nature genetics

Neuroblastoma is composed of two super-enhancer-associated differentiation states

Tim van Groningen[1,9], Jan Koster[1,9], Linda J Valentijn[1,9], Danny A Zwijnenburg[1], Nurdan Akogul[1], Nancy E Hasselt[1], Marloes Broekmans[1], Franciska Haneveld[1], Natalia E Nowakowska[1], Johannes Bras[2], Carel J M van Noesel[2], Aldo Jongejan[3], Antoine H van Kampen[3], Linda Koster[4], Frank Baas[4,8], Lianne van Dijk-Kerkhoven[5], Margriet Huizer-Smit[5], Maria C Lecca[1], Alvin Chan[1], Arjan Lakeman[1], Piet Molenaar[1], Richard Volckmann[1], Ellen M Westerhout[1], Mohamed Hamdi[1], Peter G van Sluis[1], Marli E Ebus[1,8], Jan J Molenaar[1,8], Godelieve A Tytgat[6,7], Bart A Westerman[1,8], Johan van Nes[1,10] & Rogier Versteeg[1,6,7,10]

Neuroblastoma and other pediatric tumors show a paucity of gene mutations, which has sparked an interest in their epigenetic regulation. Several tumor types include phenotypically divergent cells, resembling cells from different lineage development stages[1–4]. It has been proposed that super-enhancer-associated transcription factor (TF) networks underlie lineage identity[5,6], but the role of these enhancers in intratumoral heterogeneity is unknown. Here we show that most neuroblastomas include two types of tumor cells with divergent gene expression profiles. Undifferentiated mesenchymal cells and committed adrenergic cells can interconvert and resemble cells from different lineage differentiation stages. ChIP–seq analysis of isogenic pairs of mesenchymal and adrenergic cells identified a distinct super-enhancer landscape and super-enhancer-associated TF network for each cell type. Expression of the mesenchymal TF PRRX1 could reprogram the super-enhancer and mRNA landscapes of adrenergic cells toward a mesenchymal state. Mesenchymal cells were more chemoresistant *in vitro* and were enriched in post-therapy and relapse tumors. Two super-enhancer-associated TF networks, which probably mediate lineage control in normal development, thus dominate epigenetic control of neuroblastoma and shape intratumoral heterogeneity.

Neuroblastoma is a pediatric tumor of the peripheral adrenergic lineage, which is derived from the neural crest. During embryogenesis, cells delaminate from the neural crest, migrate ventrally and differentiate into adrenaline- or noradrenaline-producing cells. Neuroblastomas typically express enzymes for the adrenaline-synthesis route. High-stage neuroblastomas usually go into complete remission upon therapy but often relapse as therapy-resistant disease. Although seminal studies of the neuroblastoma cell line SK-N-SH identified two phenotypically divergent subclones[7], no consensus view of intratumoral heterogeneity in neuroblastoma has emerged. We have generated new neuroblastoma cell lines in neural stem cell medium. Cell lines derived from the same patient showed divergent phenotypes and mRNA profiles. One of the differentially expressed genes was the marker for normal and tumor stem cells *PROM1* (encoding CD133)[8–10], whose expression was either absent or high in the cell lines (**Supplementary Fig. 1a–d**). We studied three isogenic pairs of CD133$^+$ and CD133$^-$ cell lines, each derived from one patient with neuroblastoma (AMC691, AMC700 and AMC717; **Supplementary Table 1**)[11]. Whole-genome sequencing and array comparative genomic hybridization (CGH) confirmed shared genomic defects in the cell line pairs and original tumors, verifying the isogenic origin of the cell lines[12]. CD133$^+$ cells grew attached, formed lamellipodia and were motile, whereas CD133$^-$ cells propagated as semiattached spheres and did not migrate (**Supplementary Fig. 1a,e**). The distinct mRNA profiles of CD133$^-$ and CD133$^+$ cells were confirmed by means of protein blot analysis (**Fig. 1a,b**). CD133$^-$ cells expressed genes involved in adrenergic differentiation, like *PHOX2A*, *PHOX2B* and *DBH*, and correspond to classic neuroblastoma cells with an adrenergic lineage identity[13]. We refer to these cells as adrenergic (ADRN) neuroblastoma cells. In contrast, CD133$^+$ cells had high mRNA and protein levels of the mesenchymal markers *SNAI2*, *VIM* (vimentin) and *FN1* (fibronectin), and we refer to these cells as mesenchymal (MES) tumor cells (**Fig. 1b**). mRNA profiling of the phenotypically divergent subclones SH-EP2 and SH-SY5Y derived from the SK-N-SH cell line[7] identified similar differences, with SH-EP2 cells conforming to MES-type cells

[1]Department of Oncogenomics, Academic Medical Center, Amsterdam, the Netherlands. [2]Department of Pathology, Academic Medical Center, Amsterdam, the Netherlands. [3]Department of Bioinformatics, Academic Medical Center, Amsterdam, the Netherlands. [4]Department of Genome Diagnostics, Academic Medical Center, Amsterdam, the Netherlands. [5]Department of Clinical Genetics, VU Medical Center, Amsterdam, the Netherlands. [6]Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands. [7]Department of Pediatric Oncology, Emma Children's Hospital, Academic Medical Center, Amsterdam, the Netherlands. [8]Present addresses: Department of Clinical Genetics, LUMC, Leiden, the Netherlands (F.B.), Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands (M.E.E. and J.J.M.) and Department of Neurosurgery, Cancer Center Amsterdam, Amsterdam, the Netherlands (B.A.W.). [9]These authors contributed equally to this work. [10]These authors jointly directed this work. Correspondence should be addressed to J.v.N. (w.j.vannes@amc.uva.nl) or R. Versteeg (r.versteeg@amc.uva.nl).

Received 19 December 2016; accepted 19 May 2017; published online 26 June 2017; doi:10.1038/ng.3899

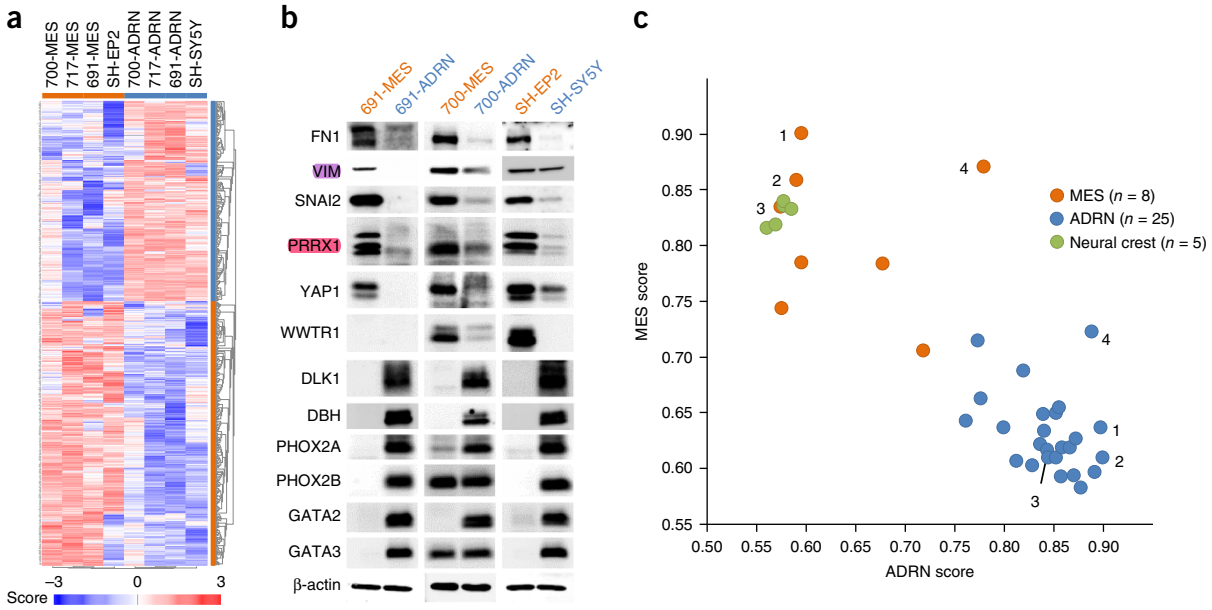

Figure 1 Identification of MES- and ADRN-type neuroblastoma cells. (**a**) Unsupervised clustering analysis using mRNA expression for MES- and ADRN-signature genes of isogenic pairs of MES and ADRN cell lines. The four MES cell lines cluster on the left, and the four ADRN cell lines cluster on the right. Genes from the MES (orange) and ADRN (blue) signatures are shown as rows. (**b**) Protein blot analysis of mesenchymal (FN1, VIM, SNAI2, PRRX1, YAP1, WWTR1) and adrenergic (DLK1, DBH, PHOX2A, PHOX2B, GATA2, GATA3) markers in cell line pairs. Multiple gels and filters were used (details and uncropped images are provided in **Supplementary Data**). (**c**) MES and ADRN signature scores for the 4 cell line pairs, 25 neuroblastoma cell lines and 5 neural crest–derived cell lines. The cell line pairs are numbered 1–4 (derived from patients AMC717, AMC691, SH and AMC700, respectively), with the MES and ADRN counterparts labeled (orange and blue, respectively). *k*-means analysis of the most differentially expressed genes (the top 500, 1,000 or 1,500) allotted the 33 neuroblastoma cell lines into two clusters corresponding to a MES (orange) or ADRN (blue) score. The five human neural crest–derived cell lines are shown in green.

and SH-SY5Y cells resembling ADRN-type cells. Using all four isogenic cell line pairs, we established a 485-gene MES mRNA signature and a 369-gene ADRN mRNA signature (**Supplementary Table 2**). Unsupervised clustering using either the genes with the highest s.d. in expression or the MES- and ADRN-signature genes grouped the eight cell lines according to the MES and ADRN phenotypes, thus supporting the fundamental nature of this classification (**Fig. 1a** and **Supplementary Fig. 1f**).

The genes in the mRNA signatures were used to calculate MES and ADRN mRNA scores for 33 neuroblastoma cell lines, including new and established lines. The results showed a gradient for MES and ADRN scores (**Fig. 1c**). *k*-means clustering of all cell lines using either genes from the MES and ADRN signatures or genes with the highest s.d. in expression gave identical results and identified 8 MES- and 25 ADRN-type cell lines (**Fig. 1c** and **Supplementary Fig. 1g**). Of the established cell lines, neither ADRN- nor MES-type lines expressed CD133, possibly owing to serum culture conditions. The signature scores of the MES cells were similar to those of human neural crest–derived cell lines[14], suggesting that these cells correspond to precursors of the adrenergic lineage (**Fig. 1c**).

To assess the relationship between the genetically similar but phenotypically divergent cell lines, we investigated whether MES and ADRN cells can interconvert. We FACS sorted $CD133^+$ and $CD133^-$ cells from the heterogeneous cell line AMC700B. Single $CD133^-$ cells were clonally expanded, and, after prolonged culture, seven clones were found to include 0.8 to 34% $CD133^+$ cells (**Supplementary Fig. 2a**). Single $CD133^+$ cells lacked clonogenic potential, but sorted and batch-cultured $CD133^+$ cells included 51% $CD133^-$ cells after eight passages. We tested *in vivo* tumorigenic and transdifferentiation potential by means of subcutaneous inoculation of mice with 2×10^5 FACS-sorted $CD133^+$ or $CD133^-$ AMC700B cells. $CD133^-$ cells were more tumorigenic than $CD133^+$ cells (**Supplementary Fig. 2b**), but all tumors were heterogeneous, as assessed by immunohistochemical (IHC) analysis for vimentin and DBH (**Supplementary Fig. 2c**). MES and ADRN cells can therefore transdifferentiate into one another *in vitro* and *in vivo*.

A growing body of evidence implicates super-enhancers as key elements defining cell identity[5,6,15], but the relationship of these enhancers to intratumoral heterogeneity is unknown. We performed ChIP–seq analysis of the H3K27ac and H3K4me3 histone modifications for four MES and five ADRN neuroblastoma cell lines, including three isogenic cell line pairs. The super-enhancers in these cell lines were defined by unstitched ROSE analysis of RSEG-called peaks (excluding prominent H3K4me3 signals; Online Methods). We selected super-enhancers that were found in two or more cell lines ($n = 1{,}662$) and performed an unsupervised clustering analysis of the H3K27ac signal measured over a region of 100 kb around the center of each super-enhancer. The nine cell lines consistently clustered into two groups, with the ADRN lines separated from the MES lines. The 1,662 super-enhancers were ordered according to differences in H3K27ac intensity in the MES and ADRN groups (limma test) (**Fig. 2a**). All four MES cell lines had highly similar super-enhancer patterns, which strongly contrasted with the super-enhancer patterns of the five ADRN cell lines. The same contrast was found within each cell line pair. There were 286 super-enhancers that had significantly stronger signal in MES-type cells and 276 super-enhancers that had significantly stronger signal in ADRN-type cells (limma test, false discovery rate (FDR) < 0.1). Many of the MES-specific super-enhancers corresponded to super-enhancers identified in neural crest–derived cells (**Fig. 2a** and **Supplementary Table 3**)[16]. We analyzed whether and how these

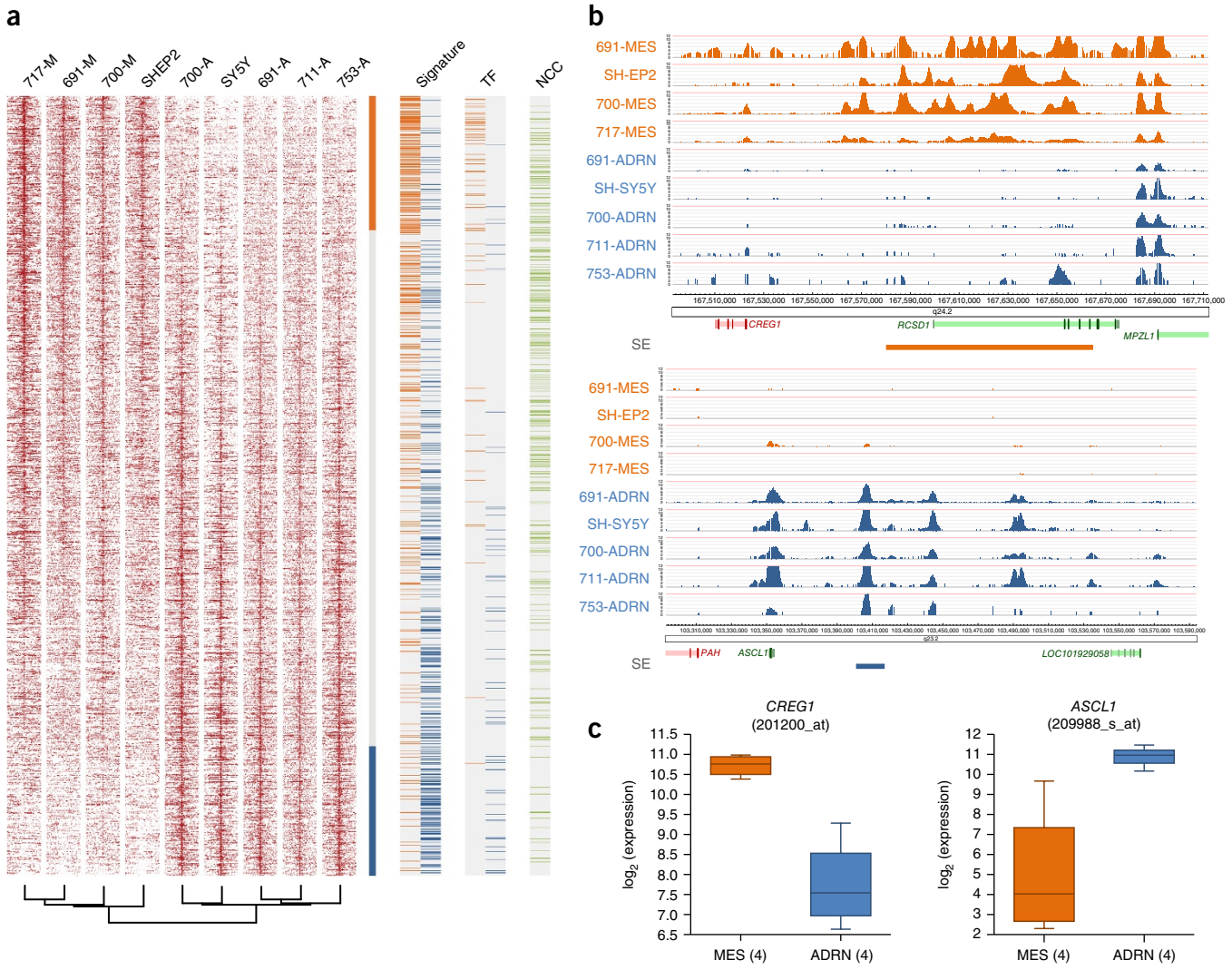

Figure 2 Super-enhancer landscapes differ in MES and ADRN cell lines and are associated with MES and ADRN signatures. (**a**) Super-enhancer landscape of four MES and five ADRN cell lines. Super-enhancers found in two or more cell lines (n = 1,662) were used to cluster samples (Ward2). The x axis represents 0.5 Mb up- and downstream of the center of the super-enhancer. Super-enhancers are ordered on the basis of their ability to distinguish MES from ADRN lines and vice versa (limma test); those with significant limma test results (FDR < 0.1) are indicated by the orange and blue vertical bars. To the right, MES- and ADRN-signature genes within 500 kb of a super-enhancer are represented by orange and blue stripes, respectively, and the same is shown specifically for TF genes. On the far right, neural crest cell (NCC) super-enhancers[16] overlapping the neuroblastoma super-enhancers are highlighted. (**b**) Examples of MES- and ADRN-specific enhancer regions associated with the MES- and ADRN-specific TFs *CREG1* (top) and *ASCL1* (bottom), respectively. H3K27ac profiles are shown for all nine cell lines. The y axis represents the number of reads per 20 million mapped reads. Super-enhancer (SE) regions are indicated below by bars. (**c**) Box plots of mRNA expression for *CREG1* and *ASCL1* in the MES and ADRN counterparts of the four cell line pairs of isogenic origin. Whiskers denote the interval within 1.5 times the interquartile range (box edges) of the median (center line).

super-enhancers were associated with genes from the MES and ADRN mRNA signatures. Genes with MES-specific expression patterns were significantly associated with MES-specific super-enhancers, measured within either 0.5 or 1.0 Mb ($P < 10^{-4}$, binomial test). A similar association was found for genes with ADRN-specific expression patterns and ADRN-specific super-enhancers (**Fig. 2a** and **Supplementary Table 3**). These findings suggest that super-enhancers indeed control part of the cell-type-specific gene expression. Many gene associations with super-enhancers were confirmed by FANTOM5 data (**Supplementary Table 3**)[17]. Strong ADRN-specific super-enhancers were associated with known markers of adrenergic differentiation, such as *DBH*, *CHGA* and *DLK1* (ref. 18). The genes with MES-specific expression changes associated with MES-specific super-enhancers included, for example, *WNT5A*, *IGFBP2*, *FN1* and *IL13RA1* (**Fig. 2b,c**, **Supplementary Fig. 3** and **Supplementary Table 3**).

We further assessed which TF genes with cell-type-specific expression were associated with super-enhancers. After validation of the H3K27ac patterns, 20 MES TF genes were identified with strong MES-specific super-enhancers, for example, *MEOX1*, *MEOX2*, *SIX1*, *SIX4*, *SOX9*, *SMAD3* and *WWTR1*. Similarly, we identified 18 ADRN super-enhancer-associated TF genes, including *ASCL1*, *EYA1*, *GATA3*, *HAND1* and *SIX3* (**Figs. 2b,c** and **3**, **Supplementary Fig. 3** and **Supplementary Table 4**). Overall, 34% of the cell-type-specific TF genes were associated with a cell-type-specific super-enhancer. These TFs were often highly expressed (**Supplementary Fig. 4**), and their associated super-enhancers belonged to the group of H3K27ac

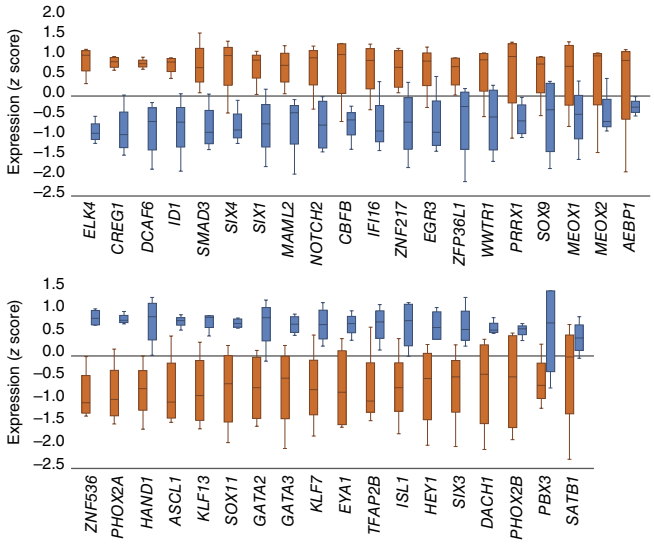

Figure 3 Expression of super-enhancer-associated lineage TFs in MES and ADRN cell lines. Box plots are shown of *z* scores corresponding to mRNA expression values for core super-enhancer-associated lineage TFs of MES-type cells (top) and ADRN-type cells (bottom). MES and ADRN cell line pairs of isogenic origin are plotted in groups of MES-type cells (691-MES, 700-MES, 717-MES, SH-EP2; orange) and corresponding ADRN-type cells (691-ADRN, 700-ADRN, 717-ADRN, SH-SY5Y; blue). Whiskers denote the interval within 1.5 times the interquartile range (box limits) of the median (center line).

peaks with the strongest signal among the super-enhancer profiles. These super-enhancer-associated TF genes are therefore candidate master genes for each of the two cell types. Induced pluripotent stem cell (iPSC) research has suggested that cell identity is imposed by a core set of TFs that mutually bind one another's super-enhancers, thus creating a feed-forward loop and resulting in very high expression[5,6,15]. We identified *GATA3* as an ADRN super-enhancer-associated TF. ChIP–seq data for GATA3 (ref. 19) confirmed prominent binding of GATA3 to its own associated super-enhancer, as well as to the super-enhancers of most other ADRN TF genes, supporting the idea of a feed-forward network among the ADRN super-enhancer-associated TFs (**Supplementary Fig. 5**). Most of the identified super-enhancer and TF gene combinations showed cell-type-specific super-enhancer patterns. However, although *PHOX2B* and *PHOX2A* had preferential or exclusive ADRN-specific expression, their associated super-enhancers could also be prominent in MES-type cells (**Supplementary Fig. 6**). Apparently, these TFs with a documented role in adrenergic differentiation are epigenetically marked in MES cells.

The core set of super-enhancer-associated TF genes in iPSCs were marked by their reprogramming potential[20]. We explored whether MES super-enhancer-associated TFs can reprogram ADRN cells into MES cells. Several of these TFs have been implicated in epithelial–mesenchymal transitions (EMT). We chose to test the homeobox TF PRRX1 (**Figs. 1b** and **3**), which can induce EMT in breast cancer cells on its own[21]. Inducible PRRX1 expression in the SK-N-BE(2)-C cell line and two more adrenergic cell lines resulted in a motile and mesenchymal phenotype, accompanied by induction of the MES marker SNAI2 and repression of the ADRN markers PHOX2B and DBH (**Supplementary Fig. 7a**,**b**). We followed the cell fate of SK-N-BE(2)-C cells over a 28-d period of induced PRRX1 expression. mRNA profiling showed step-by-step reprogramming to an induced mesenchymal state within 14 d after the initiation of PRRX1 expression (**Fig. 4a**). The profiles showed a striking upregulation of MES-specific genes and downregulation of some, but not all, ADRN-specific genes (**Fig. 4b**). The kinetics of downregulation for the ADRN-specific genes varied, with downregulation occurring by 8 h for some genes and several days later for other genes (**Supplementary Fig. 7c**). ChIP–seq analysis of H3K27ac at day 12 of PRRX1 expression showed a shift toward the MES super-enhancer pattern (**Fig. 4c–e**). Reprogramming of ADRN cells by PRRX1 therefore shifts the super-enhancer and gene expression profiles to a MES state, corroborating the tight regulatory underpinning of each cell type. The role of PRRX1 is not symmetric, as short hairpin RNA (shRNA) or CRISPR–Cas9-mediated silencing of *PRRX1* did not trigger ADRN differentiation in three MES cell lines (**Supplementary Fig. 7d–g**).

The distinct MES and ADRN phenotypes *in vitro* prompted us to analyze the presence of MES cells in neuroblastoma *in vivo*. A series of 122 neuroblastomas of all stages had MES and ADRN mRNA signatures placing them in between the MES and ADRN counterparts of the cell line pairs (**Fig. 5a**). This suggested that neuroblastomas are a mixture of both cell types with, for most tumors, a balance toward ADRN cells. Tumors with high MES signature scores cannot be explained by high stromal infiltration, as the tumor cell content in these tumors was very high and comparable to that of ADRN tumors (**Supplementary Fig. 8a**,**b**). Intratumoral heterogeneity was directly tested by IHC analysis of PRRX1 expression. Most tumors showed nuclear PRRX1 staining in a minority of cells (data not shown). Double IHC for PRRX1 and MYCN in six stage 4 neuroblastomas with high MYCN protein levels detected 0.5–10% double-positive cells, confirming the neuroblastoma identity of the PRRX1$^+$ cells (**Supplementary Fig. 9a**,**b**). Tissue microarrays of 31 neuroblastomas of all stages were analyzed by double IHC for PRRX1 and MAML3, which we used as a pan-neuroblastoma marker (Online Methods and **Supplementary Fig. 9c**,**d**). Double-positive cells were detected in at least 80% of tumors, at a low but variable frequency (median 1.85%, range 0 to 70%; **Fig. 5b**,**c**). The variation between stage 1 and 2 versus stage 3 and 4 tumors was not significant (Wilcoxon rank-sum test). Further validation of these data and assessment of whether PRRX1$^+$ cells express the full complement of MES-specific genes have to await single-cell RNA–seq analyses of neuroblastoma.

The mRNA and IHC analyses did not show any correlation between the mesenchymal character of tumors and clinical or molecular parameters (**Supplementary Fig. 8** and data not shown). Accordingly, *PRRX1* expression in our series and in other neuroblastoma series did not correlate with prognosis (**Supplementary Fig. 8i–l**)[22].

We investigated the clinical relevance of both cell types. Mesenchymal cells of various tumor types are often relatively resistant to chemotherapy[23]. *In vitro*, MES cells were more resistant to the standard neuroblastoma drugs cisplatin, doxorubicin and etoposide than their ADRN counterparts (**Supplementary Fig. 10a**). Upon diagnosis, high-risk neuroblastoma is treated with chemotherapy followed by resection of the tumors. Viable tumor cells in four rest lesions analyzed had a large majority of PRRX1$^+$ cells (**Supplementary Fig. 10b**). We compared two of these samples with the corresponding pretreatment biopsies, identifying strong enrichment for PRRX1$^+$ cells after therapy (**Supplementary Fig. 10b**). We could also compare primary and relapse tumors in two patients. Both relapse tumors, occurring 4 and 5 years after diagnosis, showed

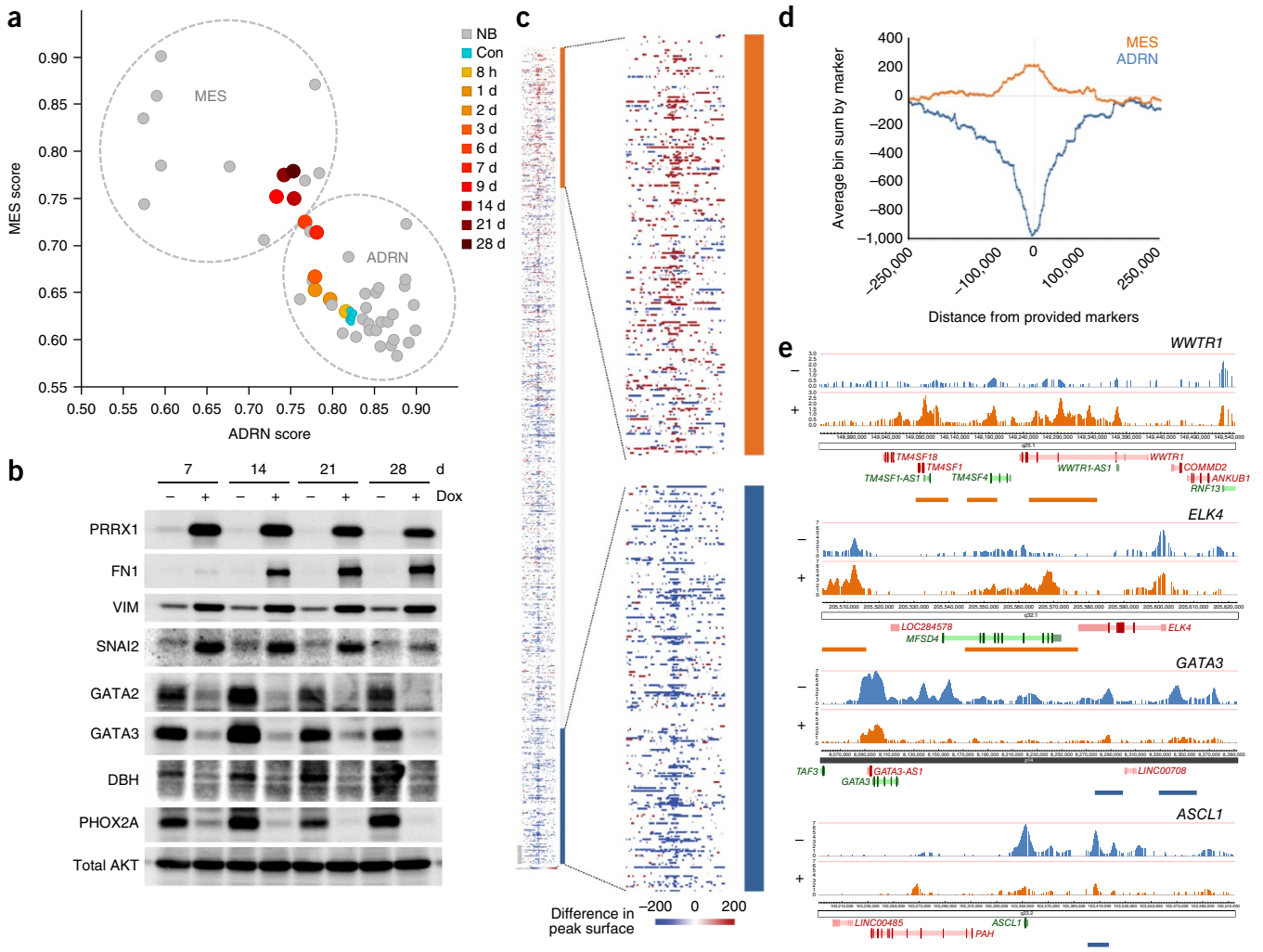

Figure 4 PRRX1 expression induces a transition toward MES mRNA and super-enhancer profiles. (**a**) 28-d time-course analysis of MES and ADRN signature scores for SK-N-BE(2)-C cells with doxycycline-inducible PRRX1 expression. Early to late time points of PRRX1 induction are indicated by the yellow-to-red gradient; non-induced control (Con) cells are shown in blue. The 33 neuroblastoma (NB) cell lines (gray) are depicted for reference, with dashed circles indicating MES and ADRN classifications as in **Figure 1c**. (**b**) Protein blot analysis of cell line SK-N-BE(2)-C with (+) or without (–) doxycycline-induced PRRX1 expression, with cells followed for 28 d. The MES markers FN1, VIM and SNAI2 and the ADRN markers GATA2, GATA3, DBH and PHOX2A were analyzed. Total AKT is used as a loading control. Dox, doxycycline. Uncropped images are shown in **Supplementary Data**. (**c**) Changes in H3K27ac signal for all 1,662 super-enhancer regions in SK-N-BE(2)-C cells upon induction of PRRX1 expression. The changes in the H3K27ac signal of the super-enhancers are indicated by the red (increased) or blue (decreased) signal of the genomic regions (horizontal bars). Expanded views are shown of the regions of the super-enhancers that distinguish MES and ADRN lines (orange and blue vertical bars, respectively) (right). (**d**) Summed changes in the H3K27ac signal of super-enhancers significant in distinguishing MES cells or ADRN cells (limma test, $P = 0.1$), centered around the middle of super-enhancers. (**e**) Examples of H3K27ac profiles in SK-N-BE(2)-C cells with (+) or without (–) 12 d of doxycycline-induced PRRX1 expression. Shown are examples of MES-specific super-enhancer-associated TF genes (*WWTR1*, *ELK4*) and ADRN-specific super-enhancer-associated TF genes (*GATA3*, *ASCL1*). ADRN- and MES-specific super-enhancers are shown by blue and orange horizontal bars, respectively.

decreased DBH staining and a strong increase in the proportion of PRRX1$^+$ cells (**Fig. 5d**). MES neuroblastoma cells are therefore more resistant to chemotherapy *in vitro* and can be enriched in post-chemotherapy and relapse biopsies *in vivo*, suggesting that therapy exerts selective pressure.

Here we showed that intratumoral heterogeneity in neuroblastoma is not a random process but is brought about by consistent regulatory programs. These programs are master regulators of two divergent phenotypes, which can nevertheless interconvert. This plasticity and the relative resistance of MES cells to therapeutic agents may facilitate escape from current therapies. Targeted therapy directed at MES cells may offer a prospect to impede relapse.

URLs. FANTOM5 data, http://enhancer.binf.ku.dk/presets/hg19_enhancer_promoter_correlations_distances_cell_type.txt.gz.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was supported by grants from KiKa (projects 11, 33, 66), the Dutch Cancer Society KWF (UVA 2010-4878), the Tom Voûte Fund/KiKa, Villa Joep and

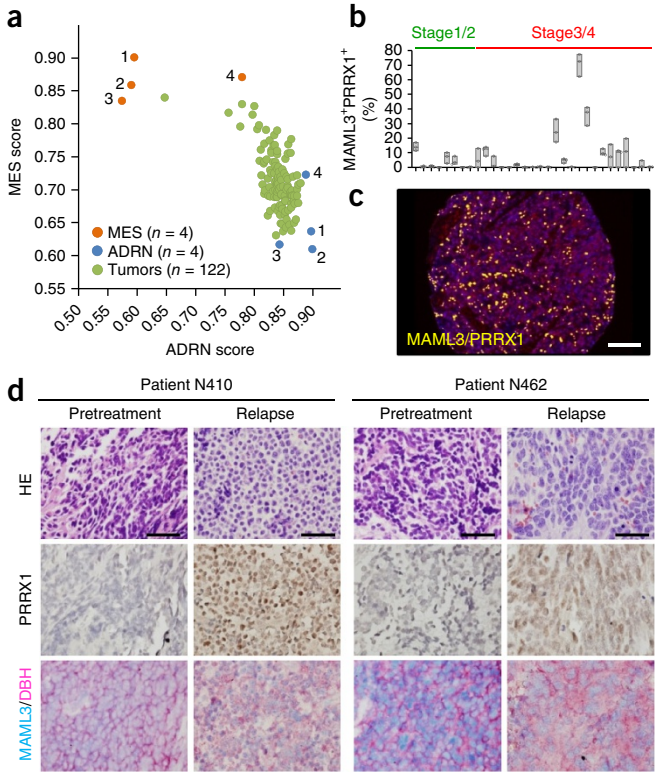

Figure 5 Neuroblastomas include a subset of PRRX1$^+$ tumor cells. (**a**) Signature scores of a series of primary neuroblastoma tumors (green). The isogenic pairs of MES (orange) and ADRN (blue) cell lines (**Fig. 1c**) are shown as reference. (**b**) Box plots showing the frequency of MAML3$^+$ PRRX1$^+$ cells in INSS stage 1–4 neuroblastomas. Cells were counted in up to three sections per tumor. The difference between stage 1 or 2 and stage 3 or 4 neuroblastomas is not significant ($P = 0.46$, two-sided Wilcoxon rank-sum test). (**c**) Representative false-color image of double IHC for MAML3 (blue) and PRRX1 (red) in a stage 4 neuroblastoma. MAML3$^+$PRRX1$^+$ double-positive cells are yellow. Scale bar, 100 µm. (**d**) Comparison of primary (pretreatment) and relapse tumors from patients N410 (left) and N462 (right). Tumor biopsies were analyzed by IHC for the MES marker PRRX1 (middle) or by double IHC for the ADRN marker DBH (red) and the pan-neuroblastoma marker MAML3 (blue) (bottom). Scale bars, 50 µm. HE, hematoxylin and eosin staining.

the European Research Council (ERC-Adv no. 340735). We are indebted to Chris van der Loos, who sadly passed away, for sharing expertise on immunohistochemistry.

AUTHOR CONTRIBUTIONS

T.v.G., J.K., L.J.V., J.v.N. and R. Versteeg conceived the study, analyzed the data and wrote the manuscript. D.A.Z., P.M. and R. Volckmann performed bioinformatics analyses. N.A., N.E.H., M.B., F.H., N.E.N., M.C.L., A.C., A.L., P.G.v.S. and M.E.E. performed part of the experiments and analyses. J.B. and C.J.M.v.N. performed histopathological analyses of tumor samples. A.J. and A.H.v.K. assisted with data analysis. L.K. and F.B. performed library preparation and sequencing, and L.v.D.-K. and M.H.-S. performed sequencing. E.M.W., M.H., J.J.M. and B.A.W. supervised part of the study. G.A.T. analyzed clinical data.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
2. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).
3. Patel, A.P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
4. Suvà, M.L. *et al.* Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).
5. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
6. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
7. Ross, R.A., Spengler, B.A. & Biedler, J.L. Coordinate morphological and biochemical interconversion of human neuroblastoma cells. *J. Natl. Cancer Inst.* **71**, 741–747 (1983).
8. Singh, S.K. *et al.* Identification of human brain tumour initiating cells. *Nature* **432**, 396–401 (2004).
9. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
10. Zhu, L. *et al.* Multi-organ mapping of cancer risk. *Cell* **166**, 1132–1146.e7 (2016).
11. Bate-Eya, L.T. *et al.* Newly-derived neuroblastoma cell lines propagated in serum-free media recapitulate the genotype and phenotype of primary neuroblastoma tumours. *Eur. J. Cancer* **50**, 628–637 (2014).
12. Molenaar, J.J. *et al.* Sequencing of neuroblastoma identifies chromothripsis and defects in neuritogenesis genes. *Nature* **483**, 589–593 (2012).
13. Zwiller, J., Treska-Ciesielski, J., Mack, G. & Mandel, P. Uptake of noradrenaline by an adrenergic clone of neuroblastoma cells. *Nature* **254**, 443–444 (1975).
14. de Pontual, L. *et al.* Epistasis between RET and BBS mutations modulates enteric innervation and causes syndromic Hirschsprung disease. *Proc. Natl. Acad. Sci. USA* **106**, 13921–13926 (2009).
15. Young, R.A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
16. Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633–648 (2012).
17. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
18. van Limpt, V.A. *et al.* High delta-like 1 expression in a subset of neuroblastoma cell lines corresponds to a differentiated chromaffin cell type. *Int. J. Cancer* **105**, 61–69 (2003).
19. Oldridge, D.A. *et al.* Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature* **528**, 418–421 (2015).
20. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
21. Ocaña, O.H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–724 (2012).
22. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).
23. Arumugam, T. *et al.* Epithelial to mesenchymal transition contributes to drug resistance in pancreatic cancer. *Cancer Res.* **69**, 5820–5828 (2009).

ONLINE METHODS

Cell culture and derivation of MES- and ADRN-type cell lines. Patient-derived neuroblastoma cell lines were cultured in neural stem cell medium as described[11]. Tumor material was obtained after informed consent. MES- and ADRN-type cell line pairs were established as follows; 691-MES and 691-ADRN were described previously (AMC691T and AMC691B, respectively)[11]. The parental AMC700B cell line (heterogeneous for CD133 positivity, between 20–40% $CD133^+$) was double FACS sorted into $CD133^-$ and $CD133^+$ populations, which were further cultured to establish the 700-MES and 700-ADRN cell lines. The 700-MES line maintained 80–90% $CD133^+$ cells, while the 700-ADRN line usually maintained a low frequency of 0–5% $CD133^+$ cells. Early-passage parental AMC717T cells showed heterogeneous growth of floating spheres and attached cells with lamellipodia. Separate culturing of these phenotypically different cells resulted in the establishment of the 717-MES (>95% $CD133^+$) and 717-ADRN (usually around 7% $CD133^+$) lines. Serum-cultured cell lines were maintained as previously described[24]. Cell line identity was routinely verified using short tandem repeat (STR) analysis. Cells were routinely checked by PCR for the presence of mycoplasma.

ChIP–seq analysis. ChIP was performed as previously described[25]. In brief, cultured cells were fixed with 1% formaldehyde. Nuclei were isolated, and the DNA was sheared to 200- to 300-bp fragments. Histone-bound DNA was precipitated using antibodies against H3K27ac (4729, Abcam) and H3K4me3 (04-745, Millipore). Cross-linking was reversed, and DNA was purified using the Qiagen PCR purification kit (Qiagen) and quantified with Quant-IT Picogreen (Invitrogen). The DNA was used to generate sequencing libraries according to the manufacturer's procedure (Life Technologies). The DNA was end polished and dA tailed, and adaptors with barcodes were ligated. The fragments were amplified (eight cycles) and quantified with a Bioanalyzer (Agilent).

Libraries were prepped with the 5500W Flowchip v2 kit (Life Technologies) and sequenced on SOLiD Wildfire (Illumina), resulting in 50-bp reads. Alternatively, libraries were sequenced using the HiSeq PE cluster kit v4 (Illumina) with the HiSeq 2500 platform (Illumina), resulting in 125-bp reads. ChIP–seq data can be accessed from GEO (accession number GSE90805).

ChIP–seq data processing. FastQ files were aligned to the human genome (NCBI38/hg19) using Bowtie v1.1.1 with the following arguments (-m 10500 --strata --best -t -v 2). Peaks were called using the RSEG algorithm[26] with the following arguments (rseg-diff -d k41.hg19.bed -m 2 -v -i 20 -c human.hg19.size.bed -o). For H3K27ac ChIP–seq profiles, the ROSE algorithm[5] was executed on the RSEG-called peaks to define super-enhancer status using the following arguments (-s 0 -t 0). For cases with *MYCN* amplification, peaks overlapping the *MYCN* amplicon were removed for the ROSE analysis, as their inclusion tended to disproportionally skew the super-enhancer distribution. As H3K27ac signals are present in both enhancers and actively transcribed promoter regions, we assessed H3K4me3 signals using ROSE on peaks that had been defined in the H3K27ac profile. In this way, we established the H3K4me3 signal in H3K27ac-defined regions. Peaks that contained an H3K4me3 signal higher than 1,000 ROSE units per kilobase were discarded in analyses in which super-enhancer overlap was investigated, as we considered these peaks to represent transcription-associated regions rather than enhancer regions. To assess the super-enhancer landscape across the different samples, we superimposed all super-enhancers that were obtained from ROSE analyses of the samples, excluding the transcription-associated regions described above. Superimposed enhancers where at least 2 of the 9 samples were represented were selected and used for further analysis. For every superimposed super-enhancer, the RSEG sum of the flanking 100 kb on each side of the center was determined for all nine samples. These values were normalized to 10 million per sample, $\log_{10}$ transformed and subsequently z scored per super-enhancer. This matrix was then used to perform Ward2 clustering on the Euclidean distance matrix to produce the ordering of the samples. This generated a clear separation into two groups matching the MES and ADRN cell lines. For visualization, we created super-enhancer heat maps (RSEG signals ±500,000 bp, binned per 5,000 bases) for each sample, ordering the enhancers by the limma-tested differential signal between the two groups. To generate the color intensity scale, we used the factors introduced in the normalization step described above.

For detailed visualization purposes, aligned BAM files for samples were processed by MACS2, extending reads to 200 bases. Pileups were binned by 25 bp and visualized on the R2 genomics analysis and visualization platform (http://r2.amc.nl/).

To assess whether genes within a gene expression signature were associated with a superimposed super-enhancer, we recorded all super-enhancers that were within 0.5 Mb or 1.0 Mb of the genes in the signature. We also assessed the relationship between enhancers and gene expression by linking genes from our mRNA signatures to FANTOM5 data at FDR = 10^{-5} (see URLs).

Gene expression analysis and composition of the MES and ADRN signatures. Neuroblastoma tumors and cell lines were profiled using Affymetrix HG U133 plus 2.0 arrays, essentially as described[24]. Microarray data and analysis tools are available from the R2 genomics analysis and visualization platform (http://r2.amc.nl/). Expression profiles for a neuroblastoma cell line panel (accession number GSE28019) and neuroblastoma tumors (GSE16476) are available from GEO. Data for human neural crest cells were previously published[14] and can be accessed from GEO (GSE14340). Expression profiles for the MES- and ADRN-type cells used in this study are available from GEO (GSE90805). Time-course Affymetrix mRNA data from SK-N-BE(2)-C-pIND-*PRRX1* cells are available from GEO (accession number GSE90805).

The signatures for mesenchymal and adrenergic cells were generated from mRNA profiles for four cell line pairs of isogenic origin. Genes with significantly different expression[27] with a minimum of one present call and minimum expression of 50 units (in at least one of the pair members) were selected from each MES–ADRN cell line pair. Lists of regulated genes from the four pairs were merged, and genes were selected for consistent down- or upregulation. Genes consistently regulated in at least three of the four cell line pairs and that had a minimum pairwise expression difference of 100 in at least one cell line pair were included. This analysis identified 485 MES-specific and 369 ADRN-specific genes (**Supplementary Table 1**).

Gene signature score. To convert the MES and ADRN gene signatures into a single value per sample, we created a signature score as follows: within each sample, we ranked all genes according to their expression level (using HugoOnce to represent each gene by only one reporter). For each gene, we established the rank order in this list and calculated the percentile relative to the full list. The average of the percentiles for all MES or for all ADRN signature genes was used as a signature score. These values (signature scores) can readily be compared across samples. The method is implemented and available on the R2 platform.

FACS analysis and interconversion assays. Cell cultures were dissociated using non-enzymatic dissociation (NeuroCult dissociation kit, Stem Cell Technologies, 5707). FACS analysis of CD133 (PROM1) was performed using anti-CD133-APC (Miltenyi Biotec, 130-090-854) and control IgG2b-APC (Miltenyi Biotec, 130-092-217) according to the manufacturer's instructions.

For single-cell FACS sorting, dissociated cells were stained using anti-CD133/2 (293C3)-PE (Miltenyi Biotec, 130-090-853) and IgG2b-PE isotype control (Miltenyi Biotec, 130-092-215). Individual cells were plated into 384-well plates (BD Biosciences, 353270). Plating of individual cells was visually confirmed. After single-cell colony outgrowth, semiadherent spheres were serially passaged when cultures were close to confluence.

***In vivo* tumorigenicity.** Tumor growth assays were performed in female NOD-scid mice (NOD.CB17-Prkdc/NCrHsd, 6–15 weeks old, 20–30 g) (Harlan Laboratories) that were subcutaneously injected with 200,000 FACS-sorted $CD133^+$ or $CD133^-$ AMC700B cells in 200 µl of a 50% Matrigel (BD, 354234) solution in PBS. Mice were randomized before injection with tumor cells. A sample size of $n = 7$ mice was chosen to reach statistical power of at least 80%. After tumor outgrowth to up to 1 cm in diameter, mice were sacrificed and tumors were fixed overnight in 10% neutral-buffered formalin and embedded in paraffin. Histological analyses were performed with investigator blinding to mouse group. All experiments were conducted after obtaining ethical approval from the DEC (animal experiments committee) of the AMC.

Immunohistochemistry and multispectral imaging. Sections (4 μm) from formalin-fixed, paraffin-embedded neuroblastoma tumors (three tissue samples of each tumor) were analyzed by standard IHC protocols or dual-color IHC as previously described[28].

For IHC, the primary antibodies used were to MAML3 (IHC-00447 from Bethyl, citrate pH 6.0, 1:2,500 with DAB+/VBlue), TH (clone F-11, sc-25269 from Santa Cruz Biotechnology, citrate pH 6.0, 1:10,000 with VRed), DBH (8586 from Cell Signaling, Tris-EDTA pH 9.0, 1:25 with VRed), MYCN (clone B8.4.B, sc-53993 from Santa Cruz Biotechnology, citrate pH 6.0, 1:500 with VBlue, 1:1,000 with DAB+) and PRRX1 (HPA051084 from Sigma, citrate pH 6.0, 1:500 with DAB+/VBlue and 1:100 with VRed). The rabbit and mouse primary antibodies were detected using biotin-free species-specific HRP-conjugated and AP-conjugated polymers (BrightVision One, Immunologic). HRP activity was developed with DAB+ (K3468, Dako).

Dual-color IHC was performed as previously described[28]. AP activity in red was developed with the Vector Red kit (SK-5100, Vector Laboratories). AP activity in blue was developed with the Vector Blue kit (SK-5300, Vector Laboratories). Together with the red–blue double-AP staining, 0.1% Methyl green (M6776, Sigma) in 0.1 M sodium acetate buffer pH 5.2 was used as a nuclear counterstain. Spectral data sets from dual-color IHC were acquired using a Leica DM5000B camera (Leica Microsystems) and analyzed with Nuance 3.0.1 Multispectral Imaging analysis software (Cambridge Research Instrumentation). Negative-control sections were processed in an identical manner but omitting the first enzymatic visualization step and the second primary antibody to check for cross-reaction between the antibodies; these sections showed no staining.

Protein blot analysis. Protein blotting was performed according to standard protocols. Cells were lysed in RIPA buffer. Protein was transferred to nitrocellulose membrane (GE Healthcare, RPN203D). Membranes were blocked for 1 h at room temperature, incubated at 4 °C overnight with primary antibody and incubated for 1 h at room temperature with secondary antibody in 2% PBA (GE Healthcare, RPN418), 5% ELK or 5% OBB (LI-COR, 829-31080) in PBS with 0.1% Tween-20 (Sigma, P1379). Primary antibodies from Cell Signaling were to AKT (4691), DBH (8586), VIM (5741), SNAI2 (SLUG; 9585), YAP1 (4912), WWTR1 (TAZ; 4883), DLK1 (2069), GATA2 (4595) and GATA3 (5852). Other primary antibodies were to FN1 (AF1918, R&D Systems), PHOX2A (sc-81978, Santa Cruz), PHOX2B (sc-376997, Santa Cruz), PRRX1 (LS-B2380, Life Span BioSciences) and β-actin (ab6276, Abcam).

The secondary antibodies used for chemiluminescent detection were donkey anti-rabbit-HRP (GE Healthcare, NA 9340V), donkey anti-sheep/goat-HRP (Bio-Rad, STAR88P) and sheep anti-mouse-HRP (GE Healthcare, NXA931). Chemiluminescent detection was performed using the ECL Prime Western Blotting kit (GE Healthcare, RPN2232), and signal was developed on an ImageQuant LAS 4000 (GE Healthcare, 28-9558-10). For infrared fluorescence detection, membranes were incubated with donkey anti-rabbit-IRDye 800CW (Rockland, 611-731-127) or goat anti-mouse-IRDye 680RD (LI-COR, 926-68070) secondary antibody and scanned on an Odyssey Infrared Imaging System (LI-COR, LIC-9201-00).

Full-length gel images can be found in **Supplementary Data**.

Migration assays. Transwell migration assays were performed in 24-well Transwell inserts (8-μm pore size, Greiner) essentially as described[24]. For cells cultured in neural stem cell medium, 100,000 cells were seeded in a Transwell and 2% fetal calf serum was added as a chemoattractant. Quadruplicate Transwells were used for each experimental condition. Two-sided *t* tests assuming equal variance were used to determine significance. Experiments were repeated at least twice.

Cell viability assays. Cells were plated in 48-well or 96-well plates, 2 d before the addition of cisplatin, doxorubicin or etoposide to the culture medium. Seeding densities were optimized for each cell line so that the control, non-treated cells would reach 70–80% confluence by the end of the experiment. Chemotherapeutic agents were added for 96 or 120 h. Metabolic activity was measured by 3-(4,5-dimethylthiazol-2-*yl*)-2,5-diphenyltetrazolium bromide (MTT) assay. DNA content was measured by CyQUANT assay (Life Technologies) according to the manufacturer's instructions with minor modifications. Quadruplicate measurements were used for each experimental condition. Two-sided *t* tests assuming equal variance were used to determine significance. Experiments were repeated at least twice.

Inducible transgene expression. Transgenic cell lines with inducible PRRX1 expression were generated by cloning the cDNA encoding human *PRRX1* isoform A into the pINDUCER (pIND) system (Invitrogen). The neomycin resistance gene from pIND was replaced with the blasticidin resistance gene. Constructs were sequence verified, packaged and used for transduction at a multiplicity of infection (MOI) of 3. The 691-ADRN, SK-N-BE(2)-C and SH-SY5Y cell lines were transduced with pIND-*PRRX1A* and selected using blasticidin (5 μg/ml). Expression of the *PRRX1A* transgene was induced by the addition of doxycycline to the culture medium at a final concentration of 100 ng/ml. Parental SK-N-BE(2)-C cells were obtained from ATCC and checked by PCR for the presence of mycoplasma.

Pattern analysis of PRRX1 target genes. Genes downregulated by PRRX1 ($n = 945$, $\log_2$ fold 1 regulated) were separated into four groups using *k*-means clustering.

CRISPR–Cas9 genome editing and mRNA silencing of *PRRX1*. Control guide RNA (gRNA) sequences and *PRRX1* gRNA sequences (**Supplementary Table 5**) were cloned into lentiCRISPR v2 (refs. 29,30). lentiCRISPR v2 was a gift from F. Zhang (Broad Institute) (Addgene plasmid 52961). Lentiviruses containing lentiCRISPR v2-control or lentiCRISPR v2-*PRRX1* plasmids were generated and used to infect GI-MEN cells or a mesenchymal subclone of SK-N-AS cells. Puromycin-resistant single-cell-derived clones were obtained by limiting dilution. Clones with inactivating mutations in *PRRX1* and knockout of PRRX1 protein were selected for further analysis. Affymetrix mRNA data obtained following CRISPR–Cas9-mediated *PRRX1* knockdown in GI-MEN and SK-N-AS cells are available from GEO (accession number GSE90805). Parental SK-N-AS and GI-MEN cells were obtained from ATCC and checked by PCR for the presence of mycoplasma.

Lentivirus-mediated inducible expression of shRNAs targeting *PRRX1*. *PRRX1* shRNA sequences (**Supplementary Table 5**) targeting both *PRRX1* isoforms were selected and cloned into pLKO-puro-IPTG-3xLacO (Sigma). An shRNA targeting *lacZ* (**Supplementary Table 5**) was used as a control. pLKO-3xLacO-shRNA constructs were packaged into lentiviruses, which were used to transduce 691-MES cells. shRNA expression was induced by the addition of isopropyl-β-D-thio-galactoside (IPTG) at a final concentration of 1 μM. Affymetrix mRNA data from 691-MES cells expressing *PRRX1* shRNA are available from GEO (accession number GSE90805).

Statistics. The statistical analyses used throughout this paper are specified in the appropriate results paragraphs and sections of the Online Methods.

Data availability. ChIP–seq and mRNA expression data from this study were deposited in the Gene Expression Omnibus (GEO) under accession number GSE90805.

24. van Nes, J. *et al.* A NOTCH3 transcriptional module induces cell motility in neuroblastoma. *Clin. Cancer Res.* **19**, 3485–3494 (2013).
25. Bunt, J. *et al.* Joint binding of OTX2 and MYC in promotor regions is associated with high gene expression in medulloblastoma. *PLoS One* **6**, e26058 (2011).
26. Song, Q. & Smith, A.D. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics* **27**, 870–871 (2011).
27. Kennedy, R.E., Archer, K.J. & Miles, M.F. Empirical validation of the S-Score algorithm in the analysis of gene expression data. *BMC Bioinformatics* **7**, 154 (2006).
28. van der Loos, C.M. Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging. *J. Histochem. Cytochem.* **56**, 313–328 (2008).
29. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
30. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

doi:10.1038/ng.3899