SCIENTIFIC DATA

OPEN

Data Descriptor: Transcriptomic profiling of 39 commonly-used neuroblastoma cell lines

Jo Lynne Harenza[1], Maura A. Diamond[1], Rebecca N. Adams[2], Michael M. Song[3], Heather L. Davidson[3], Lori S. Hart[1], Maiah H. Dent[1], Paolo Fortina[2], C. Patrick Reynolds[3] & John M. Maris[1,4]

Neuroblastoma cell lines are an important and cost-effective model used to study oncogenic drivers of the disease. While many of these cell lines have been previously characterized with SNP, methylation, and/or mRNA expression microarrays, there has not been an effort to comprehensively sequence these cell lines. Here, we present raw whole transcriptome data generated by RNA sequencing of 39 commonly-used neuroblastoma cell lines. These data can be used to perform differential expression analysis based on a genetic aberration or phenotype in neuroblastoma (e.g., *MYCN* amplification status, *ALK* mutation status, chromosome arm 1p, 11q and/or 17q status, sensitivity to pharmacologic perturbation). Additionally, we designed this experiment to enable structural variant and/or long-noncoding RNA analysis across these cell lines. Finally, as more DNase/ATAC and histone/transcription factor ChIP sequencing is performed in these cell lines, our RNA-Seq data will be an important complement to inform transcriptional targets as well as regulatory (enhancer or repressor) elements in neuroblastoma.

Design Type(s)	cell type comparison design
Measurement Type(s)	transcription profiling assay
Technology Type(s)	RNA sequencing
Factor Type(s)	cell line
Sample Characteristic(s)	Homo sapiens ● neuroblastoma cell line ● epithelial cell line ● embryonic brain

[1]Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [2]Cancer Genomics and Bioinformatics Laboratory, Sidney Kimmel Cancer Center, Philadelphia, Pennsylvania 19107, USA. [3]Cancer Center, Texas Tech University Health Sciences Center School of Medicine, Lubbock, Texas 79430, USA. [4]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. Correspondence and requests for materials should be addressed to J.M.M. (email: maris@email.chop.edu).

Background & Summary

An estimated 15,780 children were diagnosed with cancer in 2014 In the United States, and per year globally, this number is nearly 250,000 (ref. 1). Although the 5-year survival rate of pediatric cancers is ~80%, many of the most commonly diagnosed childhood cancers: brain tumors, Wilms tumor, rhabdomyosarcoma, and high-risk neuroblastoma, have devastatingly low rates of survival[1], demonstrating the continued need for research progress in these areas. Here, we focus on neuroblastoma, the most common extracranial solid tumor in children. This disease has an estimated incidence of 1 in 8,000 to 10,000 births[2] and a 5-year survival rate of >95% for children in the low and intermediate risk groups. However, children with high-risk disease have only a 40% likelihood of survival[2]. Culturing of neuroblastoma cell lines dates back to the 1940s (ref. 3), during which the sole purpose of culturing was for diagnosis. However, producing cell lines from neuroblastoma tumors quickly became routine (see review[4]) and today, they are commonly-used, highly-characterized models used in laboratories across the world. Neuroblastoma cell lines nicely model a tumor's histopathology, gene expression, aneuploidy, and drug sensitivity, thus they are routinely used to investigate oncogenes or signaling pathways pharmacologically (drug screens, drug sensitivity/resistance) and/or genetically (siRNA, shRNA, CRISPR).

The genomics of neuroblastoma cell lines have been previously characterized using SNP[5], methylation[5,6], and/or mRNA expression microarrays[7–9], however, there has not been an effort to profile a large panel of these cell lines with high-throughput sequencing techniques. The motivation behind this study was to comprehensively profile the mRNA and non-coding RNA transcriptome of commonly-used neuroblastoma cell line models with a major goal of using this information as a complement to the epigenomic data currently available and the many data in the process of being generated. Integration of RNA expression patterns with histone and/or transcription factor chromatin immunoprecipitation (ChIP) sequencing is necessary for inferring transcriptional regulatory events. Neuroblastomas can be classified into various groups based on genetic lesions, for example: MYCN copy number amplification, harboring an activating *ALK* mutation, harboring a chromosomal loss (e.g.,: 1p, 3p, 11q) or gain (17q), *TERT* rearrangements (for review of neuroblastoma genomics, see ref. 10). Utilizing a panel of cell lines which harbor a mixture of these characteristics enables differential expression analyses on the basis of a genetic lesion, mutation of interest, or expression of a gene of interest.

These data have reuse value to inform selection of cell lines for experimental investigation of putative neuroblastoma oncogenes and/or tumor suppressors. For example, choice of knock-down or over-expression studies require *a priori* knowledge of basal expression of the gene of interest for rational experimental design. These data allow the experimenter to quickly determine which cell lines are high, mid, or low expressers of a gene of interest without requiring tedious quantitative, real-time PCR analysis or western blotting of multiple cell lines prior to initiating a gene over-expression or knockdown experiment.

Here, we describe transcriptome-wide profiling of 39 neuroblastoma cell lines, the hTERT-immortalized retinal pigmented epithelial cell line, RPE-1, and pooled human fetal brain tissue. Careful and stringent technical design at each experimental stage has allowed generation of a high-quality RNA-Seq dataset which has tremendous reuse value for the neuroblastoma community. An overview of the study design is depicted in Fig. 1. Briefly, cell lines were thawed, grown, and collected at 60–80% confluency over a two-month period. Once all cell lines were pelleted, RNA extractions were performed, quality of RNA inspected, and RNA sequencing was performed. Raw FASTQ files were generated and are publicly-available for reuse (see Data Citation 1). Additionally, we provide a processed file of gene-level mRNA abundances for each sample. We anticipate this data being a valuable tool for the neuroblastoma research community as we continue investigation into oncogenomic mechanisms of this disease.

Methods

Cell lines and culturing

Cell line stocks were obtained from the Children's Oncology Group (COG) Cell Culture and Xenograft Repository at Texas Tech University Health Sciences Center (www.COGcell.org), the American Type Culture Collection (Manassas, VA), or the Children's Hospital of Philadelphia (CHOP) cell line bank. Several of the COG-derived cell lines were established direct-to-culture in parallel with a patient-derived xenograft model[11] that are being characterized separately (see Table 1 (available online only)). All cell culturing for this experiment was performed at CHOP. Each cell line was thawed for 2–3 min in a 37 °C water bath, added to a 15 ml tube containing its respective growth medium, and pelleted by centrifugation at $300 \times g$ for 3 min. The supernatant was discarded to remove the DMSO-containing freezing medium. Cells were re-suspended in 1 ml of growth medium and transferred into a T75 flask containing an additional 10 ml of growth medium. Once cells were ~70–80% confluent, they were transferred to a 150 mm dish. At ~70–80% confluency, cells were split into two 150 mm dishes and at ~70–80% confluency, each dish of cells was pelleted, washed 3x with 1X PBS, and frozen at −80C until nucleic acids were extracted. See Table 1 (available online only) for a complete listing of cell lines, whether a matched patient-derived xenograft (PDX) exists, and their growth medium. Cell lines appended with 'nb' were grown in serum-free neurobasal medium. The following were purchased from
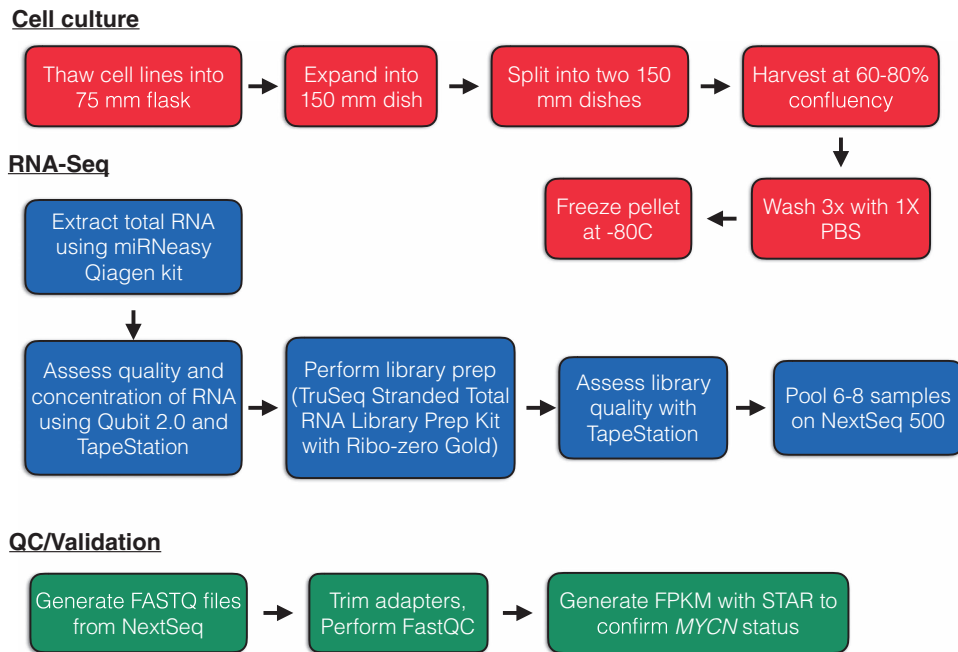
Cell culture

Thaw cell lines into 75 mm flask → Expand into 150 mm dish → Split into two 150 mm dishes → Harvest at 60-80% confluency

RNA-Seq

Extract total RNA using miRNeasy Qiagen kit

Freeze pellet at -80C ← Wash 3x with 1X PBS

Assess quality and concentration of RNA using Qubit 2.0 and TapeStation → Perform library prep (TruSeq Stranded Total RNA Library Prep Kit with Ribo-zero Gold) → Assess library quality with TapeStation → Pool 6-8 samples on NextSeq 500

QC/Validation

Generate FASTQ files from NextSeq → Trim adapters, Perform FastQC → Generate FPKM with STAR to confirm *MYCN* status

Figure 1. Experimental and data analysis workflow. Cell lines were thawed and cultured to ~60–80% confluence before passaging and finally, pelleting. RNA was extracted, sequencing performed, and data analysis performed as described.

Thermo Fisher Scientific (Waltham, MA): Iscove's IMDM (Cat# 12440053), RPMI 1640 with 25 mM HEPES (Cat# 22400089), Neurobasal-A Medium (Cat# 10888022), L-glutamine (Cat# 25030081), antibiotic/antimycotic (Cat# 15240062), 50X B-27 serum-free supplement (Cat# 18504044), 100X N-2 supplement (Cat# 17502048). The following growth factors were purchased from VWR (Radnor, PA): rhFGF (fibroblast growth factor, Cat# PAG5071) and rhEGF (epidermal growth factor, Cat# PAG5021). Insulin/Transferrin/Selenium (ITS) premix culture supplement was purchased from Corning Life Sciences (Tewksbury, MA, Cat# 354351). Hyclone Fetal bovine serum was purchased from Fisher Scientific (Cat# SH30071.03) and the lot remained consistent across the different medium formulations throughout the duration of the experiment. Of note, SK-N-BE(2)-C is a subclone derived from the parental SK-N-BE(2) cell line[12] and SH-SY5Y was derived from the SH-SY subclone of the parental SK-N-SH cell line[13].

Throughout the duration of the study, randomization was implemented to ensure unbiased data production. Cell lines were thawed in random order, nucleic acid extractions were performed randomly, and library preps and sequencing were performed randomly. Phenotypic characteristics of each cell line were also assessed as quality control during the cell growth stage. No unusual morphologies or growth rates were noted.

DNA extraction and STR profiling

From separate cell pellets, DNA was extracted using the DNeasy Blood & Tissue Kit (Cat# 69504, Qiagen, Valencia, CA). DNA was quantitated using the Nanodrop 1000 (Thermo Fisher Scientific) and Short Tandem Repeat (STR) profiling employed either the AmpFLSTR Identifiler PCR Amplification kit (Applied Biosystems, Foster City, CA) by the Children's Hospital of Philadelphia Nucleic Acids and Protein Core or the PowerPlex Fusion kit (Promega, Madison, WI) by Guardian Forensic Sciences (Abington, PA). All cell line STRs matched publicly-available references listed at http://strdb.cogcell.org/.

RNA extraction

Control human fetal brain total RNA (Cat# 636526, Lot#1605061A) was purchased from Clontech Laboratories (Mountain View, CA). This RNA was a pool of normal brain tissue from 21 spontaneously aborted male/female Caucasian fetuses of ages 26–40 weeks and was isolated using a modified guanidinium thiocyanate method[14]. For all cell lines, RNA was extracted using the miRNeasy Mini kit (Cat# 217004) from Qiagen (Valencia, CA) according to the manufacturer's protocol. RNA purity was assessed using the Nanodrop 2000 (Thermo Fisher Scientific) and quantitated with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Quality and RNA integrity numbers (RINs) were assessed using the TapeStation 2200 (Agilent Technologies, Santa Clara, CA). Each cell line RNA sample had a RIN $\geq$ 8.7 and the RIN for the fetal brain RNA was 7.6, thus all RNA was of high quality.

Library preparation and RNA sequencing

Libraries were prepared using 1 ug RNA according to the TruSeq Stranded Total RNA Sample Preparation guide (Part# 15031048 Rev. E, October 2013, Illumina, San Diego, CA). Ribosomal RNA removal was performed using the Gold rRNA Removal Mix per Illumina's recommendations. Quality of each library assessed with the Agilent TapeStation 2,200. Six to eight libraries were pooled ($N = 6–8$) and sequenced using v2 chemistry, 2×100 bp, on one high-output flow-cell of an Illumina NextSeq 500 to achieve at least 50 million paired reads per sample. Upon run completion, libraries were demultiplexed, Illumina adapters trimmed, and FASTQ files were generated using the Illumina NextSeq Control Software version 2.02.

Sequencing quality control

First, sample reads were concatenated for each paired read group. Next, FASTQC V0.11.4 (Babraham Institute, available for download at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was run on all samples and inspected for sequencing quality. Next, Picard tools version 1.140 (Broad Institute, Cambridge, MA, available for download at https://github.com/broadinstitute/picard/releases/tag/1.140) was used to calculate insert sizes for GEO according to the following parameters:

```
$ java -jar picard.jar CollectInsertSizeMetrics INPUT=
Aligned.sortedByCoord.out.bam OUTPUT=filename
```

Alignment and generation of counts

The Spliced Transcripts Alignment to Reference (STAR) version 2.4.2a aligner (available for download at https://github.com/alexdobin/STAR/releases/tag/STAR_2.4.2a)[15] was used to index the full hg19 genome fasta file from UCSC using the following parameters:

```
$ STAR --runMode genomeGenerate --runThreadN 16 --genomeDir idx_dir
--genomeFastaFiles ucsc.hg19.fa --sjdbGTFfile refSeq_hg19_2016-03-
03.gtf --sjdbOverhang 100
```

The GTF file was downloaded using the **genePredToGtf** command from the kent utility (available for download at http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/):

```
$ genePredToGtf hg19 knownGene knownGene.gtf
```

Next, sequences were aligned and counts per gene were generated using the following parameters in two-pass mode:

```
$ STAR --runMode alignReads --runThreadN 16 --twopassMode Basic --
twopass1readsN -1 --chimSegmentMin 15 --chimOutType WithinBAM --geno-
meDir dir --genomeFastaFiles ucsc.hg19.fa --readFilesIn R1.fastq.gz
R2.fastq.gz --readFilesCommand zcat --outSAMtype BAM SortedByCoordi-
nate --outFileNamePrefix $cellline. --quantMode TranscriptomeSAM Gen-
eCounts --sjdbGTFfile refSeq_hg19_2016-03-03.gtf --sjdbOverhang 100
```

Alignment resulted in an average of 66 million uniquely-mapped reads per sample. STAR two-pass mode alignment was chosen as it has been shown to have 99% alignment accuracy and has nearly 20x faster processing speed compared with TopHat2 and similar processing speed as HISAT two-pass mode[16].

Generation of FPKM

A custom R script was used to generate gene fragments per kilobase of exons per million reads (FPKM) from the count data produced from STAR. The Genomic Features Package version 1.22.13 (available for download at https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html) was used with R Version 3.2.2 (Fire Safety) to make the transcriptome database and figures were produced using ggplot2 version 2.1.0 (http://ggplot2.org/).

Differential expression analyses

Differential expression of genes based on *MYCN* amplification status was performed separately for cell lines and primary neuroblastoma tumor samples using the R package, DESeq2 (version 1.10.1)[17]. FASTQ files and *MYCN* status for patient tumors were obtained with consent through the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Consortium (see Data Citation 2, https://ocg.cancer.gov/programs/target/data-matrix). Next, the differentially-expressed genes' $\log_2$-transformed mean expression and the $\log_2$ fold-change were correlated between the cell lines and patient samples.

Code availability

R scripts for generation of FPKM and differential expression analyses are available for download at: https://github.com/marislab/NBL-cell-line-RNA-seq.

Data Records

All raw RNA-sequencing data (paired FASTQ files) as well as the processed FPKM matrix from this study have been deposited into the Gene Expression Omnibus (GEO) under Accession Number GSE89413 (see Data Citation 1). For associated specimen metadata, see Table 1 (available online only) and for
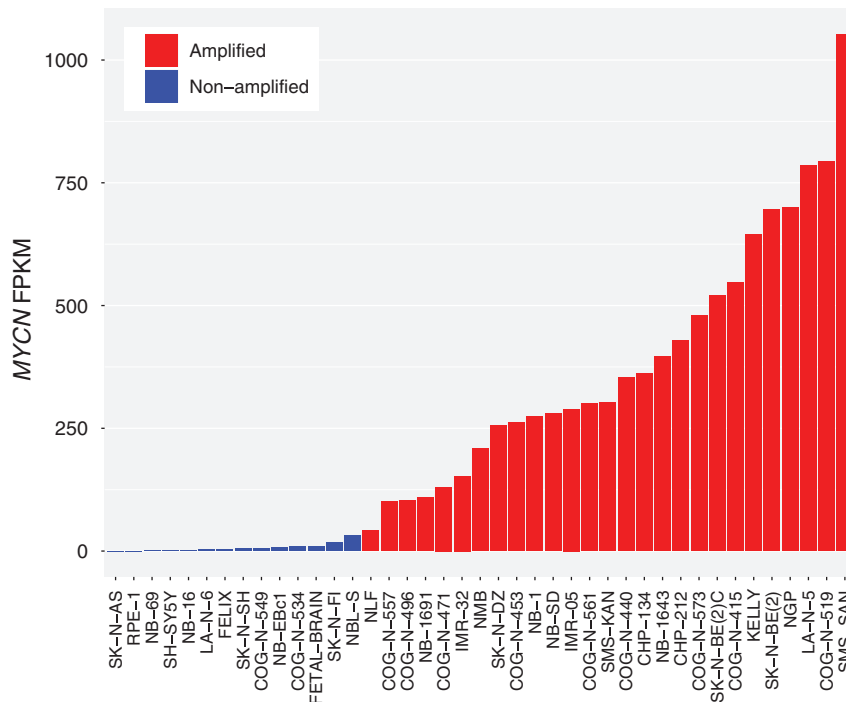
Figure 2. **Validation of *MYCN* genomic amplification status in neuroblastoma cell lines.** Plotted are rank-ordered *MYCN* FPKM values for the human fetal brain sample and each cell line, colored by known *MYCN* copy number status. These data validate known *MYCN* amplification status for each cell line.

associated assay metadata, see Table 2 (available online only). Raw single nucleotide polymorphism (SNP) array IDAT files and processed Genome Studio files for 27 of the cell lines have been deposited into GEO under Accession Number GSE89968 (see Data Citation 3). Together, these data make up the GEO Super Series GSE89969.

Technical Validation

As a technical validation of our RNA-Seq data, we generated FPKM for all genes (See Methods and Data Citation 1) and compared *MYCN* FPKM with each cell line's known copy number amplification status across cell lines (Fig. 2 and Table 3 (available online only)) . Of note, the tumor from which the NLF cell line was derived was *MYCN* copy number amplified by the fluorescence *in situ* hybridization, however, it is not amplified at the protein level[18] and therefore, as expected, has the lowest *MYCN* FPKM of all cell lines designated as *MYCN* amplified. All cell lines were concordant with known *MYCN* amplification status.

Next, for both cell lines and neuroblastoma patient data, we performed differential expression analyses based on *MYCN* genomic amplification status using the R package, DESeq2 (ref. 17). We correlated the DESeq2 base mean of the common differentially-expressed genes ($N = 2,395$) between cell lines and primary patient tumors, which were significantly correlated (Fig. 3a, Pearson's $R = 0.824$, $t = 71.131$, $df = 2,393$, 95% $CI = 0.811–0.836$, $P < 2.2$ e-16). The fold changes of these genes were also significantly correlated between the cell lines and patient samples (Fig. 3b, Pearson's $R = 0.73$, $t = 52.231$, $df = 2,393$, 95% $CI = 0.711–0.748$, $P < 2.2$ e-16), not only supporting the technical validity of our dataset, but also emphasizing the utility of these cell lines as a surrogate model for neuroblastoma.

Finally, we correlated non-differentially expressed genes (DESeq2 p-adjusted > 0.20) between the cell lines and patient tumors ($N = 6,523$). As expected, base mean expression of the genes correlated significantly (Fig. 3c, Pearson's $R = 0.829$, $t = 119.74$, $df = 6,521$, 95% $CI = 0.821–0.837$, $P < 2.2$ e-16). While correlating fold-change yields a significant *P*-value because of the large number of genes analyzed, it is clear that the relationship is weak, as the correlation is close to zero (Fig. 3d, Pearson's $R = 0.052$, $t = 4.1,766$, $df = 6,521$, 95% $CI = 0.027–0.076$, $P < 3$ e-5). This is expected, as all fold-changes of non-DE genes are close to zero.

Usage Notes

All raw FASTQ files and the associated FPKM matrix file can be downloaded from the Gene Expression Omnibus (GEO) under Accession Number GSE89413. STAR-Fusion (https://github.com/STAR-Fusion/STAR-Fusion) enables detection of fusion transcripts. Alternative gene expression analyses can be
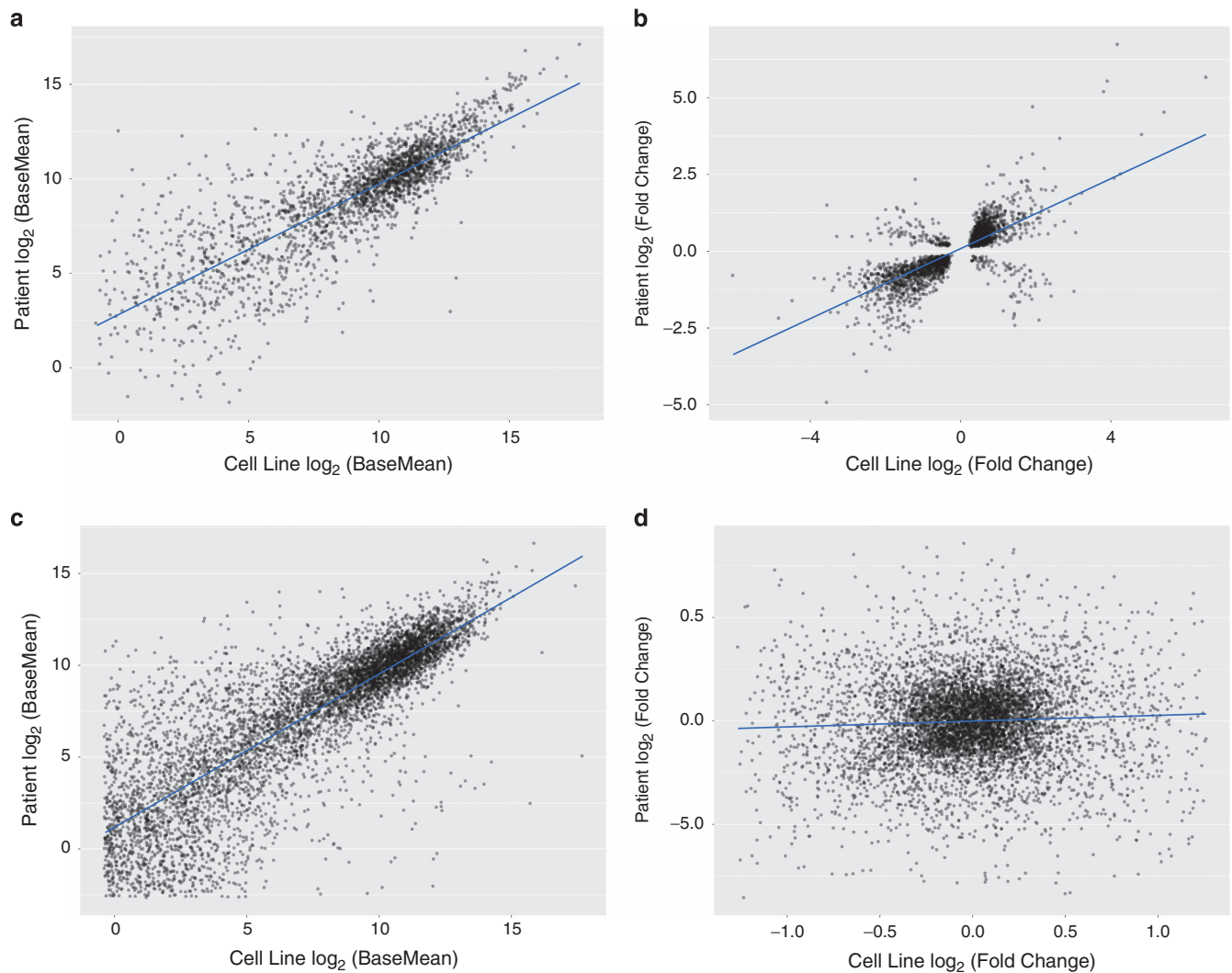
Figure 3. Concordance of differentially-expressed genes between neuroblastoma cell lines and primary tumors. (**a**) Across the neuroblastoma cell lines, 3,940 genes were differentially-expressed (DE) based on *MYCN* amplification status and of those, 2,395 were differentially-expressed based on *MYCN* amplification status in primary tumors and were significantly correlated (Pearson's $R = 0.824$, $P < 2.2$ e-16). (**b**) The fold changes of these DE genes were significantly correlated between the cell line dataset and the patient tumor dataset (Pearson's $R = 0.73$, $P < 2.2$ e-16). (**c**) A significant correlation between the common 6,523 genes that were not DE in cell lines and tumors was observed (Pearson's $R = 0.829$, $P < 2.2$ e-16). (**d**) As expected, correlation of the non-DE genes' fold changes was close to zero (Pearson's $R = 0.052$, $P < 3$ e-5).

performed using RSEM[19] and/or transcript level analyses can be performed using kallisto[20]. Use of kallisto will also allow quantification of non-coding RNA abundances. Differential expression analyses may be performed using the common R packages, limma[21] or DESeq2 (ref. 17). Differentially expressed gene lists can be explored for enrichment in signaling pathways using Ingenuity Pathway Analysis (Qiagen, http://www.ingenuity.com/products/ipa) and/or gene ontologies using ToppGene[22] or the Gene Ontology Consortium tool[23]. Finally, these expression data can be integrated with epigenomics datasets (e.g.: ChIP-Seq, DNase-Seq/ATAC-Seq, Histone ChIP-Seq) to infer transcriptional regulation or repression.

References

1. American Childhood Cancer Organization. *Special Section: Cancer in Children & Adolescents. ACS Special Report* 25–42 (2014).
2. Maris, J. M., Hogarty, M. D., Bagatell, R. & Cohn, S. L. Neuroblastoma. *Lancet* **369,** 2106–2120 (2007).
3. Murray, M. R. & Stout, A. P. Distinctive Characteristics of the Sympathicoblastoma Cultivated in Vitro: A Method for Prompt Diagnosis. *Am. J. Pathol.* **23,** 429–441 (1947).
4. Thiele, C. J. in *Human Cell Culture* 1, 21–53 (Human Cell Culture Lancaster, 1999).

5. Henrich, K.-O. *et al.* Integrative Genome-Scale Analysis Identifies Epigenetic Mechanisms of Transcriptional Deregulation in Unfavorable Neuroblastomas. *Cancer Res.* **76,** 5523–5537 (2016).
6. Decock, A., Ongenaert, M., Van Criekinge, W., Speleman, F. & Vandesompele, J. DNA methylation profiling of primary neuroblastoma tumors using methyl-CpG-binding domain sequencing. *Sci. Data* **3,** 160004–160011 (2016).
7. Cole, K. A., Huggins, J. & LaQuaglia, M. RNAi screen of the protein kinome identifies checkpoint kinase 1 (CHK1) as a therapeutic target in neuroblastoma. *Proc. Natl. Acad. Sci* **108,** 3336–3341 (2011).
8. Whiteford, C. C. *et al.* Credentialing preclinical pediatric xenograft models using gene expression and tissue microarray analysis. *Cancer Res.* **67,** 32–40 (2007).
9. Keshelava, N. *et al.* Histone deacetylase 1 gene expression and sensitization of multidrug-resistant neuroblastoma cell lines to cytotoxic agents by depsipeptide. *J. Natl. Cancer Inst.* **99,** 1107–1119 (2007).
10. Bosse, K. R. & Maris, J. M. Advances in the translational genomics of neuroblastoma: From improving risk stratification and revealing novel biology to identifying actionable genomic alterations. *Cancer* **122,** 20–33 (2016).
11. Kang, M. H. *et al.* National Cancer Institute pediatric preclinical testing program: model description for in vitro cytotoxicity testing. *Pediatr. Blood Cancer.* **56,** 239–249 (2011).
12. Ciccarone, V., Spengler, B. A., Meyers, M. B. & Biedler, J. L. Phenotypic diversification in human neuroblastoma cells: expression of distinct neural crest lineages. *Cancer Res.* **49,** 219–225 (1989).
13. 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).
14. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156–159 (1987).
15. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
16. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).
17. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
18. Hart, L. S. *et al.* Preclinical Therapeutic Synergy of MEK1/2 and CDK4/6 Inhibition in Neuroblastoma. *Clin. Cancer Res.* doi:10.1158/1078-0432.CCR-16-1131 (2016).
19. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
20. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnol* **34,** 525–527 (2016).
21. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
22. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37,** W305–W311 (2009).
23. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
24. Carr, J. *et al.* High-resolution analysis of allelic imbalance in neuroblastoma cell lines by single nucleotide polymorphism arrays. *Cancer Genet. Cytogenet.* **172,** 127–138 (2007).
25. Nair, P. N., McArdle, L., Cornell, J., Cohn, S. L. & Stallings, R. L. High-resolution analysis of 3p deletion in neuroblastoma and differential methylation of the SEMA3B tumor suppressor gene. *Cancer Genet. Cytogenet.* **174,** 100–110 (2007).
26. Wang, K. *et al.* Integrative genomics identifies LMO1 as a neuroblastoma oncogene. *Nature* **469,** 216–220 (2011).
27. Vandesompele, J. *et al.* Identification of 2 putative critical segments of 17q gain in neuroblastoma through integrative genomics. *Int. J. Cancer* **122,** 1177–1182 (2008).

Data Citations

1. Harenza, J., Diamond, M. A., Hart, L. S. & Maris, J. M. *Gene Expression Omnibus* GSE89413 (2016).
2. *NBCI Bioproject* PRJNA89523 (2009).
3. Harenza, J., Diamond, M. A. & Maris, J. M. *Gene Expression Omnibus* GSE89968 (2016).

Acknowledgements

Author Contributions

J.H. and J.M.M. conceived the experiment. J.H. designed the experiment, collected samples, performed QC and data analysis, and wrote the manuscript. J.H. and L.S.H. cultured the cell lines. M.A.D. extracted RNA samples and prepared them for sequencing. RNA performed library preps and sequenced the samples under the guidance of P.F. From separate cell pellets, J.H., M.A.D., and M.H.D. extracted DNA for STR typing. M.J.S., H.L.D., and C.P.R. established and validated COG cell lines. All authors read and approved of the final manuscript.

Additional Information

Tables 1, 2 and 3 are only available in the online version of this paper.

Competing interests: The authors declare no competing financial interests.

How to cite this article: Harenza, J. L. *et al.* Transcriptomic profiling of 39 commonly-used neuroblastoma cell lines. *Sci. Data* 4:170033 doi: 10.1038/sdata.2017.33 (2017).

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