Supplementary Table 1. Genotype and allele frequencies of *MAFA* +1377 G/T (Gly346Cys) in type 1 diabetes (all cases).

	Control n=263 (%)		All cases n=139 (%)		OR [95%CI]	p value
Genotype						
G / G	236	(89.7)	134	(96.4)		
G / T	27	(10.3)	5	(3.6)	0.33 [0.13-0.83]	0.0196
Allele						
G (Gly)	499	(94.9)	273	(98.2)		
T (Cys)	27	(5.1)	5	(1.8)	0.34 [0.13-0.85]	0.0222

Fisher's exact probability test

Supplementary Table 2. Genotype and allele frequencies of *MAFA* +1377 G/T (Gly346Cys) in high-risk type 1 diabetes

	Control n=263 (%)		High-risk cases n=124(%)		OR [95%CI]	p value
Genotype						
G / G	236	(89.7)	121	(98.0)		
G / T	27	(10.3)	3	(2.0)	0.22[0.07-0.69]	0.0083
Allele						
G (Gly)	499	(94.9)	245	(99.0)		
T (Cys)	27	(5.1)	3	(1.0)	0.23 [0.08-0.71]	0.0095

*High-risk cases: type 1 diabetes patients with high-risk *IDDM2* allele (-23Hph+/+)

Fisher's exact probability test

Supplementary Table 3. Haplotype association of *MAFA* with susceptibility to type 1 diabetes (all cases)

	Polymorphism		Frequency			
Haplotype	+69 VNTR	+1377 G/T	All cases	Control	OR [95%CI]	p value
MAFA-ht1	1R	G (Gly)	0.727	0.700	1.12[0.72-1.74]	NS
MAFA-ht2	non 1R	G (Gly)	0.254	0.249	1.02[0.65-1.61]	NS
MAFA-ht3	1R	T (Cys)	0.018	0.051	0.33 [0.13-0.86]	<0.03

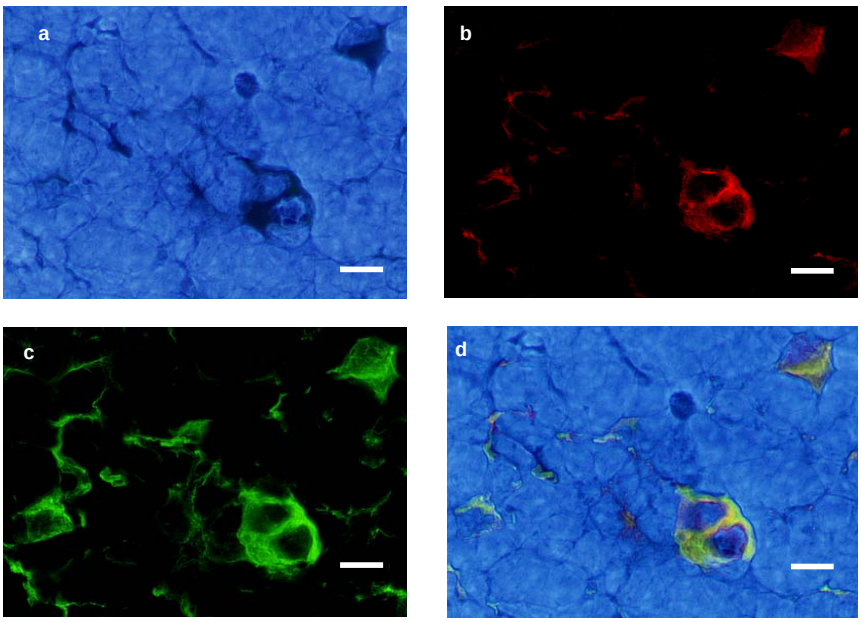
Supplementary Table 4. Genotype and allele frequencies of *MAFA* +1377 G/T (Gly346Cys) in type 2 diabetes

	Control n=425 (%)		Cases n=347 (%)		OR [95%CI]	p value
Genotype						
G / G	392	(92.2)	329	(94.8)		
G / T	33	(7.8)	18	(5.2)	1.54 [0.85-2.77]	NS
Allele						
G (Gly)	817	(96.1)	676	(97.4)		
T (Cys)	33	(3.9)	18	(2.6)	1.52 [0.85-2.71]	NS

Fisher's exact probability test

Supplementary Figure 1.

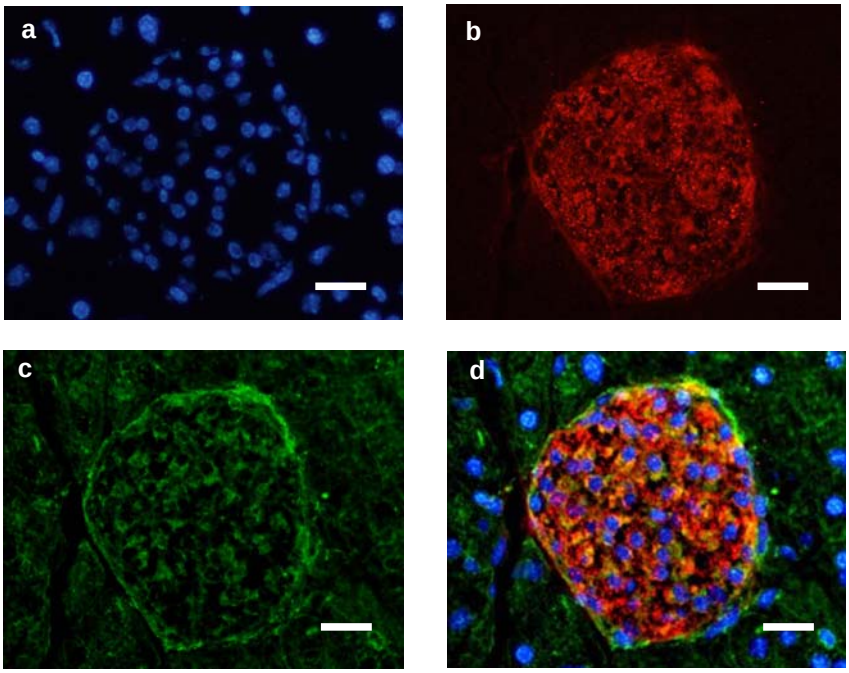
A



Co-localization of insulin and cytokeratin in thymic medulla (C3H female, 7 days old) (a) Nuclear staining by DAPI, (b) anti-insulin antibody, (c) anti-cytokeratin antibody, and (d) merged image.

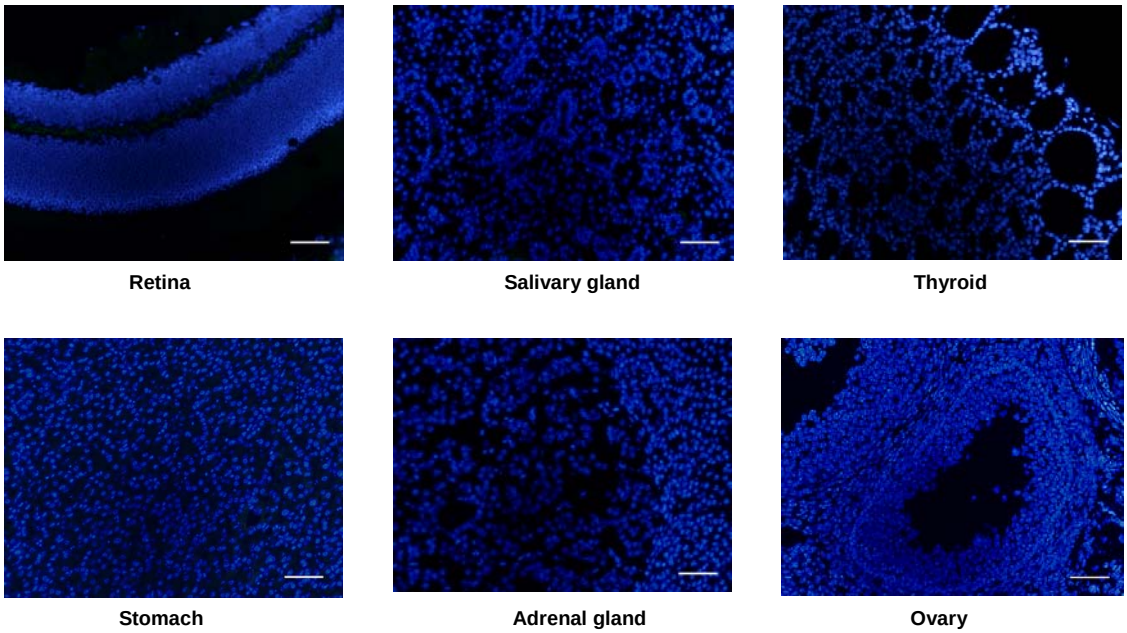
Bar scale: 10 μm

B



(a) Nuclear staining by DAPI, (b) anti-insulin antibody, (c) serum from Mafa-/- mice, and (d) merged image.

C

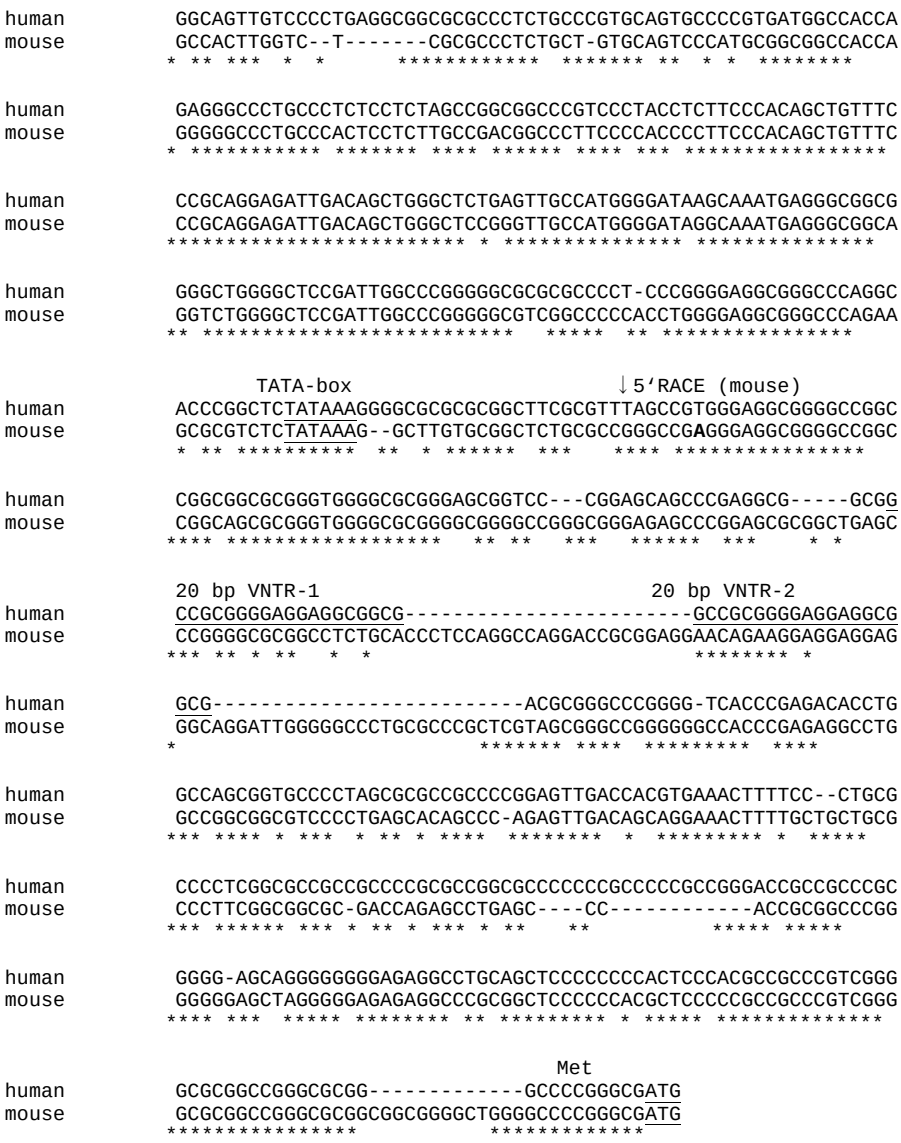


No autoantibody was detected in retina, salivary glands, thyroid, stomach, adrenal gland, and ovary by staining with Mafa -/- serum.

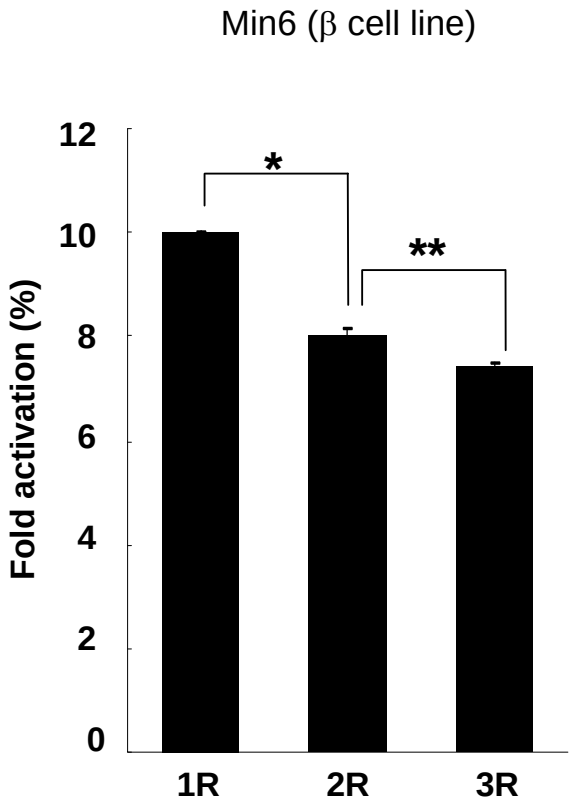
Blue: DAPI for nuclear staining
Green: mouse IgG for autoantibody

Bar scale: 50 μm

Supplementary Figure 2.

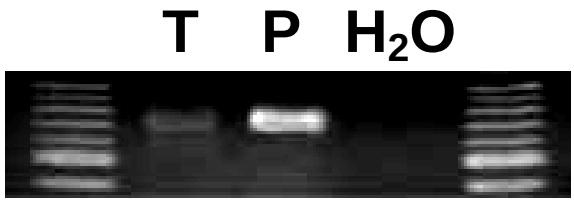


Supplementary Figure 3.



Mean ±SEM, * p<0.0001,**p<0.005 ANOVA

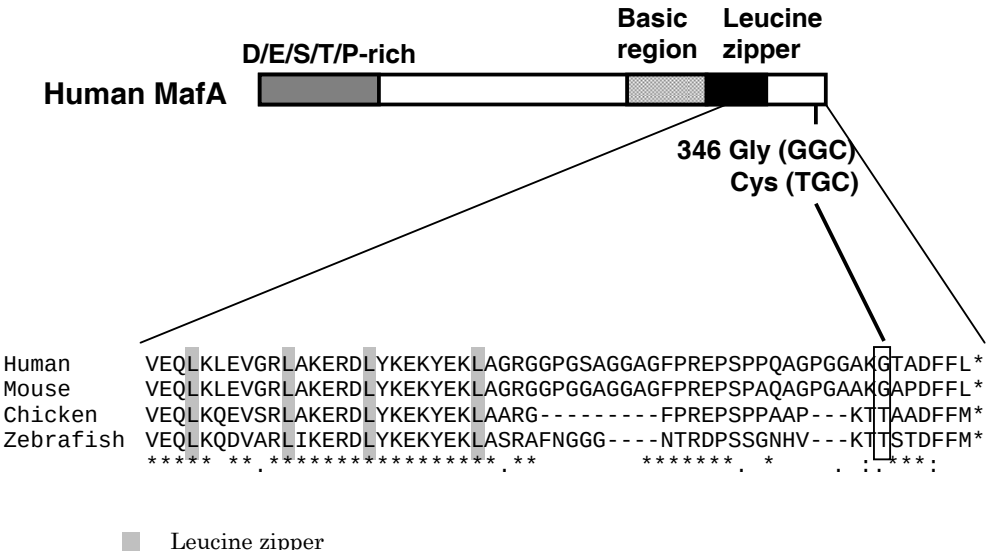
Supplementary Figure 4.



T: human thymic cDNA library

P: human pancreatic cDNA library

Supplementary Figure 5.



Supplementary materials and methods

cDNA was amplified by PCR with the following primers:

Aire	(Forward)	5’-GCGATGCAGTGTGTGTGGCGATGG-3’
	(Reverse)	5’-TCTGGATGGCCCACTGCAGGATGC-3’
Hprt	(Forward)	5’-CTCGAAGTGTTGGATACAGC-3’
	(Reverse)	5’-TGGCCTATAGGCTCATAGTG-3’
Pdx1	(Forward)	5’-CTCGCTGGGATCACTGGAGCA-3’
	(Reverse)	5’-GCTTTGGTGGATTTCATCCACGG-3’
Neurod1	(Forward)	5’-CTTGGCCAAAAACTACATTTGG-3’
	(Reverse)	5’-GGAGTAGGGGTGTACCGGGAA-3’
Ins1	(Forward)	5’-CCATCAGCAAGCAGGTCATTGTTT-3’
	(Reverse)	5’-GCGGGACTTGGGTGTGTAGAAGAA-3’
Ins2	(Forward)	5’-CCATCAGCAAGCAGGTTATTGTTTC-3’
	(Reverse)	5’-CAGCTCCAG TTGTGCCACTTGTG-3’

Immunohistochemical staining

Universal negative control for rabbit primary antigen (DAKO Japan Co., Kyoto, Japan) was used for insulin staining as recommended by DAKO, and rabbit polyclonal IgG (Abcam Co., Tokyo, Japan) was used for MafA staining as isotype control.

Genomic sequence of mouse Mafa

The transcription start site of mouse *Mafa* was determined by rapid amplification of cDNA ends (5'RACE) analysis of polyA RNA isolated from Min6 cells (mouse insulinoma cell line). Several clones, starting from 398bp upstream of the first ATG and 27bp downstream of the TATA box, were obtained. The 3' untranslated region (UTR, 1343bp) of mouse *Mafa* was estimated by searching the expressed sequence tag (EST) containing both the 3' flanking sequence of the *Mafa* open reading frame (ORF) and the poly A site from a database. Genomic DNA was extracted from liver tissue by a standard phenol-chloroform method. A total of 17 pairs of primers were designed using Primer3 Input (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi), so that the promoter region (approximately 2400bp upstream from the transcription initiation site), an exon, and the 3′-flanking region of the *Mafa* gene were covered by 17 segments.

Newly identified polymorphisms of Mafa in NOD mouse

Ten SNPs (-1788T/C, -1722C/T, -1716T/C, -1507G/T, -1201A/G, -1113C/T, -993G/A, -851T/C, -818C/A, and -493C/T relative to the transcription start site), a 1bp deletion (-1319del) and two insertions (-1582insCCA and -1236insTTGA) in the promoter region, one SNP (2437A/G) and one 2bp deletion (2727del) in the 3' untranslated region, and one SNP (3640A/C) in the 3' flanking region were newly identified in the NOD mouse.

Subjects for case-control studies

The diagnosis of type 1 diabetes was defined based on both clinical features and laboratory data. All the patients were ketosis-prone and lacked endogenous insulin secretion as judged by a C-peptide level of <3.3 nmol/day, and needed four or more insulin injections per day. Individuals with type 2 diabetes were diagnosed according to World Health Organization criteria. Control subjects had normal glucose tolerance and no family history of diabetes or other autoimmune diseases. All individuals with type 1 diabetes or type 2 diabetes and control subjects were of Japanese origin and resided in the Osaka area (western Japan). Glutamic acid decarboxylase antibody (GAD Ab) was measured by a commercially available RIA kit using ^{125}I-labeled recombinant human GAD65 as a tracer reagent (Cosmic, Tokyo, Japan). Samples were defined as GAD Ab positive above a threshold of 10 units/ml as suggested in previous reports (Maruyama et al. *J Clin Endocrinol Metab* 93:2115-21, 2008).

Screening for human MAFA polymorphisms

Genomic DNA was extracted from peripheral blood leukocytes using proteinase K as

described previously, and polymorphisms in the complete *MAFA* gene were detected by sequencing. The transcription start site of *MAFA* was estimated by alignment of sequences between human and mouse (341bp upstream to first ATG and 29bp downstream to TATA box, Supplementary Fig. 2). The 3' UTR (1280bp) of human *MAFA* was estimated by searching EST containing both the 3' flanking sequence of *MAFA* ORF and the poly A site from a database. Primers purchased from Vector Asian Pacific Corporation were designed to determine the sequence of the complete *MAFA* gene. Polymerase chain reaction (PCR) amplification was performed in a total reaction volume of 20 μl containing the following reagents: 4 ng genomic DNA, 5 pmol of each primer, 0.5 unit LA Taq polymerase (Takara, Shiga, Japan), 2 x PCR buffer containing 5mM $MgCl_2$, and 8mM dNTP mixture. Amplification was carried out in a Takara thermal cycler. PCR products were visualized by 2% agarose gel electrophoresis with ethidium bromide staining, and purified using a gel extraction kit (QIAGEN) for mice and PCR96 Cleanup plates (Millipore) for humans, before the sequence reaction. Sequencing was performed using an ABI PRISM Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems).

Newly identified polymorphisms were located at the following positions (numbers are relative to transcription start site, Fig. 4A); -2078ins/del (2bp), -2065T/C, -1764C/A, -1761T/C, -1719C/G, -1670G/A, -1486T/C, -1123ins/del (1bp), -955C/T, -784ins/del (103bp), -282G/T, -263G/A, -205ins/del (1bp), -109C/G, -18ins/del (2bp), +41A/G, +69VNTR (GCCGCGGGGAGGAGGCGGCG, 1 repeat (1R) to 3 repeats (3R)), +132A/G, +153T/A, +1377G/T(Gly346Cys), +1430G/T, +1630T/A, +1716C/T, and +2279T/A.

Genotyping of human MAFA polymorphisms

Gene Scan analysis was performed for genotyping of +69VNTR. The forward primer was labeled with FAM for the polymerase chain reaction, and analyzed by an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Restriction fragment length polymorphism analysis using an ApaLI (New England Biolab, MA, USA) was performed for genotyping the Gly346Cys polymorphism, and the result was confirmed by using a Taqman System (Applied Biosystems) in some of the subjects. All genotyping data were double-scored to minimize error. The overall genotype call rate was 98.1% (+69VNTR) and 96.9% (Gly346Cys). The genotype distribution of each polymorphism was compatible with Hardy-Weinberg equilibrium ($p>0.05$, chi-squared test).

Luciferase assay

To construct the reporter plasmid, m-Mafa-p-luc, the promoter region of mouse Mafa (-2500~first ATG) was amplified from Balb/c or NOD genomic DNA by PCR. The fragments were inserted into KpnI-HindII sites of the pGL2-basic plasmid (Promega). Min6 cells were grown in a 24-well plate and transfected with a total of 3.2 μg of plasmids (2.8 μg luciferase plasmid, 0.4 μg pER-Rluc), using 8 μl Lipofectamine2000. The construction of 3xMARE/RBGP-luc is described in a previous paper {Kataoka, 1996 #84}. The expression plasmid for human MAFA with both alleles of the Gly346Cys variant was constructed by inserting a *NotI* fragment excised from pGEM-T-easy/h-*MAFA* into the pHygEF2 mammalian expression vector. To construct hemagglutinin (HA)-tagged human *MAFA*, a double-stranded oligonucleotide encoding the HA epitope tag sequence was inserted into the 5'-*EcoR*I site of pGEM-T-easy/h-*MAFA*. The resultant HA-h-*MAFA* fragment (XbaI-SpeI) was inserted into pHygEF2 to obtain pHygEF2/HA-h-*MAFA*. PT67 cells (derived from NIH3T3 cells) grown in a 24-well plate were transfected with a total of 1.5 μg plasmids (0.2 μg luciferase plasmid, 0.4 μg pHygEF2 vector, 0.8 μg expression plasmids and 0.1 μg pEF-*R*luc) using 7.5 μl Superfect (Qiagen). Cells were harvested 24 h after transfection. To construct the reporter plasmid, CMV-h-MAFA-p-luc, the human MAFA promoter region containing the 1R, 2R or 3R allele (-130~+321) was amplified from human genomic DNA by PCR, and the CMV immediate early enhancer was conjugated with upstream of the fragments. These fragments were inserted into the pGL2-basic plasmid (Promega). TEC1C6 (thymic medullary epithelial cell line) cells grown in a 24-well plate were transfected with a total of 1.6 μg plasmids (0.4 μg luciferase plasmid, 0.8 μg expression plasmid, and 0.4 μg pEF-Rluc) using 1.6 μl Lipofectamine2000 (Invitrogen). Twenty four hours after transfection in each experiment, firefly and Renilla luciferase activities were measured using a Dual Luciferase Assay System (Promega).