

# The genomic structure and chromosomal localization of the mouse STAT3 gene

Wei Shi, Masahiro Inoue, Masashi Minami, Kiyoshi Takeda, Makoto Matsumoto, Yoichi Matsuda<sup>2,3</sup>, Tadamitsu Kishimoto<sup>1</sup> and Shizuo Akira<sup>4</sup>

Division of Immunology, Institute for Molecular and Cellular Biology, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565, Japan

<sup>1</sup>Department of Internal Medicine III, Osaka University Medical School, 2-2 Yamadaoka, Suita, Osaka 565, Japan

<sup>2</sup>Genome Research Group, National Institute of Radiological Sciences, 4-9-1 Anagawa, Inage-ku, Chiba, Chiba 263, Japan

<sup>3</sup>Present address: Laboratory of Animal Genetics, School of Agricultural Sciences, Nagoya University, Chikusa-ku, Nagoya, Aichi 464-01, Japan

<sup>4</sup>Present address: Department of Biochemistry, Hyogo College of Medicine, 1-1 Mukogawacho, Nishinomiya, Hyogo 663, Japan

**Keywords:** cytokine, genomic DNA, IL-6, signal transduction, promoter region

## Abstract

**A variety of cytokines induce the tyrosine phosphorylation of signal transducers and activators of transcription (STATs). Activation of the same STAT proteins by distinct cytokines and activation of different STAT proteins by each cytokine are thought to contribute to redundancy and pleiotropy of cytokine actions respectively. STAT3 is rapidly tyrosine phosphorylated in response to IL-6, ciliary neurotrophic factor, oncostatin M, leukemia inhibitory factor, IL-11, granulocyte colony stimulation factor and epidermal growth factor. In this report we have isolated and characterized the mouse genomic structure of STAT3. The mouse STAT3 gene consisted of 24 exons which spanned >37 kb. The structure of the mouse STAT3 gene was almost identical to that of the human STAT2 gene, including the number and size of exons, indicating that the exon–intron organization had already been accomplished before these two genes duplicated, and then these genes evolved to respond to different ligands. By molecular linkage analysis with interspecific backcross mice the STAT3 gene mapped at 1.4 cM proximal to *D11Mit59* on mouse chromosome 11. The promoter region contained potential regulatory elements such as GATA, NF-IL-6, PEBP2, Sp-1, AP-2 binding sites, cAMP response element, CAAT box and E-box. Transient expression of constructs harboring the 5' flanking region of the STAT3 gene fused to the luciferase gene showed that a 160 bp sequence upstream of the transcription start site conferred a basal and an IL-6-inducible promoter activity.**

## Introduction

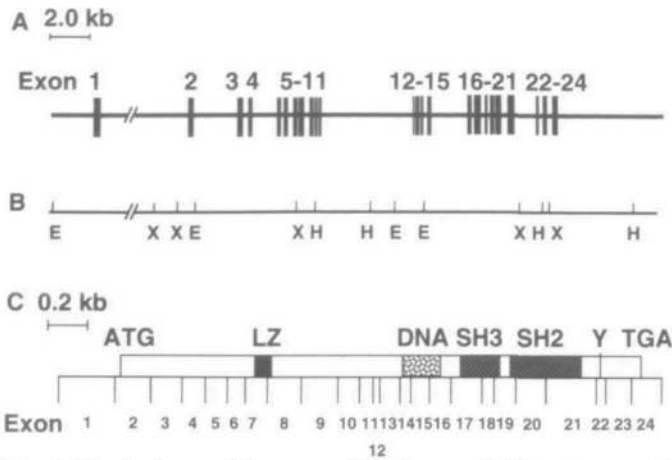
STAT proteins (signal transducers and activators of transcription) are latent transcription factors that contain *src* homology 2 (SH2), SH3-like domains and a C-terminal tyrosine phosphorylation site (1). These proteins become activated by phosphorylation on a specific tyrosine residue in response to a variety of cytokines and growth factors (2,3). The tyrosine phosphorylation of cytoplasmic STATs is mediated by the Janus protein kinase (JAK) family, which currently

consists of JAK1, JAK2, JAK3 and Tyk2. In the case of IL-6 and ciliary neurotrophic factor (CNTF) which share signal transducing receptor component gp130, JAK1, JAK2 and Tyk2 can be activated, although the selection of the kinases varies depending on the cell line. In the absence of JAK1, the phosphorylation of gp130 and the activation of STAT3 were greatly reduced, suggesting that JAK1 is essential for the gp130-mediated signaling pathway (4).

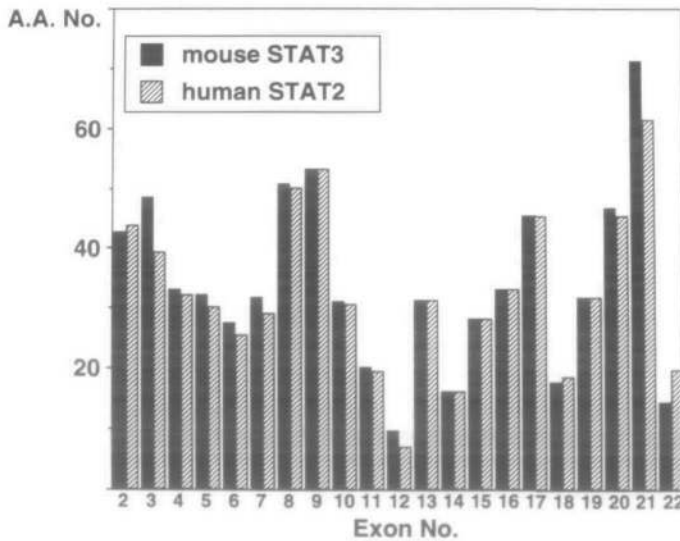
Correspondence to: S Akira

Transmitting editor: T. Watanabe

Received 28 March 1996, accepted 16 April 1996

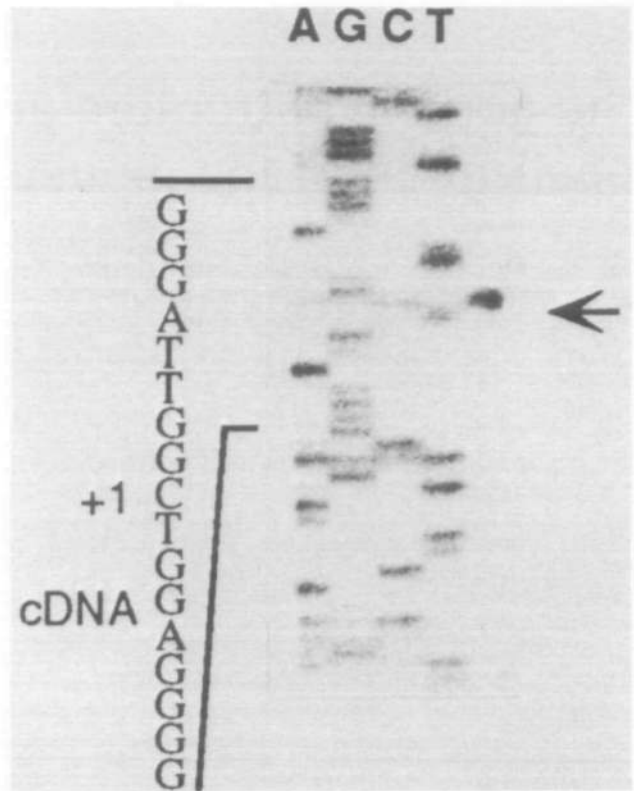


**Fig. 1.** Physical map of the mouse STAT3 gene (A) The structure of the STAT3 gene is represented with regard to the organization of the exons and introns. (B) The restriction map shows cleavage sites for restriction enzymes, *Xba*I (X), *Eco*RI (E) and *Hind*III (H) (C) The exons are scaled in mouse STAT3 cDNA. Translation initiates in exon 2. STAT3 leucine zipper domain (between amino acids 203 and 224) is scattered in exon 7 and 8, SH3 domain is in exon 17, 18 and 19, SH2 domain is split into exon 19, 20 and 21. Termination codon is in exon 24



**Fig. 2.** Comparison of exon length of STAT3 and STAT2. Amino acid numbers in each exon of mouse STAT3 and human STAT2 were compared.

Unstimulated STAT proteins are present in monomeric forms. When cells are stimulated with ligands, STATs associate with phosphorylated tyrosine residues of an activated receptor complex through the SH2 domain of STATs and become phosphorylated on the tyrosine residues by JAK kinases. The phosphorylated STATs then form homo- or heterodimers, and rapidly translocate to the nucleus, where they bind specific DNA sequences and regulate the transcription of the target genes (5). At present, six different members of STATs (STAT1 to STAT6) have been cloned and are involved in specific cytokine regulation (6–12).



**Fig. 3.** Determination of the transcription initiation site by primer extension. A <sup>32</sup>P-labeled 35-mer oligonucleotide was annealed with 20 µg of total RNA from mouse liver and extended using AMV reverse transcriptase (lane 1). The nucleotide sequencing ladder extended with the same primer is for calculation of the size of the extension product. The specific extension product for STAT3 mRNA is indicated by an arrow

In response to IFN- $\alpha/\beta$ , STAT1 and STAT2 are tyrosine phosphorylated, whereas STAT1 is tyrosine phosphorylated in response to IFN- $\gamma$  (13,14). STAT3 is predominantly phosphorylated in IL-6, leukemia inhibitory factor (LIF), CNTF, oncostatin M and IL-11 that share gp130 (15). In addition, granulocyte colony stimulating factor (G-CSF) (16), which utilizes a receptor highly related to gp130, also induces phosphorylation of STAT3. STAT4 and STAT6 are phosphorylated by IL-12 and IL-4 respectively. STAT5 is phosphorylated by a variety of cytokines including prolactin, IL-2, IL-3, IL-5 and granulocyte macrophage colony stimulating factor.

Recently it has been shown that selectivity of STAT activation is determined by a specific docking interaction between STATs and the phosphorylated tyrosine-containing sequences in the signal-transducing receptor components. STAT1 docks to the sequence YDKPH in the IFN- $\gamma$  receptor, while STAT3 docks to a consensus sequence YxxQ in the gp130 and LIF receptor cytoplasmic domains (17,18). In addition to the obligatory requirement of tyrosine phosphorylation for dimerization and DNA binding, phosphorylation of a C-terminal serine residue is required for maximal transcriptional activation of STATs (19).

The complete genomic structure of human STAT2 and

**Table 1.** Exon–intron and junction sequences of the mouse STAT3 gene

<u>Exon</u>			<u>Intron</u>			<u>Exon</u>
No.	Size (base)	Junction (cDNA position)	No.	Size kb	Junction (cDNA position)	No.
1	235	GACAGgtgaa	1	>15	tacagTCGAG	2
2	152	GACTGgtgag	2	2.50	cctagGGCAG	3
3	144	TGCAGgtgag	3	0.54	tacagAGCAG	4
4	99	CCCAGgtgag	4	1.88	ggcagCAAGG	5
5	96	TGCAGgtaat	5	0.20	tgacagGATCT	6
6	72	AGGAGgtaag	6	0.30	cttagACATG	7
7	95	GGAGAgtaag	7	0.14	aacagAGCAT	8
8	152	AACTGgtaag	8	0.51	ggcagGATAA	9
9	159	AAGAGgtaac	9	0.10	tccagTGCCCT	10
10	93	GTCAGgtggg	10	0.16	aatagGTTGC	11
11	59	GATAAgtaag	11	4.50	tacagAGACT	12
12	31	AGAGGgttaa	12	0.09	tttagGTCTC	13
13	94	ACCTGgtcgg	13	0.11	tccagACCCCT	14
14	48	GTGATgtaag	14	0.43	tgacagGCCTC	15
15	84	TAGAGgtgag	15	2.10	tacagACCCA	16
16	99	CCAAGgtgag	16	0.30	aatagAACGT	17
17	136	CCTAGgtttc	17	0.50	tccagGGCCT	18
18	53	GCAAAGttga	18	0.14	tccagGAAAA	19
19	95	GAAGGgtaag	19	0.07	tgacagGTACA	20
20	140	CAGTGgtaag	20	0.48	gacagGCAAG	21
21	213	AGGTAgttgt	21	1.34	tgacagGTGCT	22
22	43	ACACCgtaag	22	0.48	tacagAACGA	23
23	113	GTTTGgtgag	23	0.40	tgacagAGTCC	24
24	137	ATATCgtctc				

partial genomic structure of human STAT1 have been demonstrated (20). Here we report the complete mouse genomic organization of STAT3 and chromosomal location as well as the analysis of the promoter region of STAT3.

**Methods**

*Genomic library screening*

A mouse genomic DNA library derived from 129Sv liver in λFIX II (Stratagene, La Jolla, CA) was screened with the <sup>32</sup>P-labeled mouse STAT3 cDNA probe. Four different parts of STAT3 cDNA (5'-untranslated region, 5'-untranslated region, and amino acids 1–28, 280–440 and 626–770) were used for probe. Filters blotted with recombinant phage DNA were incubated in hybridization buffer (2.5×Denharts, 1% SDS, 10 mM EDTA, pH 8.0, 100 µg/ml of salmon sperm DNA, 6×SSC) with the labeled probe for 16 h at 65°C and then washed with 0.1×SSC/0.1% SDS, and autoradiographed.

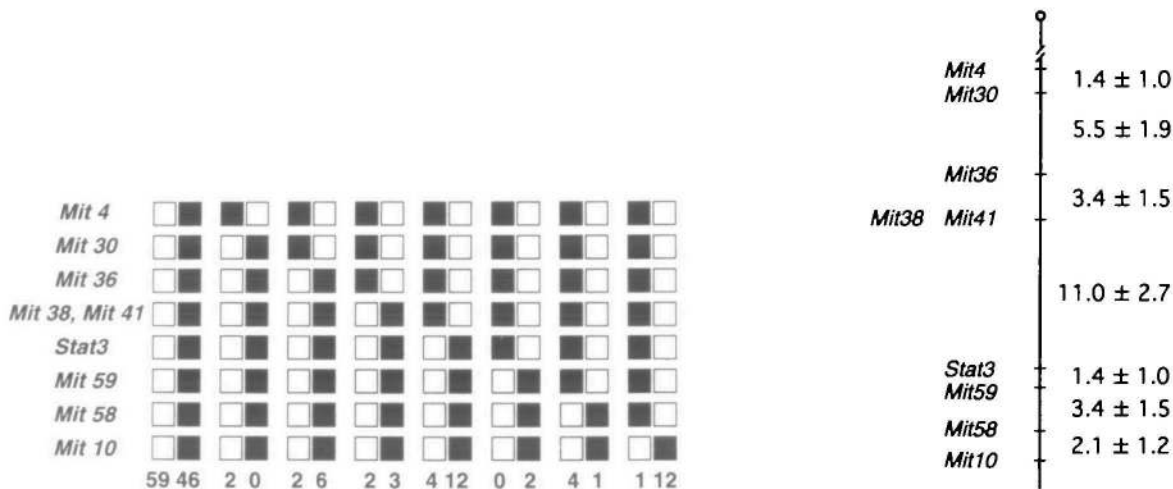
*Subcloning and sequencing*

To identify genomic fragments containing exons, bacteriophage DNA purified from each clone was digested by several restriction enzymes, separated by electrophoresis on 0.6% agarose gels and blotted on GeneScreen Plus (NEN Research Products, Boston, MA). The blot was hybridized with <sup>32</sup>P-labeled STAT3 cDNA. Appropriate restriction fragments were

subcloned into the plasmid vector, pBluescript KS<sup>+</sup> (Stratagene, La Jolla, CA). Sequencing was performed by using the Sequenase Version 2.0 DNA System (Amersham, Arlington Heights, IL) or The ABI 373A automated DNA sequencer (Foster City, CA) according to the manufacturer's instructions.

*Primer extension analysis*

The transcription initiation site was determined by a primer extension system (Promega, Madison, WI). The primer used for the reaction was a 35-mer oligonucleotide (5'-GGAA-CCCCAGGTTCCCCCTCCCTGTCTACACTGCA-3') complementary to the sequence in exon 1 of the mouse STAT3 gene, and was radiolabeled at the 5' end with T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol). The labeled primer was added to 25 µg of total RNA isolated from HepG2 cells or mouse liver in the presence of 50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol and 0.5 mM spermidine. The reaction mixtures were incubated at 60°C for 30 min and annealed at 30°C for 2 h. Avian myeloblastosis virus reverse transcriptase (1 U) was added and primer extension was performed at 42°C for 105 min. The products were electrophoresed on a 6% polyacrylamide gel containing 7 M urea and the length of the primer-extended product was determined by comparison with a dideoxynucleotide sequencing of the promoter region performed with an oligonucleotide complementary to exon 1.



**Fig. 6.** (a) Segregation of the STAT3 gene locus on mouse chromosome 11. Closed boxes represent the homozygous C57BL/6 pattern and open boxes represent the F<sub>1</sub> pattern. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. (b) Partial chromosome linkage maps showing the chromosomal locations of the STAT3 gene. Recombination distances between loci in cM are shown to the right of the chromosome.

6×*M. spretus*)F<sub>1</sub> females and *M. spretus* males. We examined the concordance of the RFLV segregation identified by Southern blot hybridization with the segregation of eight microsatellite marker loci, *D11Mit4*, *D11Mit30*, *D11Mit36*, *D11Mit38*, *D11Mit41*, *D11Mit59*, *D11Mit58* and *D11Mit10*, and the segregations of mouse *Stat3* in the backcross mice are shown with the microsatellite DNA markers as anchors in Fig. 6(a).

Comparative pairwise loci analysis showed that gene order and genetic distances in cM ± SE on chromosome 11 is: centromere (Hc11)–*D11Mit4*–1.4 ± 1.0 (2/146)–*D11Mit30*–5.5 ± 1.9 (8/146)–*D11Mit36*–3.4 ± 1.5 (5/146)–*D11Mit38*, *D11Mit41*–11.0 ± 2.7 (16/146)–*Stat3*–1.4 ± 1.0 (2/146)–*D11Mit59*–3.4 ± 1.5 (5/146)–*D11Mit58*–2.1 ± 1.2 (3/146)–*D11Mit10*–telomere (Fig. 6b) (22,23).

## Discussion

We have isolated and characterized the mouse STAT3 gene as well as the chromosomal localization. The mouse STAT3 gene consisted of 24 exons which spanned >37 kb long. The STAT3 gene mapped on mouse chromosome 11. When compared between the human STAT2 and mouse STAT3 genes, the exon–intron junction and the length of exons were almost identical, in spite of 38% amino acid homology between these proteins. As conserved exonic organization of STAT1 and STAT2 genes suggested (20), this result may indicate that the exon–intron organization had already been accomplished before these two genes duplicated, and then these genes evolved to respond to different ligands. During preparation of this manuscript, Copeland *et al.* (24) showed that the mouse STAT3 gene mapped to the distal region of chromosome 11 as demonstrated in our paper.

The C-terminal domain of the STATs is highly divergent. Within the carboxy region a potential mitogen-activated protein kinase site is found and phosphorylation of this site has been shown to be involved in the transcriptional activation. In the case of STAT1, a splice variant (STAT1β) lacks the C-terminal

38 amino acids including the MAPK site (25). This variant can be phosphorylated, translocate to the nucleus and bind to DNA, but cannot activate gene transcription. Similar variants have been identified for STAT2, STAT3 and STAT5 (26–28). In the case of STAT3, a variant named STAT3β lacks the C-terminal 55 amino acids which are replaced by seven other amino acids. Our present exon–intron boundary analysis show that this variant is an alternative splicing form in which exon 22 and an as yet unidentified exon are combined.

Although the nucleotide sequence around the transcription initiation site does not contain a TATA box-like sequence, the 5′-flanking region of the mouse STAT3 gene contained several potential *cis*-acting regulatory sequences such as binding sites for GATA, CRE, NF-IL-6, E-box, PEBP2, Sp-1, Pit-1 and AP-2. We previously observed that STAT3 gene expression was augmented in hepatocytes in response to IL-6. We demonstrated in this study that the 160 bp region from +38 to –165 conferred the IL-6 responsive promoter activity in a hepatoma cell line. At present two IL-6 responsive elements, a C/EBP binding site and a STAT3 binding site, are identified from the analysis of acute phase protein gene regulation (29). However, within the 160 bp region we could not find these binding motifs. This may indicate that other transcription factors are involved in the regulation of the IL-6-mediated STAT3 induction. Since STAT3 is expressed in embryonic stem cells which do not contain many transcription factors, it is also interesting to examine if there are any differences in the transcription factors involved in STAT3 expression in embryonic stem cells and in hepatocytes. Further studies are needed to elucidate the mechanism of STAT3 gene expression.

In conclusion, we succeeded in the isolation and characterization of the mouse STAT3 gene. Information of the exact exon–intron boundary of STAT3 obtained in this study will be helpful for screening novel splicing isoforms of STAT3 messenger in certain physiological and pathological conditions.

**Acknowledgements**

We thank Ms K. Kubota for her assistance and Dr T. Sugiyama for providing us with total RNA of HepG2.

**Abbreviations**

CNTF	ciliary neurotrophic factor
EGF	epidermal growth factor
G-CSF	granulocyte colony stimulating factor
GRE	glucocorticoid responsive elements
JAK	Janus protein kinase
LIF	leukemia inhibitory factor
PEBP2	polyomavirus enhancer-binding proteins
RFLV	restriction fragment length variant
SH2	Src homology 2
STAT	signal transducers and activators of transcription

**References**

- 1 Cohen, G. B., Ren, R. and Baltimore, D. 1995. Modular binding domains in signal transduction proteins. *Cell* 80:237
- 2 Ihle, J. N., Witthuhn, B. A., Quelle, F. W., Yamamoto, K., Thierfelder, W. E., Kreider, B. and Silvennoinen, O. 1994. Signaling by cytokine receptor superfamily: JAKs and STATs *Trends Biochem Sci.* 19:222
- 3 Darnell, J. E., Jr, Kerr, I. M. and Stark, G. M. 1994. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science* 264:1415.
- 4 Guschin, D., Rogers, N., Briscoe, J., Witthuhn, B., Watling, D., Horn, F., Pellegrini, S., Yasukawa, K., Heinrich, P., Stark, G. R., Ihle, J. N. and Korr, I. M. 1995. A major role for the protein tyrosine kinase JAK-1 in the JAK/STAT signal transduction pathway in response to interleukin-6. *EMBO J.* 14:1421.
- 5 Campbell, G. S., Meyer, D. J., Raz, R., Levy, D. E., Schwartz, J. and Carter-Su, C. C. 1995. Activation of acute phase response factor (APRF)/Stat3 transcription factor by growth hormone. *J. Biol. Chem.* 270:3974.
- 6 Fu, X.-Y., Schindler, C., Improt, T., Aebersold, R. and Darnell, J. E., Jr. 1992. The proteins of ISGF-3: the interferon- $\alpha$ -induced transcriptional activators define a gene family involved signal transduction. *Proc Natl Acad Sci. USA* 89:7840
- 7 Schindler, C., Fu, X.-Y., Improt, T., Aebersold, R. and Darnell, J. E., Jr. 1992. Proteins of transcription factor ISGF-3: one gene encodes the 91 and 84 kDa ISGF-3 proteins that are activated by interferon- $\alpha$ . *Proc. Natl Acad Sci. USA* 89:7836.
- 8 Akira, S., Nishio, Y., Inoue, M., Wang, X. J., Wei, S., Matsusaka, T., Yoshida, K., Sudo, T., Naruto, M. and Kishimoto, T. 1994. Molecular Cloning of APRF, a novel IFN-stimulated gene factor 3 p91-related transcription factor involved in the gp130-mediated signaling pathway. *Cell* 77:63
- 9 Hou, J. Z., Schindler, U., Henzel, W. J., Ho, T. C., Brasseur, M. and Meknight, S. L. 1994. An interleukin-4-induced transcription factor: IL-4 Stat. *Science* 265:1701.
- 10 Wakao, H., Gouilleux, F. and Groner, B. 1994. Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. *EMBO J.* 13:2182

- 11 Yamamoto, K., Quelle, F. W., Thierfelder, W. E., Kreider, B. L., Gilbert, D. T., Jenkins, N. A., Copeland, N. G., Silvennoinen, O. and Ihle, J. N. 1994. STAT4, a novel  $\gamma$  interferon activation site-binding protein expressed in early myeloid differentiation. *Mol. Cell. Biol.* 14:4342.
- 12 Zhong, Z., Wen, Z. and Darnell, J. E., Jr. 1994. Stat3 and Stat4: members of the family of signal transducers and activators of transcription. *Proc. Natl Acad Sci. USA* 91:4806
- 13 Levy, D. E., Kessler, D. S., Pine, R. and Darnell, J. E., Jr. 1989. Cytoplasmic activation of ISGF3, the positive regulator of interferon- $\alpha$ -stimulated transcription, reconstituted *in vitro*. *Genes Dev.* 3:1362.
- 14 Reich, N., Evans, B., Levy, D. E., Fahey, D., Knight, E., Jr and Darnell, J. E., Jr. 1987. Interferon-induced transcription of a gene encoding a 15 kDa protein depends on an upstream enhancer element. *Proc Natl Acad Sci. USA* 84:6394
- 15 Migone, T.-S., Lin, J.-X., Cereseto, A., Mulloy, J. C., O'Shea, J. J., Franchini, G. and Leonard, W. J. 1995. Constitutively activated Jak-STAT pathway in T cells transformed with HTLV-1. *Science* 269:79
- 16 Tian, S.-S., Lamb, P., Seidel, H. M., Stein, R. B. and Rosen, J. 1994. Rapid activation of the STAT3 transcription factor by granulocyte colony-stimulating factor. *Blood* 84:1760
- 17 Greenlund, A. C., Farrar, M. A., Viviano, B. L. and Schreiber, R. D. 1994. Ligand-induced IFN $\gamma$  receptor tyrosine phosphorylation couples the receptor to its signal transduction system (p91). *EMBO J.* 13:1591.
- 18 Stahl, N., Farruggella, T. J., Boulton, T. G., Zhong, Z., Darnell, J. E. and Yancopoulos, G. D. 1995. Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* 267:1349.
- 19 Wen, Z., Zhong, Z. and Darnell, J. E., Jr. 1995. Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* 82:241
- 20 Yan, R. Q., Qureshi, S., Zhong, Z., Wen, Z. and Darnell, J. E., Jr. 1995. The genomic structure of the STAT gene: multiple exons in coincident sites in Stat1 and Stat2. *Nucleic Acids Res.* 23:459.
- 21 Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459.
- 22 Lossie, A. C., Macphie, M., Buchberg, A. M. and Camper, S. A. 1994. Mouse chromosome 11. *Mammalian Genome* 5:S164.
- 23 Lyon, M. F. and Kirby, M. C. 1995. Mouse chromosome atlas. *Mouse Genome* 92:23.
- 24 Copeland, N. G., Gilbert, D. J., Schindler, C., Zhong, Z., Wen, Z., Darnell, J. E., Mui, A. L.-F., Miyajima, A., Quelle, F. W., Ihle, J. N. and Nenkins, N. A. 1995. Distribution of the mammalian Stat gene family in mouse chromosomes. *Genomics* 29:225
- 25 Ruff-Jamison, S., Chen, K. and Cohen, S. 1993. Induction by EGF and Interferon- $\gamma$  of tyrosine phosphorylated DNA binding proteins in mouse liver nuclei. *Science* 261:1733
- 26 Sugiyama, T., Nishio, Y., Kishimoto, T. and Akira, S. 1996. Identification of alternative splicing form of Stat2. *FEBS Lett.* 381:191.
- 27 Schaefer, T. S., Sanders, L. K. and Nzthans, D. 1995. Cooperative transcriptional activity of Jun and Stat3 $\beta$ , a short form of Stat3. *Proc. Natl Acad. Sci. USA* 92:9097
- 28 Ihle, J. N. 1996. STATs: signal transducers and activators of transcription. *Cell* 84:331.
- 29 Akira, S. and Kishimoto, T. 1992. IL-6 and NF-IL6 in acute-phase response and viral infection. *Immunol. Rev.* 127:25.