Remarkably stable copy-number profiles in osteosarcoma revealed using single-cell DNA sequencing

Sanjana Rajan[1,2*], Simone Zaccaria[3,4,5*], Matthew V. Cannon[2], Maren Cam[2], Amy C. Gross[2], Benjamin J. Raphael[3,6#], Ryan D. Roberts[2,7,8#]

INSTITUTIONS:

1. Molecular, Cellular, and Developmental Biology Program, The Ohio State University, Columbus, Ohio, USA.
2. Center for Childhood Cancers and Blood Diseases, Abigail Wexner Research Institute at Nationwide Children's Hospital, Ohio, USA.
3. Department of Computer Science, Princeton University, Princeton, NJ, USA.
4. Computational Cancer Genomics Research Group, University College London Cancer Institute, London, UK
5. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK
6. Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA.
7. The Ohio State University James Comprehensive Cancer Center, Columbus, Ohio, USA.
8. Division of Pediatric Hematology, Oncology, and BMT, Nationwide Children's Hospital, Columbus, Ohio, USA

*Contributed equally

#Contributed equally

Financial support: This work was supported by funding provided by: NIH/NCI (K08CA201638, RDR), St Baldrick's Foundation Scholar Award (RDR), Hyundai Hope on Wheels Young Investigator Award (RDR), CancerFree Kids Foundation (RDR), Steps for Sarcoma Foundation

(RDR), Sarcoma Foundation of America (RDR), a Pelotonia Fellowship (SR), and an NIH CTSA Grant UL1TR002733.

Correspondence:

Ryan D. Roberts

Division of Pediatric Hematology, Oncology, and BMT

Nationwide, Children's Hospital

700 Children's Drive, Columbus, OH 43205

Email: Ryan.Roberts@nationwidechildrens.org

Phone: 614-722-29966; Fax: 614-355-2927

Disclosure of potential conflicts of interest: The authors declare no potential conflicts of interest.

Text words: 3559

Figures: 3

Supplementary figures and files: 6 figures, 2 files

Abstract

Osteosarcoma is an aggressive malignancy characterized by high genomic complexity. Identification of few recurrent mutations in protein coding genes suggests that somatic copy-number aberrations (SCNAs) are the genetic drivers of disease. Models around genomic instability conflict-it is unclear if osteosarcomas result from pervasive ongoing clonal evolution with continuous optimization of the fitness landscape or an early catastrophic event followed by stable maintenance of an abnormal genome. We address this question by investigating SCNAs in 12,019 tumor cells obtained from expanded patient tissues using single-cell DNA sequencing, in ways that were previously impossible with bulk sequencing. Using the CHISEL algorithm, we inferred allele- and haplotype-specific SCNAs from whole-genome single-cell DNA sequencing

data. Surprisingly, we found that, despite extensive genomic aberrations, cells within each tumor exhibit remarkably homogeneous SCNA profiles with little sub-clonal diversification. Longitudinal analysis between two pairs of patient samples obtained at distant time points (early detection, relapse) demonstrated remarkable conservation of SCNA profiles over tumor evolution. Phylogenetic analysis suggests that the bulk of SCNAs was acquired early in the oncogenic process, with few new events arising in response to therapy or during adaptation to growth in distant tissues. These data suggest that early catastrophic events, rather than sustained genomic instability, drive formation of these extensively aberrant genomes. Overall, we demonstrate the power of combining single-cell DNA sequencing with an allele- and haplotype-specific SCNA inference algorithm to resolve longstanding questions regarding genetics of tumor initiation and progression, questioning the underlying assumptions of genomic instability inferred from bulk tumor data.

Introduction

Osteosarcoma is the most common primary bone tumor affecting children and adolescents[1]. Nearly always high grade and aggressive, this disease exhibits extensive structural variation (SV) that results in a characteristically chaotic genome[2–4]. With few recurrent point mutations in protein coding regions, osteosarcoma is characterized by widespread somatic copy-number aberrations (SCNAs), the likely genomic driver of disease[5]. Indeed, osteosarcoma is the prototype tumor whose study led to the discovery of chromothripsis[6,7], a mutational process that causes the shattering of one or few chromosomes leading to localized genomic rearrangements causing extreme chromosomal complexity[8]. However, genomic complexity in osteosarcoma often goes beyond alterations caused by the canonical processes associated with chromothripsis [9,10]. Many have reasonably interpreted chromosomal complexity to be evidence of sustained chromosomal instability (CIN) in other cancers, often with additional supporting evidence[11–14]. Indeed, cancer-sequencing studies have identified the presence of extensive SCNAs as a marker for CIN[13].

There are two dominant models of copy number evolution in cancer: sequential acquisition of chromosomal changes where ongoing selection of advantageous phenotypes fosters intra-tumor heterogeneity, and punctuated evolution with rapid, discrete bursts where a small number of catastrophic events parsimoniously explain the origin of extreme complexity in many cancer genomes[15,16]. While a gradual tumor evolution model would help explain the high degree of intra-tumoral transcriptional heterogeneity exhibited by osteosarcoma[17,18], recent observations showing conservation of SCNA profiles between primary and metastatic lesions in osteosarcoma patients seem to support a general conservation of a highly aberrant genome[5,19]. Alterations in TP53 enable maintenance of an altered genome[16] and interestingly, Chen et al.[3] showed that over 90% of osteosarcomas have at least one allele of TP53 that is mutated. Overall, it remains unclear if SCNA profiles in osteosarcoma demonstrate intra-tumor heterogeneity emerging from continuous cycles of diversification and fitness optimization, or result from an early catastrophic event followed by maintenance of an abnormal, but otherwise stable, genome.

Most existing cancer sequencing studies have addressed questions regarding cancer evolution using bulk tumor sequencing, often from a single time point[20]. However, investigating ongoing clonal evolution from bulk sequencing data remains particularly challenging, as each bulk tumor sample is an unknown mixture of millions of normal and distinct cancer cells[21–24]. The emergence of single-cell genomic DNA sequencing technologies now permits scalable and unbiased whole-genome single-cell DNA sequencing of thousands of individual cells in parallel[21,25], providing the ideal framework for analyzing intra-tumor genomic heterogeneity and SCNA evolution. Recent computational advancements – most notably the CHISEL algorithm[24] – enable inference of allele- and haplotype-specific SCNAs in individual cells and sub-populations of cells from low coverage (<0.05x) single cell DNA sequencing. This allows assessment of intra-tumoral SCNA heterogeneity, identification of allele-specific alterations and reconstruction of the evolutionary

history of a tumor from thousands of individual cancer cells obtained at a single or multiple time points during tumor progression.

Here, we leverage these approaches to determine whether the widespread SCNAs in osteosarcoma result from ongoing genomic instability, providing a mechanism for tumor growth and evolution. Using expanded patient tissue samples, our studies revealed widespread aneuploidy and SCNAs in 12,019 osteosarcoma cells from ten tumor samples. Contrary to what we expected, we found negligible intra-tumor genomic heterogeneity, with remarkably conserved SCNA profiles both between the individual cells within a tumor and from tumors collected from the same patients at different therapeutic time points. These findings suggest that the widespread patterns of genomic SVs in osteosarcoma are acquired early in tumorigenesis, and the resulting patterns of SVs and SCNAs are stably preserved within an individual tumor, across treatment time, and through the metastatic bottleneck.

Results

Individual cells within a tumor bear surprisingly homogeneous copy number profiles despite extensive SCNAs

Single-cell DNA sequencing was performed on 12,019 tumor cells from expanded patient tissue samples. These patient tissues were obtained from diagnostic biopsies of localized primary tumors (n = 3), from post-chemotherapy resection procedures (n = 2), or from relapsed metastatic lung lesions (n = 4), representing the full spectrum of disease progression (Supplementary File S1, Supplementary File S2). With the exception of OS-17, a well-established model of metastatic osteosarcoma[26], all patient tissues were expanded for a single passage in mice as either subcutaneous flank tumors or as primary tibial (tibia) tumors to obtain sufficient tissue to perform single cell DNA sequencing (300-2500 single cells per sample; supplementary ***Figure S1***). We

used CHISEL[24] to identify allele- and haplotype-specific SCNAs from low-coverage (0.01-0.03x per cell) sequencing data.

Consistent with previous reports[6,27], these osteosarcoma genomes showed a high degree of aneuploidy and extensive SCNAs across the entire genome (***Figure 1***). In the presence of ongoing clonal evolution, we would expect to observe the presence of distinct groups of cancer cells with different complements of SCNAs within the same tumor, as shown in recent single-cell studies of different cancer types[21,24,28]. Surprisingly, in each of the ten samples investigated, we identified one dominant clone that comprised nearly all cells (78-100%) in each of the samples (supplementary ***Figure S1***). Moreover, we confirmed that the cells with noisy copy number profiles, that were discarded by CHISEL, bear SCNAs similar to the dominant clones identified in each sample – thus no rare clones with distinct copy number profiles were discarded (supplementary ***Figure S2***). Interestingly, we found that a substantial fraction of the overall copy-number changes involved allele-specific SCNAs, including copy-neutral LOHs (i.e., allele-specific copy numbers {2, 0}) that would have been missed by previous analyses of total copy numbers.

Genome-wide ploidy of single cells showed high variability, ranging from 1.5 to 4, demonstrating high degree of aneuploidy (supplementary ***Figure S3***). Consistent with the high levels of aneuploidy, we identified the presence of whole-genome doubling (WGD) across nearly all cancer cells of six tumors (NCH-OS-8, NCH-OS-17, NCH-OS-11, SJOS046149_X2, SJOS003939_X2 and SJOS003939_X1; ***Figure 1***A-C, H-J). Interestingly, whole-genome copy-number profiles were largely consistent when comparing two sets of paired patient tissue samples obtained at distant time points during tumor progression. The first set being NCH-OS-4 obtained after neoadjuvant therapy (two rounds of MAP chemotherapy), and NCH-OS-7 obtained after second relapse post-extensive treatment. We noted similar SCNV profiles irrespective of tumor growth location (orthotopic primary or subcutaneous flank tumor; ***Figure 1***E-G). The second set of paired primary and metastatic lesions (SJOS003939_X1, SJOS003939_X2) also showed SCNV profiles

that were highly similar, suggesting a high degree of conservation of genomic aberration profiles over therapeutic time. Overall, we observed surprising homogeneity within cancer cells sequenced from the same tumor. Even in tumors where small proportions of cells (5-20%) are classified as part of small subclones, these subclonal cells are only distinguished by few SCNAs in a small number of chromosomes. Thus, despite the high levels of aneuploidy and massive SCNAs identified in all ten samples, these osteosarcoma cells demonstrated negligible levels of intra-tumor heterogeneity.

Osteosarcoma cells harbor extensive SCNAs that mostly correspond to deletions.

The occurrence of WGD events is correlated with high levels of aneuploidy and higher frequency of SCNAs[29]. Recent reports have identified that WGDs serve as a compensatory mechanism for cells to mitigate the effects of deletions[30]. We investigated cell ploidy and fraction of genome affected by SCNAs (aberrant), amplifications, deletions, and subclonal CNAs between tumors affected by WGDs (NCH-OS-8, NCH-OS-17, NCH-OS-11, SJOS046149_X2, SJOS003939_X2 and SJOS003939_X1) and tumors not affected by WGDs (NCH-OS-10, NCH-OS-4 and NCH-OS-7). Osteosarcoma cells in all analyzed tumors demonstrate extensive SCNAs, affecting more than half of the genome in almost every cell. We found that the fraction of genome affected by SCNAs ranged from 50-70% on average (***Figure 2***A, supplementary ***Figure S4***A). This result might not be surprising for tumors affected by WGDs, however, we observed that tumors not affected by WGD had a high fraction of aberrant genome as well (higher than 50% on average; ***Figure 2***A). This aberrant fraction is substantially higher than has been reported for other cancer types[29].

We observed a clear enrichment of deletions among the identified SCNAs across all cancer cells. The fraction of genome affected by amplifications is 0-40% on average in every tumor, while the fraction of the genome affected by deletions is 40-100% on average across all cancer cells in every tumor (***Figure 2***B). This result is not particularly surprising for tumors with WGD events,

and is consistent with a recent study of Lopez et al.[30] that demonstrated a similar correlation in non-small-cell lung cancer patients. However, in the osteosarcoma tumors analyzed in this study, we found that cancer cells in non-WGD tumors are similarly affected by a high fraction of deletions (***Figure 2***B). Importantly, we observed that >80% of all cells in all but two of our samples harbored LOH events at the TP53 locus (in-line with frequency previously reported[3]) (supplementary ***Figure S5***). This substantiates the correlation between LOH of TP53 and high levels of genomic instability (including the occurrence of WGDs) reported in recent studies[13,30,31], and suggests that these events might have a critical role in the maintenance of a highly aberrant genomic state. Notably, CHISEL identified 50% of the samples to harbor copy-neutral LOH alterations at the TP53 locus that would have been missed by total copy number analyses (supplementary ***Figure S5***).

Interestingly, subclonal SCNAs that are only present in subpopulations of cancer cells and likely occurred late in the evolutionary process are relatively rare across all analyzed osteosarcoma tumors irrespective of WGD status (with a frequency of 0-20% in most cancer cells; ***Figure 2***C, supplementary ***Figure S5****B*). Note the only exceptions to this observation correspond to cells in NCH-OS-11, a sample with overall higher noise and variance, and a minority of cells in two other tumors (SJOS046149_X2, SJOS003939_X2) inferred to be pre-WGD cells, likely from errors in the inference of cell ploidy (supplementary ***Figure S4***B). Indeed, the average fraction of SCNAs in SJOS046149_X2, SJOS003939_X2 is lower than 20%. Overall, we observed that osteosarcoma cells investigated in these ten samples, whether passaged in cell culture over a few generations (OS-17), treatment naïve or exposed to extensive chemotherapy, bear high levels of aneuploidy marked with extensive deletions and negligible subclonal diversification, irrespective of WGD status.

Longitudinal single-cell sequencing shows modest evolution of SCNA from diagnosis to relapse

Increased aneuploidy has previously been associated with chromosomal instability (CIN) and accelerated tumor evolution[13,32], though some have suggested that this observation specifically applies to tumors that exhibit not only high levels of SCNA, but also high levels of subclonal SCNA[33]. To determine whether osteosarcoma exhibits CIN, we examined a pair of samples, NCH-OS-4 and NH-OS-7, collected at diagnosis and at relapse from the same patient to determine whether SCNAs remained stable or showed signs of significant instability/evolution. These were expanded for a single cycle in Icr-SCID mice either orthotopically (within the tibia) or subcutaneously (in the flank). We used CHISEL to jointly analyze 4238 cells from these paired tumor samples and to infer corresponding allele- and haplotype-specific SCNAs (***Figure 3***A). Based on existing evolutionary models for SCNAs, we reconstructed a phylogenetic tree that describes the evolutionary history of the different tumor clones identified in these tumors (***Figure 3***B). The result from this phylogenetic analysis confirmed our findings in two ways. First, we found that the evolutionary ordering of the different clones in the phylogenetic tree is concordant with the longitudinal ordering of the corresponding samples (***Figure 3***B): the tumor clones identified in the early sample (NCH-OS-4) correspond to ancestors of all the other tumor clones identified in later samples (NCH-OS-7-tib and NCH-OS-7-flank). Second, we observed that most of the SCNAs accumulated during tumor evolution are truncal, indicating that these aberrations are accumulated early during tumor evolution in the most common tumor ancestor and are shared across all the extant cancer cells (***Figure 3***B). In fact, only three large events distinguish the most common ancestor of all tumor cells (identified in NCH-OS-4) from the most common ancestor of only the relapse cancer cells: gain of chromosome 14, gain of chromosome 16q (resulting in copy-neutral LOH), and deletion of one allele of chromosome 18 (resulting in LOH). Many have previously shown that patient tissue expanded in a murine host for few generations maintain highly concordant genomic profiles with the parent tissue. Indeed, clones identified from patient tissue expanded in two different environments (NCH-OS-7-tib and NCH-OS-7-flank) are highly concordant and distinguished by only few focal SCNAs. Hierarchical clustering of clones identified

in another paired primary and metastatic lesion (SJOS003939_X1, SJOS003939_X2) also demonstrated negligible subclonal diversification (supplementary **Figure S*6***). Indeed, each of the two samples were dominated by one major clone, further strengthening our finding of unusually stable chromosomal aberrations in osteosarcoma cells over tumor evolution. Overall, using two separate paired tumor samples, we validated our finding of negligible heterogeneity in SCNA profiles both within tumors and across two distant time points in tumor evolution.

Discussion

Osteosarcoma tumors have highly chaotic genomic landscapes, dominated by SVs and SCNAs[3]. Genomic complexity in osteosarcoma and other cancers has often been interpreted as sign of ongoing genomic instability, suggesting that these tumors gradually accumulate changes that facilitate tumor growth and progression. The advent of single-cell DNA sequencing has allowed us to investigate intra-tumor genomic heterogeneity and tumor evolution in ways that were previously impossible with bulk sequencing methods[34–37]. In this study, we investigated SCNA profiles in osteosarcoma tumors at a single-cell resolution using expanded patient tissue. Cells within a tumor demonstrated surprisingly little cell-to-cell variability in SCNA profiles, challenging previous reports of intra-tumor genomic heterogeneity in osteosarcoma[17]. Using the CHISEL algorithm[24], we identified high levels of aneuploidy and extensive genomic aberrations – especially deletions – in osteosarcoma tumors. While many have previously shown WGD as a mechanism to mitigate the effect of deletions[30], we identified extensive deletions even in tumors not characterized by WGDs. This finding suggests that WGDs might not be essential for the stable maintenance of an aberrant genome but only a facilitator.

Analyzing two longitudinal sets of paired samples, we showed that osteosarcoma tumors maintain stable SCNA profiles from diagnosis to relapse. The clones identified on expansion in different microenvironments differed by only few SCNAs. In particular, phylogenetic analysis identified that the most recent common ancestor of these related samples harbored most of the observed

SCNAs, suggesting that genomic aberrations likely arise earlier in tumorigenesis followed by stable clonal expansion (clonal stasis)[16]. Multiple studies in osteosarcoma and other cancers have equated high SCNA with an ongoing genomic instability[27,38]. In contrast, our findings support the hypothesis that mechanisms leading to genomic alterations early in tumorigenesis are reverted and followed by an apparent stability. These findings are consistent with recent work identifying conservation of SCNA profiles between metastatic and primary osteosarcoma lesions[39]. Stable propagation of an aberrant genome requires non-functional p53[16] and recent work identified that p53 is inactivated in nearly all osteosarcoma tumors[3]. In this study, we identified that >85% of the sequenced cells in all but two samples harbored LOH of TP53, including widespread copy-neutral LOH alterations at the TP53 locus that were likely missed in previous total copy-number analyses.

Expanding patient tissue for one cycle in an animal host proved exceptionally useful for generating high-quality single-cell suspensions of sufficient quantity. Some may have concerns whether mouse-specific evolution selects for sub clonal populations. However, many have shown that while mouse-specific evolution occurs over multiple passages in patient derived xenograft models, early passages maintain high fidelity to patient tissue even in the most rapidly evolving models[40]. Therefore, this approach may represent a productive compromise between facilitating multiple lines of research on tissues with limited availability (rare diseases), and convenience while retaining fidelity to actual patient tumors.

While a much larger sample size of patient tissues is needed to capture the full heterogeneity of osteosarcoma seen in the human disease, our initial findings of clonal stasis in osteosarcoma has started to shed some light on the complex evolutionary history of this cancer type and could have important implications for tumor evolution, patient diagnosis and treatment of osteosarcoma. If genomic aberrations are indeed highly stable throughout the tumor mass, multi-region sampling may not be necessary to assess SCNAs as biomarkers. An early catastrophe model suggest that cancer cells are genomically hardwired early in tumor growth to become invasive, metastatic, or

resistant to therapy. This suggests that oncologists and scientists should focus efforts to understand tumor evolution on other mechanisms of evolution, such as transcriptional and epigenetic changes. While this study is agnostic of other alterations (such as SNVs) as a source of genomic variation, few recurrent mutations in osteosarcoma have been identified, despite extensive genetic analysis[5,38,41]. Interestingly, we and others have shown that osteosarcomas demonstrate intra-tumor transcriptional heterogeneity[17,18]; combined with the apparent absence of underlying genomic heterogeneity reported here, this suggests an intriguing possibility of an epigenetic basis of regulation.

Overall, our study highlights the power of combining single-cell DNA sequencing with an allele- and haplotype-specific SCNA analysis[24] to study the genomic heterogeneity and tumor evolution with the granularity missed using other methods and has implications in our understanding of tumor biology in other cancers characterized by extensive structural variation.

Methods

Experimental model – Expanded patient tissues and murine studies

Expanded patient tissue. Patient samples NCH-OS-4, NCH-OS-7, NCH-OS-8, NCH-OS-10 and NCH-OS-11 were obtained from patients consented under an Institutional Review Board (IRB)-approved protocol IRB11-00478 at Nationwide Children's Hospital (Human Subject Assurance Number 00002860). Germline whole genome sequencing (WGS) was generated from patient blood collected under IRB approved protocol IRB11-00478. Patient samples SJOS046149_X1, SJOS046149_X2, SJOS003939_X1 and SJOS031478_X2, with matched normal WGS were received from St. Jude's Children's Research Hospital through the Childhood Solid Tumor Network[42]. The OS-17 PDX was established from tissue obtained in a primary femur biopsy performed at St. Jude's Children's Research Hospital in Memphis and was a gift from Peter Houghton[26].

Murine Studies. Viable tissue fragments from patient tissue were expanded in C.B-17/IcrHsd-Prkdc^{scid} mice as subcutaneous tumors following approved IACUC protocols. These tumors were allowed to grow to 300-600 mm^3 before harvest. Passage 1 expanded tissue was used for all samples, with the exception of OS-17 (p18). For primary tibial tumors, single cell suspensions were injected intra-tibially. These tumors were harvested once they grew to 800 mm3.

Single-cell suspension and DNA library generation

Tumors harvested from mice were processed using the human tumor dissociation kit (Miltenyi Biotec, 130-095-929) with a GentleMacs Octo Dissociator with Heaters (Miltenyi Biotec, 130-096-427). Single cell suspensions in 0.04% BSA-PBS of dissociated tumor tissues were generated and frozen down using the 10X freezing protocol for CNV. The frozen down single-cell suspensions were processed using the Chromium Single Cell DNA Library & Gel Bead Kit (10X genomics #1000040) according to the manufacturer's protocol with a target capture of 1000-2000 cells. These barcoded single-cell DNA libraries were sequenced using the NovaSeq 6000 System using paired sequencing with a 100b (R1), 8b (i7) and 100b (R2) configuration and a sequencing coverage ranging from 0.01X to 0.05X (~0.02X on average) per cell. Germline WGS was performed on NovaSeq SP 2x150BP.

Single-cell CNV calling using CHISEL

Paired-end reads were processed using the Cell Ranger DNA Pipeline (10× Genomics), obtaining a barcoded BAM file for every considered single-cell sequencing dataset. As described previously[24], the pipeline consists of barcode processing and sequencing-reads alignment to a reference genome, for which we used hg19. We applied CHISEL (v1.0.0) to analyze each generated barcoded BAM file using the default parameters and by increasing to 0.12 the expected error rate for clone identification in order to account for the lower sequencing coverage of the analysed data[24]. In addition, we provided CHISEL with the available matched-normal germline

sample from each patient and phased germline SNPs according to the recommended pipeline by using Eagle2 through the Michigan Imputation Server with the Haplotype Reference Consortium (HRC) reference panel (v.r1.1 2016). CHISEL inferred allele- and haplotype-specific copy numbers per cell and used these results to group cells into distinct tumor clones, while excluding outliers and likely noisy cells. To determine fraction of aberrant genome (genome affected by SCNAs), we defined aberrant as any non-diploid genomic region (i.e., allele-specific copy numbers different than {1, 1}) in tumors not affected by WGDs (NCH-OS-10, NCH-OS-4, and NCH-OS-7) or any non-tetraploid genomic region (i.e., allele-specific copy numbers different than {2, 2}) in tumors affected by WGDs (NCH-OS-8, NCH-OS-17, NCH-OS-11, SJOS046149_X2, SJOS003939_X2 and SJOS003939_X1). We defined deletions as previously described in cancer evolutionary studies[23,43–45]. We say that a genomic region in a cell is affected by a deletion when any of the two allele-specific copy numbers inferred by CHISEL is lower than the expected allele-specific copy number (1 for non-WGD tumors or 2 for tumors affected by WGD). Conversely, a genomic region is amplified when any of the two allele-specific copy numbers is higher than expected.

Reconstruction of copy-number trees

We reconstruct copy-number trees for tumor samples NCH-OS-4 (tibia), NCH-OS-7 (flank) and NCH-OS-7 (tibia), to describe the phylogenetic relationships between distinct tumor clones inferred by CHISEL based on SCNAs using the same procedure proposed in previous studies[24]. Briefly, we reconstructed the trees using the maximum parsimony model of interval events for SCNAs[43,44] and the copy-number profiles of each inferred clones. These copy number profiles were obtained as the consensus across the inferred haplotype-specific copy numbers derived by CHISEL for all the cells in the same clone, where we also considered the occurrence of WGDs predicted by CHISEL. We classified copy-number events as deletions (i.e., del), as LOH which

are deletions resulting in the complete loss of all copies of one allele (loh), as copy-neutral LOH which are LOHs in which the retained allele is simultaneously amplified, and as gains (gain).

Data and code availability

All the processed data, scripts and results from CHISEL are available on GitHub at https://github.com/kidcancerlab/sc-OsteoCNAs

Figures

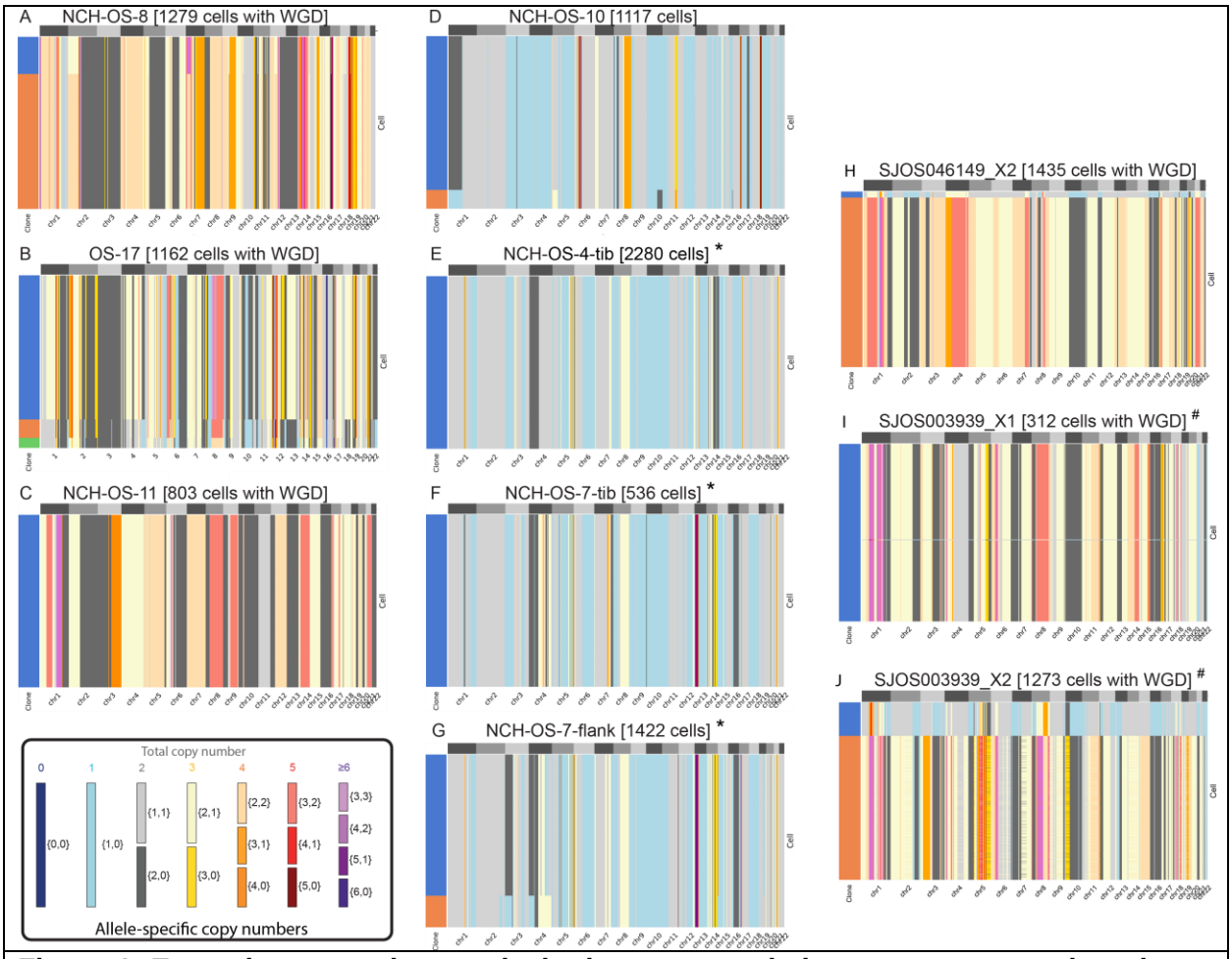

Figure 1. Extensive genomic complexity in ten expanded osteosarcoma patient tissue samples using single-cell DNA sequencing. Allele-specific copy numbers (heatmap colors) are inferred by using the CHISEL algorithm[24] from each of ten datasets including 300-2300 single cancer cells from osteosarcoma tumors. In each dataset, cancer cells are grouped into clones (colors in leftmost column) by CHISEL based on the inferred allele-specific copy numbers. Corrected allele-specific copy-numbers are correspondingly obtained by consensus. Note that cells classified as noisy by CHISEL have been excluded. '*' and '#' represent samples obtained from the same patient.

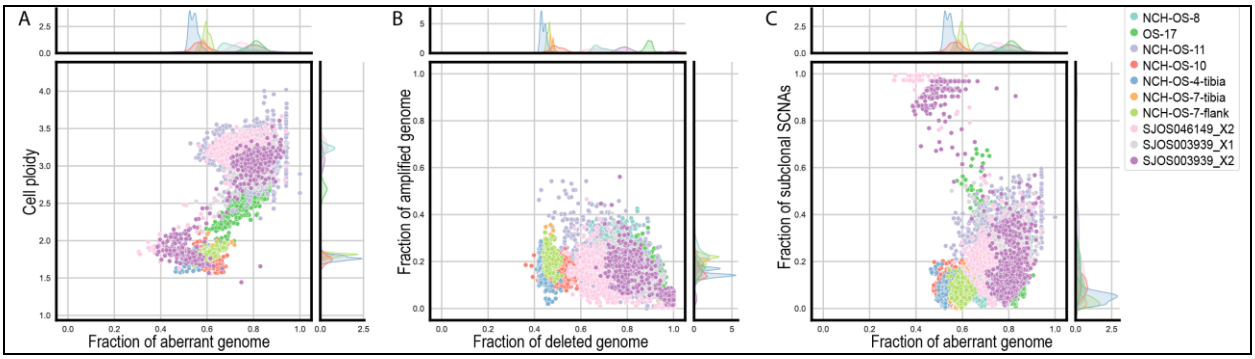

Figure 2. Osteosarcoma cancer cells exhibit extensive genetic alterations, especially deletions, but a relatively low level of heterogeneity. (**A**) Ploidy (y-axis) and fraction of aberrant genome (x-axis) of every cell (point) across the ten analyzed datasets (colors). The kernel density of the marginal distributions of each value is reported accordingly in every plot. (**B**) Fraction of genome affected by deletions (x-axis) vs. fraction of genome affected by amplifications (y-axis) of every cell (point) across the ten analyzed datasets (colors). (**C**) Fraction of aberrant genome (x-axis) and fraction of subclonal SCNAs (i.e. fraction of the genome with SCNAs different than the most common clone for the same region across all cells in the same dataset, y-axis) of every cell (point) across the ten analyzed datasets (colors).

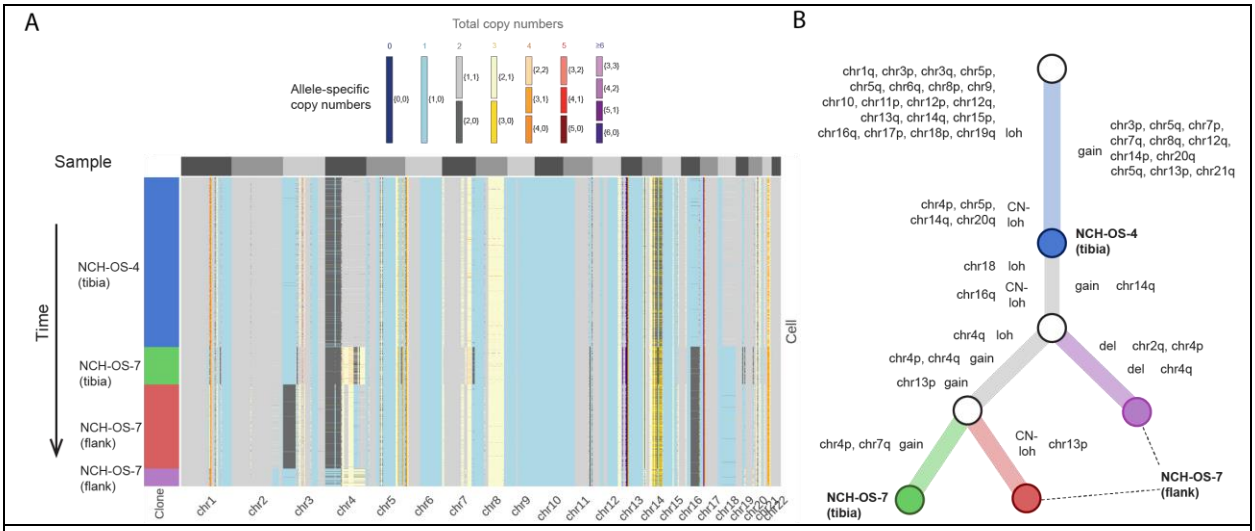

Figure 3. Phylogenetic reconstruction of tumor evolution is consistent with longitudinal ordering of matched tumor samples and reveals conservation of SCNA profiles. (A) Allele-specific copy numbers (heatmap colors) across all autosomes (columns) have been inferred by CHISEL jointly across 4238 cells (rows) in 3 tumor samples from the same patient: 1 pre-treatment sample (NCH-OS-4 tibia) and two post-treatment samples (NCH-OS-7 tibia and NCH-OS-7 flank). CHISEL groups cells into 4 distinct clones (blue, green, red, and purple) characterized by different complements of SCNAs. **(B)** Phylogenetic tree describes the evolution in terms of SCNAs for the four identified tumor clones. The tree is rooted in normal diploid clone (white root) and is characterized by two unobserved ancestors (white internal nodes). Edges are labelled with the corresponding copy-number events that occurred and transformed the copy-number profile of the parent into the profile of the progeny. The four tumor clones (blue, green, red, and purple) are labelled according to the sample in which they were identified.

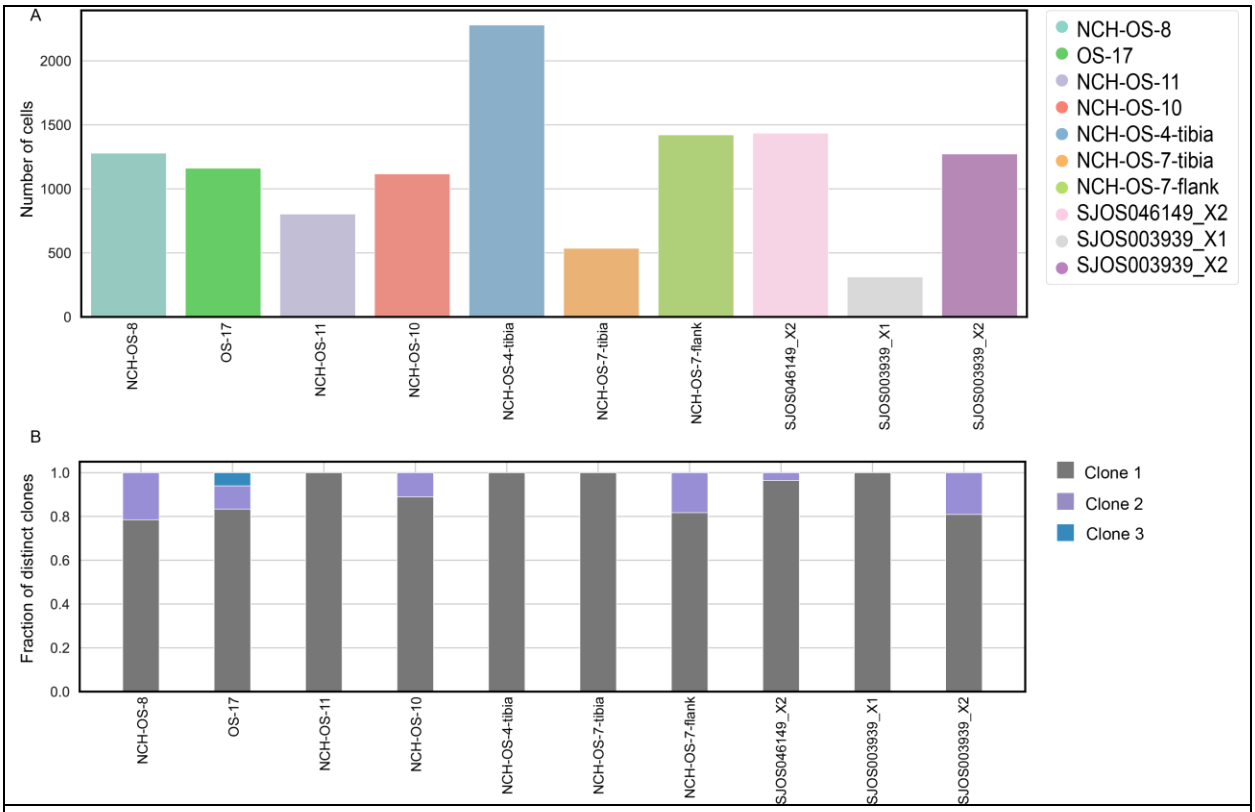

Figure S1. Majority of samples have one dominant clone, irrespective of total number of cells analyzed. (A) Number of cells sequenced and analyzed across all ten samples considered in this study. (**B**) Fraction of distinct clones identified in each dataset. Grey identifies most prevalent clone in each sample.

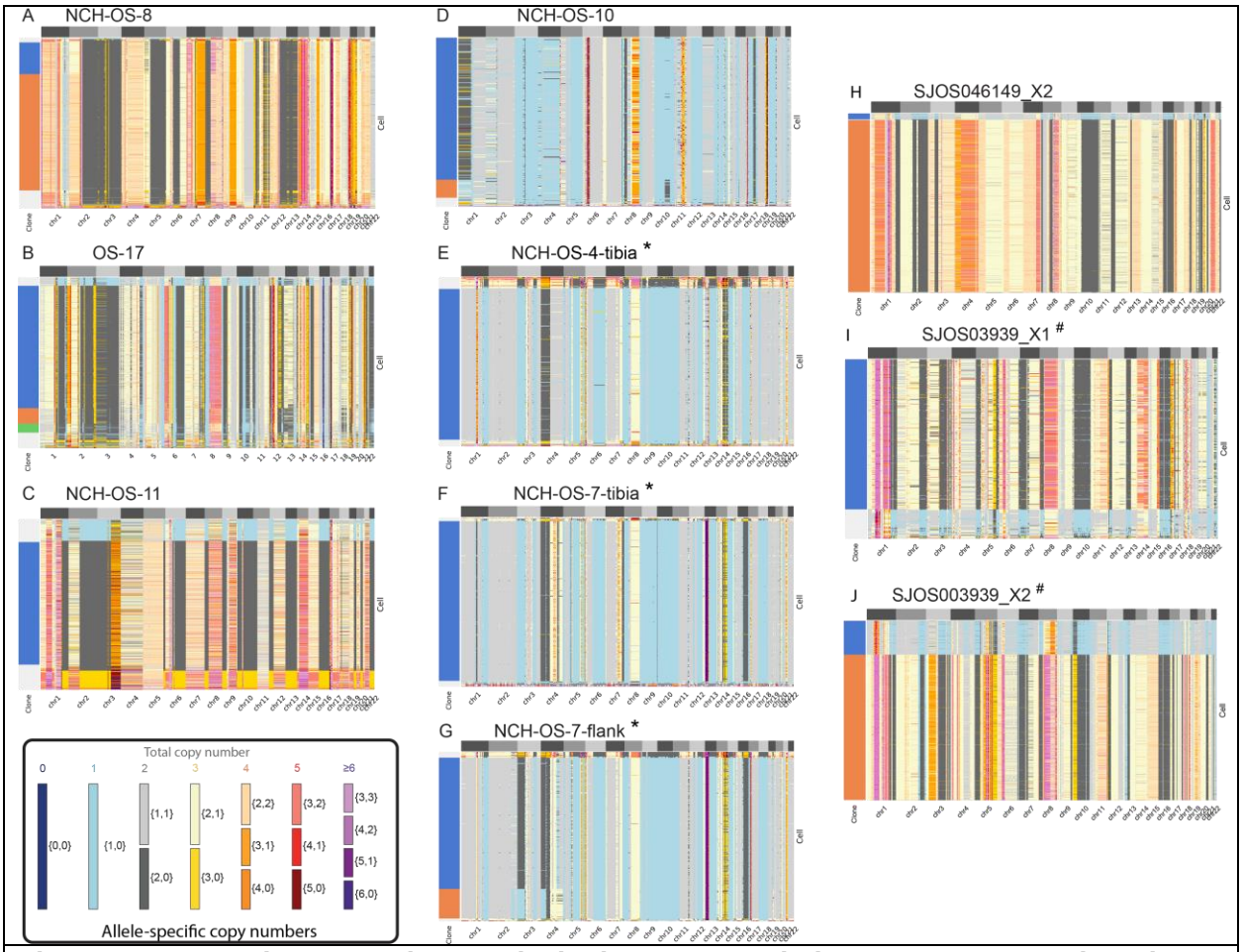

Figure S2. Extensive genomic complexity in ten expanded osteosarcoma patient tissue samples using single-cell DNA sequencing. Uncorrected single-cell allele-specific copy numbers (heatmap colors) are inferred by using the CHISEL algorithm[24] from each of ten datasets including 300-2300 single cancer cells from osteosarcoma tumors. In each dataset, cancer cells are grouped into clones (row colors) by CHISEL based on the inferred allele-specific copy numbers. Note that cells classified as noisy by CHISEL have been excluded. '*' and '#' represent samples obtained from the same patient.

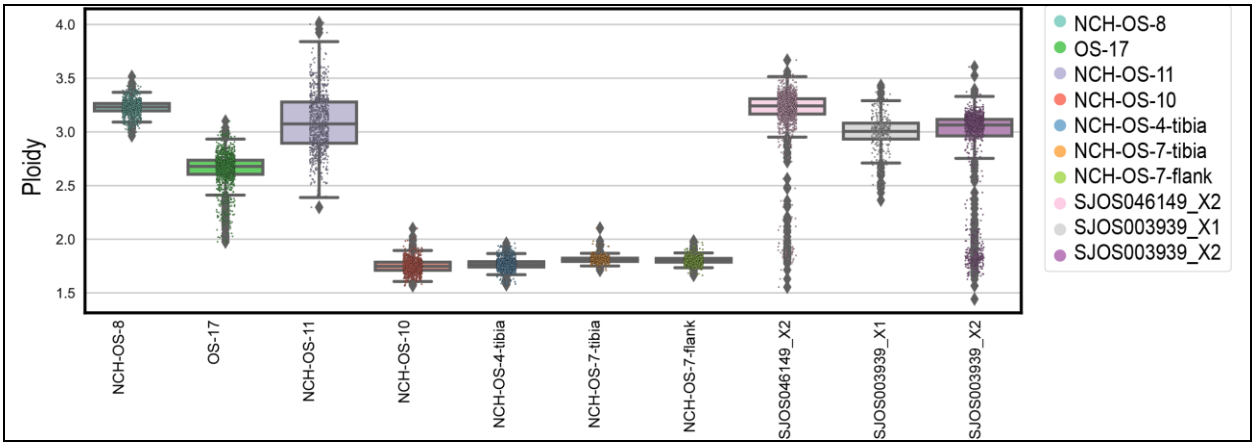

Figure S3. Osteosarcoma tumors demonstrate different and high levels of aneuploidy. Boxplots representing ploidy of individual cells (dots) in each sample (colors).

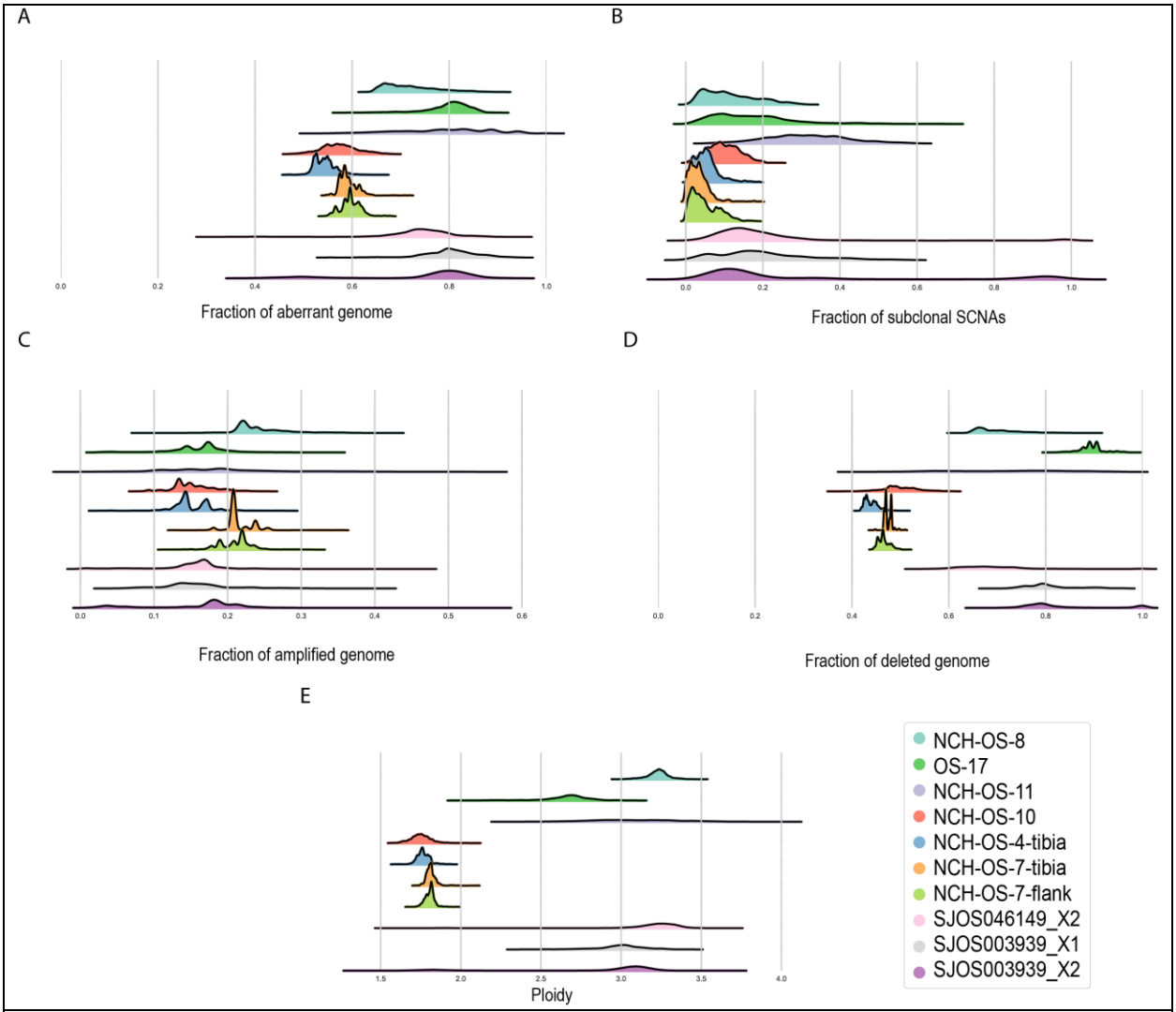

Figure S4. Extensive levels of aneuploidy, dominated by deletions, but low levels of subclonal diversification detected in osteosarcoma datasets. Density plots are computed across all cells in each analyzed sample (colors) to represent the distribution of the fraction of aberrant genome (**A**) fraction of subclonal SCNAs (**B**) fraction of amplified genome (**C**) fraction of deleted genome (**D**), and cell ploidy (**E**).

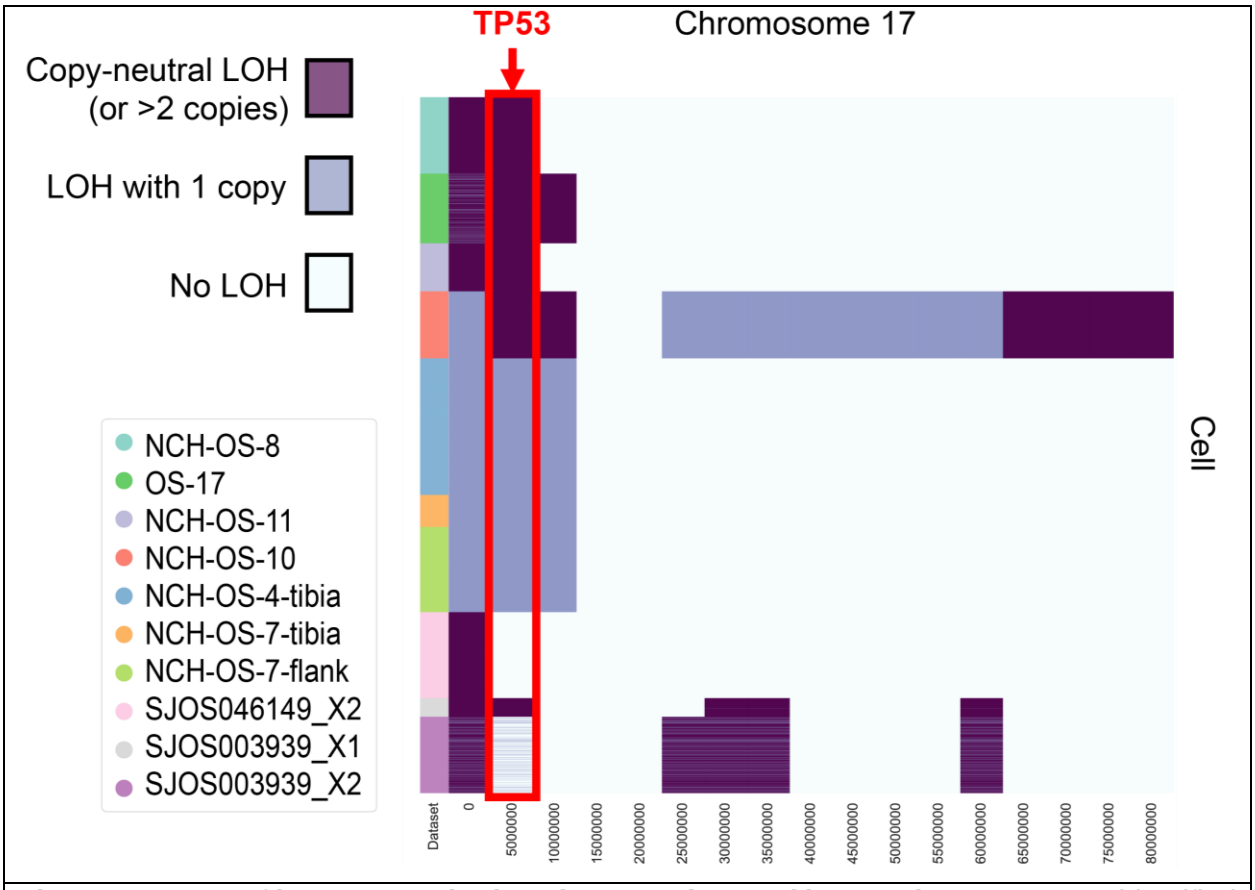

Figure S5. Loss of heterozygosity (LOH) events detected in TP53 locus. CHISEL identified LOH (light violet) and copy-neutral LOH (dark purple) events across all cells in the identified samples (colors) and along 5 Mb genomic regions of chromosome 17. The genomic region including gene TP53 is highlighted (red).

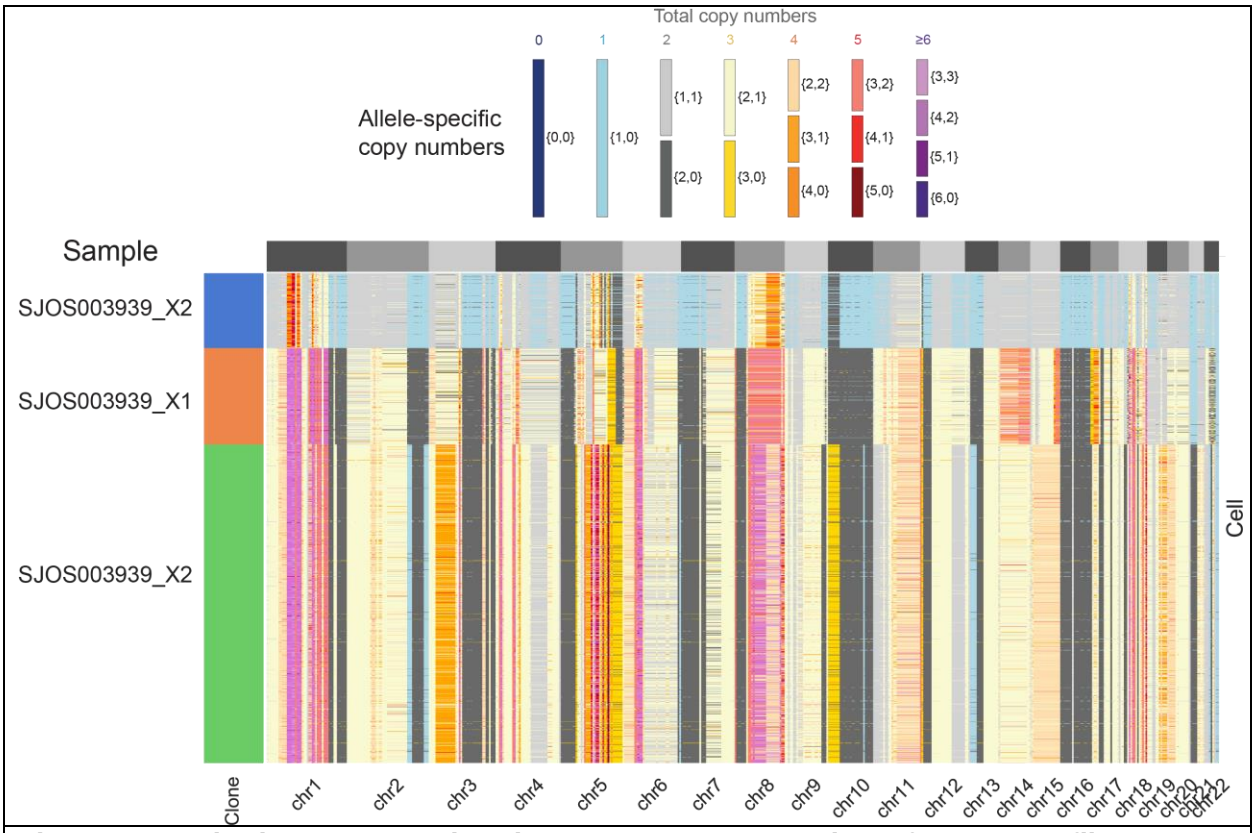

Figure S6. Paired tumor samples demonstrate conservation of SCNA profiles across therapeutic time points. Allele-specific copy numbers (heatmap colors) across all autosomes (columns) inferred by CHISEL in a pair of primary and metastatic tumor samples (SJOS003939_X1, SJOS003939_X2). CHISEL groups cells into three distinct clones (blue, green, orange) characterized by different complements of SCNAs. While there are small differences, majority of the SCNA profiles of clonal populations identified in each sample are conserved.

References

1. Casali, P. G. *et al.* Bone sarcomas: ESMO-PaedCan-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **29**, iv79–iv95 (2018).
2. Bridge, J. A. *et al.* Cytogenetic findings in 73 osteosarcoma specimens and a review of the literature. *Cancer Genet. Cytogenet.* **95**, 74–87 (1997).
3. Chen, X. *et al.* Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep.* **7**, 104–112 (2014).
4. Squire, J. A. *et al.* High-resolution mapping of amplifications and deletions in pediatric osteosarcoma by use of CGH analysis of cDNA microarrays. *Genes, Chromosom. Cancer* **38**, 215–225 (2003).
5. Sayles, L. C. *et al.* Genome-informed targeted therapy for osteosarcoma. *Cancer Discov.* **9**, 46–63 (2019).
6. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
7. Meyerson, M. & Pellman, D. Cancer genomes evolve by pulverizing single chromosomes. *Cell* vol. 144 9–10 (2011).
8. Li, Y. *et al.* Constitutional and somatic rearrangement of chromosome 21 in acute lymphoblastic leukaemia. *Nature* **508**, 98–102 (2014).
9. Behjati, S. *et al.* Recurrent mutation of IGF signalling genes and distinct patterns of genomic rearrangement in osteosarcoma. *Nat. Commun.* **8**, 1–8 (2017).
10. Perry, J. A. *et al.* Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5564–E5573 (2014).
11. Zhao, Y. *et al.* Single-cell RNA sequencing reveals the impact of chromosomal instability on glioblastoma cancer stem cells. *BMC Med. Genomics* **12**, 79 (2019).
12. Bakker, B. *et al.* Single-cell sequencing reveals karyotype heterogeneity in murine and human malignancies. (2016) doi:10.1186/s13059-016-0971-7.
13. Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126–132 (2020).
14. Bach, D. H., Zhang, W. & Sood, A. K. Chromosomal instability in tumor initiation and development. *Cancer Research* vol. 79 3995–4002 (2019).
15. Umbreit, N. T. *et al.* Mechanisms generating cancer genome complexity from a single cell division error. *Science (80-.).* **368**, (2020).
16. Gao, R. *et al.* Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
17. Zhou, Y. *et al.* Single-cell RNA landscape of intratumoral heterogeneity and immunosuppressive microenvironment in advanced osteosarcoma. *Nat. Commun.* **11**, 1–17 (2020).
18. Rajan, S. *et al.* Osteosarcoma tumors maintain intratumoral heterogeneity, even while

adapting to environmental pressures that drive clonal selection. *bioRxiv* 2020.11.03.367342 (2020) doi:10.1101/2020.11.03.367342.

19. 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).

20. Wang, Y. & Navin, N. E. Advances and Applications of Single Cell Sequencing Technologies. *Mol. Cell* **58**, 598 (2015).

21. Laks, E. *et al.* Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing. *Cell* **179**, 1207-1221.e22 (2019).

22. Tarabichi, M. *et al.* A practical guide to cancer subclonal reconstruction from DNA sequencing. *Nat. Methods 2021 182* **18**, 144–155 (2021).

23. Zaccaria, S. & Raphael, B. J. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat. Commun. 2020 111* **11**, 1–13 (2020).

24. Zaccaria, S. & Raphael, B. J. Characterizing allele- and haplotype-specific copy numbers in single cells with CHISEL. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0661-6.

25. Andor, N. *et al.* Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. *NAR Genomics Bioinforma.* **2**, (2020).

26. Houghton, P. J. *et al.* The pediatric preclinical testing program: Description of models and early testing results. *Pediatr. Blood Cancer* (2007) doi:10.1002/pbc.21078.

27. Martin, J. W., Squire, J. A. & Zielenska, M. The genetics of osteosarcoma. *Sarcoma* **2012**, 11 (2012).

28. Minussi, D. C. *et al.* Breast tumours maintain a reservoir of subclonal diversity during expansion. *Nature* **592**, 302–308 (2021).

29. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).

30. López, S. *et al.* Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat. Genet.* **52**, 283–293 (2020).

31. Bielski, C. M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).

32. Passerini, V. *et al.* The presence of extra chromosomes leads to genomic instability. *Nat. Commun.* **7**, 1–12 (2016).

33. Sheltzer, J. M. A transcriptional and metabolic signature of primary aneuploidy is present in chromosomally unstable cancer cells and informs clinical prognosis. *Cancer Res.* **73**, 6401–6412 (2013).

34. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–95 (2011).

35. Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).

36. Navin, N. E. The first five years of single-cell cancer genomics and beyond. *Genome Research* vol. 25 1499–1507 (2015).

37. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet. 2016 173* **17**, 175–188 (2016).

38. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).

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

40. Ben-David, U. *et al.* Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat. Genet. 2017 4911* **49**, 1567–1575 (2017).

41. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).

42. Stewart, E. *et al.* Orthotopic patient-derived xenografts of paediatric solid tumours. *Nature* **549**, 96–100 (2017).

43. Schwarz, R. F. *et al.* Phylogenetic Quantification of Intra-tumour Heterogeneity. *PLOS Comput. Biol.* **10**, e1003535 (2014).

44. El-Kebir, M. *et al.* Complexity and algorithms for copy-number evolution problems. *Algorithms Mol. Biol. 2017 121* **12**, 1–11 (2017).

45. ZaccariaSimone, El-KebirMohammed, W., K. & J., R. Phylogenetic Copy-Number Factorization of Multiple Tumor Samples. *https://home.liebertpub.com/cmb* **25**, 689–708 (2018).