ORIGINAL RESEARCH

Pathway Analysis of ChIP-Seq-Based NRF1 Target Genes Suggests a Logical Hypothesis of their Involvement in the Pathogenesis of Neurodegenerative Diseases

Jun-ichi Satoh, Natsuki Kawana and Yoji Yamamoto

Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Kiyose, Tokyo, Japan.
Corresponding author email: satoj@my-pharm.ac.jp

Abstract: Nuclear respiratory factor 1 (NRF1) serves as a transcription factor that activates the expression of a wide range of nuclear genes essential for mitochondrial biogenesis and function, including mitochondrial respiratory complex subunits, heme biosynthetic enzymes, and regulatory factors involved in the replication and transcription of mitochondrial DNA. Increasing evidence indicates that mitochondrial function is severely compromised in the brains of aging-related neurodegenerative diseases. To identify the comprehensive set of human NRF1 target genes potentially relevant to the pathogenesis of neurodegenerative diseases, we analyzed the NRF1 chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) dataset retrieved from the Encyclopedia of DNA Elements (ENCODE) project. Overall, we identified 2,470 highly stringent ChIP-Seq peaks on protein-coding genes in SK-N-SH human neuroblastoma cells. They were accumulated in the proximal promoter regions with an existence of the NRF1-binding consensus sequence. The set of ChIP-Seq-based NRF1 target genes included known NRF1 targets such as EIF2S1, EIF2S2, CYCS, FMR1, FXR2, E2F6, CD47, and TOMM34. By pathway analysis, the molecules located in the core pathways related to mitochondrial respiratory function were determined to be highly enriched in NRF1 target genes. Furthermore, we found that NRF1 target genes play a pivotal role in regulation of extra-mitochondrial biological processes, including RNA metabolism, splicing, cell cycle, DNA damage repair, protein translation initiation, and ubiquitin-mediated protein degradation. We identified a panel of neurodegenerative disease-related genes, such as PARK2 (Parkin), PARK6 (Pink1), PARK7 (DJ-1), and PAELR (GPR37) for Parkinson's disease, as well as PSENEN (Pen2) and MAPT (tau) for Alzheimer's disease, as previously unrecognized NRF1 targets. These results suggest a logical hypothesis that aberrant regulation of NRF1 and its targets might contribute to the pathogenesis of human neurodegenerative diseases via perturbation of diverse mitochondrial and extra-mitochondrial functions.

Keywords: Alzheimer's disease, binding sites, ChIP-Seq, GenomeJack, NRF1, Parkinson's disease

Introduction

Mitochondria play a pivotal role in adenosine triphosphate (ATP) synthesis by oxidative phosphorylation in nearly all eukaryotic cells. Mitochondrial function is severely compromised in the brains of aging-related neurodegenerative diseases, such as Parkinson's disease (PD), Alzheimer's disease (AD), and Huntington's disease (HD).[1,2] The precise genetic and molecular mechanisms accounting for mitochondrial dysfunction in human neurodegenerative diseases, however, remain largely unknown. Mitochondrial DNA (mtDNA), a small circular DNA, encodes only 13 respiratory chain subunits, 22 transfer RNAs, and 2 ribosomal RNAs. In contrast, all proteins essential for mitochondrial transcription, translation, replication, and protein import and assembly are coded by a series of nuclear genes.[3] Nuclear respiratory factor 1 (NRF1), alternatively named alpha-palindromic-binding protein (α-PAL), serves as a basic region-leucine zipper (bZIP) transcription factor which activates the expression of nuclear genes essential for mitochondrial biogenesis and function.[3]

NRF1 target genes include a wide range of genes coding for five mitochondrial respiratory complex subunits, heme biosynthetic enzymes, and regulatory factors involved in the replication and transcription of mtDNA.[3] NRF1 is composed of the N-terminal serine phosphorylation domain and the nuclear localization signal (NLS), the C-terminal transactivation domain, and the central DNA binding domain which recognizes the palindromic sequence defined as YGCGCAYGCGCR, where Y and R indicate a pyrimidine or purine nucleotide, respectively.[3] A homodimer of NRF1 binds to target gene promoters.[4] Phosphorylation of NRF1 at the N-terminal serine residues induced in response to the serum enhances DNA binding and transactivation abilities.[4]

The human NRF1 gene shows a substantial homology to developmental transcription factors, such as P3A2 expressed in sea urchins, the *erect wing (ewg)* protein in *Drosophila, not really finished (nrf)* in Zebrafish, and *Nrf1* in mice.[5–8] The homozygous *Nrf1* knockout mice revealed an embryonic lethal phenotype and NRF1−/− blastocysts exhibited reduced levels of mtDNA.[7] Disruption of NRF1 ortholog functions causes a severe neurological defect in *Drosophila* and Zebrafish, suggesting a crucial role of NRF1 in maintenance of the nervous system function.[6,8]

Because mitochondria contain a number of enzymes pivotal for redox homeostasis, even a subtle defect in oxidative phosphorylation causes not only fatal energy failure in neurons but also generates a large amount of reactive oxygen species (ROS) highly toxic to vulnerable neurons in the central nervous system (CNS).[1] Therefore, defective mitochondrial physiology plays a crucial role in the pathogenesis of neurodegenerative diseases, such as PD, AD and HD in which mitochondrial function is compromised at the early stage of the diseases.[1,2] With respect to a crucial role of NRF1 in the maintenance of mitochondrial function, it is highly important to characterize the comprehensive set of NRF1 target genes in human neural cells involved in mitochondrial function relevant to the pathogenesis of human neurodegenerative diseases.

Recently, the next-generation sequencing (NGS) technology has rapidly advanced the genome research. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq), an NGS application, provides a highly efficient method for global profiling of DNA-binding proteins, histone modifications, and nucleosomes on a genome-wide scale.[9] ChIP-Seq endowed with higher resolution, less noise, and greater coverage of the genome, compared with microarray-based ChIP-Chip, provides an invaluable tool for characterizing the comprehensive gene regulatory networks.[9,10] Since ChIP-Seq produces large amounts of high-throughput data, it is often hampered by the difficulty in interpretation of biologically meaningful implications. To overcome this problem, we could illustrate the cell-wide map of the complex molecular interactions with the help of the knowledgebase of well-annotated molecular pathways to establish a logical hypothesis accounting for cellular biological processes.[11]

No previous studies have characterized the comprehensive set of ChIP-Seq-based NRF1 target genes in human neural cells. At present, only one study is available, which determined the genome-wide NRF1 target genes by ChIP-Chip in T98G human glioblastoma cells.[12] Recently, the Encyclopedia of DNA Elements (ENCODE) project based on an international collaborative research was established to identify a comprehensive list of functional elements in the human genome, including genes, transcripts, and transcriptional regulatory elements

(encodeproject.org/ENCODE).[13] A panel of high-throughput functional genomic data produced by the ENCODE project are currently available through a freely accessible database following the ENCODE Consortium Data Release restriction. They serve as a high-quality resource for the scientific community to find novel biological insights into the complex gene regulation in the human genome. In the present study, we attempted to analyze the NRF1 ChIP-Seq dataset retrieved from the database of the ENCODE project. Overall, we identified 2,470 ChIP-Seq-based reliable NRF1 targets in SK-N-SH human neuroblastoma cells, including the genes essential for mitochondrial function and those closely related to neurodegeneration. Based on these observations, we would propose an original hypothesis that deregulation of NRF1 target genes are potentially involved in the pathogenesis of human neurodegenerative diseases.

Methods
ChIP-Seq dataset of NRF1 binding sites
To identify a comprehensive set of NRF1 target genes, we investigated a ChIP-Seq dataset retrieved from DDBJ Sequence Read Archive (DRA) under the accession number of SRP007993. The ChIP-Seq experiments were performed for the ENCODE project (encodeproject.org/ENCODE) by researchers in Dr. Michel Snyder's Laboratory, Stanford University. The data were opened to public on September 13, 2012. In these experiments, SK-N-SH human neuroblastoma cells (ATCC HTB-11) cultured in Eagle's Minimum Essential Medium supplemented with 10% fetal bovine serum without any special treatments (encodeproject.org/ENCODE/cellTypes.html) were fixed with formaldehyde to crosslink NRF1-DNA complexes. Sonicated nuclear lysates were processed for immunoprecipitation with a mouse monoclonal anti-NRF1 antibody (ab55744; Abcam) (SRX186635) or normal IgG for input signal (SRX186629). NGS libraries constructed from size-selected and adapter-ligated ChIP DNA fragments were processed for deep sequencing at a 36 bp read length on Genome Analyzer (Illumina).

First, we evaluated the quality of NGS short reads by searching them on the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Then, we mapped them on the human genome reference sequence version hg19 by using a mapping tool of AvadisNGS (Agilent Technologies) or by the Bowtie 0.12.7 program (bowtie-bio.sourceforge.net). Subsequently, we identified statistically significant peaks of mapped reads by using the Model-based Analysis of ChIP-Seq (MACS) program under the highly stringent condition of fold enrichment (FE) ≥ 20 and the false discovery rate (FDR) ≤ 0.01, according to the methods described previously.[14] The genes corresponding to MACS peaks were determined by a neighboring gene analysis tool of AvadisNGS within a distance of 5000 bp from peaks to genes. Next, we identified more detailed genomic locations of MACS peaks by importing the processed data composed of individual peaks one by one into a novel genome viewer named GenomeJack v1.4 (Mitsubishi Space Software).[15] The genomic locations of the peaks were categorized into the following: the promoter region defined by the location within 5,000 bp upstream from the 5′ end of genes, the 5′ untranslated region (5′UTR), the exon, the intron, the 3′UTR, the location within 5,000 bp downstream from the 3′ end of genes (3′down), and intergenic regions outside these. In the technical aspects, we defined NRF1 target genes as those having ChIP-Seq peaks within the distance of 5,000 bp upstream or downstream from 5′ or 3′ ends of the genes whose sequences are defined by RefSeqGene (http://www.ncbi.nlm.nih.gov/refseq/rsg). The consensus motif sequences surrounding a 400 bp-length sequence of the peak summits, extracted from the hg19 sequence track of GenomeJack, were identified by the MEME-ChIP program.

Molecular pathway analysis
To identify the pathways biologically relevant to ChIP-Seq-based NRF1 target genes, we imported the corresponding Entrez Gene IDs into the Functional Annotation tool of Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (david.abcc.ncifcrf.gov).[16] DAVID identifies the most relevant pathway constructed by Kyoto Encyclopedia of Genes and Genomes (KEGG), composed of the genes enriched in the given set with an output of statistical significance evaluated by the modified Fisher's exact test and corrected by Benjamini-Hochberg multiple comparison test. KEGG (http://www.kegg.jp) is a publicly accessible knowledgebase containing manually curated reference pathways that cover a wide range of metabolic, genetic, environmental, and

cellular processes, as well as human diseases. It is currently composed of 271,665 pathways generated from 451 reference pathways.[17]

We also imported Entrez Gene IDs into Ingenuity Pathways Analysis (IPA) (Ingenuity Systems; http://www.ingenuity.com) in the setting of 140 molecules per net for greater than 1,000 imported genes and 70 molecules per net for smaller than 1,000 imported genes. IPA is a commercial knowledgebase that contains approximately 3,000,000 biological and chemical interactions and functional annotations 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 or networks by the Fisher's exact test.

Results

Identification of 2,470 ChIP-Seq-based NRF1 target genes

First, we evaluated the quality of short read NGS data derived from NRF1 ChIP-treated DNA, composed of 50,295,303 reads, and IgG ChIP-treated DNA, composed of 39,245,635 reads. The quality scores across all bases exceeded 30 on FastQC, indicating that these data showed an acceptable quality for downstream analysis (Supplementary Fig. 1, panels a and b). After mapping them on hg19 by AvadisNGS, we identified all 2,792 stringent ChIP-Seq peaks that met the highly stringent criteria of fold enrichment (FE) $\geq$20 and false discovery rate (FDR) $\leq$0.01 by MACS in order to reduce the detection of false positive binding sites, if at all possible.[14] To process them for the pathway analysis, we omitted the peaks located in non-coding genes and uncategorized genes (n = 322). Finally, we identified 2,470 ChIP-Seq peaks on protein-coding genes. The genomic locations of the peaks were determined by using GenomeJack. Fragile X mental retardation protein 1 (FMR1) (Fig. 1, panels a and b) represents an example of well-known NRF1 target genes.[18]

The summits of the peaks of top 200 genes were located in the promoter (68.0%), 5′UTR (18.5%), exon (2.0%), intron (7.5%), 3′UTR (0.5%), and 3′down (3.5%) by GenomeJack, suggesting that the binding regions are accumulated in the proximal promoter regions. The comprehensive list of 2,470 NRF1 target genes is shown in Supplementary Table 1. The top 10 most significant genes are listed in Table 1. Importantly, among 691 NRF1 target genes reported by a previous ChIP-Chip study of T98G human glioblastoma cells,[12] the set of 495 genes (71.6%) (Supplementary Table 1, underline), including eukaryotic translation initiation factor 2, subunit 1 alpha (EIF2S1), ataxia telangiectasia mutated (ATM), cyclin B1 (CCNB1), and proliferating cell nuclear antigen (PCNA), were overlapped with the set of 2,470 ChIP-Seq-based NRF1 target genes, supporting the validity of our results. However, a panel of neurodegenerative disease-related genes identified in SK-N-SH cells, such as PARK2 (Parkin), PARK7 (DJ-1), PAELR (GPR37), MAPT (tau), and TAF4 (TAFII130), were not included in ChIP-chip-based NRF1 targets of T98G cells.

By motif analysis using MEME-ChIP, 99 genes (99%) among top 100 genes exhibited an existence of the NRF1-binding element comprising (T/C)GCGCA(C/T)GCGC(A/G/C) (E-value = 2.6e-195) (Fig. 2), validating the specificity of mapping of ChIP-Seq short reads to genomic regions with the NRF1-binding consensus sequence motif.

Molecular pathways of ChIP-Seq-based NRF1 target genes

Next, we studied molecular pathways of the set of 2,470 NRF1 target genes by using pathway analysis tools of bioinformatics. By using DAVID, we identified functionally associated gene ontology (GO) terms (Supplementary Table 2). The most significant GO terms included "RNA processing" (GO:0006396; P = 6.59E-12) for biological process, "intracellular organelle lumen" (GO:0070013; P = 1.08E-38) for cellular component, and "nucleotide binding" (GO:0000166; P = 3.86E-10) for molecular function.

By using KEGG, NRF1 target genes showed top 3 closest associations with "RNA degradation" (hsa03018; P = 0.00032 corrected by multiple comparison test), "Cell cycle" (hsa04110; P = 0.00041) (Fig. 3), and "Spliceosome" (hsa03040; P = 0.00238), suggesting that NRF1 target genes play a key role in RNA metabolism, splicing, and cell cycle control (Table 2). Importantly, KEGG pathways defined by "Aminoacyl-tRNA biosynthesis" (hsa00970; P = 0.01104) involved in protein translation,
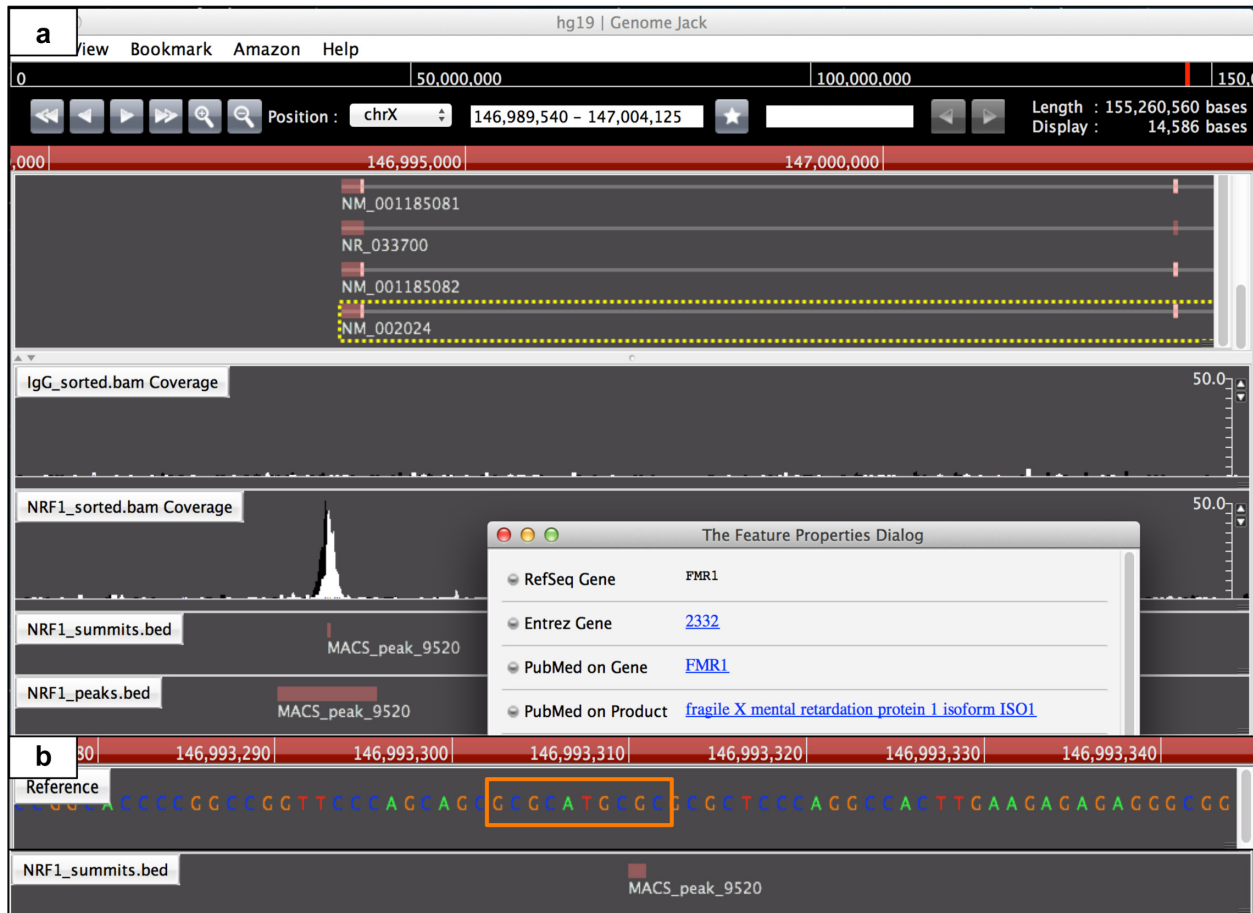
Figure 1. Genomic locations of ChIP-Seq peaks in FMR1. By analyzing the ChIP-Seq dataset of NRF1-binding sites, we identified totally 2,470 stringent peaks showing fold enrichment (FE) ≥20 and FDR ≤ 0.01. The genomic locations of the peaks were determined by importing the processed data into GenomeJack. An example of fragile X mental retardation protein 1 (FMR1; Entrez Gene ID 2332), where a MACS peak numbered 9520 in the NRF1_ sorted.bam coverage lane is located in the promoter region of FMR1 (panel a) with a NRF1-binding consensus sequence motif highlighted by orange square (panel b).

"Oxidative phosphorylation" (hsa00190; $P = 0.01545$) involved in mitochondrial respiratory chain, "Base excision repair" (hsa03410; $P = 0.01947$) involved in DNA damage repair, and "Ubiquitin mediated proteolysis" (hsa04120; $P = 0.04020$) involved in protein degradation were also listed as significant pathways (Table 2). We also found that the set of 495 genes overlapped between T98G cells by ChIP-Seq and SK-N-SH cells by ChIP-Seq showed a significant relationship with "Aminoacyl-tRNA biosynthesis" (hsa00970; $P = 0.01597$), "Splicesome" (hsa03040; $P = 0.02612$), and "Oxidative phosphorylation" (hsa00190; $P = 0.02917$), whereas the set of 196 genes identified only in T98G did not show significant relationship with KEGG pathways.

Importantly, NRF1 target genes showed a significant relationship with "Parkinson's disease" pathway on KEGG (hsa05012; $P = 0.03561$) (Fig. 2, Table 2),

where PARK2 (Parkin), PARK6 (Pink1), PARK7 (DJ-1), and PAELR (GPR37) were matched with the previously unrecognized set of NRF1 target genes. We also found that the set of 28 genes (16.5%) among 170 Alzheimer's disease-related genes (Supplementary Fig. 2) and 36 genes (21.2%) among 183 Huntington's disease-related genes (Supplementary Fig. 3) on KEGG pathways represented ChIP-Seq-based NRF1 target genes, although the enrichment did not reach the statistical significance. However, these observations raise the possibility that aberrant regulation of NRF1 and its targets might contribute to neurodegenerative processes not only of PD but also of AD and HD.

Finally, we studied molecular pathways of 2,470 ChIP-Seq-based NRF1 target genes by using the core analysis tool of IPA (Table 3). We identified "Mitochondrial dysfunction" ($P = 2.80E-06$) (Fig. 5)

Table 1. Top 10 ChIP-Seq-based NRF1 target genes.

Chromosome	Start	End	FE	FDR	Entrez gene ID	Gene symbol	Gene name	Peak location	Consensus sequence
chr6	169613018	169613018	109.41	0	7058	THBS2	Thrombospondin 2	3'down	+
chr17	21029838	21030524	108.60	0	25979	DHRS7B	Dehydrogenase/reductase (SDR family) member 7B	Promoter	+
chr14	39583182	39583824	101.02	0	8487	GEMIN2	Gem (nuclear organelle) associated protein 2	5'UTR	+
chr1	1709580	1710275	100.63	0	65220	NADK	NAD kinase	Promoter	+
chrX	152864501	152865114	99.83	0	92002	FAM58A	Family with sequence similarity 58, member A	Promoter	+
chr6	44094936	44095311	98.40	0	64928	MRPL14	Mitochondrial ribosomal protein L14	5'UTR	+
chr11	777285	777883	98.33	0	347862	PDDC1	Parkinson disease 7 domain containing 1	Promoter	+
chr3	43327657	43328320	95.70	0	54861	SNRK	SNF related kinase	5'UTR	+
chr4	967074	967797	95.50	0	1609	DGKQ	Diacylglycerol kinase, theta 110 kDa	Promoter	+
chr22	39052359	39052916	93.25	0	25776	CBY1	Chibby homolog 1 (Drosophila)	Promoter	+

Notes: From the ChIP-Seq dataset, we identified 2,470 NRF1 target genes in SK-N-SH cells based on stringent criteria that satisfy both false discovery rate (FDR) ≦0.01 and fold enrichment (FE) ≧20. Top 10 genes based on FE are listed with the chromosome, the position (start, end), FE, FDR, Entrez Gene ID, Gene Symbol, Gene Name, location of the peak summit, and presence of the consensus sequence.

and "Regulation of eIF4 and p70S6K signaling" ($P = 7.50E-06$) (Fig. 6) as the most significant canonical pathways associated with the set of imported genes. Again, these results indicate that NRF1 target genes play a key role in maintenance of mitochondrial function and protein translation initiation. Furthermore, the canonical pathways defined as "Protein ubiquitination pathway" ($P = 2.85E-05$), "DNA double-strand break repair by non-homologous end joining" ($P = 3.28E-05$), and "EIF2 signaling" ($P = 3.55E-05$) were also listed as significant pathways. This supports the notion that NRF1 target genes regulate a wide range of extra-mitochondrial biological processes, including protein translation, degradation, and DNA damage repair. The set of 495 genes overlapped between T98G and SK-N-SH cells showed a highly significant relationship with "Mitochondrial dysfunction" ($P = 9.53E-07$), while the set of 196 genes identified only in T98G were less significantly related to "Mitochondrial dysfunction" ($P = 3.90E-03$).

Discussion

Mitochondrial function is severely compromised in the brains of aging-related neurodegenerative diseases, such as PD, AD, and HD.[1,2] However, at present, the precise genetic and molecular mechanisms and causes responsible for mitochondrial dysfunction in human neurodegenerative diseases remain mostly unknown. NRF1 which activates transcription of nuclear-encoded mitochondrial proteins plays an integrative role in nuclear-mitochondrial interactions.[3] ChIP-Seq showed a great advantage over ChIP-chip in resolution and coverage of the genome when we characterize the comprehensive gene regulatory networks.[9,10] We, for the first time, characterized the comprehensive set of 2,470 ChIP-Seq-based NRF1 target genes in SK-N-SH human neuroblastoma cells. The binding sites were concentrated in the promoter and 5′UTR regions with an existence of the NRF1-binding consensus sequence.

The set of ChIP-Seq-based NRF1 target genes included a battery of known NRF1 targets, such as eukaryotic translation initiation factor 2, subunit 1 alpha (EIF2S1),[19] eukaryotic translation initiation factor 2, subunit 2 beta (EIF2S2),[20] cytochrome c, somatic (CYCS),[21] fragile X mental retardation 1 (FMR1),[18] fragile X mental retardation,

Figure 2. NRF1-binding consensus sequence motif. The consensus motif sequences were identified by importing a 400 bp-length sequence surrounding the summit of MACS peaks of top 100 genes based on fold enrichment into the MEME-ChIP program.

autosomal homolog 2 (FXR2),[22] E2F transcription factor 6 (E2F6),[23] CD47,[24] and translocase of outer mitochondrial membrane 34 (TOMM34).[25] By GO analysis with DAVID, the most significant GO terms of 2,470 NRF1 target genes were "RNA processing,"

"intracellular organelle lumen," and "nucleotide binding." By using pathway analysis with KEGG and IPA, we found that the molecules located in the core pathways related to mitochondrial respiratory function, designated as "Oxidative phosphorylation" in
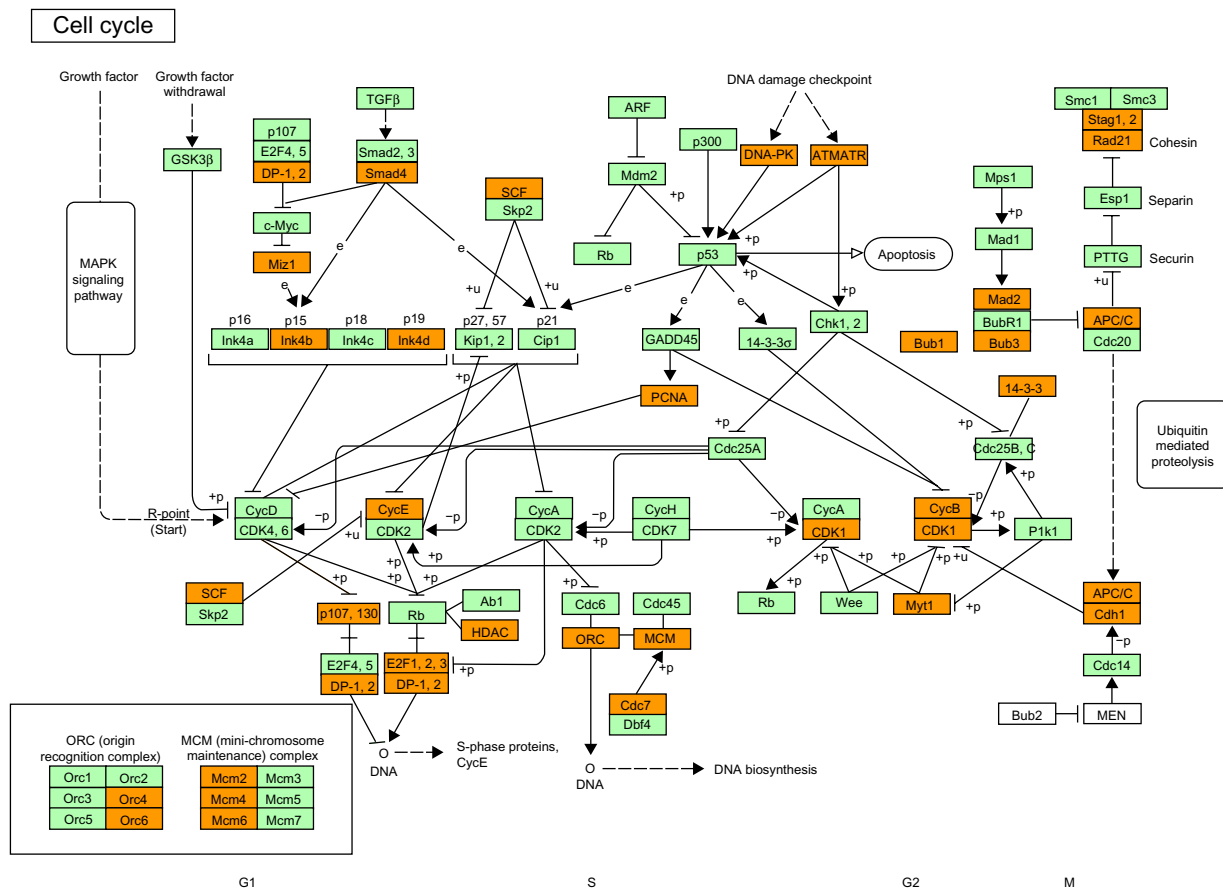

Figure 3. KEGG "cell cycle" pathway relevant to NRF1 target genes. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Functional Annotation tool of DAVID. It extracted the KEGG "Cell cycle" pathway (hsa04110) as the second rank significant pathway as listed in Table 2. NRF1 target genes are highlighted by orange.

Table 2. Top 10 KEGG pathways relevant to 2,470 ChIP-Seq-based NRF1 target genes genes.

Rank	Category	Genes in the Pathway	*P*-value
1	hsa03018: RNA degradation	CNOT1, CNOT4, DCPS, ENO1, EXOSC1, EXOSC10, EXOSC2, EXOSC5, EXOSC6, HSPA9, HSPD1, LSM1, LSM2, LSM4, LSM7, MPHOSPH6, NAA38, PAPD7, PARN, PATL1, TTC37, ZCCHC7	0.00032
2	hsa04110: cell cycle	ANAPC1, ANAPC11, ANAPC13, ANAPC2, ANAPC7, ATM, BUB1, BUB3, CCNB1, CCNE1, CDC7, CDK1, CDKN2B, CDKN2D, E2F1, FZR1, HDAC2, MAD2l1, MCM2, MCM4, MCM6, ORC4L, ORC6L, PCNA, PKMYT1, PRKDC, RAD21, RBL2, RBX1, SMAD4, STAG1, TFDP1, YWHAH, YWHAZ, ZBTB17	0.00041
3	hsa03040: spliceosome	ACIN1, BAT1, CCDC12, DDX42, DHX15, HNRNPA1, HNRNPM, HNRNPU, HSPA8, LSM2, LSM4, LSM7, NAA38, PRPF3, RBM17, SF3A2, SF3B1, SF3B2, SFRS1, SFRS13A, SFRS4, SFRS9, SLU7, SNRNP70, SNRPD2, SNW1, SYF2, THOC3, THOC4, U2AF1, U2AF2, USP39, ZMAT2	0.00238
4	hsa00970: aminoacyl-tRNA biosynthesis	FARSB, HARS, HARS2, KARS, MARS2, MTFMT, NARS2, PSTK, SARS, SARS2, SEPSECS, TARS, VARS2, WARS2, YARS2	0.01104
5	hsa00190: oxidative phosphorylation	ATP5G2, ATP5I, ATP5O, ATP6V0D1, ATP6V0E2, ATP6V1C1, ATP6V1D, ATP6V1F, ATP6V1G1, COX15, COX5B, COX6A2, COX6C, NDUFA11, NDUFA3, NDUFA6, NDUFA8, NDUFB10, NDUFB4, NDUFS6, NDUFS8, NDUFV3, PPA1, SDHA, SDHB, SDHD, UQCR10, UQCR11, UQCRB, UQCRC1, UQCRFS1	0.01545
6	hsa03410: base excision repair	APEX2, LIG3, MBD4, MPG, NEIL2, NEIL3, PARP3, PCNA, POLD1, POLE4, TDG, UNG, XRCC1	0.01947
7	hsa00240: pyrimidine metabolism	AK3, CANT1, CTPS, DHODH, DTYMK, ENTPD5, ENTPD6, NT5C, NT5C3, POLA1, POLD1, POLE4, POLR1A, POLR1C, POLR2A, POLR2F, POLR2H, POLR2I, POLR3B, POLR3H, POLR3K, RRM2B, TYMP, UCK1	0.02517
8	hsa05012: Parkinson's disease	ATP5G2, ATP5O, COX5B, COX6A2, COX6C, CYCS, GPR37, NDUFA3, NDUFA6, NDUFA8, NDUFB10, NDUFB4, NDUFS6, NDUFS8, NDUFV3, PARK2, PARK7, PINK1, SDHA, SDHB, SDHD, SLC25A6, UQCR10, UQCR11, UQCRB, UQCRC1, UQCRFS1, VDAC2, VDAC3	0.03561
9	hsa04120: ubiquitin mediated proteolysis	ANAPC1, ANAPC11, ANAPC13, ANAPC2, ANAPC7, BTRC, CDC34, FANCL, FZR1, HERC4, KEAP1, PARK2, PIAS4, RBX1, RFWD2, SIAH1, SOCS1, STUB1, TCEB, TRIM32, TRIM37, UBA2, UBE2D3, UBE2D4, UBE2K, UBE2M, UBE2N, UBE4B, VHL, WWP1	0.04020
10	hsa04722: neurotrophin signaling pathway	AKT2, ARHGDIA, CRK, CRKL, CSC42, CSK, HRAS, IRAK4, IRS1, MAGED1, MAP2K7, MAPK12, MAPK7, PLCG1, PRDM4, RAC1, RAF1, RELA, RHOA, RIPK2, RPS6KA5, RPS6KA6, SHC1, SHC4, SOS1, SOS2, YWHAH, YWHAZ	0.04082

Notes: By importing Entrez Gene IDs of 2,470 ChIP-Seq-based NRF1 target genes into the Functional Annotation tool of DAVID, KEGG pathways showing significant relevance to the set of imported genes were identified. They are listed with *P*-value corrected by Benjamini-Hochberg multiple comparison test.

KEGG and "Mitochondrial dysfunction" in IPA, are enriched in NRF1 target genes. They include a panel of cytochrome c oxidase (COX), NADH dehydrogenase, succinate dehydrogenase, and ubiquinol-cytochrome c reductase genes. Furthermore, we found that NRF1 target genes play a pivotal role in regulation of extra-mitochondrial biological processes, such as RNA metabolism, splicing, cell cycle, DNA damage repair, protein translation initiation, and ubiquitin-mediated protein degradation, all of which are pivotal for neuronal and non-neuronal cell growth, differentiation, and survival. However, these biological processes are most remarkably deregulated in neurons in humans affected by neurodegenerative diseases closely associated with mitochondrial dysfunction.[26,27] It is worthy to note that stable expression of NRF1 induces neurite outgrowth in IMR32 human neuroblastoma cells.[28]

Accumulating evidence indicated that NRF1 target genes are not restricted to the genes involved in mitochondrial function.[3] Previous studies with a non-comprehensive approach identified several NRF1-binding genes encoding components of signaling pathways, and metabolic enzymes, along with the genes essential for chromosome maintenance and
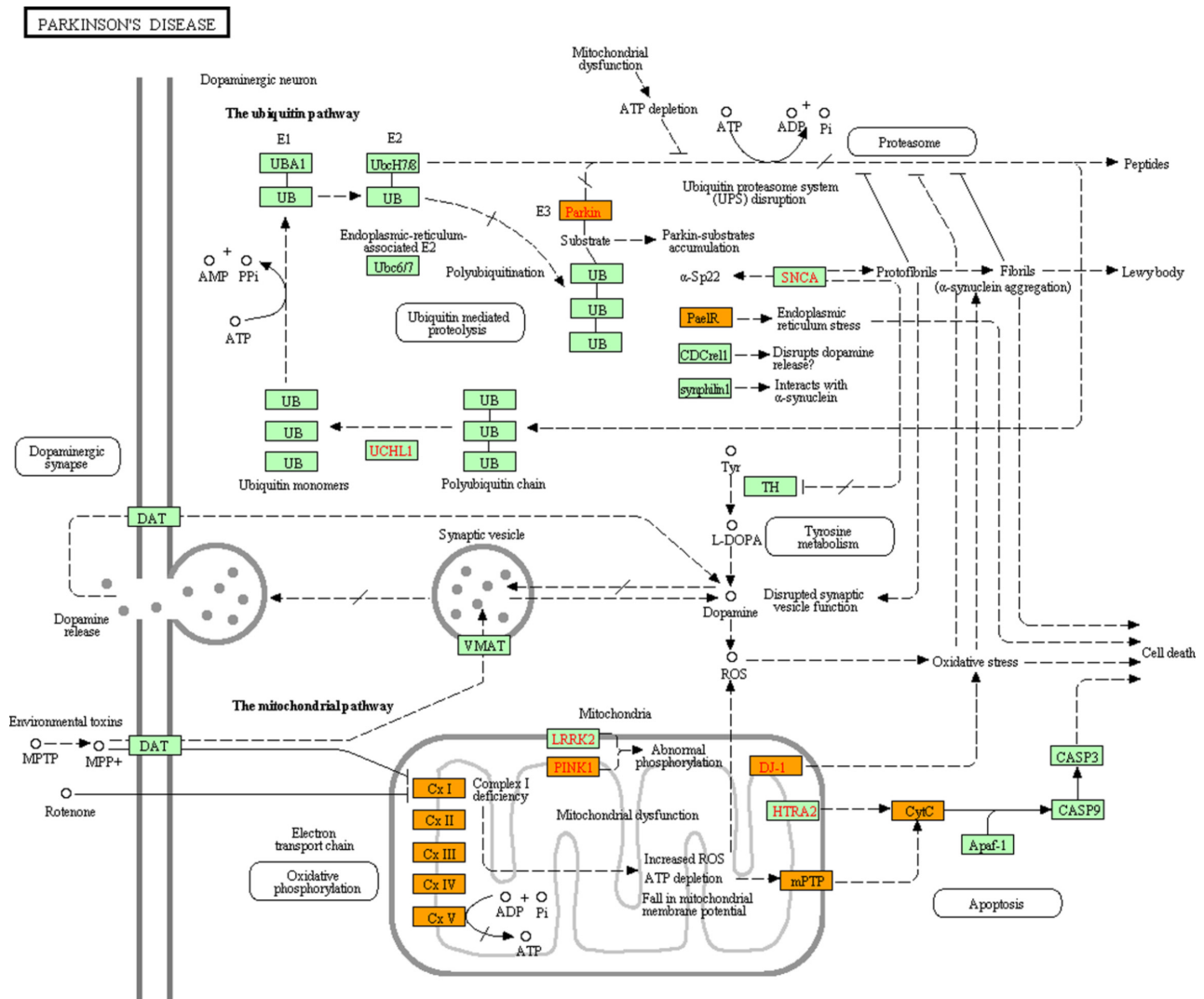
Figure 4. KEGG "Parkinson's disease" pathway relevant to NRF1 target genes. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Functional Annotation tool of DAVID. It extracted the KEGG "Parkinson's disease" pathway (hsa04110) as the eighth rank significant pathway as listed in Table 2. NRF1 target genes are highlighted by orange.

nuclei acid metabolism.[5,29] More recently, a ChIP-chip study of T98G cells showed that a number of NRF1 target genes, particularly related to DNA replication, mitosis, and cytokinesis, were also recognized by the growth regulatory transcription factor E2F, suggesting that NRF1 plays an active role in regulation of a subset of E2F-responsive genes.[12] The small interfering RNA targeting NRF1 reduced expression of a panel of NRF1 targets and E2F targets, indicating that NRF1 generally functions as a transcriptional activator.[3,12]

We found that a subset of NRF1 target genes show a significant relationship with the "Parkinson's disease" pathway on KEGG where PARK2 (Parkin), PARK6

(Pink1), PARK7 (DJ-1), and PAELR (GPR37), along with numerous mitochondrial respiratory complex components, are located in the PD pathway. PD clinically causes progressive motor and autonomic dysfunction, and is pathologically characterized by a loss of nigral dopaminergic neurons linked with accumulation of cytoplasmic inclusions named Lewy bodies immunoreactive for α-synuclein. Approximately 10% of PD patients are derived from genetic mutations.[30] Mutations in PARK2, PARK6, and PARK7, all of which are closely associated with mitochondrial dysfunction, cause autosomal recessive juvenile PD. Parkin plays a pivotal role in clearance of damaged mitochondria by mitophagy.[31] Pink1 attaches to the

Table 3. Top 10 IPA canonical pathways relevant to 2,470 ChIP-Seq-based NRF1 target genes genes.

Rank	Category	Genes in the pathway	*P*-value
1	Mitochondrial dysfunction	AIFM1, ATP5G2, ATPAF2, COX15, COX5B, COX6A2, COX6C, CPT1C, CYB5A, CYCS, DHODH, GSR, MAPK12, NDUFA11, NDUFA12, NDUFA3, NDUFA6, NDUFA8, NDUFB10, NDUFB4, NDUFS6, NDUFS8, NDUFV3, PARK2, PARK7, PINK1, PRDX3, PRDX5, PSENEN, SDHA, SDHB, SDHD, UQCR10, UQCR11, UQCRB, UQCRC1, UQCRFS1	2.80E-06
2	Regulation of eIF4 and p70S6K signaling	AGO1, AKT2, ATM, EIF1, EIF1AX, EIF2S1, EIF2S2, EIF3E, EIF3K, EIF3L, EIF4E, EIF4G1, FAU, HRAS, IRS1, MAPK12, MKNK1, PAIP2, PPP2CA, PPP2R1A, PPP2R3B, PPP2R5C, RAF1, RPS12, RPS13, RPS16, RPS21, RPS23, RPS29, RPS5, RPS6KB1, RPS9, SHC1, SOS1, SOS2	7.50E-06
3	Mitotic roles of polo-like kinase	ANAPC1, ANAPC11, ANAPC13, ANAPC2, ANAPC7, CCNB1, CDC7, CDK1, FBXO5, FZR1, HSP90AA1, PKMYT1, PLK4, PPP2CA, PPP2R1A, PPP2R3B, PPP2R5C, PRC1, RAD21, SLK	1.94E-05
4	Protein ubiquitination pathway	ANAPC1, ANAPC11, ANAPC2, BAG1, BTRC, CDC34, DNAJB12, DNAJB2, DNAJB6, DNAJB9, DNAJC13, DNAJC25, DNAJC28, DNAJC3, DNAJC5, FZR1, HSP90AA1, HSPA14, HSPA4L, HSPA8, HSPA9, HSPD1, HSPE1, MED20, PARK2, PSMA2, PSMA3, PSMA4, PSMC4, RBX1, SACS, STUB1, TCEB2, THOP1, UBE2D3, UBE2D4, UBE2M, UBE2N, UBE2V1, UBE4B, USO1, USP10, USP11, USP12, USP21, USP22, USP28, USP37, USP39, VHL	2.85E-05
5	DNA double-strand break repair by non-homologous end joining	ATM, LIG3, PRKDC, RAD50, WRN, XRCC1, XRCC5, XRCC6	3.28E-05
6	EIF2 signaling	AGO1, AKT2, ATM, EIF1, EIF1AX, EIF2AK1, EIF2S1, EIF2S2, EIF3E, EIF3K, EIF3L, EIF4E, EIF4G1, FAU, HRAS, PPP1CB, RAF1, RPL10, RPL13A, RPL14, RPL19, RPL26L1, RPL36, RPL5, RPL7L1, RPL8, RPLP1, RPS12, RPS13, RPS16, RPS21, RPS23, RPS29, RPS5, RPS9, SHC1, SOS1, SOS2, UBA52	3.55E-05
7	DNA methylation and transcriptional repression signaling	CHD3, DNMT1, HDAC2, MBD3, MECP2, RBBP4, RBBP7, SAP18, SAP30	1.19E-04
8	Remodeling of epithelial adherens junctions	ACTB, ACTN4, APC, ARF6, ARPC1A, ARPC2, ARPC5, CBLL1, CTNNA2, DNM1L, DNM3, HGS, MAPRE3, SRC, TUBA1B, TUBB, TUBB4B, TUBG1	2.73E-04
9	tRNA charging	FARSB, HARS, HARS2, KARS, MARS2, NARS2, SARS, SARS2, TARS, VARS2, WARS2, YARS2	5.00E-04
10	Estrogen receptor signaling	CTBP1, ERCC2, GTF2A1, GTF2F1, GTF2H4, HDAC3, HRAS, MED18, MED20, MED4, PELP1, POLR2A, POLR2F, POLR2H, POLR2I, PRKDC, RAF1, SHC1, SMARCA4, SOS1, SOS2, SRC, TAF12, TAF4, TAF5, TAF6, TAF7	5.08E-04

Notes: By importing Entrez Gene IDs of 2,470 ChIP-Seq-based NRF1 target genes into the Core Analysis tool of IPA, canonical pathways showing significant relevance to the set of imported genes were identified. They are listed with *P*-value evaluated by Fisher's exact test.

outer mitochondrial membrane, where it facilitates recruitment of Parkin.[32] DJ-1 capable of interacting with α-synuclein, Parkin, and Pink1 acts as a central regulator of anti-oxidant defense.[33] GPR37, alternatively named Parkin-associated endothelin-like receptor (PAELR), encodes an orphan G protein-coupled receptor which serves as a substrate for an E3 ubiquitin-protein ligase Parkin when PAELR is expressed in an unfolded and insoluble form.[34]

We also found that a substantial subset of NRF1 target genes, mostly those related to mitochondrial respiratory function, are accumulated on "Alzheimer's disease" and "Huntington's disease" pathways of KEGG. AD is the most common cause of dementia worldwide. It affects the elderly population and is characterized by the hallmark pathology of widespread deposition of amyloid-β (Aβ) on senile plaques and formation of neurofibrillary tangles (NFT) containing
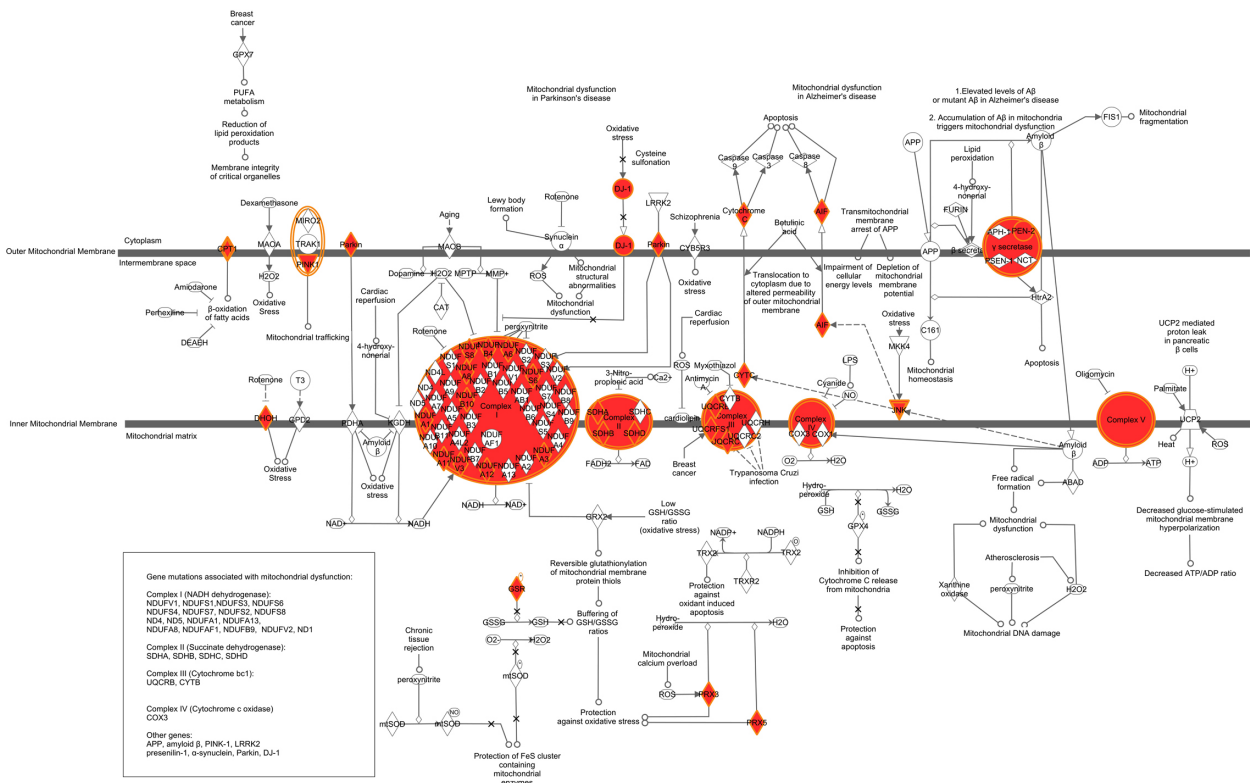
Figure 5. IPA "mitochondrial dysfunction" pathway relevant to NRF target genes. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Core Analysis tool of IPA. It extracted the "Mitochondrial dysfunction" pathway as the first rank significant pathway as listed in Table 3. NRF1 target genes are highlighted by red.

hyperphosphorylated tau in degenerating neurons. Importantly, Aβ, produced via sequential cleavage of amyloid precursor protein (APP) by β-secretase and γ-secretase, often accumulates on mitochondria and inhibits respiratory complex IV, thereby increasing ROS production and the frequency of mtDNA mutation.[35] We identified PSENEN (PEN2), a component of the γ-secretase complex and MAPT (tau) as two principal NRF1 target genes. Furthermore, a recent study showed that enhanced phosphorylation of eukaryotic initiation factor 2 α-subunit (eIF2α; EIF2S1), a prototype of NRF1 targets, inhibits protein translation, being responsible for defective synaptic plasticity in AD brains.[36]

HD, an intractable disease causing chorea and psychiatric disturbances due to expansion of a CAG trinucleotide repeat in the huntingtin (HTT) gene, is characterized by a loss of long projection neurons in the striatum and the cortex. Decreased activities of mitochondrial respiratory complexes II, III, and IV are found in the caudate nucleus of HD brains.[37] HTT interacts with a battery of transcription factors and coactivators important for diverse biological functions, such as p53, CREB-binding protein, and Sp1. We identified HTT-interacting TAF4 (TAFII130) as one of NRF1 targets, which are involved in the dopamine D2 receptor expression in cooperation with Sp1.[38] Notably, mutant HTT causes disruption of mitochondrial function by inhibiting expression of peroxisome proliferator-activated receptor gamma, coactivator 1 alpha (PGC-1α) which serves as a principal coactivator of NRF1.[39]

The present study has a major limitation in the interpretation of the results as we analyzed only a single ChIP-Seq dataset without experimental validation, retrieved from the ENCODE project, in which no biological replicates are currently available. However, by intensive in silico data mining from the pilot study, we could raise a logical hypothesis suggesting that aberrant regulation of NRF1 and its targets might contribute to neurodegenerative processes underlying PD, AD, and HD via perturbation of diverse mitochondrial and extra-mitochondrial functions. To evaluate this hypothesis, it is highly important to investigate a heretofore unrecognized role of NRF1 and NRF1 targets in PD, AD, and HD brains and model animals.
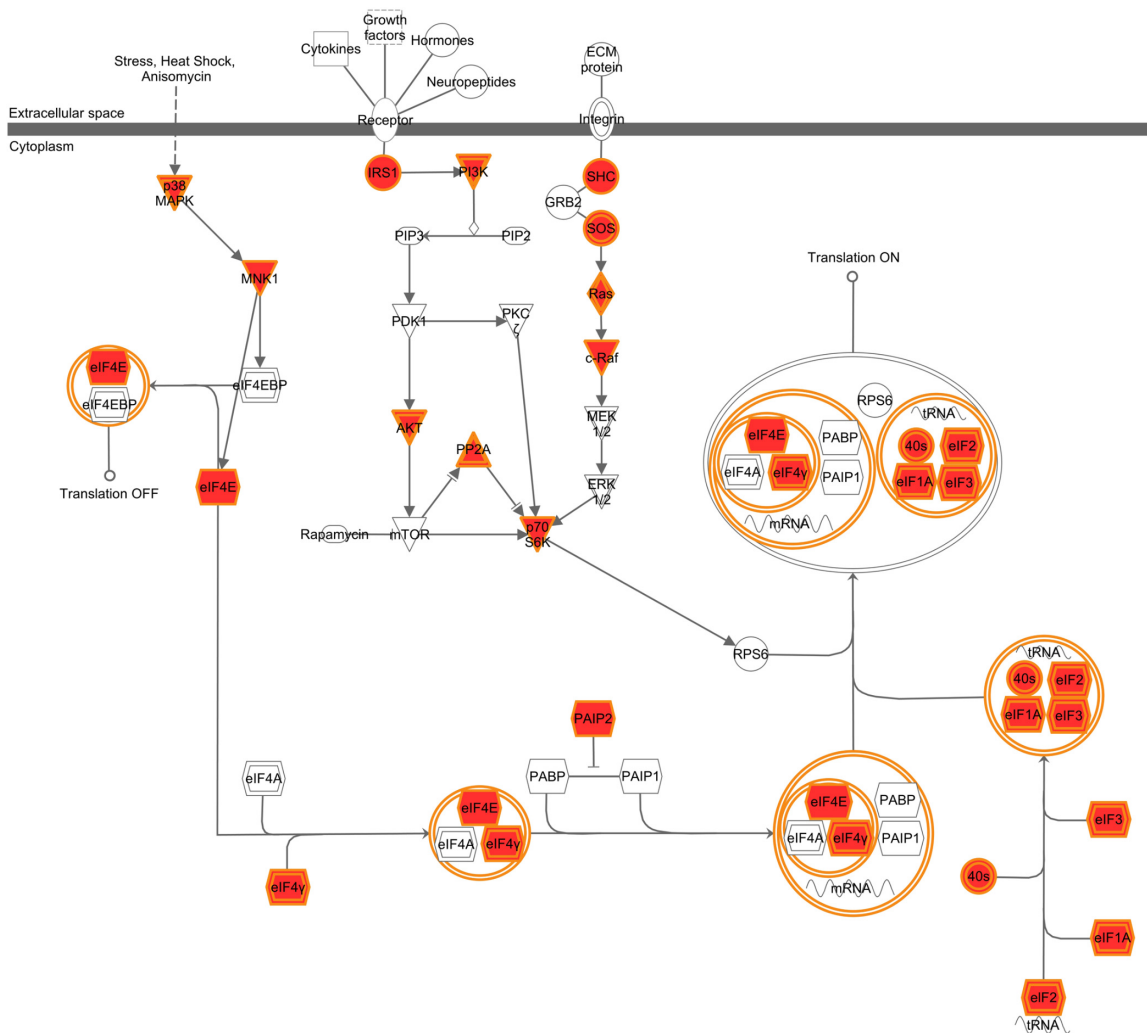
Figure 6. IPA "regulation of eIF4 and p70S6K signaling" pathway relevant to NRF target genes. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Core Analysis tool of IPA. It extracted the "Regulation of eIF4 and p70S6K signaling" pathway as the second rank significant pathway as listed in Table 3. NRF1 target genes are highlighted by red.

Conclusions

We identified 2,470 ChIP-Seq-based NRF1 target genes in SK-N-SH human neuroblastoma cells. The molecular pathways of these genes involve regulation of RNA metabolism, splicing, cell cycle, DNA damage repair, protein translation initiation, and ubiquitin-mediated protein degradation, along with mitochondrial respiratory function. We identified a panel of neurodegenerative disease-related genes, such as PARK2 (Parkin), PARK6 (Pink1), PARK7 (DJ-1), PAELR (GPR37), PSENEN (Pen2), and MAPT (tau) as previously unrecognized NRF1 targets. These results would suggest an original hypothesis of involvement of NRF1 and target genes in the pathogenesis of human neurodegenerative diseases.

Acknowledgements

Funding

Competing Interests

Author(s) disclose no potential conflicts of interest.

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 Files
Supplementary Figure 1

FastQC analysis of ChIP-Seq data. FASTQ format files are derived from short read NGS data of NRF1 ChIP-treated DNA (panel a) and IgG ChIP-treated DNA (panel b). They were imported into the FastQC program. The per base sequence quality score is shown with the median (red line), the mean (blue line), and the interquatile range (yellow box).

Supplementary Figure 2

KEGG "Alzheimer's disease" pathway. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Functional Annotation tool of DAVID. The set of 28 genes (16.5%) among 170 genes located in the KEGG "Alzheimer's disease" pathway (hsa05010) corresponded to ChIP-Seq-based NRF1 target genes highlighted by orange.

Supplementary Figure 3

KEGG "Huntington's disease" pathway. Entrez Gene IDs of 2,470 NRF1 target genes were imported into the Functional Annotation tool of DAVID. The set of 36 genes (21.2%) among 183 genes located in the KEGG "Huntington's disease" pathway (hsa05016) corresponded to ChIP-Seq-based NRF1 target genes highlighted by orange.

Supplementary Table 1

The comprehensive set of 2,470 ChIP-Seq-based NRF1 target genes.

Supplementary Table 2

Top 10 GO terms relevant to 2,470 ChIP-Seq-based NRF1 target genes genes.

References

1. Lin MT, Beal MF. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature*. 2006;443(7113):787–95.
2. McInnes J. Insights on altered mitochondrial function and dynamics in the pathogenesis of neurodegeneration. *Transl Neurodegener*. 2013;2(1):12.
3. Scarpulla RC. Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol Rev*. 2008;88(2):611–38.
4. Gugneja S, Scarpulla RC. Serine phosphorylation within a concise amino-terminal domain in nuclear respiratory factor 1 enhances DNA binding. *J Biol Chem*. 1997;272(30):18732–9.
5. Virbasius CA, Virbasius JV, Scarpulla RC. NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. *Genes Dev*. 1993;7(12A):2431–45.
6. Becker TS, Burgess SM, Amsterdam AH, Allende ML, Hopkins N. Not really finished is crucial for development of the zebrafish outer retina and encodes a transcription factor highly homologous to human Nuclear Respiratory Factor-1 and avian Initiation Binding Repressor. *Development*. 1998;125(22):4369–78.
7. Huo L, Scarpulla RC. Mitochondrial DNA instability and peri-implantation lethality associated with targeted disruption of nuclear respiratory factor 1 in mice. *Mol Cell Biol*. 2001;21(2):644–54.
8. DeSimone SM, White K. The Drosophila erect wing gene, which is important for both neuronal and muscle development, encodes a protein which is similar to the sea urchin P3A2 DNA binding protein. *Mol Cell Biol*. 1993;13(6):3641–9.
9. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*. 2009;10(10):669–80.
10. Gerstein MB, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature*. 2012;489(7414):91–100.
11. Satoh J. Bioinformatics approach to identifying molecular biomarkers and networks in multiple sclerosis. *Clin Exp Neuroimmunol*. 2010;1(3):127–40.
12. Cam H, Balciunaite E, Blais A, et al. A common set of gene regulatory networks links metabolism and growth inhibition. *Mol Cell*. 2004;16(3):399–411.
13. The ENCODE Project Consortium. A user's guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol*. 2011;9(4):e1001046.
14. Ramagopalan SV, Heger A, Berlanga AJ, et al. defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res*. 2010;20(10):1352–60.
15. Satoh J, Tabunoki H. A comprehensive profile of ChIP-Seq-based STAT1 target genes suggests the complexity of STAT1-mediated gene regulatory mechanisms. *Gene Regul Syst Bio*. 2013;7:41–56.
16. 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.
17. Kanehisa M. Molecular network analysis of diseases and drugs in KEGG. *Methods Mol Biol*. 2013;939:263–75.
18. Kumari D, Usdin K. Interaction of the transcription factors USF1, USF2, and α-Pal/Nrf-1 with the FMR1 promoter. Implications for Fragile X mental retardation syndrome. *J Biol Chem*. 2001;276(6):4357–64.
19. Jacob WF, Silverman TA, Cohen RB, Safer B. Identification and characterization of a novel transcription factor participating in the expression of eukaryotic initiation factor 2α. *J Biol Chem*. 1989;264(34):20372–84.
20. Chiorini JA, Miyamoto S, Harkin SJ, Safer B. Genomic cloning and characterization of the human eukaryotic initiation factor-2β promoter. *J Biol Chem*. 1999;274(7):4195–201.

21. Evans MJ, Scarpulla RC. Interaction of nuclear factors with multiple sites in the somatic cytochrome c promoter. Characterization of upstream NRF-1, ATF, and intron Sp1 recognition sequences. *J Biol Chem*. 1989;264(24):14361–8.

22. Mahishi L, Usdin K. NF-Y, AP2, Nrf1 and Sp1 regulate the fragile X-related gene 2 (FXR2). *Biochem J*. 2006;400(2):327–35.

23. Kherrouche Z, De Launoit Y, Monte D. The NRF-1/α-PAL transcription factor regulates human E2F6 promoter activity. *Biochem J*. 2004;383(Pt 3): 529–36.

24. Chang WT, Huang AM. α-Pal/NRF-1 regulates the promoter of the human integrin-associated protein/CD47 gene. *J Biol Chem*. 2004;279(15): 14542–50.

25. Blesa JR, Prieto-Ruiz JA, Abraham BA, Harrison BL, Hegde AA, Hernández-Yago J. NRF-1 is the major transcription factor regulating the expression of the human TOMM34 gene. *Biochem Cell Biol*. 2008;86(1):46–56.

26. Jeppesen DK, Bohr VA, Stevnsner T. DNA repair deficiency in neurodegeneration. *Prog Neurobiol*. 2011;94(2):166–200.

27. Hegde AN, Upadhya SC. Role of ubiquitin-proteasome-mediated proteolysis in nervous system disease. *Biochim Biophys Acta*. 2011;1809(2):128–40.

28. Chang WT, Chen HI, Chiou RJ, Chen CY, Huang AM. A novel function of transcription factor α-Pal/NRF-1: increasing neurite outgrowth. *Biochem Biophys Res Commun*. 2005;334(1):199–206.

29. Efiok BJ, Chiorini JA, Safer B. A key transcription factor for eukaryotic initiation factor-2α is strongly homologous to developmental transcription factors and may link metabolic genes to cellular growth and development. *J Biol Chem*. 1994;269(29):18921–30.

30. Houlden H, Singleton AB. The genetics and neuropathology of Parkinson's disease. *Acta Neuropathol*. 2012;124(3):325–38.

31. Youle RJ, Narendra DP. Mechanisms of mitophagy. *Nat Rev Mol Cell Biol*. 2011;12(1):9–14.

32. Narendra DP, Jin SM, Tanaka A, et al. PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol*. 2010;8(1):e1000298.

33. Kim RH, Smith PD, Aleyasin H, et al. Hypersensitivity of DJ-1-deficient mice to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyrindine (MPTP) and oxidative stress. *Proc Natl Acad Sci U S A*. 2005;102(14):5215–20.

34. Imai Y, Soda M, Inoue H, Hattori N, Mizuno Y, Takahashi R. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. *Cell*. 2001;105(7):891–902.

35. Manczak M, Anekonda TS, Henson E, Park BS, Quinn J, Reddy PH. Mitochondria are a direct site of Aβ accumulation in Alzheimer's disease neurons: implications for free radical generation and oxidative damage in disease progression. *Hum Mol Genet*. 2006;15(9):1437–49.

36. Ma T, Trinh MA, Wexler AJ, et al. Suppression of eIF2α kinases alleviates Alzheimer's disease-related plasticity and memory deficits. *Nat Neurosci*. 2013;16(9):1299–305.

37. Gu M, Gash MT, Mann VM, Javoy-Agid F, Cooper JM, Schapira AH. Mitochondrial defect in Huntington's disease caudate nucleus. *Ann Neurol*. 1996;39(3):385–9.

38. Dunah AW, Jeong H, Griffin A, et al. Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science*. 2002;296(5576): 2238–43.

39. Cui L, Jeong H, Borovecki F, Parkhurst CN, Tanese N, Krainc D. Transcriptional repression of PGC-1α by mutant huntingtin leads to mitochondrial dysfunction and neurodegeneration. *Cell*. 2006;127(1):59–69.