nature genetics

Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters

Tiffany Hung[1,2], Yulei Wang[3], Michael F Lin[4,5], Ashley K Koegel[1,2], Yojiro Kotake[6–8], Gavin D Grant[9], Hugo M Horlings[10], Nilay Shah[11], Christopher Umbricht[12], Pei Wang[13], Yu Wang[3], Benjamin Kong[3], Anita Langerød[14], Anne-Lise Børresen-Dale[14,15], Seung K Kim[2,13], Marc van de Vijver[10], Saraswati Sukumar[11], Michael L Whitfield[9], Manolis Kellis[4,5], Yue Xiong[6], David J Wong[1,16] & Howard Y Chang[1,2,16]

Transcription of long noncoding RNAs (lncRNAs) within gene regulatory elements can modulate gene activity in response to external stimuli, but the scope and functions of such activity are not known. Here we use an ultrahigh-density array that tiles the promoters of 56 cell-cycle genes to interrogate 108 samples representing diverse perturbations. We identify 216 transcribed regions that encode putative lncRNAs, many with RT-PCR–validated periodic expression during the cell cycle, show altered expression in human cancers and are regulated in expression by specific oncogenic stimuli, stem cell differentiation or DNA damage. DNA damage induces five lncRNAs from the *CDKN1A* promoter, and one such lncRNA, named *PANDA*, is induced in a p53-dependent manner. *PANDA* interacts with the transcription factor NF-YA to limit expression of pro-apoptotic genes; *PANDA* depletion markedly sensitized human fibroblasts to apoptosis by doxorubicin. These findings suggest potentially widespread roles for promoter lncRNAs in cell-growth control.

Mammalian genomes are more pervasively transcribed than previously expected[1–4]. In addition to protein-coding genes, many types of non-coding RNAs (ncRNAs) are transcribed. Small regulatory ncRNAs, including small interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-associated RNAs (piRNAs), function in genome defense and post-transcriptional regulation[5–7]. Near transcriptional start sites (TSS), divergent transcription by RNA polymerase can generate small ncRNAs ranging from 20 to 200 nucleotides, which have been variously named promoter-associated small RNAs (PASRs), transcription-initiation RNAs (tiRNAs) and TSS-associated RNAs (TSSa-RNAs)[8–11]. However, it remains uncertain if these ncRNAs are functional or just represent byproducts of RNA polymerase infidelity[12,13]. lncRNAs vary in length from several hundred bases to tens of kb; lncRNAs may be located in isolation from protein coding genes (long intergenic ncRNAs, or lincRNAs), or they may be interspersed nearby or within protein coding genes[14,15]. Moreover, recent evidence suggest that active enhancer elements are also transcribed to produce ncRNAs[16,17].

Although evidence for function of lncRNAs as a group is lacking, several lncRNAs have been implicated in transcriptional regulation. Two prime examples are in the genomic loci of cell-cycle genes. In the *CCND1* (encoding cyclin D1) promoter, an ncRNA transcribed 2 kb upstream of *CCND1* is induced by ionizing radiation and regulates transcription of *CCND1* in *cis* by forming a ribonucleoprotein repressor complex[18]. This ncRNA binds to and allosterically activates the RNA-binding protein TLS (translated in liposarcoma), which inhibits histone acetyltransferases, resulting in repression of *CCND1* transcription. A second example is an antisense ncRNA gene *CDKN2B-AS1* (also known as *p15AS* or *ANRIL*) that overlaps with the *p15* coding sequence, and *CDKN2B-AS1* expression is increased in human leukemias with an inverse correlation with *p15* expression[19,20]. *CDKN2B-AS1* can transcriptionally silence *p15* directly as well as through induction of heterochromatin formation. Many well-studied lncRNAs, such as those involved in dosage compensation and imprinting, regulate gene expression in *cis*[21], but other lincRNAs, such as *HOTAIR* and *linc-p21*, can regulate the activity of distantly located genes in *trans*[22–24]. Inspired by these examples, we hypothesized that the genomic loci of cell-cycle genes may harbor other functional ncRNAs that have yet to be discovered.

[1]Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, USA. [3]Life Technologies, Foster City, California, USA. [4]The Broad Institute, Cambridge, Massachusetts, USA. [5]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [6]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [7]Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [8]Department of Biochemistry 1, Hamamatsu University School of Medicine, Higashi-ku, Hamamatsu, Japan. [9]Department of Genetics, Dartmouth Medical School, Hanover, New Hampshire, USA. [10]Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands. [11]Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [12]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [13]Department of Developmental Biology, Stanford University School of Medicine, Stanford, California, USA. [14]Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello, Oslo, Norway. [15]Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [16]These authors contributed equally to this work. Correspondence should be addressed to D.J.W. (davewong@stanford.edu) or H.Y.C. (howchang@stanford.edu).

Received 19 August 2010; accepted 6 May 2011; published online 5 June 2011; doi:10.1038/ng.848

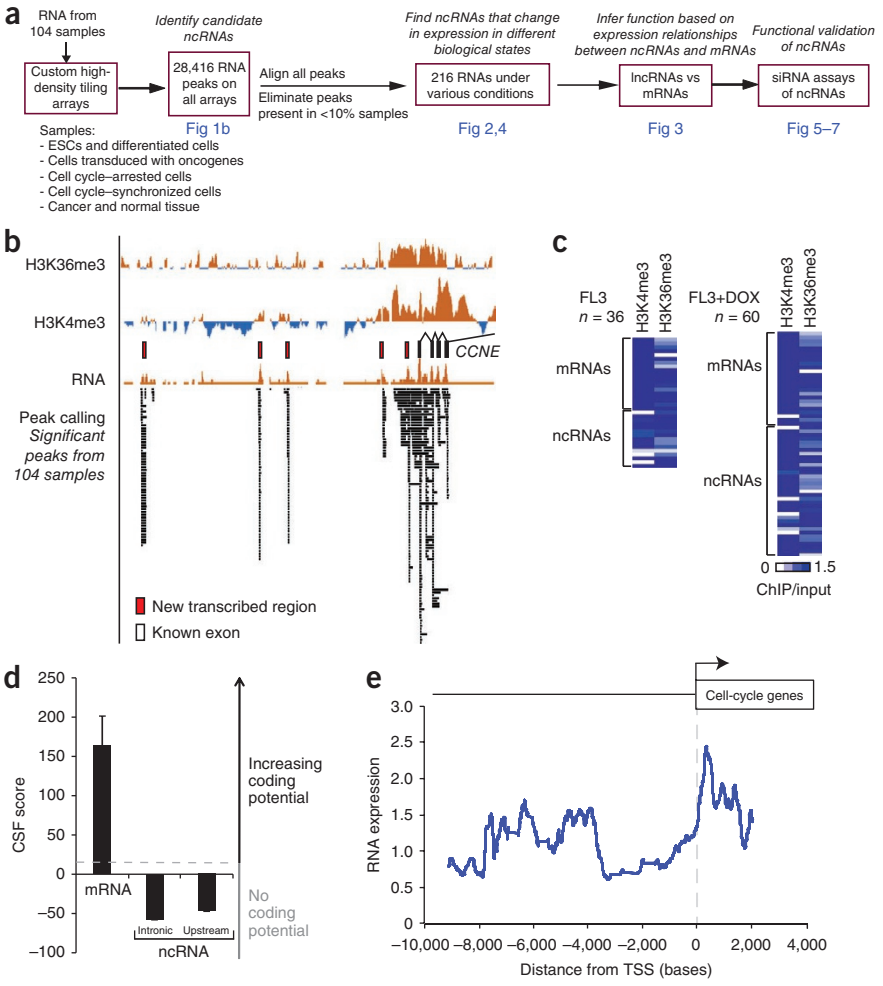

Figure 1 Identification of ncRNAs near and within cell-cycle genes. (**a**) Flow chart of the strategy for systematic discovery of cell-cycle ncRNAs. (**b**) Representative tiling array data. RNA hybridization intensity and H3K36me3 and H3K4me3 ChIP-chip signals relative to the input at the *CCNE1* locus in human fetal lung fibroblasts. The predicted transcripts are shown in red boxes. Known mRNA exons are shown in black boxes. Each bar represents a significant peak from one of the 108 array channels. (**c**) Chromatin state at the transcribed regions. The average ChIP-chip signal relative to the input calculated across transcriptional peaks expressed in human fetal lung fibroblasts with or without doxorubicin treatment. (**d**) Codon substitution frequency (CSF) analysis. Graph of the average evolutionary CSF of the exons of coding genes and their predicted transcripts. CSF < 10 indicates no protein coding potential. (**e**) Transcriptional landscape of cell-cycle promoters. We aligned all cell-cycle promoters at the TSS and calculated the average RNA hybridization signal across the 12-kb window. The output represents a 150-bp running window of average transcription signals across all 54 arrays. See also **Supplementary Table 1** and **Supplementary Figure 1**.

In this study, we create an ultrahigh-resolution tiling microarray to interrogate the transcriptional and chromatin landscape around the TSSs of 56 cell-cycle genes, including genes encoding all cyclins, cyclin-dependent kinases (CDKs) and cyclin-dependent kinase inhibitors (CDKIs). We analyze a diverse collection of cells and tissue samples that interrogate distinct perturbations in cell-growth control. Our results reveal a map of extensive and choreographed noncoding transcription and identify a specific set of lncRNAs that function in the DNA damage response.

RESULTS

Extensive noncoding transcription near cell-cycle genes

To systematically discover functional ncRNAs in the regulatory region of human cell-cycle genes, we created a tiling array that interrogates at 5-nt resolution across 25 kb of the 9p21 locus (which encompasses *CDKN2A* (*p16*), *p14ARF* and *CDKN2B* (*p15*)), as well as from 10 kb upstream to 2 kb downstream of each TSS from 53 cell-cycle genes to include those that encode all known cyclins, CDKs and CDKIs (**Fig. 1a** and **Supplementary Table 1**). These genes are also critical for fundamental biological processes such as senescence, self-renewal, DNA damage response and tumor formation[25–27]. Thus, we hybridized 54 pairs of polyadenylated RNAs from various human cells that were altered or perturbed through cell-cycle synchronization, DNA damage, differentiation stimuli, oncogenic stimuli or carcinogenesis (**Supplementary Table 2**).

A peak calling algorithm searched for statistically significant signals above background and detected contiguous regions (peaks) of at least 50 bp. We then compiled statistically significant transcripts from all 108 channels of the 54 arrays, clustered all transcripts that overlapped by a minimum of 50 bases and identified clusters that were present in at least 10% of the samples. Averaging the signal intensity across all probes in a peak produced a quantitative estimate of transcript abundance. Despite possible 3′ bias caused by polyadenylated RNA selection, our procedure detected exon 1 transcription from the majority of cell-cycle coding genes (41 of the 56), showing that this custom tiling array can detect previously reported transcribed regions. In each individual sample, we detected an average of 73 of the 216 transcribed regions (with a range of 14–189 transcribed regions) that did not overlap with known exons of the 56 cell-cycle genes (**Supplementary Fig. 1**; an example of the *CCNE1* locus in human fetal lung fibroblasts is shown in **Fig. 1b**). Across all 108 samples, we identified a total of 216 discrete transcribed regions (**Supplementary Table 3**). The average transcript length was 234 nt (with a range of 50–1,494 nt). One hundred seventy one of the 216 (79%) previously unidentified transcribed regions were located 5′ of the TSS of the cell-cycle genes ('upstream'), 40 of the 216 (19%) were located within introns ('intronic'), and 5 of the 216 (2%) were located downstream of the 3′ end of *CDKN2A*.

Genes actively transcribed by RNA polymerase II are marked by trimethylation of histone H3 on lysine 4 (H3K4me3) and lysine 36 of histone H3 (H3K36me3), which reflect gene starts and bodies, respectively[28]. These chromatin marks can be used to identify noncoding transcription[14]. In a subset of our samples, we determined whether the 216 transcribed regions were similarly marked for active transcription by performing chromatin immunoprecipitation followed by hybridization to our custom tiling array (ChIP-chip). This analysis confirmed that the chromatin state at a majority of the newly

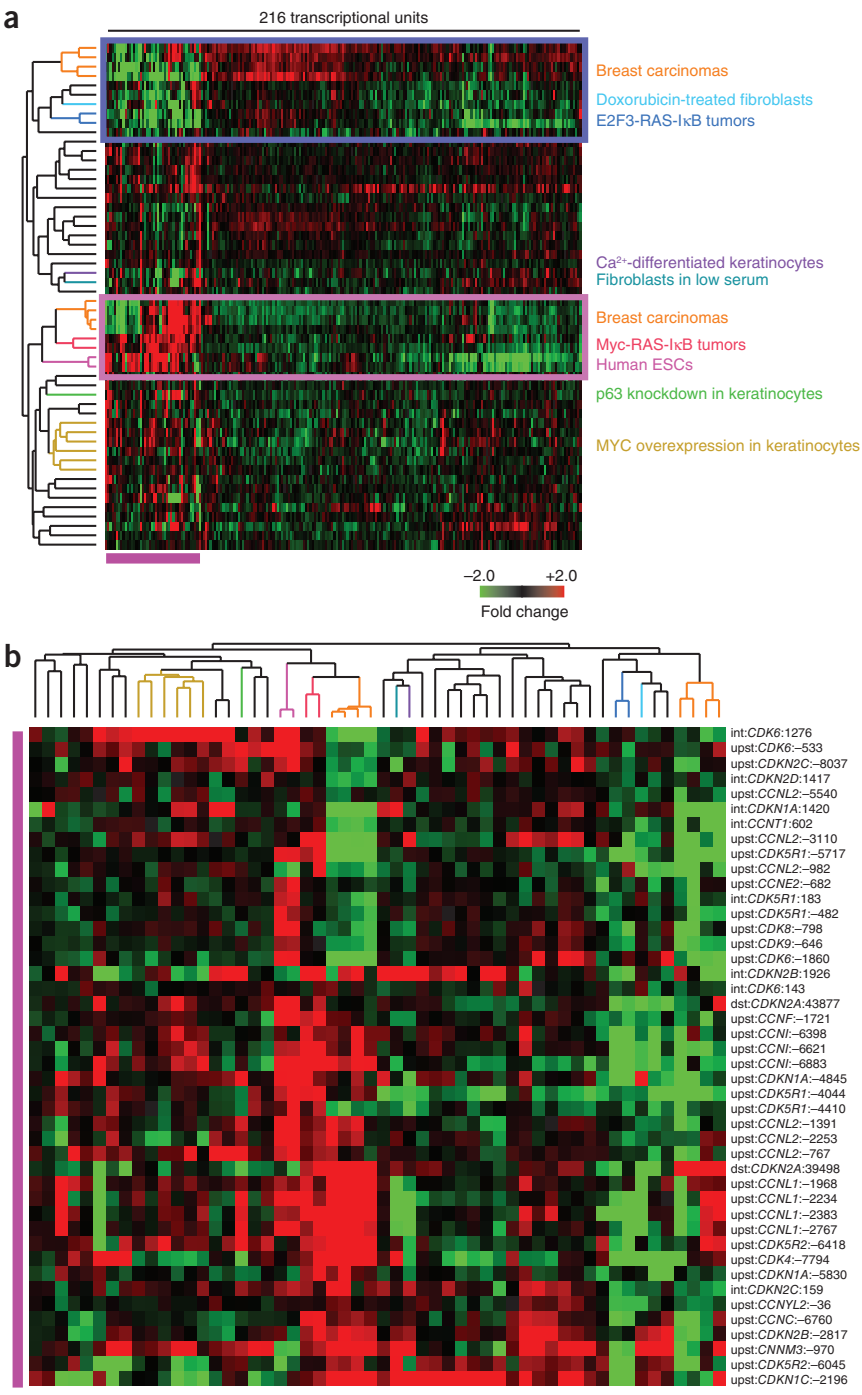

Figure 2 ncRNA expression across diverse cell cycle perturbations. (**a**) Hierarchical clustering of 216 predicted ncRNAs across 54 arrays, representing 108 conditions. Red indicates that the cell cycle perturbation induced transcription of the ncRNA. Green indicates that the cell cycle perturbation repressed transcription of the ncRNA. Black indicates no significant expression change. (**b**) Close up view of the ncRNAs in cluster 1. See also **Supplementary Tables 2,3**.

defined transcripts was enriched in both H3K4me3 and H3K36me3 (**Fig. 1b,c**). Using EpiGRAPH analysis to query our transcripts against approximately 900 published genomic attributes[29], the 216 putative transcribed regions were enriched for H3K4me3 ($P < 10^{-9}$) and RNA polymerase II binding ($P < 10^{-7}$), providing further evidence that these genomic regions are actively transcribed.

To determine whether the 216 transcripts may encode previously unknown protein-coding exons or noncoding RNAs, we used a codon substitution frequency (CSF) analysis to assess for characteristic evolutionary signatures of protein-coding sequences across 21 sequenced mammalian genomes[30]. As expected, the transcribed regions that coincided with annotated exons had high CSF scores. However, over 86% of the new transcribed regions had CSF scores well below the threshold of known protein-coding genes and resembled known ncRNAs (**Fig. 1d** and **Supplementary Table 3**), suggesting that most of the new regions do not have protein-coding potential. BLAST analysis confirmed that the majority of the transcripts are not known protein-coding genes (**Supplementary Table 3**). Furthermore, none of the transcripts intersect known pre-miRNAs, C/D box small nucleolar RNAs, H/ACA box small nucleolar RNAs or small Cajal-body specific RNAs as annotated in the UCSC genome browser. Thereafter, we refered to these transcribed regions as long noncoding RNAs (lncRNAs). We aligned the RNA hybridization signals at all 56 protein-coding loci of all 108 samples relative to their TSS (**Fig. 1e**). As expected, we found a peak immediately downstream of the TSS corresponding to exon 1 of the protein-coding gene. In addition, we found enrichment of non-coding transcription in the region 4–8 kb upstream of the TSS. Thus, unlike the previously described PASRs, tiny RNAs and TSSaRNAs, which are primarily located within 100 bp of the TSS, the majority of these ncRNAs are longer and are not clustered immediately around the TSS.

Expression patterns of ncRNAs suggest specific biological functions

Next, we examined the biological conditions that regulate expression of these ncRNAs in order to infer possible biological functions. We assembled a matrix of the expression changes of the 216 new transcribed regions across all 54 perturbations and hierarchically clustered the genes and samples (**Fig. 2a** and **Supplementary Table 4**). Of the 216 new transcribed regions, 92 (43%) had at least a twofold change in expression detected on the tiling array in at least one of the perturbations, suggesting that a large subset of the transcribed regions may have functional roles. The samples that had the most transcripts with at least twofold expression change were the embryonic stem cells (ESC) relative to day 152 fetal pancreas (40 of 216) and invasive ductal breast carcinomas relative to normal (as many as 35 of 216), suggesting that a subset of these lncRNAs may play a role in self-renewal and carcinogenesis (**Fig. 2a**). Notably, lncRNA expression profiles of keratinocytes with knockdown of P63, which

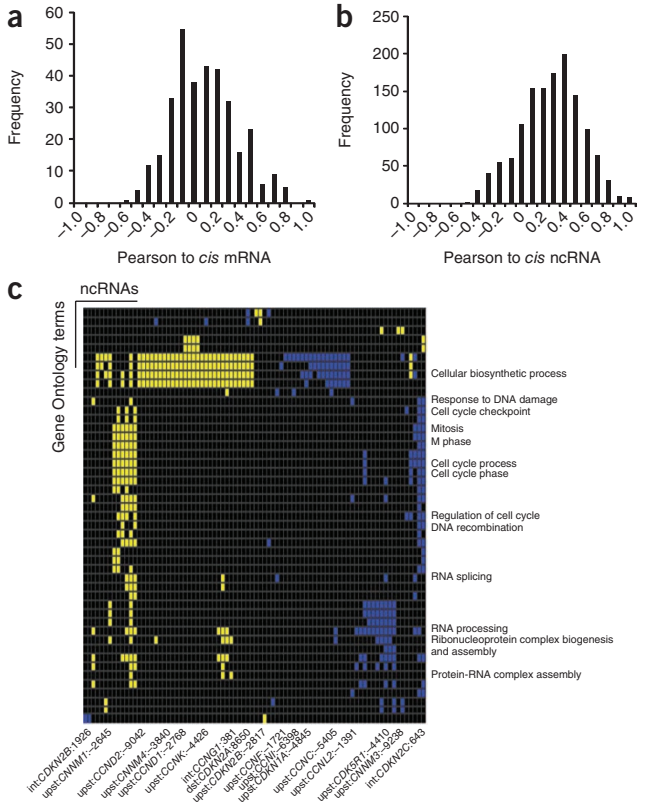

Figure 3 Functional associations of ncRNAs. (**a**) lncRNA expression patterns do not correlate with those of the mRNAs in *cis*. Histogram of Pearson correlations between each of the 216 ncRNAs and the *cis* mRNA across 108 samples. (**b**) lncRNA expression patterns have a positive correlation with neighboring lncRNA transcripts. Histogram of Pearson correlations between each of the 216 ncRNAs and nearby transcripts on the same locus across 108 samples. (**c**) Genes co-expressed with lncRNAs are enriched for functional groups in the cell cycle and in DNA damage response. Module map of lncRNA gene sets (columns) versus Gene Ontology Biological Processes gene sets (rows) across 17 samples ($P < 0.05$, false discovery rate <0.05). A yellow entry indicates that the Gene Ontology gene set is positively associated with the lncRNA gene set. A blue entry indicates that the Gene Ontology gene set is negatively associated with the lncRNA gene set. A black entry indicates no significant association. Representative enriched Gene Ontology gene sets are listed.

inhibits keratinocyte differentiation, clustered with that of ESC, suggesting that these ncRNAs may have a role in the undifferentiated state. Expression patterns from five keratinocyte samples that were transduced with the oncogene *MYC* alone or in combination with other oncogenes relative to controls clustered together, showing that *MYC* has a dominant effect on ncRNA expression. MYC-RAS-IκBα transduced human keratinocytes activate an ESC-like mRNA gene expression program and acquire properties of cancer stem cells[31]. Notably, the lncRNA expression profile of MYC-RAS-IκBα cells clustered with that of ESCs (**Fig. 2**), suggesting a shared lncRNA signature for embryonic and cancer stem cells. In contrast, the E2F3-RAS-IκBα transduced keratinocytes, which do not express the ESC-like mRNA gene expression program, had an inverse pattern of expression for the majority of lncRNAs. In addition, eight primary human invasive ductal breast carcinomas split into two different groups based on their lncRNA profiles: four of the cancers clustered with the ESCs and MYC-RAS-IκBα tumors, and the other four clustered with the E2F3- RAS-IκBα tumors, suggesting that these tumor models mimic the expression pattern of not only mRNAs but also these lncRNAs in *bona fide* human cancers.

The 216 lncRNAs are divided into three main clusters based on their expression pattern across all samples (**Fig. 2**). Notably, cluster 1 is composed of lncRNAs that are strongly induced in ESCs, keratinocytes with P63-knockdown and MYC-RAS-IκB tumors relative to differentiated cells and GFP-RAS-IκB tumors, which we interpret to be a 'stemness cluster' (**Fig. 2b**). Notably, each cluster is composed of many of the ncRNAs from the same genomic locus, suggesting that multiple adjacent ncRNAs are either coordinately regulated in a shared response or are spliced together as exons of one transcript. High correlation of the dynamic expression patterns of these ncRNAs and different biological and cellular conditions suggest that these ncRNAs may be functional in the cell cycle, in self-renewal and in cancer.

A gene co-expression map infers *trans* regulatory mechanisms and biological functions

Multiple lncRNAs, including *p15AS* and the lncRNA upstream of *CCND1*, have been shown to regulate the transcription of the nearby coding gene. To determine whether gene-proximal lncRNAs are typically correlated with the expression of the nearest mRNA, we conducted whole-genome expression arrays on 17 samples that were also examined on our tiling array and calculated pairwise Pearson correlations between the expression patterns of each cell-cycle promoter lncRNA versus every mRNA genome wide. Notably, there was no significant correlation or anti-correlation between most of the 216 lncRNAs and the nearby protein-coding mRNA, suggesting that most of the lncRNAs may not function in *cis* to activate or repress nearby mRNA expression (**Fig. 3a**). Quantitative RT-PCR (qRT-PCR) analysis of lncRNAs and neighboring 5′ and 3′ mRNAs in 34 additional samples confirmed these findings (**Supplementary Fig. 2**). In contrast, we found that the median correlation between two ncRNAs of the same locus was positive, supporting our hypothesis that neighboring ncRNAs may be coordinately regulated, positively regulate each other and/or are exons of the same transcript (**Fig. 3b**).

Given that expression of the 216 ncRNAs does not generally correlate with the mRNA in *cis*, we further explored the genes and pathways that they may regulate using a guilt-by-association approach[14]. For each lncRNA, we defined a co-expression gene set as the group of mRNAs that are positively or negatively correlated with that lncRNA across the 17 samples ($R > 0.5$ or $R < 0.5$, respectively) (**Supplementary Fig. 3**). We then constructed a gene module map[32] of the association of each lncRNA co-expression gene set versus the Gene Ontology Biological Processes gene set and performed biclustering to identify lncRNAs that are associated with distinct Gene Ontology terms (**Fig. 3c**). This analysis revealed multiple sets of lncRNAs that are associated with biological processes including cell cycle, DNA recombination, ribonucleoprotein complex biogenesis and assembly, RNA splicing, and response to DNA damage. Thus, despite having limited correlation in expression to their neighboring protein-coding gene, the expression patterns of these lncRNAs are still strongly related to the cell cycle. We constructed a similar module map with curated gene sets of metabolic and signaling pathways as well as biological and clinical states from the Molecular Signatures Database (MSigDB c2 collection)[33]. This module map confirmed the enrichment for cell-cycle–related sets (for example, Cell Cycle Brentani or Cell Cycle KEGG). In addition, enriched modules included several poor prognosis breast cancer gene sets (BRCA estrogen receptor negative, BRCA prognosis negative and BRCA1 overexpressed up), DNA-damage–related gene sets (UVA/UVB), several oncogenic signatures

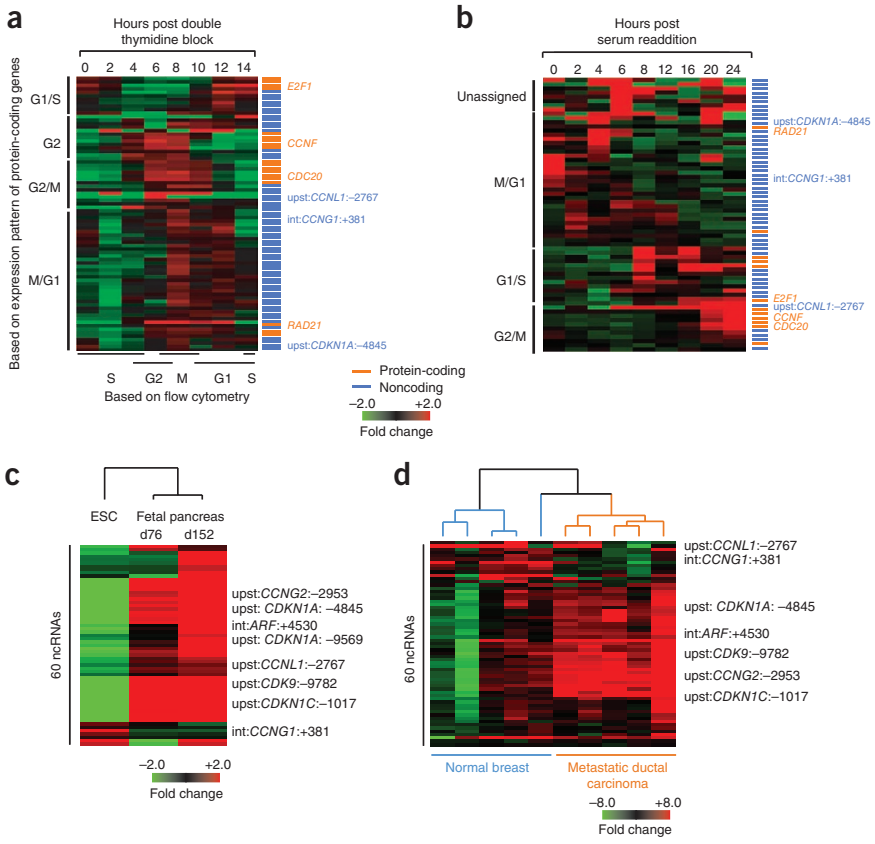

Figure 4 Validated expression of ncRNAs in cell cycle progression, ESC differentiation and human cancers. We generated custom TaqMan probes and used them to interrogate independent biological samples for lncRNA expression. (**a**,**b**) Periodic expression of lncRNAs (blue) during synchronized cell cycle progression in HeLa cells (**a**) and foreskin fibroblasts (**b**). Cell cycle phases were confirmed by fluorescence-activated cell sorting and expression of genes with known periodic expression in the cell cycle (orange). (**c**) Regulated expression of lncRNAs in human ESCs compared to fetal pancreas. D, day. (**d**) Differential expression of lncRNAs in normal breast epithelium compared to breast cancer samples.

(RAS, MYC) and stem-cell gene sets (hematopoietic stem cell, neural stem cell) (**Supplementary Fig. 4**).

Validation of ncRNA expression in cell cycle, ESC differentiation, cancer and DNA damage response

To validate these inferred functional associations, we designed qRT-PCR assays for 60 of the 216 new transcribed regions (53 upstream and 7 intronic) to obtain a more quantitative measure of these lncRNAs across different conditions. Expression in HeLa cells synchronized in cell cycle progression by double thymidine block showed that most of the lncRNA have periodic expression peaking at different phases of the cell cycle (**Fig. 4a**)[34]. Parallel analysis in primary human fibroblasts synchronized by serum stimulation confirmed the peak cell cycle phase of 74% of the lncRNAs with periodic expression pattern during the cell cycle (**Fig. 4b**). Next, comparison of human ESCs and fetal pancreas at days 76 and 152 showed that a majority of these lncRNAs are regulated during differentiation (**Fig. 4c**). In addition, unsupervised clustering of lncRNA expression patterns in five metastatic breast cancers and five normal mammary tissues readily distinguished the five metastatic breast cancers from the normal mammary tissues (**Fig. 4d**). Some of the lncRNAs, including upst:*CCNL1*:–2,767 and int:*CDKN1A*:+885 (**Supplementary Table 3**), are repressed in the metastatic breast cancers relative to normal mammary tissues, whereas others, including upst:*CDKN1A*:–4,845, upst:*CDKN2B*:–2,817 and int:*ARF*:+4,517, are induced. Thus, the majority of these lncRNAs have periodic expression in the cell cycle and are differentially expressed in different states of cell differentiation and cancer progression.

Our co-expression maps predicted associations of several lncRNAs with DNA damage response pathways (**Fig. 3c** and **Supplementary Fig. 3**). In support of this finding, doxorubicin-treated human fetal lung fibroblasts showed at least two-fold change in 12 of the 216 ncRNAs on the tiling array and by qRT-PCR (**Fig. 2**). Notably, 2 of those 12 ncRNAs were located 5′ of the TSS of the canonical p53 target gene *CDKN1A* (upst:*CDKN1A*:–1,210 and upst:*CDKN1A*:–4,845), and, similar to the *CDKN1A* mRNA, were induced by doxorubicin (**Fig. 5a**). In addition, a third lncRNA at the *CDKN1A* locus, upst:*CDKN1A*:–800, was also induced by doxorubicin but was not included in the 216 lncRNAs because it was only expressed in one of the 108 samples, the doxorubicin-treated fibroblasts. In order to confirm whether these lncRNAs may be responsive to DNA damage, we measured the expression changes of 60 lncRNAs predicted in the DNA damage pathway (as well as upst:*CDKN1A*:–800) by quantitative RT-PCR in human fetal lung fibroblasts treated with doxorubicin over a 24-h time course. Most of the lncRNAs were either markedly induced or repressed by doxorubicin, and all five of the tested lncRNAs surrounding the *CDKN1A* TSS were induced, including the three that were previously detected on the tiling array (**Fig. 5b**). Notably, several lncRNAs upstream of *CDKN1A* are induced more rapidly and with substantially higher magnitude than *CDKN1A* upon DNA damage. Upst:*CDKN1A*:–4,845 is induced up to 40-fold upon DNA damage (**Fig. 5c**). These variations in expression patterns within the same locus suggest that the lncRNAs in the *CDKN1A* locus may play distinct roles in the DNA damage response from the *CDKN1A* protein, p21.

PANDA: a long ncRNA involved in the DNA-damage response

To investigate the functional relevance of these lncRNAs at the *CDKN1A* locus, we selected upst:*CDKN1A*:–4,845, hereafter termed *PANDA* (P21 associated ncRNA DNA damage activated), for further analysis. *PANDA* is located approximately 5 kb upstream of the *CDKN1A* TSS, coincides with a cluster of previously annotated expressed sequence tags and is evolutionarily conserved (**Supplementary Fig. 5**). Although the *PANDA* locus intersects a computationally predicted pseudogene of *LAP3*, qRT-PCR showed that *PANDA* was specifically induced by DNA damage, whereas *LAP3* expression did not significantly change, confirming that the change in expression detected by the tiling array was not caused by cross hybridization with *LAP3* (**Supplementary Fig. 6**). Furthermore, the CSF score of *PANDA*, 9.3, indicated very low protein-coding potential compared to *LAP3* (with a CSF range of 117–1,343 for its 13 exons). Rapid amplification of the 5′ and 3′ complementary DNA

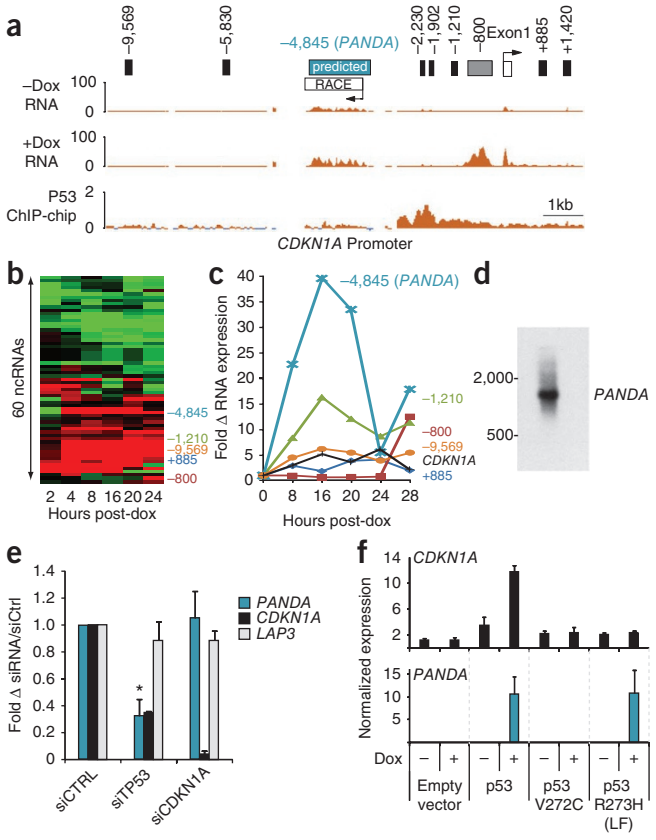

Figure 5 ncRNAs at the *CDKN1A* locus are induced by DNA damage. (**a**) At the top is a map of all detected transcripts at the CDKN1A promoter. In the middle two tracks are examples of RNA hybridization intensity in the control or in 24 h doxorubicin (dox) treated (200 ng/ml) human fetal lung fibroblasts. Note that we did not observe all DNA-damage–inducible transcripts in one single time point. At the bottom, the p53 ChIP-chip signal relative to input confirmed the p53 binding site immediately upstream of the *CDKN1A* TSS after DNA damage. The RACE clone of upst:*CDKN1A*:−4,845 closely matches the predicted transcript on the tiling array. See also **Supplementary Figure 7**. (**b**) Quantitative RT-PCR of lncRNAs shows coordinate induction or repression across a 24 h time course of doxorubicin treatment. A cluster of lncRNAs transcribed from the *CDKN1A* locus are induced. (**c**) Expression of transcripts from the *CDKN1A* locus over a 24 h time course after doxorubicin treatment of normal human fibroblasts (FL3). See also **Supplementary Figure 6**. (**d**) RNA blot of *PANDA* confirms a transcript size of 1.5 kb. (**e**) Doxorubicin induction of *PANDA* requires p53 but not CDKN1A. Mean ± s.d. are shown; $*P < 0.05$ relative to siCTRL (control siRNA) determined by student's *t*-test. (**f**) Expression of wild-type p53 in p53-null H1299 cells restores DNA damage induction of *CDKN1A* and *PANDA*. The p53 (p.Val272Cys) loss-of-function mutant fails to restore induction, whereas a gain-of-function Li-Fraumeni allele, p53 (p.Arg273His), selectively retains the ability to induce PANDA.

ends (RACE) and RNA blot analysis revealed a 1.5-kb transcript that is divergently transcribed from *CDKN1A*, antisense of the predicted *LAP3* pseudogene (**Fig. 5d** and **Supplementary Fig. 7**). Thus, *PANDA* is a 5′-capped and polyadenylated non-spliced lncRNA that is transcribed antisense to *CDKN1A*.

Because p53 is a positive regulator of *CDKN1A* during the DNA damage response, we asked whether p53 also regulates *PANDA* expression. ChIP-chip analysis confirmed the p53 binding site immediately upstream of the *CDKN1A* TSS (**Fig. 5a**)[35]. *PANDA* and *CDKN1A* are diametrically situated 2.5 kb from this intervening p53 binding site, which supports the possibility of p53 co-regulation. Indeed, siRNA-mediated knockdown of p53 before DNA damage inhibited the induction of *PANDA* by 70% 24 h after DNA damage (**Fig. 5e** and **Supplementary Fig. 8**), which is similar to its effect on *CDKN1A*. In contrast, RNA interference of *CDKN1A* had no effect on *PANDA* expression, indicating that PANDA is not a linked transcript of *CDKN1A* nor is PANDA expression dependent on p21. *PANDA* level shows a trend of lower expression in human primary breast tumors harboring inactivating mutation in *TP53* as determined by exon 2–11 DNA sequencing[36] (**Supplementary Fig. 9a**). Further, complementation of p53-null H1299 lung carcinoma cells by wild-type p53—but not the loss-of-function p53 (p.Val272Cys) mutant—restored DNA damage-inducible expression of PANDA (**Fig. 5f**). Notably, a gain-of-function p53 (p.Arg273His) mutant, observed in Li-Fraumeni syndrome[37], abrogated the ability to induce *CDKN1A* but selectively preserved the ability to induce *PANDA* (**Fig. 5f**). We also observed selective induction of *PANDA* without concordant *CDKN1A* expression in metastatic ductal carcinomas but not in normal breast tissue (**Supplementary Fig. 9b**).

Next, we addressed whether *PANDA* affects the DNA damage response. We transduced human fetal fibroblasts (FL3) with custom siRNAs targeting *PANDA* and then applied doxorubicin for 24 h following the knockdown (**Fig. 6a**). Global gene expression analysis showed that 224 genes were induced and 193 genes were repressed at least twofold by *PANDA* knockdown (**Fig. 6b**). Genes induced by *PANDA* knockdown were significantly enriched for those involved in apoptosis, such as the Gene Ontology terms 'cell death' ($P < 0.04$) and 'apoptosis' ($P < 0.03$) (**Fig. 6b**). qRT-PCR confirmed that *PANDA* depletion induced several genes encoding canonical activators of apoptosis, including *APAF1*, *BIK*, *FAS* and *LRDD* (**Fig. 6c**). On the other hand, expression of neither *CDKN1A* itself nor *TP53* was affected by *PANDA* depletion (**Fig. 6d**), suggesting that *PANDA* is a P53 effector that acts independently of $p21^{CDKN1A}$.

DNA damage in human fibroblasts triggers p53-dependent G1 arrest but not apoptosis[38,39]. Consistent with this finding, doxorubicin treatment in FL3 cells exposed to control siRNA had little to no apoptosis as measured by TUNEL. In contrast, *PANDA* knockdown resulted in fivefold to sevenfold increased TUNEL-positive cells (**Fig. 6e**,**f**). Immunoblot analysis of PARP, a caspase substrate and marker of apoptosis, revealed PARP cleavage only in *PANDA*-depleted cells (**Fig. 6g**). In contrast, six additional siRNAs targeting other transcripts within the *CDKN1A* promoter had no effect on apoptosis (data not shown; **Supplementary Fig. 10**). Thus, PANDA knockdown sensitized fibroblasts to DNA-damage–induced apoptosis. Altogether, these data suggest that in parallel with p53-mediated induction of *CDKN1A* for cell cycle arrest, p53-mediated induction of *PANDA* delimits apoptosis.

Core promoters of cell death genes downstream of p53 are distinguished from other p53 target genes by the binding site for the transcription factor NF-YA[40], and we reasoned that *PANDA* may affect NF-YA function. RNA chromatography[41] using purified, *in vitro* transcribed PANDA RNA, but not a 1.2-kb *LacZ* mRNA fragment, specifically retrieved NF-YA from cellular lysates of human fibroblasts induced by DNA damage (**Fig. 7a**). PANDA did not retrieve other chromatin modification complexes that can bind other lncRNAs, such as EZH2 or LSD1 (refs. 42,43), or p21, illustrating the specificity of the interaction. Immunoprecipitation of NF-YA from doxorubicin-treated primary human lung fibroblasts specifically retrieved endogenous

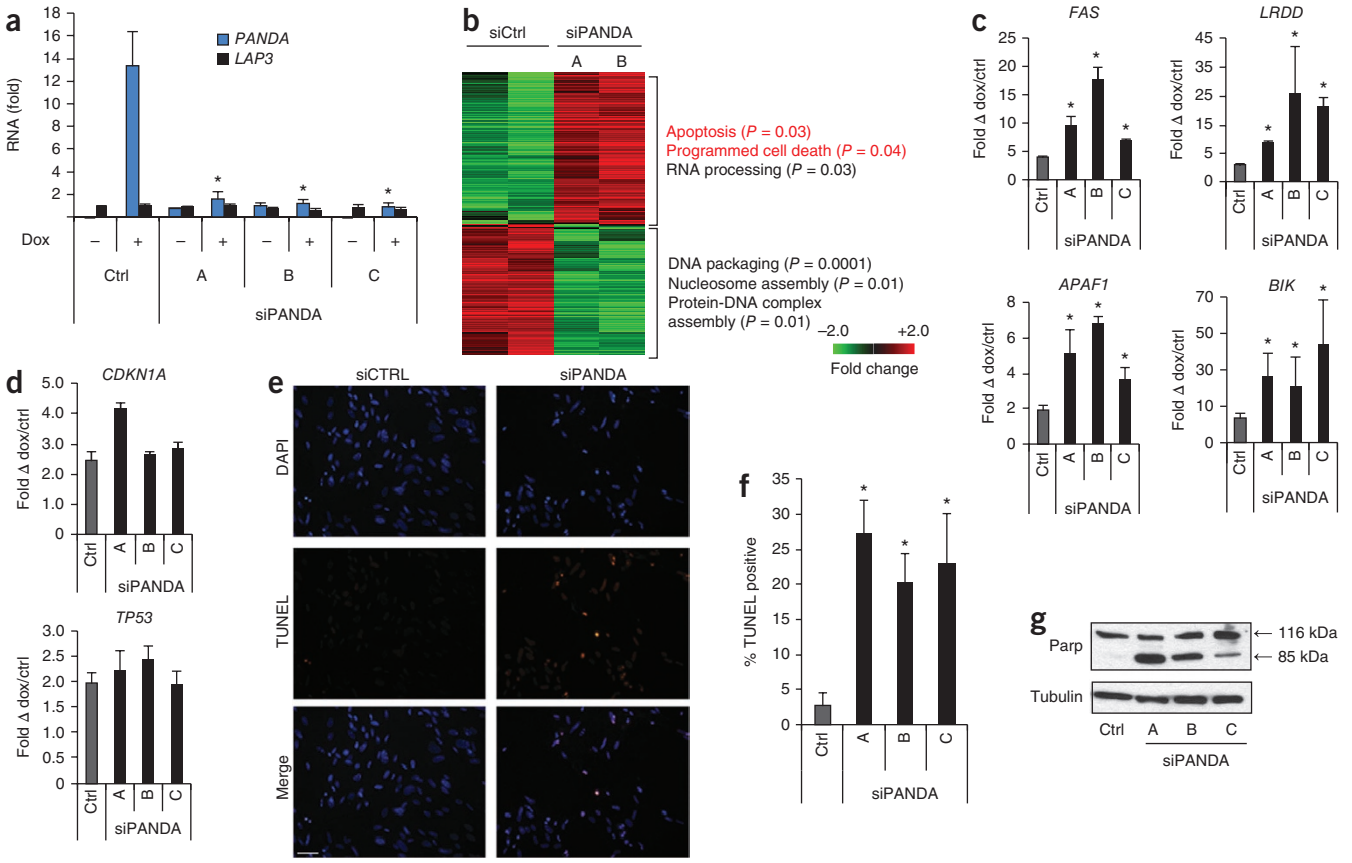

Figure 6 *PANDA* lncRNA regulates the apoptotic response to DNA damage. (**a**) siRNA knockdown of PANDA in the presence of DNA damage with doxorubicin in human fibroblasts (FL3). Custom siRNAs specifically target *PANDA* with no discernable effect on the *LAP3* mRNA. Mean ± s.d. are shown in all bar graphs. $*P < 0.05$ compared to siCTRL for all panels determined by Student's *t*-test. (**b**) Heat map of gene expression changes with siPANDA relative to control siRNA after 24 h of doxorubicin treatment in FL3 cells. (**c**) Quantitative RT-PCR of canonical apoptosis pathway genes revealed induction with siPANDA relative to control siRNA after 28 h of doxorubicin treatment (in FL3 cells). (**d**) Quantitative RT-PCR of *CDKN1A* and *TP53* in FL3 cells revealed no reduction in expression with siPANDA relative to control siRNA. (**e**) TUNEL immunofluorescence of control and siPANDA FL3 fibroblasts after 28 h of doxorubicin treatment. Scale bar, 20 μm. (**f**) Quantification of three independent TUNEL assays. $P < 0.05$ for each siPANDA sample compared to siCTRL determined by student's *t*-test. (**g**) Protein blot of PARP cleavage in control and PANDA siRNA FL3 fibroblasts after 24 h of doxorubicin treatment.

PANDA (**Fig. 7b**). NF-YA is a nuclear transcription factor that activates the p53-responsive promoter of *FAS* upon DNA damage[40]. Depletion of PANDA substantially increased NF-YA occupancy at target genes, including *CCNB1*, *FAS*, *BBC3* (also known as *PUMA*) and *PMAIP1* (also known as *NOXA*) (**Fig. 7c**). Moreover, concomitant knockdown of NF-YA and PANDA substantially attenuated induction of apoptotic genes and apoptosis as measured by TUNEL, indicating that NF-YA is required in part for cell death triggered by loss of PANDA (**Fig. 7d,e**). Thus, PANDA binding to NF-YA may evict or prevent NF-YA binding to chromatin. These data suggest that DNA damage activates p53-mediated transcription at *CDKN1A* and *PANDA* that functions synergistically to mediate cell cycle arrest and survival. *CDKN1A* mRNA produces p21 to mediate arrest, whereas PANDA impedes NF-YA activation of apoptotic gene expression program (**Fig. 8**).

DISCUSSION

Recent studies have revealed that a surprisingly large fraction of mammalian genomes is transcribed. In addition to small noncoding RNAs, long noncoding RNAs can be produced from gene promoters and enhancers, as well as stand-alone intergenic loci[14,15,17]. New approaches are needed that not only identify ncRNAs but also provide insight into their potential biological function. Using an ultrahigh-resolution tiling array, we interrogated the transcriptional landscape at cell-cycle promoters in 108 samples that represent diverse perturbations. The ability to interrogate numerous and diverse biological samples in a rapid and economical fashion is advantageous for at least two reasons. First, many of the noncoding transcripts are induced only in highly specific conditions and may have been missed if only a few conditions were surveyed. Of the 216 new noncoding transcribed regions we identified, on average, only 73 of these are transcribed in any one biological sample. Second, comparison of lncRNA profiles amongst these diverse samples highlighted unexpected similarities in cell cycle promoter states among distinct perturbations. For instance, we identified a similarity of promoter states among ESCs, tumors induced by MYC and epithelial progenitors depleted of the differentiation regulator p63. Likewise, authentic human tumors can be classified based on the similarity of their promoter states to those of cells with defined oncogenic perturbation.

Noncoding transcription through regulatory elements may affect gene activity in a variety of ways. The act of transcription may open compacted chromatin over regulatory sequences or compete with transcription factor binding (so called transcriptional interference). In addition, the ncRNA product may modulate neighboring gene expression in *cis*[21,44], affect distantly located genes in *trans*[22] or even

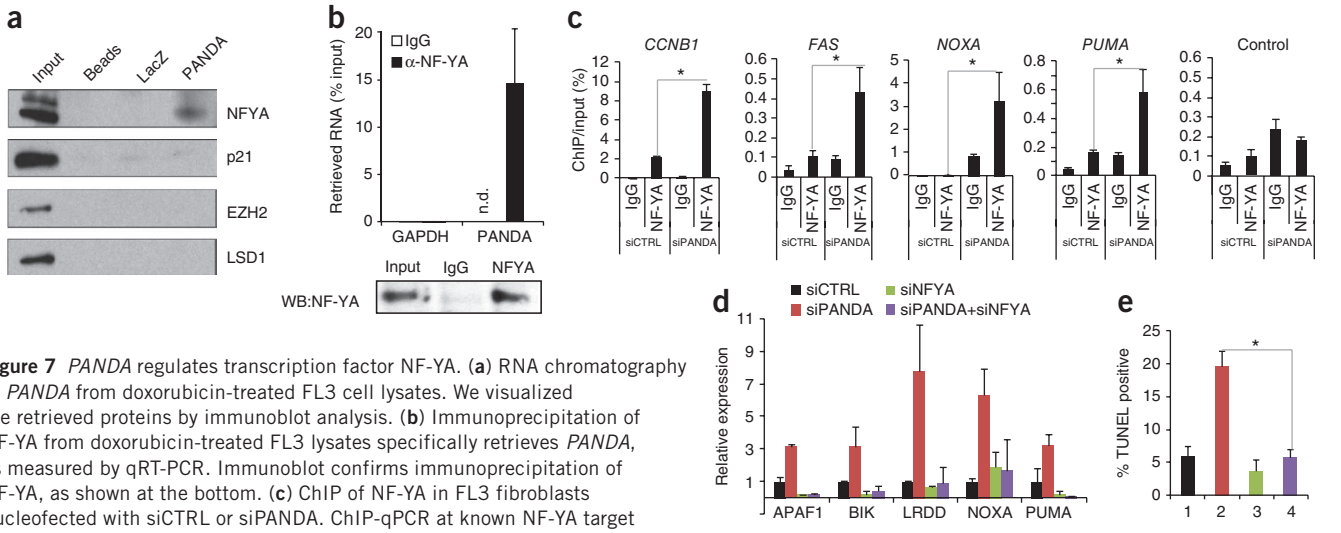

Figure 7 *PANDA* regulates transcription factor NF-YA. (**a**) RNA chromatography of *PANDA* from doxorubicin-treated FL3 cell lysates. We visualized the retrieved proteins by immunoblot analysis. (**b**) Immunoprecipitation of NF-YA from doxorubicin-treated FL3 lysates specifically retrieves *PANDA*, as measured by qRT-PCR. Immunoblot confirms immunoprecipitation of NF-YA, as shown at the bottom. (**c**) ChIP of NF-YA in FL3 fibroblasts nucleofected with siCTRL or siPANDA. ChIP-qPCR at known NF-YA target sites on promoters of *CCNB1*, *FAS*, *NOXA*, *BBC3* (*PUMA*) or a control downstream region in the *FAS* promoter lacking the NF-YA motif. Mean ± s.d. are shown in all bar graphs. $*P < 0.05$ determined by Student's *t*-test. (**d**) Concomitant knockdown of NF-YA attenuates induction of apoptotic genes by PANDA depletion, as measured by qRT-PCR. For knockdown efficiency see **Supplementary Figure 11**. (**e**) Concomitant knockdown of NF-YA rescues apoptosis induced by PANDA depletion. Quantification of TUNEL staining is shown. The legend for this panel is as in **d**.

serve as a target for regulation by small regulatory RNAs[45,46]. Because these different mechanisms predict distinct relationships between levels of ncRNAs and cognate mRNAs, we compared ncRNA and mRNA expression profiles across our samples. We found that most promoter ncRNAs are neither positively nor negatively correlated in expression with their neighboring mRNA but are rather correlated in expression with genes located elsewhere in the genome. The genes co-expressed (and presumably co-regulated) with promoter ncRNAs function in specific biological pathways, including cell cycle, DNA damage response and stem cell differentiation, and have been associated with cancer prognosis. qRT-PCR analysis further validated that many of these ncRNAs are differentially expressed in the cell cycle and in human cancers, and are regulated in response to DNA damage or ESC differentiation. These findings suggest that cell-cycle ncRNAs may participate in gene regulation in *trans*. In addition, noncoding transcription of cell-cycle promoters may be a form of regulatory anticipation or feedback to modulate the chromatin state of cell-cycle promoters.

Our results suggest that the human genome is organized into genomic units that code for multiple transcripts that function in the same biological pathways (**Fig. 8**). Forty nine of 56 cell-cycle protein-coding gene loci have at least one detected lncRNA and an average of four lncRNAs within 10 kb upstream and 2 kb downstream of the TSS. At the *CDKN1A* promoter, five lncRNAs, similar to the *CDKN1A* mRNA itself, are induced by DNA damage. One of these lncRNAs, which we named *PANDA*, is a non-spliced 1.5-kb ncRNA that is transcribed antisense to *CDKN1A* and is induced with faster kinetics than *CDKN1A*. Loss-of-function and complementation experiments show that *PANDA* induction during DNA damage is p53 dependent. In contrast, depletion of *CDKN1A* or depletion of *PANDA* had no effect on the other's response to DNA damage, indicating that their induction by p53 occurs in parallel. *PANDA* inhibits the expression of apoptotic genes by sequestering the transcription factor NF-YA from occupying target gene promoters. Whereas *CDKN1A* encodes a cell cycle inhibitor to mediate cell cycle arrest, *PANDA* promotes cell survival by impeding the apoptotic gene expression program. This linkage can be apparently exploited by tumors: the ability of the Li-Fraumeni gain-of-function P53 mutant p.Arg273His to selectively retain *PANDA* induction instead of *CDKN1A* in effect uncouples cell survival from cell cycle arrest, which was similarly observed in metastatic ductal carcinomas. Thus, lncRNAs like *PANDA* may provide new explanations for human cancer susceptibility.

Intriguingly, a recent study identified a distinct long intergenic noncoding RNA located 15 kb upstream of *CDKN1A*, named *lincRNA-p21*, that is induced by p53 and mediates p53-dependent gene repression[24]. Thus, the regulatory sequence upstream of *CDKN1A* drives the expression of multiple coding and noncoding transcripts that cooperate to regulate the DNA damage response (**Fig. 8**). These findings provide a vivid example that shows the blurring boundary between 'genes' and 'regulatory sequences'[47].

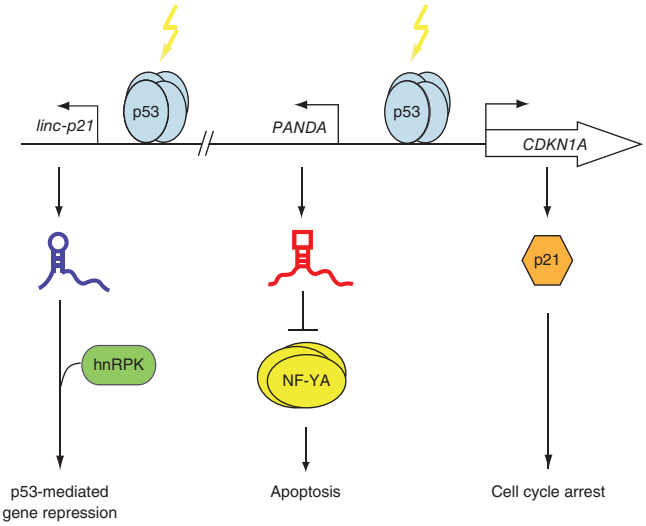

Figure 8 Model of coding and noncoding transcripts at the *CDKN1A* locus coordinating the DNA damage response. After DNA damage, p53 binding at the *CDKN1A* locus coordinately activates transcription of *CDKN1A* as well as noncoding transcripts PANDA and linc-p21. CDKN1A mediates cell cycle arrest, PANDA blocks apoptosis through NF-YA, and linc-p21 mediates gene silencing through recruitment of hnRPK.

Our study provides an initial catalog of lncRNAs in cell-cycle promoters that may play diverse functions. At a minimum, promoter ncRNA expression provides a convenient means of tracking the chromatin state of promoters, which may be of use in cancer biology and regenerative medicine. Future studies are needed to pinpoint the functions of these and likely other ncRNAs emanating from regulatory sequences.

URLs. Genomica, http://genomica.weizmann.ac.il/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Tiling and microarray data are available at Gene Expression Omnibus (GSE28631). Sequence for human *PANDA* RNA has been deposited with GenBank under the accession number JF803844.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J. Rinn, M. Guttman and A. Regev for discussions, L. Attardi for careful reading of the manuscript and P. Khavari for reagents. Y.W., B.K. and Yu Wang are employees of Life Technologies. This work was supported by grants from the US National Institutes of Health (NIH)/National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) (K08-AR054615 to D.J.W.), NIH/National Cancer Institute (NCI) (R01-CA118750 to H.Y.C. and R01-CA130795 to M.L.W.), the Juvenile Diabetes Research Foundation (S.K.K. and H.Y.C.) and the American Cancer Society (H.Y.C.). H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute. T.H. is supported by the Stanford Graduate Fellowship, the National Science Foundation (NSF) Graduate Research Fellowship and the Department of Defense (DoD) National Defense Science & Engineering Graduate Fellowship (NDSEG).

AUTHOR CONTRIBUTIONS

H.Y.C. and D.J.W. initiated the project. H.Y.C., D.J.W. and T.H. designed the experiments. T.H. performed the experiments and the computational analysis. Yulei Wang, Yu Wang and B.K. conducted high-throughput TaqMan RT-PCRs. M.F.L. and M.K. contributed CSF analysis. The following authors contributed samples or reagents: A.K.K., Y.K., G.D.G., H.M.H., N.S., C.U., P.W., A.L., S.K.K., M.v.d.V., A.-L.B.-D., S.S., M.L.W. and Y.X. The manuscript was prepared by H.Y.C., T.H. and D.J.W. with input from all co-authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242–2246 (2004).
2. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
3. Calin, G.A. *et al.* Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215–229 (2007).
4. Carninci, P. Non-coding RNA transcription: turning on neighbours. *Nat. Cell Biol.* **10**, 1023–1024 (2008).
5. Mattick, J.S. & Makunin, I.V. Small regulatory RNAs in mammals. *Hum. Mol. Genet.* **14**Spec No 1, R121–R132 (2005).
6. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5**, 522–531 (2004).
7. Hutvagner, G. & Simard, M.J. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* **9**, 22–32 (2008).
8. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).
9. Seila, A.C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
10. Taft, R.J. *et al.* Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* **41**, 572–578 (2009).
11. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
12. Ponjavic, J., Ponting, C.P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556–565 (2007).
13. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* **14**, 103–105 (2007).
14. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
15. Katayama, S. *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309**, 1564–1566 (2005).
16. Kim, T.K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
17. De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
18. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature* **454**, 126–130 (2008).
19. Pasmant, E. *et al.* Characterization of a germ-line deletion, including the entire *INK4/ARF* locus, in a melanoma-neural system tumor family: identification of *ANRIL*, an antisense noncoding RNA whose expression coclusters with *ARF*. *Cancer Res.* **67**, 3963–3969 (2007).
20. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* **451**, 202–206 (2008).
21. Lee, J.T. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* **23**, 1831–1842 (2009).
22. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
23. Gupta, R.A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
24. Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
25. Sherr, C.J. & Roberts, J.M. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.* **13**, 1501–1512 (1999).
26. Hall, M. & Peters, G. Genetic alterations of cyclins, cyclin-dependent kinases, and Cdk inhibitors in human cancer. *Adv. Cancer Res.* **68**, 67–108 (1996).
27. Johnson, D.G. & Walker, C.L. Cyclins and cell cycle checkpoints. *Annu. Rev. Pharmacol. Toxicol.* **39**, 295–312 (1999).
28. Rando, O.J. & Chang, H.Y. Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* **78**, 245–271 (2009).
29. Bock, C., Halachev, K., Buch, J. & Lengauer, T. EpiGRAPH: user-friendly software for statistical analysis and prediction of (epi)genomic data. *Genome Biol.* **10**, R14 (2009).
30. Lin, M., Jungreis, I. & Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein-coding and non-coding regions. *Nature Precedings* published online, doi:10.1038/npre.2010.4784.1 (18 August 2010).
31. Wong, D.J. *et al.* Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333–344 (2008).
32. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* **36**, 1090–1098 (2004).
33. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
34. Whitfield, M.L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977–2000 (2002).
35. Wei, C.L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
36. Geisler, S. *et al.* Influence of *TP53* gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. *Cancer Res.* **61**, 2505–2512 (2001).
37. Olive, K.P. *et al.* Mutant *p53* gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
38. Agarwal, M.L., Agarwal, A., Taylor, W.R. & Stark, G.R. *p53* controls both the G2/M and the G1 cell cycle checkpoints and mediates reversible growth arrest in human fibroblasts. *Proc. Natl. Acad. Sci. USA* **92**, 8493–8497 (1995).
39. Di Leonardo, A., Linke, S.P., Clarkin, K. & Wahl, G.M. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. *Genes Dev.* **8**, 2540–2551 (1994).
40. Morachis, J.M., Murawsky, C.M. & Emerson, B.M. Regulation of the p53 transcriptional response by structurally diverse core promoters. *Genes Dev.* **24**, 135–147 (2010).
41. Michlewski, G. & Caceres, J.F. RNase-assisted RNA chromatography. *RNA* **16**, 1673–1678 (2010).
42. Khalil, A.M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
43. Tsai, M.C. *et al.* Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
44. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
45. Han, J., Kim, D. & Morris, K.V. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci. USA* **104**, 12422–12427 (2007).
46. Schwartz, J.C. *et al.* Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.* **15**, 842–848 (2008).
47. Mattick, J.S. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**, 930–939 (2003).

ONLINE METHODS

Tiling array design and RNA hybridization. A custom tiling array (Roche NimbleGen) was designed at 5-bp resolution across 25 kb of the 9p21 region (which encompasses *CDKN2A*, *P14ARF* and *CDKN2B*), as well as from 10 kb upstream to 2 kb downstream of each TSS from 53 other cell-cycle genes, including those encoding cyclins, CDKs and CDKIs (**Supplementary Table 1**). In addition, the *HOXA* and *HOXD* loci were placed on the array as a control. Briefly, RNA was amplified (MessageAmp Kit, Ambion), reverse transcribed (RETROscript Kit, Ambion), labeled and hybridized according to the standard NimbleGen protocol.

Peak calling. Robust multichip average normalized single channel data from each array were subjected to peak calling using the NimbleScan program (Roche NimbleGen) with a window size of 50. Peaks with a peak score greater than ten were considered significant transcriptional units. Peak calls from all 55 array samples were clustered using Galaxy[2,48], and only transcripts present in a minimum of 10% of the samples were considered for further analysis. Transcripts were annotated as follows: 'genomic location (upstream of TSS of cell-cycle protein-coding gene, upst; exon of cell-cycle protein–coding gene, exon; intron of cell-cycle protein–coding gene, int; downstream of cell-cycle protein coding gene, dst)'; 'gene symbol of nearest mRNA'; 'distance from TSS'.

Measuring protein-coding potential. To assess the coding potential of the new transcribed regions, we evaluated the evolutionary signatures in their alignments with orthologous regions in 20 other sequenced placental mammalian genomes using the codon substitution frequencies (CSF) method[30,49,50], which has also been applied to assess new transcribed regions in mouse[14]. CSF produces a score for any region in the genome considering all codon substitutions observed within its alignment, based on the relative frequency of similar substitutions in known coding and noncoding regions. Briefly, CSF performs a statistical comparison between two empirical codon models[51], one estimated from alignments of known coding regions and the other based on noncoding regions, and reports a likelihood ratio that quantifies whether the protein-coding model is a better explanation while controlling for the overall level of sequence conservation[30].

Module map analysis. We generated a module map of the ncRNAs versus the protein-coding genes by computing the Pearson correlations for all pairwise combinations based on expression across 17 different samples. This map was clustered and visualized using the program Genomica (see URLs). For each ncRNA, we then defined gene sets of the protein-coding genes that had a Pearson correlation that was greater than or less than 0.5 with that ncRNA. To determine functional associations, we then generated a module map of these ncRNA gene sets with Gene Ontology Biological Processes gene sets (**Fig. 3c**) and with curated gene sets of metabolic and signaling pathways and biological and clinical states from the Molecular Signatures Database (MSigDB c2 collection) (**Supplementary Fig. 4**)[33]. The *P* value of enrichment was determined by the hypergeometric distribution, and a false discovery rate (FDR) calculation was used to account for multiple hypothesis testing ($P < 0.05$, FDR < 0.05).

Tissue samples and cells. Informed consent was obtained for tissue donation, and we obtained approval from institutional review boards of Stanford University, Johns Hopkins University and Netherlands Cancer Institute. Human primary breast tumors from The Netherlands Cancer Institute[52] and normal breast tissues and metastatic breast tumors from the Johns Hopkins University Rapid Autopsy Program are as described[23]. Human fetal pancreata were obtained from the Birth Defects Research Laboratory, University of Washington. Staged fetal pancreata were processed within 24 h of receipt, minced, washed and processed for RNA isolation using standard methods. Human fetal lung fibroblasts FL3 (Coriell AG04393) or foreskin fibroblasts (ATCC CRL2091) were cultured in 10% FBS (Hyclone) and 1% penicillin-streptomycin (Gibco) at 37 °C in 5% CO2.

***PANDA* cloning and sequence analysis.** 3′ and 5′ RACE was performed using the FirstChoice RLM-RACE Kit (Ambion). RNA was extracted from 200 ng/ml doxorubicin (Sigma)-treated human fetal lung fibroblasts, polyA was selected using the Poly(A)Purist MAG kit (Ambion) and RLM-RACE was performed according to the standard manufacturer's protocol.

RT-PCR. Total RNA was extracted from cells using the TRIzol reagent (Invitrogen) and the RNeasy Mini Kit (Qiagen), and genomic DNA was eliminated using TURBO DNA-free (Ambion). RT-PCR using 50–250 ng of total RNA was performed using the One-Step RT-PCR Master Mix (Applied Biosystems) using TaqMan Gene Expression Assays and normalized to GAPDH. Strand-specific RT-PCR for *PANDA* was performed using the One-Step RT-PCR Master Mix SYBR Green (Stratagene)).

TaqMan custom ncRNA assays. A panel of TaqMan custom ncRNA assays was developed targeting 60 of the 219 new transcribed regions using the 'single-exon' design mode. The transcript specificity and genome specificity of all TaqMan assays were verified using a position-specific alignment matrix to predict potential cross reactivity between designed assays and genome-wide nontarget transcripts or genomic sequences. For gene expression profiling of these ncRNAs across different conditions, complementary DNAs (cDNA) were generated from 50 ng of total RNA using the High Capacity cDNA Reverse Transcription Kit (Life Technologies). The resulting cDNA was subjected to a 14-cycle PCR amplification followed by real-time PCR reaction using the manufacturer's TaqMan PreAmp Master Mix Kit Protocol (Life Technologies). Two replicates were run for each gene for each sample in a 384-well format plate on the 7900HT Fast Real-Time PCR System (Life Technologies). PPIA was used as an endogenous control for normalization across different samples.

RNA blot. We obtained 5 µg of polyA RNA using an RNeasy Kit (QIAGEN) and PolyA Purist Mag (Ambion). RNA blot was performed using NorthernMax Kit (Ambion) following the standard manufacturer's protocol. Probes were generated with full length *PANDA* using the Prime-It RmT Random Primer Labeling Kit (Agilent).

Antibodies. The following antibodies were used for chromatin immunoprecipitation assays: anti-H3K4me3 (Abcam ab8580), anti-H3K35me3 (Abcam ab9050) and anti-p53 (Abcam ab28). Protein blots were performed using anti-PARP (Cell Signal 9542), anti-B-tubulin (Abcam ab6046), anti LSD1 (ab17721), anti EZH2 (Cell Signal AC22), anti p21 (Santa Cruz Biotech) and anti NF-YA (Santa Cruz Biotech H-209).

RNA interference. Human fetal lung fibroblasts were transfected with 50 nM of ON-TARGETPlus siRNAs (Dharmacon) targeting *PANDA* (**Supplementary Table 5**). Validated siRNAs for mRNAs were obtained from Ambion (**Supplementary Table 5**).

TUNEL. TUNEL assays were performed using the *in situ* Cell Death Detection Kit, TMR Red (Roche). Human fetal lung fibroblasts were cultured on chamber slides (Lab-Tek), treated with 200 ng/ml doxorubicin (Sigma) for 24 h, fixed with methanol at −20 °C for 10 min and incubated with the TUNEL labeling mixture for 1 hour at 37 °C. Slides were then washed with PBS and mounted in Prolong Gold antifade reagent with DAPI (Invitrogen) and imaged at 20× magnification.

RNA immunoprecipitation. Ten million cells were treated with 200 ng/ml doxorubicin for 16 h, trypsinized and crosslinked with 1% formaldehyde for 10 min, followed by the addition of 0.125 M glycine for 5 min. After two PBS washes, cells were lysed with 2× volume of Buffer A (10 mM HEPES pH 7.5, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 1 mM PMSF) for 15 min on ice at 150 r.p.m. NP-40 was added to a final concentration of 0.25% for 5 min on ice. Lysates were centrifuged for 3 min at 2,000 r.p.m., and the supernatant (cytosol) was collected. Next, an equal volume of Buffer C as was used of Buffer A was added to the pellet for 20 min with frequent vortex (20 mM HEPES pH 7.5, 10% glycerol, 0.42 M KCl, 4 mM MgCl2, 0.5 mM DTT, 1 mM PMSF). Nuclear lysates were dounced for 5 s using a motorized pestle and sonicated for 7 min using a Diagenode Sonicator (30 s on, 30 s off, power setting H). Nuclear and cytoplasmic lysates were combined and centrifuged for 15 min at 13,000 r.p.m. Supernatants were transferred into micro spin columns (Pierce 89879), and 2 µg of antibody was added and incubated overnight. We washed 10 µl of Protein A/G UltraLink Resin (Pierce 53132)

doi:10.1038/ng848

three times with RIP wash buffer (50 mM TrisHcl pH 7.9, 10% glycerol, 100 mM KCl, 5 mM $MgCl_2$, 10 mM B-me and 0.1% NP-40) and added it to the immunoprecipitation reaction for 1 h at 4 °C. Samples were washed four times with RIP wash buffer and two times with 1 M RIPA (50 mM Tris pH 7.4, 1 M NaCl, 1 mM EDTA, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 0.5 mM DTT and 1 mM PMSF). Beads were resuspended in 200 µl 150 mM RIPA (50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate, 0.5 mM DTT and 1 mM PMSF) plus 5 µl Proteinase K (Ambion) and incubated for 1 h at 45 °C. We added 1 ml of TRIzol to the sample, and RNA was extracted using the RNEasy Mini Kit (QIAGEN) with the on column DNAse digest (QIAGEN).

RNAse mediated RNA chromatography. RNAse mediated RNA chromatography[41] was performed as previously described with the following modifications: 6 pmols of RNA (*PANDA* or a 1.2-kb fragment of *LacZ*) were used per reaction. RNA was folded (90 °C for 2 min, ice for 2 min), supplied with RNA structure buffer (Ambion) and shifted to room temperature (22–25 °C) for 20 min before conjugation to beads. RNAse digestion was performed with 5 µl of RNase A/T1 cocktail (Ambion) and 2 µl of RNase V1 (Ambion).

Cellular lysates were prepared as follows: 10 million doxorubicin treated cells (16 h) were incubated in 200 µl PBS, 600 µl H20 and 200 µl nuclear lysis buffer (1.28 M sucrose; 40 mM Tris-HCl pH 7.5; 20 mM MgCl2; 4% Triton X-100) on ice for 20 min. Nuclei were pelleted by centrifugation at 2,500*g* for 15 min. The nuclear pellet was resuspended in 1 ml RIP buffer (150 mM KCl, 25 mM Tris pH 7.4, 0.5 mM DTT, 0.5% NP40, 1 mM PMSF and protease inhibitor (Roche Complete Protease Inhibitor Cocktail Tablets)). Resuspended nuclei were sheared using a motorized douncer for 5 s. Nuclear membrane and debris were pelleted by centrifugation at 18,000*g* for 10 min.

Chromatin immunoprecipitation (ChIP). ChIP was performed as previously described[53]. qPCR primers for FAS and CCNB1 and FAS-control NF-YA binding sites were obtained from Morachis *et al.*[40] Primers for PUMA and BAX were designed to surround the NF-YA consensus motif CCAAT (**Supplementary Table 5**).

48. Taylor, J., Schenck, I., Blankenberg, D. & Nekrutenko, A. Using galaxy to perform large-scale interactive data analyses. *Curr. Protoc. Bioinformatics* Chapter 10, Unit 10.5 (2007).
49. Lin, M.F. *et al.* Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Res.* **17**, 1823–1836 (2007).
50. Lin, M.F., Deoras, A.N., Rasmussen, M.D. & Kellis, M. Performance and scalability of discriminative metrics for comparative gene identification in 12 *Drosophila* genomes. *PLOS Comput. Biol.* **4**, e1000067 (2008).
51. Kosiol, C., Holmes, I. & Goldman, N. An empirical codon model for protein sequence evolution. *Mol. Biol. Evol.* **24**, 1464–1479 (2007).
52. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
53. Rinn, J.L., Bondre, C., Gladstone, H.B., Brown, P.O. & Chang, H.Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2**, e119 (2006).