Molecular Cell, Volume *63*

Supplemental Information

A Long Noncoding RNA Regulates Sister Chromatid Cohesion

Francesco P. Marchese, Elena Grossi, Oskar Marín-Béjar, Sanjay Kumar Bharti, Ivan Raimondi, Jovanna González, Dannys Jorge Martínez-Herrera, Alejandro Athie, Alicia Amadoz, Robert M. Brosh, Jr., and Maite Huarte

SUPPLEMENTAL INFORMATION

A long noncoding RNA regulates sister chromatid cohesion

Francesco P. Marchese[1,2], Elena Grossi[1,2], Oskar Marín-Béjar[1,2], Sanjay Kumar Bharti[3], Ivan Raimondi[1,2], Jovanna González[1,2], Dannys Jorge Martínez-Herrera[1,2], Alejandro Athie[1,2], Alicia Amadoz[1,2,4], Robert M. Brosh Jr[3] and Maite Huarte[1,2,*]

[1]Center for Applied Medical Research (CIMA), Department of Gene Therapy and Regulation of Gene Expression, University of Navarra, 55 Pio XII Ave., 31008, Pamplona, Spain.

[2]Institute of Health Research of Navarra (IdiSNA), Pamplona, Spain.

[3]Laboratory of Molecular Gerontology, National Institute on Aging, National Institutes of Health, NIH Biomedical Research Center, 251 Bayview Blvd, Baltimore, MD 21224 USA.

[4]Current address: Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, 46012, Spain.

*Corresponding author: maitehuarte@unav.es

SUPPLEMENTAL FIGURES

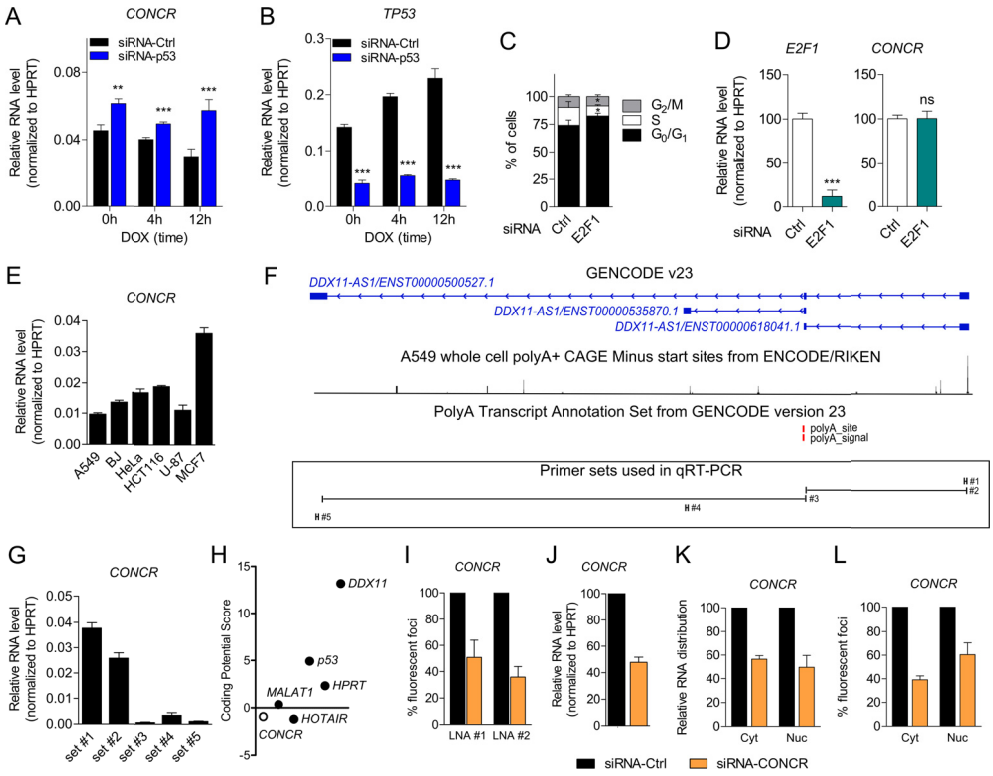

Figure S1, related to Figure 1

(A-B) CONCR and TP53 expression levels determined by qRT-PCR of A549 cells transfected with a control siRNA (siRNA-Ctrl) or with a siRNA targeting TP53 (siRNA-p53), either untreated or treated with the DNA damaging drug doxorubicin (DOX) for 4 and 12 hours. Graph shows mean ± SEM of three independent experiments.

(C-D) Analysis of cell cycle phase distribution by flow cytometry of propidium iodide stained A549 cells either transfected with a control siRNA (Ctrl) or with a siRNA targeting E2F1. (D) E2F1 and CONCR RNA levels determined by qRT-PCR. Graph shows mean ± SEM of three independent experiments.

(E) CONCR relative expression level in different cell lines (A549, BJ, HeLa, HCT116, U-87 and MCF7) determined by qRT-PCR. Graph shows mean ± SD of two independent experiments.

(F) CONCR (DDX11-AS1) isoforms as annotated in GENCODE version 23; position of the 5' and 3' ends by 5' cap gene expression analysis (CAGE) and polyA site and signal; positions along CONCR sequence of the different primer sets (#1 to #5) used in the qRT-PCRs shown in (G).

(G) CONCR relative expression level in A549 determined by qRT-PCR using different primer sets (#1 to #5). Graph shows mean ± SD of three independent experiments.

(H) CONCR coding potential evaluated using the coding potential calculator generated by (Kong et al., 2007).

(I) Percentage of fluorescent foci detected by RNA-FISH in A549 transfected with a control siRNA (siRNA-Ctrl) or with a combination of two siRNAs targeting CONCR (siRNA-CONCR) using two independent LNA probes (LNA #1 and #2). Fluorescent foci were quantified imaging and counting approximately one hundred cells per condition. Graph shows mean ± SD of two independent experiments.

(J) CONCR knockdown efficiency was determined by qRT-PCR. A fraction of cells analysed in (I) was used. Graph shows mean ± SD of two independent experiments.

(K-L) Percentages of CONCR knockdown efficiencies in the cytoplasm (Cyt) and nucleus (Nuc) of A549 cells, analysed by subcellular fractionation (K) or RNA-FISH (L).

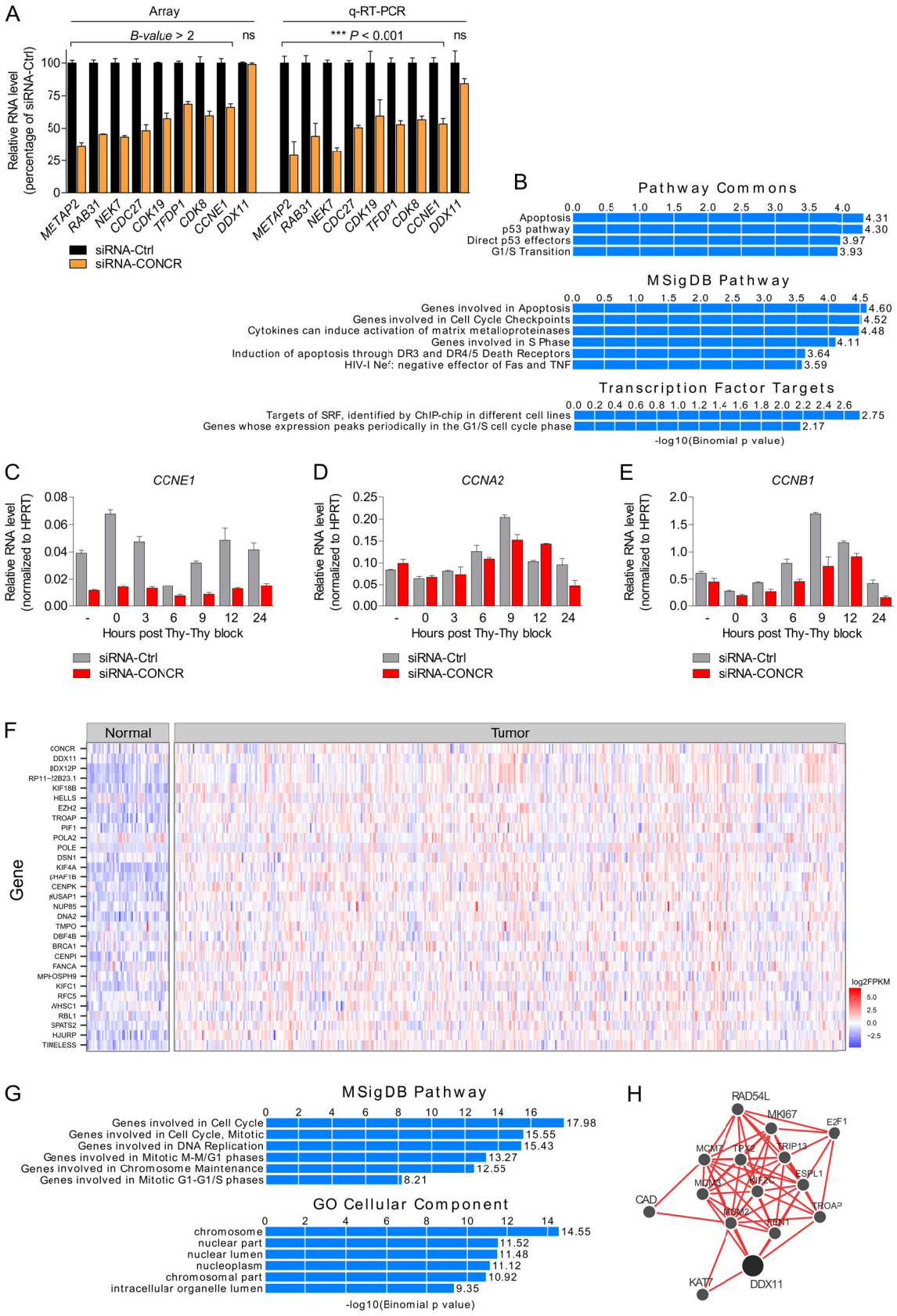

Figure S2, related to Figure 3

(A) Gene expression analysis of A549 cells depleted of CONCR. A549 were transfected with a control siRNA (siRNA-Ctrl) or with a combination of two siRNAs targeting CONCR (siRNA-

CONCR) and RNA levels analysed by microarray (left) or qRT-PCR (right). Graph shows mean ± SEM of three independent experiments. Significance was determined by LIMMA for the array and by two-tailed unpaired t test for the qRT-PCR.
(B) GREAT functional and pathway analyses of the genes differentially expressed.
(C-E) A549 cells were either transfected with a control siRNA (Ctrl) or with two siRNAs targeting CONCR in combination. G1/S synchronized cells were obtained by double thymidine block procedure. Normal medium was then used for the release and cells collected at the different time points indicated for CCNE1, CCNA2 and CCNB1 expression analysis by qRT-PCR. Graphs show mean ± SD of two independent experiments.
(F) Correlation analysis of CONCR expression level in both normal and lung adenocarcinoma samples of the TCGA-LUAD (lung adenocarcinoma) dataset (Cancer Genome Atlas Research, 2014). Top CONCR co-expressed genes are plotted. Graph shows expression values, as log_2FPKM, in a colour scheme for both normal (n = 54) and lung adenocarcinoma (n = 441) samples obtained from the TCGA project database.
(G) GREAT functional and pathway analyses of genes co-expressed (correlation value > 0.45) with CONCR in the LUAD dataset.
(H) Multi-tissue functional interaction network of DDX11 obtained from GIANT (giant.princeton.edu) (Greene et al., 2015).

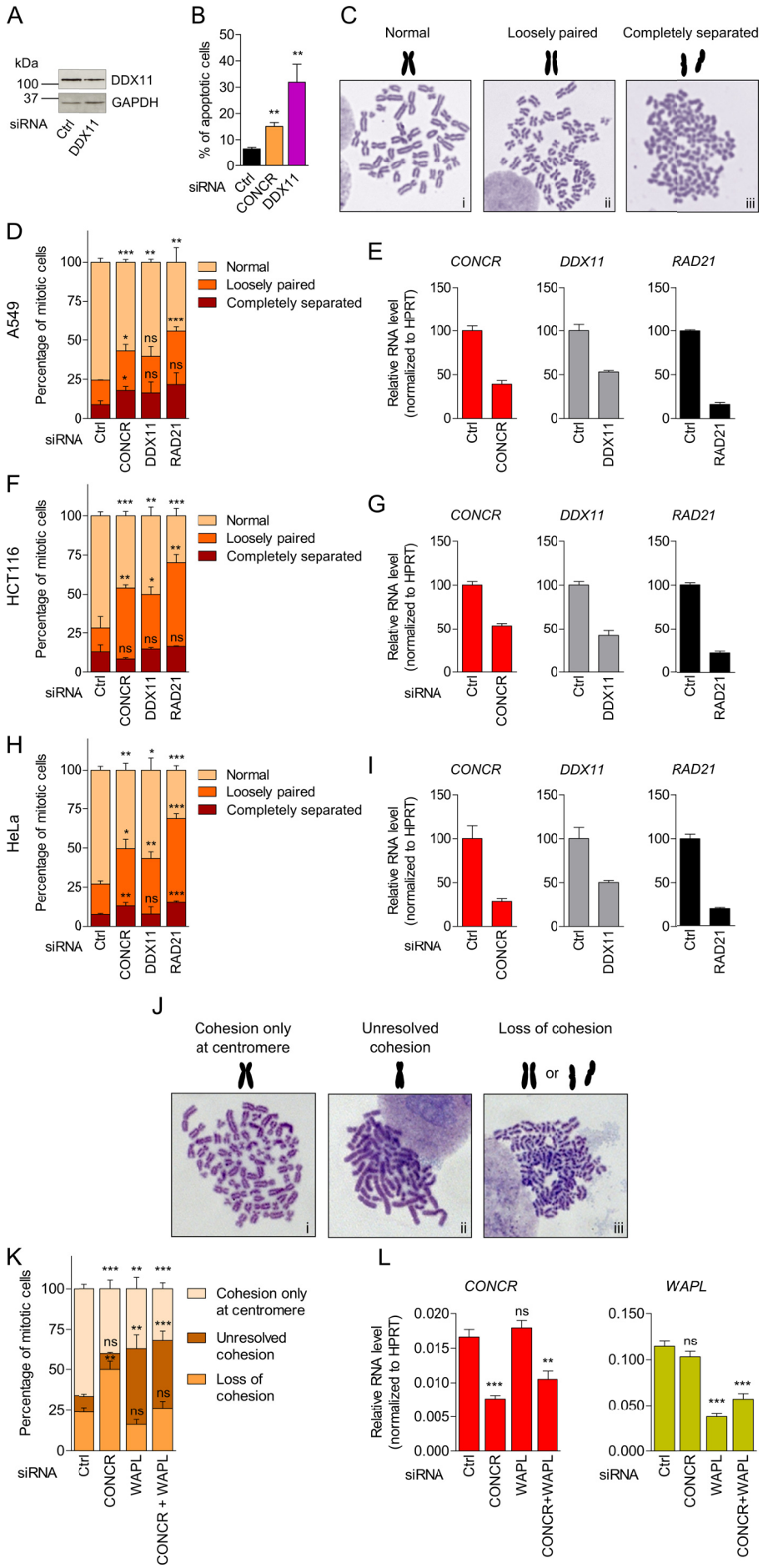
A
kDa
100
37
DDX11
GAPDH
siRNA
Ctrl
DDX11
B
% of apoptotic cells
siRNA
Ctrl
CONCR
DDX11
C
Normal
Loosely paired
Completely separated
i
ii
iii
D
A549
Percentage of mitotic cells
Normal
Loosely paired
Completely separated
siRNA
Ctrl
CONCR
DDX11
RAD21
E
CONCR
DDX11
RAD21
Relative RNA level (normalized to HPRT)
F
HCT116
G
H
HeLa
I
J
Cohesion only at centromere
Unresolved cohesion
Loss of cohesion
or
K
Cohesion only at centromere
Unresolved cohesion
Loss of cohesion
CONCR + WAPL
L
WAPL
CONCR+WAPL

Figure S3, related to Figure 4

(A) DDX11 levels in A549 cells transfected with a siRNA-control (Ctrl) or a pool of two siRNAs targeting DDX11.
(B) Percentage of apoptotic cells determined by flow cytometry of annexin V and 7-AAD stained cells.
(C) Representative images of chromosome spreads showing normal X-shaped conformation (i) or cohesion defects in the sister chromatid, classified as “loosely paired” (ii) or “completely separated” (iii).
(D-I) Cells were transfected with a control siRNA (Ctrl) or with siRNA targeting CONCR, DDX11 or RAD21.
(D) Percentage of A549 mitotic cells showing normal sister chromatid cohesion or cohesion defects classified as “loosely paired” or “completely separated”. At least fifty metaphases per condition were scored and each experiment blindly scored twice. Graph shows mean ± SD of two independent experiments.
(E) CONCR, DDX11 and RAD21 knockdown efficiencies determined by qRT-PCR.
(F-G) As (D) and (E) in HCT116 cells.
(H-I) As (D) and (E) in HeLa cells.
(J-L) A549 cells were transfected with a control siRNA (Ctrl) or with siRNA targeting either CONCR, WAPL or both (CONCR+WAPL).
(J) Representative images of chromosome spreads showing different cohesion phenotypes, classified as “cohesion only at centromere” (i) “unresolved cohesion” (ii) or “loss of cohesion” (iii) as quantified in (K). At least fifty metaphases per condition were scored and each experiment blindly scored twice. Graph shows mean ± SD of three independent experiments.
(L) CONCR, and WAPL knockdown efficiencies determined by qRT-PCR.

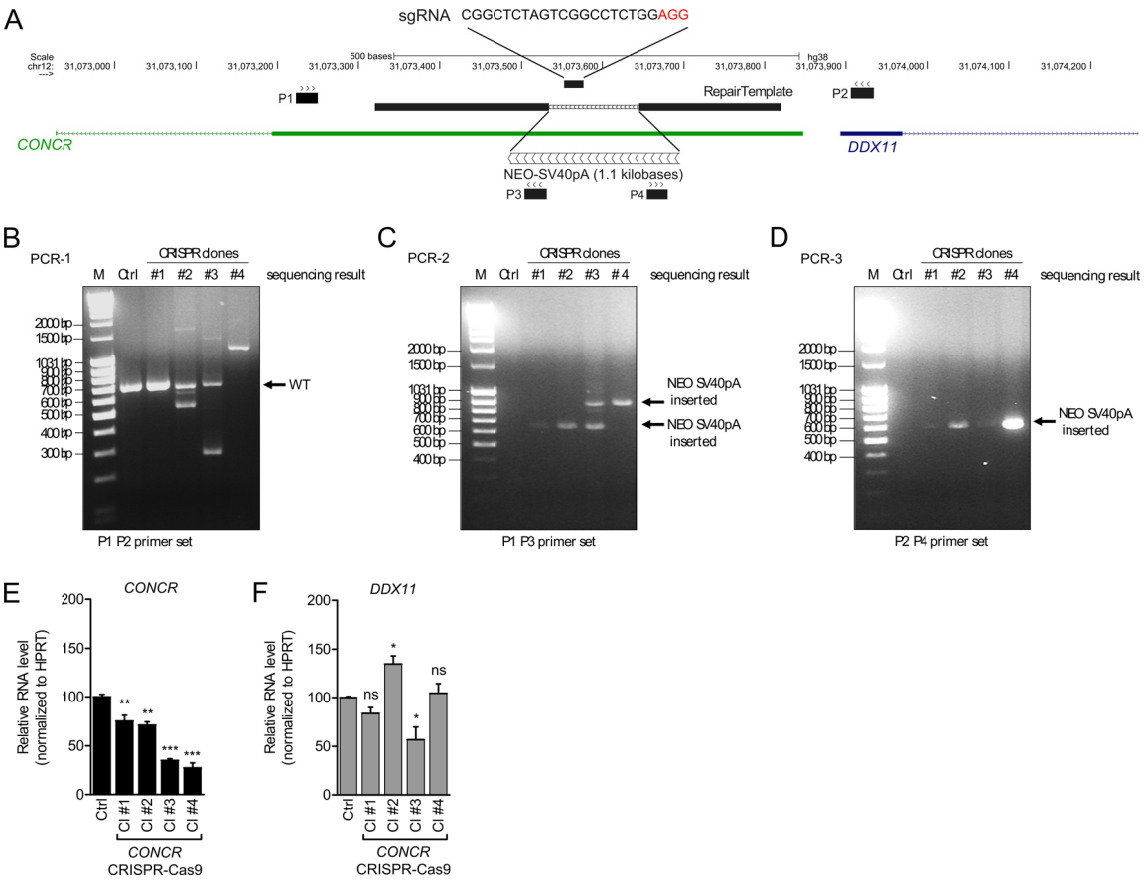

Figure S4, related to Figure 4

(A) Schematic of the editing strategy of the CONCR genomic locus by CRISPR-Cas9. The position and sequence of the sgRNA is shown at the top. The repair template used to insert the NEO-SV40pA sequence at the site of cleavage is shown; flanking sequences to the cleavage site are drawn as black lines and the NEO-SV40pA sequence is drawn in white with arrows indicating the orientation. The position and orientation of the primers (P1 to P4) used for the clones screening by PCR (B-D) is shown with black lines and arrows.

(B-D) Screening of the CRISPR clones by PCR and sequencing. (B) Genomic DNA was amplified by PCR using the primer set P1-P2. The bands indicated with the arrow were excised and sequenced confirming the WT sequence. (C) Genomic DNA was amplified by PCR using the primer set P1-P3. The bands indicated with the arrow were excised and sequenced confirming the insertion of the NEO-SV40pA sequence. (D) Genomic DNA was amplified by PCR using the primer set P2-P4. The bands indicated with the arrow were excised and sequenced confirming the insertion of the NEO-SV40pA sequence.

(E-F) CONCR and DDX11 relative levels were determined by qRT-PCR. Graphs show mean ± SD of three independent experiments.

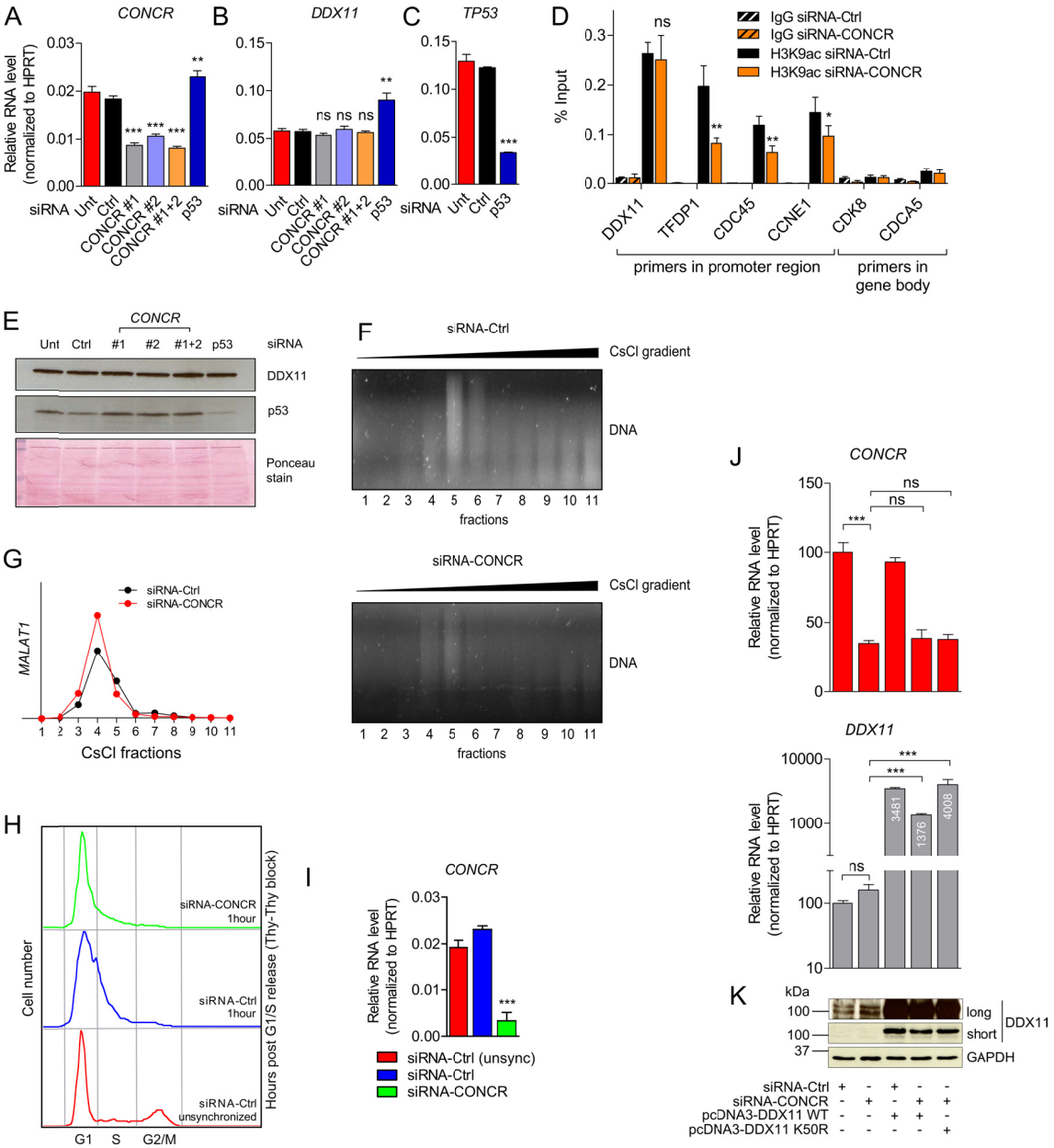

Figure S5, related to Figure 5 and Figure 6

(A-C) A549 cells were left untransfected (Unt), transfected with a control siRNA (Ctrl), with two siRNAs targeting CONCR, separately (#1 and #2) or in combination (#1+2), or with a siRNA targeting TP53. CONCR, DDX11 and TP53 relative levels were determined by qRT-PCR. Graphs show mean ± SEM of three independent experiments. Significance was determined comparing to the siRNA-Ctrl.

(D) H3K9ac ChIP-qRT-PCR in control (siRNA-Ctrl) and CONCR depleted cells (siRNA-CONCR).

(E) DDX11 and p53 levels were determined by western blotting. Ponceau staining was used as loading control.

(F-G) A549 cells were transfected with a siRNA-control (Ctrl) or a pool of two siRNAs targeting CONCR and cross-linked cell extracts were fractionated by CsCl density-gradient centrifugation. Protein, DNA and RNA were then isolated (details in experimental procedures). (F) DNA distribution analysed by ethidium bromide staining of agarose gel. (G) MALAT1 qRT-PCR showing enrichment of the lncRNA in the chromatin fractions (fractions 4 and 5).

(H-I) A549 cells were either transfected with a control siRNA (Ctrl) or with two siRNAs targeting CONCR in combination. G1/S synchronized cells were obtained by double thymidine block procedure. Normal medium was then used for the release and cells used for the DDX11 ChIP-seq shown in Figure 5. In parallel, aliquots were collected for cell cycle analysis (H) and RNA expression analysis by qRT-PCR (I). Graph shows mean ± SD of two independent experiments, while in (H) cell cycle profiles of one representative experiment are shown.
(J-K) CONCR and DDX11 levels determined by qRT-PCR (J) and western blot (K) in A549 cells transfected with the indicated combinations of siRNAs and DDX11 expression plasmids. Graphs show mean ± SD of two independent experiments.

Figure S6, related to Figure 5

(A-B) Representative images of the DDX11 ChIP-sequencing. Entire chromosome 16 (A) and chromosome 22 (B). *Top* to *bottom*: Chromosome schematic of two representative regions of chr16 and chr22; DDX11 ChIP-seq signal in control (Ctrl) and CONCR depleted cells (KD); regions with differential binding of DDX11 (Ctrl *vs* KD); signals of DNA replicating regions of BJ cells in G1 to G2 phases of the cell cycle as reported in (Hansen et al., 2010); G1 replication enriched regions common to BJ, GM06990, H0287 and BG02 cell types as reported in (Hansen et al., 2010).

SUPPLEMENTAL TABLES

Table S1. RNA-seq of HCT116 p53+/+ and p53-/-: list of mRNAs and lincRNAs higher in p53-/-, Related to Figure 1.

Table S2. Statistics and TCGA data used in the study, related to Figure 2.

Table S3. Microarray analysis of A549 cells depleted of CONCR, related to Experimental Procedures.

Table S4. CONCR gene expression correlation in lung adenocarcinoma, related to Experimental Procedures.

Table S5. DDX11 ChIP-seq analysis, related to Figure 5.

Table S6. List of oligonucleotides used in the study, related to Figure 1, 2, 3, 4, 5, 6.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Cross-linked chromatin fractionation by CsCl density-gradient centrifugation

All steps were performed as previously described (Dellino et al., 2013; Schwartz et al., 2005) with minor modification. A549 cells were cross-linked with formaldehyde to a final concentration of 1% for 10 minutes at room temperature. The reaction was stopped by addition of glycine solution to a final concentration of 125 mM. Cells were harvested by scraping, washed once with PBS and resuspended in sonication buffer (50 mM Tris-HCl pH 8; 10 mM EDTA; 1% SDS; protease inhibitors). The suspension was sonicated (three cycles of 15 sec ON, 30 sec OFF) at medium constant power. The clear lysate was mixed with 6,81 g of CsCl, and the sample volume adjusted to 12 ml with gradient buffer (10mM Tris-HCl pH 7.5; 1 mM EDTA, 0.5 mM EGTA; 0.5% N-lauroylsarcosine). Centrifugation was at 20°C for 72 h at 35900 rpm in a Beckman SW40Ti rotor on a Beckman Optima L100XP. After centrifugation, 12

fractions were collected, crosslinking was reversed at 65°C overnight and protein, RNA and DNA purified by TCA, TRIzol and phenol-chloroform precipitation respectively.

Chromatin immunoprecipitation (ChIP)

RNAi and G1/S synchronization were performed as described above. Briefly, cells were cross-linked with 1% formaldehyde and chromatin prepared by sonication to obtain DNA fragments between 100 bp and 500 bp. DDX11 or H3K9ac bound DNA was then immunoprecipitated using the anti-DDX11 antibody (ab66971, Abcam) or the anti-H3K9ac antibody (ab4441, Abcam), decrosslinked and extracted for analysis by sequencing or qRT-PCR.

ChIP sequencing (ChIP-seq)

For DDX11 ChIP-seq libraries were prepared using the NEBNext Ultra DNA Library Prep Kit (Illumina) and sequenced using the NextSeq 500 system (Illumina) with 50 M reads per sample in average. Unfiltered 75-bp Illumina reads were aligned to the human reference genome (NCBI Build 37, hg19) using Bowtie2 (Langmead and Salzberg, 2012) with the default parameters. FeatureCounts v1.5.0 (Liao et al., 2014) was used to assign and quantify the ChIP-Seq reads aligned to the 867 consensus G1 replication-enriched regions common between BJ, GM06990, H0287 and BG02 cells as defined by the previously published Repli-seq study (Hansen et al., 2010) (Table S5). The regions differentially expressed (logFC ≠ 0) between *CONCR* knockdown and control ChIP-seq data were identified using the Bioconductor package LIMMA (Smyth, 2004).

ATP hydrolysis assays

DDX11 recombinant protein was purified using a protocol previously described (Capo-Chichi et al., 2013). Briefly, pcDNA3-His_6-DDX11-3xFLAG plasmid containing human DDX11 cDNA was transfected into 293T cells using Lipofectamine 2000. Nuclear extract was incubated with anti-

FLAG affinity resin (Sigma) and eluted with 3X FLAG peptide. Purified recombinant DDX11 migrated as a single band on Coomassie-stained SDS-PAGE. *CONCR* cDNA clone was obtained from OriGene Technologies (cDNA FLJ39041 fis, clone NT2RP7010109) and sequence corresponding to ENST00000618041.1 was subcloned by PCR into pcR4-TOPO (Invitrogen). *CONCR* and control RNA (antisense sequence of *CONCR*) were then obtained by *in vitro* transcription with T3 or T7 RNA polymerase respectively. Effect of *CONCR* and control RNA on ATP hydrolysis was measured using [γ-^{32}P] ATP (PerkinElmer Life Sciences) and analysis by thin layer chromatography on polyethyleneimine-cellulose plates (Mallinckrodt Baker). The standard reaction mixture (20 µl total volume) contained 80 nM or the indicated concentration of RNA in 25 mM Hepes pH 7.5, 25 mM potassium acetate, 1 mM magnesium acetate, 1 mM DTT, 100 µg/ml bovine serum albumin, 250 µM [γ-^{32}P] ATP, and 40 nM DDX11 protein. Reaction mixtures were incubated at 37 °C followed by quench with 50 mM EDTA (final concentration). The reaction mixture was spotted onto a polyethyleneimine-cellulose TLC plate and resolved using 0.5 M LiCl, 1 M formic acid as the carrier solvent. The TLC plate was exposed to a phospho rimaging cassette for 1 h, visualized using a PhosphorImager, and analyzed with ImageQuant software (GE Healthcare Life Sciences).

RNA sequencing (RNA-seq)

Paired-end and strand-specific RNA-seq libraries were prepared from purified poly-A+ RNA from untreated and 5-FU-treated p53$^{+/+}$ and p53$^{-/-}$ HCT116 cells for 4 and 12 h, sequenced and analysed as previously described (Sanchez et al., 2014).

Western blotting

Briefly, protein concentrations were estimated by Bradford assay. For western blotting, proteins were separated by SDS-PAGE and transferred to nitrocellulose. The membranes were blocked and probed with the primary antibody, washed and probed with the required

secondary antibody HRP-conjugated. Detection was performed by ECL according to the manufacturer's instructions. The antibodies used were as follows: anti-DDX11 (ab66971, Abcam), anti-WDR5 (ab56919, Abcam), anti-FEN1 (ab17994, Abcam), anti-H3 (4499, Cell Signaling), anti-p53 (sc-126, Santa Cruz Biotechnology), anti-GAPDH (ab9484, Abcam).

RNA immunoprecipitation (RIP)

RNA immunoprecipitation was performed following cross-linking of cells. Cells were cross-linked with 0.5% formaldehyde and lysed with nuclear isolation buffer (1.28 M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 4 % Triton X-100). Nuclei were pelleted by centrifugation at 2,500 x g for 15 min. Nuclear pellet was resuspended in RIP buffer (150 mM KCl, 25 mM Tris-HCl pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP-40, supplemented with 1X cOmplete Protease Inhibitor Cocktail (Roche) and SUPERaseIN (Ambion) 10 U/ml) and divided into two fractions for IgG control and IP. Nuclei were then mechanically sheared using a Dounce homogenizer. The anti-DDX11 antibody (ab66971, Abcam) was then added to the nuclear extract and incubated overnight at 4°C with gentle rotation. 50 µl of protein G magnetic beads were added and incubated for 2 h at 4°C with gentle rotation. Beads were collected using a magnet, washed four times wth RIP buffer and immunoprecipitated RNA was finally eluted by heating the beads at 70°C for 15 minutes and extracted using TRIzol.

Statistical analysis

Unless specified otherwise in the Figure legend significance was determined by two-tailed unpaired *t* test using GraphPad Prism (ns, not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

SUPPLEMETAL REFERENCES

Cancer Genome Atlas Research, N. (2014). Comprehensive molecular profiling of lung adenocarcinoma. Nature *511*, 543-550.
Capo-Chichi, J.M., Bharti, S.K., Sommers, J.A., Yammine, T., Chouery, E., Patry, L., Rouleau, G.A., Samuels, M.E., Hamdan, F.F., Michaud, J.L*., et al.* (2013). Identification and biochemical characterization of a novel mutation in DDX11 causing Warsaw breakage syndrome. Hum Mutat *34*, 103-107.
Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res *23*, 1-11.
Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C*., et al.* (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat Genet *47*, 569-576.
Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A *107*, 139-144.
Kong, L., Zhang, Y., Ye, Z.Q., Liu, X.Q., Zhao, S.Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res *35*, W345-349.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.
Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.
Sanchez, Y., Segura, V., Marin-Bejar, O., Athie, A., Marchese, F.P., Gonzalez, J., Bujanda, L., Guo, S., Matheu, A., and Huarte, M. (2014). Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. Nat Commun *5*, 5812.
Schwartz, Y.B., Kahn, T.G., and Pirrotta, V. (2005). Characteristic low density and shear sensitivity of cross-linked chromatin containing polycomb complexes. Mol Cell Biol *25*, 432-439.
Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.