

# Single nucleotide polymorphism (SNP) discovery and linkage mapping of bovine cytokine genes

William M. Grosse, Steven M. Kappes, William W. Laegreid, John W. Keele, Carol G. Chitko-McKown, Michael P. Heaton

United States Department of Agriculture, Agricultural Research Service, Roman L. Hruska U.S. Meat Animal Research Center (MARC), P.O. Box 166, State Spur 18D, Clay Center, Nebraska 68933, USA

Received: 30 April 1999 / Accepted: 12 July 1999

**Abstract.** Polymorphic markers at bovine gene loci facilitate the integration of cattle genetic maps with those of humans and mice. To this end, 31 single nucleotide polymorphism (SNP) markers were developed for seven bovine chemokine genes. Loci were amplified from bovine genomic DNA by the polymerase chain reaction, and candidate amplicons were sequenced to determine their identity. Amplified loci from 24 founding parents and select progeny from a beef cattle reference population were sequenced and analyzed for SNPs. SNP haplotype alleles were determined by examining segregation patterns and used to establish the locus position on the bovine linkage map. Loci for growth-related proteins (GRO3, GRO1, and GROX) were clustered with the related CXC chemokine genes, interleukin (IL) 8, and epithelial cell inflammatory protein 1, at 84 cM from the centromeric end of the bovine chromosome (BTA) 6 linkage group. Bovine loci for a cluster of IL8 receptors, a stromal cell-derived factor 1, interferon- $\gamma$ , and tumor necrosis factor- $\alpha$  were mapped at 90, 55, 59, and 34 cM, respectively, from the centromeric ends of the BTA 2, 28, 5, and 23 linkage groups. The positions of these bovine loci were compared with those of orthologous loci on the human map to refine the boundaries of conserved synteny. These seven loci provide examples of SNP development in which the efficiency was largely dependent on the availability of bovine genomic or cDNA sequence. The polymorphic nature of these SNP haplotype markers suggests that they will be useful for mapping complex traits in cattle, such as resistance to infectious disease.

## Introduction

Infectious diseases are a significant source of economic loss to the cattle industry (NASS 1996). In addition, cattle infected with foodborne pathogens such as *Escherichia coli* O157:H7 and *Salmonella typhimurium* DT104 have become an emerging human health issue (Glynn et al. 1998; Mead and Griffin 1998). A potential method for reducing the impact of infectious disease is to increase host genetic resistance to infection. Among the genes likely to influence the magnitude, duration, and transmission of infection are those that encode cytokines. This group of proteins plays a pivotal role in mobilizing inflammatory cells in response to infectious challenges (Krakauer et al. 1999). Moreover, specific allelic variants of cytokines and their receptors profoundly influence infection phenotypes in humans (Smith et al. 1997; Altare et al. 1998; Martin et al. 1998; Winkler et al. 1998; Knight et al. 1999).

In an effort to identify bovine genes that are modulated during infection, we previously reported two pro-inflammatory cytokine transcripts whose abundance increased in epithelial cells upon ex-

posure to *E. coli* O157:H7 lipopolysaccharide (Heaton et al. 1999). Both cDNAs encoded members of the CXC (or  $\alpha$ ) chemokine subfamily, specifically interleukin (IL)-8 and epithelial cell inflammatory protein (ECIP)-1. We hypothesize that allelic variation in chemokines, cytokines, or their receptors may be associated with variation in response to infection in beef cattle populations. Testing this hypothesis requires genetic markers for identifying allelic variants at each gene locus. In humans, there are 11 known CXC chemokines and more than 150 cytokines and receptors (Vaddi et al. 1997; Thomson 1998). This paper describes the identification of 31 single nucleotide polymorphism (SNP) markers for a set of seven bovine cytokine loci and their use in linkage mapping.

## Materials and methods

**Primer design and synthesis.** Primers (Table 1) were designed from DNA and protein sequence information available in the National Center for Biotechnology Information databases (<http://www.ncbi.nlm.nih.gov/>) and the Genome Database (<http://www.gdb.org/>). Sequences were analyzed for primer design by OLIGO 5.0 (National Biosciences Inc., Plymouth, Minn.) and were synthesized on a 1000 M DNA synthesizer (Beckman Coulter, Inc., Fullerton, Calif.). Amplification conditions for each locus were optimized, where possible, to produce single amplicons of the predicted length. Non-coding gene regions were selected for polymerase chain reaction (PCR) based on the observation that introns and intergenic regions tend to have more sequence variation than exon regions (Turker et al. 1993; Nickerson et al. 1998). In all cases, primers were designed to amplify a portion of known sequence to help establish the relationship between amplicon and gene identity.

**Amplification and sequencing.** The standard amplification reaction contained 200 ng of genomic DNA, 2  $\mu$ M each of sense and antisense primer, 0.1 mM of each dNTP, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.1% vol/vol Triton X-100, 10 mM TrisHCl (pH 9.0), and 4 U of Taq polymerase in a total volume of 50  $\mu$ l. PCR was performed with a PTC-200 DNA engine (MJ Research, Watertown, Mass.) with temperature ramp times of 2.5°C/s. Reactions were denatured at 94°C for 2 min and subjected to 35 cycles of denaturation at 94°C for 30 s, annealing at the appropriate temperature (Table 1), and a 72°C extension for 1 min. After cycling, an additional 2-min incubation at 72°C was included before storage at 4°C. A portion of each amplified product was analyzed by agarose gel electrophoresis in buffer containing 90 mM Tris-borate (pH 8.0), 2 mM ethylenediamine tetraacetic acid, and 0.1  $\mu$ g/ml ethidium bromide. Unless otherwise indicated, amplification products were purified with silica membrane columns (QIAGEN Inc., Valencia, Calif.). Sequencing reactions were assembled according to manufacturer's instructions with BigDye terminator chemistry, and resolved on an ABI PRISM 377 DNA sequencer (PE Applied Biosystems, Foster City, Calif.).

The 2.5-kb stromal cell-derived factor (SDF1) 1 PCR product that spanned intron 2 was amplified with high fidelity DNA polymerase (Gibco BRL, Gaithersburg, Md.). Cycling conditions were 94°C for 1 min followed by 34 cycles of 94°C for 15 s, 55°C for 20 s, and 68°C for 10 min.

**Table 1.** Oligonucleotide primers, annealing temperatures, and amplicon lengths.

Locus	Primer orientation	Amplification primer sequences <sup>a</sup>	Annealing temperature (°C)	Amplicon length (bp)	Amplicon	Amplicon haplotype accession numbers <sup>b</sup>	Reference sequence
<i>GRO3</i>	Sense	<b><i>CCCATGGTTAAGAAAATCATC</i></b>	60	829	bGRO3DS3	AF140644-647	U95811
	Antisense	<b><i>GTTTTCCACCTGGTCAGTTAG</i></b>					
<i>GRO1</i>	Sense	<b><i>TGCCAACTGATCAAGAGAG</i></b>	60	456	bGRO1DS2	AF140641-643	U95812
	Antisense	<b><i>CATTCCCTACATTAACAGTG</i></b>					
<i>GROX</i>	Sense	<b><i>CCCATGGTTAAGAAAATCATC</i></b>	60	912	bGROXDS1	AF140660-664	U95813
	Antisense	<b><i>CCATCTCTCTTGATCAGTTGG</i></b>					
<i>IL8R</i>	Sense	<b><i>ACGCACGCTGACCCAGAA</i></b>	60	523	bIL8RABDS1	AF140648-652	U19947
	Antisense	<b><i>GAGGAGTCCGTGGCGAAACT</i></b>					
<i>SDF1</i>	Sense	<b><i>CAACACTCCAAACTGTGCCCTTCAGA</i></b>	55	2500	bSDF1DS1	na <sup>c</sup>	U16752
	Antisense	<b><i>AACAAATTTGAAAGAGGTCATGAG</i></b>					
<i>SDF1</i>	Sense	<b><i>CTCCCATCCCCAGACTA</i></b>	65	605	bSDF1DS2	AF140665-671	bSDF1DS1
	Antisense	<b><i>TGTTCTTCAGCCTTGCCCTGTG</i></b>					
<i>IFNG</i>	Sense	<b><i>CGGGAGATTGCTTTCATTTTC</i></b>	65	892	bIFNGDS3	AF140653-655	Z54144
	Antisense	<b><i>GGTTAGATTTTGGCGACAGGT</i></b>					
<i>TNFA</i>	Sense	<b><i>CTGGAGAAGTGGGGGTCA</i></b>	60	639	bTNFADS1	AF140656-659	Z14137
	Antisense	<b><i>TTAGAAATGGGAGGGCTTTAT</i></b>					

<sup>a</sup> Sequences listed 5' to 3'. Primers used for DNA sequencing are bold and italicized.

<sup>b</sup> Each haplotype sequence has been submitted to GenBank.

<sup>c</sup> The full sequence of this amplicon was not determined.

After cycling, a final extension step of 68°C for 2 min was added before storage at 4°C. The 2.5-kb product for *SDF1* was gel purified and sequenced with the amplification primers. Sufficient sequence information was obtained from this product to design new primers for amplifying the 603-bp fragment used for SNP identification (Table 1).

**SNP identification.** The U.S. Meat Animal Research Center (MARC) reference population (Bishop et al. 1994) was used to identify SNP candidates. Loci from founding parents were amplified, sequenced, and compared to identify putative SNPs. Sequences were edited with Factura 2.0.1 (PE Applied Biosystems), and potential heterozygous bases were noted. Sequences from reference animals were aligned with the Clustal option of the ABI Sequence Navigator 1.0.1 software (PE Applied Biosystems) to identify candidate SNPs. These markers were genotyped in reference progeny by DNA sequencing to verify Mendelian inheritance patterns and SNP haplotypes within the amplicon locus.

**Linkage mapping.** SNP haplotype markers were placed in linkage groups based on twopoint LOD scores and ordered within groups by multipoint analysis with the ALL, TWOPOINT, FIXED, and CHROMPIC options of CRI-MAP Version 2.4 (Green et al. 1990). Additional marker information is present in the MARC cattle genome mapping database (<http://www.marc.usda.gov>).

## Results

**SNP markers for bovine *GRO* genes.** The third intron of the bovine *GRO3* gene, predicted from human sequence, was targeted for amplification and DNA sequencing (Fig. 1A). Analysis of the 829-bp *GRO3* amplification product indicated that the last 26-bp from the 3' end were 100% identical to those of exon 4 from *GRO3*. This combination of 26 nucleotides (nt) in *GRO3* is not found in *GRO1*, *GRO2*, *ECIP1*, or any other GenBank sequence. Intron 3 of the bovine *GRO3* amplicon was 803-bp and contained approximately 265 bp of a 560-bp short interspersed element (SINE) sequence similar to Bov-2 (Lenstra et al. 1993). By comparison, the 874-bp intron 3 of *ECIP1* contains a complete 188-bp SINE sequence similar to Bov-t1 (Lenstra et al. 1993). These results suggest that the 829-bp PCR product was amplified from the bovine *GRO3* gene and its intron structure is similar to *ECIP1*.

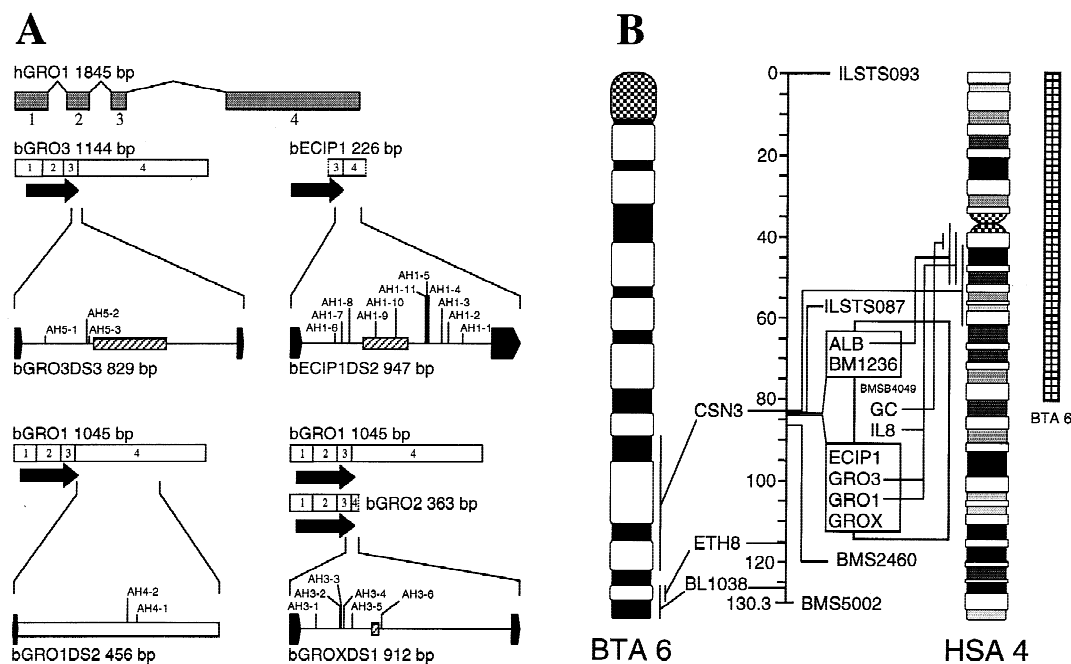
The density of SNPs in the *GRO3* intron 3 was roughly comparable to that in *ECIP1*; three variable sites were present in the first 300 bp. Four heritable sets of SNP haplotype alleles were determined by examining segregation patterns in progeny (Table

2). Twopoint analysis with 198 informative meioses placed the *GRO3* locus 83.8 cm from the centromeric end of the bovine chromosome (BTA) 6 linkage group (Fig. 1B). This position was indistinguishable from *IL8* and *ECIP1* with no recombinations detected in 49 and 46 coinformative meioses, respectively. Although more coinformative meioses are required for ordering these loci, these results suggested that the *GRO3* locus is tightly linked to *IL8* and *ECIP1* on BTA6 (LOD scores of 12.03 and 12.64, respectively).

To obtain similar markers for other bovine *GRO* genes, a 456-bp portion of the *GRO1* putative exon 4 was selected for amplification (Fig. 1A). Comparison of the amplified bovine DNA sequence with the 3' untranslated region (UTR) of bovine *GRO1* (Accession No. U95812) revealed 99% sequence identity (three conflicting sites). The 456-bp amplicon sequence did not display significant homology with the 3' UTR of bovine *GRO3* or *ECIP1* (BLAST search E value > 0.5). The 3' UTR sequence for bovine *GRO2* (Accession No. U95813) was not available for comparison. The high degree of identity with the 3' UTR of bovine *GRO1* suggested that the 456-bp amplicon was from the *GRO1* gene.

SNP analysis of the 456-bp *GRO1* amplicon revealed the presence of two polymorphisms in this exon region (Fig. 1A), neither of which corresponded to the sites conflicting with bovine *GRO1* (Accession No. U95812). Three SNP haplotype alleles were defined for the *GRO1* amplicon and used in twopoint linkage analysis with 149 informative meioses to place the *GRO1* locus 83.8 cM from the centromeric end of the BTA 6 linkage group (Table 2). Multipoint analysis showed that the *GRO1* map position was indistinguishable from that of *IL8*, *ECIP1*, and *GRO3*, with no recombinations detected in 10, 26, and 84 coinformative meioses, respectively (Fig. 1B). These results indicated that the putative bovine *GRO1* gene was tightly linked to the CXC chemokine gene cluster on BTA 6.

After bovine SNP markers were developed for *GRO3* and *GRO1*, efforts were directed towards *GRO2*. The high degree of similarity between bovine *GRO1* and *GRO2* cDNA sequences presented a significant challenge in designing primers to amplify a single gene. Among the 363-bp of sequence available for *GRO2*, there were only four nt positions and one amino acid residue that differed when compared with *GRO1*. To circumvent this difficulty, the aim was to amplify across intron 3 from both *GRO1* and *GRO2* and clone individual loci for sequencing. Sequence from cloned products would then be used for designing new sets of intron-



**Fig. 1.** Genetic maps of GRO-family genes and BTA 6. Panel A: Physical maps of human GRO1 genomic DNA, bovine GRO-like cDNA, and bovine GRO-like genomic DNA. The symbol legend is as follows: shaded rectangles, genomic exon regions; lines connecting shaded rectangles, introns; open rectangles, cDNA sequence; dashed vertical lines, ends of partial cDNA sequences; solid arrows, coding sequence; hatched rectangles, SINE elements. These SNP marker names in the MARC cattle database (<http://sol.marc.usda.gov>) begin with an "AH" prefix and their relative position is indicated by vertical lines within the amplicon. The physical map of ECIP1 (Heaton et al. 1999) is presented for comparison. Panel B: Linkage map and ideograms of BTA 6 (Kappes et al. 1997) and comparison with HSA 4 (Francke 1994). Physical assignment of loci to chromosomes is indicated

with vertical lines adjacent to assigned region. For simplicity, the linkage map includes only terminal markers, gene markers, and markers adjacent to GRO-loci. A more detailed and interactive map may be accessed at [http://sol.marc.usda.gov/genome/cattle/htmls/chromosome\\_list.html](http://sol.marc.usda.gov/genome/cattle/htmls/chromosome_list.html). No recombinations were observed between markers within a box; thus, their order is arbitrary. The relative order of *ECIP1*, *GRO3*, *GRO1*, and *GROX* with respect to ALB and BM1236 was determined by a recombination detected in one calf. The remaining loci in the box (*IL8*, *GC*, and *BMSB4049*) were not informative in this animal. The hatched vertical rectangle indicates the region of HSA4 known to correspond to BTA 6 (Solinas-Toldo et al. 1995; Chowdhary et al. 1996).

specific amplification primers. However, the amplification reaction generated a single 912-bp product (designated *GROX*). Sequence analysis of *GROX* showed no sites with 100% heterozygosity in the reference population, as would be expected if two closely related genes were simultaneously amplified. Moreover, *GROX* contained 66 of 67-bp of sequence from the putative exons 3 and 4 that were identical to *GRO1* and *GRO2*. The site in *GROX* conflicting with both *GRO1* and *GRO2* was identical in 24 reference parents and corresponded to a substitution of asparagine for lysine at position 96. This substitution is significant because the lysine 96 residue is conserved in 25 other GRO-like genes from various mammalian species (data not shown). The *GROX* amplicon contained a 846-bp intron, including a partial (54-bp) Bov-t1 SINE sequence (Fig. 1A). When a BLAST search against the non-redundant GenBank database was performed with the *GROX* intron (non-SINE portion), the lowest E value returned was 2.5. These results suggest that the *GROX* product was amplified from a single gene or pseudogene that is distinct from the known GRO-like bovine genes.

The 5' portion of the *GROX* amplicon was analyzed for SNPs, and six polymorphisms were identified in the first 500 bp in the reference population (Table 2). Five SNP haplotype alleles were defined and used in twopoint linkage analysis with 122 informative meioses to place the *GROX* locus in the BTA 6 linkage group. Multipoint analysis showed that the *GROX* position was indistinguishable from that of *ECIP1*, *GRO3*, and *GRO1*, with no recombinations detected in 26, 56, and 96 coninformative meioses, respectively (Fig. 1B). There were no coinformative meioses with *IL8*. Taken together, these results indicate that the *GROX* locus is clustered with the CXC chemokines on BTA 6.

*SNP markers for a cluster of IL8 receptor genes from bovine cDNA sequence.* The identification of multiple SNPs in each of the GRO loci analyzed indicated that the reference population was highly suitable for SNP marker development and linkage mapping of bovine chemokine genes. A query of the publicly available databases for other bovine genes in the CXC chemokine family revealed a sequence for only IL8 receptor B (GenBank Accession No. U19947). The IL8 receptors A and B (*CXCR1* and *CXCR2*, respectively) are two distinct seven-transmembrane, G-protein coupled receptors that interact directly with a variety of CXC chemokines to induce neutrophil chemotaxis and margination (Murphy 1997). In humans, both receptors and several pseudogenes have been mapped to human chromosome (HSA) 2q33-2q36 (Holmes et al. 1991; Morris et al. 1992; Lloyd et al. 1993; Mollereau et al. 1993; Sprenger et al. 1994). The cDNA sequence for bovine IL8RB was used for designing several primer pairs. The most robust amplification yielded a 523-bp fragment that was nested within the coding sequence of the predicted third exon (Fig. 2A). Sequence analysis of the 523-bp product revealed 16 sites that were heterozygous in all cattle tested (70 parents and progeny), suggesting that multiple IL8R-like loci were present in the amplified product.

In addition to the 16 invariant heterozygous sites, there were four sites that varied only in certain parents, and their alleles segregated in progeny according to Mendelian patterns. When compared with bovine IL8RB, three of the sites corresponded to silent mutations in codons, whereas the fourth site variant corresponded to a glutamine substitution for the histidine residue at position 245. Because the IL8R loci were clustered in the human genome, we hypothesized that the four SNPs might be useful for

**Table 2.** Summary of SNP and haplotype linkage information.

Locus (Chr, cM <sup>a</sup> )	Amplicon	Marker	Informative meioses	Allele definition <sup>b</sup>	Linked markers with LOD scores >3								
					Number	Range	Nearest marker <sup>c</sup>						
<i>GRO3</i> (6, 83.8)	bGRO3DS3	<i>GRO3-1</i>	198	1 ATT 2 GTT 3 ATG 4 ACT	50	3.22 to 47.89	<i>BMSB4049</i>						
			85	CTAGG-R-TTTAG									
			78	TAAAC-Y-GCTTT									
			82	TTCAT-K-AAAAT									
			149	1 AT 2 GA 3 AA									
<i>GRO1</i> (6, 83.8)	bGRO1DS2	<i>GRO1-1</i>	71	ACTTA-W-AGTTA	50	3.01 to 32.39	<i>BMSB4049</i>						
			138	TTCAA-R-GTTTA									
			122	1 TCG(T) <sup>d</sup> CA 2 TAA(T) <sup>d</sup> CA 3 GCA(T) <sup>d</sup> GA 4 TCG(T) <sup>d</sup> CC 5 GCA( ) <sup>d</sup> GA									
<i>GROX</i> (6, 83.8)	bGROXDS1	<i>GROX-1</i>	88	GACTA-K-AGTCA	38	3.01 to 24.38	<i>GRO3-1</i>						
			50	AAGGT-M-ATTAR									
			74	MATTA-R-TCCGC									
			63	TGGTG-(T) <sup>d</sup> -CAAAA									
			27	CTCCT-S-GCTCT									
			8	GAAGC-M-GAGTT									
			<i>IL8R</i> (2, 90.3)	bIL8RDS1				<i>IL8R-1</i>	71	1 GCCC 2 GCCA 3 AGCC 4 ACCC 5 GCAA	18	3.61 to 16.39	<i>TEXAN-4</i>
									37	CTGAT-M-GYRGA			
									13	AGCAC-M-GGGCC			
									50	GGCA-S-AAGCA			
57	CCACC-R-TATTC												
92	1 CTCC( ) <sup>e</sup> C 2 CCCC(C) <sup>e</sup> C 3 CTCG( ) <sup>e</sup> C 4 GTCC( ) <sup>e</sup> C 5 CCCC( ) <sup>e</sup> C 6 CCTC( ) <sup>e</sup> C 7 CCTC( ) <sup>e</sup> T												
<i>SDF1</i> (28, 54.9)	bSDF1DS2	<i>SDF1-1</i>	52	TCCAT-S-AAAAT	18	3.16 to 22.58	<i>BM2515</i>						
			62	CCCTC-Y-GGTTC									
			49	CTGGC-Y-GCAGG									
			28	CACCC-S-CC <sup>g</sup> ( ) <sup>e</sup> TGG									
			13	CCSCC <sup>h</sup> (C) <sup>e</sup> -TGGAG									
			37	GCCCA-Y-TCCTA									
			<i>IFNG</i> (5, 59.3)	bIFNGDS3				<i>IFNG-1</i>	81	2 C(ACT) <sup>g</sup> CCCTA 3 T(ACT) <sup>g</sup> ATTCTG 4 C( ) <sup>g</sup> CCCTA	19	3.79 to 12.40	<i>BMS1617</i>
									71	AGATC-Y-GTGTT			
									10	AAAST-(ACT) <sup>g</sup> -GAACC			
									64	GATAG-M-AGTGA			
64	AGTGA-Y-GGAGA												
64	AGAGC-Y-TCAGT												
64	AGTGG-Y-AAATT												
<i>TNFA</i> (23, 33.5)	bTNFADS1	<i>TNF-1</i>	91	1 GTG 2 GTC 3 ACG 4 GCG	31	3.60 to 21.98	<i>BMS2275</i>						
			42	ACAGG-R-GCTCT									
			54	GGAAA-Y-TGGAG									
			59	CTGCT-S-TGCTG									

<sup>a</sup> Distance of locus from centromere. Total linkage group sizes for BTA 6, 2, 28, 5, and 23 were 130.3, 122.2, 54.9, 122.9 and 67.0 cM, respectively.

<sup>b</sup> The haplotype allele definition is presented as a set of sequential SNP alleles in the same order as shown in Figs. 1A through 5A. The MARC haplotype allele definition code is listed to left of each haplotype allele. The SNP alleles are defined by five flanking nucleotides on either side.

<sup>c</sup> Marker with highest LOD score and zero recombination frequency.

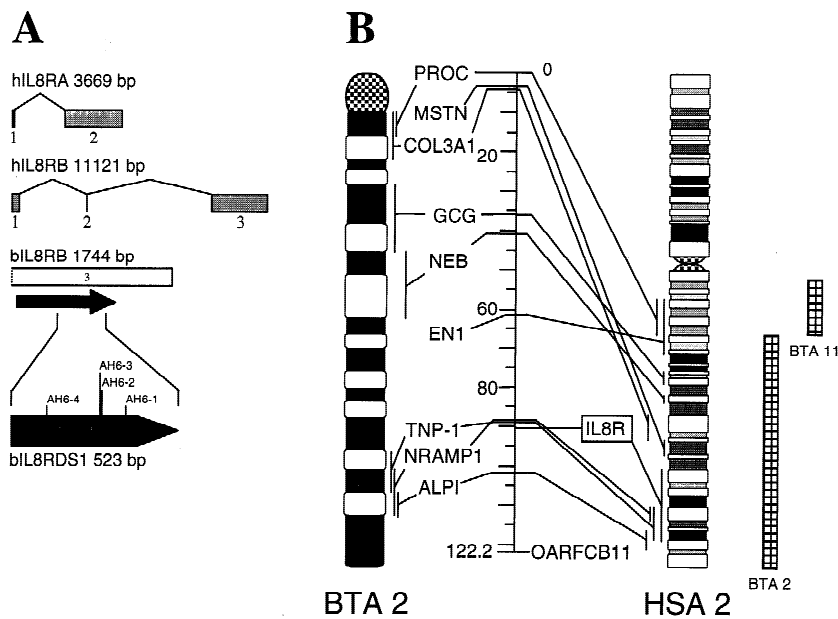
<sup>d</sup> Site of thymine nucleotide deletion.

<sup>e</sup> Site of cytosine nucleotide insertion.

<sup>f</sup> The "C\*" nucleotide has only been observed by mass spectrometry analysis and is not polymorphic in this population (data not shown).

<sup>g</sup> Site of ACT-nucleotide deletion.





**Fig. 2.** Genetic maps of *IL8R* genes and BTA2. The symbol legends are described in Fig. 1. Panel A: Physical maps of human *IL8RA* and *IL8RB* genomic DNA, partial bovine *IL8RB* cDNA, and bovine *IL8R* cluster amplicon. Panel B: Linkage and physical maps of BTA 2 and comparison with HSA 2.

linkage mapping in spite of the amplicon heterogeneity. Five SNP haplotypes were defined for this locus (tentatively named *IL8R*) and used in twopoint linkage analysis with 63 informative meioses to place *IL8R* 90.3 cM from the centromere of BTA 2 (Table 2). This result indicated that the genes or pseudogenes amplified at this locus are tightly linked. Multipoint analysis suggested that the *IL8R* locus is 0.7 and 1.3 cM telomeric from transition protein (TNP) 1 and natural resistance-associated macrophage protein (NRAMP) 1, respectively (Fig. 2B).

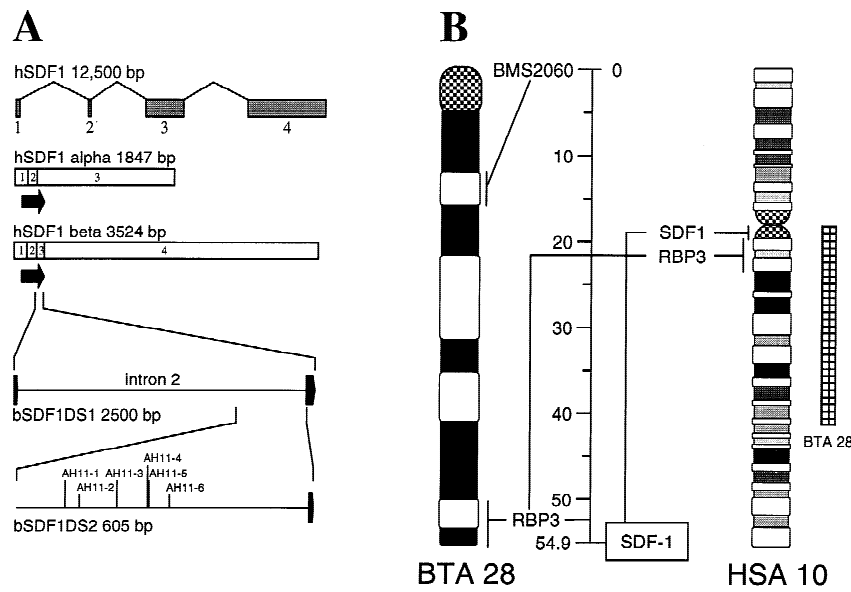
*SNP markers for a bovine SDF1 gene from human DNA sequence.* Because there were no orthologous bovine sequences available for the remaining known human CXC chemokines [PF4, NAP2, IP10, ENA78, GCP2, MIG, and SDF1 (Vaddi et al. 1997)], a different strategy was used to develop markers for a member of this gene family. The human SDF1 gene (Accession No. U16752) was selected to test this strategy because allelic variants in SDF1 have been implicated in resistance to infectious disease (Magierowska et al. 1998; Winkler et al. 1998). Primers were designed to amplify across exons 2 and 3 of the human sequence (Fig. 3A), and the annealing temperature of the PCR was lowered to stabilize imperfect duplex formations. Amplification of bovine genomic DNA produced multiple fragments including one that was close in size to the 2.5-kb human SDF1 amplification product. Sequence analysis of the 3' end revealed that 39 of 41 bp were identical to human SDF1 exon 3 (BLAST search E value  $3e-10$ ). The 5' end of the 2.5-kb fragment did not contain enough exon sequence for comparison with the human SDF1 exon 2. Based on these results, it was concluded that the 2.5-kb bovine amplicon was from an SDF1 gene.

To facilitate the search for SNPs, a more robust pair of bovine-specific amplification primers were developed to produce a 605-bp amplicon that was nested within the 2.5-kb SDF1 intron 2 (Fig. 3A). Six variable sites were observed in this 605-bp region when sequences from 24 founding parents of the reference population were compared (Table 2). Seven SNP haplotype alleles were defined and used in twopoint linkage analysis with 92 informative meioses to place the SDF1 locus on the telomeric end of the BTA 28 linkage group (Fig. 3B). Multipoint analysis suggested that SDF1 is currently the endmost telomeric marker on BTA 28, extending this linkage group 2.9 cM beyond the marker for retinol-binding protein [RBP3 (Crawford et al. 1995)]. In humans, SDF1

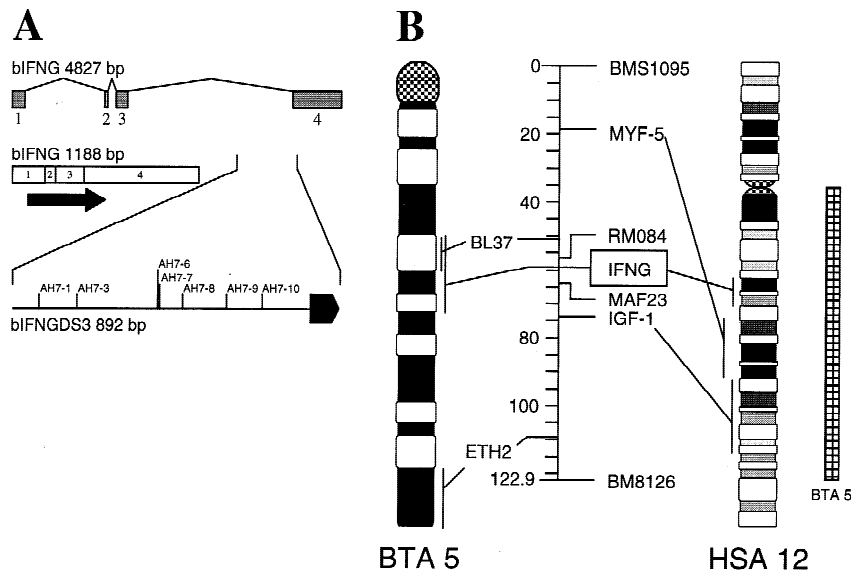
and RBP3 have been mapped to HSA 10q11.1 and 10q11.2, respectively (Farrer et al. 1988; Nakamura et al. 1988; Shirozu et al. 1995). Thus, SDF1 and RBP3 define the location and orientation of a 2.9-cM syntenic region conserved between cattle and humans.

*SNP markers for interferon (IFN)- $\gamma$  from bovine genomic DNA sequence.* *IFN- $\gamma$*  represents a cytokine gene for which bovine genomic sequence is publicly available (Table 1). A primer pair was designed to amplify a 892-bp fragment consisting of sequence from intron 3 and exon 4 of bovine *IFN- $\gamma$*  (Fig. 4A). Sequence comparison of this product identified six SNPs and one 3-bp insertion/deletion site. All seven of these polymorphisms segregated according to Mendelian inheritance patterns in the reference population. Although seven polymorphic sites were observed, only three SNP haplotypes were present (Table 2). Twopoint linkage analysis with 81 informative meioses placed the *IFN- $\gamma$*  locus 59.3 cM from the centromere of BTA 5 (Fig. 4B). Multipoint analysis refined the *IFN- $\gamma$*  position to give the most likely marker order: *RM084*, *IFN- $\gamma$* , *MAF23*. The *IFN- $\gamma$*  locus is currently the only locus linked to BTA 5 and physically assigned to both cattle (Chaudhary et al. 1993) and human chromosomes (Zimonjic et al. 1995). The nearest gene markers flanking *IFN- $\gamma$*  are myogenic factor (MYF) 5 and insulin-like growth factor (*IGF-1*). Both MYF5 and *IGF-1* lie distal to *IFN- $\gamma$*  on the human physical map (Morton et al. 1986; Cupelli et al. 1996), indicating a potential rearrangement of gene order between these segments of BTA 5 and HSA 12.

*SNP markers for the tumor necrosis factor (TNF)- $\alpha$  locus from bovine genomic DNA sequence.* *TNF- $\alpha$*  represents a cytokine gene for which multiple bovine genomic sequences were available. Three 4.3-kb segments of genomic sequence from two *Bos taurus* animals (Accession No. Z14137 and AF011926) and one *Bos indicus* animal (Accession No. AF011927) were compared to identify 37 potential SNPs (data not shown). A primer pair was designed to amplify a 639-bp region 5' to *TNF- $\alpha$* , which contained nine of the 37 predicted nucleotide differences (Fig. 5A). Sequence analysis of the reference population revealed three SNPs that were inherited according to the expected patterns. However, only one polymorphism corresponded to the set of nine predicted SNPs. The other two polymorphisms were not predicted from the comparison



**Fig. 3.** Genetic maps of SDF1 genes and BTA28. The symbol legends are described in Fig. 1. Panel A: Physical maps of human SDF1 genomic and cDNA, and nested bovine SDF1 amplicons. Panel B: Linkage and physical maps of BTA 28 and comparison with HSA 10.



**Fig. 4.** Genetic maps of the bovine *IFN- $\gamma$*  gene and BTA5. The symbol legends are described in Fig. 1. Panel A: Physical maps of bovine *IFN- $\gamma$*  genomic, cDNA, and amplicon from intron 3. Panel B: Linkage and physical maps of BTA 5 and comparison with HSA 12.

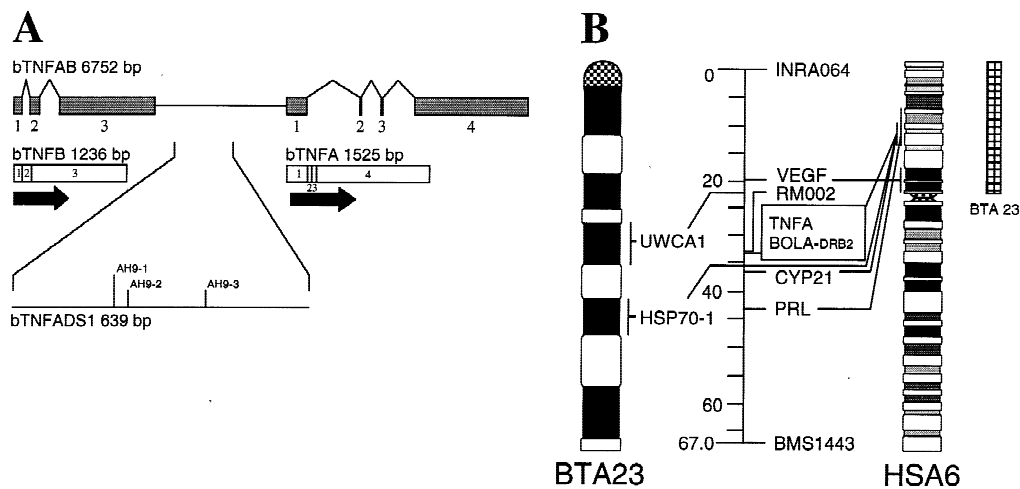
of *TNF- $\alpha$*  sequences in GenBank. Four SNP haplotype alleles were defined and used in twopoint linkage analysis with 91 informative meioses to place the *TNF- $\alpha$*  locus 33.5 cM from the centromere on the BTA 23 linkage group (Table 2). Multipoint analysis indicated that the most likely marker order was: *RM002*, *TNF- $\alpha$* , *BoLA-DRB2* (Fig. 5B).

**Discussion**

Members of the CXC chemokine family are among those genes modulated in the early phase of bovine epithelial cell exposure to *E. coli* endotoxin (Heaton et al. 1999). The chemokines, their receptors, and other members of the cytokine superfamily represent rational gene targets for developing genetic markers to study their association with infection phenotypes. The present results show that comparing bovine DNA sequences from 24 reference animals is an efficient means for discovering SNPs in bovine cytokine loci, as all seven amplicons selected contained sequence variation in the reference population. Searching for bovine SNPs in the public sequences did not increase the SNP discovery rate for

the *TNF- $\alpha$*  locus in the reference population. This is in contrast to the success reported with human expressed sequence tag and genomic DNA data sets (Gu et al. 1998; Buetow et al. 1999). The variable sites identified in public bovine *TNF* sequences were either artifacts of conflicting sequence information or the SNPs were not present in this reference population.

Prediction of gene identity for the GRO3, GRO1, GROX, IL8R, and SDF1 amplicon sequences was based on their comparison with public exon sequence and synteny inferred from human maps. In one example, sequence comparison showed that neither of the two target genes were amplified and, thus, a new designation, GROX, was tentatively assigned. The predictions are provisional because the relationship between amplicon identity and gene identity is ambiguous when only exon sequence is available for primer design. Exon sequences may be highly conserved among gene family members and pseudogenes and, thus, a complete set of sequences for related genes and pseudogene in the mapped region may be required for conclusive identification of some amplicons. Another strategy for increasing the accuracy of assigning amplicon identity is to target known intron/intergenic



**Fig. 5.** Genetic maps of the *TNF- $\alpha$*  gene and BTA23. The symbol legends are described in Fig. 1. Panel A: Physical maps of bovine *TNF- $\alpha$*  and *TNF- $\beta$*  genomic and cDNA, and amplicon from intergenic region between

sequences like those of *IFN- $\gamma$*  and *TNF- $\alpha$* . These regions are less conserved among gene families and therefore likely to be unique within the genome.

A recent report demonstrated that SNPs were present in the bovine CXC chemokine genes *IL8* and *ECIP1* and mapped to BTA 6 (Heaton et al. 1999). Linkage analysis for GRO-like genes presented here indicate that a significant portion of the CXC chemokine cluster appears to be conserved between cattle and humans. The human GRO genes are tightly linked to *IL8* on HSA 4 in a nine-gene cluster of CXC chemokines (Modi and Chen 1998). Our results were also consistent with somatic cell hybrids mapping results in which two bovine GRO genes were localized to BTA 6 (Modi et al. 1998). The placement of *SDF1* on the telomeric end of BTA 28 represents a new bovine assignment, extends the linkage group, and defines a small syntenic region conserved between cattle and human genomes. These loci, together with *IL8* and *ECIP1*, comprise 30 SNPs and 29 SNP haplotype alleles for 6 CXC chemokine genes.

Because developing bovine amplification primers from human sequence was sometimes difficult, bovine genomic DNA sequence made available via public databases (e.g., *IFN- $\gamma$*  and *TNF- $\alpha$* ) was an important resource in the SNP discovery process. *IFN- $\gamma$*  is a potent activator of macrophages and monocytes and greatly augments their antibacterial activities (Boehm et al. 1997). Though the *IFN- $\gamma$*  gene has been localized to BTA 5q22-5q24 by fluorescence in situ hybridization (Chaudhary et al. 1993), genetic markers for linkage analysis have not been described. *TNF- $\alpha$*  is considered a major inflammatory mediator and, in humans, variation in the *TNF- $\alpha$*  gene promoter has been associated with several infectious diseases (Hill 1998; Knight et al. 1999). SNP markers for these genes may be useful for evaluating whether variants are associated with similar traits in cattle.

In addition to determining linkage map position and marker order, analysis of SNP segregation patterns in reference families provided three important results. First, it allowed definition of SNP haplotype alleles within an amplicon. Unlike the individual SNPs that are biallelic, the set of SNP alleles constituting a haplotype is not limited to two alleles. The *SDF1* amplicon, for example, contained six SNP haplotypes in the reference population. The use of SNP haplotypes in linkage analysis typically reduced the number of cases in which parents and progeny had identical heterozygous genotypes. These cases (termed "switches") present computational difficulties for linkage analysis algorithms like CRI-MAP. A second advantage of analyzing segregation patterns in reference fami-

*TNF- $\alpha$*  and *TNF- $\beta$* . Panel B: Linkage and physical maps of BTA 23 and comparison with HSA 6.

lies was the detection of apparent heterozygous sites that result from sequencing "artifacts". These artifacts were occasionally apparent even when a high-quality sequence was obtained from both DNA strands. The artifacts were revealed when non-Mendelian inheritance patterns were observed in reference progeny. A third benefit of segregation analysis was the discovery of unamplified alleles in certain families. This phenomenon caused animals with heterozygous haplotypes to appear homozygous. Unamplified alleles are postulated to occur when SNPs were present in the primer binding region. Mismatches between the primer and template may destabilize DNA duplex formation and reduce the amplification of specific haplotype alleles. For these reasons, it is imperative that SNP haplotype loci be extensively evaluated in families with defined pedigrees prior to their application in populations with poorly defined genetic backgrounds.

**Acknowledgments.** We thank Dr. Timothy P.L. Smith for critical comments regarding this manuscript; James Wray for integration of SNP information with the MARC livestock database; Stacy McHale for technical assistance in marker development; Stephen Simcox, Linda Flathman, and Robert Lee for technical assistance in genomic DNA preparation and DNA sequencing; and Joan Rosch and Sherry Kluver for secretarial assistance. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

## References

- Altare F, Durandy A, Lammas D, Emile JF, Lamhamedi S et al. (1998) Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* 280, 1432-1435
- Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SL et al. (1994) A genetic linkage map for cattle. *Genetics* 136, 619-639
- Boehm U, Klamp T, Groot M, Howard JC (1997) Cellular responses to interferon-gamma. *Annu Rev Immunol* 15, 749-795
- Buetow KH, Edmonson MN, Cassidy AB (1999) Reliable identification of large numbers of candidate SNPs from public EST data. *Nat Genet* 21, 323-325
- Chaudhary R, Chowdhary BP, Johansson M, Gustavsson I (1993) The gene for bovine interferon gamma (IFNG) maps to the q22-q24 bands of chromosome 5 in cattle. *Hereditas* 119, 11-14
- Chowdhary BP, Fronicke L, Gustavsson I, Scherthan H (1996) Comparative analysis of the cattle and human genomes: detection of ZOO-FISH and gene mapping-based chromosomal homologies. *Mamm Genome* 7, 297-302

- Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW et al. (1995) An autosomal genetic linkage map of the sheep genome. *Genetics* 140, 703–724
- Cupelli L, Renault B, Leblanc-Straceski J, Banks A, Ward D et al. (1996) Assignment of the human myogenic factors 5 and 6 (MYF5, MYF6) gene cluster to 12q21 by in situ hybridization and physical mapping of the locus between D12S350 and D12S106. *Cytogenet Cell Genet* 72, 250–251
- Farrer LA, Castiglione CM, Kidd JR, Myers S, Carson N et al. (1988) A linkage group of five DNA markers on human chromosome 10. *Genomics* 3, 72–77
- Francke U (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenet Cell Genet* 65, 206–218
- Glynn MK, Bopp C, Dewitt W, Dabney P, Mokhtar M et al. (1998) Emergence of multidrug-resistant *Salmonella enterica* serotype *typhimurium* DT104 infections in the United States. *N Engl J Med* 338, 1333–1338
- Green P, Falls K, Crooks S (1990) Documentation for CRI-MAP version 2.4 (St. Louis, MO: Washington University School of Medicine)
- Gu Z, Hillier L, Kwok PY (1998) Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat* 12, 221–225
- Heaton MP, Laegreid WW, Beattie WW, Smith TPL, Kappes SM (1999) Identification and genetic mapping of bovine chemokine genes expressed in epithelial cells. *Mamm Genome* 10, 128–133
- Hill AV (1998) The immunogenetics of human infectious diseases. *Annu Rev Immunol* 16, 593–617
- Holmes WE, Lee J, Kuang WJ, Rice GC, Wood WI (1991) Structure and functional expression of a human interleukin-8 receptor. *Science* 253, 1278–1280
- Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS et al. (1997) A second-generation linkage map of the bovine genome. *Genome Res* 7, 235–249
- Knight JC, Udalova I, Hill AV, Greenwood BM, Peshu N et al. (1999) A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. *Nat Genet* 22, 145–150
- Krakauer T, Vilcek J, Oppenheim JJ (1999) Proinflammatory cytokines. In *Fundamental Immunology*, WE Paul, ed. (Philadelphia: Lippincott-Raven), pp. 775–811
- Lenstra JA, van Boxtel JA, Zwaagstra KA, Schwerin M (1993) Short interspersed nuclear element (SINE) sequences of the *Bovidae*. *Anim Genet* 24, 33–39
- Lloyd A, Modi W, Sprenger H, Cevario S, Oppenheim J et al. (1993) Assignment of genes for interleukin-8 receptors (IL8R) A and B to human chromosome band 2q35. *Cytogenet Cell Genet* 63, 238–240
- Magierowska M, Lepage V, Boubnova L, Carcassi C, de Juan D et al. (1998) Distribution of the CCR5 gene 32 base pair deletion and SDF1-3'A variant in healthy individuals from different populations. *Immunogenetics* 48, 417–419
- Martin MP, Dean M, Smith MW, Winkler C, Gerrard B et al. (1998) Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science* 282, 1907–1911
- Mead PS, Griffin PM (1998) *Escherichia coli* O157:H7. *Lancet* 352, 1207–1212
- Modi WS, Amarante MR, Hanson M, Womack JE, Chidambaram A (1998) Assignment of the mouse and cow CXC chemokine genes. *Cytogenet Cell Genet* 81, 213–216
- Modi WS, Chen ZQ (1998) Localization of the human CXC chemokine subfamily on the long arm of chromosome 4 using radiation hybrids. *Genomics* 47, 136–139
- Mollereau C, Muscatelli F, Mattei MG, Vassart G, Parmentier M (1993) The high-affinity interleukin 8 receptor gene (IL8RA) maps to the 2q33-q36 region of the human genome: cloning of a pseudogene (IL8RBP) for the low-affinity receptor. *Genomics* 16, 248–251
- Morris SW, Nelson N, Valentine MB, Shapiro DN, Look AT et al. (1992) Assignment of the genes encoding human interleukin-8 receptor types 1 and 2 and an interleukin-8 receptor pseudogene to chromosome 2q35. *Genomics* 14, 685–691
- Morton CC, Byers MG, Nakai H, Bell GI, Shows TB (1986) Human genes for insulin-like growth factors I and II and epidermal growth factor are located on 12q22–q24.1, 11p15, and 4q25–q27, respectively. *Cytogenet Cell Genet* 41, 245–249
- Murphy PM (1997) Neutrophil receptors for interleukin-8 and related CXC chemokines. *Semin Hematol* 34, 311–318
- Nakamura Y, Lathrop M, Bragg T, Leppert M, O'Connell P et al. (1988) An extended genetic linkage map of markers for human chromosome 10. *Genomics* 3, 389–392
- NASS (1996) *Cattle and Calves Death Loss 1995* (Washington, D.C.: U.S. Department of Agriculture, Animal and Plant Health Inspection Service, National Agricultural Statistics Service)
- Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG et al. (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 19, 233–240
- Shirozu M, Nakano T, Inazawa J, Tashiro K, Tada H et al. (1995) Structure and chromosomal localization of the human stromal cell-derived factor 1 (SDF1) gene. *Genomics* 28, 495–500
- Smith MW, Dean M, Carrington M, Winkler C, Huttley GA et al. (1997) Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study. *Science* 277, 959–965
- Solinas-Toldo S, Lengauer C, Fries R (1995) Comparative genome map of human and cattle. *Genomics* 27, 489–496
- Sprenger H, Lloyd AR, Lautens LL, Bonner TI, Kelvin DJ (1994) Structure, genomic organization, and expression of the human interleukin-8 receptor B gene. *J Biol Chem* 269, 11065–11072
- Thomson AW (1998) *The Cytokine Handbook*, 3rd ed. (San Diego: Academic Press)
- Turker MS, Cooper GE, Bishop PL (1993) Region-specific rates of molecular evolution: a fourfold reduction in the rate of accumulation of "silent" mutations in transcribed versus nontranscribed regions of homologous DNA fragments derived from two closely related mouse species. *J Mol Evol* 36, 31–40
- Vaddi K, Keller M, Newton RC (1997) *The Chemokine Factsbook* (San Diego: Academic Press)
- Winkler C, Modi W, Smith MW, Nelson GW, Wu X et al. (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant. *Science* 279, 389–393
- Zimonjic DB, Rezanka LJ, Evans CH, Polymeropoulos MH, Trent JM et al. (1995) Mapping of the immune interferon gamma gene (IFNG) to chromosome band 12q14 by fluorescence in situ hybridization. *Cytogenet Cell Genet* 71, 247–248