

Article

NSD1 Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint

Giuseppina Conteduca ^{1,†}, Davide Cangelosi ^{2,†}, Simona Coco ³, Michela Malacarne ¹, Chiara Baldo ¹, Alessia Arado ¹, Rute Pinto ¹, Barbara Testa ¹ and Domenico A. Coviello ^{1,*}

¹ Laboratory of Human Genetics, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy; giusy_conteduca@alice.it (G.C.); michelamalacarne@gaslini.org (M.M.); chiarabaldo@gaslini.org (C.B.); alessiaarado@gaslini.org (A.A.); ruteluisapinto@gmail.com (R.P.); testabarbara@libero.it (B.T.)

² Clinical Bioinformatics Unit, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy; davidecangelosi@gaslini.org

³ Lung Cancer Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; simona.coco@hsanmartino.it

* Correspondence: domenicocoviello@gaslini.org; Tel.: +39-010-5636-3977

† These authors contributed equally to this work.

Abstract: An increasing amount of evidence indicates the critical role of the *NSD1* gene in Sotos syndrome (SoS), a rare genetic disease, and in tumors. Molecular mechanisms affected by *NSD1* mutations are largely uncharacterized. In order to assess the impact of *NSD1* haploinsufficiency in the pathogenesis of SoS, we analyzed the gene expression profile of fibroblasts isolated from the skin samples of 15 SoS patients and of 5 healthy parents. We identified seven differentially expressed genes and five differentially expressed noncoding RNAs. The most upregulated mRNA was stratifin (*SFN*) (fold change, 3.9, Benjamini–Hochberg corrected $p < 0.05$), and the most downregulated mRNA was goosecoid homeobox (*GSC*) (fold change, 3.9, Benjamini–Hochberg corrected $p < 0.05$). The most upregulated lncRNA was lnc-C2orf84-1 (fold change, 4.28, Benjamini–Hochberg corrected $p < 0.001$), and the most downregulated lncRNA was lnc-C15orf57 (fold change, -0.7 , Benjamini–Hochberg corrected $p < 0.05$). A gene set enrichment analysis reported the enrichment of genes involved in the KRAS and E2F signaling pathways, splicing regulation and cell cycle G2/M checkpoints. Our results suggest that *NSD1* is involved in cell cycle regulation and that its mutation can induce the down-expression of genes involved in tumoral and neoplastic differentiation. The results contribute to defining the role of *NSD1* in fibroblasts for the prevention, diagnosis and control of SoS.

Keywords: G2/M checkpoint; Sotos syndrome; gene expression; *NSD1*; noncoding RNAs



Citation: Conteduca, G.; Cangelosi, D.; Coco, S.; Malacarne, M.; Baldo, C.; Arado, A.; Pinto, R.; Testa, B.; Coviello, D.A. *NSD1* Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint. *Life* **2022**, *12*, 988. <https://doi.org/10.3390/life12070988>

Academic Editors: Sanjeev Gupta and Bijan K. Dey

Received: 26 May 2022

Accepted: 1 July 2022

Published: 2 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The nuclear receptor SET domain protein 1 (*NSD1*) histone methyltransferase has been identified as a protein that interacts with several nuclear receptors, and it may also act as a bifunctional transcriptional cofactor, playing a dual role in transcription [1,2]. The mono- and di-methylation of histone 3 lysine 36 (H3K36) and lysine 168 (H3K168) have been proposed to be the major cellular *NSD1* substrates [3,4]. In addition, *NSD1* (NM 022455.4) can act as a tumor suppressor gene [5,6]. The structure of the *NSD1* protein is characterized by two nuclear interaction domains (NIDs); two proline–tryptophan–tryptophan–proline (PWWP) domains; five plant homeodomains (PHD); an atypical plant homeodomain (C5HCH) finger; and a catalytic domain (CD), which is composed of pre-SET, subset of SET and post-SET-containing proteins. A so-called ‘associated with SET’ (AWS) domain is sometimes found instead of the pre-SET domain, containing Su(var)3–9, enhancer-of-zeste, Trithorax (SET) and a post-SET domain [7].

Germline mutations in the *NSD1* gene cause Sotos syndrome (SoS) (OMIM 117550), a rare genetic disease with a prevalence of 1/14.000 births. It is characterized by overgrowth,

macrocephaly, advanced bone age, characteristic facial features (long face) and intellectual delay [8–10]. SoS is due to an *NSD1* gene haploinsufficiency that can be caused by several different mechanisms (truncating mutations, missense mutations, splice-site mutations, partial gene deletions and 5q35 microdeletions) [11,12]. Intragenic mutations account for 80–85% of SoS cases among European and American populations; 5q35 microdeletions are present in 10–15% of European and American cases and in more than 50% of Japanese cases [11].

In SoS patients bearing germline *NSD1* haploinsufficiency, a deregulation of the mitogen-activated protein kinase (MAPK)/ERK signaling pathway has been observed downstream of KRAS activation, and this has been postulated to contribute to accelerated skeletal outgrowth [13]. Even if different functions of *NSD1* have been documented, the molecular mechanisms that induce the phenotypic characteristics of SoS remain largely unknown.

The Human Genome Project showed that approximately 76% of the human genome is transcribed, with about 70% of transcripts corresponding to noncoding RNAs [14,15]. Noncoding RNAs can be divided into two groups depending on their lengths. Small RNAs are defined as being shorter than 200 nucleotides (nt), which include microRNA, and RNAs longer than 200 nt are referred to as long noncoding RNAs (lncRNAs). lncRNAs and microRNAs have roles in different mechanisms of gene regulation [16–23].

Moreover, somatically acquired alterations are frequently detected in various human cancers [24–34], and RNAi-mediated knock-down seems to increase the proliferation of tumor cells, suggesting the role of *NSD1* as a tumor suppressor [5]. Studies have also described *NSD1* as a potential biomarker for drug resistance in tumors.

In this study, we generated genome-wide transcription profiles for fibroblasts taken from SoS patients and their sex-matched healthy parents in order to identify which lncRNAs, microRNAs, mRNAs and downstream pathways are perturbed by *NSD1* mutations in SoS syndrome. Our results underline the involvement of *NSD1* in the G2/M cell cycle checkpoint, which prevents cells from entering mitosis when DNA is damaged, stopping cell proliferation. We also observed that *NSD1* mutations can induce dysregulation in the expression of noncoding RNA.

2. Materials and Methods

2.1. Patients

This study was conducted with the approval of the Ethical Committee of the Liguria Region (Approval #OG01IGG, 12 July 2021). Written informed consent was obtained from all participants. Fifteen individuals with the classic clinical features of SoS and with molecular diagnoses of *NSD1* pathogenic variants (point mutation or deletion) were enrolled in this study. As controls, five parents with the same sexes as their co-respective SoS patients, selected on the basis of skin biopsy availability, were also enrolled. The details are presented in Tables S1 and S2. The fibroblast cell lines were deposited at the Gaslini Biobank.

2.2. Cell Culture

Skin fibroblasts were obtained via punch biopsy of the forearm skin. Cells were maintained in the RPMI-1640 medium (Life Technologies, Grand Island, NY, USA), supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin (Euroclone S.p.a, Milan, Italy). Cell lines were tested for mycoplasma with a mycoplasma detection kit (Lonza, Basel, CH, Switzerland). In all experiments, cells with between 2 and 15 passages were used.

2.3. Analysis of *NSD1* Pathogenetic Variants

DNA was extracted from the fibroblasts using QIAamp[®] DNA Blood kit (Qiagen, Milan, Italy) according to the manufacturer's protocol. PCR amplifications were performed using platinum-Taq DNA polymerase (Thermo Fisher Scientific, Carlsbad, CA, USA) and specific primers for the 23 different *NSD1* exons (see Table S3). The PCR conditions

were a single denaturation step at 95 °C for 3 min followed by 40 cycles of 95 °C/30 s, 58 °C/30 s and 72 °C (1 min/kb), with a final extension step at 72 °C for 5 min. PCR products were sequenced using the ABI BigDye Terminator Ready Reaction Mix (Thermo Fisher Scientific, Foster City, CA, USA) and analyzed on an ABI 3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. The sequences were aligned with Seqscape analysis software V.2.5 (Thermo Fisher Scientific, Foster City, CA, USA). To verify variants with deletions of the 5q35.3 region, array-CGH was performed with CGH 8 × 60 K (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's protocol. The data were analyzed with the Agilent Cytogenomics 4.0.3.12 software (Agilent Technologies, Santa Clara, CA, USA). All genomic positions were reported according to the human genome assembly (GRCh37/hg19).

2.4. Gene Expression Profiling

The total RNA was extracted using Trizol reagent (Thermo Fisher Scientific, MA, USA). The total RNA from each sample was quantified with the NanoDrop ND-1000, and RNA integrity was assessed by an Agilent 2100 bioanalyzer. Microarray hybridization was performed using the Agilent One Color microarray gene expression kit and the SurePrint G3, 8 × 60 K Human Gene Expression V3 array (Agilent Technologies, Santa Clara, CA, USA) following established protocols. Briefly, cyanine-3-CTP-labeled cRNA was hybridized onto SurePrint G3 microarray chips (Design ID: G4851C), which contained 50,599 probes for 32,776 human mRNAs and 17,438 human lncRNAs; these were derived from authoritative databases, including RefSeq, Ensemble, GenBank and the Broad Institute. After hybridization and washing, processed slides were scanned with a Agilent microarray scanner (Agilent Technologies). Raw data were extracted using Feature Extraction (version 12.0.1.1; Agilent Technologies). Next, data preprocessing, including normalization and filtering, was carried out with the Genespring software (version 14.3; Agilent Technologies). Raw data were normalized by a 75-percentile shift, log₂-transformed and shifted to the median of all samples. The microarray data were deposited in GEO with the accession number GSE204775 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE204775>, accessed on 24 May 2022). Only samples in which quality control was excellent were used for subsequent analyses in order to reduce potential biases introduced by analyzing low-quality specimens.

2.5. Quantitative Real-Time RT-PCR Validation of Microarray Gene Expression Patterns

To confirm the significant modulation of the differentially expressed genes resulting from microarray data analysis, we performed real-time quantitative PCR using gene-specific primer (Table S3) and ran each experiment in triplicate for analysis robustness. Briefly, cDNA was synthesized from 400 ng of the total RNA using the Advantage RT cDNA Kit (Clontech, Mountain View, CA, USA) following the manufacturer's instructions. Specifically, samples were incubated at 42 °C for 90 min, followed by 2 min at 90 °C. Quantitative real-time PCR was performed using LightCycler 480 SYBR Green I Master (Roche Diagnostics, Mannheim, Germany) in a 15 µL reaction mixture. β-actin was used as an endogenous reference to normalize gene expression values with the 2^{-ΔΔCt} method [35]. PCR products were also confirmed via sequence analysis using the ABI BigDye Terminator Ready Reaction Mix (Life Technologies, Carlsbad, CA, USA) and analyzed on an ABI 3130XL genetic analyzer (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol.

2.6. Bioinformatic Analysis

Using a bioinformatic analysis, based on the normalized fluorescence signal values of the lncRNA/mRNA probes, differentially expressed mRNAs and lncRNAs were identified by examining fold changes as well as by *p*-values calculated via Student's *t*-tests between the SoS patients and the healthy controls using the "Statistical Analysis" module from the analysis section, provided by the Genespring software (version 14.3; Agilent Technologies)

with default parameters. The PO value was adjusted with the Benjamini–Hochberg method to account for the increased probability of false positive findings derived from multiple comparisons. A fold change of ≥ 2.0 or ≤ 0.5 and a p -adjusted value of < 0.05 were considered to be significantly up- or downregulated. In silico protein–protein functional interactions among the differentially expressed genes were assessed with the STRING database (<http://stringdb.org>, accessed on 24 May 2022) [36] using default parameters. An extension of the network was analyzed to assess potential indirect interactions between the differentially expressed genes. Using and Graphpad Prism 8 software, a Bonferroni-adjusted significance level of 0.05 was calculated to account for the increased possibility of obtaining false positive results.

A gene set enrichment analysis (GSEA) [37] was used to assess the enrichment of functionally related gene sets in SoS samples [37]. Chemical and genetic perturbations (C2.CGP), hallmark (H), gene ontology biological processes (C5.GO.BP) or gene set collections retrieved from the Molecular Signature Database (MSigDB) v7.4 [38] were used to perform enrichment analyses. GSEA calculated an enrichment score (ES) and a normalized enrichment score (NES) for each gene set and estimated the statistical significance of the NES with an empirical permutation test using 1000 gene permutations in order to obtain the nominal p -value (NOM p -value). When multiple gene sets are evaluated, GSEA adjusts the estimate of the significance level to account for multiple hypothesis testing. To this end, GSEA computed the false discovery rate q -value (FDR q -value), measuring the estimated probability that the normalized enrichment score represented a false positive finding. We considered gene sets containing between 15 and 250 genes. Gene sets with nominal p -values lower than 0.05 and FDR q -values lower than 0.05 were considered significantly enriched.

3. Results

3.1. Comparison of the Gene Expression Profile of Fibroblast Lines between SoS and Control Individuals

Fifteen patients with SoS and five healthy parents were enrolled in the study. The characteristics of the enrolled individuals are reported in Tables S1 and S2. A flowchart summarizing the main steps of these analyses is shown in Figure 1. RNA obtained from dermal fibroblasts from those SoS patients with confirmed NSD1 alterations and their sex-matched controls were analyzed using Agilent SurePrint G3 microarrays. Gene expression profiles were collected into one dataset for subsequent analysis.

Quality control was carried out on all samples to exclude potential artifacts derived from sample preparation and processing. One sample from the SoS patient group was filtered out because it did not pass the conformity test (Supplementary Figure S1). A differential expression analysis identified five significantly differentially expressed noncoding RNAs (four lncRNA and one microRNA) (Table 1) and seven probe sets relative to the mRNAs (Table 2). Four probe sets relative to the noncoding RNAs were upregulated (*MIR646*, *lnc-C2orf84-1*, *lnc-C00665* and *lnc-C20orf197-3*) and one was downregulated (*lnc-C15orf57*, fold change, -0.7 , $p < 0.05$). The most upregulated lncRNA was associated with *lnc-C2orf84-1* (fold change, 4.28, $p < 0.001$; Table 1). Among the probe sets relative to mRNAs, four were upregulated (*NOS3*, *CD19*, *SFN* and *ZNF883*) and three were downregulated (*NDRG2*, *GSC* and *SORBS1*). The most upregulated probe set was associated with the stratifin (*SFN*) gene (fold change, 3.9, $p < 0.05$), and the most downregulated was associated with the goosecoid homeobox (*GSC*) gene (fold change, 3.9, $p < 0.05$; Table 1).

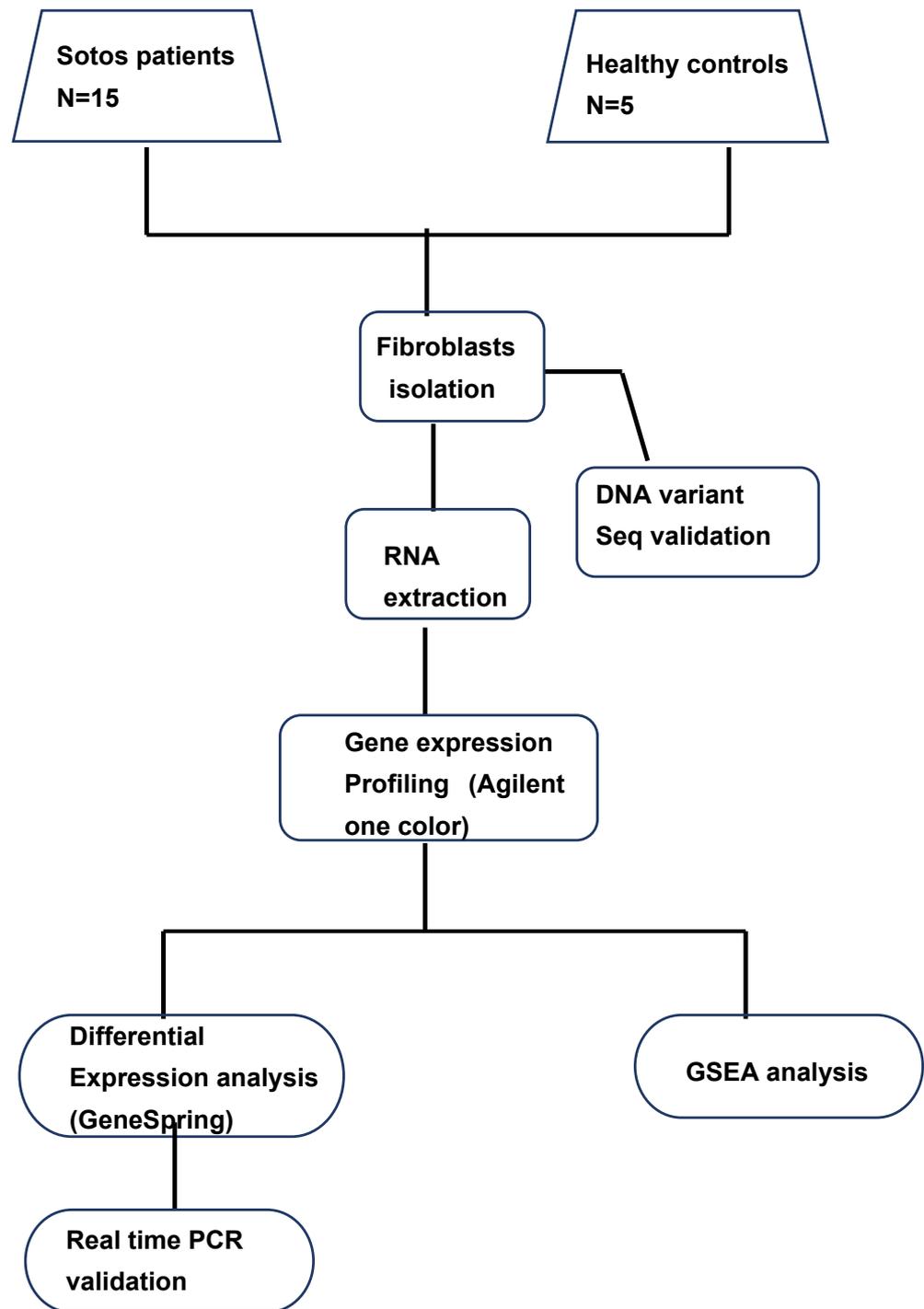


Figure 1. Schematic representation of the whole study strategy used. A GeneSpring differential expression analysis assessed the significant modulation of target genes between the SoS patient group and the healthy control. Differentially expressed genes were validated by real-time PCR. A GSEA pathway analysis identified the most significantly altered biological processes and pathways.

Table 1. Differentially expressed probe sets relative to noncoding RNAs (ncRNAs) in SoS samples.

Nr	Probe Set ID	Gene Symbol	Gene Name	Seq. Name	Log ₂ Fold Change	p-Value	p-Value Adjusted
1	A_22_P00002837	Lnc-C20orf197-3	long intergenic nonprotein-coding RNA	lnc-C20orf197-3:13	3.2	4.95×10^{-7}	0.007
2	A_22_P0000283	Lnc-C2orf84-1	long intergenic nonprotein	lnc-C2orf84-1:1	3.2	4.95×10^{-7}	0.007
3	A_22_P00002715	MIR646HG	MIR646 host gene	ENSG00000228340	2.8	8.64×10^{-6}	0.042
4	A_24_P16214	LINC00665	long intergenic nonprotein-coding RNA 665	ENST00000427868	1.2	2.57×10^{-6}	0.016
5	A_32_P98975	C15orf57	chromosome 15 open reading frame 57	NM_052849	−0.7	1.40×10^{-5}	0.047

Table 2. Significant differentially expressed probe sets relative to mRNAs in SoS samples.

Nr	Probe Set ID	Gene Symbol	Gene Name	Seq. Name	Log ₂ Fold Change	p-Value	p-Value Adjusted
1	A_33_P3389286	SNF	stratifin	NM_006142	3.99	1.62×10^{-6}	0.015
2	A_23_P113572	CD19	CD19 molecule	NM_001770	3.36	1.28×10^{-6}	0.014
3	A_33_P3305790	NOS3	nitric oxide synthase 3 (endothelial cell)	NM_000603	2.4	1.41×10^{-7}	0.0027
4	A_22_P0001791	ZNF883	zinc finger protein 883	NM_001101338	1.33	4.95×10^{-6}	0.026
5	A_33_P3334515	NDRG2	NDRG family member 2	NM_001282213	−1.24	1.06×10^{-5}	0.047
6	A_24_P317907	SORBS1	sorbin and SH3 domain containing 1	NM_001034954	−3.1	1.14×10^{-5}	0.047
7	A_23_P76774	GSC	goosecoid homeobox	NM_173849	−3.98	1.38×10^{-5}	0.047

To assess the biological connection between the differentially expressed genes, we performed a network analysis using the STRING-DB software [36]. We extended the network to assess likely indirect interactions between differentially expressed genes. The resulting network, reported in Figure 2, showed that, statistically, the significantly modulated genes did not display any direct interaction with NSD1, but it did report an interaction with a subset of other gene products, including KRAS and pTEN, which are known to interact with NSD1. These findings indicate that NSD1 is an important regulator of different genes involved in cell differentiation and proliferation.

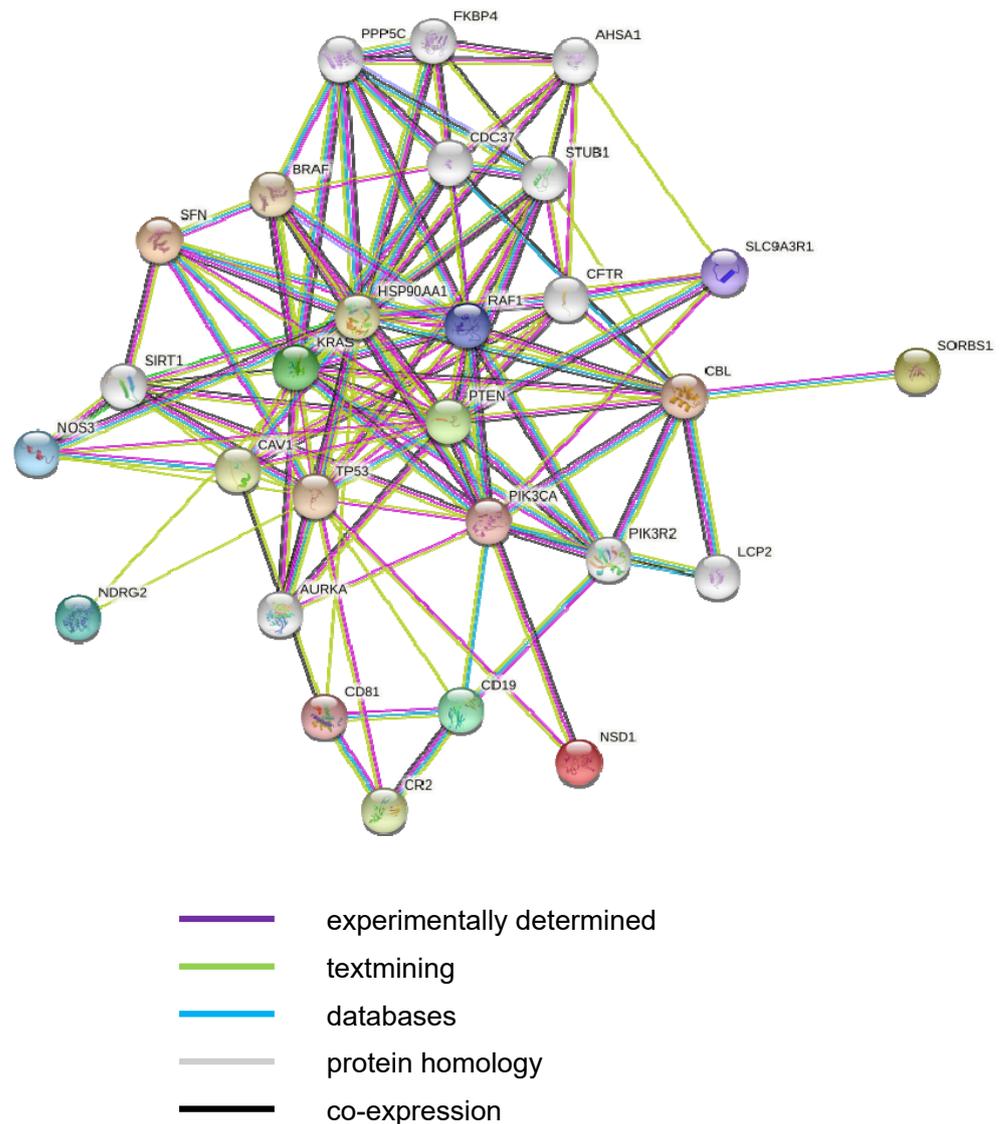
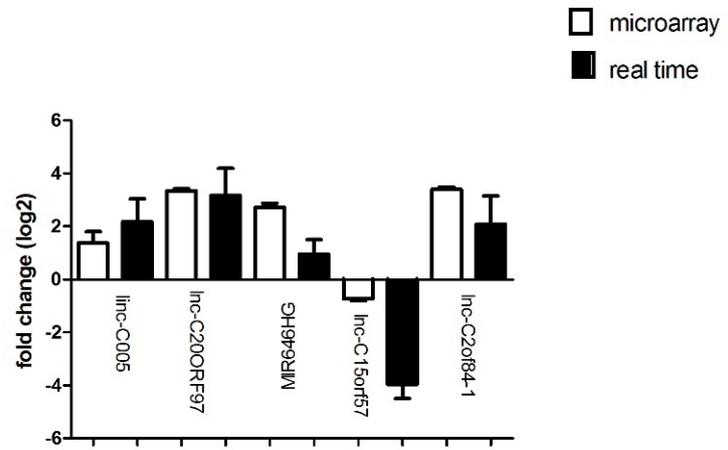


Figure 2. Functional interaction network among differentially expressed genes between SoS patients and healthy controls.

3.2. Validation of Microarray Results by Real-Time Quantitative PCR

All seven genes and six lncRNAs that were found to be significantly modulated in the microarray analysis were subjected to further validation using real-time PCR. Interestingly, the qPCR results confirmed all of the deregulations in the genes and lncRNAs assessed by microarray analysis (Figure 3a,b). In particular, a Student's *t*-test analysis, by applying a 0.05 cutoff on the Benjamini–Hochberg-corrected *p*-values, showed that the expressions of *MIR646* ($p < 0.05$), *lnc-C2orf84-1* ($p < 0.05$), *lnc-C00665* ($p < 0.001$), *lnc-C20orf197-3* ($p < 0.001$), *NOS3* ($p < 0.05$), *ZNF883* ($p < 0.001$), *CD19* ($p < 0.05$) and *SFN* ($p < 0.05$) were upregulated and that those of *GSC* ($p < 0.05$), *lnc-C15orf57* ($p < 0.001$), *NDRG2* ($p < 0.001$) and *SORBS1* ($p < 0.05$) were downregulated in SoS patients.

a



b

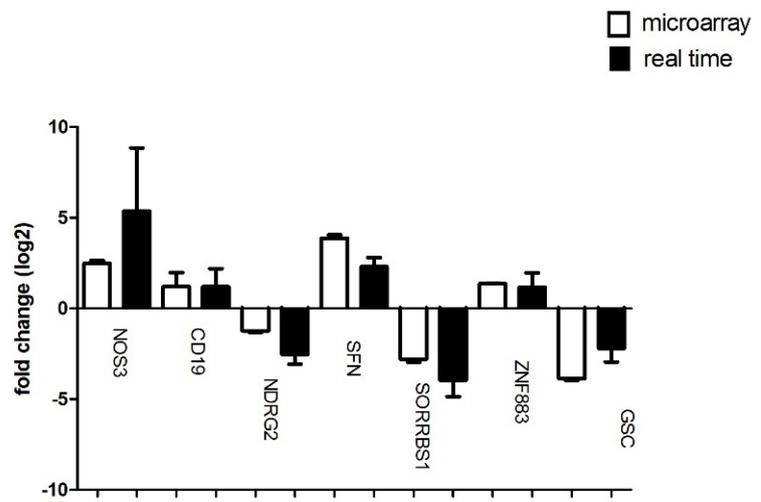
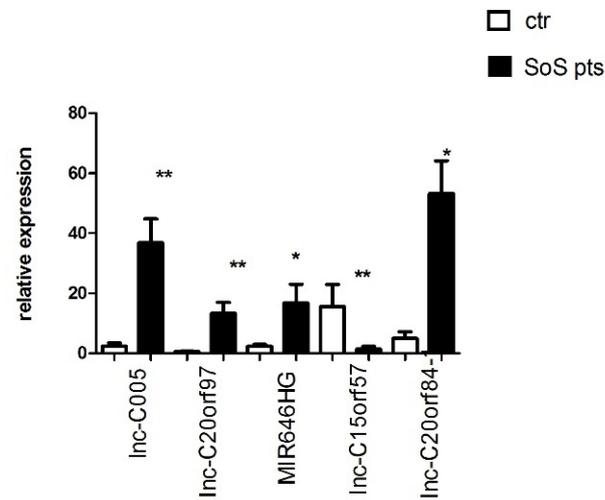


Figure 3. Cont.

c



d

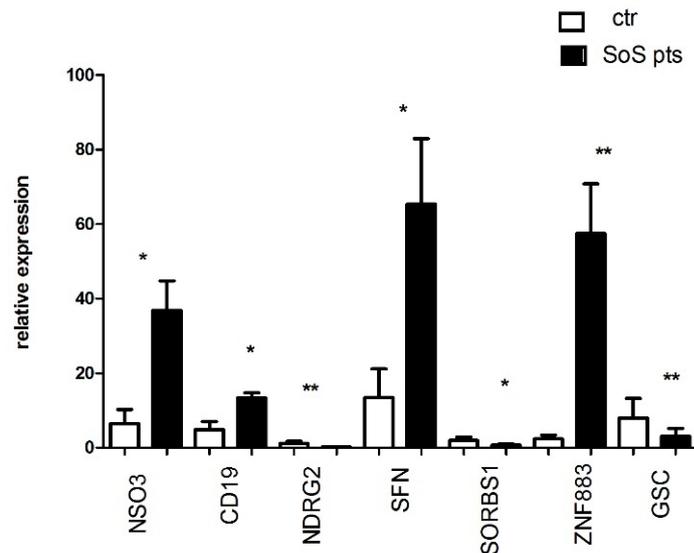


Figure 3. Real-time PCR validation of differentially expressed genes. (a) Quantitative expression measurements of five noncoding RNAs (i.e., *linc-C15orf57*, *MIR646*, *linc-C2orf84-1*, *linc-C00665* and *linc-C20orf197-3*) and (b) seven mRNAs (i.e., *NOS3*, *ZNF883*, *CD19*, *SFN*, *GSC*, *NDRG2* and *SORBS1*). The fold change represents the average difference in expression level of the respective genes between the SoS fibroblasts and the controls. The black bars depict the fold changes detected with the microarray, whereas the white bars show the average fold change with respect to the triplicate real-time PCR assessments. The fold changes indicating a downregulated expression are represented with negative values. The error bars represent the standard error of the mean. Significant levels were at $p < 0.05$. (c) Compared to healthy controls, five noncoding RNAs (*linc-C005*, *linc-C20orf97*, *MIR646HG*, *linc-C15orf57* and *linc-C20orf84-1*) and (d) seven mRNAs (*NOS3*, *ZNF883*, *CD19*, *SFN*, *NDRG2*, *SORBS1* and *GSC*) were selected. The black bars depict healthy control (Ctr), and the white bars represent the Sotos syndrome patients (SoS pts). The results were consistent with the findings obtained from the microarray analysis. Data are presented as the mean \pm standard deviation. * $p < 0.05$, ** $p < 0.01$.

3.3. Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) is a computational technique that assesses the coordinated gene expression modulation of functionally related genes (gene sets) between two groups [37]. To perform this analysis, we used gene sets included in the chemical and genetic perturbation, hallmark (H) and C2 gene ontology collections from the MSigDB database [38]. The analysis showed the 89 significantly enriched biological processes and pathways in the NSD1-mutated fibroblasts of SoS patients, as compared to the healthy controls; Table S4 reports the complete list of significant processes and pathways. The most statistically significant over-represented gene sets were related to the cell cycle and proliferation, cell differentiation, P53-mediated cell cycle arrest, cellular senescence and cancer (Table 3). In particular, upregulated genes were mainly involved in the regulation of the meiotic cell cycle, the negative regulation of nuclear division (Supplementary Figure S2A,B) and kinetochore organization. The most statistically significant underrepresented gene sets were related to the E2F pathway (Supplementary Figure S2), pediatric cancer markers, neoplastic transformation via the STAT3 pathway, MYC and TFRC targets, the cell cycle G2/M checkpoint, TNFA-signaling via NFkB, epithelial mesenchymal transition, the apoptosis process caused by CDKN1A via TP53 and neoplastic transformation KRAS signaling pathways (Supplementary Figure S3A,B).

Table 3. Selected significantly enriched gene sets in SoS samples found via GSEA.

Pathway or Process Description ^a	Number of Genes ^b	FDR q -Value ^c	Type of Regulation ^d	Ontology ^e
Cancer Process				
Florio neocortex basal radial glia DN	151	0.00	DOWN	CGP
Kong E2F3 targets	77	0.00	DOWN	CGP
Kobayashi EGFR-signaling 24 h DN	213	0.00	DOWN	CGP
Whitfield cell cycle literature	39	0.00	DOWN	CGP
Whiteford pediatric cancer markers	102	0.00	DOWN	CGP
Nakayama soft-tissue tumors PCA2 UP	76	0.00	DOWN	CGP
Zhan multiple myeloma pr UP	35	0.00	DOWN	CGP
Pyeon HPV positive tumors UP	79	0.00	DOWN	CGP
Sotiriou breast cancer grade 1 vs. 3 UP	127	3.65×10^{-4}	DOWN	CGP
Chiang liver cancer subclass proliferation UP	146	8.15×10^{-4}	DOWN	CGP
Farmer breast cancer cluster 2	29	0.001	DOWN	CGP
Villanueva liver cancer KRT19 UP	147	0.002	DOWN	CGP
Rodrigues thyroid carcinoma DN	66	0.002	DOWN	CGP
Finetti breast cancer kinome red	15	0.002	DOWN	CGP
Riggi Ewing sarcoma progenitor DN	158	0.002	DOWN	CGP
Azare neoplastic transformation by STAT3 DN	16	0.002	DOWN	CGP
West adrenocortical tumor markers UP	103	0.005	DOWN	CGP
Li Wilms tumor anaplastic UP	15	0.010	DOWN	CGP
Li prostate cancer epigenetic	27	0.014	DOWN	CGP
GAL leukemic stem cell DN	188	0.014	DOWN	CGP
Rickman head and neck cancer B	37	0.018	DOWN	CGP
Poola invasive breast cancer UP	228	0.019	DOWN	CGP
Sengupta nasopharyngeal carcinoma UP	240	0.024	DOWN	CGP
SMID breast cancer luminal A UP	68	0.025	DOWN	CGP
Vantveer breast cancer BRCA1 UP	28	0.027	DOWN	CGP
Winnepenninckx melanoma metastasis UP	134	0.029	DOWN	CGP
Ferreira Ewing's sarcoma unstable vs. stable UP	131	0.029	DOWN	CGP
Lopes methylated in colon cancer UP	22	0.047	DOWN	CGP
Chiaroni neoplastic transformation KRAS UP	113	0.047	DOWN	CGP

Table 3. Cont.

Pathway or Process Description ^a	Number of Genes ^b	FDR q -Value ^c	Type of Regulation ^d	Ontology ^e
Cell Cycle and Proliferation Process				
Fischer G2 M cell cycle	191	0.00	DOWN	CGP
Croonquist NRAS signaling DN	55	0.00	DOWN	CGP
Molenaar targets of CCND1 and CDK4 DN	48	0.00	DOWN	CGP
Ishida E2F targets	39	0.00	DOWN	CGP
Reichert mitosis LIN9 targets	21	0.00	DOWN	CGP
Graham normal quiescent vs. normal dividing DN	74	1.15×10^{-4}	DOWN	CGP
Odonnell TFRC targets DN	110	2.18×10^{-4}	DOWN	CGP
Graham cml dividing vs. normal quiescent UP	157	0.002	DOWN	CGP
Eguchi cell cycle RB1 targets	17	0.002	DOWN	CGP
Yu MYC targets UP	40	0.003	DOWN	CGP
Zhou cell cycle genes in IR response 24HR	103	0.003	DOWN	CGP
Odonnell targets of MYC and TFRC DN	36	0.006	DOWN	CGP
Alcalay AML by NPM1 localization DN	158	0.010	DOWN	CGP
Zhou cell cycle genes in IR response 6HR	70	0.011	DOWN	CGP
Plasari TGFB1 targets 10HR DN	205	0.014	DOWN	CGP
Chicas RB1 targets growing	196	0.02	DOWN	CGP
ULE-splicing via NOVA2	35	0.02	DOWN	CGP
Benporath proliferation	124	0.033	DOWN	CGP
Whitfield cell cycle G2 M	162	0.034	DOWN	CGP
Liang silenced by methylation 2	50	0.034	DOWN	CGP
Nojima SFRP2 targets DN	19	0.034	DOWN	CGP
Kamminga EZH2 targets	37	0.037	DOWN	CGP
Graham CML quiescent_VS normal quiescent UP	74	0.040	DOWN	CGP
GOBP heat generation	15	0.014	UP	GO BP
GOBP regulation of meiotic cell cycle	31	0.07	UP	GO BP
GOBP negative regulation of nuclear division	42	0.016	UP	GO BP
GOBP regulation of feeding behavior	20	0.021	UP	GO BP
GOBP positive regulation of organic acid transport	29	0.017	UP	GO BP
GOBP regulation of nuclear division organization	106	0.046	UP	GO BP
GOBP kinetochore organization	16	0.041	UP	GO BP
HALLMARK G2M checkpoint	172	0.001	DOWN	H
HALLMARK KRAS-signaling UP	174	0.0001	DOWN	H
HALLMARK mitotic spindle	163	0.008	DOWN	H
HALLMARK inflammatory response	159	0.006	DOWN	H
HALLMARK E2F targets	167	0.005	DOWN	H
HALLMARK apoptosis	130	0.017	DOWN	H
HALLMARK KRAS-signaling DN	166	0.032	DOWN	H
HALLMARK allograft rejection	160	0.036	DOWN	H
HALLMARK TNFA-signaling via NFkB	163	0.040	DOWN	H
Cell Differentiation Process				
Boquest stem cell DN	186	0.05	DOWN	CGP
Sarrío epithelial mesenchymal transition UP	149	0.006	DOWN	CGP
Le neuronal differentiation DN	16	0.034	DOWN	CGP
P53-Mediated Cell Cycle Arrest and Cellular Senescence Process				
Wu apoptosis by CDKN1A via TP53	47	0.002	DOWN	CGP
Tang senescence TP53 targets DN	48	0.00	DOWN	CGP

a. Official name of the biological process. b. Number of genes in the biological process. c. The FDR q -value estimates the significance of the enrichment of a biological process or a pathway. FDR q -values ≤ 0.05 are considered acceptable. d. Type of regulation of the genes involved in a process or a pathway. e. Name of the ontology defining a biological process or pathway. GO BP: gene ontology biological process; H: hallmark pathways; CGP: chemical and genetic perturbations pathways.

Then, an interaction network was generated using the STRING database. Nodes represent gene products, and edges represent protein–protein associations. Only the associations with an evidence score higher than 0.3 are shown; their colors indicate different kinds of interaction evidence (key, bottom right).

4. Discussion

SoS is an autosomal dominant disorder with manifestations characterized by tall stature, facial dysmorphism and mental retardation [1–5].

Currently, the pathogenetic cause of SoS is thought to be due to haploinsufficiency in the *NSD1* gene [6]. However, the molecular mechanisms of action are still unknown. At present, an increasing number of studies have indicated that *NSD1* is involved in the regulation of a wide variety of biological processes [27–32]. The *NSD1* gene is expressed in most tissues from different organisms. Elevated *NSD1* levels are detected in normal brains, pancreases and male reproductive tracts, in hematopoietic organs, bone marrow and lymphoid tissues [9]. Significant *NSD1* mRNA expression in bone marrow polymorphonuclear cells, CD4, CD8 and NK has also been reported by genevisible.com [10].

NSD1 defects induce the alternate methylation of H3K36, blocking cellular differentiation and promoting oncogenesis [24]. Reduced *NSD1* activity has been observed in head and neck cell cancers, while abnormal DNA-promoter hypermethylation is associated with renal clear-cell carcinoma [25,27–30]. Furthermore, *NSD1* mutations are present in hematological malignancies, and a recurrent t(5;11)(q35;p15) chromosomal translocation is associated with aggressive pediatric acute myeloid leukemia (AML). The resulting fusion protein with leukemogenic activity is composed of the N-terminus of nucleopore 98 (NUP98) and the C-terminal *NSD1* SET domain [32–34]. The abnormal expression of *NSD1* in different cancers has also been detected, which is associated with tumorigenesis, survival and chemoresistance. In pancreatic ductal adenocarcinoma, a high level of *NSD1* is significantly correlated with the advanced clinical stage [39]. Moreover, *NSD1* gene-silencing inhibits the proliferative, migratory and invasive abilities of hepatocellular carcinoma cells [40]. Furthermore, upregulation of *NSD1* might improve oncogenic initiation through the reinforced methylation of H3K36 [41].

In the present study, the expression of lncRNAs and mRNAs in SoS patients and the healthy control was investigated using a microarray analysis to reveal their potential role in the pathogenesis of SoS. We decided to compare the expression profiles of SoS patients with those of their healthy parents, as they share approximately 50% of their DNA; this allowed us to minimize differences caused by other genes variants with respect to gene expression [42,43].

In total, five noncoding RNAs and seven mRNAs were identified as being differentially expressed. GSEA analyses were used to explore the possible biological functions and potential mechanisms of the mRNAs and noncoding RNAs in SoS.

The results of the present study demonstrated that these biological processes, including G2/M checkpoints, can control transitions between cell cycle phases, cell senescence, and meiotic and mitotic division and are among the most significantly enriched mRNAs in SoS. Most of these functions are involved in neoplastic disease and tumor development, and this is consistent with previous studies [24–34,39–41]. Furthermore, we observed that the KRAS signaling pathway, the E2F target and apoptosis by CDKN1A via TP53-signaling were remarkably downregulated in the SoS samples compared to the healthy controls. On the other hand, genes involved in the regulation of nuclear division, meiotic cell cycle and kinetochore organization were remarkably upregulated in SoS patients.

lncRNAs have long been considered the only transcriptional noise; however, several lncRNAs analyzed with microarrays showed greater expression in immortalized lymphocytes [44]. lncRNAs were also found to be upregulated in neuroblastoma cell line LAN-I [45,46]. Different studies have shown that lncRNA alterations contribute to neuronal and neurodegenerative disease development [19,47,48]. lncRNAs may provide basic

information that could be used to understand pathways related to the disease course of SoS and to find more effective targeted therapies.

The results of the present study revealed that, among the differentially expressed genes, SFN was the most upregulated mRNA (FC = 3.9) in SoS samples. The SFN protein is ubiquitously expressed and exerts many relevant intracellular functions, such as the control of cell cycle and apoptosis, the regulation of signal transduction pathways, cellular trafficking, cell proliferation and differentiation, and protein folding and processing, among others [49]. Both p53 and SFN prevent DNA errors during mitosis [40–52]. Therefore, according to the reported evidence, our study suggests that NSD1 may be an important protein for cell cycle and mitotic translation, regulating SFN G2/M checkpoint expression.

In addition, in our study, *GSC*, *SORBS1* and *NDRG2* mRNA were down-expressed in Sotos patients compared to the healthy controls. These three genes encode proteins that are associated with the actin cytoskeleton, the formation of actin stress fibers and focal adhesions, and receptor tyrosinase kinase-signaling, and they control different aspects of morphogenesis and cell differentiation [53–55]. Some studies have reported that the expression of *GSC* inhibits the proliferation of PC12 cells and that—as with *NSD1*—*GSC* mediates the regulation of erythropoiesis functions [56]. *NDRG2* is frequently downregulated in cancer, and it plays an important role in the control of tumor growth and metastasis [57]. *SORBS1* expression has been detected in colorectal cancer cell lines, and its overexpression increases the proliferation and migration abilities of tumor cells [58].

We found a downregulation of the *SORBS1*, *NDRG2* and *GSC* genes in the SoS samples, similar to the RNA expression profile of the mesenchymal cell. The down-expression of these three genes induces the epithelial–mesenchymal transition [59,60].

Furthermore, *GSC* acts in mesenchyme-derived tissues during craniofacial development, and it has been reported that mice that are homozygous for the deletion of the homeobox gene *gooseoid* can have multiple craniofacial defects, and various bone and cartilage malformations [61,62]. Therefore, we propose that *GSC* down-expression may contribute to the advanced bone age and characteristic facial features of Sotos syndrome, including macrocephaly and long face.

Epilepsy is one of the clinical manifestations found in SoS patients. We found an over-expression of *ZNF833*, a gene that is correlated with the progression of epilepsy in rats [63].

In conclusion, we propose that *NSD1* may participate in Sotos syndrome's mechanism of action by regulating the function of lncRNA, inducing the down-expression of *GSC*, *NDRG2* and *SORBS1* and the up-expression of *SFN* and *ZNF883*.

These expression signatures may be useful tools for screening and monitoring the disease and for predicting its prognosis.

Additional investigations should be performed to establish and clarify the detailed molecular mechanism of action behind the lncRNA-mediated regulation of potential coding genes in SoS. Further studies with a larger sample size should also be performed to verify the relevant results.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/life12070988/s1>, Figure S1: Quality control evaluation sample 19; Figure S2: Enrichment plots of representative over represented biological processes in SoS fibroblasts samples by GSEA; Figure S3: Enrichment plots of representative down represented of C2,CGP and hallmark gene sets in SoS fibroblasts samples by GSEA; Table S1: Sotos patients enrolled; Table S2: Healthy parents enrolled; Table S3: Primers used for validation of microarray results by real-time PCR; Table S4: GSEA analysis.

Author Contributions: Conceptualization, D.A.C.; methodology, G.C., C.B. and A.A.; software, D.C.; validation, D.C. and M.M.; formal analysis, G.C. and R.P.; investigation, G.C.; resources, D.A.C. and C.B.; data curation, D.C. and S.C.; writing—original draft preparation, G.C. and D.C.; writing—review and editing, D.A.C.; visualization, S.C. and B.T.; supervision, D.A.C.; project administration, G.C. and D.C.; funding acquisition, D.A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the AssiGulliver Italian Association; Fondazione Sardegna; Italian Ministry of Health, Ricerca Corrente 2020 and Ricerca Corrente 2021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This study was approved by the Ethical Committee of the Liguria Region (Approval #OG01IGG on 12 July 2021), and written informed consent was provided by patients and/or their parents or legal guardians.

Data Availability Statement: The gene expression profiles of the patients and healthy controls are accessible in the Gene Expression Omnibus repository (GEO accession: GSE204775).

Acknowledgments: We thank the Biobank of the Laboratory of Human Genetics of the Giannina Gaslini Institute, Genoa (Italy) for the cell line provided and the patients for their participation in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, N.; Vom Baur, J.M.; Garnier, T.; Lerouge, J.; Vonesch, L.; Lutz, Y.; Chambon, P.; Losson, R. Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. *EMBO J.* **1998**, *17*, 3398–3412. [[CrossRef](#)] [[PubMed](#)]
2. Wang, X.; Yeh, S.; Wu, G.; Hsu, C.L.; Wang, L.; Chiang, T.; Yang, Y.; Guo, Y.; Chang, C. Identification and characterization of a novel androgen receptor coregulator ARA267- α in prostate cancer cells. *J. Biol. Chem.* **2001**, *276*, 40417–40423. [[CrossRef](#)] [[PubMed](#)]
3. Wagner, E.J.; Carpenter, P.B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 115–126. [[CrossRef](#)]
4. Kudithipudi, S.; Lungu, C.; Rathert, P.; Happel, N.; Jeltsch, A. Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chem. Biol.* **2014**, *21*, 226–237. [[CrossRef](#)] [[PubMed](#)]
5. Berdasco, M.; Ropero, S.; Setien, F.; Fraga, M.; Lapunzina, P.; Losson, R.; Alaminos, M.; Cheung, N.-K.; Rahman, N.; Esteller, M. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21830–21835. [[CrossRef](#)]
6. Bui, N.; Huang, J.K.; Bojorquez-Gomez, A.; Licon, K.; Sanchez, K.S.; Tang, S.N.; Beckett, A.N.; Wang, T.; Zhang, W.; Shen, J.P.; et al. Disruption of NSD1 in Head and Neck Cancer Promotes Favorable Chemotherapeutic Responses Linked to Hypomethylation. *Mol. Cancer Ther.* **2018**, *17*, 1585–1594. [[CrossRef](#)]
7. Amin, N.; Nietlispach, D.; Qamar, S.; Coyle, J.; Chiarparin, E.; Williams, G. NMR backbone resonance assignment and solution secondary structure determination of human NSD1 and NSD2. *Biomol. NMR Assign.* **2016**, *10*, 315–320. [[CrossRef](#)]
8. Tatton-Brown, K.; Douglas, J.; Coleman, K.; Baujat, G.; Cole, T.R.; Das, S.; Horn, D.; Hughes, H.E.; Temple, I.K.; Faravelli, F.; et al. Genotype-phenotype associations in Sotos syndrome: An analysis of 266 individuals with NSD1 aberrations. *Am. J. Hum. Genet.* **2005**, *77*, 193–204. [[CrossRef](#)]
9. The Human Protein Atlas. Available online: <https://www.proteinatlas.org/ENSG00000165671-NSD1/tissue> (accessed on 13 July 2021).
10. GeneCards: The Human Gene Database: NSD1. Available online: https://www.genecards.org/cgi-bin/carddisp.pl?gene=NSD1#protein_expression (accessed on 13 July 2021).
11. Tatton-Brown, K.; Rahman, N. Sotos syndrome. *Eur. J. Hum. Genet.* **2007**, *15*, 264–271. [[CrossRef](#)]
12. Tatton-Brown, K.; Cole, T.R.P.; Rahman, N. Sotos syndrome. In *GeneReviews*[®] [Internet]; Pagon, R.A., Adam, M.P., Ardinger, H.H., Wallace, S.E., Bean, L.J.H., Mefford, H.C., Gripp, K.W., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 2014; pp. 1993–2017.
13. Visser, R.; Landman, E.B.; Goeman, J.; Wit, J.M.; Karperien, M. Sotos syndrome is associated with deregulation of the MAPK/ERK-signaling pathway. *PLoS ONE* **2012**, *7*, e49229. [[CrossRef](#)]
14. Lee, H.; Zhang, Z.; Krause, K.M. Long Noncoding RNAs and Repetitive Elements: Junk or Intimate Evolutionary Partners? *Trends Genet.* **2019**, *35*, 892–902. [[CrossRef](#)] [[PubMed](#)]
15. Kashi, K.; Henderson, L.; Bonetti, A.; Carninci, P. Discovery and functional analysis of lncRNAs: Methodologies to investigate an uncharacterized transcriptome. *Biochim. Biophys. Acta* **2016**, *1859*, 3–15. [[CrossRef](#)] [[PubMed](#)]
16. Hangauer, M.J.; Vaughn, I.W.; McManus, M.T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS. Genet.* **2013**, *6*, e1003569. [[CrossRef](#)] [[PubMed](#)]
17. Li, C.J.; Hong, T.; Tung, Y.T.; Yen, Y.P.; Hsu, H.C.; Lu, Y.L.; Chang, M.; Nie, Q.; Chen, J.A. MicroRNA filters Hox temporal transcription noise to confer boundary formation in the spinal cord. *Nat. Commun.* **2017**, *8*, 14685. [[CrossRef](#)]
18. Hoye, M.L.; Koval, E.D.; Wegener, A.J.; Hyman, T.S.; Yang, C.; O'Brien, D.R.; Miller, R.L.; Cole, T.; Schoch, K.M.; Shen, T.; et al. MicroRNA Profiling Reveals Marker of Motor Neuron Disease in ALS Models. *J. Neurosci.* **2017**, *37*, 5574–5586. [[CrossRef](#)] [[PubMed](#)]
19. Wang, C.; Wang, L.; Ding, Y.; Lu, X.; Zhang, G.; Yang, J.; Zheng, H.; Wang, H.; Jiang, Y.; Xu, L. LncRNA Structural Characteristics in Epigenetic Regulation. *Int. J. Mol. Sci.* **2017**, *18*, 2659. [[CrossRef](#)]

20. Engreitz, J.M.; Haines, J.E.; Perez, E.M.; Munson, G.; Chen, J.; Kane, M.; McDonel, P.E.; Guttman, M.; Lander, E.S. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **2016**, *539*, 452–455. [[CrossRef](#)]
21. Vangoor, V.R.; Gomes-Duarte, A.; Pasterkamp, R.J. Long non-coding RNAs in motor neuron development and disease. *J. Neurochem.* **2021**, *156*, 777–801. [[CrossRef](#)]
22. Anderson, D.M.; Anderson, K.M.; Chang, C.L.; Makarewich, C.A.; Nelson, B.R.; McAnally, J.R.; Kasaragod, P.; Shelton, J.M.; Liou, J.; Bassel-Duby, R.; et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **2015**, *160*, 595–606. [[CrossRef](#)]
23. Liu, L.; Fang, F. Long Noncoding RNA Mediated Regulation in Human Embryogenesis, Pluripotency, and Reproduction. *Stem Cells Int.* **2022**, *2022*, 8051717. [[CrossRef](#)]
24. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **2015**, *517*, 576–582. [[CrossRef](#)] [[PubMed](#)]
25. Bianco-Miotto, T.; Chiam, K.; Buchanan, G.; Jindal, S.; Day, T.K.; Thomas, M.; Pickering, M.A.; O’Loughlin, M.A.; Ryan, N.K.; Raymond, W.A.; et al. Global Levels of Specific Histone Modifications and an Epigenetic Gene Signature Predict Prostate Cancer Progression and Development. *Cancer Epidemiol. Biomark. Prev.* **2010**, *19*, 2611–2622. [[CrossRef](#)] [[PubMed](#)]
26. Spurr, L.; Li, M.; Alomran, N.; Zhang, Q.; Restrepo, P.; Movassaghi, M.; Trenkov, C.; Tunnessen, N.; Apanasovich, T.; Crandall, K.A.; et al. Systematic pan-cancer analysis of somatic allele frequency. *Sci. Rep.* **2018**, *8*, 7735. [[CrossRef](#)] [[PubMed](#)]
27. Su, X.; Zhang, J.; Mouawad, R.; Comp erat, E.; Roup r t, M.; Allanic, F.; Malouf, G.G. NSD1 Inactivation and SETD2 Mutation Drive a Convergence toward Loss of Function of H3K36 Writers in Clear Cell Renal Cell Carcinomas. *Cancer Res.* **2017**, *77*, 4835–4845. [[CrossRef](#)] [[PubMed](#)]
28. Mohammad, F.; Helin, K. Oncohistones: Drivers of pediatric cancers. *Genes Dev.* **2017**, *31*, 2313–2324. [[CrossRef](#)]
29. Schuhmacher, M.; Kusevic, D.; Kudithipudi, S.; Jeltsch, A. Kinetic Analysis of the Inhibition of the NSD1, NSD2 and SETD2 Protein Lysine Methyltransferases by a K36M Oncohistone Peptide. *ChemistrySelect* **2017**, *2*, 9532–9536. [[CrossRef](#)]
30. Farhangdoost, N.; Horth, C.; Hu, B.; Bareke, E.; Chen, X.; Li, Y.; Coradin, M.; Garcia, B.A.; Lu, C.; Majewski, J. Chromatin dysregulation associated with NSD1 mutation in head and neck squamous cell carcinoma. *Cell Rep.* **2021**, *34*, 108769. [[CrossRef](#)]
31. Rajagopalan, K.N.; Chen, X.; Weinberg, D.N.; Chen, H.; Majewski, J.; Allis, C.D.; Lu, C. Depletion of H3K36me2 recapitulates epigenomic and phenotypic changes induced by the H3.3K36M oncohistone mutation. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021795118. [[CrossRef](#)]
32. Jaju, R.J. A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* **2001**, *98*, 1264–1267. [[CrossRef](#)]
33. Ntziachristos, P.; Abdel-Wahab, O.; Aifantis, I. Emerging concepts of epigenetic dysregulation in hematological malignancies. *Nat. Immunol.* **2016**, *17*, 1016–1024. [[CrossRef](#)]
34. Tauchmann, S.; Almosailekh, M.; Schwaller, J. NSD1 in erythroid differentiation and leukemogenesis. *Mol. Cell Oncol.* **2020**, *7*, 1809919. [[CrossRef](#)] [[PubMed](#)]
35. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))}. *Method* **2001**, *25*, 402–408. [[CrossRef](#)]
36. Jensen, L.J.; Kuhn, M.; Stark, M.; Chaffron, S.; Creevey, C.; Muller, J.; Doerks, T.; Julien, P.; Roth, A.; Simonovic, M.; et al. STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **2009**, *37*, D412–D416. [[CrossRef](#)] [[PubMed](#)]
37. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)] [[PubMed](#)]
38. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsd ttir, H.; Tamayo, P.; Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **2011**, *27*, 1739–1740. [[CrossRef](#)] [[PubMed](#)]
39. Ettl, M.; Zhao, L.; Schechter, S.; Shi, J. Expression and prognostic value of NSD1 and SETD2 in pancreatic ductal adenocarcinoma and its precursor lesions. *Pathology* **2019**, *51*, 392–398. [[CrossRef](#)] [[PubMed](#)]
40. Zhang, S.; Zhang, F.; Chen, Q.; Wan, C.; Xiong, J.; Xu, J. CRISPR/Cas9-mediated knockout of NSD1 suppresses the hepatocellular carcinoma development via the NSD1/H3/Wnt10b signaling pathway. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 467. [[CrossRef](#)]
41. Lucio-Eterovic, A.K.; Singh, M.M.; Gardner, J.E.; Veerappan, C.S.; Rice, J.C.; Carpenter, P.B. Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16952–16957. [[CrossRef](#)]
42. Miko, I. Gregor Mendel and the principles of inheritance. *Nat. Educ.* **2008**, *1*, 134.
43. Chial, H. Mendelian genetics: Patterns of inheritance and single-gene disorders. *Nat. Educ.* **2008**, *1*, 63. [[CrossRef](#)]
44. Ronchetti, D.; Manzoni, M.; Agnelli, L.; Vinci, C.; Fabris, S.; Cutrona, G.; Matis, S.; Colombo, M.; Galletti, S.; Taiana, E.; et al. lncRNA profiling in early-stage chronic lymphocytic leukemia identifies transcriptional fingerprints with relevance in clinical outcome. *Blood Cancer J.* **2016**, *6*, e468. [[CrossRef](#)] [[PubMed](#)]
45. Hu, Y.; Sun, H.; Hu, J.; Zhang, X. lncRNA DLX6-AS1 Promotes the Progression of Neuroblastoma by Activating STAT2 via Targeting miR-506-3p. *Cancer Manag. Res.* **2020**, *12*, 7451–7463. [[CrossRef](#)] [[PubMed](#)]
46. Utnes, P.; L kke, C.; Fl egstad, T.; Einvik, C. Clinically Relevant Biomarker Discovery in High-Risk Recurrent Neuroblastoma. *Cancer Inform.* **2019**, *18*, 1176935119832910. [[CrossRef](#)] [[PubMed](#)]

47. Ang, C.E.; Trevino, A.E.; Chang, H.Y. Diverse lncRNA mechanisms in brain development and disease. *Curr. Opin. Genet. Dev.* **2020**, *65*, 42–46. [[CrossRef](#)] [[PubMed](#)]
48. Deng, Q.; Ma, L.; Chen, T.; Yang, Y.; Ma, Y.; Ma, L. NF- κ B 1-induced LINC00665 regulates inflammation and apoptosis of neurons caused by spinal cord injury by targeting miR-34a-5p. *Neurol Res.* **2021**, *43*, 418–427. [[CrossRef](#)]
49. Khalil, A.M.; Guttman, M.; Huarte, M.; Garber, M.; Raj, A.; Rivea Morales, D.; Thomas, K.; Presser, A.; Bernstein, B.E.; Van Oudenaarden, A.; et al. Many human large intergenic noncoding RNAs associate with chromatin modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11667–11672. [[CrossRef](#)] [[PubMed](#)]
50. Mattick, J.S.; Gagen, M.J. The evolution of controlled multitasked gene networks: The role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.* **2001**, *18*, 1611–1630. [[CrossRef](#)]
51. Popadin, K.; Gutierrez Arcelus, M.; Dermitzakis, E.T.; Antonarakis, S.E. Genetic and epigenetic regulation of human lincRNA gene expression. *Am. J. Hum. Genet.* **2013**, *93*, 1015–1026. [[CrossRef](#)]
52. Luk, S.C.; Siu, S.W.; Lai, C.K.; Wu, Y.J.; Pang, S.F. Cell Cycle Arrest by a Natural Product via G2/M Checkpoint. *Int. J. Med. Sci.* **2005**, *2*, 64–69. [[CrossRef](#)]
53. Sawada, K.; Konishi, Y.; Tominaga, M.; Watanabe, Y.; Hirano, J.; Inoue, S.; Kageyama, R.; Blum, M.; Tominaga, A. Goosecoid suppresses cell growth and enhances neuronal differentiation of PC12 cells. *J. Cell Sci.* **2000**, *113*, 2705–2713. [[CrossRef](#)]
54. Foletta, V.C.; Prior, M.J.; Stupka, N.; Carey, K.; Segal, D.H.; Jones, S.; Swinton, C.; Martin, S.; Cameron-Smith, D.; Walder, K.R. NDRG2, a novel regulator of myoblast proliferation, is regulated by anabolic and catabolic factors. *J. Physiol.* **2009**, *587 Pt 7*, 1619–1634. [[CrossRef](#)]
55. Cho, W.C.; Jang, J.E.; Kim, K.H.; Yoo, B.C.; Ku, J.L. SORBS1 serves a metastatic role via suppression of AHNAK in colorectal cancer cell lines. *Int. J. Oncol.* **2020**, *56*, 1140–1151. [[CrossRef](#)] [[PubMed](#)]
56. Konishi, Y.; Tominaga, M.; Watanabe, Y.; Imamura, F.; Goldfarb, A.; Maki, R.; Blum, M.; De Robertis, E.M.; Tominaga, A. GOOSECOID inhibits erythrocyte differentiation by competing with Rb for PU.1 binding in murine cells. *Oncogene* **1999**, *18*, 6795–6805. [[CrossRef](#)] [[PubMed](#)]
57. Golestan, A.; Ghaderi, A.; Mojtahedi, Z. Effects of NDRG2 Overexpression on Metastatic Behaviors of HCT116 Colorectal Cancer Cell Line. *Adv. Pharm. Bull.* **2017**, *7*, 661–664. [[CrossRef](#)] [[PubMed](#)]
58. Xu, Y.; Xin, R.; Sun, H.; Long, D.; Li, Z.; Liao, H.; Xue, T.; Zhang, Z.; Kang, Y.; Mao, G. Long Non-coding RNAs LOC100126784 and POM121L9P Derived from Bone Marrow Mesenchymal Stem Cells Enhance Osteogenic Differentiation via the miR-503-5p/SORBS1 Axis. *Front. Cell. Dev. Biol.* **2021**, *9*, 723759. [[CrossRef](#)] [[PubMed](#)]
59. Moradi Monfared, M.; Alizadeh Zarei, M.; Rafiei Dehbidi, G.; Behzad Behbahani, A.; Arabsolghar, R.; Takhshid, M.A. NDRG2 Regulates the Expression of Genes Involved in Epithelial Mesenchymal Transition of Prostate Cancer Cells. *Iran. J. Med. Sci.* **2019**, *44*, 118–126.
60. Kang, K.W.; Lee, M.J.; Song, J.A.; Jeong, J.Y.; Kim, Y.K.; Lee, C.; Kim, T.H.; Kwak, K.B.; Kim, O.J.; An, H.J. Overexpression of goosecoid homeobox is associated with chemoresistance and poor prognosis in ovarian carcinoma. *Oncol. Rep.* **2014**, *32*, 189–198. [[CrossRef](#)]
61. Ulmer, B.; Tingler, M.; Kurz, S.; Maerker, M.; Andre, P.; Mönch, D.; Campione, M.; Deißler, K.; Lewandoski, M.; Thumberger, T.; et al. A novel role of the organizer gene Goosecoid as an inhibitor of Wnt/PCP-mediated convergent extension in Xenopus and mouse. *Sci. Rep.* **2017**, *7*, 43010. [[CrossRef](#)]
62. Rivera-Perez, J.A.; Wakamiya, M.; Behringer, R.R. Goosecoid acts cell autonomously in mesenchyme-derived tissues during craniofacial development. *Development* **1999**, *126*, 3811–3821. [[CrossRef](#)]
63. Gong, L.; Yang, P.; Hu, L.; Zhang, C. MiR-181b suppresses the progression of epilepsy by regulation of lncRNA ZNF883. *Am. J. Transl. Res.* **2020**, *12*, 2769–2780.