Genome-wide analysis of somatic copy number alterations and chromosomal breakages in osteosarcoma

Short title: Somatic copy number alterations in osteosarcoma

Jan Smida[1,2,3#©], Hongen Xu[4#], Yanping Zhang[4#], Daniel Baumhoer[5], Sebastian Ribi[5], Michal Kovac[5], Irene von Luettichau[3], Stefan Bielack[6], Valerie B. O'Leary[1], Christine Leib-Mösch[7], Dmitrij Frishman[4,8,9§©] , Michaela Nathrath [2,3,10§]

[1]Institute of Radiation Biology, Helmholtz Zentrum Munich - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany;

[2]Clinical Cooperation Group Osteosarcoma, Helmholtz Zentrum Munich - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany;

[3]Pediatric Oncology Center, Department of Pediatrics, Technical University of Munich and Comprehensive Cancer Center, Koelner Platz 1, 80804 Munich, Germany;

[4]Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technical University of Munich, Maximus-von-Imhof-Forum 3, 85354 Freising, Germany;

[5]Bone Tumour Reference Center, Institute of Pathology, University Hospital Basel, Schoenbeinstrasse 40, 4031 Basel, Switzerland;

[6]Pediatrics 5 (Oncology, Hematology, Immunology), Klinikum Stuttgart Olgahospital, Kriegsbergstrasse 62, 70174 Stuttgart, Germany;

[7]Institute of Virology, Helmholtz Zentrum Munich - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany;

[8]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum Munich - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, 85764 Neuherberg, Germany;

[9]St Petersburg State Polytechnic University, Polytekhnicheskaya, 29, 195251 St Petersburg, Russia;

[10]Department of Pediatric Hematology and Oncology, Klinikum Kassel, 34125 Kassel, Germany.

The authors wish it to be known that, in their opinion, these authors should be regarded as joint First Authors.

§ Shared senior authorship

© To whom correspondence should be addressed. Dmitrij Frishman, Tel: +49 8161712134; Fax: +49 8161712186; Email: d.frishman@wzw.tum.de. Correspondence may also be addressed to Jan Smida, Tel: +49 8931873797; Fax: +49 8931873381; Email: smida@helmholtz-muenchen.de.

BAF: B allele frequency; COSMIC: Catalogue of Somatic Mutations in Cancer; CTLP: Chromothripsis Like Pattern; CytoScan HD: CytoScan High Density; DSBs: Double Strand Breaks; G4: G-quadruplexes; GISTIC: Genomic Identification of Significant Targets In Cancer; Indels: Insertions/Deletions; LRR: Log R Ratio; LTR: Long Terminal Repeat; OS: Osteosarcoma; SCNAs: Somatic Copy Number Alterations; SNP: Single Nucleotide Polymorphism; SNP-FASST2: SNP-Fast Adaptive States Segmentation Technique 2; UCSC: University of California, Santa Cruz.

Article category:

Tumor Markers and Signatures

Novelty and Impact:

A comprehensive assessment of somatic copy number alterations (SCNAs) was performed using whole-genome CytoScan HD arrays in 160 osteosarcoma (OS) samples. Genes or regions frequently targeted by SCNAs were identified. Breakage analysis revealed OS specific unstable regions in which well-known OS tumor suppressor genes, including *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP* are located. A complex breakage pattern – chromothripsis – was suggested as a widespread phenomenon in OS and is predictive of OS patient clinical outcome.

Financial disclosure [1]

Conflict of interests [2]

---

[2] The authors declare that they have no competing interests.

Abstract

Osteosarcoma (OS) is the most common primary malignant bone tumor in children and adolescents. It is characterized by highly complex karyotypes with structural and numerical chromosomal alterations. The observed OS-specific characteristics in localization and frequencies of chromosomal breakages strongly implicate a specific set of responsible driver genes or a specific mechanism of fragility induction. In this study, a comprehensive assessment of somatic copy number alterations (SCNAs) was performed in 160 OS samples using whole-genome CytoScan High Density arrays (Affymetrix, Santa Clara, CA). Genes or regions frequently targeted by SCNAs were identified. Breakage analysis revealed OS specific unstable regions in which well-known OS tumor suppressor genes, including *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP* are located. Certain genomic features, such as transposable elements and non-B DNA-forming motifs were found to be significantly enriched in the vicinity of chromosomal breakage sites. A complex breakage pattern – chromothripsis – has been suggested as a widespread phenomenon in OS. It was further demonstrated that hyperploidy and in particular chromothripsis were strongly correlated with OS patient clinical outcome. The revealed OS-specific fragility pattern provides novel clues for understanding the biology of osteosarcoma.

Introduction

Osteosarcoma (OS) is the most common primary malignant bone tumor in adolescents and young adults [1, 2]. It is characterized by a complex karyotype with a high degree of aneuploidy and numerous structural aberrations such as somatic copy number alterations (SCNAs) and genomic rearrangements [3-5]. Curative treatment of OS is based on multi-agent chemotherapy in addition to complete surgery. For patients with localized extremity disease 10-year event-free survival rates reach approximately 60% [6], but have plateaued during the past decades. Further improvement in cure rates will most likely depend on an increased knowledge about the underlying molecular mechanisms of this disease.

Although several predictors, such as gene expression profiles [7] and chromosomal alteration staging systems [4] have been proposed to anticipate tumor response to chemotherapy, common markers of prognostic and therapeutic value remain to be identified. Genomic instability, a hallmark of most cancers, including OS [8, 9], is either driven by positive selection or originates from sequence-specific unstable regions [8]. Chromosomal fragile sites are specific genomic locations that appear as gaps or breaks on metaphase chromosomes under replication stress [10]. This can be induced by endogenous or exogenous sources, and result in the generation of DNA double strand breaks (DSBs) and genomic instability [11]. A variety of molecular pathways are involved in DSB repair, and, in the case of deficient repair, copy number alterations result.

To identify SCNAs, array-based copy number profiling has been utilized as an alternative to next generation sequencing due to its lower consumption of precious biopsy material. DNA copy number profiling was generally opted for over gene expression, as it provided relatively stable profiles enabling differentiation of clinically relevant genetic subgroups [12]. However, the analysis of whole genome array data for tumor samples can be difficult due to the fact that the total DNA amount in a cancer cell can differ significantly from a diploid state, and tumor tissues often contain some proportion of normal cells [13]. SCNAs have the potential to inactivate tumor suppressor genes or activate oncogenes, and consequently play fundamental roles in gene regulation and pathobiological processes in cancer [14]. Analyses of SCNA

4

data generated in recent years have provided insights into driver genes for many tumor types [14, 15]. However, the enormous complexity of genomic aberrations in OS has made it challenging to identify recurrent alterations and genes driving tumorigenesis [3, 5]. Furthermore, in OS the identification of driver genes has been hindered by intra- and inter-tumor heterogeneity and limited sample availability [5, 16-18]. Despite such complications, we and others have revealed recurrent genomic loss in regions containing tumor suppressor genes such as *LSAMP*, *CDK2NA*, *RB1*, and *TP53* and most frequent gains at sites including the oncogene *MYC* and the gene *RUNX2* - an important player in osteogenic differentiation [5, 16-19].

Apart from their genomic instability, osteosarcomas show a disease specific SCNA pattern. The phenomenon of chromothripsis represents an important mechanism of carcinogenesis that differs from progressive accumulation of genomic rearrangements. The simultaneous fragmentation of distinct chromosomal regions (breakpoints showing a specific, non-random distribution) and subsequent imperfect reassembly of those fragments leads to a specific SCNA pattern (chromothripsis like pattern, CTLP). The initial discovery indicated that chromothripsis is a widespread phenomenon, which can be seen in 2% - 3% of all cancers, most notably in 25% of bone cancers [20]. There is strong evidence for an association between chromothripsis and poor outcome in different cancer types, including multiple myeloma [21], neuroblastoma [22] and Sonic-Hedgehog medulloblastoma [23]. Although the mechanisms governing chromothripsis are largely unknown, it has important implications for our understanding of cancer and disease [24], as such detailed analyses of chromothripsis-like patterns may shed light on OS development and progression.

Herein, copy number profiles derived from 160 pre-therapeutic osteosarcoma biopsies have been analysed using whole-genome CytoScan High Density (CytoScan HD) arrays (Affymetrix, Santa Clara, CA). Integration of SCNAs for each sample was performed in order to identify potential genes driving OS oncogenesis. Previously found OS driver genes were identified as well as other OS-related genes. Chromosomal breakages were found to be spatially clustered in certain locations, termed 'broken regions', harboring the regarded OS tumor suppressor genes *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP*. Furthermore, chromosomal breakages in these regions occurred early and were influenced by local genomic context. Most noteworthy, both aneuploidy and CTLP occurrence were found to be correlated with

clinical outcome of OS patients.

Methods

Tissue samples and patient characteristics

For CytoScan HD array analysis, a set of 160 fresh-frozen tissue samples derived from pretherapeutic biopsies was used. All biopsies were evaluated by an experienced bone pathologist who confirmed the tumor content to be > 70 % per sample. The patient cohort samples were obtained according to the guidelines and approval of the Research Ethics Board at the Faculty of Medicine of the Technical University of Munich (Technische Universität München, Reference 1867/07) and local ethical committee of Basel, Switzerland (Ethikkommission beider Basel EKBB, www.ekbb.ch, Reference 274/12). The descriptive characteristics of this collection are summarized in Table 1. The vast majority of the investigated samples (n=141) are classified as high-grade osteosarcoma. The patients were treated between 1990 and 2012 according to the protocols of the Cooperative German-Austria-Swiss Osteosarcoma Study Group [25] (reviewed and approved by the appropriate ethics committees) after informed consent was obtained.

SCNA calling, driver gene identification, and tumor subclone decomposition

DNA from frozen osteosarcoma tissue was analysed using the Affymetrix CytoScan HD platform. The raw data are available in the ArrayExpress database [26] under accession number E-MTAB-4815. Nexus copy number software version 7.5 (obtained from BioDiscovery, Inc.) was used to process CEL files. Copy number alterations were determined by the Single Nucleotide Polymorphism Fast Adaptive State Segmentation Technique 2 (SNP-FASST2) algorithm together with a quadratic correction implemented in Nexus. Sample- and chromosome-specific thresholds defining copy number gain, copy number loss, high copy gain, and homozygous copy loss (Supplementary Table 1) were based on true diploid regions in individual tumor samples (performed using Nexus with subsequent manual curation by experts from BioDiscovery, Inc.). SCNAs with fewer than 20 informative probes were excluded from further consideration. GISTIC 2.0 (Genomic Identification of Significant Targets

In Cancer) integrated in the Nexus copy number software was utilised to identify potential driver SCNAs and genes by evaluating the frequency and amplitude of observed events [27].

Subclone structures were reconstructed for each tumor sample based on SCNA calling data from Nexus copy number software. The SubcloneSeeker software [28] was utilized to decompose tumor subclone structures. In this study, a subclone was defined as a collection of cells in the tumor sample that contained the same set of SCNAs. The segmental mean values of each segment generated by SNP-FASST2 was used as input for the SubcloneSeeker software [28] to reconstruct the clonal structures for each patient. The *segtxt2db* and *ssmain* applications were employed to cluster the segments based on their cell prevalence values and to enumerate the clonal structures. The results were exported using the '*treeprint*' utility. We refer to the SCNAs that occurred at the root node of the subclone tree as 'clonal' SCNAs and to all others as 'sub-clonal'.

Definitions of chromosomal breakages and their association with genomic features

We defined genomic starts and ends of SCNAs as SCNA breakpoints although their exact chromosomal positions could not be determined. Breakpoints situated upstream of the first or downstream of the last CytoScan HD probe on the same chromosome as well as those located in telomeres or centromeres were ignored. We defined a genomic position to be a chromosomal break when the Log_2 signal value alteration between two adjacent genomic segments (from centromere to telomere) was > 0.3.

An association was determined between chromosomal breakages and multiple genomic features as obtained from public databases and published studies or as identified in the current study. All genomic coordinates of the features correspond to the human genome assembly hg19 and, when necessary, the University of California, Santa Cruz (UCSC) liftOver tool was used to convert the hg18 coordinates to hg19 [29]. Specifically, chromosomal coordinates for Alu repeats, DNA transposons, L1 and long terminal repeat (LTR) retrotransposons, exons, and conserved elements (the PhyloP46wayPrimates table) were downloaded from UCSC Genome Browser [29]. Non-B DNA motifs were obtained from non-B DB v2.0 [30]. Common fragile sites were found to be tissue- and cell-type specific [31]. As tissue-specific data was not available, we obtained genomic coordinates for common fragile sites and non-fragile regions

7

from a previous study [32]. We defined nucleotide substitution (or insertions/deletions, indels) rate as the ratio of the total number of substitutions (or indels) to the total number of nucleotides in the human-chimpanzee alignments (from UCSC Genome Browser).

The density of SCNA breakpoints, chromosomal breaks or genomic features (*i.e.* item) were defined as the ratio of total base pairs belonging to the item against the total length of the genomic region. The subdivision of the genome, shuffling and feature density calculation were performed using BEDTools [33] and in-house Perl scripts.

Detection of chromothripsis-like patterns in osteosarcoma

To detect chromothripsis-like patterns (CTLPs) the algorithm described in [34] was applied to identify clustering of copy number changes in the genome. Default settings were used except for the parameter of Log_2 signal value difference between two adjacent segments (set to 0.2). CTLP samples were determined by the evidence of the copy number switching its status at least 12 times (SwitchNo $\geq$ 12) and Log_{10} of likelihood ratio greater than 8 (Log_{10} LR $\geq$ 8) within a single chromosome.

Estimation of tumor purity and ploidy

SNP-based DNA microarrays allow simultaneous measurement of the allele-specific copy number at many different SNP loci in the genome. For each probeset, the log R ratio (LRR) reflects the ratio of total signal intensity for both alleles against expected signals, and the B allele frequency (BAF) is an estimate of the relative proportion of one of the alleles with respect to the total signal intensity. LRR and BAF values were derived using the *affy2sv* R package [35] together with the Affymetrix Power Tools. A total of 873 normal samples downloaded from the study [36] (Gene Expression Omnibus accession number: GSE59150) were also processed using *affy2sv*. The resulting LRR and BAF were used as input for the GPHMM algorithm (version 1.4) [37] to obtain an estimation of normal cell contamination and absolute copy number of genomic segments for each sample. Population frequency of the B allele file required for running GPHMM was created using the Perl script compile_pfb.pl in PennCNV [38], with BAF values from the 873 normal samples as input. Another required file - GC model file (GC content flanking SNP markers) - was generated using the Perl script cal_gc_snp.pl in PennCNV [38]. Tumor ploidy was further determined following the

8

protocol described in [39]. Specifically, the chromosome arm count in a tumor genome was estimated based on the absolute copy number of genomic segments in the pericentric region. The copy number of the corresponding arm was set to the absolute copy number of the segments in the pericentric region if its size was ≥ 1.5 Mb. Otherwise, if the size of the pericentric segments was < 1.5 Mb, the copy number of the chromosome arm was approximated by the average copy number of all segments on that chromosome arm. Tumor ploidy was assigned for each tumor sample based on chromosome counts and the DNA index, defined as the average copy number of the tumor genome divided by 2. Tumor ploidy was set at 2 (near-diploid genome) for chromosome counts < 60 and DNA index < 1.3, and set at 4 (near-tetraploid genome) for chromosome counts ≥ 60 and DNA index ≥ 1.3 [40].

Results

Overview of SCNAs in osteosarcoma

The SCNA landscape of pre-treatment tissue samples (n = 160) from osteosarcoma patients (characteristics of whom are provided in Table 1) was profiled using Affymetrix CytoScan HD arrays. Three samples were excluded from copy number analysis due to insufficient data quality. A genome-wide frequency plot of SCNAs is shown in Figure 1. In our collections, the median size of the SCNAs was 1.2 Mb with the OS genome having on average 209 SCNA events. Regional gains and losses of various sizes were observed, ranging from entire chromosomes to minor genomic segments. Many oncogenes and tumor suppressor genes were located within these sites. No significant correlation was noted between the total SCNA number, size, or median in relation to age or gender. An apparent correlation trend was evident for total SCNA size and survival, although perhaps due to insufficient power this did not reach significance.

GISTIC analysis and tumor subclone decomposition uncover key driver genes affected by SCNAs in osteosarcoma

GISTIC 2.0 [27] is a tool to identify genes targeted by SCNAs that may drive cancer development. The X and Y chromosomes were excluded from the analysis and were analysed separately in gender specific subsets of OS patients. GISTIC identified 88

9

regions significantly altered in 157 OS samples (Figure 2; genomic locations of these regions are listed in Supplementary Table 2). The annotation of GISTIC regions revealed 101 targeted genes (listed in Supplementary Table 3), of which the vast majority (74 transcripts) were protein-coding genes. Nine genes listed in the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) [41] - namely *NOTCH2*, *PDGFRA*, *CDK4*, *CCNE1*, and *RUNX1* were located in copy-number gain regions, while *CDKN2A*, *FLI1*, *TP53*, and *ATRX* were identified in copy-number loss regions. *TP53* and *ATRX*, often targeted by SCNAs, have been reported by us and others as important driver genes in OS [16, 42, 43]. Besides these well-known OS driver genes, GISTIC regions contained several other OS-related genes, such as *RUNX2* and *DLG2* [16, 44].

Analysis also revealed novel or recently described genes - *FOXN1* and *WWOX*. *FOXN1* (17q11.2) is the main transcriptional regulator of thymic epithelial cell development, differentiation, and function [45]. Although it directly or indirectly regulates expression of a broad variety of genes, it has not been found to date associated with cancer and, in particular OS. The *WWOX* gene (16q23.1) spans a common fragile site FRA16D, associated with DNA instability in cancer [46]. Recently, a series of reports demonstrated the relevance of reduced or absent *WWOX* expression in various cancer types, including OS, presumably due to chromosomal deletions and translocations within the *WWOX* gene, highlighting an essential role for *WWOX* in tumor suppression and genomic stability [47-49]. Besides the tumor suppressor and pro-apoptotic activity of *WWOX* in OS, its role in osteogenic differentiation and interaction with *RUNX2* has recently been elucidated [50].

A malignant tumor often consists of genetically distinct cell populations, referred to as tumor sub-clones, with each possessing a specific mutation subset. Determination of the order in which SCNA mutations occur is a powerful means for identifying genes with fundamental roles in oncogenesis. SubcloneSeeker [28] succeeded in inferring subclone structures for 99.4 % of tumors (156 out of 157). The mean number of predicted subclone structures for each tumor was 8.5 (ranging from 1 to 45, Supplementary Table 4). Thirty-six tumors had greater than 10 possible subclone structures, which may be due to the complex nature of such tumor samples. Next, an investigation was undertaken as to whether or not SCNAs overlapping with putative genes (identified by GISTIC) were clonal events. Previously reported findings as

revealed by alternative approaches were confirmed, to show that even for the well-known OS driver genes such as *TP53* and *RB1*, the majority (~90%) of SCNAs were subclonal events [43]. Thirty-four tumors had clonal SCNAs overlapping one to ten driver genes, such as *TP53*, *RB1*, *DLG2*, *WWOX*, *TERT*, *FOXN1*, *APC*, *PTEN*, *LSAMP*, *ATRX*, and *CDKN2A*. No single gene had clonal SCNAs in the majority of tumors.

Breakage analyses reveal osteosarcoma-specific unstable regions

DNA breakage is a prerequisite for cancer-associated genomic aberrations, including amplifications, deletions, inversions, and translocations. The genomic start and end of SCNAs were defined as breakpoints with a precision of ~ 1 kb (average inter-probe distance for CytoScan HD Array is < 1 kb). Since whole genome arrays have reduced ability for inversion and/or translocation detection, the chromosomal breakage landscape was investigated, which strongly indicated the prevalence of genomic rearrangements. The criterion for considering a SCNA breakpoint as a chromosomal break was based on the Log_2 signal value alteration between two adjacent genomic segments > 0.3 (Supplementary Figure S1), which is more stringent than the cutoff of 0.23 previously used [51]. In total, 62,172 SCNA breakpoints and 19,810 chromosomal breaks were identified in 157 OS samples. The number of chromosomal breaks per sample ranged from 17 to 425, with a median value of 114. The number of breaks per mega base ranged from 4 (chromosome 2) to 14 (chromosome 17). In order to further examine the landscape of chromosomal breaks across different chromosomes, each chromosome was divided into non-overlapping 1 Mb regions following gap exclusion in the genome assembly and the density of chromosomal breaks per block calculated. Results showed that 2% of genomic regions (61/3060, Supplementary Table 5) were significantly enriched for chromosomal breaks (Bonferroni corrected *P*-values < 0.1). Out of these 'broken regions', 11% are located within common fragile sites, while 46% overlapped with non-fragile sites [32], indicating apparent OS-specific instability characteristics.

Some of the OS-associated tumor suppressor genes [19], including *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP*, but no known OS oncogenes, were located in these broken regions (Figure 3). To determine the evolutionary order in which SCNAs occurred in these areas, a comparison was made with clonal SCNAs obtained by the SubcloneSeeker analysis. An enrichment of clonal SCNAs was found in these broken

regions compared to randomly generated ones (10662 *vs* 4579, *P*-value=0), implicating chromosomal breakage as a clonal event of early occurrence in tumorigenesis.

In order to identify genes prone to breakage in OS, we compared the distribution of actual chromosomal breaks to a background distribution obtained by shuffling the position of chromosomal breaks 1,000 times. This approach, while limited by a degree of uncertainty in calling the location of chromosomal breaks (due to the inter-probe distance characteristic of CytoScan HD arrays), can nevertheless provide clues as to which genes are prone to breakage in OS. A total of 343 genes were found to harbor chromosomal breaks significantly more frequently than would be expected by chance (Bonferroni corrected *P*-values < 0.01). Of these, 24 genes (listed in Table 2) have been previously shown to be associated with OS (*DLG2*, *WWOX*, *TP53*, *RB1*, *LSAMP*, *PTEN*, and *APC* [19]) and other tumors (*DMD*, *EYA1*, *SCAPER*, *WNK1*, *KANSL1*, *TP63*, *FOXN1*, and *CHM*) and found by GISTIC analysis. *TP53* was selected to demonstrate the distribution of chromosomal breaks along the gene. As seen in Supplementary Figure S2 the largest number of chromosomal breaks was located in the first intron of this gene [16, 42].

Chromosomal breakage in osteosarcoma is dependent on local genomic context

To examine whether chromosomal breakages in OS were associated with the local genomic context, we investigated the joint distributions of chromosomal breaks, SCNA breakpoints and multiple genomic features within a 1Mb genomic window. Previous studies have shown that DNA breakage can be induced by DNA structures such as non-B DNA conformations, including Cruciform, G-quadruplexes (G4), Slip, Triplex, and Z-DNA, and by highly homologous genomic repeats, such as L1 and Alu [52-54]. Further features considered in this analysis were common fragile sites, evolutionarily conserved elements, substitution rate, indel rate and exon density which have been associated with SCNA breakpoints [53, 55, 56]. As expected, SCNA breakpoints and chromosomal breakage are highly correlated (P-value < 2.20×10^{-16}, Spearman rho = 0.76). In addition, it was also noted that SCNA breakpoints and chromosomal breaks were significantly correlated with diverse genomic properties, including Alu, L1, Cruciform, G4, Slip, Triplex, Z-DNA, exon density, and indel rate (Bonferroni corrected *P*-values<0.01; Supplementary Table 6).

We further examined the association of genomic properties to chromosomal breaks at a higher resolution. Specifically, windows of 10 kb, 20 kb, 50 kb, and 100 kb centred around each chromosomal break were analysed with subsequently merging of overlapped windows. The density of each feature was computed and determined as to whether the feature was enriched compared to the remaining regions. Compared with random expectation, the vicinity of chromosomal breaks was significantly enriched for several genomic features, including genomic repeats, non-B DNA conformation forming motifs, conserved elements, exon density, substitution rate and indel rate (Table 3; Bonferroni corrected P-values<0.01, Mann-Whitney test). These genomic features have been associated with SCNA breakpoints in different cancer types [55], suggesting that OS is similar to other cancers in regards to chromosomal breakage occurrence. Of note, common fragile sites were not preferentially associated with chromosomal breaks at any genomic resolution investigated in this study (Table 3), indicating that OS has perhaps very specific breakage characteristics that include already known common fragile sites as well as unique sites of instability.

Clinical implications of chromothripsis-like patterns and hyperploidy

Applying the CTLP detecting algorithm to the OS SCNA dataset, a total of 87 chromosomes from 52 patients passed the threshold and were termed CTLP cases. CTLP occurred in 33.1% of patients within this dataset, implying that chromothripsis is a widespread phenomenon in OS. This incidence rate was largely consistent with a previous study of a small sample size of bone cancers [20]. CTLPs had a tendency to occur frequently on chromosomes 8 (11.5%) and 17 (9.2%). The OncoPrint shown in Figure 4 provides an overview of SCNAs in specific genes and CTLP affecting individual samples. Chromosomal aberrations in *TP53* occurred in 88% (46/52) of CTLP patients, compared to 56% (59/105) of non-CTLP cases (P-value = 1.0×10^{-4}, two-tailed Fisher's exact test). We analysed three genes - *RB1*, *WWOX* and *DLG2* - that frequently harbor structural variation in OS [16]. Chromosomal alterations in *RB1* occur in 73% (38/52) of CTLP cases, but only in 48% (50/105) of non-CTLP samples (P-value = 3.5×10^{-3}, two-tailed fisher's exact test). Chromosomal aberrations in *WWOX* occur in 85% (44/52) and 66% (69/105) CTLP and non-CTLP samples respectively (P-value=1.4×10^{-2}, two-tailed fisher's exact test). Finally, 83% (43/52) of CTLP cases harboured aberrations in *DLG2*, compared to 57% (60/105) of non-CTLP cases (P-value=1.3×10^{-3}, two-tailed fisher's exact test). These observations indicate

13

that chromosomal aberrations in *TP53*, *RB1, WWOX* and *DLG2* genes are strongly associated with chromothripsis-like patterns in OS.

Furthermore, an investigation of the association between chromothripsis-like patterns and clinical data was performed [57]. As follow-up clinical data was available for 114 patients, CTLP was detected in 33% (38/114) of this cohort. Notably, as shown in Figure 5a, Kaplan-Meier analysis revealed that patients with CTLP patterns in their tumors showed significantly curtained survival expectancies compared to those without CTLP (log-rank test, *P*-value=7.06×10^{-4}).

A successful estimation was made of tumor ploidy and content for 90.4% (142 /157) of samples using the GPHMM algorithm. These osteosarcoma biopsies were estimated to have on average 37.5% normal tissue contamination with a median ploidy of 2.7n. Following the procedures for chromosome number estimation (as described in the *Methods*), the distribution of chromosome numbers was plotted in 142 samples to clearly demonstrate a two ploidy status of the tumor genome (Figure 5b). Near-diploid was defined for tumors with chromosome number < 60 and DNA index <1.3 (see *Methods* for details), without consideration for SCNAs presence or absence in tumors. Near-tetraploid tumors had greater chromothripsis events than those classified as near-diploid (Figure 5c, P-value=4.60×10^{-3}, Fisher's exact test). This was compatible with results from a recent study linking chromothripsis with hyperploidy [58]. Patients with tumors exhibiting near-tetraploid genomes had poorer survival compared to patients having tumors with estimated ploidy of ~ 2 (Figure 5d).

Discussion

Rarity and genomic complexity, as well as marked intra- and inter-tumoral heterogeneity, have challenged the molecular characterization of osteosarcoma etiology [19]. Given the difficulty in acquiring a large cohort of samples in this rare tumor, we integrated DNA copy number profiles of 160 pretherapeutic biopsies to identify recurrent genomic changes and driver genes. Genome-wide profiles were performed using an Affymetrix CytoScan HD platform, which has the highest resolution of SNP and non-polymorphic probes for detecting human chromosomal alterations. Copy number analyses confirmed high genomic instability in OS biopsies, with the vast majority of samples (82%) exhibiting highly complex altered genomes.

The unstable genome in the majority of OS is probably due to the deficiency in homologous recombination repair [43]. The *BRCA1/2* (important players in homologous recombination pathway) deficiency associated characteristics in single base substitutions, and large-scale genome instability signatures are evident in more than 80% of OS [43].

Using GISTIC, we identified a number of genes frequently targeted in OS, including already known driver genes (e.g. *TP53* and *ATRX*) as well as other OS-related genes, such as *WWOX*. *WWOX* is a putative tumor suppressor gene encompassing a common fragile site FRA16D, which is a frequent target of chromosomal rearrangement in multiple cancers. The absence or reduced expression of *WWOX* have been linked to poor prognosis in a wide variety of cancers, particularly in ovarian cancer and OS [59, 60]. In previous reports by others, the function loss of WWOX has been linked to chromosomal deletions and translocations as well as loss of expression [47, 49]. In this study, we showed that 32% of OS samples have at least one chromosomal break within the *WWOX* gene, supporting the *WWOX* inactivation by chromosomal rearrangements. We further showed that the *WWOX* gene was located in 'broken regions' (discussed below) with SCNAs and chromosomal breaks in those regions more likely to be of early occurrence. These results are consistent with the hypothesis that loss of WWOX expression is an early event in OS pathogenesis [49].

Genome-wide analysis revealed that chromosomal breaks are not randomly distributed and clustered in 'broken regions'. About half of these regions overlapped with non-fragile sites, strongly suggestive of OS-specific fragility. Our observations comply with the findings that unstable sites are tissue specific [31]. It is noteworthy that OS-associated tumor suppressor genes including *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP* [19] are situated in 'broken regions'. SCNAs in those regions were more likely to be clonal events as opposed to those expected by chance. The early occurrence of breakages and the presence of multiple tumor suppressor genes in such regions may explain the complex and aggressive nature of OS.

We further revealed that SCNA breakpoints and chromosomal breaks were significantly correlated with diverse genomic properties, including Alu, L1, cruciform, G4, slip, triplex, Z-DNA, conserved elements, exon density, and indel rate. Genomic repeats such as L1 and Alu are interspersed throughout the human genome at high copy numbers, and non-allelic homologous recombination events between different

copies lead to duplications, deletions, and inversions [61]. Repetitive DNA motifs may fold into non-B DNA conformation, thereby serving as chromosomal targets for DNA repair and recombination leading to the formation of structural variations including CNVs, inversions and translocations [62]. Therefore, it could be speculated that breakages probably occur at OS-specific instability sites with the potential to form stable secondary structures (*i.e.* non-B DNA structures) and to consequently stall the replication fork.

Based on 20 patients including 9 osteosarcomas and 11 chordomas, Stephens et al. [20] estimated that 25% of bone cancers were associated with chromothripsis. In our dataset, chromothripsis-like patterns occurred in about one third of patients suggestive that chromothripsis is a widespread OS phenomenon. Massive genomic rearrangement raised by chromothripsis apparently represents an important mechanism of carcinogenesis, as distinct from progressive accumulation. Although the underlying cause of chromothripsis is not fully understood, several hypotheses have been recently proposed [20, 63, 64]. Firstly, chromothripsis might occur by ionizing radiation induced DNA damage at a short or long stretch of the chromosome [20, 63]. Secondly, telomere attrition may cause dicentric chromosomes which persist through mitosis developing into chromatin bridges that further generate single-stranded DNA and trigger DNA repair [20, 65]. Thirdly, abortive apoptosis has also been considered as a possible mechanism [63], but it does not provide a reasonable explanation for the localization of DNA shattering [64]. Fourthly, premature chromosome compaction, in which chromosomes are induced to undergo chromosome condensation before completing DNA replication, results in shattering of the incompletely replicated chromosome [66]. An appealing explanation for chromothripsis is that the localized damage could occur in one or two chromosomes (or chromosome part) physically isolated from other chromosomes [67]. The so-called nuclear structure 'micronuclei' are widely observed in cancer cell lines. Taking advantage of live cell imaging and single-cell genome sequencing, Zhang *et al.* demonstrated that chromatid fragmentation and subsequent reassembly occur in the micronucleus and can generate localized genomic rearrangements, some of which recapitulate all features of chromothripsis [68]. Investigation of the association between lesions in specific genes and chromothripsis will offer some insights into the impact of chromothripsis in cancerogenesis. Our analysis indicates that SCNAs in the *TP53*, *RB1* and *DLG2* are strongly associated

16

with chromothripsis-like patterns in OS. Among them, *DLG2* frequently shows breakages in OS and may be a preferential target for chromothripsis and breakage [16]. *RB1* is significantly copy-number altered in OS, while the other candidate, *TP53*, has already been linked to chromothripsis in medulloblastoma [23]. Utilizing an *in vitro* cell-based system, chromothripsis has been recently linked to hyperploidy [58]. Indeed, we have shown that compared with diploid tumors, those which are hyperploid had a greater chance to harbour chromothripsis events and less favourable outcomes.

Conclusions

A comprehensive characterization of SCNAs in a large cohort (n = 160) of osteosarcoma samples was undertaken in this study. Almost all (98%) of the analysed OS samples were of sufficient quality for data analysis. A high degree of aneuploidy and large-scale copy number alterations in OS were confirmed. Using GISTIC, a number of genes that are frequently targeted in OS were identified, of which *TP53*, *ATRX*, *FOXN1*, and *WWOX* are already known tumor suppressors associated with OS and other tumor types. Genome-wide analysis of chromosomal breaks revealed a tendency for confinement to genomic regions harbouring OS-associated tumor suppressor genes including *TP53*, *RB1*, *WWOX*, *DLG2*, and *LSAMP*. Breakage susceptibility in OS was found to be largely dependent on local genomic context. A complex breakage pattern – chromothripsis – is suggested as a widespread OS phenomenon correlated with patient survival. By unlocking a definitive OS instability pattern, a specific code is hereby revealed shedding an important light on the biology of osteosarcoma.

Authors' contributions

JS, CLM, DF and MN designed and initiated the study. DB, IvL, SB and MN collected OS samples and the corresponding clinical data. JS, SR and MK performed SNP array analysis. HX, YZ and DF conceived the bioinformatics part of the project and DF supervised it. HX and YZ did the bioinformatics analysis. VBO performed gene expression and edited the manuscript. JS, HX, YZ, DF and MN wrote the manuscript with all other authors contributing to its completion.

Acknowledgements

References

1. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the Surveillance, Epidemiology, and End Results Program. Cancer 2009;115:1531-1543.

2. Bielack SS, Kempf-Bielack B, Branscheid D, et al. Second and subsequent recurrences of osteosarcoma: presentation, treatment, and outcomes of 249 consecutive cooperative osteosarcoma study group patients. J. Clin. Oncol. 2009;27:557-565.

3. Bayani J, Zielenska M, Pandita A, et al. Spectral karyotyping identifies recurrent complex rearrangements of chromosomes 8, 17, and 20 in osteosarcomas. Genes Chromosomes Cancer 2003;36:7-16.

4. Smida J, Baumhoer D, Rosemann M, et al. Genomic alterations and allelic imbalances are strong prognostic predictors in osteosarcoma. Clin. Cancer. Res. 2010;16:4256-4267.

5. Kuijjer ML, Rydbeck H, Kresse SH, et al. Identification of osteosarcoma driver genes by integrative analysis of copy number and gene expression data. Genes Chromosomes Cancer 2012;51:696-706.

6. Luetke A, Meyers PA, Lewis I, et al. Osteosarcoma treatment - where do we stand? A state of the art review. Cancer Treat. Rev. 2014;40:523-532.

7. Man T-K, Chintagumpala M, Visvanathan J, et al. Expression profiles of osteosarcoma that can predict response to chemotherapy. Cancer Res. 2005;65:8142-8150.

8. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability--an evolving hallmark of cancer. Nat. Rev. Mol. Cell Biol. 2010;11:220-228.

9. Martin JW, Squire JA, Zielenska M. The genetics of osteosarcoma. Sarcoma 2012;2012:627254.

10. Durkin SG, Glover TW. Chromosome fragile sites. Annu. Rev. Genet. 2007;41:169-192.

11. Zeman MK, Cimprich KA. Causes and consequences of replication stress. Nat. Cell Biol. 2014;16:2-9.

12. Smeets SJ, Harjes U, van Wieringen WN, et al. To DNA or not to DNA? That is the question, when it comes to molecular subtyping for the clinic! Clin. Cancer. Res. 2011;17:4959-4964.

13. Van Loo P, Nordgard SH, Lingjaerde OC, et al. Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U. S. A. 2010;107:16910-16915.

14. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. Nature 2010;463:899-905.

18

15. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. Nat. Genet. 2013;45:1134-1140.

16. Chen X, Bahrami A, Pappo A, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. Cell Rep 2014;7:104-112.

17. Perry JA, Kiezun A, Tonzi P, et al. Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. Proc. Natl. Acad. Sci. U. S. A. 2014;111:E5564-E5573.

18. Poos K, Smida J, Maugg D, et al. Genomic heterogeneity of osteosarcoma - shift from single candidates to functional modules. PLoS One 2015;10:e0123082.

19. Kansara M, Teng MW, Smyth MJ, et al. Translational biology of osteosarcoma. Nat. Rev. Cancer 2014;14:722-735.

20. Stephens P, Greenman C, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27-40.

21. Magrangeas F, Avet-Loiseau H, Munshi NC, et al. Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. Blood 2011;118:675-678.

22. Molenaar JJ, Koster J, Zwijnenburg DA, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. Nature 2012;483:589-593.

23. Rausch T, Jones DTW, Zapatka M, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell 2012;148:59-71.

24. Maher CA, Wilson RK. Chromothripsis and human disease: piecing together the shattering process. Cell 2012;148:29-32.

25. Bielack SS, Kempf-Bielack B, Delling G, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. J. Clin. Oncol. 2002;20:776-790.

26. Kolesnikov N, Hastings E, Keays M, et al. ArrayExpress update-simplifying data submissions. Nucleic Acids Res. 2015;43:D1113-D1116.

27. Mermel CH, Schumacher SE, Hill B, et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12:R41.

28. Qiao Y, Quinlan AR, Jazaeri AA, et al. SubcloneSeeker: a computational framework for reconstructing tumor clone structure for cancer variant interpretation and prioritization. Genome Biol. 2014;15:443.

29. Meyer LR, Zweig AS, Hinrichs AS, et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res. 2013;41:D64-D69.

30. Cer RZ, Donohue DE, Mudunuri US, et al. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic Acids Res. 2013;41:D94-D100.

31. Sarni D, Kerem B. The complex nature of fragile site plasticity and its importance in cancer. Curr. Opin. Cell Biol. 2016;40:131-136.

32. Fungtammasan A, Walsh E, Chiaromonte F, et al. A genome-wide analysis of common fragile sites: What features determine chromosomal instability in the human genome? Genome Res. 2012;22:993-1005.

19

33. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841-842.

34. Cai H, Kumar N, Bagheri HC, et al. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. BMC Genomics 2014;15:82.

35. Hernandez-Ferrer C, Quintela Garcia I, Danielski K, et al. affy2sv: an R package to pre-process Affymetrix CytoScan HD and 750K arrays for SNP, CNV, inversion and mosaicism calling. BMC Bioinformatics 2015;16:167.

36. Uddin M, Thiruvahindrapuram B, Walker S, et al. A high-resolution copy-number variation resource for clinical and population genetics. Genet. Med. 2014;17:747-752.

37. Li A, Liu Z, Lezon-Geyda K, et al. GPHMM: an integrated hidden Markov model for identification of copy number alteration and loss of heterozygosity in complex tumor samples using whole genome SNP arrays. Nucleic Acids Res. 2011;39:4928-4941.

38. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17:1665-1674.

39. Popova T, Manié E, Stern M. Genomic Signature of Homologous Recombination Deficiency in Breast and Ovarian Cancers. Bio-protocol 2013;3:e814.

40. Popova T, Manié E, Rieunier G, et al. Ploidy and large-scale genomic instability consistently identify basal-like breast carcinomas with BRCA1/2 inactivation. Cancer Res. 2012;72:5454-5462.

41. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43:D805-D811.

42. Ribi S, Baumhoer D, Lee K, et al. TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome. Oncotarget 2015;6:7727-7740.

43. Kovac M, Blattmann C, Ribi S, et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA deficiency. Nat Commun 2015;6:8940.

44. Martin JW, Zielenska M, Stein GS, et al. The Role of RUNX2 in Osteosarcoma Oncogenesis. Sarcoma 2011;2011:282745.

45. Weiner L, Han R, Scicchitano BM, et al. Dedicated epithelial recipient cells determine pigmentation patterns. Cell 2007;130:932-942.

46. Mangelsdorf M, Ried K, Woollatt E, et al. Chromosomal fragile site FRA16D and DNA instability in cancer. Cancer Res. 2000;60:1683-1689.

47. Aqeilan RI, Abu-Remaileh M, Abu-Odeh M. The common fragile site FRA16D gene product WWOX: roles in tumor suppression and genomic stability. Cell. Mol. Life Sci. 2014;71:4589-4599.

48. Schrock MS, Huebner K. WWOX: a fragile tumor suppressor. Exp. Biol. Med. (Maywood) 2015;240:296-304.

49. Yang J, Cogdell D, Yang D, et al. Deletion of the WWOX gene and frequent loss of its protein expression in human osteosarcoma. Cancer Lett. 2010;291:31-38.

50. Del Mare S, Aqeilan RI. Tumor Suppressor WWOX inhibits osteosarcoma metastasis by modulating RUNX2 function. Sci. Rep. 2015;5:12959.

51. Zheng S, Fu J, Vegesna R, et al. A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. Genes Dev. 2013;27:1462-1472.

52. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics 2008;1:4.

53. De S, Michor F. DNA secondary structures and epigenetic determinants of cancer genome evolution. Nat. Struct. Mol. Biol. 2011;18:950-955.

54. Zhou W, Zhang F, Chen X, et al. Increased genome instability in human DNA segments with self-chains: homology-induced structural variations via replicative mechanisms. Hum. Mol. Genet. 2013;22:2642-2651.

55. Li Y, Zhang L, Ball RL, et al. Comparative analysis of somatic copy-number alterations across different human cancer types reveals two distinct classes of breakpoint hotspots. Hum. Mol. Genet. 2012;21:4957-4965.

56. Zhang Y, Xu H, Frishman D. Genomic determinants of somatic copy number alterations across human cancers. Hum. Mol. Genet. 2016;25:1019-1030.

57. Forment JV, Kaidi A, Jackson SP. Chromothripsis and cancer: causes and consequences of chromosome shattering. Nat. Rev. Cancer 2012;12:663-670.

58. Mardin BR, Drainas AP, Waszak SM, et al. A cell-based model system links chromothripsis with hyperploidy. Mol. Syst. Biol. 2015;11:828.

59. Nunez MI, Rosen DG, Ludes-Meyers JH, et al. WWOX protein expression varies among ovarian carcinoma histotypes and correlates with less favorable outcome. BMC Cancer 2005;5:64.

60. Kurek KC, Del Mare S, Salah Z, et al. Frequent Attenuation of the WWOX Tumor Suppressor in Osteosarcoma Is Associated with Increased Tumorigenicity and Aberrant RUNX2 Expression. Cancer Res. 2010;70:5577-5586.

61. Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. Semin. Cancer Biol. 2010;20:211-221.

62. Zhao J, Bacolla A, Wang G, et al. Non-B DNA structure-induced genetic instability and evolution. Cell. Mol. Life Sci. 2010;67:43-62.

63. Tubio JMC, Estivill X. Cancer: When catastrophe strikes a cell. Nature 2011;470:476-477.

64. Leibowitz ML, Zhang CZ, Pellman D. Chromothripsis: A New Mechanism for Rapid Karyotype Evolution. Annu. Rev. Genet. 2015;49:183-211.

65. Maciejowski J, Li Y, Bosco N, et al. Chromothripsis and Kataegis Induced by Telomere Crisis. Cell 2015;163:1641-1654.

66. Meyerson M, Pellman D. Cancer genomes evolve by pulverizing single chromosomes. Cell 2011;144:9-10.

67. Crasta K, Ganem NJ, Dagher R, et al. DNA breaks and chromosome pulverization from errors in mitosis. Nature 2012;482:53-58.

68. Zhang CZ, Spektor A, Cornils H, et al. Chromothripsis from DNA damage in micronuclei. Nature 2015;522:179-184.

Figure Legends

Figure 1

Genome-wide frequency plot of somatic copy number alterations in 157 osteosarcoma samples. Copy number losses and gains are in red and blue respectively.

Figure 2

Significantly altered regions and genes contained therein with copy number alterations in osteosarcoma as identified by GISTIC analysis.

Figure 3

The genomic landscape of chromosomal breaks and associated genes in osteosarcoma. The outermost circle represents chromosomes and cytogenetic bands. The next circle represents known OS driver genes and other genes as listed in Table 2. The third circle represents 'broken regions'. The innermost circle shows common fragile sites and non-fragile regions in red and blue respectively.

Figure 4

OncoPrint showing the distribution of SCNAs (CN gain and CN loss) for genes *TP53*, *RB1*, *DLG2* and *WWOX* and chromothripsis-like pattern (CTLP) in osteosarcoma patients (column). Each bar represents a sample. Green bars indicate samples with CTLP. Red and blue bars indicate samples with CN loss and gain for a specific gene, respectively. Gray bars represent samples without CTLP or without CN changes for a specific gene. The numbers on the left show what percentage of samples is affected by CTLP or CN changes for a specific gene.

Figure 5

Clinical implications of chromothripsis and ploidy. (a) Kaplan-Meier survival curves for chromothripsis-like patterns (CTLPs) *versus* non-CTLP cases. The P-value is based on the log-rank test; (b) Distribution of chromosome numbers in 142 osteosarcoma samples, displaying the two ploidy status of tumor genomes; (c) Association of the ploidy status with chromothripsis; (d) Kaplan-Meier survival curves for near-tetraploid samples *versus* near-diploid samples. The P-value is based on the log-rank test.

Supplementary Materials

Supplementary Figure S1

Schematic illustration of chromosomal breaks. "d" means Log_2 value changes between two adjacent genomic segments at a specific genomic position. Genomic position with $d > 0.3$ was defined as a "chromosomal break".

Supplementary Figure S2

The distribution of chromosomal breaks around the *TP53* gene in 157 osteosarcoma samples. The red bars after sample identifiers denote the genomic locations of chromosomal breaks.

Supplementary Tables 1-6

Provided in Supplementary Tables.xlsx

Tables

Table 1 Clinical characteristics of 157 osteosarcoma patients.

Descriptive statistics		
Sex	**n=157**	
Male	83	
Female	74	
Age at diagnosis (years)	**n=157**	
Average	20.08	
Median	15	
Range	3-85	
Metastases	**n=143**	
Yes	61	
No	82	
Observation period (months)	**n=147**	
Average	64.5	
Median	56.2	
Range	0.24-204.5	
Response to neoadjuvant treatment	**n=128**	
Good	64	
Poor	64	
Survival	**n=130**	
Alive	90	
Deceased	40	
Event (relapse or death)	**n=143**	
Yes	60	
No	83	
Overall survival	5-year: 74.8%	10-year: 62.9%
Grouped by event status	**5-year**	**10-year**
Event	25.5%	27.3%
Grouped by response to chemotherapy	**5-year**	**10-year**
Good response	90.2%	83.6%
Poor response	66.7%	61.1%

Table 2 Genes frequently targeted by chromosomal breaks in OS that were previously shown to associate with OS or other tumors.

Gene	Chromosome	Start	End	OMIM	Count	% OS
DLG2	11	83166055	85338314	603583	113	27.39
WWOX	16	78133309	79246564	605131	102	31.85
DMD	X	31137344	33357726	300377	71	17.83
EYA1	8	72109667	72274467	601653	62	20.38
SCAPER	15	76640526	77176217	611611	61	19.75
ERBB4	2	212240441	213403352	600543	43	12.74
FHIT	3	59735035	61237133	601153	42	8.28
WNK1	12	862088	1020618	605232	40	14.01
KANSL1	17	44107281	44302740	612452	40	21.66
LRP1B	2	140988995	142889270	608766	39	12.74
TP53	17	7571719	7590868	191170	34	19.75
TP63	3	189349215	189615068	603273	34	10.83
USP34	2	61414589	61697849	615295	29	11.46
TERT	5	1253286	1295162	187270	28	10.19
FOXN1	17	26850958	26865175	600838	25	15.92
NF2	22	29999544	30094589	607379	25	6.37
RB1	13	48877882	49056026	614041	24	8.28
NEGR1	1	71868624	72748277	613173	21	7.01
CHM	X	85116184	85302566	300390	21	7.01
LSAMP	3	115521209	116164385	603241	19	8.92
PTEN	10	89623194	89728532	601728	11	3.82
APC	5	112043201	112181936	611731	10	3.18
RET	10	43572516	43625797	164761	8	4.46
FANCA	16	89803958	89883065	607139	6	2.55

All genomic coordinates are based on human genome assembly hg19; Count: the total number of chromosomal breaks found in gene regions; % OS: percent of OS samples affected by chromosomal breaks; gene names previously associated with OS [19] are in bold; gene names identified by GISTIC analysis in this study are in italics.

Table 3 Correlation between chromosomal breaks and genomic features

Genomic features	Enrichment in genomic regions centered at chromosomal breaks			
	10 kb	20 kb	50 kb	100 kb
Alu	+	+	+	+
DNA transposons	+	+	+	+
L1	+	+	+	+
LTR retrotransposons	+	+	+	+
Cruciform		+	+	+
G4	+	+	+	+
Slip	+	+	+	+
Triplex			+	+
Z-DNA		+	+	+
Conserved elements	+	+	+	
Exon density			+	+
Common fragile sites				
Substitution rate	+	+	+	+
Indel rate	+	+	+	+

+ denotes enrichment of genomic features in genomic windows centered at chromosomal breaks (Bonferroni corrected P-values <0.01).
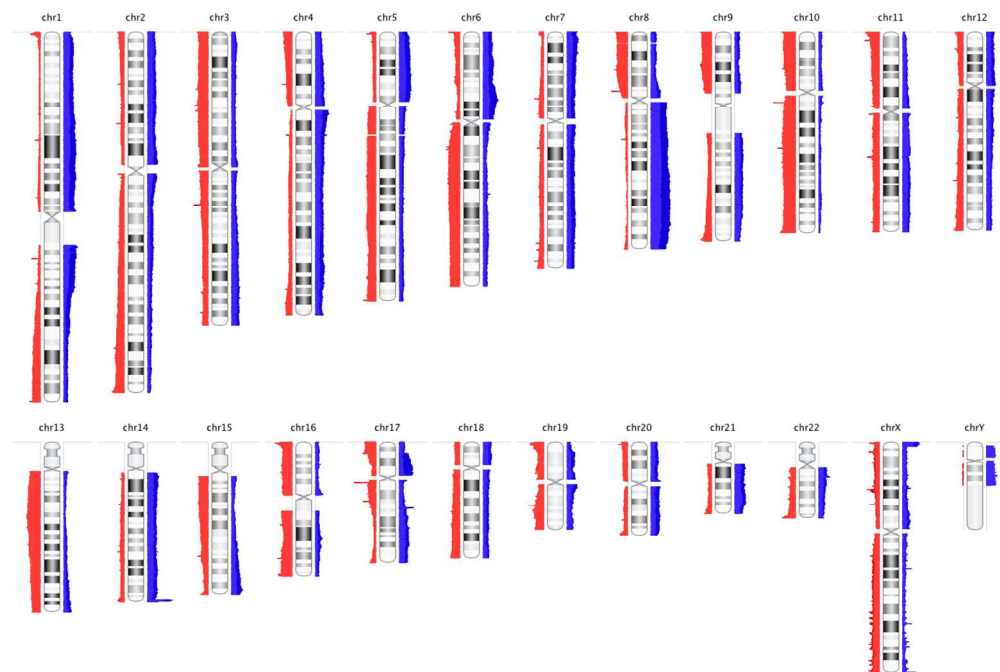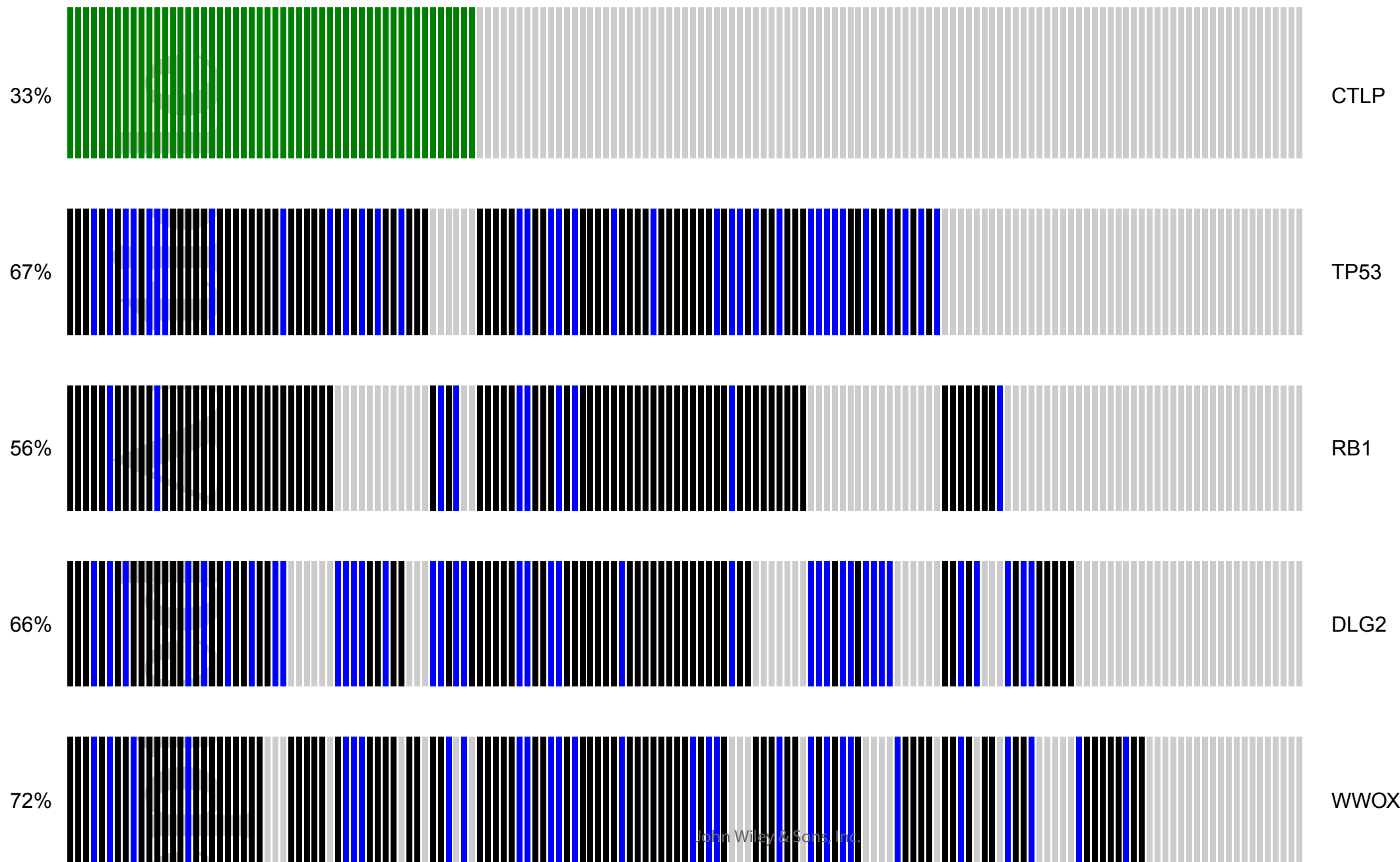
Genome-wide frequency plot of somatic copy number alterations in 157 osteosarcoma samples. Copy number losses and gains are in red and blue respectively.
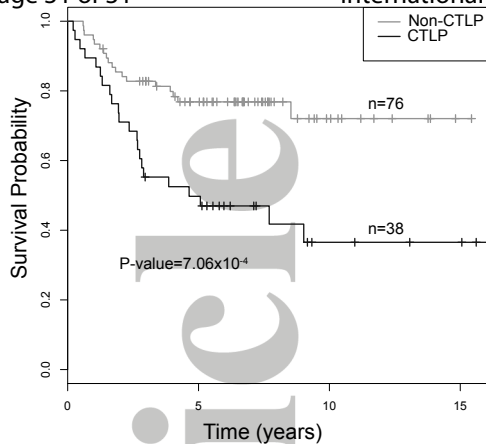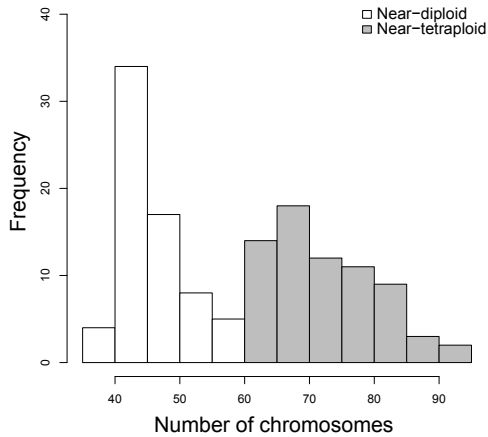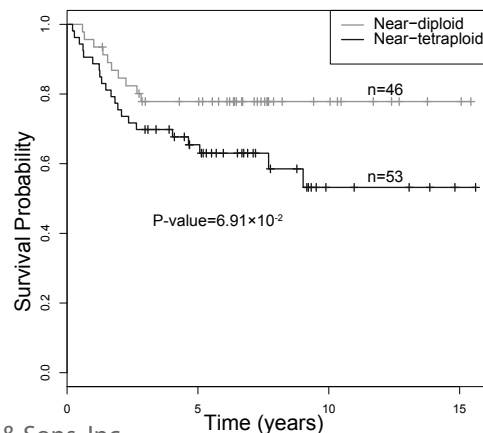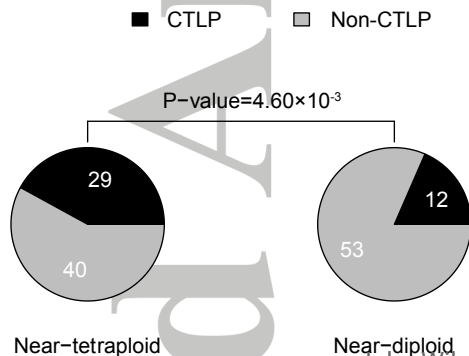
118x82mm (300 x 300 DPI)

Chr.	Loss	Gain	Regions	Transcripts

CN Gain — **NOTCH2**
CN Gain — SETDB1, **CERS2**, ANXA9, FAM63A, PRUNE
CN Gain — LCE1D
CN Loss — NME7
CN Loss
CN Loss
CN Loss

CN Gain — CD8B, RMND5A
CN Gain — ANKRD36
CN Loss — SNED1, MTERF4, *MTERFD2*
CN Loss — CTDSPL
CN Loss
CN Loss — TP63
CN Loss

CN Gain — ATP10D, CORIN
CN Gain — **PDGFRA**
CN Loss — UGT2B15
CN Gain
CN Loss

CN Gain — *OSMR-AS1*
CN Loss — BTNL8
CN Loss

CN Gain — RUNX2
CN Loss

CN Gain — SDK1
CN Loss — *TCRBV2S1, TCRVB, BV03S1J2.2, BV6S4-BJ2S2, PRSS3P2*, PRSS2
CN Loss — DPP6
CN Loss
CN Loss

CN Gain — *ADAM5*
CN Loss — *ADAM3A*
CN Gain — *LOC101929268*
CN Loss — EYA1
CN Loss — MTDH
CN Gain — *BC042052, CASC11*
CN Loss — MTAP, **CDKN2A**
CN Loss — KIAA1217
CN Loss — ANXA8
CN Loss — C10orf11
CN Loss — TRIM22, OR52N5, TRIM5
CN Loss
CN Loss — DLG2
CN Gain — ANGPTL5, *KIAA1377,* C11orf70
CN Loss — **FLI1**
CN Loss — WNK1
CN Gain

CN Gain — AGAP2, TSPAN31, *MIR6759*, **CDK4**, *DM110804*, MARCH9, CYP27B1, METTL1, METTL21B, TSFM, AVIL, *MIR26A2,* CTDSP2, *AK130110*
CN Loss — ANKS1B
CN Loss

CN Gain — *KIAA0125, ADAM6*
CN Gain — *BC042994*
CN Loss
CN Loss
CN Loss — SCAPER
CN Gain — PGPEP1L, *AL109706*, SYNM, TTC23, *HSP90B2P*, LRRC28
CN Loss
CN Loss — WWOX
CN Loss — **TP53**
CN Gain — MPRIP
CN Loss — FOXN1
CN Loss

CN Gain — KANSL1, *KANSL1-AS1*
CN Gain
CN Loss — DYM
CN Loss — FGF22, RNF126
CN Loss — INSR
CN Gain — **CCNE1**
CN Loss — ARHGEF1
CN Loss — SIRPB1
CN Gain

CN Gain — **RUNX1**
CN Loss

CN Gain — *DKFZp667J0810, MIR650*
CN Loss

1.0e-5   1   1.0e-5

48

33%   CTLP

67%   TP53

56%   RB1

66%   DLG2

72%   WWOX

CN Gain
CN Loss
CTLP

(a)

(b)

(c)

(d)