Oncogene (2003) 22, 3307–3318

www.nature.com/onc

# Large-scale identification and characterization of human genes that activate NF-κB and MAPK signaling pathways

Akio Matsuda[1,2,*], Yutaka Suzuki[2], Goichi Honda[1], Shuji Muramatsu[1], Osamu Matsuzaki[1], Yukiko Nagano[1], Takahiro Doi[3], Kunitada Shimotohno[3], Takeshi Harada[4], Eisuke Nishida[4], Hiroshi Hayashi[1] and Sumio Sugano[2]

[1]*Laboratory for Biology, Institute for Life Science Research, Health Care Company, ASAH1 KASEI Corporation, 2-1 Samejima, Fuji-shi, Shizuoka 416-8501, Japan;* [2]*Laboratory of Genome Structure Analysis, Human Genome Center, Institute of Medical Science, the University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan;* [3]*Department of Viral Oncology, Institute for Virus Research, Kyoto University, 53 Kawara-cho Shogo-in, Sakyo-ku, Kyoto 606-8507, Japan;* [4]*Department of Cell and Developmental Biology, Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan*

**We have carried out a large-scale identification and characterization of human genes that activate the NF-κB and MARK signaling pathways. We constructed full-length cDNA libraries using the oligo-capping method and prepared an arrayed cDNA pool consisting of 150 000 cDNAs randomly isolated from the libraries. For analysis of the NF-κB signaling pathway, we introduced each of the cDNAs into human embryonic kidney 293 cells and examined whether it activated the transcription of a luciferase reporter gene driven by a promoter containing the consensus NF-κB binding sites. In total, we identified 299 cDNAs that activate the NF-κB pathway, and we classified them into 83 genes, including 30 characterized activator genes of the NF-κB pathway, 28 genes whose involvement in the NF-κB pathways have not been characterized and 25 novel genes. We then carried out a similar analysis for the identification of genes that activate the MARK pathway, utilizing the same cDNA resource. We assayed 145 000 cDNAs and identified 57 genes that activate the MARK pathway. Interestingly, 27 genes were overlapping between the NF-κB and the MAPK pathways, which may indicate that these genes play cross-talking roles between these two pathways.**
*Oncogene* (2003) **22,** 3307–3318. doi:10.1038/sj.onc.1206406

**Keywords:** full-length cDNA; signaling pathways; NF-κB; MAPK; oligo-capping

## Introduction

Cells respond to diverse extracellular stimuli by transducing extracellular signals from cell-surface receptors to cytoplasmic molecules. Within the cell, complex networks of intermediate molecules sequentially transfer the signals from the plasma membrane into the nucleus and eventually activate transcription factors that induce new gene expression. It has been a central theme of molecular biology to delineate the signaling pathways involved in a particular biological response of the cell and to identify the genes involved in the pathways.

The signaling pathway, which leads to the eventual activation of transcription factor NF-κB, plays a pivotal role in immune and inflammatory responses. NF-κB is activated in response to TNFα, IL-1, bacterial lipopolysaccharide (LPS) and various growth factors (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id = 164011). Recent intensive efforts have delineated how extracellular signals are transmitted to NF-κB in some detail (Verma and Stevenson, 1997; Karin, 1999). In most circumstances, NF-κB is sequestered in the cytoplasm bound to any of the inhibitors collectively called IκB (Jacobs and Harrison, 1998). Various external stimuli lead to the phosphorylation of IκB by IκB kinases (IKKs), which results in its ubiquitination and proteolytic degradation at the proteasome. Then NF-κB, which is released from IκB, is allowed to translocate into the nucleus, bind to its cognate DNA elements and induce the transcription of various cytokines, their receptors and other proinflammatory genes such as TNFα, IL-2, IL-6, IL-8, IL-2R, IFNγ and VCAM-1 (Baeuerle and Baltimore, 1996; Baldwin, 1996; Ghosh *et al*., 1998; Barkett and Gilmore, 1999). Since greater knowledge of the NF-κB pathway is expected to shed light on the pathogenesis of human inflammatory diseases such as rheumatoid arthritis and asthma and in certain types of cancer and may identify new targets for pharmacological intervention, the molecular mechanism of this signaling pathway has attracted widespread interest (Yin *et al*., 1998; Baldwin, 2001; Reed, 2001).

However, even in this relatively well-characterized signaling pathway, the mechanisms by which distinct pathways from various receptors converged onto IκB are not fully understood and are still being actively investigated. Recent evidence suggests that crosstalk with other pathways plays an important role in the

*Correspondence: A Matsuda;
E-mail: matsuda.ab@om.asahi-kasei.co.jp
The nucleotide sequences reported in this paper have been deposited in the DDBJ, EMBL and Genbank database under the Accession numbers AB97000–AB97053
Received 28 October 2002; revised 17 January 2003; accepted 20 January 2003

regulation of the NF-κB pathway as well. For example, inhibition of the kinase activity of p38, which is the most downstream kinase of the MAPK pathway, significantly inhibits NF-κB-dependent gene expression, as the translocation of NF-κB to the nucleus remains intact (Stein e*t al*., 1993; Beyaert *et al*., 1996; Schmitz *et al*., 2001; Schulze-Osthoff *et al*., 1997). Although much progress has been made, our knowledge of most of the signaling pathways and their crosstalk is still incomplete.

In order to enhance our understanding of various signaling pathways and their crosstalk, genomewide identification and characterization of genes involved in each of the pathways should be invaluable. Although recent papers using microarray analysis have begun to describe changes in gene expression patterns invoked by activation of a certain transcription factor(s) or a signaling pathway(s), almost no paper has focused on the molecules that are directly involved in the signaling pathway itself. For that purpose, we initiated a systematic approach to analyse signaling pathways. In order to carry out such an analysis of the NF-κB pathway, we constructed full-length cDNA libraries using the oligo-capping method, which we developed previously (Maruyama and Sugano, 1994; Suzuki *et al*., 1997), and prepared an arrayed cDNA pool consisting of 150 000 cDNA clones randomly isolated from the libraries. Each of the cDNAs was transiently introduced into human embryonal kidney (HEK) 293 cells and assayed as to whether it activated NF-κB in a luciferase reporter gene assay. In order to facilitate the reporter gene assay, it was essential to use full-length cDNA libraries, since cDNA libraries constructed by conventional methods mostly consist of truncated cDNAs that lack the initiator ATG and are thus useless for directing protein expression. We then applied the same approach for the analysis of the MAPK pathway. Here, we report our first large-scale identification and characterization of human genes that activate the NF-κB and the MAPK signaling pathways.

## Results and discusssion

### *Construction of full-length cDNA libraries*

We constructed full-length cDNA libraries from human normal lung fihroblast (HNLF) cells cultured without (Library 1) or with (Library 2) previous exposure to IL-4 and TNFα. In order to cover the longer mRNA population, we also constructed a size-fractionated cDNA library using Library 1 (Library 3). For characterization of the libraries, we determined the one-pass sequences of 96 clones isolated from each library. Overall, the libraries contained about 70% full-length cDNAs, with the average cDNA insert sizes being 1.5 kb for Libraries 1 and 2 and 3.3 kb for Library 3 (Table 1). Since each cDNA insert was cloned downstream of a eucaryotic promoter. SRα, which is the promoter of SV40 large T antigen, it could be directly expressed when introduced into mammalian cells. We randomly isolated 150 000 cDNA clones from the libraries (60 000, 60 000 and 30 000 from Libraries 1, 2 and 3, respectively) and prepared an arrayed cDNA pool.

### *Genomewide identification of the full-length cDNAs that activate the NF-κB pathway*

In order to identify genes that activate the NF-κB pathway, each of the cDNAs was introduced into HEK293 cells and examined as to whether it activated the transcription of NF-κB in a luciferase reporter gene assay. We used pNF-κB-Luc, which contains four tandemly repeated NF-κB binding sites, as a reporter plasmid. The cDNA clones that induced the luciferase activity more than fourfold compared to the mock insert plasmid were defined as 'positive' clones. In total, we isolated 299 'positive' cDNAs out of 150 000 cDNA clones.

The isolated cDNA clones were sequenced from both the ends and BLAST searches were performed against the Genbank nonredundant (nr) database. Among the 299 cDNAs, 245 were cDNAs of 'Named' genes and 54 were cDNAs of 'Novel' genes. These cDNAs could be clustered into 58 nonredundant 'Named' genes and the 25 'Novel' genes. In order to further characterize the 'Named' genes, we searched the literature for evidence about whether their involvement in the NF-κB pathway had been reported so far. Among the 58 'Named' genes, 30 had been reported to activate NF-κB ('Characterized Genes'). Regarding the other 28 genes, their involvement in the NF-κB pathway had not been reported ('Uncharacterized Genes'). The complete lists of the genes belonging to each category are shown in Tables 2 ('Characterized Genes'), 3 ('Uncharacterized Genes') and 4 ('Novel Genes').

**Table 1** Full-length cDNA libraries used for the screening

| | | *Known* | | | | | |
|---|---|---|---|---|---|---|---|
| | *Total* | *Full* | *Truncated* | *Not specified* | *% Full* | *Novel* | *Insert size (kb)* |
| Library 1 | 96 | 42 | 33 | 6 | 66% | 26 | 1.6 |
| Library 2 | 96 | 53 | 22 | 7 | 71% | 14 | 1.4 |
| Library 3 | 96 | 39 | 17 | 23 | 71% | 17 | 3.3 |

As for each of the libraries, 96 cDNAs were randomly selected and their 5′-end sequences were determined. The determined cDNA sequences were searched against RefSeq (http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html) with the cutoff value of 1.0e−100. Among the cDNAs that matched RefSeq cDNAs ('Known'), the cDNAs which covered the reported translation initiator ATG were tentatively categorized as 'Full'. Those that lacked the initiator ATGs were categorized as 'Truncated'. The cDNAs that corresponded to different splice isoforms of the reported ones were categorized as 'Not-specified'

**Table 2** Complete list of the screened 'Characterized Genes'

| *Gene* | *Fold induction* | *Library 1* | *Library 2* | *Library 3* | *Total* | *Locus ID* | *OM1M* | *Gene function* |
|---|---|---|---|---|---|---|---|---|
| TNF receptor | 680 | 18 | 15 | 1 | 34 | 7132 | 191190 | Cytokine receptor |
| MyD88 | 430 | 7 | 10 | 2 | 19 | 4615 | 602170 | Adaptor |
| TRADD | 298 | 5 | 3 | 0 | 8 | 8717 | 603500 | Adaptor |
| TRAF6 | 285 | 0 | 1 | 0 | 1 | 7189 | 602355 | Adaptor |
| FADD | 266 | 2 | 3 | 0 | 5 | 8772 | 602457 | Adaptor |
| RICK | 200 | 2 | 3 | 0 | 5 | 8767 | 603455 | Adaptor, kinase |
| Thrombin receptor | 195 | 1 | 0 | 0 | 1 | 2149 | 187930 | GPCR |
| LTBR | 153 | 4 | 1 | 1 | 6 | 4055 | 600979 | Cytokine receptor |
| RhoA | 136 | 38 | 19 | 1 | 58 | 387 | 165390 | Small G protein |
| Edg2 | 139 | 9 | 4 | 0 | 13 | 1902 | 602282 | GPCR |
| Caspase 8 | 122 | 1 | 0 | 0 | 1 | 841 | 601763 | Proteinase |
| Rel | 117 | 1 | 0 | 0 | 1 | 5966 | 164910 | Nuclear factor |
| TRAIL | 93 | 0 | 10 | 0 | 10 | 8743 | 603598 | Cytokine |
| CIIKS/Act1 | 83 | 1 | 0 | 0 | 1 | 10757 | NR | Adaptor |
| CD40 | 68 | 0 | 1 | 0 | 1 | 958 | 109535 | Cytokine receptor |
| RIP | 57 | 0 | 1 | 0 | 1 | 8737 | 603453 | Adaptor |
| ICE | 53 | 1 | 0 | 0 | 1 | 834 | 147678 | Protease |
| TRAF5 | 49 | 0 | 1 | 0 | 1 | 7188 | 602356 | Adaptor |
| HTR2B | 43 | 2 | 1 | 0 | 3 | 3357 | 601122 | GPCR |
| CIPER/Bcl-10 | 36 | 1 | 0 | 0 | 1 | 8915 | 603517 | CARD |
| TAB2 | 24 | 1 | 1 | 0 | 2 | 23118 | 605101 | Adaptor |
| IKK-ε | 23 | 0 | 0 | 1 | 1 | 9641 | 605048 | Kinase |
| RelA | 21 | 0 | 1 | 0 | 1 | 5970 | 164014 | Nuclear factor |
| 1AP2 | 20 | 2 | 0 | 0 | 2 | 329 | 601712 | Protease inhibitor |
| Nodl/CARD4 | 17 | 0 | 0 | 1 | 1 | 10392 | 605980 | CARD |
| FLICE-like IP | 14 | 1 | 1 | 0 | 2 | 8837 | 603599 | Protenase inhibitor |
| MEKK3 | 12 | 0 | 0 | 1 | 1 | 4215 | NR | Kinase |
| NAK/TBK1 | 11 | 1 | 1 | 0 | 2 | 29110 | 604834 | Kinase |
| FAS Ligand | 11 | 6 | 1 | 0 | 7 | 356 | 134638 | Cytokine |
| TRA1LR2 | 6 | 1 | 0 | 2 | 3 | 8795 | 603612 | Cytokine receptor |

The genes previously characterized as activators of the NF-κB pathway and identified in this study were listed. The fold inductions were calculated as the induced luciferase activities compared to the mock insert plasmid (pME18S-FL3). The numbers of the cDNAs isolated from the corresponding libraries were shown in the third to the sixth columns and the length of the full-length cDNAs and the deduced amino acid were in the seventh and the eighth columns. The LocusLink ID, OMIM ID and the reported gene functions are also presented in the ninth to the 11th columns. NR: No record

For all of the 'Novel Genes' and 'Uncharacterized Genes', whose roles in the NF-κB pathway have not been characterized, we repeated the reporter gene assay and confirmed that the fold induction of the luciferase gene activity was increased in proportion to the amount of introduced plasmid. The expected correlation was observed in most cases (Supplementary Information Figure 1). In the rare cases, in which such dose dependency was not observed, the corresponding genes were excluded from the data set (data not shown). In order to further confirm that the identified genes activate the NF-κB, not just the κB element in the reporter plasmid, we introduced dominant-negative-type mutant of IKKβ, which is the most common convergent point of this signaling pathway. The results showed that, in all cases, the fold induction invoked by these genes could be suppressed by the mutant IKKβ (Supplementary Information Figure 2). This observation strongly supported that the genes listed in Tables 2–4 are biological relevant activators of the NF-κB. It was also suggested that the genes identified in this study modulate their activity upstream to the IKKβ.

### *Identification of 'Characterized Genes'*

We first overviewed what kinds of genes were included among the 'Characterized Genes'. The strongest fold induction was observed for the cDNAs of the TNF receptor gene, which has been reported to activate the NF-κB pathway when overexpressed (Pahl, 1999). A total of 34 TNF receptor cDNAs with an average fold induction of 680 were isolated. The most frequently isolated cDNAs were those for the RhoA gene. In all, 58 RhoA cDNAs were isolated from this assay. Recent studies demonstrated that RhoA triggers the translocation of the NF-κB into the nucleus and induces the gene expression from NF-κB-dependent promoters (Montaner *et al*., 1998). For each gene in Table 2, the ID no. of the OMIM database is listed to provide access to the references that describe how the corresponding gene is involved in the NF-κB pathway and what experimental evidence supports its involvement (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id= 'OMIMID'). Genes that are abundantly expressed but explicitly have no relation to the NF-κB pathway, such as cDNAs of the EF-1 alpha gene and the ribosomal protein genes, showed fold induction of less than 4 (data not shown). From these results, we concluded that our system works effectively for the screening of full-length cDNAs of genes involved in the NF-κB pathway.

### *Characterization of the 'Uncharacterized Genes'*

As shown in Table 3, 28 'Uncharacterized Genes' were identified as potential activators of NF-κB. It is

**Table 3** Complete list of the screened 'Uncharacterized Genes'

| *Gene* | *Fold induction* | *Library 1* | *Library2* | *Library 3* | *Total* | *Locus ID* | *OMIM* | *Gene function* |
|---|---|---|---|---|---|---|---|---|
| MLT | 150 | 0 | 1 | 0 | 1 | 10892 | 604860 | Unknown |
| RFP2 | 133 | 1 | 3 | 0 | 4 | 10206 | 605661 | RING finger |
| BST2 | 125 | 0 | 18 | 0 | 18 | 684 | 600534 | Unknown |
| Filamin | 92 | 0 | 0 | 2 | 2 | 2316 | 300017 | Actin binding protein |
| RoRet | 87 | 0 | 2 | 8 | 10 | 10475 | NR | RING finger |
| Vamp-associated protein | 80 | 3 | 1 | 0 | 4 | 9218 | 605703 | Unknown |
| GLVR1 | 56 | 1 | 0 | 0 | 1 | 6574 | 137570 | Channel |
| SNK | 51 | 3 | 1 | 0 | 4 | 10769 | NR | Kinase |
| 14 kDa lectin | 44 | 1 | 0 | 0 | 0 | 3956 | 150570 | Unknown |
| ECM1 | 37 | 1 | 0 | 0 | 1 | 1893 | 602201 | Extracellular protein |
| APOLlll/CG12-1 | 35 | 3 | 1 | 0 | 4 | 80833 | NR | Apolipoprotein |
| RhoB | 34 | 1 | 0 | 0 | 1 | 388 | 165370 | Small G protein |
| EF-1delta | 25 | 0 | 1 | 0 | 1 | 1936 | 130592 | Translation factor |
| ECT2 | 19 | 1 | 0 | 0 | 1 | 1894 | 600586 | Oncogene |
| NEK6 | 16 | 2 | 0 | 0 | 2 | 10783 | 604884 | Kinase |
| FK506BP | 14 | 1 | 0 | 0 | 1 | 2280 | 186945 | Adaptor |
| SECTM1 | 14 | 0 | 0 | 1 | 1 | 6398 | 602602 | Unknown |
| GJA1 | 13 | 1 | 0 | 0 | 1 | 2697 | 121014 | Gap junction protein |
| PIG7 | 13 | 2 | 4 | 0 | 6 | 9516 | 603795 | Nuclear factor |
| TFG/TRK-T3 | 11 | 1 | 0 | 0 | 1 | 10342 | 602498 | Unknown |
| PMR1 | 10 | 1 | 0 | 0 | 1 | 27032 | 604384 | $Ca^{2+}$ pump |
| SEC61A1 | 10 | 0 | 0 | 17 | 17 | 29927 | NR | Protein transport |
| RhoC | 9 | 1 | 3 | 0 | 4 | 389 | 165380 | Small G protein |
| Heme oxygenase | 9 | 1 | 0 | 0 | 1 | 3162 | 141250 | Enzyme |
| PP2C | 7 | 1 | 0 | 0 | 1 | 5494 | 606108 | Phosphatase |
| PP5 | 6 | 0 | 1 | 0 | 1 | 5536 | 600658 | Phosphatase |
| Ecalectin | 5 | 2 | 3 | 0 | 5 | 3965 | 601279 | Eosinophil chemoattractant |
| CTL2 | 5 | 0 | 0 | 1 | 1 | 57153 | 606106 | Transporter |

NR: No record

intriguing that several cancer-related genes are included in this category. For example, the MLT gene was originally identified as a gene disrupted by the chromosomal translocation. t(l1:18)(q21;q21), which recurrently occurs in B-cell lymphomas of mucosa-associated lymphoid tissue (MALT) type (Akagi *et al*., 1999). The translocation fuses the MLT gene and the API2 gene, generating a chimeric protein, API2-MLT. In another MALT lymphoma translocation, t(1;14)(p22;q32), the Bcl-10 gene is mutated. Interestingly, all of these genes, MLT, API2 and Bcl-10, were identified as activators of NF-κB in our study (Tables 2 and 3). Considering that there is no apparent difference in histology, immunophenotype or clinical behavior between MALT with t(11;18)(q21;q21) and t(1;14)(p22;q32), it is possible that these genes exert their oncogenic activities via a similar mechanism. In addition to these genes, we identified several uncharacterized cancer-related genes, such as BST-2, which is a surface antigen preferentially overexpressed on multiple myeloma cells (Ishikawa *et al*., 1995) and the TFG/TRK3 gene, which is associated with papillary thyroid carcinoma (Mencinger *et al*., 1997). This reflects the fact that aberrant activation of NF-κB is involved in the oncogenesis of various cancer types.

In the course of this study, reports appeared describing the cloning of the full-length cDNA of the MALT gene appeared (Uren *et al*., 2000; Liu *et al*., 2001). According to these studies, the NF-κB-activation activity of the MALT gene is possessed not by full-length form of the MALT protein but by the N-terminal-truncated form. The N-terminal part thus should he essential for the canonical control functions of this gene. Since we performed the reporter gene assay using the NF-κB activation activity as an indicator, the truncated form may have been selectively identified. This also might have been the case for several other genes, although the overall fullness of each library was around 70%.

In addition to the possible cancer-related genes, genes involved in various aspects of cellular functions were isolated from our screening. PIG7 is a transcription factor (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id = 603795: further references are therein) and NEK6 and SNK are kinases (http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 10769: http://www.ncbi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 604884) RORET and RFP2 are uncharacterized proteins containing RING - finger domains (http://ww.nebi.nlm.nih.gov/LocusLink/LocRpt.cgi?l = 10475: http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id = 605661), which have been observed in a number of genes involved in the ubiquitin–proteosome-dependent protein degradation pathway (Joazeiro *et al*., 1999: Joazeiro and Weissman, 2000; Xie and Varshavsky, 2001: also see Table 4). Although further analyses will be required to determine how each gene affects the NF-κB pathway, the data presented here should lay groundwork for the genome-wide understanding of the NF-κB signaling network.

### *Characterization of the 'Novel Genes'*

For each of the 'Novel Genes', we completely sequenced the cDNAs and determined the amino-acid sequences

**Table 4** Complete list of the screened 'Novel Genes'

| *Gene* | *Fold induction* | *Library 1* | *Library 2* | *Library 3* | *Total* | *mRNA length* | *Amino acid length* | *Human fully sequenced cDNAs hit* | *Human ESTs hit* | *In-frame stop* | *Genomic locations* | *Predicted subcellular sorting signal* | *Motif (CDD)* | *Genbank ID* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clone 20[a] | 196 | 1 | 0 | 0 | 1 | 1617 | 184 | MGC20791 (id.) | 81 | No | 4q26 (ctg12749) | NS | FHA | AB097011 |
| Clone 502 | 116 | 1 | 0 | 0 | 1 | 2717 | 235 | AF070530 (52%) | 32 | Yes | 5q22.3 (ctg595) | NS | TIR | AB097022 |
| Clone 002N | 81 | 0 | 0 | 1 | 1 | 3168 | 999 | FLJ25919 (id.) | 141 | No | NA | Mitochondrial | ZZ,ANKx5 | AB097000 |
| Clone 373 | 80 | 3 | 3 | 0 | 6 | 702 | 127 | BC007211 (id.) | 59 | Yes | 1p31.2(ctg12483) | NS | ND | AB097018 |
| Clone 413 | 60 | 1 | 0 | 0 | 0 | 2636 | 336 | FLJ10861 (id.) | 171 | No | 13q31.1(ctg15214) | Mitochondrial | ND | AB097019 |
| Clone 330 | 57 | 4 | 3 | 0 | 7 | 1498 | 198 | HSPC186 (75%) | 193 | No | 11p15.4(ctg5305) | Signal peptide | ND | AB097017 |
| Clone 164 | 46 | 5 | 2 | 0 | 7 | 1864 | 221 | BC004317 (id.) | >250 | Yes | 5q31.1 (ctg13420) | NS | Frizzled | AB097010 |
| Clone 209 | 40 | 1 | 0 | 0 | 1 | 2426 | 622 | FLJ10852 (id.) | 58 | Yes | 11p15.1(ctg3235) | NS | NKx2 zf-DHI | AB097014 |
| Clone 266 | 34 | 0 | 1 | 0 | 1 | 2401 | 352 | FLJ12875 (id.) | 210 | No | 1p36.12(ctg18251) | NS | Ring finger | AB097015 |
| Clone 015N | 25 | 0 | 0 | 2 | 2 | 3450 | 910 | MGC17838 (id.) | 57 | Yes | 19p13.2(ctg15174) | NS | Rho-GEF,PH | AB097001 |
| Clone 055N | 20 | 0 | 0 | 1 | 1 | 2815 | 172 | DKFZp564K0822 (id.) | >250 | No | 20p13 (ctg20fin1) | Signal peptide | ND | AB097004 |
| Clone 107 | 19 | 1 | 0 | 0 | 1 | 1551 | 245 | BC017655 (id.) | 220 | No | 17q25.3 (ctg15983) | NS | ND | AB097006 |
| Clone 200 | 19 | 1 | 1 | 0 | 2 | 1085 | 76 | MGC23866 (id.) | 250 | Yes | NA | Signal peptide | ND | AB097012 |
| Clone 120 | 17 | 4 | 3 | 0 | 7 | 1727 | 137 | FLJ31766 (id.) | >250 | Yes | 3p21.31(ctg17803) | Signal peptide | ND | AB097007 |
| Clone 90 | 16 | 3 | 1 | 0 | 4 | 1903 | 455 | BC003187 (id.) | 179 | Yes | 1q21.1(ctg14357) | Signal peptide | ND | AB097024 |
| Clone 031N | 15 | 0 | 0 | 1 |  | 3337 | 540 | FLJ13737 (id.) | 83 | Yes | 7p11.2(ctg30) | NS | ND | AB097003 |
| Clone 205 | 12 | 1 | 0 | 0 | 1 | 2948 | 622 | KIAA0946 (id.) | 31 | No | 12q21.1 (ctg18855) | NS | NKx3 zf-DHI | AB097013 |
| Clone 130 | 10 | 1 | 0 | 0 | 1 | 1524 | 257 | BC007438 (id.) | 232 | No | 17q21.31 (ctg16031) | Mitochondrial | ND | AB097008 |
| Clone 156 | 7 | 1 | 0 | 0 | 1 | 1496 | 107 | ND | 237 | No | 7p22.2 (ctg15064) | NS | ND | AB097009 |
| Clone 321 | 7 | 0 | 1 | 0 | 1 | 1793 | 245 | BC0012389 (id.) | >250 | No | Xq22.1 (ctg15556) | Mitochondrial | ND | AB097016 |
| Clone 470 | 6 | 1 | 1 | 0 | 2 | 2976 | 138 | DKFZp761G2423 (id.) | 186 | No | 12p12.1 (ctg14210) | NS | ND | AB097020 |
| Clone 48 | 6 | 1 | 0 | 0 | 1 | 1898 | 392 | BC024288 (id.) | >250 | No | 6p21.1 (ctg15907) | Signal peptide | DUF6 | AB097021 |
| Clone 023N | 5 | 0 | 0 | 1 | 1 | 3579 | 951 | FLJ20475 (id) | 100 | No | 1p36.23 (ctg1565) | Mitochondrial | C2 | AB097002 |
| Clone 102 | 5 | 2 | 1 | 0 | 3 | 1448 | 322 | MGC915 (id.) | 154 | Yes | 5q12.3 (ctg12639) | NS | zf-CXXC | AB097005 |
| Clone 502H | 225 | – | – | – | 1* | 2164 | 712 | ND | 48 | No | 19p13.3 (ctg18730) | NS | TIR | AB097023 |

NA: not assigned; NS: not significant ND: not detected

Each of the cDNA sequences were searched against the databases of fully sequenced cDNAs and the dbEST (the ninth and the 10th columns). As for fully sequenced cDNAs, the IDs of hit cDNAs and their identifies are shown, id.: identical. As for ESTs, the numbers of hit ESTs are shown. Whether a cDNA contained any terminator codon upstream to the translation initiator ATG is shown in the 11th column. The genomic positions, the presences of the predicted subcellular localization signals and the protein motifs are shown in the 12th to the 14th columns. As for the rest of the columns, refer to the legend of Table 2

[a]During the preparation of the manuscript, this gene was reported to activate NF-κB

which they encoded. Using the deduced amino-acid sequences of 'Novel Genes', we searched the CDD protein motif database using PSI-BLAST at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). In Clone 502, the TIR domain was observed (Figure 1). The TIR domain was originally found as a conserved cytoplasmic domain of Toll-like receptors (TLRs), which include the LPS receptors, TLR2 and TLR4, and the IL-1R superfamily. Several reports have shown that the TIR domain mediates homophilic protein–protein interactions between the signaling molecules of this pathway. So far, three subtypes have been reported in the TIR-containing protein superfamily, namely, the IL-1R type, TLR type and MyD88 type (Kopp and Medzhitov, 1999). The former two consists of membrane proteins and the last one consists of cytoplasmic proteins. Since Clone 502 contains no transmembrane domain, this protein should belong to the MyD88-type subfamily. Until very recently, MyD88, which is a critical mediator for signaling from IL-1RI, TLR2 and TLR4, was the only known human gene in this subfamily. While the present study was in progress, a new member, Mal, was identified and shown to play an indispensable role in transmitting the signal from TLR4 to cellular responses, including NF-κB activation (Fitzgerald *et al*., 2001). Clone 502 should be a new member of this family that is also involved in the NF-κB pathway, which plays a pivotal role in the innate immune response.

We further searched Genbank for other new members of the TIR domain-containing gene family. A homology search using the cDNA sequence of Clone 502 hit another TIR domain-containing protein. We carried out the molecular cloning and the complete sequencing of the full-length cDNA of this gene and designated it as 'Clone 502 Homolog' (Figure 1a). We performed the luciferase reporter gene assay for Clone 502 Homolog as for other clones and found that this clone showed more than 200-fold induction of the luciferase activity (Figure 1b). Moreover, this fold induction was suppressed by introducing dominant-negative-type mutant IKKβ (Figure 1c). Therefore, we concluded that this new gene also belongs to the TIR domain-containing gene family involved in the NF-κB pathway. It is also noteworthy that the simple combination of the homology search and the reporter gene assay expanded our data set.

Several homologous groups were found among 'Novel Genes'. Clone 413 was 64% homologous to Clone 164. Besides, a mouse gene, the Nedd4 WW-binding protein 5 gene, showed 97% homologous to Clone 164. This close mouse homolog was reported to interact with a ubiquitin–protein ligase, Nedd4 (Jolliffe *et al*., 2000). In all of these genes, the PY motif, which is essential for the interaction with Nedd4, was conserved. The fact that these two related genes have been identified independently as activators of the NF-κB pathway suggests that the Nedd4-mediated protein-degradation pathway should be also involved in the regulation of the NF-κB pathway.

Clones 205 and 209 showed 69% overall homology and their exon–intron structures are highly conserved. The lengths of the exons were completely identical through third to eighth and 10th exons (the genomic sequence corresponding to the region downstream of the 10th exon of Clone 205 has not yet been determined), which strongly suggests the duplication of an ancestral gene occurred during evolution. Consistent with their similar genomic organization, both of them contained a Zn-finger domain at similar positions. However, while Clone 205 contains four ankyrin repeats, which are

**a**

```
TIR: Clone502      ----------IL-HAEDDTDEALRVQNLLQDDFGIKPGIIFAEMPC--GKQHLQNLIDAVNGSAWTILLLTENF--LRDTWCN
TIR: Clone502H     ----------YNFVIIHARADEHIALRVREKLEA-LGVPDGATFCEDFQVPGKGELSGLQDAIDHSAFIILLLTSN----------
TIR: IL1L1         ----------E---H-FVHQLPDVLENKCGYTL-C---IY--GRDML--PGEDVVTAVETTIRKSRRHIFIL-TPQITHNKEFAY
TIR: IL1R1         AYILYPKTVG-EGSTSDCD---I-FVFKVLPEVLEKQCGYKL-F---IY--GRDDY--VGEDIVEVINENVKKSRRHIFIL-VRETSGFSWLGG
TIR: IL1RAPL1      AYLSYTKVDP-DQWNQETGEE-ERFALEILPDMLEKHYGYKL-F---IP--DRDLI--PTGPYIEDVARCVDQSKRLIIVM-TPNYVVRRGWSI
TIR: IL1RAPL2      AYLSYTKVDQ-DTLDCDNPEE-EQFALEVLPDVLEKHYGYKL-F---IP--ERDLI--PSGPYMEDLTRYVEQSRRLIIVL-TPDYILRRGWSI
TIR: MyD88         ----------AFICYCPSDIQFVQEMIRQLEQTNYRLKLCVSD--RDVLPGTCVWSIASELIEK-RCRRMVVVVSDDY--LQSKECD
TIR: TIRAP         ----------VCHSEEDLVAAQDLVSYLEGSTASLR-LFLQL-RDATPGGAIVSELCQALSSSHCRVLLITP--GF--LQDEWCK
TIR: TLR1          ----------AFISYSGHDSFWVKNE-LLPNLEK-EGM-QI-----CLHERNFV--PGKSIVENIITCIEKSY--KSIFVLSPNFVQSEWCH
TIR: TLR2          ----------AFVSYSERDAYWVENL-MVQELEN-FNP-PFKL---CLHKRDFI--PGKWIIDNIIDSIEKSH--KTVFVLSENFVKSEWCK
TIR: TLR3          ----------AYIIHAYKDKDWVENH-FSSMEKEDQS-LK--FCLE--ERDFE--AGVFELEAIVNSIKRSR--KIIFVITHHLLKDPLCK
TIR: TLR4          ----------AFVIYSSQDEDWVRNE-LVKNLEE-GVP----PFQLCLHYRDFI--PGVAIAANIIHEG-FHKSRKVIVVVSQHFIQSRWCI
TIR: TLR5          ----------AYLCFSSKDFTWVQNA-LLKHLDT-QYS-DQNRFNLCFEERDFV--PGENRIANIQD-AIWN-ARKIVCLVSRHFLRDGWCL
TIR: TLR6          ----------AFISYSEHDSAWNKSE-LVPYLEK-EDI-QI-----CLHERNFV--PGKSIVENIINCIEKSY--KSIFVLSPNFVQSEWCH
TIR: TLR7          ---------AFIVYDTKDPAVTEWVLAE-LVAKLED-PRE-KHFNL--CLEERDWL--PGQPVLENLSQSIQLSK--KTVFVMTDKYAKTENFK
TIR: TLR8          ---------AYISYDTKDASVTDWVINE-LPYHLEE-SRD-KNVLL--CLEERDWD--PGLAIIDNLMQSINQSK--KTVFVLTKKYAKSWNFK
TIR: TLR10         ----------AFISYSEHDSLWVKNE-LIPNLEK-EDG-SILI---CLYESYFD--PGKSISENIVSFIEKSY--KSIFVLSPNFVQNEWCH
```

**Figure 1** (**a**) TIR domains observed in Clones 502 and 502H are aligned with the TIR domains of other genes. The amino acids conserved throughout the TIR domains are shown by the black boxes. The amino acids conserved between Clones 502 and 502H are shown by the open boxes. TLR: Toll-like receptor. (**b**) Fold induction of the luciferase gene activity by Clone 502 (left panel) and Clone 502H (right panel). The empty vector or vector carrying the full-length cDNA of Clones 502, Clone 502H or IKKa, which is a known activator of the NF-κB pathway, was introduced into HEK293 cells and subjected to the reporter gene assay. The doses of the introduced plasmid are shown in the boxes. (**c**) The fold inductions invoked by Clone 502 (left panel) and Clone 502H (right panel) were suppressed by a dominant-negative-type mutant IKKβ. Indicated amount of the mutant IKKβ were introduced together with the 70 ng of each of the plasmids. The fold suppression was calculated compared to the luciferase activity without the cotransfection of the mutant IKKβ designated as 1. NFκB-p65 and MyD88 are negative and positive controls, which are irrespective of or are subjected to the modulation of the mutant IKKβ, respectively

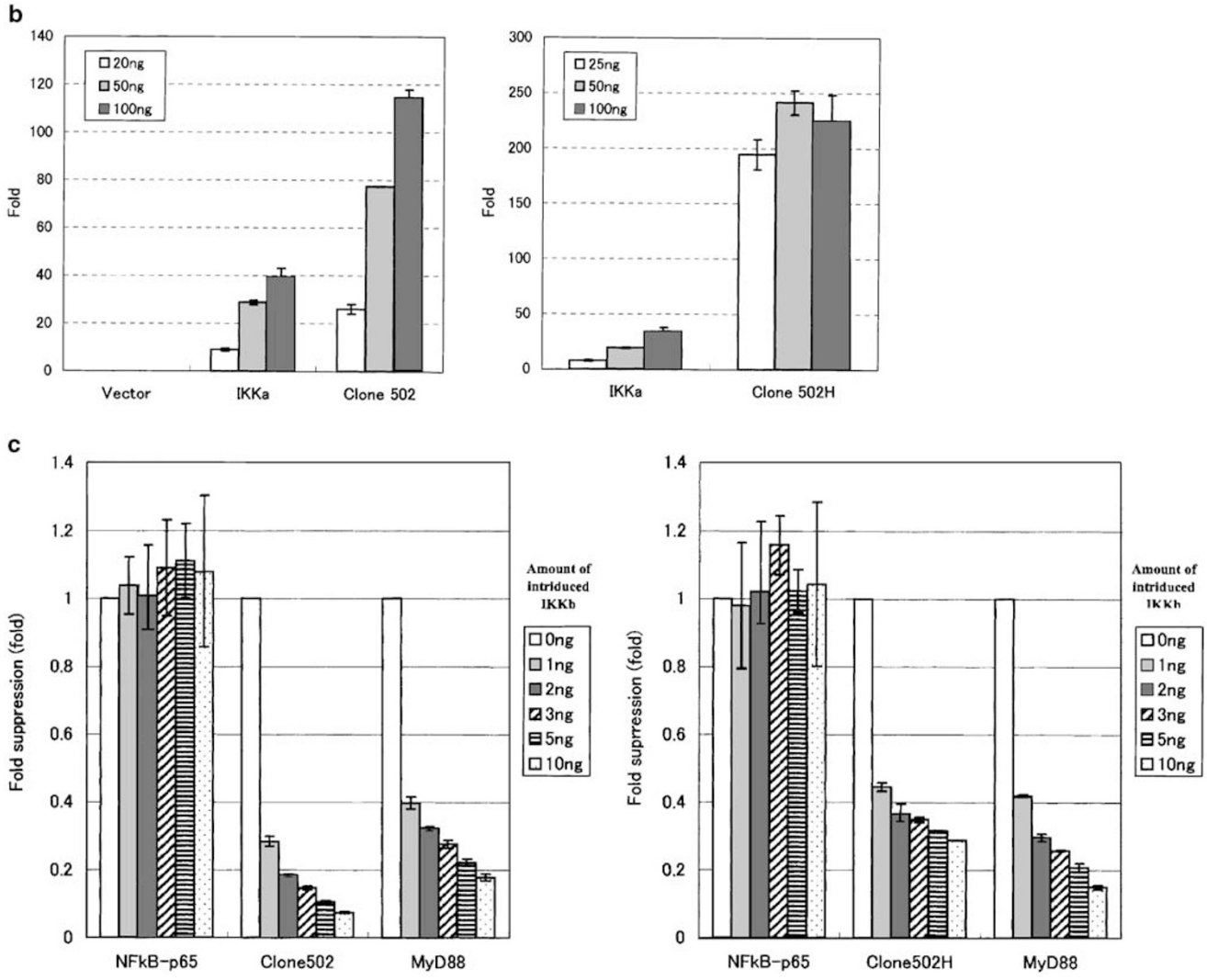


**Figure 1** (Continued)

thought to play an important role in protein–protein interactions in signaling molecules (Sedgwick and Smerdon, 1999), Clone 209 has only two ankyrin repeats. These results imply that Clones 205 and 209 should correspond to genes of a very closely related gene family, but have different functions.

A BLAST search using the cDNA sequences of 'Novel Genes' also revealed that some of them have homologs in other organisms such as mouse, *C. elegans*, fruit fly and yeast (Table 5). In these organisms, several kinds of genomewide analyses for characterizing gene functions, including systematic knockout studies and protein–protein interaction mapping, are underway (Barstead, 2001; Gavin *et al*., 2002). It will be extremely informative to see what kinds of phenotype are reported for the disruptants of the homologs of these genes or with what proteins they interact in order to understand in what context the genes identified here are involved in the NF-κB pathway.

*Detailed determination of the transcriptional start sites and characterization of the potential NF-κB binding sites in the promoters*

Since it is not always possible to obtain useful information on gene functions solely by analyses of the cDNAs, we attempted to analyse the transcriptional regulation of the genes corresponding to the cDNAs identified here as well. Fortunately, since the cDNAs which we had isolated were full-length cDNAs, their 55-ends corresponded to the transcriptional start sites (TSSs) of the mRNAs. Thus, using the almost completed public human genome sequence data, we could map the 5′-ends of the cDNAs we identified onto the human genome and identify the adjacent promoter regions. Using the retrieved promoter sequences, the consensus sequences of transcription factor binding sites were searched using TRANSFAC (Ver 5.1: Wingender *et al*., 2001). Considering that the promoters of most

**Table 5** Homologs of the 'Novel Genes'

| *Gene* | *M. musculus* | *D. melanogaster* | *C. elegans* | *S. cerevisiae* | *S. pombe* | *A. thaliana* |
|---|---|---|---|---|---|---|
| Clone 205 | 40% (447aa) | 48% (576aa) | 26% (458aa) | 30% (506aa) | 28% (541aa) | 29% (161aa) |
| Clone 209 | 87% (446aa) | 36% (619aa) | 26% (565aa) | 28% (518aa) | 26% (546aa) | 25% (235aa) |
| Clone 266 | 89% (352aa) | 29% (335aa) | ND | ND | ND | 26% (346aa) |
| Clone 330 | 85% (164aa) | ND | 26% (193aa) | ND | ND | ND |
| Clone 90 | 96% (455aa) | 57% (434aa) | 56% (450aa) | ND | ND | 43% (447aa) |
| Clone 48 | 76% (432aa) | 42% (363aa) | 34% (370aa) | ND | 24% (286aa) | 34% (276aa) |
| Clone 002N | 78% (961aa) | 37% (1047aa) | 29% (353aa) | ND | ND | 25% (346aa) |
| Clone 15N | 64% (490aa) | 33% (368aa) | 30% (370aa) | ND | ND | ND |
| Clone 023N | 84% (596aa) | 29% (797aa) | 29% (483aa) | ND | ND | ND |

ND: Not detected
For each of the homologs identified from the indicated organisms, the sequence identity and the aligned amino-acid length is shown. For the search of the homologs, BLASTP was performed against Genbank with the cutoff value of $1 \times 10^{-10}$. The hits with the top scores within each of the organisms are shown

human genes have not been identified because of the technical difficulty of the conventional methods (273 human genes are recorded in the Eukaryotic Promoter Database; Perier *et al*., 2001), another significant advantage in our approach is that we can identify the promoters of the isolated genes without additional laborious experiments.

Figure 2a illustrates the results of such the analyses on the BST-2 gene and the TRAIL gene. Determination of the TSS and identification of the promoter have been previously reported for both of these genes (Ohtomo *et al*., 1999; Baetu *et al*., 2001). The TSSs determined in this study were located just proximal to the previously determined ones and promoter sequences identical with those previously identified were successfully retrieved from the genomic sequences using our procedure, although the exact positions of the TSSs we identified were highly divergent from each other in both genes, suggesting the slipperiness of the transcriptional initiation events in these genes (Suzuki *et al*., 2001). In Figure 2a, each TSS identified in this study is represented by an arrow. The consensus sequence of the STAT binding sites in the promoter of the BST-2 gene and the NF-κB binding sites in the promoter of the TRAIL gene are also highlighted in the corresponding panels. These promoter elements may be responsible for the fact that all of the cDNAs of these genes were isolated only from Library 2 (all 18 cDNAs of the BTS gene and all 10 cDNAs of the TRAIL gene; see Tables 2 and 3), which was constructed from HNLF cells stimulated by TNFα and IL-4 (Kurata *et al*., 1999; Baetu *et al*., 2001).

Intriguingly, the consensus sequence of the NF-κB binding site was found in the promoter regions of five 'Characterized Genes' and three 'Novel Genes' (Figure 2b, c). For these genes, it is possible that a positive feedback loop of the transcriptional regulation exists. Indeed, it has been recently reported that the CD40 gene, which activates the NF-κB signaling pathway, is subject to positive expression regulation by NF-κB (Hinz *et al*., 2001). The promoters of the genes that contained the NF-κB binding sites may serve as useful clues for the functional characterization of these genes. It would also be interesting to compare the promoters of these genes with their expression profiles as studies employing microarray analysis begin to identify candidate target genes that may be activated by NF-κB (Dong *et al*., 2001). Although further experimental validation, such as electromobility shift assay, on each of these genes should be indispensable before concluding that they are really downstream targets of NF-κB, it is significant that we were able to produce possible candidates at the same time with identifying genes activating the pathway itself. Detailed analyses of both the transcriptional regulation of each gene and of its protein functions should be performed complementarily with each other for functional characterization of the feedback/feedforward regulatory network of this signaling pathway.

### *Crosstalk between the NF-κB pathway and the MAPK pathway*

We also carried out a similar analysis for the identification of genes that activate the MAPK pathway. As is the case of the NF-κB, the cDNAs that induced the luciferase activity more than fourfold compared to the mock insert plasmid were selected. In this case, full-length cDNAs of p38 and JNK were introduced together with the plasmid. This was done because our preliminary experiments showed that activators of the p38/JNK pathway, which are subpathways of the MAPK pathway, could not be identified without supplying exogenous P38 or JNK, while genes activating Erk pathway, which is the other known subpathway of the MAPK pathway, could be identified without introducing Erk-1/2. In total, we examined 145 000 cDNA clones and identified 57 genes that activate the MAPK pathway (Table 6; for further details, see Materials and methods). The identified cDNAs were categorized and analysed in the same way as in the case of the NF-κB pathway (Supplementary Information Table 1). We observed that many of the genes that had been previously reported to activate the MAPK pathway, such as Raf (MAPKKK), Tak (MAPKKK), Cot (MAPKKK) and MKK6 (MAPKK) were included among the 'Characterized Genes'. Again, genes that explicitly have no relation to the pathway were not

identified. These observations showed that our system also works for the analysis of the MAPK pathway. Although it is still possible that some of these genes are identified irrespective of the activation of the MAPK pathway, it is likely that its frequency is low and the genes listed in Supplementary Information Table 1 are mostly biologically relevant activators of the MAPK pathway.

Among the identified genes, 27 genes were overlapping, activated both the NF-κB pathway and the MAP pathway ('Overlapping Genes' in Table 6). This result suggests that both pathways utilize these overlapping signaling molecules. Signaling through a common mediator could allow for coordinated responses downstream of certain receptors and may further provide cells with the opportunity to regulate the relative amounts of the signals that the two pathways receive at a single point. Consistent with this idea, emerging evidence indicates that many signaling molecules, including TRAFs, MyD88 and CIKS, are shared

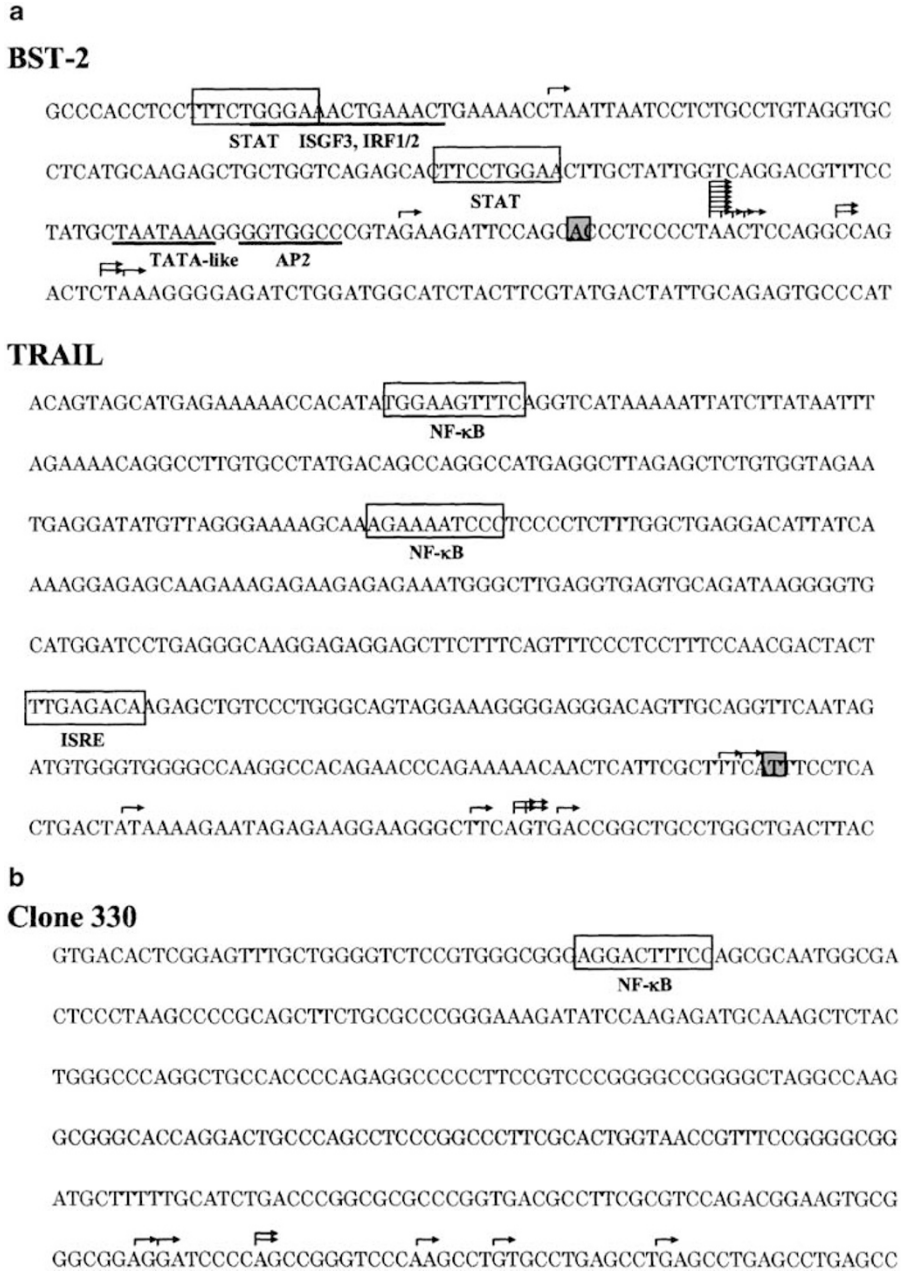


**Figure 2** Putative promoter structures of the screened Genes. Transcriptional start sites identified in this study are represented by arrows. Previously reported TSSs are shown by shaded boxes. Transcription factor binding sites predicted in the corresponding promoters are shown by open boxes. (**a**) The promoter region of the BST-2 gene and the TRAIL gene. (**b**) The promoter region of Clone 330. (**c**) Predicted NF-κB sites in the promoters of the screened genes. When slippery transcriptional start sites were observed, the mostupstream site was selected as the representative

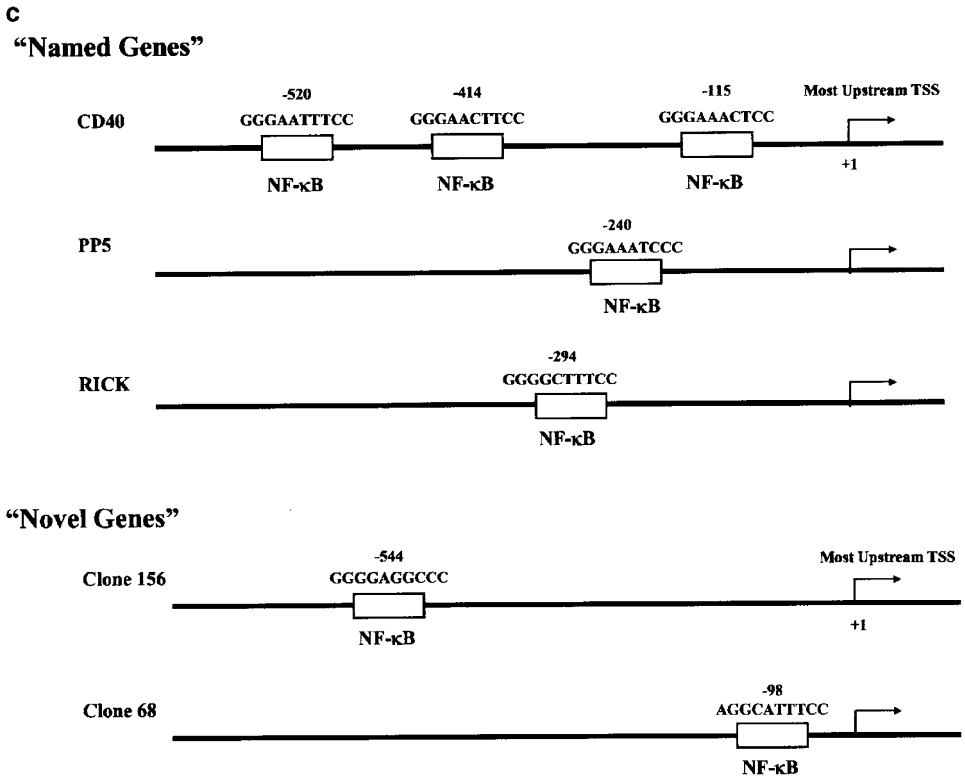


**Figure 2** (Continued)

**Table 6** Screened genes that activate the MAPK pathway

| | *# MAPK activating genes* | *# Overlapping genes* |
|---|---|---|
| 'Characterized Genes' | 28 | 13 |
| 'Uncharacterized Genes' | 11 | 2 |
| 'Novel Genes' | 18 | 12 |
| Total | 57 | 27 |

The numbers of screened genes that activated the MAPK pathway or both of the MAPK pathway and the NF-κB pathway are shown

by these two pathways. A recent study also demonstrated that MEKK3, which is one of the most important modulators of the MAPK pathway, is essential for TNF-induced NF-κB activation as well (Yang *et al*., 2001). Actually. TRAF 5, MyD88, CIKS and MEKK3 were identified as 'Overlapping Genes' in the present study (Figure 3). The idea that the NF-κB pathway and the MAPK pathway has multiple points of signaling convergence should be examined further by detailed analysis of the genes noted in Table 6.

## Conclusion

Here we reported large-scale identification and characterization of human genes that activate the NF-κB signaling pathway. We also demonstrated that our approach could be applied for the analysis of the MAPK signaling pathways too. A significant number of full-length cDNAs of genes were identified, some of which contained important motifs and others may play crosstalking roles between these two important signaling pathways. The information described here should lay the firm groundwork for the future analyses on each of the genes as to how they are involved in these pathways.

Our approach has two major advantages. First, we could facilitate the reporter gene assay by using cDNA clones isolated from full-length cDNA libraries. Considering that the cDNA libraries used for this study contained about 70% full-length cDNAs, the efficiency of the assay should have been increased by more than 10-fold compared to the efficiency with conventional cDNA libraries, since the frequency of full-length cDNAs in a conventional library is usually estimated to be several percent (for details see Suzuki *et al*., 2001). Utilization of full-length cDNA libraries should be especially critical when transfection into mammalian cells and reporter gene assays are used, since this strictly limits the screening size. Secondly, since we started with full-length cDNA libraries, no further efforts, including several laborious rounds of the 5′RACE procedure, were required to isolate a full-length cDNA. Moreover, because the 5′-ends of the screened cDNAs should correspond to the exact mRNA start sites, the adjacent promoter sequences could be identified as well, which

should give us important clues about the transcriptional regulation of the genes.

A drawback in our approach is that we owe the screening method to the overexpression of the exogeneously introduced cDNAs. This implicates that some of the identified cDNAs might be artifacts, since the overexpression can cause an imbalance among components of the signaling pathway, so that the pathway gets artificially activated by a gene that is not a normal component of the pathway. Although dominant-negative-type mutant IKK$\beta$ suppressed the activation of the pathway invoked by those genes (Figure 1c and Supplementary Information Figure 2), biochemical analyses on each of the identified genes should he indispensable before concluding that it is really a physiological activator of the pathway. To the contrary, it is also possible that we have missed the cDNAs that are really involved in the pathways but could not be activated by overexpression. However, it should be noteworthy that many of the previously characterized activators of the pathways were included in our screened genes and the genes that have explicitly no correlation to the pathways were seldom identified. These observations should validate the usability and fidelity of our method at least to some extent.

Another advantage of our approach is that it is applicable for essentially all of the signaling pathways. Moreover, on the other hand, we are now rapidly generating almost complete collection of full-length cDNAs of human genes, of which there are estimated to be 30 000–40 000 kinds (Strausberg *et al*., 1999; submitted by Ota *et al*). Using this full-length cDNA resource for the future screening, further improvement of the efficiency and coverage of the analysis can be expected. The genes involved in various signaling pathways that are known to eventually activate a particular transcription factor would be identified one by one. Such a comprehensive analyses of signaling pathways and their crosstalks should create a bridge from the rapidly accumulating cDNA sequence information to the genomewide understanding of human gene networks.

## Materials and methods

### *Construction of full-length cDNA libraries and arrayed cDNA pool*

NHLF cells were cultured in MEM containing 10% FCS. At 24 h before the cells were harvested, they were stimulated with 0.5 ng/ml IL-4 and 5 ng/ml TNF$\alpha$. Total RNA was isolated using ISOGEN (Wako) and RNeasy (QIAGEN) according to the manufacturer's instructions. PolyA + RNAs were purified using oligo-dT cellulose (MRC) according to standard methods. Using 50 $\mu$g of polyA + RNA, the cap structure of the mRNA was replaced with a 5′-oligo-ribonucleotide by the oligo-capping method, which consists of three enzymatic reaction steps. First, bacterial alkaline phosphatase (BAP) hydrolyses the phosphate of the 5′-ends of truncated mRNAs whose cap structures have been broken down. Then, tobacco acid pyrophosphatase (TAP) removes the cap structure, leaving a phosphate at the 5′-end. Finally, T4 RNA ligase, which requires a phosphate at the 5′-end as its substrate, selectively ligates a 5′-oligo to the 5′-end that originally had a cap structure. Using oligo-capped mRNA, first-strand cDNA was synthesized with oligo dT adapter primer. After alkaline degradation of the template, first-strand cDNA was amplified by 12 rounds of PCR reactions and cloned into a plasmid vector, pME18S-FL3 (Genbank Accession #AB009864), in an orientation-defined manner. For further details of the procedure, see reference Suzuki *et al*. (1997); for more detailed procedure, see http://cdna.ims.u-tokyo.ac.jp/matsuda_etal/Protocol_TextFigs.pdf.

We randomly isolated 150 000 cDNA clones from the full-length cDNA libraries (60 000, 60 000 and 30 000 for Libraries 1, 2 and 3, respectively) and used them to construct an arrayed cDNA pool. Plasmid DNAs were purified using QIAwell 96 Ultra Plasmid Kits (QIAGEN) according to the manufacturer's instructions.

### *Transient transfection and luciferase reporter gene assay*

In each well of a 96-well microtiter plate, $1 \times 10^4$ HEK293 cells were cultured and 50 ng of the plasmid DNAs were transfected together with 25 ng of the reporter plasmid using 0.3 $\mu$l of Fugene 6 (Roche). At 24 h after the transfection, cells were harvested and assayed using the Dual Luc System (Promega) according to the manufacturer's instructions. The cDNAs were transfected into the cells as quadruples (four cDNAs per well). For each group of cDNAs that showed more than fourfold induction of the luciferase activity, a second round of this reporter gene assay was performed in the same manner. Each reporter gene assay was performed in triplicate. The dominant-negative-type mutant IKK$\beta$ was produced as an kinase-deficient form containing an amino-acid substitution of 44 (K → M) as described previously (Mercuric *et al*., 1997) and indicated amount was introduced.

For analysis of the NF-κB pathway, pNF-κB-Luc (Stratagene) was used as the reporter plasmid. For analysis of the MAPK pathway, the cDNAs were transfected together with full-length cDNAs of p38, JNK and pFA2-Elk1 (Stratagene), which contains the GAL4-Elk1-fusion gene. A plasmid containing the GAL4 binding sequence upstream of the luciferase gene, pFR-Luc (Stratagene) was used as the reporter plasmid. Briefly, cDNAs involved in the MAPK pathway leading to the activation of p38, JNK or Erk1/2 were expected to activate the Elk-1 transcription activation activity and thereby increase the fold-induction of the luciferase activity. The validation of the screening was performed as is done in the case of the NF-κB pathway.

### *Computational analyses*

The BLAST searches were performed using the NCBI BLAST (http://www.ncbi. nlm.nih.gov/BLAST/; Altschul *et al*., 1997). For the protein motif database, CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) was used. Genomic alignments of the cDNAs were determined using Sim4 (Florea *et al*., 1998) and Golden Path (http://www.genome.ucsc.edu/; as of 22 December 2001). TRANSFAC was from hltp://www.biobase.de/ (Rel. 5.1). For the search of transcription factor binding sites, we employed cutoff values of 0.90 and 0.95 for the matrix similarity and the core similarity, respectively. The promoter sequences from nucleotides −1000 to 0 compared to the TSS were searched. Subcellular sorting signals of the proteins were predicted using iPSORT (http://www.hypothesiscreator.net/iPSORT/; Bannai *et al*., 2001).

**Acknowledgements**
We thank K Ishizawa, M Yagi, T Sugahara, Nao Suzuki, S Suzuki, S Yoneta, Nori Suzuki and R Ooishi for their excellent screening and sequencing work. We are thankful to M Shirota and K Abe for their technical support and JM Sugano for helpful discussions. We are also grateful to E Nakajima for critical reading of the manuscript.

## References

Akagi T, Motegi M, Tamura A, Suzuki R, Hosokawa Y, Suzuki H, Ota H, Nakamura S, Morishima Y, Taniwaki M and Seto M. (1999). *Oncogene*, **18,** 5785–5794.
Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ. (1997). *Nucleic Acids Res*., **25,** 3389–3402.
Baetu TM, Kwon H, Sharma S, Grandvaux N and Hiscott J. (2001). *J. Immunol.*, **167,** 3164–3173.
Baeuerle PA and Baltimore D. (1996). *Cell*, **87,** 13–20.
Baldwin Jr AS. (1996). *Annu. Rev. Immunol*., **14,** 649–683.
Baldwin Jr AS. (2001). *J. Clin. Invest.*, **107,** 3–6.
Bannai H, Tamada Y, Maruyama O, Nakai K and Miyano S. (2001). Proceedings of the 14th International FLAIRS Conference. AAAI Press, pp. 233–238.
Barkett M and Gilmore T. (1999). *Oncogene*, **18,** 6910–6924.
Barstead R. (2001). *Curr. Opin. Chem. Biol*., **5,** 63–66.
Beyaert R, Cuenda A, Vanden Berghe W, Plaisance S, Lee JC, Haegeman G, Cohen P and Fiers W. (1996). *EMBO J.* **15,** 1914–1923.
Dong G, Loukinova E, Chen Z, Gangi L, Chanturita TI, Liu ET and Van Waes C. (2001). *Cancer Res*., **61,** 4797–4808.
Fitzgerald KA, Palsson-McDermott EM, Bowie AG, Jefferies CA, Mansell AS, Brady G, Brint E, Dunne A, Gray P, Harte MT, McMurray D, Smith DE, Sims JE, Bird TA and O'Neill LA. (2001). *Nature*, **413,** 78–83.
Florea L, Hartzell G, Zhang Z, Rubin GM and Miller W. (1998). *Genome Res.*, **8,** 967–974.
Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Radia M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G and Superti-Furga G. (2002). *Nature*, **415,** 141–147.
Ghosh S, May MJ and Kopp EB. (1998). *Annu. Rev. Immunol.*, **16,** 225–260.
Hinz M, Loser P, Mathas S, Krappmann D, Dorken B and Scheidereit C. (2001). *Blood*, **97,** 2798–2807.
Ishikawa J, Kaisho T, Tomizawa H, Lee BO, Kobune Y, Inazawa J, Oritani K, Itoh M, Ochi T and Ishihara K. (1995). *Genomics*, **26,** 527–534.
Jacobs MD and Harrison SC. (1998). *Cell*, **95,** 749–758.
Joazeiro CA, Wing SS, Huang H, Leverson JD, Hunter T and Liu YC. (1999). *Science*, **286,** 309–312.
Joazeiro CA and Weissman AM. (2000). *Cell*, **102,** 549–552.
Jolliffe CN, Harvey KF, Haines BP, Parasivam G and Kumar S. (2000). *Biochem. J*., **351,** 557–565.
Karin M. (1999). *Oncogene*, **18,** 6867–6874.
Kopp EB and Medzhitov R. (1999). *Curr. Opin. Immunol*., **11,** 13–18.
Kurata H, Lee HJ, O'Garra A and Arai N. (1999). *Immunity*, **11,** 677–688.
Liu H, Ye H, Dogan A, Ranaldi R, Hamoudi RA, Bearzi I, Isaacson PG and Du MQ. (2001). *Blood*, **98,** 1182–1187.
Maruyama K and Sugano S. (1994). *Gene* **138,** 171–174.
Mencinger M, Panagopoulos I, Andreasson P, Lassen C, Mitelman F and Aman P. (1997). *Genomics*, **41,** 327–331.
Mercuric F, Zhu H, Murray BW, Shevchenko A, Bennett BL, Li J, Young DB, Barhosa M, Mann M, Manning A and Rao A. (1997). *Science*, **278,** 860–866.
Montaner S, Perona R, Saniger L and Lacal JC. (1998). *J. Biol. Chem.*, **273,** 12779–12785.
Ohtomo T, Sugamata Y, Ozaki Y, Ono K, Yoshimura Y, Kawai S, Koishihara Y, Ozaki S, Kosaka M, Hirano T and Tsuchiya, M. (1999). *Biochem. Biophys. Res. Commun.*, **258,** 583–591.
Pahl HL. (1999). *Oncogene*, **18,** 6853–6866.
Perier RC, Praz V, Junier T, Bonnard C and Bucher P. (2000). *Nucleic Acids Res.*, **28,** 302–303.
Reed JC. (2001). *Trends. Mol. Med*., **7,** 314–319.
Sambrook J, Fritsh EF and Maniatis T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
Schmitz ML, Bacher S and Kracht M. (2001). *Trends. Biochem. Sci*., **26,** 186–190.
Schulze-Osthoff K, Ferrari D, Riehemann K and Wesselborg S. (1997). *Immunobiology*, **198,** 35–49.
Sedgwick SG and Smerdon SJ. (1999). *Trends. Biochem. Sci*., **24,** 311–316.
Stein B, Baldwin AS Jr, Ballard DW, Greene WC, Angel P and Herrlich P. (1993). *EMBO J*., **12,** 3879–3891.
Strausberg RL, Feingold EA, Klausner RD and Collins FS. (1999). *Science*, **286,** 455–457.
Suzuki Y, Yoshitomo K, Maruyama K, Suyama A and Sugano S. (1997). *Gene*, **200,** 149–156.
Suzuki Y, Taira H, Tsunoda T, Mizushima JS, Sese J, Hata H, Ota T, Isogai T, Tanaka T, Morishita S, Okubo O, Sakaki Y, Nakamura Y, Suyama A and Sugano S. (2001). *EMBO Rep*., **2,** 388–393.
Uren AG, O'Rourke K, Aravind LA, Pisabarro MT, Seshagiri S, Koonin EV and Dixit VM. (2000). *Mol. Cell*, **6,** 961–967.
Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, Pruss M, Schacherer F, Thiele S and Urbach S. (2001). *Nucleic Acids Res.*, **29,** 281–283.
Verma IM and Stevenson J. (1997). *Proc. Natl. Acad. Sci. USA*, **94,** 11758–11760.
Xie Y and Varshavsky A. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 3056–3061.
Yang J, Lin Y, Guo Z, Cheng J, Huang J, Deng L, Liao W, Chen Z, Liu Z and Su B. (2001). *Nat Immunol*., **2,** 620–624.
Yin MJ, Yamamoto Y and Gaynor RB. (1998). *Nature*, **396,** 77–80.