**IMPORTANT NOTE:** The following is a thorough annotation of a subset of clusters we identified within our transcriptional heterogeneity analysis project. Currently we are considering publishing this series of documents independently of the main paper, as an online resource for users to reference when diving into our pan-cancer atlas. These will be then updated with time as new and better information emerges. Alternatively, a stripped-down version could be published as supplemental information.

**T001 Neuroblastoma**

A total of 180 samples were labelled as neuroblastoma in our reference dataset, 162 of which are from TARGET while the rest are from different sources within the TreeHouse initiative. RACCOON identified four separate subclusters within the neuroblastoma parent class (Fig. 5a). They exhibit different molecular profiles and clinical outcomes, roughly overlapping with microarray expression clusters described in literature1.

**T062 NEBLA MES ERBB2**, the smallest cluster, contains a molecular signature which corresponds to microarray cluster p43 and comprises Children’s Oncology Group (COG) samples reported as high-risk. It is characterized by underexpression of the neuroblastoma predisposition genes *PHOX2B* (median LogFC= -2.58, FDR ≤ 2.93e-13), *MYCN* (median LogFC= -3.30, FDR ≤ 3.11e-09), as well as the cell-cycle genes *BIRC5* (median LogFC= -5.92, FDR ≤ 2.90e-16) and CCND1 (median LogFC= -2.13, FDR ≤ 4.84e-12), compared to its sibling clusters. **T062 NEBLA MES ERBB2** is also characterized by overexpression of *ERBB2* (*HER2*) (median LogFC= 1.32, FDR < 8.16e-08), which has been demonstrated to be a favourable prognostic factor2. Enrichment of *ERBB2* signalling is also seen in this cluster (medNES ≥ 1.31, KW adj. p-val = 1.25e-13, Dunn adj. p-val > 1.00e-02)3,4. Neuroblastomas expressing *ERBB2* have increased differentiation, immunoreactivity, and patients show improved overall survival compared to patients with tumors with either low or no *ERBB2* expression2. Indeed, we observe statistically significantly higher immune infiltration and lower stemness in this cluster (.38 median score, KW adj. p-val=1.39e-10 and .75 median score and KW adj. p-val = 1.75e-12 respectively, see Methods for details on these scores), with respect to all the other classes (Fig. 5c, 5f). Furthermore, **T062** contains the majority of nodular ganglioneuroblastomas (7/12, *χ*2 p-val = 3.849e-05) (Fig. S22b) and shows significant overexpression of ganglioneuroblastoma marker *ERRB3* (median LogFC= 5.7, FDR ≤ 3.484e-15), as well as enrichment for ERBB network gene sets (medNES ≥ 1.27, KW p-val = 1.95e-09, Dunn adj. p-val < 1.00e-04)5. It also exhibits enrichment of glial cell developmental pathways (medNES ≥ 1.31, KW p-val = 2.55e-12, Dunn adj. p-val < 1.00e-03) and adrenal gland developmental gene sets (medNES ≥ 1.30, KW adj. p-val= 2.14e-12, Dunn adj. p-val < 1.00e-02) 3,4.

**T063 NEBLA ADR NTRK1**, the most populous subgroup, corresponds to microarray cluster p13, and is characterized by overexpression of *NTRK1* with respect to its sibling clusters (median LogFC= 1.51, FDR < 5.78e-4). Patients with tumors within this class are significantly younger than ones in other NEBLA clusters (KW FDR = 1.25e-05). It contains all samples classified as low and intermediate COG risk by TARGET (*χ*2 p-val = 1.04e-08), and contains all patients classified as stage 4s (*χ*2 p-val = 2.76e-07) and stage 3 (*χ*2 p-val = 3.58e-2) (Fig. S22b). It is significantly enriched in patients with tumors with favourable histology (*χ*2 p-val = 3.30e-08), and also contains the only intermixed ganglioneuroblastoma tumor referenced in the TARGET cohort (Fig. S22b). **T063** shows enrichment of gene sets related to sympathetic nervous system development (medNES = 1.08, KW adj. p-val = 1.97e-17, Dunn adj. p-val < 1.00e-02) and chromaffin cells (medNES = 1.11, adj. p-val = 4.49e-17, Dunn adj. p-val < 1.00e-04)3,4, suggesting this cluster may be defined by sympathoadrenal differentiation. It carries low immune infiltration (median score 0.29) (Fig. 5c) and high stemness (median score 0.77) (Fig. 5f).

The two remaining clusters, **T064 NEBLA MYCN** and **T065 NEBLA ADR TERT**, are exclusively comprised of samples marked as COG high-risk (Fig. S22b), and overlap with microarray clusters p3 and p21, respectively. Both clusters exhibit overexpression of *BIRC5* compared to **T062 NEBLA ERBB2** and **T063 NEBLA ADR NTRK1** (**T064** logFC 1.74, FDR = 3.33e-05; **T065** logFC = 2.05, FDR = 9.47e-07). **T064 NEBLA MYCN** is characterized by a statically significant overexpression of *MYCN* (median LogFC= 1.51, FDR ≤ 5.78e-04), and contains the majority of samples flagged as *MYCN* amplified by TARGET (*χ*2 p-val = 7.31e-15) (Fig. 5a). It is also defined by the underexpression of *NTRK1* (logFC = -3.25, FDR = 1.99e-09). Though the majority of **T064 NEBLA MYCN** samples are marked by TARGET as *MYCN*-amplified, 37.5% of samples (n = 9/34) are annotated as non-amplified. However, gene set analysis with downstream *MYCN* targets from literature6 shows continued enrichment of *MYCN* targets in these samples despite the absence of *MYCN* amplification (with all samples, medNES = 1.22, KW adj. p-val = 8.64e-17, Dunn adj. p-val < 1.00e-04, with only *MYCN*-non amplified tumors in T064, medNES = 1.07, KW adj. p-val = 7.62e-11) (Fig. 5e, 5h). In line with previous studies, which identified a correlation between *MYCN* amplificated tumors and mitosis-karryohexis index (MKI)7, we observe significantly more samples carrying high MKI (13/33, χ2 p-val = 1.03e-02) in **T064** when compared to the other classes.

Both **T064 NEBLA MYCN** and **T065 NEBLA ADR TERT** are characterized by significant *TERT* overexpression compared to **T062 NEBLA ERBB2** and **T063 NEBLA ADR NTRK1**. Previous studies have explored the associations between telomere maintenance and prognosis in neuroblastoma, identifying three mutually exclusive pathways which are enriched in high risk tumors: *ATRX* upregulation, *MYCN* amplification, and *TERT* rearrangements, each of which result in the overexpression of *TERT*8,9. Indeed, both **T064** and **T065** have enrichment of alternative telomere lengthening pathways (KW adj. p-val < 2.06e-14)10,11(Fig. S22a). *TERT* rearrangements are associated with the upregulation of *SLC6A18* and *SLC6A19*, genes neighbouring *TERT* on the distal side of its breakpoint. Both these genes were significantly upregulated in **T065 NEBLA ADR TERT** (SLC6A18, median logFC= 3.77, FDR ≤ 3.78e-06; SLC6A19, median logFC= 3.88, FDR < 2.96e-03), but not in **T064 NEBLA MYCN,** suggesting **T065 NEBLA ADR TERT** may be comprised of *TERT*-rearranged neuroblastomas. *CCND1* amplification has been observed concurrently with *TERT* rearrangements in neuroblastomas12 and is highly upregulated in **T065** (median logFC= 1.09, FDR ≤ 5.67e-06). We find no significant differences in expression of *ATRX* between clusters. **T065** exhibits the lowest expression of gene sets related to adrenal development (medNES = 0.39, KW adj. p-val = 2.14e-12, Dunn adj. p-val < 1.00e-04)3,4, as well as low expression of mature chromaffin markers such as *EPAS1* (median LogFC= -1.09, FDR ≤ 4.412e-02)13, suggesting this cluster is formed of poorly differentiated neuroblastomas. To further support this hypothesis, we observe here the highest median stemness score (0.81) among all classes, while a non-negligible immune infiltration score is also observed (.45) (Fig. 5c, 5e).

Hypermethylation of the *TERT* locus in high-risk neuroblastomas has been reported in literature14. In line with this observation, **T065 NEBLA ADR TERT** shows enrichment for DNA methylation pathways (medNES ≥ 1.04, KW = 2.79e-14), and numerous histone modification gene sets: notably methylation of *H3K4*, a transcriptional inducer (medNES ≥ 1.02, KW p-val = 2.97e-13), and methylation of *H3K9*, a known silencer of tumor suppressors (medNES ≥ 1.15, KW p-val = 1.78e-12)3,4,15,16. Furthermore, **T065** is highly enriched for PRC2 complex activity (medNES ≥ 1.06, KW = 1.15e-14, Dunn adj. p-val < 1.00e-03)17. Though PRC2 activity is usually examined in the context of *MYCN* amplification18–20, this data supports recent evidence of a PRC2 signature independent of *MYCN* amplification in high-risk neuroblastoma21.

Both **T064 NEBLA MYCN** and **T065 NEBLA ADR TERT** show a characteristic enrichment of COSMIC signature 18 gene set (KW adj. p-val ≤ 4.87e-12)22, associated with reactive oxygen species, when compared to **T061** and **T062** (Fig. S22a)**.** This signature has been suggested to be causative of point mutations in neuroblastoma and has been associated with *MYCN* amplification, and increased expression of electron-transport, ribosomal, and mitochondrial genes. The latter, in particular, follows from a 17q gain, a prognostic marker for poor outcome22,23. We observe significant enrichment of chromosome 17q gene sets in **T065 NEBLA ADR TERT** (medNES ≥ 1.20, KW adj p-val ≤ 5.86e-04)24. Partial loss of 11q (q21-25), associated with *TERT* rearrangements in literature25, is also present in **T065** (medNES ≤ 6.56e-01, KW p-val ≤ 1.03e-05).

Our data support the existence of two major phenotypes with very poor outcome in canonically high-risk neuroblastoma, one driven by *MYCN* activation, the other by *TERT* activation independent of *MYCN*. While genomic rearrangements for samples in **T065** were not reported, neuroblastomas lacking genomic rearrangements at the *TERT* locus, but expressing a *TERT* high phenotype, have been reported in literature25,26. We speculate **T065** may also include samples with non-lesional *TERT* activation, potentially involving gain of 17q and loss of 11q.

The four neuroblastoma classes also show a significant segregation of samples by ploidy level. **T063** contains most hyperdiploid tumors (34/46, *χ*2 p-val = 4.01e-03) and consequently has the highest median ploidy value (1.285, KW adj. p-val = 6.56e-03) (Fig. S22b). **T062 NEBLA ERBB2** and **T064 NEBLA MYCN** have the lowest median value (1.00 both), with the former having a majority of diploid members (9/12, *χ*2 p-val = 4.01e-03) (Fig. S22b).

Furthermore, we observe a significant separation between the Kaplan-Meier fitted curves of overall survival rates (OS, available only for TARGET data, lrt p-val = 1.36e-02 at 4948 days) (Fig. 5b). As expected, patients with tumors in **T064 NEBLA MYCN** have the poorest outcome, followed by **T065 NEBLA ADR TERT**, **T062 NEBLA ERBB2** and finally **T062 NEBLA ADR NTRK1**. This is consistent with literature: improved survival was documented for *ERBB2* overexpressing neuroblastomas,2 although here observed only against other COG high-risk samples.

Recent work investigated linage and developmental differences across neuroblastomas and identified two major groups defined by distinct expression modules driven: a sympathoadrenal identity and neural-crest cell-like (NCC-like)/mesenchymal identity. These developmental states are mediated epigenetically through the action of of super-enhancer and super-enhancer related transcriptional factor networks. Neuroblastomas can move from one to the other identity under selective pressure, induced by therapy or epigenetic alterations and often contain intermixed populations27,28. We thus decided to search for overlaps between these developmental identities and our clusters.

Interestingly, we observe the characteristic signature of both lineages in all clusters although expressed to different degrees. **T062 NEBLA ERBB2** in particular is committed to an NCC-like linage as shown by high expression of NCC-like and mesenchymal markers (medNES = 1.57, KW adj. p-val = 3.69e-07, Dunn adj. p-val < 1.00e-03) against all other classes (Fig. 5g, Fig. S22a). These in turn show enrichment noradrenergic and sympathoadrenal gene sets (KW adj. p-val ≤ 1.19e-09)27–29(Fig. 5g, Fig. S22a).

Samples belonging to **T065 NEBLA ADR TERT** seem to be the most committed to the sympathoadrenal specification (Dunn adj. p-val < 0.05 against **T062** and **T064)** (Fig. S22b). **T064 NEBLA MYCN** shows high variation in the values of its enrichment scores for both linages (ssGSEA) (Fig. 5g, Fig. S22b). The expression profile downstream of *MYCN* amplification may have overridden the original identity signal, or alternatively mixed-lineage populations are common in *MYCN* amplified samples.

**Bibliography**

1. Abel, F. *et al.* A 6-gene signature identifies four molecular subgroups of neuroblastoma. *Cancer Cell Int.* **11,** 9 (2011).

2. Izycka-Swieszewska, E. *et al.* Prognostic significance of HER2 expression in neuroblastic tumors. *Mod. Pathol.* **23,** 1261–1268 (2010).

3. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

4. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47,** D330–D338 (2019).

5. Schaefer, C. F. *et al.* PID: the pathway interaction database. *Nucleic Acids Res.* **37,** D674-9 (2009).

6. Valentijn, L. J. *et al.* Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. *Proc. Natl. Acad. Sci. USA* **109,** 19190–19195 (2012).

7. Teshiba, R. *et al.* Age-dependent prognostic effect by Mitosis-Karyorrhexis Index in neuroblastoma: a report from the Children’s Oncology Group. *Pediatr. Dev. Pathol.* **17,** 441–449 (2014).

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

9. Duan, X.-F. & Zhao, Q. TERT-mediated and ATRX-mediated Telomere Maintenance and Neuroblastoma. *J. Pediatr. Hematol. Oncol.* **40,** 1–6 (2018).

10. Nabetani, A. & Ishikawa, F. Alternative lengthening of telomeres pathway: recombination-mediated telomere maintenance mechanism in human cells. *J. Biochem.* **149,** 5–14 (2011).

11. Jassal, B. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **48,** D498–D503 (2020).

12. Fransson, S. *et al.* Whole-genome sequencing of recurrent neuroblastoma reveals somatic mutations that affect key players in cancer progression and telomere maintenance. *Sci. Rep.* **10,** 22432 (2020).

13. Westerlund, I., Shi, Y. & Holmberg, J. EPAS1/HIF2α correlates with features of low-risk neuroblastoma and with adrenal chromaffin cell differentiation during sympathoadrenal development. *Biochem. Biophys. Res. Commun.* **508,** 1233–1239 (2019).

14. Olsson, M., Beck, S., Kogner, P., Martinsson, T. & Carén, H. Genome-wide methylation profiling identifies novel methylated genes in neuroblastoma tumors. *Epigenetics* **11,** 74–84 (2016).

15. Ke, X.-X. *et al.* Inhibition of H3K9 methyltransferase G9a repressed cell proliferation and induced autophagy in neuroblastoma cells. *PLoS One* **9,** e106962 (2014).

16. Durinck, K. & Speleman, F. Epigenetic regulation of neuroblastoma development. *Cell Tissue Res.* **372,** 309–324 (2018).

17. Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2,** 117–120 (2001).

18. Corvetta, D. *et al.* Physical interaction between MYCN oncogene and polycomb repressive complex 2 (PRC2) in neuroblastoma: functional and therapeutic implications. *J. Biol. Chem.* **288,** 8332–8341 (2013).

19. Tsubota, S. *et al.* PRC2-Mediated Transcriptomic Alterations at the Embryonic Stage Govern Tumorigenesis and Clinical Outcome in MYCN-Driven Neuroblastoma. *Cancer Res.* **77,** 5259–5271 (2017).

20. Chen, L. *et al.* CRISPR-Cas9 screen reveals a *MYCN*-amplified neuroblastoma dependency on EZH2. *J. Clin. Invest.* (2018).

21. Yang, X. H., Tang, F., Shin, J. & Cunningham, J. M. Incorporating genomic, transcriptomic and clinical data: a prognostic and stem cell-like MYC and PRC imbalance in high-risk neuroblastoma. *BMC Syst. Biol.* **11,** 92 (2017).

22. Brady, S. W. *et al.* Pan-neuroblastoma analysis reveals age- and signature-associated driver alterations. *Nat. Commun.* **11,** 5183 (2020).

23. Kucab, J. E. *et al.* A compendium of mutational signatures of environmental agents. *Cell* **177,** 821–836.e16 (2019).

24. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48,** D682–D688 (2020).

25. Roderwieser, A. *et al.* Telomerase is a prognostic marker of poor outcome and a therapeutic target in neuroblastoma. *JCO Precis. Oncol.* 1–20 (2019). doi:10.1200/PO.19.00072

26. Ackermann, S. *et al.* A mechanistic classification of clinical phenotypes in neuroblastoma. *Science* **362,** 1165–1170 (2018).

27. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49,** 1408–1413 (2017).

28. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.* **49,** 1261–1266 (2017).

29. Tomolonis, J. A., Agarwal, S. & Shohet, J. M. Neuroblastoma pathogenesis: deregulation of embryonic neural crest development. *Cell Tissue Res.* **372,** 245–262 (2018).