SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Tumor Samples

Twenty high-grade intermedullary osteosarcoma samples with matched normal tissue from 19 patients were subjected to the whole genome sequence (WGS) analysis. The samples included 16 untreated primary and 4 metastatic tumors, the latter obtained from 3 patients, of whom two had metastatic disease at presentation and one had tumor recurrences. The tumors occurred in 11 males and 8 females ranging in age from 8-22 years of age (median age, 14 years). H&E slides of the tumors were retrieved from the Pathology Archives at St. Jude Children's Research Hospital for review and the tumors were classified in the following histologic subtypes: osteoblastic (10 tumors), mixed pattern (3 tumors), telangiectatic (3 tumors), fibroblastic (2 tumor), chondroblastic (1 tumor), and small cell (1 tumor). Clinical features of the validation cohort are provided in Suppl. Table 1. For the p53 analysis cohort, 38 samples from 31 additional patients were analyzed for alterations in the p53 pathway. The clinicopathologic data and the results of the molecular genetic assays on validation cohort are provided in Supplemental Table 4.

TP53 Immunostaining

The corresponding formalin-fixed, paraffin-embedded (FFPE) tissue blocks for each specimen were cut at 4 micron thickness. Immunohistochemical staining was performed using an antibody directed against p53 protein (DO-7, DAKO, 1:50) and processed with standard heat-induced epitope retrieval (Ventana CC1) and the Ventana IVIEW detection systems. p53 nuclear staining was scored using a previously published scoring system(Papai et al., 1997) with a minor modification as follows: samples with no staining

cells were scored as negative; samples with <5% p53 immunopositive cells were scored as rare; samples with 5-25% immunopositive cells were scored as 1+; samples with 26-50% immunopositive cells were scored as 2+; samples with >50% immunopositive cells were scored as 3+.

TP53 and MDM2 Fluorescence In Situ Hybridization Studies

The fluorescence in situ hybridization (FISH) probe sets were designed using bacterial artificial chromosome (BAC) clones according to the UCSC Genome Bioinformatics database (http://genome.ucsc.edu). The *TP53* dual-color break-apart FISH assay was developed by using RP11-1081A10 BAC clone, flanking the 3′ end of the gene– labeled with Rhodamine (red fluorochrome)– and the RP11-709J3 BAC clone, flanking the 5′ end of the gene– labeled with AlexaFluor-488 (green fluorochrome). In addition, a *TP53* FISH probe mixture was constructed to enumerate the gene copy using the RP11-89D11 BAC clone, spanning the entire sequence of *TP53* that was labeled with green fluorochrome and the control RP11-64J19 that was labeled with red fluorochrome. FISH for *MDM2* was set up using the RP11-611O2 BAC clone spanning the entire gene (red fluorechrome) and a probe targeting *ATF1* at 12q13.1 as the control probe (green fluorochrome. DNA was isolated from BAC clones (BACPAC Resources, Oakland, CA) according to a modified Qiagen (Valencia, CA) extraction protocol. The probes were labeled by nick translation using a modification of the manufacturer's protocol (Life Technologies, Inc., Carlsbad, CA). FISH analysis was performed on 4 micron-thick FFPE tissue sections using the previously published methods(Bahrami et al., 2012).

Hybridization signals were evaluated in 200 interphase nuclei of each sample. FISH images were captured and processed as previously described(Bahrami et al., 2012).

Telomeres were analyzed in the discovery cohort using 3 different methods. The whole genome sequencing (WGS) data was analyzed for telomere length (described below) for all 20 tumors and matched normal tissue in the discovery cohort. Quantitative PCR (described below) was performed to validate the results from WGS analysis for all 10 tumors in the discovery cohort with *ATRX* mutations and an additional 4 samples with wild type *ATRX* to serve as controls. Matched normal DNA was used as internal control for each patient's sample. Telomere FISH (see below) was performed on all tissue samples in the discovery cohort that had available FFPE material (non-decalcified tumor tissue). All of the samples that were analyzed by telomere FISH were also analyzed for ATRX protein expression by immunohistochemistry (described below).

Whole-Genome Sequencing, RNA-Seq and Exome Seq

Using a paired-end sequencing approach, we sequenced DNA from 20 tumors and their matching germline control DNA with an average of 30× haploid coverage per genome. Single nucleotide variations (SNVs) and insertions/deletions (indels) were identified independently algorithms by Washington University Genome Sequencing Center (WUGSC) and St. Jude Children's Research Hospital (SJCRH) using different approaches. The results generated were then compared and a final candidate SNV and indel list was developed for experimental validation.

At WUGSC, SNVs were found by Somatic Sniper that defines high quality somatic predictions as those sites with a somatic score greater than 40 and an average

mapping quality greater than 40. The predicted SNVs are compared to the most current version of dbSNP(Sherry et al., 2001) (build 129-130). For SNVs, we require both positional and allele match. In addition we also compared the predicted SNVs to SNPs found in CEU and YRI trios as described(Ding et al.). All predicted SNVs were filtered through a SNV false-positive filter developed at the Genome Institute that is based on a standard set of criteria including mapping quality score, average supporting read length, average position of the variant in the read, strand bias and the presence of homopolymer. Indels were called using modified SAMtools(Li et al., 2009) indel-calling algorithm as described(Ding et al.), Pindel(Ye et al., 2009) and GATK(Zerbino and Birney, 2008).

At SJCRH, putative sequence variants including SNVs and indels were initially detected by running the variation detection module of Bambino(Edmonson et al.) using the following three parameters: (1) a high quality threshold for pooled tumor and matching normal bam files (min-quality=20, min-flanking-quality=20, min-alt-allele-count=3, min-minor-frequency=0, broad-min-quality=10, mmf-max-hq-mismatches=4, mmf-min-quality=15, mmf-max-any-mismatches=6; (2) a low quality threshold for pooled tumor and matching normal bam files (min-quality=10, min-flanking-quality=10, min-alt-allele-count=2, min-minor-frequency=0, broad-min-quality=10); and (3) a high tolerance for the number of mismatches for normal bam file alone (min-quality=20, min-flanking-quality=15, min-alt-allele-count=2, min-minor-frequency=0, mmf-max-hq-mismatches=15, mmf-min-quality=15, mmf-max-any-mismatches=20). In addition to Bambino, putative indels were also found by a *de novo* assembly process which construct contigs using unmapped reads and re-map them to the reference genome followed by a Smith-Waterman alignment to detect indels. In this process, unmapped reads include (1)

unmapped reads whose mate are mapped to the genome; (2) reads with indels in CIGAR (Compact Idiosyncratic Gapped Alignment Report) string; (3) reads with at least 4 high-quality (quality value >=20) mismatches; and (4) reads with high-quality (quality value at least 20) soft-clipped bases in the CIGAR string. All putative sequence variants were further assessed to determine their accuracy and somatic original using the processes described below. Velvet(Zerbino and Birney, 2008), BLAT(Kent, 2002) and SIM(Huang et al., 1990) were the three programs used for assembly, mapping, and Smith-Waterman alignment, respectively.

A putative somatic sequence mutation determined by SJCRH process was collected based on the following criteria: (1) the variant site is absent in the normal-only analysis; (2) Fisher's exact test *P* value indicates that the number of reads harboring non-reference allele is significantly higher in tumor; (3) the non-reference allele frequency in normal is <=5%; and (4) mutant alleles present in both orientations. Higher *P* value and absence of non-reference allele in normal is required for a variant to be considered somatic if it matches dbSNP build 130 or is located in an unmappable region (determined by recurrence of 75mers across the reference genome) or is inside a polynucleotide repeat. Substitution variants are classified into four categories based on combination of their *P* value and sequence quality scores: High quality, high *P* value; high quality, low *P* value; low quality, high *P* value; low quality, low *P* value. *P* value refers to the *P* value of Fisher's exact test comparing the distribution of the alternative allele in tumor and normal. High *P* value, $P<0.05$; low *P* value, $0.05<P<0.10$. A final review process re-maps and re-aligns the reads harboring the non-reference allele to the reference genome to filter potential false positive calls introduced by mapping in repetitive regions and

alignment artifacts. For putative somatic indels, the review process re-aligns all reads in tumor and normal at the indel site to a mutant allele template sequence constructed by substituting the wild-type allele with the indel. Presence of reads in normal that cover the mutant allele is considered a germline variant. Structural variations including the 5 deletions in *ATRX* were detected using the CREST algorithm (Wang et al., 2011) and CONSERTING algorithm. The data have been deposited in EBI with accession number: EGAS00001000263.

Paired-end reads from mRNA-seq were aligned to the following 4 database files using BWA (0.5.5) aligner (4): (i) human NCBI Build 37 reference sequence, (ii) RefSeq, (iii) a sequence file that represents all possible combinations of non-sequential pairs in RefSeq exons, and (iv) AceView flat file downloaded from UCSC and representing transcripts constructed from human EST. The final BAM file was constructed by selecting the best alignment in the four databases. SV detection was carried out using CREST (1) and deFuse (5) as well as a novel algorithm that searched for the predicted junction breakpoints from detected SVs in matching WGS samples.

For exome sequencing, OS DNA libraries were prepared from 1 ug of WGA material from matched samples using the Illumnia TruSeq DNA library prep kit following the recommended manufacturer's protocol. Libraries were analyzed on an Agilent Bioanalyzer to inspect quality of each library construction. Germline and diagnostic library samples were independently pooled and applied for exome capture using the Illumina TruSeq Exome Enrichment kit as described by the manufacturer. Captured libraries were then clustered on the Illumina c-bot and were sequenced on an

Illumina HiSeq 2000 platform with 100 base pair end multiplexed reads at an equivalent of 3 samples per lane.

We used cghMCR (an R implementation of a modified version of GISTIC analysis(Aguirre et al., 2004)) to find common regions of copy number alterations. To identify genes of significant DNA copy number alterations, we defined the genes with Segments Of Gain Or Loss (SGOL) scores above the 3 standard deviations of the mean SGOL scores of all 'gains' scores as significantly amplified genes. The genes with SGOL scores below the 3 standard deviations of the mean SGOL scores of all 'losses' scores were sselected as significantly deleted genes.

Sequence Validation

For enrichment of the regions containing putative alterations, genomic coordinates of the putative WGS targets were used to order Nimbelgen Seqcap EZ solution bait sets (Roche). The library construction and target enrichment was performed per manufacturer's instructions using repli-G (Qiagen) whole genome amplified DNA. Enriched targets were sequenced on the Illumina platform using paired end 100 cycle sequencing. The resulting data was converted to FASTQ files using CASAVA 1.8.2 (Illumina) and mapped with BWA prior to pipeline analysis.

Statistical Methods

Kaplan-Meier method was used to estimate the overall survival and event-free survival curves. Log-rank test was performed to test the significant difference of survival curves

between *TP53* missense mutation group and the *TP53* truncating mutation group in SAS version 9.2.

We used the MuSiC software(Dees et al., 2012) to identify significantly mutated genes with point mutations. For significantly mutated genes with SVs, the "background" base-level mutation rate for SVs in each tumor under the null hypothesis that SV breakpoints were distributed randomly within the genome and the number of tumors mutated by SVs for a specific gene follows the Poisson binomial distribution under the null hypothesis. The significance level was estimated from the Le Cam's theorem.

Telomere Analysis

Telomere length was predicted in silico by counting the number of next-generation sequencing reads containing the telomeric-repeat sequence TTAGGG (Castle et al., 2010). The resulting number of reads was normalized to the average genomic coverage, and the difference in diagnostic and germline telomeric sizes was calculated. Telomere length was validated in vitro in NBs expressing an *ATRX* aberration as described previously (Cawthon, 2002; O'Callaghan et al., 2008). Briefly 15-20ng of diagnostic and germline WGA amplified DNA was subject to qPCR using two sets of primers in separate reactions, one to amplify telomeric sequence and one to amplify a common gene; *36B4 (RPLP0)*. Ct values obtained were compared to those of two standard curves, a telomeric standard curve performed on known quantities of a telomeric 84mer and one using an oligomer of *36B4 (RPLP0)*. All reactions were performed in triplicate with both tumor and germline DNA and both assays on the same plate. All reactions were carried out using Brilliant III Ultra-Fast SYBR Green master mix (Agilent) on a Stratagene

Mx3000 thermal cycler using a melting temperature of 60ºC. This allowed us to determine the telomere length in Kb per diploid genome. The forward primer for telomere analysis was:

5'- CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGGTTTGGGTT-3'

The reverse primer for telomere analysis was:

5'-GGCTTGCCTTACCCTTACCCTTACCCTTACCCTTACCC-3'

The forward primer for the internal control *36B4 (RPL0)* gene was:

5'- CAGCAAGTGGGAAGGTGTAATCC-3'

The reverse primer for the internal control *36B4 (RPL0)* gene was:

5'- CCCATTCTATCATCAACGGGTACAA-3'

The standard used to generate the standard curve for telomeres was:

5'-$(TTAGGG)_{14}$-3'

The standard used to generate the standard curve for the internal control *36B4 (RPL0)* was:

5'- CAGCAAGTGGGAAGGTGTAATCCGTCTCCACAGACAAGGCCAGGACTCG TTTGTACCCGTTGATGATAGAATGGG-3'

ATRX Immunohistochemistry

Formalin-fixed, paraffin-embedded tissues were cut into 4-μm-thick sections and immunostained with a polyclonal antibody against ATRX (1:600; Sigma-Aldrich) by using heat-induced epitope retrieval and Leica Polymer Refine Detection Kit (Leica Microsystems) on a Leica Bond system after 15-minute antibody incubation.

Telomere FISH

Interphase FISH was performed on 4-μm-thick, formalin-fixed, paraffin-embedded tissue sections. The Cy3-labeled TelG probe (PNAbio) was co-denatured with the target cells on a hotplate at 90 °C for 12 minutes. The slides were incubated for 48 hours at 37 °C and then washed in 4 M Urea/2× SSC at 45 °C for 5 minutes. Nuclei were counterstained with DAPI (200 ng/mL) (Vector Labs).

Tumor Purity Estimations

For germline heterogeneous SNPs, loss of heterozygosity (LOH) measures the absolute difference between the mutant allele fraction in tumor and that in germline sample (0.5). LOH is the result of copy number alterations and/or copy neutral-LOH in tumor cells. Compared to copy number gains (a single copy gain in 100% tumor results in a LOH value of 0.167), regions with copy number loss showed stronger LOH (a single copy loss in 100% tumor result in a LOH value of 0.5). Consequently, we used LOH signals in copy neutral or heterozygous copy number loss regions (CNA value between [-1, 0]) to estimate tumor purity for all WGS samples. Briefly, a single copy loss in x% tumor cells resulted in an estimated CNA value of $-\frac{x}{100}$ and a LOH value of $\frac{x}{400-2x}$. Assuming the remaining LOH signal came from CN-LOH (CN-LOH in x% tumor cell resulted in a LOH value of $\frac{x}{200}$), the tumor content in a region could be estimated as the sum of the fraction with copy number loss and the fraction with CN-LOH by: $-CNA + 2 * \left(LOH - \frac{-CNA}{4-2CNA}\right)$. Using tumor content estimates from various regions within the genome, we performed an unsupervised clustering analysis using the *mclust* package (version 3.4.8) in R (version 2.11.1). The tumor purity of the sample was defined as the highest cluster center value among all clusters.

Purity Adjusted Mutant Allele Fraction (MAF) Estimation

MAF for validated SNVs was estimated as $\frac{\#Mutant\ reads}{(\#Total\ reads)\times(tumor\ purity)}$ using deep sequencing data. The frequency of SV was determined by a process of re-mapping all reads at breakpoints to both SV and non-SV templates using a BWA Smith-Waterman based approach. To do this, we use the assembled consensus sequence from CREST result as SV template. From comparison, a pool of non-SV templates were constructed by including: 1, directly pull out the flanking sequences of 100 bp of each side of the breakpoint from reference genome (GRCh37-Lite); 2, assemble non-SV reads around the breakpoint from the bam file, where non-SV reads were defined by: any non-duplicate, non-softclipped, reads that contains at least 10 bases mapped on each side of a breakpoint, and requiring for at least one of two sides of that breakpoint, all bases are mapped in the read within a 10 bp continuous window immediately next to the breakpoint. We then extracted all reads at both SV break points, together with any unmapped or partially mapped (soft-clipped) reads within 4 kb of the breakpoints, and perform a pair-wise mapping and comparison for the SV and each of the non-SV templates to determine the status of individual read. Reads only covering the breakpoint in the SV template, but not in the non-SV template, are considered as SV supporting reads, and verse versa. If there are any reads covering breakpoint in both SV and normal template, we calculated a local alignment score within a 10-bp window of the breakpoint from SV and normal templates, and chose the template with higher score. In the end, the statuses of every read from all pair-wise comparisons were summarized to generate a consensus status. Any reads with conflicting statuses, i.e., called as SV in one run and

non-SV in another run will be considered as "unknown". The SV mutant allele frequency was calculate by the ratio of number of SV reads to the total number of SV, non-SV, and unknown reads.

Tumor Heterogeneity Estimation

We used all validated autosomal SNVs satisfying the following criteria in heterogeneity analysis:

1) In copy neutral region (Log2ration between (-0.1, 0.1) in CNV analysis).
2) Not in regions with LOH (LOH value < 0.12 + min (0.08 purity*0.1).
3) With MAF > 0.05 or mutant allele count > 2.

We drew the kernel density estimate plot for MAFs of the qualifying SNVs using the *density* function in the *stat* package in R. For samples with at least 50 qualifying SNVs, we also estimated the number of significant peaks and the relative MAF component for each peak (peaks with less than 5 SNVs, peaks with less than 1% SNVs, and peaks with excessive variance were ignored). A sample with heterogeneity shows density peaks at a MAF smaller than 0.5 (the expected MAF assuming heterogeneous SNVs).

Kataegis Analysis

Kataegis analysis was performed on all validated Tier1-3 SNVs and SV breakpoints for each sample. The intermutation distance for a SNV was calculated as the distance to its nearest neighbor. For each SNV, its distance to the nearest validated SV breakpoint was also calculated. We defined microclusters of kataegis as clusters that contain at least 5 consecutive SNVs with inter-variant distance less than 10 kb. Mutant allele frequency

(MAF) was estimated for SNVs with at least 20X coverage in tumor BAMs based on deep sequencing of custom capture validation.

We derived copy number of SNVs from the CONSERTING analysis. For each SNV, the CNV segment covering the SNV was identified and the corresponding CN was estimated after tumor purity adjustment and rounded to the nearest integer.

Statistical evaluation of chromothripsis in OS tumors analyzed by WGS

Chromothripsis was described as localized chromosome shattering and repair occurs in a single event. The initial criterion is oscillation between restricted CNV states (Stephens *et al*. 2011(Stephens et al., 2011)), which were found in 4 OS tumors in this study. Most recently, Korbel and Campbell (Korbel and Campbell, 2013) proposed four potential criteria for assessing chromothripsis: 1) clustering of breakpoints; 2) randomness of DNA fragment joins; 3) randomness of DNA fragment order; and 4) ability to walk the derivative chromosome. Since randomness of DNA fragment order (Criterion 3) was not entirely valid even in Korbel and Campbell's own analysis, we decided not to evaluate this feature. For the 4 tumors in Supplementary Table 5, we performed Bartlett's goodness-of-fit test for exponential distribution to assess whether the distribution of SV breakpoints in each tumor departs from the null hypothesis of random distribution. A significant departure from random distribution supports clustering of SV breakpoints. To evaluate whether there is any bias in the DNA fragment joints categorized by the SV types (*i.e*. deletion, tandem duplication, head-to-head re-arrangements and tail-to-tail re-

arrangements), we applied goodness-of-fit test separately for inter- and intra-chromosomal events with a minimum of 5 SVs. A significant *p* value suggests biased fragment joins, which would *not* support chromothripsis. When both inter- and intra-chromosomal data are available, we reported the lower *p* value to represent a more conservative assessment of the random distribution for DNA fragment joints.

The significant chromothripsis regions were chromosome 14 in SJOS002_D (*p=2.09E-09)*, chromosome 17 in SJOS003_D (*p=9.65E-05)*, chromosome 6 in SJOS005_D (*p=1.75E-90)* and chromosome 13 in SJOS010_M (*p=2.21E-35)*.

REFERENCES

Aguirre, A. J., Brennan, C., Bailey, G., Sinha, R., Feng, B., Leo, C., Zhang, Y., Zhang, J., Gans, J. D., Bardeesy, N., *et al.* (2004). High-resolution characterization of the pancreatic adenocarcinoma genome. Proceedings of the National Academy of Sciences of the United States of America *101*, 9067-9072.

Bahrami, A., Dalton, J. D., Krane, J. F., and Fletcher, C. D. (2012). A subset of cutaneous and soft tissue mixed tumors are genetically linked to their salivary gland counterpart. Genes, chromosomes & cancer *51*, 140-148.

Castle, J. C., Biery, M., Bouzek, H., Xie, T., Chen, R., Misura, K., Jackson, S., Armour, C. D., Johnson, J. M., Rohl, C. A., and Raymond, C. K. (2010). DNA copy number, including telomeres and mitochondria, assayed using next-generation sequencing. BMC genomics *11*, 244.

Cawthon, R. M. (2002). Telomere measurement by quantitative PCR. Nucleic acids research *30*, e47.

Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., Mooney, T. B., Callaway, M. B., Dooling, D., Mardis, E. R., *et al.* (2012). MuSiC: identifying mutational significance in cancer genomes. Genome research *22*, 1589-1598.

Ding, L., Ellis, M. J., Li, S., Larson, D. E., Chen, K., Wallis, J. W., Harris, C. C., McLellan, M. D., Fulton, R. S., Fulton, L. L., *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature *464*, 999-1005.

Edmonson, M. N., Zhang, J., Yan, C., Finney, R. P., Meerzaman, D. M., and Buetow, K. H. Bambino: a variant detector and alignment viewer for next-generation sequencing data in the SAM/BAM format. Bioinformatics *27*, 865-866.

Huang, X. Q., Hardison, R. C., and Miller, W. (1990). A space-efficient algorithm for local similarities. Comput Appl Biosci *6*, 373-381.

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.

Korbel, J. O., and Campbell, P. J. (2013). Criteria for inference of chromothripsis in cancer genomes. Cell *152*, 1226-1236.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
O'Callaghan, N., Dhillon, V., Thomas, P., and Fenech, M. (2008). A quantitative real-time PCR method for absolute telomere length. BioTechniques *44*, 807-809.
Papai, Z., Feja, C. N., Hanna, E. N., Sztan, M., Olah, E., and Szendroi, M. (1997). P53 Overexpression as an Indicator of Overall Survival and Response to Treatment in Osteosarcomas. Pathology oncology research : POR *3*, 15-19.
Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res *29*, 308-311.
Stephens, P. J., Greenman, C. D., Fu, B., Yang, F., Bignell, G. R., Mudie, L. J., Pleasance, E. D., Lau, K. W., Beare, D., Stebbings, L. A*., et al.* (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27-40.
Wang, J., Mullighan, C. G., Easton, J., Roberts, S., Heatley, S. L., Ma, J., Rusch, M. C., Chen, K., Harris, C. C., Ding, L*., et al.* (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nature methods *8*, 652-654.
Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865-2871.
Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res *18*, 821-829.