

1 Comprehensive analysis of mutational signatures in pediatric cancers

2

3 Venu Thatikonda^{1,2,3}, S. M. Ashiqul Islam⁴, Barbara C. Jones^{1,3,5,6}, Susanne N. Gröbner^{1,2,3},
4 Gregor Warsow⁷, Barbara Hutter^{3,8,9}, Daniel Huebschmann^{3,5,8,10}, Stefan Fröhling^{3,8,11}, Mirjam
5 Blattner-Johnson^{1,6}, David T.W. Jones^{1,3,6}, Ludmil B. Alexandrov⁴, Stefan M. Pfister^{1,2,3,5*},
6 Natalie Jäger^{1,2,3*}

7

8 1. Hopp Children's Cancer Center Heidelberg (KITZ), Heidelberg, Germany
9 2. Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Heidelberg, Germany
10 3. German Cancer Consortium (DKTK), 69120 Heidelberg, Germany
11 4. Department of Cellular and Molecular Medicine and Department of Bioengineering, Moores Cancer Center,
12 UC San Diego, La Jolla, 92093, California, USA
13 5. Department of Pediatric Oncology, Hematology and Immunology, Heidelberg University Hospital,
14 Heidelberg, Germany
15 6. Pediatric Glioma Research Group, German Cancer Research Center (DKFZ), Heidelberg, Germany
16 7. Omics IT and Data Management Core Facility (W610), DKFZ, Heidelberg, Germany
17 8. NCT Molecular Diagnostics Program, NCT Heidelberg and DKFZ, 69120 Heidelberg, Germany
18 9. Division of Applied Bioinformatics, DKFZ, Heidelberg, Germany
19 10. Pattern Recognition and Digital Medicine, Heidelberg Institute for Stem cell Technology and Experimental
20 Medicine (HI-STEM), Heidelberg, Germany
21 11. Division of Translational Medical Oncology, NCT Heidelberg and DKFZ, 69120 Heidelberg, Germany

22

23 * corresponding authors

24

25

26 **Abstract**

27 Analysis of mutational signatures can reveal the underlying molecular mechanisms of the
28 processes that have imprinted the somatic mutations found in a cancer genome. Here, we
29 present a pan-cancer mutational signatures analysis of single base substitutions (SBS) and
30 small insertion and deletions (ID) in pediatric cancers encompassing 537 whole genome
31 sequenced tumors from 20 molecularly defined cancer subtypes. We identified only a small
32 number of mutational signatures active in pediatric cancers when compared to the previously
33 analyzed adult cancers. Further, we report a significant difference in the proportion of
34 pediatric tumors which show homologous recombination repair defect signature SBS3
35 compared to prior analyses. Correlating genomic alterations with signature activities, we
36 identified an association of *TP53* mutation status with substitution signatures SBS2, SBS8,
37 SBS13 and indel signatures ID2 and ID9, as well as chromothripsis associated with SBS8,
38 SBS40 and ID9. This analysis provides a systematic overview of COSMIC v.3 SBS and ID
39 mutational signatures active across pediatric cancers, which is highly relevant for
40 understanding tumor biology as well as enabling future research in defining biomarkers of
41 treatment response.

42

43

44 **Main**

45 Childhood cancers have a lower incidence rate when compared to the overall incidence of
46 adult cancers. Nevertheless, cancer remains one of the leading causes of death by disease
47 amongst children¹. Growing research evidence suggests that childhood cancers are
48 significantly different in terms of molecular features and therapy response when compared to
49 their adult counterparts. Most prominently, recent pan-cancer analyses revealed that pediatric
50 cancers show a significantly lower mutation burden in contrast to common adult cancers^{2,3}.
51 However, the knowledge of underlying mutational processes that contribute to the somatic
52 mutation burden and tumor development in pediatric cancer is still limited.

53

54 Somatic mutations including single base substitutions, small insertions and deletions (indels),
55 copy number changes, and other genomic rearrangements can be caused solely by
56 endogenous processes (*e.g.*, defects in DNA repair, errors in DNA replication, damage due to
57 reactive oxygen species, *etc.*), or by the additional influence of exogenous causes such as
58 exposure to ultra-violet light, tobacco smoking, and numerous others⁴. Previous pan-cancer

59 analyses mainly focused on adult cancer types, employed mathematical models based on
60 nonnegative matrix factorization (NMF) to identify patterns of somatic mutations, termed
61 mutational signatures and utilized these patterns to infer underlying mutational processes⁵.
62 Since then, mutational signatures have been used to understand tumor development, to
63 identify gene alterations associated with mutational processes and importantly, as biomarkers
64 for predicting treatment response^{6,7,8,9,10,11,12}. A total of 45 single base substitution signatures
65 (SBS) and 17 small insertion/deletion (ID) signatures were identified in a recent analysis as
66 part of the pan-cancer analysis of whole genomes (PCAWG) consortium^{13,14}, updating and
67 expanding the previous set of reference mutational signatures (*i.e.*, mutational signatures in
68 COSMIC v.2).

69

70 Previous studies of individual pediatric tumor types and pan-cancer approaches analyzed
71 mutational signatures as part of molecular tumor landscape analyses using the COSMIC v.2
72 reference signatures^{15,16,17,18,19,20,23}. Here, we carried out an extensive analysis encompassing
73 signatures of single base substitution and, for the first time, signatures of small
74 insertion/deletions in 537 whole genome sequenced pediatric tumor-normal pairs. These were
75 subsequently compared with COSMIC v.3¹⁴ signatures to identify overlap with the latest set
76 of known mutational signatures.

77

78

79 Results

80

81 Somatic mutation frequencies across twenty pediatric cancer types

82 A dataset of 537 whole genome sequenced tumor-normal pairs (the PedPanCan (PPC-WGS)
83 cohort, <https://www.kitz-heidelberg.de/en/research/datacommons/pedpancan/>) was compiled
84 from a previously published study², spanning 20 molecularly defined entities of childhood
85 cancer (Fig. 1a). Single base substitutions (SBS) and small insertions and deletions (IDs) were
86 identified using an updated in-house mutation calling pipeline as compared with the previous
87 analysis, with minor variations mostly in the calling strategy of indels (Supp. Fig. 1a). In total,
88 across the cohort we identified 2,712,521 SBS and 270,928 IDs. The number of SBS
89 mutations per megabase (median: 0.24; range: 0.0035 - 645.70) and IDs (median: 0.033;
90 range: 0.00142- 46.64) was highly variable both across individual tumors as well as across
91 different cancer types (Fig. 1a, Supp. Table 1). The lowest overall somatic mutation burden
92 was observed in pilocytic astrocytoma (median: 0.0429 mutations per megabase) and the
93 highest in osteosarcoma (median: 1.17). Although the indel mutation burden per tumor was
94 low, numbers of SBS and IDs were significantly correlated across tumors (Fig. 1b). As
95 previously described in different childhood and adulthood cancer types, the mutation burden
96 of both SBS and IDs is also clearly correlated with age in this pediatric cohort (Supp. Fig. 1b).
97 A small number (n=3) of tumors classified as high-grade glioma with germline alterations in
98 DNA mismatch repair genes (*MSH6*, *PMS2*) showed a hyper-mutator phenotype², with 78.98
99 SBS and 2.63 ID mutations/Mb, respectively (Fig. 1b).

100

101 SBS and ID signature activities

102 To extract mutational signatures active in this pediatric pan-cancer cohort, we generated 96-
103 context SBS and 83-context ID mutational catalogues (Fig. 2a) and utilized an approach based
104 on nonnegative matrix factorization, as previously described^{5,21,14}. Due to the very low
105 number of double base substitutions per pediatric tumor (data not shown), we only extracted
106 signatures from SBS and IDs.

107

108 In total, 27 SBS signatures were extracted that matched the COSMIC v.3 SBS signatures with
109 a minimum cosine similarity of 0.9 (Fig. 2b, Supp. Fig. 2a). Amongst these, SBS1 and SBS5
110 were present in 97.5% and 96.1% of samples across the cohort, respectively (Fig. 2b, Supp.
111 Table 2, Supp. Table 3). As described in adult cancers and a small fraction of pediatric brain
112 tumors, the clock-like nature of SBS1 and SBS5 was also observed in this cohort, with a

113 significant correlation of signature activity with age at diagnosis⁷ (Supp. Fig. 3a). An
114 additional signature with unknown etiology, namely SBS40, is similar to SBS5 (cosine
115 similarity =0.83)¹⁴. SBS40 was found to be active in five pediatric cancer types and 20% of
116 tumors across the cohort (Fig. 2b, Supp. Table 3) and was also correlated with age at
117 diagnosis (Supp. Fig. 3a), suggesting it may be an additional clock-like signature.

118

119 SBS2 and SBS13 were reported to be mainly due to the activity of APOBEC enzymes¹⁴ and
120 in this cohort they were found to be present only in adrenocortical carcinoma (ACC), high-
121 grade glioma and osteosarcoma (Fig. 2b, Supp. Table 2). In a cross-cohort comparison,
122 mutations attributed to SBS2 and SBS13 were significantly higher in *TP53* germline mutated
123 tumors compared to tumors with somatic or no *TP53* mutations (Supp. Fig. 3b). This
124 observation is in line with previous studies which identified a link between p53 loss and
125 elevated *APOBEC3B* expression^{22,23}. However, not all *TP53* germline mutated tumors in this
126 cohort showed SBS2 or SBS13 activity, including the *TP53*-defined subtype of SHH-
127 medulloblastoma (Supp. Table 2 & Table 13).

128

129 Ultraviolet light (UV) exposure signatures SBS7a and SBS7b¹⁴ were identified in tumors of
130 hypodiploid B-cell acute lymphoblastic leukemia (B-ALL-HYPO) (Fig. 2b), but not in non-
131 aneuploid tumors. UV light affected tumors show an enrichment of dipyrimidine substitution
132 mutations (CC>TT). Consistently, SBS7a and SBS7b positive tumors show a significantly
133 higher number of CC>TT doublet base substitutions when compared to SBS7a and SBS7b
134 negative tumors in the B-ALL-HYPO subtype (Supp. Fig. 3c). Although multiple independent
135 studies have recently identified signatures apparently linked with UV light exposure in
136 pediatric B-ALL^{3,45}, the exact mechanism how it contributes to leukemogenesis, and whether
137 UV light is really the cause or rather the signature is mimicked by another process, remains to
138 be elucidated. In 1.5% of tumors (8/537; ETMR and Group4 medulloblastoma) we identified
139 signature SBS4 (Fig. 2b, Supp. Table 2), which is proposed to be the result of DNA damage
140 caused by tobacco smoking¹⁴. To the best of our knowledge, these tumors were derived from
141 untreated, primary pediatric tumors and we hypothesize that the C>A mutation pattern
142 consistent with SBS4 may potentially be due to a different exogenous mutational process
143 generating a similar signature of mutations as tobacco smoking. No common germline or
144 somatic mutations were identified in these eight brain tumors with SBS4 activity.

145

146 Other relatively frequent signatures across cancer types included SBS8 and SBS18 (Fig. 2b).
147 Signature 8 from COSMIC v.2 was proposed to be due to homologous recombination repair
148 pathway gene mutations². We did not identify any such mutations in SBS8 positive pediatric
149 tumors in this analysis. However, the number of mutations attributed to SBS8 was
150 significantly higher in tumors with germline or somatic *TP53* mutations compared to wildtype
151 tumors (Supp. Fig. 3d). Recent evidence suggests that the SBS8 signature is due to the DNA
152 damage caused by late replication errors²⁴. SBS18 is the result of DNA damage caused by
153 reactive oxygen species¹⁴ and is observed in 12 out of 20 cancer types analyzed in this cohort
154 (Fig. 2b). Amongst these 12 cancer types, the highest fraction of tumors with SBS18 signature
155 activity were observed in neuroblastoma (85%, NB) and rhabdomyosarcoma (83%, RMS)
156 (Supp. Table 4). Further, we found high activity of SBS18 in MYCN-amplified tumors in a
157 cross-cohort analysis (Supp. Fig. 3h). However, in a per-cancer-type analysis of NB, RMS
158 and Group3 medulloblastoma, tumors in which MYCN amplification is common, we did not
159 identify a significant SBS18 association with MYCN status.

160

161 Mutations attributed to signature SBS9 are due to DNA damage caused by polymerase eta
162 activity and were observed only in Burkitt's lymphoma (BL) in this cohort. As previously
163 described, SBS9 activity is found in BL with immunoglobulin gene hypermutations³⁶ (Supp.
164 Fig. 3e). Signature 10 from COSMIC v.2 has been split into SBS10a and SBS10b in the
165 COSMIC v.3 signatures¹⁴. Mutations from high-grade glioma hyper-mutator tumors were
166 attributed to both SBS10 signatures, as well as SBS14 and SBS15 (Fig. 2b). In Group4
167 medulloblastoma, we identified a novel SBS signature (termed Group4MB-SBS96D) with
168 elevated T>A mutations in the context of CTT/TTT (Supp. Fig. 3f, Supp. Table 5). This
169 signature contributed a large fraction of somatic point mutations to two tumors (3,623 in
170 ICGC_MB174 and 1,458 in ICGC_MB175, respectively). However, we did not identify any
171 molecular features common and specific to these tumors. Further investigation on the
172 processing of tumor material revealed that the surgery for these two patients was performed
173 on the same day in the same hospital, and therefore we suspect that the Group4MB-SBS96D
174 signature is an additional artefact signature.

175

176 In addition, we identified SBS21 in high-grade glioma, which is the result of defective DNA
177 mismatch repair, SBS36 in ETMR and Group4 medulloblastoma, which is the result of
178 defective base excision repair, and SBS44 in Group4 medulloblastoma, which is also caused
179 by defective DNA mismatch repair. For these signatures, we have not identified any

180 associated consistent genetic alteration enriched across tumors. However, individual tumors
181 harbor non-synonymous somatic mutations in different genes involved in DNA mismatch
182 repair and nucleotide excision repair pathways (e.g. *TP53*, *POLD1*, *SMG1*, *TP73*, *SWI5*;
183 Supp. Table 6). The possibility that tumors with SBS21, SBS36 and SBS44 activity might
184 have epigenetic alterations affecting known genes in repair pathways and/or mutations in
185 genes that play a role in repair pathways, but for which their function is not yet well
186 characterized cannot be excluded.

187

188 Next, we sought to identify any associations between SBS signature activity, chromothripsis
189 and genomic instability. In a cross-cohort analysis, mutations attributed to SBS8 and SBS40
190 were significantly higher in chromothriptic tumors (Supp. Fig. 3g and Supp. Table 12 for an
191 overview of samples with chromothripsis). Individual cancer type analysis revealed that
192 SBS40 activity in SHH-subtype medulloblastoma was significantly different depending on the
193 presence of chromothripsis in the tumor genome (Supp. Fig. 3g). Genomic instability,
194 quantified here as the total number of structural variants (deletions, duplications, inversions
195 and translocations) identified in a tumor, is highly correlated with SBS2, SBS13 and SBS40
196 in a cross-cohort analysis (Supp. Fig. 3i). A separate per-cancer type analysis revealed a high
197 correlation of genomic instability and SBS40 activity in SHH and Group4 medulloblastoma,
198 as well as in neuroblastoma (Supp. Fig. 3i). These observations may indicate that mechanisms
199 involved in genomic rearrangements contribute to mutational processes underlying APOBEC
200 signatures and SBS40.

201

202 A total of nine small insertion/deletion (ID) signatures were identified from our PedPanCan
203 cohort that matched with the COSMIC v.3 ID signatures (Fig. 2b, Supp. Table 7). As also
204 observed in the PCAWG mutational signature analysis, ID1, ID2, ID5 and ID9 were active
205 across tumors of multiple cancer types (Fig. 2b, Supp. Fig. 2b)¹⁴. ID1 and ID2, which are the
206 result of DNA damage induced by replication slippage, were present in ~95% and ~61% of
207 the tumors in this cohort, respectively (the most prevalent ID signatures in this cohort).
208 Although ID5 and ID9 were present in multiple cancer types, mutations were attributed to
209 these signatures in only a small percentage of tumors (Supp. Table 8, Supp. Table 9). ID8, a
210 signature caused by the potential damage induced by DNA double-strand break repair by non-
211 homologous end joining was present in 6% (n=26) of the whole cohort. The large cohort
212 analysis of tumors as part of PCAWG revealed that ID8 activity was correlated with age at
213 diagnosis, suggesting a clock-like behavior of this signature¹⁴. However, we did not observe

214 such correlation in our cohort, potentially due to the very low number of tumors with ID8
215 activity. ID12, a signature with unknown etiology was identified in a small fraction of high-
216 grade glioma and Group3 medulloblastoma. Manual review of aligned reads and variant calls
217 of ID12 assigned indels showed that these mutations were mostly present in low-mappability,
218 repeat-rich regions of the genome, leading to the hypothesis that ID12 might present an
219 artefact signature in this cohort at least.

220

221 Next, we sought to identify potential associations of indel signature activity with age, *TP53*
222 mutation status, chromothripsis and genomic instability. Of the 9 ID signatures active in our
223 cohort, we observed that ID1 and ID2 signature activity is highly correlated with age at
224 diagnosis (Supp. Fig. 4a), revealing their clock-like behavior in pediatric cancers as observed
225 in adult cancers¹⁴. The number of mutations attributed to ID2 and ID9 (of unknown etiology)
226 were significantly different depending on the *TP53* mutation status. *TP53* germline/somatic
227 mutated tumors showed higher activity of ID2 and ID9 compared with wildtype tumors
228 (Supp. Fig 4b). Similarly, a significantly higher number of mutations were attributed to ID2
229 and ID9 in chromothriptic tumors compared to non-chromothriptic tumors (Supp. Fig. 4c). A
230 similar difference was observed in adult cancers²⁵. The total number of structural variants was
231 highly correlated with ID9 signature activity across the cohort. A per-cancer type analysis
232 revealed that ID9 and genomic instability were highly correlated in RMS (Supp. Fig. 4d). In
233 addition, ID3 and ID4 activities were observed in a small fraction of tumors (Supp. Table 8),
234 however, we did not identify any common genomic alterations across these tumors.

235

236 Finally, we performed a correlation analysis of SBS and ID signature activities across the
237 cohort to understand if any of the SBS and ID signatures co-occurred. We observed a high
238 correlation amongst clock-like signatures SBS1, SBS5 and ID1, ID2 (Fig. 2c). In addition,
239 ID3, a tobacco smoking related signature was highly correlated with SBS31 and SBS35, both
240 of which are due to DNA damage caused by platinum treatment (Fig. 2b, c).

241

242 In summary, we extracted 27 SBS signatures and 9 ID signatures in our pediatric pan-cancer
243 cohort that overlap COSMIC v.3 signatures. The total number of identified signatures is lower
244 than those identified in adult cancers¹⁴ and a large fraction of mutations (57% and 42% of
245 SBS mutations; 85% and 100% of ID mutations in non-hyper mutated and hyper-mutated
246 samples, respectively) were attributed to clock-like signatures such as SBS1, SBS5, ID1 and

247 ID2. A significant difference in terms of mutations attributed to SBS2, SBS8 and SBS13 as
248 well as ID2 and ID9 was observed to be depending on *TP53* mutation status.

249

250 **Homologous recombination repair defect signature activities**

251 The homologous recombination (HR) repair pathway is an error-free mechanism to repair
252 DNA double-strand breaks²⁶. Genomic alterations in components of the HR pathway, mainly
253 in the *BRCA1/2* genes, lead to a characteristic pattern of single base substitution mutations
254 and large deletions at microhomology regions, a phenotype frequently termed ‘BRCAness’²⁷.
255 Previous studies focusing mainly on breast and ovarian cancers have identified a strong
256 correlation between *BRCA1/2* biallelic pathogenic mutations and activity of signature 3
257 (COSMIC v.2). However, a significant fraction of these tumors showed signature 3 activity
258 without any identifiable alterations in HR pathway components^{5,11,28,14}. Discerning the
259 activity of HR defect COSMIC v.3 signatures SBS3 and ID6 in tumors is important as it has
260 previously been shown to be associated with the therapeutic response to platinum and poly
261 ADP-ribose polymerase (PARP) inhibitor treatment^{29,12,30,31,27,28,46}.

262

263 Previous analyses of mutational signatures in pediatric cancers, including our own published
264 analysis², have identified a significant proportion of Signature.3 (COSMIC v.2), mostly in
265 tumors without any HR pathway gene defects². However, the current analysis with COSMIC
266 v.3 signatures identified only a small fraction (2.23% of the whole cohort) of tumors with
267 SBS3 signature activity (Fig. 3a,b). This marked difference is most likely the result of
268 previous Signature 3 mutations now being attributed to “flat” signatures (e.g., SBS5 and
269 SBS40) of the updated and refined COSMIC v.3 mutational signatures. In addition, there is a
270 difference in the approach compared to the initial signature analysis, as we assume SBS1 and
271 SBS5 as background signatures and only add SBS3 if it improves the cosine similarity with at
272 least 0.02. However, none of these tumors with SBS3 showed the associated ID6 signature
273 activity, represented by a high fraction of long deletions at microhomology regions (Fig. 2a,
274 Fig. 3b). This lack of ID6 prompted us to test whether our current variant calling pipeline
275 could be penalizing deletions at microhomologies by assigning low confidence, for example,
276 or if pediatric cancers in general have very low numbers of microhomology-associated
277 deletions as an inherent property. In order to test this further, we whole-genome sequenced
278 five tumors from the INFORM registry³² with known (likely) pathogenic germline mutations
279 in *BRCA1/2* based on ClinVar. Amongst these, INF_R_1076 had a *BRCA2* pathogenic
280 homozygous mutation and INF_R_025 a *BRCA2* compound heterozygous mutation (both

281 patients also had a clear phenotype, i.e. Fanconi anaemia), while the remaining tumors had
282 heterozygous (likely) pathogenic *BRCA1/2* mutations (Fig. 3c) without any second hit in the
283 tumor. Mutational signature analysis of these five pediatric tumors revealed that only the two
284 tumors with compound heterozygous and homozygous *BRCA2* mutations, i.e. biallelic
285 inactivation, showed SBS3 and ID6 signature activity (Fig. 3c, Supp. Table 10). In addition,
286 we analyzed 22 whole genome sequenced adult tumors from the NCT-MASTER precision
287 oncology program (<https://www.nct-heidelberg.de/master>) with respect to ID signatures, for
288 which somatic mutations were called with the same DKFZ in-house pipeline. Amongst these,
289 half of the tumors (n=11) had known *BRCA1/2* deficiency and showed clear ID6 activity
290 (Supp. Fig. 5a, Supp. Table 11). These results indicate that our variant calling pipeline does
291 not systematically miss the deletions at microhomology regions and confirm the importance
292 of biallelic *BRCA1/2* inactivation for the presence of predictive HR defect-associated
293 mutational signatures, that is SBS3 and ID6.

294 In the Gröbner *et al.* publication², the initial analysis of this pediatric pan-cancer cohort, six
295 patients were identified with a pathogenic/likely pathogenic *BRCA2* germline variant (Supp.
296 Table 13). Five of these patients were heterozygous in the germline and without evidence for
297 biallelic inactivation in the tumor, while one case was compound heterozygous in the
298 germline for *BRCA2* (SJMB012, 8 year old male with SHH-medulloblastoma). For the latter
299 case, one *BRCA2* germline variant (ENSP00000439902.1:p.Val2407GlufsTer60) is not
300 currently reported in ClinVar and the other one as likely pathogenic (variationID: 421014).
301 We did not identify SBS3 or ID6 activity in any of these six tumors.

302
303 Next, we sought to understand whether a proportion of long deletions (>5bp) at
304 microhomologies (MH) are present in pediatric tumors, albeit at very low levels. In order to
305 understand how MH-associated deletions are represented in ID6 positive adult tumors, we
306 divided the PCAWG whole genome tumors (n=2,776) into different categories depending on
307 the presence of SBS3 and ID6. Then we compared the proportion of MH-associated deletions
308 of tumors in these categories with our pediatric cancer cohort. This analysis revealed that
309 pediatric tumors overall have very few MH-associated deletions compared with adult tumors
310 and a significantly lower fraction compared with ID6 positive tumors (Fig. 3d). However, in
311 our cohort, one osteosarcoma tumor (SJOS004) showed ID6 signature without any activity of
312 the SBS3 signature (Fig. 3e). Amongst its SBS signatures, a high fraction of SBS40 mutations
313 were observed. SBS3 and SBS40 are relatively flat signatures and have a similar feature
314 distribution (cosine similarity of 0.88). For this reason, the possibility of mis-assigning SBS3

315 mutations to SBS40 cannot be excluded. However, no HR pathway mutations were identified
316 in this osteosarcoma tumor.

317

318 Amongst the tumors with SBS3 signature activity in our cohort, 58% (7/12) belong to the
319 Group3 subgroup of medulloblastoma (MB). MB tumors are known to follow a linear
320 increase of somatic mutation burden with age³³. Presumably, mutations in tumors that do not
321 follow such a correlation could be contributed by mutational processes other than the
322 ubiquitous clock-like signatures. In this analysis, we identified only one tumor (SJMB008)
323 with an active SBS3 signature as an outlier of age versus mutation burden correlation,
324 suggesting that most of the other SBS3-positive Group3 MBs follow the typical linear
325 increase in mutation burden with age (Supp. Fig. 5b), and therefore the possibility of mis-
326 assigning mutations from clock-like signatures (especially SBS5 and SBS40) to SBS3 cannot
327 be excluded. Next, we compared expression of HR pathway genes such as *BRCA1/2* and
328 *PALB2* between SBS3 positive and negative tumors in Group 3 MBs, as there could be other
329 mechanisms acting to inactivate the expression of these genes. We observed only a slight
330 difference in *PALB2* expression (Supp. Fig. 5c), but not in the other genes and also no
331 significant difference in promoter DNA methylation of these genes (data not shown).

332

333 In summary, we identified only a small percentage (2.23%) of pediatric tumors in our cohort
334 with SBS3 (COSMIC v.3) activity, the so-called HR deficient phenotype, compared to ~54%
335 of tumors with Signature 3 (COSMIC v.2) activity in our previous analysis of the same
336 cohort². Furthermore, in tumors with SBS3 signature activity, we could not identify any
337 genomic alterations or loss of expression of HR pathway genes. The genome of these SBS3
338 positive tumors was relatively stable with very few outliers (Supp. Fig. 6). The indel signature
339 ID6, which is characterized by long deletions (>5bp) at microhomology sequences, is a strong
340 predictor of HR deficiency and was not identified in our cohort except for one osteosarcoma
341 tumor. Further analysis of n=5 INFORM and n=22 NCT-MASTER whole genome sequenced
342 tumors suggests the importance of genetic biallelic *BRCA1/2* inactivating events in the
343 generation of HR deficient mutational signatures.

344

345 **Signature.P1 similarity to COSMIC v.3 signatures**

346 Previous mutational signature analysis of this cohort based on the COSMIC v.2 reference
347 signatures identified a novel substitution signature, called Signature P1, which featured
348 elevated mutations of C>T in the context of CCC/CCT². Signature P1 was active in the

349 pediatric brain tumors atypical teratoid/rhabdoid tumors (ATRT) and ependymoma. In the
350 current analysis, the Signature P1 profile was compared to all identified COSMIC v.3 SBS
351 signatures and a high similarity was observed with signature SBS31 (Fig. 4a,b). SBS31 is the
352 result of DNA damage caused by platinum treatment. In the present analysis, SBS31 activity
353 was identified in ATRT, ependymoma and ETMR tumors (Fig. 4c). These results suggest that
354 Signature.P1 is not, as previously hypothesized, a pediatric specific mutational signature, but
355 rather a treatment-associated signature identified in a small fraction of tumors that were
356 annotated as treatment naïve. These patients had likely been treated prior to genomic analysis,
357 and at least for one ATRT sample (H049-JVCT; high SBS35 activity, Suppl. Table 2) we
358 could follow up with the sample source which confirmed that this tumor was a recurrence and
359 not a primary tumor.
360

361 Discussion

362 In this study, we re-analyzed 537 whole genome sequenced tumor-normal pairs from 20
363 different molecularly defined entities of childhood cancers to identify and refine the
364 mutational signatures of the underlying mutational processes. We examined single base
365 substitution (SBS) and small insertion/deletion (ID) in depth and showed that a relatively
366 small number of mutational processes operate in pediatric cancers compared with adult
367 cancers. Amongst the identified 27 SBS and 9 ID signatures, etiologies for more than half of
368 these signatures have been described either by experimental approaches or association
369 analyses in previous studies in adult cancers. In this cohort, a large fraction of SBS (45.4%)
370 and ID (93.2%) mutations across multiple cancer types were attributed to clock-like
371 substitution signatures SBS1 and SBS5, as well as to clock-like indel signatures. Signatures
372 such as SBS8 are relatively frequent across cancer entities, but the etiology remains unknown
373 and recent evidence suggests this signature is due to DNA damage induced by late replication
374 errors²⁴. Other signatures without known etiology were identified in a small fraction of tumors
375 from specific cancer types such as pilocytic astrocytoma (namely, SBS12 and SBS23) and
376 Burkitt's lymphoma (namely, SBS17a, SBS17b). However, we did not identify any genomic
377 alteration common to these tumors.

378
379 Previous mutational signature analyses identified Signature 3 (equivalent to the current SBS3)
380 along with other features, with or without discernable alterations in *BRCA1/2* genes, as a
381 potential biomarker for HR deficiency and platinum plus PARPi-based treatment
382 response^{34,35}. In pediatric cancers, germline predisposition or somatic alterations in *BRCA1/2*
383 genes are infrequent^{2,3} and only a small fraction (2.23%) of tumors showed SBS3 activity in
384 this analysis, but without the complementing indel signature ID6. Microhomology (MH)-
385 associated deletions, a strong feature of ID6, were significantly more rare in pediatric than in
386 adult cancers. A plausible explanation for this observation could be that the total somatic
387 mutation burden of childhood cancers is very low² and the likelihood of mutations occurring
388 at MH sequences will therefore also be small. Polak *et al.*¹¹ identified a signature that is
389 similar to the current SBS3 in breast cancer tumors without alterations in *BRCA1/2* genes, but
390 identified alterations in other HR pathway components (e.g. *PALB2* or *RAD51C*). In this
391 pediatric cancer cohort, we have not identified genomic alterations in other HR pathway
392 genes that could potentially result in the observed SBS3 signature. While there might be other
393 epigenetic mechanisms to inactivate HR pathway genes, in Group3 medulloblastoma (for
394 which complementary omics data is available^{33,47}) we have not observed promoter

395 hypermethylation or reduced expression of HR pathway genes. Furthermore, we do not
396 discount the possibility of mis-assigning SBS5 and SBS40 signature mutations to SBS3. In
397 addition, a recent study suggests that SBS3 is likely not as specific as previously believed and
398 that the identification of HR deficiency should rely on multiple orthogonal mutational
399 signatures, not only on SBS3⁴⁸. In order to validate the biomarker efficacy of SBS3 to predict
400 PARP inhibitor treatment response in pediatric cancers, experiments involving PDX models
401 of pediatric tumors are currently ongoing (for example in the “BRCAddict” project,
402 <https://www.transcanfp7.eu/index.php/abstract/brcaddict.html>). Recent studies focusing on
403 Ewing sarcoma³⁷ and ETMR brain tumors³⁸ identified the presence of R-loops, DNA-RNA
404 hybrid structures, to be correlated with PARP inhibitor response. In the future, mutational
405 signature SBS3 combined with other features such as indel signature ID6 and/or R-loops
406 should be investigated in preclinical models to assess PARP inhibitor response and to
407 ultimately define optimized biomarkers.

408

409 In summary, although some cancer entities are under-represented in this cohort, we believe
410 this analysis on the mutational signature repertoire in childhood cancers provides a valuable
411 resource for further understanding of tumor biology and aids future research in defining
412 biomarkers of treatment response.

413

414 **METHODS**

415

416 **Whole-genome sequencing data, alignment, and variant calling**

417 All whole genome sequencing data analyzed in this study was collected from Gröbner, Worst
418 et al., 2018². Briefly, FASTQ data was aligned to reference genome GRCh37/hg19 with
419 BWA-MEM (v 0.7.8)³⁹. Single base substitutions were called using an updated samtools⁴⁰
420 based DKFZ in-house pipeline (0.1.19) and indels were called using Platypus (0.8.1.1).

421

422 **Somatic Mutation Frequencies**

423 Single base substitution and small insertion/deletion mutation burden was calculated as the
424 total number of mutations identified per megabase of the genome. For whole genome
425 sequencing data, the total number of mutations was divided by 2800 (effective human genome
426 size in megabases that can be assessed by whole-genome sequencing).

427

428 **Analysis of mutational signatures using SigProfiler**

429 In order to extract mutational signatures, 96-context SBS and 83-context ID mutational
430 catalogues were prepared using SigProfilerMatrixGenerator (version 1.0.24)⁴¹ for each of the
431 20 cancer types. These mutational catalogues were used as input to SigProfilerExtractor
432 (version 1.0.19)⁴² to extract signatures and attribute mutations to each signature in every
433 individual tumor. Briefly, SigProfilerExtractor applies a nonnegative matrix factorization
434 algorithm in multiple iterations (n=100) and for each iteration the software minimizes a
435 generalized Kullback-Leibler divergence constrained for non-negativity. The optimal number
436 of mutational signatures was selected based on highest average stability and lowest average
437 sample cosine distance. After the optimal number of signatures was estimated, attribution of
438 mutations to each signature in each sample involved finding the minimum of the Frobenius
439 norm of a constrained function using a nonlinear convex optimization programming solver
440 based on the interior point algorithm²¹.

441

442 **Structural variants and copy-number profiles**

443 Structural variants per tumor were identified as described previously using the DELLY⁴³
444 ICGC pan-cancer analysis workflow ([https://github.com/ICGC-TCGA-
445 PanCancer/pcawg_delly_workflow](https://github.com/ICGC-TCGA-PanCancer/pcawg_delly_workflow)). Copy-numbers were estimated using the ACEseq (allele-
446 specific copy-number estimation from sequencing) too, based on binned tumor-control
447 coverage ratio and B-allele frequencies (BAF)⁴⁴.

448

449 **Statistical analyses and code availability**

450 All downstream statistical analyses were performed using the R statistical programming
451 language (version 4.0.3) and all code is available at

452 <https://github.com/KiTZ-Heidelberg/Signatures-Manuscript>

453

454

455

456

457 **Acknowledgements**

458 This project was supported by funding from the ADDRESS consortium (Grant 01GM1909E,
459 BMBF). L.B.A. is an Abeloff V scholar and he is personally supported by a Packard
460 Fellowship for Science and Engineering. We acknowledge the DKFZ's ODCF and GPCF core
461 facilities for supporting the genomic sequencing and data processing.

462

463 **Contributions**

464 V.T., S.M.P. and N.J. conceptualized the study. S.M.A.I. and L.B.A. contributed to software
465 development and implementation. Analyses were performed by V.T., N.J., G.W., S.M.A.I and
466 L.B.A. Data visualization was conducted by V.T. Data were collected and curation of data
467 was conducted by V.T., B.C.J., S.N.G., B.H., D.H., S.F., M.B.-J., and D.T.W.J. The original
468 draft was written by V.T. and N.J., and all co-authors reviewed and edited the manuscript.

469

470 **Corresponding authors**

471 Correspondence to Stefan Pfister and Natalie Jäger.

472

473

474

475 **REFERENCES**

476

- 477 1. Pui, C., Gajjar, A. *et al.* Challenging issues in pediatric oncology. *Nat. Rev. Clin. Oncol* **540**, 540–549
478 (2011).
- 479 2. Gröbner, S., Worst, B., Weischenfeldt, J. *et al.* The landscape of genomic alterations across childhood
480 cancers. *Nature* **555**, 321–327 (2018)
- 481 3. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid
482 tumours. *Nature* **555**, 371–376 (2018).
- 483 4. Stratton, M. R., Campbell, P. J. & Futreal, & P. A. The cancer genome. *Nature* **458**, (2009).
- 484 5. Alexandrov, L., Nik-Zainal, S., Wedge, D. *et al.* Signatures of mutational processes in human cancer.
485 *Nature* **500**, 415–421 (2013)
- 486 6. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993
487 (2012).
- 488 7. Alexandrov, L., Jones, P., Wedge, D. *et al.* Clock-like mutational processes in human somatic cells. *Nat*
489 *Genet* **47**, 1402–1407 (2015).
- 490 8. Schulze, K., Imbeaud, S., Letouzé, E. *et al.* Exome sequencing of hepatocellular carcinomas identifies
491 new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505–511 (2015).
- 492 9. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences.
493 *Nature* **534**, 47–54 (2016).
- 494 10. Petljak, M. & Alexandrov, L. B. Understanding mutagenesis through delineation of mutational
495 signatures in human cancer. *Carcinogenesis* **37**, 531–540 (2016).
- 496 11. Polak, P. *et al.* A mutational signature reveals alterations underlying deficient homologous
497 recombination repair in breast cancer. *Nat. Genet.* **49**, 21 (2017).
- 498 12. Keung, M., Wu, Y. & Vadgama, J. PARP Inhibitors as a Therapeutic Agent for Homologous
499 Recombination Deficiency in Breast Cancers. *J. Clin. Med.* **8**, 435 (2019).
- 500 13. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
- 501 14. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101
502 (2020).
- 503 15. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups
504 with Distinct Enhancer Landscapes. *Cancer Cell* **29**, 379–393 (2016).
- 505 16. Northcott, P., Buchhalter, I., Morrissy, A. *et al.* The whole-genome landscape of medulloblastoma
506 subtypes. *Nature* **547**, 311–317 (2017)
- 507 17. Kovac, M. *et al.* Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of BRCA
508 deficiency. *Nat. Commun.* **6**, 1–9 (2015).
- 509 18. Kool, M. *et al.* Genome sequencing of SHH medulloblastoma predicts genotype-related response to
510 smoothed inhibition. *Cancer Cell* **25**, 393–405 (2014).
- 511 19. Kunz, J. B. *et al.* Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection,
512 acquisition of mutations and promoter hypomethylation. *Haematologica* **100**, 1442–1450 (2015).
- 513 20. Jones, D. T. W. *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat.*
514 *Genet.* **45**, 927–932 (2013).

- 515 21. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering
516 Signatures of Mutational Processes Operative in Human Cancer. *Cell Rep.* **3**, 246–259 (2013).
- 517 22. Periyasamy, M. *et al.* p53 controls expression of the DNA deaminase APOBEC3B to limit its potential
518 mutagenic activity in cancer cells. *Nucleic Acids Res.* **45**, 11056–11069 (2017).
- 519 23. Wang, S., Jia, M., He, Z. & Liu, X.-S. APOBEC3B and APOBEC mutational signature as potential
520 predictive markers for immunotherapy response in non-small cell lung cancer. *Oncogene* **37**, 3924–3936
521 (2018).
- 522 24. Singh, V.K., Rastogi, A., Hu, X. *et al.* Mutational signature SBS8 predominantly arises due to late
523 replication errors in cancer. *Commun Biol* **3**, 421 (2020).
- 524 25. Voronina, N., Wong, J.K.L., Hübschmann, D. *et al.* The landscape of chromothripsis across adult cancer
525 types. *Nat Commun* **11**, 2320 (2020).
- 526 26. Jasin, M. & Rothstein, R. Repair of Strand Breaks by Homologous Recombination. (2013)
527 doi:10.1101/cshperspect.a012740
- 528 27. Byrum, A. K., Vindigni, A. & Mosammaparast, N. Defining and Modulating ‘BRCAness’. *Trends in*
529 *Cell Biology* **29**, 740–751 (2019).
- 530 28. Davies, H. *et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational
531 signatures. *Nat. Med.* **23**, 517–525 (2017).
- 532 29. Póti, Á. *et al.* Correlation of homologous recombination deficiency induced mutational signatures with
533 sensitivity to PARP inhibitors and cytotoxic agents. *Genome Biol.* **20**, 240 (2019).
- 534 30. Chen, A. PARP inhibitors: its role in treatment of cancer. *Chinese journal of cancer* **30**, 463–471 (2011).
- 535 31. Tutt, A. *et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2
536 mutations and advanced breast cancer: A proof-of-concept trial. *Lancet* **376**, 235–244 (2010).
- 537 32. Worst, B. C. *et al.* Next-generation personalised medicine for high-risk paediatric cancer patients – The
538 INFORM pilot study. *Eur. J. Cancer* **65**, 91–101 (2016).
- 539 33. Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**,
540 100–105 (2012).
- 541 34. Ma, J., Setton, J., Lee, N. Y., Riaz, N. & Powell, S. N. The therapeutic significance of mutational
542 signatures from DNA repair deficiency in cancer. *Nature Communications* **9**, (2018).
- 543 35. Zhao, E. Y. *et al.* Personalized Medicine and Imaging Homologous Recombination Deficiency and
544 Platinum-Based Therapy Outcomes in Advanced Breast Cancer. *Clin Cancer Res* **23**, (2017).
- 545 36. Machado, H. *et al.*, Genome-wide mutational signatures of immunological diversification in normal
546 lymphocytes. (2021) bioRxiv 2021.04.29.441939; doi: <https://doi.org/10.1101/2021.04.29.441939>
- 547 37. Gorthi, A. *et al.* EWS-FLI1 increases transcription to cause R-Loops and block BRCA1 repair in Ewing
548 sarcoma. *Nature* **555**, 387–391 (2018).
- 549 38. Lambo, S. *et al.* The molecular landscape of ETMR at diagnosis and relapse. *Naure* **576**, 576(7786):274-
550 280 (2019).
- 551 39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
552 *Bioinformatics* **25**, 1754–1760 (2009).
- 553 40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Appl. NOTE* **25**, 2078–
554 2079 (2009).
- 555 41. Bergstrom, E. N. *et al.* SigProfilerMatrixGenerator: A tool for visualizing and exploring patterns of

- 556 small mutational events. *BMC Genomics* **20**, 685 (2019).
- 557 42. Ashiqul Islam, S. M. *et al.* Uncovering novel mutational signatures by de novo extraction with. (2021)
- 558 doi:10.1101/2020.12.13.422570, BioRxiv
- 559 43. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis.
- 560 *Bioinformatics* **28**, 333–339 (2012).
- 561 44. Kleinheinz, K. *et al.* ACEseq-allele specific copy number estimation from whole genome sequencing.
- 562 (2017) doi:10.1101/210807, BioRxiv
- 563 45. Li, B. *et al.*, Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic
- 564 leukemia. *Blood*. 2020; 135(1): 41–55.
- 565 46. Gröschel, S., Hübschmann, D., Raimondi, F. *et al.* Defective homologous recombination DNA repair as
- 566 therapeutic target in advanced chordoma. *Nat Commun* **10**, 1635 (2019).
- 567 47. Hovestadt, V., Jones, D., Picelli, S. *et al.* Decoding the regulatory landscape of medulloblastoma using
- 568 DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
- 569 48. Degasperi, A., Amarante, T.D., Czarnecki, J. *et al.* A practical framework and online tool for mutational
- 570 signature analyses show intertissue variation and driver dependencies. *Nat Cancer* **1**, 249–263 (2020).
- 571
- 572
- 573
- 574

575 **Figure Legends**

576

577 **Figure 1: Mutation burden of SBSs and small indels (IDs) across 20 pediatric tumor**

578 **types.**

579 a) Single base substitution (SBS) and small insertion/deletion (ID) mutation burden of the

580 pediatric pan-cancer whole-genome sequenced (PPC-WGS) cohort. The numbers of samples

581 for each tumor type are shown next to the labels. Each dot represents one tumor sample.

582 Tumor types are ordered by the median numbers of single-base substitutions. b) Correlation

583 between SBS and ID mutations per megabase across the cohort.

584

585 **Figure 2: SBS and ID signature activity across pediatric cancers**

586 a) Example profiles of SBS and ID signatures extracted from the PPC-WGS cohort that are

587 similar to COSMIC v.3 SBS3 and ID6 signatures. b) The number of mutations contributed by

588 each mutational signature to the PPC-WGS tumors. Circle size indicates fraction of tumors

589 with signature activity in a cancer type and the color indicates the median number of

590 mutations per megabase due to a signature in a specific entity. c) Correlation between SBS

591 and ID signature activities.

592

593 **Figure 3: HR defect signatures activity in pediatric cancers**

594 a) Fraction of all PPC-WGS cohort samples that show activity of Signature.3 (COSMIC v.2)

595 and SBS3 (COSMIC v.3). b) Normalized SBS and ID signature activities in tumors with

596 SBS3 signature activity c) BRCA1/2 mutation status of 5 INFORM whole genome sequenced

597 tumors and their SBS and ID signature activities (MB_SHH – Sonic hedgehog subgroup of

598 medulloblastoma; HGG – high-grade glioma; DIPG – diffuse intrinsic pontine glioma; RMS -

599 rhabdomyosarcoma) d) Comparison of microhomology deletion proportion between PCAWG

600 and PPC-WGS cohorts e) SBS, ID signature activities, microhomology deletion proportions
601 and genome wide copy-number profile of osteosarcoma tumor SJOS004.

602

603 **Figure 4: Similarity between Signature.P1 and COSMIC v.3 SBS reference signatures**

604 a) Cosine similarity between novel Signature.P1 from Gröbner, Worst et al., 2018 and

605 COSMIC v.3 SBS signatures. b) 96-context mutational profiles of Signature.P1 and SBS31.

606 c) SBS31 signature activity in the PPC-WGS cohort.

607

608

609 **Supp. Figure 1: Correlation of SBS and ID mutations with age at tumor diagnosis**

610 a) Comparison of number of mutation calls identified between 2018 (in Gröbner, Worst et al.,

611 2018) and the updated 2020 DKFZ in-house pipeline. b) Correlation of SBS and ID mutations

612 with age across the PPC-WGS cohort.

613

614 **Supp. Figure 2: Normalized mutational signature activities across the cohort per sample**

615 a) Normalized SBS signature activities across the cohort b) Normalized ID signature activities

616

617 **Supp. Figure 3: Association of SBS signature activities and genomic alterations**

618 a) Correlation of SBS1, SBS5 and SBS40 signature activities and age at diagnosis. b)

619 Association of *TP53* mutation status and SBS2 and SBS13 signature activities c) Difference

620 in number of CC>TT double base substitutions between SBS7a/b positive and negative

621 tumors of B-ALL-HYPO tumors. d) Association of *TP53* mutation status and SBS8 signature

622 activity e) Number of IGHV mutations and SBS9 signature activities in Burkitt's lymphoma

623 samples.

624 f) Profile of the novel SBS signature identified in Group4-subgroup of medulloblastoma

625 (Group4MB-SBS96D) g) Association of chromothripsis and SBS8 and SBS40 signature

626 activities. h) Association of MYCN-amplification status and SBS18 signature activity i)

627 Correlation of genomic instability and SBS2, SBS13 and SBS40 signature activities.

628

629 **Supp. Figure 4: Association of ID signature activities and genomic alterations**

630 a) Correlation of ID1, ID2, ID5 signature activities and age at diagnosis b) Association of

631 TP53 mutation status and ID1, ID2 and ID9 signature activities c) Association of

632 chromothripsis and ID2, ID9 signature activities d) Correlation of genomic instability and ID9

633 signature activity

634

635 **Supp. Figure 5: ID signatures in adult cohort and downstream analysis of SBS3 positive tumors**

636 a) ID signature activities in adult tumors of NCT-MASTER b) Correlation of SBS mutation

637 burden and age at diagnosis in medulloblastoma c) HR pathway gene expression difference

638 between SBS3 positive and negative tumors of Group3 subgroup of medulloblastoma d) HR

639 pathway gene expression in INFORM whole genome sequenced tumors

640

641 **Supp. Figure 6: Genome-wide copy-number profiles of SBS3 positive tumors from the PedPanCan cohort and INFORM tumors (n=5)**

642 TCN = Tumor Copy Number. Red indicates loss of genomic material, green indicates gains.

643 Plots were generated using the ACEseq tool.

644

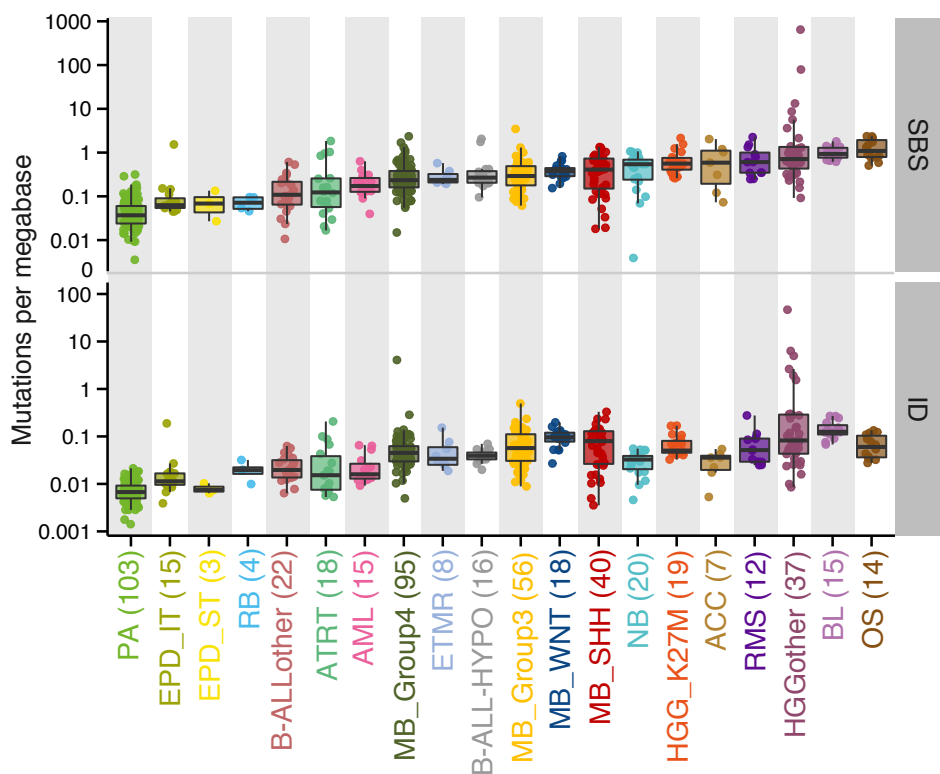
645

646

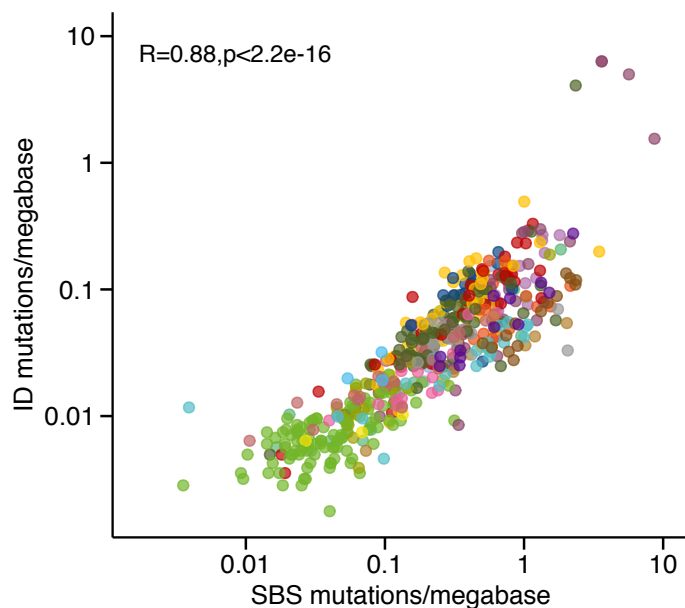
Fig. 1: Tumor mutation burden of pediatric cancers

a

PPC-WGS cohort (n=537)



b



- ACC, Adrenocortical carcinoma
- AML, Acute myeloid leukemia
- ATRT, Atypical teratoid rhabdoid tumors
- B-ALL-HYPO, B-cell acute lymphoblastic leukemia, hypodiploid
- B-ALLother, B-cell acute lymphoblastic leukemia, non-hypodiploid
- BL, Burkitt's lymphoma
- EPD_IT, Ependymoma infratentorial
- EPD_ST, Ependymoma supratentorial
- ETMR, Embryonal tumors with multilayered rosettes
- HGG_K27M, High-grade glioma K27M

- HGGothen, High-grade glioma K27wt
- MB_Group3, Medulloblastoma Group3
- MB_Group4, Medulloblastoma Group4
- MB_SHH, Medulloblastoma SHH
- MB_WNT, Medulloblastoma WNT
- NB, Neuroblastoma
- OS, Osteosarcoma
- PA, Pilocytic astrocytoma
- RB, Retinoblastoma
- RMS, Rhabdomyosarcoma

Fig. 2: SBS and ID signatures in pediatric cancers

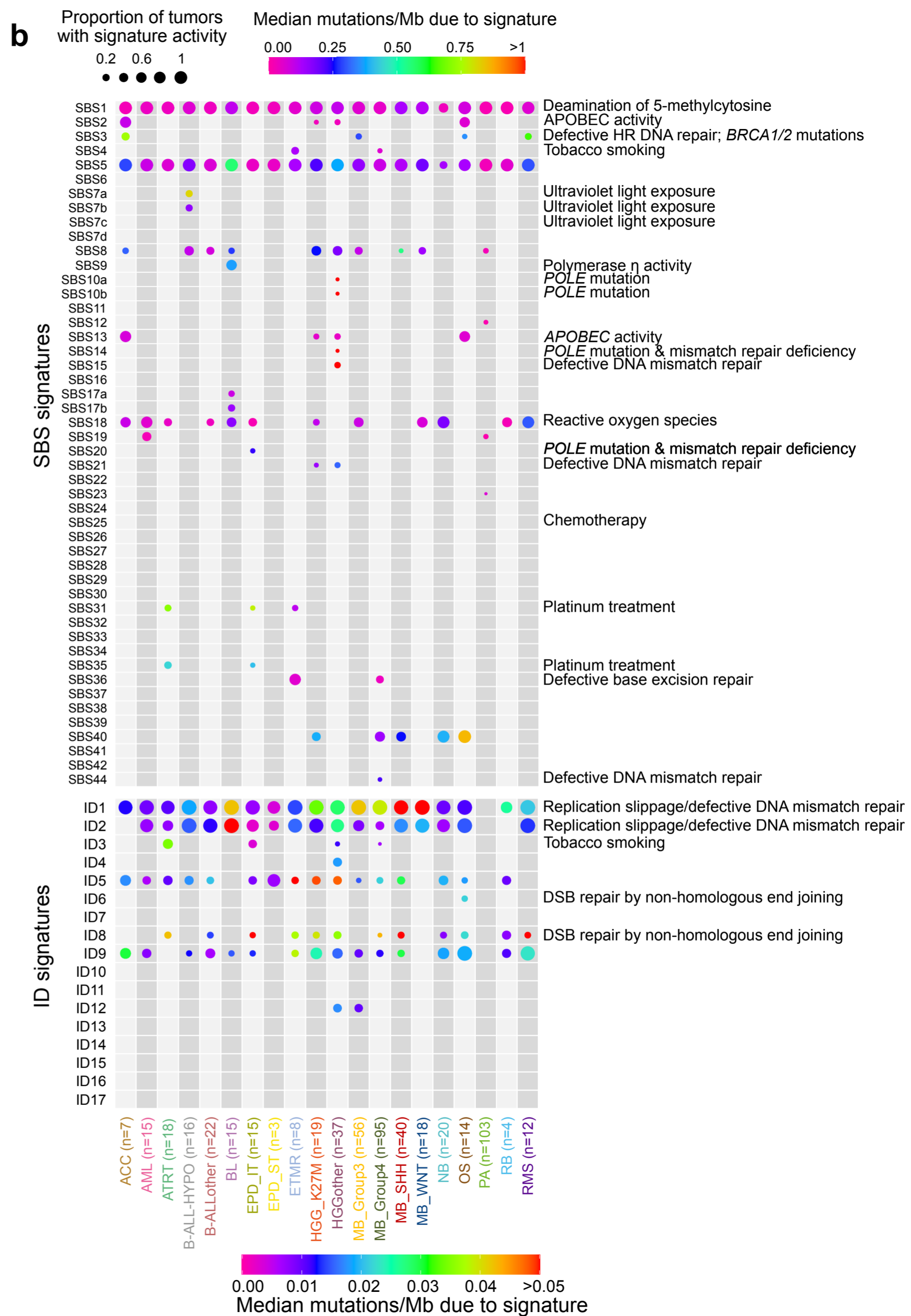
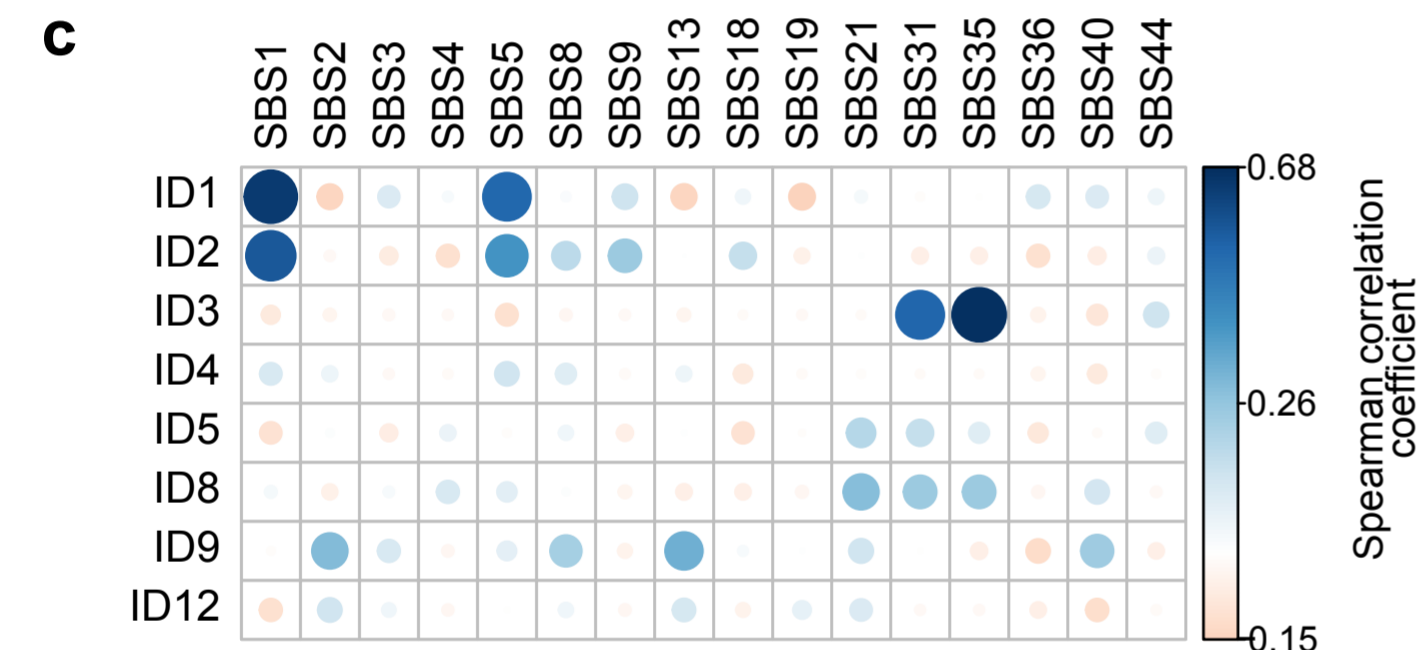
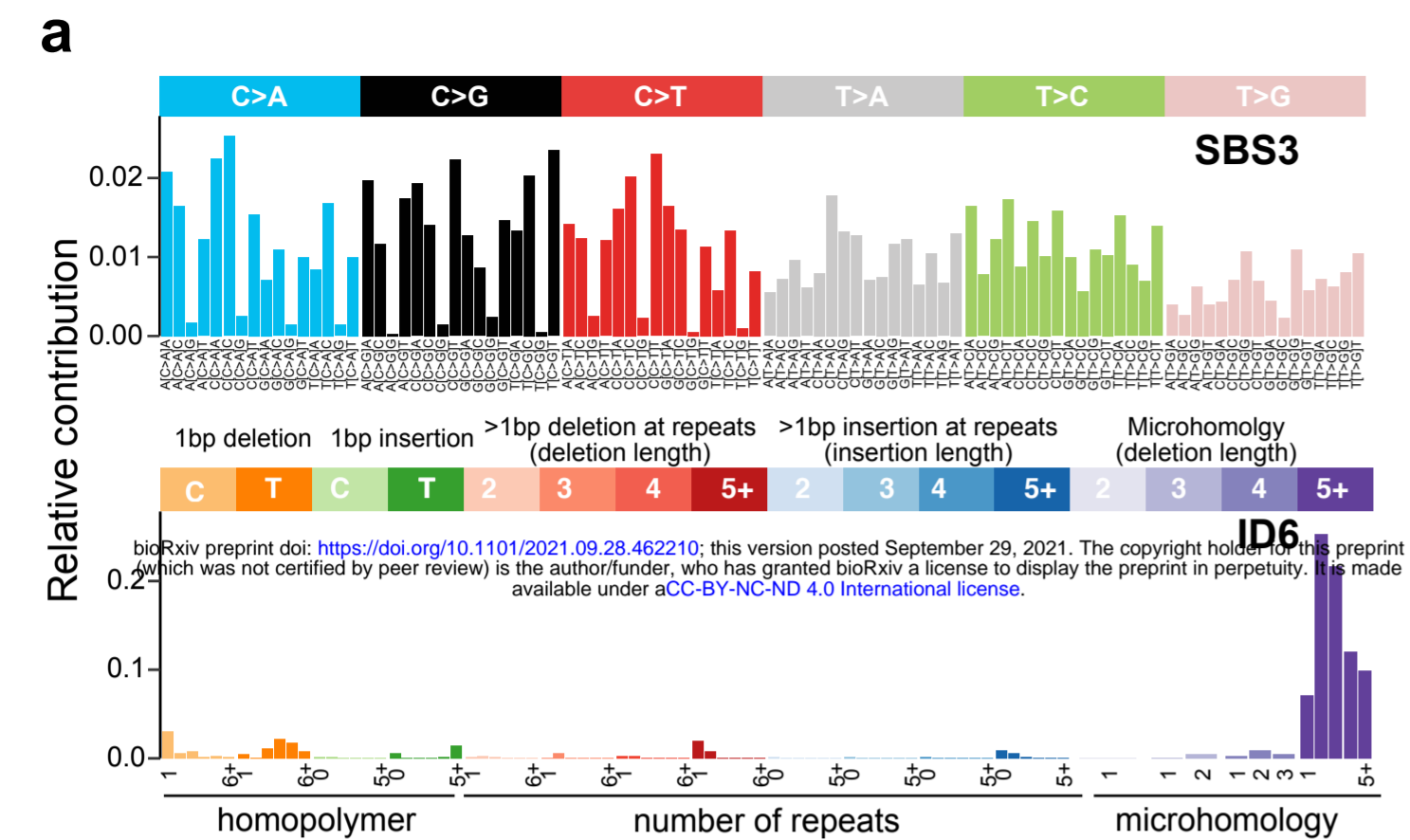
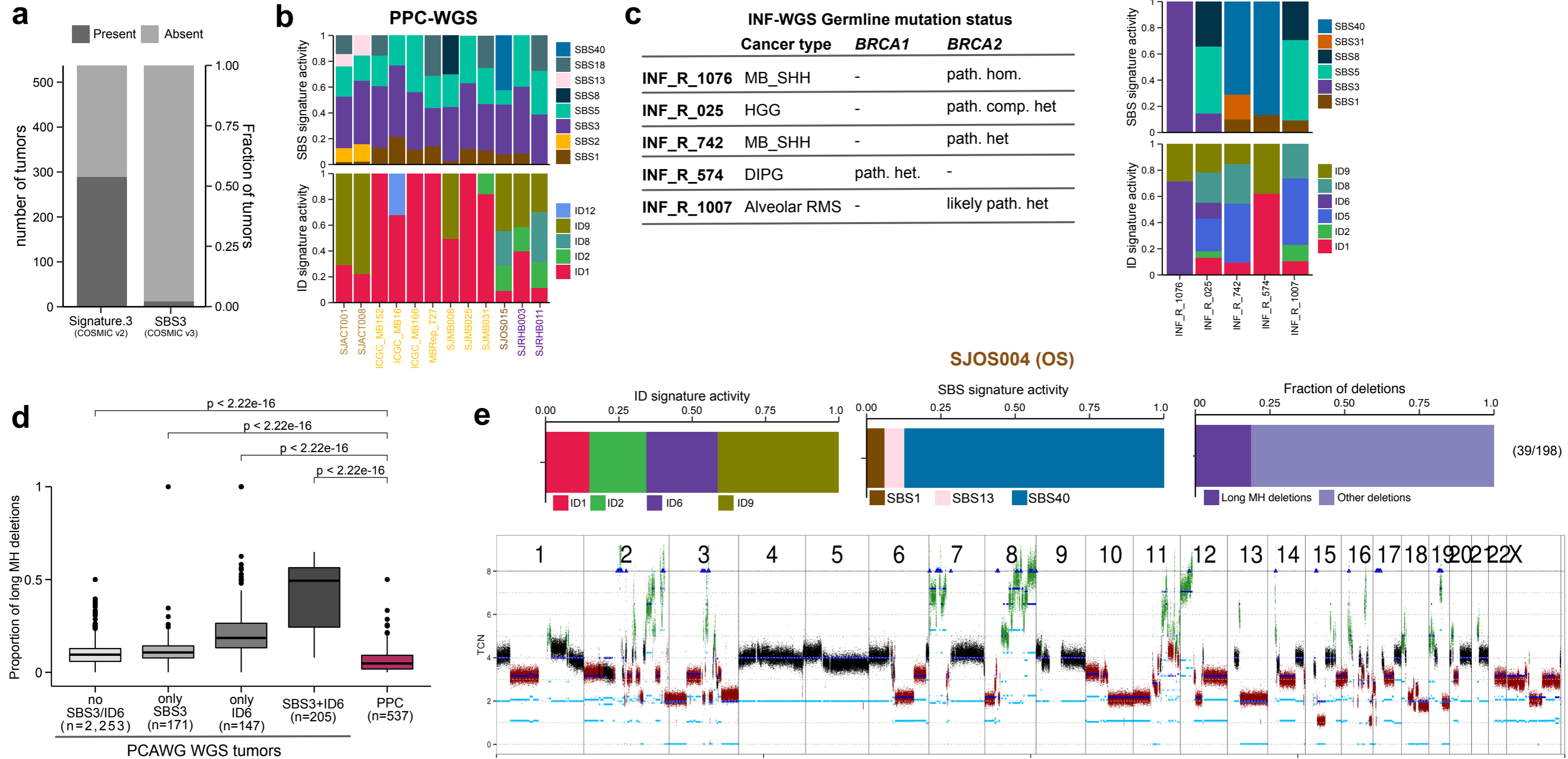
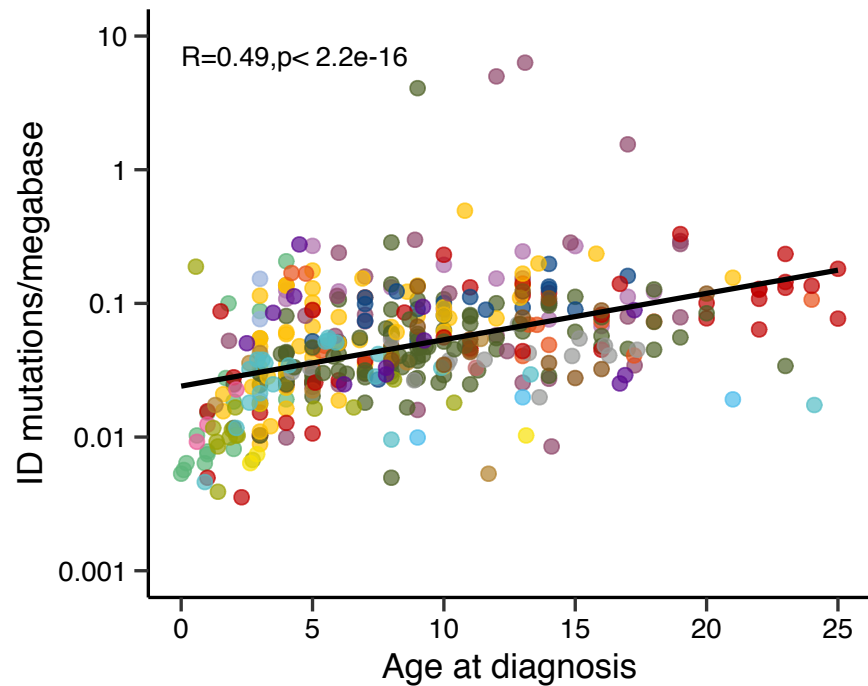
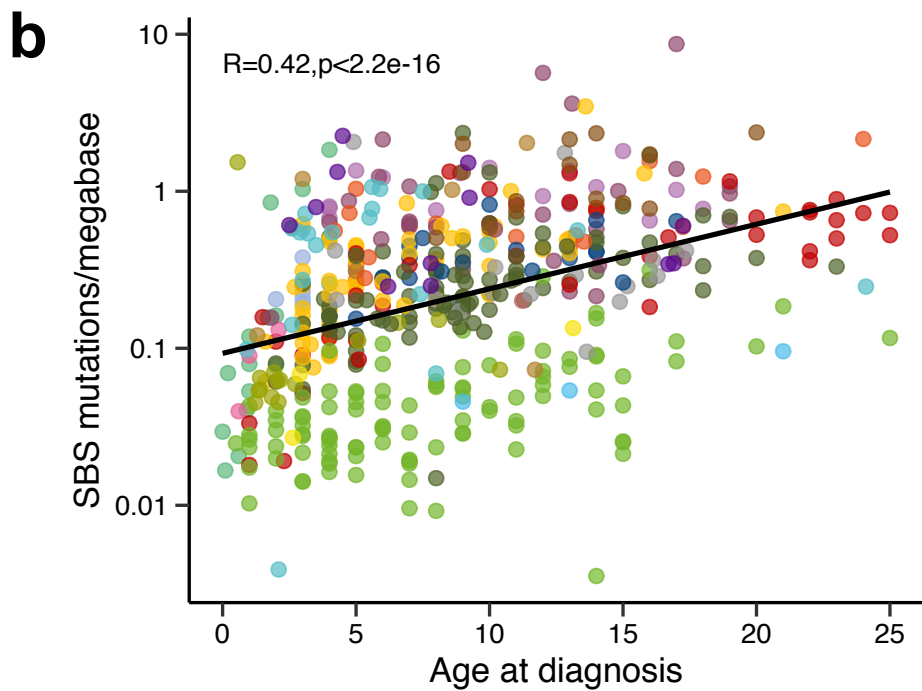
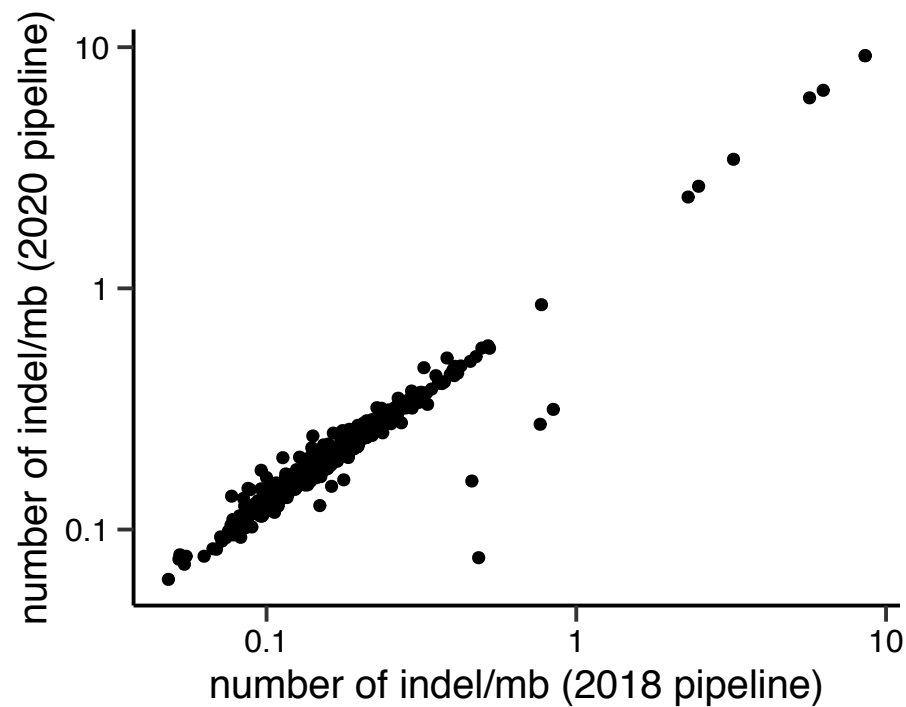
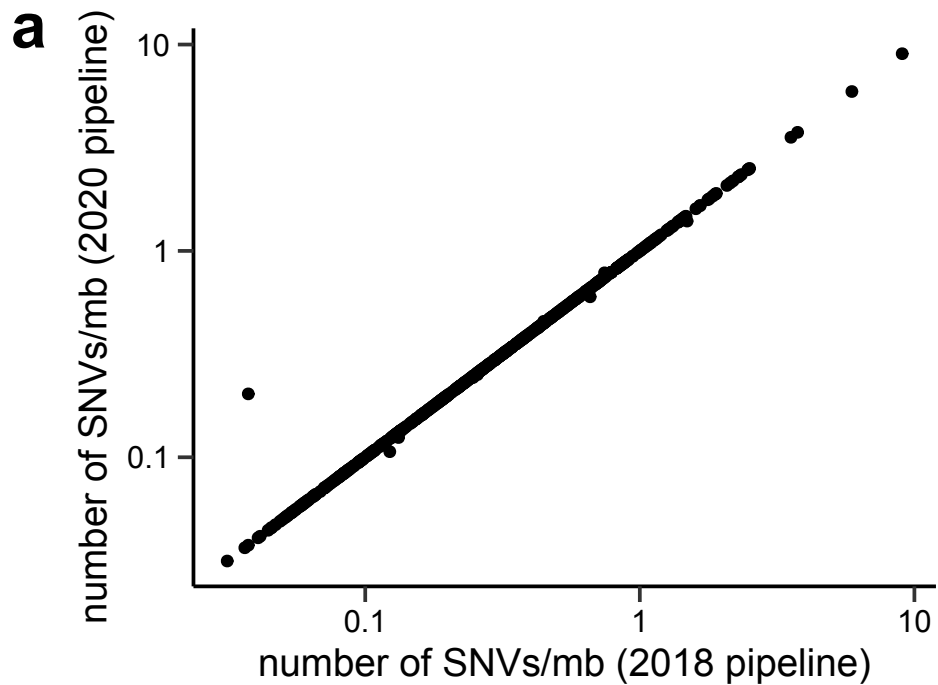


Fig. 3: HR defect signatures activity in pediatric cancers

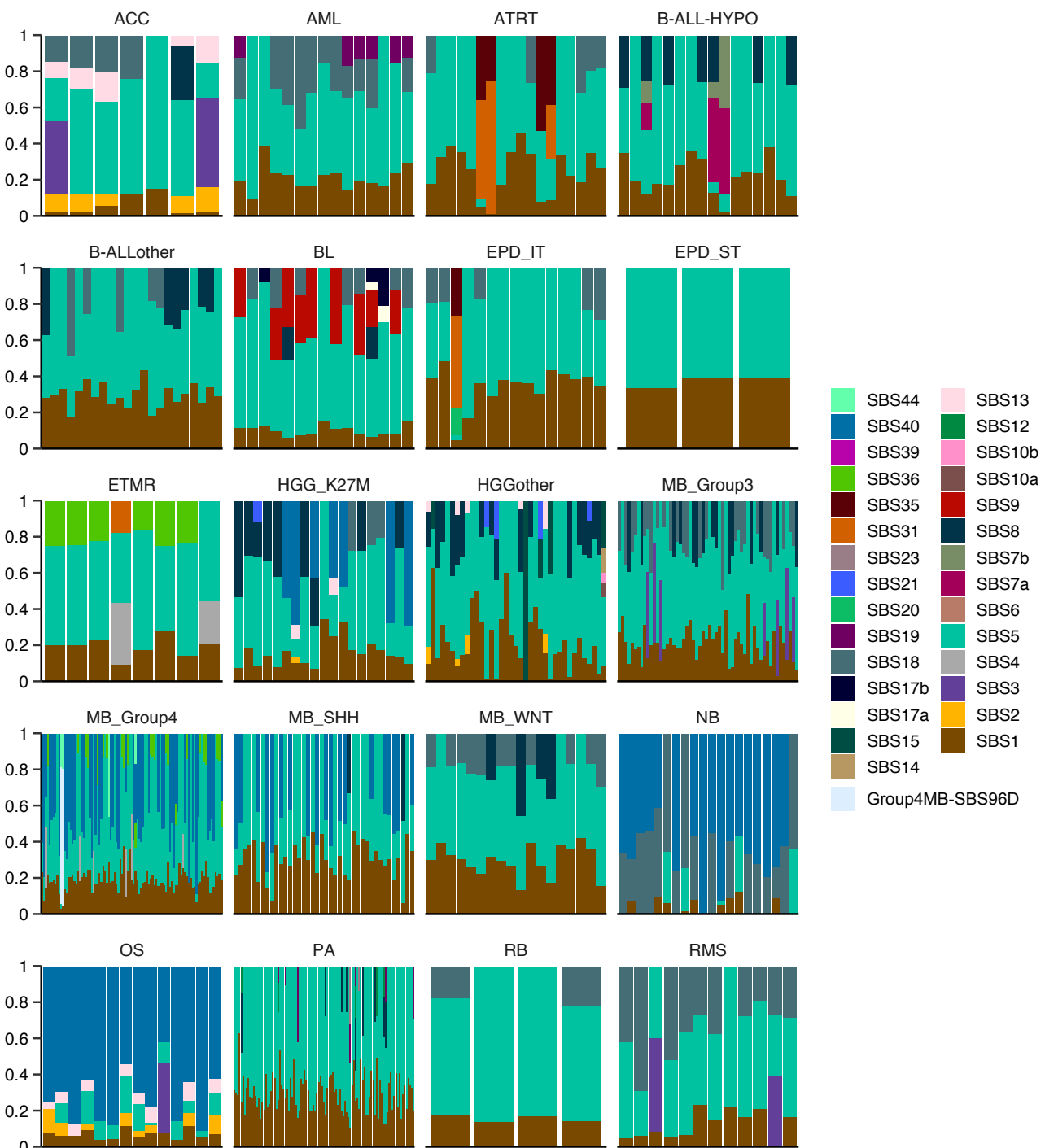


Supp Fig. 1:

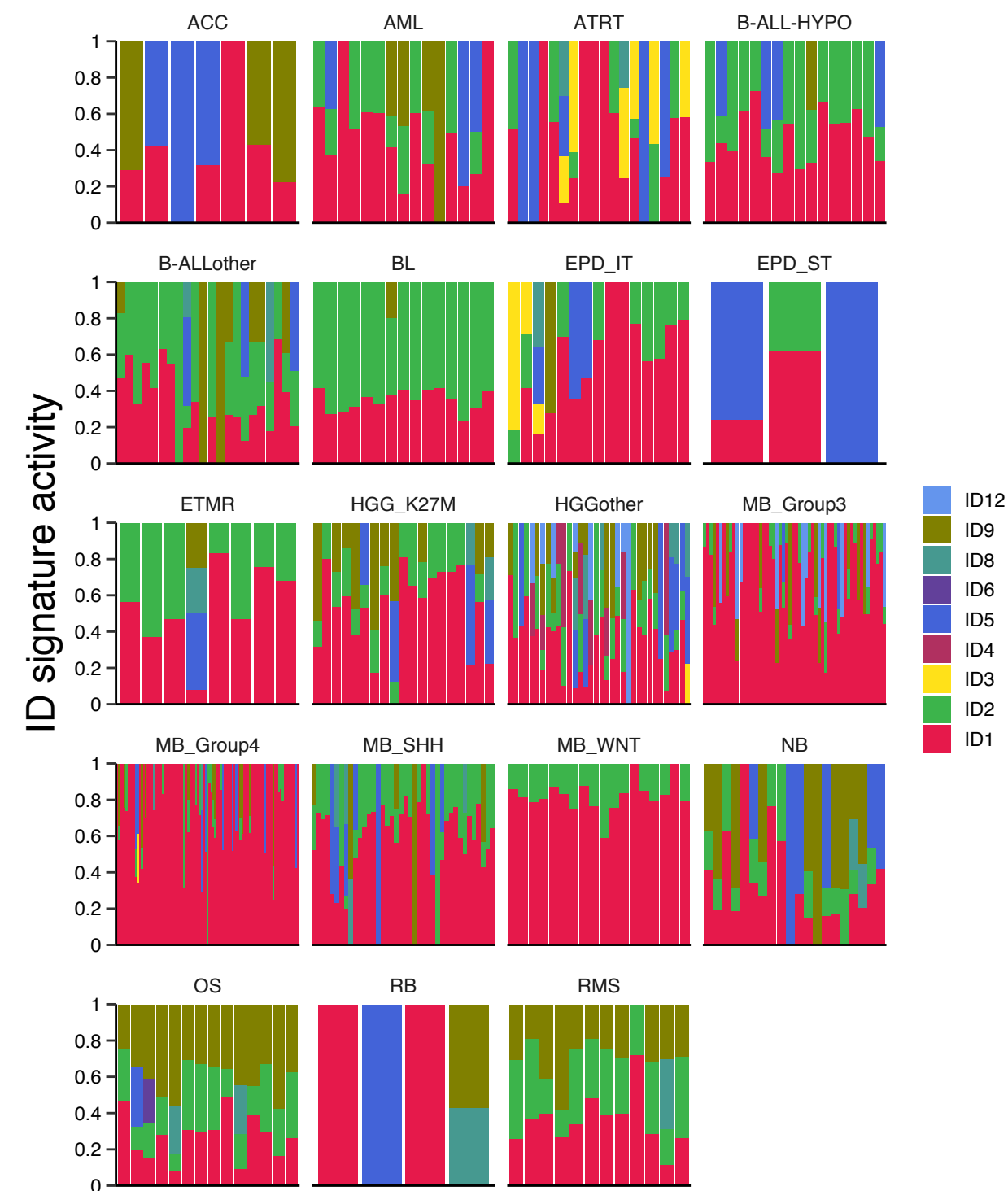


Supp Fig. 2:

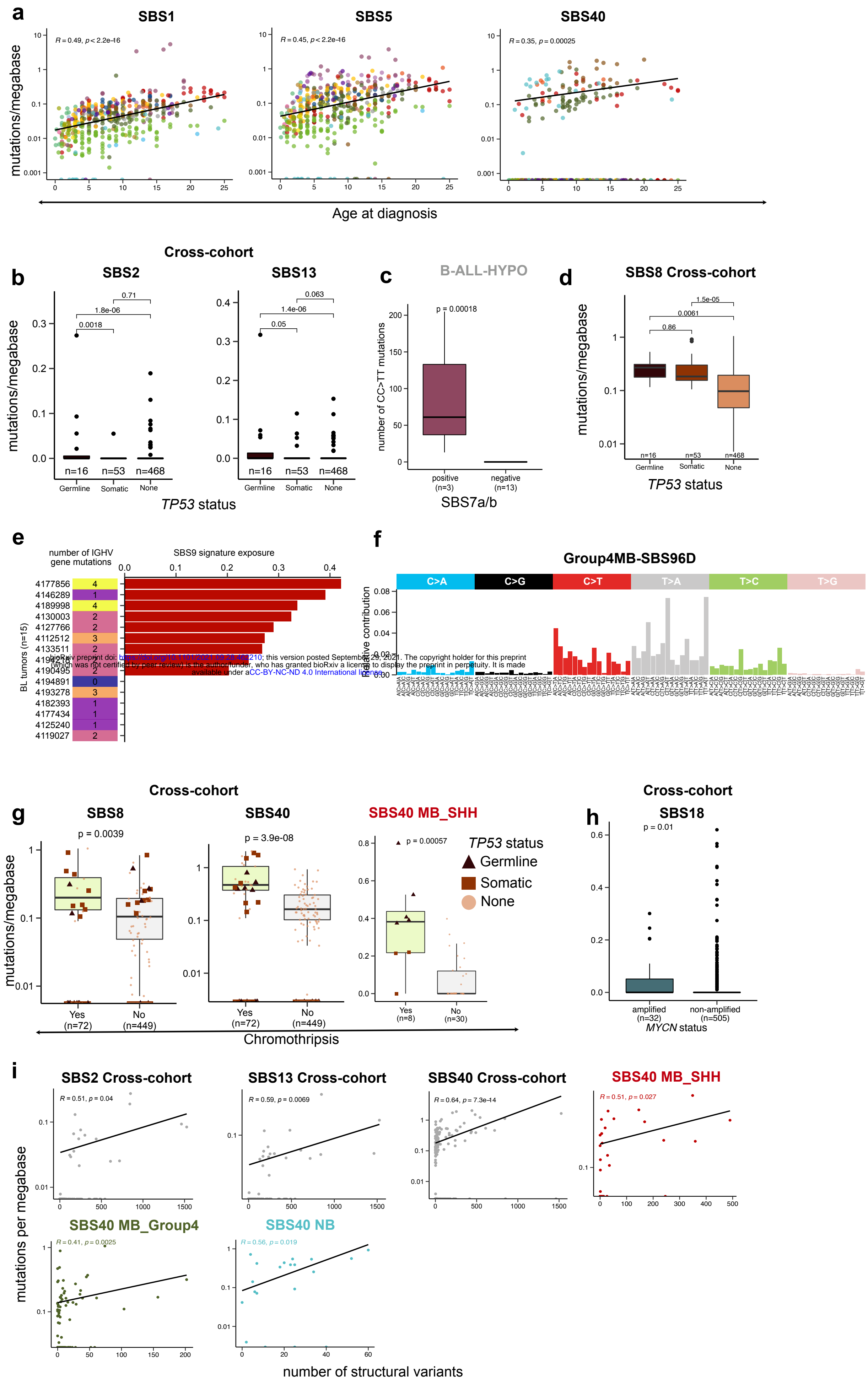
a



b

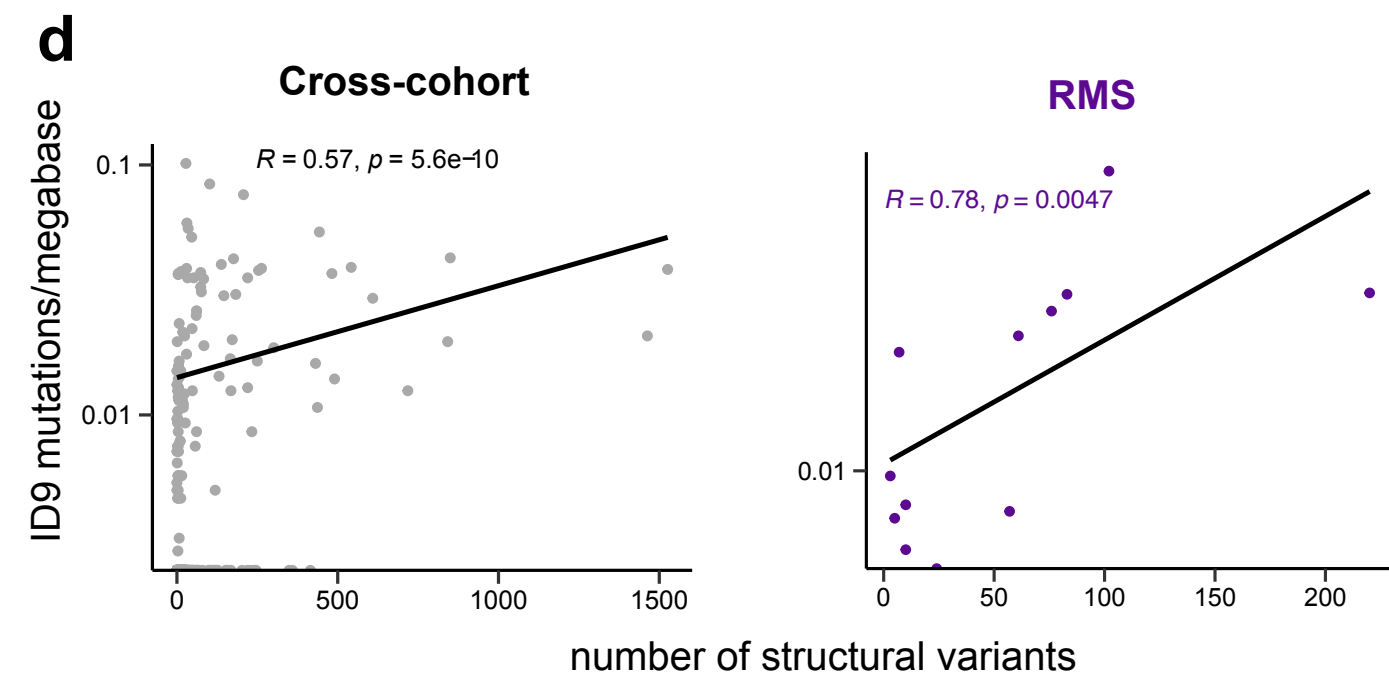
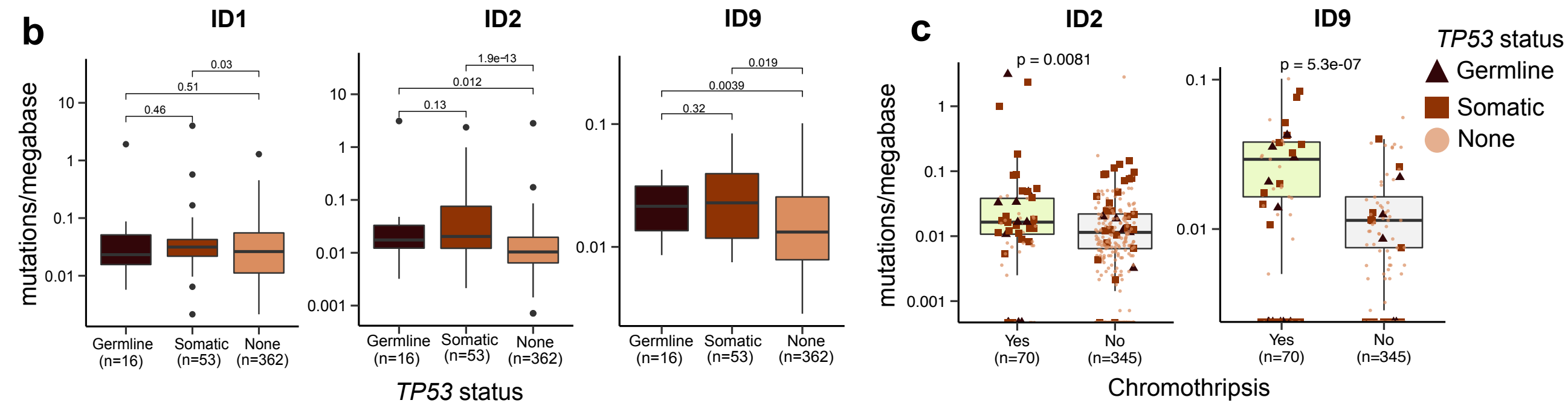
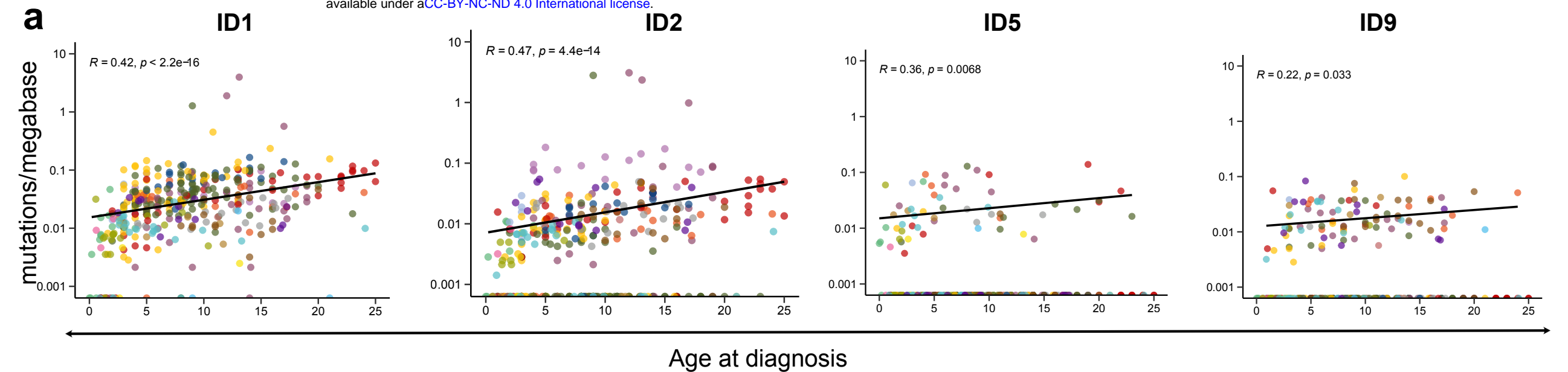


Supp Fig. 3:

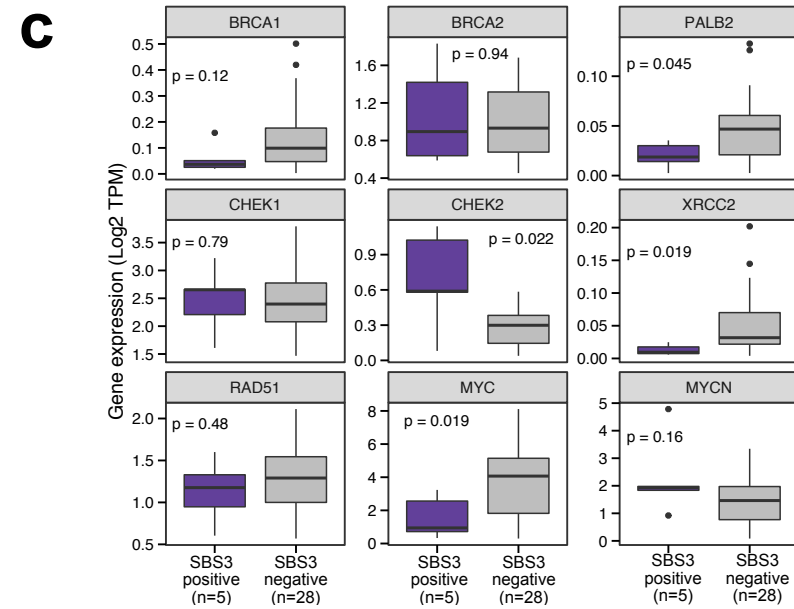
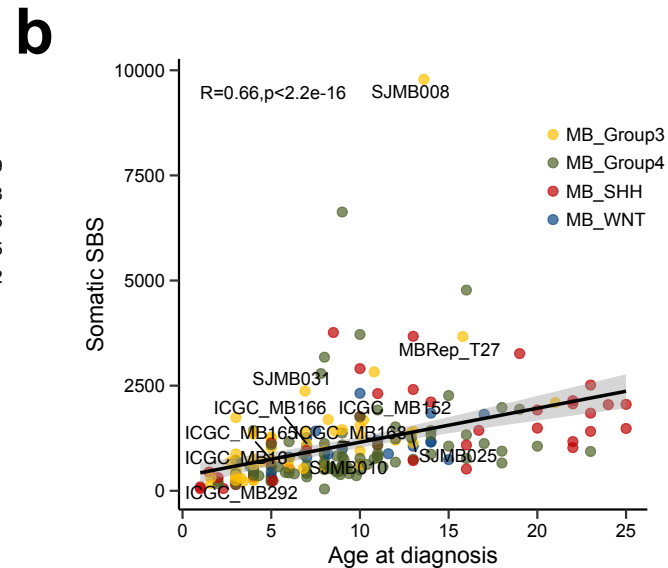
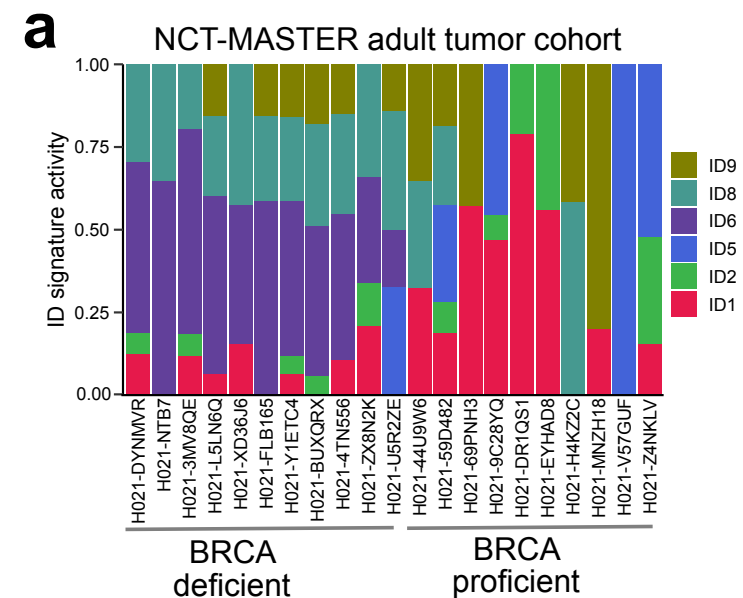


Supp Fig. 4:

bioRxiv preprint doi: <https://doi.org/10.1101/2021.09.28.462210>; this version posted September 29, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supp Fig. 5:



d

Expression in INFORM tumors (TPM)

	Cancer type	<i>BRCA1</i>	<i>BRCA2</i>	<i>PALB2</i>
INF_R_1076	MB_SHH	20.41	2.38	22.44
INF_R_025	HGG	9.53	2.03	10.74
INF_R_742	MB_SHH	22.06	4.33	19.54
INF_R_574	DIPG	13.68	4.38	13.77
INF_R_1007	Alveolar RMS	19.41	10.03	19.71

Supp Fig. 6:

