DOI: 10.1016/j.freeradbiomed.2015.12.005

Unique cistrome defined as CsMBE is strictly required for Nrf2-sMaf heterodimer function in cytoprotection

Akihito Otsuki[1], Mikiko Suzuki[2,*], Fumiki Katsuoka[3], Kouhei Tsuchida[1], Hiromi Suda[1], Masanobu Morita[1], Ritsuko Shimizu[4], and Masayuki Yamamoto[1,3,*]

[1]Department of Medical Biochemistry, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan
[2]Center for Radioisotope Sciences, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan
[3]Tohoku Medical Megabank Organization, Tohoku University, Sendai 980-8573, Japan
[4]Department of Molecular Hematology, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan

***Correspondence:**
Masayuki Yamamoto
Department of Medical Biochemistry, Tohoku University Graduate School of Medicine
2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan
Phone +81-22-717-8084; Fax +81-22-717-8090
E-mail: masiyamamoto@med.tohoku.ac.jp
Mikiko Suzuki
Center for Radioisotope Sciences, Tohoku University Graduate School of Medicine
2-1 Seiryo-machi, Aoba-ku, Sendai, Miyagi, 980-8575, Japan
Phone +81-22-717-8088; Fax +81-22-717-8090
E-mail: suzukimikiko@med.tohoku.ac.jp

Abstract

Nrf2-small Maf (sMaf) heterodimer is essential for the inducible expression of cytoprotective genes upon exposure to oxidative and xenobiotic stresses. While the Nrf2-sMaf heterodimer recognizes DNA sequences referred to as the antioxidant/electrophile responsive element (ARE/EpRE), we here define these DNA sequences collectively as CNC-sMaf binding element (CsMBE). In contrast, large and small Maf proteins are able to form homodimers that recognize the Maf recognition element (MARE). CsMBE and MARE share a conserved core sequence but they differ in the 5'-adjacent nucleotide neighboring the core. Because of the high similarity between the CsMBE and MARE sequences, it has been unclear how many target binding sites and target genes are shared by the Nrf2-sMaf heterodimers and Maf homodimers. To address this issue, we introduced a substitution mutation of alanine to tyrosine at position 502 in Nrf2, which rendered the DNA-binding domain structure of Nrf2 similar to Maf, and generated knock-in mice expressing the $Nrf2^{A502Y}$ mutant. Our chromatin immunoprecipitation-sequencing analyses showed that binding sites of $Nrf2^{A502Y}$-sMaf were dramatically changed from CsMBE to MARE *in vivo*. Intriguingly, however, one-quarter of the $Nrf2^{A502Y}$-sMaf binding sites also bound Nrf2-sMaf commonly and *vice versa*. RNA-sequencing analyses revealed that $Nrf2^{A502Y}$-sMaf failed to induce expression of major cytoprotective genes upon stress stimulation, which increased the sensitivity of $Nrf2^{A502Y}$ mutant mice to acute acetaminophen toxicity. These results demonstrate that the unique cistrome defined as CsMBE is strictly required for the Nrf2-sMaf heterodimer function in cytoprotection and that the roles played by CsMBE differ sharply from those of MARE.

Highlights

- Substitution of Ala-502 to Tyr renders the DNA-binding of Nrf2 similar to that of Maf
- Sequence recognition of $Nrf2^{A502Y}$ shifts from CsMBE to MARE
- $Nrf2^{A502Y}$ fails to induce major cytoprotective genes upon stress stimulation
- Recognition of CsMBE by Nrf2-sMaf is required for the cytoprotective function
- $Nrf2^{A502Y}$ mutant mice are susceptible to oxidative and xenobiotic stresses

Keywords

Nrf2, Maf, CsMBE, MARE

Introduction

Nrf2 (NF-E2-related factor 2) is a CNC (cap 'n' collar) family transcription factor that regulates inducible expression of an array of cytoprotective genes [1-3]. Nrf2 activates target genes in a stress-dependent manner through forming a heterodimer with small Maf proteins (sMaf). Under normal conditions, Nrf2 protein is constitutively trapped by Keap1 (Kelch-like ECH-associated protein 1) and is degraded through the proteasome pathway in the cytoplasm [4, 5]. Oxidative and electrophilic stresses inactivate Keap1 and stabilize Nrf2 [3]. The stabilized Nrf2 is translocated into nucleus and activates expression of target genes that encode enzymes/proteins scavenging of reactive oxygen species (ROS) or related to detoxification of xenobiotics and drug metabolism.

The CNC family transcription factors, including NF-E2 p45, Nrf1, Nrf2, Nrf3, Bach1 and Bach2, form heterodimers with the sMaf family of transcription factors, MafF, MafG and MafK [6-9]. The CNC-sMaf heterodimers bind to a consensus DNA sequences, which are called various names, such as antioxidants response element (ARE) [10, 11], electrophile response element (EpRE) [12], and NF-E2 binding element [13], via their basic region-leucine zipper (bZip) structure. We have compared these binding sequences and found that they show a common consensus sequence, 5'-(A/G)TGA(G/C)nnnGC-3', but these recognition elements are partially distinct from the element bound by Maf homodimers. Therefore, in this study we refer to the sequence recognized by CNC-sMaf, including the ARE, the EpRE and the NF-E2 binding element, as CNC-sMaf binding element (CsMBE). Of note, the CsMBE sequence shares substantial overlap with that of the Maf recognition element (MARE), a palindromic motif 5'-TGCTGA(G/C)TCAGCA-3' (underline shows overlapping sequence with CsMBE) that binds homodimers of large Maf proteins (c-Maf, MafA/L-Maf, MafB and Nrl) and sMaf proteins [14-16]. Because of the significant overlap, there has been substantial confusion in the cistrome dynamics or the binding sequence

selection by the CNC-sMaf heterodimers and Maf homodimers.

CsMBE and MARE harbor TRE (phorbol 12-O-tetradecanoate-13-acetate (TPA)-responsive element; TGA(G/C)TCA) or binding site for AP-1 in the middle of the motifs [17]. MARE harbors GC at the 5' of TRE, while CsMBE retains A/G at the position. It is interesting to note that the presence of a GC dinucleotide adjacent to the TRE stabilizes MafG homodimer binding [11, 15, 16]. Indeed, the surface plasmon resonance (SPR)-based protein-DNA interaction studies revealed that the GC sequence is essential for recognition by sMaf proteins [18]. Structural analysis of MafG revealed that Arg-57, Asp-61 and Tyr-64 of the basic region of MafG are important for the recognition of the GC sequence [19, 20]. These three residues are highly conserved in the large Maf and sMaf family proteins and their ancestors [21-24].

While Arg-57 and Asp-61 of MafG are conserved in the basic region of the CNC family proteins, the residue of Nrf2 corresponding to Tyr-64 of MafG is converted to alanine residue (Ala-502). The alanine residue is highly conserved among the CNC family transcription factors (Nrf1, Nrf2, Nrf3, NF-E2 p45, Bach1 and Bach2) and ancestors of CNC family proteins; SKN-1 (Skinhead family member-1) in *Caenorhabditis elegans* [25], CncC in *Drosophila melanogaster* [26] and Nfe2l2a in *Danio rerio* [27, 28]. We and others previously found that the alanine and tyrosine in the basic region are critical residues to determine the unique binding preference of Nrf2-sMaf heterodimer and Maf homodimer to CsMBE and MARE, respectively [29, 30]. A heterodimer of an Nrf2 mutant generated by replacing the Ala-502 residue with a tyrosine residue ($Nrf2^{A502Y}$) and MafG displays binding preferences similar to MafG homodimer [29]. Since both $Nrf2^{A502Y}$ and MafG require the GC sequence in the TRE flanking region, the $Nrf2^{A502Y}$-sMaf heterodimer displays similar high-affinity binding to the palindromic MARE to that of MafG homodimer.

To assess the contribution of CsMBE recognition by Nrf2-sMaf to cytoprotective

function, we generated Nrf2^{A502Y} mutant knock-in mice using a genome-editing technique. Utilizing peritoneal macrophages from the Nrf2^{A502Y} mutant knock-in mice, we performed comprehensive analyses of Nrf2 binding sites [chromatin immunoprecipitation (ChIP)-sequencing (ChIP-Seq)] and gene expression profiles [RNA-sequencing (RNA-Seq)]. To our surprise, we found that the Nrf2^{A502Y} mutant fails to support the expression of three-quarters of the electrophile-inducible cytoprotective genes, including glutathione conjugation- and hydrogen peroxide degradation-related enzyme genes, inducible expression of which are normally supported by wild-type Nrf2. Meanwhile, the Nrf2^{A502Y} mutant still retains the ability to support the expression of one-quarter of the electrophile-inducible genes under the Nrf2 regulation. These results thus unequivocally demonstrate that CsMBE and MARE, binding sequences for CNC-sMaf heterodimer and Maf homodimer, respectively, generate distinct sets of gene regulations. Specific recognition of the CsMBE by the Nrf2-sMaf heterodimer is critical for the inducible expression of Nrf2 target genes, which play key roles in the cytoprotection against ROS and toxic electrophiles.

Materials and methods

Generation of *Nrf2*^{A502Y} knock-in mice. A plasmid expressing single-guide RNA (sgRNA) and Cas9 was constructed as described previously [31]. Plasmid vector pX330 [32] expressing Cas9 and gRNA was digested with BbsI and a pair of oligo DNA recognizing Nrf2 targeting site (5'-AAG TCG CCG CCC AGA ACT GT-3') was ligated to the linealized vector. Donor oligo DNA encoding substitution from alanine to tyrosine was designed as follows; 5'-ATC CGA GAT ATA CGC AGG AGA GGT AAG AAT AAA GTC TAC GCC CAG AAC TGT AGG AAA AGG AAG CTG GAG-3'. The plasmid and donor DNA were co-injected into BDF1 fertilized eggs. We obtained two lines of *Nrf2*^{A502Y} knock-in mice. All mice were handled according to Regulations for Animal Experiments and Related Activities at Tohoku University.

Mouse genotyping. Genomic DNA was extracted from a piece of tail. The DNA samples were genotyped by using TaqMan SNP Genotyping Assay System (Applied Biosystems). Wild type (WT) alleles were detected by 2'-chloro-7'-phenyl-1, 4-dichliro-6-carboxyfluorescein (VIC)-labeled probes, and mutant alleles were detected by 6-carboxyfluorescein (FAM)-labeled probes.

Genomic DNA and cDNA sequencing. Genomic DNA was extracted from a piece of tail. RNA extractions from brain, thymus, lung, heart, liver, pancreas, spleen, kidney, esophagus, skeletal muscle were conducted for cDNA synthesis. Targeted region of *Nrf2* gene was amplified by PCR using following primers; forward, 5'- AAG ACA AAC ATT CAA GCC GC-3'; reverse, 5'- GCT TTT GGG AAC AAG GAA CA-3'. The amplicon was sequenced using ABI 3100 sequencer. The primer sequence for sequence was 5'-GCT TTT GGG AAC AAG GAA CA-3'.

Peritoneal macrophage isolation and cell culture. 7-8 weeks of mice were received an intraperitoneal injection of 4% thioglycolate broth. Four days later, macrophages collected by intraperitoneal lavage were cultured in RPMI 1640 medium containing 10% fetal bovine serum and 1% penicillin-streptomycin [33]. For analysis of Nrf2-induced state, the macrophages were treated with 100-μM diethylmaleate (DEM). To test the cell viability after 12 hours of menadione treatment, Cell Count Reagent SF (nacalai tesque) was used. DEM and menadione were from Wako Pure Chemicals and Sigma-Aldrich, respectively.

Flow cytometry analysis. The cells were stained with antibodies to Gr-1, Mac1, and F4/80 conjugated with FITC, APC, and PE, respectively. These antibodies were from eBioscience. The stained cells were analyzed with FACSCanto II and the data analyses were performed with FlowJo software (Tree Star).

RNA extraction and quantitative RT-PCR. RNA was extracted with Sepasol-RNA I Super G (nacalai tesque) and reverse-transcribed with ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO) according to the manufacturer's instruction. Quantitative PCR was run on ABI7300 (Applied Biosystems). We used the following primers and probe to detect mRNA levels; Nrf2, forward primer, 5'-CAA GAC TTG GGC CAC TTA AAA GAC-3'; reverse primer, 5'-AGT AAG GCT TTC CAT CCT CAT CAC-3'; probe 5'-AGG CGG CTC AGC ACC TTG TAT CTT GA-3', 18S rRNA, forward primer, 5'-CGG CTA CCA CAT CCA AGG AA-3'; reverse primer, 5'-GCT GGA ATT ACC GCG GCT-3'; and Taqman probe, 5'-TGC TGG CAC CAG ACT TGC CCT C-3'.

Immunoblot analyses. Peritoneal macrophages were treated with 100-μM DEM for 3 hours.

Nuclear lysate for immunoblot was prepared using NE-PER Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific). 5 μg of nuclear lysate was subjected to immunoblot using anti-Nrf2 [34] and anti-Lamin B (M-20) (Santa-Cruz; sc-6217) antibodies. The densitometries of image were analyzed with ChemiDoc MP Imaging System (Bio-Rad), and normalized to Lamin B intensity.

ChIP-Seq analysis. For ChIP-Seq analysis, the peritoneal macrophages were treated with 100-μM DEM for 4 hours as described [11] with minor modifications. ChIP was performed with anti-Nrf2 antibody (Cell Signaling Technology; D1Z9C). DNA libraries were prepared from 1.5 or 2 ng of ChIP and input samples quantified with Qubit Fluorometer (Life Technologies), using Mondrian SP+ and Ovation SP Ultralow DR Multiplex System (TaKaRa). The constructed libraries were amplified by PCR and DNA fragments in 300-600 bp in size were yielded with AMPure XP Kit (BECKMAN COULTER). Prepared samples were quantified by quantitative MiSeq (qMiSeq) method [35], followed by high throughput sequencing using HiSeq2500 (Illumina) to generate 101 base-single reads. Three biological replicates of ChIPed DNA and Input DNA prepared from each genotype and ChIP-Seq analyses were conducted with these samples.

ChIP-Seq data analyses. The sequenced reads were mapped to the mouse genome (mm9) using Bowtie2 software [36]. The mapped tags were visualized by using Integrative Genomics Viewer [37]. Peak calling was performed using a model-based analysis of ChIP-seq (MACS) version 1.4.2 [38]. DNA motif construction was performed using MEME-ChIP version 4.10.0 [39]. Extraction of ARE motifs was performed using R based script.

RNA-Seq analysis. Total RNA was prepared by using RNeasy Mini Kit (QIAGEN) and 1.5 μg of total RNA was used for further steps. Isolation of poly(A)-tailed RNA and library construction were performed using Sureselect Strand Specific RNA Sample Prep Kit (Agilent Technologies). The libraries were sequenced using NextSeq500 (Illumina) for 86 cycles of single read. Three biological replicates were performed in each genotype.

RNA-Seq data analyses. TopHat [40] was used for mapping of RNA-Seq data, and Cufflinks version 2.1.1 [41] was used for quantifying the expression level of each gene as fragments per kilobase of exon per million fragments (FPKM) with default parameters. The differentially expressed genes were identified using Cuffdiff version 2.1.1, threshold of q value < 0.05. The KEGG pathway analysis was performed using DAVID Bioinformatics Resource 6.7（http://david.abcc.ncifcrf.gov/). The KEGG pathway significantly enriched were defined as p value < 0.05. The p values were corrected using Benjamini-Hochberg procedure. The gene set analysis was performed using the Gene Set Enrichment Analysis (GSEA) software [42]. The gene set was created by using data described in references [11, 43].

Acetaminophen (APAP) induced liver injury model. Following 16-hours fasting, 10-12 weeks male mice were treated with 125-mg/kg or 200-mg/kg APAP by intraperitoneal injection and sacrificed 6 hours after dosing. APAP was purchased from Sigma-Aldrich. Using blood serum obtained from posterior vena cava of anesthetized animal, plasma alanine transaminase (ALT) and asparate transaminase (AST) were determined using FUJI DRI-CHEM 7000V (FUJIFILM). Liver sample were fixed in 10% formalin solution and stained with hematoxylin and eosin (HE).

Accession number. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [44] and are accessible through GEO Series accession number GSE75177 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75177).

Results

Generation of *Nrf2*^{A502Y} knock-in mice. Dimetric transcription factors that contain Maf protein can bind various cis-acting element sequences. Whilst Maf homodimers recognize MARE sequences, CNC-sMaf heterodimers recognize CsMBE (Fig. 1A). Molecular basis of this *cis*-element selection resides in the structural difference in Maf and CNC transcription factors, and substituting an amino acid residue modifies this specificity. Substitution of Nrf2 Ala-502 residue to tyrosine brings in a significant difference in cistrome, and Nrf2^{A502Y} becomes recognizing Maf-oriented sequence. Accordingly, the recognition sequence specificity of Nrf2-sMaf heterodimer changes from CsMBE to MARE [29] (Figs. 1A and 1B). However, in biological context *in vivo*, the importance of Nrf2 binding specificity to CsMBE has not been fully evaluated.

To examine how germline modification to Nrf2^{A502Y} influences the Nrf2-sMaf activity, we generated *Nrf2*^{A502Y} knock-in mice using the CRISPR/Cas9 technology. To this end, we designed a guide RNA (gRNA) containing 20 nucleotides capable of recognizing the *Nrf2* target site followed by a protospacer adjacent motif (PAM) to recruit Cas9 to the target site (Fig. 1C). We generated a plasmid expressing both Cas9-encoding mRNA and the gRNA [31]. We next introduced 69-mer oligo-DNA including mutations from GCC to TAC resulting in substitution of the 502nd alanine to tyrosine (A502Y) for homologous recombination. We co-injected both plasmid and oligo-DNA into fertilized eggs. We obtained 24 pups. To verify homologous recombination of genomic DNA, we sequenced the targeted regions, and identified two pups carrying mono-allelic A502Y mutation. We then crossed these Nrf2^{A502Y} founder mice with wild-type mice and established two lines of knock-in substitution mice. Through genomic DNA sequencing analyses, we confirmed both TAC (encoding tyrosine) and GCC (encoding alanine) in the heterozygous (*Nrf2*$^{AY/+}$) offspring (Fig. 1D, middle panel).

To examine whether Nrf2^{A502Y} is expressed in the *Nrf2*$^{AY/+}$ mice, we prepared RNA samples from various tissues of the *Nrf2*$^{AY/+}$ mice, and synthesized Nrf2 cDNA and sequenced. We detected comparable level of TAC and GCC in all the tissues of the *Nrf2*$^{AY/+}$ mice examined, indicating successful homologous recombination of Nrf2^{A502Y} (Fig. 1E). We further crossed *Nrf2*$^{AY/+}$ mice and obtained homozygous (*Nrf2*$^{AY/AY}$) mice (Fig. 1D, lower panel). Body-weight-gain of both male and female *Nrf2*$^{AY/AY}$ mice is comparable with that of the wild-type mice (Fig. 1F) and the mice were fertile.

***Nrf2*$^{AY/AY}$ macrophages are more susceptible to the cytotoxic effect of xenobiotics.** To examine whether *Nrf2*$^{AY/AY}$ mice preserves cytoprotective activities assisted by Nrf2, we employed the thioglycolate-elicited peritoneal macrophage system [33]. We injected thioglycolate into *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice and harvested peritoneal macrophages (Fig. 2A). Almost all cells obtained from both *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ peritoneal lavage exhibited Mac1^{+}Gr1^{-}F4/80^{+} surface markers, indicating that macrophage induction was comparable between *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice (Fig. 2B).

We next treated the peritoneal macrophages, harvested both from *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice, with an electrophilic Nrf2 inducer DEM (Fig. 2A). We found that the Nrf2 mRNA level of *Nrf2*$^{AY/AY}$ macrophages was comparable to that of wild type under the basal and DEM-induced conditions (Fig. 2C). Furthermore, comparable level of Nrf2 protein was accumulated in the nucleus under the DEM-treated condition in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages (Figs. 2D and 2E). These results indicate that stress-responsiveness of *Nrf2*$^{AY/AY}$ macrophages were not significantly different from that of *Nrf2*$^{+/+}$, regarding intercellular Nrf2 localization and the abundance.

To test cytoprotective function of Nrf2^{A502Y}, we then examined susceptibility of *Nrf2*$^{AY/AY}$ macrophages against menadione, which is a free radical-generating compound and is

well-established stressor for testing roles played by Nrf2 in the oxidative stress response [45]. Cell viabilities of the *Nrf2*$^{AY/AY}$ macrophages were lower than those of the *Nrf2*$^{+/+}$ macrophages (Fig. 2F), indicating that the *Nrf2*$^{AY/AY}$ macrophages were more susceptible to toxicity of xenobiotics than the *Nrf2*$^{+/+}$ macrophages.

ChIP-Seq analyses of Nrf2 and Nrf2^{A502Y} reveal their preference of binding sequences. Since the *Nrf2*$^{AY/AY}$ macrophages were more susceptible to toxicity of xenobiotics, we assumed that Nrf2^{A502Y} might fail to recognize CsMBE in the regulatory regions of Nrf2 target genes and therefore fail to induce their expression. To confirm preferences of binding sequences of Nrf2 and Nrf2^{A502Y} *in vivo*, we performed ChIP-Seq analyses using an anti-Nrf2 antibody on the DEM-treated peritoneal macrophages derived from *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice. The ChIP-Seq analyses were performed using three biological replicates from each genotype. We defined Nrf2 and Nrf2^{A502Y} binding peaks as peaks called in three or two samples in the three replicates [46]. We obtained 1062 peaks for Nrf2 binding sites and 1304 peaks for Nrf2^{A502Y} binding sites (Fig. 3A). Of the 1062 Nrf2 binding sites, 669 peaks were recognized only by Nrf2. We thus designated the 669 sites as “WT-specific” sites. Meanwhile, we identified 911 peaks to which only Nrf2^{A502Y} bound. We named these 911 sites as “AY-specific” sites. Of these Nrf2 and Nrf2^{A502Y} peaks, 393 peaks overlapped between both Nrf2 and Nrf2^{A502Y}. We designated the 393 sites that both Nrf2 and Nrf2^{A502Y} bound to as “Common” sites. Typical peak profiles for Nrf2, Nrf2^{A502Y} and the overlap of Nrf2 and Nrf2^{A502Y} are shown in Figure 3B. We found that the Nrf2 binding to the common sites showed higher probability of binding than that of WT- and AY-specific sites, suggesting that Nrf2 and Nrf2^{A502Y} binding to Common sites is tighter than that to WT- and AY-specific sites (Fig. 3C). These results thus demonstrate that the alanine to tyrosine substitution of Nrf2 502 position elicits marked conformation change, so that target-binding

sites of Nrf2^{A502Y} *in vivo* are largely different from those of wild-type Nrf2.

To determine consensus binding motifs for Nrf2^{A502Y} and Nrf2 in WT-specific, Common and AY-specific sites, we extracted sequences within ±150 bp of each peak center and performed *de novo* motif analysis. Core sequences of TRE (position 1-7) neighbored by 3' GC motif (position 8 and 9) appeared to be similar in WT-specific, Common and AY-specific sites (Fig. 3D). Consistent with our previous report [11], nucleotides A or G (A/G) at 5'-end neighboring to TRE core sequence (position 0) was enriched in WT-specific sites (*E-value* = 5.8×10^{-724}), conforming our original observation that Nrf2 recognizes CsMBE. On the other hand, the most enriched nucleotide at position 0 of AY-specific sites was C, showing that the binding preference of Nrf2^{A502Y} mimics that of sMaf homodimer or MARE *in vivo* (*E-value* = 2.5×10^{-521}). Of note, we did not detect enrichment of a specific base at position 0 in Common sites by *de novo* motif analysis (*E-value* = 1.2×10^{-365}).

To analyze the nucleotide at position 0 in detail, we extract core motifs (position 1-9) within ±150 bp of each peak center and examined frequency of bases at position 0. Nucleotides A/G and C were enriched at position 0 of TRE in WT-specific and AY-specific sites, respectively, showing a good agreement with *de novo* motif analysis (Fig. 3E). On the other hand, we found that A, G, or C but not T were enriched at position 0 in Common sites, showing that Common sites exhibit DNA preference of either WT-specific or AY-specific sites at position 0 (Fig. 3E).

A number of previous papers show that TMA sequence located the 5' side of the CsMBE (position –5 to –3 in Fig. 3F, M represents A or C) influences activation of genes containing the element [11, 47-50]. Therefore, we examined prevalence of the TMA motif in 5' region of WT-specific, Common and AY-specific sites. We found that the TMA-motif, especially TCA-motif, was observed in 6.9% and 8.0% of CsMBE of WT-specific and Common sites, respectively. On the other hand, TMA sequence was not enriched in AY-specific sites (3.0%

of motif, Fig. 3F). These results suggest that TCA at position –5 to –3 may support the binding of Nrf2-sMaf heterodimer to CsMBE but not support the binding of Maf homodimer to MARE *in vivo*.

Impairment of transcriptional activity in Nrf2^{A502Y} macrophages. Since the preference of binding sequences of Nrf2^{A502Y} shifted from CsMBE to MARE, it is expected that Nrf2^{A502Y} might support expression of a distinct gene set from that supported by Nrf2. To examine this issue, we performed RNA-Seq analysis and compared gene expression profiles in peritoneal macrophages from *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice between basal and DEM-induced conditions. We found that expression levels of 1402 genes were significantly changed upon DEM stimulation in the *Nrf2*$^{+/+}$ macrophages, in which 696 genes were upregulated and 706 genes were downregulated (Fig. 4A). On the other hand, we found that expression levels of 402 genes were changed upon DEM stimulation in *Nrf2*$^{AY/AY}$ macrophages, in which 148 genes were upregulated and 254 genes were downregulated (Fig. 4B). Thus, the numbers of upregulated and downregulated genes were strikingly decreased in *Nrf2*$^{AY/AY}$ macrophages compared to those in *Nrf2*$^{+/+}$ macrophages.

Furthermore, the majority (309 out of 402 genes) of upregulated and downregulated genes [referred to as differentially expressed genes (DEGs)] in *Nrf2*$^{AY/AY}$ macrophages overlapped with DEGs in the *Nrf2*$^{+/+}$ macrophages (Figs. 4C). As shown in Figure 4D, we identified 1093 DEGs (586 and 507 genes were upregulated and downregulated, respectively) observed only in the *Nrf2*$^{+/+}$ macrophages (WT-specific DEGs), 309 DEGs (110 and 199 genes were upregulated and downregulated, respectively) observed in both *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages (Common DEGs), 93 DEGs (38 and 55 genes were upregulated and downregulated, respectively) observed only in the *Nrf2*$^{AY/AY}$ macrophages (AY-specific DEGs). The number of WT-specific DEGs was much larger than those of

common and AY-specific DEGs. These results unequivocally demonstrate that Nrf2^{A502Y} lacks the induction and repression abilities for the majority of Nrf2 target genes.

To annotate upregulated and downregulated genes in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages, we performed a KEGG pathway analysis. We found that known Nrf2-dependent pathways such as glutathione metabolism and pentose phosphate pathway were enriched in genes upregulated specifically in the *Nrf2*$^{+/+}$ macrophages (Fig. 4E). On the other hand, inflammation-related pathways such as chemokine signaling pathway, focal adhesion and leukocyte transendothelial migration pathways were enriched in the gene group specifically downregulated in the *Nrf2*$^{+/+}$ macrophages, showing very good agreement with the recent findings that Nrf2 regulates anti-inflammatory genes [51, 52].

Nrf2^{A502Y} fails to induce major cytoprotective genes. Since known Nrf2-dependent pathways were enriched in the gene set upregulated specifically in the *Nrf2*$^{+/+}$ macrophages, we next examined whether Nrf2^{A502Y} failed to induce known Nrf2 target genes. Our gene set enrichment analysis (GSEA) showed that differentially expressed genes only in the *Nrf2*$^{+/+}$ macrophages contained known Nrf2 target genes (Fig. 5A). We found that expression levels of genes related to quinone detoxification (*Nqo1*), glutathione (GSH) conjugation (*Gstm1* and *Gstp1*), GSH synthesis (*Gss*, *Gclm* and *Gclc*), GSH reduction (*Gsr*), hydrogen peroxide degradation (*Cat*), and pentose phosphate pathway (*Taldo1*) were induced specifically in the *Nrf2*$^{+/+}$ macrophages (Fig. 5B). In contrast, DEM induction of these genes was abrogated almost completely in the Nrf2^{A502Y} macrophages. Expression levels of genes related to heme degradation (*Hmox1*), transcription factor (*Mafg*), and autophagy (*Sqstm1*) were induced both in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages. These results indicate that Nrf2^{A502Y} lost ability to induce major cytoprotective genes.

To assess whether Nrf2 and Nrf2^{A502Y} directly regulate these genes, we examined

binding peaks of Nrf2 and Nrf2^{A502Y} ChIP-Seq in the proximity of these genes. Expectedly, the genes that were induced specifically in the *Nrf2*$^{+/+}$ macrophages, including *Nqo1*, *Gclm*, *Gss* and *Cat*, harbored WT-specific peaks (Fig. 5D). In addition, the genes that were induced both in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages, including *Mafg*, *Hmox1* and *Sqstm1*, harbored Common peaks (Fig. 5E). These results support our contention that differences in the Nrf2-sMaf cistrome indeed affect the gene expression profiles.

CsMBE recognition of Nrf2-sMaf is required for liver protection from APAP toxicity. To examine whether Nrf2^{A502Y} mutant mice are more susceptible to toxicity than wild-type mice, we finally examined susceptibility of *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice to acetaminophen (APAP) toxicity. We intraperitoneally administered low-dose (125 mg/kg) and high-dose (200 mg/kg) APAP to *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice, which were fasted for 16 hours beforehand. We analyzed these mice 6-hours after re-feeding (Fig. 7A).

Levels of liver damage indicators, ALT and AST, in the *Nrf2*$^{AY/AY}$ mice were significantly higher than those in the *Nrf2*$^{+/+}$ mice in the low-dose examination (Fig. 7B). While there were some fluctuations perhaps due to toxicity in the high-dose examination, the results showed reproducibility. Histological analysis revealed that liver damage in the *Nrf2*$^{AY/AY}$ mice was more severe than those in the *Nrf2*$^{+/+}$ mice (Fig. 7C). These results thus demonstrate that the Nrf2^{A502Y} mutant mice are more susceptible to the acute toxicity of APAP than wild-type mice. Taken together, this study supports the notion that Nrf2-sMaf specifically recognizes CsMBE sequences, which is necessary to the cytoprotective function.

Discussion

Since CsMBE and MARE share common core sequence, it has been uncertain how many target binding sites and target genes are shared by the Nrf2-sMaf heterodimers and Maf homodimers. In this study, we wish to clarify this issue, and have generated a knock-in line of mice expressing $Nrf2^{A502Y}$ mutant. As summarized in Figure 7, we first verified that CsMBE is substantially different from MARE *in vivo*, despite of their similarity in terms of DNA sequences. Of note, while Nrf2-sMaf prefers A/G nucleotide at 5'-flanking region of the core sequence (CsMBE, left side), $Nrf2^{A502Y}$-sMaf prefers C at that position (right side), similar to the Maf homodimer-binding site (MARE), demonstrating that the sequence recognition of $Nrf2^{A502Y}$ shifts drastically from CsMBE to MARE in peritoneal macrophages *in vivo*. The common binding sites of Nrf2 and $Nrf2^{A502Y}$ do not show preference between A/G/C nucleotides at the position. Of note, RNA-Seq data revealed that $Nrf2^{A502Y}$ substantially lost the ability to support the expression of majority of the cytoprotective genes and, showing very good agreement with the results, $Nrf2^{A502Y}$ mutant mice are severely susceptible to the APAP toxicity. Based on these results, we conclude that the Nrf2-sMaf heterodimers have acquired the CsMBE recognition during molecular evolution, and this progress is critical for the cytoprotective functions of our body.

An ancestor of CNC family proteins is SKN-1 in *Caenorhabditis elegans* [25, 53, 54]. SKN-1 regulates a set of cytoprotective genes responding to oxidative stress as is the case for vertebrate Nrf2 [53]. Of note, despite of the functional similarity to Nrf2, SKN-1 recognizes ATGA(G/A) motif as a monomer. On the other hand, a CNC family protein in *Drosophila melanogaster*, CncC, forms a heterodimer with sMaf protein, Maf-S, and together recognize CsMBE [(A/G)TGA(G/C)nnnGC] [23]. The SKN-1 binding motif is conserved within CsMBE (underlined), indicating the CNC transcription factors acquired ability to recognize extended *cis*-element by forming heterodimer with sMaf proteins in

process of the molecular evolution. We surmise that the acquired long *cis*-acting element enables vertebrates to execute the strict gene regulation in their huge genomes through competing with the other transcription factors sharing partly the cis-element. Furthermore, differences of CsMBE and MARE reduce mutual interference between CNC-sMaf heterodimers and Maf homodimers, resulting in selective activation of the genes required in the response against oxidative and xenobiotic stresses.

Several reports indicate the importance of TMA motif at 5' flanking region of CsMBE [47, 48, 50]. Showing good agreement with the studies, we found that TMA, especially TCA, motifs are enriched in WT-specific and Common sites but not in AY-specific sites. While the TCA motif appears to play important roles in Nrf2-sMaf heterodimer binding to CsMBE, it still remains unclear which factor recognizes the TCA motif and how the binding of Nrf2 to CsMBE is stabilized upon the presence of the TCA motif.

On the other hand, since $Nrf2^{A502Y}$-sMaf heterodimers recognize MARE, $Nrf2^{A502Y}$-sMaf has the potential to affect MARE-dependent transcription of large Maf and sMaf homodimers. Large Maf proteins (c-Maf, MafA/L-Maf, MafB and Nrl) play critical roles in maintenance of homeostasis and ontogeny, including lens development, glucose homeostasis and macrophage differentiation [53, 55-57]. Although accumulation of $Nrf2^{A502Y}$ may interfere these biological effects, we did not observe obvious abnormality in the $Nrf2^{AY/AY}$ mice except for high susceptibility to oxidative and xenobiotic stresses. These may be due to two reasons. First, Keap1 constitutively degrades Nrf2 and $Nrf2^{A502Y}$ under normal conditions, so that without a challenge of chemical Nrf2-inducers $Nrf2^{A502Y}$ does not accumulate massively in our body. The pharmacological induction of Nrf2 and $Nrf2^{A502Y}$ accumulation by DEM or other inducers is transient. Second, compared with the pharmacological induction, genetic induction by knocking-out or knocking-down of Keap1 is potent and constitutive, and therefore results in adverse effects [58, 59]. We surmise that

455 analyses using genetic induction of Nrf2 may provide clues about this issue.

456 In addition to WT-specific (CsMBE) and AY-specific (MARE) binding sites, we
457 identified the Common binding sites that bind both Nrf2-sMaf and Nrf2^{A502Y}-sMaf.
458 Nevertheless, it still remains enigmatic why both Nrf2-sMaf and Nrf2^{A502Y}-sMaf
459 heterodimers are able to bind to the Common sites. Our results show that the peaks
460 belonging the Common binding sites are highly enrichment compared with both WT-specific
461 and AY-specific binding sites, suggesting that the Common sites possess beneficial genomic
462 conditions for binding of these factors, such as adjacent interacting motifs of other
463 transcription factors that stabilize the binding of Nrf2-sMaf and sMaf homodimers.

464 We previously identified a competitive regulation between NF-E2 p45 and sMaf
465 heterodimer (p45-sMaf) and sMaf homodimer in mouse megakaryocytes [60]. In the study,
466 we identified that transgenic overexpression of sMaf severely repressed the p45-sMaf target
467 gene expression and resulting proplatelet formation. Our present results further support the
468 notion that binding of Nrf2-sMaf heterodimer to the Common genes seems to be competed
469 with the sMaf homodimer and *vise-versa*, and the competition between Nrf2-sMaf
470 heterodimer and sMaf homodimer affects expression profile of the Common genes. On the
471 other hand, WT-specific genes seem to be regulated by the Nrf2-sMaf heterodimer without
472 the competition.

473 In summary, we conclude that this study provides fundamental information that
474 enlightens the elaborate transcriptional regulation of a subset of cytoprotective genes. The
475 Nrf2-sMaf heterodimer sustains expression of the genes that are critical for cytoprotection
476 against oxidative and xenobiotic stresses. In order to achieve a quick response against these
477 stresses, it is crucial to select the target genes properly and timely, and CsMBE ensures the
478 Nrf2-sMaf heterodimer to specify the proper genes without the interference or competition
479 with large Maf and sMaf homodimers. Meanwhile, heterodimers of sMaf and CNC family

480 transcription factors including Nrf1, Nrf2, Nrf3 and NF-E2 p45 regulates different target
481 genes via CsMBE recognition, indicating that the specificity of their gene regulation also
482 exists on different level (*e.g.* co-activator/co-repressor selection or epigenetic regulation) in
483 addition to simple recognition of DNA sequences [61]. Further analyses are necessary to
484 elucidate mechanisms how these transcription factors select specific target genes and exert
485 their diverse biological functions.
486

Acknowledgments

We thank Ms. Eriko Naganuma and the Biomedical Research Core of Tohoku University Graduate School of Medicine for technical support. This work was supported in part by MEXT/JSPS KAKENHI (24249015, 26111002 and 15H02507 to M.Y., and 26461395 to M.S.), AMED-CREST (to M.Y.), MEXT [a research program of the Project for Development of Innovative Research on Cancer Therapeutics (P-Direct)], the Naito Foundation, the Mitsubishi Foundation, and the Takeda Science Foundation (to M.Y.). A.O. was a JSPS Research Fellow.

REFERENCES

[1] Itoh, K.; Chiba, T.; Takahashi, S.; Ishii, T.; Igarashi, K.; Katoh, Y.; Oyake, T.; Hayashi, N.; Satoh, K.; Hatayama, I.; Yamamoto, M.; Nabeshima, Y. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem Biophys Res Commun* **236:**313-322; 1997.

[2] Taguchi, K.; Motohashi, H.; Yamamoto, M. Molecular mechanisms of the Keap1–Nrf2 pathway in stress response and cancer evolution. *Genes Cells* **16:**123-140; 2011.

[3] Motohashi, H.; Katsuoka, F.; Engel, J. D.; Yamamoto, M. Small Maf proteins serve as transcriptional cofactors for keratinocyte differentiation in the Keap1-Nrf2 regulatory pathway. *Proc Natl Acad Sci U S A* **101:**6379-6384; 2004.

[4] Itoh, K.; Wakabayashi, N.; Katoh, Y.; Ishii, T.; Igarashi, K.; Engel, J. D.; Yamamoto, M. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev* **13:**76-86; 1999.

[5] Kobayashi, A.; Kang, M. I.; Okawa, H.; Ohtsuji, M.; Zenke, Y.; Chiba, T.; Igarashi, K.; Yamamoto, M. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. *Mol Cell Biol* **24:**7130-7139; 2004.

[6] Igarashi, K.; Kataoka, K.; Itoh, K.; Hayashi, N.; Nishizawa, M.; Yamamoto, M. Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins. *Nature* **367:**568-572; 1994.

[7] Itoh, K.; Igarashi, K.; Hayashi, N.; Nishizawa, M.; Yamamoto, M. Cloning and characterization of a novel erythroid cell-derived CNC family transcription factor heterodimerizing with the small Maf family proteins. *Mol Cell Biol* **15:**4184-4193; 1995.

[8] Johnsen, O.; Skammelsrud, N.; Luna, L.; Nishizawa, M.; Prydz, H.; Kolstø, A. B. Small Maf proteins interact with the human transcription factor TCF11/Nrf1/LCR-F1. *Nucleic Acids Res* **24:**4289-4297; 1996.

[9] Oyake, T.; Itoh, K.; Motohashi, H.; Hayashi, N.; Hoshino, H.; Nishizawa, M.; Yamamoto, M.; Igarashi, K. Bach proteins belong to a novel family of BTB-basic leucine zipper transcription factors that interact with MafK and regulate transcription through the NF-E2 site. *Mol Cell Biol* **16:**6083-6095; 1996.

[10] Rushmore, T. H.; Morton, M. R.; Pickett, C. B. The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. *J Biol Chem* **266:**11632-11639; 1991.

[11] Hirotsu, Y.; Katsuoka, F.; Funayama, R.; Nagashima, T.; Nishida, Y.; Nakayama, K.; Engel, J. D.; Yamamoto, M. Nrf2-MafG heterodimers contribute globally to antioxidant and metabolic networks. *Nucleic Acids Res* **40:**10228-10239; 2012.

[12] Friling, R. S.; Bensimon, A.; Tichauer, Y.; Daniel, V. Xenobiotic-inducible expression of murine glutathione S-transferase Ya subunit gene is controlled by an electrophile-responsive element. *Proc Natl Acad Sci U S A* **87:**6258-6262; 1990.

[13] Mignotte, V.; Eleouet, J. F.; Raich, N.; Romeo, P. H. Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. *Proc Natl Acad Sci U S A* **86:**6548-6552; 1989.

[14] Kataoka, K.; Noda, M.; Nishizawa, M. Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol Cell Biol* **14:**700-712; 1994.

[15] Kataoka, K.; Fujiwara, K. T.; Noda, M.; Nishizawa, M. MafB, a new Maf family transcription activator that can associate with Maf and Fos but not with Jun. *Mol Cell Biol*

14:7581-7591; 1994.
[16] Kerppola, T. K.; Curran, T. A conserved region adjacent to the basic domain is required for recognition of an extended DNA binding site by Maf/Nrl family proteins. *Oncogene* **9:**3149-3158; 1994.
[17] Kataoka, K.; Noda, M.; Nishizawa, M. Maf nuclear oncoprotein recognizes sequences related to an AP-1 site and forms heterodimers with both Fos and Jun. *Mol Cell Biol* **14:**700-712; 1994.
[18] Yamamoto, T.; Kyo, M.; Kamiya, T.; Tanaka, T.; Engel, J. D.; Motohashi, H.; Yamamoto, M. Predictive base substitution rules that determine the binding and transcriptional specificity of Maf recognition elements. *Genes Cells* **11:**575-591; 2006.
[19] Kurokawa, H.; Motohashi, H.; Sueno, S.; Kimura, M.; Takagawa, H.; Kanno, Y.; Yamamoto, M.; Tanaka, T. Structural basis of alternative DNA recognition by Maf transcription factors. *Mol Cell Biol* **29:**6232-6244; 2009.
[20] Kusunoki, H.; Motohashi, H.; Katsuoka, F.; Morohashi, A.; Yamamoto, M.; Tanaka, T. Solution structure of the DNA-binding domain of MafG. *Nat Struct Biol* **9:**252-256; 2002.
[21] Consortium, C. e. S. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282:**2012-2018; 1998.
[22] Li, M. A.; Alls, J. D.; Avancini, R. M.; Koo, K.; Godt, D. The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila. *Nat Cell Biol* **5:**994-1000; 2003.
[23] Rahman, M. M.; Sykiotis, G. P.; Nishimura, M.; Bodmer, R.; Bohmann, D. Declining signal dependence of Nrf2-MafS-regulated gene expression correlates with aging phenotypes. *Aging Cell* **12:**554-562; 2013.
[24] Takagi, Y.; Kobayashi, M.; Li, L.; Suzuki, T.; Nishikawa, K.; Yamamoto, M. MafT, a new member of the small Maf protein family in zebrafish. *Biochem Biophys Res Commun* **320:**62-69; 2004.
[25] Blackwell, T. K.; Bowerman, B.; Priess, J. R.; Weintraub, H. Formation of a monomeric DNA binding domain by Skn-1 bZIP and homeodomain elements. *Science* **266:**621-628; 1994.
[26] McGinnis, N.; Ragnhildstveit, E.; Veraksa, A.; McGinnis, W. A cap 'n' collar protein isoform contains a selective Hox repressor function. *Development* **125:**4553-4564; 1998.
[27] Kobayashi, M.; Itoh, K.; Suzuki, T.; Osanai, H.; Nishikawa, K.; Katoh, Y.; Takagi, Y.; Yamamoto, M. Identification of the interactive interface and phylogenic conservation of the Nrf2-Keap1 system. *Genes Cells* **7:**807-820; 2002.
[28] Timme-Laragy, A. R.; Karchner, S. I.; Franks, D. G.; Jenny, M. J.; Harbeitner, R. C.; Goldstone, J. V.; McArthur, A. G.; Hahn, M. E. Nrf2b, novel zebrafish paralog of oxidant-responsive transcription factor NF-E2-related factor 2 (NRF2). *J Biol Chem* **287:**4609-4627; 2012.
[29] Kimura, M.; Yamamoto, T.; Zhang, J.; Itoh, K.; Kyo, M.; Kamiya, T.; Aburatani, H.; Katsuoka, F.; Kurokawa, H.; Tanaka, T.; Motohashi, H.; Yamamoto, M. Molecular basis distinguishing the DNA binding profile of Nrf2-Maf heterodimer from that of Maf homodimer. *J Biol Chem* **282:**33681-33690; 2007.
[30] Dlakić, M.; Grinberg, A. V.; Leonard, D. A.; Kerppola, T. K. DNA sequence-dependent folding determines the divergence in binding specificities between Maf and other bZIP proteins. *EMBO J* **20:**828-840; 2001.
[31] Mashiko, D.; Fujihara, Y.; Satouh, Y.; Miyata, H.; Isotani, A.; Ikawa, M. Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci Rep* **3:**3355; 2013.

[32] Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:**819-823; 2013.
[33] Ishii, T.; Itoh, K.; Takahashi, S.; Sato, H.; Yanagawa, T.; Katoh, Y.; Bannai, S.; Yamamoto, M. Transcription factor Nrf2 coordinately regulates a group of oxidative stress-inducible genes in macrophages. *J Biol Chem* **275:**16023-16029; 2000.
[34] Maruyama, A.; Tsukamoto, S.; Nishikawa, K.; Yoshida, A.; Harada, N.; Motojima, K.; Ishii, T.; Nakane, A.; Yamamoto, M.; Itoh, K. Nrf2 regulates the alternative first exons of CD36 in macrophages through specific antioxidant response elements. *Arch Biochem Biophys* **477:**139-145; 2008.
[35] Katsuoka, F.; Yokozawa, J.; Tsuda, K.; Ito, S.; Pan, X.; Nagasaki, M.; Yasuda, J.; Yamamoto, M. An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Anal Biochem* **466:**27-29; 2014.
[36] Langmead, B.; Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:**357-359; 2012.
[37] Thorvaldsdóttir, H.; Robinson, J. T.; Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14:**178-192; 2013.
[38] Zhang, Y.; Liu, T.; Meyer, C. A.; Eeckhoute, J.; Johnson, D. S.; Bernstein, B. E.; Nusbaum, C.; Myers, R. M.; Brown, M.; Li, W.; Liu, X. S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9:**R137; 2008.
[39] Machanick, P.; Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27:**1696-1697; 2011.
[40] Trapnell, C.; Pachter, L.; Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:**1105-1111; 2009.
[41] Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:**511-515; 2010.
[42] Subramanian, A.; Tamayo, P.; Mootha, V. K.; Mukherjee, S.; Ebert, B. L.; Gillette, M. A.; Paulovich, A.; Pomeroy, S. L.; Golub, T. R.; Lander, E. S.; Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102:**15545-15550; 2005.
[43] Hayes, J. D.; Dinkova-Kostova, A. T. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends Biochem Sci* **39:**199-218; 2014.
[44] Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:**207-210; 2002.
[45] Higgins, L. G.; Kelleher, M. O.; Eggleston, I. M.; Itoh, K.; Yamamoto, M.; Hayes, J. D. Transcription factor Nrf2 mediates an adaptive response to sulforaphane that protects fibroblasts in vitro against the cytotoxic effects of electrophiles, peroxides and redox-cycling agents. *Toxicol Appl Pharmacol* **237:**267-280; 2009.
[46] Yang, Y.; Fear, J.; Hu, J.; Haecker, I.; Zhou, L.; Renne, R.; Bloom, D.; McIntyre, L. M. Leveraging biological replicates to improve analysis in ChIP-seq experiments. *Comput Struct Biotechnol J* **9:**e201401002; 2014.
[47] Nioi, P.; McMahon, M.; Itoh, K.; Yamamoto, M.; Hayes, J. D. Identification of a novel Nrf2-regulated antioxidant response element (ARE) in the mouse NAD(P)H:quinone oxidoreductase 1 gene: reassessment of the ARE consensus sequence. *Biochem J* **374:**337-348; 2003.

[48] Wasserman, W. W.; Fahl, W. E. Functional antioxidant responsive elements. *Proc Natl Acad Sci U S A* **94:**5361-5366; 1997.
[49] Katsuoka, F.; Motohashi, H.; Engel, J. D.; Yamamoto, M. Nrf2 transcriptionally activates the mafG gene through an antioxidant response element. *J Biol Chem* **280:**4483-4490; 2005.
[50] Favreau, L. V.; Pickett, C. B. The rat quinone reductase antioxidant response element. Identification of the nucleotide sequence required for basal and inducible activity and detection of antioxidant response element-binding proteins in hepatoma and non-hepatoma cell lines. *J Biol Chem* **270:**24468-24474; 1995.
[51] Itoh, K.; Mochizuki, M.; Ishii, Y.; Ishii, T.; Shibata, T.; Kawamoto, Y.; Kelly, V.; Sekizawa, K.; Uchida, K.; Yamamoto, M. Transcription factor Nrf2 regulates inflammation by mediating the effect of 15-deoxy-Delta(12,14)-prostaglandin j(2). *Mol Cell Biol* **24:**36-45; 2004.
[52] Kobayashi, E.; Suzuki, T.; Yamamoto, M. Roles nrf2 plays in myeloid cells and related disorders. *Oxid Med Cell Longev* **2013:**529219; 2013.
[53] An, J. H.; Blackwell, T. K. SKN-1 links C. elegans mesendodermal specification to a conserved oxidative stress response. *Genes Dev* **17:**1882-1893; 2003.
[54] Carroll, A. S.; Gilbert, D. E.; Liu, X.; Cheung, J. W.; Michnowicz, J. E.; Wagner, G.; Ellenberger, T. E.; Blackwell, T. K. SKN-1 domain folding and basic region monomer stabilization upon DNA binding. *Genes Dev* **11:**2227-2238; 1997.
[55] Takeuchi, T.; Kudo, T.; Ogata, K.; Hamada, M.; Nakamura, M.; Kito, K.; Abe, Y.; Ueda, N.; Yamamoto, M.; Engel, J. D.; Takahashi, S. Neither MafA/L-Maf nor MafB is essential for lens development in mice. *Genes Cells* **14:**941-947; 2009.
[56] Zhang, C.; Moriguchi, T.; Kajihara, M.; Esaki, R.; Harada, A.; Shimohata, H.; Oishi, H.; Hamada, M.; Morito, N.; Hasegawa, K.; Kudo, T.; Engel, J. D.; Yamamoto, M.; Takahashi, S. MafA is a key regulator of glucose-stimulated insulin secretion. *Mol Cell Biol* **25:**4969-4976; 2005.
[57] Nakamura, M.; Hamada, M.; Hasegawa, K.; Kusakabe, M.; Suzuki, H.; Greaves, D. R.; Moriguchi, T.; Kudo, T.; Takahashi, S. c-Maf is essential for the F4/80 expression in macrophages in vivo. *Gene* **445:**66-72; 2009.
[58] Wakabayashi, N.; Itoh, K.; Wakabayashi, J.; Motohashi, H.; Noda, S.; Takahashi, S.; Imakado, S.; Kotsuji, T.; Otsuka, F.; Roop, D. R.; Harada, T.; Engel, J. D.; Yamamoto, M. Keap1-null mutation leads to postnatal lethality due to constitutive Nrf2 activation. *Nat Genet* **35:**238-245; 2003.
[59] Taguchi, K.; Maher, J. M.; Suzuki, T.; Kawatani, Y.; Motohashi, H.; Yamamoto, M. Genetic analysis of cytoprotective functions supported by graded expression of Keap1. *Mol Cell Biol* **30:**3016-3026; 2010.
[60] Motohashi, H.; Katsuoka, F.; Shavit, J. A.; Engel, J. D.; Yamamoto, M. Positive or negative MARE-dependent transcriptional regulation is determined by the abundance of small Maf proteins. *Cell* **103:**865-875; 2000.
[61] Ohtsuji, M.; Katsuoka, F.; Kobayashi, A.; Aburatani, H.; Hayes, J. D.; Yamamoto, M. Nrf1 and Nrf2 play distinct roles in activation of antioxidant response element-dependent genes. *J Biol Chem* **283:**33554-33562; 2008.

Figure legends

Figure 1. Generation of Nrf2^{A502Y} knock-in mice. (A) Cis-element recognition by Nrf2-sMaf heterodimers and Maf homodimers. Nrf2-sMaf heterodimer recognizes CsMBE, while Maf homodimer recognizes MARE. An amino acid substitution of Nrf2 502nd alanine to tyrosine is expected to change the recognition specificity of Nrf2-sMaf heterodimers from CsMBE to MARE. Critical GC sequence of MARE in the 5' of TRE and corresponding A/G sequence of CsMBE are underlined, and described as lock-and-key models in the scheme. (B) Domain structure of Nrf2 protein. The 502nd residue of alanine (A, blue letter) positioned at Nrf2-ECH homology 1 (Neh1) domain is replaced by tyrosine (Y, red letter). (C) Cas9/guide RNA (gRNA)-targeting site in *Nrf2* gene. The sequences of gRNA and the donor DNA co-injected for targeting mutagenesis are underlined. The proto spacer adjacent motif (PAM) sequence is indicated by green. The 502nd alanine residue (blue) and corresponding tyrosine residue (red) are shown. (D) Representative sequences of Nrf2 targeting region of genomic DNA from *Nrf2*$^{+/+}$, *Nrf2*$^{AY/+}$, and *Nrf2*$^{AY/AY}$ mice. The codons encoding alanine (GCC) and tyrosine (TAC) are underlined. (E) Representative sequences of Nrf2 targeting region of cDNA in *Nrf2*$^{AY/+}$ mice. (F) The growth curves of *Nrf2*$^{+/+}$ (+/+, blue circle, male n=5, female n=8) and *Nrf2*$^{AY/AY}$ (AY/AY, pink square, male n=6, female n=6) male and female mice. Data represent the mean ± standard deviations (S.D.).

Figure 2. *Nrf2$^{AY/AY}$* macrophages are susceptible to cytotoxic effects of xenobiotics. (A) Scheme for induction of peritoneal macrophages and Nrf2 accumulation. (B) Cellular surface marker profiles of *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages. The cells recovered from *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ mice were stained with Mac1, Gr-1, and F4/80 antibodies. (C) Nrf2 mRNA levels in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages under basal (white) and DEM-treated (gray) states. The abundance of each mRNA was normalized to rRNA.

713 Average values for *Nrf2*$^{+/+}$ macrophages under basal state were set to 1. *Nrf2*$^{+/+}$ n=4,
714 *Nrf2*$^{AY/AY}$ n=6. (D and E) Protein levels in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages under basal
715 (white) and DEM-treated (gray) states. Average values for *Nrf2*$^{+/+}$ macrophages under basal
716 state were set to 1. Lamin B was used as a loading control. *Nrf2*$^{+/+}$ n=6, *Nrf2*$^{AY/AY}$ n=5. (F)
717 Relative viabilities of *Nrf2*$^{+/+}$ (WT, blue circle, n=5) and *Nrf2*$^{AY/AY}$ (AY, red square, n=8)
718 macrophages under menadione treatment. Note that *Nrf2*$^{AY/AY}$ macrophages are prone to die
719 upon the menadione treatment. Graph data represent the mean ± SD. Student's *t*-test
720 (two-tailed), **P*<0.05, n.s., not significant.
721
722 **Figure 3. A502Y mutation of Nrf2 converts the binding preference of Nrf2 *in vivo*.** (A)
723 Venn diagram showing the overlap between the Nrf2 (blue circle) and Nrf2^{A502Y} (red circle)
724 binding sites. The numbers in parenthesis show total numbers of Nrf2 or Nrf2^{A502Y} binding
725 sites. (B) Representative binding peaks of Nrf2 and Nrf2^{A502Y}. The representative histograms
726 of tag count observed in ChIP-ed and Input samples from WT-specific, Common, and
727 AY-specific sites are shown. Scale bars, 1 kb. (C) Fold-enrichment (ChIP-ed/Input) values
728 of Nrf2 (left panel) and Nrf2^{A502Y} (right panel) binding at WT-specific (WT-sp.), Common
729 (Comm.), and AY-specific (AY-sp.) sites. In the box plots, bottom and top of the boxes
730 correspond to the 25th and 75th percentiles and the internal band is the median. The bars
731 outside the boxes indicate the highest and lowest data within 1.5 interquartile ranges of the
732 upper and lower quartiles, respectively. Wilcoxon rank sum test, *** p<0.001. (D) The
733 motifs enriched in Nrf2 and Nrf2^{A502Y} binding sites. These motifs were identified using *de*
734 *novo* motif-discovery algorithm MEME-ChIP version 4.10.0. (E) Frequency of nucleotides
735 at position 0 in WT-specific (WT-sp.), Common (Comm.) and AY-specific (AY-sp.) sites.
736 Note that AY-sp. prefers C at this position, as is the case for MARE, while WT-sp. prefers
737 A/G at this position. (F) Frequency of TMA-containing CsMBE in WT-specific, Common,

and AY-specific sites. The consensus sequence of the TMA-containing CsMBE is shown in top panel. The TMA motif (position –5 to –3, M=A or C) is indicated in red.

Figure 4. Substitution of Nrf2 to Nrf2^{A502Y} abrogates electrophilic stress response in mice. (A and B) Scatter plots comparing transcript levels at basal (x-axis) and DEM-induced (y-axis) states in the *Nrf2*$^{+/+}$ (WT, panel A) and *Nrf2*$^{AY/AY}$ (AY, panel B) macrophages. We found that transcript levels of 696 and 706 genes were significantly upregulated and downregulated by DEM (q<0.05), respectively, in WT mouse macrophages. Similarly, transcript levels of 149 and 254 genes were significantly upregulated and downregulated by DEM (q<0.05), respectively, in *Nrf2*$^{AY/AY}$ mouse macrophages. These genes are plotted. The numbers of upregulated (UP) and downregulated (DOWN) genes are shown at the upper left and lower right, respectively. (C) Venn diagram showing the overlap between differentially expressed genes (DEGs). DEGs in the *Nrf2*$^{+/+}$ (WT) and *Nrf2*$^{AY/AY}$ (AY) macrophages are shown in blue and red, respectively. The numbers in parenthesis show total numbers of DEGs. Note that the genes responding to DEM, irrespective to the upregulated or downregulated, are significantly reduced in *Nrf2*$^{AY/AY}$ macrophages. (D) Heat maps showing relative expression levels of WT-specific (left), Common (middle) and AY-specific (right) DEGs. The numbers of upregulated (UP) and downregulated (DOWN) genes are shown above the heat maps. (E) KEGG biological pathways enriched in WT-specific DEGs.

Figure 5. Nrf2^{A502Y} mutant fails to induce Nrf2 target genes. (A) GSEA histogram of the gene set containing well-known Nrf2 target genes. We compared a gene set that contains Nrf2-target genes differentially expressed by DEM specifically in the *Nrf2*$^{+/+}$ macrophages but not in Nrf2^{A502Y} macrophages with a known Nrf2-dependent gene set [11, 43]. The enrichment score (ES) and the nominal p value are indicated. The gene expression spectrum

763 (red to blue) is shown in the bottom of the histogram. Note that known Nrf2 target genes are
764 contained in the DEM-upregulated gene group in the *Nrf2*$^{+/+}$ macrophages. (B and C)
765 Transcriptome analyses of representative genes induced by DEM. Typical Nrf2-target genes
766 that are induced specifically in *Nrf2*$^{+/+}$ (WT, n=3) macrophages, but not in the *Nrf2*$^{AY/AY}$
767 (AY, n=3) macrophages, as shown in panel B. Note that there exist a group of genes that are
768 induced by DEM both in the *Nrf2*$^{+/+}$ and *Nrf2*$^{AY/AY}$ macrophages, as shown in panel C.
769 Expression levels of the genes are expressed as FPKM (fragments per kilobase of exon per
770 million fragments). Graph data represent the mean ± SD. Paired t-test, **p*<0.05. (D and E)
771 Representative histograms of Nrf2 and Nrf2^{A502Y} occupancy in proximal region of
772 representative Nrf2 target genes. WT-specific binding and common binding are shown in
773 panels D and E, respectively.
774
775 **Figure 6. Nrf2^{A502Y} mice are sensitive to APAP-induced liver injury.** (A) Scheme for the
776 analysis of sensitivity of Nrf2^{A502Y} mice to acetaminophen (APAP) toxicity. (B) AST (left)
777 and ALT (right) levels in the plasma of *Nrf2*$^{+/+}$ (WT, blue circle) and *Nrf2*$^{AY/AY}$ (AY, red
778 square) mice. The plots and bars show individual values and means, respectively. Student's
779 *t*-test (two-tailed), **p*<0.05. (C) Liver pathology of high-dose APAP-treated *Nrf2*$^{+/+}$ and
780 *Nrf2*$^{AY/AY}$ mice. HE staining of representative liver sections of *Nrf2*$^{+/+}$ (WT) and *Nrf2*$^{AY/AY}$
781 (AY) mice are shown. Scale bars, 100 μm.
782
783 **Figure 7. Schematic diagram of target recognition by Nrf2 and Nrf2^{A502Y}.** Nrf2-sMaf
784 prefers CsMBE harboring an A/G nucleotide at the 5'-flanking region of the core sequence
785 (left side), while Nrf2^{A502Y}-sMaf prefers MARE harboring a C nucleotide at that position
786 (right side), similar to sMaf homodimer. The common binding sites of Nrf2 and Nrf2^{A502Y} do
787 not show preference between A/G/C nucleotides at that position (middle). Majority of the

788 cytoprotective genes (*e.g.,* detoxifying and antioxidant genes) are regulated by Nrf2 in a
789 CsMBE-dependent manner. Nrf2^{A502Y} fails to recognize CsMBE and therefore fails to
790 induce cytoprotective genes, which results in a weak defense in the *Nrf2*$^{AY/AY}$ mice.

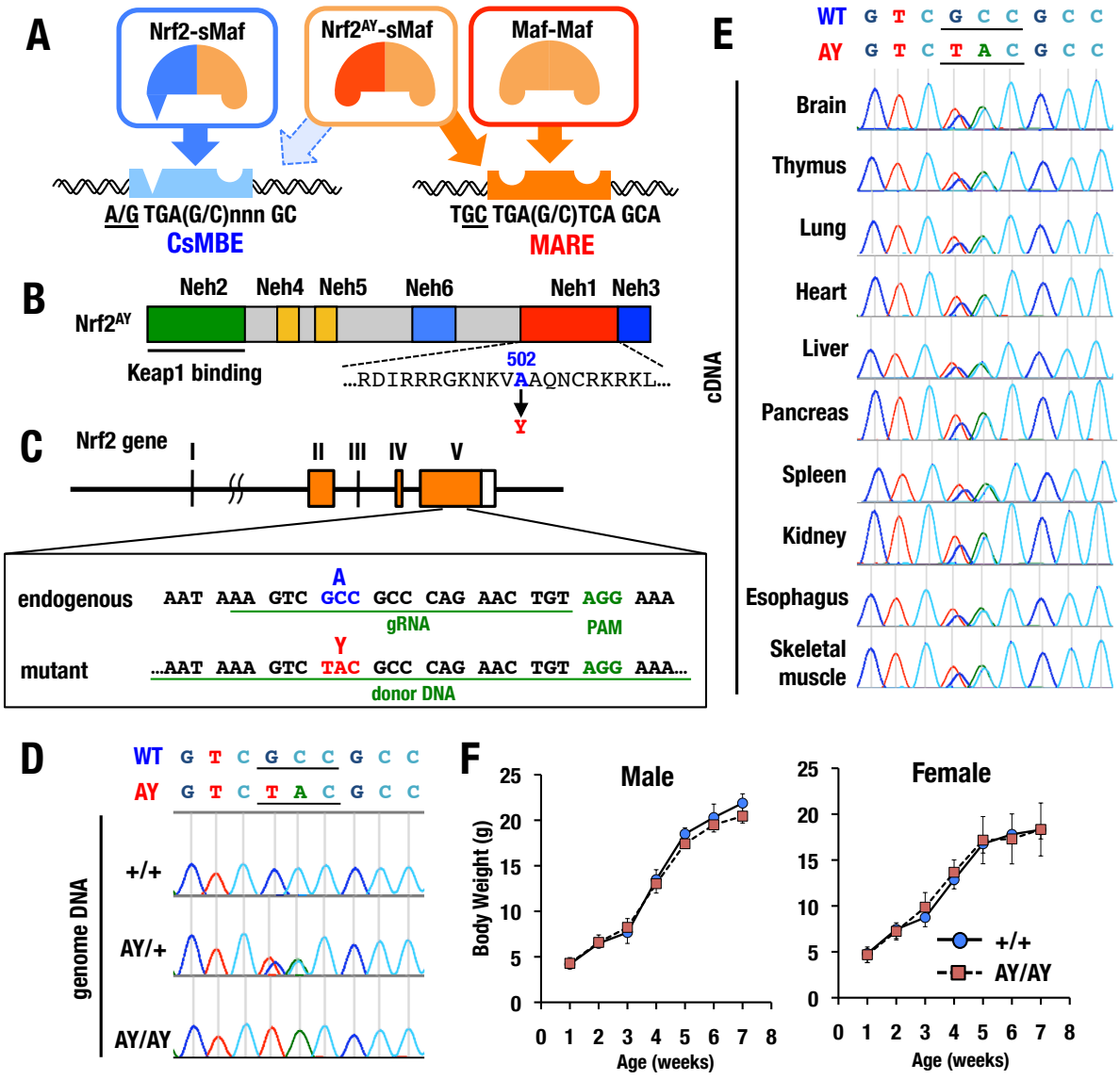

A
Nrf2-sMaf
Nrf2AY-sMaf
Maf-Maf
A/G TGA(G/C)nnn GC
CsMBE
TGC TGA(G/C)TCA GCA
MARE
B
Nrf2AY
Neh2
Neh4
Neh5
Neh6
Neh1
Neh3
Keap1 binding
502
...RDIRRRGKNKVAAQNCRKRKL...
Y
C
Nrf2 gene
I
II
III
IV
V
endogenous
AAT AAA GTC GCC GCC CAG AAC TGT AGG AAA
A
gRNA
PAM
mutant
...AAT AAA GTC TAC GCC CAG AAC TGT AGG AAA...
Y
donor DNA
D
WT G T C G C C G C C
AY G T C T A C G C C
genome DNA
+/+
AY/+
AY/AY
E
WT G T C G C C G C C
AY G T C T A C G C C
cDNA
Brain
Thymus
Lung
Heart
Liver
Pancreas
Spleen
Kidney
Esophagus
Skeletal muscle
F
Male
Female
Body Weight (g)
Age (weeks)
+/+
AY/AY

Figure 1

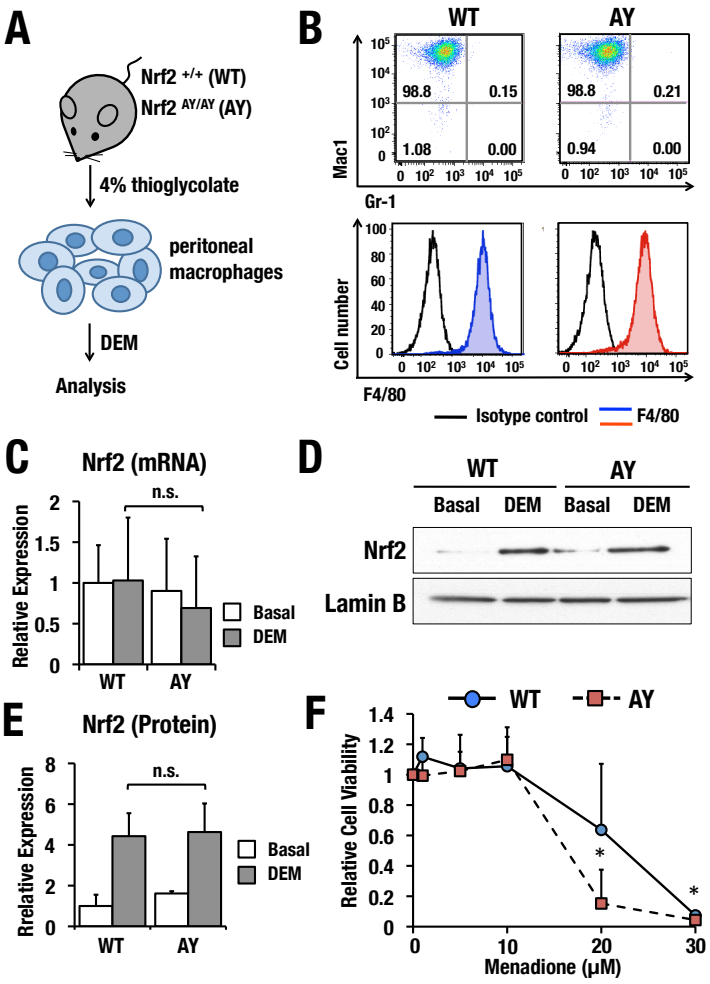

A
Nrf2 +/+ (WT)
Nrf2 AY/AY (AY)
4% thioglycolate
peritoneal macrophages
DEM
Analysis
B
WT
AY
98.8
0.15
1.08
0.00
98.8
0.21
0.94
0.00
Mac1
Gr-1
Cell number
F4/80
Isotype control
F4/80
C
Nrf2 (mRNA)
n.s.
Relative Expression
Basal
DEM
WT
AY
D
WT
AY
Basal
DEM
Basal
DEM
Nrf2
Lamin B
E
Nrf2 (Protein)
n.s.
Rrelative Expression
Basal
DEM
WT
AY
F
WT
AY
Relative Cell Viability
Menadione (μM)

Figure 2

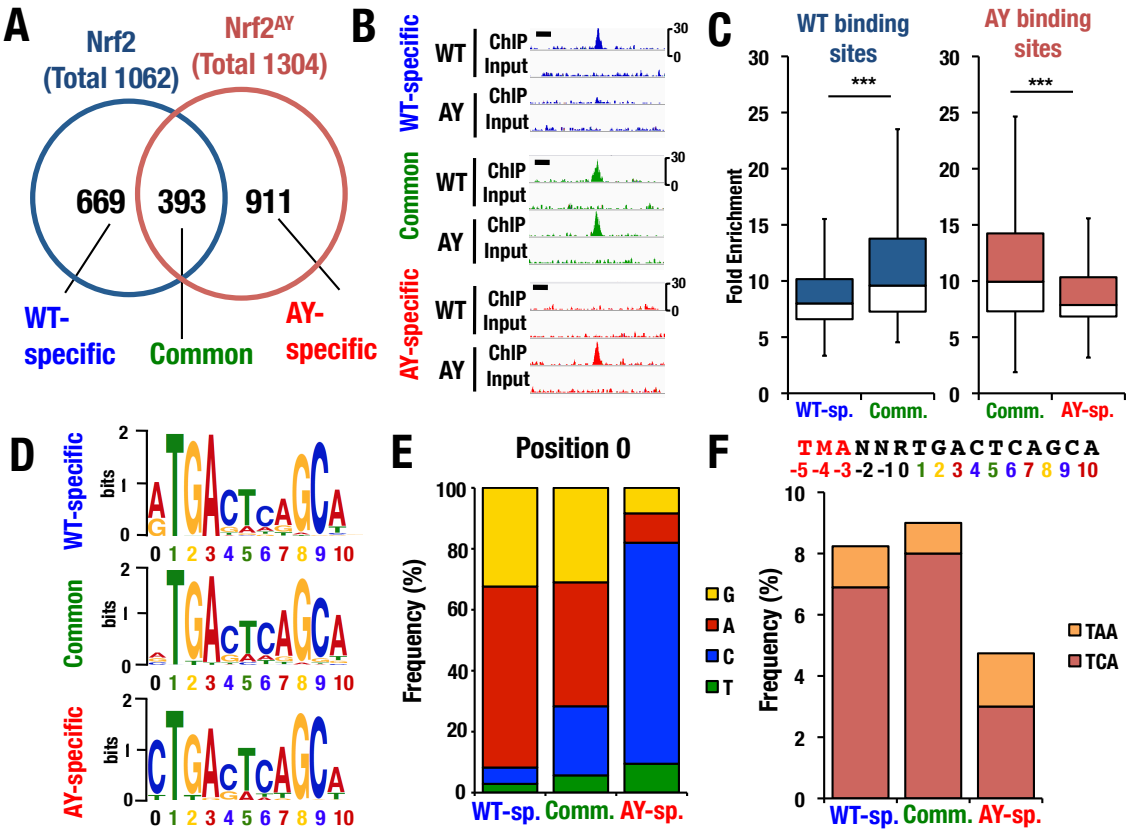
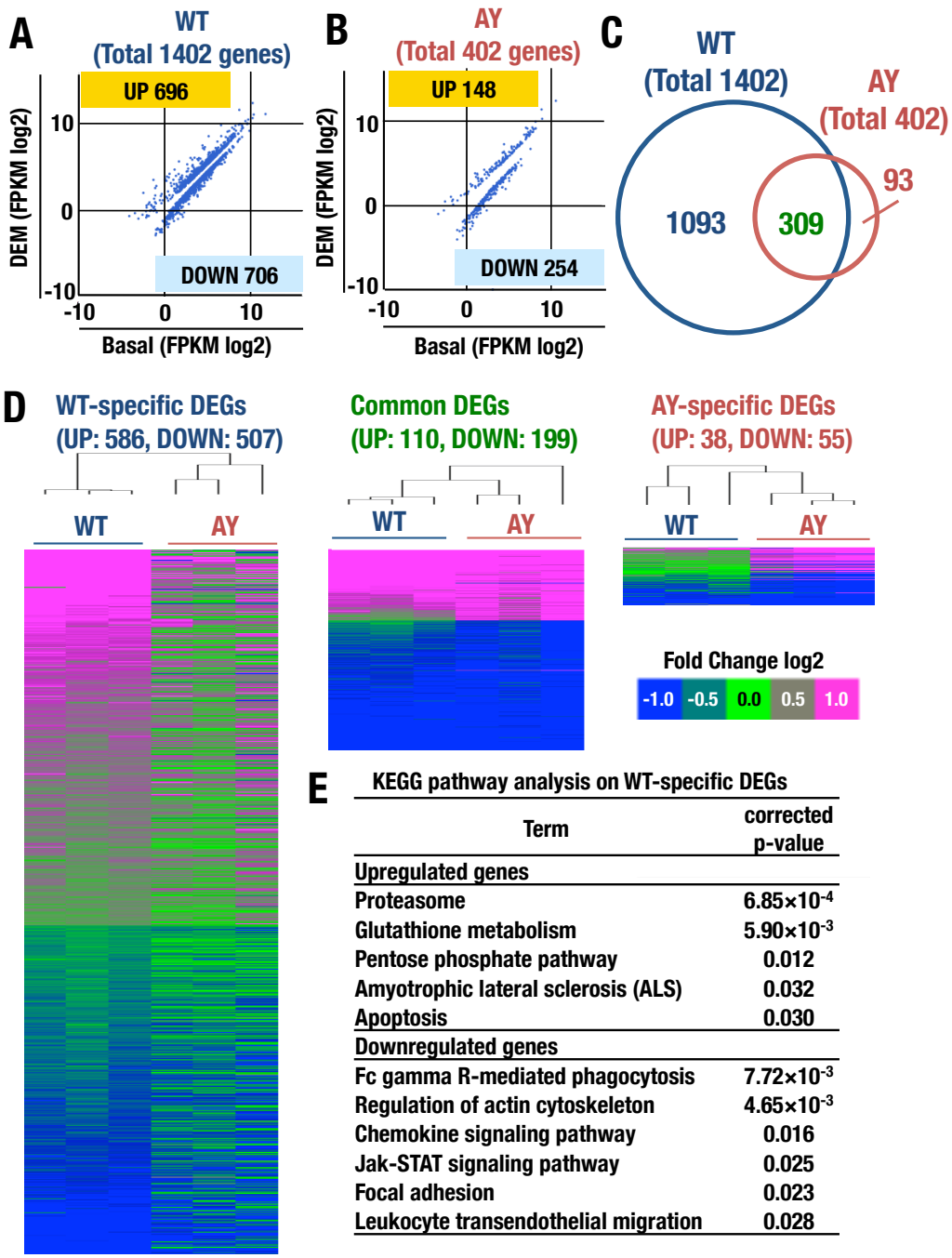

A
Nrf2
(Total 1062)
Nrf2AY
(Total 1304)
669
393
911
WT-
specific
Common
AY-
specific
B
WT-specific
Common
AY-specific
WT
AY
ChIP
Input
30
0
C
WT binding
sites
AY binding
sites
***
Fold Enrichment
WT-sp.
Comm.
AY-sp.
D
WT-specific
Common
AY-specific
bits
0 1 2 3 4 5 6 7 8 9 10
E
Position 0
Frequency (%)
G
A
C
T
F
TMANNRTGACTCAGCA
-5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Frequency (%)
TAA
TCA

Figure 3



E KEGG pathway analysis on WT-specific DEGs

Term	corrected p-value
Upregulated genes	
Proteasome	6.85×10^{-4}
Glutathione metabolism	5.90×10^{-3}
Pentose phosphate pathway	0.012
Amyotrophic lateral sclerosis (ALS)	0.032
Apoptosis	0.030
Downregulated genes	
Fc gamma R-mediated phagocytosis	7.72×10^{-3}
Regulation of actin cytoskeleton	4.65×10^{-3}
Chemokine signaling pathway	0.016
Jak-STAT signaling pathway	0.025
Focal adhesion	0.023
Leukocyte transendothelial migration	0.028

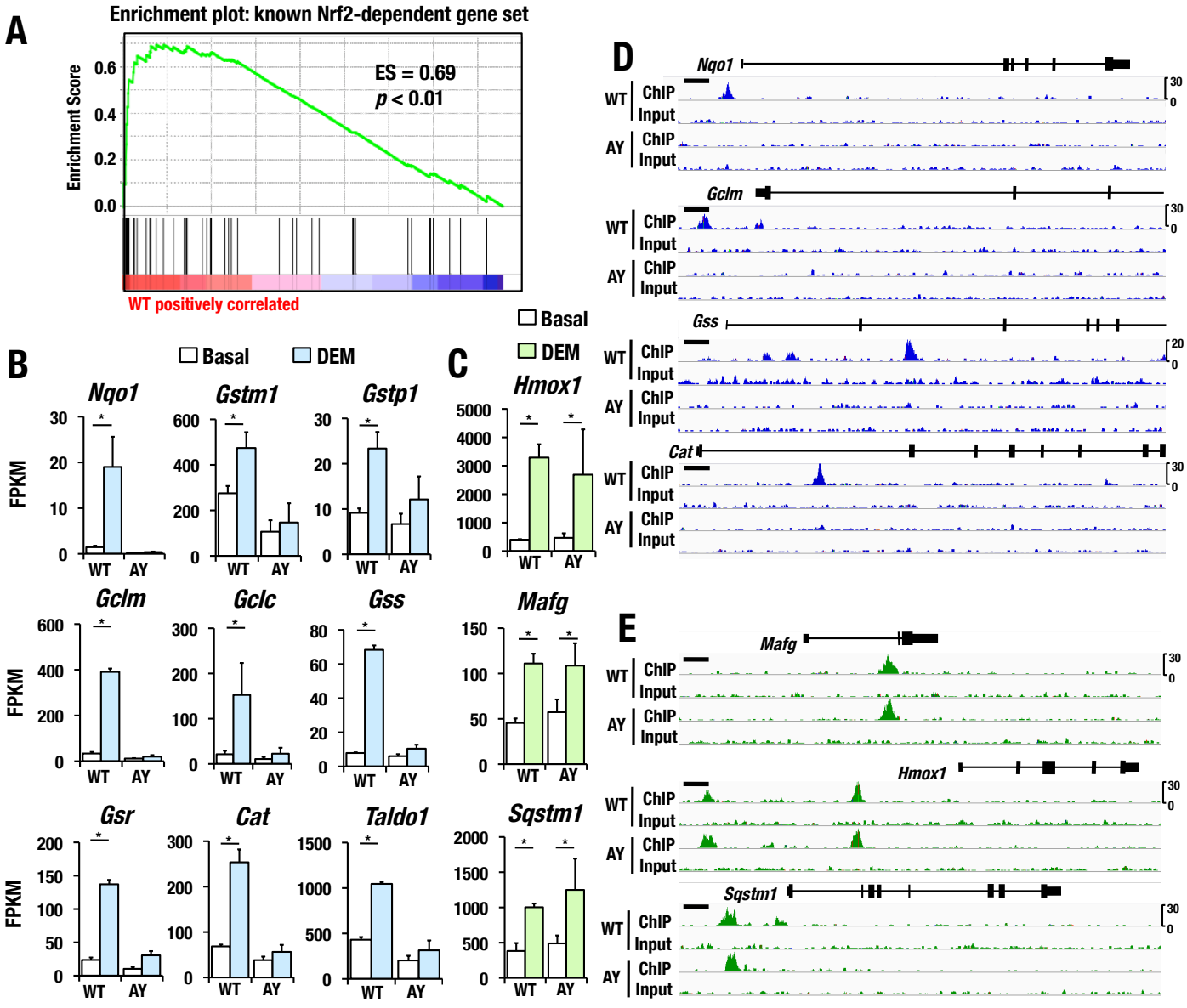

A
Enrichment plot: known Nrf2-dependent gene set
Enrichment Score
0.6
0.4
0.2
0.0
ES = 0.69
p < 0.01
WT positively correlated
B
Basal
DEM
Nqo1
Gstm1
Gstp1
Gclm
Gclc
Gss
Gsr
Cat
Taldo1
FPKM
WT
AY
C
Basal
DEM
Hmox1
Mafg
Sqstm1
D
Nqo1
Gclm
Gss
Cat
WT
AY
ChIP
Input
E
Mafg
Hmox1
Sqstm1

Figure 5

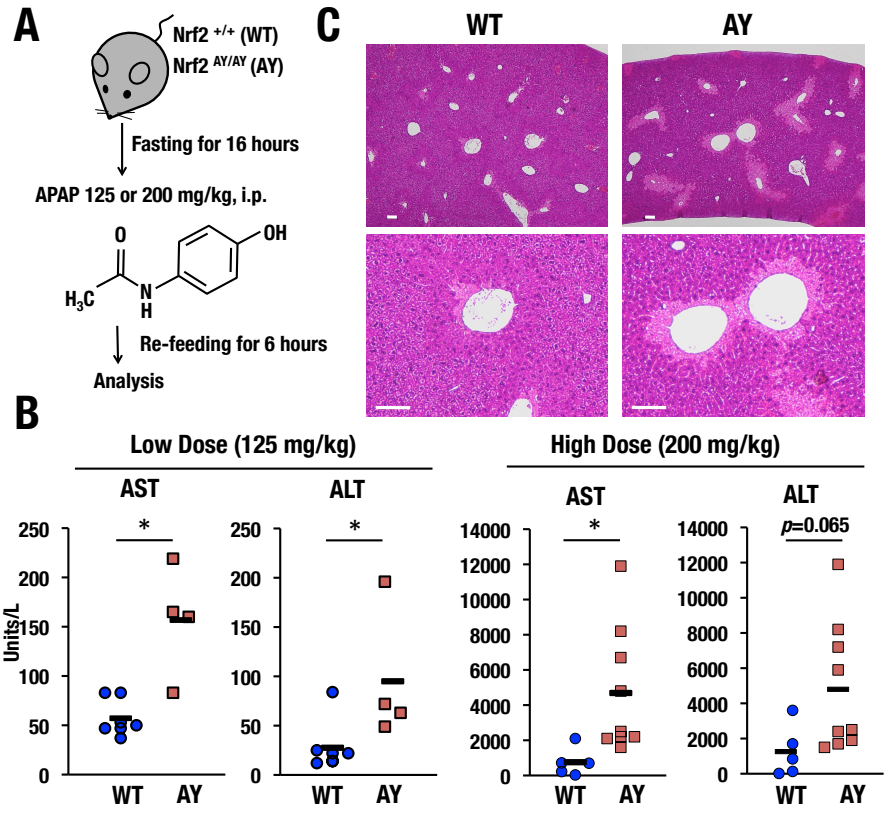

A
Nrf2 +/+ (WT)
Nrf2 AY/AY (AY)
Fasting for 16 hours
APAP 125 or 200 mg/kg, i.p.
O
OH
H3C
N
H
Re-feeding for 6 hours
Analysis
C
WT
AY
B
Low Dose (125 mg/kg)
High Dose (200 mg/kg)
AST
ALT
AST
ALT
*
*
*
p=0.065
Units/L
250
200
150
100
50
0
14000
12000
10000
8000
6000
4000
2000
0
WT
AY

Figure 6

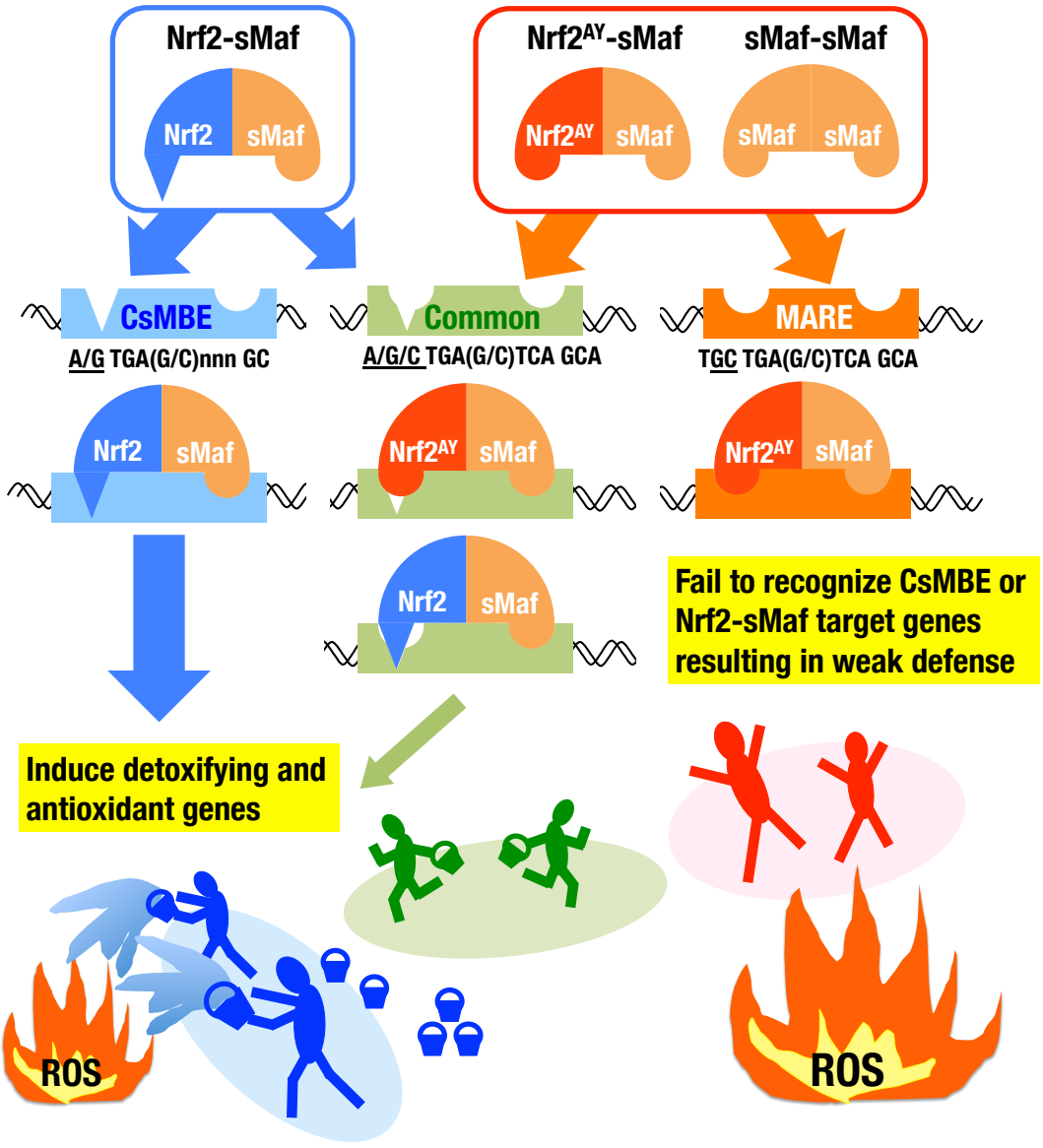

Nrf2-sMaf
Nrf2
sMaf
Nrf2AY-sMaf
sMaf-sMaf
Nrf2AY
sMaf
sMaf
sMaf
CsMBE
A/G TGA(G/C)nnn GC
Common
A/G/C TGA(G/C)TCA GCA
MARE
TGC TGA(G/C)TCA GCA
Nrf2
sMaf
Nrf2AY
sMaf
Nrf2AY
sMaf
Nrf2
sMaf
Fail to recognize CsMBE or Nrf2-sMaf target genes resulting in weak defense
Induce detoxifying and antioxidant genes
ROS
ROS

Figure 7