SCIENTIFIC DATA

Check for updates

OPEN

DATA DESCRIPTOR

Epigenomic profiling of neuroblastoma cell lines

Kristen Upton[1,8], Apexa Modi[1,2,8], Khushbu Patel[3], Nathan M. Kendsersky[1,4], Karina L. Conkrite[1], Robyn T. Sussman[1], Gregory P. Way[2], Rebecca N. Adams[5], Gregory I. Sacks[1], Paolo Fortina[5], Sharon J. Diskin[1,6], John M. Maris[1,6] & Jo Lynne Rokita[1,3,7]

Understanding the aberrant transcriptional landscape of neuroblastoma is necessary to provide insight to the underlying influences of the initiation, progression and persistence of this developmental cancer. Here, we present chromatin immunoprecipitation sequencing (ChIP-Seq) data for the oncogenic transcription factors, MYCN and MYC, as well as regulatory histone marks H3K4me1, H3K4me3, H3K27Ac, and H3K27me3 in ten commonly used human neuroblastoma-derived cell line models. In addition, for all of the profiled cell lines we provide ATAC-Seq as a measure of open chromatin. We validate specificity of global MYCN occupancy in MYCN amplified cell lines and functional redundancy of MYC occupancy in MYCN non-amplified cell lines. Finally, we show with H3K27Ac ChIP-Seq that these cell lines retain expression of key neuroblastoma super-enhancers (SE). We anticipate this dataset, coupled with available transcriptomic profiling on the same cell lines, will enable the discovery of novel gene regulatory mechanisms in neuroblastoma.

Background & Summary

An estimated 15,780 children in the United States will be diagnosed with cancer in 2019[1]. While 80% of pediatric cancer patients overcome this disease, 20% of children do not survive, and survivors often have multiple side effects of therapy[1]. Neuroblastoma accounts for more than 7% of malignancies in patients under 15 years of age and approximately 12% of all pediatric cancer-related deaths (for review see[2]). Neuroblastoma shows wide phenotypic variability, with tumors arising in children diagnosed under the age of 18 months often spontaneously regressing with little or no treatment, but patients diagnosed at an older age or with unfavorable genomic features often showing a relentlessly progressive and widely metastatic disease pattern despite intensive, multimodal therapy (for review see[2–4]). Ninety-eight percent of low-risk neuroblastoma disease are currently cured[5], however, the survival rate for patients with high-risk neuroblastoma remains less than 50%[6]. Relapsed high-risk neuroblastoma is typically incurable[7], and thus these children require improved therapeutic options.

A major prognostic factor predicting the severity, risk, and inferior outcome for neuroblastoma patients is amplification of the proto-oncogene *MYCN*. *MYCN* amplification occurs in nearly 20% of all neuroblastomas, and approximately 50% of patients with high-risk disease[8,9]. It is a truncal genomic event, and typically stable across the spectrum of therapy and disease recurrence. MYCN, along with structural and binding homologues MYC and MYCL, are members of the MYC transcription factor family[10] and have been implicated in transcriptional regulation of proteins involved in cell growth[11], proliferation[12], and ribosome biogenesis[12]. Mounting evidence has also indicated that MYCN and MYC are functionally redundant[13–15]. However, the protein expression of MYC and MYCN appears to be mutually exclusive. For example, neuroblastoma tumors with *MYCN* amplification typically lack or have low *MYC* mRNA expression[9]. The strong influence of MYCN on the progression and

[1]Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA. [2]Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA, 19104, USA. [3]Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA. [4]Pharmacology Graduate Group, University of Pennsylvania, Philadelphia, PA, 19104, USA. [5]Cancer Genomics and Bioinformatics Laboratory, Sidney Kimmel Cancer Center, Philadelphia, Pennsylvania, 19107, USA. [6]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA. [7]Center for Data-Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA. [8]These authors contributed equally: Kristen Upton, Apexa Modi. ✉e-mail: maris@email.chop.edu; rokita@email.chop.edu

a

b

c
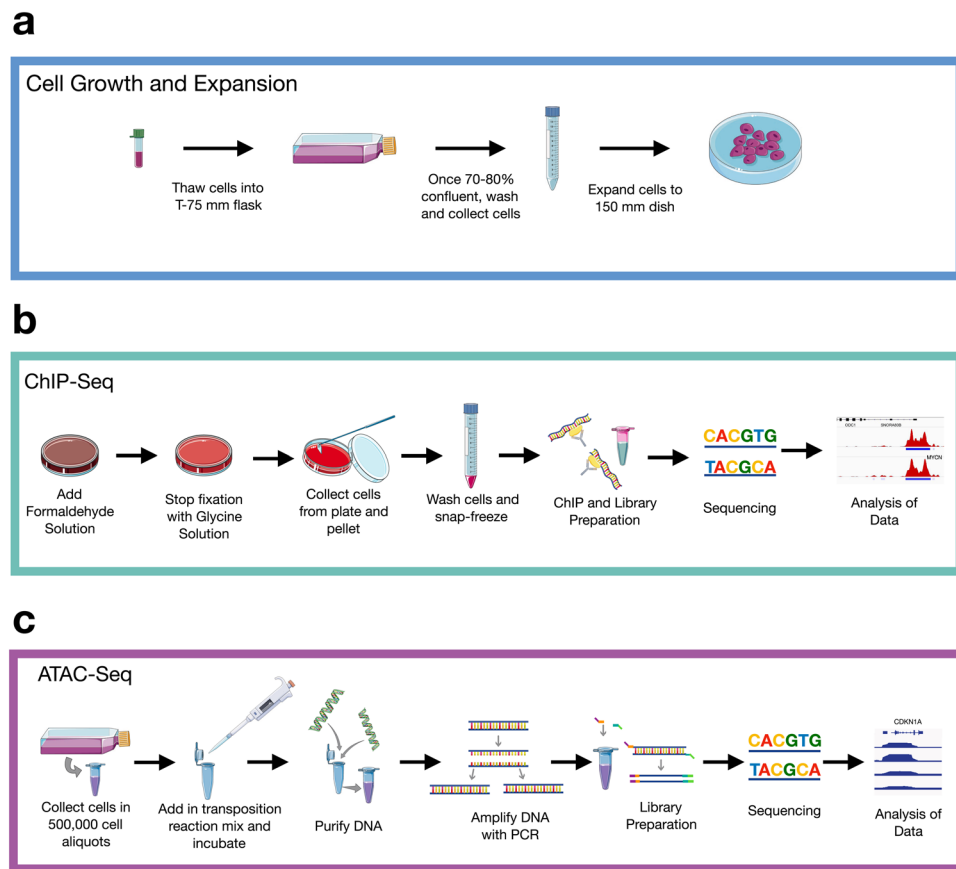
Fig. 1 Experimental workflow. (**a**) Cells were thawed, grown, and expanded until 70–80% confluency in a 150 mm dish. (**b**) For ChIP-Seq, cells were fixed, collected, and frozen (N = 1 biological replicate per cell line). Libraries were prepared, sequenced, and data analyzed. (**c**) For ATAC-Seq (n = 14 samples with n = 2 biological replicates), cells were incubated in a transposition reaction, DNA was purified, and amplified with limited PCR. Libraries were prepared, sequenced, and analyzed. Diagram was created using Servier Medical ART (https://smart.servier.com/).

metastasis of neuroblastoma makes it a key target for therapy, but due to its global transcriptional activity, it is necessary to develop a better understanding of which of its gene targets directly influence oncogenesis.

To better understand the regulatory effects of MYC family proteins in neuroblastoma, we performed ChIP-Seq data for MYCN in six neuroblastoma cell lines with *MYCN* amplification, MYC in four neuroblastoma cell lines without *MYCN* amplification, and H3K27Ac, H3K27me3, H3K4me1, and H3K4me3 histone modifications along with ATAC-Seq in all ten neuroblastoma cell lines (with ATAC data in four additional lines also reported here). All of the cell lines here also have RNA sequencing data freely available[16].

Methods
Online Table 1 summarizes which assays were performed for each cell line, and an overview of the workflow is shown in Fig. 1.

Cell growth and expansion. The cell lines used to collect this data were obtained from multiple sources: the Children's Oncology Group (COG) Cell Culture and Xenograft Repository at Texas Tech University Health Sciences Center (www.cccells.org), the American Type Culture Collection (Manassas, MA), or the Children's Hospital of Philadelphia (CHOP) cell line bank. All of the cell growth and preparations were done at CHOP. The neuroblastoma cell lines were cultured using media (Table 1) and methods as previously described[16]. Briefly, cells were thawed by floating in a 37 °C water bath for 2–3 minutes. Cells were then added to a 15 mL conical tube, containing 5 mL of the appropriate growth media, and centrifuged at $300 \times g$ for 5 minutes at room temperature (RT). Media was then carefully aspirated off, and the pelleted cells were resuspended in 1 mL of media before being transferred to a 75 mm2 flask containing 10 mL of growth media. Cells were incubated at 37 °C with a 5.0% CO_2 concentration. When cells reach 70–80% confluency, media was aspirated off and cells were gently washed with 1X PBS. Following aspiration of the PBS, 3 mL of the appropriate detachment solution (noted in Table 1) was added and the flask was incubated at 37 °C for 2–5 minutes. Cells were then gathered by tilting the plate at a 45° angle and washing with at least 4 mL of the appropriate growth media, and transferred to a 15 mL conical. After centrifugation for 5 minutes at $300 \times g$. Media was aspirated off, and the pellet was resuspended in 1 mL of growth media and transferred to a 150 mm cell culture dish containing 19 mL of growth media. Cells were incubated at

Cell Line	Organism	Cancer Histology	Matched PDX	MYCN Status	Growth Media	Detachment Method
COG-N-415	Homo sapiens	Neuroblastoma	COG-N-415x	Amplified	IMDM, 20% FBS, 2 mM L-glutamine, 1:1000 ITS Premix Supplement	0.02% Versene
KELLY	Homo sapiens	Neuroblastoma	N/A	Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
LA-N-5	Homo sapiens	Neuroblastoma	N/A	Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
NB-1643	Homo sapiens	Neuroblastoma	NB-1643	Amplified	IMDM, 20% FBS, 1% Penicillin/ Streptomycin, 2 mM L-glutamine	0.02% Versene
NB-69	Homo sapiens	Neuroblastoma	N/A	Non-Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
NB-LS	Homo sapiens	Neuroblastoma	N/A	Non-Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
NGP	Homo sapiens	Neuroblastoma	N/A	Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
SK-N-AS	Homo sapiens	Neuroblastoma	SK-N-AS (xenograft)	Non-Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.05% Trypsin/EDTA
SK-N-BE(2)-C	Homo sapiens	Neuroblastoma	N/A	Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
SK-N-FI	Homo sapiens	Neuroblastoma	N/A	Non-Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.02% Versene
SK-N-SH	Homo sapiens	Neuroblastoma	N/A	Non-Amplified	RPMI 1640, 10% FBS, 1% Penicillin/ Streptomycin, 2 mM L-Glutamine	0.05% Trypsin/EDTA

Table 1. Neuroblastoma cell line information. Listed are the cell lines used in this study, their MYCN amplification status, and culturing media information.

37 °C with a 5.0% CO_2 concentration until reaching 70–80% confluency. Necessary materials and reagents are listed in Online Table 2.

Immunoblotting. Whole cell lysates were prepared using a mixture of cell lysis buffer (Cell Signaling, #9803), PSMF (Cell Signaling, 8553S), Phosphatase Inhibitor Cocktail 2 (P5726, Sigma Aldrich), Phosphatase Inhibitor Cocktail 3 (P0044, Sigma Aldrich), and PBS (Gibco, 14190–136). Cells were resuspended in lysis buffer and kept on ice for 15 minutes. Cells were then spun at *14,000 × g* at 4 °C for 15 minutes. The supernatant was collected and protein concentration was quantified using the Pierce BCA Protein Assay kit (Thermo Scientific, #23225). Next, 20 µg of protein was loaded using 4X Laemmli sample buffer (BioRad, #1610747) and separated on a 4–15% Criterion™ TGX™ Precast Midi Protein Gel (#5671085), and transferred to an Immobilon Membrane (Cat No. IPVH00010, 0.45 µm pore size). The membrane was blocked in 5% non-fat milk in Tris-buffered saline and Tween-20 (TBS-T) at room temperature for one hour. Incubation with primary antibody was overnight, rocking at 4 °C. Membranes were then washed three times for 10 mins in TBS-T, and then incubated with HRP-labeled Rabbit secondary antibody at room temperature for one hour (1:2000–1:5000; Millipore, AP132P). The membranes were then developed using chemiluminescence (SuperSignal West Femto, Thermo Fischer Scientific). The primary antibodies used were: N-MYC (1:1000; Cell Signaling, #9405S), MYC (1:800; Cell Signaling #5605), and β-Actin (1:5000; Cell Signaling, #4967S).

ChIP-Seq protocol. The ChIP-Seq Protocol is separated into four sections: Cell Fixation, Chromatin Immunoprecipitation (ChIP) and Library Preparation, Library Sequencing, and ChIP-Seq Analysis. Of note, the MYCN ChIP-Seq for Kelly and NGP cell lines was performed using a varied procedure and is noted in a separate section within this protocol. Necessary materials and reagents are listed in Online Table 2.

Cell fixation.   Cells were grown as described in Cell Growth and Expansion section of protocol to 70–80% confluence in 150 mm tissue culture plates in 20 mL of media. The Formaldehyde solution (Online Table 2) was freshly prepared. Cells were removed from incubation and 1/10th of the growth media volume of the Formaldehyde Solution was added to the existing media in the plate (i.e. if the current volume of the plate is 20 mL of media, 2 mL of Formaldehyde Solution would be added). The solution was gently swirled, and then rocked at RT for 15 minutes. To stop the fixation, 1/20th the current volume of the Glycine Solution (Online Table 2) was added to the plate (i.e. if the current volume in the plate is 22 mL then 1.1 mL of Glycine Solution should be added). The plate was gently swirled to mix, and then allowed to sit at RT for 5 minutes. Following this incubation, a cell scraper was used to collect the cells, and then all cells and solution were transferred to a 50 mL conical on ice. From this point forward, all samples were kept on ice. The 50 mL conical was centrifuged at 800 × g at 4 °C for 10 minutes to pellet the cells. Supernatant was removed and discarded, and the cells were resuspended with 10 mL of chilled, sterile PBS. Centrifugation of the tube at 800 × g at 4 °C for 10 minutes was repeated. The supernatant was removed and discarded, and the cells were resuspended with 10 mL of chilled, sterile PBS with 100 uL of PMSF. The tube was centrifuged at 800 × g at 4 °C for 10 minutes, the supernatant was removed, and then the cells were snap frozen on dry ice and stored at −80 °C. The cells were then shipped to Active Motif on dry ice following the instructions listed at on the Sample Submission Form, downloaded from www.activemotif. com/sample-submission.

ChIP and library preparation by active motif.   Chromatin immunoprecipitation was completed by Active Motif. Full methods are proprietary. Chromatin was isolated using a lysis buffer and membranes were disrupted with a dounce homogenizer. The lysates were then sonicated with Active Motif's EpiShear probe sonicator (#53051) and cooled sonication platform (#53080) to an average fragment length 300–500 bp. A portion of the sample was collected as the Input DNA, treated with RNase, proteinase K, and incubated to reverse crosslinking. The DNA was then collected by ethanol precipitation. The Input DNA was resuspended and concentration was quantified by a NanoDrop spectrophotometer. Extrapolation of this concentration to the original chromatin volume allowed for quantitation of the total chromatin yield. Aliquots of the fixed chromatin were used in the immunoprecipitation were precleared with protein A agarose beads (Invitrogen, #15918014). Genomic DNA regions of interest were isolated using specific ChIP antibodies (Online Table 2). Antibody DNA complexes were isolated using additional protein A agarose beads, and the crosslinked DNA, antibody, and bead complexes were washed. The cross-linked DNA was eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Reverse crosslinking was done in an overnight incubation at 65 °C, and ChIP DNA was purified with a phenol-chloroform extraction and ethanol precipitation.

Illumina sequencing libraries were prepared from the ChIP and Input DNAs using the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation using Active Motif's custom liquid handling robotics pipeline. Samplers were amplified with a 15 cycle PCR amplification and then quantified before being shipped to the Jefferson Cancer Genomics Laboratory at the Kimmel Cancer Center for sequencing.

MYCN ChIP-Seq: Kelly and NGP cell lines.   Chromatin immunoprecipitation was performed on adherent cells as described in Bosse *et al.*[17]. Of note, a different MYCN antibody was used than listed in Bosse *et al.*, 2017 (Santa Cruz B8.4B, sc-53993). Cells were grown as described in Cell Growth and Expansion section of protocol to 70–80% confluence in 150 mm tissue culture plates in 20 mL of media. To the existing media, 415 mL of 37% formaldehyde (final concentration of 0.75%) was added, and rocked for 10 minutes at RT. To this, 1.5 mL of 2.5 M glycine (Online Table 2) (final concentration of 0.18 M) was added to inactivate the formaldehyde, and the plate was rocked for an additional 5 min. Cells were lysed with a volume of FA Lysis Buffer (Online Table 2) equivalent to 5 pellet volumes. Beads were washed 3 times in ChIP Wash Buffer (Online Table 2) and one time with Final Wash Buffer (Online Table 2). Libraries were constructed using NEB Ultra Kit following the manufacturer's instructions. Libraries were sequenced as single-end, 50 bp reads on a MiSeq to a depth of ~50 M reads by the Children's Hospital of Philadelphia Nucleic Acid and PCR Core.

ChIP library sequencing for ChIP.   Sequencing was conducted by the Jefferson Cancer Genomics Laboratory at the Kimmel Cancer Center. Samples were quality control tested using an Agilent High Sensitivity Screen Tape to determine average fragment length. The concentration of each library was measured using a High Sensitivity Qubit Quantification kit, and samples were diluted to an appropriate amount for the loading protocol (4 nM or less). Samples were normalized to the same nanomolar concentration, and libraries were pooled together in equal amounts. Samples were diluted to 1.51 pM in Low EDTA TE Buffer. Samples were then sequenced as single-end, 75 bp reads to an average depth of ~30 M reads on a NextSeq. 500.

ATAC-Seq protocol.   The following ATAC-Seq protocol was adapted from Buenrostro, *et al.*[18]. This protocol consists of four parts: Cell Preparation, Transposition Reaction and Purification, PCR Amplification, qPCR, and Library Preparation. Primer 1 and Primer 2 were custom synthesized by Integrated DNA Technologies (IDT), using sequences provided in Buenrostro, *et al.*, 2015. Note: ATAC-Seq for NB-69 and NGP was performed using a slightly varied procedure and is noted in a separate section.

Cell preparation.   Cells were grown as described in Cell Growth and Expansion section of protocol to 70–80% confluence in a 75 mm² tissue culture flasks in 10 mL of media. Following detachment and pelleting, cells were resuspended in 1.0 mL of the appropriate growth media. Cells were triturated until they were in a homogenous single-cell suspension. Using an automated cell counter, the volume for 500,000 cells was determined and aliquoted into a sterile 1.5 mL Eppendorf tube containing 500 μL of sterile 1X PBS. Cells were centrifuged at 500 × g for 5 minutes at 4 °C. The supernatant was carefully aspirated, and the cells were resuspended in 500 mL of sterile

Primer Name	Sequence	Concentration	Purification
Primer 1 (Ad1)	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG	25 nm	STD
Primer2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT	25 nm	STD
Primer2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT	25 nm	STD

Table 2. Primer sequences used in ATAC-Seq. ATAC-Seq primer sequences. Primer sequences used in ATAC-Seq to amplify transposed DNA.

1X PBS. Centrifugation was repeated and cells were resuspended in 500 mL of cold lysis buffer by gently pipetting up and down, and then immediately centrifuged at $500 \times g$ for 10 minutes at 4 °C. The supernatant was carefully removed and discarded. The pellet was immediately resuspend in 50 μL of nuclease free water by gently pipetting up and down, and the protocol immediately continued on to Transposition Reaction and Purification section.

Transposition reaction and purification. The pellet was placed on ice. The following reagents were prepared and combined: transposition reaction mix (25 μL TD (2X reaction buffer from Nextera Kit), 2.5 μL TDE1 (Nextera Tn5 Transposase from Nextera Kit), 17.5 μL nuclease-free water, and 5.0 μL of resuspended DNA/protein from the final step in *Cell Preparation* (resuspended pellet in 50 μL of nuclease free water). The transposition reaction was incubated in a thermocycler at 37 °C for 30–35 minutes. The reaction was immediately purified using Qiagen MinElute PCR Purification Kit, and the transposed DNA was eluted in 10.5 μL of elution buffer (Buffer EB from the MinElute Kit consisting of 10 mM Tris-Cl (pH 8)). The eppendorf tube containing purified DNA was para-filmed, and stored at −20 °C. *NOTE:* This can act as a good stopping point, however these DNA fragments are not PCR amplifiable if melted at this point.

PCR amplification. Primer sequences are shown in Table 2. To amplify the Transposed DNA, the following were combined into a 0.2 mL PCR tube: 10 μL transposed DNA, 10 μL nuclease-free H$_2$O, 2.5 μL 25 mM PCR Primer 1 (Ad1), 2.5 μL 25 mM Barcoded PCR Primer 2 (Ad2.X, X being the unique number of samples), and 25 μL NEBNext High-Fidelity 2X PCR Master Mix. The thermal cycle was as follows:

$$\begin{array}{rl} \text{1 cycle:} & \text{5 min 72 °C} \\ & \text{30 sec 98 °C} \\ \text{5 cycles:} & \text{10 sec 98 °C} \\ & \text{30 sec 63 °C} \\ & \text{1 min 72 °C} \end{array}$$

The five minute extension in the first cycle is critical to allow extension on both ends of the primer after transposition, thereby generating amplifiable fragments. This ensures that downstream quantitative PCR (qPCR) quantitation will not change the complexity of the original library.

qPCR. To reduce the GC and size bias in PCR, the appropriate number of PCR cycles (N) was determined using qPCR, allowing us to stop prior to saturation. The samples were kept in the thermocycler following the PCR Amplification reaction, and the qPCR side reaction was run. In a 0.2 mL PCR tube the following were added: 5 μL

of DNA PCR amplified DNA, 2 μL of nuclease free H_2O, 1 μL of 6.25 mM Custom Nextera PCR Primer (Ad1), 1 μL of 6.25 mM Custom Nextera PCR Primer 2 (Ad2.X), 1 μL 9X SYBR Green I, and 5 μL NEBNext High-Fidelity 2X PCR Master Mix. This sample was run in the qPCR instrument with the following cycles:

1 cycle:	30 sec 98 °C
20 cycles:	10 sec 98 °C
	30 sec 63 °C
	1 min 72 °C

To calculate the additional number of cycles needed, a linear plot of Rn versus cycle was generated. This determined the cycle number (N) that corresponds to one-third of the maximum fluorescent intensity.

The remaining 45 mL PCR reaction was run to the cycle number (N) determined by qPCR. Cycles are as follows:

1 cycle:	30 sec 98 °C
N cycles:	10 sec 98 °C
	1 min 72 °C
	30 sec 63 °C

The amplified library was purified using Qiagen MinElute PCR Purification Kit after the additional PCR. The purified library was eluted in 20 μL of elution buffer (Buffer EB from the MinElute Kit consisting of 10 mM Tris-Cl (pH 8)). It is important to make sure that the column is dry prior to adding elution buffer to avoid ethanol contamination of final library. The amplified library was purified using AMPure XP beads at a 1.8x ratio to get rid of adapter dimers, using 80% ethanol for the wash steps. Sample was eluted in 50 μL of nuclease free H_2O. The concentration of the DNA eluted from the column should be about 30 nM.

Library preparation. The quality of the purified libraries was assessed using a Bioanalyzer High-Sensitivity DNA Analysis kit (Agilent). If libraries contained predominant peaks around 1000 bp, SPRI beads were used to remove these fragments. This was accomplished by first, with a new vial of SPRI beads, performing size selection with various ratios to ensure larger peaks are removed. For example, ratios could include 0.4X, 0.45X, 0.5X. Choose the ratio that removes 1000 bp fragments, but leaves 800 bp fragments. Libraries were eluted in 20 μL of nuclease-free water, and sequenced as described below.

Sequencing for ATAC-Seq by Beijing Genomics International (BGI). Sequencing was conducted by Beijing Genomics International at the Children's Hospital of Philadelphia. Samples were quality control tested using an Agilent High Sensitivity Screen Tape to confirm average fragment sizes were ~180, 380, 580, 780, and 980 bp. The concentration of each library was measured using a High Sensitivity Qubit Quantification kit, to ensure they were 5.5 nM. Samples were normalized and libraries were pooled together in equal amounts. Samples were then sequenced as paired-ends, 100 bp to an average depth of 80 M reads on a HiSeq. 2500.

ATAC-Seq NB-69 and NGP cell lines via Active Motif. Cells were grown as described in Cell Growth and Expansion section of protocol to 70–80% confluence in a 75 mm² tissue culture flasks in 10 mL of media. Following detachment and pelleting, cells were resuspended in 1.0 mL of the appropriate growth media. Cells were triturated into a homogenous single-cell suspension. Using an automated cell counter, the volume for 100,000 cells was determined, and aliquoted into a sterile 1.5 mL eppendorf tube containing 500 μL of sterile 1X PBS. Cells were then centrifuged at $500 \times g$ for 5 minutes at 4 °C. The supernatant was carefully aspirated off, and the cells were resuspended in 500 μL of growth media with 5% DMSO. The sample was transferred to a 1.7 mL microfuge tube on ice. Cells were frozen with a slow cooling to minimize cell lysis. Samples were shipped on dry ice to Active Motif (1914 Palomar Oaks Way, Ste 150, Carlsbad, CA 92008) following the instructions listed at on the Sample Submission Form, downloaded from www.activemotif.com/sample-submission. Samples were prepared and sequenced following Active Motif's ATAC-Seq proprietary protocol. Cells were thawed in a 37 °C water bath, pelleted, washed with cold PBS, and tagmented as previously described[18], with some modifications based on[19]. Cell pellets were resuspended in lysis buffer, pelleted, and tagmented using the enzyme buffer provided in the Nextera Library Prep Kit (Illumina). Tagmented DNA was then purified using the MinElute PCR purification kit (Qiagen), amplified with 10 cycles of PCR, and purified using Agencourt AMPure SPRI beads (Beckman Coulter). The resulting material was quantified using the KAPA Library Quantification Kit for Illumina platforms (KAPA Biosystems) and sequenced with PE42 sequencing on the NextSeq. 500 sequencer (Illumina).

ChIP-Seq data analysis. FASTQ quality was assessed using FastQC v0.11.4 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and sequences were adapter- and quality-trimmed using default parameters for Trim Galore v.0.4.0 and CutAdapt v.1.12[20,21]. MultiQC v1.4 was used to aggregate FastQC results across all samples, with the report available on Figshare[22]. Since multiple sequencers were used, FASTQ phred sequencing scores[23] were calculated using a perl script (https://raw.githubusercontent.com/douglasgscofield/bioinfo/master/scripts/phredDetector.pl). This value was used as input into the alignment algorithm. The bwa v.0.7.12 samse[24] was used to align the reads to hg19 reference genome and Picard tools v.2.17.9-SNAPSHOT[25] was used to remove duplicates. Fragment sizes were estimated using MaSC 1.2.1[26] and these values were used as input into the –extsize argument of MACS2 v.2.1.1[27] for narrow peak calling (transcription factors) or broad peak calling (histone marks). Broad peaks were called significant using a q-value (minimum False Discovery Rate) cut off of 0.10 and narrow peaks at a q-value cutoff of 0.05. Results were returned in units of signal per million reads to get

normalized peak values. Repetitive centromeric, telomeric and satellite regions known to have low sequencing confidence were removed using blacklisted regions defined by the ENCODE project: http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/wgEncodeHg19ConsensusSignalArtifactRegions.bed.gz. The resulting filtered peakfiles were used as input into Homer v4.10.4 for gene annotation and motif analysis.

ATAC-Seq data analysis.   Samples were quality-controlled and trimmed as described in Chip-Seq Analysis. FASTQ files were aligned using bwa aln for BGI samples (100 bp reads) and bwa mem for Active Motif samples (42 bp reads). Reads with mapping quality <10 were discarded. Biological duplicate BAMs were merged using Picard v.2.17.9-SNAPSHOT. Broad peaks were called using –extsize 200, –shift 100, –nomodel. Results were returned in units of signal per million reads to get normalized peak values. Finally, repetitive centromeric, telomeric and satellite regions known to have low sequencing confidence were removed using merged blacklisted regions defined by the ENCODE project: http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/wgEncodeHg19ConsensusSignalArtifactRegions.bed.gz.

ChIP-Seq quality control metrics.   We investigated three metrics to assess ChIP-seq quality. To calculate enrichment of reads within peaks we determined the FRiP score using deeptools2[28]. The FRiP score is defined as the fraction of reads that fall within a peak divided by the total number of reads. To measure read enrichment independent of peak calling we calculated the NSC (normalized strand cross-correlation) and the RSC (relative strand cross-correlation) using phantompeakqualtools[29,30] as part of the ENCODE ChIP-seq processing pipeline. All ChIP-Seq data passed quality control and results are reported in Online Table 3.

ATAC-Seq quality control metrics.   To compare reproducibility between ATAC-seq biological replicates we performed irreproducible discovery rate (IDR) analysis using scripts downloaded from https://github.com/nboley/idr. Peaks passing the suggested threshold (IDR < = 0.05%) between two replicates were kept. The ratio between the number of peaks between true replicates (Nt) and pooled pseudoreplicates (Np) was calculated. In accordance with ENCODE guidelines, we confirmed that at least 50% of true replicate IDR analysis based peaks (Nt) were identified in the IDR comparison of pseudoreplicates (Np): Np/Nt < 2. A similar analysis was done with self-pseudoreplicates (N1 and N2). We confirmed that the ratio between Np/Nt or N1/N2 was <2. All ATAC-seq data passed IDR results and are reported in Online Table 4. Peakfiles resulting from IDR analysis are available from FigShare[31].

Super-enhancer calling and comparison.   Super-enhancers (SEs) were called from H3K27Ac BAM files using the default parameters of LILY (https://github.com/BoevaLab/LILY), which includes correction for copy number variation inherently present in cancer samples. Enhancers were classified into SEs, enhancers, and promoters and annotated using Homer v4.10.4. Scripts to run LILY can be found on Github (https://github.com/marislab/epigenomics-data-descriptor). SEs were also called from H3K27Ac MACS2 peaks using ROSE v.0.1 (https://bitbucket.org/young_computation/rose/src/master/) using default parameters and annotated using Homer v4.10.4. SEs which overlapped with the MYCN locus (hg19, chr2:16080683-16087129) were removed from the analysis. SE genes which we annotated as transcription factors[32] were used for comparison to two literature studies[33,34].

Heatmap preparation.   The 5,000 most significant (sorted by highest -$\log_{10}$(p-value) and -$\log_{10}$(q-value)) MYCN peaks for each of the five MYCN amplified cell line were intersected using bedtools. Heatmaps were generated for regions +/−4 kb from the transcription start site (TSS) for the 5,046 peaks common to at least four MYCN amplified cell lines. Heatmaps were created for LA-N-5 and NB-69 at loci annotated as enhancers, SEs, and promoters-TSS by LILY. All ChIP-seq heatmaps were created using deepTools 3.2.0 package plotHeatmap tool[28]. The code and parameters used to generate heatmaps can be found on GitHub (https://github.com/marislab/epigenomics-data-descriptor).

Cell line authentication.   All cell lines were STR-authenticated by Guardian Forensic Sciences (Abington, PA) using the GenePrint 24 (Promega, #B1870).

Data Records

Raw, concatenated FASTQ files were deposited in Sequence Read Archive under the SRA study accessions SRP223941[35], SRP223977[36], and SRP223942[37]. Processed BIGWIG files for all sequencing data were deposited into the Gene Expression Omnibus (GEO) under SuperSeries Accession Number GSE138315[38]. MYCN and MYC ChIP-Seq data for the Kelly and NGP cell lines were deposited into GEO under Accession Number GSE94782[39], all other MYCN and MYC ChIP-Seq were deposited under Accession Number GSE138295[40], histone ChIP-Seq data were deposited under Accession Number GSE138314[41], and ATAC-Seq data were deposited under Accession Number GSE138293[42]. Homer motif analysis and motif files are available on FigShare[22,31,43].

Technical Validation

Prior to selecting cell lines for MYCN and MYC profiling, we assessed RNA expression (Fig. 2a,b) and protein expression (Fig. 2c,d) across a subset of neuroblastoma cell lines. NB-LS, while *MYCN* non-amplified, has substantial MYCN RNA and protein expression[44], but was not chosen, as we restricted MYCN ChIP-Seq to *MYCN* amplified cell lines plus one negative control. SK-N-BE(2)-C, a *MYCN* amplified cell line, showed high *MYCN* mRNA expression, but surprisingly low protein expression, and thus was excluded. The remaining cell lines had concordant *MYCN* and *MYC* mRNA and protein expression, thus, COG-N-415, KELLY, NB-1643, LA-N-5, and NGP were chosen for MYCN ChIP-Seq while NB-69, SK-N-AS, and SK-N-SH were chosen for MYC ChIP-Seq. As additional controls, we performed MYCN ChIP-Seq in the *MYCN* non-amplified line NB-69,
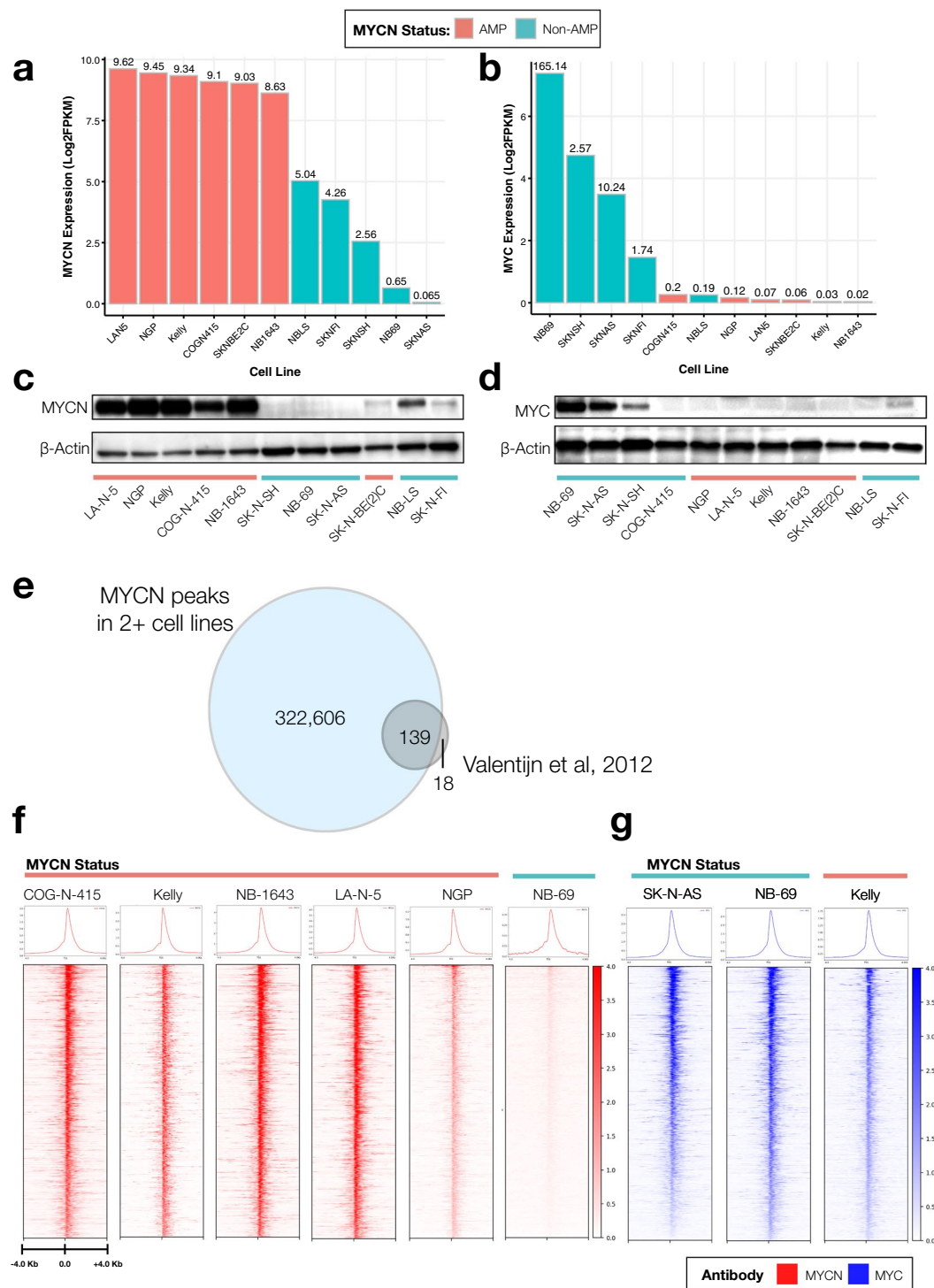
Fig. 2 Comparison of MYCN and MYC binding based on MYCN amplification status. The log$_2$ FPKM mRNA expression of *MYCN* (**a**) and *MYC* (**b**) in all neuroblastoma cell lines assayed herein. Protein expression for MYCN (**C**) and MYC (**d**) in a subset of cell lines. (**e**) Comparison of unique MYCN peaks called in at least two of the MYCN ChIP-Seq samples (blue) to known MYCN regulated genes (gray)[45] demonstrates concordance of MYCN ChIP-Seq to MYCN ChIP-ChIP. (**f**) The top 1,335 MYCN peaks (p < 0.05, q < 0.05) are plotted as ChIP-Seq heatmaps for five MYCN amplified cell lines (COG-N-415, Kelly, NB-1643, LAN-5, and NGP) and one MYCN non-amplified cell line (NB-69). All heat map densities ranges from +/−4.0 Kb from the TSS, with average signal plots shown above. (**g**) MYC ChIP-Seq heat maps for the same peaks in two MYCN non-amplified lines (SK-N-AS, and NB-69) and one MYCN amplified cell line (Kelly) showing redundant binding of MYC in non-amplified cell lines.
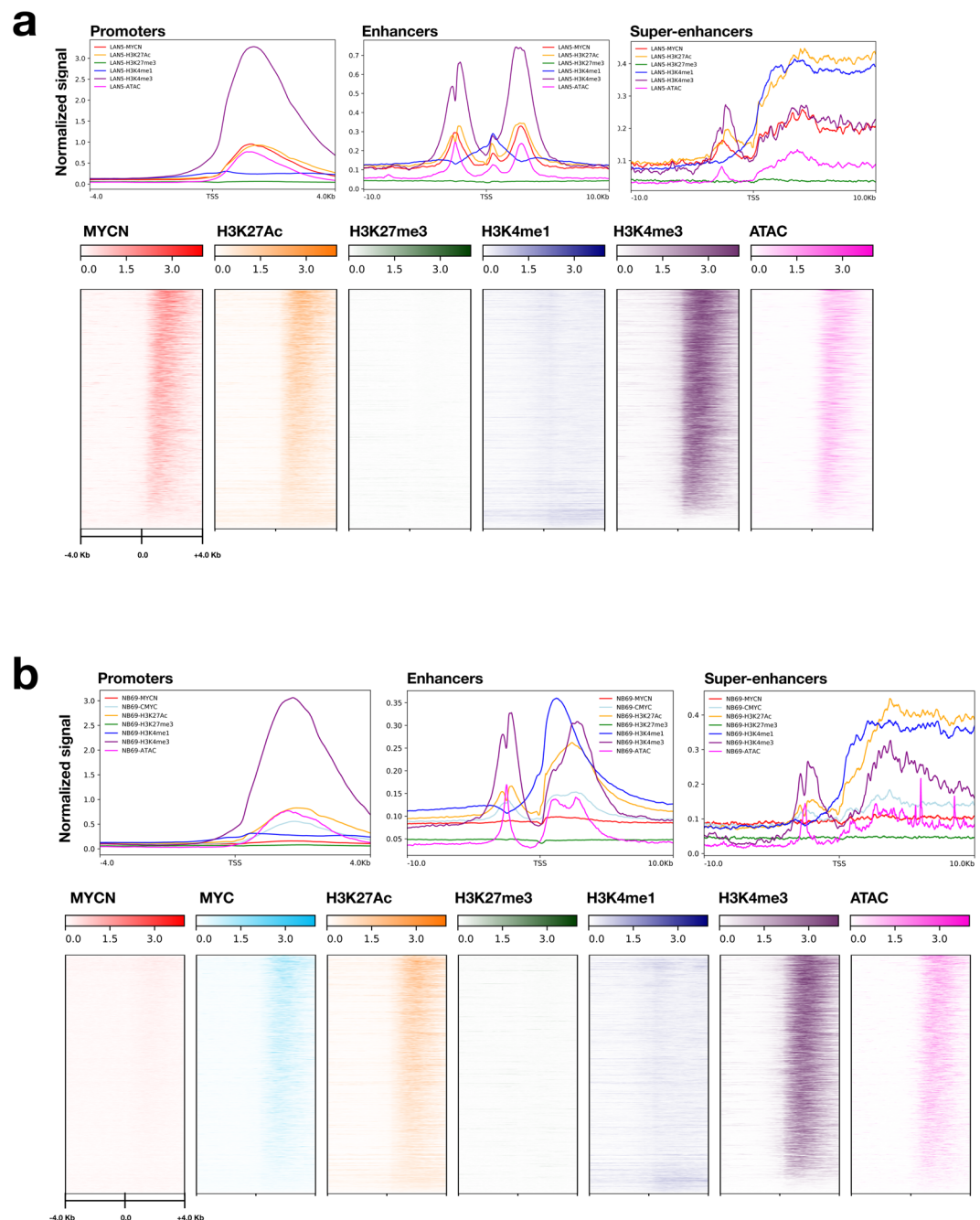
and MYC ChIP-Seq on the *MYCN* amplified cell line KELLY. To validate the MYCN and MYC ChIP-Seq antibodies, we first intersected loci bound by MYCN in two or more cell lines and of the 157 MYCN transcriptional targets previously reported using ChIP-on-ChIP[45], found 139 loci occupied by the MYCN via ChIP-Seq (Fig. 2e). Next, we integrated the top 5,000 MYCN peaks from each *MYCN* amplified cell line. We generated heatmaps for the peaks (1,335) which overlapped in all five cell lines (as defined in **Heatmap Preparation**) and depict occupancy of MYCN (Fig. 2f) and MYC (Fig. 2g) at these sites. As expected, the MYCN amplified cell lines COG-N-415, KELLY, NB-1643, LA-N-5, and NGP show similar binding profiles, while the negative control MYCN non-amplified line NB-69 depicted an absence of binding for MYCN at the same loci. Importantly, Homer motif analysis of the 34,906 target sequences bound by MYC in NB-69 were significantly enriched (Benjamini q-value < 0.001) for the canonical CACGTG e-box motif, while this motif was absent from the 112 target sequences found in the NB-69 MYCN ChIP-Seq sample. We observed MYC bound to the same loci in the MYCN non-amplified cell lines, SK-N-AS, SK-N-SH, and NB-69 as well as the *MYCN* amplified and low MYC-expressing line KELLY (Fig. 2g), and observed shared CACGTG motif binding for both MYCN and MYC in KELLY, supporting the notion of redundant functionality of MYC family protein members. To further validate both the specificity and functional redundancy of the MYCN and MYC ChIP-Seq, we assessed MYCN and MYC binding to transcriptional targets of an 18-gene MYC family (MYCN/MYC/MYCL1) activity signature[46] in KELLY (MYCN and MYC) and SKNBE(2)C (MYCN) cell lines alongside six non-MYC family core regulatory TFs (*ASCL1, GATA3, HAND2, ISL1, PHOX2B, TBX2*) from publicly-available ChIP-Seq data (*ASCL1:* GEO accession number GSE120074 and *GATA3, HAND2, ISL1, PHOX2B, TBX2:* GEO accession number GSE94824) reprocessed with our pipeline (see Methods). Supplemental Fig. 2 shows the binding patterns for four of the 18 genes: *APEX1, NME1, ENO1*, and *ODC1*. *APEX1, NME1*, and *ENO1* are not bound by the six non-MYC family core regulatory TFs (*ASCL1, GATA3, HAND2, ISL1, PHOX2B, TBX2*), while *ASCL1* shows binding at *ODC1* because it recognizes the e-box motif, CANNTG. Altogether, these data demonstrate specificity of MYCN and MYC antibodies and functional redundancy of MYCN and MYC proteins.

Next, we evaluated genome-wide binding densities of the histone antibodies and assessed open chromatin by plotting binding of one *MYCN* amplified cell line LA-N-5 (Fig. 3a), and one *MYCN* non-amplified cell line NB-69 (Fig. 3b). Of note, cell-line specific promoters are located in regions of open chromatin and strongly occupied by narrow regions of H3K4me3 and devoid of H3K27me3 and H3K4me1, as expected. The majority of promoters are also occupied by MYCN in LA-N-5 and MYC in NB-69. Enhancers have bivalent marking of MYCN, H3K4me3, H3K27Ac, open chromatin, and absence of H3K27me3. SEs are broadly marked by MYCN, H3K4me3, H3K27Ac, H3K4me1, and open chromatin.

Finally, we used our H3K27Ac ChIP-Seq data to compare SE prediction of cell line lineage in our dataset compared to those reported in two other publications describing the SE landscape in neuroblastoma (Fig. 4 and Supplemental Fig. 1). Boeva and colleagues identified 4,791 SE-associated genes in Table S3 to identify core regulatory transcriptional circuitry in neuroblastoma using 25 cell lines[33]. Four cell lines were common to our study: SK-N-BE(2)C, SK-N-FI, SK-N-AS, and NB-69. Therefore, to validate our H3K27Ac ChIP-Seq, we utilized the same algorithm (LILY, see Methods) to call SEs from our H3K27Ac data, and restricted comparison analyses to genes defined as transcription factors (TFs), as defined by core regulatory circuitry[32]. We annotated 396 of the SEs reported by Boeva and colleagues as transcription factors and found 59–85% concordance of our TF SE calls (Supplemental Fig. 1). While a majority of SEs called in each of our cell lines was concordant with Boeva and colleagues, the high variance in total number of SEs called likely stems from the diversity of cell lines in both studies, as well as pipeline processing and filtering parameters. We were unable to directly compare methods without their code and raw data readily available. Thus, we additionally compared our TF SE calls to those from an independent neuroblastoma study[34] which used the ROSE algorithm (see Methods) and reported smaller SE genesets (Online Table 5) driving the lineage-specific mesenchymal (MES, N = 20 TFs) and adrenergic (ARDN, N = 18 TFs) subtypes. To mimic the analysis performed by van Gronigen and colleagues, we ran ROSE on our H3K27Ac ChIP-Seq data and removed any peaks which overlapped the *MYCN* locus (see Methods) to account for false SE calls due to *MYCN* amplification. There were no common neuroblastoma cell lines between van Gronigen and colleagues study and the lines used in our study. We assessed the number of MES or ADRN SE-associated TFs detected in each of our study and found between five and eight ADRN SEs were detected using ROSE (Fig. 4a) and between five and 11 ADRN SEs were detected using LILY (Fig. 4b). SK-N-SH has a known MES subtype; its subclone, SH-SY-5Y, was profiled as MES by van Gronigen and colleagues. Combining the calls, we were able to significantly (Fisher's exact test, p < 0.05) validate ADRN subtypes in eight of the ten cell lines we profiled (Fig. 4c). Interestingly, SK-N-AS contains SEs from both subtypes and thus may reflect a heterogeneous cell line. Specific SEs are reported per algorithm per cell line in Online Table 5. As further validation, we re-analyzed publicly-available SK-N-SH H3K27Ac (Biosample SAMN05733860, Run SRR5338927) and SK-N-SH Input (Biosample SAMN05733844, Run SRR5471111) ChIP-Seq data (GEO accession GSM2534162) using the same peak-calling and SE pipelines used on our data (see Methods). We observed enhancer binding (H3K27Ac) and open chromatin (ATAC) at the same loci we observe strong MYC occupancy (Supplemental Fig. 1A). Further, we assessed concordance of SEs called in SK-N-SH with those previously reported and found 76% of TF SEs called in SK-N-SH in common with those from Boeva, et. al, similar to our findings.
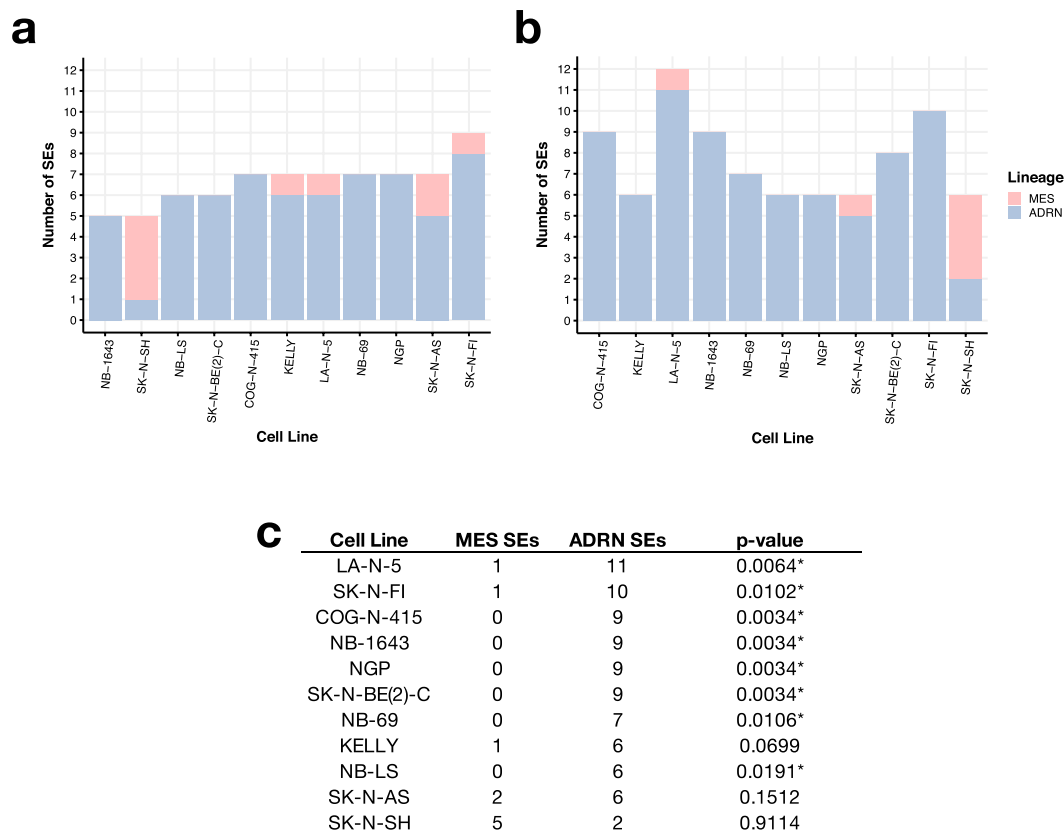
Together, we have validated both MYCN and MYC ChIP-Seq antibodies for use in ChIP-Seq, as well as genome-wide occupancy profiles for histone markers and open chromatin across a cohort of neuroblastoma cell lines. We ran two algorithms (LILY and ROSE) and compared our data to two independent datasets to validate reproducibility of lineage-specific SEs in neuroblastoma cell lines. Finally, we demonstrate integration of publicly-available H3K27Ac data from SK-N-SH with our MYC ChIP-Seq and ATAC-Seq data, and show reproducibility of SE calls between the publicly-available data and two independent reports. These data should be a valuable resource to the childhood cancer and MYC research communities.

Fig. 3 Validation of ChIP-Seq promoter, enhancer, and open chromatin occupancy. Binding densities of MYCN, MYC, histone antibodies, and open chromatin for promoter regions of the MYCN-amplified cell line LA-N-5 (**a**) and the MYCN non-amplified cell line NB-69 (**b**) are depicted ($+/-4.0$ Kb from gene TSS) and distinct profiles are shown for promoters ($+/-4.0$ Kb from gene TSS), enhancers ($+/-10.0$ Kb from gene TSS), and SEs ($+/-10.0$ Kb from gene TSS). For LA-N-5: $N_{promoters-TSS} = 4,662$, $N_{enhancers} = 25,601$, $N_{SE} = 826$, and for NB-69: $N_{promoters-TSS} = 4,718$, $N_{enhancers} = 31,769$, $N_{SE} = 667$.

Usage Notes

Here, we provide raw FASTQ and bigwigs for a comprehensive, validated ChIP-Seq (MYCN, MYC, H3K27Ac, H3K27me3, H3K4me3, and H3K4me1) and ATAC-Seq neuroblastoma cell line dataset which can be coupled with our previous RNA-Seq profiling dataset[16] to interrogate novel transcriptional regulation in this disease. For example, the H3K27me3 ChIP-Seq can be used to identify genes being repressed via the PRC2 complex, while H3K27Ac and H3K4me1 ChIP-Seq can be used to interrogate promoter-enhancer mechanisms. CSI-ANN can be used to integrate histone ChIP-Seq data to predict regulatory DNA segments[47], and IM-PET can use the results from CSI-ANN to predict enhancer-promoter interactions without the need for Hi-C data[13]. Additionally, chromatin states can be inferred[48,49], and these data can be later integrated with whole exome or genome sequencing

a

b

Lineage
- MES
- ADRN

c

Cell Line	MES SEs	ADRN SEs	p-value
LA-N-5	1	11	0.0064*
SK-N-FI	1	10	0.0102*
COG-N-415	0	9	0.0034*
NB-1643	0	9	0.0034*
NGP	0	9	0.0034*
SK-N-BE(2)-C	0	9	0.0034*
NB-69	0	7	0.0106*
KELLY	1	6	0.0699
NB-LS	0	6	0.0191*
SK-N-AS	2	6	0.1512
SK-N-SH	5	2	0.9114

Fig. 4 Super-enhancer calls correctly determine neuroblastoma cell line lineage (**a**) Total number of lineage-specific SEs called from ROSE and (**b**) LILY per cell line (ADRN = adrenergic, MES = mesenchymal subtype). (**c**) Total number of lineage-specific SEs called when combining ROSE and LILY results. One-tailed fisher's exact test p-values are listed (*denotes significance at an α level of 0.05).

data or genome-wide association studies to identify molecular alterations driving transcriptional regulatory marked by histone marks or open chromatin.

All data are openly-available from GEO as described in the Data Records section.

Code availability
Code for IDR analysis, SE calling, filtering, heatmap generation is available on Github[50].

References
1. ACCO. Childhood Cancer Statistics, https://www.acco.org/childhood-cancer-statistics/ (2016).
2. Matthay, K. K. *et al*. Neuroblastoma. *Nat Rev Dis Primers* **2**, 16078 (2016).
3. Maris, J. M. Recent advances in neuroblastoma. *N. Engl. J. Med.* **362**, 2202–2211 (2010).
4. Ambros, P. F. *et al*. International consensus for neuroblastoma molecular diagnostics: report from the International Neuroblastoma Risk Group (INRG) Biology Committee. *Br. J. Cancer* **100**, 1471–1482 (2009).
5. Twist, C. J. *et al*. Maintaining Outstanding Outcomes Using Response- and Biology-Based Therapy for Intermediate-Risk Neuroblastoma: A Report From the Children's Oncology Group Study ANBL0531. *J. Clin. Oncol.*, JCO1900919 (2019).
6. Pinto, N. R. *et al*. Advances in Risk Classification and Treatment Strategies for Neuroblastoma. *J. Clin. Oncol.* **33**, 3008–3017 (2015).
7. Simon, T. *et al*. Treatment and outcomes of patients with relapsed, high-risk neuroblastoma: Results of German trials. *Pediatr. Blood Cancer* **56**, 578–583 (2011).
8. Gherardi, S., Valli, E., Erriquez, D. & Perini, G. MYCN-mediated transcriptional repression in neuroblastoma: the other side of the coin. *Front. Oncol.* **3**, 42 (2013).
9. Rickman, D. S., Schulte, J. H. & Eilers, M. The Expanding World of N-MYC-Driven Tumors. *Cancer Discov.* **8**, 150–163 (2018).
10. Pistoia, V., Morandi, F., Pezzolo, A., Raffaghello, L. & Prigione, I. MYCN: from oncoprotein to tumor-associated antigen. *Front. Oncol.* **2**, 174 (2012).
11. Coller, H. A. *et al*. Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc. Natl. Acad. Sci. USA* **97**, 3260–3265 (2000).
12. Ji, H. *et al*. Cell-type independent MYC target genes reveal a primordial signature involved in biomass accumulation. *Plos One* **6**, e26057 (2011).
13. He, S., Liu, Z., Oh, D.-Y. & Thiele, C. J. MYCN and the epigenome. *Front. Oncol.* **3**, 1 (2013).
14. Malynn, B. A. *et al*. N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes Dev.* **14**, 1390–1399 (2000).

15. Chappell, J. & Dalton, S. Roles for MYC in the establishment and maintenance of pluripotency. *Cold Spring Harb. Perspect. Med.* **3**, a014381 (2013).
16. Harenza, J. L. *et al.* Transcriptomic profiling of 39 commonly-used neuroblastoma cell lines. *Sci Data* **4**, 170033 (2017).
17. Bosse, K. R. *et al.* Identification of GPC2 as an Oncoprotein and Candidate Immunotherapeutic Target in High-Risk Neuroblastoma. *Cancer Cell* **32**, 295–309.e12 (2017).
18. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–9 (2015).
19. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
20. Krueger, F. *TrimGalore*. (Github).
21. Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
22. Rokita, J. L. & Modi, A. MutiQC Report for Neuroblastoma Cell Line ATAC-Seq and ChIP-Seq. *figshare* https://doi.org/10.6084/m9.figshare.11929881 (2020).
23. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
25. Alec Wysoker, K. T., McCowan, M., Homer, N. & Fennell, T. A set of tools (in Java) for working with next generation sequencing data in the BAM. https://broadinstitute.github.io/picard/.
26. Ramachandran, P., Palidwor, G. A., Porter, C. J. & Perkins, T. J. MaSC: mappability-sensitive cross-correlation for estimating mean fragment length of single-end short-read sequencing data. *Bioinformatics* **29**, 444–450 (2013).
27. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
28. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–5 (2016).
29. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
30. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).
31. Rokita, J. L. & Modi, A. Neuroblastoma ATAC-Seq IDR peakfiles. *figshare* https://doi.org/10.6084/m9.figshare.11929995 (2020).
32. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
33. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).
34. van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat. Genet.* **49**, 1261–1266 (2017).
35. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP223941 (2019).
36. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP223977 (2019).
37. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP223942 (2019).
38. Rokita, J. L. Epigenomic profiling of neuroblastoma cell lines. *Gene Expression Omnibus* https://identifiers.org/geo:GSE138315 (2019).
39. Sussman, R. T. & Rathi, K. S. Genome-wide binding of MYCN protein in MYCN-amplified neuroblastoma cell lines. *Gene Expression Omnibus* https://identifiers.org/geo:GSE94782 (2017).
40. Rokita, J. L., Upton, K., Patel, K. & Maris, J. M. MYCN and MYC ChIP-Seq profiling in neuroblastoma cell lines. *Gene Expression Omnibus* https://identifiers.org/geo:GSE138295 (2019).
41. Rokita, J. L., Upton, K. & Maris, J. M. Histone ChIP-Seq of neuroblastoma cell lines. *Gene Expression Omnibus* https://identifiers.org/geo:GSE138314 (2019).
42. Rokita, J. L., Upton, K. & Maris, J. M. ATAC-Seq of neuroblastoma cell lines. *Gene Expression Omnibus* https://identifiers.org/geo:GSE138293 (2019).
43. Rokita, J. L. Homer Motif Analysis. *figshare* https://doi.org/10.6084/m9.figshare.11929872.v6 (2020).
44. Cohn, S. L. *et al.* Prolonged N-myc protein half-life in a neuroblastoma cell line lacking N-myc amplification. *Oncogene* **5**, 1821–1827 (1990).
45. Valentijn, L. J. *et al.* Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. *Proc. Natl. Acad. Sci. USA* **109**, 19190–19195 (2012).
46. Jung, M. *et al.* A Myc Activity Signature Predicts Poor Clinical Outcomes in Myc-Associated Cancers. *Cancer Res.* **77**, 971–981 (2017).
47. Firpi, H. A., Ucar, D. & Tan, K. Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinformatics* **26**, 1579–1586 (2010).
48. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
49. Sohn, K.-A. *et al.* hiHMM: Bayesian non-parametric joint inference of chromatin state maps. *Bioinformatics* **31**, 2066–2074 (2015).
50. Patel, K., Rokita, J. L. & Modi, A. *marislab/epigenomics-data-descriptor: Release v1.0.0. Zenodo* https://doi.org/10.5281/zenodo.3701677 (2020).

Acknowledgements

Author contributions

Conceptualization: J.L.R., K.U. and A.M. Methodology: J.L.R., K.U., A.M., K.P., R.T.S., R.N.A., P.F. and G.P.W. Validation: J.L.R., K.U., A.M., K.P., K.L.C. and A.M. Formal Analysis: J.L.R., K.U., K.P. and A.M. Investigation: J.L.R., K.U., K.P. and A.M. Resources: J.M.M., J.L.R. and S.J.D. Data Curation: J.L.R., K.P. and G.I.S. Writing - Original Draft: J.L.R., K.U., K.P. and A.M. Writing - Review & Editing: J.L.R., K.L.C., K.U., G.P.W., J.M.M. and A.M. Visualization: J.L.R., K.U., K.P. and A.M. Supervision: J.L.R. Funding Acquisition: J.L.R. and J.M.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41597-020-0458-y.

Correspondence and requests for materials should be addressed to J.M.M. or J.L.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.