Data in Brief

Multi-omic profiling of MYCN-amplified neuroblastoma cell-lines

Erik Dassi [a,*], Valentina Greco [a], Viktoryia Sidarovich [a], Paola Zuccotti [a], Natalia Arseni [a], Paola Scaruffi [b], Gian Paolo Tonini [c], Alessandro Quattrone [a]

[a] Laboratory of Translational Genomics, Centre for Integrative Biology, University of Trento, Italy
[b] Center of Physiopathology of Human Reproduction, Unit of Obstetrics and Gynecology, IRCCS A.O.U. San Martino IST, Genoa, Italy
[c] Neuroblastoma Laboratory, Pediatric Research Institute, Fondazione Città della Speranza, Padua, Italy

ARTICLE INFO

ABSTRACT

Neuroblastoma is the most common pediatric cancer, arising from the neural crest cells of the sympathetic nervous system. Its most aggressive subtype, characterized by the amplification of the MYCN oncogene, has a dismal prognosis and no effective treatment is available. Understanding the alterations induced by the tumor on the various layers of gene expression is therefore important for a complete characterization of this neuroblastoma subtype and for the discovery of new therapeutic opportunities. Here we describe the profiling of 13 MYCN-amplified neuroblastoma cell lines at the genome (copy number), transcriptome, translatome and miRome levels (GEO series GSE56654, GSE56552 and GSE56655). We provide detailed experimental and data analysis procedures by means of which we derived the results described in [1].

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications

Organism/cell line/tissue	Homo sapiens neuroblastoma cell lines: CHP-126, CHP-134, CHP-212, IMR-32, KELLY, LAN-1, MHH-NB-11, SIMA, SK-N-BE2 and SK-N-DZ, STA-NB-1, STA-NB-7 and STA-NB-10
Sex	Female, male, male, male, female, male, male, male, male, female, female, male, male
Sequencer or array type	Agilent Whole Human Genome Microarray 4x44K G4112F for GSE56552, Agilent Human Genome CGH Microarray 244A for GSE56654 and Agilent Human miRNA Microarray 2.0 G4470B for GSE56655
Data format	Raw and analyzed
Experimental factors	NA
Experimental features	All samples are derived from MYCN-amplified neuroblastoma cell lines
Consent	NA
Sample source location	NA

1. Direct link to deposited data

Transcriptome and translatome profiling: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56552

Genome (copy number) profiling: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56654

miRome profiling: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56655

* Corresponding author.

2. Experimental design, materials and methods

2.1. Experimental design

We performed one copy number (aCGH), transcriptome (total RNA), translatome (RNA associated to polysomes) and miRome (microRNA expression) profiling for each cell line, thus obtaining 13 samples for each profiled level (except for the miRome, consisting of 11 samples) [1]. The experimental design is also depicted in Fig. 1.

2.2. Cell culture

Cell lines were grown at 37 °C in a 5%-CO_2 humidified atmosphere. CHP-134, IMR-32, KELLY, LAN-1, SK-N-BE2 and -DZ were purchased from the ECACC (Salisbury, UK). CHP-126, MHH-NB-11 and SIMA were purchased from the DSMZ (Braunschweig, Germany). CHP-212 was purchased from the ATCC, while STA-NB-1, -7 and -10 were kindly provided by Dr. Peter F. Ambros (CCRI, Vienna, Austria). Samples were collected from cell lines at early passages (n = 3 to 6) to avoid the insurgence of culture-induced alterations; all lines were also checked for Mycoplasma and other potential infections. All the cell lines were searched in the Database of Cross-contaminated or Misidentified Cell Lines (http://iclac.org/), and none were flagged as cross-contaminated or misidentified.

2.3. aCGH microarrays

Total DNA was isolated by means of the DNA Blood and Tissue Extraction Kit (Qiagen), following the manufacturer's protocol. Array-
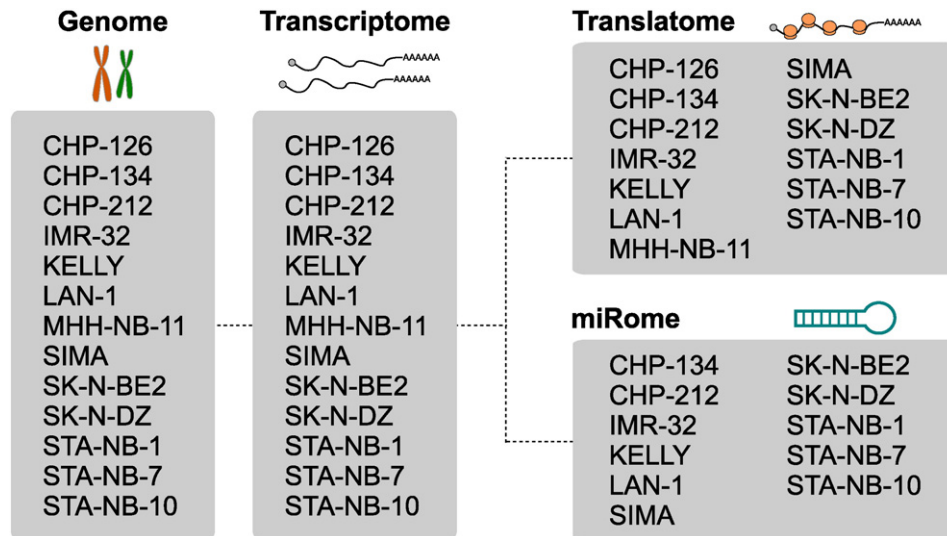
Fig. 1. Experimental design. List the cell lines employed in the omic profiling at the various levels (genome, transcriptome, translatome and miRome) included in this study.

CGH was performed using Human Genome CGH 1x244K microarrays slides (G4411B, Agilent Technologies). Briefly, 3 μg of genomic DNA was labeled with Cy3-dUTP or Cy5-dUTP (Perkin Elmer) using the BioPrime DNA Labeling kit (Life Technologies); labeled reference and tumor DNAs were then hybridized in a SureHyb gasket (Agilent Technologies) at 65 °C for 40 h. Eventually, the slides were washed and scanned immediately after by means of a G2565BA scanner (Agilent Technologies) using the two color scan setting for 244 k slides; TIFF images were processed with the FeatureExtractor software (Agilent Technologies).

2.4. Transcriptome microarrays

Total RNA was first extracted from cells at 80% confluence by means of the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol; it was then quantified, and its quality was assessed by using the RNA 6000 Nano assay on the 2100 Bioanalyzer (Agilent Technologies); an RNA integrity number threshold of 7 was used to select samples to be included in the study; 500 ng of starting material was eventually employed for each sample and Cyanine-3 labeled cRNA was produced. The samples were hybridized on Human GE 4x44K v2 microarrays (G4112F, Agilent Technologies) in a SureHyb gasket (Agilent Technologies) at 65 °C for 17 h: 1.5 μg of Cy3-labeled cRNA was used for this step. The hybridized slides were then washed and scanned immediately after by using a G2565BA scanner (Agilent Technologies); TIFF images were processed with the FeatureExtractor software (Agilent Technologies).

2.5. Translatome microarrays

The translatome was studied with the polysomal profiling technique, which allows retrieving the mRNAs which are associated to the polysomes, and thus actively translated. Cells at 80% confluence were incubated for 3 min with 10 μg/ml cycloheximide at 37 °C. The plates were then placed on ice, the medium was removed, and the cells were washed twice with PBS + 10 μg/ml cycloheximide. The cells were lysed directly on the plates with 300 μl of lysis buffer (10 mM MgCl$_2$, 10 mM NaCl, 10 mM Tris–HCl (pH 7.5), 0.2 U/ml RNase inhibitor (Fermentas), 1 mM DTT, 1% Triton X-100, 1% sodium deoxycholate and 10 μg/ml cycloheximide), scraped and transferred to a tube. The resulting extracts were then centrifuged for 5 min at 12,000 g/4 °C to remove nuclei and cellular debris. The resulting lysate was directly layered on a 15–50% linear sucrose gradient containing 30 mM Tris–HCl (pH 7.5), 100 mM NaCl and 10 mM MgCl$_2$ and eventually centrifuged on a SW41 rotor (Beckman Coulter) for 100 min at 180,000 g. Fractions

of 1 ml volume were then collected by monitoring the absorbance at 254 nm and were treated with 0.1 mg/ml proteinase K for 2 h at 37 °C. Only polysome-containing fractions (thus corresponding to actively translated mRNAs) were collected: all the fractions after the ribosomal 80S peak were considered to be polysomal. RNA was eventually extracted with phenol–chloroform, then precipitated with isopropanol and resuspended in 30 μl RNase-free water; the resulting RNA was assessed for RNA integrity and profiled on Human GE 4x44K v2 microarrays (G4112F, Agilent Technologies) as described for transcriptome microarrays.

2.6. miRome microarrays

MicroRNAs from cells at 80% confluence were isolated by means of the miRNeasy Micro Kit (Qiagen), following the manufacturer's protocol. Resulting material was then quantified and quality-assessed using the Small RNA Assay on a 2100 Bioanalyzer (Agilent Technologies); an RNA integrity number threshold of 7 was used to select samples to be included in the study. The microarray profiling was performed using 100 ng of starting material for each sample. The RNA was Cyanine-3-labeled and samples were then hybridized on Human miRNA Microarrays 2.0 (G4470B, Agilent Technologies) in a SureHyb gasket (Agilent Technologies) at 55 °C for 20 h. Slides were then washed and immediately scanned using a G2565BA scanner (Agilent Technologies); TIFF images were eventually processed with the FeatureExtractor software (Agilent Technologies).

2.7. aCGH data analysis

Arrays were first loaded in R by means of limma [2]. The raw signals were centered around the median and the MANOR [3] package was used to correct the data for global intensity trends, local spatial biases, to check the signal-to-noise ratio and the replicate probes consistency. Probes retained by these quality checks were then processed with the DNAcopy package (bioconductor.org/packages/release/bioc/html/DNAcopy.html): an outlier smoothing correction was applied, followed by segmentation with the CBS algorithm and a post-segmentation normalization to fit data to its most likely zero-value. Default parameters were used. Genomic segments were assigned a loss/gain status and recurrent aberrations were obtained through KCsmart [4], employing a p-value threshold of 0.05 and 1000 permutations to find the significance threshold. Alterations can then be were plotted with Circos [5]. The R [6] source code employed to analyze aCGH microarrays is provided in the Supplementary File 1.

2.8. Transcriptome and translatome data analysis

Arrays were first loaded in R [6] by means of the Agi4x44PreProcess package (bioconductor.org/packages/2.13/bioc/html/Agi4x44PreProcess.html). Probes were filtered by the following criteria: retained probes must have at least 25% of the samples with intensity well above background; retained probes must have at least 75% of the samples with sufficient spot diameter and a signal-to-noise ratio well above the negative controls; retained probes must not be saturated or be an outlier. The remaining probes were then median-summarized and their signal was log2-converted; quantile normalization was eventually applied by means of the same package. Non-expressed genes, defined as having expression levels below the 25th percentile, were filtered out. Additional analyses were performed as follows: hierarchical clustering with the hclust function (with Ward linkage rule and Pearson distance metric); PCA analysis with the prcomp function and k-means clustering with the kmeans function (n = 3). Differential representation of genes between the transcriptome and the transaltome was computed with RankProd [7], SAM (cran.r-project.org/web/packages/samr) and the T-test [8]. The corrected p-value threshold was 0.05 for T-test (Benjamini–Hochberg correction), RankProd (percentage of false positives) and SAM (false discovery rate). The source code employed to analyze transcriptome and translatome microarrays is provided in the Supplementary File 2.

2.9. miRome data analysis

Arrays were first loaded into R by means of the AgiMicroRna package [9]. Replicated probes were summarized and all probes were filtered according to the following criteria: probes must be expressed in at least 75% of the samples and must have a signal which is well above negative controls in at least 25% of the samples. Signals were then log2-converted, and quantile normalization was applied by means of the same package. Ingenuity Pathway Analysis (www.ingenuity.com) was then be used on the expressed miRNAs list to perform functional enrichments and pathway analysis. The source code employed to analyze miRome microarrays is provided in the Supplementary File 3.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.11.012.

Funding

References

[1] E. Dassi, V. Greco, V. Sidarovich, P. Zuccotti, N. Arseni, P. Scaruffi, G.P. Tonini, A. Quattrone, Translational compensation of genomic instability in neuroblastoma. Sci. Rep. 5 (2015) 14364.
[2] G.K. Smyth, Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 3 (2004) (Article3).
[3] P. Neuvial, P. Hupe, I. Brito, S. Liva, E. Manie, C. Brennetot, F. Radvanyi, A. Aurias, E. Barillot, Spatial normalization of array-CGH data. BMC Bioinformatics 7 (2006) 264.
[4] J.J. de Ronde, C. Klijn, A. Velds, H. Holstege, M.J. Reinders, J. Jonkers, L.F. Wessels, KC-SMARTR: an R package for detection of statistically significant aberrations in multi-experiment aCGH data. BMC Res. Notes 3 (2010) 298.
[5] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics. Genome Res. 19 (2009) 1639–1645.
[6] R Core Team, R: a language and environment for statistical computing. 2014.
[7] F. Hong, R. Breitling, C.W. McEntee, B.S. Wittner, J.L. Nemhauser, J. Chory, RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 22 (2006) 2825–2827.
[8] L. Tian, S.A. Greenberg, S.W. Kong, J. Altschuler, I.S. Kohane, P.J. Park, Discovering statistically significant pathways in expression profiling studies. Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 13544–13549.
[9] P. Lopez-Romero, Pre-processing and differential expression analysis of agilent microRNA arrays using the AgiMicroRna bioconductor library. BMC Genomics 12 (2011) 64.