

1 **Lineage origin and microenvironment shape neuroblastoma**
2 **transcriptional state and plasticity**

3 *Nora Fresmann^{1,2}, Julia Köppke³, Anton Gauert³, Anis Senoussi¹, Pedro Olivares-Chauvet^{1,4},*
4 *Marie Schott⁵, Lennart Höfer³, Anton G. Henssen^{3,6,7}, Nikolaus Rajewsky^{2,5,7,8,9,10,11}, Bastiaan*
5 *Spanjaard^{3,6,*}, Anja I. H. Hagemann^{3,7,*}, Jan Philipp Junker^{1,2,11,*}*

6

7 *1 Laboratory for Quantitative Developmental Biology, Berlin Institute for Medical Systems*
8 *Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin,*
9 *Germany*

10 *2 Charité – Universitätsmedizin Berlin, Germany*

11 *3 Department of Pediatric Oncology and Hematology, Charité – Universitätsmedizin Berlin,*
12 *Germany*

13 *4 Department of Systems Immunology, Weizmann Institute of Science, Rehovot, Israel*

14 *5 Laboratory for Systems Biology of Regulatory Elements, Berlin Institute for Medical Systems*
15 *Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin,*
16 *Germany*

17 *6 Experimental and Clinical Research Center of the MDC and Charité Berlin, Berlin, Germany*

18 *7 German Cancer Consortium (DKTK), Heidelberg, Germany*

19 *8 National Center for Tumor Diseases (NCT), Site Berlin, Berlin, Germany*

20 *9 NeuroCure Cluster of Excellence, Berlin, Germany*

21 *10 German Center for Child and Adolescent Health (DZKJ), Site Berlin, Germany*

22 *11 German Center for Cardiovascular Research (DZHK), Site Berlin, Berlin, Germany*

23

24 **correspondence: bastiaan.spanjaard@charite.de, anja.heeren-hagemann@charite.de,*
25 *janphilipp.junker@mdc-berlin.de*

26

27 **Abstract**

28 Neuroblastoma, a neural-crest-derived malignancy of the peripheral nervous system, is
29 a devastating pediatric disease, characterized by high intra- and intertumoral
30 heterogeneity. While expression of several tumor expression modules correlates with
31 poor patient survival, the determinants of their emergence and plasticity remain elusive.
32 Here, we systematically dissected neuroblastoma transcriptional heterogeneity and
33 measured how tumor expression programs are determined by early developmental
34 signaling versus local tumor environment. To achieve this, we combined single-cell
35 transcriptomics with high-throughput lineage tracing and tumor cell transplantations in
36 zebrafish models of high-risk neuroblastoma. We observed transcriptional programs
37 determined by the cell of origin, including an ALK-activated state linked to poor disease
38 prognosis in humans – in contrast to plastic states associated with physiological
39 processes. Even lineage-determined tumor states can be reprogrammed upon
40 exposure to a developmental signaling environment, indicating high plastic potential *in*
41 *vivo* and a crucial role for the signals received in early tumorigenesis for tumor
42 phenotype.

43

44 **Main text**

45 *Introduction*

46 Transcriptional heterogeneity and phenotypic plasticity are increasingly recognized as
47 drivers of tumorigenesis, metastatic dissemination and treatment evasion¹⁻³. Phenotypic
48 differences between tumor cells can be described by their transcriptional states, which
49 are defined by expression of gene modules - groups of co-regulated genes that
50 comprise both cell identity-specific as well as physiological programs^{4,5}. Plasticity, the
51 ability of a cell to switch between different transcriptional states, is crucial during e.g.
52 development, but is largely lost in fully differentiated cells in healthy tissues. Cancer
53 cells override these rules, exhibiting the capability to switch between different
54 phenotypes⁶. A key question is how the gene expression programs they access relate
55 to their cell of origin and capacity for phenotypic plasticity.

56 Recent pan-cancer studies suggest that tumor cells access a common set of
57 transcriptional programs related to general cellular processes, such as stress response
58 or cell cycle^{4,5}. In contrast to this, cell identity programs are cancer type-specific and
59 derive from the cell type of origin and developmentally related cell types⁷. Efforts to link
60 tumor cell lineage and state in cancer animal models have elucidated that a single cell
61 can give rise to a complex tumor with diverse cell identities, e.g. alveolar type I, type II
62 and gastric-like states in lung adenocarcinomas⁸⁻¹⁰. This plastic behavior arises from
63 the interplay of cell-intrinsic mechanisms, including genetic and epigenetic state and
64 interaction with the tumor microenvironment.

65 The link between differentiation and tumor cell state is particularly relevant for
66 pediatric cancers, which arise from developmental precursor cells and where cell of
67 origin and early developmental environment can profoundly influence tumor behavior¹¹.
68 Neuroblastoma (NB) is a childhood cancer with heterogeneous disease progression,
69 high metastatic potential and low survival rates for high-risk patients^{12,13}. NB arises from
70 cells of the developing sympathoadrenal lineage with the earliest tumorigenic events
71 occurring in the first trimester of pregnancy¹⁴. Amplification of the MYCN oncogene is
72 observed in 20 % of NB patients and is a strong predictor for poor prognosis^{15,16}.
73 MYCN-amplification is an early event in NB-formation and studies have shown that
74 MYCN alone can induce NB in neural crest derivatives^{14,17-19}. Despite the strong
75 changes that sympathoadrenal cells undergo during development, the impact of the cell
76 of origin on the tumor state of MYCN-driven NB remains unknown.

77 In cell culture, NB has been shown to exist in two different, interconvertible tumor
78 states, being either adrenergic or mesenchymal²⁰⁻²². Adrenergic NB cells are
79 neuroblastic and express sympathoadrenal genes and enzymes for neurotransmitter
80 biosynthesis²⁰. In contrast, mesenchymal NB cells rather resemble non-neuronal neural
81 crest derivatives such as smooth muscle or Schwann cells in their gene expression
82 profile. MYCN, together with the transcriptional co-activator LMO1, has been shown to
83 reinforce the adrenergic core regulatory circuit of transcription factors and to thus keep
84 NB cells in an undifferentiated neuronal progenitor state^{22,23}. Extensive single-cell RNA-
85 sequencing (scRNA-seq) studies on patient samples and healthy fetal adrenal glands
86 have shown that NB cells are also mostly adrenergic *in vivo*, but show transcriptional

87 heterogeneity associated with their lineage, resembling developmental cell types such
88 as neuroblasts, chromaffin cells or earlier precursor states^{24–28}. While these studies
89 show that NBs display transcriptional heterogeneity associated with disease risk, little is
90 known about the regulation of these states *in vivo*. In particular, it remains unclear to
91 which degree their activation and plasticity are shaped by the state of the cell of origin.
92 Studying these questions in patient data is challenging due to limited experimental
93 accessibility and confounding factors, including population genetic and tumor genome
94 diversity and technical batch effects. Thus, we currently cannot distinguish between the
95 following two scenarios: 1) Transcriptional programs are stable over long periods of time
96 and are hence indicative of the earliest events in neoplastic transformation in a specific
97 cellular origin; and 2) Transcriptional programs are highly plastic and represent attractor
98 states in gene expression space between which tumor cells can readily transition.

99 Here, we address these questions using well-established zebrafish models of
100 NB^{29,30}, in which we find substantial heterogeneity of tumor expression programs. In
101 these transgenic lines, human *MYCN* is specifically activated in sympathoadrenal cells,
102 leading to growth of tumors that histopathologically resemble human NBs. This well-
103 controlled system allows measurement of transcriptional cell states together with clonal
104 structure via high-throughput lineage tracing. We found that tumors in this model are
105 composed of clones from multiple cells of origin, allowing us to directly measure the
106 influence of lineage on gene expression within individual tumors. We identified a range
107 of NB transcriptional states that are either related to general cellular processes or to
108 specific cell identity-programs. We found that cell-identity related states tend to be
109 determined by the cell of origin and are subsequently stably expressed within cells of
110 one clone, suggesting an important role for the developmental state of the cell of origin
111 for tumor state. By transplanting primary zebrafish NB cells into zebrafish embryos, we
112 showed that even stably activated transcriptional states can be reprogrammed when
113 exposed to developmental signals. This highlights the role of developmental signals
114 received during early tumorigenesis for tumor cell state and potentially disease severity.

115

116

117 *Results*

118 *Dissecting tumor expression heterogeneity by multiplexed single-cell transcriptomics*

119 We used single-cell transcriptomics to analyze the diversity of tumor transcriptional
120 states in the two established transgenic zebrafish lines that closely reiterate the
121 pathogenesis of human neuroblastoma $tg[dbh:MYCN, dbh:EGFP]$ and $tg[dbh:MYCN,$
122 $dbh:LMO1, dbh:EGFP, dbh:mCherry]$ ^{29,30} (hereafter called MYCN and MYCN;LMO1,
123 respectively). To enable later assessment of the influence of cell lineage on
124 transcriptional state, we combined scRNA-seq with high-throughput lineage tracing
125 using CRISPR/Cas9 induced lineage barcodes^{31–33} (Fig. 1A, details described later in
126 Fig. 3). Single-cell expression profiles and lineage barcodes were read out jointly by
127 scRNA-seq in tumors from adult fish, using the cell hashing method MULTI-seq³⁴. With
128 this cell hashing approach, we jointly processed up to 14 tumors in one scRNA-seq run,
129 thereby minimizing experimental batch effects (Fig. 1A, Table S1).

130 In agreement with previous reports, we found efficient induction of tumors
131 between 6 and 15 weeks post fertilization (wpf), with faster induction in the
132 MYCN;LMO1 line (Fig. 1B). As expected, the fish developed tumors in the interrenal
133 gland (IRG), which is equivalent to the human adrenal gland, and the superior cervical
134 ganglion; we refer to these tumor locations as lateral. Additionally, only MYCN;LMO1
135 fish developed tumors in the arch-associated complex (AAC)^{18,29,35}, a location we refer
136 to as ventral (Fig. S1A). Thus, *LMO1* expression does not only increase MYCN-driven
137 tumor penetrance, but also enables tumorigenic transformation in additional
138 sympathoadrenal progenitor populations that *MYCN* expression alone cannot transform.
139 Tumors in both locations showed the typical small round blue cell phenotype of
140 neuroblastoma (NB)³⁶ (Fig. 1C). In total we sequenced the transcriptomes of 141,812
141 single cells from 60 tumors and 9405 cells from three healthy control samples (head
142 kidneys with IRG). Whenever possible, we split larger tumors into multiple samples to
143 gain sub-tumor resolution (Table S2). We detected 101,872 NB cells, expressing
144 sympathoadrenal and known NB markers *phox2bb*, *hand2*, *dbh* and tumor transgenes
145 (Fig. 1D, Fig. S1, Table S3). In the stromal and immune compartment of tumor samples,
146 we detected 39,940 cells, including various kidney cell types (e.g. kidney tubule and

147 multiciliated cells³⁷) and steroidogenic interrenal cells from the IRG (Fig. S1B-C).
148 Immune cells were likely derived from tumor immune invasion, but also partially from
149 the hematopoietic tissue in the zebrafish kidney marrow, equivalent to human bone
150 marrow (Fig. S1B).

151 Within individual MULTI-seq runs, technical batch effects were minimal with
152 tumor microenvironmental (TME) cells from different tumor samples intermixing in
153 clusters and on the UMAP (Fig. 1E). By contrast, tumor cells exhibited substantial inter-
154 and intra-tumor transcriptional heterogeneity, reflected by clusters composed of cells
155 from one or only few samples as well as cells of one tumor spread across multiple
156 clusters. This is reminiscent of the pronounced phenotypic heterogeneity observed
157 between patient samples, often driven by genetic variation. However, we detected
158 neither additional mutations in whole exome sequencing nor copy number variants
159 when inferring copy numbers from scRNA-seq data (Fig. S2). This suggests that the
160 detected differences are largely transcriptional and may therefore originate from the cell
161 of origin of the tumor or the influence of the tumor niche.

162 To better understand the overall transcriptional profile of NB cells, we performed
163 differential expression analysis between NB cells and all other cells. This showed that
164 NB cells are characterized by clear expression of the tumor transgenes and known
165 adrenergic genes (*dbh*, *hand2*, *elavl3*; signature ‘NB_markers’) (Fig. 1F, Table S4). The
166 100 most highly expressed genes in NB cells comprised almost exclusively ribosomal
167 genes (signature ‘ribosomal_genes’, Fig. 1G, Table S4), in line with the reported
168 increase in ribosomal biogenesis induced by MYCN^{38,39}. We further calculated
169 expression of the human cell line-derived adrenergic and mesenchymal NB signatures.
170 In line with findings in human primary tumors, we found the adrenergic signature to be
171 expressed in zebrafish NB cells, whereas the corresponding mesenchymal signature
172 was more highly expressed in TME cells (Fig. 1G). Taken together, these findings show
173 that zebrafish NB transcriptomes are overall adrenergic and shaped by MYCN-activity.

174 Beyond the overarching NB cell transcriptomic profile, we next sought to
175 investigate the observed transcriptional heterogeneity within the population (Fig. 1E) by

176 performing a systematic *de novo* analysis of tumor gene expression programs in our
177 zebrafish NB data.

178

179 *The spectrum of MYCN-driven NB transcriptional programs*

180 Gene expression in tumor cells has been shown to be composed of multiple gene
181 expression programs, which can be active to different degrees in individual cells^{4,5,40}.
182 Non-negative matrix factorization (NMF)-based approaches have previously been
183 shown to reliably detect groups of co-varying genes, typically called gene modules in
184 heterogeneous scRNA-seq data of malignant cells^{4,5}.

185 In order to identify gene modules that capture both intra- and inter-tumor
186 expression variation, we performed NMF on three different levels (Fig. 2A). In the first
187 instance, we ran NMF on NB cells from individual tumors separately, resulting in a list of
188 modules representing expression variation between NB cells within each tumor. We
189 grouped these modules by their similarity in gene content to derive recurrently activated
190 consensus modules (Fig. S3A, Methods^{4,5,38}). We then repeated this analysis using NB
191 cells from individual MULTI-seq runs as input. This approach allowed the identification
192 of gene modules that are differentially activated between tumors without suffering from
193 batch effects in the data. Lastly, we also ran NMF on the whole NB cell population from
194 all samples. While this method is more prone to capture technical, in addition to
195 biological variation, it is useful to identify gene modules that are only activated in few
196 samples or cells. We annotated modules by associated GO-terms and the functional
197 annotation of individual genes contributing to them (Tables S4-5). All three approaches
198 showed some overlap in detected gene modules (Fig. S3B). We therefore compiled a
199 final combined list of modules that are largely non-overlapping in terms of their gene
200 content and which represent the spectrum of NB cell states (Fig. 2B, Table S6,
201 Methods).

202 The resulting 17 gene modules comprise programs related to sympathoadrenal
203 tissue development and thus the cell identity of the lineage NB is derived from (Fig. 2C,
204 termed sympathoadrenal-specific; e.g. *catecholamine_production*,
205 *early_adrenergic_development*, *immature_neuronal*). In addition, the list comprised

206 modules associated with general cellular processes, which are not cell type specific
207 (termed cellular process; e.g. *interferon_signaling*, *stress_response*, *cell_cycle*). Of
208 note, the module *immature_neuronal* includes the gene *alk*, increased activation of
209 which is associated with poor prognosis in patients^{41–43}. We found that the detected
210 modules varied across individual cells and tumors, validating that they capture some of
211 the observed expression heterogeneity (Fig. 2C, Fig. S3C). Some modules also showed
212 a clear dependence on tumor location, especially *immature_neuronal*, which was only
213 activated in ventral tumors; conversely, *catecholamine_production* was more highly
214 activated in lateral tumors.

215 Multiple modules contain the *MYCN* transgene (*early_adrenergic_development*,
216 *neuron_projection*, *immature_neuronal*), supporting the notion that *MYCN* drives
217 several distinct zebrafish NB cell transcriptional states. Expression of the
218 *early_adrenergic_development* and *immature_neuronal* modules was particularly
219 associated with *MYCN* expression levels and the expression of published *MYCN*-driven
220 genes in the zebrafish data (Fig. 2D, Fig. S4A)⁴⁴, while *catecholamine_production* was
221 more weakly associated with *MYCN* expression. In contrast, PRC2-target gene
222 expression showed a negative association with expression of the detected modules,
223 consistent with *MYCN*-mediated gene silencing via *EZH2*⁴⁵. Notably, we did not observe
224 a correlation between expression of *MYCN* or downstream targets and the *cell_cycle*
225 module. This may be explained by the cell cycle-independent nature of the *MYCN*-
226 upregulated signature as well as the *dbh*-driven *MYCN* transgene, which is likely not
227 transcribed in a cell cycle-dependent manner.

228 To further test the relevance of these modules for human cancer, we compared
229 them to gene modules from recent cancer studies and scored their expression in
230 published human NB datasets. As expected, programs like *stress_response* and
231 *interferon_signaling* resembled general programs activated in many cancers^{4,5} (Fig.
232 S4B). Conversely, multiple sympathoadrenal-specific zebrafish modules overlapped
233 with broad adrenergic / neuronal programs derived from human NB^{20,28,46,47} (Fig. S4C).
234 Notably, the zebrafish *immature_neuronal* module overlapped with a human *ALK*-
235 activated program and the *ribosomal_genes* module matched known *MYCN*-driven
236 modules. We then scored expression of the zebrafish NB modules in a published

237 scRNA-seq dataset of human NB²⁵ and found variable activation across modules (Fig.
238 S4D). In a compendium of published human NB bulk RNA-seq datasets from the
239 TARGET and SEQC cohorts^{12,48–50}, the modules *cell_cycle*, *ribosomal_genes*, and
240 *immature_neuronal* scored significantly higher in *MYCN*-amplified and high-risk tumors
241 (Fig. 2E Fig. S4E), underscoring their relevance for high-risk disease. In contrast, the
242 *catecholamine_production* module was more highly expressed in low-risk and non-
243 *MYCN*-amplified tumors, consistent with a more differentiated neuroendocrine state.
244 The *immature_neuronal* module further showed variable expression in both *ALK*-
245 mutated and non-mutated tumors (Fig. 2F, Fig. S4F), with a slightly, albeit non-
246 significant, higher expression in the former, further indicating it may represent an *ALK*-
247 activated state.

248 Together, these findings demonstrate that the zebrafish-derived modules capture
249 key oncogenic and differentiation-associated programs relevant to human NB biology.
250 We hypothesized that activation of some modules might be clonally determined, for
251 instance by the cell of origin that a tumor cell was derived from, while others may be
252 regulated in a niche-dependent manner. We further speculated that the degree of
253 lineage determination might be higher for modules related to sympathoadrenal
254 development compared to modules related to general cellular processes. In order to test
255 this, we next analyzed the clonal structure of zebrafish NB tumors using lineage tracing.

256

257 *High-throughput lineage tracing identifies multiple cells of origin per tumor*

258 To experimentally measure the influence of lineage on transcriptional state, we
259 combined scRNA-seq with high-throughput lineage tracing using CRISPR/Cas9 induced
260 lineage barcodes^{31–33}, which are created by injection of Cas9 and sgRNAs targeting
261 lineage recording sites into zebrafish embryos at the one-cell stage (Fig. 3A, Methods).
262 In our system, lineage barcodes are created on multiple integrations of a cassette of
263 three Cas9 target sites in the 3'-untranslated region (UTR) of a dsRed transgene (Fig.
264 S5A-C) as well as in the 3'-UTRs of seven endogenous genes (Fig. 3A). Lineage
265 barcodes mostly present as small insertions or deletions around the Cas9 target site
266 (Fig. 3B). We measured that lineage barcodes are created within the first 8 hours post

267 fertilization (hpf) (Fig. 3C). This precedes the onset of *dbh* expression at 14 hpf⁵¹ (Fig.
268 3D). Consequently, activation of the tumor-inducing *dbh:MYCN* transgene occurs after
269 lineage barcode creation, ensuring that all progeny of a given transformed cell inherit
270 the same lineage barcodes¹⁸ (Fig. 3A). By contrast, tumor cells derived from different
271 cells of origin can be distinguished based on differing lineage barcodes.

272 Joint analysis of single-cell transcriptomes and lineage barcodes revealed that
273 NB cells and TME cells typically have separate lineage scar profiles, in line with the
274 distinct lineage origins of these cell types (Fig. 3E). Furthermore, we found that the NB
275 cells of a single tumor were composed of cells with multiple distinct lineage barcode
276 profiles and were hence derived from multiple cells of origin (Fig. 3E). We attribute this
277 to the strong effect of *MYCN* in our genetic models, which induces tumorigenic
278 transformation in many cells. We hereafter refer to these groups of cells from different
279 origins as clones. We clustered NB cells from all 38 tumors with lineage information into
280 clones according to their lineage barcode pattern across all lineage reporter sites,
281 focusing on maximizing clonal resolution (Methods, Fig. S5D-G). We typically found
282 between 2 and 6 NB cell clones (and hence cells of origin) per tumor (Fig. 3F, S5H-L).
283 We found that ventral and lateral tumors originating from the same fish always had
284 completely distinct clonal composition, indicating different lineage origins between these
285 two tumor types (Fig. S6A). The multi-cellular origin of individual tumors now allows us
286 to study gene expression differences between NB cell clones in a shared environment,
287 and thus assess lineage effects on transcriptional state independently of confounding
288 effects related to differences between individual tumors.

289

290 *Clonal analysis reveals differential plasticity of tumor states*

291 We next sought to use this approach for quantifying to which degree transcriptional
292 heterogeneity within one tumor is driven by different cells of origin. In an extreme
293 scenario, tumor states are fixed by the cell of origin, with no expression state transitions
294 and thus no plasticity, leading to co-segregation of cells by gene expression and clone
295 (Fig. 4A). In the opposite scenario, cells are highly plastic, and tumor states are
296 independent of the cell of origin.

297 To assess the clonal determination of specific states, we first examined
298 expression of the modules identified in Fig. 2 in the different clones of selected
299 individual tumors. We found that some modules, such as *cell_cycle*, were expressed at
300 similar levels across the different clones of a tumor, while sympathoadrenal-specific
301 modules (e.g. *catecholamine_production*) tended to vary between the clones of a tumor
302 (Fig. 4B, Fig. S6A). For a systematic analysis of tumor state association with clonal
303 origin, we next calculated the differential module expression between all clone pairs
304 within the same (sub-)tumor (Methods, Fig. 4C-D, S6B). The inter-clone differences
305 reported by this analysis correspond to the effective state determination by the clonal
306 origin. We observed considerable differences in lineage-determination between the
307 gene modules, with the sympathoadrenal-specific modules *catecholamine_production*,
308 *immature_neuronal* and *neurogenesis* being particularly lineage-determined and thus
309 different between clones (63 %, 56 % and 53% significant comparisons respectively).
310 By contrast, modules related to cellular activity (e.g. *interferon_signaling*, *cell_cycle*,
311 *stress_response*) tended to be less lineage-determined (25 %, 28 % and 38 %
312 significant comparisons respectively). Conversely, when we calculated differential
313 module expression between groups of cells from the same clone residing in different
314 sublocations of the same tumor, we found more frequent expression differences for
315 cellular process modules than for sympathoadrenal-specific modules, further indicating
316 that activation of the former is more context-dependent, for example on the location of a
317 cell in the tumor (Fig. S6B).

318 The inter-clone expression differences we observed suggest that
319 sympathoadrenal-specific modules tend to be determined by their clonal origin and will
320 likely be activated at a similar level across cells within a clone. To test this hypothesis,
321 we next determined the intra-clone expression variance of the gene modules. Indeed,
322 we found a tendency for sympathoadrenal-specific modules to exhibit lower intra-clone
323 variance than modules associated with general cellular processes (Fig. 4E, Methods).
324 Together, these two metrics (inter-clone module expression difference and intra-clone
325 variance) suggest stable activation after lineage-determination for sympathoadrenal-
326 specific programs and context-dependent activation for general cellular process
327 programs (Fig. 4F).

328 To exclude the possibility that an uneven distribution of clones within a tumor
329 would confound the observed inter-clonal expression differences, we performed Open-
330 ST⁵²-based spatial transcriptomics to assess the localization of clones. All major cell
331 types identified in scRNA-seq data were also found in the spatial data and we found
332 most of the previously detected modules activated in spots across the tumor mass (Fig.
333 4G, Fig. S7). To identify clones in the spatial transcriptomics data, we extracted clone-
334 specific lineage barcode sequences from scRNA-seq data of the same tumor and
335 analyzed their distribution in space (Fig. 4G, Fig. S8A-C). All six NB cell clones
336 identified in the scRNA-seq data were found in the spatial data. Cells from all these
337 clones were spread across a large area of the tumor section and were well intermixed.
338 Co-occurrence of distinct lineage barcodes in spatial spots confirmed that cells from
339 different clones often occupied the same neighborhoods (Fig. S8D). This example
340 indicates that clones show expression differences even when they occupy the same
341 region of a tumor, emphasizing the importance of the cell of origin for emergence of
342 inter-clonal expression differences.

343 Spatial transcriptomics further allows the analysis of effects of intercellular
344 interactions on tumor cell transcriptional states, such as for example interferon signaling
345 activation close to immune cells⁴. Spatial analysis revealed that TME composition has
346 an overall effect on NB transcriptional states with most NB expression modules
347 correlating positively with TME cell presence, whereas the *ribosomal_genes* module
348 was enriched in the tumor core and inversely associated with presence of most TME
349 cell types (Fig. S9A-B, Methods). We also observed a correlation between the
350 abundance of multiple macrophage populations as well as cytotoxic T-cells and the
351 activation of the *interferon_signaling* module, indicating TME cell types can influence
352 context-dependent modules (Fig. S9C-E). Overall, this analysis confirms that the niche
353 impacts tumor cell transcriptional state, both regarding tumor area (core vs. periphery)
354 as well as presence of certain immune cell populations. However, in contrast to lineage
355 origin, it does not explain the specific activation of most modules.

356 In summary, we performed a direct experimental measurement of tumor state
357 plasticity based on lineage tracing, which revealed different levels of regulation across
358 the spectrum of NB gene modules. These range from cellular process modules that are

359 dynamically regulated in cells regardless of their clonal origin to more stable modules,
360 which are determined by the cell of origin. Among the stably clonally activated modules,
361 we identified the low-risk-associated *catecholamine_production* module, as well as the
362 high-risk-associated *alk-positive immature_neuronal* module. This suggests that such
363 states related to developmental programs and with potential relevance for disease
364 progression are either inherited from the cell of origin or established very early in
365 tumorigenesis.

366

367 *Tumor cell transplantation into an embryonic environment*

368 We next wanted to explore the limits of tumor state stability, and we hypothesized that
369 exposure to a developmental signaling environment might induce reprogramming of
370 lineage-determined states in tumor cells. To test this, we transplanted batches of
371 around 150 lineage-barcoded cells from primary tumors into wildtype zebrafish embryos
372 at the blastula stage. Transplantation of cells from the same clones into multiple
373 embryos enabled their subsequent recovery from different hosts at several timepoints.
374 We profiled cells with scRNA-seq directly after primary tumor dissociation and at two
375 engraftment stages: Tumor cells were FACS-enriched and sequenced together with
376 larval host cells 2 days post transplantation (dpt; early allografts) and whole graft tumors
377 were processed several months post transplantation (mpt; late allografts) to track clonal
378 states over time (Fig. 5A). We processed and transplanted cells from multiple tumors at
379 once, thereby increasing the number of clones observed per experiment and providing
380 sufficient material for the transplantation. Zebrafish NB cells spread throughout the host
381 larvae after transplantation and started re-growing tumors two to six months later (Fig.
382 5B, Fig. S10A-B), most commonly in the orbital cavity of the eye, near the arch-
383 associated complex (heart-proximal) and the superior cervical ganglion, all of which are
384 sites populated by neural crest derivatives or known to be NB primary or metastatic
385 sites^{53,54} (Fig. 5C). Integrated analysis of scRNA-seq data of all primary tumors and
386 graft samples revealed that tumor cells from the grafts, identified by the presence of
387 lineage barcodes, cluster together with the primary tumor cells (Fig. 5D-E, Fig. S10C-E).

388 Thus, NB cells retain a broad adrenergic identity, but may activate different programs
389 within this profile.

390

391 *Exposure to embryonic environment drives re-emergence of expression plasticity*

392 To compare gene expression modules before and after transplantation, we
393 repeated the NMF-based gene module identification on both early and late graft tumor
394 cells (Methods, Fig. S10F). We found that the vast majority of modules represented
395 processes also identified in the primary tumors, indicating that they do not reflect novel
396 transcriptional states induced by engraftment (Fig. S10G). Analysis of module
397 expression across time points and locations revealed a generally high degree of
398 transcriptional variation (Fig. 5F), both in cellular process and in sympathoadrenal-
399 specific modules. We thus continued to use the list of modules identified in primary
400 tumors for all further analyses

401 To better understand state plasticity upon transplantation, we tracked individual
402 clones over time using lineage barcodes. We assigned a graft cell to a primary tumor
403 clone based on shared lineage barcode patterns, allowing us to track 9 clones over
404 three and 18 clones over two time points (Methods, Fig. S11A-B). Overall, grafted NB
405 cells transiently upregulated genes related to interferon signaling and the cell cycle,
406 reflecting a short-term cellular response to the early developmental environment (Fig.
407 6A, Fig. S11C). Interestingly, we observed multiple cases in which lineage-determined
408 modules such as *catecholamine_production* or *immature_neuronal* were down- or
409 upregulated upon transplantation, indicating reprogramming of stable states through
410 transplantation. We then looked at an individual clone in more detail: Clone 6_8 in
411 dataset #1 – derived from a lateral MYCN;LMO1 tumor – which had contributed to
412 seeding of multiple graft tumors in different host fish and tumor locations. Depending on
413 the specific graft tumor, cells from this clone exhibited varying expression levels of
414 programs previously found to be stably expressed within a single primary tumor clone
415 (e.g. *catecholamine_production*), further suggesting that even such stable states can be
416 reprogrammed upon transplantation (Fig. 6B).

417 To substantiate these findings, we next performed a systematic analysis of
418 differential module expression between primary tumor and late allografts using all
419 clones captured in these two timepoints. Unlike the analysis in Fig. 4C which was
420 focused on pairwise comparison of clones within one time point, we now computed the
421 expression differences within individual clones across time points (Fig. 6C-D, Fig.
422 S11D). The difference score reported by this analysis corresponds to the sensitivity of
423 expression programs to transplantation. The modules *interferon_signaling* and
424 *immature_neuronal* stood out by having particularly large expression differences
425 between primary tumor and late allografts. Comparison with the intermediate “early
426 allograft” time point showed that *interferon_signaling* expression is rapidly upregulated
427 after transplantation, whereas the *immature_neuronal* program is markedly altered only
428 in the late allograft tumors, suggesting activation by reprogramming later than two dpt
429 (Fig. S11D).

430 The observed state plasticity prompted us to investigate if module regulation
431 within the graft tumors still largely follows the same general principles as in primary
432 tumors – namely that sympathoadrenal-specific programs are clonally determined within
433 a single tumor (Fig. 4F,G). Therefore, we compared module expression scores between
434 distinct clones that were found in the same graft tumor. This revealed that
435 sympathoadrenal-specific programs remain clonally determined after transplantation
436 (Fig. 6E, Fig. S11E). Similarly, the relative levels of intra-clone variance of the different
437 modules are maintained after transplantation (Fig. 6E, Fig. S11E).

438 Overall, this suggests that the lineage-determined cellular state undergoes
439 reprogramming during a bottle-neck after transplantation, likely within a developmental
440 time window. During this phase, individual transplanted clonal cells can alter their
441 expression of previously stably activated modules. Subsequently, module activation
442 stabilizes and is maintained in the graft tumor, where cells from the same clone
443 consistently express lineage-determined modules and diverge from other clones with
444 distinct primary tumor- and transplantation-signaling exposure histories. This highlights
445 the importance of the developmental signaling environment for NB state determination.

446 Comparing the two measures obtained in these experiments – lineage
447 determination (Fig. 4C) and sensitivity to transplantation into an embryonal environment
448 (Fig. 6C) – revealed different scenarios, exemplified by the following four cases (Fig.
449 6F): The module *immature_neuronal* is highly lineage determined in primary tumors, but
450 also highly sensitive to transplantation, possibly due to sensitivity to a developmental
451 signaling environment. The module *catecholamine_production* is highly lineage
452 determined and displays relatively low sensitivity to transplantation, which indicates
453 overall low plasticity. By contrast, the module *interferon_signaling* has low lineage
454 determination in primary tumors and is highly sensitive to transplantation, in line with a
455 strong dependence on the local environment. Finally, the module *cell_cycle* is neither
456 lineage determined nor sensitive to transplantation and rather represents an
457 autonomous cellular program.

458 Taken together, this analysis shows that tumor modules do not only vary in their
459 lineage determination, but also in the degree to which they are influenced by local
460 developmental signaling and the tumor niche. Importantly, expression states
461 consistently activated or completely absent in the primary tumor clone can be
462 (de)activated upon transplantation, after which they are stable again.

463

464 *Discussion*

465 Making use of well-controlled experimental conditions in the zebrafish model, our
466 systematic analysis of intra- and inter-tumor heterogeneity allowed us to determine the
467 spectrum of *MYCN*-driven NB tumor states consisting of 17 distinct gene expression
468 modules, which represent both lineage-specific and general biological processes that
469 are co-opted by tumor cells.

470 Overall, zebrafish NB cells show strong signs of known *MYCN*-driven gene
471 expression, such as activation of cell cycle genes, ribosomal genes and adrenergic
472 core-regulatory circuit factors. However, expression of the identified programs varied
473 strongly depending on clonal or tissue context, indicating that *MYCN* has differential
474 effects based on the specific tissue context, even in this controlled system. We found
475 that expression of some programs is associated with high-risk disease, namely a

476 ribosomal gene module and the *immature_neuronal* module, which resembles an ALK-
477 activated state^{38,46}. Activating mutations in *ALK* and high levels of *ALK* expression are
478 associated with poor prognosis in human NB. *ALK* mutations are most frequently
479 observed in *MYCN*-amplified NB, where *ALK* prohibits *MYCN*-induced apoptosis and
480 induces transcription of *MYCN* itself, potentially leading to further stabilization of this
481 cooperative state^{18,42,43,55–58}. In the zebrafish model, we observe this state in primary
482 tumors in the AAC (ventral) region. Importantly, tumorigenic transformation in the AAC
483 requires both *MYCN* and *LMO1* transgenes, suggesting that the cell state of origin in
484 this location is less permissive to *MYCN*-driven tumorigenesis, but can specifically give
485 rise to the *immature_neuronal* state. This program is also newly activated in graft
486 tumors after transplantation into an embryonic environment, suggesting that it is
487 determined by lineage factors and highlighting that such an aggressive, but potentially
488 targetable state can emerge via non-genetic mechanisms *in vivo*⁵⁹.

489 In our clonal analysis, we found that modules associated with general cellular
490 processes tend to exhibit high plasticity, while modules associated with
491 sympathoadrenal tissue development - and thus likely the cell state of origin - are
492 mostly lineage-determined. Assessment of clones in space showed that cells from
493 multiple clones are spatially intermixed and their differential states are thus determined
494 by their clonal origin and not their spatial position in the tumor tissue. The distinct clonal
495 states may represent remnants of the cell state of origin or, alternatively, represent
496 divergent states that can be induced by *MYCN* during neoplastic transformation that
497 subsequently stabilize. Their stable activation at an early timepoint in tumorigenesis
498 suggest that the differentiation state and location of the cell at tumorigenic
499 transformation are important factors in determining tumor phenotype. It is conceivable
500 that specific tumor cell states arise from distinct effects of *MYCN*-overexpression on
501 different cell of origin states or that tumor cell state diversity is shaped by the
502 differentiation status of the cell of origin, with less differentiated progenitors giving rise to
503 a more diverse tumor cell population. Addressing these questions will be important for
504 advancing our understanding and the treatment of NB.

505 Transplantations allowed us to determine how tumor cells from the same cell of
506 origin behave in a different environment. Interestingly, some cellular process, but also

507 some sympathoadrenal-specific modules were particularly affected by transplantation.
508 Within fully developed graft tumors, many sympathoadrenal-specific programs are once
509 again clonally determined, indicating that state reprogramming only occurs in a distinct
510 time window, after which cellular states stabilize. Overall, this analysis showed that
511 even the most stable tumor states can be reprogrammed upon exposure to a different
512 signaling environment. Mechanisms of plastic adaptation have also recently been
513 proposed to drive NB metastatic dissemination and emergence of drug tolerant persister
514 cells^{3,60,61}. Thus, understanding the exact cues that lead to state transitions and the cell-
515 intrinsic mechanisms enabling them will be important steps in finding ways to block
516 transitions into more aggressive states.

517 *Limitations*

518 While we did not detect mutations by WES, we cannot rule out that undetected
519 mutations may influence the observed states. However, the observed sustained
520 polyclonality and the high plasticity of tumor states upon transplantation suggests that
521 modules are largely determined in a non-genetic manner. Furthermore, we only
522 consider MYCN-driven NB, an important high-risk disease type, which may however
523 behave differently from other NB subtypes.

524 Here we used spatial transcriptomics to assess the intermixing of clones within
525 tumors. Larger number of samples and more advanced analysis of spatial gene
526 expression patterns would be required to investigate the effect of the TME in more
527 detail.

528 While we can associate tumor states with human disease progression, we do not
529 have a direct measurement of a tumor state driving tumor aggression. Additionally, the
530 set-up of our allografting experiment is not suited for a quantitative assessment of
531 selective advantage of individual tumor states. To address such questions, it would be
532 necessary to analyze tumor state changes upon treatment e.g. in patient-derived
533 xenografts and to identify strategies for inducing targeted state transitions.

534

535

536 *Acknowledgements*

537 We thank Jana Richter for help with cloning and zebrafish line establishment, and Luca
538 Bramè and Jutta Proba for help with allografting experiments. We acknowledge Janita
539 Mintcheva for support and helpful discussions in the establishment of the Open-ST
540 workflow for zebrafish tissues. We also thank Anika Neuschulz and Frank Westermann
541 for helpful feedback on the manuscript. Furthermore, we would like to thank Chris
542 McGinnis and Prof. Zev Gartner's laboratory for generously providing a batch of
543 reagents for MULTI-seq. The transgenic lines *tg[dbh:MYCN,dbh:EGFP]* and
544 *tg[dbh:LMO1,dbh:mCherry]* were a kind gift from Thomas Look (Dana Faber Cancer
545 Institute, Boston, USA). We would further like to acknowledge support from the
546 MDC/BIMSB protein and genomics facilities as well as the zebrafish facilities at
547 MDC/BIMSB and Charité.

548 This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German
549 Research Foundation) within the Collaborative Research Center CRC1588 (project
550 number 493872418), the Forschergruppe FOR5628 (project number 513752256) and
551 the Bruno and Helene Jöster Foundation within the collaborative research grant "Tumor
552 Evolution and Plasticity in Childhood Cancer - TEP-CC". A.S. was supported by an
553 EMBO Long-Term Fellowship (ALTF 972-2022) and a Marie Skłodowska-Curie Actions
554 Postdoctoral Fellowship (ID 101106181 sc-LAB2FATE) from the European Union's
555 Horizon Europe programme.

556

557 *Data and code availability:*

558 The sequencing data generated in this study is deposited in the Gene Expression
559 Omnibus under GSE301974 (scRNA-seq data), GSE301860 (lineage tracing data from
560 scRNA-seq), GSE301869 (spatial transcriptomics data). Custom code will be uploaded
561 and made available through Github at [https://github.com/junker-](https://github.com/junker-lab/neuroblastoma_plasticity)
562 [lab/neuroblastoma_plasticity](https://github.com/junker-lab/neuroblastoma_plasticity).

563

564

565 *Author contributions:*

566 J.P.J., B.S. and N.F. conceived and designed the project. P.O.-C. and N.F. designed,
567 established and tested the dsRedLinrecorder transgene and fish line. J.K. and L.H.
568 performed histology and fluorescence microscopy. N.F. performed bulk and single-cell
569 sequencing experiments with support from J.K. and A.G.. N.F. developed the
570 computational lineage data analysis approach and performed single-cell transcriptome
571 and lineage data analysis. N.F. generated whole exome sequencing data and B.S.
572 performed the data analysis. A.S. performed spatial transcriptomics experiments and
573 analysis with support from M.S.. A.S. developed the experimental and computational
574 approaches for lineage barcode detection in spatial transcriptomics with support from
575 N.F.. J.K., A.G., L.H. and A.I.H.H. established, optimized and performed tumor cell
576 transplantations and supported extraction of allograft tumor cells for sequencing. M.S.
577 was supported by N.R.. J.P.J., B.S., A.I.H.H. and A.G.H. supervised the project. J.P.J.,
578 B.S. and A.I.H.H. guided experiments and J.P.J. and B.S. guided data analysis. N.F.
579 and J.P.J. wrote the manuscript in close interaction with B.S., A.I.H.H. and A.G.H. and
580 with input from all authors.

581

582 *Declaration of interest*

583 A.G.H. is a founder of Econic Biosciences Ltd. B.S. was employed by Econic
584 Biosciences during the conduct of this work.

585

586 **Methods**

587

588 Ethics statement.

589 All experiments were performed in accord with the legal authorities approved license 'G
590 0325/19' and were conducted in accordance with the European Community Council
591 Directive of November 24, 1986 (86/609/EEC).

592

593 Zebrafish lines.

594 Zebrafish (*Danio rerio*) were raised and maintained according to standard protocols at
595 28°C with a 14/10 hour light-dark cycle⁶². Experiments in this study used the zebrafish
596 wild-type strain AB. For lineage tracing, we used the transgenic tg[ubi:zebrabow-M]⁶³,
597 tg[bActin2:dsRed_LinRecorder] and tg[hsp70:dsRed_LinRecorder] (created in this
598 study). Transgenic neuroblastoma lines used were tg[dbh:MYCN, dbh:EGFP] and
599 tg[dbh:LMO1, dbh:mCherry]^{29,30}.

600 Adult zebrafish of random sex were included in an experiment, when they had apparent
601 growth of a dissectible tumor. Fish were euthanized immediately before tumor
602 dissection by hypothermic shock as described by Wallace et al.⁶⁴.

603

604 Histological staining and imaging

605 Specimens were fixed in 4% phosphate-buffered formaldehyde (Labochem, L01-
606 LC64701) for 48 hours, then washed in PBS and decalcified in 0.25M EDTA for 24
607 hours. Sections were mounted on microscope slides, deparaffinized, rehydrated, and
608 stained with hematoxylin (Agilent Dako, CS70030-2) and eosin (Sigma-Aldrich) in
609 Coplin jars following this protocol: 5 min xylene, 10 min xylene, 5 min 95% EtOH, 2x 2
610 min 80% EtOH, 5 min ddH₂O, 6 min hematoxylin, dip into tap water, 6 min running tap
611 water, 3 min eosin, dip into tap water, 1 min 70% EtOH, 1 min 80% EtOH, 1 min 95%
612 EtOH, 2x 1 min 100% EtOH, 3 x 5 min xylene. Sections were mounted with Eukitt quick-
613 hardening medium (Sigma-Aldrich, 03989) and a glass coverslip.

614 Cloning of the lineage tracing recording cassette.

615 The dsRedExpress coding sequence was sequence-optimized for zebrafish. We
616 designed the recording cassette by placing three 23 bp sequences (including PAM)
617 from RFP that had been tested for CRISPR/Cas9 editing before in an array interspersed
618 by 7 bp spacers³¹. Sequences containing multiple restriction sites were placed on either
619 side of the cassette. We further added the 10x capture sequence ¹⁶⁵ downstream of the
620 cassette to enable more efficient capturing of transcripts in the 10x 3' GE assay. Both
621 sequences were synthesized by Twist Bioscience. The dsRed and the recording
622 cassette were then inserted consecutively into the MCS of pME with the NEBBuilder
623 HiFi DNA Assembly (NEB, E2621L), after linearizing the vector with KpnI+HindIII or
624 BamHI+XbaI, respectively. The transgene was then cloned downstream the bActin2- or
625 hsp70-promoter and upstream of a polyA-signal by Gateway LR reaction (Thermo
626 Fisher, 11791020) into a pDest carrying the Tol2-TIRs for insertion of transgenes into
627 the zebrafish genome. Finally, two integration barcodes were inserted into the
628 transgene flanking the recording cassette. A 7-base-pair random sequence with a stable
629 G in the middle flanked by 20-basepair overhangs complementary to the integration site
630 was obtained as a single-stranded DNA oligo (IDT). The plasmid was linearized with
631 EcoRI upstream of the cassette and the barcode oligo was inserted using the
632 NEBBuilder HiFi DNA Assembly. A plasmid library with high barcode diversity was
633 isolated from transformed bacteria. This process was repeated to integrate a second
634 barcode downstream of the recording cassette after linearizing with NruI. The final
635 plasmid library was subjected to Sanger sequencing to confirm a near-complete
636 insertion of the barcode and the presence of barcode-diversity in the library.

637

638 Generation of a lineage tracing line.

639 The plasmid (at 6.25 ng/ μ l) was injected together with Tol2-mRNA (25 ng/ μ l) in 2 nl
640 droplets into the yolk of wildtype zebrafish at the one cell stage. Larvae showing strong
641 and widely spread red fluorescence in the body at 48 hpf to 5 dpf were selected for
642 raising. Adult founders producing large fractions of red-fluorescent offspring were
643 selected. Individual larvae were genotyped by next generation sequencing to identify

644 integration barcodes and thereby number of integrations per larva. Two pairs of one
645 male and one female founder each conferring several integrations with distinct
646 integration barcodes to their offspring were bred to produce two F1-lines. F1 fish were
647 mated with the transgenic NB lines for the lineage tracing experiments shown here.

648

649 Cas9 and sgRNA injections for high-throughput lineage tracing and bulk barcode
650 creation dynamics.

651 For experimental batches 1 and 2 (Fig. S5H), we used a similar approach for
652 sgRNA/Cas9 preparation as described before^{31,32}. For experimental batch 3 Alt-R
653 crRNAs (100 μ M) were ordered from IDT and annealed to the constant tracrRNA (100
654 μ M, IDT) individually by incubation of the mix at 85 °C for five minutes, followed by
655 cooling at room temperature and subsequently on ice. Hybridized gRNAs for different
656 targets were then pooled in multiple batches at the desired ratio. SpCas9 (MDC protein
657 facility) was diluted to 26.8 μ M in Cas9 freezing buffer (20 mM Tris-HCl pH7.5, 600 mM
658 KCl, 20 % glycerol; all nuclease-free). Each gRNA-pool was mixed with an equal
659 volume of diluted spCas9 and guides were loaded onto the protein by incubation at 37
660 °C for five minutes. Batches of loaded ribonucleoprotein complexes (RNPs) were then
661 mixed at the desired ratio, aliquoted and frozen at -80 °C. 2 nl of freshly thawed RNPs
662 were injected into one cell stage offspring of a cross between a transgenic lineage
663 tracing line and tg[dbh:MYCN, dbh:EGFP, dbh:LMO1, dbh:mCherry]. Successfully
664 injected specimens were selected after 48 hours based on loss of pigmentation induced
665 by cleavage of the tyr-CDS by Cas9 with a gRNA included in each mix.

666

667 Sample collection and library preparation for bulk lineage barcode creation dynamics.

668 Zebrafish (tg[bActin2:dsRed_LinRecorder]) were injected with 2 nl RNPs at the one-cell
669 stage. Batches of embryos or larvae were transferred to Eppendorf tubes at selected
670 time points. For RNA-extraction, zebrafish samples were homogenized in TriZol reagent
671 (Thermo Fisher, 15596026) and RNA was extracted following the manufacturer's
672 instructions. Up to 3 μ g of RNA were used as input for reverse transcription with

673 SuperScript IV (Thermo Fisher, 18090010) and a poly-dT primer as used for
674 CELseq2⁶⁶, but with a 22 bp UMI. After second strand synthesis, cDNA from multiple
675 samples was pooled and cleaned up as in the CELseq2 protocol⁶⁶. Each target
676 sequence of interest was enriched by a three-round nested PCR approach with
677 NEBNext High Fidelity Master Mix (NEB, M0541) using target-specific primers, thereby
678 also introducing the overhang sequences needed for sequencing. For DNA-extraction,
679 zebrafish samples were incubated in 50 µl lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM
680 EDTA, 0.3 % Tween-20, 0.3 % Triton-100; all nuclease-free) with 1 mg/ml proteinase K
681 (Invitrogen, 25530049) at 55 °C overnight, followed by 30 minutes at 85 °C to inactivate
682 proteinase K. gDNA was precipitated using isopropanol, washed twice with 70 %
683 ethanol and finally resuspended in 10 mM Tris-HCl pH 8.0. gDNA was directly used as
684 input for a two-round nested PCR approach using gene-specific primers. All libraries
685 were sequenced on Illumina NextSeq2000 200 bp kits (R1 28 cycles, R2 155 cycles).

686

687 Analysis of lineage scar creation dynamics.

688 Sequencing reads from individual libraries were demultiplexed using bcl2fastq
689 (v.2.19.0.316). A custom python script was used to further demultiplex reads from each
690 individual sample based on barcodes introduced during PCR and sequenced on a non-
691 index read. FASTQ files were aligned to references of the target genes using bwa mem
692 (v.0.7.17)⁶⁷. Sequences with less than three reads were removed. Reads were
693 shortened to a sequence identifier 30 bases around the expected Cas9 cut site for
694 endogenous targets and to a 92 base sequence identifier covering all three target genes
695 for dsRedLinRecorder. Sequence identifiers matching the reference sequence were
696 classified as wildtype, while all other sequence identifiers were classified as edited by
697 Cas9. These assignments were used for calculation of the fraction of wildtype UMIs or
698 wildtype reads for RNA-based and DNA-based libraries, respectively.

699

700

701

702 Whole exome sequencing.

703 Genomic DNA from tumor and control tissue (clipped fins of the same fish) was
704 extracted with the DNeasy Blood & Tissue Kit (Qiagen, 69504) and WES libraries were
705 constructed with the SureSelect XT HS2 DNA Reagent Kit (Agilent, G9981A) following
706 manufacturer's instructions with zebrafish-specific probes (SSXT Zebrafish All Exon,
707 Agilent 5190-5450). Libraries were sequenced on the NovaSeq system (R1 150 cycles,
708 R2 150 cycles).

709 FASTQ files were trimmed and aligned to the zebrafish genome GRCz11 using bwa
710 mem and read duplicates were removed. Somatic mutations for a tumor sample taking
711 into account the individual's matched healthy tissue were called using GATK's
712 Mutect2⁶⁸. These variants were used to calculate variant allele frequencies.

713 Segmental copy number variants were called with the CNVkit software⁶⁹. The genome
714 was first split into bins containing an equal number of bait-targets (excluding
715 centromeric and telomeric regions), in which reads were piled up. Pile-ups were
716 compared between tumor and matched control samples. For calling CNVs the in-built
717 circular binary segmentation approach was used^{70,71}.

718

719 Tissue dissociation.

720 Tumors were excised from adult zebrafish, carefully cleaned from non-fluorescent
721 tissues and placed into ice-cold PBS. The tissue was then minced into small pieces.
722 Tissue fragments were pelleted and resuspended in dissociation solution. Primary
723 tumors used for scRNA-seq only were dissociated in 0.01% papain (Sigma-Aldrich,
724 1495005), 0.1% dispase II (Sigma-Aldrich, D4693-1G), 0.01% DNase I (AppliChem
725 GmbH, A3778) and 12.4mM MgSO₄ in calcium- and magnesium-free Hank's balanced
726 salt solution at room temperature for 25 min with trituration through a micropipette tip
727 every 5 minutes. Primary tumors used for transplantations as well as larval samples and
728 allograft tumors were dissociated in 500 µl 30 mg/ml type II Collagenase (Sigma-
729 Aldrich, C2-28) (approx. 700 U/ml), 25 mM HEPES, 10% FCS. The tissue was
730 incubated at 37 °C for 30 minutes with trituration through a micropipette tip every 5

731 minutes. 200 µl 2U/ml dispase II (Sigma-Aldrich, D4693-1G, in 50 mM HEPES, 150 mM
732 NaCl, pH7.4) was added for the last five minutes of dissociation. Following dissociation
733 with either protocol, the opaque cell suspension was pipetted on a 5mL polystyrene
734 Falcon® round-bottom tube with a 30 µm-mesh cell strainer cap (Corning, 352003) pre-
735 filled with 500 µl ice-cold PBS with 1 % BSA and centrifuged at 500 g at 4 °C for 5
736 minutes. Cells were washed once with 500 to 1000 µl ice-cold PBS depending on the
737 cell pellet size. Cells were then pelleted again and resuspended in the desired volume
738 for downstream processing. Live and dead cells were counted manually using a
739 Neubauer counting chamber and Trypan Blue as a stain for dead cells. The dissociation
740 protocol used for each tumor sample is listed in Table S2.

741

742 MULTI-seq labelling.

743 MULTI-seq lipid-modified oligos (LMOs) and co-anchor were kindly gifted by Zev
744 Gartner's lab and later acquired from Sigma Aldrich (LMO001)³⁴. After counting cells in
745 the suspension, an equal number of live cells from each tumor (200 to 500 thousand,
746 depending on material availability) were incubated in 100 µl to 200 µl 200 nM LMO-
747 Barcode-oligo mix in PBS on ice for five minutes. Then the co-anchor was added and
748 mixed into the cells at 200 nM, followed by another 5 minute incubation on ice. The
749 labeling reaction was quenched by addition of 800 µl PBS with 1 % BSA. Cells were
750 pelleted and washed once before pooling in a large volume of PBS with 1 % BSA.
751 Before droplet encapsulation or FACS cells were again passed through a 30 µm
752 strainer.

753

754 Gene expression, MULTI-seq barcode and lineage library preparation.

755 Single cells from individual tumors or samples pooled after MULTI-seq labelling were
756 counted manually using a counting chamber and Trypan Blue as a stain for dead cells.
757 Cells were then used for scRNA-seq library preparation with the 10x 3' GE kit (V3.0 or
758 3.1), aiming for 10,000 cells per library. The gene expression library was prepared
759 following the 10x protocol. For MULTI-seq runs, the small fraction of the cDNA

760 containing most barcode molecules was cleaned up in addition to the large cDNA
761 fraction used for the gene expression library. To this end, the supernatant remaining
762 during cDNA clean-up was incubated with SPRIselect beads (Beckman Coulter
763 B23318, final ratio 3.2X) and isopropanol (final ratio of 1.8X) and cDNA was eluted in
764 EB (Qiagen, 19086) after two ethanol washes. The small cDNA fraction was then used
765 for library preparation using a two-step PCR protocol with primers amplifying the
766 MULTI-seq barcode oligos specifically. For specific amplification of genetic lineage
767 barcodes, we used an approach similar to the one previously described^{31,32}. Briefly, a
768 targeted library was prepared for each lineage tracing target gene using a three-round
769 nested PCR approach with gene-specific primers and 100 ng 10x cDNA as input
770 material. All libraries were sequenced on Illumina NextSeq500, NextSeq2000 or
771 NovaSeq. 10x gene expression libraries and MULTI-seq LMO libraries were sequenced
772 with a minimum Read-2 length of 90 cycles. Lineage tracing libraries were sequenced
773 with a minimum Read-2 length of 120 cycles.

774

775 Sequencing data processing and mapping.

776 All zebrafish single cell sequencing data was demultiplexed using bcl2fastq
777 (v.2.19.0.316). Gene expression data was mapped to a zebrafish transcriptome
778 (GRCz11, Ensembl release 92) extended to include all transgenes present in the fish
779 lines used (*MYCN*, *LMO1*, *EGFP*, *mCherry*, *dsRedLinRecorder*) using Cell Ranger
780 (v.7.0.0). Lineage tracing libraries were further processed using custom pipelines as
781 described below.

782

783 Zebrafish NB transcriptome analysis.

784 Gene expression data was analyzed using Seurat v.4.0.0. Cells with over 10 %
785 mitochondrial transcripts (genes named 'mt-') were removed. Datasets derived from one
786 individual sample using the standard 10x workflow were filtered to only contain cells
787 with at least 250 distinct genes and 500 UMIs (Table S1). Cells from MULTI-seq
788 datasets were filtered more leniently and subsequently assigned to a sample of origin

789 using the classification approach described in the following section. Cells that could not
790 be assigned a sample of origin or that were classified as doublets in this process were
791 removed, thereby ensuring that only Bonafide cells are kept for downstream analysis.
792 The data was then processed using the standard Seurat pipeline with log-normalization,
793 scaling and identification of highly variable genes. The list of highly variable genes was
794 filtered to remove batch-associated genes (genes expressed in 80 % of cells from a
795 dataset and $\log_2FC > 0.2$ compared to cells from all other datasets) and known cell
796 cycle markers⁷² (translated to zebrafish genes using <https://www.flyrnai.org/diopt>) to
797 reduce their impact on dimensionality reduction and clustering. This was followed by
798 PCA and selection of a suitable number of components to use for Louvain clustering
799 and UMAP. Differentially upregulated genes were used to assign a cell type to each
800 cluster (Table S3). (Sub-) cell type assignment was further refined iteratively by sub-
801 clustering all non-NB cells, followed by a further separation of blood / immune cells and
802 other stromal cells (Fig. S1B). cNMF⁷³ was used to identify gene modules active in each
803 cell subset. Modules and differentially upregulated genes were used to assign cell types
804 and specific functions. Marker genes for cell type classification were mainly taken from
805 two previous zebrafish scRNA-seq studies^{37,74}.

806

807 MULTI-seq sample demultiplexing.

808 MULTI-seq libraries were first processed using the deMULTIplex³⁴ pipeline to obtain a
809 count matrix of observations of each sample barcode in each cell. Cells were then
810 assigned to a sample of origin based on three different classification approaches. First,
811 the deMULTIplex R package (<https://github.com/chris-mcginnis-ucsf/MULTI-seq>) was
812 used to classify cells. Second, a manual thresholding approach was used, where cells
813 are assigned to a sample of origin, if the associated barcode is found in them at a high
814 frequency. Cells that passed the threshold for multiple barcodes were labelled as
815 doublets, whereas cells below the thresholds for all barcodes were labelled negative. In
816 the third approach, cells were classified according to the dominant barcode found in
817 them. Cells were assigned to a sample, if the associated barcode had at least 2.3 times
818 as many molecules as the second-most highly detected barcode and if the molecules

819 from the dominant barcode made up over 40 % of the total molecules found. Cells that
820 did not fulfill these criteria were classified as doublets or negative if their total amount of
821 barcode molecules was above or below the mean of total barcode molecules across all
822 cells, respectively. To make sure cell-sample-assignment was stringent, a consensus of
823 the three approaches was taken. Only cells with a matching assignment in at least two
824 of the classification approaches were classified according to this label. Cells with
825 ambiguous classification across the three approaches were labelled doublets. Doublets
826 and negative cells were removed from further analysis.

827

828 Identification of gene expression modules in NB cells and identification of consensus
829 modules.

830 To capture both intra- and inter-tumor variation in gene expression, we performed NMF
831 on NB cells on three levels: individual tumor samples, individual MULTI-seq datasets
832 and the entire dataset. Modules from all individual tumor samples or MULTI-seq runs
833 were aggregated into recurring consensus modules following approaches similar to
834 those described in^{4,5,38}. Count matrices and highly variable genes lists as determined in
835 Seurat for NB cells from a) an individual tumor sample, b) an individual MULTI-seq run
836 or c) all samples were passed as input to cNMF⁴¹. We ran cNMF with output module
837 numbers (k) from 5 to a) 25 for individual samples and MULTI-seq runs or b) 50 for the
838 entire NB cell dataset. We ran cNMF with 200 iterations for each sample, from which
839 the algorithm builds a consensus result and a measure of stability for the results
840 obtained for a given k. A suitable number of modules (k) displaying good stability across
841 NMF-iterations was chosen for each sample, mostly in the range of 5 to 15 modules per
842 sample or MULTI-seq run.

843 For consensus module generation, all modules from individual tumor samples (or all
844 MULTI-seq runs) were compared to each other in terms of their gene content using
845 Pearson correlation across all genes with a positive z-scored contribution to one of the
846 modules. (Figure S3A). Modules with a Pearson's R of at least 0.1 with at least two
847 other modules were selected. Modules that passed this filter were clustered using
848 hdbscan from the dbscan package⁷⁵ and a consensus module was generated for each

849 cluster by keeping genes found in over 25 % of all modules in that group. Each
850 consensus module was named according to function and modules were further merged,
851 if they were assigned the same function as well as modules with strong overlap in gene
852 content. Finally, ambiguous genes that were found in multiple modules were removed:
853 For each module, a gene that is found in a more prominent position in another modules
854 is flagged. One such duplicate was allowed, but all duplicates from the second onwards
855 were removed. The final list of modules was compiled by adding MULTI-seq derived
856 modules to those derived from individual tumors, followed by modules derived from
857 whole dataset NMF (Figure S3B). Modules that overlapped strongly with existing
858 modules as well as modules that were only very spuriously expressed or could not be
859 assigned a function were removed (Table S5).

860

861 Gene module expression scoring.

862 Two different approaches were used for gene module expression scoring. To classify
863 cells into those that express a module and those that do not and to thus be able to
864 determine the fraction of cells expressing a module in a given population, we scored the
865 expression following the approach published by Barkley et al.⁴. Briefly, the average
866 centered expression of the module and 1000 random gene lists was calculated. The
867 expression score for the module of interest was defined as $-\log_{10}(\text{fraction of random}$
868 $\text{modules higher than module of interest})$ and rescaled linearly to [0,1]. Here, a score
869 higher than 0.5 means that the module of interest scored higher than half of the random
870 gene sets and this value was used as a cut-off to determine whether a module is
871 expressed in a cell. To derive expression scores that can be used for differential module
872 expression calculations, scores were calculated using AUCCell⁷⁶. AUCCell uses the “Area
873 Under the Curve” (AUC) to calculate whether a given list of genes is enriched in the
874 expressed genes of a cell. The AUCCell score is relative measure of gene module
875 expression and can thus be used to compare module expression levels between cells.
876 Scores were calculated in NB cells from each tumor sample or early allograft dataset
877 separately.

878

879 Translation of zebrafish to human orthologs and vice versa.

880 Zebrafish gene modules were input into DIOPT (<https://www.flyrnai.org/diopt>) to get a
881 list of orthologs. Only orthologs with the following metrics were kept: rank = high OR
882 rank = moderate, best score = yes, DIOPT-score ≥ 10 (max = 19). Mitochondrial genes
883 were manually translated using the same DIOPT criteria to make sure that current gene
884 name versions are used. Human modules were similarly translated using DIOPT
885 keeping matches with rank = high. Importantly, all comparisons between zebrafish and
886 human gene modules were carried out in 'human gene space', as assigning a single
887 human gene match to a zebrafish gene is more reliable than vice versa, due to the
888 genome duplication in teleosts.

889

890 Human cancer gene expression modules.

891 Published gene expression modules were retrieved from the indicated sources. In Fig.
892 2D and Fig. S4A, the signature 'MYCN_upregulated' corresponds to the signature
893 determined as constitutively upregulated MYCN-targets across cell cycle phases in Ryl
894 et al⁴⁴. The PRC2-targets signature consists of the intersection of the 300 top EZH2-
895 bound genes in a ChIP-seq assay from two human NB cell lines as reported in Chen et
896 al⁴⁵.

897

898 Human NB scRNA-seq analysis.

899 FASTQ files from scRNA-seq of sixteen primary NBs from Dong et al.²⁵ were
900 downloaded from GEO (GSE137804) and mapped to the human reference genome
901 (GRCh38, Ensembl release 98) with Cell Ranger (v.4.0.0). Doublets were identified
902 using Scrublet⁷⁷ and removed in addition to cells with less than 250 genes or 600 UMIs
903 detected in the following analysis performed with Seurat v.4.0.0. Each dataset was
904 downsampled to contain a maximum of 7000 cells. Then, all samples were integrated
905 using the Seurat-integrated reciprocal PCA approach⁷⁷, followed by Louvain clustering
906 and dimensionality reduction for data visualization on a UMAP. Gene module
907 expression scores were calculated on non-integrated data for each tumor dataset

908 separately after single-cell data was normalized and scaled following the Seurat
909 workflow. Module expression was calculated as described above using the approach
910 described by Barkley et al.⁴ to obtain an expression score between 0 and 1.

911

912 Human bulk RNA-seq analysis.

913 Bulk gene expression data for 498 NB samples from the SEQC cohort was retrieved
914 from GEO (GSE49711) as $\log_2(1 + \text{FPKM})$. Genes expressed in less than four samples
915 were removed. Gene expression data from the TARGET-NBL-cohort was retrieved from
916 the GDC data portal (<https://portal.gdc.cancer.gov/>) as STAR gene counts. Information
917 on *ALK* mutational status was retrieved from Brady et al.⁴⁸. Entrez gene IDs were
918 translated to gene symbols. If this introduced duplicated gene symbols, the one with the
919 higher variance was kept. Genes with less than 10 counts in less than five samples
920 were removed. Counts were then normalized with a variance standardized
921 transformation (VST) as implemented in DESeq2 (v.1.30.1)⁷⁹.

922 In all datasets, expression scores were calculated following an approach implemented
923 by Decoene et al.⁸⁰, similar to the approach developed for single cell data by Tirosh et
924 al.⁸¹. To test for significance, pairwise Wilcoxon rank sum tests between expression
925 scores in groups were carried out and significance was adjusted for multiple
926 comparisons using Bonferroni correction.

927

928 Mapping and filtering of lineage barcode libraries.

929 The lineage tracing libraries for endogenous targets were processed as previously
930 described³². Briefly fastqs were aligned to individual references of the endogenous
931 targets using bwa mem (v.0.7.17). Reads associated with a valid cell barcode present in
932 the transcriptome library were kept. Subsequently, scar sequences were filtered to
933 remove PCR and sequencing errors or sequences arising from doublets, following the
934 assumption that there can be at most two distinct alleles of an endogenous target within
935 a cell. Cells, in which one or two sequences made up 80 % or more of the gene-specific
936 transcripts were kept and only these top sequences were kept.

937 The lineage tracing libraries for transgenic targets were processed similar to our
938 previously described approach³¹. Briefly, sequencing reads were aligned to individual
939 transgene references using bwa mem and only reads associated with a valid cell
940 barcode in the transcriptome data were kept. Scar sequences with only one read were
941 removed. Following this, for each combination of cell barcode and UMI, only the scar
942 sequence with the highest number of reads was kept. Furthermore, sequences derived
943 from sequencing errors were reduced by comparing sequences found within a cell and
944 removing those that had a low Hamming distance to others and a comparatively low
945 read number. Finally, scar sequences with a relatively low number of reads (determined
946 by distribution of reads for all scar sequences that passed previous filtering steps) were
947 removed.

948 Sequences derived from the dsRedLinRecorder were subsequently split by integration
949 ID. Only sequences carrying a valid integration ID barcode were kept. These were
950 determined as sequences that contain a G in the middle position and have a
951 considerably higher number of reads than invalid sequences. As a given transgene
952 integration in an individual cell can only carry a single scar sequence, ambiguous
953 sequences and cells were removed. Only cells, in which a single scar sequence
954 contributed to 60 % or more of the detected transcripts, were kept. In these cells, only
955 the sequence with the highest number of UMIs was kept. Cells and sequences that
956 passed these filters were used as input for the clone calling pipeline.

957

958 Clone calling.

959 Clone calling is illustrated in Fig. S5C-G and was performed for each fish individually, as
960 the possibility that scar sequences are created in multiple fish hampers a joint analysis.
961 Clone determination starts out with a separate analysis of each endogenous target
962 gene, similar to our previously developed approach³². Briefly, the sequences were
963 shortened to a sequence-ID of 30 bases around the CRISPR target site. Sequence-IDs
964 that were only observed once across all cells were removed. First, only cells with two
965 distinct alleles (one being wildtype is allowed) were kept. For each sequence-ID, the
966 fraction of the total observations of this ID that occur in cells together with a given other

967 sequence-ID is determined. If at least 80 % of observations of one or both sequence-
968 IDs were found in a specific combination, this combination is kept. If possible, a
969 hierarchy of a sequence-ID that was created first ('parent scar') and one that was
970 created later ('child scar'), was determined on the fractions of co-occurrence.
971 Subsequently, cells with only a single sequence ID are assigned to a group defined by a
972 combination of sequence IDs, provided that this single ID could be unambiguously
973 matched to that combination. Finally, cells that have multiple UMIs of wildtype sequence
974 IDs only are labelled as wildtype cells. Cells and sequence-ID combinations that passed
975 these filters are passed on to clone calling based on all target genes.

976 For final clone calling, information from all targets is merged, with input for transgenic
977 targets directly taken from scar filtering step. Here, each endogenous target is
978 represented with one joint combination of two sequence-IDs. Each transgene
979 integration (as distinguished by integration ID barcode sequence) is input as an
980 individual target gene. Wildtype sequences are excluded. Each contributing sequence
981 (or sequence combination) is hereafter called a seq-ID. The overlap in associated cell
982 barcodes is calculated (Jaccard index) for each seq-ID pair and a threshold of 0.3 was
983 determined to derive a binary adjacency matrix for all seq-IDs. This is used as input for
984 an undirected graph, serving as a basis for clustering of seq-IDs. Cell barcodes and
985 seq-IDs are aggregated by cluster, but this often leaves several distinct clusters with
986 overlap in cell barcodes. Therefore, two overlap fractions are calculated for each cluster
987 pair: the two clusters' cell barcode intersection size divided by the total cell barcode
988 count of either one of the clusters. The higher of the two values is kept. Based on this
989 adjacency matrix, clusters are flagged for merging using an overlap threshold, set to 0.8
990 by default (i.e. 80 % of cells associated with one cluster are also associated with the
991 other cluster). A cluster is flagged as ambiguous, if it overlaps with multiple other
992 clusters, but those clusters share barcodes with each other only below a secondary
993 threshold (default 0.6). Such ambiguity can arise from two scenarios: a) a lineage
994 barcode (or combination) was created multiple times in independent events or b) a
995 cluster is defined by lineage barcode(s) ('parent') that were created early and overlap
996 with multiple clusters represented by lineage barcode(s) that were created later ('child').
997 To account for the latter case and to avoid removal of many cells that only carry an

998 early lineage barcode, larger clusters can optionally be treated as a ‘parent’ and smaller
999 ‘child’ clusters can be merged into it. This option leads to the establishment of lower
1000 resolution clusters as used for analysis in Fig. 6 (Fig. S10A). If this option is not
1001 activated, ambiguous clusters are removed, e.g. two independent ‘child clusters’ would
1002 be kept, while the ‘parent cluster’ overlapping with both of them is removed, leading to
1003 higher resolution clustering (as used for analysis in Fig. 3 and 4). Remaining clusters
1004 flagged for merging are merged. Cell barcode overlap between these merged clusters is
1005 again determined by Jaccard similarity and if new clusters with significant barcode
1006 overlap (default Jaccard index of 0.3) have emerged, these are marked as ambiguous
1007 and removed. The remaining clusters are the final clones. Finally, cell barcodes that
1008 were assigned to multiple clones are removed.

1009 Once clones had been determined, a seq-ID is classified as being clone-specific, if 90
1010 % of cells it was observed in came from one specific clone in a given experiment.
1011 Clone-specific seq-IDs were later used as identifiers to match allograft-cells to primary
1012 tumor clones.

1013

1014 Differential module expression analysis.

1015 Differential module expression analysis was performed in a pairwise manner between
1016 cells from two different groups. In the analysis of primary tumors, cells were grouped
1017 according to clone and tumor (sub-)sample to allow for comparison of clones within one
1018 tumor location and of cells from one clone in different sub-samples. In the analysis of
1019 allograft tumor cells, cells from the early allograft were grouped by clone only, whereas
1020 cells from the late allograft tumors were again grouped by clone and (sub-)sample to
1021 enable comparison between clones across time in one late graft tumor or between
1022 several graft tumors. For all comparisons, only groups of at least 10 cells were
1023 considered, where the two combined groups contained at least 25 cells. The differential
1024 expression score for each module was calculated based on module AUCell expression
1025 scores using a Wilcoxon rank sum test (as implemented in the Seurat function
1026 FindMarkers). To assess significance, group assignments of the tested cells were
1027 randomly shuffled 1000 times, while preserving group sizes, and the differential

1028 expression test was repeated for each shuffle. All differential expression values were
1029 ranked and significance p was determined as the rank of the test group of interest
1030 relative to the 1000 permutations and the p-value was calculated as this rank divided by
1031 1,000. This means that $p < 0.05$ is equivalent to the test group of interest exceeding 95
1032 % of random outcomes. In order to make results comparable between different
1033 modules, the differential expression score was normalized to the overall expression
1034 level of the module. To this end, the differential expression score for the comparison of
1035 interest was divided by the mean differential expression score of all random
1036 permutations for that module. Only results with $p < 0.05$ were considered. Furthermore,
1037 comparisons, in which a given module was expressed in less than 5 % of cells in both
1038 clones were removed.

1039

1040 Calculation of module expression variance.

1041 In order to obtain module expression variance measures that are comparable between
1042 different modules, we used the expression-standardized variance values generated by
1043 Seurat's FindVariableFeatures function (selection.method = 'vst') (Seurat v.4.0.0). First,
1044 counts for all genes in a module were summed to get raw module expression scores
1045 that were added to the gene count matrix. The counts matrix containing genes and gene
1046 module counts was then log-normalized using Seurat's NormalizeData function. Log-
1047 normalized expression values were used as input for the FindVariableFeatures function,
1048 which fits the mean-variance relationship across genes and rescales observed
1049 variances by the expected variance at a given mean expression level. Expression-
1050 standardized variance was calculated per clone or other group of cells of interest.
1051 Groups of less than 30 cells were removed as well as groups of cells, which only had
1052 module expression in less than 5 % of cells.

1053

1054

1055

1056 Tissue processing, library preparation and data preprocessing for spatial
1057 transcriptomics of zebrafish tumors

1058 Open-ST spatial transcriptomics and sequencing: Dissected entire tumors or pieces of
1059 tumor tissue (if another piece was used for scRNA-seq) were embedded in optimal
1060 cutting temperature compound (O.C.T., Tissue-Tek, 4583) in plastic cryomolds. The
1061 filled mold was frozen by placing it on a flat metal surface cooled down with dry ice.
1062 Frozen samples were subsequently stored at -80 °C. Cryosections were then cut at 10
1063 µm thickness and mounted on Open-ST capture areas. Tissue handling and spatial
1064 barcoding were performed following the Open-ST protocol⁵². Brightfield images of H&E-
1065 stained sections were acquired using a Keyence BZ-X700 to assist with downstream
1066 image registration and background removal. Following cDNA elution, the whole
1067 transcriptome library was prepared following the Open-ST protocol. Gene-specific
1068 libraries were generated using a two-round nested PCR approach using target-specific
1069 primers. All final products were size-selected on a BluePippin HT system (Sage
1070 Science) Spatial transcriptomics libraries were sequenced on an Illumina NextSeq 2000
1071 using a 200-cycle kit (R1: 37 cycles, R2: 191 cycles).

1072 Alignment and generation of count matrices: Raw spatial transcriptomics data were
1073 processed and aligned using SpaceMake⁸², which produced a gene-by-spot count
1074 matrix from sequencing reads. Individual tiles were stitched, and expression was
1075 aggregated on a hexagonal grid with 5 µm diameter bins using custom Python code.

1076 Image-based spot filtering: The brightfield image of the tissue section was used to
1077 create two images: one inverted in Fiji for alignment and one thresholded (black and
1078 white) for spatial filtering. Using a custom Python code, the first image was aligned to a
1079 synthetic transcriptomic image rendered by showing the number of spots aggregated
1080 into each hexagon of the grid. Manual landmarks were selected on both optical and
1081 spatial transcriptomic images. An affine or homography transformation was computed
1082 using OpenCV and applied to the binarized version of the optical image. Only spots
1083 falling within foreground tissue regions were retained for downstream analysis.

1084 Transcriptomic filtering and normalization: Following image-based subsetting, spatial
1085 transcriptomic data were filtered to remove spots with fewer than 5 detected genes and

1086 genes expressed in fewer than 5 spots. Total counts per spot were normalized to
1087 10,000 and log-transformed using log_{1p}. Highly variable genes were selected using
1088 Scanpy's variance-based method, retaining the top 2,500 genes.

1089

1090 Spatial analysis of modules

1091 Cell type label transfer: Single-cell RNA-seq reference data were integrated on shared
1092 highly variable genes using Scanpy's implementation of Harmony⁸³, and spatial spots
1093 were projected into a shared PCA space. For each spot, a k-nearest neighbor model
1094 was used to infer a probability distribution over reference cell types, resulting in per-spot
1095 soft cell type scores.

1096 Module scoring: Gene modules were quantified by computing the fraction of total
1097 expression per spot attributable to each module. For each spot, expression of all valid
1098 module genes was summed and divided by the total spot-wise expression.

1099 Spatial correlation: To assess local co-variation of cell type or module scores, values
1100 were smoothed across spatial neighborhoods defined by a fixed Euclidean distance
1101 using a cKDTree search. Pearson correlations were then computed between smoothed
1102 scores across spots.

1103 Proximity to tissue boundaries: To evaluate spatial positioning relative to tissue borders,
1104 connected component analysis was used to define tissue regions. We manually
1105 retained the three largest regions of the tumors. A Euclidean distance transform was
1106 applied to compute each spot's distance to the nearest external boundary, and
1107 Spearman correlations were computed between distance and module scores within
1108 each region.

1109

1110 Spatial clonal analysis

1111 Spatial gene-specific libraries were used for spatial clonal analysis: Scar gene barcodes
1112 were extracted from read 1 and assigned to spatial coordinates based on their title and
1113 lane identifiers. Read 2 was aligned to reference scar genes using bwa mem, and a
1114 spatial barcode-scar sequence count matrix was constructed. Barcode matrices were

1115 filtered using the same imaging-derived tissue mask applied previously. Clone identities
1116 inferred from matched scRNA-seq data were transferred to spatial spots by aligning
1117 shared scar sequences (seqIDs) across modalities. Clone presence at each spot was
1118 binarized by thresholding to ≥ 1 supporting read per scar gene.

1119

1120 Allogeneic transplantations into zebrafish embryos.

1121 The pool of dissociated tumor cells from multiple tumors was divided between the
1122 workflow for scRNA-seq and allogeneic transplantation. For transplantation, the cells
1123 from multiple tumors were counted and mixed. Cells were centrifuged through a 20 μm
1124 mesh filter at 500 g at 4 °C for 5 minutes. The supernatant was removed almost entirely
1125 and cells were resuspended in a tiny volume of PBS to keep the suspension very
1126 dense. Glass pipettes with a 20 μm outer diameter (BioMedical Instruments) connected
1127 to an air-pressure injector (IM-400) were used to inject 100 - 150 cells into each embryo
1128 4 h after fertilization. Embryos were allowed to recover in E3 medium (5mM NaCl,
1129 0.17mM KCl, 0.33mM CaCl₂, 0.33 mM MgSO₄, pH 7.4) supplemented with 1%
1130 penicillin/streptomycin for 1 h at 28 °C before manually sorting for transplantation
1131 success on a M165 FC stereomicroscope (Leica Microsystems).

1132

1133 Isolation of tumor cells from larvae via fluorescence activated cell sorting (FACS)

1134 Larvae at 2 dpt (days post transplantation) were collected in batches of 15 to 20 and
1135 placed on ice for 10 minutes. Larvae were then washed once in ice-cold HBSS and
1136 dissociated with collagenase II and dispase as described above. If multiple larval
1137 batches were processed at once, cells were then labelled with MULTI-seq barcodes as
1138 described above. After quenching of the labelling reaction and thorough washing with
1139 PBS with 1 % BSA, cells were resuspended in PBS with 0.05 % BSA for FACS sorting
1140 (BD Biosciences FACSAria III). Cells were selected by first gating for GFP-
1141 fluorescence, followed by side- and forward-scatter gating. Cells were sorted into a
1142 cooled 1.5 ml eppendorf tube pre-filled and coated overnight with 500 μl PBS with 2 %
1143 fetal calf serum. Cells were sorted with a flow rate below 4 and an event-rate below

1144 2000 events per second. Every 20 minutes, the sort was halted and the source cell
1145 suspension was vortexed. The receiving tube was closed and inverted before placing it
1146 back into the sorter. Sorting was done until either a) 50000 cells had been sorted into
1147 the tube, b) at least 20000 cells had been sorted and the cell suspension was close to
1148 running out, c) sorting time reached two hours. In the two latter scenarios, the gate
1149 sorting for fluorescence was inactivated, so that all live cells were sorted until 40000
1150 total cells had been sorted into the tube. After sorting the receiving tube was inverted a
1151 few times before centrifugation at 4 °C, 1000 g for 5 minutes. Most of the supernatant
1152 was removed, leaving an estimated 10 µl around the cell pellet. The cells were
1153 resuspended after addition of 50 µl PBS and subsequently counted in a counting
1154 chamber prior to single cell encapsulation.

1155

1156 Assignment of allograft cells to primary tumor clone.

1157 Lineage tracing target-specific libraries were processed as described for primary
1158 tumors. Sequences from endogenous target genes were shortened to a sequence-ID of
1159 30 bases around the CRISPR target site. Cells with two distinct sequence-IDs for a
1160 given target were selected. For transgenic targets, valid integration ID barcodes were
1161 extracted and other sequences were removed. Thereafter, sequences were filtered, so
1162 that each cell retained at most one sequence per integration ID as described for the
1163 primary tumors. Combined sequence IDs from the endogenous targets as well as
1164 individual sequence-IDs from each transgene integration were used to match allograft
1165 cells to a primary tumor clone defined by one or more of these sequence-IDs. Here, we
1166 used primary tumor clones called at slightly lower resolution to increase the number of
1167 allograft cells that could be assigned to a clone (Fig. S11A-B). Ambiguous assignments
1168 of one graft cell to multiple primary clones were removed.

1169

1170 References

- 1171 1. Marin-Bejar, O. *et al.* Evolutionary predictability of genetic versus nongenetic resistance to
1172 anticancer drugs in melanoma. *Cancer Cell* **39**, 1135–1149.e8 (2021).
- 1173 2. Moorman, A. *et al.* Progressive plasticity during colorectal cancer metastasis. *Nature* **637**,
1174 947–954 (2025).
- 1175 3. Roux, C. *et al.* Dynamic Plasticity Systems Direct Early Adaptation to Treatment in
1176 Neuroblastoma. Preprint at <https://doi.org/10.1101/2023.12.07.570359> (2023).
- 1177 4. Barkley, D. *et al.* Cancer cell states recur across tumor types and form specific interactions
1178 with the tumor microenvironment. *Nat. Genet.* **54**, 1192–1201 (2022).
- 1179 5. Gavish, A. *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand
1180 tumours. *Nature* **618**, 598–606 (2023).
- 1181 6. Ciriello, G. *et al.* Cancer Evolution: A Multifaceted Affair. *Cancer Discov.* **14**, 36–48 (2024).
- 1182 7. Moiso, E. *et al.* Developmental Deconvolution for Classification of Cancer Origin. *Cancer*
1183 *Discov.* **12**, 2566–2585 (2022).
- 1184 8. Yang, D. *et al.* Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor
1185 evolution. *Cell* **185**, 1905–1923.e25 (2022).
- 1186 9. Quinn, J. J. *et al.* Single-cell lineages reveal the rates, routes, and drivers of metastasis in
1187 cancer xenografts. *Science* **371**, eabc1944 (2021).
- 1188 10. Simeonov, K. P. *et al.* Single-cell lineage tracing of metastatic cancer reveals selection of
1189 hybrid EMT states. *Cancer Cell* **39**, 1150–1162.e9 (2021).
- 1190 11. Filbin, M. & Monje, M. Developmental origins and emerging therapeutic opportunities for
1191 childhood cancer. *Nat. Med.* **25**, 367–376 (2019).
- 1192 12. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284
1193 (2013).
- 1194 13. Schulte, J. H. & Eggert, A. Neuroblastoma. *Crit. Rev. Oncog.* **20**, 245–270 (2015).
- 1195 14. Körber, V. *et al.* Neuroblastoma arises in early fetal development and its evolutionary
1196 duration predicts outcome. *Nat. Genet.* **55**, 619–630 (2023).
- 1197 15. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *The Lancet* **369**,
1198 2106–2120 (2007).
- 1199 16. Westermann, F. *et al.* Distinct transcriptional MYCN/c-MYC activities are associated with
1200 spontaneous regression or malignant progression in neuroblastomas. *Genome Biol.* **9**,
1201 (2008).
- 1202 17. Rickman, D. S., Schulte, J. H. & Eilers, M. The Expanding World of N-MYC–Driven Tumors.
1203 *Cancer Discov.* **8**, 150–163 (2018).
- 1204 18. Zhu, S. *et al.* Activated ALK Collaborates with MYCN in Neuroblastoma Pathogenesis. *Cancer*
1205 *Cell* **21**, 362–373 (2012).
- 1206 19. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**,
1207 5075–5082 (2017).

- 1208 20. Van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated
1209 differentiation states. *Nat. Genet.* **49**, 1261–1266 (2017).
- 1210 21. Van Groningen, T. *et al.* A NOTCH feed-forward loop drives reprogramming from adrenergic
1211 to mesenchymal state in neuroblastoma. *Nat. Commun.* **10**, 1530 (2019).
- 1212 22. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional
1213 circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).
- 1214 23. Wang, L. *et al.* ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic
1215 neuroblastoma core regulatory circuitry. *Nat. Commun.* **10**, (2019).
- 1216 24. Jansky, S. *et al.* Single-cell transcriptomic analyses provide insights into the developmental
1217 origins of neuroblastoma. *Nat. Genet.* **53**, 683–693 (2021).
- 1218 25. Dong, R. *et al.* Single-Cell Characterization of Malignant Phenotypes and Developmental
1219 Trajectories of Adrenal Neuroblastoma. *Cancer Cell* **38**, 716-733.e6 (2020).
- 1220 26. Kildisiute, G. *et al.* Tumor to normal single-cell mRNA comparisons reveal a pan-
1221 neuroblastoma cancer cell. *Sci. Adv.* (2021).
- 1222 27. Olsen, T. K. *et al.* Joint single-cell genetic and transcriptomic analysis reveal pre-malignant
1223 SCP-like subclones in human neuroblastoma. *Mol. Cancer* **23**, 180 (2024).
- 1224 28. Yuan, X. *et al.* Single-cell profiling of peripheral neuroblastic tumors identifies an aggressive
1225 transitional state that bridges an adrenergic-mesenchymal trajectory. *Cell Rep.* **41**, 111455
1226 (2022).
- 1227 29. Tao, T. *et al.* The pre-rRNA processing factor DEF is rate limiting for the pathogenesis of
1228 MYCN-driven neuroblastoma. *Oncogene* **36**, 3852–3867 (2017).
- 1229 30. Zhu, S. *et al.* LMO1 Synergizes with MYCN to Promote Neuroblastoma Initiation and
1230 Metastasis. *Cancer Cell* **32**, 310-323.e5 (2017).
- 1231 31. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR–
1232 Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
- 1233 32. Mitic, N. *et al.* Dissecting the spatiotemporal diversity of adult neural stem cells. *Mol. Syst.*
1234 *Biol.* **20**, 321–337 (2024).
- 1235 33. Chan, M. M. *et al.* Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82
1236 (2019).
- 1237 34. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using
1238 lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
- 1239 35. Pei, D. *et al.* Distinct Neuroblastoma-associated Alterations of PHOX2B Impair Sympathetic
1240 Neuronal Differentiation in Zebrafish Models. *PLoS Genet.* **9**, e1003533 (2013).
- 1241 36. Finegold, M. J., Triche, T. J. & Askin, F. B. Neuroblastoma and the differential diagnosis of
1242 small-, round-, blue-cell tumors. *Hum. Pathol.* **14**, 569–595 (1983).
- 1243 37. Tang, Q. *et al.* Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at
1244 single-cell resolution using RNA sequencing. *J. Exp. Med.* **214**, 2875–2887 (2017).

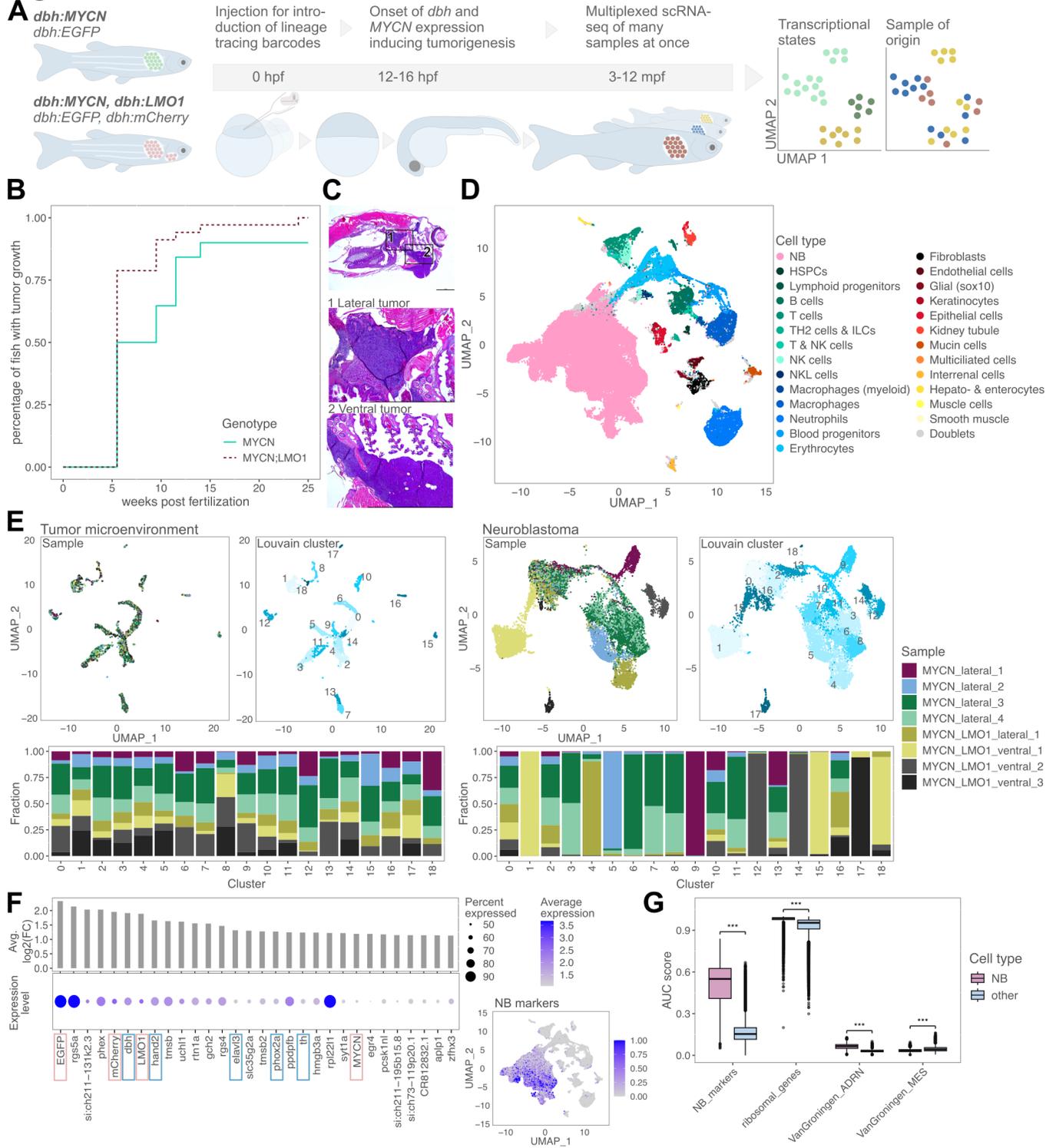
- 1245 38. Stöber, M. C. *et al.* Intercellular extrachromosomal DNA copy-number heterogeneity drives
1246 neuroblastoma cell state diversity. *Cell Rep.* **43**, 114711 (2024).
- 1247 39. Hald, Ø. H. *et al.* Inhibitors of ribosome biogenesis repress the growth of MYCN-amplified
1248 neuroblastoma. *Oncogene* **38**, 2800–2813 (2019).
- 1249 40. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for
1250 Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
- 1251 41. Duijkers, F. A. M. *et al.* High Anaplastic Lymphoma Kinase Immunohistochemical Staining in
1252 Neuroblastoma and Ganglioneuroblastoma Is an Independent Predictor of Poor Outcome.
1253 *Am. J. Pathol.* **180**, 1223–1231 (2012).
- 1254 42. Schulte, J. H. *et al.* High ALK Receptor Tyrosine Kinase Expression Supersedes ALK Mutation
1255 as a Determining Factor of an Unfavorable Phenotype in Primary Neuroblastoma. *Clin.*
1256 *Cancer Res.* **17**, 5082–5092 (2011).
- 1257 43. Mossé, Y. P. *et al.* Identification of ALK as a major familial neuroblastoma predisposition
1258 gene. *Nature* **455**, 930–935 (2008).
- 1259 44. Ryl, T. *et al.* Cell-Cycle Position of Single MYC-Driven Cancer Cells Dictates Their
1260 Susceptibility to a Chemotherapeutic Drug. *Cell Syst.* **5**, 237-250.e8 (2017).
- 1261 45. Chen, L. *et al.* CRISPR-Cas9 screen reveals a MYCN-amplified neuroblastoma dependency on
1262 EZH2. *J. Clin. Invest.* **128**, 446–462 (2017).
- 1263 46. Chapple, R. H. *et al.* An integrated single-cell RNA-seq map of human neuroblastoma
1264 tumors and preclinical models uncovers divergent mesenchymal-like gene expression
1265 programs. *Genome Biol.* **25**, 161 (2024).
- 1266 47. Sengupta, S. *et al.* Mesenchymal and adrenergic cell lineage states in neuroblastoma
1267 possess distinct immunogenic phenotypes. *Nat. Cancer* **3**, 1228–1246 (2022).
- 1268 48. Brady, S. W. *et al.* Pan-neuroblastoma analysis reveals age- and signature-associated driver
1269 alterations. *Nat. Commun.* **11**, 5183 (2020).
- 1270 49. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy,
1271 reproducibility and information content by the Sequencing Quality Control Consortium.
1272 *Nat. Biotechnol.* **32**, 903–914 (2014).
- 1273 50. Heath, A. P. *et al.* The NCI Genomic Data Commons. *Nat. Genet.* **53**, 257–262 (2021).
- 1274 51. Sur, A. *et al.* Single-cell analysis of shared signatures and transcriptional diversity during
1275 zebrafish development. *Dev. Cell* **58**, 3028-3047.e12 (2023).
- 1276 52. Schott, M. *et al.* Open-ST: High-resolution spatial transcriptomics in 3D. *Cell* **187**, 3953-
1277 3972.e26 (2024).
- 1278 53. DuBois, S. G. *et al.* Metastatic Sites in Stage IV and IVS Neuroblastoma Correlate With Age,
1279 Tumor Biology, and Survival. *J. Pediatr. Hematol. Oncol.* **21**, (1999).
- 1280 54. Rocha, M., Singh, N., Ahsan, K., Beiriger, A. & Prince, V. E. Neural crest development:
1281 insights from the zebrafish. *Dev. Dyn.* **249**, 88–111 (2020).

- 1282 55. Lamant, L. *et al.* Expression of the ALK Tyrosine Kinase Gene in Neuroblastoma. *Am. J.*
1283 *Pathol.* **156**, 1711–1721 (2000).
- 1284 56. Somasundaram, D. B. *et al.* ALK expression, prognostic significance, and its association with
1285 MYCN expression in MYCN non-amplified neuroblastoma. *World J. Pediatr.* **18**, 285–293
1286 (2022).
- 1287 57. Berry, T. *et al.* The ALKF1174L Mutation Potentiates the Oncogenic Activity of MYCN in
1288 Neuroblastoma. *Cancer Cell* **22**, 117–130 (2012).
- 1289 58. Schönherr, C. *et al.* Anaplastic Lymphoma Kinase (ALK) regulates initiation of transcription
1290 of MYCN in neuroblastoma cells. *Oncogene* **31**, 5193–5200 (2012).
- 1291 59. Bergaggio, E. *et al.* ALK inhibitors increase ALK expression and sensitize neuroblastoma cells
1292 to ALK.CAR-T cells. *Cancer Cell* **41**, 2100-2116.e10 (2023).
- 1293 60. Villalard, B. *et al.* Neuroblastoma plasticity during metastatic progression stems from the
1294 dynamics of an early sympathetic transcriptomic trajectory. *Nat. Commun.* **15**, 9570 (2024).
- 1295 61. Grossmann, L. D. *et al.* Identification and Characterization of Chemotherapy-Resistant High-
1296 Risk Neuroblastoma Persister Cells. *Cancer Discov.* **14**, 2387–2406 (2024).
- 1297 62. Westerfield, M. *A Guide for the Laboratory Use of Zebrafish (Danio Rerio)*. (University of
1298 Oregon Press, Eugene, 2007).
- 1299 63. Pan, Y. A. *et al.* Zebrafish: multispectral cell labeling for cell tracing and lineage analysis in
1300 zebrafish. *Development* **140**, 2835–2846 (2013).
- 1301 64. Wallace, C. K. *et al.* Effectiveness of Rapid Cooling as a Method of Euthanasia for Young
1302 Zebrafish (*Danio rerio*). *J Am Assoc Lab Anim Sci* **57**, 58–63 (2018).
- 1303 65. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture
1304 and targeted sequencing. *Nat. Biotechnol.* **38**, 954–961 (2020).
- 1305 66. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome*
1306 *Biol.* **17**, 77 (2016).
- 1307 67. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1308 Preprint at <https://doi.org/10.48550/arXiv.1303.3997> (2013).
- 1309 68. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing
1310 next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
- 1311 69. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number
1312 Detection and Visualization from Targeted DNA Sequencing. *PLOS Comput. Biol.* **12**,
1313 e1004873 (2016).
- 1314 70. Olshen, A. B. *et al.* Parent-specific copy number in paired tumor–normal studies using
1315 circular binary segmentation. *Bioinformatics* **27**, 2038–2046 (2011).
- 1316 71. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the
1317 analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
- 1318 72. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation
1319 programs upon aging of hematopoietic stem cells. *Genome Res.* **25**, 1860–1872 (2015).

- 1320 73. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular
1321 activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).
- 1322 74. Rubin, S. A. *et al.* Single-cell analyses reveal early thymic progenitors and pre-B cells in
1323 zebrafish. *J Exp Med* **219**, e20220038 (2022).
- 1324 75. Hahsler, M., Piekenbrock, M. & Doran, D. **dbSCAN** : Fast Density-Based Clustering with *R. J.*
1325 *Stat. Softw.* **91**, (2019).
- 1326 76. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat.*
1327 *Methods* **14**, 1083–1086 (2017).
- 1328 77. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell
1329 Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).
- 1330 78. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21
1331 (2019).
- 1332 79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
1333 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 1334 80. Decoene, I., Herpelinck, T., Geris, L., Luyten, F. P. & Papantoniou, I. Engineering bone-
1335 forming callus organoid implants in a xenogeneic-free differentiation medium. *Front. Chem.*
1336 *Eng.* **4**, 892190 (2022).
- 1337 81. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell
1338 RNA-seq. *Science* **352**, 189–196 (2016).
- 1339 82. Sztanka-Toth, T. R., Jens, M., Karaikos, N. & Rajewsky, N. Spacemake: processing and
1340 analysis of large-scale spatial transcriptomics data.
- 1341 83. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony.
1342 *Nat. Methods* **16**, 1289–1296 (2019).
- 1343

1344 **Figures**

Fig. 01



1345 **Fig. 1: Multiplexed scRNA-seq to understand NB intra- and inter-tumor**
1346 **transcriptional heterogeneity.**

1347 A) Overview of the main experiment: The two transgenic zebrafish lines MYCN and
1348 MYCN;LMO1 grow NB tumors in the interrenal gland (lateral site) and the latter
1349 also in the arch-associated complex (ventral site). We inject Cas9 and sgRNAs to
1350 later analyze cellular lineage relationships. We performed multiplexed scRNA-seq
1351 with MULTI-seq to gather single cell transcriptome data from several tumors at the
1352 same time, allowing inter-sample analysis with reduced batch effects. hpf = hours
1353 post fertilization, mpf = months post fertilization.

1354 B) Tumor incidence curves (monitoring between six and 25 wpf) for the two transgenic
1355 zebrafish models.

1356 C) H&E-stained sagittal section of a 3 mpf MYCN;LMO1 fish with magnified views of
1357 the lateral tumor in the middle and ventral tumor in the bottom image.

1358 D) UMAP of the entire dataset (around 150 thousand cells) colored by cell types: NB
1359 cells in pink, blood cells in green and blue hues, other stromal cells in brown to
1360 yellow hues. Cells with ambiguous marker gene expression are termed doublets
1361 and are shown in grey.

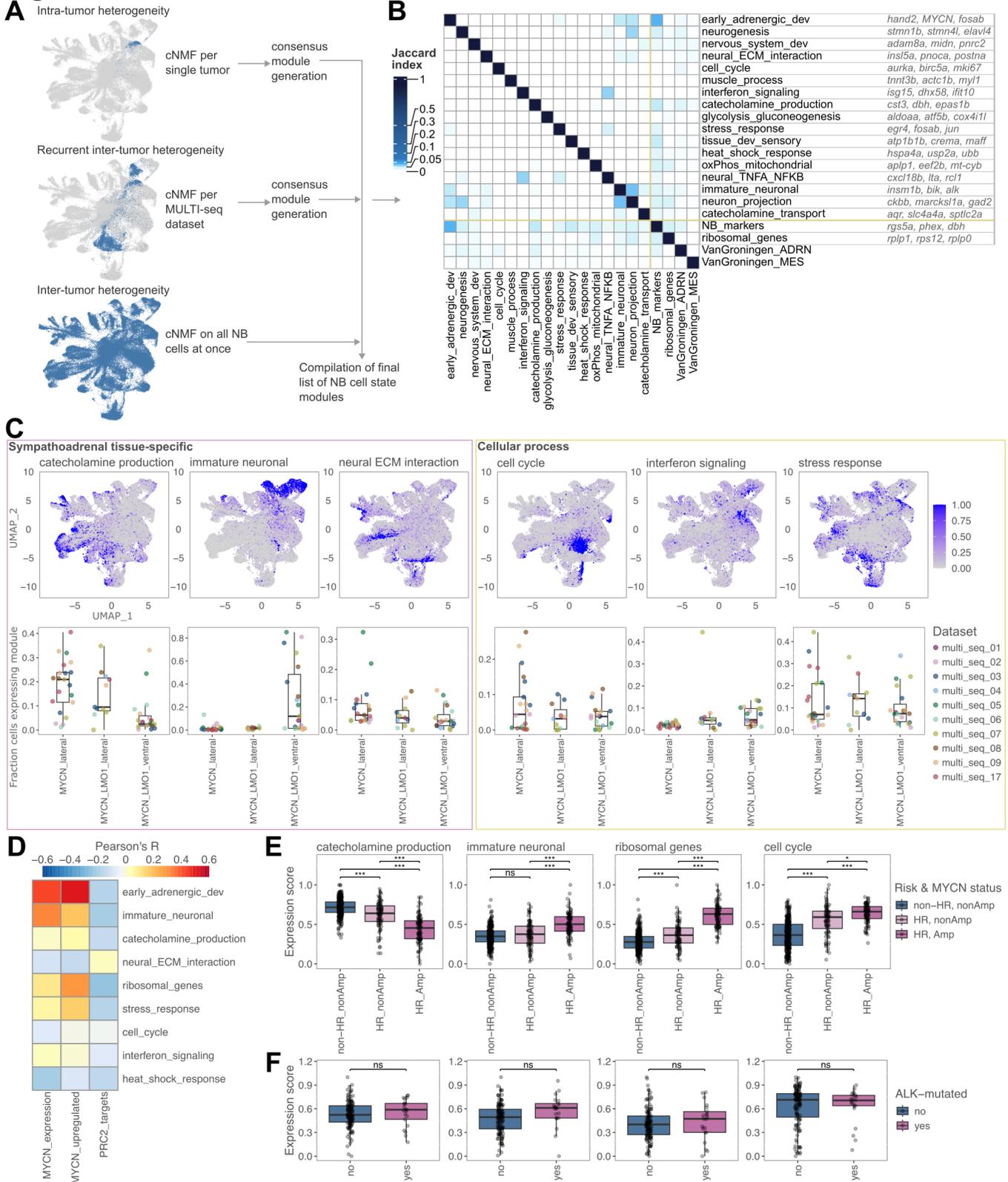
1362 E) 3451 TME cells (left) and 16482 NB cells (right) from one individual MULTI-seq
1363 run (multi_seq_09) sub-clustered and colored by tumor sample of origin or Louvain
1364 cluster in the UMAP. Barplots show the fraction of cells in each cluster made up of
1365 cells from a given sample.

1366 F) Top positively differentially expressed genes in NB cells when compared to all
1367 other cell types. Known adrenergic genes and transgenes are highlighted by blue
1368 and pink boxes, respectively. UMAP of all cell types shows expression score for
1369 the 46 most differentially expressed genes (signature 'NB_markers').

1370 G) Expression scores for the genes most differentially upregulated ('NB_markers' as
1371 in F) or most highly expressed ('ribosomal_genes') in NB cells, as well as for
1372 adrenergic and mesenchymal gene signatures derived from human NB cell lines.
1373 Significance of differential expression between NB cells and all other cells was
1374 determined with Wilcoxon rank sum test (***) $p < 0.0001$.

1375

Fig. 02

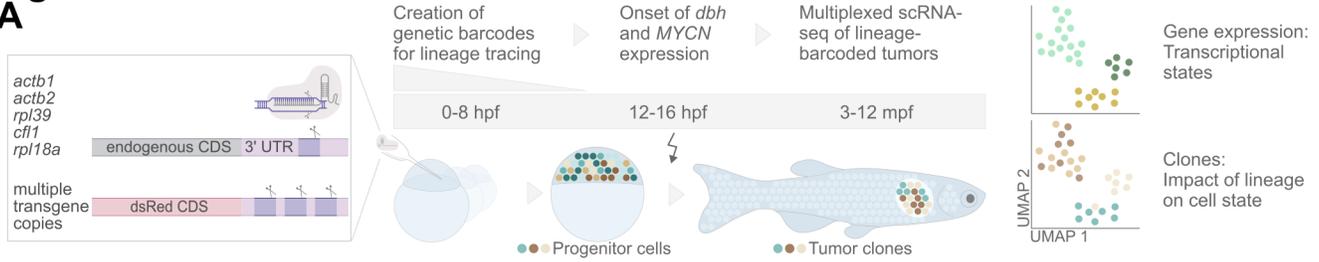


1376 **Fig. 2: The spectrum of transcriptional states of MYCN-driven zebrafish NB.**

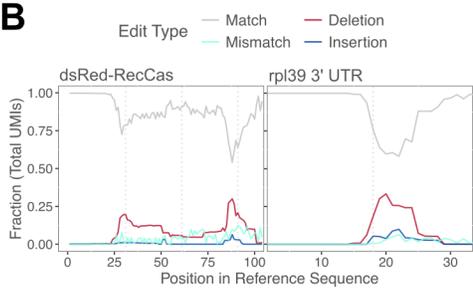
- 1377 A) Gene expression module identification process: A three-fold approach to detect
1378 both modules describing intra-tumor heterogeneity (expressed in a subset of cells
1379 in most tumors) and inter-tumor heterogeneity (expressed only in some tumors and
1380 absent in others). Modules from all approaches were summarized to derive
1381 consensus-modules shown in B.
- 1382 B) Summary of final list of modules compiled from all three approaches in A with
1383 Jaccard index of gene content overlap for all modules as well as the NB differential
1384 and high expression signatures ('NB_markers' and 'ribosomal_genes' as in Fig.
1385 1G) and the human adrenergic and mesenchymal NB signatures (as in Fig. 1G).
1386 The top three genes contributing to each gene module are shown on the right.
- 1387 C) UMAPs of NB cells with expression scores for the indicated modules. Box plots
1388 and jitter show the fraction of cells that express a given module per tumor. Modules
1389 were grouped into sympathoadrenal-specific (pink box) and general cellular
1390 processes (yellow box).
- 1391 D) Pearson correlation between expression scores of zebrafish NB-derived gene
1392 modules (rows) and expression of *MYCN* ('MYCN_expression') or expression
1393 scores for known human MYCN-driven or PRC2-target genes across all zebrafish
1394 NB cells.
- 1395 E) Expression of selected zebrafish NB modules in bulk RNA-seq data from the
1396 SEQC NB cohort (n = 498), grouped by risk factor (HR = high-risk, non-HR = non-
1397 high-risk) and MYCN-status (amplified or non-amplified). Significance determined
1398 with Wilcoxon rank sum test (ns = non-significant, * p < 0.01, ** p < 0.001, *** p <
1399 0.0001).
- 1400 F) Expression of selected zebrafish NB modules in bulk RNA-seq data from the
1401 TARGET NB cohort (n = 151), grouped by ALK-status (mutated or wildtype).
1402 Significance tested and denoted as in E.
- 1403

Fig. 03

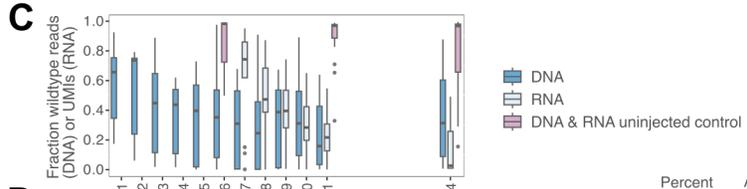
A



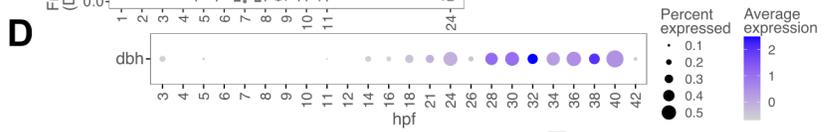
B



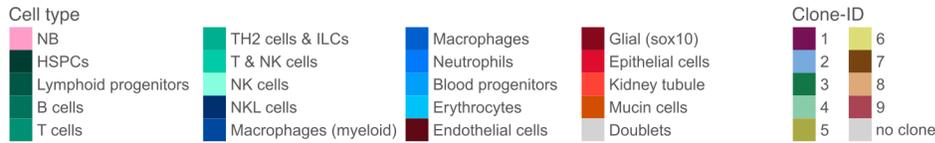
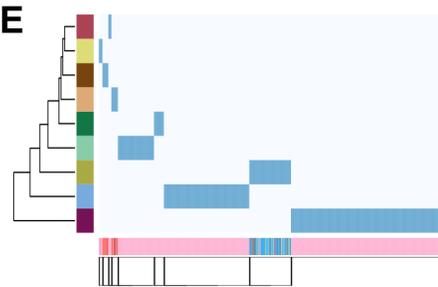
C



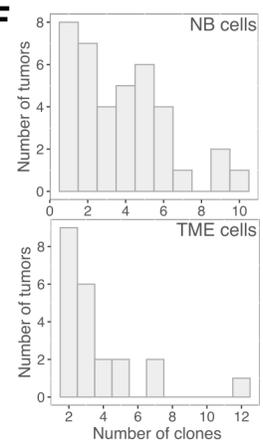
D



E



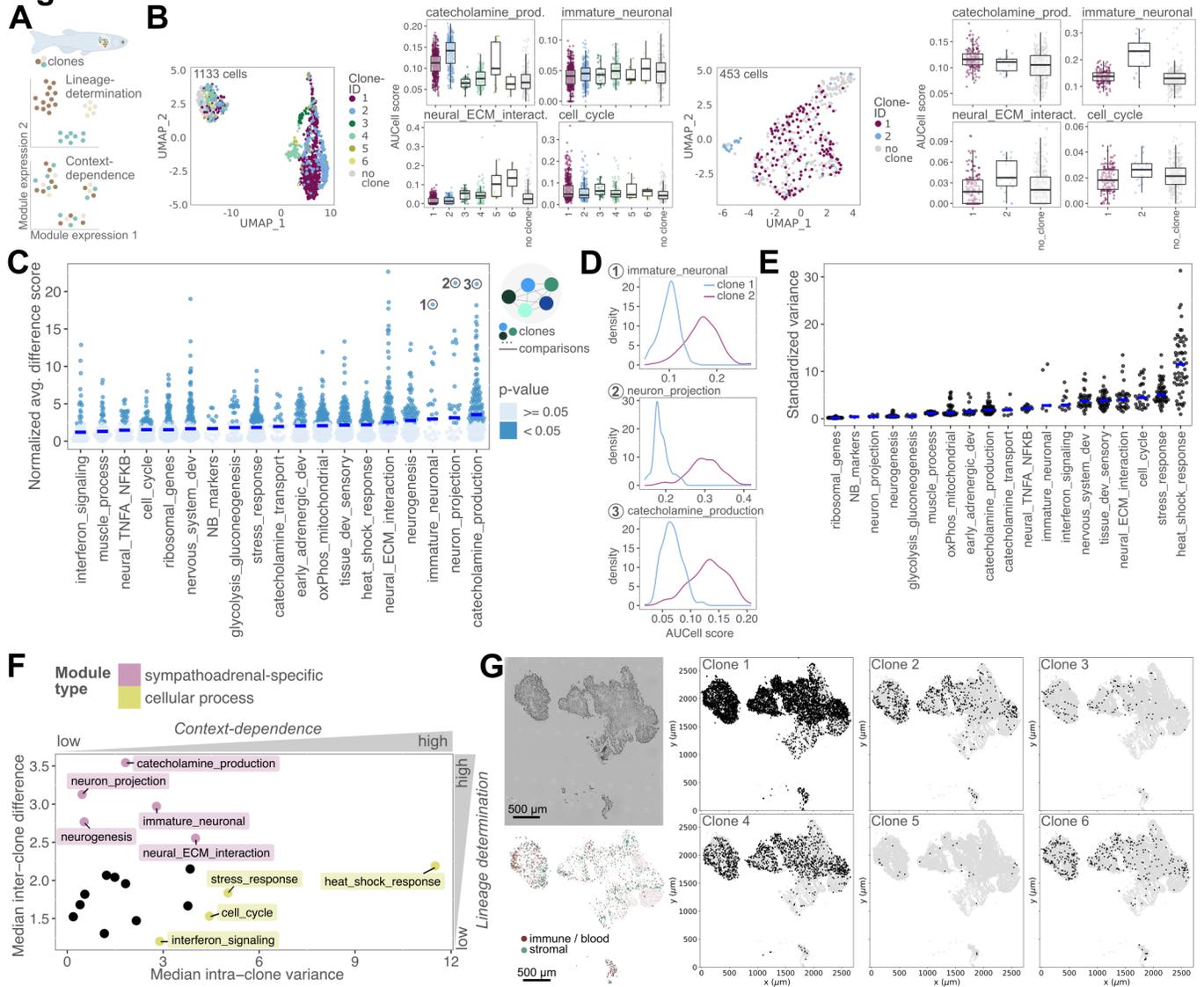
F



1404 **Fig. 3: Tracing tumor cell clones derived from distinct cells of origin.**

- 1405 A) Experimental strategy for early developmental lineage tracing via CRISPR/Cas9.
1406 Cas9 and sgRNAs targeting a transgene carrying three target sites in the 3' UTR
1407 as well as single target sites in the 3' UTR of the listed highly expressed
1408 endogenous genes are injected at the one-cell stage. The repair of Cas9-induced
1409 DNA-lesions results in the formation of indels used as lineage barcodes. Lineage
1410 barcodes are read out together with the cell transcriptomes, when tumors are
1411 processed via scRNA-seq.
- 1412 B) Base position plots for one NB MULTI-seq dataset showing the fraction of UMIs
1413 with edits in a given base in the transgenic recording cassette locus (dsRed-
1414 RecCas, left) and the endogenous *rpl39* 3' UTR locus (15 bases upstream and
1415 downstream of the expected Cas9 cut-site). Border between target (spacer) and
1416 PAM-sequence for each target is marked by a vertical grey line.
- 1417 C) Lineage barcode creation dynamics show a rapid loss of uncut alleles and thus
1418 introduction of lineage barcodes in the first hours of development after
1419 CRISPR/Cas9 injection. Lineage barcode introduction on DNA saturates around 5
1420 hpf and unedited RNA is largely replaced at 11 hpf. Samples from at least two
1421 separate injection experiments were taken for each time point and assay.
- 1422 D) Expression dynamics of *dbh* in scRNA-seq of zebrafish development (Danicell
1423 atlas⁵¹). *dbh* expression is only observed in few individual cells prior to broad
1424 activation at around 14 hpf.
- 1425 E) Hierarchical clustering of 1099 cells from a single tumor (MYCN_lat_m9_1) based
1426 on assigned clone-IDs (dark blue heatmap color: assigned to clone). Cells are
1427 annotated by cell type at the bottom of the heatmap. We found clear lineage splits
1428 between one immune/blood cell clone and two stromal cell clones and identified
1429 six distinct NB cell clones. The UMAPs show cells from the same tumor including
1430 those that could not be assigned to a clone (total of 1410 cells) colored by cell type
1431 or clone-ID.
- 1432 F) Number of clones detected in NB cells (top) or the TME cell compartment (bottom)
1433 per individual tumor (n = 38).
- 1434

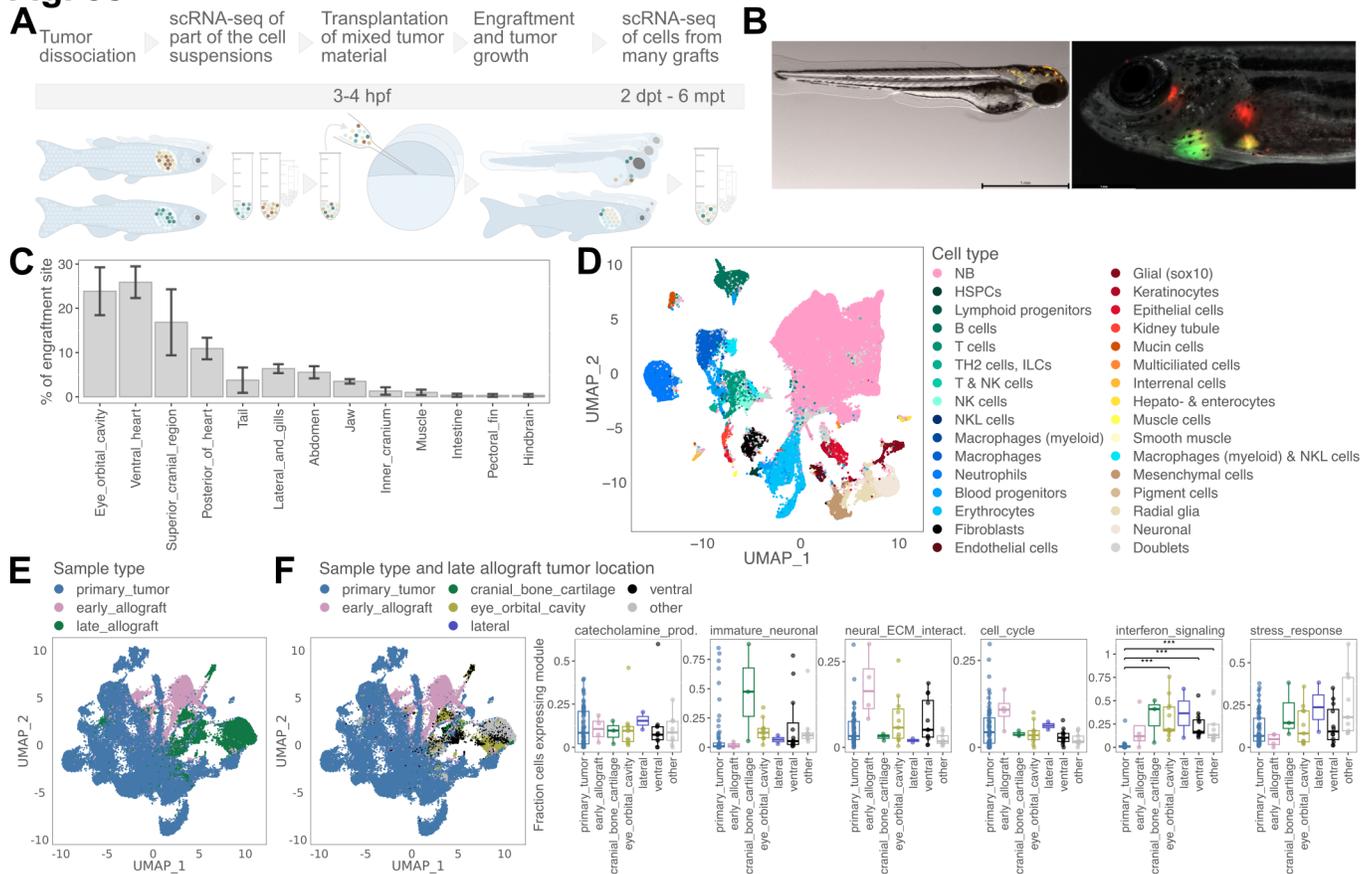
Fig. 04



1435 **Fig. 4: Clonal analysis of NB reveals lineage-determined transcriptional programs.**

- 1436 A) Opposing scenarios of tumor state regulation: In the first scenario, module
1437 expression is determined by the cell of origin (i.e. clone, indicated by color). In
1438 contrast, if tumor states depend on the current conditions in the tumor niche, cells
1439 dynamically take on different tumor states, regardless of their origin.
- 1440 B) UMAP showing NB cells from one tumor colored by the assigned clone for a lateral
1441 tumor (left) and a ventral tumor (right). The AUCell-determined expression scores
1442 for four modules in NB cells grouped by clone is shown right to the UMAP.
- 1443 C) Differential module expression scores for comparisons between pairs of clones
1444 coming from the same tumor (= one dot) with the median differential expression
1445 score indicated by a blue line. Only pairs with detectable module expression in at
1446 least one clone were included (Methods), ranging from 18 pairs for module
1447 'NB_markers' to 188 pairs for module 'ribosomal_genes'. The distributions of
1448 module expression scores for three example comparisons are shown in D.
- 1449 D) Distribution of AUCell-determined expression scores for the pairs of clones
1450 compared in the highlighted examples in C.
- 1451 E) Standardized intra-clone variance of gene modules expression with the median
1452 variance indicated by a blue line. Only clones with detectable module expression
1453 were included (Methods), ranging from 2 clones for module 'NB_markers' to 63
1454 clones for module 'ribosomal_genes'.
- 1455 F) Summary plot showing median inter-clone difference (as in C) versus median intra-
1456 clone variance (as in E) for each module, highlighting sympathoadrenal-specific
1457 and cellular process associated modules.
- 1458 G) Clonal distribution in space using spatial transcriptomics data. Left: Microscopic
1459 image of the tumor section and indication of spots assigned to immune or blood
1460 cell and stromal cell types. Other spots are mainly NB cells. Right: Spatial outline
1461 of the tumor section highlighting spots, in which lineage barcode sequences
1462 representative of the indicated clones (derived from scRNA-seq data) were found.
1463

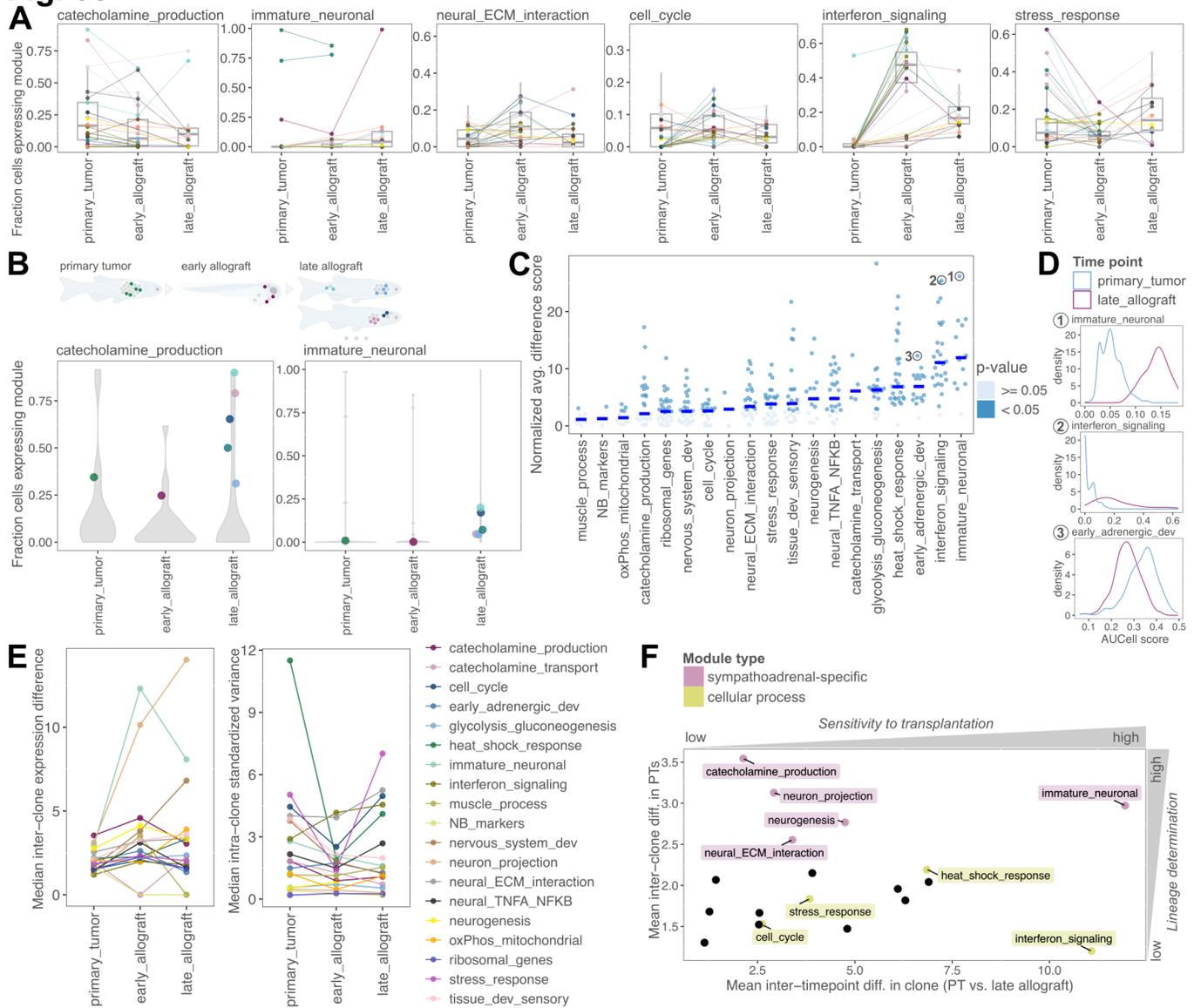
Fig. 05



1464 **Fig. 5: Tumor cell transplantation into an embryonic environment.**

- 1465 A) Experimental workflow for transcriptional profiling of clones over multiple
1466 transplantation time points: Lineage-barcoded cells from multiple tumors are
1467 extracted and a part is sequenced, while the rest is pooled and transplanted into
1468 many wildtype zebrafish blastulae. Transplanted cells are extracted for sequencing
1469 two days or months after transplantation (d/mpt).
- 1470 B) Fish engrafted with a mix of MYCN (green) and MYCN;LMO1 (red and green)
1471 tumor cells three days after transplantation (left) or 2.5 months after transplantation
1472 (right).
- 1473 C) Prevalence of tumors in different engraftment sites three months after
1474 transplantation in three separate transplantation experiments. Error bars denote
1475 the standard error of the mean.
- 1476 D) UMAP of all cells from primary tumors, 2 dpt allografts and late allograft tumors
1477 colored by cell type (around 208 thousand cells). Cell types derived from healthy
1478 tissues of the host larvae at 2 dpt are colored in beige tones (mesenchymal cells,
1479 pigment cells, radial glia, neuronal).
- 1480 E) UMAP of NB cells from all time points (around 131 thousand cells) colored by
1481 sample type (primary tumor, early allograft, late allograft).
- 1482 F) UMAP of NB cells from all time points colored by sample type (primary tumor, early
1483 allograft) or by tumor location for samples derived from late allografts. Fraction of
1484 cells per tumor expressing the indicated module with samples grouped as in the
1485 UMAP. Significance is only shown for significant comparisons as determined by
1486 pairwise Wilcoxon rank sum test (* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$).
- 1487

Fig. 06



1488 **Fig. 6: Exposure to embryonic environment drives re-emergence of expression**
1489 **plasticity.**

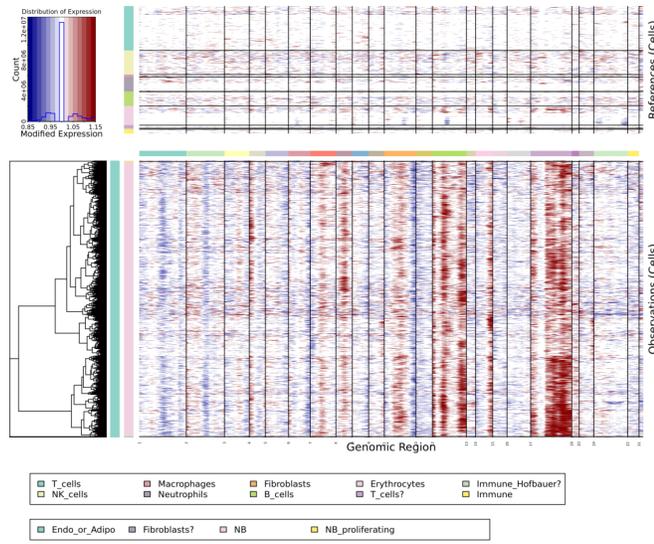
- 1490 A) Fraction of cells per clone and time point that express a given module. Each dot
1491 shows a clone in a specific time point and groups of cells from the same clone in
1492 different time points are connected via lines.
- 1493 B) Fraction of cells from a clone in a time point and location expressing the indicated
1494 modules. Each dot represents cells of the clone in distinct locations (as illustrated
1495 in the sketch above the plot: primary tumor, early allograft timepoint, multiple late
1496 allograft timepoint tumors). Grey violins show background distribution of module
1497 expression fraction for all clones in that dataset.
- 1498 C) Differential module expression scores for comparisons between cells from one
1499 clone found in two different timepoints (one dot = comparison between group of
1500 cells from one clone in primary tumor with group of cells of the same clone in late
1501 allograft tumor). Only pairs of cell groups with detectable module expression in at
1502 least one group were included (Methods), ranging from 2 pairs for module
1503 'neuron_projection' to 19 pairs for module 'stress_response'. Module expression
1504 score distributions for three examples are shown in D.
- 1505 D) AUCell-determined module expression distribution for the respective module in the
1506 cells from one clone in the primary tumor and the late allograft tumor is shown for
1507 three examples highlighted with circles in C.
- 1508 E) Median inter-clone difference (left) and median intra-clone variance (right) for each
1509 module across clones in the three sampled timepoints. Inter-clone difference was
1510 calculated for pairs of clones found in the same tumor or allograft sample. For both
1511 measures, only clones with detectable expression of a given module were
1512 included, as described in Fig. 4C and E.
- 1513 F) Summary plot showing median module expression difference between cells from
1514 one clone in two time points (primary tumor and late allograft, as in C) versus
1515 median inter-clone difference in the primary tumor (as in Fig. 4C). Highlighted
1516 modules are classified as sympathoadrenal-specific or cellular process-
1517 associated.
1518

1519 **Fig S1: Information on zebrafish NB models, cell type composition and annotation.**

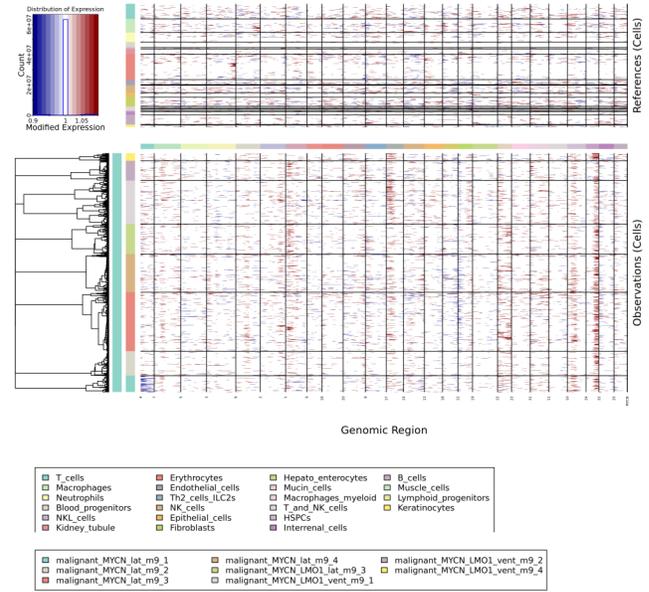
- 1520 A) Stereomicroscopic images of two MYCN;LMO1 fish with a lateral tumor (both fish)
1521 and a ventral tumor (fish on the right).
- 1522 B) UMAP showing sub-clustered immune and blood cell types (upper plots) or sub-
1523 clustered stromal cell types (lower plots) with detailed sub-cell type annotation (left)
1524 and indication of the sample of origin (right). In the plots showing samples of origin,
1525 only the colors used for control samples are highlighted, as these contributed a
1526 large amount of immune and blood cells that separate slightly from tumor-derived
1527 cells on the UMAP. These cells are derived from the hematopoietic tissue of the
1528 head kidney.
- 1529 C) Dotplot showing expression level and prevalence of the top four cell type marker
1530 genes (as determined by differential expression analysis, Table S3) for each final
1531 cell type.
- 1532 D) UMAP of all NB cells (left) and all other cell types (middle and right) colored by the
1533 sample of origin (left and middle) or cell type (right). In the plots showing samples
1534 of origin, only the colors used for control samples are highlighted, as these
1535 samples show larger differences compared to TME cells.
1536

Fig. S02

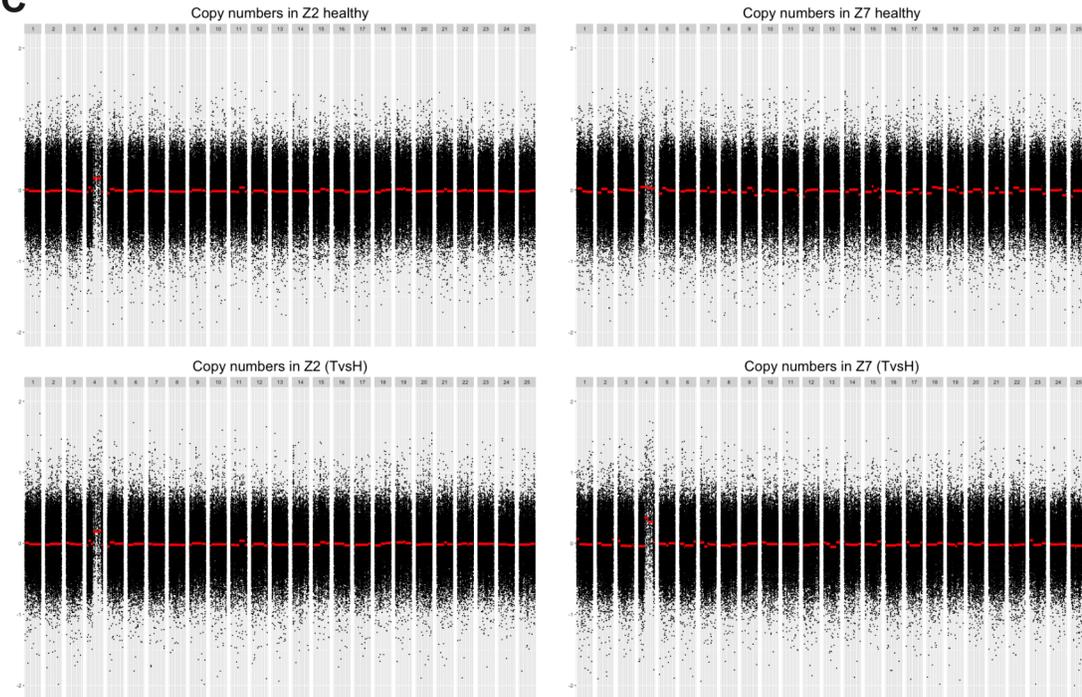
A Human NB dataset (Dong et al. T92)



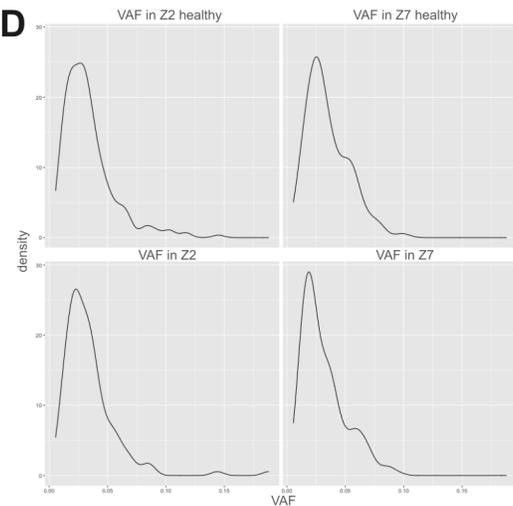
B Zebrafish NB tumors from dataset multi_seq_09



C



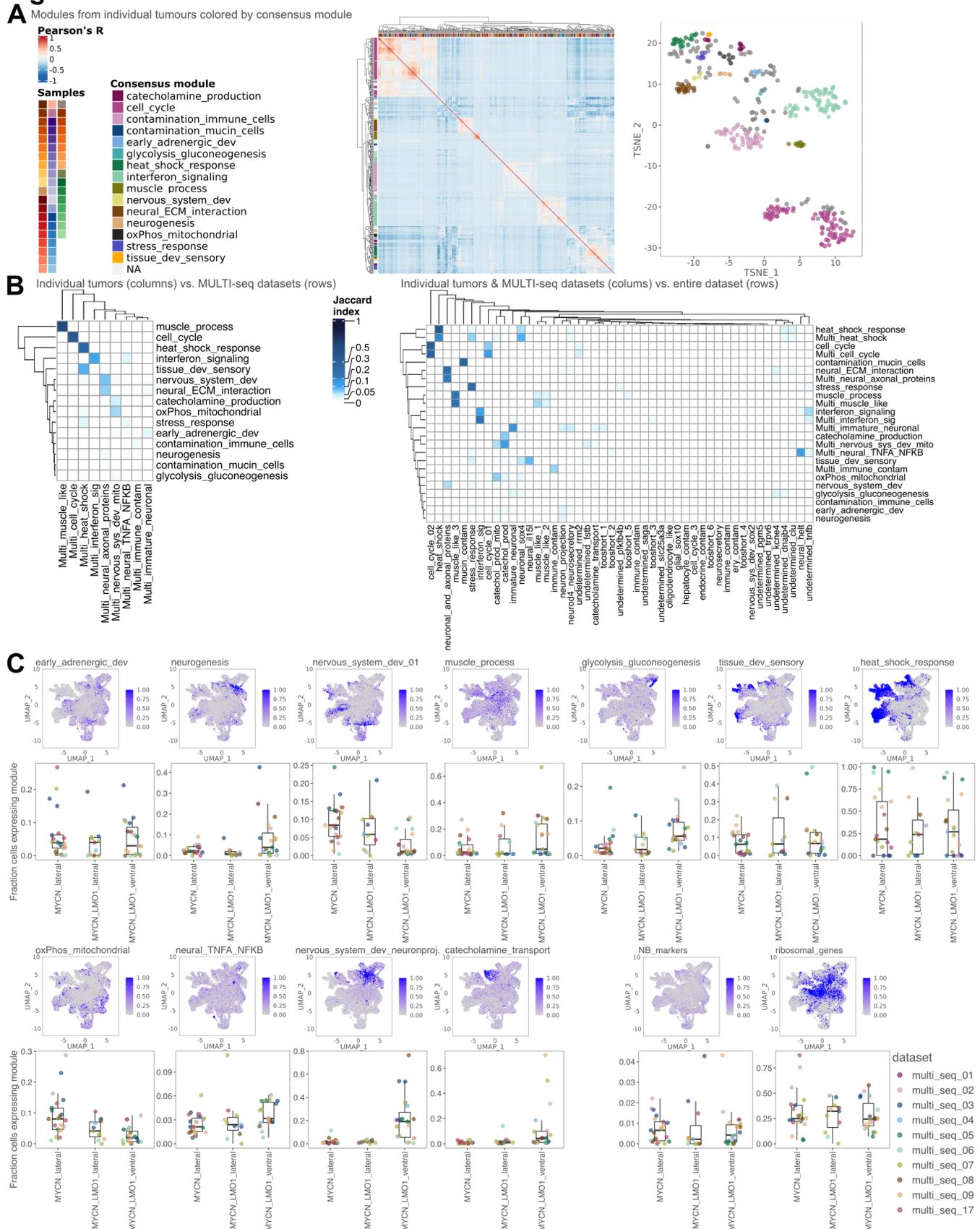
D



1537 **Fig S2: Genetic analysis of zebrafish NB based on inferred CNVs and whole exome**
1538 **sequencing.**

- 1539 A) inferCNV result for a human NB scRNA-seq dataset (T92 from Dong et al.).
1540 Reference cells and background CNV probability profiles are shown at the top. The
1541 tumor cells in the bottom part of the plot show strong signals of CNVs, e.g. for an
1542 amplification of part of chromosome 17, which is frequently observed in NB.
- 1543 B) inferCNV result for all zebrafish NB tumors from one MULTI-seq run
1544 (multi_seq_09). Reference cells and background CNV probability profiles are
1545 shown at the top. Cells from different tumors are grouped and marked with a
1546 distinct color in the legend on the left. There is no strong CNV-signal.
- 1547 C) WES copy number profiles for two samples (Z2 on the left, Z7 on the right). CNV
1548 profiles are shown for healthy control tissue taken from the fin (top) and tumor
1549 tissue (bottom).
- 1550 D) WES variant allele frequencies for the Z2 and Z7 healthy control tissue and tumor
1551 tissue.
1552

Fig. S03



1553 **Fig S3: Detection of NB gene expression modules using a three-level NMF**
1554 **approach.**

1555 A) Heatmap and T-SNE showing modules detected in the analysis of individual
1556 tumors plotted according to their pairwise correlation to each other (Pearson).
1557 Modules in the heatmap are annotated by the individual tumor sample they were
1558 derived from (color legend in top row) and by the consensus module they were
1559 assigned to after clustering with HDBScan (color legend on the left side). Modules
1560 in the T-SNE plot are colored according to the consensus module they were
1561 assigned to after clustering with HDBScan. Modules that were not assigned to any
1562 consensus module are labelled NA and colored in grey.

1563 B) Overlap in terms of gene content (measured by Jaccard index) for all consensus
1564 modules derived from the analysis of individual datasets (as in A), MULTI-seq
1565 datasets or classical cNMF analysis of all NB cells from the entire dataset (Table
1566 S4).

1567 C) UMAPs of NB cells with expression scores for those of the 17 curated consensus
1568 modules not shown in the main figure 2 as well as the signatures for NB cell
1569 differential and NB high expression ('NB_markers' and 'ribosomal_genes',
1570 respectively). Box plots and jitter plots show fraction of cells expressing a given
1571 module per tumor. Tumors are grouped by genotype (MYCN or MYCN;LMO1) and
1572 primary tumor location (lateral or ventral).
1573

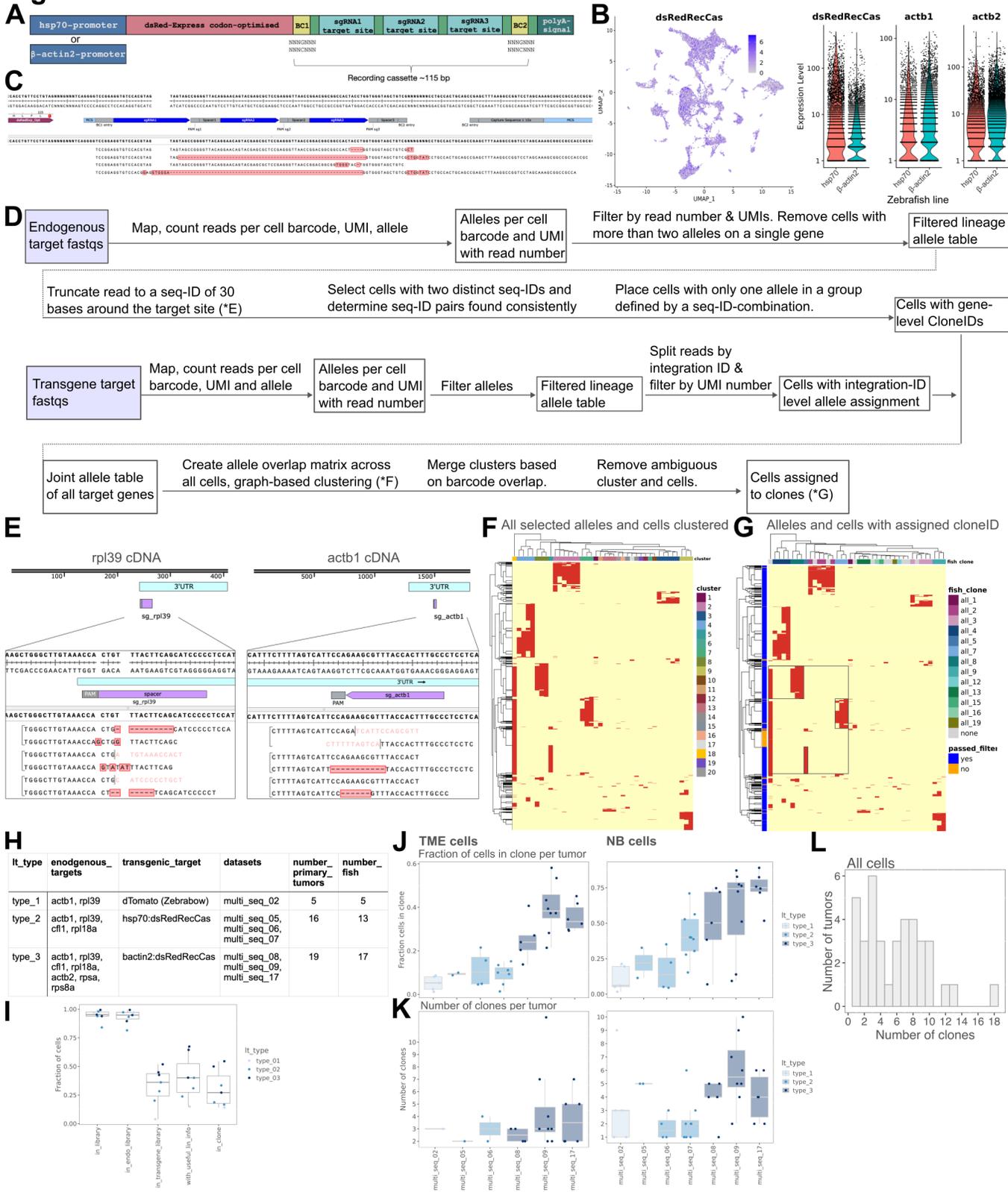
Fig. S04



1574 **Fig S4: Similarity between zebrafish NB and human NB gene expression programs**
1575 **and expression of zebrafish NB modules in human NB.**

- 1576 A) Correlation between expression scores of all zebrafish NB-derived gene modules
1577 (rows) and expression of *MYCN* ('*MYCN_expression*') or expression scores for
1578 human *MYCN*-driven or *PRC2*-target genes across all zebrafish NB cells as in Fig.
1579 2D.
- 1580 B) Gene content overlap as measured as Jaccard index between zebrafish NB gene
1581 modules (rows) and gene expression modules derived from analyses of many
1582 human cancer types (columns, named by study of origin and module function).
- 1583 C) Gene content overlap as measured as Jaccard index between zebrafish NB
1584 modules (rows) and gene expression modules derived from human NBs (columns,
1585 named by study of origin and module function).
- 1586 D) UMAP of integrated human NB cells from 16 tumors (Dong et al.) colored by risk
1587 (low, intermediate, high) and *MYCN*-status (non-amplified or amplified). Upper
1588 heatmap shows the fraction of cells from a given group of tumors that express the
1589 indicated zebrafish NB modules (columns). Bottom heatmap shows the average
1590 expression score for the indicated modules across all NB cells of a given group.
- 1591 E) Expression score of zebrafish NB modules in bulk RNA-seq data from the SEQC
1592 NB cohort (n = 498), grouped by risk (HR = high-risk, non-HR = non-high-risk) and
1593 *MYCN*-status (non-amplified or amplified). Significance of inter-group differences
1594 in E and F were tested with a Wilcoxon rank sum test (ns = non-significant, * p <
1595 0.01, ** p < 0.001, *** p < 0.0001).
- 1596 F) Expression score of selected zebrafish NB modules in bulk RNA-seq data from the
1597 TARGET NB cohort (n = 151), grouped by *ALK* mutational status.
1598

Fig. S05



1599 **Fig S5: Lineage tracing target design, data processing, clone calling and clone**
1600 **detection statistics**

- 1601 A) Lineage tracing cassette design: Three Cas9-target sites are encoded in the 3'
1602 UTR of a dsRed-gene sequence, optimized for expression in zebrafish, flanked by
1603 two 7 basepair barcode sequences (BC1 and BC2) that function as integration IDs
1604 to distinguish multiple genome insertions of the transgene.
- 1605 B) Expression levels of dsRed-RecCas (considering reads from the full length of the
1606 gene) and endogenous *actb1* and *actb2* in scRNA-seq data derived from zebrafish
1607 larvae of the hsp70:dsRedLinRecorder or bActin2:dsRedLinRecorder lines. Violin
1608 plots show expression of the transgene and two highly expressed endogenous
1609 genes in the two lines.
- 1610 C) Sequence of dsRed-RecCas with aligned CRISPR/Cas9-edited example reads.
1611 Edits include small alterations at one cut site as well as larger deletions induced
1612 by cutting of multiple target sites.
- 1613 D) Workflow for pre-processing and filtering of lineage target sequencing data from
1614 endogenous and transgenic targets as well as clone calling. E-G show data
1615 representations of an example dataset at various stages of the workflow and are
1616 referenced in the corresponding steps in the flowchart.
- 1617 E) Pairs of sequence-IDs found in three different cells on *rpl39* (left) and *actb1* (right)
1618 3' UTRs.
- 1619 F) Heatmap of cells (rows) and alleles (columns) clustered by their presence (red).
1620 Data represents all cells from one fish that passed filtering for at least one target
1621 gene and alleles (for transgene) or allele combinations (for endogenous targets)
1622 from all target genes (columns). Alleles were clustered based on cell barcode
1623 overlap with clusters indicated above the heatmap.
- 1624 G) Heatmap as in F. Final clone-assignments are indicated above the heatmap with
1625 alleles that did not pass filtering shown in grey. Cells that could unambiguously be
1626 assigned to a single clone are highlighted in blue on the left side of the heatmap.
1627 The heatmap represents maximum resolution clone calling. One allele is shared
1628 between many cells from multiple different clones (large box and smaller box
1629 insets). Cells that only carry the allele that is shared between the clones (which is
1630 thus ambiguous) are not assigned to a clone and are excluded (orange legend on
1631 the side of the heatmap). In order to increase the number of cells and lineage
1632 barcodes used, the clone calling resolution can be lowered, so that all cells in the
1633 large box would be merged into one clone.
- 1634 H) Conditions used for lineage tracing injections in three experimental rounds.
1635 Different endogenous 3' UTRs and transgenic loci were targeted in the conditions.
1636 Adult fish from one injection round were used for one or multiple MULTI-seq runs
1637 (datasets).

- 1638 I) Fraction of cells per MULTI-seq dataset (dots) with lineage information at different
1639 steps of the filtering and clone calling pipeline (in_library = cells with at least two
1640 UMIs with more than one read each in all targeted libraries; in_endo_library = cells
1641 with at least two UMIs with more than one read each in targeted libraries for
1642 endogenous genes; in_transgene_library = cells with at least two UMIs with more
1643 than one read each in targeted library for transgene target; with_useful_lin_info =
1644 cells with a valid combination of alleles on any endogenous target and/or a valid
1645 lineage barcode (uncut) on a transgenic target; in_clone = cells that could be
1646 assigned to a high resolution clone). It_types as in H.
- 1647 J) Fraction of cells assigned to a high-resolution clone per tumor (dots) grouped by
1648 MULTI-seq dataset considering only TME cells or only NB cells (right). It_types as
1649 in H.
- 1650 K) Number of high-resolution clones per tumor (dots) grouped by MULTI-seq dataset
1651 considering only TME or only NB cells (right). It_types as in H.
- 1652 L) Number of clones detected considering all cells (TME and NB) per individual tumor
1653 (as in Fig. 3F for NB and TME cells separately).
1654

Fig. S06



1655 **Fig. S6: Lineage-dependence of gene expression in zebrafish NB tumors.**

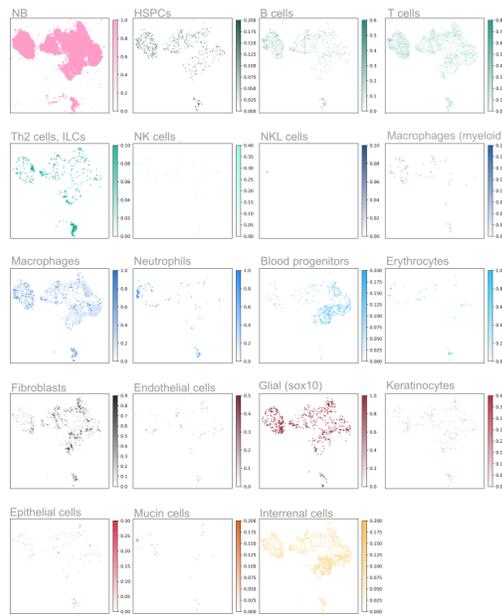
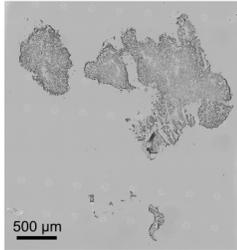
1656 A) UMAPs of NB cells taken from three example fish colored by the tumor location or
1657 the clone-ID. Fish 17_5 and 08_2 had a ventral and lateral tumor each that show
1658 clear separation on the UMAP as well as distinct clonal composition. Expression
1659 scores of selected gene modules are shown as box- and jitter-plots. Inter-clone
1660 expression differences are particularly strong for clones found in distinct tumor
1661 locations.

1662 B) Fraction of significantly different module expression scores between groups of
1663 cells in four different comparisons. First pairwise differences between clones in a
1664 single tumor (sub-)sample are shown (as in Fig. 4C). Comparison of one clone vs.
1665 cells from all other clones in the same (sub-)sample show similar results.
1666 Comparison between cells from the same clone found in different sub-locations of
1667 the tumor shows overall lower differences. Comparisons between different clones
1668 found in the same fish, but in ventral and lateral locations show overall larger
1669 differences.

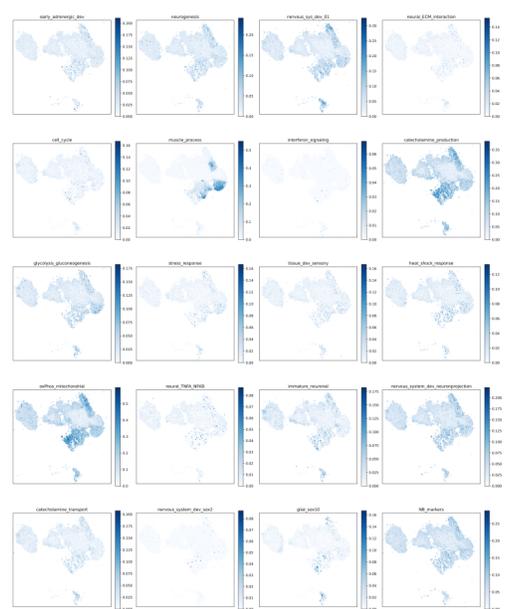
1670

Fig. S07

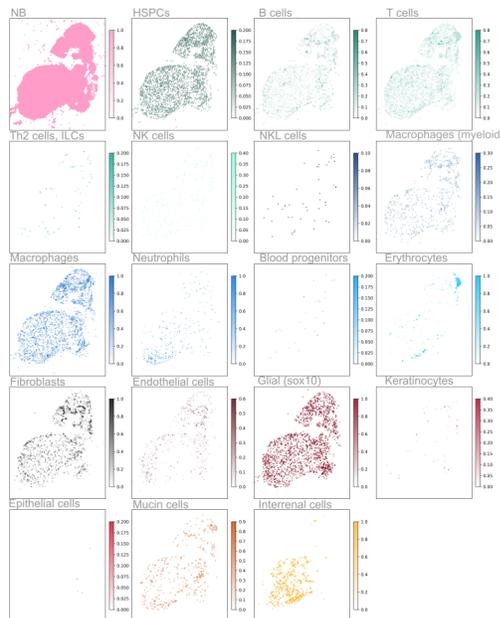
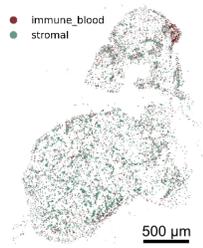
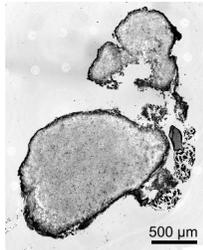
A Tumor 1 - MYCN lateral



B



C Tumor 2 - MYCN lateral



D



1671 **Fig S7: Spatial transcriptomics of two lateral MYCN-tumors.**

1672 A) Light microscopic image of the section taken for Open-ST from tumor 1 as well as
1673 scores for the indicated cell types in 5 μm spots, using cell type marker genes
1674 derived from scRNA-seq data.

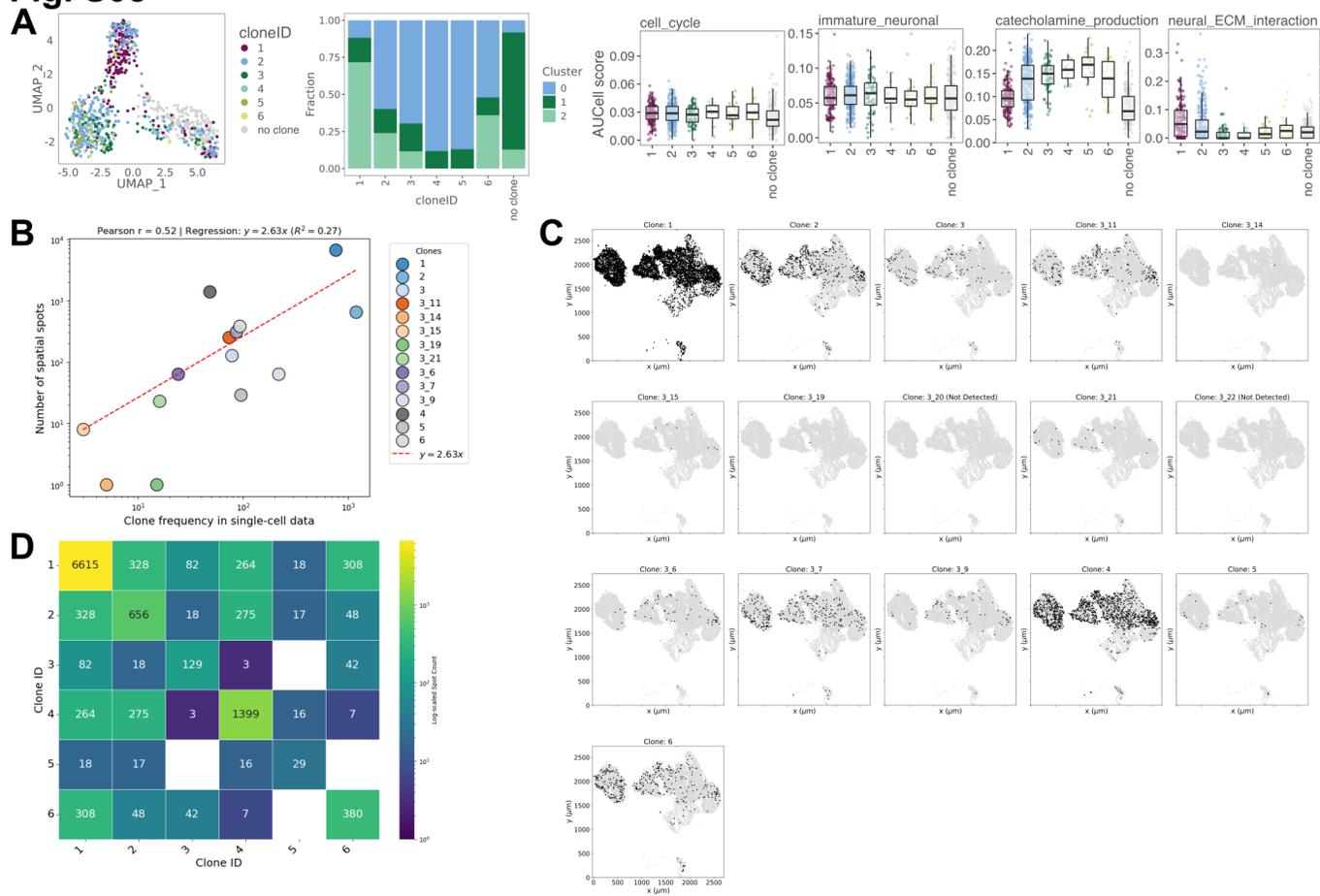
1675 B) Expression score for NB gene expression modules in 5 μm spots across the tumor
1676 section.

1677 C) Light microscopic image of the section taken for Open-ST from tumor 2 as well as
1678 scores for the indicated cell types in 5 μm spots, using cell type marker genes
1679 derived from scRNA-seq data.

1680 D) Expression score for all NB gene expression modules in 5 μm spots across the
1681 tumor section.

1682

Fig. S08



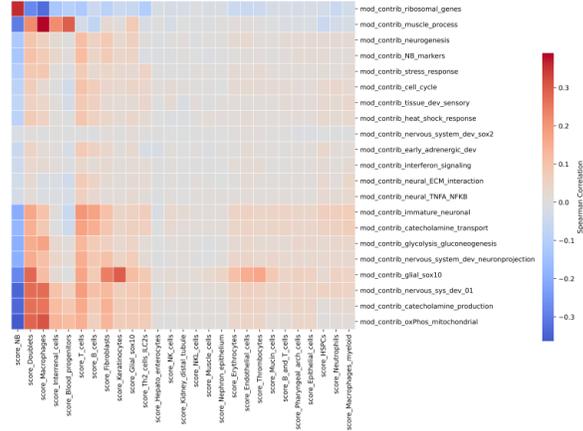
1683 **Fig S8: Spatial transcriptomics with lineage tracing.**

- 1684 A) UMAP of single cell transcriptomes derived from scRNA-seq from the same tumor
1685 that was profiled with spatial transcriptomics. Cells are colored according to the
1686 clone that they were assigned to. The barplot of fraction of cells from each clone
1687 assigned to a given Louvain cluster shows that clones occupy different areas in
1688 transcriptional space. Boxplots and jitter plots show the expression score for the
1689 indicated modules per clone.
- 1690 B) Comparison between number of cells in the scRNA-seq dataset that were
1691 assigned to a clone (x-axis) and number of spatial spots, in which a clone-specific
1692 sequence was found (y-axis) shows an overall agreement in the relative clone
1693 sizes across both data modalities.
- 1694 C) Spatial outline of the tumor section highlighting spots, in which lineage barcode
1695 sequences representative of the indicated clones (derived from scRNA-seq data)
1696 were found. Clones 1 to 6 were defined as large NB cell clones in the scRNA-seq
1697 data and are also shown in Fig. 4G. The other clones shown here either had very
1698 few NB cells or were completely composed of TME cells in the scRNA-seq data.
- 1699 D) Heatmap showing spatial overlap of clones. The numbers indicate the number of
1700 spots, in which at least one read for each of two clonal lineage barcodes was found.
1701 The diagonal shows the total number of spots in which at least one lineage barcode
1702 of a particular clone was found. Most clones overlap in space to a certain extent
1703 with differences between clone pairs, e.g. almost 80 % of all spots with clone 6
1704 labels also have a clone 1 label; in contrast, only about 20 % of all spots with clone
1705 4 labels also have a clone 1 label and clone 4 and clone 6 have little overlap in
1706 space.
1707

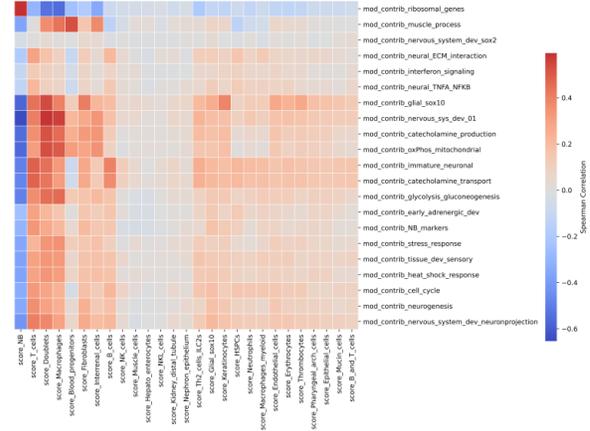
Fig. S09

A Spatial correlation of NB module expression and cell type scores

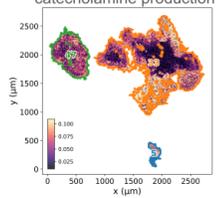
distance threshold: 6 μm



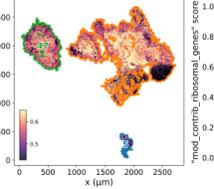
distance threshold: 20 μm



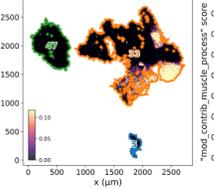
B catecholamine production



ribosomal genes

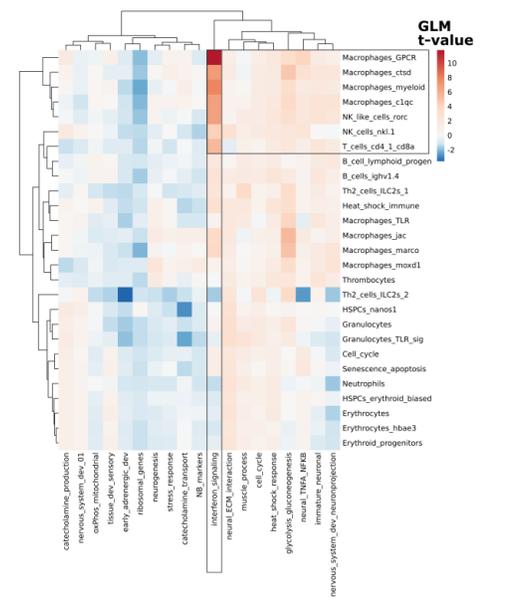


muscle process

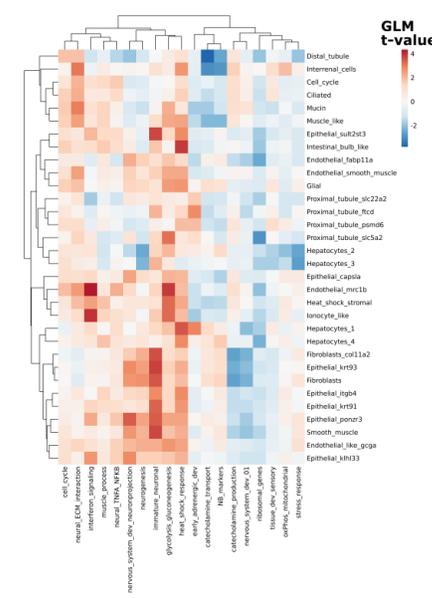


C quasi-Poisson GLM: NB cell fraction ~ TME cell fraction + tumor location + dataset + dissociation

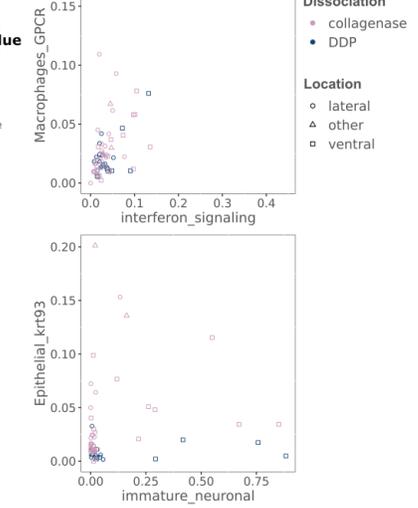
Association of prevalence of NB modules and immune/blood modules



D Association of prevalence of NB modules and stromal modules



E

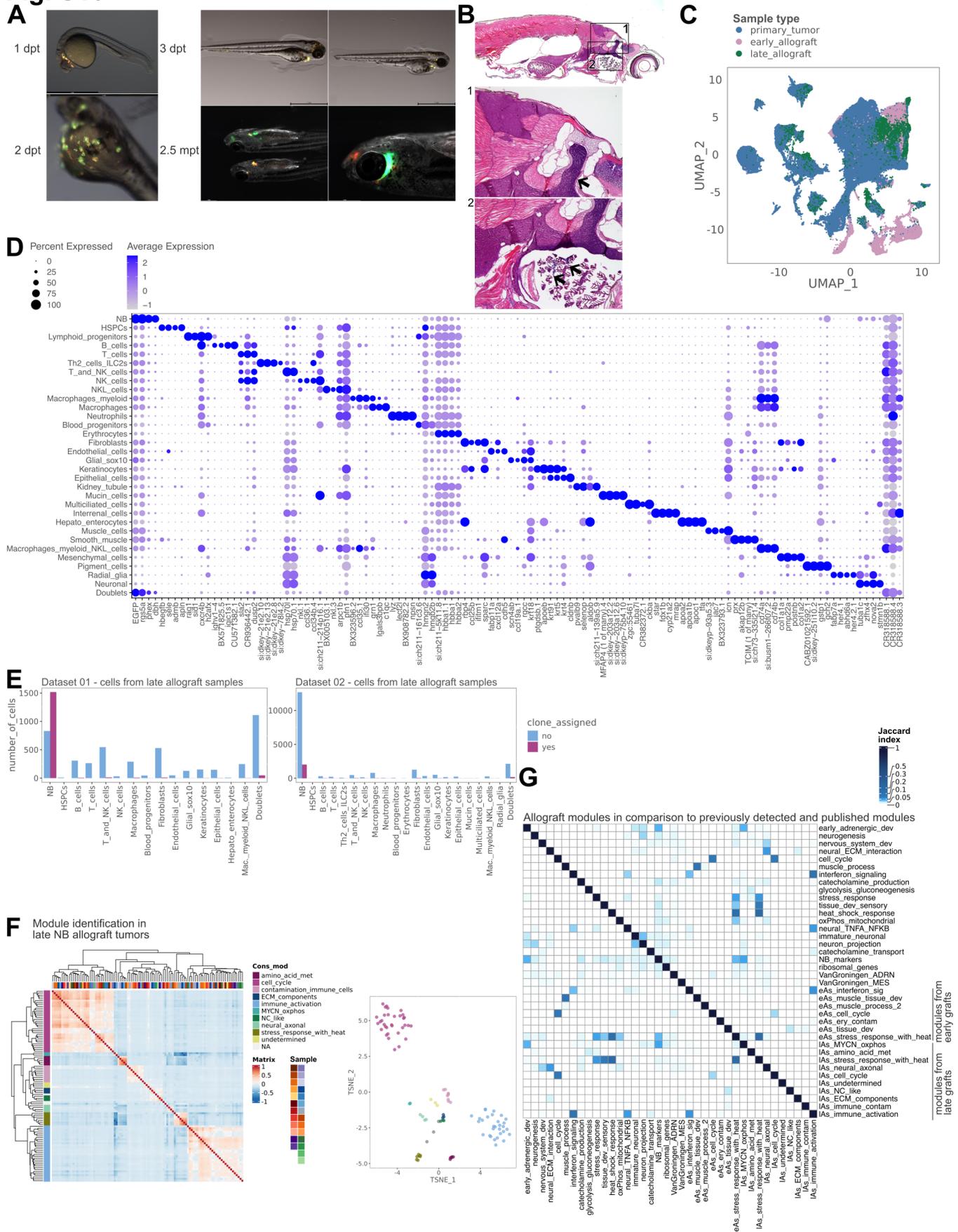


1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745

Fig S9: Association of NB gene expression program activation with presence of TME cell types in spatial transcriptomics and scRNA-seq data.

- A) Co-localization of NB cells expressing distinct modules and TME cell types in spatial transcriptomics: Correlation of activation of NB gene expression programs and scores for TME cell types in spots that are directly neighboring each other (left) or that are found in a slightly larger neighborhood (right). Spots directly neighboring each other (6 μm) have an average of 6.53 neighbors, while the larger neighborhood (20 μm) contains an average of 56.69 neighbors for one spot.
- B) Spatial distribution of module expression relative to the tumor border for three modules (*ribosomal_genes*, *catecholamine_production*, *muscle_process*). The tumor border was defined from the microscopic image and is indicated in orange, green or blue depending on tumor region. Next to expression scores on the tissue section, spots are plotted according to their distance from the tumor border and their expression of a given module. *Ribosomal_genes* expression is positively correlated with distance from the border, *catecholamine_production* is inversely correlated with it and *muscle_process* doesn't show a systematic association.
- C) Co-occurrence of NB cells in a specific state and individual TME cell types in scRNA-seq data: Heatmap showing t-values of association of fraction of TME cells with expression of modules representative of immune / blood cells (rows) and fraction of NB cells expressing NB modules (columns) in a generalized linear model considering the dataset, dissociation method and tumor location as covariates. Only modules with at least one significant association after correction for multiple testing are shown. Association of interferon signaling activation in NB cells and immune cells expressing macrophage-like programs is highlighted as an example of putative functional association (see scatterplot in E).
- D) Heatmap showing t-values of association of fraction of TME cells expressing modules representative of stromal cells (rows) and fraction of NB cells expressing NB modules (columns) in a generalized linear model (as in C). Only modules with at least one significant association are shown.
- E) Scatter plot showing fraction of NB cells expressing the module *interferon_signaling* and fraction of TME cells expressing the module 'Macrophages_GPCR' per tumor (top) shows a visible association of the two values. Tumors are colored by dissociation protocol used and shaped according to the tumor location. Association of TME cells expressing the module 'Epithelial_krt93' and NB cells with *immature_neuronal* activation looks less clear (bottom).

Fig. S10



1746 **Fig S10: Tumor cell allogeneic transplantation and scRNA-seq of allograft tumors.**

- 1747 A) Example images of fluorescent tumor cells (mCherry- and GFP-channels overlaid
1748 on the light microscopic image) in host fish from 1 day post transplantation (dpt) to
1749 2.5 months post transplantation (mpt).
- 1750 B) H&E-stained sagittal section of a 3 months old host fish with graft tumors. The
1751 insets highlight (1) a large lateral tumor mass close to the superior cervical
1752 ganglion and (2) tumor cell foci along the gills.
- 1753 C) UMAPs of all cells from all timepoints colored by sampling time point, showing that
1754 a large population of putative NB cells from the early and late allograft stage cluster
1755 together with primary tumor NB cells.
- 1756 D) Dotplot showing top four marker genes determined by differential expression for
1757 each cell type in the dataset with cells from all timepoints (shown in C).
- 1758 E) Detection of NB-clone-specific lineage barcodes (derived from primary tumors) in
1759 cells of late allograft tumors in two separate experiments, in which data from all
1760 three timepoints was collected (primary tumor, early allograft, late allograft). Mainly
1761 graft cells identified as NB cells based on gene expression were assigned to
1762 primary tumor clones, while relatively few other cell types carried lineage barcodes
1763 also found in the primary tumor.
- 1764 F) Module detection in late allograft tumors. Heatmap and T-SNE showing modules
1765 detected in the analysis of individual late allograft tumors plotted according to their
1766 pairwise correlation to each other (Pearson). Modules in the heatmap are
1767 annotated by the individual late graft tumor sample they were derived from (color
1768 legend in top row) and by the consensus module they were assigned to after
1769 clustering with HDBScan (color legend on the left side). Modules in the T-SNE are
1770 colored according to the consensus module they were assigned to after clustering
1771 with HDBScan. Modules that were not assigned to any consensus module are
1772 labelled NA and colored in grey.
- 1773 G) Summary of final list of modules derived from the primary tumor samples, the
1774 human adrenergic and mesenchymal gene signatures and the modules derived
1775 from early graft NB cells (modules with prefix 'eAs') or late graft tumors (modules
1776 with prefix 'lAs'). Jaccard index shows gene content overlap for all modules.
1777

1778 **Fig. S11: Clonal dynamics and gene module expression changes across**
1779 **allografting timepoints**

- 1780 A) Heatmaps of all cells from one tumor (columns) that passed filtering for at least
1781 one target gene versus alleles (for transgene) or allele combinations (for
1782 endogenous targets) from all target genes (rows). Alleles are colored by the clone
1783 they are definitive of left of the heatmap. Alleles that did not pass filtering are shown
1784 in grey. Cells that could unambiguously be assigned to a single clone are
1785 highlighted in orange (left heatmap) or blue (right heatmap). The left heatmap
1786 represents clone calling, which maximizes resolution in the assignment of cells to
1787 clones. This leads to the exclusion of cells, which only carry lineage barcodes that
1788 were created early, in favor of splitting subgroups of this population into smaller
1789 clones based on lineage barcodes created later (black boxes). The heatmap on
1790 the right shows the result of lower resolution clone calling as used for the analyses
1791 in Fig. 5 and 6. Here, retaining the large group of cells excluded in the high-
1792 resolution analysis is favored (black box). Many cells are placed into one clone
1793 regardless of the additional lineage barcodes they may have obtained later.
- 1794 B) Clonal dynamics for one experiment with cells from all three timepoints (primary
1795 tumor, early allograft, late allograft). Barplot shows composition of primary tumors,
1796 early graft samples and late graft tumors by clone. Very small clones were grouped
1797 ('summed_small_clones'). Cells that could not be assigned to a clone, but to a
1798 primary tumor fish of origin are shown in black ('fish_def_lineage_info'). Cells that
1799 only carry lineage barcodes that are ambiguous in terms of their origin are shown
1800 in dark grey ('ambiguous_lineage_info'). Cells with only uncut lineage reads or
1801 lacking lineage info entirely are shown in lighter grey hues. Left UMAP shows cells
1802 colored according to fish of origin (for primary tumors) or assigned fish of origin
1803 (for graft samples) and sampling time point (PTs = primary tumor, eA = early
1804 allograft, IA = late allograft). Right UMAP shows cells colored according to
1805 clone/fish of origin (for primary tumors) or assigned clone/fish of origin (for graft
1806 samples) and sampling time point.
- 1807 G) Fraction of cells per clone and time point that express a given module. Each dot
1808 shows a clone in a specific time point and groups of cells from the same clone in
1809 different time points are connected via lines (as in Fig. 6A).
- 1810 C) Differential module expression scores for comparisons between cells from one
1811 clone found in two different time points (as indicated above the plots) with the blue
1812 bars indicating the median. Summary plot on the right shows median values of
1813 differential expression for each timepoint and module.
- 1814 D) Scatterplots as in Fig. 4F for early allograft (left) and late allograft (right) timepoints.
1815 Plots show median inter-clone difference in module expression (between clones
1816 within one early allograft sample or late allograft tumor) and median intra-clone
1817 variance for cells from individual clones found in one early or late graft sample.