Oncogenes hijack a constitutively active *TP53* promoter in osteosarcoma

Karim H. Saba[1], Louise Cornmark[1]†, Michal Kovac[2]†, Linda Magnusson[1], Jenny Nilsson[1], Hilda van den Bos[3], Diana C. J. Spierings[3], Mahtab Bidgoli[4], Tord Jonson[4], Vaiyapuri P. Sumathi[5], Otte Brosjö[6], Johan Staaf[7], Floris Foijer[3], Emelie Styring[8], Michaela Nathrath[9,10], Daniel Baumhoer[2]‡, Karolin H. Nord[1]*‡

Affiliations

[1]Department of Laboratory Medicine, Division of Clinical Genetics, Lund University, Lund, Sweden.

[2]Bone Tumour Reference Centre at the Institute of Pathology, University Hospital and University of Basel, Basel, Switzerland.

[3]European Research Institute for the Biology of Ageing, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands.

[4]Department of Clinical Genetics and Pathology, Laboratory Medicine, Medical Services, Skåne University Hospital, Lund, Sweden.

[5]Department of Musculoskeletal Pathology, Royal Orthopaedic Hospital, Birmingham, UK.

[6]Department of Orthopedics, Karolinska University Hospital, Stockholm, Sweden.

[7]Department of Clinical Sciences, Division of Oncology and Pathology, Lund University, Lund, Sweden.

[8]Department of Orthopedics, Lund University, Skåne University Hospital, Lund, Sweden.

[9]Children's Cancer Research Centre and Department of Pediatrics, Klinikum rechts der Isar, Technische Universität München, Munich, Germany.

[10]Depertment of Pediatric Oncology, Klinikum Kassel, Kassel, Germany.

†These authors contributed equally to this work.

‡D.B. and K.H.N. jointly directed this work.

*Correspondence to: karolin.hansen_nord@med.lu.se

5

Abstract

The malignant bone tumor osteosarcoma harbors an extreme number of chromosome rearrangements. How such massive DNA errors confer a competitive advantage to a cancer cell has remained an enigma. Osteosarcoma typically presents mutations disrupting normal *TP53* gene function, frequently in the form of structural rearrangements that separate the promoter region from the coding parts of the gene. To unravel the consequences of a *TP53* promoter relocated in this manner, we performed in-depth genetic analyses of osteosarcoma biopsies (n=148) and cell models. We show that *TP53* structural variations not only facilitate further chromosomal alterations, but also allow the constitutively active *TP53* promoter to upregulate putative oncogenes erroneously placed under its control. Paradoxically, many of the induced genes are part of the *TP53*-associated transcriptome, suggesting a need to counterbalance the initial loss of function. Our findings demonstrate how the promoter region of a tumor suppressor gene can functionally turn into an oncogenic driver.

Introduction

Osteosarcoma is the most common primary malignancy of the skeleton. The majority of osteosarcomas develop in children and adolescents, often in close proximity to the active growth plate of long bones[1]. After the age of 25, there is an incidence plateau followed by a second, smaller incidence peak in elderly individuals. During the 1980s, the introduction of multidrug chemotherapy dramatically improved the survival rate for osteosarcoma patients. Clinical outcome has improved little since then, however. The overall survival rate remains at 60-80% for localized disease and below 40% for disseminated osteosarcoma[1]. Osteosarcoma typically displays a very large number of numerical and structural chromosome aberrations[2-7]. Hitherto, there are no reports on genetic alterations specific to this disease and a consistent genetic pattern between patients is lacking. Because of this, identification of novel therapeutic targets is particularly challenging for this disease.

Large-scale sequencing efforts have consistently demonstrated that a vast majority of osteosarcomas have loss-of-function mutations in the *TP53* gene[3,4]. In addition to inactivating single nucleotide variants, at least 50% of pediatric osteosarcomas show structural variations in this gene[3,4,6]. These rearrangements separate the promoter region from the coding parts of *TP53*, often resulting in loss of the latter. The promoter region is not lost, but instead relocated enabling the erroneous activation of genes other than those originally under its control. Transfer of promoter activity is a known driver in neoplasia, commonly denoted promoter swapping/switching or enhancer hijacking. Promoter substitution has been shown to operate in bone tumors other than osteosarcoma, *e.g.* in chondromyxoid fibroma and aneurysmal bone cyst where strong promoters are juxtaposed to the entire coding sequences of the *GRM1*

and *USP6* genes, respectively[8,9]. Previously reported promoter substitutions in neoplasia have typically involved promoters assumed to be constitutively active in the cell-of-origin[10]. Here, we use the complex genome of osteosarcoma to test the novel hypothesis that acquired genetic damage can activate a transferred promoter of a tumor suppressor to drive oncogenesis.

Results

Ectopic localization of the *TP53* promoter is more common in young osteosarcoma patients

To make a detailed assessment of the role of *TP53* rearrangements in osteosarcoma, we first subjected a discovery cohort of conventional osteosarcomas from pediatric (age <18 y, *n*=15) and adult (age range 18-81 y, *n*=21) patients to whole genome mate pair sequencing, which is a powerful technology to identify structural genomic alterations. The majority of samples analyzed in this cohort were chemotherapy-treated resection specimens. We found structural rearrangement of *TP53* in 13/36 cases (Figure 1A, Supplementary Tables 1 and 2). We then analyzed an independent validation cohort of treatment-naïve diagnostic biopsies from conventional osteosarcomas, again including both pediatric (age <18 y, *n*=20) and adult (age range 18-59 y, *n*=16) patients. In the validation cohort, structural rearrangement of *TP53* was found in 16/36 cases (Figure 1B, Supplementary Tables 3 and 4). We extended our validation cohort and analyzed genome-wide DNA copy number profiles based on SNP array analyses from treatment-naïve conventional osteosarcomas (age range 3-74 y, *n*=108). For more than one-third of these patients we had matched whole genome mate pair sequencing data (Supplementary Table 3). By integrating array and sequencing data, we identified a subset of

cases with a copy number profile of chromosome arm 17p that we termed '*TP53* promoter gain'. We defined this pattern as copy number loss, or copy number neutral loss of heterozygosity, of whole or parts of the *TP53* coding region coupled to concurrent relative copy number gain of the *TP53* promoter region along with regions of the proximal part of chromosome arm 17p (Figure 1C). We found *TP53* promoter gain in 16/108 cases (15%; Figure 1D). Both *TP53* structural variation, determined by whole genome mate pair sequencing, and *TP53* promoter gain, determined by SNP array analysis, were non-randomly associated with young age of onset (Figure 1E and F, Supplementary Tables 1 and 3). In an additional 24 of the 108 SNP array analyzed cases (22%), we detected a copy number shift within the nearest measuring points downstream and upstream relative to *TP53*, but lacking at least one criterion for *TP53* promoter gain (Supplementary Table 3). Based on whole genome sequencing information, we know that the most likely outcome in this category as well is transposition of the *TP53* promoter region (Figure 1A and B). Collectively, we identified transposition of the *TP53* promoter in 40% of conventional osteosarcomas, *i.e.*, 29/72 by DNA mate pair sequencing and 40/108 by SNP array analysis (Supplementary Tables 1 and 3). This was associated with a high number of chromosome breaks genome-wide as exemplified in Figure 1G.

Recurrent transposition of the *TP53* promoter suggested that it regulates other genes in a fashion that favors tumor development, through gene fusion or promoter swapping events[10]. To test this, we assessed gene expression levels by performing RNA sequencing of conventional osteosarcomas (age range 4-81 y, *n*=66) and, as control, benign osteoblastomas (*n*=13; Supplementary Tables 1 and 3). To evaluate the presence of *TP53* promoter gain, we analyzed DNA copy numbers in cases of the discovery cohort from which material was

available (*n*=12). To determine if *TP53* structural variations were present among multiple samples from the same tumor, we analyzed five osteosarcomas sampled across several regions and time points. In these cases, we compared paired-end whole genome sequencing data from diagnostic biopsies, resection specimens, and/or metastases (*n*=11). To evaluate the proportion of *TP53* structural variations among individual cells from the same tumor, we finally applied single cell low-pass whole genome sequencing to cryopreserved cells from four osteosarcomas. By integrating the obtained high-resolution genomic data with matched transcriptomic information, we found that transposition of the *TP53* promoter is an early event that results in deregulation of several well-known or putative oncogenes. Below we provide several lines of evidence for this mechanism.

Transposition of the *TP53* promoter is a single early event that can spark genome-wide rearrangements and oncogene amplification

In a subset of osteosarcomas, DNA sequencing supported intra- and interchromosomal events (inversions, insertions or translocations) that transposed the *TP53* promoter without compromising chromosome stability (Figure 2A-C, Supplementary Tables 1 and 3). In these cases, we detected no further rearrangements involving the *TP53* promoter or its partner region. In another subset of osteosarcomas, transposition of the *TP53* promoter was the initiating event that generated unstable, most likely dicentric, derivative chromosomes (Figure 2D-F, Supplementary Figures 1 and 2, Supplementary Tables 1 and 3). In osteosarcoma, such derivative chromosomes repeatedly break and rejoin with multiple partner chromosomes[11,12]. This amplifies both the *TP53* fusion and additional genomic regions of potential importance for osteosarcoma progression, such as regions on chromosomes 6, 12 and 17 (Figure 2F).

Notably, this sequence of events is different from chromothripsis and multi-way translocations, which in other subtypes of bone tumors are known to generate gene fusions (Figure 2G and H)[13]. We found no evidence for the generation of *TP53* structural variants or *TP53* gene fusions through one massive burst of genome rearrangements in osteosarcoma. Instead, the genomic footprint of *TP53* gene fusions in osteosarcoma mimics that of oncogene amplification through breakage-fusion-bridge cycles, found in *e.g.* low-grade osteosarcoma with ring chromosomes and *MDM2* amplification (Figure 2I). Thus, according to our model, transposition of the *TP53* promoter is an early spark for genome-wide rearrangements in osteosarcoma. Results from whole genome sequencing of multi-sampled bulk and single cell tumor DNA supported this model. *TP53* fusion positive osteosarcomas harbored their respective fusions in all investigated diagnostic biopsies, post chemotherapy resection specimens and metastases, as well as in all investigated individual tumor cells (Figure 2J and K, Supplementary Figures 2A-F and 3-5, Supplementary Tables 3 and 5).

The bidirectional *TP53* promoter induces the expression of *WRAP53* and oncogenes *in vivo*

To determine if the ectopically located *TP53* promoter regulates genes of functional importance for tumorigenesis, we categorized the *TP53* structural variants based on their copy number state and features of the partner region (Supplementary Figure 6). Out of 30 osteosarcomas with a known *TP53* promoter partner region, 16 displayed at least one partner gene known to be involved in human cancer and/or bone development (Supplementary Tables 1 and 3). To asses if the *TP53* promoter induces the expression of its respective partner genes *in vivo*, we measured the expression levels for both the partner genes and the gene *WRAP53*. The *TP53* promoter normally induces the latter[14], wherefore its elevated expression

level was used as a proxy for adequate representation of neoplastic cells. Figure 3 displays three representative osteosarcomas that harbor whole or parts of *ROR2*, *MAP4K4* and *E2F3*, respectively, placed under the *TP53* promoter. *TP53* exon 1 and partner gene exons placed under the *TP53* promoter showed higher expression levels than exons excluded from the fusions. In Supplementary Tables 1 and 3, we display the matched genomic and transcriptomic data for all detected *TP53* gene fusions. Taken together, these data unequivocally demonstrate that the transposed *TP53* promoter is active in osteosarcoma and that it induces the expression of genes important for tumor and bone development.

Cisplatin evokes oncogene expression through the *TP53* promoter *in vitro*

As a proof-of-concept, we modelled the above findings *in vitro*. First, we created a *TP53*$^{-/-}$ mesenchymal cell line (BJ-5ta) by CRISPR genome editing and single cell cloning. Second, we constructed a vector containing the *TP53* promoter region fused to the coding DNA sequence of *ROR2* (*TP53-ROR2*). As control, we used the same vector but without the *TP53* promoter region (*ROR2*). Third, we exposed *TP53*$^{-/-}$ cells harboring either *TP53-ROR2* or *ROR2* to the DNA damaging agent cisplatin. We found that the *TP53*$^{-/-}$ background, even in the absence of cisplatin, was sufficient to activate the *TP53* promoter and elicit expression of a gene placed under its control (Figure 4A). Induced DNA damage through cisplatin treatment further increased the expression level of the *TP53* promoter partner gene. Thus, in a *TP53*$^{-/-}$ background, a constitutively active *TP53* promoter can induce expression of an oncogene transposed into its vicinity in a fashion accentuated by additional genetic damage.

Genes erroneously placed under the *TP53* promoter are involved in a regulatory network orchestrated by the tumor protein p53

None of the identified *TP53* promoter partner genes were recurrent. We therefore speculated that they are part of a shared network. We noted that several of the *TP53* promoter partner genes, as well as previously suggested target genes in osteosarcoma[6], are putative members of a network regulated by *TP53*[15,16]. Intrigued by this, we performed comparative gene expression analyses of the above-mentioned *TP53*$^{-/-}$ mesenchymal cell line and its wild type counterpart (Figure 4B). Knockdown of *TP53* resulted in significantly reduced expression levels of approximately 3000 genes (Figure 4B and C; $P < 0.01$, Student's *t* test, Supplementary Table 6). The cell model that we used lacked expression of *ROR2* (Figure 4D). This enabled us to test if forced expression of *ROR2* through *TP53-ROR2* in *TP53*$^{-/-}$ cells affected the expression of *TP53*-orchestrated genes. Indeed, induction of *ROR2* under the *TP53* promoter rescued the expression of 1347 of the 3000 most significantly downregulated genes (45%; Figure 4D, Supplementary Table 6). This reverted the global gene expression profile of *TP53*$^{-/-}$ cells towards that of wild type cells (Figure 4B). These 1347 genes included several genes that we had identified as *TP53* promoter partner genes, as well as other genes of potential importance for osteosarcomagenesis such as *COPS3* (Figure 4E and F). The latter is located 10 Mb proximal to *TP53* in chromosome arm 17p and is often co-amplified with the *TP53* promoter region (Figure 2F). Because of its commonly increased copy number state, previous reports have even suggested that *COPS3* is the target for 17p amplification in osteosarcoma[6]. Of the identified *TP53* promoter partner genes, 16 were part of the *TP53*-regulated transcriptome and 20 were induced by *TP53-ROR2* (Figure 4G and H). This set of genes covered 18 of the 30 osteosarcomas with a known *TP53* promoter partner gene. Collectively, these findings suggest

that loss of *TP53* is not beneficial for osteosarcoma, unless there is a simultaneous gain-of-function to save parts of the *TP53*-regulated transcriptome. This may seem paradoxical, but it is important to stress that the response to *TP53*-regulated signaling pathways may be very different in a *TP53* null cell compared to a normal cell[17,18].

In summary, we show that genome rearrangements early in osteosarcoma development silence the *TP53* gene thereby deregulating many *TP53*-associated genes. However, structural variations in *TP53* not only silence *TP53*, but also result in oncogene amplification and hijacking of the active *TP53* promoter by a variety of genes that are part of the *TP53*-orchestrated network. The induction of such genes reverts the aberrant gene expression profile of *TP53*$^{-/-}$ cells towards that of normal cells. Their behavior will not be normal though, as the cancer cells have acquired the detrimental abilities to survive and proliferate despite ongoing genetic damage.

Discussion

The first reports on *TP53* structural rearrangements in osteosarcoma date back to the late 1980s and early 1990s[19,20]. Already then, the clustering of alterations to *TP53* intron 1 was noted and it was speculated that 'rearrangements of p53 in osteosarcoma could activate a second as yet unidentified gene'[19]. During the following decades, efforts from several research groups confirmed these rearrangements, and genomic patterns similar to what we here term '*TP53* promoter gain' were reported in osteosarcoma and subtypes of soft tissue sarcomas[21]. In parallel, somatic structural variations affecting *TP53* were also found in subsets of leukemias and carcinomas, including chronic myelogenous leukemia[22-24], lung cancer[25] and

prostate cancer[26-29]. Such variants inevitably silence the *TP53* gene, but evidence for a concomitant gain-of-function mechanism has not been reported. There may be two probable reasons for not recognizing such a mechanism in previous studies. First, the *TP53* promoter is a promiscuous fusion partner that induces the expression of many different genes. This, however, does not exclude an important functional outcome. There are numerous examples of interchangeable partners of gene fusions that are disease-specific, strongly indicating that activation of a specific pathway, in one way or the other, is the key feature for transformation[10,30]. Second, the *TP53* gene fusions in osteosarcoma involve transfer of promoter activity. Although a well-recognized concept in neoplasia, its detection requires access to matched high quality genomic and transcriptomic data. We generated a unique combined dataset for a large series of pediatric and adult osteosarcomas, sampled across several regions and time points. This enabled us to show for the first time that a promoter activated by genetic damage can induce cancer-driving genes transposed into its vicinity. Genes induced by the *TP53* promoter region in this fashion are often part of the tumor protein p53 pathway. Their induction in *TP53*$^{-/-}$ cells restores parts of the lost pathway, and we speculate that this compensatory mechanism may be crucial for cancer cells to survive loss of p53. Importantly, we found this phenomenon to occur in all tumor cells of *TP53*-rearranged osteosarcomas. This makes it a particularly meaningful mechanism to explore further for therapeutic applications. Massive intratumor heterogeneity, as typically present in osteosarcoma, poses major limitations for the use of so-called personalized or precision medicine[31]. This is because treatment directed towards targets that are not present in all cancer cells will eradicate only sensitive clones and leave resistant clones unaffected. The latter will be able to thrive and eventually kill the patient. Thus, for a targeted strategy to be effective long-term against a genetically heterogeneous tumor, the target must be present in

all tumor cells and required for cell proliferation or survival. We argue that *TP53* gene fusions fulfil these criteria. Our findings encourage further exploration of this phenomenon in osteosarcoma and other genetically complex cancers.

Methods

Subject information and tumor material

Fresh-frozen tumor biopsies from 148 conventional osteosarcomas were subjected to genomic analyses. The clinical features were typical of conventional osteosarcoma patients. The age of the patients ranged from 3-81 years with a median age of 15 years and a mean age of 20 years, and there were 68 females and 80 males. Detailed information is displayed in Supplementary Tables 1 and 3. For comparison, we included osteoblastomas (*n*=13), a chondromyxoid fibroma, a phosphaturic mesenchymal tumor of bone and a parosteal osteosarcoma. All tumor material was obtained after informed consent, and the study was approved by the Regional Ethics Committee of Lund University and the Ethikkommission beider Basel (reference 274/12).

DNA and RNA extractions

Fresh-frozen tumor biopsies were dismembered and homogenized using a Mikro-Dismembrator S (Sartorius AG, Goettingen, Germany). The material was optimally split into two fractions, one used for immediate DNA extraction and the other, when available, was stored in Qiazol at -80°C for later RNA extraction (Qiagen, Hilden, Germany). DNA was

extracted using the DNeasy Blood & Tissue Kit including the optional RNase A treatment (Qiagen). DNA quality and concentration were measured using a NanoDrop ND-1000 and a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA). The material stored in Qiazol was heated at 65°C for 5 min and RNA was extracted using the RNeasy Lipid Tissue Kit including the optional DNase digestion (Qiagen). RNA quality and concentration were assessed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), and a NanoDrop ND-1000 (Thermo Fisher Scientific).

Whole genome mate pair sequencing for detection of structural variations

To detect structural chromosomal abnormalities, mate pair libraries were prepared for sequencing using the Nextera Mate Pair Library Preparation Kit (Illumina, San Diego, CA). This was done according to the manufacturer's instructions except for the number of shearing cycles, which were increased to three cycles. Paired-end 76 base pair reads were generated using an Illumina NextSeq 500 sequencing instrument. Sequencing depth was on average 3.2x (mapping coverage 2.13x) and the mean insert size was 3.0 kb, resulting in a median spanning coverage of 63.2x of the human genome (mean 63.1x, range 5.2x-119.1x). All samples were sequenced with high quality and yield; between 12.4 and 115.5 million read pairs were obtained per sample and the average quality scores were 31.3-34.1. Sequencing reads were trimmed using NxTrim v 0.4.2 and subsequently aligned against the GRCh37/hg19 build using the Borrows-Wheeler Aligner v 0.7.15[32]. To identify structural rearrangements, the sequence data were analyzed using Integrative Genomics Viewer[33,34], as well as the structural variant callers TIDDIT v 2.2.6, Delly2 v 0.7.8 and Manta v 1.2.2[35-37]. Structural alterations were considered true when identified by at least two of the three variant callers.

Whole genome paired-end sequencing of multi-sampled osteosarcomas

Whole genome paired-end sequencing was performed using the Agilent SureSelect v3 library preparation kit (Agilent Technologies, Santa Clara, CA). Paired-end 150 base pair reads were generated using an Illumina HiSeq 2500 sequencing instrument. Sequencing depth was on average 13.4x (mapping coverage 14.1x) and the mean insert size was 0.34 kb, resulting in a median spanning coverage of 14.5x of the human genome (mean 14.3x, range 5.2x-40.9x). Sequencing reads were aligned against the GRCh37/hg19 build using the Borrows-Wheeler Aligner v 0.7.15. To identify structural rearrangements, the sequence data were analyzed as described above. It is important to stress that whole genome paired-end sequencing is a less optimal technique to detect structural variations, compared with mate pair sequencing, and therefore requires a higher sequencing depth. The reason is the higher spanning coverage of the human genome obtained by mate pair sequencing, due to the analyzed DNA fragments being approximately one order of a magnitude larger. In the present study, the median spanning coverage for mate pair data was 63.2x compared to 14.5x for paired-end data.

Genome-wide DNA copy number and loss of heterozygosity analyses

SNP array analysis was used for combined DNA copy number and loss of heterozygosity investigation. DNA was extracted according to standard procedures from fresh frozen tumor biopsies and hybridized to CytoScan HD arrays, following protocols supplied by the manufacturer (Thermo Fisher Scientific). Somatic copy number alterations in a proportion of the cases were published by Smida *et al.* 2017[5]. Data analysis was performed using the

Chromosome Analysis Suite v 3.3.0.139 (Thermo Fisher Scientific), detecting imbalances by visual inspection, and by segmenting $\log_2$ values using the R package 'copynumber', available via Bioconductor. The inbuilt pcf function was used with a strict gamma value of 160 to create copy number segments and the plotFreq function was used to create the frequency plot of losses and gains on chromosome 17. The threshold for gain was set as a $\log_2$ value of 0.18 and the threshold for loss as -0.18. SNP positions were based on the GRCh37/hg19 sequence assembly. '*TP53* promoter gain' is defined as copy number loss, or copy number neutral loss of heterozygosity, of whole or parts of the *TP53* coding region coupled to concurrent relative copy number gain of the *TP53* promoter region along with regions of the proximal part of chromosome arm 17p.

Visualization of structural and copy number variations using circos plots

Circos plots were generated using the R package 'RCircos', by integrating genomic copy number data obtained from either SNP array analysis or whole genome sequencing and structural variant data based on whole genome sequencing and the TIDDIT algorithm described above. Copy number segments based on SNP array data were generated as described above with the exception of a less stringent gamma value, which was set to 12. Copy number segments based on sequencing data were generated using CNVkit[38].

Whole genome low-pass sequencing of single cells

Whole genome sequencing of cryopreserved primary osteosarcoma cells was performed as described in detail previously[39]. In brief, library preparation was performed using a modified

single cell whole genome sequencing protocol and 77 base pair single reads were generated using a NextSeq 500 sequencing instrument (Illumina). From each assessed tumor, 93 individual cells were sequenced at an average depth of 0.01x. Copy number analysis was performed using AneuFinder[40], and bin positions were based on the GRCh38/hg38 sequence assembly.

RNA sequencing for detection of gene fusions and expression levels

Total RNA was enriched for polyadenylated RNA using magnetic oligo(dT) beads. Enriched RNA was prepared for sequencing using the TruSeq RNA Sample Preparation Kit v2 according to the manufacturer's protocol (Illumina). Paired-end 151 base pair reads were generated from the cDNA libraries using an Illumina NextSeq 500 instrument. Sequencing reads were aligned to the GRCh37/hg19 build using STAR v 2.5.2b[41]. For comparison of relative gene expression levels, data were normalized using Cufflinks with default settings[42], and visualized using the Qlucore Omics Explorer version 3.5 (Qlucore AB, Lund, Sweden). FusionCatcher v 1.0 and STAR-Fusion v 1.4.0 were used to identify candidate fusion transcripts from the sequence data[43].

PCR and Sanger sequencing

Genomic PCR, RT-PCR and nested RT-PCR for detection of the *TP53-ROR2, TP53-SUZ12, TP53-NDEL1*, and *TP53-RTBDN* gene fusions were performed as previously described[44]. Amplified fragments were purified from an agarose gel and Sanger sequencing was performed by GATC

Biotech (Eurofins Genomics, Ebersberg, Germany). BLASTN software (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used for the analysis of sequence data.

Cell model to determine *TP53-ROR2* responsiveness to DNA damage

A promoter-less vector (pSMPUW Universal Lentiviral Expression Vector, Cell Biolabs, Inc., San Diego, CA) containing the *TP53-ROR2* fusion was constructed (GenScript, Piscataway, NJ). The *TP53* promoter was represented by the first 2000 bp upstream of *TP53* together with exon 1 and the first 500 bp of intron 1 of *TP53*. These *TP53* sequences were fused to the last 500 bp of *ROR2* intron 1 and the coding sequences of *ROR2* exons 2-13. This hybrid sequence is denoted *TP53-ROR2* and thus contains the complete coding sequence of *ROR2* transcript variant 002 (ENST00000375715.1) under the control of the *TP53* promoter. A vector containing the same *ROR2* sequences but lacking *TP53* sequences was used as control. CRISPR-mediated knockout of *TP53* in BJ-5ta cells was performed as described elsewhere[45]. In brief, hCas9 and a guide RNA for *TP53* exon 6 were transduced into *TERT*-immortalized human foreskin fibroblast BJ-5ta cells (ATCC, LGC Standards, Middelsex, UK). The BJ-5ta cell line was used in the experiments immediately after purchase and was tested negative for mycoplasma. Antibiotic resistance-selected cells were single cell cloned and analyzed for mutations with the Surveyor mutation detection kit (Integrated DNA Technologies, Inc., Coralville, IA). Clones with detected mutations were validated for homozygous or compound heterozygous mutations with Sanger sequencing or Nextera sequencing (Illumina). This confirmed a 19 bp deletion in *TP53* exon 6 in one clone. Large genomic copy number alterations in this clone were investigated by CytoScan HD array analysis (Thermo Fisher Scientific), revealing a hemizygous deletion of proximal 17p, with a break in *WRAP53*, in all cells. Thus, one *TP53*

allele was deleted and the remaining allele harbored a frame-shift mutation, resulting in complete knockout of this gene. This BJ-5ta *TP53*$^{-/-}$ clone was transduced with the *TP53-ROR2* and *ROR2* vectors, respectively. Cell cultures were exposed to the DNA damaging agent cisplatin at concentrations ranging from 1-5 µM. Cells were harvested for RNA extraction four days following cisplatin treatment. The relative expression levels of *TP53* (Hs01034254_g1) and *ROR2* (Hs00896174_m1) were investigated using RT-qPCR and TaqMan Gene Expression assays (Thermo Fisher Scientific). The *TBP* (Hs99999910_m1) gene was used as an endogenous control. Calculations were performed using the comparative *Ct* method (i.e., ΔΔ*Ct*). The experiment was performed in biological triplicates with each replicate including technical triplicates per sample. Samples were assayed on a 7500 RT-PCR system (Thermo Fisher Scientific).

Cell model to analyze gene regulatory networks orchestrated by *TP53*

BJ-5ta wild type cells, BJ-5ta hCas9 positive cells transduced with guide RNA empty vector control (gEV), BJ-5ta *TP53*$^{-/-}$ cells, and BJ-5ta *TP53*$^{-/-}$ cells transduced with the *TP53-ROR2* vector, described above, were cultured and harvested for RNA extraction. RNA was sequenced and analyzed as described above. Unsupervised correlation-based principal component analysis was performed using the Qlucore Omics Explorer. The first three principal components are displayed, and each sample is connected with its *k* nearest neighbors. Individual gene expression levels are displayed using box plots. The box ranges from the 25th to the 75th percentile. The dotted line represents the 50th percentile (the median). The box whiskers are set at the lowest data point value still within 1.5 times the box range of the lower box limit, and at the highest data point value still within 1.5 times the box range of the upper

box limit. Outliers, represented by circles, are defined as data point values falling outside of the box whisker limits.

Statistical calculations

Statistical calculations were performed using the two-tailed Mann-Whitney *U* test or the Student's *t* test.

Data availability

Sequencing data have been deposited at the European Genome-phenome Archive (EGA) under the accession number EGAS00001003842.

Acknowledgements

We thank The Center for Translational Genomics at Lund University for technical support. This work was supported by the Swedish Childhood Cancer Fund, the Swedish Cancer Society, the Swedish Research Council, the Faculty of Medicine at Lund University, the Åke Wiberg Foundation, the Royal Physiographic Society (Lund, Sweden), and the Crafoord Foundation. D.B. and M.K. were supported by the Swiss National Science Foundation, the Hemmi Stiftung, the Foundation of the Basel Bone Tumour Reference Centre, the Gertrude von Meissner Stiftung, the Susy-Rückert Gedächtnis-Stiftung, the Nora van Meeuwen-Häflinger Stiftung, and the Stiftung für krebskranke Kinder, Regio Basiliensis.

Authors' contributions

K.H.S. and K.H.N. conceived and designed the experiments. S.V., O.B., E.S., M.N. and D.B. contributed tumor material and clinical information. K.H.S., M.K., and L.M. performed DNA extractions and whole-genome sequencing analyses. K.H.S., J.S. and K.H.N. analyzed DNA sequencing data and interpreted the results. K.H.S., M.K., M.B., T.J., D.B. and K.H.N. performed SNP array analyses and interpreted the results. K.H.S., J.N., and L.M performed RNA extractions and RNA sequencing experiments. K.H.S. and K.H.N. carried out bioinformatic analyses of RNA sequencing data and interpreted the results. H.v.d.B., D.C.J.S. and F.F. applied low-pass whole genome sequencing on single cells. K.H.S., L.C., L.M. and J.N. conducted genomic PCR, RT-PCR, Sanger sequencing and RT-qPCR experiments, designed lentiviral vectors and performed the *in vitro* experiments. K.H.S. and K.H.N. prepared the manuscript with contributions from all other authors.

Competing interests

The authors declare no competing interests.

Figure 1

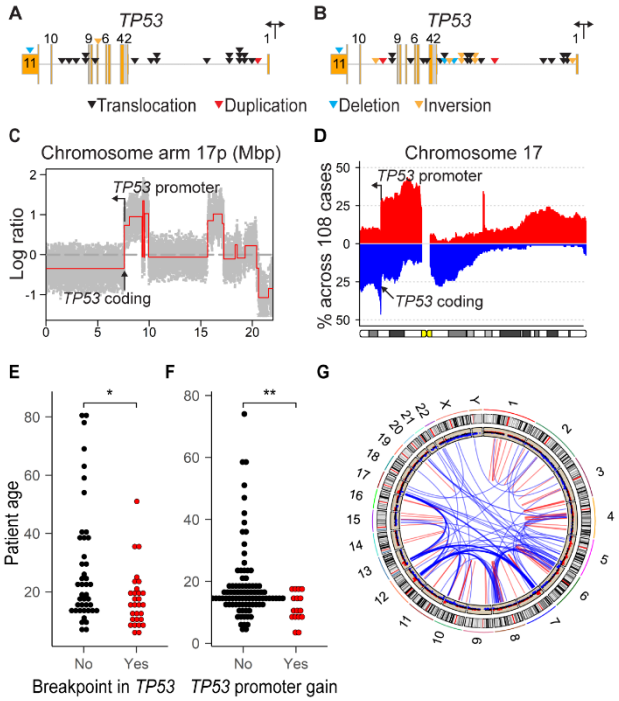


Figure 1. Structural variation in *TP53* is associated with young age at onset and a high number of chromosomal breaks genome-wide. A, Schematic representation of *TP53* structural variation in a discovery (*n*=36) **B**, and a validation osteosarcoma cohort (*n*=36). **C**, DNA copy number profile of 17p in a representative osteosarcoma with gain of the *TP53* promoter region. **D**, Frequency plot of genomic copy number gain (red) and loss (blue) for chromosome 17 across conventional osteosarcomas (*n*=108). **E**, Age distribution of osteosarcoma patients without (*n*=43) and with (*n*=29) *TP53* structural variants as determined by DNA mate pair sequencing. **P* < 0.05, two-tailed Mann-Whitney *U* test. **F**, Age distribution of osteosarcoma patients without (*n*=92) and with (*n*=16) *TP53* promoter gain as determined by SNP array analysis. ***P* < 0.01, two-tailed Mann-Whitney *U* test. **G**, Circos plot showing

genome rearrangements in a representative osteosarcoma with structural variation in *TP53*. Red and blue lines denote intra- and interchromosomal events, respectively.

Figure 2

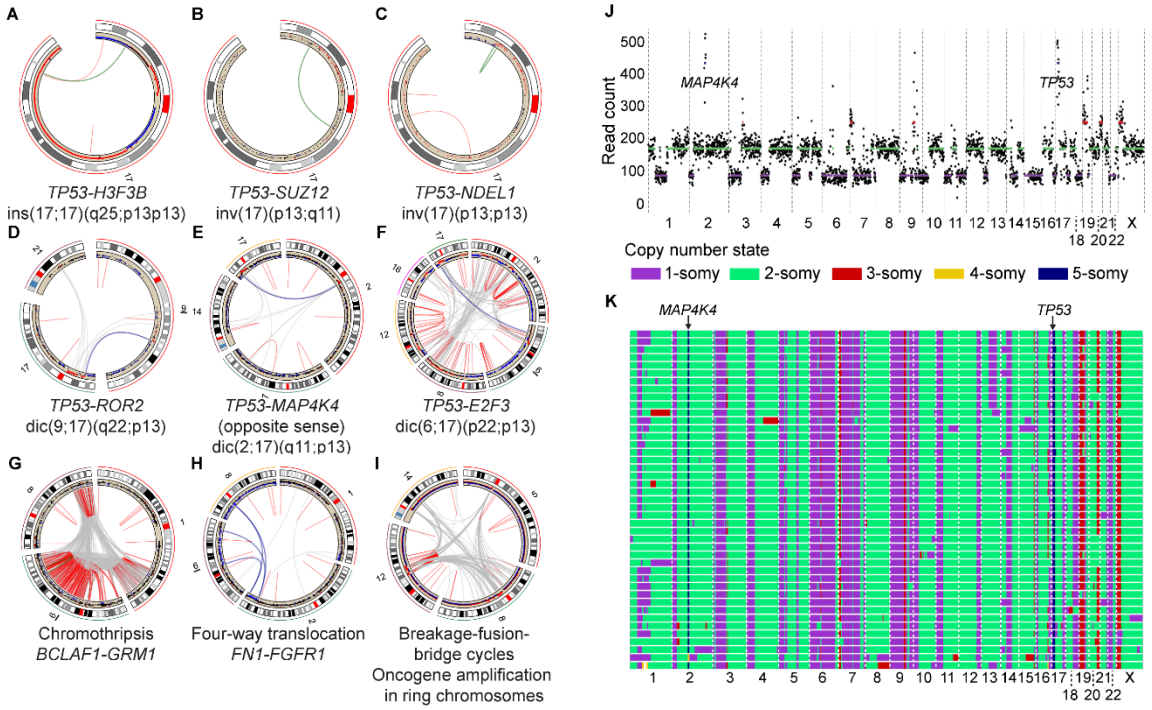


Figure 2. Transposition of the *TP53* promoter is a single early event that can spark genome-wide rearrangements and oncogene amplification. A-C, Intrachromosomal events resulting in *TP53* gene fusions (green lines). **D-F**, Interchromosomal events resulting in *TP53* gene fusions (blue lines). The derivative dicentric chromosomes repeatedly break and rejoin with multiple partner chromosomes. Exemplified are the genomic footprints of **G**, chromothripsis in a chondromyxoid fibroma **H**, a multi-way translocation in a phosphaturic mesenchymal tumor of bone and **I**, breakage-fusion-bridge cycles in a parosteal osteosarcoma. **J**, Genomic copy numbers in a representative individual cell from an osteosarcoma with a *TP53-MAP4K4* fusion. **K**, Heat map of genomic copy numbers across all 43 sequenced individual tumor cells of the *TP53-MAP4K4* fusion positive case. Each row of copy number states represents a single cell.

Figure 3

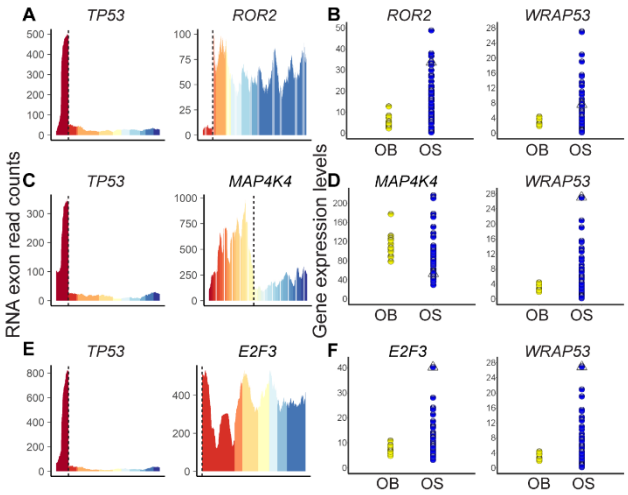


Figure 3. The bidirectional *TP53* promoter induces the expression of *WRAP53* and oncogenes *in vivo*. A, Exon expression levels in Case 9, in which *TP53* intron 1 is fused to *ROR2* exons 2-9. **B**, Normalized gene expression levels. **C**, Exon expression levels in Case 22, in which *TP53* intron 1 is fused to *MAP4K4* exons 1-15, including coding regions for the kinase domain, in the opposite direction. **D**, Normalized gene expression levels, including all exons of *MAP4K4* in Case 22. **E**, Exon expression levels in Case OS046, in which *TP53* intron 1 is fused to regions upstream the complete coding sequence of *E2F3*. **F**, Normalized gene expression levels. Different colors mark individual exons. Dotted lines indicate the fusion points. Triangles mark the case under investigation. OB = osteoblastoma, OS = osteosarcoma.

Figure 4

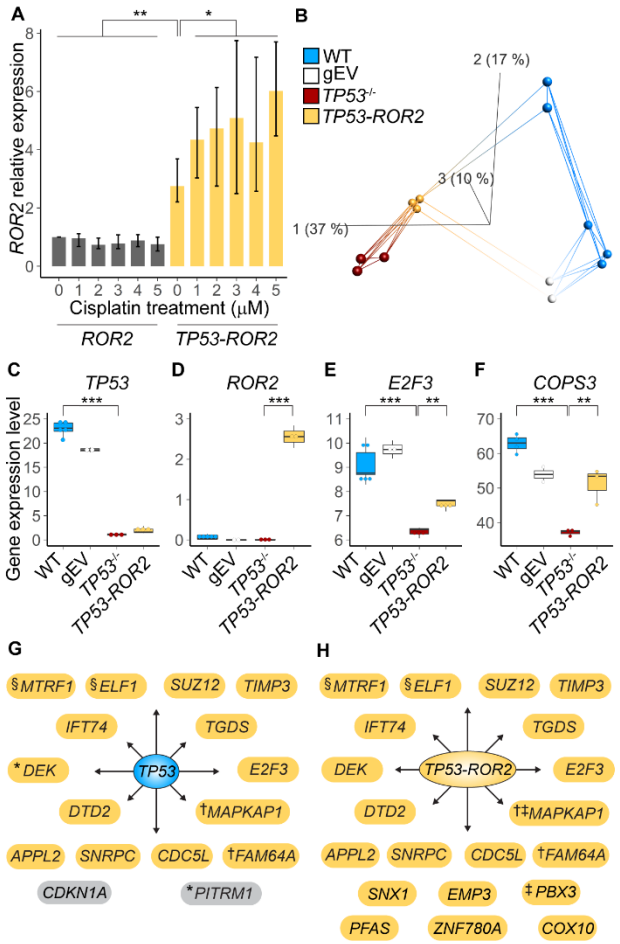


Figure 4. A *TP53* null background constitutively activates the *TP53* promoter, which elicits downstream partner genes that normally are part of a *TP53*-regulated network. A, *ROR2* relative expression levels after cisplatin treatment in BJ-5ta *TP53*$^{-/-}$ cells transduced with either *ROR2*- or *TP53-ROR2*-containing promoter-less vectors. n = 3 biological replicates, mean ± range, *P < 0.05, **P < 0.01, two-tailed Mann-Whitney U test. **B**, Unsupervised principal component analysis based on global gene expression levels in BJ-5ta cells. Each sample is connected with its five nearest neighbors. WT = BJ-5ta wild type cells, gEV = BJ-5ta cells harboring 'guide RNA empty vector', *TP53*$^{-/-}$ = *TP53*$^{-/-}$ BJ-5ta cells, *TP53-ROR2* = *TP53*$^{-/-}$ BJ-5ta

cells harboring *TP53-ROR2*. **C-F**, Gene expression levels for representative genes in BJ-5ta cells. $^{**}P < 0.01$, $^{***}P < 0.001$, Student's *t* test. **G**, *TP53* promoter partner genes with significantly reduced expression levels in BJ-5ta *TP53*$^{-/-}$ compared with BJ-5ta wild type cells. **H**, *TP53* promoter partner genes significantly induced by the *TP53-ROR2* fusion in BJ-5ta *TP53*$^{-/-}$ cells. *[†]Partner genes of different *TP53* fusions in the same tumor, and [‡§]different partner genes under the same *TP53* promoter are marked.

References

1 Gianferante, D. M., Mirabello, L. & Savage, S. A. Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy. *Nat Rev Endocrinol* **13**, 480-491 (2017).

2 Kovac, M. *et al.* Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. *Nat Commun* **6**, 8940 (2015).

3 Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep* **7**, 104-112 (2014).

4 Lorenz, S. *et al.* Unscrambling the genomic chaos of osteosarcoma reveals extensive transcript fusion, recurrent rearrangements and frequent novel TP53 aberrations. *Oncotarget* **7**, 5273-5288 (2016).

5 Smida, J. *et al.* Genome-wide analysis of somatic copy number alterations and chromosomal breakages in osteosarcoma. *Int J Cancer* **141**, 816-828 (2017).

6 Behjati, S. *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun* **8**, 15936 (2017).

7 Negri, G. L. *et al.* Integrative genomic analysis of matched primary and metastatic pediatric osteosarcoma. *J Pathol* **249**, 319-331 (2019).

8 Nord, K. H. *et al. GRM1* is upregulated through gene fusion and promoter swapping in chondromyxoid fibroma. *Nat Genet* **46**, 474-477 (2014).

9 Oliveira, A. M. *et al.* USP6 (Tre2) fusion oncogenes in aneurysmal bone cyst. *Cancer Res* **64**, 1920-1923 (2004).

10 Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* **15**, 371-381 (2015).

11 Gisselsson, D. *et al.* Telomere dysfunction triggers extensive DNA fragmentation and evolution of complex chromosome abnormalities in human malignant tumors. *Proc Natl Acad Sci U S A* **98**, 12683-12688 (2001).

12 Gisselsson, D. *et al.* Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc Natl Acad Sci U S A* **97**, 5357-5362 (2000).

13 Anderson, N. D. *et al.* Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* **361** (2018).

14 Polson, A. & Reisman, D. The bidirectional p53-Wrap53beta promoter is controlled by common cis- and trans-regulatory elements. *Gene* **538**, 138-149 (2014).

15 Miled, C., Pontoglio, M., Garbay, S., Yaniv, M. & Weitzman, J. B. A genomic map of p53 binding sites identifies novel p53 targets involved in an apoptotic network. *Cancer Res* **65**, 5096-5104 (2005).

16 Li, Y. *et al.* Genome-wide analysis of the p53 gene regulatory network in the developing mouse kidney. *Physiol Genomics* **45**, 948-964 (2013).

17 Galanos, P. *et al.* Chronic p53-independent p21 expression causes genomic instability by deregulating replication licensing. *Nat Cell Biol* **18**, 777-789 (2016).

18 Georgakilas, A. G., Martin, O. A. & Bonner, W. M. p21: A Two-Faced Genome Guardian. *Trends Mol Med* **23**, 310-319 (2017).

19 Miller, C. W. *et al.* Frequency and structure of p53 rearrangements in human osteosarcoma. *Cancer Res* **50**, 7950-7954 (1990).

20 Masuda, H., Miller, C., Koeffler, H. P., Battifora, H. & Cline, M. J. Rearrangement of the p53 gene in human osteogenic sarcomas. *Proc Natl Acad Sci U S A* **84**, 7716-7719 (1987).

21 Kaur, S., Larramendy, M. L., Vauhkonen, H., Böhling, T. & Knuutila, S. Loss of *TP53* in sarcomas with 17p12 to approximately p11 gain. A fine-resolution oligonucleotide array comparative genomic hybridization study. *Cytogenet Genome Res* **116**, 153-157 (2007).

22 Kelman, Z. *et al.* Rearrangements in the p53 gene in Philadelphia chromosome positive chronic myelogenous leukemia. *Blood* **74**, 2318-2324 (1989).

23 Mashal, R. *et al.* Rearrangement and expression of p53 in the chronic phase and blast crisis of chronic myelogenous leukemia. *Blood* **75**, 180-189 (1990).

24 Hernandez, A. *et al.* p53 gene rearrangements in chronic myelocytic leukemia. *Ann Hematol* **66**, 81-83 (1993).

25 George, J. *et al.* Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**, 47-53 (2015).

26 Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297-303 (2017).

27 Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359-364 (2017).

28 Sirohi, D. *et al. TP53* structural variants in metastatic prostatic carcinoma. *PLoS One* **14**, e0218618 (2019).

29 Quigley, D. A. *et al.* Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769 e759 (2018).

30 Mitelman, F., Johansson, B. & Mertens, F. (Eds.) Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2020). https://mitelmandatabase.isb-cgc.org

31 Tannock, I. F. & Hickman, J. A. Limits to Personalized Cancer Medicine. *N Engl J Med* **375**, 1289-1294 (2016).

32 Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-2851 (2014).

33 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).

34 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).

35 Eisfeldt, J., Vezzi, F., Olason, P., Nilsson, D. & Lindstrand, A. TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. *F1000Res* **6**, 664 (2017).

36 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339 (2012).

37 Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222 (2016).

38 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).

39 van den Bos, H. *et al.* Quantification of Aneuploidy in Mammalian Systems. *Methods Mol Biol* **1896**, 159-190 (2019).

40 Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. *Genome Biol* **17**, 115 (2016).

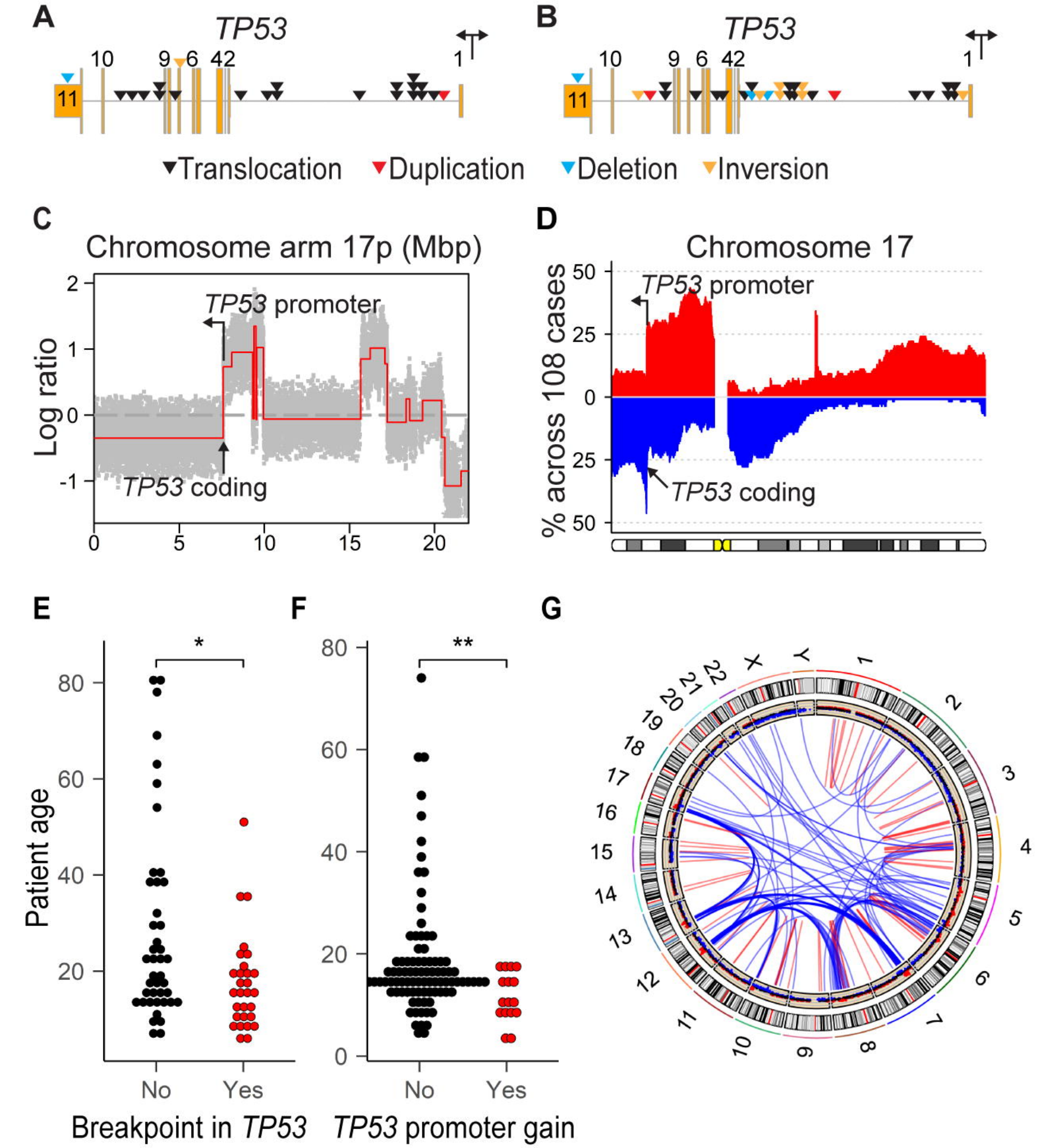
41 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

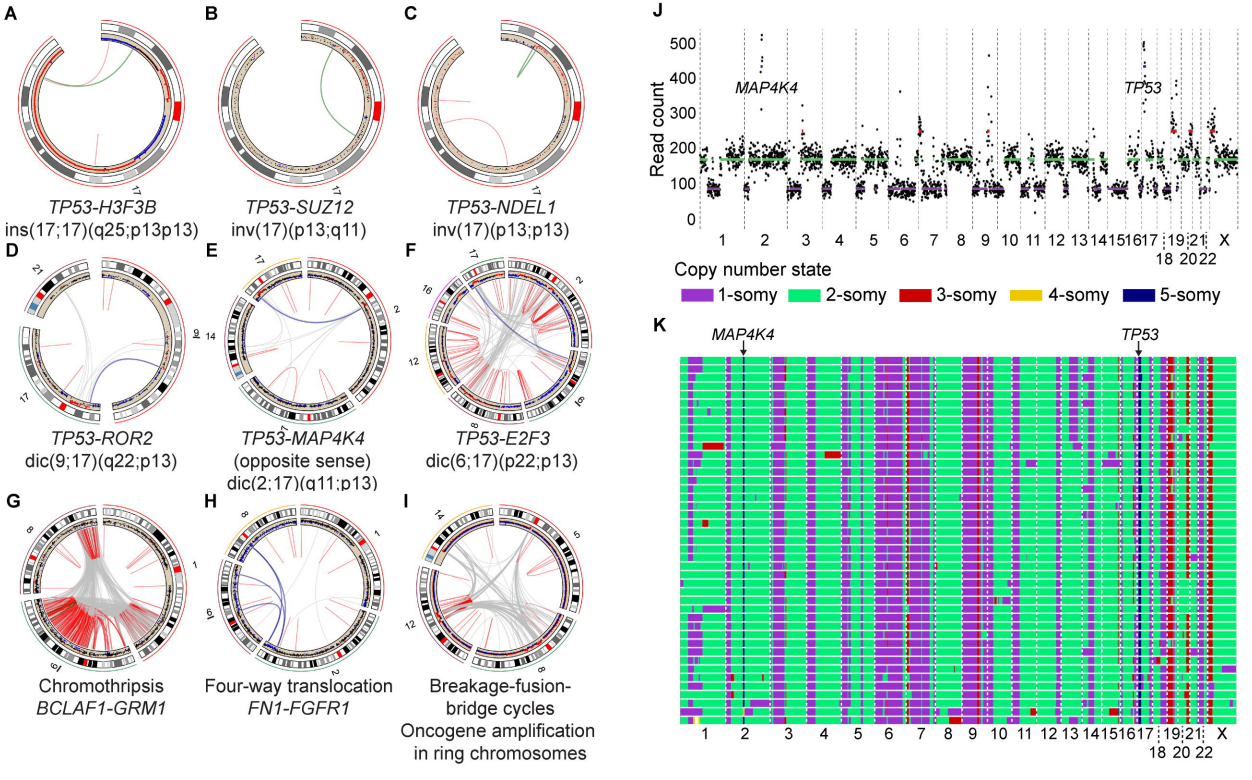
42 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515 (2010).

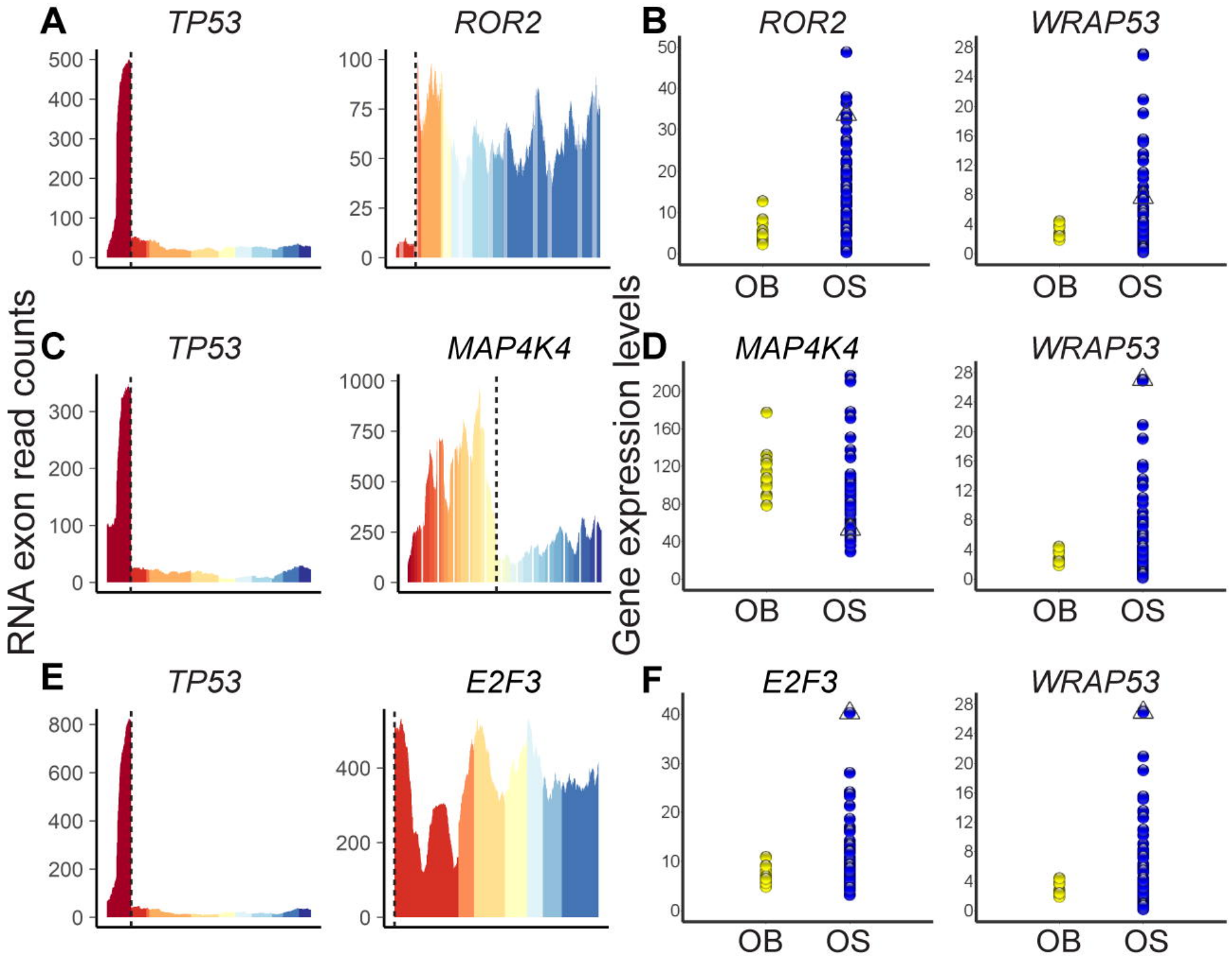
43 Haas, B. J. *et al.* Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol* **20**, 213 (2019).

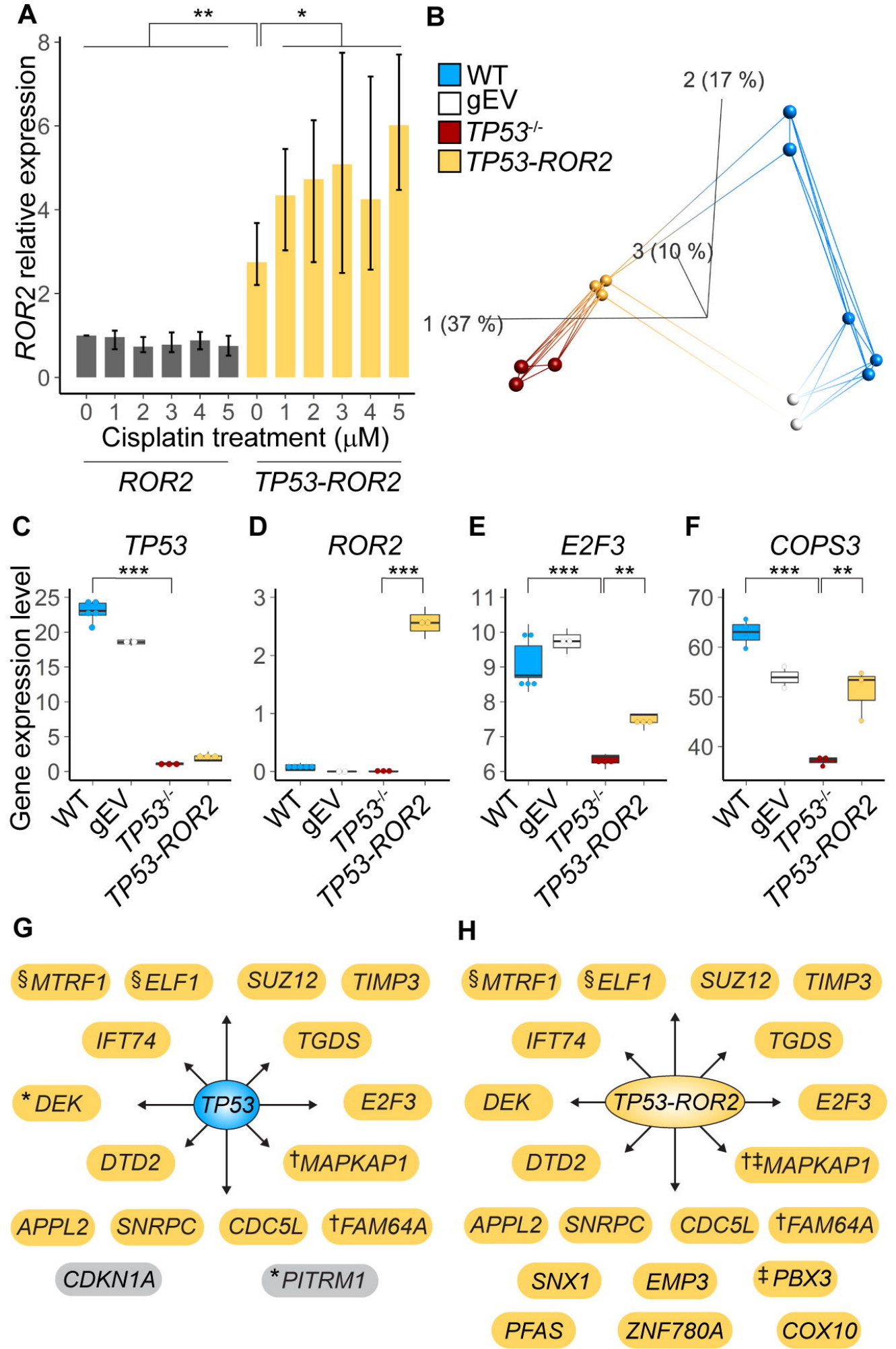
44 Saba, K. H. *et al.* Genetic profiling of a chondroblastoma-like osteosarcoma/malignant phosphaturic mesenchymal tumor of bone reveals a homozygous deletion of *CDKN2A*, intragenic deletion of *DMD*, and a targetable *FN1-FGFR1* gene fusion. *Genes Chromosomes Cancer* **58**, 731-736 (2019).

45 Karlsson, J. *et al.* Four evolutionary trajectories underlie genetic intratumoral variation in childhood cancer. *Nat Genet* **50**, 944-950 (2018).


A
TP53
10
9
6
42
1
11
B
TP53
10
9
6
42
1
11
Translocation
Duplication
Deletion
Inversion
C
Chromosome arm 17p (Mbp)
Log ratio
TP53 promoter
TP53 coding
D
Chromosome 17
% across 108 cases
TP53 promoter
TP53 coding
E
Patient age
No
Yes
Breakpoint in TP53
F
No
Yes
TP53 promoter gain
G


A
B
C
TP53-H3F3B
ins(17;17)(q25;p13p13)
TP53-SUZ12
inv(17)(p13;q11)
TP53-NDEL1
inv(17)(p13;p13)
D
E
F
TP53-ROR2
dic(9;17)(q22;p13)
TP53-MAP4K4
(opposite sense)
dic(2;17)(q11;p13)
TP53-E2F3
dic(6;17)(p22;p13)
G
H
I
Chromothripsis
BCLAF1-GRM1
Four-way translocation
FN1-FGFR1
Breakage-fusion-
bridge cycles
Oncogene amplification
in ring chromosomes
J
Read count
500
400
300
200
100
0
MAP4K4
TP53
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Copy number state
1-somy
2-somy
3-somy
4-somy
5-somy
K
MAP4K4
TP53
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X


A
TP53
ROR2
B
ROR2
WRAP53
C
TP53
MAP4K4
D
MAP4K4
WRAP53
E
TP53
E2F3
F
E2F3
WRAP53
RNA exon read counts
Gene expression levels
OB
OS


A
**
*
ROR2 relative expression
0
2
4
6
8
0 1 2 3 4 5 0 1 2 3 4 5
Cisplatin treatment (μM)
ROR2
TP53-ROR2
B
WT
gEV
TP53-/-
TP53-ROR2
2 (17 %)
3 (10 %)
1 (37 %)
C
TP53
***
Gene expression level
0
5
10
15
20
25
WT
gEV
TP53-/-
TP53-ROR2
D
ROR2
***
0
1
2
3
WT
gEV
TP53-/-
TP53-ROR2
E
E2F3
***
**
6
7
8
9
10
11
WT
gEV
TP53-/-
TP53-ROR2
F
COPS3
***
**
40
50
60
70
WT
gEV
TP53-/-
TP53-ROR2
G
§MTRF1
§ELF1
SUZ12
TIMP3
IFT74
TGDS
*DEK
TP53
E2F3
DTD2
†MAPKAP1
APPL2
SNRPC
CDC5L
†FAM64A
CDKN1A
*PITRM1
H
§MTRF1
§ELF1
SUZ12
TIMP3
IFT74
TGDS
DEK
TP53-ROR2
E2F3
DTD2
†‡MAPKAP1
APPL2
SNRPC
CDC5L
†FAM64A
SNX1
EMP3
‡PBX3
PFAS
ZNF780A
COX10