nature cell biology

Genome-scale RNAi profiling of cell division in human tissue culture cells

Ralf Kittler[1,6], Laurence Pelletier[1,7], Anne-Kristine Heninger[1], Mikolaj Slabicki[1], Mirko Theis[1], Lukasz Miroslaw[1], Ina Poser[1], Steffen Lawo[7], Hannes Grabner[1,2], Karol Kozak[1,2], Jan Wagner[1,2], Vineeth Surendranath[1], Constance Richter[1], Wayne Bowen[3], Aimee L. Jackson[4], Bianca Habermann[1,5], Anthony A. Hyman[1] and Frank Buchholz[1,8]

Cell division is fundamental for all organisms. Here we report a genome-scale RNA-mediated interference screen in HeLa cells designed to identify human genes that are important for cell division. We have used a library of endoribonuclease-prepared short interfering RNAs for gene silencing and have used DNA content analysis to identify genes that induced cell cycle arrest or altered ploidy on silencing. Validation and secondary assays were performed to generate a nine-parameter loss-of-function phenoprint for each of the genes. These phenotypic signatures allowed the assignment of genes to specific functional classes by combining hierarchical clustering, cross-species analysis and proteomic data mining. We highlight the richness of our dataset by ascribing novel functions to genes in mitosis and cytokinesis. In particular, we identify two evolutionarily conserved transcriptional regulatory networks that govern cytokinesis. Our work provides an experimental framework from which the systematic analysis of novel genes necessary for cell division in human cells can begin.

The use of RNA-mediated interference (RNAi) for gene silencing has provided a powerful tool for loss-of-function studies in a variety of metazoans[1]. In combination with high-throughput assays, genome-wide RNAi studies in invertebrates have uncovered novel gene functions in various biological processes[2]. Several of these studies were aimed at the identification of genes essential for cell division and cell cycle progression in *Caenorhabditis elegans* and cultured *Drosophila* cells[3–7]. Because defective regulation of cell division provokes human disease, particularly leukaemias and other cancers[8–10], a global survey of genes essential for cell division in human cells would not only advance the understanding of a fundamental biological process but may also deliver novel diagnostic and therapeutic targets for cancer.

We and others have developed endoribonuclease-prepared short interfering RNAs (esiRNAs) as potent and specific mediators of RNAi[11] in mammalian cells. We previously conducted a proof-of-concept screen validating the use of esiRNAs for the identification of genes required for cell division in HeLa cells[12]. The use of highly complex pools of siRNAs decreases off-target effects[13], which have been recognized as a major hurdle in RNAi screens[14–16].

Here we use an improved esiRNA library with genome-scale coverage[13]. We performed high-throughput DNA content analysis for the initial detection of cell cycle arrest and altered ploidy phenotypes, which were further analysed by fluorescence imaging, flow cytometry and time-lapse video microscopy. Using hierarchical clustering we grouped the identified genes on the basis of their multiparameter signatures into phenotypic classes, as a prelude for more detailed bioinformatic and functional analyses of their specific functions in mitosis and cytokinesis. Using this approach we uncovered a multitude of new genes implicated in these processes, including two evolutionarily conserved transcriptional regulatory networks that govern cytokinesis in mammalian cells.

RESULTS

Primary screen and validation

To identify cell division genes in HeLa cells, we developed a rapid and robust DNA content assay amenable to the high-throughput detection of phenotypes induced by RNAi (Fig. 1). We used an esiRNA library targeting 17,828 genes (Supplementary Information, Table S1) to transfect the cells and stained chromatin with propidium iodide 72 h after transfection. The DNA content histograms obtained were analysed and scored as a hit when a distance with a highly significant difference (more than 3 s.d.) from that of the negative controls was observed. Using these criteria we nominated 2,146 genes that altered cell cycle progression or ploidy on knockdown (Fig. 1, and Supplementary Information, Table S1).

[1]Max Planck Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, D-01307 Dresden, Germany. [2]Technology Development Studio, Pfotenhauerstrasse 108, D-01307 Dresden, Germany. [3]TTP LabTech Ltd, Melbourn Science Park, Melbourn, Hertfordshire SG8 6EE, UK. [4]Rosetta Inpharmatics LLC, 401 Terry Avenue N, Seattle, Washington 98109, USA. [5]Scionics Computer Innovation GmbH, Pfotenhauerstrasse 110, D-01307 Dresden, Germany. [6]Present address: Department of Human Genetics, University of Chicago, CLSB, 920 E. 58th Street, Chicago, Illinois 60637, USA. [7]Present address: Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [8]Correspondence should be addressed to F.B. (e-mail: buchholz@mpi-cbg.de)

Received 22 June 2007; accepted 9 october 2007; published online 11 November 2007; DOI: 10.1038/ncb1659

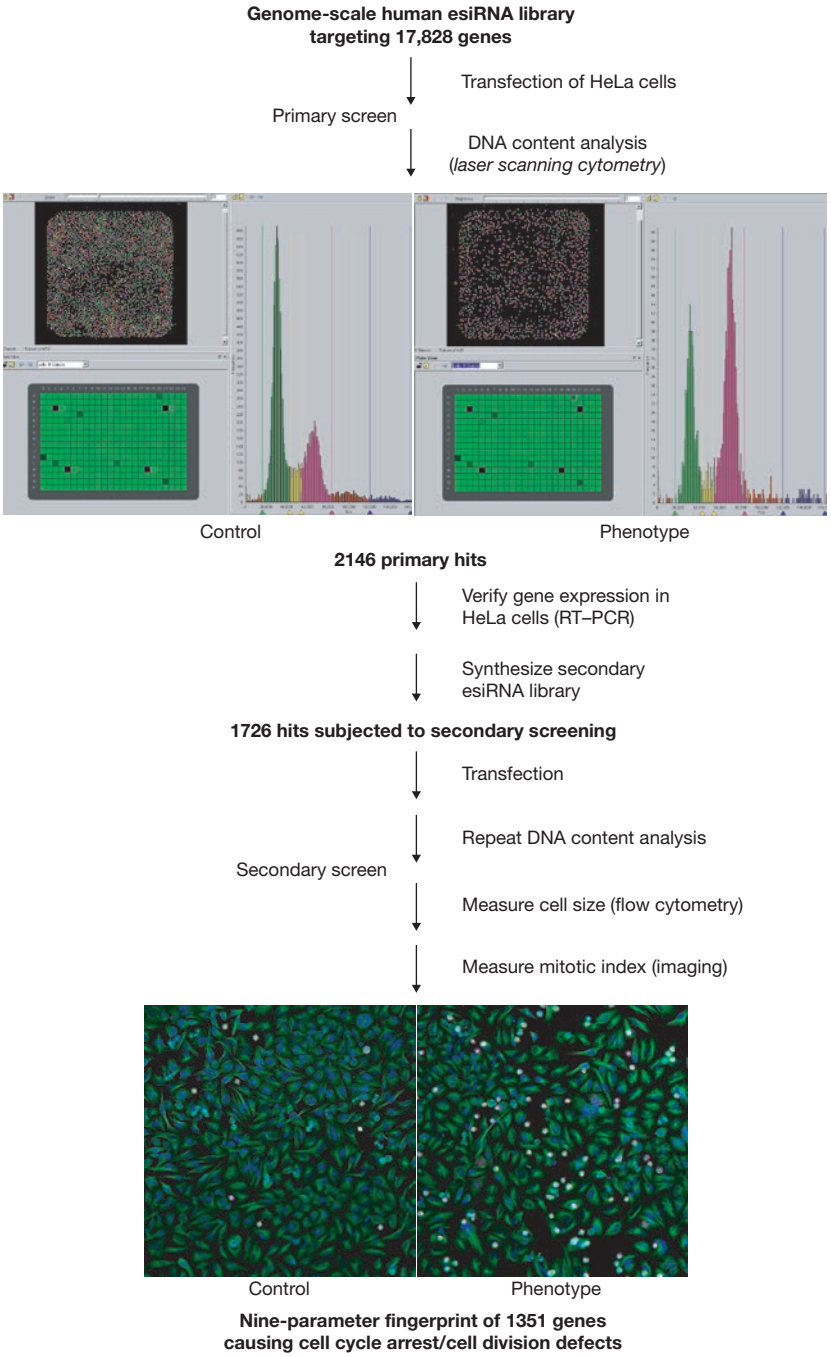

Figure 1 Screening strategy. Flow diagram of the genome-scale screening strategy for genes essential for cell cycle progression in human cells. Important steps are shown.

High-throughput screens typically produce false positives due to experimental errors. In addition, RNAi screens can contain false positives due to off-target effects[14–16]. To minimize the number of false positives in our screen we developed a three-step validation protocol. First, we analysed the expression status of the primary hits in HeLa cells; second, we resynthesized the esiRNAs and repeated the DNA content analysis for the genes that passed the first validation step. Third, we generated and assayed a second non-overlapping esiRNA for the genes that passed the second validation step. This cumulative validation procedure ensures that the data set we present is of high quality (for details on the hit validation see Methods and Supplementary Methods).

A second important aspect for assessing the quality of our RNAi data set is the evaluation of the rate of false negatives, which is an indicator for the saturation level of the screen. We used the analyses of subunit coverage for protein complexes/pathways to test screening efficiency (Supplementary Information, Table S2). This analysis indicated an overall high degree of efficiency (72%) for our screen, although the

Table 1 Overlap between RNAi screens

	Overlap (percentage)			
	***H. sapiens* (this study)**	***H. sapiens*[27]**	***D. melanogaster*[3]**	***C. elegans*[7]**
H. sapiens (this study)	–			
H. sapiens[27]	10	–		
D. melanogaster[3]	38	12	–	
C. elegans[7]	36	10	ND	–

For two human cell lines the overlap was calculated as the percentage from the putative human orthologues of genes that scored in the fly and in the worm RNAi screens. ND, not determined.

detection rate varied between 31% and 100% for individual complexes, which probably reflects differences in the stability of the corresponding complex subunits (for details on the evaluation of the rate of false negatives see Supplementary Methods).

Secondary screen and profile analysis of RNAi phenotypes

We extracted seven parameters from the DNA content analyses of the 1,351 hits that passed the second validation step: G1, S, G2/M, 8*N* (polyploid) DNA content, cell number, the percentage of dead cells and debris (subG1; that is, a DNA content of less than 2*N*) and the percentage of cells with 4*N*–8*N* DNA, which provided a measure for aneuploidy (Fig. 2a, b). We performed two additional assays to further characterize the cell cycle progression and cell division defects. For the first assay we determined the mitotic index for each hit gene to distinguish between cells in G2 and those in mitotic arrest. For the second assay we determined cell size by flow cytometry as an indicator of cell growth. Reproducibility for all parameters was documented by a linear regression analysis of all duplicate plate pairs, which yielded a high degree of reproducibility for the cell cycle parameters determined (Fig. 2c–h). Our screen therefore allowed the accurate detection and quantification of different types of cell cycle progression and division defects (Fig. 2i–p).

The secondary screen produced a nine-parameter fingerprint for each phenotype. We used these signatures to assign the corresponding genes to functional groups through hierarchical clustering, on the assumption that phenotypic profiles should reflect similar functions. On the basis of the gene clusters obtained (Fig. 3a) and manual inspection, we grouped the genes into four phenotypic classes: G0/G1 arrest, S arrest, G2 arrest, and cell division defects (Supplementary Information, Table S1). To annotate the identified genes we searched for functional data in two databases. We identified 217 genes previously associated with cell cycle progression, and assigned a novel function in cell cycle progression to 252 previously uncharacterized genes (Supplementary Information, Table S1). Furthermore, we identified 882 known genes that had previously been implicated in functions other than cell cycle progression and cell division. We grouped the hit genes into 12 functional categories (Supplementary Information, Table S1) that included processes relevant for cell cycle regulation such as transcriptional regulation, protein (de)ubiquitination and protein (de)phosphorylation.

We tested the reliability of phenotype grouping by searching our hit list for known cell cycle regulators and found many of them in the expected cluster (Fig. 3a). More than 50% (721) of all detected phenotypes (Fig. 3b) led to an increase in the number of cells with 2*N* DNA content, suggestive of either G1 arrests or exit from the cell cycle into G0. This phenotypic class contained known regulators for G1/S transition such as *CCNE1*, *MYC*, *E2F* and *TFDP1*. RNAi-mediated knockdown of 157 genes with a G0/G1 arrest caused a decrease in cell size indicative of impaired growth, whereas RNAi of 93 genes led to an increase in cell size, which may reflect a primary defect in G1/S transition. Similarly, we detected among the G2 arrest phenotypes the key regulator for G2/M transition, *CDC2* (*CDK1*). The S arrest cluster contained many known genes implicated in DNA replication (such as *CDC45L*, *PCNA* and those encoding replication protein A subunits and DNA polymerases) and dNTP synthesis (such as *RRM1* and *RRM2*). A delay in S phase was also observed for the knockdown of histones and genes that are implicated in histone expression (such as *SLBP* and *NPAT*) and chromatin assembly (such as *CHAF1A*). This finding may reflect the tight coordination of histone and DNA synthesis that is necessary for the proper replication of chromatin[17,18]. The cell division phenotypes contained a wealth of known regulators of mitosis and cytokinesis, for example *PLK1*, *KIF11*, *CDC20*, *ESPL1* (Separase), *SGOL1* (Shugoshin), *CDCA5* (Sororin), *CENPE*, *KNTC2* (Hec), *STK6* (Aurora A), *ch-TOG*, *TPX2*, *AURKB* (Aurora B), *INCENP*, *ANLN* (Anillin), *RACGAP1* (MgcRacGAP), *ECT2*, *PRC1* and *KIF23* (MKLP-1).

The congruence of genes with known cell cycle and cell division functions with the four phenotypic classes indicates that these clusters indeed reflect specific processes. We can therefore propose functions for uncharacterized genes by analysing their phenotypic profiles. For example, the silencing of genes implicated in DNA replication caused a delay in S phase. It is therefore plausible to assume similar functions for previously uncharacterized genes with the same phenotypic profile. For example, BLAST analysis of the predicted gene C17ORF41, which resulted in a pronounced S-phase delay, revealed weak protein sequence similarity ($E = 5 \times 10^{-14}$) to *RFC1* (encoding replication factor C1). RFC1 is essential for the loading of proliferating-cell nuclear antigen (PCNA) onto the 3′ ends of primer DNA to form a DNA sliding clamp that keeps DNA polymerase engaged at the replication fork[19]. Strikingly, *PCNA* was identified as the closest phenotypic profile neighbour of C17ORF41 (Fig. 3c). Although a weak sequence similarity of C17ORF41 to *RFC1* alone would be insufficient for functional annotation, in combination with RNAi profiling it enables us to assign a function for C17ORF41 in DNA replication with high confidence, potentially as a DNA clamp loader.

High-resolution analysis of cell division phenotypes

We selected the genes for which cluster analysis suggested a role in cell division for a more detailed functional characterization. Of these 289 genes, 223 caused an altered ploidy on knockdown. These genes may have special relevance for cancer biology, considering that genome instability and altered ploidy have been implicated in tumorigenesis[10,20].

In general, there are two major events that lead to altered ploidy. These are mitotic defects that impair spindle assembly and chromosome segregation, and cytokinesis defects preventing the proper division of the cytoplasm. We first analysed the phenotypic profiles of three well-known

mitotic and cytokinesis regulators (Fig. 4a–d) and used these signatures as core patterns to group the 223 genes into a cytokinesis defect cluster of 131 genes and a mitotic defect cluster of 92 genes (Fig. 4e–f, and Supplementary Information, Table S3). To evaluate the reliability of this grouping we analysed randomly 68 phenotypes by video microscopy of a HeLa cell line stably expressing histone–GFP (green fluorescent protein). Of the predicted mitotic defects, 91% showed defects in prometaphase or metaphase (in most cases accompanied by an arrest in these phases) from which some cells exited with unequal or no chromosome segregation. Of the predicted cytokinesis defects, 96% showed normal progression through mitosis until telophase, when the cells failed to divide the cytoplasm (Supplementary Information, Table S3). Thus, we can distinguish between gene functions in chromosome segregation and cytokinesis with high confidence by phenotypic profiling. For 184 genes, bioinformatic analyses revealed novel functions in cytokinesis or mitosis (Supplementary Information, Table S3).

Because defects in spindle assembly and cytokinesis have been linked to aneuploidy and cancer progression, we searched the literature for the 223 genes causing altered ploidy on knockdown and found that 51 (23%) of them had putative roles in tumorigenesis (Supplementary Information, Table S3). We analysed one of these genes, *CASC5*, in more detail because of its striking RNAi phenotype. *CASC5* is a fusion partner of the *MLL* oncogene in acute lymphoblastic leukaemia and is upregulated in certain types of lung cancer[21–23]. Knockdown of *CASC5* resulted in a severe chromosome congression defect leading to an unequal distribution of chromosomes to daughter cells (Fig. 5e–h, and Supplementary Information, Movie S1). The cells also progressed significantly faster through mitosis (Fig. 5i) and showed a marked decrease in the spindle checkpoint proteins Bub1 and BubR1 at the kinetochores of prometaphase cells (Fig. 5j–u). This phenotype is consistent with a defect in kinetochore function, which impairs both microtubule capture and the activation of the spindle assembly checkpoint. Interestingly, *CASC5* was recently identified in two proteomic studies as a component of the human kinetochore (as KIAA1570 and AF15q14) and was suggested to be the putative orthologue of the *KNL-1* gene in *C. elegans* and *Spc105* in *Saccharomyces cerevisiae*[24,25]. The *C. elegans* orthologue is required for the targeting of multiple components of the outer kinetochore, which is consistent with a 'kinetochore-null' phenotype observed on RNAi-mediated knockdown[26]. The loss-of-function phenotype we now report therefore shows that human *CASC5* is an essential component of the mammalian kinetochore. Because of the severe chromosome segregation defect and the inactivation of the spindle checkpoint on knockdown, we speculate that *CASC5* has a role in early tumorigenesis by causing aneuploidy.

Cross-species RNAi comparisons

Genome-wide data sets for cell cycle progression and/or cell division have previously been reported for *C. elegans* and cultured *Drosophila* cells[3,7]. We combined these results with data for human tissue culture cells presented here on HeLa cells and for U2OS cells[27] for a cross-species comparison of requirements for cell division. Although all four screens are not saturating (Supplementary Information, Table S2) and some of the genes of these data sets might have a different function or have become genetically redundant in different lineages, a cross-species comparison of cell cycle/cell division data should enable the identification of conserved regulatory pathways and functional modules.

We first analysed the global overlap between the screens and found that 38% of the fly genes and 36% of the worm genes with clear human orthologues were also identified in our screen (Table 1, and Supplementary Information, Table S4). These genes are therefore likely to have highly conserved gene functions in the cell cycle and cell division and to encode proteins with a high degree of sequence conservation. Second, we performed a more detailed cross-species analysis of cytokinesis defects, because for this phenotypic class the data source for cross-species comparisons is the largest available[3–5,28]. We compared these data sets with our screening data to identify conserved regulatory pathways for cytokinesis. This analysis revealed numerous motor proteins and central spindle components, confirming the conserved role of these proteins in cytokinesis (Supplementary Information, Table S4).

Identification of conserved transcriptional regulators for cytokinesis

Interestingly, the cross-species comparison of cytokinesis genes yielded genes encoding putative chromatin-binding proteins implicated in transcriptional repression. The knockdown of *NCOR2*, which is a co-repressor of various transcription factors[29] resulted in a cytokinesis defect (Supplementary Information, Movie S2). A similar phenotype was reported for the *D. melanogaster* homologue *Smr*[4] reflecting a high degree of functional conservation of *NCOR2* in cytokinesis. The analysis of *NCOR2* profile neighbours suggested functional interaction with *TBL1X* and *MLL5* (Fig. 3d). Bioinformatic analyses identified three putative *S. cerevisiae* homologues, *SNT1* (*NCOR2*), *SET3* (*MLL5*) and *SIF2* (*TBL1X*), encoding proteins that form the histone deacetylase-recruiting SET3 complex (SET3C) for the repression of meiosis-specific genes, including cytokinesis genes[30]. Strikingly, NCOR2 was recently found to interact with TBL1X and MLL5 in nuclear extracts of HeLa cells, substantiating the existence of a human SET3 complex (www.nursa.org/10.1621/datasets.01002). On the basis of these findings we propose a human SET3 complex composed of the NCOR2, MLL5 and TBL1X proteins (Fig. 3e), which might regulate either the expression of genes involved in cytokinesis or the chromatin structure of mitotic chromosomes.

We also identified a putative transcriptional regulatory complex containing the protein DKFZp686L1814 (LIN54). The silencing of the corresponding gene caused cytokinesis defects (Fig. 6a, b, and Supplementary Information, Movie S3) and a weak mitotic arrest (Supplementary Information, Table S1). We used the cross-species RNAi rescue approach[31] to confirm the role of *LIN54* in cytokinesis (Fig. 6e, f) and to show that GFP-tagged LIN54 protein localizes to the nucleus, which is consistent with its role in transcriptional regulation (Fig. 6c, d). A cross-species comparison revealed that the putative *D. melanogaster* orthologue *MIP120* is a component of the dREAM complex (*Drosophila* RBF, E2F2 and Myb interacting proteins containing complex)[32,33], whereas the *C. elegans Lin-54* is a component of the DRM complex (DP, Rb and MuvB containing complex)[34]. The dREAM and DRM complexes have been implicated in regulating Rb/E2F target genes[32–34], many of which are cell cycle-related genes, including genes implicated in cytokinesis, in both *Drosophila* and human cells[35,36]. Furthermore, we found that *Arabidopsis thaliana Tso1*, which has previously been shown to be essential for cytokinesis[37], shares similarity to human *LIN54*. A previous study has shown a role of some dREAM complex subunits in DNA replication in *Drosophila*[38]. To determine whether the cell division defects observed

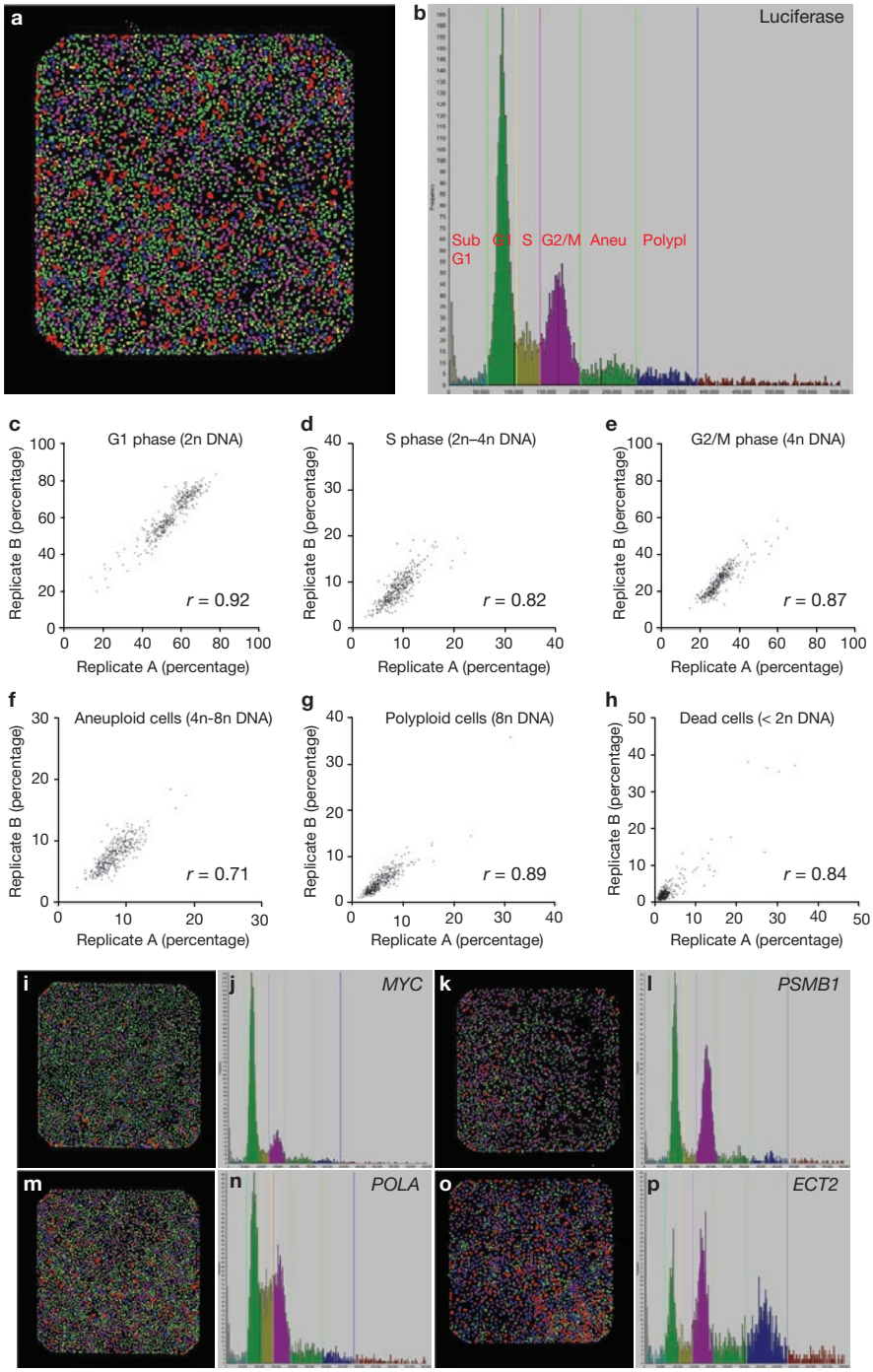

Figure 2 High-throughput detection and reproducibility of cell cycle parameters by DNA content analysis. (**a**) Propidium-iodide-stained nuclei of cells in a single well of a 384-well plate were analysed by laser scanning cytometry 72 h after transfection with control esiRNA targeting firefly luciferase. (**b**) Fluorescence intensities were plotted in a DNA-content histogram and used for gating of indicated cell cycle populations. (**c–h**) Scatter plots of a replicate 384-well plate pair for the percentages of DNA content of G1-phase (**c**), S-phase (**d**), G2/M-phase (**e**), aneuploid (**f**), polyploid (**g**) and dead (**h**) cells. The *r* values indicate the average linear correlation coefficient for replicate plate pairs. (**i–p**) The high reproducibility of DNA content analysis allowed the reliable detection of multiple cell cycle defects, for example G1 arrest (**i**, **j**), G2/M arrest (**k**, **l**), S arrest (**m**, **n**) and accumulation of polyploid cells (**o**, **p**) exemplified by the genes whose names are shown in the DNA content histograms. The reproducibility of the determined cell number, mitotic index and cell size was documented by linear regression analysis as for the other six parameters shown in **c–h**. We determined the following *r* values (indicated in parentheses) for cell number (0.90), mitotic index (0.78) and cell size (0.66). SubG1, debris/dead cells; aneu, aneuploid cells; polypl, polyploid cells.

on LIN54 knockdown were caused indirectly by defective replication, we monitored the dynamics of the key replication protein PCNA in LIN54-depleted cells and control cells (Supplementary Information, Movies S4 and S5). We did not observe S-phase progression defects on LIN54 depletion, suggesting that the observed cell division defect is not caused by defects in DNA replication. Interestingly, a recent study has

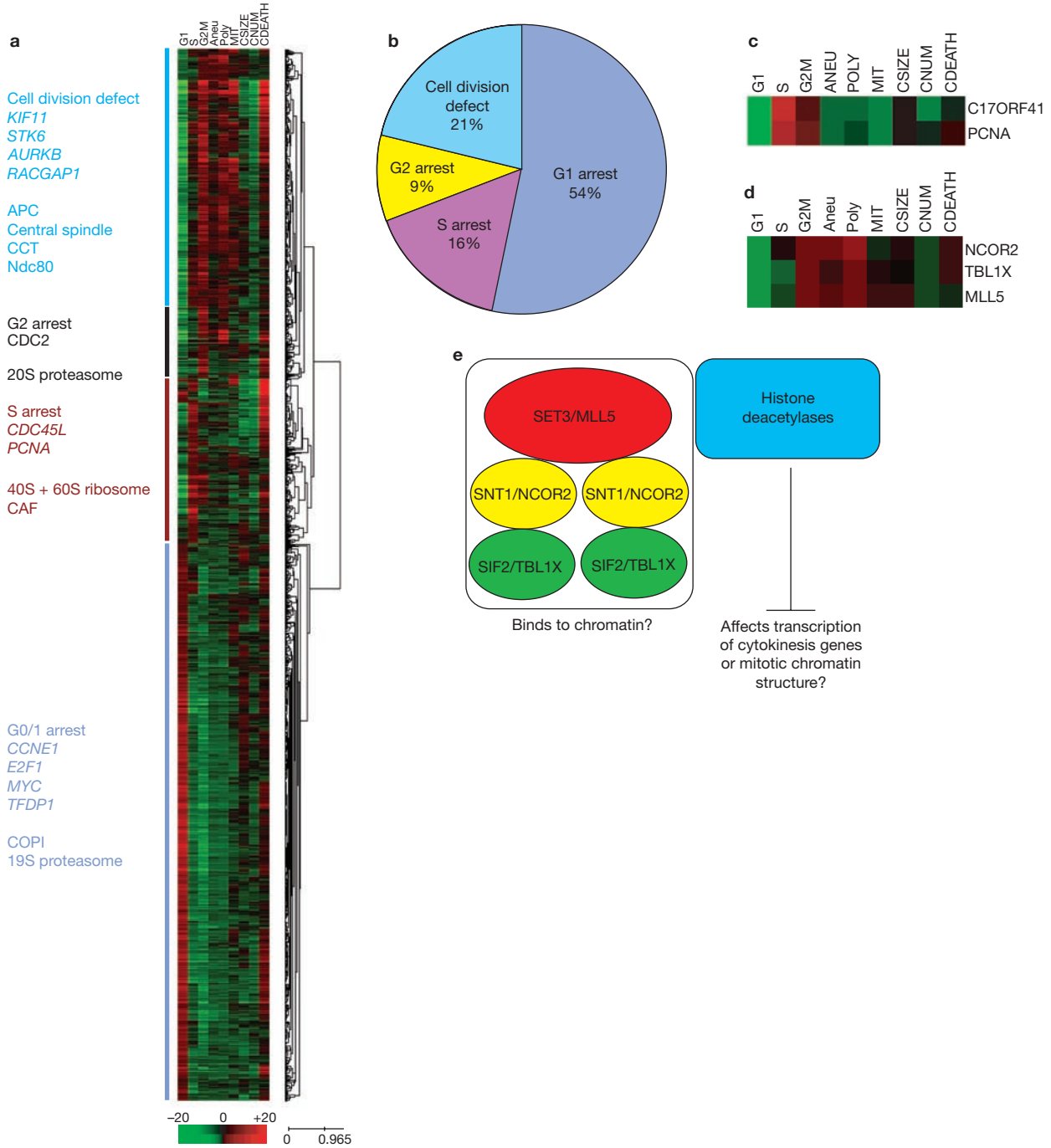

Figure 3 Phenotypic profiling of cell cycle defects. **a**, Hierarchical clustering of cell cycle defects based on their phenotypic signatures. The coloured lines to the left of the heat map indicate the major clusters. The predominantly represented cell cycle defect class is written in the same colour, with example genes of these clusters shown. Exemplified protein complexes for which subunits were found in specific clusters are shown. **b**, Classification of 1,351 cell cycle defects based on hierarchical clustering and subsequent manual inspection. **c**, *PCNA* has the most similar phenotypic profile to C17ORF41, potentially reflecting functional interaction. **d**, Profile analysis of NCOR2, TBL1X and MLL5. Profile comparison between the three genes reveals a similar pattern that may indicate functional interaction. **e**, Graphical depiction of the yeast SET3 complex (based on ref. 30) with their predicted human homologues NCOR2, TBL1X and MLL5 form a similar repressor complex like their putative yeast orthologues and a model how this complex regulates the expression of genes implicated in cytokinesis. G2M, G2/M phase; aneu, aneuploidy; poly, polyploidy; MIT, mitotic index; CSIZE, cell size; CNUM, cell number; CDEATH, dead cells; APC, anaphase-promoting complex; CCT, chaperonin-containing T-complex polypeptide 1; Ndc80, Nuf2–Ndc80 complex; CAF, chromatin assembly factor 1; COPI, coatomer I complex.

shown that the human dREAM complex binds to E2F target promoter regions and represses cell-cycle-dependent genes in quiescent T98G cells[39]. If dREAM is a general repressor of cell cycle genes, one might expect that the destruction of the complex would activate the expression of cell cycle genes, somewhat contradicting the observed RNAi phenotype. To test the role of *LIN54* in the transcriptional regulation

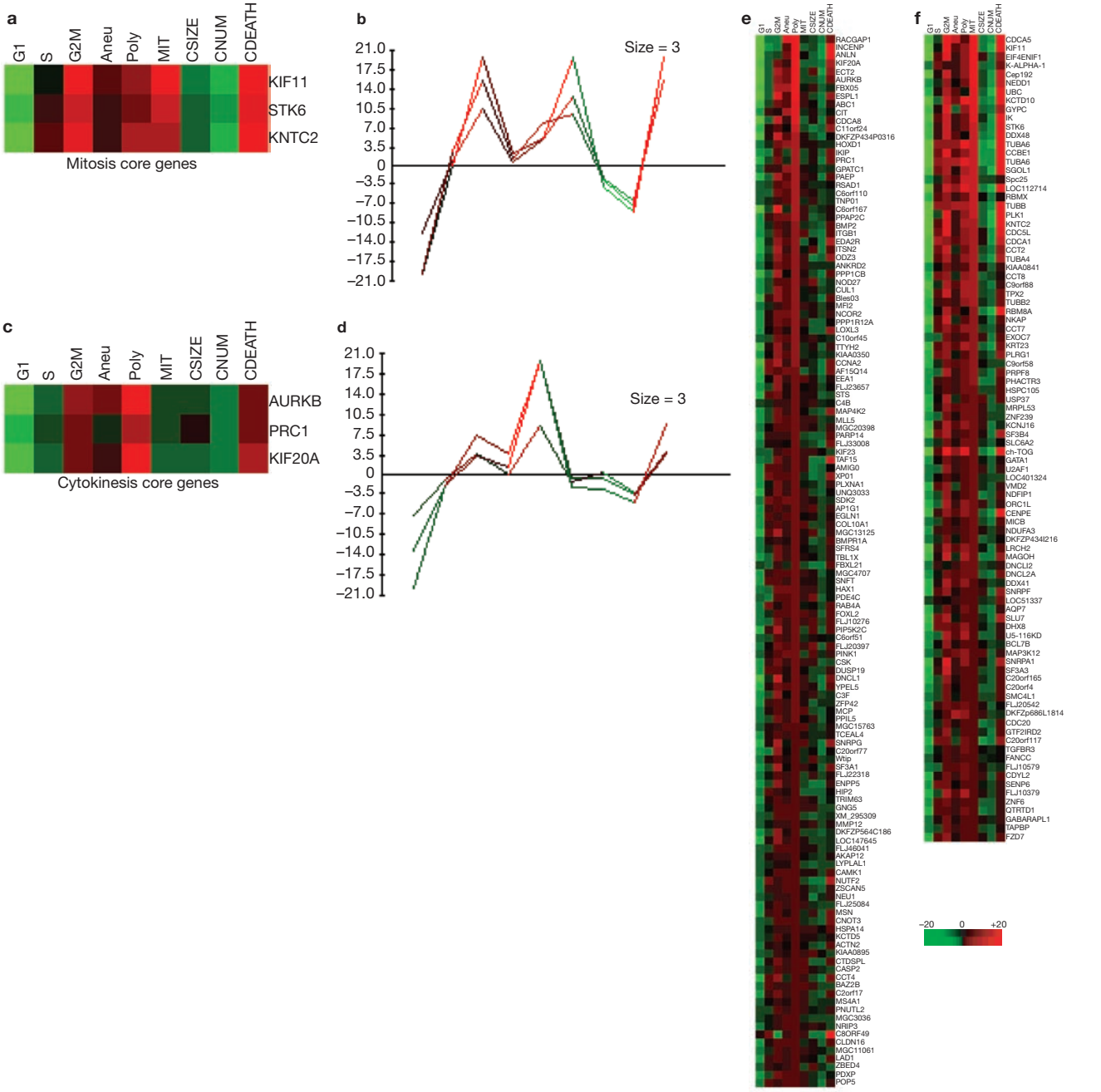

Figure 4 Profile analysis of cell division defects. (**a–d**) Profile comparison between the known mitotic genes *KIF11*, *STK6* and *KNTC2* (**a**, **b**) and the cytokinesis genes *AURKB*, *PRC1* and *KIF20A* (**c**, **d**). (**e**, **f**) On the basis of these core patterns the genes were grouped into a cytokinesis defect cluster (**e**) and a mitosis defect cluster (**f**). Abbreviations in the heat maps are the same as those in Fig. 3.

of cell cycle genes in proliferating cells, we performed gene expression array analyses after LIN54 depletion. Intriguingly, depletion of LIN54 resulted in the downregulation of many cell cycle genes (Fig. 6g, h, and Supplementary Information, Table S5). In particular, genes required for cytokinesis were rapidly downregulated, which is consistent with the observed phenotype. Strikingly, 82% of these genes for which chromatin immunprecipitation data are available[39] showed LIN54 binding in their promoter regions, suggesting that these genes are direct targets of LIN54 (Supplementary Information, Table S5). We tested esiRNAs for two other known subunits of the human dREAM complex (LIN37 and LIN52), which were not present in the screened esiRNA library. These two esiRNAs caused a highly significant increase in the 8*N* DNA content, as was observed for LIN54 (Supplementary Information, Table S5). Collectively, these findings indicate that LIN54 acts in a complex that is a transcriptional activator of cell division genes in proliferating cells, in contrast to quiescent cells, in which LIN54 is a transcriptional repressor.

DISCUSSION

Here we used a genome-scale RNAi screening strategy to identify genes required for cell division. Recent reports on the occurrence of false

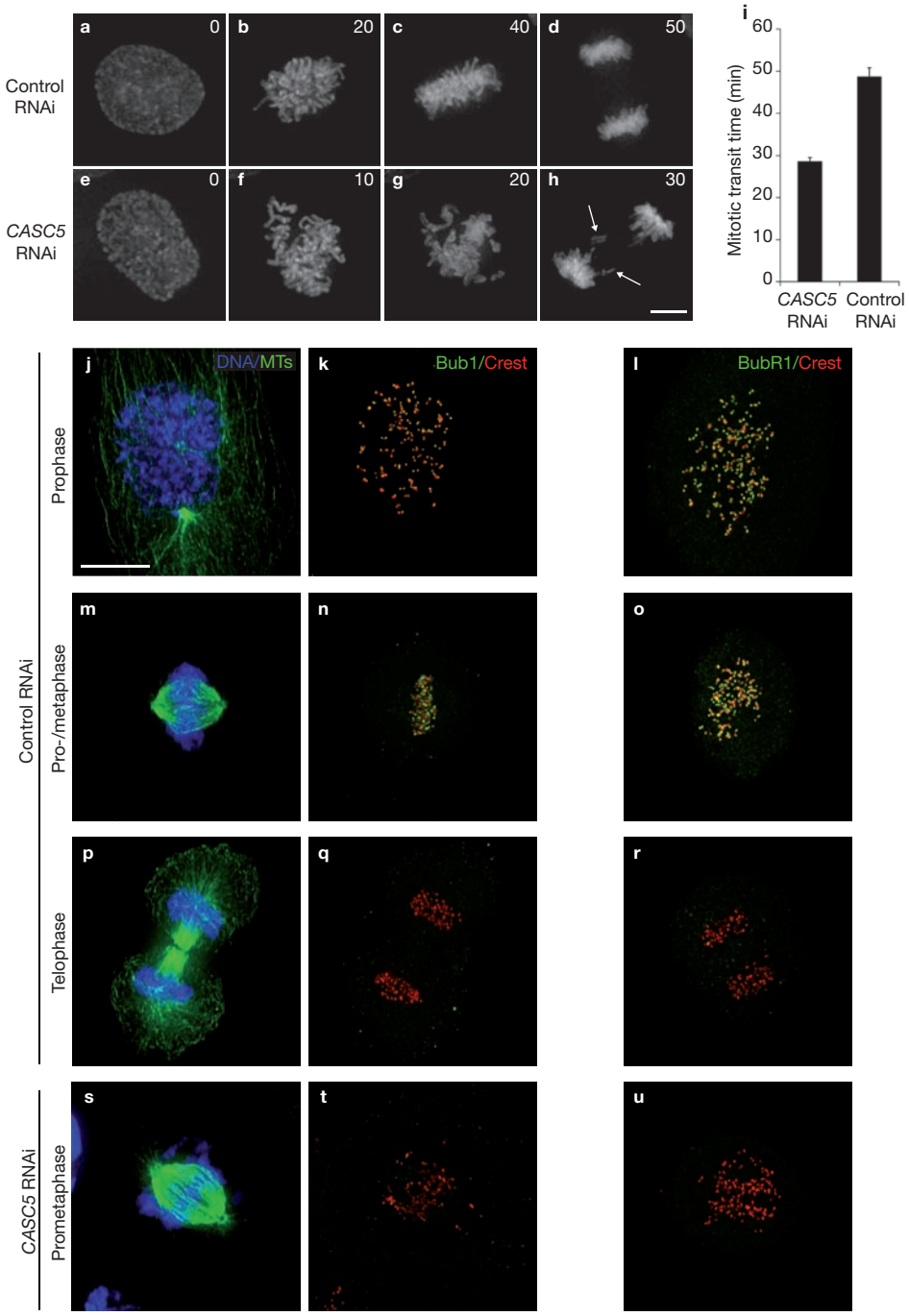

Figure 5 Mitotic phenotype of CASC5 depletion. (**a–h**) High-resolution video microscopy of chromosome dynamics during mitosis in cells transfected with esiRNA targeting firefly luciferase (**a–d**) or *CASC5* (**e–h**). The time-lapse sequence starts 60 h after transfection of a HeLa cell line stably expressing histone–GFP. Numbers indicate minutes after the start of the time-lapse sequence. The arrows in **h** indicate lagging chromosomes. (**i**) Quantification of mitotic transit times from prophase to anaphase for *CASC5* RNAi and control RNAi using esiRNA targeting firefly luciferase. In all, 21 mitoses were analysed for *CASC5* RNAi and 35 for control RNAi, starting 60 h after transfection. Bars indicate means and error bars indicate s.e.m. *P* was calculated as 5×10^{-9} with a two-tailed Mann–Whitney *U*-test. (**j–u**) Loss of Bub1 and BubR1 localization to centromeres on depletion of CASC5. HeLa cells were transfected with esiRNA targeting firefly luciferase (control) and *CASC5*, and were imaged by three-dimensional deconvolution microscopy 48 h after transfection for DNA (blue), microtubules (green), Bub1 (green) or BubR1 (green) and CREST (red). Merged images of DNA/microtubules and of Bub1/CREST and BubR1/CREST at different stages of mitosis in control (**j–r**) and in prometaphase *CASC5* esiRNAi treated cells (**s–u**) are shown. Note that on depletion of CASC5 the checkpoint proteins Bub1 and BubR1 no longer localize to centromeres in prometaphase cells. A total of 50 mitoses of CASC5-depleted cells and control cells were analysed. A loss of centromeric Bub1 localization was observed for 40 and 3 mitoses, and a loss of centromeric BubR1 localization was observed for 39 and 7 mitoses in CASC5-depleted and control cells, respectively. Scale bars, 5 μm (**a–h**); 10 μm (**j–u**).

positives in RNAi screens have raised concerns about the validity of large RNAi data sets[40]. When we compared the genes identified in a recently published siRNA-based screen for cell cycle regulators in human cells[27] with our study, we observed an overlap of only 10% between the two data sets (Table 1). Interestingly, the coverage of protein complex subunits/pathway components and the overlap between the genes identified in mammals and in the fly and worm RNAi screens are lower when using siRNA than when using esiRNA (Table 1, and Supplementary Information, Tables S2 and S4). In the siRNA-based screen, U2OS cells were used in assays similar to those in our study to identify genes required for cell cycle progression. The use of two different cell lines may account for some of the differences in the identified genes required for cell cycle progression or cell division, because many genes implicated in G1 progression and G1/S transition are probably cell-type-dependent (see also below). However, it is unlikely that this divergence alone accounts for all differences because we observed only a slight increase in the overlap rate (from 10% to 12.5%) when we restricted our comparison to genes that resulted in S-phase arrest or cell division defects and were verified by a second esiRNA (Supplementary Information, Table S4). In addition, the comparison with the *C. elegans* and *Drosophila* screens revealed a stronger overlap with our screen. Another possibility would be that the differences observed are caused by the use of different silencing reagents. For our screen we used a well-characterized library of esiRNAs and we carefully validated the genes identified. The characterization of the library revealed a high silencing efficiency and a 12-fold decrease in off-target effects at the mRNA level compared with chemically synthesized siRNA[13]. The high quality of the esiRNA resource translated into a high degree of screening efficiency in this study, as judged by a high saturation rate, and might have avoided many false positives due to off-target effects among the primary hits. The subsequent three-step hit validation eliminated many of the remaining false positives, which ensured an overall high quality of the reported data.

We identified hundreds of genes that on RNAi cause cell cycle arrest and/or cell division defects in HeLa cells. This cell line is the generic experimental system used in many laboratories to study the mammalian cell cycle. As are many other tissue culture cells, HeLa cells are transformed, and it is therefore likely that we have missed genes that are essential in untransformed cells or may have identified genes with a specialized function in transformed cells. In particular, steps controlling progression through G1 and entry into S phase are largely dependent on the cell type and context[9]. Many of the genes uncovered in this category may therefore be specific for HeLa cells, and a knockdown in a different cell line may produce a different phenotype. In addition, many of the reported genes causing a G0/G1 arrest are probably required for cell growth or stress response, with no direct function in cell cycle regulation or progression. In contrast to G1 progression and G1/S transition, mitosis and cytokinesis follow canonical steps that should vary little from cell line to cell line. Hence, many genes identified in this category may be newly identified participants that will also be important for cell division in most mammalian cells. Thus, the large number of newly described gene functions from this study will be a rich starting point for future work in research on the cell cycle and cancer.

We have demonstrated the potential of phenotypic RNAi profiling in assigning potential functions to novel genes and multiprotein complexes in combination with bioinformatics, cross-species comparisons and mining of proteomic data. This was highlighted by the identification of two putatively conserved transcriptional networks governing cytokinesis. For model organisms with genome-wide data sets of protein interactions such as *C. elegans*, the power of this approach for the identification of functional modules has recently been demonstrated[41]. We therefore expect that the comparative analysis of large-scale RNAi data sets with the growing number of interaction data for mammalian proteomes[42] will facilitate the identification of additional functional networks governing mammalian cell division. □

METHODS

Generation of a genome-scale esiRNA library. esiRNAs were synthesized as described previously[13,43]. Normalized esiRNAs were arrayed into 384-well plates for genome-scale screening.

Cell-based screening. esiRNAs (15 ng) were reversely transfected with Oligofectamine (Invitrogen, Karlsruhe, Germany) in 384-well tissue culture plates (Greiner, Frickenhausen, Germany). For DNA content analysis, the cells were fixed and stained with propidium iodide (Molecular Probes, Leiden, The Netherlands) and scanned with an Acumen Explorer microplate cytometer (TTP LabTech, Melbourn, UK). The resulting DNA-content histograms were manually gated with the Acumen Explorer software (TTP LabTech) to quantify the cell number and the percentages of cells with subG1, G1, S, G2/M phase, 4*N*–8*N* and 8*N* DNA content.

For determination of the mitotic index the cells were fixed and incubated with primary antibodies against tubulin (Serotec, Düsseldorf, Germany), phospho-histone H3 (Cell Signaling, Danvers, MA) and pericentrin (Abcam, Cambridge, UK) and subsequently with the secondary antibodies donkey anti-mouse Alexa 647, donkey anti-rabbit Alexa 555 and donkey anti-rat Alexa 488 secondary antibodies at 0.4 μg ml^{-1} (Molecular Probes) and 1 μg ml^{-1} 4′,6-diamidino-2-phenylindole (DAPI; Sigma, Taufkirchen, Germany). Images were acquired on the CellWoRx system (Applied Precision, Issaquah, WA) and the mitotic index was calculated automatically.

For measurement of cell size we used the forward scatter determined by fluorescence-activated cell sorting analysis of detached cells with a FACSCalibur (BD Biosciences, Heidelberg, Germany) using CellQuestPro software.

Hit selection and validation; clustering analyses. Each determined parameter was normalized for each well by using the mean and s.d. of eight negative controls (esiRNAs targeting firefly and *Renilla* luciferase) for an individual 384-well plate. We scored hits in the primary screen when the averages for the G1, S, G2M phase, or polyploid DNA content were greater than 3, indicating statistically highly significant hits.

To minimize the number of false positives in our screen we validated the primary hits, for which we considered two major sources of false positives: first, experimental errors, and second, off-target effects caused by the knockdown of unintended transcripts. To exclude these false positives from our primary hit list we performed three consecutive analyses. First, we analysed the expression status of the primary hits in HeLa cells. We reasoned that true positives must be expressed in HeLa cells whereas primary hits that are not expressed probably represent false positives. We therefore performed RT–PCR for the primary hits for complementary DNA prepared from the HeLa cells or *in silico* analysis of microarray data (see Supplementary Methods) and found that 1,748 out of 1,964 (89%) of the tested primary hits were expressed. Second, to eliminate experimental errors, we resynthesized esiRNAs for these genes and repeated the DNA content analysis. We reasoned that phenotypes resulting from experimental errors should not be reproducible in an independent experiment using the original esiRNA as silencing trigger. We were able to reproduce 78% of 1,726 tested initial phenotypes, yielding 1,351 candidate genes (Supplementary Information, Table S1). Third, we generated a second non-overlapping esiRNA for 1,254 of the primary hits to target the same transcript and tested their effect using the DNA content assay. For the remaining 97 genes the design of second esiRNAs was not possible. Using this approach we were able to validate 743 cell cycle arrest/cell division defect phenotypes (59%) in HeLa cells (Supplementary Information, Table S1). The cumulative validation procedure used here therefore ensures that the data set presented here is of high quality and contains a limited number of false positives.

Hierarchical clustering of the 1,351 phenotypic profiles was performed with EPCLUST (http://ep.ebi.ac.uk/EP/EPCLUST) using correlation-measure-based

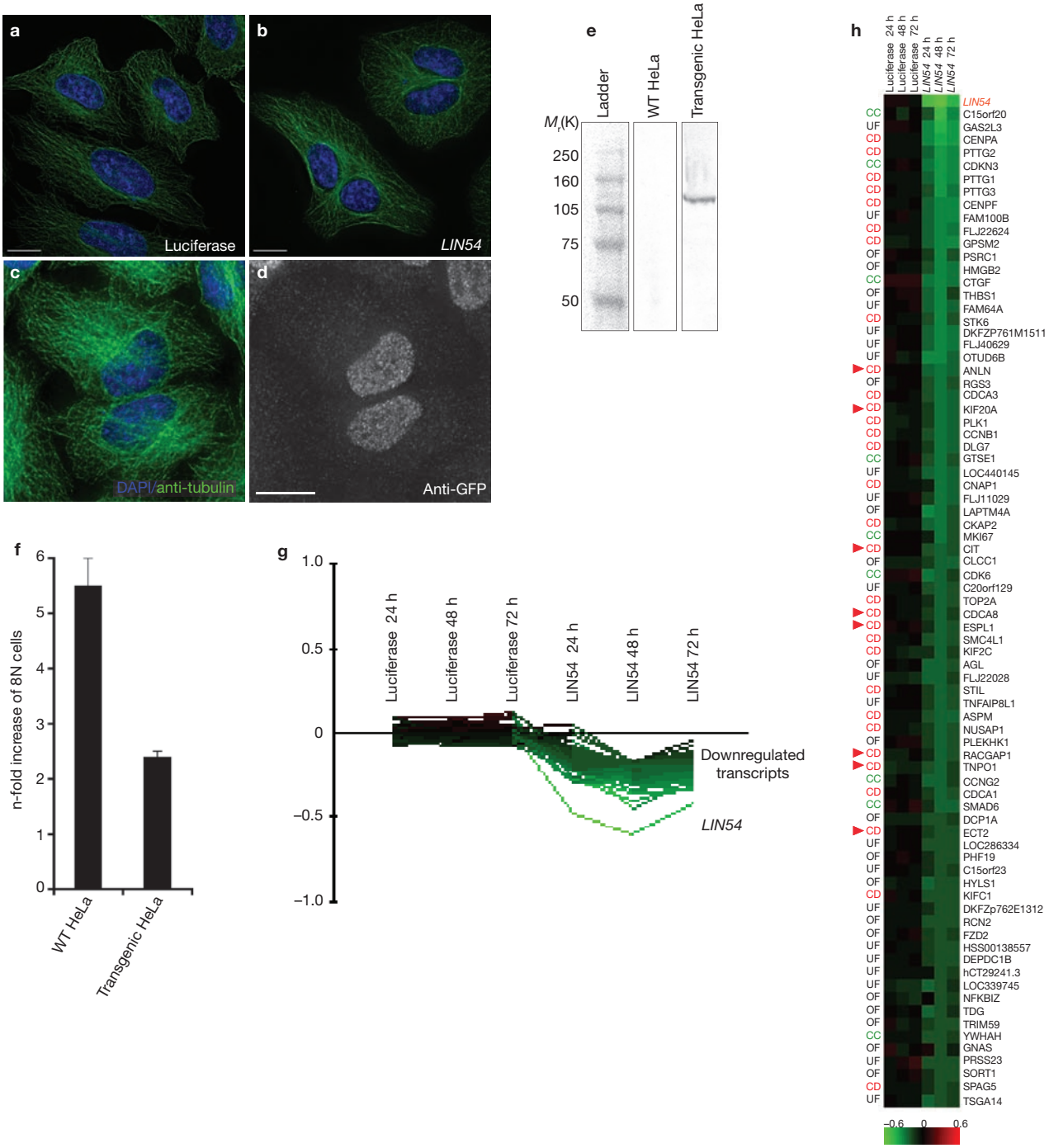

Figure 6 Characterization of LIN54 function in cell division. (**a**, **b**) Cytokinesis defect of LIN54 depletion. HeLa cells were transfected with esiRNA targeting firefly luciferase (**a**) as negative control and LIN54 (**b**) and were imaged by three-dimensional deconvolution microscopy 72 h after transfection for tubulin (green) and DNA (blue). Note the occurrence of binucleate cells on *LIN54* knockdown. (**c**–**e**) Characterization of LIN54–GFP. HeLa cells containing the mouse LIN54 BAC tagged with GFP stained for DNA (blue) and microtubules (green) (**c**), and GFP (**d**). Scale bars, 10 μm. (**e**) Western blot of wild-type (WT) and transgenic HeLa cell lysates with an anti-GFP antibody shows the presence of the tagged protein with the predicted relative molecular mass of 110,000. (**f**) Cross-species rescue of the cytokinesis defect. Wild-type HeLa cells or HeLa cells stably expressing the mouse LIN54–GFP were subjected to *LIN54* and luciferase RNAi. At 72 h after transfection, the increase in octaploid (8*N*) cells compared with the negative control was determined. Error bars represent s.d. *P* was calculated as 0.0008 ($n = 4$) with a two-tailed *t*-test based on a nearly normal distribution of the 8*N* DNA content for the negative controls in the screen (data not shown). (**g**, **h**) Downregulated transcripts on knockdown of *LIN54*. At 24, 48 and 72 h after transfection of *LIN54* and luciferase esiRNA, RNA was isolated and subjected to microarray analysis. A total of 78 genes with significantly decreased transcript levels ($P < 0.01$, at least 1.5-fold change after 48 h) are shown in a line diagram (**g**) and a heatmap (**h**). Genes causing cytokinesis defects when silenced in our screen are indicated with red arrowheads. Genes are annotated as follows: CC, known cell cycle function; OF, other function; UF, unknown function.

distances and average linkage clustering. As input for EPCLUST we used the averaged normalized values of the nine determined parameters that were cut off at +20 and −20.

Time-lapse microscopic assay. Transfections were performed as described above with a HeLa cell line stably expressing histone–GFP. Time-lapse fluorescence microscopy for 68 cell division defects was performed with a ScanR system (Olympus, Hamburg, Germany). For high-resolution time-lapse imaging, HeLa cells stably expressing histone–GFP were grown and transfected in eight-well LAB-TEK II chambered coverglasses (Nalge Nunc International, Wiesbaden-Biebrich, Germany). Cells were imaged 48 h after transfection on a DeltaVision RT system (Applied Precision) and a CoolSnap HQ charge-coupled device (CCD) camera (Roper Scientific, Tucson, AZ).

Immunofluorescence and microscopy for CASC5 and LIN54 phenotypes. Cells were transfected with esiRNAs targeting LIN54 and CASC5 (for esiRNA target sequences see Supplementary Information, Table S1). Cells grown on coverslips were fixed in ice-cold methanol at −20 °C for 5–8 min. After being blocked in 0.2% gelatin from cold-water fish (Sigma) in PBS (PBS/FSG) for 15–30 min, coverslips were incubated for 30 min with primary antibodies in PBS/FSG at the following concentrations: 1:1,000 rat anti-α-tubulin (Serotec); 1:1,000 mouse anti-hBub1 (Immuquest, Seamer, UK); 1:1,000 mouse anti-hBubR1 (Immuquest); 1:500 human anti-CREST (Europa Bioproducts Ltd, Cambridge, UK) or 1:1,000 sheep anti-GFP. After washes with 0.2% PBS/FSG, cells were incubated with 1:500 dilution of secondary antibodies for 30 min (donkey anti-mouse, rabbit, rat or sheep conjugated to Alexa 488, Alexa 594, Alexa 647; Molecular Probes). Coverslips were counterstained with 1 μg ml^{-1} DAPI to visualize chromatin. After being washed with 0.2% FSG/PBS, coverslips were mounted on glass slides by inverting them into mounting solution. Three-dimensional image datasets were acquired on an imaging system (DeltaVision; Applied Precision) equipped with a microscope (model IX71; Olympus), a CCD camera (CoolSNAP HQ2 1,024 × 1,024; Roper Scientific), and 60×, numerical aperture 1.42, plan-Apochromat objectives using 1 × 1 binning. *Z*-stacks (35 sections, 0.2 μm apart for each optical section) were collected and deconvolved computationally with the SoftwoRx software package (v3.6; Applied Precision).

Tagging of bacterial artificial chromosomes (BACs). The BACs RP23-96B7 (harbouring *mLIN54*) and RP23-428F16 (harbouring *mPCNA*), were obtained from the BACPAC Resources Center (http://bacpac.chori.org). A localisation and affinity purification (LAP) cassette[44] was inserted as a carboxy-terminal fusion using recombineering. Isolated BAC DNA was transfected and selected for stable integration as described[31].

Western blotting. Whole-cell lysates of 100,000 cells stably transfected with mLIN54-LAP were subjected to SDS–PAGE (NuPage 4–12% Bis-Tris; Invitrogen), blotted, and subsequently incubated with an anti-GFP antibody (1:5,000 dilution; Roche, Mannheim, Germany). After incubation with anti-mouse IgG conjugated with horseradish peroxidase (Bio-Rad, Munich, Germany), bands were detected with enhanced chemiluminescence (ECL) Western Blotting Detection Reagents (Amersham, Piscataway, NJ). As a ladder, Full-Range Rainbow Molecular Weight Marker (10–250 kDa; RPN800; Amersham) was used.

Microarray experiments. RNA isolated from esiRNA-transfected cells was hybridized against RNA from mock-transfected cells. Transfection, total RNA purification and processing, microarray hybridization and analyses were performed as described previously[13].

A more detailed description of the experimental procedures and data analysis is provided in Supplementary Methods.

Accession number. Gene Expression Omnibus (GEO): GSE9176.

Note: Supplementary Information is available on the Nature Cell Biology website.

ACKNOWLEDGEMENTS

We thank S. Payne for support with laser scanning cytometry; E. Krausz for support with automated transfection; D. Pinchev for help with the CASC5 experiments; E. Tanaka, F. Stewart and W. Zachariae for critical reading of an earlier version of the manuscript; P. Goodwin and C. Brown for development of the DeltaVision/CellWorX screening platform; and all members of the Buchholz laboratory and our colleagues in Mitocheck and SMP-RNAi for discussions. This study was funded by the Max Planck Society, the Bundesministerium für Bildung und Forschung grant SMP-RNAi under the framework of NGFN-2 (01GR0402) and the EU-FP6 grant FunGenES (LSHG-CT-2003-503494). During the course of this work R.K. and L.P. were supported by postdoctoral fellowships of the Human Frontier Science Program, and L.P. was supported by the Samuel Lunenfeld Research Institute.

AUTHOR CONTRIBUTIONS

R.K and L.P. contributed equally to this work. R.K., L.P. and F.B. performed project planning, experimental work, data analysis and wrote the manuscript. A.-K.H., M.S., M.T., L.M., I.P., S.L., H.G., K.K., J.W., V.S., C.R., W.B., A.L.J. and B.H. performed experimental work and data analysis. A.A.H. performed data analysis.

Published online at http://www.nature.com/naturecellbiology/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hannon, G. J. RNA interference. *Nature* **418,** 244–251 (2002).
2. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nature Rev. Genet.* **7,** 373–384 (2006).
3. Bjorklund, M. *et al.* Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature* **439,** 1009–1013 (2006).
4. Eggert, U. S. *et al.* Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.* **2**, e379 (2004).
5. Bettencourt-Dias, M. *et al.* Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432,** 980–987 (2004).
6. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408,** 331–336 (2000).
7. Sönnichsen, B. *et al.* Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* **434,** 462–469 (2005).
8. Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432,** 316–323 (2004).
9. Massagué, J. G1 cell-cycle control and cancer. *Nature* **432,** 298–306 (2004).
10. Rajagopalan, H. & Lengauer, C. Aneuploidy and cancer. *Nature* **432,** 338–341 (2004).
11. Buchholz, F., Kittler, R., Slabicki, M. & Theis, M. Enzymatically prepared RNAi libraries. *Nature Methods* **3,** 696–700 (2006).
12. Kittler, R. *et al.* An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432,** 1036–1040 (2004).
13. Kittler, R. *et al.* Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature Methods* **4,** 337–344 (2007).
14. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21,** 635–637 (2003).
15. Birmingham, A. *et al.* 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods* **3,** 199–204 (2006).
16. Jackson, A. L. *et al.* Widespread siRNA 'off-target' transcript silencing mediated by seed region sequence complementarity. *RNA* **12,** 1179–1187 (2006).
17. Nelson, D. M. *et al.* Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. *Mol. Cell. Biol.* **22,** 7459–7472 (2002).
18. Ye, X. *et al.* Defective S phase chromatin assembly causes DNA damage, activation of the S phase checkpoint, and S phase arrest. *Mol. Cell* **11,** 341–351 (2003).
19. Shiomi, Y. *et al.* ATP-dependent structural change of the eukaryotic clamp-loader protein, replication factor C. *Proc. Natl Acad. Sci. USA* **97,** 14127–14132 (2000).
20. Storchova, Z. & Pellman, D. From polyploidy to aneuploidy, genome instability and cancer. *Nature Rev. Mol. Cell Biol.* **5,** 45–54 (2004).
21. Takimoto, M. *et al.* Frequent expression of new cancer/testis gene D40/AF15q14 in lung cancers of smokers. *Br. J. Cancer* **86,** 1757–1762 (2002).
22. Chinwalla, V. *et al.* A t(11;15) fuses MLL to two different genes, AF15q14 and a novel gene MPFYVE on chromosome 15. *Oncogene* **22,** 1400–1410 (2003).
23. Kuefer, M. U. *et al.* Characterization of the MLL partner gene AF15q14 involved in t(11;15)(q23;q14). *Oncogene* **22,** 1418–1424 (2003).
24. Cheeseman, I. M. *et al.* A conserved protein network controls assembly of the outer kinetochore and its ability to sustain tension. *Genes Dev.* **18,** 2255–2268 (2004).
25. Obuse, C. *et al.* A conserved Mis12 centromere complex is linked to heterochromatic HP1 and outer kinetochore protein Zwint-1. *Nature Cell Biol.* **6,** 1135–1141 (2004).
26. Desai, A. *et al.* KNL-1 directs assembly of the microtubule-binding interface of the kinetochore in *C. elegans*. *Genes Dev.* **17,** 2421–2435 (2003).
27. Mukherji, M. *et al.* Genome-wide functional analysis of human cell-cycle regulators. *Proc. Natl Acad. Sci. USA* **103,** 14819–14824 (2006).
28. Echard, A., Hickson, G. R., Foley, E. & O'Farrell, P. H. Terminal cytokinesis events uncovered after an RNAi screen. *Curr. Biol.* **14,** 1685–1693 (2004).
29. Jepsen, K. & Rosenfeld, M. G. Biological roles and mechanistic actions of co-repressor complexes. *J. Cell Sci.* **115,** 689–698 (2002).
30. Pijnappel, W. W. *et al.* The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev.* **15,** 2991–3004 (2001).
31. Kittler, R. *et al.* RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proc. Natl Acad. Sci. USA* **102,** 2396–2401 (2005).
32. Lewis, P. W. *et al.* Identification of a *Drosophila* Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev.* **18,** 2929–2940 (2004).

33. Korenjak, M. *et al.* Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell* **119,** 181–193 (2004).
34. Harrison, M. M., Ceol, C. J., Lu, X. & Horvitz, H. R. Some *C. elegans* class B synthetic multivulva proteins encode a conserved LIN-35 Rb-containing complex distinct from a NuRD-like complex. *Proc. Natl Acad. Sci. USA* **103,** 16782–16787 (2006).
35. Dimova, D. K., Stevaux, O., Frolov, M. V. & Dyson, N. J. Cell cycle-dependent and cell cycle-independent control of transcription by the *Drosophila* E2F/RB pathway. *Genes Dev.* **17,** 2308–2320 (2003).
36. Ren, B. *et al.* E2F integrates cell cycle progression with DNA repair, replication, and G_2/M checkpoints. *Genes Dev.* **16,** 245–256 (2002).
37. Hauser, B. A., He, J. Q., Park, S. O. & Gasser, C. S. TSO1 is a novel protein that modulates cytokinesis and cell expansion in *Arabidopsis*. *Development* **127,** 2219–2226 (2000).
38. Beall, E. L. *et al.* Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* **420,** 833–837 (2002).
39. Litovchick, L. *et al.* Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. *Mol. Cell* **26,** 539–551 (2007).
40. Echeverri, C. J. *et al.* Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nature Methods* **3,** 777–779 (2006).
41. Gunsalus, K. C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436,** 861–865 (2005).
42. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437,** 1173–1178 (2005).
43. Kittler, R., Heninger, A. K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nature Methods* **2,** 779–784 (2005).
44. Cheeseman, I. M. & Desai, A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci STKE* **2005,** pl1 (2005).

nature cell biology

SUPPLEMENTARY METHODS
Detailed description of the experimental procedures and data analysis.

TABLES
Table S1: All esiRNA sequences and screening data, annotation and validation of hit genes

Table S2: Analysis of complex subunit/pathway component coverage for the estimation of the detection rate

Table S3: Analysis of genes altering ploidy upon knockdown

Table S4: Cross-species comparison between cell cycle/division RNAi screens

Table S5: Analysis of LIN54 function and LIN37/LIN52 RNAi phenotypes

MOVIES
Movie S1: Phenotype for CASC5 knockdown, luc (luciferase, negative control)

Movie S2: Phenotype for NCOR2 knockdown, luc (luciferase, negative control)

Movie S3: Phenotype for LIN54 (DKFZP686L1814) knockdown, luc (luciferase, negative control)

Movie S4: PCNA dynamics in LIN54-depleted cells

Movie S5: PCNA dynamics in luciferase-depleted cells (negative control for Movie S4)

Supplementary Methods

Cell-based screening. 15 ng esiRNAs in 5 µl TE buffer were mixed with 5 µl OptiMEM (Invitrogen) containing 0.2 µl Oligofectamine (Invitrogen) in a 384 well tissue culture plate (Greiner) using a FREEDOM EVO open platform (SN 446, TECAN) with a 8 needle liquid handling system (LiHa), 384 teflon coated steel head (TeMo) with active 384 well wash station and a robot manipulator arm (RoMA). After 20 minutes of incubation at room temperature 1000 HeLa cells in 40 µl medium (DMEM, 12.5% FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin; Invitrogen) were seeded onto the transfection mix using a WellMate dispensing system (Matrix). All assays were performed 72 hours after transfection.

For DNA content analysis cells were fixed in ice-cold 100% ethanol for 2 hours, rehydrated for 15 minutes in PBS, and stained for 30 minutes in PBS containing 10 µg/ml propidium iodide (Molecular Probes) and 100 µg/ml RNase A (Qiagen). The stained cells were scanned with an Acumen Explorer (TTP LabTech). The resulting DNA content histograms were manually gated with the Acumen Explorer software (TTP LabTech) to quantify the percentages of cells with subG1, G1, S, G2/M phase, 4N-8N and 8N DNA content.

For determining the mitotic index, cells were fixed by adding 50 µl of 8% PFA (Sigma), and incubated for 20 minutes at room temperature. Cells were permeabilised and blocked in PBS containing 0.1% Triton X-100 (Sigma) and 0.2% gelatine (Sigma). The cells were incubated for 60 minutes at room temperature in the presence of primary antibodies against tubulin (Serotec), phospho-histone H3 (Cell Signaling) and pericentrin

(Abcam) at 1 µg/ml, washed 3 times in PBS containing 0.2% fish skin gelatine, incubated for 60 minutes with donkey anti-mouse Alexa647, donkey anti-rabbit Alexa555 and donkey anti-rat Alexa488 secondary antibodies at 0.4 µg/ml (Molecular Probes), and 1 µg/ml DAPI (Sigma). After two washes with PBS the cells were stored in 100 µl PBS. Images were acquired on the CellwoRx system (Applied Precision) equipped with a 10x objective. 6 images were acquired per well containing typically a total of more than 2000 cells. For the quantification of the mitotic index, we used the software program esiImage. The mitotic index was calculated by dividing the number of phosphorylated histone H3 positive cells (mitotic cells) by the number of DAPI positive nuclei (total cell number).

For measurement of cell size, cells were detached by adding PBS containing 5 mM EDTA (Sigma), incubated for 15 minutes at 37°C, and fixed with 1% PFA. Forward scatter (FSC) was determined by FACS analysis with a FACSCalibur (BD Biosciences) using CellQuestPro software as a measure of cell size.

DNA content analyses and hit selection. Because the low number of cells analysed in an individual well of a 384-well plate did not allow automated peak modeling, e.g. by ModFit, a manual procedure was utilized to gate the individual cell cycle phases. In the primary screen, gates were defined for each 384-well plate by eye-observation of the overlayed DNA content histograms of the eight negatives control wells, and the same gates were used for the analysis of all other wells of the same plate. Although the peak positions of many wells on a plate were slightly shifted relative to the peaks of the control samples, this gating procedure allowed a highly reproducible quantification of G1 and G2 DNA content, and a moderately reproducible quantification of S phase and 8N DNA

content. However, the subG1 and 4N-8N DNA content could not be accurately determined with this procedure. In the secondary screen, the accuracy of gating was improved by gating the histograms based on the individual peak positions of each well. This laborious procedure markedly increased the reproducibility of S and 8n DNA content and enabled a reliable quantitification of subG1 and 4n-8n DNA content.

Each determined DNA content parameter as well as the mitotic index and the cell size were normalized for each well using mean and standard deviation of eight negative controls (esiRNAs targeting firefly and renilla luciferase) for an individual 384-well plate. These normalized values represent the distance between an individual score and the mean of the negative controls in units of the standard deviation of the negative controls. We averaged these normalized values from the two replicates, and identified hits in the primary screen when the averages for the G1, S, G2M phase were at least greater than 3 indicating statistically highly significant hits. We also scored genes as hits when the decrease in the cell number was highly significant, i.e. averages at least smaller than – 3. Because of the peak shifting effect described above we used a cut-off of +10 for the averages of the polyploid DNA content. In the secondary screen we scored a primary hit as verified when the averaged normalized value of the affected cell cycle phase in the primary screen was at least +2 (indicating statistical significance) in the secondary screen. In addition, we scored primary hits with an averaged normalized values +1.5 >> +2.0 (approaching statistical significance) in the secondary screen as verified low confidence hits that are highlighted in yellow in Table S1. In several cases primary hits that showed an increase in one phase, e.g. S phase, were not reproducible in the secondary screen but exhibited a significant increase in another phase, e.g. G2. Many of

these cases were due to peak shifting effects in the primary screen, i.e. the inspection of the primary DNA content histogram revealed the same phenotype that was observed observed in the secondary. For all these cases, the histograms of the primary and secondary screen were manually inspected, and scored as verified hits only when the same phenotype was apparent.

Analysis of the expression status of candidate genes in HeLa. For 1466 of the hit genes identified in the primary screen we tested the presence of a transcript in HeLa experimentally by RT-PCR. RNA from HeLa cells was extracted using the RNeasy Mini Kit (Qiagen) including a DNaseI digest. cDNA was synthesized with SuperScript III reverse transcriptase (Invitrogen) using an oligo(dT) primer (Invitrogen). PCR was performed with transcript-specific primers. PCR products were separated on a 2% agarose gel. We detected a single RT-PCR product of the predicted size for 1357 hit genes (93%). The expression status of 498 hit genes, for which no transcript-specific primers were available was analyzed *in silico*. For this purpose we used expression data of the NCI60 cell lines (including HeLa) from the SYMATLAS database (http://wombat.gnf.org/index.html)[1] and expression data generated by Rosetta Inpharmatics. Using the A/P calls as an indicator for the presence of a transcript we identified 391 genes expressed in HeLa (79 %). We note that many of the 107 genes flagged absent in HeLa had a relatively low computed RMA expression value in all other NCI60 cell lines and also in other tissues we tested *in silico*. Therefore, the *in silico* determined number of expressed hit genes might be artificially low in comparison to the experimental results because A/P calling might have excluded low-abundant transcripts.

The expression status of 84 genes could not be tested, because neither transcript-specific primers nor gene expression data were available. The expression status of the genes is indicated in Table S1.

Determination of false-positive and false-negative rate. Inspection of the 41% of primary hits not validated with a second esiRNA suggests that they are not necessarily due to off-target effects but rather to less efficient silencing efficacy. In fact, many genes with demonstrated roles in cell division (e.g. *Shugoshin*, *CENP-E*)[2,3] did not reproduce the phenotype obtained in the primary screen when using a second esiRNAs (see also Table S1). In such cases, it is therefore likely that insufficient silencing when using the second esiRNA is responsible for the lack of phenotype. Because the esiRNAs used in the primary screen were selected to represent the most efficient esiRNA for gene silencing[4], many of the second esiRNAs were predicted to be less potent silencers. It is therefore conceivable that a significant proportion of the genes we could not confirm reflect insufficient knockdown rather than off-target gene silencing. In order to estimate the percentage of phenotypes we missed with the second esiRNAs, we analyzed the detection rate of subunits of multiprotein complexes/pathway components (see also next paragraph). This analysis indicates a false-negative rate of 19% (Table S2), which is likely due to insufficient knockdown, and suggest that the actual false-positive rate of the genes passing the first two filters is lower than 41%. Thus, assuming that the false-positive rate of our screen is relatively low after the first two validation steps and that the third validation step would remove a substantial number of true-positives we decided to further analyze the 1351 primary hits that remained after the second validation step. However, we have marked all hits verified with an independent esiRNA as validated hits, while the genes that could not be verified or tested with a second esiRNA have been marked as not validated hits (see Table S1). We also estimated the false-positive rate due

to experimental errors by randomly sequencing of 305 PCR products used for the esiRNA generation. 99% of these templates displayed the expected sequence and hence about 1% of false-positives in our screen may originate from this artefact.

A second important aspect for assessing the quality of our RNAi data set is the evaluation of the false-negative rate, which is an indicator for the saturation level of the screen. Ideally, this rate would be inferred from the coverage of all genes known to be involved in cell division or cell cycle progression. However, the calculation of a saturation score is complicated, because there is no ultimate list of essential cell cycle and division genes. Although a set of clearly essential genes could be defined when focusing only on some key regulators, e.g *CDK1* and *PLK1*, many known cell cycle-associated genes would have to be excluded because of genetic redundancy, e.g. D-type and E-type cyclins and Cdk2/4/6 were shown to be dispensable in mice[5]. Thus, a control set for measuring the rate of false-negatives would be biased to some extent, and the calculated efficiency could be a matter of debate. An alternative, unbiased approach for estimating the saturation of RNAi screens has been recently used for the analysis of genome-wide fly and worm studies that also aimed on the identification of cell cycle and/or cell division genes. These studies demonstrated that depletion of individual subunits of protein complexes resulted in identical RNAi phenotypes for most if not all subunits of a complex[6,7]. The analyses of ten known human protein complexes in our data set demonstrated similar phenotypes for 97% of the identified subunits of the same complex (Table S2). When analysing the subunit coverage for the ten complexes, we detected on average 72% of all subunits of these complexes present in the screened library indicating a high degree of efficiency for our screen. The high correlation of phenotypes between

complex subunits also highlights the potential for the identification of functional networks in cell cycle regulation, when RNAi data are compared with protein interaction data.

Time-lapse microscopic assay. Transfections were performed as described above using a HeLa cell line stably expressing Histone-GFP. Time-lapse fluorescence microscopy for 68 cell division defects was performed by aquiring frames every 6 minutes over a 48 hours period starting 24 hours after transfection using a Cell^R system (Olympus) equipped with a 10X objective and an incubator chamber (Evotec). For high-resolution time-lapse imaging, HeLa cells stably expressing Histone-GFP were grown and transfected in 8-well LAB-TEK II chambered coverglass (Nalge Nunc International). Cells were imaged 48 hours after transfection on a DeltaVision RT (Applied Precision) system heated at 37°C and equipped with an Olympus 60x objective and a CoolSnap HQ CCD camera (Roper Scientific). Z-stacks (30 sections, 1 μm apart, 35ms exposure per section, 10% ND filter) were acquired every 10 minutes over 24 hours. Images were computationally deconvolved using the SoftWorX software package version 3.4.4 (Applied Precision) and shown as maximal projections.

To analyse a possible effect of LIN54 depletion on S phase, we measured the length of PCNA dynamics in PCNA-BAC transgenic HeLa cells. For this purpose, we tagged the mouse PCNA gene (present on BAC PR23-428F16) with GFP (as part of the LAP tag) on the C-terminus and transfected cells stably expression the PCNA-LAP fusion protein with LIN54 or control (Luc) esiRNAs. 24 hours after transfection cells were imaged on a DeltaVision RT (Applied Precision) as above for 48 hours and images

were acquired every 15 minutes. 12 cells were analysed each for the time where PCNA foci became visible to the end of cell division. Of the 12 cells imaged, none and seven showed a cytokinesis defect for the control and LIN54 transfected cells, respectively.

Bioinformatic analyses. For functional annotation we performed a batch download of gene ontology data from the public database SOURCE (http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch)[8]. In addition, we searched the Proteome BioKnowledge® Library (http://www.proteome.com/control/tools/proteome) for genes that had no gene ontology entry in SOURCE. Based on their annotations we grouped the genes manually into the following categories: metabolism (anabolic and catabolic pathways, including glycosylation), transcription (general transcription, i.e. RNA polymerase subunits), transcriptional regulation (positive and negative regulation of transcription), translation (ribosomal proteins, biogenesis of ribosomes, translation factors), RNA metabolism (mRNA splicing, RNA modication and degradation), transport (channels, membrane trafficking, nucleo-cytosolic transport), (de)phosphorylation (kinases, phosphatases and their accessory proteins), (de)ubiquitination (proteins implicated in ubiquitin metabolism), protein folding (e.g. chaperons, chaperonins), cytoskeleton (components and modulators of the actin/microtubule cytoskeleton, including motor proteins), other (all genes that did not match any of the categories above) and unknown (genes without annotations). These annotations are indicated in Table S1.

For the analysis of the overlap between the genome-wide human and the fly and worm RNAi screens[6,7], we identified the likely human orthologues of *C. elegans* and *D. melanogaster* genes using both BLAST similarity and the InParanoid database

(http://inparanoid.cgb.ki.se). We then calculated the ratio between the human orthologues that also scored in our screen and the total number of the human orthologues present in our library. (Table S4).

For the comparison of our data with previous cytokinesis RNAi screens in *D. melanogaster*[7,9-11] (Table S4) we first identified the putative homologues of the reported *Drosophila* genes among our 1351 hit genes. For this purpose we either used the published information of the human orthologues in the corresponding screening papers or the predicted human orthologues. We identified putative orthologues with similar phenotypes manually by comparing our phenotypic signatures with the published fly RNAi data.

For the identification of the homologues of C17ORF41, *LIN54*, *MLL5*, *NCOR2* and *TBL1X* we used the BLAST similarity of the predicted amino acid sequences that were refined with the Smith-Waterman algorithm as implemented in the Proteome BioKnowledge® Library. For *TBL1X* we identified *S. cerevisae SIF2* (E=1e-40), for *MLL5* we identified *S. cerevisiae SET3* (E=9e-11), and for *NCOR2* we identified *S. cerevisiae SIF2* (E=3e-13) as the most likely homologues. All three homologues were verified by reciprocal BLAST. Likewise, we identified for human *LIN54* the putative *D. melanogaster* homologue *MIP120* (E=9e-49) and the *A. thaliana* homologue *Tso1* (E=3e-20). The GenBank accession numbers of the amino acid sequences are as follows: human *MLL5* (NP_891847), *S. cerevisiae* (CAA82101), human *NCOR2* (NP_006303), *S. cerevisae SNT1* (CAC42983), human *TBL1X* (NP_005638), *S. cerevisiae SIF2* (CAA85058), human DKFZp686L1814 (NP_919258), *D. melanogaster MIP120* (NP_610879), *A. thaliana Tso1* (NP_566718), human C17orf41 (NP_079133) and human

RFC1 (NP_002904).

Previous implications of genes exhibiting a mitotic/cytokinesis defect upon knockdown in cancer were identified by literature search in PubMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed) or by data mining in the Proteome BioKnowledge® Library and are indicated with a reference in Table S3.

Supplementary References

1. Su, A.I. et al. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* 99, 4465-70 (2002).
2. Tanudji, M. et al. Gene silencing of CENP-E by small interfering RNA in HeLa cells leads to missegregation of chromosomes after a mitotic delay. *Mol Biol Cell* 15, 3771-81 (2004).
3. Salic, A., Waters, J.C. & Mitchison, T.J. Vertebrate shugoshin links sister centromere cohesion and kinetochore microtubule stability in mitosis. *Cell* 118, 567-78 (2004).
4. Kittler, R. et al. Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods* 4, 337-44 (2007).
5. Sherr, C.J. & Roberts, J.M. Living with or without cyclins and cyclin-dependent kinases. *Genes Dev* 18, 2699-711 (2004).
6. Sonnichsen, B. et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* 434, 462-9 (2005).
7. Bjorklund, M. et al. Identification of pathways regulating cell size and cell-cycle progression by RNAi. *Nature* 439, 1009-13 (2006).
8. Diehn, M. et al. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. *Nucleic Acids Res* 31, 219-23 (2003).
9. Bettencourt-Dias, M. et al. Genome-wide survey of protein kinases required for cell cycle progression. *Nature* 432, 980-7 (2004).
10. Echard, A., Hickson, G.R., Foley, E. & O'Farrell, P.H. Terminal cytokinesis events uncovered after an RNAi screen. *Curr Biol* 14, 1685-93 (2004).
11. Eggert, U.S. et al. Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol* 2, e379 (2004).