White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/195022/

Version: Submitted Version

Preprint:
Saldana-Guerrero, I.M., Montano-Gutierrez, L.F., Hafemeister, C. et al. (Submitted: 2022) Neuroblastoma-associated chromosomal aberrations drive cell identity loss in human neural crest via disruption of developmental regulators. [Preprint - bioRxiv] (Submitted)

https://doi.org/10.1101/2022.11.21.515753

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS University of Sheffield UNIVERSITY of York

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

37 # Introduction

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

46 Neuroblastoma (NB) is the most common extra-cranial solid tumour in infants and an
47 archetypal "developmental cancer"[7–9]. NB tumours are usually found in adrenal glands or sympathetic
48 ganglia, tissues derived from the trunk neural crest (NC) lineage during embryonic development[10,11],
49 and studies using transgenic animal models and transcriptome analysis have anchored NB
50 tumourigenesis in impaired sympathoadrenal differentiation of trunk NC cells[12–22]. CNVs such as gains
51 of the long arms of chromosomes 17 (chr17q) and 1 (chr1q) have been identified in the majority (up to
52 65%) of NB tumours[23–27]. These often co-occur with amplification of the *MYCN* oncogene[23,27–31] (at
53 least one CNV in >95% of *MYCN*-amplified tumours[32]), suggesting they may jointly contribute to
54 tumourigenesis. However, despite our advanced understanding of the genetic and developmental origin
55 of NB, it remains unclear to date how CNVs disrupt embryonic cell differentiation and lead to
56 tumourigenesis.

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

68

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

69 # Results

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

72 We have previously described an efficient strategy for the *in vitro* production of human trunk NC,
73 sympathoadrenal progenitors, and sympathetic neurons from hESCs[33,34]. Our protocol involves
74 treatment with defined cocktails of pathway agonists/antagonists that induce neuromesodermal-potent
75 axial progenitors (NMPs) at day 3 of differentiation (D3)[35], and subsequently steer NMPs toward trunk
76 NC cells (D9) and their sympathoadrenal derivatives (>D14). At D19, the protocol yields
77 catecholamine-producing sympathetic neurons marked by peripherin-expressing axons[33] (**Figs. 1a, S1**).

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

95 To validate that cell types in our hESC-derived model resembled their *in vivo* counterparts, we
96 bioinformatically compared our data to cell types from the adrenal glands of human embryos[16] (**Figs.
97 1e, S2h,i**). This mapping confirmed cell populations with SCP-like, sympathoblast-like (SYM), and
98 mesenchymal features (MES). Other foetal, non-NC lineages (e.g., endothelia) or late neuroendocrine
99 derivatives (e.g., chromaffin cells) were absent. Intriguingly, the mapping also revealed two distinct
100 groups of SCP-like cells at D9 (C4-C5) and D14-D19 (C8). Quantification of SCP signature strength
101 indicated that cluster C8 comprised SCPs closely matching those seen in foetal adrenal glands, while
102 clusters C4-C5 matched the cells in this reference to a lesser extent (**Fig. 1f**). This may be because the
103 latter corresponds to an immature NC or early SCP state not captured in the advanced developmental
104 stages in the reference[16] (weeks 6-14 post conception). Moreover, C11 SYM cells matched the

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

110 Opposing gradients of overlapping SYM and MES signatures (**Fig. 1g**) and RNA velocity (**Fig.**
111 **1h**) also indicated that the intermediate cluster C10 comprised cells at a fork in trunk NC development,
112 as observed in mice[46]. Since MES- and SYM-like cells have also been described in NB cell lines[48–52],
113 we sought to study C9-C11 cells along a pseudotime trajectory (**Figs. 1i, S3; Table S3**). This analysis
114 revealed transcripts associated with NB and epithelial-to-mesenchymal transition (EMT) marking an
115 intermediate MES-SYM state, e.g., WNT-antagonist *SFRP1*, transcriptional regulators *NR2F1/2*[53], and
116 chemotaxis gene *RARRES2*[54].

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

122

123 **CNVs and *MYCN* cumulatively disrupt human trunk NC differentiation**

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

142 **2b**). Altogether, 17q1q and 17q1qMYCN cells matched WT cells of earlier developmental stages,
143 suggesting they were delayed in their differentiation (**Fig. 2c**).
144 Next, we tested whether the cell types induced from mutant hESCs still truthfully recapitulated
145 *in vivo* cell types as seen for WT. Mapping mutant cells onto the same human embryonic adrenal gland
146 reference[16] identified fewer SYM- and MES-like cells in 17q1q and 17q1qMYCN (**Figs. 2d,e**). For
147 cells mapped to the respective cell types, we observed a stronger SCP signature in 17q and 17q1q, while
148 the expression of MES/SYM genes was weaker (**Fig. 2f**). In 17q1qMYCN, the expression of all
149 signatures was weak, suggesting a failure to fully specify the expected cell types (**Figs. 2d-g**).
150 Consistently, antibody staining for SOX10 and HOXC9 and flow cytometry revealed depletion of
151 SOX10+ trunk NC cells in 17q1qMYCN cultures (**Fig. 2h**). The reduced ability of 17q1qMYCN hESCs
152 to differentiate toward trunk NC derivatives was also reflected by their failure to generate
153 PERIPHERIN-positive neuronal axons (**Fig. S4d**) and morphological changes such as rounding up and
154 formation of tight, dome-like colonies (**Fig. S4e**).
155 Differential analysis identified 721 (17q vs. WT), 1,043 (17q1q vs. WT), and 3,116
156 (17q1qMYCN vs. WT) differentially expressed genes (DEGs) at D9 (**Table S4**). As expected, many
157 upregulated genes were located within the known CNVs (43% within chr17q for 17q cells; 23% within
158 chr17q and 25% within chr1q for 17q1q cells; **Fig. S5a**). Pathway analysis revealed an enrichment of
159 genes related to E2F and MYC targets as well MTORC1 signalling components for DEGs on chr17q
160 (e.g., *BRCA1*, *NME1*), and of apoptosis-related and members of the p53 pathway on chr1q (e.g., the
161 anti-apoptotic regulator *MCL1*; **Figs. 3a-c; Table S5**). These perturbed pathways may contribute to
162 deregulation of expression outside CNVs (e.g., upregulation of oxidative phosphorylation and
163 downregulation of G2-M checkpoint-related genes in 17q1qMYCN; **Fig. 3a**), and therefore to the
164 changes in cell phenotypes observed.
165 To better resolve the molecular impact of each mutation, we integrated all datasets into a joint
166 projection of WT and aberrant trunk NC differentiation (**Figs. 3d, S5b-h**; **Table S6**). The strongest
167 changes were found in 17q1qMYCN, which formed disconnected cell clusters not normally produced
168 in our protocol, including a subpopulation expressing genes indicative of vascular/endothelial
169 differentiation (e.g., *CD34, KDR, PECAM1*; **Figs. S5i-l**). To delineate the stepwise alteration of
170 transcriptional programmes, we placed cells from D9 on a spectrum from WT to 17q1qMYCN by
171 scoring each cell between 0 and 1 based on the fraction of mutant cells among its gene expression
172 neighbours ("mutation score"; **Fig. 3e**). This allowed us to identify four sets of genes (D9_1–D9_4)
173 correlated with mutations (**Figs. 3f, Fig. S5m,n; Table S7**): Gain of CNVs led to a loss of genes
174 involved in trunk NC differentiation and cell death (e.g., chromaffin cell-associated gene *PEG3* and
175 WNT-antagonist *SFRP1*[58]) and induction of NB-associated transcripts such as *MSX2* and *CNTNAP2*[59]
176 (**Fig. 3f**). *MYCN* overexpression in 17q1qMYCN repressed genes related to NC development (e.g.,
177 *TFAP2A/B* and nuclear receptors *NR2F1/2*[38,60,61]), and additionally triggered the induction of NOTCH
178 target *HES7* and genes associated with metabolic changes/amino acid level regulation (e.g., *NR1D1,*

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

188

189 **Impaired trunk NC differentiation correlates with acquisition of tumourigenic hallmarks**

190 We next examined whether CNVs/*MYCN* amplification led to the acquisition of tumourigenic
191 hallmarks. We first carried out cell cycle analysis of trunk NC cells (D9) generated from each hESC
192 line by monitoring EdU (5-ethynyl-2´-deoxyuridine) incorporation via flow cytometry. We observed a
193 significant increase of cells in S-phase in 17q1qMYCN ($P=0.0001$, two-way ANOVA with Tukey
194 correction; **Fig. 4a**) indicating a faster cell cycle, consistent with NB tumours and cell lines[65].
195 Conversely, no significant difference was found between 17q or 17q1q cells and WT.
196 Immunofluorescence analysis of Ki-67 expression further showed that 17q1qMYCN cultures exhibited
197 a higher proliferation rate by D14 (SCP/sympathoblast stage) compared to their CNV-only counterparts
198 (17q1q vs 17q1qMYCN, $P = 0.0001$; 17q vs 17q1qMYCN, $P < 0.0001$; **Fig. 4b**).

199 We next tested how CNVs/*MYCN* influenced colony formation, another hallmark of
200 tumourigenesis. Low-density plating of trunk NC cells (D9) and image analysis showed increased
201 clonogenicity in 17q1qMYCN, while 17q or 17q1q cells behaved like WT (**Fig. 4c**). At D9,
202 17q1qMYCN cells formed large, tight, and dome-like colonies, while the 17q1q colonies were smaller,
203 spread out, and contained differentiated cells with neurites. Time-lapse imaging showed that the 17q1q
204 cultures were composed of cells with a higher degree of motility compared to 17q1qMYCN (**Videos**
205 **S1, S2**).

206 Finally, we examined the behaviour of 17q1qMYCN- and WT-derived trunk NC cells *in vivo*
207 by labelling the cells with a fluorescent dye (CellTrace Violet) and injecting them into the perivitelline
208 space of zebrafish larvae on day 2 post fertilisation. We found that 17q1qMYCN cells survived longer
209 in zebrafish than WT, which had diminished in number at day 1 post injection (dpi) and were completely
210 absent at 3 dpi (**Fig. 4d**). In contrast, 17q1qMYCN cells survived until 3 dpi with 16% of larvae even
211 showing an increase in xenotransplant size. For comparison, injection of cells from a *MYCN*-amplified
212 NB cell line (SK-N-BE2C-H2B-GFP[66]) resulted in engraftment with subsequent tumour cell growth in
213 84% of larvae (**Fig. S6**).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

214 Together, our results demonstrate that CNV-carrying trunk NC cells transit into a pre-
215 tumourigenic state under the influence of *MYCN* overexpression, as reflected by the acquisition of
216 altered cellular properties reminiscent of cancer hallmarks.
217
218 ***In vitro* differentiation of mutant hESCs captures NB tumour cell heterogeneity**
219 We asked whether the *in vitro* NB-like development could provide insights into NB heterogeneity. To
220 this end, we first categorised our full *in vitro* reference dataset (of WT and mutant cells) into 17 mega-
221 clusters based on gene expression and similarity to *in vivo* cell types (**Fig. 5a**, **S7**). And interrogated the
222 expression of common NB diagnostic markers (**Fig. 5b**). For example, high levels of *B4GALNT1*
223 (involved in the production of the ganglioside GD2) and *CHRNA3* were seen in SYM-like cells of
224 D14/D19, and *DDC* and *DBH* were specific to cells at the intersection of SCP and SYM identity (**Fig.**
225 **5b**). In contrast, *NCAM1* was widely expressed in all NC cells (**Fig. 5b**). These data indicate that typical
226 NB diagnostic markers report distinct developmental stages of human trunk NC/sympathoadrenal
227 specification *in vitro*. The combination of multiple markers may thus provide clues to tumour origin.
228 Next, we tested whether cells in our model indeed reflected cell states in tumours. To this end,
229 we collected scRNA-seq data from eleven *MYCN*-amplified NB tumours from three independent
230 sources[15,17]. For each dataset, we curated *MYCN*$^+$ tumour cells and bioinformatically mapped these to
231 our reference (**Fig. 5c**). For example, this approach matched cells of tumour dataset *Dong_230*[17] to
232 SYM-like, SCP/SYM transitional, sensory neuron-like, and 17q1qMYCN-like mega-clusters (**Fig. 5d**).
233 While tumour cells appeared karyotypically homogeneous (including a chr17q gain) and all expressed
234 *MYCN*, the expression of NB markers differed substantially among those mapped to different mega-
235 clusters **(Fig. 5e**). Interestingly, cells with high *DBH* expression mapped to the SCP/SYM transitional
236 mega-cluster, consistent with *DBH* expression *in vitro (***Fig. 5b***)*. Mapping to the *in vitro* reference
237 helped detect such *DBH*$^+$, SCP/SYM-like cells in at least three other tumours (**Fig. S8**), suggesting it
238 may be a recurrent cell subtype that can be directly modelled *in vitro*. We also detected the
239 differentiation markers (cp. **Fig. 1d**) in distinct tumour cell subsets, e.g., *PHOX2B*, *GATA3,* and *ASCL1*
240 marked cells at the SCP/SYM junction, while tumour cells mapped to clusters consisting of
241 17q1qMYCN cells ("HDMYCN") had a decreased expression of most markers, consistent with the
242 eroded cell identity observed earlier (**Fig. 5f**).
243 Extending the *in vitro* reference mapping to all eleven tumours portrayed differences between
244 datasets, such as one tumour with *MYCN*+ endothelial-like cells similar to those that emerged from
245 *MYCN*-induced developmental distortion *in vitro* (**Figs. 5g, S8**). Interestingly, we found SYM-like and
246 SCP/SYM-like cells in almost all datasets, and even rare populations of *MYCN*+ MES cells in two
247 tumours. Jointly, these observations demonstrate that our *in vitro* model generates cell types that
248 transcriptionally resemble NB cell subpopulations and that it facilitates the systematic dissection of
249 intra-tumour heterogeneity in NB tumours.
250

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

251 **CNVs and *MYCN* disrupt the reconfiguration of NC regulatory circuits during differentiation**
252 NB tumours and cell lines are marked by a 're-wiring' of non-coding regulatory elements (e.g.,
253 enhancers) giving rise to tumour-specific regulatory circuitries[48,49,67–71]. We therefore hypothesised that
254 disruption of developmental TFs also underpins the aberrant differentiation observed in our mutant
255 hESCs (**Figs. 2-4**) and employed the assay for transposase-accessible chromatin followed by
256 sequencing (ATAC-seq)[72] to profile chromatin accessibility in the same samples used for scRNA-seq
257 analysis (n = 51; **Table S1**). Chromatin accessibility serves as a proxy for the dynamic regulatory
258 activity during differentiation. For instance, the promoters of hESC regulator *POU5F1* and trunk NC
259 regulator *SOX10* were accessible only at D0 and D9, respectively (**Fig. 6a**), and the *PHOX2B* promoter
260 had reduced accessibility in 17q1q and 17q1qMYCN at D19 (**Fig. 6b**).
261 Unsupervised analysis of chromatin patterns on a global level showed that WT and 17q hESCs
262 changed consistently throughout differentiation (**Fig. 6c**). In contrast, 17q1q and 17q1qMYCN
263 appeared not to follow the same path as WT in this low-dimensional projection, in line the
264 differentiation defects observed in our previous analyses (cp. **Figs. 2b,c**). To delineate chromatin
265 changes in detail, we performed differential accessibility analysis between all differentiation stages per
266 cell line and between all cell lines at matched stages (**Tables S8, S9**). As in our DEG analysis, we found
267 an increasing number of regions with altered accessibility in 17q (n = 477 regions), 17q1q (n = 2,825),
268 and 17q1qMYCN (n = 6,683; **Fig. 6d**). In total, there were 45,580 regions with differential accessibility
269 in at least one comparison, which we divided into nine chromatin modules R1-R9 (**Fig. 6e**). Modules
270 R1-R7 reflect differentiation order, e.g., regions in module R1 were mostly accessible at D0, and R6
271 comprises regions accessible at D14 and D19. Most changes observed in mutant hESCs derivatives fell
272 within these modules (**Figs. S9a,b**). 17q1q and 17q1qMYCN cells failed to close chromatin that is
273 usually specific to D3 and D9 (R4, R5) and conversely to open chromatin of late sympathoadrenal
274 differentiation (R6, R7; **Fig. S9c**). Additionally, modules R8 and R9 comprised regions with reduced
275 and increased accessibility in mutant hESC derivatives, respectively, independent of differentiation
276 stage.
277 We sought to annotate our chromatin modules by looking for overlaps with genomic regions
278 accessible *in vivo*[73–75] (**Table S10**). In line with our transcriptome data, we found a stepwise change
279 toward chromatin resembling differentiated tissues, e.g., neural tissues in R4/R5 and mesenchyme in
280 R6/R7 (**Figs. S9d-f**). Next, we examined genes near the chromatin modules (**Fig. 6f**). For each module,
281 we found enrichments of specific marker genes identified in our scRNA-seq analysis of WT trunk NC
282 differentiation (i.e., clusters C1-C11 from **Figs. 1b,c**). For example, chromatin module R7 (accessible
283 in late differentiation stages, lost in mutants) was linked to clusters C9/C10 (MES-like cells). Moreover,
284 we examined TF binding motifs in each module to identify potential upstream regulators (**Fig. 6g**). Also
285 here, we found an enrichment of known regulators of each developmental stage, e.g., TFs associated
286 with trunk NC in R3/R5 (e.g., SOX10) and with sympathetic development in R6/R7 (e.g.,
287 PHOX2A/B)[46]. Finally, we found enriched overlaps of modules R7, R8, and R9 with super-enhancers

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

288 associated with subsets of NB tumours[68] with mesenchymal characteristics, with non-*MYCN*-amplified
289 low-risk tumours, and with non-*MYCN*-amplified high-risk tumours, respectively (**Fig. 6h**). No
290 significant overlap was found with super-enhancers specific for *MYCN*-amplified NB.
291 Together, our results suggest a systematic reprogramming of chromatin throughout trunk NC
292 differentiation. In cells with CNVs and *MYCN* overexpression, this orderly reconfiguration of chromatin
293 was severely disrupted, providing a plausible mechanism for the observed developmental defects.
294

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

296 Finally, we sought to investigate the links between CNV/*MYCN*-induced changes in chromatin
297 dynamics, gene-regulatory networks, and the distorted differentiation trajectories observed at the
298 transcriptional level. In our scRNA-seq analyses, we found a stepwise alteration of expression from WT
299 to 17q1qMYCN at D9 comprising four gene sets: D9_1–D9_4 (cp. **Figs. 3a,b**). We hypothesised that
300 these mutation-linked gene sets were also regulated by specific TFs and therefore employed an
301 algorithm to identify TF targets based on correlated expression patterns[76] (**Fig. 7a**). This analysis
302 identified NR1D1 and TFAP4 as putative TF targets of MYCN (**Figs. 7b,c, S10a,b; Table S11**). *TFAP4*
303 inhibition leads to differentiation of *MYCN*-amplified neuroblastoma cells[77,78], and the nuclear receptor
304 *NR1D1* has been shown to correlate with *MYCN* amplification in NB patients[62,63], supporting the
305 validity of the inferred target genes.
306 We intersected the inferred lists of TF targets with the mutation-linked gene sets (D9_1–D9_4)
307 and found an enrichment (**Fig. 7d; Table S12**) of MYCN, TFAP4, and NR1D1 targets in D9_4 (highly
308 expressed in 17q1qMYCN). Conversely, gene sets D9_1 and D9_3 (expressed in WT/17q/17q1q) were
309 enriched for targets of TFs expected at this stage of differentiation, e.g., SOX10/4, TFAP2A/B, and
310 nuclear receptor NR2F1. The expression of targets of these TFs increased or decreased along the
311 mutational spectrum, corroborating the association of these TFs with the mutations (**Fig. 7e**). While
312 most TF targets switched expression rapidly with *MYCN* overexpression, others showed a continuous
313 pattern with up-/down-regulation already detectable in 17q and 17q1q, e.g., targets of anterior axial
314 marker HOXB1[79,80] and EGR3 (up), or of sensory neurogenesis regulator NEUROD1[46] (down). To aid
315 interpretation, we visualised cell-line-specific interactions between TFs and targets as edges in
316 connected network diagrams (**Figs. 7f, S10c**). These diagrams showcased the emergence of a new
317 subnetwork of TFs in 17q1qMYCN that centred on MYCN and incorporated TFs like NR1D1, TFAP4,
318 and EGR3. In contrast, a subnetwork involving NC-related TFs such as SOX10, SOX4, and TFAP2B
319 was lost in these cells. Intriguingly, downregulation of TFs linked to sensory neuronal development
320 (NEUROD1, ONECUT1) was visible already in 17q cells (**Fig. 7f**). In 17q1q, we additionally observed
321 upregulation of TFs including HOXB1, EGR3, and FOS (**Fig. 7f**).
322 In summary, our data suggest a subtle rewiring of gene-regulatory networks in CNV-carrying
323 hESCs, which may be linked to the depletion of sensory NC derivatives and increased early SCP

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

324 signature found in our single-cell analyses (cp. **Fig. 2**). Overexpression of *MYCN* resulted in a switch
325 in favour of known NC-linked TFs downstream of MYCN, including NR1D1 and TFAP4.
326

327 Discussion

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

334 We provide a comprehensive knowledge base of transcriptomic and epigenetic changes in this
335 model on a temporal (i.e., during differentiation) and a genetic (i.e., with different mutations) axis. Our
336 data show that CNVs impair trunk NC differentiation and potentiate an SCP-like gene signature. In this
337 aberrant cell state, overexpression of *MYCN* (mimicking *MYCN* amplification commonly found along
338 with chr17q/chr1q gains in NB tumours) leads to a complete derailment of sympathoadrenal lineage
339 specification, ectopic induction of endothelial cells, and a proliferative, tumour-like cellular phenotype.
340 We also found that acquisition of NB-linked CNVs correlated with upregulation of anti-apoptotic and
341 DNA-damage-repair-related genes and speculate that this may provide an early selective advantage
342 facilitating subsequent oncogenic transformation triggered by *MYCN* overexpression, which in this case
343 acts as a 'second hit' on top of CNVs. This is consistent with previous studies showing that *MYCN*
344 overexpression alone is associated with increased apoptosis in early sympathoadrenal cells[81,82] and
345 triggers tumourigenesis only in combination with additional mutations[13,14,83]. Considering that NB-
346 associated CNVs regularly arise *in vitro*, our findings once again highlight the importance of rigorous
347 monitoring of hESC cultures prior to their use in disease modelling or cell therapy[55].

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

358 Complementing earlier studies using cell lines and animal models[12–14,18,19,22], recent single-cell
359 transcriptomic analyses of NB tumours[15–17] corroborated an origin of NB from neuroblastic, SCP-like

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

360 progenitors, and highlighted intra-tumour heterogeneity comprising subtypes of tumour cells with
361 adrenergic and mesenchymal properties. In our *in vitro* experiments, we also observed cells expressing
362 signatures of both cell types, suggesting that our model could be useful to experimentally investigate
363 the transition between these and other NB-relevant cell types, providing a new scope into their role in
364 therapy resistance[85]. Furthermore, *MYCN* overexpression (in conjunction with CNVs) in nascent trunk
365 NC cells was sufficient to drive tumourigenic traits, suggesting that in some cases NB initiation might
366 occur before SCP/neuroblast emergence and that acquisition of an SCP-like identity may also be a
367 consequence of mutations in earlier stages rather than the origin.

368 Our hESC-based model provides a tractable system for analysing tumour initiation events
369 within disease-relevant human embryonic cell-like populations. However, in this study, we focused on
370 cell-intrinsic transcriptional regulation since our cultures lack tumour-relevant, non-NC cell types (e.g.,
371 immune cells or Schwann cells) and do not recapitulate the structural and physical properties of the
372 human tumour micro- and macroenvironment[86–89]. In the future, it will be possible to combine our
373 system with 3D co-culture approaches with defined cell types or to use biomimetic scaffolds to emulate
374 cell-cell interactions and extrinsic environmental influences.

375 In conclusion, this study unravels the developmental effects of NB-associated mutations and
376 proposes the progressive corruption of gene-regulatory networks by CNVs as an early step toward
377 tumour initiation by selection of undifferentiated progenitor phenotypes. Transformation is then
378 triggered by a second hit with *MYCN* overexpression, which tilts cells toward increased proliferation
379 and formation of aberrant cell types.

380

381 **Acknowledgements**

382 We would like to thank the Biomedical Sequencing Facility at the CeMM Research Center for
383 Molecular Medicine of the Austrian Academy of Sciences for assistance with next-generation
384 sequencing, and Bettina Brunner-Herglotz (CCRI) for her technical assistance. We are also grateful to
385 Igor Adameyko and Polina Kameneva (Medical University of Vienna) and Sofie Mohlin (Lund
386 University) for critical reading of the manuscript. Funding: A.T.: Biotechnology and Biological
387 Sciences Research Council (New Investigator Research Grant, BB/P000444/1), European Union
388 Horizon 2020 Framework Programme (H2020-EU.1.2.2; project 824070), Children's Cancer and
389 Leukaemia Group/Neuroblastoma UK/Little Princess Trust (CCLGA 2019 28, CCLGA 2020 19,
390 together with H.B.), Medical Research Council (MRC; MR/V002163/1). F.H.: Alex's Lemonade Stand
391 Foundation for Childhood Cancer (ALSF; 20-17258), Austrian Science Fund (FWF; TAI 454, TAI
392 732). E.M.P.: FWF (P32001-B, P34832-B). M.C.B.: Austrian Academy of Sciences (25905). M.D.:
393 Austrian Research Promotion Agency (FFG) project 7940628 533 (Danio4Can), ALSF (20-17258).
394 I.B.: UK Regenerative Medicine Platform (MR/R015724/1), MRC (MR/X000028/1). M.F.: ALSF (20-
395 17258). S.T.M.: Vienna Science and Technology Fund (WWTF; project LS18-111), FWF (P35841-B
396 MAPMET), H2020 (project 826494 PRIMAGE). I.F.: FWF (P35072).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

397 **Author contributions**

398 I.S.G. and L.M.G. championed the experimental and computational work on this study, respectively.
399 Formal contributions in authorship order (CrediT taxonomy): Conceptualization: A.T., F.H.; Data
400 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.,
401 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.;
402 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.,
403 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.,
404 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.;
405 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.,
406 L.M.G., C.St., F.H.; Writing – original draft: I.S.G., L.M.G., A.T., F.H.; Writing – review & editing:
407 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.,
408 S.T.M., M.F., A.T., F.H.

409

410 **Declaration of interests**

411 The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

412 # Methods

413

414 **Human embryonic stem cell (hESC) cell culture and differentiation**

415

416 *Cell lines and cell culture*

417 We employed H7 hESCs as a karyotypically normal, female WT control[36]. Use of hES cells has been
418 approved by the Human Embryonic Stem Cell UK Steering Committee (SCSC15-23). Their isogenic
419 chr17q counterparts carry a gain in chromosome 17q (region q27q11) via an unbalanced translocation
420 with chromosome 6[56,99]. The chr17q1q hESC line was clonally derived, after its spontaneous emergence
421 following the genetic modification of chr17q hESCs. The chr17q1q-MYCN hESC line was generated
422 by introducing a TetOn-PiggyBac plasmid (PB-TRE3G-MYCN, plasmid#104542, Addgene) carrying
423 the wild-type version of the *MYCN* gene[100] via nucleofection using the Lonza 4D-Nucleofector System
424 as per the manufacturer's instructions (Amaxa 4D-Nucleofector Basic Protocol for Human Stem Cells).
425 All cell lines were tested regularly for mycoplasma and expression of pluripotency markers. hESCs
426 were cultured routinely in feeder-free conditions at 37°C and 5% CO_2 in E8 media[101] complemented
427 with GlutaMax (Cat# 35050061, Thermo Fisher Scientific) on Vitronectin (Cat# A14700, Thermo
428 Fisher Scientific) as an attachment substrate. All hESC lines described in this manuscript are available
429 upon request and completion of a Material Transfer Agreement.

430

431 *Differentiation toward trunk neural crest*

432 hESC differentiation toward trunk NC and its derivatives was performed using a modified version of
433 the protocol described previously[33,34]. Briefly, hESCs were harvested using StemPro Accutase Cell
434 Dissociation Reagent (Cat# A1110501, Thermo Fisher Scientific) and plated at 60,000 cells/cm^2 in
435 N2B27 medium supplemented with 20 ng/ml of FGF2 (Cat# 233-FB/CF, R&D) and 4 µM of CHIR
436 99021 (Cat# 4423, Tocris) and 10 µM of Rock Inhibitor (Y-27632) (Cat# A11001, Generon). The
437 N2B27 medium consisted of 50:50 DMEM F12 (Merck Life Science / Neurobasal medium (Gibco) and
438 1x N2 supplement (Cat# 17502048, Invitrogen), 1x B27 (Cat#17504044, Invitrogen), 1x GlutaMAX
439 (Cat# 35050061, Thermo Fisher Scientific), 1x MEM Non-essential amino acids (NEAA;
440 Cat#11140050, Thermo Fisher Scientific), 50 µM 2-Mercaptoethanol (Cat# 31350010, Thermo Fisher
441 Scientific). After 24 hours, media was refreshed removing the Rock Inhibitor and cells were cultured
442 for a further 2 days in FGF2/CHIR to generate NMPs. NMPs at D3 were then re-plated at 50,000
443 cells/cm^2 in neural crest inducing medium consisting of DMEM/F12, 1x N2 supplement, 1x GlutaMAX,
444 1x MEM NEAA, the TGF-beta/Activin/Nodal inhibitor SB-431542 (2 µM, Cat# 1614, Tocris),
445 CHIR99021 (1 µM, Cat# 4423, Tocris), BMP4 (15ng/ml, Cat# PHC9534, Thermo Fisher Scientific),
446 the BMP type-I receptor inhibitor DMH-1 (1 µM, Cat# 4126, Tocris), 10 µM of Rock Inhibitor (Y-
447 27632) on Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Cat# A1413202,
448 Thermo Fisher Scientific). 48 hours later (D5), media was replaced removing the Rock Inhibitor. Media
449 was refreshed at D7 and D8. On D5, the expression of MYCN was induced by supplementing the neural
450 crest media with 100ng/ml of Doxycycline (Cat# D3447, Merck). On D9, cells were re-plated at
451 100,000 cells/cm^2 in plates coated with Geltrex (Thermo Fisher Scientific) in the presence of medium
452 containing BrainPhys (Cat# 05790, Stem Cell Technologies), 1x B27 supplement (Cat# 17504044,
453 Invitrogen), 1x N2 supplement (Cat# 17502048, Invitrogen), 1x MEM NEAA (Cat# 11140050, Thermo
454 Fisher Scientific) and 1x Glutamax (Cat# 35050061, Thermo Fisher Scientific), BMP4 (50 ng/ml, Cat#
455 PHC9534, Thermo Fisher Scientific), recombinant SHH (C24II) (50 ng/ml, Cat# 1845-SH-025, R and
456 D) and purmorphamine (1.5 µM, Cat# SML0868, Sigma) and cultured for 5 days (d14 of
457 differentiation). For further sympathetic neuron differentiation, D14 cells were switched into a medium
458 containing BrainPhys neuronal medium (Stem Cell Technologies), 1x B27 supplement (Invitrogen), 1x

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

459 N2 supplement (Invitrogen), 1x NEAA (Thermo Fisher Scientific) and 1x Glutamax (Thermo Fisher
460 Scientific), NGF (10 ng/ml, Cat#450-01 Peprotech), BDNF (10 ng/ml, Cat# 450-02, Peprotech) and
461 GDNF (10 ng/ml, Cat# 450-10, Peprotech).
462
463 *Immunostaining*
464 Cells were fixed using 4% PFA (P6148, Sigma-Aldrich) at room temperature for 10 minutes, then
465 washed twice with PBS (without Ca^{2+}, Mg^{2+}) to remove any traces of PFA and permeabilised using a
466 PBS supplemented with 10% FCS, 0.1% BSA and 0.5% Triton X-100 for 10 minutes. Cells were then
467 incubated in blocking buffer (PBS supplemented with 10% FCS and 0.1% BSA) for 1 hour at RT or
468 overnight at 4°C. Primary and secondary antibodies were diluted in the blocking buffer; the former were
469 left overnight at 4°C and the latter for 2 hours at 4°C on an orbital shaker. Samples were washed twice
470 with blocking buffer between the primary and secondary antibodies. Hoechst 33342 (H3570,
471 Invitrogen) was added at a ratio of 1:1000 to the secondary antibodies' mixture to label nuclei in the
472 cells. We used the following primary antibodies SOX10 (D5V9L) (Cell Signalling, 89356S,1:500);
473 HOXC9 (Abcam, Ab50839,1:50); MYCN (Santa Cruz, Sc-53993, 1:100); PHOX2B (Santa Cruz, SC-
474 376997, 1:500); MASH1 (ASCL1) (Abcam, Ab211327, 1:100 or Santa Cruz, SC-374104, 1:500); Ki67
475 (Abcam, Ab238020, 1:100); PERIPHERIN (Sigma-Aldrich, AB1530, 1:400); TH (Santa Cruz, 25269,
476 1:500). Secondary antibodies: Goat anti-Mouse Affinipure IgG+IgM (H+L) AlexaFluor 647 (Stratech
477 (Jackson ImmunoResearch) 115-605-044-JIR, Polyclonal 1:500); Donkey anti-Rabbit IgG (H+L)
478 Alexa Fluor 488 (Invitrogen, A-21206, 1:1000).
479
480 *Intracellular flow cytometry staining*
481 Cells were detached and resuspended as single cells using StemPro Accutase Cell Dissociation Reagent
482 (Cat# A1110501, Thermo Fisher Scientific) and then counted. Next, 10 million cells/ml were
483 resuspended in 4% PFA at room temperature for 10 minutes. Then cells were washed once with PBS
484 (without Ca^{2+}, Mg^{2+}) and pelleted at 200g. Cells were resuspended in PBS at 10 million/ml and used
485 for antibody staining. Permeabilisation buffer (0.5% Triton X-100 in PBS with 10% FCS and
486 0.1%BSA) was added to each sample, followed by incubation at room temperature for 10 minutes.
487 Samples were then washed once with staining buffer (PBS with 10% FCS and 0.1% BSA) and pelleted
488 at 200g. Then samples were resuspended in staining buffer containing pre-diluted primary antibodies:
489 SOX10 (D5V9L) (1:500; 89356S, Cell Signalling); HOXC9 (1:50; Ab50839, Abcam). The samples
490 were left at 4°C on an orbital shaker overnight. Then, the primary antibodies were removed, and samples
491 were washed two times with staining buffer. After washings, staining buffer with pre-diluted secondary
492 antibody was added to the samples and incubated at 4°C for 2 hours. The secondary antibodies used
493 were Goat anti-Mouse Affinipure IgG+IgM (H+L) AlexaFluor 647 (Stratech (Jackson
494 ImmunoResearch) 115-605-044-JIR, Polyclonal 1:500); Donkey anti-Rabbit IgG (H+L) Alexa Fluor
495 488 (Invitrogen, A-21206, 1:1000). Finally, samples were washed once with staining buffer,
496 resuspended in staining buffer and analysed using the BD FACSJazz flow cytometer. A secondary
497 antibody-only sample was used as a control to set the gating.
498
499 *Cell cycle analysis*
500 The 5-ethynyl-2´-deoxyuridine (EdU) assay was performed following the manufacturer's instructions
501 (Thermo Fisher Scientific, C10633 Alexa Fluor 488). We used 10µM of Edu for a 2-hour incubation.
502 Cells were analysed in the flow cytometer (BD FACSJazz) using the 405 nm laser to detect the Hoechst
503 staining and 488 nm to detect the EdU staining.
504

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

505 *Low-density plating*
506 Day 9 trunk NC cells derived from hESCs as described above were harvested and plated at a density of
507 500 cells/cm^2 in plates pre-coated with Geltrex LDEV-Free Reduced Growth Factor Basement
508 Membrane Matrix (Cat# A1413202, Thermo Fisher Scientific) in the presence of DMEM/F12 (Sigma-
509 Aldrich), 1x N2 supplement, 1x GlutaMAX, 1x MEM NEAA, the TGF-beta/Activin/Nodal inhibitor
510 SB-431542 (2 μM, Tocris), CHIR99021 (1 μM, Tocris), BMP4 (15ng/ml, Thermo Fisher Scientific),
511 the BMP type-I receptor inhibitor DMH-1 (1 μM, Tocris) and ROCK inhibitor Y-27632 2HCl (10 μM).
512 The culture medium was replaced the following day with medium containing BrainPhys (Stem Cell
513 Technologies), 1x B27 supplement (Invitrogen), 1x N2 supplement (Invitrogen), 1x NEAA (Thermo
514 Fisher Scientific) and 1x Glutamax (Thermo Fisher Scientific), BMP4 (50 ng/ml, Thermo Fisher
515 Scientific), recombinant SHH (C24II) (50 ng/ml, R and D) and Purmorphamine (1.5 μM, Sigma). Plates
516 were then incubated at 37°C at 5% CO2. The media was refreshed every 48 hours. After 5 days of
517 culture, cells were fixed (PFA 4%/10min) and stained with Hoechst 33342 (Cat# H3570, Invitrogen)
518 for 5 minutes. Colonies were detected using an InCell Analyser 2200 (GE Healthcare) at a 4X
519 magnification. Images were processed using Cell Profiler.
520
521 *Time-lapse imaging*
522 Time-lapse images of differentiating 17q1q and 17q1q-MYCN trunk NC cultures were taken every
523 hour, from 0 to 84 hours. Imaging started 24 hours after plating using the BioStation CT system (Nikon).
524
525
526 **Zebrafish experiments**
527
528 *Cell preparation for xenotransplantation*
529 Pre-differentiated neural crest cells were frozen on D7 during their *in vitro* differentiation as described
530 above, shipped, and subsequently thawed in DMEM at room temperature. All cells were retrieved in
531 complete neural crest media as described above and plated onto Geltrex-coated wells in the presence of
532 Rock inhibitor (50μM) for 24 hours. 17q1q cells were additionally treated with doxycycline (100ng/ml)
533 to induce *MYCN* expression. On D8, media were refreshed, and respective doxycycline treatment was
534 continued but Rock inhibitor was discontinued. On D9, cells were collected for xenografting
535 experiments and labeled with CellTrace™ Violet (Invitrogen, Thermo Fisher Scientific) for imaging.
536 For this, cells were harvested with Accutase (PAN-Biotech) and resuspended at a concentration of $1*10^6$
537 cells/ml in PBS. CellTrace™ Violet was added to a final concentration of 5 μM for an incubation time
538 of 10 minutes at 37°C in the dark. The cell-staining mixture was filled up with 5 volumes of DMEM
539 supplemented with 10% FBS and the suspension was incubated for 5 min. After gentle centrifugation
540 (5 min, 500 g, 4°C) the collected cells were resuspended in fresh DMEM medium supplemented with
541 10% FBS and incubated at 37°C for 10 min. Adhering/ clumping cells were separated via a 35 μm cell
542 strainer. The cell number was adjusted to a concentration of 100 cells/ml in PBS. The freshly stained
543 cells were kept on ice until transplantation. SK-N-BE2C-H2B-GFP cells[66] (a kind gift of F.
544 Westermann) were cultured in RPMI 1640 medium with GlutaMAX™ (Cat# 61870044, Thermo Fisher
545 Scientific) supplemented with 10 % (v/v) fetal bovine serum (Cat# F7524500ML, Sigma), 80 units/ml
546 penicillin, 80 μg/ml streptomycin (Cat# 15140122, Thermo Fisher Scientific), 1 nM sodium pyruvate
547 (Cat# P0443100, PAN-Biotech), 25 mM Hepes buffer (PAN-Biotech) and 8 μl/ml G418. For zebrafish
548 xenotransplantations, the GFP-labelled cells were harvested and resuspended in PBS at a density of
549 10^5/μl as described above.
550

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

599 4 samples by differentiation stage (from three independent replicate differentiation experiments). Each
600 pooled group was processed using the 10X Genomics Single Cell 3' v3.1 workflow following the
601 manufacturer's instructions. Enriched barcode libraries were indexed following the MULTI-seq
602 workflow[106]. After quality control, libraries were sequenced on the Illumina NovaSeq S4 platform in
603 2x150bp paired-end mode. **Table S1** includes an overview of sequencing data and performance metrics.
604
605 *Raw data processing and alignment*
606 Raw sequencing data were processed with the *CellRanger* v5.01 software (10x Genomics) for cell-level
607 demultiplexing and alignment to the human reference transcriptome (*refdata-gex-GRCh38-2020-A*
608 assembly provided by 10x Genomics; parameters: *--expect-cells=15000 --r1-length 28*). Following
609 initial data processing, all subsequent analyses were performed in R (v4.0.3) using Bioconductor
610 packages and the *Seurat*[93,107,108] (v4.1.1) package.
611
612 *Default basic processing*
613 We applied processing of scRNA-seq data in many instances across this manuscript. Unless parameters
614 are otherwise specified, the default processing of scRNA-seq counts involved the following steps.
615 Counts were normalised for read depth using Seurat's *SCTransform*[109] (parameters:
616 *method="glmGamPoi"*; *variable.features.n=5000*), followed by *RunPCA* (keeping the top 50
617 components), and inference of cell neighbourhoods by *FindNeighbors* on the PCA reduction. Finally,
618 Uniform Manifold Approximation and Projection (UMAP) was performed using Seurat's *RunUMAP*
619 function with default parameters.
620
621 *Quality control*
622 We assessed quality of cells via two complementary pipelines. We first assessed technical covariates
623 and characteristic expression profiles separately per dataset. Here, we kept cells with less than 15%
624 mitochondrial UMI counts, and at least 500 detected genes and applied basic scRNA-seq processing
625 and clustering of the cells (*SCTransform*[109] v0.3.3, parameters: *method="glmGamPoi"*; *RunPCA*
626 keeping 30 dimensions, clustering with default parameters). We used clusters devoid of markers or
627 characterised by markedly higher mitochondrial expression, to derive a library-specific UMI count
628 threshold to further remove low-quality or empty cells (thresholds: $\log_{10}$(nCount_RNA) ≥ G1_GEX:
629 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:
630 3.8, G10_GEX: 3.6, G11_GEX: 3.6, G12_GEX: 3.7, G13_GEX: 3.9). After combining all datasets, we
631 performed am additional technical assessment globally. Here, empty and doublet droplets were flagged
632 with *Emptydrops*[110] (v1.17.3; default parameters) and *scDblFinder*[111] (v1.4.0; parameters: *dbr=0.1*),
633 respectively. We retained only cells with *Emptydrops* FDR>0.01, doublet score smaller than 0.1, and
634 a mitochondrial gene content less than or equal to 10%.
635
636 *Sample demultiplexing*
637 To demultiplex cells belonging to different pooled samples, we counted MULTI-seq barcodes[106] and
638 used two complementary methods: the *MULTIseqDemux* function from *Seurat* (parameters:
639 *autoThresh=TRUE, maxiter=5*) and a custom Gaussian finite mixture model (GMM) that identified
640 per pair of barcodes four groups of cells: G1) positive for barcode 1, G2) positive for barcode 2, G3)
641 negative for both barcodes, and G4) positive for both barcodes (*cb_demux_gmm* function, *canceRbits*
642 v0.1.6). Briefly, for each cell and barcode pair, we calculated the mean and relative differences
643 (difference over the mean) of $\log_{10}$-transformed counts (pseudo-count 10 added). In the first iteration,
644 we used a 1-dimensional mixture model (*mclust* package[112] v5.4.9; *MClustSSC* method; parameters:
645 *modelNames="E", G = 3*) with relative differences as input and the following training data: the 50
646 cells with highest/lowest relative difference as positives for G1 and G2, 50 cells closest to the mean of

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

647 G1 and G2 (i.e., undecided between both groups) as G3. Based on the resulting classification, we
648 generated synthetic G4 training data by sampling cells assigned to G1 and G2 and combining their
649 barcode counts (barcode 1 from G1 cells, barcode 2 from G2 cells). These synthetic doublet cells were
650 added to the training data and a final 2-dimensional mixture model (parameters: *modelNames="VVV",*
651 *G = 4*) with a relative difference and mean as input was used for classification. Only cells unequivocally
652 assigned to one sample by both methods were retained. Only cells classified as singlets by both
653 approaches were retained.
654
655 *Normalisation, clustering, and marker gene analysis for the main dataset*
656 Raw UMI counts were normalised using Seurat's *SCTransform*[109] (parameters: *variable.genes.n=5000,*
657 *method="glmGamPoi", vars.to.regress="ccvar"*) to account for differences in sequencing depth and
658 cell cycle phase (the variable "*ccvar*" variable was calculated as the difference of S and G2/M scores
659 using Seurat's *CellCycleScoring* method with default parameters). To integrate data from three
660 independent differentiation experiments (replicates; **Table S1**), datasets were integrated using
661 *Harmony*[113] (v1.0; parameters: *dims=1:30, group.by.vars = "replicate"*). Nearest neighbours were
662 identified using Seurat's *FindNeighbors* function (parameters: *k=70*) on principal components and a
663 Uniform Manifold Approximation and Projection (UMAP) was calculated using Seurat's *RunUMAP*
664 function (parameters: *n.neighbors=70, min.dist=0.5, dims=1:30*). For the wild-type UMAP (cp. **Fig.**
665 **1**), the dataset was limited to only wild-type cells, where neighbours were recalculated, and the first 10
666 principal components were considered sufficient to capture relevant cell types and transitions among
667 wild-type cells. Clusters were defined using Seurat's *FindClusters* (parameters [full dataset]:
668 *resolution=6,* parameters [WT-only]: *resolution=0.2*). For the wild-type dataset, neighbouring clusters
669 that differed by cell cycle gene expression, but otherwise shared functional markers were merged
670 manually and relabelled to reflect differentiation order. Finally, markers for each cluster were identified
671 using *Seurat*'s *FindMarkers* function (parameters: *method="wilcox", min.logfc=0, min.cells=0.1*),
672 with each cluster compared to all the other cells in the dataset. Genes with an adjusted P-value less
673 than 0.05 and an average $\log_2$ fold change greater than 1 were considered cluster markers after masking
674 all ribosomal and mitochondrial genes (**Table S2**). To compare mutant and wild-type cells, we filtered
675 the integrated dataset to cells from D9 and identified pairwise DEGs ($P_{adj} \leq 0.005$, $|\log_2\text{FoldChange}| >$
676 0.25) between each mutant condition and WT using the *FindMarkers* function (parameters:
677 *test.use="wilcox", only.pos=FALSE, logfc.threshold=0*). Up- and down-regulated DEGs on chr1q, on
678 chr17q, and outside either CNV were then tested separately to identify significant overlaps with
679 MSigDB HALLMARK[91] gene sets using the hypergeometric test implemented in the *hypeR*[92] package
680 (v1.10.0). DEGs and enriched pathways are listed in **Tables S4** and **S5**. To obtain markers for the 167
681 full integrated dataset (WT and mutant) clusters (**Fig. S5b**), we used a more sensitive approach suitable
682 for fewer cells per cluster. Briefly, the cells on each side of the comparison were randomly binned into
683 three pseudo-replicates and the counts were summed to create pseudo-bulk data. These counts were
684 then used as input for *edgeR*[114] (v3.32.1, test type = QLF, default parameters). After processing each
685 cluster, we removed genes with negative fold change, and calculated FDR values. We sorted the results
686 by p-value and discarded genes with FDR > 0.05. The same approach was used to obtain markers for
687 mega-clusters. All non-ribosomal/non-mitochondrial DEGs are reported in **Table S6**.
688
689 *Pseudotime trajectory analysis*
690 Pseudotime trajectories were inferred using *Slingshot*[90] (v1.8.0; default parameters). In each trajectory,
691 we filtered the dataset to adjacent cell clusters where apparent continuities were observed between
692 related cell types (cp. **Figs. 1e,i, S3**). The filtered datasets were then reprocessed using the basic
693 scRNA-seq processing workflow as described above and the first two principal components were used
694 to find trajectories between two extreme clusters. If more than one trajectory was found, the longest

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

695 trajectory (spanning the most cells) was selected. Genes whose expression was associated with the
696 trajectories were identified with the generalised additive model and association test as implemented in
697 *tradeSeq*[97] (v1.4.0; parameters: *knots=3* for MES-SYM and for SCP_SYM (D9), and *knots=6 for MES-*
698 *SCP-SYM (D14)*., The top genes with the highest Wald statistic were selected for reporting (**Table S3**),
699 with different number of genes are shown for each trajectory maximise the number of legible receptors,
700 ligands, and TFs (based on the human transcription factors database[115] and *CellTalkDB*[116]) in **Fig. 1** and
701 **Fig. S3.**
702
703 *Cross-dataset annotation, label transfer, and signature scores*
704 To map data between scRNA-seq datasets, we used label transfer with *Azimuth* and *AzimuthReferences*
705 packages[93] using a Dockerfile provided by the developers (*satijalab/azimuth:0.4.2*). Both query and
706 reference datasets were processed using the default basic scRNA-seq processing workflow as described
707 above and subjected to the Azimuth mapping workflow (*FindTransferAnchors*, *TransferData*,
708 *IntegrateEmbeddings*, *NNTransform*, and *MappingScore* functions; default parameters), using the 50
709 first principal components from both datasets. To visualise cell mappings, we used "glasswork plots",
710 in which the UMAP of the reference was used to define the coordinates of convex hulls for each cluster.
711 Query cells mapping to each cluster were plotted at random positions within their cognate reference
712 cluster hull to mitigate overplotting bias when many cells mapped to a small neighbourhood. The
713 following mappings were performed in this study:
714 1. Human foetal adrenal reference datasets[15,16] onto WT-only (**Figs. 1e, S2h**) and full *in vitro*
715 (**Figs. 2d,e, S7**) scRNA-seq references. Upon obtaining consistent results for both (**Fig. S2i**),
716 the reference provided by Kameneva *et al.* was used throughout the analysis, because of the
717 curated cell type markers they provided (**Fig. S2h**). These gene signatures were also quantified
718 with Seurat's *AddModuleScore* function (default parameters) in **Figs. 1f,g, 2f**.
719 2. Our mutant scRNA-seq data onto the wild-type reference (**Figs. 2b,c**).
720 3. NB tumour data onto our integrated reference (**Figs. 5c-g, S8**) and to the human foetal adrenal
721 reference[16]. See additional details about these datasets and processing in the section *"Pre-*
722 *processing and mapping of NB tumour data"* below.
723
724 *Mutation score analysis*
725 To calculate the mutation score, we encoded each cell's genotype as a number *G* based on the genetic
726 lineage of hESC lines: *G(WT)=0*, *G(17q)=1*, *G(17q1q)=2*, and *G(17q1qMYCN)=3*. We then calculated
727 the mutation score *m* as the mean *G* of the cell's *K* nearest neighbouring cells (K = 70) in the
728 neighbourhood graph (see *"Normalization, clustering, and marker gene analysis"*). Division by 3
729 yielded a score between 0 and 1. Intuitively, the mutation score of a cell indicates whether a cell
730 phenotypically resembles wild-type cells or cells with a given number of relevant alterations
731 independent of its own genotype. To find genes correlated with the mutation score, we calculated
732 Spearman correlations with gene expression in three settings: (i) correlation for each gene with *m* in all
733 cells; (ii) correlation for each gene with *m* leaving out the 17q1qMYCN cells, to emphasise subtle
734 correlations with CNVs; and (iii) correlation for each gene and the neighbourhood entropy (Shannon
735 entropy of all genotype scores *G* of the *K* nearest neighbours), to find genes expressed in phenotypes
736 achieved by two or more mutants. All non-duplicate absolute correlations (calculated using R's *cor.test*,
737 parameters: *method="spearman", exact=TRUE*) were subject to Bonferroni correction and ranked.
738 The top-100 correlated genes ($p \leq 0.05$) per differentiation stage (D9, D14, D19) are reported in **Table**
739 **S3**.
740

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

788 the Illumina HiSeq 4000 platform in 1x50bp single-end mode. **Table S1** includes an overview of the
789 sequencing data and performance metrics.
790
791 _Raw data processing, alignment, and quality control_
792 Quant-seq adapter fragments were trimmed using *bbduk* (from *bbmap* v38.87; parameters: *k=13*
793 *ktrim=r useshortkmers=t mink=5 qtrim=r trimq=10 minlength=20*) prior to alignment to the human
794 reference genome (*refdata-gex-GRCh38-2020-A* assembly provided by 10x Genomics for maximum
795 compatibility with scRNA-seq analyses) using *STAR*[117] v2.7.3a (parameters: *--outFilterType BySJout -*
796 *-outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --*
797 *outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20 --*
798 *alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMmapqUnique 60 --*
799 *outSAMunmapped Within*). Read counts per gene were calculated using *mmquant*[118] v1.3 (parameters:
800 *-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).
801
802
803 **Chromatin accessibility mapping**
804
805 _Library generation and sequencing_
806 ATAC-seq was performed as described previously[72]. Briefly, 20,000 to 50,000 cells were lysed in the
807 transposase reaction mix (12.5 µl 2xTD buffer, 2 µl TDE1 [Illumina], 10.25 µl nuclease-free water, and
808 0.25 µl 1% digitonin [Promega]) for 30 min at 37 °C. Following DNA purification with the MinElute
809 kit (Qiagen) eluting in 12 µl, 1 µl of eluted DNA was used in a quantitative PCR (qPCR) reaction to
810 estimate the optimum number of amplification cycles. The remaining 11 µl of each library were
811 amplified for the number of cycles corresponding to the Cq value (i.e., the cycle number at which
812 fluorescence has increased above background levels) from the qPCR using custom Nextera primers.
813 Library amplification was followed by SPRI (Beckman Coulter) size selection to exclude fragments
814 larger than 1,200 bp. Libraries concentration was measured with a Qubit fluorometer (Life
815 Technologies), and libraries were quality checked using a 2100 Bioanalyzer (Agilent Technologies).
816 Libraries were sequenced by the Biomedical Sequencing Facility at CeMM using the Illumina HiSeq
817 4000 platform in 1x50bp single-end mode. **Table S1** includes an overview of the sequencing data and
818 performance metrics.
819
820 _Raw data processing, alignment, and quality control_
821 Raw sequencing data were processed using *PEPATAC*[119] (v0.9.5; default parameters) including
822 alignment to the human genome (*refdata-cell ranger-atac-GRCh38-1.2.0* assembly provided by 10x
823 Genomics for maximum compatibility with scRNA-seq analyses). Following initial data processing, all
824 subsequent analyses were performed in R (v4.1.2) using Bioconductor packages. After discarding low-
825 quality data (NRF<0.65 or PBC1<0.7 or PBC2<1 or FRiP<0.025), we removed peaks overlapping
826 blacklisted regions from ENCODE
827 (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/hg38.blacklist.bed.gz) and
828 merged overlapping peaks across all ATAC-seq datasets to create a common set of consensus genomic
829 regions for subsequent analysis (**Table S8**). Next, we quantified for each input dataset the number of
830 reads overlapping these consensus peaks using *featureCounts*[120] (*Rsubread* v2.8.1).
831
832 _Differential accessibility analysis and chromatin modules_
833 Raw read counts were loaded into DESeq2[95] (v1.34.0; default parameters, design:
834 *~lane+batch+sample_group*) for normalization (variance-stabilizing transformation) and differential
835 analysis. In doing so, we estimated count size factors for normalization excluding regions on

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

836 chromosomes with known chromosomal aberrations (i.e., chr1, chr17) to avoid overcompensation due
837 to differences in global signal strength. We queried all pairwise comparisons of sample groups stratified
838 by cell line / condition stratified (time-wise differences, e.g., WT-D3 vs. WT-D0) and between
839 conditions stratified by stage (condition-wise differences, e.g., 17q-D9 vs. WT-D9) and recorded all
840 significantly differentially accessible regions ($P_{adj} \leq 0.005$, $|log_2FoldChange| \geq log_2(1.5)$; parameters:
841 *pAdjustMethod="BH", lfcThreshold=log2(1.5), independentFiltering=TRUE;* **Table S9**). To define
842 chromatin regulatory modules, we focused on time-wise differences in WT hESCs (n = 30,749 regions),
843 which we subdivided into six chromatin modules (R1-R6) by hierarchical clustering with cosine
844 distance using the Ward criterion (parameter: *method = "ward.D2"*). To associate ATAC-seq regions
845 with putative target genes, we used the *GenomicRanges*[121] package (v1.46.1) to assign each region to
846 all genes (using the *refdata-gex-GRCh38-2020-A* gene annotation provided by 10x Genomics) with
847 overlapping promoters (transcription start side) or to distal genes whose promoter within a maximum
848 distance of 250kb whose expression was significantly correlated with the region's accessibility. To this
849 end, we calculated the correlation coefficient between normalised read counts in our ATAC-seq data
850 with the normalised read counts in matching samples of our RNA-seq data (mean per stage and
851 condition). We calculated an empirical P-value by shuffling RNA/ATAC assignments (10 repetitions)
852 and retained associations with a P-value $\leq$ 0.05. Annotated regulatory regions from the analysis of
853 ATAC-seq data are listed in **Table S8**.
854
855 *Overlap enrichment analysis for chromatin modules*
856 To characterize the chromatin modules, we interrogated overlaps with genomic regions or associated
857 genes using the hypergeometric test implemented in the *hypeR*[92] package (v1.10.0). We looked at three
858 types of overlaps: (a) Annotated reference regions from the DNase hypersensitivity index[73], from the
859 Cis-element Atlas[74], from the Enhancer Atlas[75], and NB subgroup-specific super-enhancers[68], which all
860 catalogue regulatory elements active in different cell or tissue types. (b) Matches to known TF motifs
861 from the *HOCOMOCO* database[96] (v11). Here, we downloaded motifs from the *HOCOMOCO* website
862 (*HOCOMOCOv11_full_annotation_HUMAN_mono.tsv*) and used *motifmatchr* (v1.16.0) to scan the
863 DNA sequences underlying each genomic region for matches. Regions with at least one match to the
864 motif were recorded as potential binding sites. (c) Marker genes from our scRNA-seq analysis of WT
865 hESC differentiation (**Table S2**). For this purpose, genomic regions were associated with genes as
866 described above. In each case, we used the entire set of all analysed genomic regions as a background
867 for the enrichment analysis, and we considered overlaps with an FDR-corrected P-value less than 0.005
868 (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
869 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)
870 as significant. All enrichment results are reported in **Table S10**.
871
872 *Identification of transcription factor targets*
873 To identify putative target genes of TFs, we used *GRNboost2*[76] (*arboreto* library v0.1.6, with Python
874 v3.8) to identify genes whose expression could be predicted from the expression of each TF. We tested
875 all TFs in the *HOCOMOCO* database[96] for which at least one motif could be identified in our dataset.
876 We found that stronger association values were reported for stem-cell-related factors, likely because of
877 a proportional overrepresentation of this developmental stage in our dataset. To alleviate this effect and
878 create more balanced data to build our networks on, we downsampled our dataset to no more than 500
879 cells per cluster and took the average importance value of eight random samples forward for further
880 analysis. Putative targets with high importance values but without a supporting nearby ATAC-seq peak
881 with a motif matching the respective TF were considered indirect targets and discarded from the target
882 gene sets. We found that the range of importance values varied between TFs. We therefore calculated
883 a TF-specific threshold on the importance score to define target genes. To this end, we ranked

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

884 importance values and used the *changepoint* package (v2.2.3; default parameters) to identify the first
885 point at which the mean values of the curve of importance values changed (disregarding the top 1%
886 highest importance values which often were outliers and disrupted this analysis). The resulting target
887 gene sets were divided into putative activating and inhibiting interactions by the sign of the Pearson
888 correlation coefficient r of the respective TF-target pairs (using the mean correlation value of the same
889 eight random samples as used for *GRNboost2*). Interactions with $|r|<0.1$ were discarded. To calculate
890 the average expression of target genes in cells and to identify significant overlaps between target genes
891 and gene sets D9_1 – D9_4 (**Table S7**), we used only activated targets ($r>0.1$) and the Seurat module
892 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%),
893 respectively. All target gene sets are reported in **Table S11** and all enrichment results in **Table S12**.
894

895 *Gene-regulatory network visualisation*
896 For the visualisation of gene-regulatory networks, we used the *igraph* package (v1.2.9). A directed
897 graph was constructed from edges between genes in the gene sets D9_1, D9_2, D9_3, or D9_4 (**Table**
898 **S7**) and TFs found enriched in the overlap with these genes (**Fig. 7f**). The same automated graph layout
899 (function *layout_nicely()*) was used to draw mutant-specific network diagrams. To generate mutant-
900 specific networks (**Fig. 7h**), we selected cells of cells derived at D9 and parameterised node colour to
901 indicate the mean scaled expression of the genes in those cells and node size to indicate the mean scaled
902 TF target score (Seurat module score) for TFs or the mean scaled expression for non-TFs. To simplify
903 plots, we only labelled TFs with positive mean scaled expression values (>0.05) and manually
904 aggregated many overlapping values, but all node labels are shown in **Fig. S10c**.
905
906

907 ## Data availability
908 Raw and processed single-cell RNA-seq, RNA-seq, and ATAC-seq data will be deposited at the Gene
909 Expression Omnibus (GEO). Public scRNA-seq data from NB tumours used in this study are available
910 under the following accession codes: GSE147821[16] and GSE137804[17] (Gene Expression Omnibus), and
911 EGAS00001004388[15] (European Genome-Phenome Archive).
912
913

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

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

917 # References

918 1. Marshall, G. M. *et al.* The prenatal origins of cancer. *Nat Rev Cancer* **14**, 277–289 (2014).
919 2. Pritchard-Jones, K. Genetics of childhood cancer. *Br Med Bull* **52**, 704–23 (1996).
920 3. Scotting, P. J., Walker, D. A. & Perilongo, G. Childhood solid tumours: a developmental disorder. *Nat*
921 *Rev Cancer* **5**, 481–8 (2005).
922 4. Maris, J. M. & Denny, C. T. Focus on embryonal malignancies. *Cancer Cell* **2**, 447–50 (2002).
923 5. Gröbner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–
924 327 (2018).
925 6. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid
926 tumours. *Nature* **555**, 371–376 (2018).
927 7. Matthay, K. K. *et al.* Neuroblastoma. *Nat Rev Dis Primers* **2**, 16078 (2016).
928 8. Schulte, J. H. & Eggert, A. Neuroblastoma. *Crit Rev Oncog* **20**, 245–270 (2015).
929 9. Zeineldin, M., Patel, A. G. & Dyer, M. A. Neuroblastoma: When differentiation goes awry. *Neuron* **110**,
930 2916–2928 (2022).
931 10. Nakagawara, A. Neural crest development and neuroblastoma: the genetic and biological link. *Prog*
932 *Brain Res* **146**, 231–242 (2004).
933 11. Ponzoni, M. *et al.* Recent advances in the developmental origin of neuroblastoma: an overview. *Journal*
934 *of Experimental and Clinical Cancer Research* **41**, 1–28 (2022).
935 12. Weiss, W. A., Aldape, K., Mohapatra, G., Feuerstein, B. G. & Bishop, J. M. Targeted expression of
936 MYCN causes neuroblastoma in transgenic mice. *EMBO Journal* **16**, 2985–2995 (1997).
937 13. Cohen, M. A. *et al.* Formation of Human Neuroblastoma in Mouse-Human Neural Crest Chimeras. *Cell*
938 *Stem Cell* **26**, 579-592.e6 (2020).
939 14. Olsen, R. R. *et al.* MYCN induces neuroblastoma in primary neural crest cells. *Oncogene* **36**, 5075–
940 5082 (2017).
941 15. Jansky, S. *et al.* Single-cell transcriptomic analyses provide insights into the developmental origins of
942 neuroblastoma. *Nat Genet* **53**, 683–693 (2021).
943 16. Kameneva, P. *et al.* Single-cell transcriptomics of human embryos identifies multiple sympathoblast
944 lineages with potential implications for neuroblastoma origin. *Nat Genet* **53**, 694–706 (2021).
945 17. Dong, R. *et al.* Single-Cell Characterization of Malignant Phenotypes and Developmental Trajectories
946 of Adrenal Neuroblastoma. *Cancer Cell* **38**, 716-733.e6 (2020).
947 18. Hanemaaijer, E. S. *et al.* Single-cell atlas of developing murine adrenal gland reveals relation of
948 Schwann cell precursor signature to neuroblastoma phenotype. *Proc Natl Acad Sci U S A* **118**, (2021).
949 19. Kildisiute, G. *et al.* Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer
950 cell. *Sci Adv* **7**, eabd3311 (2021).
951 20. de Preter, K. *et al.* Human fetal neuroblast and neuroblastoma transcriptome analysis confirms
952 neuroblast origin and highlights neuroblastoma candidate genes. *Genome Biol* **7**, 1–17 (2006).
953 21. Althoff, K. *et al.* A Cre-conditional MYCN-driven neuroblastoma mouse model as an improved tool for
954 preclinical studies. *Oncogene* **34**, 3357–3368 (2015).
955 22. Molenaar, J. J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in
956 neuritogenesis genes. *Nature* **483**, 589–593 (2012).
957 23. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nat Genet* **45**, 279–284 (2013).
958 24. Bown, N. *et al.* Gain of chromosome arm 17q and adverse outcome in patients with neuroblastoma. *N*
959 *Engl J Med* **340**, 1954–1961 (1999).
960 25. Gilbert, F. *et al.* Human neuroblastomas and abnormalities of chromosomes 1 and 17. *Cancer Res* **44**,
961 5444–9 (1984).
962 26. Lastowska, M. *et al.* Comprehensive genetic and histopathologic study reveals three types of
963 neuroblastoma tumors. *J Clin Oncol* **19**, 3080–3090 (2001).
964 27. Mazzocco, K. *et al.* Genetic abnormalities in adolescents and young adults with neuroblastoma: A report
965 from the Italian Neuroblastoma Group. *Pediatr Blood Cancer* **62**, 1725–1732 (2015).
966 28. Bogen, D. *et al.* The genetic tumor background is an important determinant for heterogeneous MYCN-
967 amplified neuroblastoma. *Int J Cancer* **139**, 153–163 (2016).
968 29. O'Neill, S. *et al.* MYCN amplification and 17q in neuroblastoma: evidence for structural association.
969 *Genes Chromosomes Cancer* **30**, 87–90 (2001).
970 30. Huang, M. & Weiss, W. A. Neuroblastoma and MYCN. *Cold Spring Harb Perspect Med* **3**, (2013).
971 31. Schleiermacher, G. *et al.* Accumulation of segmental alterations determines progression in
972 neuroblastoma. *J Clin Oncol* **28**, 3122–3130 (2010).
973 32. Schleiermacher, G. *et al.* Segmental chromosomal alterations have prognostic impact in neuroblastoma:
974 a report from the INRG project. *Br J Cancer* **107**, 1418 (2012).
975 33. Frith, T. J. R. *et al.* Human axial progenitors generate trunk neural crest cells in vitro. *Elife* **7**, (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

976 34. Frith, T. J. R. & Tsakiridis, A. Efficient Generation of Trunk Neural Crest and Sympathetic Neurons
977 from Human Pluripotent Stem Cells Via a Neuromesodermal Axial Progenitor Intermediate. *Curr*
978 *Protoc Stem Cell Biol* **49**, e81 (2019).
979 35. Wymeersch, F. J., Wilson, V. & Tsakiridis, A. Understanding axial progenitor biology in vivo and in
980 vitro. *Development* **148**, (2021).
981 36. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–7
982 (1998).
983 37. Gogolou, A. *et al.* Early anteroposterior regionalisation of human neural crest is shaped by a pro-
984 mesodermal factor. *Elife* **11**, 2021.09.24.461516 (2022).
985 38. Mitchell, P. J., Timmons, P. M., Hébert, J. M., Rigby, P. W. J. & Tjian, R. Transcription factor AP-2 is
986 expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev* **5**, 105–119 (1991).
987 39. Ma, Q., Fode, C., Guillemot, F. & Anderson, D. J. Neurogenin1 and neurogenin2 control two distinct
988 waves of neurogenesis in developing dorsal root ganglia. *Genes Dev* **13**, 1717–1728 (1999).
989 40. Hirsch, M. R., Tiveron, M. C., Guillemot, F., Brunet, J. F. & Goridis, C. Control of noradrenergic
990 differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system.
991 *Development* **125**, 599–608 (1998).
992 41. Tsarovina, K. *et al.* The Gata3 Transcription Factor Is Required for the Survival of Embryonic and
993 Adult Sympathetic Neurons. *Journal of Neuroscience* **30**, 10833–10843 (2010).
994 42. Kastriti, M. E. *et al.* Schwann cell precursors represent a neural crest-like state with biased
995 multipotency. *EMBO J* **41**, e108780 (2022).
996 43. Pattyn, A., Morin, X., Cremer, H., Goridis, C. & Brunet, J. F. Expression and interactions of the two
997 closely related homeobox genes Phox2a and Phox2b during neurogenesis. *Development* **124**, 4065–4075
998 (1997).
999 44. Xie, M. *et al.* Schwann cell precursors contribute to skeletal formation during embryonic development
1000 in mice and zebrafish. *Proc Natl Acad Sci U S A* **116**, 15068–15073 (2019).
1001 45. McGonnell, I. M. & Graham, A. Trunk neural crest has skeletogenic potential. *Curr Biol* **12**, 767–771
1002 (2002).
1003 46. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science (1979)*
1004 **364**, (2019).
1005 47. la Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
1006 48. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat*
1007 *Genet* **49**, 1408–1413 (2017).
1008 49. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation
1009 states. *Nat Genet* **49**, 1261–1266 (2017).
1010 50. van Groningen, T. *et al.* A NOTCH feed-forward loop drives reprogramming from adrenergic to
1011 mesenchymal state in neuroblastoma. *Nat Commun* **10**, 1530 (2019).
1012 51. Ciccarone, V., Spengler, B. A., Meyers, M. B., Biedler, J. L. & Ross, R. A. Phenotypic diversification in
1013 human neuroblastoma cells: expression of distinct neural crest lineages. *Cancer Res* **49**, 219–25 (1989).
1014 52. Biedler, J. L., Helson, L. & Spengler, B. A. Morphology and growth, tumorigenicity, and cytogenetics
1015 of human neuroblastoma cells in continuous culture. *Cancer Res* **33**, 2643–52 (1973).
1016 53. Khalil, B. D. *et al.* An NR2F1-specific agonist suppresses metastasis by inducing cancer cell dormancy.
1017 *Journal of Experimental Medicine* **219**, (2021).
1018 54. Tümmler, C. *et al.* Inhibition of chemerin/CMKLR1 axis in neuroblastoma cells reduces clonogenicity
1019 and cell viability in vitro and impairs tumor growth in vivo. *Oncotarget* **8**, 95135 (2017).
1020 55. Halliwell, J., Barbaric, I. & Andrews, P. W. Acquired genetic changes in human pluripotent stem cells:
1021 origins and consequences. *Nature Reviews Molecular Cell Biology 2020 21:12* **21**, 715–728 (2020).
1022 56. Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells.
1023 *Nat Biotechnol* **22**, 53–54 (2004).
1024 57. Kerosuo, L. *et al.* Enhanced expression of MycN/CIP2A drives neural crest toward a neural stem cell-
1025 like fate: Implications for priming of neuroblastoma. *Proc Natl Acad Sci U S A* **115**, E7351–E7360
1026 (2018).
1027 58. Handel, A. E. *et al.* Developmental dynamics of the neural crest–mesenchymal axis in creating the
1028 thymic microenvironment. *Sci Adv* **8**, 9844 (2022).
1029 59. Szemes, M. *et al.* A Wnt-BMP4 Signaling Axis Induces MSX and NOTCH Proteins and Promotes
1030 Growth Suppression and Differentiation in Neuroblastoma. *Cells* **9**, (2020).
1031 60. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic
1032 remodeling during neural crest specification. *Genome Res* **30**, 35–48 (2020).
1033 61. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human
1034 neural crest. *Cell Stem Cell* **11**, 633–648 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1035 62. Moreno-Smith, M. *et al.* Restoration of the molecular clock is tumor suppressive in neuroblastoma.
1036 *Nature Communications 2021 12:1* **12**, 1–16 (2021).
1037 63. Altman, B. J. *et al.* MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell Metab* **22**,
1038 1009–1019 (2015).
1039 64. Cho, H. *et al.* Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β.
1040 *Nature* **485**, 123–127 (2012).
1041 65. Ryl, T. *et al.* Cell-Cycle Position of Single MYC-Driven Cancer Cells Dictates Their Susceptibility to a
1042 Chemotherapeutic Drug. *Cell Syst* **5**, 237-250.e8 (2017).
1043 66. Wrobel, J. K. *et al.* Rapid In Vivo Validation of HDAC Inhibitor-Based Treatments in Neuroblastoma
1044 Zebrafish Xenografts. *Pharmaceuticals (Basel)* **13**, 345 (2020).
1045 67. Durbin, A. D. *et al.* EP300 Selectively Controls the Enhancer Landscape of MYCN-Amplified
1046 Neuroblastoma. *Cancer Discov* **12**, 730–751 (2022).
1047 68. Gartlgruber, M. *et al.* Super enhancers define regulatory subtypes and cell identity in neuroblastoma.
1048 *Nat Cancer* **2**, 114–128 (2021).
1049 69. Zeid, R. *et al.* Enhancer invasion shapes MYCN-dependent transcriptional amplification in
1050 neuroblastoma. *Nat Genet* **50**, 515–523 (2018).
1051 70. Jahangiri, L. *et al.* Core regulatory circuitries in defining cancer cell identity across the malignant
1052 spectrum. *Open Biol* **10**, 200121 (2020).
1053 71. Zimmerman, M. W. *et al.* Retinoic acid rewires the adrenergic core regulatory circuitry of childhood
1054 neuroblastoma. *Sci Adv* **7**, (2021).
1055 72. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of
1056 frozen tissues. *Nat Methods* **14**, 959–962 (2017).
1057 73. Meuleman, W. *et al.* Index and biological spectrum of human DNase I hypersensitive sites. *Nature 2020*
1058 *584:7820* **584**, 244–251 (2020).
1059 74. Zhang, K. *et al.* A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985-
1060 6001.e19 (2021).
1061 75. Gao, T. & Qian, J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell
1062 types across nine species. *Nucleic Acids Res* **48**, D58–D64 (2020).
1063 76. Moerman, T. *et al.* GRNBoost2 and Arboreto: Efficient and scalable inference of gene regulatory
1064 networks. *Bioinformatics* **35**, 2159–2161 (2019).
1065 77. Xue, C. *et al.* MYCN promotes neuroblastoma malignancy by establishing a regulatory circuit with
1066 transcription factor AP4. *Oncotarget* **7**, 54937–54951 (2016).
1067 78. Boboila, S. *et al.* Transcription factor activating protein 4 is synthetically lethal and a master regulator
1068 of MYCN-amplified neuroblastoma. *Oncogene* **37**, 5451–5465 (2018).
1069 79. Gouti, M., Briscoe, J. & Gavalas, A. Anterior Hox Genes Interact with Components of the Neural Crest
1070 Specification Network to Induce Neural Crest Fates. *Stem Cells* **29**, 858 (2011).
1071 80. Arenkiel, B. R., Tvrdik, P., Gaufo, G. O. & Capecchi, M. R. Hoxb1 functions in both motoneurons and
1072 in tissues of the periphery to establish and maintain the proper neuronal circuitry. *Genes Dev* **18**, 1539
1073 (2004).
1074 81. Kramer, M., Ribeiro, D., Arsenian-Henriksson, M., Deller, T. & Rohrer, H. Proliferation and Survival of
1075 Embryonic Sympathetic Neuroblasts by MYCN and Activated ALK Signaling. *J Neurosci* **36**, 10425–
1076 10439 (2016).
1077 82. Mobley, B. C. *et al.* Expression of MYCN in Multipotent Sympathoadrenal Progenitors Induces
1078 Proliferation and Neural Differentiation, but Is Not Sufficient for Tumorigenesis. *PLoS One* **10**,
1079 e0133897 (2015).
1080 83. Berry, T. *et al.* The ALK(F1174L) mutation potentiates the oncogenic activity of MYCN in
1081 neuroblastoma. *Cancer Cell* **22**, 117–130 (2012).
1082 84. Svaren, J. & Meijer, D. The molecular machinery of myelin gene transcription in Schwann cells. *Glia*
1083 **56**, 1541–1551 (2008).
1084 85. Gautier, M., Thirant, C., Delattre, O. & Janoueix-Lerosey, I. Plasticity in neuroblastoma cell identity
1085 defines a noradrenergic-to-mesenchymal transition (Nmt). *Cancers (Basel)* **13**, (2021).
1086 86. Lazic, D. *et al.* Landscape of Bone Marrow Metastasis in Human Neuroblastoma Unraveled by
1087 Transcriptomics and Deep Multiplex Imaging. *Cancers (Basel)* **13**, 4311 (2021).
1088 87. Verhoeven, B. M. *et al.* The immune cell atlas of human neuroblastoma. *Cell Rep Med* **3**, 100657
1089 (2022).
1090 88. Costa, A. *et al.* Single-cell transcriptomics reveals shared immunosuppressive landscapes of mouse and
1091 human neuroblastoma. *J Immunother Cancer* **10**, (2022).
1092 89. Weiss, T. *et al.* Schwann cell plasticity regulates neuroblastic tumor cell differentiation via epidermal
1093 growth factor-like protein 8. *Nat Commun* **12**, (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1094 90. Street, K. *et al.* Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC*
1095 *Genomics* **19**, 477 (2018).
1096 91. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell*
1097 *Syst* **1**, 417 (2015).
1098 92. Federico, A. & Monti, S. hypeR: an R package for geneset enrichment workflows. *Bioinformatics* **36**,
1099 1307–1308 (2020).
1100 93. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
1101 94. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.
1102 *Science (1979)* **344**, 1396–1401 (2014).
1103 95. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq
1104 data with DESeq2. *Genome Biol* **15**, 550 (2014).
1105 96. Kulakovskiy, I. v. *et al.* HOCOMOCO: Towards a complete collection of transcription factor binding
1106 models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252–D259
1107 (2018).
1108 97. van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing
1109 data. *Nat Commun* **11**, 1–13 (2020).
1110 98. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection
1111 and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873 (2016).
1112 99. Price, C. J. *et al.* Genetically variant human pluripotent stem cells selectively eliminate wild-type
1113 counterparts through YAP-mediated cell competition. *Dev Cell* **56**, 2455-2470.e10 (2021).
1114 100. Randolph, L. N., Bao, X., Zhou, C. & Lian, X. An all-in-one, Tet-On 3G inducible PiggyBac system for
1115 human pluripotent stem cells and derivatives. *Sci Rep* **7**, (2017).
1116 101. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat Methods* **8**,
1117 424–429 (2011).
1118 102. Fior, R. *et al.* Single-cell functional and chemosensitive profiling of combinatorial colorectal therapy in
1119 zebrafish xenografts. *Proc Natl Acad Sci U S A* **114**, E8234–E8243 (2017).
1120 103. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013)
1121 doi:10.48550/arxiv.1303.3997.
1122 104. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA
1123 sequencing data. *Nat Genet* **43**, 491–501 (2011).
1124 105. van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis
1125 Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-11.10.33 (2013).
1126 106. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-
1127 tagged indices. *Nat Methods* **16**, 619–626 (2019).
1128 107. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21 (2019).
1129 108. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data
1130 across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
1131 109. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using
1132 regularized negative binomial regression. *Genome Biol* **20**, 576827 (2019).
1133 110. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell
1134 RNA sequencing data. *Genome Biol* **20**, 1–9 (2019).
1135 111. Germain, P. L., Robinson, M. D., Lun, A., Garcia Meixide, C. & Macnair, W. Doublet identification in
1136 single-cell sequencing data using *scDblFinder*. *F1000Research 2022 10:979* **10**, 979 (2022).
1137 112. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density
1138 Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289–317 (2016).
1139 113. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat*
1140 *Methods* **16**, 1289–1296 (2019).
1141 114. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
1142 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
1143 115. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
1144 116. Shao, X. *et al.* CellTalkDB: a manually curated database of ligand–receptor interactions in humans and
1145 mice. *Brief Bioinform* **22**, (2021).
1146 117. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
1147 118. Zytnicki, M. mmquant: How to count multi-mapping reads? *BMC Bioinformatics* **18**, 1–6 (2017).
1148 119. Smith, J. P. *et al.* PEPATAC: an optimized pipeline for ATAC-seq data analysis with serial alignments.
1149 *NAR Genom Bioinform* **3**, lqab101 (2021).
1150 120. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning
1151 sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
1152 121. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput*
1153 *Biol* **9**, e1003118 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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).

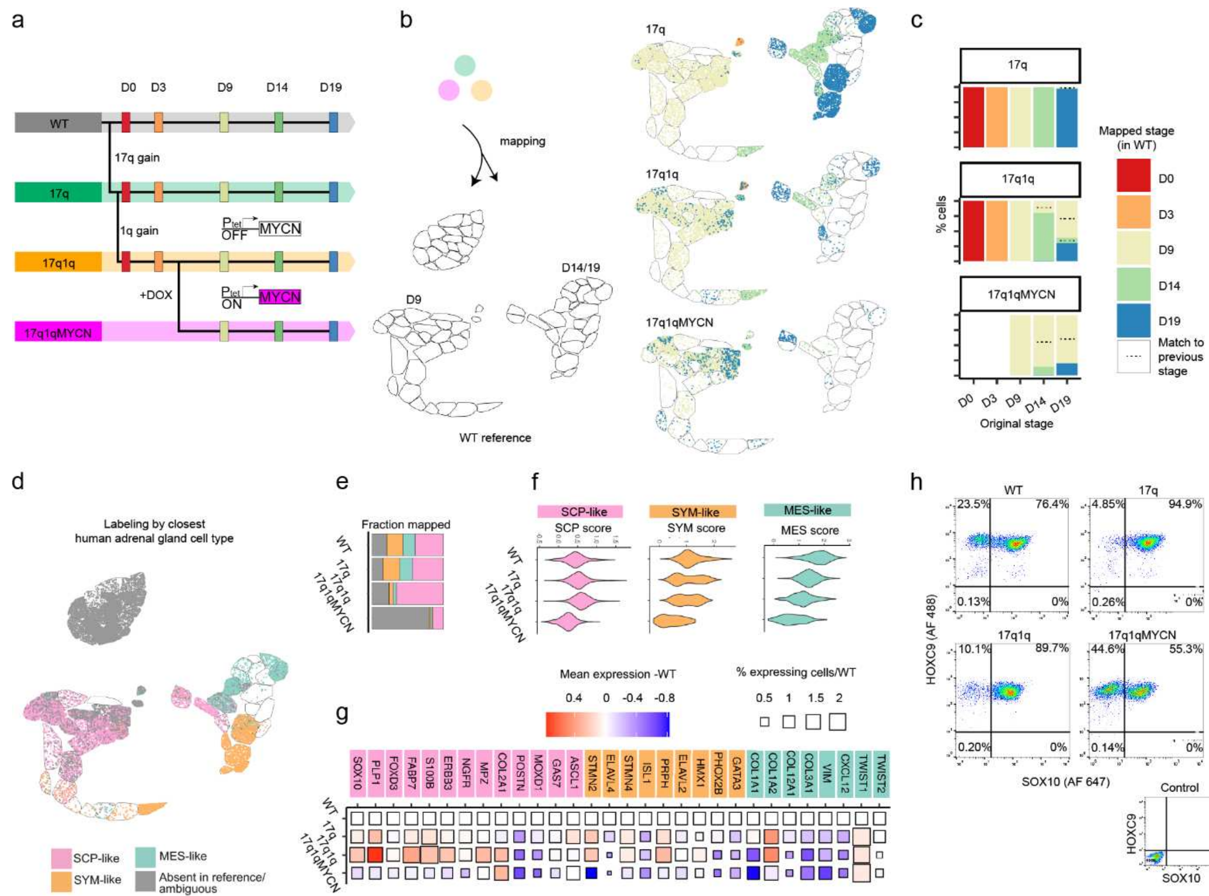
c) Bubble plot indicating the mean expression (colour) and fraction of cells expressing (size) each of the HOX genes per cluster (from **panel (b)**). Genes have been ordered from sacral to cranial axis specification.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1170 **d)** Heatmap of gene markers for each cluster in **panel (b)**. Selected genes have been highlighted
1171 and UMAPs indicate the expression level of canonical markers for stem (*POU5F1*), neural crest
1172 (*SOX10*), mesenchymal (*FN1*), and sympathetic (*PHOX2B*) cells. All marker genes are reported
1173 in **Table S2**.
1174 **e)** Cells from D9-D19 of **panel (b)** labelled by their closest matching cell type from the human
1175 embryonic adrenal gland reference[16] via label transfer. Cells in grey could not be verified with
1176 markers (**Fig. S2h**) or could not be assigned to a single type. Pseudotime trajectories for these
1177 panels can be found on **panel (i)** and **Fig. S3**.
1178 **f)** Cells from **panel (e)** coloured by the strength of their SCP marker signature (Seurat module
1179 score) in red.
1180 **g)** Same as above but visualising simultaneously SYM (blue) and MES (red) marker signature.
1181 Cells with overlapping marker signatures appear in grey/purple tones.
1182 **h)** RNA velocities calculated for the cells in **panel (e)** using Velocyto[47].
1183 **i)** Slingshot[90] pseudotime trajectory (top) for MES- and SYM-like cells in clusters C9-C11,
1184 coloured as in **panel (e)**. Cells were reprocessed and trajectory was calculated on the first two
1185 principal components. The heatmap (bottom) depicts the top 140 genes associated with the
1186 trajectory. Selected genes associated with the intermediate cells are highlighted. All trajectory-
1187 associated genes are reported in **Table S3**.

1188 **Abbreviations**: HPSC, human pluripotent stem cells; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform
1189 Manifold Approximation and Projection; C1-C11, cell clusters; SCP, Schwann cell progenitor; SYM,
1190 sympathoblast; MES, mesenchymal.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1191
1192
1193 **Figure 2. Copy number variants and overexpression of *MYCN* impair the specification of trunk**
1194 **NC derivatives.**

1195 **a)** Scheme depicting the different hESC genetic backgrounds employed and the timing of
1196 Doxycycline (Dox)-induced MYCN overexpression in the context of our trunk NC
1197 differentiation system.

1198 **b)** scRNA-seq data from all mutant cells (17q, 17q1q, 17q1qMYCN at all developmental stages)
1199 were mapped to the wild-type trunk NC reference (illustration on the left side). Glasswork
1200 UMAP plots (right side) depicting the destination clusters in the WT reference for cells of the
1201 17q, 17q1q and 17q1qMYCN conditions. Mutant cells are coloured by stage to emphasise
1202 mismatches with WT.

1203 **c)** Barplots summarising the mappings from **panel (b)** for derivatives of each hESC line (top to
1204 bottom). The position on the x-axis indicates the stage at which the cell sample was collected
1205 and the colour of the bar the stage to which each cell was mapped. Mismatched mappings to
1206 earlier developmental stages are indicated with three dots ("…").

1207 **d)** Glasswork UMAPs as in **panel (b)** coloured by closest-matching cell type in the human
1208 embryonic adrenal gland reference[16]. The category "other" comprises other cell types in the
1209 reference dataset and low-confidence mappings.

1210 **e)** Percentage of cells mapped to each cell type in **panel (d)** split by cell line.

1211 **f)** Violin plots indicating the strength of the SCP/SYM/MES (left to right) gene expression
1212 signature (Seurat module score) for cells mapped to the respective cell type, split by cell line.

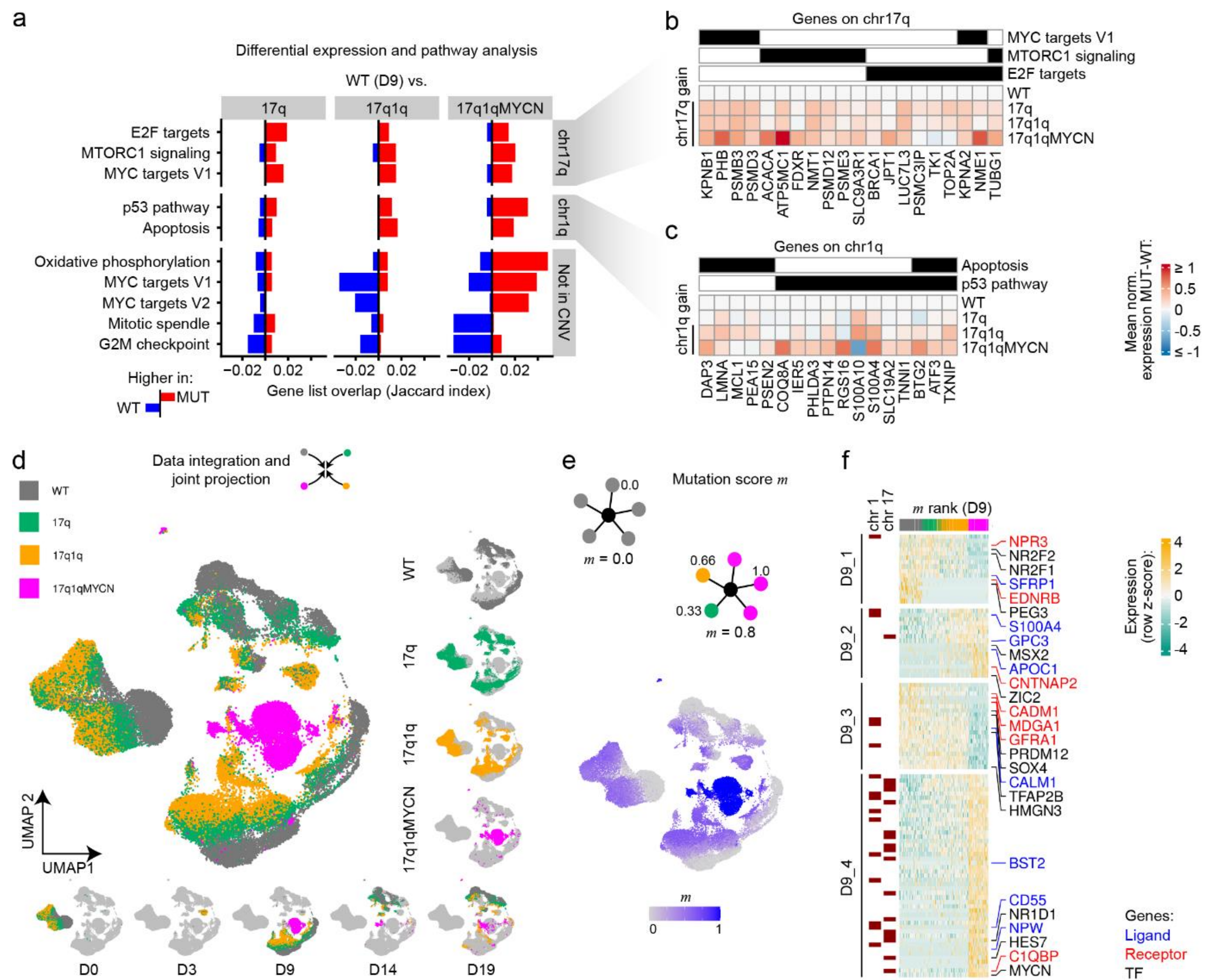
1213 **g)** Plot indicating the change in mean expression (colour) and the percentage of cells expressing
1214 the gene (size) for each gene in the signatures from **panel (e)** relative to WT. WT squares (= 1)
1215 are shown for reference.

1216 **h)** Flow cytometric analysis of the expression of the trunk NC markers HOXC9 and SOX10 in D9
1217 cultures obtained from hESCs marked by the indicated NB-associated lesions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1218 **Abbreviations**: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold
1219 Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES,
1220 mesenchymal.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1221
1222
1223 **Figure 3. Copy number variants and overexpression of *MYCN* alter the expression of metabolic**
1224 **and developmental pathways.**

1225 **a)** MSigDB hallmark pathways[91] enriched (hypergeometric test, *hypeR*[92]; $P_{adj} \leq 0.05$, $|\log_2 \text{odds}| > $
1226 $\log_2(2)$, at least 6 genes in overlap) in differentially expressed genes (DEGs) in trunk NC cells
1227 from each mutant cell line compared to WT at D9 (left to right). The overlap between up- and
1228 down-regulated DEGs with the pathway genes is indicated as a positive (red colour bars) or
1229 negative (blue colour) number, respectively. We additionally distinguished between DEGs
1230 located on chromosome arms chr17q, chr1q, or anywhere else in the genome to analyse
1231 potential direct and indirect effects of CNVs (split from top to bottom). All differentially
1232 expressed genes and pathway enrichments are available in **Tables S4 and S5**.

1233 **b)** DEGs located on chromosome arm chr17q from the enriched pathways shown in **panel (a)**.
1234 The heatmap indicates the mean normalised expression difference between each indicated
1235 mutant cell line and WT (at D9).

1236 **c)** As **panel (b)**, but for DEGs on chr1q and the respective enriched pathways.

1237 **d)** UMAP of scRNA-seq data of scRNA-seq data from wild-type and mutant hESCs (see **Fig. 2a**)
1238 throughout differentiation to trunk neural crest and sympathoadrenal derivatives. Separate
1239 UMAPs indicating cells belonging to each of the four cell lines (left) and each of the five
1240 developmental stages sampled (bottom) are shown.

1241 **e)** Illustration (top) of the calculation of mutation scores *m* (k-nearest neighbour (KNN)
1242 mutational average) as average score of each cell's neighbours. In this calculation, each
1243 neighbour weighs in by its cell line (0 = WT, 1/3 = 17q, 2/3 = 17q1q, 1 = 17q1qMYCN) such
1244 that the mutation score allows ordering cells from WT to MYCN mutation. The actual scores
1245 are shown overlaid on the UMAP from **panel (d)** (bottom).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1246 **f)** Heatmap showing the expression of top 100 genes highly correlated to the mutation score *m*
1247 from **panel (e)** across all cells from D9. Genes have been divided into four groups by
1248 hierarchical clustering, and selected TFs, receptors, and ligands are highlighted. All correlated
1249 genes are reported in **Table S7**. Genes located on chr17q or chr1q are indicated.

1250 **Abbreviations**: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold
1251 Approximation and Projection; m, mutation score; TF, transcription factor.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

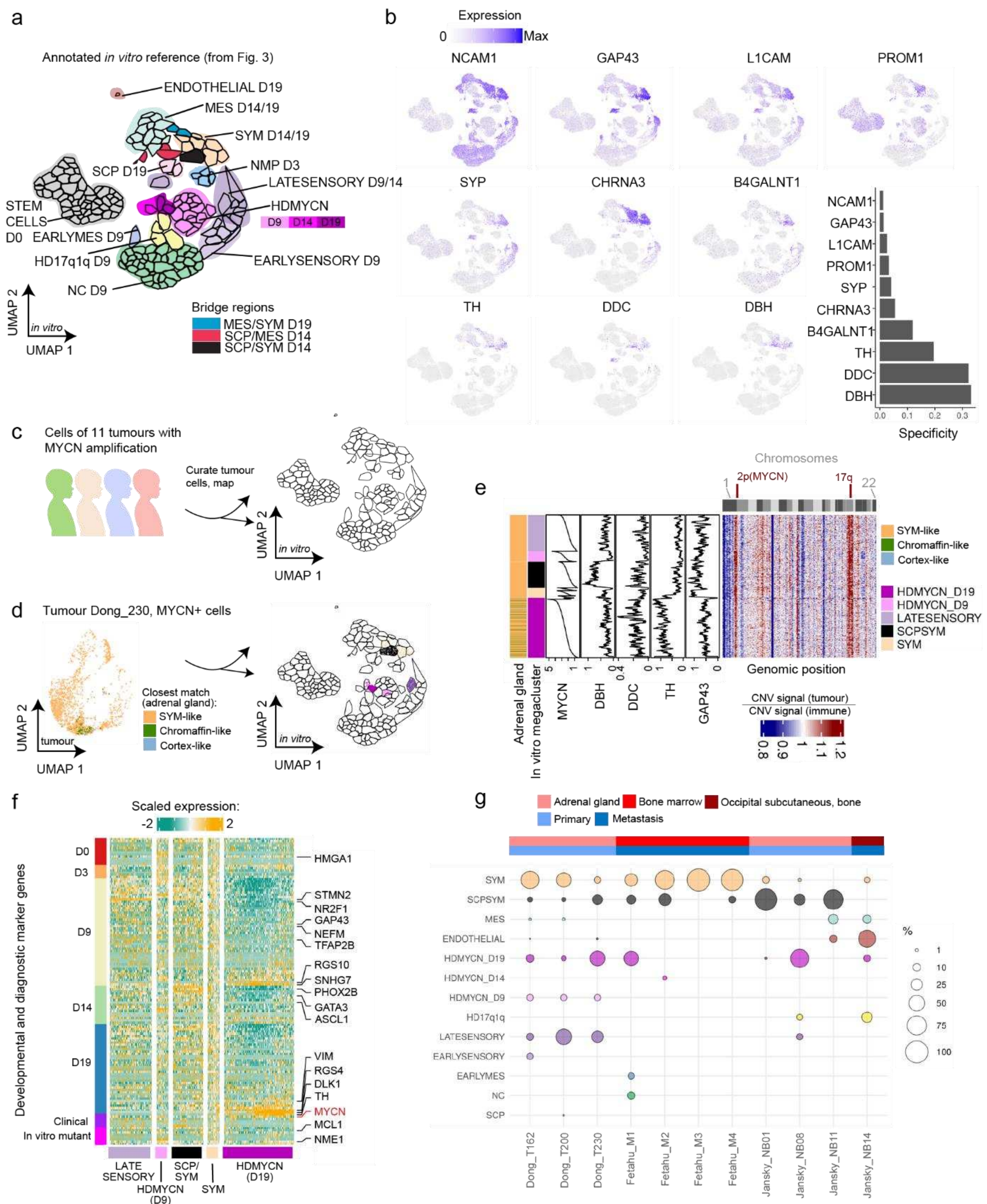
1252
1253
1254 **Figure 4. Impaired trunk NC specification correlates with acquisition of tumourigenic hallmarks.**
1255 **a)** Flow cytometric analysis of cell cycle in D9 cultures obtained from hESCs marked by the
1256 indicated NB-associated lesions. Top: Representative FACS plots. Bottom: Percentage of cells
1257 found in each of the different stages of the cell cycle (G1, S, G2) corresponding to indicated
1258 NB-associated lesions (n = 3 biological replicates, error bars= standard deviation two-way
1259 ANOVA). P values in comparisons: G1 (17q vs 17q1qMYCN, p = 0.0012 = **; 17q1q vs

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1260 17q1qMYCN, p= 0.0004=***), S (17q vs 17q1qMYCN, p = 0.0008 = ***; 17q1q vs
1261 17q1qMYCN, p= 0.0001=***)
1262 **b)** Immunofluorescence analysis (green) of the expression of the cell proliferation marker KI-67
1263 in D9 (top) and D14 (bottom) cultures obtained from hESCs marked by the indicated NB-
1264 associated lesions. Cell nuclei were counterstained using Hoechst 33342 (blue). Scoring of the
1265 percentages of KI-67-positive cells is also shown (n = 3 biological replicates, error bars=
1266 standard deviation, ordinary one-way ANOVA with Tukey correction). P values in
1267 comparisons: D9 (WT vs. 17q p= 0.0072= **; 17q vs 17q1qMYCN p= 0.0005=***; 17q1q vs.
1268 17q1Qmycn p= 0.0125=*), D14 (WT vs 17q p=0.0255=*; WT vs. 17q1qMYCN p= 0.0034=**;
1269 17q vs. 17q1qMYCN p<0.0001=****; 17q1q vs. 17q1qMYCN p= 0.0001=***)
1270 **c)** Top: Representative brightfield images of cell/colony morphology following a low-density
1271 plating assay using cells marked by the indicated NB-associated lesions after 84 hours. Bottom:
1272 Comparison of the number of colonies formed by cells marked by the indicated NB-associated
1273 lesions following plating at low density. (n= 3 biological replicates, error bars= SD, Ordinary
1274 One-way ANOVA test with Tukey correction). P values in comparisons: WT vs. 17q1qMYCN
1275 p= 0.0278=*; 17q vs. 17q1qMYCN p= 0.022=*; 17q1q vs. 17q1qMYCN p= 0.0421=*. The
1276 fold change per experiment was calculated for 17q1qMYCN with 17q1q acting as the control.
1277 (n= 3 biological replicates, error bars= SD, Two-tailed unpaired t test. P value = 0.0041=**)
1278 **d)** Representative images of zebrafish xenografted with WT or 17q1qMYCN cells labelled with
1279 CellTrace Violet at 1 day post injection (dpi). Quantification of the area covered by WT or
1280 17q1qMYCN cells in zebrafish xenografts at 1 and 3 dpi. While 17q1qMYCN cells persist, WT
1281 cells are not maintained in zebrafish xenografts. Xenografts with WT cells (n = 11), and
1282 17q1qMYCN cells (n = 51).

1283 **Abbreviations**: p ≤ 0.05= *, p ≤ 0.01= **, p ≤ 0.001= ***, p ≤ 0.0001= ****.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1284
1285
1286 **Figure 5. Comparison to hESC-based trunk NC differentiation resolves structured heterogeneity**
1287 **across neuroblastoma tumours**

1288 **a)** UMAP of the integrated trunk NC developmental reference, divided into high-resolution
1289 clusters that resolve rare subpopulations. These clusters were grouped into interpretable "mega-
1290 clusters" that shared characteristics like developmental stage, condition, and closest-matching
1291 human adrenal gland cell type (**Fig. S7**).

1292 **b)** UMAPs as in **panel (a)** showing expression of widely used diagnostic markers for NB. Markers
1293 were ranked from the least specific (appearing in many regions of the UMAP) to the most
1294 specific (limited to few clusters) according to a specificity score (bar plot, bottom right). *DBH*,
1295 *DDC* and *TH* score as highly specific to the SCP/SYM region of the dataset.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1296 **c)** Schematic overview of the analysis of tumour cells. We curated *MYCN*+ cancer cells from 11
1297 MYCN-amplified NB tumour samples[15,17] from three studies and mapped them onto our
1298 reference (cp. **panel (a)**) using Azimuth[93]. Mapping is represented as tumour cells falling into
1299 sectors of the *in vitro* reference (depicted as convex hulls of each cluster in the glasswork plot).
1300 **d)** Low-dimensionality UMAP projection of the *MYCN*+ cells of tumour *Dong_T230*[17] coloured
1301 by their closest matching human adrenal gland cell type, showing continuous groups of SYM-
1302 like cells. Mapping to the *in vitro* trunk NC dataset resolves classification into diverse subtypes
1303 (see following panels).
1304 **e)** Heatmap depicting gene expression in *MYCN*+ tumour cells of dataset *Dong_T230* from **panel**
1305 **(d).** Values are inferCNV[94] copy number estimations per gene, relative to hematopoietic and
1306 immune cells in the sample ordered by genomic position and chromosome (1-22). Cells (one
1307 per row) are shown ordered by mega-cluster and Louvain cluster, therein ordered by *MYCN*
1308 levels. Annotations (left to right): closest *in-vivo* adrenal gland cell type[16], *in vitro* mega-cluster,
1309 and a sliding-window moving average (w=20 cells) of depth-normalised levels of selected
1310 diagnostic NB markers. Mappings of other tumours datasets are shown in **Fig. S8**.
1311 **f)** Heatmap displaying expression of *MYCN*+ tumour cells for selected differentiation markers.
1312 Genes were selected to include highly variable markers (ordered by day, D0-D19, from **Fig.**
1313 **1d, Table S2**), mutant cluster markers (“invitro mutant”, **Table S6**), and diagnostic NB markers
1314 (“clinical”, **panel (b)**). Heatmap columns are *Dong_T230* cells[17] from **panels (c,d)** seriated
1315 within mega-clusters.
1316 **g)** Bubble plots showing the relative percentage (bubble size) of high-confidence mappings
1317 (prediction score >=0.6) of MYCN+ tumour cells onto each mega-cluster for cells from 11
1318 tumour datasets, processed and curated as described in **panel (c)**. Tumour sample covariates
1319 (tissue of origin (red/brown) and sample type (blue/dark blue)) are depicted in the annotation
1320 above.

1321 **Abbreviations**: WT, wild-type H7 hESCs; UMAP, Uniform Manifold Approximation and Projection;
1322 SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal; M1-M25, “mutant” cell
1323 clusters.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

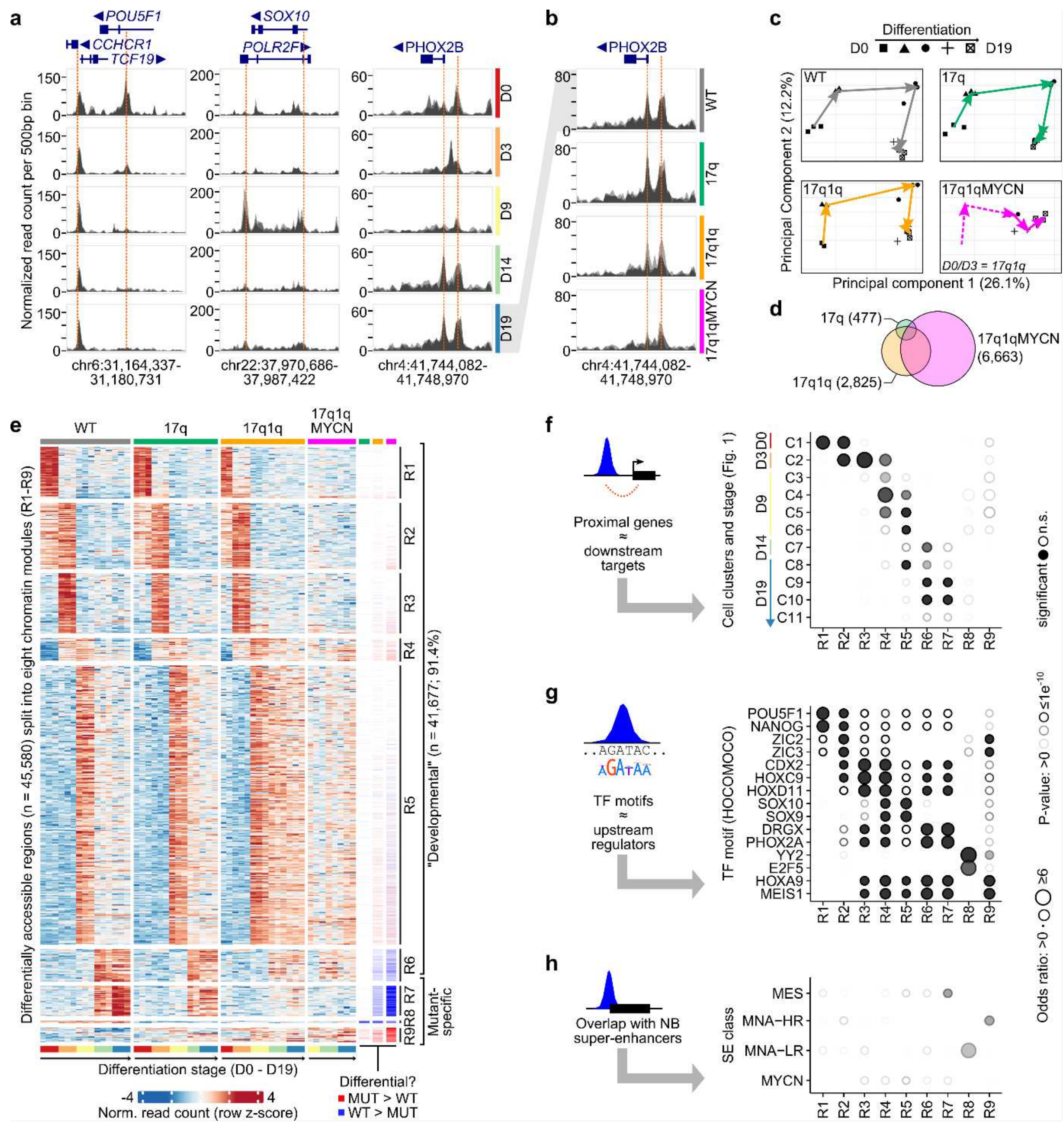


Figure 6. Differentiation of wild-type and mutant hESCs is associated with epigenetic changes in nine distinct chromatin modules.

- **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]; $P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq \log_2(1.5)$) in mutant hESCs compared to WT-hESCs. Numbers indicate the total number of regions per cell line aggregated over all developmental stages.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1340 **e)** Heatmaps showing normalised read counts for all differentially accessible regions (columns)
1341 in any pairwise comparison of two stages or conditions (*DEseq2*[95]; $P_{adj} \leq 0.005$,
1342 $|\log_2 FoldChange| \geq \log_2(1.5)$; $n_{total} = 45{,}580$). Regions have been divided into nine non-
1343 overlapping modules (R1–R9) by hierarchical clustering. Three annotation columns are shown
1344 to the right indicating regions called down- (blue) and up-regulated (red) in each mutant hESC.
1345 All regions and differential analysis results are reported in **Tables S8 and S9**.
1346 **f)** Enrichment analysis of co-localisation of regions belonging to the nine chromatin modules
1347 (from left to right; cp. **panel (e)**) and nearby genes identified as markers of differentiating cell
1348 populations in our scRNA-seq analysis (cp. **Fig. 1b**). The size and transparency of circles
1349 indicates the odds ratio and P-value, respectively (hypergeometric test, *hypeR*[92]). Significant
1350 results are indicated with filled circles ($P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq \log_2(1.5)$, frequency ≥
1351 2.5%). All results are shown in the figure and also reported in **Table S10**.
1352 **g)** Enrichment analysis for overlaps between chromatin modules and known TF motifs
1353 (*HOCOMOCO* database[96], v11). The plots are as in **panel (f)**, with the exception that only
1354 overlaps with $P_{adj} \leq 0.0000001$, $|\log_2 FoldChange| \geq \log_2(2)$, and frequency ≥ 2.5% were marked
1355 as significant. The top results per module are shown and all results are reported in **Table S10**.
1356 **h)** Enrichment analysis of overlaps between regions belonging to the nine chromatin modules and
1357 super-enhancers specific to certain NB subgroups[68]. Plots as in **panel (f)**.

1358 **Abbreviations**: D0/3/9/14/19, day 0/3/9/14/19; WT, wild-type H7 hESCs; MUT, a "mutant" hESC line
1359 (17q, 17q1q, or 17q1qMYCN); R1-R9, chromatin region modules; NMP, neuromesodermal-potent
1360 axial progenitors; SYM, sympathoblast; sig., significant.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

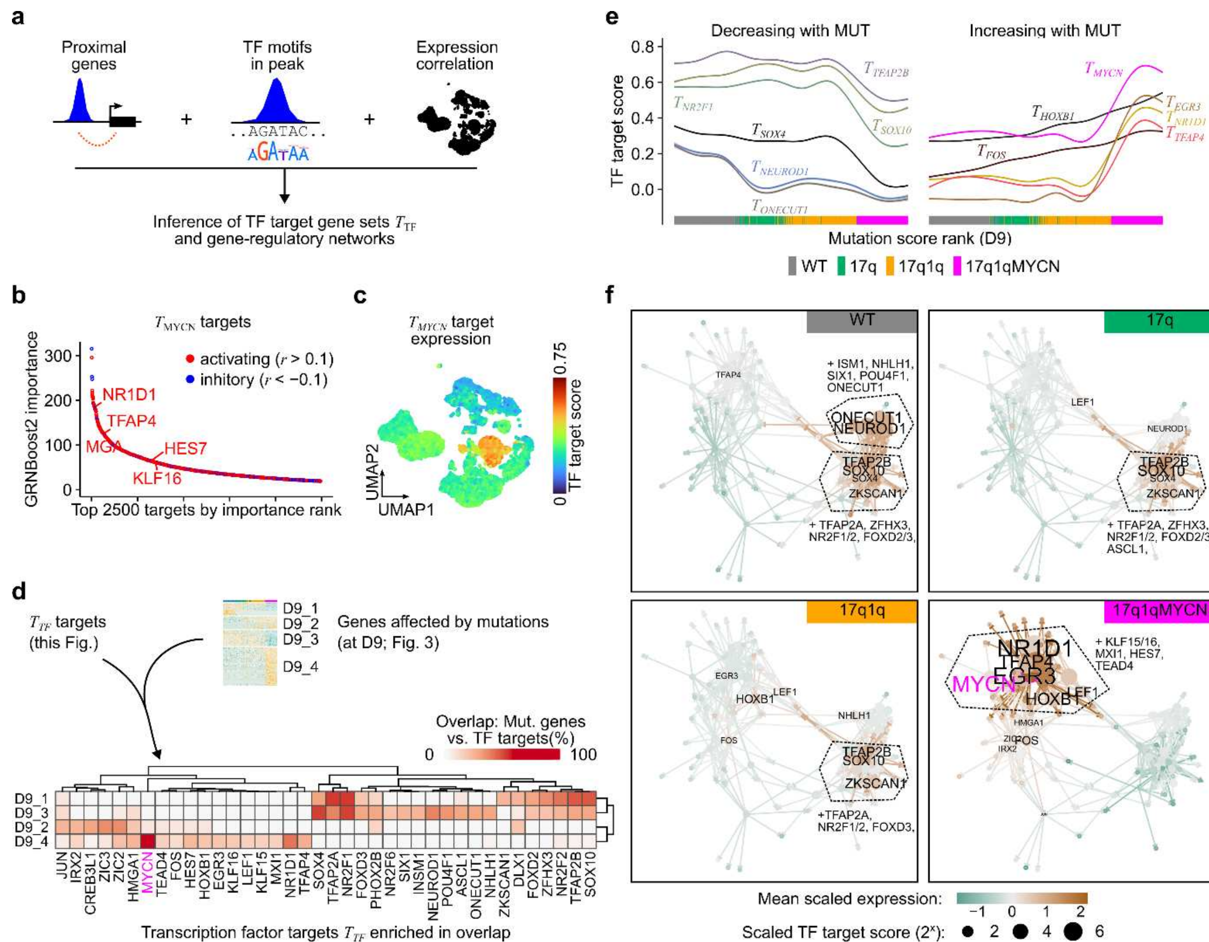


Figure 7. Copy number changes facilitate *MYCN*-mediated blockage of differentiation via developmental transcription factor networks.

- **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]; $P_{adj} \leq 0.0005$, $|\log_2 FoldChange| \geq \log_2(2)$, 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

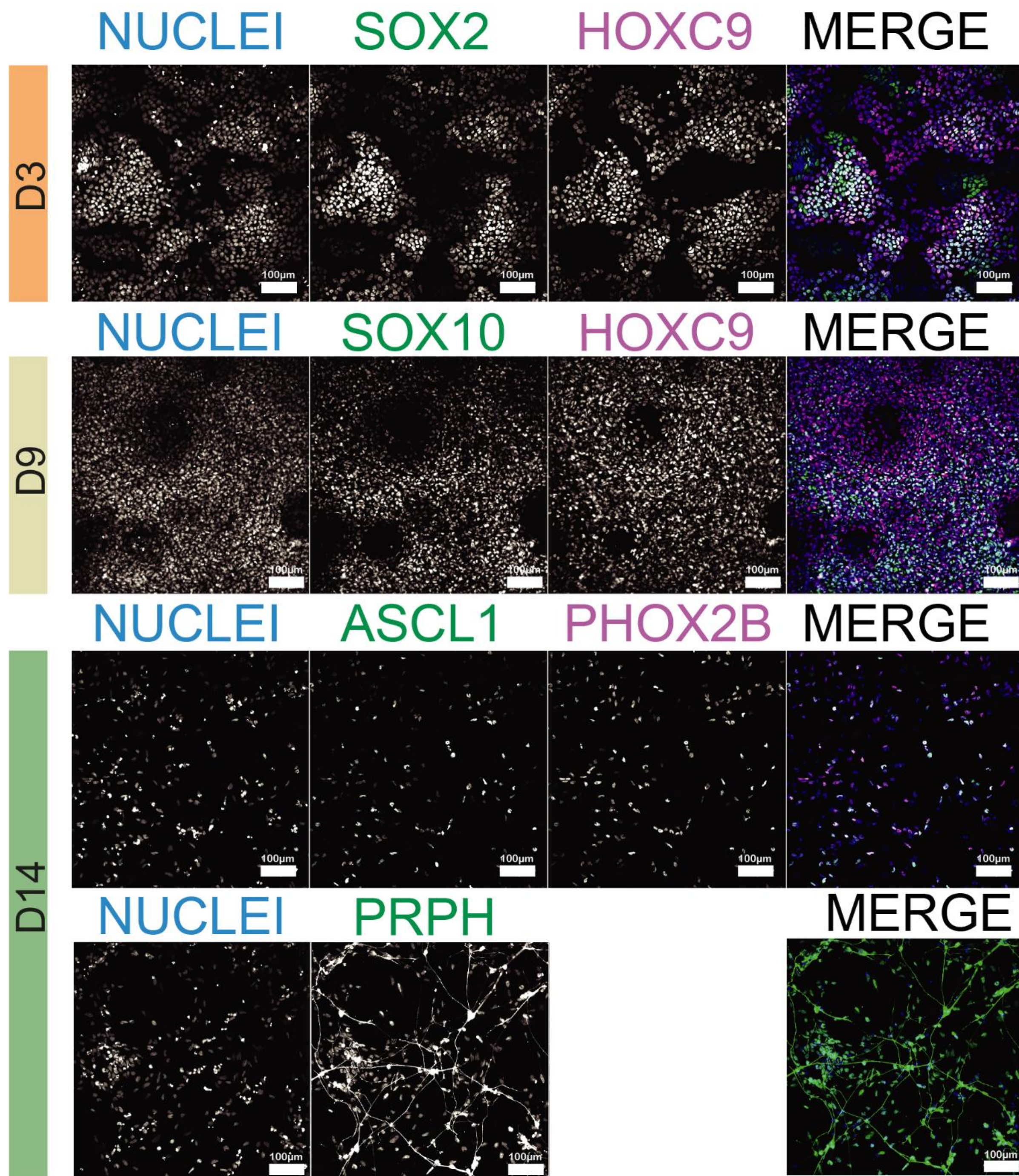
1388 these diagrams, each node represents a TF or target gene, and each edge is a link between a TF
1389 and a target. We made these networks specific to cells from each condition (WT, 17q, 17q1q,
1390 17q1qMYCN) by using colour to indicate the mean scaled expression of each gene in the
1391 respective cells at D9 (edges coloured by source TF) and node size to indicate the mean scaled
1392 T_{TF} target score of each TF. Only labels of TFs with positive scaled expression are shown and
1393 selected groups of TFs have been merged for visualisation. All network node labels are shown
1394 in **Fig. S10c**.

1395 **Abbreviations**: D0/3/9/14/19, day 0/3/9/14/19; R1-R9, chromatin region modules; TF, transcription
1396 factor; WT, wild-type H7 hESCs; MUT, a "mutant" hESC line (one of: 17q, 17q1q, 17q1qMYCN);
1397 sig., significant; r, Pearson correlation coefficient; act., activating (positive correction); inh., inhibitory
1398 (negative correlation).

1399

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

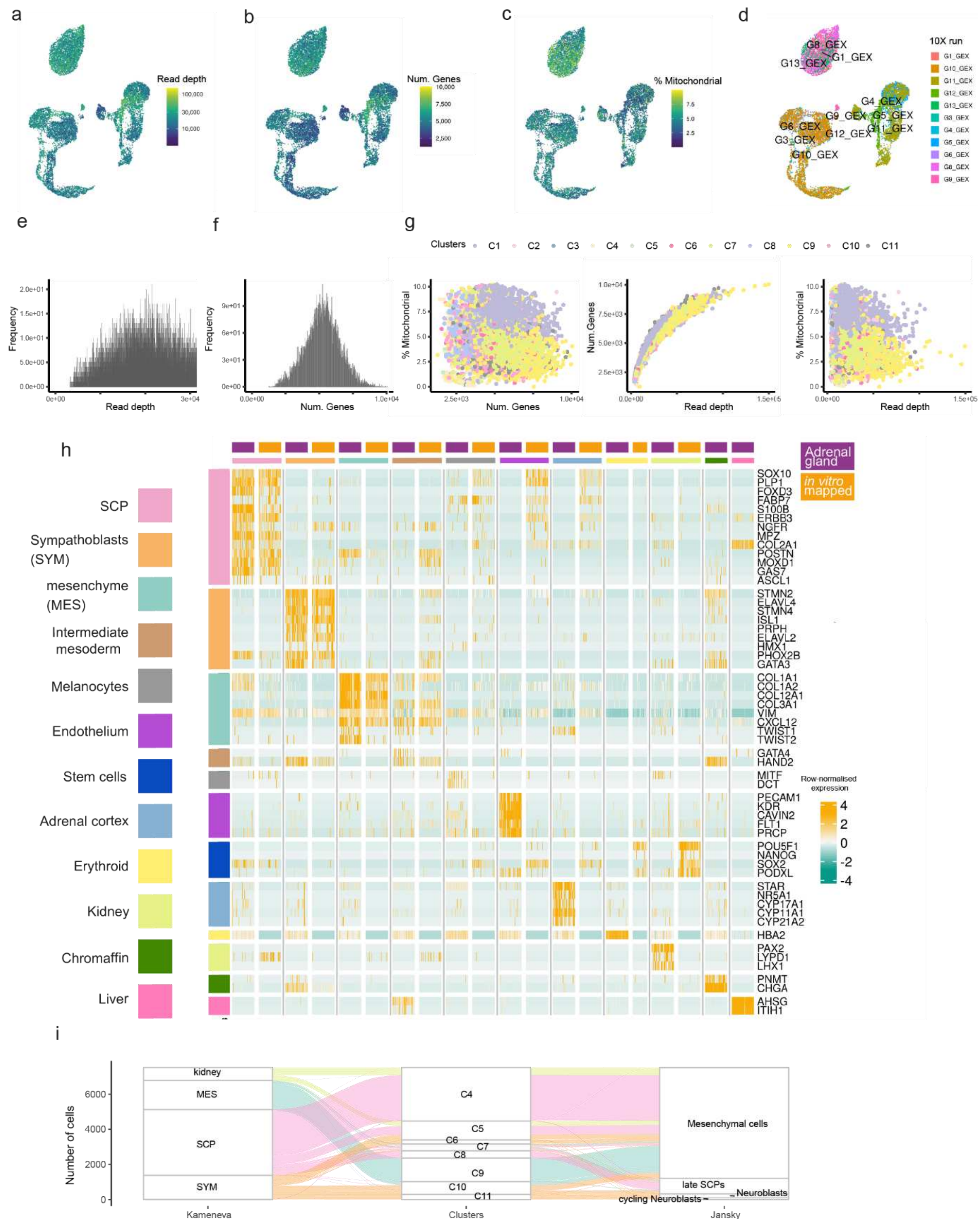
1400 ## Supplemental figures



1401
1402
1403 **Supplemental Figure S1 (related to Fig. 1). Immunofluorescence analysis of hESCs during trunk**
1404 **NC differentiation.**
1405 Immunofluorescence analysis of the expression of indicated markers at different time points during the
1406 differentiation of hESCs toward trunk NC and its derivatives.
1407 **Abbreviations**: D3/9/14, day 0/3/9/14.

1408

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1409

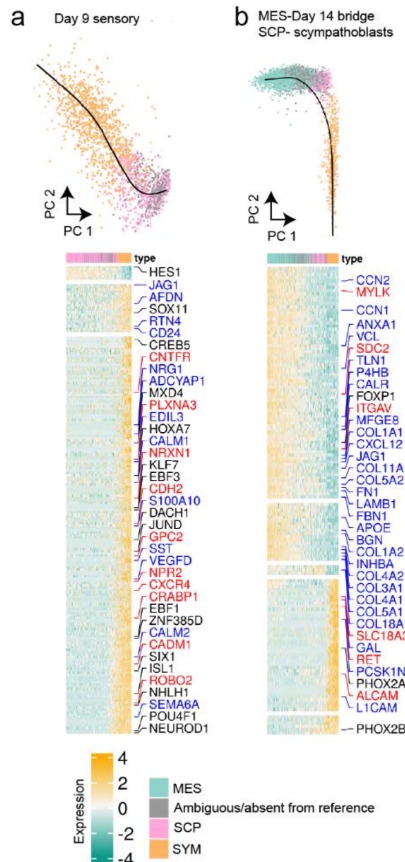
1410

1411 **Supplemental Figure S2 (related to Fig. 1). Quality control and reference mapping of single-cell**
1412 **RNA-seq data from wild-type hESC trunk neural crest differentiation.**
1413 **a-d)** UMAP plots showing quality covariates for the wild-type hESC dataset in **Figure 1**.
1414 **e-f)** Histograms depicting the distribution of read depth (**panel (e)**) and number of genes detected
1415 (**panel (f)**) per cell barcode after quality control filtering.
1416 **g)** Scatterplots comparing all quality control covariates (shown on panels (**a-c**)) from the same cell,
1417 plotted versus each other, coloured by wild-type clusters (C1-C11).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1418 **h)** Side-to-side comparison of cell type marker expression in 200 annotated cells selected at random
1419 from the human adrenal gland reference[16] versus the top 200 high-confidence cells mapped to
1420 the same cell types in our WT *in vitro* trunk NC dataset. Rows are cell-type marker genes.
1421 Columns are cells first divided by cell type (separated with a grey line), then by dataset of origin
1422 (adrenal gland: purple, in vitro: orange). Expression values are depth-normalised per
1423 experiment and row-scaled globally. Known stem cell markers were added to trace where the
1424 stem cell population would be spuriously mapped to, in this case kidney, an indicator that cells
1425 not found in the reference may be mapped to kidney. No *in vitro* cells were mapped to
1426 chromaffin or liver identities, leading to the absence of the respective *in vitro* columns. Cells
1427 erroneously mapped, absent from the reference, or lacking relevant cell type markers were
1428 classified as "other" and coloured grey in **Figure 1e**.
1429 **i)** Alluvial plots comparing the mappings between cells in the *in vitro* dataset compared to two
1430 adrenal gland reference datasets[15,16]. Each "stream" indicates a group of cells that were mapped
1431 consistently to one cell identity in the Kameneva *et al.* reference (also indicated in colour). For
1432 example, cells that were labelled MES in the paper (cluster C9) also mapped to mesenchymal
1433 cells in the Jansky *et al.* reference, cells that mapped to SYM (clusters C10, C11) mapped to
1434 cycling neuroblasts and to neuroblasts. Cells that we labelled as SCPs split into cells that
1435 mapped to mesenchymal cells (C4, C5) and late SCPs (C8) in Jansky *et al.*, consistent with our
1436 observations that the former may represent a less mature, early SCP state (see main text).
1437
1438 **Abbreviations**: D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold Approximation and
1439 Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES, mesenchymal.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1440

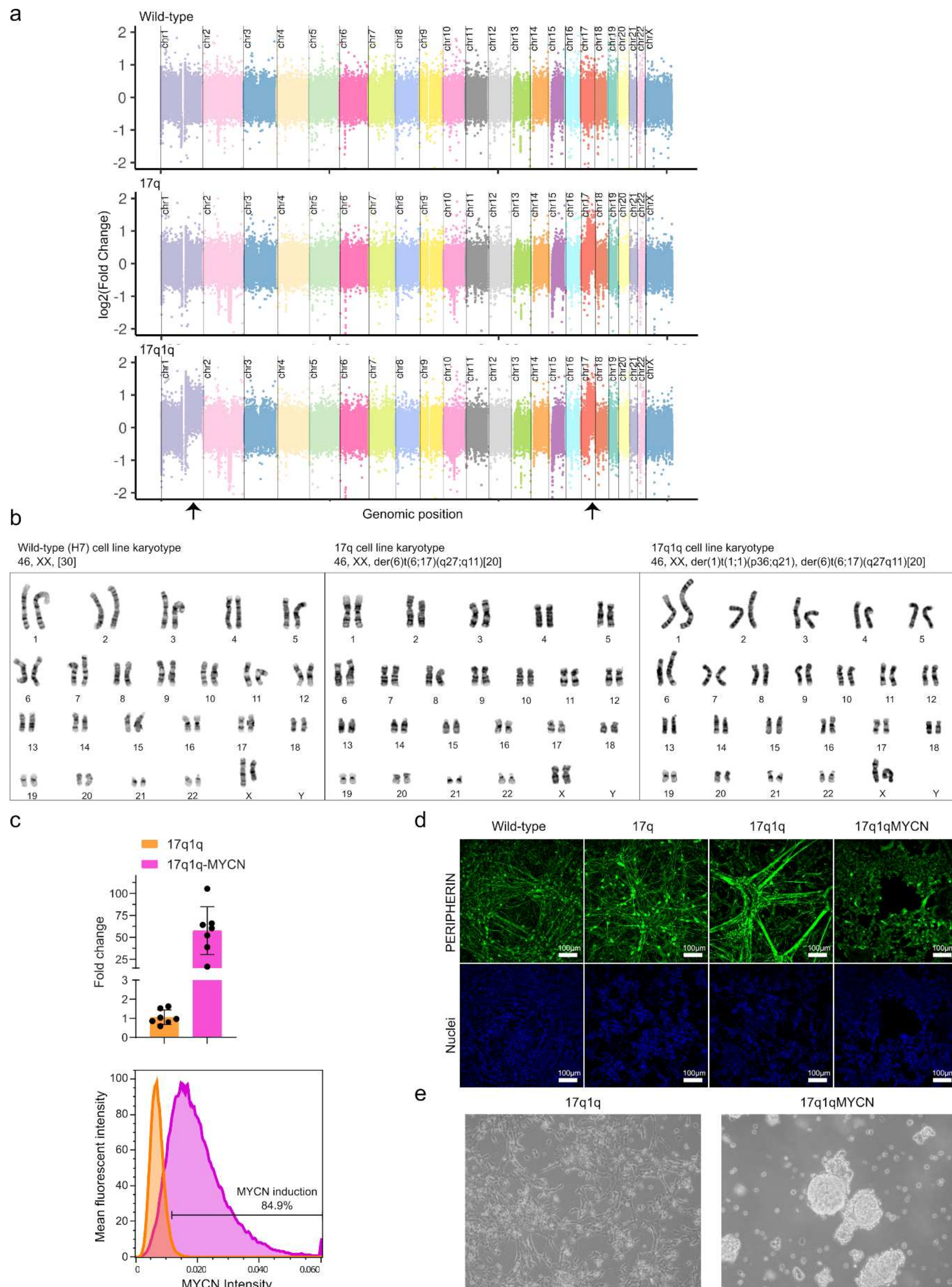
1441

1442 **Supplemental Figure S3 (related to Fig. 1). Trajectories connecting gradients of transcriptionally**
1443 **similar cells at different developmental stages.**

1444 *Slingshot*[90] pseudotime trajectories (top) for wild-type clusters C5-C6 (**panel (a)**) and C7, C8, C9, and
1445 C11 (**panel (b)**). Cells were separated from the main dataset and reprocessed (see basic scRNA-seq
1446 processing in Methods), and trajectories were calculated on the first two principal components with C5
1447 cells and D14 cells (b), respectively, declared as start of the trajectory. Heatmaps (bottom) show the top
1448 140 genes with the strongest association with the trajectory as ranked by *tradeSeq*'s Wald test[97].
1449 Highlighted genes are all TFs (black), receptors (red) and ligands (blue) found in the association test.

1450 **Abbreviations**: PC, principal component; SCP, Schwann cell progenitor; SYM, sympathoblast; MES,
1451 mesenchymal; TF, transcription factor.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



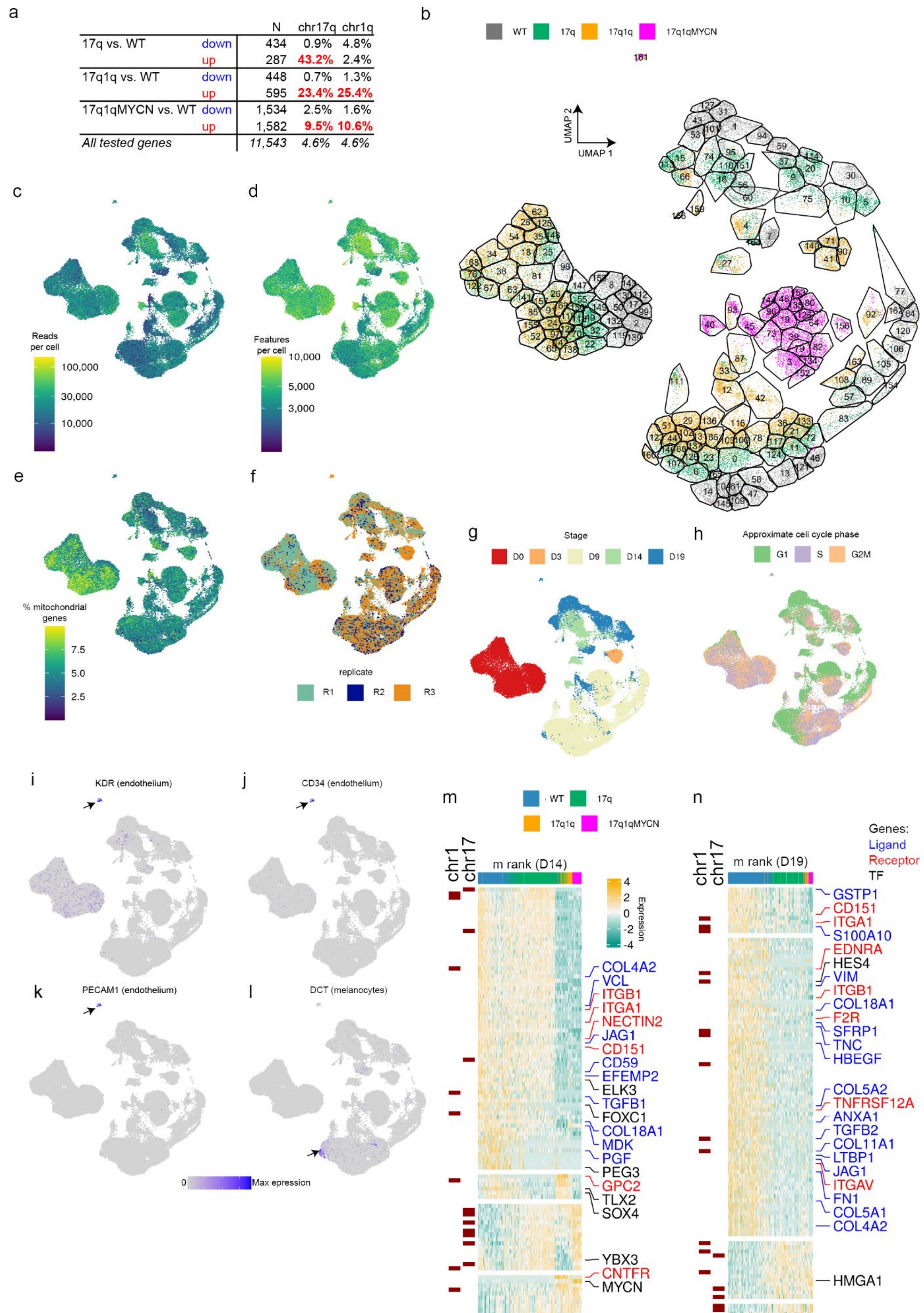
1452
1453
1454 **Supplemental Figure S4 (related to Fig. 2). Genetic and phenotypic characterization of mutant**
1455 **hESC lines.**
1456 **a)** Plots of the output of CNVkit[98] depicting $\log_2$ fold change of whole-exome sequencing reads
1457 relative to the genome average. CNVs can be seen for chr1q and chr17q.
1458 **b)** Cytogenetic analysis of indicated hESC lines used in the study.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1459 **c)** Analysis of MYCN expression at the transcript (top) and protein (bottom) level in D9
1460 17q1qMYCN cultures after Doxycycline treatment at day 5 vs untreated control.
1461 **d)** Immunofluorescence analysis of PERIPHERIN expression in D19 cultures following
1462 differentiation of hESCs with the indicated genotypes. Cell nuclei were counterstained using
1463 Hoechst 33342.
1464 **e)** Representative brightfield images of D14 cultures following differentiation of hESCs with the
1465 indicated genotypes.

1466 **Abbreviations**: WT, wild-type H7 hESCs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



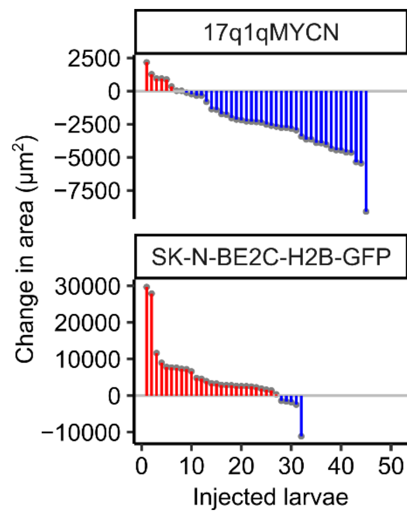
1467
1468
1469 **Supplemental Figure S5 (related to Fig. 3). scRNA-seq analysis of differentiating wild-type and**
1470 **mutant hESCs.**

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1471 **a)** Overview of the number of differentially expressed genes (DEGs) in 17q, 17q1q, and
1472 17q1qMYCN cells at D9 of differentiation compared to wild-type. The total number of DEGs
1473 is given (N), and the percentage of those genes that are located on chromosome arms chr17q or
1474 chr1q are indicated. Percentage values >5% have been highlighted (which also correspond to
1475 upregulated DEGs within known CNVs).
1476 **b)** High-resolution cell clusters defined for the full *in vitro* TNC dataset scRNA-seq dataset
1477 generated in this study. Cluster marker genes are reported in **Table S6**.
1478 **c-h)** QC covariate plots: reads per cell (**panel (c)**), features per cell (**panel (d)**), % mitochondrial
1479 genes (**panel (e)**), replicates (one of up to three repeat experiments; **panel (f)**), developmental
1480 stage (sampling day; **panel (g)**), and computationally inferred cell cycle stage (**panel (h)**).
1481 **i-l)** Visualisations of endothelial (*KDR*, *CD34*, *PECAM1*) and melanocyte (*DCT*) gene expression
1482 across the full *in vitro* UMAP. Arrows highlight the small, high-intensity cluster 161 which
1483 expresses all endothelial markers, and cluster 160 which mapped to SCPs, but expressed the
1484 melanocyte marker.
1485 **m, n)** Heatmaps containing the genes correlated or anti-correlated with the mutation score *m* and
1486 related measures (see Methods) for D14 (**panel (m)**) and D19 (**panel (n)**). Transcription factors
1487 (black), receptors (red) and ligands (blue) have been highlighted.

1488 **Abbreviations**: WT, wild-type H7 hESCs; D0/3/9/14/19, day 0/3/9/14/19; UMAP, Uniform Manifold
1489 Approximation and Projection; SCP, Schwann cell progenitor; SYM, sympathoblast; MES,
1490 mesenchymal; m rank, mutation score rank.

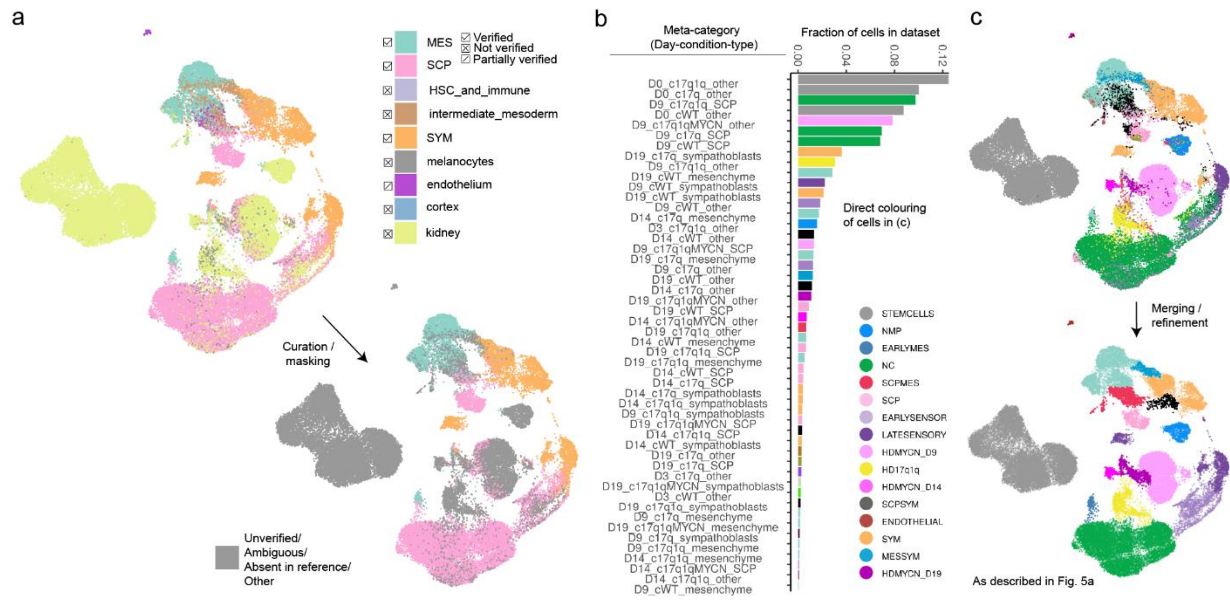
bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1491
1492
1493 **Supplemental Figure S6 (related to Fig. 4). Survival of xenotransplanted 17q1qMYCN cells and**
1494 **an NB cell line in zebrafish larvae.**
1495 Waterfall plots depicting the change in tumour area for 17q1qMYCN at D9 of differentiation (~ NC
1496 stage) and SK-N-BE2C-H2B-GFP[66] cells in zebrafish xenografts from 1dpi to 3dpi.

1497 **Abbreviations**: dpi, day post injection.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



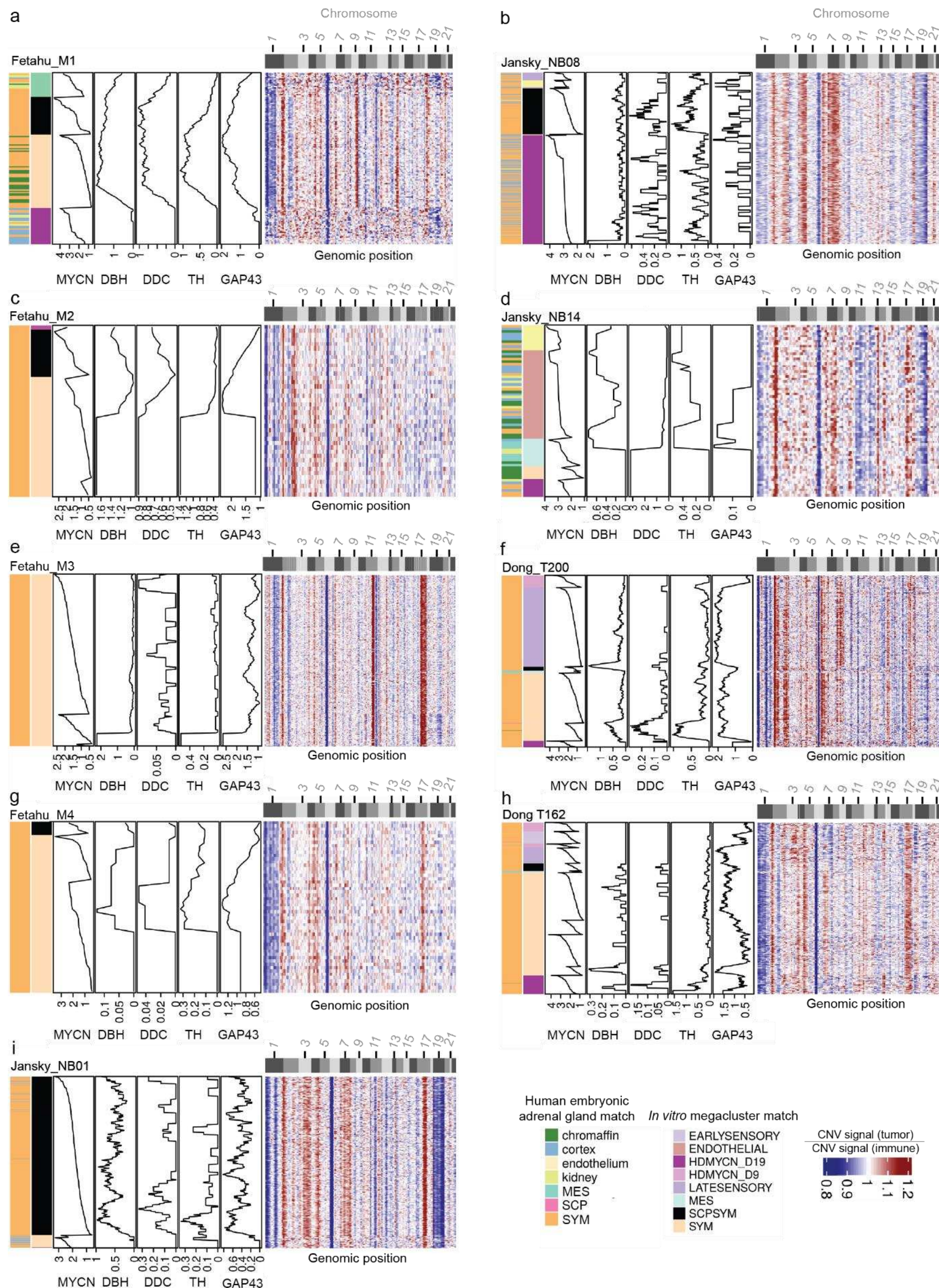
1498
1499

1500 **Supplemental Figure S7 (related to Fig. 5). Classification of cells in mega-clusters.**

1501 Cells of the full *in vitro* trunk NC dataset (UMAP shown in **panel (a)**) containing all the integrated
1502 conditions (WT, 17q, 17q1q, and 17q1qMYCN) just as in Fig. 3d were classified via label transfer[93]
1503 using the human embryonic adrenal gland cell reference[16]. The classification was then curated and
1504 verified via expression of relevant cell type markers. Cells with markers of the matching types were
1505 classified as verified (panel (a), boxes with a tick). Labelled cells that did not express markers of the
1506 matching cell type, or well-defined cell types were clearly misclassified (such as stem cells classified
1507 as kidney), or cell types in neighbourhoods of highly mixed matches, or unlabelled cells were declared
1508 as not verified (panel (a), crossed boxes). Cells matching to one type, which expressed the markers in
1509 one UMAP region but not in other (such as endothelial cells) were classified as partially verified (panel
1510 (a), tickboxes with diagonal lines) All unverified cells were henceforth classified as other. Using the
1511 curated cell labels ("type"), we combined information from each cell's stage, condition, and type and
1512 classified cells into meta-categories (**panel (b)**) revealing a wide distribution of meta-category
1513 frequencies. We coloured the meta-categories as described in **panel (b)**, by merging clusters
1514 heuristically when changing the condition and stage did not alter a defined (non-other) type. Groups of
1515 several clusters (**panel (c)**, top) were refined by incorporating nearby satellite cells from other meta-
1516 categories and merging clusters with the same meta-category to obtain the mega-clusters shown at the
1517 bottom of **panel (c)**, and in figure **Fig. 5a**.

1518 **Abbreviations**: SCP, Schwann cell precursors; HSC, hematopoietic stem cells; SYM, sympathoblasts;
1519 UMAP, Uniform Manifold Approximation and Projection; WT, wild-type.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



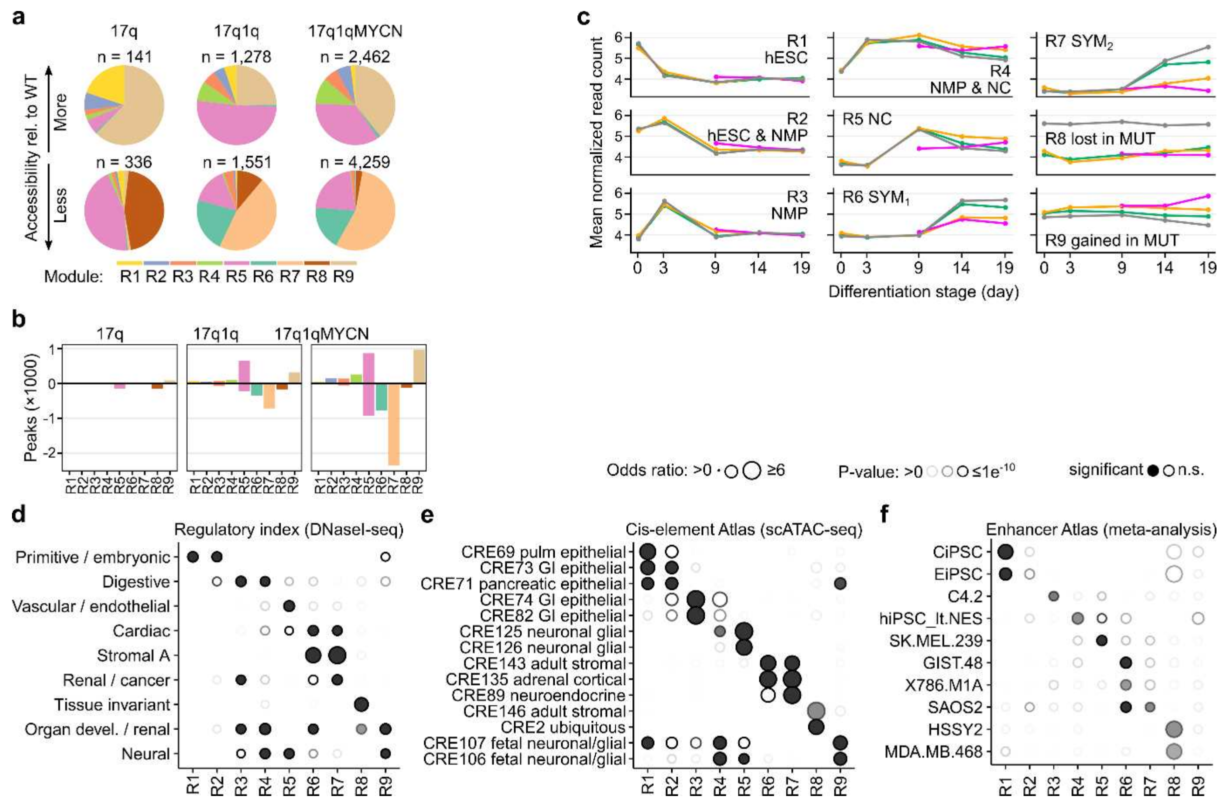
1520
1521

1522 **Supplemental Figure S8 (related to Fig. 5). Analysis of *MYCN*-amplified tumour cells and**
1523 **mapping to *in vitro* trunk NC differentiation.**
1524 *inferCNV*[94] profile heatmaps such as the one in **Fig. 5e** for the remaining 10 tumour datasets[15,17] not
1525 shown in **Fig. 5**. Each row (*MYCN*+ tumour cells) and each column (genes, ordered by genomic
1526 position), indicate the intensity of the CNV signal relative to non-tumour, HSC/immune cells from the

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1527 same sample. All samples were curated and processed as described in **Fig. 5c-e** and mapped both to the
1528 human embryonic adrenal gland reference[16] and the full *in vitro* trunk NC differentiation reference
1529 (**Figs. 5a**, **S5b, S7**). Cells are ordered first by matching mega-cluster, then by matching Louvain cluster
1530 (visualised in **Fig. S5b**), and then by *MYCN* levels within each cluster. Annotations are (left to right):
1531 matching adrenal gland cell type, matching mega-cluster, *MYCN*, *DBH*, *DDC*, *TH*, and *GAP43* levels,
1532 respectively. Annotation of the chromosomes can be found on top of the heatmap of tumour dataset
1533 *Fetahu_M1* and other annotations in the supporting panel.
1534 **Abbreviations**: CNV, copy number variant; NC, trunk neural crest.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1535
1536

1537 **Supplemental Figure S9 (related to Fig. 6). Chromatin accessibility in differentiating wild-type**
1538 **and mutant hESCs.**

1539 **a)** Pie charts indicating the fraction of differentially accessible regions belonging to chromatin
1540 modules (**Fig. 5b**) in mutant hESCs compared to WT (left to right). Up- and down-regulated
1541 regions are shown separately (top vs. bottom row). The total number (n) of regions in each
1542 category is indicated.

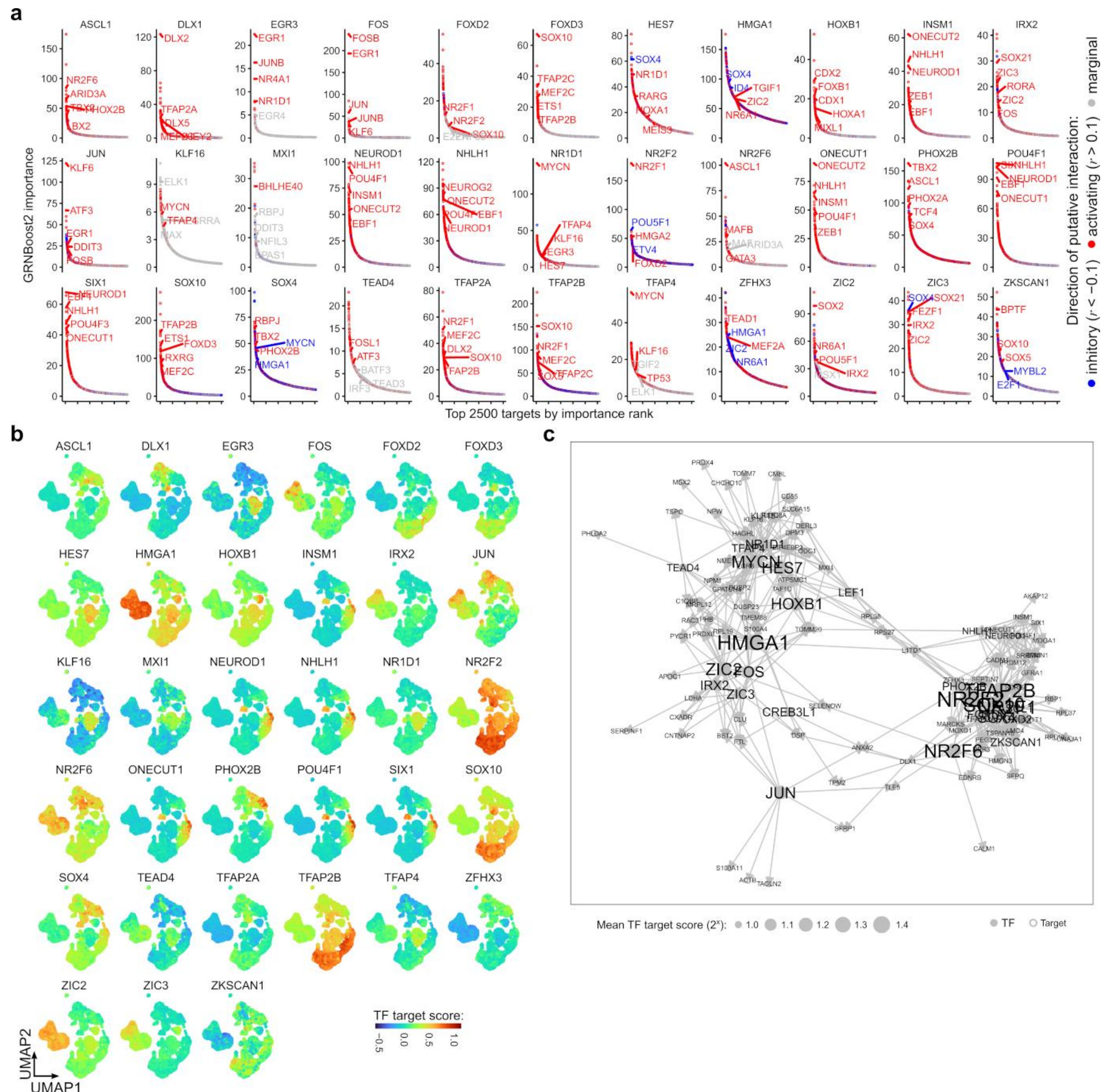
1543 **b)** Barplots indicating the number of up- (positive numbers) and down-regulated (negative
1544 numbers) peaks from each comparison of mutant hESCs vs WT, split by chromatin module
1545 (R1-R9).

1546 **c)** Line plots summarizing the dynamics of accessibility per module and cell line throughout
1547 differentiation. Each data point indicates the mean normalised read count.

1548 **d-f)** Enrichment analysis of overlaps between regions belonging to the nine chromatin modules
1549 (from left to right) and annotated reference regions from the Regulatory Index[73] (based on
1550 DNaseI-seq; **panel (d)**), Cis-element Atlas[74] (based on scATAC-seq analysis; **panel (e)**) and
1551 the Enhancer Atlas[75] (based on a meta-analysis of many different data; **panel (f)**). The size and
1552 transparency of circles indicate the odds ratio and P-value, respectively (hypergeometric test,
1553 *hypeR*[92]). Significant results are indicated with filled circles ($P_{adj} \leq 0.005$, $|\log_2 FoldChange| \geq$
1554 $\log_2(1.5)$, frequency $\geq 2.5\%$). The top enrichments per stage have been selected for visualization
1555 (all results are shown in panel **e**) and all results are reported in **Table S10**.

1556 **Abbreviations**: WT, wild-type H7 hESCs; R1-R9, chromatin modules identified in **Fig. 6e**; n.s., not
1557 significant.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



1558
1559
1560 **Supplemental Figure S10 (related to Fig. 7). Analysis of transcription factor target sets and gene-**
1561 **regulatory networks.**
1562 **a)** Top 2500 targets of selected TFs as predicted by *GRNboost2* algorithm[76] based on our scRNA-
1563 seq data. Putative targets without support in our ATAC-seq data (motif for TF in ≥1 peak near
1564 the gene) have been removed. We also calculated the Pearson correlation coefficient (*r*)between
1565 each TF and target gene to determine the direction of the putative interaction (*r*>0.1,
1566 "activating"; *r*<-0.1, "inhibitory"; others, "marginal"). The top TFs in the target lists have been
1567 highlighted. TF target gene sets are reported in **Table S11**.
1568 **b)** Average expression (Seurat module score) of the target gene sets (matching "activating" targets
1569 of the TFs in panel **(a)** in our integrated scRNA-seq dataset (cp. **Fig. 3d**).
1570 **c)** Gene-regulatory networks diagram visualizing putative TF to target interactions for the genes
1571 in gene sets D9_1 to D9_4 (cp. **Fig. 3e,f**) and enriched TF targets (cp. **panels (a,b)**). In these
1572 diagrams, each node represents a TF or target gene, and each edge is a link between a TF and
1573 a target. Node size is proportional to the mean target score of the indicated TFs (fixed size for
1574 non-TF nodes).
1575 **Abbreviations**: TF, transcription factor; r, Pearson correlation coefficient; D14/19, day 14/19.
1576

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.21.515753; this version posted November 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1577 # Supplemental tables and files

1578

1579 **Supplemental Table S1 (related to Figs. 1, 2, 3, 6, 7). scRNA-seq, RNA-seq, and ATAC-seq**
1580 **dataset overview**
1581 **Supplemental Table S2 (related to Fig. 1). scRNA-seq cluster marker genes (WT)**
1582 **Supplemental Table S3 (related to Fig. 1). scRNA-seq markers of SCP-SYM-MES transition**
1583 **states**
1584 **Supplemental Table S4 (related to Fig. 3). scRNA-seq MUT vs. WT differentially expressed**
1585 **genes**
1586 **Supplemental Table S5 (related to Fig. 3). scRNA-seq MUT vs. WT enrichment results**
1587 **Supplemental Table S6 (related to Fig. 3). scRNA-seq cluster marker genes (WT+MUT)**
1588 **Supplemental Table S7 (related to Fig. 3). scRNA-seq genes correlated to mutations**
1589 **Supplemental Table S8 (related to Fig. 6). ATAC-seq regions (peaks) and chromatin modules**
1590 **Supplemental Table S9 (related to Fig. 6). ATAC-seq differential accessibility analysis**
1591 **Supplemental Table S10 (related to Fig. 6). ATAC-seq chromatin module enrichment results**
1592 **Supplemental Table S11 (related to Fig. 7). Transcription factor target genes**
1593 **Supplemental Table S12 (related to Fig. 7). Transcription factor enrichments**

1594

1595 **Supplemental Video S1 (related to Fig. 5). Time-lapse imaging of 17q1q hESCs in low-density**
1596 **culture**
1597 **Supplemental Video S2 (related to Fig 5). Time-lapse imaging of 17q1qMYCN hESCs in low-**
1598 **density culture**