Genomic Organization and Sequence of the Human NRAMP Gene: Identification and Mapping of a Promoter Region Polymorphism

Jenefer M. Blackwell, C. Howard Barton, Jacqueline K. White, Susan Searle, Anne-Marie Baker, Hazel Williams, and Marie-Anne Shaw
University of Cambridge Clinical School, Department of Medicine, Addenbrooke's Hospital, Cambridge, United Kingdom

ABSTRACT

Background: Murine *Nramp* is a candidate for the macrophage resistance gene *Ity/Lsh/Bcg*. Sequence analysis of human NRAMP was undertaken to determine its role in man.
Materials and Methods: A yeast artificial chromosome carrying NRAMP was subcloned and positive clones sequenced. The transcriptional start site was mapped using 5′ RACE PCR. Polymorphic variants were amplified by PCR. Linkage analysis was used to map NRAMP.
Results: NRAMP spans 12kb and has 15 exons encoding a 550 amino acid protein showing 85% identity (92% similarity) with *Nramp*. Two conserved PKC sites occur in exon 2 encoding the Pro/Ser rich SH3 binding domain, and in exon 3. Striking sequence similarities (57 and 53%) were observed with yeast mitochondrial proteins, SMF1 and SMF2, especially within putative functional domains: exon 6 encoding the second transmembrane spanning domain, site of the murine susceptibility mutation; and exon 11 encoding a conserved transport motif. No mutations comparable to the murine susceptibility mutation were found. The transcriptional initiation site mapped 148 bp 5′ of the translational initiation codon. 440bp of 5′ flanking sequence contained putative promoter region elements: 6 interferon-γ response elements, 3 W-elements, 3 NFκB binding sites and 1 AP-1 site. Nine purine-rich GGAA core motifs for the myeloid-specific PU.1 transcription factor were identified, two combining with imperfect AP1-like sites to create PEA3 motifs. TATA, GC and CCAAT boxes were absent. A possible enhancer element containing the Z-DNA forming dinucleotide repeat t(gt),ac(gt),ac(gt),g was polymorphic (4 alleles; n=4,9,10,11), and was used to map NRAMP to 2q35.
Conclusions: This analysis provides important resources to study the role of NRAMP in human disease.

INTRODUCTION

The gene encoding the natural resistance-associated macrophage protein, *Nramp*, was identified as a candidate for the murine macrophage resistance gene *Ity/Lsh/Bcg* based (a) on its macrophage-restricted expression and (b) on the presence of a common mutation in all susceptible mouse strains (1,2). The deduced amino acid sequence encodes a polytopic integral membrane protein, with structural features common to prokaryotic and eukaryotic transporters, and a 20 amino acid consensus sequence showing identity with a conserved binding protein-dependent transport motif of a non-ATP binding class of membrane transporter molecules. The original murine pre-B cell-derived cDNA (1) was recently shown to lack sequence from exons 1 and 2 (3). Exon 2 may be of particular importance in regulating *Nramp* function since it encodes a proline/serine rich domain typical of SH3 binding domains found in signaling/cytoskeletal molecules, and used in assembly of the phagocyte

Address correspondence and reprint requests to: J. M. Blackwell, Department of Medicine, Level 5, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK.

Molecular Medicine, Volume 1, Number 2, January 1995 194-205

NADPH oxidase complex (4). Exons 1, 2, and 3 also introduce three additional protein kinase C (PKC) phosphorylation sites (3).

In mice, *Ity/Lsh/Bcg* regulates the activation of macrophages for nitric oxide (NO)-mediated antimicrobial activity against intracellular pathogens, and exerts a range of pleiotropic effects in vitro (reviewed in 5–10) including regulation of the following: KC, IL-1β and inducible NO synthase (iNOS) mRNA; surface MHC class II, 5′ nucleotidase and AcM.1 antigen expression; and TNFα release, oxidative burst, and tumouricidal activity. In vivo, the gene has a dramatic effect (reviewed in Ref. 5) on early T cell-independent regulation of *Salmonella typhimurium, Leishmania donovani,* and mycobacterial infections (*Mycobacterium bovis; M. lepraemurium;* and *M. intracellulare*), as well as on the later development of an interferon-γ generating CD4-positive T cell response (11). These in vivo effects presumably reflect synergy between the many pleiotropic effects of the gene on macrophage function. Hence, although human macrophages do not appear to use iNOS-generated NO for antimicrobial activity (12), a human homolog (NRAMP) for murine *Nramp* might nevertheless play a role in regulating macrophage priming/activation and hence be important in any disease involving defective macrophage function. To facilitate the search for human disease associations with NRAMP, this paper presents an analysis of the sequence and genomic organization of the human NRAMP gene and includes identification of a promoter region polymorphism which might be important in regulating NRAMP expression.

MATERIALS AND METHODS

Genomic Sequencing of NRAMP

A human yeast artificial chromosome (YAC) AM11/D3/14, obtained by screening the ICRF (13) library with a VIL1 probe (14) and containing the entire human NRAMP sequence (15), was sublconed into λEMBL3 (Stratagene Ltd., Cambridge, U.K.) and screened with the full-length murine *Nramp* cDNA λ8.1 (3). Two overlapping clones, λ3 and λB.1, containing the full-length NRAMP sequence, were digested with *Pst*I, sublconed into pBluescript II SK (Stratagene Ltd.), and re-screened with the full-length murine cDNA probe (3). Exon-positive clones were selected for sequence analysis, with gaps being filled by sequencing fragments prepared by PCR between identified exons. Exons were identified by comparison of human genomic sequence with mouse (1,3) or human cDNA sequences. Human cDNA sequence was obtained by reverse transcription (RT) and PCR amplification of RNA prepared from the human monocyte-derived THP1 cell line (16). Where appropriate, PCR products were cloned into the pCR vector (Invitrogen Corp., Abingdon, U.K.) for sequence analysis from at least two independent clones. Clones corresponding to the 3′ region were not originally isolated by screening with the murine cDNA. A fragment was generated by 3′ rapid amplification of cDNA ends (RACE) (17) from polydT adaptor primed THP1 cDNA. cDNA was amplified using the adaptor primer in combination with two nested primers selected from exon 13 (GTGCTGCCCATCCTCACG; GAGTTTGCCAA TGGCCTG). A suitable genomic clone was prepared by amplification of a fragment from both λ3 and the YAC AM11/D3/14 using exon 13 primers and a primer (GGACGAGAAGGGAACT AG) designed from the 3′ end of the RACE product. The 5′ end of the RNA was mapped by 5′ RACE involving RNA ligase-dependent ligation of a blocked anchor primer to the 3′ end of random hexamer primed reverse transcribed THP1 RNA. Amplification using an anchor primer and two NRAMP-specific nested antisense primers (AAGAAGGTGTCCACAATGGTG, CGGTTTTGTGTCTGGGAT) yielded a single NRAMP product. The product was TA cloned, and three clones were subjected to sequence analysis to determine the transcriptional initiation site and sequence of the most proximal exon that failed to hybridize to any mouse cDNA probe. This facilitated further analysis of the 5′ flanking region, the sequence for which was obtained from a 1.6 kb *Pst*I fragment that contained sequence homologous to the 5′ RACE product.

Analysis of Sequence Data

Nucleotide and amino acid sequence comparisons were made using the BESTFIT program online to the CRC Resource Center, U.K. Amino acid sequences for murine and human NRAMP were aligned with yeast SMF1 and SMF2 (18) using the multiple sequence alignment program Clustal V (19).

Direct Cycle Sequencing Across Exons 4–6 of Human NRAMP

Primers (GACAGGCAAGGACTTGGGT and AAG AAGGTGTCCACAATGGTG) were designed for RT/PCR amplification of a 200 bp product between exons 4 and 6 of human NRAMP, using RNA purified from peripheral blood mononuclear cells. This product spans the region of murine *Nramp* which carries the susceptibility mutation. PCR products were purified with a Qiagen PCR purification kit (Hybaid Ltd., Teddington, U.K.), and subjected to direct cycle sequence analysis using the Circum Vent Thermal Cycle Dideoxy DNA Sequencing Kit (New England Biolabs, CP Laboratories, Bishop's Stortford, U.K.) with an internal sequencing primer (CATCTCTACTACCCCAAGGTGC). Direct cycle sequence analysis was performed on 19 individuals: 8 visceral leishmaniasis patients, 9 unaffected individuals taken from the same families, and 2 nonendemic British controls. Endemic samples were from Brazil (4 affected; 5 unaffected) and the Sudan (4 affected; 4 unaffected).

Primer Design and PCR Analysis of a 5′ gt Repeat Using Human Genomic DNAs

PCR products of 780–794 bp were amplified from genomic DNA using primers located −365 bp 5′ of the transcription start site (GAGGGGTC TTG GAACTCCA) and within intron 1 (CACCTT CTCCGGCAGCCC). This product was reamplified to generate 108–122 bp products using the 5′ primer and an end-labelled (γ^{32}PdATP; ICN Biomedicals Ltd., Thame, U.K.) internal reverse primer TACCCCATGACCACACCC. The products were resolved by denaturing polyacrylamide gel electrophoresis and sized using a sequencing ladder. PCR products corresponding to different allelic forms were directly sequenced as described above.

Family Linkage Studies

A set of 36 multicase families of leprosy, tuberculosis, and visceral leishmaniasis from our study site in Brazil (20) were used to determine linkage between a polymorphic gt repeat in the 5′ promoter region of human NRAMP and previously mapped 2q34–q35 markers (15,20). Two-point linkage analyses were carried out between NRAMP and the markers (TNP1, IL8RB, VIL1, DES) using LINKAGE (21) on-line to the CRC Resource Center. Gene frequencies for the NRAMP alleles were calculated from a sample of 72 genetically independent individuals from the Brazilian study site.

RESULTS

Sequence and Genomic Organization of Human NRAMP

The sequencing of exon-positive clones isolated by hybridization with a full-length cDNA allowed for the identification of the complete sequence (EMBL accession numbers x82015 and x82016) of the human 2q homologue (NRAMP) of the murine chromsome 1-derived *Nramp* gene. Analysis of exon sequence from a region 440 bp 5′ of the transcriptional initiation site to the termination codon allowed for the complete exon-intron organization to be elucidated (Table 1). Human NRAMP is encoded by 15 exons and, in constrast to the 548 amino acid murine macrophage isoform (3), contains 550 amino acids (Fig. 1). This 550 amino acid polypeptide is initiated from a translational codon within exon 1 in the context of a weak (1/6) Kozak (22)-consensus. The next, more distal codon found at M68 has a 2/6 Kozak consensus. However, we propose that like the murine macrophage form (3), the more proximal initiation codon will be utilized. This is reinforced by the striking (100%) sequence conservation for residues 51–67 (Fig. 1), indicating a requirement for the maintenance of sequence for function. The discrepancy in size between murine (548) and human (550) genes results from the inclusion of three additional residues within exon 2 causing a PTS duplication, with the nonduplicated form representing a rare variant in Brazilian (15) and British (unpublished data) pedigrees. In addition, the human gene exhibits a single amino acid deletion relative to the mouse within the poorly conserved last exon. Overall amino acid identity with murine *Nramp* was 86% (92% with conserved substitutions). Exons exhibiting highest sequence identity (100%) include exons 4, 6, and 7, with exon 11 displaying 98% identity. These exons encode TM1, the first extracellular domain, TM2 and TM3, and the conserved transport motif. It is of interest that TM2, containing the murine susceptibility-associated mutation (1,2) is well conserved, suggesting that this domain plays an important functional role which cannot tolerate amino acid substitutions. NRAMP was aligned with murine *Nramp* and with the two yeast mi-

TABLE 1. Intron (four flanking nucleotides)/exon (amino acids) boundaries and sizes (bp) for the 15 exons of human NRAMP identified by genomic sequence analysis of YAC-derived clones.

Exon Number	Size (bp)	Intron/Exon Boundaries	%AA Identity (Mouse)
		Met Thr G	50
EXON 1	155	ATG . . 145bp . . ATG ACA G gtga	
		ly Asp Lys . . (43aa) . . . Lys Pro	68
EXON 2	143	acag GT GAC AAG . . . . . AAA CCG gtgg	
		Gly Thr . . (37aa) . . . Phe Lys	95
EXON 3	123	acag GGC ACC . . . . . TTC AAA gtaa	
		Leu Leu . . (36aa) . . . Pro Lys	100
EXON 4	120	acag CTT CTC . . . . . CCt AAG gtgg	
		Val Pro . . . (31aa) . . Ala Gly Ar	91
EXON 5	107	tcag GTG CCC . . . . . GCT GGA CG gtac	
		g Ile Pro . . (19aa) . . . Asn Tyr G	100
EXON 6	71	tcag A ATC CCA . . . . . AAC TAC G gtgg	
		ly Leu Arg . . (18aa) . . . Tyr Gln	100
EXON7	68	gtag GG CTG CGG . . . . . TAT GAG gtag	
		Tyr Val . . (48aa) . . . Val Lys	88
EXON 8	156	gcag TAT GTG . . . . . GRC AAG gtag	
		Ser Arg . . (49aa) . . . Ala Ala	87
EXON 9	159	gtag TCT CGA . . . . . GCT CGC gtga	
		Phe Asn . . (26aa) . . . Gln Gly	80
EXON 10	90	gcag TTC AAC . . . . . CAG GGG gtga	
		Gly Val . . (36aa) . . . Met Glu	98
EXON 11	120	gcag GGC GTG . . . . . ATG GAG gtag	
		Gly Phe . . (46aa) . . . Leu Leu	94
EXON 12	150	ccag GGC TTC . . . . . CTG CTG gtga	
		Leu Pro . . (20aa) . . . Asn Gly Le	84
EXON 13	74	ccag CTC CCG . . . . . AAT CCG CT gtga	
		u Leu Asn . . (47aa) . . . Tyr Leu	73
EXON 14	154	ccag G CTG AAC . . . . . TAC CTG gtac	
		Val Trp . . (34aa) . . . Ter	67
EXON 15	108	ccag GTC TGG . . . . . TAG	

Amino acid sequence identity with murine *Nramp* is shown for each exon.

tochondrial membrane proteins, SMF1 and SMF2, using the multiple sequence alignment program Clustal V (Fig. 1). SMF1 and SMF2, which show 49% identity (70% similarity) with each other, show 30% (57%) and 29% (53%) identities (similarities), respectively, with human NRAMP. This parallels the 30% (58%) and 30% (53%) identities (similarities) we reported (8) for murine *Nramp*. Regions of most striking sequence identity between all four proteins were found predominantly within the hydrophobic regions, although high identities were also found in exons 3, 4, 5, and 6, and for the conserved transport motif from exon 11. Within exon 6, the YAC-derived amino-acid human sequence exhibited a Gly at residue 172, corresponding to the position of the Gly→Asp susceptibility mutation at codon 169 of the murine sequence. Although the two SMF genes do not encode a similar Gly, they encode residues that do not introduce neg-

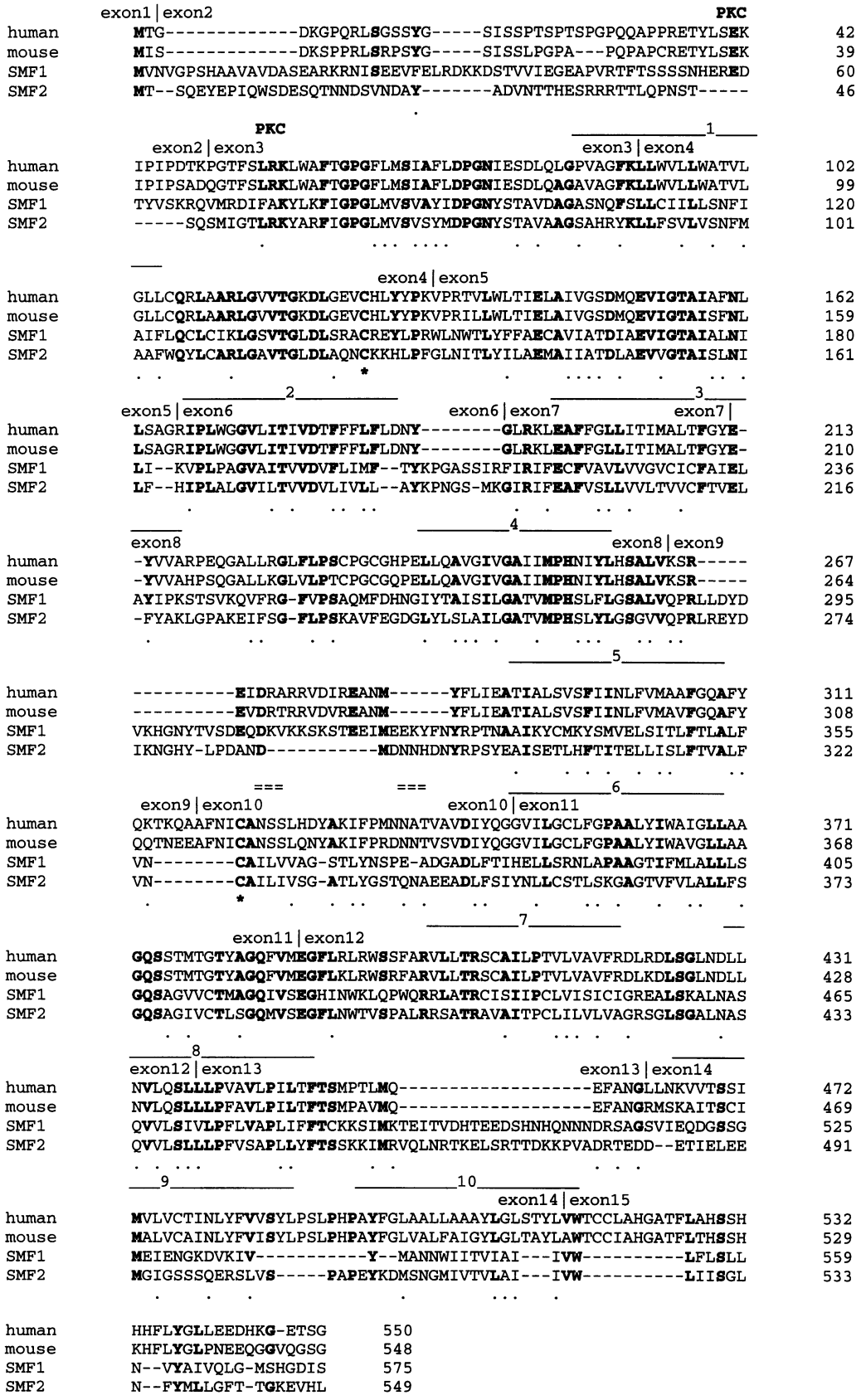

FIG. 1. Clustal V multiple sequence alignment for the deduced amino acid sequence for human NRAMP, murine *Nramp* clone λ 8.1 (3), and the yeast mitochondrial proteins SMF1 and SMF2 (18).

Residues showing 3/4 or 4/4 identities across the four proteins are shown in bold. For the NRAMP sequence: exon boundaries are indicated above the sequence; (PKC) consensus sites (S/T-X-R/K) for protein kinase C phosphorylation; (===) consensus sites for N-linked glycosylation; and putative membrane spanning domains (after Ref. 1) are overlined and numbered on the sequence. (*) cysteine residues conserved across all four proteins; (·) conserved substitutions.

atively charged residues found in the susceptible allele of mice. As before (3,15), matches with other proteins (Fig. 2) in the sequence databases were observed over exon 2 which contains a putative SH3 binding domain; and over the region of exon 11 containing the conserved binding protein-dependent transport motif (1). The latter was highly conserved (7/20 identity; 11/20 similarity) in murine/human NRAMP, the yeast proteins, and in two expressed sequence tags from *Oryza sativa* (rice) and *Arabidopsis thaliana*. SMF1 and SMF2 do not demonstrate high identity over the proline/serine rich sequence of exon 2 but do have consensus (S/T-X-R/K) sequences (one in SMF1; two in SMF2) for PKC-dependent phosphorylation. Human NRAMP has two PKC consensus sites (in exons 2 and 3, Fig. 1) in this region, compared with three in the murine gene. The location of the distal site in SMF2 matches precisely with human NRAMP site 2/murine *Nramp* site 3, whereas the site in SMF1 is located eight residues upstream. A pair of cysteine residues are conserved in all four genes: (1) in the first extracellular loop domain; and (2) in the third extracellular domain which also contains two sites for N-linked glycosylation in the human and murine genes. Charged residues are conserved across all four proteins within the transmembrane spanning domains 1, 2, 3, 4, and 7 (Fig. 1), except for a Lys→Ser substitution in the first transmembrane domain of SMF1.

Analysis of the Murine Mutation Site in Visceral Leishmaniasis Patients and Controls

To determine whether a mutation homologous to the murine disease susceptibility Gly→Asp mutation occurs in man, RT/PCR and direct cycle sequencing was performed on RNA from visceral leishmaniasis patients and controls from Brazil and the Sudan. All 19 human samples, whether from affected or unaffected individuals, encoded a Gly at this position.

Analysis of the 5′ Promoter Region of Human NRAMP

A 1654 bp *Pst*I fragment subcloned from λB.1 contained exons 1 and 2, and also provided 440 bp of sequence 5′ of the transcription start site (Fig. 3). The transcription start site is located 148 bp 5′ of the ATG initiation codon. A series of predicted promoter region elements also occur 5′ of the transcription start site, including a possible Z-DNA forming (23,24) dinucleotide repeat $t(gt)_5ac(gt)_5ac(gt)_9g$ located −317 to −274 bp 5′ of the transcription start site. On either side of the Z-DNA forming dinucleotide repeat are a series of matches to inducible promoter element consensus sequences. These include six interferon-γ response elements, 1 × 3′→5′ showing 8/8 matches to the consensus sequence $CT^G/_T{}^G/_TANN^C/_T$ (25,26), 3 × 5′→3′ showing 7/8 matches, 2 × 3′→5′ showing 7/8 matches; three W-elements (also known as H-, E-, W-, S-, or Z-boxes), 1 × 3′→5′ showing 8/8 matches to the consensus sequence $^A/_TGNA^C/_AC^C/_T{}^G/_T$ (25), 2 × 5′→3′ with 7/8 matches; an AP1 site showing 6/7 matches to the consensus sequence TGACTCA (27); and three NFκB binding sites, 2 × 5′→3′ and 1 × 3′→5′, each showing 7/10 matches to the consensus sequence $GGG^G/_A{}^C/A/_TT^C/_T{}^C/_TCC$ (28). Nine purine-rich GGAA core motifs (two on the antisense strand) for the myeloid-specific PU.1 transcription factor (29,30) also occur across this region, two of which combine with imperfect AP1-like sites to create PEA3 motifs (31), and another two are juxtaposed. Strings of heat shock transcription factor (HSTF) motifs (NGAAN or NTTCN) (32) were also present, although their order and phase are not consistent with currently functional elements. TATA, GC, and CCAAT boxes were not found within the 440 bp 5′ flanking sequence.

Mapping of a Polymorphic Repeat in the 5′ Promoter Region

The presence of a gt repeat in the 5′ region of the YAC-derived NRAMP sequence stimulated further analysis of this region to determine whether a polymorphism was present in human population samples. Four alleles were observed in Brazilian families (Fig. 4): allele 1 = 122 bp; allele 2 = 120 bp; allele 3 = 118 bp; and allele 4 = 108 bp. Direct sequence analysis confirmed that the polymorphism was located in the largest cluster of gt repeats. Hence, allele 1 = $t(gt)_5ac(gt)_5ac(gt)_{11}g$; allele 2 = $t(gt)_5ac(gt)_5ac(gt)_{10}g$, allele 3 = $t(gt)_5ac(gt)_5ac(gt)_9g$; and allele 4 = $t(gt)_5ac(gt)_5ac(gt)_4g$. Gene frequencies determined on 72 genetically independent Brazilians were 0.021 (allele 1), 0.326 (allele 2), 0.646 (allele 3), and 0.007 (allele 4), providing an overall heterozygosity score of 0.476. Linkage analysis generated positive (>3) LOD scores (Table 2) for linkage between NRAMP and the four closest markers TNP1 (proximal) and IL8RB, VIL1, and DES (distal), consistent with physical mapping

```
(A)  Gene name                                  Conserved amino acid sequence              % Identity to
                                                                                            human NRAMP
                                                          *                   * PKC
HUMAN NRAMP                                  DKGPQRLSGSSYGSISSPTSPTSPGPQQAPPRETYLSEKIPI
MOUSE Nramp                                  DKSPPRLSRPSYGSISS   LPGPAPQPAPCRETYLSEKIPI    72%(28/39)
MICROTUBULE-ASSOCIATED PROTEIN 4               P       S G++S    PTS GP  + PR + L+  I      32%(14/43)
CYTOKINE RECEPTOR COMMON β CHAIN                  L G   G +S  + P+SP P                     45%( 9/20)
PHOSPHOLIPASE C-β 3                                        S P+S  +P P  A PR               47%( 8/17)
ZYXIN REGION I (162-178)                                   S+P  PT P P +  P+               41%( 7/17)
ZYXIN REGION II (265-283)                                     SP+ P  +  PP  TY  +          36%( 7/19)
ADENYLYL CYCLASE-ASSOCIATED PROTEIN                        S+P   TSP P++A  +E              39%( 9/23)
ANKYRIN B REGION I (1812-1844)                    +S SS    SP SP++   + +P   +  +EK         30%(10/33)
ANKYRIN B REGION II (1837-1871)                    S SS    SP SP++   + +P   T      P+      28%(10/35)
PROTEIN PHOSPHATASE SAL6                                GS  +P+SPTS  P                     57%( 8/14)
PTDINS-3-KINASE P85α                                  +Y     P +   PGP    PR               35%( 7/22)

CONSENSUS                                    XXXPXXLSXSSYGSXSSPXSPTSPGPXXAPPRETYLXEKIPX
                                                          *                   *
```

```
                                                                                  6
(B)   Human NRAMP    AFGQAFYQKTKQAAFNICANSSLHDYAKIFPMNNATVAVDIYQGGVILGCLFGPAALYIW
      Mouse Nramp    VFGQAFYQQTNEEAFNICANSSLQNYAKIFPRDNNTVSVDIYQGGVILGCLFGPAALYIW
      SMF 1          LFTLALFVN--------CAILVVAG-STLYNSPE-ADGADLFTIHELLSRNLAPAAGTIF
      SMF 2          LFTVALFVN--------CAILIVSG-ATLYGSTQNAEEADLFSIYNLLCSTLSKGAGTVF
      O. sativa      VSGT------------VCNATNLSPEDAV-----KCSDLTLDSSSFLLRNVLGKSSATVY
      A. thaliana    VSGA------------VCNAPNLSPEDRA-----NCEDLDLNKASFLLRNVVGKWSSKLF
                     .            *    .                  .      .     .   .   .

                                                                      7
      Human NRAMP    AIGLLAAGQSSTMTGTYAGQFVMEGFLRLRWSSFARVLLTRSCAILPTVLVAVFRDLRDL
      Mouse Nramp    AVGLLAAGQSSTMTGTYAGQFVMEGFLKLRWSRFARVLLTRSCAILPTVLVAVFRDLKDL
      SMF1           MLALLLSGQSAGVVCTMAGQIVSEGHINWKLQPWQRRLATRCISIIPCLVISICIGREAL
      SMF2           VLALLFSGQSAGIVCTLSGQMVSEGFLNWTVSPALRRSATRAVAITPCLILVLVAGRSGL
      O. sativa      GVALLASGQSSTITGTYAGQYVMQGFLDIKMKQWLRNLMTRSIAIVPSLIVSII------
      A. thaliana    AIALLASGQSSTITGTYAGQYVMLGFLDLRLEPW--------------------------
                       ..   .   . .   .       .        ..    . .   ... .
```

FIG. 2.

(a) Results of amino acid database searches for exon 2 identifying a number of sequence matches with the Pro/Ser-rich putative SH3 binding domain of NRAMP; (+) a conserved amino acid. Residues showing four or more identities are in bold. Multiple sequence alignments allowed for the generation of a consensus motif over this region (double underlining). Also shown is the PKC site on S40, and tyrosine residues (*) on either side of the consensus motif. (b) Clustal V multiple sequence alignment for human NRAMP, mouse *Nramp*, SMF1 and SMF2, and the expressed sequence tags (40) of *Oryza sativa* (rice; accession number d15268) and *Arabidopsis thaliana* (accession number z30530) genes, reading Frames 1 and 2 respectively. Residues showing ≥4/6 identities across the six proteins are in bold. Membrane spanning domains 6 and 7 for NRAMP are overlined and numbered on the sequence. The 20 amino acid conserved transport motif (1) is indicated by double overlining. All six proteins show identities (similarities) of 7/20 (11/20) across the transport motif. *Cysteine residues conserved across all 6 proteins; (·) conserved substitutions.

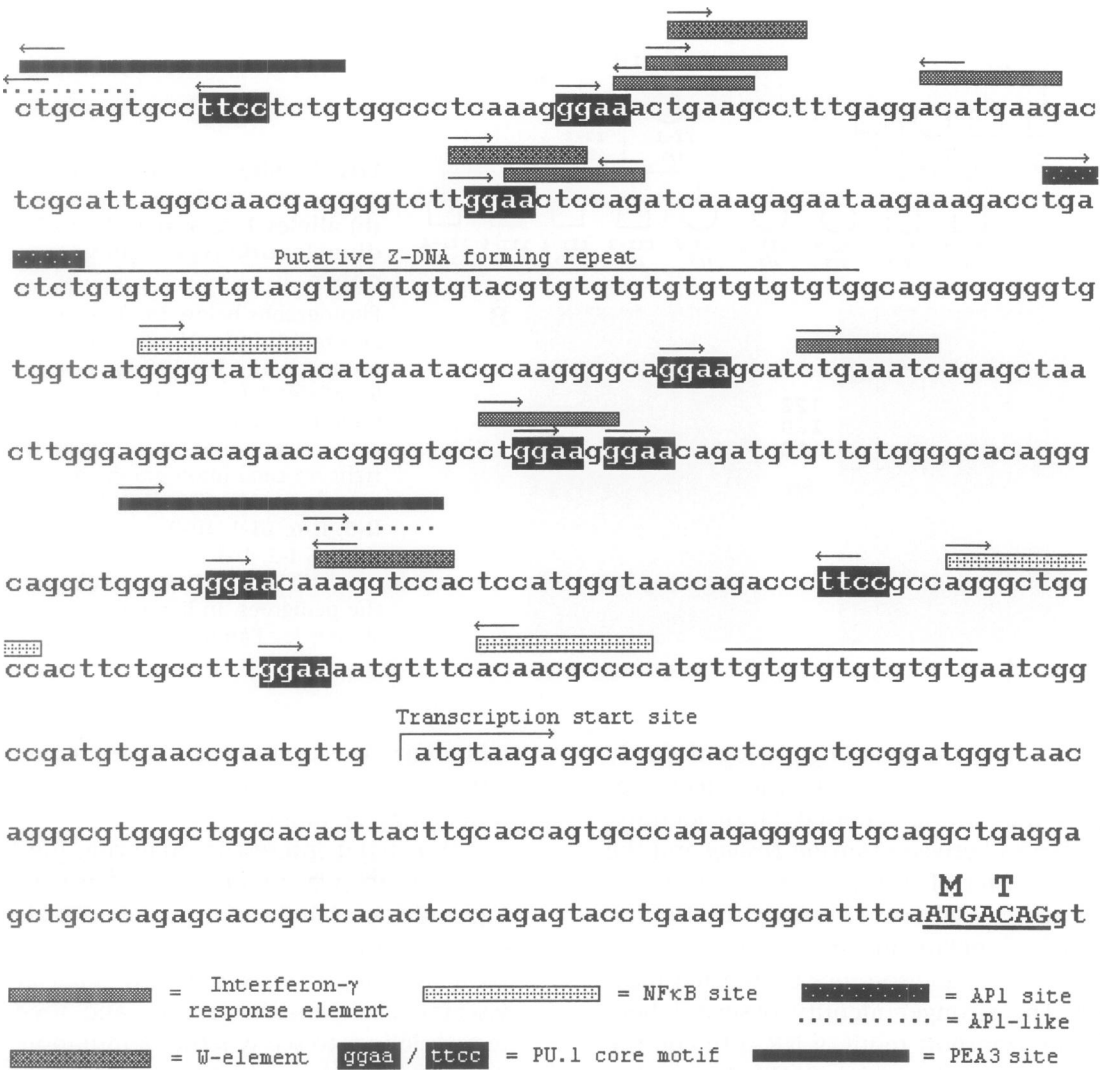

FIG. 3. 440 bp of putative promoter region human NRAMP sequence 5′ of the transcription start site

The transcription start site is located 148 bp 5′ of the ATG initiation codon, as indicated. Putative promoter region elements identified by inspection (indicated above the sequence) include a possible Z-DNA forming dinucleotide repeat $t(gt)_5ac(gt)_5ac(gt)_9g$; 6 interferon-γ response elements; three W-elements; one AP1 site; three NFκB binding sites; and nine purine-rich GGAA core motifs (two on the antisense strand) for the myeloid-specific PU.1 transcription factor, two of which combine with imperfect AP1-like sites to create PEA-3 consensus motifs. Strings of heat shock transcription factor (HSTF) motifs (NGAAN or NTTCN) also occur across the 440 bp sequence (not marked).

data (15) placing NRAMP 130 kb proximal to IL8RB, and confirming that this particular polymorphism occurs in the 2q35 copy of NRAMP rather than in a related sequence (33) mapping to a region in mice homologous to 6q27 in man.

DISCUSSION

Genomic sequence analysis presented here demonstrates that the human NRAMP gene located on chromosome 2q35 has a genomic size of 12 kb and contains 15 exons. The amino acid sequence deduced from nucleotide sequencing of the 15 exons shows that, like murine *Nramp*, NRAMP encodes a polytopic integral membrane protein containing both a conserved transport motif (1) and a putative SH3 binding domain (3). Over the 20 amino acid transport motif, strong sequence identity (7/20 residues; 11/20 with conserved substitutions) was observed between NRAMP (*Nramp*), the two yeast proteins SMF1/2, and the expressed sequence tags from rice and *Arabidopsis*, suggesting that this is a func-

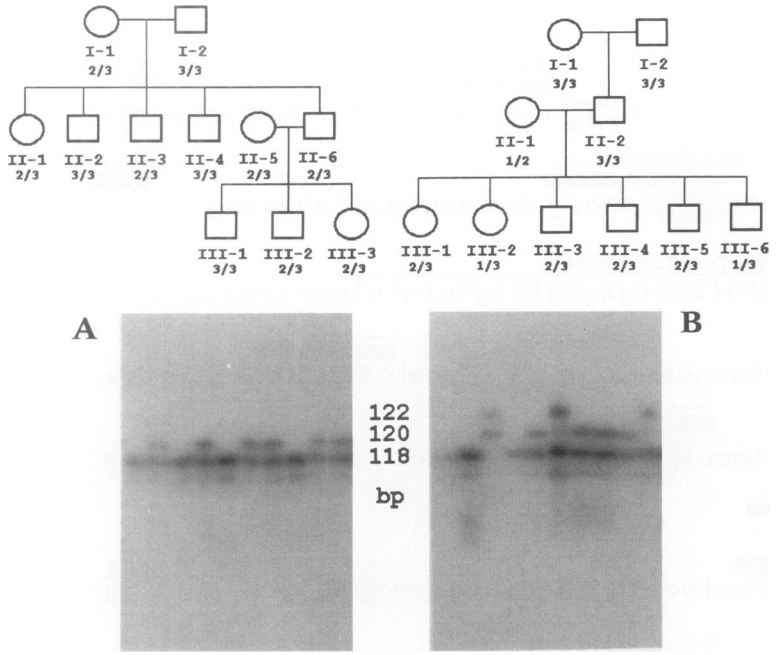

FIG. 4. Shows two families segregating for (a) alleles 2 and 3, or (b) alleles 1, 2, and 3 of the 5′ dinucleotide repeat polymorphism

Photographs below the families show autoradiographs of polymorphic PCR products (122 bp, 120 bp, and 118 bp for alleles 1 to 3, respectively) separated by denaturing polyacrylamide gel electrophoresis. Lanes from left to right on each photograph show individuals (a) I-2, II-1, II-2, II-3, II-4, II-5, II-6, III-1, III-2, and III-3; and (b) I-1, I-2, II-1, II-2, III-1, III-2, III-3, III-4, III-5, and III-6, as indicated on the pedigrees. Individual I-1 is not shown for Family a.

tionally important motif among phylogenetically distinct organisms. Interestingly, these identities are higher than those reported (4/20 identity; 6/20 similarity) between murine *Nramp* and the nitrate transporter of *Aspergillus nidulans*, which led Vidal and coworkers (1) to hypothesise that *Nramp* might function in direct delivery of nitrates into the phagolysosomes of infected macrophages. The stronger identity observed here between the transport motif of NRAMP and the yeast mitochondrial proteins SMF1/2, together with the striking overall similarity between the yeast and human/murine genes, suggests that NRAMP may be a functional homolog to the yeast mitochondrial genes. The yeast genes encode hydrophobic molecules that influence processing enhancing protein-dependent protein import into mitochondria, possibly at the level of translocation (18). Complementation experiments with yeast mutants might therefore reveal more about the molecular mechanism of *Nramp* function. Sequence similarity between NRAMP (*Nramp*) and SMF1/2 was poor over the proline/serine rich putative SH3 binding domain. This is perhaps not unexpected as these are modular structures that occur in a variety of otherwise unrelated proteins involved in signaling and/or cytoskeletal attachment (3). Hence, this modular motif may be a recent addition to the NRAMP molecule related to its macrophage-restricted function, and we might expect that other more ubiquitously expressed NRAMP-like molecules will occur. A second *Nramp*-related sequence has already been mapped in the mouse (33), and others may be found.

Our major interest in analyzing the human NRAMP gene was to provide the basis to screening multicase families for mycobacterial (tuberculosis and leprosy) and leishmanial infections. As a first step, we examined a small group of visceral leishmaniasis patients and their unaffected siblings to see whether a mutation similar to the murine susceptibility-associated mutation (1,2) could be found. As might have been predicted, exon 6 encoding the second membrane spanning domain is highly conserved between murine and human sequences, as well as with the yeast genes, suggesting that this is a functionally important domain. No mutations were found within this region in the 19 human samples examined by direct cycle sequencing. Similarly, a polymorphic variant identified by us (15) in the putative SH3 binding domain occurred at very low frequency, suggesting that this too might be a region of the macrophage-expressed NRAMP molecule which, although recently acquired in evolutionary terms, may be critical to its function and intolerant to nonconservative substitutions.

The 440 bp of promoter region sequence identified here is of particular interest with respect to macrophage-restricted expression of the NRAMP gene, and provides a different approach to analyzing polymorphisms which might influ-

TABLE 2. Peak LOD scores for pairwise linkage analysis between NRAMP and previously mapped (15,20) 2q34 (TNP1) and 2q35 (IL8RB, VIL1, DES) markers calculated for 36 Brazilian families.

Marker Intervals	*n*	Peak LOD Score	RF
TNP1-NRAMP	14	10.49	0.026
TNP1-IL8RB	9	6.02	0.032
TNP1-VIL1	15	9.84	0.001
TNP1-DES	19	11.45	0.046
NRAMP-IL8RB	11	3.56	0.072
NRAMP-VIL1	15	10.94	0.001
NRAMP-DES	20	8.94	0.051
IL8RB-VIL1	10	5.80	0.065
IL8RB-DES	12	10.03	0.035
VIL1-DES	14	9.47	0.059

RF = recombination fraction (M = F) at which the peak LOD score was obtained. *n* = number of families contributing to the analysis.

ence expression rather than cause structural changes to the molecule. Identification of PU.1 and PEA3/AP1-like response elements is consistent with haematopoietic-restricted gene expression (31,34,35). Although earlier studies (1,3) suggest that murine *Nramp* is constitutively expressed in macrophages, the inducible promoter region elements identified in the human sequence suggest that expression may be regulated by macrophage priming/activation stimuli. In particular, interferon-γ and W-elements are common to other genes (e.g., MHC class II, [25]; FcγRI [26]; iNOS [36]) inducible in macrophages. AP1 and NFκB sites also occur in the promoter regions of other macrophage-expressed proteins (e.g., tissue factor [27]; iNOS [36]) and are required for LPS and TNF inducibility, AP1 acting to stabilise and maintain NFκB activity (27). Given the many functional observations (reviewed in Refs. 5,8–10) demonstrating that the *Ity/Lsh/Bcg* (candidate *Nramp*) phenotype is so closely allied to the interferon-γ/LPS macrophage activation pathway, it will be important to determine the functional relevance of these elements to tissue-specific expression of NRAMP in different macrophage subpopulations. This may be particularly relevant to previous observations demonstrating that the *Lsh* gene phenotype is differentially expressed in different macrophage subpopulations (37,38) and that interaction with extracellular matrix elicits different levels of TNFα in bone marrow-derived macrophages from congenic resistant and susceptible mice (39). Although their order and phase were not consistent with currently functional elements, it was of interest that strings of HSTF elements were also found in the promoter region of human NRAMP. These may represent ancestral elements related to the mitochondrial activity/expression of the yeast SMF1 and SMF2 genes.

Another interesting feature of the 5′ flanking region of human NRAMP was the presence of a putative Z-DNA forming dinucleotide repeat $t(gt)_5ac(gt)_5ac(gt)_ng$. A distinct class of binding proteins exists in eukaryotes which interact exclusively with DNA in the Z-conformation, and roles in both positive and negative regulatory signaling have been attributed to this form of DNA (reviewed in Ref. 23). It was particularly intriguing that a polymorphism in this repeat unit was observed in human genomic DNA samples. The fact that the putative Z-DNA forming repeat is flanked on either side by other promoter region response elements suggests that this polymorphism may be functionally important in determining gene expression, if not on the basis of its own role as a transcriptional regulator, at least because it will influence the juxtaposition of other response elements. The level of heterozygosity (0.476) in the Brazilian population studied here made this a useful marker for genetic linkage analysis between NRAMP and other 2q markers. However, the number of alleles was small compared with other repeat (e.g., microsatellite) polymorphisms, suggesting that the generation of further polymorphic variants across this repeat may not be tolerated in evolutionary terms. This polymorphism may therefore be of functional relevance in further analysis of the association between NRAMP and disease. Our own analysis of association between NRAMP and leprosy, TB, or visceral leishmaniasis in the Brazilian population from which linkage data was derived is in progress. Such studies will also need to take account of mutations/polymorphisms across coding region sequences. The data provided in this study will provide some of the tools required for further functional and genetic analysis of diseases in humans involving defective macrophage function.

ACKNOWLEDGMENTS

This work was supported by grants from The Wellcome Trust. We acknowledge the assistance of Christopher Peacock in collecting and processing blood samples from the Brazilian families, and Drs Jeffrey Shaw, Fernando Silveira, Luzio Ramos, and Zea Lins-Lainson of the Instituto Evandro Chagas, Belem, Brazil for all their help in making the Brazilian family study possible. Dr. A. M. El Hassan kindly provided samples from visceral leishmaniasis families in Sudan.

REFERENCES

1. Vidal SM, Malo D, Vogan K, Skamene E, Gros P. (1993) Natural resistance to infection with intracellular parasites: Isolation of a candidate for *Bcg*. *Cell* **73:** 469–485.
2. Malo D, Vogan K, Vidal S, et al. (1994) Haplotype mapping and sequence analysis of the mouse *Nramp* gene predict susceptibility to infection with intracellular parasites. *Genomics* **23:** 51–61.
3. Barton CH, White JK, Roach TIA, Blackwell JM. (1994) NH_2-terminal sequence of macrophage-expressed natural resistance-associated macrophage protein (*Nramp*) encodes a proline/serine-rich putative *Src* homology 3-binding domain. *J. Exp. Med.* **179:** 1683–1687.
4. Finan P, Shimizu Y, Gout I, et al. (1994) An SH3 domain and proline-rich sequence mediate an interaction between two components of the phagocyte NADPH oxidase complex. *J. Biol. Chem.* **269:** 13752–13755.
5. Blackwell JM. (1989) (Convenor, 27th Forum in Immunology) The macrophage resistance gene Lsh/Ity/Bcg. *Res. Immunol.* **140:** 767–828.
6. Blackwell JM, Roach TIA, Atkinson SE, Ajioka JW, Barton CH, Shaw M-A. (1991) Genetic regulation of macrophage priming/activation: The *Lsh* gene story. *Immunol. Lett.* **30:** 241–248.
7. Blackwell JM, Toole S, King M, Dawda P, Roach TI, Cooper A. (1988) Analysis of Lsh gene expression in congenic B10.L-Lshr mice. *Curr. Top. Microbiol. Immunol.* **137:** 301–309.
8. Blackwell JM, Barton CH, White JK, et al. (1994) Genetic regulation of leishmanial and mycobacterial infections: The *Lsh/Ity/Bcg* gene story continues. *Immunol. Lett.* **43:** 99–107.
9. Schurr E, Radzioch D, Malo D, Gros P, Skamene E. (1991) Molecular genetics of inherited susceptibility to intracellular parasites. *Behring. Inst. Mitt.* **88:** 1–12.
10. Schurr E, Malo D, Radzioch D, et al. (1991) Genetic control of innate resistance to mycobacterial infections. *Immunol. Today* **12:** A42–A45.
11. Kaye PM, Blackwell JM. (1989) Lsh, antigen presentation and the development of CMI. *Res. Immunol.* **140:** 810–815.
12. Schneeman M, Schoedon G, Hofer S, Blau N, Guerrero L, Schaffer A. (1993) Nitric oxide is not a constituent of the antimicrobial armature of human mononuclear phagocytes. *J. Infect. Dis.* **167:** 1358–1363.
13. Larin Z, Monaco AP, Lehrach H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. U.S.A.* **88:** 4123–4127.
14. Pringault E, Arpin M, Garcia A, Finidori J, Louvard D. (1986) A human villin cDNA clone to investigate the differentiation of intestinal and kidney cells in vivo and in culture. *E.M.B.O. J.* **5:** 3119–3124.
15. White JK, Shaw M-A, Barton CH, et al. (1994) Genetic and physical mapping of 2q35 in the region of NRAMP and IL8R genes: Identification of a polymorphic repeat in exon 2 of NRAMP. *Genomics.* **24:** 295–302.
16. Tsuchiya S, Yamabe M, Yamaguchi Y, Kobayashi Y, Konno T, Tada K. (1980) Establishment and characterization of a human acute monocyte leukemia cell line (THP-1). *Int. J. Cancer* **26:** 171–176.
17. Edwards JBDM, Delort J, Mallet J. (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: A new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by in vitro amplification. *Nucleic Acids Res.* **19:** 5227–5232.
18. West AH, Clark DJ, Martin J, Neupert W, Hartl FU, Horwich AL. (1992) Two related genes encoding extremely hydrophobic proteins suppress a lethal mutation in the yeast mitochondrial processing enhancing protein. *J. Biol. Chem.* **267:** 24625–24633.
19. Higgins DG, Sharpe PM. (1988) CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73:** 237–244.
20. Shaw M-A, Atkinson S, Dockrell H, et al. (1993) An RFLP map for 2q33-q37 from multicase mycobacterial and leishmanial disease families: No evidence for an *Lsh/Ity/Bcg* gene homologue influencing susceptibility to leprosy. *Ann. Hum. Genet.* **57:** 251–271.
21. Lathrop GM, Lalouel JM. (1984) Easy calcu-

lations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.* **36:** 460–465.
22. Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44:** 283–292.
23. Rich A, Nordheim A, Wang AH-J. (1984) The chemistry and biology of left-handed Z-DNA. *Ann. Rev. Biochem.* **53:** 791–846.
24. Hamada H, Kakunaga T. (1982) Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298:** 396–398.
25. Yang Z, Sugawara M, Ponath PD, et al. (1990) Interferon gamma response region in the promoter of the human DPA gene. *Proc. Natl. Acad. Sci. U.S.A.* **87:** 9226–9230.
26. Pearse RN, Feinman R, Ravetch JV. (1991) Characterization of the promoter of the human gene encoding the high-affinity IgG receptor: Transcriptional induction by gamma-interferon is mediated through common DNA response elements. *Proc. Natl. Acad. Sci. U.S.A.* **88:** 11305–11309.
27. Mackman N, Brand K, Edgington TS. (1991) Lipopolysaccharide-mediated transcriptional activation of the human tissue factor gene in THP-1 monocytic cells requires both activator protein 1 and nuclear factor kB binding sites. *J. Exp. Med.* **174:** 1517–1526.
28. Lenardo MJ, Baltimore D. (1989) NF-kB: A pleiotropic mediator of inducible and tissue-specific gene control. *Cell* **58:** 227–229.
29. Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA. (1990) The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. *Cell* **61:** 113–124.
30. Karim FD, Urness LD, Thummel CS, et al. (1990) The ETS-domain: A new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Gene. Dev.* **4:** 1451–1453.
31. Cassady AI, Stacey KJ, Nimmo KA, et al. (1991) Constitutive expression of the urokinase plasminogen activator gene in murine RAW264 macrophages involves distal and 5′ non-coding sequences that are conserved between mouse and pig. *Nucleic Acids Res.* **19:** 6839–6847.
32. Lis J, Wu C. (1992) Heat shock factor. In: McKnight SL, Yamamoto KR (eds). *Transcriptional Regulation.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp. 907–930.
33. Dosik JK, Barton CH, Holiday DL, Krall MM, Blackwell JM, Mock BA. (1994) An Nramp-related sequence maps to mouse chromosome 17. *Mammalian Genome* **5:** 458–460.
34. Feinman R, Qiu WQ, Pearse RN, et al. (1994) PU.1 and an HLH family member contribute to the myeloid-specific transcription of the Fc-gamma-RIIIA promoter. *E.M.B.O. J.* **13:** 3852–3860.
35. Zhang D-E, Hetherington CJ, Chen H-M, Tenen DG. (1994) The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Mol. Cell. Biol.* **14:** 373–381.
36. Xie Q-w, Whisnant R, Nathan C. (1993) Promoter of the mouse gene encoding calcium-independent nitric oxide synthase confers inducibility by interferon gamma and bacterial lipopolysaccaride. *J. Exp. Med.* **177:** 1779–1784.
37. Crocker PR, Davies EV, Blackwell JM. (1987) Variable expression of the murine natural resistance gene *Lsh* in different macrophage populations infected in vitro with *Leishmania donovani*. *Parasite Immunol.* **9:** 705–719.
38. Davies EV, Singleton AM, Blackwell JM. (1988) Differences in Lsh gene control over systemic Leishmania major and Leishmania donovani or Leishmania mexicana mexicana infections are caused by differential targeting to infiltrating and resident liver macrophage populations. *Infect. Immun.* **56:** 1128–1134.
39. Formica S, Roach TIA, Blackwell JM. (1994) Interaction with extracellular matrix proteins influences *Lsh/Ity/Bcg* (candidate *Nramp*) gene regulation of macrophage priming/activation for tumour necrosis factor α and nitrite release. *Immunology* **82:** 42–50. factor α and nitrite release. *Immunology* **82:** 42–50.
40. Boguski MS, Lowe TMJ, Tolstoshev CM. (1993) dbEST—Database for "expressed sequence tags". *Nature Genet.* **4:** 332–333.

Contributed by K. Peters on October 13, 1994.