

1 Drivers underpinning the malignant transformation of giant cell 2 tumour of bone

3 Matthew W. Fittall^{1,2,3}, Peter Ellery^{2,4}, Iben Lyskjær², Patrick Lombard², Jannat Ijaz³,
4 Anna-Christina Strobl⁴, Dahmane Oukrif², Maxime Tarabichi^{1,3}, Martin Sill^{5,6},
5 Christian Koelsche^{6,7}, Jonas Demeulemeester^{1,8}, Grace Collord³, Roberto
6 Tirabosco⁴, Fernanda Amary⁴, Peter J. Campbell³, Stefan Pfister^{6,7}, David T.W.
7 Jones^{6,7}, Nischalan Pillay^{2,4}, Peter Van Loo^{1,8}, Sam Behjati^{#3,9}, Adrienne M.
8 Flanagan^{#2,4}

9 [#]These authors jointly directed and contributed equally to this research.

10 Affiliations

11 ¹The Francis Crick Institute, London, NW1 1AT, UK

12 ²University College London Cancer Institute, London, WC1E 6DD, UK

13 ³Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

14 ⁴Department of Histopathology, Royal National Orthopaedic Hospital NHS Trust,
15 Stanmore, UK

16 ⁵Division of Biostatistics, German Cancer Research Center (DKFZ), 69120
17 Heidelberg, Germany

18 ⁶Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ) and
19 German Cancer Consortium (DKTK), 69120 Heidelberg, Germany

20 ⁷Department of Neuropathology, Institute of Pathology, University Hospital
21 Heidelberg, 69120 Heidelberg, Germany

22 ⁸Department of Human Genetics, University of Leuven, Leuven, Belgium

23 ⁹Department of Paediatrics, University of Cambridge, Cambridge, CB2 0QQ, UK

24 Correspondence and requests for material should be directed to S.B. (email:

25 sb31@sanger.ac.uk) and A.M.F. (email: a.flanagan@ucl.ac.uk)

26 **Drivers underpinning the malignant transformation of giant cell**
27 **tumour of bone**

28 **The rare benign giant cell tumour of bone (GCTB) is defined by an almost**
29 **unique G34W oncohistone mutation in the H3.3 histone gene. Here we reveal**
30 **the genomic and methylation patterns underlying the rare clinical phenomena**
31 **of benign metastases and malignant transformation of GCTB.**

32 Giant cell tumour of bone (GCTB) is a locally destructive benign tumour, prone to
33 local recurrence. They present predominantly at the site of the mature
34 epiphysis/epimetaphysis of the long bones, particularly the distal femur and proximal
35 tibia in the 3rd and 4th decades of life¹. GCTB is defined by a near universal (96%)
36 pathognomonic H3F3A G34W missense mutation²⁻⁶. Two unexplained phenomena
37 in GCTB are of particular interest, namely that lung metastases occur despite the
38 absence of malignant histological features in either the primary or metastatic lesions⁷
39 and secondly that the characteristic H3.3 mutation is occasionally found in primary
40 malignant bone tumours which often share features with conventional GCTB^{4,8}. We
41 set out to explore the genomic events underlying these phenomena using whole
42 genome sequencing and genome-wide methylation profiling using methylation array
43 and whole genome bisulfite sequencing.

44 We started our investigation by performing whole genome sequencing (WGS) on
45 seven primary malignant bone tumours possessing an H3.3 mutation, one case of
46 metastatic GCTB, and nine conventional GCTB for which we had frozen tissue with
47 corresponding germline DNA from blood samples (**Supplementary Table 1**). We
48 used the analysis pipeline of the *Cancer Genome Project* to generate catalogues of

49 somatic mutations, indels, structural variants and copy number changes and a
50 previously reported strategy to identify putative drivers (**Methods**)⁹.

51 Benign GCTB genomically resemble other benign mesenchymal tumours (**Figure**
52 **1a**). They possess few somatic changes of any type and no plausible driver
53 mutations other than the canonical H3.3 mutation (medians: 640 substitutions, 43
54 indels, 7 structural variants; **Figure 1a and Supplementary Figure 1**). In contrast,
55 we found that malignant bone tumours with the same H3.3 mutation possess
56 genomic features resembling osteosarcoma. They possessed an increased burden
57 of somatic variants, and broadly divided into two groups: 3/7 tumours exhibited a
58 modest increase in mutations (medians: 1815 substitutions, 86 indels, 21 structural
59 variants) and the remaining four possessed a greater mutation burden (medians:
60 4177 substitutions, 205 indels, 108 structural variants; Figure 1a and
61 **Supplementary Figure 1**).

62 Unlike osteosarcoma, malignant H3.3-mutated bone tumours are enriched with
63 mutations suggesting telomere dysfunction. Two tumours had mono-allelic G>A
64 mutations 124bp upstream of the *TERT* transcription start site, reported to increase
65 promoter binding¹⁰. Another sample, PD3788d, had a complex rearrangement event,
66 resembling chromothripsis, encompassing *TERT*, resulting in the juxtaposition of the
67 gene *MEGF10* with the *TERT* promoter (**Supplementary Figure 2**). *MEGF10* is
68 reported to be under the control of a super-enhancer in the dbSUPER database¹¹.
69 Two other malignant samples, PD4922e and PD30985a, were identified as having
70 markedly elongated telomeres (**Figure 1b, Supplementary Figure 1 and Methods**),
71 a finding consistent with the Alternative Lengthening of Telomeres (ALT)
72 mechanism, which is usually mutually exclusive with *TERT* alteration^{12,13}. In keeping

73 with this pattern of ALT, recently reported in other sarcoma types¹⁴, these tumours
74 possessed highly rearranged genomes, the telomeres of which comprised
75 conventional ('TTAGGG') repeats and had undergone loss of heterozygosity at the
76 *RB1* locus (**Supplementary Figure 3**). In total, 5/7 malignant tumours had evidence
77 of a *TERT*-mutated phenotype. In contrast, only *TERT* amplifications have previously
78 been reported in osteosarcoma⁹. The remaining two malignant tumours both
79 harbored biallelic losses of an additional histone lysine demethylase, *KDM4B* or
80 *KDM5A* (**Supplementary Figure 3**). All malignant tumours had thus acquired at
81 least one additional driver mutation in addition to the G34W.

82 The degree of aneuploidy observed in 3/7 malignant tumours, against a backdrop of
83 almost ubiquitously diploid GCTBs (**Supplementary Figure 4**), allowed the ordering
84 and timing, in real-time, of the most significant mutational events. In all three cases
85 (PD4922e, PD30985a, and PD3788d), whole genome duplication (WGD) had
86 occurred in adulthood, but several years prior to diagnosis. Chromothripsis had
87 occurred subsequent to this. In 2/3 samples (PD4922e, PD30985a), with informative
88 data, the H3.3 histone mutation had also been duplicated, demonstrating its
89 occurrence as an earlier mutational event prior to WGD (**Supplementary Figure 5**).
90 This is consistent with the progression of these malignant tumours from GCTBs.

91 We next investigated the 'benign metastasising GCTB'¹. In contrast to malignant
92 tumours, the morphology of both the metastases and primary lesion (PD38329a/c/d)
93 was that of a conventional GCTB which was reflected in the low mutational burden
94 and the absence of additional driver mutations (**Figure 1b and Supplementary**
95 **Figure 1**). Leveraging the independent sampling across these three tumour samples
96 increased the power to define the clonality of mutations. Clonal mutations are those

97 found in all tumours cells whereas those in only a fraction of cells are considered
98 subclonal. The primary tumour (PD38329a) and the two metastases (PD38329c and
99 PD38329d) each possessed a group of private mutations to only that tumour sample.
100 Furthermore, one cluster of mutations (**Supplementary Figure 6**) was found to be
101 common but subclonal in all samples. This suggests that both metastases were
102 seeded by at least one cell possessing those mutations and at least one cell that did
103 not, a process known as polyclonal seeding.

104 To explore the epigenetic differences between malignant and benign H3.3-mutated
105 bone tumours we collected additional tumours for DNA methylation array analysis
106 (**Methods**). This collection (n = 121), included some of the sequenced samples,
107 osteosarcomas without H3.3 mutations, and chondroblastomas, a benign tumour
108 with an alternative H3.3 mutation, *H3F3B*:p.K27M (**Supplementary Table 2**).
109 Unsupervised clustering based on the most variable methylation probes
110 recapitulated the diagnostic groups (**Figure 2a and Supplementary Figure 7**).
111 Furthermore, while closely related to conventional GCTB, the malignant H3.3-mutant
112 tumours formed a distinct clade. The benign metastasising samples clustered with
113 the benign GCTB group.

114 To unpick the methylation differences underlying the separate clustering of the
115 benign and malignant H3.3 mutated tumours, differentially methylated regions
116 (DMRs) were identified. This revealed focal changes in a small number of specific
117 gene promoters (**Methods**). Of 74 DMRs identified, 56 were located around gene
118 transcriptions start sites (**Supplementary Figure 8**). The most statistically significant
119 DMR was also the only one identified in a plausible cancer driver gene, *CCND1*,
120 which encodes Cyclin D1. Differential methylation spanned a promoter region of

121 1500bp either side of the transcription start site, a finding validated by bisulfite
122 sequencing (**Supplementary Figure 9**). Comparing the mean methylation level
123 across this promoter region between different bone and soft tissue tumour types
124 revealed that hypermethylation at this site is specific to GCT (**Figure 2d**). Malignant
125 histone-mutated tumours and chondrosarcomas were the only tumour types with a
126 similar degree of *CCND1* promoter methylation. *CCND1* promoter methylation was
127 concordant with unsupervised methylation clustering groups (**Figure 1** and **Figure**
128 **2**). Beyond this, methylation differences were enriched at non-enhancer intergenic
129 sites, however those that affected genes did not consistently alter gene pathways. At
130 a broader scale part of the cluster of histone genes on chromosome 6 was focally
131 hypermethylated in malignant tumours, suggesting additional epigenetic driver
132 events (**Supplementary Figure 10**).

133 Using comprehensive genomic and methylation profiling, we report the driver events
134 associated with malignant or metastatic progression of GCTBs. Malignant H3.3
135 mutated tumours are characterized by a methylation profile that is related but distinct
136 from conventional GCTB. Histone mutation predates the development of aneuploidy
137 in malignant tumours, which still occurs some years prior to diagnosis. Malignant
138 progression requires additional genetic mutations endowing either telomere or
139 epigenetic dysfunction, and possible additional epigenetic changes altering clusters
140 of histone genes. This combination of genomic and epigenomic findings could
141 potentially distinguish benign from malignant GCTs, thereby predicting aggressive
142 behavior in challenging diagnostic cases (**Figure 1b** and **Supplementary Table 3**): it
143 also permits malignant GCTB to be classified on a molecular basis distinguishing it
144 from other primary malignant bone tumours. The absence of additional genetic
145 events in metastatic, but histologically benign GCTB, and the presence of polyclonal

146 seeding supports the longstanding hypothesis that they represent a thrombotic
147 event.

148 **Methods**

149 **Patient samples**

150 Patients provided their written and informed consent to provide samples for this
151 study via the UCL Biobank for Health and Disease, based at the Royal National
152 Orthopaedic Hospital. This was approved by the National Research Ethics Service
153 (NRES) Committee Yorkshire & The Humber – Leeds East (15/YH/0311). DNA was
154 extracted from areas of fresh frozen tissue selected by bone pathologists
155 (A.M.F./R.T./F.A./P.E.). Matched normal DNA was acquired from blood samples.

156 **SNP and Methylation Array**

157 SNP analysis was performed using the Illumina HumanOmni2.5 BeadChip. DNA
158 Methylation analysis was performed on either Illumina Infinium
159 HumanMethylation450 or MethylationEPIC arrays. Pre-processing and quality
160 control were performed (**Supplementary Methods**).

161 **Array analysis**

162 SNP array copy number profiles were produced using ASCAT (v2.5.1). Methylation
163 array-based copy number profiles were generated using the conumee package
164 (v1.18.0) and a bespoke adaptation of the principles utilised by ASCAT
165 (**Supplementary Methods**). Unsupervised clustering was performed by hierarchical
166 clustering using the 5,000 most variable probes across samples after scrutiny for

167 batch effects (**Supplementary Methods**). Differentially methylated probes and
168 regions were detected using the ChAMP package (v2.14.0). Gene set enrichment
169 analysis (GSEA) was performed using an adapted approach from the ebBayes
170 function in the ChAMP package¹⁵. Bespoke analysis for regional differences in
171 methylation were performed using Circular Binary Segmentation (CBS) functions
172 from the DNACopy package (v1.58.0) based on a signal-noise ratio for each
173 methylation probe (**Supplementary Methods**).

174 **Whole genome bisulfite sequencing**

175 Whole genome bisulfite sequencing was performed on seven of the samples:
176 PD30981a, PD30982a, PD30984a, PD30985a, PD3788d, PD3795d and PD4915d.
177 Oxidative bisulfite conversion and library preparation was done using the Cambridge
178 Epigenetix Truemethyl Whole Genome kit following manufacturer's instructions. The
179 efficiency of bisulfite treatment was determined using control probes; 90.6% of 5-
180 methylcytosine remaining unconverted and 100% of unmethylated cytosines were
181 converted into thymines. Hydroxymethylation was not detected using this kit as
182 91.7% of 5-hydroxymethyl cytosine were converted to thymine. Libraries were
183 sequenced on an Illumina HiSeqX using a 150bp paired end run. Both mapping of
184 reads to GRCh37 and methylation calling was done using Bismark
185 (<https://github.com/FelixKrueger/Bismark>).

186 **Whole genome sequencing**

187 For whole genome sequencing the Illumina (Illumina, Chesterford, UK) no-PCR
188 library protocol was used to construct short insert 500bp libraries, prepare flowcells
189 and generate clusters. Whole genome sequencing was performed using the Illumina

190 HiSeq 2000 or 2500 platform, using 100 bp paired-end libraries. Samples PD37332,
191 PD3788, PD3795, PD38328, PD38329, PD4915, PD4922 were sequenced using the
192 XTen platform using 150 bp paired-end libraries.

193 **Variant detection, validation and analysis**

194 The Cancer Genome Project (Wellcome Trust Sanger Institute) variant calling
195 pipeline was used to call somatic mutations (versions as below). All variant calling
196 algorithms were used with standard settings with limited post-processing filtering and
197 variants were analysed using a previously documented strategy⁹ (**Supplementary**
198 **Methods**). Variants were considered as potential drivers if they presented in
199 established cancer genes (Chapter 3 COSMIC v82, Chapter 4 COSMIC v85).
200 Tumour suppressor coding variants were considered if they were annotated as
201 functionally deleterious by the VAGrENT algorithm
202 (<http://cancerit.github.io/VAGrENT/>). Disruptive rearrangements or homozygous
203 deletions of tumour suppressors were also considered. Additionally, homozygous
204 deletions were required to be focal (<1 Mb in size). Mutations in oncogenes were
205 considered driver events if they were located at previously reported hot spots (point
206 mutations) or amplified the intact gene. Amplifications also had to be focal (<1 Mb)
207 and result in at least 5 copies in diploid genomes, or 4 copies more than the modal
208 major copy number in genome duplicated samples.

209 **Copy number scoring**

210 A sample was considered Whole Genome Duplicated (WGD) when the modal total
211 copy number was >2. The baseline total copy number was considered as 4 for WGD
212 samples and 2 for others. Autosomal copy number segments were then scored as

213 the difference from this baseline; no difference (0), total copy number of 0
214 (homozygous deletion, 2), total copy number ≥ 3 + baseline (amplification, 2), other
215 score not equal to baseline (1). Scores were normalised relative to the length of the
216 chromosome, summed and then divided by the theoretical maximum (44).
217 Aneuploidy was validated using image cytometry on nuclear extracts from formalin
218 fixed tissue sections (**Supplementary Methods**).

219 **Mutation clustering, phylogenetic reconstruction and timing analyses**

220 The algorithm DPCLust (v2.2.6) and its pre-processing pipeline (v1.0.8) were used to
221 cluster mutations according to fraction of cancer cells (Cancer Cell Fraction, CCF) in
222 which they were found, as described previously²⁰ (**Supplementary Methods**).
223 Phylogenetic reconstruction was performed using the pigeon-hole principle²⁰. In
224 brief, subclones were designated to be nested within a clone or another subclone if
225 their combined CCF exceeded that of their parent.

226 Initial timing analysis required the transformation of individual mutation allele
227 frequencies into mutation copy number. This was performed using the equation:

$$MCN = \frac{VAF(\rho \times TCN + 2(1 - \rho))}{\rho}$$

228 MCN is Mutation Copy Number, ρ is the sample purity, TCN is the local total copy
229 number.

230 For WGD timing, deamination (clock-like, C>T mutations at CpG dinucleotides)
231 mutations were selected from regions of balanced gain (2+2) or LOH (2+0). A
232 probabilistic approach to WGD timing was taken with confidence intervals generated

233 by bootstrapping sampling iterations of the underlying mutations (**Supplementary**
234 **Methods**).

235

236 **Data availability**

237 DKFZ raw methylation array data (IDAT files) were acquired from the authors of
238 Koelsche *et al*, the data having been acquired as described²².

239 The authors declare that all data supporting the findings of this study are available
240 within the article and its supplementary files or from the corresponding author on
241 reasonable request. Sequencing data have been deposited at the European
242 Genome-Phenome Archive (<http://www.ebi.ac.uk/ega/>) that is hosted by the
243 European Bioinformatics Institute.

244 **Acknowledgements**

245 This work was supported by funding from: The Tom Prince Cancer Trust, the UK
246 Medical Research Council grant (MR/M00094X/1), The Wellcome Trust, Skeletal
247 Cancer Action Trust UK, the Royal National Orthopaedic Hospital NHS Trust R&D
248 Department, the Rosetrees and Stoneygate Trusts (M46-F1), the Pathological
249 Society of Great Britain and Ireland (PL) and the Bone Cancer Research Trust.
250 A.M.F is a NIHR senior investigator. AMF and N.P. were supported by the National
251 Institute for Health Research, UCLH Biomedical Research Centre and the UCL
252 Experimental Cancer Centre. M.W.F, J.D., M.T., and P.V.L. are supported by the
253 Francis Crick Institute, which receives its core funding from Cancer Research UK
254 (FC001202), the UK Medical Research Council (FC001202), and the Wellcome Trust
255 (FC001202). Personal fellowships have been granted to S.B. (Wellcome Trust
256 Intermediate Clinical Research Fellowship, St. Baldrick's Foundation Robert J.
257 Arceci International Innovation Award), I.L. (Lundbeck Foundation, award 2018-
258 3018), P.J.C. (Wellcome Trust Senior Clinical Research Fellowship), N.P. (CRUK
259 Clinician Scientist Fellowship (grant number 18387). and E.M. (CRUK Career
260 Development Fellow), M.W.F. (from a CRUK accelerator award C422/A21434), J.D.
261 (Research Foundation – Flanders, FWO Postdoctoral Fellowship; European Union's
262 Horizon 2020 Research and Innovation programme, MSCA 703594-DECODE), M.T.
263 (European Union's Horizon 2020 Research and Innovation programme postdoctoral
264 fellowship, MSCA 747852-SIOMICS). P.V.L. is a Winton Group Leader in recognition
265 of the Winton Charitable Foundation's support towards the establishment of The
266 Francis Crick Institute. We are grateful to the RNOH Musculoskeletal Pathology
267 Biobank team for consenting patients and accessing samples. We thank all the
268 patients for participating in our research and the clinical teams involved in their care.

269 **Author Contributions**

270 AMF conceived the project. M.W.F. performed the data analyses to which P.L., P.E.,
271 I.L. and N.P. contributed some preliminary analyses. J.I. performed whole genome
272 bisulfite analysis. A-C.S. performed DNA extraction. D.O. contributed the image
273 cytometry analysis. A.M.F., P.E., R.T. and F.A. curated and reviewed the samples,
274 clinical data and/or provided clinical expertise. M.S., C.K., I.L. and D.T.W.J.
275 contributed to the methylation analyses. M.T. contributed to the methylation copy
276 number analysis and the timing analyses. J.D. contributed to discussions. A.M.F.,
277 P.V.L., and S.B. directed the research. M.W.F, P.V.L., S.B., and A.M.F. wrote the
278 manuscript.

279 **Competing interests**

280 The authors declare no competing financial or non-financial interests.

281 **Corresponding author**

282 Correspondence to Adrienne M Flanagan and Sam Behjati

283

284 **Figure Legends**

285 **Figure 1. Landscape of H3.3 mutant tumours.** a) Mutational burden of
286 samples in comparison with selected other mesenchymal tumours: osteoblastoma²³,
287 chondroblastoma⁵, chondrosarcoma²⁴ (*exome data only; SVs not shown), and
288 osteosarcoma⁹. b) The genomic and methylation classification of sequenced
289 tumours. From top to bottom: clinical diagnoses and age, unsupervised methylation
290 cluster assignment, *CCND1* promoter methylation status (hypermethylation is
291 defined as a mean *CCND1* promoter methylation beta value >0.2 and a tileplot of
292 curated drivers, clinical outcomes are shown underneath (more detailed clinical
293 outcomes are shown in **Supplementary Table 1**). Note sample PD38328a had
294 undergone deletion of the H3F3A locus, which had been present on the pre-
295 resection biopsy (**Supplementary Figure 11**)

296 **Figure 2. Methylation changes of H3.3 mutant tumours.**

297 a) Hierarchical (unrooted) clustering of tumours. Leaves are coloured by diagnosis
298 and the methylation clusters annotated with shaded ovals. b) Analysis of methylation
299 differences between malignant ("M") and benign ("G") tumours (n=12 and 40
300 respectively) across chromosome 11 (upper) and across *CCND1* (lower). Raw
301 (black) and segmented (green) signal-noise-ratio (SNR; >0 shows greater
302 methylation in malignant tumours) are plotted. Blue ticks in the upper plot represent
303 DMRs. In the lower plot raw methylation beta values across *CCND1* are shown for
304 each sample. The underlying schematic represents the *CCND1* gene body (grey)
305 and the predicted promoter (green). c) Mean *CCND1* promoter across the clustered
306 samples from a) and d) a variety of other tissues.

307 **Supplementary Legends**

308 **Supplementary Table 1. Clinical table of sequenced cases**

309 **Supplementary Table 2. Clinical table for methylation and SNP array cases**

310 **Supplementary Table 3. Clinically indeterminate cases**

311 **Supplementary Figure 1. Telomere lengths and mutation burdens for**
312 **sequenced tumours**

313 **Supplementary Figure 2. *TERT* rearrangement in PD3788d**

314 **Supplementary Figure 3. *KDM4B* homozygous deletion and *RB1* loss of**
315 **heterozygosity**

316 **Supplementary Figure 4. SNP and Methylation array-based copy number**
317 **scores**

318 **Supplementary Figure 5. Mutation timing in malignant samples**

319 **Supplementary Figure 6. Mutation clustering in metastatic samples**

320 **Supplementary Figure 7. Tumour methylation clustering with other malignant**
321 **bone tumours**

322 **Supplementary Figure 8. Selected gene promoters identified as differentially**
323 **methylated regions**

324 **Supplementary Figure 9. Whole genome bisulfite sequencing confirmation of**
325 ***CCND1* differential methylation**

326 **Supplementary Figure 10. Differential methylation in the HIST1 cluster**

327 **Supplementary Figure 11. The loss of H3.3 mutation in sample PD38328a**

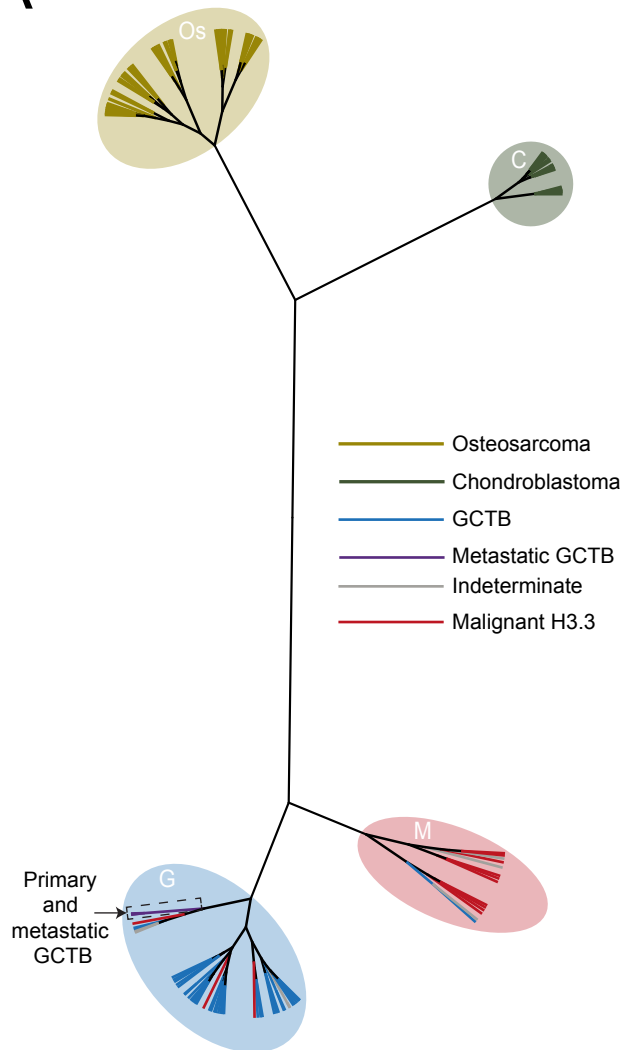
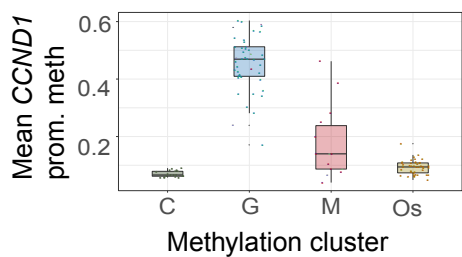
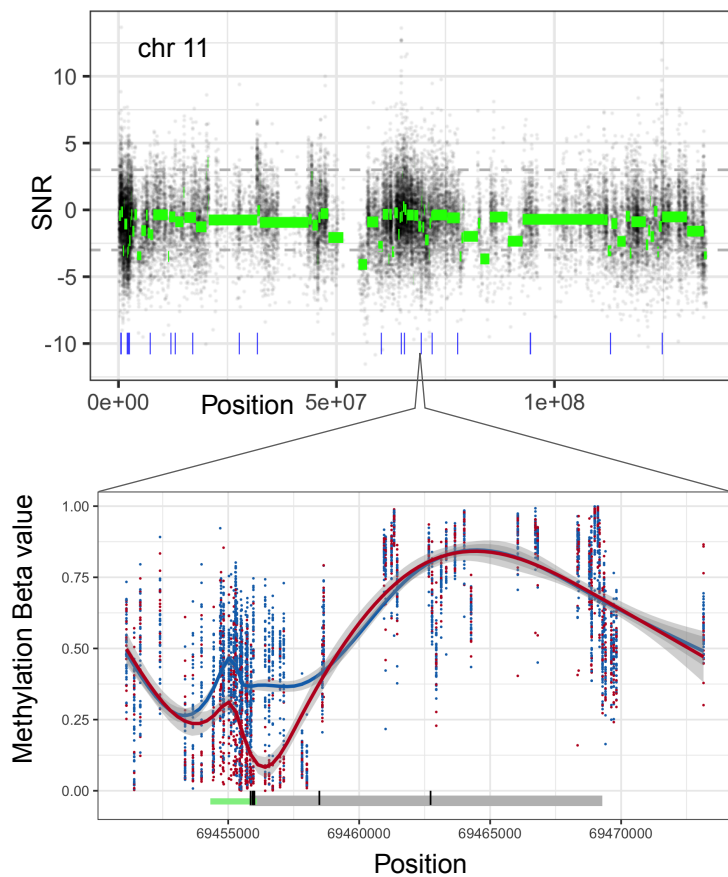
328

329

330 **References**

- 331 1. *WHO Classification of Soft Tissue and Bone Tumours* (IARC Publications, 2020).
- 332 2. Presneau, N. *et al. J Pathol Clin Res* **1**, 113-23 (2015).
- 333 3. Ogura, K. *et al. Genes Chromosomes Cancer* **56**, 711-718 (2017).
- 334 4. Amary, F. *et al. Am J Surg Pathol* **41**, 1059-1068 (2017).
- 335 5. Behjati, S. *et al. Nat Genet* **45**, 1479-82 (2013).
- 336 6. Yamamoto, H. *et al. Hum Pathol* **73**, 41-50 (2018).
- 337 7. Muheremu, A. & Niu, X. *World J Surg Oncol* **12**, 261 (2014).
- 338 8. Bertoni, F., Bacchini, P. & Staals, E.L. *Cancer* **97**, 2520-9 (2003).
- 339 9. Behjati, S. *et al. Nat Commun* **8**, 15936 (2017).
- 340 10. Horn, S. *et al. Science* **339**, 959-61 (2013).
- 341 11. Khan, A. & Zhang, X. *Nucleic Acids Res* **44**, D164-71 (2016).
- 342 12. Barthel, F.P. *et al. Nature Genetics* **49**, 349 (2017).
- 343 13. Bryan, T.M., Englezou, A., Dalla-Pozza, L., Dunham, M.A. & Reddel, R.R. *Nature Medicine* **3**,
344 1271-1274 (1997).
- 345 14. PCAWG. *Nature* **578**, 82-93 (2020).
- 346 15. Morris, T.J. *et al. Bioinformatics* **30**, 428-30 (2014).
- 347 16. Li, H. & Durbin, R. *Bioinformatics* **25**, 1754-60 (2009).
- 348 17. Jones, D. *et al. Curr Protoc Bioinformatics* **56**, 15 10 1-15 10 18 (2016).
- 349 18. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. *Bioinformatics* **25**, 2865-71 (2009).
- 350 19. Dentre, S.C., Wedge, D.C. & Van Loo, P. *Cold Spring Harb Perspect Med* **7**(2017).
- 351 20. Nik-Zainal, S. *et al. Cell* **149**, 994-1007 (2012).
- 352 21. Ding, Z. *et al. Nucleic Acids Research* **42**, e75-e75 (2014).
- 353 22. Koelsche, C. *et al. Clin Sarcoma Res* **7**, 9 (2017).
- 354 23. Fittall, M.W. *et al. Nat Commun* **9**, 2150 (2018).
- 355 24. Tarpey, P.S. *et al. Nat Genet* **45**, 923-6 (2013).

356

A**C****B****D**