Title: Genomic features of NF1-associated peripheral nerve sheath tumors: a cohort analysis from the Johns Hopkins NF1 biospecimen repository.

Authors: Jineta Banerjee [1*], Yang Lyu [2*], Stavriani C. Makri [3*], Alexandra J. Scott [1], Lindy Zhang [3,4], Ana Calizo [5], Kai Pollard [6], Kuangying Yang [2], John M. Gross [7], Jiawan Wang [3], Carlos G. Romo [3,8], Robert Allaway [1], Jaishri O. Blakeley [3,8], Angela C. Hirbe [2], Christine A. Pratilas [3,4]

*These authors contributed equally.

Affiliations:

Sage Bionetworks, Seattle, WA, USA (1);

Division of Oncology, Department of Internal Medicine, Siteman Cancer Center, Washington University St. Louis (2);

Departments of Oncology (3), Pediatrics (4), Pathology (7), and Neurology (8), Johns Hopkins University School of Medicine;

Cellular and Molecular Medicine Program, Johns Hopkins University (5);

Personal Genome Diagnostics (6).

Abstract

Background: Neurofibromatosis type 1 (NF1) is a prevalent inherited neurocutaneous condition that predisposes to the development of peripheral nerve sheath tumors (PNST) including cutaneous neurofibromas (CNF), plexiform neurofibromas (PNF), atypical neurofibromatous neoplasms with unknown biological potential (ANNUBP), and malignant peripheral nerve sheath tumors (MPNST). Historically, therapeutic progress for PNF and MPNST has been limited in part due to restricted availability of primary tissues from patients with NF1. The successful advancement of therapeutic development for NF1-associated PNST necessitates ongoing efforts in the systematic acquisition and analysis of human tumor specimens and their corresponding model systems.

Methods: Patients with clinically or genetically confirmed NF1 having a clinically indicated surgical resection or biopsy of any NF1-associated tumor were invited to participate in an institutional review board (IRB) approved study for the collection and sharing of tissues and specimens. Tumors were assessed by the study pathologist, and banked in the laboratory as flash frozen tissues, paraffin embedded blocks or slides, DNA and RNA, or single cell suspensions. Efforts were made to create cell lines and patient derived xenografts (PDX) from primary human tissues. Clinical data for participating patients were fully annotated in a database that corresponds to banked tissue specimens. Applications for access to biospecimens, genomic data, and disease models, as well as de-identified clinical and molecular data are reviewed and approved with IRB oversight, to allow internal and external sharing to promote research collaboration.

Results: Since the inception of the JH NF1 biospecimen repository in 2016, 357 unique samples have been banked (from 183 unique patients) and include PNF (n=89), ANNUBP

(n=6), MPNST (n=62), CNF (n=103), and diffuse neurofibroma (diffuse NF, n=44). Xenografts have successfully been generated from seven MPNST samples and cell lines have been generated from three PNF and seven MPNST. RNA sequencing (RNAseq) and whole genome sequencing (WES) data were generated from 73 and 114 primary human tumor samples, respectively. These pre-processed data, standardized for immediate computational analysis, are accessible through the NF Data Portal, allowing immediate interrogation. Our analysis herein highlights key genetic variants and alterations in gene expression patterns, linked to pathways implicated in the pathology of the NF1-associated tumor types represented in the dataset. This work also combines new sample data with previously released samples, offering a comprehensive view of the entire cohort sequenced to date. Somatic variants in genes including *NF1*, *SUZ12* and *LRP1* and *LRP2* were identified in MPNST. Enrichment of RAS-RTK signaling pathways was identified through analysis of variants in both PNF and MPNST, however, MPNST demonstrated unique enrichment in pathways associated with extracellular matrix organization and cell cycle regulation.

Conclusion: Analysis of primary human tissue samples is critical for identification of therapeutically relevant molecular alterations. As a dedicated effort to systematically bank tumor samples from people with NF1 undergoing surgery, in collaboration with molecular geneticists and computational biologists who seek to advance understanding of NF1 biology, the Johns Hopkins NF1 biospecimen repository offers access to samples and genomic data to the NF1 research community to promote advancement of NF1-related therapies.

67 **Materials and Methods:**

68 Tumor sample collection and sequencing

69 *Patient enrollment*

70 All research was conducted according to a Johns Hopkins Hospital (JHH) institutional
71 review board (IRB)-approved protocol, #IRB00096544.

72 Patients with clinically or genetically confirmed NF1 having a clinically indicated surgical
73 resection or biopsy of any NF1-associated tumor (including, but not limited to, CNF, PNF,
74 ANNUBP, MPNST, other NF1-associated neoplasm) were invited to participate. Written
75 informed consent was obtained from all patients. The JHH IRB-approved consent form
76 includes a statement about the voluntary nature of the research, a description of
77 corresponding clinical data that will be collected for future analysis, and a statement that
78 samples may be shared beyond JHH. Blood was collected from most patients on the day
79 of surgery.

80 Medical records were retrospectively reviewed for demographic information and
81 information related to the NF1 clinical condition (genetic diagnosis, family history, age of
82 diagnosis), characteristics (phenotypic findings, symptoms), and tumor burden (number
83 and size of NF1-associated tumors). These clinical data are stored in a password
84 protected and de-identified database at JHH.

85

86 *Tumor preservation and quality control*

87 Surgical specimens were couriered to surgical pathology immediately after resection. The
88 study neuropathologist performed immediate inspection of the tumor to ensure that

89 adequate tumor existed for clinical diagnostic needs. Upon approval, tumor pieces were
90 sampled for banking and transported to the research laboratory in harvest medium (RPMI
91 with 20% fetal bovine serum, supplemented with 1% penicillin-streptomycin and
92 glutamine).

93 Specimens were sized into 5-10mm aliquots under sterile conditions in a biosafety
94 cabinet. Individual aliquots were placed into 10% neutral buffered formalin, cell freezing
95 media (Sigma #C6295), RNAlater (Qiagen #76104), and/ or placed into an empty vial and
96 snap frozen on dry ice. Formalin-fixed paraffin-embedded tissues were stained with
97 hematoxylin and eosin (H&E). H&E slides from each specimen were reviewed by the
98 study neuropathologist to confirm an accurate diagnosis of each preserved specimen.

99

100 *Tumor sample preparation for DNA sequencing*

101 Genomic DNA from blood and tumor was extracted using the QIAGEN DNeasy Blood &
102 Tissue kit. DNA library was constructed using KAPA HyperPrep Kits for NGS DNA Library
103 Prep (Roche #7962363001). Exomes were captured by IDT xGen Exome Hyb Panel v1.0
104 (IDT #1056115) and the libraries were sequenced by NovaSeq6000 S4 with ~100X
105 coverage for normal samples and tumor samples.

106

107 *Tumor sample preparation for RNA sequencing*

108 RNA was isolated from tumor by using TRIZOL. Samples containing at least 100ng total
109 RNA with RNA Integrity Number (RIN) > 6.5 were sequenced. For RNA extraction, the
110 first batch of samples were prepared with MGI/GTAC RiboErase method. Depletion was

111 performed with KAPA RiboErase HMR (Roche #07962274001). Library preparation was
112 performed with a WUSTL in-house preparation protocol. For the second batch, samples
113 were prepared with Illumina TruSeq® Stranded Total RNA Library Prep Gold kit
114 (#20020599) which includes RiboZero Gold depletion. Library preparation was performed
115 according to the manufacturer's protocol. Fragments from both batches were sequenced
116 on an Illumina NovaSeq-6000 using paired-end reads extending 150 bases.

117

118 Whole exome sequencing data processing for somatic variant calling and analysis

119 Raw fastq data files were quality checked using FastQC v0.11.9 and a report was
120 generated using MultiQC v1.8. Fastq files were aligned to GRCh38 using BWA 0.7.17-
121 r1188 (1). Duplicates were marked using GATK MarkDuplicates, and bases recalibrated
122 using GATK BaseRecalibrator and GATK ApplyBQSR (GATK v4.3.0.0) (2-4). Somatic
123 single nucleotide variants (SNVs) were then called using Strelka2 software (Strelka
124 v2.9.10) (5) and Mutect2 (GATK v4.4.0.0) (6). Strelka 2 was shown to have high precision
125 and recall for SNVs compared to others as tested in the precision FDA challenge (5, 7).
126 The variants were annotated using Variant Effect Predictor (VEP v99.2) and converted to
127 mutation annotation format (MAF) files using vcf2maf (vcf2maf v1.6.21). All of these steps
128 were completed on Nextflow Tower running the standardized nf-core pipeline sarek v3.1.2
129 (8).

130

131 Variant calling from samples with paired normal in a separate batch

132 The samples contained in syn52659564 and **Supplementary Table 1** were special cases
133 where the tumor samples and the paired normal samples were sequenced in two different
134 batches with different sequencing library kits and thus had two different browser
135 extensible data (BED) files for target capture. For the above samples, first the BED file
136 from JH_batch1 was lifted over from hg19 to hg38 coordinates using the UCSC liftOver
137 tool (9) and the liftover chain file named 'hg19ToHg38.over.chain.gz' and then sorted
138 using the sort function from bedtools suite. Then a common region BED file was
139 generated using the intersect function in bedtools suite with at least 50% overlap in
140 intervals between the BED file from JH_batch1 and WU_batch1. The common region
141 BED file (syn52594249) was then used to call somatic variants from these tumor-normal
142 pairs using the sarek v3.2.2 pipeline and Strelka2 and Mutect2 variant callers.

143

144 Consensus variant calling using SomaticSeq

145 To avoid detection of false positive variant calls in the samples, we identified and reported
146 the variant calls that had consensus between Strelka2 and Mutect2 callers. To identify
147 these consensus calls we used a previously benchmarked method called SomaticSeq
148 (10). Once consensus calls for SNVs and insertions and deletions (indels) were identified,
149 the variants in respective variant call formats (VCFs), were annotated using vcf2maf in
150 the sarek v3.2.2 pipeline to generate a merged MAF file that was used for further analysis.
151 The consensus variants were filtered for high confidence calls (PASS filtered) with
152 population allele frequency below 0.05 to identify robust variant calls excluding common
153 variants.

154

155 RNAseq data processing and analysis

156 Raw fastq files were processed using nf-core/rnaseq (v3.11.2) (11) and quantified using
157 salmon: v1.10.1 (12). ComBat from sva R package (v3.42.0) was used for batch
158 correction (13). DESeq2 (v1.34.0) was applied to call the differentially expressed genes
159 (|fold change| > 1.2, adjust $P < 0.05$), and ggplot2 R package (v3.3.6) was applied to draw
160 volcano plots. Principal component analysis (PCA) and uniform manifold approximation
161 and projection (UMAP) (14) were applied for the visualization of samples. Gene set
162 enrichment analysis (GSEA) was conducted by using R-package fgsea (v1.20.0) [13].
163 ggplot2 (v3.3.6) was used for the gene volcano plot and pathway enrichment plot [14].
164 Pheatmap (v1.0.12) was applied to plot gene expression heatmap. Singscore (v1.14.0)
165 was applied to calculate single sample gene signature score (15).

166

167 Evaluation of relatedness in genomic and transcriptomic data

168 Somalier (v0.2.17) (16) was used to evaluate the relatedness between WES and RNAseq
169 samples. The *extract* function was used to identify informative sites from the WES cram
170 files and RNAseq bam files, and then the *relate* function was used to calculate statistics
171 on the number of shared genotypes and the coefficient of relatedness between all
172 samples. One WES sample, JH-2-009-2578C-A, did not contain data at a sufficient
173 number of polymorphic sites to calculate relatedness to other samples and was excluded
174 from further analysis. The pheatmap R package (v1.0.01) was used to generate
175 heatmaps of the relatedness comparing each primary human WES sample exactly once

to each primary human RNAseq sample and the ggplot2 package was used to generate scatter plots. Cell line and PDX samples were excluded from the heatmap.

Results

Efforts to generate a genomically and clinically annotated NF1 biospecimen repository, for the purpose of sharing resources, to promote collaborative research, and make specimens and genomic data available to the NF1 research community, were previously reported (17). As of December 2023, the Johns Hopkins NF1 biospecimen repository includes 357 banked tumor tissues from 183 unique patients with NF1. A full specimen inventory is available online (https://bit.ly/nf1biobank) and investigators can request specimens by providing a statement of scientific rationale and intended use. More than 65 requests for sharing of available specimens, models, and/ or genomic data have been fulfilled to date.

Patient and sample information

Clinical and demographic information on the patients whose tumors have been genomically characterized, including sex and age by range, are included in **Table 1**. Specific ages at time of tumor collection are not provided to protect patient identities given the rarity of some NF1-associated conditions. Sample and file information can be found in Synapse.org datasets (WES: syn53132831.1, RNAseq: syn53133024.1; for more details, refer to the Data Availability selection) and accessed on the NF Data Portal for further analysis.

198

199 Relatedness of WES and RNAseq data from individual patients was confirmed as a
200 measure of quality control.

201 As a quality and validation step, we used Somalier (16) to evaluate relatedness between
202 WES and RNAseq samples, comparing each WES sample exactly once to each RNAseq
203 sample. We found that all sample comparisons that originated from different individuals
204 had a relatedness coefficient <0.9. Based on this and the clinical annotations of the
205 samples, we set a threshold of relatedness coefficient ≥0.9 to be the confirmation that a
206 pair of WES and RNAseq samples were indeed from the same individual (**Figure 1A**).

207 With this threshold in place, all but three corresponding samples from the same patient
208 were found to meet the definition of relatedness (**Figure 1A,B**). One RNAseq sample
209 from a recurrent tumor, JH-2-079-CB92G-FG226, had a relatedness coefficient of 0.86 to both
210 the normal WES sample from the same individual and to a WES sample from a different tumor in
211 the same individual. However, it had high relatedness (0.997) to the WES sample generated from
212 the same recurrent tumor. Two additional RNAseq samples that originated from one
213 individual had lower than expected relatedness to other samples from that individual. On
214 examination of the clinical annotations, we found that JH-2-002-GAF53-2E57B was an
215 RNAseq sample of a patient-derived cell line and JH-2-002-GAF53-3BAGC was an
216 RNAseq sample of a xenograft developed from the tumor sample. This distinction
217 suggested that the threshold of relatedness coefficient determined by our analysis was
218 sensitive enough to detect differences between primary samples and the cell lines and
219 xenografts derived from them.

220 Unbiased exploration of somatic variants identifies NF1 among the top genes with
221 variants using WES

222 WES data from the human tumor samples were haplotype matched to their respective
223 normal samples (**Supplementary Table 2**). To explore the variant landscape of the
224 human tumor samples from the biobank, we consulted six highly cited reports (17-22) to
225 generate a list of 60 priority genes that are deemed to be of key biological importance in
226 NF1-MPNST. Given the differences in detecting limits of variant caller algorithms, we
227 used multiple variant callers to identify single nucleotide polymorphisms (SNPs) and short
228 indels for these genes of interest in the tumor samples. Variants detected using Strelka2
229 (**Figure 2A,D,E)** exceeded those detected using Mutect2 (**Figure 2B,F,G**). The median
230 number of variants found in the samples using Strelka2 and Mutect2, respectively, are
231 shown in **Figure 2D** (median =129) and **2F** (median=11) and highlight that the majority of
232 the variants detected were missense variants. The top ten genes with variants detected
233 using Strelka2 and Mutect2 callers are shown as **Figures 2E,G**. To find variants that were
234 identified regardless of the variant caller used, we identified consensus variant calls with
235 SomaticSeq – (**Figure 2C,H,I)**. Samples in which no high-confidence variants were
236 detected using at least one variant caller were excluded from further analysis, and those
237 that had consensus calls for variants using SomaticSeq are visualized in **Figure 2C**. A
238 median of 6 variants including missense, nonsense, splice-site, nonstop, and translation
239 start-site variants were found per sample when only consensus calls were considered
240 (**Figure 2H)**. In the consensus calls, *NF1, SUZ12, LRP1, and LRP2* genes were found
241 within the top 10 genes with variants with moderate or high impact on gene function
242 among the samples (**Figure 2I**).

243 Given that PNF and MPNST tumor types represented the greatest number of samples in
244 our cohort, and PNF may undergo malignant transformation to MPNST, we focused on
245 exploring the consensus genomic variants identified in these two tumor types with
246 attention to their similarities, and differences that may relate to malignant transformation.

247

248 Somatic variants in known oncogenic pathways identified in PNF and MPNST samples

249 We identified somatic SNVs in the *NF1* gene in 31% of PNF samples and 26% of MPNST
250 samples examined. The genetic loci of *NF1* variants identified in our samples and the
251 predicted effect on protein expression are shown schematically alongside a protein
252 structure representation of the *NF1* gene (**Figure 3A**, PNF shown at top, MPNST shown
253 at bottom). **Figure 3B** represents a diagrammatic rendition of known functional domains
254 in the NF1 protein for illustration, modified from (23).

255

256 Additionally, we compared the presence of variants in known oncogenic pathways from
257 The Cancer Genome Atlas (TCGA) (24) to the consensus genetic variants found in our
258 cohort. Specifically, we tested the enrichment of genes where genetic variants were found
259 in PNF samples and in MPNST samples separately to understand which pathways may
260 be affected by the presence of variants in each of these tumor types. SNVs in our cohort
261 of PNF samples identified enrichment of variants in genes related to the RAS-RTK
262 pathway (**Figure 3C**). Fourteen out of 36 PNF samples had variants in at least one gene
263 involved in RAS-RTK signaling pathways. In MPNST samples, however, five out of 19
264 MPNST samples had variants in at least one gene related to the RAS-RTK pathway, and

265 in several samples, genetic variants related to the Hippo, TP53, MYC, and PI3K pathways
266 were enriched (**Figure 3D).**
267
268 MPNSTs demonstrate enrichment in pathways associated with extracellular matrix
269 organization and cell cycle regulation

270 We employed PCA and UMAP to generate a clustering visualization, based on the
271 expression of all genes in all tumor samples (**Figure 4A,B**). The majority of CNF tumors
272 fell into a distinct cluster, appearing less related to MPNST, indicating divergent gene
273 expression profiles between the two tumor types. Some PNF and MPNST samples
274 however, clustered nearby each other, possibly indicating a range of biological behavior
275 in PNF, which are known to have a propensity for malignant progression in some patients,
276 or possibly MPNST with less aggressive biology or potential for metastasis. Interrogation
277 of the clinical outcomes for specific patients from whom these PNF were obtained would
278 take years of follow up and is therefore outside the scope of this analysis. For other tumor
279 types, strong discernible patterns were not evident, due to the constraints imposed by
280 limited sample numbers.

281 To further characterize the MPNST and PNF cohorts, we evaluated the expression of
282 genes previously implicated in MPNST biology [29]. We found that *ERBB3, S100B, CNP,*
283 *PMP22, PDGFB,* and *TGFBR2* were significantly downregulated in MPNST compared to
284 PNF, while *CCNB2, TWIST1, COL6A3, PTK7, GAS1,* and *TGFB2* were significantly
285 upregulated in MPNST (**Figure 4C,D**) (|fold change| > 1.2 with adjusted $P < 0.05$),
286 consistent with previous reports [29, 30].

287 To elucidate the underlying pathways implicated in MPNST biology, we employed gene
288 set enrichment analysis (GSEA) on all genes differentially expressed between MPNST
289 and PNF in our cohort (**Figure 5**). Our analysis revealed alterations in six pathways within
290 Hallmark gene sets in MPNST versus PNF (adjusted $P < 0.01$) (**Figure 5A and**
291 **Supplementary Table 3**). Notably, the three most statistically significant pathways
292 involved upregulation of genes related to epithelial to mesenchymal transition, G2M
293 checkpoint, and myogenesis (25). Volcano plots depicting the involved genes are shown
294 in **Figure 5C,D,E** and heatmaps showing corresponding gene expression are shown in
295 **Figure S1 A,B,C.** Differentially expressed genes in these pathways are listed in
296 **Supplementary Tables 5, 6, 7** (adjusted $P < 0.01$ & $|FC|>1.2$). Additionally, we evaluated
297 pathways within Reactome gene sets (26) and identified 23 pathways that were
298 significantly different in MPNST vs. PNF (adjusted $P < 0.01$) (**Figure 5B and**
299 **Supplementary Table 4**). The three most statistically significant pathways were
300 associated with upregulation of genes involved in extracellular matrix (ECM) organization,
301 ECM degradation, and collagen degradation. Genes involved in these pathways are
302 shown in the volcano plots in **Figure 5F,G,H** and corresponding heatmaps of gene
303 expression are shown in **Figure S1 D,E,F**. Differentially expressed genes in these
304 pathways are listed in **Supplementary Tables 8, 9, 10** (adjusted $P < 0.01$ and $|FC| >$
305 1.2). Taken together, these findings suggest that MPNST exhibit a stronger mesenchymal
306 phenotype compared to PNF and that MPNST display increased ECM and activated
307 collagen degradation, potentially fostering tumor metastasis (27-29).
308

309 Further, as expected, we observed significant activation of several pathways previously
310 reported in the literature, which have clear links to MPNST biology including "H3K27me3-
311 associated", "metastasis-associated", and "mTOR-associated" pathways in MPNST
312 (**Supplementary Figure 2**). The single sample gene signature score for each of these
313 gene sets was significantly higher in MPNST versus PNF (**Supplementary Figure 3,**
314 $p<0.01$, Wilcox test).

315

316 **Discussion**

317 The Johns Hopkins NF1 biospecimen repository continues to preserve human tissues
318 from patient volunteers with NF1 undergoing surgical resection of any tumor or tissue for
319 the purposes of: 1) analysis of current and emerging diagnostic criteria for PNST; 2)
320 histologic and genomic characterization of these samples; and 3) creating a publicly
321 accessible, clinically and genomically annotated biospecimen repository that enables
322 therapeutic discovery research for NF1 across internal and external collaborators. The
323 goal of this project is to collect and store blood and tumor samples with accompanying
324 clinical data from 15-20 patients per year, including those with CNF, PNF, ANNUBP,
325 MPNST, and other NF1-associated neoplasms. We expect ongoing accrual at this rate,
326 given the volume of patients with NF1 seen at the Johns Hopkins Comprehensive NF
327 Center annually and patient interest. At present, we have banked 357 tissue samples
328 from a total of 183 unique patients with NF1 since the inception of the biospecimen
329 repository in 2016. These tissues, as well as cell lines and PDX models derived from
330 human tissue collection, and accompanying genomic data, represent publicly available
331 resources which are shared routinely with investigators upon request and free of charge,

332 to promote ongoing collaborative science focused on therapeutic discovery in the NF1
333 research community. A total of 187 samples (plus 71 corresponding normal samples)
334 from banked tissues have been sequenced, including whole exome on a total of 114
335 tumors (33 MPNST, 4 ANNUBP, 57 PNF, 9 diffuse NF, and 11 CNF) and RNA sequencing
336 on a total of 73 tumors (18 MPNST, 1 ANNUBP, 38 PNF, 9 diffuse NF, 6 CNF, nodular
337 NF 1). These genomic data are now publicly available to NF1 researchers to support an
338 array of important lines of scientific inquiry.

339 With this publication, we are releasing curated datasets of pre-processed WES and
340 RNAseq data that are ready for general re-use. These datasets are processed using
341 standardized publicly vetted processing workflows to reduce the efforts of data re-use.
342 Multiple variant callers and well known RNAseq quantification methods have been used
343 to generate datasets designed for maximum reusability. Additionally, these datasets are
344 well annotated with clinical metadata to enable appropriate re-analysis of the data as
345 needed.

346 Here, we present an analysis of 133 primary tumor samples sequenced using WES (82
347 total: MPNST 24; ANNUBP 1; PNF 45; diffuse NF 1; CNF 10, nodular NF 1) and RNAseq
348 (51 total: MPNST 14; ; PNF 29; diffuse NF 1; CNF 6, nodular NF 1). All existing RNAseq
349 and WES raw data for these tumors, as well as 4 MPNST cell lines and 2 PDX, have been
350 deposited in the NF Data Portal. The analysis presented herein is a preliminary
351 demonstration of whole exome and RNA sequencing analysis, meant to provide the
352 scientific community with an overview of available samples and their potential utility.
353 Several aspects of NF1 associated tumors were confirmed in this work. First, as expected
354 the *NF1* gene was the most commonly mutated gene in MPNST and PNF. We note that

355 variants in the *NF1* gene were detected in 35% of the samples rather than the expected
356 100% of samples, given their origin from people with the NF1 clinical condition. The
357 incomplete detection of *NF1* variants is likely due to a combination of factors: 1) detection
358 sensitivity of WES is limited by lack of coverage on intronic regions; 2) WES analysis is
359 not sensitive for microdeletion or copy number changes; 3) somatic variant analysis will
360 be unable to call a variant in tumor samples with loss of heterozygosity of the *NF1* gene
361 since both tumor and normal samples will have the same genetic variants; and 4) variants
362 will not be called with high confidence in tumor samples with low tumor cell purity. These
363 limitations suggest that targeted high depth clinical sequencing may be more sensitive in
364 capturing *NF1* variants in these tumors (30-33). The inability to detect *NF1* variants in all
365 samples despite a clinical diagnosis highlights the current limitations with WES
366 technology and the many ways in which the *NF1* gene can be altered resulting in the
367 variable clinical presentation of NF1. Second, we demonstrated that the main pathway in
368 which oncogenic variants are detected in PNF involves RTK-RAS signaling. In MPNST
369 samples, the bulk of genetic variants affected not only RAS-RTK signaling, but also the
370 Hippo and TP53 pathways. This finding agrees with prior work indicating that progression
371 to MPNST involves genetic changes beyond those affecting RAS biology. Third, through
372 our RNAseq analysis, we demonstrated an increase in activation of pathways involving
373 extracellular matrix organization and degradation as well as cell cycle checkpoints in
374 MPNST compared to PNF. These biologic insights are in accordance with the aggressive
375 nature of MPNST. While it would be validating to find that these expression signatures
376 are even further enhanced in recurrent or metastatic MPNST, compared to those that

377 remained localized or did not recur, the numbers of samples included herein limit the
378 ability to achieve meaningful significance in this analysis.

379 In summary, this study showcases the release of WES and RNA sequencing data from
380 human NF1 associated peripheral nerve sheath tumors, along with pre-processed data
381 in analysis-ready format to facilitate reuse without additional efforts. Our exploration of
382 the data confirms findings that are concordant with available literature regarding the
383 genomics of NF1 associated tumors and validate this biospecimen repository as a rich
384 resource for the scientific community.

385

386

Author Contribution:

Authorship was determined using ICMJE recommendations.

Conceptualization: JB, ACH, JOB, CAP

Data curation: JB, YL, SCM, AC, KP, LZ

Formal Analysis: JB, YL, AJS

Funding acquisition: JOB, RJA, JB, CAP

Investigation: all authors

Methodology: JB, YL, SCM, AC, KP, LZ, AJS

Project administration: ACH, JOB, CAP

Resources: JB, YL,

Software: JB, YL, AJS

Supervision: ACH, JOB, RJA, CAP

Validation: not applicable

Visualization: JB, YL, AJS

Writing: original draft: JB, YL, SCM, AJS, ACH, CAP

Writing: review & editing: all authors

Code and Data Availability:

All the figures in this article are generated through reproducible R code which is available on Github (https://github.com/nf-osi/biobank-release-2). All raw data files were processed for analyses using nf-core pipelines available through https://nf-co.re/ and specific versions used in this study have been noted in the methods section. All the raw data files are available on Synapse.org (WES: dataset format: syn53132831.1, RNAseq: dataset format: syn53133024.1). All processed data visualized in this article are available on Synapse.org (WES: Somatic Variants Mutect2 – syn53149144.2, Somatic Variants Strelka2 – syn53149128.1, RNAseq quantification files: syn53140231.1, RNAseq counts files: syn53141534.1).

418 **Table 1.** List of samples with RNAseq and/ or WES, including age (in in years by range),
419 sex, and type of NF1-associated PNST, included in the current analysis of samples and
420 for which sequencing data are available (inclusive of samples released with Pollard et al,
421 *Sci Data*, 2020 (17)). WES and RNA sequencing are indicated by an “x” and “y” symbol,
422 respectively.

Patient ID	Age at collection	Sex	Normal (blood)	CNF	Nodular NF	Diffuse NF	PNF	ANNUBP	MPNST
JH-2-001		F	x				x, y		
JH-2-002		M	x*				x*, y		x*, y*
JH-2-003		M					y		
JH-2-004		F	x				x, y		
JH-2-006		F	x				x, y		
JH-2-007		M		y					
JH-2-009		M	x				x		x*, y*
JH-2-010		M	x				x, y		
JH-2-012		M	x				x, y		
JH-2-013		F	x				x, y		
JH-2-014		M	x				x, y		
JH-2-015		M	x				x*		x, y
JH-2-016		F	x				x, y		x, y
JH-2-017		M	x				x, y		
JH-2-019		F	x				x, y		
JH-2-021		M	x		x, y				
JH-2-023		M	x				x		x*, y*
JH-2-026		F	x				x, y		
JH-2-029		M	x	x, y					
JH-2-031		M	x						x*, y*
JH-2-032		F	x$				x, y		
JH-2-044		F	x$				x, y		
JH-2-045		F	x$	x, y					x
JH-2-054		M	x				x*		
JH-2-055		F	x				x		x*, y
JH-2-057		F	x				x		
JH-2-060		M	x				x*		
JH-2-061		F	x				x, y		
JH-2-068		M	x				x*, y		
JH-2-072		M	x						x

JH-2-074		M	x				x		
JH-2-075		M	x	x			x, y		
JH-2-076		F	x	x			x, y		
JH-2-077		F	x				x		
JH-2-079		F	x						x*, y*
JH-2-082		M	x				x, y		x, y
JH-2-085		M	x				x, y		
JH-2-086		M	x				x, y		
JH-2-089		M	x				x, y		
JH-2-090		M	x			x, y		x	
JH-2-091		M	x				x, y		
JH-2-092		F	x	x*,y*					
JH-2-093		M	x	x*, y					
JH-2-094		F	x				x, y		x
JH-2-095		M	x				x		
JH-2-096		M	x						x
JH-2-099		F	x				x		
JH-2-100		F	x				x, y		
JH-2-102		F	x	x					
JH-2-103		M	x						x
JH-2-104		M	x				x, y		
JH-2-107		M					y		
JH-2-111		M	x				x, y		
JH-2-113		F	x						x

Key:	Age (years)
	0-10
	11-20
	21-30
	31-40
	41-50
	>50

x WES

y RNAseq

* more than one same sample was sequenced \

$ normal tissue was sequenced (not blood)

429 **Figure legends**

430 **Figure 1.** Relatedness between WES and RNAseq samples. A) A scatter plot of the
431 relatedness coefficient versus the number of sites with the same genotype between two
432 samples, as calculated by Somalier. Comparisons between two samples annotated as
433 being from the same individual are shown in red and those from different individuals are
434 colored blue. The shape of the point indicates the tissue of origin of the samples
435 compared. The vertical line indicates the cutoff value (≥ 0.9) used to determine whether
436 two samples are from the same individual based on the relatedness coefficient. B) A
437 heatmap of the relatedness coefficient of pairwise comparisons between WES and
438 RNAseq data from primary human samples as shown in (A), excluding cell lines or
439 xenografts derived from these samples. RNAseq samples are shown on the x-axis and
440 WES samples are on the y-axis. The relatedness coefficient is displayed according to the
441 color scale bar shown in the figure.

442

443 **Figure 2.** Summary of somatic variants detected in PNST (10 CNF, 1 nodular NF, 1
444 diffuse NF, 36 PNF, 1 ANNUBP, and 19 MPNST). A) Oncoplot of variants in selected
445 genes of interest (see methods) in the cohort of biobank patients detected using Strelka2.
446 B) Oncoplot of variants in selected genes of interest in the cohort of biobank patients
447 detected using Mutect2. C) Oncoplot of the consensus of variants in selected genes of
448 interest in the cohort of biobank patients detected using SomaticSeq. The status of *NF1*
449 functional inactivation in the patients was determined using either clinical genetic testing
450 or clinical diagnosis and is provided using colored bars in the bottom panel of each
451 oncoplot. (D, F, H) Barplot showing the number of variants per sample, by SNV class

452 (missense, nonsense, splice site, nonstop or translation start site). (E, G, I) Bar plot of
453 the top 10 genes with variants of moderate or high impact in the cohort.

454

455 **Figure 3.** Somatic variants in plexiform neurofibromas and malignant peripheral nerve
456 sheath tumors. A) Lollipop plot showing different positions and protein consequences of
457 variants in the *NF1* gene as detected in this cohort using WES. B) Schematic
458 representation of neurofibromin protein with various functional domains. Domains: CBD
459 caveolin-binding domains, CSRD cysteine and serine rich domain, CTD C-terminal
460 domain, GRD RAS-GAP-related domain, HLR HEAT-like repeats, LRD leucine-rich
461 domain, PH pleckstrin homology, SBR syndecan-binding region, TBD tubulin-binding
462 domain (adapted from Mo et al, 2022, with permission) (23). C) Plots showing top
463 oncogenic pathways that are affected by genomic variants in PNF samples. D) Plots
464 showing top oncogenic pathways that are affected by genomic variants in MPNST
465 samples.

466

467 **Figure 4.** A) Principal component analysis (PCA) plot illustrating the distribution of all
468 tumors based on RNAseq gene counts. B) Uniform manifold approximation and projection
469 (UMAP) plot depicting the tumor distribution using RNAseq gene counts. C) Volcano plot
470 highlighting differentially expressed genes associated with MPNST compared to PNF
471 ($|FC| > 1.2$ and $P < 0.05$). D) Heatmap displaying the expression patterns of differentially
472 expressed genes associated with MPNST in comparison to PNF.

473

474 **Figure 5.** Pathway enrichment analysis in MPNST versus PNF (RNAseq). A) Dot plot
475 showing significant pathways in Hallmark gene sets in MPNST versus PNF (adjusted *P*
476 < 0.01). B) Dot plot showing significant pathways in Reactome gene sets in MPNST
477 versus PNF (adjusted *P* < 0.01 & |FC| > 1.2). C,D,E) Volcano plots for top three significant
478 Hallmark pathways (epithelial-mesenchymal transition, G2M checkpoint, and
479 myogenesis) highlighting differentially expressed genes in MPNST vs. PNF (adjusted *P*
480 < 0.01). F,G,H) Volcano plots for top three significant Reactome pathways (extracellular
481 matrix organization, collagen degradation, and extracellular matrix degradation)
482 highlighting differentially expressed genes in MPNST vs. PNF (adjusted *P* < 0.05 &
483 |FC|>1.2).
484
485
486

487 **Supplementary Materials**

488 **Supplementary figures**

489 **Supplementary Figure 1.** Heatmap of gene expression in significant pathways in
490 MPNST vs. PNF tumors. Heatmap illustrating the expression of differentially expressed
491 genes in significant Hallmark pathways (A,B,C – epithelial-mesenchymal transition, G2M
492 check-point, and myogenesis) and Reactome pathways (D,E,F – extra cellular matrix
493 organization, collagen degradation, and extracellular matrix degradation) in MPNST vs.
494 PNF tumors.

495 **Supplementary Figure 2.** Volcano plot and heatmap of gene expression in three
496 pathways with established biological activity in MPNST vs. PNF tumors ("H3K27me3-
497 associated", "metastasis-associated", and "mTOR-associated"). Genes with fold change
498 > 2 (adjusted $P < 0.05$) are labeled in the volcano plots (A,B,C) and gene expression
499 patterns of differentially expressed genes in MPNST vs. PNF (adjusted $P < 0.05$ & |FC|
500 >1.2|) are shown in heatmap (D,E,F).

501 **Supplementary Figure 3.** Boxplot of the single sample gene signature score (singscore)
502 in three significant pathways in MPNST vs. PNF tumors as shown in Fig S2. These three
503 pathways as shown in aggregate score, are upregulated in MPNST compared to PNF
504 (A,B,C, Wilcox test, $P < 0.05$).

505

506

507 **Supplementary tables**

508 **Supplementary table 1.** Tumor samples with paired-normal samples sequenced in two
509 different batches.

510 **Supplementary table 2.** HLA typing results of WES data.

511 **Supplementary table 3.** Hallmark pathways that meet significance thresholds in
512 MPNST vs. PNF (adjusted $P < 0.05$).

513 **Supplementary table 4.** Reactome pathways that meet significance thresholds in
514 MPNST vs. PNF (adjusted $P < 0.05$).

515 **Supplementary table 5.** Differentially expressed genes in "HALLMARK_EPITHELIAL_
516 MESENCHYMAL_TRANSITION" in MPNST vs. PNF (adjusted $P < 0.05$, $|FC| > 1.2$).

517 **Supplementary table 6.** Differentially expressed genes in
518 "HALLMARK_G2M_CHECKPOINT" in MPNST vs. PNF (adjusted $P < 0.05$ & $|FC| >$
519 1.2).

520 **Supplementary table 7.** Differentially expressed genes in
521 "HALLMARK_MYOGENESIS" in MPNST vs. PNF (adjusted $P < 0.05$ & $|FC| > 1.2$).

522 **Supplementary table 8.** Differentially expressed genes n
523 "REACTOME_EXTRACELLULAR_ MATRIX_ORGANIZATION" in MPNST vs. PNF
524 (adjusted $P < 0.05$ & $|FC| > 1.2$).

525 **Supplementary table 9.** Differentially expressed genes in "REACTOME_COLLAGEN_
526 DEGRADATION" in MPNST vs. PNF (adjusted $P < 0.05$ & $|FC| > 1.2$).

527 **Supplementary table 10**. Differentially expressed genes in
528 "REACTOME_DEGRADATION_OF_ THE_EXTRACELLULAR_MATRIX" in MPNST vs.
529 PNF (adjusted $P < 0.05$ & $|FC| > 1.2$).

530

531

532

533

534

535 **Acknowledgements**

536 Funding: This publication was supported by funding from the Neurofibromatosis
537 Therapeutic Acceleration Program (NTAP) (https://www.n-tap.org/) at the Johns Hopkins
538 University School of Medicine, to C.A.P. and to A.C.H (Grant ID: 212027, 152009). Its
539 contents are solely the responsibility of the authors and do not necessarily represent the
540 official views of The Johns Hopkins University School of Medicine. Bioinformatics
541 analyses were supported by funding from NTAP to R.J.A. and J.B (Grant ID: 229101).
542 The funders were involved in reviewing the manuscript and providing feedback.

543

544

References

1. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM 2013 [
2. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
3. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8.
4. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1110):11 0 1- 0 33.
5. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15(8):591-4.
6. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.
7. Kandoth C, Gao JJ, Qwang M, Mattioni M, Struck A, Boursin Y, et al. mskcc/vcf2maf: vcf2maf v1.6.16 2018 [Available from: doi:10.5281/zenodo.593251.
8. Garcia M, Juhos S, Larsson M, Olason PI, Martin M, Eisfeldt J, et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res. 2020;9:63.
9. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. Brief Bioinform. 2013;14(2):144-61.
10. Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, et al. An ensemble approach to accurately detect somatic mutations using SomaticSeq. Genome Biol. 2015;16(1):197.
11. Patel HE, P., Peltzer, A. Botvinnik, O., Manning, J., Sturm, G., Garcia, M. Moreno, D. Vemuri, P., nf-core bot, Pantano, L., Binzer-Panchal, M., Zepper, M., Syme, R., Kelly, G., Hanssen, F., rfenouil, Espinosa-Carrasco, J., marchoeppner, Miller, E., Zhou, P., Guinchard, S., Gabarnet, G., Mertes, C., Straub, D., DiTommaso, P. nf-core/rnaseq: nf-core/rnaseq v3.13.2 - Cobalt Colt 2023 [updated November 21, 2023. Version 3.13.2:[
12. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-9.
13. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882-3.
14. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction: arXiv; 2018 [
15. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. BMC Bioinformatics. 2018;19(1):404.

589 16. Pedersen BS, Bhetariya PJ, Brown J, Kravitz SN, Marth G, Jensen RL, et al.
590 Somalier: rapid relatedness estimation for cancer and germline studies using efficient
591 genome sketches. Genome Med. 2020;12(1):62.
592 17. Pollard K, Banerjee J, Doan X, Wang J, Guo X, Allaway R, et al. A clinically and
593 genomically annotated nerve sheath tumor biospecimen repository. Sci Data.
594 2020;7(1):184.
595 18. Zhang M, Wang Y, Jones S, Sausen M, McMahon K, Sharma R, et al. Somatic
596 mutations of SUZ12 in malignant peripheral nerve sheath tumors. Nat Genet.
597 2014;46(11):1170-2.
598 19. Lee W, Teckie S, Wiesner T, Ran L, Prieto Granada CN, Lin M, et al. PRC2 is
599 recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath
600 tumors. Nat Genet. 2014;46(11):1227-32.
601 20. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2
602 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies.
603 Nature. 2014;514(7521):247-51.
604 21. Sohier P, Luscan A, Lloyd A, Ashelford K, Laurendeau I, Briand-Suleau A, et al.
605 Confirmation of mutation landscape of NF1-associated malignant peripheral nerve
606 sheath tumors. Genes Chromosomes Cancer. 2017;56(5):421-6.
607 22. Cortes-Ciriano I, Steele CD, Piculell K, Al-Ibraheemi A, Eulo V, Bui MM, et al.
608 Genomic Patterns of Malignant Peripheral Nerve Sheath Tumor (MPNST) Evolution
609 Correlate with Clinical Outcome and Are Detectable in Cell-Free DNA. Cancer Discov.
610 2023;13(3):654-71.
611 23. Mo J, Moye SL, McKay RM, Le LQ. Neurofibromin and suppression of
612 tumorigenesis: beyond the GAP. Oncogene. 2022;41(9):1235-51.
613 24. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al.
614 Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell. 2018;173(2):321-37
615 e10.
616 25. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The
617 Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst.
618 2015;1(6):417-25.
619 26. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al.
620 The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50(D1):D687-
621 D92.
622 27. Brockman QR, Scherer A, McGivney GR, Gutierrez WR, Voigt AP, Isaacson AL,
623 et al. PRC2 loss drives MPNST metastasis and matrix remodeling. JCI Insight.
624 2022;7(20).
625 28. Thomson CS, Pundavela J, Perrino MR, Coover RA, Choi K, Chaney KE, et al.
626 WNT5A inhibition alters the malignant peripheral nerve sheath tumor microenvironment
627 and enhances tumor growth. Oncogene. 2021;40(24):4229-41.
628 29. Errico A, Stocco A, Riccardi VM, Gambalunga A, Bassetto F, Grigatti M, et al.
629 Neurofibromin Deficiency and Extracellular Matrix Cooperate to Increase Transforming
630 Potential through FAK-Dependent Signaling. Cancers (Basel). 2021;13(10).
631 30. Koboldt DC. Best practices for variant calling in clinical sequencing. Genome
632 Med. 2020;12(1):91.

633 31. Viray H, Li K, Long TA, Vasalos P, Bridge JA, Jennings LJ, et al. A prospective,
634 multi-institutional diagnostic trial to determine pathologist accuracy in estimation of
635 percentage of malignant cells. Arch Pathol Lab Med. 2013;137(11):1545-9.
636 32. Smits AJ, Kummer JA, de Bruin PC, Bol M, van den Tweel JG, Seldenrijk KA, et
637 al. The estimation of tumor cell percentage for molecular testing by pathologists is not
638 accurate. Mod Pathol. 2014;27(2):168-74.
639 33. Dudley J, Tseng LH, Rooper L, Harris M, Haley L, Chen G, et al. Challenges
640 posed to pathologists in the detection of KRAS mutations in colorectal cancers. Arch
641 Pathol Lab Med. 2015;139(2):211-8.

642

Figure 1

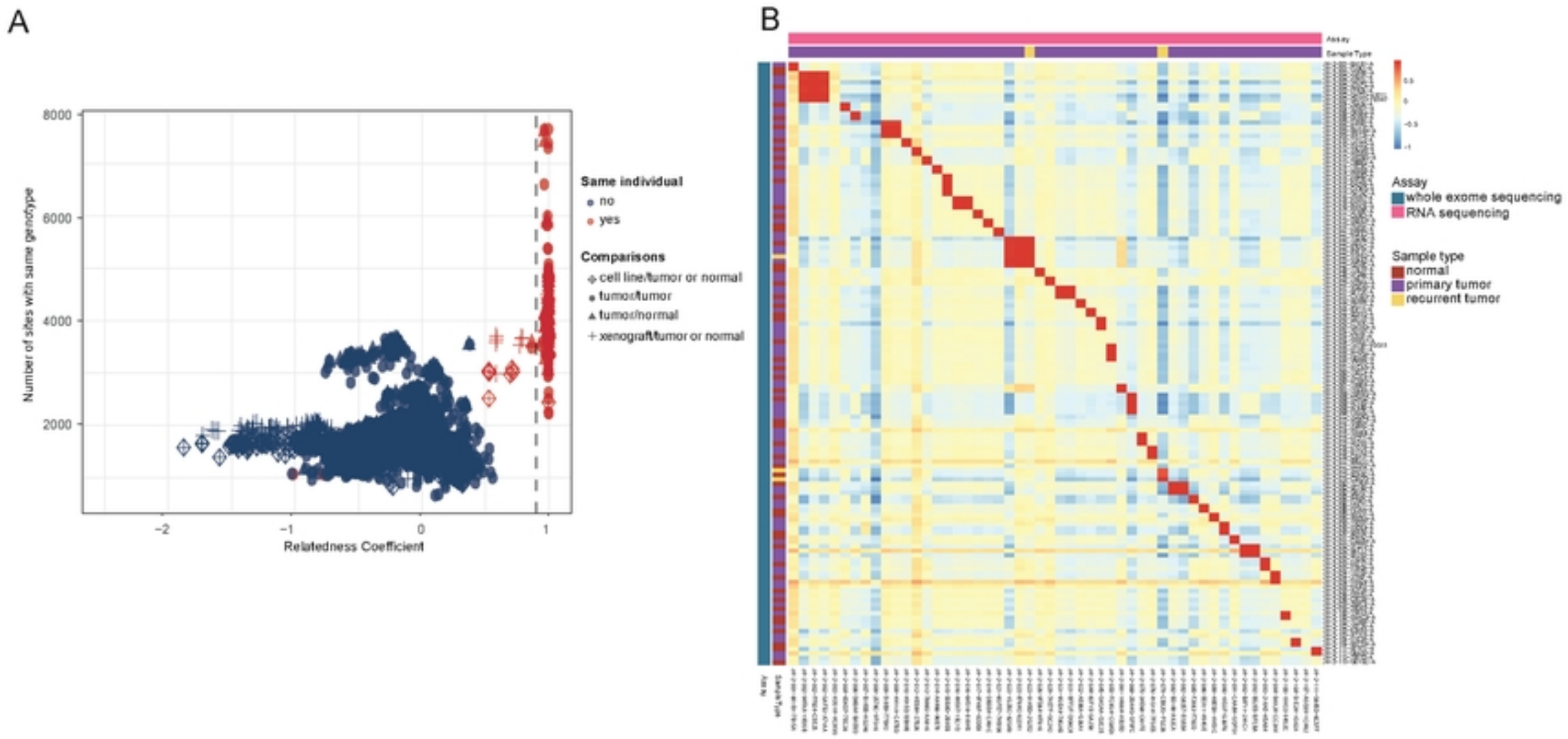
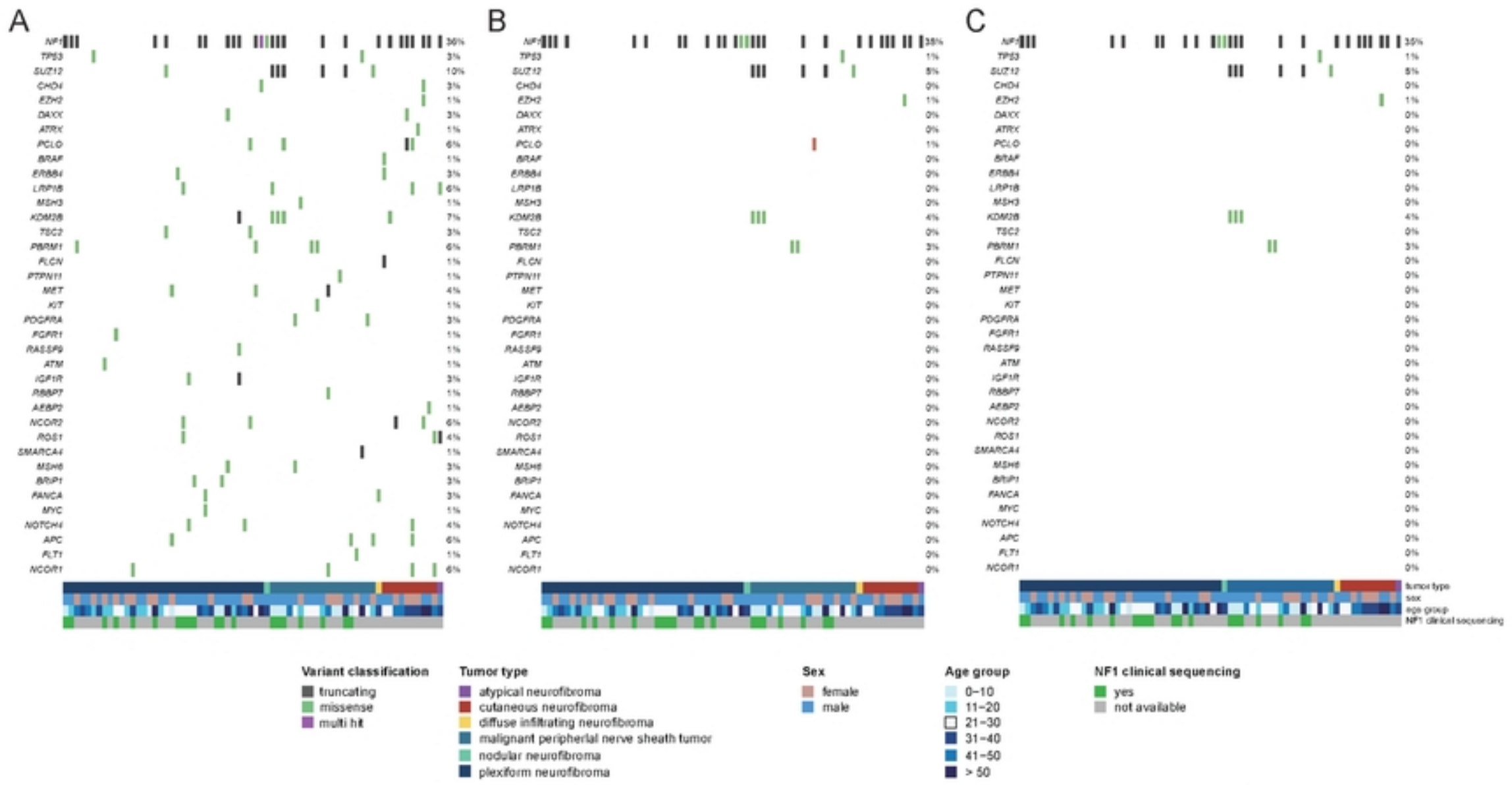

A
B
Number of sites with same genotype
8000
6000
4000
2000
−2
−1
0
1
Relatedness Coefficient
Same individual
no
yes
Comparisons
cell line/tumor or normal
tumor/tumor
tumor/normal
xenograft/tumor or normal
Assay
whole exome sequencing
RNA sequencing
Sample type
normal
primary tumor
recurrent tumor

Figure 2


A
B
C
Variant classification
truncating
missense
multi hit
Tumor type
atypical neurofibroma
cutaneous neurofibroma
diffuse infiltrating neurofibroma
malignant peripherial nerve sheath tumor
nodular neurofibroma
plexiform neurofibroma
Sex
female
male
Age group
0–10
11–20
21–30
31–40
41–50
> 50
NF1 clinical sequencing
yes
not available

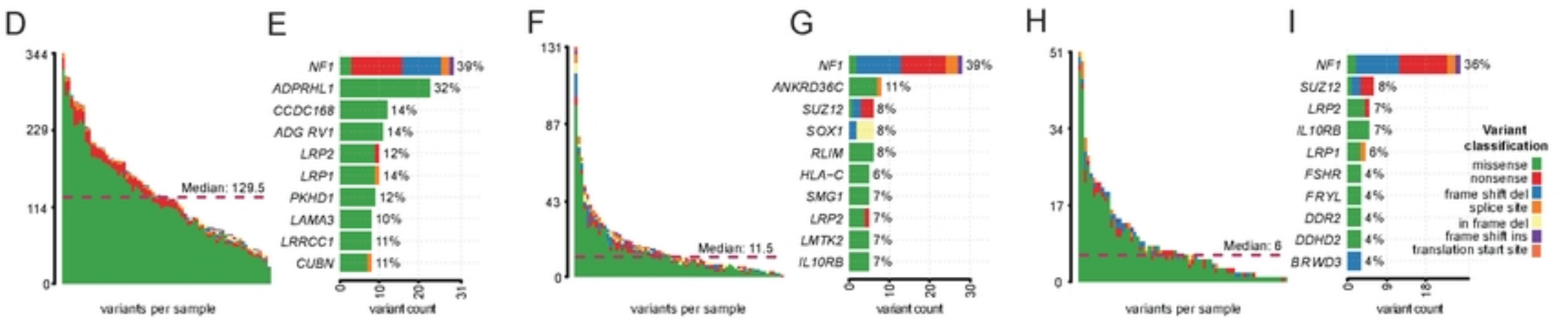

D
Median: 129.5
variants per sample
E
NF1 39%
ADPRHL1 32%
CCDC168 14%
ADG RV1 14%
LRP2 12%
LRP1 14%
PKHD1 12%
LAMA3 10%
LRRCC1 11%
CUBN 11%
variant count
F
Median: 11.5
variants per sample
G
NF1 39%
ANKRD36C 11%
SUZ12 8%
SOX1 8%
RLIM 8%
HLA-C 6%
SMG1 7%
LRP2 7%
LMTK2 7%
IL10RB 7%
variant count
H
Median: 6
variants per sample
I
NF1 36%
SUZ12 8%
LRP2 7%
IL10RB 7%
LRP1 6%
FSHR 4%
FRYL 4%
DDR2 4%
DDHD2 4%
BRWD3 4%
variant count
Variant classification
missense
nonsense
frame shift del
splice site
in frame del
frame shift ins
translation start site

Figure 3

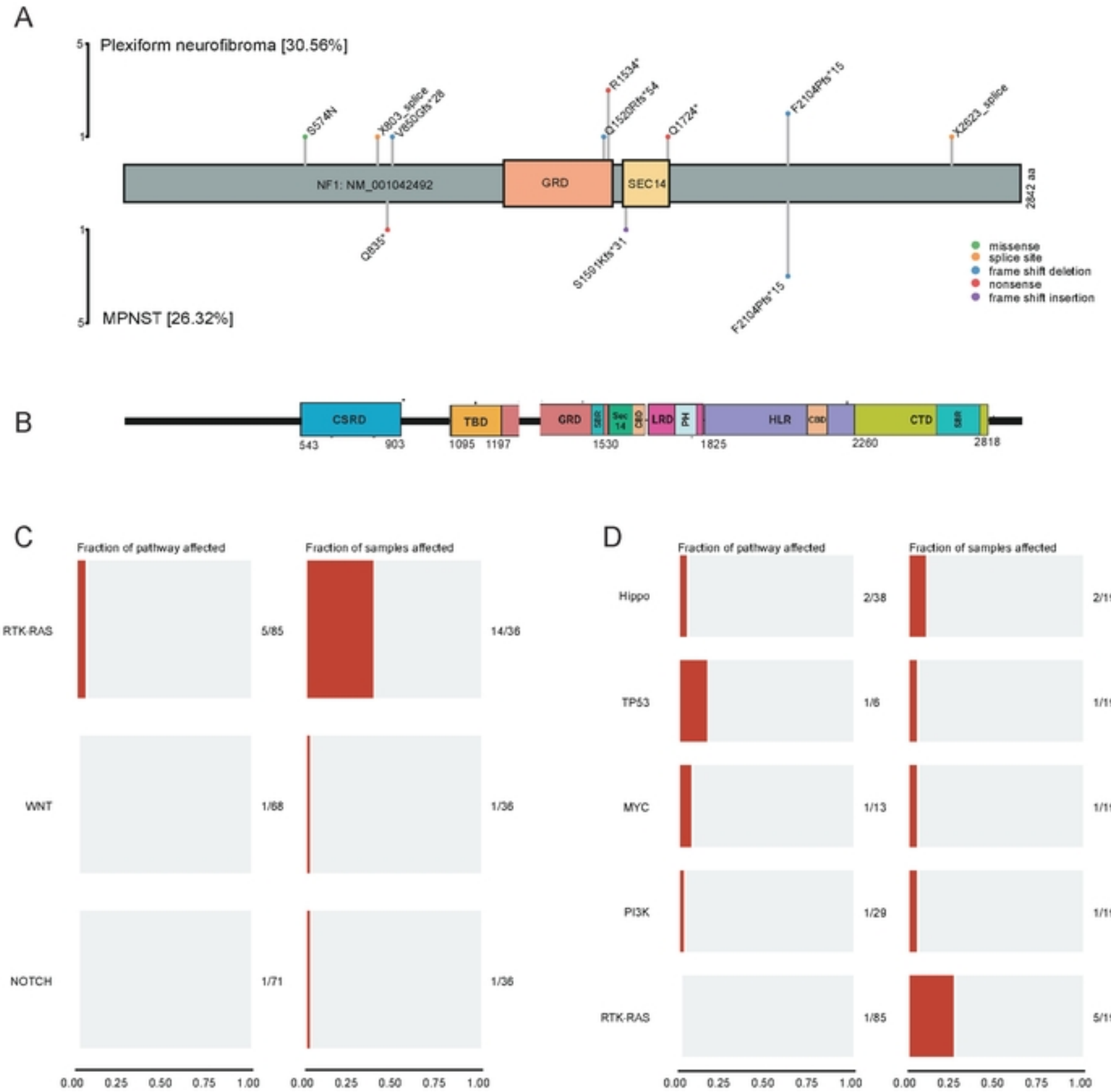
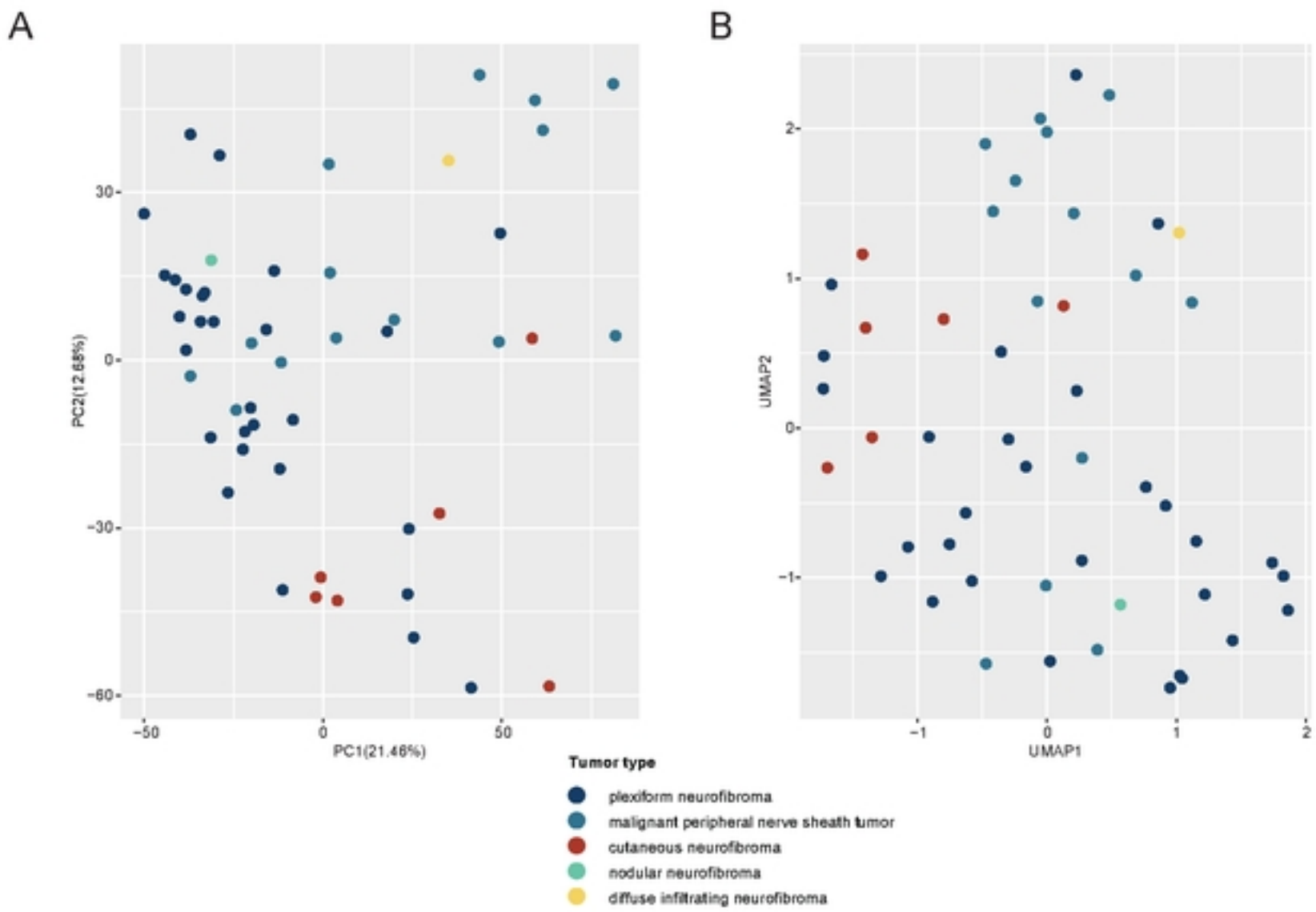

A
Plexiform neurofibroma [30.56%]
MPNST [26.32%]
S574N
X803_splice
V650Gfs*28
R1534*
Q1520Rfs*54
Q1724*
F2104Pfs*15
X2623_splice
NF1: NM_001042492
GRD
SEC14
2842 aa
Q835*
S1591Kfs*31
F2104Pfs*15
missense
splice site
frame shift deletion
nonsense
frame shift insertion
B
CSRD
TBD
GRD
SEC 14
LRD
PH
HLR
CTD
543
903
1095
1197
1530
1825
2260
2818
C
Fraction of pathway affected
Fraction of samples affected
RTK-RAS
5/85
14/36
WNT
1/68
1/36
NOTCH
1/71
1/36
0.00
0.25
0.50
0.75
1.00
D
Fraction of pathway affected
Fraction of samples affected
Hippo
2/38
2/19
TP53
1/6
1/19
MYC
1/13
1/19
PI3K
1/29
1/19
RTK-RAS
1/85
5/19

Figure 4


A
B
PC2(12.68%)
PC1(21.46%)
UMAP2
UMAP1
Tumor type
plexiform neurofibroma
malignant peripheral nerve sheath tumor
cutaneous neurofibroma
nodular neurofibroma
diffuse infiltrating neurofibroma

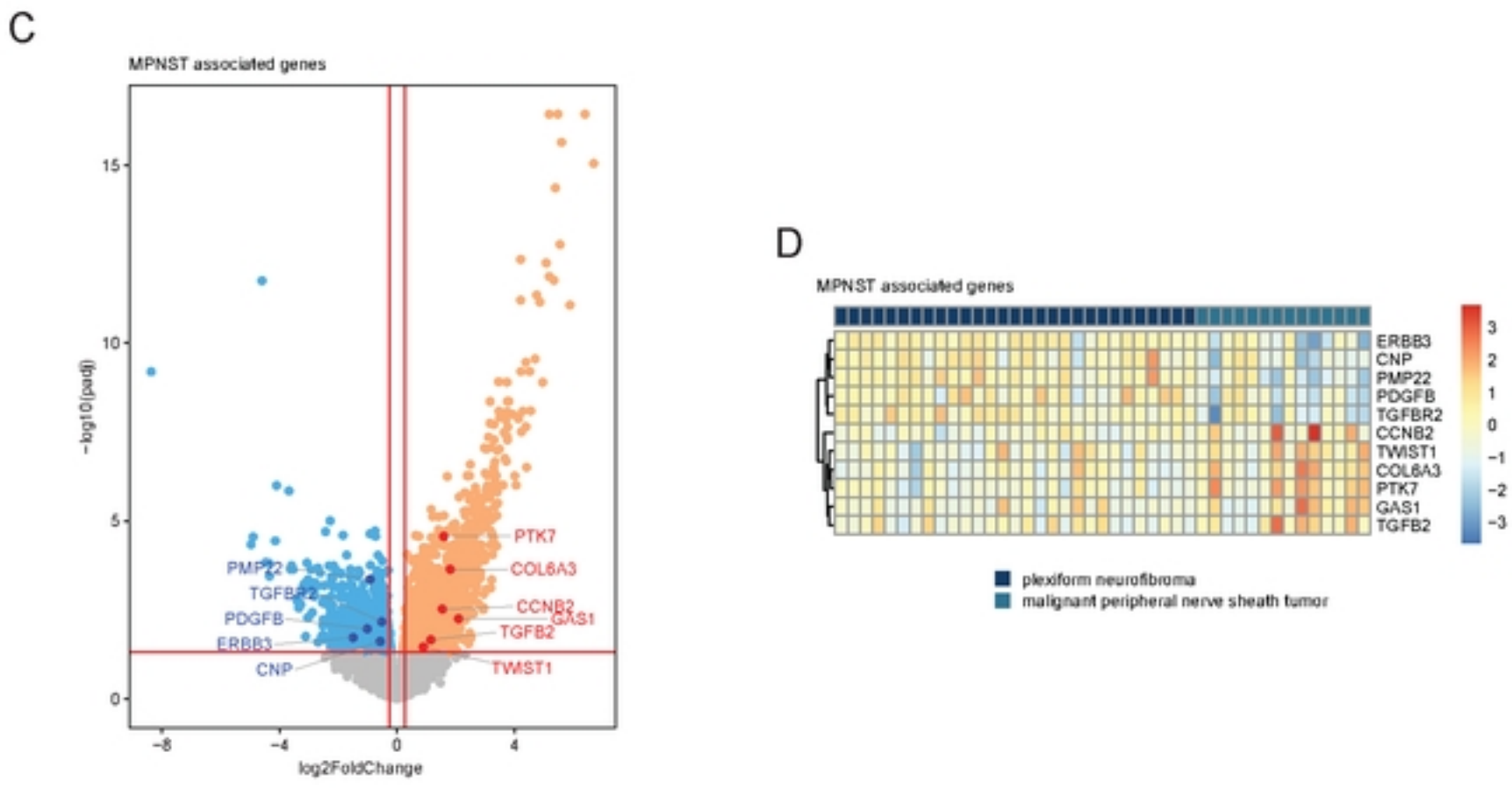
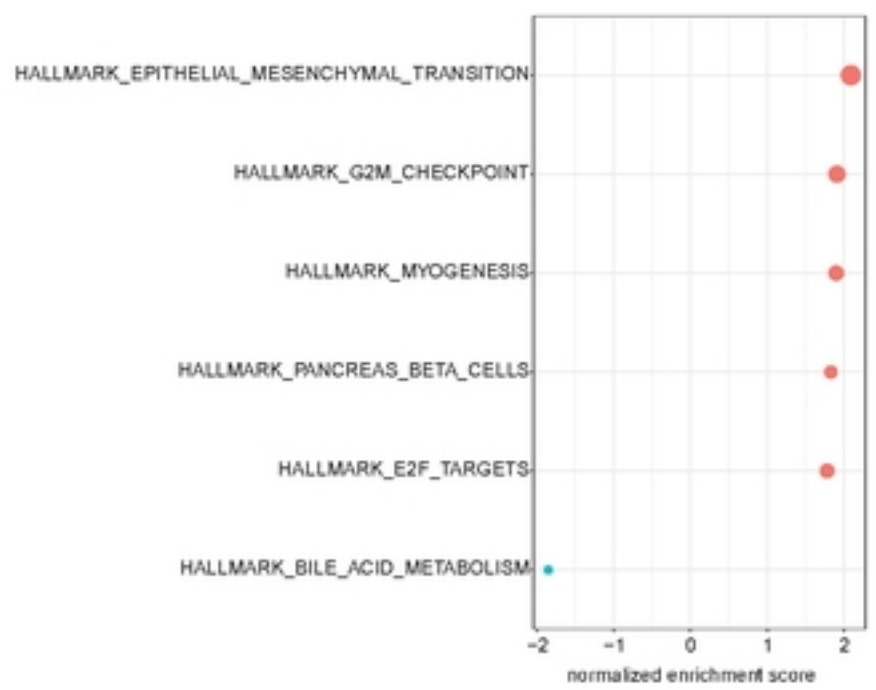
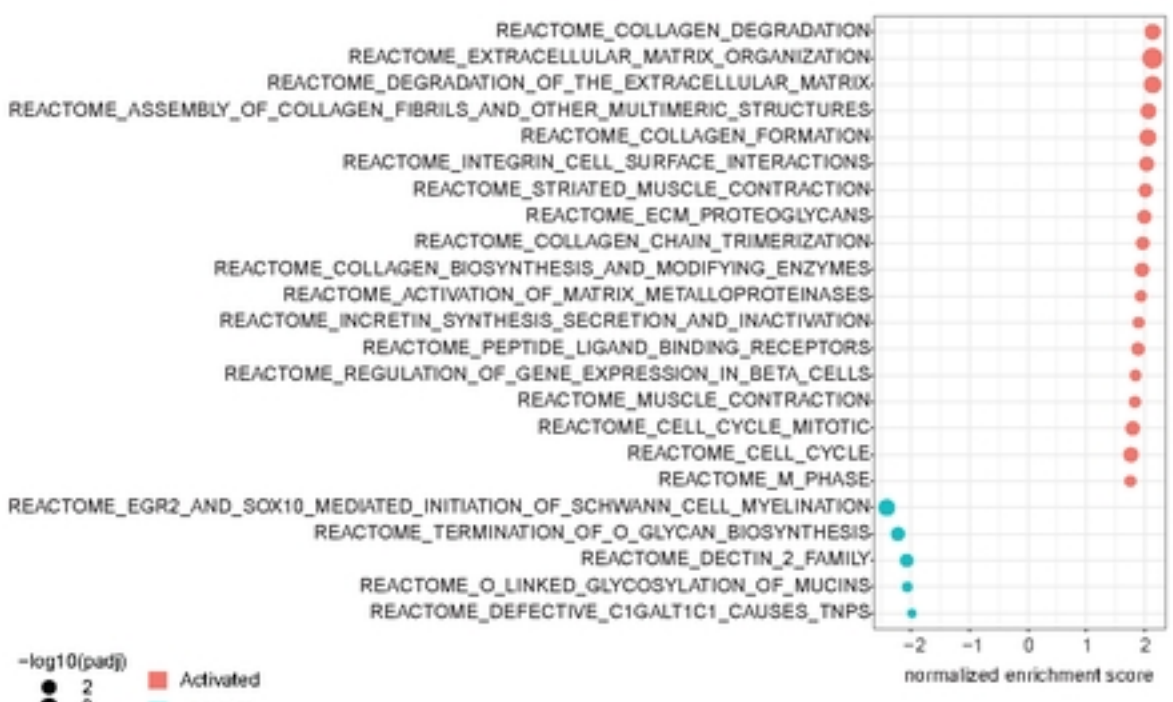
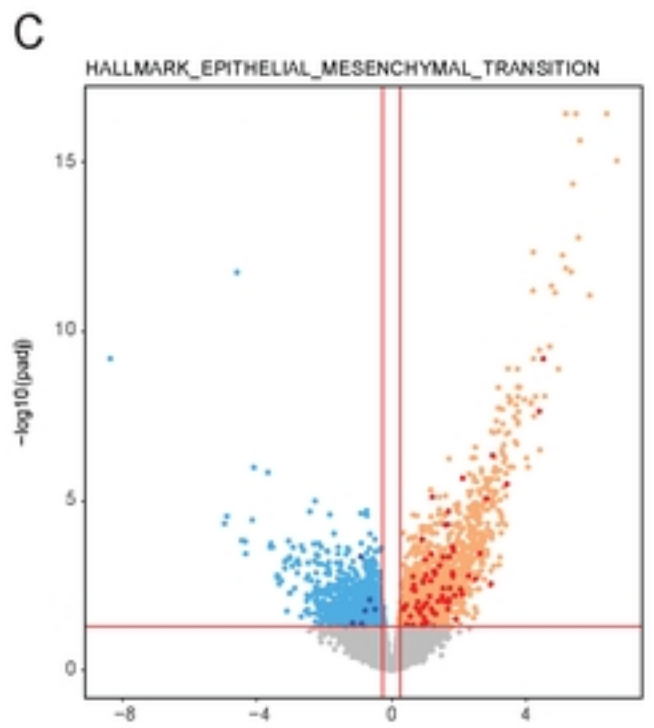
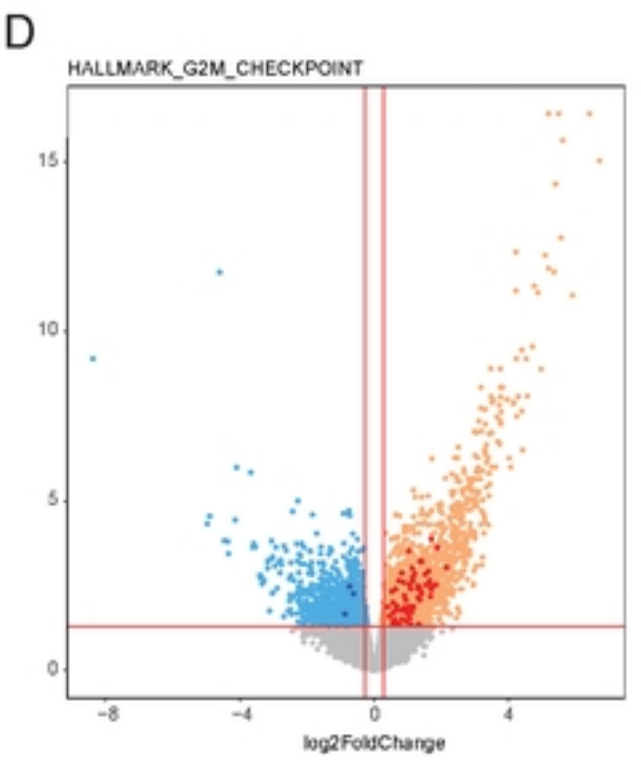

C
MPNST associated genes
−log10(padj)
log2FoldChange
PTK7
COL6A3
CCNB2
GAS1
TGFB2
TWIST1
PMP22
TGFBR2
PDGFB
ERBB3
CNP
D
MPNST associated genes
ERBB3
CNP
PMP22
PDGFB
TGFBR2
CCNB2
TWIST1
COL6A3
PTK7
GAS1
TGFB2
plexiform neurofibroma
malignant peripheral nerve sheath tumor

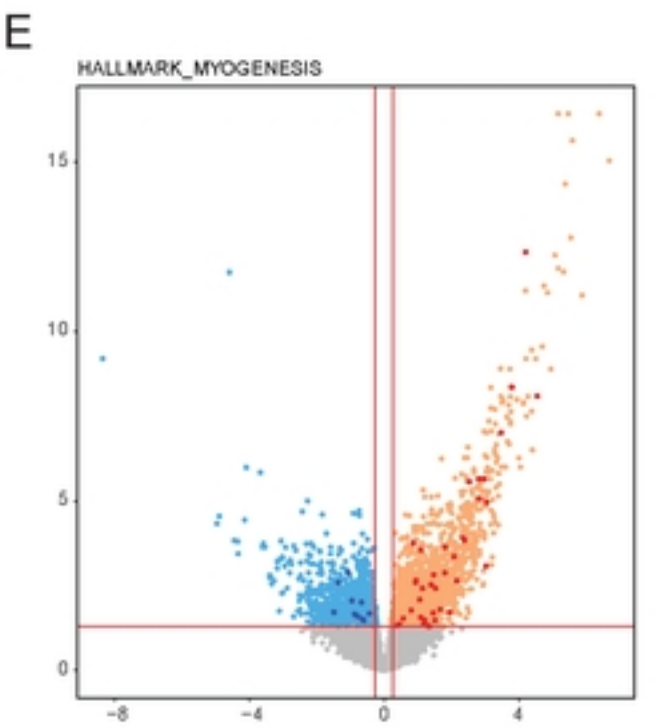
Figure 5

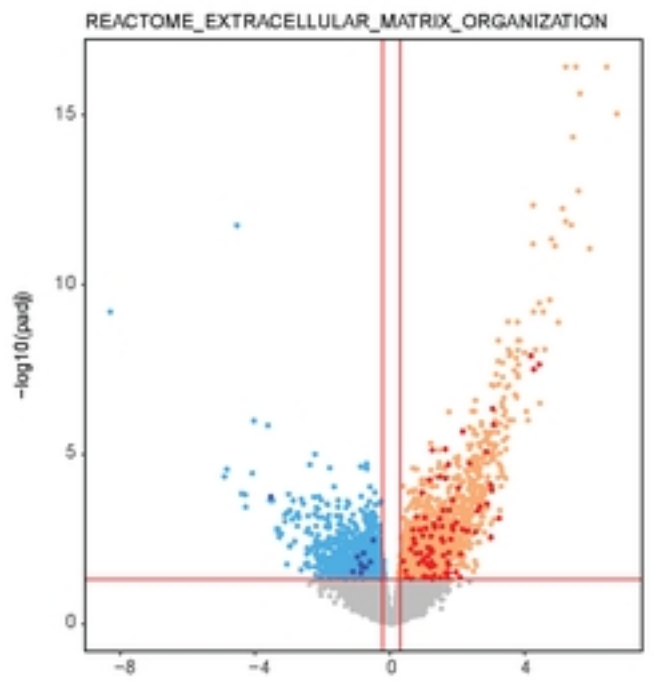

A
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_G2M_CHECKPOINT
HALLMARK_MYOGENESIS
HALLMARK_PANCREAS_BETA_CELLS
HALLMARK_E2F_TARGETS
HALLMARK_BILE_ACID_METABOLISM
−2 −1 0 1 2
normalized enrichment score

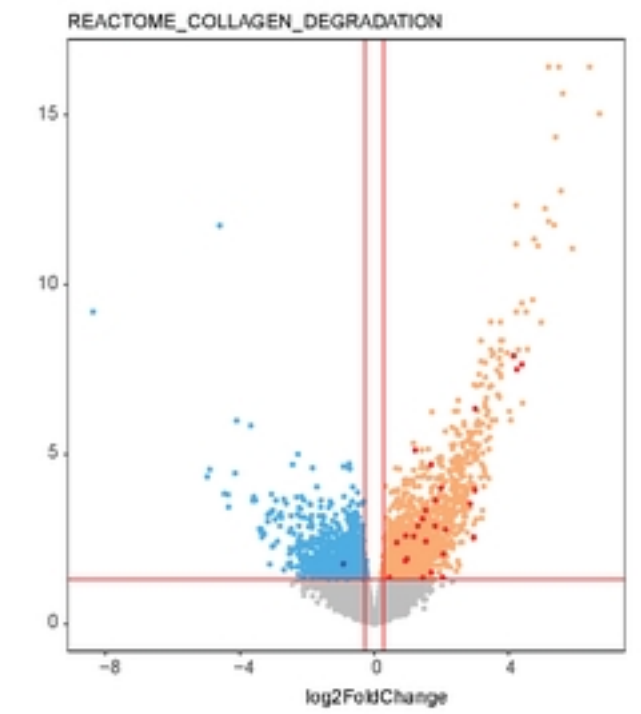

B
REACTOME_COLLAGEN_DEGRADATION
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX
REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES
REACTOME_COLLAGEN_FORMATION
REACTOME_INTEGRIN_CELL_SURFACE_INTERACTIONS
REACTOME_STRIATED_MUSCLE_CONTRACTION
REACTOME_ECM_PROTEOGLYCANS
REACTOME_COLLAGEN_CHAIN_TRIMERIZATION
REACTOME_COLLAGEN_BIOSYNTHESIS_AND_MODIFYING_ENZYMES
REACTOME_ACTIVATION_OF_MATRIX_METALLOPROTEINASES
REACTOME_INCRETIN_SYNTHESIS_SECRETION_AND_INACTIVATION
REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS
REACTOME_REGULATION_OF_GENE_EXPRESSION_IN_BETA_CELLS
REACTOME_MUSCLE_CONTRACTION
REACTOME_CELL_CYCLE_MITOTIC
REACTOME_CELL_CYCLE
REACTOME_M_PHASE
REACTOME_EGR2_AND_SOX10_MEDIATED_INITIATION_OF_SCHWANN_CELL_MYELINATION
REACTOME_TERMINATION_OF_O_GLYCAN_BIOSYNTHESIS
REACTOME_DECTIN_2_FAMILY
REACTOME_O_LINKED_GLYCOSYLATION_OF_MUCINS
REACTOME_DEFECTIVE_C1GALT1C1_CAUSES_TNPS
−2 −1 0 1 2
normalized enrichment score
−log10(padj)
2
3
4
Activated
Inhibited

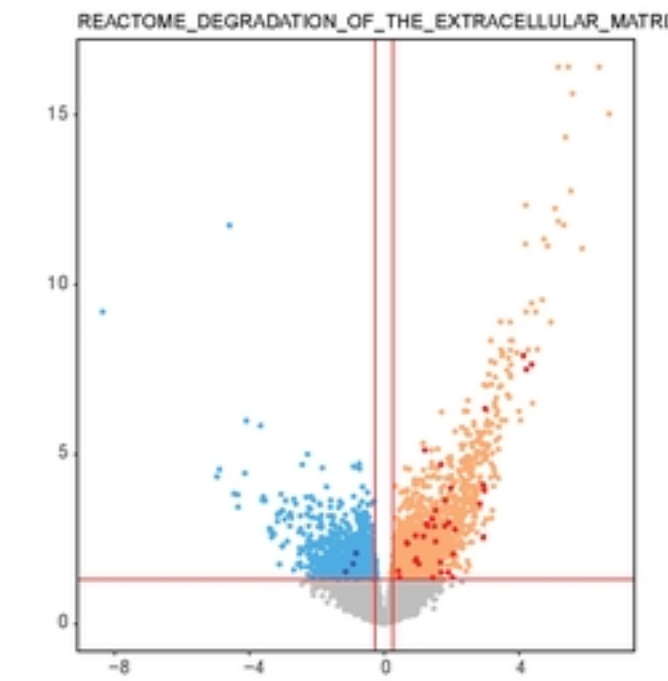

C
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
−log10(padj)
15
10
5
0
−8 −4 0 4


D
HALLMARK_G2M_CHECKPOINT
15
10
5
0
−8 −4 0 4
log2FoldChange


E
HALLMARK_MYOGENESIS
15
10
5
0
−8 −4 0 4


F
REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION
−log10(padj)
15
10
5
0
−8 −4 0 4


G
REACTOME_COLLAGEN_DEGRADATION
15
10
5
0
−8 −4 0 4
log2FoldChange


H
REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX
15
10
5
0
−8 −4 0 4