

Cytokine-Related Genes Identified From the RIKEN Full-Length Mouse cDNA Data Set

Vladimir Brusic,^{1,2,10} Rekha S. Pillai,¹ Diego G. Silva,^{3,4} Nikolai Petrovsky,^{2,3} RIKEN GER Group^{5,8} and GSL Members,^{6,9} and Christian Schönbach⁷

¹Institute for Infocomm Research, Singapore 119613; ²Medical Informatics Centre, Division of Science and Design, University of Canberra, Bruce ACT 2617, Australia; ³Autoimmunity Research Unit, The Canberra Hospital, Garran ACT 2606, Australia; ⁴John Curtin School of Medical Research, Australian National University, Canberra ACT 2601, Australia; ⁵Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan; ⁶Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan; ⁷Biomedical Knowledge Discovery Team, Bioinformatics Group, RIKEN Genomic Sciences Center (GSC), Yokohama 230-0045, Japan

To identify novel cytokine-related genes, we searched the set of 60,770 annotated RIKEN mouse cDNA clones (FANTOM2 clones), using keywords such as cytokine itself or cytokine names (such as interferon, interleukin, epidermal growth factor, fibroblast growth factor, and transforming growth factor). This search produced 108 known cytokines and cytokine-related products such as cytokine receptors, cytokine-associated genes, or their products (enhancers, accessory proteins, cytokine-induced genes). We found 15 clusters of FANTOM2 clones that are candidates for novel cytokine-related genes. These encoded products with strong sequence similarity to guanylate-binding protein (GBP-5), interleukin-1 receptor-associated kinase 2 (IRAK-2), interleukin 20 receptor α isoform 3, a member of the interferon-inducible proteins of the Ili 200 cluster, four members of the membrane-associated family I-8 of interferon-inducible proteins, one p27-like protein, and a hypothetical protein containing a Toll/Interleukin receptor domain. All four clones representing novel candidates of gene products from the family contain a novel highly conserved cross-species domain. Clones similar to growth factor-related products included transforming growth factor β -inducible early growth response protein 2 (TIEG-2), TGF β -induced factor 2, integrin β -like 1, latent TGF-binding protein 4S, and FGF receptor 4B. We performed a detailed sequence analysis of the candidate novel genes to elucidate their likely functional properties.

[Supplemental material is available online at www.genome.org]

Cytokines are polypeptides secreted by immune cells that function as humoral regulators modulating functional activities of cells and tissues including cellular interactions and processes. Cytokine genes often encode functional variants through alternative splicing, resulting in proteins that may have slightly different biological activities (Ibelgaufts 1999). Cytokines are typically mutually unrelated, but some may be grouped in families.

Activated cytokine-producing cells often release many different cytokines, resulting in complex networks of interacting signals. Cytokines are crucial in immune cell activation, cell-to-cell communication, signal transduction, mitosis, cell survival, death, and transformation. In addition to their involvement in initiation and regulation of immune responses, cytokines are also important in embryogenesis, organ development, and neuronal processes (Ibelgaufts 1999). Clinical uses of cytokines include application in hematopoiesis (Smith 1990), tumor immunotherapies (Cao et al. 1998; Rosenberg 2001; Homey et al. 2002), bone marrow transplan-

tation (Miyazaki and Kato 1999), gene therapy (Jeschke et al. 2001), and vaccine formulations (Knutson and Disis 2001).

The function of a particular cytokine can be assessed in vivo in transgenic or gene knock-out animals by studying the effects of an excess or lack of a particular cytokine gene product. It is of great importance for cytokine research, therefore, that the counterparts of human cytokine, cytokine receptors, and cytokine-related genes are identified in animals (e.g., the mouse) suitable for genetic manipulation. Because of the remarkable similarity between human and mouse genomes and their closely related biochemical, physiological, and pathological pathways, the mouse is an extremely useful animal model for immunological studies.

The FANTOM2 collaboration (Okazaki et al. 2002) involved functional annotation of a set of 60,770 RIKEN mouse cDNA clones (FANTOM2 clones). The FANTOM2 data set, the most complete picture of the mouse transcriptome to date, provides an ideal opportunity for identification of novel mouse genes, including cytokine-related genes. The annotation process included data from microarray expression, mapping and protein interactions, and sequence similarity search analyses. The annotation was performed by large-scale, computerized annotation combined with manual curation by human experts. We analyzed the FANTOM2 annotations, specifically looking for novel and existing mouse cytokines and

⁸Takahiro Arakawa.

⁹Piero Carninci, Jun Kawai, and Yoshihide Hayashizaki.

¹⁰Corresponding author.

E-MAIL vladimir@i2r.a-star.edu.sg; FAX 65-6774-8056.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1016503>.

cytokine-related genes. In this study, we focused on genes related to interferons (IFN), interleukins (IL), growth factors (epidermal growth factors [EGF], fibroblast growth factors [FGF], and transforming growth factors [TGF]), and small inducible cytokines (SIC). Some cytokines, such as tumor necrosis factors (TNF), belong to superfamilies that also include non-cytokines. Because this study was based on matching cytokine names from descriptions of database entries, cytokine families that belong to broader superfamilies were excluded from the analysis. Further study is required for a comprehensive analysis of all known cytokines and cytokine-related genes that have not been included in this work. The FANTOM2 clones contained more than 100 previously identified cytokine and cytokine-related genes. We also identified several candidates for novel mouse cytokine-related genes.

RESULTS

Of the 60,770 FANTOM2 clones, 256 had annotations that contained a cytokine name. A total of 222 clones had high-sequence identity (>96% over 50% length of the reference sequence) to 108 known mouse genes, as listed in the major databases. Of these, clones represented 27 known interferon-related genes (Table 1). The known mouse genes in the FANTOM2 set comprised 1 interferon (IFN- γ), 2 IFN receptors (IFN $\alpha\beta$ receptor, and IFN $\alpha\beta$ receptor 2), 20 inducible/activatable genes, and 4 other IFN-related genes. Search of the protein translation of the GenBank database (Benson et al. 2002, as of September 2002) revealed that there were at least 15 nonredundant entries of various mouse IFNs, 10 IFN receptors, 190 IFN inducible or activatable proteins, and 60 IFN regulatory proteins.

A total of 77 clones represented 39 known interleukin-related genes (Table 2). The known mouse genes in the FANTOM2 set comprised 11 interleukins (IL) or chains thereof (IL-1 δ , IL-1 ϵ , IL-6, IL-7, IL-1F9, IL-16, IL-17B, IL-20, IL-23p19, and IL-25), 19 interleukin receptors (ILR-11, ILR-1-1, ILR-2 α , ILR-2 β , ILR-2 γ , ILR-4 α , ILR-6 α , ILR-7, ILR-9, ILR-10 β , ILR-12 β 1, ILR-12 β 2, ILR-13 α 1, ILR-13 α 2, ILR-15 α , ILR-17, ILR-18-1, ILR-21, and ILR-23 α p19), and 9 other IL-related genes. Search of the protein translation of the GenBank database revealed that there are at least 40 nonredundant entries of mouse ILs (including variants and subunits), 50 IL receptors, and 80 other IL-related proteins. These results are in agreement with an earlier finding (Staudt and Brown 2000), which stated that only a half of human ILs, mainly those widely expressed, were found in the EST libraries. The ILs reliant on the activation of immune cells, namely IL-2, IL-3, IL-5, IL-9, IL-11, IL-12 β , IL-14, and IL-17, were lacking in the EST libraries.

A total of 61 clones represented 23 known growth factor (EGF, FGF, and TGF)-related genes (Table 3). The known mouse genes in the FANTOM2 set comprised 5 growth factors (EGF, FGF 14, FGF 15, TGF β , and TGF β 68 kD) and 18 other growth factor-related genes. Search of the protein translation of the GenBank database revealed that there are at least 140 nonredundant entries of various mouse growth factor (EGF, FGF, and TGF)-related genes (including variants and subunits).

Finally, 30 clones represented 19 known small inducible cytokine-related genes in mouse (Table 4). Search of the protein translation of the GenBank database revealed that there are at least 38 nonredundant entries of various small inducible cytokines in mouse.

Clones representing 17 known mouse cytokine-related genes were derived from cDNA libraries from the non-obese diabetic (NOD) mouse. All of the remaining cytokine-related clones were derived from the C57BL/6 mouse strain. Six known genes were only found from the NOD mouse libraries. Clones representing a total of 11 known mouse genes were found in both C57BL/6 and NOD mouse cDNA libraries. The clones representing known cytokine-related genes were derived from a broad variety of RIKEN cDNA libraries. These libraries were generated from various tissues and developmental stages. Because this analysis focused on name matching from both the FANTOM2 data set and GenBank database, we can estimate that FANTOM2 clones cover ~20% of existing known cytokine-related genes. Cytokines induce cascades of cytokine networks as well as of regulatory molecules. Often it is unclear whether the induced molecule acts as a cytokine or has some other regulatory role. In this study, we consider all clones as cytokine related and did not discriminate between the two groups.

We have also identified 15 candidate novel cytokine or cytokine-related genes (Table 5), including 7 candidate novel mouse IFN-related genes (13 clones), 3 candidates for novel IL-related genes (5 clones), and 4 candidates for novel growth factor-related genes (3 TGF- and 1 EGF-related) (16 clones). All of the clones representing candidate novel genes were derived from the C57BL/6 mouse and showed \leq 96% identity over >50% length of the reference sequence.

Interferon-related Novel Gene Candidates

The clones 1110004C05, 4930507H06, A330075D07, and 1110036C17 represent four putative novel members of the membrane-associated family 1-8 of interferon-inducible proteins (Fig. 1). Sequences of these clones showed similarity to a variety of interferon-induced genes and proteins in human (Lewin et al. 1991), rat (Hayzer et al. 1992), cattle (Pru et al. 2001), and trout (SPTR Q8QFL3). Members of the 1-8 family of interferon (IFN)-inducible genes encode proteins thought to modulate cellular growth and adhesion wound healing, and inflammatory responses. It was hypothesized that they may function as membrane transport proteins (Pru et al. 2001). The 1-8 proteins participate in multimeric complexes involved in the transduction of homotypic and antiproliferative adhesion signals. Sequences of all four clones of novel candidate members of the 1-8 family contain a novel (not found in motif databases), motif DxxxWSxxxxxxNxx-CLGxxAxxxSxKSRDRKMVGdxxGAXxxxxTA that is conserved across species (fish, rat, mouse, and human). Furthermore, sequence of the clone 1110004C05 (Fig. 1) contains the motif YxxxxCLxI that conserves critical residues found in the active sites of human E2 ubiquitin conjugating enzymes (Pru et al. 2001). Sequences of the clones 4930507H06 and 1110036C17 contain similar motif FxxxxCLxI, whereas the remaining clone A330075D07 does not contain the corresponding motif. Each of the four candidate new members of the 1-8 family showed highest similarity to a distinct database entry and therefore each represents a distinct transcript (data not shown). The amino acid differences between the members of the 1-8 family at carboxy- and amino-termini may contribute to subtle functional differences or cellular localization.

Sequence of the clone 2310061N23 showed similarity to the gene for the IFN- α -inducible protein p27-h found in hamster cardiomyopathic tissue (Denovan-Wright et al. 2000) and human gene TLH29 (Q9H2X8) from liver tissue (Fig. 2). The

Table 1. The List of 54 Clones Representing 27 Known Interferon-Related Genes in Mouse Found in the FANTOM2 Data Set

| | Clone ID | Mouse strain | cDNA library | Gene name |
|----|------------|--------------|-------------------------------|--|
| 1 | 5830458K16 | C57BL/6 | Thymus | 28-kD interferon α responsive protein (5830458K16RIK protein) homolog (<i>Mus musculus</i>) |
| | 1810013G24 | C57BL/6 | Pancreas | |
| 2 | 0710005K22 | C57BL/6 | Brain | Genes associated with retinoid-IFN-induced mortality 19 |
| | 0710005K22 | C57BL/6 | Brain | |
| | 2700054G14 | C57BL/6 | Embryo11 | |
| | 2410016H22 | C57BL/6 | ES | |
| 3 | A430075E22 | C57BL/6 | N0 Thymus | Interferon (α and β) receptor |
| 4 | 1100001K09 | C57BL/6 | 18-day embryo | |
| | 2510009P07 | C57BL/6 | Embryo 13-Liver | Interferon (α and β) receptor 2 |
| | E430025C02 | NOD | (NOD)N2 Thymic cells | |
| 6 | A730048F03 | C57BL/6 | N7 Cerebellum | Interferon-activated gene 202A (noncoding RNA) |
| 7 | A430077M01 | C57BL/6 | N0 Thymus | Interferon-activated gene 205 |
| | A530058P16 | C57BL/6 | Aorta and vein | Interferon-inducible protein 203 (IFI-203) (interferon-inducible protein P203) |
| | A530093G11 | C57BL/6 | Aorta and vein | |
| | E430077M01 | NOD | (NOD)N2 Thymic | Interferon consensus sequence-binding protein |
| 8 | 9030603A05 | C57BL/6 | Colon | |
| | 9830117K07 | C57BL/6 | Bone | |
| 9 | A830021G01 | C57BL/6 | N10 Cortex | |
| 10 | F830002I10 | NOD | Activated spleen from NOD.Cz | |
| | F830032O20 | NOD | Activated spleen from NOD.Cz | Interferon γ |
| 11 | 0610039K12 | C57BL/6 | Kidney | |
| | E430012H16 | NOD | (NOD)N2 Thymic cells | Interferon γ induced GTPase |
| 12 | 5730557A06 | C57BL/6 | Embryo 8 | |
| 13 | 2010205D02 | C57BL/6 | Small intestine | Interferon γ inducible protein 30 |
| | 1110003K24 | C57BL/6 | 18-day embryo | |
| | 2310040F23 | C57BL/6 | Tongue | Interferon γ inducing factor-binding protein |
| | 2310075I12 | C57BL/6 | Tongue | |
| | 2310047H18 | C57BL/6 | Tongue | |
| 14 | 0610011H23 | C57BL/6 | Kidney | Interferon inducible protein 1 |
| 15 | 9830146E22 | C57BL/6 | Bone | |
| | E430001A19 | NOD | (NOD)N2 Thymic | Interferon regulatory factor 2 |
| 16 | F730005C13 | C57BL/6 | B6 derived | |
| | F830028K14 | NOD | Activated spleen | Interferon regulatory factor 4 |
| 17 | A130064P06 | C57BL/6 | N16 Thymus | |
| | E230028I05 | C57BL/6 | P2 oviduct | Interferon regulatory factor 6 |
| 18 | 0610039E12 | C57BL/6 | Kidney | |
| | 1110062H15 | C57BL/6 | 18-day embryo | Interferon regulatory factor 7 |
| | A430014N21 | C57BL/6 | N0 Thymus | |
| | F830035B02 | NOD | Activated spleen | |
| 19 | 2010008K16 | C57BL/6 | Small intestine | Interferon-induced 35-kD protein homolog (IFP 35) homolog (<i>Mus musculus</i>) |
| 20 | 2010002M12 | C57BL/6 | Small intestine | Interferon-induced protein with tetratricopeptide repeats 1 |
| | F830004H17 | NOD | Activated spleen | |
| 21 | A130072L05 | C57BL/6 | N16 Thymus | Interferon-induced protein with tetratricopeptide repeats 2 |
| | A630085F01 | C57BL/6 | N3 Thymus | |
| 22 | 5730412C13 | C57BL/6 | Embryo 8 | Interferon-induced protein with tetratricopeptide repeats 3 (IFIT-3) glucocorticoid-attenuated response gene 49 |
| | 5031412D17 | C57BL/6 | Ovary & Uterus (Preg. 11 day) | |
| 23 | F830009I11 | NOD | Activated spleen | |
| 24 | A430077M01 | C57BL/6 | N0 Thymus | interferon-inducible GTPase |
| 25 | 0610030F02 | C57BL/6 | Kidney | Interferon-inducible protein 203 (<i>Mus musculus</i>) |
| 26 | 2900034J12 | C57BL/6 | Hippocampus | Interferon-related developmental regulator 2 |
| 27 | 1600023I01 | C57BL/6 | Placenta | Interferon-stimulated protein (15 kD) |
| | 2010006M08 | C57BL/6 | Small intestine | Interferon-stimulated protein (20 kD) |
| | 2010009P14 | C57BL/6 | Small intestine | |
| | 2010107M23 | C57BL/6 | Small intestine | |

p27-h gene showed increased expression in cardiac and skeletal muscle tissue undergoing active necrosis. These sequences are similar (data not shown) to the human IFN- α -inducible protein p27 (GenBank NM_005532) shown to have increased expression in breast carcinoma (Rasmussen et al. 1993). The p27 protein is relatively small (11 kD) and highly hydrophobic. The p27-like mRNA is present in multiple tissues treated with the TNF, and it was speculated that these genes may be linked to inflammation related to the necrotic

processes. However, the specific function of p27 has yet to be determined.

Clone A430075K09 represents a potential new member of a group of structurally related interferon-inducible proteins, the Ifi 200 cluster, that regulate biological activities of IFNs. The IFNs are immune response modulating cytokines that exhibit antiviral effects, induce expression of class I or class II major histocompatibility complex proteins, and help activate macrophages, natural killer cells, and neutrophils. Se-

Table 2. The List of 77 Clones Representing 39 Known Interleukin-Related Genes Found in the FANTOM2 Data Set

| | Clone ID | Mouse strain | cDNA library | Gene name |
|----|------------|--------------|-----------------------------------|--|
| 1 | 2210418I04 | C57BL/6 | Stomach | Interleukin 1 family, member 5 (δ) |
| | 4632413N13 | C57BL/6 | Skin (Neonate 0 day) | |
| | 2310041K07 | C57BL/6 | Tongue | |
| | 2310063B08 | C57BL/6 | Tongue | |
| 2 | 1110033G16 | C57BL/6 | 18-day embryo | Interleukin 1 family, member 6 (ε) |
| | 9330131B06 | C57BL/6 | Diencephalon | |
| | 9330171A08 | C57BL/6 | Diencephalon | |
| 3 | A330066C15 | C57BL/6 | Spinal cord | Interleukin 1 receptor accessory protein |
| | B230304B05 | C57BL/6 | C.quadrigemina | |
| | C630028F01 | C57BL/6 | Hippocampus | |
| 4 | F630041P17 | NOD | NOD-derived | Interleukin 1 receptor antagonist |
| | 4632427H17 | C57BL/6 | Skin | |
| 5 | C130072K02 | C57BL/6 | E16 Head | Interleukin 1 receptor, type I |
| | E230023C01 | C57BL/6 | P2 oviduct | |
| 6 | B230327C20 | C57BL/6 | C.quadrigemina | Interleukin 1 receptor-associated kinase |
| | D830047P08 | C57BL/6 | N16 heart | |
| 7 | 9030010H19 | C57BL/6 | Colon | Interleukin 1 receptor-like 1 |
| 8 | A530048E20 | C57BL/6 | Aorta and vein | Interleukin 10 receptor, β |
| 9 | 1500012D04 | C57BL/6 | Cerebellum | Interleukin 10-related T cell-derived inducible factor |
| 10 | A430092F05 | C57BL/6 | N0 Thymus | Interleukin 12 receptor, β 1 |
| 11 | A430056E01 | C57BL/6 | N0 Thymus | Interleukin 12 receptor, β 2 |
| | A530070O09 | C57BL/6 | Aorta and vein | |
| 12 | C430004G12 | C57BL/6 | E7 whole body | Interleukin 13 receptor, α 1 |
| | G430044I06 | C57BL/6 | Mixed ^a | |
| 13 | F830008L10 | NOD | Activated spleen | Interleukin 13 receptor, α 2 |
| 14 | D630025K20 | C57BL/6 | N0 kidney | Interleukin 15 |
| | A530045E03 | C57BL/6 | Aorta and vein | |
| 15 | A530048H17 | C57BL/6 | Aorta and vein | Interleukin 15 receptor, α chain |
| | A630048I20 | C57BL/6 | N3 Thymus | |
| 16 | A130095H14 | C57BL/6 | N16 Thymus | Interleukin 16 |
| 17 | 9830138F10 | C57BL/6 | Bone | Interleukin 17 receptor |
| | A530085G20 | C57BL/6 | Aorta and vein | |
| 18 | 1110006O16 | C57BL/6 | 18-day embryo | Interleukin 17B |
| 19 | E030029C19 | C57BL/6 | N0 lung | Interleukin 18 receptor 1 |
| | F630101C07 | NOD | NOD-derived | |
| 20 | E030025K24 | C57BL/6 | N0 lung | Interleukin 18 receptor accessory protein |
| | E430021B09 | NOD | (NOD) N2 Thymic | |
| 21 | A130010M06 | C57BL/6 | N16 Thymus | Interleukin 2 receptor, α chain |
| | 5430409O05 | C57BL/6 | Head (Neonate 6 day) | |
| 22 | 9530010F16 | C57BL/6 | U.bladder | Interleukin 2 receptor, β chain |
| | A130027N13 | C57BL/6 | N16 Thymus | |
| 23 | 7530415H19 | C57BL/6 | Eyeball (adult) | Interleukin 2 receptor, γ chain |
| 24 | A430056E07 | C57BL/6 | N0 Thymus | Interleukin 20 |
| 25 | A830047G15 | C57BL/6 | N0 Thymus | Interleukin 21 receptor |
| 26 | 1110033L24 | C57BL/6 | N10 Cortex | Interleukin 23, α subunit p19 |
| | 2210408F01 | C57BL/6 | 18-day embryo | |
| 27 | 2010016N09 | C57BL/6 | Stomach | Interleukin 25 |
| | 5830455M19 | C57BL/6 | Small intestine | |
| 28 | 9330197K12 | C57BL/6 | Thymus | Interleukin 4 receptor, α |
| | A630083M23 | C57BL/6 | Diencephalon | |
| 29 | E430003K23 | NOD | N3 Thymus | Interleukin 6 |
| | F830019G17 | NOD | (NOD) N2 Thymic | |
| 30 | 9530096I01 | C57BL/6 | Activated spleen from NOD.Cz Idd3 | Interleukin 6 receptor, α |
| | B230372K20 | C57BL/6 | U.bladder | |
| 31 | C230031H18 | C57BL/6 | C.quadrigemina | Interleukin 7 |
| | A430091K05 | C57BL/6 | N0 Cerebellum | |
| 32 | A530098I02 | C57BL/6 | N0 Thymus | Interleukin 7 receptor |
| | A630007F02 | C57BL/6 | Aorta and vein | |
| 33 | D430026C11 | C57BL/6 | N3 Thymus | Interleukin 9 receptor |
| | A430089D02 | C57BL/6 | E13 lung | |
| 34 | A530022C19 | C57BL/6 | N0 Thymus | Interleukin enhancer-binding factor 2 |
| | A630041A17 | C57BL/6 | Aorta and vein | |
| 35 | A630076J23 | C57BL/6 | N3 Thymus | Interleukin 9 receptor |
| | E430023N04 | NOD | N3 Thymus | |
| 36 | 2700043J07 | C57BL/6 | (NOD) N2 Thymic | Interleukin enhancer-binding factor 2 |
| | 3000002D13 | C57BL/6 | Embryo 11 | |
| | 6230405A16 | C57BL/6 | Embryo 12-head | |
| | 6330441H14 | C57BL/6 | Head (embryo 11) | |
| 37 | B130010B06 | C57BL/6 | Medulla | Interleukin enhancer-binding factor 2 |
| | | C57BL/6 | Partenogenetic | |

(continued)

Table 2. *Continued*

| | Clone ID | Mouse strain | cDNA library | Gene name |
|----|------------|--------------|---------------------|--|
| 35 | 5730447J21 | C57BL/6 | Embryo 8 | Interleukin enhancer-binding factor 3 |
| | C130034D23 | C57BL/6 | E16 Head | |
| | E430021F21 | NOD | (NOD) N2 Thymic | |
| 36 | 3200002B17 | C57BL/6 | Embryo (14+17)-Head | Interleukin-four induced gene 1 |
| | E330015O10 | C57BL/6 | P2 ovary | |
| 37 | E130317D06 | C57BL/6 | N0 eyeball | Nuclear factor, interleukin 3, regulated |
| 38 | C130076J06 | C57BL/6 | E16 Head | <i>Mus musculus</i> IL-1F9 (IL1F9) |
| 39 | 1110054H05 | C57BL/6 | 18-day embryo | <i>Mus musculus</i> similar to Interleukin enhancer-binding factor 1 |

^aMixed library (AL,AK,AD,AF,AG,AC,AE,AQ,AP,AH,AA,AO,AB,AI,AJ,AN,AM).

quence of the clone A430075K09 shows similarity (84.9%ID, 425 amino acids, 100% of the total length) to the Ifi 205, the new member of the Ifi 200 family (Table 5 and Fig. 4, supplemental research data; available online at www.genome.org). Little is known about the precise function of the Ifi 205. However, it contains the PAAD/DAPIN/Pyrim domain (Pfam 02758), which is present in other members of the Ifi 200 family, indicating a possible protein-protein interaction site. This domain was predicted to have the same six α -helices-containing fold as the well-described Death Domain (Pfam 00531). The death domain is present in the TNFR-1 and FAS/APO1 receptors involved in cell death (Hofmann and Tschopp 1995). The potential functions of our candidate novel Ifi 205 factor, therefore, is the inhibition of cell proliferation and the arrest of cell growth, possibly through modulation of the activity of several transcription factors (Gribaudo et al. 1999).

Sequence of the clone 5330409J06 showed a weak similarity (Table 5 and Fig. 5, supplemental research data) to the guanylate-binding protein 5—GBP5 (P26376), published only as a database entry (GenBank AF288815). To our knowledge, the function and structure of GBP5 has not been described or published. In humans, the members of the GBP family are the most abundant class of proteins induced by IFN- γ (Prakash et al. 2000). GBPs have the ability to hydrolyze guanosine triphosphate (GTP) to both mono- (GMP) or di-phosphate (GDP) products (Schwemmler and Staeheli 1994) involved in cellular signal transduction. However, little is known about either the specific biological function of GBPs or their cellular localization (Vestal et al. 2000). Members of the GBP family were reported to have antiviral properties (Anderson et al. 1999), play a role in the macrophage activation processes (Han et al. 1998), and inhibit proliferation of endothelial cells (Guenzi et al. 2001). The clone 5330409J06 contains GBP (Pfam 002263) and GBP-C (Pfam 02841) domains. It is therefore a strong candidate for a novel gene—the mouse guanylate-binding protein mGBP-5.

Interleukin-Related Novel Gene Candidates

Sequence of the clone B430113A10 contains Toll/Interleukin receptor TIR domain structure and shows similarity (88% identity in 123 AA overlap) to the hypothetical 14.3-kD protein in long-tailed macaque (Table 5 and Fig. 6, supplemental research data). The TIR domain (Pfam 01582) is an intracellular-signaling domain found in MyD88, interleukin 1 receptor, and the Toll receptor. The Toll/Interleukin receptor domain is associated with a family of proteins that are involved in early responses to injury and infection (e.g., interleukin-1 receptor) (Slack et al. 2000). The Toll-like receptors in insects

and mammals mediate innate defensive responses against a variety of microbial products (Anderson 2000). The clone B430113A10 thus represents a hypothetical protein potentially involved in immune responses.

Sequence of the clone 4732448K15 is similar (69% identity in full-length sequence of 598 AA) to the human IL-1 receptor-associated kinase-2 (IRAK-2), a proximal mediator of IL-1 signaling (Table 5 and Fig. 7, supplemental research data). IRAK-2 binds to the IL-1 receptor and is involved in IL-1-dependent signaling (Muzio et al. 1997). The clone 4232448K15 contains a Death Domain (Pfam 00531) (Hofmann and Tschopp 1995) present in many apoptosis pathway proteins. It also contains a catalytic Protein Kinase domain (Hanks and Hunter 1995) that has a conserved catalytic core common in both serine/threonine and tyrosine protein kinases. It is therefore a strong candidate for defining a novel gene—the mouse interleukin-1 receptor-associated kinase 2—mIRAK-2.

Sequence of the clone E230031K19 is weakly similar (76.3% identity over 93.3% length) to the human interleukin 20 receptor α isoform 3 (Table 5 and Fig. 8, supplemental research data). The conceptual translation (amino acids 1–292, PFAM 01108) contains a tissue factor motif (Muller et al. 1996) that is involved in blood coagulation. IL-20 was reported to be overexpressed in transgenic mice that have skin abnormalities similar to psoriatic lesions in humans (Dumoutier and Renaud 2002).

Growth Factor-Related Novel Gene Candidates

The FANTOM2 clone D430010O15 is a representative of a cluster (TIEG2-like) that showed similarity to human TGF β -inducible early growth response protein 2 (TIEG2) (Table 5 and Fig. 9, supplemental research data). The conceptual translation of 9830142A17 showed a slightly higher similarity (75% identity) to the human, than to mouse TIEG2 (O89091) (74% identity). TIEG2 is ubiquitously expressed in human tissues, enriched in pancreas and muscle (Cook et al. 1998). Similarly, the TIEG2-like FANTOM2 clones were derived from eight different libraries. The TIEG2-like clone has a conserved zinc finger domain. TIEG2 is involved in cell growth, and its overexpression in hamster ovary cells inhibits cell proliferation (Cook et al. 1998).

Sequence of the clone 4921501K24 showed similarity—93.7% identity over the full length—to (Table 5 and Fig. 10, supplemental research data) to TGF- β -induced factor 2 (TGIF2), a homeobox gene of the TALE superclass, thought to play a role in the development and progression of some ovarian tumors through a gene-amplification mechanism (Imoto et al. 2000). TGIF2 was shown to repress TGF- β -responsive

Table 3. The List of 61 Clones Representing 23 Known Growth Factor-Related Genes Found in the FANTOM2 Data Set

| | Clone ID | Mouse strain | cDNA library | Gene name |
|----|------------|--------------|--------------------------|--|
| 1 | D730020G24 | C57BL/6 | L2 mammary | Epidermal growth factor |
| | D430013M15 | C57BL/6 | E13 lung | |
| | E430016A08 | NOD | (NOD) N2 Thymic cells | |
| 2 | 1300005M11 | C57BL/6 | Liver | Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3) homolog (<i>Mus musculus</i>) |
| | 1300008I23 | C57BL/6 | Liver | |
| | 1300003K07 | C57BL/6 | Liver | |
| 3 | C630023M09 | C57BL/6 | Hippocampus | Epidermal growth factor receptor pathway substrate 15 |
| | 9430061P05 | C57BL/6 | Upper body | |
| 5 | 9430070A01 | C57BL/6 | E12 Upper body | Epidermal growth factor receptor pathway substrate 15, related sequence |
| | | | | Epidermal growth factor receptor pathway substrate 15, related sequence |
| 6 | 9830147J04 | C57BL/6 | Bone | Epidermal growth factor receptor pathway substrate 15, related sequence |
| | | | | Epidermal growth factor receptor pathway substrate 15, related sequence |
| 7 | 5230401N02 | C57BL/6 | Xiphoid | Epidermal growth factor-containing fibulin-like extracellular matrix protein 1 |
| | | | | Epidermal growth factor-containing fibulin-like extracellular matrix protein 2 |
| 9 | 0610042A18 | C57BL/6 | Kidney | Eukaryotic translation initiation factor 3 subunit 2 (EIF-3 β) (EIF P36) (TGF- β receptor interacting protein) (TRIP-1) homolog (<i>Mus musculus</i>) |
| | 2410126D05 | C57BL/6 | ES | |
| | 2700045C12 | C57BL/6 | Embryo 11 | |
| | 4930503E24 | C57BL/6 | Testis | |
| | 1190024F07 | C57BL/6 | 18-day embryo | |
| 10 | 2010004G08 | C57BL/6 | Small intestine | Fibroblast growth factor (acidic) intracellular-binding protein |
| | D930040K23 | C57BL/6 | E15 head | |
| | 2810001F12 | C57BL/6 | Embryo 10 + Embryo 11 | |
| | 2010005N21 | C57BL/6 | Small intestine | |
| | 3010027N18 | C57BL/6 | Embryo 12-Head | |
| | G430069F17 | C57BL/6 | Mixed ^a | |
| | 5730551C10 | C57BL/6 | Embryo 8 | |
| | D030003J18 | C57BL/6 | Embryo 9 | |
| | D230045H09 | C57BL/6 | E12 eyeball | |
| | 0610042J09 | C57BL/6 | Kidney | |
| 12 | 2810484E09 | C57BL/6 | Embryo 10 + Embryo 11 | Fibroblast growth factor inducible 14 |
| | | | 1110020J11 | |
| 14 | 9630030D05 | C57BL/6 | N16 | Latent transforming growth factor β -binding protein 1 |
| | E330034H12 | C57BL/6 | P2 ovary | |
| | C730041O04 | C57BL/6 | SV40T antigen | |
| | 9830146M04 | C57BL/6 | Bone | |
| | E230007P04 | C57BL/6 | P2 oviduct | |
| 15 | 1810044G20 | C57BL/6 | Pancreas | Latent transforming growth factor β -binding protein 3 |
| | A230048E14 | C57BL/6 | Hypothalamus | |
| | B430105M07 | C57BL/6 | Adipose | |
| 16 | E130107B17 | C57BL/6 | N0 eyeball | MEGF 12 homolog (<i>Mus musculus</i>) |
| 17 | E130317D06 | C57BL/6 | N0 eyeball | Nuclear factor, interleukin 3, regulated |
| 18 | 1300004O05 | C57BL/6 | Liver | TGF β -inducible early growth response |
| | A730095A08 | C57BL/6 | N7 Cerebellum | |
| 19 | A430010P14 | C57BL/6 | N0 Thymus | Transforming growth factor β 1-induced transcript 1 |
| | 1100001I05 | C57BL/6 | 18-day embryo | |
| | A530093G23 | C57BL/6 | Aorta and vein | |
| | D430029G15 | C57BL/6 | E13 lung | |
| | E030027O16 | C57BL/6 | N0 lung | |
| 20 | 1110002A05 | C57BL/6 | E13 lung | Transforming growth factor β -regulated gene 1 |
| | 2610001N22 | C57BL/6 | Embryo 10 | |
| | 2610029J18 | C57BL/6 | Embryo 10 | |
| 21 | D230045E03 | C57BL/6 | E12 eyeball | Transforming growth factor, β -induced, 68 kD |
| 22 | 1700105F20 | C57BL/6 | Testis | Transforming growth factor, β receptor I |
| | 4930438I19 | C57BL/6 | Testis | |
| | 1700105F20 | C57BL/6 | Testis | |
| 23 | 5730530B02 | C57BL/6 | Embryo 8 | Hypothetical 30.1-kD protein (TGF β inducible nuclear protein TINP1) (hairly cell leukemia protein) (<i>Homo sapiens</i>) |
| | 6720481N02 | C57BL/6 | Extra testis (embryo 12) | |
| | 2310004E21 | C57BL/6 | Tongue | |
| | 2810027P06 | C57BL/6 | Embryo 10 + Embryo 11 | |
| | 3200001B21 | C57BL/6 | Embryo (14 + 17)-Head | |
| | 0610010G24 | C57BL/6 | Kidney | |
| | 5730530B02 | C57BL/6 | Embryo 8 | |

^aMixed library (AL,AK,AD,AF,AG,AC,AE,AQ,AP,AH,AA,AO,AB,AI,AJ,AN,AM).

Table 4. The List of 30 Clones Representing 19 Known Small Inducible Cytokine-Related Genes Found in the FANTOM2 Data Set

| | Clone ID | Mouse strain | cDNA library | Gene name |
|----|------------|-----------------|--------------------------------|---|
| 1 | 1810065I01 | C57BL/6 | Pancreas | Small inducible cytokine A1 |
| 2 | 2310074D07 | C57BL/6 | Tongue | Small inducible cytokine A11 |
| | 2310011L12 | C57BL/6 | Tongue | |
| 3 | 2700043D19 | C57BL/6 | Embryo 11 | Small inducible cytokine A12 |
| 4 | 1600023B02 | C57BL/6 | Placenta | Small inducible cytokine A27 |
| | 1600002M16 | C57BL/6 | Placenta | |
| 5 | 4921506E08 | C57BL/6 | Testis | Small inducible cytokine A28 |
| 6 | 1010001B04 | C57BL/6 | Heart | Small inducible cytokine A5 |
| 7 | 0610030C20 | C57BL/6 | Kidney | Small inducible cytokine A6 |
| | 2010308M07 | C57BL/6 | Small intestine | |
| 8 | 8430433J09 | C57BL/6 | Lung (embryo 16) | Small inducible cytokine A7 |
| 9 | 1810063B20 | C57BL/6 | Pancreas | Small inducible cytokine A8, MCP-2 |
| 10 | 1810031C02 | C57BL/6 | Pancreas | Small inducible cytokine A17 |
| 11 | A430095C13 | C57BL/6 | N0 Thymus | Small inducible cytokine A20 |
| 12 | A930010L22 | C57BL/6 | Retina | Small inducible cytokine B, member 5 |
| 13 | A630052H15 | C57BL/6 | N3 Thymus | Small inducible cytokine B (Cys-X-Cys), member 9 |
| | A630095J17 | C57BL/6 | Thymus | |
| 14 | A430052M02 | C57BL/6 | N0 Thymus | Small inducible cytokine B (Cys-X-Cys), member 11 |
| | A430101A17 | C57BL/6 | N0 Thymus | |
| | C730001K22 | C57BL/6 | SV40T antigen transgenic mouse | |
| 15 | 4631412M08 | C57BL/6 (EST_a) | Skin (Neonate 0 day) | Small inducible cytokine B (Cys-X-Cys), member 13 |
| 16 | 1110012L01 | C57BL/6 (EST_a) | 18-day embryo | Small inducible cytokine B (Cys-X-Cys), member 14 |
| | 1110031L23 | C57BL/6 | 18-day embryo | |
| | 1200006I23 | C57BL/6 | Lung | |
| | 2810410F08 | C57BL/6 | Embryo 10 + Embryo 11 | |
| | 3300001J19 | C57BL/6 | Embryo 17-Head | |
| 17 | 3110001P14 | C57BL/6 | Embryo 13-Head | Small inducible cytokine B, member 15 |
| 18 | A430076J21 | C57BL/6 | N0 Thymus | C, member 1 (lymphotactin) |
| 19 | 6430571D05 | C57BL/6 | Olfactory brain | Small inducible cytokine D, 1 |
| | A430078J23 | C57BL/6 | N0 Thymus | |

transcription through interaction with a TGIF-binding site in DNA (Melhuish et al. 2001).

Little is known about the integrin β -like 1 homolog (with EGF-like repeats), that showed similarity to the clone B930011D01 (Table 5 and Fig. 11, supplemental research data). The EGF-like domain is characteristic for extracellular domain of membrane-bound proteins (InterPro entry IPR000561). This product may possibly function as a cell-adhesion receptor.

The conceptual translation of clone E330012C08 showed similarity (Table 5 and Fig. 12, supplemental research data) to the human latent TGF- β -binding protein 4S, LTBP-4 (O75412). The LTPB-4 is a high molecular weight calcium-binding protein with EGF-like repeats. The cDNA of the LTBP-4 was found in several different forms generated by alternative splicing (Saharinen et al. 1998), consistent with the suspected frame shift positions found in the clone E330012C08. TGF- β proteins are multifunctional growth factors excreted as inactive precursors forming large protein complexes. These complexes include, among others, LTBPs which are believed to regulate both local tissue deposition of TGF- β and the related signaling (Sterner-Kock et al. 2002). The disruption of LTBP-4 in mice leads to ailments such as severe pulmonary emphysema, cardiomyopathy, and colorectal cancer (Sterner-Kock et al. 2002).

Finally, the conceptual translation of clone 9130025L13 showed similarity (Table 5 and Fig. 13, supplemental research data) to the human FGF receptor 4B—FGFR4b (Q9BXY2). FGF receptors are involved in multiple hormonal and proliferative processes of hormonal cells (Yu et al. 2002). A variant of the

FGF receptor 4, related to the FGFR4b was found to increase tumor cell motility (Bange et al. 2002).

DISCUSSION

The mouse is the most popular model organism for both basic and applied immunological studies, including the study of cytokines. Our analysis of the FANTOM2 data set focused on the subset of transcripts that showed similarity to the cytokine-related products, and which had cytokine names or abbreviations in the name or description fields of the corresponding database entries (the cytokine-name criterion). Our study is only partial because databases include cytokine-related entries of genes and proteins that do not have an explicit cytokine name in the name or description fields. However, this study is representative because it focused on a well-defined subset of cytokine-related genes from which we can make an estimate of the coverage of cytokine-related genes in the FANTOM2 data set.

In comparison with database entries that conform to the cytokine-name criterion, we estimate that the ratio of cytokine-related products represented in the FANTOM2 data set is of the order of 20% of the total known cytokine-related products. This estimate is not surprising, because cytokines are inducible and often tissue-specific products. In addition, even when induced, the expression level of cytokines is often very low. The RIKEN libraries, although diverse, may not represent the optimal conditions required for induction of cytokines and their related genes. The regulatory processes involving cytokines are complex processes involving activation and suppression of multiple genes (Doly et al. 1998; Munshi et al.

Table 5. The List of Candidate Novel Cytokine Related Genes Represented in the FANTOM2 Data Set

| Clone ID | DDBJ ^a | cDNA library | Similarity ^b | Similar gene name |
|----------|---|--|---|--|
| 1 | 1110004C05 ^c 1810060G19 1810061A10 | AK003507 18-day embryo Pancreas Pancreas | 89.8, 100.0 (145 AA) 89.1, 100.0 89.8, 100.0 | Interferon-inducible protein (<i>Rattus norvegicus</i>). Identical to 1110004C05RIK protein (Q9CQW9) |
| 2 | 4930507H06 ^c 4933438K12 | AK076846 Testis Testis | 73.1, 75.9 (145 AA) 64.8, 89.0 | Interferon-inducible protein (<i>Rattus norvegicus</i>). Identical to 4933438K12RIK protein (Q9D3R8) |
| 3 | A330075D07 ^c | AK039626 Spinal cord | 52.8, 81.8 (145 AA) | Interferon-induced transmembrane protein 2 (interferon-inducible protein 1-8D) (<i>Homo sapiens</i>) |
| 4 | 1110036C17 ^c | AK004121 18-day embryo | 72.9, 78.1 (145 AA) | Homolog to interferon-inducible protein |
| 5 | 2310061N23 ^c | AK010014 Tongue | 78.0, 97.6 (84 AA) | α -interferon inducible protein (fragment) (<i>Mesocricetus auratus</i>) |
| 6 | A430075K09 ^c 9830003L02 | AK040191 N0 Thymus Bone | 84.9, 100.0 (425 AA) 77.3, 77.3 | Interferon-activatable protein 205 (IFI-205), also known as D3 protein (<i>Mus musculus</i>). Contains PAAD/DAPIN/ Pysin domain (Pfam 02758) and the associated HIN-200/IF120x domain. |
| 7 | 5330409J06 ^c 9530054H01 E030025M22 | AK030414 Pituitary gland U. bladder N0 lung | 69.0, 93.6 65.1, 29.3 69.0, 93.6 (551 AA) | IF-induced GUANYLATE-BINDING PROTEIN 5 (<i>Homo sapiens</i>) |
| 8 | B430113A10 ^c | AK080873 Adipose | 87.6, 100.0 (123 AA) | Hypothetical Toll/Interleukin receptor TIR domain structure containing protein |
| 9 | 4732448K15 ^c 9130227N12 6330415L08 | AK028733 Skin (Neonate 10 day) Cecum Medulla oblongata | 68.8, 100.0 (598 AA) 75.5, 24.2 64.9, 100.0 | Interleukin-1 receptor-associated kinase 2 (<i>Homo sapiens</i>) |
| 10 | E230031K19 ^c | AK054215 P2 oviduct | 76.3, 93.3 (556 AA) | Interleukin 20 receptor α isoform 3 (<i>Homo sapiens</i>) |
| 11 | 9830142A17 C330038O18 C730023J20 D430010O15 ^c A230003P18 D630011E14 E230024J15 E430037H02 | AK084913 Bone ES cells SV40T antigen transgenic E13 Hypothalamus N0 kidney P2 oviduct (NOD)N2 Thymic cells | 75.2, 99.8 (512 AA) 75.0, 99.8 75.4, 99.8 75.0, 99.8 75.2, 99.8 75.0, 99.8 75.2, 99.8 75.1, 99.8 | Transforming growth factor- β -inducible early growth response protein 2 (TGF β -inducible early growth response protein 2) (TIEG-2) (Kreppel-like factor 11) (<i>Homo sapiens</i>) |
| 12 | 4921501K224 ^c 5730599O09 | AK076631 Testis Embryo 8 | 93.7, 100.0 (237 AA) 89.9, 100.0 | DJ977B1.4.1 (isoform 1) (TGIF2) (TGF β -induced factor 2) (TALE family homeobox) homolog (<i>Homo sapiens</i>) |
| 13 | B930011D01 ^c | AK047014 N10 Cerebellum | 88.2, 84.4 (153 AA) | Integrin, β -like 1 (with EGF-like repeat domains) homolog (<i>Mus musculus</i>) |
| 14 | 2310046A13 ^c E330012C08 A730031D03 | AK054297 Tongue P2 ovary N7 Cerebellum | 88.3, 26.7 88.8, 96.5 (1505 AA) 95.4, 32.5 | Latent transforming growth factor- β -binding protein 4S homolog (<i>Homo sapiens</i>) |
| 15 | 9130025L13 ^c 9530005F04 | AK033606 Cecum U. bladder | 77.1, 99.5 (423 AA) 62.8, 99.5 | FGF receptor 4B (<i>Homo sapiens</i>) |

Thirty-one clones representing fourteen candidate new genes.

^aDDBJ accession no. of the representative transcript.

^bSimilarity was defined by two parameters, first the percent identity and second the percent of the length of the reference sequence (similar-gene) having similarity to the FANTOM2 clone. The number in the parentheses in the SIMILARITY field shows the length of the alignment or the maximum length of multiple alignment. All clones were derived from the mouse strain C57BL/6, except for clone E430037H02, which was derived from NOD mouse.

^cRepresentative transcripts.

1999; Vanden Berghe et al. 2000). This illustrates the need for creating more specific cDNA libraries that dynamically capture gene expression profiles for the study of the cellular processes involving cytokines.

It is not always possible, particularly in large-scale studies, to derive clear conclusions and explanations of the data. Our selection of the threshold that distinguishes novel from the known genes or gene candidates is arbitrary. The clone representing integrin β -like homolog (Fig. 11, supplemental research data) has only 10 amino acid differences to its reference sequence. Such clones may, therefore, represent variants or merely the inter-strain differences of the same transcript. Three of the clones representing the 1-8 family of IFN-inducible proteins (Fig. 1) do not appear to be full length and may represent poor-quality sequence data. We should also be aware of the possible problems with the reference data. The Ifi

205 reference sequence (Fig. 4, supplemental research data) has a triple repeat of the sequence AGSTAQ, which is missing in the corresponding FANTOM2 clone. This may possibly represent a sequencing artifact in the reference sequence. Conversely, the FANTOM2 clone having sequence similarity to human latent TGF- β binding protein (Fig. 12, supplemental research data) contains P+R-rich regions that are not present in the reference sequence and may represent artifacts. Mapping FANTOM2 clones to the mouse genome sequence at Ensembl (Hubbard et al. 2002) indicated that some of the FANTOM2 clones represented clear hits, whereas others, (especially Ifi 200 clones) had multiple hits. The multiple hits results may be explained by possible artifacts or poor sequence data, but may also represent the complexity of transcriptional regulation of inducible genes warranting further investigation.

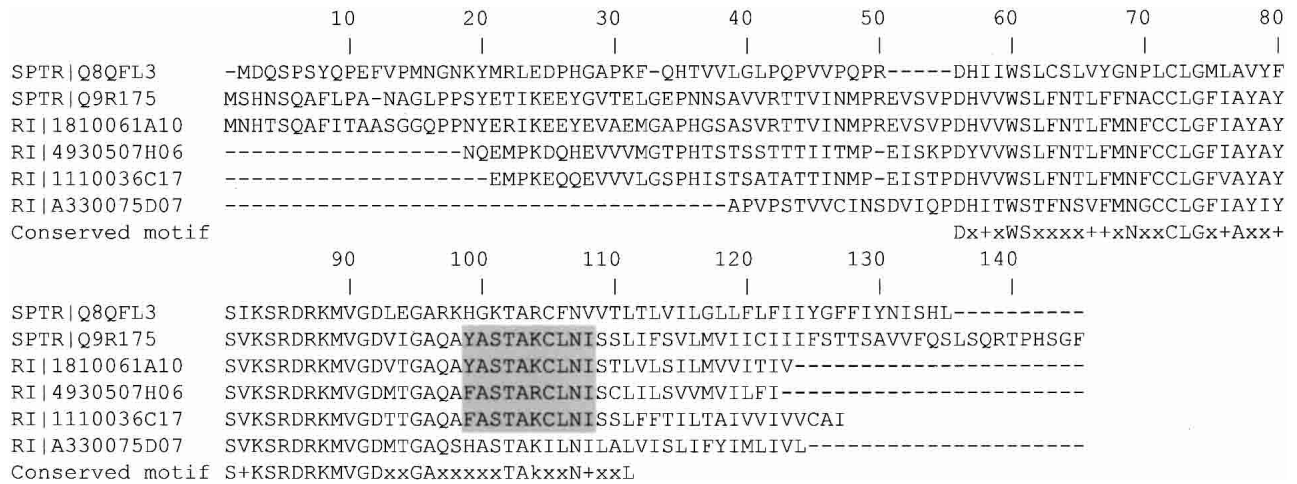


Figure 1 The alignment of conceptual translations of the four candidate novel mouse genes of the family 1-8 of IFN-inducible proteins with representative sequences from rainbow trout (Q8QFL3) and rat (Q9R175). A conserved motif characteristic for the members of interferon-induced 1-8 family of proteins in mice, rats, humans, cattle, and trout is shown with the alignment. Capital letters in the consensus motif stand for identical residues, + for conserved residues (aliphatic—I, V, L, or M; positively charged—K or R; aromatic—F, Y, or W), and x for any residue. The ubiquitin-conjugating enzyme E2 active site motif in 1810061A10 [25] and a similar motif in 4930507H06 are shaded. SPTR stands for SWISS-PROT/TrEMBL database [15], and RI stands for RIKEN clone.

Computational analysis has been applied previously to identification of novel cytokines from EST libraries, such as cardiotrophin-like cytokine (Shi et al. 1999). We have shown that this process can be enhanced by using the elimination strategy that involves multiple consecutive steps as follows: generation of multiple cDNA libraries, large-scale computational annotation, human curation, identification of relevant subsets of sequences, and individual clone analysis. This process involves a number of elimination steps—it is therefore critical that the loss of information is kept at a reasonable level. Too stringent criteria for elimination of clones from further analysis would lead to a small number of clones having a high-quality and high-confidence annotation. On the other hand, this would lead to the elimination of clones that represent results of complex transcriptional events, modifications, or variants. Too lenient criteria for the inclusion of clones for further analysis may lead to the proliferation of artifacts and errors into sequence data. To minimize the proliferation of errors, we excluded a number of clones from the analysis, including those that contained stop codons (data not shown).

We identified 15 candidates for novel cytokine-related genes. Because many cytokine-related genes do not contain cytokine names in their description, this should represent only a fraction of the novel cytokine-related candidate genes captured by the FANTOM2 library. Further strategies will involve making comprehensive lists of cytokine-related genes followed by comparison of these genes with the existing cDNA libraries and future cDNA libraries designed for capturing specific gene expression during the immune processes.

METHODS

The FANTOM2 set of full-length mouse cDNA clones, referred to as clones in this text, contains 60,770 sequences with the total length of almost 120 million base pairs and more than 12 million conceptually translated amino acids. The FANTOM2 clones were functionally annotated using automatic computational annotation followed by expert human curation (Okazaki et al. 2002). The overall process of identification of novel candidate genes is shown in Fig. 3).

Sequence Analysis

The sequences of the cDNA clones were compared with the genetic, protein, motif and protein domain databases. The genetic databases used for the analysis of clones included GenBank (Benson et al. 2002), LocusLink (Wheeler et al. 2002), UniGene (Wheeler et al. 2002), Mouse Genome Database (MGD) (Blake et al. 2002), and TIGR databases (Quackenbush et al. 2001). The comparisons to the genetic databases were performed using the nucleotide to nucleotide sequence local alignment algorithm BLASTN (Altschul et al. 1990). The protein databases used for the analysis of FANTOM2 clones included SWISS-PROT and TrEMBL (Bairoch and Apweiler 2000), and PIR (Wu et al. 2002). The comparisons were performed using the program FASTY (Pearson et al. 1997) that compares a DNA sequence with protein sequences in a database. FASTY translates a DNA sequence in three frames and aligns the translations to each sequence in the protein database, permitting gaps and frame shifts. Finally, the FANTOM2 clones were compared with the SCOP (Lo Conte et al. 2002), Pfam (Bateman et al. 2002), and InterPro (Apweiler et al. 2001) databases for identification of potential protein domains and functional sites, and to the MDS (Maximum Den-

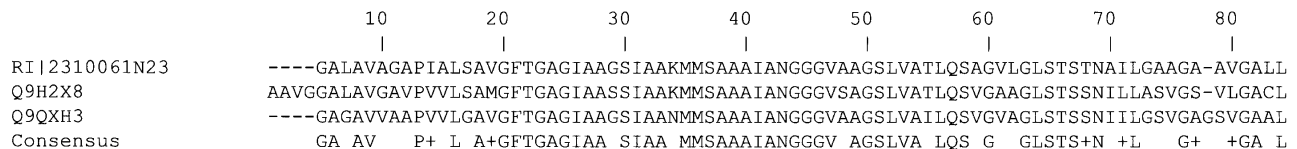


Figure 2 Candidate novel gene represented by clone 2310061N23 similar to the fragments of IFN-α inducible proteins in human (Q9H2X8) and hamster (Q9QXH3). (+) Conserved replacements.

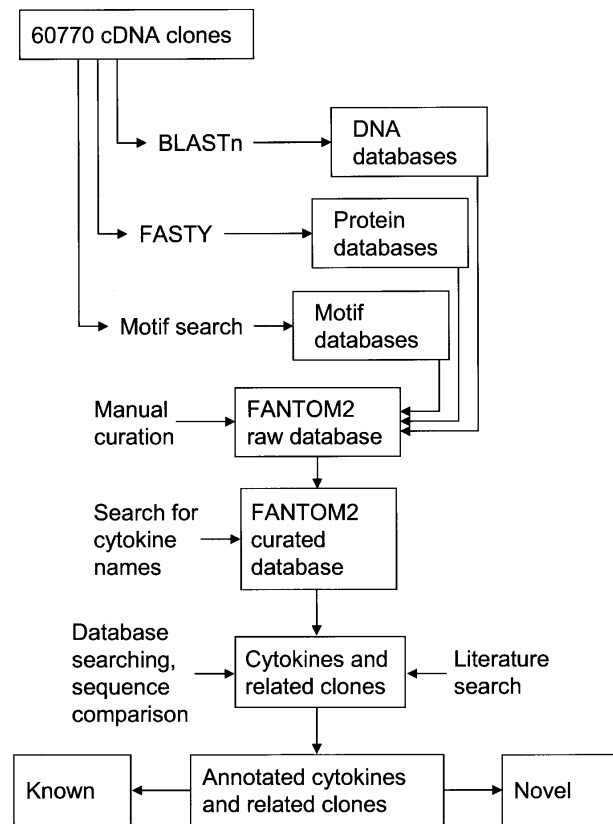


Figure 3 The overall process of identification and functional annotation of cytokine-related FANTOM2 clones.

sity Subgraph) Motif Database (Kawaji et al. 2002) for the presence of potential protein motifs. Unless specified otherwise, the gene products references throughout the text correspond to SWISS-PROT and TrEMBL (SPTR) primary accession numbers.

Functional Annotation and Identification of FANTOM2 Clones

Functional annotation of clones was performed in three steps as follows: data mining, manual annotation, and sequence analysis of the selected clones followed by literature text searching. First, the data-mining step involved a large-scale sequence comparison using BLASTN or FASTY against all of the selected databases. The results of these searches were accessible to the annotators via the FANTOM2 graphical interface, providing a convenient tool for the second step, the manual annotation. The most likely coding sequences (CDS) were determined by comparison of the results of four prediction programs (ProCrest, Decoder, rsCDS, and NCBI cds). Additional tools, such as ClustalW (Thompson et al. 1994) were available for facilitation of the functional annotation. The annotation results were then summarized and searched by keywords cytokine, interferon, IF, IFN, interleukin, IL, transforming growth factor, TGF, epidermal growth factor, EGF, fibroblast growth factor, and FGF. The clones that showed 50%–95% identity to the known cytokine-related genes were considered as candidate novel genes and were further analyzed using the programs BLAST or FASTA. The clone A530093G11 was also included in analysis, because its conceptual translation showed 96% identity over the 50% of the reference sequence length. Finally, literature searches in the

PubMed (Wheeler et al. 2002) were performed for identification of key publications. The possible functions of the candidate novel genes were elucidated from literature sources.

ACKNOWLEDGMENTS

Diego G. Silva is the recipient of a scholarship from the Canberra Hospital Salaried Specialists Private Practice Fund.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.

Anderson, K.V. 2000. Toll signaling pathways in the innate immune response. *Curr. Opin. Immunol.* **12**: 13–19.

Anderson, S.L., Carton, J.M., Lou, J., Xing, L., and Rubin, B.Y. 1999. Interferon-induced guanylate binding protein-1 (GBP-1) mediates an antiviral effect against vesicular stomatitis virus and encephalomyocarditis virus. *Virology* **256**: 8–14

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D., et al. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**: 37–40.

Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**: 45–48.

Bange, J., Prechtel, D., Cheburkin, Y., Specht, K., Harbeck, N., Schmitt, M., Knyazeva, T., Muller, S., Gartner, S., Sures, I., et al. 2002. Cancer progression and tumor cell motility are associated with the FGFR4 Arg (388) allele. *Cancer Res.* **62**: 840–847.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etmiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**: 276–280.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., and Wheeler, D.L. 2002. GenBank. *Nucleic Acids Res.* **30**: 17–20.

Blake, J.A., Richardson, J.E., Bult, C.J., Kadin, J.A., Eppig, J.T., and the Mouse Genome Database Group. 2002. The Mouse Genome Database (MGD): The model organism database for the laboratory mouse. *Nucleic Acids Res.* **30**: 113–115.

Cao, L., Kulmburg, P., Veelken, H., Mackensen, A., Mezes, B., Lindemann, A., Mertelsmann, R., and Rosenthal, F.M. 1998. Cytokine gene transfer in cancer therapy. *Stem Cells* **16**: 251–260.

Cook, T., Gebelein, B., Mesa, K., Mladek, A., and Urrutia, R. 1998. Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor- β -inducible Sp1-like zinc finger-encoding genes involved in the regulation of cell growth. *J. Biol. Chem.* **40**: 25929–25936.

Denovan-Wright, E.M., Ferrier, G.R., Robertson, H.A., and Howlett, S.E. 2000. Increased expression of the gene for α -interferon-inducible protein in cardiomyopathic hamster heart. *Biochem. Biophys. Res. Commun.* **267**: 103–108.

Doly, J., Civas, A., Navarro, S., and Uze, G. 1998. Type I interferons: Expression and signalization. *Cell. Mol. Life Sci.* **54**: 1109–1121.

Dumoutier, L. and Renauld, J.C. 2002. Viral and cellular interleukin-10 (IL-10)-related cytokines: From structures to function. *Eur. Cytokine Netw.* **13**: 5–15.

Gribaudo, G., Ravaglia, S., Guandalini, L., Riera, L., Gariglio, M., and Landolfo, S. 1997. Molecular cloning and expression of an interferon-inducible protein encoded by gene 203 from the gene 200 cluster. *Eur. J. Biochem.* **249**: 258–264.

Gribaudo, G., Riera, L., De Andrea, M., and Landolfo, S. 1999. The antiproliferative activity of the murine interferon-inducible Ifi 200 proteins depends on the presence of two 200 amino acid domains. *FEBS Lett.* **456**: 31–36.

Guenzi, E., Topolt, K., Cornali, E., Lubeseder-Martellato, C., Jorg, A., Matzen, K., Zietz, C., Kremmer, E., Nappi, F., Schwemmler, M., et al. 2001. The helical domain of GBP-1 mediates the inhibition of endothelial cell proliferation by inflammatory cytokines. *EMBO J.* **20**: 5568–5577.

Han, B.H., Park, D.J., Lim, R.W., Im, J.H., and Kim, H.D. 1998. Cloning, expression, and characterization of a novel guanylate-binding protein, GBP3 in murine erythroid progenitor cells. *Biochim. Biophys. Acta* **1384**: 373–386.

Hanks, S.K. and Hunter, T. 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: Kinase (catalytic) domain structure and classification. *FASEB J.* **9**: 576–596.

- Hayzer, D.J., Brinson, E., and Runge, M.S. 1992. A rat β -interferon-induced mRNA: Sequence characterization. *Gene* **117**: 277–278.
- Hofmann, K. and Tschopp, J. 1995. The death domain motif found in Fas (Apo-1) and Tnf receptor is present in proteins involved in apoptosis and axonal guidance. *FEBS Lett.* **371**: 321–323.
- Homey, B., Muller, A., and Zlotnik, A. 2002. Chemokines: Agents for the immunotherapy of cancer? *Nature Rev. Immunol.* **2**: 175–184.
- Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. 2002. The Ensembl genome database project. *Nucleic Acids Res.* **30**: 38–41. (www.ensembl.org).
- Ibelgauffs, H. 1999. Cytokine Online Pathfinder Encyclopaedia. (www.copewithcytokines.de)
- Imoto, I., Pimkhaokham, A., Watanabe, T., Saito-Ohara, F., Soeda, E., and Inazawa, J. 2000. Amplification and overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell lines. *Biochem. Biophys. Res. Commun.* **276**: 264–270.
- Jeschke, M.G., Herndon, D.N., Baer, W., Barrow, R.E., and Jauch, K.W. 2001. Possibilities of non-viral gene transfer to improve cutaneous wound healing. *Curr. Gene Ther.* **1**: 267–278.
- Kawaji, H., Schonbach, C., Matsuo, Y., Kawai, J., Okazaki, Y., Hayashizaki, Y., and Matsuda, H. 2002. Exploration of novel motifs derived from mouse cDNA sequences. *Genome Res.* **12**: 367–378.
- Knutson, K.L. and Disis, M.L. 2001. Expansion of HER2/neu-specific T cells ex vivo following immunization with a HER2/neu peptide-based vaccine. *Clin. Breast Cancer* **2**: 73–79.
- Lewin, A.R., Reid, L.E., McMahon, M., Stark, G.R., and Kerr, I.M. 1991. Molecular analysis of a human interferon-inducible gene family. *Eur. J. Biochem.* **199**: 417–423.
- Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C., and Murzin, A.G. 2002. SCOP database in 2002: Refinements accommodate structural genomics. *Nucleic Acids Res.* **30**: 264–267.
- Melhuish, T.A., Gallo, C.M., and Wotton, D. 2001. TGIF2 interacts with histone deacetylase 1 and represses transcription. *J. Biol. Chem.* **276**: 32109–32114.
- Miyazaki, H. and Kato, T. 1999. Thrombopoietin: Biology and clinical potentials. *Int. J. Hematol.* **70**: 216–225.
- Muller, Y.A., Ultsch, M.H., and de Vos, A.M. 1996. The crystal structure of the extracellular domain of human tissue factor refined to 1.7 Å resolution. *J. Mol. Biol.* **256**: 144–159.
- Munshi, N., Yie, Y., Merika, M., Senger, K., Lomvardas, S., Agalioti, T., and Thanos, D. 1999. The IFN- β enhancer: A paradigm for understanding activation and repression of inducible gene expression. *Cold Spring Harb. Symp. Quant. Biol.* **64**: 149–159.
- Muzio, M., Ni, J., Feng, P., and Dixit, V.M. 1997. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* **278**: 1612–1615.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563–573.
- Pearson, W.R., Wood, T., Zhang, Z., and Miller, W. 1997. Comparison of DNA sequences with protein sequences. *Genomics* **46**: 24–36.
- Prakash, B., Renault, L., Praefcke, G.J., Herrmann, C., and Wittinghofer, A. 2000. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *EMBO J.* **19**: 4555–4564.
- Pru, J.K., Austin, K.J., Haas, A.L., and Hansen, T.R. 2001. Pregnancy and interferon- γ upregulate gene expression of members of the 1-8 family in the bovine uterus. *Biol. Reprod.* **65**: 1471–1480.
- Quackenbush, J., Cho, J., Lee, D., Liang, F., Holt, I., Karamycheva, S., Parvizi, B., Perte, G., Sultana, R., and White, J. 2001. The TIGR Gene Indices: Analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Res.* **29**: 159–164.
- Rasmussen, U.B., Wolf, C., Mattei, M.G., Chenard, M.P., Bellocq, J.P., Chambon, P., Rio, M.C., and Basset, P. 1993. Identification of a new interferon- α -inducible gene (p27) on human chromosome 14q32 and its expression in breast carcinoma. *Cancer Res.* **53**: 4096–4101.
- Rosenberg, S.A. 2001. Progress in the development of immunotherapy for the treatment of patients with cancer. *J. Intern. Med.* **250**: 462–475.
- Saharinen, J., Taipale, J., Monni, O., and Keski-Oja, J. 1998. Identification and characterization of a new latent transforming growth factor- β -binding protein, LTBP-4. *J. Biol. Chem.* **273**: 18459–18469.
- Schwemmle, M. and Staeheli, P. 1994. The interferon-induced 67-kDa guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *J. Biol. Chem.* **269**: 11299–11305.
- Shi, Y., Wang, W., Yourey, P.A., Gohari, S., Zakauskas, D., Zhang, J., Ruben, S., and Alderson, R.F. 1999. Computational EST database analysis identifies a novel member of the neurotrophic cytokine family. *Biochem. Biophys. Res. Commun.* **262**: 132–138.
- Slack, J.L., Schooley, K., Bonner, T.P., Mitcham, J.L., Qwarnstrom, E.E., Sims, J.E., and Dower, S.K. 2000. Identification of two major sites in the type I interleukin-1 receptor cytoplasmic region responsible for coupling to pro-inflammatory signaling pathways. *J. Biol. Chem.* **275**: 4670–4678.
- Smith, B.R. 1990. Regulation of hematopoiesis. *Yale J. Biol. Med.* **63**: 371–380.
- Staudt, L.M. and Brown, P.O. 2000. Genomic views of the immune system. *Annu. Rev. Immunol.* **18**: 829–859.
- Sterner-Kock, A., Thorey, I.S., Koli, K., Wempe, F., Otte, J., Bangsow, T., Kuhlmeier, K., Jin, S., Keski-Oja, J., and Melchner, H. 2002. Disruption of the gene encoding the latent transforming growth factor- β binding protein 4 (LTBP-4) causes abnormal lung development, cardiomyopathy, and colorectal cancer. *Genes & Dev.* **16**: 2264–2273.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Vanden Berghe, W., Vermeulen, L., De Wilde, G., Bosscher, K., Boone, E., and Haegeman, G. 2000. Signal transduction by tumor necrosis factor and gene regulation of the inflammatory cytokine interleukin-6. *Biochem. Pharmacol.* **60**: 1185–1195.
- Vestal, D.J., Gorbacheva, V.Y., and Sen, G.C. 2000. Different subcellular localizations for the related interferon-induced GTPases, MuGBP-1 and MuGBP-2: Implications for different functions? *J. Interferon Cytokine Res.* **20**: 991–1000.
- Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L., et al. 2002. Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res.* **30**: 13–16.
- Wu, C.H., Huang, H., Arminski, L., Castro-Alvear, J., Chen, Y., Hu, Z.Z., Ledley, R.S., Lewis, K.C., Mewes, H.W., Orcutt, B.C., et al. 2002. The Protein Information Resource: An integrated public resource of functional annotation of proteins. *Nucleic Acids Res.* **30**: 35–37.
- Yu, S., Asa, S.L., and Ezzat, S. 2002. Fibroblast growth factor receptor 4 is a target for the zinc-finger transcription factor Ikaros in the pituitary. *Mol. Endocrinol.* **16**: 1069–1078.

Received November 27, 2002; accepted in revised form March 6, 2003.