

## Review Article

# Genomic Structure and Variation of Nuclear Factor (Erythroid-Derived 2)-Like 2

**Hye-Youn Cho**

Laboratory of Respiratory Biology, National Institute of Environmental Health Sciences, National Institutes of Health, 111 TW Alexander Dr., Building 101, MD D-201, Research Triangle Park, NC 27709, USA

Correspondence should be addressed to Hye-Youn Cho; cho2@niehs.nih.gov

Received 11 February 2013; Accepted 22 April 2013

Academic Editor: Mi-Kyoung Kwak

Copyright © 2013 Hye-Youn Cho. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

High-density mapping of mammalian genomes has enabled a wide range of genetic investigations including the mapping of polygenic traits, determination of quantitative trait loci, and phylogenetic comparison. Genome sequencing analysis of inbred mouse strains has identified high-density single nucleotide polymorphisms (SNPs) for investigation of complex traits, which has become a useful tool for biomedical research of human disease to alleviate ethical and practical problems of experimentation in humans. Nuclear factor (erythroid-derived 2)-like 2 (*NRF2*) encodes a key host defense transcription factor. This review describes genetic characteristics of human *NRF2* and its homologs in other vertebrate species. *NRF2* is evolutionally conserved and shares sequence homology among species. Compilation of publically available SNPs and other genetic mutations shows that human *NRF2* is highly polymorphic with a mutagenic frequency of 1 per every 72 bp. Functional at-risk alleles and haplotypes have been demonstrated in various human disorders. In addition, other pathogenic alterations including somatic mutations and misregulated epigenetic processes in *NRF2* have led to oncogenic cell survival. Comprehensive information from the current review addresses association of *NRF2* variation and disease phenotypes and supports the new insights into therapeutic strategies.

## 1. Overview

The gene nuclear factor (erythroid-derived 2)-like 2 (*NFE2L2*) or more commonly the used synonym nuclear factor erythroid 2- (NF-E2-) related factor 2 (*NRF2*) and its mouse homolog (*Nfe2l2*, *Nrf2*) encode a ubiquitous transcription factor belonging to the basic leucine zipper (bZIP) protein family [1, 2]. *NRF2* modulates downstream genes by binding to their *cis*-regulatory module antioxidant response elements (AREs). *NRF2* targets include ARE-bearing effector genes such as reactive oxygen species (ROS) scavenging enzymes (e.g., superoxide dismutases, SODs), phase-2 defense enzymes (e.g., glutathione-S-transferase, GST; heme oxygenase-1, HO-1), drug efflux pumps (e.g., multidrug resistance proteins, MRPs), and various interacting and indirectly modulated proteins [3–6]. The *NRF2*-ARE pathway has emerged in mechanisms of human diseases in which oxidative stress is implicated. Importantly,

three lines of gene-targeted (knockout) mice were generated by Drs. M. Yamamoto (*Nfe2l2<sup>tm1Mym</sup>*), Y. W. Kan (*Nfe2l2<sup>tm1Ywk</sup>*), and P. A. Ney (*Nfe2l2<sup>tm1Ney</sup>*) [7–9], and 124 gene-trapped or gene-targeted cell lines have been established ([http://www.informatics.jax.org/searches/allele\\_report.cgi?markerID=MGI:108420](http://www.informatics.jax.org/searches/allele_report.cgi?markerID=MGI:108420)). During the last decade or more, wide application of the knockout mice to human disease models has led to new insights into disease pathogenesis and therapeutic potential (Figure 1).

Kelch-like ECH-activating protein 1 (KEAP1 for humans, Keap1 for mice, or iNrf2 for rats) is a cytoplasmic suppressor of *NRF2* and is critical in *NRF2* homeostasis and activity [10]. Substantial efforts have led to the discovery of the molecular mechanisms of KEAP1-mediated *NRF2* regulation. In unstressed conditions, the *NRF2*-bound KEAP1 homodimer is complexed to a ubiquitin ligase (Cullin 3-based E3 ligase), which polyubiquitinates *NRF2* for proteasomal degradation and maintains *NRF2* homeostasis (20 min half-life of cellular

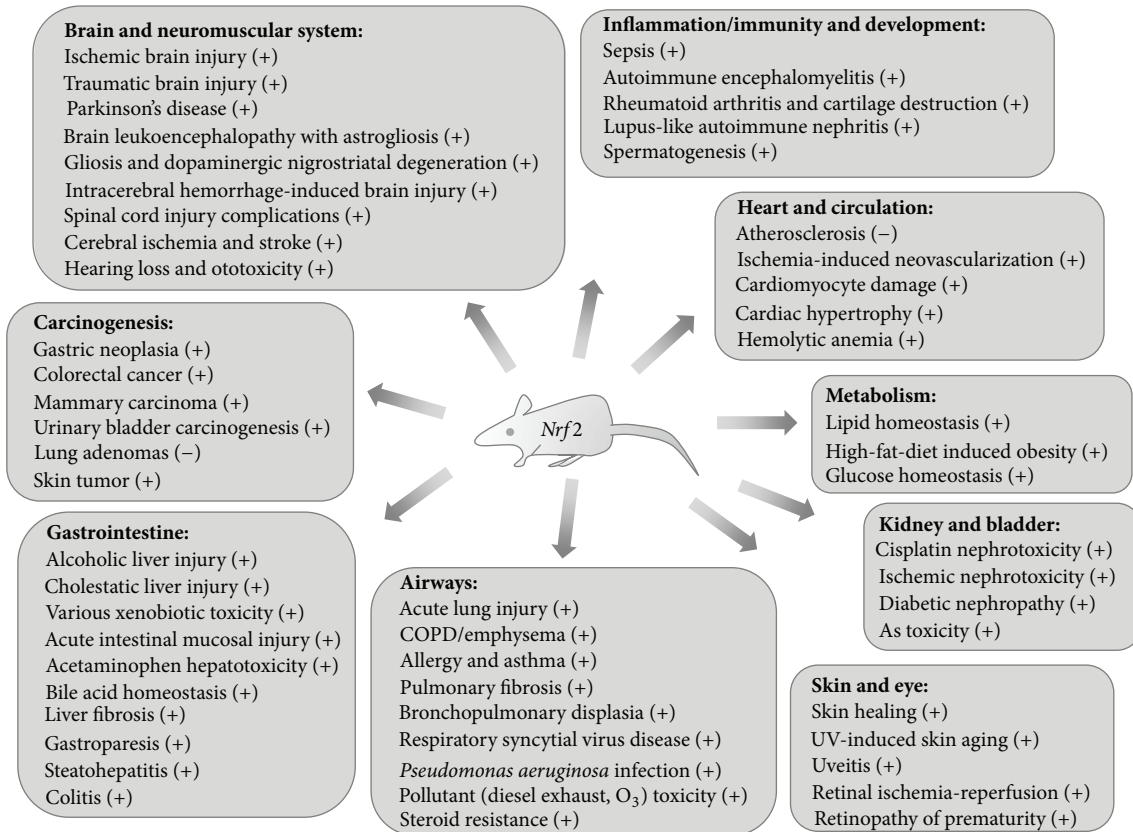


FIGURE 1: Role of Nrf2 in human pathogenesis learned from model studies using mice genetically deficient in *Nrf2*. (+) indicates protective or beneficial effects and (-) indicates aberrant roles.

NRF2 [11]). However, modifications of KEAP1 (e.g., cysteine residues) and NRF2 (e.g., serine residues) under stressed conditions activate NRF2 by liberating it from a “hinge and latch” NRF2-KEAP1 affinity binding, allowing its nuclear translocation [12].

In the current review, I address genetic aspects of human *NRF2* and its homologs in other vertebrate species. Sequence variations in human *NRF2* and murine *Nrf2* including single-nucleotide polymorphisms (SNPs) were collected from public databases and compiled. Mutations that have been associated with disease risks are defined. Nongenetic variations including somatic mutations and epigenetic modifications are also described. Although the current review does not deal with mutations in other species, recent characterization of *nrf2* mutant zebrafish which were hypersensitive to environmental toxicants [13] also provides a useful investigational tool.

## 2. Sequence of NF-E2-Related Factor 2 and Cross-Species Homology

Homology scores of gene (coding DNA sequences, cds) and protein across 10 species were compared with human *NRF2*. The highest sequence homology (98%-99%) was with chimpanzee and rhesus monkey while the lowest similarity was found with zebrafish (Table 1). While there is approximately

about 83% homology in cds and protein sequences of humans and rodents, 5'-untranslated regions (5'-UTR, UTR-5) of these strains extend differentially (114 bp in human, 233 bp in mouse, and 82 bp in rat), and the human 5'-UTR does not share significant sequence homology with either rat or mouse (the rat is 94% homologous with the 3' portion of mouse 5'-UTR).

Human *NRF2* is located in the cytogenetic band 2q31.2 of chromosome 2 spanning 178,095,031–178,129,859 bp as a complementary sequence (gene ID: 4780, Table 1). Murine *Nrf2* maps as a complementary sequence to chromosome 2 C3 (44.75 centimorgan) and spans 75,675,519–75,704,641 (gene ID: 18024, Table 1). The complete cds of *NRF2* is 2,859 bp, and there are 14 transcript variants reported ([http://useast.ensembl.org/Homo\\_sapiens](http://useast.ensembl.org/Homo_sapiens)). Mouse *Nrf2* mRNA spans 2,469 bp, and another variant has been reported ([http://useast.ensembl.org/Mus\\_musculus](http://useast.ensembl.org/Mus_musculus)). The human *NRF2* protein (ID: NP\_006155) contains 605 amino acid (aa) residues with molecular weight of 67.7 kDa (isoform 1), and total 12 isoforms are published (the National Center for Biotechnology Information, NCBI, <http://www.ncbi.nlm.nih.gov/refseq/rsg/>; e!Ensemble, <http://useast.ensembl.org/index.html>; UniProt Consortium, <http://www.uniprot.org/uniprot/Q16236>, <http://www.uniprot.org/uniprot/Q60795>). Mouse *Nrf2* protein (ID: NP\_035032) comprises 597 aa at 66.8 kDa (Table 1). Structurally, there are 6 NRF2-ECH

TABLE 1: Gene orthology of NF-E2-related factor 2 across the species.

Species	Human <i>Homo sapiens</i>	Cat <i>Felis catus</i>	Chimpanzee <i>Pan troglodytes</i>	Dog <i>Canis lupus familiaris</i>	Cattle <i>Bos taurus</i>
Official full name					
Author name		nuclear factor erythroid 2-related factor 2, NF-E2-related factor 2, NFE2-related factor 2	nuclear factor (erythroid-derived 2)-like 2		
Author name	HEBP1	—	—	—	—
Gene synonyms	NFE2L2, NRF2	NFE2L2, NRF2	NFE2L2, NRF2	NFE2L2, NRF2	NFE2L2, NRF2
Chromosome map	2q31.2	Cl	2B	36	2
Genome sequence (NCBI Build)	NC_000002.11 (GRCCh37p10)	NC_108730.1 (Felis.catus-6.2)	NC_006470.3 (Pan.troglodytes-2.1.4)	NC_006618.3 (CanFam3.1)	AC_000159.1 (Bos.taurus.UMD_3.1)
Primary source	HGNC: 7782	—	—	—	BOS_1716
Gene ID	4780	101098812	742622	478813	497024
mapping (orientation *)	178095031-178129859 (-)	165356974-165392728 (-)	181721941-181756837 (-)	20989206-21021893 (-)	19659540-19692048
Ensemble gene ID (mapping)	ENSG00000116044 (178092323-178257425)	—	ENSPTRG00000012677 (181721949-181756285)	ENSCAFG000000013506 (20987569-21022230)	ENSBTAG000000019255 (19659540-19692048)
RefSeq ID, mRNA size	NM_006164.4 (GI: 372620346), 2859 bp	XM_003990893.1 (GI: 410968909), predicted, 2394 bp	XM_00145876.2 (GI: 332814815), predicted, 2850 bp	XM_535975.3 (GI: 345797173), predicted, 2570 bp	NM_001011678.2 (GI: 147904941), 2409 bp
Protein ID (aa/MW <sup>s</sup> )	NP_006155.2 (605/67696)	XP_003990942.1 (606/67797)	XP_001145876.2 (605/67686)	XP_535975.1 (601/67087)	NP_001011678.2 (607/67813)
UniProt ID	Q16236	—	A2T6Y9	—	Q5NUA6.2
Variants <sup>#</sup>	14 transcripts, 12 isoforms	3 transcripts, 3 isoforms	5 transcripts, 5 isoforms	2 transcripts, 2 isoforms	—
Homology% gene/protein	100/100	89.3/91.4	99.9/99.9	89.3/88.8	90.5/89.1

TABLE 1: Continued.

Species	Rhesus monkey <i>Macaca mulatta</i>	Chicken <i>Gallus gallus</i>	Zebrafish <i>Danio rerio</i>	Mouse <i>Mus musculus</i>	Rat <i>Rattus norvegicus</i>
Official full name		nuclear factor (erythroid-derived 2)-like 2		nuclear factor, erythroid-derived 2, like 2	
Author name					
Author name	HEBPI				
Gene synonyms	<i>NFE2L2, NRF2, Mmu.966</i>	<i>NFE2L2, NRF2</i>	<i>nfe2l2, Nrf2, Nrf2a,</i> wu:fc15g09, wu:ff67e03	<i>Nfe2l2, Nrf2, Al194320</i>	<i>Nfe2l2, Nrf2</i>
Chromosome map	12	7	9	2C3 (44.75 cM)	3q23
Genome sequence (NCBI Build)	NC_007869.1	NC_006094.3	NC_007120.5 (Zv9)	NC_000068.7 (GRCm38.pl)	NC_005102.3 (Rnor.5.0)
Primary Source	—	CGNC: 49604	ZFIN: ZDB-GENE-030723-2	MGI: 108420	RGD: 620360
Gene ID mapping (orientation*)	<b>707606</b>	<b>396014</b>	<b>360149</b>	<b>18024</b>	<b>83619</b>
Ensemble gene ID (mapping)	40848940-40884600 (-)	15304546-15320356	1643746-1670288	75675519-75704641 (-)	69041647-69069070 (-)
RefSeq ID, mRNA size	ENSMUTG00000001861 (40848948-40884174)	ENSGALG00000009240 (16897956-16922637)	ENSARG00000042824 (1643631-1672715)	ENSMUSG00000015839 (75675513-75704641)	ENSRNOG00000001548 (58366693-58394118)
Protein ID (aa/MW <sup>§</sup> )	NM_001257607.1 (GI: 383872375), 2335 bp	NM_205117.1 (GI: 45384113), 2555 bp	NM_182889.1 (GI: 33504557), 2149 bp	NM_010902.3 (GI: 76573877), 2469 bp	NM_031789.2 (GI: 402692377), 2305 bp
UniProt ID Variants <sup>#</sup>	NP_001244536.1 (606/67703) F7GPD8	NP_990448.1 (582/65160) Q90834	NP_878309.1 (586/65757) Q8JIMI	NP_035032.1 (597/66770) Q60795	NP_113977.1 (597/66825) O54968
Homology <sup>%</sup> gene/protein	2 transcripts, 2 isoforms 98.8/99.0	1 transcript, 1 isoform 72.6/67.4	1 transcript, 1 isoform 54.6/49.1	1 transcript 83.4/82.5	1 transcript, 1 isoform 83.8/83.2

Refer to <http://www.ncbi.nlm.nih.gov/homologene/2412> and reference [14] for cross-species homology. Details of homology scores and sequences can be also obtained by blast against human sequence (<http://blast.ncbi.nlm.nih.gov>).<sup>#</sup> Further information on transcript variants and protein isoforms are viewable by gene ID at NCBI (<http://www.ncbi.nlm.nih.gov/refseq/rsg>) and eEnsemble (<http://useast.ensemble.org/index.html>) or by protein ID at UniProt (<http://www.uniprot.org>).<sup>§</sup>MW: predicted molecular weight from NCBI (MW in UniProt varies slightly). RefSeq: reference sequence, aa: amino acids, and \* gene in complement orientation.

TABLE 2: Protein domains of NF-E2-related factor 2.

Domains	Amino acid positions*		Predicted functions
	Human (605 aa)	Mouse (597 aa)	
Neh2	16–89 DLG motif 17–32, ETGF motif 77–82	16–89	KEAP1 repression through DLG/ETGF motif-DC motif binding, fast, redox-sensitive proteasomal degradation.
Neh4	111–134	111–134	Translocation and transactivation. Phosphorylation or CBP binding.
Neh5	182–209	172–201	
Neh6	337–394 (or 338–388)	328–385	Degron motif-associated constitutive turnover, slow, redox-insensitive (Keap1-independent).
Neh1	435–568 basic motif 503–518 leucine zipper 525–539	427–560 basic motif 494–509 leucine zipper 517–531	Dimerization for nuclear translocation, DNA binding through basic motif-leucine zipper.
Neh3	569–605	561–597	CHD6 binding, stability, or transactivation.

\*Varies slightly among publications.

homology (Neh) domains configuring the protein sequence of either species (Table 2), and the potential functions of each region, particularly the highly conserved KEAP1-binding Neh2 and DNA- (ARE-) binding Neh1 domains, have been intensively investigated [12, 14, 15].

### 3. Genetic Variation of NF-E2-Related Factor 2 in Human and Mouse

*3.1. Evolution, Genome Sequence, and Polymorphism Discovery in Human and Mouse.* While rare and monogenic Mendelian diseases are inheritable mutations in a single gene [16], many common diseases are complex traits, and the disease phenotypes are affected by variants in multiple genetic loci. Recent advancements in high-throughput technology have enabled sequencing of entire mammalian genomes [17–19], and information on DNA sequence and variation has facilitated the study of complex traits of human disorders. Genome-wide association studies (GWAS) examine whether SNPs are associated with important disease traits and ascertains “at-risk” genotypes that are significantly more prevalent in the affected group than in the nonaffected group. The HapMap Project (<http://hapmap.ncbi.nlm.nih.gov/>) has mapped combinations of alleles at specific loci (haplotypes), that is, common patterns of sequence variation in several human populations. It has supported efficient mapping of multiple loci for complex traits in GWAS. Candidate gene approaches based on findings from GWAS of similar disorders have also been useful for determining the potential genetic mechanisms of diseases.

The evolutionary divergence of human and mouse lineages occurred for roughly 75 million years, and their genome sequences have been altered by nearly one substitution (or deletion or addition) for every two nucleotides [20]. However, this slow evolution process resulted in a high degree of conservation across the two species, which allows alignment of orthologous sequences: >90% of the human and mouse genomes is partitioned into corresponding regions of conserved synteny, and at the nucleotide level approximately 40% of the human genome is aligned to the mouse [20]. Due to this

fact, biomedical studies of human genes are complemented by experimental manipulation of corresponding mouse genes, and they have aided functional understanding of genes in human health. Following the 2003 completion of the Human Genome Project of approximately 3.1 giga base pairs (Gbp), the Mouse Genome Project assembled the complete genome sequence of one strain (C57BL/6J; 2,716,965,481 bp) in 2011. Using this reference strain, whole genome sequencing data across 16 additional inbred strains were done (<http://www.sanger.ac.uk/resources/mouse/genomes/>, [21]). Discovery of high-density SNPs in the mouse genome supports evolutionary history of the strain and provides a tool to investigate models of human disease processes that cannot often be practically achieved through direct human studies.

*3.2. Genetic Mutations in Human NRF2.* Human *NRF2* codes three major isoforms of protein (Figure 2). Transcript variant 2 (NM\_001145412.2, 2746 bp) has an alternate promoter, 5'-UTR, and a downstream start codon, compared to variant 1 (NM\_006164.4, 2859 bp). It encodes an isoform 2 missing N-terminal 16 aa (NP\_001138884 or Q16236-2, 589 aa) relative to isoform 1 (NP\_006155.2 or Q16236, 605 aa). Isoform 3 (NP\_001138885.1 or Q16236-3, 582 aa) is encoded by transcript variant 3 (NM\_001145413.2; 2,725 bp) and lacks an internal segment relative to isoform 2 due to an alternate in-frame splice site in the 3' coding region. In public databases, more than 583 sequence mutations are reported in *NRF2* (34,827 bp) and 7,000 bp upstream (Table S1; data acquired as of December, 2012) (See Table S1 in Supplementary Material available on line at <http://dx.doi.org/10.1155/2013/286524>). *NRF2* locates on 178,130,354–178,129,304 bp of GRCh37.p10 Primary Assembly, and Figure 2 shows sequences of proximal promoter (–1 to –500), partial mRNA variant 1 including 5'-UTR (exon 1, up to TSS), and protein isoform 1 (NP\_006155, encoded by variant 1 NM\_006164.4: 556–2,373 bp). Based on the current assembly and sequence update, previous promoter positions –686 [22]/–653 [23] are identified as –214; –684 [22]/–651 [23] as –212; and –650 [22]/–617 [23] as –178. Overall frequency of *NRF2* SNPs and other mutations is about 1 per 72 bp. The genetic mutations include 37 in



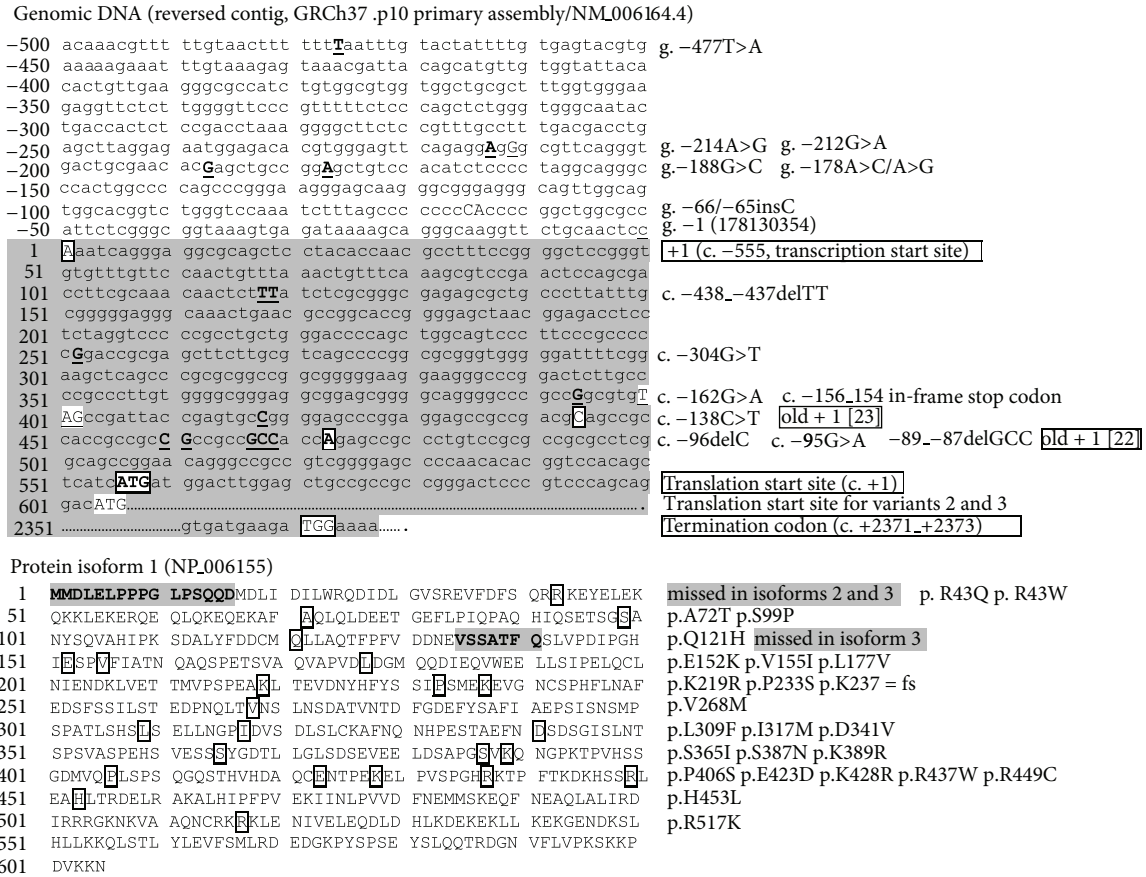


FIGURE 2: DNA (variant 1 partial promoter and exons 1 and 5) and protein (isoform 1) sequence of Human *NRF2*. SNPs and amino acid residues for nonsynonymous SNPs are marked.

the 5' flanking promoter and 59 in exons (Tables 3 and S1). Among exon SNPs, 26 are nonsynonymous (Cns) mutations. A triplet repeat variation (rs143406266; GCC<sub>4</sub> versus GCC<sub>5</sub>, previously published as -20/-6) in the 5'-UTR was uniquely identified in Asian populations [22, 24].

### 3.3. SNPs, Haplotypes, and Association with Disease Risk.

The use of gene knockout mice in model systems has provided potential insights into the role of *NRF2* in the pathogenesis of various human disorders (see Figure 1). Recent epidemiological and association studies have revealed significant associations of *NRF2* sequence variations with disease risks, which further supports *NRF2* as a susceptibility gene. Most of the phenotype-associated variants are in the promoter region and presumed to be involved in *NRF2* gene regulation. Table 4 summarizes *NRF2* SNP and/or haplotype alleles that have been associated with oxidant-related disease risks. Interestingly, there is no evidence for exon SNPs as at-risk alleles. For convenience and consistency, intronic and 3' distal SNP alleles are presented as chromosome contig (HGVS) alleles while promoter and exon SNPs are presented as reversed contig alleles throughout the text.

*Pulmonary Diseases.* *NRF2* SNPs in the promoter and intron 1 sequences have been investigated for their potential associations with risk of pulmonary critical disorders including acute lung injury (ALI), cigarette smoke-induced chronic obstructive pulmonary disease (COPD), and asthma. A heterozygous C/A SNP at -178 position (rs6721961T>C or T>G, previously -617 or -650) significantly increased the risk for developing ALI following major trauma in European and African-American populations (odds ratio, OR 6.44; 95% confidence interval, CI 1.34–30.8; allelic frequency = 11.9% at 21/180) [23]. Promoter activity of the A allele (A/C or A/A) determined *in vivo* and *in vitro* was significantly lower than C/C allele at that locus (-178 in an ARE-like motif) indicating that it is a functional SNP for autoregulation [23]. The -178G/G were also nominally associated with ALI-related 28-day mortality following systemic inflammatory response syndrome [37]. In a Japanese cohort, SNP haplotype (rs2001350T/rs6726395A/rs1962142A/rs2364722A/rs6721961T) co-taining the -178A/A homozygote was associated with an annual decline of rapid forced expiratory volume in one second (FEV<sub>1</sub>) in relation to cigarette-smoking status [34]. In addition, a promoter and 5'-UTR SNP haplotype consisting of -214 G allele (52%, rs35652124, previously -686/-653),

TABLE 3: Genetic mutations in promoter and exons of human *NFE2L2*.

ID	Map on chromosome 2* (HGVS name)	Position in <i>Nfe2l2</i> * (HGVS name)	Regions (position from mRNA)*	Variation class and consequences (HGVS name) <sup>†</sup>	Minor allele frequency (MAF)/ MA counts* (cohort size)	MAF sources
rs16865105	g.178136629A>C	c.-555-6770T>G	5' Flanking (-6770)	SNP	C = 0.1928/421	1000 Genomes
rs7557529	g.178135097C>T	c.-555-5238G>A	5' Flanking (-5238)	SNP	na	1000 Genomes
rs6750320	g.178131796C>T	c.-555-1937G>A	5' Flanking (-1937)	SNP	T = 0.0302/66	1000 Genomes
rs181162518	g.178131774T>C	c.-555-1915A>G	5' Flanking (-1915)	SNP	C = 0.0005/1	1000 Genomes
rs190044775	g.178131746C>T	c.-555-1887G>A	5' Flanking (-1887)	SNP	T = 0.0005/1	1000 Genomes
rs185117338	g.178131704A>G	c.-555-1845T>C	5' Flanking (-1845)	SNP	G = 0.0005/1	1000 Genomes
rs149947189	g.178131697C>T	c.-555-1838G>A	5' Flanking (-1838)	SNP	T = 0.0009/2	1000 Genomes
rs139771244	g.178131625A>G	c.-555-1766T>C	5' Flanking (-1766)	SNP	G = 0.0009/2	1000 Genomes
rs6747203	g.178131604C>G	c.-555-1745G>C	5' Flanking (-1745)	SNP	G = 0.006/13	1000 Genomes
rs193101749	g.178131504T>C	c.-555-1645A>G	5' Flanking (-1645)	SNP	C = 0.0046/10	1000 Genomes
rs190630762	g.178131495G>C	c.-555-1636C>G	5' Flanking (-1636)	SNP	C = 0.0009/2	1000 Genomes
rs183764402	g.178131366C>A	c.-555-1507G>T	5' Flanking (-1507)	SNP	A = 0.0005/1	1000 Genomes
rs191222964	g.178131211G>A	c.-555-1352C>T	5' Flanking (-1352)	SNP	A = 0.0009/2	1000 Genomes
rs187137522	g.178131165T>G	c.-555-1306A>C	5' Flanking (-1306)	SNP	G = 0.0009/2	1000 Genomes
rs182620359	g.178131158A>G	c.-555-1299T>C	5' Flanking (-1299)	SNP	G = 0.0005/1	1000 Genomes
rs4893819 (rs61433302)	g.178131134C>T	c.-555-1275G>A	5' Flanking (-1275)	SNP	C = 0.4263/931	1000 Genomes
rs191547130	g.178131017C>T	c.-555-1158G>A	5' Flanking (-1158)	SNP	T = 0.0005/1	1000 Genomes
rs188422217	g.178131003A>G	c.-555-1144T>C	5' Flanking (-1144)	SNP	G = 0.0014/3	1000 Genomes
rs143047764	g.178130865A>G	c.-555-1006T>C	5' Flanking (-1006)	SNP	G = 0.0069/15	1000 Genomes
rs74432849	g.178130766C>A	c.-555-907G>T	5' Flanking (-907)	SNP	na	
rs11679252	g.178130691C>G	c.-555-832G>C	5' Flanking (-832)	SNP	na	
rs12993217	g.178130516A>G	c.-555-657T>C	5' Flanking (-657)	SNP	na	
rs115644826	g.178130442T>A	c.-555-583A>T	5' Flanking (-583)	SNP	A = 0.0151/33	1000 Genomes
rs140803524	g.178130431G>A	c.-555-572C>T	5' Flanking (-572)	SNP	A = 0.0046/10	1000 Genomes
rs77684420	g.178130427T>C	c.-555-568A>G	5' Flanking (-568)	SNP	C = 0.0339/74	1000 Genomes
rs183651094	g.178130336A>T	c.-555-477T>A	5' Flanking (-477)	SNP	C = 0.190 (84)	[24]
rs35652124 <sup>§</sup> (rs76952243)	g.178130073T>C	c.-555-214A>G	5' Flanking (-214)	SNP	T = 0.0018/4	1000 Genomes
rs6706649 <sup>§</sup>	g.178130071C>T	c.-555-212G>A	5' Flanking (-212)	SNP	C = 0.3512/767	1000 Genomes
rs150648896	g.178130047C>G	c.-555-188G>C	5' Flanking (-188)	SNP	T = 0.429 (84)	[24]
rs6721961 <sup>§</sup> (rs117801448)	g.178130037T>C, T>G	c.-555-178A>C, A>G	5' Flanking (-178)	SNP	T = 0.413/181 <sup>‡</sup>	[22]
					C = 0.338/27 <sup>‡</sup>	[23]
					C = 0.351/769	[25]
					T = 0.078/170	1000 Genomes
					T = 0.048 (84)	[24]
					T = 0.048/21 <sup>‡</sup>	[22]
					T = 0.075/3 <sup>‡</sup>	[23]
					G = 0.0023/5	1000 Genomes
					G = 0.006 (84)	[24]
					T = 0.150/328	1000 Genomes
					T = 0.283/124 <sup>‡</sup>	[22]
					T = 0.313/25 <sup>‡</sup>	[23]
					T = 0.321 (84)	[24]

TABLE 3: Continued.

ID	Map on chromosome 2* (HGVS name)	Position in Nfe2l2# (HGVS name)	Regions (position from mRNA)#	Variation class and consequences (HGVS name)†	Minor allele frequency (MAF)/ MA counts‡ (cohort size)	MAF sources
rs201345604	g.178129924_178129925insG	c.-555-66_-555-65insC	5' Flanking (-66/-65)	Insertion	G = 0.0179/39	1000 Genomes
rs200432479	g.178129741_178129742delAA	c.-438_-437delTT	exon 1/UTR-5' (118-119)	Deletion	- = 0.0037/8 - = 0.006 (84)	1000 Genomes [24]
rs75485459	g.178129608C>A	c.-304G>T	Exon 1/UTR-5' (252)	SNP	na	
rs192086766	g.178129466C>T	c.-162G>A	Exon 1/UTR-5' (394)	SNP	T = 0.022/48	1000 Genomes
—	g.178129442G>A	c.-138C>T	Exon 1/UTR-5' (418)	SNP	A = 0.012 (84)	[24]
rs71668246	g.178129400delG	c.-96delC	Exon 1/UTR-5' (460)	Deletion	na	
rs187291840	g.178129399C>T	c.-95G>A	Exon 1/UTR-5' (461)	SNP	T = 0.0549/120	1000 Genomes
rs143406266	g.178129391_178129393delGGC	c.-89_-87delGCC	Exon 1/UTR-5' (467-469)	Deletion	- = 0.644/282* - = 0.589 (84)	[22] [24]
rs182428269	g.178098918G>A	c.127C>T	Exon 2 (682)	Cns (p.Arg43Trp)	A = 0.0005/1	1000 Genomes
rs35248500	g.178098917C>T	c.128G>A	Exon 2 (683)	Cns (p.Arg43Gln)	T = 0.006/13	1000 Genomes
rs1135118	g.178098831C>T	c.214G>A	Exon 2 (769)	Cns (p.Ala72Thr)	T*	[23]
rs199691660	g.178098829A>T	c.216T>A	Exon 2 (771)	Cs (p.Ala72=)	na	
—	g.178098769G>A	c.276C>T	Exon 2 (831)	Cs (p.Ile 92=)	T*	[23]
rs5031039	g.178098750A>G	c.295T>C	Exon 2 (850)	Cns (p.Ser99Pro)	G = 0*	[23]
rs200239262	g.178098017C>G	c.363G>C	Exon 3 (918)	Cns (p.Gln121His)	na	
rs183034165	g.178098008T>C	c.372G>A	Exon 3 (927)	Cs (p.Ala124=)	T = 0.0009/2 T = 0.006 (84)	1000 Genomes [24]
rs199970826	g.178097996C>T	c.384G>A	Exon 3 (939)	Cs (p.Pro128=)	T = 0.0005/1	1000 Genomes
rs201992337	g.178097260C>T	c.454G>A	Exon 4 (1009)	Cns (p.Glu152Lys)	na	
rs201589693	g.178097251C>T	c.463G>A	Exon 4 (1018)	Cns (p.Val151Ile)	T = 0.0005/1	1000 Genomes
rs35577826	g.178097185A>C	c.529T>G	Exon 4 (1084)	Cns (p.Leu177Val)	C = 0.0014/3 A < 0.005*	1000 Genomes [23]
rs181513314	g.178096710C>T	c.621G>A	Exon 5 (1176)	Cn (p.Leu207=)	T = 0.0005/1	1000 Genomes
rs60132461	g.178096675T>C	c.656A>G	Exon 5 (1211)	Cns (p.Lys219Arg)	C = 0.0018/4	1000 Genomes
rs139187151	g.178096634G>A	c.697C>T	Exon 5 (1252)	Cns (p.Pro233Ser)	A = 0.0005/1 A = 0.012* (84)	1000 Genomes [24]
rs35557421	g.178096620delT	c.711delA	Exon 5 (1266)	Frame shift/deletion (p.Lys237 = fs)	na	
rs34154613	g.178096529C>T	c.802G>A	Exon 5 (1357)	Cns (p.Val268Met)	T = 0.0018/4	1000 Genomes
rs141363120	g.178096406G>A	c.925C>T	Exon 5 (1480)	Cns (p.Leu309Phe)	A = 0.0037/8	1000 Genomes
rs201661476	g.178096380A>C	c.951T>G	Exon 5 (1506)	Cns (p.Ile317Met)	na	



TABLE 3: Continued.

ID	Map on chromosome 2* (HGVS name)	Position in Nfe2l2# (HGVS name)	Regions (position from mRNA)*	Variation class and consequences (HGVS name)†	Minor allele frequency (MAF)/ MA counts‡ (cohort size)	MAF sources
rs199673454	g.178096309T>A	c.1022A>T	Exon 5 (1577)	Cns (p.Asp341Val)	A = 0.0005/1	1000 Genomes
rs35007548	g.178096299G>A	c.1032C>T	Exon 5 (1587)	Cs (p.Ser344=)	A = 0.0009/2	1000 Genomes
rs200209692	g.178096287T>C	c.1044A>G	Exon 5 (1599)	Cs (p.Leu348=)	C = 0.0005/1	1000 Genomes
—	g.178096237C>A	c.1094G>T	Exon 5 (1649)	Cns (p.Ser365Ile)	A = 0.125/273 A = 0.006* (84)	PubMed [24]
rs201214197	g.178096171C>T	c.1160G>A	Exon 5 (1715)	Cns (p.Ser387Asn)	T = 0.0005/1	1000 Genomes
rs200494292	g.178096165T>C	c.1166A>G	Exon 5 (1721)	Cns (p.Lys389Arg)	na	
rs186171287	g.178096115G>A	c.1216C>T	Exon 5 (1771)	Cns (p.Pro406Ser)	A = 0.0005/1	1000 Genomes
rs182276775	g.178096062C>A	c.1269G>T	Exon 5 (1824)	Cns (p.Glu423Asp)	A = 0.0005/1	1000 Genomes
rs201560221	g.178096048T>C	c.1283A>G	Exon 5 (1838)	Cns (p.Lys428Arg)	na	
rs189238236	g.178096043A>G	c.1288T>C	Exon 5 (1843)	Cs (p.Leu430=)	G = 0.0005/1	1000 Genomes
rs184287392	g.178096022G>A	c.1309C>T	Exon 5 (1864)	Cns (p.Arg437Trp)	A = 0.0005/1	1000 Genomes
rs201871588	g.178095986G>A	c.1345C>T	Exon 5 (1900)	Cns (p.Arg449Cys)	na	
rs181294188	g.178095985T>C	c.1346G>A	Exon 5 (1901)	Cns (p.Arg449His)	T = 0.0009/2	1000 Genomes
rs201690466	g.178095973T>A	c.1358A>T	Exon 5 (1913)	Cns (p.His453Leu)	A = 0.0005/1	1000 Genomes
rs1057044 (rs52789869)	g.178095781C>T	c.1550G>A	Exon 5 (2105)	Cns (p.Arg517Lys)	T* [23]	
rs200750800	g.178095603A>G	c.1728T>C	Exon 5 (2283)	Cs (p.Tyr576=)	na	
rs200175942	g.178095567A>G	c.1764T>C	Exon 5 (2319)	Cs (p.Asp588=)	G = 0.0005/1	1000 Genomes
rs77547666	g.178095495G>C	c.*18C>G	Exon 5/UTR-3' (2391)	SNP	C = 0.0069/15	1000 Genomes
rs73031353	g.178095425T>C	c.*88A>G	Exon 5/UTR-3' (2461)	SNP	C = 0.0018/4	1000 Genomes
rs6759443	g.178095345T>C	c.*168A>G	Exon 5/UTR-3' (2541)	SNP	C = 0.0041/9	1000 Genomes
rs188674558	g.178095279C>A	c.*234G>T	Exon 5/UTR-3' (2607)	SNP	A = 0.0005/1	1000 Genomes
rs77685897	g.178095247A>G	c.*266T>C	Exon 5/UTR-3' (2639)	SNP	G = 0.0009/2	1000 Genomes
rs1057092	g.178095162T>G	c.*351A>C	Exon 5/UTR-3' (2724)	SNP	na	
rs3197704	g.178095162T>G	c.*351A>C	Exon 5/UTR-3' (2724)	SNP	na	
rs184701151	g.178095159T>C	c.*354A>G	Exon 5/UTR-3' (2727)	SNP	C = 0.0027/6	1000 Genomes
rs11543307	g.178095153A>G	c.*360T>C	Exon 5/UTR-3' (2733)	SNP	na	
rs111874043	g.178095146A>G	c.*367T>C	Exon 5/UTR-3' (2740)	SNP	na	
rs34012004	g.178095102A>C	c.*411T>G	Exon 5/UTR-3' (2784)	SNP	C = 0.071 [24]	

TABLE 3: Continued.

ID	Map on chromosome 2* (HGVS name)	Position in <i>Nfe2l2</i> * (HGVS name)	Regions (position from mRNA) <sup>#</sup>	Variation class and consequences (HGVS name) <sup>†</sup>	Minor allele frequency	
					(MAF)/ MA counts* (cohort size)	MAF sources
rs201481890	g.178095090_178095091insT	c.*422_*423insA	Exon 5/UTR-3' (2795–2796)	Insertion	na	na
rs3082500	g.178095089_178095090delTT, delTTinsT	c.*423_*424delA AinsA	Exon 5/UTR-3' (2796–2797)	Deletion, insertion	na	na
rs71792546 (rs71796710)	g.178095079delT	c.*425delA	Exon 5/UTR-3' (2798)	Deletion	na	na
rs1057106	g.178095078A>C, A>T	c.*435T>A, T>G	Exon 5/UTR-3' (2808)	SNP	na	na
rs34176791	g.178095076A>C	c.*437T>G	Exon 5/UTR-3' (2810)	SNP	C = 0.0023/5	1000 Genomes
rs35911553	g.178095045C>T	c.*468G>A	Exon 5/UTR-3' (2841)	SNP	T = 0.0069/15	1000 Genomes

Sequence variations in upstream and exons of human *NFE2* from 655 variations available in public database as of December, 2012 (583 active, some SNPs merged,  $\geq 4$  SNPs cited in PubMed). \* NCBI reference sequence NC\_000002.11 (Homo sapiens chromosome 2, GRCh37p10 primary assembly) spanning 178,095,033–178,129,859 bp (complement, 34,827 bp for exons and introns), 5' - Flanking regions start at 178,129,860 bp (-1) in transcript variant 1. HGVS: Human Genome Variation Society. <sup>#</sup> Positions in variant 1 (NM\_006164.4), 2859 bp. <sup>§</sup> SNPs cited in PubMed. Exon 1: 178,129,859–178,129,260 (600 bp, TTS = 178,129,304), exon 2: 178,098,999–178,098,733 (267 bp), exon 3: 178,098,067–178,097,978 (90 bp), exon 4: 178,097,311–178,097,120 (192 bp), and exon 5: 178,096,736–178,095,033 (1704 bp). <sup>†</sup> Protein amino acid (aa) residues in isoform 1 (NP\_006155.2, 605 aa). Cns: coding-nonsynonymous. Cs: coding-synonymous. <sup>‡</sup> heterozygosity detected. na: not available.

TABLE 4: Human *NRF2* SNPs associated with disease risk.

ID	Map on chromosome 2 (HGVS name)	Region/class	Disease association and references	Risk alleles and statistics	Ethnic group (number of case)
rs7557529	g.178135097C>T	-5238G>A	Parkinson's disease (2010 [26])	G = 0.718/252 <sup>‡</sup> in haplotype OR = 0.90 or 0.40; CI = 0.60–1.40 or 0.3–0.6	Swedish/Polish Caucasian (357)
rs2886162	g.178133165A>G	-3306T>C	Breast cancer survival (2012 [27])	T/T = 0.324 OR = 1.687; CI = 1.105–2.75	Finland KBPC (452)
rs35652124 (rs57695243)	g.178130073T>C	-214A>G (previously -653 [23] or -686 [22])	Chronic gastritis, gastric ulcer (2007 [28]) P14 methylation in gastric cancer <i>H. pylori</i> infection (2008 [29]) Gastric cancer in <i>H. pylori</i> -negative cases (2008 [30]) Ulcerative colitis (2008 [31]) Lupus with nephritis in female (2010 [25]) Parkinson's disease (2010 [26]) COPD (2010 [32])	G <sup>‡</sup> G <sup>‡</sup> in haplotype for ulcer (OR = 2.52; CI = 1.19–5.45) G in haplotype OR = 2.90; CI = 1.14–7.36 A in haplotype P = 0.022 A/G (OR = 0.45; CI = 0.22–0.93) G (OR = 2.57; CI = 1.01–6.60) G/A OR = 1.81; CI = 1.04–3.12 A = 0.884/312 <sup>‡</sup> in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.30–0.60 G = 0.52 <sup>‡</sup> hazard ratio = 0.95; CI = 0.91–0.99 (Haplotype)	Japanese (159) Japanese (209) Japanese (209) Japanese (89) Mexican mestizo (362) Swedish/Polish Caucasian (357) German (69)
rs706649	g.178130071C>T	-212G>A (previously -651 [23] or -684 [22])	Gastric ulcer (2007 [28]) P14 methylation in gastric cancer <i>H. pylori</i> infection (2008 [29]) Gastric cancer in <i>H. pylori</i> -negative cases (2008 [30]) Ulcerative colitis (2008 [31]) Maternal acetaminophen and asthma (2010 [33])	G <sup>‡</sup> in haplotype OR = 2.52; CI = 1.19–5.45 G in haplotype OR = 2.90; CI = 1.14–7.36 G in haplotype, P = 0.022 A/G (OR = 0.45; CI = 0.22–0.93) G (OR = 2.57; CI = 1.01–6.60) A = 0.232/1137 <sup>‡</sup> OR = 1.73; CI = 1.22, 2.45	Japanese (159) Japanese (209) Japanese (209) Japanese (89) UK ALSPAC (>4000 mothers, >5000 children)

TABLE 4: Continued.

ID	Map on chromosome 2 (HGVS name)	Region/class	Disease association and references	Risk alleles and statistics	Ethnic group (number of case)
rs6721961 (rs117801448)	g.178130037T>C, T>G	-178A>G, A>C (previously -617 [23] or -650 [22])	COPD (2010 [32])  Parkinson's disease (2010 [26])  Acute lung injury following trauma (2007 [23])  Annual FEV <sub>1</sub> decline (2011 [34])  Vitiligo (2008 [35])  COPD (2010 [32])  Parkinson's disease (2010 [26])  Postmenopausal venous thromboembolism (2011 [36]) Breast cancer survival and NRF2 protein expression (2012 [27]) Acute lung injury-related mortality following systemic inflammatory response syndrome (2012 [37])  Infection-induced asthma (2012 [38])	G = 0.98* in haplotype hazard ratio = 0.95; CI = 0.91–0.99 G = 0.972/343* in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.3–0.6  C/A = 0.119* OR = 6.44; CI = 1.34–30.80  A = 0.082 in haplotype  C/A, A/A OR = 1.724; CI = 1.35–2.21 G = 0.73* in haplotype hazard ratio = 0.95; CI = 0.91–0.99 C = 0.980/346* in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.30–0.60 A = 0.333/11 OR = 2.5; CI = 3.70–85.70 A/A OR = 4.656; CI = 1.35–16.06 G/G OR = 9.73; CI = 1.27–74.80 A/C OR = 0.437; CI = 0.28–0.80  GCC <sub>4</sub> = 0.53* hazard ratio = 0.95; CI = 0.91–0.99  T = 0.878/309* in haplotype OR = 0.9 or 0.4; CI = 0.6–1.4 or 0.3–0.6  CI = –63.60~–17.80, C = 0.525* Also as haplotype A = 0.082 in haplotype G = 0.401/1578* in haplotype	German (69)  Swedish/Polish Caucasian (357)  Caucasian/African-American (164) Japanese (915)  Chinese (300) German (69)  Swedish/Polish Caucasian (357)  Caucasian (161) Finland KBGP (452)  Caucasian (750) Hungarian Caucasian/Gypsy (307)  Taiwanese (69)  Swedish/Polish Caucasian (357)  Netherlands (2542) Japanese (915) Netherlands (2542)
rs143406266	g.178129391.178 129393delGGC	Exon 1 (467–469GCC)	COPD (2010 [32])	GCC <sub>4</sub> = 0.53* hazard ratio = 0.95; CI = 0.91–0.99	Taiwanese (69)
rs2886161	g.178127839T>C*	Intron 1	Parkinson's disease (2010 [26])	T = 0.878/309* in haplotype OR = 0.9 or 0.4; CI = 0.6–1.4 or 0.3–0.6	Swedish/Polish Caucasian (357)
rs2364723	g.178126546G>C	Intron 1	Basal and smoker FEV <sub>1</sub> (2009 [39])	CI = –63.60~–17.80, C = 0.525* Also as haplotype	Netherlands (2542)
rs2364722	g.178124787A>G	Intron 1	Annual FEV <sub>1</sub> decline (2011 [34])	A = 0.082 in haplotype	Japanese (915)
rs13001694	g.178118990A>G	Intron 1	Basal and smoker FEV <sub>1</sub> (2009 [39])	G = 0.401/1578* in haplotype	Netherlands (2542)

TABLE 4: Continued.

ID	Map on chromosome 2 (HGVS name)	Region/class	Disease association and references	Risk alleles and statistics	Ethnic group (number of case)
rs1806649 (rs58745895)	g.178118152C>T	Intron 1	Breast cancer [40] Basal and smoker FEV <sub>1</sub> (2009 [39]) Parkinson's disease (2010 [26]) Particulate matter and asthma/COPD admission (2012 [41])	T with NQO1/NOS3/HOI risk alleles: OR = 1.56; CI = 0.97–2.51 T = 0.263/1119* in haplotype CI = -87.30–(-1.70) T = 0.422/148* in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.30–0.60 C with low vitamin C level (OR = 3.1; CI = 1.50–6.30)	Caucasian and others (505) Netherland (2542) Swedish/Polish Caucasian (357) UK (209)
rs4243387 (rs60038464)	g.178117765C>T	Intron 1	Basal and smoker FEV <sub>1</sub> (2009 [39])	T = 0.091/425* in haplotype	Netherland (2542)
rs1962142 (rs58448508)	g.178113484A>G	Intron 1	Annual FEV <sub>1</sub> decline (2011 [34]) Breast cancer NRF2 and ARE expression (2012 [27])	A = 0.082 in haplotype A (P = 0.036)	Japanese (915) Finland KBBCP (452)
rs6726395 (rs57309289)	g.178103229A>G	Intron 1	Smoking-related FEV <sub>1</sub> decline and annual FEV <sub>1</sub> decline (2011 [34]) Basal and smoker FEV <sub>1</sub> (2009 [39])	G = 0.884* A = 0.082 in haplotype G = 0.464/1764* in haplotype	Japanese (915) Netherland (2542)
rs2001350 (rs17515179, rs60883775)	g.178100425C>T	Intron 1	Annual FEV <sub>1</sub> decline (2011 [34]) Parkinson's disease (2010 [26])	T = 0.082 in haplotype T = 0.986/350* in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.30–0.60	Japanese (915) Swedish/Polish Caucasian (357)
rs10183914 (rs58731187, rs61374844)	g.178097666C>T	Intron 3	Parkinson's disease (2010 [26])	T = 0.536/188* in haplotype OR = 0.9 or 0.4; CI = 0.60–1.40 or 0.30–0.60	Swedish/Polish Caucasian (357)
rs2706110	g.178092162T>C	3' Flanking	Breast cancer (2012 [27])	T/T OR = 2.079; CI = 1.18–3.68	Finland KBBCP (452)
rs2588882	g.178087165G>T	3' Flanking	Infection-induced asthma (2012 [38])	T/G OR = 0.290; CI = 0.13–0.62	Hungarian Caucasian/Gypsy (307)

-686 in reference [22] = -653 in reference [23] = currently -214; -684 in reference [22] = -651 in reference [23] = currently -212; -651 in reference [22] = -617 in reference [23] = currently -178. Chromosome contig (intron, 3' flanking) or reversed (5' flanking promoter) alleles in bold have been used in the text and Table. OR: odds ratio. CI: 95% confidence interval.



–212G/G (98%, rs67006649, previously –684/–651), –178C allele (73%), and GCC<sub>4</sub> (53%) was predicted to increase respiratory failure development (hazard ratio = 0.95, CI 0.91–0.99) in German COPD patients [32]. Significant interaction was also identified between an intronic SNP G allele (rs6726395, g.178103229A>G, 88.4% frequency) and smoking status on FEV<sub>1</sub> decline, relative to the reference A/A allele, in the above Japanese cohort [34]. Siedlinski et al. [39] reported that the C/C genotype of another intronic SNP (rs2364723, g.178126546G>C) was associated with a lower FEV<sub>1</sub> level compared to the wild-type genotype (G/G) in two Netherland cohorts (CI, –63.6–17.8, frequency = 0.525, and pooled cohort size = 2,542). This SNP alone or as a haplotype with 4 more intronic SNPs (rs13001694G/rs1806649T/rs4243387T/rs6726395G) was also associated with high FEV<sub>1</sub> levels in individuals that ever smoked [39]. In a Hungarian population of childhood asthma, SNPs at –178 (C/A) and 3' flanking (rs2588882T/G) loci were inversely associated with infection-induced asthma (OR 0.437; CI 0.28–0.80, OR 0.290; CI 0.13–0.62, resp.), and these SNPs significantly influenced an asthma-environmental pollution interaction [38]. The intronic SNP rs1806649 (C>T) was associated but not significantly with an increased risk of hospitalization during high-level particulate matter (PM<sub>10</sub>) periods in asthma or COPD patients ( $n = 209$ ) of the United Kingdom (UK) [41]. Asthma and COPD admission rates were related to the increase in environmental PM<sub>10</sub> concentration. Importantly, effects of interaction between prenatal stress and *NRF2* SNPs on descendant pulmonary health were investigated by the Avalon Longitudinal Study in the UK: maternal smoking during pregnancy was not associated with lung function change determined by maximum mild expiratory flow (FEF<sub>25–75</sub>) or with asthma incidence in school-aged children, and this relation was not modified by *NRF2* SNP genotypes [42]. However, early gestation acetaminophen exposure significantly influenced the risk of asthma and wheezing at the age of 7 years in >4,000 mothers and >5,000 children [33]. When maternal copies of the –212A allele were present, association with asthma (1,137/4,891; OR 1.73, CI 1.22–2.45) and wheezing (1,149/4,949; OR 1.53, 95% CI 1.06–2.20) was significantly increased [33].

**Gastrointestinal Disorders.** While there was no evidence in lung cancer cases, studies in Japanese populations suggested a potential association of *NRF2* variations with gastrointestinal tumorigenesis. *Helicobacter pylori* (*H. pylori*) causes gastritis which can lead to gastric atrophy and cancer. In gastric epithelium from the Japanese cancer cohorts (39 gastric cancers, 46 controls), *H. pylori* infection was positively correlated with aberrant CpG island methylation of tumor suppressor genes (e.g., p14), and –214G/–212G or –214A/–212G *NRF2* haplotype was significantly associated with increased (OR 2.90; 95% CI 1.14–7.36) or decreased (OR 0.33; 95% CI 0.13–0.88) risk of the CpG methylation, respectively, in the *H. pylori*-infected patients [29]. Further study from the same investigators determined that –214A/–212G allele carriers had significantly ( $P = 0.022$ ) reduced risk of gastric cancer in *H. pylori*-negative cases [30]. The –214A/G–212A/G genotypes were negatively associated

(OR 0.45, CI 0.22–0.93), and the –214G–212G genotypes were positively associated (chronic continuous phenotype; OR 2.57, CI 1.01–6.60) with ulcerative colitis (89 patients, 141 controls) in a Japanese population [31].

**Autoimmune Disorders.** Systemic lupus erythematosus (SLE) is a long-term autoimmune disease more frequently found in females than in males. It affects organs including skin, joints, kidneys, and brain, and nephritis is an aggressive characteristic in some patients. Genome-wide association studies in humans identified a suggestive quantitative trait locus near *NRF2* [43]. A study of a Mexican Mestizo population (362 patients with childhood-onset SLE, 379 controls, and 212 nephritis diagnosed) determined that lupus with nephritis was significantly (OR 1.81, CI 1.04–3.12) associated with the –214G/A SNP in females [25]. The same SNPs were not closely associated with SLE risk in a Japanese cohort [22]. Vitiligo is a skin condition in which there is a loss of brown color (pigment) from areas of skin, resulting in irregular white patches. It is thought to be an autoimmune disease caused by loss of cells (melanocytes) that produce brown pigment. A study indicated that the –178A allele increased the risk of vitiligo dose-dependently (OR 1.724, 95% CI 1.35–2.21 for C/A; OR 2.902, CI 1.62–5.19 for A/A) [35].

**Female Disorders.** It is well known that estrogen metabolites (e.g., catechols) cause ROS formation suggesting correlation of *NRF2* and downstream effectors in postmenopausal mammary cancer. In a study of a Finish population (Kuopio Breast Cancer Project,  $n = 452$  patients, 370 controls), the –178A/A homozygous genotype (OR 4.656; CI = 1.35–16.06) and 3' flanking rs2706110 (T/T; OR 2.079, CI 1.18–3.68) genotype were associated with increased risk of breast cancer, while the 5' flanking –3,306T/T homozygous allele was significantly associated with lower survival (frequency = 71/219, OR 1.687, CI 1.105–2.75) [27], suggesting that *NRF2* genetic polymorphisms affect susceptibility and outcome of the patients. The –178A allele carriers together with intronic rs1962142A allele carriers were associated with lowered tissue levels of *NRF2* proteins [27]. In postmenopausal women, the –178A allele (OR 17.9; 95% CI 3.70–85.70) appeared to modify the risk of venous thromboembolism caused by oral estrogen therapy (A/A or A/C frequency = 33.3%) as demonstrated by the French ESTHER study (161 cases, 474 controls) [36]. An intronic rs1806649C>T SNP did not associate with breast cancer risk in postmenopausal women [40]. However, when this SNP and other at-risk alleles of ARE-responsive genes (*NQO1*, *HO-1*, *NOS3*) were combined, there was a significant gene-dose effect on the breast cancer risk [40]. Although coding region SNPs in *NRF2* and *KEAP1* were identified in the Japanese endometrial adenocarcinoma patients, no association of *NRF2* SNPs with the disease was found [44].

**Neurodegenerative Diseases.** Oxidative stress is known to be involved in Parkinson's disease (PD) presumably due to production of ROS from high-dopamine metabolism and low levels of antioxidants in the substantia nigra of the brain. Investigators found a protective *NRF2* haplotype consisting of four 5' flanking SNPs (–5238G/–214A/–212G/–178C)

and 4 intronic SNPs (rs2886161A/rs1806649A/rs2001350A/rs10183914A) from Swedish (OR 0.9, CI 0.60–1.40) and Polish (OR 0.4, CI 0.30–0.60) populations (total 165 + 192 PD cases, 190 + 192 controls) [26]. The investigators also suggested that *NRF2* haplotype alleles were associated with 2 years earlier age of Alzheimer's disease (AD) onset, 4 years earlier age of posterior subcapsular cataract surgery, and 4 years later age of cortical cataract surgery while they were not significantly related to AD or age-related cataract risk [45].

**3.4. Genetic Mutations in Mouse *Nrf2*.** Tsang et al. [46] compiled 673 SNPs in 55 mouse strains and constructed their phylogenetic tree to correlate and clarify the origins of strains based on the assembled mouse genome sequence and SNP data [20, 47, 48]. Recently, using the complete genome sequence of C57BL/6J (B6) mouse as a reference, high-density SNP screening in other laboratory strains or in panels of strains has been published (see [17]). Although millions of mouse SNPs (>10,089,892 as of December 2012) and haplotype mappings from more than 120 strains have been published as valuable references for dissecting the genetic basis of complex traits [49–51], little attention has been paid to polymorphisms of *Nrf2* and their correlation with disease phenotypes.

Figure 3 demonstrates the proximal promoter region (–1 to –950)/5'-UTR (exon 1, up to TSS) and protein sequence of mouse *Nrf2* based on GRCm38.p1 Primary Assembly (75,704,641–75,675,513 bp), mRNA variant 1, and protein (NP\_035032, encoded by NM\_010902:234–2,027 bp) sequences. Genetic variations in the *Nrf2* genome of inbred strains collected from public databases are listed in Tables S2 and 5. (See Supplementary Table S2) Overall, 968 genetic mutations are compiled for *Nrf2* gene and 5 kb upstream/2 kb downstream regions: 785 SNPs between B6 and another 16 strains were acquired from the Mouse Phenome Database (MPD, <http://phenome.jax.org/db/q?rtn=snp/ret1>), and additional SNPs and other mutations were acquired from NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/snp/?term=mus+musculus%20nfe2l2>). In total, 132 mutations are in the promoter (37 in proximal 1 kb), 49 in exons (38 in coding region, 19 Cns), 727 in introns, and 60 in the 3' flanking region. Excluding mutations in the 5' and 3' flanking sequences, murine *Nrf2* sequences appear to be more highly variable (1 variation per 37.5 bp) than much of the mouse genome which has an approximate frequency of one SNP per every 245 bp ([http://www.informatics.jax.org/mgihome/homepages/stats/all\\_stats.shtml#allstats\\_snp](http://www.informatics.jax.org/mgihome/homepages/stats/all_stats.shtml#allstats_snp)).

*Nrf2* was found to be a susceptibility gene from genome-wide linkage analysis in a murine model of hyperoxia-induced ALI [52]. A promoter SNP –103T>C (previously published as –336T>C) in *Nrf2* was found and predicted to add an additional Sp1 binding site in hyperoxia-susceptible B6 mice, but not in resistant C3H/HeJ mice [52]. Genotypes from the SNP and from simple-sequence length polymorphism markers of the *Nrf2* locus (D2Mit248 and D2Mit94) cosegregated in the B6C3F<sub>2</sub> mouse cohort [52], and *Nrf2* deficient mice were significantly more susceptible to ALI

sub-phenotypes caused by hyperoxia than similarly exposed wild-type mice, supporting *Nrf2* as a contributor to the phenotypic traits [53]. Although no other functional analyses on *Nrf2* SNPs or haplotype association studies have been conducted in inbred mice, strains bearing haplotypes such as multiple Cns in functional domains (e.g., F71L, L451V, H543Q, and L575M) may be useful to elucidate the role of *Nrf2* in differential susceptibility to oxidative diseases.

#### 4. Oncogenic Somatic Mutations in Human NF-E2-Related Factor 2

Somatic mutation is a change in the DNA of somatic cells that affects derived cells but is not inherited by offspring. Efforts to discover somatic mutations have provided insight into mutagenesis and cancer development. Lung cancer, particularly non-small cell lung cancer (NSCLC), is the leading cause of cancer death worldwide. Somatic mutations of *NRF2* and *KEAP1* discovered in lung cancer patients have determined the oncogenic potential of *NRF2* [54, 55]. *KEAP1* somatic mutations were associated with its reduced protein levels in lung cancer tissues and cells [56, 57]. Investigations of NSCLC in various ethnic populations as well as cancers in gastrointestinal, breast, and prostate have coordinately demonstrated that multiple Cns somatic mutations in *KEAP1* cause dysfunction of the translated protein and in turn constitutive activation of *NRF2*, increasing risk of neoplasia and chemoresistance [12, 55, 58, 59]. Somatic mutations of *NRF2* have been detected in various cancer tissues (largely squamous cell carcinomas) in Asian populations (Table 6). *NRF2* mutations were significantly associated with NSCLC cases (squamous cell lung carcinoma, adenocarcinoma) of the Japanese (10.7%, [54]), the Chinese (23%, [60]), and the Koreans (8%, [61]) as well as with lung cancer cell lines. Smoking history was also correlated with mutation occurrence in all of the studies [54, 60, 61]. In addition to lung cancers, laryngeal squamous carcinoma (13% in [61]), esophageal squamous cancer (ESC, 22% in [60], 11.4% in [61]), head and neck cancers (25% in [54]), skin (1/17 case in [61]), and oral cancer cell lines had somatic changes in *NRF2*. In contrast to wide-spread *KEAP1* mutations, mutations in *NRF2* were clustered in DLG/ETGE motifs of the Neh2 domain, which are critical in the “hinge and latch” model of *KEAP1* binding [12]. Similar to *KEAP1* somatic mutations, it has been postulated that *NRF2* mutations in cancer cells lead to *NRF2* accumulation by suppressing its ubiquitination or *KEAP1* binding, which eventually confers malignant potential and resistance to chemotherapy.

Most variable sites in *NRF2* included aa residues 29 (Asp, D), 31 (Gly, G), 77 (Asp, D), and 79 (Glu, E) (Table 6). Residue 33 (Ser, S) in the Neh2 domain is mutated by either genetic or somatic processes (Figure 4). Cns in the EDGF motif of *NRF2* was experimentally determined to impair recognition of *KEAP1* [54]. *NRF2* mutations were significantly correlated with increased (2.5-fold) copy number (31% of mutants versus 3% wild types) in Japanese NSCLC cases [63]. Aberrant mutation of *NRF2* also led to increased expression of downstream effectors including *RagD* known to



FIGURE 3: DNA (partial promoter and exons 1 and 5) and protein sequence of mouse Nrf2. SNPs and amino acid residues for non-synonymous SNPs are marked. Promoter regions bearing 5 CpG islands are underlined.

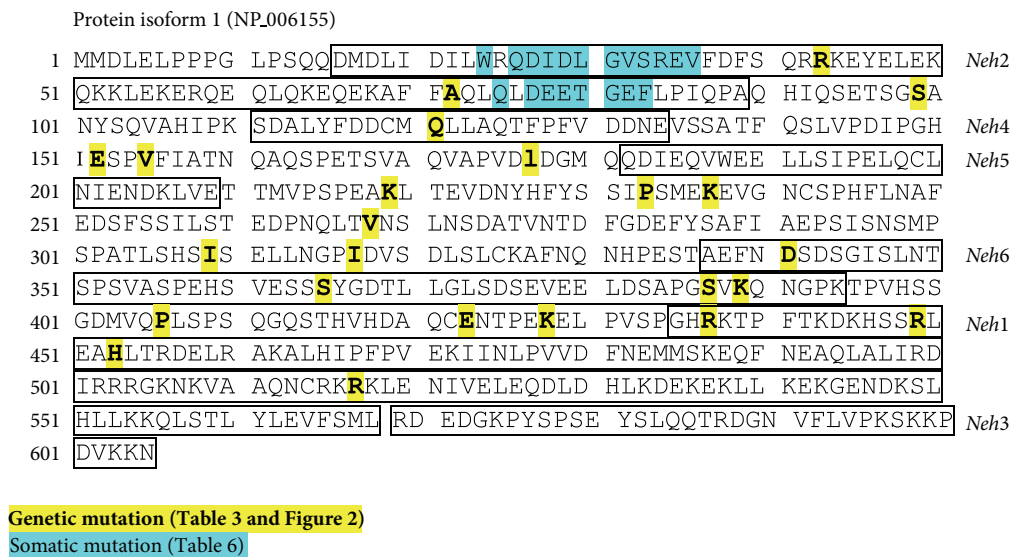


FIGURE 4: Genetic and somatic mutation loci in human NRF2 protein.







TABLE 5: Continued.

SNP ID	Chr2 location (bp)	SNP allele	Region	Location from TSS/reversed SNP allele	Consequences	Strains																		
						1	B6	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
rs27978452	75677145	C/T (G/A)	Exon5	843	Cns A204T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T		
rs225047937	75677086–75677078	GAGATCGAT del (ATGCAATCIC del)	Exon5	902–910	Cds-indel (S224/I225/S226)	GAG ATC GAT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
rs241469868	75676998	T/A (G/T)	Exon5	990	Cn T253S	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs221602571	75676923	A/C (T/G)	Exon5	1065	Cns S278A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs252576472	75676876	G/A (C/T)	Exon5	1112	Cs S293=	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs250802933	75676841	T/C (A/G)	Exon5	1147	Cns D305G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs225425698	75676792	C/T (G/A)	Exon5	1196	Cs P321=	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
rs13459064	75676732	C/T (G/A)	Exon5	1256	Cs T341=	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C	T	C
rs27978453	75676717	C/T (G/A)	Exon5	1271	Cs A346=	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
rs214335034	75676636	A/G (T/C)	Exon5	1352	Cs D373=	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs249583644	75676603	A/C (T/G)	Exon5	1385	Cs P384=	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs27978454	75676592	G/A (C/T)	Exon5	1396	Cns P388L	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs215384328	75676589	G/C (C/G)	Exon5	1399	Cns A389G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs251728286	75676571	G/A (C/T)	Exon5	1417	Cns T395I	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs247602334	75676567	T/C (A/G)	Exon5	1421	Cs V396=	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs234216231	75676530	T/C (A/G)	Exon5	1458	Cns M409V	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs212904337	75676526	C/T (G/A)	Exon5	1462	Cns R410H	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C
rs252650779	75676522	T/C (A/G)	Exon5	1466	Cs E411=	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs231273560	75676516	T/C (A/G)	Exon5	1472	Cs Q413=	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs257886949	75676413	G/T (C/A)	Exon5	1575	Cs R448=	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs241537608	75676404	G/C (C/G)	Exon5	1584	Cns L451V	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs227619071	75676312	T/C (A/G)	Exon5	1676	Cs Q481=	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
rs258913831	75676290	A/G (T/C)	Exon5	1698	Cs L489=	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A
rs225593319	75676126	A/T (T/A)	Exon5	1862	Cns H543Q	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs4223233	75676105	G/A (C/T)	Exon5	1883	Cs S550=	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs4223232	75676032	G/T (C/A)	Exon5	1956	Cns L575M	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs4223231	75675816	C/G	UTR-3	2172	NC	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs1346860	75675682	A/T	UTR-3	2306	NC	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

Sequence variations in mouse *Nrf2* were obtained from Mouse Phenome Database (<http://phenomejax.org/SNP>) and NCBI SNP database (<http://www.ncbi.nlm.nih.gov/SNP>). NCBI reference sequence is *Mus musculus* strain C57BL/6J chromosome 2, GRChm38.p1 (NC\_000068.7, GI: 372099108, 75,704,641–75,675,519,123 bp). Total 968 genetic mutations are reported in *Nrf2* gene and 5 kb upstream (≥75,704,642; –1) and 2 kb downstream (≤75,675,512) sequences as of January 2003. All SNPs and alleles are presented as genomic contig (reversed sequence indicated). Exon 1: 75,704,641–75,704,364 (278 bp, TTS = 75,704,408), exon 2: 75,679,431–75,679,165 (267 bp), exon 3: 75,678,579–75,678,490 (90 bp), exon 4: 75,677,713–75,677,546 (168 bp), and exon 5: 75,677,184–75,678,849 (1,666 bp). Protein amino acid (aa) residues in NP\_000615.2 (597 aa). Cs: coding-synonymous. Cns: coding-nonsynonymous. Amino acid A-alanine, D-aspartic acid, E-glutamic acid, F-phenylalanine, G-glycine, H-histidine, I-isoleucine, L-leucine, and M-methionine. N-asparagine, P-proline, Q-glutamine, R-arginine, S-serine, T-threonine, and V-valine. \* Errors in databases fixed. NC: not confirmed.

TABLE 6: *NRF2* somatic mutations revealed in various human cancers.

Domain	Locus	Amino acid residues		DNA mutation	Cancer types (cases)	References
		Wild type	Mutant			
DLG motif	24	W (Trp)	C (Cys)	c.72G>C/G>T	NSCLC, neck, ESC	[54, 62]
			K (Lys)	c.72T>C	ESC	[62]
	26	Q (Gln)	E (Glu)	c.76C>G	NSCLC, ESC	[54, 62]
			G (Gly)*	c.80G>A	NSCLC	[60]
	27	D (Asp)	Y (Tyr)	c.79G>T	ESC	[61]
			T (Thr)	c.83C>T	NSCLC	[54]
	29	D (Asp)	G (Gly)	c.86A>G	Head and neck, ESC	[54, 62]
			H (His)	c.85G>C	NSCLC, larynx	[60, 61]
	30	L (Leu)	F (Phe)	c.88C>T	NSCLC, ESC	[54, 62]
	31	G (Gly)	A (Ala)	c.92G>C	NSCLC, ESC, skin	[54, 60–62]
			T (Thr)	c.95T>G	NSCLC	[54]
	32	V (Val)	del	c.93_95delAGT	ESC	[61]
			S-R-E-V	S-R-E-V-S-R-E-V*	c.97_108dupAGTCGAGCCGTA	ESC
34	R (Arg)	Q (Gln)	c.101G>A	NSCLC	[54, 60, 61]	
		P (Pro)	c.101G>C	NSCLC	[63]	
ETGF motif	75	Q (Gln)	H (His)	c.225A>C	Head and neck, ESC	[54, 62]
			V (Val)	c.230A>T	NSCLC, ESC	[54, 60, 62]
	77	D (Asp)	G (Gly)	c.230A>G	ESC	[62]
			A (Ala)	c.230A>C	NSCLC	[61]
			N (Asn)	c.229G>A	Larynx	[61]
	78	E (Glu)	K (Lys)	c.232G>A	NSCLC, ESC	[54, 62]
			K (Lys)	c.235G>A	NSCLC, ESC	[54, 61, 62]
	79	E (Glu)	Q (Gln)	c.235G>C	NSCLC, ESC	[54, 60–62]
			G (Gly)	c.236A>G	Larynx	[61]
			E-E	c.234_236dupAGA	ESC	[61]
	80	T (Thr)	K (Lys)	c.239C>A/C>G	NSCLC, ESC	[54, 61, 62]
			P (Phe)	c.238A>C	ESC	[62]
	80	T (Thr)	I (Ile)	c.239C>T	Head and neck	[54]
			A (Ala)	c.238A>G	NSCLC	[63]
			V (Val)	c.242G>T	ESC	[61]
	81	G (Gly)	D (Asp)	c.242G>A	NSCLC, ESC	[60–62]
			D (Asp)	c.246A>T	ESC, oral cancer cell line	[54, 62]
G (Gly)			c.245A>G	NSCLC	[54]	
82	E (Glu)	Q (Gln)*	c.244G>C	NSCLC, ESC	[60, 61]	
		V (Val)*	c.245A>T	ESC	[61]	
		L (Leu)*	c.247T>C	NSCLC	[60]	

NSCLC: non-small cell lung cancer, ESC: esophageal squamous cancer, and \* errors in reference fixed. Number of cases: 82 NSCLC and 10 ESC in [62]; 125 NSCLC, 70 ESC, 23 larynx, and 17 skin in [61]; 103 NSCLC and 12 head and neck in [54]; 90 NSCLC in [63]; 103 NSCLC in [60].

be involved in squamous lung cancer cell proliferation [64], suggesting that the mutation is functional and overcomes KEAP1 inhibition. Singh et al. [65] determined *in vitro* that RNAi-mediated depletion of *NRF2* in lung cancer cells enhanced ROS production and susceptibility to cell death by ionizing radiation. These studies support the concept that elevated *NRF2* and ARE responsiveness provides cancer cells with proliferative advantage for malignant transformation

and undue protection from anti-cancer therapy. Oncogenic epidermal growth factor receptor (EGFR) signaling is recently found to be critical in *NRF2*-mediated proliferation of NSCLC cells [66].

Collectively, “gain of function” mutations in *NRF2* that reduce KEAP1 recognition are suggested to be predictive markers for poor responsiveness to chemotherapy and radiation therapy. Although *NRF2*-mediated cellular defense

processes are essential in the initiation stage, enhanced NRF2-ARE activity in advanced stages of cancer development may create a favorable intracellular environment for tumor cell growth and survival [67, 68]. In this context, NRF2 may be a potential molecular target for the treatment of radio-resistance cancers, especially those that have “loss of function” mutations in *EGFR*, *KRAS*, or *KEAP1* as well as “gain of function” mutations in *NRF2*.

## 5. Epigenetic Alterations of NF-E2-Related Factor 2

Epigenetic modifications are alterations of molecules interacting with genes without changes to the primary DNA sequence. They include post-translational modification of histones, DNA methylation events, chromatin conformational changes, and alterations to noncoding regulatory RNAs. Epigenetic alterations are stable and often inheritable but are reversible and may affect expression of the gene. Dysregulation or defects in epigenetic processes, particularly hypermethylation of tumor suppressor gene promoters (e.g., CpG islands) or histone modifications, are thought to be associated with carcinogenesis. Investigators have reported hypermethylation in CpG islands of *KEAP1* which were associated with reduced *KEAP1* expression in human cancers from lung, prostate, colon, and so forth, [69–72]. Similar to somatic mutations, epigenetic changes on *KEAP1* impaired the function of its encoded protein leading to constitutive NRF2 activation.

Supporting the role of “pathogenic mutations” in *NRF2*, expression of Nrf2 and downstream Nqo1 was suppressed in prostate tumors of mice (transgenic adenocarcinoma of mouse prostate, TRAMP). Among 15 promoter CpG islands located between –942 and –654 (c.–1175\_c.–1132 and c.–1059\_c.–887, gap in c.–1131\_c.–1060; see Figure 3), hypermethylation of the first 5 CpG islands (–942.–899 and c.–1175\_c.–1132) was significantly associated with tumorigenesis [73]. Moreover, treatment with inhibitors for DNA methyltransferase and histone deacetylase restored Nrf2 expression in these tumor cells [73]. A dietary phytochemical curcumin known as a DNA hypomethylation agent restored epigenetically silent *Nrf2* expression through CpG demethylation in carcinogen-induced mouse tumor cells [74].

The whole genome epigenetic datasets for 5 species are publicly accessible at NCBI Epigenomics [75, 76]. The human *NRF2* epigenome of primary cells (breast, penis) and H1 stem cell line as well as mouse *Nrf2* CpG island methylation data for sperm, blood, and cerebellum are currently available (<http://www.ncbi.nlm.nih.gov/epigenomics>). Although no direct evidence of disease-associated epigenetic modulation has been identified in human *NRF2*, various phytochemical NRF2 agonists such as sulforaphane and curcumin have shown their roles in DNA methylation and histone modification (see reviews by Lee and colleagues, e.g., [77]). Taken together, epigenetic modifications of the *NRF2/KEAP1* axle are predicted to cause dysregulation of ARE-mediated cellular defense leading to deleterious health effects, and

phytochemical antioxidants as epigenetic modulators for *NRF2* are suggested to be useful in cancer prevention.

## 6. Conclusions

*NRF2* is evolutionally conserved with high-sequence homology in many species. However, it is a highly mutable gene, and numerous genetic variants have been discovered in human ethnic groups. Importantly, certain SNPs or haplotypes have been identified in various diseases as “at-risk” alleles and are related to functional alterations. In addition to genetic variations, multiple somatic mutations identified in the *KEAP1* recognition domain of *NRF2* in cancer cells have been found to be oncogenic due to dysregulation of NRF2 homeostasis by its excess “gain of function”. Epigenetic alteration of the *NRF2* is under investigation and is predicted to have pathogenic influences as learned from mouse and phytochemical agonist studies. Continuous updates of *Nrf2* allelic variants in inbred mouse strains will provide a useful tool for effective experimental designs for models of oxidative disorders to provide insight into the disease mechanisms and intervention strategies.

## Abbreviations

AD:	Alzheimer’s disease
ALI:	acute lung injury
ARE:	antioxidant response element
bZIP:	basic leucine zipper
cds:	coding DNA sequences
CI:	confidence interval
COPD:	chronic obstructive pulmonary disease
ESC:	esophageal squamous cancer
FEV <sub>1</sub> :	forced expiratory volume in one second
Gbp:	giga base pairs
GST:	glutathione S-transferase
GWAS:	genome-wide association study
HGVS:	Human Genome Variation Society
HO-1:	heme oxygenase-1
<i>H. pylori</i> :	<i>Helicobacter pylori</i>
KEAP1:	Kelch-like ECH activating protein 1
MPD:	Mouse Phenome Database
MRP:	multidrug resistance protein
NCBI:	National Center for Biotechnology Information
Neh:	NRF2-ECH homology
Nfe2l2:	nuclear factor (erythroid-derived 2)-like 2
Nrf2:	NF-E2-related factor 2
NSCLC:	nonsmall cell lung cancer
NQO1:	NAD(P)H:quinone oxidoreductase 1
OR:	odds ratio
PD:	Parkinson’s disease
PM:	particulate matter
ROS:	reactive oxygen species
SLE:	systemic lupus erythematosus
SNP:	single nucleotide polymorphism
SOD:	superoxide dismutase
URT:	untranslated region.

## Disclosure

Author's contribution to the Work was done as part of the Author's official duties as a NIH employee and is a Work of the United States Government. Therefore, copyright may not be established in the United States.

## Conflict of Interests

The author declares that there is no conflict of interests.

## Acknowledgments

The research related to this paper was supported by the Intramural Research Program of the National Institutes of Health of the National Institute of Environmental Health Sciences (NIEHS). Drs. Steven Kleeberger and Stephanie London at the NIEHS provided excellent critical review of this paper. The author thanks Mrs. Jacqui Marzec for her helpful comments and English editing.

## References

- [1] J. Y. Chan, M. C. Cheung, P. Moi, K. Chan, and Y. W. Kan, "Chromosomal localization of the human NF-E2 family of bZIP transcription factors by fluorescence in situ hybridization," *Human Genetics*, vol. 95, no. 3, pp. 265–269, 1995.
- [2] K. Itoh, K. Igarashi, N. Hayashi, M. Nishizawa, and M. Yamamoto, "Cloning and characterization of a novel erythroid cell-derived CNC family transcription factor heterodimerizing with the small Maf family proteins," *Molecular and Cellular Biology*, vol. 15, no. 8, pp. 4184–4193, 1995.
- [3] W. O. Osburn and T. W. Kensler, "NRF2 signaling: an adaptive response pathway for protection against environmental toxic insults," *Mutation Research*, vol. 659, no. 1-2, pp. 31–39, 2008.
- [4] D. Papp, K. Lenti, D. Modos, D. Fazekas, Z. Dul, D. Turei et al., "The NRF2-related interactome and regulome contain multifunctional proteins and fine-tuned autoregulatory loops," *FEBS Letters*, vol. 586, no. 13, pp. 1795–1802, 2012.
- [5] X. Wang, D. J. Tomso, B. N. Chorley et al., "Identification of polymorphic antioxidant response elements in the human genome," *Human Molecular Genetics*, vol. 16, no. 10, pp. 1188–1200, 2007.
- [6] D. Malhotra, E. Portales-Casamar, A. Singh et al., "Global mapping of binding sites for NRF2 identifies novel targets in cell survival response through CHIP-Seq profiling and network analysis," *Nucleic Acids Research*, vol. 38, no. 17, pp. 5718–5734, 2010.
- [7] K. Chan, R. Lu, J. C. Chang, and Y. W. Kan, "NRF2, a member of the NFE2 family of transcription factors, is not essential for murine erythropoiesis, growth, and development," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 24, pp. 13943–13948, 1996.
- [8] K. Itoh, T. Chiba, S. Takahashi et al., "An NRF2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements," *Biochemical and Biophysical Research Communications*, vol. 236, no. 2, pp. 313–322, 1997.
- [9] F. Martin, J. M. van Deursen, R. A. Shivdasani, C. W. Jackson, A. G. Troutman, and P. A. Ney, "Erythroid maturation and globin gene expression in mice with combined deficiency of NF-E2 and Nrf-2," *Blood*, vol. 91, no. 9, pp. 3459–3466, 1998.
- [10] K. Itoh, N. Wakabayashi, Y. Katoh et al., "KEAP1 represses nuclear activation of antioxidant responsive elements by NRF2 through binding to the amino-terminal Neh2 domain," *Genes and Development*, vol. 13, no. 1, pp. 76–86, 1999.
- [11] A. Kobayashi, M. I. Kang, H. Okawa et al., "Oxidative stress sensor KEAP1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of NRF2," *Molecular and Cellular Biology*, vol. 24, no. 16, pp. 7130–7139, 2004.
- [12] K. Taguchi, H. Motohashi, and M. Yamamoto, "Molecular mechanisms of the KEAP1-NRF2 pathway in stress response and cancer evolution," *Genes to Cells*, vol. 16, no. 2, pp. 123–140, 2011.
- [13] K. Mukaigasa, L. T. Nguyen, L. Li, H. Nakajima, M. Yamamoto, and M. Kobayashi, "Genetic evidence of an evolutionarily conserved role for NRF2 in the protection against oxidative stress," *Molecular and Cellular Biology*, vol. 32, no. 21, pp. 4455–4461, 2012.
- [14] J. Maher and M. Yamamoto, "The rise of antioxidant signaling—the evolution and hormetic actions of NRF2," *Toxicology and Applied Pharmacology*, vol. 244, no. 1, pp. 4–15, 2010.
- [15] S. C. Lo, X. Li, M. T. Henzl, L. J. Beamer, and M. Hannink, "Structure of the KEAP1:NRF2 interface provides mechanistic insight into NRF2 signaling," *The EMBO Journal*, vol. 25, no. 15, pp. 3605–3617, 2006.
- [16] V. A. McKusick, "Mendelian inheritance in man and its online version, OMIM," *American Journal of Human Genetics*, vol. 80, no. 4, pp. 588–604, 2007.
- [17] B. Yalcin, D. J. Adams, J. Flint, and T. M. Keane, "Next-generation sequencing of experimental mouse strains," *Mammalian Genome*, vol. 23, no. 9-10, pp. 490–498, 2012.
- [18] D. R. Bentley, S. Balasubramanian, H. P. Swerdlow et al., "Accurate whole human genome sequencing using reversible terminator chemistry," *Nature*, vol. 456, no. 7218, pp. 53–59, 2008.
- [19] P. S. G. Chain, D. V. Grafham, R. S. Fulton et al., "Genome project standards in a new era of sequencing," *Science*, vol. 326, no. 5950, pp. 236–237, 2009.
- [20] R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal et al., "Initial sequencing and comparative analysis of the mouse genome," *Nature*, vol. 420, pp. 520–562, 2002.
- [21] T. M. Keane, L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin et al., "Mouse genomic variation and its effect on phenotypes and gene regulation," *Nature*, vol. 477, pp. 289–294, 2011.
- [22] T. Yamamoto, K. Yoh, A. Kobayashi et al., "Identification of polymorphisms in the promoter region of the human NRF2 gene," *Biochemical and Biophysical Research Communications*, vol. 321, no. 1, pp. 72–79, 2004.
- [23] J. M. Marzec, J. D. Christie, S. P. Reddy et al., "Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury," *The FASEB Journal*, vol. 21, no. 9, pp. 2237–2246, 2007.
- [24] H. Fukushima-Uesaka, Y. Saito, K. Maekawa et al., "Genetic variations and haplotype structures of transcriptional factor NRF2 and its cytosolic reservoir protein KEAP1 in Japanese," *Drug Metabolism and Pharmacokinetics*, vol. 22, no. 3, pp. 212–219, 2007.
- [25] E. J. Córdova, R. Velázquez-Cruz, F. Centeno, V. Baca, and L. Orozco, "The NRF2 gene variant, -653G/A, is associated with



- nephritis in childhood-onset systemic lupus erythematosus," *Lupus*, vol. 19, no. 10, pp. 1237–1242, 2010.
- [26] M. von Otter, S. Landgren, S. Nilsson et al., "Association of *NRF2*-encoding *NFE2L2* haplotypes with Parkinson's disease," *BMC Medical Genetics*, vol. 11, no. 1, article 36, 2010.
- [27] J. M. Hartikainen, M. Tengstrom, V. M. Kosma, V. L. Kinnula, A. Mannermaa, and Y. Soini, "Genetic polymorphisms and protein expression of *NRF2* and sulfiredoxin predict survival outcomes in breast cancer," *Cancer Research*, vol. 72, pp. 5537–5546, 2012.
- [28] T. Arisawa, T. Tahara, T. Shibata et al., "The relationship between *Helicobacter pylori* infection and promoter polymorphism of the *NRF2* gene in chronic gastritis," *International Journal of Molecular Medicine*, vol. 19, no. 1, pp. 143–148, 2007.
- [29] T. Arisawa, T. Tahara, T. Shibata et al., "The influence of promoter polymorphism of nuclear factor-erythroid 2-related factor 2 gene on the aberrant DNA methylation in gastric epithelium," *Oncology Reports*, vol. 19, no. 1, pp. 211–216, 2008.
- [30] T. Arisawa, T. Tahara, T. Shibata et al., "*NRF2* gene promoter polymorphism and gastric carcinogenesis," *Hepato-Gastroenterology*, vol. 55, no. 82–83, pp. 750–754, 2008.
- [31] T. Arisawa, T. Tahara, T. Shibata et al., "*NRF2* gene promoter polymorphism is associated with ulcerative colitis in a Japanese population," *Hepato-Gastroenterology*, vol. 55, no. 82–83, pp. 394–397, 2008.
- [32] C. C. Hua, L. C. Chang, J. C. Tseng, C. M. Chu, Y. C. Liu, and W. B. Shieh, "Functional haplotypes in the promoter region of transcription factor *NRF2* in chronic obstructive pulmonary disease," *Disease Markers*, vol. 28, no. 3, pp. 185–193, 2010.
- [33] S. O. Shaheen, R. B. Newson, S. M. Ring, M. J. Rose-Zerilli, J. W. Holloway, and A. J. Henderson, "Prenatal and infant acetaminophen exposure, antioxidant gene polymorphisms, and childhood asthma," *The Journal of Allergy and Clinical Immunology*, vol. 126, no. 6, pp. 1141.e7–1148.e7, 2010.
- [34] H. Masuko, T. Sakamoto, Y. Kaneko et al., "Lower FEV1 in non-COPD, nonasthmatic subjects: association with smoking, annual decline in FEV1, total IgE levels, and TSLP genotypes," *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 6, pp. 181–189, 2011.
- [35] C. P. Guan, M. N. Zhou, A. E. Xu et al., "The susceptibility to vitiligo is associated with NF-E2-related factor 2 (*NRF2*) gene polymorphisms: a study on Chinese Han population," *Experimental Dermatology*, vol. 17, no. 12, pp. 1059–1062, 2008.
- [36] J. Bouligand, O. Cabaret, M. Canonico et al., "Effect of *NFE2L2* genetic polymorphism on the association between oral estrogen therapy and the risk of venous thromboembolism in postmenopausal women," *Clinical Pharmacology & Therapeutics*, vol. 89, no. 1, pp. 60–64, 2011.
- [37] D. S. O'Mahony, B. J. Glavan, T. D. Holden, C. Fong, R. A. Black, G. Rona et al., "Inflammation and immune-related candidate gene associations with acute lung injury susceptibility and severity: a validation study," *PLoS ONE*, vol. 7, no. 12, Article ID e51104, 2012.
- [38] I. Ungvari, E. Hadadi, V. Virag et al., "Relationship between air pollution, *NFE2L2* gene polymorphisms and childhood asthma in a Hungarian population," *Journal of Community Genetics*, vol. 3, no. 1, pp. 25–33, 2012.
- [39] M. Siedlinski, D. S. Postma, J. M. A. Boer et al., "Level and course of FEV1 in relation to polymorphisms in *NFE2L2* and *KEAP1* in the general population," *Respiratory Research*, vol. 10, article 73, 2009.
- [40] C. C. Hong, C. B. Ambrosone, J. Ahn et al., "Genetic variability in iron-related oxidative stress pathways (*NRF2*, *NQO1*, *NOS3*, and *HO-1*), iron intake, and risk of postmenopausal breast cancer," *Cancer Epidemiology Biomarkers and Prevention*, vol. 16, no. 9, pp. 1784–1794, 2007.
- [41] C. Canova, C. Dunster, F. J. Kelly, C. Minelli, P. L. Shah, C. Caneja et al., "PM10-induced hospital admissions for asthma and chronic obstructive pulmonary disease: the modifying effect of individual characteristics," *Epidemiology*, vol. 23, no. 4, pp. 607–615, 2012.
- [42] A. J. Henderson, R. B. Newson, M. Rose-Zerilli, S. M. Ring, J. W. Holloway, and S. O. Shaheen, "Maternal *NRF2* and glutathione-S-transferase polymorphisms do not modify associations of prenatal tobacco smoke exposure with asthma and lung function in school-aged children," *Thorax*, vol. 65, no. 10, pp. 897–902, 2010.
- [43] C. Xing, A. L. Sestak, J. A. Kelly et al., "Localization and replication of the systemic lupus erythematosus linkage signal at 4p16: interaction with 2p11, 12q24 and 19q13 in European Americans," *Human Genetics*, vol. 120, no. 5, pp. 623–631, 2007.
- [44] T. F. Wong, K. Yoshinaga, Y. Monma, K. Ito, H. Niikura, S. Nagase et al., "Association of *KEAP1* and *NRF2* genetic mutations and polymorphisms with endometrioidometrial adenocarcinoma survival," *International Journal of Gynecological Cancer*, vol. 21, no. 8, pp. 1428–1435, 2011.
- [45] M. von Otter, S. Landgren, S. Nilsson et al., "*NRF2*-encoding *NFE2L2* haplotypes influence disease progression but not risk in Alzheimer's disease and age-related cataract," *Mechanisms of Ageing and Development*, vol. 131, no. 2, pp. 105–110, 2010.
- [46] S. Tsang, Z. Sun, B. Luke et al., "A comprehensive SNP-based genetic analysis of inbred mouse strains," *Mammalian Genome*, vol. 16, no. 7, pp. 476–480, 2005.
- [47] C. M. Wade, E. J. Kulbokas III, A. W. Kirby et al., "The mosaic structure of variation in the laboratory mouse genome," *Nature*, vol. 420, no. 6915, pp. 574–578, 2002.
- [48] T. Wiltshire, M. T. Pletcher, S. Batalov et al., "Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 6, pp. 3380–3385, 2003.
- [49] K. A. Frazer, E. Eskin, H. M. Kang et al., "A sequence-based variation map of 8.27 million SNPs in inbred mouse strains," *Nature*, vol. 448, no. 7157, pp. 1050–1053, 2007.
- [50] T. P. Maddatu, S. C. Grubb, C. J. Bult, and M. A. Bogue, "Mouse phenome database (MPD)," *Nucleic Acids Research*, vol. 40, pp. D887–D894, 2007.
- [51] A. Kirby, H. M. Kang, C. M. Wade et al., "Fine mapping in 94 inbred mouse strains using a high-density haplotype resource," *Genetics*, vol. 185, no. 3, pp. 1081–1095, 2010.
- [52] H. Y. Cho, A. E. Jedlicka, S. P. M. Reddy, L. Y. Zhang, T. W. Kensler, and S. R. Kleeberger, "Linkage analysis of susceptibility to hyperoxia *NRF2* is a candidate gene," *American Journal of Respiratory Cell and Molecular Biology*, vol. 26, no. 1, pp. 42–51, 2002.
- [53] H. Y. Cho, A. E. Jedlicka, S. P. M. Reddy et al., "Role of *NRF2* in protection against hyperoxic lung injury in mice," *American Journal of Respiratory Cell and Molecular Biology*, vol. 26, no. 2, pp. 175–182, 2002.
- [54] T. Shibata, T. Ohta, K. I. Tong et al., "Cancer related mutations in *NRF2* impair its recognition by *KEAP1*-Cul3 E3 ligase and promote malignancy," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 36, pp. 13568–13573, 2008.



- [55] A. Singh, V. Misra, R. K. Thimmulappa et al., “Dysfunctional *KEAPI-NRF2* interaction in non-small-cell lung cancer,” *PLoS Medicine*, vol. 3, no. 10, article e420, 2006.
- [56] T. Ohta, K. Iijima, M. Miyamoto et al., “Loss of *KEAPI* function activates *NRF2* and provides advantages for lung cancer cell growth,” *Cancer Research*, vol. 68, no. 5, pp. 1303–1309, 2008.
- [57] B. Padmanabhan, K. I. Tong, T. Ohta et al., “Structural basis for defects of *KEAPI* activity provoked by its point mutations in lung cancer,” *Molecular Cell*, vol. 21, no. 5, pp. 689–700, 2006.
- [58] T. Shibata, A. Kokubu, M. Gotoh et al., “Genetic alteration of *KEAPI* confers constitutive *NRF2* activation and resistance to chemotherapy in gallbladder cancer,” *Gastroenterology*, vol. 135, no. 4, pp. 1358.e4–1368.e4, 2008.
- [59] N. J. Yoo, H. R. Kim, Y. R. Kim, C. H. An, and S. H. Lee, “Somatic mutations of the *KEAPI* gene in common solid cancers,” *Histopathology*, vol. 60, no. 6, pp. 943–952, 2012.
- [60] Y. Hu, Y. Ju, D. Lin, Z. Wang, Y. Huang, S. Zhang et al., “Mutation of the *NRF2* gene in non-small cell lung cancer,” *Molecular Biology Reports*, vol. 39, no. 4, pp. 4743–4747, 2012.
- [61] Y. R. Kim, J. E. Oh, M. S. Kim et al., “Oncogenic *NRF2* mutations in squamous cell carcinomas of oesophagus and skin,” *The Journal of Pathology*, vol. 220, no. 4, pp. 446–451, 2010.
- [62] T. Shibata, A. Kokubu, S. Saito et al., “*NRF2* mutation confers malignant potential and resistance to chemoradiation therapy in advanced esophageal squamous cancer,” *Neoplasia*, vol. 13, pp. 864–873, 2011.
- [63] H. Sasaki, M. Shitara, K. Yokota, Y. Hikosaka, S. Moriyama, M. Yano et al., “Increased *NRF2* gene (*NFE2L2*) copy number correlates with mutations in lung squamous cell carcinomas,” *Molecular Medicine Reports*, vol. 6, no. 2, pp. 391–394, 2012.
- [64] H. Sasaki, M. Shitara, K. Yokota, Y. Hikosaka, S. Moriyama, M. Yano et al., “RagD gene expression and *NRF2* mutations in lung squamous cell carcinomas,” *Oncology Letters*, vol. 4, pp. 1167–1170, 2012.
- [65] A. Singh, M. Bodas, N. Wakabayashi, F. Bunz, and S. Biswal, “Gain of *NRF2* function in non-small-cell lung cancer cells confers radioresistance,” *Antioxidants and Redox Signaling*, vol. 13, no. 11, pp. 1627–1637, 2010.
- [66] T. Yamadori, Y. Ishii, S. Homma, Y. Morishima, K. Kurishima, K. Itoh et al., “Molecular mechanisms for the regulation of *NRF2*-mediated cell proliferation in non-small-cell lung cancers,” *Oncogene*, vol. 31, pp. 4768–4777, 2012.
- [67] A. K. Bauer, H. Y. Cho, L. Miller-Degraff, C. Walker, K. Helms, J. Fostel et al., “Targeted deletion of *NRF2* reduces urethane-induced lung tumor development in mice,” *PLoS ONE*, vol. 6, no. 10, Article ID e26590, 2011.
- [68] H. Y. Cho and S. R. Kleeberger, “*NRF2* protects against airway disorders,” *Toxicology and Applied Pharmacology*, vol. 244, no. 1, pp. 43–56, 2010.
- [69] R. Wang, J. An, F. Ji, H. Jiao, H. Sun, and D. Zhou, “Hypermethylation of the *KEAPI* gene in human lung cancer cell lines and lung cancer tissues,” *Biochemical and Biophysical Research Communications*, vol. 373, no. 1, pp. 151–154, 2008.
- [70] P. Zhang, A. Singh, S. Yegnasubramanian et al., “Loss of kelch-like ECH-associated protein 1 function in prostate cancer cells causes chemoresistance and radioresistance and promotes tumor growth,” *Molecular Cancer Therapeutics*, vol. 9, no. 2, pp. 336–346, 2010.
- [71] N. Hanada, T. Takahata, Q. Zhou et al., “Methylation of the *KEAPI* gene promoter region in human colorectal cancer,” *BMC Cancer*, vol. 12, article 66, 2012.
- [72] L. A. Muscarella, R. Barbano, V. D’Angelo et al., “Regulation of *KEAPI* expression by promoter methylation in malignant gliomas and association with patient’s outcome,” *Epigenetics*, vol. 6, no. 3, pp. 317–325, 2011.
- [73] S. Yu, T. O. Khor, K. L. Cheung et al., “*NRF2* expression is regulated by epigenetic mechanisms in prostate cancer of TRAMP mice,” *PLoS ONE*, vol. 5, no. 1, Article ID e8579, 2010.
- [74] T. O. Khor, Y. Huang, T. Y. Wu, L. Shu, J. Lee, and A. N. Kong, “Pharmacodynamics of curcumin as DNA hypomethylation agent in restoring the expression of *NRF2* via promoter CpGs demethylation,” *Biochemical Pharmacology*, vol. 82, no. 9, pp. 1073–1078, 2011.
- [75] I. M. Fingerman, L. McDaniel, X. Zhang et al., “NCBI epigenomics: a new public resource for exploring epigenomic data sets,” *Nucleic Acids Research*, vol. 39, supplement 1, pp. D908–D912, 2011.
- [76] I. M. Fingerman, X. Zhang, W. Ratzat, N. Husain, R. F. Cohen, and G. D. Schuler, “NCBI epigenomics: what’s new for 2013,” *Nucleic Acids Research*, vol. 41, no. 1, pp. D221–D225, 2013.
- [77] J. H. Lee, T. O. Khor, L. Shu, Z. Y. Su, F. Fuentes, and A. N. Kong, “Dietary phytochemicals and cancer prevention: *NRF2* signaling, epigenetics, and cell death mechanisms in blocking cancer initiation and progression,” *Pharmacology & Therapeutics*, vol. 137, no. 2, pp. 153–171, 2012.