

# Genomic Patterns of Malignant Peripheral Nerve Sheath Tumor (MPNST) Evolution Correlate with Clinical Outcome and Are Detectable in Cell-Free DNA



Isidro Cortes-Ciriano<sup>1</sup>, Christopher D. Steele<sup>2</sup>, Katherine Piculell<sup>3</sup>, Alyaa Al-Ibraheemi<sup>4</sup>, Vanessa Eulo<sup>5</sup>, Marilyn M. Bui<sup>6</sup>, Aikaterini Chatzipli<sup>7</sup>, Brendan C. Dickson<sup>8,9</sup>, Dana C. Borcharding<sup>10</sup>, Andrew Feber<sup>11</sup>, Alon Galor<sup>12</sup>, Jesse Hart<sup>13</sup>, Kevin B. Jones<sup>14</sup>, Justin T. Jordan<sup>15</sup>, Raymond H. Kim<sup>16,17</sup>, Daniel Lindsay<sup>18</sup>, Colin Miller<sup>1</sup>, Yoshihiro Nishida<sup>19</sup>, Paula Z. Proszek<sup>11</sup>, Jonathan Serrano<sup>20</sup>, R. Taylor Sundby<sup>21</sup>, Jeffrey J. Szymanski<sup>22</sup>, Nicole J. Ullrich<sup>23</sup>, David Viskochil<sup>24</sup>, Xia Wang<sup>25</sup>, Matija Snuderl<sup>20</sup>, Peter J. Park<sup>7</sup>, Adrienne M. Flanagan<sup>2,18</sup>, Angela C. Hirbe<sup>10</sup>, Nischalan Pillay<sup>2,18</sup>, David T. Miller<sup>3</sup>; for the Genomics of MPNST (GeM) Consortium

## ABSTRACT

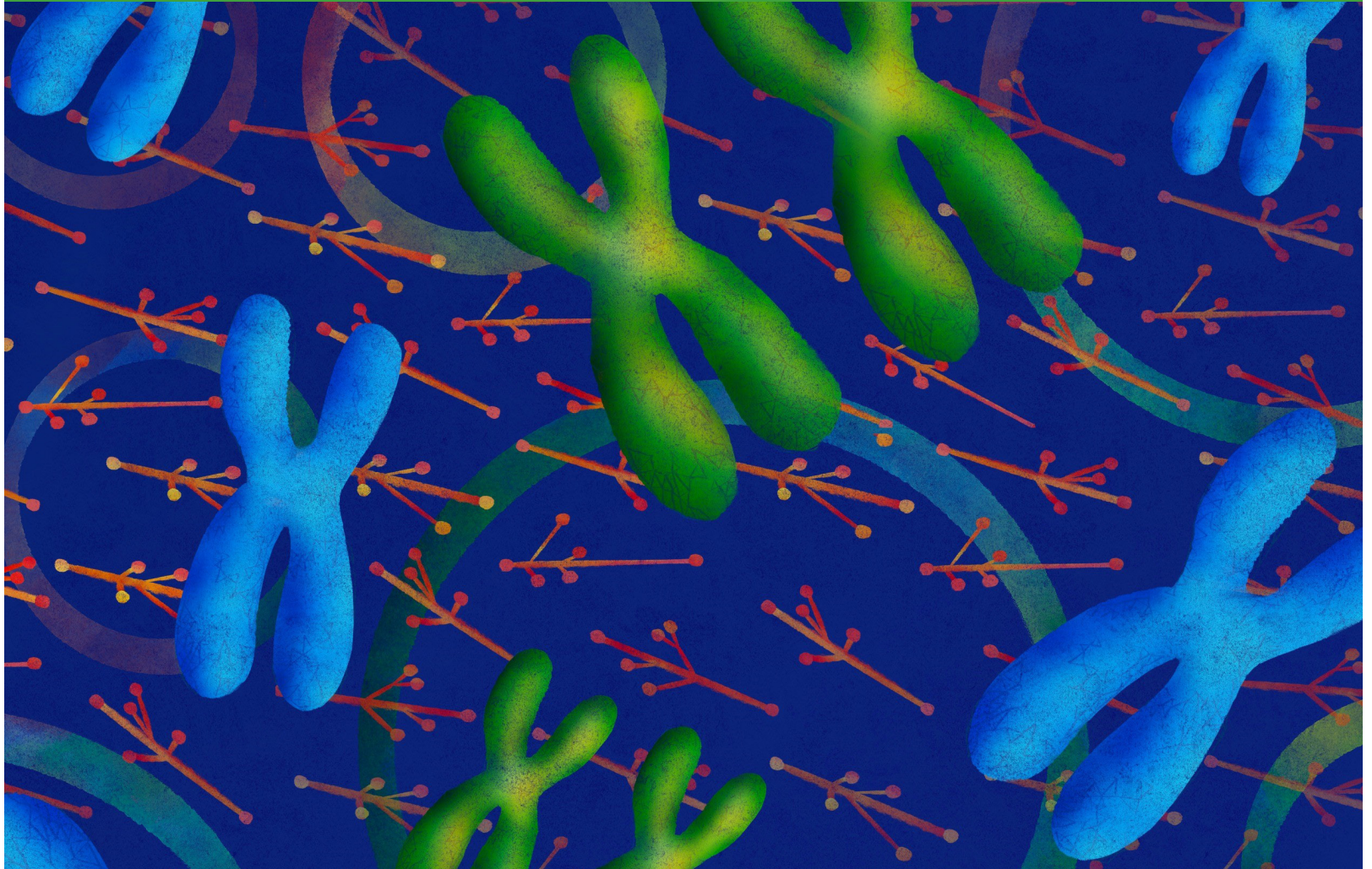
Malignant peripheral nerve sheath tumor (MPNST), an aggressive soft-tissue sarcoma, occurs in people with neurofibromatosis type 1 (NF1) and sporadically. Whole-genome and multiregional exome sequencing, transcriptomic, and methylation profiling of 95 tumor samples revealed the order of genomic events in tumor evolution. Following biallelic inactivation of *NF1*, loss of *CDKN2A* or *TP53* with or without inactivation of polycomb repressive complex 2 (PRC2) leads to extensive somatic copy-number aberrations (SCNA). Distinct pathways of tumor evolution are associated with inactivation of PRC2 genes and H3K27 trimethylation (H3K27me3) status. Tumors with H3K27me3 loss evolve through extensive chromosomal losses followed by whole-genome doubling and chromosome 8 amplification, and show lower levels of immune cell infiltration. Retention of H3K27me3 leads to extensive genomic instability, but an immune cell-rich phenotype. Specific SCNAs detected in both tumor samples and cell-free DNA (cfDNA) act as a surrogate for H3K27me3 loss and immune infiltration, and predict prognosis.

**SIGNIFICANCE:** MPNST is the most common cause of death and morbidity for individuals with NF1, a relatively common tumor predisposition syndrome. Our results suggest that somatic copy-number and methylation profiling of tumor or cfDNA could serve as a biomarker for early diagnosis and to stratify patients into prognostic and treatment-related subgroups.

<sup>1</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom. <sup>2</sup>Research Department of Pathology, University College London Cancer Institute, Bloomsbury, London, United Kingdom. <sup>3</sup>Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts. <sup>4</sup>Department of Pathology, Boston Children's Hospital, Boston, Massachusetts. <sup>5</sup>Division of Oncology, Department of Internal Medicine, University of Alabama at Birmingham, Birmingham, Alabama. <sup>6</sup>Department of Pathology, Moffitt Cancer Center & Research Institute, Tampa, Florida. <sup>7</sup>Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts. <sup>8</sup>Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. <sup>9</sup>Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada. <sup>10</sup>Division of Oncology, Departments of Internal Medicine and Pediatrics, Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri. <sup>11</sup>Clinical Genomics Translational Research, Institute of Cancer Research, Royal

Marsden NHS Foundation Trust, London, United Kingdom. <sup>12</sup>Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom. <sup>13</sup>Department of Pathology, Lifespan Laboratories, Rhode Island Hospital, Providence, Rhode Island. <sup>14</sup>Departments of Orthopaedics and Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. <sup>15</sup>Pappas Center for Neuro-oncology, Massachusetts General Hospital, Boston, Massachusetts. <sup>16</sup>Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, Sinai Health System, Toronto, Ontario, Canada. <sup>17</sup>Hospital for Sick Children, University of Toronto, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. <sup>18</sup>Department of Histopathology, Royal National Orthopaedic Hospital, NHS Trust, Middlesex, United Kingdom. <sup>19</sup>Department of Rehabilitation Medicine, Nagoya University Hospital, Nagoya, Aichi, Japan. <sup>20</sup>Department of Pathology, New York University Langone Health, Perlmutter Cancer Center, New York City, New York. <sup>21</sup>Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of





## INTRODUCTION

Malignant peripheral nerve sheath tumor (MPNST) is an aggressive soft-tissue sarcoma of peripheral nerves that arises both sporadically and in 8% to 13% of individuals with neurofibromatosis type 1 (NF1; refs. 1, 2). NF1 is a common cancer predisposition disorder affecting 1:2,500 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 not only is 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 tumor types, including histologically benign plexiform neurofibromas (PN). 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 tumors 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

Health, Bethesda, Maryland. <sup>22</sup>Division of Cancer Biology, Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri. <sup>23</sup>Department of Neurology, Boston Children's Hospital, Boston, Massachusetts. <sup>24</sup>Division of Medical Genetics, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah. <sup>25</sup>GeneHome, Department of Individualized Cancer Management, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida.

**Note:** I. Cortes-Ciriano, A.M. Flanagan, A.C. Hirbe, N. Pillay, and D.T. Miller contributed equally to leading the consortium.

**Corresponding Author:** David T. Miller, Boston Children's Hospital, Boston, MA 02115. Phone: 617-355-8221; E-mail: david.miller2@childrens.harvard.edu  
Cancer Discov 2023;13:654-71

**doi:** 10.1158/2159-8290.CD-22-0786

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2023 The Authors; Published by the American Association for Cancer Research



mutations in *SUZ12*, *EED*, or *EZH2* (10, 12–14). MPNSTs with PRC2 inactivation undergo loss of trimethylation at lysine 27 of histone H3 (H3K27me<sub>3</sub>; ref. 15), which has been associated with a worse prognosis (16, 17). However, these events are not found across all MPNSTs, and variability in histologic features adds to the challenge of accurately classifying these tumors, thereby limiting clinical decision-making and therapeutic development.

Previous genomic analyses of MPNSTs were based on small sample size studies and limited pathology characterization (12, 13, 18–20). 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 analyzing a large collection of MPNSTs with detailed clinical and pathologic characterization at a multiomic level (21). Because H3K27me<sub>3</sub> 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

### Tumor Sample Collection, Characterization, and Stratification

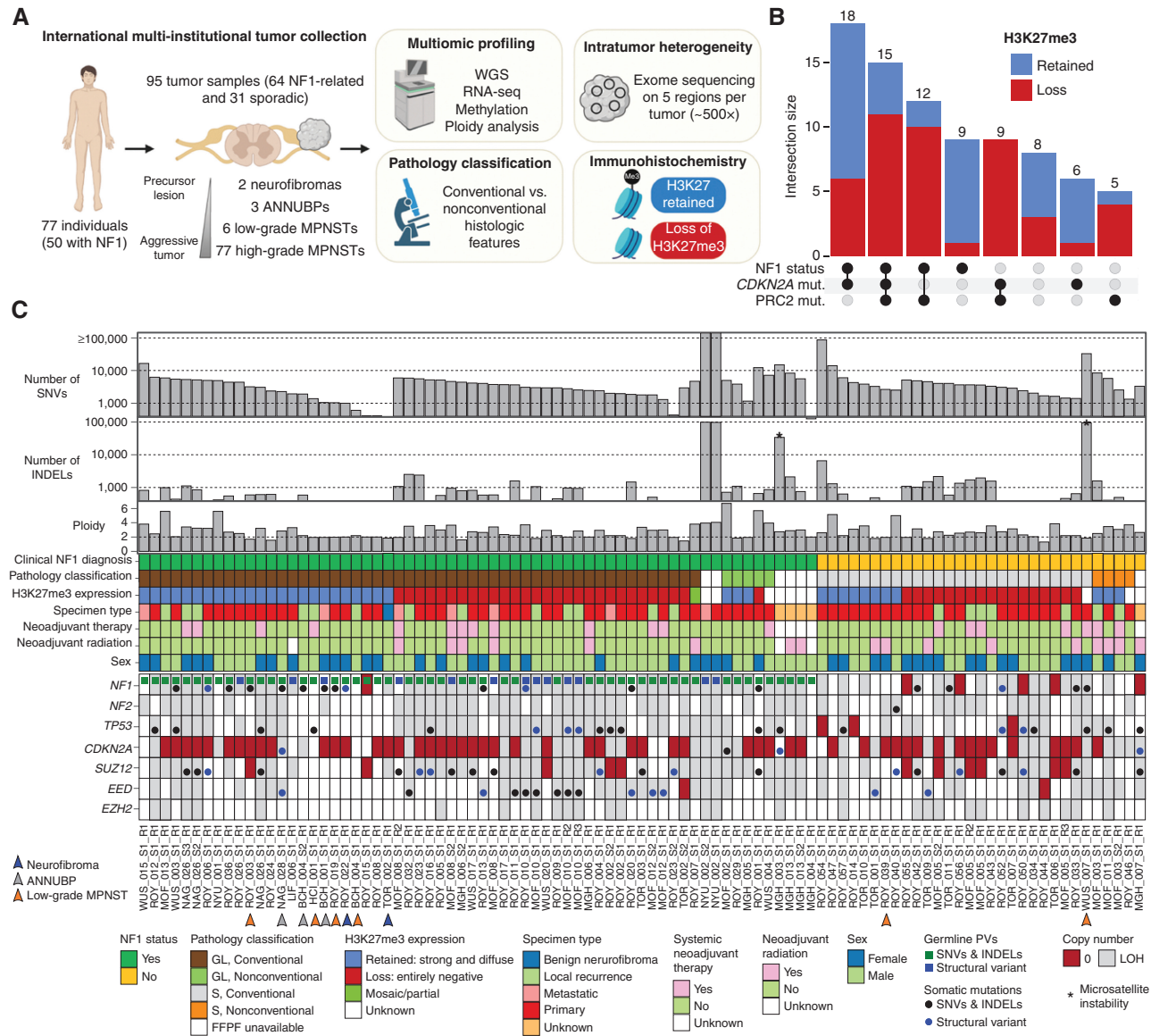
Through international multi-institutional collaboration, we collected and analyzed 95 samples from 90 fresh-frozen tumors (61 *NF1*-related; 29 sporadic) and matched blood samples from 77 individuals (50 with an *NF1* clinical phenotype and 27 sporadic) using whole-genome sequencing (WGS), whole-transcriptome sequencing [RNA sequencing (RNA-seq)], and whole-genome DNA methylation arrays (Fig. 1A; Supplementary Table S1). This collection, which is approximately 10 times larger than any prior WGS or multiomic studies of MPNSTs (12, 13), includes 72 high-grade MPNSTs, 6 low-grade MPNSTs, 3 atypical neurofibromatous neoplasms of uncertain biological potential (ANNUBP), 2 additional tumors (described below, and designated as neurofibromas in Fig. 1) and 7 cases without formalin-fixed, paraffin-embedded (FFPE) slides available for review. Cases were selected based on a pathologic diagnosis of MPNST or precursor lesion (9), availability of clinical information, H3K27me<sub>3</sub> immunoreactivity data, and availability of material for multiomic analysis (Methods). Both neurofibromas retained in the cohort were *NF1*-related cases with conventional histologic features (Fig. 1A). One case was a precursor lesion from an MPNST subject in this cohort; the other was a neurofibroma with atypia originally diagnosed as a low-grade MPNST by the providing institution. Regarding the six low-grade MPNSTs (4 *NF1*-related, 2 sporadic), all cases had conventional histologic features and H3K27me<sub>3</sub> immunoreactivity data available (5

of 6 cases had retained H3K27me<sub>3</sub>). Using the WGS data, we identified an *NF1* germline pathogenic variant in all 50 individuals with an *NF1* clinical phenotype [8 with pathogenic microdeletions and complex structural variants (SV) undetectable by exome sequencing and 2 with deep intronic mutations that generated cryptic exons; Supplementary Table S1 and Methods]. Next, we confirmed biallelic inactivation of *NF1* in 56/64 (88%) tumor samples arising within the tumors of individuals with *NF1*. *NF1* inactivation in 49 of the 56 (87.5%) tumors occurred through the loss of the wild-type allele, and biallelic inactivation was identifiable more often in tumors from people with *NF1* as compared with sporadic tumors, 10/31 (32%,  $P < 0.01$ , Fisher exact test; Supplementary Table S1). Four pathologists performed a consensus classification by assigning the tumors to 4 categories based on the presence or absence of germline *NF1* pathogenic variants (germline vs. sporadic) and conventional and nonconventional histopathologic features of MPNST (Supplementary Fig. S1A–S1D). In addition, tumors were classified according to the retention or loss of H3K27me<sub>3</sub> using IHC (ref. 9; H3K27me<sub>3</sub> retained vs. H3K27me<sub>3</sub> loss; Methods). Diagnosis using previously published machine learning classifiers of sarcomas based on genome-wide DNA methylation patterns (22) was only informative for 48 of 95 tumor samples (51%), highlighting the need for combining pathology and genomics data for accurate diagnosis (Supplementary Fig. S1E).

### Somatic Genomic Landscape of MPNSTs

To identify drivers of MPNST development, we searched for genes harboring recurrent loss- or gain-of-function mutations (Methods). We detected somatic biallelic inactivation of *CDKN2A* in 40/64 (63%) *NF1*-related MPNSTs and in 17 of 31 (55%) sporadic MPNSTs (Fig. 1B and C), primarily caused by nonrecurrent SVs and complex rearrangements of the *CDKN2A* locus, which in some tumor samples led to the concomitant inactivation of *NF1* and the loss of nearby genes (e.g., *MTAP* in 25% of tumors). We identified loss of H3K27me<sub>3</sub> in 45 of 82 (55%) of MPNSTs with IHC data available (Fig. 1B and C). *SUZ12* was the gene from the PRC2 complex with the highest frequency of biallelic inactivation (27/95 tumors with WGS data, 28%), followed by *EED* (16/95, 17%). Biallelic inactivation of *SUZ12* and *EED* were mutually exclusive events (Fig. 1C). Furthermore, the loss of H3K27me<sub>3</sub> was not associated with hypermethylation of these genes. Double-hit mutations were found in *TP53* in 20 of 95 (21%) tumors, and replicative immortality was primarily driven by *TERT* overexpression (Supplementary Fig. S2A and S2B). We did not detect recurrent gene fusions or genes recurrently mutated besides *NF1*, *CDKN2A*, PRC2 complex genes and *TP53* (Fig. 1C and Methods), suggesting that these are the main drivers of MPNST tumorigenesis.

We performed *de novo* mutational signature analysis of single-base substitutions (SBS) and small insertions and deletions (INDEL). Similar mutational processes were identified in MPNSTs irrespective of H3K27me<sub>3</sub> status (Supplementary Fig. S3A). Mutational signature SBS5, a clocklike signature found in human cancers (23) and histologically normal tissues (24), was the predominant signature in both groups, contributing a median of 58% of single-nucleotide variants (SNV). The burden of mutations and SVs was comparable between both



**Figure 1.** Genomic and pathology classification of MPNSTs. **A**, Overview of the experimental design and technologies utilized in this study. The GeM cohort consists of 95 tumor samples from 90 tumors (61 NF1-related and 29 sporadic), including 7 tumor samples without FFPE available for pathology review and confirmation of main tumor diagnosis. **B**, Relationship between NF1 status, biallelic inactivation of CDKN2A, and the PRC2 complex. The numbers on top of the bars indicate the total number of tumors in each group. **C**, Clinical information, histopathologic 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 biological potential; CN, copy number; GL, germline; INDEL, insertions and deletion; PV, pathogenic variant; RNA-seq, RNA sequencing; S, sporadic; SNV, single-nucleotide variants.

groups, with the exception of four hypermutated cases, two of which showed microsatellite instability (Fig. 1C; Supplementary Fig. S3B–S3E). SV signature analysis identified five distinct signatures previously detected in undifferentiated soft-tissue sarcomas and breast carcinomas (refs. 25, 26; Supplementary Fig. S3F and S3G), which had comparable activity in both groups (Supplementary Fig. S3A). The frequency of chromothripsis, which was detected in 47 of 95 (49%) of tumors, was also comparable between both groups (19/37 and 22/45 of tumors with retention and loss of H3K27me3, respectively;  $P > 0.05$ , Fisher exact test). Whole-genome doubling (WGD) events, which were validated using image cytometry analysis

(Methods), were detected in 22 of 37 (59%) of tumors with H3K27me3 retained and in 33 of 45 (73%) of tumors with H3K27me3 loss ( $P > 0.05$ , Fisher exact test). Together, these results indicate that MPNSTs are characterized 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 tumors stratified into two distinct



groups, largely corresponding to the retention and loss of H3K27me3 (Methods and Supplementary Fig. S4A). Cluster assignments were consistent across data types (Supplementary Fig. S4A). Classification by H3K27me3 immunoreactivity was associated with overall survival in individuals with NF1, but not in sporadic MPNSTs (Supplementary Fig. S4B and S4C). No differences in overall survival were observed when stratifying tumors on the basis of conventional or unconventional histology (Methods) by univariate or multivariate analysis accounting for age, sex, and tumor 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 tumor 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 and B; Supplementary Fig. S4D and S4E). 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 tumors with H3K27me3 loss, which is consistent with the role of PRC2 in neural development (ref. 13; Supplementary Fig. S4E–S4G). Additional data showing expression of immune signatures in MPNSTs are presented in Supplementary Fig. S5. Together, these results indicate that MPNSTs arising among individuals with NF1 stratify into two clinically and biologically distinct groups characterized 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 somatic copy-number alterations (SCNA) 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 of genome-wide LOH (Fig. 2C–E; Supplementary Fig. S6A–S6D and Methods). LOH of selected chromosomes, including 1p, 10, 11, 16, 17, and 22, was frequently detected in MPNSTs with loss of H3K27me3 (Fig. 2E). In one extreme H3K27me3 loss case (Fig. 2F), we observed near-haploidization characterized by genome-wide LOH excluding a few chromosomes with retention of heterozygosity, including chromosome 8, and a ploidy of 1.3. In most other cases, near-haploidization was followed by WGD (Fig. 2G). Copy-number signature analysis (27) identified 18 distinct signatures (Supplementary Fig. S7A) and confirmed enrichment of signatures related to LOH in

MPNSTs with loss of H3K27me3 (Fig. 2E; Supplementary Fig. S7B). In contrast, tumors with retained H3K27me3 were enriched in signatures associated with a diploid genome and genomic instability ( $P = 0.029$ , two-sided Mann-Whitney test; Supplementary Fig. S7B). Together, these results indicate that distinct copy-number patterns are associated with H3K27me3 status.

Our analysis of MPNST copy-number profiles offers a basis for comparison with other types of cancer. To this aim, we next compared MPNSTs to diverse cancer types from The Cancer Genome Atlas (TCGA) at the copy-number level. Gains in chromosome 8q have also been detected in a number of other solid tumors (28–30). Numerous cancer-related genes are located on chromosome 8q and have been implicated in cancer progression, including genes such as *C-MYC*, *RAD21*, *UBR5*, and *HEY1* (28, 31–33). Of the >700 protein-coding genes on chromosome 8, 199 genes are differentially expressed in the tumors with H3K27me3 loss, including *RAD21*, which has been linked with the mitigation of replication stress caused by the oncogenic *EWS-FLI1* fusion in Ewing sarcomas with chromosome 8q amplification (28), *UBR5*, and *HEY1* (Supplementary Table S2 and Methods). Near-haploidization has been observed across diverse cancer types, including undifferentiated soft-tissue sarcomas, adrenocortical carcinoma, gliomas, chondrosarcomas, and acute lymphoblastic leukemia, and is often linked with a poorer prognosis (25, 34–38). In pan-cancer analysis, MPNSTs with loss of H3K27me3 show the highest levels of genome-wide LOH, including near-haploidization (Supplementary Fig. S8A–S8C). However, concomitant retention of chromosome 8 heterozygosity and genome-wide LOH as observed in MPNSTs with loss of H3K27me3 (Fig. 2A) has not been linked with a distinct biological process previously. This new copy-number configuration provides a potential new avenue to investigate the evolutionary constraints leading to genome-wide LOH (39).

### Copy-Number Profiles Are Detected in Cell-Free DNA and Predict Patient Outcome

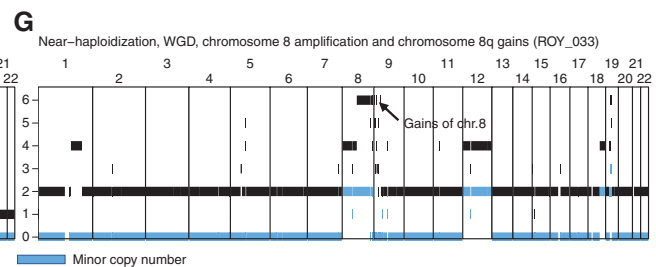
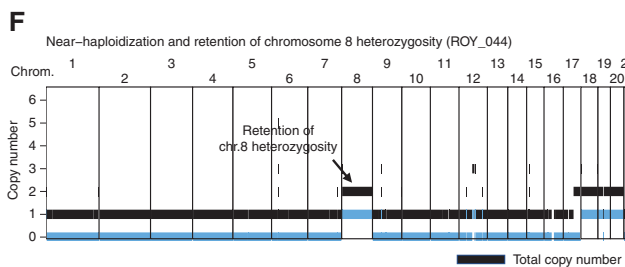
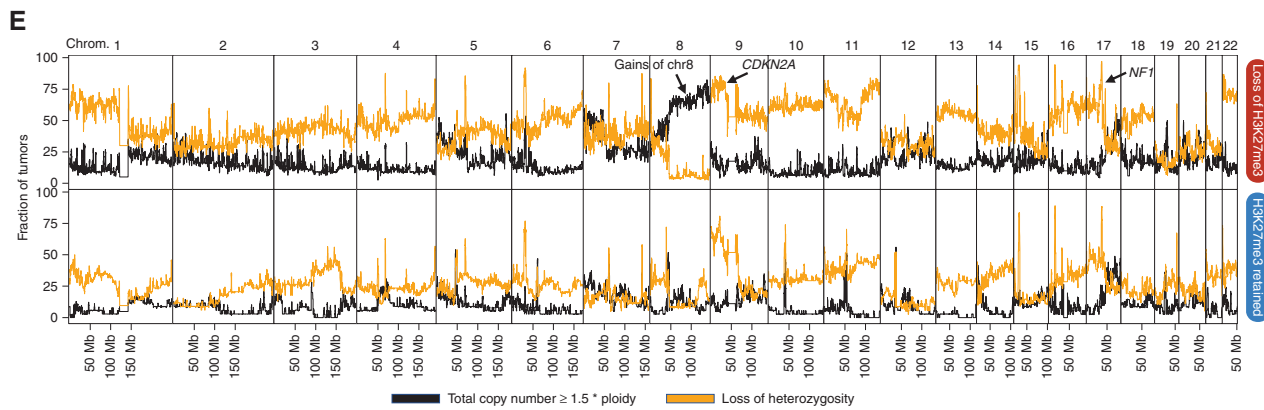
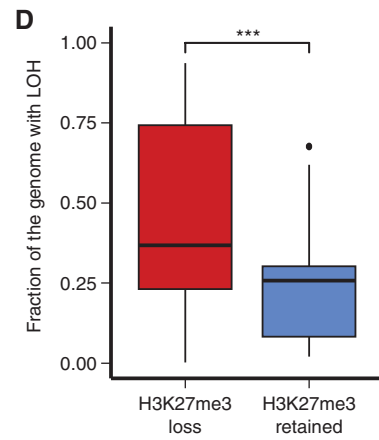
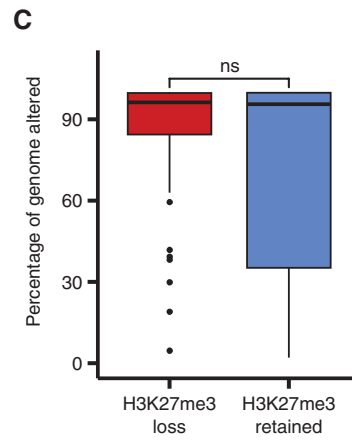
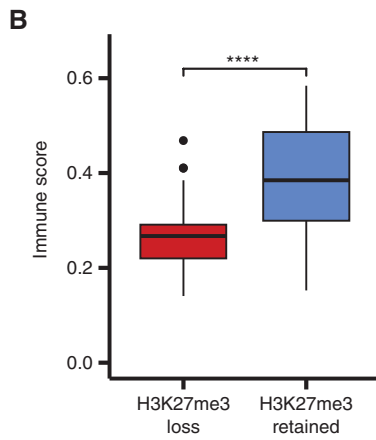
Given the strong association between distinct copy-number profiles with H3K27me3 status and immune infiltration, we evaluated the clinical relevance of SCNA patterns to predict overall survival in comparison with the predictive ability based on clinical data and H3K27me3 status (Methods). For this analysis, we focused on high-grade MPNSTs arising in individuals with NF1 to increase statistical power. Among the set of chromosome arms showing recurrent copy gain, copy loss, or LOH, LOH of chromosome 5q was the most predictive feature of poor prognosis (Fig. 3A and B). Other copy-number aberrations predictive of poor prognosis are LOH of chromosomes 11q, 7p, and 22q, and amplification of chromosome 2q and 9q ( $P < 0.05$ ;

**Figure 2.** MPNSTs stratify into two clinically and biologically distinct groups according to H3K27me3 status. **A**, Pathways related to immune activation are downregulated in tumors with loss of H3K27me3 estimated. **B**, Immune score analysis reveals decreased immune infiltration according to H3K27me3 status. **C**, Tumors with retention and loss of H3K27me3 are characterized by high levels of somatic copy-number alterations. **D**, Tumors with loss of H3K27me3 show significantly higher levels of genome-wide LOH as compared with tumors with retention of H3K27me3. **E**, Genome-wide analysis of the fraction of tumors showing amplification (black) and LOH (orange) based on H3K27me3 status. Representative copy-number profiles for tumors with loss of H3K27me3 showing near-haploidization (**F**) followed by WGD and chromosome 8q amplification (**G**). Box plots in **B–D** show the median, first, and third quartiles (boxes), and the whiskers encompass observations within a distance of  $1.5 \times$  the interquartile range from the first and third quartiles. Amp., amplification; chrom., chromosome.

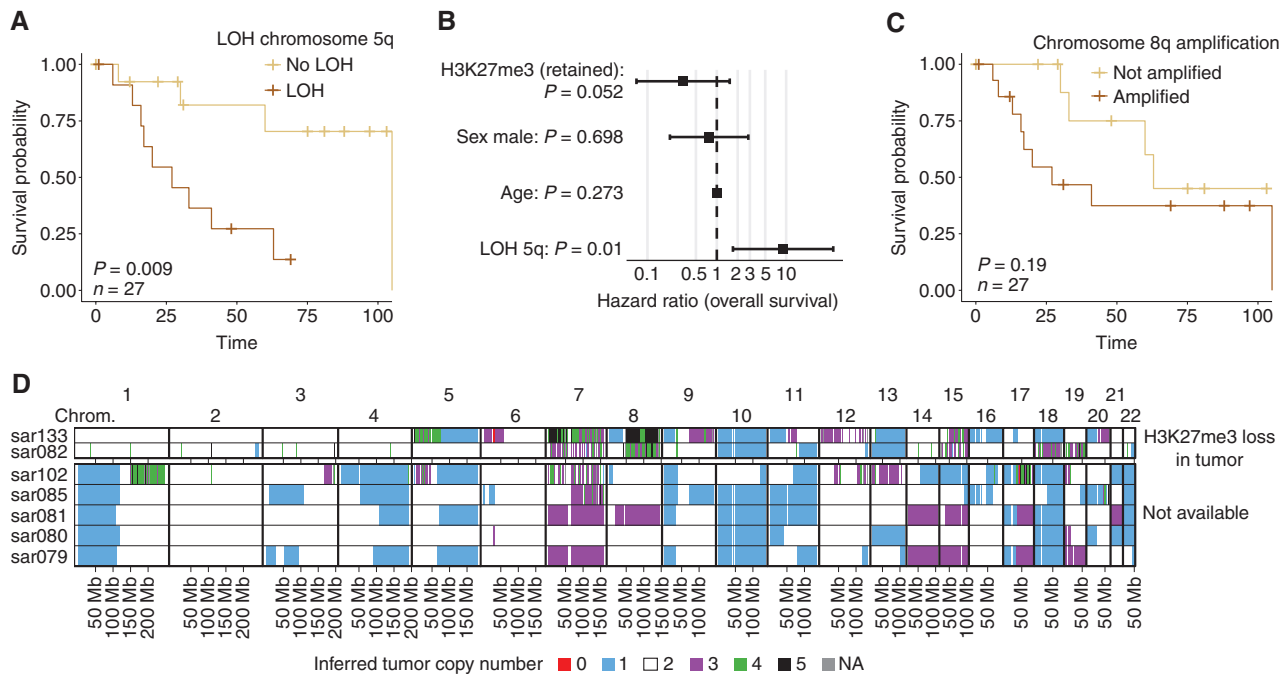
**A** Top 20 pathways downregulated in MPNSTs with H3K27me3 loss

GO:0050778 positive regulation of immune response	7.57
GO:0002252 immune effector process	7.57
GO:0002253 activation of immune response	6.7
GO:0002250 adaptive immune response	4.8
GO:0002757 immune response-activating signal transduction	4.69
GO:0002443 leukocyte mediated immunity	4.69
GO:0002764 immune response-regulating signaling pathway	4.6
GO:0002460 adaptive immune response based on somatic recombination	4.51
GO:0042110 T-cell activation	4.51
GO:0002449 lymphocyte mediated immunity	3.84
GO:0046649 lymphocyte activation	3.76
GO:0050863 regulation of T-cell activation	3.48
GO:0050865 regulation of cell activation	3.48
GO:0050870 positive regulation of T-cell activation	3.38
GO:0002694 regulation of leukocyte activation	3.36
GO:0050867 positive regulation of cell activation	3.26
GO:0002696 positive regulation of leukocyte activation	3.18
GO:0019221 cytokine-mediated signaling pathway	3.18
GO:0034097 response to cytokine stimulus	3.18
GO:0071345 cellular response to cytokine stimulus	3.18

-log<sub>10</sub> Q







**Figure 3.** Recurrent copy-number aberrations predict patient outcome and are detected in cfDNA 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 WGS of cfDNA from patients with MPNST.

log-rank test). There appeared to be a trend toward worse prognosis among tumors showing amplification of chromosome 8q, although not reaching statistical significance (Fig. 3C). Next, we investigated whether copy-number patterns associated with H3K27me3 loss and predictive of survival can be detected in cell-free DNA (cfDNA) collected from a separate cohort of patients ( $n = 14$ ) with NF1-related MPNST at different time points and analyzed by ultra-low-pass WGS (ref. 40; Methods). Overall, we detected copy-number aberrations recurrently found in MPNSTs (Fig. 2E), including loss of chromosomes 1p, 5q, 4q, and 22, and chromosome 8 amplification (Fig. 3D; Supplementary Figs. S9–S11). These results thus suggest that copy-number profiles detected in cfDNA could serve as a noninvasive surrogate for H3K27me3 status and prognostication, which warrants further investigation in larger cohorts.

### Timing of Mutations in MPNST Evolution

We next sought to determine the relative timing of the genomic alterations that occur during MPNST development across time and space. To this aim, we first investigated which alterations are present in low-grade MPNSTs and MPNST precursor lesions (either benign neurofibromas or transitional lesions referred to as ANNUBPs). We detected biallelic inactivation of *NF1* and *CDKN2A* caused by nonrecurrent complex SVs in 2 of 2 neurofibromas, 3 of 3 ANNUBPs, and 4 of 6 low-grade MPNSTs (Supplementary Fig. S12). *EED* was mutated by SVs in one neurofibroma and one low-grade MPNST (Supplementary Fig. S12). Analysis of two anatomically distant tumors from the same individual, a benign neurofibroma and a high-grade MPNST, revealed *CDKN2A*

inactivation in both tumors by independent SVs, suggesting convergent evolution and that *CDKN2A* loss is the second genomic alteration after *NF1* inactivation in MPNST pathogenesis (Fig. 4A–C). Together, these results indicate that somatic *CDKN2A* inactivation is an early event in MPNST evolution, which is necessary but not sufficient for the transformation of neurofibromas into MPNSTs.

Given the high rate of WGD in MPNSTs, we next investigated when WGD occurs during MPNST development. Timing analysis of SCNAs revealed that WGD is consistently a late clonal event in tumors with loss of H3K27me3 and is common in the background of a near-haploid genome (Fig. 4D). In tumors with H3K27me3 loss, WGD is followed by additional gains of chromosome 8q, likely as isochromosome 8q (Methods; ref. 41). In contrast, in tumors with retention of H3K27me3, WGD occurs earlier in tumor development over a wider period of time (two-sided Mann-Whitney *U* test,  $P < 0.0001$ ; Fig. 4D). These results suggest that WGD events occur at different stages during tumor evolution depending on H3K27me3 status.

To elucidate the patterns of intratumor heterogeneity, we performed multiregional high-depth exome sequencing of five regions from 21 and 15 MPNSTs with loss and retention of H3K27me3, respectively (Fig. 4E–I). Chromosome 8 amplification and WGD were found across multiple spatially distant regions in tumors with loss of H3K27me3 (Supplementary Fig. S13A–S13D). The late occurrence of WGD events in tumors with H3K27me3 loss suggests that chromosomal losses, including near-haploidization, may be associated with low fitness that is rescued by the WGD event, which triggers clonal expansion and malignant transformation. Rapid tumor

enlargement fosters intratumor heterogeneity of mutations and clonal competition, as evidenced by the accumulation of subclonal mutations and variable levels of histopathologic heterogeneity in spatially distant regions (Fig. 4H–I; Supplementary Fig. S13A–S13D), and might explain the aggressive clinical features of tumors with loss of H3K27me3.

Collectively, our integrative analysis of multiomic 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 tumor evolution include WGD and further amplifications of chromosome 8q, which might be accompanied by additional chromosomal instability (Fig. 5; Supplementary Fig. S14). In contrast, tumors with H3K27me3 retention evolve through extensive chromosome instability and chromothripsis and display more heterogeneous karyotypes (Fig. 5; Supplementary Fig. S15).

## DISCUSSION

This integrative multiomic analysis of *NF1*-related and sporadic MPNSTs has expanded the understanding of MPNST development. The international collaboration permitted the collection of a large number of fresh-frozen specimens of this rare tumor along with paired normal samples, and with comprehensive pathologic characterization. Multiomic analysis, coupled with multiregional 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 individualized treatment approaches. Although the focus has been to improve understanding of MPNST development, either in the setting of *NF1* or as a sporadic tumor, 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 the correlation of molecular signatures with tumor status. Our WGS analysis revealed that the inactivation of *CDKN2A* is caused by nonrecurrent SVs prior to the transition from benign to malignant tumors, suggesting that these are random events that increase the fitness of mutant cells. Other loci may contribute to tumor development in a subset of tumors, which may also represent therapeutic targets that could be exploited at an early stage of tumor 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).

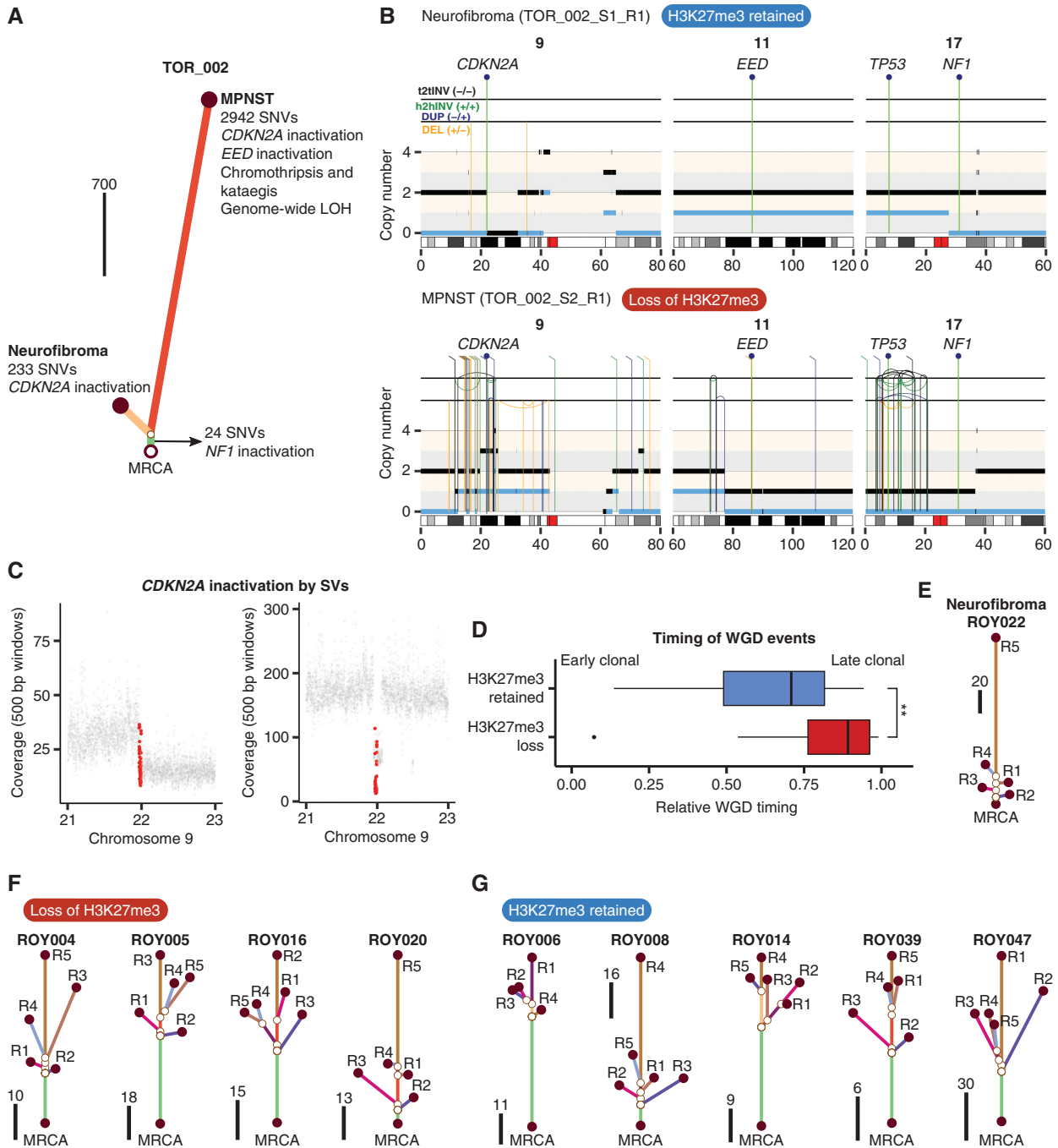
Second, analysis of WGS data revealed that the inactivation of the PRC2 complex, which follows the inactivation

of *CDKN2A*, is mediated by nonrecurrent 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 tumor 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 tumor DNA facilitates tumor subclassification into categories predictive of clinical behavior. Transcriptomic data revealed that tumors with retention of H3K27me3 also show upregulation for markers of immune infiltration and activation of the adaptive immune system, suggesting that this subset of tumors 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 unrecognized complex pattern of divergent tumor evolution on the basis of PRC2 loss. In tumors 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 tumor evolution include WGD and further amplifications of chromosome 8q. Although 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. Tumors 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.

We acknowledge several limitations of this study. We relied on retrospectively collected samples and data in order to achieve a relatively large sample size within a reasonable timeframe, resulting in some variability in the treatment approaches and types of clinical data available. Although we achieved our goal in terms of sample size, a larger overall collection of tumors would have mitigated the limited statistical power we had for assessing correlations between genomic and clinical variables, a limitation caused by the relatively small “bin” sizes when sorting tumors into subcategories based on different variables such as *NF1* germline status, H3K27 methylation status, and status for loss or gain of specific chromosomes. A larger collection of transitional/precursor tumors would have facilitated a better understanding of the genomic events occurring early in tumor development. Also, related to the understanding of early genomic events, we used bulk sequencing and tested multiple regions on a subset of 10 tumors. Single-cell sequencing would improve our understanding of tumor heterogeneity and evolution. In addition, bulk RNA-seq data suggested that pathways related to immune infiltration were upregulated in tumors that retained H3K27 methylation. Additional research is needed, through single-cell RNA-seq and other methods, in order to better understand the role of immune cells in



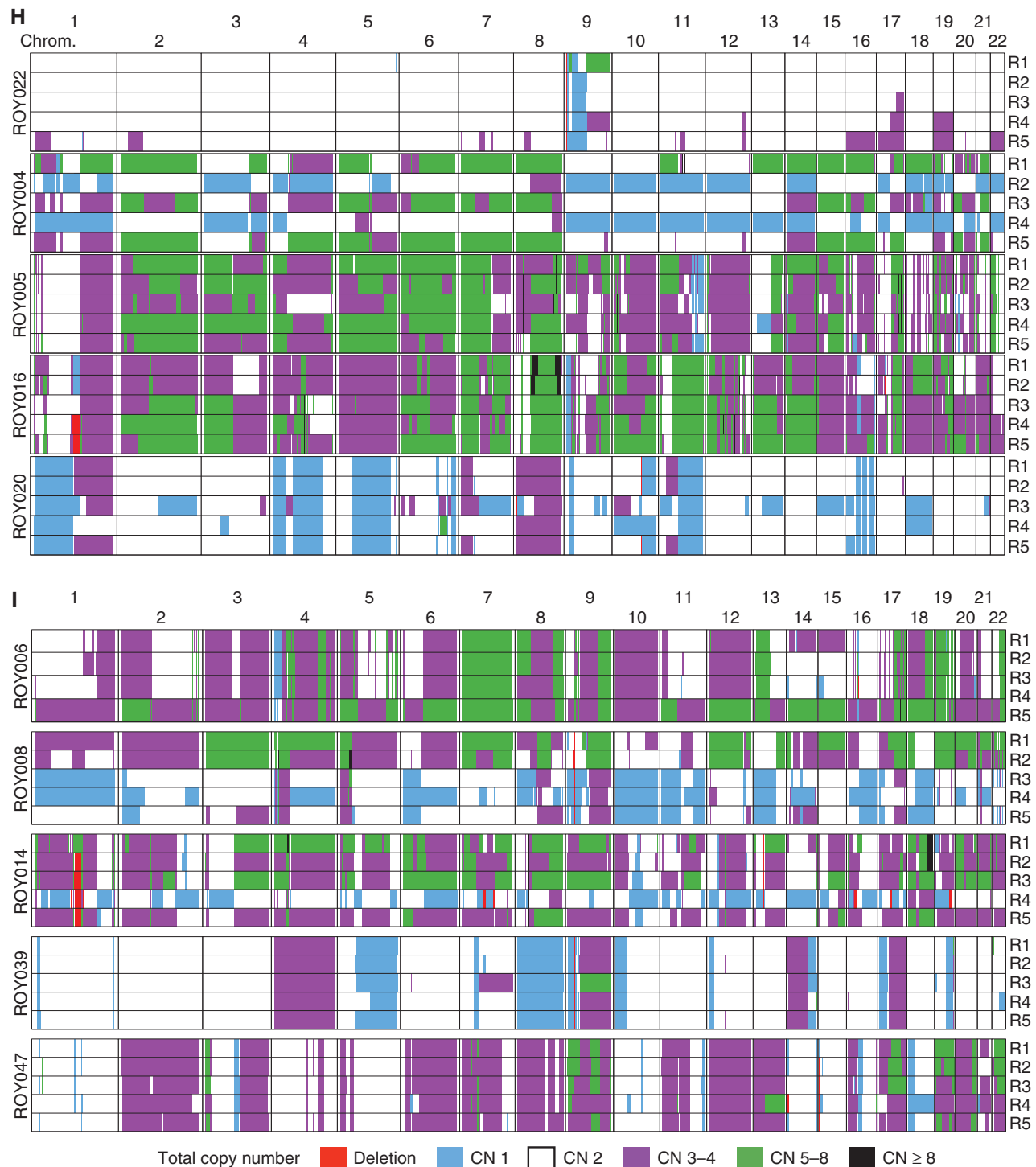


**Figure 4.** MPNST evolution and intratumor heterogeneity. **A**, Phylogenetic tree inferred from SNVs detected using WGS data for a neurofibroma and a high-grade MPNST sample from individual TOR\_002. Scale bar, 700 SNVs. **B**, Rearrangement profiles of a neurofibroma (top) and a high-grade MPNST (bottom) from patient TOR\_002. The genomic position of CDKN2A, EED, TP53, and NF1 is shown at the top. Both tumors show biallelic inactivation of CDKN2A and NF1 mediated by SVs. Total and minor copy-number values are shown in black and blue, respectively. Somatic events common to both tumors (NF1 deletion and 24 SNVs) and private to each of them are shown. **C**, Sequencing coverage calculated at 500 bp windows at the CDKN2A locus for the neurofibroma (top) and the high-grade MPNST (bottom). Dots corresponding to windows mapping to CDKN2A are shown in red. **D**, Difference in the timing of 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 multiregional sequencing data from fresh-frozen samples for a neurofibroma (**E**), MPNSTs with loss (**F**), and retention (**G**) of H3K27me3. Each region (R) is arbitrarily named with a number. The scale bars indicate the number of SNVs. (continued on following page)

different subsets of MPNSTs and the mechanisms underpinning immune evasion.

In summary, our results provide the foundation for an MPNST clinical care model that takes into account the

genomic architecture of the tumor to provide a more accurate diagnosis, prognosis, treatment approach, and surveillance. At the level of diagnosis, our comparison of pathologic classification to genomic data clearly demonstrates an inability

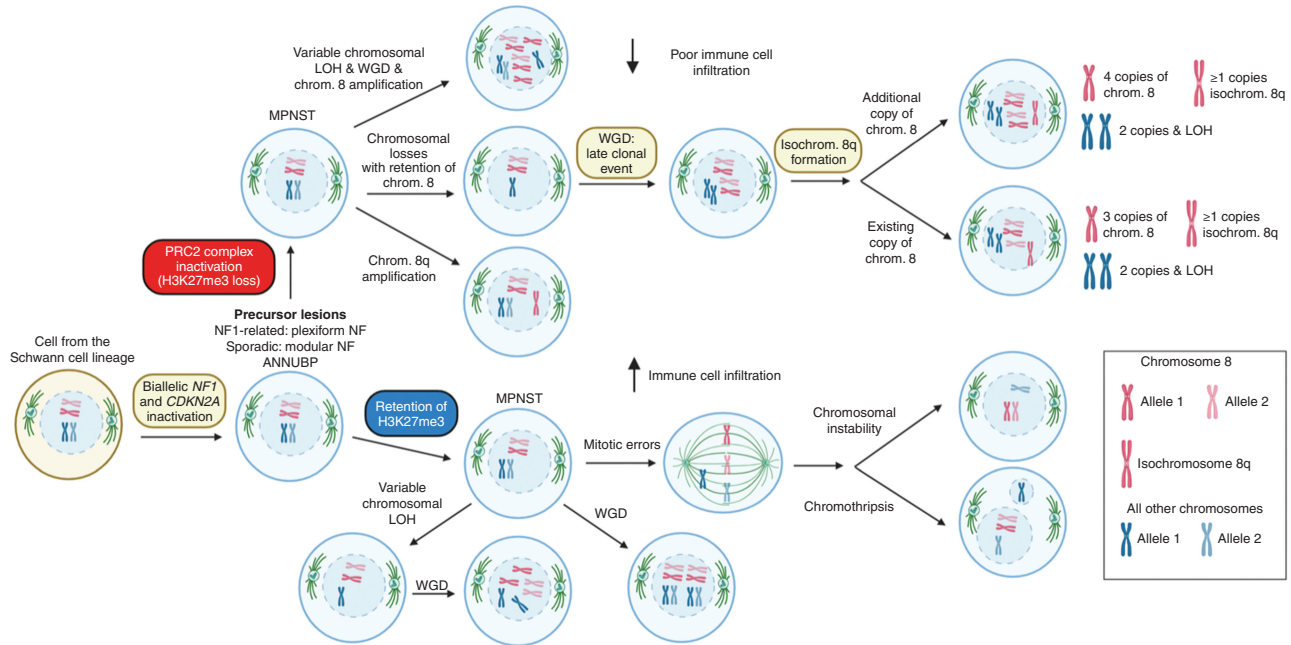


**Figure 4. (Continued) H,** Genome-wide total copy-number profiles estimated using whole-exome sequencing data for a neurofibroma and MPNSTs with loss of H3K27me3. **I,** Genome-wide total copy-number profiles estimated using whole-exome sequencing data for MPNSTs with retention of H3K27me3. Box plots in **D** show the median, first, and third quartiles (boxes), and the whiskers encompass observations within a distance of  $1.5 \times$  the interquartile range from the first and third quartiles. Chrom., chromosome; MRCA, most recent common ancestor.

to identify subsets of tumors based solely on pathology characterization. Our analysis revealed specific genomic patterns of MPNST evolution that enabled classification of tumor subtypes that are more predictive of prognosis than MPNST classification based on clinical or pathologic data alone. As

such, genomic analysis in the clinical setting could benefit all patients with an MPNST through more accurate categorization of these tumors prior to the initiation of either surgical or medical therapy, facilitating a more accurate prognosis. Furthermore, a classification that includes genomic information





**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. chrom., chromosome; isochrom., isochromosome; PRC2, polycomb repressive complex 2.

could offer an opportunity to assign patients to more specific treatment protocols that correlate with H3K27me3 status, as compared with current practice, which is largely determined by tumor grade (i.e., low or high). To this end, our data confirmed that H3K27me3 status correlated with survival and demonstrated that genomic copy-number profiles could serve as a biomarker for H3K27me3 status. Future work will focus on determining whether H3K27me3 status can predict response to therapy. The results support recent data showing detection of cell-free tumor DNA in plasma from patients with an MPNST (40). As such, the results provide new opportunities for improved clinical management in the form of tumor 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). Given the highly rearranged genomes in these tumors, effective medical therapy may remain elusive. Ultimately, early detection followed by surgical resection may prove the most effective approach for patients with an MPNST.

## METHODS

### Human Subjects

All investigations involving human subjects were performed after approval by an institutional review board and in accordance with the principles of ethical research guidelines described in the U.S. Common Rule. Informed written consent was obtained from each subject. The Dana-Farber Cancer Institute Institutional Review Board (IRB) determined that this research met the criteria for exemption from IRB review and was categorized as Secondary Use research.

The GeM Consortium is composed of investigators from academic centers and hospitals (“member sites”) with eligible subjects and existing tissue banking protocols with consent for biospecimen and clinical data

collection; genetic testing, including but not limited to whole-genome/exome sequencing; sharing of specimens/data with outside institutions, researchers, etc. The GeM coordinating center at Boston Children’s Hospital (BCH) coordinated sharing of retrospectively collected specimens from existing tumor banks and pathology archives at member sites to perform molecular characterization of MPNST and related tumors.

All specimens and data were stripped of Protected Health Information using the Safe Harbor Method and coded. The prefix of each tumor sample ID included in the final cohort in Fig. 1 ( $n = 95$ ) is an abbreviation of the member site that contributed the required specimens for analysis: Royal National Orthopaedic Hospital at University College, London (“ROY,”  $n = 43$ ), Moffitt Cancer Center, Tampa (“MOF,”  $n = 12$ ), Mount Sinai Hospital, Toronto (“TOR,”  $n = 9$ ), Massachusetts General Hospital, Boston (“MGH,”  $n = 8$ ), Washington University School of Medicine, St. Louis (“WUS,”  $n = 6$ ), Nagoya University, Nagoya (“NAG,”  $n = 4$ ), Boston Children’s Hospital/Dana-Farber Cancer Institute, Boston (“BCH,”  $n = 3$ ), New York University Langone Medical Center, New York (“NYU,”  $n = 3$ ), Huntsman Cancer Institute at University of Utah, Salt Lake City (“HCI,”  $n = 1$ ), Lifespan Laboratories, Rhode Island Hospital, Providence (“LIF,”  $n = 1$ ).

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

Comprehensive clinical and pathology report data for each participant were collected by the international MPNST Registry at Washington University School of Medicine (WUSM). Data include demographic information, disease course, tumor size/anatomic location, histologic/immuno-histochemical characteristics, diagnostic imaging, surgical procedures, systemic treatment information, neoadjuvant therapy, toxicity, clinical outcomes, and survival.

### Processing of Fresh-Frozen Tumor and Paired Normal Samples

The GeM Consortium's Standard Operating Procedure (SOP) for the processing of tissue for pathology review and molecular analysis was modeled on the Royal National Orthopaedic Hospital's SOP for the 100,000 Genomes Project, founded by England's National Health Service in 2012, as previously described (21). Fresh-frozen tissue sections [hematoxylin and eosin (H&E); 5  $\mu$ m] were assessed to select the most viable areas comprising high-quality tumor samples in terms of cellularity, lack of necrosis, and areas with little contaminating nonneoplastic tissue. Fresh-frozen sections of high cellularity (at least 40%) and with less than 20% necrosis were used for nucleic acid isolation. After pathologist assessment, two or four tubes of fresh-frozen tumor tissue scrolls (sectioned at 10  $\mu$ m each) were collected depending on whether the tumor was of high cellularity (10,000+ cells) or low cellularity (<4,000 cells), respectively. Each tube (Sarstedt–Biosphere Safeseal Tube 2.0 mL) of fresh-frozen tumor sections for DNA isolation contained 20 curls of tissue in a buffer solution. Two tubes (Sarstedt–Biosphere Safeseal Tube 1.5 mL) with 6 curls each of fresh-frozen scrolls (sectioned at 10  $\mu$ m each) in 1 mL TRIzol were collected for RNA isolation, regardless of tumor cellularity. Four additional multiregional samples were taken from a subset of 9 fresh-frozen NF1-related MPNST specimens to assess intratumor heterogeneity by performing 500 $\times$  exome sequencing, whole-transcriptome sequencing by RNA-seq, and methylation profiling. Normal tissue samples were confirmed to be free of tumor by a pathologist.

Fresh-frozen tumor and germline DNA samples were extracted, and quality was assessed at two pathology hubs, with approximately 50% of GeM samples processed at each site. BCH used the Maxwell RSC Tissue DNA Kit for extraction (Promega) and the Quantus Fluorometer and Picogreen (Promega) for quantification; Royal National Orthopaedic Hospital used QIAamp's DNA Mini and Blood Maxi kits (Qiagen) for extraction (Thermo Fisher Scientific) and the Qubit High Sensitivity and Nanodrop (Thermo Fisher Scientific) for quantification and qualification, respectively. All tumor/normal specimens with >400 ng total DNA (Picogreen, Promega) and a DIN  $\geq$  6.0 (TapeStation, Agilent) were included in this study. Fresh-frozen tumor and normal nerve samples in TRIzol were stored at  $-80^{\circ}\text{C}$  until transfer to the Broad Institute on dry ice for RNA extraction (AllPrep DNA/RNA/miRNA Universal Kit; Qiagen). RNA quality and insert size was assessed by Caliper LabchipGXII (PerkinElmer, Billerica, MA) producing a RQS value (equivalent to RIN). RNA quantity was determined by Quant-iT RiboGreen (Thermo Fisher Scientific). Samples that met the minimum required input (250 ng of purified total RNA with RIN  $\geq$  6.0) were included in this study.

### Processing of FFPE Samples

FFPE tissue blocks were collected for each fresh-frozen tumor which underwent multiomic profiling and were used for pathology review, IHC (MYF-4, H3K27me3, S100, and Sox10) and additional molecular characterization [i.e., ploidy assay and multiregional whole-exome sequencing (WES)]. FFPE H&E slides from available blocks were reviewed by a pathologist to identify the block most representative (i.e., morphologically similar) to the frozen material which underwent multiomic profiling. These H&E sections were digitally scanned (40 $\times$ ) at each pathology hub by the Deep Lens Biomedical Imaging Team and uploaded to the VIPER digital pathology platform (Deep Lens, Inc.). Multiregional 500 $\times$  exome sequencing was performed on 5 samples each from 36 FFPE NF1-related MPNST specimens to assess intratumor heterogeneity.

### Pathology Review

All cases sent for multiomic profiling underwent expert pathology review in the cloud-based digital imaging platform VIPER hosted by Deep Lens. Pathology reports were reviewed for the patient's NF1

status and lesion's proximity to a nerve, exposure to neoadjuvant systemic or radiation therapies, and anatomic location. The number of slides available for each case ranged from 1 to 5 slides, including the following stains: H&E, H3K27me3, MYF-4, S100, and/or Sox10. Additional IHC (HMB45 and Melan-A) was performed to exclude melanoma. Of the 64 subjects with available data on previous radiotherapy (83% of GeM cohort), only 4 subjects with MPNST (2 NF1-related; 2 sporadic) reported "Yes" to previous radiotherapy (6% of subjects with available data). These patients all indicated that their radiation was directed at their MPNST, not a different type of cancer. Given that knowledge of prior radiation is limited, it is unclear if any of the tumors in our cohort were radiation induced. Of the 64 subjects for which we have these data, none were associated with prior radiation.

Five pathologists with expertise in soft-tissue pathology were involved in the central review process (AA, MMB, BCD, AF, and DL). Each pathologist reviewed the most representative FFPE H&E slide from each case via a digital image platform and recorded results in a case review form (CRF). The CRF included the following data: percent tumor necrosis, degree of cytological atypia, tumor cellularity and purity, mitotic count, presence of lymphocytic infiltrate and atypical mitosis, and final tumor diagnosis. Possible diagnoses included neurofibroma, neurofibroma with atypia, cellular neurofibroma, ANNUBP, low-grade MPNST, conventional high-grade MPNST (hyper- and hypocellular fascicles with/without geographic necrosis), conventional MPNST with heterologous elements, nonconventional MPNST with associated conventional low-grade areas, nonconventional MPNST, spindle cell/undifferentiated sarcoma, melanoma, and other (e.g., dermatofibrosarcoma protuberans, synovial sarcoma, epithelioid MPNST, spindle cell rhabdomyosarcoma, carcinoma, etc.). The histologic classification of tumors was based on the proposed consensus classification by Miettinen and colleagues (45). ANNUBP is defined as neurofibroma with at least two of the following: (i) increase in cellularity, (ii) loss of neurofibroma architecture, (iii) fewer than 3 mitosis per 10 high-powered field (HPF), and (iv) cytologic atypia. Low-grade MPNST is defined as features of ANNUBP with 3 to 9 mitosis per 10 HPF and lacking necrosis. High-grade MPNST is defined by either mitosis more than 10 per HPF or 3 to 9 mitosis per 10 HPF combined with necrosis. The pathologists also reviewed IHC slides and recorded their results (i.e., stain expression, percent estimate) in a CRF.

For the second round of pathology reviews, the pathologists met via conference call and screen sharing to review all the cases and discuss the findings and proposed consensus tumor diagnosis. As the group worked through the cases, the study coordinator recorded their consensus tumor diagnosis (KP). A third round of reviews was held to review digital slides for all cases in which there was a discrepancy between tumor diagnosis and the results of the methylation assay sarcoma classification. For the purposes of collating tumor diagnosis with multiomic data, tumor diagnoses were collapsed into five main categories based on germline NF1 status and histologic features of the neurofibroma: germline/conventional, germline/nonconventional, sporadic/conventional, sporadic/nonconventional, and equivocal. Conventional MPNST were tumors that had classic morphology and were composed of relatively monomorphic spindle cells in intersecting fascicles with alternating hypercellular and hypocellular zones. Nonconventional tumors were composed predominantly of large epithelioid cells with marked pleomorphism and occasional round cell morphology which are not typical features in "classic" MPNST. Equivocal cases were excluded from the genomics analysis unless otherwise stated. Neurofibromas were classified using the criteria outlined in histopathologic evaluation of atypical neurofibromatous tumors and their transformation into MPNST in patients with NF1 (45).

### WGS

Between 350 and 500 ng of tumor DNA were used to prepare DNA sequencing libraries using the TruSeq DNA PCR Free 350bp kit, which were sequenced on the Illumina NovaSeq6000 sequencing



machine using S4 flow cells to generate  $2 \times 151$  paired-end reads. Quality control for all steps was performed using a DNA Screen-tape on the Agilent TapeStation system. Raw sequencing reads were mapped to the GRCh38 build of the human reference genome using BWA-MEM (46) version 0.7.17-r1188. Aligned reads in BAM format were processed following the Genome Analysis Toolkit (GATK, version 4.1.8.0) Best Practices workflow to remove duplicates and recalibrate base quality scores (47). NGSCheckMate was utilized using default options to verify that sequencing data from tumor-normal pairs corresponded to the same individual (48).

## WES

Four additional multiregional cores were taken from the most representative FFPE block for a subset of 10 cases for WES. DNA extraction from FFPE tissue samples was performed using the Covaris truXTRAC FFPE DNA kit with the E220 Evolution-focused ultrasonicator system, and Prepito truXTRAC DNA FFPE kit with the Chemagic Prepito-D. Any excess paraffin was trimmed before sectioning an FFPE tissue block, or after the section had been cut from the FFPE block, to maintain an optimal ratio of 80% tissue to 20% paraffin (or higher). Cores of 1.22-mm diameter were taken; any sections or cores longer than 10 mm in length were cut in half before being loaded into the microtubes. The total mass of FFPE samples processed per extraction was between 2 and 5 mg. The quantity and purity of the DNA were assessed using the Qubit dsDNA Assay kit and Nanodrop, respectively (Thermo Fisher). The DNA fragment size of each sample was estimated using an Agilent 2200 TapeStation (Agilent Technologies). Exome libraries were prepared using the Roche Kapa HyperPlus kit, captured using the HyperExome kit Roche, and sequenced on Illumina NovaSeq6000, to generate  $2 \times 151$  paired-end reads.

## Tumor DNA Ploidy Assay

Additional scrolls were cut (50  $\mu\text{m}$ ) from the most representative FFPE block for all cases for a ploidy assay. Of the 95 frozen cases that qualified for histologic assessment by pathology review, 14 had insufficient FFPE-derived tumor DNA qualifications for the ploidy assay. Quantification of DNA content was performed as previously described (25, 49). Briefly, 50-mm FFPE sections were deparaffinized and rehydrated, and cytoplasmic digestion with protease type VIII (Sigma P5380) was utilized to create nuclear suspensions. Nuclear suspensions underwent filtering, cytospinning, DNA hydrolysis (5 M HCl), and staining with Feulgen (Schiff's fuchsin-sulphite reagent; Sigma S5133). Individual nuclei were classified as "normal," "lymphocyte," "plasma," "fibroblast," and "tumor" using the PWS Classifier software (50). Nuclei were manually revised to select high-confidence members of the "normal," "lymphocyte," and "tumor" categories. The median integrated optical density (IOD) value of the combined "normal" and "lymphocyte" categories was taken as the IOD level for a diploid cell ( $\text{IOD}_{2n}$ ) in each sample. The maximum density value of the IOD of the major tumor subclone was taken as the tumor clonal IOD level ( $\text{IOD}_t$ ). Tumor ploidy was calculated as  $2(\text{IOD}_t/\text{IOD}_{2n})$ .

## Detection of Pathogenic Germline NF1 Alterations

The germline short variant discovery workflow from GATK (version 4.1.8.0; ref. 47), Strelka2 (version 2.9.2; ref. 51), and VarDict (version 1.8.2; ref. 52) were used to detect germline SNVs and INDELS in the WGS data from the matched normal samples. In brief, Strelka2 and VarDict were run using default parameter values. In the case of GATK, intermediate GVCF files were generated for each sample using HaplotypeCaller in GVCF mode. Next, GVCF files for all samples were consolidated into a single GVCF file using the GATK functionality CombineGVCFs using default options. Finally, all samples were jointly genotyped using GenotypeGVCFs and default options.

Annovar (version 2018Apr16) was used to annotate all variants (53). Truncating variants detected by at least one algorithm were considered to be pathogenic, and missense variants were reviewed for literature support. All variants were also manually curated through visual inspection of raw sequencing reads using BamSnap (54). Deep intronic mutations generating cryptic splice sites were validated at the RNA-seq level and through a literature review. To detect germline SVs, we followed two strategies. First, we called SVs in all matched normal samples using Manta (version 1.6.0; ref. 55), LUMPY (version 0.2.13; ref. 56), SvABA (version 1.1.3; ref. 57), and Delly (version 0.8.3; refs. 57, 58). Breakpoints mapping to the *NF1* locus, including 100 Kb upstream and downstream, were manually curated by inspecting the raw sequencing reads. Second, to identify *NF1* microdeletions without SV support, we computed the B-allele frequency profile for chromosome 17 for each sample using the heterozygous SNPs from dbSNP (build 151). In addition, we computed the sequencing coverage at 500-bp windows using mosdepth (59). To call an *NF1* microdeletion, the B-allele frequency (BAF) and coverage profiles at the *NF1* locus were required to clearly deviate from 0.5 and decrease by a factor of 2, respectively.

## Detection of Somatic SNVs and INDELS

Somatic SNVs were detected using Mutect2 (GATK version 4.1.8.0), MuSE (version v1.0rc; ref. 60), and Strelka2 (version 2.9.2) using default options. Somatic INDELS were called using Strelka2 (version 2.9.2) and Mutect2 (GATK version 4.1.8.0) using default options. Each algorithm was run independently on each tumor sample using the matched normal sample from the same individual as the control. INDEL calls were left-aligned using the GATK functionality LeftAlignAndTrimVariants. The calls generated by each algorithm were merged using the Python library mergevcf (<https://github.com/ljduersi/mergevcf>). To maximize specificity, only mutations called by at least two algorithms were considered for further analysis. Mutations overlapping known polymorphisms in dbSNP (build 151) were only considered for further analysis if the coverage in the matched normal sample was at least 20 reads and none of them supported the mutation. Missense variants predicted to be deleterious by MetaLR and MetaSVM as implemented in Annovar (version 2018Apr16) were considered pathogenic. To identify driver genes, we used the R package dNdScv (61).

## Mutational Signature Analysis

Mutational signatures were extracted *de novo* using nonnegative matrix factorization (NMF) as implemented in the R package NMF (62). In brief, for each rank in the set  $\{2, 12\}$ , 2,000 factorizations were run. The rank maximizing the consensus cophenetic correlation coefficient, and for which the residual sum of squares (RSS) showed a clear inflection point, was selected as the optimal rank (63). The identified signatures were compared against the COSMIC v3.2 catalog of signatures and assigned to existing ones using a cosine similarity cutoff of 0.85 using the R package MutationalPatterns (64). Next, the contribution of the set of extracted signatures to each sample was estimated using the function `fit_to_signatures` from MutationalPatterns. For this analysis, signatures appearing in hypermutated tumors were considered only when analyzing the tumors.

## Detection of Microsatellite Instability

Mutations at microsatellite loci were detected using MSIprofiler (<https://github.com/parklab/MSIprofiler>) as previously described (65). Only repeats with at least 10 sequencing reads in both the normal and tumor sample were considered for further analysis. The Kolmogorov-Smirnov test was used to compare the distributions of repeat lengths. The level of significance was set at 0.01 after false discovery rate (FDR) correction.

### Copy-Number Analysis

For WGS data, the software packages FACETS (version 0.12.1) and ascatNGS (version 2.5.1) were used to detect somatic copy-number alterations and to compute purity and ploidy estimates for all tumor samples (66, 67). Both tools were run using default options. The distribution of cancer cell fraction and mutation copy-number values for somatic SNVs computed using the copy-number values were manually reviewed in all cases to assess the quality of the ploidy and purity calls. In cases where inconsistencies were identified, such as the lack of clonal SNVs, a different combination of purity and ploidy values was selected to refit the copy-number profile. The ploidy and purity values were selected to best match the ploidy ascertained from image cytometry ploidy analysis. Timing analysis of WGD events was performed as previously described (25). For WES data, FACETS (version 0.12.1) was used to compute copy number, ploidy, and purity calls for each tumor region.

### Copy-Number Signature Analysis

For WGS data, allele counts were generated for SNP positions lifted over to hg38 (1,837,383 SNPs) using alleleCounter (<https://github.com/cancerit/alleleCount>). For WES data, allele counts were instead generated for all hg38 dbSNP v153 common SNPs (12,140,046 SNPs). LogR and BAF values were calculated for SNPs in all normal samples. SNP positions with a high frequency of undetermined genotype ( $0.68 < \text{BAF} < 0.9$  or  $0.1 < \text{BAF} < 0.32$ ) were excluded from analysis. Further, any SNPs closer than 75 bp to a neighboring SNP were excluded from the analysis. Lastly, SNPs with a high variance in logR were excluded from the analysis ( $\text{SNP logR variance} > 2 * \text{median SNP logR variance}$ ). Following SNP removal and cleaning, normal samples with a high variance in logR were removed from the dataset (samples with  $> 1.5 * \text{median sample logR variance}$ ). LogR and BAF values were calculated for all tumor samples with a corresponding retained normal sample. Tumor LogR values were corrected for GC content, and copy-number profiles were generated using ASCAT (68), with a segmentation penalty of 70, designating SNPs with  $0.2 < \text{BAF} < 0.8$  in the corresponding normal as heterozygous. For WES samples, segmentation was jointly performed for all samples originating from a single patient using the function *asmultipcf*. Following copy-number calling, copy-number profiles were summarized as vectors of counts of segments categorized into 48 copy-number classes based on combinations of LOH status (homozygous deletions, LOH, heterozygous), total copy number (0, 1, 2, 3–4, 5–8 or 9+) and segment length (0–100 kb, 100 kb–1 Mb, 1–10 Mb, 10–40 Mb, >40 Mb). For homozygous deletions, segment lengths were restricted to 0 {100 kb, 100 kb–1 Mb, >1 Mb}. See (27) for a full description of the development of these categories.

For WGS data, NMF as implemented in sigProfilerExtractor was used to extract *de novo* copy-number signatures from sequential subsets of the dataset as previously described (69): all samples ( $n = 81$ ), excluding hypersegmented samples ( $n_{\text{segs}} < 1,000$ ;  $n = 79$ ), remaining poorly explained samples (cosine similarity  $< 0.85$ ;  $n = 35$ ), samples with a high proportion of LOH ( $p_{\text{LOH}} >$ ;  $n = 12$ ). In each case, the *de novo* signatures were decomposed into previously identified pan-cancer copy-number signatures (27) using the function sigProfilerSingleSample. For hypersegmented samples, a novel signature was identified, for which the attribution was taken from the all-samples extraction. For the remaining samples, the decomposition that best explained the data (highest cosine similarity) was retained. For WES data, samples were decomposed into previously identified pan-cancer copy-number signatures using sigProfilerSingleSample (27).

### Detection of Somatic SVs

Somatic SVs were detected in WGS data using Manta (version 1.6.0), LUMPY (version 0.2.13), SvABA (version 1.1.3), and Delly (version 0.8.3). Each algorithm was run independently on each tumor using the bulk sequencing data from the same individual as control. The calls generated

by each algorithm were merged using the Python library mergevcf allowing 200 bp of slop at the breakpoints. Only calls generated by at least two algorithms were kept for further analysis. Intrachromosomal SVs were classified into four groups [deletion (DEL); duplication (DUP); head-to-head inversion (h2hINV); and tail-to-tail inversion (t2tINV)] depending on the read orientation at the breakpoints following the notation established by PCAWG (70). SVs detected in at least two tumor samples from different individuals were discarded, as these are likely germline polymorphisms or artifacts. Finally, SVs with at least one breakpoint mapping to telomeres, centromeres, heterochromatin regions, or to blacklisted regions by the ENCODE project were removed (71).

### Detection of Chromothripsis

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

### Rearrangement Signatures

Intrachromosomal structural variants were classified into categories according to the SV type as determined from the read orientation at the breakpoints, i.e., deletions, duplications, and inversion (h2hINV and t2tINV SVs were grouped together); and the size of the genomic region bridged by the breakpoints: 10 to 10,000 bp, 10 to 100 Kb, 100 Kb to 1 Mb, 1 to 10 Mb, and >10 Mb. All translocations were grouped into a single category. In addition, SVs were further stratified based on whether they mapped to an SV cluster. To define SV clusters, we used piecewise constant regression applied on the interbreakpoint distance values sorted by genomic coordinates. We required at least 10 breakpoints per segment ( $k_{\text{min}} = 10$ ) and used a value of 25 for the gamma parameter, which controls the smoothness of the segmentation (26). SVs with at least one breakpoint mapping to a segment with an average interbreakpoint distance smaller than 10% of the average interbreakpoint distance were considered to be involved in a cluster. Therefore, SVs were classified into 32 categories: 30 for intrachromosomal SVs and 2 for translocations, depending on whether at least one of the breakpoints participated in a cluster of SVs as defined above. Rearrangement signatures were extracted using NMF as implemented in the R package NMF (62). As in the case of SBS, the rank in the set {1...12} maximized the consensus cophenetic correlation coefficient and for which the RSS showed an inflection point that was considered optimal. In addition, the RSS difference obtained for factorizations performed on the original and randomized data using the function *randomize* was compared to ensure that the results obtained were not due to spurious correlations (73).

### mRNA Sequencing

At least 250 ng of purified total RNA with RIN  $\geq 6.0$  were required. Library preparation was conducted using the Illumina TruSeq Strand Specific Large Insert RNA kit (50 M pairs) v1. Thermo Fisher ERCC RNA controls were added prior to Poly(A) selection, providing additional control for variability, including quality of the starting material, level of cellularity, RNA yield, and batch-to-batch variability. Libraries were sequenced on the Illumina HiSeq4000 sequencing machine at the Broad Institute (Cambridge, MA). Sequencing reads were mapped to the transcriptome using the STAR aligner (version 2.7.4a; ref. 74). Gene-expression counts were generated using HTSeq (v.0.6.1p1; ref. 75) and normalized to transcripts per kilobase million (TPM). GENCODE

v22 was used as the gene annotation reference. Differential expression analysis was performed using DESeq2 (76). Gene fusions were detected using ARriba (version 2.1.0; ref. 77), TopHat-fusion (version 2.1.0; ref. 78), EricScript (version 0.5.5; ref. 79), and STAR-Fusion (version 1.10.1; ref. 80). Only those fusions called by at least two algorithms or with WGS support were considered for further analysis. Immune infiltration was estimated using consensusTME using the sarcoma cell-type-specific genes derived from TCGA (81). Immune gene sets were obtained from previous studies (82–84).

### Methylation Array Profiling

Tumor DNA derived from fresh-frozen tumors and paired normal nerve samples was used for whole-genome DNA methylation array profiling using the HumanMethylationEPIC beadchip platform as described previously (85). In brief, 350 ng of DNA were bisulfite converted using the Zymo EZ DNA methylation Gold kit (Zymo Research Corp.) as per the manufacturer's recommendations. Bisulfite-converted samples were processed and hybridized to the Infinium HumanMethylationEPIC beadchip arrays according to the manufacturer's recommendations. Raw Methylation EPIC data were uniformly processed using the R package minfi (version 1.34.0; ref. 86). Probes that did not satisfy a detection *P* value of 0.01 were discarded using the function `detP`. Probes mapping to known polymorphisms were identified using the function `dropLociWithSnps` and were discarded. Functional normalization after background correction using normal-exponential deconvolution using out-of-band probes (noob), as implemented in the function `preprocessFunnorm`, was applied to normalize methylation intensities (87, 88). The function `getBeta` was used to convert methylation intensities to Beta values, which were used for further analysis. Quality assessment was performed through visual inspection of the distribution of B values for all samples. For methylation-based diagnosis, iDAT files were analyzed using the open-source DNA methylation sarcoma classifier as described previously (22).

### Integrative Analysis of Multi-Omic Data

We performed clustering of each data type separately using ConsensusClusterPlus (89). The top 5,000 most variable genes (normalized counts) and the top 10,000 methylation CpG sites (Beta values) based on median absolute deviation were used for clustering analysis. Clustering was performed using the hierarchical clustering algorithm with 1,000 resamples, Pearson correlation as the distance metric, and a maximum number of 10 clusters. Both the analysis of methylation and RNA-seq data revealed two main groups that were consistent across data types. The cluster assignments computed using ConsensusClusterPlus were used for Cluster-Of-Clusters Analysis (COCA; ref. 90). COCA, as implemented in the R package `coca` (91), was run on the same datasets used for ConsensusClusterPlus clustering and using the *k*-means algorithm as the clustering method.

### Analysis of cfDNA Sequencing Data

Sequencing reads were filtered to keep read pairs corresponding to DNA fragments between 80 and 160 bp. Genome-wide sequencing coverage was estimated by counting the number of reads mapping to nonoverlapping windows of 1 Mb using `readCounter` (HMM Copy). IchorCNA (92) was used to estimate the tumor fraction in cfDNA and genome-wide copy-number profiles. Only samples with a tumor fraction of at least 5% were considered for further analysis. IchorCNA was run using default parameter values without accounting for subclonal events.

### Survival Analysis

Survival analysis was performed using the Cox proportional-hazards model as implemented in the R package `survival` (version 2.30; ref. 93). Significance was assessed by the likelihood ratio test using a cutoff for a statistical significance of 0.05. We considered only

MPNSTs with conventional histology for survival analysis. The proportional-hazards assumption was tested using the `cox.zph` function from the R package `survival` (93).

### Phylogenetic Analysis of Exome Sequencing Data

We constructed a binary matrix for each tumor with rows indexed by somatic SNVs and columns by tumor regions such that the *i,j* entry in each matrix was set to one if mutation *i* is present in region *j*, and to zero otherwise. To reconstruct phylogenetic trees for FFPE cases, we used the parsimony ratchet method (`pratchet`) as implemented in the R package `phangorn` (94). The `acctran` method was used to infer branch lengths for all trees. In the case of fresh-frozen samples, we used the maximum parsimony and neighbor-joining algorithms as implemented in the R package `MesKit` (95) to estimate consensus phylogenetic trees for each tumor using 100 bootstrap resamples.

### Data Availability

Raw sequencing data have been deposited at the European Genome-phenome Archive, which is hosted by the EBI and the CRG, under the dataset accession number EGAD00001008608. Ultra-low-pass WGS data of cell-free samples are available under controlled access data at <https://doi.org/10.7303/syn23651229>.

### Code Availability

The code used to process the raw data is available upon request.

### Authors' Disclosures

D.C. Borcherding reports grants from the Department of Defense and Children's Hospital-GeM Consortium during the conduct of the study; nonfinancial support from SpringWorks Therapeutics outside the submitted work; and U.S. provisional patent applications #63/277,331 and 18/053,935 pending. J.T. Jordan reports grants from Boston Children's Hospital during the conduct of the study; personal fees from Navio Theragnostics, Recursion Pharmaceuticals, Alexion Pharmaceuticals, CEC Oncology, Texas Health Resources, the American Academy of Neurology, Elsevier, the Children's Tumor Foundation, EMP Advisory, Atheneum Partners, GLG, Guidepoint Global, and Health2047 and grants from The Burke Foundation outside the submitted work; and is a board member for NF Northeast (nonprofit), a board member and scientific advisory board chair for the Neurofibromatosis Network (nonprofit), a member of the Children's Tumor Foundation Clinical Care Advisory Board (nonprofit), a board member for ABA Academy (nonprofit), and an editorial board member for *Neurology*. R.H. Kim reports grants from the Children's Tumor Foundation during the conduct of the study, as well as grants from the Children's Tumor Foundation outside the submitted work. X. Wang reports grants from NF Research Initiatives (Genomics of MPNST) during the conduct of the study. A.C. Hirbe reports personal fees from AstraZeneca and SpringWorks Therapeutics, and grants from Tango Therapeutics outside the submitted work. D.T. Miller reports personal fees from AstraZeneca outside the submitted work. No disclosures were reported by the other authors.

### Authors' Contributions

**I. Cortes-Ciriano:** Conceptualization, resources, data curation, formal analysis, supervision, investigation, visualization, methodology, writing—original draft, writing—review and editing. **C.D. Steele:** Conceptualization, data curation, formal analysis, investigation, methodology, writing—original draft, writing—review and editing. **K. Piculell:** Resources, data curation, methodology, writing—original draft, project administration, writing—review and editing. **A. Al-Ibraheemi:** Conceptualization, resources, formal analysis, investigation, methodology, writing—review and editing. **V. Eulo:** Resources, data curation, methodology,



writing–review and editing. **M.M. Bui:** Conceptualization, resources, formal analysis, investigation, writing–review and editing. **A. Chatzipli:** Writing–review and editing. **B.C. Dickson:** Resources, formal analysis, investigation, writing–review and editing. **D.C. Borchering:** Resources, data curation, writing–review and editing. **A. Feber:** Data curation, writing–review and editing. **A. Galor:** Data curation, formal analysis, investigation, writing–review and editing. **J. Hart:** Resources, writing–review and editing. **K.B. Jones:** Resources, writing–review and editing. **J.T. Jordan:** Resources, writing–review and editing. **R.H. Kim:** Resources, writing–review and editing. **D. Lindsay:** Conceptualization, resources, methodology, project administration, writing–review and editing. **C. Miller:** Data curation, writing–review and editing. **Y. Nishida:** Resources, writing–review and editing. **P.Z. Proszek:** Data curation, writing–review and editing. **J. Serrano:** Resources, data curation, formal analysis, investigation, writing–review and editing. **R.T. Sundby:** Resources, writing–review and editing. **J.J. Szymanski:** Resources, writing–review and editing. **N.J. Ullrich:** Writing–review and editing. **D. Viskochil:** Resources, writing–review and editing. **X. Wang:** Resources, writing–review and editing. **M. Snuderl:** Conceptualization, resources, data curation, formal analysis, investigation, writing–original draft, writing–review and editing. **P.J. Park:** Conceptualization, data curation, formal analysis, investigation, writing–review and editing. **A.M. Flanagan:** Conceptualization, resources, data curation, formal analysis, supervision, investigation, methodology, writing–original draft, writing–review and editing. **A.C. Hirbe:** Conceptualization, resources, data curation, formal analysis, supervision, investigation, methodology, writing–original draft, writing–review and editing. **N. Pillay:** Conceptualization, data curation, formal analysis, supervision, investigation, writing–original draft, writing–review and editing. **D.T. Miller:** Conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, writing–original draft, project administration, writing–review and editing.

## Acknowledgments

This research was funded by the Neurofibromatosis Research Initiative (NRFI) at Boston Children’s Hospital, made possible by an anonymous donation. We thank S. Cano, J. Da, C. Fletcher, M. Fournier, J. Hart, J. Johnson, D. Lucente, J. Chul Min Kim, P. Nielsen, T. Restrepo, M. Roche, and A. Silva for technical assistance in the collection and processing of specimens. We thank C. Clinton and A. Ward for their guidance and support. I. Cortes-Ciriano thanks EMBL for funding. N. Pillay is funded through a Cancer Research UK grant (grant no. 18387). N. Pillay and A.M. Flanagan 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. C.D. Steele is funded through Cancer Research UK and the Neurofibromatosis Research Initiative (NRFI) 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. Snuderl). We thank the patients and their families for their participation in this study. Some figures were generated using BioRender.com.

## Group Members: The GeM Consortium

David T. Miller (Consortium Coleader), Katherine Piculell (Consortium Coordinator), Alyaa Al-Ibraheemi, Dana C. Borchering, Marilyn M. Bui, Aikaterini Chatzipli, Isidro Cortes-Ciriano (Consortium Coleader), Chris Davies, Brendan C. Dickson, Vanessa Eulo, Adrienne M. Flanagan (Consortium Coleader), Alon Galor, Anthony Griffin, James F. Gusella, Jesse Hart, Angela C. Hirbe (Consortium Coleader), Kevin B. Jones, Justin T. Jordan, George Jour, Raymond H.

Kim, Daniel Lindsay, Colin Miller, Yoshihiro Nishida, Peter J. Park, Nischalan Pillay (Consortium Coleader), Anat O. Stemmer-Rachamimov, Tomohisa Sakai, Jonathan Serrano, Diane Shao, Matija Snuderl, Christopher D. Steele, Nicole J. Ullrich, David Viskochil, and Xia Wang.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 USC section 1734.

## Note

Supplementary data for this article are available at Cancer Discovery Online (<http://cancerdiscovery.aacrjournals.org/>).

Received July 19, 2022; revised November 9, 2022; accepted December 16, 2022; published first January 4, 2023.

## REFERENCES

- Evans DGR, Baser ME, McLaughran J, Sharif S, Howard E, Moran A. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. *J Med Genet* 2002;39:311–4.
- Kolberg M, Høland M, Agesen TH, Brekke HR, Liestøl K, Hall KS, et al. Survival meta-analyses for >1800 malignant peripheral nerve sheath tumor patients with and without neurofibromatosis type 1. *Neuro Oncol* 2013;15:135–47.
- Kallionpää RA, Uusitalo E, Leppävirta J, Pöyhönen M, Peltonen S, Peltonen J. Prevalence of neurofibromatosis type 1 in the Finnish population. *Genet Med* 2018;20:1082–6.
- Miller DT, Freedenberg D, Schorry E, Ullrich NJ, Viskochil D, Korf BR. Health supervision for children with neurofibromatosis type 1. *Pediatrics* 2019;143:e20190660.
- Gutmann DH, Ferner RE, Listerick RH, Korf BR, Wolters PL, Johnson KJ. Neurofibromatosis type 1. *Nat Rev Dis Primers* 2017;3:1–17.
- Ratner N, Miller SJ. A RASopathy gene commonly mutated in cancer: the neurofibromatosis type 1 tumour suppressor. *Nat Rev Cancer* 2015;15:290–301.
- Philpott C, Tovell H, Frayling IM, Cooper DN, Upadhyaya M. The NF1 somatic mutational landscape in sporadic human cancers. *Hum Genomics* 2017;11:13.
- Pemov A, Li H, Patidar R, Hansen NF, Sindiri S, Hartley SW, et al. The primacy of NF1 loss as the driver of tumorigenesis in neurofibromatosis type 1-associated plexiform neurofibromas. *Oncogene* 2017;36:3168–77.
- WHO classification of tumours editorial board. Soft tissue and bone tumours international agency for research on cancer. 2020;3:254–57.
- Nielsen GP, Stemmer-Rachamimov AO, Ino Y, Moller MB, Rosenberg AE, Louis DN. Malignant transformation of neurofibromas in neurofibromatosis 1 is associated with CDKN2A/p16 inactivation. *Am J Pathol* 1999;155:1879–84.
- Magallón-Lorenz M, Fernández-Rodríguez J, Terribas E, Creus-Batchiller E, Romagosa C, Estival A, et al. Chromosomal translocations inactivating CDKN2A support a single path for malignant peripheral nerve sheath tumor initiation. *Hum Genet* 2021;140:1241–52.
- Zhang M, Wang Y, Jones S, Sausen M, McMahon K, Sharma R, et al. Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors. *Nat Genet* 2014;46:1170–2.
- Lee W, Teckie S, Wiesner T, Ran L, Prieto-Granada CN, Lin M, et al. PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. *Nat Genet* 2014;46:1227–32.
- Brohl AS, Kahen E, Yoder SJ, Teer JK, Reed DR. The genomic landscape of malignant peripheral nerve sheath tumors: diverse drivers of Ras pathway activation. *Sci Rep* 2017;7:1–5.
- Lyskjær I, Lindsay D, Tirabosco R, Steele CD, Lombard P, Strobl A-C, et al. H3K27me3 expression and methylation status in histological variants of malignant peripheral nerve sheath tumours. *J Pathol* 2020;252:151–64.

16. Cleven AHG, Sannaa GAA, Bruijn IB, Ingram DR, van de Rijn M, Rubin BP, 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* 2016;29:582–90.
17. Ito Y, Kohashi K, Endo M, Yoshimoto M, Ishihara S, Toda Y, et al. Clinicopathological and prognostic significance of H3K27 methylation status in malignant peripheral nerve sheath tumor: correlation with skeletal muscle differentiation. *Virchows Arch* 2021;479:1233–44.
18. Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* 2017;171:950–65.
19. Lemberg KM, Wang J, Pratilas CA. From genes to -omics: the evolving molecular landscape of malignant peripheral nerve sheath tumor. *Genes*. 2020;11:691.
20. Pemov A, Li H, Presley W, Wallace MR, Miller DT. Genetics of human malignant peripheral nerve sheath tumors. *Neurooncol Adv* 2020;2:50–61.
21. Miller DT, Cortés-Ciriano I, Pillay N, Hirbe AC, Snuderl M, Bui MM, et al. Genomics of MPNST (GeM) consortium: rationale and study design for multi-omic characterization of NF1-associated and sporadic MPNSTs. *Genes* 2020;11:387.
22. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahn F, Reuss DE, et al. Sarcoma classification by DNA methylation profiling. *Nat Commun* 2021;12:498.
23. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. *Nature* 2020;578:94–101.
24. Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, et al. The mutational landscape of human somatic and germline cells. *Nature* 2021;597:381–6.
25. Steele CD, Tarabichi M, Oukrif D, Webster AP, Ye H, Fittall M, et al. Undifferentiated sarcomas develop through distinct evolutionary pathways. *Cancer Cell* 2019;35:441–56.
26. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 2016;534:47–54.
27. Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. *Nature* 2022;606:984–91.
28. Su XA, Ma D, Parsons JV, Replogle JM, Amatruda JF, Whittaker CA, et al. RAD21 is a driver of chromosome 8 gain in Ewing sarcoma to mitigate replication stress. *Genes Dev* 2021;35:1–17.
29. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* 2018;33:676–89.
30. Dehner C, Moon CI, Zhang X, Zhou Z, Miller C, Xu H, et al. Chromosome 8 gain is associated with high-grade transformation in MPNST. *JCI Insight* 2021;6:e146351.
31. Liao L, Song M, Li X, Tang L, Zhang T, Zhang L, et al. E3 ubiquitin ligase UBR5 drives the growth and metastasis of triple-negative breast cancer. *Cancer Res* 2017;77:2090–101.
32. Yang Y-F, Wang Y-Y, Hsiao M, Lo S, Chang Y-C, Jan Y-H, 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* 2020;485:27–37.
33. Maddipati R, Norgard RJ, Baslan T, Rathi KS, Zhang A, Saeid A, et al. MYC levels regulate metastatic heterogeneity in pancreatic adenocarcinoma. *Cancer Discov* 2021;12:542–61.
34. Gibbons B, MacCallum P, Watts E, Rohatiner AZ, Webb D, Katz FE, et al. Near haploid acute lymphoblastic leukemia: seven new cases and a review of the literature. *Leukemia* 1991;5:738–43.
35. Comprehensive pan-genomic characterization of adrenocortical carcinoma. *Cancer Cell* 2016;29:723–36.
36. Bigner SH, Mark J, SSC Jr, Eng LF, Bigner DD. A serially transplantable human giant cell glioblastoma that maintains a near-haploid stem line. *Cancer Genet Cytogenet* 1985;18:141–53.
37. Arbajian E, Köster J, Vult von Steyrn 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* 2018;31:93–100.
38. Bovée JV, van Royen M, Bardoel AF, Rosenberg C, Cornelisse CJ, Cleton-Jansen AM, et al. Near-haploidy and subsequent polyploidization characterize the progression of peripheral chondrosarcoma. *Am J Pathol* 2000;157:1587–95.
39. Nichols CA, Gibson WJ, Brown MS, Kosmicki JA, Busanovich JP, Wei H, et al. Loss of heterozygosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nat Commun* 2020;11:1–14.
40. Szymanski JJ, Sundby RT, Jones PA, Srihari D, Earland N, Harris PK, 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* 2021;18:e1003734.
41. Shen H, Shih J, Hollern DP, Wang L, Bowlby R, Tickoo SK, et al. Integrated molecular characterization of testicular germ cell tumors. *Cell Rep* 2018;23:3392–406.
42. Kryukov GV, Wilson FH, Ruth JR, Paulk J, Tsherniak A, Marlow SE, et al. MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science* 2016;351:1214–8.
43. Rodon Ahnert J, Perez CA, Wong KM, Maitland ML, Tsai F, Berlin 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* 2021;39:3019.
44. De Raedt T, Brems H, Wolkenstein P, Vidaud D, Pilotti S, Perrone F, et al. Elevated risk for MPNST in NF1 microdeletion patients. *Am J Hum Genet* 2003;72:1288–92.
45. Miettinen MM, Antonescu CR, Fletcher CDM, Kim A, Lazar AJ, Quezado MM, et al. Histopathologic evaluation of atypical neurofibromatous tumors and their transformation into malignant peripheral nerve sheath tumor in patients with neurofibromatosis 1—a consensus overview. *Hum Pathol* 2017;67:1–10.
46. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
47. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. Sebastopol (CA): O'Reilly Media; 2020.
48. Lee S, Lee S, Ouellette S, Park W-Y, Lee EA, Park PJ. NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. *Nucleic Acids Res* 2017;45:e103.
49. Hedley DW, Friedlander ML, Taylor IW, Rugg CA, Musgrove EA. Method for analysis of cellular DNA content of paraffin-embedded pathological material using flow cytometry. *J Histochem Cytochem* 1983;31:1333–5.
50. Dunn JM, Mackenzie GD, Oukrif D, Mosse CA, Banks MR, Thorpe S, et al. Image cytometry accurately detects DNA ploidy abnormalities and predicts late relapse to high-grade dysplasia and adenocarcinoma in Barrett's oesophagus following photodynamic therapy. *Br J Cancer* 2010;102:1608–17.
51. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* 2018;15:591–4.
52. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, McEwen R, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res* 2016;44:e108.
53. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
54. Kwon M, Lee S, Berselli M, Chu C, Park PJ. BamSnap: a lightweight viewer for sequencing reads in BAM files. *Bioinformatics* 2021;37:263–4.
55. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 2016;32:1220–2.
56. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 2014;15:R84.
57. Wala JA, Bandopadhyay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res* 2018;28:581–91.
58. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 2012;28:i333–9.

59. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 2018;34:867–8.
60. Fan Y, Xi L, Hughes DST, Zhang J, Zhang J, Futreal PA, et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol* 2016;17:178.
61. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. *Cell* 2017;171:1029–41.
62. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. *BMC Bioinf* 2010;11:367.
63. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci USA*. 2004;101:4164–9.
64. Blokzijl F, Janssen R, van Bostel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* 2018;10:33.
65. Cortes-Ciriano I, Lee S, Park W-Y, Kim T-M, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* 2017;8:1–12.
66. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 2016;44:e131.
67. Raine KM, Van Loo P, Wedge DC, Jones D, Menzies A, Butler AP, et al. ascatNgs: identifying somatically acquired copy-number alterations from whole-genome sequencing data. *Curr Protoc Bioinformatics* 2016;56:15.9.1–15.9.17.
68. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. 2010;107:16910–5.
69. Ashiqul Islam SM, Wu Y, Díaz-Gay M, Bergstrom EN, He Y, Barnes M, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. *Cell Genomics* 2022;2:100179.
70. Pan-cancer analysis of whole genomes. *Nature* 2020;578:82–93.
71. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;9:1–5.
72. Cortés-Ciriano I, Lee J-K, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* 2020;52:331–41.
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* 2008;6:275–92.
74. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
75. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166–9.
76. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
77. Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* 2021;31:448–60.
78. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* 2011;12:R72.
79. Benelli M, Pescucci C, Marseglia G, Severgnini M, Torricelli F, Magi A. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* 2012;28:3232–9.
80. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol* 2019;20:213.
81. Jiménez-Sánchez A, Cast O, Miller ML. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res* 2019;79:6238–46.
82. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 2015;160:48–61.
83. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 2017;355:eaaf8399.
84. House IG, Savas P, Lai J, Chen AX, Oliver AJ, Teo ZL, et al. Macrophage-derived CXCL9 and CXCL10 are required for antitumor immune responses following immune checkpoint blockade. *Clin Cancer Res* 2020;26:487–504.
85. Serrano J, Snuderl M. Whole genome DNA methylation analysis of human glioblastoma using Illumina BeadArrays. *Methods Mol Biol* 2018;1741:31–51.
86. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014;30:1363–9.
87. Triche TJ Jr, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA methylation BeadArrays. *Nucleic Acids Res* 2013;41:e90.
88. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* 2014;15:503.
89. Wilkerson MD, Neil Hayes D. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;26:1572–3.
90. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
91. Cabassi A, Kirk PDW. Multiple kernel learning for integrative consensus clustering of omic datasets. *Bioinformatics* 2020;36:4789–96.
92. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* 2017;8:1324.
93. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer Science & Business Media; 2013.
94. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2010;27:592–3.
95. Liu M, Chen J, Wang X, Wang C, Zhang X, Xie Y, et al. MesKit: a tool kit for dissecting cancer evolution of multi-region tumor biopsies through somatic alterations. *Gigascience* 2021;10:giab036.