SCIENTIFIC REPORTS

OPEN

Distinct isoforms of Nrf1 diversely regulate different subsets of its cognate target genes

Received: 10 July 2018
Accepted: 11 January 2019
Published online: 27 February 2019

Meng Wang, Lu Qiu, Xufang Ru, Yijiang Song & Yiguo Zhang

The single *Nrf1* gene has capability to be differentially transcripted alongside with alternative mRNA-splicing and subsequent translation through different initiation signals so as to yield distinct lengths of polypeptide isoforms. Amongst them, three of the most representatives are Nrf1α, Nrf1β and Nrf1γ, but the putative specific contribution of each isoform to regulating ARE-driven target genes remains unknown. To address this, we have herein established three cell lines on the base of the Flp-In T-REx system, which are allowed for the tetracycline-inducibly stable expression of Nrf1α, Nrf1β and Nrf1γ. Consequently, the RNA-Sequencing results have demonstrated that a vast majority of differentially expressed genes (i.e. >90% DEGs detected) were dominantly up-regulated by Nrf1α and/or Nrf1β following induction by tetracycline. By contrast, the other DEGs regulated by Nrf1γ were far less than those regulated by Nrf1α/β (i.e. ~11% of Nrf1α and ~7% of Nrf1β). However, further transcriptomic analysis revealed that the tetracycline-induced expression of Nrf1γ significantly increased the percentage of down-regulated genes in total DEGs. These statistical data were further validated by quantitative real-time PCR. The experimental results indicate that distinct Nrf1 isoforms make diverse and even opposing contributions to regulating different subsets of target genes, such as those encoding 26S proteasomal subunits and others involved in various biological processes and functions. Collectively, Nrf1γ acts as a major dominant-negative inhibitor competitively against Nrf1α/β activity, such that a number of DEGs regulated by Nrf1α/β are counteracted by Nrf1γ.

Nuclear factor-erythroid 2-related factor 1 (Nrf1, also called Nfe2l1) acts as a transcription factor belonging to the cap'n'collar (CNC) basic-region leucine zipper (bZIP) family, which is indispensable for maintaining both cellular homoeostasis and organ integrity during normal development and growth, as well as the adaptative responses to other pathophysiological processes[1–3]. It is important to note that the unique function of Nrf1 is finely tuned by a steady-state balance between production of the CNC-bZIP protein (i.e. translation of transcripts) and concomitantly (i.e. post-transcriptional and post-translational) processing in order to give rise to distinct multiple isoforms (called proteoforms, with different and even opposing abilities) before being turned over. These distinct proteoforms of Nrf1 are postulated to together confer on the host robust cytoprotection against a vast variety of cellular stresses through coordinately regulating distinctive subsets of important homoeostatic and developmental genes. The transcriptional expression of such key genes are driven by antioxidant response elements (AREs) and/or other *cis*-regulating consensus sequences, some of which are conversed with the activating protein-1 (AP-1) binding site, within these gene promoter regions.

The single gene of *Nrf1* is allowed for differential transcriptional expression to yield multiple mRNA transcripts (between ~1.5 kb and ~5.8 kb) and subsequently alternative translation into distinct polypeptide isoforms (between ~25-kDa and ~140-kDa), which are determined to be differentially distributed in embryonic, fetal and adult tissues, including liver, brain, kidney, lung, heart, skeletal muscle, bone, testis, ovary, placenta and others[4–7]. Amongst such isoforms, the full-length Nrf1 (designated Nrf1α) is yielded by the first translation initiation signal within the main open reading frame (ORF) of alternatively spliced mRNA transcripts, in which the exon 4 [encoding the peptide ^{242}VPSGEDQTALSLEECLRLLEATCPFGENAE^{271}, that was named the Neh4L (Nrf2-ECH homology 4-like) region] is removed from its long isoform TCF11 (transcription factor 11) in the human[5]. Albeit Nrf1 lacks the Neh4L region, it was identified to retain a strong transactivation activity that is largely similar to the TCF11 ability[8].

The Laboratory of Cell Biochemistry and Topogenetic Regulation, College of Bioengineering and Faculty of Sciences, Chongqing University, No. 174 Shazheng Street, Shapingba District, Chongqing, 400044, China. Correspondence and requests for materials should be addressed to Y.Z. (email: yiguozhang@cqu.edu.cn)

By contrast with Nrf1α, the short isoform Nrf1β [which was early designated as LCR-F1 (locus control region-factor 1)] is determined to be generated through the in-frame translation that is initiated by an internal perfect Kozak' starting signal (5′-puCCATGG-3′), which is situated within and around the four methionine codons between positions 289 and 297 in the mouse[4,5,9]. By bioinformatic analysis, it is thus inferred that Nrf1β lacks the N-terminal domain (NTD, aa 1–124) and its adjacent acidic domain 1 (AD1, aa 125–296)[10,11]. Later, Nrf1β is also determined to exhibit a weak transactivation activity[6,12–14], but stimulation of Nrf1β activity appears to be dependent on distinct stressors that had been administrated in different cell lines[13–15]. Furthermore, a small dominant-negative isoform, called Nrf1γ[12,13], is produced by another potential in-frame translation starting at the putative methionine of position 584, as well as by the putative endoproteolytic processing of longer Nrf1 proteins. When generation of Nrf1γ is blocked, the transactivation activity of Nrf1β is significantly increased[12]. On the contrary, when Nrf1γ is forcedly expressed, the consequence enables it to make a possible interference with the functional assembly of each of the active transcription factors (i.e. Nrf1α or Nrf2) with its heterodimeric partner (i.e. sMaf and other bZIP proteins), in order to down-regulate expression of AP1-like ARE-driven target genes[12,13].

To date, it is, however, unknown how each isoform of Nrf1 contributes to its unique role in regulating expression of ARE-driven cytoprotective genes against various physiopathological stresses. To address this issue, we have herein created distinct three stable inducible expressing cell lines, i.e., HEK293C$^{Nrf1\alpha}$, HEK293D$^{Nrf1\beta}$ and HEK293E$^{Nrf1\gamma}$, which are allowed for stably tetracycline (Tet)-induced expression of human Nrf1α, Nrf1β, and Nrf1γ, respectively. The similarities and differences of structural domains of these three isoforms were shown diagrammatically (Fig. 1A). Subsequently, different changes in the transcriptomic expression mediated by Nrf1α, Nrf1β, or Nrf1γ were analyzed by RNA-Sequencing (RNA-Seq), some of which were further validated by quantitative PCR experiments. Collectively, we first discovered that both Nrf1α and Nrf1β make major dominant contributions to Nrf1-mediated transcriptional expression of cognate target genes. Thereby, a vast majority of differentially expressed genes (DEGs) are up-regulated by Nrf1α and/or Nrf1β upon either stable expression induced by Tet treatment. By sharp contrast, an array of similar or different DEGs regulated by Nrf1γ are far less than those genes regulated by Nrf1α or Nrf1β. This demonstrates that Nrf1γ exerts a distinguishable effect from the other two isoforms Nrf1α/β, on the transcriptional expression of Nrf1-target genes, because its inducible expression by Tet significantly increased the percentage of down-regulated genes among the DEGs detected. Overall, this work provides a further better understanding of distinctions in the transcriptional regulation of cognate Nrf1-target genes by its different isoforms.

Results

Identification of differentially expressed genes in distinct Nrf1 isoforms-expressing cell lines by RNA-Sequencing. To gain an in-depth insight into distinct contributions of human Nrf1 isoforms (i.e. Nrf1α, Nrf1β and Nrf1γ) to the precision regulation of different subsets of its target genes, each of these isoform-expressing cell lines was herein established by Flp recombinase-mediated integration on the base of the Flp-In T-REx-293 host cells. These cell lines had been transfected with each of pcDNA5/FRT/TO-V5 expression constructs for distinct cDNA sequences encoding Nrf1α, Nrf1β and Nrf1γ (as shown schematically in Fig. 1A) before being selected, whilst the empty expression vector-transfected host HEK293 cells served as an internal negative control (i.e. HEK293^{control}) in the parallel experiments. The inducible expression of Nrf1α, Nrf1β and Nrf1γ was under control by Tet for distinct experimental requirements, before which the positively-selected clones of cell lines were maintained within double antibiotics 150 μg/ml hygromycin B and 15 μg/ml blasticidin. The resulting selected cell lines were named HEK293^{control}, HEK293C$^{Nrf1\alpha}$, HEK293D$^{Nrf1\beta}$ and HEK293E$^{Nrf1\gamma}$, respectively. Thus they were also abbreviated to control, Nrf1α, Nrf1β, and Nrf1γ, as shown in the subsequent experimental results. Of note, the stable expression of distinct isoforms was monitored by interaction of Tet with its repressor TetR leading to release from the Tet operator and then induction of interested gene transcription.

As anticipated, after 12-h treatment of HEK293^{control}, HEK293C$^{Nrf1\alpha}$, HEK293D$^{Nrf1\beta}$ and HEK293E$^{Nrf1\gamma}$ cell lines with 1 μg/ml of Tet, they were harvested, followed by immunoblotting (Fig. 1B,C) and quantitative PCR (Fig. 1D–F) to determine their protein and mRNA levels, respectively. The immunoblotting with antibodies against Nrf1 and its C-terminal V5 tag showed that the inducible expression of Nrf1α, Nrf1β and Nrf1γ was validated by treatment of Tet in their respective cell lines. Notably, Nrf1α was exhibited as two major close isoforms of between ~140-kDa and ~130-kDa (i.e. glycoprotein and deglycoprotein) alongside with additional two minor processed isoforms of between ~100-kDa and ~90-kDa (Fig. 1B,C). Nrf1β displayed as a major protein of ~70-kDa, together with another minor processed polypeptide of ~50-kDa (called Nrf1β2). By contrast, Nrf1γ expressed as a minor ~36-kDa protein, along with a major processed ~25-kDa polypeptide (also called Nrf1δ).

These Tet-inducibly isoform-expressed and negative control cell lines were subjected to extraction of total RNAs before being sequenced (i.e. RNA-Seq). An overview of the primary sequencing data was then depicted in Table S1. Following the processing and analysis of the RNA-Seq data, the gene expression levels were calculated by the RPKM (Reads Per kb per Million reads) method[16]. Of note, the DEGs were screened out, with the threshold of *P*-value in multiple tests at a false discovery rate (FDR) $\leq$0.001 (ref.[17]), along with an absolute value of Log_2 (fold change) $\geq$1, and identified by calculating each gene expression in sample groups versus controls as indicated in the Stat Chart (Fig. 1G). By comparison to the control cells, 1001 genes showed significant changes (Table S2) in their transcriptional expression as accompanied by stable Nrf1α-expression in Tet-inducible HEK293C$^{Nrf1\alpha}$ cells, of which 957 genes (i.e. 95.6%) were up-regulated (Fig. 1G, *red bar*) but the other 44 genes (i.e. 4.4%) were down-regulated (*green bar*). And, transcriptional expression of as many as 1675 genes was also significantly regulated upon stable expression of Nrf1β in the Tet-inducible HEK293D$^{Nrf1\beta}$ cells, 94.8% of which (i.e. 1588 genes) were up-regulated and the other 5.2% (i.e. 87 genes) were down-regulated (Fig. 1G, and also Table S3). By striking contrast, only 113 genes were significantly transcriptionally changed in the stable Tet-inducibly Nrf1γ-expressing HEK293E$^{Nrf1\gamma}$ cells. This appeared to be just about one-tenth of the number of those genes measured in either HEK293C$^{Nrf1\alpha}$ or HEK293D$^{Nrf1\beta}$ cells. Amongst the 113 genes regulated by Nrf1γ, besides only 27 genes (i.e.

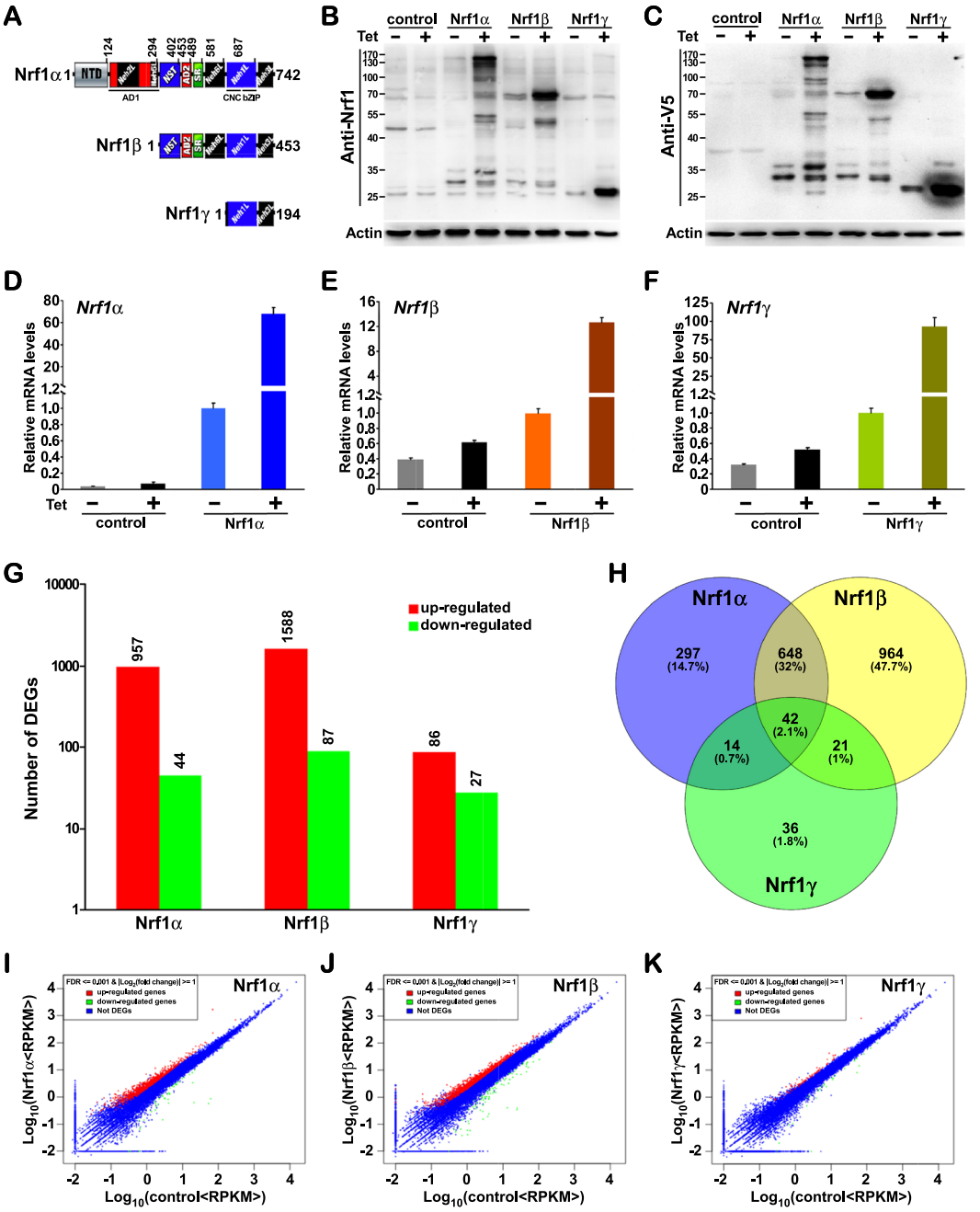

Figure 1. Identification of distinct Nrf1 isoforms-expressing cell lines along with global statistical analysis of RNA-Seq data. (**A**) Shows schematics of the similarities and differences between structural domains of Nrf1α, Nrf1β, and Nrf1γ. (**B**,**C**) Total lysates of each cell lines that had been treated with 1 μg/ml of tetracycline (+) or not (−) were subjected to protein separation by 12% Laemmli SDS-PAGE gels running in the pH 8.9 Tris-glycine buffer, and visualization by immunoblotting with Nrf1 or V5 antibody to identify Tet-inducible expression of distinct Nrf1 isoforms. (**D**–**F**) These cell lines were further identified by quantitative real-time PCR. (**G**) Shows an overview of differentially expressed genes (DEGs) indicated in the stat chart of RNA-seq data. (**H**) Shows the Venn chart with the common or unique DEGs among sequenced samples after being normalized by control samples. (**I**–**K**) Pairwise scatter plots are used to identify global changes and trends in gene expression among three conditions.

23.9%) were down-regulated, the other 86 genes (i.e. 76.1%) were still up-regulated (Fig. 1G, and Table S4). This implies a possible counteraction of Nrf1γ competitively against Nrf1α/β down-regulation of some genes.

After normalization by the control, each isoform-specific or their common DEGs amongst three Nrf1 isoforms were statistically shown with the Venn diagram (Fig. 1H). Their scatterplots were useful to identify global changes with distinct trends in gene expression between pairs of conditions (Fig. 1I–K). Overall, these data demonstrate that Nrf1β regulates the greatest number of DEGs amongst these three isoforms, followed by Nrf1α, whereas the number of DEGs regulated by Nrf1γ is far less than that of DEGs regulated by either of the former two isoforms.

Functional annotation of differentially expressed genes regulated by each of distinct Nrf1 isoforms. Further data analysis of thousands of DEGs involved in distinct biological processes is an important downstream task following RNA-Seq to understand relevant meanings of those 'interested' genes regulated by each isoform of Nrf1. Of note, the gene ontology (GO) is an internationally-standardized functional classification system of genes, which offers a dynamically-updated controllable vocabulary with a strictly defined concept to describe comprehensively properties of distinct genes and their products in any organisms, and also covers three major domains, including cellular component, molecular function and biological process. The enrichment analysis of GO provides all relevant terms that are significantly enriched in the DEGs, by comparison to the genomic background, and then filters the DEGs that correspond to potential biological functions. The *d*atabase for *a*nnotation, *v*isualization and *i*ntegrated *d*iscovery (DAVID) of bioinformatic resources consists of an integrated biological knowledgebase with analytic tools, which is used for systematic extraction of biological features and/or meanings associated with large lists of genes[18]. Therefore, in order to investigate the relationship and difference in gene regulation by amongst these three Nrf1 isoforms, the enriched functional annotation of GO terms (Tables S5–S7) was mapped with the DEGs regulated by each isoform. The involved DEGs in different terms were identified by the DAVID tool and listed according to their enrichment *P*-values. Those genes are then enabled for an interaction with each other to play some key roles in the certain biological functions. Thus, the results of pathway enrichment analysis (Tables S8–S10) was obtained on the base of the major public pathway-related database KEGG (Kyoto Encyclopedia of Genes and Genomes[19]). This is helpful to further understand biological functions of distinct subsets of genes regulated by each Nrf1 isoform.

The top 20 of significantly enriched functional annotation terms (Fig. 2) are represented by different GO classes and KEGG pathways. Within the GO-classified biological processes, the Nrf1α-regulated DEGs are highly enriched in terms of transcriptional regulation, RNA metabolism, cell cycle, macromolecule catabolic process, cellular response to stress, protein localization, metabolic process and chromosome organization (Fig. 2A). Almost all these enriched biological process terms of Nrf1α-regulated DEGs were also significantly enriched in other DEGs regulated by Nrf1β. Besides similar terms, Nrf1β-regulated DEGs were also involved in intracellular signaling cascade, phosphorylation, transcriptional regulation of RNA polymerase-II promoter and biosynthetic process (Fig. 2E). By contrast, just the fewer numbers of DEGs were regulated by Nrf1γ, and hence only a few of genes were mapped to distinct groups of biological processes, such as regulation of cell proliferation, response to wounding, protein metabolic process, blood vessel development, peptide cross-linking and plasma membrane long-chain fatty acid transport (Fig. 2I). As for some cellular components, those DEGs regulated by Nrf1α and Nrf1β are also highly enriched in terms of non-membrane-bounded organelles, membrane-enclosed lumen, organellic lumen, juxtanuclear lumen, nucleoplasm, nucleolus, cytosol, cytoskeletal part, Golgi apparatus, endomembrane system and extrinsic to membrane (Fig. 2B,F). These DEGs were also highly enriched for ion-binding, nucleotide-binding, ATP-binding and transcription regulator activity in molecular function groups (Fig. 2C,G). Notably, there are more about 30% of DEGs regulated by Nrf1β (versus control) mapped to each of the same GO term as that regulated by Nrf1α, implying that both regulate distinct subsets of more Nrf1-target genes. By sharp contrast, only a few of genes regulated by Nrf1γ were mapped to couple cellular component groups, like non-membrane-bounded, organelle, membrane-enclosed lumen and nucleolus (Fig. 2J). Meanwhile, the Nrf1γ-regulated DEGs were also not significantly enriched in molecular function terms (Fig. 2K). As such, not all of the GO terms mapped with Nrf1γ-regulated DEGs also appear in the former two groups, implying that several genes regulated by Nrf1γ are distinguishable from those regulated by Nrf1α/β. The enriched GO terms which are associated specifically to each isoform, were sorted in Sheet 2 in the Supplementary Tables S5–S7.

Furtherly, the results obtained from relevant pathway enrichment analysis (Tables S8–S10) revealed that DEGs regulated by three Nrf1 isoforms are highly enriched in the terms of human diseases (i.e. dilated cardiomyopathy, hepatitis C, hypertrophic cardiomyopathy, measles, non-small cell lung cancer, pathways in cancer, pancreatic cancer, salmonella infection, small cell lung cancer, viral myocarditis), metabolism (e.g. alanine, aspartate and glutamate metabolism, biosynthesis of unsaturated fatty acids, inositol phosphate metabolism, lysine degradation, purine metabolism) and genetic information processing (e.g. DNA replication, homologous and non-homologous end-joining recombination, protein processing in the endoplasmic reticulum, ubiquitin-mediated proteolysis, RNA transport and degradation, and spliceosome). These DEGs are also involved in pathways of cellular processes (e.g. cell cycle, endocytosis, regulation of actin cytoskeleton and tight junction), organismal systems (e.g. axon guidance, NOD-like receptor signaling, T cell receptor signaling, vascular smooth muscle contraction) and environmental information processing (e.g. mTOR-mediated and TGF-β signaling pathways). Collectively, (Fig. 2D,H,L) showed the top 20 of statistical pathway enrichment within DEGs regulated by each isoform of Nrf1. Of note, Nrf1 is also essential for maintaining cellular homoeostasis and organ integrity in multifunctional responses to a variety of endogenous and exogenous stimulators during normal development and growth. Thus, the number of DEGs was further calculated and thus mapped to distinct pathways responsible for carbohydrate metabolism, lipid metabolism, cancers, signal transduction, proteasome and redox signaling (Fig. 2M). These results demonstrate distinct Nrf1 isoform-specific regulation of different subsets of target genes, which are diversely involved in multiple physio-pathological processes. Further functional annotation also indicates that three Nrf1 isoforms have many similarities in exerting certain biological functions, but there still exist a few of differences amongst these three isoforms. They are thus required for coordinated cooperation together in a given organism to play more similarly overlapping, but also different and even opposing, biological functions of Nrf1. For example, Nrf1α and Nrf1β, but not Nrf1γ, may make main contributions to the physiological function of Nrf1, whilst Nrf1γ acts, at least in part, as a dominant-negative inhibitor of Nrf1.

Distinct expression profiles of the top DEGs regulated by Nrf1α, Nrf1β and Nrf1γ. To investigate distinctions in the target gene expression patterns regulated by between Nrf1α, Nrf1β and Nrf1γ, the differentially expressed data of each isoform-specific and/or their sharing top 10 of RPKM DEGs were extracted

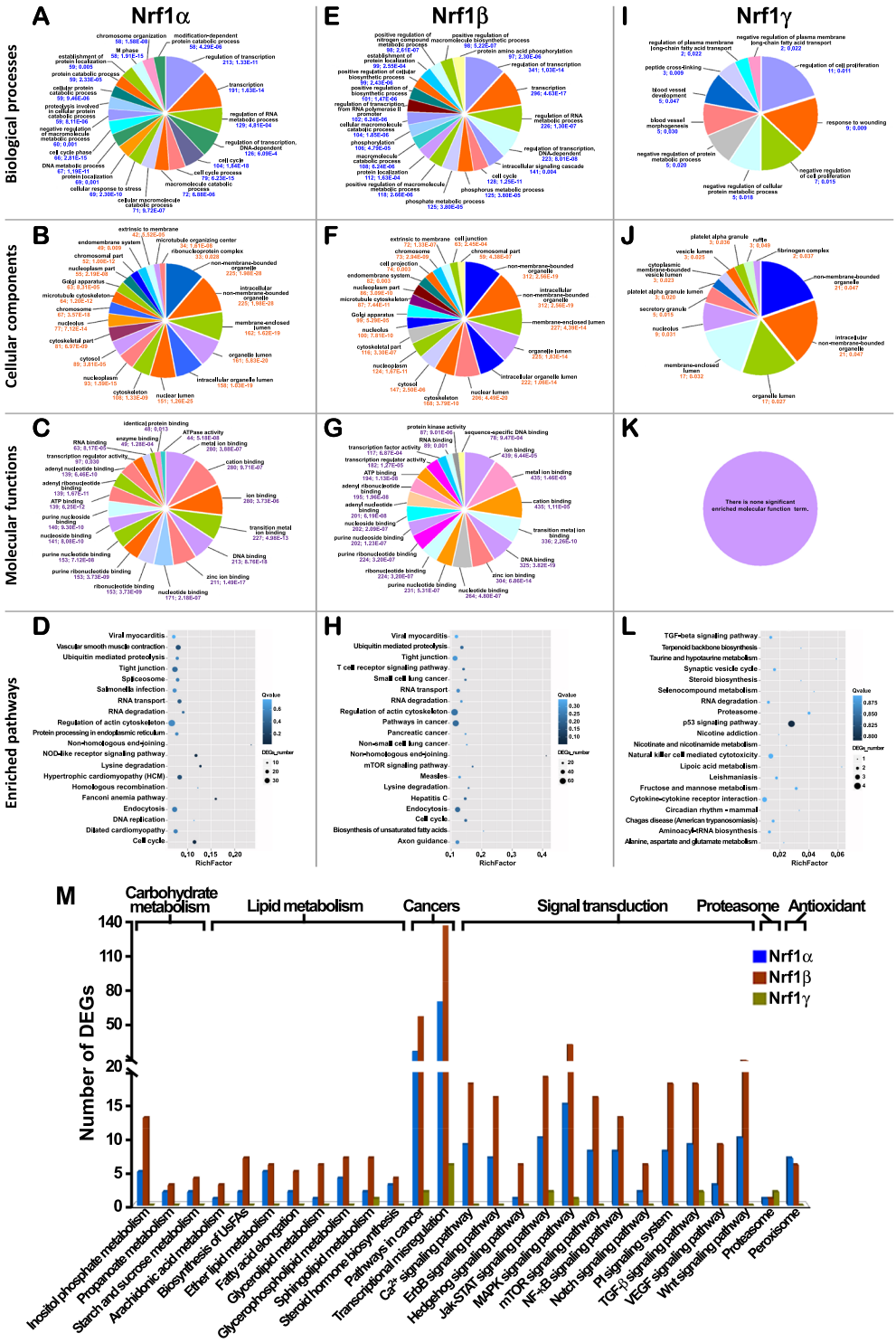

Figure 2. Functional annotation of DEGs regulated by three Nrf1 isoforms. By GO analysis of Nrf1α-regulated DEGs, multilevel distribution is shown for the GO categories: "biological process" (**A**), "cellular component" (**B**) and "molecular function" (**C**). Similar categories were shown in the cases of Nrf1β-regulated DEGs (**E**–**G**) and Nrf1γ-regulated DEGs (**I**–**K**). The significant enriched GO terms presented in the pie charts are ranked by the number of DEGs, and the numerical values below each term are represented by the number of DEGs and *P*-value. The top-ranked 20 pathways enriched according to the *Q* value were plotted for Nrf1α-regulated DEGs (**D**), Nrf1β-regulated DEGs (**H**), and Nrf1γ-regulated DEGs (**L**). In the scatter plots, the rich factor is a ratio of DEGs' number annotated in one pathway term to all genes' number annotated in this pathway term. The *Q* value is corrected *P*-value ranging from 0 to 1, and less *Q* value means greater intensiveness. Furthermore, the number of DEGs mapped to the pathways, that were associated with carbohydrate metabolism, lipid metabolism, cancers, signal transduction, proteasome and redox signaling, was calculated in the stat chart (**M**).

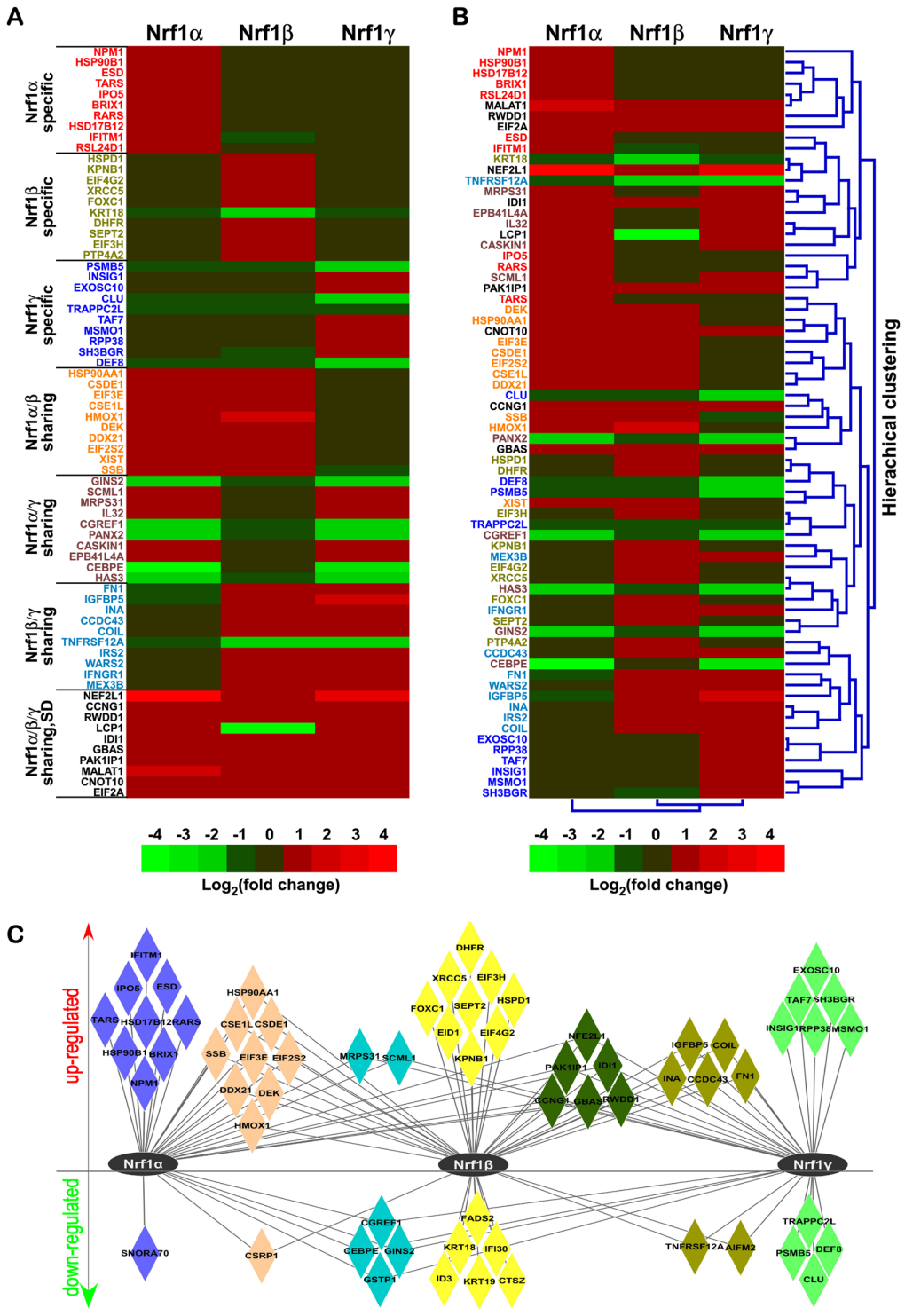

Figure 3. Similar but different expression patterns of top-ranked DEGs under regulated by three Nrf1 isoforms. (**A**,**B**) The heatmaps with distinct clusters were created for specific and/or sharing top 10 DEGs regulated by different isoforms of Nrf1. Differences in expression of distinct genes are shown in different colour backgrounds as scaled according to the Log_2 (fold changes). SD (single difference) in *LCP1*. (**C**) The network of the unique and/or common top-ranked DEGs amongst three sample groups after being normalized by control values.

(Table S11). These unique and/or common DEGs were shown in the heatmap (Fig. 3A) and also clustered by HemI[20] (Fig. 3B). According to both the RPKM values and fold changes in gene expression, the unique and/or some common top-ranked DEGs were picked out to build a gene-regulatory network among three Nrf1 isoforms (Fig. 3C, and Table S12). As shown in the heatmap, all the top 10 RPKM Nrf1α-specific genes showed up-regulation, whilst 9 of the top 10 RPKM Nrf1β-specific genes were up-regulated, but with only one gene (i.e.

KRT18 (keratin 18)) being down-regulated by Nrf1β. By contrast, 6 of the top 10 RPKM Nrf1γ-specific genes were up-regulated, whereas the other 4 genes were down-regulated by this isoform (Fig. 3A). The top 10 RPKM Nrf1α/β-shared regulatory genes were up-regulated by both Nrf1α and Nrf1β, but not by Nrf1γ, whilst the top 10 RPKM Nrf1α/γ-coordinated genes were up- and down-regulated at a 5:5 ratios by Nrf1α and Nrf1γ rather than Nrf1β (Fig. 3A). In addition, 9 of the top 10 RPKM Nrf1β/γ-shared regulatory genes showed up-regulation, but only one (i.e. *TNFRSF12A* (tumor necrosis factor receptor superfamily, member 12A)) was down-regulated. Another 9 of the top 10 RPKM genes were commonly up-regulated by three Nrf1 isoforms, with only an exception of *LCP1* (lymphocyte cytosolic protein 1) that was significantly down-regulated by Nrf1β but up-regulated by the other two isoforms (Fig. 3A). Collectively, the global statistical results of DEGs (as described above in Fig. 1F) also revealed that Nrf1β specifically regulated a maximum number of DEGs, when compared to those regulated by Nrf1α or Nrf1γ. However, Nrf1α contributed to a maximum ratio of the up-regulated to total DEGs, whereas Nrf1γ was rather attributable to a maximum ratio of the down-regulated to total DEGs.

Interaction networks of Nrf1-related genes upon expression of its different isoforms. Since typical functions of genes are exerted in critical biological process through interaction networks, study of gene family interaction networks are useful to investigate potential gene functions[21]. As illustrated in Fig. 4A, two Nrf1-mediated networks were constructed, in which some interactive proteins were identified by BioGRID (*upper three panels*)[22] and STRING (*lower three panels*)[23], respectively. Distinct expression profiles of these putative genes involved in the networks were extracted from distinct RNA-Seq data sets, which were reflected with different gradient colors in accordance with fold changes (Fig. 4A, and Tables S13 and S14). In the network identified with the BioGRID database, 12 of 19 genes showed a similar regulation trend, whereas other 7 genes displayed different regulation trends among three Nrf1 isoforms. Within these different regulation trends, three genes *MAFG* (v-maf musculoaponeurotic fibrosarcoma oncogene homolog G), *BRD9* (bromodomain containing 9) and *FBXW7* (F-box and WD repeat domain containing 7) were differentially regulated by Nrf1γ from other two isoforms, another three genes *RUSC2* (RUN and SH3 domain containing 2), *CAPN1* (calpain 1) and *HCFC1* (host cell factor C1) were differentially regulated by Nrf1β, the remaining one gene *C8ORF33* (chromosome 8 open reading frame 33) was differentially regulated by Nrf1α. By comparison of the network identified with the STRING database, 5 of 11 genes showed an uniform regulation trend, whilst the other 6 genes were differently regulated: i) three genes *MAFG*, *NRF1* (nuclear respiratory factor 1) and *FBXW7* were differentially regulated by Nrf1γ from other two isoforms; ii) two genes *C3ORF35* (chromosome 3 open reading frame 35) and *MAFK* (v-maf musculoaponeurotic fibrosarcoma oncogene homolog K) were differentially regulated by Nrf1β; and iii) only one gene *SP7* (Sp7 transcription factor) was differentially regulated by Nrf1α. Collectively, the interaction network analysis indicates that three isoforms of Nrf1 could diversely regulate its target genes. Such changes in the expression abundance of one isoform may also influence its overall transcriptional regulation of Nrf1-target genes. In fact, transcriptional expression of different subsets of target genes (driven by AP1-like AREs) was principally attributable to the precision regulation by distinct functional heterodimers of CNC-bZIP family members (i.e. Nrf1, Nrf2, Nrf3, Bach1 and Bach2) with small Maf or other bZIP factors[24]. Such transcription of these factors was also further compared to determine distinct Nrf1-specific effects (Fig. 4B, and Table S15).

Intriguingly, the N-terminal Neh2 (Nrf2–ECH homology 2) domain of Nrf2, which contains a redox-sensitive Keap1 (Kelch-like ECH-associated protein 1)-binding degron targeting the homoeostatic CNC-bZIP protein to ubiquitin ligase cullin-3 and Rbx1-dependent proteasomal degradation pathway[25–28], is represented by the Neh2-like region within Nrf1α[24], but the latter CNC-bZIP protein is not regulated by Keap1 (ref.[10]). This notion also appeared to be further supported by the observation that both the mRNA levels (Fig. 4C) and protein levels (Fig. 4D,E) of Keap1 were differentially expressed under distinct genetic backgrounds of knockout (i.e. *Nrf1α*$^{-/-}$ or *Nrf2*$^{-/-}$), siRNA (i.e. *siNrf2*) or constructive activation of Nrf2 mutant (i.e. *caNrf2*$^{\Delta N}$). In turn, these results also unravel that significant changes in the expression of Keap1 are accompanied by altered expression of Nrf2, but not Nrf1α. This implies a feedback circuit existing between Nrf2 (rather than Nrf1) and its negative regulator Keap1.

Distinct effects of Nrf1 isoforms on proteasomal expression in response to stimulation by proteasomes inhibitor. The ubiquitin-proteasome system (UPS) is crucial for eukaryotic cells to adjust its capacity of protein degradation to changing proteolytic requirements, because these proteins are marked with polyubiquitin chains targeting for their degradation by the 26S proteasome, an ATP-dependent complex consisting of the 20S proteolytic particle capped by one or two of the 19S regulatory particles[29]. Importantly, coordinated regulation of proteasomes by Nrf1, but not Nrf2, occurs through proteasome-limited proteolytic processing of the former CNC-bZIP protein into a mature active factor to mediate Nrf1-target proteasomal gene expression in the 'bounce-back' response to relative lower doses of proteasomal inhibitors[30–33]. Nrf1-mediated induction of proteasomes subunits results in significant increases in mRNA expression levels of all proteasomal subunits only upon exposure to lower concentrations of proteasomal inhibitors, but this feedback compensatory response is prevented by high concentrations of proteasomal inhibitors[33,34]. Therefore, there exists a bidirectional regulatory feedback circuit between Nrf1 and the proteasome[24,31–35]. However, our data revealed that almost none of proteasomal subunits were differentially expressed at basal levels, because they appeared to be unaffected by stably-induced expression of any one of Nrf1 isoforms (Fig. 5A, and Table S16).

To address this, we validated Nrf1-stimulated induction of proteasomal subunits by its inhibitors. Experimental cells that had pretreated with 1 μg/ml of Tet alone or plus a low (0.01 μmol/L) or a high (10 μmol/L) concentration of bortezomib (BTZ) for 16 h were subjected to further determination of mRNA expression levels of some proteasomal subunits, including *PSMA1*, *PSMA4*, *PSMB7*, *PSMC2* and *PSMD12* by real-time qPCR (Fig. 5B–G). As expected, the results demonstrated that all the mRNA levels of these proteasomes examined were increased following exposure of Nrf1α- or Nrf1β-expressing cells to 0.01 μmol/L of BTZ. Conversely, such the increased expression was significantly prevented by additional exposure of Nrf1α- or Nrf1β-expressing cells to

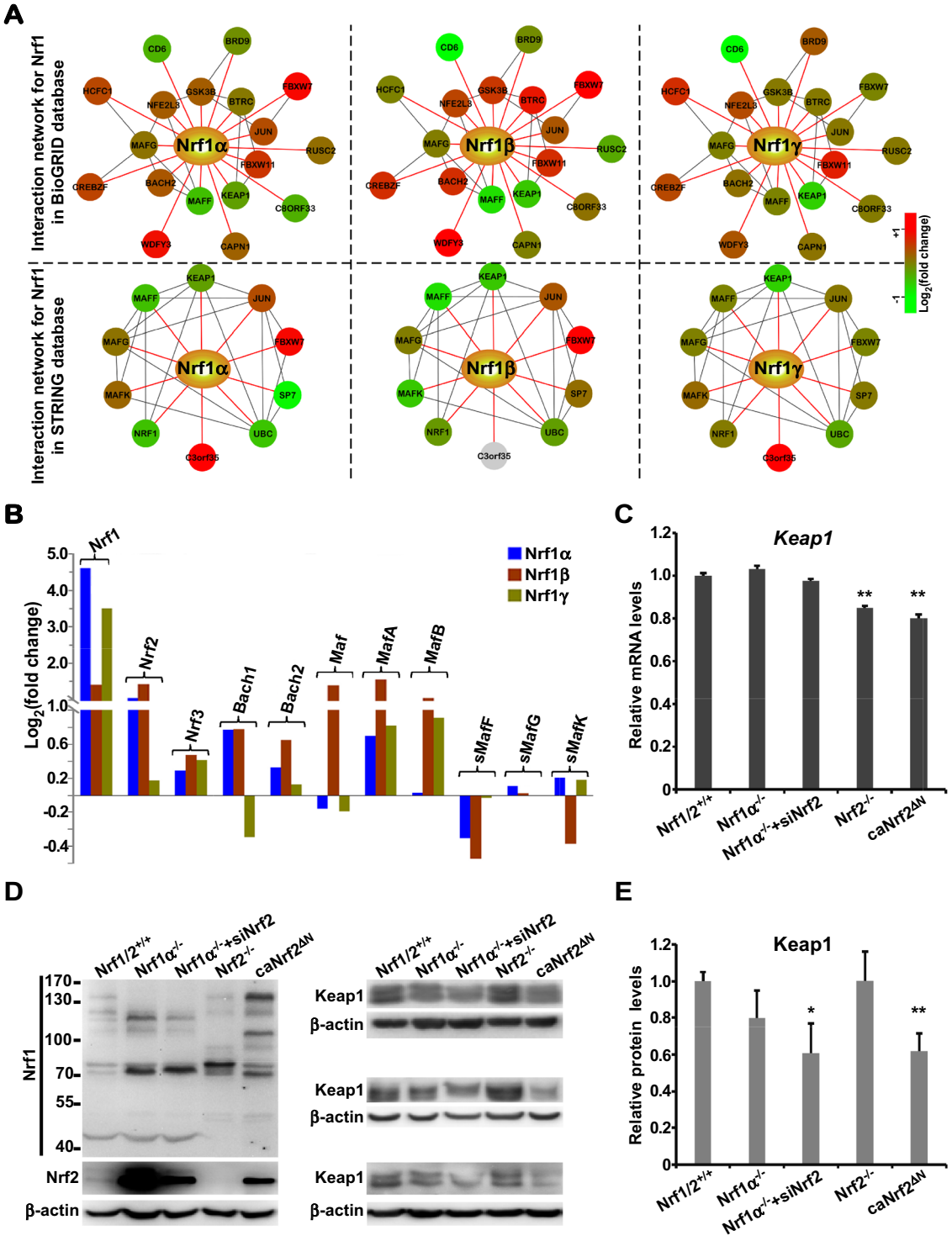

Figure 4. Nrf1-related genes in interaction networks and their expression regulated by different Nrf1 isoforms. (**A**) Two interaction networks for Nrf1 were established on the base of the BioGRID database and STRING database. The interactor nodes showing up-regulation or down-regulation were marked in red or green, respectively. Such genes were marked with gradient colours from green to red according to the Log_2 (fold changes), but gray indicates that a gene was not expressed. (**B**) The expression comparison of Nrf1, Nrf2 and Nrf3, Bach1 and Bach2, together with small Maf family members. (**C**–**E**) Expression of Keap1 (Kelch-like ECH-associated protein 1) at its mRNA levels (**C**) and protein levels (**D**,**E**) was determined in the presence or absence of Nrf1α and/or Nrf2. The expression of Nrf1α and Nrf2 at protein levels was shown in different gels at the same amount of samples (the left side of **D**). Then the relative expression of Keap1 at its protein levels (**E**) was calculated by the mean fold change relative to control based on the grayscale values of three independent experiments (the right side of **D**). The error bars indicate mean ± SD (n = 3), with significant decreases (*$p < 0.05$, **$p < 0.01$) being calculated relative to the corresponding control values.

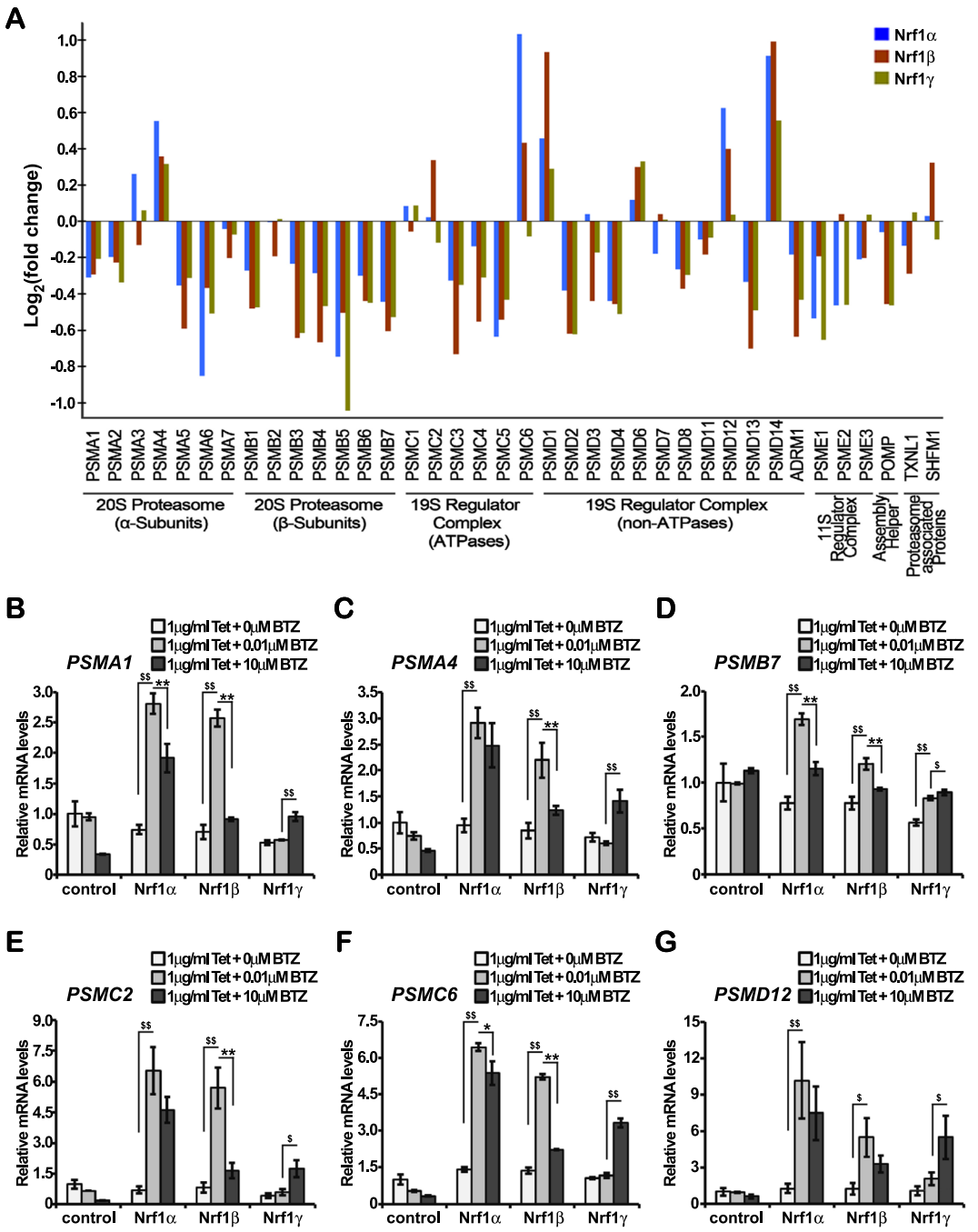

Figure 5. Distinct effects of Nrf1 isoforms on proteasomal expression in distinct responses to its inhibitor. (**A**) The RNA-Seq levels of proteasomal subunit genes were selected for distinct isoform-specific regulation relative to control values. (**B**–**G**) When compared with experimental controls, Nrf1α-, Nrf1β- and Nrf1γ-expressing cells were treated with1 μg/ml of Tet alone or plus a lower (0.01 μmol/L) or higer (10 μmol/L) doses of bortezomib (BTZ) for 16 h. The regulatory effects of Nrf1 isoforms, along with distinct concentrations of BTZ, on 26S proteasomal gene transcription were then analyzed by real-time qPCR. The data are represented by a mean ± SD of three independent experiments, with significant increases ($p < 0.05, $$p < 0.01) and significant decreases (*p < 0.05, **p < 0.01) being determined, relative to their corresponding control values.

10 μmol/L of BTZ. Intriguingly, the parallel experimentations of Nrf1γ-expressing cells revealed that almost no increases in the BTZ-stimulated expression of those aforementioned proteasomes except *PSMB7* were detected, even upon exposure to the low concentration at 0.01 μmol/L of BTZ, as consistent with the notion that Nrf1γ is likely to act as a dominant-negative inhibitor of Nrf1. However, it is full of curiosity about the finding that the high concentration (10 μmol/L) of BTZ enabled for significant stimulation of Nrf1γ to increase mRNA expression levels of all the examined proteasomal subunits, albeit the detailed mechanism(s) remains to be further explored.

Different regulatory effects of distinct Nrf1 isoforms on the downstream target genes. To confirm the conclusion drawn from the RNA-Seq data, some downstream genes of Nrf1 were selected for further experimental validation by qRT-PCR analysis (Fig. 6). Such the known Nrf1-target genes included those encoding HO1 (heme oxygenase 1)[36–38], GCLM (glutamate-cysteine ligase, modifier subunit), GCLC (glutamate-cysteine ligase, catalytic subunit)[39–42], MT1E (metallothionein 1E)[43], PGC-1β (peroxisome proliferator-activated receptor gamma, coactivator 1 beta)[44] and LPIN1 (lipin1)[45]. As expected, the experimental results revealed that these Nrf1-target genes exhibited different trends to be expressed specifically in distinct isoform-expressing stable cells (Fig. 6D–H), each of which is distinctive from those measured in the other two isoforms-expressing cell lines, but with an exception that *HO1* was up-regulated by all three isoforms (Fig. 6C).

Next, other ARE-driven downstream genes were also selected, to be potentially regulated by Nrf1, based on the RNA-Seq data according to RPKM and Log_2 (fold change), and then determined by real-time qPCR. As shown in (Fig. 6I–L), those genes encoding NPM1 (nucleophosmin 1), ESD (esterase D), IPO5 (importin 5) and IFITM1 (interferon induced transmembrane protein 1) were differentially up-regulated by Nrf1α. They were also significantly increased by Nrf1β, albeit their increased levels appeared to be less than those regulated by Nrf1α. It is, therefore, inferred that Nrf1β is not a dominant-negative inhibitor competitively against Nrf1α, this is consistent with the notion previously reported by our group[12,46–48]. Furthermore, it is, to our surprise, that mRNA levels of *ESD* and *IPO5* were significantly increased by Nrf1γ (Fig. 6J,K), even though their fold changes were less than those measured from the other two cases of Nrf1α and Nrf1β.

The additional four genes, such as *HSPD1* (heat shock 60 kDa protein 1), *KPNB1* (karyopherin subunit beta 1), *FOXC1* (forkhead box C1) and *ELOVL5* (ELOVL fatty acid elongase 5) were differentially up-regulated by Nrf1β, at less increased levels than equivalents regulated by Nrf1α and Nrf1γ (Fig. 6M–P). Moreover, both *ID3* (inhibitor of DNA binding 3, dominant negative helix-loop-helix) and *KRT19* (keratin 19) were differentially down-regulated by Nrf1β (Fig. 6Q,R), whilst *KRT19* was up-regulated by Nrf1γ. Lastly, the gene *TRAPPC2L* (trafficking protein particle complex 2-like) was down-regulated by Nrf1α and Nrf1γ, but up-regulated by Nrf1β (Fig. 6S). Notably, these genes are all known to exert their respective biological functions and be involved in different pathways (see Tables S5–S10). Further bioinformatic search demonstrated that all the above-described putative Nrf1-target genes contains more than one of the highly-conserved ARE motifs and/or its reversed sequences (Fig. 6A,B). Thus, it is inferred that these ARE-driven genes are likely to be differentially regulated by distinct Nrf1 isoforms, but this warrants further experiments to elucidate which ARE motifs are functionally responsible for an Nrf1-specific isoform to regulate a given gene.

Discussion

Differential expression of different subsets of cognate genes is dependent on their enhancers and promoter regions containing distinct *cis*-regulatory consensus sequences (e.g. AP-1-like AREs)[24]. The transcriptional expression of AREs-driven genes is thus determined by differential recruitment of Nrf1, Nrf2 and Nrf3, in different combinations with each of their heterodimeric partners (e.g. sMaf, c-Jun, JunD or c-Fos), to target gene promoters. Of note, Nrf1 and Nrf2 are two important CNC-bZIP transcription factors expressed ubiquitously in various vertebrate tissues and thus elicit their putative combinational or competitive functions. Relative to the well-known water-soluble Nrf2, less attention has hitherto been drawn to the membrane-bound Nrf1[24]. However, major discoveries that had been made in the past twenty-five years have convincingly revealed that Nrf1, but not Nrf2, has been shown to be indispensable for maintaining cellular homoeostasis and organ integrity during normal development and healthy growth, as well as a vast variety of other patho-physiological processes. Importantly, several significant pathological phenotypes were developed in different transgenic mice (expressing distinct mutants of loss-of-function of Nrf1), including embryonic lethality, fetal anemia, lipid metabolic disorder, obesity, fatty liver, NASH, liver cancer, neurodegenerative diseases, hyperinsulinemia, diabetes, Warbug effect with high glycolysis. In addition, other mice expressing gain-of-function mutants of Nrf1 displayed glucose metabolic disorder, insulin resistance, diabetes and reduced body-weight. Thus it is inferable that the functional activity of Nrf1 is finely tuned, to a robust homeostatic extent, by a steady-state balance between its production and the concomitant processing into distinct isoforms before being turned over, which are together coordinated to confer on the host cytoprotection against a variety of cellular stresses.

Accumulating evidence has also unraveled that over eleven of distinct Nrf1 isoforms are produced from the single *Nfe2l1/Nrf1* gene, though differentially expressed, in different mammalian species. These isoforms are synthesized by translation through distinct initiation signals (i.e. the first or internal start ATG codons) embedded in different lengths of open reading frames, portions of which can be alternatively spliced from intact or longer transcripts. The resulting variations in the abundance of each isoform may not only influence the whole transcriptional functions of Nrf1 to regulate distinct subsets of cognate target genes and also contribute to the nuance in between distinct pathological phenotypes[24]. Therefore, it is of crucial important to determine differences in transcriptional regulation of cognate genes mediated by each Nrf1 isoform. Although this is hard, our present study has identified differential expression profiles of distinct target genes regulated by Nrf1α, Nrf1β and Nrf1γ alone or in their cooperation respectively. Here, we have further determined differences in the transcriptional regulation of Nrf1-target genes by between each Nrf1 isoforms. Notably, Nrf1α and Nrf1β are two major isoforms contributing to the main Nrf1-mediated transcription of downstream genes at RNA levels, such that a vast majority of differentially expressed genes are up-regulated by Tet-inducible expression of these two isoforms. On the contrary, stably Tet-inducible expression of Nrf1γ as a putative dominant-negative inhibitor is likely to interfere with the putative functional assembly of active transcription factors (Nrf1α, Nrf1β, and even Nrf2), leading to down-regulation of several key genes, some of which are up-regulated by Nrf1α and Nrf1β. These findings are consistent with our previous reports[12,46–48]. Collectively, these findings are very helpful to elucidate which isoforms of Nrf1 contribute to different transcription of distinct subsets of target genes that are involved in those significant pathological phenotypes. Thus, this study has provided three cell models to facilitate the future

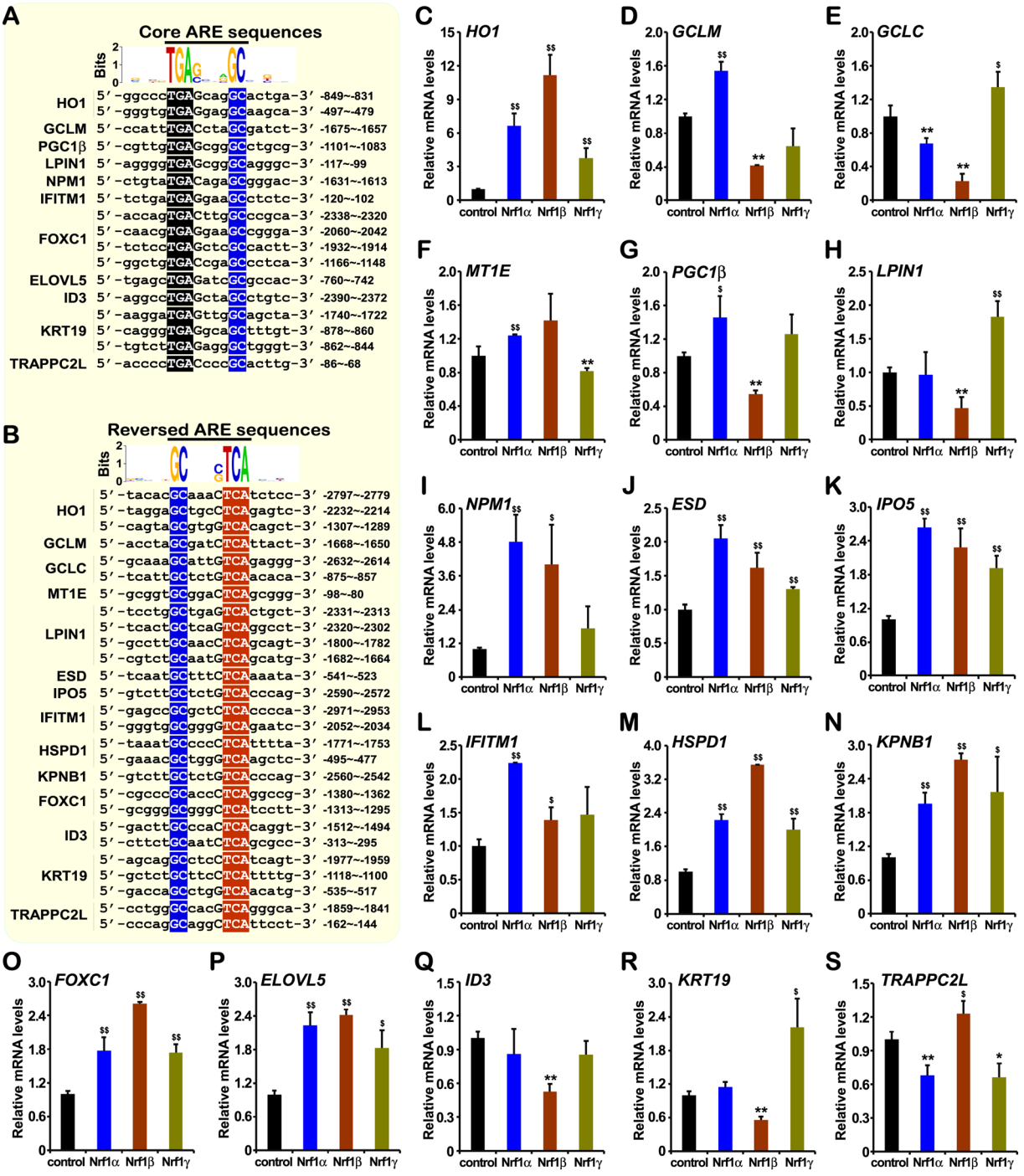

Figure 6. Distinct isoform-specific regulation of Nrf1-target ARE-driven genes. (**A**,**B**) The core and reversed ARE sequences are encompassed in the promotor regions of these Nrf1-regulated genes. When compared with experimental controls, Nrf1α-, Nrf1β- and Nrf1γ-expressing cells were induced with 1 μg/ml of Tet for 12 h. (**C**–**H**) Expression of those known Nrf1-target downstream genes, such as *HO1, GCLC, GCLM, MT1E, PGC1*β and *LPIN1*, was determined by real-time qPCR. (**I**–**S**) Expression of other potential downstream genes of Nrf1 that were selected on the base of RNA-Seq data was further validated by real-time qPCR. The data are represented by a mean ± SD of three independent experiments, with significant increases ($p < 0.05, $$p < 0.01) and significant decreases (*p < 0.05, **p < 0.01) being indicated, relative to their corresponding controls.

development of Nrf1 isoform-specific targets for chemoprevention against relevant diseases (i.e. cancer, neurodegenerative diseases, and diabetes)[49].

Furthermore, it should also be noted that we have presented most of the data about Nrf1γ in the present study to further support our previous notion[12,46–48], which revealed that this low molecular weight isoform acts as a dominant-negative inhibitor of Nrf1 competitively against the functional heterodimeric assembly of either an

active transcription factor (i.e. Nrf1α, Nrf1β, even Nrf2 and Nrf3) or another homologous trans-repressor (i.e. Bach1 and Bach2) with one of their cognate partner sMaf or other bZIP proteins (e.g. c-Jun, c-Fos, or ATF4). Therefore, it is plausible that either transactivation or transrepression of distinct subsets of similar and/or different target genes driven by AP1-like ARE/EpRE batteries depends possibly on a nuance between different temperospatially-assembled heterodimeric complexes of these CNC-bZIP factors with their partners. This is to say, understandably, that the dominant-negative form of Nrf1γ is much likely to counteract (or interfere with) the putative activity of its prototypic factors Nrf1α/β to transactivate or transrepress their downstream genes. For this reason, it is thus deduced that if some genes are down-regulated by Nrf1α/β, this down-regulation is abolished or even reversed to allow for their activation by Nrf1γ. In addition, there also exists an exception that a few number of Nrf1-target genes (as shown in Fig. 6) are positively regulated by Nrf1γ, seemingly as done by Nrf1α/β. It cannot be theoretically ruled out that Nrf1γ might also have a similar potential effect to that elicited by sMaf factors, which can still form a functional heterodemeric assembly with an activator or co-activator albeit it lacks a *bona fide* transctivation domain, but this remains to be further determined in the future experiments. For instance, it is known that some genes driven by AREs or its homologous consensus sequences are up-regulated by Bach1 or Bach2, although they also lack *de facto* transctivation domain and hence are also generally accepted as two major transrepressor amongst the CNC-bZIP family[50,51].

Materials and Methods

Chemicals and antibodies. All chemicals were of the highest quality commercially available. Hygromycin-B and blasticidin were purchased from Invitrogen Ltd, which served as double screening drugs to select the positive clones by final concentrations of 150 μg/ml and 15 μg/ml, respectively. The inducible reagent tetracycline hydrochloride was from Sangon Biotech Co (Shanghai, China) and used at a final concentration of 1 μg/ml. The proteasome inhibitor bortezomib was purchased from ApwxBio (USA). The antibody against endogenous Nrf1 proteins was acquired from our lab (i.e. Zhang's indicated in this study[52]), whereas antibody against KEAP1 was purchased from Sangon Biotech (Shanghai, China). Mouse monoclonal antibody against the V5 epitope was from Invitrogen Ltd, whilst anti-β-actin and secondary antibodies were obtained from Zhongshan Jinqiao Co (Beijing, China).

Expression constructs and cell culture. The cDNA fragments encoding three Nrf1 isoforms Nrf1α, Nrf1β and Nrf1γ were cloned into pcDNA5/FRT/TO-V5 expression vector, before being transfected into the host cell line Flp-In T-REx-293. The empty expression vector-transfected host cell served as a negative control. The positive isoform-expressing clones were selected by using 150 μg/ml hygromycin B and 15 μg/ml blasticidin and then stable expression of Nrf1α, Nrf1β and Nrf1γ was induced by addition of tetracycline. All cell lines used in this study were cultured in DMEM medium supplemented with 10% FBS and maintained in the 37 °C incubator with 5% CO_2.

Western blotting. Experimental cells were harvested in a denatured lysis buffer (2 mM Tris pH 7.5, 5 mM NaCl, 0.5 mM Na_2EDTA, 0.04 mM DTT, 0.5% SDS) containing 2 μg/ml protease inhibitor cocktail (Roche, Germany). The protein concentration of lysates was quantified by using BCA Protein Assay Reagent (Beyotime, Beijing, China). Equal amounts of protein prepared from cell lysates were loaded into each electrophoretic well so as to be separated by SDS-PAGE, followed by visualization by immunoblotting with each of indicated antibodies as described previously[53], and β-Actin was served as an internal control to verify amounts of proteins that were loaded in each well.

The RT-qPCR Analysis. Total RNAs were extracted from experimental cells by using an RNAsimple total RNA kit (Tiangen, Beijing, China). Then, 1.5 μg of total RNAs were used as a template for the subsequent synthesis of cDNA by using a RevertAid first strand cDNA synthesis kit (Thermo Fisher Scientific, USA). The resulting cDNA products (15 ng) served as the templates of quantitative real-time PCR within 5 μl of the MixGoTaq qPCR Master Mix (Promega, USA). Each of RT-qPCR with distinct pairs of primers (listed in the Supplementary Table S17) was performed in the following conditions: activation at 95 °C for 30 s, followed by 40 cycles of 10 s at 95 °C, and 30 s at 60 °C. These PCR reactions were carried out in at least 3 independent experiments that were each performed triplicate. The comparative Ct method was employed for quantification of indicated mRNA expression levels before being normalized to β-Actin.

Bioinformatics analysis of RNA-Seq data. Each isoform-expressing cell lines (i.e. Nrf1α, Nrf1β and Nrf1γ), along with the negative control cells, were allowed for growth in 6-well plates, before being induced by tetracycline (1 μg/ml) for 12 h. Such similar three independent experiments were conducted to prepare the samples on the same experimental conditions. Total RNAs were extracted by using an RNAsimple total RNA kit (Tiangen, Beijing, China) and the integrity of RNAs was checked by an Agilent Bioanalyzer 2100 system (Agilent technologies, Santa Clara, CA). For each sample, equal amounts of total RNAs from three independent experiments were pooled for RNA-Seq. Subsequently, RNA-Seq was carried out by Beijing Genomics Institute (BGI, Shenzhen, China) on an Illumina HiSeq 2000 sequencing system (Illumina, San Diego, CA) after the sample library products are ready for sequencing.

Before sequencing analysis, all mRNAs were fragmented into short fragments (about 200 bp) and the low-quality reads in which its percentage (base with a quality value ≤5) is greater than 50% were removed during data filtering. After examining the sequence quality and removing the "dirty" raw reads, which contain low quality reads and/or adaptor sequences, clean reads (the high quality sequences after data cleaning) were generated and stored as the FASTQ format[54]. Then, the clean reads were mapped to the reference human genome (GRCh37/hg19 from UCSC database) by using SOAP2 (ref.[55]), and gene expression levels were calculated by using the RPKM (Reads Per Kilobase of feature per Million mapped reads) method[16]. The expression of genes regulated by

each Nrf1 isoforms, relative to the control sample, was calculated as Log_2 (fold change), with a *P*-value calculated corresponding to the differential gene expression test and FDR (False Discovery Rate, which is a method to determine the threshold of *P*-value in multiple tests), and the differential expressed genes (DEGs) were further identified by the Poisson distribution model method (PossionDis)[56], which was developed referring to "The significance of digital gene expression profiles"[17,57] by BGI. Similar methods have also been applied in recent publications[58–60]. Both FDR $\leq$0.001 and the absolute value of Log_2 (fold change) $\geq$1 were herein taken as the threshold to identify differentially expressed genes. To give a better understanding of potential functions of the DEGs, both GO and KEGG pathway analysis were performed by using the online tools DAVID (https://david.ncifcrf.gov/) and KEGG (http://www.kegg.jp/) databases, respectively. In addition, the putative interaction networks of Nrf1-related genes were searched from the databases of BioGRID (https://thebiogrid.org/) and STRING (https://string-db.org/), before being annotated with sequencing results by the Cytoscape software[61]. The sequencing data have been submitted to NCBI SRA (PRJNA501789). For detailed descriptions, please see relevant supplementary contents at http://www.genomics.cn/en/index.

Statistical analysis. The 'wet' experimental data provided in this study were represented as the mean ± SD and were analyzed using the Student's *t*-test or Fisher's exact test, as appropriate. The resulting value of $p < 0.05$ was considered as a significant difference. In addition, statistical determination of the 'dry' sequencing analysis was aforementioned.

Data Availability

All the data that have been provided during this study are included in this publication with supplementary information documents. If interested, the datasets generated and/or analyzed during the present study are available from the corresponding author on reasonable request.

References

1. Sykiotis, G. P. & Bohmann, D. Stress-Activated Cap'n'collar Transcription Factors in Aging and Human Disease. *Sci Signal* **3**, https://doi.org/10.1126/scisignal.3112re3 (2010).
2. Gegotek, A. & Skrzydlewska, E. CNC proteins in physiology and pathology. *Postep Hig Med Dosw* **69**, 729–743 (2015).
3. Zhang, Y. *Molecular and cellular control of the Nrf1 transcription factor: An integral membrane glycoprotein*. (VDM Verlag Dr. Müller, 2009).
4. Chan, J. Y., Han, X. L. & Kan, Y. W. Cloning of Nrf1, an Nf-E2-Related Transcription Factor, by Genetic Selection inYeast. *P Natl Acad Sci USA* **90**, 11371–11375, https://doi.org/10.1073/pnas.90.23.11371 (1993).
5. Luna, L. *et al*. Molecular-Cloning of a Putative Novel Human Bzip Transcription Factor on Chromosome 17q22. *Genomics* **22**, 553–562, https://doi.org/10.1006/geno.1994.1428 (1994).
6. Caterina, J. J., Donze, D., Sun, C. W., Ciavatta, D. J. & Townes, T. M. Cloning and Functional-Characterization of Lcr-F1 - a Bzip Transcription Factor That Activates Erythroid-Specific, Human Globin Gene-Expression. *Nucleic Acids Res* **22**, 2383–2391, https://doi.org/10.1093/nar/22.12.2383 (1994).
7. Murphy, P. & Kolsto, A. B. Expression of the bZIP transcription factor TCF11 and its potential dimerization partners during development. *Mech Develop* **97**, 141–148, https://doi.org/10.1016/S0925-4773(00)00413-5 (2000).
8. Myhrstad, M. C. *et al*. TCF11/Nrf1 overexpression increases the intracellular glutathione level and can transactivate the gamma-glutamylcysteine synthetase (GCS) heavy subunit promoter. *Biochim Biophys Acta* **1517**, 212–219 (2001).
9. Mckie, J., Johnstone, K., Mattei, M. G. & Scambler, P. Cloning and Mapping of Murine Nfe2l1. *Genomics* **25**, 716–719, https://doi.org/10.1016/0888-7543(95)80015-E (1995).
10. Zhang, Y. G., Crouch, D. H., Yamamoto, M. & Hayes, J. D. Negative regulation of the Nrf1 transcription factor by its N-terminal domain is independent of Keap1: Nrf1, but not Nrf2, is targeted to the endoplasmic reticulum. *Biochem J* **399**, 373–385, https://doi.org/10.1042/Bj20060725 (2006).
11. Husberg, C., Murphy, P., Martin, E. & Kolsto, A. B. Two domains of the human bZIP transcription factor TCF11 are necessary for transactivation. *J Biol Chem* **276**, 17641–17652, https://doi.org/10.1074/jbc.M007951200 (2001).
12. Zhang, Y. *et al*. The C-terminal domain of Nrf1 negatively regulates the full-length CNC-bZIP factor and its shorter isoform LCR-F1/Nrf1beta; both are also inhibited by the small dominant-negative Nrf1gamma/delta isoforms that down-regulate ARE-battery gene expression. *PloS one* **9**, e109159, https://doi.org/10.1371/journal.pone.0109159 (2014).
13. Novotny, V., Prieschl, E. E., Csonga, R., Fabjani, G. & Baumruker, T. Nrf1 in a complex with fosB, c-jun, junD and ATF2 forms the AP1 component at the TNF alpha promoter in stimulated mast cells. *Nucleic Acids Res* **26**, 5480–5485, https://doi.org/10.1093/nar/26.23.5480 (1998).
14. Prieschl, E. E. *et al*. A novel splice variant of the transcription factor Nrf1 interacts with the TNFα promoter and stimulates transcription. *Nucleic Acids Res* **26**, 2291–2297 (1998).
15. Schultz, M. A. *et al*. Nrf1 and Nrf2 Transcription Factors Regulate Androgen Receptor Transactivation in Prostate Cancer Cells. *PloS one* **9**, https://doi.org/10.1371/journal.pone.0087204 (2014).
16. Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621–628, https://doi.org/10.1038/nmeth.1226 (2008).
17. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* **29**, 1165–1188 (2001).
18. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57, https://doi.org/10.1038/nprot.2008.211 (2009).
19. Hooper, S. D. & Bork, P. Medusa: a simple tool for interaction graph analysis. *Bioinformatics* **21**, 4432–4433, https://doi.org/10.1093/bioinformatics/bti696 (2005).
20. Deng, W. K., Wang, Y. B., Liu, Z. X., Cheng, H. & Xue, Y. HemI: A Toolkit for Illustrating Heatmaps. *PloS one* **9**, https://doi.org/10.1371/journal.pone.0111988 (2014).
21. Tohge, T. & Fernie, A. R. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc* **5**, 1210–1227, https://doi.org/10.1038/nprot.2010.82 (2010).
22. Chatr-aryamontri, A. *et al*. The BioGRID interaction database: 2017 update. *Nucleic Acids Res* **45**, D369–D379, https://doi.org/10.1093/nar/gkw1102 (2017).
23. Szklarczyk, D. *et al*. STRINGv10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447–D452, https://doi.org/10.1093/nar/gku1003 (2015).
24. Zhang, Y. G. & Xiang, Y. C. Molecular and cellular basis for the unique functioning of Nrf1, an indispensable transcription factor for maintaining cell homoeostasis and organ integrity. *Biochem J* **473**, 961–1000, https://doi.org/10.1042/Bj20151182 (2016).

25. Itoh, K. *et al*. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Gene Dev* **13**, 76–86, https://doi.org/10.1101/Gad.13.1.76 (1999).
26. McMahon, M., Itoh, K., Yamamoto, M. & Hayes, J. D. Keap1-dependent proteasomal degradation of transcription factor Nrf2 contributes to the negative regulation of antioxidant response element-driven gene expression. *J Biol Chem* **278**, 21592–21600, https://doi.org/10.1074/jbc.M300931200 (2003).
27. Kobayashi, A. *et al*. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate for proteasomal degradation of Nrf2. *Mol Cell Biol* **24**, 7130–7139, https://doi.org/10.1128/Mcb.24.16.7130-7139.2004 (2004).
28. Tong, K. I., Kobayashi, A., Katsuoka, F. & Yamamoto, M. Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism. *Biol Chem* **387**, 1311–1320, https://doi.org/10.1515/Bc.2006.164 (2006).
29. Finley, D. Recognition and Processing of Ubiquitin-Protein Conjugates by the Proteasome. *Annu Rev Biochem* **78**, 477–513, https://doi.org/10.1146/annurev.biochem.78.081507.101607 (2009).
30. Kahn, N. W., Rea, S. L., Moyle, S., Kell, A. & Johnson, T. E. Proteasomal dysfunction activates the transcription factor SKN-1 and produces a selective oxidative-stress response in Caenorhabditis elegans. *Biochem J* **409**, 205–213, https://doi.org/10.1042/Bj20070521 (2008).
31. Steffen, J., Seeger, M., Koch, A. & Kruger, E. Proteasomal Degradation Is Transcriptionally Controlled by TCF11 via an ERAD-Dependent Feedback Loop. *Mol Cell* **40**, 147–158, https://doi.org/10.1016/j.molcel.2010.09.012 (2010).
32. Radhakrishnan, S. K. *et al*. Transcription Factor Nrf1 Mediates the Proteasome Recovery Pathway after Proteasome Inhibition in Mammalian Cells. *Mol Cell* **38**, 17–28, https://doi.org/10.1016/j.molcel.2010.02.029 (2010).
33. Sha, Z. & Goldberg, A. L. Proteasome-Mediated Processing of Nrf1 Is Essential for Coordinate Induction of All Proteasome Subunits and p97. *Curr Biol* **24**, 1573–1583, https://doi.org/10.1016/j.cub.2014.06.004 (2014).
34. Xiang, Y. *et al*. Mechanisms controlling the multistage post-translational processing of endogenous Nrf1alpha/TCF11 proteins to yield distinct isoforms within the coupled positive and negative feedback circuits. *Toxicology and applied pharmacology* **360**, 212–235, https://doi.org/10.1016/j.taap.2018.09.036 (2018).
35. Radhakrishnan, S. K., den Besten, W. & Deshaies, R. J. p97-dependent retrotranslocation and proteolytic processing govern formation of active Nrf1 upon proteasome inhibition. *Elife* **3**, 10.7554 (2014).
36. Jirkovsky, E. *et al*. Early and delayed cardioprotective intervention with dexrazoxane each show different potential for prevention of chronic anthracycline cardiotoxicity in rabbits. *Toxicology* **311**, 191–204, https://doi.org/10.1016/j.tox.2013.06.012 (2013).
37. Farmer, S. C., Sun, C. W., Winnier, G. E., Hogan, B. L. M. & Townes, T. M. The bZIP transcription factor LCR-F1 is essential for mesoderm formation in mouse development. *Gene Dev* **11**, 786–798, https://doi.org/10.1101/Gad.11.6.786 (1997).
38. Chen, L. Y. *et al*. Nrf1 is critical for redox balance and survival of liver cells during development. *Mol Cell Biol* **23**, 4673–4686, https://doi.org/10.1128/Mcb.23.13.4673-4686.2003 (2003).
39. Zhang, H. Q., Court, N. & Forman, H. J. Submicromolar concentrations of 4-hydroxynonenal induce glutamate cysteine ligase expression in HBE1 cells. *Redox Rep* **12**, 101–106, https://doi.org/10.1179/135100007X162266 (2007).
40. Zhang, H. Q. *et al*. Nrf2-regulated phase II enzymes are induced by chronic ambient nanoparticle exposure in young mice with age-related impairments. *Free Radical Bio Med* **52**, 2038–2046, https://doi.org/10.1016/j.freeradbiomed.2012.02.042 (2012).
41. Chepelev, N. L. *et al*. Competition of nuclear factor-erythroid 2 factors related transcription factor isoforms, Nrf1 and Nrf2, in antioxidant enzyme induction. *Redox Biol* **1**, 183–189, https://doi.org/10.1016/j.redox.2013.01.005 (2013).
42. Leung, L., Kwong, M., Hou, S., Lee, C. & Chan, J. Y. Deficiency of the Nrf1 and Nrf2 transcription factors results in early embryonic lethality and severe oxidative stress. *J Biol Chem* **278**, 48021–48029, https://doi.org/10.1074/jbc.M308439200 (2003).
43. Ohtsuji, M. *et al*. Nrf1 and Nrf2 Play Distinct Roles in Activation of Antioxidant Response Element-dependent Genes. *J Biol Chem* **283**, 33554–33562, https://doi.org/10.1074/jbc.M804597200 (2008).
44. Hirotsu, Y., Hataya, N., Katsuoka, F. & Yamamoto, M. NF-E2-Related Factor 1 (Nrf1) Serves as a Novel Regulator of Hepatic Lipid Metabolism through Regulation of the Lipin1 and PGC-1 beta Genes. *Mol Cell Biol* **32**, 2760–2770, https://doi.org/10.1128/Mcb.06706-11 (2012).
45. He, J. *et al*. Lipin-1 regulation of phospholipid synthesis maintains endoplasmic reticulum homeostasis and is critical for triple-negative breast cancer cell survival. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **31**, 2893–2904, https://doi.org/10.1096/fj.201601353R (2017).
46. Zhang, Y. G., Lucocq, J. M. & Hayes, J. D. The Nrf1 CNC/bZIP protein is a nuclear envelope-bound transcription factor that is activated by t-butyl hydroquinone but not by endoplasmic reticulum stressors. *Biochem J* **418**, 293–310, https://doi.org/10.1042/Bj20081515 (2009).
47. Zhang, Y. G., Ren, Y. G., Li, S. J. & Hayes, J. D. Transcription Factor Nrf1 Is Topologically Repartitioned across Membranes to Enable Target Gene Transactivation through Its Acidic Glucose-Responsive Domains. *PloS one* **9**, https://doi.org/10.1371/journal.pone.0093458 (2014).
48. Zhang, Y. G. *et al*. The selective post-translational processing of transcription factor Nrf1 yields distinct isoforms that dictate its ability to differentially regulate gene expression. *Sci Rep-Uk* **5**, https://doi.org/10.1038/Srep12983 (2015).
49. Yuan, J., Zhang, S. & Zhang, Y. Nrf1 is paved as a new strategic avenue to prevent and treat cancer, neurodegenerative and other diseases. *Toxicology and applied pharmacology* **360**, 273–283, https://doi.org/10.1016/j.taap.2018.09.037 (2018).
50. Sun, J. *et al*. Hemoprotein Bach1 regulates enhancer availability of heme oxygenase-1 gene. *The EMBO journal* **21**, 5216–5224 (2002).
51. Muto, A. *et al*. The transcriptional programme of antibody class switching involves the repressor Bach2. *Nature* **429**, 566–571, https://doi.org/10.1038/nature02596 (2004).
52. Zhang, Y. G. & Hayes, J. D. Identification of topological determinants in the N-terminal domain of transcription factor Nrf1 that control its orientation in the endoplasmic reticulum membrane. *Biochem J* **430**, 497–510, https://doi.org/10.1042/Bj20100471 (2010).
53. Zhang, Y. G. *et al*. Involvement of the acid sphingomyelinase pathway in UVA-induced apoptosis. *J Biol Chem* **276**, 11775–11782, https://doi.org/10.1074/jbc.M006000200 (2001).
54. Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L. & Rice, P. M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res* **38**, 1767–1771, https://doi.org/10.1093/nar/gkp1137 (2010).
55. Li, R. Q. *et al*. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967, https://doi.org/10.1093/bioinformatics/btp336 (2009).
56. Zhao, L., Zhang, X. M., Qiu, Z. Y. & Huang, Y. De Novo Assembly and Characterization of the Xenocatantops brachycerus Transcriptome. *Int J Mol Sci* **19** (2018).
57. Audic, S. & Claverie, J. M. The significance of digital gene expression profiles. *Genome Res* **7**, 986–995 (1997).
58. Van Belleghem, J. D., Clement, F., Merabishvili, M., Lavigne, R. & Vaneechoutte, M. Pro- and anti-inflammatory responses of peripheral blood mononuclear cells induced by Staphylococcus aureus and Pseudomonas aeruginosa phages. *Sci Rep-Uk* **7** (2017).
59. Xu, J. *et al*. The Entomopathogenic Fungi Isaria fumosorosea Plays a Vital Role in Suppressing the Immune System of Plutella xylostella: RNA-Seq and DGE Analysis of Immunity-Related Genes. *Front Microbiol* **8**, 1–14 (2017).
60. Zhu, L. M. *et al*. Transcriptomics Analysis of Apple Leaves in Response to Alternaria alternata Apple Pathotype Infection. *Front Plant Sci* **8** (2017).
61. Cline, M. S. *et al*. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* **2**, 2366–2382, https://doi.org/10.1038/nprot.2007.324 (2007).

Acknowledgements

The study was supported by the National Natural Science Foundation of China (key programs 91129703, 91429305 and project 31270879) awarded to Prof. Yiguo Zhang (University of Chongqing, China), and in part funded by Chongqing University postgraduates' innovation project (No. CYB15024) awarded to Mr. Lu Qiu.

Author Contributions

M.W. performed most of bioinformatic analyses and experiments, collected the resulting data and prepared draft of this manuscript with most figures and supplementary tables. L.Q. helped M.W. together with molecular cloning to create expression constructs and performed western blotting of Keap1. X.-F.R. prepared the RNA-sequencing samples. Y.-J.S. helped M.W. with functional annotation of differentially expressed genes. Y.-G.Z. designed this study, analyzed all the data, helped to prepare all figures, wrote and revised the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-39536-0.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.