

ChIP-Seq Data Mining: Remarkable Differences in NRSF/REST Target Genes between Human ESC and ESC-Derived Neurons

Jun-ichi Satoh, Natsuki Kawana and Yoji Yamamoto

Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Tokyo 204–8588, Japan.

ABSTRACT: The neuron-restrictive silencer factor (NRSF) is a zinc finger transcription factor that represses neuronal gene transcription in non-neuronal cells by binding to the consensus repressor element-1 (RE1) located in regulatory regions of target genes. NRSF silences the expression of a wide range of target genes involved in neuron-specific functions. Previous studies showed that aberrant regulation of NRSF plays a key role in the pathological process of human neurodegenerative diseases. However, a comprehensive set of NRSF target genes relevant to human neuronal functions has not yet been characterized. We performed genome-wide data mining from chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) datasets of NRSF binding sites in human embryonic stem cells (ESC) and the corresponding ESC-derived neurons, retrieved from the database of the ENCODE/HAIB project. Using bioinformatics tools such as Avadis NGS and MACS, we identified 2,172 NRSF target genes in ESC and 308 genes in ESC-derived neurons based on stringent criteria. Only 40 NRSF target genes overlapped between both data sets. According to motif analysis, binding regions showed an enrichment of the consensus RE1 sites in ESC, whereas they were mainly located in poorly defined non-RE1 sites in ESC-derived neurons. Molecular pathways of NRSF target genes were linked with various neuronal functions in ESC, such as neuroactive ligand-receptor interaction, CREB signaling, and axonal guidance signaling, while they were not directed to neuron-specific functions in ESC-derived neurons. Remarkable differences in ChIP-Seq-based NRSF target genes and pathways between ESC and ESC-derived neurons suggested that NRSF-mediated silencing of target genes is highly effective in human ESC but not in ESC-derived neurons.

KEYWORDS: ChIP-seq, data mining, ESC, GenomeJack, Huntington's disease, human neurons, NRSF, REST

CITATION: Satoh et al. ChIP-Seq Data Mining: Remarkable Differences in NRSF/REST Target Genes between Human ESC and ESC-Derived Neurons. *Bioinformatics and Biology Insights* 2013;7:357–368 doi: 10.4137/BBI.S13279.

TYPE: Original Research

FUNDING: This work was supported by grants from the JSPS KAKENHI (C22500322, C25430054), Japan.

COMPETING INTERESTS: Author(s) disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

CORRESPONDENCE: satoj@my-pharm.ac.jp

Introduction

The neuron-restrictive silencer factor (NRSF), also known as repressor element-1 (RE1) silencing transcription factor (REST), is a zinc finger transcription factor that represses neuronal gene transcription in non-neuronal cells by binding to the consensus RE1 site, also known as neuron-restrictive silencer element (NRSE), located in regulatory regions of target genes.¹ NRSF silences the expression of a wide range of target genes involved in neuron-specific functions encoding ion channels, synaptic vesicle proteins, neurotransmitter receptors, neurotransmitter-synthesizing enzymes, neuropeptides, cell adhesion molecules, and cytoskeletal components.² When NRSF represses the transcription of target genes, it recruits

the repressor complex, composed of the REST corepressor COREST, transcription regulator SIN3A, methyl CpG binding protein MECP2, histone deacetylases, histone modifying enzymes, and the components of the SWI-SNF chromatin remodeling complexes.³

NRSF acts as a master negative regulator of neurogenesis in the central nervous system (CNS) during development. It is expressed at variable levels in most nonneuronal cell types, while the expression levels are reduced in neuronal progenitor cells and more greatly suppressed in differentiated neurons in adult rodent brains.⁴

NRSF plays a pivotal role in the control of proper timing of neuronal differentiation in neural progenitor cells.⁵ Sustained



expression of NRSF inhibits radial migration of neural progenitor cells by targeting doublecortin in the developing rat brain.⁶ Overexpression of the NRSF competitor REST-VP16 induces neuronal differentiation by direct activation of NRSF target genes in mouse neural stem cells.⁷

Because most previous studies on NRSF employed rodent models,^{3–7} the functional roles of NRSF remain largely unknown in the human CNS during development and maturation, except for a previous study showing that blocking NRSF function induces apoptosis in human medulloblastoma cells, although the target genes of NRSF were not characterized.⁸

Increasing evidence indicates that aberrant regulation of NRSF is involved in the pathological process of human neurodegenerative diseases, such as Huntington's disease (HD) and Alzheimer's disease (AD). Wild-type huntingtin (HTT), the causative gene of HD, binds to and sequesters NRSF protein in the cytoplasm, thereby inhibiting the transcriptional repression of BDNF, a cardinal target of NRSF.⁹ In contrast, mutant HTT protein with a polyglutamine repeat expansion causes nuclear accumulation of NRSF and persistently represses the expression of neuronal NRSF target genes in HD brains.⁹ Furthermore, the binding ability of NRSF to NRSE is enhanced in lymphocytes of HD patients.¹⁰ NRSF protein levels are elevated in AD brains correlated with reduced expression of the choline acetyltransferase gene.¹¹ These observations suggest that to clarify the role of NRSF in the pathogenesis of human neurodegenerative diseases, it is important to characterize the comprehensive list of NRSF target genes relevant to human neuronal functions and dysfunctions.

Recently, the next-generation sequencing (NGS) technology had a great impact on the field of genome research. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) serves as a highly efficient NGS method for genome-wide profiling of DNA-binding proteins, histone modifications, and nucleosomes, with high resolution and great coverage of the genome.¹² Increasing amounts of NGS data have been deposited in public databases, such as the Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) and the European Nucleotide Archive (ENA) (<http://www.ebi.ac.uk/ena/home>). Since these databases contain extremely high-throughput experimental data, it is often difficult to extract meaningful biological implications. Recent advances in systems biology enable illustration of a cell-wide map of complex molecular interactions by using the literature-based knowledgebase of molecular pathways.¹³ The logically arranged molecular networks construct the entire system characterized by robustness, which maintains the proper function of the system in the face of genetic and environmental perturbations. Therefore, the integration of extremely high dimensional NGS data with underlying molecular pathways and networks by using bioinformatics tools of knowledgebase offers a rational approach for characterizing the pathway- and network-based molecular mechanisms of gene regulation on the whole genome scale.¹³

To identify genome-wide NRSF target genes relevant to human neuronal functions, we compared ChIP-Seq-based NRSF binding sites between human embryonic stem cells (ESC) and the corresponding ESC-derived neurons and studied the difference in molecular pathways and networks of NRSF target genes. We found remarkable differences in NRSF target genes and pathways between ESC and ESC-derived neurons, suggesting that NRSF-mediated silencing of target genes is highly effective in human ESC but not in ESC-derived neurons.

Methods

ChIP-Seq dataset of NRSF target genes. To identify a comprehensive set of NRSF target genes in human ESC and ESC-derived neurons, we investigated FASTQ-formatted files of ChIP-Seq datasets retrieved from the DDBJ Sequence Read Archive (DRA) (<http://trace.ddbj.nig.ac.jp/DRASearch>) under the accession number of SRX100410, SRX190260, SRX100409, and SRX190329. The datasets were originally prepared for the Encyclopedia of DNA Elements (ENCODE) project by Dr Richard Myers, HudsonAlpha Institute for Biotechnology (HAIB). In their experiments, human H1 ESC cells established from inner cell mass, presenting with a normal karyotype, were processed to induce neuronal differentiation. Both ESC and ESC-derived neurons were processed for ChIP with mouse monoclonal anti-NRSF antibody (Protein Expression Center, Caltech). Next, NGS libraries constructed from size-selected ChIP DNA fragments were processed for deep sequencing at a read length composed of 30–36 bp for ESC or 40–50 bp for neurons on Genome Analyzer (Illumina, San Diego, CA, USA). In both cell types, the corresponding libraries termed RevXlinkChromatin made from sonicated chromatin, whose formaldehyde cross-links are reversed without immunoenrichment, were utilized for input control.

First, we evaluated the quality of short reads by searching them on the FastQC program. Next, we removed the reads of insufficient quality by filtering them out by the FASTX-toolkit. After cleaning the data, we mapped them on the human genome reference sequence version hg19 using a mapping tool of Avadis NGS (Agilent Technologies, Santa Clara, CA, USA) or using the Bowtie program. We then identified significant peaks by using the Model-based Analysis of ChIP-Seq (MACS) program in the stringent condition that satisfies the false discovery rate (FDR) ≤ 0.01 and fold enrichment (FE) ≥ 20 in order to reduce the detection of false-positive binding sites if possible as described previously.¹⁴ We characterized the genomic location of binding peaks using a peak-finding tool of Avadis NGS that classifies the locations into the upstream region, 5' untranslated region (5'UTR), exon, intron, and 3'UTR. To identify the exact location of binding sites, we imported the Bowtie and MACS-processed data into the GenomeJack genome viewer (Mitsubishi Space Software) as described previously.¹⁴



The genes corresponding to MACS peaks were determined by a neighboring gene analysis tool of Avadis NGS within a distance of 5000 bp from peaks to genes. The consensus sequence motifs were identified by importing a 400 bp-length sequence surrounding the summit of MACS peaks derived from top 100 genes based on FE into the MEME-ChIP program. The utility of QC tools, Bowtie, MACS, and MEME-ChIP is described elsewhere.^{15–18} The information on websites and analysis conditions of bioinformatics tools is summarized in Supplementary Table 1.

Molecular pathway analysis. To identify biologically relevant molecular pathways and networks, we imported Entrez Gene IDs of NRSF target genes into the Functional

Annotation tool of Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7.¹⁹ DAVID extracts gene ontology (GO) terms enriched in the set of imported genes and identifies relevant pathways constructed by Kyoto Encyclopedia of Genes and Genomes (KEGG) or the Protein Analysis Through Evolutionary Relationships (PANTHER) classification system. The results are followed by statistical evaluation with the modified Fisher's exact test corrected by multiple comparison tests. In DAVID analysis data, we considered P -value < 0.05 by Bonferroni correction as significant. KEGG is a publicly accessible knowledgebase that contains manually curated 272,733 pathways that cover a wide range of metabolic, genetic, environmental, and cellular processes,

Table 1. Top 20 Most Enriched ChIP-Seq-based NRSF target genes in human ESC.

CHROMOSOME	START BASE POSITION	END BASE POSITION	FE	FDR	ENTREZ GENE ID	GENE SYMBOL	GENE NAME
chr14	24981675	24982925	170.00177	0	1215	CMA1	chymase 1, mast cell
chr6	152167485	152168597	170.00177	0	2099	ESR1	estrogen receptor 1
chr2	166094555	166096160	170.00177	0	6326	SCN2A	sodium channel, voltage-gated, type II, alpha subunit
chr14	79627693	79628461	170.00177	0	9369	NRXN3	neurexin 3
chr9	19781808	19782570	170.00177	0	25769	SLC24A2	solute carrier family 24 (sodium/potassium/calcium exchanger), member 2
chr10	60936329	60937937	170.00177	0	84457	PHYHIPL	phytanoyl-CoA 2-hydroxylase interacting protein-like
chr2	163694658	163698116	170.00177	0	90134	KCNH7	potassium voltage-gated channel, subfamily H (eag-related), member 7
chr19	57148425	57151677	170.00177	0	147670	LOC147670	uncharacterized LOC147670
chr3	68355362	68356382	170.00177	0	407738	FAM19A1	family with sequence similarity 19 (chemokine (C-C motif)-like), member A1
chr6	167227855	167229954	169.15176	0	6196	RPS6KA2	ribosomal protein S6 kinase, 90kDa, polypeptide 2
chr6	108508703	108509966	169.15176	0	7101	NR2E1	nuclear receptor subfamily 2, group E, member 1
chr6	32072697	32073801	169.15176	0	7148	TNXB	tenascin XB
chr19	10203673	10204816	169.15176	0	55337	C19orf66	chromosome 19 open reading frame 66
chr6	119411155	119412478	169.15176	0	79632	FAM184A	family with sequence similarity 184, member A
chrX	153140408	153141899	167.45175	0	3897	L1CAM	L1 cell adhesion molecule
chr19	4171564	4174043	167.45175	0	84699	CREB3L3	cAMP responsive element binding protein 3-like 3
chr14	58331291	58333220	167.45175	0	341880	SLC35F4	solute carrier family 35, member F4
chr19	51983546	51984430	167.45175	0	729767	CEACAM18	carcinoembryonic antigen-related cell adhesion molecule 18
chr8	69084217	69085604	167.36694	0	80243	PREX2	phosphatidylinositol-3,4,5-trisphosphate-dependent Rac exchange factor 2
chr4	147564574	147565930	166.8817	0	5458	POU4F2	POU class 4 homeobox 2

From ChIP-Seq datasets, we identified 2,172 NRSF target genes in human ESC based on stringent criteria that satisfy both false discovery rate (FDR) ≤ 0.01 and fold enrichment (FE) ≥ 20 . Top 20 genes based on FE are listed with the chromosome, the start base position, the end base position, FE, FDR, Entrez Gene ID, Gene Symbol, and Gene Name.



as well as human diseases. PANTHER is a publicly accessible knowledgebase that includes the information on 176 signaling and metabolic pathways manually curated by expert biologists.

We also imported Entrez Gene IDs of NRSF target genes into the Core Analysis tool of Ingenuity Pathways Analysis (IPA) (Ingenuity Systems). IPA is a commercial knowledgebase that contains approximately 3,000,000 biological and chemical interactions with definite scientific evidence. By uploading the list of Gene IDs, the network-generation algorithm identifies focused genes integrated in global molecular pathways and networks. IPA calculates the score *P*-value that reflects the statistical significance of association between the genes and the pathways and networks by Fisher's exact test. For IPA network and pathway analysis data, we considered *P*-value < 0.05 by Fisher's exact test as significant.

Results

Identification of ChIP-Seq-based NRSF target genes in human ESC and ESC-derived neurons. After mapping short reads on hg19 and detecting peaks by MACS on Avadis NGS, we initially identified 36,096 ChIP-Seq peaks of NRSF target genes in ESC and 45,108 peaks in ESC-derived neurons. After performing the neighboring gene search within a distance of 5,000 bp from peaks to genes, these were reduced to 12,757 genes in ESC and 15,802 genes in ESC-derived neurons, respectively. These results indicate that approximately 65% of ChIP-Seq peaks were located in intergenic regions far from the nearest genes. Upon application of stringent criteria that satisfied both $FDR \leq 0.01$ and fold enrichment (FE) ≥ 20 , the results were further reduced to 2,697 genes in ESC and 388 genes in ESC-derived neurons, respectively. After omitting non-protein coding genes, we extracted 2,172 protein-coding genes in ESC (FE = 59.4 ± 38.8) and 308 genes

Table 2. Top 20 Most Enriched ChIP-Seq-based NRSF target genes in human ESC-derived neurons.

CHROMOSOME	START BASE POSITION	END BASE POSITION	FE	FDR	ENTREZ GENE ID	GENE SYMBOL	GENE NAME
chrX	54833719	54835596	59.98119	0	10916	MAGED2	melanoma antigen family D, 2
chr22	20066889	20069179	58.29808	0	54487	DGCR8	DiGeorge syndrome critical region gene 8
chr10	70090608	70093965	54.11434	0	3189	HNRNPH3	heterogeneous nuclear ribonucleoprotein H3 (2H9)
chr10	70090608	70093965	54.11434	0	64081	PBLD	phenazine biosynthesis-like protein domain containing
chr4	1403924	1404873	51.01328	0.007168459	54729	NKX1-1	NK1 homeobox 1
chr11	66187302	66189955	48.899265	0	266743	NPAS4	neuronal PAS domain protein 4
chr5	176264343	176265229	48.634613	0.003686636	90249	UNC5A	unc-5 homolog A (C. elegans)
chr19	23868322	23870535	48.258587	0	171392	ZNF675	zinc finger protein 675
chr5	167860686	167861828	46.708134	0	23286	WWC1	WW and C2 domain containing 1
chrX	102469463	102471360	46.172604	0.002554278	56271	BEX4	brain expressed, X-linked 4
chr4	26320438	26324056	46.108536	0	3516	RBPJ	recombination signal binding protein for immunoglobulin kappa J region
chr20	33102940	33105165	44.41789	0	83658	DYNLRB1	dynein, light chain, roadblock-type 1
chr1	230939858	230940883	42.362473	0.004624277	10753	CAPN9	calpain 9
chr9	131936079	131941155	40.959568	0	389792	IER5L	immediate early response 5-like
chr2	70120481	70121863	39.22451	0.004094166	11017	SNRNP27	small nuclear ribonucleoprotein 27kDa (U4/U6.U5)
chr11	65263253	65278757	39.090984	0	100996467	LOC100996467	metastasis-associated lung adenocarcinoma transcript 1-like
chr2	169765195	169767746	38.41547	0	57818	G6PC2	glucose-6-phosphatase, catalytic, 2
chr3	195497495	195498727	38.410645	0.002728513	4585	MUC4	mucin 4, cell surface associated
chr2	134279356	134280482	38.321316	0	344148	NCKAP5	NCK-associated protein 5
chr20	44561770	44564554	37.885845	0	63935	PCIF1	PDX1 C-terminal inhibiting factor 1

From ChIP-Seq datasets, we identified 308 NRSF target genes in human ESC-derived neurons based on stringent criteria that satisfy both false discovery rate (FDR) ≤ 0.01 and fold enrichment (FE) ≥ 20 . Top 20 genes based on FE are listed with the chromosome, the start base position, the end base position, FE, FDR, Entrez Gene ID, Gene Symbol, and Gene Name.



in ESC-derived neurons ($FE = 26.0 \pm 6.8$). The top 20 genes based on FE are listed in Table 1 for ESC and Table 2 for ESC-derived neurons. The complete lists are shown in Supplementary Table 2 for ESC and Supplementary Table 3 for ESC-derived neurons.

Importantly, when compared with 985 ChIP-Seq-based NRSF target genes in Jurkat human T cells identified in a previous study,²⁰ 405 genes (41.1%) of Jurkat overlapped with NRSF target genes of human ESC, supporting the validity of our analysis (Supplementary Table 4). In contrast, only 40 genes overlapped between human ESC and ESC-derived neurons.

These included ST8SIA3, NPAS4, EP300, NKX1-1, ARID2, SLCA7 A14, MEX3B, GPR26, ZSWIM5, SEZ6 L2, ASPHD1, DDX39B, CELSR3, GAD2, SPARTA16, MPHOSPH10, MCEE, RAB28, FAM117B, KCNMA1, MRP63, SKA3, TMEM187, LATS1, PGK1, RPS17, LRP6, H3F3B, EIF4 A2, MGEA5, RPL10, KIF21 A, FUS, AZIN1, WHSC2, BTG1, CCDC88 A, WWC1, TMED2, and GLCE (underlined in Supplementary Table 2 and Supplementary Table 3). The network analysis of 40 genes by IPA indicated that the set of 14 genes (35%), such as ARID2, MEX3B, SEZ6 L2, GAD2, RAB28, SKA3, RPS17, RPL10,

Table 3. Top 5 IPA canonical pathways of NRSF target genes in human ESC and ESC-derived neurons.

RANK	CATEGORY	FOCUSED MOLECULES	RATIO	P-VALUE
Human ESC				
1	Corticotropin releasing hormone signaling	CAMK4, BDNF, MAPK11, <u>EP300</u> , GLI3, UCN, PRKAR1B, SMO, NOS2, PRKD3, MAP2K1, GNAS, PRKCQ, PTCH1, UCN3, GNAQ, CRH, ITPR1, MEF2B, CALM1, GNAI3, ADCY9, PLCG2, PRKCD, GNAO1, ADCY1, NR4A1, GUCY1A2, PRKAG2, MEF2C, ADCY7, OPN1SW	32/138 (0.232)	7.05E-08
2	Role of NFAT in cardiac hypertrophy	CAMK4, CAMK1D, PIK3R5, PLCH2, MAPK11, <u>EP300</u> , HDAC6, TGFB1, PRKAR1B, PIK3R2, GNB1L, MAP2K1, PRKD3, CAMK2B, SRC, HDAC9, HDAC4, PRKCQ, GNAS, HDAC8, GRB2, GNAQ, PIK3C2G, SLC8A3, SLC8A2, ITPR1, MEF2B, GNAI3, CALM1, NKX2-5, ADCY9, PLCB4, HDAC3, PRKCD, PLCG2, ADCY1, MAPK10, PRKAG2, MEF2C, SLC8A1, RCAN2, ADCY7, OPN1SW	43/209 (0.206)	1.48E-07
3	CREB signaling in neurons	GRIN2A, CAMK4, GRID2, GNA11, PIK3R5, POLR2B, GNA14, PLCH2, <u>EP300</u> , GRIN2C, PRKAR1B, PIK3R2, GNB1L, MAP2K1, PRKD3, CAMK2B, POLR2J2/POLR2J3, GRIN1, GRIN2B, GNAS, PRKCQ, GRM8, GRB2, GRM1, GRIK3, PIK3C2G, GNAQ, ITPR1, GRM7, GNAI3, CALM1, ADCY9, PLCB4, PRKCD, PLCG2, GNAO1, ADCY1, PRKAG2, RPS6KA1, ADCY7, OPN1SW	41/206 (0.199)	3.64E-07
4	Axonal guidance signaling	DPYSL2, PAPP2, KLC1, RAC2, MYL10, PLCH2, NTN1, SEMA6D, ECE2, ABLIM3, UNC5D, ADAM23, ABLIM2, PLXNB3, PRKD3, RTN4R, PAPP2, PRKCQ, PTCH1, VEGFC, L1CAM, SRGAP3, ADAM12, MYL12B, PLCG2, PRKCD, GNAO1, PAK7, OPN1SW, NRP1, FYN, RGS3, BDNF, GNA11, PIK3R5, EGF, GNA14, ABLIM1, SEMA6C, EPHA8, WNT7A, GLI3, GLIS1, SDC2, EFNB1, NTRK1, SMO, PRKAR1B, ARPC3, PIK3R2, RASSF5, ROBO2, SHANK2, GNB1L, MAP2K1, ACE, UNC5C, SEMA3E, PLXNA1, GNAS, PLXNC1, ARHGEF12, NRP2, GRB2, WNT2B, PIK3C2G, GNAQ, EPHA3, GNAI3, EPHA10, PLCB4, TUBB6, NTRK3, ADAM10, PRKAG2, NFATC2, SEMA3C, BMP6	78/471 (0.166)	4.24E-07
5	Neuropathic pain signaling in dorsal horn neurons	GRIN2A, CAMK4, CAMK1D, BDNF, PIK3R5, PLCH2, GPR37, GRIN2C, PRKAR1B, PIK3R2, PRKD3, CAMK2B, GRIN2B, SRC, GRIN1, KCNN2, PRKCQ, GRM8, GRM1, PIK3C2G, ITPR1, GRIN3A, GRM7, PLCB4, PRKCD, PLCG2, PRKAG2	27/108 (0.25)	2.02E-06
ESC-derive human neurons				
1	EIF2 signaling	RPL24, RPS28, RPS2, <u>EIF4A2</u> , <u>RPS17</u> , RPS29, RPL15, RPL5, RPL7A, EIF5, <u>RPL10</u> , EIF3A, RPS15, RPS5, RPL13	15/200 (0.075)	1.98E-07
2	Regulation of eIF4 and p70S6K signaling	RPS29, RPS2, RPS28, RPS15, EIF3A, PPP2R3B, PPP2R5C, <u>EIF4A2</u> , <u>RPS17</u> , RPS5	10/175 (0.057)	1.10E-04
3	mTOR signaling	RPS29, TSC1, RPS2, RPS28, RPS15, EIF3A, PPP2R3B, PPP2R5C, <u>EIF4A2</u> , <u>RPS17</u> , RPS5	11/211 (0.052)	2.07E-04
4	ERK5 signaling	FOS, YWHAG, YWHAE, YWHAZ, ATF4, <u>EP300</u>	6/66 (0.091)	4.74E-04
5	ERK/MAPK signaling	FOS, YWHAG, <u>H3F3B</u> , DUSP1, CRKL, YWHAZ, ATF4, PPP2R3B, PPP2R5C, <u>EP300</u>	10/208 (0.048)	7.46E-04

By importing Entrez Gene IDs of ChIP-Seq-based NRSF target genes into the Core Analysis tool of IPA, we identified canonical pathways with significant relevance to the imported genes. Top 5 pathways based on *P-value* are listed with the category, focused molecules, ratio, and *P-value* of Fisher's exact test. The NRSF target genes shared between ESC and ESC-derived neurons are underlined.

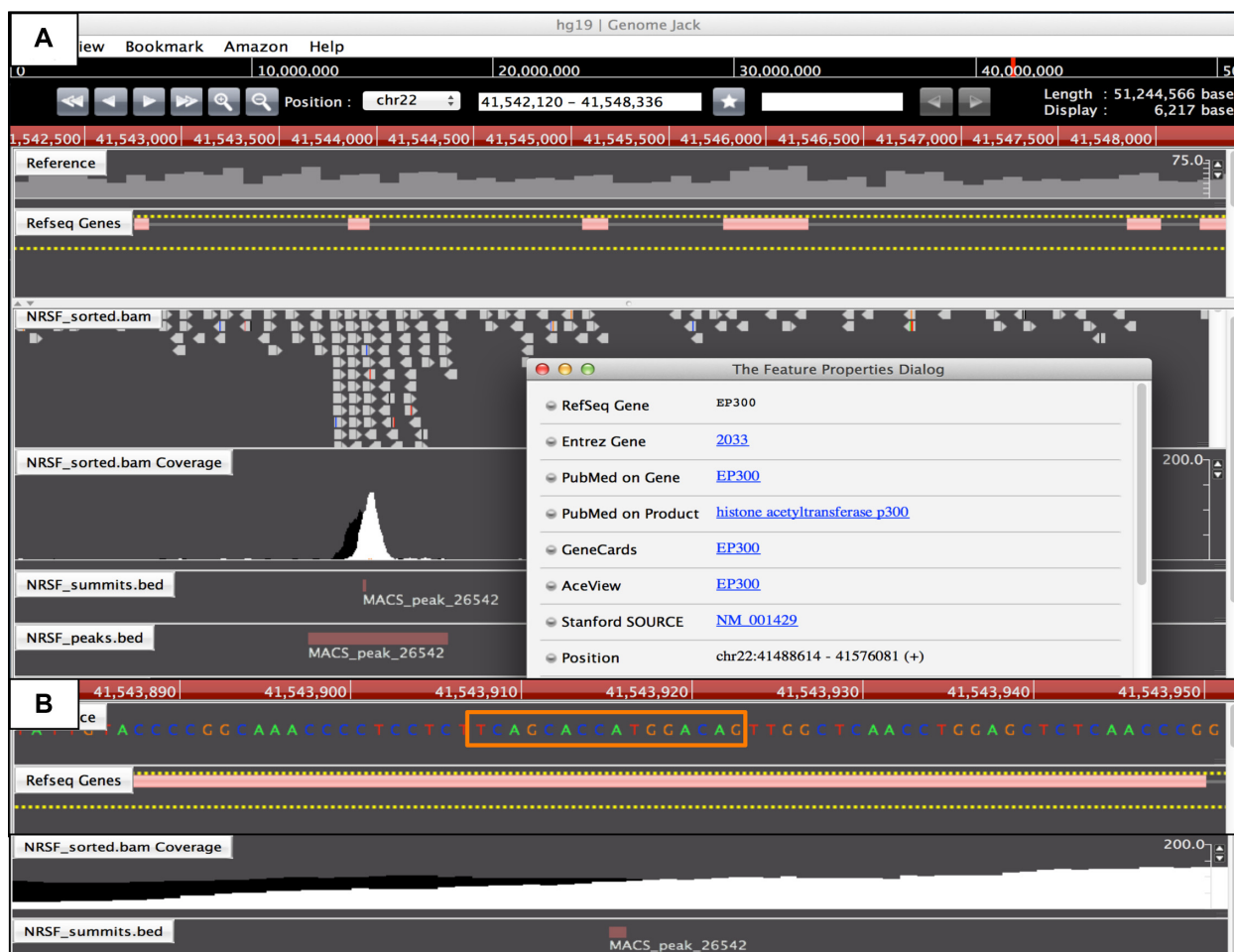


Figure 1. The genomic location of ChIP-Seq-based NRSF target gene EP300 in ESC. An example of NRSF target gene in ESC named histone acetyltransferase p300 (EP300) is shown on GenomeJack, where a MACS peak numbered 26542 is located in the exon of EP300. The panels (A, B) represent (A) the overview and (B) the close-up view. The canonical REST motif is highlighted by orange rectangle.

KIF21A, FUS, AZIN1, WHSC2 (NEFLA), CCDC88A, and TMED2, belong to the molecules interacting with UBC (ubiquitin C) (Supplementary Fig. 1).

Among 2,172 genes detected in ESC, the peaks were located in the upstream region (8.6%), 5'UTR (12.8%), exon (21.9%), intron (53.4%), and 3'UTR (3.3%). Among 308 genes detected in ESC-derived neurons, the peaks were located in the upstream region (1.9%), 5'UTR (30.9%), exon (27.4%), intron (36.2%), and 3'UTR (3.7%). Thus, the intron serves as a major NRSF-binding site in both ESC and ESC-derived neurons. The more precise location was determined by importing processed data into the GenomeJack genome viewer (Fig. 1 for ESC and Fig. 2 for ESC-derived neurons).

According to motif analysis with MEME-ChIP of top 100 genes based on FE, the most significant consensus sequences in ESC were characterized by a well-defined REST motif (RE1) composed of 5'AG[CG]ACCA[TC]GGACAG3' (80 sites; E-value = 5.1×10^{-254}) and a left half-site of RE1 composed of 5'[TG]TTCAGCAC[CT]3' (21 sites; E-Value = 2.9×10^{-11}) (Fig. 3A, B). In contrast, the

most significant consensus sequences in ESC-derived neurons represented a poorly defined non-RE1 motif composed of 5'[GA][AGC][GA][AG][ATG][GA]G[CA]GG[C/G][AG]G[CGT][GA][GA][CAG][AG][GC][TA][GA][GA][GA][AG][CAG][GA][GC]A3' (63 sites; E-Value = 6.7×10^{-17}) and an RE1-like motif composed of 5'[CG][CT][TG]TC[AT][GT]CACCG[CT][GA]GACAGC[GA]C[CT][ACT][TG]3' (11 sites; E-Value = 1.7×10^{-4}) (Fig. 3C, D). These results suggest that NRSF binds chiefly to the RE1 consensus sites in ESC by direct binding, whereas NRSF often binds to the promiscuous non-RE1 sites in ESC-derived neurons possibly by indirect binding.

Molecular pathways of ChIP-Seq-based NRSF target genes in human ESC and ESC-derived neurons. Next, we studied molecular pathways and networks of ChIP-Seq-based NRSF target genes by using bioinformatics tools of the comprehensive knowledgebase. First, by importing the sets of Entrez Gene IDs of NRSF target genes, DAVID identified statistically significant functionally associated GO terms and related KEGG and PANTHER pathways. The top 3 most significant GO terms, regardless of their categories in ESC,

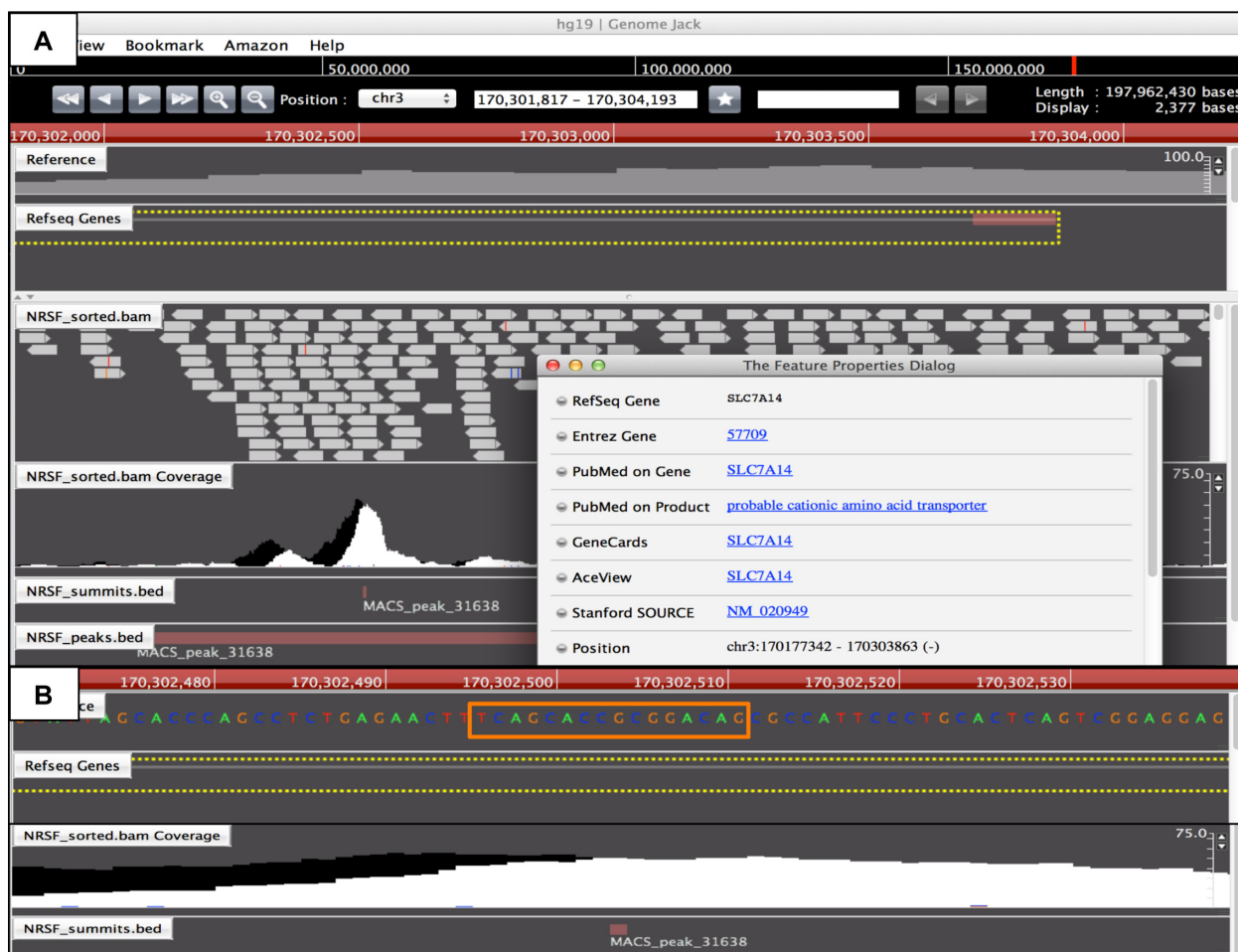


Figure 2. The genomic location of ChIP-Seq-based NRSF target gene SLC7 A14 in ESC-derived neurons. An example of NRSF target gene in ESC-derived neurons named solute carrier family 7, member 14 (SLC7A14) is shown on GenomeJack, where a MACS peak numbered 31638 is located in the first intron of SLC7A14. The panels (A, B) represent (A) the overview and (B) the close-up view. The canonical REST motif is highlighted by orange rectangle.

included “neuron projection” (GO: 0043005; $P = 1.21 \times 10^{-8}$), “calcium ion binding” (GO: 0005509; $P = 1.21 \times 10^{-7}$), and “hemophilic cell adhesion” (GO: 0007156; $P = 3.72 \times 10^{-6}$). In ESC, only one KEGG pathway, termed “Neuroactive ligand-receptor interaction” (hsa04080; $P = 0.0343$; Fig. 4), and only one PANTHER pathway, termed “Wnt signaling pathway” (P00057; $P = 0.0399$), were considered to be significantly relevant to NRSF target genes. In contrast, the top 3 most significant GO terms, regardless of their categories in ESC-derived neurons, included “nuclear lumen” (GO: 0031981; $P = 8.72 \times 10^{-8}$), “intracellular non-membrane-bounded organelle” (GO: 0043232; $P = 5.41 \times 10^{-7}$), and “non-membrane-bounded organelle” (GO: 0043228; $P = 5.41 \times 10^{-7}$). In ESC-derived neurons, only one KEGG pathway termed “Ribosome” was significant (hsa03010; $P = 3.53 \times 10^{-5}$) and none of the PANTHER pathways reached the statistical significance after Bonferroni correction.

Finally, we imported the list of Entrez Gene IDs of NRSF target genes into IPA. It extracted the network defined as “Cell-To-Cell Signaling and Interaction, Cell

Signaling, Nervous System Development and Function” ($P = 1.00 \times 10^{-39}$) and “Endocrine System Disorders, Gastrointestinal Disease, Hereditary Disorder” ($P = 1.00 \times 10^{-39}$) in ESC, and “Cell-mediated Immune Response, Cellular Movement, Developmental Disorder” ($P = 1.00 \times 10^{-71}$) in ESC-derived neurons as the most significant functionally associated molecular network. IPA also identified canonical pathways relevant to the set of imported genes (Table 3). The top 5 most significant pathways in ESC included “Corticotropin releasing hormone signaling” ($P = 7.05E-08$), “Role of NFAT in cardiac hypertrophy” ($P = 1.48 \times 10^{-7}$), “CREB signaling in neurons” ($P = 3.64 \times 10^{-7}$; Fig. 5), “Axonal guidance signaling” ($P = 4.24 \times 10^{-7}$), and “Neuropathic pain signaling in dorsal horn neurons” ($P = 2.02 \times 10^{-6}$). In contrast, the top 5 most significant pathways in ESC-derived neurons were composed of “EIF2 signaling” ($P = 1.98 \times 10^{-8}$), “Regulation of eIF4 and p70S6K signaling” ($P = 1.10 \times 10^{-4}$), “mTOR signaling” ($P = 2.07 \times 10^{-4}$), “ERK5 signaling” ($P = 4.74 \times 10^{-4}$), and “ERK/MAPK signaling” ($P = 7.46 \times 10^{-4}$). When the top 5 IPA pathways were compared between both cell types,

Signaling" ($P = 2.79 \times 10^{-6}$), "Glutamate Receptor Signaling" ($P = 2.18 \times 10^{-5}$), "Calcium Signaling" ($P = 2.84 \times 10^{-4}$), "Synaptic Long Term Depression" ($P = 5.25 \times 10^{-4}$), and "Circadian Rhythm Signaling" ($P = 5.41 \times 10^{-4}$) as the top 5 significant pathways. These observations suggested that the pathways involved in neuronal functions are enriched in NRSF target genes, and they are shared substantially between ESC and T lymphocytes.

We found that NRSF binding sites are located in various genomic regions, among which the intron region serves as the most major site in both ESC and ESC-derived neurons except for intergenic regions, which was consistent with previous observations.²³ Importantly, NRSF effectively inhibits transcription of target genes, regardless of the location or the orientation of its binding sites in 5'UTRs, introns, exons, upstream of the transcription start, and downstream of the coding region distant from the enhancer and the promoter.²⁴ Furthermore, only a subset of NRSE sequences are constitutively occupied by NRSF in non-neuronal cells, and the occupancy depends largely on endogenous levels of NRSF.²²

The canonical NRSE site reflects a 21-bp consensus sequence composed of 5'TTCAGCACCACGGA-CAGCGCC3', which is divided into the left half-site and the right half-site separated by a spacer of two nucleotides.²³ However, NRSF binding sites with highly degenerate sequences are often found throughout the genome.²⁵ Through motif analysis, we found that most NRSF binding sites in ESC are present in the most significant RE1 sites (E-value = 5.1×10^{-254}), while they are located mostly in non-RE1 sites and the minority is located in less significant RE1 sites (E-value = 1.7×10^{-4}) in ESC-derived neurons. The possibility exists that the lower expression levels of NRSF in ESC-derived neurons may make it difficult to detect specific ChIP-Seq peaks. However, a highly sensitive monoclonal anti-NRSF antibody was utilized to detect a trace amount of NRSF protein for both ESC and ESC-derived neurons. Therefore, our observations suggest that NRSF binds chiefly to RE1 sites in ESC by direct binding, whereas it often binds to promiscuous non-RE1 sites in ESC-derived neurons by indirect binding via an interaction with other DNA binding

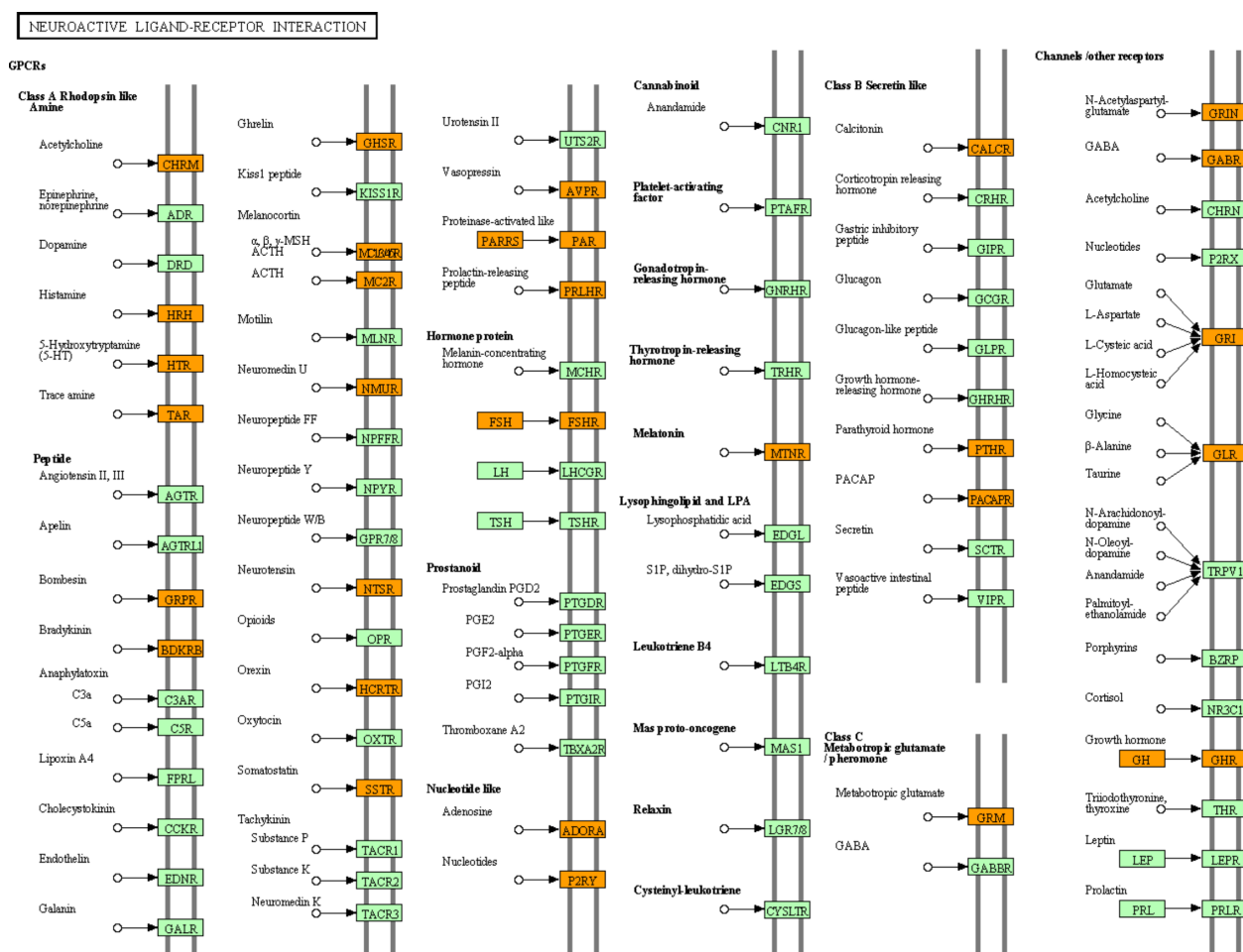


Figure 4. KEGG pathway of ChIP-Seq-based NRSF target genes in ESC. Entrez Gene IDs of 2,172 ChIP-Seq-based NRSF target genes in ESC were imported into the Functional Annotation tool of DAVID. It extracted statistically significant KEGG pathways relevant to the set of imported genes. KEGG pathway termed "Neuroactive ligand-receptor interaction" (hsa04080; $P = 0.0343$) is shown. NRSF target genes are highlighted by orange.

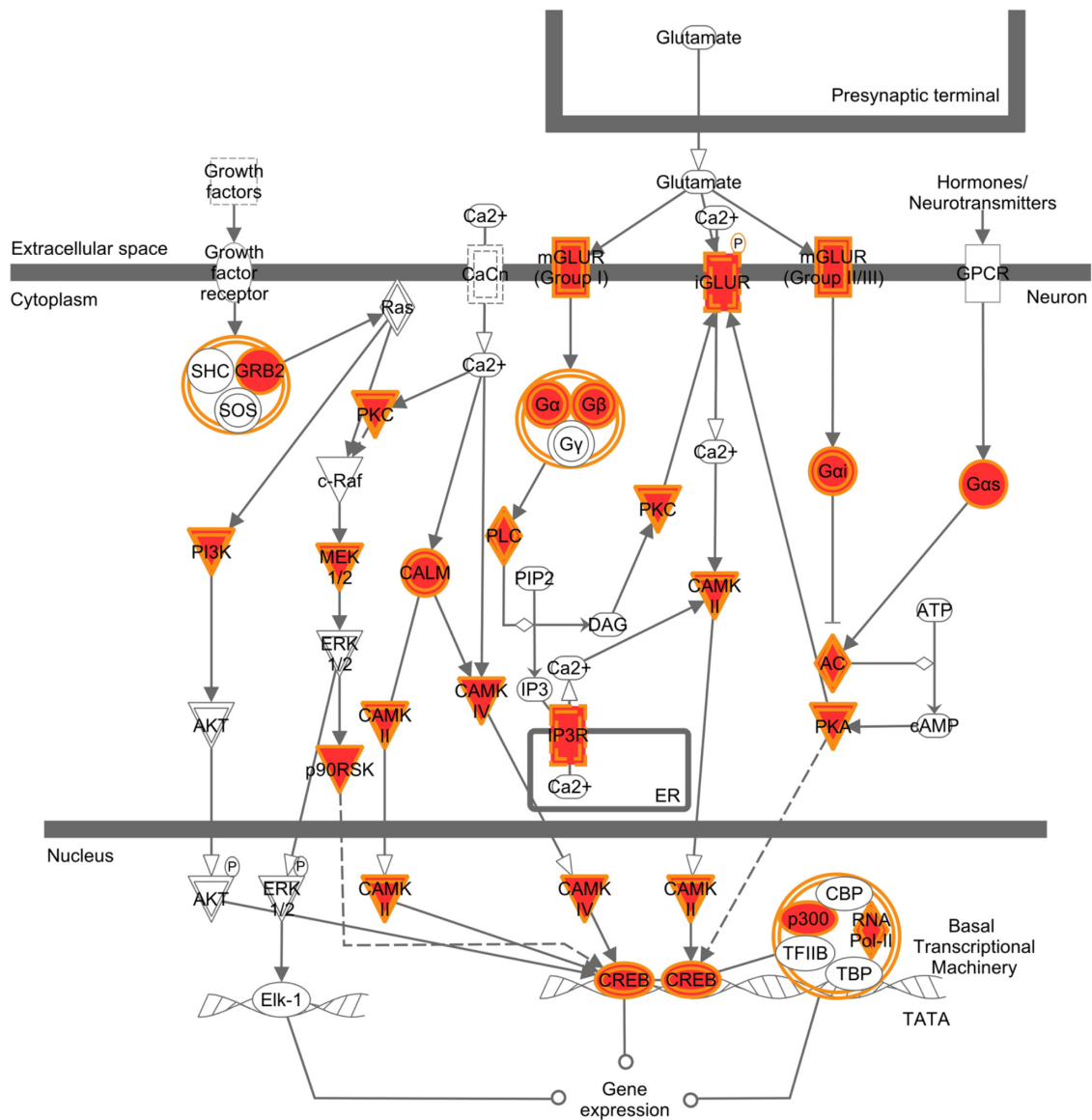


Figure 5. IPA pathways of ChIP-Seq-based NRSF target genes in ESC. Entrez Gene IDs of 2,172 ChIP-Seq-based NRSF target genes in ESC were imported into the Core Analysis tool of IPA. It extracted statistically significant canonical pathways relevant to the set of imported genes. The third rank pathway termed “CREB signaling in neurons” ($P = 3.64 \times 10^{-7}$) in Table 3 is shown. NRSF target genes are highlighted in red.

proteins, as described previously.²³ Notably, noncanonical binding sites are often highly polymorphic, where the left and right half-site motifs are separated by variable-length spacer sequences.^{25,26} Furthermore, numerous ChIP-Seq-based NRSF binding sites contain only partial motifs composed of a single half-site.²³ NRSF is composed of a DNA binding domain located within the cluster of eight zinc fingers, and two repressor domains at the N- and C-termini, which are capable of interacting with corepressor complex components.³ Importantly, as neuronal progenitor cells differentiate into neurons, NRSF and its corepressor complex gradually dissociate from RE1 sites, triggering activation of neuronal gene expression.²⁷ These observations may explain why only a limited subset of NRSF target genes overlapped between human ESC and ESC-derived neurons.

By molecular pathway analysis, we identified the pathways termed “Neuroactive ligand-receptor interaction”, “Wnt signaling pathway”, “CREB signaling in neurons”, and “Axonal guidance signaling”, all of which are closely related to neuronal functions, as the pathways relevant to NRSF target genes in ESC. Among these, downregulation of CREB signaling may represent a highly important function of NRSF in non-neuronal cells. CREB is promptly activated in response to a wide range of extracellular stimuli, such as growth factors, peptide hormones, and neuronal activity, which activates protein kinase A (PKA), mitogen-activated protein kinases (MAPKs), and Ca²⁺/calmodulin-dependent protein kinases (CAMKs).²⁸ Phosphorylated CREB upregulates a panel of genes essential for neuronal differentiation and plasticity, including BDNF, TrkB, somatostatin, and c-fos.²⁸ NRSF inhibits the CREB

signaling on activation of cocaine and amphetamine regulated transcript (CART) in SK-N-SH neuroblastoma cells, being in line with our observations.²⁹ Furthermore, NRSF expression is positively regulated by CREB in small-cell lung cancer (SCLC) cell lines, serving as an autoregulatory loop in the NRSF-CREB signaling pathway.³⁰ Importantly, NRSF is upregulated in Alzheimer's disease (AD) brains,¹¹ and CREB signaling is deregulated in the hippocampus of AD brains.³¹ With respect to the "Axonal guidance signaling" pathway of NRSF target genes, overexpression of NRSF in differentiated neurons causes axonal guidance errors in the developing check spinal cord.³²

In contrast to the molecular pathways in ESC, we identified the pathways linked to regulation of protein synthesis termed "Ribosome", "EIF2 signaling", "Regulation of eIF4 and p70S6K signaling", and "mTOR signaling" as the major pathways relevant to NRSF target genes in ESC-derived neurons. They converge on the PI3K/Akt/mTOR signaling pathway pivotal for cellular proliferation and survival in all cell types.³³ These observations suggest that NRSF target genes play a key role in maintenance of neuronal functions, which are suppressed in ESC, whereas they are not restricted to neuron-specific functions in ESC-derived neurons. Notably, NRSF negatively regulates WNT signaling pathway, providing a mechanism responsible for maintenance of pluripotency in ESC.³⁴

Increasing evidence indicates that aberrant regulation of NRSF plays a pivotal role in the pathogenesis of cancer, cardiac hypertrophy, ischemia, epilepsy, and neurodegeneration.^{9–11,35–37} NRSF acts as an essential regulator of maintenance of normal cardiac structure, since transgenic mice expressing a dominant-negative mutant NRSF in hearts exhibited dilated cardiomyopathy.³⁸ Through IPA analysis, we found the pathway termed "Role of NFAT in cardiac hypertrophy" as the second-ranked significant pathway for NRSF target genes in ESC (Table 3). In Huntington's disease (HD) brains, the inability of mutant huntingtin (HTT) to retain NRSF in the neuronal cytoplasm inhibits the expression of NRSF target genes essential for neuronal functions.⁹ We identified both NRSF itself and HTT as potential NRSF target genes in human ESC, because they have degenerate RE1 motifs of NRSF binding sites (Supplementary Fig 2 for NRSF/REST and Supplementary Fig 3 for HTT). These observations suggested the existence of a complex autoregulatory loop in the network of the HTT-NRSF axis.

Conclusions

Using genome-wide data mining from ChIP-Seq datasets, we identified 2,172 NRSF target genes in human ESC and 308 NRSF target genes in ESC-derived neurons. The binding regions were accumulated in the RE1 sites in human ESC, whereas they were located mainly in non-RE1 sites in ESC-derived neurons. The molecular pathways of NRSF target genes were relevant to the maintenance of neuronal functions in ESC, whereas they were not limited

to neuron-specific functions in ESC-derived neurons. Remarkable differences in ChIP-Seq-based NRSF target genes and pathways between ESC and ESC-derived neurons suggested that NRSF-mediated silencing of target genes is highly effective in human ESC but not in ESC-derived neurons.

Acknowledgements

The authors thank Ms. Aki Takaoka for her invaluable help.

Author Contributions

JS designed the methods, analyzed the data, and drafted the manuscript. NK and YY helped with the data analysis. All authors have read and approved the final manuscript.

DISCLOSURES AND ETHICS

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

Supplementary Data

Supplementary Table 1.

Supplementary Table 2.

Supplementary Table 3.

Supplementary Table 4.

Supplementary Figure 1. The molecular network of NRSF target genes shared between human ESC and ESC-derived neurons. Entrez Gene IDs of 40 NRSF target genes overlapping between ESC and ESC-derived neurons were imported into the Core Analysis tool of IPA. It extracted the functional network defined as "Cellular Assembly and Organization, Cellular Function and Maintenance, Developmental Disorder" ($P = 1.00 \times 10^{-35}$), composed of 14 molecules interacting with UBC (ubiquitin C), as the most significant network.

Supplementary Figure 2. The genomic location of ChIP-Seq-based NRSF target gene REST in ESC. An example of NRSF target gene named RE1-silencing transcription factor (REST) is shown on GenomeJack, where a MACS peak numbered 30234 is located in the first intron of REST transcript variant NM_005612. The panels (a, b) represent (a) the overview and (b) the close-up view. The degenerate REST motif is highlighted in an orange rectangle.

Supplementary Figure 3. The genomic location of ChIP-Seq-based NRSF target gene HTT in ESC. An example of NRSF target gene named huntingtin (HTT) is shown on GenomeJack, where a MACS peak numbered 29571 is located in the 3'UTR of HTT. The panels (a, b) represent (a) the overview and (b) the close-up view. The degenerate REST motif is highlighted in an orange rectangle.



REFERENCES

- Schoenherr CJ, Anderson DJ. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science*. 1995;267(5202):1360–3.
- Schoenherr CJ, Paquette AJ, Anderson DJ. Identification of potential target genes for the neuron-restrictive silencer factor. *Proc Natl Acad Sci U S A*. 1996;93(18):9881–6.
- Yu HB, Johnson R, Kunarso G, Stanton LW. Coassembly of REST and its cofactors at sites of gene repression in embryonic stem cells. *Genome Res*. 2011;21(8):1284–93.
- Covey MV, Streb JW, Spektor R, Ballas N. REST regulates the pool size of the different neural lineages by restricting the generation of neurons and oligodendrocytes from neural stem/progenitor cells. *Development*. 2012;139(16):2878–90.
- Chen ZF, Paquette AJ, Anderson DJ. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet*. 1998;20(2):136–42.
- Mandel G, Fiondella CG, Covey MV, Lu DD, Loturco JJ, Ballas N. Repressor element 1 silencing transcription factor (REST) controls radial migration and temporal neuronal specification during neocortical development. *Proc Natl Acad Sci U S A*. 2011;108(40):16789–94.
- Su X, Kameoka S, Lentz S, Majumder S. Activation of REST/NRSF target genes in neural stem cells is sufficient to cause neuronal differentiation. *Mol Cell Biol*. 2004;24(18):8018–25.
- Lawinger P, Venugopal R, Guo ZS, et al. The neuronal repressor REST/NRSF is an essential regulator in medulloblastoma cells. *Nat Med*. 2000;6(7):826–31.
- Zuccato C, Tartari M, Crotti A, et al. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet*. 2003;35(1):76–83.
- Marullo M, Valenza M, Mariotti C, Di Donato S, Cattaneo E, Zuccato C. Analysis of the repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy of non-neuronal genes in peripheral lymphocytes from patients with Huntington's disease. *Brain Pathol*. 2010;20(1):96–105.
- González-Castañeda RE, Sánchez-González VJ, Flores-Soto M, Vázquez-Camacho G, Macías-Islas MA, Ortiz GG. Neural restrictive silencer factor and choline acetyltransferase expression in cerebral tissue of Alzheimer's Disease patients: A pilot study. *Genet Mol Biol*. 2013;36(1):28–36.
- Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009;10(10):669–80.
- Satoh J. Bioinformatics approach to identifying molecular biomarkers and networks in multiple sclerosis. *Clinical and Experimental Neuroimmunology. Neuroimmunol*. 2000;1(3):127–40.
- Satoh J, Tabunoki H. Molecular network of chromatin immunoprecipitation followed by deep sequencing-based vitamin D receptor target genes. *Mult Scler*. 2013;19(8):1035–45.
- Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012;7(2):e30619.
- Hatem A, Bozdağ D, Toland AE, Çatalyürek ÜV. Benchmarking short sequence mapping tools. *BMC Bioinformatics*. 2013;14:184.
- Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
- Machanic P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 2011;27(12):1696–7.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
- Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007;316(5830):1497–502.
- Westbrook TF, Hu G, Ang XL, et al. SCFbeta-TRCP controls oncogenic transformation and neural differentiation through REST degradation. *Nature*. 2008;452(7185):370–4.
- Bruce AW, Donaldson IJ, Wood IC, et al. Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A*. 2004;101(28):10458–63.
- Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res*. 2008;36(16):5221–31.
- Thiel G, Lietz M, Cramer M. Biological activity and modular structure of RE-1-silencing transcription factor (REST), a repressor of neuronal genes. *J Biol Chem*. 1998;273(41):26891–9.
- Zhang C, Xuan Z, Otto S, et al. A clustering property of highly-degenerate transcription factor binding sites in the mammalian genome. *Nucleic Acids Res*. 2006;34(8):2238–46.
- Johnson R, Richter N, Bogu GK, et al. A genome-wide screen for genetic variants that modify the recruitment of REST to its target genes. *PLoS Genet*. 2012;8(4):e1002624.
- Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell*. 2005;121(4):645–57.
- Lonze BE, Ginty DD. Function and regulation of CREB family transcription factors in the nervous system. *Neuron*. 2002;35(4):605–23.
- Zhang J, Wang S, Yuan L, et al. Neuron-restrictive silencer factor (NRSF) represses cocaine- and amphetamine-regulated transcript (CART) transcription and antagonizes cAMP-response element-binding protein signaling through a dual NRSE mechanism. *J Biol Chem*. 2012;287(51):42574–87.
- Kreissler A, Strissel PL, Strick R, Neumann SB, Schumacher U, Becker CM. Regulation of the NRSF/REST gene by methylation and CREB affects the cellular phenotype of small-cell lung cancer. *Oncogene*. 2010;29(43):5828–38.
- Satoh J, Tabunoki H, Arima K. Molecular network analysis suggests aberrant CREB-mediated gene regulation in the Alzheimer disease hippocampus. *Dis Markers*. 2009;27(5):239–52.
- Paquette AJ, Perez SE, Anderson DJ. Constitutive expression of the neuron-restrictive silencer factor (NRSF)/REST in differentiating neurons disrupts neuronal gene expression and causes axon pathfinding errors in vivo. *Proc Natl Acad Sci U S A*. 2000;97(22):12318–23.
- Kazemi S, Mounir Z, Baltzis D, et al. A novel function of eIF2alpha kinases as inducers of the phosphoinositide-3 kinase signaling pathway. *Mol Biol Cell*. 2007;18(9):3635–44.
- Johnson R, Teh CH, Kunarso G, et al. REST regulates distinct transcriptional networks in embryonic and neural stem cells. *PLoS Biol*. 2008;6(10):e256.
- Calderone A, Jover T, Noh KM, et al. Ischemic insults derepress the gene silencer REST in neurons destined to die. *J Neurosci*. 2003;23(6):2112–21.
- Westbrook TF, Martin ES, Schlabach MR, et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell*. 2005;121(6):837–48.
- McClelland S, Flynn C, Dubé C, et al. Neuron-restrictive silencer factor-mediated hyperpolarization-activated cyclic nucleotide-gated channelopathy in experimental temporal lobe epilepsy. *Ann Neurol*. 2011;70(3):454–64.
- Kuwahara K, Saito Y, Takano M, et al. NRSF regulates the fetal cardiac gene program and maintains normal cardiac structure and function. *EMBO J*. 2003;22(23):6310–21.