Appl Biochem Biotechnol (2013) 170:301–319
DOI 10.1007/s12010-013-0155-9

Exon Array Analysis of Alternative Splicing of Genes in SOD1G93A Transgenic Mice

Ming Hu · Yansu Guo · Huifang Chen · Weisong Duan · Chunyan Li

Received: 11 September 2012 / Accepted: 18 February 2013 /
Published online: 19 March 2013

Abstract Alternative splicing is a common strategy for creating functional diversities of proteins. While conventional identification of splice variants generally targets individual genes in amyotrophic lateral sclerosis, we present a novel exon-centric array that allows genome-wide identification of splice variants and concurrently provides analysis of gene expression. Compare 1 was asymptomatic SOD1G93A transgenic mice with nontransgenic littermates; compare 2 was symptomatic with asymptomatic transgenic mice. RT-PCR was performed to validate. Pathway and GO analysis were performed on abnormal genes. These findings could guide us to demonstrated the potential influence of mutant human CuZn-SOD1 and of splicing regulation in pathological processes.

Keywords Alternative splicing · Exon array · Gene expression · Pathway analysis · GO analysis · Amyotrophic lateral sclerosis

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative adult disease characterized by fatal paralysis in both brain and spinal cord motor neurons. ALS can

Electronic supplementary material The online version of this article (doi:10.1007/s12010-013-0155-9) contains supplementary material, which is available to authorized users.

M. Hu · Y. Guo · H. Chen · W. Duan · C. Li (✉)
Department of Neurology, The Second Hospital of Hebei Medical University, Shijiazhuang, Hebei 050000, People's Republic of China
e-mail: chunyanli5@yahoo.com.cn

Y. Guo · W. Duan · C. Li
Institute of Cardiocerebrovascular Disease, Shijiazhuang, Hebei 050000, People's Republic of China

M. Hu
Department of Neurology, Hebei General Hospital, Shijiazhuang, Hebei 050051, People's Republic of China

Present Address:
M. Hu · H. Chen
Hebei General Hospital, Shijiazhuang, Hebei 050051, People's Republic of China

be induced by inherited mutations in the gene encoding the ubiquitously expressed enzyme Cu/Zn superoxide dismutase 1 (SOD1) [1]. The role of the normal SOD1 function—conversion of toxic superoxide into less damaging hydrogen peroxide—in ALS pathogenesis remains unclear. So we choose lumbar spinal cord of an amyotrophic lateral sclerosis mouse model to do this research for the purpose of taking a general look at transcriptional level of hSOD1-G93A transgenic mice.

Alternative splicing (AS) is known to be prominent in many important physiological processes, such as cell differentiation, apoptosis, and development, and is especially prevalent in the nervous system [2]. Pre-mRNA splicing, a step in mRNA transcription, is the process by which intervening sequences (introns) are precisely removed and the flanking functional sequences (exons) are joined together. Following splicing, the exons are always joined together in the same order in which they lie in DNA. The colinearity of gene and protein is maintained between the individual exons and the corresponding parts of the protein chain. The absence or presence of exons can directly affect the distinct proteins structurally and functionally [3]. In human cells, analysis of expressed sequence tag (EST) databases and DNA microarray experiments indicate that over 60 % of all genes is alternatively spliced [4]. Though alternative splicing is prevalent in the nervous system, little is known in ALS.

A high density exon-centric microarray, Affymetrix GeneChip® Mouse Exon 1.0 ST Array, which covers over 1 million exon clusters with approximately four probes per exon and roughly 40 probes per gene was used in this study. It enables two complementary levels of analysis—gene expression and alternative splicing. One is "gene-level" expression analysis in which multiple probes on different exons are summarized into an expression value of all transcripts from the same gene. The other is "exon-level" analysis on a whole-genome scale, which opens the door to detect specific alterations in exons that may play a central role in disease mechanism and etiology. So the information is more creditable [5].

In this study, we applied exon array to investigate the alternative splicing events and gene expression profiling in hSOD1-G93A transgenic mice.

Materials and Methods

Animals

Transgenic female mice expressing mutant human CuZn-SOD1 with a $Gly^{93} \rightarrow Ala$ substitution and their nontransgenic normal female littermates were generated by breeding male hemizygous carriers (B6SJL-Tg(SOD1G93A)1Gur/J) to female B6SJLF1 hybrids, both of which were purchased from the Jackson Laboratories (Bar Harbor, ME, USA). Transgenic progeny and their normal littermates were identified by polymerase chain reaction (PCR) amplification of tail DNA [6]. We collected 9 female mice, of which 6 were derived from transgenic mice (3 30-day-old in asymptomatic stage and 3 approximately 120 days of age in symptomatic stage) and 3 were derived from littermates (30 days of age) without hSOD1-G93A, namely 30A, 120A, and 30C, to be used in this study. Animals were killed at two time points to obtain fresh lumbar spinal cord tissues (L3 to L5 segment) for analysis. To monitor disease progression, the animals were visually inspected daily. Behavioral testing was measured twice a week starting at 90 days of age (Supplementary Material 1). We selected hSOD1-G93A mouse whose performance began to decline steadily in the

rotarod test and any hindlimb extension was absent as symptomatic stage (approximately 120 days old).

The animals were housed in an asepsis room with 12 h light cycle and free access to water and diet. All experiments were carried out in accordance with the regulation of laboratory animal management promulgated by the Ministry of Science and Technology of the People's Republic of China (1988, No. 134).

Samples of Exon Array Preparation and Array Hybridization

All mice were killed under deep anesthesia with 10 % chloral hydrate and then fresh specimens of lumbar spinal cord tissues were rapidly dissected. The specimens were immediately put into RNAlater (RNAlater RNA Stabilization Reagent, Qiagen), and additional specimens were instantly frozen in liquid nitrogen and then stored at −80 °C until use. Total RNA stored in RNAlater was extracted for exon array using RNeasy Mini kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Purity and integrity of the RNA were assessed on the Agilent 2100 bioanalyzer with the RNA 6000 Nano LabChip® reagent set (Agilent Technologies, Santa Clara, CA, USA). The RNA specimens were quantified spectrophotometrically. For each experimental condition, three independent samples were used for exon array analysis

This exon array experiment was performed in CapitalBio Corp. (Beijing, China) using Affymetrix GeneChip® Mouse Exon 1.0ST arrays. Manufacturer's instructions (Affymetrix, Inc.) were followed for biotinylation and synthesis of single-stranded DNA targets, hybridization, washing, and scanning steps [5].

Data Analysis Algorithms

The Affymetrix Expression Console Software V1.0 was used to process gene- and exon-level expression signal estimates. Signal estimates were derived from the CEL files of the nine samples by quantile sketch normalization using the "Probe Logarithmic Intensity Error Estimation" (PLIER) algorithm for probeset (exon-level) intensities and Iter-PLIER (a variant algorithm) for gene-level intensities. Candidate exons for alternative splicing were detected using a Splicing Index. Presence/absence of exons was determined by "Detection Above Background (DABG)," using surrogate GC mismatch intensities [5]. Details were shown in Supplementary Material 2.

Filtering

At the gene level, after variance stabilization and log transformation, the fold change and P value were calculated using control samples as base values. We set fold change >2 or <0.5 and P value<0.05 as cutoff values for significantly differentially expressed genes.

At the exon level, the standard by which we filtered the exons as alternative splicing events is shown below: (a) SI (Splicing Index) >2 means this exon expression in sample 1 is higher than in sample 2, indicating the "exon inclusion" event; SI<0.5 means this exon expression in sample 2 is higher than in sample 1, indicating the "exon skipping" event. (b) P value<0.05. (c) Focus on exons that NI (Normalized Intensity) is near 1.0 in the group predicted to have a higher inclusion rate and near 0 in the group predicted to have a lower inclusion rate. (d) Limit search to genes that have very large differential expression. (e) Limit search to high-confidence exons. (f) Filter probes with unusually low variance.

The formulas are shown below:
Exon Level=Transcription×Splicing

$$Gene-level\ Normalized\ Intensity(NI)=\frac{\text{Probe set intensity}}{\text{Expression level of the ``gene''}}$$

$$\text{Splicing index value(SI)}=\log_2\frac{\text{NI (Sample1)}}{\text{NI (Sample 2)}}$$

GO and Pathway Analysis of the Filtered Genes

The functions and pathways of these significant change genes were analyzed using CapitalBio® Molecule Annotation System (CB-MAS) V4.0 available through CapitalBio Corp. (http://bioinfo.capitalbio.com/mas), which included the gene ontology (GO) hierarchy analysis to group genes functionally and pathway analysis to group genes into respective pathway. Three GO terms which are biological process, cell component, and molecular function were evaluated. Pathways analysis also includes three sheets: KEGG, BioCarta, and GenMAPP. In this research, we chose molecular function and KEGG to demonstrate.

Validation by Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

RT-PCR was used to verify the levels of phosphoinositide-3-kinase gamma polypeptide (PIK3CG, NM_020272) and NADPH oxidase 2 (NOX2, NM_007807) mRNA and to check whether Fyn-binding protein (FYB, AK030905) has alternative splicing. The RNA was extracted from the lumbar spinal cord tissues stored at −80 °C using Trizol reagent (Invitrogen, Carlsbad, CA), following the manufacturer's instructions. RNA concentration and purity were determined spectrophotometrically at 260 and 280 nm Reverse transcription used reagents all from Promega following the manufacturer's instructions.

Target sequences were obtained from Affymetrix. Forward and reverse primers were 5′-g c t t a g a g g a c g a t g a c g t t-3′ and 5′-g g a t a g g a c t g t g g g a t c a g-3′ for PIK3CG (524 bp product), 5′-t g a a c g a a t t g t a c g t g g a c-3′ and 5′-a c t t g a g a a t g g a g g c a a a g-3′ for NOX2 (593 bp product), and 5′-a a t g g g t g c c t t g t t g c c-3′ and 5′-g c t t c t g a t c c t c t t g c c t t c-3′ for the exon 19 of FYB (539 and 226 bp product when aberrant splicing). Forward and reverse primers of exon 19 were designed adjacent to constitutive exons, and these primers amplified specific bands of differentially expressed transcripts. The densitometric values were normalized with respect to the values of glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Results

Differentially Expressed Genes

In 30A vs.30C, 202 differentially expressed genes (DEGs; $P<0.05$) passed our signal threshold and were identified for subsequent analysis among 14039 transcripts on the chip. There were four genes changed more than 1.5-fold, including Fc receptor-like S scavenger receptor (Fcrls, NM_030707). Only 1 gene without gene annotation was upregulated more than 2-fold (definitely 2.8-fold). In 120A vs. 30A, 2869 DEGs ($P<0.05$) were identified in 14314 transcripts, consisting of 263 (9.17 %) filtered genes upregulated more than 2-fold

and 71 (2.47 %) filtered genes downregulated more than 2-fold. Four genes regulated more than 1.5-fold in 30A vs. 30C were detected in 120A vs. 30A, but they had no statistical significance ($P>0.05$). It can be seen that at least 330 significantly changed genes were emerging with the progression of disease process. Take comparison of 120A and 30A groups as an example, the significant DEGs was visualized using Cluster 3.0 software with TreeView tools after hierarchical, average linkage clustering (Fig. 1a). The DEGs in 120A vs. 30A were described using volcano plot (Fig. 1b). One hundred DEGs including top 90 upregulated genes and bottom 10 downregulated genes in 120A vs. 30A were shown (Tables 1 and 2). In addition, 16.2 % (54/334) of the transcripts which were changed significantly contained alternatively spliced exons in 120A vs. 30A, indicating the possibility of combinatorial regulation of transcription and splicing.

Alternative Splicing Events

Using the "Splicing Index" algorithm described in the methods, 630 probe select regions (PSRs) were identified as alternatively spliced exons that belonged to 563 alternatively spliced transcripts (ASTs) in 120A vs. 30C, and only 85 PSRs were identified as alternatively spliced exons that belonged to 85 ASTs in 30A vs. 30C. In 120A vs. 30A group, 537 PSRs with higher expression were classified as "exon inclusion" events, while the other 93 PSRs with lower expression were considered as "exon skipping" events. In 30A vs. 30C group, compared with normal mice, 61 PSRs were classified as "exon inclusion" events and 24 PSRs were considered as "exon skipping" events. In 120A vs. 30A, we filtered 46 genes whose exon's NI near 1.0 in one group predicted to have a higher inclusion rate and near 0 in

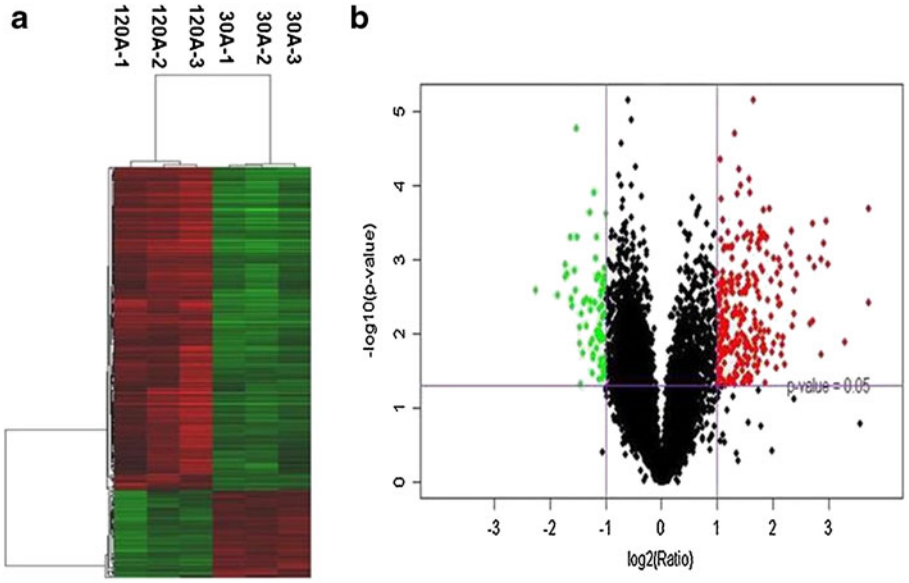

Fig. 1 **a** Clustering display of exon array data of 120A vs. 30A groups. The differentially expressed genes, consisting of 263 genes upregulated more than 2-fold and 71 genes downregulated more than 0.5-fold in 120A vs. 30A groups, were determined using Cluster 3.0 software and visualized with Tree View tools after hierarchical clustering. **b** Volcano Plot reflected the comparison of the 120A and 30A groups. Each *dot* represents the log2 values for the transcript cluster from the Exon Array (*x* axis). *Green* and *red dot signs* indicate that the transcript cluster was significantly downregulated or upregulated

Table 1 Upregulated genes in lumbar spinal cords of 120A vs. 30A mice (top 90)

transcript_cluster_id	Genebank no.	Gene name	Fold change	*P* value
6777309	NM_017372	Lyz2 // lysozyme 2	13.142	0.004
6767782	NM_013532	Lilrb4 // leukocyte immunoglobulin-like receptor	13.084	0.000
7015521	NM_007807	Cybb//cytochrome b-245, beta polypeptide	9.835	0.013
6946055	NM_053110	Gpnmb//glycoprotein (transmembrane) nmb	7.939	0.001
7010355	NM_001044384	Timp1 // tissue inhibitor of metalloproteinase 1	7.730	0.000
6868098	NM_028595	Ms4a6c // membrane-spanning 4-domains	7.529	0.001
6790294	NM_011337	Ccl3 // chemokine (C-C motif) ligand 3	7.277	0.001
6906967	NM_009264	Sprr1a // small proline-rich protein 1A	7.252	0.019
6855051	NM_009780	C4b // complement component 4B (Childo blood group)	6.701	0.001
6990715	NM_153098	Cd109 // CD109 antigen	6.537	0.007
6797579	NM_009252	Serpina3n // serine (or cysteine) peptidase inhibitor	6.527	0.000
6824838	NM_019984	Tgm1 // transglutaminase 1	6.373	0.007
6789325	NM_009853	Cd68 // CD68 antigen	6.259	0.001
6959584	NM_011662	Tyrobp // TYRO protein tyrosine kinase binding protein	5.385	0.002
6964382	NM_021334	Itgax // integrin alpha X	5.228	0.005
6925984	NM_013706	Cd52 // CD52 antigen	5.181	0.003
6899374	NM_011313	S100a6 // S100 calcium binding protein A6 (calcyclin)	5.104	0.008
6976237	NM_008278	Hpgd // hydroxyprostaglandin dehydrogenase 15 (NAD)	5.055	0.001
6926167	NM_007572	C1qa//complement component 1 q subcomponent alpha polypeptide	5.045	0.000
6764049	NM_010185	Fcer1g // Fc receptor, IgE, high affinity I, gamma polypeptide	4.969	0.001
6875132	NM_011701	Vim // vimentin	4.729	0.001
6888324	NM_009776	Serping1//serine (or cysteine) peptidase inhibitor, clade G, member 1	4.630	0.026
6755210	NM_013489	Cd84 // CD84 antigen	4.556	0.011
6932603	NM_013470	Anxa3 // annexin A3	4.475	0.014
6833736	NM_011019	Osmr // oncostatin M receptor	4.432	0.019
6957059	NM_009779	C3ar1 // complement component 3a receptor 1	4.407	0.004
6926166	NM_007574	C1qc // complement component 1, q subcomponent, C chain	4.401	0.001
6856201	NM_020568	S3-12 // plasma membrane associated protein,S3-12	4.396	0.002
6818696	NM_010705	Lgals3 //lectin, galactose binding, soluble 3	4.321	0.004
6910173	NM_010516	Cyr61 // cysteine rich protein 61	4.302	0.001
6926165	NM_009777	C1qb // complement component 1, q subcomponent, beta polypeptide	4.241	0.003
6792657	NM_011150	Lgals3bp // lectin, galactoside-binding, soluble, 3 binding protein	4.235	0.011
6768985	NM_008404	Itgb2 // integrin beta 2	4.214	0.002
6885873	NM_008491	Lcn2 // lipocalin 2	4.180	0.029

Table 1 (continued)

transcript_cluster_id	Genebank no.	Gene name	Fold change	*P* value
4512397	none	none	4.058	0.001
6769597	NM_010512	Igf1 // insulin-like growth factor 1	4.032	0.014
6878713	NM_021398	Slc43a3 // solute carrier family 43, member 3	3.969	0.017
6976901	NM_008509	Lpl // lipoprotein lipase	3.922	0.006
6991714	NM_011254	Rbp1 // retinol binding protein 1	3.899	0.002
6750440	NM_008342	Igfbp2 // insulin-like growth factor binding protein 2	3.847	0.000
6857183	NM_011723	Xdh // xanthine dehydrogenase	3.789	0.009
6796121	NM_023275	Rhoj //ras homolog gene family, member J	3.748	0.003
6933625	NM_011854	Oasl2 // 2'-5' oligoadenylate synthetase-like 2	3.715	0.000
6917393	NM_010686	Laptm5 // lysosomal-associated protein transmembrane 5	3.715	0.001
6789420	NM_023158	Cxcl16 // chemokine (C-X-C motif) ligand 16	3.702	0.025
6893630	NM_022325	Ctsz // cathepsin Z	3.687	0.001
6957410	NM_020008	Clec7a //C-type lectin domain family 7, member a	3.635	0.046
4930880	none	none	3.631	0.016
6985924	NM_007806	Cyba //cytochrome b-245, alpha polypeptide	3.603	0.009
6806245	NM_010745	Ly86 // lymphocyte antigen 86	3.571	0.000
6850135	NM_010380	H2-D1//histocompatibility 2,D region locus 1	3.570	0.014
6762603	NM_001111316	Ptprc // protein tyrosine phosphatase, receptor type, C	3.560	0.017
5217948	none	none	3.552	0.000
6851324	NM_010130	Emr1//EGF-like module containing, mucin-like, hormone receptor-like sequence 1	3.538	0.001
6946993	NM_007599	Capg //capping protein (actin filament), gelsolin-like	3.504	0.012
6856290	NM_009778	C3 // complement component 3	3.469	0.015
6949613	NM_175628	A2m // alpha-2-macroglobulin	3.461	0.007
5422319	none	none	3.456	0.001
6957044	NM_031159	Apobec1 //apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1	3.431	0.001
4939205	none	none	3.425	0.000
6850308	NM_001005485	Olfr111 // olfactory receptor 111	3.423	0.009
6765218	NM_007498	Atf3 // activating transcription factor 3	3.398	0.000
6850834	NM_031254	Trem2 // triggering receptor expressed on myeloid cells 2	3.390	0.001
6957144	NM_008479	Lag3 // lymphocyte-activation gene 3	3.388	0.003
6962376	NM_009982	Ctsc // cathepsin C	3.373	0.001
6750546	NM_013612	Slc11a1//solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1	3.372	0.023
6756541	NM_001111059	Cd34 // CD34 antigen	3.321	0.003
6886045	NM_010283	Ggta1 // glycoprotein galactosyltransferase alpha 1, 3	3.318	0.027
6991264	NM_007801	Ctsh // cathepsin H	3.318	0.001
6917365	NM_028266	Col16a1 // collagen, type XVI, alpha 1	3.314	0.011

Table 1 (continued)

transcript_cluster_id	Genebank no.	Gene name	Fold change	*P* value
6934119	NM_001042489	Hvcn1 // hydrogen voltage-gated channel 1	3.307	0.033
4539257	none	none	3.285	0.002
5543277	none	none	3.253	0.017
6917120	NM_007782	Csf3r // colony stimulating factor 3 receptor (granulocyte)	3.245	0.008
6809655	NM_008533	Cd180 // CD180 antigen	3.221	0.016
6868058	NM_023044	Slc15a3 // solute carrier family 15, member 3	3.215	0.002
6817381	NM_008873	Plau // plasminogen activator, urokinase	3.212	0.007
6833513	NM_153505	Nckap1l // NCK associated protein 1 like	3.208	0.002
6789743	NM_008878	Serpinf2// serine (or cysteine) peptidase inhibitor, clade F, member 2	3.202	0.013
7020802	NM_133211	Tlr7 // toll-like receptor 7	3.194	0.006
6907971	NM_007651	Cd53 // CD53 antigen	3.187	0.021
6899683	NM_021281	Ctss // cathepsin S	3.184	0.005
6872781	NM_009890	Ch25h // cholesterol 25-hydroxylase	3.181	0.012
6986722	NM_008605	Mmp12 // matrix metallopeptidase 12	3.160	0.027
4696497	none	none	3.150	0.002
6784345	NM_008175	Grn // granulin	3.148	0.000
6905289	NM_008536	Tm4sf1 // transmembrane 4 superfamily member 1	3.083	0.000
6977260	NM_010442	Hmox1 // heme oxygenase (decycling) 1	3.075	0.016
6824841	NM_026819	Dhrs1 // dehydrogenase/ reductase (SDR family) member 1	3.051	0.001
6913042	NM_027450	Glipr2 // GLI pathogenesis-related 2	3.050	0.009

Gene expression levels are expressed as means of fold change, which is calculated by dividing the signals of each ALS animal lumbar spinal cord tissues sample by those of control samples, in the 3 symptomatic transgenic mice and 3 30-day-old asymptomatic transgenic mice

the other group predicted to have a lower inclusion rate to show (Table 3). The same went for 30A vs. 30C, 13 genes were filtered, including lengsin (Lgsn, NM_153601), RNA-binding motif protein 15B (Rbm15b, BC052180), tryptophan hydroxylase 2 (Tph2, NM_173391), colony stimulating factor 2 receptor beta 2 (Csf2rb2, NM_007781), prolactin (Prl, NM_011164), polycystic kidney disease 1-like 3 (Pkd1l3, NM_001039700), and other 7 unnamed genes. Furthermore, 12.06 % of alternatively spliced genes was changed at least 2-fold in 120A vs. 30A. In addition, only 7 of 85 ASTs in 30A vs. 30C were detected in 120A vs. 30A; that is, 556 ASTs were emerging with the progression of disease process.

GO and Pathway Analysis of Transcription and Splicing

GO and pathway analysis were performed on 4 DEGs (fold change$\geq$1.5, $P<0.05$) and 13 ASTs in 30A vs. 30C using MAS V4.0 with P value cutoff of less than 10^{-3}. Only Fcrls in four DEGs was shown to be related to receptor activity and no pathway result was found. Six in 13 ASTs showed GO analysis results, including Rbm15b, Pkd1l3, Csf2rb2, Prl, Tph2, and gluld1, which were related to nucleotide binding, ion channel activity, receptor activity, hormone activity, monooxygenase activity, and catalytic activity, respectively. Furthermore,

Table 2 Downregulated genes in lumbar spinal cords of 120A vs. 30A mice (bottom 10)

transcript_cluster_id	Genebank no.	Gene name	Fold change	*P* value
6983639	NM_020259	Hhip // Hedgehog-interacting protein	0.342	0.003
6775355	NM_010255	Gamt // guanidinoacetate methyltransferase	0.340	0.001
6987403	NM_010700	Ldlr // low density lipoprotein receptor	0.330	0.004
6954557	NM_009762	Smyd1 // SET and MYND domain containing 1	0.325	0.003
6985925	NM_138656	Mvd // mevalonate (diphospho) decarboxylase	0.321	0.000
6935927	NM_020010	Cyp51 // cytochrome P450, family 51	0.309	0.002
6804893	NM_010917	Nid1 // nidogen 1	0.303	0.002
7005797	NM_145942	Hmgcs1 // 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1	0.301	0.001
6972405	NM_009876	Cdkn1c // cyclin-dependent kinase inhibitor 1C (P57)	0.274	0.003
6842587	NM_139134	Chodl // chondrolectin	0.209	0.003

Gene expression levels are expressed as means of fold change, which is calculated by dividing the signals of each ALS animal lumbar spinal cord tissues sample by those of control samples, in the 3 symptomatic transgenic mice and 3 30-day-old asymptomatic transgenic mice

Csf2rb2 and Prl belonged to cytokine–cytokine receptor interaction and Jak-STAT signaling pathway based on KEGG library. Likewise, 100 DEGs and 46 ASTs in 120A vs. 30A were analyzed, which were shown in details (DEGs in Supplementary Table 1, 2 and ASTs in Table 4).

Taking into account the molecular function of the 100 DEGs in 120A vs. 30A, there is a GO mapping to describe (Fig. 2a). With regard to the GO analysis, it is of note that the DEGs are significantly enriched in "protein binding, oxidoreductase activity and hydrolase activity" functional annotation. The 100 DEGs in 120A vs. 30A participate in 34 pathways. We chose KEGG data to show the 24 top enrichment pathways of these differentially expressed genes in 120A vs. 30A (Fig. 2b). Take "Leukocyte transendothelial migration" pathway for example, it includes NOX2 and PIK3CG (Supplementary Fig. 1). We found that this pathway contained not only 9 significantly upregulated genes, but also 2 exons inclusion events. Their change and change tendency have been validated by RT-PCR or supported by literature to be congruous with the exon array system [7–11]. Genes affected at both transcriptional and splicing levels appeared in this pathway. Otherwise, NOX2 appears in 5 pathways and PIK3CG appears in 28 pathways based on the three databases. Simultaneously, we can obtain the homologous gene's correlation graph which was generated to illustrate the correlativity of two or more genes within networks (Supplementary Fig. 2). The number in this graph represents the number of pathways. This graph contains 14 genes and PIK3CG is the core, indicating that PIK3CG may function synergistically in this module in response to development of ALS. Based on this exon array system, we can focus on these genes and pathways to explore the pathogenesis of ALS.

RT-PCR Validation

Two DEGs exhibiting significant differences in expression were detected verified by RT-PCR. Exon array showed that NOX2 gene was upregulated 9.83-fold (P=0.0130) and PIK3CG gene was upregulated 2.4-fold (P=0.0136) in 120A vs. 30A, and no change in

Table 3 Alternatively spliced transcripts in lumbar spinal cords of 120A vs. 30A mice

probeset_id	SI	*P* value	NI (120A)	NI (30A)	Genebank no.	Gene name
5194403	4.174	0.039	1.125	0.270	AK036058	Fam184b // family with sequence similarity 184, member B
4915139	20.661	0.037	1.096	0.053	NM_009890	Ch25h // cholesterol 25-hydroxylase
4532147	4.914	0.012	1.069	0.218	NM_031881	Nedd4l // neural precursor cell expressed, developmentally down-regulated gene 4-like
4326991	3.615	0.018	1.029	0.285	NM_201645	Ugt1a1// UDP glucuronosyltransferase 1 family, polypeptide A1
5131806	5.925	0.007	0.984	0.166	BC048546	cDNA sequence BC048546
5376282	5.247	0.042	0.970	0.185	AK030905	Fyb // FYN binding protein
5269315	4.361	0.004	0.968	0.222	NM_021398	Slc43a3 // solute carrier family 43, member 3
4696937	3.797	0.025	0.926	0.244	NM_001037937	Depdc6 // DEP domain containing 6
5400282	3.853	0.021	0.918	0.238	AF067063	cDNA sequence AF067063
4900355	5.167	0.021	0.907	0.176	NM_008329	Ifi204 // interferon activated gene 204
4854957	3.310	0.050	0.904	0.273	NM_033601	Bcl3 // B-cell leukemia/lymphoma 3
5129426	3.150	0.006	0.895	0.284	NM_010501	Ifit3 // interferon-induced protein with tetratricopeptide repeats 3
5580427	3.285	0.012	0.868	0.264	NM_001033335	Serpina3f // serine (or cysteine) peptidase inhibitor, clade A, member 3 F
4499870	3.353	0.038	0.864	0.258	NM_007942	Epo // erythropoietin
4672780	3.045	0.002	0.843	0.277	NM_011324	Scnn1a // sodium channel, nonvoltage-gated 1 alpha
5419168	4.278	0.001	0.807	0.189	NM_008670	Naip1 // NLR family, apoptosis inhibitory protein 1
4902375	3.552	0.001	0.806	0.227	NM_010870	Naip5 // NLR family, apoptosis inhibitory protein 5
4966972	3.020	0.003	0.800	0.265	NM_172574	Pqlc3 // PQ loop repeat containing
5243284	2.681	0.001	0.796	0.297	AB124611	cDNA sequence AB124611
5288280	3.826	0.036	0.796	0.208	NM_008331	Ifit1 // interferon-induced protein with tetratricopeptide repeats 1
4312244	4.529	0.014	0.788	0.174	NM_023143	C1r//complement component 1, r subcomponent
5458700	3.408	0.001	0.786	0.231	NM_008670	Naip1 // NLR family, apoptosis inhibitory protein 1
4350893	3.219	0.032	0.785	0.244	NM_001077707	Shprh // SNF2 histone linker PHD RING helicase
5211881	2.893	0.050	0.769	0.266	NM_001003961	Dnmt3b // DNA methyltransferase 3B
5030634	3.615	0.009	0.762	0.211	NM_001081957	predicted gene, OTTMUSG00000000971

Table 3 (continued)

probeset_id	SI	*P* value	NI (120A)	NI (30A)	Genebank no.	Gene name
4797244	2.809	0.015	0.746	0.266	NM_148950	Pknox2 // Pbx/knotted 1 homeobox 2
5237269	4.141	0.004	0.745	0.180	NM_175369	Ccdc122 // coiled-coil domain containing 122
4643974	2.491	0.044	0.735	0.295	NM_001136084	Tph1 // tryptophan hydroxylase 1
5130777	3.468	0.018	0.735	0.212	NM_028231	Kcnmb2 // potassium large conductance calcium-activated channel, subfamily M, beta member 2
4682361	2.450	0.025	0.733	0.299	NM_175164	Arhgap26 // Rho GTPase activating protein 26
4439412	3.254	0.007	0.730	0.224	NM_145144	Aif1l // allograft inflammatory factor 1-like
4550653	3.341	0.029	0.727	0.218	NM_010678	Aff3 // AF4/FMR2 family, member 3
4346201	3.453	0.004	0.720	0.209	NM_181542	Slfn10 // schlafen 10
4311349	4.961	0.034	0.714	0.144	NM_001033450	Mnda// myeloid cell nuclear differentiation antigen
5608885	3.141	0.032	0.710	0.226	NM_010043	Des // desmin
4434708	2.411	0.032	0.710	0.295	NM_016701	Nes // nestin
5423079	4.305	0.030	0.710	0.165	NM_146198	Slc5a11 // solute carrier family 5 (sodium/glucose cotransporter), member 11
5471544	3.630	0.021	0.709	0.195	NM_178891	Prmt6 // protein arginine N-methyltransferase 6
5494226	3.977	0.041	0.706	0.178	NM_172603	Phf11 // PHD finger protein 11
4886642	2.415	0.014	0.703	0.291	NM_175198	Prox2 // prospero homeobox 2
5004135	0.400	0.007	0.296	0.740	NM_031380	Fstl3 // follistatin-like 3
5189849	0.334	0.015	0.284	0.853	NM_022883	Lpin3 // lipin 3
5151153	0.387	0.014	0.276	0.713	NM_134163	Mbnl3 // muscleblind-like 3 (Drosophila)
4692267	0.315	0.009	0.255	0.811	U19265	Gcnt1 // glucosaminyl (N-acetyl) transferase 1, core 2
4826122	0.279	0.000	0.222	0.796	NM_007403	Adam8 // a disintegrin and metallopeptidase domain 8
5462176	0.217	0.006	0.152	0.702	NM_133754	Fblim1 // filamin binding LIM protein 1

Forty-six alternatively spliced transcripts was filtered which exons' NI near 1 in SYMPTOM vs 30A group predicted to have a higher inclusion rate and near 0 in the group predicted to have a lower inclusion rate

Table 4 GO and pathway analysis of 46 alternatively spliced transcripts in lumbar spinal cords of 120A vs. 30A mice

Genebank no.	Protein	Event	Molecular function	KEGG
AK036058	Fam184b	Inclusion	–	–
NM_009890	Ch25h	Inclusion	Metal ion binding, iron ion binding, cholesterol 25-hydroxylase activity	–
NM_031881	Nedd4l	Inclusion	Ligase activity, protein binding, acid–amino acid ligase activity	Ubiquitin mediated proteolysis
NM_201645	Ugt1a1	Inclusion	–	–
BC048546	cDNA sequence BC048546	Inclusion	–	–
AK030905	Fyb	Inclusion	–	–
NM_021398	Slc43a3	Inclusion	–	–
NM_001037937	Depdc6	Inclusion	–	–
AF067063	cDNA sequence AF067063	Inclusion	–	–
NM_008329	Ifi204	Inclusion	–	–
NM_033601	Bcl3	Inclusion	Transcription factor activity, transcription factor binding	–
NM_010501	Ifit3	Inclusion	–	–
NM_001033335	Serpina3f	Inclusion	Serine-type endopeptidase inhibitor activity	–
NM_007942	Epo	Inclusion	Hormone activity, protein binding, erythropoietin receptor binding	Jak-STAT signaling pathway, hematopoietic cell lineage, cytokine–cytokine receptor interaction
NM_011324	Scnn1a	Inclusion	WW domain binding, protein binding, ion channel activity	Taste transduction
NM_008670	Naip1	Inclusion	Nucleotide binding, nucleoside-triphosphatase activity	–
NM_010870	Naip5	Inclusion	–	–
NM_172574	Pqlc3	Inclusion	–	–
AB124611	cDNA sequence AB124611	Inclusion	–	–
NM_008331	Ifit1	Inclusion	Binding	–
NM_023143	C1r	Inclusion	–	–
NM_008670	Naip1	Inclusion	Nucleotide binding, nucleoside-triphosphatase activity	–

Table 4 (continued)

Genebank no.	Protein	Event	Molecular function	KEGG
NM_001077707	Shprh	Inclusion	ATP binding, hydrolase activity, helicase activity, metal ion binding	–
NM_001003961	Dnmt3b	Inclusion	Metal ion binding, methyltransferase activity, protein binding, zinc ion binding	Methionine metabolism
NM_001081957	predicted gene	Inclusion	Protease inhibitor activity	–
NM_148950	Pknox2	Inclusion	Sequence-specific DNA binding, actin monomer binding, transcription factor activity, actin filament binding	–
NM_175369	Ccdc122	Inclusion	–	–
NM_001136084	Tph1	Inclusion	Monooxygenase activity, metal ion binding, tryptophan 5-monooxygenase activity	Tryptophan metabolism
NM_028231	Kcnmb2	Inclusion	Ion channel activity, calcium-activated potassium channel activity	–
NM_175164	Arhgap26	Inclusion	Protein binding, GTPase activator activity	–
NM_145144	Aif1l	Inclusion	Calcium ion binding	–
NM_010678	Aff3	Inclusion	–	–
NM_181542	Slfn10	Inclusion	ATP binding	–
NM_001033450	Mnda	Inclusion	–	–
NM_010043	Des	Inclusion	Protein binding, structural constituent of cytoskeleton	Cell communication
NM_016701	Nes	Inclusion	Structural molecule activity	Cell communication
NM_146198	Slc5a11	Inclusion	Transporter activity, sodium ion binding, sugar porter activity, symporter activity	–
NM_178891	Prmt6	Inclusion	Methyltransferase activity, transferase activity	–
NM_172603	Phf11	Inclusion	Protein binding, zinc ion binding, beta-lactamase activity	–
NM_175198	Prox2	Inclusion	Transcription regulator activity, DNA binding	–
NM_031380	Fstl3	Skipping	Protein binding, activin binding, activin inhibitor activity	–
NM_022883	Lpin3	Skipping	Phosphatidate phosphatase activity	–
NM_134163	Mbnl3	Skipping	Metal ion binding, nucleic acid binding, zinc ion binding	–

Table 4 (continued)

Genebank no.	Protein	Event	Molecular function	KEGG
U19265	Gcnt1	Skipping	–	–
NM_007403	Adam8	Skipping	Metal ion binding, metalloendopeptidase activity, zinc ion binding, peptidase activity	–
NM_133754	Fblim1	Skipping	Metal ion binding, protein binding, zinc ion binding	–

GO analysis of alternatively spliced transcripts was focused on molecular function, and pathway analysis was performed based on KEGG library

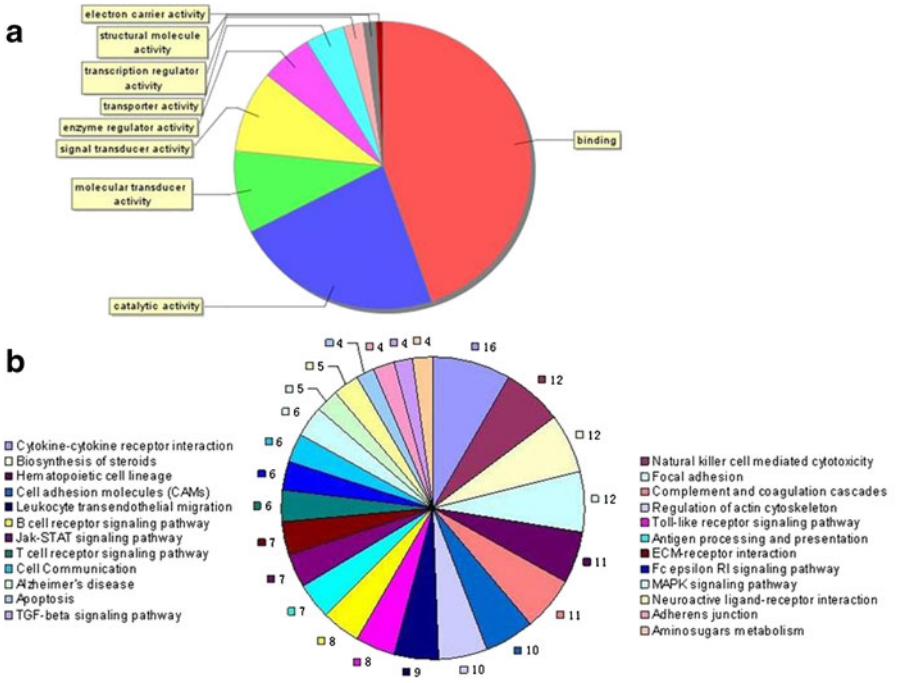

Fig. 2 **a** Go mapping classification of the molecular function of the 100 differentially expressed genes in 120A vs. 30A with MAS software (http://bioinfo.capitalbio.com/mas). **b** The top 24 statistically significant ($P<0.001$) KEGG pathway are shown and the *numbers* indicate the numbers of the genes involved in this pathway

30A vs. 30C. The expression change and changed tendency of NOX2 and PIK3CG genes validated by RT-PCR were consistent with the results acquired by the exon array (Fig. 3). One alternative splicing exon was selected for validation by RT-PCR. From the exon array, the exon 19 of FYB gene (SI=5.25, $P=0.042$) was highly included in 120A (NI=0.9704) and lower expressed in 30A (NI=0.185), which is consistent with the results of the RT-PCR

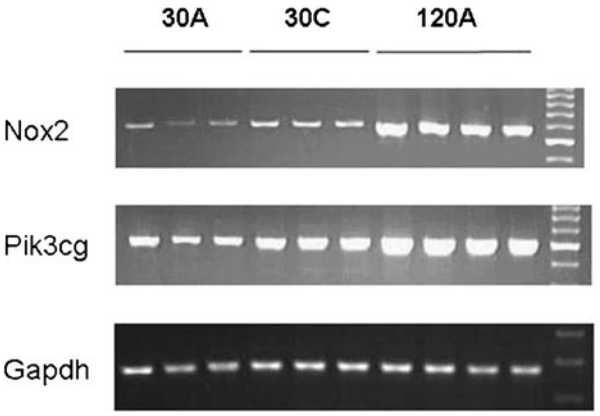

Fig. 3 RT-PCR validation of differentially expressed genes in 30A ($n=3$), 30C ($n=3$), and 120A ($n=4$) samples from lumbar spinal cord. Representative bands of NOX2 and PIK3CG mRNA expression in lumbar spinal cord. GAPDH was used as an internal control

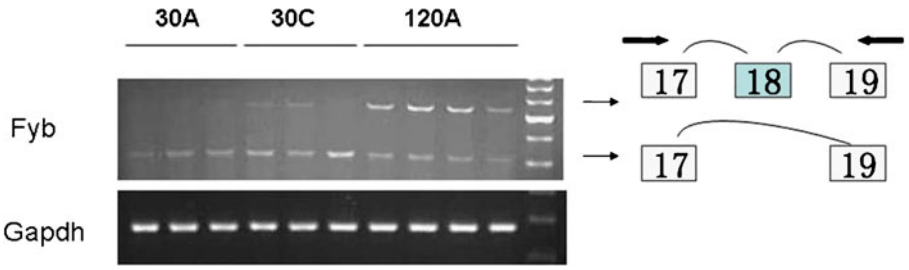

Fig. 4 RT-PCR validation of alternatively spliced genes in 30A (n=3), 30C (n=3), and 120A (n=4) samples from lumbar spinal cord. RT-PCR was performed with primers hybridizing to neighboring constitutively expressed exons. Gene symbols are listed to the left gel, and sequence-validated exon structures are shown to the right with primer postions indicated by *arrows*. This is a "exon inclusion" event

(Fig. 4). All these results suggest that the exon array system maybe reliable and effective enough to detect differential expression at both transcriptional and splicing levels.

Discussion

Alternative splicing is a common strategy for creating functional diversities of proteins that have cell and developmentally specific functions. Given the important role for splicing, it is not surprising that a recent estimate has proposed that 50 % to 60 % of mutations linked to disease affect splicing [12]. Some previous researches support the critical morbigenous contribution [13] of exon splicing to ALS in particular, such as TDP-43, FUS/TLS, and EAAT2 [14–16]. But they only focused on single gene. Exon arrays provide the most comprehensive coverage of the genome, including empirically supported and predicted transcribed sequences, enabling the discovery of previously unidentified novel events. Most bioinformatics studies [17] rely on identifying ESTs that come from the same gene and looking for differences between them that are consistent with alternative splicing. But given that only a few ESTs have been sequenced for most genes, it seems possible that even more alternative splicing exists that is not yet detectable in the available ESTs. Moreover, there are many researches on gene expression profiling of ALS not only in hSOD1-G93A mouse but also in patients [18]; however, application of GeneChip® Mouse Exon 1.0 ST array on lumber spinal cord of ALS mice at different stages that allows genome-wide identification of differential splice variation, and concurrently provides a flexible and inclusive analysis of gene expression was few. While conventional identification of splice variants generally targets individual genes, we present this exon array to take a more extensive and deeper look in ALS caused by hSOD1-G93A mutation.

Complex Regulation of Transcription and Splicing in hSOD1-G93A Mouse

In 120A vs. 30A, 12.06 % of alternatively spliced genes was upregulated or downregulated at least 2-fold, and 25.24 % of alternatively spliced genes was upregulated or downregulated by at least 1.5-fold. Alternative splicing can expand the protein repertoire and influence protein function by altering protein domains. The affected domains in the coding regions of alternatively spliced exons confirmed the existence of changes in the transcriptome and proteome resulting from alterations in the domain architecture of biological networks [19]. It was hypothesized that splicing may modify transcription activity [20] and can change RNA stability which in turn may affect gene expression [21], while transcription may change the

splicing efficiency [22]. It is also possible that different expression levels of the upstream genes such as splicing factors facilitate or inhibit splicing machinery by influencing spliceosome assembly or the *cis*-elements during the splicing process. These two aspects of regulations may result in high degree of correlation between splicing patterns and transcriptional expression [23]. Evidence from the splicing classifications and the overlap between the two levels suggests a combinatorial regulation. Especially, the research of alternatively spliced exons was analyzed based on the coding regions, which would be more significant.

Functional and Pathway Analysis in hSOD1-G93A Mouse

According to the exon array system, the gene function and pathway analysis of DEGs and ASTs revealed the direct or indirect effect of mutant SOD1 in ALS. With regard to the GO analysis, it is of note that the DEGs are significantly enriched in "protein binding, oxidoreductase activity and hydrolase activity" functional annotation. Since genes are usually functionally organized into pathways, it is necessary to explore the gene regulation in terms of the pathways involved. As shown in Supplementary Fig. 1, the "Leukocyte transendothelial migration" pathway includes NOX2/gp91phox (also known as cytochrome b-245, beta polypeptide that is Cybb), PIK3CG and FYB. The change of the PI3K/Rac1/NOX2/ROS pathway contained by the "Leukocyte transendothelial migration" pathway was detected at gene level by exon array system and RT-PCR. FYB gene undergoes alternative splicing in ALS model, which may inference that it maybe affect the function of some signaling molecules implicated in T cell receptor (TCR) signaling, including PI3K, Fyn, and so on [24, 25]. The combination of previous researches and exon array suggested that PIK3CG, NOX2, and FYB play important roles in ALS. Based on this exon array system, we can focus on these genes and pathways to explore the pathogenesis of ALS.

Cascade Amplification Reaction in hSOD1-G93A Mouse

At the gene level, there were 202 DEGs ($P<0.05$) in 30A vs. 30C, consisting of only one gene upregulated more than 2-fold. In 120A vs. 30A, we detected 2869 DEGs ($P<0.05$), consisting of 263 (9.17 %) genes with more than 2-fold upregulation and 71 (2.47 %) genes with downregulation. Four genes which were 1.5-fold changed in 30A vs. 30C were all detected in 120A vs. 30A, but they have no statistical significance because of $P>0.05$. It can be seen from that at least 330 significantly changed genes were emerging with the progression of the disease process. Likewise, 556 ASTs were emerging. At asymptomatic stage of ALS, there is no significant change between transgenic mice and non-transgenic mice at transcriptional level, but the transgenic mice at symptomatic stage differ significantly from asymptomatic stage. This suggests that the mutant SOD1 may be involved in a number of key steps in the pathological cascade of events leading to neuronal injury and this kind of ALS may be a cumulative result of mutant SOD1. It may revoke the cascade amplification reaction. The changes of NOX2 and PIK3CG are good examples. They exist in "Leukocyte transendothelial migration" pathway, which were upregulated accompanied by time lapse. From Supplementary Fig. 1, we can see that many cytokines are induced, as measured by the exon array. These DEGs and ASTs provide a new direction for further study of possible pathogenic mechanism and treatment of ALS at transcriptional level. We cannot neglect aging-associated changes, but it is not the key point in morbility process of transgenic mice. The lifespan of littermate mice without mutant SOD1 reaches at least 600 days; however, the transgenic mice would die at approximately 150 days old. Moreover, exon array did not detect the change of several transcription factors

(e.g., Ink4, LASS2, some zinc-binding proteins, IP3R, and some collagen isoforms) that might be implicated in aging.

Conclusion

In this study, hSOD1-G93A mice were used as ALS model to explore the occurrence and development processes of ALS, followed by a genome-wide expression profiling of transcription and splicing by exon array system. Functional and pathway analysis of gene level and exon level demonstrated the importance of transcriptional and splicing regulation in physiological and pathological processes. There is increasing evidence for alterations in RNA splicing as being associated with not only ALS, but also related forms of motor neuron diseases. Because very few studies have investigated splicing regulation in hSOD1-G93A mice, elucidating the underlying transcriptional mechanisms associated with these phenomena is critical for a better understanding of the pathogenesis of ALS.

Acknowledgments This study was supported in part by grants from Natural Science Foundation of China (30870882, 30900460, and 81171210) and a grant from Science and Technological department of Hebei Province (11966122D).

References

1. Deng, H. X., Hentati, A., Tainer, J. A., Iqbal, Z., Cayabyab, A., Hung, W. Y., et al. (1993). *Science, 261*, 1047–1051.
2. Schwerk, C., & Schulze-Osthoff, K. (2005). *Molecular Cell, 19*, 1–13.
3. Matlin, A. J., Clark, F., & Smith, C. W. (2005). *Nature Reviews Molecular Cell Biology, 6*, 386–398.
4. Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour, C. D., et al. (2003). *Science, 302*, 2141–2144.
5. Affymetrix Inc.: Affymetrix White Papers: GeneChip® Exon Array Design Technical Note; GeneChip® Whole Transcript (WT) Sense Target Labeling Assay Manual v 4.0; Exon Probeset Annotations and Transcript Cluster Groupings v1.0; Alternative Transcript Analysis Methods for Exon Arrays v1.1; Identifying and Validating Alternative Splicing Events; Gene Signal Estimates from Exon Arrays v1.0; Exon Array Background Correction v1.0; Guide to Probe Logarithmic Intensity Error (PLIER) Estimation; Statistical Algorithms Reference Guide [http://www.affymetrix.com/support/technical/whitepapers.affx].
6. Gurney, M. E., Pu, H., Mauro, C., Chiu, A. Y., Dal Canto, M. C., Polchow, C. Y., et al. (1994). *Science, 264*, 1772–1775.
7. Mehranpour, P., Wang, S. S., Blanco, R. R., Li, W., Song, Q., et al. (2009). *Cardiovascular and Hematological Agents in Medicinal Chemistry, 7*, 251–259.
8. Kielland, A., Blom, T., Nandakumar, K. S., Holmdahl, R., Blomhoff, R., & Carlsen, H. (2009). *Free Radical Biology and Medicine, 47*, 760–766.
9. Boillée, S., Yamanaka, K., Lobsiger, C. S., Copeland, N. G., & Jenkins, N. A. (2006). *Science, 312*, 1389–1392.
10. Kieran, D., Sebastia, J., Greenway, M. J., King, M. A., Connaughton, D., et al. (2008). *The Journal of Neuroscience, 28*, 14056–14061.
11. Lin, L. C., Wang, Y. H., Hou, Y. C., Chang, S., Liou, K. T., Chou, Y. C., et al. (2006). *Journal of Pharmacy and Pharmacology, 58*, 129–135.
12. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigó, R. (2005). *FEBS Letters, 579*, 1900–1903.
13. Blencowe, B. J. (2006). *Cell, 126*, 37–47.
14. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., et al. (2006). *Science, 314*, 130–133.
15. Kwiatkowski, T. J., Jr., Bosco, D. A., Leclerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., et al. (2009). *Science, 323*, 1205–1208.

16. Münch, C., Ebstein, M., Seefried, U., Zhu, B., Stamm, S., Landwehrmeyer, G. B., et al. (2002). *Journal of Neurochemistry, 82*, 594–603.
17. Kan, Z., Rouchka, E. C., Gish, W. R., & States, D. J. (2001). *Genome Research, 11*, 889–900.
18. Ferraiuolo, L., De Bono, J. P., Heath, P. R., Holden, H., Kasher, P., Channon, K. M., et al. (2009). *Journal of Neurochemistry, 109*, 1714–24.
19. Resch, A., Xing, Y., Modrek, B., Gorlick, M., Riley, R., & Lee, C. (2004). *Journal of Proteome Research, 3*, 76–83.
20. Nobuhiro, F., Yoshitaka, N., Popiel, H. A., Hiroki, K., Masamitsu, Y., & Tatsushi, T. (2005). *FEBS Letters, 579*, 3842–3848.
21. Hollams, E. M., Giles, K. M., Thomson, A. M., & Leedman, P. J. (2002). *Neurochemical Research, 27*, 957–980.
22. Rosonina, E., Bakowski, M. A., McCracken, S., & Blencowe, B. J. (2003). *The Journal of Biological Chemistry, 278*, 43034–43040.
23. Hang, X., Li, P., Li, Z., Qu, W., Yu, Y., Li, H., et al. (2009). *BMC Genomics, 10*, 126.
24. Musci, M. A., Hendricks-Taylor, L. R., Motto, D. G., Paskind, M., Kamens, J., Turck, C. W., et al. (1997). *The Journal of Biological Chemistry, 272*, 11674–11677.
25. Wang, H., Wei, B., Bismuth, G., & Rudd, C. E. (2009). *Proc Natl Acad Sci, 106*, 12436–12441.