PLOS ONE

RESEARCH ARTICLE

Gene signature associated with benign neurofibroma transformation to malignant peripheral nerve sheath tumors

Marta Martínez[1]*, Carlos O. S. Sorzano[1,2], Alberto Pascual-Montano[1¤], Jose M. Carazo[1]

1 Biocomputing Unit, Nacional Center for Biotechnology (CSIC), Campus Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain, **2** Bioengineering Lab., Universidad CEU San Pablo, Campus Urb. Montepríncipe, Boadilla del Monte, Madrid, Spain

¤ Current address: PerkinElmer España, S. L., Tres Cantos, Madrid, Spain
* mmmtnez@cnb.csic.es

Citation: Martínez M, Sorzano COS, Pascual-Montano A, Carazo JM (2017) Gene signature associated with benign neurofibroma transformation to malignant peripheral nerve sheath tumors. PLoS ONE 12(5): e0178316. https://doi.org/10.1371/journal.pone.0178316

Editor: Marta M. Alonso, Universidad de Navarra, SPAIN

Received: February 21, 2017

Accepted: May 11, 2017

Published: May 24, 2017

Data Availability Statement: Starting datasets analyzed for the current study are available in the repositories ArrayExpress (E-MEXP-353 and E-TABM-69), https://www.ebi.ac.uk/arrayexpress/, and GEO (GSE14038, GSE21714, GSE39764, GSE41747, and GSE66743), https://www.ncbi.nlm.nih.gov/geo/. All relevant data derived from this work are within the paper and its Supporting Information files.

Funding: This work was supported by Grant 2013-04-005 from the Children's Tumor Foundation (US)

Abstract

Benign neurofibromas, the main phenotypic manifestations of the rare neurological disorder neurofibromatosis type 1, degenerate to malignant tumors associated to poor prognosis in about 10% of patients. Despite efforts in the field of (epi)genomics, the lack of prognostic biomarkers with which to predict disease evolution frustrates the adoption of appropriate early therapeutic measures. To identify potential biomarkers of malignant neurofibroma transformation, we integrated four human experimental studies and one for mouse, using a gene score-based meta-analysis method, from which we obtained a score-ranked signature of 579 genes. Genes with the highest absolute scores were classified as promising disease biomarkers. By grouping genes with similar neurofibromatosis-related profiles, we derived panels of potential biomarkers. The addition of promoter methylation data to gene profiles indicated a panel of genes probably silenced by hypermethylation. To identify possible therapeutic treatments, we used the gene signature to query drug expression databases. Trichostatin A and other histone deacetylase inhibitors, as well as cantharidin and tamoxifen, were retrieved as putative therapeutic means to reverse the aberrant regulation that drives to malignant cell proliferation and metastasis. This *in silico* prediction corroborated reported experimental results that suggested the inclusion of these compounds in clinical trials. This experimental validation supported the suitability of the meta-analysis method used to integrate several sources of public genomic information, and the reliability of the gene signature associated to the malignant evolution of neurofibromas to generate working hypotheses for prognostic and drug-responsive biomarkers or therapeutic measures, thus showing the potential of this *in silico* approach for biomarker discovery.

Introduction

Neurofibromatosis type 1 disease (NF1; Online Mendelian Inheritance in Man/OMIM—database #162200) is a rare chronic neurological disorder caused by a deficient autosomal dominant genetic background, which affects 1 in 3000 live births [1]. Alterations in the tumor

(http://www.ctf.org/) to AP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

suppressor gene neurofibromin *(NF1)* enhance expression of the Ras signaling pathway, which is involved in the evolution of many cancers [2]. Patients develop various anomalies in skin, eyes and skeleton, as well as in the cardiovascular, endocrine and nervous systems. In the peripheral nervous system, disorders typically manifest as benign neurofibromas (NF). Dermal neurofibromas (dNF) arise from small cutaneous nerves, whereas plexiform neurofibromas (pNF) have a deeper location within larger nerves; dNF and pNF gene expression patterns are indistinguishable [3]. In ~10% of NF1 patients, pNF can degenerate to malignant peripheral nerve sheath tumors (MPNST). About 50% of MPNST cases associate to NF1 disease, whereas the other 50% appear sporadically. Whether there are significant biological differences between sporadic and NF1-associated MPNST cases is debated [4–7].

The likelihood of MPNST development in NF1 patients depends on several risk factors [8], and appropriate prediction of pNF evolution would help to stratify patients and to choose the best early treatment. Despite recent advances based on studies of concomitant alterations in genes other than *NF1*, gene copy number alteration, epigenetic changes and gene expression, no prognostic biomarkers are available. Unlike other types of sarcoma, MPNST show a wide spectrum of chromosomal alterations [9]. In both sporadic and NF1-associated MPNST, amplifications are more frequent than deletions and affect the distal part of chromosome arm 17q [10]. Deletions mainly involve band p21 of chromosome 9, thus driving a dose reduction of the kinase inhibitor CDKN2A [11], and the proximal part of 17q, where *NF1* appears to be co-deleted with *SUZ12*, which encodes a member of the epigenetic regulator polycomb repressor complex 2 (PRC2). Genetic modifications due to PRC2 silencing are suggested as biomarkers for NF1 patient stratification [12].

Hypermethylation of tumor suppressor genes and hypomethylation of oncogenes are epigenetic changes reported for most cancers. A whole-methylome analysis identified 3690 genes probably associated with MPNST development and progression; among these were genes that encode CDKN2A and the tumor suppressors SOX10 and RASSF1, a Ras association domain family member [13]. *RASSF1A* silencing by promoter methylation is a biomarker of NF1-associated MPNST patients with poor prognosis [14].

Genome-wide RNA expression studies encompass patient samples [3,6,10,15,16], human cell cultures including NF- and MPNST-derived cell lines [1,3], and mouse models [17] that replicate human NF histology [18]. Gene expression profiles serve to identify disease biomarkers, such as *BIRC5*, *TOP2A* and *TK1*, that categorize MPNST patients with poor prognosis after surgery [10], or might also help to identify new therapeutic agents through drug repositioning, *i.e.*, use of tested drugs to treat new disease indications [19]. NFFinder and other bioinformatics tools compare gene signatures to seek potential repurposing medicines in the context of orphan diseases [20]. Although a unique average gene signature associated to changes from NF to MPNST would be desirable as input for NFFinder, differences in sample nature, array platforms, and hybridization protocols hamper direct comparison of results among studies, which explains the current lack of attention to genomic data integration in neurofibromatosis.

The combination of data from public databases such as the Gene Expression Omnibus (GEO) and ArrayExpress [21,22] and development of high-throughput technologies has led generalization of data integration by meta-analysis in genomic research. By using statistical tools to combine independent studies, microarray meta-analysis approaches extract consistent average gene expression signatures as well as interaction networks [23]. Robust, unsupervised meta-analysis approaches are currently being developed based on projections of high-dimensional data into a low-dimensional space to infer dependency among data [24]. These approaches nonetheless require a relatively large sample number [25] and connections among the different -omics data, which precludes their use in studies for which information is limited.

Combining pre-calculated values (P-values and effect sizes) led to a useful method to integrate heterogeneous data, qualitative and quantitative, for analysis of diabetes mellitus and Down syndrome [23,26]. This method, based on adding a score determined for each gene in each experiment using a formula similar to that applied in correlation studies, weights the size of the effect with reproducibility for all sample replicates and the statistical significance of the differentially expressed genes. As a starting point to identify biomarkers in the context of neurofibromatosis, we used a slight modification of this formula to integrate public genomic data and obtained a robust ranked score gene signature associated to the NF transition to MPNST.

Results

Experimental sets used to define the gene expression signature associated to NF-to-MPNST transition

To define a unique MPNST vs. NF gene signature that integrates data from diverse expression studies, we sought cited accessions in GEO and ArrayExpress databases that include MPNST and NF samples, and found five microarray studies (four human and one murine). These studies are referenced in Fig 1 and the heterogeneity of these data sets is detailed in Table 1 (bold). Despite the different number of genes represented by each microarray platform, all of them satisfied the selection criteria, segregated properly NF and MPNST samples, and supported relevant works that identified potential biomarkers and therapeutic targets [3,10,16,17] or discriminated neurofibromas and MPNST from other mesenchymal tumors [15]. Based on expression profile similarity [3], we grouped dNF and pNF samples from studies E-TABM-69, GSE41747 (human) and GSE66743, respectively. We also grouped segregated sporadic and NF1-associated MPNST samples from the GSE66743 study, as there were no differentially expressed genes between samples [10]. The final number of MPNST and NF samples compared

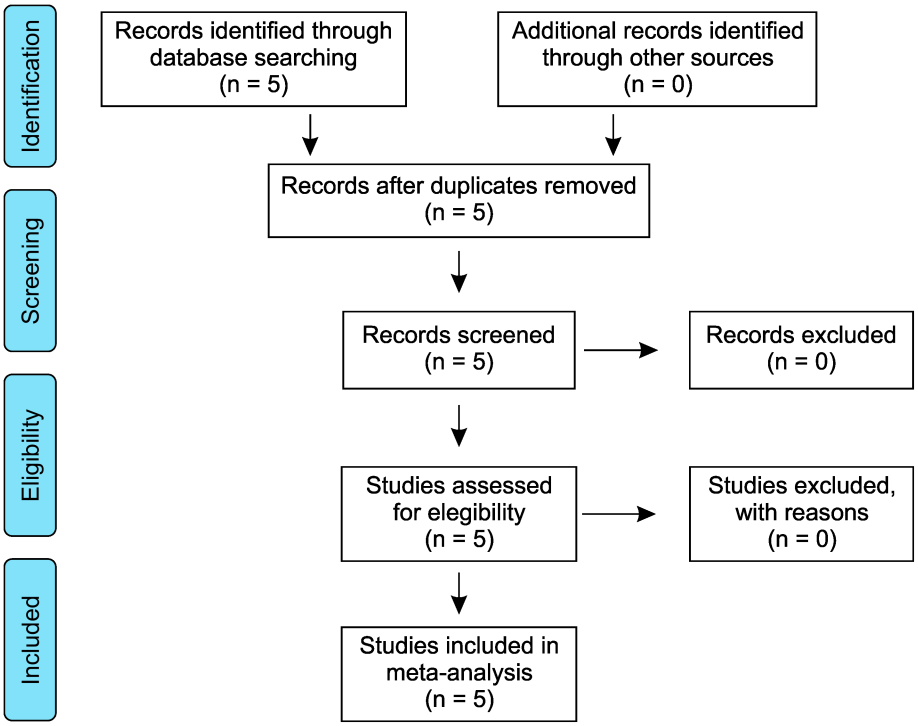


Fig 1. Prisma flow diagram.

https://doi.org/10.1371/journal.pone.0178316.g001

Table 1. Microarray studies selected from public databases and included in the MPNST vs. NF meta-analysis.

Tissue	Organism	NF vs. control			MPNST vs. control			MPNST vs. NF			Reference	Platform	Probe Number
		Accession	Samples		Accession	Samples		Accession	Samples				
			NF	control		MPNST	control		MPNST	NF			
Cell cultures	Human	GSE14038[1]	22	10	GSE14038[1]	13	10	GSE14038[1]	22	10	[3]	Affy U133 Plus 2.0	54675
					GSE39764[1]	3	3				[1]	Agilent-014850 4x44K	45015
Nerve tumors	Human	GSE41747[1]	26	3	GSE41747[1]	6	3	**E-MEXP-353[2]**	**4**	**14**	[15]	Affy U133A	22283
								E-TABM-69[2]	**4**	**16**	[16]	Agilent 011521 G4110A	19061
								GSE41747[1,3]	**6**	**26**	[17]	Affy U133 Plus 2.0	54675
								GSE66743[1]	**30**	**8**	[10]	ABI Version 2	33025
	Mouse	GSE41747[1]	15	15	GSE41747[1]	18	15	**GSE41747[1]**	**18**	**15**	[17]	Affy 430 2.0	45101

[1]Accession number from GEO database.

[2]Accession number from ArrayExpress database.

[3]Data included in GSE14038 accession.

https://doi.org/10.1371/journal.pone.0178316.t001

is shown in Table 1; see Materials and methods for details of sample selection and preprocessing, and probe mapping to ENSEMBL genes as common identifiers.

MPNST vs. NF gene expression signature

To integrate the five data sets, we calculated a score for each gene in each experiment (Fig 2, Microarray processing, and S1 Table), based on a formula used to combine data from heterogeneous sources [23]. To obtain a single score (s_i) for each gene, we added individual scores (s_{ij}) across human and mouse data sets (Fig 2, Meta-analysis). The final MPNST vs. NF gene signature contained 579 unique ENSEMBL human genes with non-null score and absolute median logFC value >0.99. The 336 up- and 243 downregulated genes included in this signature are highlighted in Table A in S2 Table (bold), embedded in the larger list from which the gene signature was filtered. The list includes the starting 7064 unique ENSEMBL human genes (4059 up- and 3005 downregulated) with non-null score in at least one independent study for which s_{ij} was computed. The comparison of non-null score profiles of this unfiltered integrative gene list and the lists derived from each individual study revealed comparable and overlapping patterns of gene scores (S1 Fig). Genes with the highest absolute scores in each study were at both ends of the unfiltered list, and were thus included in the final gene signature. However, some mice genes showed opposite behavior to that set in human, which corroborated differences in transcriptional responses between human and mouse models, particularly in neurodegenerative diseases [27]. Given the difficulty of interspecies comparison of results, we restricted the number of murine genes used to compute final scores (for score computation details, see Materials and methods).

The most promising biomarkers, 20 up- and 20 downregulated genes with the highest and the lowest score values from the MPNST vs. NF signature, were extracted from Table A in S2 Table (Table 2). In addition to gene score values, we determined attributes for estimating the relevance of each gene in the list, such as the median value of effect size, computed regarding the mean (logFC) or the median (logFC_m) across studies, as well as to assess the bias across the studies, like the number of studies in which each gene is represented, the inclusion or exclusion of

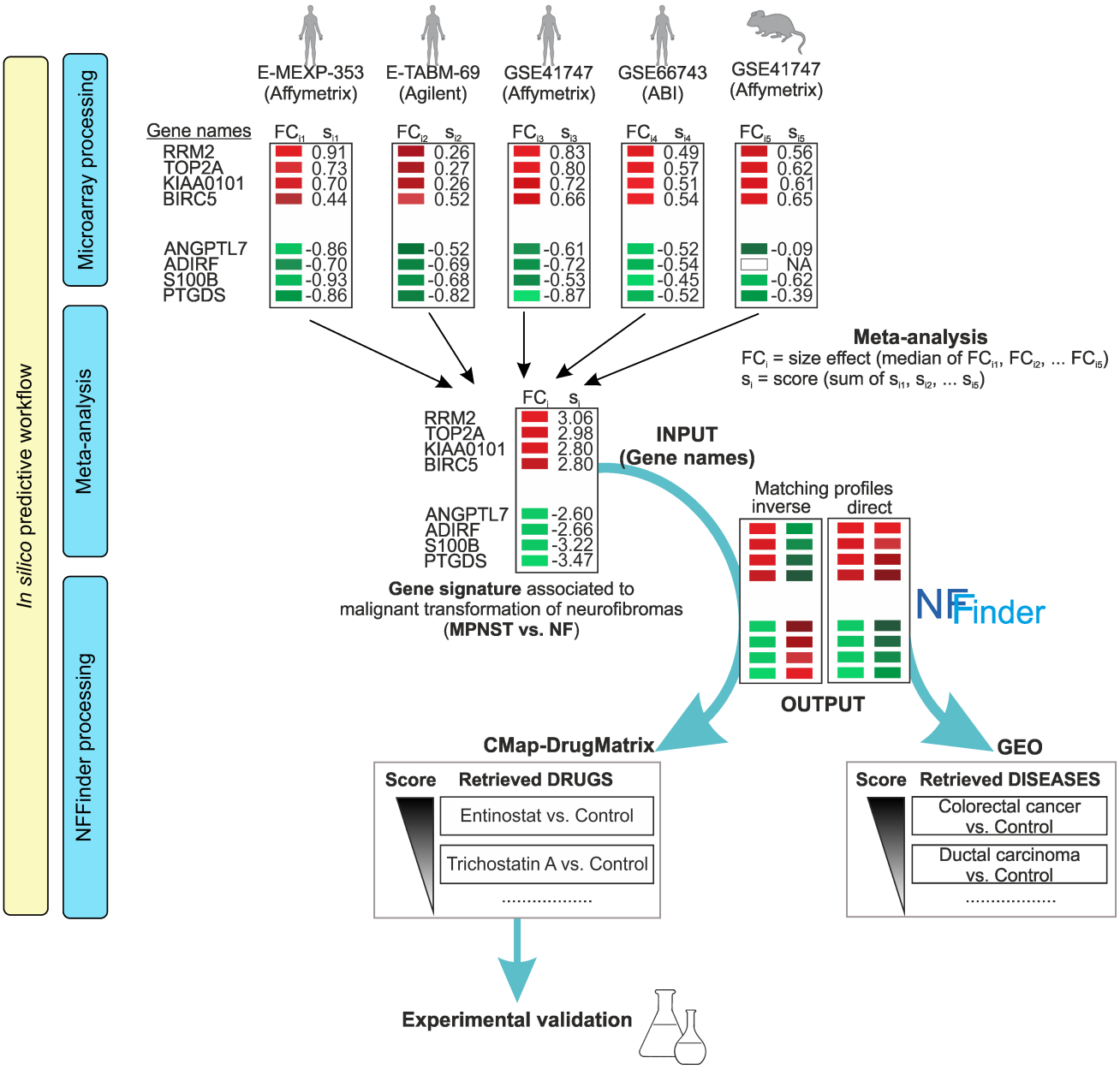


Fig 2. General schema of the integration of studies to generate the MPNST vs. NF gene signature and its processing by NFFinder. The upper panel shows the five microarray studies used to compute the MPNST vs. NF gene signature by meta-analysis (central panel). Score values calculated for each gene in each study (s_{ij}) and for each signature gene (s_i) are shown next to rectangles that indicate the respective gene size effect or fold change (FC_{ij} and FC_i); red for upregulated genes and green for downregulated ones. The lower panel describes the results obtained from NFFinder when GEO and CMap-DrugMatrix databases are interrogated for direct or inverse matching of gene expression patterns, respectively, by using the signature gene names as input. Experimental validation should verify the hypotheses generated by this in-silico predictive workflow.

https://doi.org/10.1371/journal.pone.0178316.g002

mouse data, and the Bhattacharyya distance (BD) ratio, which replaced the entropy used by Rasche et al. [23] as an estimator of the homogenous contribution of each study to the final score (see Materials and methods) (Table A in S2 Table). The relationships between gene scores and additional attributes (Results A in S1 Appendix) and the final contribution of each individual experiment to the gene signature (Results B in S1 Appendix) are also evaluated.

Table 2. List of 20 genes with the highest and the lowest scores of the MPNST vs. NF signature[1].

ENSGENE[2]	hgnc_symbol[3]	chrom_name[4]	band	score	logFC[5]	logFC_m[6]	studies[7]	mouse[8]	BD_ratio[9]
ENSG00000171848	RRM2	2	p25.1	3.06	4.19	4.47	5	1	13.68
ENSG00000131747	TOP2A	17	q21.2	2.98	3.73	3.94	5	1	10.01
ENSG00000166803	KIAA0101	15	q22.31	2.80	4.02	4.08	5	1	8.80
ENSG00000089685	BIRC5	17	q25.3	2.80	3.13	3.22	5	1	5.24
ENSG00000137804	NUSAP1	15	q15.1	2.59	3.67	3.75	5	1	16.50
ENSG00000185686	PRAME	22	q11.22	2.50	3.43	3.65	4	0	7.08
ENSG00000149948	HMGA2	12	q14.3	2.44	3.54	3.57	5	1	21.31
ENSG00000117724	CENPF	1	q41	2.43	3.08	3.20	5	1	9.38
ENSG00000198901	PRC1	15	q26.1	2.43	3.05	3.14	5	1	16.71
ENSG00000157456	CCNB2	15	q22.2	2.34	2.99	3.11	5	1	14.17
ENSG00000128045	RASL11B	4	q12	2.19	3.39	3.81	4	0	17.65
ENSG00000156076	WIF1	12	q14.3	2.17	4.70	4.99	4	0	22.10
ENSG00000134057	CCNB1	5	q13.2	2.16	2.54	2.48	5	1	20.67
ENSG00000066279	ASPM	1	q31.3	2.15	2.93	3.02	5	1	11.88
ENSG00000088325	TPX2	20	q11.21	2.14	2.96	3.19	5	1	23.98
ENSG00000123975	CKS2	9	q22.2	2.07	2.21	2.30	5	1	17.55
ENSG00000143476	DTL	1	q32.3	2.07	3.20	3.30	5	1	26.58
ENSG00000170312	CDK1	10	q21.2	2.01	3.12	3.17	4	1	17.50
ENSG00000007062	PROM1	4	p15.32	1.99	4.19	4.54	4	0	12.37
ENSG00000115163	CENPA	2	p23.3	1.94	2.77	2.93	5	1	18.19
ENSG00000021300	PLEKHB1	11	q13.4	-1.66	-2.96	-3.13	5	1	35.86
ENSG00000109846	CRYAB	11	q23.1	-1.68	-3.13	-3.20	5	1	38.39
ENSG00000100146	SOX10	22	q13.1	-1.74	-2.81	-2.93	5	1	35.15
ENSG00000100307	CBX7	22	q13.1	-1.79	-2.33	-2.37	5	1	18.08
ENSG00000149218	ENDOD1	11	q21	-1.82	-2.40	-2.44	4	1	18.78
ENSG00000197766	CFD	19	p13.3	-1.87	-2.98	-2.53	5	1	43.75
ENSG00000148180	GSN	9	q33.2	-1.90	-2.60	-2.58	5	1	7.053
ENSG00000134121	CHL1	3	p26.3	-1.94	-3.63	-3.36	5	1	32.93
ENSG00000172005	MAL	2	q11.1	-1.97	-3.05	-3.24	5	1	16.49
ENSG00000174944	P2RY14	3	q25.1	-2.01	-3.39	-3.47	5	1	35.95
ENSG00000108381	ASPA	17	p13.2	-2.06	-2.29	-2.40	5	1	31.63
ENSG00000168477	TNXB	6	p21.32	-2.07	-3.13	-3.34	5	1	26.58
ENSG00000127951	FGL2	7	q11.23	-2.08	-3.92	-3.75	5	1	27.43
ENSG00000196616	ADH1B	4	q23	-2.08	-3.86	-3.88	5	1	38.97
ENSG00000071991	CDH19	18	q22.1	-2.15	-4.21	-4.01	5	1	33.78
ENSG00000147588	PMP2	8	q21.13	-2.33	-4.39	-4.80	5	1	14.94
ENSG00000171819	ANGPTL7	1	p36.22	-2.60	-3.85	-4.12	5	1	18.30
ENSG00000148671	ADIRF	10	q23.2	-2.66	-3.81	-4.00	4	0	3.22
ENSG00000160307	S100B	21	q22.3	-3.22	-4.31	-4.38	5	1	11.91
ENSG00000107317	PTGDS	9	q34.3	-3.47	-4.89	-4.85	5	1	7.88

[1]The complete MPNST vs. NF gene signature of 579 genes, embedded in the unfiltered list of 7064 genes, is bold-highlighted in Table A in S2 Table.

[2]ENSEMBL gene ID.

[3]Gene symbol from HUGO Gene Nomenclature Committee.

[4]Human chromosome name.

[5]Median of $logFC_{ij}$ computed for each gene across the studies.

[6]Median of $logFC\_m_{ij}$ computed for each gene across the studies.

[7]Number of studies included in the MPNST vs. NF meta-analysis.

[8]Inclusion (1) or exclusion (0) of mouse data in MPNST vs. NF meta-analysis. Exclusion may be due to the absence of mouse data or because current data differed from human data.

[9]Bhattacharyya distance ratio.

https://doi.org/10.1371/journal.pone.0178316.t002

In silico search for therapeutic drugs to reverse malignant phenotype

Analysis of similar or opposite gene signatures is one of the most important applications of gene signatures for generating hypotheses for the study of NF1 and other rare diseases. We used NFFinder [20] to explore repurposing of drugs that might reverse the NF1-associated MPNST malignant phenotype by inspecting CMap and DrugMatrix databases for gene expression patterns opposite to the MPNST vs. NF gene signature (Fig 2, NFFinder processing). The complete score-ranked list of drugs with pval <0.005 can be seen at goo.gl/IdyV1N; Table A in S3 Table shows the first 50 drug entries retrieved. In the top two positions was Entinostat MS-275, a histone deacetylase (HDAC) inhibitor selective for class I HDAC. The non-specific HDAC inhibitor Trichostatin A (TSA) appeared in 39 of the top 50 entries retrieved. Two other HDAC inhibitors, the class I-selective HC-toxin and the non-specific HDAC inhibitor Scriptaid, were also on the shortlist. Other anti-cancer compounds were rifabutin, an antibiotic effective against lung cancer cells, PNU-0251126, which correlated positively with drugs for leukemia treatment, the protein phosphatase 2A inhibitor cantharidin, which induces cell death, the anti-inflammatory steroid medrysone, the topoisomerase II inhibitor ellipticine, a potent antineoplastic agent, and the non-steroidal selective estrogen receptor (ER) modulator tamoxifen, used to treat ER-positive breast cancer. To identify conditions similar to the NF-to-MPNST transformation that might share therapeutic treatments, we used NFFinder to search for disease gene signatures from GEO experiments with expression patterns resembling the MPNST vs. NF gene signature; results are shown in goo.gl/36xlB1 and the first 50 GEO studies are summarized in Table B in S3 Table. The most similar diseases were other types of cancer (58%), of which 73% were solid tumors, especially in prostate and breast, 20% were leukemias and lymphomas, and 7%, tumor cell lines. We also found premalignant neoplasias of epithelial tissue in endometrium and kidney (12%), lipid metabolism conditions (6%), pulmonary diseases (4%), muscular dystrophy (4%) and neuronal conditions (2%).

The experimental validation of the therapeutic effectiveness of drugs retrieved by NFFinder to treat NF1-associated MPNST was reported for HDAC inhibitors [28], cantharidin [29] and tamoxifen [30], and clinical trials have been suggested for all of them. The conclusion of our in-silico workflow (Fig 2) with these experimental data confirms the predictive value of the MPNST vs. NF gene signature and its usefulness to identify robust biomarkers and therapeutic agents in neurofibromatosis disease.

Functional characterization of the MPNST vs. NF gene signature

We provide detailed lists of GO term enrichment of MPNST vs. NF up- and downregulated genes (Tables A to F in S4 Table), as well as over-represented pathways derived by analysis of KEGG (Tables G and H), Wiki (Tables I and J), and Reactome (Tables K and L) databases. Results C in S1 Appendix details genes in the pathways identified.

For upregulated genes, DNA replication and cell cycle pathways were shared in the three pathway databases; 70% of GO terms associated to these genes were thus involved mainly in processes that underlie positive regulation of cell proliferation, mitosis and meiosis. We also found terms related to morphogenesis and development of skeletal, nervous, cardiovascular and digestive systems. GO terms directly related to malignancy of proliferating cells included collagen catabolic process, which can affect extracellular matrix (ECM) organization, a process involved in epithelial to mesenchymal transition (EMT) and cell migration. The KEGG ECM-receptor interaction pathway was also over-represented, which supports this observation. In accordance with GO terms and pathways, components from nucleus and cytoplasm were associated predominantly with upregulated genes.

For the downregulated genes, all three pathway databases identified the immunity pathway linked to complement activation, including regulatory elements and complement components. Downregulated genes also appeared to control peripheral nervous system development, particularly *via* myelination and axonogenesis. Other downregulated genes participated in the response to steroid hormones and in cell migration, chemotaxis, and cell adhesion to another cell or to a substrate such as the ECM. Unlike upregulated genes, downregulated genes accumulated GO terms related to plasma membrane and its intrinsic components, including elements from proteinaceous ECM involved in cell junction formation. When we analyzed the unfiltered list of 3005 downregulated genes, from which we derived the MPNST vs. NF downregulated gene signature, we found that specific KEGG pathways in cancer were over-represented (Table P in S4 Table).

In accordance with this functional characterization, most upregulated genes from Table 2 are involved in cell cycle progression, have been linked with carcinogenic processes in which they are upregulated, and many have also been described as diagnostic biomarkers (S5 and S6 Tables). HMGA2 is the only gene product that is an architectural transcription factor, although some other genes act as regulatory elements in mitosis, especially cyclins CCNB2, CCNB1, and kinase CDK1. As anticipated, and in contrast to the mitosis-related products of upregulated genes, which were found mainly in nucleus and cytoskeleton, the 20 most-downregulated gene products (Table 2) act mainly in the extracellular space. Whereas upregulated genes are overexpressed in cancer, some downregulated genes can be silenced or overexpressed, and thus have dual roles in cancer, as tumor suppressors or cell proliferation promoters. Some of these genes appear to be involved in signal transduction and development in the nervous system *(PLEKHB1*, *SOX10*, *CHL1*, *PTGDS)*, in maintenance of structural integrity of cells and tissues, e.g., ECM formation or myelin synthesis *(CRYAB*, *CBX7*, *GSN*, *MAL*, *TNXB*, *CDH19*, *PMP2*, *ANGPTL7*, *ADIRF)*, or to have various roles in the immune response *(ENDOD1*, *P2RY14*, *FGL2*, *S100B)*. SOX10 and S100B coregulate Schwann cell proliferation and myelination [31]. CBX7 is involved in epigenetic transcriptional repression. The metabolic proteins ASPA and ADH1B are also in the most-downregulated group. ASPA participates in increasing the pool of acetate, an essential precursor for histone acetylation reactions. ADH1B oxidizes alcohol, thus also helping to generate acetate precursors, as well as retinol, an early step in synthesis of retinoic acid, a basic molecule in epithelial tissue growth and differentiation.

Chromosome distribution of genes in the MPNST vs. NF signature

To establish whether there are significant distribution differences in human chromosomes for genes in the MNPST vs. NF signature (336 up- and 243 downregulated genes), we computed frequencies of the number of genes per chromosome arm from the gene signature and the whole human genome (Fig 3). Upregulated genes accumulated significantly in chromosome arms 4q, 7p, 8q and 17q, whereas downregulated genes were over-represented mainly in 1q, 3p, 5q and 11q. Another binomial distribution identified chromosome bands in which both up- and downregulated genes were over-represented. The highest concentration of upregulated genes was found in bands 3q25.33 and 15q15.1, with others concentrated in bands 7p15.3 and Xq22.1; the largest number of bands with over-representation of upregulated genes was seen in arm 15q and the distal part of 17q. The highest significant concentration of downregulated genes was found in bands 1q24.3 and 19q13.12; arm 1q showed the largest number of over-represented bands. The function of genes in over-represented chromosome regions is shown in S7 Table. Enrichment in GO biological process terms coincided with the functions of up- and downregulated genes in the MPNST vs. NF signature, as described above.

Comparison of chromosome region enrichment from the MPNST vs. NF gene signature with that in S2 Fig for the unfiltered gene list indicated some shared over-represented regions;

chromosome arms 7p, 8q and 17q accumulated upregulated genes on the unfiltered list, whereas arms 3p and 11q concentrated downregulated genes. The distal part of chromosome arm 17q showed the largest number of over-represented bands. In addition, the unfiltered list showed over-represented upregulated genes in arms 1q, 2p, 2q, 6p, 7q, 19p and Xq, and down-regulated genes in 1p, 9q, 10q, 14q, 17p and 20p.

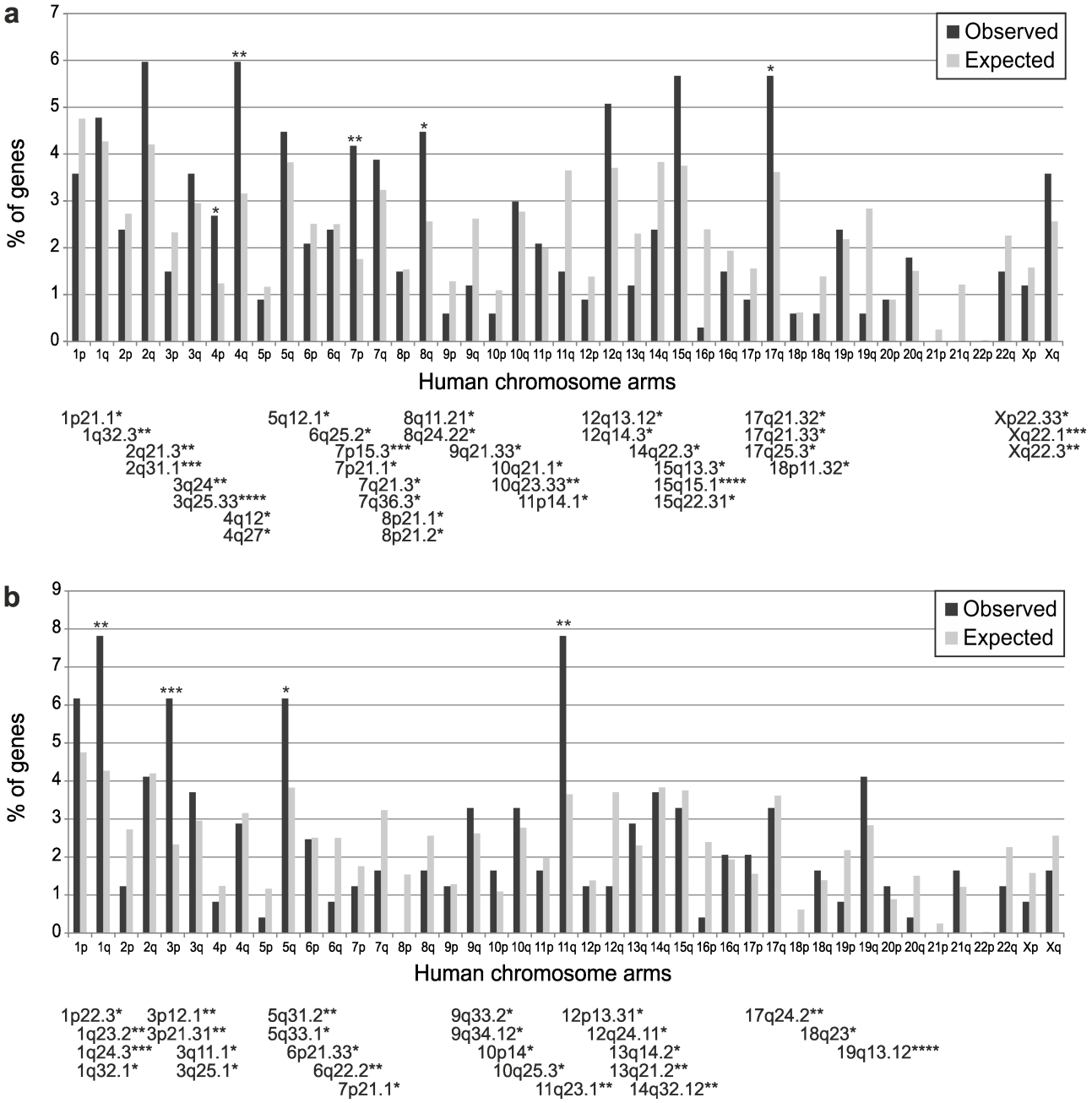


Fig 3. Chromosome distribution of the MPNST vs. NF gene signature. The gene signature distribution was calculated from the 336 genes with positive score **(a)**, and from the 243 genes with negative score **(b)**. Bar diagrams compare the observed distribution of MPNST vs. NF gene percentage in the human chromosome arms (dark bars) with the expected distribution according to the human ENSEMBL database (light bars). Statistical significance of the gene signature over-represented chromosome arms is above the bars. Over-represented human chromosome bands in the MPNST vs. NF gene signature are shown below each chart. Their statistical significance is shown at the top right side of band names. (****) $P(X \geq x) < 0.0001$, (***) $0.0001 < P(X \geq x) < 0.001$, (**) $0.001 < P(X \geq x) < 0.01$, (*) $0.01 < P(X \geq x) < 0.05$.

https://doi.org/10.1371/journal.pone.0178316.g003

Panels of biomarkers with similar expression profiles in the context of NF1 disease

Given the greater potential prognostic capacity of biomarker panels compared to individual biomarkers, we built robust gene profiles to group functionally related genes with similar expression patterns. To obtain gene profiles, we extended the gene score methodology to identify other average gene signatures in the NF1 context. We performed five additional differential expression analyses; Table 1 groups the accession numbers of the studies used to determine gene signatures for the five new comparisons. Two of these new comparisons involved nerve tumors, whereas the other three comprised cell cultures. For tumor tissue, we carried out two meta-analyses that integrated human and mouse nerve tumors by comparing NF vs. control (nerve tissue) and MPNST vs. control. For cell cultures, we made three comparisons. Only the comparison of MPNST cell lines vs. control (NHSC; normal human Schwann cells) integrated data from two studies; for the other two comparisons, primary NFSC (neurofibroma Schwann cells) vs. control (NHSC) and MPNST cell lines vs. primary NFSC, we used standard differential expression analyses, as only one study was available. These additional gene signatures were added as Tables B to J in S2 Table. Their composition, and a comparison with the MPNST vs. NF gene signature are detailed in Results D in S1 Appendix.

To search for function-related genes in the context of neurofibromatosis, we sought genes with similar expression patterns in the six comparisons of NF, MPNST and control phenotypes in cell cultures and nerve tumors; S8 Table shows the list of 2209 genes obtained by gathering signatures of these comparisons. Comparisons (clustered by tissue type) and genes grouped independently, and dendrograms were determined of the comparisons and of the genes, grouped in 46 clusters represented by their self-organizing tree algorithm (SOTA) centroid vectors (Fig 4). Divergence between sample tissue types coincided with dissimilarities in the functional characterization of genes in the cell culture and nerve tumor signatures (Results D in S1 Appendix). As predicted, the second line of sample divergence separated MPNST samples (vs. NF or vs. control) from NF vs. control comparisons. Results E in S1 Appendix evaluates gene pattern homogeneity in each cluster.

Although almost every cluster contained genes from the MPNST vs. NF tumor tissue signature (highlighted in S8 Table), the most populated clusters (>15 genes) appeared in four cluster areas, from which were derived four biomarker panels that included the majority of the 20 most up- and downregulated genes (Fig 4). Cluster 3, which includes genes involved in development of the peripheral nervous system, contained six genes of the 20 most-downregulated (Table 2), *CRYAB*, *SOX10*, *FGL2*, *CDH19*, *PMP2* and *S100B*; *SOX10* and *S100B* are markers used to diagnose neural crest-derived tumors [32]. Cluster 17 incorporated three of the most-downregulated genes, *PLEKHB1*, *GSN* and *CHL1*. Cluster 25 included 153 genes and 181 GO terms, and thus concentrated the highest functional enrichment. This cluster, with adjacent clusters 23 and 24, accumulated the majority of genes involved in cell proliferation. Cluster 25 included 17 of the 20 genes with the highest scores from the MPNST vs. NF comparison (Tables 2 and S8), particularly the genes that encode the putative NF1 prognostic markers *TOP2A* and *BIRC5* [10]. Finally, cluster 34 contained five of the 20 most-downregulated genes in the MPNST vs. NF signature, *CBX7*, *P2RY14*, *TNXB*, *ADIRF* and *PTGDS*. Details for gene accumulation in clusters from the MPNST vs. NF signature, and for the five additional signatures, are shown in S3 Fig.

Panel of hypermethylated biomarkers in the context of NF1 disease

Aberrant epigenetic changes are associated to most cancers, particularly hypomethylation of oncogenes and hypermethylation of tumor suppressors. To establish relationships between

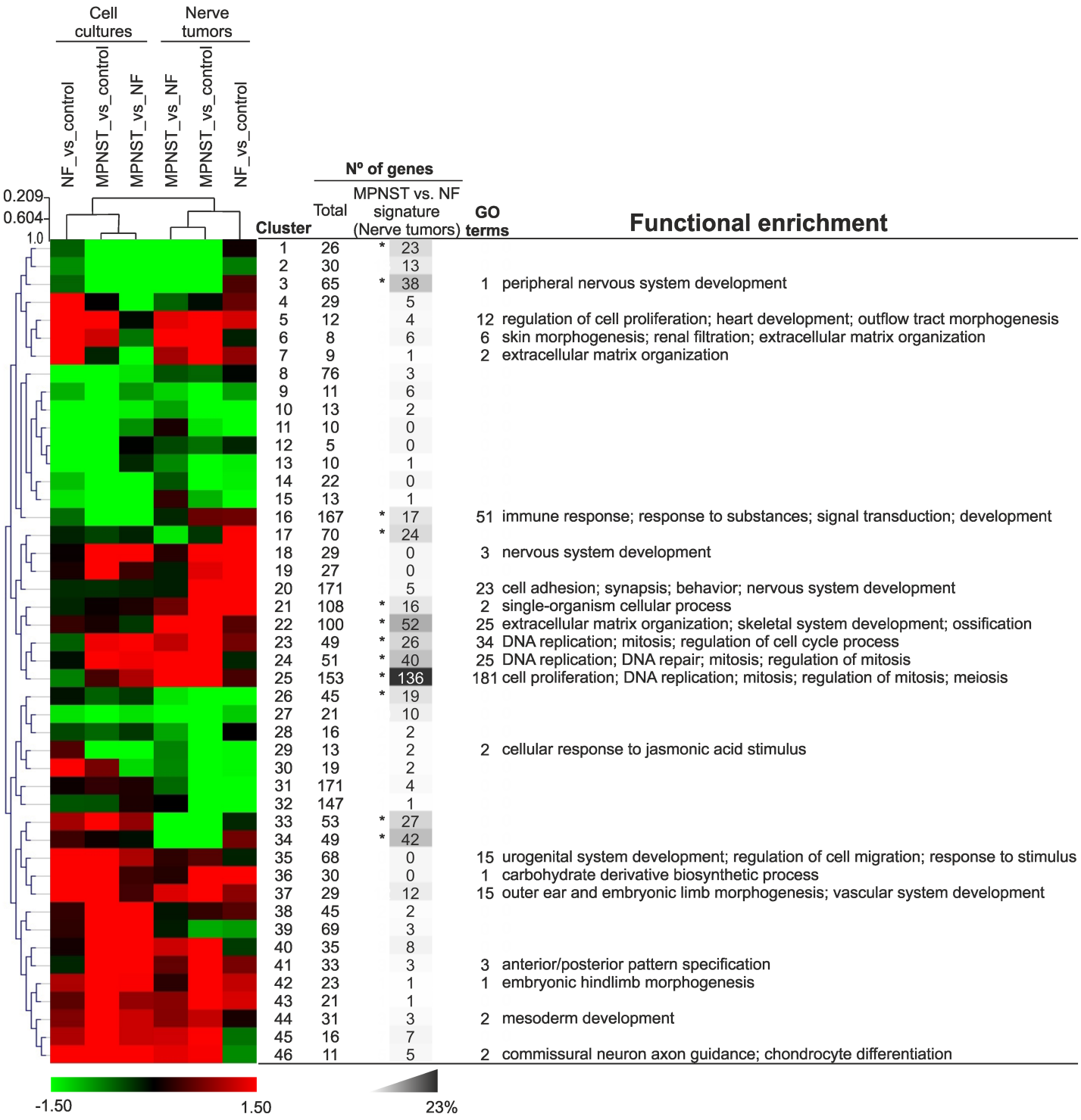


Fig 4. Clustering of phenotype comparisons and of NF1-related genes. Comparisons among MPNST, NF and control phenotypes were grouped through hierarchical clustering. Cell culture and nerve tumor comparisons are shown on the top of the upper dendrogram, and the node height scale is detailed on the left of this tree. The hierarchical relationship among gene clusters obtained by grouping the logFC_m values of 2209 NF1-related genes by the Self-Organizing Tree Algorithm (SOTA) is represented by the dendrogram on the left. Clusters are described by their SOTA centroid vectors. Color scale of logFC_m values is shown below. The right side of the Fig details the number of genes in each cluster, the gene number of the MPNST vs. NF signature in each cluster (*, > 15 genes), the percentage of genes of this signature in each cluster (grey scale shown below), the number of biological process GO terms over-represented in each cluster, and the summary of the GO term enrichment as functional characterization of clusters. A complete list of terms is shown in S9 Table.

https://doi.org/10.1371/journal.pone.0178316.g004

gene expression and aberrant promoter methylation, we associated our six gene expression signatures with the methylation status of gene promoters. In GEO and ArrayExpress databases, we sought studies that provide raw methylation data relative to NF1 evolution. GEO accession GSE21714 genome-wide methylome analysis compared immunoprecipitated DNA from 10 pooled MPNST samples, 10 pooled NF samples, and 6 pooled control Schwann cell samples [13]; for the process used to identify methylated promoters, see Materials and methods. The edgeR.logFC and edgeR.adj.pval (<0.1) values obtained for MPNST vs. NF (Table A in S2 Table), MPNST vs. control (Table B), and NF vs. control (Table C) methylome comparisons showed 428, 594 and 9 differentially methylated gene promoters, respectively (right columns); most were hypermethylated. MPNST vs. NF and MPNST vs. control comparisons showed 181 promoters in common, and MPNST vs. control and NF vs. control Schwann cells indicated 4 shared promoters.

Hypermethylation of the *RASSF1A* promoter identified a NF1-associated MPNST subgroup with poor prognosis [14]; we therefore tested *RASSF1* promoter methylation status as a control of our analysis. The *RASSF1* promoter was differentially hypermethylated in MPNST vs. NF and in MPNST vs. control Schwann cells, but not in NF vs. control Schwann cells. To find other potential regulatory elements with aberrant promoter methylation, we searched for other downregulated genes with hypermethylated promoters in MPNST vs. NF and MPNST vs. control Schwann cell comparisons. Using the unfiltered gene lists for all comparisons (Table J in S2 Table), we selected genes whose expression pattern in tumor tissue included a negative MPNST vs. NF score, as well as logFC values < -0.5 for MPNST vs. NF and MPNST vs. control comparisons. Genes selected also showed edgeR.logFC and edgeR.adj.pval promoter methylation values >1.5 and <0.1, respectively. We found 56 gene promoters that fit these selection conditions. The *RASSF1* gene and 9 of the selected genes found in the MPNST vs. NF gene signature are listed in Table 3; the full list is given in S10 Table.

Discussion

MPNST vs. NF gene expression signature

Here we report the integrative MPNST vs. NF gene signature associated to the NF-to-MPNST transition in the context of neurofibromatosis type 1 disease. For this signature, filtered from a larger list of all differentially expressed genes from five studies comparing MPNST to NF, we included genes with the highest scores, size effects and consensus among individual studies, and used it as a core to study biomarkers and drugs that might control evolution to malignity in NF1 patients.

The 20 highest and 20 lowest score-ranked genes suggested as promising biomarkers (Table 2) are all implicated in carcinogenic processes and many are suggested biomarkers. Among the top 20 genes in the MPNST vs. NF signature are 4 genes reported as induced in the transition from benign NF to MPNST (S11 Table); their products are the topoisomerase TOP2A, needed for correct chromosome segregation in mitosis, the apoptosis inhibitor BIRC5, a member of the mitotic chromosome passenger complex, the architectural transcription regulator HMGA2, and TPX2, which is essential for correct mitotic spindle assembly and activates the AURKA kinase to control cell cycle progression. Among the 20 most-downregulated genes are 4 NF- and/or MPNST-associated genes (S12 Table), the transcriptional regulator SOX10, needed for neural crest multipotent cell and peripheral nervous system development, the extracellular matrix glycoprotein TNXB, the cell-cell adhesion Ca^{2+}-dependent cadherin CDH19, specific to myelin-forming cells, and the Ca^{2+}-binding S100B, whose inhibition is associated to MPNST transformation [33].

The *NF1* gene, the basic reference of reliability for the gene signature generated, had a null score and was thus absent from the MPNST vs. NF signature. This suggested that most samples

Table 3. MPNST vs. NF signature genes potentially down-regulated by promoter hypermethylation[1].

ENSGENE[2]	hgnc symbol[3]	ch[4]	band	NF vs. control				MPNST vs. control						MPNST vs. NF					
				Cell cultures		Nerve tumors		Cell cultures		Nerve tumors				Cell cultures		Nerve tumors			
				score	logFC[5]	score	logFC[5]	score	logFC[5]	score	logFC[5]	*edgeR*[6]		score	logFC[5]	score	logFC[5]	*edgeR*[6]	
												logFC	*adj.pval*					*logFC*	*adj.pval*
ENSG00000068028	RASSF1	3	p21.31	0	-0.36	-0.08	-0.47	-0.07	-0.91	0	-1.07	3.35	1.64e-03	-0.03	-0.70	-0.09	-0.70	2.77	1.30e-02
ENSG00000159200	RCAN1	21	q22.12	0	-0.93	0	0.73	-0.76	-3.18	0	-1.11	2.99	1.88e-07	-0.19	-2.37	-0.64	-2.08	2.41	1.28e-02
ENSG00000197971	MBP	18	q23	-0.56	-4.38	-1.28	-4.28	-1.69	-5.89	-1.54	-7.13	3.80	6.60e-59	-0.02	-2.99	-0.69	-2.96	3.41	1.06e-53
ENSG00000175785	PRIMA1	14	q32.12	0	-2.20	0	-0.32	-0.58	-2.75	-0.67	-2.40	3.09	5.52e-07	0	-1.43	-0.71	-2.43	2.87	4.01e-05
ENSG00000152061	RAGAP1L	1	q25.1	0	0.50	0	-0.40	0.06	1.11	-0.24	-1.94	2.55	5.11e-06	0	0.86	-0.73	-1.54	2.24	1.74e-05
ENSG00000213088	ACKR1	1	q23.2	0	0.07	0	-0.77	0	-0.02	-0.78	-3.15	2.03	2.30e-04	0	0	-1.00	-2.84	3.33	1.04e-02
ENSG00000158887	MPZ	1	q23.3	0	-1.83	0	-1.45	-1.22	-4.66	-1.01	-5.09	4.37	1.95e-08	-0.70	-5.51	-1.11	-2.49	2.27	3.46e-03
ENSG00000070731	ST6GALNAC2	17	q25.1	-0.18	-1.97	-0.25	-1.26	-1.33	-4.69	-0.69	-2.93	2.74	4.47e-06	-0.47	-3.54	-1.56	-2.81	2.01	1.31e-02
ENSG00000132470	ITGB4	17	q25.1	0	0.02	0	-0.62	0	0.16	-0.55	-2.29	2.16	4.77e-02	0	-0.12	-1.62	-3.13	1.90	6.70e-05
ENSG00000160307	S100B	21	q22.3	0	-1.40	0	-0.72	-1.81	-8.33	-1.06	-4.94	2.37	6.82e-04	-0.86	-7.66	-3.22	-4.31	2.56	2.02e-02

[1]Data extracted and ordered from gene list showed in Table J in S2 Table. The complete list of genes potentially down-regulated by hypermethylation in MPNST vs. NF and vs. control comparisons is shown in S10 Table.

[2]ENSEMBL gene ID.

[3]Gene symbol from HUGO Gene Nomenclature Committee.

[4]Human chromosome name.

[5]Median of $logFC_{ij}$ computed for each gene across the studies comparing two phenotypes.

[6]*edgeR.logFC* and *edgeR.adj.pval* computed by the analysis of promoter methylation. NA values of *edgeR.logFC* and *edgeR.adj.pval* were omitted for NF vs. control comparison.

https://doi.org/10.1371/journal.pone.0178316.t003

of sporadic MPNST had somatic deactivating *NF1* mutations, which corroborates previous studies [34]. Our results coincide with gene expression data that compare MPNST with benign NF, although given the strict score threshold, some known MPNST-related genes on the unfiltered gene list were lacking in the gene signature, such as *SOX9*, which encodes a developmental transcription factor [3], and *TNC*, which is involved in axon guidance during development [16]. Other cancer-related genes were absent from the signature for similar reasons, such as *CDH1*, which codes for the tumor suppressor cadherin 1. CDH1 silencing is a marker of the EMT associated to cell proliferation, invasion and metastasis [35].

We found that despite some exceptions, our results for over-represented chromosome regions in the MPNST vs. NF gene signature corroborated previous data on chromosome region enrichment. In addition, we identified other regions such as chromosome arm 3p, in which many downregulated genes are concentrated. A detailed comparison of these regions and reported aberrant chromosome modifications is shown in Discussion A in S1 Appendix.

Biomarker panels that share NF1-related gene expression profiles

In addition to the study of individual genes as potential biomarkers, tests of groups of genes in the MPNST vs. NF signature can improve prospects for diagnosis or prognosis. Kolberg et al. [10] defined a prognostic test for post-tumor resection MPNST patients, based on expression of three proteins encoded by genes located in the distal part of chromosome 17q *(BIRC5*, *TK1*, *TOP2A)*. These genes were included in a 31-gene cell cycle progression (CCP) signature that is a robust predictor of clinical outcome for prostate cancer patients [36].

Given the limitation of the MPNST vs. NF score-ranked gene signature for grouping genes with similar function, we built profiles that integrated other NF1-related signatures. These signatures, obtained from tumor tissue and cultured cells, were based on one or two studies each and were thus less robust than our original MPNST vs. NF, built from five studies. Genes were grouped by their profiles using SOTA, which allows hierarchical clustering by a neural network. We obtained a large number of SOTA clusters containing homogeneous groups of functionally related genes that shared similar expression profiles.

Of the upregulated genes in the MPNST vs. NF signature, most grouped to Cluster 25 and are linked to cell cycle progression. Mechanisms that regulate the complex process of cell cycling are precisely regulated, and deregulation drives to aberrant cell proliferation and cancer development [37]. Not surprisingly, many genes from the CCP signature [38] were included in the cluster 25. Our group of the 20 most-upregulated genes contained 9 of the 31 CCP signature genes (*RRM2*, *TOP2A*, *KIAA0101*, *BIRC5*, *NUSAP1*, *CENPF*, *PRC1*, *ASPM*, *DTL*), and 14 more CCP genes were upregulated in the MPNST vs. NF gene signature (*FOXM1*, *TK1*, *CDC20*, *BUB1B*, *PBK*, *CDKN3*, *ASF1B*, *CEP55*, *DLGAP5*, *RAD51*, *KIF11*, *KIF20A*, *PTTG1*, *CDCA8*). Due to its potential as MPNST biomarker [39], we propose the inclusion of *HMGA2* in any similar list built to interrogate the expression status of genes involved in cell cycle progression in MPNST prognosis. A panel from cluster 3 should contain at least *CRYAB*, *SOX10*, *FGL2*, *CDH19*, *PMP2*, and *S100B*. Cluster 17 genes to be evaluated would be *PLEKHB1*, *GSN* and *CHL1*. A cluster 34 panel should include at least *CBX7*, *P2RY14*, *TNXB*, *ADIRF* and *PTGDS* genes.

The divergence between expression values in comparisons of cell cultures and nerve tumors indicated clear differences due to the nature of the samples. Many essential characteristics of cancers in which the ECM plays a fundamental role, such as those involving EMT, invasion and metastasis, cannot be appropriately mimicked by cultured cells. In contrast to MPNST, which contain a mixture of different cell types, cultured cells derive from a single cell type. Differences between cultured cells and nerve tumors, exemplified by *SOX9*, *SUZ12*, *EGFR*, *SPP1* and *BMP2* genes, are detailed in Discussion B in S1 Appendix.

Genes potentially silenced by hypermethylation of their CpG-island promoter region

To identify potential biomarkers whose expression is inhibited by hypermethylation, we sought genes that were downregulated and showed promoter hypermethylation in tumor tissue for MPNST vs. NF and MPNST vs. control comparisons; 42% of the differentially methylated promoters were shared, most of which were hypermethylated. This result correlated with gene expression data, as more than 58% of genes from the MPNST vs. NF showed similar behavior in the MPNST vs. control comparison. In addition to the tumor suppressor *RASSF1*, whose differential promoter hypermethylation status and expression inhibition are associated in MPNST and other cancers [14], we found additional regulatory genes. The genes identified by common expression and methylation patterns were implicated mainly in cancer, which corroborates the importance of epigenetic changes in the control and evolution of cancerous processes; they were also linked to immune response, nervous system development, lipid metabolism, metabolic energy balance and detoxification (S10 Table). Many act as mediators in signal transduction pathways, have a structural role, or interact with other proteins to control various biological processes related to cell proliferation and apoptosis. Most genes in the MPNST vs. NF signature that are potentially silenced by promoter hypermethylation are associated with structural functions in the peripheral nervous system (*PRIMA1*, *ST6GALNAC2*, *ITGB4*, *MBP*, *MPZ*) or with signal transduction pathways (*RABGAP1L*, *ACKR1*, *S100B*) (S10 Table, bold). Differentially methylated genes are discussed in detail in Discussion C in S1 Appendix.

From MPNST vs. NF gene signature to working hypotheses: Therapeutic applications of HDAC inhibitors, cantharidin and tamoxifen

To identify diseases or biological problems similar to malignant NF evolution, or to define potential therapeutic drugs that could reverse malignant phenotype, we used NFFinder to compare the MPNST vs. NF signature with similar or contrasting signatures. Gene signatures in GEO studies most similar to that of MPNST vs. NF were associated with cancer, mainly solid tumors, followed by other diseases that share several phenotypic alterations with neurofibromatosis. Most signatures in CMap and DrugMatrix databases that contrasted from the MPNST vs. NF signature indicated that the reported as effective compounds to treat MPNST, cantharidin and tamoxifen, and especially HDAC inhibitors, could potentially reverse the malignant phenotype. Our in-silico prediction corroborates these data, which indicates the reliability of our gene signature as representative of NF evolution to malignancy.

HDAC inhibitors control gene expression by blocking deacetylation of histone and non-histone proteins. These inhibitors modify the chromatin condensation status [40] and also control several chromatin structure-independent processes that alter gene expression, such as transcription factor activity, miRNA expression, and signal transduction [41]. Some HDAC inhibitors act as epigenetic regulators by modifying the DNA methylation status, which reveals crosstalk between acetylation and methylation [42]. This ability of HDAC inhibitors to reverse epigenetic aberrations make them effective therapeutic agents in cancer, as well as in neurological and immune disorders [43]. In cancer, HDAC inhibitors impair cell proliferation, neoangiogenesis and metastasis, and increase differentiation and apoptosis. Blockade of tumor angiogenesis nonetheless hinders drug delivery and limits the use of these inhibitors for solid tumors [44]. Combination therapies are thus necessary.

Based on the sensitivity of Ras signaling tumors to HDAC inhibitors, López *et al.* assessed these compounds in MPNST, both *in vitro* and in tumor xenografts [28]. The strong reaction of NF1-associated MPNST cell lines to HDAC inhibitors led the authors to suggest their

therapeutic value for inclusion in clinical trials. NFFinder also found that the pan-HDAC inhibitor TSA can replace the effective combination of synergy-acting compounds PD-901 and JQ1 to kill MPNST cells and shrink tumors [12,20]. Robust biomarkers are nonetheless needed to predict the effectiveness of TSA and other HDAC inhibitors, alone or in combination, in clinical trials for treatment of NF1-associated MPNST.

Among the 20 most up- and downregulated, several genes alter their expression in response to HDAC inhibitors such as the upregulated genes *TOP2A*, *BIRC5*, *HMGA2*, *CCNB1*, *CCNB2*, *TPX5* and *CDK1* [45–49], and the downregulated genes *SOX10* [50] and *GSN* [51]. As components of the epigenetic regulation system, HMGA2 and GSN are main targets of HDAC inhibitors, and suggest their utility as predictive treatment markers.

HMGA2 controls gene transcription directly *via* chromatin remodeling, or indirectly, by altering the binding affinity of regulators and nuclear proteins to DNA through protein-protein interactions. Involved in the control of fetal development, HMGA2 also has a central role in tumor growth and metastasis. HMGA2 protein levels rise acutely in malignancies [52], and its overexpression correlates with poor prognosis in colon, lung, pancreas, ovary and gastric cancers [53,54]. Besides use of *HMGA2* as a diagnostic/prognostic biomarker in NF progression to malignity, the reported silencing of HMGA2 with HDAC inhibitors justifies its use as a biomarker of treatment effectiveness of these drugs [55].

In addition to its structural role as one of the most abundant actin-binding proteins, the multifunctional regulator gelsolin GSN is involved in apoptosis and regulates processes related to pathological states such as amyloidosis, inflammation, Alzheimer's disease, cardiovascular diseases, cancer and aging [56]. In cancer, GSN has a dual effect as a promoter of cell growth and invasion [57] and as a tumor suppressor that inhibits metastasis. Its tumor suppressor effect is reported for most cancers, and it is downregulated in all of them [58]. GSN transcriptional repression is associated with epigenetic control through DNA methylation and histone deacetylation, and addition of HDAC inhibitors increases GSN expression [51], which supports its use as a biomarker for the effectiveness of MPNST treatment with HDAC inhibitors.

HMGA2 and GSN, as well as EZH2 and CBX7, respective members of PRC2 and PRC1 epigenetic repressor complexes, probably impaired during NF malignant transformation, are involved in the control of cell proliferation and metastatic phenotype by regulation of the tumor suppressor protein CDH1 and other markers of EMT, whose expression is reversed after treatment with HDAC inhibitors, particularly TSA [59]. HDAC inhibitor activity counteracts CBX7 silencing and EZH2 protein overexpression over the *CDH1* promoter, as discussed in Discussion D in S1 Appendix.

In addition to genes involved in epigenetic regulation as possible HDAC inhibitor targets, we found two other genes related to acetate metabolism and availability (*ASPA*, *ADH1B*) among the 20 most-underexpressed genes of the MPNST vs. NF signature. Supplementation with acetate precursors as coadjuvant chemotherapy to complement two metabolic pathways that involve ASPA and ADH1B in acetate synthesis is considered in Discussion E in S1 Appendix.

Cantharidin and tamoxifen were also retrieved by NFFinder and confirmed experimentally as effective drugs to inhibit MPNST cultured cell proliferation and survival. The protein phosphatase 2A inhibitor cantharidin was found by screening from a library of 472 small bioactive compound library [29] and shown to avoid growth of NF1-associated MPNST cultured cells, though additional studies should clarify the relevance of cantharidin *in vivo*.

Due to MPNST are not sex steroid hormone sensitive, the ER modulator tamoxifen inhibits MPNST cultured cell growth independently of ER [30,60]. Combination therapies of tamoxifen and trifluoperazine have shown to be effective on treatment of sporadic and NF1-associated MPNST, suggesting the quick repurposing of these drugs for clinical and prophylactic

uses. Tamoxifen, which induces NF1-associated MPNST cell death mediated by autophagy in a K-Ras degradation dependent-process [61], might replace cloroquine and combine with HDAC inhibitors to induce apoptosis-oriented autophagy in "resistant" sporadic MPNST cells [28]. Tamoxifen and HDAC inhibitors might thus boost productive autophagy in a similar way as the combination of tamoxifen with panobinostat has been proposed for treatment of other solid tumors like hepatocellular carcinoma [62].

Meta-analysis method to define the MPNST vs. NF gene signature

Various meta-analysis approaches have been developed to integrate data from independent studies. Due to their inherent difficulties, most meta-analyses restrict the inclusion of studies to a small number of different platforms or to a single platform. Integration of genomics data for rare diseases presents even greater challenges. Scanty heterogeneous data, common for rare diseases, might hide consistent information and patterns potentially present across studies. The meta-analysis method used here, based on previously described gene scores [23,26], identifies differentially expressed genes between two conditions, by integrating data from platforms of distinct size. Although we used a similar formula, our gene score differed slightly from that published in five aspects; a) scores were computed for genes as up- and downregulated, b) the logFC computed, based on the mean of expression ratios, was replaced by the logFC_m based on the median value of expression ratios, which increased score robustness, c) logFC_m values were normalized to the interval [0,1] for upregulated and [-1,0] for downregulated genes to avoid expression ratio bias among studies, d) the use of MAD (median absolute deviation) rather than standard error values to compute the penalty term due to expression deviation among replicates, which resulted in lower values for the penalty term and thus in more stringent scores, and e) stringency was reinforced by a third condition imposed for calculation of scores (B factor >0). Computation of the gene final score by the addition of each gene study score might favor genes represented in a higher number of platforms. The last filter step selected 10% of the genes on the complete list, which yielded the MPNST vs. NF signature; this is nonetheless a more balanced representation of the gene list for each of the starting studies, as it contains a larger proportion of genes common to various studies.

In addition to problems frequently encountered when comparing transcriptomes, integration of human and mouse data has drawbacks inherent to species-specific differences. As mouse models do not appear to mimic human neurodegenerative diseases at molecular level in every respect [27], we prioritized the human over the mouse transcriptome. Although doubts remained regarding the incorporation of mouse data in the meta-analysis, we included those data since robust genes expressed similarly in humans and mice could be candidates for preclinical studies. As the gene signature includes the behavior of each gene regarding the species and tissue type, it provides comprehensive information for the design of experiments.

The integrative gene signature associated to neurofibroma malignant evolution is the main contribution of this work. This signature constitutes the first step to generate working hypotheses concerning biomarkers of disease evolution and treatment effectiveness, as well as therapeutic drugs. Our *in silico* methodology used to define the signature and derive hypotheses for experimental purposes could be applied in the study of orphan diseases other than NF1.

Conclusions

Here we used a meta-analysis method based on gene scores to define the gene signature associated to the transition of benign neurofibromas to MPNST. Signature components showing the highest/lowest scores are proposed as disease biomarkers. Given the greater diagnostic and prognostic robustness of biomarker panels compared to individual genes, we clustered

functionally related genes from profiles that integrate additional NF1-related gene signatures, and established four panels derived from clusters 3, 17, 34 and cluster 25, which comprises genes involved in cell cycle progression. By further studying the epigenetic regulation of malignant transformation, we identified potential hypermethylated, silenced biomarkers, which links promoter methylation status to gene expression profiles.

The gene signature was used to search for drugs able to reverse malignant phenotype. Retrieved from NFFinder and previously tested for effectiveness, cantharidin, tamoxifen, TSA and other HDAC inhibitors have yielded promising effects as candidates for chemotherapy. We suggest *HMGA2* and *GSN* genes, two targets of HDAC inhibitors, as epigenetic biomarkers for testing the therapeutic effectiveness of HDAC inhibitors.

Materials and methods

The detailed workflow of selection and pre-processing of microarrays used to obtain the MPNST vs. NF gene signature, translation to ENSEMBL gene names, and computation of gene scores and of effect size medians, and final filters are shown in S4 Fig. Where appropriate, tables are indicated for some of the main results generated (right). The meta-analysis PRISMA checklist is included as supporting information (S2 Appendix).

Study selection

Microarray and DNA methylation studies were selected from GEO and ArrayExpress public databases using the key words NF1, MPNST, neurofibromatosis, and neurofibroma; we included high-throughput sequencing, microarray and methylation data. Additional selection criteria were 1) for microarray studies, only data from Affymetrix, Agilent, ABI and Illumina platforms were allowed, and 2) we only considered studies accepted for publication, with raw data from samples from cell cultures or nerve tumors.

Microarray data preprocessing

Each microarray was preprocessed and evaluated independently using the R/BioConductor software environment [63]. R packages used in pre-processing steps are detailed in Materials and methods A in S1 Appendix. After grouping dNF and pNF samples as NF, and sporadic and NF1-associated as MPNST, the resulting expression set was filtered to discard features with FDR values >0.05 obtained by ANOVA. The expression set derived was used to assess appropriate grouping of samples from the independent studies by principal component analysis (PCA), as shown in S5 and S6 Figs. The final derived expression set was used to analyze gene differential expression.

Probe name translation to human ENSEMBL and HUGO gene IDs and mapping in human chromosome arms

As integration of scores from different platforms required common identifiers, we translated probe names from each study to ENSEMBL human gene IDs according to Ensembl Archive release 82 (September 2015). To obtain a unique score for each gene, we selected the score of the probe with the greatest variation when the identifier was present more than once. This variation was calculated as the variance of normalized expression values of all samples from the two phenotypes compared. Mouse gene IDs were also translated to human homologous gene IDs. Human ENSEMBL gene IDs were associated with HUGO gene nomenclature IDs and mapped to human chromosomes and bands. Details of ENSEMBL gene ID association to human chromosome arms are shown in Materials and methods B in S1 Appendix.

Gene scores across studies and final individual gene score in a two-phenotype comparison

Gene scores were also computed in the R/Bioconductor environment [63]; Materials and methods C in S1 Appendix details the formula and computation method used. In an example illustrating the factors involved in gene score computation, S13 Table shows values for the differentially expressed genes obtained by comparing MPNST vs. NF from the GSE66743 study. The final score across studies of a single species was calculated as the sum of standardized individual s_{ij} scores. Final human and mouse scores were weighted for human results. Genes exclusive to mouse were ignored and mouse data that differed from human data in the expression ratio sign were discarded. The human score was calculated by adding scores from human studies. Inclusion of murine data was limited to genes for which the sign for the final human and mouse scores were identical, and final human scores were not null.

logFC and logFC_m of genes across studies

Gene logFC were calculated across the studies as the median value of $logFC_{ij}$; $logFC_{ij}$ values for each gene (i) in each individual study (j) were obtained in the phenotype comparison. Gene logFC_m were computed as the median value of $logFC\_m_{ij}$; we have defined $logFC\_m_{ij}$ as the $\log_2$ of the median value of all expression ratios between phenotypes. To avoid inconsistency, the same restriction imposed on the final score was used for logFC and logFC_m values. The signs for human median values of computed logFC/logFC_m were compared with those of mouse. If human and mouse signs were equal, the mouse logFC/ logFC_m was used to compute the final logFC/logFC_m. If the signs were different, mouse values were ignored and human values were used as the final logFC/logFC_m.

Final filters to obtain the gene signature

To avoid inconsistency, the resulting list of genes with non-null score was screened to remove genes showing score signs opposite to those of logFC values (S4 Fig, third filter). To select genes with the highest score and median size effect values, a final filter was applied to the previous gene list (S4 Fig, fourth filter). This filter screened 10% of genes with the highest scores (genes with positive score) and 10% of genes with the lowest score (genes with negative score). In both cases, the absolute logFC value should be >0.99.

Computation of bias in score values among studies: Bhattacharyya distance (BD) ratio

To determine the relative influence of individual experiments on final score values, we computed for each gene the BD between the observed distribution of score frequency and the expected discrete distribution if each study contributed equally to the final score. The lowest BD values thus correlated with similar contributions of individual studies to final score values, while the highest values indicated disparity in the contribution of individual studies. To compare BD without considering the number of studies available for each gene, we computed the BD ratio by dividing the BD by the maximum possible BD value considering the number of studies for each gene. The lowest and highest BD ratio values indicated the similar or dissimilar contribution of each study to each gene score, respectively. A BD ratio equal to zero was applied exceptionally to those genes for which only one study was available, to distinguish these genes from those for which there was more than one study, although only one showed a non-null score. Computation details are given in Materials and methods D in S1 Appendix.

Statistical analyses

To determine the statistical significance of the different gene distribution frequencies in human chromosome regions, we calculated the accumulated probability of a binomial distribution $P(X \geq x)$ using the pbinom function from the R stats package; **** $P \leq 0.0001$, *** $0.0001 < P \leq 0.001$, ** $0.001 < P \leq 0.01$, * $0.01 < P \leq 0.01$. The expected frequency of gene distribution in human chromosome arms/bands was calculated considering the association of ENSEMBL gene IDs to human chromosome arms/bands. To assess correlation between values of two vectors, Pearson's product-moment correlation was calculated using the cor.test function from the R stats package.

Gene and pathway functional enrichment

The over-representation test for GO terms from GO Ontology database release 2016_09_24 (The Gene Ontology Consortium, 2015) was performed through PANTHER v.10 (release 20160715; [64]) using ENSEMBL IDs as input, Homo sapiens (all genes in the database) as reference list, GO biological process, GO molecular function and GO cellular component complete as Annotation Data Set, and the Bonferroni correction for multiple testing ($P < 0.05$). The EnrichNet tool [65] was used to inspect KEGG, BioCarta, Wiki Pathways and Reactome databases to search for pathways in which a differentially expressed group of genes is enriched compared to the whole human genome; ENSEMBL IDs were used as input. Only pathways with a significant XD-score or Fischer q-value <0.1 were considered. The search for opposite and similar gene signatures in CMap-DrugMatrix and GEO databases in NFFinder was performed as described [20].

Gene clustering and visualization

Genes were grouped using SOTA [66] implemented in the analysis and visualization tool MultiexperimentViewer MeV v. 4.9 [67]. Parameters were fit by default, including distance (Pearson correlation), and Max. Cycles (100). Sample tree was carried out using hierarchical clustering to optimize sample leaf order. Pearson correlation was selected as the distance metric, and average linkage clustering as the linkage method.

DNA methylation analysis

Methylation data were downloaded from GEO, and inspected and analyzed using the MEDIPS R package [68]. Linux command lines, R code based on Lienhard *et al.* [68], and details of the analysis are shown in Materials and methods E in S1 Appendix.

Supporting information

S1 Table. Gene score (s_{ij}) and size effect values ($logFC_{ij}$ and $logFC\_m_{ij}$) for the five studies included in MPNST vs. NF meta-analysis.
(XLS)

S2 Table. Gene signatures in the context of neurofibromatosis type 1.
(ZIP)

S3 Table. Results retrieved from NFFinder.
(XLS)

S4 Table. Functional characterization of the MPNST vs. NF gene signature.
(XLS)

S5 Table. Functional description of the 20 genes with the highest and lowest scores in the MPNST vs. NF signature.
(PDF)

S6 Table. Characterization of the 20 genes with the highest and lowest scores in the MPNST vs. NF signature.
(PDF)

S7 Table. Functional description of genes included in the over-represented chromosome regions.
(XLS)

S8 Table. List of 2209 genes included in signatures from cell cultures (NF vs. control, MPNST vs. control, MPNST vs. NF) and tumor tissue (MPNST vs. NF, MPNST vs. control, NF vs. control).
(XLS)

S9 Table. Functional description of genes included in SOTA clusters (biological process GO term enrichment).
(XLS)

S10 Table. MPNST vs. NF unfiltered list of genes potentially down-regulated by promoter hypermethylation.
(XLS)

S11 Table. Up-regulated genes from the MPNST vs. NF signature previously related with NF and/or MPNST.
(PDF)

S12 Table. Down-regulated genes from the MPNST vs. NF signature previously related with NF and/or MPNST.
(PDF)

S13 Table. Head and tail of GSE66743 data extracted from differential expression analysis between MPNST and NF phenotypes with other factors required to compute gene scores.
(PDF)

S14 Table. Functional characterization of additional gene signatures in the context of NF1.
(XLS)

S1 Fig. Comparison of gene score profiles of the integrative MPNST vs. NF gene signature and the five individual studies. a. Plot showing the percentage of score values for up- (score> 0) and down- (score<0) regulated genes of the MPNST vs. NF unfiltered list. This list contains the score values of the 7064 unique ENSEMBL human genes sorted in x-axis from the highest to the lowest score value (Table A in S2 Table). Vertical red dot lines discriminate the first 336 up- and the last 243 downregulated genes with the highest absolute score values, included in the MPNST vs. NF gene signature. **b.** Plots with x and y axes equal to plot A showing the percentage of score values from non-null score genes represented in each of the five studies integrated in the MPNST vs. NF gene signature (E-MEXP-353, E-TABM-69, GSE41747 (human), GSE66743 and GSE41747 (mouse)). Unlike these five plots, that contain differentially expressed genes derived from the MPNST vs. NF comparison, the last plot (Control), as negative control, includes non-null score genes obtained in the NF vs. Control cell culture comparison from the GSE14038 accession.
(PDF)

S2 Fig. Chromosome distribution of the MPNST vs. NF gene signature. The distribution was calculated from 4059 genes with positive score **(a)**, and from 3005 genes with negative score **(b)**. Bar diagrams compare the observed distribution of MPNST vs. NF gene percentage in the human chromosome arms (blue bars) with the expected distribution according to the human ENSEMBL database (red bars). Statistical significance of the gene signature over-represented chromosome arms is above the bars. Over-represented human chromosome bands in the MPNST vs. NF gene signature are shown below each chart. Their statistical significance is shown at the top right side of band names. (****) $P(X \geq x) < 0.0001$, (***) $0.0001 < P(X \geq x) < 0.001$, (**) $0.001 < P(X \geq x) < 0.01$, (*) $0.01 < P(X \geq x) < 0.05$.
(PDF)

S3 Fig. Grey scale diagram that shows the percentage of genes from each gene signature included in the clustering process. Numbers over the gray scale diagram indicate the number of genes included in each cluster. The interval of color scale values is shown below the diagram. The right side of diagram details the number of genes in each cluster, the number of biological process GO terms over-represented in each cluster, and the summary of that GO term enrichment as functional characterization of clusters. A complete list of terms is shown in S9 Table.
(PDF)

S4 Fig. Workflow of selection and pre-processing of microarrays selected to obtain the gene signature MPNST vs. NF, translation to ENSEMBL gene names, and computation of gene scores and median of logFC/logFC_m. Main outputs of some individual steps appear on the right. i: Each individual gene. j: Each individual study.
(PDF)

S5 Fig. Principal components (PCA) plots obtained for each study included in MPNST vs. NF meta-analysis. The final number of probes (Table A in Results B in S1 Appendix) considered in the computation of PCA plots is shown. The legend of colored circles on the left shows sample phenotypes compared.
(PDF)

S6 Fig. Principal components (PCA) plots obtained for other studies included in additional comparisons other than tumor tissue from MPNST vs. NF. The final number of probes (Table A in Results B in S1 Appendix) considered in the computation of PCA plots is shown. Colored circles on the left show the sample phenotypes compared in the analyses. PCA plot from GSE14038 (cell cultures) includes the samples from the previously described comparison MPNST vs. NF from GSE41747 (human tumor tissue) depicted in S5 Fig.
(PDF)

S1 Appendix.
Results A: Relationships between score values and additional attributes in the MPNST vs. NF gene signature.
Results B: Contribution of individual studies to the MPNST vs. NF gene signature.
Results C: Genes included in main functional pathways associated to the MPNST vs. NF signature.
Results D: Characterization of additional NF1-related gene signatures.
Results E: Homogeneity of gene profiles in each SOTA cluster.
Discussion A: Comparison between over-represented chromosome regions in the MPNST vs. NF gene signature and previously described MPNST aberrant chromosome modifications.
Discussion B: Expression profile differences between cultured cells and nerve tumors observed

in *SOX9*, *SUZ12*, *EGFR*, *SPP1* and *BMP2* genes.
Discussion C: Panel of genes potentially silenced by hypermethylation of their CpG-island promoter region.
Discussion D: HDAC inhibitors counteract repression of CBX7 and over-expression of EZH2.
Discussion E: Supplementation with acetate precursors as coadjuvant chemotherapy.
Materials and methods A: Microarray data pre-processing.
Materials and methods B: Translation from probe names to human ENSEMBL gene IDs, HUGO IDs and mapping in human chromosome arms.
Materials and methods C: Scores of genes across studies and final score for each gene in a comparison between two phenotypes.
Materials and methods D: Computation of bias in score values among studies: Bhattacharya distance (BD) ratio.
Materials and methods E: DNA methylation analysis.
(PDF)

S2 Appendix. Meta-analysis PRISMA checklist.
(PDF)

Acknowledgments

We thank the support of the Children's Tumor Foundation.

Author Contributions

Conceptualization: AP.

Data curation: MM.

Formal analysis: MM CS.

Funding acquisition: AP.

Investigation: MM.

Methodology: MM CS.

Project administration: MM.

Software: MM.

Supervision: CS AP JC.

Visualization: MM.

Writing – original draft: MM.

Writing – review & editing: CS AP JC.

References

1. Sun D, Haddad R, Kraniak JM, Horne SD, Tainsky MA. RAS/MEK-independent gene expression reveals BMP2-related malignant phenotypes in the Nf1-deficient MPNST. Mol Cancer Res. 2013; 11: 616–27. https://doi.org/10.1158/1541-7786.MCR-12-0593 PMID: 23423222
2. Whelan JT, Hollis SE, Cha DS, Asch AS, Lee MH. Post-transcriptional regulation of the Ras-ERK/MAPK signaling pathway. J Cell Physiol. 2012; 227: 1235–41. https://doi.org/10.1002/jcp.22899 PMID: 21688267

3. Miller SJ, Jessen WJ, Mehta J, Hardiman A, Sites E, Kaiser S, et al. Integrative genomic analyses of neurofibromatosis tumours identify SOX9 as a biomarker and survival gene. EMBO Mol Med. 2009; 1: 236–248. https://doi.org/10.1002/emmm.200900027 PMID: 20049725
4. Evans DG, Baser ME, McGaughran J, Sharif S, Howard E, Moran A. Malignant peripheral nerve sheath tumours in neurofibromatosis 1. J Med Genet. 2002; 39: 311–4. Available: http://www.ncbi.nlm.nih.gov/pubmed/12011145 https://doi.org/10.1136/jmg.39.5.311 PMID: 12011145
5. Watson MA, Perry A, Tihan T, Prayson RA, Guha A, Bridge J, et al. Gene expression profiling reveals unique molecular subtypes of Neurofibromatosis Type I-associated and sporadic malignant peripheral nerve sheath tumors. Brain Pathol. 2004; 14: 297–303. Available: http://www.ncbi.nlm.nih.gov/pubmed/15446585 PMID: 15446585
6. Miller SJ, Rangwala F, Williams J, Ackerman P, Kong S, Jegga AG, et al. Large-scale molecular comparison of human Schwann cells to malignant peripheral nerve sheath tumor cell lines and tissues. Cancer Res. 2006; 66: 2584–2591. https://doi.org/10.1158/0008-5472.CAN-05-3330 PMID: 16510576
7. Castellsagué J, Gel B, Fernández-Rodríguez J, Llatjós R, Blanco I, Benavente Y, et al. Comprehensive establishment and characterization of orthoxenograft mouse models of malignant peripheral nerve sheath tumors for personalized medicine. EMBO Mol Med. 2015; 7: 608–627. https://doi.org/10.15252/emmm.201404430 PMID: 25810463
8. Thomas LE, Winston J, Rad E, Mort M, Dodd KM, Tee AR, et al. Evaluation of copy number variation and gene expression in neurofibromatosis type-1-associated malignant peripheral nerve sheath tumours. Hum Genomics. 2015; 9:3. https://doi.org/10.1186/s40246-015-0025-3 PMID: 25884485
9. Fang Y, Elahi A, Denley RC, Rao PH, Brennan MF, Jhanwar SC. Molecular characterization of permanent cell lines from primary, metastatic and recurrent malignant peripheral nerve sheath tumors (MPNST) with underlying neurofibromatosis-1. Anticancer Res. 2009; 29: 1255–62. Available: http://www.ncbi.nlm.nih.gov/pubmed/19414372 PMID: 19414372
10. Kolberg M, Høland M, Lind GE, Ågesen TH, Skotheim RI, Hall KS, et al. Protein expression of BIRC5, TK1, and TOP2A in malignant peripheral nerve sheath tumours—A prognostic test after surgical resection. Mol Oncol. 2015; 9: 1129–1139. https://doi.org/10.1016/j.molonc.2015.02.005 PMID: 25769404
11. Berner JM, Sørlie T, Mertens F, Henriksen J, Saeter G, Mandahl N, et al. Chromosome band 9p21 is frequently altered in malignant peripheral nerve sheath tumors: studies of CDKN2A and other genes of the pRB pathway. Genes Chromosomes Cancer. 1999; 26: 151–60. Available: http://www.ncbi.nlm.nih.gov/pubmed/10469453 PMID: 10469453
12. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. Nature. 2014; 514: 247–251. https://doi.org/10.1038/nature13561 PMID: 25119042
13. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, Down TA, et al. Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res. 2011; 515–524. https://doi.org/10.1101/gr.109678.110 PMID: 21324880
14. Danielsen SA, Lind GE, Kolberg M, Høland M, Bjerkehagen B, Hall KS, et al. Methylated RASSF1A in malignant peripheral nerve sheath tumors identifies neurofibromatosis type 1 patients with inferior prognosis. Neuro Oncol. 2015; 17: 63–69. https://doi.org/10.1093/neuonc/nou140 PMID: 25038505
15. Henderson SR, Guiliano D, Presneau N, McLean S, Frow R, Vujovic S, et al. A molecular map of mesenchymal tumors. Genome Biol. 2005; 6: R76. https://doi.org/10.1186/gb-2005-6-9-r76 PMID: 16168083
16. Lévy P, Ripoche H, Laurendeau I, Lazar V, Ortonne N, Parfait B, et al. Microarray-based identification of Tenascin C and Tenascin XB, genes possibly involved in tumorigenesis associated with neurofibromatosis type 1. Clin Cancer Res. 2007; 13: 398–407. https://doi.org/10.1158/1078-0432.CCR-06-0182 PMID: 17202312
17. Jessen WJ, Miller SJ, Jousma E, Wu J, Rizvi TA, Brundage ME, et al. MEK inhibition exhibits efficacy in human and mouse neurofibromatosis tumors. J Clin Invest. 2013; 123: 340–347. https://doi.org/10.1172/JCI60578 PMID: 23221341
18. Carroll SL. The Challenge of Cancer Genomics in Rare Nervous System Neoplasms: Malignant Peripheral Nerve Sheath Tumors as a Paradigm for Cross-Species Comparative Oncogenomics. Am J Pathol. 2016; 186: 464–77. https://doi.org/10.1016/j.ajpath.2015.10.023 PMID: 26740486
19. Ekins S, Williams AJ, Krasowski MD, Freundlich JS. In silico repositioning of approved drugs for rare and neglected diseases. Drug Discov Today. 2011; 16: 298–310. https://doi.org/10.1016/j.drudis.2011.02.016 PMID: 21376136
20. Setoain J, Franch M, Martínez M, Tabas-Madrid D, Sorzano COS, Bakker A, et al. NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. Nucleic Acids Res. 2015; 43: W193–9. https://doi.org/10.1093/nar/gkv445 PMID: 25940629

21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013; 41: D991–5. https://doi.org/10.1093/nar/gks1193 PMID: 23193258

22. Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, et al. ArrayExpress update—simplifying data submissions. Nucleic Acids Res. 2015; 43: D1113–6. https://doi.org/10.1093/nar/gku1057 PMID: 25361974

23. Rasche A, Al-Hasani H, Herwig R. Meta-analysis approach identifies candidate genes and associated molecular networks for type-2 Diabetes mellitus. BMC Genomics. 2008; 9: 310. https://doi.org/10.1186/1471-2164-9-310 PMID: 18590522

24. Gligorijević V, Pržulj N. Methods for biological data integration: perspectives and challenges. J R Soc Interface. 2015; 12: 20150571-. https://doi.org/10.1098/rsif.2015.0571 PMID: 26490630

25. Zhang M, Wang Y, Jones S, Sausen M, McMahon K, Sharma R, et al. Somatic mutations of SUZ12 in malignant peripheral nerve sheath tumors. Nat Genet. 2014; 46: 1170–2. https://doi.org/10.1038/ng.3116 PMID: 25305755

26. Vilardell M, Rasche A, Thormann A, Maschke-Dutz E, Pérez-Jurado LA, Lehrach H, et al. Meta-analysis of heterogeneous Down Syndrome data reveals consistent genome-wide dosage effects related to neurological processes. BMC Genomics. 2011; 12: 229. https://doi.org/10.1186/1471-2164-12-229 PMID: 21569303

27. Burns TC, Li MD, Mehta S, Awad AJ, Morgan AA. Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models. Eur J Pharmacol. 2015; 759: 101–17. https://doi.org/10.1016/j.ejphar.2015.03.021 PMID: 25814260

28. López G, Torres K, Liu J, Hernández B, Young E, Belousov R, et al. Autophagic survival in resistance to histone deacetylase inhibitors: novel strategies to treat malignant peripheral nerve sheath tumors. Cancer Res. 2011; 71: 185–96. https://doi.org/10.1158/0008-5472.CAN-10-2799 PMID: 21084276

29. Semenova G, Stepanova D, Deyev SM, Chernoff J. Medium throughput biochemical compound screening identifies novel agents for pharmacotherapy of neurofibromatosis type I. Biochimie. 2017; 135: 1–5. https://doi.org/10.1016/j.biochi.2017.01.001 PMID: 28065690

30. Byer SJ, Eckert JM, Brossier NM, Clodfelder-Miller BJ, Turk AN, Carroll AJ, et al. Tamoxifen inhibits malignant peripheral nerve sheath tumor growth in an estrogen receptor-independent manner. Neuro Oncol. 2011; 13: 28–41. https://doi.org/10.1093/neuonc/noq146 PMID: 21075781

31. Fujiwara S, Hoshikawa S, Ueno T, Hirata M, Saito T, Ikeda T, et al. SOX10 transactivates S100B to suppress Schwann cell proliferation and to promote myelination. PLoS One. 2014; 9: e115400. https://doi.org/10.1371/journal.pone.0115400 PMID: 25536222

32. Karamchandani JR, Nielsen TO, van de Rijn M, West RB. Sox10 and S100 in the Diagnosis of Soft-tissue Neoplasms. Appl Immunohistochem Mol Morphol. 2012; 20: 445–450. https://doi.org/10.1097/PAI.0b013e318244ff4b PMID: 22495377

33. Lévy P, Vidaud D, Leroy K, Laurendeau I, Wechsler J, Bolasco G, et al. Molecular profiling of malignant peripheral nerve sheath tumors associated with neurofibromatosis type 1, based on large-scale real-time RT-PCR. Mol Cancer. 2004; 3: 20. https://doi.org/10.1186/1476-4598-3-20 PMID: 15255999

34. Bottillo I, Ahlquist T, Brekke H, Danielsen SA, van den Berg E, Mertens F, et al. Germline and somatic *NF1* mutations in sporadic and NF1-associated malignant peripheral nerve sheath tumours. J Pathol. 2009; 217: 693–701. https://doi.org/10.1002/path.2494 PMID: 19142971

35. Timmerman LA, Grego-Bessa J, Raya A, Bertrán E, Pérez-Pomares JM, Díez J, et al. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. Genes Dev. 2004; 18: 99–115. https://doi.org/10.1101/gad.276304 PMID: 14701881

36. Sommariva S, Tarricone R, Lazzeri M, Ricciardi W, Montorsi F. Prognostic Value of the Cell Cycle Progression Score in Patients with Prostate Cancer: A Systematic Review and Meta-analysis. Eur Urol. 2016; 69: 107–15. https://doi.org/10.1016/j.eururo.2014.11.038 PMID: 25481455

37. Khattar V, Thottassery JV. Cks1: Structure, Emerging Roles and Implications in Multiple Cancers. J Cancer Ther. 2013; 4: 1341–1354. https://doi.org/10.4236/jct.2013.48159 PMID: 24563807

38. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol. 2011; 12: 245–55. https://doi.org/10.1016/S1470-2045(10)70295-3 PMID: 21310658

39. Hui P, Li N, Johnson C, De Wever I, Sciot R, Manfioletti G, et al. HMGA proteins in malignant peripheral nerve sheath tumor and synovial sarcoma: preferential expression of HMGA2 in malignant peripheral nerve sheath tumor. Mod Pathol. 2005; 18: 1519–1526. https://doi.org/10.1038/modpathol.3800464 PMID: 16056249

40. Lakshmaiah KC, Jacob LA, Aparna S, Lokanatha D, Saldanha SC. Epigenetic therapy of cancer with histone deacetylase inhibitors. J Cancer Res Ther. 2014; 10: 469–78. https://doi.org/10.4103/0973-1482.137937 PMID: 25313724

41. Hull EE, Montgomery MR, Leyva KJ. HDAC Inhibitors as Epigenetic Regulators of the Immune System: Impacts on Cancer Therapy and Inflammatory Diseases. BioMed Res Int. 2016; 2016: 1–15.

42. Jia H, Morris CD, Williams RM, Loring JF, Thomas EA. HDAC inhibition imparts beneficial transgenerational effects in Huntington's disease mice via altered DNA and histone methylation. Proc Natl Acad Sci U S A. 2015; 112: E56–64. https://doi.org/10.1073/pnas.1415195112 PMID: 25535382

43. Falkenberg KJ, Johnstone RW. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. Nat Rev Drug Discov. 2014; 13: 673–691. https://doi.org/10.1038/nrd4360 PMID: 25131830

44. Yoon S, Eom GH. HDAC and HDAC Inhibitor: From Cancer to Cardiovascular Diseases. Chonnam Med J. 2016; 52: 1–11. https://doi.org/10.4068/cmj.2016.52.1.1 PMID: 26865995

45. Kadia TM, Yang H, Ferrajoli A, Maddipotti S, Schroeder C, Madden TL, et al. A phase I study of vorinostat in combination with idarubicin in relapsed or refractory leukaemia. Br J Haematol. 2010; 150: 72–82. https://doi.org/10.1111/j.1365-2141.2010.08211.x PMID: 20456355

46. Feng W, Cai D, Zhang B, Lou G, Zou X. Combination of HDAC inhibitor TSA and silibinin induces cell cycle arrest and apoptosis by targeting survivin and cyclinB1/Cdk1 in pancreatic cancer cells. Biomed Pharmacother. 2015; 74: 257–264. https://doi.org/10.1016/j.biopha.2015.08.017 PMID: 26349994

47. Cornago M, Garcia-Alberich C, Blasco-Angulo N, Vall-llaura N, Nager M, Herreros J, et al. Histone deacetylase inhibitors promote glioma cell death by G2 checkpoint abrogation leading to mitotic catastrophe. Cell Death Dis. 2014; 5: e1435. https://doi.org/10.1038/cddis.2014.412 PMID: 25275596

48. Yan-Fang T, Zhi-Heng L, Li-Xiao X, Fang F, Jun L, Gang L, et al. Molecular Mechanism of the Cell Death Induced by the Histone Deacetylase Pan Inhibitor LBH589 (Panobinostat) in Wilms Tumor Cells. PLoS One. 2015; 10: e0126566. https://doi.org/10.1371/journal.pone.0126566 PMID: 26176219

49. Sarkar S, Abujamra AL, Loew JE, Forman LW, Perrine SP, Faller DV. Histone deacetylase inhibitors reverse CpG methylation by regulating DNMT1 through ERK signaling. Anticancer Res. 2011; 31: 2723–32. Available: http://www.ncbi.nlm.nih.gov/pubmed/21868513 PMID: 21868513

50. Murko C, Lagger S, Steiner M, Seiser C, Schoefer C, Pusch O. Histone deacetylase inhibitor Trichostatin A induces neural tube defects and promotes neural crest specification in the chicken neural tube. Differentiation. 2013; 85: 55–66. https://doi.org/10.1016/j.diff.2012.12.001 PMID: 23328540

51. Li GH, Arora PD, Chen Y, McCulloch CA, Liu P. Multifunctional roles of gelsolin in health and diseases. Med Res Rev. 2012; 32: 999–1025. https://doi.org/10.1002/med.20231 PMID: 22886630

52. Pallante P, Sepe R, Puca F, Fusco A. High mobility group A proteins as tumor markers. Front Med. 2015; 2: 15.

53. Mahajan A, Liu Z, Gellert L, Zou X, Yang G, Lee P, et al. HMGA2: A biomarker significantly overexpressed in high-grade ovarian serous carcinoma. Mod Pathol. 2010; 23: 673–81. https://doi.org/10.1038/modpathol.2010.49 PMID: 20228781

54. Wu J, Lai M, Shao C, Wang J, Wei JJ. STC2 overexpression mediated by HMGA2 is a biomarker for aggressiveness of high-grade serous ovarian cancer. Oncol Rep. 2015; 34: 1494–502. https://doi.org/10.3892/or.2015.4120 PMID: 26165228

55. Ruscetti M, Dadashian EL, Guo W, Quach B, Mulholland DJ, Park JW, et al. HDAC inhibition impedes epithelial-mesenchymal plasticity and suppresses metastatic, castration-resistant prostate cancer. Oncogene. 2016; 35: 3781–95. https://doi.org/10.1038/onc.2015.444 PMID: 26640144

56. Deng R, Hao J, Han W, Ni Y, Huang X, Hu Q. Gelsolin regulates proliferation, apoptosis, migration and invasion in human oral carcinoma cells. Oncol Lett. 2015; 9: 2129–2134. https://doi.org/10.3892/ol.2015.3002 PMID: 26137026

57. Ma X, Sun W, Shen J, Hua Y, Yin F, Sun M, et al. Gelsolin promotes cell growth and invasion through the upregulation of p-AKT and p-P38 pathway in osteosarcoma. Tumour Biol. 2016; 37: 7165–7174. https://doi.org/10.1007/s13277-015-4565-x PMID: 26662962

58. Ni XG, Zhou L, Wang GQ, Liu SM, Bai XF, Liu F, et al. The ubiquitin-proteasome pathway mediates gelsolin protein downregulation in pancreatic cancer. Mol Med. 2008; 14: 582–9. https://doi.org/10.2119/2008-00020.Ni PMID: 18584046

59. Wang X, Xu J, Wang H, Wu L, Yuan W, Du J, et al. Trichostatin A, a histone deacetylase inhibitor, reverses epithelial-mesenchymal transition in colorectal cancer SW480 and prostate cancer PC3 cells. Biochem Biophys Res Commun. 2015; 456: 320–6. https://doi.org/10.1016/j.bbrc.2014.11.079 PMID: 25434997

60. Brosius SN, Turk AN, Byer SJ, Longo JF, Kappes JC, Roth KA, et al. Combinatorial therapy with tamoxifen and trifluoperazine effectively inhibits malignant peripheral nerve sheath tumor growth by targeting

complementary signaling cascades. J Neuropathol Exp Neurol. 2014; 73: 1078–90. https://doi.org/10.1097/NEN.0000000000000126 PMID: 25289889

61. Kohli L, Kaza N, Coric T, Byer SJ, Brossier NM, Klocke BJ, et al. 4-Hydroxytamoxifen Induces Autophagic Death through K-Ras Degradation. Cancer Res. 2013; 73: 4395–4405. https://doi.org/10.1158/0008-5472.CAN-12-3765 PMID: 23722551

62. Di Fazio P, Waldegger P, Jabari S, Lingelbach S, Montalbano R, Ocker M, et al. Autophagy-related cell death by pan-histone deacetylase inhibition in liver cancer. Oncotarget. 2016; 7: 28998–9010. https://doi.org/10.18632/oncotarget.8585 PMID: 27058414

63. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5: R80. https://doi.org/10.1186/gb-2004-5-10-r80 PMID: 15461798

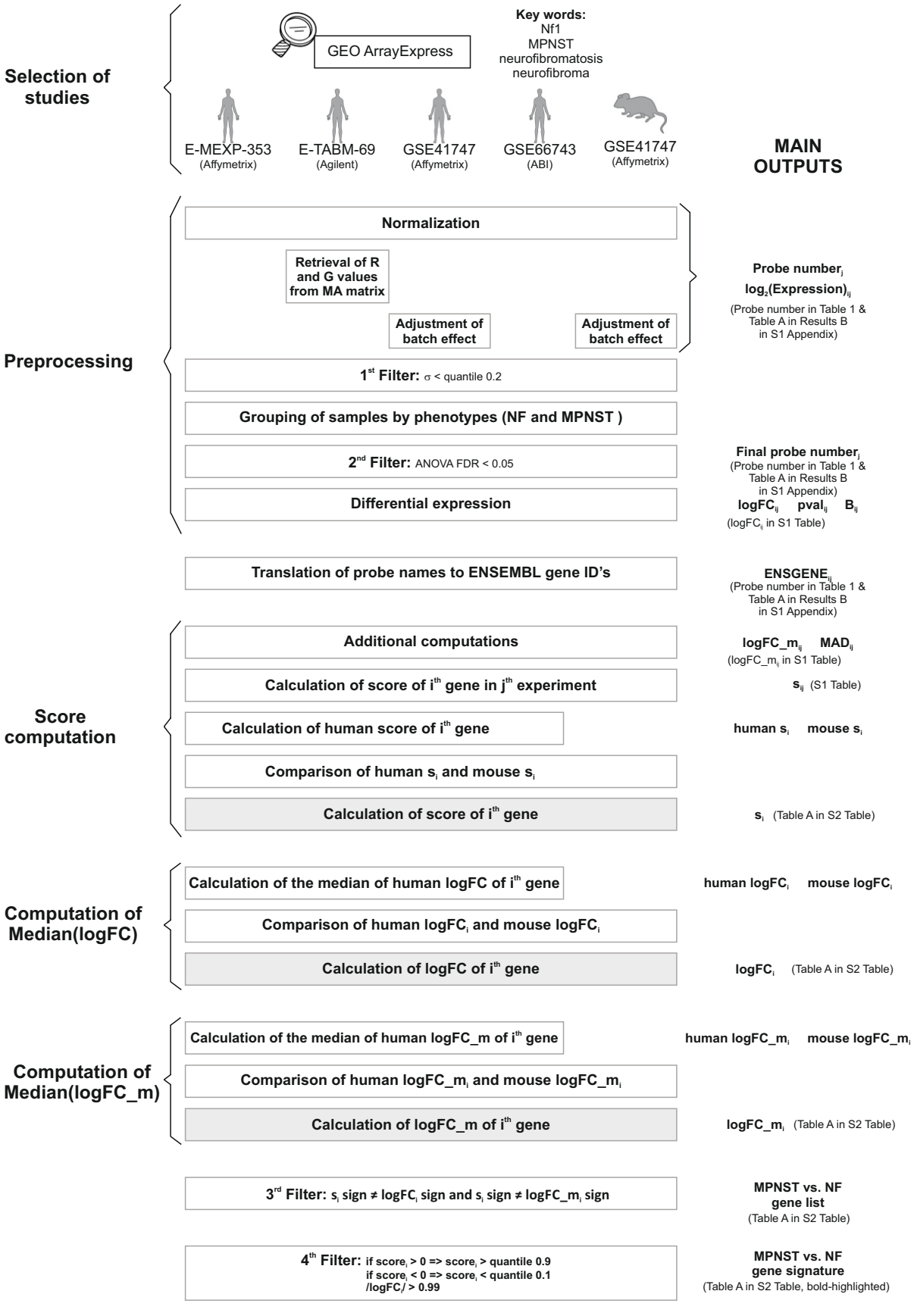
64. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2016; 44: D336–42. https://doi.org/10.1093/nar/gkv1194 PMID: 26578592

65. Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. EnrichNet: network-based gene set enrichment analysis. Bioinformatics. 2012; 28: i451–i457. https://doi.org/10.1093/bioinformatics/bts389 PMID: 22962466

66. Herrero J, Valencia A, Dopazo J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics. 2001; 17: 126–36. Available: http://www.ncbi.nlm.nih.gov/pubmed/11238068 PMID: 11238068

67. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. Biotechniques. 2003; 34: 374–8. Available: http://www.ncbi.nlm.nih.gov/pubmed/12613259 PMID: 12613259

68. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: Genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014; 30: 284–286. https://doi.org/10.1093/bioinformatics/btt650 PMID: 24227674


Selection of studies
GEO ArrayExpress
Key words:
Nf1
MPNST
neurofibromatosis
neurofibroma
E-MEXP-353
(Affymetrix)
E-TABM-69
(Agilent)
GSE41747
(Affymetrix)
GSE66743
(ABI)
GSE41747
(Affymetrix)
MAIN OUTPUTS
Preprocessing
Normalization
Retrieval of R and G values from MA matrix
Adjustment of batch effect
Adjustment of batch effect
Probe number$_i$
log$_2$(Expression)$_{ij}$
(Probe number in Table 1 & Table A in Results B in S1 Appendix)
1st Filter: σ < quantile 0.2
Grouping of samples by phenotypes (NF and MPNST)
2nd Filter: ANOVA FDR < 0.05
Final probe number$_j$
(Probe number in Table 1 & Table A in Results B in S1 Appendix)
Differential expression
logFC$_{ij}$ pval$_{ij}$ B$_{ij}$
(logFC$_{ij}$ in S1 Table)
Translation of probe names to ENSEMBL gene ID's
ENSGENE$_{ij}$
(Probe number in Table 1 & Table A in Results B in S1 Appendix)
Score computation
Additional computations
logFC_m$_{ij}$ MAD$_{ij}$
(logFC_m$_{ij}$ in S1 Table)
Calculation of score of i th gene in j th experiment
s$_{ij}$ (S1 Table)
Calculation of human score of i th gene
human s$_i$ mouse s$_i$
Comparison of human s$_i$ and mouse s$_i$
Calculation of score of i th gene
s$_i$ (Table A in S2 Table)
Computation of Median(logFC)
Calculation of the median of human logFC of i th gene
human logFC$_i$ mouse logFC$_i$
Comparison of human logFC$_i$ and mouse logFC$_i$
Calculation of logFC of i th gene
logFC$_i$ (Table A in S2 Table)
Computation of Median(logFC_m)
Calculation of the median of human logFC_m of i th gene
human logFC_m$_i$ mouse logFC_m$_i$
Comparison of human logFC_m$_i$ and mouse logFC_m$_i$
Calculation of logFC_m of i th gene
logFC_m$_i$ (Table A in S2 Table)
3rd Filter: s$_i$ sign ≠ logFC$_i$ sign and s$_i$ sign ≠ logFC_m$_i$ sign
MPNST vs. NF gene list
(Table A in S2 Table)
4th Filter: if score$_i$ > 0 => score$_i$ > quantile 0.9
if score$_i$ < 0 => score$_i$ < quantile 0.1
/logFC$_i$/ > 0.99
MPNST vs. NF gene signature
(Table A in S2 Table, bold-highlighted)

S1 Appendix

Index

Results A: Relationships between score values and additional attributes in the MPNST vs. NF gene signature

Gene logFC values across the studies from the MPNST vs. NF signature (bold-highlighted in Table A in S2 Table) were plotted regarding gene score values (Fig A (a)). A positive correlation between gene logFC and score values was inferred (Pearson's product-moment correlation 0.94; pval 2.2e-16; 95% confidence interval from 0.93 to 0.95). Red dots identify 114 genes for which mouse scores were not included in the final computation of gene scores, due to the absence of those genes in the mouse study, or because mouse and human studies showed different behavior. Most of genes ignoring mouse data located in the interval [-1, 1] of score values. Although the majority of these genes accumulated around the logFC interval [-2, 2], the highest logFC absolute values were similar in genes considering or not mouse data. Due to a high correlation between logFC and logFC_m values (Pearson's product-moment correlation 0.9962; pval 2.2e-16; 95% confidence interval from 0.9956 to 0.9969; Fig A (b)), a similar plot could be inferred when logFC_m and score values were compared.

To assess the homogenous contribution of each study to the final gene scores, we computed the Bhattacharyya distance ratio (BD-ratio). The lower was the BD-ratio, the more homogeneous was the contribution of the studies. 530 genes, 91.54% of the total gene signature, showed BD-ratios between 0 and 100. These genes, represented by more than one study, showed higher absolute scores compared with genes with BD-ratios 0 or 100, as the distribution of genes indicates considering both BD-ratio and score values (Fig A (c)). Null BD-ratios were assigned to 5 genes (less than 1%) that were in only one study, to distinguish them from 44 genes (7.60%), present in more than one study, with all score values null except one, which results in a BD-ratio of 100. This maximum value for BD-ratio remarks the unequal contribution of studies to the gene score computation of these 44 genes. A slight negative correlation existed between BD-ratios and absolute score values for genes with BD-ratios in the interval (0, 100) (Pearson's product-moment correlation -0.097; pval < 0.01968; 95% confidence interval from -0.177 to -0.016). Again, red dots denote genes excluding mouse data. Genes with absolute score values higher than 1 and lowest BD-ratios, for which individual experiments similarly contribute to the final score, mostly included mouse data in the computation of gene scores.

The number of genes included in the MPNST vs. NF signature with final score

calculated starting from 5, 4, 3, 2, or 1 study, both for up- and downregulated genes, is shown in Fig A (d). As the chart indicates, the highest amount of genes depended on 5 studies (56.31%), 4 from human and one from mouse. 26.94%, 10.88%, 5% and 0.86% of genes depended on 4, 3, 2 and 1 study, respectively. As expected from the addition of five individual scores, the highest absolute average scores were obtained for genes represented in a higher number of studies.

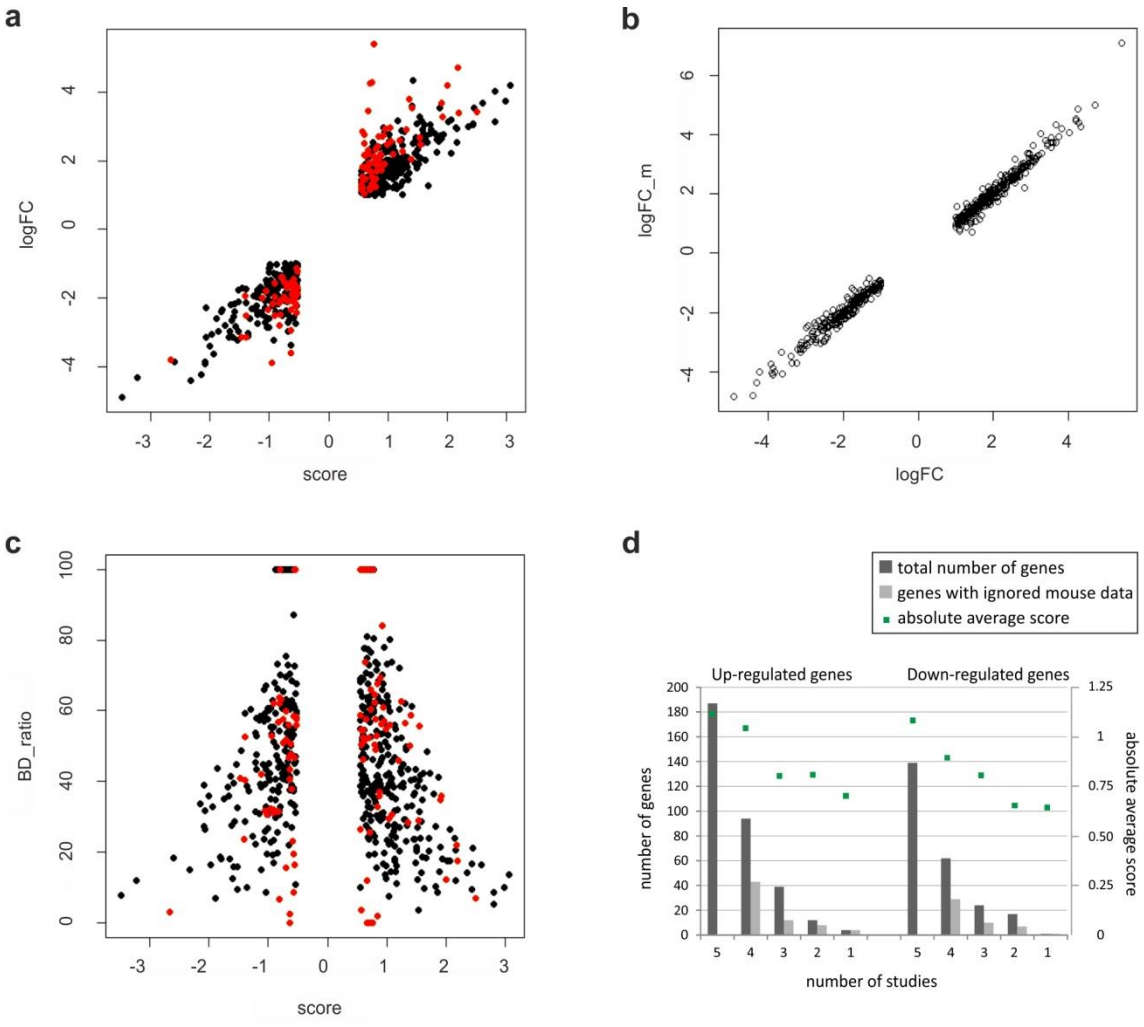


Fig A: Relationships between score values and additional attributes from MPNST vs. NF gene signature. a. The dot plot represents the 579 genes from the MPNST vs. NF gene signature. The plot compares for each gene the logFC with the score value. Red dots represent 114 genes in which mouse data were ignored in the computation of final score and only human data were taken into account. **b.** Same plot from a to compare values of logFC and logFC_m. **c.** Dot plot that compares BD_ratios with score values from genes included in MPNST vs. NF signature. The meaning of red dots was indicated in a. **d.** Chart comparing the number of genes from up- and down-regulated genes included in the MPNST vs. NF signature, and the absolute average score, with the number of studies covering those genes.

Results B: Contribution of individual studies to the MPNST vs. NF gene signature

Although the five individual studies show similar patterns of gene score distribution (S1 Fig), they contribute unequally to the MPNST vs. NF gene signature. Both human and mouse studies from the GSE41747 accession contributed with the highest number of genes to the gene signature, and studies E-TABM-69 and GSE66743 with the lowest (S1 Fig and Table A). However, the amount of genes from studies E-TABM-69 and GSE66743 present in the MPNST vs. NF gene signature considerably increased in comparison with the previous unfiltered gene list. Only 7.86 and 6.66% of genes with non-null score from individual studies E-TABM-69 and GSE66743, respectively, were in the unfiltered gene list, whereas these ratios arose to 34.71% and 39.72% in the MPNST vs. NF gene signature. The addition of the gene number from all studies that contribute to the unfiltered gene list covers 1.32 times the length of the unfiltered gene list (first column score ≠ 0 divided by 7064). In contrast, the total amount of genes from all studies included in the MPNST vs. NF gene signature covered 2.82 times the length of the gene signature (second column score ≠ 0 divided by 579). This indicates that genes included in the final gene signature relied on score values computed from a higher amount of studies than genes from the unfiltered list. This result agreed with the above described BD-ratios in which 91.53% of genes from the gene signature relied on score values computed from more than one study.

Table A: Microarray studies selected from GEO and ArrayExpress databases included in MPNST vs. NF meta-analysis

					Unfiltered MPNST vs. NF gene list (7064 genes)				MPNST vs. NF gene signature (579 genes)			
Organism	**Accession**	**Probe number**	**Final probe number**[1]	**adj.pval**[2] **<0.05** (ENSEMBL gene number)	**score ≠ 0** (ENSEMBL gene number)	**max** (adj.pval[2])	**max(logFC[2])** [score < 0]	**min(logFC[2])** [score > 0]	**score ≠ 0** (ENSEMBL gene number)	**max** (adj.pval[2])	**max(logFC[2])** [score < 0]	**min(logFC[2])** [score > 0]
Human	E-MEXP-353[3]	22283	13,487	2,670	850	0.00351	-0.5328	0.5878	293	0.00351	-0.7374	0.8759
	E-TABM-69[3]	19061	1,478	1,352	555	0.00054	-0.4648	0.4649	201	0.00052	-0.6420	0.7297
	GSE41747[4,5]	54675	20,776	7,909	3,108	0.00082	-0.3467	0.3194	512	0.00079	-0.9829	0.6547
	GSE66743[4]	33025	1,340	989	471	0.00071	-0.9010	0.8691	230	0.00071	-1.3152	0.9312
Mouse	GSE41747[4]	45101	33,061	10,982	4,307	0.00036	-0.1959	0.2262	394	0.00025	-0.4115	0.4626

[1]Obtained after the first two screening steps of preprocessing of microarray studies.

[2]adj.pval and logFC values obtained from MPNST vs. NF differential gene expression analysis with limma R package.

[3]Accession number from ArrayExpress database.

[4]Accession number from GEO database.

[5]Data included in GSE14038 accession.

The maximum adj.pval for genes with non-null score, for every study both in the gene signature and in the unfiltered list, was lower than the commonly used 0.05 to select differentially expressed genes, indicating that score selected genes were significantly

up- or downregulated, respectively, in every individual study (Table A). However, as columns max(logFC) and min(logFC) show, gene scores s_{ij} did not discriminate for high absolute logFC values. Due to the last filter step, these logFC values resulted slightly higher in the gene signature compared to the unfiltered gene list.

Results C: Genes included in main functional pathways associated to the MPNST vs. NF signature

Most upregulated genes were implicated in DNA replication and cell cycle, which are the main pathways involved in cell proliferation. Replication pathway concentrated genes involved in initiation of replication (CDC6, CDC45, CDT1, DBF4, GINS1, GINS2, MCM2, MCM4, MCM6), in elongation (RFC4), in sealing of nicks (LIG1), in control of replication (GMNN, PCNA), and in providing precursors for DNA synthesis (RRM2, and TYMS, targets of a cyclin-dependent kinase inhibitor [1], and of the chemotherapeutic agent 5-fluorouracil [2], respectively, in colorectal cancer). AURKA, AURKB, BIRC5 and CDCA8 are mitotic genes, members of the chromosomal passenger (CPC) complex, essential for alignment and segregation of chromosomes; PTTG1 regulates chromosome stability; SPDL1 and centromere proteins CENPA, CENPE, CENPF, CENPH, CENPI, CENPK, CENPN and CENPU were involved in protein recruitment to the centromere and kinetochore, mitotic progression and chromosome segregation; CASC5, KNTC1, MAD2L1, ZWINT, ZWILCH, SKA1, and NDC80 kinetochore complex components NDC80, NUF2, SPC24, SPC25, were required for chromosome segregation and spindle checkpoint activity; BUB1, BUB1B, CDK2, CHEK1, CKS1B, NEK2 appeared as mitotic checkpoint kinases, whereas PLK4 kinase regulates centriole duplication; CCNA2, CCNB1 and CCNB2 are cyclins that bind and activate mitotic kinases to promote cell cycle transitions; CDC20 and UBE2C allow cell cycle progression by destruction of mitotic cyclins; FBXO5 and SKP2 are involved in proteasome degradation of target proteins; chromosome cohesion was due to RAD21, SGOL1/SGO1 and, in meiosis, SGOL2/SGO2; NCAPG, NCAPH, NCAPD2, NCAPG2 and SMC4 are members of the condensin complex; and KIF18A, KIF18B, KIF20A, KIF23, KIF2C, and ECT2 take part in the kinesin complex. The KEGG extracellular matrix (ECM)-receptor pathway contains the potent apoptosis suppressor COMP, the motility receptor HMMR, the adhesion integrin ITGA4, and the osteopontin SPP1, relevant in interactions between the cell and the ECM. The KEGG ECM-receptor pathway is essential for cell maintenance and tissue and organ morphogenesis, involves fibril forming collagen from connective and cartilaginous tissues. We found that some genes in this pathway were shared with the Reactome pathway of

interactions with the neural adhesion immunoglobulin NCAM1.

Downregulated genes included in the immunity pathway related with complement activation incorporated regulatory elements of that pathway (CD55, A2M, PROS1, SERPING1 and CD59) in addition to complement components (C1S, C3, C4A, C4B and CFD). KEGG pathway of cell adhesion molecules (CAMs) groups glycoproteins located in the cell surface: integrins such as ITGA6 and ITGB8; immunoglobulins such as JAM2, CD58 and others related with the nervous system, such as L1CAM, the synaptic and potential tumor supressor CADM1, the brain specific receptor CADM3, the neuronal growth regulator NEGR1; selectins such as SELPLG; NLGN3, a neuroligin; MPZ, the main structural component of the peripheral myelin sheath; and cadherins CDH19 and PCDH20. Down-regulated genes also showed over-representation of circulating particles released into the extracellular space (blood microparticles).

Results D: Characterization of additional NF1-related gene signatures

In addition to the MPNST vs. NF comparison meta-analysis, we carried out other five comparisons: 2 meta-analyses to integrate nerve tumor tissue from human and mouse in comparisons NF vs. control and MPNST vs. control; 1 meta-analysis that combines human cell cultures to compare MPNST vs. control (MPNST cell lines vs. NHSC); 2 differential analyses from human cultured cells in NF vs. control (primary NFSC vs. NHSC) and MPNST vs. NF (MPNST cell lines vs. primary NFSC) comparisons. The number of genes that exhibited non-null scores in all comparisons of NF1-related signatures, both in cell cultures and in tumor tissue, is shown in Table B. The comparison MPNST vs. control showed the highest number of changes in gene expression in both types of tissue. The comparison NF vs. control showed the lowest amount of changes, especially in cultured cells.

Table B: Number of genes included in Neurofibromatosis gene signatures.

Tissue	Organism	Gene list	NF vs. control			MPNST vs. control			MPNST vs. NF		
			Up-regulated (score >0)	Down-regulated (score < 0)	Total	Up-regulated (score > 0)	Down-regulated (score < 0)	Total	Up-regulated (score >0)	Down-regulated (score < 0)	Total
Cell cultures	Human	Unfiltered list[1]	623	533	1156	4486	3900	8386	1591	1342	2933
		Gene signature[2]	63	54	117	449	390	839	160	135	295
Nerve tumors	Human/ Mouse	Unfiltered list[1]	3687	2892	6579	5204	3602	8806	4059	3005	7064
		Gene signature[2]	369	290	659	521	363	884	336	243	579

[1]Complete lists included in S2 Table.

[2]Gene lists bold-highlighted in S2 Table .

In tumor tissues, the MPNST vs. NF signature shared more genes with the gene signature MPNST vs. control than with the NF vs. control. Respective 67 and 36% of up- and downregulated genes from comparison MPNST vs. NF behaved similarly in MPNST vs. control, *i.e.* the score sign was identical in both comparisons. There were no genes showing opposite behavior, *i.e.* upregulated genes in one comparison being downregulated in the other one. Similar percentages of 66 and 48% were obtained comparing up- and downregulated genes from unfiltered lists of MPNST vs. NF and MPNST vs. control. In this case, a small percentage of 4 and 6% of up- and downregulated genes in MPNST vs. NF list showed score of opposite sign in MPNST vs. control gene list. In contrast, only 3 and 5% of up- and downregulated genes from MPNST vs. NF signature (25 and 20% in the unfiltered gene lists), respectively, appeared included in NF vs. control gene signature, indicating a lower similarity between these two signatures. Moreover, 1 and 5% of up- and downregulated genes, respectively, in the MPNST vs. NF signature exhibited opposite behavior in the NF vs. control signature. The amount of genes that showed opposite behavior in respective unfiltered gene lists increased to around 10 and 20%. The two additional tumor tissue gene signatures obtained from comparisons MPNST vs. control and NF vs. control also shared a quite large amount of genes. 23 and 38% of genes from the MPNST vs. control signature were also up- and down- regulated, respectively, in the NF vs. control gene signature. These percentages increased to 48 and 59% when we compared the unfiltered gene lists.

Compared to tumor tissue, the MPNST vs. NF signature shared a similar high number

of genes with the MPNST vs. control comparison in cell cultures. 56 and 77% of up- and down- regulated genes, respectively, from the MPNST vs. NF signature showed the same behavior in the MPNST vs. control comparison. There were no gene scores with opposite sign. Analogously, percentages 75 and 81 were computed for the unfiltered gene lists. Less than 1% of genes from MPNST vs. NF signature had scores with opposite sign in MPNST vs. control comparison. In contrast, and in agreement with tumor tissues, gene signatures from comparisons MPNST vs. NF and NF vs. control showed low similarity in cell cultures. Only 1 and 3% (4 and 13% comparing unfiltered gene lists) of respective up- and downregulated genes from MPNST vs. NF signature behaved in the same way in NF vs. control gene signature. Concerning cell culture comparisons MPNST vs. control and NF vs. control, most part of the small NF vs. control gene signature was included in the MPNST vs. control signature (60 and 59% of up and down- regulated genes, respectively). Higher percentages were observed for unfiltered signatures (68 and 83%, respectively).

To further examine similarities and differences between tumor tissue and cell cultures, we compared independently gene signatures from each comparison (MPNST vs. NF, MPNST vs. control and NF vs. control) in both types of tissue. High differences were observed between tumor tissue and cell cultures. Tumor tissue comparisons MPNST vs. NF, MPNST vs. control and NF vs. control only shared 12, 18 and 3% of genes with their respective cultured cell signatures. Although these percentages arose to 21, 42 and 6% in unfiltered gene lists, the amount of genes showing opposite behavior also increased between 3 and 5 times. Differences associated to the nature of the tissue probably contributed to such big different behavior between tumor tissue and cultured cells.

Due to the differences between cultured cells and tumor tissue in terms of gene expression, we expected analogous differences in the functional characterization of these gene signatures in the three comparisons considered: MPNST vs. NF, MPNST vs. control and NF vs. control. S14 Table summarizes biological process GO term enrichment for up- and down- regulated genes of each comparison.

As anticipated, a lot of similarities between GO terms associated to upregulated genes from MPNST vs. NF and MPNST vs. control were observed in tumor tissue. The MPNST vs. control signature showed a high number of GO terms associated to cell proliferation (52%), as well as morphogenesis and development of several systems according specification patterns, particularly nervous, skeletal, gastric, urogenital,

reproductive systems, hearth, and sense organs. Other GO terms were related to transcriptional activity and processes involved in malignancy such as collagen catabolic process, epithelial cell differentiation and cell migration (Table C in S14 Table). Unlike the MPNST vs. NF downregulated gene signature, no immunity GO terms associated to the MPNST vs. control downregulated genes were observed. This difference was still maintained when we compare the unfiltered gene lists. Instead, in common with the MPNST vs. NF signature, downregulated genes from MPNST vs. control comparison also seemed to control the peripheral nervous system development. Terms related to muscle system organization, regulation of membrane potential and regulation of phosphatidylinositol 3-kinase signaling also occurred (Table D in S14 Table).

The same differences and similarities observed between cultured cells and tumor tissue in the MPNST vs. NF comparison were observed in the MPNST vs. control comparison. GO terms related with cell proliferation from tumor tissue signatures for MPNST vs. NF and MPNST vs. control were nearly absent from signatures for these comparisons in culture cells. This difference was not observed when we evaluated the unfiltered gene lists for these two comparisons, which suggested lower scores for cell proliferation genes in cultured cells than in tumor tissue. In contrast, immune response positively associated to downregulated genes was observed for tumor tissue and cultured cells of MPNST vs. NF, but only for cultured cells of MPNST vs. control and not in the tumor tissue comparison. This difference was maintained when we compared unfiltered gene lists. In turn, many terms involving development, signal transduction, cell communication and migration can also be observed in MPNST vs. control gene signature (Table E in S14 Table). In MPNST vs. control downregulated genes, several GO terms associated to acquired immune response, development of peripheral nervous system, ECM organization and cell migration were over-represented (Table F in S14 Table).

Finally, the biggest difference in GO term representation was seen between comparisons MPNST vs. NF and NF vs. control. In tumor tissue NF vs. control comparison, most GO terms associated to upregulated genes were related with synaptic transmission (Ca^{2+} transport, membrane potential and synaptic exocytosis) (Table G in S14 Table). However, GO terms associated to cultured cells were related to growth and development, including branching morphogenesis of epithelial tubes and regulation of developmental bone morphogenetic (BMP) signaling pathway. Regulation of epithelial and mesenchymal cell proliferation, and ECM organization were also over-represented in GO terms associated to upregulated genes from cultured cells of NF vs.

control comparison (Table I in S14 Table). Considering the unfiltered lists of upregulated genes, several GO terms associated to innate immune response can be observed in NF vs. control comparison for tumor tissue and cultured cells. Instead, axonogenesis was the unique GO term linked to downregulated genes.

Results E: Homogeneity of gene profiles in each SOTA cluster

To check homogeneity of gene patterns in each cluster, Fig B illustrates the particular behavior of each gene in SOTA clusters.

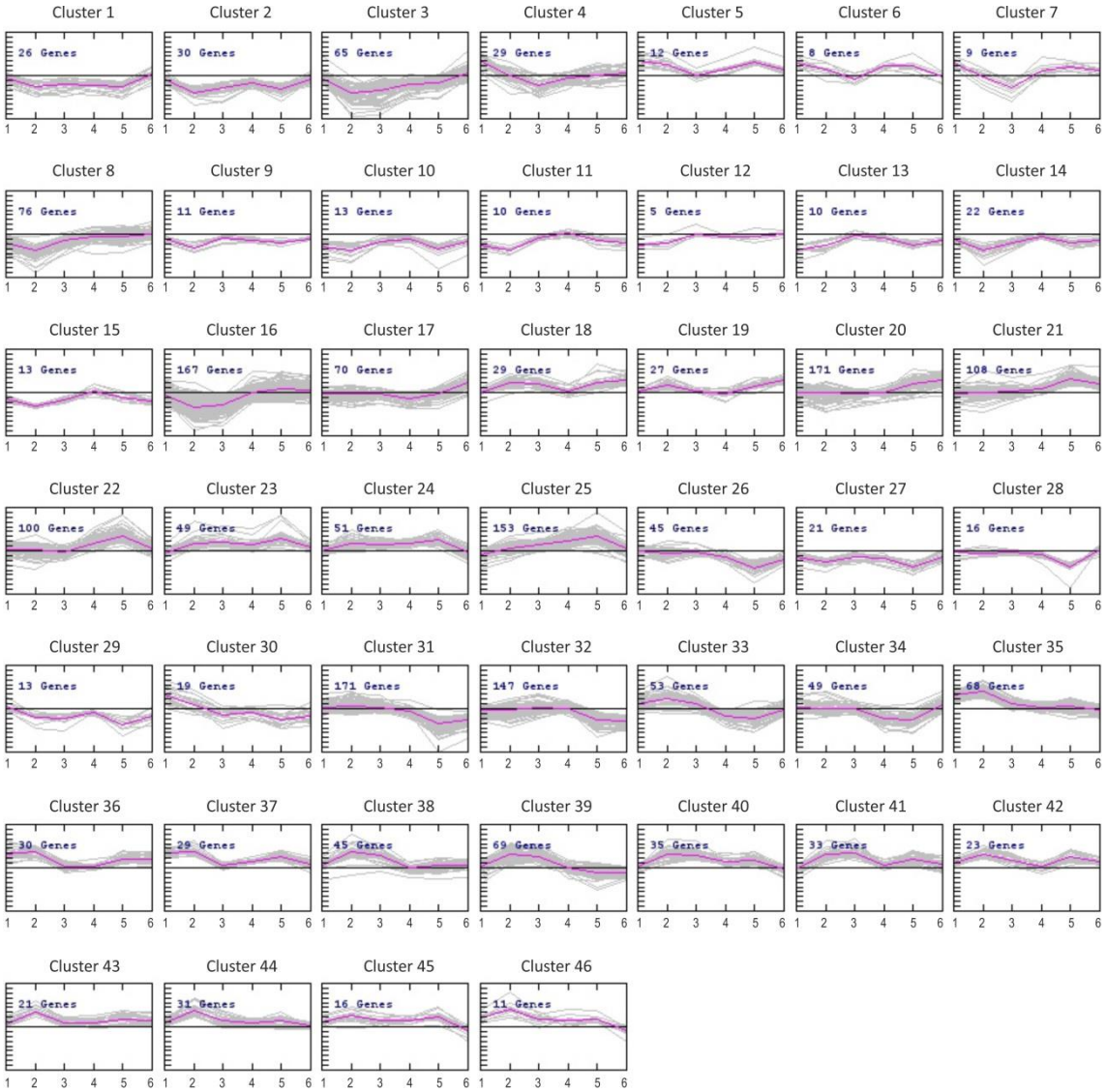


Fig B: Gene profiles in the SOTA clusters. Graphs from logFC_m values of genes grouped through SOTA in 46 clusters. Samples 1, 2 and 3 represent cell culture comparisons NF vs. control, MPNST vs. control and MPNST vs. NF, respectively. Samples 4, 5 and 6 represent nerve tumor comparisons MPNST vs. NF, MPNST vs. control and NF vs. control, respectively. Pink lines represent cluster centroids.

Although in general gene profiles fairly fit centroid vectors, some of the most populated clusters 16, 20-23 and 31 showed gene profiles that slightly differed from consensus centroids. The landscape of cluster profiles shows similarity among certain profiles according with their proximity in the SOTA tree. Additionally, these profiles remark the lack of symmetry in cell cultures regarding nerve tumors, particularly in clusters 7, 8, 16, 17, 20-22, 25, 26, 28, 30-35, 38 and 39. Cluster 33 presented the highest asymmetry between cell cultures and nerve tumor comparisons because most of genes were upregulated in cell cultures and downregulated in nerve tumors.

Discussion A: Comparison between over-represented chromosome regions in the MPNST vs. NF gene signature and previously described MPNST aberrant chromosome modifications

In order to correlate gene expression with gain or loss of chromosomic regions, we included as attribute for each gene the encoding chromosomic region. Despite differences reported in genomic aberrations (reviewed by Yang and Du [3]), our results corroborated the enrichment in chromosome regions previously described. Typical gain regions located in chromosome arms 7p, 8q and 17q, and we have found these regions over-represented, containing upregulated genes of the MPNST vs. NF gene signature and of the unfiltered list. In addition, we also identified gene enrichment in chromosome arm 4q, proposed as over-represented region [4]. The chromosome arm 15q [5], also reported as gene enriched, resulted only over-represented when we considered the unfiltered gene list, which indicates that genes included in this arm showed lower scores. Nevertheless, the over-representation of three bands from chromosome arm 15q illustrates the relevance of this region including cell proliferation genes, particularly four from the top 20 upregulated genes *(KIAA0101, NUSAP1, PRC1* and *CCNB2).* Over-representation of chromosome bands from the distal part of chromosome arm 17q included *TOP2A, BIRC5* and *TK1* genes. As this region was gained in five patients with poor outcome, a prognostic value has been associated to this amplification [6,7]. Besides these genes, *FOXM1, LOXL2* and *EYA4* were also described as upregulated and included in MPNST aberrant amplified genomic regions, in agreement with our results. In contrast with their over-expression in the gene signature, *HMMR* and *MMP13* were described as included in significant deleted chromosome regions (S11 Table).

We identified chromosome arm 3p over-represented for downregulated genes with the lowest cumulative probability and, with increasing probability, chromosome arms 1p, 5q, and 11q. Chromosome arms 1p, 3p and 11q also showed gene enrichment in the unfiltered downregulated gene list. Whereas chromosome arm 1p was reported as included in significant deleted regions, 3p and 11q were not previously described. Chromosome arm 3p includes a heterogeneous set of genes mainly involved in cell adhesion, cell communication, immune response, nervous system development and axon guidance. Some of them are related with cancer, particularly encoding genes for the Ser/Thr kinase TGFBR2 and the metalloproteinase inhibitor TIMP4. Chromosome arm 11q includes three of the 20 bottom signature genes *(PLEKHB1, CRYAB* and

ENDOD1). Unlike MPNST vs. NF gene signature, the whole unfiltered list also showed over-representation of 9p and 17p chromosome arms. Over-representation of band 17p13.1 for downregulated genes agreed with the MPNST deletion of this region encoding the tumor suppressor TP53 [8,9]. In turn, we did not find either over-representation among downregulated genes of band 9p21, the most frequently deleted region in MPNST [10]. In fact, *CDKN2A* and *CDKN2B,* located in that region, were only slightly downregulated in the unfiltered gene list. Considering reported downregulated genes related with NF and/or MPNST (S12 Table), in agreement with its downregulation, *RASSF2* is located in a significantly deleted region, whereas ITGB4, although downregulated, it locates in a MPNST amplified region, the distal part of chromosome arm 17q. *NKAIN2* and *RABGAP1L*, unexpectedly, were associated to deleted chromosomal regions in cutaneous NF1 neurofibromas.

Discussion B: Expression profile differences between cultured cells and nerve tumors observed in *SOX9*, *SUZ12*, *EGFR*, *SPP1* and *BMP2* genes

SOX9 was highly upregulated in cell culture comparisons whereas only a slight and moderate upregulation was seen in MPNST vs. NF and MPNST vs. control comparisons (S8 Table). In fact, the chosen score threshold of MPNST vs. NF gene signature avoided the inclusion of *SOX9* in this signature. In agreement with our results, Kolberg et al., [6] were unable to find upregulation for *SOX9*. *SUZ12*, the gene encoding one of the members of the silencing epigenetic complex PRC2, that in agreement with De Raedt et al. [11] was downregulated in the comparison MPNST vs. NHSC, did not show downregulation in tumor tissue. The gene encoding the epidermal growth factor receptor EGFR, reported to be amplified and upregulated in MPNST in several studies [12–14], was clearly upregulated in our data in the three comparisons from cell cultures, but it only showed to be upregulated in MPNST vs. control comparison in tumor tissue. Again, in agreement with Kolberg et al. [6], *EGFR* was not upregulated in the MPNST vs. NF comparison in tumor tissue. More paradoxical resulted the encoding gene for the osteopontin SPP1, that in agreement with Thomas et al. [15], resulted upregulated in the comparison MPNST vs. NF of tumor tissue, but it showed to be downregulated in cell cultures. In the unfiltered gene lists (Table J in S2 Table), the gene encoding the bone morphogenetic protein BMP2 was differentially up-regulated in MPNST cultured cells compared to control cells [16], but it was excluded from that respective gene signature due to the score threshold. In contrast with cultured cells, *BMP2* did not show upregulation in tumor MPNST vs. NF comparison. All these comparisons between our results and others previously reported suggest that some

differences in gene expression, imputed to intrinsic variation among tumors, might be due to the nature of tissue.

Discussion C: Panel of genes potentially silenced by hypermethylation of their CpG-island promoter region

In addition to RASSF1, other proteins seem to contribute to control the cell cycle arrest in different ways (S10 Table): the Ras oncogene family member RAB40B, involved in proteasomal degradation of target proteins; the cell growth suppressor FAM107A; the transcriptional repressor FOXS1; RHBDF1, the indirect activator of EGFR, that regulates cell survival, proliferation and migration; the mediator of cell survival NGFR; the regulator of cell cycle progression MX2; the proto-oncogene FGR, that negatively regulates cell migration and adhesion; INPP5D, the negative regulator of proliferation and survival of myeloid cells; the anti-apoptotic STAT6, an important diagnostic marker for solitary fibrous tumor [17]; the tumor suppressor CADM4; and the regulator of cell survival and apoptosis TRAF1.

As regulators of immune responses, S10 Table includes the GTPase activating protein RAP1GAP2, FGR, the proto-oncogene HCK, the arrestin ARRB1, the atypical chemokine receptor ACKR1, and the marker for neural crest Schwann cell lineage S100B. The EF-hand binding protein S100B is involved in intra and extracellular activities as regulator and signal [18]. S100B is downregulated in astrocytes by the epidermal growth factor EGF and the pro-inflammatory cytokine interferon (IFN)-γ [19]. It shows a complex transcriptional regulation that has not been addressed in the MPNST progression. S100B regulates Schwann cell proliferation and myelination with its co-regulator SOX10, described as hypermethylated [20]. However, due to our restrict threshold conditions, SOX10 was only found differentially hypermethylated in MPNST vs. control comparison. Besides S100B, other proteins involved in nervous system development and function are the signal transduction phosphatase inhibitor PPP1R1B, the nerve growth factor receptor NGFR, the scaffolding protein GRASP, a linker of receptors for phosphoinositides to neuronal proteins, the Ras protein RASGRF2, the transcription factor SOX8, that has also been related with some types of cancer [21], the axon outgrowth suppressor SLITRK2, the regulator of calcineurin RCAN1, and several proteins involved in structural functions such as PRIMA1, required to anchor acetylcholinesterase to the membrane of neuronal synapses in brain, the sialyltransferase ST6GALNAC2, the integrin ITGB4, and the myelin components MBP and MPZ.

In addition to genes *MBP, RABGAP1L, MPZ, ITGB4* and *S100B* (S12 Table), genes *PPP1R1B, FGR, NGFR* and *INPP5D* were related with MPNST and/or NF. PPP1R1B is a target of hippocampal dopamine. Reduction of dopamine levels thus decreased the PP1R1B phosphorylation levels in NF1 male mice [22]. FGR was not hyperexpressed in a neurofibrosarcoma compared to control tissue of a NF1 patient [23]. NGFR is a Schwann cell differentiation marker, downregulated in MPNST cell lines vs. control Schwann cells [24], and functional in NF cells [25]. NGFR acts as tumor suppressor that controls survival and death of neural cells, axonal growth and synaptic plasticity [26–29]. Its expression has been detected in human cancers of thyroid, stomach and liver. The NGFR hypermethylation status has been associated with cell proliferation, invasion, formation of colonies, and induced cell apoptosis in human colorectal cancer. NGFR silencing in this type of cancer reduced overall survival and disease-free survival [30]. Also in agreement with our results, loss of genomic region containing INPP5D seemed to cause silencing of INPP5D in cutaneous neurofibromas.

Not only downregulated genes resulted hypermethylated, but also many upregulated genes. Promoters of *KRT18, MEST, and WT1* genes were differentially hypermethylated in comparisons MPNST vs. NF and MPNST vs. control Schwann cells. These genes showed a similar pattern of gene upregulation. In agreement with our results, keratin KRT18 was reported as upregulated [6,31]. KRT18 is a biomarker for clinical diagnosis of cancer [32], although it has not been related with hypermethylation. In contrast, the imprinted genes *MEST* (mesoderm specific transcript) and the transcription factor *WT1 (*Wilms tumor) were related with hypermethylation. The modification in methylation levels at regulatory regions of imprinted genes, that exhibit preferential expression from a paternal allele, has been related with cancer. Aberrant methylation of *MEST* was associated to cervical cancer [33]. Upregulated in several cancers, *WT1* acts as an oncogene rather than a tumor suppressor [34] and aberrant hypermethylation of this gene was reported in hepatocellular carcinoma [35].

Discussion D: HDAC inhibitors counteract repression of CBX7 and over-expression of EZH2

The tightly regulation of cell migration and invasion by silencing CDH1 not only depends on HMGA2 and GSN, but also on the epigenetic Polycomb Repressor Complexes PRC1 and PRC2. Protein subunits of PRC1 and PRC2 control

developmental programs, whose deregulation drives to develop various types of cancer. PRC1 and PRC2, involved in epigenetic regulation of transcription by chromatin remodeling, act as transcriptional repressors by histone modification in two sequential steps. The first step involves trimethylation of histone H3 by PRC2; in the second step, PRC1 monoubiquitinates histone H2A [36]. PRC1 and PRC2 cooperate to condensate chromatin because trimethylated histones seem to recruit PRC1 to the promoter of target genes and, in addition, PRC1 and PRC2 show coordinate regulation in prostate cancer through microRNAs [37]. In MPNST some protein subunits from both PRC1 and PRC2 were differentially expressed and contribute to the inactivation of these two fundamental epigenetic regulators. The presence among top 20 downregulated genes in MPNST vs. NF signature of encoding gene for CBX7, the histone chromatin remodeling enzyme, component of the epigenetic repressor complex PRC1, suggested that PRC1 could also play a role on epigenetic regulation, probably impaired during NF malignant transformation. Our results also suggest dysfunction of PRC2 due to the over-expression of EZH2, subunit of PRC2. Loss of PRC2 has a favorable effect on transcription of genes from the Ras signaling pathway [11]. *EZH2* upregulation and *CBX7* downregulation are reported in breast tumors [36].

CBX7, a protein subunit of PRC1 is silenced in several cancers and proposed for evaluation as prognostic marker [38]. CBX7 is involved in the regulation of cell cycle and proliferation genes [39]. The loss of expression of *CBX7* drives to poor prognosis and progression to malignancy due to its essential role in epithelial to mesenchymal transition (EMT) [38]. Because CBX7 inhibits the silencing activity of HDAC2 over the *CDH1* promoter, the downregulation of CBX7 drives to EMT and to the malignant phenotype [40], not only by its lack of binding to *CDH1* promoter, but also by the lack of repression of the osteopontin gene *SPP1* (rank position 141 in Table A in S2 Table) and other genes associated with cell migration and invasion [41,42]. HDAC inhibitors avoid the silencing effect of HDAC2 over *CDH1* promoter, neutralizing the silencing of CBX7 and restoring expression levels of CDH1, which contributes to avoid metastasis. Being regulated by a member of the HMGA protein family, *CBX7* could also be an indirect target of HDAC inhibitors [38].

The upregulation of the PRC2 protein component EZH2 (rank position 114 in Table A in S2 Table), has also been reported to promote the invasive phenotype by inhibiting the expression of CDH1 by histone methylation at the *CDH1* promoter region. TSA and vorinostat HDAC inhibitors would reverse CDH1 silencing by reducing the PRC2 occupancy on the *CDH1* promoter, attenuating the metastatic phenotype [43]. PRC2

inactivation depends on co-deletion/mutation of *NF1* and *SUZ12* and/or *EED*, genes encoding two subunits of PRC2 [11]. Whereas *SUZ12* and *EED* were not significantly deregulated in MPNST vs. NF gene signature, *EZH2* appeared highly induced. This contradictory upregulation of EZH2, reported in multiple types of cancers, and correlated with poor prognosis, may be explained by a oncogenic activity of EZH2 independent of its transcriptional repression function [44], or as a consequence of the high cell proliferation rate rather than a cause of that proliferation [45]. In fact, *EZH2* expression correlated with the expression of the proliferation marker MKI67, ranking in position 32 in the gene signature. Consequently, the inhibition of EZH2 has shown antitumor effects in MPNST [46]. In agreement with De Raedt *et al.* [11], at least in MPNST showing co-deletion/mutation of *SUZ12* and *EED,* the use of EZH2 inhibitors as therapeutic agents could exacerbate the malignant phenotype. However, TSA and vorinostat HDAC inhibitors would prevent the activity of EZH2 without disturbing PRC2 protein-protein interactions and without altering the expression of PRC2 members, including HDAC1 [43].

Discussion E: Supplementation with acetate precursors as coadjuvant chemotherapy

The encoding gene for the aspartatoacylase ASPA, which metabolizes N-acetyl-L-aspartic acid (NAA) to aspartate and acetate, ranks in position 10 among negative score genes from MPNST vs. NF gene signature. Although ASPA protein is better known by its role in myelin lipid biosynthesis in central nervous system, essential to maintain intact the white matter, it is also involved in neuroblastoma, a cancer of the sympathetic nervous system in which low expression correlates with poor prognosis [47]. It is downregulated at both mRNA and protein levels in esophageal squamous cell carcinoma, glioma and glioblastoma [48–50]. The metabolic reprogramming due to the lack of ASPA causes the starvation of its final product, acetate, the essential precursor for oligodendrocyte myelination and histone acetylation. This metabolic inhibition of histone acetylation induces tumor cell transformation. The supplementation with an external source of acetate could solve the metabolic problem. In fact, a preclinical study testing the supplementation of glycerol triacetate (GTA) as a chemotherapeutic coadjuvant has been reported in glioma [51]. GTA is hydrophobic and able to cross the blood-brain barrier and plasma membranes. The cytostatic growth arrest caused by GTA was similar to the growth arrest obtained with the HDAC inhibitor vorinostat. However, this growth arrest was not associated with apoptosis or differentiation, but to the increased acetylation of proteins involved in cell cycle regulation [52].

In addition to ASPA, the enzyme ADH1B could be de-regulated in the provision of acetate. It ranks in position 7 among negative gene scores of the MPNST vs. NF signature. This enzyme catalyzes the conversion of alcohol in acetaldehyde. In a second step of alcohol degradation, the aldehyde dehydrogenase ALDH2 oxidizes acetaldehyde to acetate. Although the upregulation of ADH1B has been associated with poor prognosis in high-grade serous ovarian cancer [53], it was downregulated in tongue squamous cell carcinoma [54], and appeared upregulated or downregulated at later stages of fetal lung development [55]. The lack of ADH1B in MPNST points to a similar role as ASPA to generate acetylation precursors. Then, supplementation with GTA as coadjuvant chemotherapy could complement two metabolic pathways involved in the synthesis of acetate, involving ASPA and ADH1B. The required experimental confirmation is indispensable to validate this hypothesis.

Materials and methods A: Microarray data pre-processing

Affymetrix platform data were normalized by the function rma from affy package [56] that uses the robust multi-array average (RMA) expression measure. We normalized Agilent platform data within and between arrays by methods loess and quantiles by using the limma R package [57]. After normalization, R and G normalized values were retrieved from MA matrix values to obtain a normalized expression set, including reference values of the two color platform by using the R conversion formula `R=2^((1/2)*(M+2A)); G=2^((1/2)*(-M+2A))`. ABI platform data was normalized by quantiles through the function qnNormalize from ABarray package [58]. When sample batch information was provided in the experiment, we adjusted for batch effects the expression set obtained for each study by the function ComBat of sva package [59] by using the R code:

```
###Exprs set batch adjustment
library(sva)
Batch = targets$batch
mod = model.matrix(~as.factor(phenotype), data=targets)
Exprs_adjusted = ComBat(dat=Exprs_0, batch=Batch, mod=mod,
par.prior=TRUE, prior.plots=FALSE)
```

The resulting expression set was filtered twice to remove features showing little variation. The first filter ruled out features with expression values with standard deviation lower than quantile 0.2 by varFilter function from genefilter package [60] or a similar R script:

```
###Filtering
library(genefilter)
esetIQR <- varFilter(eset, var.func=sd, var.cutoff=0.2,
filterByQuantile=TRUE)
Exprs_filtered<-exprs(esetIQR)
###or
Exprs_filtered<-matrix(nrow= nrow, ncol= 1, apply(Exprs, 1, sd))
rownames(Exprs_filtered)<-rownames(Exprs)
Exprs_filtered<- Exprs [rownames(as.matrix(Exprs_filtered
[Exprs_filtered [,1] > quantile (Exprs_filtered [,1], 0.2),])), ]
```

R code to discard features with FDR values >0.05 obtained by ANOVA in the limma package [57]:

```
library(limma)
phenotypes<-factor(colnames(Exprs_filtered))
aof<-function(x){m<-data.frame(phenotypes,x);anova(aov(x~phenotypes,m))}
anovaresults<-apply(Exprs_filtered, 1, aof)
pvalues<-data.frame(lapply(anovaresults,function(x){x["Pr(>F)"][1,]}))
tpvalues<-t(pvalues)
colnames(tpvalues)<-"pvalue"
fdr.result<-p.adjust(tpvalues[,1], "BH")
bhtresh<-cbind(tpvalues, fdr.result)
order_bhtresh<-bhtresh[order(fdr.result),]
row.names(order_bhtresh)<-
substring(row.names(order_bhtresh),2,nchar(row.names(order_bhtresh)))
```

```
Exprs_filtered_anova<- Exprs_filtered [rownames(Exprs_filtered) %in%
rownames(order_bhtresh[order_bhtresh[,2] < 0.05,]),]
```

R code used to analyze principal components (PCA):

```
Exprs_2<- Exprs_filtered_anova
Exprs_2_t<-t(Exprs_2)
Exprs_2_t_out=prcomp(Exprs_2_t,scale=TRUE)
Cols = function(vec) {
       cols = rainbow(length(unique(vec)))
       return(cols[as.numeric(as.factor(vec))])
       }
###plot in two dimensions
par(mfrow = c(1,2))
plot(Exprs_2_t_out$x[,1:2],col = Cols(classes),pch =19,
       xlab ="PC1",ylab ="PC2")
plot(Exprs_2_t_out$x[,c(1,3)], col = Cols(classes),pch =19,
       xlab ="PC1",ylab ="PC3")
```

R code to determine differentially expressed genes with the limma package [57]:

```
library(limma)
phenotypes<-factor(colnames(Exprs_2))
###Levels: Phenotype1   Phenotype2   Phenotype3
design<-model.matrix(~0+phenotypes)  #or#
design<- modelMatrix(targets,ref="Universal_RNA")
colnames(design)<-levels(phenotypes)
fit <-lmFit(Exprs_2, design) #or# fit<-lmFit(MA, design)
cont.matrix<-makeContrasts("Phenotype1vsPhenotype2"= Phenotype1  -
Phenotype2   , "Phenotype1vsPhenotype3"= Phenotype1-Phenotype3,
"Phenotype2vsPhenotype3"= Phenotype2-Phenotype3, levels=design)
fit2<-contrasts.fit(fit, cont.matrix)
fit2<-eBayes(fit2)

###Comparison Phenotype1vsPhenotype2
toptable_coef_Phenotype1vsPhenotype2<-toptable(fit2,coef="
Phenotype1vsPhenotype2", n=nrow(fit), adjust.method="BH")
```

Materials and methods B: Translation from probe names to human ENSEMBL gene IDs, HUGO IDs and mapping in human chromosome arms

R code to translate probe names of Affymetrix U133 Plus 2.0 platform (example) by using the biomaRt package ([61]:

```
library(biomaRt)
ensembl = useMart(biomart =
"ENSEMBL_MART_ENSEMBL",dataset="hsapiens_gene_ensembl", host =
"sep2015.archive.ensembl.org")
humanensembl<-useDataset("hsapiens_gene_ensembl", mart=ensembl)
probe_to_gene<-getBM(attributes = c("affy_hg_u133_plus_2",
"ensembl_gene_id","hgnc_symbol", "chromosome_name", "band"),
mart=humanensembl)
```

Materials and methods C: Scores of genes across studies and final score for each gene in a comparison between two phenotypes

The score s_{ij} for each gene (i) in each comparison study (j) was calculated modifying the previously proposed formula [62], as follows:

$$s_{ij} = \begin{cases} S(logFC\_m_{ij})\left(1 - \dfrac{MAD_{ij}}{\left|logFC\_m_{ij}\right|}\right)(1 - pval_{ij}) & pval_{ij} < 0.1,\ \ 1 - \dfrac{MAD_{ij}}{\left|logFC\_m_{ij}\right|} > 0,\ \ B > 0 \\ 0 & otherwise \end{cases}$$

$$S(logFC\_m_{ij})\ scales\ logFC\_m_{ij} > 0 \quad between\ 0\ and\ +1\ and$$
$$logFC\_m_{ij} < 0 \quad between\ -1\ and\ 0$$

$pval_{ij}$ and B_{ij} factors for each gene in each study were obtained from computation of the differential expression between two phenotypes by using the data expression sets and the limma R package [57]. MAD_{ij} is the median deviation of each ratio between two phenotypic samples to the $logFC\_m_{ij}$. The robust $logFC\_m_{ij}$ value was calculated as the log_2 median of all expression ratios among the samples of the phenotypes compared. Phenotype expression values were obtained as 2 raised to the expression value derived from each microarray normalization step. The normalization step of positive and negative $logFC\_m_{ij}$ values regarding $logFC\_m_{ij}$ quantiles 99.95% and 0.05%, respectively, yielded scaled $S(logFC\_m_{ij})$ values normalized between -1 and +1. Values of $logFC\_m_{ij}$ > or < than quantiles 99.95% and 0.05%, respectively, were mapped to 1 or -1, respectively. R code used to compute S(logFC_mij) values:

```
scaled_logFC_median<-numeric()
for(i in 1: nrow){
       scaled_logFC_median [i]<-ifelse(logFC_median[i]>0,
       logFC_median[i]/as.numeric(quantile(logFC_median, .9995)),
       logFC_median[i]/abs(as.numeric(quantile(logFC_median, 0.0005))))}
scaled_logFC_median<- ifelse(scaled_logFC_median > 1, 1,
scaled_logFC_median)
scaled_logFC_median<- ifelse(scaled_logFC_median < -1, -1,
scaled_logFC_median)
```

The multiplication of the three factors of the formula shown above gave a s_{ij} score value between -1 and +1, positive for upregulated genes and negative for downregulated genes. Unlike the previously described score, our score retained the sign of size effect distinguishing up- and downregulated genes across the analysis. After computing gene scores in each experiment, we tested the three constraints imposed to the calculation of scores, i.e. P-values <0.1 and penalization factors due to deviations of size effect and B factors >0. If these constraints were satisfied, gene scores retained their values. Otherwise, score values were coerced to zero. R code to compute the gene scores s_{ij}:

```
score<-numeric()
for(i in 1:nrow){score[i]<-
       scaled_logFC_median [i] * (1 - P.Value[i]) * (1 -
       (MAD[i]/abs(logFC_median[i]))) }

###Conditions for score
for(i in 1:nrow){score[i]<-
       ifelse((1 - (MAD[i]/abs(logFC_median[i]))) [i] < 0, 0, score[i])}
```

```
for(i in 1:nrow){score[i]<- ifelse(P.Value[i] > 0.1, 0, score[i])}
for(i in 1:nrow){score[i]<-ifelse(B[i] < 0, 0, score[i])}
```

Materials and methods D: Computation of bias in score values among studies: Bhattacharyya distance (BD) ratio

R code (apply for all genes):

```
###nplatform: number of independent studies
###s: final score
###s1: human score 1
###s2: human score 2
###s3: human score 3
###s4: human score 4
###sm: mouse score
###max_s: maximum score (excluding mouse data)
max_s <- max(c(abs(s1), abs(s2), abs(s3), abs(s4)), na.rm=T)
###max_b: maximum score (including mouse data)
max_b <- max(c(abs(s1), abs(s2), abs(s3), abs(s4), abs(sm)), na.rm=T)
###genename: gene name
###only_human: list of genes ignoring mouse data
###Gene Bhattacharya distance
Bhattacharyya_dist<-numeric()
Bhattacharyya_dist[i]<-
            ifelse(!is.na(s) & s !=0,
                  ifelse(nplatform > 0,
                        ifelse(genename %in% only_human,

      (-1)*log((1/nplatform)*(sum(c(sqrt(abs(s1)/max_s),sqrt(abs(s2)/max_s),
      sqrt(abs(s3)/max_s), sqrt(abs(s4)/max_s)), na.rm=T))),
      (-1)*log((1/nplatform)*(sum(c(sqrt(abs(s1)/max_b),  sqrt(abs(s2)/max_b),
      sqrt(abs(s3)/max_b),      sqrt(abs(s4)/max_b),      sqrt(abs(sm)/max_b)),
      na.rm=T)))), NA), NA)
###Gene DB maximum
max_BD<-numeric()
max_BD<-ifelse(nplatform==5,(-1)*log(1/5),
              ifelse(nplatform==4, (-1)*log(1/4),
                    ifelse(nplatform==3, (-1)*log(1/3),
                          ifelse(nplatform==2, (-1)*log(1/2),
                                ifelse(nplatform==1, 0, NA)))))
###Gene BD_ratio
BD_ratio<-numeric()
BD_ratio<-ifelse(nplatform>1, (Bhattacharyya_dist *100)/max_DB,
                                    ifelse(nplatform==1, 0, NA))
```

Materials and methods E: DNA methylation analysis

SRA files were downloaded from GEO database and transformed to fasq format by the sra.toolkit program. The human genome assembly hg19 (GRCh37) used to align fastq sequences was downloaded from UCSC Genome Bioinformatics with the wget program. The index reference genome was created by the program bwa [63]. We aligned pair end sequences against the reference genome using the program bwa. Through the samtools program [64], we transformed sam files to bam files. Appropriate alignments of resulting bam files from the 6 replicates per phenotype were displayed with the Integrative Genomics Viewer (IGV; [65]) after creating an index for bam files. We inspected data by using the MEDIPS R package [66] obtaining Pearson correlations between all pairs of samples and calculating coverage saturation and calibration plots. Concerning Pearson correlation between all pairs of samples, MPNST samples correlated better among them (median 0.95) than NF and control phenotype samples (medians 0.87 and 0.89, respectively). Comparing pairs of phenotypes, high correlation was observed between NF and control samples (median 0.86) whereas these two phenotypes showed lower correlation with MPNST (median 0.65 between MPNST and NF, and median 0.59 between MPNST and control samples). To ensure that the covering of sequences along the whole genome was sufficient to compute differential DNA methylation between pairs of phenotypes, we applied a coverage saturation analysis to individual samples. As saturation plots show in Fig C, estimated correlation between artificial subsets of sequencing data was higher than 0.9 for every sample. MPNST samples showed estimated correlation values of 0.97; samples of NF exhibited 0.92 of estimated correlation; and 0.93 was the value for control replicates. These results guaranteed enough sequencing depth and allowed us to conclude that the selected window size of 200 was appropriate to accomplish the following analysis. The calibration plots shown in Fig D indicated that the normalization step regarding CpG density assured an effective MeDIP enrichment. We also compared the resulting bam files for differential methylation by MEDIPS package. We substituted stacked reads by only one representative to avoid false positives in the comparison of conditions. Translation of coordinates to promoters was carried out downloading coordinates from the UCSC Table Browser, selecting the human assembly Feb. 2009 (GRCh37/hg19), group Genes and Gene Predictions, track RefSeq Genes, BED output format, and get BED after selecting upstream by 1500 bases (promoter).

```
###DNA methylation analysis
###Download and reprocessing sra files(command line in linux)
      ###sra-toolkit program: download fastq files
```

```
      prefetch –v SRR0427XX fasq.dump –outdir/opt/fastq/ --split-
files/home/sra_files/SRR0427XX.sra
      ###wget program: download reference genome hg19
      wget
http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/chromFa.tar.gz
      ###bwa program: reference genome index
      bwa index –p bwa index –p hg19bwaidx –a bwtsw hg19.fa
      ###bwa program: alignment of individual sequences to reference genome
      bwa mem hg19bwaidx SRR0427XX_1.fastq SRR0427XX_2.fastq > SRR0427XX_aln-
pe.sam
      ###samtools program:  sam files to bam files
      samtools view –bhuS SRR0427XX_aln-pe.sam | samtools sort –
SRR0427XX_sorted
      ###samtools program: index for bam files
      samtools index SRR0427XX_aln-pe.bam


###Files in the R working directory:
#SRR042748_sorted.bam
#SRR042750_sorted.bam
#SRR042752_sorted.bam
#SRR042754_sorted.bam
#SRR042756_sorted.bam
#SRR042758_sorted.bam
#SRR042760_sorted.bam
#SRR042762_sorted.bam
#SRR042764_sorted.bam
#SRR042766_sorted.bam
#SRR042768_sorted.bam
#SRR042770_sorted.bam
#SRR042772_sorted.bam
#SRR042774_sorted.bam
#SRR042776_sorted.bam
#SRR042778_sorted.bam
#SRR042780_sorted.bam
#SRR042782_sorted.bam

###Read the input and create the MEDIPS sets for the three different
phenotypes (MPNST, NF and Normal)
library(MEDIPS)
library(BSgenome.Hsapiens.UCSC.hg19)
BSgenome="BSgenome.Hsapiens.UCSC.hg19"
uniq=1e-3
extend=300
shift=0
ws=200
bam_file_MPNST<-c("SRR042748_sorted.bam", "SRR042750_sorted.bam",
"SRR042752_sorted.bam", "SRR042754_sorted.bam", "SRR042756_sorted.bam",
"SRR042758_sorted.bam")
bam_file_NF<-c("SRR042760_sorted.bam",    "SRR042762_sorted.bam",
"SRR042764_sorted.bam",   "SRR042766_sorted.bam", "SRR042768_sorted.bam",
"SRR042770_sorted.bam")
bam_file_Normal<-c("SRR042772_sorted.bam",   "SRR042774_sorted.bam",
"SRR042776_sorted.bam",  "SRR042778_sorted.bam", "SRR042780_sorted.bam",
"SRR042780_sorted.bam", "SRR042782_sorted.bam")
MPNST_MeDIP = lapply(X=bam_file_MPNST, FUN= MEDIPS.createSet, BSgenome =
BSgenome, extend = extend, shift = shift, uniq = TRUE, window_size = ws)
NF_MeDIP = lapply(X=bam_file_NF, FUN= MEDIPS.createSet, BSgenome = BSgenome,
extend = extend, shift = shift, uniq = TRUE, window_size = ws)
Normal_MeDIP = lapply(X=bam_file_Normal, FUN= MEDIPS.createSet, BSgenome =
BSgenome, extend = extend, shift = shift, uniq = TRUE, window_size = ws)

###Local density of CpGs considering the genome and window parameters
CS_MPNST = MEDIPS.couplingVector(pattern = "CG", refObj = MPNST_MeDIP[[1]])

###Exploring data and quality control of samples: Obtaining Pearson
correlations between all pairs of samples
cor.matrix = MEDIPS.correlation(MSets=c(MPNST_MeDIP, NF_MeDIP, Normal_MeDIP),
```

```
plot =T, method ="pearson")

###Exploring data and quality control of samples: Calibration plots for MPNST
MeDIP set (calibration plots were obtained in a similar way for NF and Normal
MeDIP sets)
for (i in 1:length(MPNST_MeDIP)){
        png(paste("D:/working_directory/Suppl_Figures/MPNST",
"/Suppl.Fig1_calibration_MPNST_", i, ".png", sep=""))
        MEDIPS.plotCalibrationPlot(MSet=MPNST_MeDIP[[i]], CSet=CS_MPNST)
        dev.off()
}

###Exploring data and quality control of samples: Saturation analysis for
MPNST MeDIP set (saturation analyses were carried out in a similar way for NF
and Normal MeDIP sets)
for (i in 1:length(MPNST_MeDIP)){
  png(paste("D:/working_directory/Suppl_Figures/MPNST",
"/Suppl.Fig2_saturation_cancer_", i, ".png", sep=""))
  sr=MEDIPS.saturation(file=bam_file_MPNST[i], BSgenome = BSgenome, uniq =
TRUE, extend =extend, shift= shift, window_size = ws, nit = 10, nrit = 1,
empty_bins = TRUE, rank = FALSE)
  MEDIPS.plotSaturation(sr)
  dev.off()
}

###Differential methylation analysis between MPNST and NF (comparisons between
MPNST vs. Normal and NF vs. Normal were analyzed in a similar way)
results_MPNST_vs_NF = MEDIPS.meth(MSet1 = cancer_MeDIP, MSet2 = benign_MeDIP,
CSet = CS_cancer, p.adj = "bonferroni", diff.method = "edgeR",
                                  MeDIP = T, CNV = F, minRowSum = 10)


###Selecting significant windows for MPNST vs. NF comparison (similarly were
obtained significant windows for MPNST_vs_Normal and NF_vs_Normal):
mr.edgeR_MeDIP_MPNST_vs_NF.s = MEDIPS.selectSig(results = results_MPNST_vs_NF,
p.value = 0.1, adj = T, ratio = NULL, bg.counts = NULL, CNV = F)
###Merging hypermethylated regions
DMR_MPNST_vs_NF.s.gain =
mr.edgeR_MeDIP_MPNST_vs_NF.s[which(mr.edgeR_MeDIP_MPNST_vs_NF.s[,
grep("logFC", colnames(mr.edgeR_MeDIP_MPNST_vs_NF.s))] > 0), ]
DMR_MPNST_vs_NF.s.gain.m = MEDIPS.mergeFrames(frames = DMR_MPNST_vs_NF.s.gain,
distance = 1)
###Merging hypomethylated regions
DMR_MPNST_vs_NF.s.loss =
mr.edgeR_MeDIP_MPNST_vs_NF.s[which(mr.edgeR_MeDIP_MPNST_vs_NF.s[,
grep("logFC", colnames(mr.edgeR_MeDIP_MPNST_vs_NF.s))] < 0), ]
DMR_MPNST_vs_NF.s.loss.m = MEDIPS.mergeFrames(frames = DMR_MPNST_vs_NF.s.loss,
distance = 1)
###Extraction of data from regions of interest (ROIs)
columns = names(results_MPNST_vs_NF)[grep("counts|rpkm|edgeR",
names(results_MPNST_vs_NF))]
rois_MPNST_vs_NF.s.gain.m = MEDIPS.selectROIs(results = results_MPNST_vs_NF,
rois = DMR_MPNST_vs_NF.s.gain.m, columns = columns, summarize = "avg")
rois_MPNST_vs_NF.s.loss.m = MEDIPS.selectROIs(results = results_MPNST_vs_NF,
rois = DMR_MPNST_vs_NF.s.loss.m, columns = columns, summarize = "avg")



###Annotation
#Limiting hypermethylated ROIs to chromosomes 1 to 22 and X (analogously for
hypomethylated)
rois_MPNST_vs_NF.s.gain.m_select<-
rois_MPNST_vs_NF.s.gain.m[rois_MPNST_vs_NF.s.gain.m$chr=="chr1"|rois_MPNST_vs_
NF.s.gain.m$chr=="chr2"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr3"|rois_MPNST_vs_NF.s.gain.m$chr=="chr4"|
```

```
rois_MPNST_vs_NF.s.gain.m$chr=="chr5"|rois_MPNST_vs_NF.s.gain.m$chr=="chr6"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr7"|rois_MPNST_vs_NF.s.gain.m$chr=="chr8"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr9"|rois_MPNST_vs_NF.s.gain.m$chr=="chr10"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr11"|rois_MPNST_vs_NF.s.gain.m$chr=="chr12"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr13"|rois_MPNST_vs_NF.s.gain.m$chr=="chr14"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr15"|rois_MPNST_vs_NF.s.gain.m$chr=="chr16"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr17"|rois_MPNST_vs_NF.s.gain.m$chr=="chr18"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr19"|rois_MPNST_vs_NF.s.gain.m$chr=="chr20"|

rois_MPNST_vs_NF.s.gain.m$chr=="chr21"|rois_MPNST_vs_NF.s.gain.m$chr=="chr22"|

rois_MPNST_vs_NF.s.gain.m$chr=="chrX",]


###Promoter annotation
#File of coordinates from promoters download from UCSC Table Browser:
""promoter_1500_hg19"
promoter_1500_hg19<-read.csv("promoter_1500_hg19", sep="\t", header=FALSE)
promoter_1500_hg19<-promoter_1500_hg19[,1:4]
colnames(promoter_1500_hg19)<-c("chr","start", "end","promoter")
aaa<-as.character()
for(i in 1:57111){aaa[i]<-gsub("_up_1500_chr(([0-9]+)|X)_([0-9]+)_[f-r]", "",
as.character(promoter_1500_hg19$promoter[i]))}
promoter_1500_hg19_df<-cbind(aaa, promoter_1500_hg19[,1:3])
colnames(promoter_1500_hg19_df)[1]<-"refseq_mrna"

#Adding RefSeq promoter annotation to ROIs (hypermethylated, analogously for
hypomethylated)
rois_MPNST_vs_NF.s.gain.m_anot = MEDIPS.setAnnotation(regions =
rois_MPNST_vs_NF.s.gain.m_select, annotation = promoter_1500_hg19_df)
id<-list()
for(i in 1:((length(colnames(rois_MPNST_vs_NF.s.gain.m_select))-36){id[[i]]<-
rois_MPNST_vs_NF.s.gain.m_anot[!is.na(rois_MPNST_vs_NF.s.gain.m_anot[,36+i]),c
(1:3, 32:36, 36+i)]
colnames(id[[i]])[9]<-"ENSPROMOTER"}
rois_MPNST_vs_NF.s.gain.m_anot_promoter <- do.call("rbind", id)

#Let unique values of edgeR.logFC and edgeR.logCPM for each RefSeq annotated
promoter (hypermethylated; analogously for hypomethylated)
groups<-split(rois_MPNST_vs_NF.s.gain.m_anot_promoter,
rois_MPNST_vs_NF.s.gain.m_anot_promoter[,9])
for(i in(1:length(groups))){groups[[i]]<-
groups[[i]][abs(groups[[i]]$edgeR.logFC)==max(abs(groups[[i]]$edgeR.logFC)),][
1,]}
rois_MPNST_vs_NF.s.gain.m_anot_promoter <- do.call("rbind", groups)
rois_MPNST_vs_NF.s.gain.m_anot_promoter<-
rois_MPNST_vs_NF.s.gain.m_anot_promoter[order(rois_MPNST_vs_NF.s.gain.m_anot_p
romoter$edgeR.adj.p.value, decreasing=F),]

#Translate RefSeq annotation to ENSEMBL annotation
library(biomaRt)
ensembl = useMart(biomart =
"ENSEMBL_MART_ENSEMBL",dataset="hsapiens_gene_ensembl", host =
"feb2014.archive.ensembl.org")
humanensembl<-useDataset("hsapiens_gene_ensembl", mart=ensembl)
humangenes_mrna<-getBM(attributes = c("ensembl_gene_id", "refseq_mrna",
"hgnc_symbol"), mart=humanensembl)
humangenes_ncrna<-getBM(attributes = c("ensembl_gene_id", "refseq_ncrna",
"hgnc_symbol"), mart=humanensembl)

colnames(humangenes_mrna)[2]<-"ENSPROMOTER"
```

```
colnames(humangenes_ncrna)[2]<-"ENSPROMOTER"

rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes_1<-
merge(rois_MPNST_vs_NF.s.gain.m_anot_promoter, humangenes_mrna,
by="ENSPROMOTER")
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes_2<-
merge(rois_MPNST_vs_NF.s.gain.m_anot_promoter, humangenes_ncrna,
by="ENSPROMOTER")
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes<-
rbind(rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes_1,
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes_2)
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes<-
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes[!is.na(rois_MPNST_vs_NF.s.gain.m
_anot_promoter_genes$ensembl_gene_id),]
#Let unique values of edgeR.logFC and edgeR.logCPM for each ENSEMBL annotated
promoter (hypermethylated; analogously for hypomethylated)
groups<-split(rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes,
rois_MPNST_vs_NF.s.gain.m_anot_promoter_genes[,10])
for(i in(1:length(groups))){groups[[i]]<-
groups[[i]][abs(groups[[i]]$edgeR.logFC)==max(abs(groups[[i]]$edgeR.logFC)),][
1,]}
rois_MPNST_vs_NF.s.gain.m_anot_promoter_transl <- do.call("rbind", groups)
rois_MPNST_vs_NF.s.gain.m_anot_promoter_transl<-
rois_MPNST_vs_NF.s.gain.m_anot_promoter_transl[order(rois_MPNST_vs_NF.s.gain.m
_anot_promoter_transl$edgeR.adj.p.value, decreasing=F),]
```

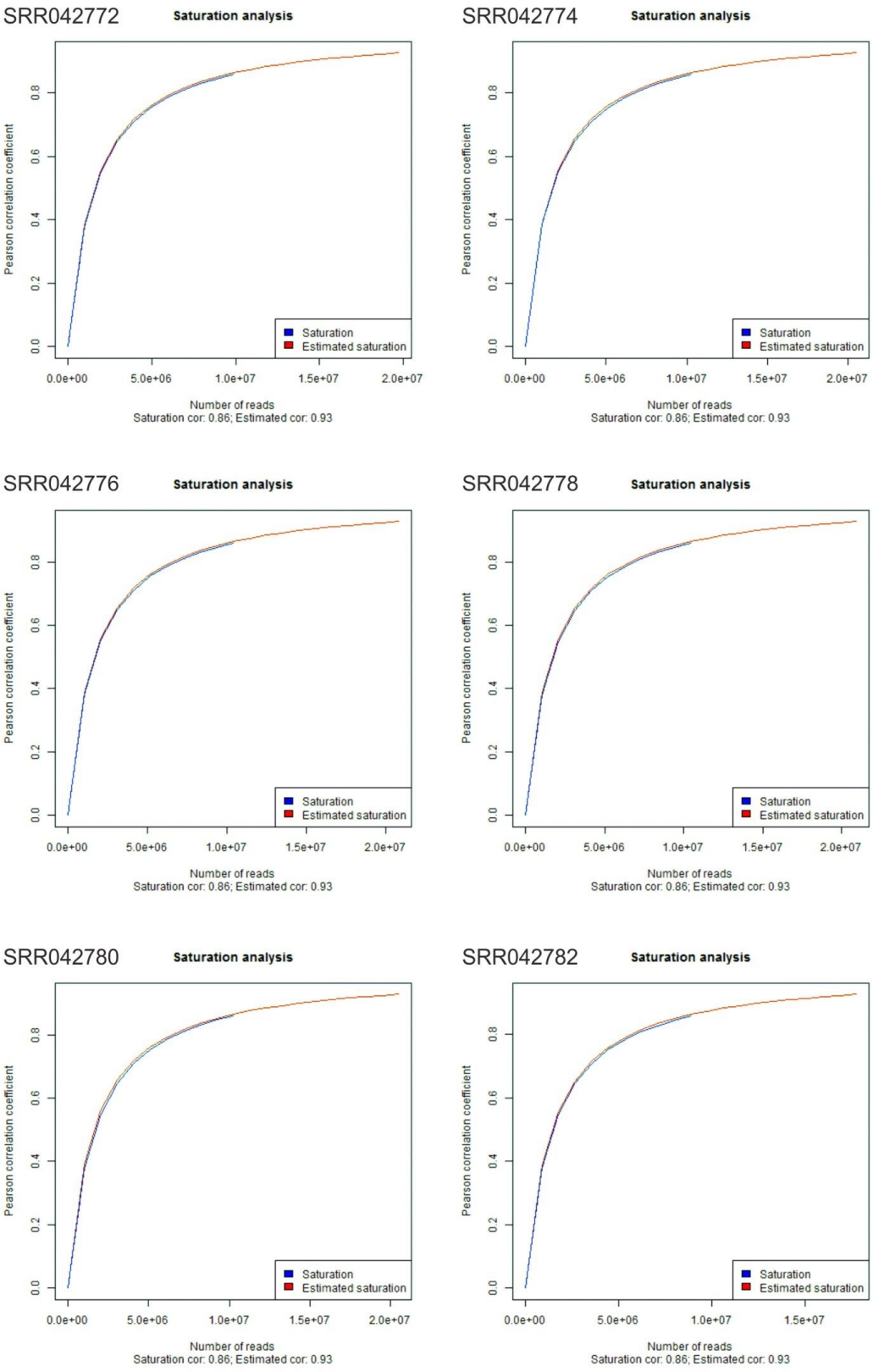


Fig C (a): Saturation plots of control Schwann cells' samples.

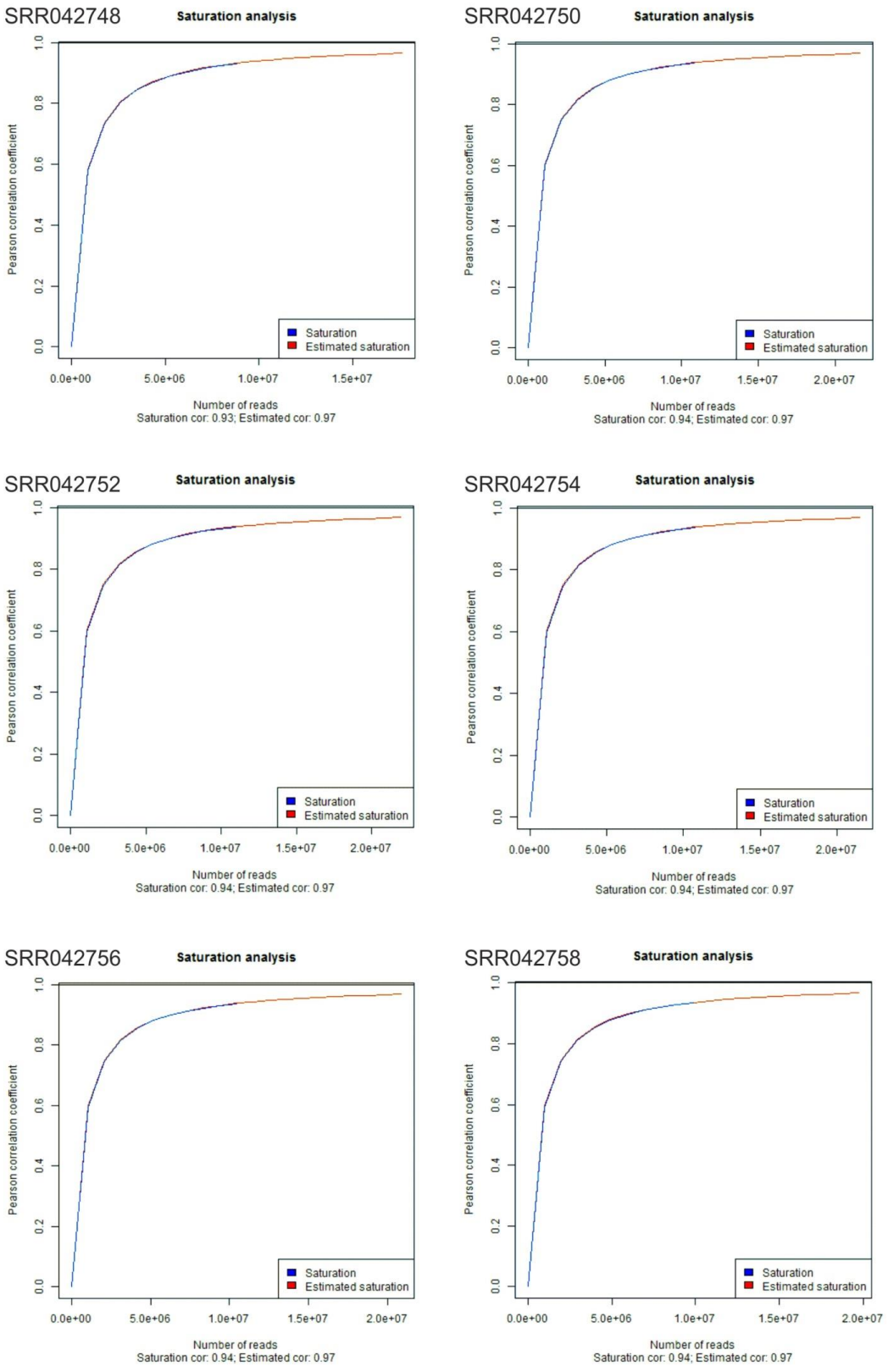


Fig C (b): Saturation plots of MPNST samples.

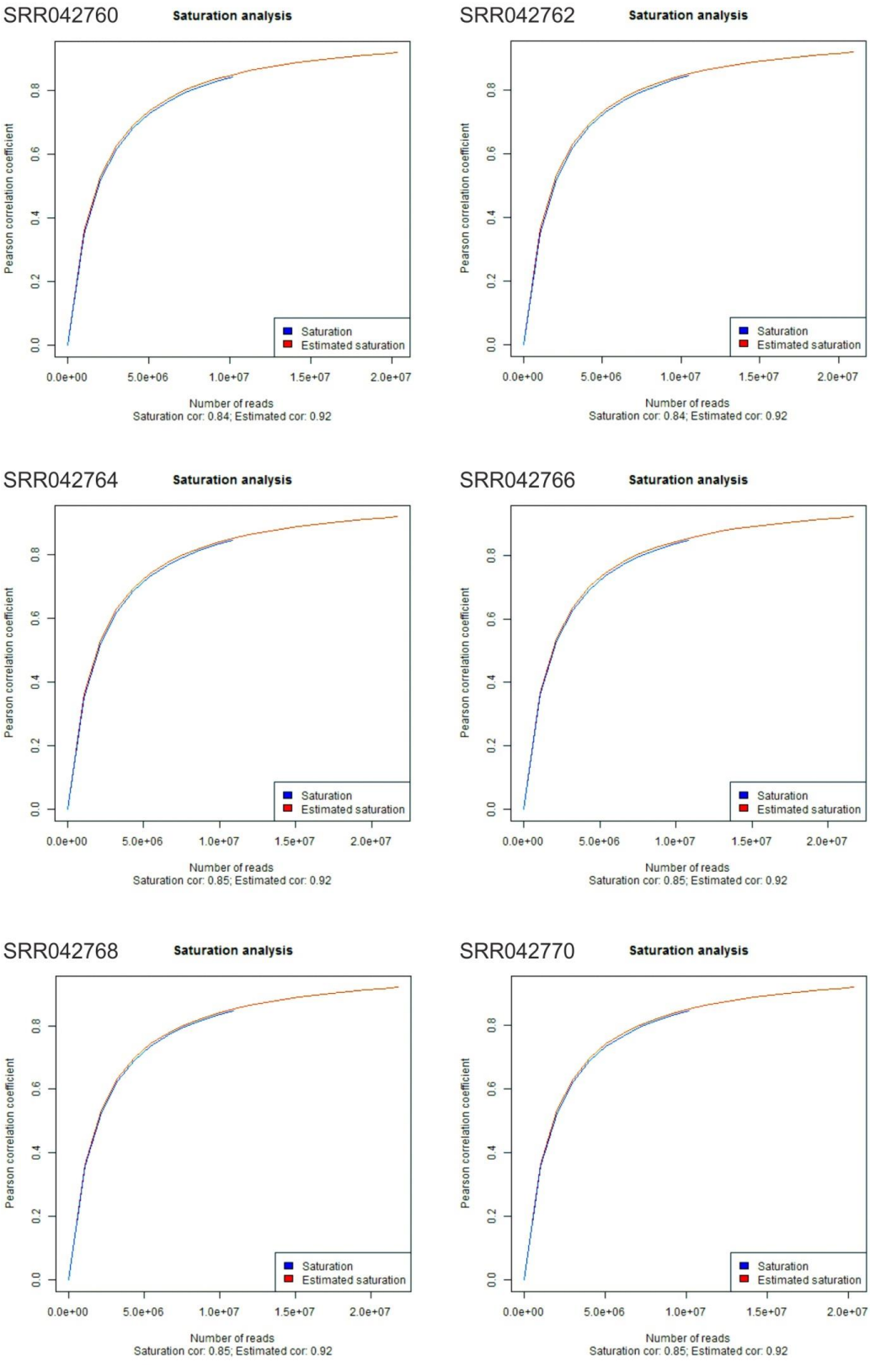


Fig C (c): Saturation plots of NF samples.

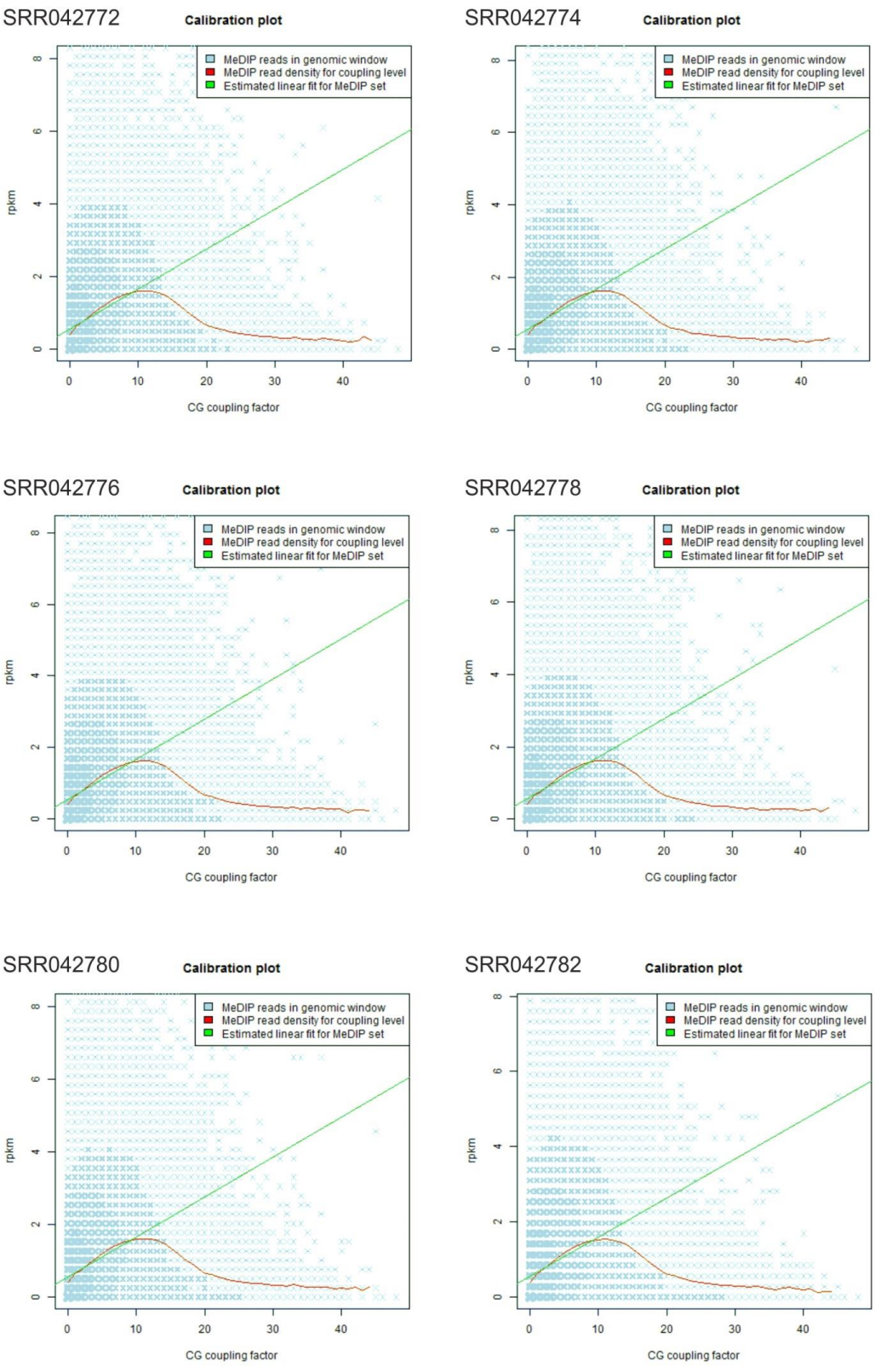


Fig D (a): Calibration plots of control Schwann cells' samples.

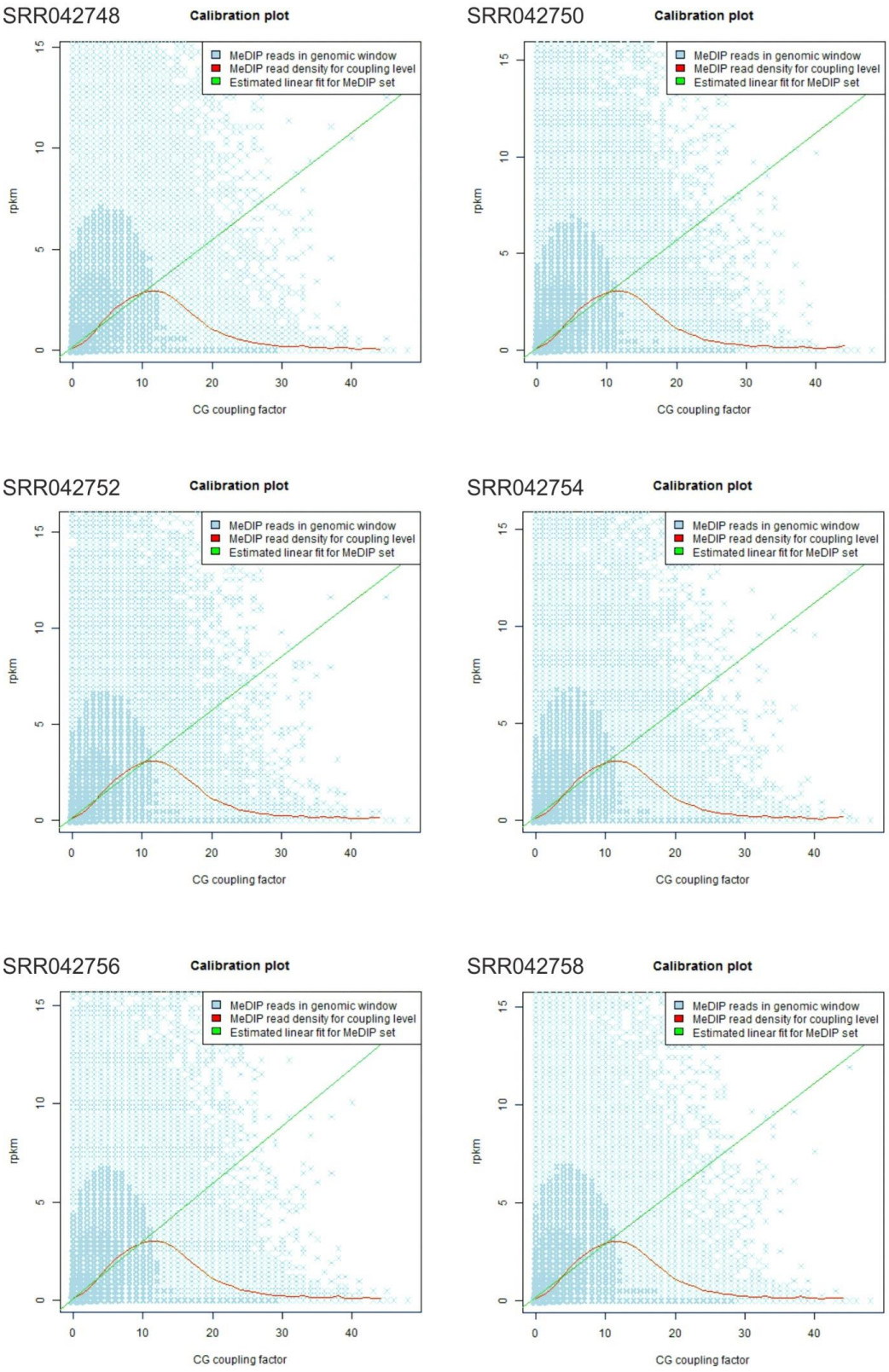


Fig D (b): Calibration plots of control Schwann cells' samples.

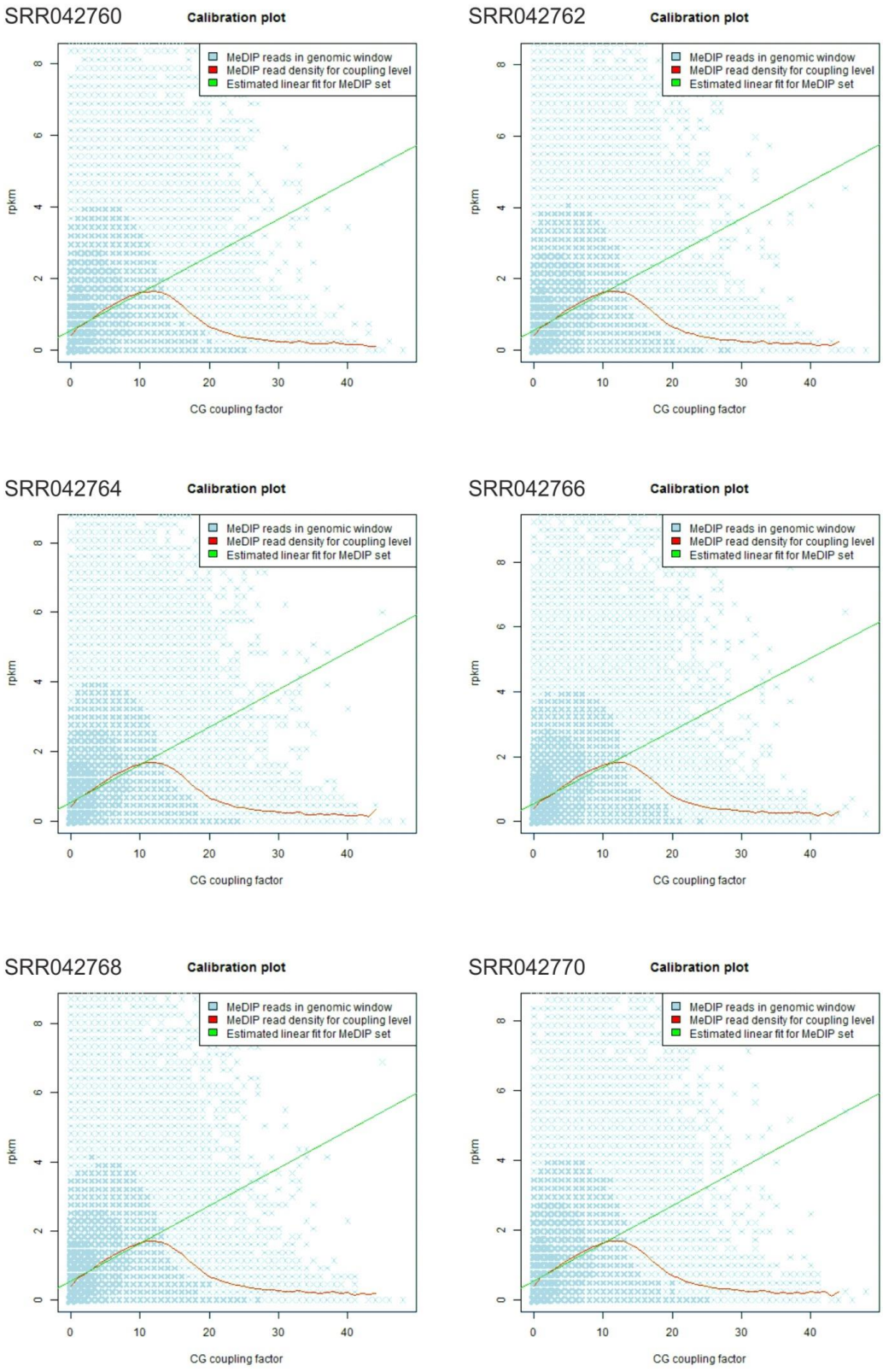


Fig D (c): Calibration plots of NF samples.

References

1. Hsieh Y-Y, Chou C-J, Lo H-L, Yang P-M. Repositioning of a cyclin-dependent kinase inhibitor GW8510 as a ribonucleotide reductase M2 inhibitor to treat human colorectal cancer. Cell Death Discov. 2016;2: 16027. doi:10.1038/cddiscovery.2016.27
2. Lecomte T, Ferraz J-M, Zinzindohoué F, Loriot M-A, Tregouet D-A, Landi B, et al. Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. Clin Cancer Res. 2004;10: 5880–8. doi:10.1158/1078-0432.CCR-04-0169
3. Yang J, Du X. Genomic and molecular aberrations in malignant peripheral nerve sheath tumor and their roles in personalized target therapy. Surg Oncol. 2013;22: e53–e57. doi:10.1016/j.suronc.2013.06.003
4. Upadhyaya M, Spurlock G, Thomas L, Thomas NST, Richards M, Mautner VF, et al. Microarray-based copy number analysis of neurofibromatosis type-1 (NF1)-associated malignant peripheral nerve sheath tumors reveals a role for Rho-GTPase pathway genes in NF1 tumorigenesis. Hum Mutat. 2012;33: 763–776. doi:10.1002/humu.22044
5. Schmidt H, Taubert H, Meye A, Würl P, Bache M, Bartel F, et al. Gains in chromosomes 7, 8q, 15q and 17q are characteristic changes in malignant but not in benign peripheral nerve sheath tumors from patients with Recklinghausen's disease. Cancer Let. 2000. doi:10.1016/S0304-3835(00)00426-2
6. Kolberg M, Høland M, Lind GE, Ågesen TH, Skotheim RI, Sundby Hall K, et al. Protein expression of BIRC5, TK1, and TOP2A in malignant peripheral nerve sheath tumours - A prognostic test after surgical resection. Mol Oncol. 2015;9: 1129–1139. doi:10.1016/j.molonc.2015.02.005
7. Kresse SH, Skårn M, Ohnstad HO, Namløs HM, Bjerkehagen B, Myklebost O, et al. DNA copy number changes in high-grade malignant peripheral nerve sheath tumors by array CGH. Mol Cancer. 2008;7: 48. doi:10.1186/1476-4598-7-48
8. Legius E, Dierick H, Wu R, Hall BK, Marynen P, Cassiman JJ, et al. TP53 mutations are frequent in malignant NF1 tumors. Genes Chromosom Cancer. 1994;10: 250–5.
9. Lothe RA, Smith-Sørensen B, Hektoen M, Stenwig AE, Mandahl N, Sæter G, et al. Biallelic inactivation ofTP53 rarely contributes to the development of malignant peripheral nerve sheath tumors. Genes Chromosom Cancer. 2001;30: 202–206. doi:10.1002/1098-2264(2000)9999:9999<::AID-GCC1079>3.0.CO;2-5

10. Sabah M, Cummins R, Leader M, Kay E. Loss of p16INK4A Expression Is Associated With Allelic Imbalance/Loss of Heterozygosity of Chromosome 9p21 in Microdissected Malignant Peripheral Nerve Sheath Tumors. Appl Immunohistochem Mol Morphol. 2006;14: 97–102. doi:10.1097/01.pai.0000143787.80564.f5
11. De Raedt T, Beert E, Pasmant E, Luscan A, Brems H, Ortonne N, et al. PRC2 loss amplifies Ras-driven transcription and confers sensitivity to BRD4-based therapies. Nature. 2014;514: 247–251. doi:10.1038/nature13561
12. Holtkamp N, Malzer E, Zietsch J, Okuducu AF, Mucha J, Mawrin C, et al. EGFR and erbB2 in malignant peripheral nerve sheath tumors and implications for targeted therapy. Neuro Oncol. 2008;10: 946–57. doi:10.1215/15228517-2008-053
13. Rahrmann EP, Moriarity BS, Otto GM, Watson AL, Choi K, Collins MH, et al. Trp53 Haploinsufficiency Modifies EGFR-Driven Peripheral Nerve Sheath Tumorigenesis. Am J Pathol. 2014;184: 2082–2098. doi:10.1016/j.ajpath.2014.04.006
14. Tabone-Eglinger S, Bahleda R, Côté J-F, Terrier P, Vidaud D, Cayre A, et al. Frequent EGFR Positivity and Overexpression in High-Grade Areas of Human MPNSTs. Sarcoma. 2008;2008: 849156. doi:10.1155/2008/849156
15. Thomas LE, Winston J, Rad E, Mort M, Dodd KM, Tee AR, et al. Evaluation of copy number variation and gene expression in neurofibromatosis type-1-associated malignant peripheral nerve sheath tumours. Hum Genomics. 2015;9: 3. doi:10.1186/s40246-015-0025-3
16. Sun D, Haddad R, Kraniak JM, Horne SD, Tainsky MA. RAS/MEK-independent gene expression reveals BMP2-related malignant phenotypes in the Nf1-deficient MPNST. Mol Cancer Res. 2013;11: 616–27. doi:10.1158/1541-7786.MCR-12-0593
17. Cheah AL, Billings SD, Goldblum JR, Carver P, Tanas MZ, Rubin BP. STAT6 rabbit monoclonal antibody is a robust diagnostic tool for the distinction of solitary fibrous tumour from its mimics. Pathology. 2014;46: 389–95. doi:10.1097/PAT.0000000000000122
18. Sorci G, Giovannini G, Riuzzi F, Bonifazi P, Zelante T, Zagarella S, et al. The Danger Signal S100B Integrates Pathogen– and Danger–Sensing Pathways to Restrain Inflammation. PLoS Pathog. Public Library of Science; 2011;7: e1001315. doi:10.1371/JOURNAL.PPAT.1001315
19. Donato R, Sorci G, Riuzzi F, Arcuri C, Bianchi R, Brozzi F, et al. S100B's double life: Intracellular regulator and extracellular signal. Biochim Biophys Acta.

2009;1793: 1008–1022. doi:10.1016/j.bbamcr.2008.11.009

20. Feber a, Wilson G, Zhang L, Presneau N, Idowu B, Down T, et al. Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. 2011; 515–524. doi:10.1101/gr.109678.110.
21. Xie C, Han Y, Liu Y, Han L, Liu J. miRNA-124 down-regulates SOX8 expression and suppresses cell proliferation in non-small cell lung cancer. Int J Clin Exp Pathol. 2014;7: 6534–42.
22. Diggs-Andrews KA, Brown JA, Gianino SM, Rubin JB, Wozniak DF, Gutmann DH. Sex Is a major determinant of neuronal dysfunction in neurofibromatosis type 1. Ann Neurol. 2014;75: 309–316. doi:10.1002/ana.24093
23. Rowley PT, Kosciolek B, Bader JL. Oncogene expression in neurofibromatosis. Ann N Y Acad Sci. 1986;486: 327–32.
24. Miller SJ, Rangwala F, Williams J, Ackerman P, Kong S, Jegga AG, et al. Large-scale molecular comparison of human Schwann cells to malignant peripheral nerve sheath tumor cell lines and tissues. Cancer Res. 2006;66: 2584–2591. doi:10.1158/0008-5472.CAN-05-3330
25. Hoshi N, Yamaki T, Hiraki H, Natsume T, Saitoh A, Watanabe K, et al. Functional nerve growth factor receptor in von Recklinghausen neurofibromatosis: an immunocytochemical and short-term culture study. Pathol Int. 1996;46: 1–8.
26. Dechant G, Barde Y-A. The neurotrophin receptor p75NTR: novel functions and implications for diseases of the nervous system. Nat Neurosci. 2002;5: 1131. doi:10.1038/NN1102-1131
27. Roux PP, Barker PA. Neurotrophin signaling through the p75 neurotrophin receptor. Prog Neurobiol. 2002;67: 203–233. doi:10.1016/S0301-0082(02)00016-3
28. Underwood CK, Coulson EJ. The p75 neurotrophin receptor. Int J Biochem Cell Biol. 2008;40: 1664–1668. doi:10.1016/j.biocel.2007.06.010
29. Vicario A, Kisiswa L, Tann JY, Kelly CE, Ibáñez CF, Beg A, et al. Neuron-type-specific signaling by the p75NTR death receptor is regulated by differential proteolytic cleavage. J Cell Sci. 2015;128: 1507–17. doi:10.1242/jcs.161745
30. Yang Z, Chen H, Huo L, Yang Z, Bai Y, Fan X, et al. Epigenetic inactivation and tumor-suppressor behavior of NGFR in human colorectal cancer. Mol Cancer Res. 2015;13: 107–19. doi:10.1158/1541-7786.MCR-13-0247
31. Patel A V, Eaves D, Jessen WJ, Rizvi TA, Ecsedy JA, Qian MG, et al. Ras-driven transcriptome analysis identifies aurora kinase A as a potential malignant peripheral nerve sheath tumor therapeutic target. Clin Cancer Res. 2012;18:

5020–30. doi:10.1158/1078-0432.CCR-12-1072

32. Weng Y-R, Cui Y, Fang J-Y. Biological functions of cytokeratin 18 in cancer. Mol Cancer Res. 2012;10: 485–93. doi:10.1158/1541-7786.MCR-11-0222
33. Vidal AC, Henry NM, Murphy SK, Oneko O, Nye M, Bartlett JA, et al. PEG1/MEST and IGF2 DNA methylation in CIN and in cervical cancer. ClinTransl Oncol. 2014;16: 266–72. doi:10.1007/s12094-013-1067-4
34. Huff V. Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. Nat Rev Cancer. 2011;11: 111–21. doi:10.1038/nrc3002
35. Mžik M, Chmelařová M, John S, Laco J, Slabý O, Kiss I, et al. Aberrant methylation of tumour suppressor genes WT1, GATA5 and PAX5 in hepatocellular carcinoma. Clin Chem Lab Med. 2016;54: 1971–1980. doi:10.1515/cclm-2015-1198
36. Meseure D, Vacher S, Alsibai KD, Nicolas A, Chemlali W, Caly M, et al. Expression of ANRIL-Polycomb Complexes-CDKN2A/B/ARF Genes in Breast Tumors: Identification of a Two-Gene (EZH2/CBX7) Signature with Independent Prognostic Value. Mol Cancer Res. 2016;14: 623–33. doi:10.1158/1541-7786.MCR-15-0418
37. Cao Q, Mani R-S, Ateeq B, Dhanasekaran SM, Asangani IA, Prensner JR, et al. Coordinated regulation of polycomb group complexes through microRNAs in cancer. Cancer Cell. NIH Public Access; 2011;20: 187–99. doi:10.1016/j.ccr.2011.06.016
38. Pallante P, Sepe R, Puca F, Fusco A. High mobility group A proteins as tumor markers. Front Med. 2015;2: 15. doi:10.3389/fmed.2015.00015
39. Forzati F, Federico A, Pallante P, Abbate A, Esposito F, Malapelle U, et al. CBX7 is a tumor suppressor in mice and humans. J Clin Invest. 2012;122: 612–23. doi:10.1172/JCI58620
40. Federico A, Pallante P, Bianco M, Ferraro A, Esposito F, Monti M, et al. Chromobox protein homologue 7 protein, with decreased expression in human carcinomas, positively regulates E-cadherin expression by interacting with the histone deacetylase 2 protein. Cancer Res. American Association for Cancer Research; 2009;69: 7079–87. doi:10.1158/0008-5472.CAN-09-1542
41. Sepe R, Formisano U, Federico A, Forzati F, Bastos AU, D'Angelo D, et al. CBX7 and HMGA1b proteins act in opposite way on the regulation of the SPP1 gene expression. Oncotarget. 2015;6: 2680–92. doi:10.18632/oncotarget.2777
42. Pallante P, Sepe R, Federico A, Forzati F, Bianco M, Fusco A. CBX7 modulates the expression of genes critical for cancer progression. PLoS One. 2014;9: e98295. doi:10.1371/journal.pone.0098295

43. Cao Q, Yu J, Dhanasekaran SM, Kim JH, Mani R-S, Tomlins SA, et al. Repression of E-cadherin by the polycomb group protein EZH2 in cancer. Oncogene. 2008;27: 7274–84. doi:10.1038/onc.2008.333
44. Tan J, Yan Y, Wang X, Jiang Y, Xu HE. EZH2: biology, disease, and structure-based drug discovery. Acta Pharmacol Sin. 2014;35: 161–74. doi:10.1038/aps.2013.161
45. Wassef M, Michaud A, Margueron R. Association between EZH2 expression, silencing of tumor suppressors and disease outcome in solid tumors. Cell Cycle. 2016;15: 2256–62. doi:10.1080/15384101.2016.1208872
46. Zhang P, Yang X, Ma X, Ingram DR, Lazar AJ, Torres KE, et al. Antitumor effects of pharmacological EZH2 inhibition on malignant peripheral nerve sheath tumor through the miR-30a and KPNB1 pathway. Mol Cancer. 2015;14: 55. doi:10.1186/s12943-015-0325-1
47. Long P, Stradecki H, Minturn J. Differential aminoacylase expression in neuroblastoma. Int J Cancer. 2011;129: 1322–30.
48. Yan W, Shih J. Identification of unique expression signatures and therapeutic targets in esophageal squamous cell carcinoma. BMC Res Notes. 2012;5: 73.
49. Long PM, Moffett JR, Namboodiri AMA, Viapiano MS, Lawler SE, Jaworski DM. N-Acetylaspartate (NAA) and N-Acetylaspartylglutamate (NAAG) Promote Growth and Inhibit Differentiation of Glioma Stem-like Cells. J Biol Chem. 2013;288: 26188–26200. doi:10.1074/jbc.M113.487553
50. Long PM, Tighe SW, Driscoll HE, Fortner KA, Viapiano MS, Jaworski DM. Acetate Supplementation as a Means of Inducing Glioblastoma Stem-Like Cell Growth Arrest. J Cell Physiol. 2015;230: 1929–1943. doi:10.1002/jcp.24927
51. Tsen AR, Long PM, Driscoll HE, Davies MT, Teasdale BA, Penar PL, et al. Triacetin-based acetate supplementation as a chemotherapeutic adjuvant therapy in glioma. Int J Cancer. 2014;134: 1300–1310. doi:10.1002/ijc.28465
52. Long P, Moffett J, Namboodiri A. N-acetylaspartate (NAA) and N-acetylaspartylglutamate (NAAG) promote growth and inhibit differentiation of glioma stem-like cells. J Biol Chem. 2013;288: 26188–26200.
53. Tucker SL, Gharpure K, Herbrich SM, Unruh AK, Nick AM, Crane EK, et al. Molecular biomarkers of residual disease after surgical debulking of high-grade serous ovarian cancer. Clin Cancer Res. 2014;20: 3280–8. doi:10.1158/1078-0432.CCR-14-0445
54. Ye H, Yu T, Temam S, Ziober BL, Wang J, Schwartz JL, et al. Transcriptomic dissection of tongue squamous cell carcinoma. BMC Genomics. 2008;9: 69. doi:10.1186/1471-2164-9-69

55. Kopantzev EP, Monastyrskaya GS, Vinogradova T V., Zinovyeva M V., Kostina MB, Filyukova OB, et al. Differences in gene expression levels between early and later stages of human lung development are opposite to those between normal lung tissue and non-small lung cell carcinoma. Lung Cancer. 2008;62: 23–34. doi:10.1016/j.lungcan.2008.02.011
56. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20: 307–315. doi:10.1093/bioinformatics/btg405
57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47. doi:10.1093/nar/gkv007
58. Sun YA. ABarray: Microarray QA and statistical data analysis for Applied Biosystems Genome Survey Microrarray (AB1700) gene expression data. 2006.
59. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28: 882–3. doi:10.1093/bioinformatics/bts034
60. Gentleman, R., Carey, V., Huber, W., Hahne F. genefilter: genefilter: methods for filtering genes from high-throughput experiments. Available: https://www.bioconductor.org/packages/devel/bioc/manuals/genefilter/man/genefilter.pdf
61. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4: 1184–91. doi:10.1038/nprot.2009.97
62. Rasche A, Al-Hasani H, Herwig R. Meta-analysis approach identifies candidate genes and associated molecular networks for type-2 Diabetes mellitus. BMC Genomics. 2008;9: 310. doi:10.1186/1471-2164-9-310
63. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26: 589–95. doi:10.1093/bioinformatics/btp698
64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–9. doi:10.1093/bioinformatics/btp352
65. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Br Bioinform. 2013;14: 178–192. doi:10.1093/bib/bbs017
66. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: Genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014;30: 284–286.

doi:10.1093/bioinformatics/btt650