



OPEN

## Joint changes in RNA, RNA polymerase II, and promoter activity through the cell cycle identify non-coding RNAs involved in proliferation

Siv Anita Hegre<sup>1,7</sup>, Helle Samdal<sup>2,7</sup>, Antonin Klima<sup>2</sup>, Endre B. Stovner<sup>2,3</sup>, Kristin G. Nørsett<sup>2,6</sup>, Nina Beate Liabakk<sup>1</sup>, Lene Christin Olsen<sup>1,4,5</sup>, Konika Chawla<sup>1,4</sup>, Per Arne Aas<sup>1</sup> & Pål Sætrum<sup>1,2,3,4</sup>✉

Proper regulation of the cell cycle is necessary for normal growth and development of all organisms. Conversely, altered cell cycle regulation often underlies proliferative diseases such as cancer. Long non-coding RNAs (lncRNAs) are recognized as important regulators of gene expression and are often found dysregulated in diseases, including cancers. However, identifying lncRNAs with cell cycle functions is challenging due to their often low and cell-type specific expression. We present a highly effective method that analyses changes in promoter activity, transcription, and RNA levels for identifying genes enriched for cell cycle functions. Specifically, by combining RNA sequencing with ChIP sequencing through the cell cycle of synchronized human keratinocytes, we identified 1009 genes with cell cycle-dependent expression and correlated changes in RNA polymerase II occupancy or promoter activity as measured by histone 3 lysine 4 trimethylation (H3K4me3). These genes were highly enriched for genes with known cell cycle functions and included 57 lncRNAs. We selected four of these lncRNAs—*SNHG26*, *EMSLR*, *ZFAS1*, and *EPB41L4A-AS1*—for further experimental validation and found that knockdown of each of the four lncRNAs affected cell cycle phase distributions and reduced proliferation in multiple cell lines. These results show that many genes with cell cycle functions have concomitant cell-cycle dependent changes in promoter activity, transcription, and RNA levels and support that our multi-omics method is well suited for identifying lncRNAs involved in the cell cycle.

Genome-wide gene expression studies have revealed that several genes are regulated in a cell cycle-specific manner<sup>1–5</sup>. Many of these genes are involved in basic cellular processes, such as cell cycle control, DNA repair, DNA replication, and chromosome segregation<sup>2,6</sup>. One example is the cyclins, which in complex with cyclin-dependent kinases (CDKs), control cell cycle progression. The cyclins are periodically expressed throughout the cell cycle; the E-type cyclins *CCNE1* and *CCNE2* are G1/S-specific<sup>2,7,8</sup>, whereas the B-type cyclins *CCNB1* and *CCNB2* are G2/M-specific<sup>2,9,10</sup>.

Long non-coding RNAs (lncRNAs) have emerged as important regulators of gene expression at the epigenetic, transcriptional, and translational level, and are recognized as key modulators in several cancers as well as neurological, autoimmune, and cardiovascular diseases. lncRNAs are more than 200 nucleotides long with little or no protein coding potential, and they generally have a more cell type-specific expression pattern compared to

<sup>1</sup>Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. <sup>2</sup>Department of Computer Science, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. <sup>3</sup>K.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. <sup>4</sup>Bioinformatics Core Facility-BioCore, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. <sup>5</sup>The Central Norway Regional Health Authority, St. Olavs Hospital HF, Trondheim, Norway. <sup>6</sup>Present address: Department of Biomedical Laboratory Science, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway. <sup>7</sup>These authors contributed equally: Siv Anita Hegre and Helle Samdal. ✉email: pal.satrom@ntnu.no

mRNAs<sup>11</sup>. lncRNAs are classified into five main categories according to where they are encoded in the genome in relation to mRNAs: sense, antisense, bi-directional, intergenic, and intronic. They are able to regulate the gene expression at the transcriptional level by acting as signals, guides, scaffolds, or decoys<sup>12</sup>. Most lncRNAs are transcribed by RNA Polymerase II (Pol II) and are poly-adenylated and 5'-capped like mRNAs<sup>13</sup>. A curated knowledgebase of lncRNAs from existing databases and published literature indicates that there are more than 268,000 human lncRNA transcripts, and only a few of them have known functional roles<sup>14</sup>.

Several lncRNAs are involved in the cell cycle, possibly through the regulation of other well-known cell cycle regulators like the cyclins, p53, retinoblastoma protein (RB), CDKs, and the CDK inhibitors<sup>15</sup>. Many lncRNAs have a cyclic expression profile, and the majority peak during the G1 phase. The cell cycle phase-specific expression of lncRNAs may be consistent with their phase-specific function. For instance, the expression of Metastasis Associated Lung Adenocarcinoma Transcript 1 (*MALAT1*) is highest during G1/S transition and mitosis, which is consistent with its independent function during these phases<sup>16</sup>. A known cell cycle-associated lncRNA is the growth arrest-specific transcript 5 (*GAS5*), which is found downregulated in several cancers where its overexpression results in cell cycle arrest or apoptosis. In prostate cancer *GAS5* functions as a tumor suppressor that inhibits proliferation by targeting the CDK inhibitor p27<sup>17</sup>. Another known cell cycle-associated lncRNA is the zinc finger antisense 1 (*ZFAS1*). *ZFAS1* can act as an oncogene in some cancer types<sup>18,19</sup>, and as a tumor suppressor in others<sup>20</sup>, possibly depending on both type of tissue and state of progression.

The function of the majority of lncRNAs is still unknown, as only about 500–1500 have been functionally characterized. As lncRNAs can bind DNA, RNAs, and proteins, they can regulate the gene expression at the transcriptional, post-transcriptional, and epigenetic level. They can serve as signaling molecules, as sponges by binding microRNAs (miRNAs), and inhibit miRNA-induced degradation of mRNAs, as guides by recruiting transcription regulators, or act as scaffolds by binding proteins to regulate gene expression<sup>21</sup>. As RNA molecules, lncRNAs need a physical proximity to exert their function. Thus, the subcellular localization of lncRNAs provides important information regarding their potential function<sup>22</sup>. For example, nuclear enriched lncRNAs can act as transcriptional and epigenetic regulators, but they are unlikely to have any coding potential since translation occurs in the cytoplasm. In general, most lncRNAs demonstrate a stronger nuclear localization than mRNAs<sup>23</sup>. Moreover, a higher cell type specificity means that targeting lncRNAs supposedly have less side-effects than targeting protein coding genes<sup>24</sup>.

Some lncRNAs can regulate gene transcription by modulating histone modifications<sup>25</sup>, although little is known about how lncRNAs are transcriptionally regulated<sup>26</sup>. Histone modifications such as H3K4me3 and H3K27me3 are considered key epigenetic regulators of transcription. H3K4me3 is a mark of actively transcribed genes and H3K27me3 is associated with silenced genes<sup>27</sup>. Previous studies have demonstrated that RNA sequencing (RNA-seq) combined with chromatin immunoprecipitation sequencing (ChIP-seq) are useful for detecting transcriptional fluctuations by correlating gene expression with changes in histone modifications<sup>28,29</sup>. A study from Wan et al combined RNA-seq with ChIP-seq and identified differential peaks for H3K4me3 and H3K27me3 around the promoter area and at enhancer regions of differentially expressed lncRNAs in an Alzheimer's disease mouse model compared to control, suggesting that most of these lncRNA genes were transcriptionally regulated by histone modifications<sup>30</sup>. Since the majority of lncRNAs are spatially and temporally regulated and expressed, ChIP-seq is a sensitive method for capturing these changes by identifying enriched peak regions of histone modifications and other transcriptional regulators<sup>26</sup>.

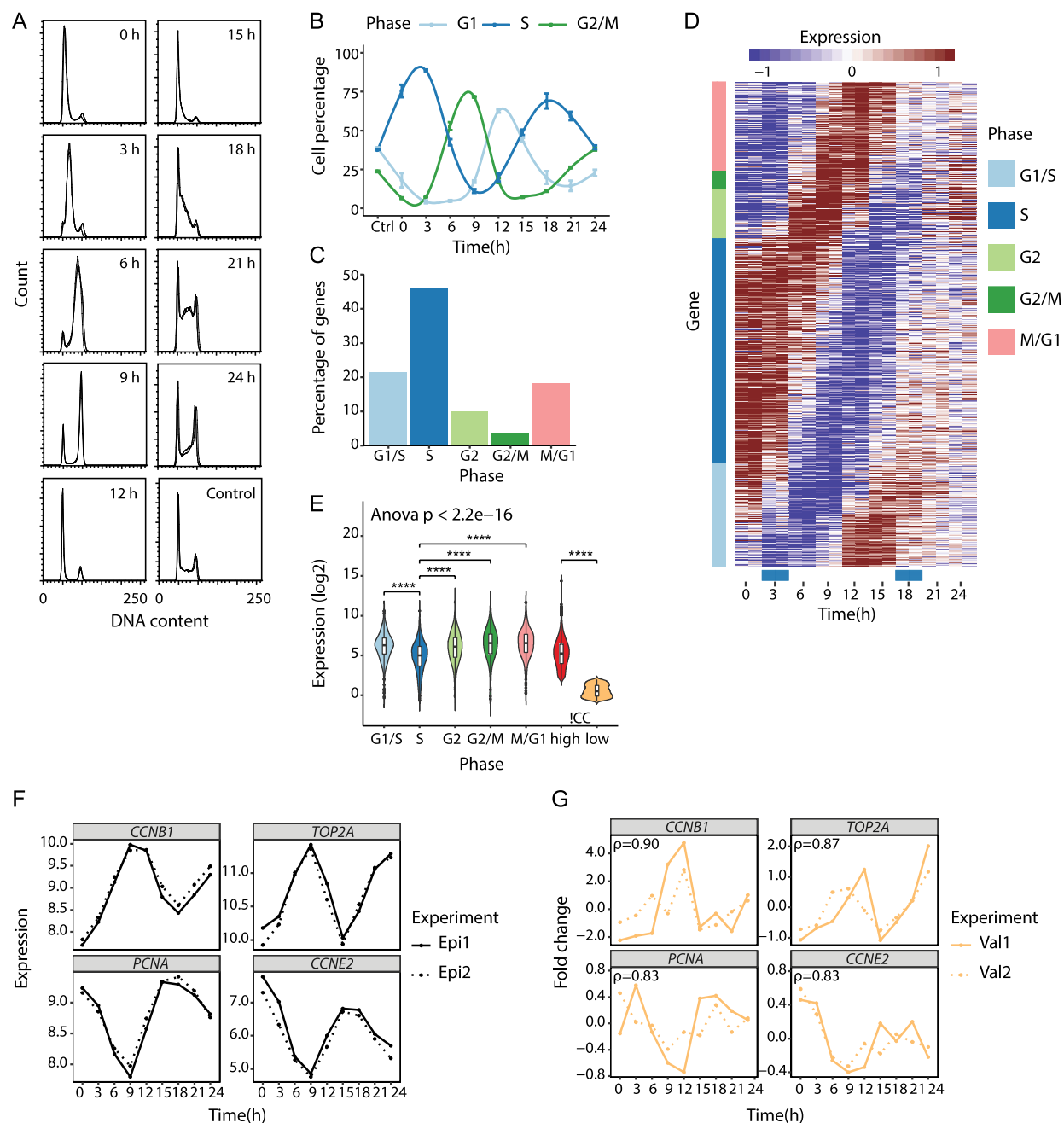
Building on our previous work, which identified protein coding genes with tissue-specific cell cycle-dependent expression<sup>1</sup>, we set out to identify lncRNAs with cell cycle-dependent expression and potential cell cycle functions. By combining RNA-seq and Pol II, H3K4me3, and H3K27me3 ChIP-seq data from synchronized HaCaT cells, we identified genes where expression and ChIP-seq signal were correlated and varied depending on the cell cycle. Genes with high correlation to Pol II or H3K4me3 were strongly enriched for cell cycle functions. From the RNA-seq data we identified 99 lncRNAs with cell cycle-dependent expression profiles; 57 of these had highly correlated Pol II or H3K4me3 signals. We selected four lncRNAs for further functional characterization and showed that knockdown of these lncRNAs affected cell cycle phase distributions and reduced cell proliferation in multiple cell lines.

## Results

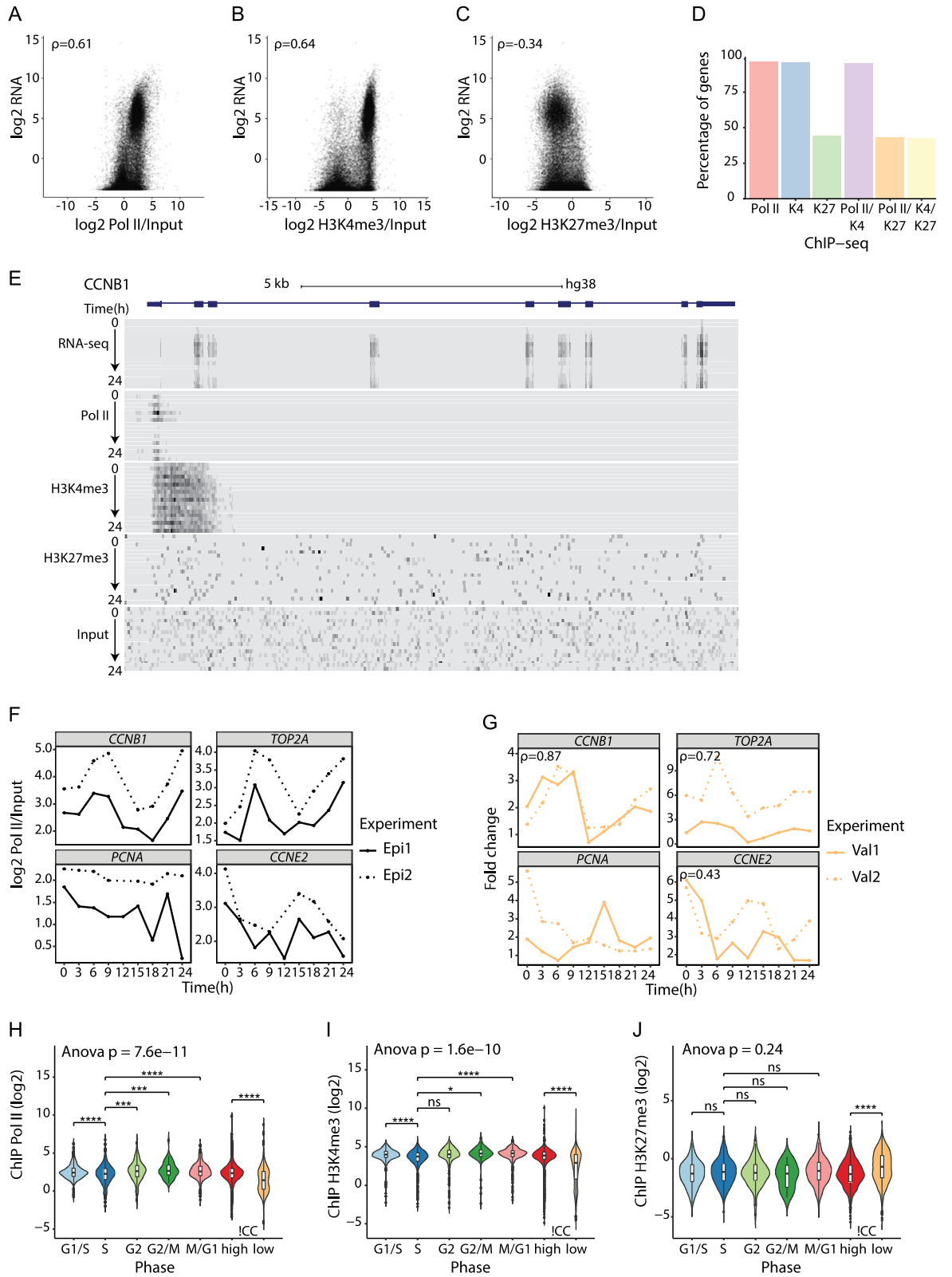
**Total RNA sequencing of HaCaT cells identifies cell cycle genes.** Our group previously published a microarray-based study of cell cycle synchronized HaCaT cells identifying a set of genes with cell cycle-dependent expression and strong enrichment for known cell cycle functions<sup>1</sup>. Although our microarray-based study identified several genes with significant periodic expression patterns during the cell cycle, the microarrays precluded the detection of non-coding RNA (ncRNA) transcripts. To identify ncRNAs that are differentially expressed during cell cycle, we therefore set out to do total RNA-seq on synchronized HaCaT cells.

To study cell cycle regulated genes, it is essential to obtain a proper cell synchronization. In two independent experiments (Epi1, Epi2), we used a double thymidine block to arrest and subsequently release HaCaT cells at the G1/S transition, and collected cells every third hour for 24 h, covering approximately two cycles of DNA replication (Fig. 1A,B). Flow cytometry analyses of the cells' DNA content showed that the distributions of cells through the cell cycle were reproducible between the two independent experiments (Fig. 1A). Upon release from thymidine block, approximately 90% of the cells progressed into S phase and continued through the cell cycle (Fig. 1B). Cells gradually lost synchrony, such that 70% instead of 90% of the cells synchronously re-entered the second S phase. Based on these results, we considered the HaCaT cells to be effectively synchronized and proceeded with gene expression profiling of the two experiments by using total RNA-seq.

Genes that had a cell cycle profile were identified using PLS regression, as described in<sup>1</sup>. We identified 1803 genes with significant periodic expression patterns during the HaCaT cell cycle. These genes are referred to as



**Figure 1.** Total RNA-seq of synchronized HaCaT cells identifies cyclic gene expression patterns. **(A)** HaCaT cells were synchronized by double thymidine block in two independent experiments (Epi1, Epi2) and cell synchrony was monitored by flow cytometry of propidium iodide-stained cells. The figure shows superimposed DNA content profiles of the two replicate experiments for each time point. Horizontal axes show DNA content (arbitrary units) and vertical axes show the number of cells with the corresponding DNA content. Control is unsynchronized cells. **(B)** Percentage of cells assigned to G1, S, and G2/M phases for each of the time points analyzed. Values and error bars are averages and standard deviations ( $n = 2$ ). **(C)** Percentage of cell cycle genes assigned to G1/S (21.5%), S (46.3%), G2 (10.1%), G2/M (3.8%), and M/G1 (18.2%) phases. **(D)** Heatmap showing the expression changes of the cell cycle genes relative to their median expression. Colour bars in the gene margin (y axis) show the genes' assigned cell cycle phase; blue bars above the time points (x axis) show the time points having the highest percentage of S phase cells. **(E)** Distribution of the genes' average RNA-seq expression for cell cycle genes in G1/S ( $n = 388$ ), S ( $n = 835$ ), G2 ( $n = 183$ ), G2/M ( $n = 68$ ), and M/G1 ( $n = 329$ ) phases. \*\*\*\* $p \leq 0.0001$  (Welch's  $t$ -test,  $p$ -values for each phase group against S phase were Bonferroni corrected for multiple testing). **(F)** RNA-seq profiles for cell cycle genes *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A*. **(G)** Relative expression profiles for *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A* as measured by RT-qPCR in two new biological replicates (Val1, Val2; Figure S3). Spearman's correlation coefficients ( $\rho$ ) were calculated based on the mean RNA-seq expression per time point (Epi1, Epi2) and mean RT-qPCR fold change per time point (Val1, Val2).



**◀Figure 2.** ChIP-seq of synchronized HaCaT cells identifies dynamic H3K4me3, H3K27me3, and Pol II changes during cell cycle. (A–C) Average RNA-seq expression (genes expressed > 0) plotted against average (A) Pol II (n = 30,744), (B) H3K4me3 (n = 30,729), and (C) H3K27me3 (n = 30,732) ChIP-seq signal normalized against input. Values are Spearman's correlation coefficients ( $\rho$ ). (D) Percentage of cell cycle gene promoters containing Pol II, H3K4me3 (K4), or H3K27me3 (K27) marks and combined Pol II/K4, Pol II/K27, or bivalent K4/K27 marks. (E) The UCSC Genome browser (GRCh38/hg38) view of RNA-seq, Pol II, H3K4me3, H3K27me3, and input ChIP-seq data at the *CCNB1* (NM\_031966) gene locus. (F) Pol II ChIP-seq profiles for cell cycle genes *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A*. (G) Biological validation of Pol II ChIP-seq data for *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A*. Spearman's correlation coefficients ( $\rho$ ) were calculated based on the mean Pol II ChIP-seq signal per time point (Epi1, Epi2) and mean Pol II ChIP-qPCR fold change per time point from the new biological replicates (Val1, Val2; Figure S3). (H–J) Distribution of the genes' average (H) Pol II (n = 1737), (I) H3K4me3 (n = 1726), and (J) H3K27me3 (n = 802) ChIP-seq signal (normalized against input) for cell cycle genes in G1/S, S, G2, G2/M, and M/G1 phases. ns  $p > 0.05$ , \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\* $p \leq 0.001$ , \*\*\*\* $p \leq 0.0001$  (Welch's *t*-test, *p*-values for each phase group against S phase were Bonferroni corrected for multiple testing).

cell cycle genes. Using the Ensembl BioMart tool (human genes dataset GRCh38.p13), we found 1692 protein coding genes and 108 ncRNAs (Supplementary Dataset 1). Three genes were not detected in BioMart. We used existing annotations<sup>2</sup> to subdivide the genes into five main groups that represent G1/S, S, G2, G2/M, and M/G1 phases of the cell cycle. Of the 1803 cell cycle genes, 835 were assigned to S phase, whereas only 68 were assigned to G2/M phase (Fig. 1C); the genes' cyclic expression patterns are presented in Fig. 1D. Examining the RNA-seq expression for the 1803 cell cycle genes by their different cell cycle phases, we observed a significant difference in gene expression between different phases (analysis of variance (ANOVA)  $p$ -value <  $2e-16$ ) and that S phase genes had a significantly lower expression level than the genes expressed in the other phases (Fig. 1E). This was in accordance with our previous findings<sup>1</sup>. When examining the genes with comparable expression levels to the cell cycle genes (that is, mean expression  $\geq$  the cell cycle gene with lowest mean expression) but no significant cell cycle-dependent expression pattern (!CC genes), we found that these genes had a bimodal distribution of expression levels (Figure S1). We therefore divided this group of genes into genes with high expression level (!CC\_high, n = 9259) and genes with low expression level (!CC\_low, n = 3003) as reference sets for comparison against the cell cycle genes in further analysis.

To further examine cell cycle profiles at the single gene level, we selected four genes with cell cycle-specific expression patterns and known functions in the cell cycle: Proliferating cell nuclear antigen (*PCNA*), Cyclin E2 (*CCNE2*), DNA topoisomerase 2-alpha (*TOP2A*), and Cyclin B1 (*CCNB1*). *PCNA* is essential for DNA replication whereas *CCNE2* plays a role in the G1/S transition of the cell cycle. Thus, both *PCNA* and *CCNE2* are G1/S-specific genes. *TOP2A* is involved in processes such as chromosome condensation and chromatid separation whereas *CCNB1* is a regulatory protein involved in mitosis. Thus, both *TOP2A* and *CCNB1* are G2/M-specific genes. For all four genes, our RNA-seq data corresponded with their previously reported cell cycle-dependent expression (Fig. 1F). Technical validation by RT-qPCR confirmed the RNA-seq expression patterns (Figure S2). As a biological validation of the RNA-seq data, we did two new double thymidine block cell cycle synchronization experiments (Val1, Val2) in HaCaT cells (Figure S3). Analyses by RT-qPCR showed a high correlation between the original RNA-seq profiles and the new validation experiments (Fig. 1G). Thus, both the biological and technical validation indicated that gene expression profiling by total RNA-seq of HaCaT cells could successfully identify cell cycle genes.

**ChIP sequencing maps dynamic transcriptional responses in HaCaT cell cycle.** By gene expression profiling we identified a set of genes with cell cycle-dependent expression patterns in HaCaT cells. We wanted to further characterize the dynamic transcriptional response of these genes during the cell cycle. Specifically, we asked whether the genes' expression changes were accompanied by similar cell cycle-dependent changes in Pol II occupancy at the genes' transcription start sites (TSSs). Moreover, we asked whether the histone modifications H3K4me3, associated with actively transcribed genes, and H3K27me3, associated with silenced genes, also changed dynamically with gene expression changes. Using ChIP-seq we measured and quantified genome-wide Pol II, H3K4me3, and H3K27me3 occupancy during two independent synchronization experiments (Epi1, Epi2).

First, we performed a sanity check of our sequencing data, by correlating gene expression levels with histone modification patterns and Pol II occupancy in TSS regions for all genes expressed in our RNA-seq data. As Pol II and H3K4me3 both are linked to actively transcribed genes, we expected a positive correlation with gene expression. In contrast, we expected a negative correlation with gene expression for the H3K27me3 modification. Indeed, we found that enhanced Pol II and H3K4me3 signals correlated well with highly expressed genes (Spearman's correlation coefficient  $\rho = 0.61$  and  $\rho = 0.64$ , respectively), whereas there was a negative correlation between gene expression and H3K27me3 signal ( $\rho = -0.34$ ) (Fig. 2A–C). Similarly, dividing the genes into quantiles based on their RNA-seq expression showed that highly expressed genes had the highest Pol II and H3K4me3 signals, whereas genes that were expressed at a low level or not expressed, had low levels of Pol II and H3K4me3 (Figure S4). The opposite was observed for H3K27me3, where genes that were not expressed or expressed at a low level had the highest H3K27me3 signal.

When investigating the Pol II signal, we observed low signal at time 12 h in the Epi2 experiment (Figure S4), but western blot analysis of the global Pol II protein level throughout the cell cycle indicated no differences in Pol II protein levels between different time points in the cell cycle or between the two independent synchronization experiments (Figure S5A). Therefore, we concluded that this was not a biological but a technical issue

and decided to exclude the sample (12 h, Epi2) in the following analysis. We also investigated global H3K4me3 and H3K27me3 modification levels and found no changes during cell cycle or between the two independent synchronization experiments (Figure S5B).

As cell cycle control is a central housekeeping function, we expected the cell cycle genes we identified by RNA-seq to be enriched for H3K4me3 within their promoters, as this epigenetic mark is associated with actively transcribed genes. We also expected the genes to be enriched for Pol II indicating active promoters. Indeed, 96% of the cell cycle genes had promoter regions containing H3K4me3 or Pol II marks. Consistent with these genes being highly expressed, only 45% of the cell cycle gene promoters contained H3K27me3 marks (Fig. 2D). In 95% of the gene promoters we found combined Pol II/H3K4me3 marks, whereas 43% of the gene promoters contained combined Pol II/H3K27me3 or bivalent H3K4me3/H3K27me3 marks.

To further inspect the data quality, we visualized RNA-seq and ChIP-seq data in the UCSC Genome browser. *CCNB1* was one of the well-known cell cycle genes we identified, and we observed RNA-seq reads for all exons and Pol II ChIP-seq signals were captured around the TSS (Fig. 2E). Additionally, H3K4me3 was highly enriched at this active promoter near TSS, whereas H3K27me3 signals were at the level of the background signal (input ChIP-seq). In contrast, Myelin transcription factor 1 (*MYT1*) is a transcription factor involved in the development of the nervous system, and this gene was not expressed in our HaCaT RNA-seq data (Figure S6). The H3K27me3 histone modification was captured at the *MYT1* locus, consistent with this gene not being expressed. Pol II and H3K4me3 signals were observed, but not higher than the level of the background signal.

To characterize the cell cycle ChIP-seq profiles, we first focused on the four cell cycle genes, *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A*. The Pol II ChIP-seq profiles for *CCNB1* and *TOP2A* indicated these genes had the highest Pol II signal and active transcription at 9 and 24 h (Fig. 2F), which is consistent with both genes having increased expression in the G2/M phase of the cell cycle (Fig. 1F). Additionally, H3K4me3 signals for *CCNB1* and *TOP2A* showed the same cyclic patterns peaking in G2/M phase (Figure S7A). *CCNE2* and *PCNA* are both G1/S-specific genes, but only *CCNE2* showed the expected Pol II (Fig. 2F) and H3K4me3 (Figure S7A) profiles peaking at 0–3 and 15–18 h. As *PCNA* displayed inconsistent RNA-seq and ChIP-seq profiles, we inspected the sequencing data for the *PCNA* locus in the UCSC Genome browser (Figure S8). Interestingly, *PCNA* has two different transcript variants separated by 6,664 bp between their TSSs. As illustrated in Figure S8, RNA-seq reads were from the short transcript variant (NM\_182649) with a cyclic expression pattern peaking in the G1/S phase. In the ChIP-seq data analysis, the longest transcript variant is always selected (see “Methods”). Thus, the ChIP-seq signals we observed for *PCNA* were the weak signals from the longest transcript variant (NM\_002592). This explains the inconsistency between RNA-seq and ChIP-seq data at the *PCNA* locus, and could also be the case for other genes with multiple TSSs separated by more than 5000 bp. The H3K27me3 ChIP-seq profiles for *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A* showed negative H3K27me3 signal values, which are consistent with these genes being actively transcribed and highly expressed (Figure S7B).

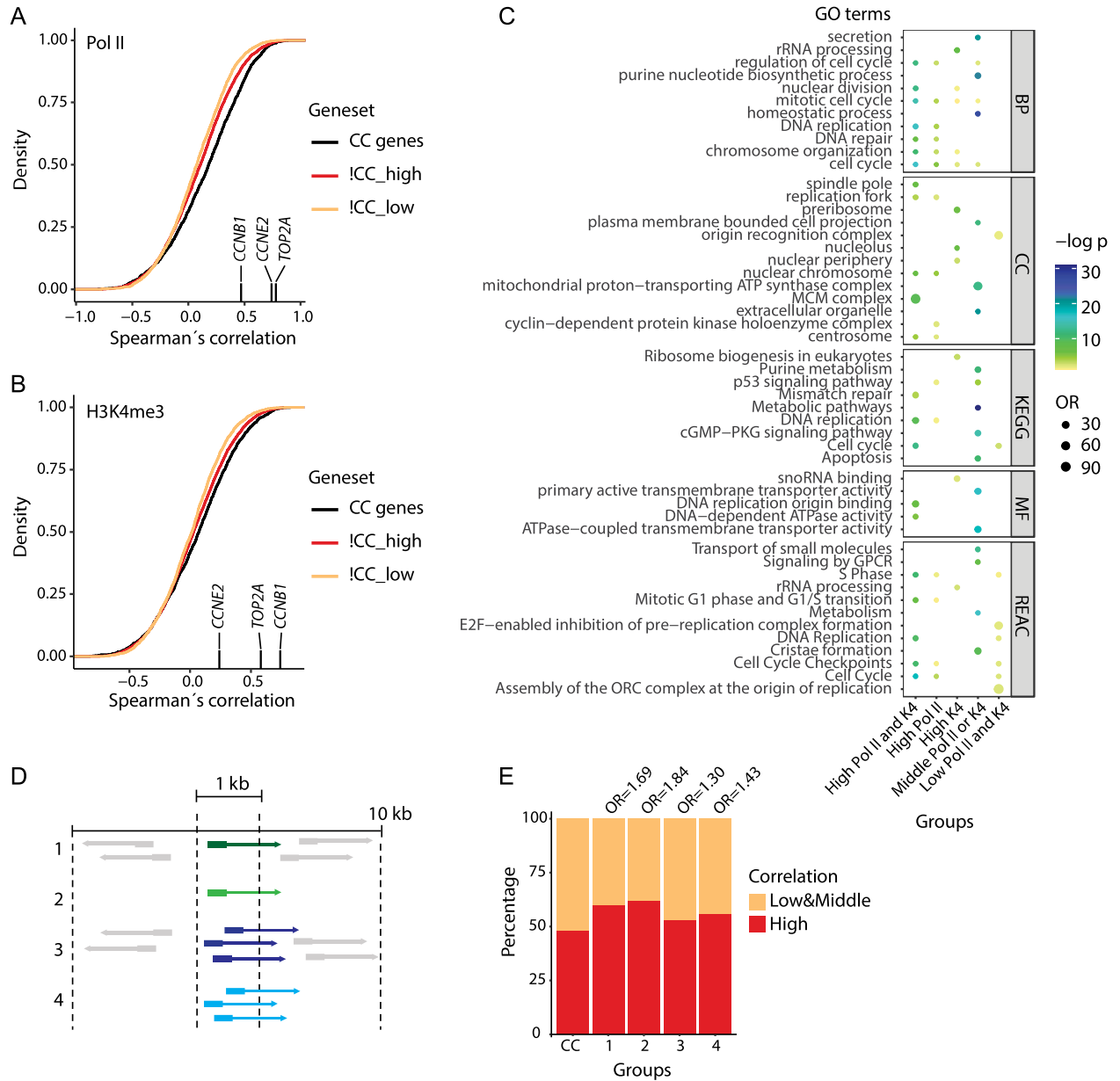
To validate the Pol II ChIP-seq data for *CCNB1*, *CCNE2*, *PCNA*, and *TOP2A* we did Pol II ChIP-qPCR from the two new cell cycle synchronization experiments (Val1, Val2) in HaCaT cells. We designed PCR primers against the TSS region for these genes (Table S1); for *PCNA* we designed primers against the short transcript variant (NM\_182649). For *CCNB1*, *TOP2A*, and *CCNE2* we found good correspondence between the original Pol II ChIP-seq experiments and the Pol II ChIP-qPCR from the new validation experiments (Fig. 2G). Moreover, the Pol II ChIP-qPCR profile for the short *PCNA* transcript showed increased expression in the G1/S phase (Fig. 2G), which is consistent with the RNA-seq (Fig. 1F) and RT-qPCR (Fig. 1G) expression profiles for *PCNA* peaking in G1/S phase (Figure S9).

Since we observed a significant difference in gene expression between the S phase genes and the genes upregulated in the other cell cycle phases (Fig. 1E), we wanted to examine the distribution of Pol II, H3K4me3, and H3K27me3 ChIP-seq signals during cell cycle. We included cell cycle genes with positive ChIP-signal in at least two time points of the cell cycle. As for gene expression, we observed a significant difference in both Pol II and H3K4me3 signals between different cell cycle phases, with ANOVA *p*-values of 7.6e–11 and 1.6e–10, respectively. And as for gene expression, the S phase genes had significantly lower Pol II and H3K4me3 signals than the genes expressed in the other phases (Fig. 2H,I), with the exception of H3K4me3 signal for genes in G2 phase, which showed no significant difference from S phase genes. As expected, genes with high expression (!CC\_high) showed significantly higher levels of Pol II and H3K4me3 signals than genes with low expression (!CC\_low). There were no significant differences in H3K27me3 signal between different cell cycle phases, and as expected, !CC\_low genes had significantly higher levels of H3K27me3 signal than !CC\_high genes (Fig. 2J).

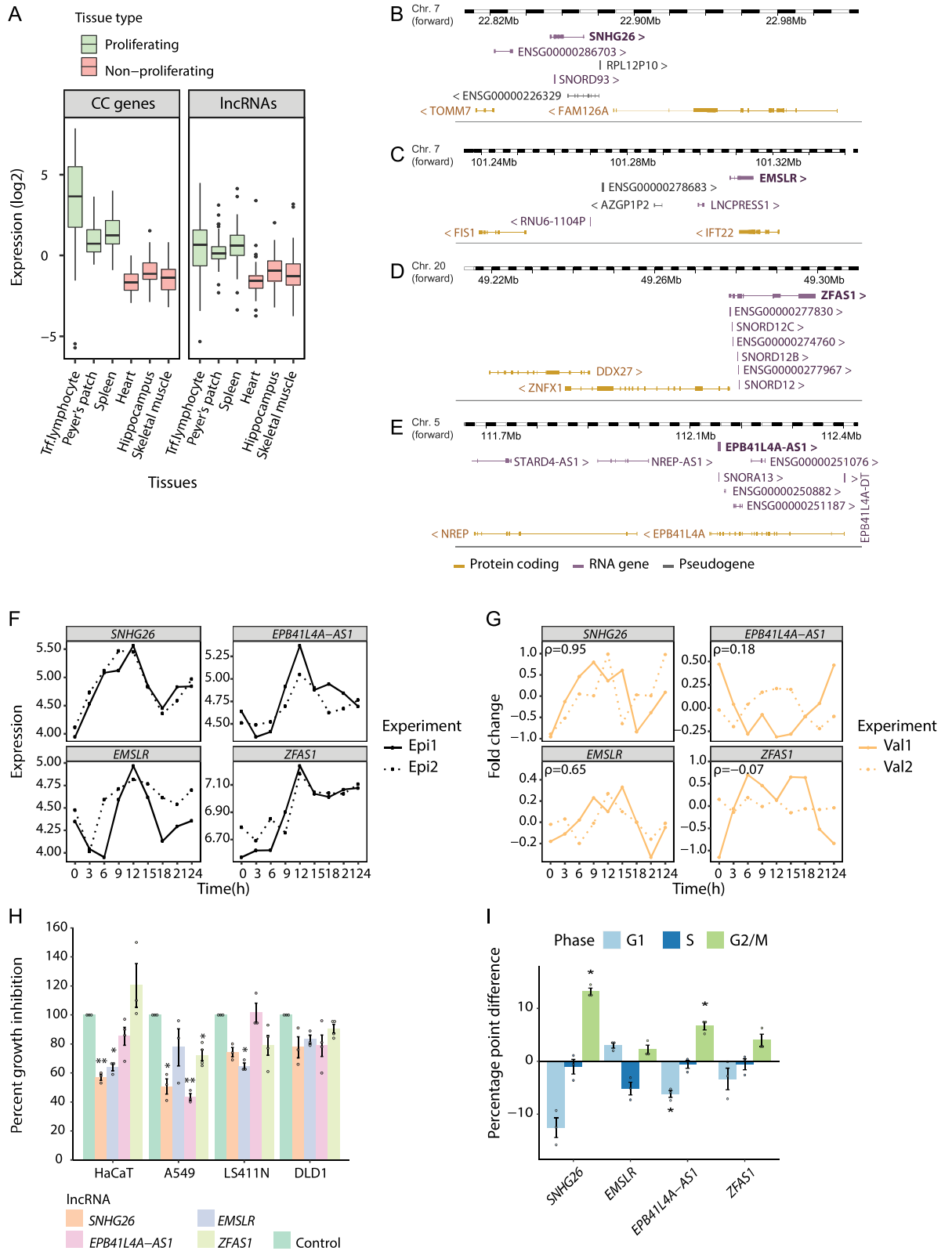
### A set of cell cycle genes is highly correlated with Pol II and H3K4me3 changes and has strong enrichment for cell cycle functions.

Having quantified gene expression and mapped histone modification patterns together with Pol II occupancy during the cell cycle, we did an integrated bioinformatic analysis of RNA-seq and ChIP-seq data for the cell cycle genes. Specifically, we asked to what extent the genes' RNA-seq expression through the cell cycle were correlated with their ChIP-seq data and whether such correlation patterns were related to the genes' functional role in the cell cycle. As a reference, we included the genes without cell cycle-dependent expression profiles, divided by their gene expression levels (!CC\_high, !CC\_low; Figure S1).

We found that all three gene sets (cell cycle (CC) genes, !CC\_high, and !CC\_low) on average showed a significant positive correlation for RNA-seq expression against Pol II ChIP-seq signal (Fig. 3A). Importantly, CC genes showed a significantly higher correlation than !CC genes. The cell cycle genes *CCNB1* (Spearman's  $\rho = 0.47$ ), *CCNE2* ( $\rho = 0.74$ ), and *TOP2A* ( $\rho = 0.78$ ) are examples of genes with high correlation values. We also found that CC genes and !CC\_high genes, but not !CC\_low genes, showed a significant positive correlation for RNA-seq expression against H3K4me3 ChIP-seq signal (Fig. 3B). Again, CC genes showed a significantly higher correlation than !CC genes. Correlation values for *CCNB1*, *CCNE2*, and *TOP2A* were  $\rho = 0.74$ ,  $\rho = 0.24$ ,



**Figure 3.** A set of cell cycle genes correlate with Pol II and H3K4me3 signals. **(A,B)** Cumulative distributions of Spearman's correlation between RNA-seq expression and ChIP-seq signal through the cell cycle for the cell cycle (CC) genes, and other highly (!CC\_high) and lowly (!CC\_low) expressed genes. Marked on the x axis are the correlation values for *CCNB1*, *CCNE2*, and *TOP2A*. **(A; Pol II)** All three gene sets had a significant ( $p$ -value  $< 2.2e-16$ ) positive correlation for RNA-seq against Pol II ChIP-seq. CC genes showed significantly higher correlation than !CC\_high ( $p = 1.49e-11$ ) and !CC\_low ( $p$ -value  $< 2.2e-16$ ). **(B; H3K4me3)** CC genes and !CC\_high had a significant positive correlation for RNA-seq against H3K4me3 ChIP-seq with  $p$ -values of  $3.83e-15$  and  $< 2.2e-16$ , respectively.  $P$ -value for the !CC\_low gene set was not significant ( $p = 0.2287$ ). CC genes showed significantly higher correlation than !CC\_high ( $p = 0.0003859$ ) and !CC\_low ( $p = 1.002e-08$ ). Significant differences were determined by Student's  $t$ -test (unpaired, two-tailed) assuming unequal variances. **(C)** GO analysis for cell cycle genes. The results show GO biological process (BP), cellular component (CC), molecular functions (MF) terms, and KEGG and REACTOME pathways significantly enriched ( $p$ -values  $< 0.05$ ) for cell cycle genes divided in five groups (Figure S11). **(D)** Cell cycle genes are divided into four groups based on the arrangement of TSSs and other nearby genes. Genes in group 1 ( $n = 127$ ) have one single TSS and a TSS from at least one other gene within 10 kb. Genes in group 2 ( $n = 61$ ) have one single TSS and no TSSs from other genes within 10 kb. Genes in group 3 ( $n = 356$ ) have multiple TSSs all within 1 kb and a TSS from at least one other gene within 10 kb. Genes in group 4 ( $n = 152$ ) have multiple TSSs all within 1 kb and no TSSs from other genes within 10 kb. **(E)** Odds ratios from Fisher's exact tests comparing the fraction of highly correlated genes in the groups (d) with all cell cycle genes with Pol II signals ( $n = 1735$ ). Group 1  $p = 0.006$ , group 2  $p = 0.026$ , group 3  $p = 0.028$ , and group 4  $p = 0.041$ .





◀ **Figure 4.** Joint RNA-seq and ChIP-seq analysis identifies lncRNA affecting cell growth and cell cycle progression. (A) Relative tissue expression (log transcript per kilobase million) of known cell cycle genes<sup>2</sup> and our cell cycle lncRNAs in selected tissues from the Genotype-Tissue Expression (GTEx) project. Gene expression values were normalized to relative expression values by subtracting the gene's average expression across all GTEx tissues. (B–E) The genomic loci of *SNHG26*, *EMSLR*, *ZFAS1*, and *EPB41L4A-AS1* from the Ensembl Genome Browser (<http://www.ensembl.org/>). (F) RNA-seq profiles for *SNHG26*, *EMSLR*, *EPB41L4A-AS1*, and *ZFAS1*. (G) Relative expression profiles for *SNHG26*, *EMSLR*, *EPB41L4A-AS1*, and *ZFAS1* as measured by RT-qPCR in two new biological replicates. Spearman's correlation coefficients ( $\rho$ ) were calculated based on the mean RNA-seq expression per time point (Epi1, Epi2) and mean RT-qPCR fold change per time point (Val1, Val2). (H) Effect of siRNA-mediated knockdown of *SNHG26* (siRNA A1; siRNA A2 for A549), *EMSLR* (siRNA R1), *EPB41L4A-AS1* (siRNA E1), and *ZFAS1* (siRNA Z1) on proliferation in four different cell lines. Data are the number of cells following siRNA treatment relative to control-treated cells (percentage of control) as measured by cell counting. Bars and error bars are mean and SEM of three or more independent replicates. (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$  (Welch's *t*-test, Bonferroni corrected for multiple testing). ANOVA *p*-values from the hierarchical, linear model: *SNHG26*  $p = 6.9e-4$ , *EMSLR*  $p = 3.92e-4$ , *EPB41L4A-AS1*  $p = 0.09$ , and *ZFAS1*  $p = 0.371$ ). (I) The distribution of cells in G1, S, and G2/M phases in response to knockdown of *SNHG26* (siRNA A1), *EMSLR* (siRNA R1), *EPB41L4A-AS1* (siRNA E1), and *ZFAS1* (siRNA Z1) in HaCaT cells. Data are the difference in percentages of G1, S, and G2/M cells of siRNA-treated HaCaT cells to those of control-treated HaCaT cells. Bars and error bars are mean and SEM of three independent replicates. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$  (Welch's *t*-test, Bonferroni corrected for multiple testing).

and  $\rho = 0.58$ , respectively. There were no significant differences between the three gene sets for the H3K27me3 modification (Figure S10).

Further, we divided the cell cycle genes in high ( $\rho > 0.2$ ), low ( $\rho < -0.2$ ), and middle ( $\rho > -0.2$  and  $\rho < 0.2$ ) correlated genes, based on the genes' Spearman correlation value. We did a gene ontology (GO) analysis for five different groups (Figure S11); high correlation for both Pol II and H3K4me3 ( $n = 400$ ), high correlation for Pol II only ( $n = 423$ ), high correlation for H3K4me3 only ( $n = 186$ ), middle correlation for Pol II or H3K4me3 ( $n = 580$ ), and low correlation for both Pol II and H3K4me3 ( $n = 125$ ). We identified non-redundant terms among the top 20 significant terms within each GO category. The results showed that genes with high correlation for Pol II and H3K4me3 were specifically enriched for cell cycle-related terms, including cell cycle regulation, DNA replication, and nuclear division (Fig. 3C, Supplementary Dataset 2). Moreover, middle correlated genes were specifically enriched for cell signalling, including p53 signalling, whereas genes with low correlation were specifically enriched for initiation of replication. For H3K27me3, genes with high or middle correlation were weakly enriched for cell cycle functions whereas genes with low correlations were weakly enriched for the ribosome biogenesis pathway (Figure S12; Supplementary Dataset 3).

The combined RNA-seq and ChIP-seq analysis indicated distinct functions for cell cycle genes based on their correlation with Pol II and H3K4me3 signals, as highly correlated genes had strong enrichment for cell cycle functions. Nevertheless, we wondered to what extent the results were affected by ambiguous mapping of ChIP-seq signals. Specifically, we wondered if the correlation was affected by whether the gene had one or multiple annotated TSSs and by whether the gene was isolated or had annotated TSSs for neighboring genes within its TSS region. Thus, we divided the cell cycle genes into four different groups based on the arrangement of TSSs and other nearby genes (Fig. 3D). Genes in group 1 had one single TSS and other genes within 10 kb, whereas genes in group 2 had one single TSS and no other genes within 10 kb. Genes in group 3 had multiple TSSs all within 1 kb and other genes within 10 kb, whereas genes in group 4 had multiple TSSs all within 1 kb and no other genes within 10 kb. Compared with all cell cycle genes (CC), a larger fraction of genes with one TSS (groups 1, 2) were highly correlated with their Pol II signal (Fig. 3E); for genes in groups 3 and 4, the fractions were between those of all CC genes and single TSS genes. Moreover, for genes with clearly defined TSS areas and with no overlap with other genes (group 2 compared to 1 and group 4 compared to 3), there was a slightly larger fraction of highly correlated genes. We found no such differences for H3K4me3 and H3K27me3 modifications (Figures S13 and S14), possibly because these signals on average cover wider genomic regions than do Pol II.

Thus, whereas ambiguous gene annotations could explain low correlation levels for some genes with cell cycle functions, including *PCNA*, genes with cell cycle-dependent expression and correlated Pol II or H3K4me3 changes were strongly enriched for known cell cycle functions. This result suggested that these highly correlated genes are prime candidates for functional follow-up, so we focused on lncRNAs among these genes.

**Combined RNA-seq and ChIP-seq analysis identifies cell cycle-associated lncRNAs.** By combining RNA profiling with ChIP-seq data, we identified 97 cell cycle lncRNAs (Table S2). Similar to a set of 57 well-described cell cycle genes<sup>2</sup>, these lncRNAs had higher expression in proliferating tissues than in non-proliferating tissues (Fig. 4A), further supporting their potential role in proliferation. Thus, the combinational use of RNA-seq and ChIP-seq data enabled us to focus on lncRNAs with strong enrichment for cell cycle functions.

We selected four lncRNAs for further functional characterization. The two lncRNA candidates Small nucleolar RNA host gene 26 (*SNHG26*) and E2F1 mRNA stabilizing lncRNA (*EMSLR*) were chosen as they had high correlation for Pol II and H3K4me3. *SNHG26* is located between the genes *TOMM7* and *FAM126A* on chromosome 7 (Fig. 4B), while *EMSLR* is located at the same chromosome between the genes *FIS1* and *IFT22* (Fig. 4C). The lncRNA ZNF1 Antisense RNA 1 (*ZFAS1*) was identified among the candidates and since it has previously been reported to affect proliferation in different types of cancer<sup>31</sup>, we chose this lncRNA as a candidate. *ZFAS1* is located between the genes *DDX27* and *ZNF1* on chromosome 20 (Fig. 4D). In addition, we chose *EPB41L4A*

Antisense RNA 1 (*EPB41L4A-AS1*) as a candidate, as this lncRNA had similar ChIP-seq characteristics as *ZFAS1* (Figure S15, Table S2). *EPB41L4A-AS1* is located between the genes *NREP* and *EPB41L4A* on chromosome 5 (Fig. 4E). Of these neighboring genes, *FAM126*, *IFT22*, and *ZNF1* had cell cycle-dependent expression (Figure S16). Tissue-specific expression for the four lncRNA candidates confirmed that all were highly expressed in proliferating tissues (Figure S17). All candidates were predicted to have a low protein coding probability, as assessed by the Coding Potential Assessment Tool<sup>32</sup> (Table S3).

Since lncRNAs are dependent on proximity to their target molecules to exert their function, the subcellular localization can provide important information about their biological role. We used the lncAtlas online tool (<https://lncatlas.crg.eu/>) to determine the subcellular localization of the four candidate lncRNAs. The lncAtlas displays the subcellular localization of lncRNAs based on a relative concentration index (RCI = concentration of a gene, per unit mass of RNA between cytoplasm and nucleus) derived from RNA-seq data sets from available cell lines and cellular compartments from Human Gencode v24<sup>33</sup>. Based on RCI from several cell lines, the four lncRNA candidates were mainly localized in either the nucleus or cytoplasm. *SNHG26* was enriched in the nucleus with an average RCI of  $-2.03$ , whereas *EMSLR*, *EPB41L4A-AS1*, and *ZFAS1* had the highest level in cytoplasm with an average RCI of 1.82, 1.12, and 1.43, respectively (Figure S18).

The RNA-seq data showed that all four lncRNA candidates had cell cycle-dependent expression profiles that peaked 12 h after release (Fig. 4F). At 12 h, the majority of the cells were in G1 phase (Fig. 1B), but when taking into account each gene's overall profile compared with those of known cell cycle genes, the genes were assigned to G2/M (*SNHG26*), M/G1 (*EMSLR*, *EPB41L4A-AS1*), and G1/S (*ZFAS1*). These profiles were confirmed by technical validation by RT-qPCR (Figure S19), but only the profiles for *SNHG26* and *EMSLR* could be validated in the two new independent double thymidine block synchronization experiments (Val1, Val2) in HaCaT cells (Fig. 4G). The phase-dependent expression profile of *SNHG26* was further confirmed by FACS analysis of HaCaT cells, showing a significantly reduced expression in S phase compared to G1 phase (Figure S20). Whereas *EMSLR* and *EPB41L4A-AS1* also showed reduced expression in S phase, these differences were not significant ( $p$ -values = 0.054 and 0.074, respectively).

To explore the functional role of the candidate lncRNAs in the cell cycle, we used siRNAs to knockdown each candidate (Figure S21) and cell counting to evaluate the effect of knockdown on proliferation. We tested four different cell lines—HaCaT, A549, LS411N, and DLD1—and used a hierarchical, linear model and ANOVA to calculate percent growth inhibition across the four cell lines assuming a random effect for each cell line (Fig. 4H). The growth in all four cell lines was significantly affected by *SNHG26* knockdown, with an average growth inhibition of 35% (Fig. 4H). Antisense oligos are supposedly more effective than siRNAs for nuclear localized transcripts, therefore we used ASO-mediated knockdown to confirm the overall growth inhibitory effect of *SNHG26* knockdown (average 55%; Figure S22). Proliferation was also significantly reduced in all cell lines in response to siRNA-mediated knockdown of *EMSLR*, with average growth inhibition of 28% and 37% for two independent siRNAs (Figs. 4H; S22). Meanwhile, the overall growth reduction in all four cell lines in response to *EPB41L4A-AS1* and *ZFAS1* knockdown was not significant. However, knockdown of *EPB41L4A-AS1* resulted in a significant growth reduction in HaCaT, A549, and DLD1 cell lines (average 31%;  $p$ -value = 0.045; Fig. 4H), but the growth of LS411N cells was not affected. An independent siRNA gave similar results, although the growth reduction in HaCaT, A549, and DLD1 was not significant ( $p$ -value = 0.070; Figure S22). In response to *ZFAS1* knockdown, the growth of the cancerous cell lines A549, LS411N, and DLD1 was significantly reduced by an average of 19% ( $p$ -value = 0.0054; Fig. 4H), which was validated using another siRNA (average 20%;  $p$ -value = 0.0016; Figure S22). Notably, both *ZFAS1* siRNAs gave a slight growth increase in HaCaT cells (20% and 6%; Figs. 4H and S22).

Finally, we investigated the distribution of HaCaT cells in different cell cycle phases in response to siRNA-mediated knockdown of the candidate lncRNAs (Figs. 4I, S23). Whereas the effects of individual siRNAs varied somewhat per phase, the overall patterns in increased and decreased percentages of cells per phase were largely consistent between the two independent siRNAs. Specifically, ANOVA showed that knockdown of *SNHG26* shifted the phases to a decrease in G1 and increase in G2/M cells; *EMSLR* shifted to increase in G1 and decrease in S; *EPB41L4A-AS1* shifted to decrease in G1 and increase in G2/M; and *ZFAS1* shifted to decrease in G1 and increase in G2/M (Figure S23). Finally, as *SNHG26* was mainly enriched in the nucleus (Figure S18), we used ASO-mediated knockdown (Figure S24) and confirmed the significant reduction of cells in the G1 phase and the significant enrichment of cells in the G2/M phase (Figure S25).

## Discussion

Whereas many lncRNAs have a tissue-specific expression, about 11% of lncRNAs are ubiquitously expressed, suggesting an involvement in cellular functions generally necessary for normal growth and development<sup>34,35</sup>. By sequencing total RNA through the cell cycle we identified 99 lncRNAs with cell cycle-dependent gene expression. Of these, 57 lncRNAs were highly correlated with changes in Pol II or H3K4me3 occupancy at their annotated TSS as measured by ChIP-seq, supporting that these lncRNAs are transcribed in a cycle-dependent manner and thereby are likely to have roles in cell proliferation. Indeed, protein coding genes with similar cell cycle-dependent expression and correlated Pol II changes were strongly enriched for cell cycle functions.

We have previously shown that a subset of protein coding cell cycle genes is cell type-specific in their expression and function<sup>1</sup>. Due to the high cell type specificity of many lncRNAs, we therefore expect that some of the lncRNAs identified as cell cycle-associated in HaCaT cells will differ somewhat in other cell types. Nevertheless, similar to protein coding genes with known cell cycle functions, the cell cycle lncRNAs had increased expression in samples from proliferating compared with non-proliferating tissues, suggesting that many of these lncRNAs will have roles in cell proliferation in multiple cell types. Indeed, several lncRNAs that are commonly dysregulated in cancer and that have known cell cycle functions, such as *GAS5*, *ZFAS1*, *LINC00963*, *DANCR*, and *MALAT1*<sup>20,36–39</sup>, were among the 99 lncRNAs with a cyclic expression profile. Moreover, three of the top

five lncRNAs (Table S2) have already been connected to proliferation and cancer in at least one functional study; *CTD-2555C10.3*<sup>40</sup>, *SNHG16*<sup>41</sup>, and *SNHG26*<sup>42</sup>. Of these, *SNHG16* is best categorized and is often found overexpressed in cancers where it is associated with poor prognosis<sup>43,44</sup>. Cao et al.<sup>41</sup> demonstrated that *SNHG16* increases proliferation in bladder cancer by epigenetic silencing of p21, a potent CDK inhibitor with several functions in the cell cycle<sup>45</sup>.

To further investigate the cell cycle-associated lncRNAs, we used siRNAs to down-regulate four candidates—*SNHG26*, *EMSLR*, *EPB41L4A-AS1*, and *ZFAS1*—and evaluated the resulting effects on cell proliferation in four cell lines (HaCaT, A549, DLD1, and LS411N) and cell cycle progression in HaCaT.

*SNHG26* was the top cell cycle-associated lncRNA candidate, based on its expression and corresponding correlation to Pol II and H3K4me3 ChIP-seq signals. Downregulation of *SNHG26* led to growth inhibition in all four cell lines and affected the cell cycle distribution of HaCaT cells by increasing the number of cells in G2/M by more than 10 percentage points. Thus *SNHG26* seems to be necessary for a normal G2/M progression, which is in line with the gene's expression profile from RNA-seq, which was most similar to those of known G2/M phase genes. *SNHG26* is also known as small nucleolar RNA host gene 26 (*SNHG26*), which belongs to a group of lncRNAs called small nucleolar RNA host genes (SNHG) that are often found upregulated in cancers, and that have oncogenic functions connected to proliferation and cell cycle progression. Zimpta et al.<sup>44</sup> reviewed the role of SNHG focusing on oncogenic properties and potential clinical applications. Overexpression of SNHG is often correlated with progression and lower overall survival in several cancers, including lung, stomach, bone, esophagus, liver, brain, and colon. Moreover, in line with our results, the oncogenic properties of SNHG can be effectively impaired by downregulating their expression using RNA interference<sup>44</sup>. *SNHG26* has been identified as dysregulated in gene expression datasets from different types of cancer<sup>46–48</sup>. The only study involving any functional characterization of *SNHG26* was a recent publication identifying *SNHG26* as a direct transcriptional target of the well-known oncogenic *c-MYC* and a mediator of MYC-driven proliferation in human lymphoid cells<sup>42</sup>. Our study is the first to report that *SNHG26* is necessary for a proper G2 to M transition during cell cycle, and is important for normal proliferation in HaCaT, A549, DLD1, and LS411N cells.

*EMSLR* expression peaked in G1 phase, and its knockdown gave growth inhibition in all four cell lines, enrichment of cells in the G1 phase by 3–7 percentage points and a corresponding reduction of cells present in the S phase. These results are in line with a previous functional study of *EMSLR* in A549 cells<sup>49</sup>. In addition to reporting similar growth inhibition and enrichment of cells in G1 phase, the study revealed that *EMSLR*, there named E2F1 mRNA stabilizing (*EMS*) lncRNA, is a direct transcriptional target of *c-MYC*. Mechanistically, their results suggest that *EMS* modulates E2F1 stability and promotes G1 to S cell cycle progression through *c-MYC*. *EMSLR* is upregulated in tissue from colon cancer patients in several datasets, and its expression is associated with poor prognosis<sup>50–52</sup>. In one study *EMSLR* was differentially expressed between patients with early and advanced stage endometrial carcinoma, where increased expression of *EMSLR* was associated with disease progression<sup>53</sup>. Previous studies together with our results, indicate an oncogenic role of *EMSLR*, possibly by interfering with the progression from G1 to the S phase of the cell cycle.

In our study, knockdown of *EPB41L4A-AS1* resulted in an overall growth inhibition of HaCaT, A549, and DLD1 cells, and the distribution of HaCaT cells in the different phases of the cell cycle was also affected with an increase of cells in the G2/M phase by 7 percentage points and a decrease in G1 by 6 percentage points. In the RNA-seq data and RT-qPCR technical validation, *EPB41L4A-AS1* expression peaked in M/G1, but this pattern was only reproduced in one of the two biological validation experiments (Val2, Fig. 4C). Several gene expression studies have identified *EPB41L4A-AS1* as both over- and underexpressed in cancer, probably depending on type of tissue and stage of progression<sup>54–56</sup>. Functional studies suggest a central role of *EPB41L4A-AS1* in metabolic reprogramming and as a repressor of the Warburg effect in placental tissue of miscarriage<sup>57</sup> and in cancer cells (cervical, breast, bladder, and liver)<sup>58</sup>. Another functional study investigating the role of *miR-146a* on the proliferation of bone marrow-derived mesenchymal stem (BMSC) cells, reported that *miR-146a* interacts with and inhibits the expression of *EPB41L4A-AS1* and *SNHG7*. Also, overexpression of *EPB41L4A-AS1* increased proliferation and affected the phase distribution of BMSCs, with a reduction of cells in G1/G0 phase and an increased percentage of cells in the S and G2/M phase<sup>59</sup>. The expression of *EPB41L4A-AS1* is higher in colorectal cancer tissue compared to normal tissue, and in line with our results, knockdown of *EPB41L4A-AS1* decreased proliferation in colorectal cancer cell lines HCT116 and SW620<sup>60</sup>. Based on previous studies, the expression of *EPB41L4A-AS1* and its biological role seem to vary in different types of cells, tissues, and stage of disease, suggesting a potential role as a biomarker for disease progression and as a therapeutic target. Our study is the first to evaluate how *EPB41L4A-AS1* affects the cell cycle and proliferation in HaCaT, A549, and DLD1 cells. The results suggest an oncogenic role of *EPB41L4A-AS1*, possibly by affecting the progression from G2 to M phase.

The function of *ZFAS1* has been evaluated in several papers, and like many other lncRNAs it varies depending on the type of cell line or tissue being investigated. *ZFAS1* was first described as a regulator of mammary development<sup>61</sup>. Later studies identified it as an oncogene upregulated in several cancers, including lung, colon, ovary, glioma, liver, and gastric cancers, but downregulated in breast cancer<sup>62</sup>. In a study from Fan et al., overexpression of *ZFAS1* resulted in G1/G0 phase arrest in two breast cancer cell lines<sup>20</sup>. In contrast, we found that *ZFAS1* knockdown had no significant effect on HaCaT proliferation and phase distributions. Instead, our results demonstrated reduced proliferation in response to knockdown in the cancerous cell lines A549, DLD1, and LS411N. Consequently, and in line with previous studies describing the oncogenic characters of *ZFAS1*, our results show that *ZFAS1*'s effects on cell proliferation are cell type-specific and may depend on other oncogenic transformations.

Three of the four lncRNAs we investigated (*SNHG26*, *ZFAS1*, and *EPB41L4A-AS1*) overlap with snoRNAs and could therefore be SNHG—the primary transcripts for these snoRNAs. Whereas our results show that knockdown of these lncRNAs affect proliferation and cell cycle progression, we acknowledge that some of these effects may be mediated by overlapping snoRNAs.

In this study we used double thymidine block to arrest and subsequently release HaCaT cells at the G1/S transition and we identified 1803 genes with significant periodic expression patterns during the cell cycle. However, all chemical synchronization procedures have disadvantages as they may perturb the normal cell cycle and potentially trigger stress-related responses and cause DNA damage<sup>63</sup>. Thus, disruptions of normal biological processes caused by the manipulations during cell synchronization can give false positive cyclic profiles and hide the expression of true cell cycle regulated genes. Although we cannot exclude such false positives among our list of candidate lncRNAs, the additional FACS and RT-qPCR analyses showed reduced expression levels in S phase for *SNHG26*, *EPB41LAAAS1*, and *EMSLR* (Figure S20), consistent with their RNA-seq profiles following the double thymidine block.

Several lncRNAs have been identified as cell cycle-associated<sup>64–67</sup>. Hung et al.<sup>66</sup> used tiling microarrays to investigate ncRNA expression in promoters of 56 cell cycle genes. Their expression data included HeLa and U2OS cells synchronized by double thymidine block, primary cells or cell lines perturbed by DNA damage, or differentiation or oncogenic stimuli, and tumors and paired normal tissues. Cell cycle-dependent expression of lncRNAs in HeLa cells synchronized by double thymidine block was also confirmed by RNA-seq<sup>67</sup>, though this study identified a different set of 39 lncRNAs with cyclic expression. Ali et al.<sup>64</sup> used RNA-seq of nascent RNAs in HeLa cells synchronized by thymidine and hydroxyurea. The data included three time points in the subsequent S-phase and their analyses identified 1145 lncRNAs with temporal expression changes. Hao et al.<sup>65</sup> used deep RNA-seq of U2OS cells synchronized to different phases of the cell cycle and identified more than 2000 lncRNAs that had phase-specific expression. Cells were synchronized to different stages by nocodazole (M) followed by mitotic shake-off (G1), and by double thymidine block (G1/S) and subsequent isolation at 4 h (S) and 8 h (G2) following release. They observed that 35–40% of the genes differentially expressed during the cell cycle consisted of lncRNAs, with the majority being highly expressed in G1 phase. In our study, 56% of the cell cycle-associated lncRNAs were assigned to S phase, whereas 19% were assigned to G1. All of our candidate lncRNAs were identified in the study by Hao et al.<sup>65</sup> where *SNHG26*, *EPB41LAA-AS1*, and *EMSLR* showed reduced expression levels in S phase, which is in line with our results. We do note the differences between these studies in the number of identified lncRNAs and in their phase-specific expression characteristics. But we also note differences in data generation, lncRNA annotations, and data analyses, which all may have contributed to these differences.

Our analyses suggest that by using a positive correlation between the RNA-seq expression profile and H3K4me3 and Pol II signal as a selection criteria, we identify genes that are actively transcribed and highly enriched for cell cycle functions. Although this is a useful method for detecting cell cycle-associated lncRNAs, cyclic lncRNAs with low correlation to ChIP-seq signal should not be dismissed, as they may also be possible candidates for cell cycle involvement. As exemplified by *PCNA*, genes with several transcripts may have a low correlation value if the wrong TSS was selected for ChIP-seq signal analysis. More sophisticated analyses that use the ChIP-seq data to identify the most likely active TSS per gene may eliminate some of these false negatives.

In summary, our results indicate that all four candidate lncRNAs tested in this study influenced both proliferation and cell cycle progression, although the degree of effect varied between the candidates. The top candidate *SNHG26*, which has a cyclic expression pattern and the highest overall correlation to Pol II and H3K4me3 ChIP-seq signal, did have a consistent effect on overall growth inhibition and cell cycle phase distribution in response to knockdown. Results from our functional evaluation support that our multi-omics method is well suited for identifying lncRNAs involved in the cell cycle.

## Methods

**Cell culture.** All cell lines were obtained from the American Type Culture Collection (ATCC) and cultivated in a humidified incubator at 37 °C and 5% CO<sub>2</sub>. The human keratinocyte cell line HaCaT was cultured in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich, D6419) supplemented with 10% fetal bovine serum (FBS, Sigma-Aldrich, F7524), 2 mM glutamine (Sigma-Aldrich, G7513), 0.1 mg/ml gentamicin (Gibco, 15710049), and 1.25 µg/ml fungizone (Sigma-Aldrich, A2942). For the lung carcinoma cell line A549 we used DMEM supplemented with 10% FBS and 2 mM glutamine, while the colorectal carcinoma cell line LS411N and the colorectal adenocarcinoma cell line DLD1 were cultivated in RPMI 1640 medium (Gibco, A1049101) supplemented with 10% FBS.

**Cell cycle synchronization.** HaCaT cells were seeded in 150-mm culture dishes ( $2 \times 10^6$  cells each dish) and were arrested in the G1/S transition by double thymidine block. Briefly, cells were treated with 2 mM thymidine for 18 h, released from the arrest for 10 h and arrested a second time with 2 mM thymidine for 18 additional hours. After blocking, media was replaced and cells were collected every third hour for 24 h, covering approximately two cell cycles. Unsynchronized cells were used as a reference sample.

**Cell cycle and fluorescence-activated cell sorting (FACS) analysis.** We used FACS analysis to determine the cell cycle phase distribution. HaCaT cells were washed twice with preheated PBS and trypsinated for 8 min before collected using cold PBS supplemented with 3% FBS. Then we centrifuged the cells at 4 °C for 5 min. The cell pellet was resuspended in 100 µl cold PBS and fixed by adding 1 ml cold (– 20 °C) methanol dropwise while vortexing at 1600 rpm and stored at 4 °C until DNA measurement. Cells were then washed with cold PBS and incubated with 200 µl of DNase-free RNase A in PBS (100 µg/ml) for 30 min at 37 °C before DNA staining with 200 µl of Propidium Iodide (PI, Sigma; 50 µg/ml) at 37 °C for 30 min. Cell cycle analyses were performed by using a BD FACS Canto flow cytometer (BD Biosciences). The excitation maximum of PI is 535 nm and the emission maximum is 617 nm. Here, PI-stained cells were excited with the blue laser (488 nm), and the PI fluorescence was detected in the Phycoerythrin (PE) channel (578 nm). PE channel (578 nm). Quantification

of cells in each phase was done with the FlowJo software and the percentage of cells assigned to G1, S, and G2/M phases was calculated.

**Total RNA-seq.** Total RNA was isolated using the mirVana miRNA Isolation Kit (ThermoFisher Scientific, AM1560) according to the manufacturer's protocol. Integrity and stability of RNA samples were assessed by Agilent 2100 Bioanalyzer (Agilent Technologies), whereas the RNA concentration and quality were measured on a NanoDrop ND-1000 UV-Vis Spectrophotometer. RNA-seq libraries were prepared using the Illumina TruSeq Stranded Total RNA with Ribo-Zero™ Human/Mouse/Rat kit, according to the manufacturer's instructions (Illumina; Supplementary Method 1). The sequencing (50 cycles single end reads) was performed on an Illumina HiSeq2500 instrument, in accordance with the manufacturer's instructions. FASTQ files were created with bcl2fastq 2.18 (Illumina).

**Identifying cell cycle genes.** RNA-seq raw reads were quality-filtered using fastq\_quality\_filter 0.0.13 ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/); parameters -Q33 -q 20 -p 80 -z), and subsequently aligned to human genome (version GRCh38.p7) with STAR 2.4.0.f1<sup>68</sup>; parameters -chimSegmentMin 30 -runThreadN 12 -outFilterMultimapNmax 20 -alignSJoverhangMin 8 -alignSJDBoverhangMin 1 -outFilterMismatchNmax 10 -outFilterMismatchNoverLmax 0.04 -alignIntronMin 20 -alignIntronMax 1,000,000). Read alignments were then feature-counted on annotated exons and summarized on genes, using htseq-count 0.6.0<sup>69</sup>; parameters -r pos -i gene\_id -t exon -s yes. The resulting raw count matrix was stripped off genes with zero-count in any of the profiles, preventing such genes from dominating the partial least squares regression (PLS) model. Specifically, out of 58,051 gtf-annotated genes in Human Gencode v24, at least a single read was present for 31,433 genes, and a total of 14,059 was left for analysis. This filtered count matrix was finally transformed to the logarithmic domain and adjusted with precision weights to reduce heteroscedasticity based on abundance using voom from the Limma package<sup>70</sup>. To identify genes with a cell cycle-dependent profile, we used PLS as previously described<sup>1</sup>, except that we used the FACS cell fraction matrix directly for the response matrix. Cell cycle phases were assigned by using the profiles of known phase-associated cell cycle genes as described in<sup>1</sup>.

**Quantitative reverse transcription PCR (RT-qPCR).** We isolated total RNA by using the mirVana miRNA Isolation Kit (ThermoFisher Scientific, AM1560) before DNA was removed using TURBO DNA-free™ Kit (Invitrogen, AM1907), according to the manufacturer's instructions. RNA concentration and quality were measured on a NanoDrop ND-1000 UV-Vis spectrophotometer. Total RNA was reverse transcribed using TaqMan reverse transcription reagents (Applied Biosystems, N8080234) followed by quantitative real-time PCR using SYBR™ Select Master Mix (Applied Biosystems, 44729199) and quantification by the Step One Real-time PCR system (Applied Biosystems). RT<sup>2</sup> lncRNA PCR assays (Qiagen, 330701) and QuantiTect primer assays (Qiagen, 249900) that were used for lncRNA and mRNA expression analysis are listed in Table S4. The relative expression of mRNAs and lncRNAs was calculated using the  $\Delta\Delta C_t$  method<sup>71</sup> with Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) as an endogenous control.

**ChIP-seq.** Chromatin immunoprecipitation (ChIP) was performed as described in Supplementary Method 2. The antibodies used for immunoprecipitation (IP) were obtained from Diagenode; anti-H3K4me3 (C1541003-50), anti-H3K27me3 (C15410195), and anti-Pol II (C15200004). As a control for successful IP, qPCR was performed using human positive and negative control qPCR primer sets from Active Motif (Table S5). Immunoprecipitated material and input chromatin were submitted to the Genomics Core Facility (GCF) at Norwegian University of Science and Technology (NTNU) for library preparation (Supplementary Method 3) and sequencing. ChIP-seq libraries were prepared using the MicroPlex Library Preparation Kit v4 (Diagenode) and the sequencing (50 cycles single end reads) was performed on an Illumina HiSeq2500 instrument, in accordance with the manufacturer's instructions (Illumina). FASTQ files were created with bcl2fastq 2.18 (Illumina).

**ChIP-seq data analysis.** The ChIP-seq FASTQ files were aligned against the reference human genome hg38 with the hisat2 aligner<sup>72</sup>. Using the -k flag, we only kept 1 primary alignment per read. We also disallowed spliced alignments. We then found the reads per kilobase million (RPKM)-coverage of each alignment file for the genes in hg38 using deepTools bamCoverage<sup>73</sup>. The coverage of each ChIP sample was divided by the average of the coverage of the input sample to provide a normalized expression per gene. For each gene in Human Gencode v24 we found the transcription start sites (TSSs) of the genes and the longest transcript variant was chosen. Then we binned the 10,000 bp area around the TSSs into bins of 50 bp. We found the counts of the reads within each bin by extending each read by 75 (half the fragment size). Then we saw which bin overlapped with the point TSS + 75. Each read was only considered to belong to one bin.

**Pol II ChIP-qPCR.** Cells were harvested and Pol II ChIP was performed as described in Supplementary Method 2, except that cells were sonicated using a Bioruptor Pico (Diagenode) for 14 cycles of 30 s ON/30 s OFF in a volume of 300  $\mu$ l. No digestion with Micrococcal Nuclease was included. For accurate fragment assessment, the shared chromatin was analyzed on a 2% agarose gel. Fragment size was optimized to be 200–500 bp. As a control for successful IP, qPCR was performed using human positive and negative control qPCR primer sets from Active Motif (Table S5). PCR primers were designed in the TSS region for selected genes (Table S1). ChIP DNA was diluted 1:2 in TE buffer and qPCR was performed on immunoprecipitated material and input chromatin. We added 2  $\mu$ l ChIP DNA and 500 nM of each primer to SYBR Select Master Mix (Applied Biosystems) in technical duplicates. Target values from all qPCR samples were normalized with matched input DNA using the

percent input method [ $100 \times 2^{(\text{Adjusted input} - \text{Ct (IP)})}$ ]. The relative expression of selected genes (Table S1) was normalized against *GAPDH* and *ACTB* (mean).

**Western blot analysis.** We used western blot analysis to validate the Pol II protein level, and H3K4me3 and H3K27me3 modification levels in double thymidine blocked synchronized HaCaT cells (Supplementary Method 4). Primary antibodies for Pol II (C15200004), H3K4me3 (C15410003), and H3K27me3 (C15410195) were obtained from Diagenode.

**RNA interference.** All cells were transfected with 20 nM siRNAs or Antisense LNA GapmeR (Antisense oligo; ASO) using Lipofectamine RNAimax (Invitrogen™, 13778030) when seeded, according to the manufacturer's protocol. Cells were harvested after 48 and/or 72 h at about 70% confluence. MISSION<sup>®</sup> siRNA Universal Negative Control #1 (Sigma, SIC001) and the negative control A Antisense LNA GapmeR (Qiagen, LG00000002) were used as controls for siRNAs and ASO, respectively. The producers and sequences of siRNAs and ASO are listed in Table S6. All cell culture experiments were performed in three or more independent experiments, and with siRNAs/ASO targeting two different sequences within the same lncRNA.

**Viability assay.** We performed cell counting using Moxi z mini automated cell counter (ORFLO Technologies) to investigate how knockdown of the lncRNA candidates affected cell growth. All four cell lines (HaCaT, A549, LS411N, and DLD1) were seeded in triplicates for each condition in a 24-well tray and counted 72 h after transfection. Each well was washed twice with preheated PBS and trypsinated for 5–10 min before the cells were resuspended in preheated growth medium and counted. We applied a two-tailed, paired Student's *t*-test to test whether the growth was significantly different ( $p < 0.05$ ) between cells transfected with negative control siRNA/ASO and a lncRNA target-specific siRNA/ASO in at least three independent experiments.

**Statistical analysis.** The results were presented as the mean values  $\pm$  standard error of mean (SEM). Differences between groups were estimated using the two-tailed Welch's *t*-test, assuming unequal variances. The one-sample *t*-test was used to compare the mean of the sample data to a known value. We used a hierarchical, linear model and ANOVA to calculate (1) the percent growth inhibition across four cell lines and (2) the distribution of HaCaT cells in different cell cycle phases in response to siRNA/ASO-mediated knockdown of the candidate lncRNAs, assuming a random effect for each cell line/siRNA/ASO. A value of  $P < 0.05$  was considered statistically significant for all tests. All data analyses were performed in R. Plots were made using the packages ggplot2 and ggpubr. Gene ontology (GO) analyses were done with the package gProfiler.

## Data availability

The RNA-seq and ChIP-seq datasets generated and analyzed during the current study are available in the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/>) under the accession number of PRJEB40813; the processed data are available as a genome browser track hub from <http://tare.medisin.ntnu.no/EpiPP/EpiPP-hub/hub.txt>.

Received: 7 April 2021; Accepted: 26 July 2021

Published online: 23 September 2021

## References

- Pena-Diaz, J. *et al.* Transcription profiling during the cell cycle shows that a subset of polycomb-targeted genes is upregulated during DNA replication. *Nucleic Acids Res.* **41**, 2846–2856. <https://doi.org/10.1093/nar/gks1336> (2013).
- Whitfield, M. L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977–2000. <https://doi.org/10.1091/mbc.02-02-0030> (2002).
- Bar-Joseph, Z. *et al.* Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl. Acad. Sci. USA.* **105**, 955–960. <https://doi.org/10.1073/pnas.0704723105> (2008).
- Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nat. Genet.* **27**, 48–54. <https://doi.org/10.1038/83751> (2001).
- Beyrouthy, M. J. *et al.* Identification of G1-regulated genes in normally cycling human cells. *PLoS ONE* **3**, e3943. <https://doi.org/10.1371/journal.pone.0003943> (2008).
- Crawford, D. F. & Piwnica-Worms, H. The G(2) DNA damage checkpoint delays expression of genes encoding mitotic regulators. *J. Biol. Chem.* **276**, 37166–37177. <https://doi.org/10.1074/jbc.M103414200> (2001).
- Lew, D. J., Dulić, V. & Reed, S. I. Isolation of three novel human cyclins by rescue of G1 cyclin (Cln) function in yeast. *Cell* **66**, 1197–1206. [https://doi.org/10.1016/0092-8674\(91\)90042-w](https://doi.org/10.1016/0092-8674(91)90042-w) (1991).
- Zariwala, M., Liu, J. & Xiong, Y. Cyclin E2, a novel human G1 cyclin and activating partner of CDK2 and CDK3, is induced by viral oncoproteins. *Oncogene* **17**, 2787–2798. <https://doi.org/10.1038/sj.onc.1202505> (1998).
- Pines, J. & Hunter, T. Isolation of a human cyclin cDNA: Evidence for cyclin mRNA and protein regulation in the cell cycle and for interaction with p34cdc2. *Cell* **58**, 833–846. [https://doi.org/10.1016/0092-8674\(89\)90936-7](https://doi.org/10.1016/0092-8674(89)90936-7) (1989).
- Jackman, M., Firth, M. & Pines, J. Human cyclins B1 and B2 are localized to strikingly different structures: B1 to microtubules, B2 primarily to the Golgi apparatus. *EMBO J.* **14**, 1646–1654 (1995).
- Yan, J. *et al.* Long noncoding RNA expression profile and functional analysis in psoriasis. *Mol. Med. Rep.* **19**, 3421–3430. <https://doi.org/10.3892/mmr.2019.9993> (2019).
- Mathy, N. W. & Chen, X. M. Long non-coding RNAs (lncRNAs) and their transcriptional control of inflammatory responses. *J. Biol. Chem.* **292**, 12375–12382. <https://doi.org/10.1074/jbc.R116.760884> (2017).
- Sun, Q., Hao, Q. & Prasanth, K. V. Nuclear long noncoding RNAs: Key regulators of gene expression. *Trends Genet.* **34**, 142–157. <https://doi.org/10.1016/j.tig.2017.11.005> (2018).
- Ma, L. *et al.* LncBook: A curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.* **47**, 2699. <https://doi.org/10.1093/nar/gkz073> (2019).
- Kitagawa, M., Kitagawa, K., Kotake, Y., Niida, H. & Ohhata, T. Cell cycle regulation by long non-coding RNAs. *Cell Mol. Life Sci.* **70**, 4785–4794. <https://doi.org/10.1007/s00018-013-1423-0> (2013).

16. Guiducci, G. & Stojic, L. Long noncoding RNAs at the crossroads of cell cycle and genome integrity. *Trends Genet.* **37**, 528–546. <https://doi.org/10.1016/j.tig.2021.01.006> (2021).
17. Luo, G. *et al.* LncRNA GAS5 inhibits cellular proliferation by targeting P27(Kip1). *Mol. Cancer Res.* **15**, 789–799. <https://doi.org/10.1158/1541-7786.Mcr-16-0331> (2017).
18. Gao, K., Ji, Z., She, K., Yang, Q. & Shao, L. Long non-coding RNA ZFAS1 is an unfavourable prognostic factor and promotes glioma cell progression by activation of the Notch signaling pathway. *Biomed. Pharmacother.* **87**, 555–560. <https://doi.org/10.1016/j.biopha.2017.01.014> (2017).
19. Fang, C. *et al.* Long non-coding ribonucleic acid zinc finger antisense 1 promotes the progression of colonic cancer by modulating ZEB1 expression. *J. Gastroenterol. Hepatol.* **32**, 1204–1211. <https://doi.org/10.1111/jgh.13646> (2017).
20. Fan, S. *et al.* Downregulation of the long non-coding RNA ZFAS1 is associated with cell proliferation, migration and invasion in breast cancer. *Mol. Med. Rep.* **17**, 6405–6412. <https://doi.org/10.3892/mmr.2018.8707> (2018).
21. Wang, K. C. & Chang, H. Y. Molecular mechanisms of long noncoding RNAs. *Mol. Cell* **43**, 904–914. <https://doi.org/10.1016/j.molcel.2011.08.018> (2011).
22. Carlevaro-Fita, J. & Johnson, R. Global positioning system: Understanding long noncoding RNAs through subcellular localization. *Mol. Cell* **73**, 869–883. <https://doi.org/10.1016/j.molcel.2019.02.008> (2019).
23. Cabili, M. N. *et al.* Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20. <https://doi.org/10.1186/s13059-015-0586-4> (2015).
24. Arun, G., Diermeier, S. D. & Spector, D. L. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol. Med.* **24**, 257–277. <https://doi.org/10.1016/j.molmed.2018.01.001> (2018).
25. Böhmrdorfer, G. & Wierzbicki, A. T. Control of chromatin structure by long noncoding RNA. *Trends Cell Biol.* **25**, 623–632. <https://doi.org/10.1016/j.tcb.2015.07.002> (2015).
26. Jiang, Q. *et al.* TF2LncRNA: identifying common transcription factors for a list of lncRNA genes from ChIP-Seq data. *Biomed. Res. Int.* **2014**, 317642. <https://doi.org/10.1155/2014/317642> (2014).
27. Sati, S., Ghosh, S., Jain, V., Scaria, V. & Sengupta, S. Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* **40**, 10018–10031. <https://doi.org/10.1093/nar/gks776> (2012).
28. Klein, H. U. *et al.* Integrative analysis of histone ChIP-seq and transcription data using Bayesian mixture models. *Bioinformatics* **30**, 1154–1162. <https://doi.org/10.1093/bioinformatics/btu003> (2014).
29. Angelini, C. & Costa, V. Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: Statistical solutions to biological problems. *Front. Cell Dev. Biol.* **2**, 51. <https://doi.org/10.3389/fcell.2014.00051> (2014).
30. Wan, G. *et al.* Transcriptional regulation of lncRNA genes by histone modification in Alzheimer's disease. *Biomed. Res. Int.* **2016**, 3164238. <https://doi.org/10.1155/2016/3164238> (2016).
31. Liu, G. *et al.* LncRNA ZFAS1 promotes growth and metastasis by regulating BMI1 and ZEB2 in osteosarcoma. *Am. J. Cancer Res.* **7**, 1450–1462 (2017).
32. Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74. <https://doi.org/10.1093/nar/gkt006> (2013).
33. Mas-Ponte, D. *et al.* LncAtlas database for subcellular localization of long noncoding RNAs. *RNA* **23**, 1080–1087. <https://doi.org/10.1261/rna.060814.117> (2017).
34. Jiang, C. *et al.* Identifying and functionally characterizing tissue-specific and ubiquitously expressed human lncRNAs. *Oncotarget* **7**, 7120–7133. <https://doi.org/10.18632/oncotarget.6859> (2016).
35. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775–1789. <https://doi.org/10.1101/gr.132159.111> (2012).
36. Chen, L. *et al.* LncRNA GAS5 regulates redox balance and dysregulates the cell cycle and apoptosis in malignant melanoma cells. *J. Cancer Res. Clin. Oncol.* **145**, 637–652. <https://doi.org/10.1007/s00432-018-2820-4> (2019).
37. Wang, L. *et al.* Linc00963: a novel, long non-coding RNA involved in the transition of prostate cancer from androgen-dependence to androgen-independence. *Int. J. Oncol.* **44**, 2041–2049. <https://doi.org/10.3892/ijo.2014.2363> (2014).
38. Jin, S. J., Jin, M. Z., Xia, B. R. & Jin, W. L. Long non-coding RNA DANCR as an emerging therapeutic target in human cancers. *Front. Oncol.* **9**, 1225. <https://doi.org/10.3389/fonc.2019.01225> (2019).
39. Tripathi, V. *et al.* Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* **9**, e1003368. <https://doi.org/10.1371/journal.pgen.1003368> (2013).
40. Zhang, X. *et al.* Novel long non-coding RNA LINC02323 promotes epithelial-mesenchymal transition and metastasis via sponging miR-1343-3p in lung adenocarcinoma. *Thorac. Cancer* **11**, 2506–2516. <https://doi.org/10.1111/1759-7714.13562> (2020).
41. Cao, X., Xu, J. & Yue, D. LncRNA-SNHG16 predicts poor prognosis and promotes tumor proliferation through epigenetically silencing p21 in bladder cancer. *Cancer Gene Ther.* **25**, 10–17. <https://doi.org/10.1038/s41417-017-0006-x> (2018).
42. Raffener, P. *et al.* An MXD1-derived repressor peptide identifies noncoding mediators of MYC-driven cell proliferation. *Proc. Natl. Acad. Sci. USA* **117**, 6571–6579. <https://doi.org/10.1073/pnas.1921786117> (2020).
43. Yang, M. & Wei, W. SNHG16: A novel long-non coding RNA in human cancers. *Onco Targets Ther* **12**, 11679–11690. <https://doi.org/10.2147/ott.S231630> (2019).
44. Zimta, A. A. *et al.* An emerging class of long non-coding RNA with oncogenic role arises from the snoRNA host genes. *Front. Oncol.* **10**, 389. <https://doi.org/10.3389/fonc.2020.00389> (2020).
45. Karimian, A., Ahmadi, Y. & Yousefi, B. Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair* **42**, 63–71. <https://doi.org/10.1016/j.dnarep.2016.04.008> (2016).
46. Liu, X. D. *et al.* Integrated analysis of lncRNA-mRNA co-expression networks in the  $\alpha$ -particle induced carcinogenesis of human bronchial epithelial cells. *Int. J. Radiat. Biol.* **95**, 144–155. <https://doi.org/10.1080/09553002.2019.1539880> (2019).
47. Hu, A. X., Huang, Z. Y., Zhang, L. & Shen, J. Potential prognostic long non-coding RNA identification and their validation in predicting survival of patients with multiple myeloma. *Tumour Biol.* **39**, 1010428317694563. <https://doi.org/10.1177/1010428317694563> (2017).
48. Bao, Z., Zhang, W. & Dong, D. A potential prognostic lncRNA signature for predicting survival in patients with bladder urothelial carcinoma. *Oncotarget* **8**, 10485–10497. <https://doi.org/10.18632/oncotarget.14441> (2017).
49. Wang, C. *et al.* Long noncoding RNA EMS connects c-Myc to cell cycle control and tumorigenesis. *Proc. Natl. Acad. Sci. USA* **116**, 14620–14629. <https://doi.org/10.1073/pnas.1903432116> (2019).
50. Zhou, M. *et al.* Recurrence-associated long non-coding RNA signature for determining the risk of recurrence in patients with colon cancer. *Mol. Ther. Nucleic Acids* **12**, 518–529. <https://doi.org/10.1016/j.omtn.2018.06.007> (2018).
51. Yamada, A. *et al.* A RNA-sequencing approach for the identification of novel long non-coding RNA biomarkers in colorectal cancer. *Sci. Rep.* **8**, 575. <https://doi.org/10.1038/s41598-017-18407-6> (2018).
52. Sun, L. *et al.* Down-regulation of long non-coding RNA RP11-708H21.4 is associated with poor prognosis for colorectal cancer and promotes tumorigenesis through regulating AKT/mTOR pathway. *Oncotarget* **8**, 27929–27942. <https://doi.org/10.18632/oncotarget.15846> (2017).
53. Sun, Y., Zou, X., He, J. & Mao, Y. Identification of long non-coding RNAs biomarkers associated with progression of endometrial carcinoma and patient outcomes. *Oncotarget* **8**, 52604–52613. <https://doi.org/10.18632/oncotarget.17537> (2017).
54. He, M., Lin, Y. & Xu, Y. Identification of prognostic biomarkers in colorectal cancer using a long non-coding RNA-mediated competitive endogenous RNA network. *Oncol. Lett.* **17**, 2687–2694. <https://doi.org/10.3892/ol.2019.9936> (2019).

55. Chen, P., Zhang, W., Chen, Y., Zheng, X. & Yang, D. Comprehensive analysis of aberrantly expressed long non-coding RNAs, microRNAs, and mRNAs associated with the competitive endogenous RNA network in cervical cancer. *Mol. Med. Rep.* **22**, 405–415. <https://doi.org/10.3892/mmr.2020.11120> (2020).
56. Zhou, M. *et al.* Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget* **7**, 12598–12611. <https://doi.org/10.18632/oncotarget.7181> (2016).
57. Zhu, Y. *et al.* Overexpression of lncRNA EPB41L4A-AS1 induces metabolic reprogramming in trophoblast cells and placenta tissue of miscarriage. *Mol. Ther. Nucleic Acids* **18**, 518–532. <https://doi.org/10.1016/j.omtn.2019.09.017> (2019).
58. Liao, M. *et al.* LncRNA EPB41L4A-AS1 regulates glycolysis and glutaminolysis by mediating nucleolar translocation of HDAC2. *EBioMedicine* **41**, 200–213. <https://doi.org/10.1016/j.ebiom.2019.01.035> (2019).
59. Cui, P. *et al.* miR-146a interacting with lncRNA EPB41L4A-AS1 and lncRNA SNHG7 inhibits proliferation of bone marrow-derived mesenchymal stem cells. *J. Cell Physiol.* **235**, 3292–3308. <https://doi.org/10.1002/jcp.29217> (2020).
60. Bin, J. *et al.* Long noncoding RNA EPB41L4A-AS1 functions as an oncogene by regulating the Rho/ROCK pathway in colorectal cancer. *J. Cell Physiol.* <https://doi.org/10.1002/jcp.29880> (2020).
61. Askarian-Amiri, M. E. *et al.* SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA* **17**, 878–891. <https://doi.org/10.1261/rna.2528811> (2011).
62. Dong, D., Mu, Z., Zhao, C. & Sun, M. ZFAS1: A novel tumor-related long non-coding RNA. *Cancer Cell Int.* **18**, 125. <https://doi.org/10.1186/s12935-018-0623-y> (2018).
63. Kurose, A., Tanaka, T., Huang, X., Tragano, F. & Darzynkiewicz, Z. Synchronization in the cell cycle by inhibitors of DNA replication induces histone H2AX phosphorylation: An indication of DNA damage. *Cell Prolif.* **39**, 231–240. <https://doi.org/10.1111/j.1365-2184.2006.00380.x> (2006).
64. Ali, M. M. *et al.* PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers. *Nat. Commun.* **9**, 883. <https://doi.org/10.1038/s41467-018-03265-1> (2018).
65. Hao, Q. *et al.* The S-phase-induced lncRNA SUNO1 promotes cell proliferation by controlling YAP1/Hippo signaling pathway. *Elife* <https://doi.org/10.7554/eLife.55102> (2020).
66. Hung, T. *et al.* Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* **43**, 621–629. <https://doi.org/10.1038/ng.848> (2011).
67. Dominguez, D. *et al.* A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. *Cell Res.* **26**, 946–962. <https://doi.org/10.1038/cr.2016.84> (2016).
68. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).
69. Anders, S., Pyl, P. T. & Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169. <https://doi.org/10.1093/bioinformatics/btu638> (2015).
70. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29. <https://doi.org/10.1186/gb-2014-15-2-r29> (2014).
71. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> method. *Methods* **25**, 402–408. <https://doi.org/10.1006/meth.2001.1262> (2001).
72. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360. <https://doi.org/10.1038/nmeth.3317> (2015).
73. Ramirez, F. *et al.* deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–165. <https://doi.org/10.1093/nar/gkw257> (2016).

## Acknowledgements

This work was supported by the Norwegian Cancer Society (grant number 2278701) and the Norwegian Research Council (grant number 230338). The library preparation and sequencing were provided in close collaboration with the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU). GCF is funded by the Faculty of Medicine and Health Sciences at NTNU and Central Norway Regional Health Authority.

## Author contributions

S.A.H. designed and performed the cell cycle synchronization experiments, isolated RNA and performed ChIP, did the bioinformatic and statistical analysis and was a major contributor in writing the manuscript. H.S. designed and performed wet lab experiments (cell cycle synchronization validation, RT-qPCR, western blot analysis, RNA interference, cell cycle and viability assays), contributed with statistical analysis and was a major contributor in writing the manuscript. A.K. analyzed RNA-seq data and identified cell cycle genes. E.B.S. analyzed ChIP-seq data. K.G.N. performed cell cycle synchronization validation experiments and performed Pol II ChIP-qPCR analysis. N.B.L. performed all FACS analysis. L.C.O. and K.C. contributed with bioinformatics and statistical analysis. P.A.A. contributed with cell cycle synchronization experiments. P.S. designed and supervised the study and edited the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-97909-w>.

**Correspondence** and requests for materials should be addressed to P.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021