Cell

Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling

Shaohui Hu,[1,4,9] Zhi Xie,[2,9] Akishi Onishi,[3,4,5] Xueping Yu,[2] Lizhi Jiang,[3,4,5] Jimmy Lin,[6] Hee-sool Rho,[1,4] Crystal Woodard,[1,4] Hong Wang,[3,4,5] Jun-Seop Jeong,[1,4] Shunyou Long,[4] Xiaofei He,[1,4] Herschel Wade,[7] Seth Blackshaw,[2,3,4,5,*] Jiang Qian,[2,8,*] and Heng Zhu[1,4,8,*]

[1]Department of Pharmacology and Molecular Sciences
[2]Department of Ophthalmology
[3]Department of Neuroscience
[4]The Center for High-Throughput Biology
[5]Institute of Cell Engineering
[6]Graduate Program in Cellular and Molecular Medicine
[7]Department of Biophysics and Biophysical Chemistry
[8]The Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[9]These authors contributed equally to this work
*Correspondence: sblack@jhmi.edu (S.B.), jiang.qian@jhmi.edu (J.Q.), heng.zhu@jhmi.edu (H.Z.)
DOI 10.1016/j.cell.2009.08.037

SUMMARY

Protein-DNA interactions (PDIs) mediate a broad range of functions essential for cellular differentiation, function, and survival. However, it is still a daunting task to comprehensively identify and profile sequence-specific PDIs in complex genomes. Here, we have used a combined bioinformatics and protein microarray-based strategy to systematically characterize the human protein-DNA interactome. We identified 17,718 PDIs between 460 DNA motifs predicted to regulate transcription and 4,191 human proteins of various functional classes. Among them, we recovered many known PDIs for transcription factors (TFs). We identified a large number of unanticipated PDIs for known TFs, as well as for previously uncharacterized TFs. We also found that over three hundred unconventional DNA-binding proteins (uDBPs)–which include RNA-binding proteins, mitochondrial proteins, and protein kinases–showed sequence-specific PDIs. One such uDBP, ERK2, acts as a transcriptional repressor for interferon gamma-induced genes, suggesting important biological roles for such proteins.

INTRODUCTION

A major challenge in the postgenome era is decoding the functional elements in the human genome. Aided by the sequencing of multiple genomes, computational approaches have identified a large number of evolutionarily conserved DNA elements that include many previously characterized *cis*-regulatory elements (Xie et al., 2005; Xie et al., 2007). Additional studies have identified DNA motifs that are highly enriched in promoters of coexpressed genes (Elemento et al., 2007; Elemento and Tavazoie, 2005; Yu et al., 2006). However, the proteins that recognize these elements cannot be reliably predicted computationally, and the target preferences of only a small minority of DNA-binding proteins have been characterized. Therefore, the identification of interaction networks among the functional elements is the next major step following the identification of the parts list in the human genome.

Protein-DNA interactions (PDIs) are perhaps the most important regulatory interactions involving these functional elements. The most intensively studied subset of PDIs is those between transcription factors (TFs) and their specific DNA target sequences. There are over 1,400 known and predicted human TFs, which fall into multiple subfamilies (Kummerfeld and Teichmann, 2006; Messina et al., 2004). Aside from the interactions between conventional TFs and DNA, the larger set of potential DNA-binding proteins has not been extensively explored. Some proteins that lack any known DNA-binding domains have been found to bind specific DNA sequences (Boggon et al., 1999; Kipreos and Wang, 1992). For instance, Arg5,6, a yeast protein which has traditionally been regarded as a metabolic enzyme with no additional biological functions, recognizes specific DNA sequences and regulates the transcription of genes in the mitochondria (Hall et al., 2004). In general, most proteins that display sequence-specific DNA binding are thought to act as TFs (Teichmann and Babu, 2004); however, some sequence-specific DNA-binding proteins play central roles in such processes as DNA replication, DNA repair, and chromosome dynamics, and are not thought to act as TFs (Petukhova et al., 2005; Tokai-Nishizumi et al., 2005; Zhu et al., 2003).

In the past, biochemical approaches have been used to characterize PDIs, but such approaches are generally laborious and slow. Recent years have witnessed the development of large-scale, unbiased technologies to characterize PDIs. These

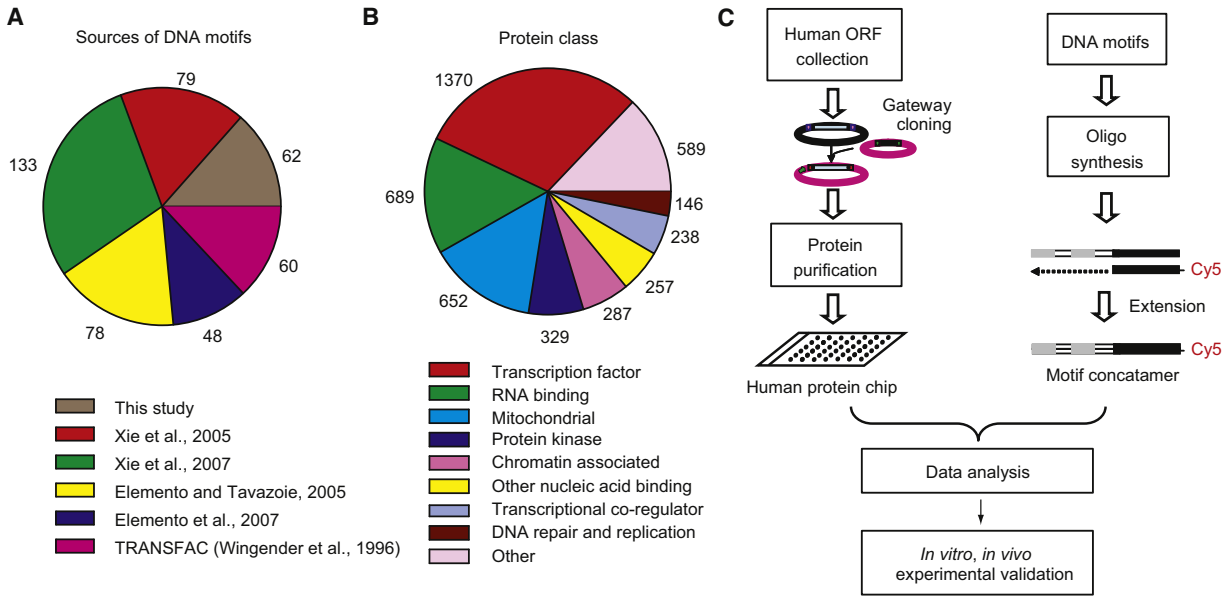


Figure 1. Overall Experimental Design for Analyzing Human PDIs

(A) Sources of the DNA motifs used for probe construction.

(B) Distribution of human proteins selected for protein microarray construction. Some proteins belong to more than one functional class and thus may be counted more than once.

(C) Overall scheme used to identify PDIs in humans using DNA probe binding to protein microarrays.

approaches can be either gene-centered, in which an individual protein is used to identify target sequences, or protein-centered, in which a DNA sequence is used to screen for uncharacterized DNA-binding proteins. Several recent large-scale, gene-centered approaches have employed the double-stranded DNA microarrays, the bacterial one-hybrid system and the yeast one-hybrid system to characterize PDIs in mice, *Drosophila*, and *C. elegans*, respectively (Berger et al., 2008; Deplancke et al., 2006; Noyes et al., 2008). Conversely, protein microarrays have been used both to characterize PDI networks (Ho et al., 2006) and to identify unconventional DNA-binding proteins in yeast (Hall et al., 2004).

In the present study, by using a microarray of 4,191 nonredundant human proteins comprising of known and predicted TFs, as well as representative proteins from other functional classes, we have systematically identified proteins that selectively bind DNA sequences that are either highly evolutionarily conserved or found in the promoters of coexpressed genes. We were able to extensively identify PDIs for known as well as previously uncharacterized human TFs, and we unexpectedly also found that many proteins of other functional classes showed sequence-specific PDIs. We further characterized the DNA binding activity of ERK2, one of these unconventional DNA-binding proteins, using in vitro and in vivo assays and demonstrated that ERK2 acts as a transcriptional repressor regulating interferon gamma signaling in mammalian cells.

RESULTS

Experimental Design

We employed a combined approach to systematically identify proteins that can specifically recognize predicted functional human DNA elements (Figure 1). First, we obtained 752 predicted DNA motifs from previously published studies (Elemento et al., 2007; Elemento and Tavazoie, 2005; Xie et al., 2005; Xie et al., 2007). Second, we used algorithms generated in our laboratories to identify different sets of DNA elements enriched in promoter sequences of tissue-specific genes (see the Supplemental Experimental Procedures available online). Third, we retrieved 60 sequences from the TRANSFAC database corresponding to experimentally-verified binding sites for known TFs (Wingender et al., 1996). After combining these three sources, we removed highly similar motif sequences using a clustering algorithm to produce 460 sequence-diverse DNA motifs with lengths ranging from 6–34 base pairs (Figure 1A, Supplemental Experimental Procedures, Figures S1 and S2, and Table S1). Double-stranded DNA (dsDNA) probes based on these sequences were then synthesized as previously described (Ho et al., 2006).

We next assembled a list of proteins that are likely to recognize these predicted DNA motifs (Table S2 and Supplemental Experimental Procedures). The proteins can be categorized into multiple functional classes (Figure 1B) (1) 1370 known and predicted TFs, representing around 80% of annotated human TFs (Ashburner et al., 2000); (2) proteins known to bind to nucleic acids but without known sequence-specific PDIs, such as RNA-binding proteins, chromatin-associated proteins, and DNA repair enzymes; (3) proteins that regulate transcription but are not known to directly bind DNA, such as transcriptional coregulators; (4) mitochondria-encoded and -targeted proteins and protein kinases, for which previous experimental evidences had suggested that these classes of protein may regulate gene expression (Hall et al., 2004; Pokholok et al., 2006); and (5) an assortment of proteins from a broad range of other functional classes (Table S3).

Human ORFs on this list were selected from the Invitrogen Ultimate ORF collection (Liang et al., 2004) or subcloned in our own laboratories. Using Gateway site-specific recombination (Hartley et al., 2000), ORFs were shuttled to a yeast expression vector that produces N-terminal GST fusions of each protein, and purified from yeast using a previously described strategy (Zhu et al., 2001). To ensure that recombinant proteins were of good quality, we performed immunoblot analysis using anti-GST antibodies, along with silver staining on a randomly selected subset of 200 proteins. Detectable levels of full-length forms of over 90% of the proteins were observed using both methods. Silver staining confirmed the absence of detectable contaminating yeast proteins after purification (Figure S3). Following printing onto nitrocellulose-coated slides (FAST), the complete protein array was probed multiple times with anti-GST antibodies, and more than 98% of the spots produced a signal above background (Figure S4). Pair-wise correlation coefficients of signal intensities ranged from 0.90–0.95 between these slides, illustrating consistency in the array quality.

Data Quality Assessment

To assess the specificity and sensitivity of our approach, we first probed the protein microarrays with three DNA motifs corresponding to consensus-binding sequences for three TFs. These motifs produced highly specific signals, binding selectively to their target proteins with minimal background (Figure 2A). We further tested the specificity of these interactions by probing the array with mutant motifs and observed that they no longer showed specific PDIs (Figure 2A). To eliminate nonspecific PDIs, we also probed the array with Cy5-labeled oligos corresponding to the T7 primer that was used to generate the dsDNA probes. We identified 134 proteins that bound this probe and excluded them from further analysis. On the basis of our earlier observation that bovine histones H3 and H4 bound intensely and nonspecifically to every DNA probe tested, we printed these proteins multiple times on each array as landmarks for orientation and as positive controls for hybridization (Figure 2B). Experimental variability for microarray hybridization was determined by conducting replicate hybridizations of the same probe to four slides. Pair-wise correlation coefficients of signal intensities ranged from 0.68–0.84 for the four slides, with greater consistency for strong signal intensities (Figure S5). On the basis of these control experiments, we concluded that our approach could detect known PDIs sensitively, specifically, and reproducibly.

Global Properties of Observed PDIs

We next used the protein array to analyze PDIs for all of the designed dsDNA motif probes. DNA-binding signals were acquired, analyzed, and normalized using the procedures described in Supplemental Experimental Procedures. From histogram analysis of each hybridization reaction, we observed that a small number of proteins showed strong positive signals with signal intensities many standard deviations (SD) above background, while the vast majority of proteins produced only small background levels of intensity (Figures 2A, 2B, and S6). To increase our confidence in our PDI identification, we applied a stringent cut-off value of 6 SD above background (Table S4).

A total of 17,718 PDIs were detected, with a median number of 30 proteins interacting with each DNA motif probe. Only a single motif did not bind specifically to any of the proteins on the array (Figure 2C). Motif length did not correlate with either the binding intensity or the number of binding proteins observed with a given motif probe (Figure S7). Many proteins on the array bound to only a few probes, while only relatively few proteins bound to a large fraction of probes, a behavior that followed a power-law distribution (Figure 2D). In fact, more than 85.7% of the proteins bound to fewer than 30 of the motifs, confirming that most of the observed PDIs are sequence-specific. For the remaining analysis performed in this study, we focus on only those proteins that fall into this class. It is notable that proteins from different functional classes showed different levels of sequence binding specificity, where RNA-binding proteins have the least sequence specific binding (Figure S8).

TF Binding Specificity

To comprehensively characterize sequence-specificity of the human TFs, we first attempted to identify consensus sequences (logos) that were preferentially bound by individual TFs. We were able to extract significant consensus sequences for 201 TFs (Table S5). These often show considerable overlap with those extracted from TRANSFAC, indicating that our approach can recover reliable consensus sequences using the test motifs (Figure 3A and Table S6). Among all consensus sequences, there are 166 for TFs which have no known binding sites listed in TRANSFAC. Our analysis considerably expands our knowledge of binding specificity of human TFs, almost doubling the number of human TFs for which consensus binding sites have been identified.

We next clustered the TFs based on the similarity of their consensus sequences (Figure 3B). Some TFs with certain DNA-binding domains (e.g., ETS, homeodomain and bHLH) showed more conserved DNA binding specificity. For example, in a clade all but one TF contain the homeodomain and recognize a TAAT consensus sequence (Figure 3B). Interestingly, we found that while some TFs in the same subfamilies showed DNA binding profiles that were distinct from other members of that gene family (e.g., zf-C2H2), many TFs with highly divergent protein sequences bound to highly similar or even identical target DNA sequences (Figure 3B and Table S7). This observation suggests that global primary protein sequence identity does not necessarily correlate with DNA binding specificity.

Finally, we examined the PDIs on the TF subfamily level. We extracted familial logos for the 12 major TF subfamilies (Figure 3C). When compared to the known familial logos from the TRANSFAC and JASPAR databases (Sandelin et al., 2004; Wingender et al., 1996), our analysis identified 8 of the 12 previously reported familial logos. Furthermore, multiple logos were identified for five subfamilies, suggesting that a considerable diversity of DNA binding specificity can be found in members of a given TF subfamily, as has recently been shown for mouse and *Drosophila* homeodomain proteins (Berger et al., 2008; Noyes et al., 2008).

The zf-C2H2 subfamily illustrates the power of our approach. This subfamily contains over 400 members, but no familial logos have been previously reported because of the limited number of confirmed PDIs. With the large number of PDIs characterized in

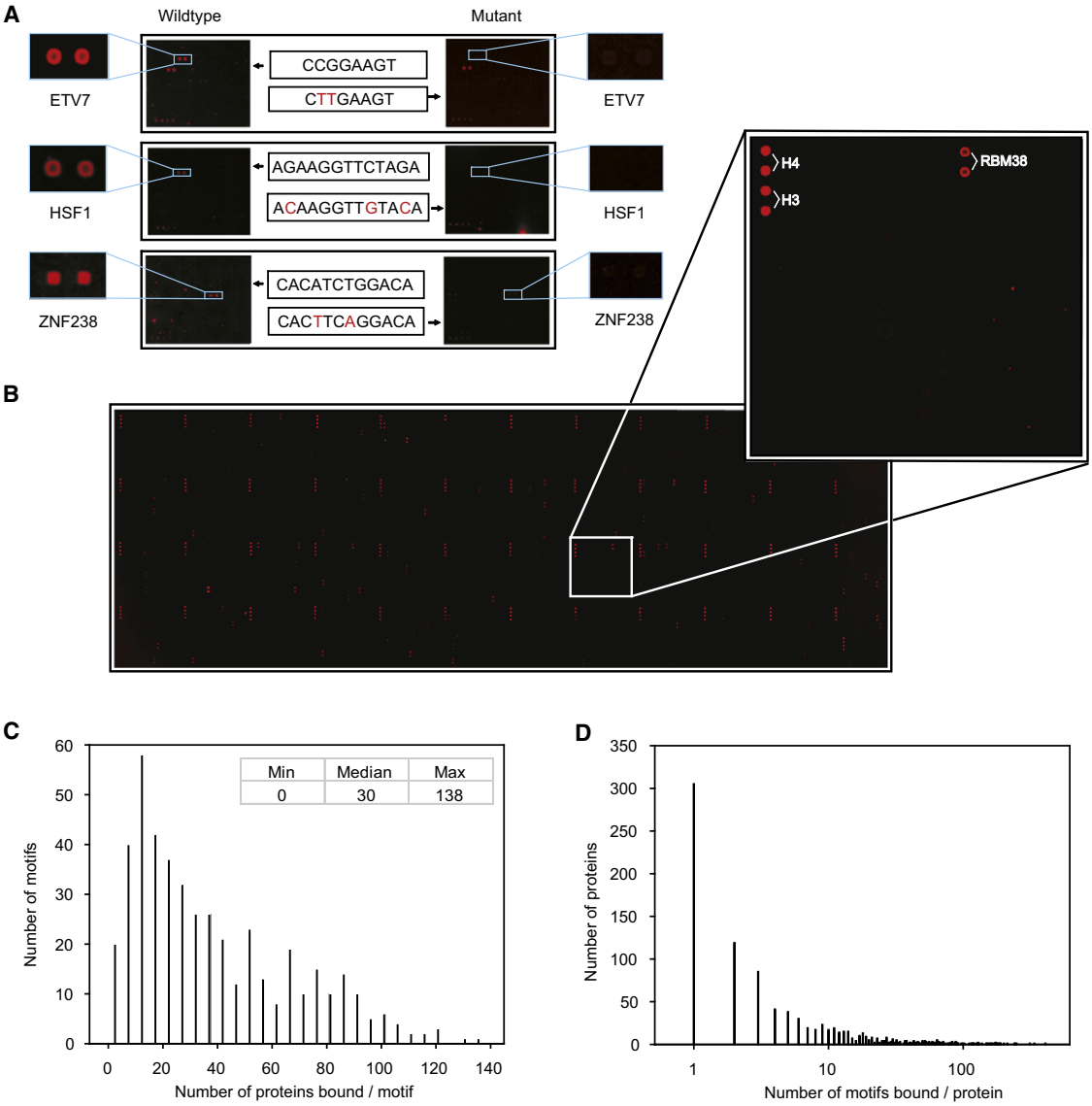


Figure 2. Human Protein-DNA Interactions Detected via Protein Microarrays

(A) Binding specificity of three previously characterized PDIs. Three Cy5-labeled, known dsDNA motifs are separately probed to the protein microarrays and can be specifically recognized by their known TFs, whereas the mutant motifs can no longer bind to their known TFs. Mutated positions are indicated in red.

(B) A typical example of a DNA-binding assay. The DNA motif selectively recognizes RBM38, a predicted RNA-binding protein (inset). Histones H3 and H4, which serve as landmarks and positive controls, are printed in duplicate at a corner of each of the 48 printed blocks.

(C) Histogram showing the number of proteins on the array that were bound by each DNA probe tested.

(D) Histogram showing the number of DNA probes bound by each protein on the array.

this study, we identified six significant logos. For the homeodomain subfamily, we identified not only the canonical consensus site, but also the atypical site recently reported for the TGIF (*Drosophila*) and Meis1 (mouse) groups (Berger et al., 2008; Noyes et al., 2008). On the other hand, only a single familial logo was identified for the NHR, ETS, and RHD subfamilies. These logos closely matched the reported familial logo for each subfamily. Finally, in the case of the Forkhead, IRF, MH1, and Myb subfamilies, we identified familial logos that did not closely resemble the reported ones.

To confirm the specificity of PDIs identified for TFs, we carried out electrophoretic mobility shift assays (EMSA) to test the PDIs for 22 annotated and nine predicted TFs. Notably, 27 of the 31 TFs tested (87.1%) demonstrated specific PDIs, indicating a low false-positive rate for the PDIs identified by protein microarray analysis (Table S8). Figure S9 shows representative examples of 9 of the subfamilies for which familial logos were identified, along with an example of a predicted TF that does not belong to any of these subfamilies. The proteins used in EMSA were tested with silver staining to eliminate the possibility

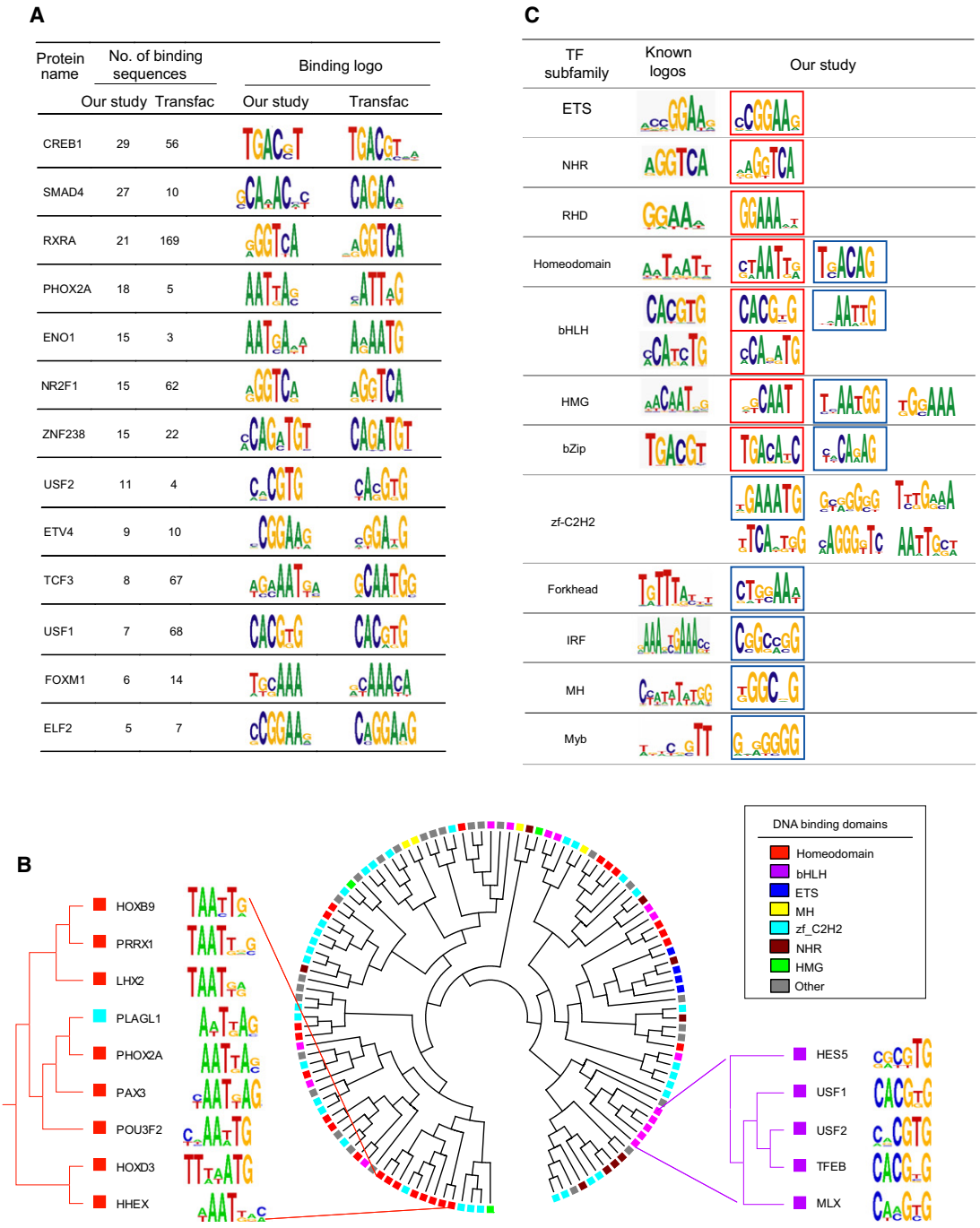


Figure 3. PDIs for Known and Predicted TFs

(A) Comparison between TF-binding logos identified in this study and those listed in TRANSFAC.

(B) Clustering of TFs based on similarity of their DNA logos identified in this study. Only TFs containing known DNA-binding domains were used to construct the cluster. Seven DNA-binding domains are explicitly indicated in the cluster and the other domains are indicated as "Other."

(C) Familial logos identified for the 12 TF subfamilies. Known logos were obtained from JASPAR database (Sandelin et al., 2004). Familial logos recovered in this study that are similar to the known familial logos are outlined in red. Logos validated with EMSA assays are outlined in blue.

of yeast protein contamination (Figure S10). For the four subfamilies (Forkhead, IRF, MH1, and Myb) that did not match the known logos, we were able to validate their logos using EMSA.

Identification of Unconventional DNA-Binding Proteins

Surprisingly, we were able to detect many PDIs between DNA motifs and proteins of other functional classes not previously known to show sequence-specific PDIs. We also extracted

Table 1. Statistics of Human PDIs Detected in This Study

		DNA-Binding Proteins			
		Complete Set[a]		High-Confidence Set[a]	
Protein Class	Total Number of Proteins	Number	Ratio (%)	Number	Ratio (%)
Known TFs	1106	456	41.2	382	34.5
Predicted TFs	264	37	14.0	20	7.6
Protein kinases	329	14	4.3	7	2.1
Chromatin-associated proteins	287	73	25.4	63	22.0
RNA-binding proteins	698	207	29.7	124	17.8
Transcriptional coregulators	238	43	18.1	25	10.5
Other nucleic acid-binding proteins	257	50	19.5	38	14.8
DNA repair & replication	146	50	34.2	42	28.8
Mitochondrial proteins	652	97	14.9	64	9.9
All other categories	589	132	22.4	42	7.1

[a] Complete set of DNA-binding proteins denotes proteins showing DNA binding activity on the protein microarrays. High-confidence set denotes proteins in the complete set which are also annotated as nuclear-localized proteins in GO database, expect for mitochondrial proteins, whose cellular localization is annotated as either nuclear and/or mitochondrial in GO.

consensus sequences for individual unconventional DNA-binding proteins (uDBPs) (Table S9) as well as significant familial logos for each functional class (Figure S11).

For each class of proteins queried, we observed different percentages of proteins showing DNA binding activity (Table 1). The percentages of proteins in different classes that showed DNA binding activity varied greatly—from 4.3% of the protein kinases to 29.7% of the RNA-binding proteins. As a comparison, 41.2% of the annotated TFs showed PDIs, the highest among all protein classes tested. In total, we identified 634 unique uDBPs (Table 1, complete set; note that some proteins belong to multiple functional classes, so that the number of proteins in each functional class listed on Table 1 adds up to more than this total number). This represents 22.4% of all the 2820 non-TF proteins tested, implying that an unexpectedly large fraction of human proteins possess sequence-specific DNA binding activity.

We noticed that some of these proteins are not known to be located in the nucleus, implying that some observed unconventional PDIs might not occur in vivo. To increase the confidence, we further refined this data set to consider only proteins annotated as having nuclear localization in the GO database (Table 1, high-confidence set). Since mitochondrial transcription is actively regulated, all PDIs annotated in GO as showing either nuclear or mitochondrial localization were considered high-confidence. Filtering our initial results in this manner, we obtained 367 unique uDBPs (the high-confidence set, Table 1 and Figure 4B).

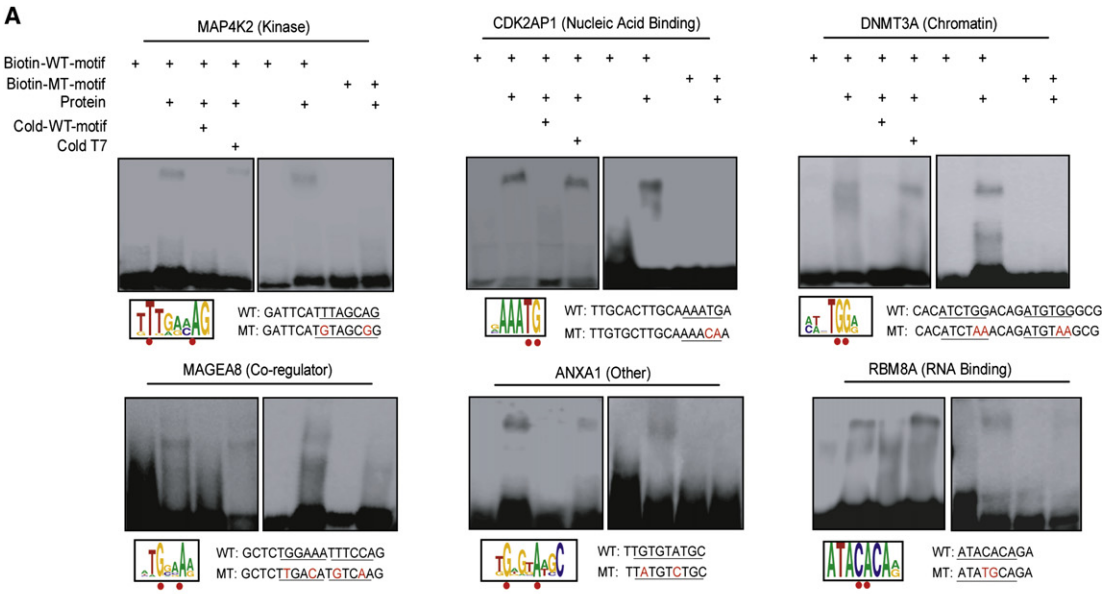
Validation of uDBPs

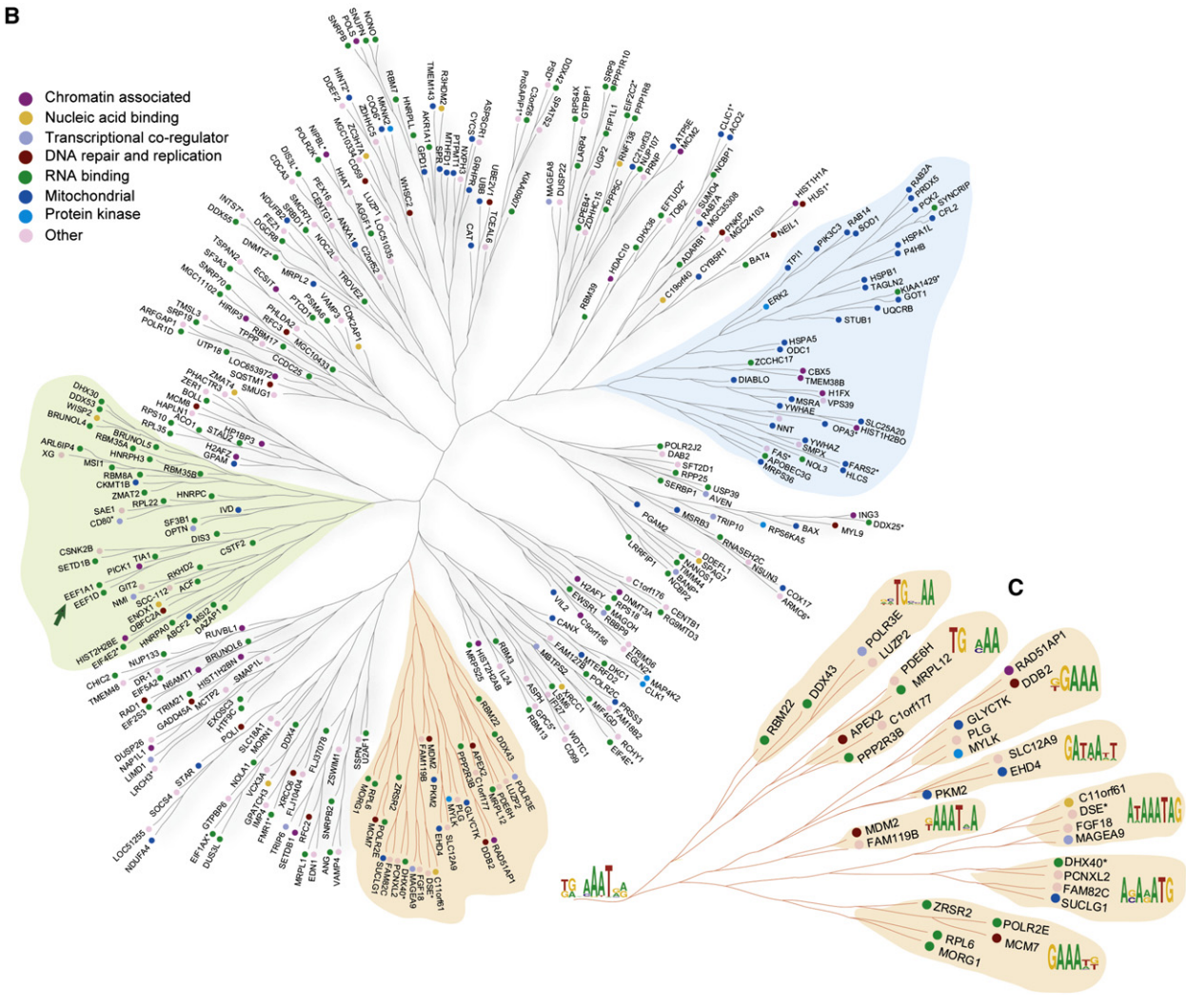
We first used EMSA assays to confirm direct binding of representative uDBPs to the corresponding DNA motifs in vitro. Over 91% (41/45) of the tested uDBPs showed direct PDIs with the corresponding DNA motifs identified from the protein microarray data (Figure 4A, Table S10). To experimentally validate the calculated familial logos, we designed mutant DNA sequences with differing sequences at two conserved nucleotide positions. Of the 13 tested proteins, 12 (92.3%) showed significant decreases in PDIs with the mutant motifs. Proteins demonstrating sequence-specific PDIs in this assay came from diverse functional categories, including mitochondrial-targeted proteins, RNA-binding proteins, and protein kinases (Figures 4A and S12). Furthermore, no contaminating yeast proteins were observed following silver-staining analysis of the purified recombinant proteins that were used for EMSA, implying that any observed PDIs are highly unlikely to result from the presence of any contaminating yeast TFs (Figure S10).

It is notable that the EMSA assays confirmed highly sequence-specific PDIs for several RNA-binding proteins, many of which were believed to bind RNA and/or DNA molecules indiscriminately. To further validate their binding specificity, we performed additional EMSA assays with single-stranded DNA (ssDNA) as competitors for two representative RNA-binding proteins. The sequence-specific PDIs showed no apparent difference with or without competition from ssDNA (Figure S13), confirming that observed specific PDIs for these RNA-binding proteins indeed result from binding to dsDNA. Taken together, these results indicate that the majority of the uDBPs identified in this study can indeed interact with DNA motifs directly and specifically.

Many uDBPs Associate with DNA In Vivo

The most surprising result to us is the observation of sequence-specific PDIs for sugar and protein kinases. To determine whether these uDBPs associate with DNA in vivo, we selected antibodies against phosphoenolpyruvate carboxykinase 2 (PCK2) and mitogen-activated protein kinase 1 (ERK2/MAPK1) to perform chromatin-immunoprecipitation (ChIP). Using primers designed to flank genomic-binding sites for these proteins predicted from our protein microarray PDI data, we obtained positive PCR products for both proteins (Figures 5D and S14), indicating that they do indeed associate with these predicted target sequences in vivo. We next conducted a thorough literature search and found that an additional 12 of the 367 uDBPs identified in this study have been shown to associate with DNA in vivo using ChIP (Table S11), although these previous studies had interpreted these data to indicate that these proteins did not directly bind DNA. More importantly, we found that ChIPed DNA products in every case included sequences that match the predicted consensus DNA-binding sites for these uDBPs. Taken together, a total of 14 uDBPs are associated in vivo with DNA fragments that contain our predicted DNA logos.


A
MAP4K2 (Kinase)
CDK2AP1 (Nucleic Acid Binding)
DNMT3A (Chromatin)
Biotin-WT-motif
Biotin-MT-motif
Protein
Cold-WT-motif
Cold T7
WT: GATTCATTTAGCAG
MT: GATTCATGTAGCGG
WT: TTGCACTTGCAAAATGA
MT: TTGTGCTTGCAAAACAA
WT: CACATCTGGACAGATGTGGGCG
MT: CACATCTAACAGATGTAAGCG
MAGEA8 (Co-regulator)
ANXA1 (Other)
RBM8A (RNA Binding)
WT: GCTCTGGAAATTTCCAG
MT: GCTCTTGACATGTCAAG
WT: TTGTGTATGC
MT: TTATGTCTGC
WT: ATACACAGA
MT: ATATGCAGA


B
Chromatin associated
Nucleic acid binding
Transcriptional co-regulator
DNA repair and replication
RNA binding
Mitochondrial
Protein kinase
Other
C
RBM22
DDX43
POLR3E
LUZP2
PDE6H
MRPL12
APEX2
C1orf177
PPP2R3B
RAD51AP1
DDB2
GLYCTK
PLG
MYLK
SLC12A9
EHD4
PKM2
MDM2
FAM119B
C11orf61
DSE*
FGF18
MAGEA9
DHX40*
PCNXL2
FAM82C
SUCLG1
ZRSR2
RPL6
MORG1
POLR2E
MCM7

Global Classification of uDBPs

Given the existence of this group of uDBPs, we set out to classify and organize these proteins. We assessed protein relatedness on the basis of the DNA motif sequences to which the proteins bound. DNA-binding profiles were constructed for each protein to include the binding intensity of the protein to each of the 460 distinct DNA-binding motifs (Supplemental Experimental Procedures). A hierarchical tree was then built based only on the similarity of the binding profiles of these unconventional DNA-binding proteins (Figure 4B). Two disparate trends were observed: On the one hand, in some clades there was a clear enrichment of proteins traditionally known to be part of a specific functional class. For example, two clades (Figure 4B, blue and green shading) were significantly over-represented for mitochondria proteins ($p < 4.78e-11$) and RNA-binding proteins ($p < 4.15e-9$), respectively. Another interesting example is that eukaryotic translation elongation factor 1 alpha 1 (EEF1A1) and delta (EEF1D), which belong to the translational elongation complex but share no sequence homology, were found to recognize similar DNA motif sequences. Such clustering indicates that some proteins that are similar either in terms of sequence homology or functional annotation may have similar DNA-binding characteristics. On the other hand, a mixture of functionally divergent proteins without sequence homology were also observed to share similar DNA-binding motifs in some clades (Figures 4B and 4C), indicating that these proteins of highly divergent structure and function may cooperate to control the same DNA-binding targets.

ERK2 Acts as a Transcriptional Repressor

As demonstrated above, many uDBPs directly and specifically bind DNA in vitro and 14 of them are found to associate with DNA in vivo. Therefore, we predicted that these uDBPs might play a physiological role in transcriptional regulation in vivo. We decided to focus on in-depth characterization of this property in ERK2, an extensively studied protein that is known to be involved in a variety of biological processes, including proliferation, differentiation, and development.

Our protein microarray-based PDI analysis revealed that ERK2 can bind to a G/CAAAG/C consensus sequence. We investigated this directly using EMSA analysis using both wild-type oligonucleotides matching the consensus site and mutant probes that departed from this consensus. We found that this binding is sequence-specific, since mutant oligonucleotides no longer showed binding activity (Figure 5A). Silver-staining analysis of ERK2 showed that no contaminating yeast proteins were observed (Figure S10). In addition, we performed EMSA assays with ERK2 protein purified from *E. coli* and still observed the sequence-specific PDI, further ruling out any possible contamination from yeast TFs (Figure S15).

To determine whether ERK2 could act as a transcriptional regulator in vivo through sequence-specific DNA binding, we next employed cell-based luciferase analysis. The corresponding wild-type and mutant motif sequences were cloned upstream of a minimal promoter in a luciferase reporter construct. We found that ERK2 tested with the wild-type motif sequence showed repression of luciferase expression in a dose-dependent manner, but showed little or no change in luciferase expression when assayed with the mutant motif, which did not bind to ERK2 protein in the EMSA assay (Figure 5B).

To identify targets of ERK2 and thereby gain clues to its function, we compared the gene-expression profiles of HeLa cells to those of the cells in which ERK2 is knocked down using siRNA (Huang et al., 2008). Because ERK2 showed a dose-dependent repression of luciferase activity in the assays described above, we collected the promoter sequences of 82 genes that showed at least a two-fold upregulation of expression following siRNA-mediated knockdown of ERK2 when compared to the control. Application of an *in silico* motif discovery algorithm to these sequences revealed a similar consensus sequence (GAAAC) to that determined by the protein microarray analysis (Figure 5C and Supplemental Experimental Procedures). In fact, the promoter regions of 78 of the 82 genes contained a total of 270 GAAAC sites, a clear indication of significant enrichment for these upregulated genes ($p = 1.5e-9$). The distribution of the ERK2-binding sites relative to the transcription start site showed a sharp peak around −90 bp, a typical distribution for many TFs (Figure 5C). ERK2 consensus sequences were not enriched in the promoter sequences of downregulated genes in ERK2 siRNA-treated cells, consistent with our observation that ERK2 represses gene expression in luciferase assays (Figure 5B).

To determine whether ERK2 binds in vivo to the promoters of any of these genes whose expression is upregulated in HeLa cells lacking ERK2 and that contain GAAAC logos upstream, 21 of these genes were tested for ERK2 binding by using ChIP. Eleven of 21 genes (52.3%) showed higher levels of immunoprecipitation with the anti-ERK2 antibody relative to controls (Figure 5D). Such enrichment was not observed for any of the six downregulated or the six unaffected genes tested (Figure S16). Thus, ERK2 associates with GAAAC sequences in vivo to regulate expression of a large number of genes.

DNA Binding Activity of ERK2 Is Independent of Kinase Activity

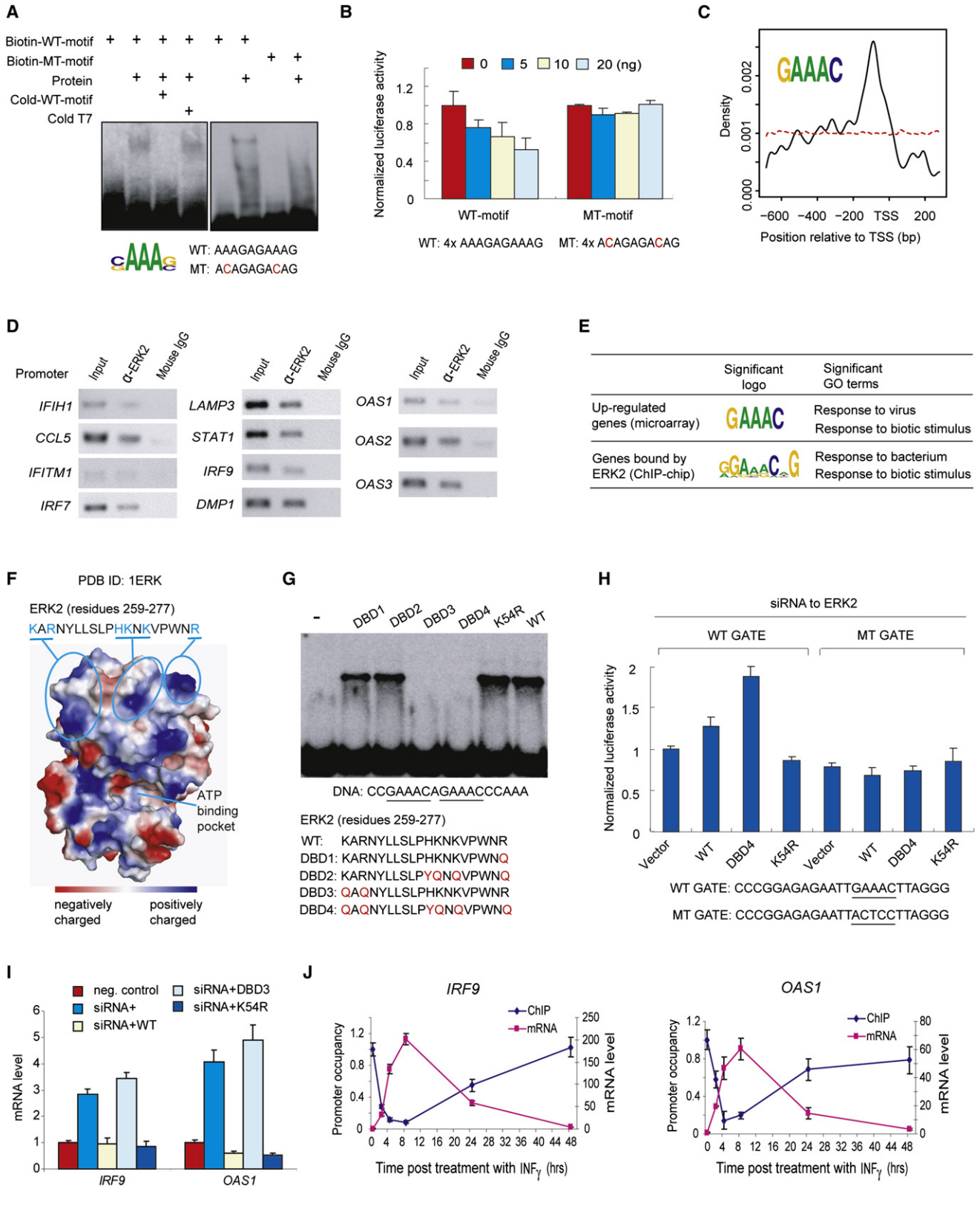
Because the protein kinase activity of ERK2 has been well studied, it is possible that its DNA binding activity serves a

Figure 4. DNA Binding Specificity of uDBPs

(A) Validation of unconventional PDIs with EMSA analysis. Representative examples are shown. Consensus sites identified in the current study for different proteins are boxed and underlined in the DNA motif sequences used for the EMSA analysis. Mutated positions are indicated in red in motif sequences used for EMSA and underscored with red dots in the predicted consensus sequences.

(B) Clustering of uDBPs based on target sequence similarity. Proteins of different function classes are color-coded. Branches highlighted in green and blue are enriched for RNA-binding and mitochondrial-targeted proteins, respectively. Asterisks indicate that multiple proteins bind to identical target sequences; in this case, a single representative protein is shown (see Table S12 for detail). The arrow indicates an example of two proteins that interact as part of a protein complex but do not share protein sequence homology.

(C) Magnified view of the orange branch in (B), where the consensus sequences for each sub-branch are shown.


A
Biotin-WT-motif
Biotin-MT-motif
Protein
Cold-WT-motif
Cold T7
WT: AAAGAGAAAG
MT: ACAGAGACAG
B
0 5 10 20 (ng)
Normalized luciferase activity
WT-motif
MT-motif
WT: 4x AAAGAGAAAG
MT: 4x ACAGAGACAG
C
GAAAC
Density
Position relative to TSS (bp)
D
Promoter
Input
α-ERK2
Mouse IgG
IFIH1
CCL5
IFITM1
IRF7
LAMP3
STAT1
IRF9
DMP1
OAS1
OAS2
OAS3
E
Significant logo
Significant GO terms
Up-regulated genes (microarray)
Response to virus
Response to biotic stimulus
Genes bound by ERK2 (ChIP-chip)
Response to bacterium
Response to biotic stimulus
F
PDB ID: 1ERK
ERK2 (residues 259-277)
KARNYLLSLPHKNKVPWNR
ATP binding pocket
negatively charged
positively charged
G
-
DBD1
DBD2
DBD3
DBD4
K54R
WT
DNA: CCGAAACAGAAACCCAAA
ERK2 (residues 259-277)
WT: KARNYLLSLPHKNKVPWNR
DBD1: KARNYLLSLPHKNKVPWNQ
DBD2: KARNYLLSLPYQNQVPWNQ
DBD3: QAQNYLLSLPHKNKVPWNR
DBD4: QAQNYLLSLPYQNQVPWNQ
H
siRNA to ERK2
WT GATE
MT GATE
Normalized luciferase activity
Vector
WT
DBD4
K54R
WT GATE: CCCGGAGAGAATTGAAACTTAGGG
MT GATE: CCCGGAGAGAATTACTCCTTAGGG
I
neg. control
siRNA+
siRNA+WT
siRNA+DBD3
siRNA+K54R
mRNA level
IRF9
OAS1
J
IRF9
OAS1
ChIP
mRNA
Promoter occupancy
mRNA level
Time post treatment with INFγ (hrs)

distinct cellular function. To explore the possibility, we examined the 82 upregulated genes for potential functional enrichment. These genes are enriched for proteins involved in response to biotic stimuli (p = 1.0e-16) and to viral infection (p = 1.0e-24) (Figure 5E). Furthermore, by analyzing the results of our ChIP-chip analysis for ERK2, we discovered a similar consensus sequence and a functional enrichment for response to biotic stimuli (p = 0.03) and response to bacterial infection (p = 0.02) (Figure 5E). These functions are not known for ERK2 in previous studies. In contrast, we found that the 53 confirmed substrates of ERK2 (Diella et al., 2008) are not enriched for the same functions (Figure S17). Thus, it is very likely that sequence-specific DNA binding activity of ERK2 is independent of its kinase activity.

To examine the structural basis of this hypothesis, we analyzed the crystal structure of ERK2 and identified one surface patch as a potential DNA-binding domain, which is comprised of three clusters of positively charged residues close to the C terminus at considerable distance from the ATP-binding pocket and the substrate groove (Figure 5F). Using site-directed mutagenesis, we investigated whether these residues might be required for sequence-specific DNA binding by ERK2. We found that mutations in DBD3 and DBD4 completely abolished sequence-specific DNA binding by ERK2 using EMSA analysis, indicating that K259 and R261 are the two key residues required for its DNA binding activity (Figure 5G). In contrast, the kinase-dead mutant (K54R) did not show any effect on DNA binding (Robinson et al., 1996). We further confirmed that the kinase activity of ERK2 was not essential for DNA binding by performing EMSA analysis with purified ERK2 proteins coexpressed with MEK1 in *E. coli*. We observed that DNA binding was unaffected by the presence of staurosporine, a kinase inhibitor (Figure S15).

ERK2 Directly Represses Expression of Interferon Gamma-Induced Genes via DNA Binding Activity

Finally, we set out to determine the physiological function of the DNA binding activity of ERK2. Interestingly, nine of the eleven genes whose promoters could be ChIPed with the anti-ERK2 antibody in HeLa cells are known to be induced by interferon. Furthermore, previous studies have shown that a transcription factor, CCAAT/enhancer-binding protein-β (C/EBP-β), binds to a so-called GATE element in the proximal promoters of one of these genes, *IRF9*, and activates its transcription upon interferon gamma (IFNγ) stimulation (Roy et al., 2000). We found that the consensus site for ERK2 is embedded in GATE element. These evidences suggest that ERK2 might be involved in IFNγ signaling via its DNA binding activity.

To test specific interactions between GATE element and the identified DNA-binding domain in ERK2, we conducted luciferase analysis in transfected HeLa cells, using a wild-type GATE element reporter and a mutant element that lacks the consensus ERK2-binding site (Weihua et al., 1997). We find that cotransfection of the siRNA-resistant wild-type *ERK2*, along with siRNAs directed against endogenous *ERK2*, did not result in a significant difference in luciferase expression compared to controls when a wild-type GATE element reporter construct is used (Figure 5H). However, the DNA-binding-deficient mutant of ERK2 led to substantially upregulated reporter expression when cotransfected with *ERK2*-targeted siRNA. In contrast, kinase-dead mutants of ERK2 efficiently repressed reporter expression. Neither wild-type nor mutant proteins showed any effect on the activity of the mutant GATE element reporter when overexpressed (Figure 5H). These results clearly demonstrated that ERK2 specifically and directly represses expression of the luciferase reporter genes driven by canonical GATE element via its DNA-binding domain in vivo.

Figure 5. ERK2 as a Transcriptional Repressor

(A) Validation of ERK2-DNA interaction with EMSA analysis. Mutated positions are indicated in red in motif sequences used for EMSA.

(B) Dose-dependent transcriptional repression by ERK2 using cell-based luciferase assays. Four tandem repeats of the wild-type (WT) motif shown to complex with ERK2 were cloned into pTK-Luc vector and cotransfected into GT1-7 cells with varying amount of plasmids that expressed ERK2. The mutant motif that abolished gel-shifting also abolished dose-dependent transcriptional repression by ERK2. Error bars represent ± SD of three independent experiments. The same error measurement is also applied to the experiments in panels (H, I, and J).

(C) Positioning distribution of ERK2-binding sites in promoters. Application of an *in silico* motif discovery algorithm to the promoter regions of 82 upregulated genes in a ERK2 knockdown experiment revealed a similar consensus sequence (inset) to that determined by the protein microarray analysis panel (A). The promoter region extends from −700 to 300 bp relative to the transcription start site (TSS). The red dashed line shows the relative position of 1000 random 5-mer DNA sequences to the TSS.

(D) In vivo validation of ERK2 and DNA interactions using ChIP coupled with PCR analysis. An anti-ERK2 monoclonal antibody was used to ChIP the endogenous ERK2 proteins in HeLa cells. Specific primer pairs were designed to PCR-amplify the promoter regions of the predicted targets of ERK2. Mouse IgG was used as a negative control for immunoprecipitation. Of the 21 upregulated genes assessed, 11 (52.3%) showed higher levels of immunoprecipitation with the anti-ERK2 antibody than with the IgG control.

(E) Comparison of consensus sites and enriched GO terms in ERK2 knockdown and ChIP-chip experiments.

(F) Structural analysis for DNA-binding domain in ERK2. Calculated using PyMol, the electrostatics surface potential of ERK2 is color-coded. A surface patch (residues 259–277) comprised of three positively charged clusters are indicated with the amino acid sequence showing above. The ATP-binding pocket is also shown.

(G) Mapping the DNA-binding domain in ERK2. Five mutant forms of ERK2 were constructed and the corresponding proteins were purified. As determined with EMSA analysis, mutations in DNA-binding-deficient (DBD) mutants 3 and 4 completely abolished the DNA binding activity, indicating that K259 and R261 are required. In contrast, K54R mutation (kinase-dead) did not affect the DNA binding activity, indicating that the two activities are independent. The DNA sequence used in the EMSA assay is also shown.

(H) Specific interactions between GATE element and the DNA-binding domain in ERK2. Using a previously reported luciferase reporter system (Weihua et al., 1997), the effects of overexpressing ERK2 in various mutant forms are monitored in cells that the endogenous ERK2 is knocked down.

(I) Regulation of IFNγ-induced gene expression by the DNA binding activity of ERK2. Changes in *IRF9* and *OAS1* expression are normalized to those in negative control cells.

(J) Dynamics of promoter occupancy by ERK2 in reverse correlation to mRNA expression levels of *IRF9* and *OAS1* after IFNγ treatment.

To further confirm the transcriptional repressor activity of ERK2 against chromosomal genes, we monitored gene expression level of two known IFNγ-induced genes, *IRF9* and *OAS1*, by overexpressing different mutant forms of ERK2 in HeLa cells. We first determined that siRNA-mediated knockdown of endogenous *ERK2* significantly de-repressed expression of *IRF9* and *OAS1* (Figure 5I). However, in cells that lack endogenous ERK2, overexpression of kinase-dead ERK2 repressed expression of *IRF9* and *OAS1* as efficiently as overexpression of wild-type ERK2, whereas overexpression of DNA-binding-deficient ERK2 did not show any significant effects (Figure 5I). These results suggest that ERK2 plays an important role in regulating expression of IFNγ-induced genes via its DNA binding activity.

The above data suggest that low expression of IFNγ-induced genes might be maintained by the occupancy of ERK2 on the promoters. Therefore, we predicted that promoter occupancy of these genes by ERK2 might inversely correlate with induction of gene expression in response to IFNγ application. Using a combination of quantitative ChIP and qRT-PCR, we measured the dynamics of promoter occupancy by ERK2 and gene expression of *IRF9* and *OAS1*. During the course of IFNγ treatment we observed that ERK2 was rapidly depleted from the promoters of *IRF9* and *OAS1* within the first four hours and the ERK2 occupancy reached its lowest level between 6 and 8 hr posttreatment. Interestingly, promoter occupancy by ERK2 gradually rose and almost fully recovered to its original level at 48 hr posttreatment. As predicted, the mRNA level of both *IRF9* and *OAS1* shows a near-perfect inverse correlation to promoter occupancy by ERK2 (Figure 5J).

DISCUSSION

The identification of many sequence-specific PDIs for both conventional TFs and uDBPs raises an interesting question; that is whether these uDBPs bind to different target sequences than do annotated TFs. While some proteins in the same functional class were found to have preferred DNA-binding profiles selective to that protein family, the overlap in the DNA motifs recognized by the TFs and uDBPs is remarkable and substantial (Figure S18), which suggests a complex landscape for human PDI networks and possible crosstalk between TFs and uDBPs. As an example, we found that ERK2 regulates expression of IFNγ-induced genes via binding to GATE element, which has also been shown to be bound by C/EBP-β (Roy et al., 2000).

Our study suggests that a crosstalk between C/EBP-β and the DNA-binding and kinase activities of ERK2 results in a negative feedback loop to tightly control the temporal expression pattern of *IRF9* and *OAS1* upon IFNγ induction. Previously, Kalvakolanu and colleagues showed that upon IFNγ induction C/EBP-β is phosphorylated by ERK1/2 to activate expression of the GATE-driven genes (Roy et al., 2002). However, this model does not explain upregulation of the GATE-driven genes when only ERK2 is knocked down in cells (Huang et al., 2008) or the suppression of *IRF9* and *OAS1* 8 hr post IFNγ-treatment (Figure 5J). Based on the discovered DNA binding activity of ERK2, a plausible explanation is that expression of the GATE-driven genes is dictated by competitive binding of C/EBP-β and ERK2 to GATE element. In untreated cells, GATE is directly bound by ERK2 via its DNA-binding domain and transcription of the downstream genes is inhibited, which explains the upregulation of those IFN-response genes when ERK2 is knocked down (Huang et al., 2008). When cells are treated with IFNγ, C/EBP-β is rapidly induced and phosphorylated by ERK1/2, which are activated by the MEKK1/MEK1 pathway (Roy et al., 2002). The activated C/EBP-β in the nucleus then rapidly competes off ERK2 bound to GATE, resulting in a rapid activation of the GATE-driven genes and a sharp decline of ERK2 occupancy at GATE (Figure 5J). As this proceeds, the concentration of nuclear ERK2 gradually increases to a level that it starts to compete off bound C/EBP-β and therefore posts a negative feedback to eventually shut down expression of these genes. Taken together, we believe that the crosstalk between the two independent ERK2 activities and C/EBP-β partially explains the dynamics of IFNγ-induced gene expression.

A significant advantage of the presented protein-centered approach is that the binding specificity of a given DNA motif can be simultaneously measured for thousands of proteins in a single assay. In our studies, we carefully selected DNA motifs that are either highly conserved during evolution or highly enriched in the regulatory regions of coexpressed genes, and thus likely to act to regulate transcription. Indeed, the fact that virtually all of the DNA motifs tested in this study bound selectively to proteins on the array suggests that these sequences are indeed involved in regulating transcription in vivo. Furthermore, our approach can examine a large variety of protein families, providing an opportunity to discover uncharacterized DNA-binding proteins. It is expected that combined with gene-centered approaches, such as protein-binding DNA microarrays and one-hybrid analysis, we will be able to precisely determine DNA binding consensus sequences for many uDBPs.

EXPERIMENTAL PROCEDURES

Probe Preparation

Double-stranded DNA probes were generated according to a protocol described previously (Ho et al., 2006).

Human ORF Cloning

Using the Gateway recombinant cloning system (Invitrogen, CA), human ORFs were shuttled from the selected entry clones of the Ultimate Human ORF Collection (Invitrogen, CA) or from the entry clones generated in our own laboratories to a yeast high-copy expression vector (pEGH-A) that produces GST-His_6 fusion proteins under the control of the galactose-inducible *GAL1* promoter. Plasmids were rescued into *E. coli* and verified by restriction endonuclease digestion. Plasmids with inserts of correct size were transformed into yeast for protein purification.

Protein Purification

Human proteins were purified as GST-His_6 fusion proteins from yeast using a high-throughput protein purification protocol as described previously (Zhu et al., 2001).

Protein Microarrays

Purified human proteins were arrayed in a 384-well format and printed on FAST slides (Whatman, Germany) in duplicate. The protein microarrays were probed with Cy5-labeled DNA motifs using a protocol similar to that previously described (Ho et al., 2006): A protein chip was blocked for 3 hr with 3% BSA in hybridization buffer (25 mM HEPES at pH 8.0, with 50 mM KGlu, 0.1% Triton X-100, 8 mM $MgAC_2$, 3 mM DTT, 4 μM poly [dA-dT], and 10% glycerol) and

then incubated with a Cy5-labeled DNA motif at a final concentration of 40 nM in hybridization buffer at 4°C overnight. The chip was washed once in cold hybridization buffer without poly (dA-dT) for 5 min and spun to dryness. The slides were finally scanned with a GenePix 4000 scanner (MDS Analytical Technologies, CA) and the binding signals were acquired using the GenePix software.

EMSAs

Each binding reaction was carried out with 100 fmol of biotinylated dsDNA probe and 2 pmol of purified protein in 20 μl of binding buffer (25 mM HEPES at pH 8.0 with 50 mM KGlu, 0.1% Triton X-100, 2 mM $MgAC_2$, 3 mM DTT, and 5% glycerol). Twenty-five pmol (a 250-fold excess) of unlabeled (cold) DNA motifs were added in the competition assays. Reactions were carried out for 30 min at room temperature, followed by overnight incubation at 4°C. Reaction mixtures were loaded onto 5% TBE polyacrylamide gels and separated at 100 V on ice until the dye front migrated two-thirds of the way to the bottom of the gel. Nucleic acids were transferred to nylon membranes and visualized with the LightShift EMSA Kit (Pierce, USA) according to the manufacturer's recommendations. All the expression clones for proteins used in EMSA were verified by DNA sequencing.

Luciferase Assays

Four tandem repeats of the DNA motif and the GATE element (Weihua et al., 1997) were subcloned into pTK-Luc vector (McKnight et al., 1981) and pGL3 vector (Promega, USA), respectively. DNA was transfected using the FugeneHD reagent (Roche, Switzerland). For the 4× DNA-motif, GT1-7 cells were cotransfected with three constructs: pTK-Luc, pCAGIG expressing ERK2, and pRL-TK (Promega, USA). For the GATE element, 3 hr after the transfection of pGL3 construct, siRNA against 3′UTR of ERK2 was tranfected using TransPass R1 reagent (NEB, USA). Cells were harvested 48 hr posttransfection for luciferase reporter assay using the Dual-Luciferase reporter assay system (Promega, USA). The luciferase activity was normalized by the internal control pRL-TK *Renilla* luciferase activity. All assays were performed in three separate experiments done in triplicate.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) was carried out on HeLa cells using a mouse anti-ERK2 antibody (Millipore, USA) or a rabbit anti-PCK2 antibody (Santa Cruz, USA) according to a protocol described previously (Nelson et al., 2006), except that the protein A-Sepharose was replaced with salmon sperm DNA/protein A-agarose (Millipore, USA). Normal mouse or rabbit IgG was used for mock IP as a negative control.

Site-Directed Mutagenesis

Site-directed mutagenesis was carried out using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene, USA) as described previously (Jensen and Weilguny, 2005).

Computational Analysis

The tissue specific motifs were identified using algorithms previously described (Yu et al., 2006), and see Supplemental Experimental Procedures for details. The procedures of protein chip data analysis include image scan, background correction, within-chip normalization, identification of positive hits, and nonspecific binding filtering. Normalization and identification of positive hits were performed using the algorithms described in Supplemental Experimental Procedures in detail. DNA-binding logos were discovered using AlignACE (Roth et al., 1998). The DNA-binding logos were aligned using the ungapped Smith-Waterman algorithm (Smith and Waterman, 1981). The clustering tree of the TF logos was built using Neighbor-joint algorithm. The tree was visualized using MEGA4 (Tamura et al., 2007). Potential DNA motifs in the promoter regions were identified using MDscan (Liu et al., 2002). The distance between the DNA-binding profiles of any two proteins in the phylogenetic tree is defined in Supplemental Experimental Procedures. The initial phylogenetic tree was constructed based on the distance information using the minimum evolution method in MEGA4. The length of the branches was log-transformed. The curved layout was built manually. The length of the branches was in some cases slightly altered when the curved layout was constructed, and therefore the length was not precisely proportional to the actual distances between binding profiles. *P* value of GO analysis was calculated using one-sided Fisher exact test corrected for multiple testing using the minimum P method of Westfall and Young (Westfall and Young, 1993) as provided in Ontologizer (Bauer et al., 2008). ChIP-chip data was analyzed using Cisgenome (Ji et al., 2008).

SUPPLEMENTAL DATA

Supplemental Data include Supplemental Experimental Procedures, 20 figures, 13 tables, and Supplemental References and can be found with this article online at http://www.cell.com/supplemental/S0092-8674(09)01111-8.

ACKNOWLEDGMENTS

We thank Drs. J. Boeke, P. Cole, J. Nathans, G. Seydoux, T. Shimogori, S. Chen, D. Griffin, S. Taverna, J. Pomerantz, and D. Zack for their comments and suggestions. We also thank Drs. K. Dalby, D. Kalvakolanu, and R. Weiner for providing reagents; and D. McClellan for editorial assistance. This work was supported by the National Institutes of Health (GM076102 to H.Z., J.Q., RR020839 to H.Z., NEI Vision Core Grant to J.Q.), a W. M. Keck Foundation Distinguished Young Investigator in Medical Research Award to S.B., a grant from the Ruth and Milton Steinbach Fund to S.B., and a generous gift from Mr. and Mrs. Robert and Clarice Smith.

Received: December 29, 2008
Revised: July 13, 2009
Accepted: August 20, 2009
Published: October 29, 2009

REFERENCES

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Bauer, S., Grossmann, S., Vingron, M., and Robinson, P.N. (2008). Ontologizer 2.0–a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics *24*, 1650–1651.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266–1276.

Boggon, T.J., Shan, W.S., Santagata, S., Myers, S.C., and Shapiro, L. (1999). Implication of tubby proteins as transcription factors by structure-based functional analysis. Science *286*, 2119–2125.

Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A.M., Grove, C.A., Martinez, N.J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J.S., Hope, I.A., et al. (2006). A gene-centered C. elegans protein-DNA interaction network. Cell *125*, 1193–1205.

Diella, F., Gould, C.M., Chica, C., Via, A., and Gibson, T.J. (2008). Phospho.ELM: a database of phosphorylation sites–update 2008. Nucleic Acids Res. *36*, D240–D244.

Elemento, O., Slonim, N., and Tavazoie, S. (2007). A universal framework for regulatory element discovery across all genomes and data types. Mol. Cell *28*, 337–350.

Elemento, O., and Tavazoie, S. (2005). Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. Genome Biol. *6*, R18.

Hall, D.A., Zhu, H., Zhu, X., Royce, T., Gerstein, M., and Snyder, M. (2004). Regulation of gene expression by a metabolic enzyme. Science *306*, 482–484.

Hartley, J.L., Temple, G.F., and Brasch, M.A. (2000). DNA cloning using in vitro site-specific recombination. Genome Res. *10*, 1788–1795.

Ho, S.W., Jona, G., Chen, C.T., Johnston, M., and Snyder, M. (2006). Linking DNA-binding proteins to their recognition sequences by using protein microarrays. Proc. Natl. Acad. Sci. USA *103*, 9940–9945.

Huang, C., Liu, L.Y., Li, Z.F., Wang, P., Ni, L., Song, L.P., Xu, D.H., and Song, T.S. (2008). Effects of small interfering RNAs targeting MAPK1 on gene expression profile in HeLa cells as revealed by microarray analysis. Cell Biol. Int. *32*, 1081–1090.

Jensen, P.H., and Weilguny, D. (2005). Combination primer polymerase chain reaction for multi-site mutagenesis of close proximity sites. J. Biomol. Tech. *16*, 336–340.

Ji, H., Jiang, H., Ma, W., Johnson, D.S., Myers, R.M., and Wong, W.H. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. Nat. Biotechnol. *26*, 1293–1300.

Kipreos, E.T., and Wang, J.Y. (1992). Cell cycle-regulated binding of c-Abl tyrosine kinase to DNA. Science *256*, 382–385.

Kummerfeld, S.K., and Teichmann, S.A. (2006). DBD: a transcription factor prediction database. Nucleic Acids Res. *34*, D74–D81.

Liang, F., Matrubutham, U., Parvizi, B., Yen, J., Duan, D., Mirchandani, J., Hashima, S., Nguyen, U., Ubil, E., Loewenheim, J., et al. (2004). ORFDB: an information resource linking scientific content to a high-quality Open Reading Frame (ORF) collection. Nucleic Acids Res. *32*, D595–D599.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat. Biotechnol. *20*, 835–839.

McKnight, S.L., Gavis, E.R., Kingsbury, R., and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell *25*, 385–398.

Messina, D.N., Glasscock, J., Gish, W., and Lovett, M. (2004). An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. Genome Res. *14*, 2041–2047.

Nelson, J.D., Denisenko, O., and Bomsztyk, K. (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat. Protoc. *1*, 179–185.

Noyes, M.B., Christensen, R.G., Wakabayashi, A., Stormo, G.D., Brodsky, M.H., and Wolfe, S.A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell *133*, 1277–1289.

Petukhova, G.V., Pezza, R.J., Vanevski, F., Ploquin, M., Masson, J.Y., and Camerini-Otero, R.D. (2005). The Hop2 and Mnd1 proteins act in concert with Rad51 and Dmc1 in meiotic recombination. Nat. Struct. Mol. Biol. *12*, 449–453.

Pokholok, D.K., Zeitlinger, J., Hannett, N.M., Reynolds, D.B., and Young, R.A. (2006). Activated signal transduction kinases frequently occupy target genes. Science *313*, 533–536.

Robinson, M.J., Harkins, P.C., Zhang, J., Baer, R., Haycock, J.W., Cobb, M.H., and Goldsmith, E.J. (1996). Mutation of position 52 in ERK2 creates a non-productive binding mode for adenosine 5′-triphosphate. Biochemistry *35*, 5641–5646.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat. Biotechnol. *16*, 939–945.

Roy, S.K., Hu, J., Meng, Q., Xia, Y., Shapiro, P.S., Reddy, S.P., Platanias, L.C., Lindner, D.J., Johnson, P.F., Pritchard, C., et al. (2002). MEKK1 plays a critical role in activating the transcription factor C/EBP-beta-dependent gene expression in response to IFN-gamma. Proc. Natl. Acad. Sci. USA *99*, 7945–7950.

Roy, S.K., Wachira, S.J., Weihua, X., Hu, J., and Kalvakolanu, D.V. (2000). CCAAT/enhancer-binding protein-beta regulates interferon-induced transcription through a novel element. J. Biol. Chem. *275*, 12626–12632.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. *147*, 195–197.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Teichmann, S.A., and Babu, M.M. (2004). Gene regulatory network growth by duplication. Nat. Genet. *36*, 492–496.

Tokai-Nishizumi, N., Ohsugi, M., Suzuki, E., and Yamamoto, T. (2005). The chromokinesin Kid is required for maintenance of proper metaphase spindle size. Mol. Biol. Cell *16*, 5455–5463.

Weihua, X., Kolla, V., and Kalvakolanu, D.V. (1997). Interferon gamma-induced transcription of the murine ISGF3gamma (p48) gene is mediated by novel factors. Proc. Natl. Acad. Sci. USA *94*, 103–108.

Westfall, P., and Young, S.S. (1993). Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment (New York: Wiley).

Wingender, E., Dietze, P., Karas, H., and Knuppel, R. (1996). TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. *24*, 238–241.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature *434*, 338–345.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. USA *104*, 7145–7150.

Yu, X., Lin, J., Zack, D.J., and Qian, J. (2006). Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. Nucleic Acids Res. *34*, 4925–4936.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., et al. (2001). Global analysis of protein activities using proteome chips. Science *293*, 2101–2105.

Zhu, X.D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J.H., and de Lange, T. (2003). ERCC1/XPF removes the 3′ overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. Mol. Cell *12*, 1489–1498.

LETTERS

HMGB proteins function as universal sentinels for nucleic-acid-mediated innate immune responses

Hideyuki Yanai[1]*, Tatsuma Ban[1]*, ZhiChao Wang[1]*, Myoung Kwon Choi[1], Takeshi Kawamura[2], Hideo Negishi[1], Makoto Nakasato[1], Yan Lu[1], Sho Hangai[1], Ryuji Koshiba[1], David Savitsky[1], Lorenza Ronfani[3], Shizuo Akira[4], Marco E. Bianchi[3], Kenya Honda[1]†, Tomohiko Tamura[1], Tatsuhiko Kodama[2] & Tadatsugu Taniguchi[1]

The activation of innate immune responses by nucleic acids is crucial to protective and pathological immunities and is mediated by the transmembrane Toll-like receptors (TLRs) and cytosolic receptors[1,2]. However, it remains unknown whether a mechanism exists that integrates these nucleic-acid-sensing systems. Here we show that high-mobility group box (HMGB) proteins 1, 2 and 3 function as universal sentinels for nucleic acids. HMGBs bind to all immunogenic nucleic acids examined with a correlation between affinity and immunogenic potential. *Hmgb1*$^{-/-}$ and *Hmgb2*$^{-/-}$ mouse cells are defective in type-I interferon and inflammatory cytokine induction by DNA or RNA targeted to activate the cytosolic nucleic-acid-sensing receptors; cells in which the expression of all three HMGBs is suppressed show a more profound defect, accompanied by impaired activation of the transcription factors interferon regulatory factor 3 (IRF3) and nuclear factor (NF)-κB. The absence of HMGBs also severely impairs the activation of TLR3, TLR7 and TLR9 by their cognate nucleic acids. Our results therefore indicate a hierarchy in the nucleic-acid-mediated activation of immune responses, wherein the selective activation of nucleic-acid-sensing receptors is contingent on the more promiscuous sensing of nucleic acids by HMGBs. These findings may have implications for understanding the evolution of the innate immune system and for the treatment of immunological disorders.

During microbial infection or tissue damage, DNA and RNA potently activate the innate and subsequent adaptive immune responses[1,2]. In mammals, TLR3, TLR7 and TLR9 recognize, respectively, double-stranded RNA, single-stranded and short double-stranded RNAs, and hypomethylated DNA[1–3], whereas the RIG-I-like receptors (RLRs), namely retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene-5 (MDA5) are best known as RNA-sensing receptors in the cytosol[4,5]. In addition, cytosolic DNA-sensing receptors, which include DNA-dependent activator of IRFs (DAI) and absent in melanoma 2 (AIM2), also trigger the innate and adaptive immune systems[6–8]. It has recently been shown that RLRs also participate in the cytosolic DNA-sensing system[3,9–11]. The hallmark of innate immune responses activated by these receptors is the induction of type-I interferons (IFNs), proinflammatory cytokines and chemokines[1], except that by AIM2, which is a critical component of the inflammasome that typically promotes the secretion of interleukin (IL)-1β (ref. 7). So far, no universal or shared mechanism of action for these nucleic-acid-receptor classes is presumed to operate in their activation.

To gain new insights into the nature of the cytosolic DNA-sensing systems, we performed an unbiased biochemical screen to identify proteins involved in DNA recognition on the basis of their direct binding to a B-form DNA, poly(dA-dT)•poly(dT-dA) (B-DNA; ref. 12); the most prominently recovered proteins were HMGB1, HMGB2 and HMGB3 (Supplementary Fig. 1a). HMGB proteins are highly expressed in the nucleus, where they regulate chromatin structure and transcription, but they are also present in the cytosol and in extracellular fluids[13,14]. HMGB1 was recently shown to participate in the activation of several immune receptors, including TLRs[13,15–18].

To examine the direct interaction of HMGBs with nucleic acids, recombinant HMGBs were purified and assayed *in vitro* for their binding to biotin-conjugated B-DNA. Both HMGB1 and HMGB2 were precipitated by immobilized B-DNA, which was inhibited in a dose-dependent manner by non-conjugated B-DNA (Supplementary Fig. 1b). HMGB1–B-DNA binding was, however, quite inefficiently inhibited by the addition of calf thymus-derived or bacteria-derived DNA, each of which only weakly activates the cytosolic-sensing pathways[8,12] (Supplementary Fig. 1b). The addition of double-stranded RNA (polyinosinic-polycytidylic acid; poly(I:C)) or single-stranded RNA (poly(U)), but not imiquimod (R837), a non-nucleic-acid agonist for TLR7, efficiently inhibited HMGB1–B-DNA binding (Supplementary Fig. 1b). In contrast, HMGB2–B-DNA binding was not affected by either poly(I:C) or poly(U) (Supplementary Fig. 1b). Thus, HMGB1 binds to these immunogenic RNAs but HMGB2 does not. Furthermore, a TLR9 agonist CpG-B oligodeoxynucleotide (ODN; refs 19, 20) and antagonist base-free phosphorothioate deoxyribose homopolymer (PS; ref. 21) were the most potent in inhibiting HMGB1–B-DNA binding, which is consistent with a previous report[16] (Supplementary Fig. 1b). In contrast, a weak TLR9 agonist base-free natural phosphodiester deoxyribose homopolymer (PD; ref. 21) showed little, if any, inhibition (Supplementary Fig. 1b). Binding of HMGB1 to biotin-conjugated poly(U), a TLR7 agonist, was also strongly inhibited by the addition of free CpG-B ODN and PS, but not by R837 (Supplementary Fig. 1c; see below). Finally, we also found that HMGB3, expressed in certain cell types[22,23], binds both DNA and RNA (Supplementary Fig. 1d). These results indicate a correlation between the affinity of a type of nucleic acid to HMGB and its immunogenicity.

To study the contribution of HMGBs to the nucleic-acid-mediated activation of innate immune responses, we first examined cells derived from gene-targeted mice for HMGB1 or HMBG2. Mouse embryonic fibroblasts (MEFs) from HMGB1-deficient mice (*Hmgb1*$^{-/-}$) showed a significant defect in messenger RNA induction for type-I IFNs, IL-6 and RANTES in response to cytosolically delivered B-DNA or poly(I:C) at all doses examined, whereas the response to lipopolysaccharide (LPS)

[1]Department of Immunology, Graduate School of Medicine and Faculty of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. [2]Laboratory for System Biology and Medicine, RCAST, University of Tokyo, Komaba 4-6-1, Meguro-ku, Tokyo 153-8904, Japan. [3]Faculty of Medicine, San Raffaele University, via Olgettina 58, 20132 Milan, Italy. [4]Laboratory of Host Defense, WPI Immunology Frontier Research Center, Osaka University, Yamada-oka 3-1, Suita, Osaka 565-0871, Japan. †Present address: Department of Microbiology and Immunology, Graduate School of Medicine, Osaka University, Yamada-oka 2-2, Suita, Osaka 565-0871, Japan.
*These authors contributed equally to this work.

remained unaffected (Fig. 1a and Supplementary Fig. 2a, b). Similar results were obtained with conventional dendritic cells (cDCs) differentiated by culturing *Hmgb1*$^{-/-}$ fetal liver with granulocyte–macrophage colony-stimulating factor (GM-CSF) (Supplementary Fig. 2c). Cytokine gene induction by *Hmgb2*$^{-/-}$ MEFs was defective when stimulated by B-DNA and not poly(I:C), which is consistent with the interaction of HMGB2 with DNA only (Fig. 1b and Supplementary Figs 1b and 2d). Accordingly, in *Hmgb1*$^{-/-}$ MEFs expressing a small interfering RNA (siRNA) that specifically targets HMGB2, type-I IFN gene induction on stimulation with B-DNA, but not with poly(I:C), was decreased further; the residual response is presumed to have been due to HMGB3 (Supplementary Fig. 2e, f).

We then examined MEFs in which the expression of all three HMGB proteins was suppressed by use of pan-HMGB siRNA vector (Supplementary Fig. 2g, h). As shown in Fig. 1c, the mRNA induction for type-I IFNs and other cytokines on stimulation with B-DNA or poly(I:C) was inhibited more strongly in MEFs expressing the pan-HMGB siRNA (HMGB-si-MEFs) than in the *Hmgb1*$^{-/-}$ and *Hmgb2*$^{-/-}$ cells tested above. Similar results were obtained on stimulation of HMGB-si-MEFs with IFN-stimulatory DNA (ISD; refs 2, 24), viral or bacterial DNA, and single-stranded RNA bearing 5′-triphosphates[25,26] (Fig. 1d and Supplementary Fig. 2i). In contrast, the induction of cytokine genes remained unaffected in HMGB-si-MEFs stimulated by LPS, indicating that the inhibition is selective to stimulations with nucleic acids (Fig. 1d and Supplementary Fig. 2i–k). Moreover, mRNA induction of the genes activated by a variety of cytokines occurred normally in these cells, indicating that gene transcription is not generally affected by the absence of HMGBs (Supplementary Fig. 2l, m). Essentially the same observations were also made in RAW264.7 macrophage and NIH/3T3 fibroblast cell lines (Supplementary Fig. 2n–q). In macrophages, cytosolic DNA induces the formation of the inflammasome by the activation of AIM2, which

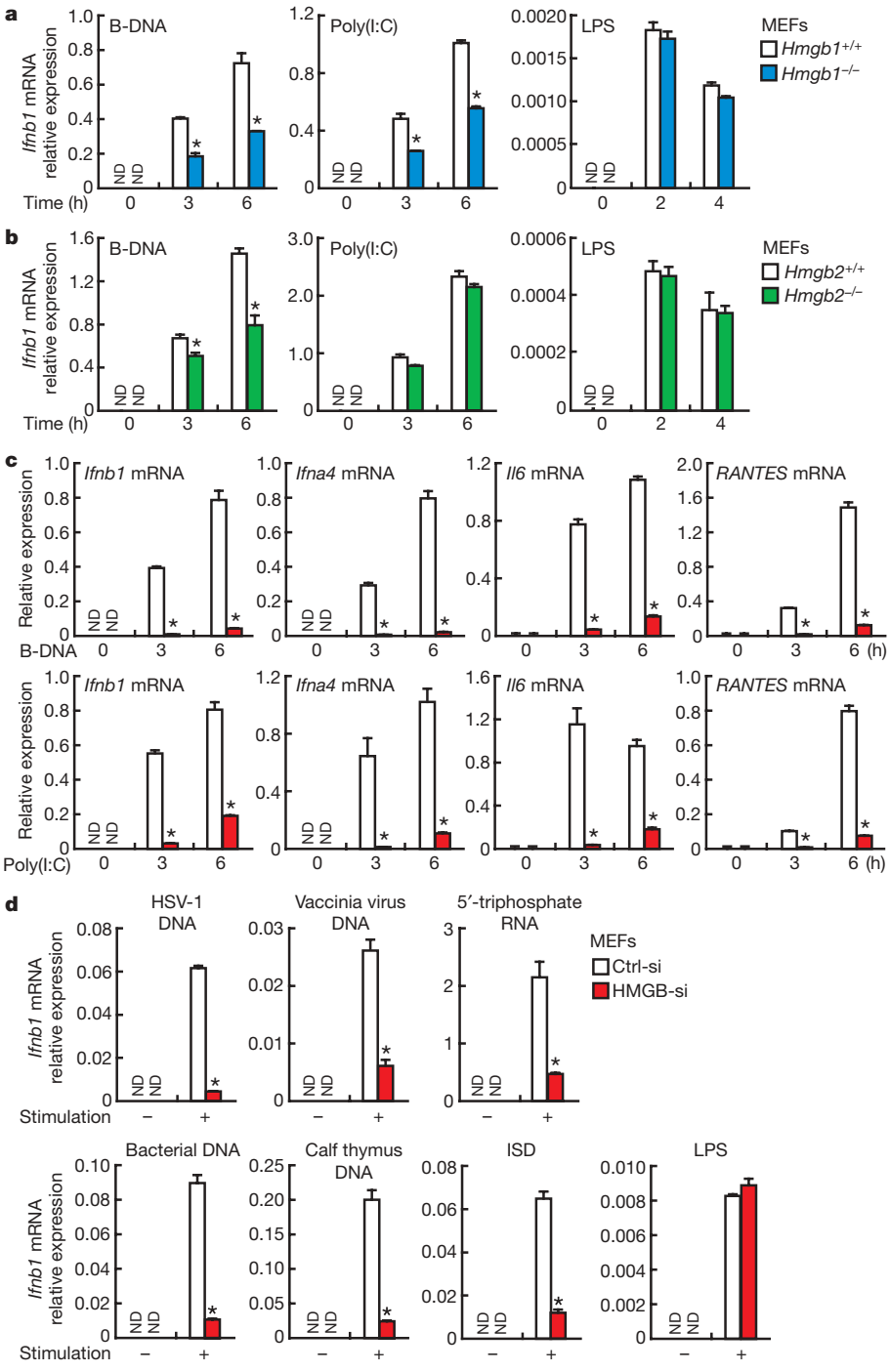


Figure 1 | The contribution of HMGBs to cytosolic DNA-mediated or RNA-mediated activation of innate immune responses. **a**, Hypoinduction of IFN-β in *Hmgb1*$^{-/-}$ cells on cytosolic delivery of DNA or RNA. *Hmgb1*$^{+/+}$ or *Hmgb1*$^{-/-}$ MEFs were stimulated with B-DNA or poly(I:C) for 6 h, or stimulated with LPS (200 ng ml^{-1}) for 2 h. The induction levels of *Ifnb1* mRNA were determined by quantitative real-time RT–PCR (qRT–PCR). Asterisk, $P < 0.01$ compared with *Hmgb1*$^{+/+}$ MEFs. Data in all panels are presented as means and s.d. ($n = 3$). ND, not detected. **b**, Hypoinduction of IFN-β in *Hmgb2*$^{-/-}$ MEFs on cytosolic delivery of DNA but not RNA. Asterisk, $P < 0.001$ compared with *Hmgb2*$^{+/+}$ MEFs. **c**, Impaired innate immune responses to cytosolic nucleic acids in MEFs deficient in all HMGBs. MEFs transduced with retrovirus expressing siRNA targeting all HMGBs (HMGB-si) or control siRNA (Ctrl-si) were lipofected with B-DNA or poly(I:C), and then mRNA expression levels of the indicated genes were measured by qRT–PCR. Asterisk, $P < 0.01$ compared with Ctrl-si-MEFs. **d**, Impaired induction of IFN-β by cytosolic delivery of nucleic acids obtained from various sources at 6 h, but not by stimulation with LPS (200 ng ml^{-1}) for 2 h in MEFs deficient in all HMGBs. Asterisk, $P < 0.01$ compared with Ctrl-si-MEFs.

triggers the secretion of IL-1β (ref. 7). The involvement of HMGBs in the DNA–AIM2–inflammasome pathway was underscored by the observation that B-DNA-induced secretion of IL-1β is significantly impaired in *Hmgb1*$^{-/-}$ fetal liver-derived macrophages and RAW264.7 cells expressing the pan-HMGB siRNA (Supplementary Fig. 3a–c). These results suggest that all three HMGBs are required for the full-blown activation of innate immune responses by cytosolic nucleic acids.

We next examined whether the signalling pathways activated through the cytosolic receptors are affected by the absence of HMGBs. The activation of the IRF3 and NF-κB transcription factors was measured in HMGB-si-MEFs or MEFs expressing *Renilla* luciferase-siRNA (Ctrl-si-MEFs) stimulated by B-DNA or poly(I:C).

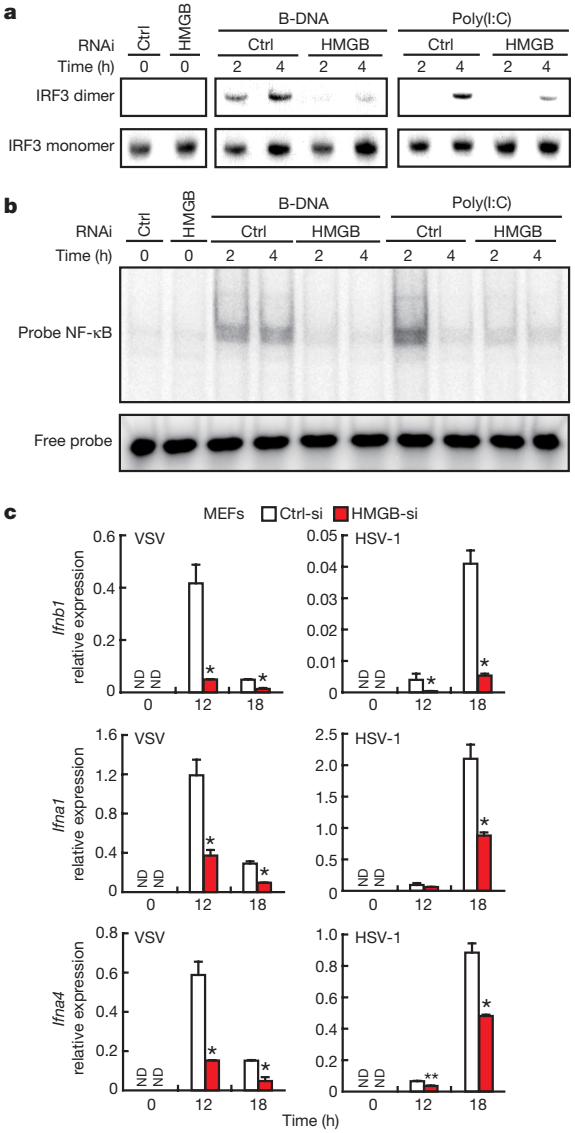


Figure 2 | A requirement for HMGBs in cytosolic nucleic-acid-receptor-mediated activation of signalling pathways and in antiviral innate immune responses. **a**, **b**, Activation status of IRF3 (**a**) and NF-κB (**b**). MEFs expressing siRNA targeting all HMGBs (HMGB-si) or control siRNA (Ctrl-si) were lipofected with B-DNA or poly(I:C). Dimerization of IRF3 was assessed by native PAGE followed by immunoblot analysis. Activation of NF-κB was analysed by electrophoretic mobility-shift assay. **c**, Induction of type-I IFNs by virus infection. MEFs expressing siRNA targeting all HMGBs (HMGB-si) or control siRNA (Ctrl-si) were infected with VSV or HSV-1. *Ifnb1*, *Ifna1* and *Ifna4* mRNA expression levels were monitored by qRT–PCR. Data in **c** are presented as means and s.d. ($n = 3$). Asterisk, $P < 0.01$; double asterisk, $P < 0.05$ compared with cells expressing Ctrl-si. ND, not detected.

As shown in Fig. 2a, the formation of dimerized IRF3, a hallmark of IRF3 activation, was strongly suppressed in HMGB-si-MEFs but robust in Ctrl-si-MEFs on stimulation by B-DNA or poly(I:C). Similarly, the activation of extracellular signal-regulated kinase (ERK) (data not shown) and NF-κB was also suppressed in HMGB-si-MEFs (Fig. 2b). Thus, these results indicate that in the context of the nucleic-acid-mediated activation of innate immune responses, HMGBs function in the cytosolic receptor–IRF3/NF-κB signalling pathways.

Because cytosolic nucleic-acid-sensing systems are crucial to antiviral immunity[1,4], we next examined the contribution of HMGBs to the virus-induced activation of innate immune responses. The gene induction of type-I IFNs and replication of vesicular stomatitis virus (VSV) or herpes simplex virus type 1 (HSV-1) in infected HMGB-si-MEFs and Ctrl-si-MEFs were measured. *Ifnb1* mRNA induced by these viruses was much lower in HMGB-si-MEFs than in Ctrl-si-MEFs (Fig. 2c). However, the suppression of *Ifna* (*Ifna1* and *Ifna4*) mRNA induction was more resistant to HMGB downregulation than *Ifnb1* mRNA (Fig. 2c), which is presumably due to the operation of one or more mechanisms that do not involve the nucleic acid–HMGB pathway[27]. Accordingly, virus replication was increased in HMGB-si-MEFs (Supplementary Fig. 4a). Similar observations were made for RAW264.7 cells expressing the pan-HMGB siRNA (Supplementary Fig. 4b, c). These data, together with those for virus-derived 'naked' DNAs (Fig. 1d), support the hypothesis that HMGB sensing of virus-derived nucleic acids is critical for the effective avocation of the antiviral innate immune responses.

Given that HMGBs (particularly HMGB1) bind to all TLR agonistic nucleic acids (Supplementary Fig. 1b–d), we next examined the roles of HMGBs in the activation of TLRs. A requirement for HMGB1 in TLR9 signalling in cooperation with the multivalent receptor RAGE (receptor for advanced glycation end-products) was reported previously[16,18]. As shown in Fig. 3a, the mRNA induction of proinflammatory cytokines was impaired, albeit partly, in *Hmgb1*$^{-/-}$ cDCs on stimulation with TLR3 (poly(I:C)) or TLR9 (CpG-B ODN), whereas mRNA levels were unaffected on stimulation with TLR4 (LPS). The residual induction is likely to be mediated by other HMGBs. Indeed, when the pan-HMGB siRNA was expressed in RAW264.7 cells, mRNA induction by poly(I:C) or CpG-B ODN was more severely inhibited (Supplementary Fig. 5). We also cultured *Hmgb1*$^{-/-}$ fetal liver with Flt3-ligand to enrich for plasmacytoid dendritic cells (pDCs), and then stimulated the cells with ligands for TLR7 and TLR9 to drive the pDC-mediated high-level production of type-I IFNs. *Ifnb1* and *Ifna4* mRNA induction by TLR9 agonist CpG-A (D19) ODN[20] was inhibited in *Hmgb1*$^{-/-}$ pDCs and CpG-A ODN-stimulated RAW264.7 cells expressing the pan-HMGB siRNA (Fig. 3b and Supplementary Fig. 5). In addition, *Ifnb1* mRNA expression was severely inhibited when *Hmgb1*$^{-/-}$ pDCs were stimulated by the TLR7 agonist poly(U), yet when stimulated by R837, which does not bind to HMGBs (Supplementary Fig. 1b), the expression levels were the same as those in wild-type pDCs (Fig. 3b). The deficiency in HMGB1 had a more profound effect on TLR7 stimulation by poly(U) than on TLR9 stimulation by CpG-A ODN (Fig. 3b), possibly because poly(U) is not bound by HMGB2 (Supplementary Fig. 1b). We speculate that, like TLR9 signalling, TLR3 and TLR7 signalling may also involve RAGE-bound HMGB1 (ref. 18). Thus, these results together underscore the role of HMGBs in the activation of all nucleic-acid-sensing TLRs.

Because CpG-B ODN and the antagonist PS bind to HMGBs with extraordinarily high affinities (Supplementary Fig. 1b, c), we tested whether these compounds interfere with innate immune responses evoked by immunogenic nucleic acids of lower affinities; we reasoned that a functional consequence of the pretreatment of HMGBs with these high-binding-affinity nucleotide analogues would 'mask' HMGBs, thereby selectively interfering with the activation of innate immune responses by other nucleic acids. Indeed, wild-type MEFs were hyporesponsive to stimulation with cytosolic B-DNA or poly(I:C) when pretreated with CpG-B ODN, which itself does not evoke a detectable

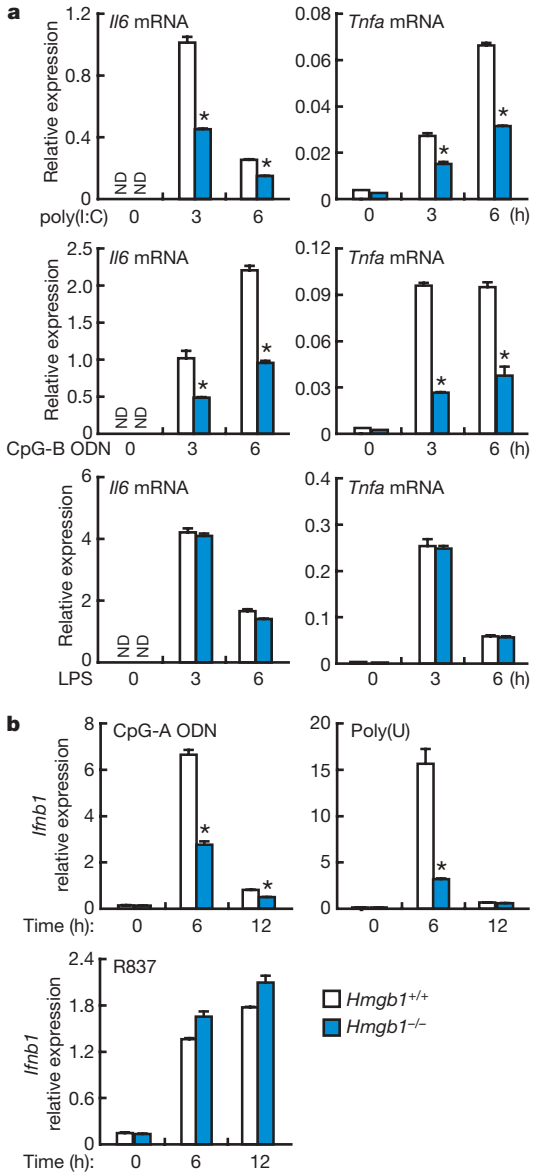


Figure 3 | A requirement for HMGB1 in the nucleic-acid-mediated activation of TLRs. *Hmgb1*$^{+/+}$ or *Hmgb1*$^{-/-}$ cDCs (**a**) and pDCs (**b**) were stimulated by the indicated TLR ligands, and then mRNA expression levels for various cytokines were measured by qRT–PCR. Data in all panels are presented as means and s.d. ($n = 3$). Asterisk, $P < 0.01$ compared with wild-type cells. ND, not detected.

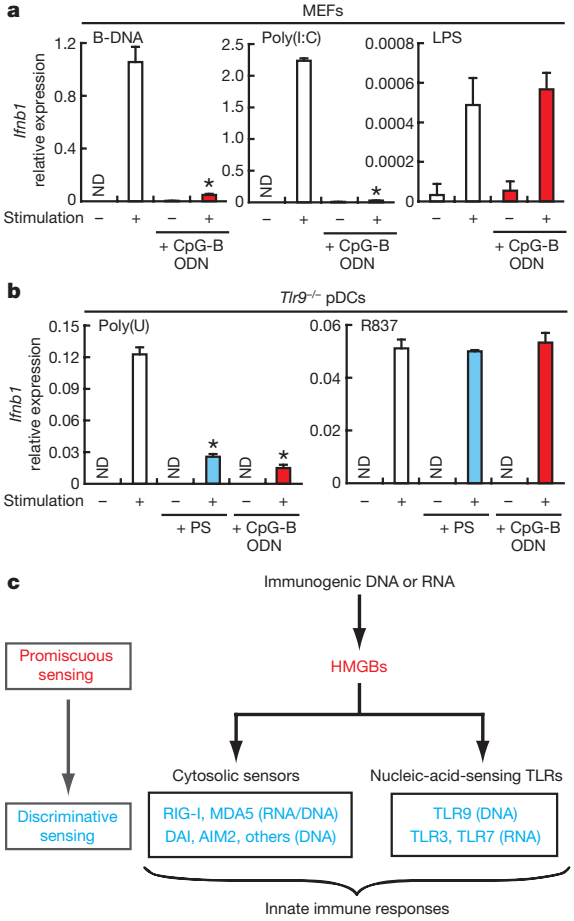


Figure 4 | Interference of HMGB-nucleic-acid-induced activation of innate immune responses by use of high-binding-affinity nucleotide analogues. **a**, MEFs were stimulated by cytosolic delivery of B-DNA, poly(I:C) or LPS with or without pretreatment for 30 min with 1 μM CpG-B ODN. *Ifnb1* mRNA expression levels were then monitored by qRT–PCR. Data in all panels are means and s.d. ($n = 3$). Asterisk, $P < 0.01$, pretreated cells versus non-pretreated cells. ND, not detected. **b**, Bone marrow-derived *Tlr9*$^{-/-}$ pDCs were either lipofected with 1 μg ml^{-1} poly(U) or stimulated with 25 μg ml^{-1} R837 for 8 h, in the absence of pretreatment or after pretreatment with 5 μM PS or 1 μM CpG-B ODN for 30 min. The mRNA expression levels of the indicated genes were measured by qRT–PCR. Asterisk, $P < 0.01$, pretreated cells versus non-pretreated cells. **c**, A schematic view of the hierarchy in the nucleic-acid-mediated activation of immune responses. All immunogenic nucleic acids bind HMGBs (promiscuous sensing), which is required for subsequent recognition by specific pattern recognition receptors (discriminative sensing) to activate the innate immune responses.

cytokine response in this cell type, whereas their response to LPS remained normal (Fig. 4a). As expected, the suppressive effect of CpG-B ODN was weaker if added after stimulation with B-DNA (Supplementary Fig. 6a). Furthermore, even in TLR9-deficient pDCs and cDCs, a profound hyporesponse was observed on pretreatment with CpG-B ODN or PS followed by stimulation by TLR3 or TLR7 with their cognate nucleic acid ligands (Fig. 4b). In contrast, activation of TLR7 by R837 was not affected by pretreatment with CpG-B ODN or PS (Fig. 4b and Supplementary Fig. 6b). Consistent with this, when B-DNA was used as an adjuvant the induction of ovalbumin-specific CD8 T-cell responses was severely impaired by the pretreatment of TLR9-deficient mice with CpG-B (M. Miyajima and T.T., unpublished observation). These results therefore lend further support to the universal role of HMGBs in the nucleic-acid-mediated innate immune responses, although strictly the possibility is not ruled out that the CpG-B ODN-mediated or PS-mediated interference with nucleic acid signalling described here may involve additional mechanisms.

We propose, and our data support, that HMGBs serve as universal sentinels for nucleic acids that are required for the full-blown, nucleic-acid-induced activation of innate immune responses mediated by the more discriminative pattern recognition receptors (Fig. 4c). What is the biological significance of the HMGB's function that we describe here? Because HMGBs are broadly found in multicellular organisms including plants[28], one may speculate that HMGBs are a remnant of a primitive sensing system for foreign nucleic acids, which has evolved along with and in support of discriminative TLRs and cytosolic receptors of the innate immune signalling system. In this regard, it is worth noting that HMGB1 associates with TLR9 (refs 16, 18) and RLRs (Z.W. and T.T., unpublished observation), suggesting that nucleic-acid-bound HMGBs have evolved to function as co-ligands for discriminative nucleic-acid receptors; in other words, the binding of nucleic acids to HMGBs is a precondition for the more efficient, subsequent recognition by and activation of TLR, RLR and cytosolic nucleic-acid receptors (Fig. 4c).

Nevertheless, it remains unclear exactly where HMGBs bind to nucleic acids and how the HMGB–nucleic acid complexes activate their respective receptor signalling cascades. *In vivo* visualization data lend support to the concept that HMGB1 functions together with RIG-I on the endosomal membrane, through which nucleic acids are released into the cytosol by low-pH-dependent membrane fusion[29,30] (Supplementary Fig. 7). In view of a previous report showing the interaction between HMGB1 and TLR9 in the endosomal compartment[16], we conjecture that endosome membrane-associated signalling may be universal to nucleic-acid-mediated innate immune responses. In addition to their role in nucleic-acid sensing described here, HMGBs serve additional functions including gene transcription, all of which may be viewed as a long-term evolutionary consequence of their ability to bind nucleic acids. Nevertheless, further study is required for a more complete understanding of the precise role and mechanism of action of these proteins. Finally, our findings have a number of interesting therapeutic implications, such as in the use of HMGB inhibitory compounds in the suppression of nucleic-acid-mediated pathological immune responses.

METHODS SUMMARY

Pull-down assay. MEFs were treated with poly(dA-dT)•(dT-dA) (B-DNA; 10 μg ml^{-1}) for 4 h before mass spectrometry analysis. After stimulation, cells were homogenized with a Dounce homogenizer (Wheaton Science) in homogenization buffer (20 mM HEPES pH 7.4, 20% glycerol, 50 mM KCl, 2 mM $MgCl_2$, 1 mM phenylmethylsulphonyl fluoride (PMSF), 10 μg ml^{-1} aprotinin, 10 μg ml^{-1} leupeptin). Cytosolic protein extract was prepared from the homogenate by centrifugation at 19,000*g* for 30 min, and then incubated with 1.4 μg ml^{-1} B-DNA, the 5′ terminus of which had been conjugated with biotin, and streptavidin (SA)-conjugated magnetic beads (Invitrogen) at 4 °C for 15 min. The 'pulled-down' sample was treated with DNase I (Invitrogen) in reaction buffer (20 mM Tris-HCl pH 8.4, 20 mM $MgCl_2$, 50 mM KCl) and the resultant supernatant was subjected to silver staining (Invitrogen) or mass spectrometry analysis. The *in vitro* pull-down assay was performed as essentially described previously[8]. Recombinant HMGB1, HMGB2 and HMGB3 proteins were first incubated for 30 min at 15–20 °C with or without competitors. The supernatants were then mixed for 30 min at 4 °C with biotin-conjugated B-DNA after preincubation with SA-conjugated magnetic beads in binding buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Nonidet P40, 100 μg ml^{-1} leupeptin, 1 mM PMSF, 1 mM Na_3VO_4). The mixture was then washed extensively with the same buffer, separated by SDS–PAGE and immunoblotted with anti-HMGB1, anti-HMGB2 and anti-HMGB3 antibodies.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 August; accepted 17 September 2009.

1. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124,** 783–801 (2006).
2. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449,** 819–826 (2007).
3. Hornung, V. *et al.* Sequence-specific potent induction of IFN-α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nature Med.* **11,** 263–270 (2005).
4. Kato, H. *et al.* Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441,** 101–105 (2006).
5. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nature Immunol.* **5,** 730–737 (2004).
6. Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451,** 725–729 (2008).
7. Schroder, K., Muruve, D. A. & Tschopp, J. Innate immunity: cytoplasmic DNA sensing by the AIM2 inflammasome. *Curr. Biol.* **19,** R262–R265 (2009).
8. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448,** 501–505 (2007).
9. Choi, M. K. *et al.* A selective contribution of the RIG-I-like receptor pathway to type I interferon responses activated by cytosolic DNA. *Proc. Natl Acad. Sci. USA* advance online publication, doi:10.1073/pnas.0910983106 (1 October 2009).
10. Chiu, Y. H., Macmillan, J. B. & Chen, Z. J. RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell* **138,** 576–591 (2009).
11. Ablasser, A. *et al.* RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III–transcribed RNA intermediate. *Nature Immunol.* **10,** 1065–1072 (2009).
12. Ishii, K. J. *et al.* A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nature Immunol.* **7,** 40–48 (2006).
13. Bianchi, M. E. & Manfredi, A. A. High-mobility group box 1 (HMGB1) protein at the crossroads between innate and adaptive immunity. *Immunol. Rev.* **220,** 35–46 (2007).
14. Scaffidi, P., Misteli, T. & Bianchi, M. E. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* **418,** 191–195 (2002).
15. Apetoh, L. *et al.* The interaction between HMGB1 and TLR4 dictates the outcome of anticancer chemotherapy and radiotherapy. *Immunol. Rev.* **220,** 47–59 (2007).
16. Ivanov, S. *et al.* A novel role for HMGB1 in TLR9-mediated inflammatory responses to CpG-DNA. *Blood* **110,** 1970–1981 (2007).
17. Kazama, H. *et al.* Induction of immunological tolerance by apoptotic cells requires caspase-dependent oxidation of high-mobility group box-1 protein. *Immunity* **29,** 21–32 (2008).
18. Tian, J. *et al.* Toll-like receptor 9-dependent activation by DNA-containing immune complexes is mediated by HMGB1 and RAGE. *Nature Immunol.* **8,** 487–496 (2007).
19. Krieg, A. M. *et al.* CpG motifs in bacterial DNA trigger direct B-cell activation. *Nature* **374,** 546–549 (1995).
20. Klinman, D. M. Immunotherapeutic uses of CpG oligodeoxynucleotides. *Nature Rev. Immunol.* **4,** 249–258 (2004).
21. Haas, T. *et al.* The DNA sugar backbone 2′ deoxyribose determines Toll-like receptor 9 activation. *Immunity* **28,** 315–323 (2008).
22. Nemeth, M. J. *et al.* Hmgb3: an HMG-box family member expressed in primitive hematopoietic cells that inhibits myeloid and B-cell differentiation. *Blood* **102,** 1298–1306 (2003).
23. Vaccari, T., Beltrame, M., Ferrari, S. & Bianchi, M. E. *Hmg4*, a new member of the *Hmg1/2* gene family. *Genomics* **49,** 247–252 (1998).
24. Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24,** 93–103 (2006).
25. Pichlmair, A. & Reis e Sousa, C. Innate recognition of viruses. *Immunity* **27,** 370–383 (2007).
26. Hornung, V. *et al.* 5′-Triphosphate RNA is the ligand for RIG-I. *Science* **314,** 994–997 (2006).
27. Paladino, P., Cummings, D. T., Noyce, R. S. & Mossman, K. L. The IFN-independent response to virus particle entry provides a first line of antiviral defense that is independent of TLRs and retinoic acid-inducible gene I. *J. Immunol.* **177,** 8008–8016 (2006).
28. Grasser, K. D., Launholt, D. & Grasser, M. High mobility group proteins of the plant HMGB family: dynamic chromatin modulators. *Biochim. Biophys. Acta* **1769,** 346–357 (2007).
29. Kampmann, T., Mueller, D. S., Mark, A. E., Young, P. R. & Kobe, B. The role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* **14,** 1481–1487 (2006).
30. Weissenhorn, W., Hinz, A. & Gaudin, Y. Virus membrane fusion. *FEBS Lett.* **581,** 2150–2155 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Reis e Sousa and J. Rehwinkel for 5′-triphosphate RNA;, Y. Kawaguchi for HSV-1 DNA; A. Katoh and M. Kidokoro for the vaccinia virus (MO) genome for ligand stimulation; J. Vilcek, R. Medzhitov, R. Schekman, J. V. Ravetch and H. Ohno for invaluable advice; and R. Takeda, M. Shishido, Y. Nakaima and T. Oh for technical assistance. This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas 'Integrative Research Toward the Conquest of Cancer,' Grant-in-Aid for Scientific Research (A), and Global Center of Excellence Program 'Integrative Life Science Based on the Study of Biosignaling Mechanisms' from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and by the Korea Science and Engineering Foundation Grant. T.B. and Z.W. are research fellows of the Japan Society for the Promotion of Science.

Author Contributions H.Y., T.B., Z.W. and M.C. designed, performed and analysed most of the experiments in this study. T. Kawamura and T. Kodama collected data and performed the mass spectrometry analysis. H.N. performed experiments on RAW264.7 cells. M.N. performed the fluorescence microscopy analysis. Y.L. and R.K. cloned the constructs. S.H. assisted with the virus infection experiments. A.S., L.R. and M.B. provided precious materials and invaluable advice. K.H., T. Tamura and T. Taniguchi provided overall coordination with respect to conception, design and supervision of the study. H.Y., T.B., D.S., T. Tamura and T. Taniguchi wrote the manuscript with comments from the co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T. Taniguchi (tada@m.u-tokyo.ac.jp).

METHODS

Mice, cells and reagents. Mice used for this study were on the genetic background of C57BL/6, except *Tlr9*$^{-/-}$ mice, which were on the Balb/c background. The generation of *Tlr9*$^{-/-}$, *Hmgb1*$^{-/-}$ and *Hmgb2*$^{-/-}$ mice were described previously[12,31,32]. MEFs, RAW264.7, NIH/3T3 and HEK293T cells, bone marrow-derived cDCs and pDCs from *Tlr9*$^{-/-}$ mice were maintained as described previously[8,33]. *Hmgb1*$^{-/-}$ macrophages, cDCs and pDCs were generated from fetal liver haematopietic progenitors (lineage-marker-negative cells purified with the MACS Lineage depletion kit from Miltenyi Biotec) by culturing fetal liver cells for 2 days in the presence of SCF (20 ng ml^{-1}), IL-3 (10 ng ml^{-1}) and IL-6 (10 ng ml^{-1}) followed by a 6-day culture in the presence of 20 ng ml^{-1} macrophage colony-stimulating factor (for macrophages), 20 ng ml^{-1} GM-CSF (for cDCs) or 100 ng ml^{-1} human Flt3-ligand (for pDCs). SCF, IL-3 and IL-6 were purchased from Peprotech. IFN-γ and TNF-α were purchased from R&D Systems. IFN-β was provided by TORAY Industries. B-DNA and calf thymus genomic DNA were purchased from Sigma. Biotin-conjugated poly(dA-dT)•poly(dT-dA) was purchased from Hokkaido System Science. Other oligo DNAs including ISD[24], CpG ODNs, fluorescein isothiocyanate-conjugated base-free phosphorothioate deoxyribose homopolymer (PS; 20-mer) and fluorescein isothiocyanate-conjugated base-free natural deoxyribose homopolymer (PD; 20-mer) were purchased from Fasmac. *E. coli* DNA and R837 were purchased from InvivoGen. Poly(U) and LPS were purchased from Sigma. Poly(I:C) was purchased from GE Healthcare Biosciences. B-DNA, poly(I:C) and other nucleic acid ligands were used at a concentration of 10 μg ml^{-1} as described previously[8], unless otherwise mentioned. The complex formation of CpG-A ODN with DOTAP (Roche) was performed as described previously[33]. MitoTracker Deep Red 633 was purchased from Invitrogen. Antibodies against the following proteins were used: HMGB1 and HMGB2 (Abcam), HMGB3 (TransGenic Inc.), IRF3 (ZM3; Zymed), β-actin (AC-15; Sigma-Aldrich), NF-κB p65 (C20; Santa Cruz Biotechnology) and phospho-STAT1 (58D6; Cell Signaling).

Plasmid constructions and protein purification. Murine HMGB cDNAs were obtained by RT–PCR on total RNA from MEFs, and then cloned into pGEX4T3 vector (GE Healthcare Biosciences) at the *Sal*I and *Not*I sites. The glutathione *S*-transferase (GST)-tagged HMGB proteins were purified by using glutathione-Sepharose beads (GE Healthcare Biosciences). HMGB proteins and GST protein were separated by treatment with thrombin proteinase (Novagen). Other expression vectors are listed in the Supplementary Information.

Immunoblot analysis. Cell lysis and immunoblot analysis were performed as described previously[8]. IRF3 dimer was assessed by native PAGE, followed by immunoblot analysis with anti-mouse IRF3 antibody as described previously[8]. The quantification of IRF3 dimer was performed by the NIH Image application. Similar results were obtained in three independent experiments.

RNA analysis. RNA extraction and reverse transcription were performed as described previously[8]. qRT–PCR analysis was performed with a Lightcycler480 and SYBR Green system (Roche Molecular Biochemicals). All data are presented as relative expression units after normalization to GAPDH. Data are presented as means and s.d. for triplicate determinations. All data were reproduced in two or more additional independent experiments. Primer sequences for murine GAPDH, IL-6, RANTES, IκB-α, IFN-α1, IFN-α4 and IFN-β have been described previously[8]. Other primer sequences are listed in the Supplementary Information.

Statistical analysis. Differences between control and experimental groups were evaluated with Student's *t*-test.

Enzyme-linked immunosorbent assay (ELISA). Murine IFN-β, IL-6 or IL-1β was measured by ELISA. IFN-β ELISA kit was purchased from PBL Biomedical Laboratories. IL-6 and IL-1β ELISA kits were obtained from R&D Systems. All data were reproduced in two additional independent experiments.

RNA interference. siRNA vectors were constructed by inserting oligonucleotides into *Eco*RI and *Hin*dIII sites of the pSUPER.retro.puro retrovirus vector[9]. The siRNA targeting sequences for murine HMGB1/2/3 (pan-HMGB siRNA), HMGB2 and *Renilla* luciferase (control) are 5′-GTATGAGAAGGATATTGCT-3′, 5′-GCGTTACGAGAAACCAGTT-3′ and 5′-GTAGCGCGGTGTATTATACA-3′, respectively. Retroviruses were prepared as described previously[9]. Transduced cells were selected by puromycin (Sigma) at 2 μg ml^{-1} (MEFs) or 4 μg ml^{-1} (RAW264.7 cells) for 48 h.

Electrophoretic mobility-shift assay. Electrophoretic mobility-shift assay was performed as described previously[8]. An oligonucleotide probe containing a consensus NF-κB binding sequence was used[8]. The presence of p65 in the NF-κB–DNA binding complex was also confirmed by detection of a supershifted band with an anti-p65 antibody (data not shown).

Viral infection. Cells were infected for 12 h with 1.0 multiplicity of infection of HSV-1 or VSV, as described previously[8]. To measure the yield of HSV-1 or VSV, a plaque-forming assay was performed as described previously[8]. All data were reproduced in two additional independent experiments. Virus preparation was described previously[8].

31. Calogero, S. *et al.* The lack of chromosomal protein Hmg1 does not disrupt cell growth but causes lethal hypoglycaemia in newborn mice. *Nature Genet.* **22,** 276–280 (1999).
32. Ronfani, L. *et al.* Reduced fertility and spermatogenesis defects in mice lacking chromosomal protein Hmgb2. *Development* **128,** 1265–1273 (2001).
33. Honda, K. *et al.* Spatiotemporal regulation of MyD88–IRF-7 signalling for robust type-I interferon induction. *Nature* **434,** 1035–1040 (2005).