Neuroblastoma-associated chromosomal aberrations drive cell identity loss in human neural crest via disruption of developmental regulators

Ingrid M. Saldana-Guerrero[1,2,3,*], Luis F. Montano-Gutierrez[4,*], Christoph Hafemeister[4,$], Dylan Stavish[1,2,$], Lisa E. Shaw[5], Irfete S. Fetahu[4], Andrea Wenninger-Weinzierl[4], Caterina Sturtzel[4], Celine Souilhol[1,2], Sophia Tarelli[1,2], Mohamed R. Shoeb[4], Marie Bernkopf[4], Polyxeni Bozatzi[4], Maria Guarini[6], Eva Bozsaky[4], Michelle C. Buri[4], Eva M. Putz[4], Peter W. Andrews[1], Ivana Barbaric[1,2], Helen E. Bryant[3], Martin Distel[4], Sabine Taschner-Mandl[4], Matthias Farlik[5], Anestis Tsakiridis[1,2,#], Florian Halbritter[4,#]

[1] Centre for Stem Cell Biology, School of Biosciences, The University of Sheffield, Sheffield, U.K.
[2] Neuroscience Institute, The University of Sheffield, Sheffield, U.K.
[3] Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, The University of Sheffield, Sheffield, U.K.
[4] St. Anna Children's Cancer Research Institute (CCRI), Vienna, Austria
[5] Medical University of Vienna, Department of Dermatology, Vienna, Austria
[6] CeMM Research Center for Molecular Medicine of the Austrian Academy of Science, Vienna, Austria

*,$ 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 are thought to arise from transformed embryonic cells, which often carry large copy number variants (CNVs). However, it remains unclear how CNVs contribute to embryonic tumourigenesis due to a lack of suitable models. Here we employ human embryonic stem cell (hESC) differentiation to assess the effects of chromosome 17q/1q gains, which are prevalent in the embryonal tumour neuroblastoma (NB). We show that CNVs impair the specification of hESC-derived trunk neural crest (NC) cells and their sympathoadrenal derivatives, the putative cells-of-origin of NB. Overexpression of *MYCN*, whose amplification co-occurs with CNVs in NB, exacerbates the differentiation block, and enables tumourigenic cell proliferation. We find links between disrupted cell states *in vitro* and tumour cells and connect these states with stepwise disruption of developmental transcription factor networks. Together, our results chart a possible route to NB and provide a mechanistic framework for the CNV-driven initiation of embryonal tumours.

Introduction

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 variants (CNVs), which co-occur with mutations in tumour suppressors or tumourigenic transcription factors (TFs)[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 adrenal glands 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[12–22]. CNVs 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[23–27]. These often co-occur with amplification of the *MYCN* oncogene[23,27–31] (at least one CNV in >95% of *MYCN*-amplified tumours[32]), 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 CNVs disrupt embryonic cell differentiation and lead to tumourigenesis.

Here, we used a human embryonic stem cell (hESC)-based model to experimentally dissect the links between NB-associated CNVs, *MYCN* amplification, and tumour initiation. We interrogated the stepwise specification of trunk NC and sympathoadrenal lineages by directed differentiation of isogenic hESC lines with chr17q/1q gains and inducible *MYCN* overexpression. We found that CNVs derail differentiation by potentiating immature NC progenitor phenotypes. Combining CNVs with *MYCN* overexpression further distorted differentiation trajectories and coincided with the acquisition of tumourigenic hallmarks. Furthermore, aberrant cell states captured in our model reflect heterogeneous cell populations in NB tumours. Finally, we discovered an extensive re-wiring of chromatin connecting the observed transcriptional and functional aberrations with a dysregulated network of developmental TFs. Collectively, our data put forward a CNV-driven distortion of trunk NC and sympathoadrenal differentiation as a priming mechanism for subsequent MYCN-induced tumour initiation.

Results

Differentiation of *in vitro*-derived human trunk NC cells reveals discrete developmental trajectories toward sensory and sympathetic neurons

We have previously described an efficient strategy for the *in vitro* production of human trunk NC, sympathoadrenal progenitors, and sympathetic neurons from hESCs[33,34]. Our protocol involves treatment with defined cocktails of pathway agonists/antagonists that induce neuromesodermal-potent axial progenitors (NMPs) at day 3 of differentiation (D3)[35], 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[33] (**Figs. 1a, S1**).

We first employed this protocol for the differentiation of karyotypically normal hESCs (H7[36]) and performed droplet-based single-cell RNA sequencing (scRNA-seq) at key differentiation stages (D0, D3, D9, D14, D19) to examine the resulting cell populations (**Table S1**). We obtained 13,665 cells passing quality control, which we allocated to 11 distinct clusters (C1-C11) (**Figs. 1b, S2a-g**). Expression of *HOX* paralogous groups 6-9 indicated early programming of a trunk axial identity starting from D3, in line with previous findings[33,37] (**Fig. 1c**). Bioinformatic analysis identified characteristic marker genes of each cluster, reflecting the progressive differentiation of trunk NC (**Fig. 1d; Table S2**). For example, cell clusters at D9 included a subpopulation (C4) expressing genes indicative of a trunk NC and early Schwann cell precursor (SCP) identity (e.g., *SOX10, TFAP2B, PLP1*[16,38]) and two interconnected entities (C5-C6) marked by dorsal root ganglion (DRG) and sensory neuron markers (*ONECUT1/2*, *NEUROG1*, *POU4F1*[39]), and NOTCH pathway components (*HES5/6*, *DLL1*). The D14 cluster (C7) was characterised by sympathoadrenal/autonomic progenitor markers (*GATA3*, *ASCL1*[40,41]). Cells at D19 consisted of three distinct fractions with mature SCP-like (cluster C8; *CRYAB*, *POSTN*, and *IGFBP5*[42]), with sympathoblast-like (C11; *PHOX2A/B*, *ELAVL4*[16,43]), or with mesenchymal (C9; *COL1A1*, *FN1*) features, in line with findings showing that trunk NC and SCPs are competent to generate mesenchyme[42,44,45]. A fourth cluster (C10) exhibited both sympathoblast and mesenchymal features, suggesting a bridge between cell identities.

To validate that cell types in our hESC-derived model resembled their *in vivo* counterparts, we bioinformatically compared our data to cell types from the adrenal glands of human embryos[16] (**Figs. 1e, S2h,i**). This mapping confirmed cell populations with SCP-like, sympathoblast-like (SYM), and mesenchymal features (MES). Other foetal, non-NC lineages (e.g., endothelia) or late neuroendocrine derivatives (e.g., chromaffin cells) were absent. Intriguingly, the mapping also revealed two distinct groups of SCP-like cells at D9 (C4-C5) and D14-D19 (C8). Quantification of SCP signature strength indicated that cluster C8 comprised SCPs closely matching those seen in foetal adrenal glands, while clusters C4-C5 matched the cells in this reference to a lesser extent (**Fig. 1f**). This may be because the latter corresponds to an immature NC or early SCP state not captured in the advanced developmental stages in the reference[16] (weeks 6-14 post conception). Moreover, C11 SYM cells matched the

autonomic sympathoblasts in the reference (**Fig. 1g**), whereas C6 had a weak SYM signature (**Fig. 1g**) and instead expressed DRG-related genes (e.g., *NEUROD1/D4*[46]; **Fig. 1d***;* **Table S2**), indicating that these cells were developing towards sensory neurons. RNA velocity[47] supported that C8 SCPs represented a terminal state emanating from sympathoadrenal progenitors (C7) of D14, while C4-C5 SCP-like cells of D9 appeared to give rise to the DRG-like cells in C6 (**Fig. 1h**).

Opposing gradients of overlapping SYM and MES signatures (**Fig. 1g**) and RNA velocity (**Fig. 1h**) also indicated that the intermediate cluster C10 comprised cells at a fork in trunk NC development, as observed in mice[46]. Since MES- and SYM-like cells have also been described in NB cell lines[48–52], we sought to study C9-C11 cells along a pseudotime trajectory (**Figs. 1i, S3; Table S3**). This analysis revealed transcripts associated with NB and epithelial-to-mesenchymal transition (EMT) marking an intermediate MES-SYM state, e.g., WNT-antagonist *SFRP1*, transcriptional regulators *NR2F1/2*[53], and chemotaxis gene *RARRES2*[54].

Together, our data indicate that differentiation of hESC-derived trunk NC cells involves two branching events: (i) an early commitment toward a DRG and sensory neuron fate under the influence of extrinsically supplemented WNT-BMP and endogenous NOTCH signalling; or (ii) the late generation of BMP-SHH-driven multipotent SCP/sympathoadrenal progenitors, which subsequently give rise to three distinct cell types: mature SCPs, MES, and SYM.

CNVs and *MYCN* cumulatively disrupt human trunk NC differentiation

Having established a reliable model of trunk NC lineages, we next asked how gains of chr17q and chr1q impacted NC development and how they interacted with overexpression of oncogene *MYCN*, which often co-occurs with CNVs in NB[23,27–32]. CNVs frequently arise spontaneously in hESC cultures due to adaptation and selection[55]. We exploited this fact, to establish clonal isogenic hESC lines from a diploid H7 background (‘WT’) carrying one or two NB-associated CNVs (**Figs. 2a, S4a,b**): (i) a gain of chromosome arm 17q11-qter (‘17q’)[56], and (ii) a gain of chr1q via unbalanced translocation with the second chromosome 1 within the 17q background (‘17q1q’). 17q1q hESCs were further engineered to include a Doxycycline (Dox)-inducible *MYCN* expression cassette to mimic *MYCN* amplification in a temporally controlled manner (‘17q1qMYCN’). In our experiments, we induced *MYCN* overexpression using Dox at D5 (when cells adopt a definitive NC identity[33]) to avoid bias toward central nervous system differentiation, as seen upon *MYCN* overexpression in earlier pre-NC progenitors[57]. Dox treatment of 17q1qMYCN resulted in robust induction of MYCN at D9 (**Fig. S4c**).

Equipped with these three ‘mutant’ hESC lines, we performed differentiation toward trunk NC and carried out scRNA-seq as described above, yielding a combined dataset comprising 45,949 cells (**Table S1**). To assess how differentiation was affected in each mutant, we bioinformatically mapped the transcriptomes of mutant cells to the reference of normal trunk NC differentiation (cp. **Fig. 1b**). While many 17q cells intertwined with all WT cell types, fewer 17q1q cells advanced beyond WT D14, and 17q1qMYCN cells rarely mapped to mature cell types except for SCP-like cells in cluster C8 (**Fig.**

2b). Altogether, 17q1q and 17q1qMYCN cells matched WT cells of earlier developmental stages, suggesting they were delayed in their 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 fewer SYM- and MES-like cells in 17q1q and 17q1qMYCN (**Figs. 2d,e**). For cells mapped to the respective cell types, we observed a stronger SCP signature in 17q and 17q1q, while the expression of MES/SYM genes was weaker (**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**). The reduced ability of 17q1qMYCN hESCs to differentiate toward trunk NC derivatives was also reflected by their failure to generate PERIPHERIN-positive neuronal axons (**Fig. S4d**) and morphological changes such as rounding up and formation of tight, dome-like colonies (**Fig. S4e)**.

Differential analysis identified 721 (17q vs. WT), 1,043 (17q1q vs. WT), and 3,116 (17q1qMYCN vs. WT) differentially expressed genes (DEGs) at D9 (**Table S4**). As expected, many upregulated genes were located within the known CNVs (43% within chr17q for 17q cells; 23% within chr17q and 25% within chr1q for 17q1q cells; **Fig. S5a**). Pathway analysis revealed an enrichment of genes related to E2F and MYC targets as well 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; Table S5**). These perturbed pathways may contribute to deregulation of expression outside CNVs (e.g., upregulation of oxidative phosphorylation and downregulation of G2-M checkpoint-related genes in 17q1qMYCN; **Fig. 3a**), and therefore to the changes in cell phenotypes observed.

To better resolve the molecular impact of each mutation, we integrated all datasets into a joint projection of WT and aberrant trunk NC differentiation (**Figs. 3d, S5b-h**; **Table S6**). The strongest changes were found in 17q1qMYCN, which formed disconnected cell clusters not normally produced in our protocol, including a subpopulation expressing genes indicative of vascular/endothelial differentiation (e.g., *CD34, KDR, PECAM1*; **Figs. S5i-l**). 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) correlated with mutations (**Figs. 3f, Fig. S5m,n; Table S7**): Gain of CNVs led to a loss of genes involved in trunk NC differentiation and cell death (e.g., chromaffin cell-associated gene *PEG3* and WNT-antagonist *SFRP1*[58]) and induction of NB-associated transcripts such as *MSX2* and *CNTNAP2*[59] (**Fig. 3f**). *MYCN* overexpression in 17q1qMYCN repressed genes related to NC development (e.g., *TFAP2A/B* and nuclear receptors *NR2F1/2*[38,60,61]), and additionally triggered the induction of NOTCH target *HES7* and genes associated with metabolic changes/amino acid level regulation (e.g., *NR1D1,*

YBX3[62–64]). Strikingly, *SFRP1* and *NR2F1/2* were also found to mark the SYM-MES transition state in WT differentiating sympathoadrenal cells (cp. **Fig. 1i**), while NOTCH-signalling has been implicated in SYM-MES plasticity in NB cell lines[50], suggesting that mutation-mediated shifts in the balance between SYM and MES fates may underlie NB tumourigenesis.

We conclude that NB-associated CNVs alter the differentiation landscape of hESC-derived trunk NC lineages by promoting SCP-like entities at the expense of mature sympathoadrenal cell types. In conjunction with *MYCN* elevation, they (a) block differentiation, (b) cause a loss of cellular identity, and (c) trigger atypical transcriptional programmes, including the ectopic generation of endothelial cells even under pro-neural differentiation culture conditions.

Impaired trunk NC differentiation correlates with acquisition of tumourigenic hallmarks

We next examined whether CNVs/*MYCN* amplification led to the acquisition of tumourigenic hallmarks. We first carried out cell cycle analysis of trunk NC cells (D9) generated from each hESC line by monitoring EdU (5-ethynyl-2´-deoxyuridine) incorporation via flow cytometry. We observed a significant increase of cells in S-phase in 17q1qMYCN (P=0.0001, two-way ANOVA with Tukey correction; **Fig. 4a**) indicating a faster cell cycle, consistent with NB tumours and cell lines[65]. Conversely, no significant difference was found between 17q or 17q1q cells and WT. Immunofluorescence analysis of Ki-67 expression further showed that 17q1qMYCN cultures exhibited a higher proliferation rate by D14 (SCP/sympathoblast stage) compared to their CNV-only counterparts (17q1q vs 17q1qMYCN, P = 0.0001; 17q vs 17q1qMYCN, P <0.0001; **Fig. 4b**).

We next tested how CNVs/*MYCN* influenced colony formation, another hallmark of tumourigenesis. Low-density plating of trunk NC cells (D9) and image analysis showed increased clonogenicity in 17q1qMYCN, while 17q or 17q1q cells behaved like WT (**Fig. 4c**). At D9, 17q1qMYCN cells formed large, tight, and dome-like colonies, while the 17q1q colonies were smaller, spread out, and contained differentiated cells with neurites. Time-lapse imaging showed that the 17q1q cultures were composed of cells with a higher degree of motility compared to 17q1qMYCN (**Videos S1, S2**).

Finally, we examined the behaviour of 17q1qMYCN- and WT-derived trunk NC cells *in vivo* by labelling the cells with a fluorescent dye (CellTrace Violet) and injecting them into the perivitelline space of zebrafish larvae on day 2 post fertilisation. 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 (**Fig. 4d**). 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[66]) resulted in engraftment with subsequent tumour cell growth in 84% of larvae (**Fig. S6**).

Together, our results demonstrate that CNV-carrying trunk NC cells transit into a pre-tumourigenic state under the influence of *MYCN* overexpression, as reflected by the acquisition of altered cellular properties reminiscent of cancer hallmarks.

In vitro differentiation of mutant hESCs captures NB tumour cell heterogeneity

We asked whether the *in vitro* NB-like development could provide insights into NB heterogeneity. To this end, we first categorised our full *in vitro* reference dataset (of WT and mutant cells) into 17 mega-clusters based on gene expression and similarity to *in vivo* cell types (**Fig. 5a**, **S7**). And interrogated the expression of common NB diagnostic markers (**Fig. 5b**). For example, high levels of *B4GALNT1* (involved in the production of the ganglioside GD2) and *CHRNA3* were seen in SYM-like cells of D14/D19, and *DDC* and *DBH* were specific to cells at the intersection of SCP and SYM identity (**Fig. 5b)**. In contrast, *NCAM1* was widely expressed in all NC cells (**Fig. 5b)**. These data indicate that typical NB diagnostic markers report distinct developmental stages of human trunk NC/sympathoadrenal specification *in vitro*. The combination of multiple markers may thus provide clues to tumour origin.

Next, we tested whether cells in our model indeed reflected cell states in tumours. To this end, we collected scRNA-seq data from eleven *MYCN*-amplified NB tumours from three independent sources[15,17]. For each dataset, we curated *MYCN*$^+$ tumour cells and bioinformatically mapped these to our reference (**Fig. 5c**). For example, this approach matched cells of tumour dataset *Dong_230*[17] to SYM-like, SCP/SYM transitional, sensory neuron-like, and 17q1qMYCN-like mega-clusters (**Fig. 5d**). While tumour cells appeared karyotypically homogeneous (including a chr17q gain) and all expressed *MYCN*, the expression of NB markers differed substantially among those mapped to different mega-clusters **(Fig. 5e)**. Interestingly, cells with high *DBH* expression mapped to the SCP/SYM transitional mega-cluster, consistent with *DBH* expression *in vitro (***Fig. 5b***)*. Mapping to the *in vitro* reference helped detect such *DBH*$^+$, SCP/SYM-like cells in at least three other tumours (**Fig. S8**), suggesting it may be a recurrent cell subtype that can be directly modelled *in vitro*. We also detected the differentiation markers (cp. **Fig. 1d**) in distinct tumour cell subsets, e.g., *PHOX2B*, *GATA3,* and *ASCL1* marked cells at the SCP/SYM junction, while tumour cells mapped to clusters consisting of 17q1qMYCN cells ("HDMYCN") had a decreased expression of most markers, consistent with the eroded cell identity observed earlier (**Fig. 5f**).

Extending the *in vitro* reference mapping to all eleven tumours portrayed differences between datasets, such as one tumour with *MYCN*+ endothelial-like cells similar to those that emerged from *MYCN*-induced developmental distortion *in vitro* (**Figs. 5g, S8**). Interestingly, we found SYM-like and SCP/SYM-like cells in almost all datasets, and even rare populations of *MYCN*+ MES cells in two tumours. Jointly, these observations demonstrate that our *in vitro* model generates cell types that transcriptionally resemble NB cell subpopulations and that it facilitates the systematic dissection of intra-tumour heterogeneity in NB tumours.

CNVs and *MYCN* disrupt the reconfiguration 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[48,49,67–71]. We therefore hypothesised that disruption of developmental TFs also underpins the aberrant differentiation observed in our mutant hESCs (**Figs. 2-4**) and employed the assay for transposase-accessible chromatin followed by sequencing (ATAC-seq)[72] to profile chromatin accessibility in the same samples used for scRNA-seq analysis (n = 51; **Table S1**). Chromatin accessibility serves as a proxy for the dynamic regulatory activity during differentiation. For instance, the promoters of hESC regulator *POU5F1* and trunk NC regulator *SOX10* were accessible only at D0 and D9, respectively (**Fig. 6a**), and the *PHOX2B* promoter had reduced accessibility in 17q1q and 17q1qMYCN at D19 (**Fig. 6b**).

Unsupervised analysis of chromatin patterns on a global level showed that WT and 17q hESCs changed consistently throughout differentiation (**Fig. 6c**). 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 (**Tables S8, S9**). As in our DEG analysis, we found an increasing number of regions with altered accessibility in 17q (n = 477 regions), 17q1q (n = 2,825), and 17q1qMYCN (n = 6,683; **Fig. 6d**). In total, there were 45,580 regions with differential accessibility in at least one comparison, which we divided into nine chromatin modules R1-R9 (**Fig. 6e**). 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 hESCs derivatives fell within these modules (**Figs. S9a,b**). 17q1q and 17q1qMYCN cells failed to close chromatin that is usually specific to D3 and D9 (R4, R5) and conversely to open chromatin of late sympathoadrenal differentiation (R6, R7; **Fig. S9c**). Additionally, modules R8 and R9 comprised regions with reduced and increased accessibility in mutant hESC derivatives, respectively, independent of differentiation stage.

We sought to annotate our chromatin modules by looking for overlaps with genomic regions accessible *in vivo*[73–75] (**Table S10**). In line with our transcriptome data, we found a stepwise change toward chromatin resembling differentiated tissues, e.g., neural tissues in R4/R5 and mesenchyme in R6/R7 (**Figs. S9d-f**). Next, we examined genes near the chromatin modules (**Fig. 6f**). 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-C11 from **Figs. 1b,c**). For example, chromatin module R7 (accessible in late differentiation stages, lost in mutants) was linked to clusters C9/C10 (MES-like cells). Moreover, we examined TF binding motifs in each module to identify potential upstream regulators (**Fig. 6g**). Also here, 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 development in R6/R7 (e.g., PHOX2A/B)[46]. Finally, we found enriched overlaps of modules R7, R8, and R9 with super-enhancers

associated with subsets of NB tumours[68] with mesenchymal characteristics, with non-*MYCN*-amplified low-risk tumours, and with non-*MYCN*-amplified high-risk tumours, respectively (**Fig. 6h**). No significant overlap was found with super-enhancers specific for *MYCN*-amplified NB.

Together, our results suggest a systematic reprogramming of chromatin throughout trunk NC differentiation. In cells with CNVs and *MYCN* overexpression, this orderly reconfiguration of chromatin was severely disrupted, providing a plausible mechanism for the observed developmental defects.

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

Finally, we sought to investigate the links between CNV/*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 found a stepwise alteration of expression from WT to 17q1qMYCN at D9 comprising four gene sets: D9_1–D9_4 (cp. **Figs. 3a,b**). We hypothesised that these mutation-linked gene sets were also regulated by specific TFs and therefore employed an algorithm to identify TF targets based on correlated expression patterns[76] (**Fig. 7a**). This analysis identified NR1D1 and TFAP4 as putative TF targets of MYCN (**Figs. 7b,c, S10a,b; Table S11**). *TFAP4* inhibition leads to differentiation of *MYCN*-amplified neuroblastoma cells[77,78], and the nuclear receptor *NR1D1* has been shown to correlate with *MYCN* amplification in NB patients[62,63], 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. 7d; Table S12**) of MYCN, TFAP4, and NR1D1 targets in D9_4 (highly expressed in 17q1qMYCN). Conversely, gene sets D9_1 and D9_3 (expressed in WT/17q/17q1q) were enriched for targets of TFs expected at this stage of differentiation, e.g., SOX10/4, TFAP2A/B, and nuclear receptor NR2F1. The expression of targets of these TFs increased or decreased along the mutational spectrum, corroborating the association of these TFs with the mutations (**Fig. 7e**). While most 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 anterior axial marker HOXB1[79,80] and EGR3 (up), or of sensory neurogenesis regulator NEUROD1[46] (down). To aid interpretation, we visualised cell-line-specific interactions between TFs and targets as edges in connected network diagrams (**Figs. 7f, S10c**). These diagrams showcased the emergence of a new subnetwork of TFs in 17q1qMYCN that centred on MYCN and incorporated TFs like NR1D1, TFAP4, and EGR3. In contrast, a subnetwork involving NC-related TFs such as SOX10, SOX4, and TFAP2B was lost in these cells. Intriguingly, downregulation of TFs linked to sensory neuronal development (NEUROD1, ONECUT1) was visible already in 17q cells (**Fig. 7f**). In 17q1q, we additionally observed upregulation of TFs including HOXB1, EGR3, and FOS (**Fig. 7f**).

In summary, our data suggest a subtle rewiring of gene-regulatory networks in CNV-carrying hESCs, which may be linked to the depletion of 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, including NR1D1 and TFAP4.

Discussion

Although CNVs 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 CNVs that are prevalent in NB (chr17q and chr1q gains). Because the NC is a transient embryonic tissue that is inaccessible after birth, hESC differentiation allowed us to experimentally study for the first time 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 CNVs 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, ectopic induction of endothelial cells, and a proliferative, tumour-like cellular phenotype. We also found that acquisition of NB-linked CNVs correlated with upregulation of anti-apoptotic and DNA-damage-repair-related genes and speculate that this may provide an early selective advantage facilitating subsequent oncogenic transformation triggered by *MYCN* overexpression, which in this case acts as a 'second hit' on top of CNVs. This is consistent with previous studies showing that *MYCN* overexpression alone is associated with increased apoptosis in early sympathoadrenal cells[81,82] and triggers tumourigenesis only in combination with additional mutations[13,14,83]. Considering that NB-associated CNVs regularly arise *in vitro*, our findings once again highlight the importance of rigorous monitoring of hESC cultures prior to their use in disease modelling or cell therapy[55].

The accumulation of NB-associated lesions correlated with a failure to reprogramme chromatin during trunk NC differentiation. Upon gain of CNVs, cells lost TFs associated with sensory differentiation (e.g., *NEUROD1*) and instead upregulated TFs associated with pro-gliogenic and Schwann cell fates such as *HOXB1*[79,80] and *EGR3*[84]. *MYCN* overexpression on top of CNVs abolished chromatin states of sympathoadrenal differentiation, and instead led to the induction of targets of NR1D1, TFAP4, and other TFs of the reported NB regulatory circruitry[48,49,67–71]. TFAP4 is a well-established downstream effector of MYCN[77,78], and NR1D1 (Rev-erbα) is a circadian rhythm and metabolic regulator, and a downstream effector of MYCN hyperactivity in NB[62,63]. Our model will facilitate the functional dissection of these TFs via loss-/gain-of-function approaches to decipher their crosstalk with *MYCN*/CNV-driven tumourigenesis.

Complementing earlier studies using cell lines and animal models[12–14,18,19,22], recent single-cell transcriptomic analyses of NB tumours[15–17] 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[85]. Furthermore, *MYCN* overexpression (in conjunction with CNVs) 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.

Our hESC-based model provides a tractable system for analysing tumour initiation events within disease-relevant human embryonic cell-like populations. However, 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[86–89]. 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 CNVs 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.

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, and Bettina Brunner-Herglotz (CCRI) for her technical assistance. We are also grateful to Igor Adameyko and Polina Kameneva (Medical University of Vienna) and Sofie Mohlin (Lund University) for critical reading of the manuscript. Funding: A.T.: Biotechnology and Biological Sciences Research Council (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.), Medical Research Council (MRC; MR/V002163/1). F.H.: Alex's Lemonade Stand Foundation for Childhood Cancer (ALSF; 20-17258), Austrian Science Fund (FWF; TAI 454, TAI 732). E.M.P.: FWF (P32001-B, P34832-B). M.C.B.: Austrian Academy of Sciences (25905). M.D.: Austrian Research Promotion Agency (FFG) project 7940628 533 (Danio4Can), ALSF (20-17258). I.B.: UK Regenerative Medicine Platform (MR/R015724/1), MRC (MR/X000028/1). M.F.: ALSF (20-17258). S.T.M.: Vienna Science and Technology Fund (WWTF; project LS18-111), FWF (P35841-B MAPMET), H2020 (project 826494 PRIMAGE). I.F.: FWF (P35072).

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., I.F., S.T.M., M.F., A.T., F.H.; Formal Analysis: L.M.G., I.S.G., C.H., C.St., M.S., F.H.; Funding acquisition: P.A., H.B., I.B., I.F., M.D., S.T.M., M.F., A.T., F.H.; Investigation: I.S.G., L.M.G., D.S., L.S., I.F., A.W.W., C.St., C.So., S.T., M.B., P.B., M.G., E.B., M.C.B.; Methodology: I.S.G., L.M.G., L.S., I.F., M.B., P.B., M.G., H.B., M.D., S.T.M., M.F., A.T., F.H.; Project administration: A.T., F.H.; Resources: H.B., M.D., S.T.M, M.F., A.T., F.H., S.T.M.; Software: L.M.G., C.H.; Supervision: E.P., H.B., M.D., S.T.M., M.F., A.T., F.H.; Visualization: I.S.G., L.M.G., C.St., F.H.; Writing – original draft: I.S.G., L.M.G., A.T., F.H.; Writing – review & editing: I.S.G., L.M.G., C.H., L.S., I.F., C.St., C.So, M.B., P.B., M.G., M.C.B., E.P., P.A., I.B., H.B., M.D., S.T.M., M.F., A.T., F.H.

Declaration of interests

The authors declare no competing interests.

Methods

Human embryonic stem cell (hESC) cell culture and differentiation

Cell lines and cell culture

We employed H7 hESCs as a karyotypically normal, female WT control[36]. Use of hES cells has been approved by the Human Embryonic Stem Cell UK Steering Committee (SCSC15-23). Their isogenic chr17q counterparts carry a gain in chromosome 17q (region q27q11) via an unbalanced translocation with chromosome 6[56,99]. The chr17q1q hESC line was clonally derived, after its spontaneous emergence following the genetic modification of chr17q hESCs. The 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[100] via nucleofection using the Lonza 4D-Nucleofector System as per the manufacturer's instructions (Amaxa 4D-Nucleofector Basic Protocol for Human Stem Cells). All cell lines were tested regularly for mycoplasma and expression of pluripotency markers. hESCs were cultured routinely in feeder-free conditions at 37°C and 5% CO_2 in E8 media[101] complemented with GlutaMax (Cat# 35050061, Thermo Fisher Scientific) on Vitronectin (Cat# A14700, 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[33,34]. 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). 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. NMPs at D3 were then re-plated at 50,000 cells/cm^2 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). 48 hours later (D5), media was replaced removing the Rock Inhibitor. Media was refreshed at D7 and D8. On D5, the expression of MYCN was induced by supplementing the neural crest media with 100ng/ml of Doxycycline (Cat# D3447, Merck). On D9, cells were re-plated at 100,000 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). 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).

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). 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). 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 the BD FACSJazz 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 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). 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). 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.

Time-lapse imaging

Time-lapse images of differentiating 17q1q and 17q1q-MYCN trunk NC cultures were taken every hour, from 0 to 84 hours. Imaging started 24 hours after plating using the BioStation CT system (Nikon).

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 doxycycline (100ng/ml) to induce *MYCN* expression. On D8, media were refreshed, and respective doxycycline 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[66] (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[102]. 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).

Whole-exome 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 and Twist Universal Adapter System (Twist). For whole-exome sequencing, the libraries were pooled and enriched with the Exome v1.3 and RefSeq spike-in capture probes (Twist) according to the manufacturer's protocols. Libraries were quality-checked on a 2100 Bioanalyzer automated electrophoresis instrument (Agilent) and diluted before sequencing by the Biomedical Sequencing Facility at CeMM on an Illumina NovaSeq SP flowcell in 2x100bp paired-end mode. Raw reads were mapped to the human reference genome (GRCh38) with *BWA-MEM*[103] (v0.7.17-r1188) before SNP and INDEL discovery and genotyping following GATK Best Practices [104,105](v4.2.0.0). Copy number analysis was done using the *CNVkit* (v0.9.1) batch pipeline[98].

Single-cell RNA sequencing

Library generation and sequencing

Single-cell suspensions were barcoded using oligo-conjugated lipids following the MULTI-seq workflow and frozen live[106]. Frozen, barcoded samples were thawed and 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 three 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[106]. After quality control, libraries were sequenced on the Illumina NovaSeq S4 platform in 2x150bp paired-end mode. **Table S1** includes an overview of sequencing data and performance metrics.

Raw data processing and alignment

Raw sequencing data were processed with the *CellRanger* v5.01 software (10x Genomics) for cell-level demultiplexing and alignment to the human reference transcriptome (*refdata-gex-GRCh38-2020-A* assembly provided by 10x Genomics; parameters: *--expect-cells=15000 --r1-length 28*). Following initial data processing, all subsequent analyses were performed in R (v4.0.3) using Bioconductor packages and the *Seurat*[93,107,108] (v4.1.1) 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*[109] (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.

Quality control

We assessed quality of cells via two complementary pipelines. We first assessed technical covariates and characteristic expression profiles separately per dataset. Here, we kept cells with less than 15% mitochondrial UMI counts, and at least 500 detected genes and applied basic scRNA-seq processing and clustering of the cells (*SCTransform*[109] v0.3.3, parameters: *method="glmGamPoi"*; *RunPCA* keeping 30 dimensions, clustering with default parameters). We used clusters devoid of markers or characterised by markedly higher mitochondrial expression, to derive a library-specific UMI count threshold to further remove low-quality or empty cells (thresholds: $\log_{10}$(nCount_RNA) ≥ G1_GEX: 3.5, G3_GEX: 4.0, G4_GEX: 3.8, G5_GEX: 3.5, G6_GEX: 4.0, G7_GEX: 4, G8_GEX: 3.9, G9_GEX: 3.8, G10_GEX: 3.6, G11_GEX: 3.6, G12_GEX: 3.7, G13_GEX: 3.9). After combining all datasets, we performed am additional technical assessment globally. Here, empty and doublet droplets were flagged with *Emptydrops*[110] (v1.17.3; default parameters) and *scDblFinder*[111] (v1.4.0; parameters: *dbr=0.1*), respectively. We retained only cells with *Emptydrops* FDR>0.01, doublet score smaller than 0.1, and a mitochondrial gene content less than or equal to 10%.

Sample demultiplexing

To demultiplex cells belonging to different pooled samples, we counted MULTI-seq barcodes[106] and used two complementary methods: the *MULTIseqDemux* function from *Seurat* (parameters: *autoThresh=TRUE, maxiter=5*) and a custom Gaussian finite mixture model (GMM) that identified per pair of barcodes four groups of cells: G1) positive for barcode 1, G2) positive for barcode 2, G3) negative for both barcodes, and G4) positive for both barcodes (*cb_demux_gmm* function, *canceRbits* v0.1.6). Briefly, for each cell and barcode pair, we calculated the mean and relative differences (difference over the mean) of $\log_{10}$-transformed counts (pseudo-count 10 added). In the first iteration, we used a 1-dimensional mixture model (*mclust* package[112] v5.4.9; *MClustSSC* method; parameters: *modelNames="E", G = 3*) with relative differences as input and the following training data: the 50 cells with highest/lowest relative difference as positives for G1 and G2, 50 cells closest to the mean of

G1 and G2 (i.e., undecided between both groups) as G3. Based on the resulting classification, we generated synthetic G4 training data by sampling cells assigned to G1 and G2 and combining their barcode counts (barcode 1 from G1 cells, barcode 2 from G2 cells). These synthetic doublet cells were added to the training data and a final 2-dimensional mixture model (parameters: *modelNames="VVV", G = 4*) with a relative difference and mean as input was used for classification. Only cells unequivocally assigned to one sample by both methods were retained. Only cells classified as singlets by both approaches were retained.

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

Raw UMI counts were normalised using Seurat's *SCTransform*[109] (parameters: *variable.genes.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 three independent differentiation experiments (replicates; **Table S1**), datasets were integrated using *Harmony*[113] (v1.0; parameters: *dims=1:30, group.by.vars = "replicate"*). Nearest neighbours were identified using Seurat's *FindNeighbors* function (parameters: *k=70*) on principal components and a Uniform Manifold Approximation and Projection (UMAP) was calculated using Seurat's *RunUMAP* function (parameters: *n.neighbors=70, min.dist=0.5, dims=1:30*). For the wild-type UMAP (cp. **Fig. 1**), the dataset was limited to only wild-type cells, where neighbours were recalculated, and the first 10 principal components were considered sufficient to capture relevant cell types and transitions among wild-type cells. Clusters were defined using Seurat's *FindClusters* (parameters [full dataset]: *resolution=6*, parameters [WT-only]: *resolution=0.2*). For the wild-type dataset, neighbouring clusters that differed by cell cycle gene expression, but otherwise shared functional markers were merged manually and relabelled to reflect differentiation order. Finally, markers for each cluster were identified using *Seurat*'s *FindMarkers* function (parameters: *method="wilcox", min.logfc=0, min.cells=0.1*), with each cluster compared to all the other cells in the dataset. Genes with an adjusted P-value less than 0.05 and an average $\log_2$ fold change greater than 1 were considered cluster markers after masking all ribosomal and mitochondrial genes (**Table S2**). To compare mutant and wild-type cells, we filtered the integrated dataset to cells from D9 and identified pairwise DEGs ($P_{adj} \leq 0.005$, $|\log_2\text{FoldChange}| > 0.25$) between each mutant condition and WT using the *FindMarkers* function (parameters: *test.use="wilcox", only.pos=FALSE, logfc.threshold=0*). Up- and down-regulated DEGs on chr1q, on chr17q, and outside either CNV were then tested separately to identify significant overlaps with MSigDB HALLMARK[91] gene sets using the hypergeometric test implemented in the *hypeR*[92] package (v1.10.0). DEGs and enriched pathways are listed in **Tables S4** and **S5**. To obtain markers for the 167 full integrated dataset (WT and mutant) clusters (**Fig. S5b**), we used a more sensitive approach suitable for fewer cells per cluster. Briefly, the cells on each side of the comparison were randomly binned into three pseudo-replicates and the counts were summed to create pseudo-bulk data. These counts were then used as input for *edgeR*[114] (v3.32.1, test type = QLF, default parameters). After processing each cluster, we removed genes with negative fold change, and calculated FDR values. We sorted the results by p-value and discarded genes with FDR > 0.05. The same approach was used to obtain markers for mega-clusters. All non-ribosomal/non-mitochondrial DEGs are reported in **Table S6**.

Pseudotime trajectory analysis

Pseudotime trajectories were inferred using *Slingshot*[90] (v1.8.0; default parameters). In each trajectory, we filtered the dataset to adjacent cell clusters where apparent continuities were observed between related cell types (cp. **Figs. 1e,i, S3**). The filtered datasets were then 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. If more than one trajectory was found, the longest

trajectory (spanning the most cells) was selected. Genes whose expression was associated with the trajectories were identified with the generalised additive model and association test as implemented in *tradeSeq*[97] (v1.4.0; parameters: *knots=3* for MES-SYM and for SCP_SYM (D9), and *knots=6 for MES-SCP-SYM (D14)*., The top genes with the highest Wald statistic were selected for reporting (**Table S3**), with different number of genes are shown for each trajectory maximise the number of legible receptors, ligands, and TFs (based on the human transcription factors database[115] and *CellTalkDB*[116]) in **Fig. 1** and **Fig. S3.**

Cross-dataset annotation, label transfer, and signature scores

To map data between scRNA-seq datasets, we used label transfer with *Azimuth* and *AzimuthReferences* packages[93] using a Dockerfile provided by the developers (*satijalab/azimuth:0.4.2*). Both query and reference datasets were processed using the default basic scRNA-seq processing workflow as described above and subjected to the Azimuth mapping workflow (*FindTransferAnchors*, *TransferData*, *IntegrateEmbeddings*, *NNTransform*, and *MappingScore* functions; default parameters), using the 50 first principal components 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. The following mappings were performed in this study:

1. Human foetal adrenal reference datasets[15,16] onto WT-only (**Figs. 1e, S2h**) and full *in vitro* (**Figs. 2d,e, S7**) scRNA-seq references. Upon obtaining consistent results for both (**Fig. S2i**), the reference provided by Kameneva *et al.* was used throughout the analysis, because of the curated cell type markers they provided (**Fig. S2h**). These gene signatures were also quantified with Seurat's *AddModuleScore* function (default parameters) in **Figs. 1f,g, 2f**.
2. Our mutant scRNA-seq data onto the wild-type reference (**Figs. 2b,c**).
3. NB tumour data onto our integrated reference (**Figs. 5c-g, S8**) and to the human foetal adrenal reference[16]. See additional details about these datasets and processing in the section *"Pre-processing and mapping of NB tumour data"* below.

Mutation score analysis

To calculate the mutation score, 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 = 70) in the 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 Spearman correlations 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 CNVs; 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 expressed in phenotypes achieved by two or more mutants. All non-duplicate absolute correlations (calculated using R's *cor.test*, parameters: *method="spearman", exact=TRUE*) were subject to Bonferroni correction and ranked. The top-100 correlated genes ($p \leq 0.05$) per differentiation stage (D9, D14, D19) are reported in **Table S3**.

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],
- four samples (3 primary adrenal, 1 relapse/occipital subcutaneous bone metastasis [Jansky_NB14]; 1 female [Jansky_NB08], 3 male [Jansky_NB01, Jansky_NB11, Jansky_NB14]; accession EGAS00001004388 [European Genome-Phenome Archive])[15],
- and four samples (all metastatic bone marrow; 2 female [Fetahu_M1, Fetahu_M3], 2 male [Fetahu_M2, Fetahu_M4]; Fetahu, Esser-Skala, *et al.*, *in revision*).

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 40% were kept for further analysis, except for dataset *Jansky_NB08* where a filter of 100 reads was used to prevent loss of all cells. To focus on *MYCN*-amplified tumour cells, we selected based on the following gene expression profile (read count > 0): *MYCN*$^{+}$/*CD45*$^{-}$/*CD34*$^{-}$/*KDR*$^{-}$/*AHSG*$^{-}$/*STAR*$^{-}$/*NR5A1*$^{-}$/*CYP17A1*$^{-}$/*PAX2*$^{-}$/*LYPD1*$^{-}$/*HBA2*$^{-}$. These markers selected mostly cells with strong CNV profiles (see below) at key genomic positions such as chr2p (*MYCN* locus). Retained cells were subjected to scRNA-seq processing as described above. Cells were then subjected to default basic scRNA-seq processing (see above) with slightly different parameters for *SCTransform* (*variable.features.n=6000*). When mapping tumour cells to our full integrated *in vitro* dataset's clusters, we additionally filtered cells with a prediction score greater than or equal to 0.6 (60% or more of each cells neighbours belong to the same cluster) to ensure high-confidence mapping, and thus we favoured specificity over sensitivity.

Inference of CNV profiles from scRNA-seq data

To infer tumour cell CNV profiles from scRNA-seq expression data, we used the *infercnv*[94] 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 previous mapping. 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.

Neuroblastoma marker specificity

To quantify how specific the expression of NB marker genes was for specific clusters, we added up the percentage of cells expressing the marker gene per cluster. If most cells in all clusters express a gene, this would yield a high sum, implying low specificity. Therefore, the inverse of *T* was used as a specificity score.

RNA sequencing

Library generation and sequencing

The amount of total RNA was quantified using the Qubit 4.0 Fluorometric Quantitation system (Thermo Fisher Scientific) and the RNA integrity number (RIN) was determined using the 2100 Bioanalyzer instrument (Agilent). RNA-seq libraries were prepared with the QuantSeq 3'mRNA-Seq Library Prep Kit (FWD) for Illumina (Lexogen). Library concentrations were quantified with the Qubit 4.0 Fluorometric Quantitation system (Life Technologies) and the size distribution was assessed using the 2100 Bioanalyzer instrument (Agilent). For sequencing, samples were diluted and pooled into libraries in equimolar amounts. Libraries were sequenced by the Biomedical Sequencing Facility at CeMM using

the Illumina HiSeq 4000 platform in 1x50bp single-end mode. **Table S1** includes an overview of the sequencing data and performance metrics.

Raw data processing, alignment, and quality control

Quant-seq adapter fragments were trimmed using *bbduk* (from *bbmap* v38.87; parameters: *k=13 ktrim=r useshortkmers=t mink=5 qtrim=r trimq=10 minlength=20*) prior to alignment to the human reference genome (*refdata-gex-GRCh38-2020-A* assembly provided by 10x Genomics for maximum compatibility with scRNA-seq analyses) using *STAR*[117] v2.7.3a (parameters: *--outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMmapqUnique 60 --outSAMunmapped Within*). Read counts per gene were calculated using *mmquant*[118] v1.3 (parameters: *-l 1 -D 10 -d 0.5 -s F -e Y -p -t 1*) and used for ATAC-seq peak to gene assignments (see below).

Chromatin accessibility mapping

Library generation and sequencing

ATAC-seq was performed as described previously[72]. 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. **Table S1** includes an overview of the sequencing data and performance metrics.

Raw data processing, alignment, and quality control

Raw sequencing data were processed using *PEPATAC*[119] (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 (v4.1.2) using Bioconductor packages. 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 (**Table S8**). Next, we quantified for each input dataset the number of reads overlapping these consensus peaks using *featureCounts*[120] (*Rsubread* v2.8.1).

Differential accessibility analysis and chromatin modules

Raw read counts were loaded into DESeq2[95] (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;* **Table S9**). To define chromatin regulatory modules, we focused on time-wise differences in WT hESCs (n = 30,749 regions), which we subdivided into six chromatin modules (R1-R6) by hierarchical clustering with cosine distance using the Ward criterion (parameter: *method = "ward.D2"*). To associate ATAC-seq regions with putative target genes, we used the *GenomicRanges*[121] 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 matching samples of our RNA-seq data (mean per stage and condition). We calculated an empirical P-value by shuffling RNA/ATAC assignments (10 repetitions) and retained associations with a P-value ≤ 0.05. Annotated regulatory regions from the analysis of ATAC-seq data are listed in **Table S8**.

Overlap enrichment analysis for chromatin modules

To characterize the chromatin modules, we interrogated overlaps with genomic regions or associated genes using the hypergeometric test implemented in the *hypeR*[92] package (v1.10.0). We looked at three types of overlaps: (a) Annotated reference regions from the DNase hypersensitivity index[73], from the Cis-element Atlas[74], from the Enhancer Atlas[75], and NB subgroup-specific super-enhancers[68], which all catalogue regulatory elements active in different cell or tissue types. (b) Matches to known TF motifs from the *HOCOMOCO* database[96] (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 (**Table S2**). 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 (for motifs: $P_{adj} \leq 0.0000001$), an absolute $\log_2$ odds greater than $\log_2(1.5)$ (for motifs: $\log_2(2)$), and a total frequency of at least 2.5% (i.e., a hit was found in at least 2.5% of all regions in the query module) as significant. All enrichment results are reported in **Table S10**.

Identification of transcription factor targets

To identify putative target genes of TFs, we used *GRNboost2*[76] (*arboreto* library v0.1.6, with Python v3.8) to identify genes whose expression could be predicted from the expression of each TF. We tested all TFs in the *HOCOMOCO* database[96] 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 eight 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 cells and to identify significant overlaps between target genes and gene sets D9_1 – D9_4 (**Table S7**), we used only activated targets ($r>0.1$) and the Seurat module score and *hypeR*[92] package (v1.10.0; selected TFs: $P_{adj} \leq 0.0005$, $|\log_2 \text{odds}| \geq \log_2(2)$, frequency $\geq 10\%$), respectively. All target gene sets are reported in **Table S11** and all enrichment results in **Table S12**.

Gene-regulatory network visualisation

For the visualisation of gene-regulatory networks, we used the *igraph* package (v1.2.9). A directed graph was constructed from edges between genes in the gene sets D9_1, D9_2, D9_3, or D9_4 (**Table S7**) and TFs found enriched in the overlap with these genes (**Fig. 7f**). The same automated graph layout (function *layout_nicely()*) was used to draw mutant-specific network diagrams. To generate mutant-specific networks (**Fig. 7h**), we selected cells of 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 **Fig. S10c**.

Data availability

Raw and processed single-cell RNA-seq, RNA-seq, and ATAC-seq data 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[16] and GSE137804[17] (Gene Expression Omnibus), and EGAS00001004388[15] (European Genome-Phenome Archive).

Code availability

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

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. *Nat Rev Cancer* **5**, 481–8 (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. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279–284 (2013).
24. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N Engl J Med* **340**, 1954–1961 (1999).
25. Gilbert, F. *et al.* Human neuroblastomas and abnormalities of chromosomes 1 and 17. *Cancer Res* **44**, 5444–9 (1984).
26. Lastowska, M. *et al.* Comprehensive genetic and histopathologic study reveals three types of neuroblastoma tumors. *J Clin Oncol* **19**, 3080–3090 (2001).
27. 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).
28. Bogen, D. *et al.* The genetic tumor background is an important determinant for heterogeneous MYCN-amplified neuroblastoma. *Int J Cancer* **139**, 153–163 (2016).
29. O'Neill, S. *et al.* MYCN amplification and 17q in neuroblastoma: evidence for structural association. *Genes Chromosomes Cancer* **30**, 87–90 (2001).
30. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb Perspect Med* **3**, (2013).
31. Schleiermacher, G. *et al.* Accumulation of segmental alterations determines progression in neuroblastoma. *J Clin Oncol* **28**, 3122–3130 (2010).
32. Schleiermacher, G. *et al.* Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *Br J Cancer* **107**, 1418 (2012).
33. Frith, T. J. R. *et al.* Human axial progenitors generate trunk neural crest cells in vitro. *Elife* **7**, (2018).

34. 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).
35. Wymeersch, F. J., Wilson, V. & Tsakiridis, A. Understanding axial progenitor biology in vivo and in vitro. *Development* **148**, (2021).
36. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–7 (1998).
37. Gogolou, A. *et al.* Early anteroposterior regionalisation of human neural crest is shaped by a pro-mesodermal factor. *Elife* **11**, 2021.09.24.461516 (2022).
38. 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).
39. Ma, Q., Fode, C., Guillemot, F. & Anderson, D. J. Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia. *Genes Dev* **13**, 1717–1728 (1999).
40. Hirsch, M. R., Tiveron, M. C., Guillemot, F., Brunet, J. F. & Goridis, C. Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. *Development* **125**, 599–608 (1998).
41. Tsarovina, K. *et al.* The Gata3 Transcription Factor Is Required for the Survival of Embryonic and Adult Sympathetic Neurons. *Journal of Neuroscience* **30**, 10833–10843 (2010).
42. Kastriti, M. E. *et al.* Schwann cell precursors represent a neural crest-like state with biased multipotency. *EMBO J* **41**, e108780 (2022).
43. 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).
44. 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).
45. McGonnell, I. M. & Graham, A. Trunk neural crest has skeletogenic potential. *Curr Biol* **12**, 767–771 (2002).
46. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (1979)* **364**, (2019).
47. la Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
48. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* **49**, 1408–1413 (2017).
49. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* **49**, 1261–1266 (2017).
50. van Groningen, T. *et al.* A NOTCH feed-forward loop drives reprogramming from adrenergic to mesenchymal state in neuroblastoma. *Nat Commun* **10**, 1530 (2019).
51. 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).
52. 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).
53. Khalil, B. D. *et al.* An NR2F1-specific agonist suppresses metastasis by inducing cancer cell dormancy. *Journal of Experimental Medicine* **219**, (2021).
54. Tümmler, C. *et al.* Inhibition of chemerin/CMKLR1 axis in neuroblastoma cells reduces clonogenicity and cell viability in vitro and impairs tumor growth in vivo. *Oncotarget* **8**, 95135 (2017).
55. 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).
56. Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* **22**, 53–54 (2004).
57. 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).
58. Handel, A. E. *et al.* Developmental dynamics of the neural crest–mesenchymal axis in creating the thymic microenvironment. *Sci Adv* **8**, 9844 (2022).
59. Szemes, M. *et al.* A Wnt-BMP4 Signaling Axis Induces MSX and NOTCH Proteins and Promotes Growth Suppression and Differentiation in Neuroblastoma. *Cells* **9**, (2020).
60. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res* **30**, 35–48 (2020).
61. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).

62. Moreno-Smith, M. *et al.* Restoration of the molecular clock is tumor suppressive in neuroblastoma. *Nature Communications 2021 12:1* **12**, 1–16 (2021).
63. Altman, B. J. *et al.* MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell Metab* **22**, 1009–1019 (2015).
64. Cho, H. *et al.* Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β. *Nature* **485**, 123–127 (2012).
65. 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).
66. Wrobel, J. K. *et al.* Rapid In Vivo Validation of HDAC Inhibitor-Based Treatments in Neuroblastoma Zebrafish Xenografts. *Pharmaceuticals (Basel)* **13**, 345 (2020).
67. Durbin, A. D. *et al.* EP300 Selectively Controls the Enhancer Landscape of MYCN-Amplified Neuroblastoma. *Cancer Discov* **12**, 730–751 (2022).
68. Gartlgruber, M. *et al.* Super enhancers define regulatory subtypes and cell identity in neuroblastoma. *Nat Cancer* **2**, 114–128 (2021).
69. Zeid, R. *et al.* Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat Genet* **50**, 515–523 (2018).
70. Jahangiri, L. *et al.* Core regulatory circuitries in defining cancer cell identity across the malignant spectrum. *Open Biol* **10**, 200121 (2020).
71. Zimmerman, M. W. *et al.* Retinoic acid rewires the adrenergic core regulatory circuitry of childhood neuroblastoma. *Sci Adv* **7**, (2021).
72. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).
73. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature 2020 584:7820* **584**, 244–251 (2020).
74. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-6001.e19 (2021).
75. 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).
76. Moerman, T. *et al.* GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
77. Xue, C. *et al.* MYCN promotes neuroblastoma malignancy by establishing a regulatory circuit with transcription factor AP4. *Oncotarget* **7**, 54937–54951 (2016).
78. 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).
79. Gouti, M., Briscoe, J. & Gavalas, A. Anterior Hox Genes Interact with Components of the Neural Crest Specification Network to Induce Neural Crest Fates. *Stem Cells* **29**, 858 (2011).
80. Arenkiel, B. R., Tvrdik, P., Gaufo, G. O. & Capecchi, M. R. Hoxb1 functions in both motoneurons and in tissues of the periphery to establish and maintain the proper neuronal circuitry. *Genes Dev* **18**, 1539 (2004).
81. 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).
82. 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).
83. Berry, T. *et al.* The ALK(F1174L) mutation potentiates the oncogenic activity of MYCN in neuroblastoma. *Cancer Cell* **22**, 117–130 (2012).
84. Svaren, J. & Meijer, D. The molecular machinery of myelin gene transcription in Schwann cells. *Glia* **56**, 1541–1551 (2008).
85. 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).
86. Lazic, D. *et al.* Landscape of Bone Marrow Metastasis in Human Neuroblastoma Unraveled by Transcriptomics and Deep Multiplex Imaging. *Cancers (Basel)* **13**, 4311 (2021).
87. Verhoeven, B. M. *et al.* The immune cell atlas of human neuroblastoma. *Cell Rep Med* **3**, 100657 (2022).
88. Costa, A. *et al.* Single-cell transcriptomics reveals shared immunosuppressive landscapes of mouse and human neuroblastoma. *J Immunother Cancer* **10**, (2022).
89. Weiss, T. *et al.* Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal growth factor-like protein 8. *Nat Commun* **12**, (2021).

90. Street, K. *et al.* Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
91. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417 (2015).
92. Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**, 1307–1308 (2020).
93. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
94. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (1979)* **344**, 1396–1401 (2014).
95. 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).
96. 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).
97. van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1–13 (2020).
98. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).
99. 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).
100. 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).
101. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**, 424–429 (2011).
102. Fior, R. *et al.* Single-cell functional and chemosensitive profiling of combinatorial colorectal therapy in zebrafish xenografts. *Proc Natl Acad Sci U S A* **114**, E8234–E8243 (2017).
103. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013) doi:10.48550/arxiv.1303.3997.
104. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–501 (2011).
105. van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-11.10.33 (2013).
106. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods* **16**, 619–626 (2019).
107. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21 (2019).
108. 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).
109. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 576827 (2019).
110. 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).
111. 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).
112. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289–317 (2016).
113. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
114. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
115. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
116. Shao, X. *et al.* CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief Bioinform* **22**, (2021).
117. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
118. Zytnicki, M. mmquant: How to count multi-mapping reads? *BMC Bioinformatics* **18**, 1–6 (2017).
119. Smith, J. P. *et al.* PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genom Bioinform* **3**, lqab101 (2021).
120. 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).
121. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, e1003118 (2013).

Figures

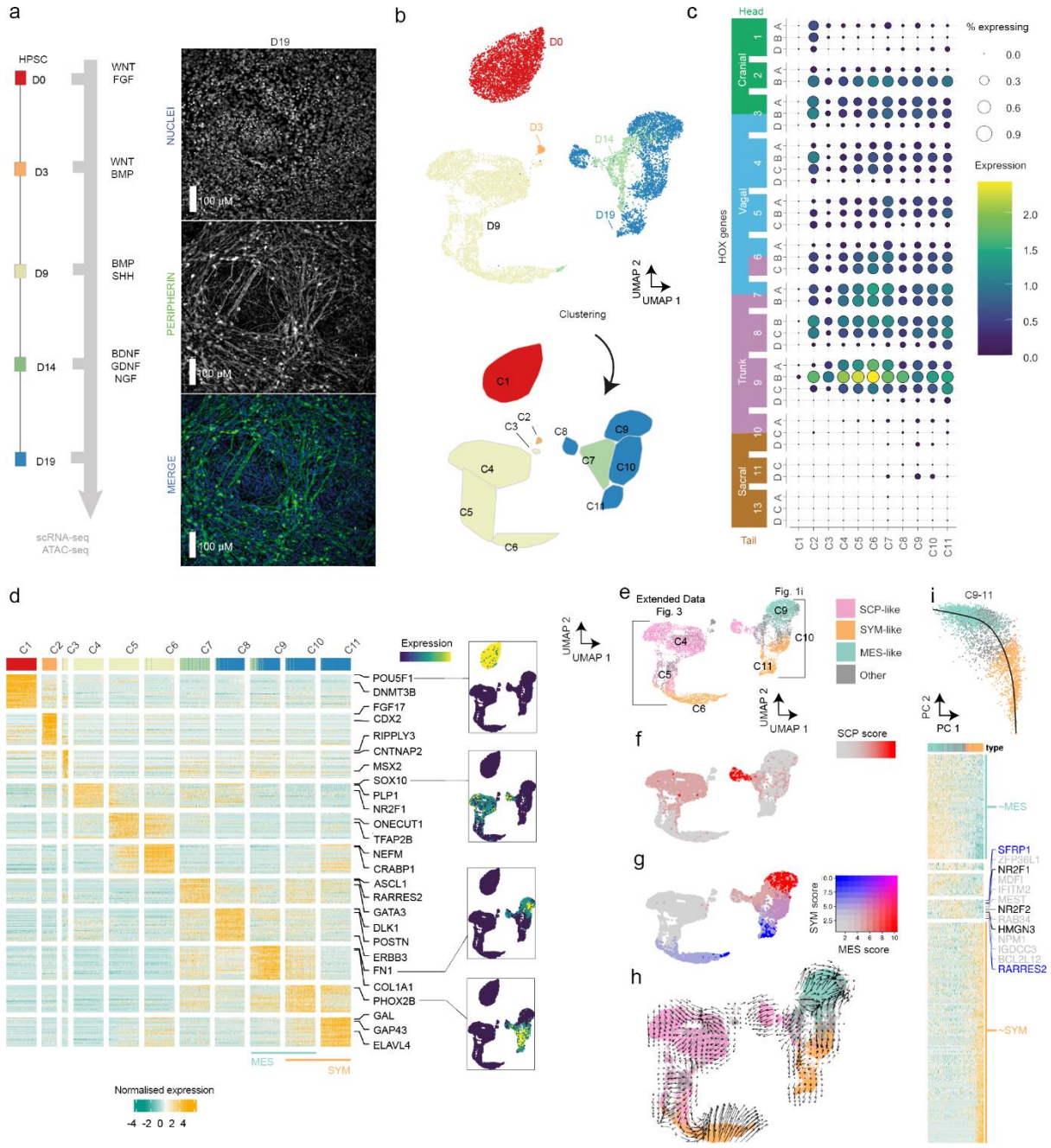


Figure 1. *In vitro* culture efficiently generates human trunk NC cells and their sympathoadrenal derivatives from hESCs.

a) Diagram depicting the extrinsically supplemented signals employed to direct hESCs toward trunk NC cells their downstream derivatives, and immunofluorescence analysis of PRPH 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 during differentiation to trunk neural crest and sympathoadrenal derivatives (top). Cells were divided into 11 distinct clusters as indicated (bottom).

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

d) 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 **Table S2**.

e) Cells from D9-D19 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 (**Fig. S2h**) or could not be assigned to a single type. Pseudotime trajectories for these panels can be found on **panel (i)** and **Fig. S3**.

f) Cells from **panel (e)** coloured by the strength of their SCP marker signature (Seurat module score) in red.

g) Same as above but visualising simultaneously SYM (blue) and MES (red) marker signature. Cells with overlapping marker signatures appear in grey/purple tones.

h) RNA velocities calculated for the cells in **panel (e)** using Velocyto[47].

i) Slingshot[90] pseudotime trajectory (top) for MES- and SYM-like cells in clusters C9-C11, coloured as in **panel (e)**. Cells were reprocessed and trajectory was calculated on the first two principal components. The heatmap (bottom) depicts the top 140 genes associated with the trajectory. Selected genes associated with the intermediate cells are highlighted. All trajectory-associated genes are reported in **Table S3**.

Abbreviations: HPSC, human pluripotent stem cells; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; C1-C11, cell clusters; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal.

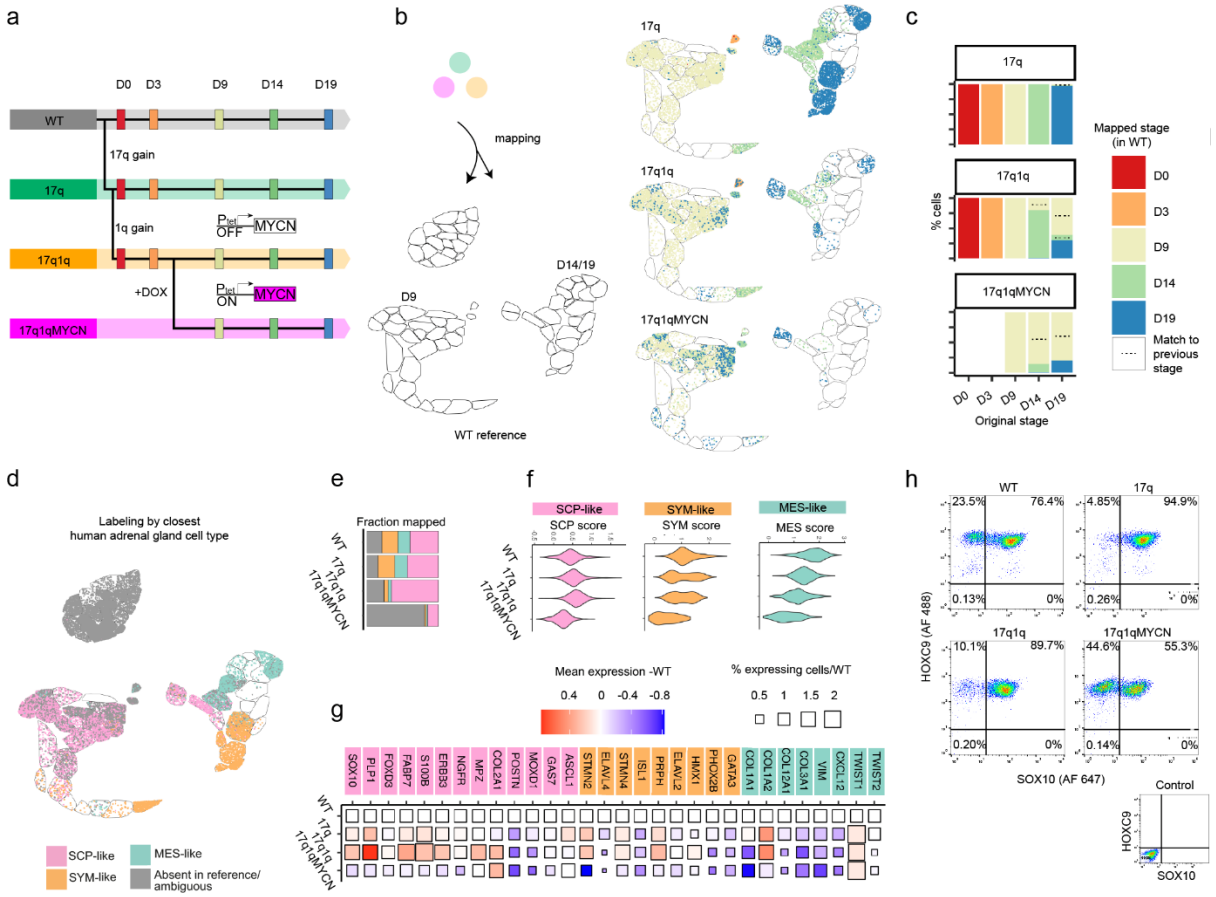


Figure 2. Copy number variants and overexpression of *MYCN* impair the specification of trunk NC derivatives.

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

b) scRNA-seq data from all mutant cells (17q, 17q1q, 17q1qMYCN at all developmental stages) were mapped to the wild-type trunk NC reference (illustration on the left side). Glasswork UMAP plots (right side) 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.

c) Barplots summarising the mappings from **panel (b)** for derivatives of each hESC line (top to bottom). The position on the x-axis indicates the stage at which the cell sample was collected and the colour of the bar the stage to which each cell was mapped. Mismatched mappings to earlier developmental stages are indicated with three dots ("…").

d) Glasswork UMAPs 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 low-confidence mappings.

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 (= 1) are shown for reference.

h) Flow cytometric analysis of the expression of the 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; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal.

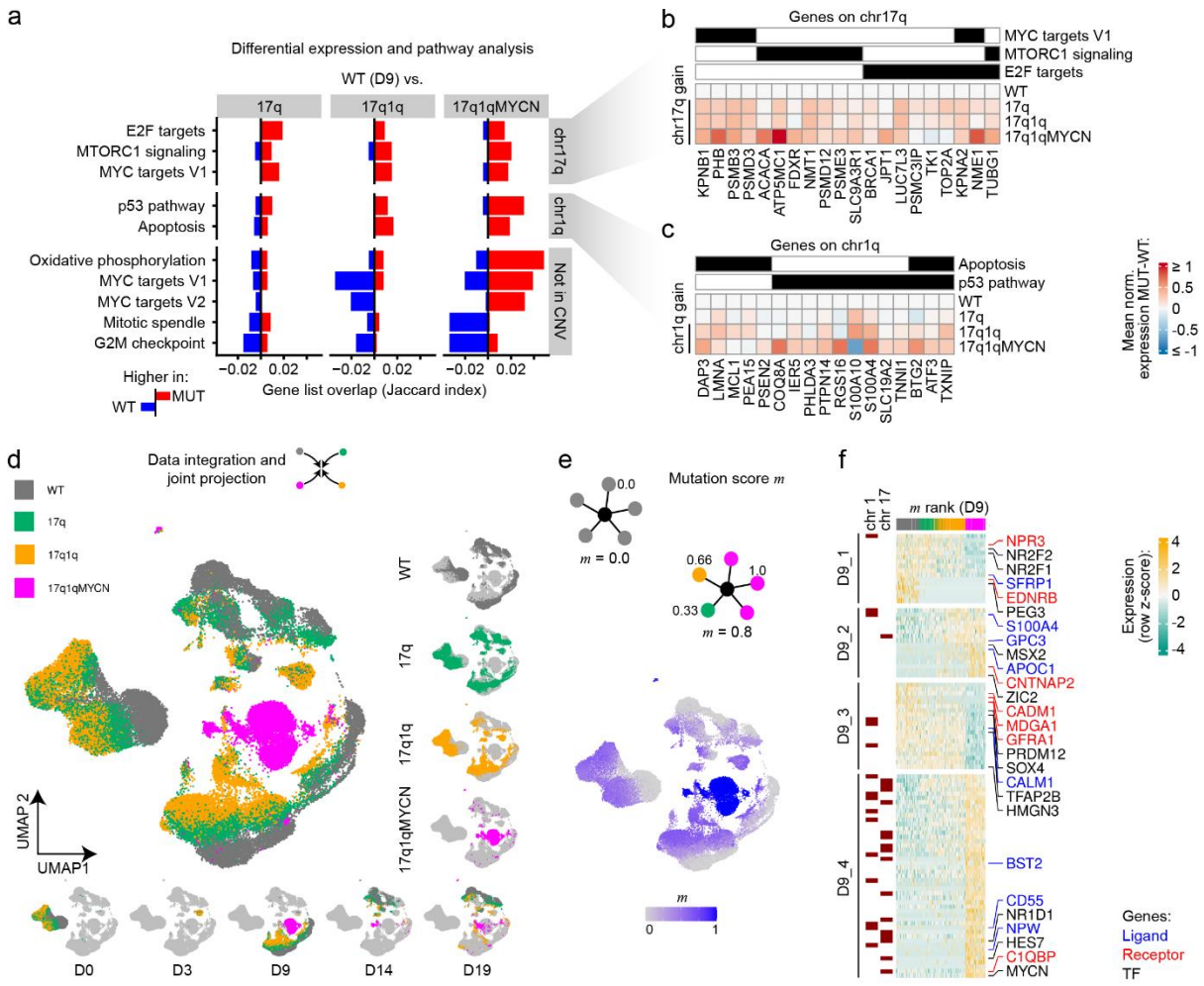


Figure 3. Copy number variants and overexpression of *MYCN* alter the expression of metabolic and developmental pathways.

a) MSigDB hallmark pathways[91] enriched (hypergeometric test, *hypeR*[92]; $P_{adj} \leq 0.05$, $|\log_2 \text{odds}| > \log_2(2)$, at least 6 genes in overlap) in differentially expressed genes (DEGs) in trunk NC cells from each mutant cell line compared to WT at D9 (left to right). The overlap between up- and down-regulated DEGs with the pathway genes is indicated as a positive (red colour bars) or negative (blue 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 CNVs (split from top to bottom). All differentially expressed genes and pathway enrichments are available in **Tables S4 and S5**.

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).

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

d) UMAP of scRNA-seq data of scRNA-seq data from wild-type and mutant hESCs (see **Fig. 2a**) throughout differentiation to trunk neural crest and sympathoadrenal derivatives. Separate UMAPs indicating cells belonging to each of the four cell lines (left) and each of the five developmental stages sampled (bottom) are shown.

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. The actual scores are shown overlaid on the UMAP from **panel (d)** (bottom).

f) Heatmap showing the expression of top 100 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 **Table S7**. 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.

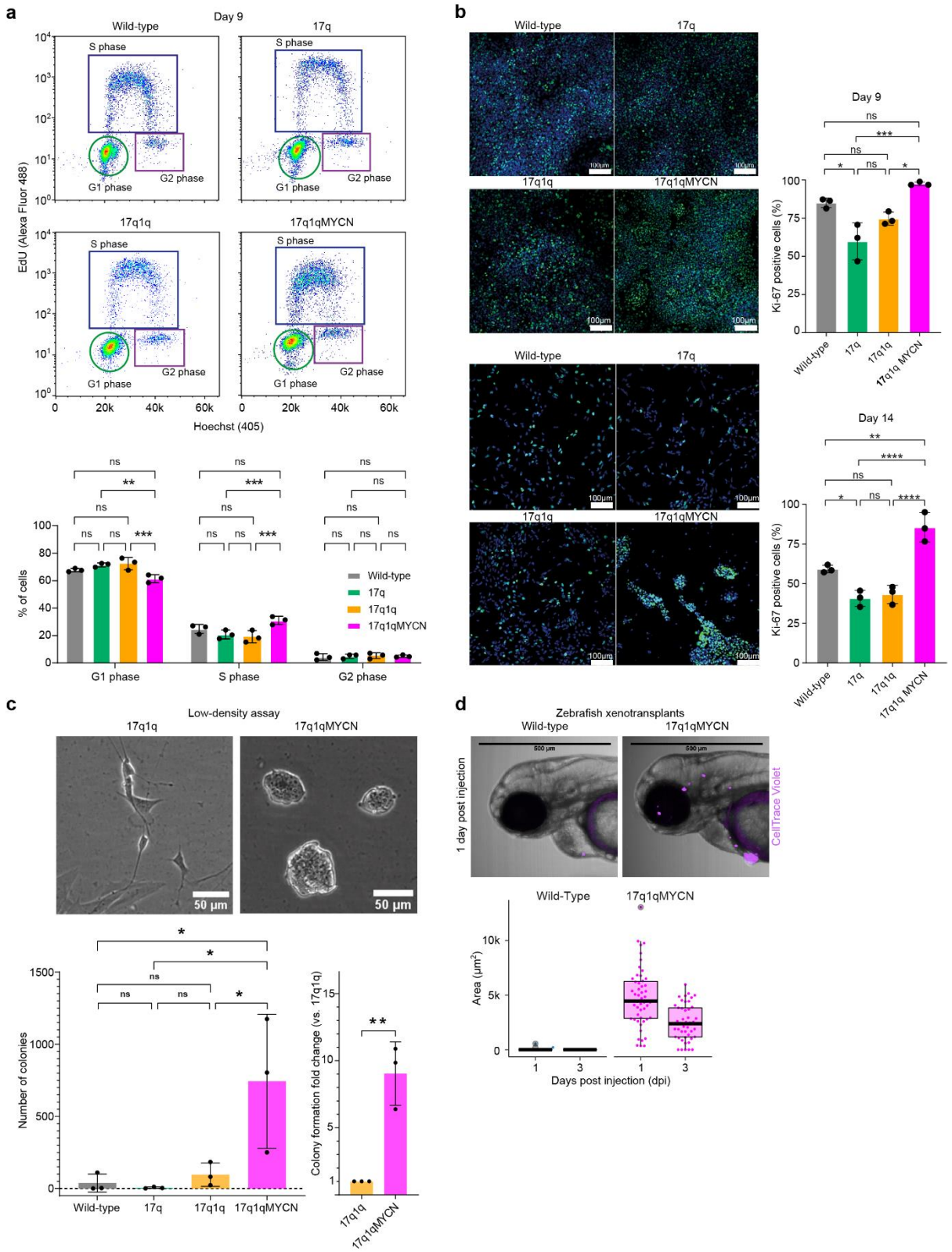


Figure 4. Impaired trunk NC specification correlates with acquisition of tumourigenic hallmarks.

a) 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 17q1qMYCN, p = 0.0012 = **; 17q1q vs

17q1qMYCN, p= 0.0004=***), S (17q vs 17q1qMYCN, p = 0.0008 = ***; 17q1q vs 17q1qMYCN, p= 0.0001=***)

b) Immunofluorescence analysis (green) of the expression of the cell proliferation marker KI-67 in D9 (top) and D14 (bottom) 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 (n = 3 biological replicates, error bars= standard deviation, ordinary one-way ANOVA with Tukey correction). P values in comparisons: D9 (WT vs. 17q p= 0.0072= **; 17q vs 17q1qMYCN p= 0.0005=***; 17q1q vs. 17q1Qmycn p= 0.0125=*), D14 (WT vs 17q p=0.0255=*; WT vs. 17q1qMYCN p= 0.0034=**; 17q vs. 17q1qMYCN p<0.0001=****; 17q1q vs. 17q1qMYCN p= 0.0001=***)

c) Top: 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. Bottom: 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 One-way ANOVA test with Tukey correction). P values in comparisons: WT vs. 17q1qMYCN p= 0.0278=*; 17q vs. 17q1qMYCN p= 0.022=*; 17q1q vs. 17q1qMYCN p= 0.0421=*. The fold change per experiment was calculated for 17q1qMYCN with 17q1q acting as the control. (n= 3 biological replicates, error bars= SD, Two-tailed unpaired t test. P value = 0.0041=**)

d) Representative images of zebrafish xenografted with WT or 17q1qMYCN cells labelled with CellTrace Violet at 1 day post injection (dpi). 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).

Abbreviations: $p \leq 0.05$= *, $p \leq 0.01$= **, $p \leq 0.001$= ***, $p \leq 0.0001$= ****.

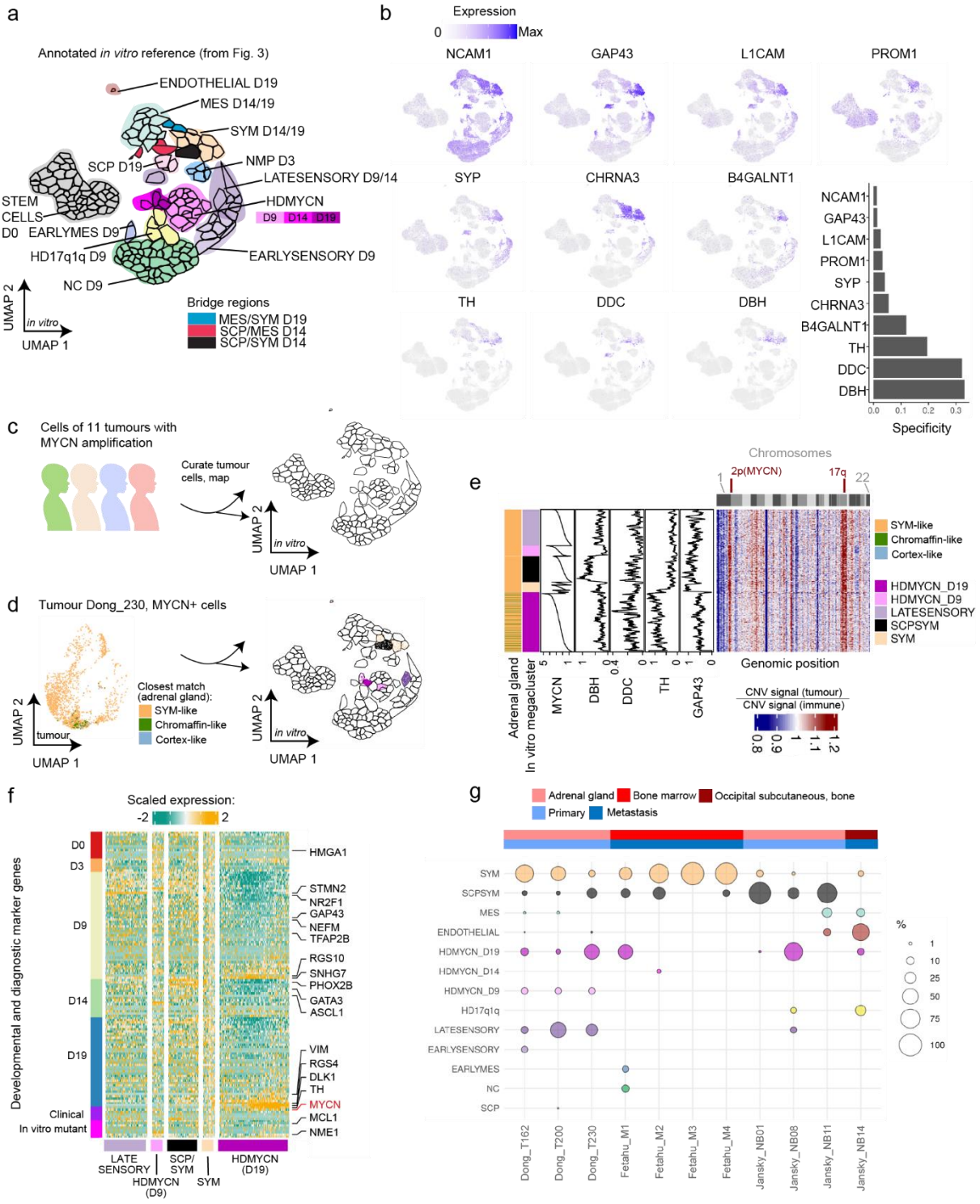


Figure 5. Comparison to hESC-based trunk NC differentiation resolves structured heterogeneity across neuroblastoma tumours

a) UMAP of the integrated trunk NC developmental reference, divided into high-resolution clusters that resolve rare subpopulations. These clusters were grouped into interpretable "mega-clusters" that shared characteristics like developmental stage, condition, and closest-matching human adrenal gland cell type (**Fig. S7**).

b) UMAPs as in **panel (a)** showing expression of widely used diagnostic markers for NB. Markers were ranked from the least specific (appearing in many regions of the UMAP) to the most specific (limited to few clusters) according to a specificity score (bar plot, bottom right). *DBH*, *DDC* and *TH* score as highly specific to the SCP/SYM region of the dataset.

c) Schematic overview of the analysis of tumour cells. We curated *MYCN*+ cancer cells from 11 MYCN-amplified NB tumour samples[15,17] from three studies and mapped them onto our reference (cp. **panel (a)**) using Azimuth[93]. Mapping is represented as tumour cells falling into sectors of the *in vitro* reference (depicted as convex hulls of each cluster in the glasswork plot).

d) Low-dimensionality UMAP projection of the *MYCN*+ cells of tumour *Dong_T230*[17] coloured by their closest matching human adrenal gland cell type, showing continuous groups of SYM-like cells. Mapping to the *in vitro* trunk NC dataset resolves classification into diverse subtypes (see following panels).

e) Heatmap depicting gene expression in *MYCN*+ tumour cells of dataset *Dong_T230* from **panel (d).** Values are inferCNV[94] 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 mega-cluster and Louvain cluster, therein ordered by *MYCN* levels. Annotations (left to right): closest *in-vivo* adrenal gland cell type[16], *in vitro* mega-cluster, and a sliding-window moving average (w=20 cells) of depth-normalised levels of selected diagnostic NB markers. Mappings of other tumours datasets are shown in **Fig. S8**.

f) Heatmap displaying expression of *MYCN*+ tumour cells for selected differentiation markers. Genes were selected to include highly variable markers (ordered by day, D0-D19, from **Fig. 1d, Table S2**), mutant cluster markers ("invitro mutant", **Table S6**), and diagnostic NB markers ("clinical", **panel (b)**). Heatmap columns are *Dong_T230* cells[17] from **panels (c,d)** seriated within mega-clusters.

g) Bubble plots showing the relative percentage (bubble size) of high-confidence mappings (prediction score >=0.6) of MYCN+ tumour cells onto each mega-cluster for cells from 11 tumour datasets, processed and curated as described in **panel (c)**. Tumour sample covariates (tissue of origin (red/brown) and sample type (blue/dark blue)) are depicted in the annotation above.

Abbreviations: WT, wild-type H7 hESCs; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal; M1-M25, "mutant" cell clusters.

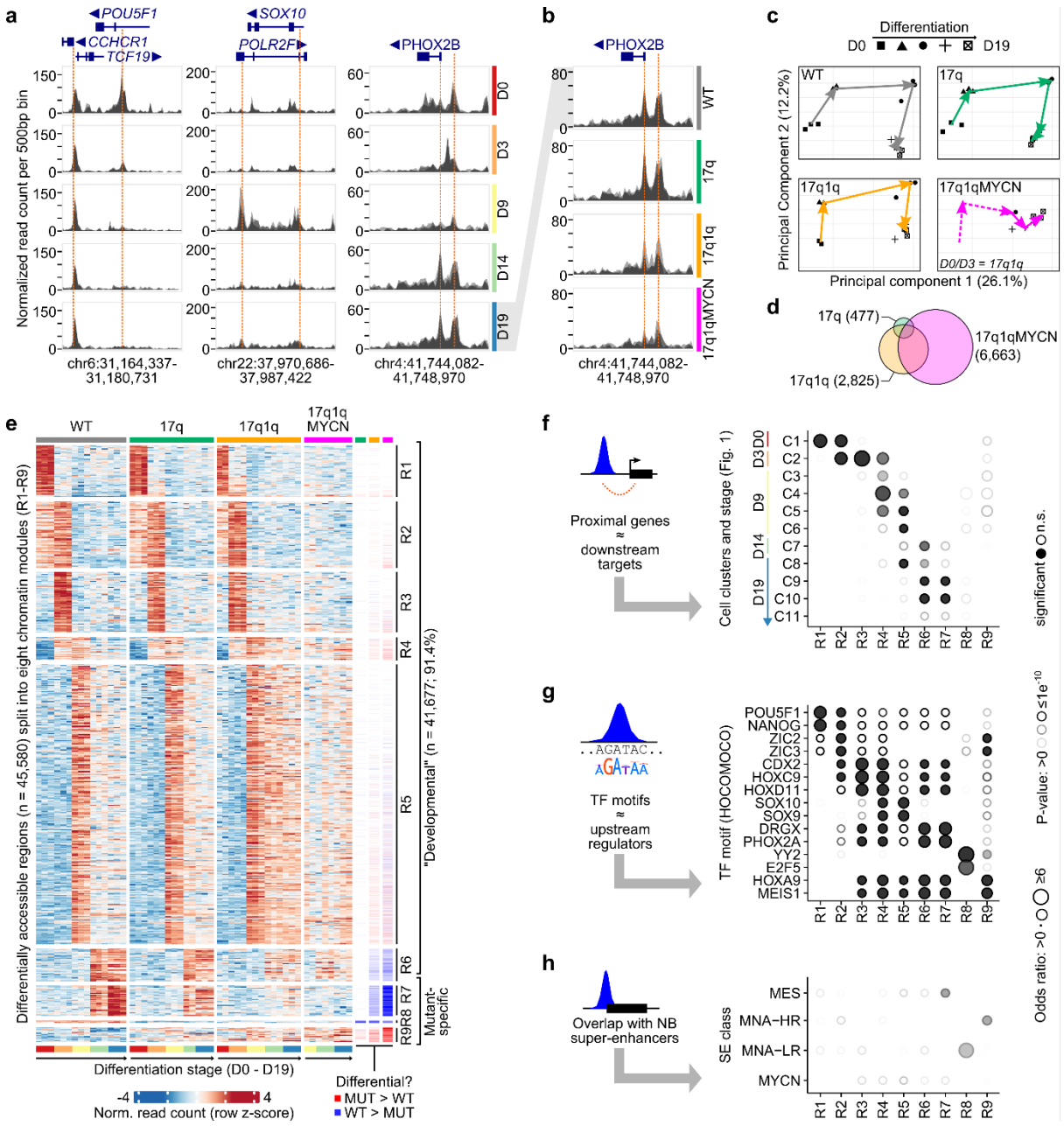


Figure 6. Differentiation of wild-type and mutant hESCs is associated with epigenetic changes in nine distinct chromatin modules.

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. Selected peaks have been highlighted manually.

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 visualizing the overlap of differentially accessible regions (*DEseq2*[95]; $P_{adj} \leq 0.005$, $|\log_2 \text{FoldChange}| \geq \log_2(1.5)$) in mutant hESCs compared to WT-hESCs. 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*[95]; $P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq \log_2(1.5)$; $n_{total} = 45{,}580$). 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 **Tables S8 and S9**.

f) Enrichment analysis of co-localisation of regions belonging to the nine chromatin modules (from left to right; cp. **panel (e)**) and nearby genes identified as markers of differentiating cell populations in our scRNA-seq analysis (cp. **Fig. 1b**). The size and transparency of circles indicates the odds ratio and P-value, respectively (hypergeometric test, *hypeR*[92]). Significant results are indicated with filled circles ($P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq \log_2(1.5)$, frequency $\geq$ 2.5%). All results are shown in the figure and also reported in **Table S10**.

g) Enrichment analysis for overlaps between chromatin modules and known TF motifs (*HOCOMOCO* database[96], v11). The plots are as in **panel (f)**, with the exception that only overlaps with $P_{adj} \leq 0.0000001$, $|\log_2 FoldChange| \geq \log_2(2)$, and frequency $\geq$ 2.5% were marked as significant. The top results per module are shown and all results are reported in **Table S10**.

h) Enrichment analysis of overlaps between regions belonging to the nine chromatin modules and super-enhancers specific to certain NB subgroups[68]. 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; NMP, neuromesodermal-potent axial progenitors; SYM, sympathoblast; sig., significant.

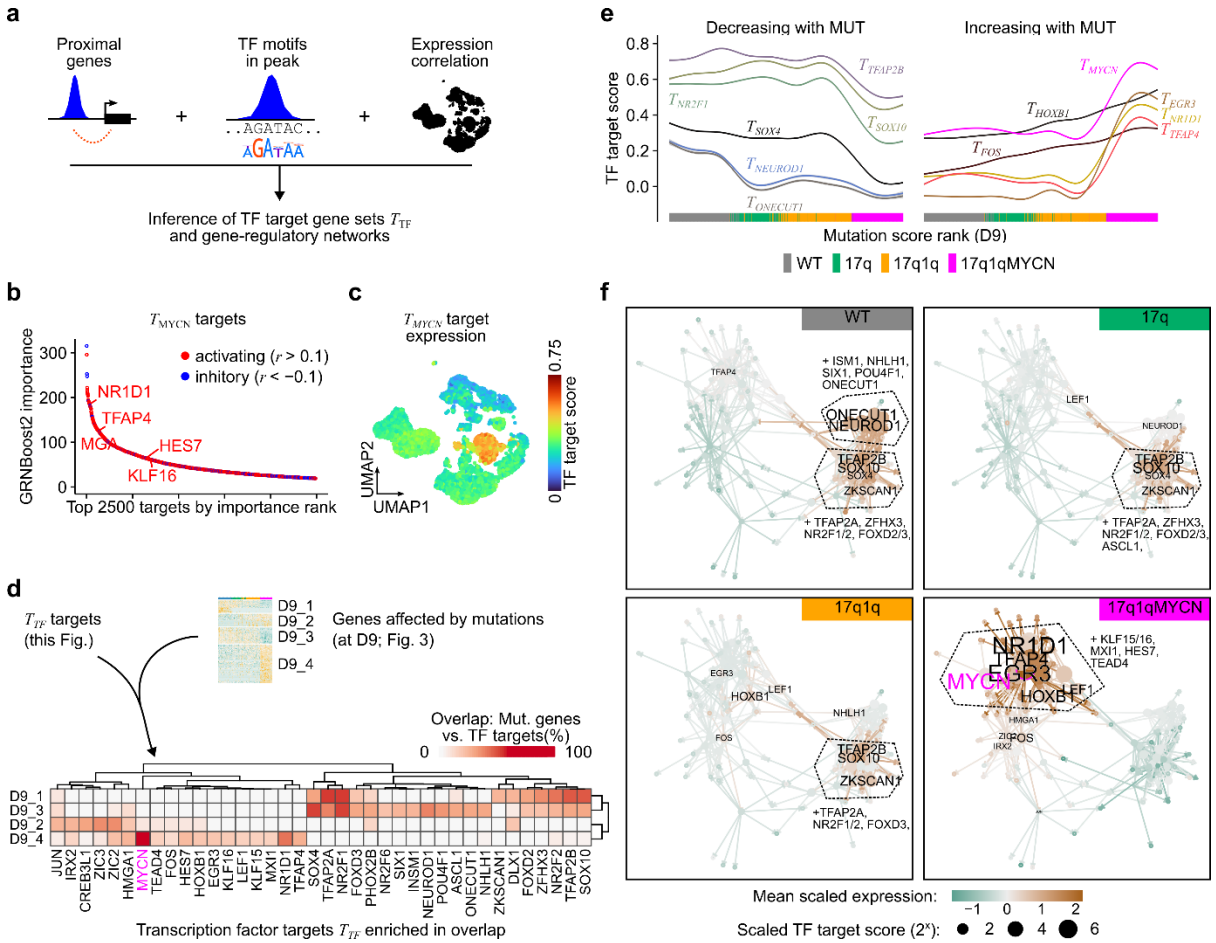


Figure 7. Copy number changes facilitate *MYCN*-mediated blockage of differentiation via developmental transcription factor networks.

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[76] to identify highly correlated TF-target gene candidates based on our scRNA-seq data.

b) Top 2500 targets of MYCN (target set T_{MYCN}) predicted by *GRNboost2*[76]. 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 in the target lists have been highlighted. TF target gene sets are reported in **Table S11**.

c) Average expression (Seurat module score) of the T_{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*[92]; $P_{adj} \leq 0.0005$, $|\log_2 \text{FoldChange}| \geq \log_2(2)$, frequency ≥ 10%). Enrichment results are also reported in **Table S12.**

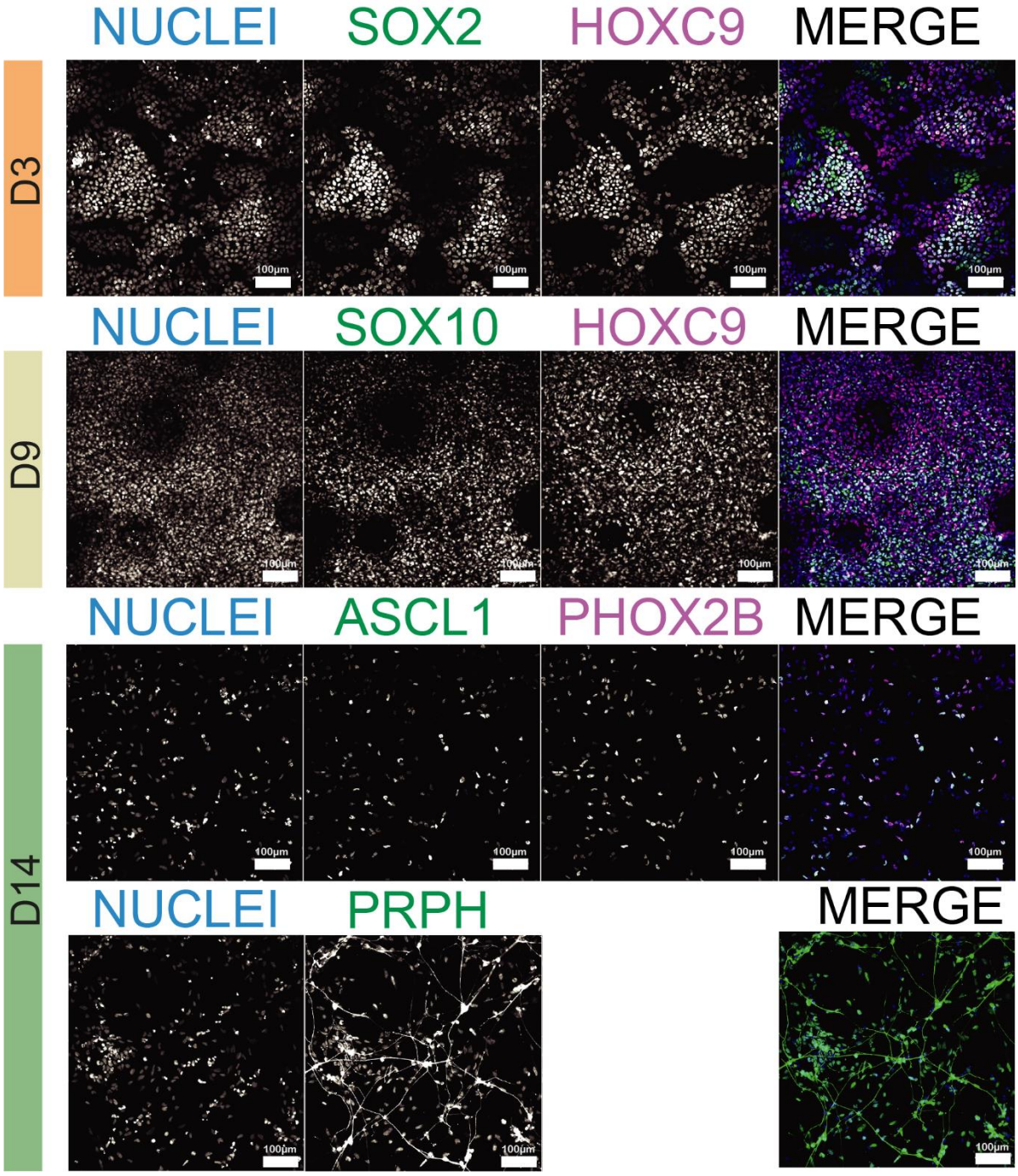
e) Smoothed line plots indicating the average T_{TF} target gene expression (Seurat module score) for selected TFs from **panel (f)**. We split the TFs into two groups corresponding to target genes losing or gaining expression along the module score spectrum. The source cell line of each data point is indicated at the bottom.

f) Gene-regulatory network diagrams visualizing putative TF to target gene regulations for the genes in gene sets D9_1 to D9_4 (cp. Fig. **3e,f**) and enriched TF targets (cp. panels **(c-e)**). In

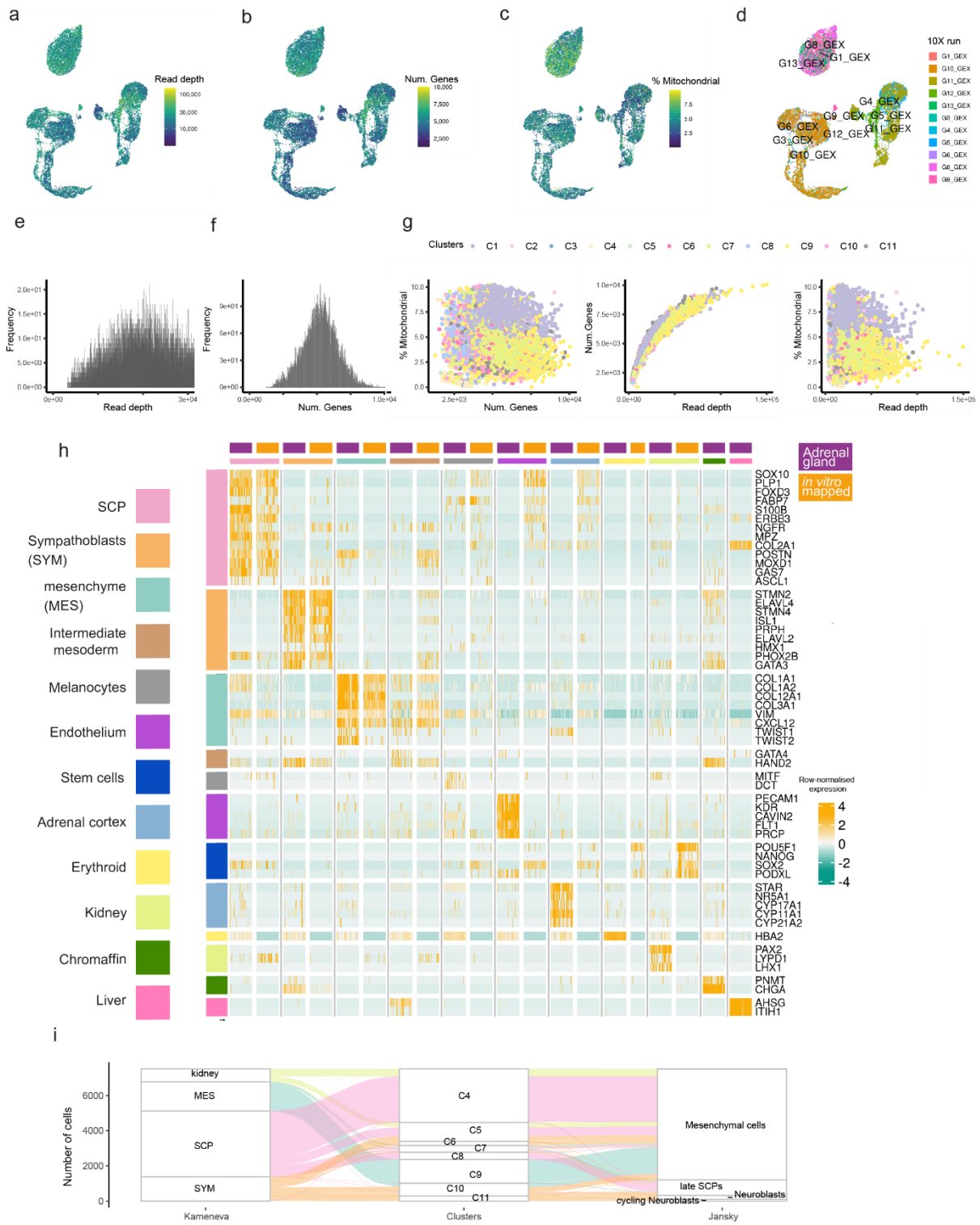
these diagrams, each node represents a TF or target gene, and each edge is a link between a TF and a target. We made these networks specific to cells from each condition (WT, 17q, 17q1q, 17q1qMYCN) by using colour to indicate the mean scaled expression of each gene in the respective cells at D9 (edges coloured by source TF) and node size to indicate the mean scaled T_{TF} target score of each TF. Only labels of TFs with positive scaled expression are shown and selected groups of TFs have been merged for visualisation. All network node labels are shown in **Fig. S10c**.

Abbreviations: D0/3/9/14/19, day 0/3/9/14/19; R1-R9, chromatin region modules; TF, transcription factor; WT, wild-type H7 hESCs; MUT, a "mutant" hESC line (one of: 17q, 17q1q, 17q1qMYCN); sig., significant; r, Pearson correlation coefficient; act., activating (positive correction); inh., inhibitory (negative correlation).

Supplemental figures



Supplemental Figure S1 (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 0/3/9/14.



Supplemental Figure S2 (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**.

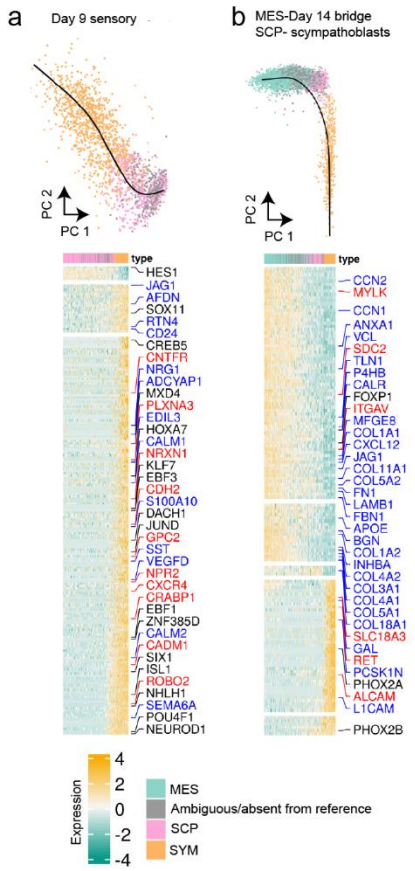
e-f) Histograms depicting the distribution of read depth (**panel (e)**) and number of genes detected (**panel (f)**) per cell barcode after quality control filtering.

g) Scatterplots comparing all quality control covariates (shown on panels (**a-c**)) from the same cell, plotted versus each other, coloured by wild-type clusters (C1-C11).

h) 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**.

i) 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 consistently to one cell identity in the Kameneva *et al.* reference (also indicated in colour). For example, cells that were labelled MES in the paper (cluster C9) also mapped to mesenchymal cells in the Jansky *et al.* reference, cells that mapped to SYM (clusters C10, C11) mapped to cycling neuroblasts and to neuroblasts. Cells that we labelled as SCPs split into cells that mapped to mesenchymal cells (C4, C5) and late SCPs (C8) in Jansky *et al*., consistent with our observations that the former may represent a less mature, early SCP state (see main text).

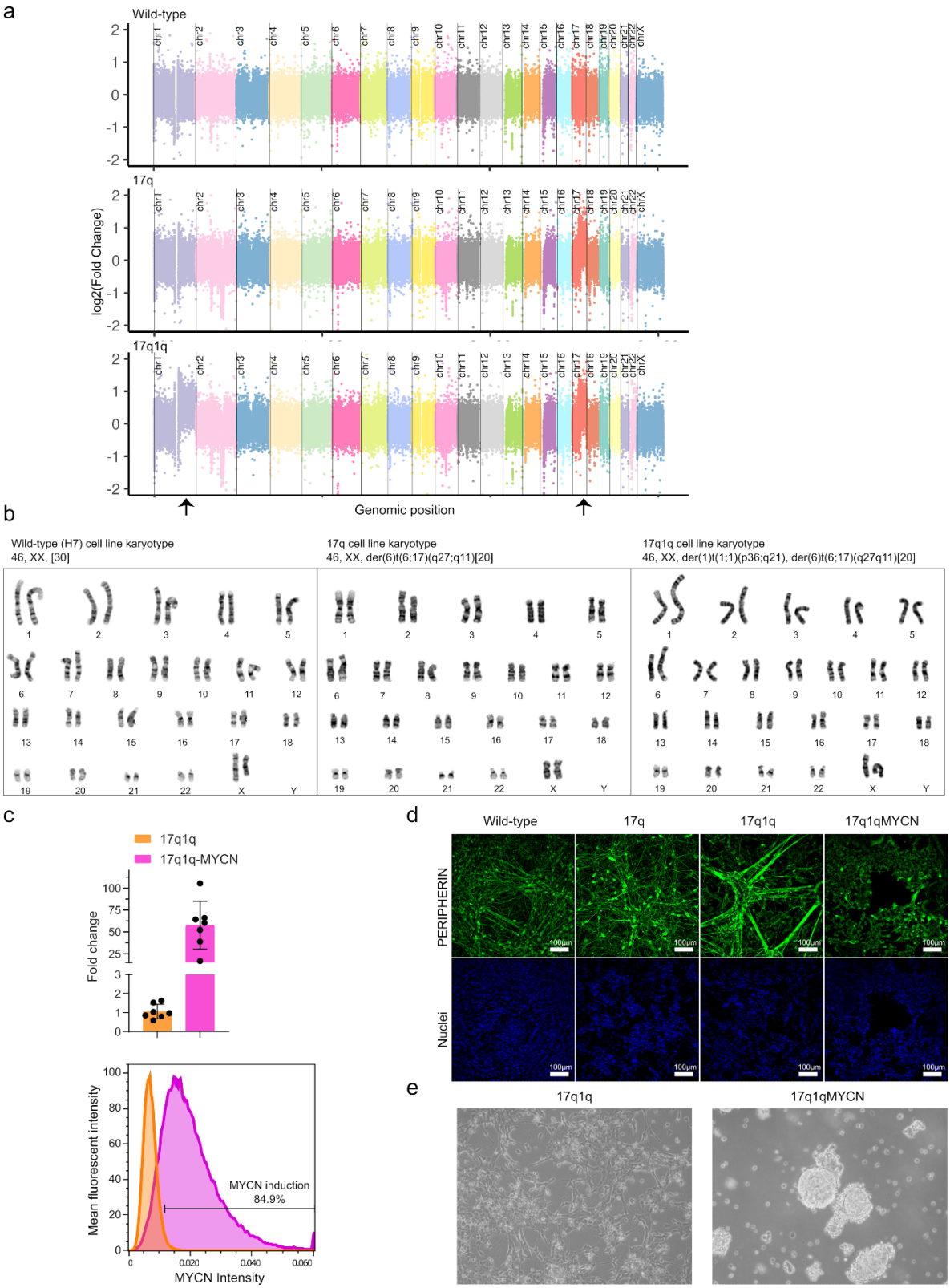
Abbreviations: D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal.



Supplemental Figure S3 (related to Fig. 1). Trajectories connecting gradients of transcriptionally similar cells at different developmental stages.

Slingshot[90] pseudotime trajectories (top) for wild-type clusters C5-C6 (**panel (a)**) and C7, C8, C9, and C11 (**panel (b)**). Cells were separated from the main dataset and reprocessed (see basic scRNA-seq processing in Methods), and trajectories were calculated on the first two principal components with C5 cells and D14 cells (b), respectively, declared as start of the trajectory. Heatmaps (bottom) show the top 140 genes with the strongest association with the trajectory as ranked by *tradeSeq*'s Wald test[97]. Highlighted genes are all TFs (black), receptors (red) and ligands (blue) found in the association test.

Abbreviations: PC, principal component; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal; TF, transcription factor.



Supplemental Figure S4 (related to Fig. 2). Genetic and phenotypic characterization of mutant hESC lines.

a) Plots of the output of CNVkit[98] depicting $\log_2$ fold change of whole-exome sequencing reads relative to the genome average. CNVs can be seen for chr1q and chr17q.

b) Cytogenetic analysis of indicated hESC lines used in the study.

c) Analysis of MYCN expression at the transcript (top) and protein (bottom) level in D9 17q1qMYCN cultures after Doxycycline treatment at day 5 vs untreated control.

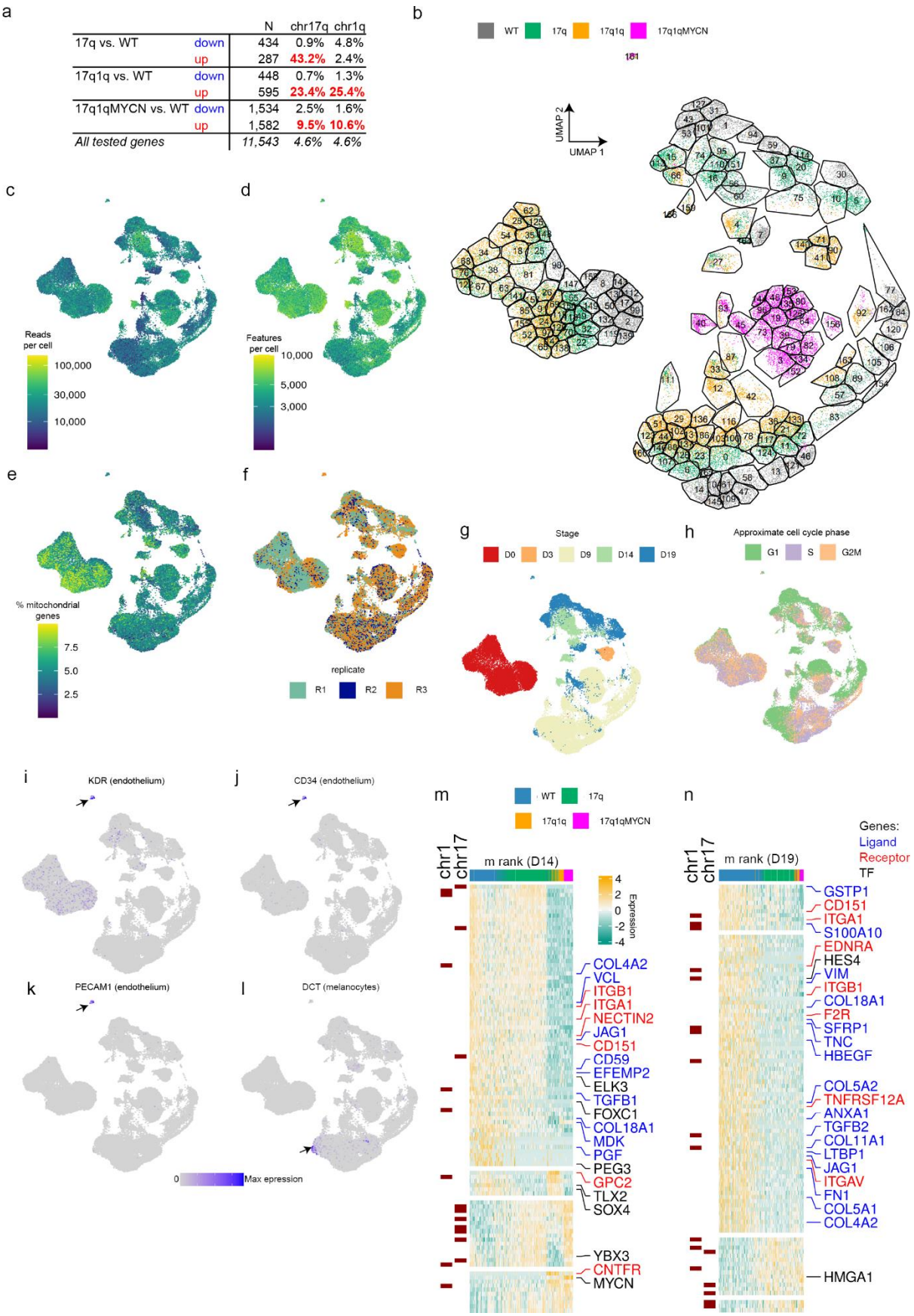
d) Immunofluorescence analysis of PERIPHERIN expression in D19 cultures following differentiation of hESCs with the indicated genotypes. Cell nuclei were counterstained using Hoechst 33342.

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

Abbreviations: WT, wild-type H7 hESCs.

a

		N	chr17q	chr1q
17q vs. WT	down	434	0.9%	4.8%
	up	287	**43.2%**	2.4%
17q1q vs. WT	down	448	0.7%	1.3%
	up	595	**23.4%**	**25.4%**
17q1qMYCN vs. WT	down	1,534	2.5%	1.6%
	up	1,582	**9.5%**	**10.6%**
All tested genes		*11,543*	*4.6%*	*4.6%*



Supplemental Figure S5 (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 CNVs).

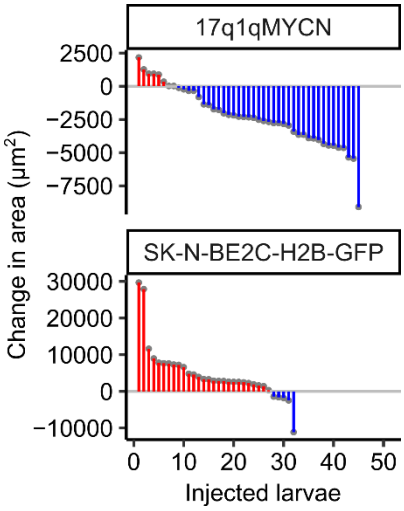
b) High-resolution cell clusters defined for the full *in vitro* TNC dataset scRNA-seq dataset generated in this study. Cluster marker genes are reported in **Table S6**.

c-h) QC covariate plots: reads per cell (**panel (c)**), 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)**).

i-l) Visualisations of endothelial (*KDR*, *CD34*, *PECAM1*) and melanocyte (*DCT*) gene expression across the full *in vitro* UMAP. Arrows highlight the small, high-intensity cluster 161 which expresses all endothelial markers, and cluster 160 which mapped to SCPs, but expressed the melanocyte marker.

m, n) Heatmaps containing the genes correlated or anti-correlated with the mutation score *m* and related measures (see Methods) for D14 (**panel (m)**) and D19 (**panel (n)**). Transcription factors (black), receptors (red) and ligands (blue) have been highlighted.

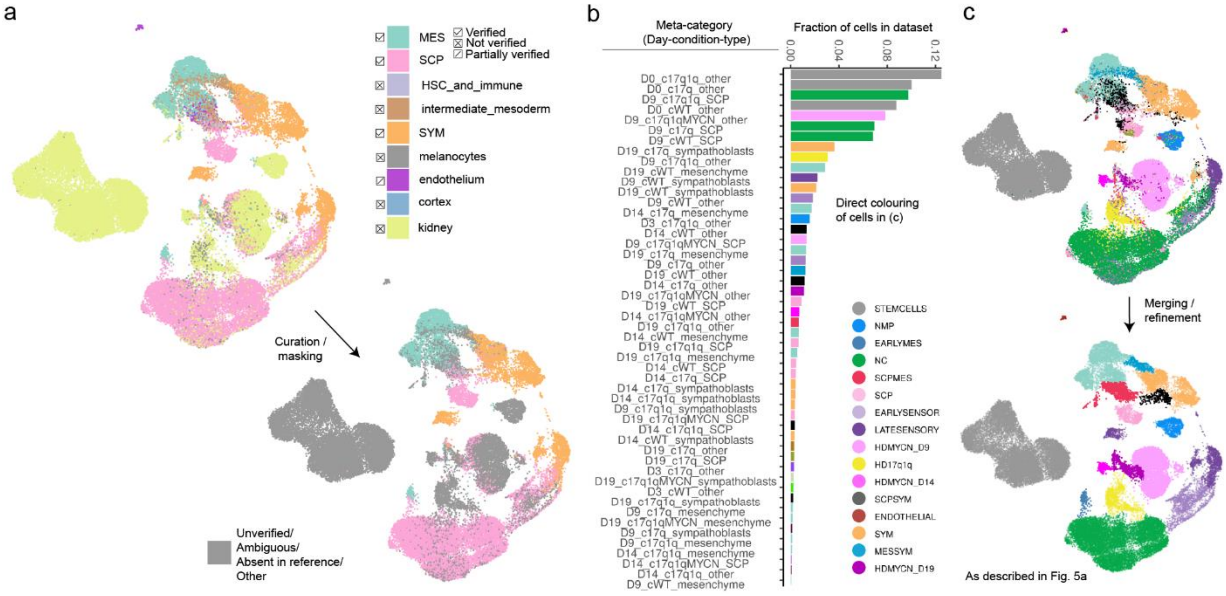
Abbreviations: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal; m rank, mutation score rank.



Supplemental Figure S6 (related to Fig. 4). Survival of xenotransplanted 17q1qMYCN cells and an NB cell line in zebrafish larvae.

Waterfall plots depicting the change in tumour area for 17q1qMYCN at D9 of differentiation (~ NC stage) and SK-N-BE2C-H2B-GFP[66] cells in zebrafish xenografts from 1dpi to 3dpi.

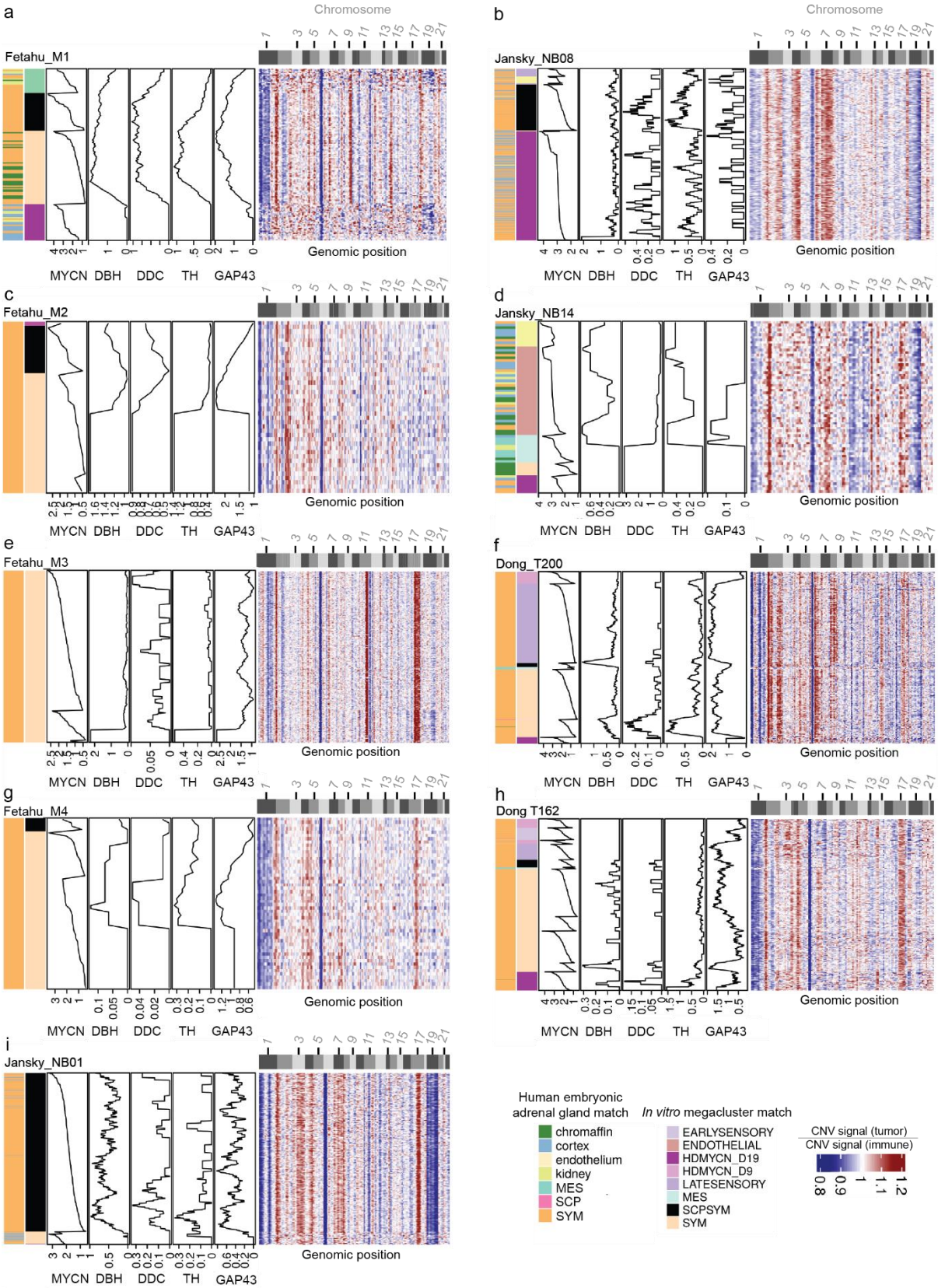
Abbreviations: dpi, day post injection.



Supplemental Figure S7 (related to Fig. 5). Classification of cells in mega-clusters.

Cells of the full *in vitro* trunk NC dataset (UMAP shown in **panel (a)**) containing all the integrated conditions (WT, 17q, 17q1q, and 17q1qMYCN) just as in Fig. 3d were classified via label transfer[93] using the human embryonic adrenal gland cell reference[16]. The classification was then curated and verified via expression of relevant cell type markers. Cells with markers of the matching types were classified as verified (panel (a), boxes with a tick). Labelled cells that did not express markers of the matching cell type, or well-defined cell types were clearly misclassified (such as stem cells classified as kidney), or cell types in neighbourhoods of highly mixed matches, or unlabelled cells were declared as not verified (panel (a), crossed boxes). Cells matching to one type, which expressed the markers in one UMAP region but not in other (such as endothelial cells) were classified as partially verified (panel (a), tickboxes with diagonal lines) All unverified cells were henceforth classified as other. Using the curated cell labels ("type"), we combined information from each cell's stage, condition, and type and classified cells into meta-categories (**panel (b)**) revealing a wide distribution of meta-category frequencies. We coloured the meta-categories as described in **panel (b)**, by merging clusters heuristically when changing the condition and stage did not alter a defined (non-other) type. Groups of several clusters (**panel (c)**, top) were refined by incorporating nearby satellite cells from other meta-categories and merging clusters with the same meta-category to obtain the mega-clusters shown at the bottom of **panel (c)**, and in figure **Fig. 5a**.

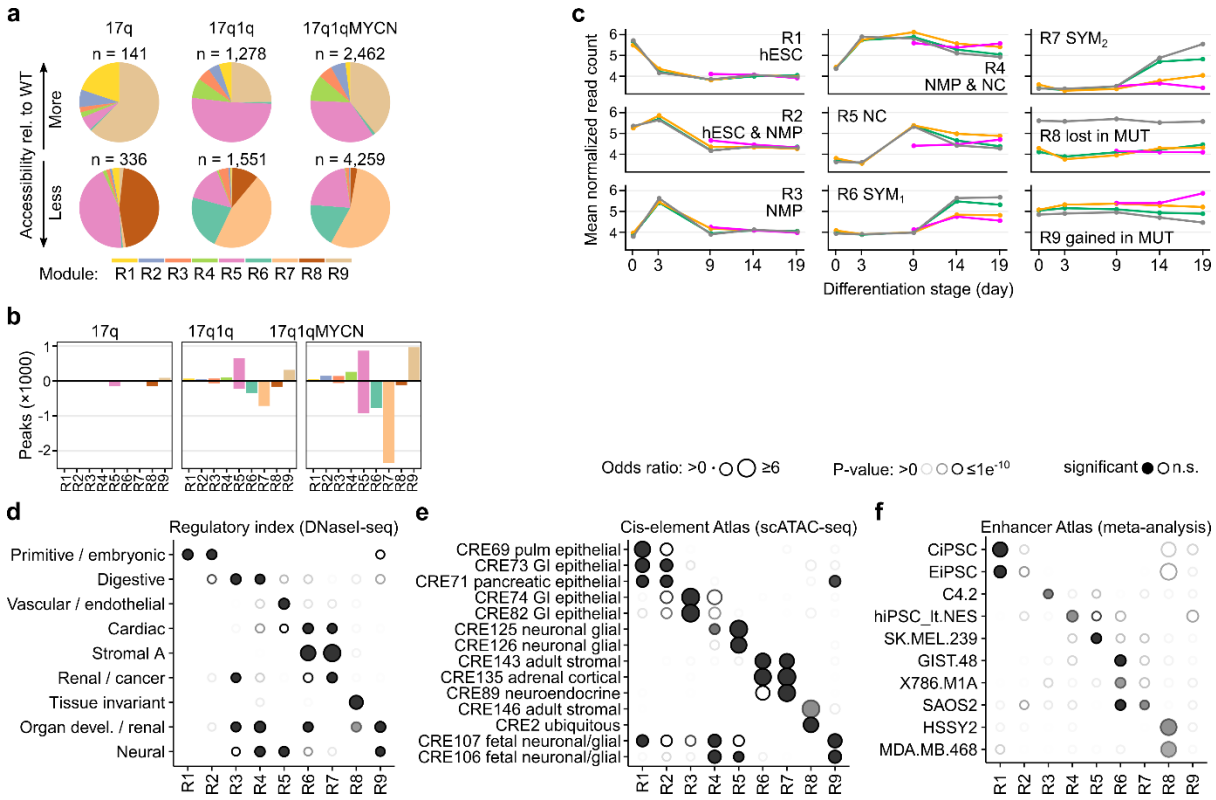
Abbreviations: SCP, Schwann cell precursors; HSC, hematopoietic stem cells; SYM, sympathoblasts; UMAP, Uniform Manifold Approximation and Projection; WT, wild-type.



Supplemental Figure S8 (related to Fig. 5). Analysis of *MYCN*-amplified tumour cells and mapping to *in vitro* trunk NC differentiation.

inferCNV[94] profile heatmaps such as the one in **Fig. 5e** for the remaining 10 tumour datasets[15,17] not shown in **Fig. 5**. Each row (*MYCN*+ tumour cells) and each column (genes, ordered by genomic position), indicate the intensity of the CNV signal relative to non-tumour, HSC/immune cells from the

same sample. All samples were curated and processed as described in **Fig. 5c-e** and mapped both to the human embryonic adrenal gland reference[16] and the full *in vitro* trunk NC differentiation reference (**Figs. 5a**, **S5b, S7**). Cells are ordered first by matching mega-cluster, then by matching Louvain cluster (visualised in **Fig. S5b**), and then by *MYCN* levels within each cluster. Annotations are (left to right): matching adrenal gland cell type, matching mega-cluster, *MYCN*, *DBH*, *DDC*, *TH*, and *GAP43* levels, respectively. Annotation of the chromosomes can be found on top of the heatmap of tumour dataset *Fetahu_M1* and other annotations in the supporting panel.
Abbreviations: CNV, copy number variant; NC, trunk neural crest.



Supplemental Figure S9 (related to Fig. 6). Chromatin accessibility in differentiating wild-type and mutant hESCs.

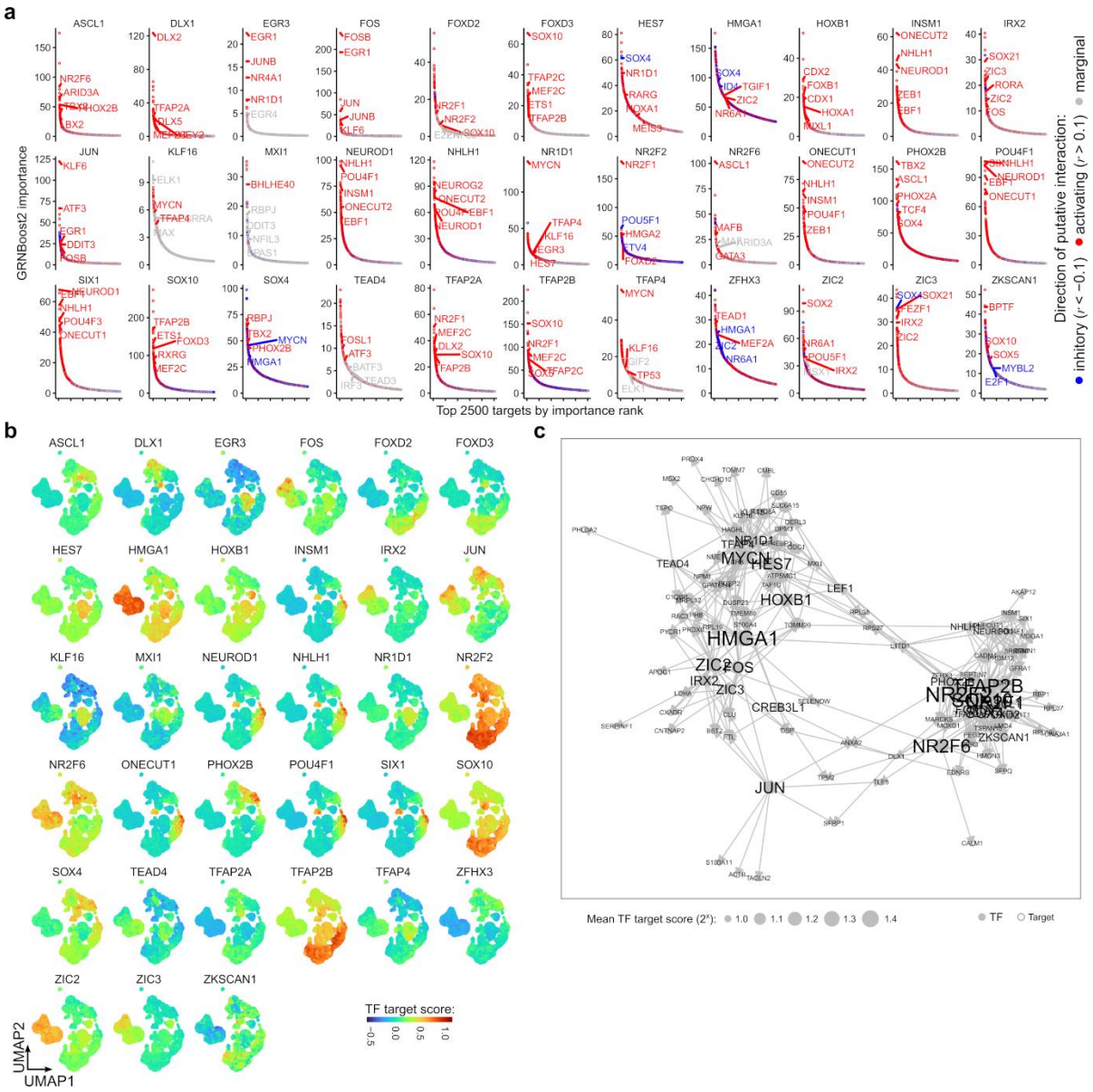
a) Pie charts indicating the fraction of differentially accessible regions belonging to chromatin modules (**Fig. 5b**) in mutant hESCs compared to WT (left to right). Up- and down-regulated regions are shown separately (top vs. bottom row). The total number (n) of regions in each category is indicated.

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[73] (based on DNaseI-seq; **panel (d)**), Cis-element Atlas[74] (based on scATAC-seq analysis; **panel (e)**) and the Enhancer Atlas[75] (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*[92]). Significant results are indicated with filled circles ($P_{adj} \leq 0.005$, $|\log_2 \text{FoldChange}| \geq \log_2(1.5)$, frequency $\geq 2.5\%$). The top enrichments per stage have been selected for visualization (all results are shown in panel **e**) and all results are reported in **Table S10**.

Abbreviations: WT, wild-type H7 hESCs; R1-R9, chromatin modules identified in **Fig. 6e**; n.s., not significant.



Supplemental Figure S10 (related to Fig. 7). Analysis of transcription factor target sets and gene-regulatory networks.

a) Top 2500 targets of selected TFs as predicted by *GRNboost2* algorithm[76] 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 **Table S11**.

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 visualizing 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. **panels (a,b)**). 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; D14/19, day 14/19.

Supplemental tables and files

Supplemental Table S1 (related to Figs. 1, 2, 3, 6, 7). scRNA-seq, RNA-seq, and ATAC-seq dataset overview
Supplemental Table S2 (related to Fig. 1). scRNA-seq cluster marker genes (WT)
Supplemental Table S3 (related to Fig. 1). scRNA-seq markers of SCP-SYM-MES transition states
Supplemental Table S4 (related to Fig. 3). scRNA-seq MUT vs. WT differentially expressed genes
Supplemental Table S5 (related to Fig. 3). scRNA-seq MUT vs. WT enrichment results
Supplemental Table S6 (related to Fig. 3). scRNA-seq cluster marker genes (WT+MUT)
Supplemental Table S7 (related to Fig. 3). scRNA-seq genes correlated to mutations
Supplemental Table S8 (related to Fig. 6). ATAC-seq regions (peaks) and chromatin modules
Supplemental Table S9 (related to Fig. 6). ATAC-seq differential accessibility analysis
Supplemental Table S10 (related to Fig. 6). ATAC-seq chromatin module enrichment results
Supplemental Table S11 (related to Fig. 7). Transcription factor target genes
Supplemental Table S12 (related to Fig. 7). Transcription factor enrichments

Supplemental Video S1 (related to Fig. 5). Time-lapse imaging of 17q1q hESCs in low-density culture
Supplemental Video S2 (related to Fig 5). Time-lapse imaging of 17q1qMYCN hESCs in low-density culture