Genomic patterns of malignant peripheral nerve sheath tumour (MPNST) evolution correlate with clinical outcome and are detectable in cell-free DNA

Cortes-Ciriano I[1*], Steele CD[2], Piculell K[3], Al-Ibraheemi A[4], Eulo V[5], Bui MM[6], Chatzipli A[7], Dickson BC[8], Borcherding DC[9], Feber A[10], Galor A[11], Hart J[12], Jones KB[13], Jordan JT[14], Kim RH[15], Lindsay D[16], Miller C[1], Nishida Y[17], Proszek P[10], Serrano J[18], Sundby RT[19], Szymanski JJ[20], Ullrich NJ[21], Viskochil D[22], Wang X[23], Genomics of MPNST (GeM) Consortium, Snuderl M[18], Park PJ[7], Flanagan AM[2,16*], Hirbe AC[9*], Pillay N[2,16*], Miller DT[3*]

1 European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, United Kingdom
2 Research Department of Pathology, University College London Cancer Institute, Bloomsbury, London, WC1E 6BT, United Kingdom
3 Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, 02115, United States of America
4 Department of Pathology, Boston Children's Hospital, Boston, Massachusetts, 02115, United States of America
5 Division of Oncology, Department of Internal Medicine, University of Alabama at Birmingham, Birmingham, Alabama, 35294, United States of America
6 Department of Pathology, Moffitt Cancer Center & Research Institute, Tampa, Florida, 33612, United States of America
7 Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, 02115, United States of America
8 Department of Laboratory Medicine and Pathobiology, University of Toronto; Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, M5G 1X5, Canada
9 Division of Oncology, Departments of Internal Medicine and Pediatrics, Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri, 63110, United States of America
10 Clinical Genomics Translational Research, Institute of Cancer Research, Royal Marsden NHS Foundation Trust, London, SM2 5NG, United Kingdom
11 University of Oxford, Wellcome Centre for Human Genetics, Oxford, OX3 7BN, United Kingdom
12 Department of Pathology, Lifespan Laboratories, Rhode Island Hospital, Providence, Rhode Island, 02903, United States of America
13 Departments of Orthopaedics and Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, 84112, United States of America
14 Pappas Center for Neuro-oncology, Massachusetts General Hospital, Boston, Massachusetts, 02114, United States of America
15 Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Sinai Health System, Hospital for Sick Children, University of Toronto, Ontario Institute for Cancer Research, Toronto, Ontario, M5G 1Z5, Canada
16 Department of Histopathology, Royal National Orthopaedic Hospital, NHS Trust, Middlesex, HA7 4LP, United Kingdom
17 Department of Rehabilitation Medicine, Nagoya University Hospital, Nagoya, Aichi, 466-8550, Japan
18 Department of Pathology, New York University Langone Health, Perlmutter Cancer Center, New York City, New York, 10016, United States of America

19 Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, United States of America
20 Division of Cancer Biology, Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri, 63108, United States of America
21 Department of Neurology, Boston Children's Hospital, Boston, Massachusetts, 02115 United States of America
22 Division of Medical Genetics, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, 84108, United States of America
23 GeneHome, Department of Individualized Cancer Management, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, 33612, United States of America

* Contributed equally to leading the consortium

Abstract

Malignant peripheral nerve sheath tumour (MPNST) is an aggressive soft-tissue sarcoma that arises in peripheral nerves. MPNST occurs either sporadically or in people with neurofibromatosis type 1 (NF1), a common cancer predisposition syndrome caused by germline pathogenic variants in *NF1*. Although MPNST is the most common cause of death and morbidity for individuals with NF1, the molecular underpinnings of MPNST pathogenesis remain unclear. Here, we report the analysis of whole-genome sequencing, multi-regional exome sequencing, transcriptomic and methylation profiling data for 95 MPNSTs and precursor lesions (64 NF1-related; 31 sporadic) from 77 individuals. Early events in tumour evolution include biallelic inactivation of *NF1* followed by inactivation of *CDKN2A* and in some cases also *TP53* and polycomb repressive complex 2 (PRC2) genes. Subsequently, both sporadic and NF1-related MPNSTs acquire a high burden of somatic copy number alterations (SCNAs). Our analysis revealed distinct pathways of tumour evolution and immune infiltration associated with inactivation of PRC2 genes and H3K27 trimethylation (H3K27me3) status. Tumours with loss of H3K27me3 evolve through extensive chromosomal losses with retention of chromosome 8 heterozygosity followed by whole genome doubling and chromosome 8 amplification. These tumours show lower levels of immune cell infiltration with low cytotoxic activity and low expression of immune checkpoints. In contrast, tumours with retention of H3K27me3 evolve through extensive genomic instability in the absence of recurrent alterations and exhibit an immune cell-rich phenotype. Specific SCNAs detected in both tumour samples and cell-free DNA (cfDNA) act as a surrogate for loss of H3K27me3 and immune infiltration, and predict prognosis. Our results suggest that SCNA profiling of tumour or cfDNA could serve as a biomarker for early diagnosis and to stratify patients into prognostic and treatment-related subgroups.

Main

Malignant peripheral nerve sheath tumour (MPNST) is an aggressive soft-tissue sarcoma of peripheral nerves that arises both sporadically and in 8-13% of individuals with neurofibromatosis type 1 (NF1)[1,2]. NF1 is a common cancer predisposition disorder affecting 1:2500 individuals worldwide caused by germline heterozygous loss-of-function pathogenic variants in the *NF1* gene[3,4], which encodes for neurofibromin, a Ras GTPase-activating

protein[5,6]. Activation of the Ras pathway is a driver event in many sporadic cancers in the general population, including up to 30% of melanomas and breast cancers, and nearly 25% of acute myeloid leukemias and glioblastomas, among others[7]. Therefore, understanding the development of cancer as a result of *NF1* loss is not only essential for improving prognostication and therapies for MPNST, but can also help to gain a deeper understanding of *NF1*-related cancer biology.

In individuals with NF1, germline inactivation of *NF1* facilitates the development of multiple tumour types, including histologically benign plexiform neurofibromas (PNs). PNs arise through loss of heterozygosity (LOH) for *NF1* in the Schwann cell lineage[8], and may develop anywhere in the peripheral nervous system. A subset of tumours subsequently undergoes genomic transformation with progression to MPNST, which is the major cause of morbidity and mortality for individuals with NF1[9]. Malignant transformation has been attributed to genetic alterations, including *NF1* inactivation followed by early loss of *CDKN2A*[10,11] and, in some cases, PRC2 inactivation as a result of somatic mutations in *SUZ12*, *EED* or *EZH2*[10,12,13]. MPNSTs with PRC2 inactivation undergo loss of trimethylation at lysine 27 of histone H3 (H3K27me3)[14], which has been associated with a worse prognosis[15,16]. However, these events are not found across all MPNSTs, and variability in histological features adds to the challenge of accurately classifying these tumours, thereby limiting clinical decision-making and therapeutic development.

Previous genomic analyses of MPNSTs were based on small sample size studies and limited pathology characterisation[12,17–20]. As a result, the timing and clinical relevance of the genomic alterations underpinning MPNST development remain poorly understood. To overcome these limitations, we established the international Genomics of MPNST (GeM) consortium with the goal of analysing a large collection of MPNSTs with detailed clinical and pathological characterisation at a multi-omic level[21]. Since H3K27me3 has been shown to correlate with clinical outcome, we sought to better understand the molecular events and downstream functional consequences of PRC2 inactivation in MPNST development. In doing so, we anticipated several opportunities: improving diagnostic accuracy, refining clinical prognosis, identifying potentially effective therapeutic intervention, and adding to the understanding of tumorigenesis for cancers related to *NF1* loss. Here, we describe the genomic landscape and key evolutionary events in MPNST development and progression, and provide evidence that this genomic landscape can be used clinically to identify subgroups of patients with varying prognoses.

Results

Tumour sample collection, characterisation and stratification

Through international multi-institutional collaboration, we collected and analysed 95 fresh-frozen tumours (64 NF1-related; 31 sporadic) and matched blood samples from 77 individuals (50 with an NF1 clinical phenotype and 27 sporadic) using whole-genome sequencing (WGS), RNA-sequencing and whole-genome DNA methylation arrays (Fig. 1a and Supplementary Table 1). This collection, which is approximately 10 times larger than any prior WGS or multi-omic studies of MPNSTs[12,17], includes 85 high-grade MPNSTs, 5 low-grade MPNSTs, 3

136 atypical neurofibromatous neoplasms of uncertain biologic potential (ANNUBPs) and 2
137 neurofibromas. Cases were selected based on a pathological diagnosis of MPNST or precursor
138 lesion[9], availability of clinical information, H3K27me3 immunoreactivity data, and availability
139 of material for multi-omic analysis (Methods). Using the WGS data, we identified an *NF1*
140 germline pathogenic variant in all 50 individuals with an NF1 clinical phenotype (8 with
141 pathogenic microdeletions and complex SVs undetectable by exome sequencing and 2 with
142 deep intronic mutations that generated cryptic exons; Supplementary Table 1, Supplementary
143 Figs. 1-3, and Methods). Next, we confirmed biallelic inactivation of *NF1* in 56/64 (88%)
144 tumour samples arising within the tumours of individuals with NF1. *NF1* inactivation in 49 of
145 the 56 (87.5%) tumours occurred through the loss of the wild-type allele, and biallelic
146 inactivation was identifiable more often in tumours from people with NF1 as compared to
147 sporadic tumours, 10/31 (32%, $P<0.01$, Fisher's exact test; Supplementary Table 1). Four
148 pathologists performed a consensus classification by assigning the tumours to 4 categories
149 based on the presence or absence of germline *NF1* pathogenic variants (germline vs sporadic)
150 and conventional and non-conventional histopathological features of MPNST (Supplementary
151 Fig. 4). In addition, tumours were classified according to the retention or loss of H3K27me3
152 using immunohistochemistry[9] (H3K27me3 retained vs H3K27me3 loss; Methods). Diagnosis
153 using previously published machine learning classifiers of sarcomas based on genome-wide
154 DNA methylation patterns[22] was only informative for 48/95 tumours (51%), highlighting the
155 need for combining pathology and genomics data for accurate diagnosis (Supplementary Fig.
156 5).
157
158 **Somatic genomic landscape of MPNSTs**
159 To identify drivers of MPNST development, we searched for genes harbouring recurrent loss
160 or gain-of-function mutations (Methods). We detected somatic biallelic inactivation of
161 *CDKN2A* in 40/64 (63%) NF1-related MPNSTs and in 17/31 (55%) sporadic cases (Fig. 1b-
162 c), primarily caused by non-recurrent SVs and complex rearrangements of the *CDKN2A* locus,
163 which in some cases led to the concomitant inactivation of *NF1* and the loss of nearby genes
164 (e.g., *MTAP* in 25% of tumours; Supplementary Fig. 6). We identified loss of H3K27me3 in
165 45/82 (55%) of MPNSTs with immunohistochemistry data available (Fig. 1b-c). *SUZ12* was
166 the gene from the PRC2 complex with the highest frequency of biallelic inactivation (27/95
167 tumours with WGS data, 28%), followed by *EED* (16/95, 17%). Biallelic inactivation of *SUZ12*
168 and *EED* were mutually exclusive events (Fig. 1c). Furthermore, the loss of H3K27me3 was
169 not associated with hypermethylation of these genes (Supplementary Fig. 7). Double-hit
170 mutations were found in *TP53* in 20/95 (21%) tumours, and replicative immortality was
171 primarily driven by *TERT* overexpression (Supplementary Fig. 8). We did not detect recurrent
172 gene fusions or genes recurrently mutated besides *NF1*, *CDKN2A*, PRC2 complex genes and
173 *TP53* (Fig. 1c and Methods), suggesting that these are the main drivers of MPNST
174 tumorigenesis.
175
176 We performed *de novo* mutational signature analysis of single-base substitutions (SBS) and
177 small indels. Similar mutational processes were identified in MPNSTs irrespective of
178 H3K27me3 status (Extended Data Fig. 1a). Mutational signature SBS5, a clock-like signature
179 found in human cancers[23] and histologically normal tissues[24], was the predominant signature

in both groups, contributing a median of 58% of SNVs. The burden of mutations and SVs was comparable between both groups, with the exception of four hypermutated cases, two of which showed microsatellite instability (Fig. 1c) (Extended Data Fig. 1b-e). SV signature analysis identified five distinct signatures previously detected in undifferentiated soft tissue sarcomas and breast carcinomas[25,26] (Extended Data Fig. 1f-g), which had comparable activity in both groups (Extended Data Fig. 1a). The frequency of chromothripsis, which was detected in 47/95 (49%) of tumours, was also comparable between both groups (19/37 and 22/45 of tumours with retention and loss of H3K27me3, respectively; $P > 0.05$, Fisher's exact test). Whole-genome doubling (WGD) events, which were validated using image cytometry analysis (Methods), were detected in 22/37 (59%) of tumours with H3K27me3 retained and in 33/45 (73%) of tumours with H3K27me3 loss ($P > 0.05$, Fisher's exact test). Together, these results indicate that MPNSTs are characterised by a high level of genomic instability irrespective of H3K27me3 status.

Molecular subgroups correlate with clinical outcome and immune infiltration

Next, we investigated transcriptomic differences between MPNSTs with different H3K27me3 status. Unbiased clustering based on transcriptomic and genome-wide methylation data confirmed that tumours stratified into two distinct groups, largely corresponding to the retention and loss of H3K27me3 (Methods and Extended Data Fig. 2a). Cluster assignments were consistent across data types (Extended Data Fig. 2a). Classification by H3K27me3 immunoreactivity was associated with overall survival in NF1 individuals, but not in sporadic MPNSTs (Extended Data Fig. 2b-c). No differences in overall survival were observed when stratifying tumours on the basis of conventional or unconventional histology (Methods) by univariate or multivariate analysis accounting for age, sex and tumour grade (Cox proportional-hazards model; $P>0.05$; log-rank test).

Differential expression analysis between both groups revealed a strong downregulation of genes related to adaptive immunity among a subcategory of MPNSTs (Fig. 2a). Specifically, H3K27me3 loss was strongly associated with high tumour cellularity estimates, decreased immune cell infiltration and adaptive immune response activation, and decreased expression of granzymes (*GZMA*, *GZMK* and *GZMH*) and immune checkpoints (*PD-L1* and *HAVCR2;* Fig. 2a-b, Extended Data Figs. 2d-e and 3). Both differential analysis of genome-wide DNA methylation patterns and gene expression revealed a significant activation of genes involved in neural development and morphogenesis pathways in tumours with H3K27me3 loss, which is consistent with the role of PRC2 in neural development[12] (Extended Data Fig. 2e,g). Together, these results indicate that MPNSTs arising among individuals with NF1 stratify into two clinically and biologically distinct groups characterised by differential levels of immune cell infiltration.

MPNST evolution is defined by recurrent patterns of copy number alterations

Next, we compared the patterns of MPNST copy number profiles to H3K27me3 status. We found focal deletions of the *CDKN2A* and *NF1* loci and a high burden of SCNAs irrespective of H3K27me3 status. However, both sporadic and NF1-related MPNSTs with loss of H3K27me3 showed a strong enrichment for chromosome 8 amplification and variable levels

224 of genome-wide LOH (Fig. 2c-e, Extended Data Fig. 4, Supplementary Fig. 9 and Methods).
225 LOH of selected chromosomes, including 1p, 10, 11, 16, 17 and 22, was frequently detected in
226 MPNSTs with loss of H3K27me3 (Fig 2e). In one extreme H3K27me3 loss case (Fig. 2f), we
227 observed near-haploidization characterised by genome-wide LOH excluding a few
228 chromosomes with retention of heterozygosity, including chromosome 8, and a ploidy of 1.3.
229 In most other cases, near-haploidization was followed by WGD (Fig. 2g and Supplementary
230 Fig. 9). Copy number signature analysis[27] identified 18 distinct signatures (Extended Data Fig.
231 5a) and confirmed an enrichment of signatures related to LOH in MPNSTs with loss of
232 H3K27me3 (Fig. 2e and Extended Data Fig. 5b). In contrast, tumours with retained H3K27me3
233 were enriched in signatures associated with a diploid genome and genomic instability
234 (P=0.029, two-sided Mann-Whitney test; Extended Data Fig. 5b). Together, these results
235 indicate that distinct copy number patterns are associated with H3K27me3 status.
236
237 Our analysis of MPNST copy number profiles offers a basis for comparison to other types of
238 cancer. To this aim, we next compared MPNSTs to diverse cancer types from TCGA at the
239 copy number level. Gains in chromosome 8q have also been detected in a number of other solid
240 tumours[28–30]. Numerous cancer-related genes are located on chromosome 8q and have been
241 implicated in cancer progression, including genes such as *C-MYC, RAD21, UBR5*, and
242 *HEY1*[28,31–33]. Of the >700 protein-coding genes on chromosome 8, 199 genes are differentially
243 expressed in the tumours with H3K27me3 loss, including *RAD21,* which has been linked with
244 the mitigation of replication stress caused by the oncogenic *EWS-FLI1* fusion in Ewing
245 sarcomas with chromosome 8q amplification[28]*, UBR5*, and *HEY1* (Supplementary Fig. 10,
246 Supplementary Table 2 and Methods). Near-haploidization has been observed across diverse
247 cancer types, including undifferentiated soft tissue sarcomas, adrenocortical carcinoma,
248 gliomas, chondrosarcomas, and acute lymphoblastic leukemia, and is often linked with a poorer
249 prognosis[25,34–38]. In pan-cancer analysis, MPNSTs with loss of H3K27me3 show the highest
250 levels of genome-wide LOH, including near-haploidization (Extended Data Fig. 6). However,
251 concomitant retention of chromosome 8 heterozygosity and genome-wide LOH as observed in
252 MPNSTs with loss of H3K27me3 (Fig. 2a) has not been linked with a distinct biological
253 process previously. This new copy number configuration provides a potential new avenue to
254 investigate the evolutionary constraints leading to genome-wide LOH[39].
255
256 **Copy number profiles are detected in cell-free DNA and predict patient outcome**
257 Given the strong association between distinct copy number profiles with H3K27me3 status and
258 immune infiltration, we evaluated the clinical relevance of SCNA patterns to predict overall
259 survival in comparison to the predictive ability based on clinical data and H3K27me3 status
260 (Methods). For this analysis, we focused on high-grade MPNSTs arising in NF1 individuals to
261 increase statistical power. Among the set of chromosome arms showing recurrent copy gain,
262 copy loss, or loss of heterozygosity (LOH), LOH of chromosome 5q was the most predictive
263 feature of poor prognosis (Fig. 3a-b). Other copy number aberrations predictive of poor
264 prognosis are LOH of chromosomes 11q, 7p and 22q, and amplification of chromosome 2q
265 and 9q (P<0.05; log-rank test). There appeared to be a trend toward worse prognosis among
266 tumours showing amplification of chromosome 8q, although not reaching statistical
267 significance (Fig. 3c). Next, we investigated whether copy number patterns associated with

268 H3K27me3 loss and predictive of survival can be detected in cell-free DNA (cfDNA) collected
269 from a separate cohort of patients with NF1-related MPNST at different time points and
270 analysed by ultra-low-pass whole-genome sequencing[40] (Methods). Overall, we detected in
271 cfDNA copy number aberrations recurrently found in MPNSTs using WGS data (Fig. 2e),
272 including loss of chromosomes 1p, 5q, 4q, and 22, and chromosome 8 amplification (Fig. 3d,
273 Supplementary Figs 11-13). These results thus suggest that copy number profiles detected in
274 cfDNA could serve as a non-invasive surrogate for H3K27me3 status and prognostication,
275 which warrants further investigation in larger cohorts.
276
277 **Timing of mutations in MPNST evolution**
278 We next sought to determine the relative timing of the genomic alterations that occur during
279 MPNST development across time and space. To this aim, we first investigated which
280 alterations are present in low-grade MPNSTs and MPNST precursor lesions (either benign
281 neurofibromas or transitional lesions referred to as atypical neurofibromatous neoplasms of
282 uncertain biologic potential; ANNUBPs). We detected biallelic inactivation of *NF1* and
283 *CDKN2A* caused by non-recurrent complex SVs in 2/2 neurofibromas, 3/3 ANNUBPs, and 4/5
284 low-grade MPNSTs (Extended Data Fig. 7). *EED* was mutated by SVs in one neurofibroma
285 and one low-grade MPNST (Extended Data Fig. 7). Analysis of two anatomically distant
286 tumours from the same individual, a neurofibroma and a high-grade MPNST, revealed
287 *CDKN2A* inactivation in both tumours by independent SVs, suggesting parallel evolution and
288 that *CDKN2A* loss is the second genomic alteration after *NF1* inactivation in MPNST
289 pathogenesis (Fig. 4a-c). Together, these results indicate that somatic *CDKN2A* inactivation is
290 an early event in MPNST evolution, which is necessary but not sufficient for the transformation
291 of neurofibromas into MPNSTs.
292
293 Given the high rate of WGD in MPNSTs, we next investigated when WGD occurs during
294 MPNST development. Timing analysis of SCNAs revealed that WGD is consistently a late
295 clonal event in tumours with loss of H3K27me3, and is common in the background of a near-
296 haploid genome (Fig. 4d). In tumours with H3K27me3 loss, WGD is followed by additional
297 gains of chromosome 8q, likely as isochromosome 8q (Methods)[41]. In contrast, in tumours with
298 retention of H3K27me3, WGD occurs earlier in tumour development over a wider period of
299 time (two-sided Mann-Whitney U test, $P<0.0001$; Fig. 4d; and see Supplementary Fig. 14 for
300 an example). These results suggest that WGD events occur at different stages during tumour
301 evolution depending on H3K27me3 status.
302
303 To elucidate the patterns of intra-tumour heterogeneity, we performed multi-regional high-
304 depth exome sequencing of five regions from 21 and 15 MPNSTs with loss and retention of
305 H3K27me3, respectively (Fig. 4e-i). Chromosome 8 amplification and WGD were found
306 across multiple spatially distant regions in tumours with loss of H3K27me3 (Extended Data
307 Fig. 8). The late occurrence of WGD events in tumours with H3K27Me3 loss suggests that
308 chromosomal losses, including near-haploidization, may be associated with low fitness that is
309 rescued by the WGD event, which triggers clonal expansion and malignant transformation.
310 Rapid tumour enlargement fosters intra-tumour heterogeneity of mutations and clonal
311 competition, as evidenced by the accumulation of subclonal mutations and variable levels of

histopathological heterogeneity in spatially distant regions (Fig. 4h-i, Extended Data Fig. 8 and Supplementary Fig. 15), and might explain the aggressive clinical features of tumours with loss of H3K27me3.

Collectively, our integrative analysis of multi-omic data revealed recurrent patterns of genomic alterations, which allow us to propose clinically relevant mechanistic models of MPNST evolution (Fig. 5). The first steps of MPNST pathogenesis irrespective of H3K27me3 status involve the biallelic inactivation of *CDKN2A,* and in some *TP53* as well*,* on an *NF1*-deficient genomic background. In MPNSTs with H3K27me3 loss, biallelic inactivation of the PRC2 complex is followed by extensive chromosomal losses, leading to variable levels of genome-wide LOH with retention of chromosome 8 heterozygosity. Late events in tumour evolution include WGD and further amplifications of chromosome 8q, which might be accompanied by additional chromosomal instability (Fig. 5 and Extended Data. Fig. 9). In contrast, tumours with H3K27me3 retention evolve through extensive chromosome instability and chromothripsis, and display more heterogeneous karyotypes (Fig. 5 and Extended Data Fig. 10).

Discussion

This integrative multi-omic analysis of NF1 and sporadic MPNSTs has expanded the understanding of MPNST development. The international collaboration permitted collection and comprehensive pathologic characterization of a large number of these rare tumours with paired normal samples, thus providing a resource of a large number of these rare tumours with paired normal samples, along with comprehensive pathologic characterization. Multi-omic analysis, coupled with multi-regional deep exome sequencing, revealed distinct genomic evolutionary pathways underlying MPNST pathogenesis that permit subclassification of MPNSTs in a way that correlates with prognosis, and ultimately may lead to individualised treatment approaches. While the focus has been to improve understanding of MPNST development, either in the setting of NF1 or as a sporadic tumour, this expanded MPNST genomic landscape also contributes to the general knowledgebase of other common cancers that share genomic features with MPNST, including loss of *NF1*.

First, we have delineated the order of genomic events leading to this aggressive form of cancer, which has revealed windows of opportunity for intervention through correlation of molecular signatures with tumour status. Our WGS analysis revealed that inactivation of *CDKN2A* is caused by non-recurrent SVs prior to transition from benign to malignant tumours, suggesting that these are random events that increase the fitness of mutant cells. Other loci may contribute to tumour development in a subset of tumours, which may also represent therapeutic targets that could be exploited at an early stage of tumour development. For example, 25% of MPNST cases showed loss of *MTAP*, a gene adjacent to *CDKN2A*. Loss of *MTAP* leads to enhanced dependence on the arginine methyltransferase *PRMT5*[42], for which there now exist drugs that target this enzyme[43].

Second, analysis of WGS data revealed that inactivation of the PRC2 complex, which follows inactivation of *CDKN2A*, is mediated by non-recurrent SVs. PRC2 loss is not necessary for

MPNST development, as it occurs in only 45% of cases; however, identification of PRC2 loss through genomic analysis of the tumour can provide clinically meaningful information. PRC2 loss correlates with H3K27me3 loss, which in turn correlates with poor prognosis among people with MPNST in the setting of NF1. Thus, genomic analysis of tumour DNA facilitates tumour subclassification into categories predictive of clinical behaviour. Transcriptomic data revealed that tumours with retention of H3K27me3 also show upregulation for markers of immune infiltration and activation of the adaptive immune system, suggesting that this subset of tumours may be more responsive to immunotherapy. Our results suggest that further research into the possibility of incorporating immunotherapy into clinical trials for a subset of MPNST patients is warranted.

Third, we identified a previously unrecognised complex pattern of divergent tumour evolution on the basis of PRC2 loss. In tumours with H3K27me3 loss, biallelic inactivation of the PRC2 complex is followed by chromosomal losses, ranging from loss of several chromosomes to near-haploidization, leading to variable levels of genome-wide LOH with retention of chromosome 8 heterozygosity. Late clonal events in tumour evolution include WGD and further amplifications of chromosome 8q. While chromosome 8 amplification has been found in diverse cancer types, such as Ewing sarcoma, concomitant amplification of chromosome 8 and extensive LOH represents a new copy number configuration specific to MPNSTs with loss of H3K27me3. Tumours with H3K27me3 retention instead evolve through extensive chromosome instability and chromothripsis, and display more heterogeneous karyotypes. Further understanding the transcriptional changes that occur with these copy number aberrations, and comparison of these patterns to other forms of cancer demonstrating similar copy number profiles, may provide insights into novel therapeutic vulnerabilities.

In summary, our results provide the foundation for an MPNST clinical care model that takes into account the genomic architecture of the tumour to provide a more accurate diagnosis, prognosis, and treatment approach. At the level of diagnosis, our comparison of pathologic classification to genomic data clearly demonstrates an inability to identify subsets of tumours based solely on pathology characterization. Our analysis revealed specific genomic patterns of MPNST evolution which enabled classification of tumour subtypes and are more predictive of prognosis than MPNST classification based on clinical or pathologic data alone. Along with more recent data showing the detection of cell-free tumour DNA in plasma from patients with MPNST[40], the results provide new opportunities for tumour surveillance among patients known to be at increased risk for MPNST development, namely patients with NF1, and in particular those with NF1 caused by a microdeletion[44]. Furthermore, genomic analysis in the clinical setting could benefit all patients with an MPNST through more accurate diagnosis and categorization of these tumours prior to the initiation of either surgical or medical therapy, facilitating a more accurate prognosis and offering an opportunity to assign patients to more specific treatment protocols.

References

1. Evans, D. G. R. *et al.* Malignant peripheral nerve sheath tumours in neurofibromatosis 1.

J. Med. Genet. **39**, 311–314 (2002).
2. Kolberg, M. *et al.* Survival meta-analyses for >1800 malignant peripheral nerve sheath tumor patients with and without neurofibromatosis type 1. *Neuro. Oncol.* **15**, 135–147 (2013).
3. Kallionpää, R. A. *et al.* Prevalence of neurofibromatosis type 1 in the Finnish population. *Genet. Med.* **20**, (2018).
4. Miller, D. T. *et al.* Health Supervision for Children With Neurofibromatosis Type 1. *Pediatrics* **143**, (2019).
5. Gutmann, D. H. *et al.* Neurofibromatosis type 1. *Nature Reviews Disease Primers* **3**, 1–17 (2017).
6. Ratner, N. & Miller, S. J. A RASopathy gene commonly mutated in cancer: the neurofibromatosis type 1 tumour suppressor. *Nat. Rev. Cancer* **15**, 290–301 (2015).
7. Philpott, C., Tovell, H., Frayling, I. M., Cooper, D. N. & Upadhyaya, M. The NF1 somatic mutational landscape in sporadic human cancers. *Hum. Genomics* **11**, 13 (2017).
8. Pemov, A. *et al.* The primacy of NF1 loss as the driver of tumorigenesis in neurofibromatosis type 1-associated plexiform neurofibromas. *Oncogene* **36**, 3168–3177 (2017).
9. WHO Classification of Tumours Editorial Board. *Soft Tissue and Bone Tumours*. (International Agency for Research on Cancer, 2020).
10. Nielsen, G. P. *et al.* Malignant transformation of neurofibromas in neurofibromatosis 1 is associated with CDKN2A/p16 inactivation. *Am. J. Pathol.* **155**, 1879–1884 (1999).
11. Magallón-Lorenz, M. *et al.* Chromosomal translocations inactivating CDKN2A support a single path for malignant peripheral nerve sheath tumor initiation. *Hum. Genet.* **140**, 1241–1252 (2021).
12. Lee, W. *et al.* PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. *Nature Genetics* vol. 46 1227–1232 (2014).
13. Brohl, A. S., Kahen, E., Yoder, S. J., Teer, J. K. & Reed, D. R. The genomic landscape of malignant peripheral nerve sheath tumors: diverse drivers of Ras pathway activation. *Sci. Rep.* **7**, 1–5 (2017).
14. Lyskjær, I. *et al.* H3K27me3 expression and methylation status in histological variants of malignant peripheral nerve sheath tumours. *J. Pathol.* **252**, 151–164 (2020).
15. Cleven, A. H. G. *et al.* Loss of H3K27 tri-methylation is a diagnostic marker for malignant peripheral nerve sheath tumors and an indicator for an inferior survival. *Mod. Pathol.* **29**, 582–590 (2016).
16. Ito, Y. *et al.* Clinicopathological and prognostic significance of H3K27 methylation status in malignant peripheral nerve sheath tumor: correlation with skeletal muscle differentiation. *Virchows Arch.* **479**, 1233–1244 (2021).
17. Zhang, M. *et al.* Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors. *Nat. Genet.* **46**, 1170–1172 (2014).
18. Cancer Genome Atlas Research Network. Electronic address: elizabeth.demicco@sinaihealthsystem.ca & Cancer Genome Atlas Research Network. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950–965.e28 (2017).
19. Lemberg, K. M., Wang, J. & Pratilas, C. A. From Genes to -Omics: The Evolving Molecular Landscape of Malignant Peripheral Nerve Sheath Tumor. *Genes* **11**, (2020).
20. Pemov, A., Li, H., Presley, W., Wallace, M. R. & Miller, D. T. Genetics of human malignant peripheral nerve sheath tumors. *Neurooncol Adv* **2**, i50–i61 (2020).
21. Miller, D. T. *et al.* Genomics of MPNST (GeM) Consortium: Rationale and Study Design for Multi-Omic Characterization of NF1-Associated and Sporadic MPNSTs. *Genes* **11**, (2020).

22. Koelsche, C. *et al.* Sarcoma classification by DNA methylation profiling. *Nat. Commun.* **12**, 498 (2021).
23. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
24. Moore, L. *et al.* The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
25. Steele, C. D. *et al.* Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell* **35**, 441–456.e8 (2019).
26. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
27. Steele, C. D. *et al.* Signatures of copy number alterations in human cancer. *bioRxiv* 2021.04.30.441940 (2021) doi:10.1101/2021.04.30.441940.
28. Su, X. A. *et al.* RAD21 is a driver of chromosome 8 gain in Ewing sarcoma to mitigate replication stress. *Genes Dev.* (2021) doi:10.1101/gad.345454.120.
29. Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).
30. Dehner, C. *et al.* Chromosome 8 gain is associated with high-grade transformation in MPNST. *JCI Insight* **6**, (2021).
31. Liao, L. *et al.* E3 Ubiquitin Ligase UBR5 Drives the Growth and Metastasis of Triple-Negative Breast Cancer. *Cancer Res.* **77**, 2090–2101 (2017).
32. Yang, Y.-F. *et al.* IMPAD1 functions as mitochondrial electron transport inhibitor that prevents ROS production and promotes lung cancer metastasis through the AMPK-Notch1-HEY1 pathway. *Cancer Lett.* **485**, 27–37 (2020).
33. Maddipati, R. *et al.* MYC levels regulate metastatic heterogeneity in pancreatic adenocarcinoma. *Cancer Discov.* (2021) doi:10.1158/2159-8290.CD-20-1826.
34. Gibbons, B. *et al.* Near haploid acute lymphoblastic leukemia: seven new cases and a review of the literature. *Leukemia* **5**, 738–743 (1991).
35. Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma. *Cancer Cell* **29**, 723–736 (2016).
36. Bigner, S. H., Mark, J., Schold, S. C., Jr, Eng, L. F. & Bigner, D. D. A serially transplantable human giant cell glioblastoma that maintains a near-haploid stem line. *Cancer Genet. Cytogenet.* **18**, 141–153 (1985).
37. Arbajian, E., Köster, J., Vult von Steyern, F. & Mertens, F. Inflammatory leiomyosarcoma is a distinct tumor characterized by near-haploidization, few somatic mutations, and a primitive myogenic gene expression signature. *Mod. Pathol.* **31**, 93–100 (2018).
38. Bovée, J. V. *et al.* Near-haploidy and subsequent polyploidization characterize the progression of peripheral chondrosarcoma. *Am. J. Pathol.* **157**, 1587–1595 (2000).
39. Nichols, C. A. *et al.* Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nat. Commun.* **11**, 1–14 (2020).
40. Szymanski, J. J. *et al.* Cell-free DNA ultra-low-pass whole genome sequencing to distinguish malignant peripheral nerve sheath tumor (MPNST) from its benign precursor lesion: A cross-sectional study. *PLoS Med.* **18**, e1003734 (2021).
41. Shen, H. *et al.* Integrated Molecular Characterization of Testicular Germ Cell Tumors. *Cell Rep.* **23**, 3392–3406 (2018).
42. Kryukov, G. V. *et al.* MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science* **351**, 1214–1218 (2016).
43. Rodon Ahnert, J. *et al.* PF-06939999, a potent and selective PRMT5 inhibitor, in patients with advanced or metastatic solid tumors: A phase 1 dose escalation study. *J. Clin. Oncol.* **39**, 3019–3019 (2021).

44. De Raedt, T. *et al.* Elevated risk for MPNST in NF1 microdeletion patients. *Am. J. Hum. Genet.* **72**, 1288–1292 (2003).

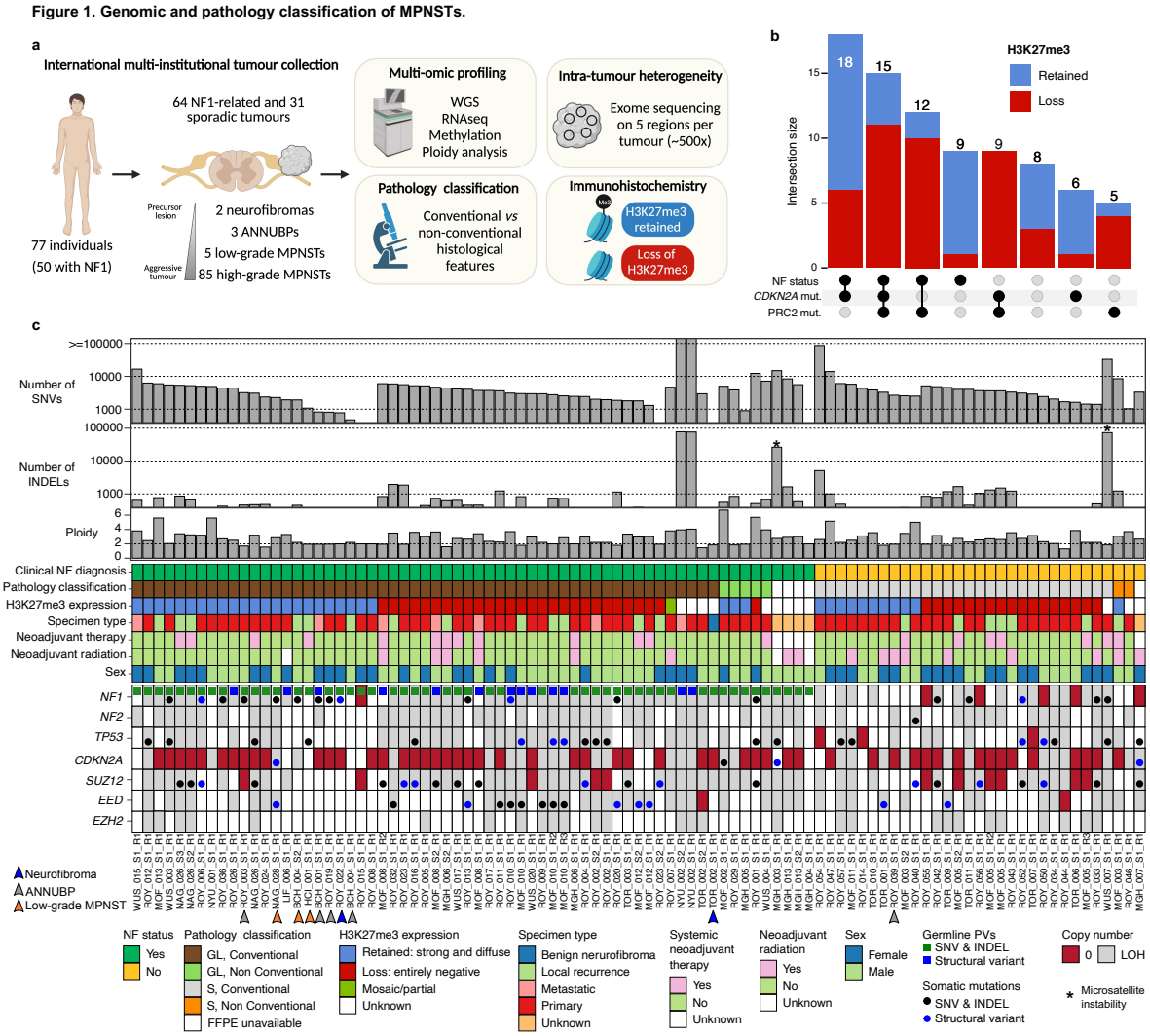


Figure 1. Genomic and pathology classification of MPNSTs. a, Overview of the experimental design and technologies utilized in this study. **b**, Relationship between NF status, biallelic inactivation of *CDKN2A* and the PRC2 complex. The numbers on top of the bars indicate the total number of tumours in each group. **c**, Clinical information, histopathological features and mutations in driver genes in the GeM cohort. Precursor lesions are marked by arrowheads. Sample labels not marked by an arrowhead correspond to high-grade MPNSTs. ANNUBP, atypical neurofibromatous neoplasms of uncertain biologic potential; CN, copy number; GL, germline; S, sporadic.

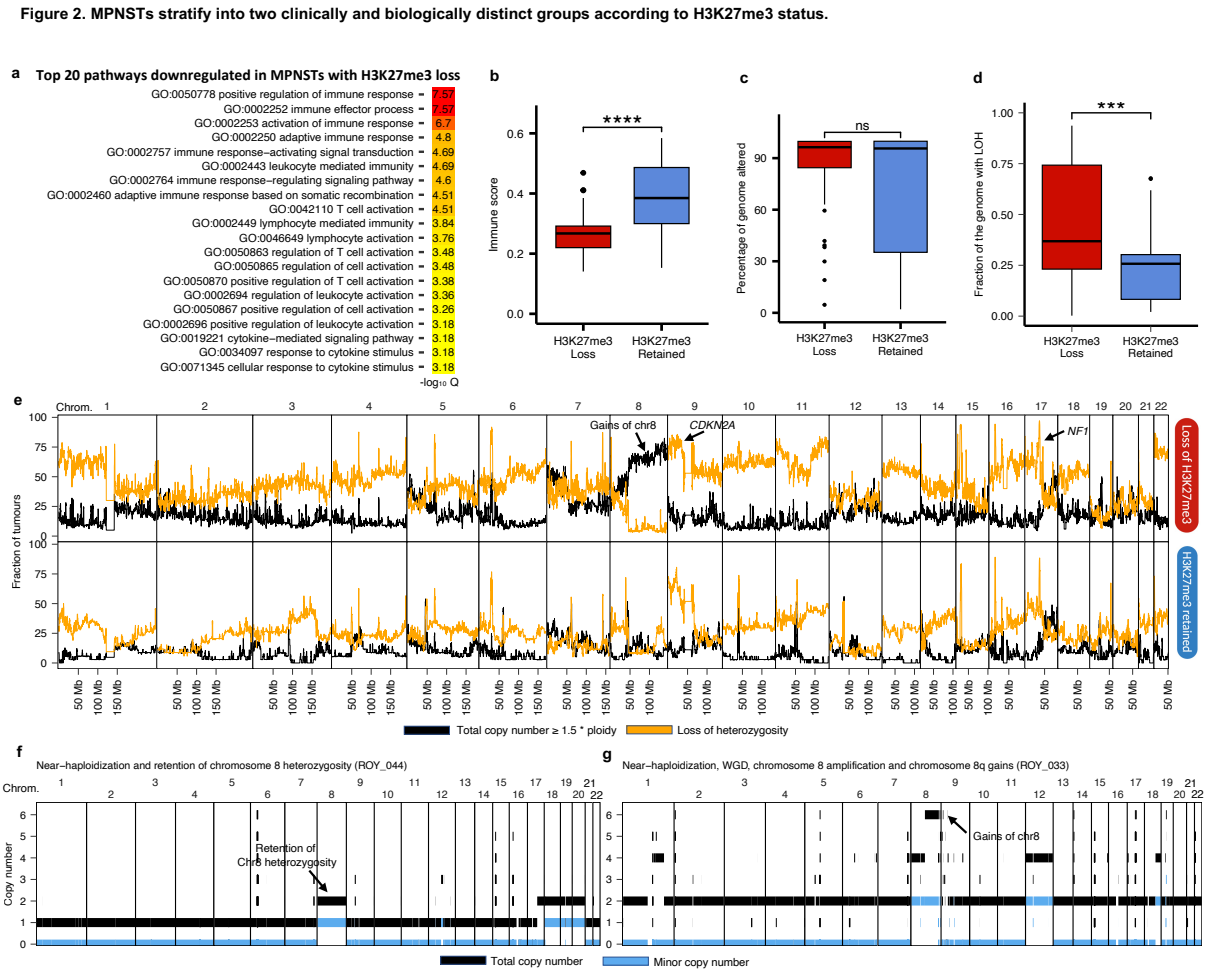


Figure 2. MPNSTs stratify into two clinically and biologically distinct groups according to H3K27me3 status. a, Pathways related to immune activation are downregulated in tumours with loss of H3K27me3 estimated. **b**, Immune score analysis reveals decreased immune infiltration according to H3K27me3 status. **c**, Tumours with retention and loss of H3K27me3 are characterised by high levels of somatic copy number alterations. **d**, Tumours with loss of H3K27me3 show significantly higher levels of genome-wide LOH as compared to tumours with retention of H3K27me3. **e**, Genome-wide analysis of the fraction of tumours showing amplification (black) and LOH (orange) based on H3K27me3 status. Representative copy number profiles for tumours with loss of H3K27me3 showing near-haploidization (**f**) followed by whole-genome doubling and chromosome 8q amplification (**g**). Box plots in **b-d** show median, first and third quartiles (boxes), and the whiskers encompass observations within a distance of 1.5× the interquartile range from the first and third quartiles. Amp., amplification; CN, copy number signature; chrom., chromosome; LOH, loss of heterozygosity; WGD, whole genome doubling.

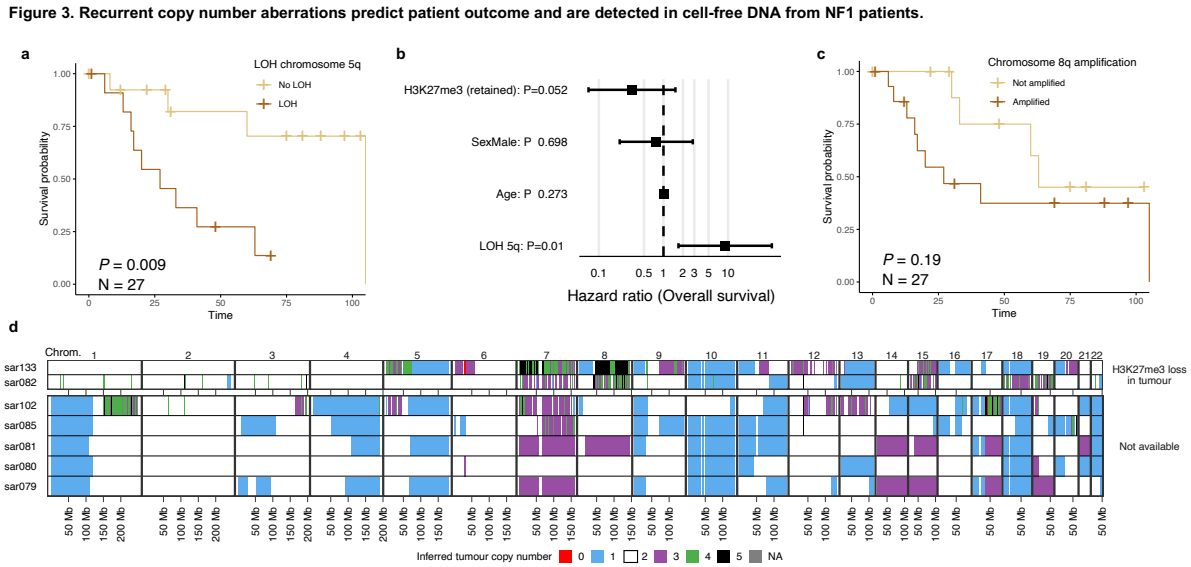
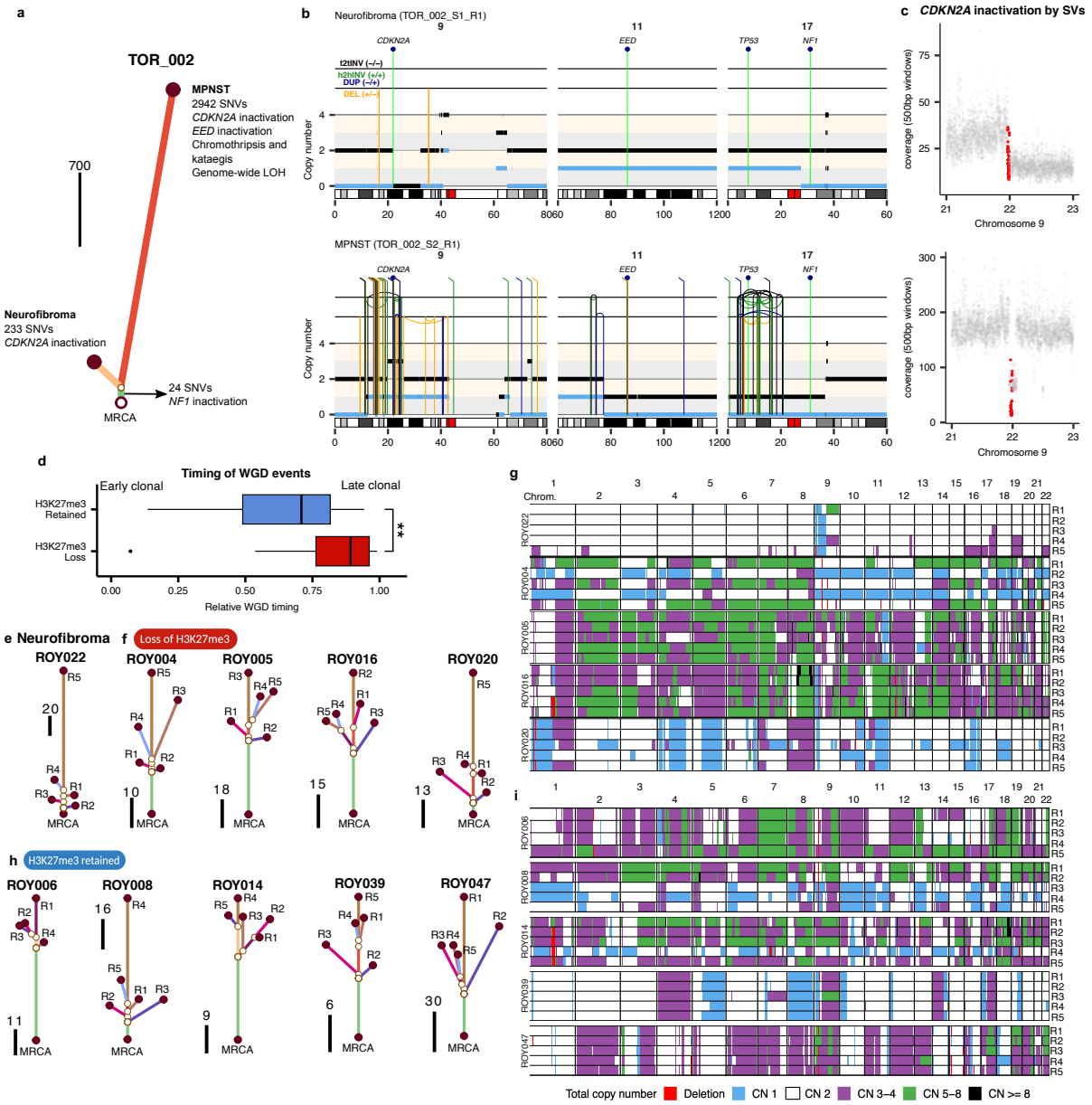


Figure 3. Recurrent copy number aberrations predict patient outcome and are detected in cell-free DNA from NF1 patients. a, Kaplan Meier plot showing the overall survival for NF1 individuals with high-grade MPNSTs stratified according to chromosome 5q LOH. **b**, Hazard ratios with 95% confidence intervals and *P* values computed using Cox proportional hazards regression. **c**, Kaplan Meier plot showing the overall survival for NF1 individuals with high-grade MPNSTs based on the amplification of chromosome 8q. **d**, Copy number profiles estimated using ultra-low-pass whole-genome sequencing of cell-free DNA from patients with MPNST.

548

Figure 4. MPNST evolution and intra-tumour heterogeneity.



549

550 **Figure 4. MPNST evolution and intra-tumour heterogeneity.** Phylogenetic tree inferred
551 from SNVs detected using whole-genome sequencing data for a neurofibroma and a high-grade
552 MPNST sample from individual TOR_002. Scale bar, 700 SNVs. **b**, Rearrangement profiles
553 of a neurofibroma (top) and a high-grade MPNST (bottom) from patient TOR_002. The
554 genomic position of *CDKN2A, EED, TP53* and *NF1* is shown at the top. Both tumours show
555 biallelic inactivation of *CDKNA* and *NF1* mediated by SVs. Total and minor copy number
556 values are shown in black and blue, respectively. Somatic events common to both tumours
557 (*NF1* deletion and 24 SNVs) and private to each of them are shown. **c**, Sequencing coverage
558 calculated at 500 bp windows at the *CDKN2A* locus for the neurofibroma (top) and the high-
559 grade MPNST (bottom). Dots corresponding to windows mapping to *CDKN2A* are shown in
560 red. **d,** Difference in the timing of whole-genome doubling (WGD) events relative to the clonal

expansion between MPNSTs with loss and retention of H3K27me3 (***$P < 0.001$; two-sided Wilcoxon rank sum test). Phylogenetic trees inferred from SNVs detected using multi-regional sequencing data from fresh-frozen samples for a neurofibroma (**e**), MPNSTs with loss (**f**) and retention (**h**) of H3K27me3. Each region (R) is arbitrarily named with a number. The scale bars indicate the number of SNVs. **g,** Genome-wide total copy number profiles estimated using whole-exome sequencing data for a neurofibroma and MPNSTs with loss of H3K27me3. **i**, Genome-wide total copy number profiles estimated using whole-exome sequencing data for MPNSTs with retention of H3K27me3. MRCA, most recent common ancestor. Box plots in **d** show median, first and third quartiles (boxes), and the whiskers encompass observations within a distance of 1.5× the interquartile range from the first and third quartiles.

Figure 5. Pathways of MPNST evolution.

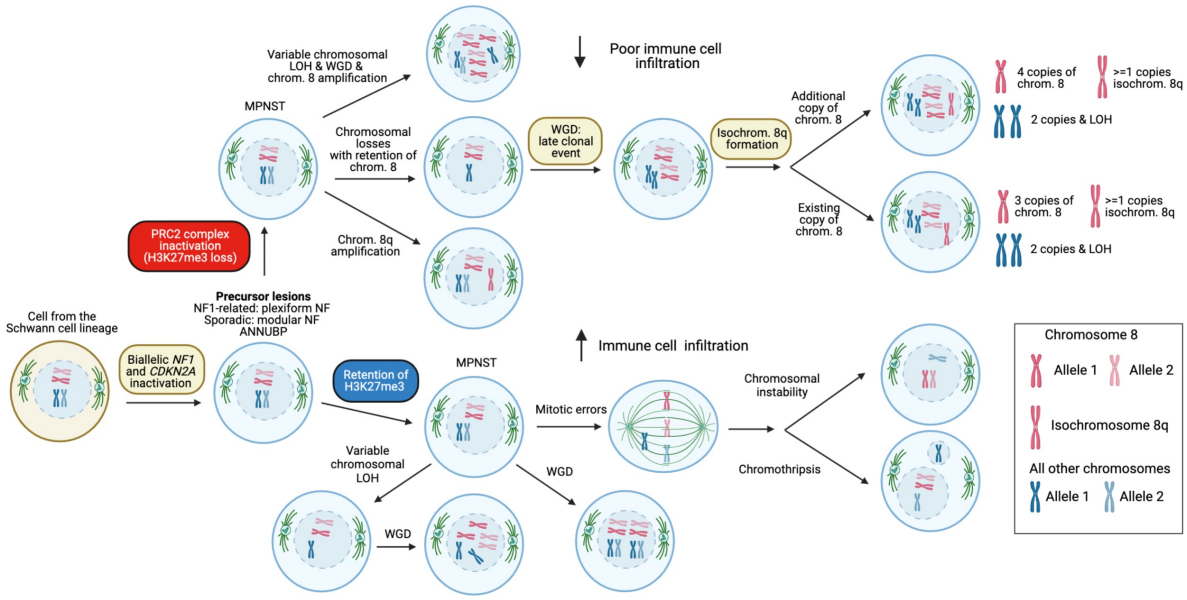


Figure 5. Pathways of MPNST evolution. Schematic representation of the order and timing of events involved in the evolution of MPNSTs. Copy number signatures predominant in specific karyotypic configurations are shown. CN, copy number signature; chrom., chromosome; isochrom., isochromosome; PRC2, polycomb repressive complex 2; WGD, whole genome doubling.

580 **Methods**
581
582 **Human Subjects**
583 The Genomics of MPNST (GeM) Consortium is comprised of investigators from academic
584 centres and hospitals ("member sites") with eligible subjects and existing tissue banking
585 protocols with consent for: biospecimen and clinical data collection; genetic testing, including
586 but not limited to whole genome/exome sequencing; sharing of specimens/data with outside
587 institutions, researchers, etc. The GeM coordinating centre at Boston Children's Hospital
588 (BCH) coordinated sharing of retrospectively-collected specimens from existing tumour banks
589 and pathology archives at member sites to perform molecular characterization of MPNST and
590 related tumours. The Dana Farber Cancer Institute IRB determined that this research met the
591 criteria for exemption from IRB review and was categorised as Secondary Use research.
592
593 All specimens and data were stripped of Protected Health Information (PHI) using the Safe
594 Harbor Method and coded. The prefix of each tumour ID included in the final cohort in Fig. 1
595 (n=90) is an abbreviation of the member site which contributed the required specimens for
596 analysis: Royal National Orthopaedic Hospital at University College, London ("ROY", n=43),
597 Moffitt Cancer Center, Tampa ("MOF", n=12), Mount Sinai Hospital, Toronto ("TOR", n=9),
598 Massachusetts General Hospital, Boston ("MGH", n=8), Washington University School of
599 Medicine, St. Louis ("WUS", n=6), Nagoya University, Nagoya ("NAG", n=4), Boston
600 Children's Hospital/Dana Farber Cancer Institute, Boston ("BCH", n=3), New York University
601 Langone Medical Center, New York (""NYU", n=3), Huntsman Cancer Institute at University
602 of Utah, Salt Lake City ("HCI", n=1), Lifespan Laboratories, Rhode Island Hospital,
603 Providence ("LIF", n=1).

604 The coordinating centre at BCH received coded material from the following types of
605 biospecimens: MPNSTs (both sporadic and NF1-related); precursor and other lesions, such as
606 benign, atypical, and plexiform neurofibroma; metastatic and local recurrence neurofibroma;
607 related specimens such as peripheral blood or tissue for germline comparison, and normal nerve
608 where available. Subjects with tumours that were unrelated to MPNST or neurofibroma and
609 subjects with specimens of insufficient quality and/or quality for pathology interpretation were
610 excluded from this study.

611 Comprehensive clinical and pathology report data for each participant was collected by the
612 international MPNST Registry at Washington University School of Medicine (WUSM). Data
613 include demographic information, disease course, tumour size/anatomical location,
614 histological/immuno-histochemical characteristics, diagnostic imaging, surgical procedures,
615 systemic treatment information, neoadjuvant therapy, toxicity, clinical outcomes and survival.

616 **Processing of fresh-frozen tumour and paired normal samples**
617 The GeM Consortium's Standard Operating Procedure (SOP) for the processing of tissue for
618 pathology review and molecular analysis was modelled on the Royal National Orthopaedic
619 Hospital's SOP for the 100,000 Genomes Project, founded by England's National Health
620 Service in 2012, as previously described[21]. Fresh-frozen tissue sections (H&E; 5 μm) were
621 assessed to select the most viable areas comprising high quality tumour samples in terms of
622 cellularity, lack of necrosis, and areas with little contaminating non-neoplastic tissue. Fresh-
623 frozen sections of high cellularity (at least 40%) and with less than 20% necrosis were used for

624 nucleic acid isolation. After pathologist assessment, two or four tubes of fresh-frozen tumour
625 tissue scrolls (sectioned at 10um each) were collected depending on whether the tumour was
626 of high cellularity (10,000+ cells) or low cellularity (<4,000 cells), respectively. Each tube
627 (Sarstedt – Biosphere Safeseal Tube 2.0 mL) of fresh-frozen tumour sections for DNA isolation
628 contained 20 curls of tissue in a buffer solution. Two tubes (Sarstedt – Biosphere Safeseal Tube
629 1.5 mL) with 6 curls each of fresh-frozen scrolls (sectioned at 10µm each) in 1 mL TRIZOL
630 were collected for RNA isolation, regardless of tumour cellularity. Four additional multi-
631 regional samples were taken from a subset of 9 fresh-frozen NF1-related MPNST specimens
632 to assess intra-tumour heterogeneity by performing 500x exome sequencing, whole
633 transcriptome sequencing by RNA-seq, and methylation profiling. Normal tissue samples were
634 confirmed to be free of tumour by a pathologist.
635
636 Fresh-frozen tumour and germline DNA samples were extracted and quality was assessed at
637 two pathology hubs, with approximately 50% of GeM samples processed at each site. Boston
638 Children's Hospital used the Maxwell RSC Tissue DNA Kit for extraction (Promega, Madison,
639 WI) and the Quantus Fluorometer and Picogreen (Promega) for quantification; Royal National
640 Orthopaedic Hospital used QIAamp's DNA Mini and Blood Maxi kits (Qiagen, Germantown,
641 MD) for extraction (Thermo Fisher Scientific, Waltham, MA) and the Qubit High Sensitivity
642 (HS) and Nanodrop (Thermo Fisher Scientific) for quantification and qualification,
643 respectively. All tumour/normal specimens with >400 ng total DNA (Picogreen, Promega) and
644 a DIN>=6.0 (TapeStation, Agilent) were included in this study. Fresh-frozen tumour and
645 normal nerve samples in TRIzol were stored at –80°C until transfer to the Broad Institute
646 (Cambridge, MA) on dry ice for RNA extraction (AllPrep DNA/RNA/miRNA Universal Kit;
647 Qiagen). RNA quality and insert-size was assessed by Caliper LabchipGXII (PerkinElmer,
648 Billerica, MA) producing a RQS value (equivalent to RIN). RNA quantity was determined by
649 Quant-iT RiboGreen (Thermo Fisher Scientific). Samples which met the minimum required
650 input (250 ng of purified total RNA with RIN >=6.0) were included in this study.
651
652 **Processing of FFPE samples**
653 Formalin-fixed, paraffin-embedded (FFPE) tissue blocks were collected for each fresh-frozen
654 tumour which underwent multi-omic profiling and were used for pathology review,
655 immunohistochemistry (MYF-4, H3K7me3, S100, Sox10) and additional molecular
656 characterization (i.e., ploidy assay and multi-regional whole-exome sequencing). FFPE
657 haematoxylin and eosin (H&E) slides from available blocks were reviewed by a pathologist to
658 identify the block most representative (i.e., morphologically similar) to the frozen material
659 which underwent multi-omic profiling. These H&E sections were digitally scanned (40x) at
660 each pathology hub by the Deep Lens Biomedical Imaging Team and uploaded to the VIPER
661 digital pathology platform (Deep Lens, Inc., Columbus, OH). Multi-regional 500x exome
662 sequencing was performed on 5 samples each from 36 FFPE NF1-related MPNST specimens
663 to assess intra-tumour heterogeneity.
664 **Pathology review**
665 All cases sent for multi-omic profiling underwent expert pathology review in the cloud-based
666 digital imaging platform VIPER hosted by Deep Lens (Columbus, OH). Pathology reports were
667 reviewed for the patient's NF1 status and lesion's proximity to a nerve, exposure to
668 neoadjuvant systemic or radiation therapies, and anatomical location. The number of slides

669 available for each case ranged from 1-5 slides, including the following stains: H&E,
670 H3K27me3, MYF-4, S100, and/or Sox10. Additional immunohistochemistry (HMB45 and
671 Melan-A) was performed to exclude melanoma.

672 The most representative FFPE H&E slide from each case was reviewed by at least one expert
673 pathologist and the following results were recorded in a case review form (CRF): percent
674 tumour necrosis, degree of cytological atypia, tumour cellularity and purity, mitotic count,
675 presence of lymphocytic infiltrate and atypical mitosis, and final tumour diagnosis based on
676 the slide image. Possible diagnoses included: neurofibroma, neurofibroma with atypia, cellular
677 neurofibroma, ANNUBP, low-grade MPNST, conventional high grade MPNST (hyper- and
678 hypo-cellular fascicles with/without geographic necrosis), conventional MPNST with
679 heterologous elements, non-conventional MPNST with associated conventional low-grade
680 areas, non-conventional MPNST, spindle cell/undifferentiated sarcoma, melanoma, and other
681 (e.g., dermatofibrosarcoma protuberans, synovial sarcoma, epithelioid MPNST, spindle cell
682 rhabdomyosarcoma, carcinoma etc.). Additional slides on which IHC was performed from each
683 case were reviewed and results were recorded in CRFs.

684 Consensus review in real-time via screen sharing and conference calls was conducted for all
685 cases in which there was a discrepant diagnosis between pathologists or diagnostics were
686 difficult based on the available H&E and IHC for each case. A second round of consensus
687 review was conducted for all cases in which there was a discrepancy between tumour diagnosis
688 by consensus review and the results of the methylation assay sarcoma classification. For the
689 purposes of collating tumour diagnosis with multi-omic data, tumour diagnoses were collapsed
690 into five main categories based on germline NF1 status and histological features of the
691 neurofibroma: germline/conventional, germline/non-conventional, sporadic/conventional,
692 sporadic/non-conventional, and equivocal. Equivocal cases were excluded from the genomics
693 analysis unless otherwise stated. Neurofibromas were classified using the criteria outlined in
694 histopathologic evaluation of atypical neurofibromatous tumours and their transformation into
695 MPNST in patients with NF1[45].

696 **Whole-genome sequencing**

697 Between 350 and 500ng of tumour DNA were used to prepare DNA sequencing libraries using
698 the TruSeq DNA PCR Free 350bp kit, which were sequenced on the Illumina NovaSeq6000
699 sequencing machine using S4 flow cells to generate 2 x 151 paired-end reads. Quality control
700 for all steps was performed using a DNA Screentape on the Agilent TapeStation system. Raw
701 sequencing reads were mapped to the GRCh38 build of the human reference genome using
702 BWA-MEM[46] version 0.7.17-r1188. Aligned reads in BAM format were processed following
703 the Genome Analysis Toolkit (GATK, version 4.1.8.0) Best Practices workflow to remove
704 duplicates and recalibrate base quality scores[47]. A minimum sequencing coverage of 30x and
705 90x was required for the normal and tumour samples, respectively. NGSCheckMate was
706 utilised using default options to verify that sequencing data from tumour-normal pairs
707 corresponded to the same individual[48].

708 **Whole-exome sequencing**

709 Four additional multi-regional cores were taken from the most representative FFPE block for
710 a subset of 10 cases for whole-exome sequencing. DNA extraction from FFPE tissue samples
711 was performed using the Covaris truXTRAC FFPE DNA kit with the E220 Evolution focused
712 ultrasonicator system, and Prepito truXTRAC DNA FFPE kit with the Chemagic Prepito-D.

713 Any excess paraffin was trimmed before sectioning an FFPE tissue block, or after the section
714 had been cut from the FFPE block, to maintain an optimal ratio of 80% tissue to 20% paraffin
715 (or higher). Cores of 1.22 mm diameter were taken; any sections or cores longer than 10 mm
716 in length were cut in half before being loaded into the microtubes. The total mass of FFPE
717 samples processed per extraction was between 2 and 5 mg. The quantity and purity of the DNA
718 was assessed using the Qubit dsDNA Assay kit and Nanodrop, respectively (ThermoFisher).
719 The DNA fragment size of each sample was estimated using an Agilent 2200 TapeStation
720 (Agilent Technologies). Exome libraries were prepared using Roche Kapa HyperPlus kit,
721 captured using the HyperExome kit Roche, and sequenced on Illumina NovaSeq6000, to
722 generate 2 x 151 paired-end reads.

723 **Tumour DNA ploidy assay**

724 Additional scrolls were cut (50 um) from the most representative FFPE block for all cases for
725 a ploidy assay. Of the 95 frozen cases that qualified for histological assessment by pathology
726 review, 14 had insufficient FFPE-derived tumour DNA qualifications for the ploidy assay.
727 Quantification of DNA content was performed as previously described[25,49]. Briefly, 50 mm
728 FFPE sections were deparaffinized and rehydrated, and cytoplasmic digestion with protease
729 type VIII (Sigma P5380) was utilised to create nuclear suspensions. Nuclear suspensions
730 underwent filtering, cytospinning, DNA hydrolysis (5 M HCl) and staining with Feulgen
731 (Schiff's fuchsin-sulphite reagent; Sigma S5133). Individual nuclei were classified as
732 "normal", "lymphocyte", "plasma", "fibroblast" and "tumour" using the PWS Classifier
733 software[50]. Nuclei were manually revised to select high confidence members of the "normal","
734 lymphocyte" and" tumour" categories. The median integrated optical density (IOD) value of
735 the combined "normal" and "lymphocyte" categories was taken as the IOD level for a diploid
736 cell (IOD_{2n}) in each sample. The maximum density value of the IOD of the major tumour
737 subclone was taken as the tumour clonal IOD level (IOD_t). Tumour ploidy was calculated as
738 $2(IOD_t/IOD_{2n})$.
739

740 **Detection of pathogenic germline *NF1* alterations**

741 The germline short variant discovery workflow from GATK (version 4.1.8.0)[47], Strelka2
742 (version 2.9.2)[51] and VarDict (version 1.8.2)[52] were used to detect germline single-nucleotide
743 variants (SNVs) and small insertions and deletions (INDELs) in the whole-genome sequencing
744 data from the matched normal samples. In brief, Strelka2 and VarDict were run using default
745 parameter values. In the case of GATK, intermediate GVCF files were generated for each
746 sample using HaplotypeCaller in GVCF mode. Next, GVCF files for all samples were
747 consolidated into a single GVCF file using the GATK functionality CombineGVCFs using
748 default options. Finally, all samples were jointly genotyped using GenotypeGVCFs and default
749 options. Annovar (version 2018Apr16) was used to annotate all variants[53]. Truncating variants
750 detected by at least one algorithm were considered to be pathogenic, and missense variants
751 were reviewed for literature support. All variants were also manually curated through visual
752 inspection of raw sequencing reads using BamSnap[54]. Deep intronic mutations generating
753 cryptic splice sites were validated at the RNAseq level and through literature review. To detect
754 germline structural variants (SVs) we followed two strategies. First, we called SVs in all
755 matched normal samples using Manta (version 1.6.0)[55], LUMPY (version 0.2.13)[56], SvABA
756 (version 1.1.3)[57], and Delly (version 0.8.3)[57,58]. Breakpoints mapping to the *NF1* locus,
757 including 100Kb upstream and downstream, were manually curated by inspecting the raw

758 sequencing reads. Second, to identify *NF1* microdeletions without SV support, we computed
759 the B-allele frequency profile for chromosome 17 for each sample using the heterozygous SNPs
760 from dbSNP (build 151). In addition, we computed the sequencing coverage at 500bp windows
761 using mosdepth[59]. To call an *NF1* microdeletion the B-allele frequency (BAF) and coverage
762 profiles at the *NF1* locus were required to clearly deviate from 0.5 and decrease by a factor of
763 2, respectively.
764
765 **Detection of somatic single-nucleotide variants and indels**
766 Somatic single-nucleotide variants (SNVs) were detected using Mutect2 (GATK version
767 4.1.8.0), MuSE (version v1.0rc)[60], and Strelka2 (version 2.9.2) using default options. Somatic
768 indels were called using Strelka2 (version 2.9.2) and Mutect2 (GATK version 4.1.8.0) using
769 default options. Each algorithm was run independently on each tumour sample using the
770 matched normal sample from the same individual as control. Indel calls were left-aligned using
771 the GATK functionality LeftAlignAndTrimVariants. The calls generated by each algorithm
772 were merged using the Python library mergevcf (https://github.com/ljdursi/mergevcf). To
773 maximise specificity, only mutations called by at least two algorithms were considered for
774 further analysis. Mutations overlapping known polymorphisms in dbSNP (build 151) were only
775 considered for further analysis if the coverage in the matched normal sample was at least 20
776 reads and none of them supported the mutation. Missense variants predicted to be deleterious
777 by MetaLR and MetaSVM as implemented in Annovar (version 2018Apr16) were considered
778 pathogenic. To identify driver genes, we used the R package dNdScv[61].
779
780 **Mutational signature analysis**
781 Mutational signatures were extracted de novo using non-negative matrix factorization (NMF)
782 as implemented in the R package NMF[62]. In brief, for each rank in the set {2,12} 2,000
783 factorizations were run. The rank maximising the consensus cophenetic correlation coefficient,
784 and for which the residual sum of squares (RSS) showed a clear inflection point, was selected
785 as the optimal rank[63]. The identified signatures were compared against the COSMIC v3.2
786 catalogue of signatures and assigned to existing ones using a cosine similarity cut-off of 0.85
787 using the R package MutationalPatterns[64]. Next, the contribution of the set of extracted
788 signatures to each sample was estimated using the function *fit_to_signature*s from
789 MutationalPatterns. For this analysis, signatures appearing in hypermutated tumours were only
790 considered when analysing the tumours.
791
792 **Detection of microsatellite instability**
793 Mutations at microsatellite loci were detected using MSIprofiler
794 (https://github.com/parklab/MSIprofiler) as previously described[65]. Only repeats with at least
795 10 sequencing reads in both the normal and tumour sample were considered for further
796 analysis. The Kolmogorov-Smirnov test was used to compare the distributions of repeat
797 lengths. The level of significance was set at 0.01 after False Discovery Rate (FDR) correction.
798
799 **Copy number analysis**
800 For WGS data, the software packages FACETS (version 0.12.1) and ascatNGS (version 2.5.1)
801 were used to detect somatic copy number alterations and to compute purity and ploidy estimates

802 for all tumour samples[66,67]. Both tools were run using default options. The distribution of cancer
803 cell fraction and mutation copy number values for somatic SNVs computed using the copy
804 number values were manually reviewed in all cases to assess the quality of the ploidy and purity
805 calls. In cases where inconsistencies were identified, such as the lack of clonal SNVs, a
806 different combination of purity and ploidy values was selected to refit the copy number profile.
807 The ploidy and purity values were selected to best match the ploidy ascertained from image
808 cytometry ploidy analysis. Timing analysis of whole-genome doubling events was performed
809 as previously described[25]. For WES data, FACETS (version 0.12.1) was used to compute copy
810 number, ploidy and purity calls for each tumour region.
811
812 **Copy number signature analysis**
813 For WGS data, allele counts were generated for SNP positions lifted over to hg38 (1,837,383
814 SNPs) using alleleCounter (https://github.com/cancerit/alleleCount). For WES data, allele
815 counts were instead generated for all hg38 dbSNP v153 common SNPs (12,140,046 SNPs).
816 LogR and BAF values were calculated for SNPs in all normal samples. SNP positions with a
817 high frequency of undetermined genotype (0.68<BAF<0.9 or 0.1<BAF<0.32) were excluded
818 from analysis. Further, any SNPs closer than 75bp to a neighbouring SNP were excluded from
819 analysis. Lastly, SNPs with a high variance in logR were excluded from analysis (SNP logR
820 variance>2*median SNP logR variance). Following SNP removal and cleaning, normal
821 samples with a high variance in logR were removed from the dataset (samples with
822 >1.5*median sample logR variance). LogR and BAF values were calculated for all tumour
823 samples with a corresponding retained normal sample. Tumour LogR values were corrected
824 for GC content, and copy number profiles were generated using ASCAT[68], with a segmentation
825 penalty of 70, designating SNPs with 0.2<BAF<0.8 in the corresponding normal as
826 heterozygous. For WES samples, segmentation was jointly performed for all samples
827 originating from a single patient using the function *asmultipcf*. Following copy number calling,
828 copy number profiles were summarised as vectors of counts of segments categorised into 48
829 copy number classes based on combinations of loss of heterozygosity (LOH) status
830 (homozygous deletions, LOH, heterozygous), total copy number (0, 1, 2, 3-4, 5-8 or 9+) and
831 segment length (0-100kb, 100kb-1Mb, 1Mb-10Mb, 10Mb-40Mb, >40Mb). For homozygous
832 deletions, segment lengths were restricted to 0-{100kb, 100kb-1Mb, >1Mb}. See [27] for a full
833 description of the development of these categories.
834
835 For WGS data, non-negative matrix factorisation as implemented in sigProfilerExtractor was
836 employed to extract de novo copy number signatures from sequential subsets of the dataset as
837 previously described[69]: all samples (n=81), excluding hypersegmented samples (nsegs<1000;
838 n=79), remaining poorly explained samples (cosine similarity<0.85; n=35), samples with a
839 high proportion of LOH (pLOH>; n=12). In each case, the de novo signatures were
840 decomposed into previously identified pan-cancer copy number signatures[27] using the function
841 *sigProfilerSingleSample*. For hypersegmented samples, a novel signature was identified, for
842 which the attribution was taken from the all-samples extraction. For the remaining samples, the
843 decomposition that best explained the data (highest cosine similarity) was retained. For WES
844 data, samples were decomposed into previously identified pan-cancer copy number signatures
845 using sigProfilerSingleSample[27].

846

847 **Detection of somatic structural variants**

848 Somatic SVs were detected in whole-genome sequencing data using Manta (version 1.6.0),
849 LUMPY (version 0.2.13), SvABA (version 1.1.3), and Delly (version 0.8.3). Each algorithm
850 was run independently on each tumour using the bulk sequencing data from the same individual
851 as control. The calls generated by each algorithm were merged using the Python library
852 mergevcf allowing 200bp of slop at the breakpoints. Only calls generated by at least two
853 algorithms were kept for further analysis. Intrachromosomal SVs were classified into four
854 groups (DEL: deletion; DUP: duplication; h2hINV; head-to-head inversion; and t2tINV; tail-
855 to-tail inversion) depending on the read orientation at the breakpoints following the notation
856 established by PCAWG[70]. SVs detected in at least two tumour samples from different
857 individuals were discarded, as these are likely germline polymorphisms or artefacts. Finally,
858 SVs with at least one breakpoint mapping to telomeres, centromeres, heterochromatin regions,
859 or to blacklisted regions by the ENCODE project were removed[71].

860

861 **Detection of chromothripsis**

862 Chromothripsis events were detected using ShatterSeek (version 0.7) using recommended cut-
863 off values[72]. We made a chromothripsis call if one of the following sets of criteria were
864 satisfied: (1) at least 6 interleaved intrachromosomal SVs, the copy number for at least 7
865 contiguous genomic segments oscillates between 2 total copy number states, the fragment joins
866 test, or either the chromosomal enrichment or the exponential distribution of breakpoints test;
867 or (2) at least 3 interleaved intrachromosomal SVs and 4 or more interchromosomal SVs, 7
868 adjacent segments oscillating between 2 total copy number states and the fragment joins test.
869 An FDR of 0.2 was used as the threshold for statistical significance.

870

871 **Rearrangement signatures**

872 Intrachromosomal structural variants were classified into categories according to (1) the SV
873 type as determined from the read orientation at the breakpoints, i.e., deletions, duplications,
874 and inversion (Ih2hINV and t2tINV SVs were grouped together); (2) and the size of the
875 genomic region bridged by the breakpoints: 10-10,000bp, 10-100Kb, 100Kb-1Mb, 1-10Mb,
876 and >10Mb. All translocations were grouped in a single category. In addition, SVs were further
877 stratified based on whether they mapped to an SV cluster. To define SV clusters, we used
878 piecewise constant regression applied on the inter-breakpoint distance values sorted by
879 genomic coordinates. We required at least 10 breakpoints per segment (kmin=10) and used a
880 value of 25 for the gamma parameter, which controls the smoothness of the segmentation[26].
881 SVs with at least one breakpoint mapping to a segment with an average inter-breakpoint
882 distance smaller than 10% the average inter-breakpoint distance were considered to be involved
883 in a cluster. Therefore, SVs were classified into 32 categories: 30 for intrachromosomal SVs
884 and 2 for translocations, depending on whether at least one of the breakpoints participated in a
885 cluster of SVs as defined above. Rearrangement signatures were extracted using NMF as
886 implemented in the R package NMF[62]. As in the case of SBS, the rank in the set {1..12}
887 maximising the consensus cophenetic correlation coefficient and for which the RSS showed an
888 inflection point that was considered optimal. In addition, the RSS difference obtained for

889 factorizations performed on the original and randomised data using the function *randomise*
890 were compared to ensure that the results obtained were not due to spurious correlations[73].
891
892 **mRNA sequencing**
893 At least 250 ng of purified total RNA with RIN >=6.0 were required. Library preparation was
894 conducted using the Illumina TruSeq Strand Specific Large Insert RNA kit (50M pairs) v1.
895 Thermo Fisher ERCC RNA controls were added prior to Poly(A) selection, providing
896 additional control for variability, including quality of the starting material, level of cellularity,
897 RNA yield, and batch to batch variability. Libraries were sequenced on the Illumina HiSeq4000
898 sequencing machine at the Broad Institute (Cambridge, MA). Sequencing reads were mapped
899 to the transcriptome using the STAR aligner (version 2.7.4a)[74]. Gene expression counts were
900 generated using HTSeq (v.0.6.1p1)[75] and normalised to transcripts per kilobase million (TPM).
901 GENCODE v22 was used as the gene annotation reference. Differential expression analysis
902 was performed using DESeq2[76]. Gene fusions were detected using ARRIBA (version 2.1.0)[77],
903 TopHat-fusion (version 2.1.0)[78], EricScript (version 0.5.5)[79], and STAR-Fusion (version
904 1.10.1)[80]. Only those fusions called by at least 2 algorithms or with WGS support were
905 considered for further analysis. Immune infiltration was estimated using consensusTME using
906 the sarcoma cell-type-specific genes derived from TCGA[81]. Immune gene sets were obtained
907 from previous studies[82–84].
908
909 **Methylation array profiling**
910 Tumour DNA derived from fresh-frozen tumour and paired normal nerve samples was used
911 for whole-genome DNA methylation array profiling using the HumanMethylationEPIC
912 beadchip platform as described previously[85]. In brief, 350ng of DNA were bisulfite converted
913 using the Zymo EZ DNA methylation Gold kit (Zymo Research Corp., Irvine, CA, USA) as
914 per manufacturer's recommendations. Bisulfite-converted samples were processed and
915 hybridised to the Infinium HumanMethylationEPIC beadchip arrays according to the
916 manufacturer's recommendations. Raw Methylation EPIC data were uniformly processed
917 using the R package minfi (version 1.34.0)[86]. Probes that did not satisfy a detection *P* value of
918 0.01 were discarded using the function *detP*. Probes mapping to known polymorphisms were
919 identified using the function *dropLociWithSnps* and were discarded. Functional normalisation
920 after background correction using normal-exponential deconvolution using out-of-band probes
921 (noob), as implemented in the function *preprocessFunnorm*, was applied to normalise
922 methylation intensities[87,88]. The function *getBeta* was used to convert methylation intensities
923 to Beta values, which were used for further analysis. Quality assessment was performed
924 through visual inspection of the distribution of B values for all samples. For methylation-based
925 diagnosis, iDAT files were analysed using the open-source DNA methylation sarcoma
926 classifier as described previously[22].
927
928 **Integrative analysis of multi-omic data**
929 We performed clustering of each data type separately using ConsensusClusterPlus[89]. The top
930 5,000 most variable genes (normalised counts) and the top 10,000 methylation CpG sites (Beta
931 values) based on median absolute deviation were used for clustering analysis. Clustering was
932 performed using the hierarchical clustering algorithm with 1,000 resamples, Pearson

933 correlation as the distance metric, and a maximum number of 10 clusters. Both the analysis of
934 methylation and RNAseq data revealed two main groups that were consistent across data types.
935 The cluster assignments computed using *ConsensusClusterPlus* were used for Cluster-Of-
936 Clusters Analysis (COCA)[90]. COCA, as implemented in the R package coca[91], was run on the
937 same data sets used for *ConsensusClusterPlus* clustering and using the *k*-means algorithm as
938 the clustering method.
939
940 **Analysis of cell-free DNA sequencing data**
941 Sequencing reads were filtered to keep read pairs corresponding to DNA fragments between
942 80 and 160bp. Genome-wide sequencing coverage was estimated by counting the number of
943 reads mapping to non-overlapping windows of 1Mb using readCounter[92]. IchorCNA[93] was
944 used to estimate the tumour fraction in cell-free DNA and genome-wide copy number profiles.
945 Only samples with a tumour fraction of at least 5% were considered for further analysis.
946 IchorCNA was run using default parameter values without accounting for subclonal events.
947
948 **Survival analysis**
949 Survival analysis was performed using the Cox proportional hazards model as implemented in
950 the R package survival (version 2.30)[94]. Significance was assessed by the likelihood ratio test
951 using a cut-off for statistical significance of 0.05. We only considered MPNSTs with
952 conventional histology for survival analysis. The proportional hazards assumption was tested
953 using the *cox.zph* function from the R package survival[94].
954
955 **Phylogenetic analysis of exome-sequencing data**
956 We constructed a binary matrix for each tumour with rows indexed by somatic SNVs and
957 columns by tumour regions such that the *i,j* entry in each matrix was set to one if mutation *i* is
958 present in region *j*, and to zero otherwise. To reconstruct phylogenetic trees for FFPE cases we
959 used the parsimony ratchet method (*pratchet*) as implemented in the R package phangorn[95].
960 The *acctran* method was used to infer branch lengths for all trees. In the case of fresh-frozen
961 samples, we used the maximum parsimony and neighbour joining algorithms as implemented
962 in the R package MesKit[96] to estimate consensus phylogenetic trees for each tumour using 100
963 bootstrap resamples.
964
965 **Data availability**
966 Raw sequencing data has been deposited at the European Genome-phenome Archive (EGA),
967 which is hosted by the EBI and the CRG, under the dataset accession number
968 EGAD00001008608. Ultra-low-pass whole-genome sequencing data of cell-free samples is
969 available under controlled access data at https://doi.org/10.7303/syn23651229.
970
971 Correspondence and requests for materials should be addressed to David T. Miller (MD, PhD):
972 david.miller2@childrens.harvard.edu
973
974 **Code availability**
975 The code used to process the raw data is available upon request.
976

977 **Methods References**

978 45. Miettinen, M. M. *et al.* Histopathologic evaluation of atypical neurofibromatous tumors
979 and their transformation into malignant peripheral nerve sheath tumor in patients with
980 neurofibromatosis 1-a consensus overview. *Hum. Pathol.* **67**, 1–10 (2017).
981 46. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
982 transform. *Bioinformatics* **26**, 589–595 (2010).
983 47. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud: Using Docker,*
984 *GATK, and WDL in Terra*. (O'Reilly Media, 2020).
985 48. Lee, S. *et al.* NGSCheckMate: software for validating sample identity in next-generation
986 sequencing studies within and across data types. *Nucleic Acids Res.* **45**, e103 (2017).
987 49. Hedley, D. W., Friedlander, M. L., Taylor, I. W., Rugg, C. A. & Musgrove, E. A.
988 Method for analysis of cellular DNA content of paraffin-embedded pathological material
989 using flow cytometry. *J. Histochem. Cytochem.* **31**, (1983).
990 50. Dunn, J. M. *et al.* Image cytometry accurately detects DNA ploidy abnormalities and
991 predicts late relapse to high-grade dysplasia and adenocarcinoma in Barrett's oesophagus
992 following photodynamic therapy. *Br. J. Cancer* **102**, (2010).
993 51. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat.*
994 *Methods* **15**, 591–594 (2018).
995 52. Lai, Z. *et al.* VarDict: a novel and versatile variant caller for next-generation sequencing
996 in cancer research. *Nucleic Acids Res.* **44**, e108 (2016).
997 53. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
998 variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
999 54. Kwon, M., Lee, S., Berselli, M., Chu, C. & Park, P. J. BamSnap: a lightweight viewer
1000 for sequencing reads in BAM files. *Bioinformatics* (2021)
1001 doi:10.1093/bioinformatics/btaa1101.
1002 55. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and
1003 cancer sequencing applications. *Bioinformatics* vol. 32 1220–1222 (2016).
1004 56. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic
1005 framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
1006 57. Wala, J. A. *et al.* SvABA: genome-wide detection of structural variants and indels by
1007 local assembly. *Genome Res.* **28**, 581–591 (2018).
1008 58. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-
1009 read analysis. *Bioinformatics* **28**, i333–i339 (2012).
1010 59. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and
1011 exomes. *Bioinformatics* **34**, 867–868 (2018).
1012 60. Fan, Y. *et al.* MuSE: accounting for tumor heterogeneity using a sample-specific error
1013 model improves sensitivity and specificity in mutation calling from sequencing data.
1014 *Genome Biol.* **17**, 178 (2016).
1015 61. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues.
1016 *Cell* **171**, 1029–1041.e21 (2017).
1017 62. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization.
1018 *BMC Bioinformatics* **11**, 367 (2010).
1019 63. Brunet, J.-P., Tamayo, P., Golub, T. R. & Mesirov, J. P. Metagenes and molecular
1020 pattern discovery using matrix factorization. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4164–
1021 4169 (2004).
1022 64. Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns:
1023 comprehensive genome-wide analysis of mutational processes. *Genome Med.* **10**, 33
1024 (2018).
1025 65. Cortes-Ciriano, I., Lee, S., Park, W.-Y., Kim, T.-M. & Park, P. J. A molecular portrait of
1026 microsatellite instability across multiple cancers. *Nat. Commun.* **8**, 1–12 (2017).

66. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).
67. Raine, K. M. *et al.* ascatNgs: Identifying Somatically Acquired Copy-Number Alterations from Whole-Genome Sequencing Data. *Curr. Protoc. Bioinformatics* **56**, 15.9.1–15.9.17 (2016).
68. Van Loo, P. *et al.* Allele-specific copy number analysis of tumors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16910–16915 (2010).
69. Ashiqul Islam, S. M. *et al.* Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *bioRxiv* 2020.12.13.422570 (2021) doi:10.1101/2020.12.13.422570.
70. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
71. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).
72. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. doi:10.1101/333617.
73. Frigyesi, A. & Höglund, M. Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. *Cancer Inform.* **6**, (2008).
74. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* vol. 29 15–21 (2013).
75. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
76. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
77. Uhrig, S. *et al.* Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res.* **31**, 448–460 (2021).
78. Kim, D. & Salzberg, S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.* **12**, R72 (2011).
79. Benelli, M. *et al.* Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* **28**, 3232–3239 (2012).
80. Haas, B. J. *et al.* STAR-fusion: Fast and accurate fusion transcript detection from RNA-Seq. *bioRxiv* (2017) doi:10.1101/120295.
81. Jiménez-Sánchez, A., Cast, O. & Miller, M. L. Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* **79**, 6238–6246 (2019).
82. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* vol. 160 48–61 (2015).
83. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, (2017).
84. House, I. G. *et al.* Macrophage-Derived CXCL9 and CXCL10 Are Required for Antitumor Immune Responses Following Immune Checkpoint Blockade. *Clin. Cancer Res.* **26**, 487–504 (2020).
85. Serrano, J. & Snuderl, M. Whole Genome DNA Methylation Analysis of Human Glioblastoma Using Illumina BeadArrays. *Methods Mol. Biol.* **1741**, 31–51 (2018).
86. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

87. Triche, T. J., Jr, Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
88. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).
89. Wilkerson, M. D. & Neil Hayes, D. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* vol. 26 1572–1573 (2010).
90. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
91. Cabassi, A. & Kirk, P. D. W. Multiple kernel learning for integrative consensus clustering of omic datasets. *Bioinformatics* **36**, 4789–4796 (2020).
92. HMMcopy. https://bioconductor.riken.jp/packages/3.0/bioc/html/HMMcopy.html.
93. Adalsteinsson, V. A. *et al.* Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
94. Therneau, T. M. & Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. (Springer Science & Business Media, 2013).
95. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2010).
96. Liu, M. *et al.* MesKit: a tool kit for dissecting cancer evolution of multi-region tumor biopsies through somatic alterations. *Gigascience* **10**, (2021).

Acknowledgements

This research was funded by the Neurofibromatosis Research Initiative (NFRI) at Boston Children's Hospital, made possible by an anonymous donation. We thank S. Cano, J. Da, M. Fournier, N. Gokgoz, J. Hart, J. Johnson, D. Lucente, T. Restrepo, M. Roche and A. Silva for technical assistance in collection and processing of specimens. We thank C. Clinton and A. Ward for their guidance and support. I.C.-C. thanks EMBL for funding. NP is funded through a Cancer Research UK grant (grant no: 18387). NP and AMF are supported by the UCLH Biomedical Research Centre and the CRUK Experimental Cancer Centre. Provision of patients' samples from the RNOH was made possible through the Royal National Orthopaedic Hospital Research and Development Department, The Rosetrees Trust, Skeletal Cancer Trust, Sarcoma UK and The Bone Cancer Research Trust over the last two decades. CDS is funded through Cancer Research UK and the Neurofibromatosis Research Initiative (NFRI) at Boston Children's Hospital-GEM consortium. DNA methylation profiling at NYU is in part supported by grants from the Friedberg Charitable Foundation, the Sohn Conference Foundation, and the Making Headway Foundation (to M.S.). We thank the patients and their families for their participation in this study.

Author Contributions

A.A., M.M.B., B.C.D., A.M.F., J.H., A.C.H., K.J., J.T.J., R.H.K., D.L., D.T.M., Y.N., M.S., D.V. and X.W. provided samples. D.C.B., V.E., A.M.F., J.H., A.C.H., J.T.J., D.L., D.T.M., Y.N., K.P., M.S., R.T.S., J.J.S., D.V. and X.W. provided clinical data. A.A., M.M.B., B.C.D., A.M.F., A.C.H., D.L., K.P., J.S. and M.S. provided technical support in specimen processing. D.C.B., V.E., A.C.H. and K.P. conducted clinical data processing. R.T.S. and J.J.S. provided annotated genomic data. A.C.H., I.C.C., A.F., A.M.F., A.G., C.M., P.J.P., N.P., P.P., J.S., M.S. and C.D.S. conducted genomic data processing. I.C.C., A.A., M.M.B., A.M.F., A.C.H., D.L.,

P.J.P., D.T.M., N.P., M.S., C.D.S. and X. W. developed the experimental design. A.C.H., I.C.C., A.A., A.C, M.M.B., A.M.F., B.C.D., A.G., A.C.H., P.J.P., D.T.M., N.P., J.S., M.S. and C.D.S. conducted data analysis and interpretation. I.C.C., A.M.F., A.C.H., D.T.M., N.P., K.P., M.S. and C.D.S. drafted the manuscript. A.A.I., D.C.B., M.B., A.C., I.C.C., B.C.D., V.E., A.F., A.M.F., A.G., A.G., J.G., J.H., A.C.H., K.B.J., J.T.J., G.J., R.H.K., D.L., C.M., D.T.M., Y.N., P.J.P., K.P.., N.P., P.P., J.S., M.S., C.D.S., R.T.S., J.J.S., N.J.U., D.V. and X.W. reviewed and approved the manuscript. I.C.C., A.M.F., A.C.H., D.T.M. and N.P. jointly supervised the work.

Competing Interest Declaration

A.A., A.C., I.C.C., B.C.D., V.E., A.F., A.M.F., A.G., J.H., K.B.J., R.H.K., D.L., C.M., P.J.P., K.P., N.P., P.P., J.S., M.S., C.D.S., R.T.S., J.J.S., D.V., and X.W. have no conflicting interests to declare.

M.B. Advisory boards for BioAtlas, Epizyme, Bristol-Myers Squibb, ContextVision, Airforia, Caris Life Sciences, and GlaxoSmithKline, consultant for AstraZeneca Pharmaceuticals LP, Foundation Medicine Inc, Visiopharm, Roche Laboratories, Inc.; B.C.D. received lab supplies from Illumina; J.F.G. consulted for Pfizer; A.C.H. Advisory boards for AstraZeneca and Springworks Consulting, received Intellisphere Research Funding from Tango Therapeutics Licensing Agreement with Deutsches Krebsforschungszentrum; J.T.J: Navio Theragnostics consultant, Recursion member scientific advisory board and consulting, Health2047 consultant, CEC Oncology speaker at a single CME conference regarding plexiform neurofibromas; D.T.M: Advisory Board, AstraZeneca; Y.N. Chairman, Japanese Society of Recklinghausen Disease; N.U. received royalties from University of Alabama, Birmingham and UpToDate, paid lecture to the advisory board of Astra Zeneca, expert testimony for Wolf, Horowitz & Etlinger, LLC

Group Authors: The Genomics of MPNST (GeM) Consortium

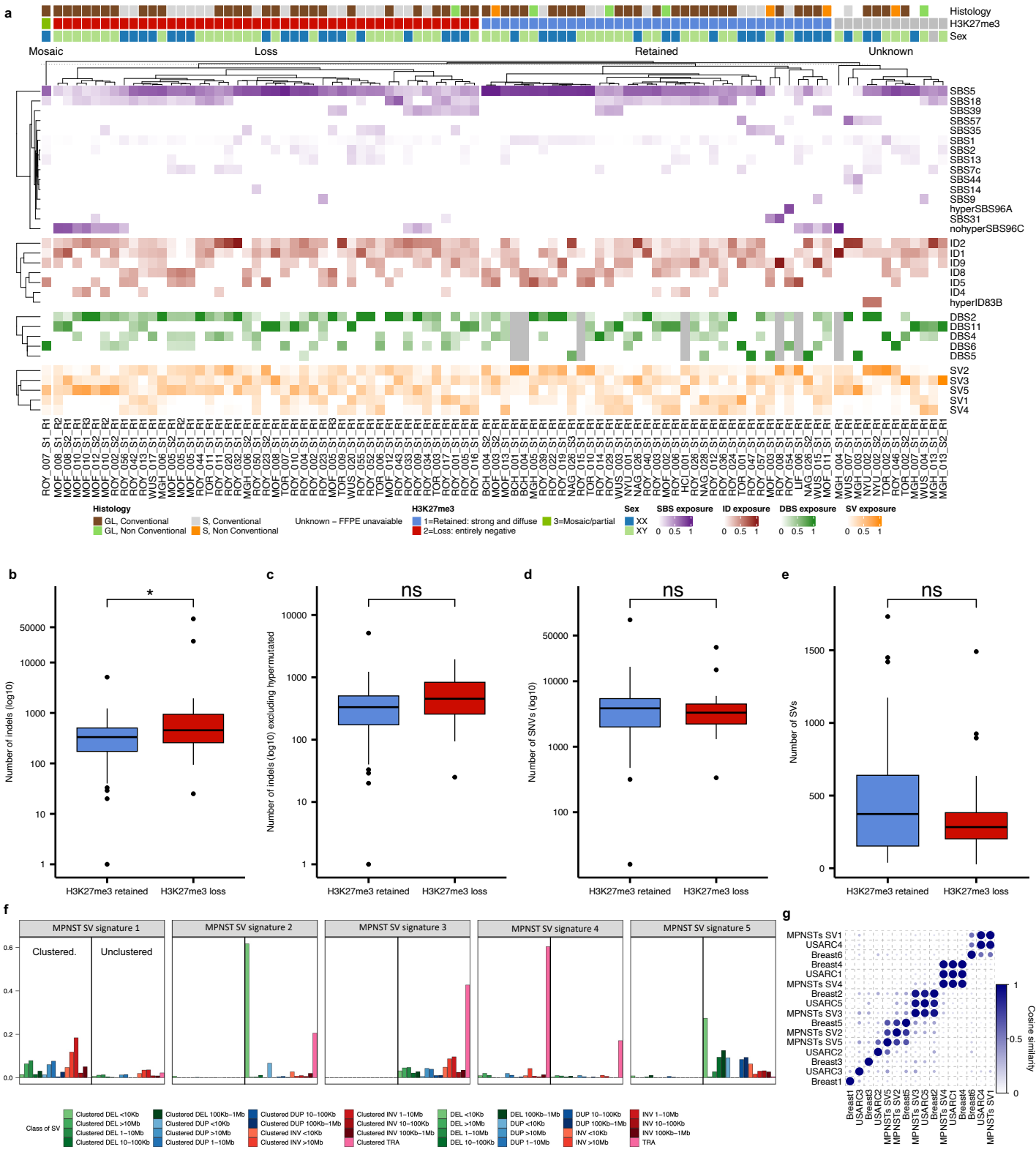
Davies C, Griffin A, Gusella JF, Jour G, Sakai T, Shao D, and Stemmer-Rachamimov AO

Supplementary Tables

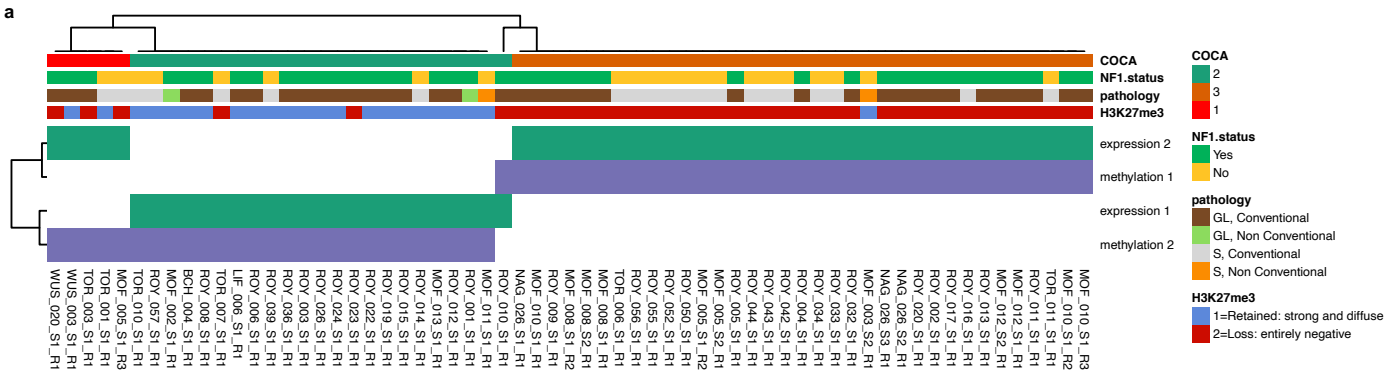
Supplementary Table 1. Clinical information for the GeM cohort.

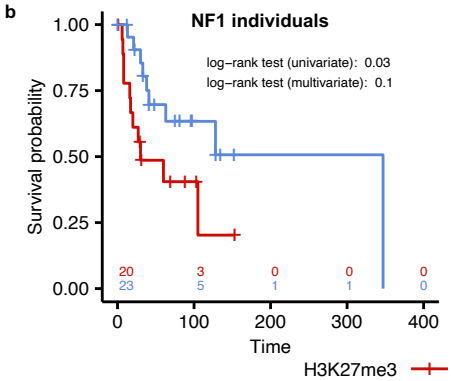
Supplementary Table 2. Differentially expressed genes between MPNSTs with loss or retention of H3K27me3 immunoreactivity.

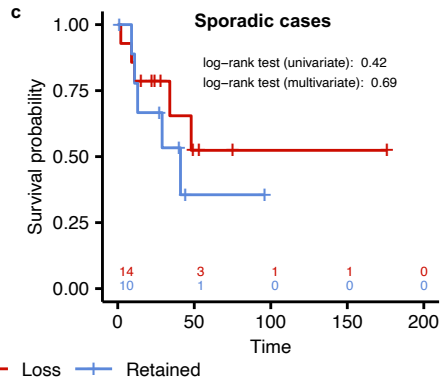
Extended Data Figure 1. Patterns and rates of somatic mutations in MPNSTs detected using WGS data. a, Mutational signatures in MPNSTs. **b,** Genome-wide number of indels. **c,** Genome-wide number of indels excluding hypermutated cases (>20,000 indels). **d** Genome-wide number of SNVs. **e,** Genome-wide number of SVs. **f**, SV signatures detected in MPNSTs. **g**, Comparison of the SV signatures detected in MPNSTs with previously reported SV signatures. Box plots in **b-d** show median, first and third quartiles (boxes), and the whiskers encompass observations within a distance of 1.5× the interquartile range from the first and third quartiles. SBS, single-base substitution signature; ID, indel signature; DBS, doublet base substitution signature; SV, structural variation signature. GL, germline; S, sporadic. ns, not significant, $P > 0.05$; *, $P < 0.05$ (two-sided Wilcoxon rank sum test).

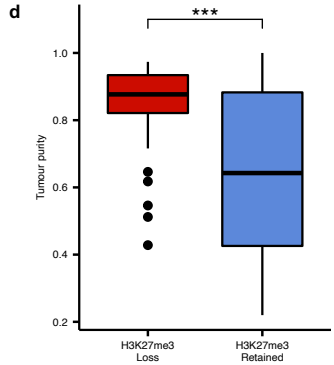


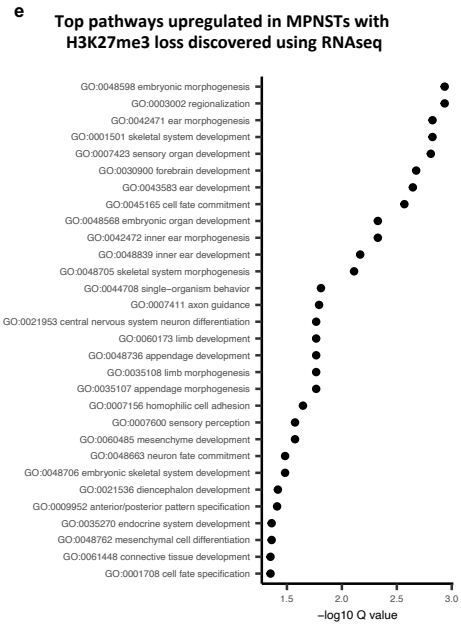
Extended Data Figure 2. Integrative molecular analysis of MPNSTs. a, Cluster-Of-Clusters Analysis (COCA) using methylation and mRNAseq data. **b,** PCA analysis of mRNAseq data. **b,** Kaplan-Meier plots of overall survival of NF1 patients based on H3K27me3 status. **c,** Kaplan-Meier plots of overall survival of patients with sporadic MPNST based on H3K27me3 status. **d,** WGS estimates of tumour purity. Gene Ontology (GO) terms upregulated (**e**) and downregulated (**f**) in MPNSTs with loss of H3K27me based on RNAseq data analysis. **g,** Gene sets enriched in MPNSTs with loss of H3K27me based on the analysis of differentially methylated regions.

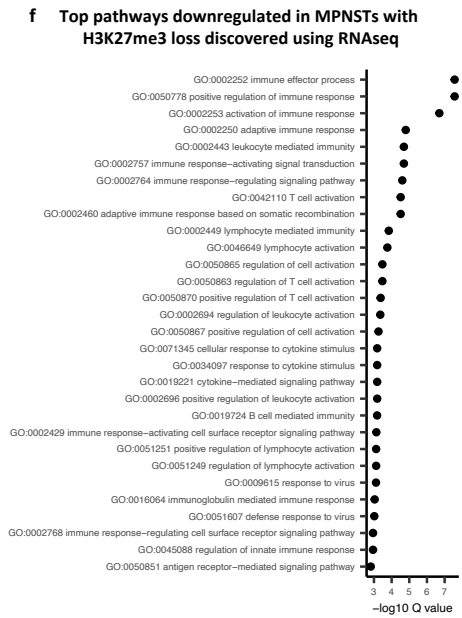


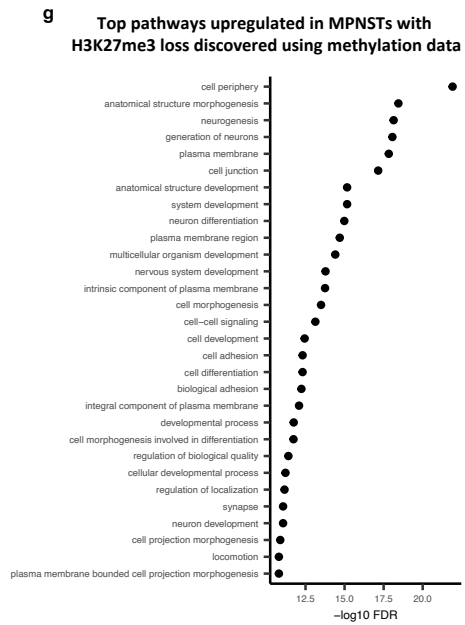




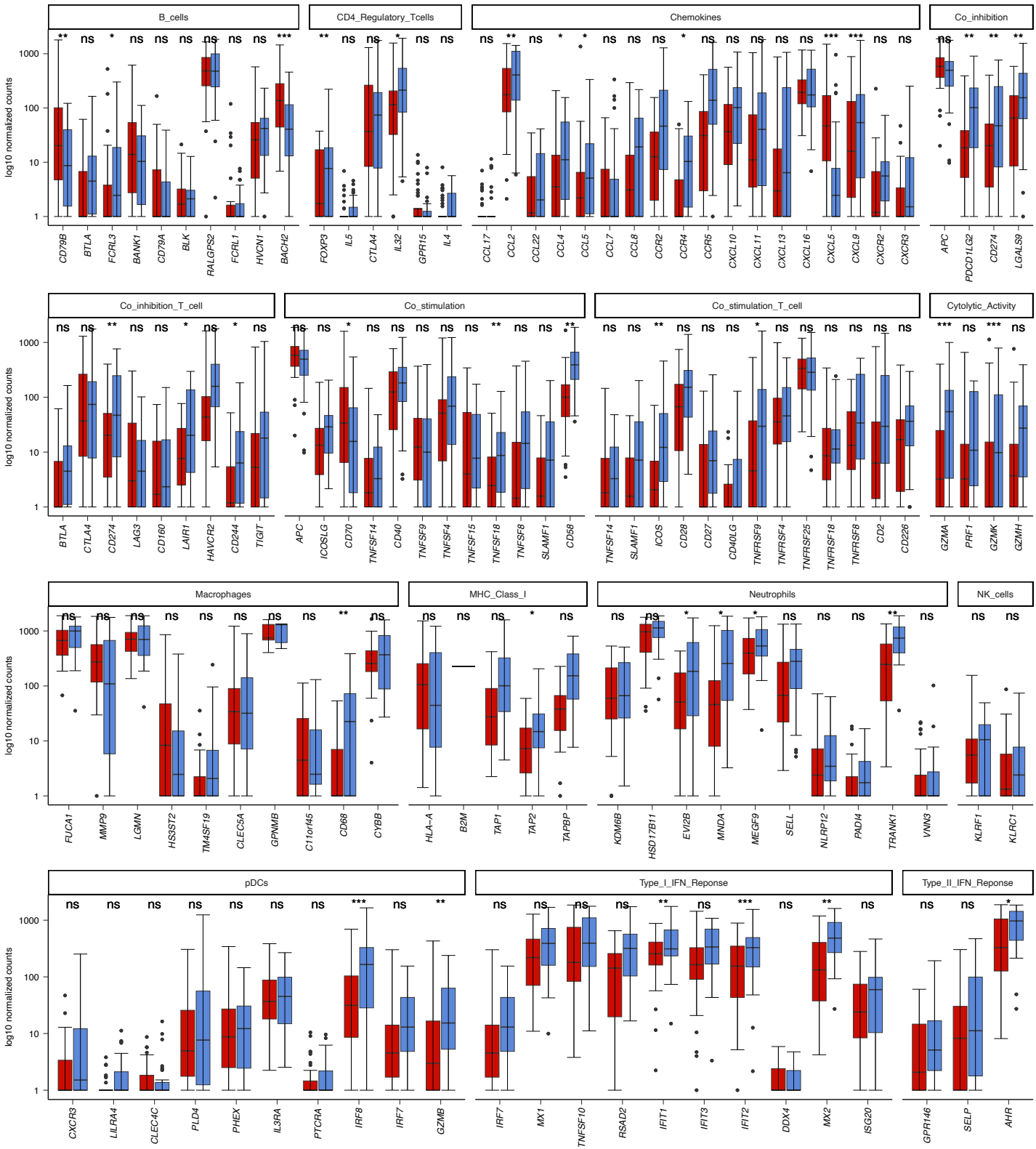




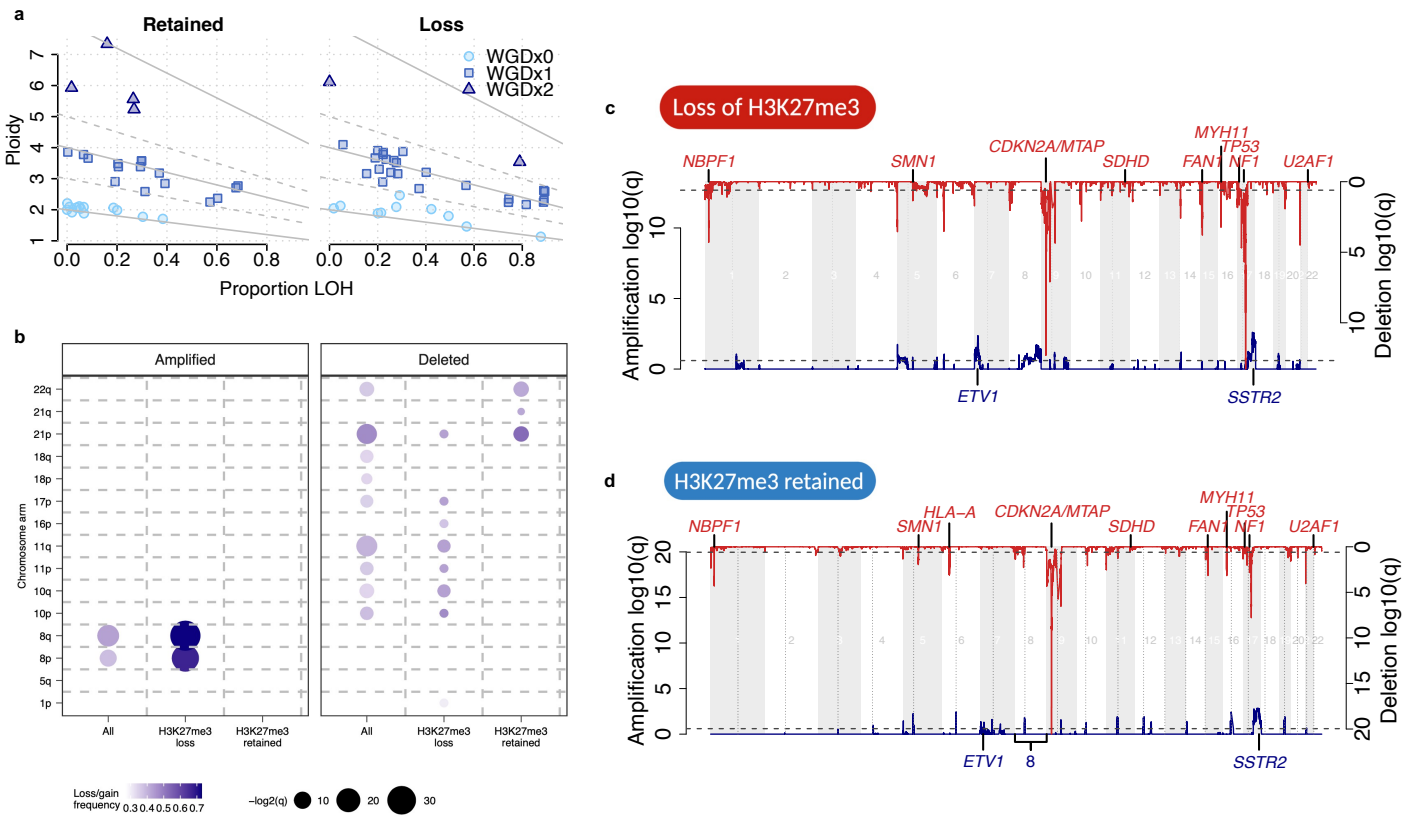




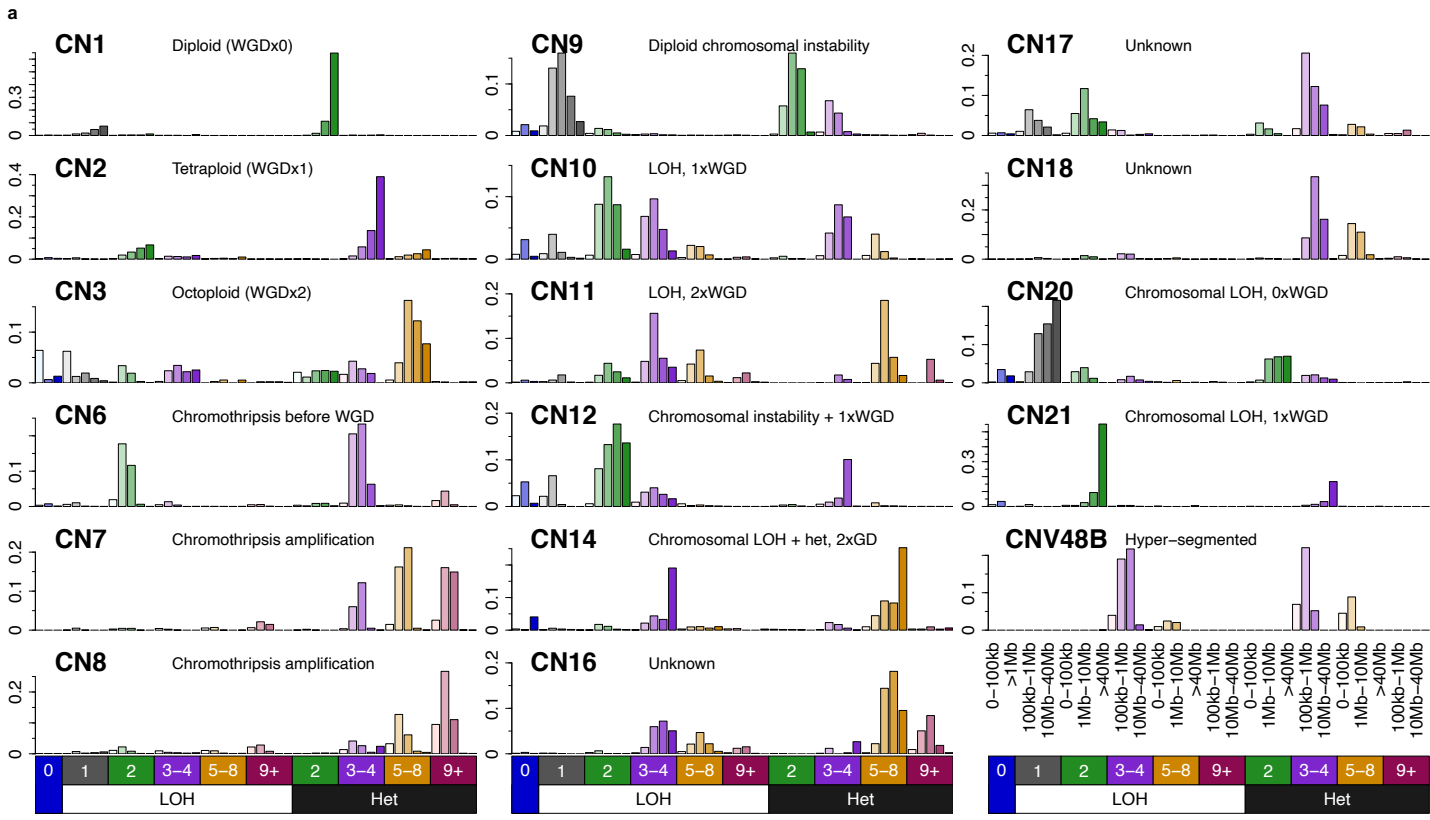
Extended Data Figure 3. Analysis of the expression of immune signatures in MPNSTs.
Box plots show median, first and third quartiles (boxes), and the whiskers encompass observations within a distance of 1.5× the interquartile range from the first and third quartiles. pDCs, plasmacytoid dendritic cells. ns, not significant $P > 0.05$; *, $P < 0.05$; **$P < 0.01$; ***$P < 0.001$ (two-sided Wilcoxon rank sum test).

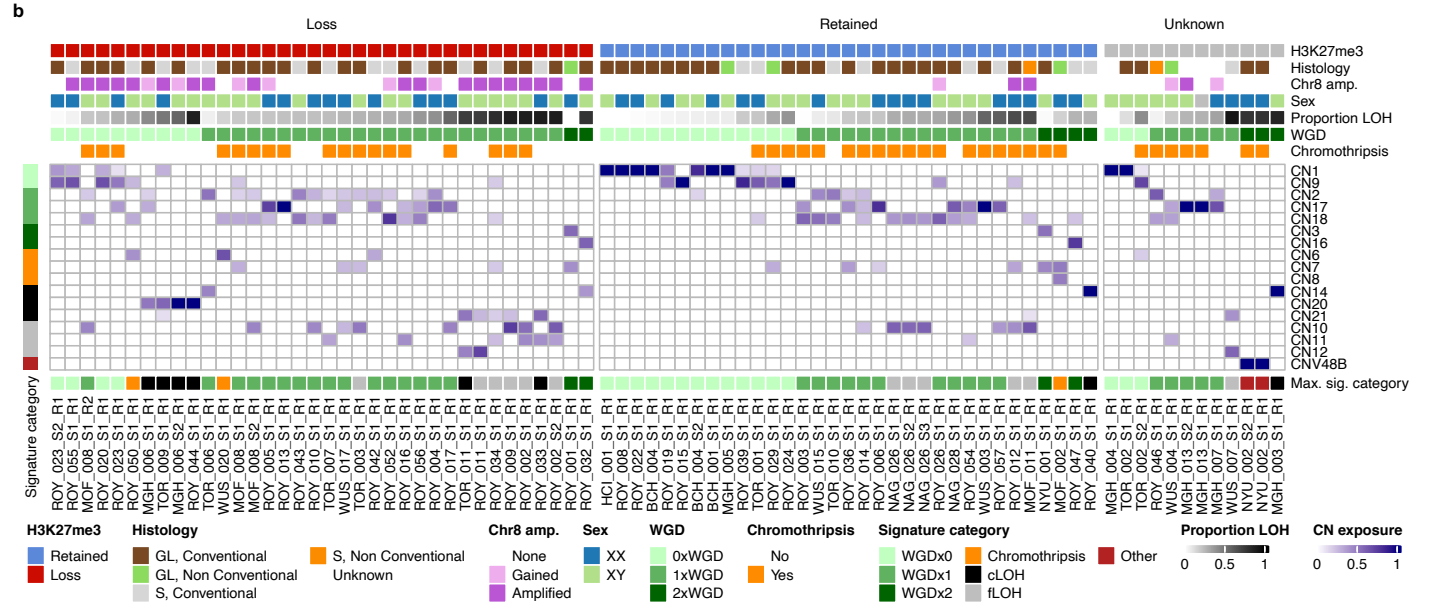


Extended Data Figure 4. Analysis of recurrent copy number aberrations and whole-genome duplication events. a, Relationship between tumour ploidy and the fraction of the genome showing loss of heterozygosity (LOH). **b,** Analysis of chromosome arms recurrently amplified or deleted in MPNSTs. GISTIC analysis of recurrent genome wide copy number amplifications (shown in red) and deletions (shown in blue) in tumours with loss (**c**) or retention (**d**) of H3K27me3, respectively. Genes mapping to segments recurrently amplified or deleted are shown at the top and bottom, respectively.

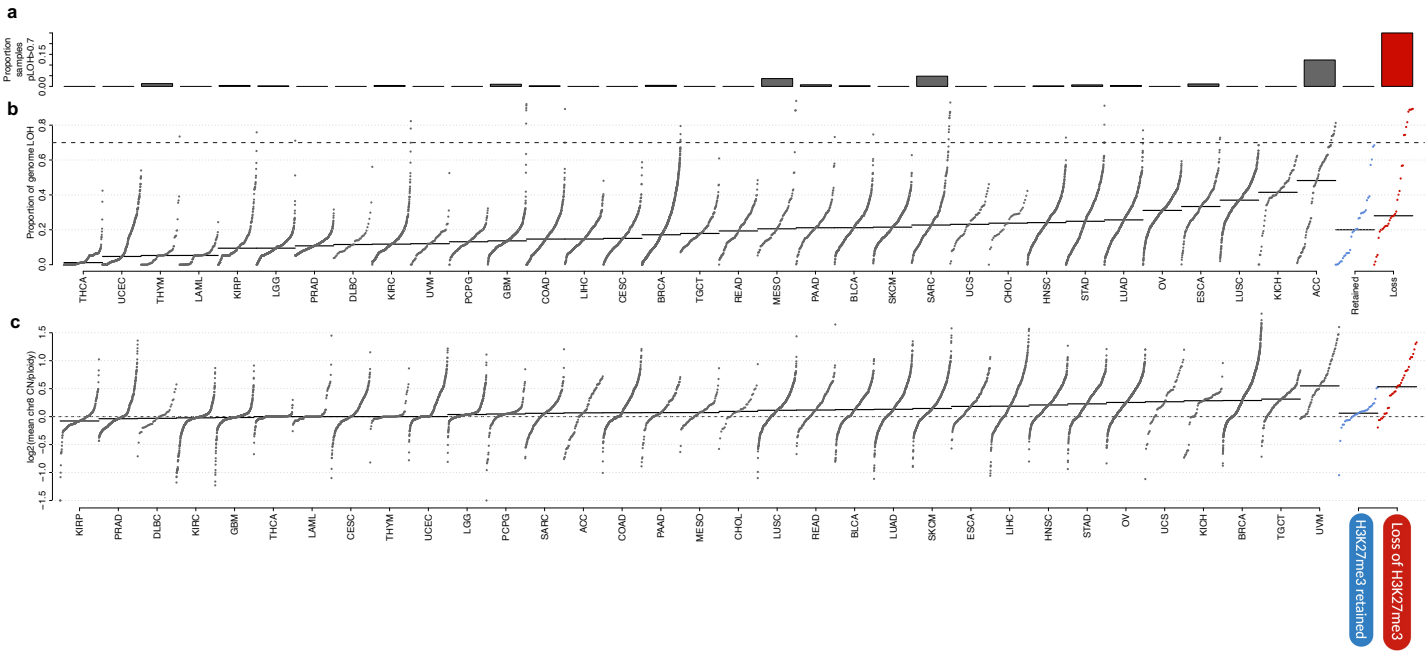


Extended Data Figure 5. Analysis of copy number signatures. a, Copy number signatures detected in MPNSTs using whole-genome sequencing data. CN, copy number signature; LOH, loss of heterozygosity; Het, heterozygous genomic segment; WGD, whole-genome doubling. **b,** The exposure of each copy number signature in each sample is shown. Samples are stratified based on H3K27me3 status. Signature abbreviations. CN1: Diploid (large heterozygous segments with CN=2); CN2: Tetraploid (large heterozygous segments with CN=3-4); CN3: Octoploid (large heterozygous segments with CN=5-8); CN4: Chromothripsis; CN5: Chromothripsis; CN6: Chromothripsis before WGD (LOH segments indicate before WGD); CN7: Chromothripsis associated amplification, after WGD; CN8: Chromothripsis associated amplification, after WGD; CN9: Diploid chromosomal instability; CN10: Extensive LOH, followed by 1x WGD; CN11: Extensive LOH, followed by 2x WGD; CN12: Extensive LOH, followed by 1x WGD, larger segment sizes than CN10; CN13: Early chromosomal LOH, with chromosomal heterozygosity (retained chrs), 1x WGD; CN14: Early chromosomal LOH, with chromosomal heterozygosity (retained chrs), 2x WGD; CN15: Associates with HRD; CN16: Unknown aetiology, 2x WGD, mostly heterozygous; CN17: Unknown aetiology; CN18: Unknown aetiology; CN19: Unknown aetiology; CN20: Early chromosomal LOH, 0x WGD; CN21: Early chromosomal LOH, 1x WGD; CN22: Early chromosomal LOH, 2x WGD; CNV48B: Extreme segmentation.

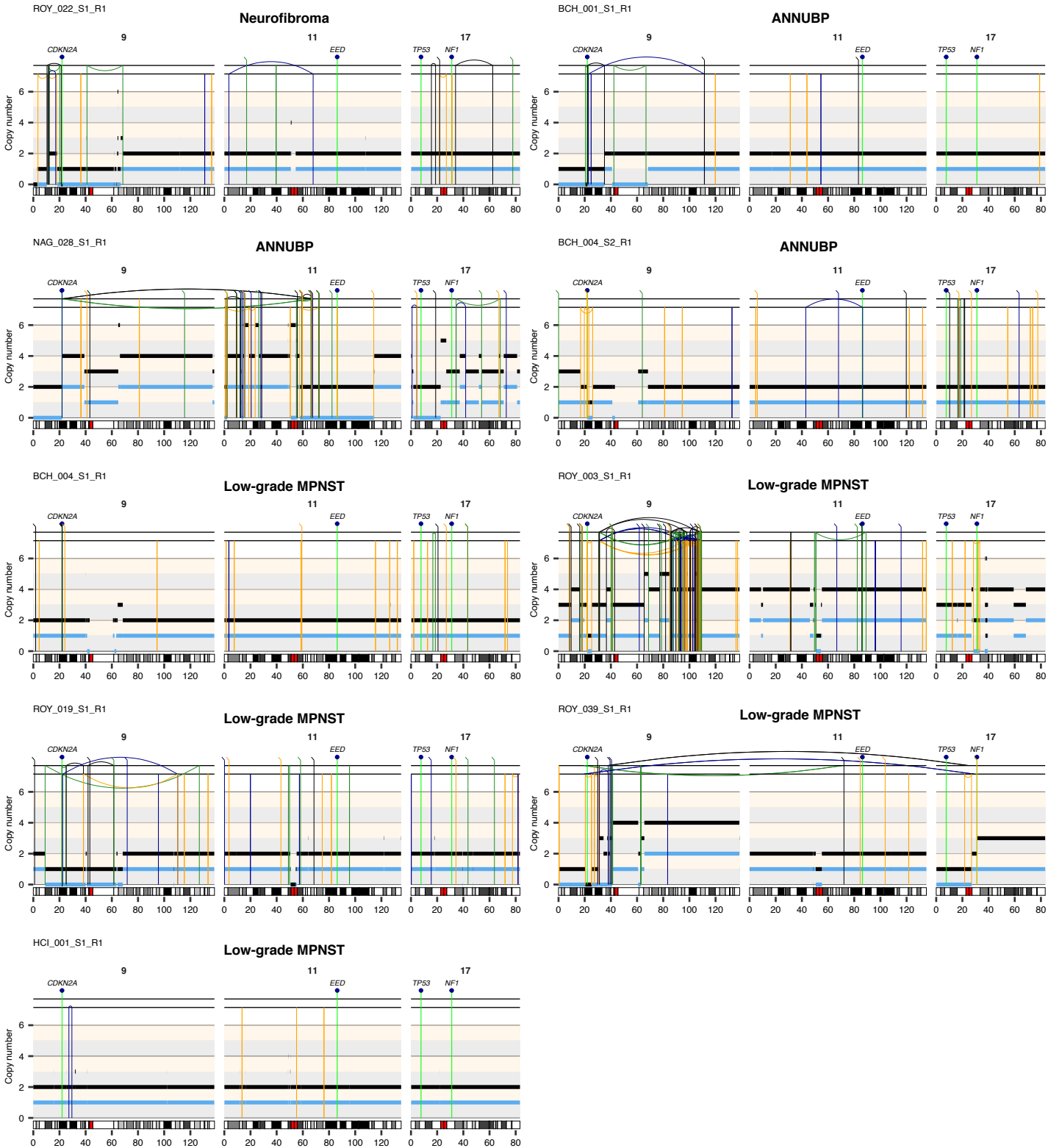




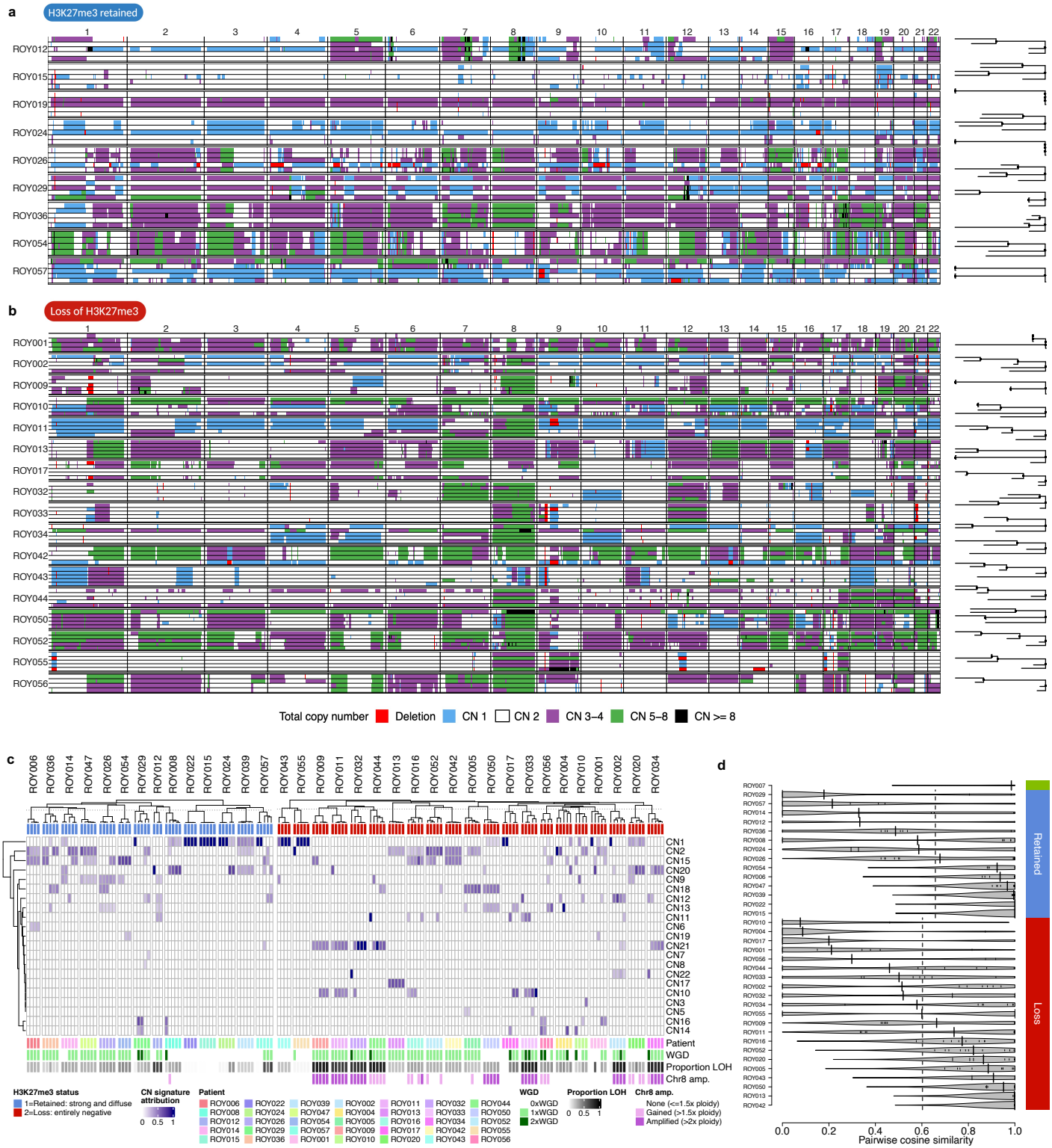
Extended Data Figure 6. Pan-cancer analysis of genome-wide loss of heterozygosity and chromosome 8 amplification. a, Fraction of tumours in TCGA showing LOH in at least 70% of the genome across 33 cancer types. **b**, Fraction of LOH in 9,699 tumours from TCGA. Cancer types are sorted with respect to the 90% percentile (black line). The grey line denotes the median fraction of LOH for each cancer type. **c**, Chromosome 8 copy number in relation to overall ploidy across TCGA. Cancer types are ordered by their median chromosome 8: ploidy ratio (horizontal lines).



Extended Data Figure 7. Analysis of rearrangements at the *CDKN2A, EED, TP53* and *NF1* loci. Rearrangement profiles of 1 neurofibroma (ROY_022_S1_R1), 3 ANNUBPs (BCH_001_S1_R1, BCH_004_S2_R1 and NAG_028_S1_R1) and 5 low-grade MPNSTs (BCH_004_S1_R1, ROY_003_S1_R1, ROY_019_S1_R1, ROY_039_S1_R1, HCI_001_S1_R1) in the GeM cohort. The position of *CDKN2A, EED, TP53* and *NF1* is shown at the top of each plot.



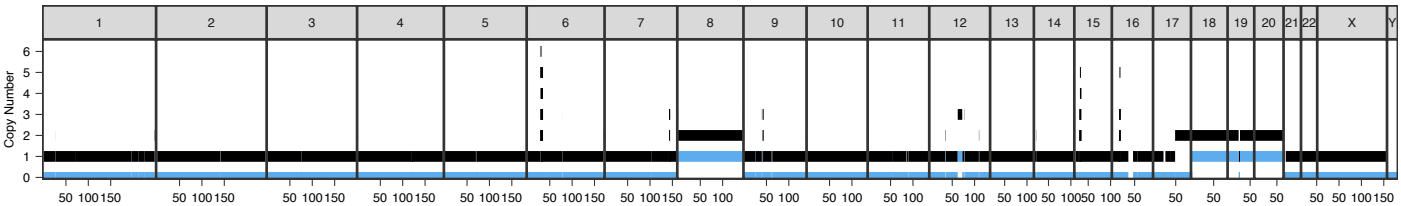
Extended Data Figure 8. Analysis of the intra-tumour heterogeneity of copy number aberrations in MPNSTs with retention of H3K27me3.
Genome-wide total copy number profiles for each region from tumours with retention (**a**) and loss (**b**) of H3K27me3 are shown. The phylogenetic trees generated using SNVs are shown on the right-hand side. **c**, Attribution of copy number signatures in multi-regional exome sequencing data (blue), clustered both within and between patients. Clinical and genetic covariates are shown at the top and bottom of the heatmap. **c**, Pairwise cosine similarity of the copy number signature attributions across regions from the same patient. Individual cosine similarities (thin black lines) and median cosine similarity (thick black lines) are denoted for individual patients. Dashed lines indicate mean pairwise cosine similarities across H3K27me3 loss and retained patients separately.



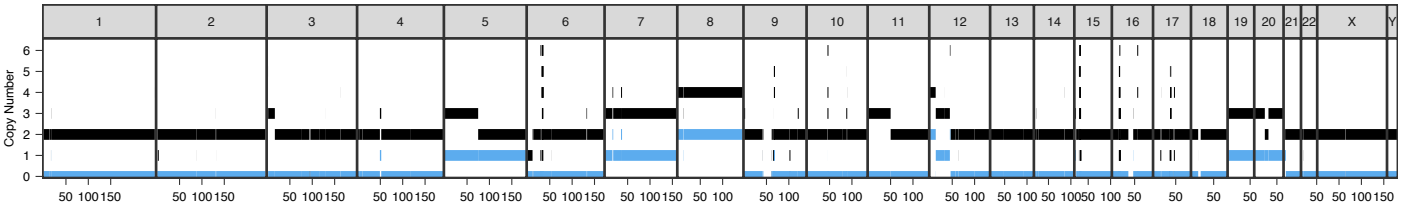
Extended Data Figure 9. Representative examples of the copy number profiles observed in samples with H3K27me3 loss. Total and minor copy number values are shown in black and blue, respectively.

Loss of H3K27me3

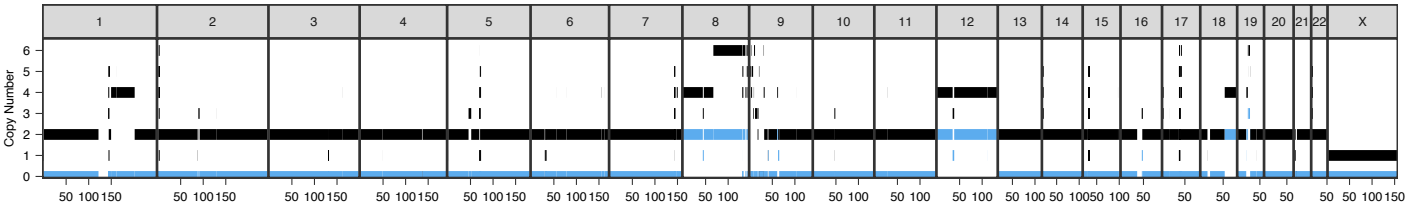
Pseudo−haploidization & chromosome 8 amplification (ROY_044)



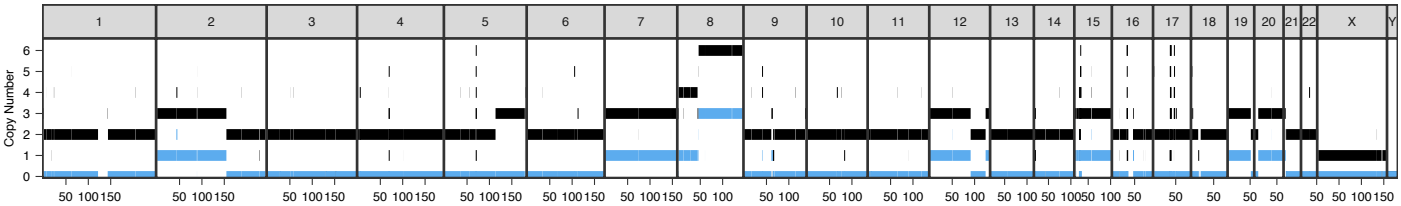
Pseudo−haploidization, WGD & chromosome 8 amplification (TOR_011)



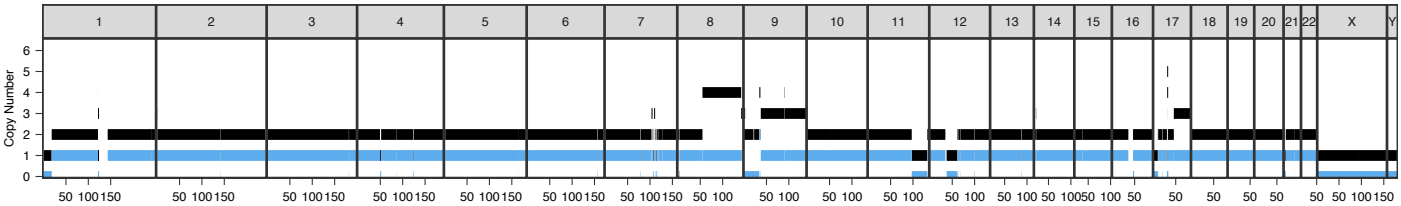
Pseudo−haploidization, WGD, chromosome 8 amplification and chromosome 8q gains (ROY_033)



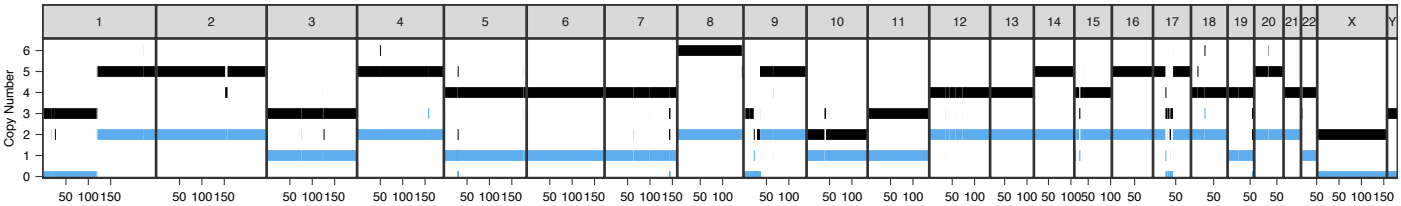
Pseudo−haploidization, WGD, chromosome 8 amplification and chromosome 8q gains (ROY_011)



Diploid & chromosome 8q gains (ROY_055)

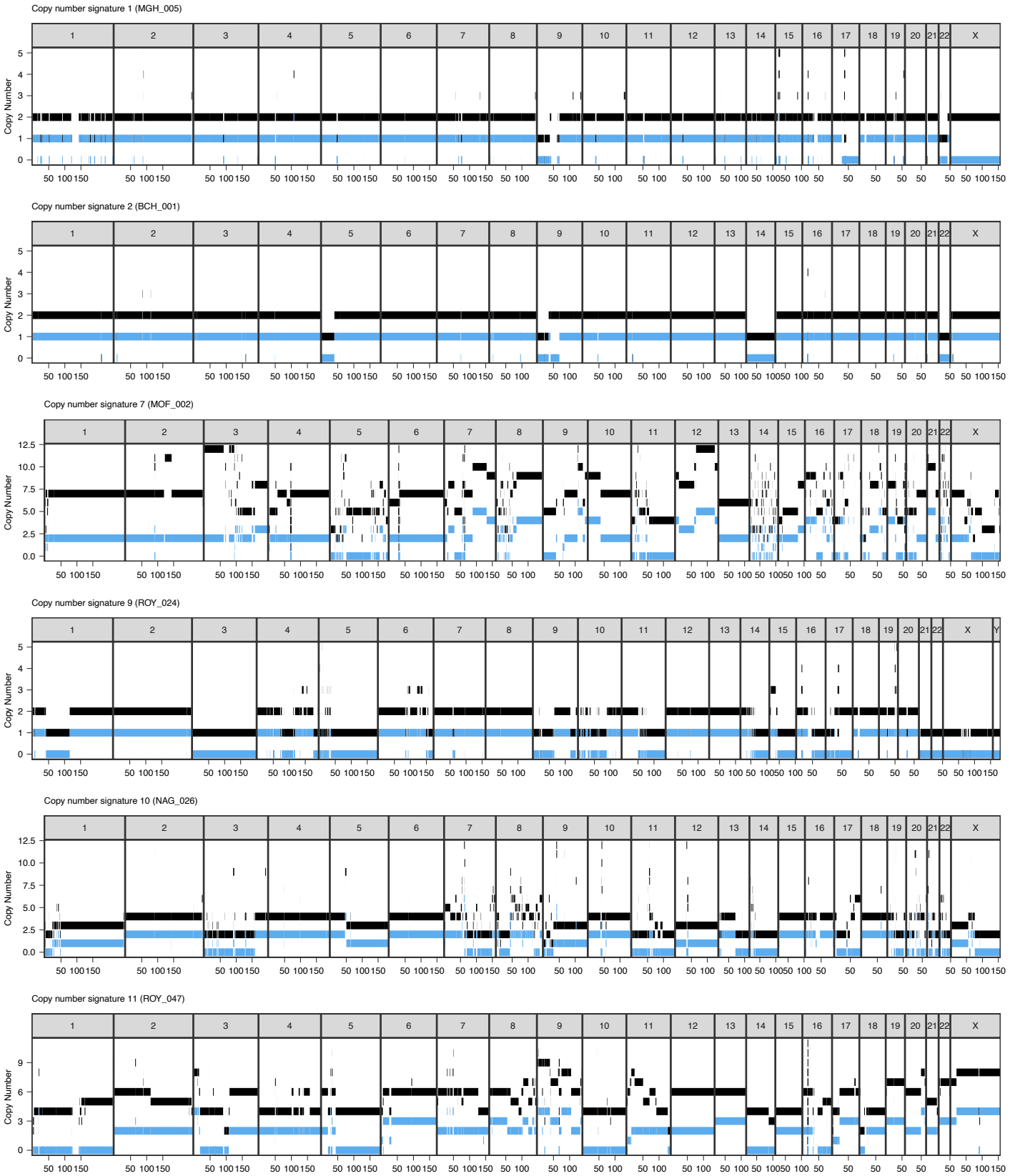


WGD & chromosome 8 gains (TOR_006)

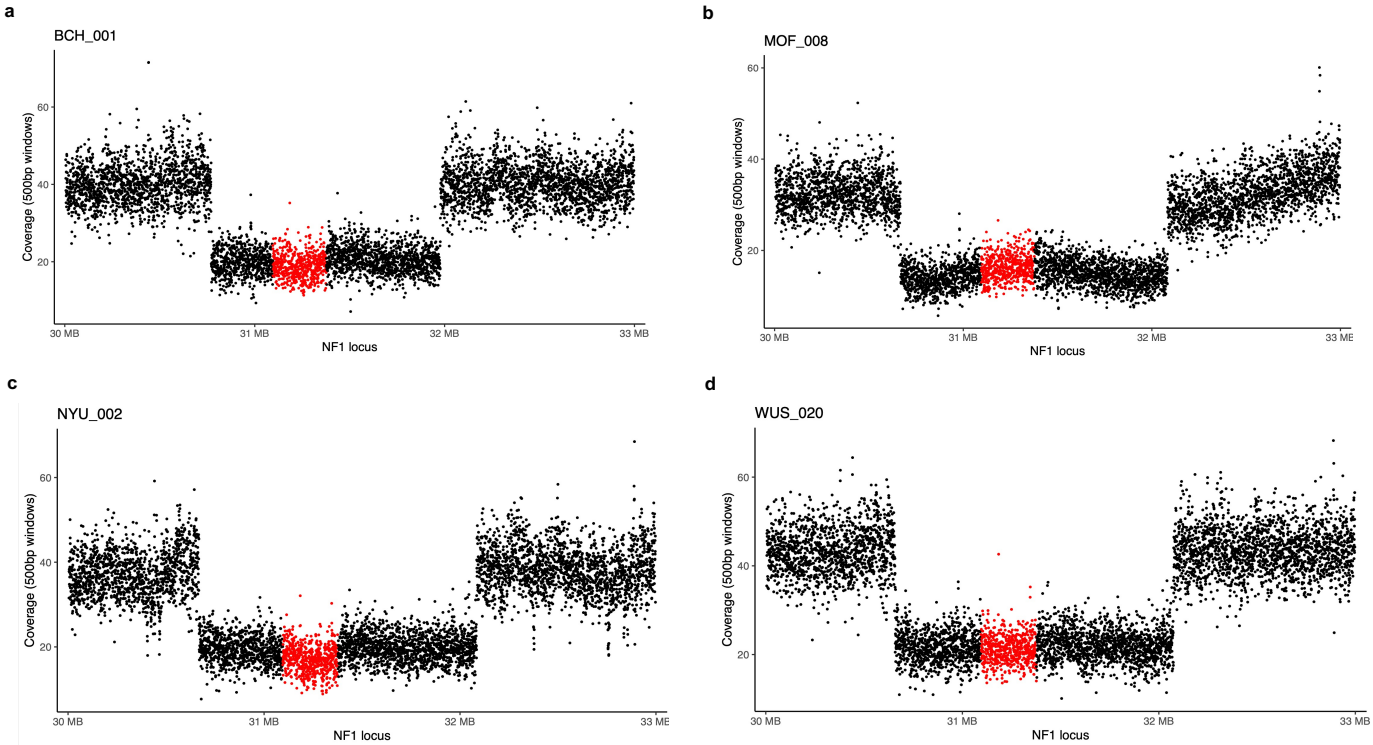


H3K27me3 retained

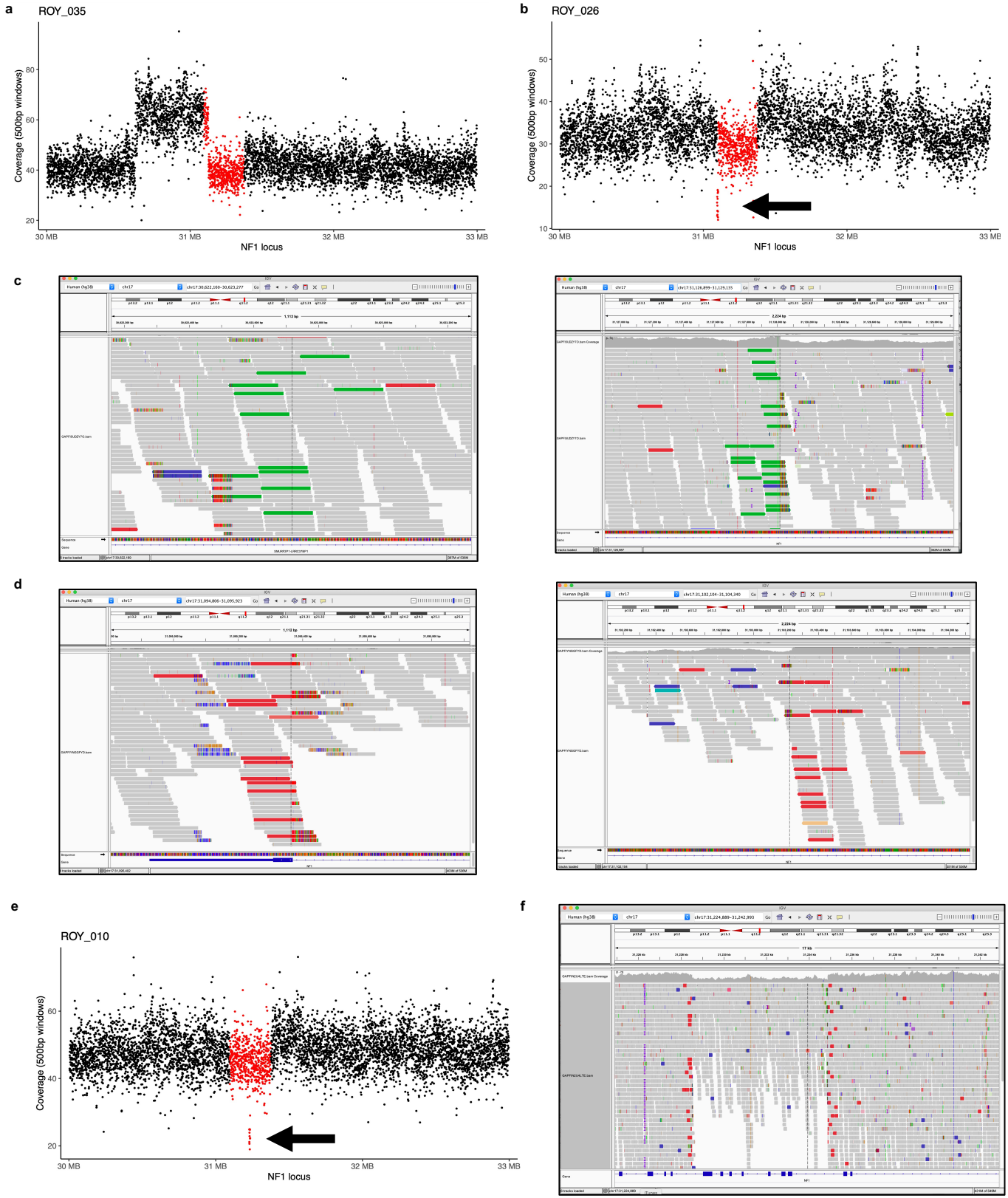
Extended Data Figure 10. Representative examples of the copy number profiles observed in samples with H3K27me3 retained. Total and minor copy number values are shown in black and blue, respectively.



Supplementary Figure 1. Analysis of germline *NF1* microdeletions. Sequencing coverage calculated at 500 bp windows at the *NF1* locus for individuals BCH_001 (**a**), MOF_008 (**b**), NYU_002 (**c**), and WUS_020 (**d**). Dots corresponding to windows mapping to *NF1* are shown in red. Genomic coordinates are shown in megabases (MB) according to GRCh38.

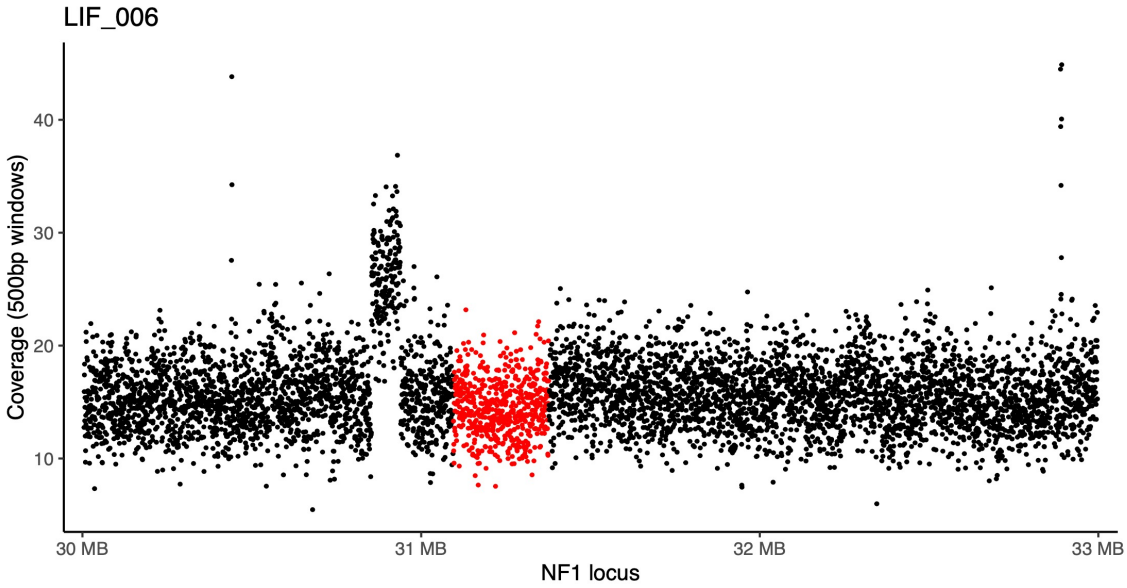


Supplementary Figure 2. Analysis of germline rearrangements in *NF1*. Sequencing coverage calculated at 500 bp windows at the *NF1* locus for individuals ROY_035 (**a**) and ROY_026 (**b**). **c,** Sequencing reads supporting the tandem duplication detected in ROY_035. **d,** Sequencing reads supporting the intragenic deletion detected in ROY_026. **e**, Sequencing coverage calculated at 500 bp windows at the *NF1* locus for individual ROY_010. **f**, Sequencing reads supporting the intragenic deletion detected in ROY_010. Dots corresponding to windows mapping to *NF1* in **a**, **b**, and **e**, are shown in red, and genomic coordinates are shown in megabases (MB) according to GRCh38. Gray = concordance and colours = discordance with GRCh38 reference in **c**, **d**, and **f**.

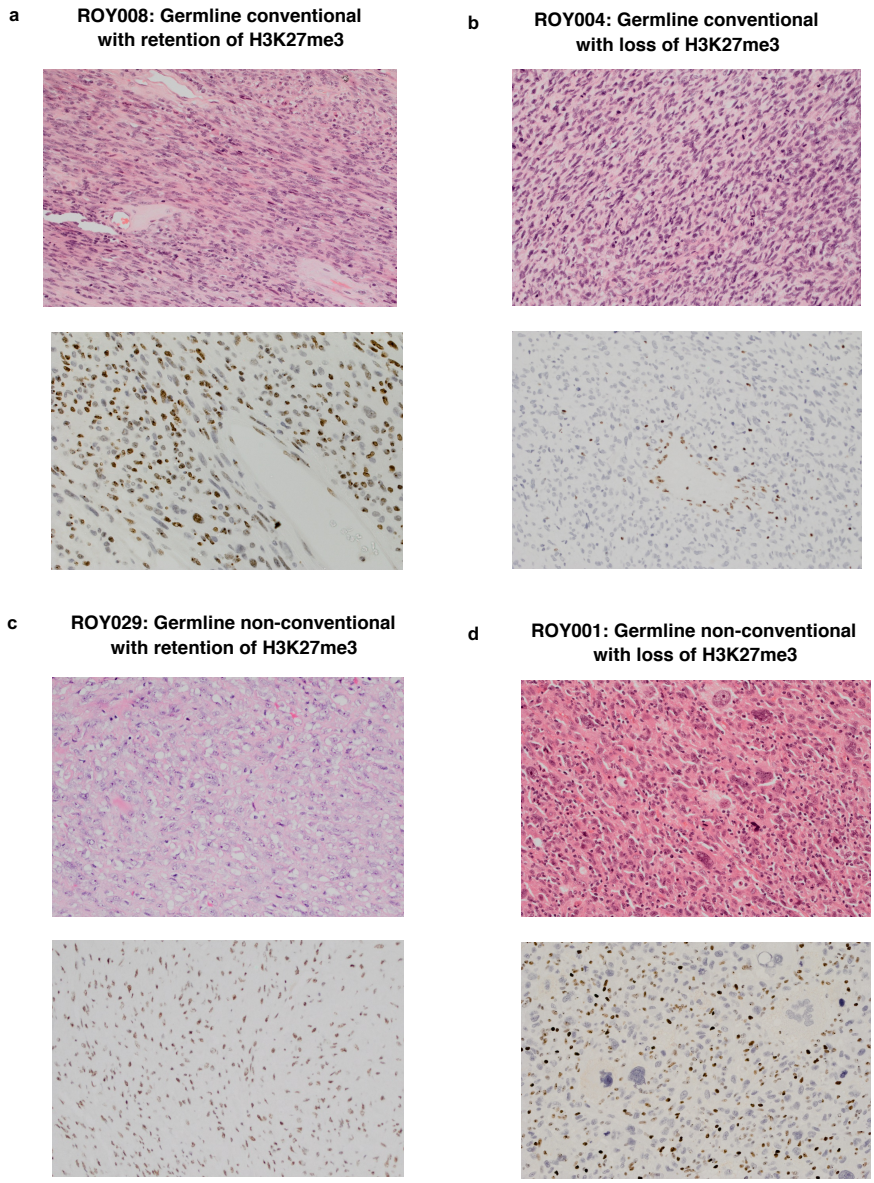


Supplementary Figure 3. Analysis of germline rearrangements upstream of *NF1*. a, Sequencing coverage calculated at 500 bp windows at the *NF1* locus for individual LIF_006. Dots corresponding to windows mapping to *NF1* are shown in red.

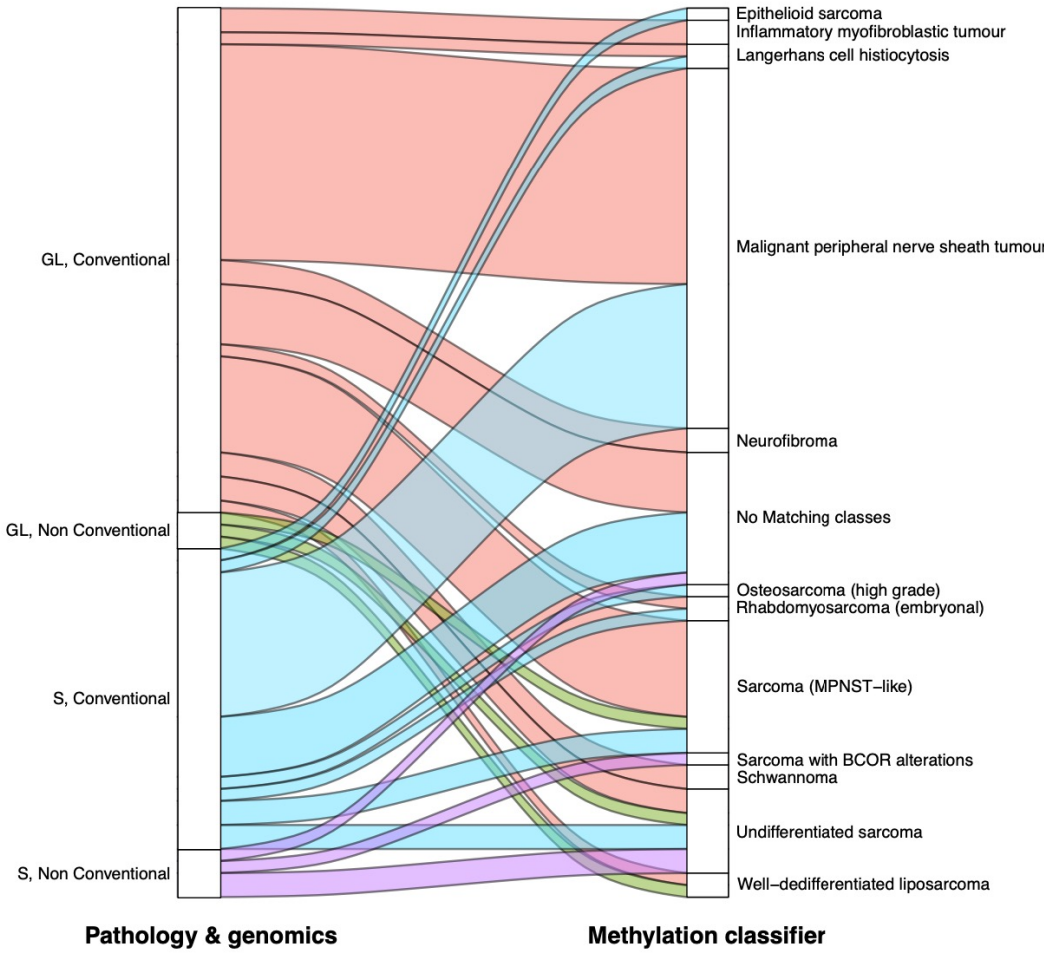
a



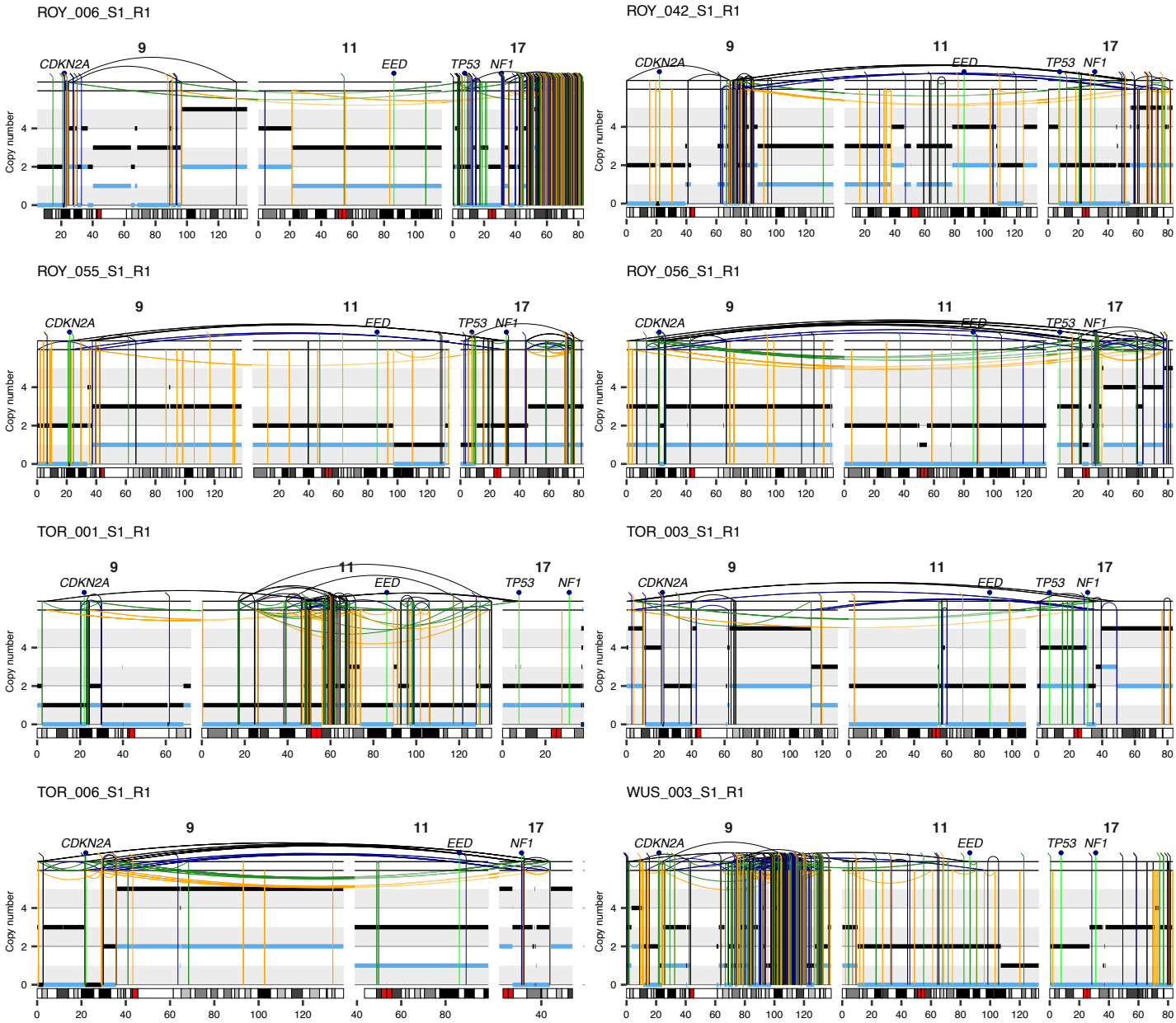
Supplementary Figure 4. Pathology classification of MPNSTs. H&E and H3K27me3 staining of four representative MPNSTs.



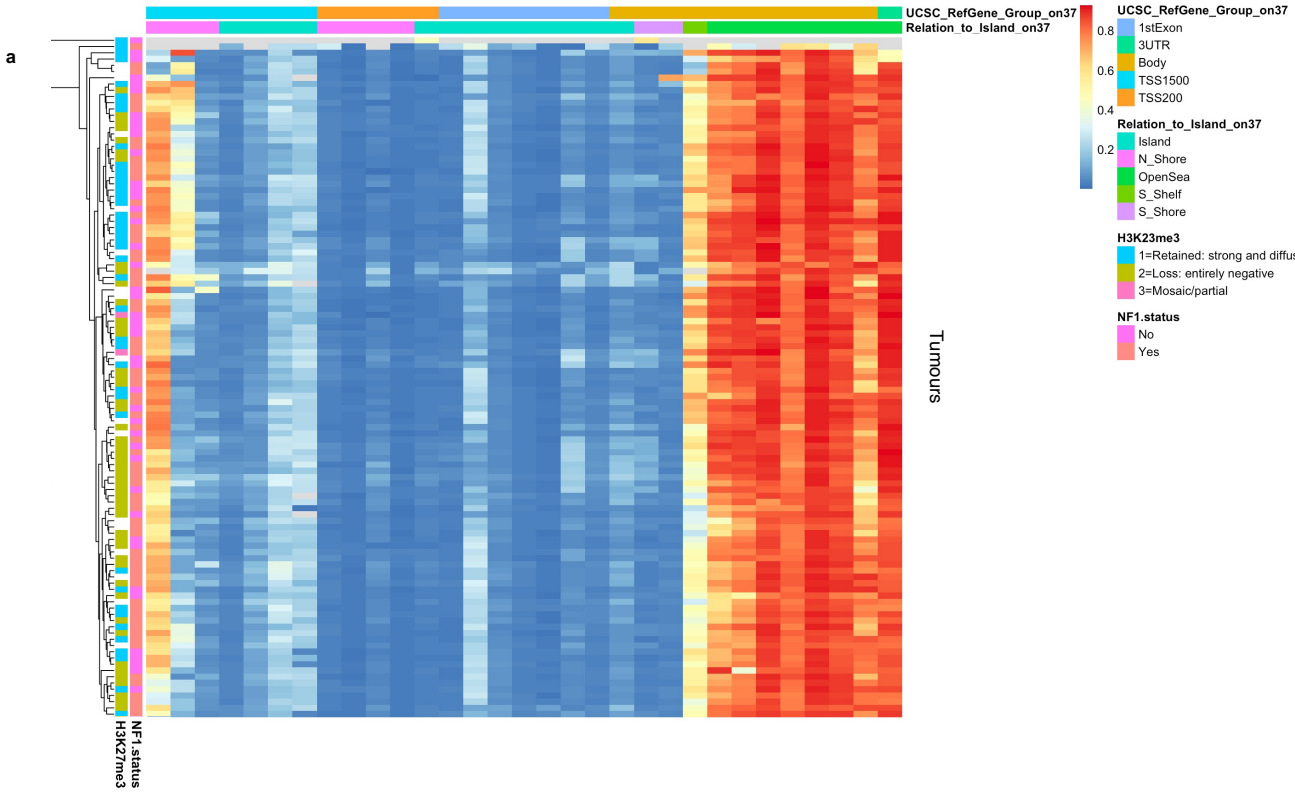
Supplementary Figure 5. Relationship between the methylation classifier results and the consensus classification established by the GeM consortium based on genomics and pathology information.

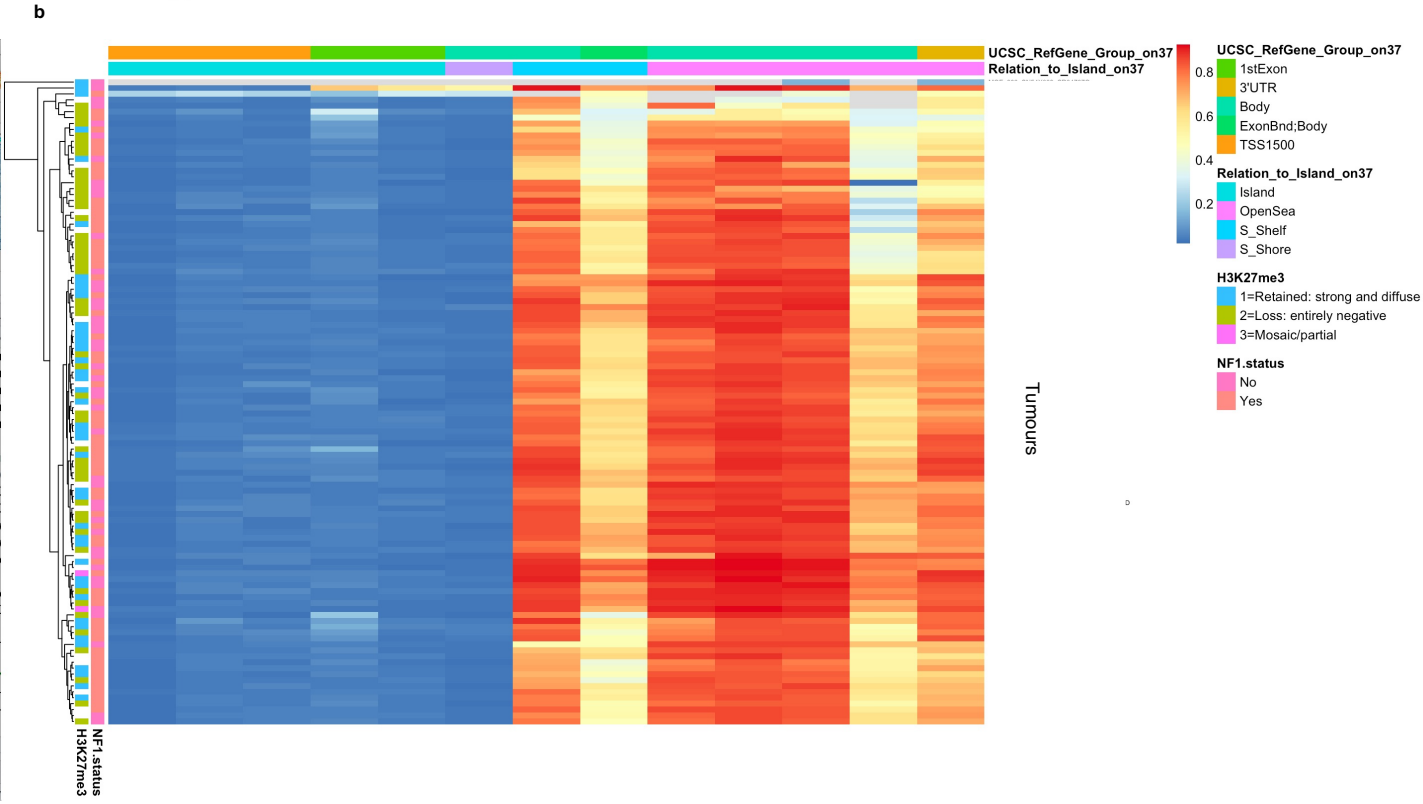


Supplementary Figure 6. Analysis of complex genome rearrangements involving the *CDKN2A, EED, TP53* and *NF1* loci. Rearrangement profiles of representative MPNSTs showing complex rearrangements of *CDKN2A, EED, TP53* and *NF1*. The position of *CDKN2A, EED, TP53* and *NF1* is shown at the top of each plot. Total and minor copy number values are shown in black and blue, respectively. The chromosome number is indicated on top of each rearrangement profile. Genomic coordinates are shown in Mbp according to GRCh38.

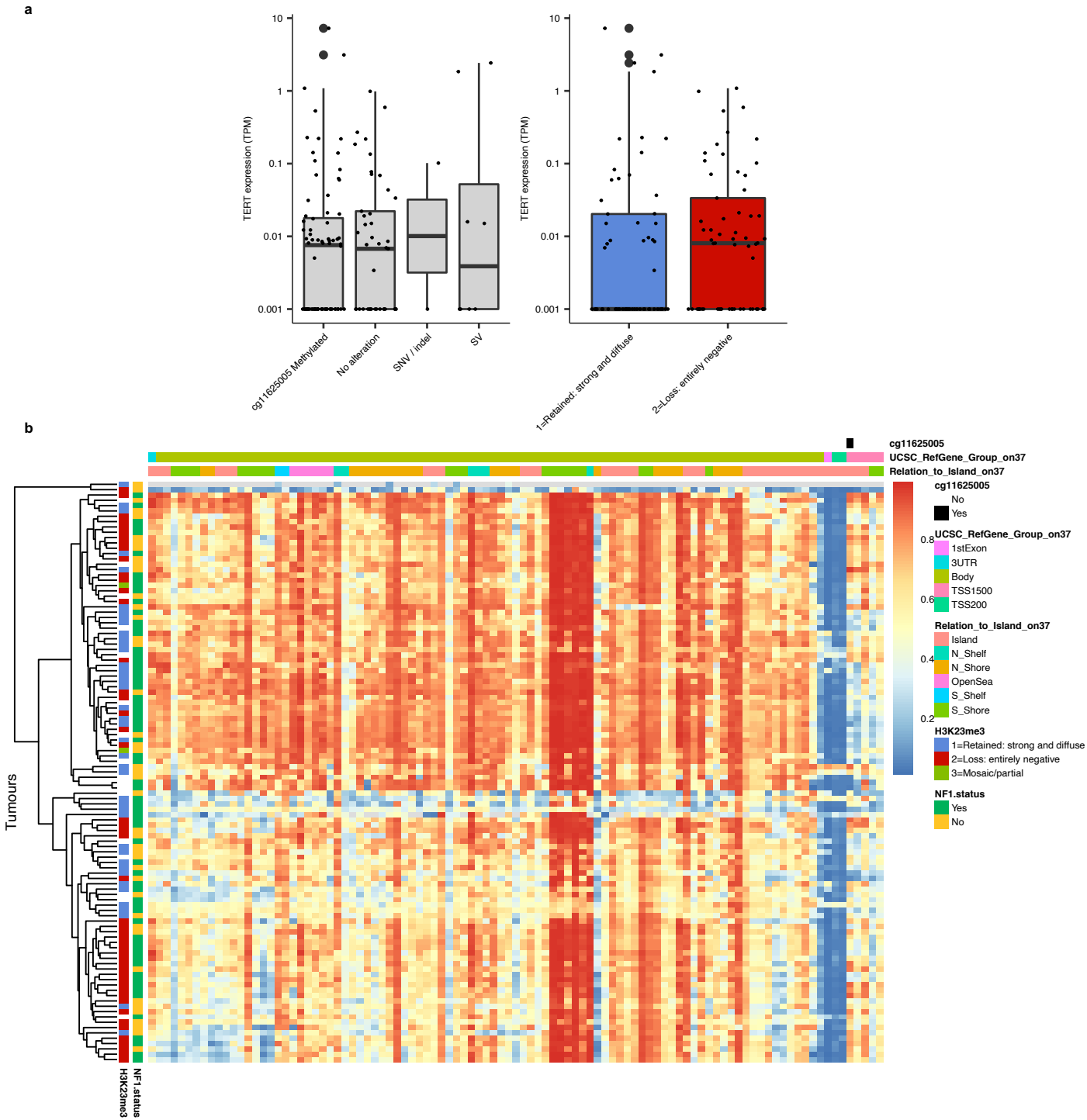


Supplementary Figure 7. Methylation profiling results for *EED* and *SUZ12*. Beta values for the probes mapping to EED (**a**) and SUZ12 (**b**) are shown.

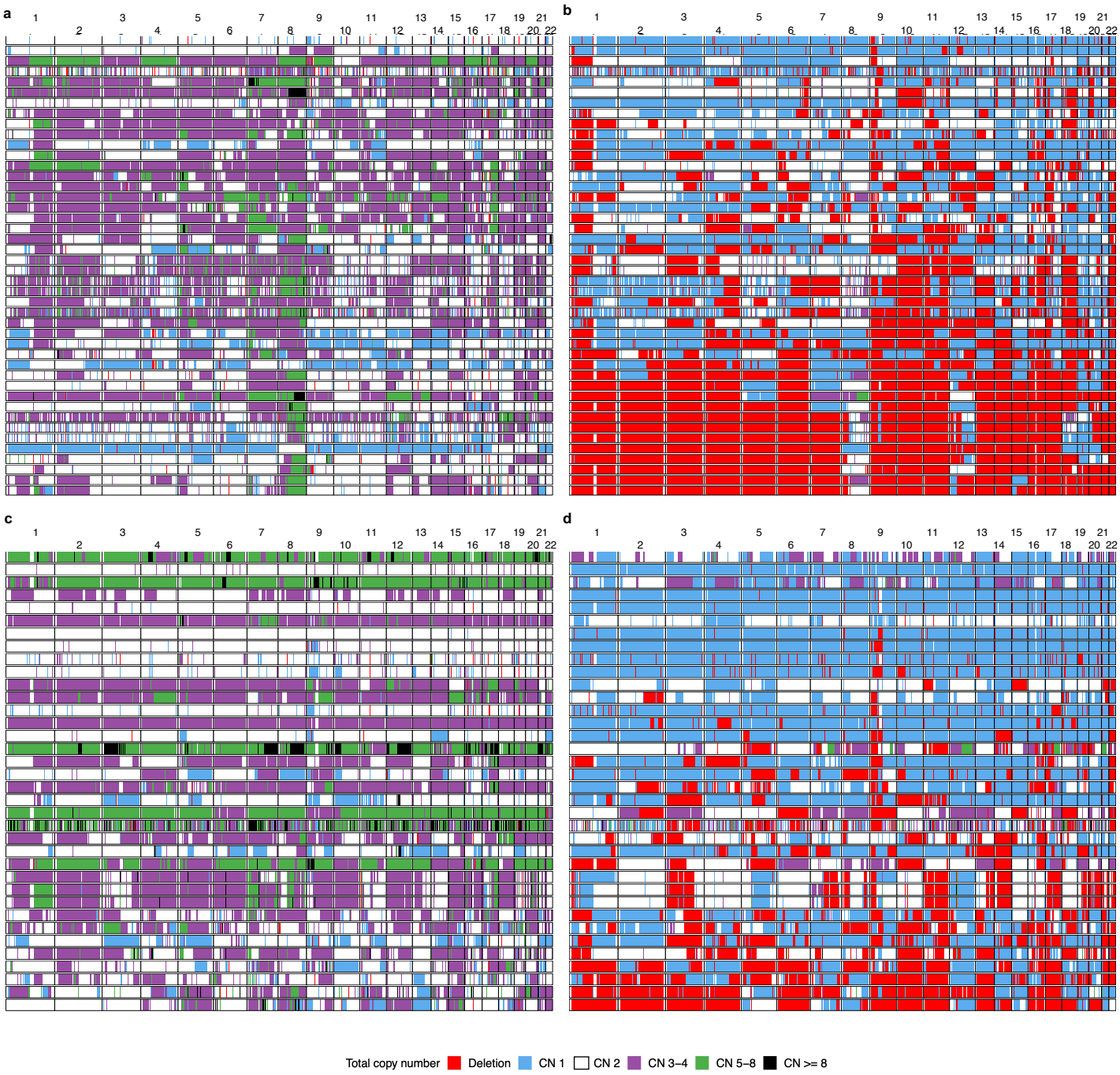




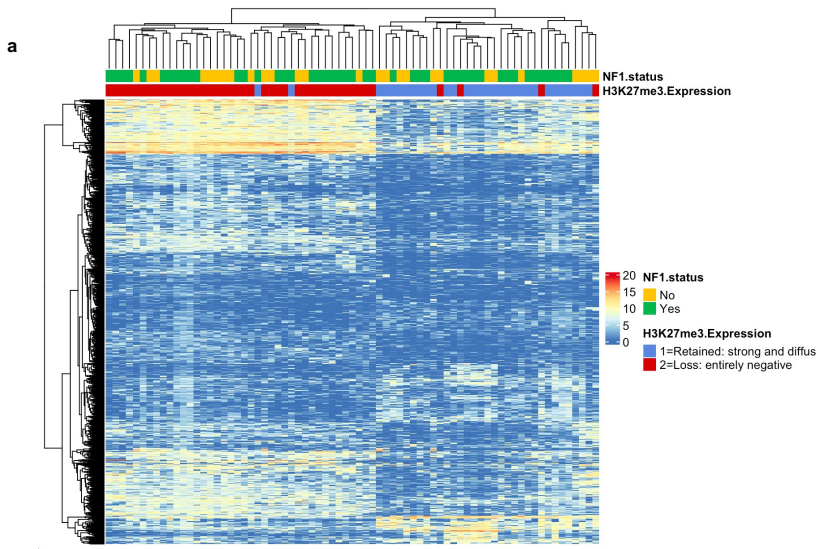
Supplementary Figure 8. Analysis of replicative immortality in MPNSTs. a, *TERT* is overexpressed in samples with *TERT* promoter mutations, SVs mapping to the *TERT* locus (+/- 100Kb), and methylation of the probe cg11625005. **b**, Beta values for the probes mapping to *TERT*. Each row corresponds to a sample and each column to a probe. The probe cg11625005 is highlighted in black.

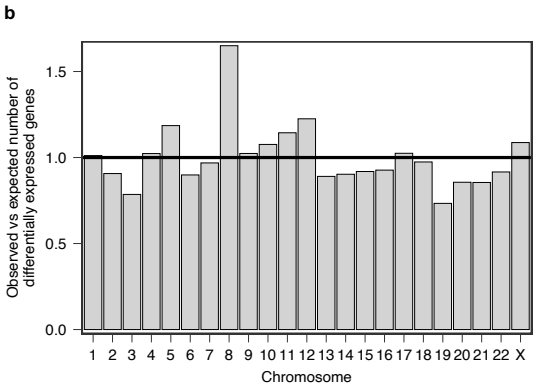


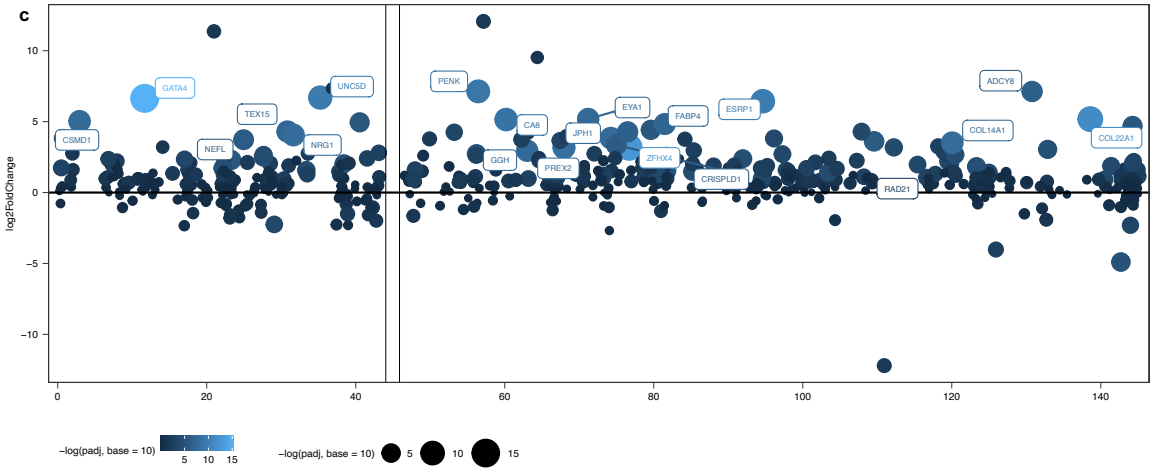
Supplementary Figure 9. Landscape of somatic copy number aberrations in MPNSTs. Total (**a**) and minor (**b**) copy number of the autosomal genome in MPNSTs with loss of H3K27me3. Total (**c**) and minor (**d**) copy number of the autosomal genome in MPNSTs with retention of H3K27me3.

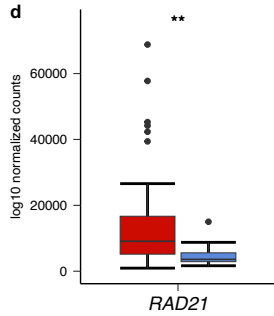


Supplementary Figure 10. Differential expression analysis. a, Heatmap showing the normalised expression counts for genes differentially expressed between MPNSTs with loss or retention of H3K27me3. **b**, Genes in chromosome 8 are enriched among the set of genes differentially expressed between MPNSTs with loss and retention of H3K27me3. **c**, Log2 fold change estimate for chromosome 8 genes for tumours with loss of H3K27me3 expression with respect to those with retained H3K27me3. Overall, chromosome 8 genes show higher expression in MPNSTs with loss of H3K27me3. The size and colour of the dots are proportional to the adjusted *P* value calculated in the differential expression analysis. Genomic coordinates are shown in megabases according to GRCh38. **d**, Differential *RAD21* expression between MPNSTs with loss or retention of H3K27me3. $^{**}P < 0.01$.

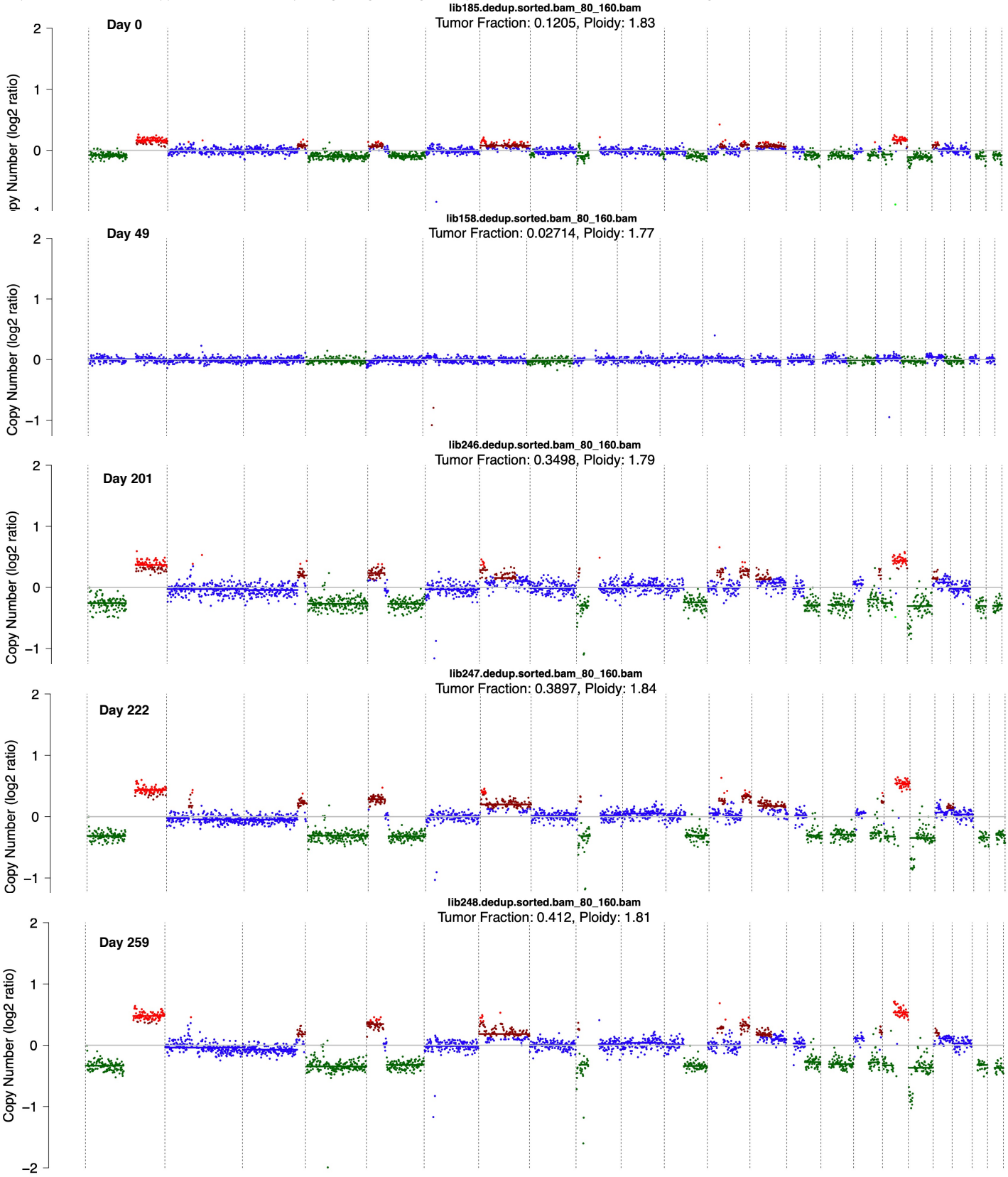




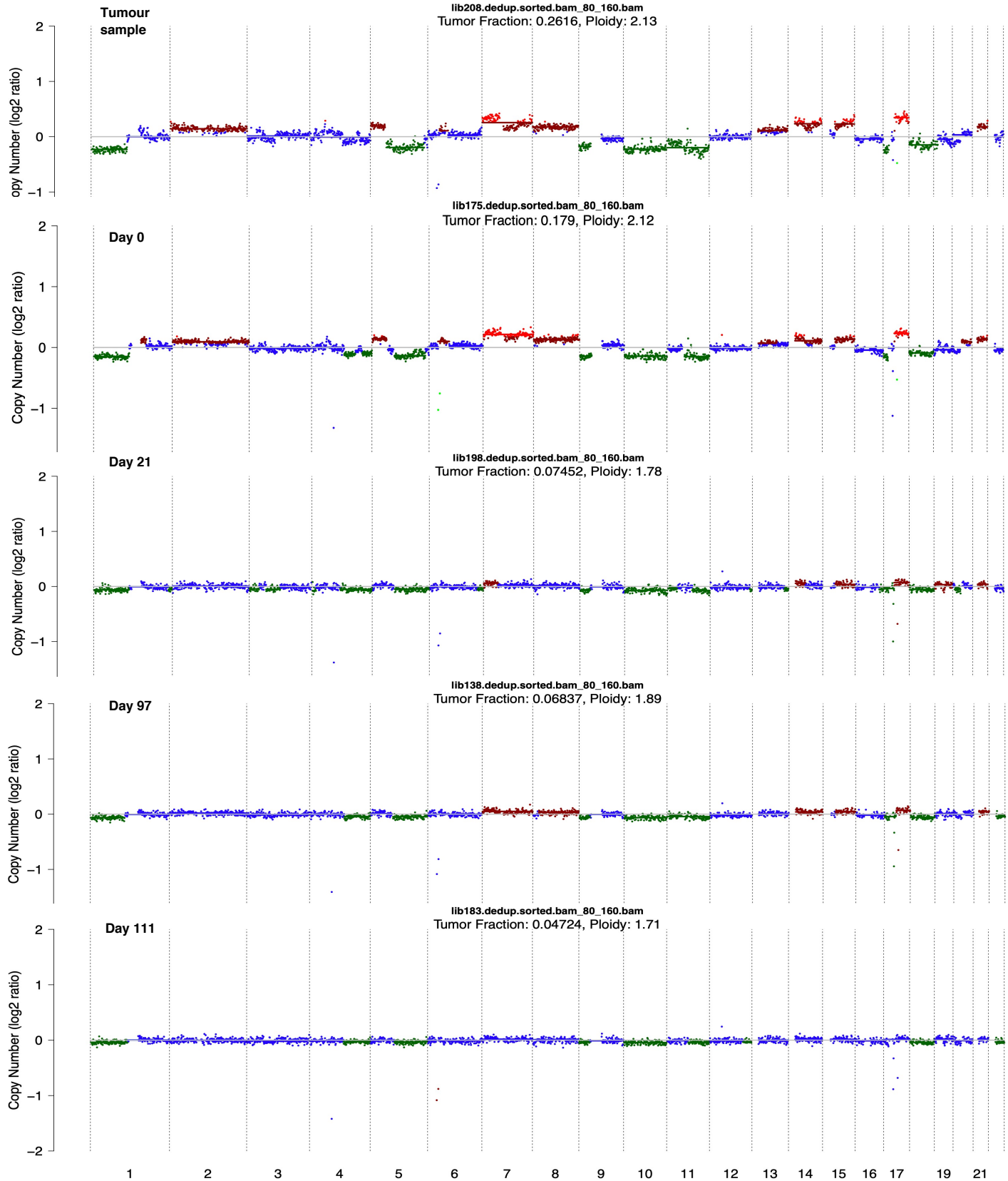




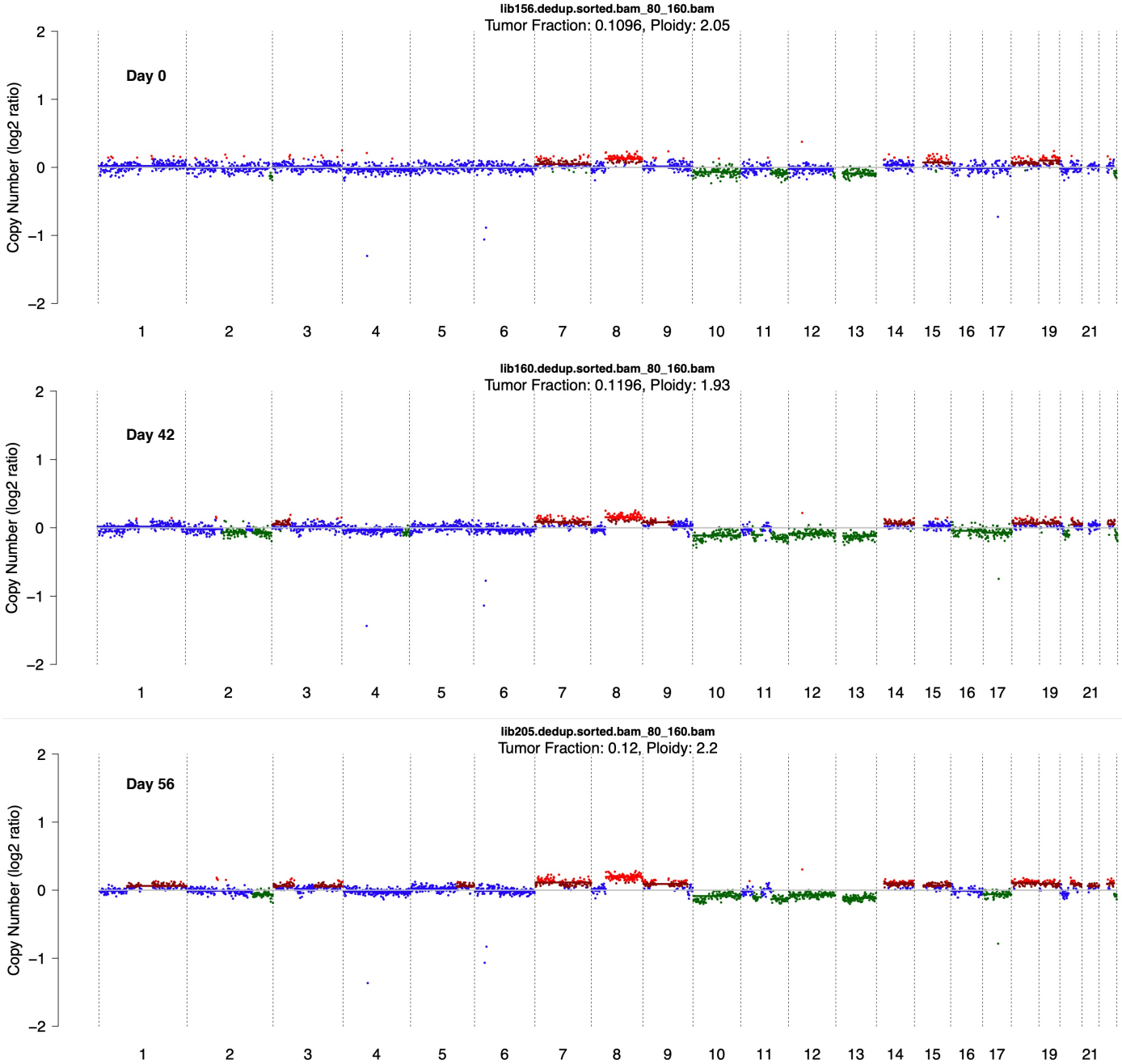
Supplementary Figure 11. Copy number profiles detected using ultra-low-pass whole-genome sequencing of cell-free DNA. Genome-wide copy number profiles computed using ichorCNA for longitudinal cell-free DNA samples from a patient with MPNST (sar102). The day of collection after clinical diagnosis is indicated in each case. The colour of each dot represents the estimated copy number for the corresponding 1-megabase region. Amplifications are shown in red and deletions in green.



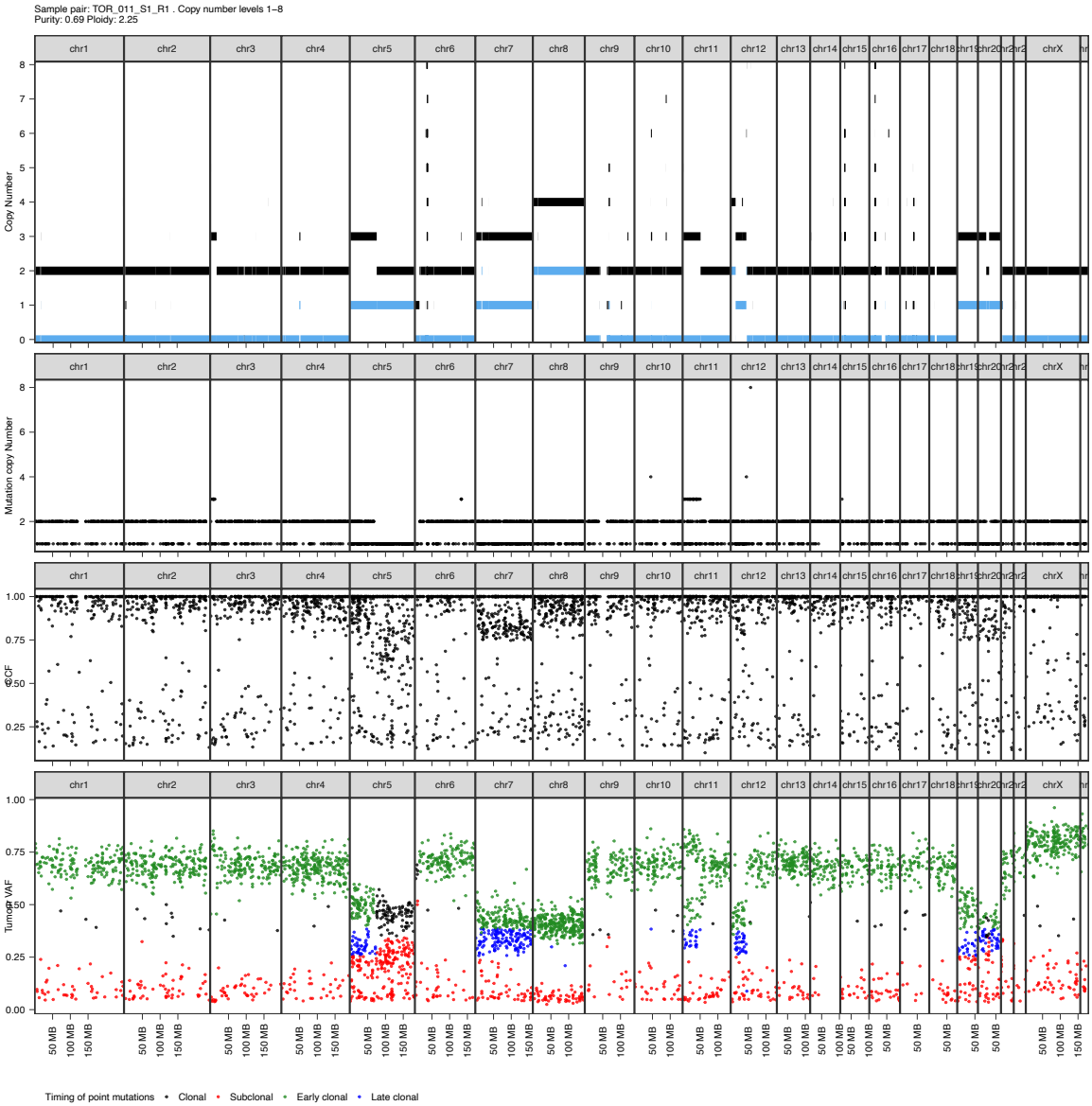
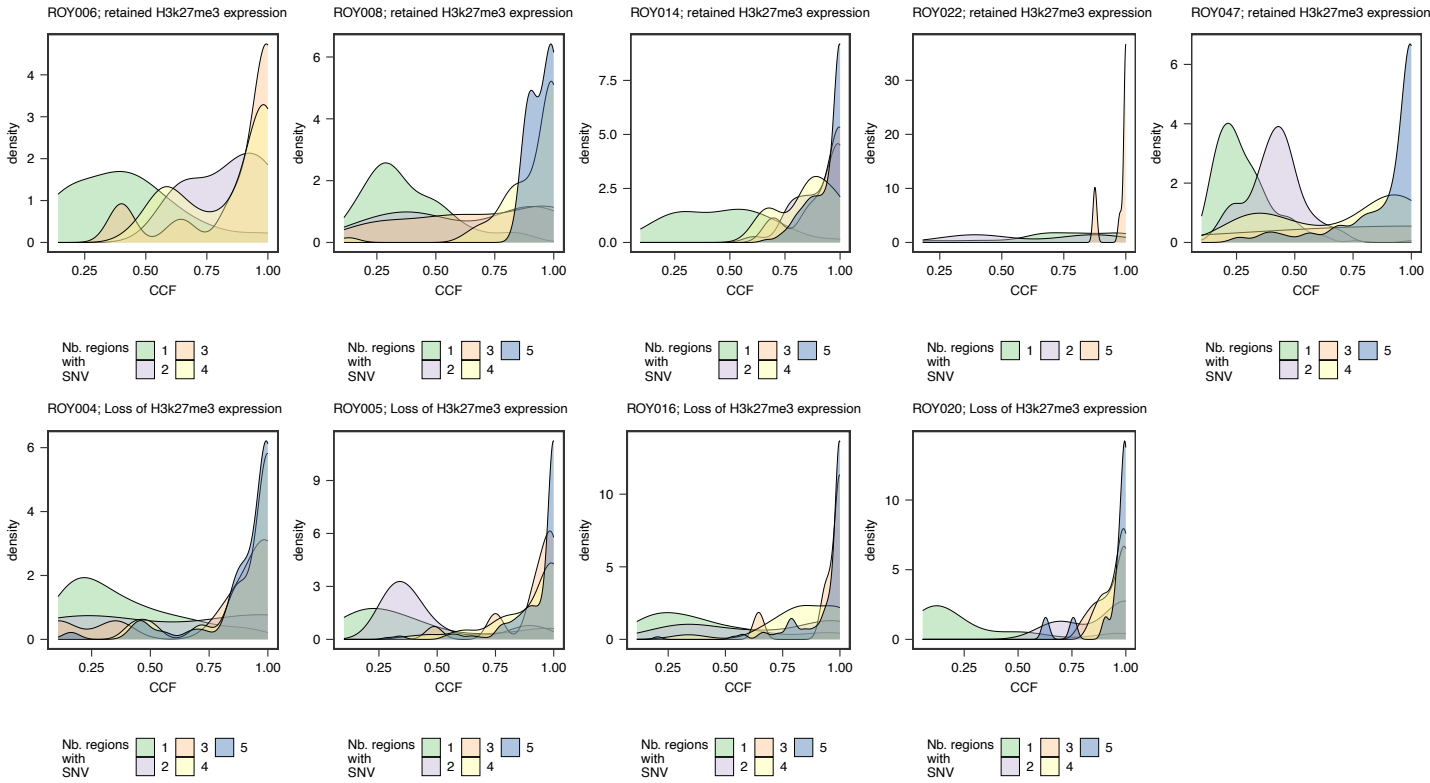
Supplementary Figure 12. Copy number profiles detected using ultra-low-pass whole-genome sequencing of cell-free DNA. Genome-wide copy number profiles computed using ichorCNA for longitudinal cell-free DNA samples from a patient with MPNST (sar081). The day of collection after clinical diagnosis is indicated in each case. The copy number profile for the tumour sample computed using the same computational and experimental methodology used for cell-free DNA analysis is shown at the top. The colour of each dot represents the estimated copy number for the corresponding 1-megabase region. Amplifications are shown in red and deletions in green.



Supplementary Figure 13. Copy number profiles detected using ultra-low-pass whole-genome sequencing of cell-free DNA. Genome-wide copy number profiles computed using ichorCNA for longitudinal cell-free DNA samples from a patient with MPNST (sar082). The day of collection after clinical diagnosis is indicated in each case. The colour of each dot represents the estimated copy number for the corresponding 1-megabase region. Amplifications are shown in red and deletions in green.



Supplementary Figure 14. Analysis of the timing of whole genome doubling in MPNSTs with loss of H3K27me3. a, Representative example of the somatic copy number profile of an MPNST (TOR_011) with loss of H3K27me3. Total and minor copy number values are shown in black and blue, respectively. **b**, Mutation copy number analysis for SNVs. Each dot represents an SNV and the y axis indicates the number of chromosomal copies in which each mutation is detected in the tumour sample. c, Cancer Cell Fraction (CCF) estimates for each SNV. d, Variant allele fraction (VAF) values in the tumour sample for each SNV coloured according to mutation timing. The large number of clonal SNVs detected in two chromosomal copies in regions with total copy number of 2 and minor copy number of 0 (that is, acquired before whole genome doubling) indicates that the whole genome doubling event occurred late in the clonal evolution of the tumour. Genomic coordinates are shown in megabases according to GRCh38.



Supplementary Figure 15. Analysis of the clonality of SNVs detected in the multi-regional exome sequencing data from fresh-frozen samples. The cancer cell fraction (CCF) values for mutations detected in one or multiple regions are shown.