TargetGeneReg 2.0: a comprehensive web-atlas for p53, p63, and cell cycle-dependent gene regulation

Martin Fischer [1,*], **Robert Schwarz** [1], **Konstantin Riege**[1], **James A. DeCaprio** [2,3] **and Steve Hoffmann**[1]

[1]Computational Biology Group, Leibniz Institute on Aging – Fritz Lipmann Institute (FLI), Beutenbergstraße 11, 07745 Jena, Germany, [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA and [3]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

ABSTRACT

In recent years, our web-atlas at www.TargetGeneReg.org has enabled many researchers to uncover new biological insights and to identify novel regulatory mechanisms that affect p53 and the cell cycle – signaling pathways that are frequently dysregulated in diseases like cancer. Here, we provide a substantial upgrade of the database that comprises an extension to include non-coding genes and the transcription factors ΔNp63 and RFX7. TargetGeneReg 2.0 combines gene expression profiling and transcription factor DNA binding data to determine, for each gene, the response to p53, ΔNp63, and cell cycle signaling. It can be used to dissect common, cell type and treatment-specific effects, identify the most promising candidates, and validate findings. We demonstrate the increased power and more intuitive layout of the resource using realistic examples.

GRAPHICAL ABSTRACT

INTRODUCTION

The cell proliferation cycle and the tumor suppressor p53 are closely linked and represent the most commonly dysregulated signaling pathways in cancer. Despite more than 40 years of research on p53 and many more on the cell cycle, we still lack a comprehensive understanding of the p53 and the cell cycle-dependent regulation of a surprisingly large number of genes. Several mechanisms have been proposed to explain the temporal regulation of hundreds of cell cycle genes (1,2) and the downstream targets of p53 (3–5), but the substantial overlap between p53 and the cell cycle render the analysis of individual genes difficult.

The expansion of publicly available high-throughput datasets has enabled a more detailed understanding of gene regulatory mechanisms and networks in recent years. We developed a meta-analysis approach to cross-validate results and to improve statistical power by integrating datasets derived from different experimental setups (6). The meta-analysis allows inferring p53 and cell cycle regulation of genes from multiple cell types and treatment conditions and derive common signature genes. It follows the intuitive idea that when multiple independent datasets agree on a finding it is more likely to be accurate and that the sum of available evidence provides the best guess for the truth. Previously, we employed this meta-analysis approach to chart the transcriptional programs of the cell cycle, human and mouse p53, the viral oncoprotein E7, and the transcription factor ΔNp63 (6–10). Key findings from these meta-analyses included that p53 serves as a transcriptional activator, while genes repressed by p53 were actually cell cycle-dependent genes (6). Specifically, while p53 up-regulates hundreds of genes directly through engaging with chromatin in physical contact with the gene locus, it also down-regulates the large group of cell cycle genes indirectly through its target gene *CDKN1A*. *CDKN1A* encodes for the cyclin-dependent kinase (CDK) inhibitor p21 that suppresses cyclin:CDK activity leading to the activation of the cell cycle *trans*-repressor complexes DREAM (DP, RB-like, E2F4 and MuvB) and RB:E2F (11–17). More-

---
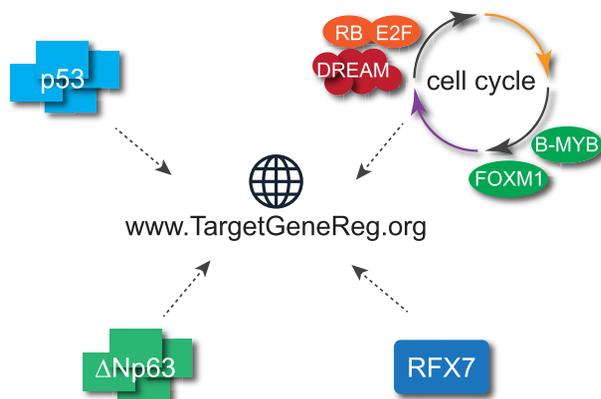
*To whom correspondence should be addressed. Tel: +49 3641656876; Fax: +49 3641656255; Email: martin.fischer@leibniz-fli.de

over, our meta-analyses revealed that the transcription factor complexes RB:E2F, DREAM, and MMB:FOXM1 controlled essentially all of the cell cycle genes. The analysis revealed a small number of genes that were specifically activated by p53 and controlled within the cell cycle by RB:E2F and DREAM (6). However, transcriptome analyses suggest that larger subnetworks of the p53 and cell cycle-dependent gene regulation networks (GRN) are yet to be understood (5,6).

The target gene regulation (TargetGeneReg) database developed from the meta-analyses was made available through a web-based atlas at www.TargetGeneReg.org (6) to enable researchers to easily scrutinize the influence of the cell cycle and p53 on any gene of interest. Through www.TargetGeneReg.org, researchers can rapidly determine common as well as treatment, cell type, and species-specific regulations, identify promising targets, and validate findings. In numerous research projects, it is necessary to establish the degree to which a given gene is directly or indirectly affected by p53 and major cell cycle signaling pathways. Its easy-to-use interface enables researchers to quickly gather evidence about the extent and frequency their genes of interest are affected by these critical regulators. TargetGeneReg has been used for understanding cell cycle regulators, their signaling cues, and their disease relevance (18–26). Moreover, our database has helped to identify pathways that respond to drugs and stress conditions (27,28), among many other applications.

Alternative resources such as the p53 BAER hub and the Cyclebase v3.0 either focus on p53-dependent regulation or the influence of the cell cycle on gene expression, respectively (29,30). However, the integration of both layers of information necessary to understand p53's contributions to target gene regulation is not readily possible using these tools. Likewise, it is of interest for many researchers to study the target gene expression in other species such as *Mus musculus*. While Cyclebase v3.0 includes cell cycle-dependent gene regulation data from other species, the p53 BAER hub does not provide information beyond *Homo sapiens*.

The TargetGeneReg resource enabled us to compare the p53 GRN between mouse and human. Surprisingly, up-regulation by p53 displayed substantial evolutionary divergence, while down-regulation of cell cycle genes by p53–p21 is well conserved (9,31). Moreover, we employed the resource to compare the GRN of p53 to its sibling ΔNp63 and, in contrast to previous reports, we demonstrated that ΔNp63 minimally affects any direct p53 target. Instead, a large number of ΔNp63 targets were cell cycle genes, but the mechanistic link between ΔNp63 and the cell cycle remained unclear (10,32). Most recently, TargetGeneReg enabled the discovery of the transcription factor RFX7, an emerging tumor suppressor, as a novel node in the p53 GRN, proposing a mechanism for how p53 regulates several targets (33). RFX7 is linked to multiple lymphoid cancers (34), such as Burkitt lymphoma where we and others identified RFX7 as a potential cancer driver (35,36).

Here, we provide a major update for TargetGeneReg through a substantial expansion of the underlying data resources to include recent RNA-seq and ChIP-seq datasets on p53 and cell cycle regulation, inclusion of data resources on ΔNp63 and RFX7, an upgrade of the website, and vi-

sualizations of expanded ChIP-seq data through the UCSC Genome Browser.

MATERIALS AND METHODS

RNA-seq analysis pipeline

We used Trimmomatic (37) v0.39 (5nt sliding window approach, 5′ leading and mean quality cutoff 20) for read quality trimming according to inspections made from FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) v0.11.9 reports. Illumina adapters as well as mono- and di-nucleotide content were clipped using Cutadapt v2.10 (38). Potential sequencing errors were detected and corrected using Rcorrector v1.0.4 (39). Ribosomal RNA (rRNA) transcripts were artificially depleted by read alignment against rRNA databases through SortMeRNA v2.1 (40). The preprocessed data was aligned to the reference genome hg38, retrieved along with its gene annotation from Ensembl v102 (41), using the mapping software segemehl (42,43) v0.3.4 with adjusted accuracy (95%) and split-read option enabled. Mappings were filtered by Samtools v1.12 (44) for uniqueness and, in case of paired-end data, properly aligned mate pairs. Differential gene expression and its statistical significance was identified using DESeq2 v1.30.0 (45). Common thresholds of $|\log_2\text{fold-change}| \geq 0.25$ and FDR <0.05 were used to identify significantly differentially expressed genes.

Microarray analysis pipeline

All microarray datasets were available at a pre-processed stage at the Gene Expression Omnibus (GEO) and we re-analyzed the datasets with GEO2R to obtain fold expression changes and Benjamini Hochberg-corrected *P*-values (FDR) (46). Gene identifiers were mapped to Ensembl Gene IDs using the Ensembl annotation data v102 (41). Common thresholds of $|\log2\text{fold-change}| \geq 0.25$ and FDR <0.05 were used to identify significantly differentially expressed genes.

Meta-analysis / generation of Expression Scores

Following our meta-analysis approach (6), *Expression Scores* for genes regulated by human and mouse p53 and ΔNp63 were calculated as the number of datasets that find the gene to be significantly up-regulated minus the number of datasets that find the gene to be significantly down-regulated by the respective transcription factor. Both, the *mouse p53* and the *ΔNp63 Expression Score* were published previously (9,10). The *p53 Expression Score 2.0* (human) contains two additional quality control measures. First, only datasets derived from at least two biological replicates have been considered for this updated score. Second, all datasets were removed that failed to identify at least 50 out of 116 direct p53 target genes that were most recurrently identified in a previous meta-analysis (3). The *Cell Cycle Expression Score* reflects the number of datasets that identified a gene as cell cycle-regulated gene. The *Cell Cycle Gene Category* is calculated by a majority vote of the nine datasets on cell cycle-dependent gene expression and is displayed for each gene that shows a *Cell Cycle Expression Score* ≥ 3. Precisely, each dataset that identified peak expression of

Table 1. Comparison of TargetGeneReg v2.0 features and resource properties to TargetGeneReg v1.0/1.1 (6,9), p53 BAER hub (26), and Cyclebase v3.0 (27). Numbers concern data from human except for when 'mouse' is indicated. While Cyclebase v3.0 contains multiple datasets on cell cycle-dependent gene regulation from various species, it contains only one dataset from human. ChIP-seq replicates were combined to single datasets for TargetGeneReg but kept separate for the p53 BAER hub. *NA* – not available, information was not provided in the respective publications

	TargetGeneReg v2.0	TargetGeneReg v1.0/1.1	p53 BAER hub	Cyclebase v3.0
# human genes	**37243**	18845	*NA*	*NA*
human genome version	**hg38**	hg19	hg19	*NA*
# p53 expression datasets	**57**	20	16	-
# p53 ChIP-seq tracks/datasets	32 / 28	15	**41**	-
p53RE prediction	yes	no	yes	-
# mouse p53 expression datasets	**15**	**15**	-	-
# mouse p53 ChIP-seq datasets	**9**	**9**	-	-
# cell cycle expression datasets	**9**	5	-	1
# DREAM ChIP-seq datasets	**17**	9	-	-
# RB ChIP-seq datasets	**6**	2	-	-
# MMB-FOXM1 ChIP-seq datasets	**22**	6	-	-
CHR and E2F motif predictions	yes	yes	-	-
# ΔNp63 expression datasets	**16**	-	-	-
# ΔNp63 ChIP-seq datasets	**20**	-	-	-
p63RE prediction	yes	-	-	-
RFX7 target gene prediction	yes	-	-	-
Genome browser visualizations	**yes**	no	**yes**	no

the gene in 'G1', 'G1/S', or 'S-phase' is grouped as 'G1/S', and peak expression in 'G2', 'G2/M', 'M', and 'M/G1' is grouped as 'G2/M'.

ChIP-seq data integration

Peak datasets and bigwigs from ChIP-seq experiments were retrieved from CistromeDB (47) ensuring a common data processing pipeline and thereby a direct comparability. Only RFX7 ChIP-seq data were taken from our recent study (33) as they were not yet available through CistromeDB. Bigwigs (ChIP-seq tracks) have been made available through track hubs for the UCSC Genome Browser (48). Notably, while ChIP-seq replicates are available as individual tracks in the track hubs, they have been jointly considered as one dataset for the generation of peak-of-peaks summaries. Precisely, when replicate experiments were available, all peaks were used that have been identified in at least two replicates. To identify overlapping and non-overlapping peaks, Bedtools 'intersect' was employed, and to generate the peak-of-peaks summaries, multiple peak files were combined using Bedtools 'multiinter' (49). The p53 and p63 ChIP-seq collections and summaries on human and mouse p53 and ΔNp63 have been published previously (9,10). Similarly, p53REs and p63REs were taken from our previous study (10).

RESULTS

Similar to TargetGeneReg v1.0 (6), TargetGeneReg v2.0 focuses on gene regulation by the tumor suppressor p53 in conjunction with the human cell cycle. An earlier upgrade (TargetGeneReg v1.1) introduced p53-dependent gene regulation and p53 binding data from mouse to TargetGeneReg (9), which, to our knowledge, is unique to the TargetGeneReg resources. Similarly, transcription factor binding data for central transcriptional cell cycle regulators, including the DREAM complex, RB, the MMB complex, and FOXM1, is unique to TargetGeneReg. The upgrade to version 2.0 not only expands the data, but also includes data

on p53's oncogenic sibling ΔNp63 and the emerging tumor suppressor RFX7 (Table 1). In the following sections we provide more detailed information on the data and the applicability of the upgraded TargetGeneReg resource.

Gene regulation by p53

In the first version of TargetGeneReg, we integrated 20 datasets on p53-dependent gene regulation (6). Since the publication of TargetGeneReg v1.0, several additional high-throughput datasets with varying resolution and experimental strategies became available. To optimally use this additional data and strengthen the power of our resource, we have adjusted our quality control regiment. Specifically, we systematically searched the GEO database for RNA-seq and microarray datasets that employed experimental strategies known to affect p53 signaling. This search included experiments involving MDM2 inhibitors (Nutlin and RG7388), genotoxic and nucleolar stress inducers (Doxorubicin, 5-FU, Actinomycin D, Daunorubicin, Etoposide, Bleomycin, Camptothecin, and UV), viral oncoproteins (SV40 LT, HPV16 E6, and HPV18 E6), exogenous TP53 expression, TNFα, and senescence (oncogene and replication-induced). For inclusion in the updated resource, we required all datasets to comprise at least two biological replicates for both treatment and control conditions. Notably, all datasets we obtained were derived from cell line models. We integrated information from 64 RNA-seq and 35 microarray datasets derived to identify significantly differentially expressed genes. To verify the effects of the selected experiments on known p53-regulated genes, we used a benchmark dataset of 116 direct and highly responsive p53 target genes identified earlier based on 16 genome-wide analyses (3). All experiments that yielded <50 significantly differentially expressed benchmark targets were removed from further analysis (see Materials and Methods). This measure ensures focus on activation of the p53 pathway by the experimental setup and sufficient power to identify specific p53 activities (Figure 1A). A total of 44 RNA-
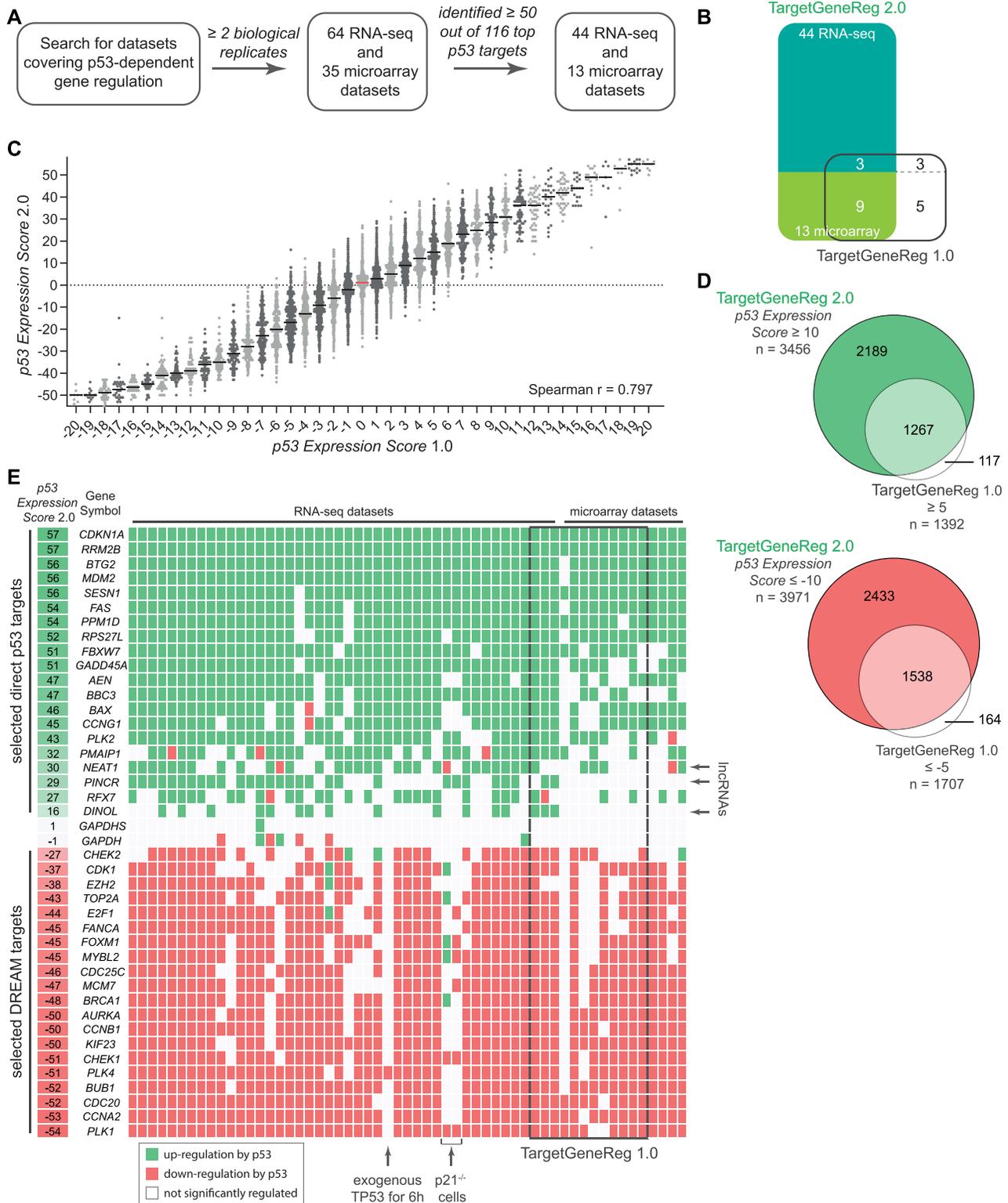
Figure 1. (**A**) Flow chart for the integration of datasets on p53-dependent gene regulation. (**B**) Datasets on p53-dependent gene regulation in TargetGeneReg v2.0 compared to TargetGeneReg v1.0. (**C**) The *p53 Expression Score* v2.0 from TargetGeneReg v2.0 compared to the *p53 Expression Score* v2.0 from TargetGeneReg v1.0 for 17 446 genes present in both databases. Genes are displayed by individual points. The median is indicated by a black line or a red line to highlight '0'. (**D**) Genes passing the recommended p53 Expression Score threshold to be considered high-recurrence genes that are up- or down-regulated by p53 in TargetGeneReg v2.0 compared to TargetGeneReg v1.0. (**E**) The p53 Expression Score v2.0 and data from the underlying 57 individual datasets visualized for 20 selected direct p53 target genes, 20 selected targets of the DREAM complex, and the non-regulated *GAPDH* genes. It is indicated whether individual datasets were generated using RNA-seq or microarray. Individual datasets that were also present in TargetGeneReg v1.0 are indicated. The lncRNAs *NEAT1, PINCR,* and *DINOL* are highlighted as they were not available in TargetGeneReg v1.0. Details on three datasets in which most DREAM targets are not down-regulated by p53 are shown.

seq and 13 microarray datasets passed this control (Figure 1B).

The *p53 Expression Score*, based on these 57 datasets, was calculated for each gene by the number of datasets yielding significant up-regulation minus the number of datasets with significant down-regulation of the gene by p53. To calculate the score, we required a gene to be sufficiently expressed in at least three datasets. A gene was deemed to be expressed when DESeq2 was able to include it in the differential expression analysis, i.e. assign $\log_2$fold-change and FDR values. A direct comparison of the updated score (*p53 Expression Score* v2.0) with the initial one (*p53 Expression Score* v1.0) exhibits a strong correlation but also suggests an improved resolution (Figure 1C). Most importantly, while the previous version was limited to 18 845 protein-coding genes from hg19, the updated resource now provides a *p53 Expression Score* for 37 243 genes from hg38. The *p53 Expression Score* v1.0 had a minimum threshold of ≥ 5 and ≤ -5 to consider genes with high confidence as being up and down-regulated by p53, respectively (6). In the case of the updated *p53 Expression Score* v2.0, respective thresholds of ≥ 10 and ≤ -10 were passed by 3456 and 3971 genes (Figure 1D). To illustrate the advantage of the updated *p53 Expression Score*, we visualized it together with the underlying individual datasets for 20 selected direct p53 target genes and 20 selected targets of the DREAM complex (Figure 1E).

Of note, some datasets show few if any down-regulated DREAM targets. This is the case when p21 (CDKN1A) negative cells were used, since they were unable to reactivate the DREAM complex efficiently. Likewise, an experiment where exogenous TP53 was induced for only 6 h, an interval too brief to reactivate the cell cycle *trans*-repressor complexes, did not lead to a down-regulation of critical cell cycle genes. Moreover, our comparison identifies p53-dependent lncRNAs such as *DINOL* (50), *PINCR* (51) and *NEAT1* (52,53) excluded from the previous version because of insufficient data (Figure 1E). In addition to information on the p53-dependent regulation of thousands of noncoding RNAs, the updated *p53 Expression Score* v2.0 provides much more detailed information on p53-dependent regulation for hundreds of genes for which the *p53 Expression Score* v1.0 was inconclusive. Consequently, 1674 and 1917 genes that previously displayed a *p53 Expression Score* v1.0 between 5 and –5 now passed the threshold of a *p53 Expression Score* v2.0 of ≥ 10 and ≤ -10, respectively, indicating a differential regulation by p53 with high recurrence.

The direct binding of p53 to the gene promoter is a crucial property of many genes up-regulated by this transcription factor (3–5). While TargetGeneReg v1.0 integrated 15 datasets on p53 genome binding (6), the updated version now integrates an expanded collection of 28 ChIP-seq datasets. While the previous version displayed only the number of datasets that identified p53 binding near a gene's TSS, the updated resource contains precise binding location information and visualizations thereof. To enable users to rapidly visualize the large number of 28 individual p53 ChIP-seq datasets, we provide a 'peak-of-peaks' data track representing a pile-up of p53 peak regions from individual datasets (Figure 2A). Therefore, the 'peak-of-peaks' track provides quick summary information on how many datasets identified p53 binding to any locus in the genome. In ad-

dition, the p53 response element (p53RE) most closely resembling a canonical p53RE is displayed for each peak-of-peaks, as described previously (10). In addition, the ChIP-seq tracks from all individual datasets can be displayed upon selection, providing a seamless visualization of cell type and treatment-specific information next to the summary data (Figure 2A and B). As established previously, any binding site with support from at least five datasets is considered to be of high recurrence (9,10). The website's 'Overview' section indicates for every gene whether it displays a high-recurrence binding site within 2.5 kb of a TSS, and whether a high-recurrence binding site is linked to the gene locus through a double-elite enhancer:gene association listed in the GeneHancer database (54).

Together, TargetGeneReg v2.0 provides information on p53-dependent gene regulation for twice as many genes from almost three times as many datasets in total and almost five times as many datasets that follow the tightened control measures. In addition, it provides almost twice as many p53 ChIP-seq datasets, predictions for the underlying p53RE, and precise location visualizations.

Cell cycle-dependent gene regulation

Cell cycle genes play essential roles in cell cycle progression and therefore are typical markers of proliferation that are dysregulated in many cancers (55). The tumor suppressor p53 down-regulates cell cycle genes to sustain cell cycle arrest. Based on TargetGeneReg v1.0, we consolidated the five cell cycle gene peak clusters defined by Whitfield *et al.* (56) to two major groups of cell cycle genes, namely G1/S and G2/M genes (6). Here, we expanded the previous resource's five datasets to include four additional datasets (Figure 3A). For all genes identified as cell cycle-dependently regulated in at least three of the nine datasets, we predicted whether the gene is a G1/S or a G2/M gene based on a majority vote by the nine datasets (see Materials and Methods). The website's 'Overview' section provides information on the number of datasets that suggest a gene to be driven by the cell cycle ('Cell Cycle Expression Score') and its classification prediction ('Cell Cycle Gene Category').

The two distinct groups of G1/S and G2/M genes are primarily characterized by E2F and CHR (cell cycle genes homology region) DNA recognition motifs in their promoters, respectively (2,6,57). The DREAM complex can bind to both E2F and CHR motifs through its respective subunits E2F4 and LIN54, while RB:E2F specifically binds G1/S cell cycle genes through E2F motifs. In contrast, the transcription factors B-MYB (also known as MYBL2) and FOXM1 associate with DREAM's LIN54-containing MuvB core complex later in the cell cycle to specifically activate the expression of G2/M genes through binding their CHR sequences (2). To allow a more comprehensive analysis of cell cycle-dependent regulation, we expanded the nine datasets on genome binding by DREAM complex components to 17. Similarly, we extended the two previous datasets on RB binding to six datasets, and the previous six datasets on MMB:FOXM1 (B-MYB:MuvB:FOXM1) binding to 22 datasets. Potential E2F and CHR motifs under respective RB and MMB:FOXM1 binding sites have been predicted using HOMER (58). The individual ChIP-
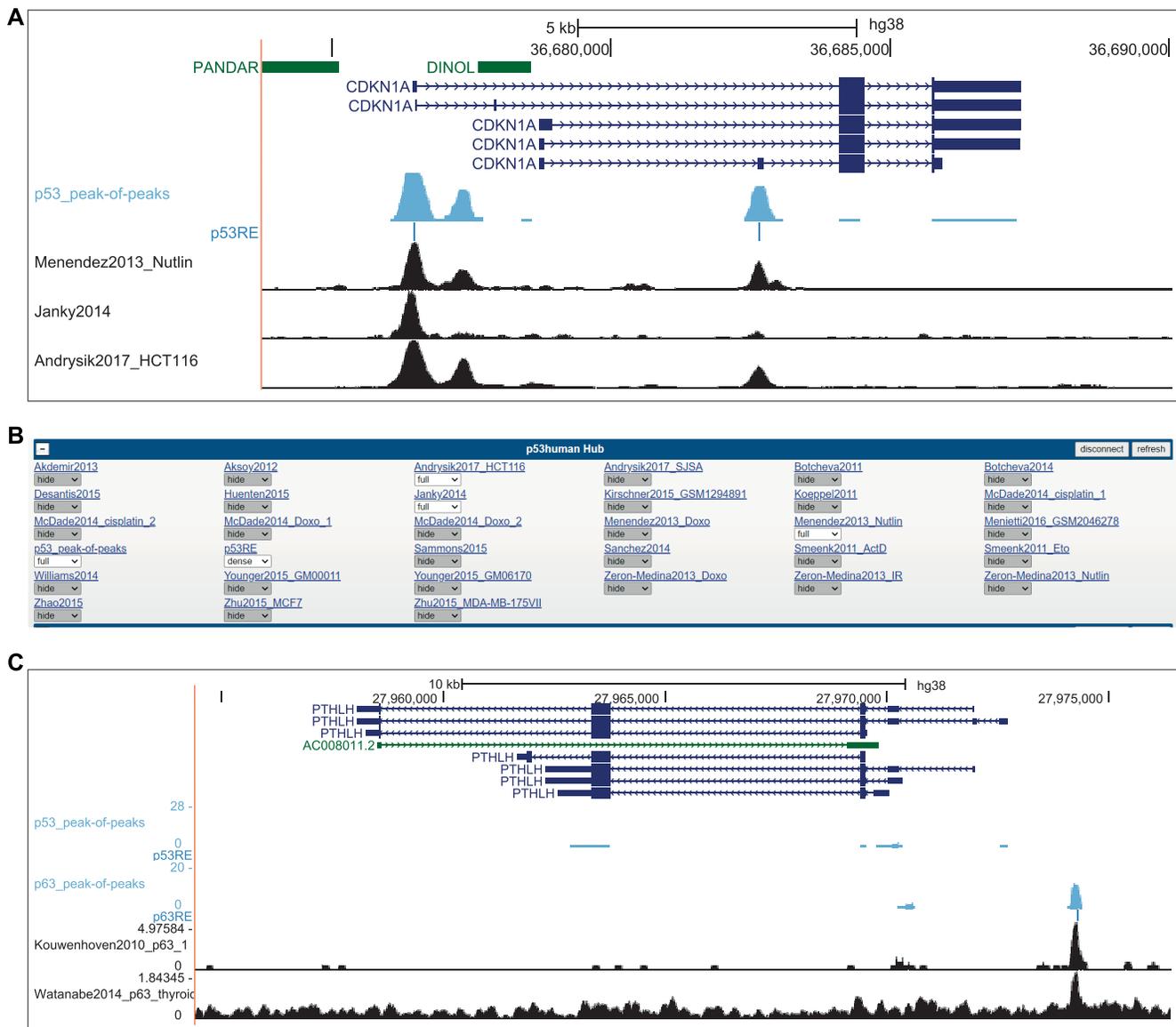
Figure 2. (**A**) Image of UCSC Genome Browser displaying the *CDKN1A* locus linked from TargetGeneReg v2.0. The blue tracks display the p53_peak-of-peaks summary and the most likely underlying p53RE that has been identified. Individual p53 ChIP-seq tracks can be selected from (**B**) the track hub that is loaded through the TargetGeneReg v2.0 linkage. Together, the visualization provides precise visualization of p53 binding sites locations and their underlying p53RE and enables seamless comparisons between summary data and individual datasets. (**C**) Image of UCSC Genome Browser displaying the *PTHLH* locus linked from TargetGeneReg v2.0. The blue tracks display the p53_peak-of-peaks and p63_peak-of-peaks summaries and the most likely underlying p53RE and p63RE. Individual p63 ChIP-seq tracks can be selected from the track hub hat is loaded through the TargetGeneReg v2.0 linkage, as shown for p53 above. *PTHLH* is a direct target of ΔNp63 but not p53, and the unique p63 binding site can be readily seen by comparing the peak-of-peaks summary binding data.

seq tracks, peak-of-peaks, and motif predictions are available through UCSC Genome Browser visualizations.

Gene regulation by mouse p53 and its difference to human p53

Previously, we employed our meta-analysis approach on mouse p53 synthesizing p53-dependent gene regulation data across 15 datasets, and we made the data available through TargetGeneReg v1.1 (9). Here, we integrated our database on p53-dependent gene regulation in mice (mm10) with the updated database on p53-dependent gene regu-

lation in humans (hg38) described above. Therefore, TargetGeneReg v2.0 includes mouse p53-dependent gene regulation data for the one-to-one orthologs of 14 712 human genes. While there is a good correlation between the *mouse p53 Expression Score* and the *human p53 Expression Score* v2.0 for genes down-regulated by p53, the correlation for up-regulated genes is poor (Figure 3B), indicating a strong and a comparably low evolutionary conservation of p53 down and up-regulated genes, respectively. Similar results have already been reported for the first *p53 Expression Score* (9,31). Precise binding data (ChIP-seq) of mouse p53 is available through links to the UCSC Genome Browser
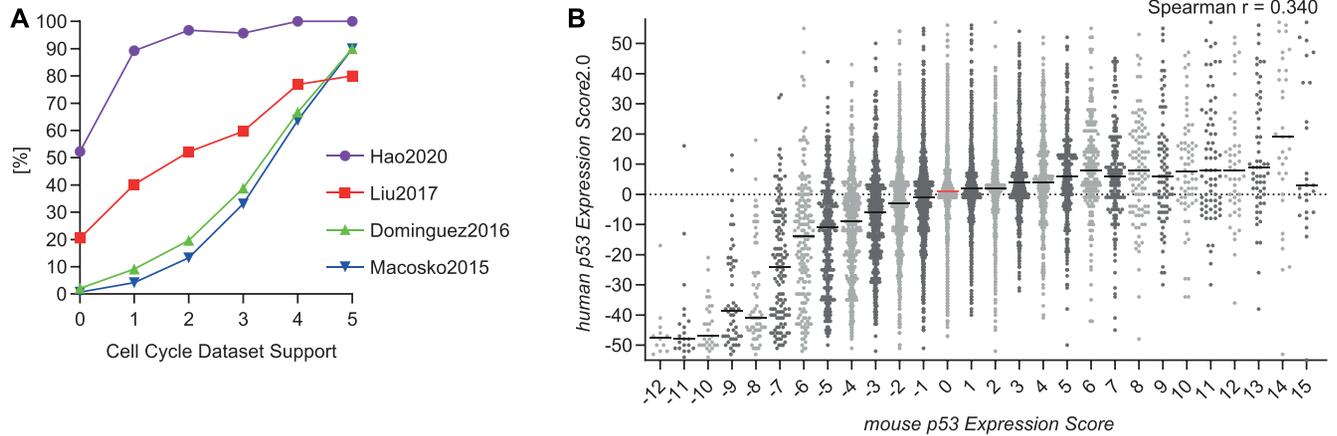
Figure 3. (**A**) Four new datasets on cell cycle-dependent gene expression have been added (63–66). Following our previous quality control (6), we tested whether the datasets were more likely to identify a gene as cell cycle gene when more datasets from TargetGeneReg v1.0 (X-axis) agreed on its cell cycle gene status. (**B**) The *mouse p53 Expression Score* (9) compared to the *human p53 Expression Score* v2.0 for 14 712 one-to-one orthologs with both scores. Genes are displayed by individual points. The median is indicated by a black line or a red line to highlight '0'.

with embedded track hubs similar to human protein binding data described above.

Gene regulation by ∆Np63

We previously employed our meta-analysis approach to provide a comprehensive resource for gene regulation by p53's sibling ∆Np63, an essential oncoprotein in squamous cell carcinomas (10). Given its relevance to cancer and close connection to p53, we integrated our ∆Np63 database comprising 16 datasets on p63-dependent gene regulation and 20 ChIP-seq datasets on p63 DNA binding into TargetGeneReg v2.0. The *∆Np63 Expression Score* and predictions of p63 targets (180 high-recurrence targets available in Table 1 from Riege *et al*. (10)) and potential p63 targets (comprising all genes bound and regulated by p63) are available through the website's 'Overview' section. DNA binding data and identified p63 response elements (p63RE) are available through UCSC Genome Browser track hubs and enable a seamless comparison between individual p63 ChIP-seq tracks, summary data thereof (p63_peak-of-peaks), and respective data from p53. *PTHLH*, for instance, is a direct target gene of ∆Np63 but not of p53 (Figure 2C).

Expanding the p53 gene regulatory network through RFX7

Complex cross-talk between signaling pathways impedes the identification of indirect gene regulatory mechanisms employed by p53. For example, following two decades of conflicting data on mechanisms of p53-dependent gene repression, p53 was found to serve as a transcriptional activator that represses genes indirectly, with its target p21 taking a predominant role through its profound influence on down-regulating the cell cycle genes (3,4,6,7,59). Recently, we identified the transcription factor and emerging tumor suppressor RFX7 as a vital node in the p53 transcriptional program. RFX7 orchestrates a subnetwork of tumor suppressor genes in response to cellular stress and p53 (33) and cooperates with p53 to inhibit the pro-survival kinases AKT

and mTORC1 (60). Given the crucial role of the novel p53-RFX7 signaling axis in the p53 gene regulatory network and its potential importance to cancer biology, we included the data on RFX7 target genes in TargetGeneReg 2.0. The website's 'Overview' section displays whether a gene has been predicted as an RFX7 target, offering a mechanistic explanation for its p53-dependent up-regulation. In addition, RFX7 ChIP-seq tracks are available through the UCSC Genome Browser visualizations.

Navigating the TargetGeneReg 2.0 website

Our updated resource is tailored to rapidly provide information on genes of interest entered in the main input field. The arguably strongest asset of the TargetGeneReg resource is summary data on gene regulation by various transcription factors generated through a synthesis of multiple individual datasets integrated by our meta-analysis approach. Therefore, the 'Overview' section situated at the top of the one-page website provides all summary information on the genes of interest that have been entered (Figure 4). Importantly, it enables a quick direct comparison of the summary data between multiple genes. More detailed information on gene regulation from the individual datasets, a pie chart illustration of the summary data, as well as volcano and box-plots of the dataset results are provided in the detailed sections below, which are equipped with sorting options to quickly identify the most relevant data. Precise transcription factor binding data visualized through the UCSC Genome Browser (48) are available through the genomic position links provided in the 'Overview' section for both human (hg38) and mouse (mm10).

DISCUSSION

The TargetGeneReg has gained a strong reputation in the p53 and cell cycle communities. It provides a deeper insight into p53 and cell cycle-dependent gene regulation mechanisms. The presented upgrade, TargetGeneReg v2.0, substantially improves this resource. Like its predecessor, the

① Start at this input field: Enter your genes of interest as gene symbols or ensembl IDs separated by commas.

② The one-page website contains several sections. Navigate through the navigation bar or by scrolling.

③ The additional sections provide more detailed information, such as data from individual datasets.

④ The overview section starts with general information on your genes of interest. Mind the genome position <u>link</u>!

⑤ Information on p53-dependent gene regulation are clustered together with the p53 network extender RFX7.

⑥ Information on cell cycle-dependent gene regulation are clustered together.

⑦ Information on ΔNp63-dependent gene regulation are clustered together.

⑧ Detailed information on mouse orthologs and gene regulation by mouse p53. Mind the genome position <u>link</u>!

Figure 4. The TargetGeneReg 2.0 website design. Start by entering your genes of interest as gene symbol or Ensembl gene ID. The overview section provides a helpful summary on the regulation of your genes of interest. Details on each point in the overview section is available through mouse-over boxes and through the 'About' section. The additional sections are available through the navigation bar in the upper left corner or by scrolling through this one-page website design. The additional sections contain more detailed information, such as data from the individual datasets and citation and history data for the resource.

resource is tailored to quickly retrieve information on the users' genes of interest and provides swift comparisons between genes and experimental conditions. TargetGeneReg v1.0 was a starting point to help the p53 and cell cycle communities to gain deep biological insights by providing reference points and a platform that enables users to quickly test whether their genes of interest are likely regulated by p53 or the cell cycle. The new version integrates more datasets, substantially improving its power. Specifically, TargetGeneReg now includes data on non-coding RNAs and provides information on the gene regulation by p53's oncogenic sibling ΔNp63 and the emerging tumor suppressor RFX7.

Visualization of the transcription factor binding data through the UCSC Genome Browser provides precise location information for the user to better interpret the potential consequences of the binding for their gene of interest. For example, p53 binding to intronic locations can induce alternative transcription start sites leading to transcript variants with shortened 5′ sequences, as reported for *MDM2* and *FBXW7* (61,62). Visualization of the strongest scoring underlying p53RE and p63RE provides an unprecedented depth of binding information.

While the summary data, such as the *Expression Scores* are particularly helpful to quickly assess the regulation of genes, it is critical to tally the characteristics of individual datasets and genes used for the generation of this summary. Importantly, a low *Expression Score* does not rule out p53 or cell cycle-dependent regulation. In addition to biological variability, such as cell line-specific differences, genes may evade the differential expression detection due to low transcript abundances, low but biologically relevant fold-changes, or methodological limitations (e.g. limited number of replicates and sequencing depth). For example, many non-coding RNAs have low expression levels and thus may have a low *p53 Expression Score* although they are actually strongly regulated by p53 (e.g. *DINOL* displayed in Figure 1E). From a statistical perspective, this situation would require the integration of more datasets to increase the statistical power. To address this limitation in spite of additional datasets, we display the 'p53 median log2FC'. The combination of a low *p53 Expression Score* and a high 'p53 median log2FC' might indicate that a gene evades differential expression detection due to a low overall expression level.

Together, with TargetGeneReg 2.0 we provide a comprehensive resource on p53-dependent regulation in humans and mice. Additional information on ΔNp63 and cell cycle-dependent gene regulation facilitates the discovery of further novel biological insights.

DATA AVAILABILITY

All data are available through the web-atlas at www. TargetGeneReg.org. Accession numbers of individual datasets are available through the website.

ACKNOWLEDGEMENTS

We gratefully acknowledge Bernd Senf from the FLI Core Facility Life Science Computing for help with the website's server backbone. We thank Arne Sahm for critical reading of the manuscript.

REFERENCES

1. Sadasivam,S. and DeCaprio,J.A. (2013) The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer*, **13**, 585–595.
2. Fischer,M. and Müller,G.A. (2017) Cell cycle transcription control: DREAM/MuvB and RB-E2F complexes. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 638–662.
3. Fischer,M. (2017) Census and evaluation of p53 target genes. *Oncogene*, **36**, 3943–3956.
4. Sullivan,K.D., Galbraith,M.D., Andrysik,Z. and Espinosa,J.M. (2018) Mechanisms of transcriptional regulation by p53. *Cell Death Differ.*, **25**, 133–143.
5. Sammons,M.A., Nguyen,T.-A.T., McDade,S.S. and Fischer,M. (2020) Tumor suppressor p53: from engaging DNA to target gene regulation. *Nucleic Acids Res.*, **48**, 8848–8869.
6. Fischer,M., Grossmann,P., Padi,M. and DeCaprio,J.A. (2016) Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res.*, **44**, 6070–6086.
7. Fischer,M., Steiner,L. and Engeland,K. (2014) The transcription factor p53: not a repressor, solely an activator. *Cell Cycle*, **13**, 3037–3058.
8. Fischer,M., Uxa,S., Stanko,C., Magin,T.M. and Engeland,K. (2017) Human papilloma virus E7 oncoprotein abrogates the p53-p21-DREAM pathway. *Sci. Rep.*, **7**, 2603.
9. Fischer,M. (2019) Conservation and divergence of the p53 gene regulatory network between mice and humans. *Oncogene*, **38**, 4095–4109.
10. Riege,K., Kretzmer,H., Sahm,A., McDade,S.S., Hoffmann,S. and Fischer,M. (2020) Dissecting the DNA binding landscape and gene regulatory network of p63 and p53. *Elife*, **9**, e63266.
11. Quaas,M., Müller,G.A. and Engeland,K. (2012) p53 can repress transcription of cell cycle genes through a p21 WAF1/CIP1-dependent switch from MMB to DREAM protein complex binding at CHR promoter elements. *Cell Cycle*, **11**, 4661–4672.
12. Fischer,M., Grundke,I., Sohr,S., Quaas,M., Hoffmann,S., Knörck,A., Gumhold,C. and Rother,K. (2013) p53 and cell cycle dependent transcription of kinesin family member 23 (KIF23) is controlled Via a CHR promoter element bound by DREAM and MMB complexes. *PLoS One*, **8**, e63187.
13. Fischer,M., Quaas,M., Wintsche,A., Müller,G.A. and Engeland,K. (2014) Polo-like kinase 4 transcription is activated via CRE and NRF1 elements, repressed by DREAM through CDE/CHR sites and deregulated by HPV E7 protein. *Nucleic Acids Res.*, **42**, 163–180.
14. Fischer,M., Quaas,M., Nickel,A. and Engeland,K. (2015) Indirect p53-dependent transcriptional repression of Survivin, CDC25C, and PLK1 genes requires the cyclin-dependent kinase inhibitor

p21/CDKN1A and CDE/CHR promoter sites binding the DREAM complex. *Oncotarget*, **6**, 41402–41417.

15. Fischer,M., Quaas,M., Steiner,L. and Engeland,K. (2016) The p53-p21-DREAM-CDE/CHR pathway regulates G2/M cell cycle genes. *Nucleic Acids Res.*, **44**, 164–174.

16. Uxa,S., Bernhart,S.H., Mages,C.F.S., Fischer,M., Kohler,R., Hoffmann,S., Stadler,P.F., Engeland,K. and Müller,G.A. (2019) DREAM and RB cooperate to induce gene repression and cell-cycle arrest in response to p53 activation. *Nucleic Acids Res.*, **47**, 9087–9103.

17. Schade,A.E., Fischer,M. and DeCaprio,J.A. (2019) RB, p130 and p107 differentially repress G1/S and G2/M genes after p53 activation. *Nucleic Acids Res.*, **47**, 11197–11208.

18. Macedo,J.C., Vaz,S., Bakker,B., Ribeiro,R., Bakker,P.L., Escandell,J.M., Ferreira,M.G., Medema,R., Foijer,F. and Logarinho,E. (2018) FoxM1 repression during human aging leads to mitotic decline and aneuploidy-driven full senescence. *Nat. Commun.*, **9**, 2834.

19. Decaesteker,B., Denecker,G., Van Neste,C., Dolman,E.M., Van Loocke,W., Gartlgruber,M., Nunes,C., De Vloed,F., Depuydt,P., Verboom,K. *et al.* (2018) TBX2 is a neuroblastoma core regulatory circuitry component enhancing MYCN/FOXM1 reactivation of DREAM targets. *Nat. Commun.*, **9**, 4866.

20. Braun,L., Brenier-Pinchart,M.-P., Hammoudi,P.-M., Cannella,D., Kieffer-Jaquinod,S., Vollaire,J., Josserand,V., Touquet,B., Couté,Y., Tardieux,I. *et al.* (2019) The Toxoplasma effector TEEGR promotes parasite persistence by modulating NF-κB signalling via EZH2. *Nat. Microbiol.*, **4**, 1208–1220.

21. Werwein,E., Cibis,H., Hess,D. and Klempnauer,K.-H. (2019) Activation of the oncogenic transcription factor B-Myb via multisite phosphorylation and prolyl cis/trans isomerization. *Nucleic Acids Res.*, **47**, 103–121.

22. Diab,A., Gem,H., Swanger,J., Kim,H.Y., Smith,K., Zou,G., Raju,S., Kao,M., Fitzgibbon,M., Loeb,K.R. *et al.* (2020) FOXM1 drives HPV+ HNSCC sensitivity to WEE1 inhibition. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 28287–28296.

23. Ng,J.C.F., Quist,J., Grigoriadis,A., Malim,M.H. and Fraternali,F. (2019) Pan-cancer transcriptomic analysis dissects immune and proliferative functions of APOBEC3 cytidine deaminases. *Nucleic Acids Res.*, **47**, 1178–1194.

24. Bainor,A.J., Saini,S., Calderon,A., Casado-Polanco,R., Giner-Ramirez,B., Moncada,C., Cantor,D.J., Ernlund,A., Litovchick,L. and David,G. (2018) The HDAC-Associated Sin3B protein represses DREAM complex targets and cooperates with APC/C to promote quiescence. *Cell Rep.*, **25**, 2797–2807.

25. Pfister,K., Pipka,J.L., Chiang,C., Liu,Y., Clark,R.A., Keller,R., Skoglund,P., Guertin,M.J., Hall,I.M. and Stukenberg,P.T. (2018) Identification of drivers of aneuploidy in breast tumors. *Cell Rep.*, **23**, 2758–2769.

26. Gonyo,P., Bergom,C., Brandt,A.C., Tsaih,S.-W., Sun,Y., Bigley,T.M., Lorimer,E.L., Terhune,S.S., Rui,H., Flister,M.J. *et al.* (2017) SmgGDS is a transient nucleolar protein that protects cells from nucleolar stress and promotes the cell cycle by regulating DREAM complex gene expression. *Oncogene*, **36**, 6873–6883.

27. Hernandez Borrero,L., Dicker,D.T., Santiago,J., Sanders,J., Tian,X., Ahsan,N., Lev,A., Zhou,L. and El-Deiry,W.S. (2021) A subset of CB002 xanthine analogs bypass p53-signaling to restore a p53 transcriptome and target an S-phase cell cycle checkpoint in tumors with mutated-p53. *Elife*, **10**, e70429.

28. Min,M. and Spencer,S.L. (2019) Spontaneously slow-cycling subpopulations of human cells originate from activation of stress-response pathways. *PLOS Biol.*, **17**, e3000178.

29. Nguyen,T.-A.T., Grimm,S.A., Bushel,P.R., Li,J., Li,Y., Bennett,B.D., Lavender,C.A., Ward,J.M., Fargo,D.C., Anderson,C.W. *et al.* (2018) Revealing a human p53 universe. *Nucleic Acids Res.*, **46**, 8153–8167.

30. Santos,A., Wernersson,R. and Jensen,L.J. (2015) Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res.*, **43**, D1140–D1144.

31. Fischer,M. (2021) Mice are not humans: the case of p53. *Trends Cancer*, **7**, 12–14.

32. Woodstock,D.L., Sammons,M.A. and Fischer,M. (2021) p63 and p53: collaborative partners or dueling rivals? *Front. Cell Dev. Biol.*, **9**, 701986.

33. Coronel,L., Riege,K., Schwab,K., Förste,S., Häckes,D., Semerau,L., Bernhart,S.H., Siebert,R., Hoffmann,S. and Fischer,M. (2021) Transcription factor RFX7 governs a tumor suppressor network in response to p53 and stress. *Nucleic Acids Res.*, **49**, 7437–7456.

34. Fischer,B.A., Chelbi,S.T. and Guarda,G. (2020) Regulatory factor X 7 and its potential link to Lymphoid cancers. *Trends Cancer*, **6**, 6–9.

35. López,C., Kleinheinz,K., Aukema,S.M., Rohde,M., Bernhart,S.H., Hübschmann,D., Wagener,R., Toprak,U.H., Raimondi,F., Kreuz,M. *et al.* (2019) Genomic and transcriptomic changes complement each other in the pathogenesis of sporadic Burkitt lymphoma. *Nat. Commun.*, **10**, 1459.

36. Grande,B.M., Gerhard,D.S., Jiang,A., Griner,N.B., Abramson,J.S., Alexander,T.B., Allen,H., Ayers,L.W., Bethony,J.M., Bhatia,K. *et al.* (2019) Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood*, **133**, 1313.

37. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

38. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

39. Song,L. and Florea,L. (2015) Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience*, **4**, 48.

40. Kopylova,E., Noé,L. and Touzet,H. (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

41. Howe,K.L., Achuthan,P., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., Bhai,J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.

42. Hoffmann,S., Otto,C., Kurtz,S., Sharma,C.M., Khaitovich,P., Vogel,J., Stadler,P.F. and Hackermüller,J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.*, **5**, e1000502.

43. Hoffmann,S., Otto,C., Doose,G., Tanzer,A., Langenberger,D., Christ,S., Kunz,M., Holdt,L.M., Teupser,D., Hackermüller,J. *et al.* (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.*, **15**, R34.

44. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

45. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.

46. Clough,E. and Barrett,T. (2016) The Gene Expression Omnibus Database. In: *Methods in Molecular Biology*. NIH Public Access, Clifton, NJ, Vol. **1418**, pp. 93–110.

47. Zheng,R., Wan,C., Mei,S., Qin,Q., Wu,Q., Sun,H., Chen,C.-H., Brown,M., Zhang,X., Meyer,C.A. *et al.* (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.*, **47**, D729–D735.

48. Navarro Gonzalez,J., Zweig,A.S., Speir,M.L., Schmelter,D., Rosenbloom,K.R., Raney,B.J., Powell,C.C., Nassar,L.R., Maulding,N.D., Lee,C.M. *et al.* (2021) The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.*, **49**, D1046–D1057.

49. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

50. Schmitt,A.M., Garcia,J.T., Hung,T., Flynn,R.A., Shen,Y., Qu,K., Payumo,A.Y., Peres-da-Silva,A., Broz,D.K., Baum,R. *et al.* (2016) An inducible long noncoding RNA amplifies DNA damage signaling. *Nat. Genet.*, **48**, 1370–1376.

51. Chaudhary,R., Gryder,B., Woods,W.S., Subramanian,M., Jones,M.F., Li,X.L., Jenkins,L.M., Shabalina,S.A., Mo,M., Dasso,M. *et al.* (2017) Prosurvival long noncoding RNA PINCR regulates a subset of p53 targets in human colorectal cancer cells by binding to Matrin 3. *Elife*, **6**, e23244.

52. Adriaens,C., Standaert,L., Barra,J., Latil,M., Verfaillie,A., Kalev,P., Boeckx,B., Wijnhoven,P.W.G., Radaelli,E., Vermi,W. *et al.* (2016) p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.*, **22**, 861–868.

53. Mello,S.S., Sinow,C., Raj,N., Mazur,P.K., Bieging-Rolett,K., Broz,D.K., Imam,J.F.C.C., Vogel,H., Wood,L.D., Sage,J. *et al.* (2017) Neat1 is a p53-inducible lincRNA essential for transformation suppression. *Genes Dev.*, **31**, 1095–1108.

54. Fishilevich,S., Nudel,R., Rappaport,N., Hadar,R., Plaschkes,I., Iny Stein,T., Rosen,N., Kohn,A., Twik,M., Safran,M. *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, **2017**, bax028.

55. Whitfield,M.L., George,L.K., Grant,G.D. and Perou,C.M. (2006) Common markers of proliferation. *Nat. Rev. Cancer*, **6**, 99–106.

56. Whitfield,M.L., Sherlock,G., Saldanha,A.J., Murray,J.I., Ball,C.A., Alexander,K.E., Matese,J.C., Perou,C.M., Hurt,M.M., Brown,P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell*, **13**, 1977–2000.

57. Müller,G.A., Stangner,K., Schmitt,T., Wintsche,A. and Engeland,K. (2017) Timing of transcription during the cell cycle: protein complexes binding to E2F, E2F/CLE, CDE/CHR, or CHR promoter elements define early and late cell cycle gene expression. *Oncotarget*, **8**, 97736–97748.

58. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.

59. Fischer,M. (2016) p21 governs p53's repressive side. *Cell Cycle*, **15**, 2852–2853.

60. Coronel,L., Häckes,D., Schwab,K., Riege,K., Hoffmann,S. and Fischer,M. (2022) p53-mediated AKT and mTOR inhibition requires RFX7 and DDIT4 and depends on nutrient abundance. *Oncogene*, **41**, 1063–1069.

61. Barak,Y., Gottlieb,E., Juven-Gershon,T. and Oren,M. (1994) Regulation of mdm2 expression by p53: Alternative promoters produce transcripts with nonidentical translation potential. *Genes Dev.*, **8**, 1739–1749.

62. Sionov,R.V., Netzer,E. and Shaulian,E. (2013) Differential regulation of FBXW7 isoforms by various stress stimuli. *Cell Cycle*, **12**, 3547–3554.

63. Macosko,E.Z., Basu,A., Satija,R., Nemesh,J., Shekhar,K., Goldman,M., Tirosh,I., Bialas,A.R., Kamitaki,N., Martersteck,E.M. *et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, **161**, 1202–1214.

64. Dominguez,D., Tsai,Y.-H., Weatheritt,R., Wang,Y., Blencowe,B.J. and Wang,Z. (2016) An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife*, **5**, e10288.

65. Liu,Y., Chen,S., Wang,S., Soares,F., Fischer,M., Meng,F., Du,Z., Lin,C., Meyer,C., DeCaprio,J.A. *et al.* (2017) Transcriptional landscape of the human cell cycle. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 3473–3478.

66. Hao,Q., Zong,X., Sun,Q., Lin,Y.-C., Song,Y.J., Hashemikhabir,S., Hsu,R.Y.C., Kamran,M., Chaudhary,R., Tripathi,V. *et al.* (2020) The S-phase-induced lncRNA SUNO1 promotes cell proliferation by controlling YAP1/Hippo signaling pathway. *Elife*, **9**, e55102.