A human neural crest model reveals the developmental impact of neuroblastoma-associated chromosomal aberrations

Ingrid M. Saldana-Guerrero[1,2,3,*], Luis F. Montano-Gutierrez[4,*], Katy Boswell[1,2,$], Christoph Hafemeister[4,$], Evon Poon[5,$], Lisa E. Shaw[6,$], Dylan Stavish[1,2,$], Rebecca A. Lea[1], Sara Wernig-Zorc[4], Eva Bozsaky[4], Irfete S. Fetahu[4], Peter Zoescher[4], Ulrike Pötschger[4], Marie Bernkopf[4,7], Andrea Wenninger-Weinzierl[4], Caterina Sturtzel[4], Celine Souilhol[1,2,8], Sophia Tarelli[1,2], Mohamed R. Shoeb[4], Polyxeni Bozatzi[4], Magdalena Rados[4], Maria Guarini[9], Michelle C. Buri[4], Wolfgang Weninger[6], Eva M. Putz[4], Miller Huang[10,11], Ruth Ladenstein[4], Peter W. Andrews[1], Ivana Barbaric[1,2], George D. Cresswell[4], Helen E. Bryant[3], Martin Distel[4], Louis Chesler[5,+], Sabine Taschner-Mandl[4,+], Matthias Farlik[6,+], Anestis Tsakiridis[1,2,#], Florian Halbritter[4,#]

[1] Centre for Stem Cell Biology, School of Biosciences, The University of Sheffield, Sheffield, UK
[2] Neuroscience Institute, The University of Sheffield, Sheffield, UK
[3] Sheffield Institute for Nucleic Acids (SInFoNiA), School of Medicine and Population Health, The University of Sheffield, Sheffield, UK
[4] St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria
[5] Division of Clinical Studies, The Institute of Cancer Research (ICR) & Royal Marsden NHS Trust, London, UK
[6] Medical University of Vienna, Department of Dermatology, Vienna, Austria
[7] Labdia Labordiagnostik GmbH, Vienna, Austria
[8] Biomolecular Sciences Research Centre, Department of Biosciences and Chemistry, Sheffield Hallam University, Sheffield, UK
[9] CeMM Research Center for Molecular Medicine of the Austrian Academy of Science, Vienna, Austria
[10] Children's Hospital Los Angeles, Cancer and Blood Disease Institutes, and The Saban Research Institute, Los Angeles, CA, USA
[11] Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

*,$,+ These authors contributed equally to this work.
# Equally contributing senior authors.

Correspondence to a.tsakiridis@sheffield.ac.uk (A.T.) or florian.halbritter@ccri.at (F.H.)

Abstract

Early childhood tumours arise from transformed embryonic cells, which often carry large copy number alterations (CNA). However, it remains unclear how CNAs contribute to embryonic tumourigenesis due to a lack of suitable models. Here we employ female human embryonic stem cell (hESC) differentiation and single-cell transcriptome and epigenome analysis to assess the effects of chromosome 17q/1q gains, which are prevalent in the embryonal tumour neuroblastoma (NB). We show that CNAs impair the specification of trunk neural crest (NC) cells and their sympathoadrenal derivatives, the putative cells-of-origin of NB. This effect is exacerbated upon overexpression of *MYCN*, whose amplification co-occurs with CNAs in NB. Moreover, CNAs potentiate the pro-tumourigenic effects of *MYCN* and mutant NC cells resemble NB cells in tumours. These changes correlate with a stepwise aberration of developmental transcription factor networks. Together, our results sketch a mechanistic framework for the CNA-driven initiation of embryonal tumours.

Keywords: neuroblastoma, copy number alterations, tumourigenesis, stem cells, neural crest

Cancers in early childhood are driven by sparse genetic aberrations arising *in utero*, which are thought to lead to defective differentiation and uncontrolled proliferation[1–4]. Most tumours harbour large genomic rearrangements and chromosomal copy number alterations (CNA), which co-occur with mutations in tumour suppressors or tumourigenic transcription factors (TF)[5,6]. The mechanistic interactions between different mutations and early developmental processes are likely foundational drivers of tumour heterogeneity. However, since visible tumours are only detected long after their initiation, early mutation-driven interactions leading to the healthy-to-tumour transition have remained largely intractable.

Neuroblastoma (NB) is the most common extra-cranial solid tumour in infants and an archetypal "developmental cancer"[7–9]. NB tumours are usually found in the adrenal gland or sympathetic ganglia, tissues derived from the trunk neural crest (NC) lineage during embryonic development[10,11], and studies using transgenic animal models and transcriptome analysis have anchored NB tumourigenesis in impaired sympathoadrenal differentiation of trunk NC cells[12–23]. CNAs such as gains of the long arms of chromosomes 17 (chr17q) and 1 (chr1q) have been identified in the majority (up to 65%) of NB tumours[24–28] and their emergence is considered an early tumourigenesis "priming" event[29]. Chr17q/1q gains often co-occur with amplification of the *MYCN* oncogene[24,28,30–33] (at least one CNA in >95% of *MYCN*-amplified tumours[34]), suggesting they may jointly contribute to tumourigenesis. However, despite our advanced understanding of the genetic and developmental origin of NB, it remains unclear to date how CNAs disrupt embryonic cell differentiation and lead to NB initiation.

Here, we used a human embryonic stem cell (hESC)-based model to experimentally dissect the links between NB-associated CNAs, *MYCN* amplification, and tumour initiation. We interrogated the stepwise specification of trunk NC and sympathoadrenal lineages using directed differentiation of isogenic hESC lines with chr17q/1q gains and inducible *MYCN* overexpression. We found that (i) CNAs derail differentiation by potentiating immature NC progenitor phenotypes. Combining CNAs with *MYCN* overexpression completely disrupted normal NC differentiation; (ii) Mutant NC cells acquired tumourigenic hallmarks *in vitro*, the capacity to form tumours in xenografts, and resemble distinct subpopulations of heterogeneous NB tumours; (iii) An extensive re-wiring of chromatin connects the observed transcriptional and functional aberrations with a dysregulated network of developmental TFs. Collectively, our data put forward a CNA-driven distortion of trunk NC and sympathoadrenal differentiation as a priming mechanism for subsequent MYCN-induced tumour initiation.

Results

Differentiation of human embryonic stem cells recapitulates key stages of trunk NC and sympathoadrenal development

To model the initiation stage and cell types relevant to NB tumourigenesis, we turned to an *in vitro* modelling approach. We have previously described an efficient strategy to produce human trunk NC, sympathoadrenal progenitors, and sympathetic neurons from hESCs[35,36]. Our protocol involves treatment with defined cocktails of signalling pathway agonists/antagonists that induce neuromesodermal-potent axial progenitors (NMPs) at day 3 of differentiation (D3)[37], and subsequently steer NMPs toward trunk NC cells (D9) and their sympathoadrenal derivatives (>D14). At D19, the protocol yields catecholamine-producing sympathetic neurons marked by peripherin-expressing axons[35] (**Fig. 1a, Supplementary Fig. 1**).

As a prerequisite for studying the effects of CNAs on trunk NC differentiation, we first needed to define a molecular roadmap of normal hESC differentiation as a control. Therefore, we employed our protocol for the differentiation of karyotypically normal hESCs (H7[38]; 46XX) and performed droplet-based single-cell RNA sequencing (scRNA-seq) at key differentiation stages (D0 ≈ hESCs, D3 ≈ NMPs, D9 ≈ trunk NC, D14 ≈ sympathoadrenal progenitors, D19 ≈ early sympathetic neuron) and intermediate/late time points (D4, D10, D12, D28) to examine the resulting cell populations (up to five replicates each; **Supplementary Table 1**). We obtained 29,857 cells that passed quality control, which we allocated to 14 distinct clusters (C1-C14) (**Fig. 1b; Supplementary Figs. 2a-g**). We bioinformatically annotated these cell clusters using two complementary approaches: (i) by identifying characteristic marker genes (**Fig. 1c; Supplementary Fig. 2h; Supplementary Table 2**), and (ii) by mapping our data to single-cell transcriptomes of trunk NC derivatives in human embryos[15,16] (**Figs. 1d-f, Supplementary Figs. 2i,j**). This strategy identified cells at different stages of trunk NC development, including NMP-like cells (marked by CDX1/2, NKX1-2, and FGF signalling-associated transcripts[37]; cluster C2 in **Fig. 1c, Supplementary Table 2**) and later cell populations of a predominantly trunk axial identity (**Supplementary Fig. 2h**) exhibiting characteristics of Schwann cell precursors (SCP), sympathoblasts (SYM), as well as mesenchymal features (MES). For example, D9 cells split into subpopulations expressing markers of trunk NC/early SCPs (C3; e.g., *SOX10*[16], **Fig. 1c**; weak SCP-like signature, **Fig. 1e**) and sensory neurons (C5; *ONECUT1*[39], **Fig. 1c**; weak SYM-like signature, **Fig. 1f**). At D14, cells started to assume a sympathoadrenal/autonomic progenitor (C8; *ASCL1*) or mesenchymal (C11; *FN1*) identity, and by day D19, we observed three distinct fractions: mature SCP-like cells (C9; *POSTN*[40]; strong SCP signature), autonomic sympathoblasts (C12-C14; *PHOX2A/B, ELAVL4*[16,41]; strong SYM signature), and MES-like cells (C11; *COL1A1, FN1*). This is in line with findings showing that trunk NC and SCPs are competent to generate mesenchyme[40,42,43]. Interestingly, we also found cells at the intersection of MES and SYM identity, as observed in mice[39] and NB cell lines[44–48] (**Supplementary Fig. 3; Supplementary Table 3**). After 4 weeks (D28), we also

observed some cells with a partial chromaffin-like cell identity (part of C14; *CHGA*$^{+}$, PMNT^{-}) (**Fig. 1d**).

Together, these data confirm that our hESC-based model successfully captures trunk NC and sympathoadrenal cells as found in embryos during the onset of NB tumourigenesis. Moreover, they reveal two major developmental branching events: (i) an early commitment of trunk NC toward a sensory neuron fate; (ii) the late generation of multipotent SCP/sympathoadrenal progenitors, which subsequently give rise to three distinct cell types: mature SCPs, MES, and SYM.

CNAs and *MYCN* cumulatively disrupt human trunk NC differentiation

Having established a reliable model of trunk NC lineages relevant for NB pathogenesis, we next asked how chr17q and chr1q gains and their interplay with *MYCN* overexpression, which often co-occurs with CNAs in NB[24,28,30–34], influence NC development. To this end, we employed two clonal isogenic hESC lines with NB-associated CNAs that were acquired by H7 hESCs ('WT'; used in **Fig. 1**) as a result of culture adaptation[49] (**Fig. 2a; Supplementary Fig. 4a**): (i) a gain of chromosome arm 17q11-qter ('17q')[50], and (ii) an additional gain of chr1q in the 17q background ('17q1q'). Whole-exome sequencing of 17q and 17q1q cells compared to the parental H7 hESCs revealed a small number of additional mutations and a loss of a small region in chromosome 2 (**Supplementary Figs. 4b; Supplementary Tables 4 and 5**). For brevity, we labelled the cell lines by their major CNAs, which overlap regions commonly gained in NB tumours[51] (**Supplementary Fig. 4c**). 17q1q hESCs were engineered to include a Doxycycline (DOX)-inducible *MYCN* expression cassette to mimic *MYCN* amplification in a temporally controlled manner ('17q1qMYCN'). DOX treatment of 17q1qMYCN resulted in robust induction of MYCN, similar to expression levels in MYCN-amplified tumours (**Supplementary Figs. 4d-f**). In our experiments, we induced *MYCN* overexpression at D5 (when cells adopt a definitive NC identity[35]) to avoid bias toward central nervous system differentiation, as seen upon *MYCN* overexpression in earlier pre-NC progenitors[52].

Equipped with these three isogenic 'mutant' hESC lines, we performed differentiation toward trunk NC and carried out scRNA-seq as described above, yielding a combined dataset comprising 95,766 cells (**Supplementary Table 1**). To assess how differentiation was affected in each mutant cell line, we first focused on stages D9, D14, and D19 for which we had data from all four experimental conditions, and bioinformatically mapped the transcriptomes of mutant cells to our reference of normal trunk NC differentiation (cp. **Fig. 1b**). While many 17q cells intertwined with all WT cell types (~98% matching the cognate WT stage), fewer 17q1q and 17q1qMYCN cells advanced beyond WT D14 (only ~48% and 22% matched with WT, respectively; **Fig. 2b**). Only ~4% of 17q1qMYCN cells mapped to mature cell types (**Fig. 2b**). Altogether, at this level of resolution, we found no evidence that 17q affected differentiation. In contrast, 17q1q and 17q1qMYCN cells matched WT cells of earlier developmental stages, suggesting impaired differentiation (**Fig. 2c**).

Next, we tested whether the cell types induced from mutant hESCs still truthfully recapitulated *in vivo* cell types as seen for WT. Mapping mutant cells onto the same human embryonic adrenal gland reference[16] identified proportionally fewer SYM- and MES-like cells in 17q1q and 17q1qMYCN (**Figs. 2d,e**). For cells mapped to the respective cell types, we observed a slightly stronger SCP signature in 17q and 17q1q, while the expression of both MES and SYM genes was weaker relative to the WT (**Fig. 2f**). In 17q1qMYCN, the expression of all signatures was weak, suggesting a failure to fully specify the expected cell types (**Figs. 2d-g**). Consistently, antibody staining for SOX10 and HOXC9 and flow cytometry revealed depletion of SOX10+ trunk NC cells in 17q1qMYCN cultures (**Fig. 2h; Supplementary Fig. 5**). The reduced ability of 17q1qMYCN hESCs to differentiate toward trunk NC derivatives was also reflected by their failure to generate PERIPHERIN-positive neuronal axons (**Supplementary Fig. 4g**). A similar, albeit milder effect was observed upon DOX-induced *MYCN* overexpression at later timepoints (**Supplementary Fig. 4h)**.

Differential analysis identified 941 (17q vs. WT), 2,039 (17q1q vs. WT), and 5,915 (17q1qMYCN vs. WT) differentially expressed genes (DEGs) at D9 (**Supplementary Table 6**). As expected, many upregulated genes were located within the known CNAs (41.4% within chr17q for 17q cells; 18.7% within chr17q and 25.6% within chr1q for 17q1q cells; **Supplementary Fig. 6a**). Pathway analysis indicated an enrichment of genes related to E2F and MTORC1 signalling components for DEGs on chr17q (e.g., *BRCA1*, *NME1*), and of apoptosis-related and members of the p53 pathway on chr1q (e.g., the anti-apoptotic regulator *MCL1*; **Figs. 3a-c; Supplementary Table 7**). Notably, genes upregulated in 17q1q also include the p53 inhibitor *MDM4*[53] (**Supplementary Table 6**). These perturbed pathways may contribute to deregulation of expression of genes outside CNAs (e.g., upregulation of MYC targets and oxidative phosphorylation, and downregulation of G2-M checkpoint-related genes in 17q1qMYCN; **Fig. 3a**). To better resolve the molecular impact of each mutation, we integrated all datasets into a joint projection of WT and aberrant trunk NC differentiation (**Fig. 3d; Supplementary Figs. 6b-h**; **Supplementary Table 8**). The strongest changes were found in 17q1qMYCN, which formed disconnected cell clusters not normally produced in our protocol. To delineate the stepwise alteration of transcriptional programmes, we placed cells from D9 on a spectrum from WT to 17q1qMYCN by scoring each cell between 0 and 1 based on the fraction of mutant cells among its gene expression neighbours ("mutation score"; **Fig. 3e**). This allowed us to identify four sets of genes (D9_1–D9_4) that were correlated with mutations (**Fig. 3f, Supplementary Figs. 7a,b; Supplementary Table 9**): Gain of CNAs led to a decrease in expression of genes (gene set D9_3, **Fig. 3f***)* involved in NC development (e.g., *TFAP2B*[54,55]) and gradual induction of genes (D9_4, **Fig. 3f**) associated with NC/NB cell migration (e.g.. *ZIC2*, *HOXD3*, *GPC3*[56–58]). *MYCN* overexpression in 17q1qMYCN further repressed genes related to NC development (D9_2; e.g., WNT-antagonist *SFRP1*[59] and nuclear receptors *NR2F1/2*[60]) and led to upregulation of MYCN targets implicated in NB (D9_1; e.g., *NME1* on 17q[61]; **Supplementary Table 9)**. Interestingly, we had also found *SFRP1* and *NR2F1* to mark the SYM-MES transition state in WT differentiating sympathoadrenal cells (cp.

Supplementary Table 3). Moreover, we found that many of genes that are upregulated in 17q1qMYCN (D9_1) were also highly expressed in NB tumours (**Supplementary Fig. 7c**).

We further sought to disentangle the relative contributions of *MYCN* overexpression and CNAs to the observed differentiation block phenotype in 17q1q cells. To this end, we generated additional cell lines derived from WT and 17q H7 hESCs by equipping each with a DOX-inducible *MYCN* construct (WTMYCN, 17qMYCN; **Supplementary Fig. 8a**). Moreover, we introduced the same inducible *MYCN* expression cassette into a second female hESC cell line[38] (H9) which had independently acquired chr17q and 1q gains (H9-WT, H9-17q1q, H9-17q1qMYCN). The differentiation trajectories of these cell lines in the presence and absence of *MYCN* overexpression were interrogated using split-pool single-cell RNA-seq. To ensure consistency, we also included the previously analysed H7 cell lines (WT, 17q, 17q1q, 17q1qMYCN) and performed 2-4 replicate experiments. We obtained a total of 45,546 cells (all D9) post-QC and mapped each dataset onto our WT reference, as we had done before (**Supplementary Fig. 8b**). Starting with gain of chr17q, we found a reduction in cells allocated to sensory neuronal differentiation (cluster C4 in **Figs. 1b,c**) and instead a slight increase in a transitional progeny (C7). With chr1q, we also saw an increase of cells in C10. On top of these changes, *MYCN* overexpression led to most cells allocating to earlier developmental stages including clusters C2 and C3 (**Supplementary Figs. 8b,c**) – reflecting the differentiation block we had observed before (cp. **Figs. 2b,c**). The observed changes were consistent for derivatives of both parental hESC lines (H7 and H9). Analysis of marker gene expression associated with the altered cell clusters (C2, C4, C7, C10) in the different mutant cell lines identified an upregulation of genes like *AZIN1* in all *MYCN*-overexpressing cells that was not active in their wild-type or CNA-only counterparts (**Supplementary Fig. 8d**). Conversely, these cells downregulated developmental regulators. For instance, even though MYCN-overexpressing cells still expressed remnant gene signatures leading them to map to differentiating WT cell clusters they downregulated genes in developmental pathways like *HHIP* in cluster C7 or *WNT1* in C10. Moreover, the neurogenic capacity of the mutant trunk NC cells (as reflected by the presence of PERIPHERIN-positive neuronal axons) was found to be disrupted by *MYCN* overexpression primarily in the presence of CNAs, with the strongest phenotype being observed in the presence of both chr17q and chr1q gains (**Supplementary Fig. 8e**), in line with our earlier findings. Collectively, these data indicate that CNAs potentiate the *MYCN*-driven block to the induction of a trunk NC/sympathoadrenal identity.

We conclude that NB-associated CNAs alter the differentiation landscape of hESC-derived trunk NC lineages by promoting transitional progenitor states at the expense of mature sympathoadrenal cell types. In conjunction with *MYCN* elevation, they block differentiation and trigger atypical transcriptional programmes incompatible with normal trunk NC development.

Impaired trunk NC differentiation correlates with acquisition of tumourigenic hallmarks

We observed that ectopic *MYCN* induction altered the morphology of cultures by D14 only in the presence of CNAs as cells lost their ability to spread out and form neurites, and 17q1qMYCN cells even formed tight, dome-like colonies (**Fig. 4a**). As this phenomenon is reminiscent of loss of contact inhibition, a cancer hallmark, we next examined whether CNAs/*MYCN* overexpression led to further cellular changes that are typical of tumourigenesis. We first carried out cell cycle analysis of trunk NC cells (D9) generated from each *MYCN*-overexpressing hESC line (from WT/17q/17q1q backgrounds) by monitoring EdU (5-ethynyl-2´-deoxyuridine) incorporation via flow cytometry. We found a significant increase in the proportion of cells in S-phase only when *MYCN* overexpression was combined with CNAs ($p = 0.0233$ and $p = 0.0073$ respectively; two-way ANOVA; **Fig. 4b**) indicating altered cell cycle and increased replication similar to NB tumours and cell lines[62–64]. Immunofluorescence analysis of Ki-67 expression further showed that 17q1qMYCN and 17qMYCN cultures exhibited a higher proliferation rate by D14 compared to their CNA-only counterparts ($p < 0.0001$ and $p = 0.0078$, respectively; two-way ANOVA; **Fig. 4c**). We next tested how CNAs/*MYCN* influenced colony formation, another hallmark of tumourigenesis. Low-density plating of trunk NC cells (D9) and image analysis showed significantly increased clonogenicity ($p = 0.0109$; two-way ANOVA) exclusively in 17q1qMYCN cells (**Fig. 4d**). DOX treatment of the unedited parental wild type and chr17q gain cell lines had no effect (**Supplementary Fig. 9a**).

Previous work has indicated that *MYCN* overexpression alone is associated with increased apoptosis in early sympathoadrenal cells[65,66] and can trigger tumourigenesis only in combination with additional mutations[67–69]. Therefore, we also examined apoptosis levels during the transition of D9 trunk NC cells toward the SCP/sympathoblast stage (D14) by assessing cleaved Caspase-3 levels using flow cytometry. We found that *MYCN* overexpression indeed resulted in a higher rate of apoptosis in the WT background, while this increase was reversed in 17q cells (**Supplementary Fig. 9b**). However, this was not the case in cultures derived from *MYCN*-overexpressing cells with chr17q1q gains, which exhibited apoptosis levels similar to their *MYCN*-overexpressing WT counterparts (**Supplementary Figs. 9b,c**). This phenomenon may be linked to the presence of both pro- and anti-apoptotic genes in chromosome arm chr1q (cp. **Fig. 3c**) as well as increased DNA damage (assessed by the presence of γH2AX foci) following *MYCN* overexpression specifically in the 17q1q background (**Supplementary Figs. 9d,e**). Interestingly, we detected lower levels of DNA damage in the absence of MYCN overexpression in 17q and 17q1q trunk NC cells at D9 compared to WT controls suggesting a potential protective effect of 17q/1q gains (**Supplementary Figs. 9d).** Moreover, we investigated whether *MYCN*-overexpressing cells from different backgrounds had acquired additional mutations during differentiation. Whole-exome sequencing analysis at D19 of differentiation did not reveal any new large CNAs and detected only few mutations (<10 mutations with variant allele frequency >0.2 between D0 and D19; **Fig. 4e**, **Supplementary Fig. 9f**; **Supplementary Table 4**). None of the discovered mutations have previously been reported in NB, leading us to conclude that the observed phenotypic changes in 17q1qMYCN

were likely a product of the CNAs and *MYCN* overexpression rather than an expansion of new clonal cell populations with additional pathognomonic mutations. Despite an increase of proliferation (cp. **Figs. 4b,c**), *MYCN* overexpression did not yield more high-frequency mutations during the short timeframe of our differentiation experiments, consistent with earlier work in human neuroepithelial stem cells in *vitro* and after xenotransplantation[70] ($p = 0.3458$, two-sided, paired Wilcoxon test, $n = 3$ per group; **Supplementary Fig. 9g**).

Finally, we sought to examine the tumourigenic potential of 17q1q-, 17q1qMYCN- and WT-derived trunk NC (D9 of differentiation) cells *in vivo* by xenografting them into immunodeficient NSG mice. We first injected aliquots of about 1 million cells subcutaneously into the recipient animals ($n = 6$ per cell line) and monitored tumour volume over time. After 3-5 weeks with continuous DOX administration, all 17q1qMYCN-injected mice developed visible tumours at the injection site (**Fig. 5a**). In contrast, neither WT- nor 17q1q-injected control animals displayed any signs of tumours for up to 16 weeks (**Fig. 5a**). Likewise, orthotopic injection into the adrenal gland ($n = 3$ mice per condition) yielded tumour growths visible by magnetic resonance imaging (MRI) after 5 weeks only when *MYCN* overexpression was induced by DOX in 17q1qMYCN-grafted animals (**Fig. 5b,c**). We found that both subcutaneous and adrenal xenograft-derived tumours consisted of undifferentiated, small round cells similar to tumours from transgenic Th-*MYCN* mice[12] (**Supplementary Fig. 10a**). Complementary to our analysis in mice, we also performed exploratory xenografts of the same cell lines in zebrafish larvae. To this end, we labelled our 17q1qMYCN and WT cells at D9 with a fluorescent dye (CellTrace Violet) and injected them into the perivitelline space of zebrafish larvae on day 2 post fertilisation. Consistent with our results in mice, we found that 17q1qMYCN cells survived longer in zebrafish than WT, which had diminished in number at day 1 post injection (dpi) and were completely absent at 3 dpi (**Supplementary Figs. 10b,c**). In contrast, 17q1qMYCN cells survived until 3 dpi, with 16% of larvae even showing an increase in xenotransplant size. For comparison, injection of cells from a *MYCN*-amplified NB cell line (SK-N-BE2C-H2B-GFP[71]) resulted in engraftment with subsequent tumour cell growth in 84% of larvae (**Supplementary Fig. 10d**).

Together, our results demonstrate that CNA-carrying trunk NC cells transit into an undifferentiated pre-tumourigenic state and acquire altered cellular properties reminiscent of cancer hallmarks, such as increased proliferation, clonogenic and tumour formation capacity under the influence of *MYCN* overexpression. Our data also suggest that CNAs enhance the pro-tumourigenic effects of *MYCN*.

In vitro differentiation of mutant hESCs captures NB tumour cell heterogeneity

Given that cells in our *in vitro* model exhibit similarities to NB cells, we asked whether our data could provide insights into cellular heterogeneity in NB tumours. To this end, we collected scRNA-seq data from ten *MYCN*-amplified NB tumours from three independent sources[15,17,23]. For each dataset, we curated $MYCN^+$ tumour cells and bioinformatically mapped these to our reference (**Fig. 6a**). For

example, this approach matched most cells from tumour dataset *Jansky_NB14*[15] to clusters C13 and C14, which correspond to late SYM-like cell states (**Fig. 6b**; cp. **Fig. 1**). A few cells also mapped to clusters C11 and C12, i.e., cells with MES-like characteristics. The observed transcriptional heterogeneity was surprising, given that most tumour cells appeared karyotypically homogeneous (including a chr17q gain) and expressed *MYCN* (**Fig. 6b**).

Extending the *in vitro* reference mapping to all ten tumours portrayed a spectrum of *MYCN*-amplified cells with a majority C13- or C14-like expression profile, and a subset of cells mapping to other differentiating trunk NC cell states (**Supplementary Figs. 11a,b**). We observed apparent differences between studies and tumour types, but to date there is only a limited amount of single-cell data from NB tumours to robustly interpret such heterogeneity. We therefore sought to interrogate a large collection of bulk RNA-seq data from NB tumours (SEQC[72,73]). We first intersected the development-related gene signatures (C1-C14 from **Fig. 1**) with marker genes identified for the tumour cells that had been mapped to those respective clusters (from all 10 investigated samples; **Fig. 6c; Supplementary Table 10**) and labelled each refined signature with an asterisk to distinguish it from the original gene signature (e.g. signature C13* contained genes such as *DLC1* and *RORA;* **Fig. 6c**). Applying these gene signatures to the NB tumour data, we found that expression signatures C5* (sensory neuron-like cells) and C13* (differentiating SYM-like cells), jointly separated *MYCN*-amplified and non-amplified tumours, as well as tumours at different clinical stages (**Fig. 6d**). The C13* signature effectively stratified tumours with a good and poor survival across the entire cohort even when corrected for INSS stage, MYCN amplification status, and age (Cox regression analysis with covariates; **Fig. 6e; Supplementary Table 11**).

Jointly, these observations demonstrate that our *in vitro* model generates cell types that transcriptionally resemble different NB cell subpopulations and that it facilitates the systematic dissection of intra-tumour heterogeneity in NB tumours.

CNAs and *MYCN* disrupt the configuration of NC regulatory circuits during differentiation

NB tumours and cell lines are marked by a 're-wiring' of non-coding regulatory elements (e.g., enhancers) giving rise to tumour-specific regulatory circuitries[44,45,74–78]. We therefore hypothesised that disruption of developmental TFs also underpins the aberrant differentiation observed in our mutant hESCs (cp. **Figs. 2, 3**) and employed the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq)[79] to profile chromatin accessibility in the same samples used for scRNA-seq analysis (n = 51; **Supplementary Table 1**). Chromatin accessibility serves as a proxy for the dynamic regulatory DNA element activity during differentiation. For instance, the promoters of the hESC regulator *POU5F1* and trunk NC regulator *SOX10* were accessible only at D0 and D9, respectively (**Fig. 7a**), while the *PHOX2B* promoter exhibited reduced accessibility in 17q1q and 17q1qMYCN cells at D19 consistent with impaired differentiation (**Fig. 7b**).

Unsupervised analysis of chromatin patterns on a global level showed that WT and 17q hESCs changed consistently throughout differentiation (**Fig. 7c**). In contrast, 17q1q and 17q1qMYCN appeared not to follow the same path as WT in this low-dimensional projection, in line the differentiation defects observed in our previous analyses (cp. **Figs. 2b,c**). To delineate chromatin changes in detail, we performed differential accessibility analysis between all differentiation stages per cell line and between all cell lines at matched stages (**Supplementary Tables 12, 13**). As in our DEG analysis, we found an increasing number of regions with altered accessibility in 17q (n = 477 regions), 17q1q (n = 2,826), and 17q1qMYCN (n = 6,663; **Fig. 7d**). In total, there were 45,583 regions with differential accessibility in at least one comparison, which we divided into nine chromatin modules R1-R9 (**Fig. 7e**). Modules R1-R7 reflect differentiation order, e.g., regions in module R1 were mostly accessible at D0, and R6 comprises regions accessible at D14 and D19. Most changes observed in mutant hESC-derivatives fell within these modules (**Supplementary Figs. 12a,b**). 17q1q and 17q1qMYCN cells failed to close chromatin that is usually specific to D9 (R4, R5) and conversely to open chromatin regions indicative of late sympathoadrenal differentiation (R6, R7; **Supplementary Fig. 12c**). Additionally, modules R8 and R9 comprised regions with reduced and increased accessibility in mutant hESC derivatives, respectively, independently of differentiation stage.

We sought to annotate our chromatin modules by looking for overlaps with genomic regions accessible in human tissues[80–82] (**Supplementary Table 14**). In line with our transcriptome data, we found a stepwise change toward chromatin resembling differentiated tissues, e.g., neural tissues in R3-R5 and mesenchyme/stroma in R6/R7 (**Supplementary Figs. 12d-f**). Next, we examined the identity of genes near the chromatin modules (**Fig. 7f**). For each module, we found enrichments of specific marker genes identified in our scRNA-seq analysis of WT trunk NC differentiation (i.e., clusters C1-C14 from **Figs. 1b,c**). For example, chromatin module R7 (accessible in late differentiation stages, lost in mutants) was linked to clusters C11/C12 (MES-like cells). Next, we examined TF binding motifs in each module to identify potential upstream regulators (**Fig. 7g**). We found an enrichment of known regulators of each developmental stage, e.g., TFs associated with trunk NC in R3-R5 (e.g., SOX10) and with sympathetic neuron development in R6/R7 (e.g., PHOX2A/B)[39]. Moreover, we found enriched overlaps of modules R2/R4/R5/R8/R9 with super-enhancers identified in mesenchymal NB cell lines or adrenergic super-enhancers (in the case of R8), depending on the source annotation used[44,45]. Furthermore, R7 and R9 overlapped super-enhancers associated with subsets of NB tumours[75] with mesenchymal characteristics and with non-*MYCN*-amplified high-risk tumours, respectively (**Fig. 7h**). No significant overlap was found with super-enhancers specific for *MYCN*-amplified NB. Finally, we examined the accessibility of each module across a range of NB cell lines (**Supplementary Fig. 12g**). As expected, we found that modules R1 and R2 (undifferentiated, early embryonic developmental stages) and modules R4 and R5 (early trunk NC to sympathoadrenal differentiation) were not accessible in NB cell lines, while modules R6-R8 (late sympathetic neurons and consistently open in mutants) were often highly accessible in cell lines. Interestingly, R3 (accessible at NMP and NC stage) was

accessible in most NB cell lines examined except in those with mesenchymal characteristics (SK-N-AS and SHEP; **Supplementary Fig. 12g**). Using data from other studies, we found that R6-R8 were also accessible in non-NB cell lines and tissues, while R3 was only found accessible in brain tissue (**Supplementary Fig. 12g**).

Together, our results suggest a systematic reprogramming of chromatin throughout trunk NC differentiation. In cells with CNAs and *MYCN* overexpression, this orderly reconfiguration of chromatin was severely disrupted in a manner similar to NB cells, providing a plausible mechanism for the link between the observed developmental defects and tumour initiation.

CNA/*MYCN*-driven cell identity loss is mediated by sets of developmental transcription factors

Finally, we investigated the links between CNA/*MYCN*-induced changes in chromatin dynamics, gene-regulatory networks, and the distorted differentiation trajectories observed at the transcriptional level. In our scRNA-seq analyses, we had recorded a stepwise alteration of expression from WT to 17q1qMYCN at D9 comprising four gene sets: D9_1 – D9_4 (cp. **Figs. 3f**), which revealed *MYCN*-driven disruptions of early NC and the sensory neuron lineage specification. We hypothesised that these mutation-linked gene sets were also regulated by distinct TFs and therefore employed an algorithm to identify TF targets based on correlated expression patterns[83] (**Fig. 8a**). This analysis identified *NR1D1* and *TFAP4* as putative TF targets of MYCN (**Figs. 8b,c; Supplementary Figs. 13a,b; Supplementary Table 15**). The nuclear receptor *NR1D1* has been shown to correlate with *MYCN* amplification in NB patients[84,85] *and TFAP4* inhibition leads to differentiation of *MYCN*-amplified neuroblastoma cells[86,87], supporting the validity of the inferred target genes.

We intersected the inferred lists of TF targets with the mutation-linked gene sets (D9_1 – D9_4) and found an enrichment (**Fig. 8d; Supplementary Table 16**) of MYCN, NR1D1, TFAP4, and ZIC2 targets in D9_1 (highly expressed in 17q1qMYCN). Conversely, the gene set D9_2 (expressed in WT/17q/17q1q) was enriched for targets of TFs expected at this stage of differentiation, e.g. SOX4/5/10, TFAP2A/B, and nuclear receptors NR2F1/2. The expression of targets of these TFs increased or decreased along the mutational spectrum, corroborating their association with the mutations (**Fig. 8e**). While many TF targets switched expression rapidly with *MYCN* overexpression, others showed a continuous pattern with up-/down-regulation already detectable in 17q and 17q1q, e.g., targets of vagal and early NC regulators HOXB3 and CDX2[88] (up), or of sensory neurogenesis regulator NEUROD1[39] (down). To aid interpretation, we visualised cell-line-specific interactions between TFs and targets as edges in connected network diagrams (**Fig. 8f; Supplementary Fig. 13c**). These diagrams showcased the emergence of a new subnetwork of TFs in 17q1qMYCN that centred on MYCN and incorporated TFs like NR1D1 and TFAP4. In contrast, a subnetwork involving NC-related TFs such as SOX10 and TFAP2A/B was lost in these cells. Intriguingly, downregulation of TFs linked to sensory neuronal development (NEUROD1, ONECUT1) was visible already in 17q cells (**Fig. 8f**), perhaps explaining why sensory neuron-like derivatives were less abundant in 17q cultures **(Fig. 2b).**

In 17q1q, we additionally observed upregulation of TFs related to early posterior NC specification including HOXB3, LEF1, and CDX2, which was partially reversed (HOXB3) upon *MYCN* overexpression (**Fig. 8f**). While many of the TFs implicated in these developmental gene-regulatory networks are not or only weakly expressed in NB tumours (**Supplementary Fig. 14a**), we found that the targets of MYCN-related TFs (based on our analysis) are highly expressed in *MYCN*-amplified tumours (**Supplementary Fig. 14b**). Our analysis also revealed that the targets of 17q/1q-related TFs strongly are expressed in groups of tumours, but we could not determine whether these contained the corresponding CNAs due incomplete annotations.

In summary, our data suggest a subtle rewiring of gene-regulatory networks in CNA-carrying hESCs, which may be linked to the depletion of mature sensory NC derivatives and increased early SCP signature found in our single-cell analyses (cp. **Fig. 2**). Overexpression of *MYCN* resulted in a switch in favour of known NC-linked TFs downstream of MYCN.

Discussion

Although CNAs are a principal genetic hallmark of paediatric cancers, it has remained difficult to determine their exact role in tumour initiation due to the lack of suitable human models. In this study, we used hESCs carrying CNAs that are prevalent in NB (chr17q and chr1q gains). The NC is a transient embryonic tissue that is inaccessible after birth; therefore, hESC differentiation allowed us to experimentally study the effects of these mutations on human sympathoadrenal progenitors, the putative cells-of-origin of NB.

We provide a comprehensive knowledge base of transcriptomic and epigenetic changes in this model on a temporal (i.e., during differentiation) and a genetic (i.e., with different mutations) axis. Our data show that chr17q/1q gains impair trunk NC differentiation and potentiate an SCP-like gene signature. In this aberrant cell state, overexpression of *MYCN* (mimicking *MYCN* amplification commonly found along with chr17q/chr1q gains in NB tumours) leads to a complete derailment of sympathoadrenal lineage specification, and a proliferative, tumour-like cellular phenotype that correlates with the emergence of NB-like tumours *in vivo*. Moreover, chr17q/1q gains were found to enhance the *MYCN*-driven differentiation block and acquisition of tumourigenic hallmarks such as proliferation, clonogenicity and resistance to apoptosis. In line with recent studies[29,89], we speculate that CNAs provide an early selective advantage manifested by the expansion of undifferentiated cells, which act subsequently as a NB-initiating entity upon a second oncogenic hit such as *MYCN* overexpression.

The accumulation of NB-associated lesions correlated with a failure to reprogram chromatin during trunk NC differentiation. Upon gain of chr17q1/q, cells lost TFs associated with sensory differentiation (e.g., *NEUROD1*) and instead upregulated vagal NC TFs like HOXB3 and WNT-related effector LEF1[88,90]. *MYCN* overexpression on top of these CNAs abolished chromatin states indicative of sympathoadrenal differentiation, and instead led to the induction of targets of NR1D1, TFAP4, and

other TFs of the reported NB regulatory circuitry[44,45,74–78]. TFAP4 is a well-established downstream effector of MYCN[86,87], and NR1D1 (Rev-erbα) is a circadian rhythm and metabolic regulator, and a downstream effector of MYCN hyperactivity in NB[84,85]. Together with the appearance of sensory-related signatures in NB tumours (C4* and C5*, **Fig. 6**) our "early *MYCN* onset" scenario reveals previously uncharted disruptions of the early sensory NC lineage, which might complement the currently prevailing model of dichotomic mesenchymal/adrenergic heterogeneity in NB[11,44–46,78,91–94]. Thus, our model will facilitate the functional dissection of these TFs via loss-/gain-of-function approaches to decipher their crosstalk with *MYCN*/CNA-driven tumourigenesis.

Complementing earlier studies using cell lines and animal models[12–14,18,19,22], recent single-cell transcriptomic analyses of NB tumours and metastases[15–17,23] corroborated an origin of NB from neuroblastic, SCP-like progenitors, and highlighted intra-tumour heterogeneity comprising subtypes of tumour cells with adrenergic and mesenchymal properties. In our *in vitro* experiments, we also observed cells expressing signatures of both cell types, suggesting that our model could be useful to experimentally investigate the transition between these and other NB-relevant cell types, providing a new scope into their role in therapy resistance[91]. Furthermore, *MYCN* overexpression (in conjunction with chr17q/1q gains) in nascent trunk NC cells was sufficient to drive tumourigenic traits, suggesting that in some cases NB initiation might occur before SCP/neuroblast emergence and that acquisition of an SCP-like identity may also be a consequence of mutations in earlier stages rather than the origin. We also observed that *MYCN* induction resulted in an apparent block of differentiation when activated at other stages. In future, our cell models will provide the means to dissect the specific effects of *MYCN* at different time points and in specific cell types.

In this study we exploited the phenomenon of culture adaptation of hESCs[49], to obtain sets of cell lines with CNAs that are commonly observed in NB in an otherwise largely isogenic background. Our detailed genetic analyses of the used cell lines revealed other mutations that had naturally arisen in these cell lines (**Supplementary Table 4**), including a point mutation in the BCL6-interacting corepressor *BCOR* (*BCOR*^{L1673F}). *BCOR* mutations have been previously observed in human induced pluripotent stem cell cultures[95,96] and NB patients with *BCOR* mutations exhibit a high frequency of CNAs[89]. *BCOR* mutations have also been reported together with CNAs in other cancers, e.g., retinoblastoma[97]. It would be tempting to speculate that *BCOR* dysfunction might facilitate the ability of cells to tolerate the emergence of certain CNAs; however, to date a causal relationship remains to be established.

Our hESC-based model provides a tractable system for analysing tumour initiation events within disease-relevant human embryonic cell-like populations. In this study, we focused on cell-intrinsic transcriptional regulation since our cultures lack tumour-relevant, non-NC cell types (e.g., immune cells or Schwann cells) and do not recapitulate the structural and physical properties of the human tumour micro- and macroenvironment[98–101]. In the future, it will be possible to combine our

system with 3D co-culture approaches with defined cell types or to use biomimetic scaffolds to emulate cell-cell interactions and extrinsic environmental influences.

In conclusion, this study unravels the developmental effects of NB-associated mutations and proposes the progressive corruption of gene-regulatory networks by CNAs as an early step toward tumour initiation by selection of undifferentiated progenitor phenotypes. Transformation is then triggered by a second hit with *MYCN* overexpression, which tilts cells toward increased proliferation and formation of aberrant cell types. Our data provide a direct link between CNAs that commonly emerge in hESC cultures with impaired differentiation and the acquisition of tumourigenic hallmarks, thus highlighting the importance of rigorous monitoring of such cultures prior to their use in disease modelling or cell therapy application in line with recent recommendations from the International Society for Stem Cell Research[49,102].

Online Methods

Human embryonic stem cell (hESC) cell culture and differentiation

Cell lines and cell culture

We employed H7 and H9 hESCs as karyotypically normal, female WT controls[38]. Use of human embryonic cells has been approved by the Human Embryonic Stem Cell UK Steering Committee (SCSC23-29). Their isogenic chr17q counterparts carry a gain in chromosome 17q (region q27q11) via an unbalanced translocation with chromosome 6 (H7) or a gain of 17q via an unbalanced translocation with chromosome 21 with breakpoints at 17q21 and 21p11.2 (H9)[50,103]. The chr17q1q hESC lines were clonally derived, after their spontaneous emergence following the genetic modification of chr17q hESCs. The H7 chr17q1q-MYCN hESC line was generated by introducing a TetOn-PiggyBac plasmid (PB-TRE3G-MYCN, plasmid#104542, Addgene) carrying the wild-type version of the *MYCN* gene[104] while the H9 chr17q1q-MYCN and H7 WT-MYCN and 17q-MYCN hESC lines were produced using a Tet-On "all-in-one" inducible expression cassette containing the TRE3G promoter driving the expression of *MYCN* with a 2A peptide-linked fluorescent reporter (mScarlet) and a pCAG promoter-driven rtTA3G transactivator[105,106]. Plasmids were introduced via nucleofection using either the Neon NxT Electroporation System (Thermo Fisher Scientific) or the Lonza 4D-Nucleofector System. In the case of the latter, the Amaxa 4D-Nucleofector Basic Protocol for Human Stem Cells was employed with the following modification: $2x10^6$ cells were transfected with 2 µg plasmid in 100 µl Nucleocuvettes. All cell lines were tested regularly for mycoplasma and expression of pluripotency markers. Karyotypic analysis was carried out using G-banding (number of cells examined = 20-30). A rapid qPCR assay was also regularly employed to detect the emergence of common CNAs such as chr17q and 1q gains in our hESC lines[107]. hESCs were cultured routinely in feeder-free conditions at 37°C and 5% CO_2 in E8 media[108] complemented with GlutaMax (Cat# 35050061, Thermo Fisher Scientific) on Vitronectin (Cat# A14700, Thermo Fisher Scientific) or on Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Cat# A1413202, Thermo Fisher Scientific) as an attachment substrate. All hESC lines described in this manuscript are available upon request and completion of a Material Transfer Agreement.

Differentiation toward trunk neural crest

hESC differentiation toward trunk NC and its derivatives was performed using a modified version of the protocol described previously[35,36]. Briefly, hESCs were harvested using StemPro Accutase Cell Dissociation Reagent (Cat# A1110501, Thermo Fisher Scientific) and plated at 60,000 cells/cm^2 in N2B27 medium supplemented with 20 ng/ml of FGF2 (Cat# 233-FB/CF, R&D) and 4 µM of CHIR 99021 (Cat# 4423, Tocris) and 10 µM of Rock Inhibitor (Y-27632) (Cat# A11001, Generon) in a volume of 300µl/cm^2. The N2B27 medium consisted of 50:50 DMEM F12 (Merck Life Science / Neurobasal medium (Gibco) and 1x N2 supplement (Cat# 17502048, Invitrogen), 1x B27 (Cat#17504044, Invitrogen), 1x GlutaMAX (Cat# 35050061, Thermo Fisher Scientific), 1x MEM Non-essential amino acids (NEAA; Cat#11140050, Thermo Fisher Scientific), 50 µM 2-Mercaptoethanol (Cat# 31350010, Thermo Fisher Scientific). After 24 hours, media was refreshed removing the Rock Inhibitor and cells were cultured for a further 2 days in FGF2/CHIR to generate NMPs (300µl/cm^2). NMPs at D3 were then re-plated at 50,000 cells/cm^2 (H7) or 40,000 cells/cm^2 (H9) in neural crest inducing medium consisting of DMEM/F12, 1x N2 supplement, 1x GlutaMAX, 1x MEM NEAA, the TGF-beta/Activin/Nodal inhibitor SB-431542 (2 µM, Cat# 1614, Tocris), CHIR99021 (1 µM, Cat# 4423, Tocris), BMP4 (15ng/ml, Cat# PHC9534, Thermo Fisher Scientific), the BMP type-I receptor inhibitor DMH-1 (1 µM, Cat# 4126, Tocris), 10 µM of Rock Inhibitor (Y-27632) on Geltrex LDEV-

Free Reduced Growth Factor Basement Membrane Matrix (Cat# A1413202, Thermo Fisher Scientific) in a volume of 300µl/cm^2. 48 hours later (D5), media was replaced removing the Rock Inhibitor. Media was refreshed at D7 and D8 increasing volume to 500µl/cm^2. On D5, the expression of MYCN was induced by supplementing the neural crest media with 100ng/ml (H7-17q1q-MYCN), 200ng/ml (H7 WT-MYCN, 17q-MYCN), or 1000 ng/ml (H9-derived lines) of Doxycycline (Cat# D3447, Merck). On D9, cells were re-plated at 150,000-250,000 cells/cm^2 in plates coated with Geltrex (Thermo Fisher Scientific) in the presence of medium containing BrainPhys (Cat# 05790, Stem Cell Technologies), 1x B27 supplement (Cat# 17504044, Invitrogen), 1x N2 supplement (Cat# 17502048, Invitrogen), 1x MEM NEAA (Cat# 11140050, Thermo Fisher Scientific) and 1x Glutamax (Cat# 35050061, Thermo Fisher Scientific), BMP4 (50 ng/ml, Cat# PHC9534, Thermo Fisher Scientific), recombinant SHH (C24II) (50 ng/ml, Cat# 1845-SH-025, R and D) and purmorphamine (1.5 µM, Cat# SML0868, Sigma) and cultured for 5 days (=D14 of differentiation) in a volume of 250µl/cm^2. Media was refreshed daily. For further sympathetic neuron differentiation, D14 cells were switched into a medium containing BrainPhys neuronal medium (Stem Cell Technologies), 1x B27 supplement (Invitrogen), 1x N2 supplement (Invitrogen), 1x NEAA (Thermo Fisher Scientific) and 1x Glutamax (Thermo Fisher Scientific), NGF (10 ng/ml, Cat#450-01 Peprotech), BDNF (10 ng/ml, Cat# 450-02, Peprotech) and GDNF (10 ng/ml, Cat# 450-10, Peprotech) for a further 5-14 days (volume of 300µl/cm^2 changing media every other day). Volume was increased up to 500µl/cm^2, depending on cell density, after day 17 of differentiation.

Immunostaining

Cells were fixed using 4% PFA (P6148, Sigma-Aldrich) at room temperature for 10 minutes, then washed twice with PBS (without Ca^{2+}, Mg^{2+}) to remove any traces of PFA and permeabilised using a PBS supplemented with 10% FCS, 0.1% BSA and 0.5% Triton X-100 for 10 minutes. Cells were then incubated in blocking buffer (PBS supplemented with 10% FCS and 0.1% BSA) for 1 hour at RT or overnight at 4°C. Primary and secondary antibodies were diluted in the blocking buffer; the former were left overnight at 4°C and the latter for 2 hours at 4°C on an orbital shaker. Samples were washed twice with blocking buffer between the primary and secondary antibodies. Hoechst 33342 (H3570, Invitrogen) was added at a ratio of 1:1000 to the secondary antibodies' mixture to label nuclei in the cells. We used the following primary antibodies SOX10 (D5V9L) (Cell Signalling, 89356S, 1:500); HOXC9 (Abcam, Ab50839,1:50); MYCN (Santa Cruz, SC-53993, 1:100); PHOX2B (Santa Cruz, SC-376997, 1:500); MASH1 (ASCL1) (Abcam, Ab211327, 1:100 or Santa Cruz, SC-374104, 1:500); Ki67 (Abcam, Ab238020, 1:100); PERIPHERIN (Sigma-Aldrich, AB1530, 1:400); TH (Santa Cruz, 25269, 1:500); cleaved Caspase 3 (Asp175) (Cell Signalling, 9661S, 1:400), yH2AX (Cell Signalling, S139/9718S, 1:400). Secondary antibodies: Goat anti-Mouse Affinipure IgG+IgM (H+L) AlexaFluor 647 (Stratech (Jackson ImmunoResearch) 115-605-044-JIR, Polyclonal 1:500); Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488 (Invitrogen, A-21206, 1:1000).

Intracellular flow cytometry staining

Cells were detached and resuspended as single cells using StemPro Accutase Cell Dissociation Reagent (Cat# A1110501, Thermo Fisher Scientific) and then counted. Next, 10 million cells/ml were resuspended in 4% PFA at room temperature for 10 minutes. Then cells were washed once with PBS (without Ca^{2+}, Mg^{2+}) and pelleted at 200g. Cells were resuspended in PBS at 10 million/ml and used for antibody staining. Permeabilisation buffer (0.5% Triton X-100 in PBS with 10% FCS and 0.1% BSA) was added to each sample, followed by incubation at room temperature for 10 minutes. Samples were then washed once with staining buffer (PBS with 10% FCS and 0.1% BSA) and pelleted at 200g. Then samples were resuspended in staining buffer containing pre-diluted primary antibodies: SOX10 (D5V9L) (1:500; 89356S, Cell Signalling); HOXC9 (1:50; Ab50839, Abcam); cleaved Caspase 3

(Asp175) (Cell Signalling, 9661S, 1:400). The samples were left at 4°C on an orbital shaker overnight. Then, the primary antibodies were removed, and samples were washed two times with staining buffer. After washings, staining buffer with pre-diluted secondary antibody was added to the samples and incubated at 4°C for 2 hours. The secondary antibodies used were Goat anti-Mouse Affinipure IgG+IgM (H+L) AlexaFluor 647 (Stratech (Jackson ImmunoResearch) 115-605-044-JIR, Polyclonal 1:500); Donkey anti-Rabbit IgG (H+L) Alexa Fluor 488 (Invitrogen, A-21206, 1:1000). Finally, samples were washed once with staining buffer, resuspended in staining buffer and analysed using a BD FACSJazz or a CytoFLEX (Beckman Coulter) flow cytometer. A secondary antibody-only sample was used as a control to set the gating.

Cell cycle analysis

The 5-ethynyl-2´-deoxyuridine (EdU) assay was performed following the manufacturer's instructions (Thermo Fisher Scientific, C10633 Alexa Fluor 488). We used 10μM of Edu for a 2-hour incubation. Cells were analysed in the flow cytometer (BD FACSJazz) using the 405 nm laser to detect the Hoechst/DAPI staining and 488 nm to detect the EdU staining.

Low-density plating

Day 9 trunk NC cells derived from hESCs as described above were harvested and plated at a density of 500 cells/cm^2 in plates pre-coated with Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Cat# A1413202, Thermo Fisher Scientific) in the presence of DMEM/F12 (Sigma-Aldrich), 1x N2 supplement, 1x GlutaMAX, 1x MEM NEAA, the TGF-beta/Activin/Nodal inhibitor SB-431542 (2 μM, Tocris), CHIR99021 (1 μM, Tocris), BMP4 (15ng/ml, Thermo Fisher Scientific), the BMP type-I receptor inhibitor DMH-1 (1 μM, Tocris) and ROCK inhibitor Y-27632 2HCl (10 μM) (300μl/cm^2). The culture medium was replaced the following day with medium containing BrainPhys (Stem Cell Technologies), 1x B27 supplement (Invitrogen), 1x N2 supplement (Invitrogen), 1x NEAA (Thermo Fisher Scientific) and 1x Glutamax (Thermo Fisher Scientific), BMP4 (50 ng/ml, Thermo Fisher Scientific), recombinant SHH (C24II) (50 ng/ml, R and D) and Purmorphamine (1.5 μM, Sigma) (250μl/cm^2). Plates were then incubated at 37°C at 5% CO2. The media was refreshed every 48 hours. After 5 days of culture, cells were fixed (PFA 4%/10min) and stained with Hoechst 33342 (Cat# H3570, Invitrogen) for 5 minutes. Colonies were detected using an InCell Analyser 2200 (GE Healthcare) at a 4X magnification. Images were processed using Cell Profiler.

DNA damage analysis

DNA damage was measured by assessing the phosphorylation state of the histone H2AX on the SerCells were fixed and immunostained using the anti-yH2AX as described above at different time points. Stained cells were imaged using the InCell Analyser 2200 (GE Healthcare) at 40X magnification. Image analysis was performed using a pipeline in CellProfiler that allowed us to detect the number of foci of yH2AX antibody per nuclei.

Quantitative real-time PCR

RNA extractions were performed using the total RNA purification kit (Norgen Biotek, 17200) according to the manufacturer's instructions. cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription kit (ThermoFisher, 4368814). Quantitative real-time PCR was performed using PowerUp SYBR master mix (ThermoFisher, A25780) and run on a QuantStudio 12K Flex (Applied Biosystems).

Mouse experiments

Cell preparation for xenotransplantation

H7 wild type, 17q1q and 17q1qMYCN hESCs were differentiated up to day 9 following the protocol described above. Cells were harvested using Accutase to create a single cell suspension, counted and resuspended with media containing Matrigel before injection.

Mice and in vivo experiments

All animal experiments were approved by The Institute of Cancer Research Animal Welfare and Ethical Review Body and performed in accordance with the UK Home Office Animals (Scientific Procedures) Act 1986, the UK National Cancer Research Institute guidelines for the welfare of animals in cancer research and the ARRIVE (animal research: reporting in vivo experiments) guidelines. Female NSG mice were obtained from Charles River and enrolled into trial at 6-8 weeks of age. Mice were maintained on a regular diet in a pathogen-free facility on a 12h light/dark cycle with unlimited access to food and water.

Subcutaneous xenograft

One million cells with 50% Matrigel were injected subcutaneously into the right flank of NSG mice (female; 6-8 weeks old) and allowed to establish a murine xenograft model. Studies were terminated when the mean diameter of the tumour reached 15mm. Tumour volumes were measured by Vernier caliper across two perpendicular diameters, and volumes were calculated according to the formula $V=4/3\pi$ [(d1+d2)/4]3; where d1 and d2 were the two perpendicular diameters. The weight of the mice was measured every 2 days. Mice were fed with either regular diet or DOX diet (chow containing 20 g of DOX per kg of diet) to induce the expression of *MYCN*.

Orthotopic (adrenal)xenograft

100,000 cells with 50% Matrigel were injected into the right adrenal gland of NSG mice (female; 6-8 weeks old) and allowed to establish a murine xenograft model. Detection of xenografted tumours was performed by magnetic resonance imaging (MRI). The weight of the mice was measured every 2 days. Mice were fed with either standard diet or DOX diet (chow containing 20 g of DOX per kg of diet) to induce the expression of *MYCN*. Magnetic resonance images were acquired on a 1 Tesla M3 small animal MRI scanner (Aspect Imaging). Mice were anesthetised using isoflurane delivered via oxygen gas and their core temperature was maintained at 37 °C. Anatomical T2-weighted coronal images were acquired through the mouse abdomen, from which tumour volumes were determined using segmentation from regions of interest (ROI) drawn on each tumour-containing slice using the Horos medical image viewer.

Pathology

Tissue sections were stained with haematoxylin and eosin (H&E) or specific antibodies (MYCN, Merck; Ki67, BD Pharmingen). Immunohistochemistry was performed using standard methods. Briefly, 5µm sections were stained with antibodies, including heat-induced epitope retrieval of specimens using citrate buffer (pH 6) or EDTA buffer.

Zebrafish experiments

Cell preparation for xenotransplantation

Pre-differentiated neural crest cells were frozen on D7 during their *in vitro* differentiation as described above, shipped, and subsequently thawed in DMEM at room temperature. All cells were retrieved in complete neural crest media as described above and plated onto Geltrex-coated wells in the presence of Rock inhibitor (50µM) for 24 hours. 17q1q cells were additionally treated with DOX (100ng/ml) to induce *MYCN* expression. On D8, media were refreshed, and respective DOX treatment was continued but Rock inhibitor was discontinued. On D9, cells were collected for xenografting experiments and labeled with CellTrace™ Violet (Invitrogen, Thermo Fisher Scientific) for imaging. For this, cells were harvested with Accutase (PAN-Biotech) and resuspended at a concentration of $1*10^6$ cells/ml in PBS. CellTrace™ Violet was added to a final concentration of 5 µM for an incubation time of 10 minutes at 37°C in the dark. The cell-staining mixture was filled up with 5 volumes of DMEM supplemented with 10% FBS and the suspension was incubated for 5 min. After gentle centrifugation (5 min, 500 g, 4°C) the collected cells were resuspended in fresh DMEM medium supplemented with 10% FBS and incubated at 37°C for 10 min. Adhering/ clumping cells were separated via a 35 µm cell strainer. The cell number was adjusted to a concentration of 100 cells/ml in PBS. The freshly stained cells were kept on ice until transplantation. SK-N-BE2C-H2B-GFP cells[71] (a kind gift of F. Westermann) were cultured in RPMI 1640 medium with GlutaMAX™ (Cat# 61870044, Thermo Fisher Scientific) supplemented with 10 % (v/v) fetal bovine serum (Cat# F7524500ML, Sigma), 80 units/ml penicillin, 80 µg/ml streptomycin (Cat# 15140122, Thermo Fisher Scientific), 1 nM sodium pyruvate (Cat# P0443100, PAN-Biotech), 25 mM Hepes buffer (PAN-Biotech) and 8 µl/ml G418. For zebrafish xenotransplantations, the GFP-labelled cells were harvested and resuspended in PBS at a density of 10^5/µl as described above.

Zebrafish strains, husbandry, and xenotransplantation

Zebrafish (*Danio rerio*) were reared under standard conditions in a 14 hours / 10 hours light cycle according to the guidelines of the local authorities (Magistratsabteilung MA58 of the municipal administration of Vienna, Austria) under licenses GZ:565304-2014-6 and GZ:534619-2014-4. For xenotransplantation experiments, the pigment mutant strain $mitfa^{b692/b692}$; $ednrba^{b140/b140}$ was used. $mitfa^{b692/b692}$; $ednrba^{b140/b140}$ embryos raised at 28°C were anaesthetised with Tricaine (0.16 g/l Tricaine (Cat# E1052110G, Sigma-Aldrich), adjusted to pH 7 with 1M Tris pH 9.5, in E3) and xenotransplanted at 2 days post fertilization (dpf) as previously described[109]. For xenotransplantation, a micromanipulator (Cat# M3301R, World Precision Instruments) holding a borosilicate glass capillary (Cat# GB100T-8P, without filament, Science Products) connected to a microinjector (FemtoJet 4i, Eppendorf) was used. Transplantation capillaries were pulled with a needle puller (P-97, Sutter Instruments) and loaded with approximately 5 µl of tumour cell suspension. Cells were injected into the perivitelline space (PVS) of larvae. Visual inspection was carried out at 2 hours post-injection on an Axio Zoom.V16 fluorescence microscope (Zeiss, Jena) and only correctly injected larvae were used in subsequent experiments and further maintained at 34°C.

Automated imaging and quantification

One day post injection (1dpi) and 3dpi xenografted larvae were anaesthetised in 1x Tricaine and embedded in a 96-well ZF plate (Hashimoto Electronic Industry) with 0.5 % ultra-low gelling agarose (Cat# A2576-25G, Sigma-Aldrich) for automated imaging on a high-content imager (Operetta CLS, PerkinElmer). Images were acquired with a 5x air objective. Exposure times for brightfield images was 40ms at 10% LED power. CellTrace Violet was recorded with an excitation of 390-420 nm at 100% LED power and detection at 430-500 nm using an exposure time of 600ms. GFP was excited with 460-

490nm and detected at 500-550nm with an exposure time of 400ms. 23 planes with a distance of 25 µm were imaged per field of view of the laterally orientated larvae to cover the whole tumour. Tumour size was quantified with Harmony Software 4.9 (PerkinElmer).

Ethical use of data

This study did not generate any new genomics data from patients. However, we performed re-analyses of previously published (sc)RNA-seq and SNP-array data that was previously collected at our institutions. The collection and research use of human tumour specimen was performed according to the guidelines of the Council for International Organizations of Medical Sciences (CIOMS) and World Health Organisation (WHO) and has been approved by the ethics board of the Medical University of Vienna (Ethikkommission Medizinische Universität Wien; EK2281/2016, 1216/2018, 1754/2022). Informed consent has been obtained from all patients or parents/guardians/legally authorized representatives. The age-adapted informed consent for the CCRI Biobank covers the use of left over materials from medically necessary surgery or biopsy, which after completion of routine diagnostic procedures is biobanked (EK1853/2016) and available for research purposes, including genetic analysis, that are further specified in EK1216/2018 and EK1754/2022: to conduct genetic and transcriptomic analysis and link to clinical data for survival analysis. All data obtained from external sources derived from studies where informed consent was given for broad research use.

Whole-exome sequencing

Library generation and sequencing

Genomic DNA (gDNA) from cell lines was isolated using a desalting method and library preparation was performed with 100ng gDNA and the Enzymatic Fragmentation (undifferentiated hESC lines; **Supplementary Figs. 4b,c**) or Enzymatic Fragmentation 2.0 (cells after differentiation; **Fig. 4e, Supplementary Figs. 9f,g**) kit, and Twist Universal Adapter System (Twist). For whole-exome sequencing, the libraries were pooled and enriched with the Exome v1.3 and RefSeq (Human Comprehensive Exome) spike-in capture probes (Twist) according to the manufacturer's protocols. Libraries were quantified (Qubit 4 Fluorometer) and quality-checked on 4200 TapeStation and 2100 Bioanalyzer automated electrophoresis instruments (Agilent) and diluted before sequencing by the Biomedical Sequencing Facility at CeMM on an Illumina NovaSeq SP flowcell in 2x100bp paired-end mode (median coverage 87.2; **Supplementary Table 1**).

Variant identification and annotation

Raw reads were processed using the nf-core *sarek*[110,111] WES pipeline version 2.7.2. Variant calling was performed in a tumour-normal matched mode, with the parental H7 line serving as the matched normal sample. Three variant callers, *Mutect2*, *Strelka*, and *Manta*[112–114], were employed for comprehensive variant identification. Resulting VCF files from *Mutect2* and *Strelka* were normalised using *bcftools* norm (v1.9)[115] and subsequently annotated using the Ensembl *Variant Effect Predictor* (VEP; v99.2)[116]. The identified variants were filtered based on the default quality control measures implemented in each tool (*FILTER* column in the VCF contains "PASS"). To identify biologically relevant variants a filtering strategy was applied that was partly inspired by MAPPYACTS[117]: (i) exclude variants for which GERMQ or STRQ Phred-scaled values are < 30; (ii) exclude variants with a population allele frequency of over 0.1% (in 1000 Genomes or gnomAD); (iii) only include variants that have any of "coding_sequence_variant", "frameshift_variant", "incomplete_terminal_codon", "inframe_deletion",

"inframe_insertion", "missense_variant", "protein_altering_variant", "start_lost", "stop_gained", "stop_lost" as Consequence; (iv) only include variants that have any of IMPACT == HIGH, SIFT == "deleterious", PolyPhen == probably_damaging or damaging[118,119]; (v) exclude variants that have a variant allele frequency <= 5%.

Copy number calling

CNAs were called by Sequenza (version 3.0.0)[120]. GC content was calculated for hg38 using *sequenza-utils gc_wiggle*. Depth ratio and B-allele frequency information was calculated using *bam2seqz* for each non-parental cell line using the parental cell line as a normal reference, single nucleotide variants (SNVs) were considered heterozygous if the allele frequency was in the range 0.4 to 0.6 (*--het* 0.4). Data was then binned using the *seqz_binning* command. Autosomes and the X chromosome were then extracted using Sequenza (*sequenza.extract*) and, as the cell lines are not contaminated with normal cells as is common place in tumour tissue samples, cellularity was tested in a range of 1 to 1.0002 to ensure a pure solution was produced by Sequenza. Copy number profiles were then plotted using ComplexHeatmaps[121]. Breakpoints were considered telomeric if they were within 1Mbp of the beginning or end of the chromosome.

Phylogenetic analysis

Mutations called by Mutect2[122] with the PASS filter and of VARIANT_CLASS SNV as annotated by VEP[116] that overlapped with the exome target panel without padding were used for phylogenetic analysis. Mutations were required to have a minimum variant allele frequency (VAF) of 0.2 to ensure only high frequency clonal mutations were included in the phylogeny. Phylogenetic trees were constructed using parsimony and the *phangorn* R package[123]. The parsimony ratchet method (*pratchet*) was used to search for the best tree and the tree was rooted on the parental cell line. Branch lengths were calculated using the *acctran* function. Distance between tree tips was calculated using the *distTips* function in the *adephylo* R package[124]. Phylogenetic trees were plotted using *ggtree*[125].

Pre-processing and analysis of NB SNP-array data

SNP-array data from tumour or bone marrow obtained at diagnosis from Austrian cases with INSS stage 4 high-risk NB[51] were re-analysed for chr17 and chr1 CNAs using VARAN-GIE (v0.2.9), yielding 88 samples with CNAs (>10kb) on at least one of these chromosomes. Genomic segments were manually curated and plotted using *ggplot2*[126] (v3.3.5). The available CNA annotations based on the human genome reference hg19. Because of this, the breakpoint annotations for our own cell lines have been brought from hg38 to hg19 using *liftOver* from the R package *rtracklayer* (v1.54.0).

Single-cell RNA sequencing (10x Genomics)

Library generation and sequencing

Single-cell suspensions were barcoded using oligo-conjugated lipids following the MULTI-seq workflow and frozen live[127] for G1-G13 (note, G2 was removed due to a technical failure), or frozen live and barcoded after thawing using the CELLPLEX (10x Genomics) workflow for G14-G27. After thawing cells were stained with DAPI. A maximum of 10,000 live cells per sample were sorted with a FACS-Aria v3 and pooled in sets of 3 or 4 samples by differentiation stage (from 3-5 independent replicate differentiation experiments). Each pooled group was processed using the 10X Genomics Single Cell 3' v3.1 workflow following the manufacturer's instructions. Enriched barcode libraries were indexed following the MULTI-seq workflow[127]. After quality control, libraries were sequenced on the

Illumina NovaSeq S4 (G1-13) or S2 (G14-27) platform in 2x150bp paired-end mode. **Supplementary Table 1** includes an overview of sequencing data and performance metrics.

Raw data processing and alignment

Raw sequencing data were processed with the *CellRanger multi* v7.1.0 software (10x Genomics) for cell-level demultiplexing and alignment to the human reference transcriptome (*refdata-gex-GRCh38-2020-A* assembly provided by 10x Genomics). Following initial data processing, all subsequent analyses were performed in R (v4.1.3) using Bioconductor packages and the *Seurat*[128–130] (v4.1.0) package.

Default basic processing

We applied processing of scRNA-seq data in many instances across this manuscript. Unless parameters are otherwise specified, the default processing of scRNA-seq counts involved the following steps. Counts were normalised for read depth using Seurat's *SCTransform*[131] v0.3.3 (parameters: *method="glmGamPoi"*; *variable.features.n=5000*), followed by *RunPCA* (keeping the top 50 components), and inference of cell neighbourhoods by *FindNeighbors* on the PCA reduction. Finally, Uniform Manifold Approximation and Projection (UMAP) was performed using Seurat's *RunUMAP* function with default parameters. Clustering was performed using FindClusters.

Quality control

For each dataset, we first assessed technical covariates and characteristic expression profiles separately. We kept cells with less than 15% mitochondrial UMI counts, and at least 1000 detected genes and applied basic scRNA-seq processing and clustering of the cells (*SCTransform*[131] v0.3.3, parameters: *method="glmGamPoi"*). Cell cycle scoring was calculated as recommended by Seurat and added as a variable to regress in SCTransform (*vars.to.regress="ccvar"*). We used clusters devoid of markers and/or characterised by abnormally high mitochondrial expression, to derive a library-specific UMI count and mitochondrial percentage threshold for high-quality cells (thresholds for counts / mitochondrial percentage: G1: 3162/10%; G3: 10000/7.5%; G4: 10000/8%; G5: 3162/10%; G6: 10000/8%; G7: 12589/8%; G8: 7943/10%; G9: 7079/10%; G10: 3981/7.5%; G11: 3981/10%; G12: 5012/10%; G13: 10000/10%; G14: 5500/13%; G15: 3500/5%; G16: 3000/8%; G17: 2000/8%; G18: 3500/10%; G19: 1800/6%; G20: 3000/15%; G21: 6000/8%; G22: 5000/6%; G23: 3000/6%; G24: 1500/6%; G25: 3500/5%; G26: 2000/10%; G27: 3000/15%). In addition, empty and doublet droplets were flagged with *Emptydrops*[132] (v1.14.2; default parameters) and *scDblFinder*[133] (v1.8.0; parameters: *dbr=0.01*), respectively. We retained only cells with *Emptydrops* FDR>0.05 and individual scDblscore cutoffs for the datasets were: G1: 0.01; G3: 0.016; G4: 0.005; G5: 0.005; G6: 0.003; G7: 0.005; G8: 0.005; G9: 0.005; G10: 0.005; G11: 0.005; G12: 0.005; G13: 0.005; G14: 0.005; G15: 0.005; G16: 0.0075; G17: 0.002; G18: 0.007; G19: 0.00375; G20: 0.01; G21: 0.007; G22: 0.007; G23: 0.0125; G24: 0.003; G25: 0.007; G26: 0.005; G27: 0.007.

Sample demultiplexing

To demultiplex cells belonging to different pooled samples, we used deMULTIplex2[134] (v1.0.1) with default parameters on each dataset using the tag counts from *CellRanger multi*. All non-singlet cells were filtered out from the dataset.

Normalisation, clustering, and marker gene analysis for the main dataset

Raw UMI counts were normalised using Seurat's *SCTransform*[131] (parameters: *variable.features.n=5000, method="glmGamPoi", vars.to.regress="ccvar"*) to account for differences in sequencing depth and cell cycle phase (the variable "*ccvar*" variable was calculated as

the difference of S and G2/M scores using Seurat's *CellCycleScoring* method with default parameters). To integrate data from 3-5 independent differentiation experiments (replicates; **Supplementary Table 1**), we used *scVI*[135] (v0.20.3; parameters: *n_epochs = 50*) using 5000 highly variable features of the input data with Python 3.11 via reticulate (v1.24). Nearest neighbours were identified using Seurat's *FindNeighbors* function (parameters: *k=30*) on the ten scVI components. The same scVI reduction was used to find a low dimensionality UMAP projection using Seurat's RunUMAP for both the WT-only (*n.neighbors=50, min.dist=.5, dims=1:8*) and full dataset (*n.neighbors: 30, min.dist=0.4, dims=1:8 method="umap-learn", metric="correlation").* Clusters on the UMAP projection were defined using Seurat's *FindClusters* (parameters [full dataset]: *resolution=0.6,* parameters [WT-only]: *resolution=0.4, algorithm=4*). Neighbouring clusters that shared functional markers were merged manually and relabelled to roughly reflect differentiation order. Finally, markers for each cluster were identified using the *FindAllMarkers2* function (*DElegate*[136] v1.1.0; parameters: *method="deseq", min_fc=1, min_rate=0.5, replicate_column="replicate"*), with each cluster compared to all the other cells in the dataset. Genes with an adjusted P-value less than 0.05 were selected as markers. (**Supplementary Tables 2, 8**). To compare mutant and wild-type cells, we filtered the integrated dataset to cells from D9 and identified pairwise DEGs ($P_{adj} \leq 0.05$, $|\log_2 FoldChange| > 0.25$) between each mutant condition and WT using the *findDE*function (Delegate v1.1.0; parameters: *group_column="condition", method="deseq", replicate_column = "day_rep"*). We discarded DEGs that were not expressed in at least 20% of cells on one side of the comparison. Up- and down-regulated DEGs on chr1q, on chr17q, and outside either CNA were then tested separately to identify significant overlaps with MSigDB HALLMARK[137] gene sets using the hypergeometric test implemented in the *hypeR*[138] package (v1.10.0). DEGs and enriched pathways are listed in **Supplementary Tables 6 and 7**.

Pseudotime trajectory analysis

Pseudotime trajectories were inferred using *Slingshot*[139] (v2.2.0; default parameters) using a filtered dataset comprising only MES-SYM clusters C11, C12, C13, and C14 (cp. **Fig. 1d; Supplementary Fig. 3**). The filtered dataset was reprocessed using the basic scRNA-seq processing workflow as described above and the first two principal components were used to find trajectories between two extreme clusters. Only one trajectory was found. Genes whose expression was associated with the trajectories were identified with the generalised additive model and association test as implemented in *tradeSeq*[140] (v1.8.0; parameters: *knots=5)*. The top genes with the highest Wald statistic were selected for reporting (**Supplementary Table 3**). Transcription factors were identified based on the human transcription factors database[141] in **Supplementary Fig. 3b.**

Cross-dataset annotation, label transfer, and signature scores

To map data between scRNA-seq datasets, we employed Seurat's label transfer workflow. Both query and reference datasets were processed using the default basic scRNA-seq processing workflow as described above and mapped (*FindTransferAnchors, TransferData, IntegrateEmbeddings, NNTransform,* and *MappingScore* functions; default parameters) using the 50 first principal components of the PCA reduction from both datasets. To visualise cell mappings, we used "glasswork plots", in which the UMAP of the reference was used to define the coordinates of convex hulls for each cluster. Query cells mapping to each cluster were plotted at random positions within their cognate reference cluster hull to mitigate overplotting bias when many cells mapped to a small neighbourhood. In this study, the following mappings were performed with the same processing and parameters:

1. Human foetal adrenal reference datasets[15,16] mapped onto WT-only (**Figs. 1d-f; Supplementary Figs. 2i,j, 3c**) and full *in vitro* (**Figs. 2d-f, 3d; Supplementary Fig. 6g**) scRNA-seq references. Upon obtaining consistent results for both (**Supplementary Fig. 2j**),

the reference provided by Kameneva *et al.* was used throughout the analysis, because of the curated cell type markers they provided (**Supplementary Fig. 2i**). These gene signatures were also quantified with Seurat's *AddModuleScore* function (default parameters) in **Figs. 1e,f, 2f**.

2. Our mutant scRNA-seq data mapped onto the wild-type reference (**Figs. 2b,c**).
3. NB tumour scRNA-seq data mapped onto our WT-only reference (**Figs. 6b,c; Supplementary Fig. 11**). See additional details about these datasets and processing in the section *"Pre-processing and mapping of NB tumour data"* below.
4. Extended data from a split-pool scRNA-seq (Parse Biosciences) mapped to the WT-only dataset (10x Genomics) (**Supplementary Fig. 8b-c**).

Validation of label transfers

WT mappings to adrenal gland references were validated by the presence of relevant markers (**Supplementary Fig. 2i**). Mutant and tumour cell mappings were not strictly curated via markers (i.e., they were allowed to deviate). When analysing markers of mapped mutant and tumour cells, cells with a prediction score of 0.4 or higher were used to minimise ambiguous mappings and maximise marker discovery. Shared markers were consistently found between the query and the cognate cells in the reference, even though their number varied (**Fig. 6c, Supplementary Fig. 8d**).

Visualising label transfers with glasswork plots

To visualise cell mappings, we used "glasswork plots", in which the UMAP of the reference was used to define the coordinates of concave hulls for each cluster (calculated with R package *concaveman* v1.1.0). Query cells mapping to each cluster were plotted at random positions within their mapped target cluster hull to mitigate overplotting. Input cell populations for the plot were downsampled evenly by condition and stage (n = 1000 cells) to avoid sampling effects.

Mutation score analysis

To calculate the mutation score, we focused on days 9, 14 and 19 as they contained samples from all conditions. We encoded each cell's genotype as a number *G* based on the genetic lineage of hESC lines: *G(WT)=0*, *G(17q)=1*, *G(17q1q)=2*, and *G(17q1qMYCN)=3*. We then calculated the mutation score *m* as the mean *G* of the cell's *K* nearest neighbouring cells (K = 30) in the scVI reduction's neighbourhood graph (see *"Normalization, clustering, and marker gene analysis"*). Division by 3 yielded a score between 0 and 1. Intuitively, the mutation score of a cell indicates whether a cell phenotypically resembles wild-type cells or cells with a given number of relevant alterations independent of its own genotype. To find genes correlated with the mutation score, we calculated Pearson correlation coefficients with gene expression in three settings: (i) correlation for each gene with *m* in all cells; (ii) correlation for each gene with *m* leaving out the 17q1qMYCN cells, to emphasise subtle correlations with CNAs; and (iii) correlation for each gene and the neighbourhood entropy (Shannon entropy of all genotype scores *G* of the *K* nearest neighbours), to find genes appearing in mixed regions. All non-duplicate absolute correlations (calculated using R's *cor.test*, parameters: *method="pearson", exact=TRUE*) were subject to Bonferroni correction and ranked. The top-300 correlated genes ($p \leq 0.05$) per differentiation stage (D9, D14, D19) are reported in **Supplementary Table 9**.

Pre-processing and mapping of NB tumour data

We collected scRNA-seq data for tumours with reported *MYCN* amplification from three sources from the stated database or the corresponding authors:

- Three samples (all primary adrenal, 2 male [Dong_T162, Dong_T230], 1 female [Dong_T200]; accession GSE137804 [Gene Expression Omnibus])[17],

- three samples (1 primary adrenal, 1 primary intraspinal, 1 relapse/occipital subcutaneous bone metastasis [Jansky_NB14]; 1 female [Jansky_NB08], 2 male [Jansky_NB01, Jansky_NB14]; accession EGAS00001004388 [European Genome-Phenome Archive])[15],
- and four samples (all metastatic bone marrow; 3 female [Fetahu_M1, Fetahu_M3, Fetahu_M4], 1 male [Fetahu_M2]; accession GSE216176 [Gene Expression Omnibus])[23].

Additional details about each dataset are available from the original research articles. In each dataset, cells with more than 500 reads per barcode and mitochondrial DNA less than 20% were kept for further analysis. We then performed an adrenal gland mapping[16] (same workflow as described above) and discarded cells mapping to the category "HSC_and_immune". This process left us with strong CNA profiles (see below) at key genomic positions such as chr2p (*MYCN* locus). Cells were then subjected to default basic scRNA-seq processing (see above) and mapped onto our WT-only reference (see above).

Inference of CNA profiles from scRNA-seq data

To infer tumour cell CNA profiles from scRNA-seq expression data, we used the *infercnv*[142] R package (v1.10.1). We first removed cells with less than 500 UMI counts. Then, we created a pan-patient healthy reference cell population by sampling from each patient 500 cells that we determined to be HSC/immune cells based on a mapping to a human embryonic adrenal gland reference[16]. For every patient, we then ran *infercnv* with the non-HSC/immune cells as the main input and the pan-patient HSC/immune cells as a reference. The *cutoff* parameter was set to 0.1, all other parameters were left at their default values.

Pre-processing and analysis of NB bulk RNA-seq data

We obtained bulk RNA-seq counts and associated metadata from patient-derived NB samples from three sources: TARGET[24] (phs000467 [Genomic Data Commons]), SEQC[72,73] (GSE49711 [Gene Expression Omnibus]) and from our institution[98,101,143–149] (labelled "CCRI" in the figures; GSE94035, GSE147635 and GSE172184 [Gene Expression Omnibus]). Open access unstranded counts from TARGET patients were obtained directly from the GDC data portal (subsection TARGET:NBL, phs000467). Counts from the CCRI patients were obtained in-house. Both CCRI and TARGET datasets were normalised using DESeq2[150] (v1.34.0) and transformed using the variance stabilizing transformation. A prenormalised $\log_2$ SEQC matrix was exponentiated, rounded to the nearest integer, and subjected to variance stabilizing transformation. In all datasets, the names of relevant marker genes were harmonised manually in case the gene was found with a different name. Each dataset was analysed separately due to differences in count quantification and normalisation. PCA projections of the normalised variables revealed mainly biological/clinical variables (and not technical variables) having major weight in the variance of the datasets. Only NB data collected at diagnosis were used for our analyses (discarding, e.g., ganglioneuroblastoma and relapse data). To quantify the *in-vitro* cluster signature strength, we used the intersection of markers found both in our *in-vitro* WT-only dataset (**Supplementary Table 2**) and in the tumour scRNA-seq datasets (**Supplementary Table 10**). We then used the function *gsva* (from *GSVA*[151] v1.42.0; parameters: *method="ssgsea"*) to calculate signature scores for each of the shared cluster signatures.

Survival analysis

We obtained survival data for the SEQC cohort from the original publication[72]. Event-free survival (EFS) was defined as time from diagnosis to any of the following events: Relapse/progression of disease, or death due to any cause and secondary malignancies. Patients without events were censored at last follow-up evaluation. EFS was estimated with Kaplan-Meier estimates and evaluated with Cox-

regression. Cluster signatures (see previous section) were dichotomized using the median-value. The results of the survival analysis are reported in **Supplementary Table 11**.

Split-pool single-cell RNA sequencing (Parse Biosciences)

Library generation and sequencing

Cells were harvested with Accutase to create a single-cell suspension, then were counted using Bio-rad Tc10 Automated cell counter in the presence of Trypan Blue Stain (Bio-rad). For cell fixation we used the Evercode Fixation v2 Kit (SKU: ECF 2001, Parse Biosciences, Seattle, USA) as per manufacturer instructions. A maximum of 5,000 cells per sample were multiplexed using the Evercode WT Mega v2 kit (Parse Biosciences). Three rounds of combinatorial barcoding were performed, and cells were then pooled and split into 16 sub-libraries (one small 5,000-cell sub-library and 15 large sub-libraries of 32,000 cells each). After DNA amplification and library prep, the small library was sequenced as part of a larger Illumina NovaSeq S4 flowcell and the 15 large sub-libraries on one dedicated NovaSeq S4 platform in 2x150 bp paired-end mode.

Raw data processing and alignment

Raw sequencing data were processed with the *split-pipe* v1.0.6p software (Parse Biosciences) for cell-level demultiplexing and alignment to the human reference transcriptome (*refdata-gex-GRCh38-2020-A* assembly provided by 10x Genomics; parameters: *-m all -c v2*). Following initial data processing, all subsequent analyses were performed in R (v4.1.3) using Bioconductor packages and the *Seurat*[128–130] (v4.1.0) package.

Basic processing, quality control, and marker analysis

We applied the *cb_filter_count_matrix* (default parameters) function from *canceRbits* (v0.1.6; default parameters) to remove cells with high mitochondrial counts (>15%), unusually high/low number of genes (< 300 genes or z-score of log(genes) not in range (-3, 3)), abnormally high/low number of reads (z-score of log(transcripts) not in range (-3, 3)), or an abnormal transcript-to-gene ratio (z-score of residuals of loess fit of "log(genes) ~ log(transcripts)" not in range (-5, 5)), and the *cb_seurat_pipeline* function (parameters: *seurat_max_pc = 15, metric = "manhattan", k_param = 20, n_neighbors = 40, cluster_res = 0.3*) to perform a standard Seurat analysis workflow including data normalisation, dimensionality reduction, and clustering. Subsequently, the data were mapped to our 10x-based WT-only reference as described above. To identify marker genes for cells mapped to different clusters of the WT reference (for cells with prediction score >=0.4) we again used DElegate::FindAllMarkers2. (DElegate v1.1.0; *parameters: method="deseq", min_fc=1, min_rate=0.5, replicate_column="replicate"*), and kept all genes with adjusted pvalue of 0.05 of less.

Chromatin accessibility mapping (ATAC-seq)

Library generation and sequencing

ATAC-seq was performed as described previously[79]. Briefly, 20,000 to 50,000 cells were lysed in the transposase reaction mix (12.5 µl 2xTD buffer, 2 µl TDE1 [Illumina], 10.25 µl nuclease-free water, and 0.25 µl 1% digitonin [Promega]) for 30 min at 37 °C. Following DNA purification with the MinElute kit (Qiagen) eluting in 12 µl, 1 µl of eluted DNA was used in a quantitative PCR (qPCR) reaction to estimate the optimum number of amplification cycles. The remaining 11 µl of each library were amplified for the number of cycles corresponding to the Cq value (i.e., the cycle number at which

fluorescence has increased above background levels) from the qPCR using custom Nextera primers. Library amplification was followed by SPRI (Beckman Coulter) size selection to exclude fragments larger than 1,200 bp. Libraries concentration was measured with a Qubit fluorometer (Life Technologies), and libraries were quality checked using a 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq 4000 platform in 1x50bp single-end mode. **Supplementary Table 1** includes an overview of the sequencing data and performance metrics.

Raw data processing, alignment, and quality control

Raw sequencing data were processed using *PEPATAC*[152] (v0.9.5; default parameters) including alignment to the human genome (*refdata-cell ranger-atac-GRCh38-1.2.0* assembly provided by 10x Genomics for maximum compatibility with scRNA-seq analyses). Following initial data processing, all subsequent analyses were performed in R (v4.1.3) using Bioconductor packages and *ggplot2*[126] (v3.3.5) and *ComplexHeatmap*[121] (v2.10.0) for plotting. After discarding low-quality data (NRF<0.65 or PBC1<0.7 or PBC2<1 or FRiP<0.025), we removed peaks overlapping blacklisted regions from ENCODE (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/hg38.blacklist.bed.gz) and merged overlapping peaks across all ATAC-seq datasets to create a common set of consensus genomic regions for subsequent analysis (**Supplementary Table 12**). Next, we quantified for each input dataset the number of reads overlapping these consensus peaks using *featureCounts*[153] (*Rsubread* v2.8.2).

Differential accessibility analysis and chromatin modules

Raw read counts were loaded into DESeq2[150] (v1.34.0; default parameters, design: *~lane+batch+sample_group*) for normalization (variance-stabilizing transformation) and differential analysis. In doing so, we estimated count size factors for normalization excluding regions on chromosomes with known chromosomal aberrations (i.e., chr1, chr17) to avoid overcompensation due to differences in global signal strength. We queried all pairwise comparisons of sample groups stratified by cell line / condition stratified (time-wise differences, e.g., WT-D3 vs. WT-D0) and between conditions stratified by stage (condition-wise differences, e.g., 17q-D9 vs. WT-D9) and recorded all significantly differentially accessible regions ($P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq \log_2(1.5)$; parameters: *pAdjustMethod="BH", lfcThreshold=log2(1.5), independentFiltering=TRUE;* **Supplementary Table 13**). To define chromatin regulatory modules, we divided time-wise differences in WT hESCs (n = 41,699 regions) into six chromatin modules (R1-R6) and condition-wise differences (n = 3,914 regions) into three chromatin modules (R7-R9) by hierarchical clustering using the Ward criterion (parameter: *method = "ward.D2"*). To associate ATAC-seq regions with putative target genes, we used the *GenomicRanges*[154] package (v1.46.1) to assign each region to all genes (using the *refdata-gex-GRCh38-2020-A* gene annotation provided by 10x Genomics) with overlapping promoters (transcription start side) or to distal genes whose promoter within a maximum distance of 250kb whose expression was significantly correlated with the region's accessibility. To this end, we calculated the correlation coefficient between normalised read counts in our ATAC-seq data with the normalised read counts in our matching scRNA-seq data (mean of cells per sample; note, the ATAC-seq was collected from the same experiments as the first replicate experiments for scRNA-seq). We calculated an empirical false discovery rate (FDR) by shuffling RNA/ATAC assignments (10 repetitions) and retained associations with a value ≤ 0.05. Annotated regulatory regions from the analysis of ATAC-seq data are listed in **Supplementary Table 12**.

Overlap enrichment analysis for chromatin modules

To characterise the chromatin modules, we interrogated overlaps with genomic regions or associated genes using the hypergeometric test implemented in the *hypeR*[138] package (v1.10.0) via the *cb_hyper* function (*canceRbits* v0.1.6; parameters: collapse = FALSE, min_size = 5, max_size = (<75% of the size of the background dataset)). We looked at three types of overlaps: (a) Annotated reference regions from the DNase hypersensitivity index[80], from the Cis-element Atlas[81], from the Enhancer Atlas[82], and NB subgroup-specific super-enhancers[75], which all catalogue regulatory elements active in different cell or tissue types. (b) Matches to known TF motifs from the *HOCOMOCO* database[155] (v11). Here, we downloaded motifs from the *HOCOMOCO* website (*HOCOMOCOv11_full_annotation_HUMAN_mono.tsv*) and used *motifmatchr* (v1.16.0) to scan the DNA sequences underlying each genomic region for matches. Regions with at least one match to the motif were recorded as potential binding sites. (c) Marker genes from our scRNA-seq analysis of WT hESC differentiation (**Fig. 1c**; **Supplementary Table 2**). For this purpose, genomic regions were associated with genes as described above. In each case, we used the entire set of all analysed genomic regions as a background for the enrichment analysis, and we considered overlaps with an FDR-corrected P-value less than 0.005 as significant. For motifs, we find the reported P-values are inflated and therefore used stricter thresholds: $P_{adj} \leq 0.0000001$, $|\log_2 \text{odds}| > \log_2(2)$. All enrichment results are reported in **Supplementary Table 14**.

Integration with published ATAC/DNaseI-seq data

To interrogate accessibility of the chromatin modules in existing data from NB cell lines we used fast gene set enrichment analysis *fgsea* (v1.20.0)[156]. We obtained ready-processed genomic coverage tracks (*wig* or *bigwig* files) from three studies profiling NB cell lines[75,157,158] (GSE138293, GSE224241, GSE136279). Additionally, we obtained data from three studies profiling breast[159] (GSE202511) and lung cancer[160] lines (GSE228832), or human tissue data[161] (https://epigenome.wustl.edu/epimap) as controls. For studies based on older genome assemblies (GSE138293, GSE224241, GSE136279, GSE228832, and EpiMap used hg19), we converted our peak coordinates to hg19 using the *liftOver* R package (v1.18.0). We then used the *GenomicRanges*[154] (v1.46.1) and *plyranges*[162] (v1.14.0) packages to identify genome segments overlapping our peaks and to aggregate the corresponding mean score reported in the coverage tracks, which were then used for gene set enrichment analysis via the *cb_fgsea* function (*canceRbits* v0.1.6; parameters: max_size = Inf).

Identification of transcription factor targets

To identify putative target genes of TFs, we used *GRNboost2*[83] (*arboreto* library v0.1.6, with Python v3.8.17 via reticulate [v1.24]) to identify genes whose expression could be predicted from the expression of each TF. We tested all TFs in the *HOCOMOCO* database[155] for which at least one motif could be identified in our dataset. We found that stronger association values were reported for stem-cell-related factors, likely because of a proportional overrepresentation of this developmental stage in our dataset. To alleviate this effect and create more balanced data to build our networks on, we downsampled our dataset to no more than 500 cells per cluster and took the average importance value of ten random samples forward for further analysis. Putative targets with high importance values but without a supporting nearby ATAC-seq peak with a motif matching the respective TF were considered indirect targets and discarded from the target gene sets. We found that the range of importance values varied between TFs. We therefore calculated a TF-specific threshold on the importance score to define target genes. To this end, we ranked importance values and used the *changepoint* package (v2.2.3; default parameters) to identify the first point at which the mean values of the curve of importance values changed (disregarding the top 1% highest importance values which often were outliers and disrupted

this analysis). The resulting target gene sets were divided into putative activating and inhibiting interactions by the sign of the Pearson correlation coefficient *r* of the respective TF-target pairs (using the mean correlation value of the same eight random samples as used for *GRNboost2*). Interactions with $|r|<0.1$ were discarded. To calculate the average expression of target genes in each cell we used only activated targets ($r>0.1$) and the Seurat module score. To identify significant overlaps between target genes and gene sets D9_1 – D9_4 (**Supplementary Table 15**), we used the *hypeR*[138] package (v1.10.0) via the *cb_hyper* function (*canceRbits* v0.1.6; parameters: collapse = FALSE, min_size = 0, max_size = Inf), considering TFs with $P_{adj} \leq 0.05$, $|\log_2 \text{odds}| \geq \log_2(4)$, and frequency ≥ 5% as significant. All target gene sets are reported in **Supplementary Table 15** and all enrichment results in **Supplementary Table 16**.

Gene-regulatory network visualisation

For the visualisation of gene-regulatory networks, we used the *igraph* package (v1.3.1). A directed graph was constructed from edges between genes in the gene sets D9_1, D9_2, D9_3, or D9_4 (**Supplementary Table 9**) and TFs found enriched in the overlap with these genes (**Fig. 8d**). The same automated graph layout (function *layout_with_fr()*) was used to draw mutant-specific network diagrams. To generate mutant-specific networks (**Fig. 8f**), we selected cells derived at D9 and parameterised node colour to indicate the mean scaled expression of the genes in those cells and node size to indicate the mean scaled TF target score (Seurat module score) for TFs or the mean scaled expression for non-TFs. To simplify plots, we only labelled TFs with positive mean scaled expression values (>0.05) and manually aggregated many overlapping values, but all node labels are shown in **Supplementary Fig. 12c**.

Data availability

Raw and processed single-cell RNA-seq and ATAC-seq will be deposited at the Gene Expression Omnibus (GEO). Public scRNA-seq data from NB tumours used in this study are available under the following accession codes: GSE147821, GSE216176, and GSE137804 (Gene Expression Omnibus), and EGAS00001004388 (European Genome-Phenome Archive). Public ATAC-seq data from NB cell lines and controls used in this study are available under accession codes: GSE138293, GSE224241, GSE136279, GSE202511, and GSE228832 (Gene Expression Omnibus), from the EpiMap website (https://epigenome.wustl.edu/epimap). Bulk RNA-seq data from NB tumours were obtained from phs000467 (Genomic Data Commons), GSE49711, GSE94035, GSE147635 and GSE172184 (Gene Expression Omnibus).

Code availability

Computer code used for the data analysis in this paper will be shared via our GitHub page (https://github.com/cancerbits).

Acknowledgements

We would like to thank the Biomedical Sequencing Facility at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences for assistance with next-generation sequencing, Bettina Brunner-Herglotz (CCRI) for her technical assistance, and Duncan Baker (Sheffield Diagnostic Genetic Services, Sheffield Children's Hospital) for carrying out karyotyping of hESC lines. We would like to acknowledge Yann Jamin and Barbara Martins da Costa (ICR, London) for help with the MRI and animal work, respectively. We are also grateful to Igor Adameyko and Polina

Kameneva (Medical University of Vienna), Sofie Mohlin (Lund University), and Christoph Bock (CeMM Research Center for Molecular Medicine) for critical reading of the manuscript. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Funding

Funding sources in authorship order: K.B.: Biotechnology and Biological Sciences Research Council (BBSRC) (DTP Studentship, BB/T007222/1); I.S.F.: Austrian Science Fund (FWF) (10.55776/P35072); M.C.B.: Austrian Academy of Sciences (25905); E.M.P.: FWF (10.55776/P32001, 10.55776/P34832); M.H.: National Institutes of Health (R00CA197484); I.B.: UK Regenerative Medicine Platform (MR/R015724/1), Medical Research Council UK (MRC) (MR/X000028/1, MR/X007979/1); M.D.: Austrian Research Promotion Agency (FFG) project 7940628 533 (Danio4Can), Alex's Lemonade Stand Foundation for Childhood Cancer (ALSF) (20-17258); L.C./E.Po.: CRUK Programme Award A28278; L. C.: HEFCE/ICR; S.T.M.: Vienna Science and Technology Fund (WWTF; project LS18-111), FWF (10.55776/P35841B MAPMET), H2020 (project 826494 PRIMAGE); M.F.: ALSF (20-17258); A.T.: BBSRC (New Investigator Research Grant, BB/P000444/1), European Union Horizon 2020 Framework Programme (H2020-EU.1.2.2; project 824070), Children's Cancer and Leukaemia Group/Neuroblastoma UK/Little Princess Trust (CCLGA 2019 28, CCLGA 2020 19, together with H.B.), MRC (MR/V002163/1); F.H.: ALSF (20-17258), FWF (10.55776/TAI454, 10.55776/TAI732, 10.55776/PAT1300223).

Author contributions

I.S.G. and L.M.G. championed the experimental and computational work on this study, respectively. Formal contributions in authorship order (CrediT taxonomy): Conceptualization: A.T., F.H.; Data curation: L.M.G., I.S.G., L.S., K.B., E.Po., S.W.Z., E.B., I.S.F., P.Z., U.P., G.C., S.T.M., M.F., A.T., F.H.; Formal Analysis: L.M.G., I.S.G., C.H., E.Po., S.W.Z., P.Z., U.P., C.St., M.S., G.C., H.B., A.T., F.H.; Funding acquisition: K.B., I.S.F., M.C.B., W.W., P.A., I.B., H.B., M.D., L.C., S.T.M., M.F., A.T., F.H.; Investigation: I.S.G., L.M.G., K.B., E.Po., L.S., D.S., R.Le., E.B., I.S.F., M.B., A.W.W., C.St., C.So., S.T., P.B., M.R., M.G., M.C.B.; Methodology: I.S.G., L.M.G., L.S., E.Po., I.S.F., M.B., P.B., M.G., G.C., H.B., M.D., L.C., S.T.M., M.F., A.T., F.H.; Project administration: A.T., F.H.; Resources: W.W., E.Pu., M.H., R.La., H.B., M.D., L.C., S.T.M, M.F., A.T., F.H.; Software: L.M.G., C.H.; Supervision: E.Pu., G.C., H.B., M.D., L.C., S.T.M., M.F., A.T., F.H.; Visualization: I.S.G., L.M.G., K.B., E.P., C.St., G.C., A.T., F.H.; Writing – original draft: I.S.G., L.M.G., A.T., F.H.; Writing – review & editing: I.S.G., L.M.G., K.B., C.H., E.Po., L.S., I.S.F., C.St., C.So, M.B., P.B., M.G., M.C.B., E.Pu., M.H., P.A., I.B., H.B., G.C., M.D., S.T.M., M.F., A.T., F.H.

Declaration of interests

The authors declare no competing interests.

References

1. Marshall, G. M. *et al.* The prenatal origins of cancer. *Nat Rev Cancer* **14**, 277–289 (2014).
2. Pritchard-Jones, K. Genetics of childhood cancer. *Br Med Bull* **52**, 704–23 (1996).
3. Scotting, P. J., Walker, D. A. & Perilongo, G. Childhood solid tumours: A developmental disorder. *Nature Reviews Cancer* vol. 5 481–488 Preprint at https://doi.org/10.1038/nrc1633 (2005).
4. Maris, J. M. & Denny, C. T. Focus on embryonal malignancies. *Cancer Cell* **2**, 447–50 (2002).
5. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).
6. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).
7. Matthay, K. K. *et al.* Neuroblastoma. *Nat Rev Dis Primers* **2**, 16078 (2016).
8. Schulte, J. H. & Eggert, A. Neuroblastoma. *Crit Rev Oncog* **20**, 245–270 (2015).
9. Zeineldin, M., Patel, A. G. & Dyer, M. A. Neuroblastoma: When differentiation goes awry. *Neuron* **110**, 2916–2928 (2022).
10. Nakagawara, A. Neural crest development and neuroblastoma: the genetic and biological link. *Prog Brain Res* **146**, 231–242 (2004).
11. Ponzoni, M. *et al.* Recent advances in the developmental origin of neuroblastoma: an overview. *Journal of Experimental and Clinical Cancer Research* **41**, 1–28 (2022).
12. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *EMBO Journal* **16**, 2985–2995 (1997).
13. Cohen, M. A. *et al.* Formation of Human Neuroblastoma in Mouse-Human Neural Crest Chimeras. *Cell Stem Cell* **26**, 579-592.e6 (2020).
14. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**, 5075–5082 (2017).
15. Jansky, S. *et al.* Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. *Nat Genet* **53**, 683–693 (2021).
16. Kameneva, P. *et al.* Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin. *Nat Genet* **53**, 694–706 (2021).
17. Dong, R. *et al.* Single-Cell Characterization of Malignant Phenotypes and Developmental Trajectories of Adrenal Neuroblastoma. *Cancer Cell* **38**, 716-733.e6 (2020).
18. Hanemaaijer, E. S. *et al.* Single-cell atlas of developing murine adrenal gland reveals relation of Schwann cell precursor signature to neuroblastoma phenotype. *Proc Natl Acad Sci U S A* **118**, (2021).
19. Kildisiute, G. *et al.* Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell. *Sci Adv* **7**, eabd3311 (2021).

20. de Preter, K. *et al.* Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. *Genome Biol* **7**, 1–17 (2006).

21. Althoff, K. *et al.* A Cre-conditional MYCN-driven neuroblastoma mouse model as an improved tool for preclinical studies. *Oncogene* **34**, 3357–3368 (2015).

22. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589–593 (2012).

23. Fetahu, I. S. *et al.* Single-cell transcriptomics and epigenomics unravel the role of monocytes in neuroblastoma bone marrow metastasis. *Nat Commun* **14**, 3620 (2023).

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

25. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N Engl J Med* **340**, 1954–1961 (1999).

26. Gilbert, F. *et al.* Human neuroblastomas and abnormalities of chromosomes 1 and 17. *Cancer Res* **44**, 5444–9 (1984).

27. Lastowska, M. *et al.* Comprehensive genetic and histopathologic study reveals three types of neuroblastoma tumors. *J Clin Oncol* **19**, 3080–3090 (2001).

28. Mazzocco, K. *et al.* Genetic abnormalities in adolescents and young adults with neuroblastoma: A report from the Italian Neuroblastoma Group. *Pediatr Blood Cancer* **62**, 1725–1732 (2015).

29. Körber, V. *et al.* Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome. *Nat Genet* **55**, 619–630 (2023).

30. Bogen, D. *et al.* The genetic tumor background is an important determinant for heterogeneous MYCN-amplified neuroblastoma. *Int J Cancer* **139**, 153–163 (2016).

31. O'Neill, S. *et al.* MYCN amplification and 17q in neuroblastoma: evidence for structural association. *Genes Chromosomes Cancer* **30**, 87–90 (2001).

32. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb Perspect Med* **3**, (2013).

33. Schleiermacher, G. *et al.* Accumulation of segmental alterations determines progression in neuroblastoma. *J Clin Oncol* **28**, 3122–3130 (2010).

34. Schleiermacher, G. *et al.* Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *Br J Cancer* **107**, 1418 (2012).

35. Frith, T. J. R. *et al.* Human axial progenitors generate trunk neural crest cells in vitro. *Elife* **7**, (2018).

36. Frith, T. J. R. & Tsakiridis, A. Efficient Generation of Trunk Neural Crest and Sympathetic Neurons from Human Pluripotent Stem Cells Via a Neuromesodermal Axial Progenitor Intermediate. *Curr Protoc Stem Cell Biol* **49**, e81 (2019).

37. Wymeersch, F. J., Wilson, V. & Tsakiridis, A. Understanding axial progenitor biology in vivo and in vitro. *Development* **148**, (2021).

38. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–7 (1998).

39. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (1979)* **364**, (2019).

40. Kastriti, M. E. *et al.* Schwann cell precursors represent a neural crest-like state with biased multipotency. *EMBO J* **41**, e108780 (2022).

41. Pattyn, A., Morin, X., Cremer, H., Goridis, C. & Brunet, J. F. Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. *Development* **124**, 4065–4075 (1997).

42. Xie, M. *et al.* Schwann cell precursors contribute to skeletal formation during embryonic development in mice and zebrafish. *Proc Natl Acad Sci U S A* **116**, 15068–15073 (2019).

43. McGonnell, I. M. & Graham, A. Trunk neural crest has skeletogenic potential. *Curr Biol* **12**, 767–771 (2002).

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

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

46. van Groningen, T. *et al.* A NOTCH feed-forward loop drives reprogramming from adrenergic to mesenchymal state in neuroblastoma. *Nat Commun* **10**, 1530 (2019).

47. Ciccarone, V., Spengler, B. A., Meyers, M. B., Biedler, J. L. & Ross, R. A. Phenotypic diversification in human neuroblastoma cells: expression of distinct neural crest lineages. *Cancer Res* **49**, 219–25 (1989).

48. Biedler, J. L., Helson, L. & Spengler, B. A. Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. *Cancer Res* **33**, 2643–52 (1973).

49. Halliwell, J., Barbaric, I. & Andrews, P. W. Acquired genetic changes in human pluripotent stem cells: origins and consequences. *Nature Reviews Molecular Cell Biology 2020 21:12* **21**, 715–728 (2020).

50. Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* **22**, 53–54 (2004).

51. Abbasi, M. R. *et al.* Impact of Disseminated Neuroblastoma Cells on the Identification of the Relapse-Seeding Clone. *Clin Cancer Res* **23**, 4224–4232 (2017).

52. Kerosuo, L. *et al.* Enhanced expression of MycN/CIP2A drives neural crest toward a neural stem cell-like fate: Implications for priming of neuroblastoma. *Proc Natl Acad Sci U S A* **115**, E7351–E7360 (2018).

53. Girish, V. *et al.* Oncogene-like addiction to aneuploidy in human cancers. *Science* **381**, eadg4521 (2023).

54. Mitchell, P. J., Timmons, P. M., Hébert, J. M., Rigby, P. W. J. & Tjian, R. Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev* **5**, 105–119 (1991).

55. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res* **30**, 35–48 (2020).

56. Howard, A. G. A. & Uribe, R. A. Hox proteins as regulators of extracellular matrix interactions during neural crest migration. *Differentiation* **128**, 26–32 (2022).

57. Akkermans, O. *et al.* GPC3-Unc5 receptor complex structure and role in cell migration. *Cell* **185**, 3931-3949.e26 (2022).

58. TeSlaa, J. J., Keller, A. N., Nyholm, M. K. & Grinblat, Y. Zebrafish Zic2a and Zic2b regulate neural crest and craniofacial development. *Dev Biol* **380**, 73–86 (2013).

59. Handel, A. E. *et al.* Developmental dynamics of the neural crest–mesenchymal axis in creating the thymic microenvironment. *Sci Adv* **8**, 9844 (2022).

60. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).

61. Adam, K., Lesperance, J., Hunter, T. & Zage, P. E. The Potential Functional Roles of NME1 Histidine Kinase Activity in Neuroblastoma Pathogenesis. *Int J Mol Sci* **21**, (2020).

62. Ryl, T. *et al.* Cell-Cycle Position of Single MYC-Driven Cancer Cells Dictates Their Susceptibility to a Chemotherapeutic Drug. *Cell Syst* **5**, 237-250.e8 (2017).

63. Lutz, W. *et al.* Conditional expression of N-myc in human neuroblastoma cells increases expression of alpha-prothymosin and ornithine decarboxylase and accelerates progression into S-phase early after mitogenic stimulation of quiescent cells. *Oncogene* **13**, 803–12 (1996).

64. King, D. *et al.* MYCN expression induces replication stress and sensitivity to PARP inhibition in neuroblastoma. *Oncotarget* **11**, 2141 (2020).

65. Kramer, M., Ribeiro, D., Arsenian-Henriksson, M., Deller, T. & Rohrer, H. Proliferation and Survival of Embryonic Sympathetic Neuroblasts by MYCN and Activated ALK Signaling. *J Neurosci* **36**, 10425–10439 (2016).

66. Mobley, B. C. *et al.* Expression of MYCN in Multipotent Sympathoadrenal Progenitors Induces Proliferation and Neural Differentiation, but Is Not Sufficient for Tumorigenesis. *PLoS One* **10**, e0133897 (2015).

67. Berry, T. *et al.* The ALK(F1174L) mutation potentiates the oncogenic activity of MYCN in neuroblastoma. *Cancer Cell* **22**, 117–130 (2012).

68. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**, 5075–5082 (2017).

69. Cohen, M. A. *et al.* Formation of Human Neuroblastoma in Mouse-Human Neural Crest Chimeras. *Cell Stem Cell* **26**, 579-592.e6 (2020).

70. Huang, M. *et al.* Engineering Genetic Predisposition in Human Neuroepithelial Stem Cells Recapitulates Medulloblastoma Tumorigenesis. *Cell Stem Cell* **25**, 433-446.e7 (2019).

71. Wrobel, J. K. *et al.* Rapid In Vivo Validation of HDAC Inhibitor-Based Treatments in Neuroblastoma Zebrafish Xenografts. *Pharmaceuticals (Basel)* **13**, 345 (2020).

72. Su, Z. *et al.* A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* **32**, 903–914 (2014).

73. Wang, C. *et al.* The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol* **32**, 926–932 (2014).

74. Durbin, A. D. *et al.* EP300 Selectively Controls the Enhancer Landscape of MYCN-Amplified Neuroblastoma. *Cancer Discov* **12**, 730–751 (2022).

75. Gartlgruber, M. *et al.* Super enhancers define regulatory subtypes and cell identity in neuroblastoma. *Nat Cancer* **2**, 114–128 (2021).

76. Zeid, R. *et al.* Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat Genet* **50**, 515–523 (2018).

77. Jahangiri, L. *et al.* Core regulatory circuitries in defining cancer cell identity across the malignant spectrum. *Open Biol* **10**, 200121 (2020).

78. Zimmerman, M. W. *et al.* Retinoic acid rewires the adrenergic core regulatory circuitry of childhood neuroblastoma. *Sci Adv* **7**, (2021).

79. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).

80. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature 2020 584:7820* **584**, 244–251 (2020).

81. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).

82. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res* **48**, D58–D64 (2020).

83. Moerman, T. *et al.* GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).

84. Moreno-Smith, M. *et al.* Restoration of the molecular clock is tumor suppressive in neuroblastoma. *Nature Communications 2021 12:1* **12**, 1–16 (2021).

85. Altman, B. J. *et al.* MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell Metab* **22**, 1009–1019 (2015).

86. Xue, C. *et al.* MYCN promotes neuroblastoma malignancy by establishing a regulatory circuit with transcription factor AP4. *Oncotarget* **7**, 54937–54951 (2016).

87. Boboila, S. *et al.* Transcription factor activating protein 4 is synthetically lethal and a master regulator of MYCN-amplified neuroblastoma. *Oncogene* **37**, 5451–5465 (2018).

88. Ana Paula Azambuja, A., Paula Azambuja, A. & Simoes-Costa, M. The connectome of neural crest enhancers reveals regulatory features of signaling systems. *Dev Cell* **56**, 1268-1282.e6 (2021).

89. Gundem, G. *et al.* Clonal evolution during metastatic spread in high-risk neuroblastoma. *Nat Genet* **55**, 1022–1033 (2023).

90. Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Lef1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev* **15**, 1688–1705 (2001).

91. Gautier, M., Thirant, C., Delattre, O. & Janoueix-Lerosey, I. Plasticity in neuroblastoma cell identity defines a noradrenergic-to-mesenchymal transition (Nmt). *Cancers (Basel)* **13**, (2021).

92. Thirant, C. *et al.* Reversible transitions between noradrenergic and mesenchymal tumor identities define cell plasticity in neuroblastoma. *Nat Commun* **14**, (2023).

93. Vayani, O. R. *et al.* Adrenergic and mesenchymal signatures are identifiable in cell-free DNA and correlate with metastatic disease burden in children with neuroblastoma. *Pediatr Blood Cancer* **71**, (2024).

94. Durbin, A. D. *et al.* Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nature Genetics* vol. 50 1240–1246 Preprint at https://doi.org/10.1038/s41588-018-0191-z (2018).

95. Rouhani, F. J. *et al.* Substantial somatic genomic variation and selection for BCOR mutations in human induced pluripotent stem cells. *Nature Genetics 2022 54:9* **54**, 1406–1416 (2022).

96. Puigdevall, P., Jerber, J., Danecek, P., Castellano, S. & Kilpinen, H. Somatic mutations alter the differentiation outcomes of iPSC-derived neurons. *Cell Genomics* **3**, 100280 (2023).

97. Aschero, R. *et al.* Recurrent Somatic Chromosomal Abnormalities in Relapsed Extraocular Retinoblastoma. *Cancers 2021, Vol. 13, Page 673* **13**, 673 (2021).

98. Lazic, D. *et al.* Landscape of Bone Marrow Metastasis in Human Neuroblastoma Unraveled by Transcriptomics and Deep Multiplex Imaging. *Cancers (Basel)* **13**, 4311 (2021).

99. Verhoeven, B. M. *et al.* The immune cell atlas of human neuroblastoma. *Cell Rep Med* **3**, 100657 (2022).

100. Costa, A. *et al.* Single-cell transcriptomics reveals shared immunosuppressive landscapes of mouse and human neuroblastoma. *J Immunother Cancer* **10**, (2022).

101. Weiss, T. *et al.* Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal growth factor-like protein 8. *Nat Commun* **12**, (2021).

102. Ludwig, T. E. *et al.* ISSCR standards for the use of human stem cells in basic research. *Stem Cell Reports* **18**, 1744 (2023).

103. Price, C. J. *et al.* Genetically variant human pluripotent stem cells selectively eliminate wild-type counterparts through YAP-mediated cell competition. *Dev Cell* **56**, 2455-2470.e10 (2021).

104. Randolph, L. N., Bao, X., Zhou, C. & Lian, X. An all-in-one, Tet-On 3G inducible PiggyBac system for human pluripotent stem cells and derivatives. *Sci Rep* **7**, (2017).

105. Loew, R., Heinz, N., Hampf, M., Bujard, H. & Gossen, M. Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol* **10**, 1–13 (2010).

106. Zhou, X., Vink, M., Klaver, B., Berkhout, B. & Das, A. T. Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther* **13**, 1382–1390 (2006).

107. Laing, O., Halliwell, J. & Barbaric, I. Rapid PCR Assay for Detecting Common Genetic Variants Arising in Human Pluripotent Stem Cell Cultures. *Curr Protoc Stem Cell Biol* **49**, e83 (2019).

108. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**, 424–429 (2011).

109. Sturtzel, C. *et al.* Refined high-content imaging-based phenotypic drug screening in zebrafish xenografts. *npj Precision Oncology 2023 7:1* **7**, 1–16 (2023).

110. Hanssen, F. *et al.* Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. *bioRxiv* **23**, 2023.07.19.549462 (2023).

111. Garcia, M. *et al.* Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Research 2020 9:63* **9**, 63 (2020).

112. Van der Auwera, G., O'Connor, B. & Safari, an O. M. Company. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. *Genomics in the Cloud* 300 (2020).

113. Kim, S. *et al.* Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods* **15**, 591–594 (2018).

114. Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).

115. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).

116. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 1–14 (2016).

117. Berlanga, P. *et al.* The European MAPPYACTS Trial: Precision Medicine Program in Pediatric and Adolescent Patients with Recurrent Malignancies. *Cancer Discov* **12**, 1266–1281 (2022).

118. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, (2013).

119. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–3814 (2003).

120. Favero, F. *et al.* Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann Oncol* **26**, 64–70 (2015).

121. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

122. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

123. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

124. Jombart, T., Balloux, F. & Dray, S. adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* **26**, 1907–1909 (2010).

125. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**, 28–36 (2017).

126. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. *Use R! series* (Springer, 2016).

127. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**, 619–626 (2019).

128. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21 (2019).

129. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

130. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).

131. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 576827 (2019).

132. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, 1–9 (2019).

133. Germain, P. L., Robinson, M. D., Lun, A., Garcia Meixide, C. & Macnair, W. Doublet identification in single-cell sequencing data using *scDblFinder*. *F1000Research 2022 10:979* **10**, 979 (2022).

134. Zhu, Q., Conrad, D. N. & Gartner, Z. J. deMULTIplex2: robust sample demultiplexing for scRNA-seq. *bioRxiv* 2023.04.11.536275 (2023) doi:10.1101/2023.04.11.536275.

135. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods 2018 15:12* **15**, 1053–1058 (2018).

136. Hafemeister, C. & Halbritter, F. Single-cell RNA-seq differential expression tests within a sample should use pseudo-bulk data of pseudo-replicates. *bioRxiv* 2023.03.28.534443 (2023) doi:10.1101/2023.03.28.534443.

137. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417 (2015).

138. Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307–1308 (2020).

139. Street, K. *et al.* Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

140. van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1–13 (2020).

141. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).

142. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (1979)* **344**, 1396–1401 (2014).

143. Rifatbegovic, F. *et al.* Neuroblastoma cells undergo transcriptomic alterations upon dissemination into the bone marrow and subsequent tumor progression. *Int J Cancer* **142**, 297–307 (2018).

144. Ladenstein, R. *et al.* [Value of prognostic factors in the Austrian A-NB87 Neuroblastoma Study]. *Klin Padiatr* **208**, 210–20 (1996).

145. Fiedler, S. *et al.* Long-Term Outcome and Role of Biology within Risk-Adapted Treatment Strategies: The Austrian Neuroblastoma Trial A-NB94. *Cancers 2021, Vol. 13, Page 572* **13**, 572 (2021).

146. Garaventa, A. *et al.* Randomized Trial of Two Induction Therapy Regimens for High-Risk Neuroblastoma: HR-NBL1.5 International Society of Pediatric Oncology European Neuroblastoma Group Study. *J Clin Oncol* **39**, 2552–2563 (2021).

147. Ladenstein, R. *et al.* Investigation of the Role of Dinutuximab Beta-Based Immunotherapy in the SIOPEN High-Risk Neuroblastoma 1 Trial (HR-NBL1). *Cancers (Basel)* **12**, (2020).

148. Ladenstein, R. *et al.* Interleukin 2 with anti-GD2 antibody ch14.18/CHO (dinutuximab beta) in patients with high-risk neuroblastoma (HR-NBL1/SIOPEN): a multicentre, randomised, phase 3 trial. *Lancet Oncol* **19**, 1617–1629 (2018).

149. Ladenstein, R. *et al.* Busulfan and melphalan versus carboplatin, etoposide, and melphalan as high-dose chemotherapy for high-risk neuroblastoma (HR-NBL1/SIOPEN): an international, randomised, multi-arm, open-label, phase 3 trial. *Lancet Oncol* **18**, 500–514 (2017).

150. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

151. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 1–15 (2013).

152. Smith, J. P. *et al.* PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genom Bioinform* **3**, lqab101 (2021).

153. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

154. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, e1003118 (2013).

155. Kulakovskiy, I. v. *et al.* HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252–D259 (2018).

156. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021) doi:10.1101/060012.

157. Upton, K. *et al.* Epigenomic profiling of neuroblastoma cell lines. *Sci Data* **7**, (2020).

158. Decaesteker, B. *et al.* SOX11 regulates SWI/SNF complex components as member of the adrenergic neuroblastoma core regulatory circuitry. *Nature Communications 2023 14:1* **14**, 1–16 (2023).

159. Xu, D. *et al.* Recapitulation of patient-specific 3D chromatin conformation using machine learning. *Cell reports methods* **3**, (2023).

160. Hariprakash, J. M. *et al.* Leveraging Tissue-Specific Enhancer-Target Gene Regulatory Networks Identifies Enhancer Somatic Mutations That Functionally Impact Lung Cancer. *Cancer Res* **84**, (2024).

161. Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature 2021 590:7845* **590**, 300–307 (2021).

162. Lee, S., Cook, D. & Lawrence, M. Plyranges: A grammar of genomic data transformation. *Genome Biol* **20**, (2019).

163. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

Figures

Figure 1

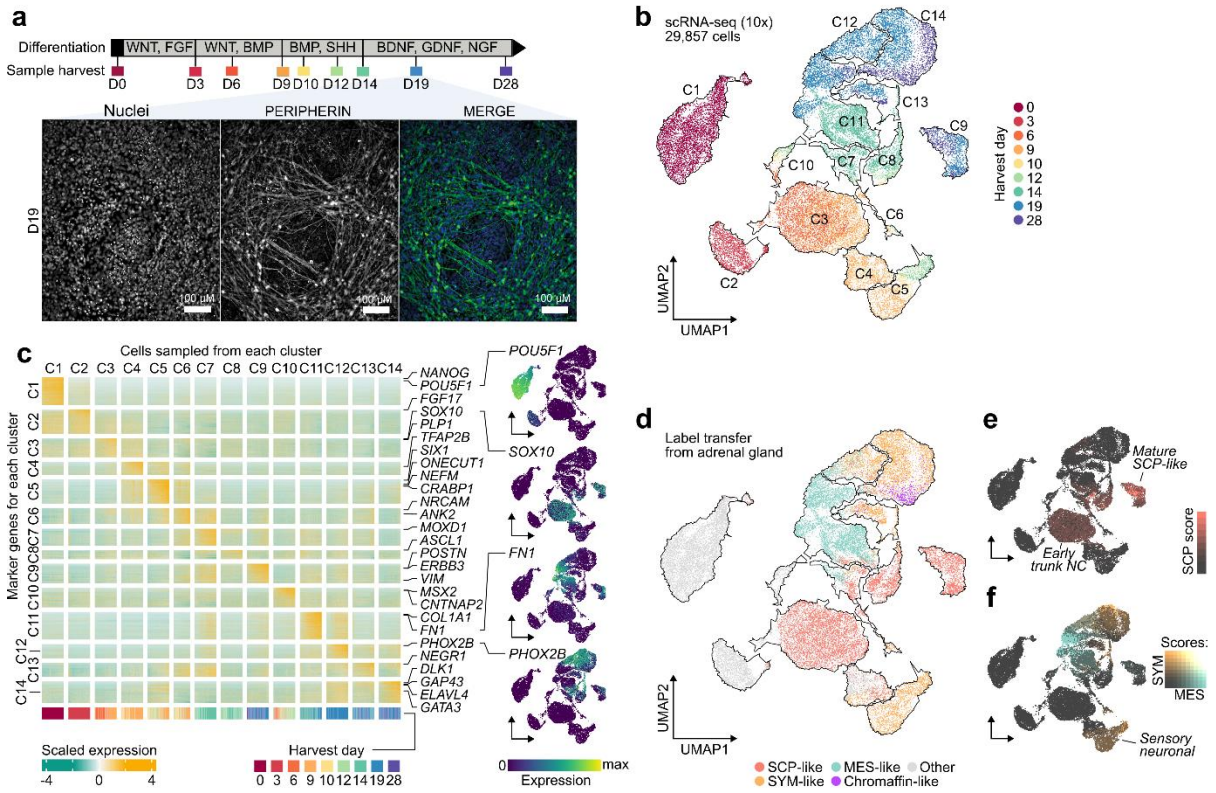


Figure 1. *In vitro* culture efficiently generates human trunk NC cells and their sympathoadrenal derivatives from hESCs.

a) Diagram depicting the extrinsically supplemented signals employed to direct hESCs toward trunk NC cells and their downstream derivatives, and immunofluorescence analysis of PERIPHERIN protein expression illustrating the generation of sympathetic neurons at D19. Cell nuclei were counterstained using Hoechst 33342.

b) UMAP of scRNA-seq data from wild-type hESCs collected at 9 stages (indicated by different colours) of differentiation to trunk neural crest and sympathoadrenal derivatives. Cells were divided into 14 distinct clusters as indicated by the contours.

c) Heatmap of gene markers for each cluster in **panel b**. Selected genes have been highlighted and UMAPs indicate the expression level of canonical markers for stem (*POU5F1*), neural crest (*SOX10*), mesenchymal (*FN1*), and sympathetic (*PHOX2B*) cells. All marker genes are reported in **Supplementary Table 2**.

d) Cells from D9-D28 of **panel b** labelled by their closest matching cell type from the human embryonic adrenal gland reference[16] via label transfer. Cells in grey could not be verified with markers (**Supplementary Fig. 2i**) or could not be assigned to a single type.

e) Cells from **panel d** coloured by the strength of their SCP marker signature (Seurat module score) in red. A high / medium score distinguishes a cluster of early SCP-like / trunk NC and a late cluster with more mature SCP-like cells.

f) Same as above but visualising simultaneously SYM (orange) and MES (teal) marker signature. Cells with overlapping marker signatures appear in grey tones, with the highest mixture in C12. An early diverging cluster of sensory neuron-like cells has a weak match to the SYM signature. A pseudotime trajectory for the MES-SYM transition in clusters C11-C14 can be found in **Supplementary Figure 3**.

Abbreviations: hESC, human embryonic stem cells; D0/3/9/10/12/14/19/28, day 0/3/9/10/12/14/19/28; UMAP, Uniform Manifold Approximation and Projection; C1-C14, cell clusters; SCP, Schwann cell precursor; SYM, sympathoblast; MES, mesenchymal.

Figure 2

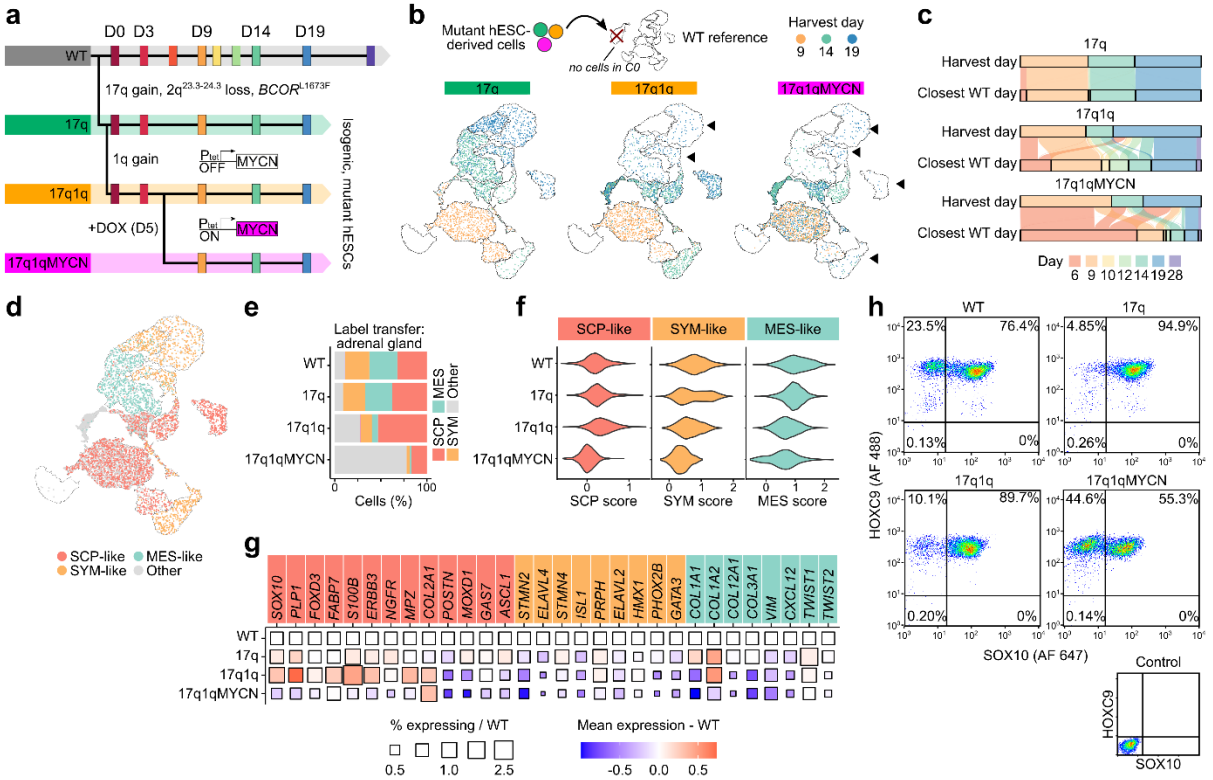


Figure 2. Copy number alterations and overexpression of *MYCN* impair the specification of trunk NC derivatives.

a) Scheme depicting the different hESC genetic backgrounds employed and the timing of DOX-induced MYCN overexpression in the context of our trunk NC differentiation system.

b) scRNA-seq data from mutant cells (17q, 17q1q, 17q1qMYCN at D9, D14, and D19) were mapped to the wild-type trunk NC reference (illustration on the left side). Glasswork UMAP plots depicting the destination clusters in the WT reference for cells of the 17q, 17q1q and 17q1qMYCN conditions. Mutant cells are coloured by stage to emphasise mismatches with WT. Fewer 17q1q and 17q1qMYCN cells map to later differentiation stages, highlighted by arrows.

c) Alluvial plots comparing the stage at which each mutant cell was harvested versus its phenotypically closest stage in the WT reference (based on label transfer as also shown in **panel b**). In each subplot, the top bar indicates the proportion of cells collected at each stage (D9, D14, D19). The bottom bar indicates the distribution of matching phenotypes in the WT for that same population of cells, and streams indicate which subpopulations flow into cognate or non-cognate WT stages. The plots suggest that cells from 17q1q and 17q1qMYCN progressively mapped to earlier stages compared to WT.

d) Glasswork UMAPs of mapped 17q, 17q1q, and 17q1qMYCN cells (as in **panel b**) coloured by closest-matching cell type in the human embryonic adrenal gland reference[16]. The category "other" comprises other cell types in the reference dataset and mappings that could not be validated by cell type markers (**Supplementary Fig. 2i**).

e) Percentage of cells mapped to each cell type in **panel d** split by cell line.

f) Violin plots indicating the strength of the SCP/SYM/MES (left to right) gene expression signature (Seurat module score) for cells mapped to the respective cell type, split by cell line.

g) Plot indicating the change in mean expression (colour) and the percentage of cells expressing the gene (size) for each gene in the signatures from **panel e** relative to WT. WT squares (size = 1, white) are shown for reference.

h) Flow cytometric analysis of trunk NC markers HOXC9 and SOX10 in D9 cultures obtained from hESCs marked by the indicated NB-associated lesions.

Abbreviations: WT, wild-type H7 hESCs; D0/3/9/10/12/14/19/28, day 0/3/9/10/12/14/19/28; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell precursor; SYM, sympathoblast; MES, mesenchymal.

Figure 3

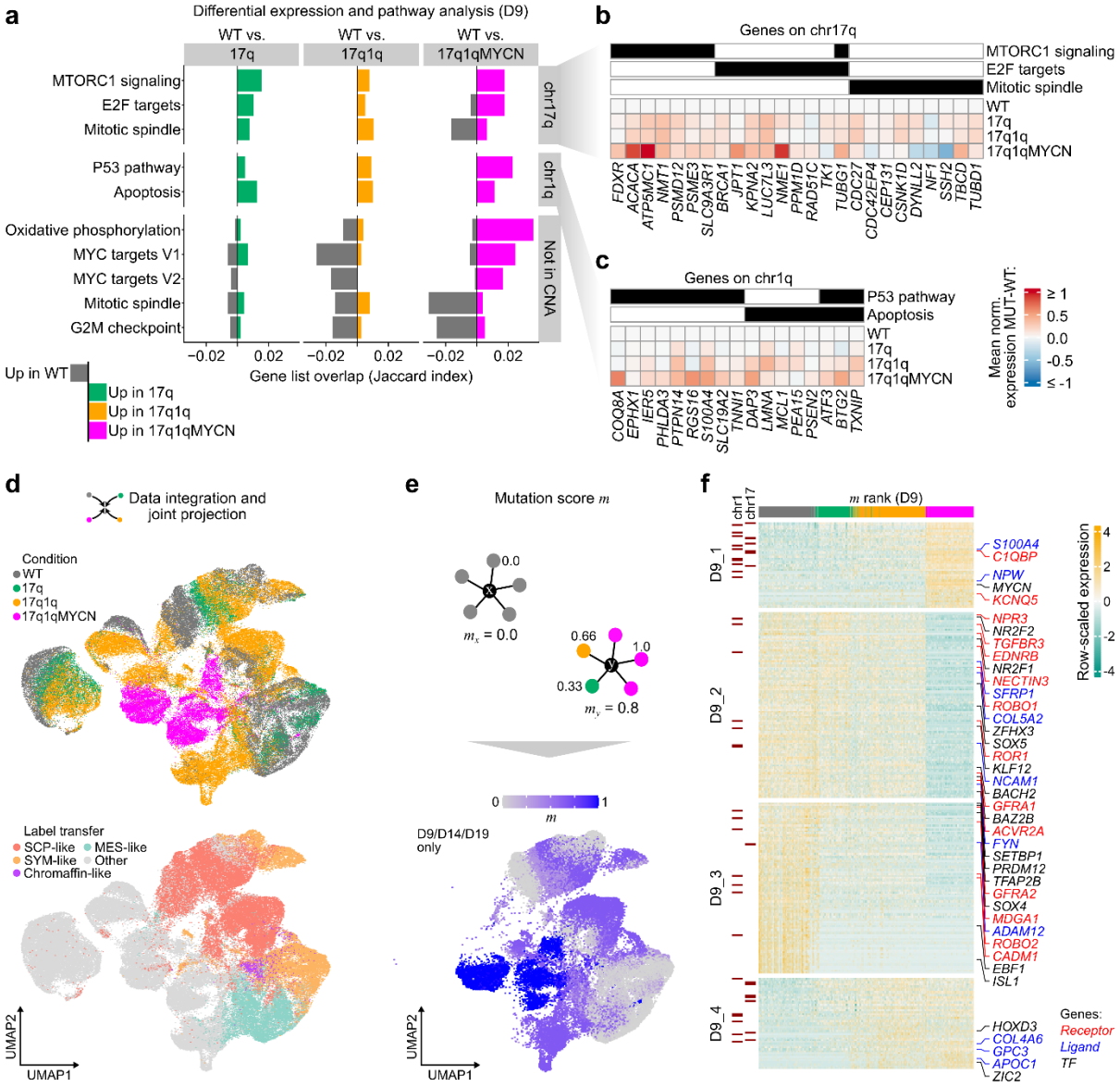


Figure 3. Copy number alterations and overexpression of *MYCN* alter the expression of metabolic and developmental pathways.

a) We performed differential expression analysis between WT and derivatives of each mutant hESC line at D9 of differentiation and performed a pathway enrichment analysis to summarise the differentially expressed genes (DEGs). Enrichment was determined by hypergeometric tests (*hypeR*[138], background: all detected genes in our scRNA-seq dataset; $P_{adj} \leq 0.1$) using pathways from MSigDB[137]. The overlap between up- and down-regulated DEGs with the pathway genes is indicated as a positive (green/orange/magenta colour bars) or negative (grey colour) number, respectively. We additionally distinguished between DEGs located on chromosome arms chr17q, chr1q, or anywhere else in the genome to analyse potential direct and indirect effects of CNAs (split from top to bottom). All differentially expressed genes and pathway enrichments are available in **Supplementary Tables 6 and 7**.

b) DEGs located on chromosome arm chr17q from the enriched pathways shown in **panel a**. The heatmap indicates the mean normalised expression difference between each indicated mutant cell line and WT (at D9). The annotation bar on top of the heatmap indicate membership (black colour) of genes in the selection pathways (MSigDB hallmark database).

c) As **panel (b)**, but for DEGs on chr1q and the respective enriched pathways.

d) UMAP of scRNA-seq data from wild-type and mutant hESCs (indicated by colour; cp. **Fig. 2a**) throughout differentiation to trunk neural crest and sympathoadrenal derivatives (top). Bottom: the same dataset coloured by closest-matching cell type in the human embryonic adrenal gland reference[16]. The category "other" comprises other cell types in the reference dataset and mappings that could not be validated by cell type markers (**Supplementary Fig. 2i**).

e) Illustration (top) of the calculation of mutation scores *m* (k-nearest neighbour (KNN) mutational average) as average score of each cell's neighbours. In this calculation, each neighbour weighs in by its cell line (0 = WT, 1/3 = 17q, 2/3 = 17q1q, 1 = 17q1qMYCN) such that the mutation score allows ordering cells from WT to MYCN mutation. Only cells from D9, D14, and D19 were used, for which data from all conditions were available. The actual scores are shown overlaid on the UMAP from **panel d** (bottom).

f) Heatmap showing the expression of top 300 genes highly correlated to the mutation score *m* from **panel e** across all cells from D9. Genes have been divided into four groups by hierarchical clustering, and selected TFs, receptors, and ligands are highlighted. All correlated genes are reported in **Supplementary Table 9**. Genes located on chr17q or chr1q are indicated.

Abbreviations: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; m, mutation score; TF, transcription factor.

Figure 4

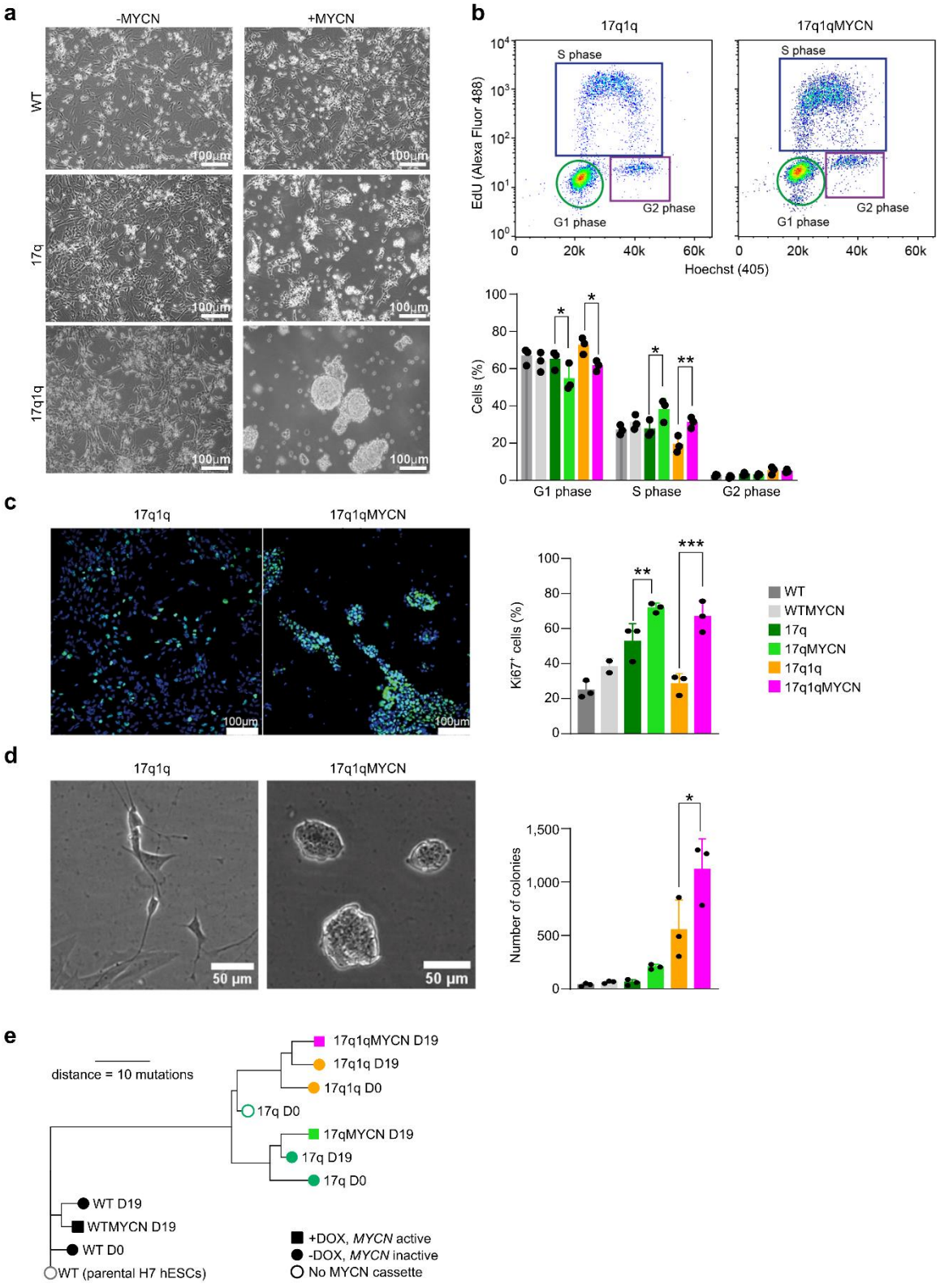


Figure 4. Impaired trunk NC specification correlates with acquisition of tumourigenic hallmarks.

a) Representative brightfield images of D14 cultures following differentiation of hESCs with the indicated genotypes.

b) Flow cytometric analysis of cell cycle in D9 cultures obtained from hESCs marked by the indicated NB-associated lesions. Top: Representative FACS plots. Bottom: Percentage of cells

found in each of the different stages of the cell cycle (G1, S, G2) corresponding to indicated NB-associated lesions (n = 3 biological replicates, error bars= standard deviation two-way ANOVA). P values in comparisons: G1 (17q vs 17qMYCN, p = 0.0266= *; 17q1q vs 17q1qMYCN, p= 0.0153=*), S (17q vs 17qMYCN, p = 0.0233= *; 17q1q vs 17q1qMYCN, p= 0.0073=**). Note only comparisons examining the effect of MYCN overexpression in different backgrounds are shown.

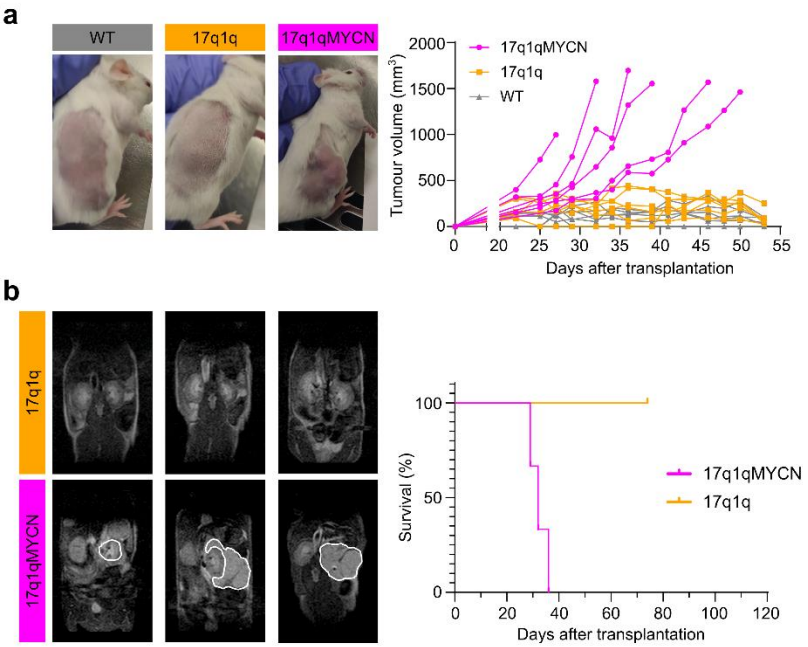
c) Immunofluorescence analysis (green) of the expression of the cell proliferation marker KI-67 in D14 (left) cultures obtained from hESCs marked by the indicated NB-associated lesions. Cell nuclei were counterstained using Hoechst 33342 (blue). Scoring of the percentages of KI-67-positive cells is also shown (right) (n = 3 biological replicates, error bars= standard deviation, ordinary two-way ANOVA with Tukey correction). P values in comparisons: 17q vs 17qMYCN, p=0.0078 =**; 17q1q vs. 17q1qMYCN p= 0.0001=***). Note only comparisons examining the effect of MYCN overexpression in different backgrounds are shown.

d) Left: Representative brightfield images of cell/colony morphology following a low-density plating assay using cells marked by the indicated NB-associated lesions after 84 hours. Right: Comparison of the number of colonies formed by cells marked by the indicated NB-associated lesions following plating at low density. (n= 3 biological replicates, error bars= SD, Ordinary Two-way ANOVA test with Tukey correction). P values in comparisons: 17q1q vs. 17q1qMYCN p= 0.0109=*. Note only comparisons examining the effect of MYCN overexpression in different backgrounds are shown.

e) Phylogenetic tree indicating the genetic relationship and distance (in number of SNVs detected by whole-exome sequencing) between different hESC lines before (D0) and after differentiation (D19). The shape of nodes indicates samples without a *MYCN* overexpression cassette (unfilled circles), with an expression cassette which has not been activated (filled circles), and with an activated (by addition of DOX from D5 onwards) *MYCN* overexpression cassette (filled squares). The colours have been chosen to match those used elsewhere in the paper, without specific meaning. The plot shows that few additional mutations occurred during differentiation since the distances between differentiated cell lines and the shared ancestor with the matching undifferentiated samples are small (<10 mutations). **Supplementary Tables 4 and 5** report SNVs and CNAs identified in our analyses.

Abbreviations: D0/5/14/19, day 0/5/14/19; WT, wild-type H7 hESCs.

Figure 5



c

Site of injection	Cell line	Penetrance
Subcutaneous	WT	0 out of 6
	17q1q	0 out of 6
	17q1qMYCN	6 out of 6
Adrenal gland	17q1q	0 out of 3
	17q1qMYCN	3 out of 3

Figure 5. hESC-derived trunk NC cells with CNAs form tumours in mice upon *MYCN* overexpression.

a) Left: Representative images of subcutaneous xenografts of trunk NC cells derived from the indicated cell lines in the presence (17q1qMYCN) and absence (WT, 17q1q) of DOX treatment. Right: Graph showing tumour size growth per mouse corresponding to xenografts of indicated cell lines (n = 6 animals per cell line).

b) Left: Representative MRI sections of mice at week 5 following xenografting of indicated cell lines in the adrenal gland and DOX treatment regimens. The white lines indicate the tumour perimeter. Right: Graph showing survival of the recipient animals after xenografting (n = 3 animals per cell line).

c) Summary of mouse xenograft experiments.

Abbreviations: DOX, doxycycline; MRI, magnetic resonance imaging.

Figure 6

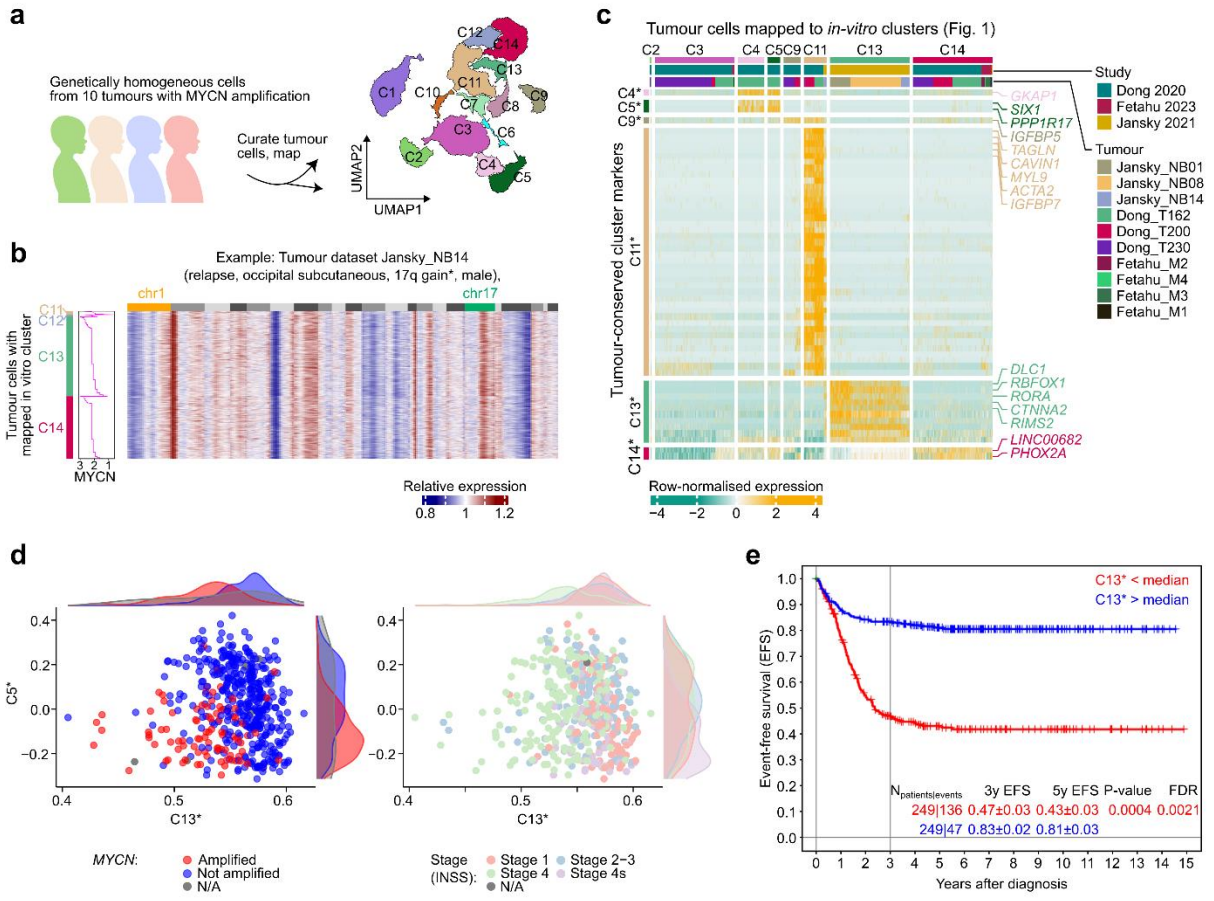


Figure 6. Comparison to hESC-based trunk NC differentiation resolves structured heterogeneity across neuroblastoma tumours.

a) Schematic overview of the analysis of tumour cells. We curated tumour cells from 10 *MYCN*-amplified NB samples[15,17,23] from three studies and mapped them onto our reference (cp. **Fig. 1**)[129]. Mapping is represented as tumour cells falling into sectors of the WT *in vitro* reference (depicted as contours of each cluster in the glasswork plot).

b) Heatmap depicting gene expression in *MYCN*+ tumour cells of dataset *Jansky_NB14*[15]. Values are inferCNV[142] copy number estimations per gene, relative to hematopoietic and immune cells in the sample ordered by genomic position and chromosome (1-22). Cells (one per row) are shown ordered by the respective cluster in our reference dataset that the tumour cells were mapped to (C11 to C14) and therein by *MYCN* expression levels (depth-normalised sliding window average, width = 20 cells). Mappings of other tumours datasets are shown in **Supplementary Fig. 11**.

c) Heatmap showing markers from gene expression signatures C4*, C5*, C9*, C13*, and C14* (rows, top to bottom) in cells from 10 tumour datasets that were mapped to our in vitro reference dataset (cp. **panel a**). Each gene expression signature is the intersection of the cluster markers in our reference dataset (as in Fig. 1) and differentially expressed genes between the respective tumour cells. No genes were found for C2* and C3*. Markers for tumour cells mapped to C2 and C3 showed no overlap with *in-vitro* cluster markers; thus, only mapped cells are shown. All genes identified in this analysis are reported in **Supplementary Table 10**.

d) Scatterplots evaluating the strength of gene expression signatures C5* and C13* (from **panel c**; calculated using *GSVA*[151]) in NB bulk RNA-seq data from SEQC[72,73]. Each dot corresponds to one tumour dataset coloured by *MYCN* amplification status (left) or clinical stage (right). The density of points (kernel density estimate) in each group is indicated in the margins of the plots.

e) Survival analysis for data from the SEQC cohort stratified by strength of the C13* expression signature. Groups were split by the median. Cox regression adjusted for age-group (< 18 months, 18-60 months, > 60 months), INSS stage 4 (yes / no), and *MYCN* amplification status (yes / no). n = 249 patients per group, or 136 [C13* low] and 47 [C13* high] events. All survival analysis results are reported in **Supplementary Table 11**.

Abbreviations: UMAP, Uniform Manifold Approximation and Projection; EFS, event-free survival.

Figure 7

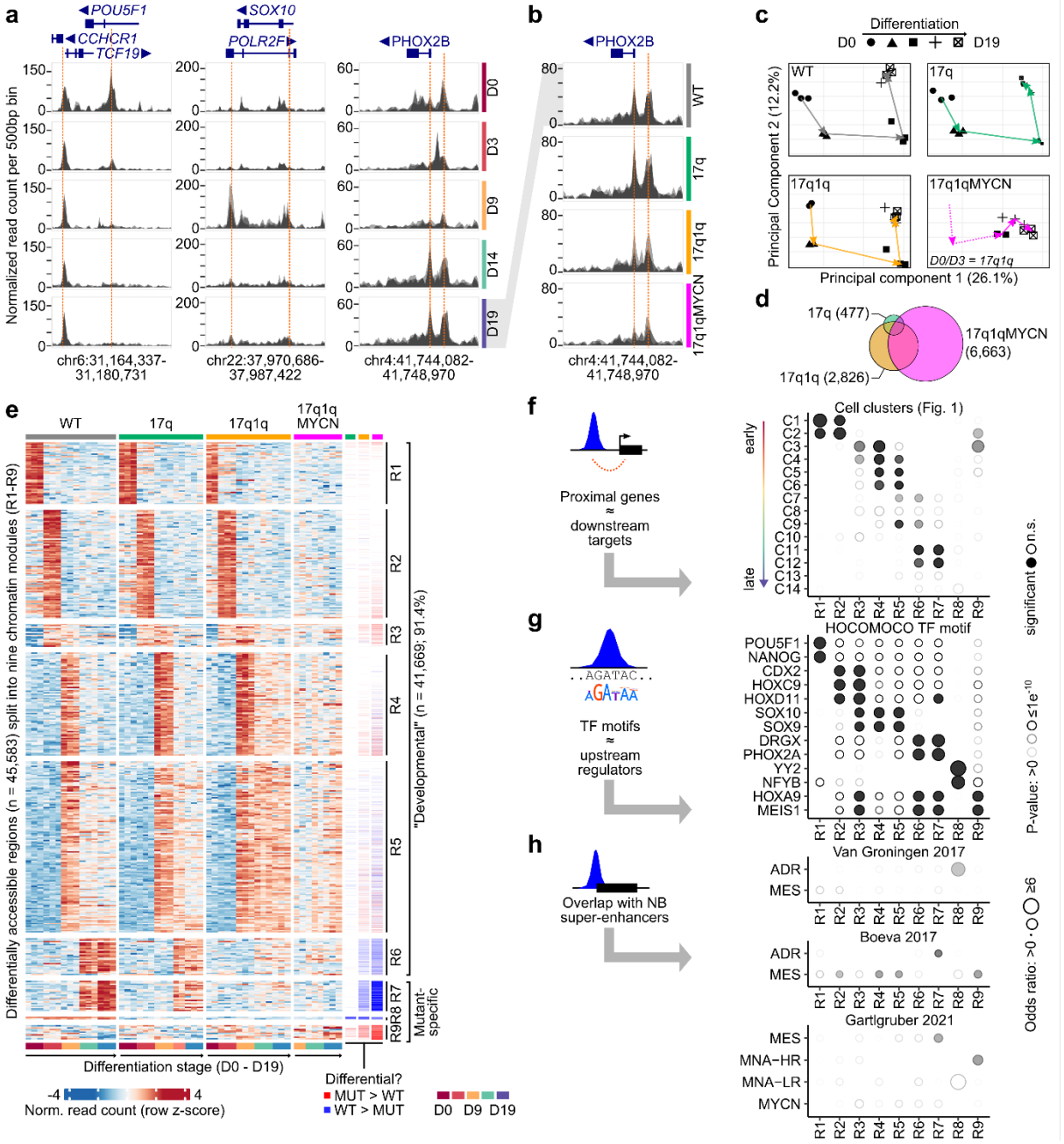


Figure 7. Differentiation of wild-type and mutant hESCs is associated with epigenetic changes in nine distinct chromatin modules.

a) ATAC-seq read coverage for wild-type hESCs at three example loci. Each area plot reports the normalised read count aggregated per genomic bin (width = 500bp). Multiple semi-transparent area plots are overlaid for each replicate. Genes within each locus are shown on top with thin/thick lines indicating introns/exons. The arrows next to gene names indicate the direction of transcription. Selected peaks have been highlighted manually (dashed red lines).

b) ATAC-seq read coverage of wild-type and mutant hESCs at D19 near the *PHOX2B* locus. Plots as in **panel a**.

c) Principal component analysis of all ATAC-seq datasets, split into four panels by condition. The geometric means of all data belonging to the same stages are connected by arrows to visualise the stepwise chromatin changes during differentiation.

d) Euler diagram visualising the overlap of differentially accessible regions in mutant hESC-derived trunk NC derivatives across all stages compared to WT (*DEseq2*[150]; $P_{adj} \leq 0.005$,

|log$_2$FoldChange| ≥ log$_2$(1.5)). Numbers indicate the total number of regions per cell line aggregated over all developmental stages.

e) Heatmaps showing normalised read counts for all differentially accessible regions (columns) in any pairwise comparison of two stages or conditions (*DEseq2*[150]; $P_{adj} \leq 0.005$, |log$_2$FoldChange| ≥ log$_2$(1.5); n_{total} = 45,583). Regions have been divided into nine non-overlapping modules (R1–R9) by hierarchical clustering. Three annotation columns are shown to the right indicating regions called down- (blue) and up-regulated (red) in each mutant hESC. All regions and differential analysis results are reported in **Supplementary Tables 12 and 13**.

f) Comparison of regions belonging to the nine chromatin modules (from **panel e**) and proximal genes from the cluster markers lists defined in our scRNA-seq analysis (cp. **Fig. 1**). An enrichment analysis for the overlaps was performed using hypergeometric tests (using *hypeR*[138]; background: all genes associated with at lest one ATAC-seq peak) and the sizes and transparency of circles indicates the odds ratio and P-value, respectively. Significant results are indicated with filled circles ($P_{adj} \leq 0.005$). All results are shown in the figure and also reported in **Supplementary Table 14**.

g) Enrichment analysis for overlaps between chromatin modules and known TF motifs (*HOCOMOCO* database[155], v11). The plots are as in **panel f**, with the exception that only overlaps with $P_{adj} \leq 0.0000001$ and |log$_2$FoldChange| ≥ log$_2$(2) were marked as significant (background: all peaks with at least one motif match). The top results per module are shown and all results are reported in **Supplementary Table 14**.

h) Enrichment analysis of overlaps between regions belonging to the nine chromatin modules and super-enhancers specific to certain NB epigenetic subtypes[44,45,75] (background: all peaks with at least one overlapping region annotated in the super-enhancer analyses). Plots as in **panel f**.

Abbreviations: D0/3/9/14/19, day 0/3/9/14/19; WT, wild-type H7 hESCs; MUT, a "mutant" hESC line (17q, 17q1q, or 17q1qMYCN); R1-R9, chromatin region modules; sig., significant; ADR, adrenergic; MES, mesenchymal; MNA-HR, not *MYCN*-amplified high-risk; MNA-LR, not *MYCN*-amplified low-risk.

Figure 8

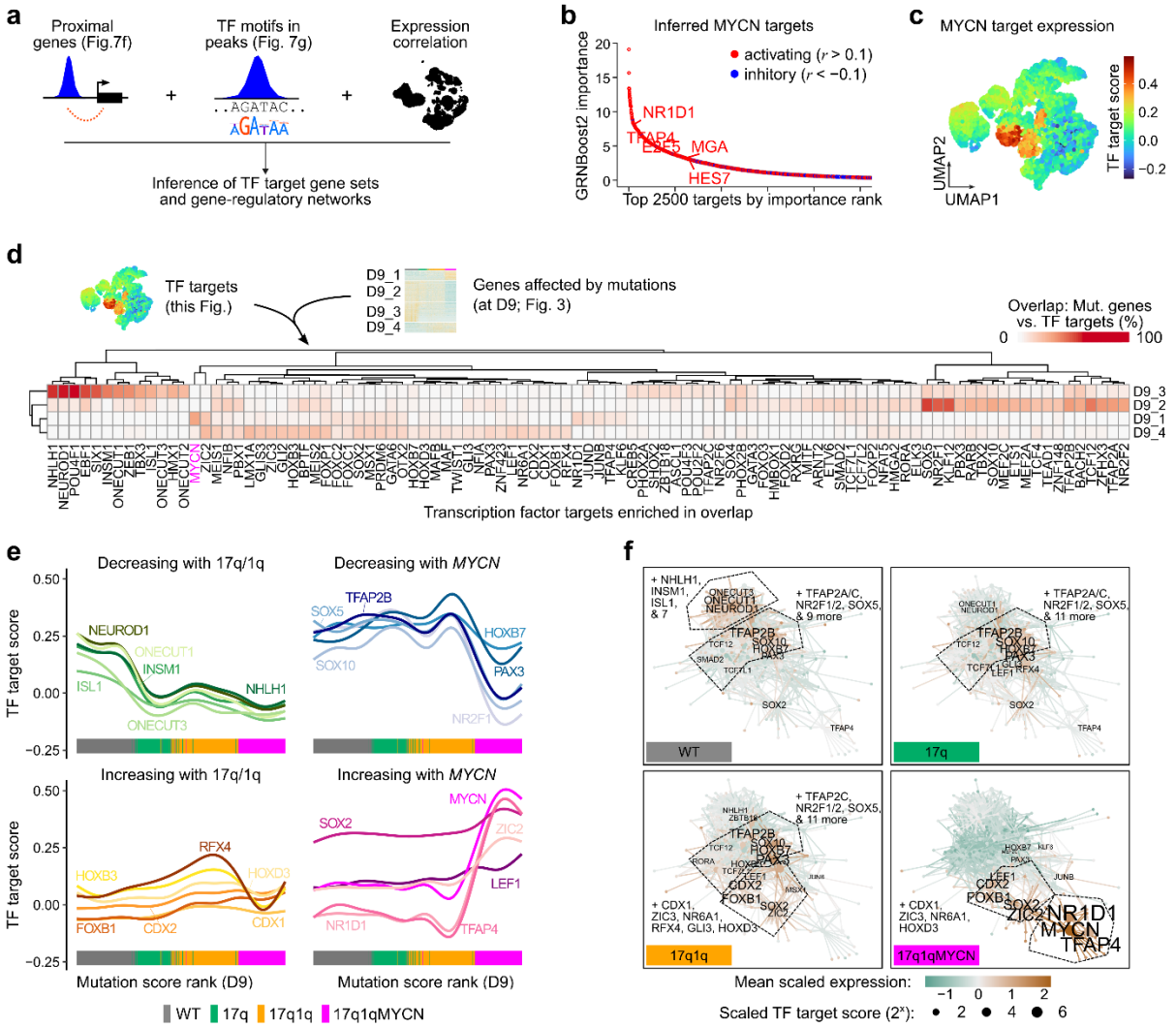


Figure 8. Copy number changes facilitate *MYCN*-mediated blockage of differentiation via developmental transcription factor networks.

a) To define putative target genes of TFs, we linked TF motifs identified in ATAC-seq peaks with proximal genes and additionally used the *GRNboost2* algorithm[83] to identify highly correlated TF-target gene candidates based on our scRNA-seq data.

b) Top 2500 targets of MYCN predicted by *GRNboost2*[83]. Putative targets without support in our ATAC-seq data (motif for TF in ≥1 peak near the gene) have been removed. We also calculated the Pearson correlation coefficient (*r*) between each TF and target gene to determine the direction of the putative interaction (*r*>0.1 = "activating", *r*<-0.1 = "inhibitory", others = "marginal"). The top 5 TFs have been highlighted. TF target gene sets are reported in **Supplementary Table 15**.

c) Average expression (Seurat module score) of the MYCN target gene set ("activated" targets from **panel d**) in our integrated scRNA-seq dataset (cp. **Fig. 3d**).

d) Heatmap displaying the percentage of genes in gene sets D9_1 to D9_4 (correlated with mutation score, cp. **Fig. 3e,f**) that overlapped with targets of the indicated TFs (one TF per column). All TF target sets with significant overlaps in at least one comparison are shown (hypergeometric test, *hypeR*[138]; $P_{adj} \leq 0.05$, $|\log_2\text{FoldChange}| \geq \log_2(4)$, frequency ≥ 5%). Enrichment results are also reported in **Supplementary Table 16.**

e) Smoothed line plots evaluating target gene expression (Seurat module score) for selected TFs from **panel f** with increasing mutations (mutation scores from **Figs. 3e,f**). We manually split

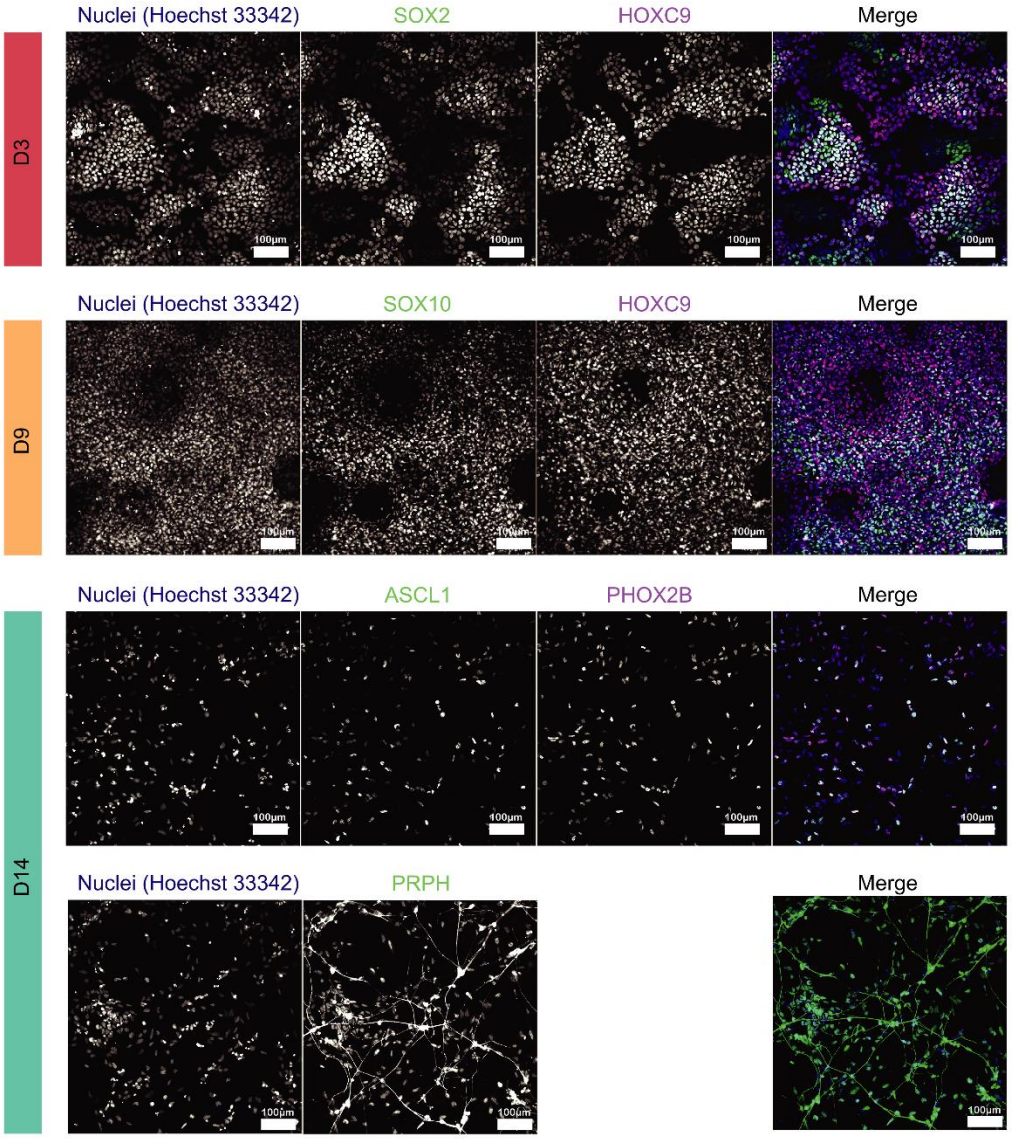
1836 the TFs into groups corresponding to target genes losing or gaining expression along the module
1837 score spectrum. The source cell line of each data point is indicated at the bottom.
1838 **f)** Gene-regulatory network diagrams visualizing putative TF-to-target relations for enriched TF
1839 targets (cp. **panels c-e**). In these diagrams, each node represents a TF or target gene, and each
1840 edge is a link between a TF and a target. We made these networks specific to cells from each
1841 condition (WT, 17q, 17q1q, 17q1qMYCN) by using colour to indicate the mean scaled
1842 expression of each gene in the respective cells at D9 (edges coloured by source TF) and node
1843 size to indicate the mean scaled T_{TF} target score of each TF. Only labels of TFs with positive
1844 scaled expression are shown and selected groups of TFs have been merged for visualisation. A
1845 network diagram with all node labels is shown in **Supplementary Fig. 13c**.

1846 **Abbreviations**: D9, day 9; R1-R9, chromatin region modules; TF, transcription factor; WT, wild-type
1847 H7 hESCs; sig., significant; r, Pearson correlation coefficient.

1848

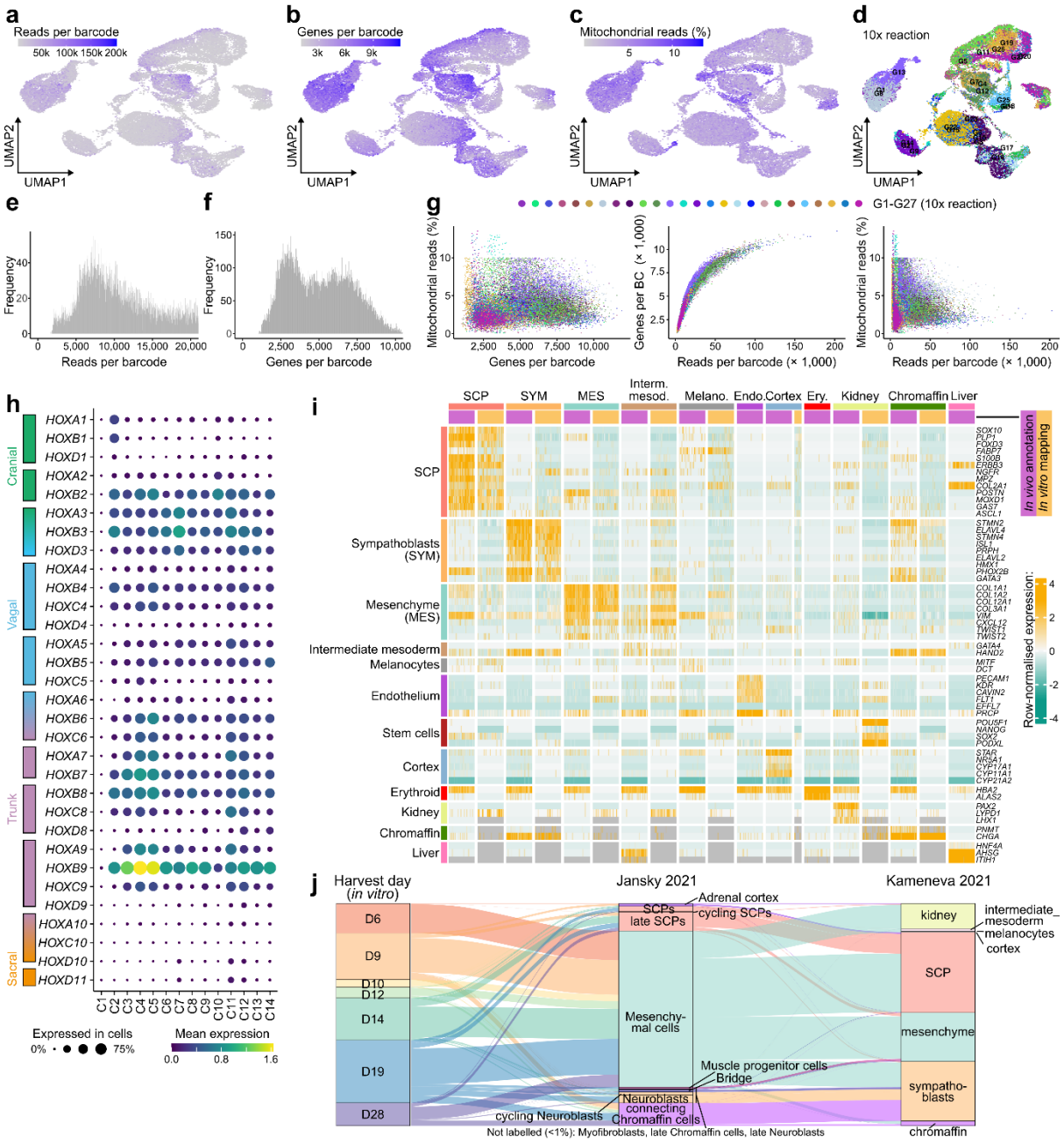
Supplementary figures

Supplementary Figure 1



Supplementary Figure 1 (related to Fig. 1). Immunofluorescence analysis of hESCs during trunk NC differentiation.
Immunofluorescence analysis of the expression of indicated markers at different time points during the differentiation of hESCs toward trunk NC and its derivatives.
Abbreviations: D3/9/14, day 3/9/14.

Supplementary Figure 2



Supplementary Figure 2 (related to Fig. 1). Quality control and reference mapping of single-cell RNA-seq data from wild-type hESC trunk neural crest differentiation.

a-d) UMAP plots showing quality covariates for the wild-type hESC dataset in **Figure 1**.

e-f) Histograms depicting the distribution of read depth (**e**) and number of genes detected (**f**) per cell barcode after quality control filtering.

g) Scatterplots comparing all quality control covariates (shown on **panels a-c**) from the same cell, plotted versus each other, coloured by 10x Genomics run (G1-G27; cp. **Fig. 1b**).

h) Bubble plot indicating the mean expression (colour) and fraction of cells expressing (size) each of the HOX genes per cluster. Genes have been ordered from cranial to sacral axis specification.

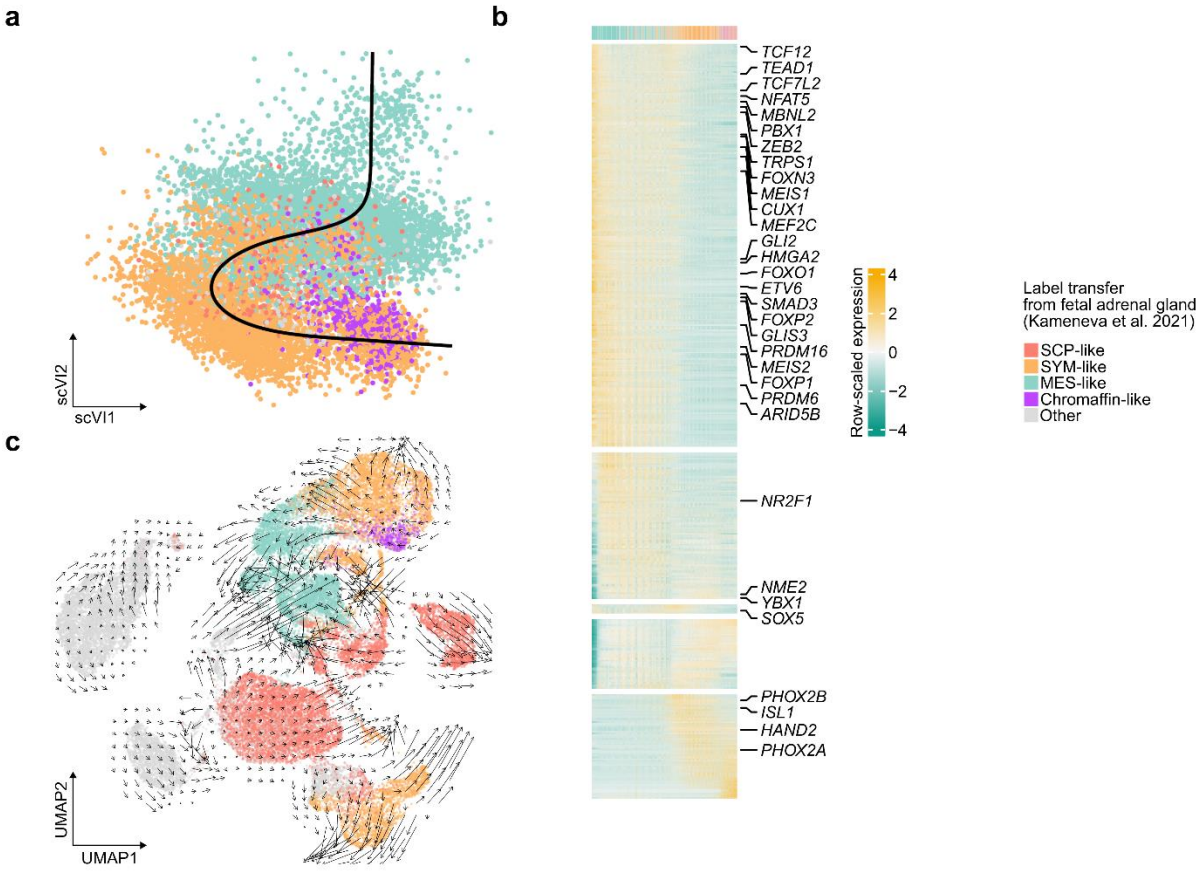
i) Side-to-side comparison of cell type marker expression in 200 annotated cells selected at random from the human adrenal gland reference[16] versus the top 200 high-confidence cells mapped to the same cell types in our WT *in vitro* trunk NC dataset. Rows are cell-type marker genes. Columns are cells first divided by cell type (separated with a grey line), then by dataset of origin (adrenal gland: purple, in vitro: orange). Expression values are depth-normalised per

experiment and row-scaled globally. Known stem cell markers were added to trace where the stem cell population would be spuriously mapped to, in this case kidney, an indicator that cells not found in the reference may be mapped to kidney. No *in vitro* cells were mapped to chromaffin or liver identities, leading to the absence of the respective *in vitro* columns. Cells erroneously mapped, absent from the reference, or lacking relevant cell type markers were classified as "other" and coloured grey in **Figure 1e**.

j) Alluvial plots comparing the mappings between cells in the *in vitro* dataset compared to two adrenal gland reference datasets[15,16]. Each "stream" indicates a group of cells that were mapped to one cell identity in the Jansky et al. (middle) and the Kameneva *et al.* (right) references (also indicated in colour). For example, cells that were labelled MES in Kameneva et al. (this is the foundation for the cell type labels used in this paper) also mapped to mesenchymal cells in the Jansky *et al.* reference, cells that mapped to SYM mapped to cycling neuroblasts and to neuroblasts. Cells that we labelled as SCPs split into cells that mapped to mesenchymal cells and late SCPs in Jansky *et al*., consistent with our observations that some of the "SCP-like" cells in our dataset represent a less mature, early SCP state (see main text).

Abbreviations: D0/3/9/10/12/14/19/28, day 0/3/9/10/12/14/19/28; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell precursor; SYM, sympathoblast; MES, mesenchymal.

Supplementary Figure 3



Supplementary Figure 3 (related to Fig. 1). Trajectories connecting gradients of transcriptionally similar cells at different developmental stages.

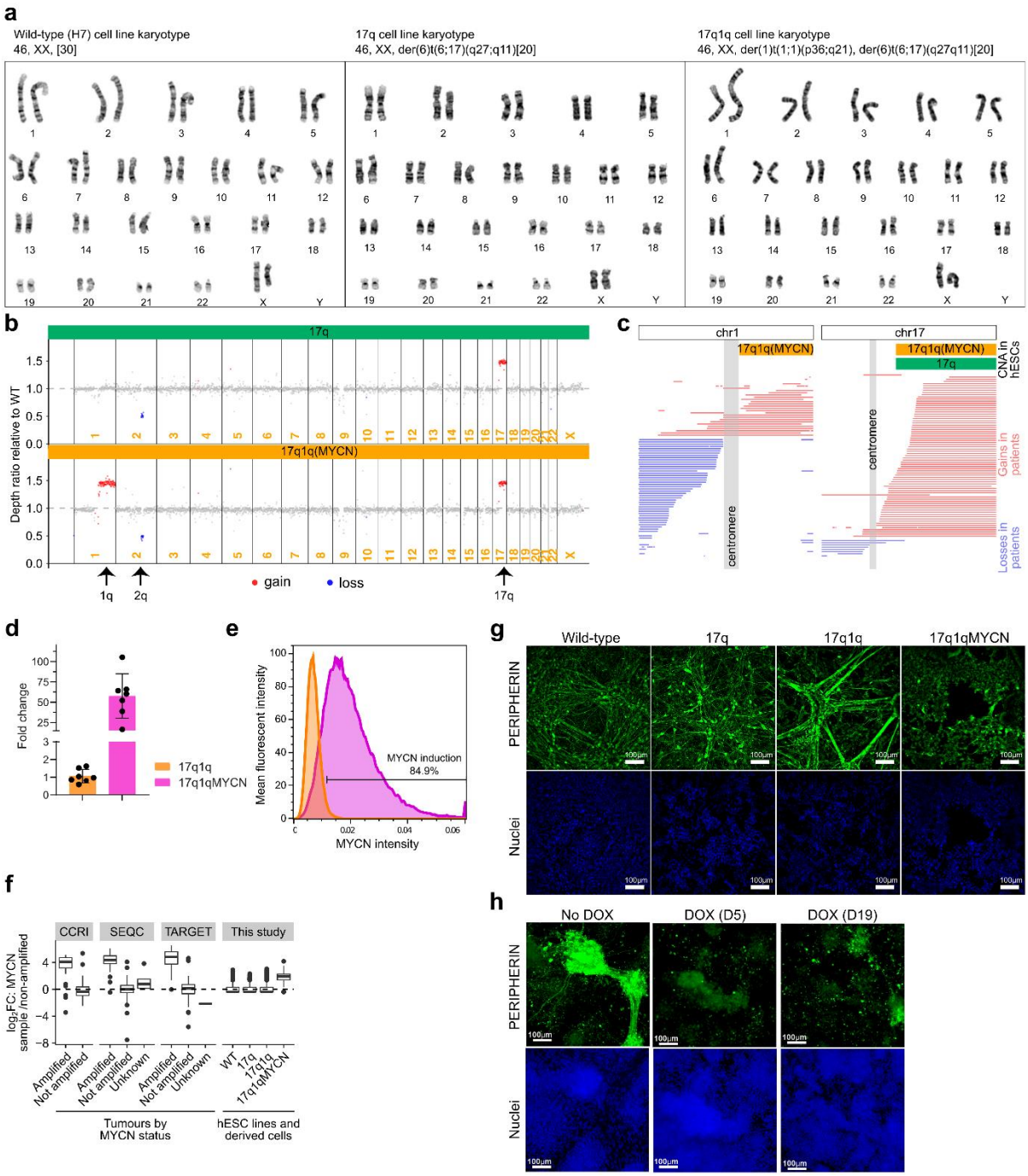
a) *Slingshot*[139] pseudotime trajectories (top) for wild-type clusters C11-C14. Cells were extracted from the main dataset and reprocessed (see "Basic scRNA-seq processing" in Methods), and trajectories were calculated on the first two scVI components.

b) Heatmap showing the top 400 genes with the strongest association with the trajectory as ranked by *tradeSeq*'s Wald test[140] (based on a *fitGAM* model with 5 knots). Highlighted genes are all the transcription factors found in the association test. **Supplementary Table 3** reports all genes found in this analysis.

c) RNA velocities calculated for the cells in **Figure 1d** using Velocyto[163].

Abbreviations: principal component; SCP, Schwann cell precursor; SYM, sympathoblast; MES, mesenchymal.

Supplementary Figure 4



Supplementary Figure 4 (related to Fig. 2). Genetic and phenotypic characterization of mutant hESC lines.

a) Cytogenetic analysis of the H7-derived hESC lines used in the study.

b) Plots of the depth ratio calculated between each sample and the parental control using Sequenza[120]. CNAs can be seen for chr1q and chr17q, and a small segment which is lost on chr2 (q23.3-24.3). CNA coordinates and SNVs in these cell lines are reported in **Supplementary Tables 4 and 5**.

c) Pile-up of segmental gains (red) and losses (blue) on chromosome 1 (left) and 17 (right) based on SNP array data of 88 NB samples[51] (thin lines) compared to CNAs in our 17q and 17q1q(MYCN) hESC lines (from panel **b**). The X axis corresponds to chromosome position. Each horizontal line depicts all the aberrations found for the same patient. For compatibility

with available annotations, coordinates in this plot are shown with respect to human genome reference GRCh37/hg19 (while all other analysis in this paper were done using hg38/GRCh38).

d,e) Analysis of MYCN expression at the transcript (**d**) and protein (**e**) level in D9 17q1qMYCN cultures after DOX treatment at day 5 vs untreated control.

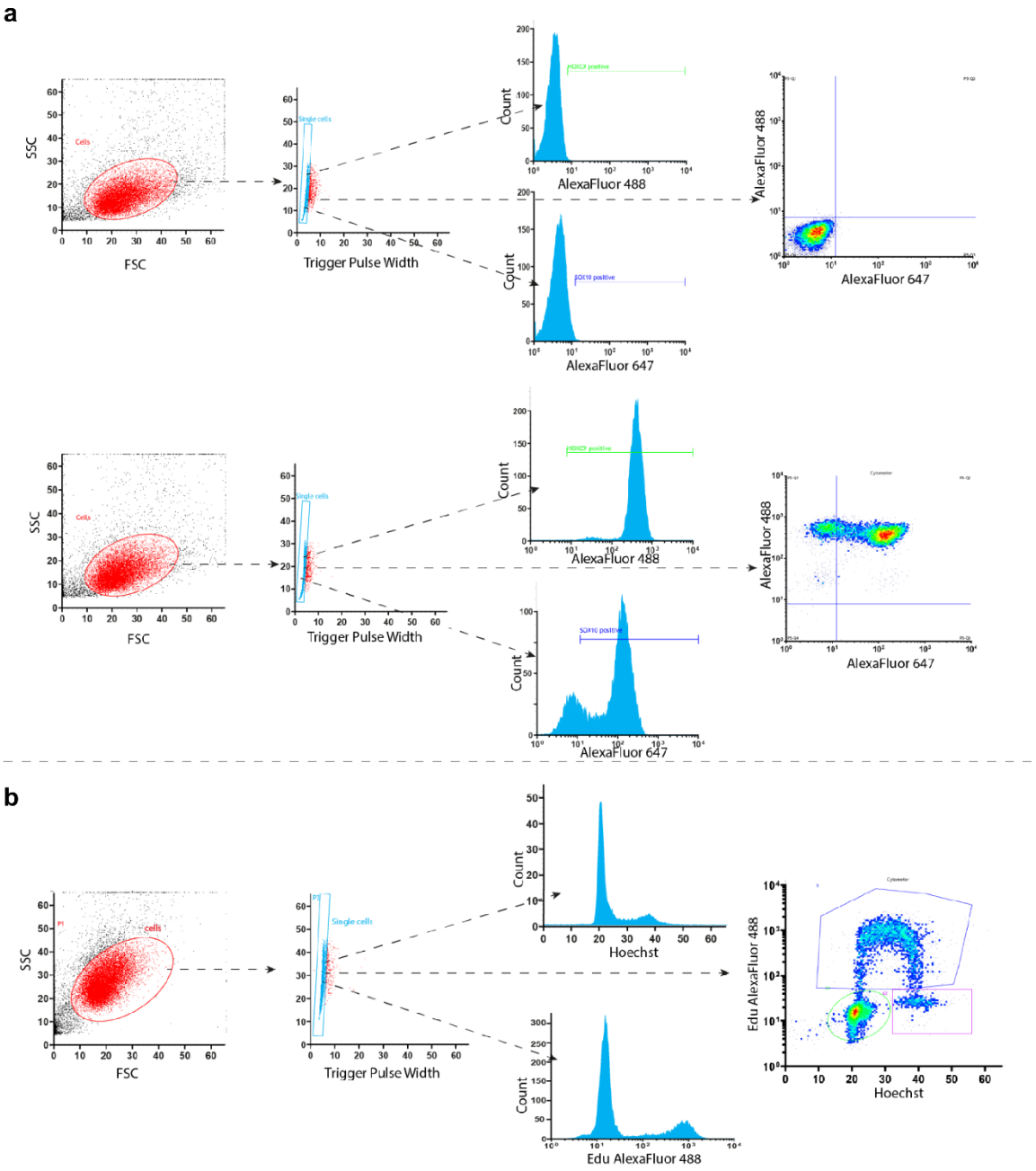
f) Comparison of *MYCN* expression in tumours and our engineered hESCs and their derivatives. The plots on the left show bulk RNA-seq data from three NB tumour compendia (CCRI, SEQC, TARGET) divided into cases with and without diagnosed *MYCN* amplification. The plot on the right ("This study") shows our scRNA-seq data divided by cell line (across all timepoints). The values in all plots are the $\log_2$ fold changes of over the mean of all non-amplified tumours (left plots) or all WT cells (right plot).

g) Immunofluorescence analysis of PERIPHERIN expression in D19 cultures following differentiation of hESCs with the indicated genotypes. Cell nuclei were counterstained using Hoechst 33342.

h) Immunofluorescence analysis of PERIPHERIN expression in D28 cultures following differentiation of 17q1q (No DOX) or 17q1qMYCN hESCs following DOX treatment at the indicated timepoints. Cell nuclei were counterstained using Hoechst 33342.

Abbreviations: WT, wild-type H7 hESCs; CNA, copy number alteration; DOX, Doxycycline.

Supplementary Figure 5



Supplementary Figure 5 (related to Fig. 2). Plots illustrating the gating strategy for the FACS performed in this manuscript.

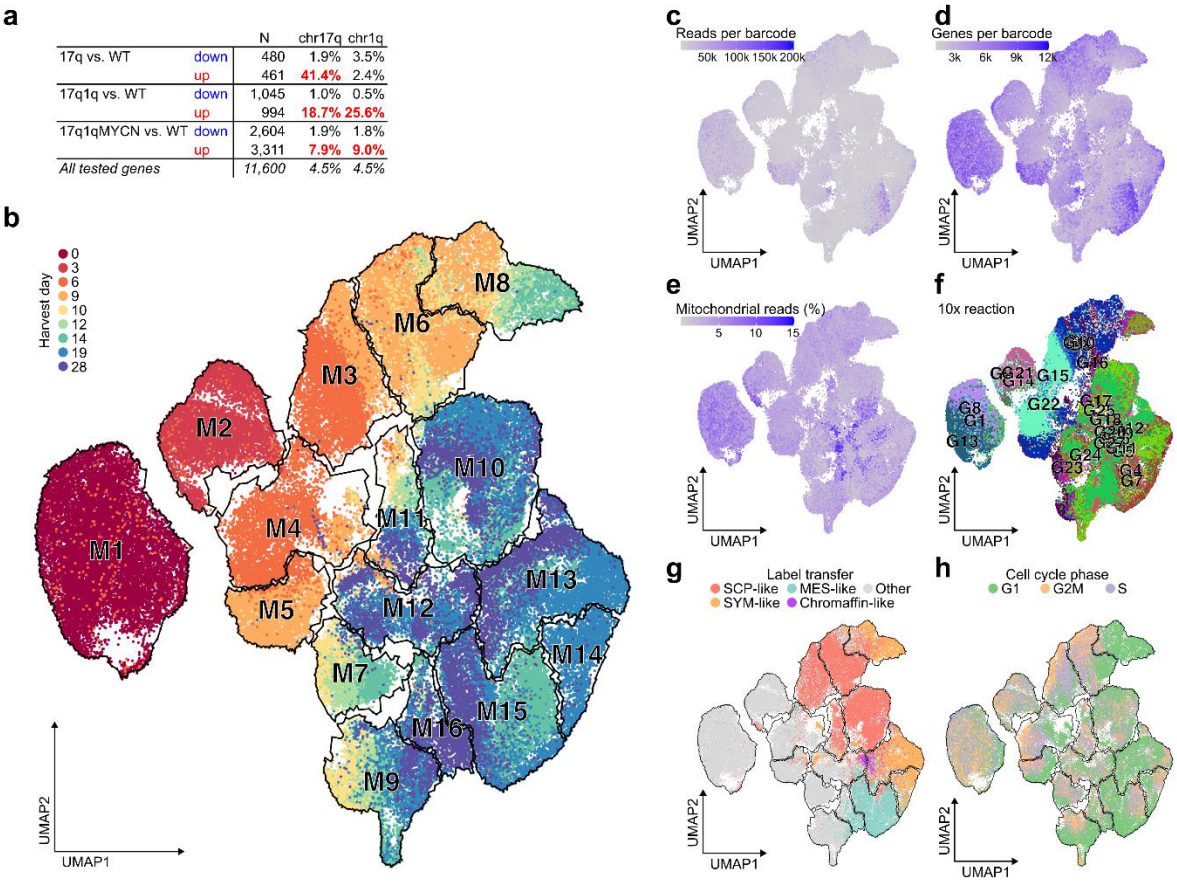
a) Secondary-only (negative) gating was set using cells differentiated at the correspondent stage stained only with secondary antibodies (detailed in the methods section). The scatter plots of Forward Scatter (FSC) and Side scatter (SSC) variables were used as the method to identify the main cell population size. Once the main population was identified, SSC and Trigger Pulse Width variables were further used to determine the single-cell population (doublet discrimination). The FITC (488) and Red (647) channels were used to set the baseline of fluorescence (upper panel) and subsequent positive gates (lower panel).

b) Cells were analysed in the flow cytometer (BD FACSJazz) using the 405 nm laser to detect the Hoechst staining and 488 nm to detect the Edu staining.

Supplementary Figure 6

a

		N	chr17q	chr1q
17q vs. WT	down	480	1.9%	3.5%
	up	461	**41.4%**	2.4%
17q1q vs. WT	down	1,045	1.0%	0.5%
	up	994	**18.7%**	**25.6%**
17q1qMYCN vs. WT	down	2,604	1.9%	1.8%
	up	3,311	**7.9%**	**9.0%**
All tested genes		*11,600*	*4.5%*	*4.5%*



Supplementary Figure 6 (related to Fig. 3). scRNA-seq analysis of differentiating wild-type and mutant hESCs.

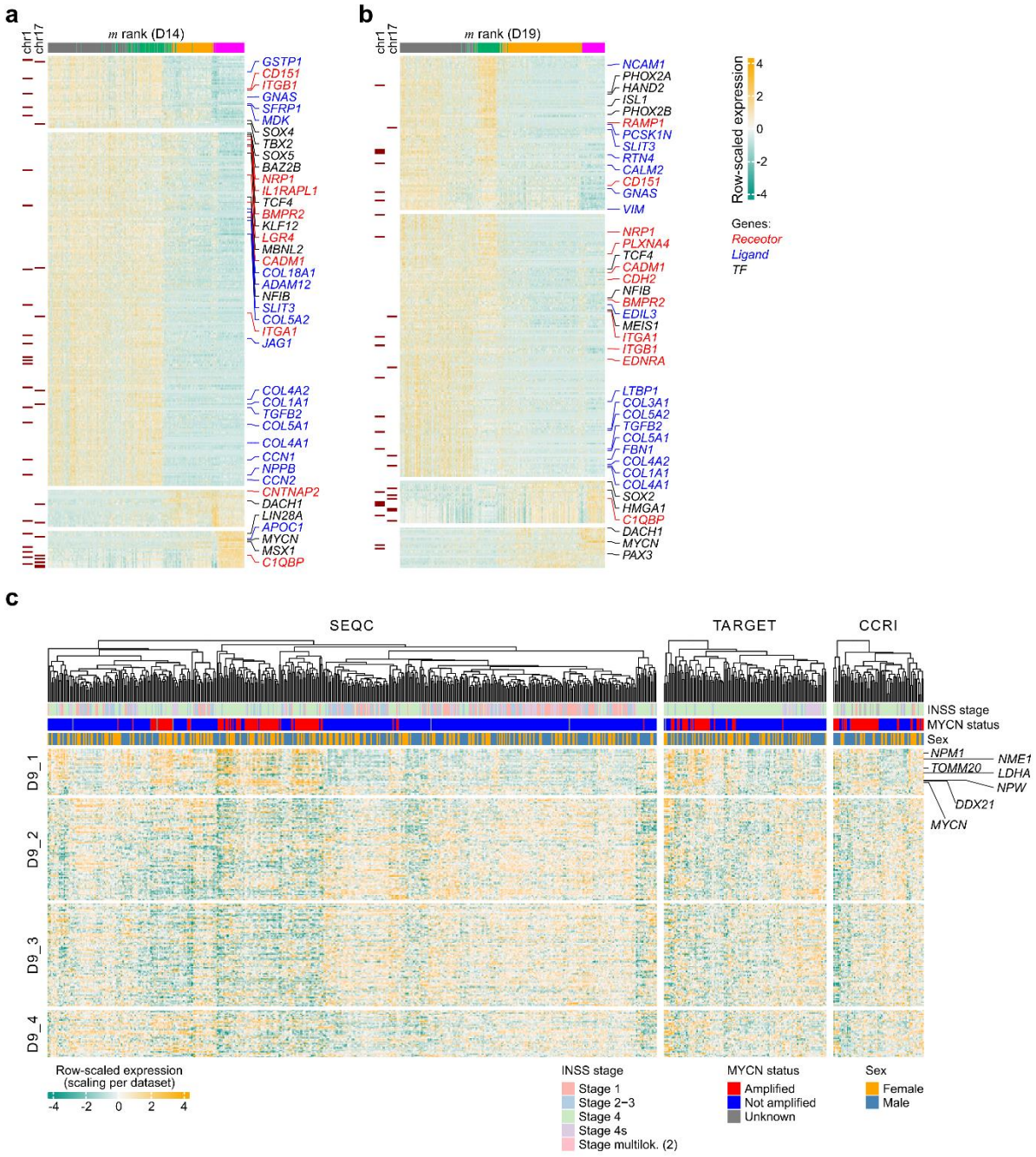
a) Overview of the number of differentially expressed genes (DEGs) in 17q, 17q1q, and 17q1qMYCN cells at D9 of differentiation compared to wild-type. The total number of DEGs is given (N), and the percentage of those genes that are located on chromosome arms chr17q or chr1q are indicated. Percentage values >5% have been highlighted (which also correspond to upregulated DEGs within known CNAs). DEGs are reported in **Supplementary Table 6**.

b) Cell clusters defined for the full *in vitro* trunk NC dataset scRNA-seq dataset generated in this study. Cluster marker genes are reported in **Supplementary Table 8**.

c-h) QC covariate plots: reads per cell (**panel (**), features per cell (**panel d**), % mitochondrial genes (**panel e**), replicates (one of up to three repeat experiments; **panel f**), developmental stage (sampling day; **panel g**), and computationally inferred cell cycle stage (**panel h**).

Abbreviations: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell precursor; SYM, sympathoblast; MES, mesenchymal.

Supplementary Figure 7



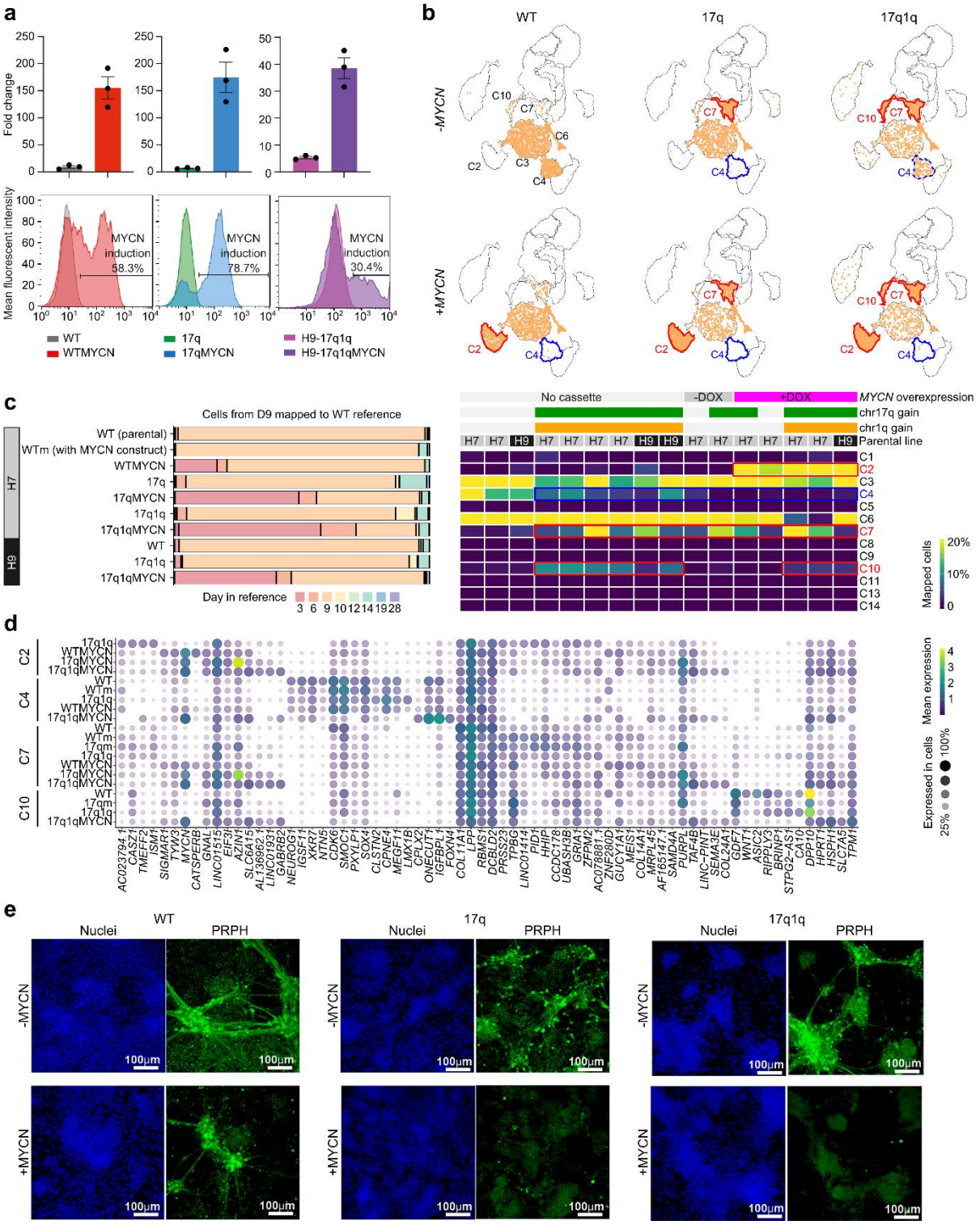
Supplementary Figure 7 (related to Fig. 3). Differential gene expression in mutant hESC-derived trunk NC and sympathoadrenal cells.

a,b) Heatmaps containing the genes correlated or anti-correlated with the mutation score *m* and for D14 (**a**) and D19 (**b**). Left annotation indicates whether the gene is on chromosome 1q or 17q, respectively. Transcription factors (black), receptors (red) and ligands (blue) have been highlighted. All mutation-score-related genes are reported in **Supplementary Table 9**.

c) Heatmap showing the expression of mutation-score-related genes (**Fig. 3f**, **Supplementary Table 9**) in public bulk RNA-seq data from three NB tumour compendia (SEQC, TARGET, CCRI). The heatmaps display the row-normalised transcript counts per gene and sample. The INSS stage, *MYCN* amplification status, and sex of each sample are indicated by the colour bars on top. Genes in D9_1 that are highly expressed in *MYCN*-amplified tumours are highlighted.

Abbreviations: TF, transcription factor; m rank, mutation score rank.

Supplementary Figure 8



Supplementary Figure 8 (related to Figs. 2,3). Comparison of multiple mutant hESC lines with the WT trunk NC differentiation reference.

a) Analysis of MYCN expression at the transcript (top) and protein (bottom) level in D9 cultures of the indicated genotypes after DOX treatment from day 5 vs untreated control.
b) scRNA-seq data mapped to the wild-type trunk NC reference (from **Fig. 1**). Top: Example glasswork UMAP plots depicting the destination clusters in the WT reference for cells of the WT, 17q, and 17q1q cells (H7 cell line) with and without *MYCN* overexpression. Clusters with an increased (red) and decreased (blue) number of mapped cells are indicated. Bottom:

Heatmap displaying the mapping of all 16 datasets including 1-4 replicates of each cell line / experimental condition. Cells were derived from two parental hESC lines (H7 and H9), with/without gains of chr17q and/or chr1q, and edited with a *MYCN* expression cassette (which can be active [+DOX] or inactive [-DOX]). Clusters with increased/decreased numbers of mapped cells have been highlighted with red/blue boxes, respectively.

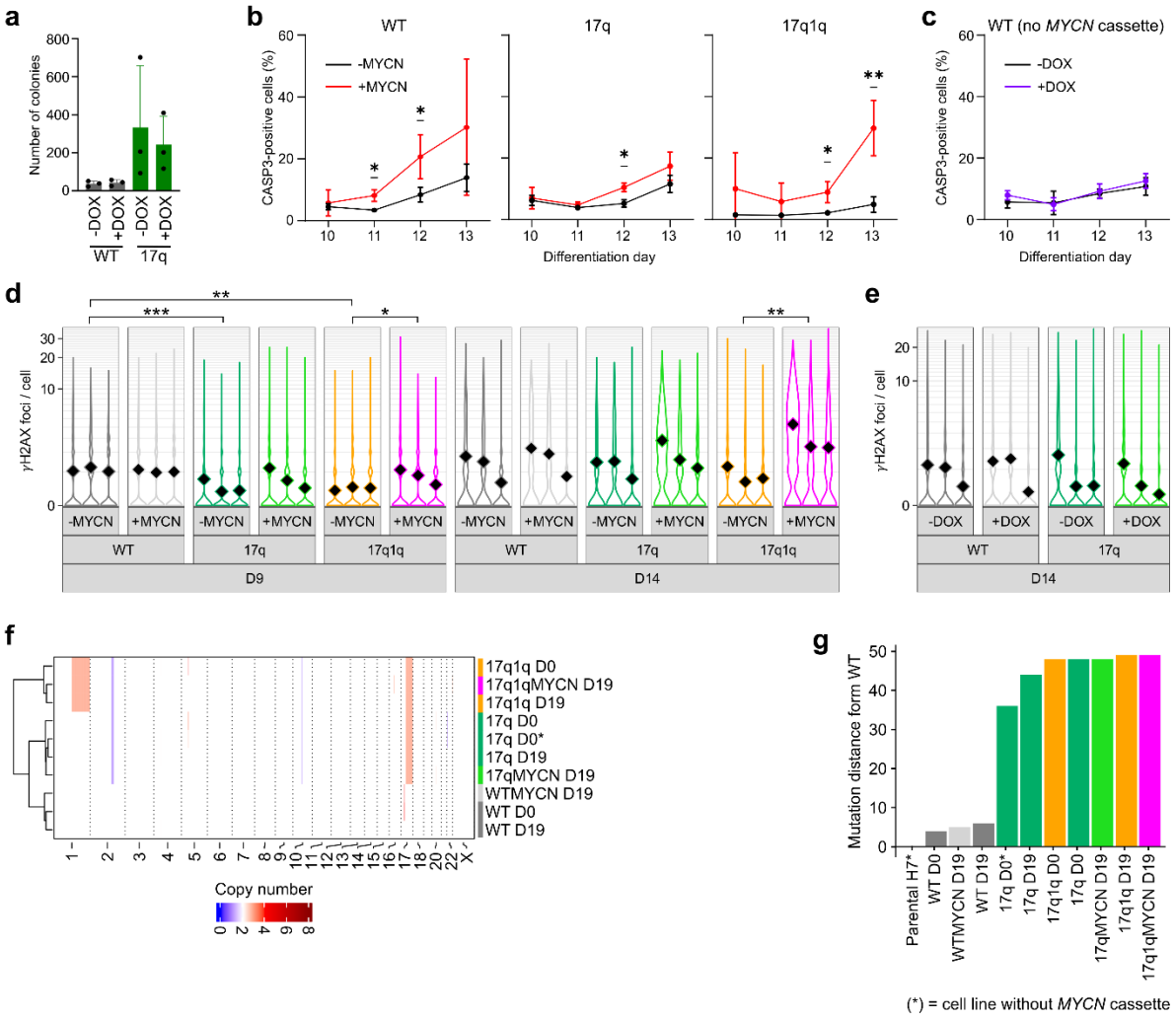
c) Barplots summarising the mappings from **panel b** for derivatives of each hESC line (top to bottom). Each bar indicates the proportion of cells (all collected at D9) that mapped to cells in the WT reference of a given developmental stage. The plots suggest that cells with *MYCN* induction mapped to earlier stages compared to WT.

d) Bubble plot showing marker genes of cells mapped to the WT cell clusters associated with the different genetic changes in **panel b (**C2, C4, C7 and C10). Only cells with positive percentages are shown. The size and colour of each circle indicate the percent of cells in the indicated group which express the gene and the average expression, respectively. Some gene expression programmes are affected by specific combinations of mutations and cell types, such as polyamine homeostasis gene *AZIN1* or P53 suppressor *PURPL* in 17qMYCN in C2 or C7.

e) Immunofluorescence analysis of PERIPHERIN (PRPH) expression in D19 cultures following differentiation of *MYCN*-overexpressing hESCs with the indicated genotypes in the presence and absence of DOX. Cell nuclei were counterstained using Hoechst 33342.

Abbreviations: DOX, Doxycycline.

Supplementary Figure 9



Supplementary Figure 9 (related to Fig. 4). Apoptosis, DNA damage, and mutations in mutant hESC-derived trunk NC differentiation.

a) Comparison of the number of colonies formed by D14 cells generated from unmodified WT and 17q cells in the presence and absence of DOX treatment with or without 17q gain (n= 3 biological replicates, error bars= SD, Ordinary One-way ANOVA test with Tukey correction). No statically significant differences were observed.

b) Time course flow cytometry analysis of cleaved Caspase-3 levels in differentiating trunk NC cells generated from hESCs of the indicated genotypes in the presence and absence of DOX treatment from day 5. Error bars= standard deviation, unpaired t-test). P values in comparisons: WTMYCN (Day 11 DOX vs NO DOX, p = 0.010797 = *; Day 12 DOX vs NO DOX, p= 0.033974 = *), 17qMYCN (Day 12 DOX vs NO DOX, p = 0.034166 = *), 17q1qMYCN (Day 12 DOX vs NO DOX, p = 0.020679 = *; Day 13 DOX vs NO DOX, p = 0.005686 = **).

c) Control comparison of Caspase-3 levels in untreated and DOX-treated WT controls. As in **panel b**.

d) Number of γH2AX foci per cell in differentiating cultures corresponding to indicated timepoints and genotypes in the presence and absence of DOX treatment. FDR values in comparisons (ANOVA with mixed effect model with negative binomial distribution): FDR *** < 0.001, ** < 0.01, * < 0.05. Comparisons performed: DOX status within a genotype and day, genotype vs WT per DOX status and day.

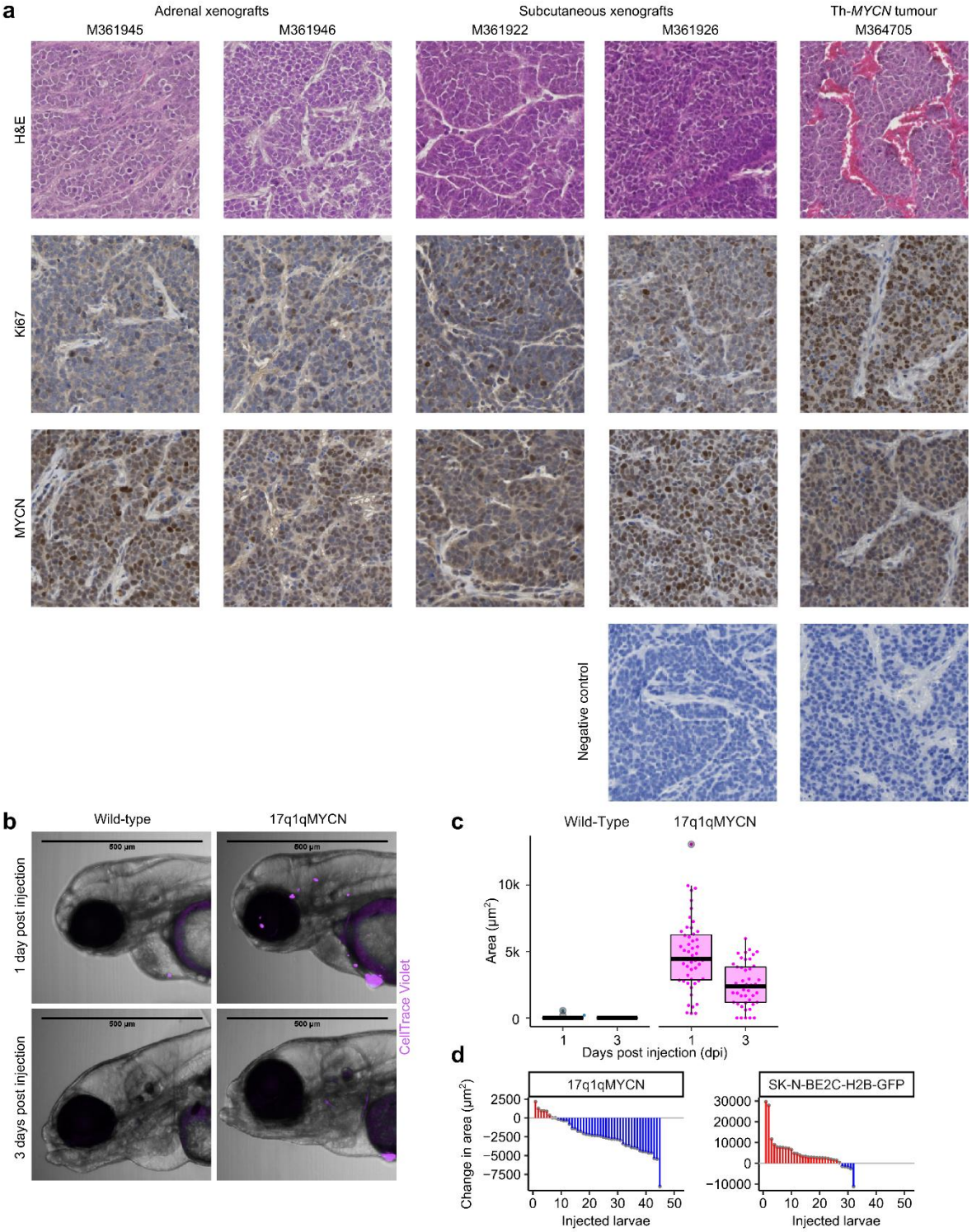
e) Control comparison of the number of γH2AX foci per cell in untreated and DOX-treated unmodified WT and 17q cells. As in **panel d**.

f) Heatmap containing CNA calls by Sequenza[120] in bins of 1Mbp. Dendrogram represents hierarchical clustering. CNA coordinates are reported in **Supplementary Table 5**.

g) Phylogenetic distance measured in mutations (SNVs) of each sample relative to the parental control using the phylogenetic analysis presented in **Fig. 4e**.

Abbreviations: WT, wild-type H7 hESCs; DOX, Doxycycline.

Supplementary Figure 10



Supplementary Figure 10 (related to Fig. 5). Mouse and zebrafish xenografts of WT, 17q1q, and 17q1qMYCN cells.

a) Representative sections from individual tumours following xenografting of 17q1qMYCN cells in the indicated locations and showing haematoxylin and eosin staining (top row), Ki67 (middle row) and MYCN expression (bottom row). Tumour sections from a Th-*MYCN* NB GEM model and negative controls are shown.

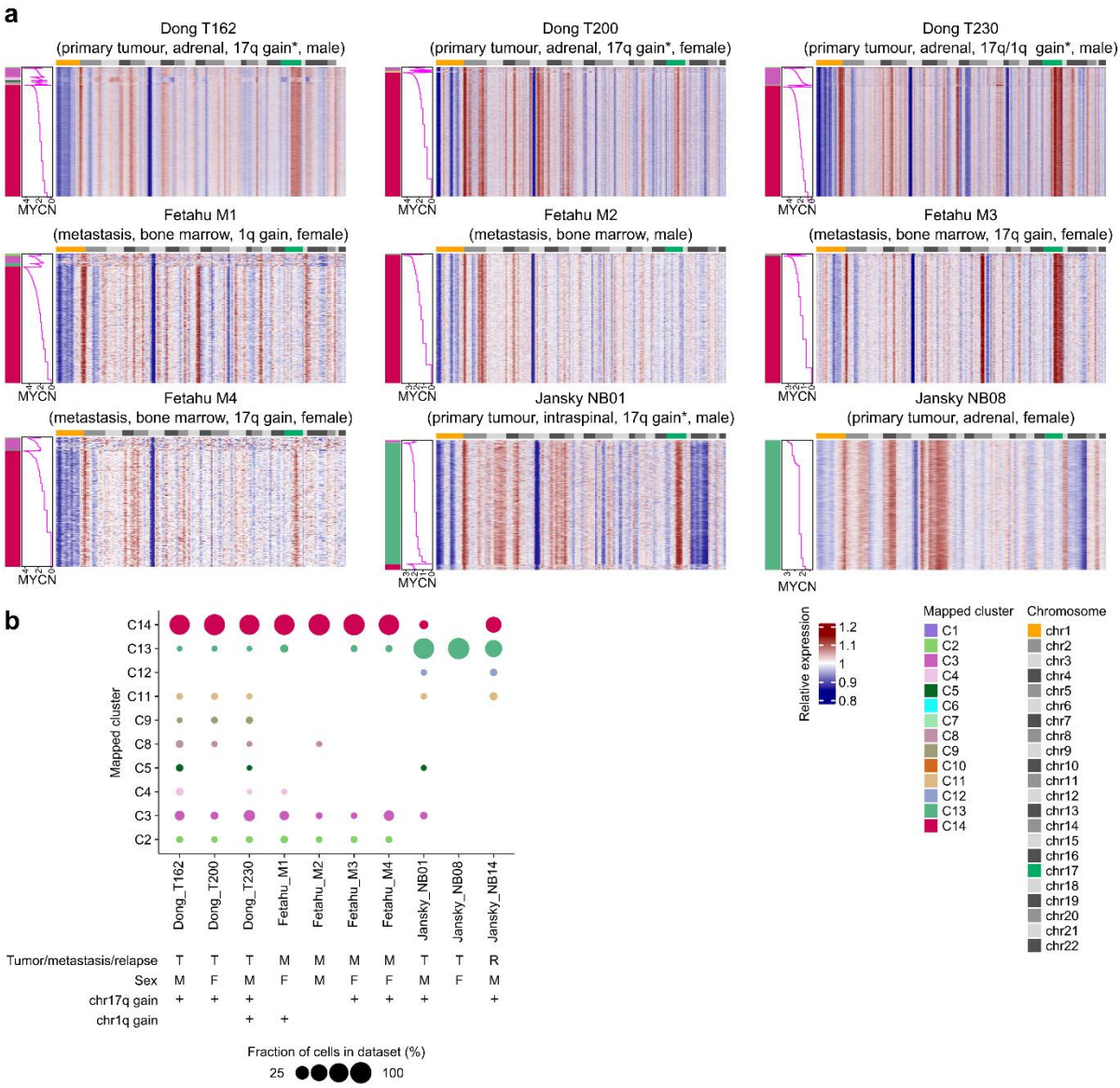
b) Representative images of zebrafish xenografted with WT or 17q1qMYCN cells labelled with CellTrace Violet at one day or three days post injection (top and bottom, respectively).

c) Quantification of the area covered by WT or 17q1qMYCN cells in zebrafish xenografts at 1 and 3 dpi. While 17q1qMYCN cells persist, WT cells are not maintained in zebrafish xenografts. Xenografts with WT cells (n = 11), and 17q1qMYCN cells (n = 51).

d) Waterfall plots depicting the change in tumour area for 17q1qMYCN at D9 of differentiation (~ NC stage) and SK-N-BE2C-H2B-GFP[71] cells in zebrafish xenografts from 1dpi to 3dpi.

Abbreviations: dpi, day post injection.

Supplementary Figure 11



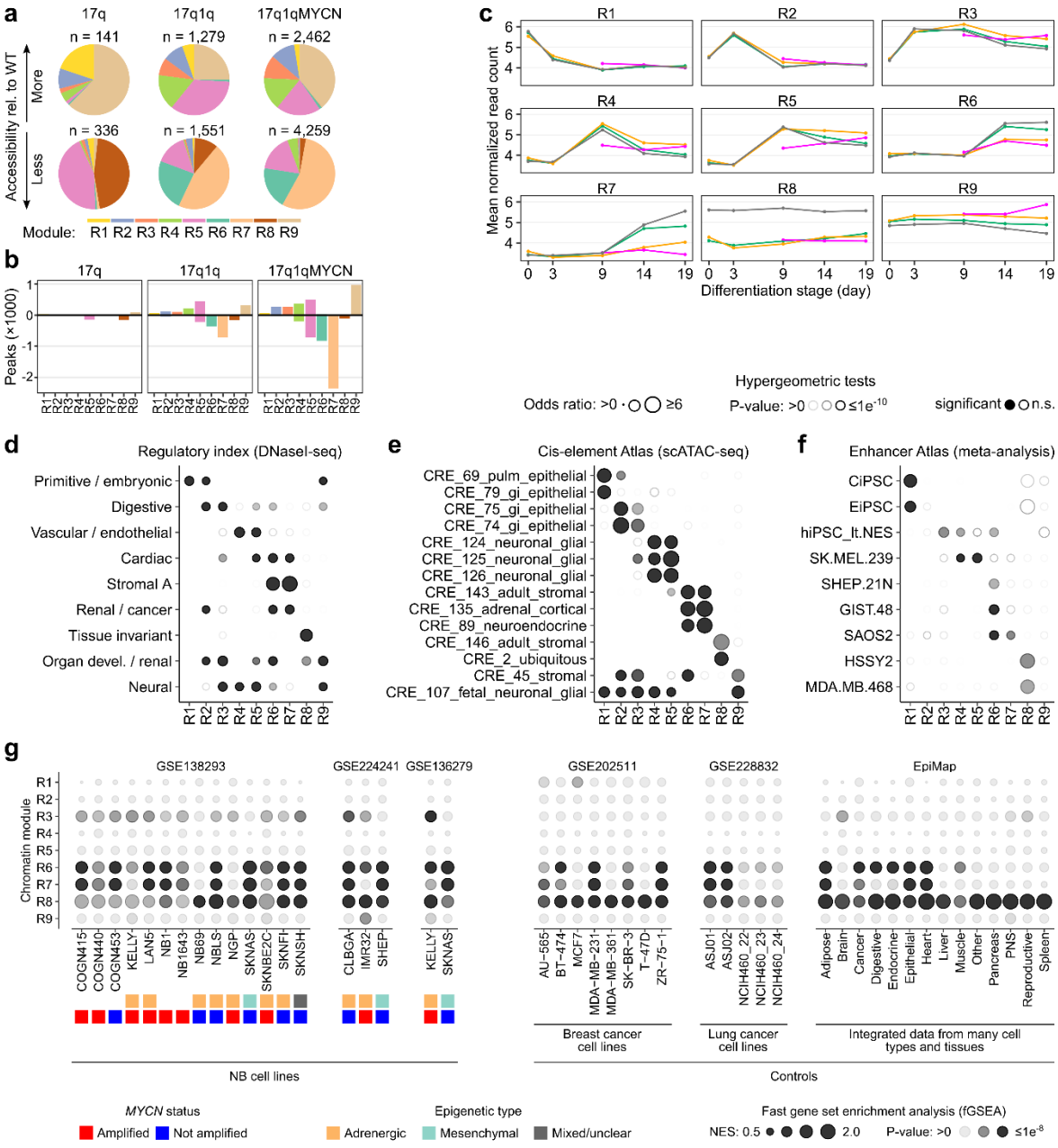
Supplementary Figure 11 (related to Fig. 6). Analysis of *MYCN*-amplified tumour cells and mapping to *in vitro* trunk NC differentiation.

a) *inferCNV*[142] profile heatmaps such as the one in **Fig. 6b** for the remaining 9 tumour datasets[15,17,23] not shown in **Fig. 6**. Each row (tumour cells) and each column (genes, ordered by genomic position), indicate the intensity of the CNA signal relative to non-tumour, HSC/immune cells from the same sample. All samples were curated and processed as described in **Fig. 6** and mapped to our wild-type trunk NC differentiation reference (**Fig. 1**). Cells are ordered first by matching cluster and then by *MYCN* levels within each cluster. Annotation of the chromosomes can be found on top of the heatmap. The tumour type (primary, metastasis, relapse), tissue, chr17q/1q gain status, and sex of each sample are indicated. The asterisk (*) marks cases in which the chr17q/1q gain status is based on the inferred inferCNV copy number profile shown in this figure panel.

b) Bubble plot summarising the mapping of tumour cells to our in vitro trunk NC reference (from **panel a** and **Fig. 6b**). The tumour type (primary, metastasis, relapse), sex, and chr17q/1q gain status are indicated. Ordering by study suggests a study-specific effect; for example, Fetahu samples are all bone marrow metastases and lack mesenchymal C11 cells.

Abbreviations: CNA, copy number alteration; T, primary tumour; M, metastasis; R, relapse; M, male; F, female

Supplementary Figure 12



Supplementary Figure 12 (related to Fig. 7). Chromatin accessibility in differentiating wild-type and mutant hESCs.

a) Pie charts indicating the fraction of differentially accessible regions belonging to chromatin modules (**Fig. 7e**) in mutant hESCs compared to WT (left to right). Up- and down-regulated regions are shown separately (top vs. bottom row). The total number (*n*) of regions in each category is indicated.

b) Barplots indicating the number of up- (positive numbers) and down-regulated (negative numbers) peaks from each comparison of mutant hESCs vs WT, split by chromatin module (R1-R9).

c) Line plots summarizing the dynamics of accessibility per module and cell line throughout differentiation. Each data point indicates the mean normalised read count.

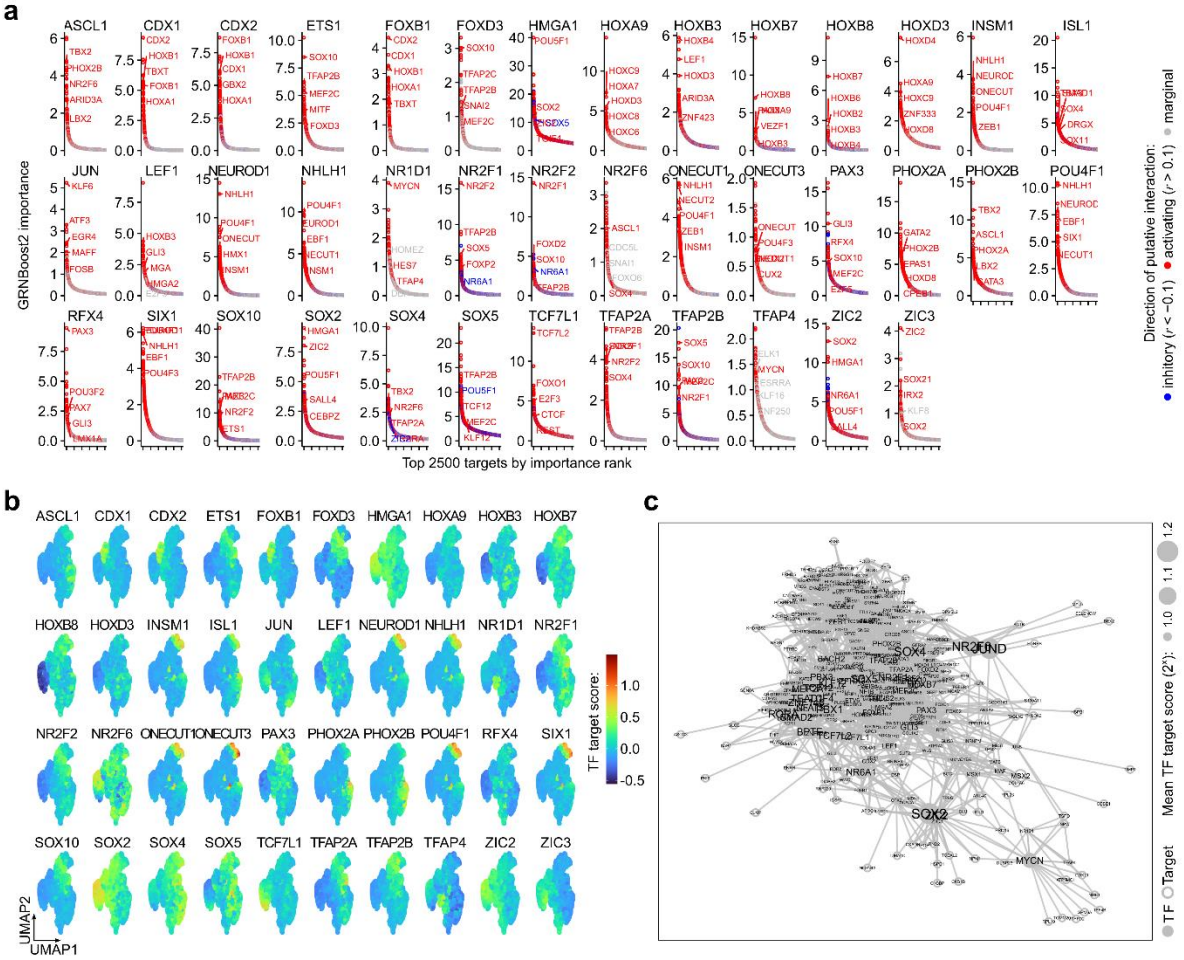
d-f) Enrichment analysis of overlaps between regions belonging to the nine chromatin modules (from left to right) and annotated reference regions from the Regulatory Index[80] (based on DNaseI-seq; **panel d**), Cis-element Atlas[81] (based on scATAC-seq analysis; **panel e**) and the

Enhancer Atlas[82] (based on a meta-analysis of many different data; **panel f**). The size and transparency of circles indicate the odds ratio and P-value, respectively (hypergeometric test, *hypeR*[138]). Significant results are indicated with filled circles ($P_{adj} \leq 0.05$). The top enrichments per stage have been selected for visualisation (all results are shown in **panel e**) and all results are reported in **Supplementary Table 14**.

g) Bubble plots showing the outputs of a fast gene set enrichment analysis (*fgsea*[156]) of open chromatin in external data compared to our chromatin modules. Each sub-panel indicates data from a different source (from left to right) including three collections of NB cell lines[75,157,158] (GSE138293, GSE224241, GSE136279), two adult cancer cell lines [159,160] (GSE202511, GSE228832), and a meta-analysis of human tissue data[161] (https://epigenome.wustl.edu/epimap). The epigenetic type[44,45] and *MYCN* amplification status of each NB cell line are indicated.

Abbreviations: WT, wild-type H7 hESCs; R1-R9, chromatin modules identified in **Fig. 7e**; NES, normalised enrichment score; n.s., not significant.

Supplementary Figure 13



Supplementary Figure 13 (related to Fig. 8). Analysis of transcription factor target sets and gene-regulatory networks.

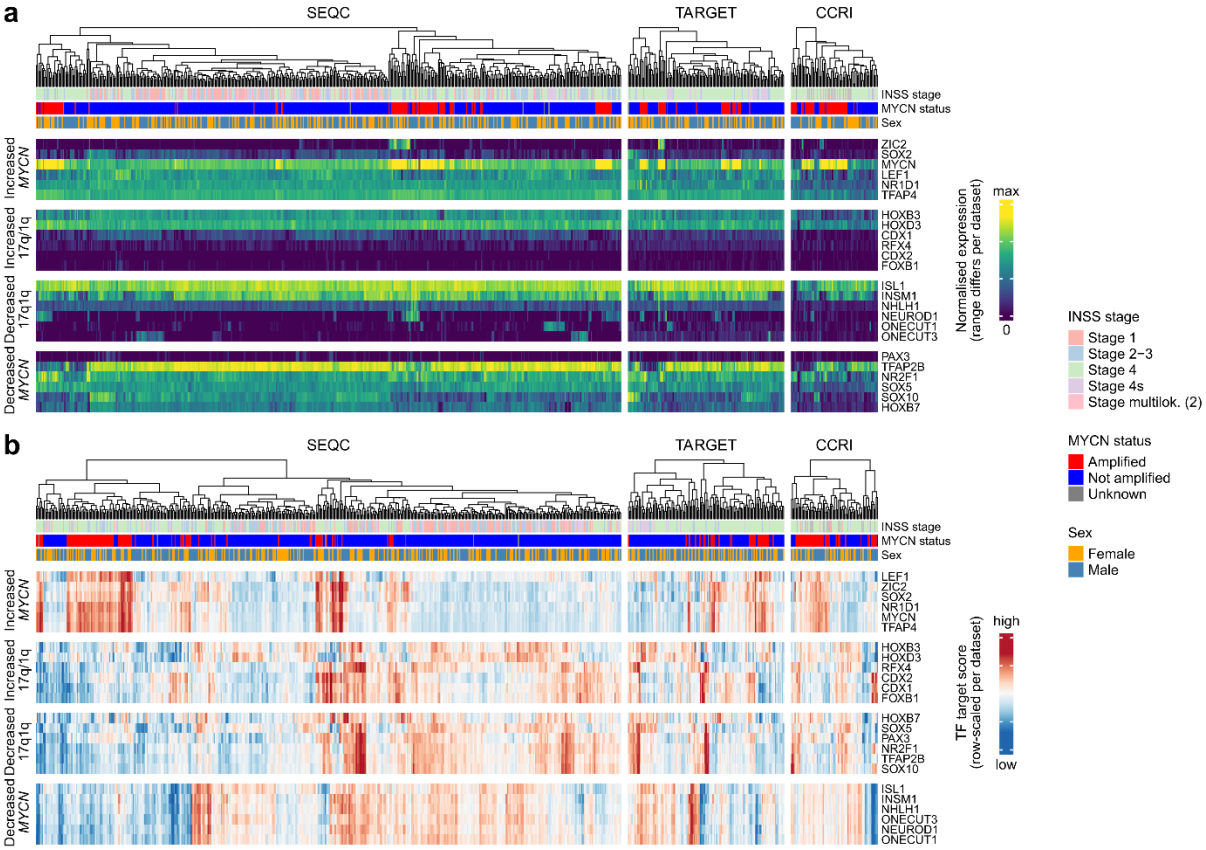
a) Top 2500 targets of selected TFs as predicted by *GRNboost2* algorithm[83] based on our scRNA-seq data. Putative targets without support in our ATAC-seq data (motif for TF in ≥1 peak near the gene) have been removed. We also calculated the Pearson correlation coefficient (*r*) between each TF and target gene to determine the direction of the putative interaction ($r>0.1$, "activating"; $r<-0.1$, "inhibitory"; others, "marginal"). The top TFs in the target lists have been highlighted. TF target gene sets are reported in **Supplementary Table 15**.

b) Average expression (Seurat module score) of the target gene sets (matching "activating" targets of the TFs in panel **a** in our integrated scRNA-seq dataset (cp. **Fig. 3d**).

c) Gene-regulatory networks diagram visualising putative TF to target interactions for the genes in gene sets D9_1 to D9_4 (cp. **Fig. 3e,f**) and enriched TF targets (cp. **Fig. 8d**) In these diagrams, each node represents a TF or target gene, and each edge is a link between a TF and a target. Node size is proportional to the mean target score of the indicated TFs (fixed size for non-TF nodes).

Abbreviations: TF, transcription factor; r, Pearson correlation coefficient.

Supplementary Figure 14



Supplementary Figure 14 (related to Fig. 8). Expression of transcription factors and their target genes in public RNA-seq datasets.

a) Expression of 24 selected TFs (from **Fig. 8e**) in public bulk RNA-seq data from three NB tumour compendia (SEQC, TARGET, CCRI). The heatmaps display the normalised transcript counts per gene and sample, colours have been scaled from 0 to the maximum per source dataset (dark blue to yellow). The INSS stage, *MYCN* amplification status, and sex of each sample are indicated by the colour bars on top. Some of the examined TFs are not or very weakly expressed in the investigated samples from tumours.

b) Heatmaps for the same data and TFs as in panel a, but here heatmap values indicate the TF target score (Seurat module score), a summary of the expression of putative target genes of each TF in the respective sample. TFs with high activity in 17q1qMYCN cells (the 6 TFs on top) are active in *MYCN*-amplified samples.

Abbreviations: INSS, International Neuroblastoma Staging System; TF, transcription factor.

Supplementary tables and files

Supplementary Table 1 (related to Figs. 1, 2, 3, 4, 7, 8). scRNA-seq, WES, and ATAC-seq dataset overview

Supplementary Table 2 (related to Fig. 1). scRNA-seq cluster marker genes (WT)

Supplementary Table 3 (related to Fig. 1). scRNA-seq markers of SCP-SYM-MES transition states

Supplementary Table 4 (related to Fig. 2, 4). Whole-exome sequencing SNVs.

Supplementary Table 5 (related to Fig. 2, 4). Whole-exome sequencing CNAs.

Supplementary Table 6 (related to Fig. 3). scRNA-seq MUT vs. WT differentially expressed genes

Supplementary Table 7 (related to Fig. 3). scRNA-seq MUT vs. WT enrichment results

Supplementary Table 8 (related to Fig. 3). scRNA-seq cluster marker genes (WT+MUT)

Supplementary Table 9 (related to Fig. 3). scRNA-seq genes correlated to mutations

Supplementary Table 10 (related to Fig. 6). Tumour-in vitro markers

Supplementary Table 11 (related to Fig. 6). Survival analyses

Supplementary Table 12 (related to Fig. 7). ATAC-seq regions (peaks) and chromatin modules

Supplementary Table 13 (related to Fig. 7). ATAC-seq differential accessibility analysis

Supplementary Table 14 (related to Fig. 7). ATAC-seq chromatin module enrichment results

Supplementary Table 15 (related to Fig. 8). Transcription factor target genes

Supplementary Table 16 (related to Fig. 8). Transcription factor enrichments