Title Page

Subclonal somatic copy number alterations emerge and dominate in recurrent osteosarcoma

Authors

Michael D. Kinnaman[1], Simone Zaccaria[2,3], Alvin Makohon-Moore[4,5,6,7,8], Brian Arnold[9,10], Max Levine[1,11,12], Gunes Gundem[1,11], Juan E. Arango Ossa[1,11], Dominik Glodzik[1,11,13], M. Irene Rodríguez-Sánchez[1,14], Nancy Bouvier[1,15], Shanita Li[1], Emily Stockfisch[1], Marisa Dunigan[16], Cassidy Cobbs[16], Umesh Bhanot[17,18], Daoqi You[1], Katelyn Mullen[4,19], Jerry Melchor[4,5], Michael V. Ortiz[1], Tara O'Donohue[1], Emily Slotkin[1], Leonard H. Wexler[1], Filemon S. Dela Cruz[1], Meera Hameed[17], Julia L. Glade Bender[1], William D. Tap[20], Paul A. Meyers[1], Elli Papaemmanuil[1,11], Andrew L. Kung[1], Christine A Iacobuzio-Donahue[4,5,6]

Affiliations

1. Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
2. Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK
3. Computational Cancer Genomics Research Group, University College London Cancer Institute, London, UK
4. Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA
5. David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA
6. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, USA
7. Hackensack Meridian Health Center for Discovery and Innovation, Nutley, NJ, USA (current affiliation)
8. Georgetown University Lombardi Comprehensive Cancer Center, Washington, DC, USA (current affiliation)
9. Department of Computer Science, Princeton University, Princeton, NJ, USA
10. Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA
11. Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA
12. Isabl, New York, NY, USA (current affiliation)
13. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA (current affiliation)
14. Wunderman Thompson Health, New York, NY, USA (current affiliation)
15. IT and Digital Initiatives, Memorial Sloan Kettering Cancer Center, New York, NY, USA (current affiliation)
16. Integrated Genomics Operation Core, Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA
17. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

18. Precision Pathology Biobanking Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA
19. Gerstner Sloan Kettering Graduate School of Biomedical Sciences, New York, NY, USA
20. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Running Title: Clonal heterogeneity in osteosarcoma

Keywords: osteosarcoma, intratumor heterogeneity, clonal evolution, somatic copy number alterations, whole-genome sequencing

Funding Sources

- **Award Group**
  - Funder: Rally Foundation
  - Principal Award Recipient(s): Michael Kinnaman
- **Award Group**
  - Funder: Hyundai Hope on Wheels
  - Principal Award Recipient(s): Michael Kinnaman
- **Award Group**
  - Funder: Conquer Cancer Foundation of the American Society of Clinical Oncology/Quad W Foundation
  - Principal Award Recipient(s): Michael Kinnaman
- **Award Group**
  - Funder: National Cancer Institute (NCI)
  - Award Id(s): K12
  - Principal Award Recipient(s): Michael Kinnaman
- **Award Group**
  - Funder: National Institutes of Health
  - Award Id(s): R00 (R00CA229979)
  - Principal Award Recipient(s): Alvin Makohon-Moore

Corresponding Authors: Michael Kinnaman, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065. Phone: 212-639-8856; Fax: 929-321-7101; E-mail: kinnamam@mskcc.org; and Christine A. Iacobuzio-Donahue, Memorial Sloan Kettering Cancer Center, 417 East 68th Street, New York, NY 10065. Phone: 646-888-2239; Fax: 646-888-3235; E-mail: iacobuzc@mskcc.org

Author's Disclosures

M. D. Kinnaman reports grants from the Rally Foundation, Hyundai Hope on Wheels, Conquer Cancer Foundation of the American Society of Clinical Oncology/Quad W Foundation, and the NCI during the conduct of the study. S. Zaccaria is a Cancer Research UK Career Development Fellow (Award Reference RCCCDF-Nov21\100005) and is also supported by Rosetrees Trust (Grant Reference M917). A. Makohon-Moore reports research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R00CA229979 (A.M-M). The content is solely the responsibility of the authors and

does not necessarily represent the official views of the National Institutes of Health. M. Levine reports equity in Isabl and has filed IP relating to whole genome analytics. D. Glodzkik is a consultant for Repare Therapeutics and equity holder in MNM Biosciences Inc. W. D. Tap reports consulting fees from Eli Lilly, Daiichi Sankyo, Deciphera, Foghorn Therapeutics, AmMAx Bio, Novo Holdings, Servier, Medpacto, Ayala Pharmaceuticals, Kowa Research Institute, Epizyme (Nexus Global Group), Bayer, Cogent Biosciences, Amgen, Physicians' Education Resource, and BioAlta on the development of novel agents for the treatment of patients with sarcomas; has a patent Companion Diagnostic for CDK4 inhibitors (14/854,329) for the treatment of liposarcoma; holds stock options in and is a member of the scientific advisory boards for Certis Oncology Solutions, a life science technology company working on orthotopic patient-derived xenografts in oncology; and holds stock options in and is a co-founder and director of Atropos Therapeutics, a company that focuses on a drug discovery platform to identify chemical matter that can modulate the formation of senescent cells in relation to ageing and oncology. None of WDT's competing interests are related to the topic discussed here. WDT is the 2022 President of the Connective Tissue Oncology Society. E. Papaemmanuil is a founder, equity holder and has a fiduciary role in Isabl, a cancer whole genome sequencing analytics company. A.L. Kung is on the Scientific Advisory Board of Emendo Biotherapeutics, Karyopharm Therapeutics, Imago BioSciences, and DarwinHealth, is co-founder and on the board of directors of Isabl Technologies, and has equity interest in Imago BioSciences, Emendo Biotherapeutics, and Isabl Technologies. This research was also supported by the Cancer Center Support grant P30 CA008748. No disclosures were reported by the other authors.

Total Word Count: 5952
Figures: 6 main figures, 10 supplemental figures

Abstract

Multiple large-scale tumor genomic profiling efforts have been undertaken in osteosarcoma, however, little is known about the spatial and temporal intratumor heterogeneity and how it may drive treatment resistance. We performed whole-genome sequencing of 37 tumor samples from eight patients with relapsed or refractory osteosarcoma. Each patient had at least one sample from a primary site and a metastatic or relapse site. We identified subclonal copy number alterations in all but one patient. We observed that in five patients, a subclonal copy number clone from the primary tumor emerged and dominated at subsequent relapses. *MYC* gain/amplification was enriched in the treatment-resistant clone in 6 out of 7 patients with more than one clone. Amplifications in other potential driver genes, such as *CCNE1*, *RAD21*, *VEGFA*, and *IGF1R*, were also observed in the resistant copy number clones. Our study sheds light on intratumor heterogeneity and the potential drivers of treatment resistance in osteosarcoma.

Significance

Subclonal copy number clones emerged and dominated in relapsed osteosarcoma, with *MYC* gain/amplification being the defining characteristic in our cohort. Selective pressure from neoadjuvant chemotherapy revealed this clone at the time of primary resection, highlighting that genomic profiling at this time may identify clones that are selected for, or determine innate resistance to primary chemotherapy.

131 **Introduction**

132 Osteosarcoma is an aggressive bone tumor which primarily affects children and young
133 adults. Patients who present with metastatic disease at diagnosis have a poor overall prognosis
134 and those with an inferior response to neoadjuvant chemotherapy have a high risk for
135 recurrence[1–3]. Multiple large-scale tumor genomic profiling efforts have been undertaken to
136 describe the genomic underpinnings and identify new potential therapeutic targets for
137 osteosarcoma[4–9]. These studies revealed that osteosarcoma, typically characterized by a high
138 degree of chromosomal instability, has a large number of chromosomal deletions, translocations,
139 and amplifications. The common alterations present in osteosarcoma primarily involve tumor
140 suppressor genes (e.g., *TP53*, *RB1*, *PTEN*), whereas targetable activating mutations are rare,
141 making it challenging to link the mutational genotype to a broadly applicable treatment
142 strategy[10,11]. However, recent studies have suggested that targeting focal gene amplifications in
143 consensus driver genes may be an effective strategy for identifying precision-based therapies[4,12].

144 Recent genomic studies in osteosarcoma have suggested that metastatic clones do not
145 correspond to the dominant clones present in the primary tumor,[13] and osteosarcoma may evolve
146 via parallel evolution,[14] with evidence for both monoclonal[15] and polyclonal synchronous
147 seeding of metastases[13,14]. Copy number alterations in consensus driver genes, such as *MYC* and
148 *CDK4* were found to be likely early events[15]. Cisplatin-induced mutagenesis has also been
149 highlighted as a potential driver of treatment resistance in recurrent osteosarcoma[15]. Despite
150 these initial insights into the clonal heterogeneity of osteosarcoma, the extent to which
151 neoadjuvant chemotherapy affects clonal selection in patients with a poor response to
152 chemotherapy and the degree to which copy number alterations evolve from diagnosis to relapse
153 remains unclear.

154 To address these open questions about clonal selection and heterogeneity in
155 osteosarcoma, we performed whole-genome sequencing of 37 spatially and temporally separated
156 tumor samples from eight patients with osteosarcoma who had a poor response to neoadjuvant
157 chemotherapy (<90% necrosis). We describe spatial intermetastatic heterogeneity and temporal
158 clonal evolutionary processes, with a focus on identifying and tracking unique copy number
159 clones from diagnosis through relapse.

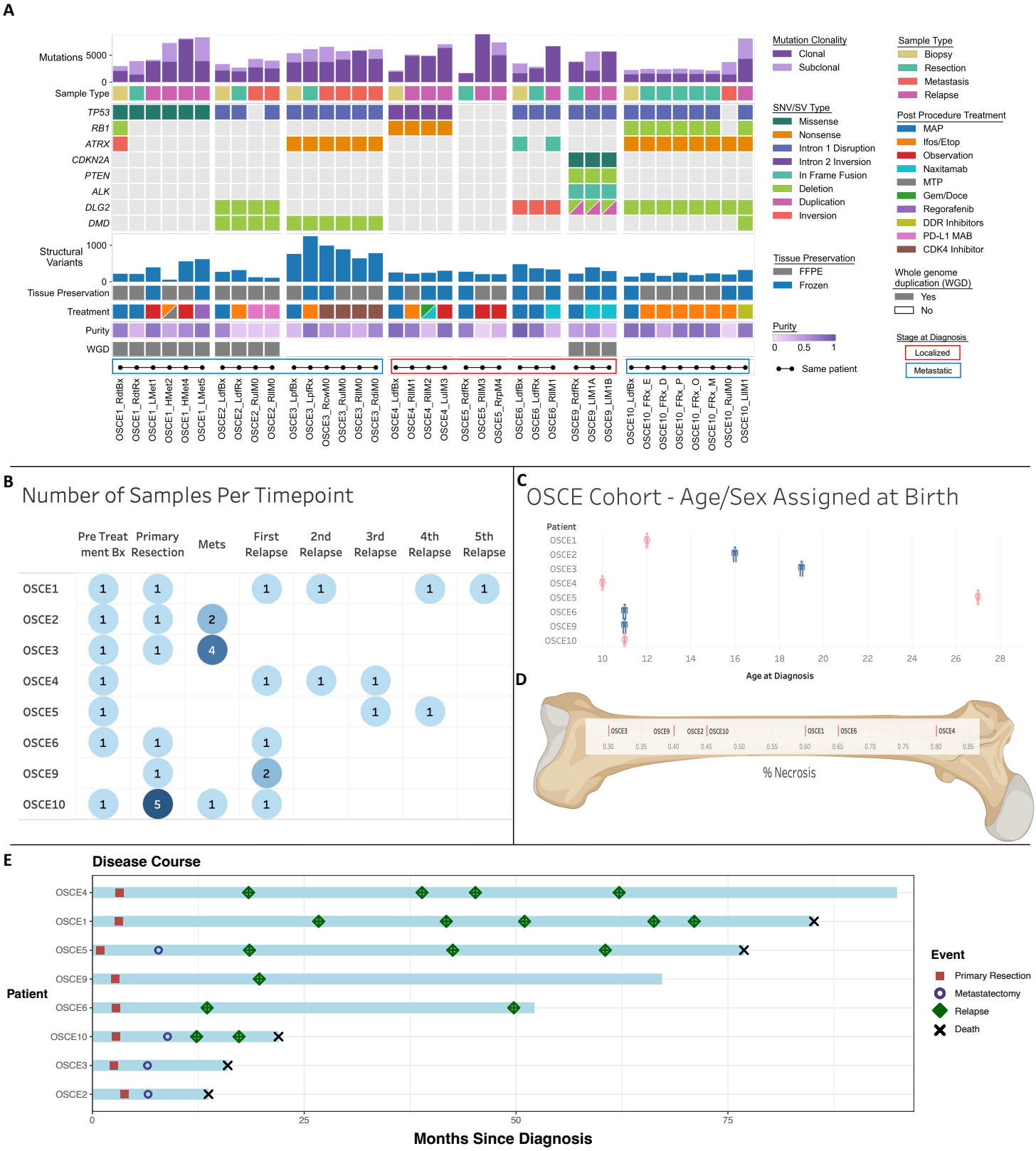
160 **Results**

161 **Analysis of Single Nucleotide Variants reveal limited driver gene heterogeneity in**
162 **temporally and spatially distinct osteosarcoma sample**

163 To analyze clonal evolution and intratumoral heterogeneity in osteosarcoma, we
164 performed whole-genome sequencing (WGS) of tumor tissues from multiple spatially and
165 temporally distinct samples from eight individuals with relapsed/refractory osteosarcoma. DNA
166 was extracted from 84 samples collected from 10 patients. After initial quality control, we
167 sequenced 62 unique tumor samples from eight patients with WGS to a target depth of 80x
168 (Supplementary Fig. 1). Of these eight patients, four had localized disease at diagnosis (OSCE4,
169 OSCE5, OSCE6, OSCE9) and four had metastatic disease at diagnosis (OSCE1, OSCE2,
170 OSCE3, OSCE10) (Fig. 1A), and the age at diagnosis was 11-27 years (four girls and four boys,
171 Fig. 1C). All patients were treated at the Memorial Sloan Kettering Cancer Center and received
172 methotrexate, cisplatin, and doxorubicin (MAP) chemotherapy (subsequent post-procedure
173 treatment, Fig. 1A). Seven of the eight patients had a poor response to neoadjuvant
174 chemotherapy (<90% necrosis at the time of primary resection, Fig. 1D), while OSCE5 had an
175 upfront resection; therefore, the response to therapy could not be evaluated.

176 After sequencing was completed, we reviewed the quality of the sequencing data (purity,
177 sequencing coverage) and determined that 37 of the 62 (59%) samples, similar to other studies[16],
178 met our quality control requirements to proceed with further downstream analysis
179 (Supplementary Fig. 1). Of these 37 samples, 17 came from primary sites, with seven of eight
180 patients having a pretreatment sample, six of which were biopsies and one (OSCE5), which was
181 a pretreatment primary resection (Fig. 1A/B). Of the seven patients who did not undergo an

Figure 1


A
Mutations
Sample Type
TP53
RB1
ATRX
CDKN2A
PTEN
ALK
DLG2
DMD
Structural Variants
Tissue Preservation
Treatment
Purity
WGD
Mutation Clonality
Clonal
Subclonal
SNV/SV Type
Missense
Nonsense
Intron 1 Disruption
Intron 2 Inversion
In Frame Fusion
Deletion
Duplication
Inversion
Tissue Preservation
FFPE
Frozen
Purity
Same patient
Sample Type
Biopsy
Resection
Metastasis
Relapse
Post Procedure Treatment
MAP
Ifos/Etop
Observation
Naxitamab
MTP
Gem/Doce
Regorafenib
DDR Inhibitors
PD-L1 MAB
CDK4 Inhibitor
Whole genome duplication (WGD)
Yes
No
Stage at Diagnosis
Localized
Metastatic
B
Number of Samples Per Timepoint
Pre Treatment Bx
Primary Resection
Mets
First Relapse
2nd Relapse
3rd Relapse
4th Relapse
5th Relapse
OSCE1
OSCE2
OSCE3
OSCE4
OSCE5
OSCE6
OSCE9
OSCE10
C
OSCE Cohort - Age/Sex Assigned at Birth
Patient
Age at Diagnosis
D
% Necrosis
E
Disease Course
Patient
Months Since Diagnosis
Event
Primary Resection
Metastatectomy
Relapse
Death

182

183 **Figure 1.** Characteristics of the patients and samples included in the analysis cohort. **A,**
184 Oncoprint of sample and patient level details for each patient. Samples from the same patient are
185 connected by dots and lines on bottom of figure and are in chronological order of time obtained
186 (earliest on left). Sample Name Key: Bx = biopsy, Rx = Resection, Sample Ending in 0 =
187 metastatic site present at diagnosis, sample ending in a number >0 indicates number of relapses.
188 L/R = laterality, d = distal, p=proximal, f=femur, t=tibia, cw=chest wall, H=heart, ul=upper lobe,
189 ll=lower lobe, di=diaphragm, rp=retroperitoneal, l=lobe. **B,** Summary of the number of samples
190 per timepoint for each patient, darker shades of blue represent higher number of samples at a
191 respective time point. **C,** Age and sex assigned at birth for each patient, patient is on the x-axis,
192 age is on the y-axis, and sex assigned at birth is plotted on the chart as blue for male and pink for
193 female. **D,** Percent necrosis at time of primary resection for each patient (Note OSCE2/OSCE10
194 both have 45% necrosis). **E,** Patients are listed on the y-axis and are ordered from longest disease
195 course at the top to shortest at the bottom of the figure. The light blue bars represent length of
196 disease course in months. Events are marked as depicted in the legend with different shapes and
197 colors and plotted along the disease course bar at the time in months that the event occurs.

198 upfront resection, all had at least one on-therapy resection sample, and one patient (OSCE10)
199 had multi-region sampling from the primary tumor (Fig. 1A/B). The other 20 samples came from
200 metastatic sites, 15 of which were from lung metastases, and 7 of the 20 were metastatic sites
201 that were present at diagnosis (Fig. 1A/B). Fresh frozen samples accounted for 15 of the 37
202 samples selected for downstream analysis, while the remaining 22 were formalin fixed paraffin
203 embedded (FFPE) samples (Fig. 1A). The median purity of fresh frozen samples was 0.75
204 compared to 0.46 of FFPE samples (Fig. 1A). All eight patients in this analysis had matched
205 normal blood samples sequenced at a target depth of 40x.

206 The clinical courses of OSCE2, OSCE3, and OSCE10 were defined as refractory disease
207 with progression on MAP (methotrexate, doxorubicin, cisplatin) chemotherapy and extremely
208 virulent disease courses, with time from diagnosis to death of 1.08, 1.3, and 1.75 years
209 respectively (Fig. 1E). OSCE1 and OSCE5 had long protracted relapsing and remitting disease
210 courses, with a time from diagnosis to death of seven and six years, respectively (Fig. 1E).
211 OSCE4 also had a relapsing and remitting disease course but is currently in remission eight years
212 after diagnosis (Fig. 1E). OSCE6 is alive 4.5 years after diagnosis but has had a recent
213 recurrence (Fig. 1E). OSCE9 is 4 years out from the initial metastatic recurrence (Fig. 1E).

214 After filtering and germline subtraction, whole-genome sequencing data identified
215 between 1684 and 16,215 single nucleotide variants (SNVs) per sample (Fig. 1A). Of these, there
216 were between 12 and 181 coding nonsynonymous SNVs per sample (median = 54,
217 Supplementary Fig. 2A). The average number of nonsynonymous SNVs in primary site samples
218 was 42, compared to 73 in metastatic or relapsed samples. SNVs were clustered across all
219 samples for each patient using the DeCiFer[17] algorithm, which determines the descendant cell
220 fraction (DCF) of all SNVs for a given cluster in each patient (analogous to the cancer cell

221 fraction but accounting for potential mutation losses[17]). Following previous approaches[17], SNVs
222 were categorized as clonal if they belonged to a cluster in a sample with a DCF ≥ 90%, and
223 subclonal if they belonged to a cluster with a DCF < 90%. The proportion of clonal SNVs ranged
224 from 35.7-100% across all samples (median = 68.2%), with relapse samples having the highest
225 proportion (median = 92.7%) of clonal SNVs compared with biopsy, resection, and metastatic
226 samples (median = 66.1%, 64.7%, and 63.6%, respectively) (Fig. 1A and Supplementary Fig.
227 1B). Likely functional driver gene SNVs[18] were identified in five of the eight patients, including
228 genes known to be frequently mutated in osteosarcoma, such as *TP53*, *ATRX*, *RB1*, and *CDKN2A*
229 (Fig. 1A)[5–9]. These driver genes were clonal (shared) across all samples for each patient, and no
230 new SNVs were likely functional drivers that were unique to any metastatic or recurrent samples
231 (Fig. 1A, Supplementary Fig. 3A).

232 SNVs were clustered across all samples for each patient using the DeCiFer[17] algorithm,
233 which determines the descendant cell fraction (DCF) of all SNVs for a given cluster in each
234 patient (analogous to the cancer cell fraction but accounting for potential mutation losses[17]).
235 Following previous approaches[17], SNVs were categorized as clonal if they belonged to a cluster
236 in a sample with a DCF ≥ 90% and subclonal if they belonged to a cluster with a DCF < 90%.
237 The proportion of clonal SNVs ranged from 35.7-100% across all samples (median = 68.2%),
238 with relapse samples having the highest proportion (median = 92.7%) of clonal SNVs compared
239 with biopsy, resection, and metastatic samples (median = 66.1%, 64.7%, and 63.6%,
240 respectively) (Fig. 1A and Supplementary Fig. 1B).

241 Since there were no new SNVs in consensus driver genes unique to relapse or metastatic
242 samples, we next examined structural variations and copy number alterations. Structural variants
243 (SVs) in consensus driver genes were shared across all samples for each patient. *TP53* structural

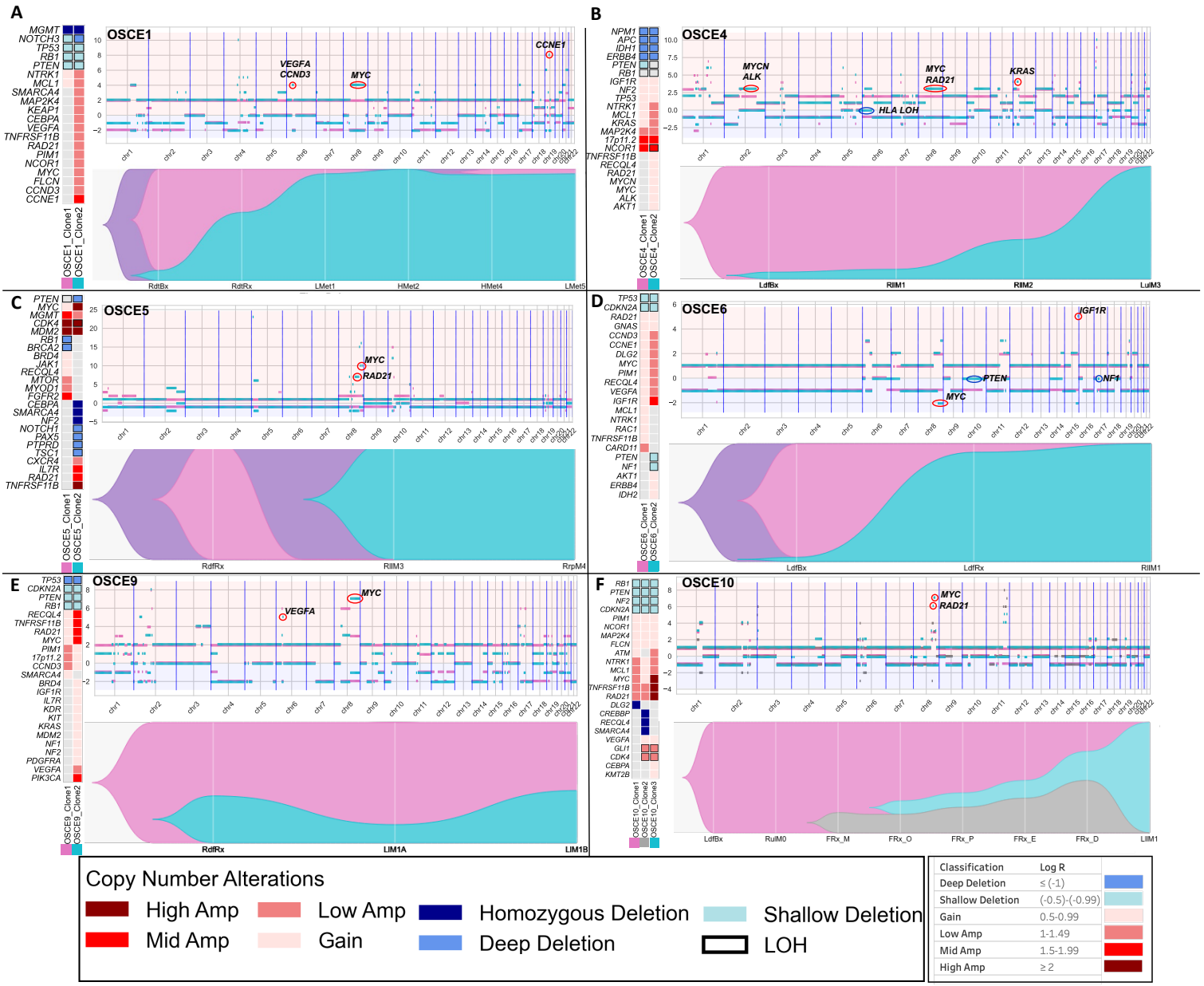
244 variants involving intron 1 were observed in 5/8 patients (OSCE2, OSCE3, OSCE6, OSCE9, and
245 OSCE10; Fig. 1A). A *TP53* intron 2 inversion was observed in OSCE4 (Fig. 1A). There was a
246 *TP53* SNV in OSCE1 and while there was no *TP53* SNV or SV found in OSCE5 (Fig. 1B), there
247 was amplification of *MDM2*, an important negative regulator of *TP53*. In OSCE1, there was a
248 deletion event in *RB1* and an inversion in *ATRX* in the pretreatment sample (RdtBx) that was not
249 seen in the primary resection or relapse samples (Fig. 1A). Disruptions in *DLG2*, a bone tumor
250 supressor gene[19], were observed in 4 patients (OSCE2, OSCE6, OSCE9, and OSCE10; Fig. 1A).
251 Deletion events in *DMD*, a gene that has been linked to aggressive behavior in human cancers,
252 and is believed to have a potential role as a tumor suppressor, were observed in the three
253 refractory cases (OSCE2, OSCE3, and OSCE9) and were present in all samples for OSCE2 and
254 OSCE3; however, they were only detected in the relapse sample in OSCE10 (Fig. 1A). OSCE9
255 was found to have a deletion event in *PTEN* and an in-frame fusion event in *ALK* (Fig. 1A).
256

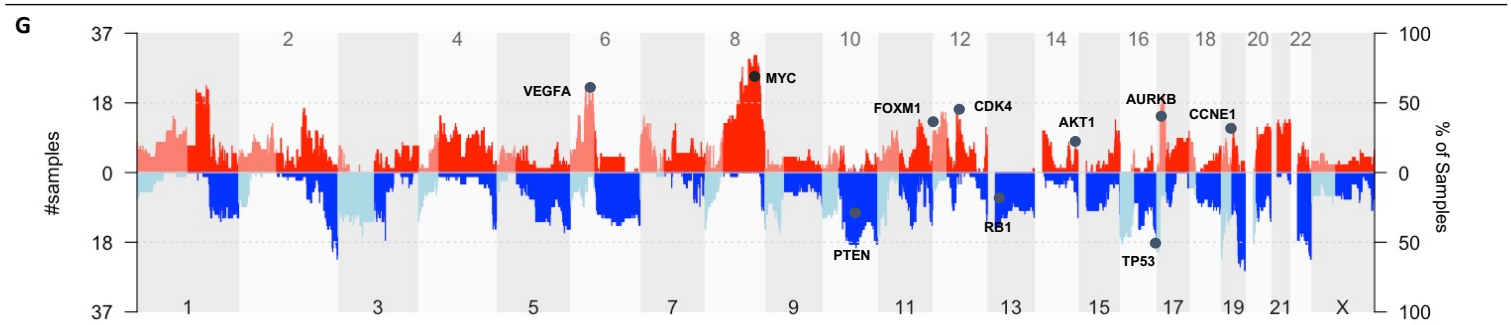
257 **Subclonal Somatic Copy Number Alterations Emerge and Dominate in Recurrent**
258 **Osteosarcoma**

259 A high prevalence of somatic copy number alterations (CNAs) in osteosarcoma has been
260 reported previously[5–9]. Therefore, we used the HATCHet[20] algorithm to infer both allele and
261 clone-specific CNAs as well as their relative proportions across multiple samples from the same
262 patient. The average tumor ploidy for each sample ranged from 1.7 in OSCE10 to 3.15 in
263 OSCE3 (Supplementary Fig. 2B). HATCHet[20] identified subclonal CNAs in all but one patient
264 (OSCE2) and whole genome duplications present at diagnosis in three of the eight patients
265 (OSCE1, OSCE2, and OSCE9; Fig. 1A). Among the seven patients with subclonal CNAs, six
266 were identified as having two major copy number clones (Fig. 2A-E, Supplementary Fig. 4A),

267 and one patient (OSCE10) had three distinct copy number clones (Fig. 2F). In four of the seven
268 patients (OSCE1, OSCE4, OSCE6, and OSCE10), multiple distinct copy-number clones were
269 identified to be simultaneously present in the primary tumor (Fig. 2A, 2B, 2D, 2F), but only one
270 of these subclones emerged and dominated at subsequent relapses. In two of these patients
271 (OSCE1, OSCE6), the emergence of this clone was observed in the primary resection sample
272 when compared to the pretreatment biopsy for each patient (Fig. 2A, 2D). In OSCE5, a new copy
273 number clone emerged in the late relapse samples, which was not identified as being present in
274 the pretreatment sample (Fig. 2C). However, when combining the analysis of mutations and
275 CNAs (see Online Methods), the dominant SNV-based clone (which shared the dominant copy
276 number profile of the late emerging copy number clone) in the relapse samples was found to be
277 sub-clonal in the pretreatment sample, providing evidence that the dominant copy number clone
278 in the relapse sample for OSCE5 was present at diagnosis (Supplementary Fig. 5A). In OSCE10,
279 no subclonal copy number aberrations were identified in the pretreatment sample; however, a
280 subclonal copy number clone was detected in the primary resection sample that emerged and
281 dominated at relapse in this patient (Fig. 2F). In summary, we found that in most cases, a minor
282 subclone present in the primary tumor emerged and dominated in patients with relapsed disease.
283 In 2/3 refractory cases (OSCE2, OSCE3), there was no subclonal copy number clone that
284 was identified in the primary that emerged and dominated in metastatic or relapse samples. In
285 OSCE2, only a single major copy number clone was identified; in OSCE3, two copy number
286 clones were identified in the pretreatment sample, with the copy number subclone emerging and
287 dominating in one of the metastatic sites but in mixed proportion in the three other metastatic
288 sites. OSCE9 did not have a pretreatment sample for comparison but did show two copy number
289 clones in mixed proportions in the primary resection sample and the two relapse sites.

Figure 2


A
OSCE1
VEGFA
CCND3
MYC
CCNE1
B
OSCE4
MYCN
ALK
MYC
RAD21
KRAS
HLA LOH
C
OSCE5
MYC
RAD21
D
OSCE6
IGF1R
PTEN
NF1
MYC
E
OSCE9
VEGFA
MYC
F
OSCE10
MYC
RAD21
Copy Number Alterations
High Amp
Mid Amp
Low Amp
Gain
Homozygous Deletion
Deep Deletion
Shallow Deletion
LOH
Classification
Log R
Deep Deletion
≤ (-1)
Shallow Deletion
(-0.5)-(-0.99)
Gain
0.5-0.99
Low Amp
1-1.49
Mid Amp
1.5-1.99
High Amp
≥ 2


G
#samples
% of Samples
VEGFA
MYC
FOXM1
CDK4
AKT1
AURKB
CCNE1
PTEN
RB1
TP53

290

291 **Figure 2.** Subclonal copy number clones emerge at relapse. **A, B, C, D, E, and F,** For each
292 patient there is a panel of three figures. The figure on the left is an oncoprint featuring clone
293 specific copy number alterations in recurrently altered genes of interest in osteosarcoma. The top
294 figure is a plot of allele specific copy number alterations for each clone with significant events
295 for each clone circled and highlighted (note y-axis scales are unique for each patient). Clone 1 is
296 the magenta clone, clone 2 is the teal clone, and clone 3 in OSCE10 is the gray clone. The major
297 allele is plotted above 0 and the minor allele is plotted below 0. The bottom figure in each panel
298 is a TimeScape plot of the prevalence of each clone at different timepoints throughout a patient's
299 disease course. **G,** Combined genome-wide copy number alterations across all patients in the
300 cohort with recurrently altered genes highlighted.

301 Determinations regarding branched vs. parallel evolution (depicted in the TimeScape
302 plots in Fig. 2A-F, Supplementary Fig. 4A) for patients with ≥2 copy number clones were based
303 on a review of loss of heterozygosity (LOH) events in the dominant metastatic or recurrent copy
304 number clone, using the rationale and methods outlined by Watkins et al[21]. Branched evolution
305 with the emergence of the treatment resistant copy number clone from the dominant pretreatment
306 copy number clone (clone 2 emerging from clone 1) was observed in three patients (OSCE4,
307 OSCE9, and OSCE10). Parallel evolution, where the pretreatment and treatment resistant copy
308 number clones share the same parent clone (clone 1 and clone 2 share the same parent clone) but
309 evolve in parallel, was seen in four patients (OSCE1, OSCE3, OSCE5, OSCE6). LOH events
310 were common in tumor suppressor genes such as *TP53, RB1,* and *PTEN* and were mostly shared
311 between clones for each patient, with a median of 77.95% of LOH events shared between all
312 clones (range 18.9%-86.82%, Supplementary Fig. 4C).
313
314 **Copy number amplifications in recurrently altered oncogenes in osteosarcoma characterize**
315 **chemoresistant copy number clones**
316 Cohort-wide copy number alterations reflected what has been previously described in
317 osteosarcoma[4], with gains and amplifications seen in *VEGFA*, *MYC*, *FOXM1*, *CDK4*, *AKT1*,
318 *AURKB*, and *CCNE1*, and loss/deletion events in *PTEN*, *RB1*, and *TP53* (Fig. 2G). In contrast to
319 previously identified SNV drivers, the relative proportions of alterations across tumor cells
320 changed with time. Clones were classified as chemoresistant if they were present in the primary
321 site and became dominant at relapse, and chemosensitive if it was the dominant clone in the
322 primary site and became subclonal or eliminated in metastatic or relapse sites. When comparing
323 the genomic alterations between copy number clones for each patient, deletion or LOH events in

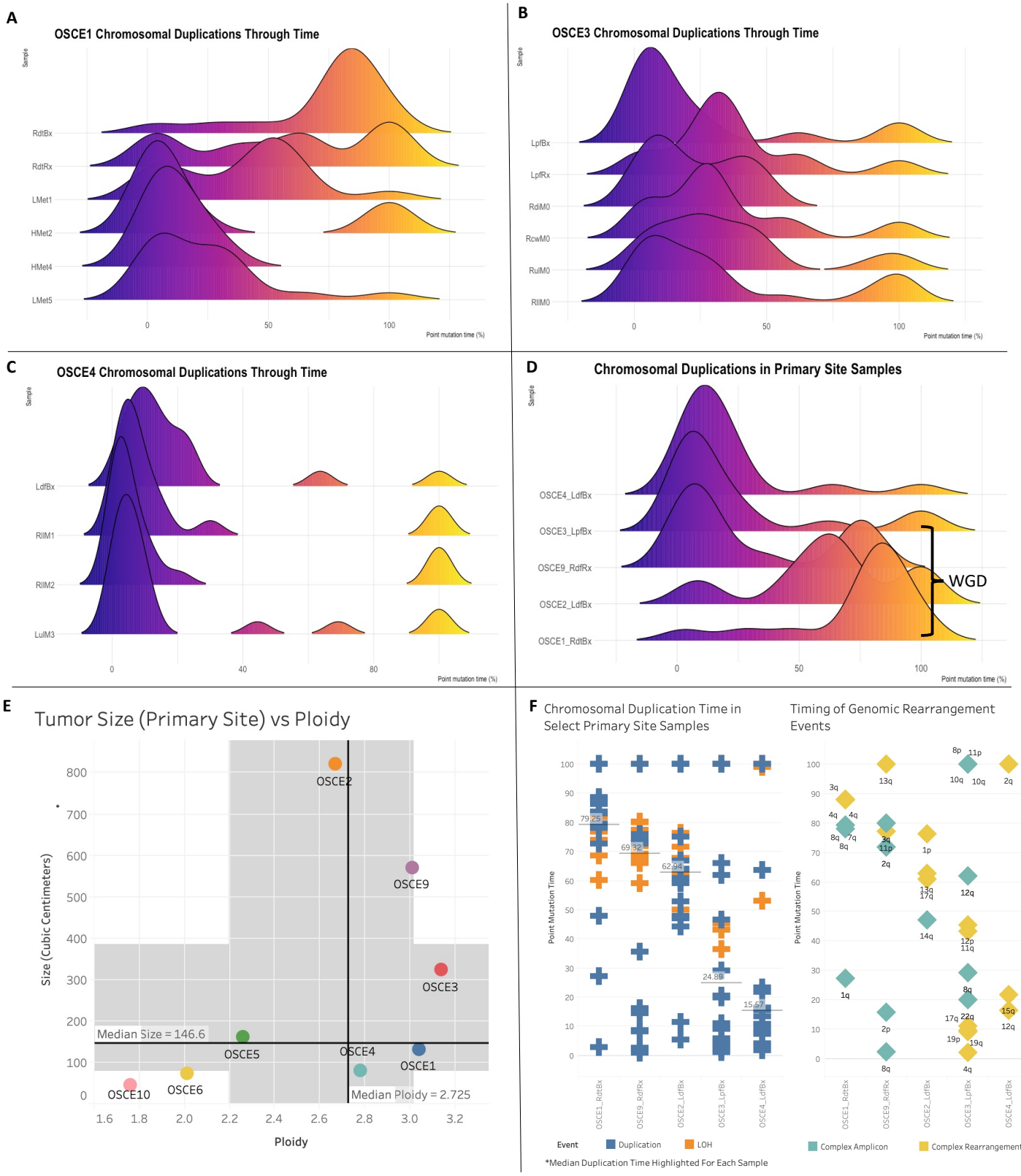
324 tumor suppressor driver genes were often clonal in patients found to have ≥ 2 copy number
325 clones (Fig. 2A-F, Supplementary Fig. 4A). In the 5 patients with clear emergence of a
326 chemoresistant copy number clone (clone 2 in OSCE1, OSCE4, OSCE5, OSCE6 and clone 3 in
327 OSCE10), the resistant clone had a higher degree of *MYC* gain or amplification then the
328 dominant chemosensitive clone at diagnosis (Fig. 2A-F, Supplementary Fig. 5B). In OSCE10,
329 high-level (log2 ≥ 2) MYC-amplified clone 3 emerged at the time of primary resection and
330 dominated at relapse (Fig. 2F). Notably, this treatment-resistant clone was present in 4/5 multi-
331 region samples from the primary resection, suggesting intratumoral heterogeneity regarding copy
332 number alterations depending on the site sampled. In addition to *MYC*, amplification of
333 *CCNE1*/*CCND3* (OSCE1, OSCE6; Fig. 2A/D), *KRAS* (OSCE2, OSCE4; Supplementary Fig. 4B,
334 Fig. 2B), *IGF1R* (OSCE6, Fig. 2D), *CDK4* (OSCE10, Fig. 2F), *VEGFA* (OSCE1, OSCE6,
335 OSCE9; Fig. 2A/D/E), and LOH at HLA (OSCE4, Fig. 2B and Supplementary Fig. 4D) uniquely
336 characterized treatment-resistant or metastatic copy number clones in this cohort.
337
338 **Subclonal selection/emergence evident at the time of primary resection**
339 Three patients within our cohort had a pretreatment sample on therapy resection, and at
340 least one relapse sample (OSCE1, OSCE6, OSCE10), which provided an opportunity to analyze
341 the effect of neoadjuvant chemotherapy on subclonal copy number selection at the time of
342 primary resection and whether this selection is reflective of the dominant clone at the time of
343 relapse. OSCE1 and OSCE6 share a similar pattern, where a copy number subclone present at
344 diagnosis emerges as the dominant clone at the time of primary resection and continues to
345 dominate at the first and subsequent relapses. Chemoresistant clone 2 was characterized by mid-
346 level (log2=1.5-1.99) *CCNE1* amplification in OSCE1 (Fig. 2A) and mid-level *IGF1R*

347 amplification in OSCE6 (Fig. 2D). A slightly different pattern emerged in the refractory case of
348 OSCE10, in which only a single copy number clone was present at diagnosis. The primary
349 resection sample underwent multi-region sequencing of five spatially separated sites, which
350 revealed the emergence of two new copy number clones, with clone 2 present in all five samples
351 and clone 3 in 4/5 samples. Clone 3 then became the dominant clone in the first relapse
352 specimen, characterized by high-level *MYC* amplification.
353

354 **Timing of Copy Number Gains Reveals Large Bursts of Copy Number Gains Before**
355 **Diagnosis**

356 The HATCHet algorithm also infers whole-genome duplication (WGD) events jointly
357 across all samples for each patient. Of the five patients with average ploidy ≥ 2.5 (OSCE1,
358 OSCE2, OSCE3, OSCE4, OSCE9), WGD was identified in OSCE1, OSCE2, and OSCE9 (Fig.
359 1A and Supplementary Fig. 2B). The timing of these WGD events as well as other chromosomal
360 duplications can be determined in "molecular time" using previously described methods[23,24]
361 which compares the number of duplicated vs non-duplicated mutations to estimate the timing of
362 each duplication (Supplementary Fig. 6). For the patients who were determined to have WGD
363 events, these appeared to be late events for each respective patient (median point mutation time
364 (pmt) range = 62-79%, Fig. 3F), compared to the two patients with ploidy ≥2.5 without WGD
365 (patients with ploidy < 2.5 did not have enough duplication or LOH events to analyze), which
366 appeared to have early synchronous duplications (median pmt range = 16-25%, Fig. 3F). The
367 two patients with the lowest ploidy (OSCE10 and OSCE6) also had the smallest primary tumors
368 (Fig. 3E), whereas all large primary tumors (≥300 cm^3) had ploidy ≥ 2.5 (OSCE2, OSCE9,

Figure 3

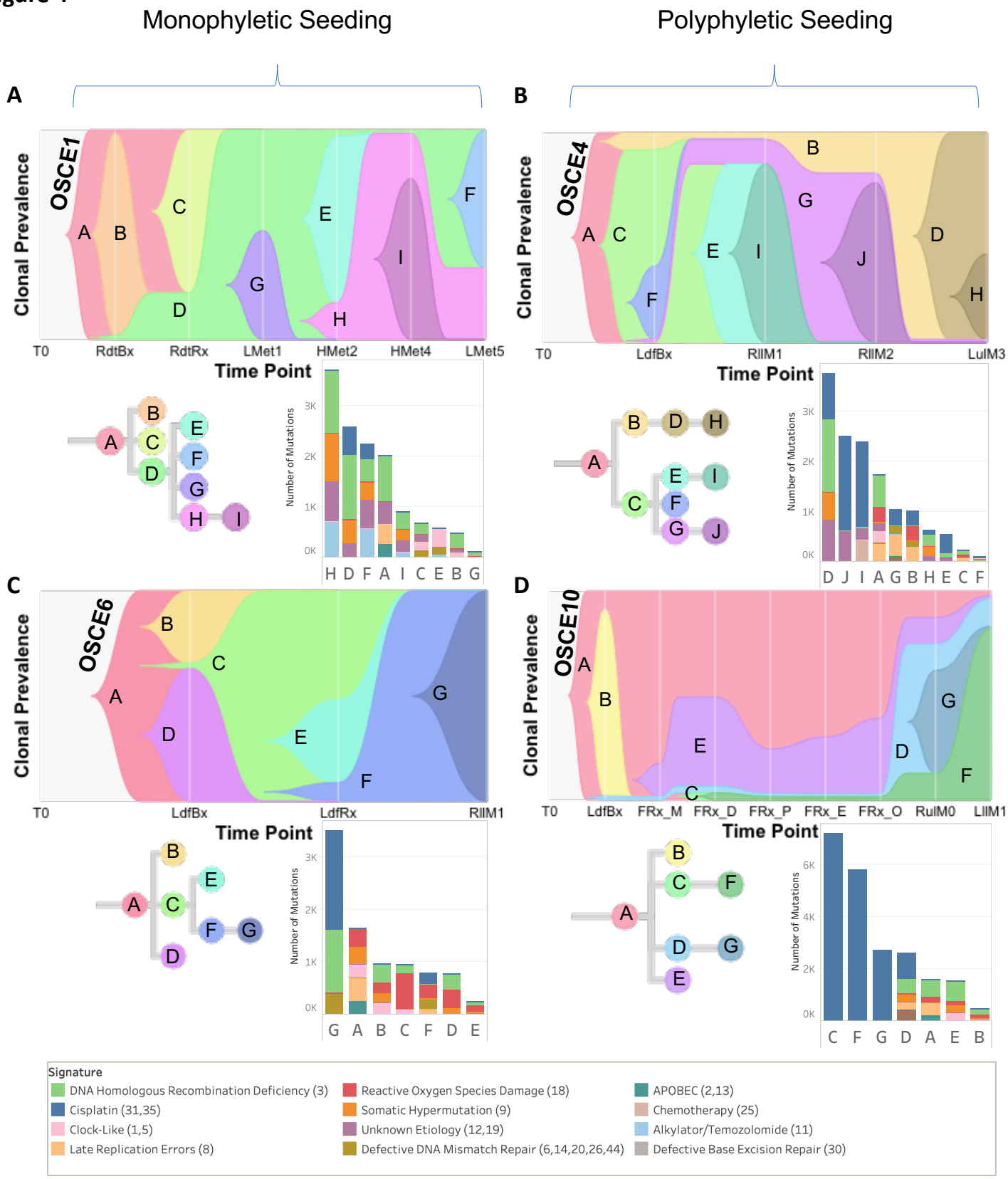

A
OSCE1 Chromosomal Duplications Through Time
Sample
RdtBx
RdtRx
LMet1
HMet2
HMet4
LMet5
0
50
100
Point mutation time (%)
B
OSCE3 Chromosomal Duplications Through Time
Sample
LpfBx
LpfRx
RdiM0
RcwM0
RulM0
RllM0
0
50
100
Point mutation time (%)
C
OSCE4 Chromosomal Duplications Through Time
Sample
LdfBx
RllM1
RllM2
LulM3
0
40
80
Point mutation time (%)
D
Chromosomal Duplications in Primary Site Samples
Sample
OSCE4_LdfBx
OSCE3_LpfBx
OSCE9_RdfRx
OSCE2_LdfBx
OSCE1_RdtBx
WGD
0
50
100
Point mutation time (%)
E
Tumor Size (Primary Site) vs Ploidy
Size (Cubic Centimeters)
Ploidy
OSCE2
OSCE9
OSCE3
OSCE5
OSCE4
OSCE1
OSCE10
OSCE6
Median Size = 146.6
Median Ploidy = 2.725
F
Chromosomal Duplication Time in Select Primary Site Samples
Timing of Genomic Rearrangement Events
Point Mutation Time
OSCE1_RdtBx
OSCE9_RdfRx
OSCE2_LdfBx
OSCE3_LpfBx
OSCE4_LdfBx
79.25
69.32
62.94
24.89
15.57
Event
Duplication
LOH
Complex Amplicon
Complex Rearrangement
*Median Duplication Time Highlighted For Each Sample

369

370 **Figure 3.** Chromosomal duplication timing analysis reconstructs evolutionary past of genomic
371 instability events. **A, B, C,** Ridgeline plots of the density of duplication events over molecular
372 time for each sample for the selected patients. Notably the highest peak in duplications occurs
373 before diagnosis. **D,** Chromosomal duplication timing of 5 primary site samples. Whole genome
374 duplication events as called by HATCHet appear to be a late event in our cohort. **E,** Plot of
375 tumor size (by volume) verse ploidy called by HATCHet. Median size (y-axis) and ploidy (x-
376 axis) values are plotted with dark black lines, with the shaded gray areas representing the range
377 between the lower and upper quartile for each metric. **F,** Plot of duplication and rearrangement
378 events in molecular time. Left side of figure is plot of duplication events in blue and LOH events
379 in orange for select primary site samples. The median duplication time is highlighted for each
380 sample. The plot on the right side of the figure are complex amplicon events in teal and complex
381 rearrangement events in yellow. Y-axis for both plots is molecular time. The samples are in the
382 same order for each plot. Each plotted event represents an affected chromosomal arm.

383 OSCE3). For the five patients with ploidy ≥ 2.5, we analyzed their earliest primary site samples
384 to assess the natural history of these duplication events in the context of tumorigenesis.
385 Most of these duplication/LOH events clustered in a single burst of events (Fig. 3D) and
386 were associated with complex rearrangement (chromoplexy, chromothripsis) or complex
387 amplicon[25] (Tyfonas, breakage fusion bridge, double minute) events (Fig. 3F). When comparing
388 longitudinal samples from the same patient (Fig. 3A-C), there was no subsequent burst that was
389 greater than the primary site burst. Across the cohort, duplication events appeared to be fixed
390 during tumorigenesis and had decreasing average molecular times when comparing
391 biopsy/resection samples with metastatic/relapse samples (Supplementary Fig. 7A/C). LOH
392 events were consistently "late" events, with a median molecular time of 82.16 across the cohort
393 (compared to 21.72 for duplication events, Supplementary Fig. 7B/D).
394
395 **Heterogenous seeding patterns observed in metastatic and recurrent osteosarcoma**
396 Clone-based phylogenies were created to explore the clonal architecture and track the
397 spatial and temporal patterns of evolution. The SNVs across all samples for each patient were
398 clustered using the DeCiFer[17] algorithm. After clustering, each SNV was assigned to a clone
399 with an estimated descendant cell fraction (DCF) per sample, and clone-based phylogenetic trees
400 were then constructed using the CALDER[26] algorithm, allowing for the assessment of modes of
401 metastatic seeding and dissemination. The median number of clones per patient was eight (range
402 = 5-12, Supplementary Fig. 8I). At the patient level, there was a heterogeneous mix of
403 dissemination patterns (Fig. 4A-D, Fig. 5A-D, and Supplementary Fig. 8A-H), with complete
404 agreement between clone and sample-based phylogenies (Supplementary Fig. 3A) regarding
405 whether metastatic dissemination was monophyletic (all metastatic clones shared a common

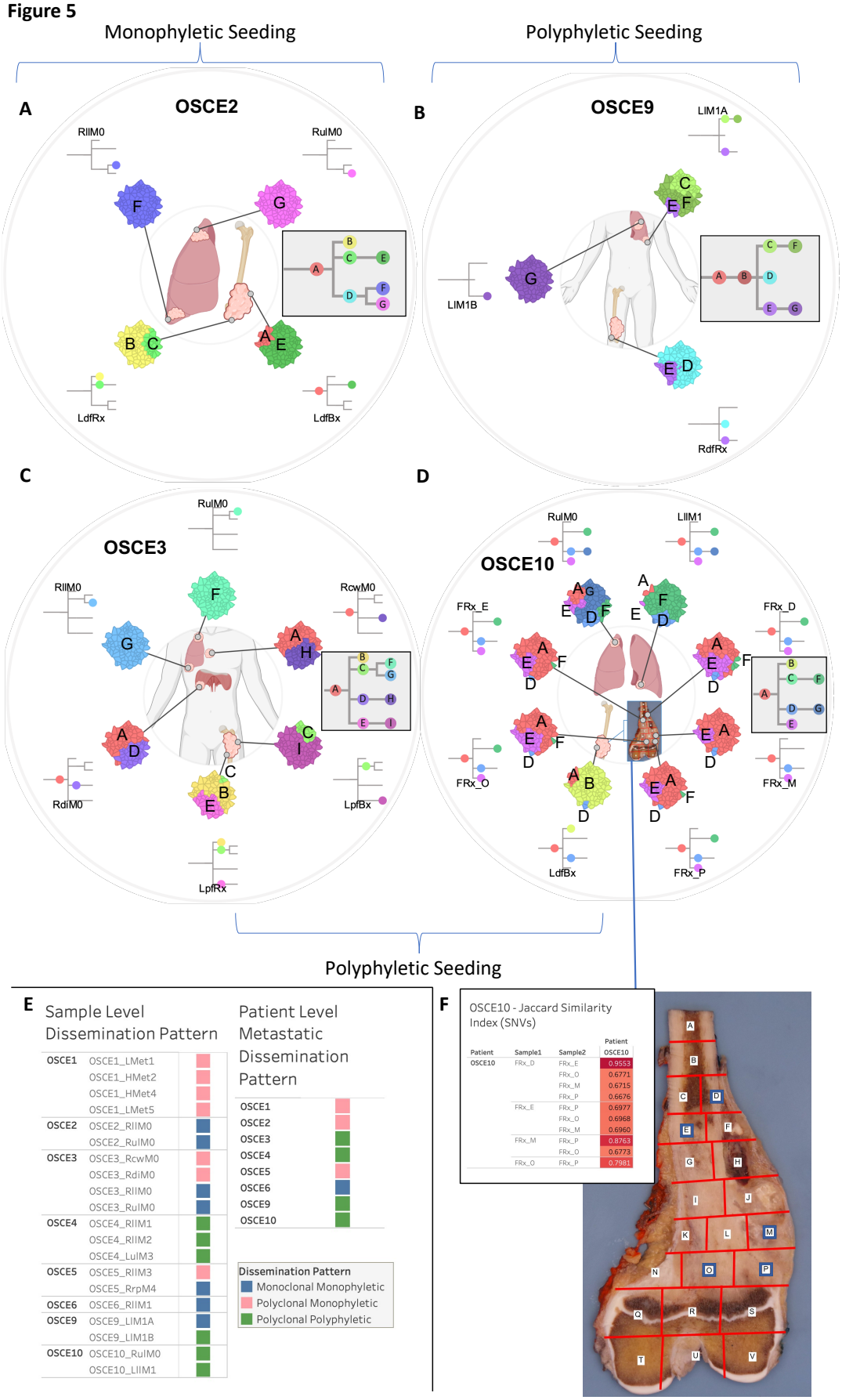
Figure 4


Monophyletic Seeding
Polyphyletic Seeding
A
B
C
D
OSCE1
OSCE4
OSCE6
OSCE10
Clonal Prevalence
Time Point
T0
RdtBx
RdtRx
LMet1
HMet2
HMet4
LMet5
LdfBx
RIIM1
RIIM2
LuIM3
LdfRx
RIIM1
FRx_M
FRx_D
FRx_P
FRx_E
FRx_O
RuIM0
LIIM1
Number of Mutations
H D F A I C E B G
D J I A G B H E C F
G A B C F D E
C F G D A E B
Signature
DNA Homologous Recombination Deficiency (3)
Cisplatin (31,35)
Clock-Like (1,5)
Late Replication Errors (8)
Reactive Oxygen Species Damage (18)
Somatic Hypermutation (9)
Unknown Etiology (12,19)
Defective DNA Mismatch Repair (6,14,20,26,44)
APOBEC (2,13)
Chemotherapy (25)
Alkylator/Temozolomide (11)
Defective Base Excision Repair (30)

407 **Figure 4.** SNV based phylogenies highlighting temporal evolution with clonal mutational
408 signature composition. **A, B, C, and D,** Upper figure in each panel is a TimeScape plot of the
409 inferred evolutionary phylogeny, highlighting clonal proportions over time. The prevalence of
410 different clones is shown over time on the vertical axis, with the different clones represented by
411 different colors. The horizontal axis represents the timepoints, which are represented by gray
412 lines. The evolutionary relationships between the clones are shown on the phylogenetic tree and
413 in the TimeScape layout. The bottom right of each panel is a stacked bar plot of the total number
414 of mutations assigned to each clone. Colors represent total number of mutations attributed to
415 each mutational signature with color legend at bottom of figure. Patient level metastatic seeding
416 patterns are denoted by the brackets at the top of the page.

417 ancestral clone: Fig. 4A, 4C, 5E; Supplementary Fig. 6A, 6C, 6D, 6G, 6H)) vs. polyphyletic in
418 origin (where ≥2 clones are present from distinct branches of phylogeny whose common
419 ancestor represents the trunk of the primary tumor tree: Fig. 4B, 4D, 5B; Supplementary Fig. 6B,
420 6D, 6F). Of note, OSCE6 was the only patient among those with a monophyletic origin that had
421 monoclonal seeding (one clone present in the sample, Fig. 4C and Supplementary Fig. 6H). The
422 seeding and dissemination patterns were also assessed for each metastatic sample (Fig. 5E). In
423 two patients (OSCE3 and OSCE5), there were examples of different modes of metastatic
424 dissemination among spatially separated metastases from the same resection. In OSCE3 for
425 example, extra-pulmonary metastases (RcwM0, RdiM0) demonstrate polyclonal (≥2 clones
426 present in the sample) monophyletic patterns of dissemination, while the pulmonary metastases
427 (RllM0, RulM0) demonstrate monoclonal monophyletic dissemination, with both samples
428 sharing the same ancestral clone (Fig. 5C, Supplementary Fig. 6B, 6I).
429
430 **Limited Heterogeneity after Induction Chemotherapy**
431 For OSCE10, we obtained a section of the primary tumor sample for multi-region
432 sequencing after the patient had received 10 weeks of induction MAP chemotherapy. The
433 specimen was mapped (Fig. 5F), DNA was extracted from the sections with viable tumor, and
434 then five sections (E, D, M, O, P) with the highest DNA quantity/quality metrics underwent
435 WGS. When examining the clonal architecture of each section (Fig. 4D/5F), each section had the
436 same four clones (A, D, E, F), with truncal clone A dominating, except for FRx_M, which did
437 not have clone F, which later became the clone that dominated at relapse. This pattern was also
438 seen with the copy number clones, where clone 3 was present in all sections except for FRx_M
439 and then dominated at relapse (Fig. 2F). When comparing Jaccard similarity coefficients based

Figure 5


Monophyletic Seeding
Polyphyletic Seeding
A
OSCE2
RIIM0
RuIM0
LdfRx
LdfBx
B
OSCE9
LIM1A
LIM1B
RdfRx
C
OSCE3
RuIM0
RIIM0
RcwM0
RdiM0
LpfBx
LpfRx
D
OSCE10
RuIM0
LIIM1
FRx_E
FRx_D
FRx_O
FRx_M
LdfBx
FRx_P
Polyphyletic Seeding
E
Sample Level Dissemination Pattern
Patient Level Metastatic Dissemination Pattern
OSCE1 OSCE1_LMet1
OSCE1_HMet2
OSCE1_HMet4
OSCE1_LMet5
OSCE2 OSCE2_RIIM0
OSCE2_RuIM0
OSCE3 OSCE3_RcwM0
OSCE3_RdiM0
OSCE3_RIIM0
OSCE3_RuIM0
OSCE4 OSCE4_RIIM1
OSCE4_RIIM2
OSCE4_LuIM3
OSCE5 OSCE5_RIIM3
OSCE5_RrpM4
OSCE6 OSCE6_RIIM1
OSCE9 OSCE9_LIM1A
OSCE9_LIM1B
OSCE10 OSCE10_RuIM0
OSCE10_LIIM1
OSCE1
OSCE2
OSCE3
OSCE4
OSCE5
OSCE6
OSCE9
OSCE10
Dissemination Pattern
Monoclonal Monophyletic
Polyclonal Monophyletic
Polyclonal Polyphyletic
F
OSCE10 - Jaccard Similarity Index (SNVs)
Patient Sample1 Sample2 Patient OSCE10
OSCE10 FRx_D FRx_E 0.9553
FRx_O 0.6771
FRx_M 0.6715
FRx_P 0.6676
FRx_E FRx_P 0.6977
FRx_O 0.6968
FRx_M 0.6960
FRx_M FRx_P 0.8763
FRx_O 0.6773
FRx_O FRx_P 0.7981

440

441 **Figure 5.** SNV based phylogenies highlighting spatial evolution and descriptions of metastatic
442 seeding patterns at the sample and patient level. **A, B, C, and D,** Spatially and in some cases
443 temporally distinct samples are indicated on the anatomic sites from where the sample
444 originated. The colors represent different clones, and the phylogenetic trees show the
445 evolutionary relationships between these clones. The prevalence of each clone at a particular site
446 is proportional to the colored area of the cellular aggregate representation. **E,** Sample and patient
447 level dissemination patterns are characterized in these charts. *Monoclonal dissemination*: single
448 subclone within the primary tumor seeds one or more metastatic lesions, *polyclonal*
449 *dissemination*: multiple distinct subclones from the primary tumor seed one or more metastatic
450 lesions, *monophyletic origin*: all metastatic clones are derived from a recent common ancestor,
451 *polyphyletic origin*: metastasizing clones are more similar to other subclones within the primary
452 tumor than they are to each other. These descriptions can be considered at the sample level,
453 focused on the clonal make up of a single metastatic site compared to the primary tumor, or
454 taken as a whole, evaluating all spatially or temporally separated samples and how they relate
455 back to the primary tumor. **F,** Multi-region sequencing was performed on a primary resection
456 sample from OSCE10. Regions D, E, M, O, P were sequenced from the specimen grid depicted.
457 A table of Jaccard similarity indexes based on shared SNVs for these samples is shown in the
458 upper left inset.

459 on SNV composition across the different sections, adjoining sections shared the highest
460 coefficients (D/E=0.96, O/P=0.79, M/P=0.88, Fig. 5F), whereas non-adjoining sections had
461 similar coefficients, regardless of distance from each other (range = 0.67-0.69, Fig. 5F). As
462 highlighted previously, a single-sample sequencing strategy that sampled from FRx_M would
463 have missed detecting the metastatic subclone that was present in the other four sections.
464

465 **Most new SNVs in relapsed disease are attributed to HRD-related SBS3 and cisplatin**
466 **mutational signatures**

467 To further characterize the potential drivers of clonal evolution, mutational signature
468 analysis was performed for each clone (Fig. 4A-D, Supplementary Fig. 6A-D). In three patients,
469 the largest clone by number of SNVs (OSCE5 – clone C, OSCE6 – clone G, and OSCE10 –
470 clone C) had over half of the SNVs attributed to the DNA-damaging effects of cisplatin
471 chemotherapy (Supplementary Fig. 9A). In four patients, the largest clone by number of SNVs
472 (OSCE1 – clone H, OSCE3 – clone A, OSCE4 – clone D, and OSCE9 – clone G) had a plurality
473 of SNVs in each clone attributed to single base substitution (SBS) 3, a genomic signature that
474 has been associated with homologous repair deficiency (HRD, Supplementary Fig. 9A). Tumors
475 with this signature are thought to have a BRCAness phenotype and exhibit features similar to
476 those cancers with germline BRCA1 or BRCA2 mutations, even though no mutations in those
477 genes have been identified. In OSCE2 and OSCE3, both refractory cases, the largest number of
478 SNVs was assigned to the truncal clone, clone A, which had a high number of SNVs (OSCE2 –
479 1731/2234, OSCE3 – 2464/4022) associated with HRD-related SBS3 and late replication errors
480 (Sig. 8). In patients with patient-level monophyletic seeding of metastases (OSCE1, OSCE2,
481 OSCE5, OSCE6), where a single ancestral metastatic clone could be identified (OSCE1–clone

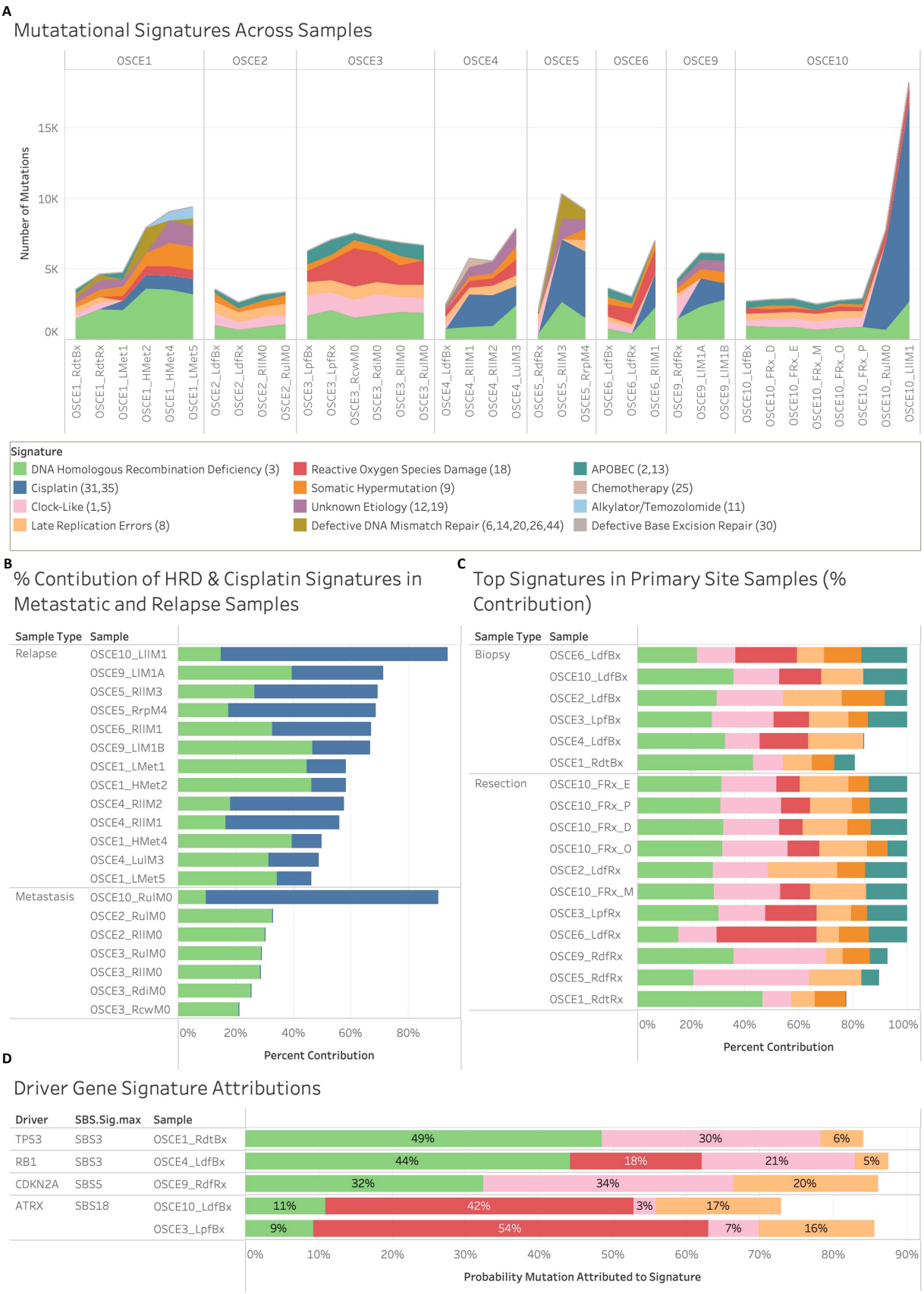
482 D, OSCE2–clone D, OSCE5–clone C, OSCE6–clone C), there were no clear patterns regarding
483 the signature composition identified (Supplementary Fig. 9B). OSCE1–clone D and OSCE2-
484 clone D had a plurality of SNVs attributed to HRD-related SBS3, whereas OSCE5-clone C had
485 the majority of SNVs attributed to cisplatin, and OSCE6 had the majority of its SNVs attributed
486 to reactive oxygen species damage (Sig. 18). When looking at doublet base substitution (DBS)
487 signatures at the clonal level, in clones with ≥10 DBS SNVs, cisplatin-associated DBS 5 was the
488 largest contributor of DBS SNVs, accounting for 50% or more of the total SNVs in 17/26 clones
489 (Supplementary Fig. 9C).
490
491 **Emergence of cisplatin and alkylator signatures helps time the formation of metastases**
492 A simple method for ascertaining whether a given metastasis arose before or after
493 treatment with cisplatin or an alkylator is to find clonal SNVs attributed to the respective
494 signature in a metastatic tumor sample[27]. When looking at the dominant clone in the first relapse
495 sample for patients with recurrent disease (OSCE1-Clone D, OSCE4-Clone E/I, OSCE5-Clone
496 C, OSCE6-Clone C/F/G, OSCE9-Clone C/G, OSCE10-Clone C/F), there was a cisplatin
497 signature present in each of these respective clones/clades, indicating that the metastases arose
498 after therapy (Fig. 4A-D, Supplementary Fig. 6C, 6D). Additionally, in OSCE1, after the patient
499 received chemotherapy with ifosfamide and etoposide, OSCE1-Clone H and OSCE1-Clone F
500 had mutations attributed to the alkylator signature, SBS11 (Fig. 4A). In the three patients with
501 refractory disease (OSCE2, OSCE3, and OSCE10), we can demonstrate which metastatic
502 samples were present at diagnosis and which developed while on therapy. Within OSCE2, Clone
503 F and Clone G were the dominant subclones in RllM0 and RulM0 respectively in OSCE2 (Fig.
504 5A, Supplementary Fig. 6A). Clone F had 181/614 mutations attributed to cisplatin and Clone G

505 had no mutations attributed to cisplatin, evidence Clone F was seeded on therapy as opposed to
506 prior to therapy. In OSCE3, there was no evidence of a cisplatin signature in any of the
507 subclones (Supplementary Fig. 6B), indicating that all metastatic sites had developed prior to
508 initiating therapy. In OSCE10, the metastatic sample (RulM0), which has polyphyletic and
509 polyclonal seeding (Fig. 4D, 5D), showed that the dominant clone/subclone pair of D/G likely
510 seeded on therapy, given that there is a cisplatin signature attributed to approximately half of the
511 mutations (1027/2644) and in all the mutations in subclone G (2727/2727).
512

513 **HRD-related SBS3 and Reactive Oxygen Species Damage Linked to Driver Gene**
514 **Mutagenesis**

515 To identify the mutational processes most likely to be the origin of truncal driver gene
516 SNVs, we calculated the likelihood that each individual SNV was caused by each signature,
517 considering the mutation category and proportion of each mutational signature in the tumor
518 genome[24]. To minimize the effect of treatment-related mutagenesis, we limited this analysis to
519 the earliest primary site sample available for patients with truncal driver SNVs. Similar to the
520 observations across all samples, HRD-related SBS3 had the highest probability of attribution in
521 the *TP53* driven OSCE1 sample and the *RB1* driven OSCE4 samples (OSCE1 probability=0.49,
522 OSCE4 probability=0.44, Fig. 6D), and a slightly lower attribution probability than the clock-
523 like signature in OSCE9 (HRD-related SBS3 probability=0.32, clock-like probability=0.34, Fig.
524 6D). Both OSCE3 and OSCE10 had truncal *ATRX* SNVs, with reactive oxygen series damage
525 accounting for the highest probability of attribution in both samples (OSCE3 probability=0.54,
526 OSCE10 probability=0.42, Fig. 6D).
527

528 **Figure 6**


A
Mutatational Signatures Across Samples
OSCE1
OSCE2
OSCE3
OSCE4
OSCE5
OSCE6
OSCE9
OSCE10
Number of Mutations
15K
10K
5K
0K
OSCE1_RdtBx
OSCE1_RdtRx
OSCE1_LMet1
OSCE1_HMet2
OSCE1_HMet4
OSCE1_LMet5
OSCE2_LdfBx
OSCE2_LdfRx
OSCE2_RlIM0
OSCE2_RulM0
OSCE3_LpfBx
OSCE3_LpfRx
OSCE3_RcwM0
OSCE3_RdiM0
OSCE3_RlIM0
OSCE3_RulM0
OSCE4_LdfBx
OSCE4_RlIM1
OSCE4_RlIM2
OSCE4_LulM3
OSCE5_RdfRx
OSCE5_RlIM3
OSCE5_RrpM4
OSCE6_LdfBx
OSCE6_LdfRx
OSCE6_RlIM1
OSCE9_RdfRx
OSCE9_LIM1A
OSCE9_LIM1B
OSCE10_LdfBx
OSCE10_FRx_D
OSCE10_FRx_E
OSCE10_FRx_M
OSCE10_FRx_O
OSCE10_FRx_P
OSCE10_RulM0
OSCE10_LIIM1
Signature
DNA Homologous Recombination Deficiency (3)
Reactive Oxygen Species Damage (18)
APOBEC (2,13)
Cisplatin (31,35)
Somatic Hypermutation (9)
Chemotherapy (25)
Clock-Like (1,5)
Unknown Etiology (12,19)
Alkylator/Temozolomide (11)
Late Replication Errors (8)
Defective DNA Mismatch Repair (6,14,20,26,44)
Defective Base Excision Repair (30)
B
% Contibution of HRD & Cisplatin Signatures in Metastatic and Relapse Samples
Sample Type
Sample
Relapse
OSCE10_LIIM1
OSCE9_LIM1A
OSCE5_RlIM3
OSCE5_RrpM4
OSCE6_RlIM1
OSCE9_LIM1B
OSCE1_LMet1
OSCE1_HMet2
OSCE4_RlIM2
OSCE4_RlIM1
OSCE1_HMet4
OSCE4_LulM3
OSCE1_LMet5
Metastasis
OSCE10_RulM0
OSCE2_RulM0
OSCE2_RlIM0
OSCE3_RulM0
OSCE3_RlIM0
OSCE3_RdiM0
OSCE3_RcwM0
0%
20%
40%
60%
80%
Percent Contribution
C
Top Signatures in Primary Site Samples (% Contribution)
Sample Type
Sample
Biopsy
OSCE6_LdfBx
OSCE10_LdfBx
OSCE2_LdfBx
OSCE3_LpfBx
OSCE4_LdfBx
OSCE1_RdtBx
Resection
OSCE10_FRx_E
OSCE10_FRx_P
OSCE10_FRx_D
OSCE10_FRx_O
OSCE2_LdfRx
OSCE10_FRx_M
OSCE3_LpfRx
OSCE6_LdfRx
OSCE9_RdfRx
OSCE5_RdfRx
OSCE1_RdtRx
0%
20%
40%
60%
80%
100%
Percent Contribution
D
Driver Gene Signature Attributions
Driver
SBS.Sig.max
Sample
TP53
SBS3
OSCE1_RdtBx
49%
30%
6%
RB1
SBS3
OSCE4_LdfBx
44%
18%
21%
5%
CDKN2A
SBS5
OSCE9_RdfRx
32%
34%
20%
ATRX
SBS18
OSCE10_LdfBx
11%
42%
3%
17%
OSCE3_LpfBx
9%
54%
7%
16%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
Probability Mutation Attributed to Signature

529 **Figure 6.** Mutational signature patterns across the cohort. **A,** Stacked line chart of mutational
530 signature contribution by total number of mutations attributed to each signature. Colors represent
531 the different signatures. **B,** Stacked bar chart of the relative contribution of HRD-related SBS3
532 (green) & cisplatin (blue) signature in metastatic and relapse samples. **C,** Stacked bar chart of the
533 relative contribution of HRD (green), clock-like (pink), reactive oxygen species damage (red),
534 late replication errors (beige), somatic hypermutation (orange), and APOBEC (teal), in primary
535 site samples. **D,** Stacked bar chart of the probability that driver gene SNVs were attributed to a
536 mutational signature. Pretreatment samples from patients with driver SNVs were included and
537 the primary resection sample from OSCE9 since no pretreatment sample was available. Colors
538 represent the different signatures.

539 **Cisplatin associated hypermutation in a case of refractory osteosarcoma**

540 When evaluating mutational signatures at the sample level, the two most prevalent SBS
541 signatures across all samples were HRD-related SBS3 and cisplatin (Fig. 6A). In the 13 relapse
542 samples, HRD-related SBS3 and cisplatin accounted for more than half of all the SNVs (Fig.
543 6B). In the 17 primary site samples, the HRD-related SBS3, clock-like (Sig. 5), late replications
544 errors (Sig. 8), ROS damage (Sig. 18), and APOBEC (Sig. 2,13) accounted for 75%–100% of all
545 mutations (Fig. 6C). In 6/7 metastatic samples, HRD-related SBS3, clock-like, late replication
546 errors, somatic hypermutation, ROS damage, and APOBEC accounted for all SNVs
547 (Supplementary Fig. 10A). In the metastatic sample thought to have emerged on therapy for
548 OSCE10 (RulM0) and the subsequent relapse sample (LllM1), SNVs attributed to cisplatin
549 accounted for 6226/7724 (86%) and 14276/18216 (78%) SNVs, respectively (Fig. 10A). In the
550 six patients with relapsed disease, the number of mutations attributed to HRD-related SBS3
551 consistently increased from diagnosis to subsequent relapses, with at least 1000 new SNVs
552 attributed to HRD-related SBS3 when comparing diagnostic and relapse samples (Supplementary
553 Fig. 10C).

554 Doublet base substitution (DBS) signatures were also evaluated across the cohort
555 (Supplementary Fig. 10B). DBS signature 5, which is associated with cisplatin, was detected in
556 21 samples; however, to filter out false positives, a threshold of ≥ 5 DBS signature 5 SNVs was
557 used to confirm the absence of the signature. Using this filter, DBS signature 5 was present in 15
558 samples, all metastatic or relapse samples, with complete overlap with the 14 samples that had
559 SBS cisplatin signatures 31 or 35. Only OSCE3_RdiM0 had a DBS cisplatin signature but not an
560 SBS cisplatin signature (Supplementary Fig. 10D).

561 **Discussion**

562 Tumor evolution and clonal heterogeneity have been increasingly recognized as major
563 causes of therapeutic resistance to current anti-neoplastic therapies[28]. These findings extend our
564 understanding of therapeutic resistance in spatially and temporally separated tumor samples from
565 8 patients with recurrent or refractory osteosarcoma. We found that while clonal driver gene
566 SNVs and structural variants remain largely unchanged over the course of tumor progression,
567 subclonal tumor populations with unique driver gene amplifications are present at diagnosis,
568 emerge after treatment, and persist as the major clone at subsequent relapses.

569 Somatic copy number alterations are now increasingly recognized for their prognostic
570 value over SNVs in multiple cancer types[29]. Oncogenic copy number alterations, while
571 heterogeneous across osteosarcoma, represent potential therapeutic targets, given the lack of
572 recurrent targetable SNVs or structural variants[4–8]. Our study revealed that in our four patients
573 who had relapsed osteosarcoma with a matched pretreatment sample (OSCE1, OSCE4, OSCE5,
574 OSCE6), there was a subclonal treatment resistant copy number clone that emerged as the
575 dominant clone in the relapsed setting. Furthermore, in the two patients with both a pre-
576 treatment sample and on-therapy primary resection (OSCE1, OSCE6), this treatment resistant
577 clone clonally expanded after 10 weeks of neoadjuvant chemotherapy. We believe this finding
578 has important implications for molecular profiling strategies in osteosarcoma, as it suggests that
579 the primary resection sample, and not the pretreatment biopsy, is more reflective of the
580 metastatic potential for a tumor than the pretreatment biopsy, due to the selection pressure of
581 neoadjuvant chemotherapy. Achieving a cure in osteosarcoma requires the extinction of all
582 cancer cells with a successful "first-strike' strategy with maximum tolerated doses of cisplatin,
583 doxorubicin, and methotrexate[30]. For patients for whom this first strike fails (poor necrosis at the

584 time of primary resection), characterizing and targeting the treatment-resistant population of
585 cancer cells using a second-strike strategy may prove to be an effective treatment strategy[30]. Our
586 work highlights that molecular profiling of primary resection samples could allow for a more
587 precise "genome-informed" approach[4], aimed at targeting resistant copy number alterations, to
588 augment MAP chemotherapy.

589 The emergence of subclonal copy number alterations in primary tumors to fully clonal
590 alterations in metastatic or recurrent samples, as demonstrated in our study in 6/7 patients with
591 recurrent/refractory disease and pretreatment samples, has been previously described in a subset
592 of adult cancers where analysis of matched primary and metastasis was performed[21]. We found
593 *MYC* gain/amplification to be enriched in the treatment-resistant clone in 6/7 patients with more
594 than one clone. Previous studies have shown that *MYC* amplification is often enriched in
595 metastatic sites[21] and has been previously associated with poor outcomes and increased cell
596 proliferation in osteosarcoma;[31–35] however, a recent study questioned its prognostic
597 significance[36]. Our study demonstrated that in patients with localized and metastatic disease at
598 diagnosis, *MYC* amplification is subclonal in pretreatment samples and emerges after 10 weeks
599 of neoadjuvant chemotherapy, highlighting the importance of sample timing when considering
600 the prognostic value of *MYC* amplification. Furthermore, as we begin to define molecular risk
601 categories within osteosarcoma, our work demonstrates that profiling of post-treatment primary
602 resection samples may reveal previous subclonal amplifications in driver oncogenes; thus, this
603 time point would be more informative when assessing metastatic potential. While multi-region
604 profiling of post-treatment resection from OSCE10 revealed limited heterogeneity among the
605 different sites, there was one site where the metastatic clone was not present, highlighting the
606 potential risk of failing to profile the metastatic clone with single-sample strategies. When

607 considering future sequencing approaches, pooling DNA/RNA extracts from
608 multiple anatomically distinct tumor regions of the primary tumor could be a cost-effective way
609 to improve DNA yield and variant detection, while providing a more complete picture of
610 intratumoral heterogeneity[37].

611 Our chromosomal duplication timing analysis revealed that gains for the same patient
612 often clustered around the same time point, regardless of whether whole-genome duplication was
613 present. These bursts of duplications occurred prior to diagnosis, and there were no comparable
614 bursts of duplications in resection, metastatic, or recurrent samples that would reflect ongoing
615 instability. In a pan-cancer cohort, synchronous bursts of copy number gain were found to occur
616 in 57% of diploid samples and 78% of WGD samples[38]. We found that these clustered
617 duplication events were associated with catastrophic complex genomic rearrangement and
618 amplicon events that occurred before diagnosis, such as chromothripsis and tyfonas. In contrast
619 to a previous multi-region osteosarcoma study[15], we demonstrated that tumor ploidy remained
620 consistent across all samples for each patient, which is likely because our copy number calls
621 were inferred jointly across all samples for each patient, which can improve ploidy estimation in
622 sample sets with wide ranges of purity[20]. These findings of pre-diagnostic duplication events
623 followed by relative genomic stability support recent work that demonstrated that early
624 catastrophic events are responsible for the structural complexity in the osteosarcoma genome, as
625 opposed to sustained evolution and instability over time[39]. Whole-genome duplication was
626 confirmed in a subset of patients and was found to be a late event in all three cases. Previous
627 pan-cancer studies have found that WGD events are typically early events in a tumor's molecular
628 time history, but are often preceded by *TP53* inactivation[21,38,40]. Late WGD events have been
629 described in a cohort of patients with hepatocellular carcinoma (HCC), and they are typically

630 associated with larger tumors, leading to the conclusion that they may be the last step prior to
631 rapid growth and expansion[24]. Our data support a macroevolutionary model of evolution in
632 osteosarcoma[41], with a large number of genomic aberrations acquired over a short period of time
633 secondary to chromosomal instability events, followed by clonal selection, as opposed to
634 ongoing evolution.

635 Large-scale genomic sequencing studies in osteosarcoma have revealed that there is
636 significant inter-tumoral heterogeneity in osteosarcoma, with shared mutations typically in tumor
637 suppressor genes rather than in targetable oncogenes[5–8]. We observed a heterogeneous mix of
638 metastatic and recurrent seeding patterns in our cohort. We observed only one example of
639 monoclonal, monophyletic dissemination in OSCE6, which is typically a result of a treatment-
640 induced bottle-necking effect. There were three cases of polyclonal/monophyletic dissemination,
641 in which multiple clones were present in the metastatic/recurrent samples, but they all shared a
642 common ancestor, and there were four cases of polyclonal polyphyletic dissemination where
643 multiple distinct clones from the primary tumor seeded a metastatic site. In previous studies, a *de*
644 *novo* induced murine model of osteosarcoma demonstrated polyclonal seeding of metastases
645 with ongoing parallel evolution[14], while studies using longitudinal and spatially separated
646 samples have yielded mixed results, demonstrating both polyclonal seeding with parallel
647 evolution[13] and monoclonal monophyletic seeding[15] in a majority of the respective cases from
648 each study. These studies were limited by the lack of pretreatment primary tumors; therefore, the
649 analysis relied on comparing metastatic and recurrent samples to post-treated primaries in many
650 cases.

651 We demonstrate that while cisplatin and HRD-related SBS3 are active mutagenic

652 processes in osteosarcoma, accounting for most new mutations in relapsed disease, we found no
653 new driver SNVs attributable to these signatures that could account for treatment resistance. The
654 HRD-related SBS3 signature was conserved across all samples in our cohort, consistent with a
655 recent pan-pediatric cancer study that found that 18/19 (95%) patients with osteosarcoma had
656 mutations attributed to HRD-related SBS3[42]. The prevalence of the HRD-related SBS3 signature
657 in osteosarcoma is comparable to BRCA1 deficient cancers, suggesting that drugs that target
658 homologous recombination deficient cells, such as PARP inhibitors, may have therapeutic value
659 in osteosarcoma, a concept currently being evaluated in a phase II clinical trial
660 (NCT04417062)[43]. A commonly cited limitation of using signature-based assays to assess HRD
661 is that they reflect the HRD state prior to sample acquisition and may not reflect the current state,
662 where HRD may have been restored[44]. Our study demonstrated that in osteosarcoma, the number
663 of mutations attributed to HRD-related SBS3 increases at each time point in patients with
664 recurrent disease, suggesting that HRD continues to be an active mutagenic process after
665 diagnosis.

666 The cisplatin signature was present in all relapse samples and metastatic sites that were
667 thought to have developed during upfront therapy. The extent of cisplatin-induced mutagenesis
668 has been previously described in osteosarcoma, where it was found that cisplatin therapy could
669 potentially increase the mutational burden by two-fold[15]. Although we also found that cisplatin
670 therapy led to large increases in mutational burden in recurrent samples, none of these mutations
671 were likely drivers of treatment resistance, which is consistent with previous studies in patients
672 with platinum-resistant ovarian cancer[45] and osteosarcoma[15]. Although we cannot account for
673 copy number alterations or structural variants induced by cisplatin, recent cell line work in

674 cisplatin exposed esophageal and liver tumors, found few copy number alterations or structural
675 variants, suggesting that cisplatin does not contribute significantly to genomic instability[46].
676 Our findings highlight that the chemoresistant population of tumor cells in osteosarcoma
677 is subclonal at diagnosis and is characterized by unique oncogenic amplifications. As our ability
678 to target these oncogenic amplifications improves, future studies aimed at identifying these
679 oncogenic drivers during upfront therapy may be an effective strategy to eliminate
680 chemoresistant tumor cells and improve survival.

681 **Methods**

682 *Patient consent and tissue processing*

683 This study was approved by the Institutional Review Board of the Memorial Sloan
684 Kettering Cancer Center (New York, NY, USA) and conducted in accordance with the
685 Declaration of Helsinki. Informed written consent was obtained from each subject or guardian.
686 Tumor samples and matched normal samples were collected from 10 patients with a
687 pathologically confirmed diagnosis of osteosarcoma, who were identified both retrospectively
688 and prospectively for those who had their tumor banked at diagnosis and at least one other time
689 point. Only patients who consented to an IRB-approved blood and tumor collection protocol
690 were eligible for tumor sequencing. Fresh tumor samples were procured from the operating room
691 in a sterile container. The tissue was processed using scalpels and divided into pea-sized pieces
692 before being stored at -80ºC. Frozen tissue samples from several patients were also available
693 through our Precision Pathology Biobanking Center and were acquired using the same protocol.
694 Additionally, in several patients, archival tumor specimens in the form of formalin-fixed
695 paraffin-embedded (FFPE) specimens were obtained from both the internal and external
696 pathology departments using the same protocol. Only FFPE samples that were not subjected to
697 harsh decalcification techniques were selected. FFPE samples that had been decalcified using
698 EDTA were deemed appropriate for further downstream analysis.

699 Each frozen tissue sample was submitted to our pathology core, where it was embedded
700 in Tissue-Tek optimum cutting temperature compound and sectioned at 5–10 mm on a Leica
701 Cryostat to create a hematoxylin and eosin–stained (H&E) slide for review. Each FFPE sample
702 was sectioned using a Leica Microtome. H&E slides were evaluated by a trained pathologist to
703 determine tumor content. After pathologic review, tumor samples were isolated via a 21-gauge

704 punch, curl biopsy, or macro-dissected from sectioned slides to remove non-neoplastic
705 components. The neoplastic component of each tumor underwent genomic DNA extraction using
706 a Qiagen DNAeasy Blood and Tissue Kit and protocol, whereas the FFPE samples were
707 extracted using a QIAamp DNA FFPE Tissue Kit and protocol.
708 PBMCs utilized for matched normal sequencing were brought up to 15mL volume in
709 cold PBS and isolated with the DNeasy Blood & Tissue Kit (QIAGEN catalog # 69504)
710 according to the manufacturer's protocol and incubated at 55°C for digestion. DNA was eluted in
711 0.5X Buffer AE.
712
713 *Whole-genome sequencing and alignment*
714 DNA quantification, library preparation, and whole-genome sequencing were performed
715 using the Integrated Genomics Operation at the Memorial Sloan Kettering Cancer Center (New
716 York, NY). After PicoGreen quantification and quality control using an Agilent BioAnalyzer,
717 131-500ng of genomic DNA was sheared using an LE220-plus Focused-ultrasonicator (Covaris
718 catalog # 500569), and sequencing libraries were prepared using the KAPA Hyper Prep Kit
719 (Kapa Biosystems KK8504) with modifications. Briefly, libraries were subjected to a 0.5 × size
720 selection using aMPure XP beads (Beckman Coulter catalog # A63882) after post-ligation
721 cleanup. Libraries with < 500 ng of genomic DNA were amplified using 5-6 cycles of PCR and
722 pooled equimolar amounts. Libraries containing at least 500 ng of genomic DNA were not
723 amplified. Samples were run on a NovaSeq 6000 in a 150bp/150bp paired end run, using the
724 NovaSeq 6000 SBS v1 Kit and an S1, S2, or S4 flow cell (Illumina). The average number of read
725 pairs per normal was 614 million and the average number of read pairs per tumor was 1.3 billion.
726

727 *Whole Genome Sequencing Pipeline*

728 Whole genome sequencing data were analyzed using the ISABL[47] platform, with
729 methods previously described in detail[48]. The additional downstream analyses are described
730 below.

731

732 *Single Nucleotide Variant Filtering*

733 For all eight patients, single nucleotide variants (SNV) were called in triplicate by
734 MuTect[49], Strelka[50], and Caveman[51]. Only mutations that had a "PASS" flag, were called by at
735 least two mutation callers and observed in less than 2% of the reads of the matched normal
736 sample with 10x coverage were considered for further analysis. It is notoriously difficult to
737 extract high quality DNA from the osteoid matrix that surrounds viable tumor cells, often leading
738 to sequenced samples with low purity. Across the eight patients, a mix of samples was prepared
739 as either fresh-frozen or FFPE. FFPE samples are thought to be inferior to fresh frozen samples,
740 as the formalin fixation process results in nucleic acid fragmentation, DNA crosslinks, and
741 deamination, leading to C>T mutation artifacts. As a result, downstream analysis of FFPE can be
742 challenging when filtering out artifacts from a true positive. Aggressive filtering of low-allele
743 frequency variants has been shown to increase SNV overlap when comparing matched FFPE and
744 FF samples[52]. Given these assumptions, a custom filtering approach was utilized to maximize the
745 ability to utilize low-purity FFPE samples within a mix of higher-quality FFPE samples and
746 fresh frozen samples for each patient. For fresh frozen samples with an estimated purity of 20%,
747 mutations with a variant allele frequency (VAF) less than 5 were filtered out. If a frozen sample
748 had a purity of less than 20%, no additional filtering was applied, given the potential for filtering
749 out subclonal mutations in an otherwise high-quality frozen tissue sample. For FFPE samples,

750 we filtered out mutations with a VAF less than 20%. These thresholds for FFPE samples were
751 then further purity adjusted, for example a sample with an estimated purity of 80%, would have
752 mutations below a VAF of 0.16 (0.8 purity * 0.2 VAF filter) initially filtered out. Although this
753 initial filtering for FFPE is strict, the advantage of having multiple samples per patient allows us
754 to utilize the mutation calls in other high-quality samples for a patient to rescue mutations that
755 may have been initially filtered in a low-quality sample. This is accomplished through a pile-up
756 rescue, where all filtered mutations for a patient are combined and then specifically searched in
757 the BAM file that was generated for each sample.
758
759 *Driver Gene Analysis*
760 All somatic variants that led to a frameshift insertion, frameshift deletion, in-frame
761 insertion, in-frame deletion, missense, nonsense, nonstop, or splice site/region mutation, or a
762 translation start site were considered. For variants identified as missense or nonsense, we
763 required the variant to be considered a likely functional driver using the LiFD tool[18], which is a
764 two-phase algorithm that pulls from various databases and bioinformatic methods to determine
765 whether a given mutation is likely to be functional. We also considered genes that were
766 significantly mutated in large pediatric cancer and osteosarcoma sequencing studies[4–9,53]. The
767 final list consisted of 639 genes.
768
769 *Evolutionary Analysis*
770 To determine the spatial and temporal dynamics of subclonal diversity within a patient,
771 we first used Treeomics v1.9.2[54] to derive phylogenies. Treeomics reconstructs the phylogeny of
772 metastatic lesions and maps subclones to their anatomical locations. Treeomics utilizes a

773 Bayesian inference model to account for error-prone sequencing and differing neoplastic cell
774 contents to calculate the likelihood that a specific variant is present or absent. Treeomics then
775 infers a global optimal tree based on mixed-integer linear programming[54].

776 The HATCHet[20] v1.0.1 algorithm was used to infer allele and clone-specific copy-
777 number aberrations, clone proportions, and whole-genome duplications (WGD) for each patient.
778 HATCHet was run using the GATK4-CNV custom pipeline, with Battenberg copy number calls
779 fitted to meet the input requirements for running the tool. Solutions were manually reviewed
780 with the creator of the tool, Simone Zaccaria PhD, to allow for advanced fine-tuning and ensure
781 that the most accurate solutions were selected.

782 The DeCiFer[17] v1.1.5 algorithm was used to cluster mutations across all samples for each
783 patient, providing descendant cell fractions (DCF) and cancer cell fractions for each cluster. The
784 copy number input for this algorithm is the output of the HATCHet algorithm. Custom state trees
785 were generated utilizing a maximum copy number between 6-8 for each patient (lower maximum
786 copy number states were selected if the runtime exceeded 48 h). After clustering of mutations,
787 the CALDER[26] v0.11 algorithm was used to infer evolutionary phylogenies. To run the DeCiFer
788 output through CALDER, the inferred cluster DCF was converted to a read count by multiplying
789 the DCF by 1000 and then dividing by 2 (because CALDER assumes that all mutations are in
790 heterozygous diploid regions). Therefore, if a mutation has 1000 reads and an inferred DCF of
791 40%, the corresponding input will have 200 variant reads and 800 reference reads. Longitudinal
792 constraints were lifted when analyzing OSCE2 and OSCE3, because all analyzed samples were
793 present at diagnosis.

794 The Palimpest[23] algorithm (version = github commit 4795da2) was used to characterize
795 and visualize mutational signatures (using Cosmic SBS/DBS v3.2) at both clone and sample

levels. This algorithm also provided information regarding the timing of duplication and loss of heterozygosity events using previously described methods[24].

Structural Variant Analysis

Structural variants were annotated using iAnnotateSV software[55]. We used svpluscnv[56] and ShatterSeek[57] to identify regions of chromothripsis. The JaBba tool[25] was used to identify regions with complex rearrangements or amplicon events. The calls from all tools were combined for further downstream analysis.

Data Visualizations

Oncoprint was created using the CoMut[58] tool Timescape[59] (https://github.com/shahcompbio/timescape) and Mapscape[59] (https://github.com/shahcompbio/mapscape) were used to visualize temporal and spatial clonal evolution. Tableau Desktop (v2021.4) was used to analyze and visualize the data with charts, bars, and line graphs. Ridgeline graphs were created using R utilizing the ggridges package (https://cran.r-project.org/web/packages/ggridges/vignettes/introduction.html). Sankey plots were created using SankeyMATIC (https://sankeymatic.com/build/). Anatomic cartoons were created using BioRender (https://biorender.com/).

Data Availability

Sequence data will be deposited at the European Genomephenome Archive (EGA), which is hosted by the European Bioinformatics Institute and the Centre for Genomic Regulation. Further information about EGA can be found at https://ega-archive.org and "The European Genome-

phenome Archive of human data consented for biomedical research"
(http://www.nature.com/ng/journal/v47/n7/full/ng.3312.html).

821 **References**

822

823 1. Kager L, Zoubek A, Pötschger U, et al. Primary metastatic osteosarcoma: Presentation
824 and outcome of patients treated on neoadjuvant Cooperative Osteosarcoma Study
825 Group protocols. *Journal of Clinical Oncology*. 2003;21(10):2011-2018.
826 doi:10.1200/JCO.2003.08.132
827 2. Kim MS, Cho WH, Song WS, Lee SY, Jeon DG. Time dependency of prognostic factors in
828 patients with stage II osteosarcomas. *Clin Orthop Relat Res*. 2007;463:157-165.
829 doi:10.1097/BLO.0B013E318142B27D
830 3. Bielack SS, Kempf-Bielack B, Delling G, et al. Prognostic Factors in High-Grade
831 Osteosarcoma of the Extremities or Trunk: An Analysis of 1,702 Patients Treated on
832 Neoadjuvant Cooperative Osteosarcoma Study Group Protocols. *Journal of Clinical*
833 *Oncology*. 2002;20(3):776-790. doi:10.1200/jco.2002.20.3.776
834 4. Sayles LC, Breese MR, Koehne AL, et al. Genome-Informed Targeted Therapy for
835 Osteosarcoma. *Cancer Discov*. Published online September 28, 2018:CD-17-1152.
836 doi:10.1158/2159-8290.CD-17-1152
837 5. Perry JA, Kiezun A, Tonzi P, et al. Complementary genomic approaches highlight the
838 PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc Natl Acad Sci U S*
839 *A*. 2014;111(51):E5564-73. doi:10.1073/pnas.1419260111
840 6. Suehara Y, Alex D, Bowman A, et al. Clinical Genomic Sequencing of Pediatric and Adult
841 Osteosarcoma Reveals Distinct Molecular Subsets with Potentially Targetable Alterations.
842 *Clinical Cancer Research*. 2019;25(21):6346-6356. doi:10.1158/1078-0432.CCR-18-4032
843 7. Behjati S, Tarpey PS, Haase K, et al. Recurrent mutation of IGF signalling genes and
844 distinct patterns of genomic rearrangement in osteosarcoma. *Nat Commun*.
845 2017;8(May):1-8. doi:10.1038/ncomms15936
846 8. Chen X, Bahrami A, Pappo A, et al. Cell Reports Report Recurrent Somatic Structural
847 Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma. Published online
848 2014. doi:10.1016/j.celrep.2014.03.003
849 9. Kovac M, Blattmann C, Ribi S, et al. Exome sequencing of osteosarcoma reveals mutation
850 signatures reminiscent of BRCA deficiency. *Nat Commun*. 2015;6(1):8940.
851 doi:10.1038/ncomms9940
852 10. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr. LA, Kinzler KW. Cancer
853 Genome Landscapes. *Science (1979)*. 2013;339(6127):1546-1558.
854 doi:10.1126/science.1235122
855 11. Meltzer PS, Helman LJ. New Horizons in the Treatment of Osteosarcoma. *New England*
856 *Journal of Medicine*. 2021;385(22):2066-2076.
857 doi:10.1056/NEJMRA2103423/SUPPL_FILE/NEJMRA2103423_DISCLOSURES.PDF
858 12. Persha HE, Kato S, De P, et al. Osteosarcoma with cell-cycle and fibroblast growth factor
859 genomic alterations: case report of Molecular Tumor Board combination strategy
860 resulting in long-term exceptional response. *J Hematol Oncol*. 2022;15(1):1-5.
861 doi:10.1186/S13045-022-01344-X/FIGURES/2
862 13. Xu H, Zhu X, Bao H, et al. Genetic and clonal dissection of osteosarcoma progression and
863 lung metastasis. *Int J Cancer*. 2018;143(5):1134-1142. doi:10.1002/ijc.31389

14. Gambera S, Abarrategi A, González-Camacho F, et al. Clonal dynamics in osteosarcoma defined by RGB marking. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-06401-z
15. Brady SW, Ma X, Bahrami A, et al. The clonal evolution of metastatic osteosarcoma as shaped by cisplatin treatment. *Molecular Cancer Research*. 2019;17(4):895-906. doi:10.1158/1541-7786.MCR-18-0620
16. Goswami RS, Luthra R, Singh RR, et al. Identification of Factors Affecting the Success of Next-Generation Sequencing Testing in Solid Tumors. *Am J Clin Pathol*. 2016;145(2):222-237. doi:10.1093/ajcp/aqv023
17. Satas G, Zaccaria S, El-Kebir M, Raphael BJ. DeCiFering the elusive cancer cell fraction in tumor heterogeneity and evolution. *Cell Syst*. Published online August 2021. doi:10.1016/j.cels.2021.07.006
18. Reiter JG, Baretti M, Gerold JM, et al. An analysis of genetic heterogeneity in untreated cancers. *Nat Rev Cancer*. 2019;19(11):639-650. doi:10.1038/s41568-019-0185-x
19. Shao YW, Wood GA, Lu J, et al. Cross-species genomics identifies DLG2 as a tumor suppressor in osteosarcoma. *Oncogene 2018 38:2*. 2018;38(2):291-298. doi:10.1038/s41388-018-0444-4
20. Zaccaria S, Raphael BJ. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *bioRxiv*. Published online 2018:496174. doi:10.1101/496174
21. Watkins TBK, Lim EL, Petkovic M, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature*. 2020;587(7832):126-132. doi:10.1038/s41586-020-2698-6
22. McGranahan N, Rosenthal R, Hiley CT, et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*. 2017;171(6):1259-1271.e11. doi:10.1016/j.cell.2017.10.001
23. Shinde J, Bayard Q, Imbeaud S, et al. Palimpsest: An R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics*. 2018;34(19):3380-3381. doi:10.1093/bioinformatics/bty388
24. Letouzé E, Shinde J, Renault V, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nature Communications 2017 8:1*. 2017;8(1):1-13. doi:10.1038/s41467-017-01358-x
25. Hadi K, Yao X, Behr JM, et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell*. 2020;183(1):197-210.e32. doi:10.1016/j.cell.2020.08.006
26. Myers MA, Satas G, Raphael BJ. CALDER: Inferring Phylogenetic Trees from Longitudinal Tumor Samples. *Cell Syst*. 2019;8(6):514-522.e5. doi:10.1016/j.cels.2019.05.010
27. Németh E, Krzystanek M, Reiniger L, et al. The genomic imprint of cancer therapies helps timing the formation of metastases. *Int J Cancer*. 2019;145(3):694-704. doi:10.1002/ijc.32159
28. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. 2017;168(4):613-628. doi:10.1016/j.cell.2017.01.018
29. Smith JC, Sheltzer JM. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *Elife*. 2018;7. doi:10.7554/ELIFE.39217

908 30. Reed DR, Metts J, Pressley M, et al. An evolutionary framework for treating pediatric
909 sarcomas. *Cancer*. 2020;126(11):2577-2587. doi:10.1002/cncr.32777
910 31. Ueda T, Healey JH, Huvos AG, Ladanyi M. Amplification of the MYC Gene in
911 Osteosarcoma Secondary to Paget's Disease of Bone. *Sarcoma*. 1997;1(3-4):131-134.
912 doi:10.1080/13577149778209
913 32. Shimizu T, Ishikawa T, Sugihara E, et al. c-MYC overexpression with loss of Ink4a/Arf
914 transforms bone marrow stromal cells into osteosarcoma accompanied by loss of
915 adipogenesis. *Oncogene*. 2010;29(42):5687-5699. doi:10.1038/onc.2010.312
916 33. Scionti I, Michelacci F, Pasello M, et al. Clinical impact of the methotrexate resistance-
917 associated genes C-MYC and dihydrofolate reductase (DHFR) in high-grade
918 osteosarcoma. *Annals of Oncology*. 2008;19(8):1500-1508. doi:10.1093/annonc/mdn148
919 34. Gamberi G, Benassi MS, Bohling T, et al. C-myc and c-fos in Human Osteosarcoma:
920 Prognostic Value of mRNA and Protein Expression. *Oncology*. 1998;55(6):556-563.
921 doi:10.1159/000011912
922 35. Ladanyi M, Park CK, Lewis R, Jhanwar SC, Healey JH, Huvos AG. Sporadic amplification of
923 the MYC gene in human osteosarcomas. *Diagn Mol Pathol*. 1993;2(3):163-167.
924 36. Taylor AM, Sun JM, Yu A, et al. Integrated DNA Copy Number and Expression Profiling
925 Identifies IGF1R as a Prognostic Biomarker in Pediatric Osteosarcoma. *Int J Mol Sci*.
926 2022;23(14):8036. doi:10.3390/ijms23148036
927 37. Litchfield K, Stanislaw S, Spain L, et al. Representative Sequencing: Unbiased Sampling of
928 Solid Tumor Tissue. *Cell Rep*. 2020;31(5):107550. doi:10.1016/J.CELREP.2020.107550
929 38. Gerstung M, Jolly C, Leshchiner I, et al. The evolutionary history of 2,658 cancers. *Nature*.
930 2020;578(7793):122-128. doi:10.1038/s41586-019-1907-7
931 39. Rajan S, Zaccaria S, Cannon M v, et al. Structurally complex osteosarcoma genomes
932 exhibit limited heterogeneity within individual tumors and across evolutionary time.
933 *bioRxiv*. Published online January 1, 2022:2021.08.30.458268.
934 doi:10.1101/2021.08.30.458268
935 40. Bielski CM, Zehir A, Penson A v., et al. Genome doubling shapes the evolution and
936 prognosis of advanced cancers. *Nat Genet*. 2018;50(8):1189-1195. doi:10.1038/S41588-
937 018-0165-1
938 41. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. *EMBO J*.
939 2021;40(18):e108389. doi:10.15252/EMBJ.2021108389
940 42. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699
941 paediatric leukaemias and solid tumours. *Nature 2018 555:7696*. 2018;555(7696):371-
942 376. doi:10.1038/nature25795
943 43. Forrest SJ, Kinnaman MD, Livingston JA, et al. Phase II trial of olaparib in combination
944 with ceralasertib in patients with recurrent osteosarcoma.
945 *https://doi.org/101200/JCO20213915_supplTPS11575*. 2021;39(15_suppl):TPS11575-
946 TPS11575. doi:10.1200/JCO.2021.39.15_SUPPL.TPS11575
947 44. Miller RE, Elyashiv O, El-Shakankery KH, Ledermann JA. Ovarian Cancer Therapy:
948 Homologous Recombination Deficiency as a Predictive Biomarker of Response to PARP
949 Inhibitors. *Onco Targets Ther*. 2022;15:1105-1117. doi:10.2147/OTT.S272199

45. Patch AM, Christie EL, Etemadmoghadam D, et al. Whole–genome characterization of chemoresistant ovarian cancer. *Nature 2015 521:7553*. 2015;521(7553):489-494. doi:10.1038/nature14410
46. Boot A, Huang MN, Ng AWT, et al. In-depth characterization of the cisplatin mutational signature in human cell lines and in esophageal and liver tumors. *Genome Res*. 2018;28(5):654-665. doi:10.1101/GR.230219.117
47. Medina-Martínez JS, Arango-Ossa JE, Levine MF, et al. Isabl Platform, a digital biobank for processing multimodal patient data. *BMC Bioinformatics*. 2020;21(1):1-18. doi:10.1186/S12859-020-03879-7/TABLES/1
48. Shukla N, Levine MF, Gundem G, et al. Feasibility of whole genome and transcriptome profiling in pediatric and young adult cancers. *Nature Communications 2022 13:1*. 2022;13(1):1-15. doi:10.1038/s41467-022-30233-7
49. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219. doi:10.1038/nbt.2514
50. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594. doi:10.1038/s41592-018-0051-x
51. Jones D, Raine KM, Davies H, et al. cgpCaVEManWrapper: Simple Execution of CaVEMan in Order to Detect Somatic Single Nucleotide Variants in NGS Data. *Curr Protoc Bioinformatics*. 2016;56:15.10.1-15.10.18. doi:10.1002/cpbi.20
52. Robbe P, Popitsch N, Knight SJL, et al. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. *Genetics in Medicine 2018 20:10*. 2018;20(10):1196-1205. doi:10.1038/gim.2017.241
53. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321-327. doi:10.1038/nature25480
54. Reiter JG, Makohon-Moore AP, Gerold JM, et al. Reconstructing metastatic seeding patterns of human cancers. *Nat Commun*. 2017;8:1-10. doi:10.1038/ncomms14114
55. Shah RH. iAnnotateSV: Annotation of structural variants detected from NGS. Published 2021. Accessed November 2, 2021. https://github.com/rhshah/iAnnotateSV
56. Lopez G, Egolf LE, Giorgi FM, Diskin SJ, Margolin AA. Svpluscnv : Analysis and Visualization of Complex Structural Variation Data . *Bioinformatics*. 2020;(October):1-3. doi:10.1093/bioinformatics/btaa878
57. Cortés-Ciriano I, Lee JJK, Xi R, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nature Genetics 2020 52:3*. 2020;52(3):331-341. doi:10.1038/s41588-019-0576-7
58. Crowdis J, He MX, Reardon B, van Allen EM. CoMut: visualizing integrated molecular information with comutation plots. *Bioinformatics*. 2020;36(15):4348-4349. doi:10.1093/BIOINFORMATICS/BTAA554
59. Smith MA, Nielsen CB, Chan FC, et al. E-scape: interactive visualization of single-cell phylogenetics and cancer evolution. *Nature Methods 2017 14:6*. 2017;14(6):549-550. doi:10.1038/nmeth.4303