1 # Clonal evolution during metastatic spread in high-risk neuroblastoma

2 Gunes Gundem[1,2,┼], Max F. Levine[1,2], Stephen S. Roberts[1], Irene Y Cheung[1], Juan S. Medina-

3 Martínez[1,2], Yi Feng[1], Juan E. Arango-Ossa[1,2], Loic Chadoutaud[2], Mathieu Rita[2], Georgios Asimomitis[2],

4 Joe Zhou[1,2], Daoqi You[1], Nancy Bouvier[1], Barbara Spitzer[1], David B. Solit[3,4], Filemon Dela Cruz[1],

5 Michael P. LaQuaglia[1], Brian H. Kushner[1], Shakeel Modak[1], Neerav Shukla[1], Christine A. Iacobuzio-

6 Donahue[5,6,7], Andrew L. Kung[1], Nai-Kong V. Cheung[1*] and Elli Papaemmanuil[1,2,┼,*]

7

8 [1]Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

9 [2]Computational Oncology Service, Department of Epidemiology & Biostatistics, Memorial Sloan
10 Kettering Cancer Center, New York, NY, USA.

11 [3]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

12 [4]Marie-Josée and Henry R. Kravis Center for Molecular Oncology, New York, NY, USA

13 [5]The David M. Rubenstein Center for Pancreatic Cancer Research, Sloan Kettering Institute,
14 Memorial Sloan Kettering Cancer Center, New York, NY, USA

15 [6]Human Oncology and Pathogenesis Program, Sloan Kettering Institute, Memorial Sloan Kettering
16 Cancer Center, New York, NY, USA

17 [7]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

18
19 *Co-last authors
20 ┼Corresponding author
21
22 **Address for correspondence:** papaemme@mskcc.org; gundem@mskcc.org

23
24
25
26
27
28
29
30
31
32

33 **Abstract (100)**

34 High-risk neuroblastoma is generally metastatic and often lethal. Using genomic profiling of 470

35 sequential and spatially separated samples from 283 patients, we characterize subtype-specific

36 genetic evolutionary trajectories from diagnosis, through progression and end-stage metastatic

37 disease. Clonal tracing timed disease initiation to embryogenesis. Continuous acquisition of

38 structural variants at disease defining loci (*MYCN*, *TERT*, *MDM2-CDK4*) followed by convergent

39 evolution of mutations targeting shared pathways emerged as the predominant feature of

40 progression. At diagnosis metastatic clones were already established at distant sites where they

41 could stay dormant, only to cause relapses years later and spread via metastasis-to-metastasis and

42 polyclonal seeding after therapy.

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58 **Introduction**

59 Neuroblastoma is an embryonal tumor arising from the developing sympathetic nervous system

60 accounting for 15% of pediatric cancer mortality[1]. Disease presentation is highly heterogeneous and

61 ranges from low-risk local-regional tumors to widely disseminated high-risk disease seen in two

62 thirds of the patients. For high-risk neuroblastoma (HR-NB), modern clinical management includes

63 multimodal chemotherapy, surgical resection, radiotherapy and immunotherapy. Nevertheless,

64 despite intensive treatment 50% of HR-NB patients still relapse with fatal outcomes[2].

65

66 Notwithstanding the metastatic nature of HR-NB, most genomic studies of disease progression

67 focused on small cohorts of paired diagnostic-relapse tumors[3-9]. Recently, broad copy number

68 aberrations (CNA) and whole-exome sequencing (WES) from multi-region biopsies suggested

69 elevated genetic heterogeneity in high-risk disease[10-12]. However, the majority of disease-defining

70 alterations in HR-NB[13,14] result from structural variants (SVs) that cannot be captured by low-

71 resolution WES/CNA analysis. Todate, the temporal and spatial genomic features of disease

72 progression as patients go through multiple lines of therapy are not well understood. Here, we

73 leverage the MSKCC neuroblastoma biobank to study a unique cohort of 470 tumors from 283

74 patients representative of HR-NB at diagnosis, consecutive relapses, and diverse metastatic sites.

75 Using a combination of whole genome (WGS) and targeted sequencing approaches we characterized

76 the composite genetic alterations associated with neuroblastoma pathogenesis and define the

77 lineage relationships during disease progression across diverse neuroblastoma molecular subtypes.

78

79 **Results**

80 **Cohort ascertainment**

81 Our cohort consisted of 470 tumors from 283 patients with predominantly stage-4 disease (87%)

82 and/or spatially and temporally separate tumors available (Fig. 1a and Supplementary Table 1).

83   Fresh-frozen surgical specimens were collected at consecutive clinical intervention time points

84   including: 110 pre-treatment diagnostic samples, 5 therapy-naive re-resections, 132 therapy

85   resection during induction chemotherapy (t-resection), 111 relapses and 112 further relapses (1-17

86   samples per patient) and spanned spatially separated disease sites including: 217 samples from

87   primary site, 150 from local-regional spread and 97 from metastases as defined by clinical

88   guidelines[15] including rare metastatic sites such brain. A web portal describing the treatment

89   timelines and clonal phylogenies for patients with two or more tumors is provided in

90   (https://master.d32nckcows37aj.amplifyapp.com/) and detailed summaries of genomic and clonal

91   evolution patterns in 45 patients with multi-WGS data are provided in Supplementary Information.

92

93   **Landscape of genomic alterations**

94   Comprehensive genomic profiling (substitutions, indels, SVs and CNAs) (Extended Data Fig. 1a-b,

95   Supplementary Table 2) identified both established and novel gene mutations linked to

96   neuroblastoma pathogenesis. The genomic landscape was representative of HR-NB[16,13,17]. Only 4

97   genes (*MYCN*, *ALK*, *TERT, ATRX)* had mutations in >10% of patients while recurrent *MDM2-CDK4* co-

98   amplification was observed in 3%[18]. Alterations in *MYCN*, *TERT*, *ATRX* and *MDM2-CDK4* defined

99   mostly non-overlapping disease subtypes explaining 51% of the cohort while *ALK* mutations were

100   shared across the cohort. 41% and 8% of the patients did not have any subtype-defining alterations

101   other than segmental (SEG-CNA) or numeric chromosome-level CNAs (NUM-CNA), respectively.

102

103   *TERT* rearrangements are common in neuroblastoma[13,14]. Here, we report *TERT*p substitutions in 9

104   patients comprising 17% of the *TERT* events. However, contrary to *TERT* SVs which are mutually

105   exclusive to other subtype defining events, *TERT*p mutations were enriched in *MYCN*-amplified

106   (*MYCN*-A) neuroblastoma and demarcated a group of *MYCN*-A patients with a trend for rapid

107   progression and death within 2 years from diagnosis (Extended Data Fig. 1c). PI3K-mTOR pathway

108 was also mutated in 5% of the patients suggesting enrichment in HR-NB compared to primary

109 neuroblastoma (<1% in Brady et al[14]). Additionally, mutations in neuroblastoma differentiation

110 genes including *RARA*, *RARB*, *PHOX2B*, *SPRY2*, *IGF2BP3* and *WNT5A* were observed in 4% of the

111 patients at relapse (Supplementary Information).

112

113 **Evolution of mutational landscape in response to therapy**

114 Analysis of genome-wide mutation landscapes revealed distinct mutational patterns at diagnosis and

115 relapse (Extended Data Fig. 2a-b and Supplementary Table 1). At diagnosis, two substitution

116 signatures, SBS40 and SBS18, were differentially enriched across molecular subtypes. Mutations

117 attributed to SBS40, similar to the clock-like signature SBS5[19], was higher in *ATRX*-mutated patients

118 and correlated with age at diagnosis, while SBS18, which is predominantly defined by C>A mutations,

119 prevailed in *MYCN*-A tumors in agreement with prior literature[14] and did not correlate with age (Fig.

120 1b and Extended Data Fig. 3a). SBS18 was first described in neuroblastoma[20] and attributed to

121 reactive oxygen species (ROS)[21]. *MYCN*-A enhances glutaminolysis in neural crest progenitor cells,

122 which in turn induces oxidative stress by ROS production[22]. Expression of glutaminolysis signature

123 was higher in *MYCN*-A tumors compared to other subtypes (Extended Data Fig. 3b). This provides a

124 plausible mechanistic link between *MYCN*-A, metabolic reprogramming and SBS18 burden. Notably,

125 the glutaminolysis gene expression signature and the rate of accumulation of SBS18 remained stable

126 during disease progression (Extended Data Fig. 3b).

127

128 Pediatric tumors are defined by low tumor mutation burden (TMB)[23]. TMB was low at diagnosis[24,23]

129 (median=0.52 muts/Mb, range=0.06-2.6) (Fig. 1c) but increased significantly during disease

130 progression (median=2.2 muts/Mb, range=0.2-9). Notably, in patients with matched

131 diagnostic/relapse tumors, TMB increased by 6- and 14-fold at first and later relapses (Fig. 1d),

132 respectively, approximating TMB ranges seen in adult tumors[20]. At diagnosis neuroblastoma is

133 characterized by an immune-cold tumor microenvironment[25-27] (TME) with poor responses to

134 immune checkpoint blockade therapy[28,29]. We evaluated whether the increased TMB during disease

135 progression presented a therapeutic opportunity mediated by putative neo-antigens[30]. However,

136 despite the increase in predicted neoantigen burden, there was no association with transcriptional

137 patterns suggestive of an immunomodulatory switch during disease progression (Extended Data Fig.

138 3c).

139

140 At relapse, increase in TMB was associated with exposure to chemotherapy-associated mutation

141 signatures[31]. Specifically, three substitution signatures (TMZ, SBS31 and SBS35) dominated by T>C

142 and C>T mutations correlated with exposure to temozolomide and platinum with evidence of strong

143 dose-response relationships[32,21,31,19] (Fig. 1e-f). At disease progression, tumors from prior radiated

144 sites had an excess of small deletions, SV deletions, reciprocal translocations and complex events[33,34]

145 (Supplementary Fig. 1). This demonstrates that therapy directly molds the mutation landscape of

146 HR-NB tumors. Thus, we next evaluated the effect of these mutation processes in the driver

147 landscape at diagnosis and during disease progression. Of 82 oncogenic substitutions from WGS data,

148 48 were present at diagnosis and 34 emerged at relapse. Notably, only 12% of the oncogenic

149 substitutions were assigned to a therapy-related signature compared to all relapse-specific SNVs

150 (34%) (Extended Data Fig. 3d).

151

152 **Timing the emergence of the initial neuroblastoma clone**

153 For each patient, clonal reconstruction of tumor phylogenies delineated the trunk marked by the

154 mutations found in 100% of malignant cells in all tumors of a patient as well as subclonal events (not

155 on trunk). Trunk represents the most recent common ancestor (MRCA) (Online Methods and

156 Supplementary Information, Supplementary Fig. 2-46 and Supplementary Tables 3-4). The number

157 of clock-like mutations on the trunk can be used to estimate the time of MRCA emergence[35-37]. Across

158   39 evaluable patients the number of truncal substitutions (trunk length) was low (median=753,

159   range=11-5801) and correlated with age at diagnosis only when disease subtype, stage, and number

160   of tumors were taken into account (Fig. 2a and Supplementary Fig. 47). Low-stage disease tended to

161   have shorter trunks (Fig. 2b) with lengths comparable to the number of substitutions detected in

162   non-malignant clones in bulk placenta also enriched for SBS18 mutations[38]. This suggests that the

163   first malignant clone emerges in similar time frames during embryogenesis. Indeed, chronological

164   timing of MRCA emergence using the clock-like SBS40 mutations[37] confidently pinpointed an

165   embryonic and post-natal origin in 6 and 8 cases, respectively. For the remaining 25 patients the

166   confidence intervals were large, owing to low numbers of mutations on the trunk (Fig. 2c). Notably,

167   MYCN-A was common in embryonic origin (4/6) while ATRX-mutant disease was enriched for post-

168   natal onset (7/8) especially in patients with *ATRX* truncating mutations (Fig. 2c).

169

Subtype-specific evolutionary trajectories underwrite disease progression

171   Analyses of multiple samples representative of the clinical course of treatment from 94 patients

172   across HR-NB subgroups identified previously unappreciated and subtype-specific evolutionary

173   trajectories for tumors with *MYCN*-A, *TERT*-SV, *ATRX* events and *MDM2-CDK4* co-amplification (Fig.

174   2d-e, Extended Data Fig. 4-5 and Supplementary Data fig. 2-46 and 48-53). While subclonal events at

175   RAS-MAPK[4,5,39] and PI3K-mTOR pathways were common across subtypes, the acquisition of SVs

176   emerged as critical events in disease evolution with a striking propensity to repeatedly target the

177   main subtype-defining driver genes *MYCN, TERT* and *ATRX* (Fig. 2d-e).

178

179   Tumor-initiating *MYCN* amplifications in neuroblastoma[40] are frequently found in extrachromosomal

180   DNA[41,42], which may result from simple or complex SVs at the trunk of the tumor phylogeny. Amongst

181   patients with multi-WGS data, we observe continuous rearrangements of the *MYCN* locus in both

182   primary (7/10 cases) and relapse sites (9/11 patients) (Fig. 2e, Fig. 3a and Supplementary Fig. 54-

183   56). During therapy, rearrangements at the *MYCN* locus continued to accumulate in 9/10 patients

184   and in four of these patients a clone with chromothripsis at the *MYCN* locus dominated across

185   metastatic sites without evidence of further diversification (Supplementary Fig. 54-56).

186

187   *TERT*-SVs are mutually exclusive with *MYCN* and *ATRX* rearrangements, thus demarcate a distinct

188   HR-NB subtype[13,17]. In our cohort *TERT*-SVs were enriched in CNS metastases (8/19 patients)[43].

189   Intriguingly, unlike *MYCN*, where an initial amplification is always present on the trunk, *TERT*-SVs

190   were predominantly (10/13 SVs) subclonal to segmental CNAs (Fig. 2d-e, Fig. 3b and Extended Data

191   Fig. 5). Similar to the *MYCN* locus, we observe continuous *TERT* rearrangements in the majority of

192   cases (5/7) during disease progression. These relapse-specific rearrangements result in increased

193   copy number and expression of *TERT* consistent with an increasing addiction to *TERT* signaling (Fig.

194   3b, Extended Data Fig. 5 and Supplementary Fig. 3, 5, 18-20 and 24-25).

195

196   *ATRX*-mutant neuroblastoma is seen in older patients with indolent disease[44,45]. In contrast to *MYCN*-

197   A and *TERT*-SV patients who were stage-4 at diagnosis, 21% of *ATRX*-mutant cases were diagnosed

198   with low-stage disease but eventually relapsed. Subclonal acquisition of *ATRX* events were common

199   (9/29 events) and seen in relapses (Extended Data Fig. 4). In two patients, parallel acquisition of

200   *ATRX* mutations were seen at distinct metastatic sites (H135089) or locoregional relapses at

201   consecutive time points (H134817) (Fig. 3c). *ATRX*-mutant cases were also enriched for SVs affecting

202   *PTPRD* (Extended Data Fig. 1e) with evidence of parallel or continuous evolution (Fig. 3c and

203   Extended Data Fig. 4).

204

205   *MDM2-CDK*4 co-amplifications were seen in patients diagnosed with low-stage disease

206   (Supplementary Table 1). Notably, the co-amplification was not mutually exclusive with *ATRX* events

207   and *TERT*-SVs suggesting an overlapping subtype (Fig. 2d-e).   For example, phylogenetic

208 reconstruction for patient H132384 mapped truncal *MDM2-CDK4* co-amplification and subclonal SVs

209 at *ATRX* and *TERT*. However, the *ATRX* and *TERT* events were acquired on two distinct subclonal

210 lineages further validating the mutually exclusivity of these events (Fig. 3b). In patients with *MDM2-*

211 *CDK4*, subclonal *ATRX* events and continuous evolution at *MDM2-CDK4* locus via incorporation and

212 over-expression of other genomic loci (*TERT*, *WNT3A*, *IGF2BP3*) were seen during disease

213 progression (Extended Data Fig. 4), which might contribute to the dismal outcomes associated with

214 *MDM2-CDK4*[18].

215

216 *TP53* mutations demarcate an ultra HR-NB subtype[46,47]. Excluding arm-level CNAs at 17p (n=19, 7%),

217 10 patients had mutations (n=10) or SVs (n=4) at the *TP53* locus specifically. In 7/10 patients these

218 mutations were bi-allelic, most frequently by initial 17p loss followed by a TP53 mutation (n=5).

219 Parallel acquisition of mutations affecting p53 pathway were observed both within the primary site

220 (H116987) and in different metastatic tumors (H134722) (Fig. 3d).

221

222 Taken together this analysis demonstrate that in neuroblastoma subtype-specific evolutionary

223 trajectories are predominantly determined by SVs targeting the main driver gene itself with specific

224 acquisition of secondary hits (e.g. *TERT*p in *MYCN*-A and *PTPRD* in ATRX-mutant) and are followed

225 by subclonal mutations in RAS-MAPK, PI3K-mTOR and p53 pathways shared across the disease

226 subtypes.

227

228 **Clonal diversification at primary site creates multiple clones with metastatic potential**

229 Patients with HR-NB are diagnosed with widely metastatic disease (bone, bone marrow, liver, lung

230 and CNS)[48]. We studied the clonal relationships amongst primary and disseminated disease in 30

231 evaluable patients (Methods, Supplementary Tables 3-5).

232

233   Whilst all resections were clonally related, subclonal heterogeneity at the primary site was seen in

234   the majority of patients in the form of subclonal CNAs and oncogenic mutations/SVs (83%, 25/30)

235   (CCF median=100%, range=4-100%) (Extended Data Fig. 4-5). This subclonal diversification in the

236   primary site creates distinct cell subpopulations with differential capacity to spread. Analysis of the

237   ensuing metastatic trajectories demonstrates that distinct primary-metastasis pairs share closer

238   lineage relationships in the tumor phylogeny than the primary sites to one another (7/9 patients,

239   Extended Data Fig. 4-5). For example, in patient H103207 (Fig. 3a), clonal structure across two

240   adrenal tumors at diagnosis and 6 metastases from CNS, lungs and liver suggests that CNS-metastatic

241   clone separated from the trunk before the adrenal primary site diversified further. Similarly, one of

242   the two adrenal tumors segregated with the lineage leading to liver and lung metastases. This

243   demonstrates that branching evolution in the primary site gives rise to multiple subclones with the

244   potential to spread. This observation held true across disease subtypes of this cohort (Extended Data

245   Fig. 4-5 and Supplementary Table 3).

246

247   **Timing of metastasis with respect to therapy**

248   Detection of therapy-related mutation signatures indicates the presence of cells that survived

249   therapy and subsequently achieved a clonal representation detectable at WGS depth. Therefore,

250   mutational signatures can be used to time emergence of clones relative to the time of therapy[50]. We

251   illustrate this point with patient H118706 (5-yo, *MYCN*-A, stage-4) for whom WGS was performed on

252   two diagnostic tumors (adrenal and liver metastasis) and 15 metastatic sites including liver and lungs

253   at autopsy following unsuccessful treatment with platinum and temozolomide (Fig. 4a). Subclonal

254   structure suggests that all the autopsy tumors came from the same *TP53*-mutant clone with strong

255   exposure to platinum signatures (48% of the substitutions) and this clone was succeeded by six

256   clones with evidence of temozolomide signature (5-19%). This suggests that the *TP53*-mutant clone

257   emerged after platinum therapy and seeded all the metastatic tumors from autopsy when

258   temozolomide therapy started.

259

260   Using the same logic, in 13 evaluable patients we studied the relative timing of the emergence of 21

261   metastasizing subclones and 26 daughter subclones (Supplementary Tables 3-4, Extended Data Fig.

262   6 and Supplementary Information). Majority of the metastasizing subclones (18/21) had no

263   evidence of therapy exposure, suggesting that disease dissemination happens before therapy

264   consistent with widely metastatic presentation at diagnosis (Extended Data Fig. 6). Importantly, this

265   observation held true not just for local spread but across distant sites including CNS metastases (¾

266   patients), which are typically not detected by imaging at diagnosis[51]. Notably, exposure to therapy-

267   related mutational signatures was seen in most (22/26) daughter subclones that emerge from the

268   initial metastasizing subclone (CCF median=80%, range=15-100%) in line with on-therapy disease

269   progression at the metastatic sites.

270

271   **Origin of late relapses traced back to early clones followed by a long period of dormancy**

272   With improved treatment and prolonged survival, consecutive relapses are increasingly observed

273   amongst HR-NB patients. For 72 patients in our cohort, we were able to study 114 clonal transitions

274   from primary to first relapse (47 patients) and/or between consecutive relapses (43 patients). We

275   observed three distinct patterns of temporal transitions (Supplementary Tables 3 and 6 and Fig. 4b-

276   c). Majority of the relapses were accompanied by accumulation of additional CNAs or mutations/SVs

277   at recurrent loci (72%) (linear and branched in Fig. 4b-c). In 24%, relapses were seeded by exactly

278   the same clone without evidence of new genetic changes and in 4% by an earlier clone in the

279   phylogeny. All three patterns were equally common from diagnosis to first relapse, amongst

280   consecutive relapses as well as across disease subtypes (Fig. 4b). This suggests that while subclones

281   with new drivers continue to emerge and replace existing ones, biological themes are preserved. In

282   31/45 patients with clonal transitions happening as a switch between different branches, the same

283   pathway was exploited by the tumor consistent with pathway-specific dependencies (Fig. 4c).

284

285   For patients with three or more consecutive relapses, we were able to capture multiple waves of

286   clonal successions and transitions across years. For example, for H134819 (3.5-yo stage-I *MYCN*-NA)

287   the same relapsing clone with a *SPRY2* deletion[52] and a high-level amplification of *IGF2BP3* via *MDM2*-

288   *CDK4* SVs was present across all five locoregional recurrences from 2.7 to 3.4 years after diagnosis,

289   with late clonal switches from an *ALK* to a *PIK3R1*-mutant subclone and finally to a subclone with

290   *ATRX* and *CDKN1C* events (Fig. 4d). By contrast, in the case of H134821 (9-yo stage-I *MYCN*-NA), the

291   first two locoregional relapses within 4-6 months from diagnosis were caused by a *PIK3CA*-mutant

292   subclone while the last two relapses 8 years later hailed from an *MTOR*-mutant clone (Fig. 4e). Both

293   of these relapsing clones were present at diagnosis suggesting that subclones can stay dormant for

294   many years, remain clinically undetectable but nonetheless maintain the potential to instigate

295   relapses at much later time points. These findings suggest that chemo-resistant clones with specific

296   driver mutations may already exist at diagnosis and, in rare cases, can lay dormant for years before

297   gaining clonal dominance as patients go through multiple lines of therapy.

298

299   **Shared lineage relationship across locoregional and distant metastases**

300   We next studied the relationships between locoregional disease and distant metastasis by evaluating

301   data from 19 patients with 69 tumors from disseminated sites (Methods, Supplementary Table 3).

302   We observe an equal number of relapsing subclones involved in locoregional extension or distant

303   metastasis (average 1.5 vs 1.3 subclones) suggesting that there is no differential propensity to extend

304   locally or metastasize. In support of this notion, the same clone from the primary site seeded both

305   locoregionally and at metastatic sites in the four patients with available tumors (30 samples)

306   (Extended Data Fig. 4-5).

307

Polyclonal and metastasis-to-metastasis seeding after therapy

309 Despite high intensity multimodal therapy, 50% of HR-NB patients will eventually progress, typically

310 in metastatic sites. To compare the clonal representation across distant metastatic sites, we studied

311 10 patients with two or more metastases (Supplementary Table 4). In most cases the same subclone

312 seeded all metastatic tumors (5/7 patients) (Fig. 4a, Fig. 5, Extended Data Fig. 4-5 and Supplementary

313 Information). This occurred in the form of polyclonal seeding[53] during or after therapy, as evidenced

314 by the presence of therapy related mutation signatures in the metastatic clones (e.g H103207,

315 H118706 and H134722) (Supplementary Information).

316

317 We also see unequivocal evidence of metastasis-to-metastasis (met-to-met) seeding. Patient

318 H118706 was diagnosed with metastatic disease in both the lungs and liver. Interestingly, analysis

319 from 16 tumors from the same metastatic sites 16 months later at autopsy demonstrated that all sites

320 were seeded by the same therapy-exposed subclone originating from the liver metastasis (Fig. 4a).

321 This suggests that the metastatic clones in lungs and liver at diagnosis were cleared by chemotherapy

322 yet a therapy-resistant subclone escaped from the liver and successfully re-seeded multiple

323 metastatic deposits across the same anatomical sites.

324

325 Notably the subclones shared across metastases were often characterized by oncogenic events such

326 as the hyper-rearranged *MYCN*-A, convergent *TP53* mutations and *ALK* mutations. This suggests that

327 certain therapy-resistant subclones with selective advantage[54] have the capacity to disseminate and

328 successfully establish clinically detectable metastatic lesions that transverse locoregional sites, lung,

329 and liver and CNS.

330

331

Discussion

Our understanding of the temporal and spatial relationships underlying evolution of neuroblastoma from disease initiation through treatment, progression and metastasis has remained obscure. Leveraging surgical samples from a unique cohort of 283 patients, we performed a comprehensive analysis of genomic evolution throughout the natural course of the disease.

Clonal reconstruction using multi-sample WGS data traced the first malignant clone and the ensuing subclonal trajectories with notable differences amongst different molecular subtypes. We estimate that MYCN-A emerges during embryogenesis while *ATRX*-mutant disease is frequently post-natal. Unlike *MYCN,* which is always targeted on the trunk, *TERT* and *ATRX* events are not always found in the first malignant clone suggesting that they may not be required for disease initiation but yet, as secondary events, define distinct and non-overlapping clinical subtypes with unique evolutionary trajectories. Once established, tumors progress across time and space via a process underwritten by continuous or parallel acquisition of SVs at disease-defining loci namely *MYCN*, *TERT*, *ATRX* and *MDM2-CDK4*. Whilst it is likely that high-level amplifications create genomic instability thereby providing fuel for continued rearrangements, subclonal events at these loci might provide survival advantage under chemotherapeutic pressure, only to become stable at their most complex forms shared across multiple metastatic sites as evidenced by the dominance of clones with the same complex but stable *MYCN* amplicon in late stage metastatic disease. Although this is in agreement with the emerging importance of evolution through extrachromosomal DNA observed in neuroblastoma and other tumor types[56,57], study of larger sample sets is warranted to further validate these observations.

Analysis of temporally and spatially separated tumors revealed unique patterns and timing of disease spread. At diagnosis, primary disease is already carrying divergent subclones with parallel drivers

357   involved in multiple waves of metastatic spread early in disease evolution. This is in agreement with

358   the idea that clonal heterogeneity is established early in neuroblastoma pathogenesis. As the disease

359   progresses, patients go through multiple lines of therapy, with each consecutive relapse

360   underwritten by the emergence of subclones with new driver mutations affecting the same set of

361   pathways. Furthermore, we show that most metastases occur prior to the commencement of therapy,

362   consistent with the clinical presentation of HR-NB. This is particularly true for CNS metastases which

363   have a low incidence[51] and are rarely detected clinically at diagnosis[51]. This suggests that CNS

364   involvement becomes detectable only after systemic disease was debulked and controlled by drugs

365   that did not cross the blood brain barrier.

366

367   Curative treatment for neuroblastoma is aimed at killing invisible metastases. Indeed we

368   demonstrated that disseminated disease could remain dormant for a long period (up to 10 years)

369   after successful treatment, whilst the patient is clinically "disease-free", but then eventually relapse

370   and spread to new sites. Despite the observed heterogeneity at the primary site, we show that late

371   metastatic spread after therapy is underwritten by a set of subclones that can spread across

372   locoregional and distant metastatic sites via met-to-met and polyclonal seeding. This suggests that

373   therapy selects for particularly resistant and aggressive subclones with superior metastatic potential.

374

375   Taken together our data built a dismal picture of neuroblastoma pathogenesis, where malignant

376   clones arise early in embryogenesis, yet rapidly diversify and spread across local and distant

377   metastatic sites prior to disease diagnosis and unravel the complex networks of disease spread

378   during relapse in response to therapy. This dynamic and rapidly evolving disease presentation has

379   important implications for inclusion of select targeted agents in upfront therapy for HR-NB patients

380   in order to improve the chance of cure[58,59].

381

Online methods

Patient cohort

Patients in this cohort were seen in the Neuroblastoma clinic at the Memorial Sloan Kettering Cancer Center (MSKCC) from 1987-2021. A written informed consent for tumor/normal sequencing and use of clinical data was taken for all patients in accordance with the ethical rules and regulations of the institutional review board at MSKCC. Additionally three patients' guardians consented to participate in the MSKCC medical donation program.

Chemotherapy regimen

Detailed treatment information was available for 56 patients for whom 256 tumors were whole-genome sequenced. Of these, 29 cases received induction chemotherapy as per an MSK protocol (N5, N7, N8 or N9) in which the first five cycles consist of cyclophosphamide/doxorubicin/vincristine (CAV) x 2, cisplatin/etoposide (PVP), CAV, PVP and CAV. 14 patients were treated according to a COG regimen similar to ANBL0532 in which the first five cycles include cyclophosphamide/topotecan x 2, PVP, CAV and PVP while 4 patients were treated according to COG3973 (CAV x 2, PVP, CAV and PVP). The rest of the 8 patients were treated with other protocols containing 3-4 rounds of platinum-based chemotherapy including 1 patient with rapid COJEC. Therapy resection (t-resection) tumors were taken during induction chemotherapy after (3-5) cycles of chemotherapy. Of the 30 patients with WGS data from a t-resection tumor, 21 tumors were exposed to only 1 round of platinum-based chemotherapy while 8 and 1 were exposed to 2 and 3 rounds, respectively.

Whole genome and transcriptome sequencing data

WGS data for this study cohort came from three different sequencing centers. 1) 45 tumors and matched normals were sequenced at St Judes Children Hospital at a median coverage of 34X for both tumor and normal (range 30-59X)[44]. Publicly available raw data for tumors and normals were

16

407 downloaded. 2) For 29 tumors and matched normals sequencing library preparation and WGS were

408 performed at the New York Genome Center as described before[60]. Tumors and matched normals

409 were sequenced to a median of 95x (range 73-300X) and 44x (29-88X), respectively. 3) For 173

410 tumors and matched normals WGS was performed at a median coverage of 83X (40-181X) and 46X

411 (range 36-89X) at MSKCC as described in Supplementary Information. Within the subset of patients

412 where clonal structure was analyzed from multi-WGS data, genome-wide coverage figures were 50-

413 100X for 85% and >100X for 13% of the tumors. Only 6 tumors (all diagnostic) from 6 different

414 patients had <50X coverage. To supplant for lower coverage in the diagnostic tumor at least one

415 additional diagnostic tumor was sequenced whenever available (5/6 patients). RNA sequencing was

416 performed in-house as described in Supplementary Information to achieve a median of 81.5 million

417 paired reads per sample.

418

419 **Targeted gene panel sequencing**

420 DNA extracted from formalin fixed paraffin embedded (FFPE) tumor and blood samples (as a

421 matched normal) were sequenced using MSK-IMPACT, an FDA-approved and New York State

422 Department of Health validated panel used to sequence patients' tumors at MSKCC. MSK-IMPACT

423 captures protein-coding exons of 468 cancer-associated genes, introns of frequently rearranged

424 genes and genome-wide copy number probes[61]. Tumor samples were sequenced at a median depth

425 of 648X, whereas peripheral blood samples at 400x. Established pipelines followed by manual review

426 were used to characterize germline and acquired somatic mutations, copy number variants (CNVs)

427 and if targeted, genomic rearrangements as previously described[61]. Clinically relevant findings were

428 annotated using OncoKb tiers 1-4[62].

429

430 **Bioinformatic analysis of WGS and RNA-seq**

17

431 Analysis of WGS and RNAseq data was executed using Isabl platform[63] and included: 1. Data quality

432 control; 2. Ensemble variant calling for germline and somatically acquired mutations from at least

433 two out of three algorithms run for each variant class and 3. Variant classification. Briefly, upon

434 completion of each sequencing run, Isabl imports paired tumor-normal FASTQ files, executes

435 alignment, quality control algorithms and generates tumor purity and ploidy estimates. For tumor

436 samples ensembl variant calling for each variant class (substitutions, insertions and deletions and

437 structural variations) was performed. High confidence somatic mutations are classified with regards

438 to their putative role in cancer pathogenesis and statistical post-processing enables the derivation of

439 microsatellite instability scores and mutation signatures[20]. RNA-seq data were independently

440 analyzed for acquired fusions and gene expression metrics.

441 Clinical relevance of mutations in common cancer genes was annotated using OncoKb, COSMIC,

442 Ensembl Variant Effect Predictor, VAGrENT, gnomAD and ClinVar databases [62,64–67].

443 Details of the variant calling and annotation can be found in the Supplementary Information.

444

445 **Gene expression analysis**

446 Gene expression for 55,390 coding and non-coding genes were ascertained in Transcripts Per Million

447 (TPM) using SALMON (v0.10.0, https://github.com/COMBINE-lab/salmon)[68]. Genes with total

448 expression less than or equal to median across all genes were filtered. Deconvolution of the RNA-seq

449 data to predict the proportion of immune and stromal cells in the tumor microenvironment was done

450 using xCell[72].

451

452 **Identification of mutation signatures for substitutions and indels**

453 De novo mutational signature analysis was performed using sigProfilerExtractor[73] with default

454 parameters. The signatures identified were compared to the COSMIC Mutational Signatures (v3.2)

455   with the addition of temozolomide signature from Kucab et al[21,74] using cosine similarity. Details of

456   mutational signature analysis can be found in the Supplementary Information.

457

458   **Identification of simple and complex rearrangement events**

459   All structural variants called in a tumor were clustered into simple events (deletion, tandem

460   duplication, unbalanced translocation, balanced translocation, reciprocal inversion) or clustered

461   events (complex with >=2 SVs) using ClusterSV [https://github.com/cancerit/ClusterSV]. The

462   algorithm groups the SVs into clusters based on the proximity of breakpoints, the number of events

463   in the region and the size distribution of those events. The resulting clusters contain SVs that are

464   significantly closer than expected given the orientation and the number of SVs in that tumor and

465   hence are expected to have happened as part of the same event. The resulting clusters were then

466   heuristically refined as described previously[75]. Independently, SV breakpoints and CN data from

467   Battenberg and/or Brass were analyzed using ShatterSeek[76] to identify chromothripsis events.

468

469   **Inference of clonal structure from WGS data**

470   For 45 patients with 170 tumors for whom two or more WGS tumors were available, clonal structure

471   was determined using genome-wide substitutions, indels, SVs and CNAs separately. Within this

472   subset of WGS data, genome-wide coverage figures were 50-100X for 85% and >100X for 13%. Only

473   6 diagnostic tumors from 6 different patients had <50X coverage. For 5/6 of these patients at least 1

474   other tumor from diagnosis was sequenced to supplant for lower coverage (Supplementary

475   Information). Phylogenies predicted from substitutions and CNAs are displayed in the

476   Supplementary Figures 2-46.

477

478   For substitutions, union of high confidence mutations called across all tumors of the patient was used

479   as input to DPClust (v0.2.2, https://github.com/Wedge-Oxford/dpclust). Mutations were filtered if

480   they 1) had depth greater than 6 standard deviations above median coverage, 2) had no CN

481   information or 3) were in genomic regions that were affected by deletion or copy-neutral LOH in a

482   subset of the tumors of the patient. DPClust algorithm 1) calculates cancer cell fraction (CCF)

483   corrected for purity and local copy number 2) performs clustering across tumors to identify the CCF

484   position for the underlying clusters and 3) assigns mutations to each cluster[77,53]. After heuristic and

485   manual curation of the clusters whenever needed, clusters that predominantly contain mutations

486   located on the chromosome were filtered. Clonal ordering of high-confidence clusters was

487   determined using clonevol (v0.99.11, https://github.com/hdng/clonevol)[78]. When there are multiple

488   possible tumor phylogenies, clones with uncertainty were indicated with a star in the phylogenetic

489   tree. Mutational signatures were computed in each cluster independently. Signature trees were

490   generated with python matplotlib (v3.1.0, https://matplotlib.org/). All steps were run using an

491   inhouse wrapper.

492   For indels the same filtering criteria as substitutions was used with an additional filtering step prior

493   to clustering. Only indels across loci with tumor depth >=40x and <=200x and with a VAF of >1%

494   were used.

495   In order to compare the CNA segments in a patient, first aberrant segments are matched to SV

496   breakpoints using the script called "match_rg_patterns_to_library.pl" from the Brass pipeline

497   [https://github.com/cancerit/BRASS/blob/dev/perl/bin/match_rg_patterns_to_library.pl].   The

498   presence of associated SVs across all the tumors (see the next section) is used to determine if CNA

499   breakpoints are shared or tumor-specific. Finally the results are manually cross-checked by

500   comparing the allele-specific and subclonal CN states for the segments as estimated by Battenberg.

501   In addition to this, a separate analysis was performed to construct phylogenies based on CNA

502   segments   only   using   MEDICC2   with   default   parameters

503   [https://bitbucket.org/schwarzlab/medicc2/src/master/].

504

505 **Determining clonal status of SVs in a patient**

506 For all the SVs identified across the tumors of a patient, a pileup procedure was performed to

507 determine the number of aberrant reads supporting the variant in each tumor as described in

508 Supplementary Information. An SV was deemed 'present' in tumors with >=2 reads supporting the

509 associated breakpoints. SV clusters were defined as groups of SVs present in the same set of tumors.

510

511 **Comparison of substitutions, indel and SV clusters**

512 Substitution and indel clusters were compared by calculating the cosine similarity of the CCF values

513 across all tumors of the patient. Clusters with a cosine similarity >0.9 were matched. When a

514 substitution cluster matched multiple indel clusters, the pair with the smallest summed CCF

515 difference across the tumors was retained. Substitution and SV clusters were compared by their

516 presence across the different tumors of the patient. That is, an SNV cluster present in samples A, B

517 and C is matched with an SV cluster present in the same set of samples.

518

519 **Inference of clonal structure from targeted sequencing data**

520 For a subset of 49 patients with 113 tumors that were sequenced with MSK-IMPACT and/or WGS,

521 analysis of clonal structure was confined to the alterations that can be captured by MSK-IMPACT.

522 This includes the substitutions and indels called within the exonic and extended splice site regions

523 of ~450 cancer genes, focal deletions in genes such as *CDKN2A, PTPRD, ATRX* and *TP53*, structural

524 variants in select introns included in targeted sequencing panel in genes such as *ALK, ATRX* and *TP53*

525 as well as arm and chromosome level CNAs. For 29 patients only targeted sequencing data were

526 available while 20 patients had one WGS tumor and at least one tumor sequenced by targeted

527 sequencing. Additional alterations detected by WGS but cannot be captured by MSK-IMPACT (i.e. SVs

528 at loci such as *TERT* and *FOXR1*) are not included in this analysis (indicated with

529 SUBCLONE_TYPE==NA in Supplementary Table 2).

21

530

531 For this patient subset, clonal structure was first analyzed with the allele-specific CNA data as

532 assessed using FACETS[79] in targeted data and Battenberg in WGS data. CNA-based phylogenies were

533 derived by comparing the genomewide CNAs of the tumors as well as using MEDICC2 with default

534 parameters. Substitutions and indels were analyzed together using DriverClone

535 [https://github.com/papaemmelab/driverclone], an inhouse algorithm designed for studying clonal

536 structure specifically from sparse targeted sequencing data. Briefly, DriverClone first derives a

537 posterior probability for the CCF of each variant, taking into account information on local ploidy,

538 coverage, tumor purity and possible genotypes. DriverClone then clusters the variants in CCF space

539 using a weighted variant graph where edges represent overlaps of posterior credible intervals

540 between variant pairs in each sample. Low weight edges are pruned. A depth-first search then finds

541 all connected components in the variant graph and retrieves clusters of variants belonging to the

542 same predicted clone. To enable probabilistic clonal ordering with few observations per clone,

543 DriverClone extends the non-parametric bootstrapping model of ClonEvol so that bootstrap samples

544 are obtained from a mixture distribution of variant posteriors. A tree enumeration algorithm

545 (originally implemented in Clonevol) then identifies all possible tumor phylogenies that fulfill the

546 appropriate biological constraints. Phylogenies predicted from this analysis are displayed in the

547 Supplementary Figs. 48-53.

548

549 **Timing the emergence of MRCA**

550 Timing of the MRCA emergence was performed in a subset of 39 patients with >=2 WGS tumors

551 where at least one tumor was from a pre-treatment diagnostic specimen or therapy resection.

552 Association between age at diagnosis and trunk length was assessed by a linear regression model

553 using R *lm* function and in a multivariate analysis using R *glm* function taking into account disease

554 subtype, stage at diagnosis and number of WGS tumors. MRCA analysis was done using a previously

555   published analysis workflow[37] in two steps 1) First patient-specific mutation rates were estimated

556   via linear mixed effect modeling with the number of mutations attributed to the clock-like mutational

557   signature SBS40. 2) Patient-specific mutation rates and the number of SBS40 mutations on the trunk

558   were used to estimate the time of emergence for MRCA applying a bootstrapping approach to

559   estimate 95% confidence intervals (CIs). MRCA was classified as 'pre-natal' if CIs overlapped the time

560   of birth and 'post-natal' otherwise.

561

562   **Analysis of truncal and subclonal somatic changes**

563   This analysis was performed in a subset of 94 patients with two or more tumors for which WGS

564   and/or targeted sequencing data was used to assess genome-wide segmental CNAs and oncogenic

565   substitution, indels, CNAs and SVs in reported ~450 cancer genes included in targeted sequencing

566   panel. Detailed analysis of *MYCN* and *TERT* loci was performed in 13 and 7 patients, respectively,

567   with two or more WGS samples available.

568

569   **Analysis of evolutionary patterns**

570   Analysis of divergence in the primary site was performed in a subset of 30 patients with two or more

571   tumors from the primary site available. Divergence is defined as acquisition of recurrent CNAs as well

572   as oncogenic mutations provided in Supplementary Table 2. Comparison of primary site to

573   disseminated disease was performed in a subset of 9 patients with two or more tumors from the

574   primary site and tumor(s) from disseminated sites. Timing of metastasis with respect to therapy was

575   performed in a subset of 13 patients for whom at least one tumor from the primary site and two or

576   more tumors from local-regional and/or distant metastatic sites were sequenced by WGS. Lineage

577   relationship between local-regional and/or distant metastatic tumors was studied in a subset of 19

578   patients with at least one tumor from the primary site as well as local-regional and/or distant

579   metastatic tumors available.

580

Author contributions

582

E.P., N.K.C and G.G. designed the study. G.G., M.F.L., J.S.M.M, J.E.A.O and J.Z. developed algorithmic

infrastructure and G.G. performed bioinformatic analysis with support from L.C., M.R. and G.A.. N.K.C,

S.S.R., B.S., M.P.L, B.H.K, S.M. and N.S. performed the clinical management of the patients. N.K.C and

N.B. performed patient consent. N.K.C oversaw biospecimen banking performed by I.Y.C and Y.F.

while D.Y. and F.D.C executed laboratory processing of PDX specimens. N.K.C. collected clinical data

for the patients. C.A.I.O. led the clinical donation program. G.G. prepared figures and tables. G.G., N.K.C.

and E.P. reviewed analysis results and interpretation of findings and wrote the manuscript with input

from D.B.S., B.H.K., S.M., C.A.I.D and A.K.. All authors reviewed and approved the manuscript for

submission.

ACKNOWLEDGEMENTS

603

CONFLICT OF INTEREST

612

613   **Data availability**

614   All data is available in dbGAP (in submission) and cbioPortal.

615   **Code availability**

616   Additional scripts and data used for generating the figures are available at

617   https://github.com/gg10/gg10-Clonal-evolution-during-metastatic-spread-in-high-risk-

618   neuroblastoma.

619

620   **Bibliography**

621   1.   Maris, J. M. Recent Advances in Neuroblastoma. *New England Journal of Medicine* vol. 362

622   2202–2211 (2010).

623   2.   London, W. B. *et al.* Historical time to disease progression and progression-free survival in

624   patients with recurrent/refractory neuroblastoma treated in the modern era on Children's

625   Oncology Group early-phase trials. *Cancer* vol. 123 4914–4923 (2017).

626   3.   Abbasi, M. R. *et al.* Impact of Disseminated Neuroblastoma Cells on the Identification of the

627   Relapse-Seeding Clone. *Clin. Cancer Res.* **23**, 4224–4232 (2017).

628   4. Eleveld, T. F. *et al.* Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations.

629   *Nat. Genet.* **47**, 864–871 (2015).

630   5. Schramm, A. *et al.* Mutational dynamics between primary and relapse neuroblastomas. *Nat.*

631   *Genet.* **47**, 872–877 (2015).

632   6. Chicard, M. *et al.* Genomic Copy Number Profiling Using Circulating Free Tumor DNA Highlights

633   Heterogeneity in Neuroblastoma. *Clin. Cancer Res.* **22**, 5564–5573 (2016).

634   7. Van Roy, N. *et al.* Shallow Whole Genome Sequencing on Circulating Cell-Free DNA Allows

635   Reliable Noninvasive Copy-Number Profiling in Neuroblastoma Patients. *Clin. Cancer Res.* **23**,

636   6305–6314 (2017).

637   8. Chicard, M. *et al.* Whole-Exome Sequencing of Cell-Free DNA Reveals Temporo-spatial

638   Heterogeneity and Identifies Treatment-Resistant Clones in Neuroblastoma. *Clin. Cancer Res.*

639   **24**, 939–949 (2018).

640   9. Fransson, S. *et al.* Whole-genome sequencing of recurrent neuroblastoma reveals somatic

641   mutations that affect key players in cancer progression and telomere maintenance. *Sci. Rep.* **10**,

642   22432 (2020).

643   10. Karlsson, J. *et al.* Four evolutionary trajectories underlie genetic intratumoral variation in

644   childhood cancer. *Nat. Genet.* **50**, 944–950 (2018).

645   11. Andersson, N. *et al.* Extensive Clonal Branching Shapes the Evolutionary History of High-Risk

646   Pediatric Cancers. *Cancer Res.* **80**, 1512–1523 (2020).

647   12. Schmelz, K. *et al.* Spatial and temporal intratumour heterogeneity has potential consequences

648   for single biopsy-based neuroblastoma treatment decisions. *Nat. Commun.* **12**, 6804 (2021).

649   13. Peifer, M. *et al.* Telomerase activation by genomic rearrangements in high-risk neuroblastoma.

650   *Nature* **526**, 700–704 (2015).

651   14. Brady, S. W. *et al.* Pan-neuroblastoma analysis reveals age- and signature-associated driver

652   alterations. *Nat. Commun.* **11**, 5183 (2020).

653   15. Monclair, T. *et al.* The International Neuroblastoma Risk Group (INRG) staging system: an INRG
654   Task Force report. *J. Clin. Oncol.* **27**, 298–303 (2009).

655   16. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284
656   (2013).

657   17. Valentijn, L. J. *et al.* TERT rearrangements are frequent in neuroblastoma and identify
658   aggressive tumors. *Nat. Genet.* **47**, 1411–1414 (2015).

659   18. Amoroso, L. *et al.* Genomic coamplification of CDK4/MDM2/FRS2 is associated with very poor
660   prognosis and atypical clinical features in neuroblastoma patients. *Genes Chromosomes Cancer*
661   **59**, 277–285 (2020).

662   19. Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**,
663   94–101 (2020).

664   20. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–
665   421 (2013).

666   21. Kucab, J. E. *et al.* A Compendium of Mutational Signatures of Environmental Agents. *Cell* **177**,
667   821–836.e16 (2019).

668   22. Wang, T. *et al.* MYCN drives glutaminolysis in neuroblastoma and confers sensitivity to an ROS
669   augmenting agent. *Cell Death & Disease* vol. 9 (2018).

670   23. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature*
671   **555**, 321–327 (2018).

672   24. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and
673   solid tumours. *Nature* **555**, 371–376 (2018).

674   25. Wei, J. S. *et al.* Clinically Relevant Cytotoxic Immune Cell Signatures and Clonal Expansion of T-
675   Cell Receptors in High-Risk MYCN-Not-Amplified Human Neuroblastoma. *Clinical Cancer*
676   *Research* vol. 24 5673–5684 (2018).

677   26. Layer, J. P. *et al.* Amplification of N-Myc is associated with a T-cell-poor microenvironment in

678   metastatic neuroblastoma restraining interferon pathway activity and chemokine expression.

679   *Oncoimmunology* **6**, e1320626 (2017).

680   27. Geoerger, B. *et al.* Pembrolizumab in paediatric patients with advanced melanoma or a PD-L1-

681   positive, advanced, relapsed, or refractory solid tumour or lymphoma (KEYNOTE-051): interim

682   analysis of an open-label, single-arm, phase 1–2 trial. *The Lancet Oncology* vol. 21 121–133

683   (2020).

684   28. Geoerger, B. *et al.* Atezolizumab for children and young adults with previously treated solid

685   tumours, non-Hodgkin lymphoma, and Hodgkin lymphoma (iMATRIX): a multicentre phase 1–

686   2 study. *The Lancet Oncology* vol. 21 134–144 (2020).

687   29. Davis, K. L. *et al.* Nivolumab in children and young adults with relapsed or refractory solid

688   tumours or lymphoma (ADVL1412): a multicentre, open-label, single-arm, phase 1-2 trial.

689   *Lancet Oncol.* **21**, 541–550 (2020).

690   30. Havel, J. J., Chowell, D. & Chan, T. A. The evolving landscape of biomarkers for checkpoint

691   inhibitor immunotherapy. *Nat. Rev. Cancer* **19**, 133–150 (2019).

692   31. Pich, O. *et al.* The mutational footprints of cancer therapies. *Nat. Genet.* **51**, 1732–1740 (2019).

693   32. Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in

694   mutation and signature frequencies. *Nat. Genet.* **51**, 1450–1458 (2019).

695   33. Kocakavuk, E. *et al.* Radiotherapy is associated with a deletion signature that contributes to

696   poor outcomes in patients with cancer. *Nat. Genet.* **53**, 1088–1096 (2021).

697   34. Behjati, S. *et al.* Mutational signatures of ionizing radiation in second malignancies. *Nat.*

698   *Commun.* **7**, 12605 (2016).

699   35. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).

700   36. Mitchell, T. J. *et al.* Timing the Landmark Events in the Evolution of Clear Cell Renal Cell Cancer:

701   TRACERx Renal. *Cell* **173**, 611–623.e17 (2018).

702   37. Rustad, E. H. *et al.* Timing the initiation of multiple myeloma. *Nat. Commun.* **11**, 1917 (2020).

703   38.   Coorens, T. H. H. *et al.* Inherent mosaicism and extensive mutation of human placentas. *Nature*

704   **592**, 80–85 (2021).

705   39.   Schleiermacher, G. *et al.* Emergence of new ALK mutations at relapse of neuroblastoma. *J. Clin.*

706   *Oncol.* **32**, 2727–2734 (2014).

707   40.   Althoff, K. *et al.* A Cre-conditional MYCN-driven neuroblastoma mouse model as an improved

708   tool for preclinical studies. *Oncogene* **34**, 3357–3368 (2015).

709   41.   Schwab, M. *et al.* Amplified DNA with limited homology to myc cellular oncogene is shared by

710   human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* vol. 305 245–248

711   (1983).

712   42.   Kohl, N. E. *et al.* Transposition and amplification of oncogene-related sequences in human

713   neuroblastomas. *Cell* **35**, 359–367 (1983).

714   43.   Cobrinik, D. *et al.* Recurrent pre-existing and acquired DNA copy number alterations, including

715   focal TERT gains, in neuroblastoma central nervous system metastases. *Genes Chromosomes*

716   *Cancer* **52**, 1150–1166 (2013).

717   44.   Cheung, N.-K. V. *et al.* Association of age at diagnosis and genetic mutations in patients with

718   neuroblastoma. *JAMA* **307**, 1062–1071 (2012).

719   45.   Franks, L. M., Bollen, A., Seeger, R. C., Stram, D. O. & Matthay, K. K. Neuroblastoma in adults and

720   adolescents: an indolent course with poor survival. *Cancer* **79**, 2028–2035 (1997).

721   46.   Ackermann, S. *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma.

722   *Science* **362**, 1165–1170 (2018).

723   47.   Carr-Wilkinson, J. *et al.* High Frequency of p53/MDM2/p14ARF Pathway Abnormalities in

724   Relapsed Neuroblastoma. *Clin. Cancer Res.* **16**, 1108–1118 (2010).

725   48.   DuBois, S. G. *et al.* Metastatic sites in stage IV and IVS neuroblastoma correlate with age, tumor

726   biology, and survival. *J. Pediatr. Hematol. Oncol.* **21**, 181–189 (1999).

727   49.   Yates, L. R. & Campbell, P. J. Evolution of the cancer genome. *Nat. Rev. Genet.* **13**, 795–806

728   (2012).

729   50. Landau, H. J. *et al.* Accelerated single cell seeding in relapsed multiple myeloma. *Nat. Commun.*

730   **11**, 3617 (2020).

731   51. Berlanga, P. *et al.* Central nervous system relapse in high-risk stage 4 neuroblastoma: The HR-

732   NBL1/SIOPEN trial experience. *Eur. J. Cancer* **144**, 1–8 (2021).

733   52. Ishida, M. *et al.* Sprouty2 regulates growth and differentiation of human neuroblastoma cells

734   through RET tyrosine kinase. *Cancer Sci.* **98**, 815–821 (2007).

735   53. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**,

736   353–357 (2015).

737   54. Keshelava, N. *et al.* Loss of p53 function confers high-level multidrug resistance in

738   neuroblastoma cell lines. *Cancer Res.* **61**, 6185–6193 (2001).

739   55. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated

740   differentiation states. *Nat. Genet.* **49**, 1261–1266 (2017).

741   56. Koche, R. P. *et al.* Extrachromosomal circular DNA drives oncogenic genome remodeling in

742   neuroblastoma. *Nat. Genet.* **52**, 29–34 (2020).

743   57. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and poor

744   outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).

745   58. Kushner, B. H. *et al.* Efficacy of naxitamab in patients with refractory/relapse (R/R) high-risk

746   neuroblastoma (HR-NB) by bone/bone marrow (BM) evaluation, potential sites of residual

747   disease. *Journal of Clinical Oncology* vol. 39 10022–10022 (2021).

748   59. Yarmarkovich, M. *et al.* Cross-HLA targeting of intracellular oncoproteins with peptide-centric

749   CARs. *Nature* (2021) doi:10.1038/s41586-021-04061-6.

750   60. Diolaiti, D. *et al.* A recurrent novel MGA–NUTM1 fusion identifies a new subtype of high-grade

751   spindle cell sarcoma. *Molecular Case Studies* vol. 4 a003194 (2018).

752   61. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical

753   sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).

754   62.   Chakravarty, D. *et al.* OncoKB: Annotation of the oncogenic effect and treatment implications of

755   somatic mutations in cancer. *Journal of Clinical Oncology* vol. 34 11583–11583 (2016).

756   63.   Medina-Martínez, J. S. *et al.* Isabl Platform, a digital biobank for processing multimodal patient

757   data. *BMC Bioinformatics* vol. 21 (2020).

758   64.   Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants.

759   *Nucleic Acids Res.* **44**, D862–8 (2016).

760   65.   Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**,

761   D777–D783 (2017).

762   66.   Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in

763   141,456 humans. *Nature* **581**, 434–443 (2020).

764   67.   McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

765   68.   Srivastava, A. *et al.* Alignment and mapping methodology influence transcript abundance

766   estimation. *Genome Biol.* **21**, 239 (2020).

767   69.   John, C. R. *et al.* M3C: Monte Carlo reference-based consensus clustering. *Sci. Rep.* **10**, 1816

768   (2020).

769   70.   Korotkevich, G. *et al.* Fast gene set enrichment analysis. doi:10.1101/060012.

770   71.   Bagaev, A. *et al.* Conserved pan-cancer microenvironment subtypes predict response to

771   immunotherapy. *Cancer Cell* **39**, 845–865.e7 (2021).

772   72.   Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity

773   landscape. *Genome Biol.* **18**, 220 (2017).

774   73.   Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering

775   signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).

776   74.   Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive

777   genome-wide analysis of mutational processes. *Genome Med.* **10**, 33 (2018).

778   75. Li, Y. *et al.* Patterns of somatic structural variation in human cancer genomes. *Nature* **578**,

779   112–121 (2020).

780   76. Cortés-Ciriano, I. *et al.* Comprehensive analysis of chromothripsis in 2,658 human cancers

781   using whole-genome sequencing. *Nat. Genet.* **52**, 331–341 (2020).

782   77. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).

783   78. Dang, H. X. *et al.* ClonEvol: clonal ordering and visualization in cancer sequencing. *Ann. Oncol.*

784   **28**, 3076–3082 (2017).

785   79. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis

786   tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).

787   80. Dentro, S. C., Wedge, D. C. & Van Loo, P. Principles of Reconstructing the Subclonal Architecture

788   of Cancers. *Cold Spring Harb. Perspect. Med.* **7**, (2017).

789   **Figure legends**

790   **Figure 1. Patient cohort and genome-wide mutational landscape. a)** Barplot shows the number

791   of tumors sequenced from 94 patients with two or more samples color-coded by type of

792   sample and sequencing performed. Patients are shown as columns organized by the disease

793   subtype which are as follows: 1) MCYN, patients with *MYCN* amplification. 2) TERT, patients

794   with *TERT* SVs. 3) ATRX, patients with *ATRX* events. 4) MDM2-CDK4, patients with *MDM2-*

795   *CDK4* co-amplifications. 5) SEG-CNA and 6) CUM-CNA, patients with segmental or numeric

796   CNAs but without the aforementioned alterations. (SV, structural variants. CNA, copy number

797   aberration. WGS, whole genome sequencing.). Information about age and stage at diagnosis is

798   provided as tile plots at the bottom. Box plots show comparison of the proportion of mutations

799   attributed to SBS18 and SBS40 **(b)** and tumor mutation burden (TMB) **(c)** in WGS data across

800   diagnostic tumors of different molecular subtypes (n=72 tumors). **d)** Box plot on the left shows

801   the increase in TMB across samples collected at diagnosis, t-resection, relapse and further

802   relapse (n=129 patients) while on the right is the fold change in TMB in relapse and further

803   relapse samples compared to the matched diagnostic tumor of the same patient (n=22

804   patients). **e)** Box plots show the proportion of substitutions attributed to SBS31, SBS35 and

805   temozolomide (TMZ) signatures across samples collected at diagnosis, t-resection, relapse and

806   further relapse (n=129 patients). **f)** Boxplots show the number of substitutions attributed to

807   therapy-related mutational signatures for tumors from patients with stage-4 disease who were

808   exposed to increasing numbers of rounds of platinum or temozolomide-based chemotherapy

809   (n=145 tumors). The data and script for Fig. 1 are available in Supplementary Table 1 and the

810   GitHub repository.

811 **Figure 2. Timing of emergence of the first malignant clone. a)** Scatter plot shows the

812   relationship between the number of single nucleotide variants (SNVs) on the trunk (trunk

813   length) and age at diagnosis (Pearson correlation). **b)** Barplots show the breakdown of upper,

814   middle and lower tertiles of the trunk length distribution by stage at diagnosis. **c)** Timeline plot

815   shows the time of diagnosis and the predicted time of emergence of the most recent common

816   ancestor (MRCA) with 95% confidence intervals for n=39 patients. Number of SBS40-

817   associated SNVs on the trunk is shown next to the patient id. Shaded area from -9 to 0 on the

818   x-axis shows the period in utero. The predicted time of emergence of the MRCA is shown with a

819   circle (pre-natal) or a triangle (post-natal). **d)** Ternary plot shows the proportion of events that

820   are shared by all tumors of a patient, seen in a subclone specific to a sample from the primary

821   site or metastatic/relapse site. Dots are color-coded red or green to indicate a tendency to be

822   shared or metastasis/relapse-specific with a size proportional to the total number of events. **e)**

823   Heatmap shows the number of truncal and subclonal genetic changes identified in 94 patients

824   with >=2 tumors. Mutations and SVs are collapsed to the affected pathways except for those

825   hitting the most recurrent disease-defining genes (*MYCN*, *TERT* and *ATRX*). Black dots indicate

826   parallel evolution while crosses indicate the loci affected by continuous subclonal SVs when

33

827   there is already an SV event on the trunk of the patient. Lowermost tile plot shows the

828   availability of multi-WGS data, number of tumors and the timepoints studied for each patient.

829   Barplot on the left shows the frequency of the events per row. +, gain. -, loss. The data and

830   script for Fig. 2 are available in Supplementary Tables 1, 2 and 4 and the GitHub repository.

831   **Figure 3. Subtype-specific evolutionary trajectories. a)** Subclonal structure for patient H103207

832   is summarized in multiple panels. Treatment timeline gives a summary of the therapy

833   administered, the sequenced tumors and the survival status at last followup. Body map shows

834   the location of the tumor sites sequenced. Subclone tree shows the lineage relationships

835   amongst the subclones identified in a patient. Subclones are designated by branches with non-

836   informative lengths. Trunk is shown in gray at the top of the tree. Terminal nodes are

837   annotated with tumors where the corresponding clone is present . Branches are annotated

838   with putative oncogenic events. Different types of *MYCN* amplicons are indicated by a number

839   after gene name (i.e. *MYCN*.1). Continuous accumulation of SVs at *MYCN* loci is indicated by

840   stars (*). On the right, *MYCN* locus for each tumor is shown with an integrated copy number

841   (CN)/structural variant (SV) plot with absolute copy number on the y-axis and SVs as arcs

842   color-coded by the subclones they were assigned to. *ALK*$^{ECD-}$, *ALK* with a deletion of exons

843   encoding the extra-cellular domain. **b)** Treatment timeline, subclone tree and body map as

844   described in Fig-3a are shown for patient H132384. Evolution at *MDM2*, *CDK4* and *TERT* loci

845   are shown in the integrated CN/SV plot as described in Fig. 3a. Barplot shows the increase in

846   expression in the tumors with *TERT* SVs. **c)** Treatment timeline, subclone tree and body maps

847   as described in Fig-3a are shown for two different *ATRX*-mutant patients. **d)** Treatment

848   timeline, subclone tree and body maps as described in Fig-3a are shown for two patients with

849   *TP53* mutations. Detailed description of each patient is provided in Supplementary

850   Information. The data for Fig. 3 are available as raw data at dbGAP and scripts are available

851   through ISABL platform.

852 **Figure 4. Timing of metastasis. a)** Subclonal structure for patient H118706 is shown. Treatment

853 timeline is as described in Fig. 3a. On the right is the signature tree with the results from the

854 subclone-specific mutational signature analysis across the subclone tree of the patient. Each

855 subclone is shown as a stacked bar plot showing the proportion of the mutations attributed to

856 the six different mutational signatures and with total length proportional to the number of

857 substitutions in the corresponding subclone. Branches are separated by a dashed line and

858 annotated with the putative oncogenic changes assigned to the corresponding subclone.

859 Terminal nodes are annotated with tumors where the corresponding clone is present.

860 Triangles denote the subclones involved in metastatic spread where gray and white indicate

861 spread before and after therapy, respectively. The id of the metastatic subclones is annotated

862 next to the corresponding branch. Body maps show the possible movement of subclones

863 involved in spread before and after therapy indicated by the subclone ids next to the body

864 maps. Subclones are color-coded according to the clonal phylogeny shown in the legend.

865 Details of the clonal and genomic for H118706 is provided in Supplementary Information. **b)**

866 Left barplot shows the proportion of different types of clonal transitions from diagnosis to first

867 relapse (n=47) and between consecutive relapses (n=67) while the right barplot gives a

868 breakdown of clonal transitions across different disease subtypes. The different types of

869 transitions are 1) 'same' where the relapse is caused by the same clone as the previous time

870 point with no new genetic changes acquired. 2) 'earlier' where the relapse is caused by an

871 earlier clone in the phylogenetic tree. 3) 'linear' where the relapse is caused by a clone with

872 new CNAs and oncogenic mutations/SVs while no such events are seen in the clone specific to

873 the previous time point. 4) 'branched' where clones from both time points have genetic

874 changes. **c)** Barplot shows the number of different types of clonal transitions where the

875 relapsing clone has genetic changes affecting the listed CNAs and oncogenic mutations

876 collapsed to pathways affected except for those mutations affecting the most frequent disease-

35

877 defining genes (*MYCN*, *TERT* and *ATRX*). The inner plot shows the pathways affected by parallel

878 events across clonal transitions that switch between lineages. **d** and **e)** Subclonal structure for

879 patients H134819 and H134821 are shown. Body maps and treatment timelines are as

880 described in Fig. 3a. Signature trees are as described in Fig. 4a. For H134821 the left body map

881 shows the local-regional spread before therapy while the right body map shows the spread 9

882 years after diagnosis. Detailed description of each patient is provided in Supplementary

883 Information. The data for Fig. 4a, d and e are available as raw data at dbGAP and scripts are

884 available through ISABL platform while data and scripts for Fig. 4b-c are available in

885 Supplementary Table 6 and the GitHub repository.

886 **Figure 5. Complex seeding patterns after therapy.** Subclonal structures for patients H103207

887 **(a)**, H134722 **(b)** and H132374 **(c)** are shown with treatment timeline and signature tree as

888 described in Fig. 3a and Fig. 4a, respectively. Detailed description of each patient is provided in

889 Supplementary Information. **a)** Body map on the left shows the seeding events before

890 diagnosis in H103207. Body map on the right depicts the polyclonal seeding after therapy

891 amongst liver and bilateral lungs involving subclones 5 and 8. **b)** The left body map shows the

892 spread before therapy. The black arc indicates the distinct subclone in R1 left lung metastasis

893 sequenced with MSK-IMPACT. The right body map shows subclones 2, 3, 4, 7 and 10 involved

894 in polyclonal seeding across locoregional and metastatic sites in H134722. *TP53* substitution

895 assigned to subclone-7 is also found in PDX modeling of R3, R4, R8, R9, R10 and R11. **c)** Body

896 maps depict two different scenarios that explain the subclonal structure in H132374: The left

897 body map shows possible polyclonal seeding in the CNS by a mixture of subclones 3 and 4. In

898 this scenario lung metastasis is caused by subclone-4 after platinum chemotherapy. Shown on

899 the right body map is the second scenario of met-to-met seeding from CNS to lung by subclone-

900 4 after therapy. The data for Fig. 5 are available as raw data at dbGAP and scripts are available

901 through ISABL platform.

902   **Extended Data Figure 1. Summary of mutation calling and genomic landscape. a)** Barplots

903   show the number of substitutions, indels and SVs identified in WGS tumors (n=247) in the

904   cohort grouped by the different disease subtypes and color-coded by the sample type. **b)**

905   Barplots show the prevalence of segmental CNAs and genes affected by mutations and SVs

906   across the cohort (n=470 tumors). Only genes affected in at least two patients are shown.

907   Bottom bar plot gives a summary of the type of mutations for each CNA/gene. CNA, copy

908   number aberration. Complex, small complex insertion/deletion. Del, small deletion. Ins, small

909   insertion. Sub, substitution. SV, structural variant. **c)** Survival plot shows the clinical outcome

910   of *MYCN*-A patients (n=68) with *TERT*p substitutions, *TERT*-SV or no *TERT* events with 95%

911   confidence intervals. P-value from coxph analysis taking into account age at diagnosis is

912   shown. **d)** Heatmap gives a summary of the co-mutation patterns in the current cohort with

913   the frequency of events in the upper triangle and odds ratios in the lower triangle. Only odds

914   ratios with p values <0.05 are colored in shades of blue for co-mutation or red for mutually

915   exclusive interactions. Significant interactions are indicated with a star or a dot according to

916   the significance level. FDR, false discovery rate. FWER, family-wise error rate. The data and

917   script for Extended Fig. 1 are available in Supplementary Tables 1-2 and the GitHub repository.

918   **Extended Data Figure 2. Summary of mutational signature analyses in the current WGS data.**

919   **a)** 96 mutational contexts for the substitution signatures identified de novo are shown as the

920   barplots on the left while the reference signatures from COSMIC.v3 are shown in the barplot in

921   the middle. Right barplot shows the prevalence of different types of indels amongst the indel

922   signatures identified de novo. **b)** Heatmap shows the proportions of substitutions at 96

923   mutational contexts for each WGS tumor (n=247) shown in rows together with sample type,

924   disease subtype, platinum, temozolomide and radiotherapy status on the left and number of

925   substitutions and indels and exposure to identified signatures in substitution and indel data on

926   the right. The data and script for Extended Fig. 2 are available at the GitHub repository.

37

927 **Extended Data Figure 3. Biological and clinical correlates of mutational patterns. a)** Scatter

928 plots show the association between exposure to SBS18 (left) or SBS40 (right) and age at

929 diagnosis amongst the diagnostic/t-resection tumors (Pearson correlation). **b)** Box plot shows

930 the mean expression of the genes in glutaminolysis signature associated with ROS

931 accumulation[22] across diagnostic tumors of different disease subtypes (left) and *MYCN*-A

932 tumors from diagnosis, t-resection and relapse and further relapses (right). Comparisons with

933 significant p-values are shown. **c)** Scatter plot on the left shows correlation between number of

934 SNVs and the number of predicted neoantigens while the scatter plot on the right shows the

935 relationship between the number of predicted neoantigens and immune infiltrates in the

936 surrounding tumor microenvironment as assessed from RNAseq (Pearson correlation). **d)**

937 Barplots show the proportion of genome-wide SNVs (left) and oncogenic driver SNVs (right)

938 attributed to different mutational signatures broken down by presence in post-therapy relapse

939 tumors. The data and script for Extended Fig. 3 are available in Supplementary Table 1 and the

940 GitHub repository.

941 **Extended Data Figure 4-5. Subclone trees for 94 neuroblastoma patients with two or more**

942 **tumors sequenced.** Each tree shows the subclonal structure in an individual patient. Patients

943 are organized according to disease subtype and the availability of tumors from primary site

944 (diagnosis/reresection/t-resection) and relapses. Branches are annotated with recurrent CNAs

945 and oncogenic mutations/SVs and colored according to the latest tumor they were identified in

946 1) blue for subclones specific to a diagnostic tumor 2) light blue for subclones seen in

947 reresections 3) green for subclones seen in a t-resection tumor 4) orange for subclones seen in

948 a relapse tumor and 5) red for subclones seen in a further relapse tumor. Subclonal events at

949 *MYCN*, *TERT* and *ATRX* loci are shown in red font. Events with which clonal status cannot be

950 determined are indicated with a question mark. Different evolutionary patterns are indicated

951 with an icon next to the patient id. Tumor sites and the type of sequencing are indicated below

952   the trees. G, whole-genome sequencing. T, targeted sequencing. B, both. For H103207,

953   H118706 and H134819 a simplified version of the tree is shown due to space. Detailed analysis

954   of subclonal structure for 94 patients is provided in Supplementary Fig. 2-46 and 48-53. The

955   data Extended Figs. 4-5 are available in Supplementary Table 5 and the scripts are available

956   through the ISABL platform.

957   **Extended Data Figure 6. Timing of metastasis.** Signatures trees as described in Fig-4a are

958   shown for 13 patients with one or more tumors from the primary site and two or more tumors

959   from locoregional and/or distant metastasis. Patients in the top row have at least one tumor

960   from distant metastatic site while patients in the bottom row have locoregional relapses only.

961   The subclones involved in disease spread from the primary are indicated with a black arrow

962   while the daughter clones are shown with red arrows. The data for Extended Fig. 6 are

963   available as raw data at dbGAP and the scripts are available through the ISABL platform.

964

965

966
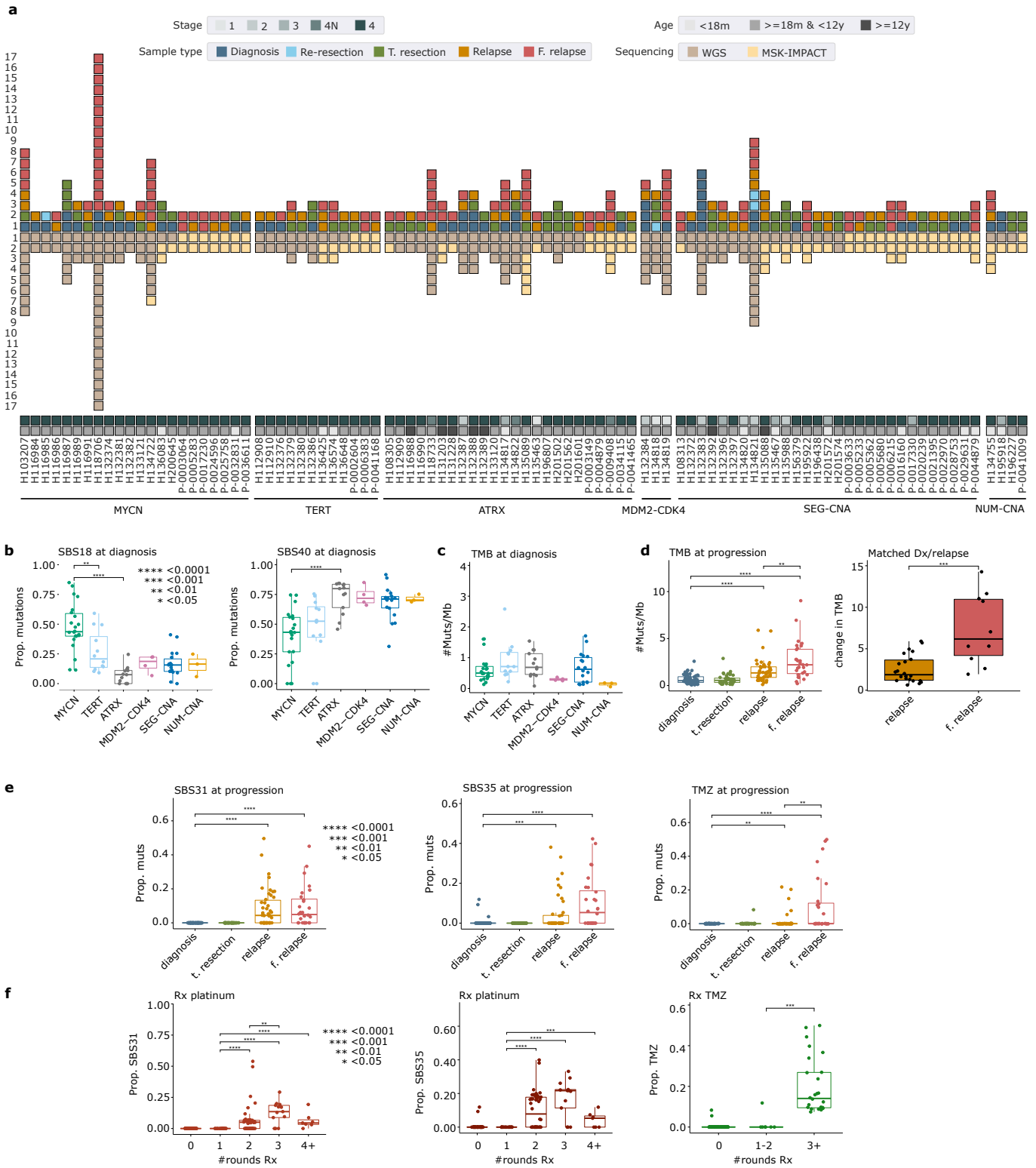
Figure 1
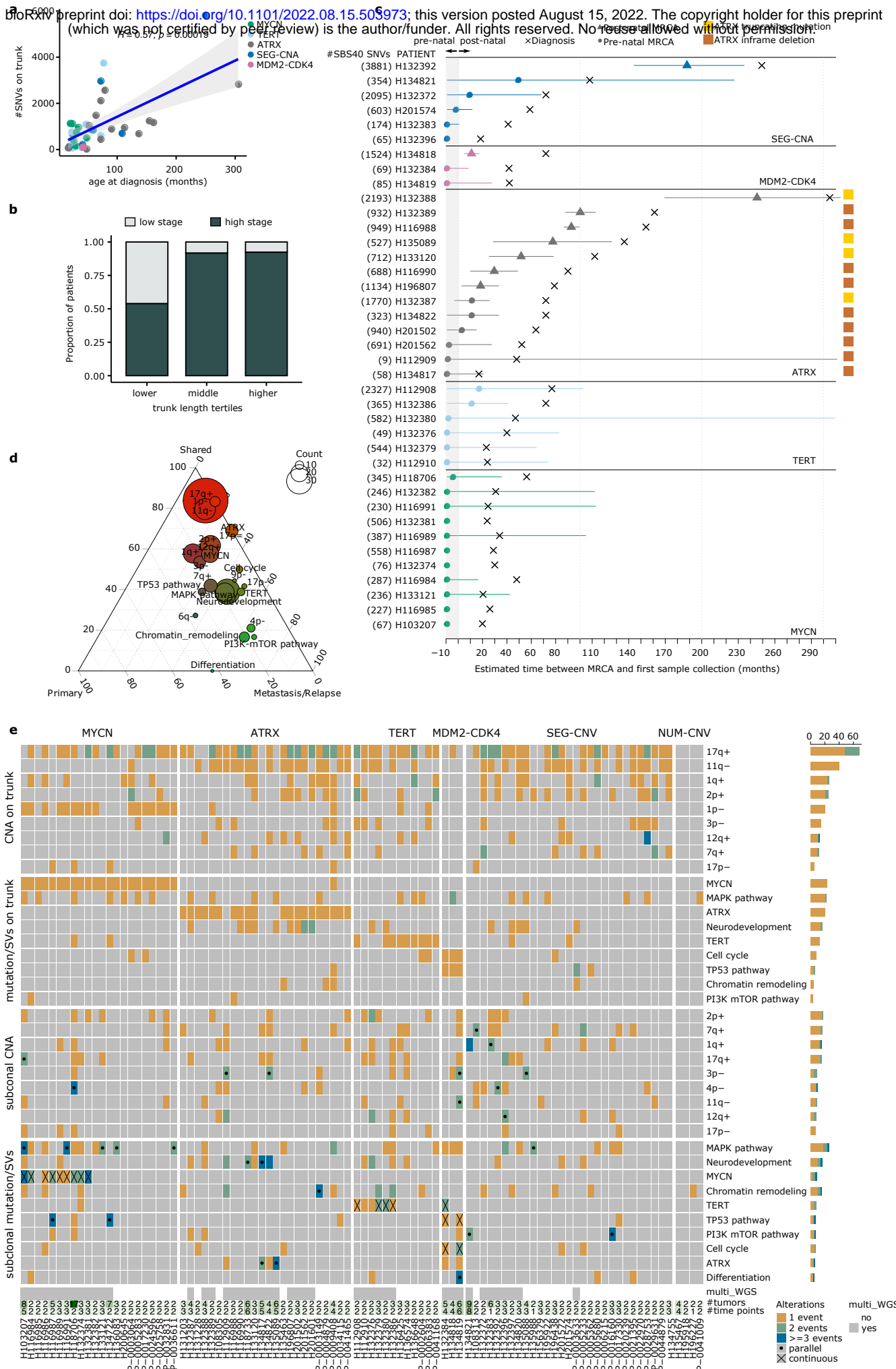
Figure 2

Figure 3

Figure 4

Figure 5

a H103207

D1/SL1 = adrenal R4 = liver (met)
R1-2 = CNS (met) R5 = retroperitoneal lymph node (local)
R3 = left lung (met) R6 = right lung (met)

b H134722

R3 = paraesophageal (local) R8 = right lung (met) (PDX)
R4 = paraspinal (local) R9 = mediastinal lymph node (local) (PDX)
R5 = pancreas (local) R10 = adrenal (PDX)
R6 = left kidney (local) R11 = retroperitoneal (local) (PDX)
R7 = right lung (met)

TP53 SNV detected in R3-4 and
R8-R11 in PDX modeling

c H132374

D1 = RP
R1 = brain(met)
R2 = lung (met)

Extended Data Figure 1

a

Extended Data Figure 1

Extended Data Figure 2

a

b
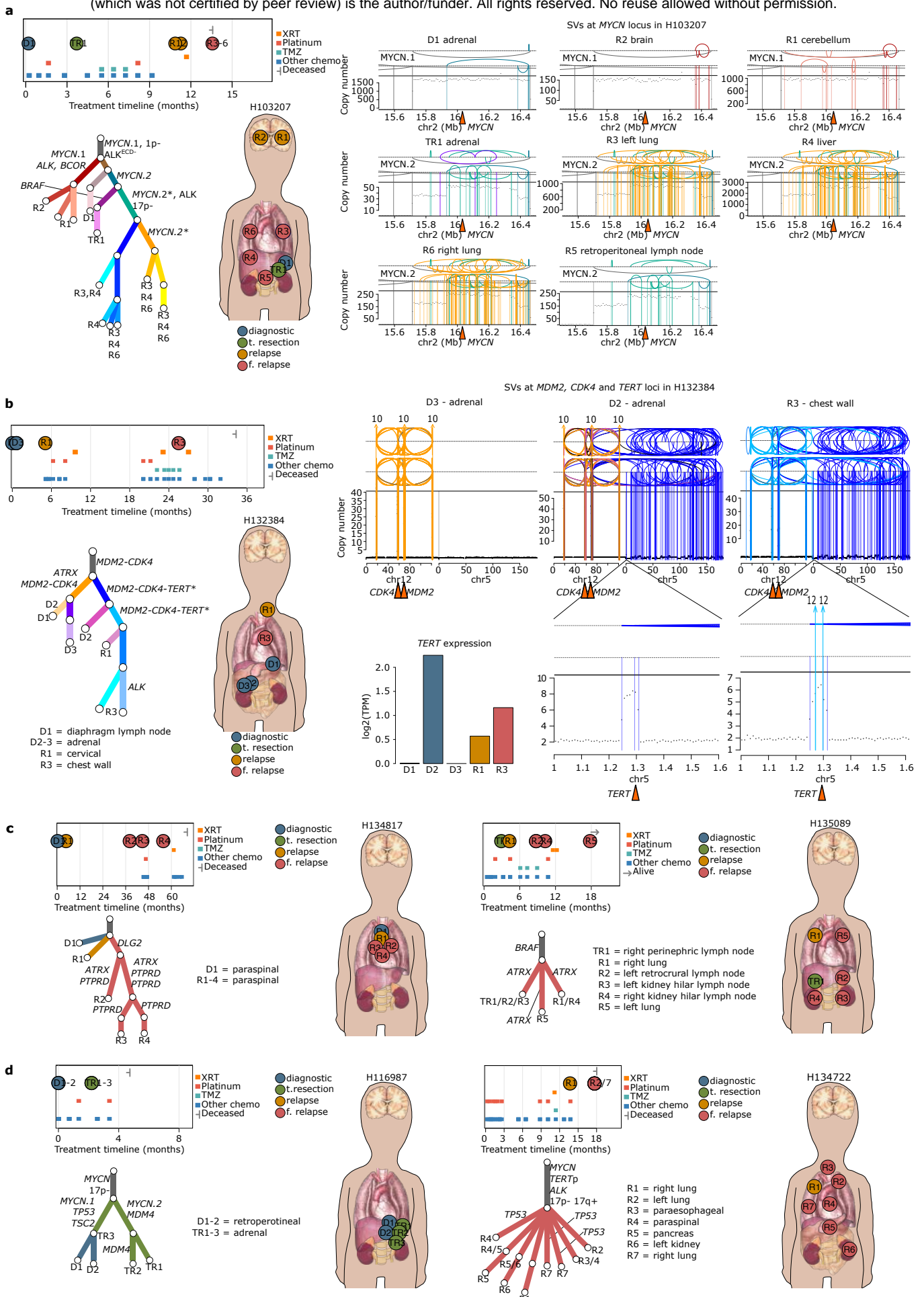
Extended Data Figure 3

Extended Data Figure 4
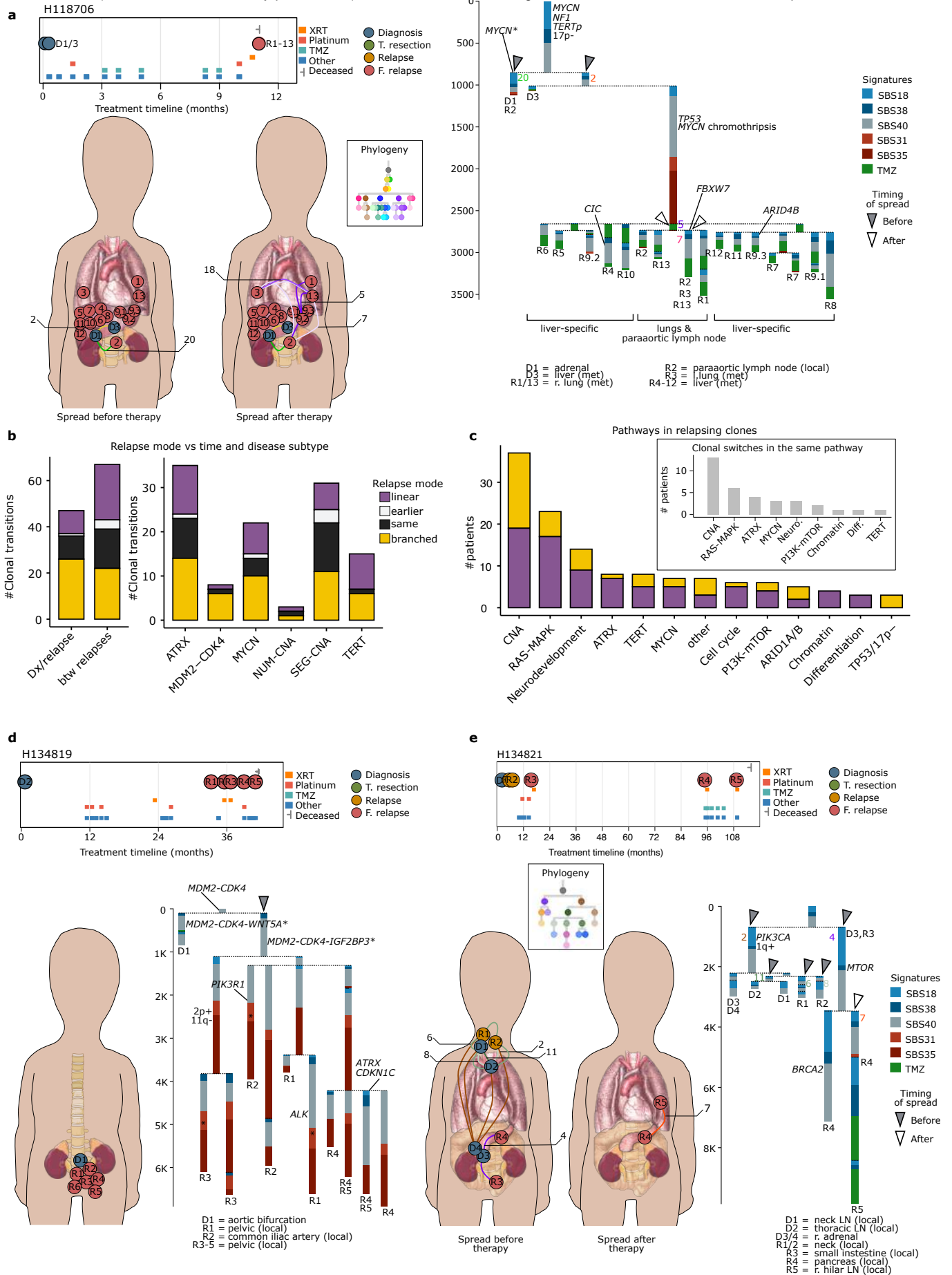
Extended Data Figure 5

Extended Data Figure 6

H103207

D1/TR1 = adrenal   R4 = liver (met)
R1/R2 = CNS (met)   R5 = retroperitoneal LN (local)
R3 = l. lung (met)   R6 = r. lung (met)

H118706

D1 = adrenal   R2 = paraortic LN (local)
D3 = liver (met)   R3 = l. lung (met)
R1/R13 = r. lung (met)   R4-12 = liver (met)

H132374

D1 = abdominal
R1 = brain (met)
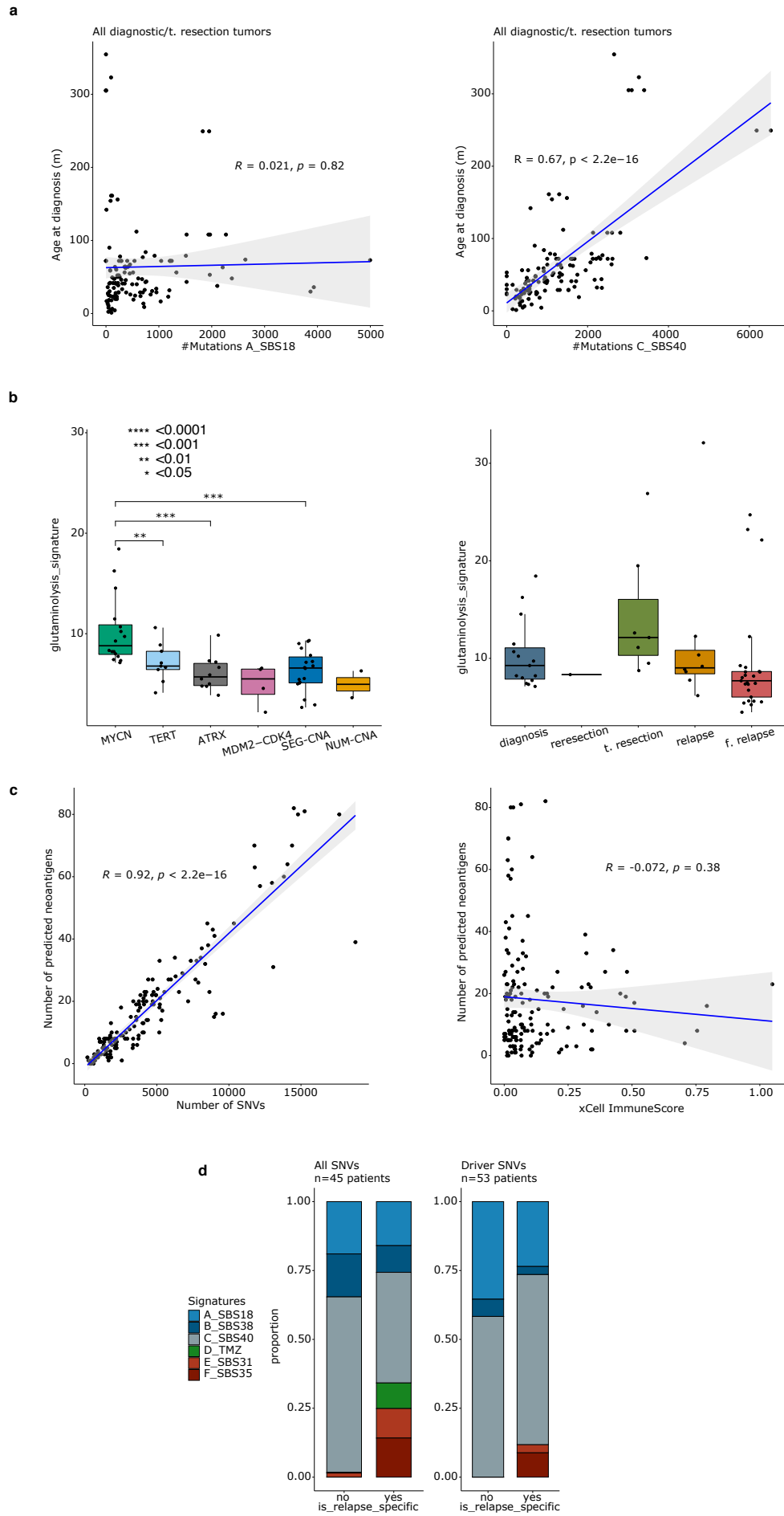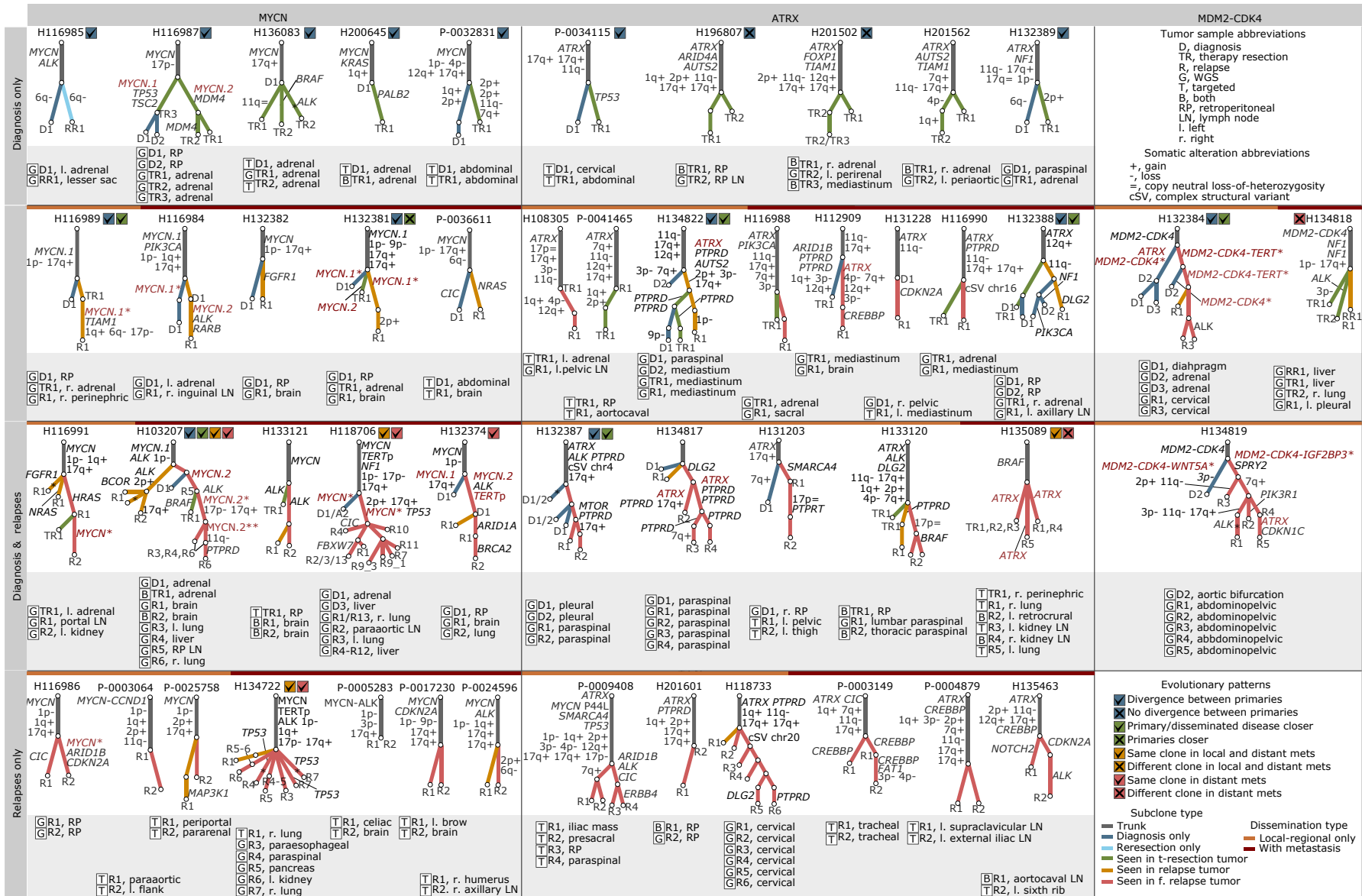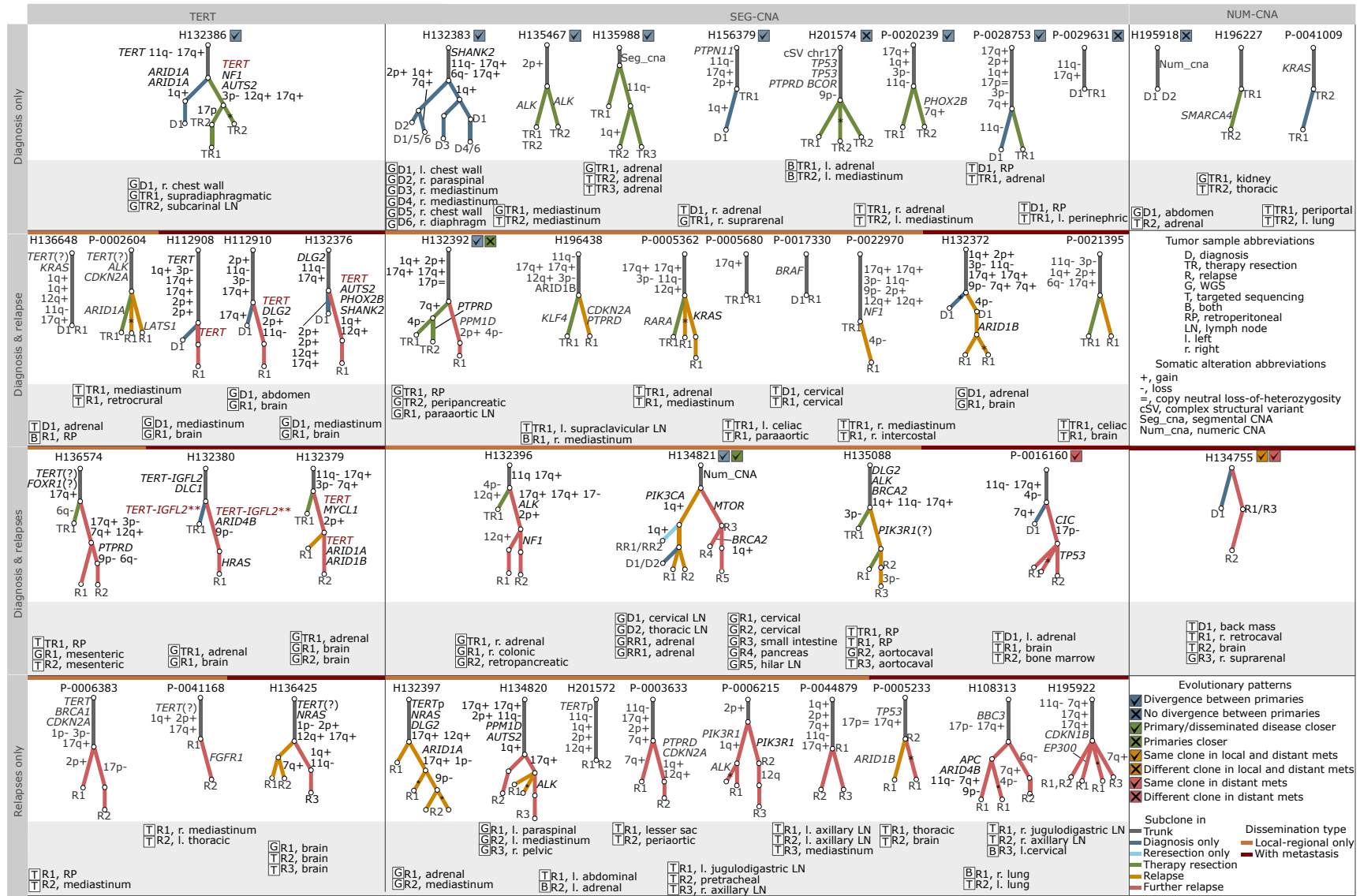R2 = lung (met)

H132379

D1 = adrenal
R1 = brain (met)
R2 = brain (met)

H133121

R1 = brain (met)
R2 = brain (met)

H135089

TR1 = r. kidney   R3 = l. kidney LN (local)
R1 = r. lung (met)   R4 = r. kidney LN (local)
R2 = l. abdominal LN (local)

H116991

TR1 = adrenal
R1 = portal LN (local)
R2 = kidney (local)

H132384

D1 = diaphragm LN
D2/3 = adrenal
R1 = cervical (local)
R3 = chest wall (local)

H132387

D1/2 = pleural
R1/2 = paraspinal (local)

H132396

TR1 = adrenal
R1 = colonic (local)
R2 = paraspinal (local)

H133120

TR1 = RP
R1 = paraspinal (local)
R2 = paraspinal (local)

H134819

D1 = abdominopelvic
R1-5 = abdominopelvic (local)

H134821

D1 = cervical LN   R3 = small intestine (local)
D2 = thoracic LN   R4 = pancreas (local)
D3/4 = r. adrenal   R5 = r. hilar LN (local)
R1/2 = cervical (local)

Legend

Signatures
- SBS18
- SBS38
- SBS40
- SBS31
- SBS35
- TMZ

Clones
- → Metastasizing
- → Daughter in metastasis

Therapy
- XRT
- Platinum
- TMZ
- Other chemo

Outcome
- Deceased
- Alive