completes almost two cycles. **(C–E)** When we select genes from the first and second dominant frequencies and perform PCA, the samples form a trajectory in latent space with a transient behavior along PC1 axis and periodic behaviors along the PC2 and PC3 axes (that also create a circular trajectory in the PC2 vs. PC3 plane).

We next selected genes from the second dominant frequency (with scores > 3) and performed PCA. We found that the samples form a circular trajectory in latent space that corresponds to almost two complete cell cycles (Figure 2B). Moreover, separating the spliced and yet-unspliced mRNA expression profiles results in two circular trajectories in latent space with a phase difference between them (Figure S2). When we selected the genes from the first and second dominant frequencies, we found that the samples form a trajectory in latent space with two patterns of behavior: a transient pattern along PC1 (Figure 2C) and an oscillatory pattern along PC2 and PC3 (Figure 2D,E).

2.3. Topic Modeling Reveals Three Temporal Components, Two of Which Are Periodic, Corresponding to the G1-S and G2-M Phases of the Cell Cycle, and a Third, Transient Component, Related to Immediate Early Response, Regulation of Cell Proliferation, and Cervical Cancer

“Topic models” are used in computer science to describe collections of documents that contain varying proportions of words from a predetermined number k of different “topics” [4,5]. By fitting a topic model to a set of documents, it is possible to discover both the k latent topics (i.e., the probability of occurrence of each word from the vocabulary within a topic) and their proportions in each individual document. In our case we assume that each cell state (gene expression profile) at a given time point is a weighted sum of k hidden components whose proportions vary with time. We therefore treat each cell state at a given time point as a document, each gene in the human genome as a word in the vocabulary, the presence of a specific RNA transcript within a cell state as the occurrence of a specific word in a document, the expression level (number of transcripts) of a gene in a cell state as the number of occurrences of a word in a document, and each latent component as a topic. Topic modeling can thus be used to find the k components (i.e., the probability of observing an RNA transcript from each gene within a component) and their proportions at each individual time point.

We fitted a topic model with $k = 3$ topics to the 14 gene expression profiles and identified three components/topics with distinct temporal patterns, which we labeled as “k1”, “k2” and “k3” (Figure 3A and Table S2). We found that topics k1 and k3 have an oscillatory behavior over the course of both cycles following release from cell cycle arrest and are periodically over-expressed at specific phases of the cell cycle (Figure 3A,B). Topic k2 however is transiently over-expressed mainly during the first cycle following release from cell cycle arrest. We confirmed that the temporal pattern of each topic is also observed in selected genes associated with it. For example, the genes *CCNE2* and *UNG* whose expression is periodic are associated with topic k1 (Figures 1B and 4A), whereas the genes *MKI67* and *TOP2A* whose expression is also periodic, but with a different phase, are associated with topic k3 (Figures 1B and 4A). Likewise, the genes *KIFC3*, *PHLDB2*, and *PLK2*, which are associated with topic k2, are transiently over-expressed following release from cell cycle arrest (Figures 3C and 4A).

To identify and characterize the topics, we performed Gene Ontology (GO) enrichment analysis. For each topic, we selected genes that are significantly over-expressed and used them as input to ToppGene [6] (Figure 4A, Tables S3–S5). We found that topic k1 over-expresses genes related to the G1-S phase of the cells cycle (e.g., DNA replication and G1 to S cell cycle control) and that topic k3 over-expresses genes that are related to the G2-M phase of the cells cycle (e.g., mitosis and regulation of the G2/M transition). In contrast, topic k2 was found to over-express genes associated with the following gene programs: (i) immediate early response [7–10] (Table S6 and Figure S3); (ii) regulation of cell proliferation (Table S4), and (iii) cervical carcinoma (Table S4).

We confirmed these observations by calculating the posterior probabilities $p(k1 | \text{gene})$, $p(k2 | \text{gene})$, and $p(k3 | \text{gene})$ for genes belonging to specific gene sets and GO terms (Figure 4C). These probabilities represent the association between these genes and each of

the three topics. Indeed, we observed that within the 67 “core” cell cycle genes, the G1-S related genes were associated with topic k1 and the G2-M related genes were associated with topic k3. Likewise, genes related to immediate early response and cervical carcinoma were more strongly associated with topic k2.

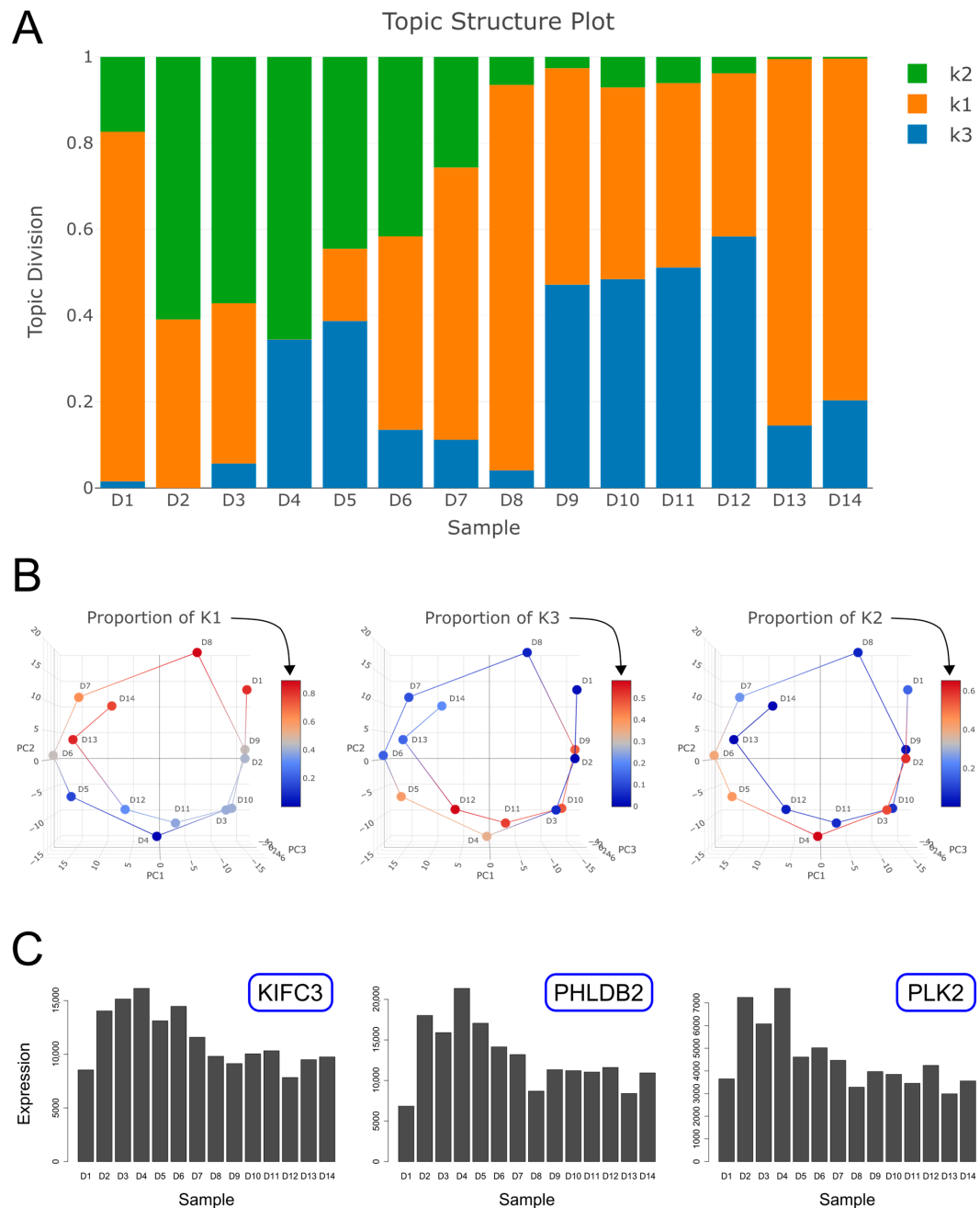


Figure 3. Topic modeling reveals three temporal components with periodic and transient behaviors over time. (A) Shown is a structure plot of the proportions of each component/topic within each sample. Topics k1 and k3 are periodically expressed over the course of both cycles (samples D1–D14), whereas topic k2 is transiently over-expressed mainly during the first cycle (samples D1–D7) following release from cell cycle arrest. (B) Shown is a visualization of the proportions of each topic along the circular trajectory in latent space representing the two cell cycles (red—high proportion, blue—low proportion). It can be seen that topics k1 and k3 are over-expressed in distinct phases during both cycles, whereas topic k2 is over-expressed during the first cycle only. Note that topic k1 is expressed mainly in the upper half of the circular trajectory, whereas topic k3 is expressed mainly in its lower half. (C) Shown are the expression levels of three representative genes that are associated with topic k2 and are transiently over-expressed following release from cell cycle arrest.

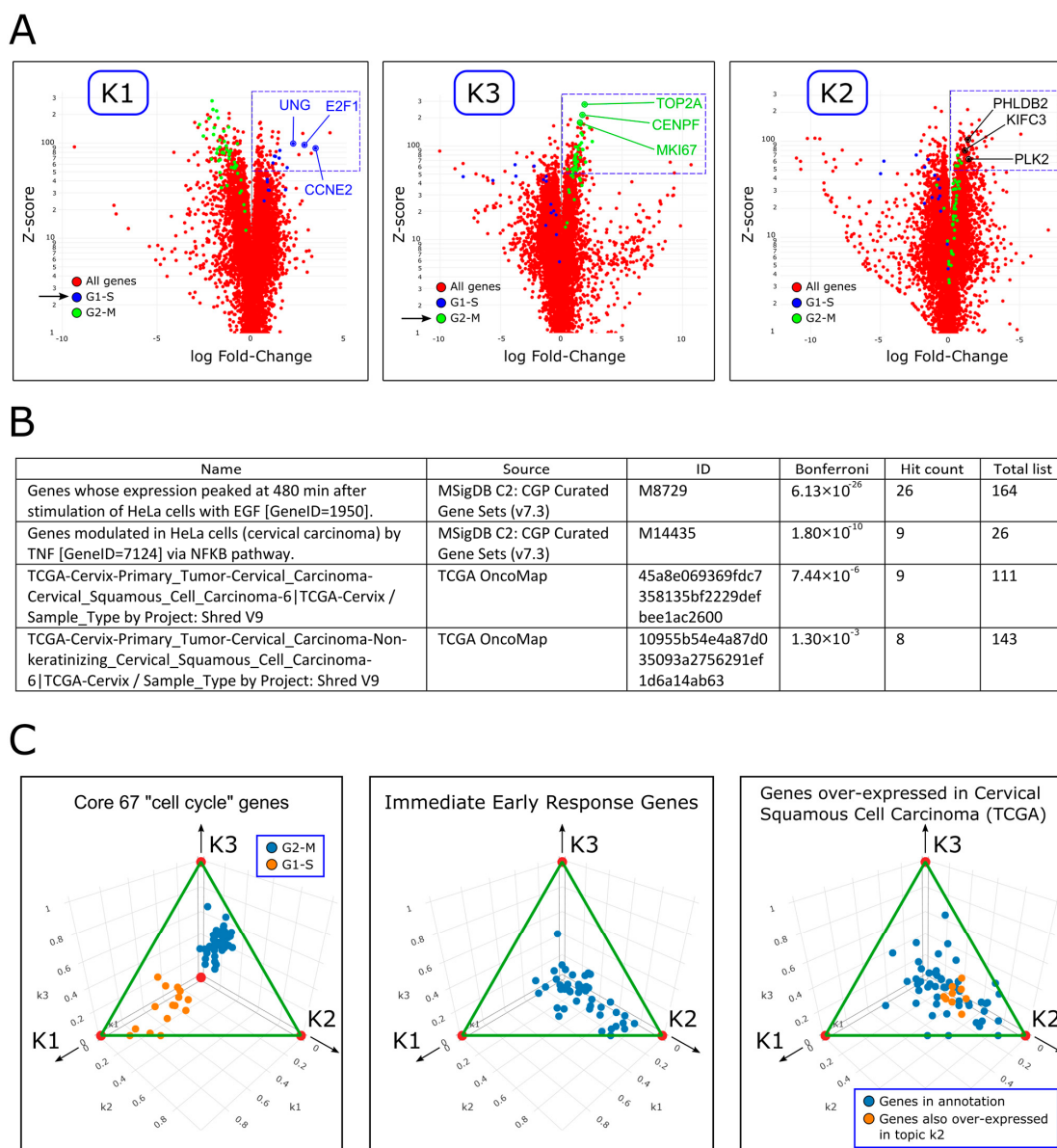


Figure 4. Gene Ontology (GO) enrichment shows that the two periodic components correspond to the G1-S and G2-M phases of the cell cycle, whereas the third transient component is related to immediate early response, regulation of cell proliferation, and cervical cancer. **(A)** Shown are volcano plots illustrating the over-expression of each gene in each of the three topics. It can be seen that genes known to be over-expressed in the G1-S phases of the cell cycle are over-expressed in topic k1 (**left panel**), genes known to be over-expressed in the G2-M phases of the cell cycle are over-expressed in topic k3 (**middle panel**), and genes that are transiently over-expressed following release from cell cycle arrest are over-expressed in topic k2 (**right panel**). **(B)** GO enrichment analysis of genes that are over-expressed in topic k2 (Z-score > 50) shows many of them are related to cervical cancer or involved in response to growth factors (Table S4). **(C)** Shown are the posterior probabilities $p(\text{topic}|\text{gene})$ for selected gene sets or specific GO terms. Each point represents a gene, and its components along the k1, k2, and k3 axes are the posterior probabilities $p(k1|\text{gene})$, $p(k2|\text{gene})$, and $p(k3|\text{gene})$, which represent the association between that gene and each of the three topics. It can be seen that topic k1 is highly associated with genes that are known to be over-expressed in the G1-S phases of the cell cycle (**left panel**) and that topic k3 is highly associated with genes that are known to be over-expressed in the G2-M phases of the cell cycle. Topic k2 is associated with genes that are related to the immediate early response program and to cervical carcinoma (**middle and right panels**). Note that all points (=genes) are located on the triangle-shaped two-dimensional simplex since the sum of the posterior probabilities for each gene equals 1.

3. Discussion

In this study we applied Fourier analysis and Topic modeling to a simple in vitro model system to identify its latent components. We identified two components with a periodic temporal behavior (topics k1 and k3), that correspond to the G1-S and G2-M phases of the cell cycle, and a third component with a transient expression pattern (topic k2) that is associated with the immediate early response gene program, regulation of cell proliferation, and cervical cancer.

Specific examples of genes that are transiently over-expressed following release from cell cycle arrest, and that have a high probability for being expressed in topic k2, are *KIFC3*, *PHLDB2*, and *PLK2* (Figures 3C and 4A). These genes are known to be associated with cervical carcinoma (Table S4). Moreover, *PLK2*, a member of the polo-like kinase family, is also known to be associated with positive regulation of the cell cycle [11] and is regarded as an immediate early response gene [10] (Figure S17). Other examples include the genes *JUN* (*C-Jun*), *FOS* (*c-Fos*), *FOSB*, and *FOSL1* (*FRA-1*), that are also transiently over-expressed following release from cell cycle arrest (Figure S13). The protein products of these genes are components of the AP-1 transcription factor complex that is thought to be involved in cell proliferation and cancer progression [12,13]. Furthermore, the genes *JUN*, *FOS*, and *FOSB* are also members of the immediate early response gene family. The genes *RELB*, *NFKB1*, and *NFKB2*, are also transiently over-expressed following release from cell cycle arrest (Figure S18). These genes are members of the NF- κ B family of transcription factors and are known to induce target genes involved in initiation and progression of cancer [14,15]. We also observed that the apoptosis inhibitors *BIRC2* and *BIRC3* [16] are transiently over-expressed following release from cell cycle arrest (Figure S27), whereas the gene *BIRC5* (Survivin), which is a regulator of the mitotic cell cycle, shows a periodic pattern of expression. Additional examples are shown in Figures S4–S27.

The association of immediate early response genes with topic k2 is not very surprising, since these genes are known to be rapidly and transiently transcribed in response to stimulation of cells by serum or growth factors such as epidermal growth factor (EGF), subsequently promoting phenotypic changes such as proliferation, differentiation, and survival [10]. In our study, it is likely that these genes facilitate cell-cycle reentry, promoting the transition from the quiescent G0 phase into the G1 phase. The association of topic k2 with cervical cancer is less straightforward. One possible explanation for this association is that, following release from cell cycle arrest, the cells enter a transient state lasting approximately twelve hours (samples D1–D7), during which they activate molecular mechanisms to resume the cell cycle. In normal differentiation, cell cycle exit is associated with lineage commitment and is typically regulated by tissue-specific mechanisms [17]. It is plausible that cell cycle reentry in HeLa cells is similarly regulated by mechanisms specific to the cervical cancer from which they were derived.

For some of the 67 “core” cell cycle genes we observed that the spliced and unspliced mRNA levels arrive at their first peak (either maximum or minimum) at the same time (Figure 1B), and only after a few hours the typical time lag between the spliced and un-spliced mRNA expression is acquired. This indicates that initially, during the few hours after release from cell cycle arrest, the spliced and un-spliced transcripts are synchronized in these genes. Additional evidence for this gradual acquisition of time lag between spliced and un-spliced mRNA can be seen for larger sets of genes with periodic behavior (Figure S2A,B). This indicates that in the transient cell state following release from cell cycle arrest, the transcriptional timing mechanisms operate in a way that is different from the normal cycling state. One possible explanation is that during this period, the primary mRNA transcribed from these genes is spliced, processed, and translated with minimal delay in order to rebuild the necessary mechanisms for cell cycle re-entry as soon as

possible. According to this hypothesis, splicing mechanisms participate in the immediate early response mechanism. Further studies at higher time resolution are required to systematically test this hypothesis and inspect the relationship between mRNA splicing and the immediate early response gene program.

4. Materials and Methods

4.1. Datasets and Preprocessing

The RNAseq dataset generated by Dominguez et al. [1,2] was downloaded from GEO (accession number GSE81485) using SRA tools (version 2.9.6). STAR [18] (version 2.7.3a) was used to align the reads to the human reference genome (hg38) and to obtain the raw counts matrix. DESeq2 [19] (version 1.36.0) was used to obtain the normalized counts matrix. No further batch-effect correction was applied since all samples were generated within the same experiment and laboratory, which minimized technical variability. Since low-count data is often disproportionately affected by technical noise (e.g., mapping errors) and is generally less suitable for correlation analysis, we removed genes whose total normalized counts across all 14 samples were below 10.

4.2. RNA Velocity

Velocyto [3] (<https://velocyto.org/>, accessed on 8 June 2018), a package for RNA velocity analysis, was used to obtain the read counts of the spliced and unspliced mRNA for each sample. The package was used to annotate and count the STAR-aligned reads. Broadly speaking, each read was annotated as “spliced” if it mapped only to exonic regions of known transcripts, or “unspliced” otherwise, i.e., if it mapped to an exon-intron boundary or entirely to an intron.

Due to processing and memory constraints, we diluted the reads to 20% using Samtools (version 1.19.2) before running Velocyto. Since read dilution is random, this should not affect genes with high expression levels, such as most cell-cycle-related genes analyzed in this study. The specific command line used to run Velocyto was: “`velocyto run-smartseq2-o./velocyto_out/-e my_analysis./bam_files_diluted_0.2/* genes.gtf`”. This was our sole use of the Velocyto package, and we did not employ any of the downstream R-based modeling tools (e.g., for estimation of the time derivative of gene expression).

Note that, since we are dealing with bulk RNA-seq rather than single-cell data, this analysis is only applicable to gene sets whose expression is assumed to be synchronized across all cells in the sample. This assumption is reasonable in our case, as the cells were synchronized using a double thymidine block. Our analysis is similar to that of La Manno et al. [3], who analyzed bulk RNA-seq time-course data from mouse liver during the circadian cycle and observed that unspliced mRNA levels of circadian-associated genes at each time point resembled the spliced mRNA levels at the subsequent time point (Figure 1e in La Manno et al. [3]).

DESeq2 normalization was performed for the spliced and unspliced counts matrixes combined.

4.3. Fourier Transform

For each gene, a Fourier analysis was performed using the R function “`periodogram()`” from the “TSA” (Time Series Analysis) R package (version 1.3.1). This function returns a vector of frequencies and a vector of scores (spectral densities) for each frequency. For each gene we found the “dominant frequency”, that is, the frequency with the highest score, and calculated a “dominant frequency score” which measures the degree to which the score of this frequency is higher than the scores of the other frequencies. Specifically, if for a particular gene the highest score is S_1 and its matching frequency is F_1 , the second

highest score is S2 and its frequency is F2, the third ranking score is S3 and its frequency is F3 etc., then the “dominant frequency” for that gene is F1, and the “dominant frequency score” is: $S1 / (S2 + S3)$. The significance of the dominant frequency scores was tested by comparing them to scores derived from randomized datasets, which were obtained by randomly shuffling the order of counts for each gene.

4.4. Topic Modeling

Topic modeling was performed on the normalized counts matrix using the R functions `fit_topic_model()`, `diff_count_analysis()`, and `structure_plot()` from the “fastTopics” R package [20] (version 0.6–135). We fitted a topic model with $k = 3$ topics that we labeled as “k1”, “k2”, and “k3”. Using these functions, we calculated the three probability distributions $p(\text{gene}|\text{k1})$, $p(\text{gene}|\text{k2})$, and $p(\text{gene}|\text{k3})$ for every gene in each one of the three topics, as well as the probabilities $p(\text{k1}|\text{sample})$, $p(\text{k2}|\text{sample})$, and $p(\text{k3}|\text{sample})$ for each topic in every one of the fourteen samples. We also tested a higher number of topics, but the results did not change significantly.

One way to visualize the association between a specific gene and each of the three topics is to calculate the “posterior probabilities” $p(\text{k1}|\text{gene})$, $p(\text{k2}|\text{gene})$, and $p(\text{k3}|\text{gene})$ and compare them. In order to do this, we used Bayes’ theorem:

$$p(\text{topic}|\text{gene}) = \frac{p(\text{gene}|\text{topic}) \cdot p(\text{topic})}{\sum_{\text{topic}} p(\text{gene}|\text{topic}) \cdot p(\text{topic})}$$

where for simplicity, we set the prior $p(\text{topic}) = 1/3$ for each topic.

4.5. Gene Ontology (GO) and Gene Set Enrichment Analysis

GO enrichment analysis was performed with Toppgene [21] (<https://toppgene.cchmc.org/>, last accessed on 8 September 2024).

Gene set enrichment analysis was performed with GSEA [22] (<https://www.gsea-msigdb.org/gsea>, last accessed on 11 June 2024). Genes were ranked in descending order according to the differences in their posterior probabilities between topic k2 and topics k1 and k3, using the option “Phenotype labels: Create on-the-fly phenotype by sample names”. The “Gene sets database” was defined as the set of immediate early genes (Table S6), the “Number of permutations” was set to 1000, the “Permutation type” was set to “gene_set”, and the “Metric for ranking genes” was set to “Diff_of_classes” (see Figure S3 for run details).

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms26199491/s1>. References [8–16,23–52] are cited in the Supplementary Materials.

Author Contributions: Study initiation and conception—T.M. and T.K.; Data analysis—T.M. and T.K.; Topic modeling—T.M., J.G. and T.K.; Other intellectual contribution—Y.T.; Manuscript writing—T.M. and T.K. All authors have read and agreed to the published version of the manuscript.

Funding: T.M., Y.T. and T.K. were supported by the Israel Science Foundation (ICORE no. 1902/12 and Grants no. 1634/13, 2017/13, and 1814/20), the Israel Ministry of Health (Grant no. 3-10146), the EU-FP7 (Marie Curie International Reintegration Grant no. 618592), the Data Science Institute at Bar-Ilan University (seed grant), the ICRF (Grant no. 19-101-PG), the Israel Ministry of Science (Grant no. 3-16220), the Israel Ministry of Justice (Veadat Haezvonot), and the Israel Cancer Association (Grant no. 20240114). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNAseq dataset generated by Dominguez et al. [1,2] is available at GEO (accession number GSE81485).

Acknowledgments: We wish to thank Shahar Alon and all the members of our lab for useful comments and suggestions. During the preparation of this work the authors used ChatGPT (GPT-4o, OpenAI, last accessed on 9 June 2025) in order to improve readability and language (e.g., refine sentence structure, wording, and grammar). After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dominguez, D.; Tsai, Y.H.; Weatheritt, R.; Wang, Y.; Blencowe, B.J.; Wang, Z. An Extensive Program of Periodic Alternative Splicing Linked to Cell Cycle Progression. *eLife* **2016**, *5*, e10288. [\[CrossRef\]](#)
2. Dominguez, D.; Tsai, Y.H.; Gomez, N.; Jha, D.K.; Davis, I.; Wang, Z. A High-Resolution Transcriptome Map of Cell Cycle Reveals Novel Connections between Periodic Genes and Cancer. *Cell Res.* **2016**, *26*, 946–962. [\[CrossRef\]](#)
3. La Manno, G.; Soldatov, R.; Zeisel, A.; Braun, E.; Hochgerner, H.; Petukhov, V.; Lidschreiber, K.; Kastrioti, M.E.; Lönnerberg, P.; Furlan, A.; et al. RNA Velocity of Single Cells. *Nature* **2018**, *560*, 494–498. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Dey, K.K.; Hsiao, C.J.; Stephens, M. Visualizing the Structure of RNA-Seq Expression Data Using Grade of Membership Models. *PLoS Genet.* **2017**, *13*, e1006599; Correction in *PLoS Genet.* **2017**, *13*, e1006759. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
6. Chen, J.; Bardes, E.E.; Aronow, B.J.; Jegga, A.G. ToppGene Suite for Gene List Enrichment Analysis and Candidate Gene Prioritization. *Nucleic Acids Res.* **2009**, *37* (Suppl. S2), W305–W311. [\[CrossRef\]](#)
7. Herschman, H.R. Primary Response Genes Induced by Growth Factors and Tumor Promoters. *Annu. Rev. Biochem.* **1991**, *60*, 281–319. [\[CrossRef\]](#)
8. O'Donnell, A.; Odrowaz, Z.; Sharrocks, A.D. Immediate early Gene Activation by the MAPK Pathways: What Do and Don't We Know? *Biochem. Soc. Trans.* **2012**, *40*, 58–66. [\[CrossRef\]](#)
9. Tullai, J.W.; Schaffer, M.E.; Mullenbrock, S.; Sholder, G.; Kasif, S.; Cooper, G.M. Immediate early and Delayed Primary Response Genes Are Distinct in Function and Genomic Architecture. *J. Biol. Chem.* **2007**, *282*, 23981–23995. [\[CrossRef\]](#)
10. Winkles, J.A. Serum- and Polypeptide Growth Factor-Inducible Gene Expression in Mouse Fibroblasts. In *Progress in Nucleic Acid Research and Molecular Biology*; Moldave, K., Ed.; Academic Press: Cambridge, MA, USA, 1997; pp. 41–78. [\[CrossRef\]](#)
11. de Cárcer, G.; Manning, G.; Malumbres, M. From Plk1 to Plk5: Functional evolution of polo-like kinases. *Cell Cycle* **2011**, *10*, 2255–2262. [\[CrossRef\]](#)
12. Casalino, L.; Talotta, F.; Cimmino, A.; Verde, P. The Fra-1/AP-1 Oncoprotein: From the 'Undruggable' Transcription Factor to Therapeutic Targeting. *Cancers* **2022**, *14*, 1480. [\[CrossRef\]](#)
13. Milde-Langosch, K. The Fos Family of Transcription Factors and Their Role in Tumourigenesis. *Eur. J. Cancer* **2005**, *41*, 2449–2461. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Da Costa, R.M.G.; Bastos, M.M.S.M.; Medeiros, R.; Oliveira, P.A. The NFκB Signaling Pathway in Papillomavirus-Induced Lesions: Friend or Foe? *Anticancer Res.* **2016**, *36*, 2073–2083.
15. Tilborghs, S.; Corthouts, J.; Verhoeven, Y.; Arias, D.; Rolfo, C.; Trinh, X.B.; van Dam, P.A. The Role of Nuclear Factor-Kappa B Signaling in Human Cervical Cancer. *Crit. Rev. Oncol. Hematol.* **2017**, *120*, 141–150. [\[CrossRef\]](#)
16. Silke, J.; Vaux, D.L. Two Kinds of BIR-Containing Protein—Inhibitors of Apoptosis, or Required for Mitosis. *J. Cell Sci.* **2001**, *114*, 1821–1827. [\[CrossRef\]](#)
17. Theilgaard-Mönch, K.; Pundhir, S.; Reckzeh, K.; Su, J.; Tapia, M.; Furtwängler, B.; Jendholm, J.; Jakobsen, J.S.; Hasemann, M.S.; Knudsen, K.J.; et al. Transcription Factor-Driven Coordination of Cell Cycle Exit and Lineage-Specification in Vivo during Granulocytic Differentiation. *Nat. Commun.* **2022**, *13*, 3595. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [\[CrossRef\]](#)
20. Carbonetto, P.; Luo, K.; Dey, K.; Hsiao, J.; Stephens, M. *fastTopics: Fast Algorithms for Fitting Topic Models and Non-Negative Matrix Factorizations to Count Data*, R Package Version 0.4–11; R Foundation: Vienna, Austria, 2021.
21. Chen, J.; Xu, H.; Aronow, B.J.; Jegga, A.G. Improved Human Disease Candidate Gene Prioritization Using Mouse Phenotype. *BMC Bioinform.* **2007**, *8*, 392. [\[CrossRef\]](#)

22. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
23. Bi, L.; Ma, F.; Tian, R.; Zhou, Y.; Lan, W.; Song, Q.; Cheng, X. AJUBA increases the cisplatin resistance through hippo pathway in cervical cancer. *Gene* **2018**, *644*, 148–154. [[CrossRef](#)]
24. Kalan, S.; Matveyenko, A.; Loayza, D. LIM Protein Ajuba Participates in the Repression of the ATR-Mediated DNA Damage Response. *Front. Genet.* **2013**, *4*, 95. [[CrossRef](#)]
25. Mehdi, H.K.; Raju, K.; Sheela, S.R. Association of P16, Ki-67, and CD44 expression in high-grade squamous intraepithelial neoplasia and squamous cell carcinoma of the cervix. *J. Cancer Res. Ther.* **2023**, *19*, S260–S267. [[CrossRef](#)] [[PubMed](#)]
26. Meyerson, M.; Harlow, E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol. Cell. Biol.* **1994**, *14*, 2077–2086. [[CrossRef](#)]
27. Huang, C.; Chen, Z.; He, Y.; He, Z.; Ban, Z.; Zhu, Y.; Ding, L.; Yang, C.; Jeong, J.; Yuan, W.; et al. EphA2 promotes tumorigenicity of cervical cancer by up-regulating CDK6. *J. Cell. Mol. Med.* **2021**, *25*, 2967–2975. [[CrossRef](#)]
28. Xie, H.; Zhao, Y.; Caramuta, S.; Larsson, C.; Lui, W.-O. miR-205 Expression Promotes Cell Proliferation and Migration of Human Cervical Cancer Cells. *PLoS ONE* **2012**, *7*, e46990. [[CrossRef](#)]
29. Wong, Y.; Cheung, T.; Tsao, G.S.; Lo, K.W.; Yim, S.; Wang, V.W.; Heung, M.M.; Chan, S.C.; Chan, L.K.; Ho, T.W.; et al. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. *Int. J. Cancer* **2006**, *118*, 2461–2469. [[CrossRef](#)] [[PubMed](#)]
30. Fernandez-Avila, L.; Castro-Amaya, A.M.; Molina-Pineda, A.; Hernández-Gutiérrez, R.; Jave-Suarez, L.F.; Aguilar-Lemarrroy, A. The Value of CXCL1, CXCL2, CXCL3, and CXCL8 as Potential Prognosis Markers in Cervical Cancer: Evidence of E6/E7 from HPV16 and 18 in Chemokines Regulation. *Biomedicines* **2023**, *11*, 2655. [[CrossRef](#)] [[PubMed](#)]
31. Kim, J.W.; Kim, Y.T.; Kim, D.K.; Song, C.H.; Lee, J.W. Expression of Epidermal Growth Factor Receptor in Carcinoma of the Cervix. *Gynecol. Oncol.* **1996**, *60*, 283–287. [[CrossRef](#)]
32. Shen, L.; Shui, Y.; Wang, X.; Sheng, L.; Yang, Z.; Xue, D.; Wei, Q. EGFR and HER2 expression in primary cervical cancers and corresponding lymph node metastases: Implications for targeted radiotherapy. *BMC Cancer* **2008**, *8*, 232. [[CrossRef](#)]
33. Zhang, L.; Chen, Q.; Hu, J.; Chen, Y.; Liu, C.; Xu, C. Expression of HIF-2 α and VEGF in Cervical Squamous Cell Carcinoma and Its Clinical Significance. *BioMed Res. Int.* **2016**, *2016*, 5631935. [[CrossRef](#)]
34. Fujimoto, J.; Ichigo, S.; Hori, M.; Hirose, R.; Sakaguchi, H.; Tamaya, T. Expression of basic fibroblast growth factor and its mRNA in advanced uterine cervical cancers. *Cancer Lett.* **1997**, *111*, 21–26. [[CrossRef](#)]
35. Yee, G.P.C.; De Souza, P.L.; Khachigian, L.M. Reducing invasion potential of cervical cancer cells via targeted knockdown of c-Jun. *J. Clin. Oncol.* **2013**, *31*, e22005. [[CrossRef](#)]
36. Prusty, B.K.; Das, B.C. Constitutive activation of transcription factor AP-1 in cervical cancer and suppression of human papillomavirus (HPV) transcription and AP-1 activity in HeLa cells by curcumin. *Int. J. Cancer* **2004**, *113*, 951–960. [[CrossRef](#)] [[PubMed](#)]
37. Liao, H.; Zhang, L.; Lu, S.; Li, W.; Dong, W. KIF3 Promotes Proliferation, Migration, and Invasion in Colorectal Cancer via PI3K/AKT/mTOR Signaling Pathway. *Front. Genet.* **2022**, *13*, 848926. [[CrossRef](#)] [[PubMed](#)]
38. Halle, M.K.; Sødal, M.; Forss, D.; Engerud, H.; Woie, K.; Lura, N.G.; Wagner-Larsen, K.S.; Trovik, J.; Bertelsen, B.I.; Haldorsen, I.S.; et al. A 10-gene prognostic signature points to LIMCH1 and HLA-DQB1 as important players in aggressive cervical cancer disease. *Br. J. Cancer* **2021**, *124*, 1690–1698. [[CrossRef](#)]
39. Wang, L.; Zhong, Y.; Yang, B.; Zhu, Y.; Zhu, X.; Xia, Z.; Xu, J.; Xu, L. LINC00958 facilitates cervical cancer cell proliferation and metastasis by sponging miR-625-5p to upregulate LRRC8E expression. *J. Cell. Biochem.* **2019**, *121*, 2500–2509. [[CrossRef](#)]
40. van de Weerd, B.C.; Medema, R.H. Polo-Like Kinases: A Team in Control of the Division. *Cell Cycle* **2006**, *5*, 853–864. [[CrossRef](#)]
41. Li, J.; Jia, H.; Xie, L.; Wang, X.; Wang, X.; He, H.; Lin, Y.; Hu, L. Association of Constitutive Nuclear Factor- κ B Activation with Aggressive Aspects and Poor Prognosis in Cervical Cancer. *Int. J. Gynecol. Cancer* **2009**, *19*, 1421–1426. [[CrossRef](#)]
42. Sun, Y.; Zhang, R.; Zhou, S.; Ji, Y. Overexpression of Notch1 is associated with the progression of cervical cancer. *Oncol. Lett.* **2015**, *9*, 2750–2756, Erratum in *Oncol. Lett.* **2021**, *21*, 134. [[CrossRef](#)]
43. Zagouras, P.; Stifani, S.; Blaumueller, C.M.; Carcangiu, M.L.; Artavanis-Tsakonas, S. Alterations in Notch signaling in neoplastic lesions of the human cervix. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6414–6418. [[CrossRef](#)]
44. Kim, H.-S.; Yoon, G.; Ryu, J.-Y.; Cho, Y.-J.; Choi, J.-J.; Lee, Y.-Y.; Kim, T.-J.; Choi, C.-H.; Song, S.-Y.; Kim, B.-G.; et al. Sphingosine kinase 1 is a reliable prognostic factor and a novel therapeutic target for uterine cervical cancer. *Oncotarget* **2015**, *6*, 26746–26756. [[CrossRef](#)] [[PubMed](#)]
45. Wang, Y.; Gao, Y.; Cheng, H.; Yang, G.; Tan, W. Stanniocalcin 2 promotes cell proliferation and cisplatin resistance in cervical cancer. *Biochem. Biophys. Res. Commun.* **2015**, *466*, 362–368. [[CrossRef](#)]

46. Caffarel, M.M.; Chattopadhyay, A.; Araujo, A.M.; Bauer, J.; Scarpini, C.G.; Coleman, N. Tissue transglutaminase mediates the pro-malignant effects of oncostatin M receptor over-expression in cervical squamous cell carcinoma. *J. Pathol.* **2013**, *231*, 168–179. [[CrossRef](#)]
47. Liu, S.; Meng, F.; Ding, J.; Ji, H.; Lin, M.; Zhu, J.; Ma, R. High TRIM44 expression as a valuable biomarker for diagnosis and prognosis in cervical cancer. *Biosci. Rep.* **2019**, *39*. [[CrossRef](#)]
48. Zhang, X.; Qin, G.; Chen, G.; Li, T.; Gao, L.; Huang, L.; Zhang, Y.; Ouyang, K.; Wang, Y.; Pang, Y.; et al. Variants in TRIM44 Cause Aniridia by Impairing PAX6 Expression. *Hum. Mutat.* **2015**, *36*, 1164–1167. [[CrossRef](#)]
49. van Rijssel, J.; van Buul, J.D. The many faces of the guanine-nucleotide exchange factor trio. *Cell Adhes. Migr.* **2012**, *6*, 482–487. [[CrossRef](#)] [[PubMed](#)]
50. Hou, C.; Zhuang, Z.; Deng, X.; Xu, Y.; Zhang, P.; Zhu, L. Knockdown of Trio by CRISPR/Cas9 suppresses migration and invasion of cervical cancer cells. *Oncol. Rep.* **2017**, *39*, 795–801. [[CrossRef](#)]
51. Lin, L.; Liu, Y.; Zhao, W.; Sun, B.; Chen, Q. Wnt5A expression is associated with the tumor metastasis and clinical survival in cervical cancer. *Int. J. Clin. Exp. Pathol.* **2014**, *7*, 6072–6078.
52. Song, Z.; Lin, Y.; Ye, X.; Feng, C.; Lu, Y.; Yang, G.; Dong, C. Expression of IL-1 α and IL-6 is Associated with Progression and Prognosis of Human Cervical Cancer. *Med. Sci. Monit.* **2016**, *22*, 4475–4481. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.