Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse

Mary E. Brunkow[1], Eric W. Jeffery[1], Kathryn A. Hjerrild[1], Bryan Paeper[1], Lisa B. Clark[1], Sue-Ann Yasayko[1], J. Erby Wilkinson[2], David Galas[3], Steven F. Ziegler[4] & Fred Ramsdell[1]

Scurfy (*sf*) is an X-linked recessive mouse mutant resulting in lethality in hemizygous males 16–25 days after birth, and is characterized by overproliferation of CD4+CD8– T lymphocytes, extensive multiorgan infiltration and elevation of numerous cytokines[1–4]. Similar to animals that lack expression of either Ctla-4 (refs. 5,6) or Tgf-β (refs. 7,8), the pathology observed in *sf* mice seems to result from an inability to properly regulate CD4+CD8– T-cell activity[3,9]. Here we identify the gene defective in *sf* mice by combining high-resolution genetic and physical mapping with large-scale sequence analysis. The protein encoded by this gene (designated *Foxp3*) is a new member of the forkhead/winged-helix family of transcriptional regulators and is highly conserved in humans. In *sf* mice, a frameshift mutation results in a product lacking the forkhead domain. Genetic complementation demonstrates that the protein product of *Foxp3*, scurfin, is essential for normal immune homeostasis.

The *sf* locus was mapped originally to a 1.7-cM interval between *DXWas70* and *Otc* in the proximal region of the mouse X chromosome[1,10]. We used an intersubspecific backcross to further localize *sf* to a 0.3-cM interval (Fig. A, see http://genetics.nature.com/supplementary_info/) and, in parallel, constructed a sequence-ready BAC contig across the entire *DXMit123–Otc* interval. Probe content mapping of 11 overlapping BACs spanning the *DXCch1–DXCch2* region, however, indicated that the 4 clones K50, K90, K60 and K70 defined the minimum tiling path (Fig. 1*a*); this region was estimated to be approximately 500 kb.

By random shotgun sequencing, we identified 20 putative genes on the 4 BACs (Fig. 1*a*), which corresponded well with a recently published map in which transcripts were identified primarily through direct cDNA selection[11]. We discovered through computational analysis an ORF with strong homology to the DNA-binding domain of the forkhead/HNF3/winged-helix family of proteins.

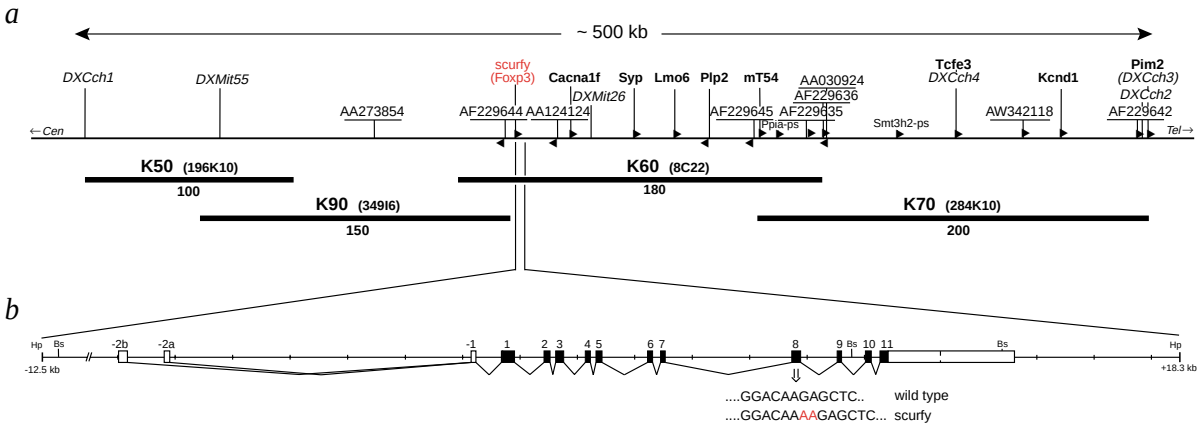

Fig. 1 Physical map of the *sf* region, genomic organization of *Foxp3* and the *sf* mutation. ***a***, The 500-kb *sf* candidate interval is limited by markers *DXCch1* and *DXCch2*. BAC clones K50, K60, K70, K90 and library ID (in parentheses) are indicated. Computational analysis of this region identified 20 putative genes, including nine known genes (in bold): *Tcfe3*, *Kcnd1*, *Pim2* and "human CMV-interacting protein", as well as the mouse orthologs for *SYP*, *LMO6*, *PLP2*, *T54* (ref. 15) and *CACNA1F* (encoding the calcium channel α1 subunit disrupted in congenital stationary night blindness[30,31]; MIM 310500); and two pseudogenes (on lowest level) related to mouse *Smt-3B* and *Ppia*. For the remaining new genes (underlined), corresponding to ESTs and/or GENSCAN predictions, RT–PCR experiments further extended and confirmed transcript structure. The transcriptional orientation of each transcript is indicated by a filled arrowhead. Polymorphic markers are shown in italics. ***b***, Genomic organization of *Foxp3* and the *sf* mutation. Coding exons are shown as filled black boxes, whereas noncoding regions are open boxes. Lines connecting exons indicate splicing events and the observed alternate polyadenylation signal is indicated by the dashed line. Note the two different 5′ noncoding exons (-2a and -2b), separated by 640 bp on the chromosome, both of which splice to a second, common noncoding exon (-1). The putative 5′ end of the most distal non-coding exon, '-2b', located 6.1 kb upstream from the first coding exon corresponds well to a promoter region predicted by GENSCAN. The *sf* mutation is the result of a 2-bp insertion in exon 8 (indicated below the schematic). Exon 8 was amplified and sequenced directly from genomic DNAs derived from the *sf* mutant, as well as four inbred (C57BL/6J, 101/RI, C3Hf/RI, 129/SvJ) and three wild-derived inbred (CAST/Ei, MOLF/Ei, SPRET/Ei) strains of mice. As expected, the 2-bp insertion was specific for *sf* mice. The *Hpa*I sites (Hp) used to generate the 30.8-kb *Foxp3* transgene, as well as the *Bst*XI sites (Bs) used for Southern-blot analysis are shown. The transgene includes the entire gene as well as 12.5 kb and 2.8 kb of 5′ and 3′ flanking sequence, respectively.

[1]*Celltech Chiroscience, Inc., Bothell, Washington, USA.* [2]*Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.* [3]*Keck Graduate Institute for Applied Life Sciences, Claremont, California, USA.* [4]*Virginia Mason Research Center, Seattle, Washington, USA. Correspondence should be addressed to M.E.B. (e-mail: marybrunkow@chiroscience.com).*

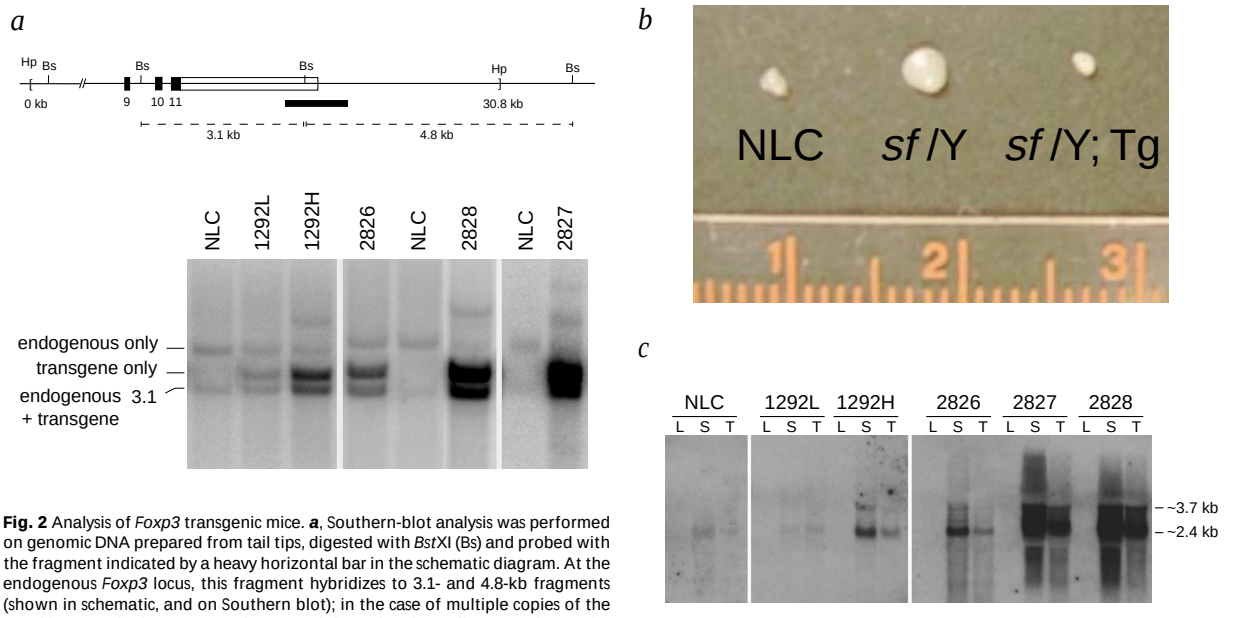

Fig. 2 Analysis of *Foxp3* transgenic mice. ***a***, Southern-blot analysis was performed on genomic DNA prepared from tail tips, digested with *Bst*XI (Bs) and probed with the fragment indicated by a heavy horizontal bar in the schematic diagram. At the endogenous *Foxp3* locus, this fragment hybridizes to 3.1- and 4.8-kb fragments (shown in schematic, and on Southern blot); in the case of multiple copies of the 30.8-kb *Hpa*I (Hp) transgene integrated in a head-to-tail array, the probe hybridizes to the same internal 3.1-kb fragment as well as an unique 3.6-kb fragment (indicated on Southern blot). Representative samples from normal littermate controls (NLC) and transgene (Tg) lines 1292L, 1292H, 2826, 2827 and 2828 are shown. Using the endogenous 4.8-kb *Bst*XI fragment as an internal control in each lane, transgene copy numbers were determined by analysis with a PhosphorImager 445 SI. The copy numbers indicated in Table 1 are the result of analyzing multiple animals from each transgenic line. Additional higher molecular weight bands most apparent in the analyses of lines 1292H, 2827 and 2828 represent the unique single-copy *Bst*XI fragments generated at the sites of transgene insertion. ***b***, Lymph nodes from NLC, *sf/Y* mutant, and *sf/Y*; Tg animals were removed at 12 days of age. Lymphoid organs appeared normal in all *sf/Y*; Tg animals examined, irrespective of transgene copy number. ***c***, *Foxp3* transgene expression level in spleen (S) and thymus (T) roughly correlates with copy number, as determined by northern-blot analysis of total RNA prepared from animals at 10 days of age. A major transcript is seen at ~2.4 kb and a minor species, at ~3.7 kb, the difference in length being due to the use of alternative polyadenylation sites. As with endogenous *Foxp3*, Tg expression is undetectable in liver (L). We loaded 20 μg total RNA in each lane.

This new transcript was not represented in the EST database, nor was it identified in the study mentioned above.

We evaluated candidates by direct sequencing of RT–PCR products obtained from mutant and normal animals. We observed a change in a protein-coding sequence in only the sequence of the new forkhead gene. The sequence derived from *sf* thymus RNA included a 2-bp insertion within the coding region (Fig. 1*b*), which was confirmed by direct sequencing of a PCR product obtained from *sf* genomic DNA. Likewise, the lack of this insertion in PCR products from genomic DNA of seven different mouse strains, as well as in the analogous cDNA from human, indicated that it was in fact unique to the *sf* mutant. The frameshift resulting from the insertion leads to a truncated gene product lacking the carboxy-terminal forkhead domain.

We confirmed the identity of the *sf* gene by functional complementation of the mutation in transgenic mice. A 30.8-kb genomic fragment containing the entire gene was used for microinjection of mouse ooctyes (Fig. 1*b*). We obtained 5 independent transgenic lines, with copy numbers ranging from 3 to approximately 70 (Fig. 2*a*). Transgenic (Tg) males were mated with *sf Otc*^{spf}/++ carrier females, and the scurfy phenotype was analyzed in male progeny. As expected, 50% of the males carried the *Otc*^{spf} (sparse-fur) allele; half of these animals also succumbed to scurfy disease by 16–18 days of age. However, approximately 50% of the *Otc*^{spf} males (also presumed to carry the *sf* mutation) survived to weaning; several of these animals were killed at approximately 4 weeks of age, and visual inspection revealed normal lymph nodes (Fig. 2*b*), spleen and liver. Southern-blot analysis of these animals indicated that they carried the transgene. None of the *Otc*^{spf} (and *sf*) males with scurfy disease were transgenic. We found complete rescue of the scurfy defect with all five transgenic lines (Table 1). Northern-blot analysis revealed elevated expression of the *sf* gene in spleen and thymus that roughly correlated with transgene copy number (Fig. 2*c* and Table 1). *sf Otc*^{spf}/Y; Tg males show no other overt phenotype besides sparse-fur, and animals that are older than one year of age are fertile. Likewise, females carrying the transgene, with or without the *sf* mutation, also seem to be normal. The lymph nodes from transgenic animals, however, were often smaller than those of their littermate controls, primarily due to a decrease in total T-cell number (Fig. 3). This was particularly acute in animals with higher expression of the transgene, but the effect was less pronounced in the spleen. In contrast, the thymus of transgenic animals seemed to be normal in number of cells and phenotype. More detailed analyses of these transgenic mice will be presented elsewhere (manuscript in preparation).

To rule out the possibility that the *sf* mutation results in a dominant-negative protein, we generated transgenic animals that overexpressed the *sf* form of the gene. In analysis of five independent lines, overexpression of the mutant gene did not result in any overt phenotype, nor did it rescue disease in scurfy mice (data not shown).

We generated sequence of full-length *sf* cDNA by assembling overlapping RT–PCR, as well as 5′- and 3′-RACE products from a number of different tissues (embryo, thymus, spleen). Comparison of cDNA with the BAC K60 sequence resulted in a gene structure (Fig. 1*b*). Multiple forms of mature message are produced as a result of alternative noncoding first exons as well as polyadenylation sites. In keeping with the recent proposal to devise a standard nomenclature for the rapidly growing family of forkhead genes[12], the gene mutated in *sf* mice has been designated *Foxp3*, and we refer to the normal gene product as scurfin.

We surveyed a large number of mouse tissues for expression of *Foxp3* using real-time RT–PCR, and the gene *Dad1* as an

Table 1 • Scurfin transgene rescues scurfy disease

Tg line	Tg copy no.	Genotype of males (no. with scurfy disease) wt non-Tg	*sf* non-Tg	wt Tg	*sf* Tg
1292L	3	1 (0)	2 (2)	2 (0)	2 (0)
1292H	9	1 (0)	1 (1)	6 (0)	4 (0)
2826	~16	0	1 (1)	4 (0)	1 (0)
2827	~70	0	2 (2)	0	3 (0)
2828	~45	1 (0)	0	8 (0)	4 (0)
	Totals	3 (0)	6 (6)	20 (0)	14 (0)

For each of the five transgenic lines, the transgene was crossed onto the *sf* mutant background; resulting male progeny were genotyped with respect to *sf* mutation (wt or *sf*) and transgene (non-Tg or Tg) status. The number of animals with scurfy disease within each genotypic class was ascertained (in parentheses). The last column demonstrates that the presence of the transgene, from any of the five lines, prevents disease in genotypically *sf* animals. Tg, transgene; wt, wild type.

endogenous reference[13] (Fig. 4). We found the highest levels of *Foxp3* expression in lymphoid organs such as thymus and spleen, consistent with transgene expression (Fig. 2*c*). Further analysis of subpopulations of purified lymphoid cells showed *Foxp3* expression in Th1 and Th2 cells, and much lower (90% decreased) levels in CD4–CD8+ and B220+ cells.

During the course of analyzing *Foxp3*, a highly conserved human sequence (JM2) was deposited in GenBank. The high degree of similarity with *Foxp3*, combined with the fact that JM2 falls within a chromosomal region with the same gene organization as *Foxp3* (refs. 11,14,15), indicates that JM2 represents the human ortholog. Because JM2 was generated through conceptual translation of genomic sequence from Xp11.23, we isolated human *FOXP3* cDNAs by RT–PCR from a variety of tissues and determined the complete coding sequence. There were a number of differences between our human *FOXP3* sequence and JM2 (Fig. 5). We have identified a single, noncoding exon located 6.1 kb upstream from the first coding exon. The longest human *FOXP3* transcript we have characterized so far includes a 5′ UTR of 188 bp, an ORF of 1,293 bp, and a 3′ UTR of 388 bp. Exon-intron boundaries are identical across the coding regions of the mouse and human genes. The complete mouse and human scurfin sequences are shown (Fig. 5).

We compared the amino acid sequences of mouse and human scurfin with other members of the forkhead/winged-helix/HNF3 family of proteins. Of all these family members, the strongest similarity was to the partial protein sequence called glutamine-rich factor[16], recently designated Foxp1. Based on these alignments, we localized the winged-helix domain of scurfin to the region between amino acid residues 337 and 420, at the C terminus of the protein (Fig. 5). The scurfin forkhead domain contains many of the sequence features thought to be important for mediating protein-DNA contacts based on co-crystallization data using HNF3-γ (ref. 17) and another predicted structure[16] (see also the accompanying report by Wildin *et al.*[18]). Recent phylogenetic analysis of all available forkhead domains from chordates has resulted in the assignment of each into 1 of 17 subfamilies[12] (see also http://www.biology.pomona.edu/fox.html). A sequence alignment including a selection of members from these various subfamilies, along with mouse and human FOXP3, is available (Fig. B, see http://genetics.nature.com/supplementary_info/). As in other family members, the scurfin winged-helix domain contains several basic residues at either end, which may serve to direct the protein to the nucleus[19,20] (Fig. 5).

Analysis of the scurfin sequence using BLAST and Motifs (GCG Wisconsin Package) also revealed the presence of a single C_2H_2 zinc-finger domain 116 amino acids upstream from the winged-helix domain. Aside from these two well-described functional motifs, the scurfin sequence is unique. In particular, we did not find a region with obvious similarity to previously described transcriptional activation domains. The identification of such a domain awaits further functional characterization of the protein.

Foxp3 encodes a new member of the forkhead family of transcriptional regulators and is required for normal T-cell function. The *Foxp3*^{sf} allele results in truncation of the protein and an apparent lack of functional activity. A number of other transcription factors are critical for various aspects of lymphocyte development or lineage commitment[21]. Previous studies have demonstrated a causal association between expression of specific factors and commitment to a specific T-cell lineage, such as Th1 versus Th2 (for example, c-Maf (ref. 22), GATA-3 (ref. 23) and T-bet (ref. 24)), or maintenance of a quiescent state in T cells (for example, LKLF; ref. 25). Our data suggest that the scurfin putative transcription factor is critical for control of immune responses and potentially T-cell number, acting as an apparent rheostat for T-cell activation, and functions predominantly in peripheral T cells.

Both the cellular and biochemical mechanisms by which scurfin controls T-cell responses are yet to be determined. Although similar in phenotype to Ctla-4 and Tgf-β deficient mice, *sf* animals express

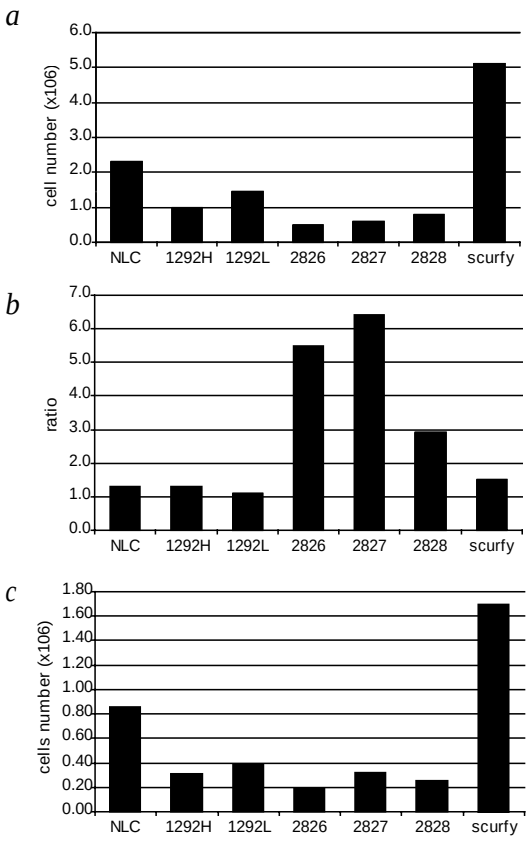

Fig. 3 Diminished T-cell numbers in lymph nodes from transgenic mice. Lymph node cells from representative normal littermate controls (NLC), *sf* mice and the various transgene lines were examined for cell number and phenotype. Total cell number (***a***) from lymph nodes of transgenic mice are consistently less than that in NLC animals, and the extent of reduction is roughly parallel to transgene expression. The ratio of CD4+8– to CD4–8+ cells in the lymph nodes is altered in high-copy transgenic mice, due in part to a decrease in the percentage of CD4–8+ cells (***b***). The total number of CD4+8– T cells in the lymph nodes of transgenic mice is less than that of NLC as well as *sf* animals (***c***). The alterations in cell numbers are consistent in at least 5 mice per transgenic line analyzed between 4 and 16 weeks of age, and the phenotype remains consistent on backcrossing to C57Bl/6 mice for at least six generations.

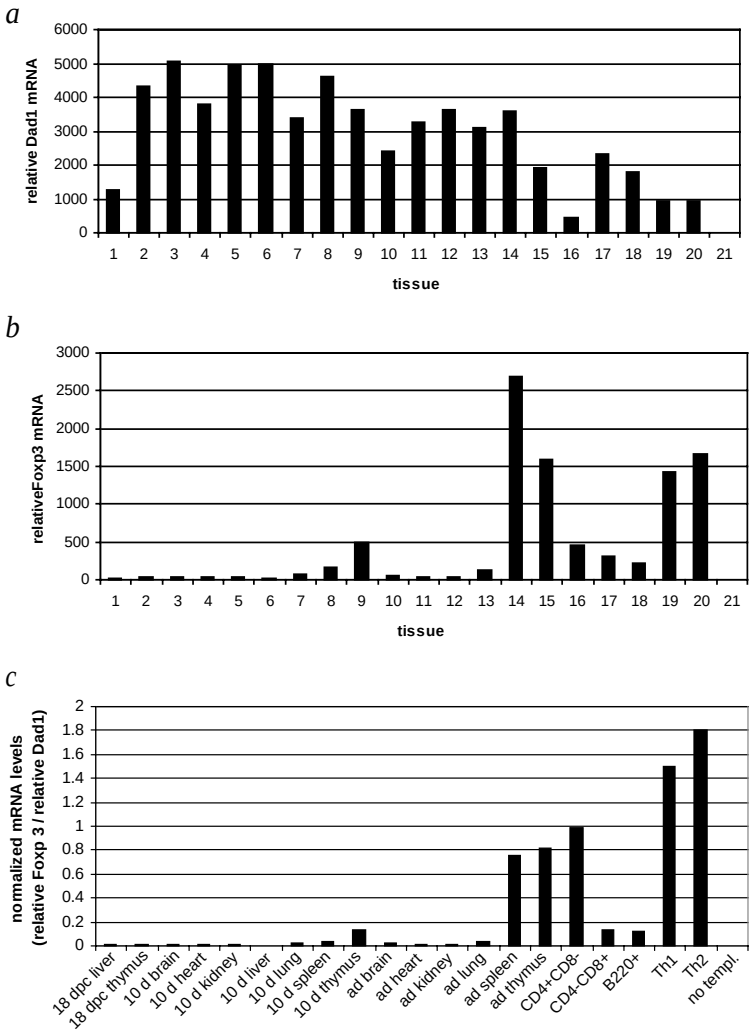

Fig. 4 Tissue distribution of *Foxp3* expression. Levels of *Foxp3* expression were determined using the standard curve method (separate tube reactions), in which *Dad1* served as endogenous reference, and the standard curve was generated from a dilution series of a standard cDNA sample. The standard curve was derived by plotting the threshold cycle (C_T) versus starting quantity, in this case expressed in arbitrary units, and raw values for each of the two genes were then calculated from the C_T of each test sample. The mean values from duplicate reactions for *Dad1* (***a***) and *Foxp3* (***b***) are shown. Normalized *Foxp3* values were then derived from the ratio of raw mean *Foxp3* to raw mean *Dad1* value for each tissue sample (***c***). Mouse tissues analyzed were 18 days post coitum liver and thymus (lanes 1, 2); brain, heart, kidney, liver, lung, spleen and thymus from day 10 animals (lanes 3–9); brain, heart, kidney, lung, spleen and thymus adult tissues (lanes 10–15); CD4+CD8–, CD4–CD8+, B220+, Th1 and Th2 purified lymphocytes (lanes 16–20); and no template control (lane 21).

both of these proteins. It remains to be determined which genes are regulated by scurfin activity, as does the mechanism by which scurfin itself is regulated. Nevertheless, the discovery that mutations in human *FOXP3* lead to a syndrome very similar to scurfy (see Wildin *et al.*[18] and Bennett *et al.*[26], this issue) suggests that the scurfy mouse provides an invaluable model system through which to better understand, and ultimately to develop treatments for, severe autoimmune-related dysfunctions.

Methods

Animals. We conducted animal studies following PHS guidelines. Double-mutant *sf Otc*^{spf} mice were maintained as described[10]. Carriers of the *Otc*^{spf} mutation were identified by amplification of genomic DNA with the primers 5′–TCTGCTGGGAGGACACCC–3′ and 5′–GGCATTATCTAAG GAGAAGCATCA–3′, and subsequent digestion with the restriction endonuclease *Mse*I (ref. 27). The original breeding stocks were obtained from Oak Ridge National Laboratory (ORNL), and *Mus musculus castaneous* animals (CAST/Ei) were obtained from The Jackson Laboratory.

Subcloning and sequencing the candidate region. We sequenced four BAC clones (CITB genomic library, Research Genetics) encompassing the *sf* critical interval by the random shotgun method. All genomic sequences were subjected to BLAST (ref. 28) analysis against all publicly available GenBank databases, and exons were predicted using GENSCAN (ref. 29). All results of the sequence analyses were deposited in a viewer developed in-house (S. Bobick and G. Brickner, pers. comm.).

cDNA cloning and mutation detection. We extracted total RNA from thymus, spleen, brain and liver of 10-d normal and *sf* mice. We used total RNA (5 μg) to generate random-primed first-strand cDNA (SuperScript Preamplification System, Gibco-BRL). For typical RT–PCR reactions, 1 μl cDNA (1/30th of total cDNA preparation) was used as template in a 25 μl reaction; typical amplification conditions on MJ Tetrad were 94 °C for 3 min, followed by 35 cycles of 94 °C for 60 s, 55 °C for 30 s, 72 °C for 60 s. The primers used for these experiments derived from all 18 transcription units located within the *sf* critical interval. In the case of known genes, we used published sequences to design primers spanning the entire cDNA. For new genes, primers were designed initially to span exons predicted by GENSCAN on partial genomic sequence contigs. RT–PCR products were either sequenced directly (using the same primers as were used for the initial amplification, as well as more internal primers as necessary) or cloned into the TA vector (Invitrogen) and sequenced from vector-specific primers. The *sf* mutation was found by sequencing an ~980-bp fragment resulting from amplification of spleen cDNA with the primers 5′–CTACCCACTGCTGGCAAATG–3′ and 5′–GGTTGTGAGGGCTCTTTGAC–3′.

For analysis of the *sf* mutation in genomic DNA, amplicons generated with primers 5′–CAGAGCCTG GTCTATACACTG–3′ and 5′–GAAGGAACTATTGC CATGGCTTC–3′ were sequenced directly using the same primers as in PCR. Cycling conditions were 94 °C for 5 min, followed by 35 cycles of 94 °C for 45 s, 62 °C for 90 s, 72 °C for 90 s.

Transgenic mice. We purified a 30.8-kb *Hpa*I fragment from BAC K60 by treatment with GELase (Epicentre Technologies) after field-inversion gel electrophoresis (FIGE) through a 1% SeaPlaque GTG agarose (FMC BioProducts) gel in 1×TAE. The *Hpa*I fragment was ligated to a linearized, blunt-ended SuperCos I (Stratagene) vector overnight, then transformed into electrocompetent DH10B. To prepare DNA for oocyte microinjection, the cosmid construct (20 μg) was digested with *Not*I and the ~31-kb insert was isolated by FIGE on a 1.1% SeaPlaque GTG agarose gel in 1×TAE. The fragment was purified as above and dialyzed extensively against microinjection buffer (10 mM Tris, 0.25 mM EDTA, pH 7.5). Microinjections of (C57BL/6×SJL)F2 hybrid oocytes were carried out by DNX Transgenic Services.

Lymph nodes from transgenic animals were removed, minced to create single-cell suspensions and analyzed for cell number and phenotype. Phenotypic characterization was performed by staining of cells with fluorescently labeled antibodies to CD4 and CD8 (Caltag) and analysis on a MoFlo flow cytometer (Cytomation).

To generate the 'scurfy' mutant transgene, the single *Eco*RV fragment (~11 kb) from the wild-type cosmid construct was first subcloned into pBluescript SK$^+$ (Stratagene). An ~4-kb *Asp*718 fragment, including exon 8 of *Foxp3*, was then subcloned into pBS SK+ to generate the template for site-directed mutagenesis. Mutagenic primers 5′–GCAGCAAGAGCTC TTTTGTCCATTGAGG–3′ and 5′–CCTCAATGGACAAAAGAGCTCTT GCTGC–3′ were used to introduce the 'scurfy' 2-bp insertion, using the

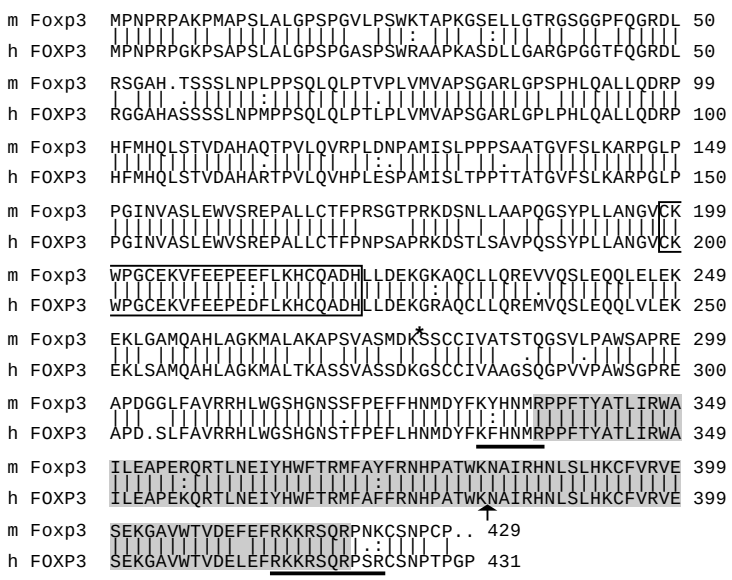

Fig. 5 Mouse and human scurfin proteins contain a highly conserved forkhead domain. Needleman-Wunsch algorithm was used to align the mouse and human sequences; the two proteins have an overall similarity index of 86%. The forkhead domain, with a similarity index of 94%, is shaded, the C_2H_2 zinc finger domain is boxed and potential nuclear localization signals are underlined. Asterisk indicates position of 2-bp insertion in *sf* mutant. One difference between the hypothetical JM2 transcript and mouse *sf* cDNAs was an in-frame insertion of 180 bp in the forkhead domain of the human sequence (between coding exons 10 and 11). The arrow below the human FOXP3 sequence indicates the site of this insertion, which would encode an additional 60 aa.

Quickchange kit (Stratagene). The cycling conditions were 95 °C for 45 s, followed by 16 cycles of 95 °C for 30 s, 55 °C for 60 s, 68 °C for 14 min, and a final elongation step of 72 °C for 5 min. After digestion with *Dpn*I and transformation into *Escherichia coli* strain DH10B, a plasmid containing the appropriate sequence alteration was identified by sequencing. The mutated *Asp*718 fragment was then re-introduced into the cosmid construct by reversing the two original subcloning steps described above. Preparation of DNA for oocyte microinjection was as above.

Full-length mouse and human cDNAs. cDNA libraries prepared from normal 10-d mouse thymus, adult mouse spleen (M.A. Gayle, pers. comm.) and mouse 15-d embryo (5′-STRETCH cDNA library, Clontech) were screened by nested PCR to obtain additional sequences 5′ and 3′ of the 980-bp *Foxp3* fragment described above. In these PCRs, vector-specific primers were paired with *Foxp3*-specific primers. For all three libraries, we used the following λgt10 vector primers: 5′–CGAGCTGCTCTATAGACTGCTGGGTAGTC–3′ nested with 5′–CCTTTTGAGCAAGTTCAGCCTGG–3′; and 5′–TTGCAT ATCGCCTCCATCAACAAAC–3′ nested with 5′–GGTGGCTTATGAGTAT TTCTTCC–3′. *Foxp3*-specific primers used for isolating 5′ end sequences were as follows: 5′–CCAGGCCACTTGCAGA–3′ nested with 5′–CAT TTGCCAGCAGTGGGTAG–3′, and 5′–GCAGCTGGGATGGTGGCAG–3′ nested with 5′–GCCCCACTTCGCAGGTCC–3′. For the 3′ end, a number of different primers were used, such as 5′–CATGGACTACTTCAAGTAC CAC–3′, and 5′–GTCAGAACACGCTCGTGTGCAC–3′, nested with 5′–GTACACACATGCAGCCCCTCC–3′. We amplified the concentrated phage library stock (1 μl) with outer primers (94 °C for 3 min, followed by 35 cycles of 94 °C for 45 s, 62 °C for 90 s, 72 °C for 90 s). This primary reaction (1 μl) was then used as template in a second round of PCR using the appropriate nested primers (same cycling conditions). We also obtained 5′ sequences by applying the RACE technique on normal thymus and spleen RNAs, following the manufacturer's instructions (Gibco-BRL 5′-RACE System v. 2.0). Full-length cDNA sequences of the alternative forms were ascertained by combining all overlapping sequences; RT–PCR using primers from the extreme 5′ and 3′ ends of *Foxp3* cDNAs confirmed the structures as shown.

We originally isolated human *FOXP3* cDNA as two overlapping PCR products from prostate cDNA (Marathon, Clontech). Primers were designed based on the sequence of JM2 and its similarity to the mouse gene: 5′–CACACTGCCCCTAGTCATGG–3′ and 5′–GCATGGCACTCAGCTTC TC–3′ to amplify the 5′ region; 5′–GATGGTACAGTCTCTGGAG–3′ and 5′–GCAAGACAGTGGAAACCTCAC–3′ for the 3′ region. To obtain additional 5′ and 3′ sequences, including the UTRs, cDNA libraries derived from peripheral blood lymphocytes (λgt10 vector, M.A. Gayle, pers. comm.) and testis (λgt11 vector, Clontech) were screened with vector-specific/gene-specific primer pairs. The longest 5′ extension was obtained from the testis library using λgt11 vector-specific primer 5′–CGGTTTCCAT ATGGGGATTGGTGGCGAC–3′ paired with the *FOXP3* anti-sense primer 5′–GGCATCCACCGTTGAGAGC–3′. In addition, the Stratagene Human Universal cDNA Library (HUCL) was screened by hybridization with a 730-bp fragment corresponding to the 5′ half of the gene, obtained by PCR with primers 5′–ATGCCCAACCCCAGGCCTGGC–3′ and 5′–CTCCAGAGACTGTACCATCTC–3′. A single polyadenylated clone was identified. The complete cDNA sequence was ascertained by combining various fragments obtained by all methods. To confirm that the intron included in the hypothetical JM2 cDNA sequence is in-frame with the rest of the coding sequence, we isolated a *FOXP3*-containing BAC from the CITB_978_SKB genomic library (Research Genetics) and directly sequenced across the appropriate region.

Foxp3 expression. Northern-blot analysis was performed on RNA extracted from normal and transgenic tissues using reagents from Ambion (Totally RNA and NorthernMax-Gly kits) and overnight transfer to HyBond N filters (Amersham). A 360-bp fragment of *Foxp3* amplified from thymus cDNA with primers 5′–CAGCTGCCTACAGTGCCCCTAG–3′ and 5′–CAT TTGCCAGCAGTGGGTAG–3′ was labeled with ^{32}P-—dCTP by random hexamer incorporation (MegaPrime kit, Amersham) and used to probe the filters. After washing at high stringency, the filters were exposed to Hyperfilm MP (Amersham Life Science).

We also examined *Foxp3* expression by real-time RT–PCR using an ABI Prism 7700 instrument. Random-primed first-strand cDNA (SuperScript Preamplification System, Gibco-BRL) was used as template in separate-tube amplification reactions. All samples were run in duplicate in each experiment. Primers were custom designed and obtained from PE Biosystems. *Foxp3* primers were 5′–GGCCCTTCTCCAGGACAGA–3′ and 5′–GCTGATCATG GCTGGGTTGT–3′, and the internal TaqMan probe was 5′–6FAM-AGCTTC ATCCTAGCGGTTTGCCTGAGAATAC-TAMRA–3′; *Dad1* primers were 5′–CCTCTCTGGCTTCATCTCTTGTGT–3′ and 5′–CCGGAGAGATGC CTTGGAA–3′, with TaqMan probe 5′–6FAM-AGCTTCATCCTAGCG GTTTGCCTGAGAATAC-TAMRA–3′. The TaqMan Universal Master Mix (PE Applied Biosystems) was used for all the reaction components, except primers, probe and template. The final primer concentrations were 300 nM for *Foxp3* and 50 nM for *Dad1*. The final probe concentration for both *Foxp3* and *Dad1* was 100 nM. Cycling conditions were 50 °C for 2 min; 95 °C for 10 min; and 40 cycles of 95 °C for 15 s, 60 °C for 1 min. The data was collected and analyzed by the ABI Prism 7700 Sequence Detection System Software, Version 1.6.4. The relative quantity was determined by the standard curve method, in which a relative standard curve of known dilutions (1×, 1:10, 1:100, 1:1,000, 1:10,000) was run with the unknown samples. The relative quantity of each unknown was determined by plotting a standard curve (threshold cycle (C_T) versus starting quantity), and calculating from the C_T of each sample the amount amplified. The normalized value was determined by dividing the relative quantity of *Foxp3* for each sample by the relative quantity of *Dad1* for that sample.

GenBank accession numbers. Mouse *Foxp3* cDNAs (with alternative noncoding 5′ exons), AF277991 and AF277992; human *FOXP3* cDNA, AF277993; human JM2 sequence, AJ005891; human genomic sequence from Xp11.23 including JM2, AF235097 and AF196779; 30.8-kb *Hpa*I genomic fragment containing mouse *Foxp3*, AF277994; *DXCch1*,

AF277995; *DXCch2*, AF277996; *DXCch3*, AF318279; *DXCch4*, AF318280; *DXCch5*, AF318281; *PLP2*, U93305; human CMV-interacting protein, X97571; Foxp1, GenPept, A49395.

Acknowledgments

We thank L. Russell for support; V. Godfrey, P. Blair and S. Witonsky for help in the initial stages of the mapping project; J. Mulligan, M. Appleby and R. Khattri for discussions; the CCH sequencing group for their diligence and efficiency; and S. Proll, M. Mortrud, D. Walker and S. Corpening for technical assistance.

Received 22 August; accepted 3 December 2000.

1. Lyon, M., Peters, J., Glenister, P., Ball, S. & Wright, E. The scurfy mouse mutant has previously unrecognized hematological abnormalities and resembles Wiskott-Aldrich syndrome. *Proc. Natl. Acad. Sci. USA* **87**, 2433–2437 (1990).
2. Kanangat, S. *et al.* Disease in the scurfy (*sf*) mouse is associated with overexpression of cytokine genes. *Eur. J. Immunol.* **26**, 161–165 (1996).
3. Blair, P. *et al.* CD4+CD8– T cells are the effector cells in disease pathogenesis in the scurfy (sf) mouse. *J. Immunol.* **153**, 3764–3774 (1994).
4. Clark, L. *et al.* Cellular and molecular characterization of the scurfy mouse mutant. *J. Immunol.* **162**, 2546–2554 (1999).
5. Tivol, E. *et al.* Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* **3**, 541–547 (1995).
6. Waterhouse, P. *et al.* Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* **270**, 985–988 (1995).
7. Shull, M. *et al.* Targeted disruption of the mouse transforming growth factor-β1 gene results in multifocal inflammatory disease. *Nature* **359**, 693–699 (1992).
8. Kulkarni, A. *et al.* Transforming growth factor β 1 null mutation in mice causes excessive inflammatory response and early death. *Proc. Natl. Acad. Sci. USA* **90**, 770–774 (1993).
9. Godfrey, V., Rouse, B. & Wilkinson, J. Transplantation of T cell-mediated, lymphoreticular disease from the scurfy (*sf*) mouse. *Am. J. Pathol.* **145**, 281–286 (1994).
10. Blair, P. *et al.* The mouse scurfy (*sf*) mutation is tightly linked to *Gata1* and *Tfe3* on the proximal X chromosome. *Mamm. Genome* **5**, 652–654 (1994).
11. Means, G., Toy, D., Baum, P. & Derry, J. A transcript map of a 2-Mb BAC contig in the proximal portion of the mouse X chromosome and regional mapping of the scurfy mutation. *Genomics* **65**, 213–223 (2000).
12. Kaestner, K., Knochel, W. & Martinez, D. Unified nomenclature for the winged-helix/forkhead transcription factors. *Genes Dev.* **14**, 142–146 (2000).
13. Hong, N. *et al.* A targeted mutation at the T-cell receptor α/δ locus impairs T-cell development and reveals the presence of the nearby antiapoptosis gene Dad1. *Mol. Cell. Biol.* **17**, 2151–2157 (1997).
14. Boyd, Y. *et al.* Mouse X chromosome. *Mamm. Genome* **7**, S313–326 (1997).
15. Schindelhauer, D. *et al.* Long-range map of a 3.5-Mb region in Xp11.23–22 with a sequence-ready map from a 1.1-Mb gene-rich interval. *Genome Res.* **6**, 1056–1069 (1996).
16. Li, C. & Tucker, P. DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/forkhead domain. *Proc. Natl. Acad. Sci. USA* **90**, 11583–11587 (1993).
17. Clark, K., Halay, E., Lai, E. & Burley, S. Co-crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
18. Wildin, R.S. *et al.* X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nature Genet.* **27**, 18–20 (2001).
19. Qian, X. & Costa, R. Analysis of hepatocyte nuclear factor-3b protein domains required for transcriptional activation and nuclear targeting. *Nucleic Acids Res.* **23**, 1184–1191 (1995).
20. LaCasse, E. & Lefebvre, Y. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res.* **23**, 1647–1656 (1995).
21. Kuo, C. & Leiden, J. Transcriptional regulation of T lymphocyte development and function. *Annu. Rev. Immunol.* **17**, 149–187 (1999).
22. Ho, I., Hodge, M., Rooney, J. & Glimcher, L. The proto-oncogene c-*maf* is responsible for tissue-specific expression of interleukin-4. *Cell* **85**, 973–983 (1996).
23. Zheng, W. & Flavell, R. The transcription factor GATA-3 is necessary and sufficient for Th2 cytokine gene expression. *Cell* **89**, 587–596 (1997).
24. Szabo, S. *et al.* A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655–669 (2000).
25. Kuo, C., Veselits, M. & Leiden, J. LKLF: A transcriptional regulator of single-positive T cell quiescence and survival. *Science* **277**, 1986–1990 (1997).
26. Bennett, C.L. *et al*. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome is caused by mutations of *FOXP3*. *Nature Genet*. **27**, 20–21 (2001).
27. Veres, G., Gibbs, R., Scherer, S. & Caskey, C. The molecular basis of the sparse fur mouse mutation. *Science* **237**, 415–417 (1987).
28. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
29. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
30. Bech-Hansen, T. *et al.* Loss-of-function mutations in a calcium-channel α1-subunit gene in Xp11.23 cause incomplete X-linked congenital stationary night blindness. *Nature Genet.* **19**, 264–267 (1998).
31. Strom, N.T. *et al.* An L-type calcium-channel gene mutated in incomplete X-linked congenital stationary night blindness. *Nature Genet.* **19**, 260–263 (1998).