Title:

Spatiotemporal dissection of the cell cycle regulated human proteome

Authors:

Diana Mahdessian[1], Devin Sullivan[1], Frida Danielsson[1], Muhammad Arif[1], Cheng Zhang[1], Lovisa Åkesson[1], Christian Gnann[1], Rutger Shutten[1], Peter Thul[1]. Oana Carja[3,4], Burcu Ayoglu[1], Adil Mardinoglu[1,2], Fredrik Pontén[5], Mathias Uhlén[1], Cecilia Lindskog[5], Emma Lundberg[1,3,4*†]

[1] Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH - Royal Institute of Technology, Stockholm, 17121, Sweden.

[2] Centre for Host–Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, SE1 9RT, United Kingdom

[3] Department of Genetics, Stanford University, Stanford, CA 94305, USA. ‡

[4] Chan Zuckerberg Biohub, San Francisco, San Francisco, CA 94158, USA.

[5] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden.

*Correspondence to: emma.lundberg@scilifelab.se

† Visiting appointment, current address.

Abstract

Here we present a spatiotemporal dissection of proteome single cell heterogeneity in human cells, performed with subcellular resolution over the course of a cell cycle. We identify 17% of the human proteome to display cell-to-cell variability, of which we could attribute 25% as correlated to cell cycle progression, and present the first evidence of cell cycle association for 258 proteins. A key finding is that the variance, of many of the cell cycle associated proteins, is only partially explained by the cell cycle, which hints at cross-talk between the cell cycle and other signaling pathways. We also demonstrate that several of the identified cell cycle regulated proteins may be clinically significant in proliferative disorders. This spatially resolved proteome map of the cell cycle, integrated into the Human Protein Atlas, serves as a valuable resource to accelerate the molecular knowledge of the cell cycle and opens up novel avenues for the understanding of cell proliferation.

Introduction

Cellular processes are, to a great extent, driven by the presence and activity of specific proteins. Essential processes, such as the cell division cycle, require precise coordination of the expression of hundreds of genes and the activity of their corresponding proteins in both time and space. The cell division cycle is tightly controlled at specific checkpoints [1,2] by regulated transcription [3-7], intricate feed-forward and feedback loops of protein post-translational modifications, and protein degradation [8-12]. Its dysregulation has devastating consequences, such as uncontrolled cell proliferation, genomic instability [13], and cancer [14,15].

Given the fundamental role of the cell cycle, its regulation with cyclins and cyclin dependent kinases (CDKs) has been extensively studied [16]. Recent efforts have focused on the investigation of genome-wide effects of cell cycle progression. Transcriptomics studies have revealed 400-1,200 human genes [17-20], and mass spectrometry-based proteomics studies have revealed 300-700 human proteins that show variation in abundance over the cell cycle [21-24]. These studies have commonly been performed in bulk, with cells sorted into synchronized populations [17,19,25-28]. This is a disruptive procedure, shown to alter gene expression [29], and perturb cellular morphology [30-32] as well as metabolism [33]. In addition, the achieved synchrony could be contaminated with cells from other phases [33-36].

Single-cell sequencing now allow the analysis of transcriptional changes without the need for synchronized cells. Recent single-cell transcriptomic studies presented the first efforts to update the decade old catalogues of periodic gene expression patterns that were based on bulk analysis [37-39]. For instance, in a study using human myxoid sarcoma cell line (MSL) cells, 472 genes with periodic expression were identified [37], of which 269 had no prior association to the cell cycle, indicating the potential of single-cell level studies to deepen our knowledge of the cell cycle.

Microscopy offers an attractive approach to study cell cycle dynamics in asynchronous cells at a single-cell level. The readout of such studies has so far been focused only on cellular growth phenotypes, as conferred by genetically encoded fluorescent indicators [40-43]. Due to technological limitations, studies of single cell variations at the proteome level have not yet been feasible. The few studies that exist [44,45] have been limited to a low number of proteins and none provides a complete view of temporal cell cycle dynamics of the human proteome with single cell resolution.

Here we report on a systematic characterization of temporal protein expression patterns with single-cell resolution in unsynchronized human cells, and present the first spatially resolved map of human proteome dynamics during the cell cycle. By leveraging the Human Protein Atlas (HPA) antibody resource [46] and the high-resolution image collection within its Cell Atlas [47], we provide a catalogue of human proteins with temporal and spatial variation correlating to cell cycle progression. This spatially resolved proteome map of the cell cycle, integrated into the HPA database, is a complement to the existing human cell cycle gene expression resources. Altogether this study has important implications for mechanistic insights into cellular proliferation as well as the contribution of its miss-regulation to tumorigenesis and disease.

Results

Single-cell variations of the human proteome

The HPA Cell Atlas aims to localize all human proteins at a subcellular level using immunofluorescence and confocal microscopy (*45*). To date, 12,390 (v.19) proteins have been localized to 33 subcellular structures. This high-resolution image collection contains protein expression in a variety of human cell lines, always non-synchronized and in log-phase growth, and provides an unprecedented resource to explore protein expression variation at single-cell level. Out of these 12,390 proteins mapped in the HPA Cell Atlas, 2,195 (17%, ***Supplementary Table 1***) showed cell-to-cell variations based on visual inspection, either in terms of variation in protein expression level or variation in spatial distribution. As exemplified in ***Figure 1A***, CCNB1, an important cell cycle regulator [48] localized to the cytosol, shows variation in abundance, whereas MRTO4, a protein with unknown function, shows spatial variation in its expression between the nucleus and nucleoli. Out of these 2,195 proteins, 69% showed similar cell-to-cell variations in more than one human cell line (***Supplementary Table 2***), as exemplified for RACGAP1 in three different cell lines (***Figure 1B***). This suggests that these proteome variations might be to a large extent controlled by preserved regulatory mechanisms. We investigate to what extent these observed protein variations represent temporally controlled expression patterns correlating to cell cycle progression.

Proteins spatiotemporally restricted to mitotic cellular structures

The cell cycle dependency of a protein can be inferred directly, if it localizes to a mitotic structure (*i.e.* kinetochores, mitotic spindle, midbody, midbody ring, cleavage furrow, or cytokinetic bridge). For example, the mitotic regulators INCENP [49] and SGO1 [50] appear at the kinetochores during mitosis; KIF20A [51] localizes to the cleavage furrow; and TACC3 [52] to the mitotic spindles (***Figure 1C***). Of the 2,195 proteins identified to show cell-to-cell variability, a total of 166 mapped to one or several of the mitotic structures (99 to cytokinetic bridge, 45 to mitotic spindle, 40 to midbody, 17 to midbody ring, 5 to kinetochores, and 3 to cleavage furrow). Among these proteins, 99 were not previously annotated to have an association with the cell cycle by a biological process (BP) term in Gene Ontology (GO) [53] or Reactome [2], nor did they have any cell cycle phenotype registered in Cyclebase [54] (***Supplementary Table 3***). Among the proteins spatiotemporally restricted to mitotic substructures were *e.g.* BIRC5, a well characterized protein essential for chromosome alignment [55], which localizes to the cytokinetic bridge as well as two other uncharacterized proteins, GLI4 and C12orf66

(***Figure 1D***). C12orf66 localizes to the lysosomes during interphase [56]. DVL3, a Wnt signaling component known to be involved in cell proliferation [57], localized to the midbody ring, which is the final bridge between dividing cells (***Figure 1D***). It is plausible to hypothesize that the proteins which localized to the mitotic spindle are involved in the process of chromosome segregation; these include KIF11 and KNSTRN, both of which are well-studied components of the mitotic spindle [58,59]. We also identified novel proteins localizing to the mitotic spindle, such as MGAT5B, a glycosyltransferase for which downregulation has been shown to inhibit cell proliferation [60]; and FKBPL, a crucial protein for response to high dose radiation stress [61] (***Figure 1D***). Altogether, these 166 proteins serve as potentially interesting targets for development of novel antimitotic drugs for cancer therapy.

Proteins with temporal expression variation correlated to cell cycle interphase progression

To determine if the observed cell-to-cell variations correlate to interphase progression, the FUCCI cell cycle marker system was used (***Figure 1E***) [42,62]. Of the 2,195 proteins identified to show cell-to-cell variability, 1,188 proteins that were expressed and exhibited variations in the U-2 OS cell line were selected for further analysis with the FUCCI system (***Supplementary Table 4***). The expression of each protein was quantified across the cell cycle by immunostaining in U-2 OS FUCCI cells. Gaussian mixture modelling was used to define three clusters representing G1, the S-transition (denoted G1/S) and the remaining S and G2 phases (denoted S/G2), and the subsequent assignment of cells to each cluster. A polar coordinate system was used to transfer the FUCCI marker information into a linear model of interphase pseudo-time (***Figure 1E***). Examples of this analysis are given in ***Figure 1F***: ANLN, a well-characterized cell cycle regulator [63], showed a significant (Kruskal Wallis $p<0.01$& FDR<0.05) increase in abundance during cell cycle progression in the nucleus. On the other hand, FAM71F, an uncharacterized protein localized to the cytosol, revealed variation that did not correlate to the cell cycle, meaning that both high and low expressing cells are present in all phases of the cell cycle. Expression of DUSP18, a member of the DUSP family [64] with no prior association to the cell cycle, was found to strongly correlate to cell cycle progression. In this analysis, staining of microtubules with alpha-tubulin in all samples served as a negative control, with no significant variation of expression during cell cycle progression.

Based on this analysis, at an FDR of 5%, we identified 298 out of 1,188 proteins (25%) to have variance in expression levels temporally correlated to cell cycle progression,

and for which the cell-cycle explained more than 10% of the variance in expression. (***Supplementary Table 5 and Supplementary Figure S1***). This cutoff was set as being significantly above the negative control. It is noteworthy that the majority of the proteins analyzed (75%) showed cell-to-cell variations that were largely unexplained by cell cycle progression. Enrichment analysis of GO BP terms was performed for the genes encoding cell cycle dependent and independent proteins. The set of genes identified as cell cycle regulated was highly enriched for functions related to chromosome organization and segregation, regulation of cell cycle processes, cytoskeleton organization, cell division and cytokinesis (***Figure 2A***). Interestingly, the set of genes, with variations not correlating to the cell cycle, was not enriched for any GO BP terms at all. This shows that the identified proteins are indeed involved in cell cycle processes whereas the proteins not correlated to cell cycle are likely involved in a variety of different biological processes.

Population distribution and fraction of variance explained by the cell cycle

To investigate the pattern of variability for these 1,188 proteins, k-means clustering was performed using the kurtosis and skewness features of the distribution of the mean intensity per cell for each protein. The mean fold-change between high and low expressing cells per protein were 7.97. Three clusters were found to represent distinct variation patterns (***Figure 2B***): Cluster 1, the largest cluster (n=1,018), contained most cell cycle dependent and independent proteins, 92% and 83%, respectively. The lower segment of Cluster 1 contained some proteins with a bimodal distribution (***Figure 2B***, exemplified by GATA6), but the majority of the proteins in this cluster had a unimodal normal distribution (***Figure 2B***, exemplified by CCNB1). Cluster 2, the second largest cluster (n=153), contained proteins with slightly skewed distribution profiles with a sharp peak distribution, as exemplified by DEF6. Cluster 3 (n=17) mostly contained proteins not correlated to the cell cycle, where the variation was highly skewed and tailed with few cells expressing the protein. These results show that cell cycle dependent variations are mostly unimodal with a normal distribution across a log-phase growing population of cells.

In addition to identifying the proteins that are regulated by the cell cycle, the single-cell resolution of our dataset allowed us to also calculate the fraction of variance that is determined by the cell cycle. To our knowledge, such analysis has been done neither at transcriptome, nor at proteome level previously. Here, the Gini index [65] was calculated and used as a metric for the variance of these 1,188 proteins (***Figure 2C***). All the proteins analyzed had a Gini index significantly higher than the negative control

(alpha tubulin) used, which serve as yet another check that we are indeed analyzing proteins with heterogeneous expression. The percentage of variance explained by the cell cycle ranged between 10%-91% (the FUCCI markers themselves were controlled at green: 80% and red: 65%) and two distinct populations were identified (***Figure 2C***): one where the variance was determined by the cell cycle (CCD), and one where the variance was independent of the cell cycle (Non-CCD). Interestingly, the majority of the observed cell cycle regulated variations appeared to be controlled by the cell cycle at a low degree (on average 21%). We hypothesize that these cell cycle regulated proteins, where the percentage of variance explained by the cell cycle is low, are important for the cross-talk between the cell cycle and other signaling processes.

Organelle specific differences in temporal cell cycle protein variations

The high subcellular resolution of our analysis allows us to study the role of subcellular localization in cell cycle regulation. We found significant differences in the localization of proteins that show cell cycle dependent or independent expression (***Figure 2D***). Proteins with variations independent of the cell cycle were significantly enriched for localization to the intermediate filaments, nucleoli, nuclear bodies, and mitochondria (binomial one sided test, $p<0.01$, mapped proteome as background), whereas proteins with cell cycle dependent variation were significantly enriched for localization to nucleoli, nuclear bodies and mitotic structures, constituting 33% of the cell cycle dependent proteins (binomial one sided test, $p<0.01$, mapped proteome as background). Half (50%) of the cell cycle dependent proteins resided in the nuclear compartment (2% nuclear speckles, 11% nuclear bodies, 24% nucleoli and 63% nucleus), not surprisingly given that one of the main functions of the nucleus is to perform and control the replication of DNA during the cell cycle.

In our analysis, we find many functionally uncharacterized proteins that share the same subcellular localization as some previously well characterized cell cycle dependent proteins (***Figure 2E***). It is plausible to assume that proteins expressed in the same organelle with similar temporal profiles may be involved in similar cell cycle processes. For example, two mitochondrial proteins with known association to cell proliferation - Pyruvate Carboxylase (PC), involved in gluconeogenesis and shown to be upregulated in several types of cancer [66-68], and XAF1, whose inhibition is known to prevent cell cycle progression [69] were both shown to peak in the S/G2 phase (0.78 and 0.80 in pseudotime, respectively). We could also identify two proteins without a prior association to the cell cycle. PC and XAF1 shared the same subcellular location and temporal expression profile as TTC21B (0.8 pseudotime) and SLIRP (0.8 pseudotime),

both with no previously described association to the cell cycle or cell proliferation. In this manner, we could associate novel and known cell cycle associated proteins with similar temporal profiles in organelles such as the cytosol, nucleus, nucleoli and the Golgi apparatus (***Figure 2E***).

Temporal protein expression patterns through interphase

We next sorted the proteins based on the time of peak expression in order to study the temporal dynamics of the cell cycle dependent proteome (***Figure 3A***). Despite G1 being the longest period of the cell cycle (G1 10.8h; G1/S 2.6h; S&G2 together 11.9h in U-2 OS FUCCI cells), the majority (85%) of the proteins peaked towards the end of the cell cycle corresponding to the S&G2 phases. This analysis enabled identification of proteins which share a highly similar temporal pattern to well-known cell cycle regulators, but with no prior association to the cell cycle. For instance, in the G1 group, well-known cell cycle dependent proteins such as ORC6 (***Figure 3B)***, required for the cell entry into S phase [70], and MCM10, required for DNA replication [71], were identified to have similar patterns as those with no prior association to the cell cycle, such as ZNF32. Recently, overexpression of ZNF32 was associated with a shorter survival time in lung adenocarcinoma cells [72,73]. The group peaking in the end of G1 contained proteins such as JUN, required for progression through the G1 phase of cell cycle [74]; the G1/S specific cyclin CCNE1[75]; and DUSP19 (***Figure 3B)***, a phosphatase whose depletion results in increased mitotic defects [76]. In the SG2 group, several known cell cycle dependent proteins were identified: CCNB1, a G2/M specific cyclin [48], AURKB, a protein involved in the regulation of alignment and segregation of the chromosomes, and BUB1B (***Figure 3B)***, a mitotic checkpoint kinase [77]. This group also contained proteins such as PAPSS1, an estrogen sulfating enzyme with no previously described association to the cell cycle, although its overexpression was reported to affect proliferation [78]. Other proteins in the SG2 group were N6AMT1, a methyltransferase [79]; PHLDB1, an uncharacterized protein; DPH2 (***Figure 3B)***, required for the synthesis of diphthamide; and FLI1, a transcription factor associated to Ewing sarcoma [80] (***Figure 3B***).

Several of the proteins identified as cell cycle dependent, such as ORC6, RBL2, BUB1B, CCNA2 and HORMAD1 have been reported to be involved in cell cycle processes, yet their temporal expression profile across the interphase, which can provide insight into their functionality, has so far remained uncharacterized (***Supplementary Figure S2***). In addition, knowledge about the temporal expression

patterns and the timing of peak expression relative to other proteins is valuable for a deeper causal understanding of the molecular effects of cell cycle progression.

An extended network of cell cycle genes

Of the 464 proteins (298 in interphase and 166 in mitotic structures) identified to correlate to cell cycle progression, 206 (44%) had a known association to the cell cycle as determined either by a GO BP term related to cell cycle processes [53] or Reactome [2], or a cell cycle phenotype registered in Cyclebase [54]. The remaining 258 proteins (56%), had no previous association to the cell cycle (***Supplementary Table 6***). To investigate whether the proteins, identified to be cell cycle regulated in this study, are connected to proteins previously known to be cell cycle regulated, we analyzed protein-protein interactions using the STRING database [81]. This analysis revealed significantly more interactions than expected for a random set of proteins of similar size (Lambda calculations PPI enrichment p-value <1e-16; 1855 interactions; 649 expected number of edges), indicating that the proteins are likely involved in similar biological processes. The known cell cycle dependent proteins were tightly clustered together and made up the core of the network, whereas the newly identified cell cycle regulated proteins formed an extended network (***Figure 3C***). For instance, KIF23 is an essential protein for the microtubule bundling during cytokinesis via its interaction with RACGAP1 [82] and it is known to oscillate temporally in the nucleus during the cell cycle [83]. In our interaction analysis (***Figure 3C),*** KIF23 showed a number of interactions with known cell cycle regulators, but also with proteins with no prior association to the cell cycle such as DRG1; MICAL3, which further interacts with the known NINL protein required for cytokinesis [84]; and RAD51AP1, which further interacts with RACGAP1 and KIF20A required for cytokinesis [85]. This implies that these three proteins with unknown function, DRG1, MICAL3, and RAD51AP1, are involved in the same process as their known interaction partners, in this case cytokinesis.

Poor overlap between the cell cycle dependent proteome and transcriptome

We performed a comparative analysis between the cell cycle regulated proteome identified in our study and the cell cycle transcriptome of U-2 OS osteosarcoma cells obtained by bulk RNA-sequencing of synchronized cells (*26*), as well as the transcriptome of another type of sarcoma cells (myxoid sarcoma cells) obtained by single-cell RNA-sequencing of non-synchronized cells (*36*). Both comparisons revealed a poor overlap of 19% and 10%, respectively (***Supplementary Table 7***). This indicates that the temporal dynamics of proteome regulation may be to a large extent maintained at a translational or post-translational level.

Gene expression patterns across tissues and cancers results in clusters reflecting proliferative activity

To further understand whether the identified proteins are functionally important for cell proliferation in a more native context than cell lines, we investigated the mRNA expression across cohorts of normal and cancer tissue. Hierarchical clustering of the transcript data from bulk RNA-sequencing of normal and cancer tissues from HPA (***Figure 4A***) resulted in four major clusters. The first cluster contained normal tissues with low proliferative activity, such as heart muscle, skeletal muscle and pancreas. The different cerebral tissues formed the second cluster, together with testis, which appeared as an outlier, most likely due to being the only sample with meiotic activity. The third cluster contained mostly normal tissues, such as kidney and breast, and showed mid-range expression level of the proliferation markers Ki67, MCM2, PCNA, CDK1 and MCM6. The fourth cluster contained mostly cancer tissues, such as skin and breast cancer, but also normal tissues with high proliferative activity, such as bone marrow, tonsil and fetal lung. The tissues in this cluster showed high expression of the abovementioned proliferation markers. Most importantly, gene expression levels were significantly higher in the proliferative tissues than the non-proliferative tissues (Kruskal Wallis test p-value $<2e^{-16}$) (***Figure 4B)***.

To further strengthen the conclusion that the novel cell cycle regulated proteins are important for cellular proliferation, we used the RNA-sequencing data from The Cancer Genome Atlas (TCGA) [86] to create genome wide co-expression networks downloaded from TCSBN [87], in which the shortest path between the novel cell cycle regulated genes identified in our study and known cell cycle genes were measured and compared to a randomly sampled set of genes. The novel genes indeed had a significantly (Kolmogorov-Smirnov one-sided test, FDR < 0.05) shorter path to the known cell cycle genes in all cancer tissues and the normal proliferative tissues such as skin, spleen and colon (***Figure 5A***), whereas there was no significant difference (Kolmogorov-Smirnov one-sided test, FDR < 0.05) of the path length in low- or non-proliferating tissues such as adipose, brain, heart and muscle tissues. This shows that even though most of these proteins are not temporally regulated at the gene expression level, their overall gene expression level is still of importance for cellular proliferation.

Genes encoding cell cycle regulated proteins often have an expression correlating to patient survival in cancer

To further test if the level of expression of genes encoding cell cycle regulated proteins is associated to cancer patient outcome, the TCGA data incorporated in the cancer pathology atlas of HPA was used [88], where genes with a statistically significant differential expression between patient populations with long and short survival were identified [86]. Genes with expression levels correlated with long survival time were denoted as favorable, and with shorter survival time were denoted as unfavorable. Globally, over half of all human genes (54%) were shown to have a prognostic association in this manner, as previously described [88]. Interestingly, prognostic genes were significantly overrepresented among the cell cycle regulated proteins identified in our study (67% prognostic) and the majority of these genes (61%) were associated with an unfavorable outcome, further supporting the hypothesis of an important role of these genes in cellular proliferation.

We next incorporated this classification into the generated co-expression networks for different human cancer tissue types. In these networks, an enrichment analysis was further subjected for each genetic community: communities were denoted as favorable, unfavorable or not enriched. All communities contained a mixture of known and novel cell cycle proteins, further strengthening their functional associations. Strikingly, these networks revealed that the association into clusters were highly different for different tumors (***Figure 5B*** and ***Supplementary Figure S3)***, with proteins being in a favorable community in one cancer type while being in an unfavorable community in another cancer type, emphasizing the complexity of cell cycle regulation from a systems perspective.

Many of the proteins identified here as cell cycle regulated are interesting candidates for in-depth studies of their roles in tumorigenesis, and for potential use as biomarkers. For instance, the gene RACGAP1, known to regulate cytokinesis, and DLGAP5, which has been reported to have a role in carcinogenesis [89-91]. In the co-expression network analysis, these genes showed interactions with known cell cycle related genes and were enriched in an unfavorable prognostic cluster in breast cancer and pancreatic cancer, respectively (***Figure 6A***). Immunohistochemical (IHC) analysis showed that these proteins are expressed at low levels in normal tissues (***Figure 6B***) and high levels in corresponding tumor tissues (***Figure 6C***). Their expression profile is shown in ***Figure 6D.*** To gain an insight into their potential pathway involvement, STRING analysis was performed (**Figure 6E**). RACGAP1 showed physical interaction with several members of the mitotic kinesin family required for cytokinesis [92], whereas

377 DLGAP5 showed direct interaction with AURKA, a protein involved in several mitotic
378 events [93].
379
380 A portion of the genes encoding proteins identified in our study (39%) were associated
381 with a favorable outcome, such as SYNE2 and FAM50B (***Figure 6A***). Comparison of
382 IHC staining of these two proteins revealed high expression in normal tissue (***Figure***
383 ***6B***), and low expression in the respective cancers (***Figure 6C***). This suggests that
384 these proteins might function in anti-tumor activities. For example, SYNE2 is a nuclear
385 membrane protein [94], for which we demonstrated temporal expression variation
386 peaking in G2. FAM50B is expressed in the nucleus in interphase and translocates to
387 the cytokinetic bridge in mitosis (***Figure 6D***). SYNE2 shows interaction with genes
388 enriched in cell cycle processes, such as STAG1, SUN2, TERF1 and TERF2 and
389 FAM50B shows a physical interaction with HDAC2 (***Figure 6E***), which is involved in
390 the regulation of cell cycle progression [95].
391
392 We conclude that these novel proteins identified to be cell cycle regulated have the
393 potential of serving as novel diagnostic or therapeutic targets for a variety of human
394 cancers.

Discussion

In this study, we find that a large extent (17%) of the human proteome displays cell-to-cell heterogeneity in terms of level of expression. We present the first temporal analysis of the cell cycle regulated human proteome in unsynchronized cells, mapped at a single cell level with subcellular resolution. Surprisingly, the majority of the variations were not correlated to the cell cycle, which opens up intriguing avenues for further exploration of the deterministic factors that might control these stochastic variations in expression.

We present 258 novel cell cycle regulated proteins, and show that despite a poor overlap with cell cycle transcriptome studies, these genes are expressed significantly higher in proliferating tissues and tumors. The poor overlap to prior transcriptome-based studies of the human cell cycle points towards massive regulation of protein levels at a translational or post-translational level. Another key finding of this study is that the variance of many cell cycle regulated proteins, in particular the newly identified proteins, are only partially explained by the cell cycle. We hypothesize that these proteins are deterministically controlled by other cellular mechanisms which open the door to further follow up work on the role of various signaling pathways in cell cycle regulation.

Finally, we demonstrate that several of the newly identified cell cycle regulated proteins may be clinically significant and have oncogenic or anti-oncogenic functions. We believe that this comprehensive dissection of the cell cycle regulated human proteome, now integrated into the HPA database, will serve as a valuable resource to accelerate studies towards a greater functional understanding of the human cell cycle, the role of these proteins in tumorigenesis and identification of novel clinical markers for cellular proliferation.

Material and Methods

Initial identification of proteins with cell-to-cell heterogeneity

Protein cell-to-cell heterogeneity was identified in the images from the Cell Atlas of the Human Protein Atlas [46] either in terms of variation in abundance, defined as the change of protein expression levels between single cells within the same field of view, or variations in spatial distribution, defined as translocation of the protein between different subcellular compartments or independent regulation of the protein in two different compartments.

Cell cultivation

U2- OS FUCCI cells were developed and kindly provided by Dr. Miyawaki [42]. These cells are endogenously tagged with two fluorescent proteins fused to cell cycle regulators to allow cell cycle monitoring; CDT1 (mKO2-hCdt1^{+}) accumulates in G1 phase, while Geminin (mAG-hGem^{+}) accumulates in S and G2 phases. Cells expressing FUCCI probes are divided into red mKO2(+)mAG(-), yellow mKO2(+)mAG(+), and green mKO2(-)mAG(+) emitting populations. The cells were cultivated in Petri dishes at 37 °C in a 5.0 % CO_2 humidified environment in McCoy's 5A (modified) medium GlutaMAX supplement, (ThermoFisher, 36600021, MA, USA) supplemented with 10% fetal bovine serum (FBS, VWR, Radnor, PA, USA). The cells were maintained sub-confluent and harvested by trypsinization at log-phase growth (60% confluency) for subsequent analysis.

Live cell imaging

U-2 OS FUCCI cells were grown on a 96-well glass bottom plates (Whatman, Cat# 7716-2370, GE Healthcare, UK, and Greiner Sensoplate Plus, Cat# 655892, Greiner Bio-One, Germany). Approximately 6,000 cells were seeded in the wells and subjected to long-term time-lapse imaging using the molecular device instrument ImageXpress Micro XL (Molecular Device) high content screening equipped with a 20 x Plan Apo objective and supported with the MetaXpress software. Three Wavelenghts were acquired; W1 transmitted light, W2 FITC-3540C filter, W3 CY3-4040C filter. Images were collected every 30 minutes over a course of 72h.

Antibodies

The rabbit polyclonal antibodies used in this study (***Supplementary Table 8***) were generated within the HPA project. The antibodies were designed to target as many different isoforms of the target protein as possible and were affinity purified using

antigen fragments [96]. Furthermore, the antibodies were validated and quality assured for sensitivity and lack of cross-reactivity using the HPA standard quality assurance including microarray analyses.

Immunostaining

Immunostaining of the cells [97] was performed in 96-well glass bottom plates (Whatman, GE Healthcare, UK, and Greiner Sensoplate Plus, Greiner Bio-One, Germany) coated with 50 µl of 12.5 µg/ml human fibronectin (Sigma Aldrich, Darmstadt, Germany). Approximately 8,000 cells were seeded in each well and incubated at 37 ºC for 24 hours. After washing with Phosphatase Buffered Saline (PBS, PH=7), cells were fixed with 40 µl 4% ice cold PFA (Sigma Aldrich, Darmstadt, Germany) dissolved in growth medium supplemented with 10 % serum for 15 minutes and permeabilized with 40 µl 0.1% Triton X-100 (Sigma Aldrich) in PBS for 3x5 minutes. Rabbit polyclonal HPA antibodies targeting the proteins of interest were dissolved to 2-4 µg/ml in blocking buffer (PBS + 4% FBS) containing 1 µg/ml mouse anti-tubulin (Abcam, ab7291, Cambridge, UK). After washing with PBS, the diluted primary antibodies were added (40 µl/well) and the plates were incubated over night at 4 ºC. After overnight incubation, wells were washed with PBS for 3x10 minutes. Secondary antibodies, goat anti-mouse Alexa405 (A31553, ThermoFisher) and goat anti-rabbit Alexa647 (A21245, ThermoFisher) diluted to 2,5 µg /ml in blocking buffer were added and the plates were incubated for 90 minutes at room temperature. After washing with PBS, all wells were mounted with PBS containing 78 % glycerol before sealed.

Image acquisition

Image acquisition was performed using ImageXpress Micro XL (Molecular Device) high content screening equipped with a 40 x Plan Apo objective and supported with the MetaXpress software for automated acquisition. Images of the four channels were acquired at room temperature from six positions per sample. Four wavelengths were acquired; W1 for the microtubules DAPI-5060C filter, W2 FITC-3540C filter, W3 CY3-4040C filter and W4 CY5-4040C for the protein of interest. The images were unbinned with a pixel size of 0.1625x0.1625 µm.

Image processing and analysis

The segmentation of each cell was performed using the Cell Profiler software [98], where the overlay of the FUCCI tags were used for the nuclei identification and the microtubule staining was used for identification of the cell outline. Size exclusion was used to prune image mitotic cells from the population.

For each cell, the green and red tag mean intensity value was used and the cells were clustered in one of the cell cycle clusters using the Gaussian Clustering. The mean intensity of the target protein was measured in one of the three main compartments; nucleus, cytosol or cell, based on the a priori-known subcellular localization of the target protein from the HPA Cell Atlas.

Statistical analysis was performed using Kruskal-Wallis statistical test to determine the p-values that significantly differed between the three cell cycle groups. An arbitrary cut-off, based on a negative control, $p<0.01$ was chosen. FDR was calculated to adjust for multiple comparisons [99].The plots were generated using R studio v1.1.423 [100]. The image montages were created using Image J and FIJI [101]. k-means clustering was performed using the features kurtosis and skewness, where each gene was assigned to a specific K-cluster. The optimal number of clusters was chosen using the Elbow method, where it looks at the percentage of variance explained as a function of the number of clusters. The bimodal distribution of the protein expression was indicated by Hartigan's dip test.

Polar-coordinate pseudo time model

In this work we utilized the FUCCI system to model cell cycle position. To generate a continuous representation of cell cycle position we utilized a polar regression based on a log-scale scatter plot of GMNN (FUCCI-green) and CDT1 (FUCCI-red) where each point represents a single cell **(*Supplementary Figure S4)***. This data was shifted such that the origin point lay at the center of mass. This allowed us to use the fractional radius of the circle could be used to estimate time for each cell as traced by a ray from the origin generating a polar regression representing continuous cell cycle position. The cell-division point was selected by using the area of lowest cell density on the polar ray from the origin. This is justified by the knowledge that M phase (where cells express neither GMNN nor CDT1 highly) is much shorter than all other phases. The selected point was validated via visual inspection of nearby cells. This allowed us to linearize the progression of time from 0 to 1 representing the fractional distance along this polar axis from 0 to 360 degrees. This fit was done on a per-plate basis to account for batch-variance observed in the data.

Moving average model

Cell-cycle correlation was measured using a moving-average model within the linearized time from the polar fit described above. A range of window sizes were tested from 5-30. The analysis proved robust to this range of window size, and results

reported are for a window size of 20 cells which was chosen to balance the robustness to outliers with potentially destroying signal.

Percent explained variance

We used the metric percent explained variance to describe the goodness of our model fit. This metric is appealing as it is scale-invariant. That is, unlike a p-value significance metric which becomes more significant as sample size increases, the percent-variance converges to a stable solution as more cells are sampled. The percent explained variance is calculated as:

$$(1)\ \%\sigma_{prot} = 1 - \frac{\sigma_{residual}}{\sigma_{total}}$$

Here, σ_{polar} represents the variance of the protein of interest for an experiment and $\sigma_{residual}$ represents the variance remaining calculated from the moving average line along the pseudo-time axis.

Periodic regression model

To model protein response over time, a novel continuous-time periodic regression model was developed. This model made the following assumptions.

1. Protein expression is smoothly differentiable
2. Protein expression in continuously dividing cells must be periodic
3. Cell cycle-dependent protein expression shows a single peak as is commonly assumed for gene expression [102,103].

To model the asymmetric nature of protein accumulation and depletion over the cell cycle we developed a sin-based equation of fit describing the expression of protein χ over the cell cycle as seen in equation (2) below.

$$(2)\qquad f(x) = b \cdot \sin(\pi \cdot x^{\alpha})^{\gamma} + C$$

Where *b* describes the magnitude and sign of response, α describes the position of extremeum, γ defines the steepness of response, and *C* defines the y-intercept. Here we use π to define the single-extremum period 0-1 as represented by the normalized relative time since division. This function is fit to the normalized protein expression in the relevant meta-compartment where protein expression is observed (nucleus, cytoplasm, or both). Parameters of these functions are bounded to ensure reasonable differentiability as follows.

$$0 < b \leq 1$$

$$\frac{1}{6} < \alpha \leq 100$$

$$\frac{1}{2} < \gamma \leq 100$$

$$0 \leq C \leq 1$$

It is worth noting that these functions do not have a stable period and may behave erratically outside the defined 0-1 interval, however they are not designed to be evaluated outside this interval.

Gene set enrichment and interaction analysis

Functional enrichment analysis for the GO domain biological process was performed using the Database for Annotation, Visualization and Integrative Discovery (DAVID) tool [104] and Cytoscape v3.6.1 [105] was used for the network visualization. Enrichment map plugin was used to visualize the results of the highly significant gene-set enrichment as a network [106].

The interaction analysis was done using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database v10.5 [81], where a medium confidence (0.4) score was used to highlight the protein-protein interaction edges.

The open sources Cyclebase v3.0 [54]; Reactome and QuickGO [107] were used for downloading the previously characterized cell cycle regulators.

RNA extraction and RNA sequencing

The RNA extraction and sequencing were performed as previously reported [46,47,88]. Briefly, for cell lines early-split samples and duplicates were used for total RNA extraction. Tissue samples were embedded in Optimal Cutting Temperature compound and stored at –80°C. HE-stained frozen sections (4 µm) were prepared from each sample using a cryostat and the CryoJane® Tape-Transfer System (Instrumedics, St. Louis, MO, USA). Three sections (10 µm) were cut from each frozen tissue block and collected in a tube for subsequent RNA extraction [108]. Total RNA was extracted from the cell lines and tissue samples using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Only samples of high-quality RNA (RNA Integrity Number ≥7.5) were used in the following mRNA sample preparation for sequencing.

A total of 172 samples from 37 tissues and organs was sequenced using Illumina Hiseq2000 and Hiseq2500, and the standard Illumina RNAseq protocol with a read length of 2x100 bases. Briefly, the reads were mapped to the human genome (GRCh37) using Tophat v2.0.8b [109]. Transcript abundance estimation was performed using Kallisto v0.42.4 [110]. For each gene, the abundance was reported in 'Transcript Per Million' (TPM) as the sum of the TPM values of all its protein-coding transcripts. For each cell line and tissue type, the average TPM value for replicate samples was used as abundance score. The threshold level to detect presence of a transcript for a particular gene was set to ≥ 1 TPM.

Co-Expression Network Analysis

The co-expression networks for different tissues and cancer were downloaded from TCSBN website [87]. The nodes (genes) in the networks were classified into three categories: i) candidate cell-cycle genes (T1), ii) known cell-cycle genes (T2) and iii) other genes (T3). Following that, the shortest path in the co-expression network was compared between each category by using simple Breadth-First Search (BFS) method. The distribution between shortest path of T1-T2 was compared with T3-T2 by FDR-Adjusted Kolmogorov-Smirnov one-sided test (FDR < 0.05).

For the next step, we then incorporated the cancer pathology data from the HPA [88] into the cancer co-expression networks. The significant prognostic property (“favorable” or “unfavorable”) was mapped into the nodes of the networks. We then employed Louvain community detection algorithm[111] to identify the communities in the network, to maximize the modularity score. For each community, we calculated hypergeometric test to understand further the behavior of each community. A community was considered as showing specific behavior if it fulfilled p-value < 0.01. Each community was mapped into one of the four categories: i) Favorable, ii) Unfavorable, iii) Both, iv) Not significant.

The aforementioned analyses were performed with in-house Python script, with Scipy module[112] for the statistical analysis and Igraph[113] for the network analysis and manipulation.

Immunohistochemical staining

Immunohistochemical (IHC) staining of tissue microarray (TMA) sections and slide scanning were performed essentially as previously described [114]. In brief, normal and cancer tissues were derived from surgical material obtained from the Department of

Pathology, Uppsala University Hospital, Uppsala, Sweden as part of the sample collection governed by the Uppsala Biobank (http://www.uppsalabiobank.uu.se/en/). All human tissue samples used in the present study were anonymized in accordance with approval and advisory report from the Uppsala Ethical Review Board (Reference # 2002-577, 2005-338 and 2007-159). Representative tissue cores (1 mm diameter) were sampled from formalin fixed and paraffin embedded (FFPE) blocks and assembled into six TMAs, containing normal tissue samples from 144 individuals, as well as cancer tissue samples from 216 individuals. TMA blocks were cut in 4 µm thick sections using waterfall microtomes (Microm HM 355S, Thermo Fisher Scientific, Freemont, CA, USA), dried in RT overnight and baked in 50°C for 12-24 hours prior to IHC staining. Automated immunohistochemistry was performed using Autostainer 480® instruments (Lab Vision, Freemont, CA, USA), followed by slide scanning using Aperio AT2 (Leica Biosystems, Wetzlar, Germany). The high-resolution images of IHC stained TMA sections were evaluated and annotated by certified pathologists (Lab SurgPath, Mumbai, India).

References

1 Barnum, K. J. & O'Connell, M. J. in *Cell Cycle Control* 29-40 (Springer, 2014).

2 Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic acids research* **39**, D691-D697 (2010).

3 Crosby, M. E. Cell cycle: principles of control. *The Yale journal of biology and medicine* **80**, 141 (2007).

4 Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323-330 (1995).

5 Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131 (1995).

6 Nurse, P. A long twentieth century of the cell cycle and beyond. *Cell* **100**, 71-78 (2000).

7 Ly, T., Endo, A. & Lamond, A. I. Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**, e04534, doi:10.7554/eLife.04534 (2015).

8 Teixeira, L. K. & Reed, S. I. Ubiquitin ligases and cell cycle control. *Annual review of biochemistry* **82**, 387-414 (2013).

9 Eifler, K. & Vertegaal, A. C. SUMOylation-mediated regulation of cell cycle progression and cancer. *Trends in biochemical sciences* **40**, 779-793 (2015).

10 King, R. W., Deshaies, R. J., Peters, J.-M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274**, 1652-1659 (1996).

11 Skaar, J. R. & Pagano, M. Control of cell growth by the SCF and APC/C ubiquitin ligases. *Current opinion in cell biology* **21**, 816-824 (2009).

12 Reed, S. I. in *Cell Cycle Regulation* 147-181 (Springer, 2006).

13 Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews cancer* **9**, 153 (2009).

14 Massagué, J. G1 cell-cycle control and cancer. *Nature* **432**, 298 (2004).

15 Hartwell, L. H. & Kastan, M. B. Cell cycle control and cancer. *Science* **266**, 1821-1828 (1994).

16 Orlando, D. A. *et al.* Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* **453**, 944, doi:10.1038/nature06955 https://www.nature.com/articles/nature06955#supplementary-information (2008).

17 Cho, R. J. *et al.* Transcriptional regulation and function during the human cell cycle. *Nature Genetics* **27**, 48, doi:10.1038/83751 (2001).

18 Rustici, G. *et al.* Periodic gene expression program of the fission yeast cell cycle. *Nature genetics* **36**, 809 (2004).

19 Whitfield, M. L. *et al.* Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Molecular Biology of the Cell* **13**, 1977-2000, doi:10.1091/mbc.02-02-0030 (2002).

20 Boström, J. *et al.* Comparative cell cycle transcriptomics reveals synchronization of developmental transcription factor networks in cancer cells. *PloS one* **12**, e0188772 (2017).

21 Lane, K. R. *et al.* Cell cycle-regulated protein abundance changes in synchronously proliferating HeLa cells include regulation of pre-mRNA splicing proteins. *PloS one* **8**, e58456 (2013).

22 Ohta, S. *et al.* The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**, 810-821 (2010).

23 Ly, T. *et al.* A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife* **3** (2014).

24 Pagliuca, F. W. *et al.* Quantitative proteomics reveals the basis for the biochemical specificity of the cell-cycle machinery. *Molecular cell* **43**, 406-417 (2011).

25 Bar-Joseph, Z. *et al.* Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proceedings of the National Academy of Sciences* **105**, 955 (2008).

26 Dominguez, D. *et al.* A high-resolution transcriptome map of cell cycle reveals novel connections between periodic genes and cancer. *Cell Research* **26**, 946, doi:10.1038/cr.2016.84

https://www.nature.com/articles/cr201684#supplementary-information (2016).

27 Grant, G. D. *et al.* Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol Biol Cell* **24**, 3634-3650, doi:10.1091/mbc.E13-05-0264 (2013).

28 Peña-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 Research* **41**, 2846-2856, doi:10.1093/nar/gks1336 (2013).

29 Cooper, S. *et al.* Membrane-elution analysis of content of cyclins A, B1, and E during the unperturbed mammalian cell cycle. *Cell Division* **2**, 28, doi:10.1186/1747-1028-2-28 (2007).

30 Mgbonyebi, O. P., Russo, J. & Russo, I. H. Roscovitine induces cell death and morphological changes indicative of apoptosis in MDA-MB-231 breast cancer cells. *Cancer research* **59**, 1903-1910 (1999).

31 Wright, J. Morphology and growth rate changes in Chinese hamster cells cultured in presence of sodium butyrate. *Experimental cell research* **78**, 456-460 (1973).

32 Maltese, W. A. & Sheridan, K. M. Isoprenylated proteins in cultured cells: subcellular distribution and changes related to altered morphology and growth arrest induced by mevalonate deprivation. *Journal of cellular physiology* **133**, 471-481 (1987).

33 Davis, P. K., Ho, A. & Dowdy, S. F. Biological methods for cell-cycle synchronization of mammalian cells. *Biotechniques* **30**, 1322-1331 (2001).

34 Cooper, S. Rethinking synchronization of mammalian cells for cell cycle analysis. *Cellular and Molecular Life Sciences CMLS* **60**, 1099-1106, doi:10.1007/s00018-003-2253-2 (2003).

35 Cooper, S., Iyer, G., Tarquini, M. & Bissett, P. Nocodazole does not synchronize cells: implications for cell-cycle control and whole-culture synchronization. *Cell and Tissue Research* **324**, 237-242, doi:10.1007/s00441-005-0118-8 (2006).

36 Cooper, S. & Shedden, K. Microarray analysis of gene expression during the cell cycle. *Cell & Chromosome* **2**, 1, doi:10.1186/1475-9268-2-1 (2003).

37 Karlsson, J., Kroneis, T., Jonasson, E., Larsson, E. & Ståhlberg, A. Transcriptomic Characterization of the Human Cell Cycle in Individual Unsynchronized Cells. *Journal of molecular biology* **429**, 3909-3924 (2017).

38 Liu, Z. *et al.* Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nature communications* **8**, 22 (2017).

39 Scialdone, A. *et al.* Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54-61 (2015).

40 Sigal, A. *et al.* Dynamic proteomics in individual human cells uncovers widespread cell-cycle dependence of nuclear proteins. *Nature Methods* **3**, 525 (2006).

41 Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721 (2010).

42 Sakaue-Sawano, A. *et al.* Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487-498 (2008).

43 Mukherji, M. *et al.* Genome-wide functional analysis of human cell-cycle regulators. *Proceedings of the National Academy of Sciences* **103**, 14819-14824 (2006).

44 Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L. & Liberali, P. Trajectories of cell-cycle progression from fixed cell populations. *Nature methods* **12**, 951 (2015).

45 Gookin, S. *et al.* A map of protein dynamics during cell-cycle progression and cell-cycle exit. *PLoS biology* **15**, e2003268 (2017).

46 Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).

47 Uhlen, M. *et al.* Towards a knowledge-based human protein atlas. *Nature biotechnology* **28**, 1248 (2010).

48 Brown, N. R. *et al.* Cyclin B and cyclin A confer different substrate recognition properties on CDK2. *Cell Cycle* **6**, 1350-1359 (2007).

49 Li, X. *et al.* Direct association with inner centromere protein (INCENP) activates the novel chromosomal passenger protein, Aurora-C. *Journal of Biological Chemistry* **279**, 47201-47211 (2004).

50 Wang, X., Yang, Y. & Dai, W. Differential subcellular localizations of two human Sgo1 isoforms: implications in regulation of sister chromatid cohesion and microtubule dynamics. *Cell Cycle* **5**, 636-641 (2006).

51 Neef, R. *et al.* Phosphorylation of mitotic kinesin-like protein 2 by polo-like kinase 1 is required for cytokinesis. *The Journal of cell biology* **162**, 863-876 (2003).

52 Gangisetty, O., Lauffart, B., Sondarva, G. V., Chelsea, D. M. & Still, I. H. The transforming acidic coiled coil proteins interact with nuclear histone acetyltransferases. *Oncogene* **23**, 2559 (2004).

53 Botstein, D. *et al.* Gene Ontology: tool for the unification of biology. *Nat genet* **25**, 25-29 (2000).

54 Santos, A., Wernersson, R. & Jensen, L. J. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic acids research* **43**, D1140-D1144 (2014).

55 Vong, Q. P., Cao, K., Li, H. Y., Iglesias, P. A. & Zheng, Y. Chromosome alignment and segregation regulated by ubiquitination of survivin. *Science* **310**, 1499-1504 (2005).

56 Wolfson, R. L. *et al.* KICSTOR recruits GATOR1 to the lysosome and is necessary for nutrients to regulate mTORC1. *Nature* **543**, 438 (2017).

57 Kikuchi, K., Niikura, Y., Kitagawa, K. & Kikuchi, A. Dishevelled, a Wnt signalling component, is involved in mitotic progression in cooperation with Plk1. *The EMBO journal* **29**, 3470-3483 (2010).

794 58 Fang, L., Seki, A. & Fang, G. SKAP associates with kinetochores and promotes
795 the metaphase-to-anaphase transition. *Cell cycle* **8**, 2819-2827 (2009).
796 59 Rapley, J. *et al.* The NIMA-family kinase Nek6 phosphorylates the kinesin Eg5
797 at a novel site necessary for mitotic spindle formation. *Journal of cell science*
798 **121**, 3912-3921 (2008).
799 60 Demetriou, M., Granovsky, M., Quaggin, S. & Dennis, J. W. Negative regulation
800 of T-cell activation and autoimmunity by Mgat5N-glycosylation. *Nature* **409**,
801 733 (2001).
802 61 Robson, T. & James, I. F. The therapeutic and diagnostic potential of FKBPL;
803 a novel anticancer protein. *Drug discovery today* **17**, 544-548 (2012).
804 62 Zielke, N. & Edgar, B. FUCCI sensors: powerful new tools for analysis of cell
805 proliferation. *Wiley Interdisciplinary Reviews: Developmental Biology* **4**, 469-
806 487 (2015).
807 63 Suzuki, C. *et al.* ANLN plays a critical role in human lung carcinogenesis
808 through the activation of RHOA and by involvement in the phosphoinositide 3-
809 kinase/AKT pathway. *Cancer research* **65**, 11314-11325 (2005).
810 64 Patterson, K. I., Brummer, T., O'brien, P. M. & Daly, R. J. Dual-specificity
811 phosphatases: critical regulators with diverse cellular targets. *Biochemical*
812 *Journal* **418**, 475-489 (2009).
813 65 O'Hagan, S., Muelas, M. W., Day, P. J., Lundberg, E. & Kell, D. B. GeneGini:
814 Assessment via the Gini Coefficient of Reference “Housekeeping” Genes and
815 Diverse Human Transporter Expression Profiles. *Cell systems* **6**, 230-244.
816 e231 (2018).
817 66 Phannasil, P. *et al.* Pyruvate carboxylase is up-regulated in breast cancer and
818 essential to support growth and invasion of MDA-MB-231 cells. *PloS one* **10**,
819 e0129848 (2015).
820 67 Sellers, K. *et al.* Pyruvate carboxylase is critical for non–small-cell lung cancer
821 proliferation. *The Journal of clinical investigation* **125**, 687-698 (2015).
822 68 Christen, S. *et al.* Breast cancer-derived lung metastases show increased
823 pyruvate carboxylase-dependent anaplerosis. *Cell reports* **17**, 837-848 (2016).
824 69 Wang, J. *et al.* Identification of XAF1 as a novel cell cycle regulator through
825 modulating G 2/M checkpoint and interaction with checkpoint kinase 1 in
826 gastrointestinal cancer. *Carcinogenesis* **30**, 1507-1516 (2009).
827 70 Semple, J. W. *et al.* An essential role for Orc6 in DNA replication through
828 maintenance of pre-replicative complexes. *The EMBO journal* **25**, 5150-5158
829 (2006).

71 Izumi, M. *et al.* The Mcm2-7-interacting domain of human mini-chromosome maintenance 10 (Mcm10) protein is important for stable chromatin association and origin firing. *Journal of Biological Chemistry*, jbc. M117. 779371 (2017).

72 Li, Y. *et al.* ZNF32 inhibits autophagy through the mTOR pathway and protects MCF-7 cells from stimulus-induced cell death. *Scientific reports* **5**, 9288 (2015).

73 Li, J. *et al.* ZNF32 contributes to the induction of multidrug resistance by regulating TGF-β receptor 2 signaling in lung adenocarcinoma. *Cell death & disease* **7**, e2428 (2016).

74 Wisdom, R., Johnson, R. S. & Moore, C. c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *The EMBO journal* **18**, 188-197 (1999).

75 Pils, D. *et al.* Cyclin E1 (CCNE1) as independent positive prognostic factor in advanced stage serous ovarian cancer patients–A study of the OVCAD consortium. *European journal of cancer* **50**, 99-110 (2014).

76 St-Denis, N. *et al.* Phenotypic and interaction profiling of the human phosphatases identifies diverse mitotic regulators. *Cell reports* **17**, 2488-2501 (2016).

77 Shin, H.-J. *et al.* Dual roles of human BubR1, a mitotic checkpoint kinase, in the monitoring of chromosomal instability. *Cancer cell* **4**, 483-497 (2003).

78 Xu, Y. *et al.* Effect of estrogen sulfation by SULT 1 E 1 and PAPSS on the development of estrogen-dependent cancers. *Cancer science* **103**, 1000-1009 (2012).

79 Figaro, S., Scrima, N., Buckingham, R. H. & Heurgué-Hamard, V. HemK2 protein, encoded on human chromosome 21, methylates translation termination factor eRF1. *FEBS letters* **582**, 2352-2356 (2008).

80 Javaheri, T. *et al.* Increased survival and cell cycle progression pathways are required for EWS/FLI1-induced malignant transformation. *Cell death & disease* **7**, e2419 (2016).

81 Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research* **43**, D447-D452 (2014).

82 Hutterer, A., Glotzer, M. & Mishima, M. Clustering of centralspindlin is essential for its accumulation to the central spindle and the midbody. *Current biology* **19**, 2043-2049 (2009).

83 Fischer, M. *et al.* p53 and cell cycle dependent transcription of kinesin family member 23 (KIF23) is controlled via a CHR promoter element bound by DREAM and MMB complexes. *PloS one* **8**, e63187 (2013).

84 Wang, Y. & Zhan, Q. Cell cycle-dependent expression of centrosomal ninein-like protein in human cells is regulated by the anaphase-promoting complex. *Journal of Biological Chemistry* **282**, 17712-17719 (2007).

85 Morita, H. *et al.* KIF20A, highly expressed in immature hematopoietic cells, supports the growth of HL60 cell line. *International journal of hematology* **108**, 607-614 (2018).

86 Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nature genetics* **45**, 1113 (2013).

87 Lee, S. *et al.* TCSBN: a database of tissue and cancer specific biological networks. *Nucleic acids research* **46**, D595-D600 (2017).

88 Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).

89 Yang, X.-M. *et al.* Overexpression of Rac GTPase activating protein 1 contributes to proliferation of cancer cells by reducing hippo signaling to promote cytokinesis. *Gastroenterology* **155**, 1233-1249. e1222 (2018).

90 Hirose, K., Kawashima, T., Iwamoto, I., Nosaka, T. & Kitamura, T. MgcRacGAP is involved in cytokinesis through associating with mitotic spindle and midbody. *Journal of Biological Chemistry* **276**, 5821-5828 (2001).

91 Schneider, M. A. *et al.* AURKA, DLGAP5, TPX2, KIF11 and CKAP5: Five specific mitosis-associated genes correlate with poor prognosis for non-small cell lung cancer patients. *International journal of oncology* **50**, 365-372 (2017).

92 Vanneste, D., Ferreira, V. & Vernos, I. (Portland Press Limited, 2011).

93 Mosquera, J. M. *et al.* Concurrent AURKA and MYCN gene amplifications are harbingers of lethal treatmentrelated neuroendocrine prostate cancer. *Neoplasia* **15**, IN1-IN4 (2013).

94 Zhen, Y.-Y., Libotte, T., Munck, M., Noegel, A. A. & Korenbaum, E. NUANCE, a giant protein connecting the nucleus and actin cytoskeleton. *J Cell Sci* **115**, 3207-3222 (2002).

95 Li, S. *et al.* HDAC2 regulates cell proliferation, cell cycle progression and cell apoptosis in esophageal squamous cell carcinoma EC9706 cells. *Oncology letters* **13**, 403-409 (2017).

96 Nilsson, P. *et al.* Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics* **5**, 4327-4337 (2005).

97 Stadler, C., Skogs, M., Brismar, H., Uhlén, M. & Lundberg, E. A single fixation protocol for proteome-wide immunofluorescence localization studies. *Journal of proteomics* **73**, 1067-1078 (2010).

98 Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology* **7**, R100 (2006).

99 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300 (1995).

100 Team, R. C. R: A language and environment for statistical computing. (2013).

101 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671 (2012).

102 de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *science* **307**, 724-727 (2005).

103 Spellman, P. T. *et al.* Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell* **9**, 3273-3297 (1998).

104 Sartor, M. A. *et al.* ConceptGen: a gene set enrichment and gene set relation mapping tool. *Bioinformatics* **26**, 456-463 (2009).

105 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

106 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS one* **5**, e13984 (2010).

107 Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045-3046 (2009).

108 Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

109 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

110 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**, 525 (2016).

111 Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008 (2008).

112 Jones, E., Oliphant, T. & Peterson, P. {SciPy}: open source scientific tools for {Python}. (2014).

113 Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9 (2006).

114 Kampf, C., Olsson, I., Ryberg, U., Sjöstedt, E. & Pontén, F. Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *Journal of visualized experiments: JoVE* (2012).

Acknowledgements

We acknowledge the entire staff of the Human Protein Atlas program. We acknowledge Dr. Hisao Masai (Tokyo Metropolitan Institute of Medical Science) for providing the stable U-2 OS FUCCI cell line. We also acknowledge Dr. Magdalena Otrocka from the National Laboratories for Chemical Biology at Karolinska Institutet (LCBKI) for access to imaging infrastructure; Dr. Petter Ranefall and Dr. Carolina Wählby for providing support for establishing the Cell Profiler pipeline. Funding was provided by the Knut and Alice Wallenberg Foundation (2016.0204) and Swedish Research Council (2017-05327) to E.L. The data is available for download in the supplementary material. Images and Cell Atlas transcriptome and proteome data is available in the Human Protein Atlas (www.proteinatlas.org/humancell).

Life Sciences Reporting Summary.

Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability statement.

The images from the Human Protein Atlas are available at: https://www.proteinatlas.org. The images from the FUCCI screening is available upon request, and will be made publicly available in the Human Protein Atlas database in the release of version 19. The RNA-sequencing data is available at www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836/

Code availability statement.

The cell profiler pipeline for image analysis and the code for generating the polar-coordinate pseudotime model and the periodic regression model are available at: https://github.com/CellProfiling/fucci_screen

Author contributions

E.L. conceived the study. D.M., D.P.S. and E.L. developed the methodology for the study. D.M., L.Å., R.S., C.G, and P.T. carried out the experimental work and contributed to the cell atlas implementation. D.M., D.P.S. and E.L. carried out data analysis and investigation. F.D. analyzed the RNA-seq data, M.A., C.Z. and A.M. carried out analysis for the co-expression network analysis and generated the corresponding figures. C.L. and F.P. provided the tissue data. D.M. and E.L. wrote the manuscript. B.A., D.P.S, O.C and P.T revised the manuscript. D.M and D.P.S. created

975 the figures. M.U. initiated the HPA project and provided antibodies. E.L. supervised
976 and administered the project and acquired funding. All authors reviewed and approved
977 the final manuscript.
978
979 **Competing interests**
980 The authors declare that they have no conflict of interest.
981
982 **Corresponding author**
983 Correspondence and requests for materials should be addressed to E.L.

Figure legends

Figure 1: Temporal dissection of cell-to-cell heterogeneity of the human proteome

In A-D the target protein is shown in green, microtubules in red and the nucleus in blue. The scalebars in A-F represents 10µm.

A: Example images of proteins with observed cell-to-cell heterogeneity in immunostained U-2 OS cells in terms of variation in protein abundance (CCNB1) and in spatial distribution (MRTO4) respectively.

B: The RACGAP1 protein shows the same type of cell-to-cell heterogeneity in several different cell types (U-2 OS, A-431 and MCF7).

C: Example images of proteins localized to one of the mitotic substructures (Kinetochores, Cytokinetic bridge, Cleavage furrow, Mitotic spindle, Midbody ring and Midbody). INCENP localized to kinetochores in MCF-7 cells, SGO1, KIF20A and TACC3 localized to the kinetochores, the cleavage furrow and the mitotic spindle in U-2 OS cells, respectively.

D: Proteins localized to the cytokinetic bridge (BIRC5, GLI4, C12orf66) midbody ring (DVL3), and mitotic spindle (KIF11, KNSTRN, MGAT5B and FKBPL) in U-2 OS cells.

E: U-2 OS FUCCI cells allow monitoring the cell cycle by expressing two fluorescently-tagged cell cycle markers, CDT1 expressed during G1 phase (red) and Geminin expressed during S and G2 phases (green) and their co-expression during G1/S transition (yellow). Intensity map of the FUCCI cells defined in three clusters representing G1, G1/S and SG2 phases by Gaussian clustering. The polar coordinate model transfers the FUCCI marker information into a linear model of pseudo-time.

F: Examples images of the analyzed proteins ANLN, FAM171F1, DUSP18 and alpha-tubulin (MT) as negative control combined with their respective boxplot, intensity plot and expression profile. In the boxplots the cells expressing the different markers (G1, G1S and SG2) are grouped and the mean intensity of the target protein is plotted. Kruskal- Wallis statistical test was used to check the significance variation across the different groups. In the intensity plot, the cells corresponding to the specific target protein is highlighted using a gradient color code of the mean intensity of the target.

Figure 2: Variation distribution and organelle proteomes

A: Gene ontology (BP) based enrichment analysis for cell cycle regulated proteins showing significantly enriched terms for the domain biological process. Each node represents a GO term and edge size corresponds to the number of genes that overlap between the two connected gene sets.

B: Scatterplot showing the three different clusters generated by K mean clustering based on Kurtosis and skewness as features for the cell cycle regulated proteins (dark blue) and the ones not correlated to cell cycle (grey).
Violin-plots and histograms showing the distinct distributions of the normalized mean intensity of each cell per protein of selected examples (GTA6; CCNB1 and DEF6).
C: Scatterplot of percentage explained variance and Gini index for each investigated protein color coded by -log10(FDR).
D: Bar plot showing the distribution of the cell cycle regulated proteins (dark blue) and the ones not correlated to cell cycle (grey) proteins to the different subcellular compartments. Asterisk marks statistically significant deviations from the mapped human proteome ($p<0.01$) based on a binomial test.
E: Examples of cell cycle correlated proteins localized to the different subcellular structures respectively: Cytosol, Mitochondria, Nucleus, Nucleoli, Nuclear sub-compartments and Secretory pathway. The scalebar represents 10µm. The target protein is shown in green and microtubules in red.

Figure 3: Temporal profiles of the cell cycle regulated human proteome
A: Heat map of the cell cycle regulated proteins showing the relative expression levels of the protein across the cell cycle. Yellow represents high expression level and blue represents low expression levels. The heatmap is sorted by the timepoint of their peak of expression.
B: Examples of selected cell cycle regulated proteins peaking in different phases of the cell cycle. ORC6 peaking in G1, DUSP19 peaking end of G1, BUB1B, DPH2 and FLI1 peaking in S&G2 phases.
C: Protein-Protein interactions network plot of the 464 CCD proteins using the STRING database. The proteins with a known association to the cell cycle (GO BP terms) are shown as squares.

Figure 4: Gene expression across normal and cancer tissues
A: Hierarchical clustering of transcript levels (TPM values) for the cell cycle regulated proteins derived from bulk RNA sequencing of various normal and cancer tissue types. The expression level of the proliferation markers MCM6, CDK1, PCNA, MCM2 and KI67 is highlighted on top, as a general measure of the proliferative activity of the tissues. Four clusters are identified; Cluster 1 contains normal tissues with low proliferative activity, 2 contains cerebral tissues with testis, 3 contains mostly normal tissues with midrange expression level of the proliferation markers and 4 contains tissues with high expression of the proliferation markers, including tumors.

B: Box plots of the average transcript level corresponding to the cell cycle regulated proteins for the four different clusters from A.

Figure 5: Co-expression networks of the cell cycle regulated proteome

A: Bar plot showing the path distance from gene co-expression networks between novel cell cycle proteins and previously known cell cycle proteins in different normal and cancer tissues.

B: Co-expression network analysis of the cell cycle regulated proteins in pancreatic, breast and colorectal cancer. The network is clustered into communities using mathematical models. Each community has been classified as favorable (green), unfavorable (red) or both based on an enrichment / hypergeometric analysis.

Figure 6: Novel cell cycle regulated proteins as potential clinical biomarkers

A: Kaplan-Meier plots showing the correlation between survival and gene expression (FPKM) for four cell cycle regulated proteins. For RACGAP1 and DLGAP5 a high expression was associated to a shorter survival (unfavorable), whereas for SYNE2 and FAM50B a high expression was associated to a longer survival (favorable). Purple and blue lines show high and low expression, respectively.

B: Images of immunohistochemically stained proteins in normal tissue. RACGAP1 in breast, DLGAP5 in pancreas, SYNE2 and FAM50B in kidney. The target protein is shown in brown and the nuclei in blue.

C: Images of immunohistochemically stained proteins in the corresponding tumor tissue as to in B. RACGAP1 in breast cancer, DLGAP5 in pancreatic cancer, SYNE2 and FAM50B in renal cancer. The target protein is shown in brown and the nuclei in blue.

D: Temporal interphase expression profile of RACGAP1, DLGAP5, SYNE2 and the localization of FAM50B to the Cytokinetic bridge during mitosis.

E: Interaction networks for each of the proteins, using a medium confidence score with a minimum interaction score of 0.4 and showing not more than 10 interactors.

Figure 1

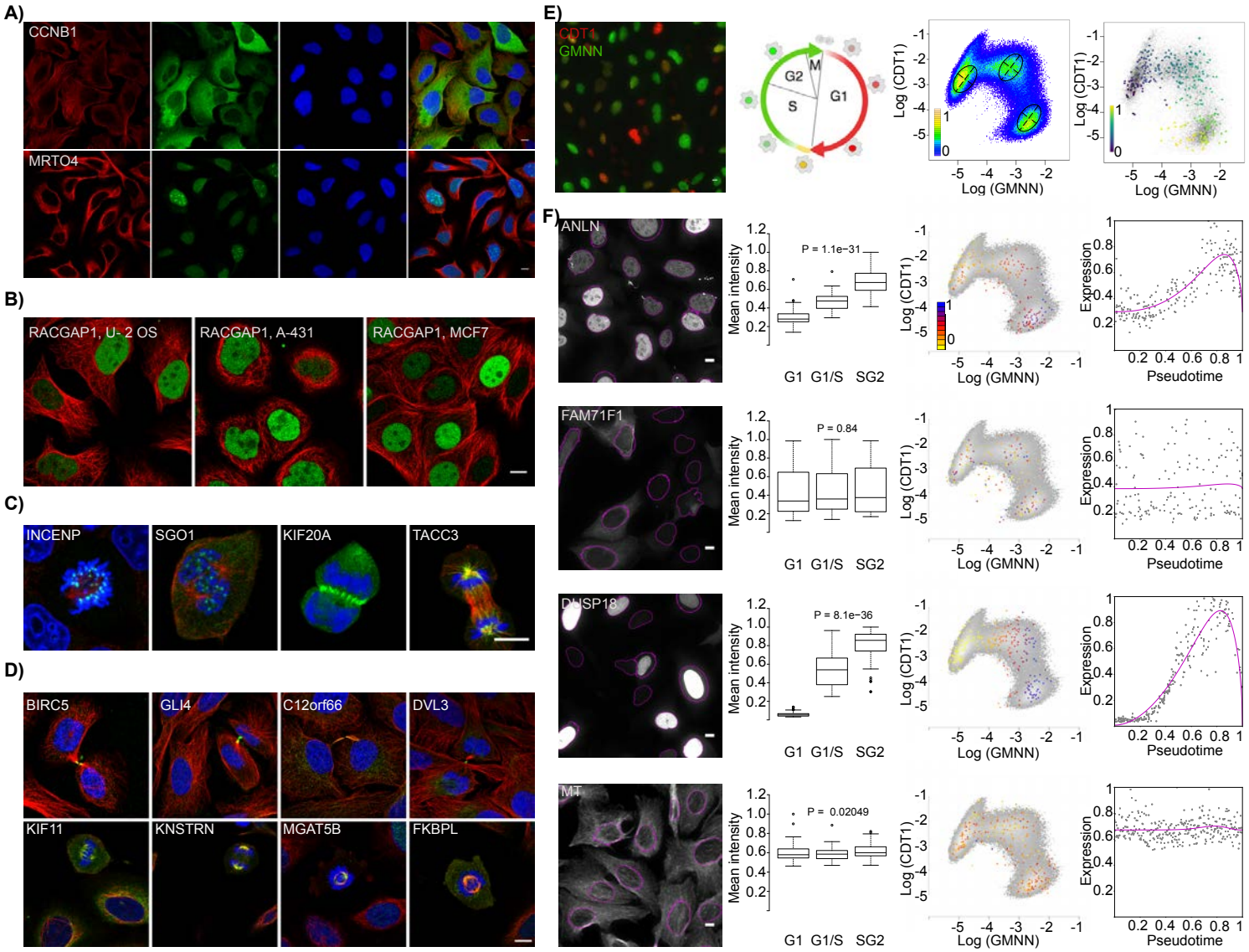

A)
CCNB1
MRTO4
B)
RACGAP1, U- 2 OS
RACGAP1, A-431
RACGAP1, MCF7
C)
INCENP
SGO1
KIF20A
TACC3
D)
BIRC5
GLI4
C12orf66
DVL3
KIF11
KNSTRN
MGAT5B
FKBPL
E)
CDT1
GMNN
G2
M
S
G1
Log (CDT1)
Log (GMNN)
F)
ANLN
FAM71F1
DUSP18
MT
Mean intensity
P = 1.1e−31
P = 0.84
P = 8.1e−36
P = 0.02049
G1 G1/S SG2
Expression
Pseudotime

Figure 2

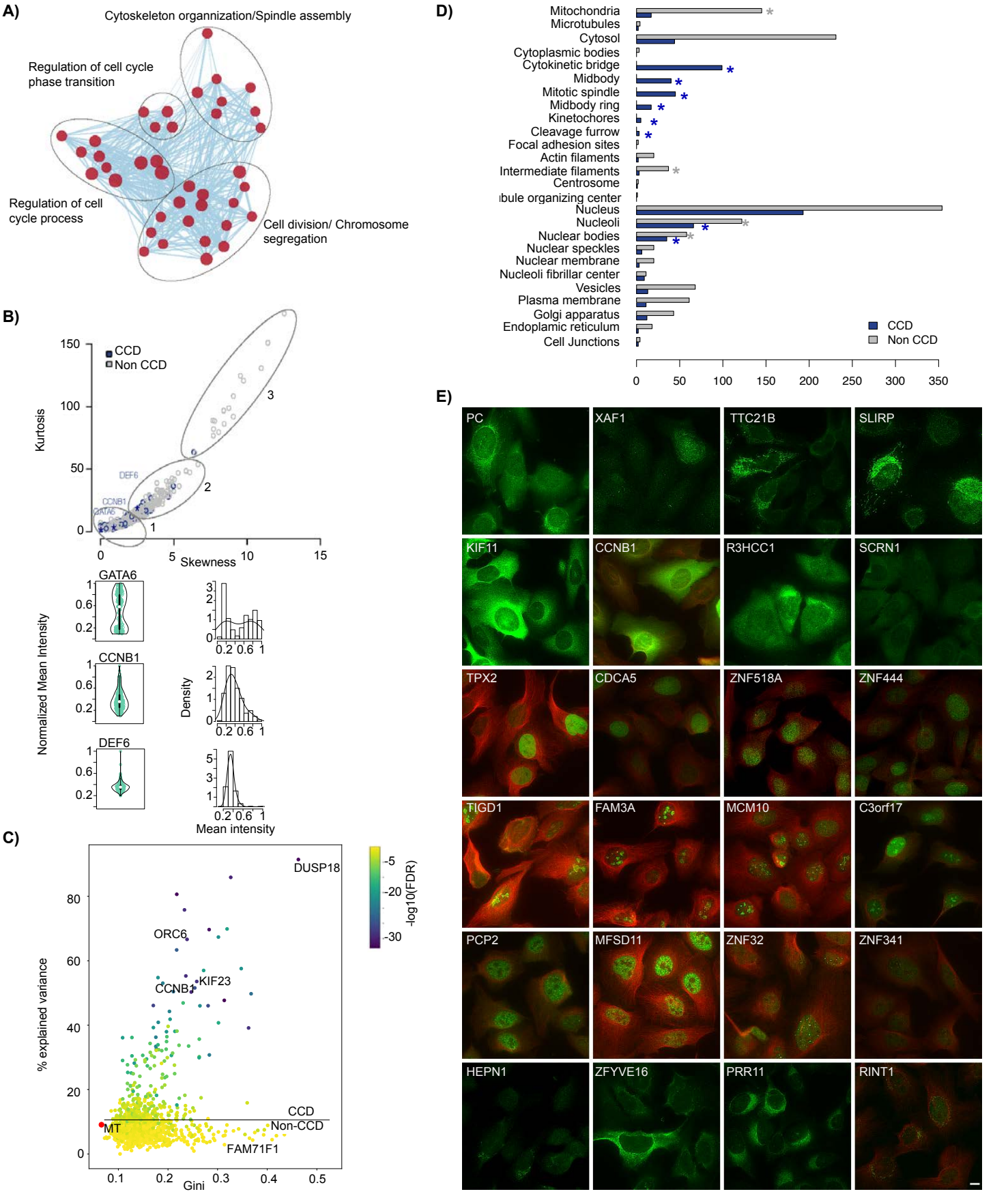
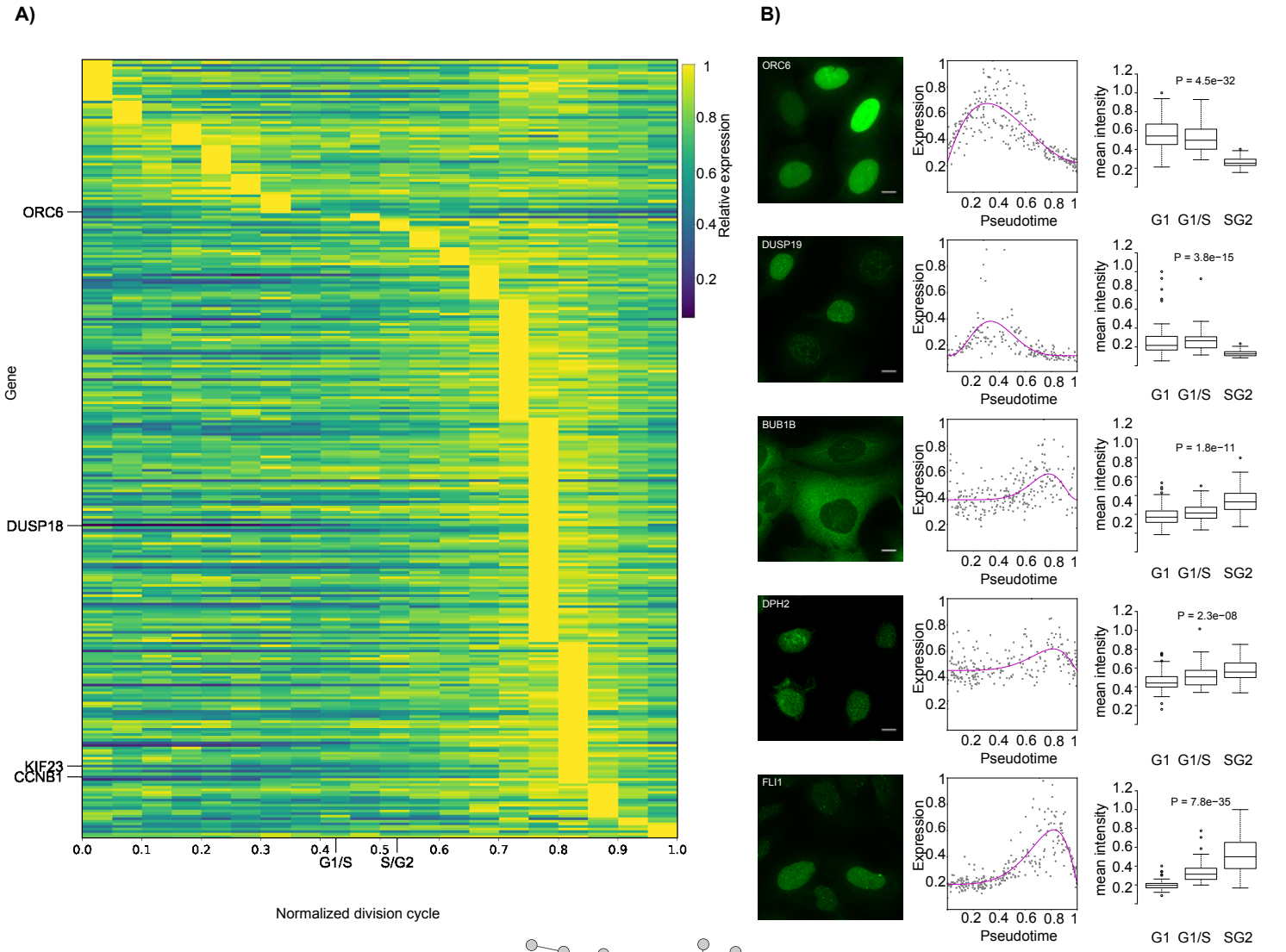

A)
Cytoskeleton organnization/Spindle assembly
Regulation of cell cycle phase transition
Regulation of cell cycle process
Cell division/ Chromosome segregation
B)
CCD
Non CCD
Kurtosis
Skewness
DEF6
CCNB1
GATA6
Normalized Mean Intensity
Density
Mean intensity
C)
DUSP18
ORC6
CCNB1
KIF23
MT
FAM71F1
CCD
Non-CCD
% explained variance
Gini
-log10(FDR)
D)
Mitochondria
Microtubules
Cytosol
Cytoplasmic bodies
Cytokinetic bridge
Midbody
Mitotic spindle
Midbody ring
Kinetochores
Cleavage furrow
Focal adhesion sites
Actin filaments
Intermediate filaments
Centrosome
Nucleus
Nucleoli
Nuclear bodies
Nuclear speckles
Nuclear membrane
Nucleoli fibrillar center
Vesicles
Plasma membrane
Golgi apparatus
Endoplamic reticulum
Cell Junctions
CCD
Non CCD
E)
PC
XAF1
TTC21B
SLIRP
KIF11
CCNB1
R3HCC1
SCRN1
TPX2
CDCA5
ZNF518A
ZNF444
TIGD1
FAM3A
MCM10
C3orf17
PCP2
MFSD11
ZNF32
ZNF341
HEPN1
ZFYVE16
PRR11
RINT1

Figure 3


A)
B)
ORC6
DUSP18
KIF23
CCNB1
Gene
Relative expression
Normalized division cycle
G1/S
S/G2
DUSP19
BUB1B
DPH2
FLI1
Expression
Pseudotime
mean intensity
P = 4.5e−32
P = 3.8e−15
P = 1.8e−11
P = 2.3e−08
P = 7.8e−35
G1
G1/S
SG2

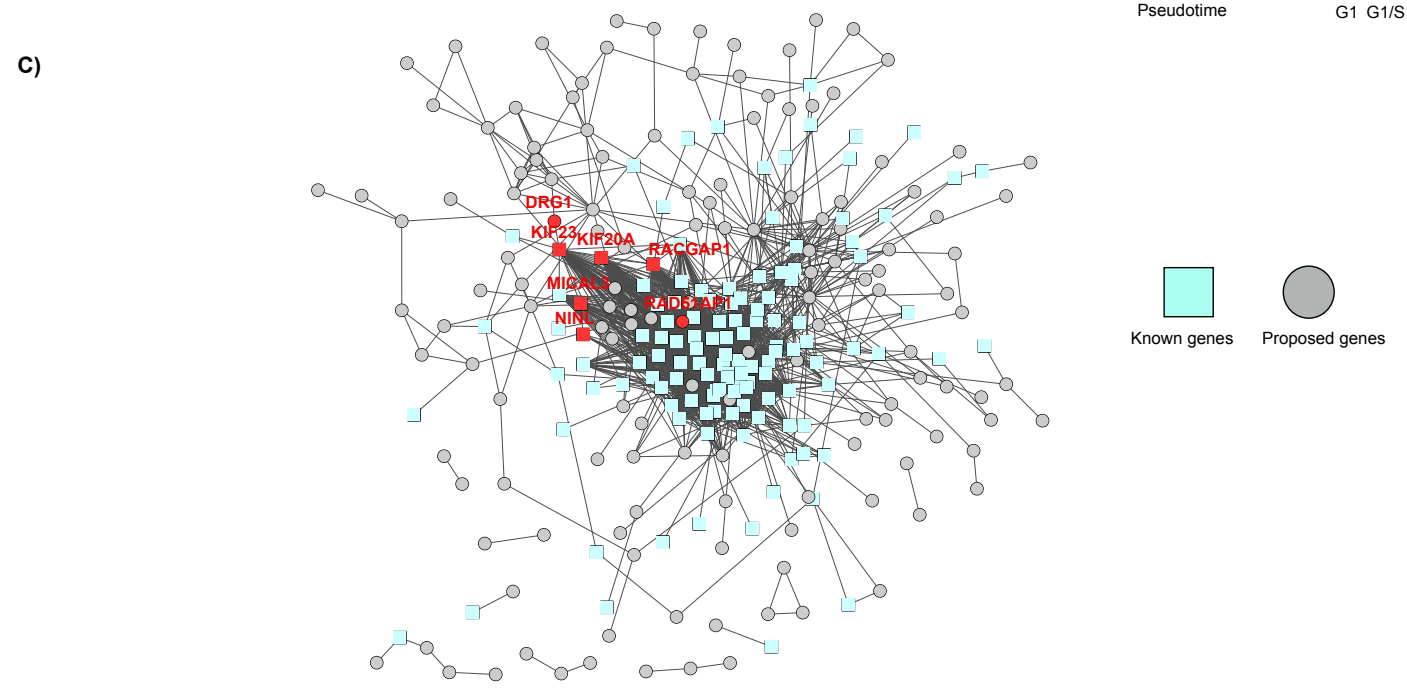
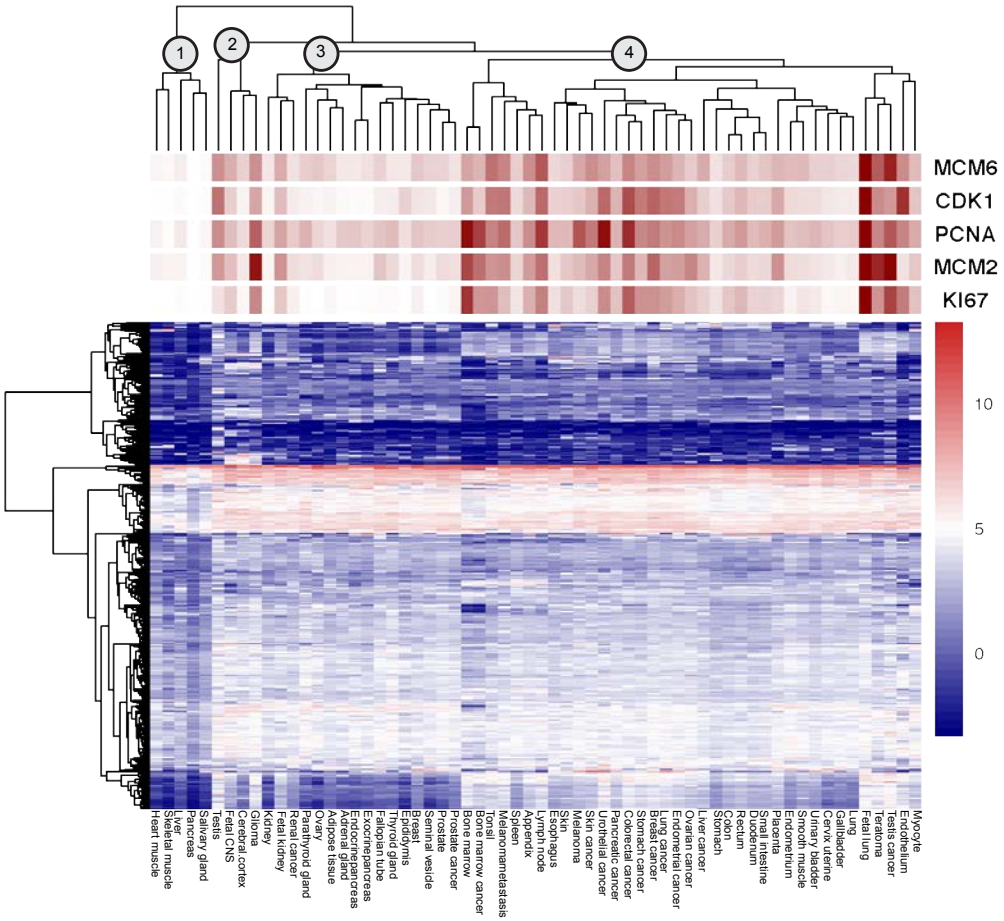

C)
DRG1
KIF23
KIF20A
RACGAP1
MICAL3
RAD51AP1
NIN
Known genes
Proposed genes

Figure 4

A)


1
2
3
4
MCM6
CDK1
PCNA
MCM2
KI67
10
5
0
Heart muscle
Skeletal muscle
Liver
Pancreas
Salivary gland
Testis
Fetal CNS
Cerebral cortex
Glioma
Kidney
Fetal kidney
Renal cancer
Parathyroid gland
Ovary
Adipose tissue
Adrenal gland
Endocrinepancreas
Exocrinepancreas
Fallopian tube
Thyroid gland
Epididymis
Bladder
Seminal vesicle
Prostate
Prostate cancer
Bone marrow
Bone marrow cancer
Tonsil
Melanoma metastasis
Spleen
Appendix
Lymph node
Skin
Esophagus
Melanoma
Skin cancer
Urothelial cancer
Pancreatic cancer
Colorectal cancer
Stomach cancer
Breast cancer
Lung cancer
Endometrial cancer
Ovarian cancer
Liver cancer
Colon
Stomach
Rectum
Duodenum
Small intestine
Placenta
Endometrium
Smooth muscle
Urinary bladder
Cervix uterine
Gallbladder
Lung
Lung lung
Teratoma
Testis cancer
Endothelium
Myocyte

B)

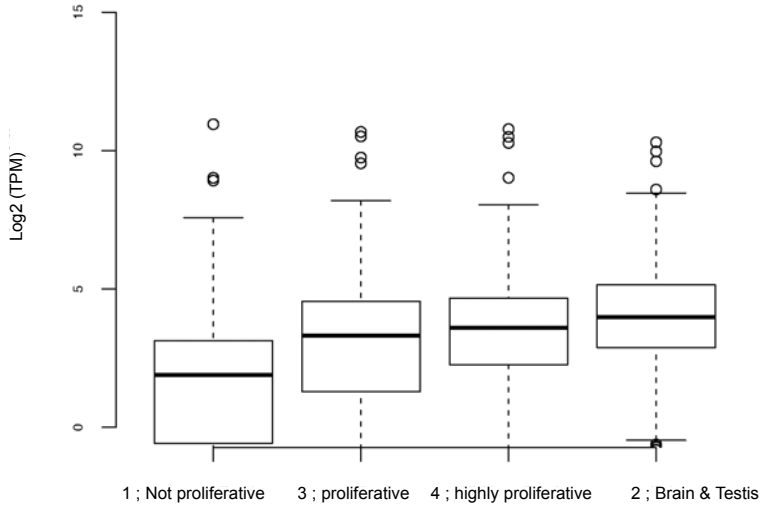

Log2 (TPM)
0
5
10
15
1 ; Not proliferative
3 ; proliferative
4 ; highly proliferative
2 ; Brain & Testis

Figure 5

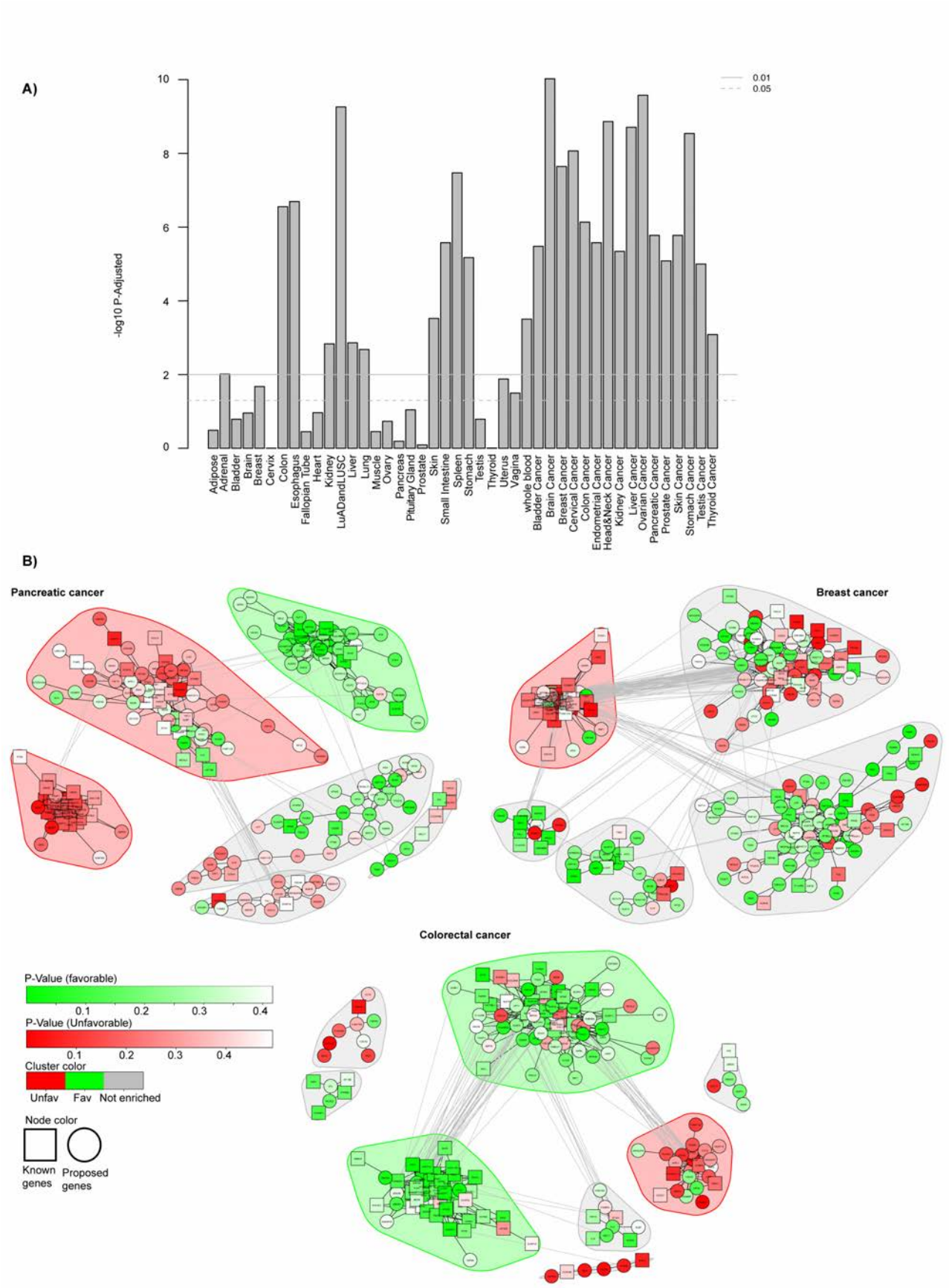
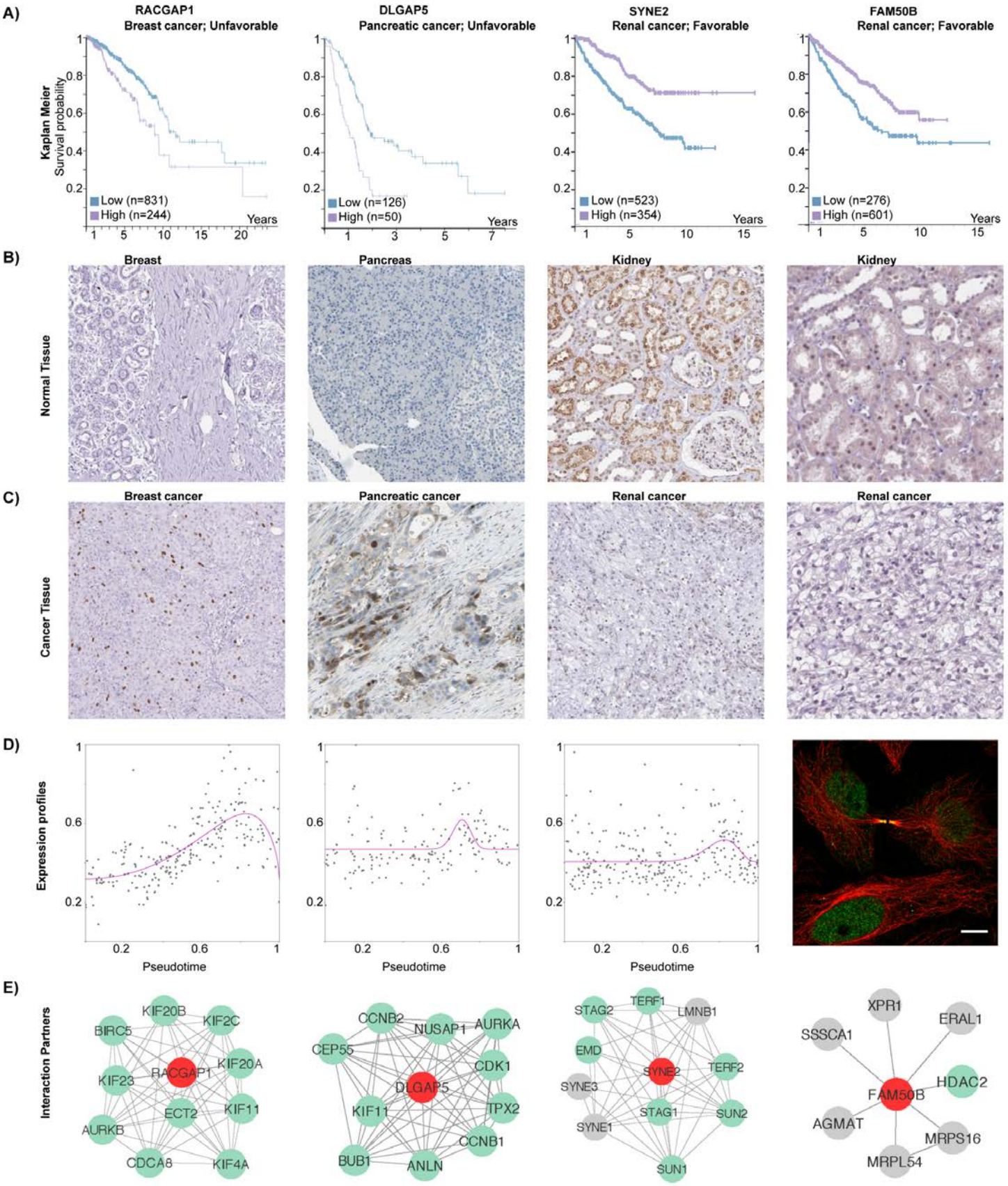

A)
-log10 P-Adjusted
0.01
0.05
Adipose
Adrenal
Bladder
Brain
Breast
Cervix
Colon
Esophagus
Fallopian Tube
Heart
Kidney
LuADandLUSC
Liver
Lung
Muscle
Ovary
Pancreas
Pituitary Gland
Prostate
Skin
Small Intestine
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
whole blood
Bladder Cancer
Brain Cancer
Breast Cancer
Cervical Cancer
Colon Cancer
Endometrial Cancer
Head&Neck Cancer
Kidney Cancer
Liver Cancer
Ovarian Cancer
Pancreatic Cancer
Prostate Cancer
Skin Cancer
Stomach Cancer
Testis Cancer
Thyroid Cancer
B)
Pancreatic cancer
Breast cancer
Colorectal cancer
P-Value (favorable)
P-Value (Unfavorable)
Cluster color
Unfav
Fav
Not enriched
Node color
Known genes
Proposed genes

Figure 6


A)
RACGAP1
Breast cancer; Unfavorable
DLGAP5
Pancreatic cancer; Unfavorable
SYNE2
Renal cancer; Favorable
FAM50B
Renal cancer; Favorable
Kaplan Meier Survival probability
Low (n=831)
High (n=244)
Low (n=126)
High (n=50)
Low (n=523)
High (n=354)
Low (n=276)
High (n=601)
Years
B)
Normal Tissue
Breast
Pancreas
Kidney
Kidney
C)
Cancer Tissue
Breast cancer
Pancreatic cancer
Renal cancer
Renal cancer
D)
Expression profiles
Pseudotime
E)
Interaction Partners
KIF20B
KIF2C
BIRC5
KIF20A
KIF23
RACGAP1
ECT2
KIF11
AURKB
CDCA8
KIF4A
CCNB2
NUSAP1
AURKA
CEP55
CDK1
DLGAP5
KIF11
TPX2
CCNB1
BUB1
ANLN
STAG2
TERF1
LMNB1
EMD
SYNE2
TERF2
SYNE3
STAG1
SUN2
SYNE1
SUN1
XPR1
ERAL1
SSSCA1
HDAC2
FAM50B
AGMAT
MRPS16
MRPL54