

Characterization and radiation hybrid mapping of expressed sequence tags from the canine brain

Monica C. Roberts,¹ Christophe Hitte,² Julie A. Hendrickson,¹ Daniel E. Hoffmann,¹ Gail H. Flickinger,¹ Mark S. Rutherford,¹ Richard Guyon,² Francis Galibert,² and James R. Mickelson¹

¹College of Veterinary Medicine, Department of Veterinary PathoBiology, University of Minnesota, 1988 Fitch Avenue, 295 Animal Science/Veterinary Medicine, St. Paul, Minnesota 55108, USA

²UMR6061 CNRS, Génétique et Développement, Faculté de Médecine, 35043 Rennes Cedex, France

Received: 12 June 2002 / Accepted: 6 November 2002

Abstract

Maps of the canine genome are now developing rapidly. Most of the markers on the current integrated canine radiation hybrid/genetic linkage/cytogenetic map are highly polymorphic microsatellite (type II) markers that are very useful for mapping disease loci. However, there is still an urgent need for the mapping of gene-based (type I) markers that are required for comparative mapping, as well as identifying candidate genes for disease loci that have been genetically mapped. We constructed an adult brain cDNA library as a resource to increase the number of gene-based markers on the canine genome map. Eighty-one percent of the 2700 sequenced expressed sequence tags (ESTs) represented unique sequences. The canine brain ESTs were compared with sequences in public databases to identify putative canine orthologs of human genes. One hundred nine of the canine ESTs were mapped on the latest canine radiation hybrid (RH) panel to determine the location of the respective canine gene. The addition of these new gene-based markers revealed three conserved segments (CS) between human and canine genomes previously detected by fluorescence *in situ* hybridization (FISH), but not by RH mapping. In addition, five new CS between dog and human were identified that had not been detected previously by RH mapping or FISH. This work has increased the number of gene-based markers on the canine RH map by approximately 30% and indicates the benefit to be gained by increasing the gene content of the current canine comparative map.

Introduction

The domestic dog is a unique species, consisting of over 300 recognized breeds, which vary widely in physical appearance and behavior (Wilcox and Walkowicz 1995). Unfortunately, an increased prevalence of heritable disorders has accompanied the selection for desirable traits in many breeds. Because of the short generation interval, a large number of full-sib offspring, and a favorable population structure resulting from high founder contributions, it is likely that the genetic basis for many of the over 350 reported genetic diseases in the dog could eventually be defined. Recent successes in the mapping of canine diseases include loci for progressive rod-cone degeneration (Acland et al. 1998), narcolepsy (Lin et al. 1999), multifocal renal cystadenocarcinoma (Jonasdottir et al. 2000), malignant hyperthermia (Roberts et al. 2001), and copper toxicosis (van de Sluis et al. 2002). Many simple heritable canine diseases share clinical features with genetically complex human diseases (Patterson 1980, 2000). Thus, in addition to positively influencing the health and well-being of dogs, analyzing canine models is likely to develop valuable information of use in understanding the basis and treatments for a number of complex human conditions (Ostrander and Giniger 1997).

Resources are becoming available with which to study the structural, comparative, and functional genomics of the dog (Breen et al. 1999; Mellersh et al. 1997, 2000; Neff et al. 1999; Priat et al. 1998; Vignaux et al. 1999; Yang et al. 1999). The most recent integrated radiation hybrid/genetic linkage/cytogenetic map of the dog genome contains 1,800 markers (Breen et al. 2001). Over 1,000 of these markers are polymorphic microsatellites that are extremely use-

Correspondence to: J. R. Mickelson; E-mail: micke001@umn.edu

ful in genetic linkage analyses. However, since the canine genome map is still relatively sparse, the physical proximity of most microsatellites to coding genes is not yet known. The assignment of 320 gene-based markers on the integrated map (Breen et al. 2001) has identified 64 large conserved segments (CS), in which the gene content of human and dog chromosomes is expected to be very similar. Nevertheless, a number of putative CS have not been precisely defined and need to be confirmed. Thus, an increased representation of protein-encoding genes on the canine genome map is necessary to provide researchers with the tools to determine what candidate genes lie in a region defined by a microsatellite marker linked to a trait of interest, and to use the dense human or rodent genome maps to full advantage in selecting candidate genes for the canine condition.

Breed-specific behavioral variation, many behavioral disorders, and a number of neurological diseases whose basis lies in abnormal regulation of central nervous system (CNS) activity are all likely to have a strong genetic component (Patterson 1980, 2000). Possible bases for genetic variation in CNS function within and between breeds of dogs include different developmental patterns of gene expression of CNS genes, different levels of gene expression within and between regions of the CNS, and polymorphisms (SNPs) within specific genes. An increasingly common means of rapidly deriving sequence information from expressed genes, and ultimately investigating regional and temporal issues in gene expression, is through the sequencing of large numbers of cDNAs, also known as expressed sequence tags (ESTs). The present limitations in EST resources for the dog will hamper efforts to define the molecular basis for genes affecting canine health, developmental biology, and behavioral biology. We report here the properties of a canine brain cDNA library used to generate several thousand ESTs from this tissue, the placement of over 100 of these ESTs on the current canine RH map (Breen et al. 2001), and the identification of several confirmed and putative areas of CS between the human and canine maps. This work significantly increases the number of coding genes on the RH map and enhances the comparative map by increasing the number of suggestive and confirmed conserved segments between the dog and humans.

Materials and methods

cDNA library construction. The brain of a normal adult dog was rapidly dissected into frontal, occipital, temporal and parietal lobes, olfactory bulb, hip-

pocampus, cerebellum, thalamus, hypothalamus, midbrain, pons, and medulla. The tissues were flash frozen in liquid nitrogen and stored at -80°C . 250 mg of each brain region was homogenized in a Brinkman Polytron with 3.5 mL of Trizol reagent (Gibco BRL). The resultant homogenate was pooled for total RNA preparation, followed by poly A⁺ RNA isolation by using oligo (dT) cellulose columns (Amersham Pharmacia). First-strand cDNA was synthesized with Gibco Superscript II RNase H-reverse transcriptase, and primed with an oligo (dT) primer that contained a 5' *NotI* site. Following second-strand synthesis, cDNA was ligated to a *SaII* adapter. The cDNAs were then digested with *NotI*, size-fractionated by column chromatography, and eluted fractions containing cDNA greater than 400 bp in length were selected. The adaptor-ligated cDNAs were directionally ligated into pSPORT1 plasmid vector (Gibco BRL).

DNA sequencing, EST editing, and database searches. Plasmid DNA was submitted for 5' end sequencing with M13 reverse primer to the Advanced Genetic Analysis Center (AGAC) at the University of Minnesota. Many clones of interest were also sequenced by using the M13 forward primer to obtain sequence from the 3' end. Sequences were trimmed to remove the pSPORT1 vector sequence and were edited by using the standard criteria within the Sequencher program (GeneCodes Corp). Here, the 5' end was trimmed until the first 25 bases contained fewer than four ambiguities; 3' end scanning then started 100 bases from the 5' end, and the sequence was trimmed at the first 25 bases containing more than four ambiguities; the leading and trailing ambiguous bases were removed. Redundant clones and clones with overlapping sequences were also identified by assembling the sequences into contigs with Sequencher. Sequence quality was assessed using Phred software (Ewing et al. 1998). The ESTs were submitted to the NCBI dbEST database (<http://www.ncbi.nlm.nih.gov:80/dbEST/>), were assigned accession numbers, and were periodically compared with other sequences in the database using both BLASTN and BLASTX software (Altschul et al. 1990) (last search date May 24, 2002). Selected EST sequences that revealed significant matches to putative orthologous human genes with *p*-values of less than or equal to e^{-10} were analyzed for PCR primer design to attempt placement on the canine RH map.

Selected PCR amplicons were sequenced to confirm their identity. Such PCR reactions were performed using 150 ng of canine genomic DNA, 22.5 pmol of each primer, 2.5 nmol each dNTP, 1.5 mM

MgCl₂, and 1 U Hot Star Taq DNA polymerase (Qiagen) in a final volume of 36 µL. Cycling conditions were 95°C for 20 min, followed by 35 cycles of 30 s at 94°C, 30 s at the primer pairs annealing temperature, and 30 s at 72°C with a final extension of 72°C for 5 min. The PCR products were purified by using the Qiaquick PCR purification kit (Qiagen) and were submitted to the AGAC for sequencing with 10 pmol of the forward EST primer used in the initial reaction.

Radiation hybrid mapping. Primers for selected ESTs were designed with the PRIME program contained within the GCG software package [Wisconsin Package Version 10.2, Genetics Computer Group (GCG), Madison, WI]. Sequence information from human, and rodent if available, was used to determine regions that are not highly conserved across species. The sequences from the 3' untranslated regions of the genes were most often used for primer design. Each primer pair was first optimized on dog and hamster DNA, as well as 1:3 dog:hamster (wt:wt) DNA admixtures to choose the annealing temperature that reproducibly amplified dog DNA without an interfering product from the hamster. Genotyping was performed on the 118 radiation hybrid cell lines from the RHDF-5000.2 panel (Vignaux et al. 1999). PCR reactions were performed in GeneMate Genius 96 well format thermocyclers (ISC) with 50 ng DNA from each clone in a final volume of 15 µL containing 10 pmol of each primer, 67 µM dNTPs, 1.5 mM MgCl₂, and 0.35 U Hot StarTaq DNA Polymerase (Qiagen). PCR cycles were 94°C for 15 or 20 min, followed by 30 cycles of 30 s at 94°C, 30 s at 54–60°C, 30 s at 72°C, and finally 5 min at 72°C. Primer sequences and annealing temperatures for each marker are provided in Table 1. PCR products were electrophoresed through 2% or 3% TBE agarose gels and visualized with ethidium bromide staining under UV light. Genotyping of all EST markers was duplicated to ensure data integrity. Only concordant data were subsequently used in the RH analysis.

The statistical analysis for placement of each marker was done with the whole RH data set by using the MultiMap software (Matise et al. 1994). Assignment of new markers to chromosomes was done by two-point linkage analysis in a two-step process. First, ESTs were assigned to chromosomes by pairwise calculation at lod score higher than 8. Second, within chromosomes, the best lod score between any new marker and mapped markers was determined. Positions within the chromosome were determined by considering the best lod score value with mapped adjacent markers.

Results

cDNA sequences. Plasmid DNA from 3332 clones was isolated. The average insert size, estimated from restriction digests of 24 clones, was approximately 2 kb. 2693 sequences, with average length of 525 bp, remained following trimming and editing of the 5' end sequence data. Post-trimming, 91% of the bases had Phred quality values greater than or equal to 20, and 54% of the bases had Phred quality values greater than or equal to 40. Of the retained sequences, 726 could be assembled into 223 contigs of overlapping and redundant sequences, leaving the remaining 1967 sequences as singletons. Therefore, there are 2190 distinct EST sequences represented in this set of randomly selected clones.

BLAST comparisons of the 2693 edited sequences to the NCBI databases found that 2161 matched to genes, ESTs, or clones (with a significance level less than e^{-10}), while 532 had no significant match. Of the sequences with significant matches, approximately 13% were derived from genes of the mitochondrial genome, 16% had the most significant match to unannotated human genomic DNA clones, and 17% had the most significant match to novel EST or cDNA sequences. The remaining canine ESTs are putative matches to specified genes. Approximately 36% of the latter canine ESTs matched ESTs from primate or mouse brain cDNA libraries, while others matched ESTs from libraries of a variety of other tissues, including, most commonly, lung, skin, kidney, and testis. We concluded that this canine brain cDNA library was of sufficient quality to use in the identification of many new EST sequences and to enable the placement of new gene markers on the canine RH map.

Radiation hybrid mapping. The RH scores of 109 canine brain ESTs were determined on the RHDF-5000.2 panel, and the data were submitted to RH analysis on the existent dog 1500 marker RH map (Breen et al. 2001). The average retention for all 109 ESTs was 0.23, corresponding to the average retention of all markers genotyped on this panel (Priat et al. 1998). The clone name and accession number, most likely human orthologous gene name, best lod score values, nearest linked marker, and chromosomal assignments for each new EST marker are reported in Table 1. Ninety-seven ESTs were assigned to chromosomes by pairwise calculation at lod score higher than 8.0, while 12 (11%) markers remained unlinked to an RH linkage group. This percentage of unlinked markers is in the range observed in other RH mapping studies by using this panel (Breen et al. 2001; Priat et al. 1998).

Table 1. Canine brain ESTs assigned to the canine RH map. All new ESTs with mapping information are presented, along with GenBank accession numbers, primer sequences and annealing temperatures, closest human match with significance (BLASTN), location of human match, and canine location with LOD score. Markers are sorted by their mapped canine chromosome location. All adjacent markers may be retrieved at "http://www-recomgen.univ-rennes1.fr/Dogs/maquette-1800.html". RH groups noted as *unassigned group* correspond to orphan RH linkage groups denoted RH01 to RH17 that have not yet been assigned to chromosomes (Breen et al. 2001).

Marker name	Accession number(s)	Primer sequences	Temp	Gene with closest matching sequence homology	Gene symbol	p-value of match	Human location	Canine location	Adjacent marker	LOD score
UMN51A03	BI398192 BQ426191	TCCAACGAGATGGATGGTG TGTTCAACAGGAAAGGCAAG	56	Heat shock transcription factor 2	HSF2	e-172	6q22.33	CFA01	C02509	17.518
UMN45G05	BI398127	TGGAGAAAAGAGAGAAGAG GTTCCCAACCTTCCAATC	58	Carbonic anhydrase XI	CAL1	4.E-26	19q13.3	CFA01	CAN2DD	15.162
UMN12E11	BQ280326 BQ426207	TTTATTGGCGTAAATGCAACC TCCCCTCTCTCTCTTC	56	Arginine methyltransferase	HRMR	1L2 e-120	19q13.3	CFA01	FH2326	11.701
UMN39B12	BI430427	CAACTCTTCCCGTGCC GCTGTGAGCTTGGTGAATC	56	Apolipoprotein E	APOE	2.E-49	19q13.2	CFA01	FH2598	15.902
UMN43E11	BI398067 BQ426223	CATAAAATAGCTGGGAAGGG AAGGAGGTCATCAGGGAGG	58	Protein phosphatase 5, catalytic subunit	PPP5C	0.E + 00	19q13.3	CFA01	FH2598	14.602
UMN44H01	BI421111	GCAACCAAGCCTGTATTC AACACCAACTCTCAATCCC	58	Calmodulin 3 (phosphorylase kinase)	CALM3	1.E-31	19q13.2-13.3	CFA01	FH2598	15.775
UMN43C09	BI398109	CGAACAAAGCGAACCGG CATCCATGTCATGTGCCCTC	58	Phosphohexose isomerase	GPI	e-154	19q13.1	CFA01	REN143K19	9.019
UMN39G03	BI397483	GGAAATTGCACATTCAAAG CAGATCCATGACAAATGCAC	56	Protein tyrosine phosphatase receptor type k	PTPRK	e-137	6q22.2-23.1	CFA01	REN97F15	14.375
UMN40H05	BI430492 BQ426222	TGTGTACTCCAGAGAGCC CCCATCTTAATCCGATGTCC	56	Myelin basic protein	MBP	e-105	18q23	CFA01 ^c	REN303I04	20.372
UMN52D03	BI405275	TGTGTAGCCGTAGCAAG ACGGAGAGAAAGCACTTAG	56	Splicing factor, arginine/serine-rich 4	SFRS4	e-150	1pter-p33	CFA02	EST-CFZ97733	19.157
UMN48G04	BI398186	ACCATCCCTGACCTGTCC CAGCCCTCCAAATATTGTCC	56	Lysophospholipase II	LYPLA2	e-116	1p36.12-1p35.1	CFA02	FUCA1	15.919
UMN44E07	BI421091	TTACACCGGACCAATCAG CAGCCAGATAAAGGAGAAG	56	6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3	PFKFB3	3.E-11	10p14-10p15	CFA02	REN303H07	17.421
UMN45H03	BI398128	GGGGCCGTGTATTATTG TGAACGTCTACTAGGAG	58	Similar to aryl hydrocarbon receptor nuclear translocator 2	ARNT2	9.E-19	15q24	CFA03	AHTH219REN	14.171
UMN41E11	BI430509	TGTGTCCAGCAAGCAC CAAACAGCGATGCAATGCACAG	56	Calcium and integrin binding protein	CIB1	2.E-68	15q25.3-15q26	CFA03	REN273H17	21.26
UMN44D01	BI421082 BQ426225	TCAGAAAGCGTACAGATAG CCGTCAAAGAATGGACAATCAG	56	WD repeat domain protein 1	WDR1	0.E + 00	4p15.32	CFA03 ^d	REN153P03	16.259
UMN51H09	BI398246	GACAGACTGACATGAACCAC TCCTTTCTTTTCTGTCTCC	56	Cyclin G associated kinase	GAK	1.E-20	4p16	CFA03 ^d	REN73P04	14.441
UMN48H11	BI398166 BQ426198	AGTTTGGAGGATGAAATGGG GTCATGTCAGAACACGACAG	56	CD74 antigen	CD74	9.E-99	5q32	CFA04	EST-CFZ97828	12.722
UMN30F02	BI398230 BQ426203	CTCACACAGTCCACAC CTTCTTTTACCAATCCCTTTC	58	FAT tumor suppressor homolog 2	FAT2	e-103	5q32-q33	CFA04	REN156O17	9.427
UMN43C08	BI398058	ACTTCCCAATCATGAAAGCC TCGCCGAAACTAATGCC	56	Sodium-dependent glutamate transporter	SLC1A3	0.E + 00	5p13	CFA04	REN204I06	18.394
UMN52A10	BI405240	AGTCCAGATGCACCCAAAG AGCCCAACAAGCCCAAG	56	Cyclin G1	CCNG1	e-155	5q32-5q34	CFA04	REN73D12	14.651
UMN44H07	BQ280364	TGACGAGAGAAATGAGAAGAC AAGGAGATCCCCCAGCAAG	58	FXYD domain-containing ion transport regulator 6	FXYD6	2.E-86	11q23.3	CFA05	THY-1	14.669
UMN48C04	BI398177 BQ426200	CCTCAGCAATTACAGGTAAC ACCGTGTCCAAAGGAGAAAC	58	Protein kinase C, zeta	PRKCZ	e-107	1p36.33-36.2	CFA05 ^d	REN137C07	15.592
UMN31B11	BI396005	CTGACCCGTAGTAAGCAATG TTACACCTTGGACAGCAAAAC	56	Na ⁺ /K ⁺ ATPase beta subunit	ATP1B1	e-117	1q22-q25	CFA07	REN04I02	16.616

UMN10C02	BM735734	TCAGCTTTCACCAAAAGGTTAG	56	Chromogranin A (parathyroid secretory protein 1)	CHGA	3.E-24	14q32	CFA08	AHTH240REN	13.84
UMN44E08	BQ426211	TTCCCTCTGCTCCAAATTC	56	Protein phosphatase 2, regulatory subunit B (B56), gamma isoform	PPP2R5C	0.E + 00	14q32.2	CFA08	COS8	20.345
UMN36F08	BQ426226	GTCCAGTGTGCTTTTC	56	Kinesin 2	KNS2	4.E-94	14q32.3	CFA08	IGHA	21.979
UMN45D06	B1397412	GCTGTGTTAGCAATCCCC	58	Eukaryotic translation Initiation factor 5	EIF5	3.E-29	14q32.33	CFA08	IGHA	17.025
UMN38B08	B1398139	GATAAGAACCCCCAAAAG	58	Reticulon 1	RTN1	5.E-52	14q21-q22	CFA08	REN68M10	15.517
UMN10A06	B1397427	TTCCATGCTGAGGTGAC	54	Pyruvate dehydrogenase kinase, isoenzyme 2	PDK2	1E-24	17q23.3	CFA09	COL1A1	9.922
UMN10F01	BQ426210	TGCACATGCTCCAAAGAC	54	Prostaglandin D synthase	PTGDS	5e-69(dog) 9e-6(hum)	9q34.2- q34.3	CFA09	DTMT	15.367
UMN39A02	BQ426212	TCTGCTTTTCCCTGAGAC	56	Calcium channel voltage-dependent L type alpha 1B subunit	CACNA1B	e-117	9q34	CFA09	DTMT	15.423
UMN41D11	BQ430416	TGAATTGACAAAAGGGAAG	56	Glial fibrillary acidic protein	GFAP	7.E-26	17q21	CFA09	KRT17	10.25
UMN52D12	BQ426184	AGAGGAGAAGTAGAGGAG	56	Hypothetical protein similar to dopamine and cAMP regulated neuronal phosphoprotein	FLJ20940	4.E-66	17q12	CFA09	RARA	14.26
UMN39H04	B1430443	TGGTTFAGAAAACAGCAG	56	Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)	SPTAN1	3.E-92	9q33-q34	CFA09	REN209L11	17.2
UMN36A05	B1397384	AAAAATGAGGGGGGAGAAC	58	Protein tyrosine phosphatase Cr. 1 PTPase precursor	PTPRR	1.E-42	12q13.2- q13.3	CFA10	C10.781	18.597
UMN51E11	B1398229	GTGTGGCAGATAGTCAATACAG	60	Potassium channel, large conductance, calcium-activated, subfamily M, beta member 4	KCNMB4	7.E-71	12q	CFA10	C10.781	14.904
UMN10F07	BQ280366	TAAGAGTTGACCCCTTAGCC	58	Glutamate receptor, metabotropic 4	GRM4	2.E-11	6q21.3	CFA12	DLADQA	20.967
UMN30D03	B1392949	ACACATCTACCTACAGACCC	58	Glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein	GRINA	e-146	8q24.3	CFA13 ^d	AHTH239	15.645
UMN39C10	BQ280362	CTCCACGCCCCAGATCAC	56	Cytochrome C1	CYCI	0.E + 00	8q24.3	CFA13 ^d	C13.365	13.446
UMN51C03	BQ426185	ATCCAGTGTGAAACCGAAG	58	Guanylate cyclase soluble 70-kDa subunit	GUC	3.E-27	4q31	CFA15	FH2278	20.388
UMN44B12	B1421074	CTCCAGTGTGAAACCGAAG	54	N-myc downstream-regulated gene	NDRG2	1.E-89	14q11.2	CFA15 ^{a,b}	REN06C11	18.522
UMN51F10	B1398236	TTTACCCCTGTGCTACCCC	56	Epidermal growth factor receptor pathway substrate 15	EPH5	2.E-67	1p32	CFA15 ^c	AHTH257	14.727
UMN39G02	B1430438	CCTCCCAACCAATCAAAATC	58	AMP kinase gamma 2 subunit	PRKAG2	e-117	7q35-q36	CFA16	HUEST- D59484	27.57
UMN41D02	B1430499	CCCTTGAGGAGAACCTGAAC	56	Neuronal membrane glycoprotein M6A	GPM6A	0.E + 00	4q34	CFA16 ^{a,b}	SPC22	16.725
UMN44D05	B1421107	GCAAGCCTGAATCTTCAAACAC	56	N-acylsphingosine amidohydrolase (acid ceramidase)	ASAH	e-113	8p22-p21.3	CFA16 ^d	REN138D05	18.227
UMN44C01	BQ426183	TTTGTCAGGTGATTCATAAG	56	B-cell CLL/lymphoma 9	BCL9	4.E-42	1q21	CFA17	IVL	18.27
UMN40F10	B1430482	CTTTCACCTTTCACAGAACC	58	Chromosome 11 open reading frame 2, similar to angiopoitin	C11orf2	0.E + 00	11q13	CFA18	C18.460	24.653
UMN43B12	BQ426221	CTTTCATGCACAGAAATGAC	60	Apoptosis Inhibitor 5	API5	e-115	11p12-q12	CFA18	COS18	20.241
UMN30G11	B1398106	CTCTGCTTTCCTCTCTTTC	58	Sotute carrier family 25, member 5	SLC25A5	e-150	Xq24-q26 ^e	CFA18	REN249N22	11.742
	BQ426204	TTTAGCATCTGTTCCACTCCC								
		AGGGGATGGGAGGTTGGG								
		ATTCGGAGCGTGGTAGCC								
		TTTGGGATTTAATAAACCCAGG								
		TGAAAGTGCAAAAAGCAAC								
		GAGTCCATCATCATGGAATG								
		CCTAGTTTTCCCTGTGAAC								

Continued on next page

Table 1. (Continued)

Marker name	Accession number(s)	Primer sequences	Temp	Gene with closest matching sequence homology	Gene symbol	p-value of match	Human location	Canine location	Adjacent marker	LOD score
UMN39D05	BI397472	GCCCTATAAAAAGAAAAAGCC AAGCTGAGCATTCCTCAACC	56	Neuronal cell adhesion molecule (NRCAM)	NRCAM	2.E-44	7q31.1-q31.2	CFA18	REN249N22	23.3
UMN48D06	BI398181	ACTGCACTGCCACATTC AGTTTCCTGTTCCCTGTC	58	Diacyl glycerol kinase zeta	DGKZ	e-104	11p11.2	CFA18	REN50L03	17.242
UMN51H04	BI398244 BQ426192	AGGGAAGAGCGAGAAAAAG TGGAATGCCATCTGGGAAC	58	Reticulon 3	RTN3	e-108	11q13	CFA18	REN64K04	12.463
UMN38G06	BI397455 BQ426202	ATCAATCACAGTAGACACTGG TTGCTTTTGGTGTAGAGTGAAG	54	Mitogen-activated protein kinase kinase 1 interacting protein 1	MAP2K1IP1	e-165	15q22.1-q22.3	CFA18 ^a	REN266I17	11.313
UMN41C08	BI430496 BQ426190	TTCCAACTCCAGACTGAC TTTCCCTCCCTTAAAAAC	58	Vacuolar protein sorting 41	VPS41	e-130	7p14-p13	CFA18 ^d	REN106H23	20.883
UMN48D07	BI398154	AATAATCTAGCTTCAACTCA GCTGTAAACCTCTTAGTGG	54	Aspartyl aminopeptidase	DNPEP	5.E-46	2q35	CFA19 ^d	REN306J16	16.101
UMN40F05	BQ280363	AAATCCGCACCTTTCCTGG TGGTAGTGGCATTTTAATGGC	56	Mannosidase alpha B (lysosomal)	MAN2B1	2e-11 (cat)	19cen-q13.1	CFA20	ACP5	17.482
UMNA1-3	BQ280365	GCTGTGTTTATTTACAGCGAG TGTGAGCTGTAAAGAAAGCC	56	Glutaminyl tRNA synthetase	QARS	e-121	3p21.3-3p21.1	CFA20	COL7A1	18.344
UMN30A02	BI389263	CCAACTTACCCAAAACCACC ACACCCGAAAGGAGACC	58	Plasma membrane Ca ²⁺ -ATPase isoform 2	ATP2B2	3.E-77	3p26-p25	CFA20	REN119P03	17.945
UMN43B11	BI1398090 BQ426195	CGACCCGAGAATAAACTTAC GGAAAGGAAAGACAAAGCAG	56	Cell division cycle 34	CDC34	e-134	19p13.3	CFA20	REN293N22	10.398
UMN39D06	BI397473 BQ426217	TAGCAGTAAACGGGTAGCAG ATGACTAGCCTCTGCCCG	58	SWI/SNF-related matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 4	SMARCA4	0.E + 00	19p13.2	CFA20	REN88D22	16.033
UMN30C06	BI389272	TCCTCTTCCAGCAGGTCC TGGAATAGCATCCAGTCCAG	56	HCF-binding transcription factor zhangfei	ZF	e-130	11q14	CFA21	REN252J02	25.58
UMN38A11	BI397424 BQ426205	GCAAAATCACTTACCAGCAAC GGCACACACATTAGGATGTT	58	Dickkopf homolog 3	DKK3	e-115	11p15.2	CFA21	REN309L01	21.916
UMN39A10	BI430418 BQ426216	GGCACACACATTAGGATGTT TTCACCCCAATGGTCCCT	56	Arginine N-methyltransferase 3	PRMT3	0.E + 00	11p14.3	CFA21	REN313E19	18.252
UMN40B10	BI430451 BQ426220	GTGTGCTCAAATGCTTAACAA GCACTTTCCAGAAAGAGCGCTT	56	Integral transmembrane protein 2B	ITM2B	e-123	13q14.3	CFA22	REN42F10	18.232
UMN38G02	BI397451	TGGGACTAACATGAGCACTG TGTGTGCAATTCCTCACTC	58	Acetyl-coenzyme A acetyltransferase 1	ACAA1	1.E-40	3p23-p22	CFA23	AHTK253	23.226
UMN38G10	BI397459	TAAGGGACCCCTGCCCTTGTG TGTGAGTTTGCACITTCGAC	58	Cholecystokinin	CCK	3.E-98	3p22-21.3	CFA23	CPH6	18.055
UMN51E03	BI398226	CTTGAATTTGGGGGAGGG GGTAAAGGGTGGGAAAAGG	58	Profilin 2	PFN2	e-105	3q25.1-3q25.2	CFA23	ZAP47	16.746
UMN45G07	BI398124	AGCCTTGCTTACCCTCACAC ACAACCTGGCTCAGCTCTTC	60	Topoisomerase (DNA) I	TOP1	9.E-87	20q11.2-20q13.1	CFA24	AHT118	20.388
UMN39H07	BI397490 BQ426219	GATAAACAACCACAGGGGA TGGCTCAACTATGTCAAAATGAA	56	Synaptosomal associated protein 25	SNAP25	0.E + 00	20p12-p11.2	CFA24	RDC8	23.464
UMN51A05	BI398194 BQ426193	GGAAGTAGACTCAGCCACAG AACACTAAACGAGTAAATGGAC	56	High-density lipoprotein-binding protein (vigilin)	HDLBP	e-136	2q37	CFA25	AHTH226REN	18.443
UMN43F10	BI398072 BQ426224	AGAAACCTCATCAACAACG CGGCTTGCAGATATACAAGAG	58	Thyroid hormone receptor interactor 12	TRIP12	0.E + 00	2q36.1	CFA25	C25.213	19.097
UMN45B07	BI398115 BQ426186	CGAGCAGCATTTTGTTC GAGTTTCTTGGCCATAACC	56	Chloride channel 3	CLCN3	e-107	4q32	CFA25 ^{ab}	EST-CFZ97779	8.446
UMN10H02	BM735752 BQ426213	ACTGCTGATTTTGTACCAAC TTGGCACTGCTCTTGGAAC	56	Epoxide hydrolase 2, cytoptasmic	EPHX2	e-132	8q21-p12	CFA25 ^d	FH2141	16.258
UMN44G12	BI421110	GTTAGGGTTGGGGAAAAAAG AGCAATCAGCCCACTTAC	56	Human mRNA AL049221 (hypothetical gene)		2.E-81	13q12	CFA25 ^d	REN228N10	22.94
UMN39H02	BI397488 BQ426218	AGTGCAGCGAAAACCTGG ATCTTGAATGGAAGGGGGGAC	58	Clathrin, heavy polypeptide-like 1	CLTCL1	0.E + 00	22q11.21	CFA26	EST-CFZ97756	13.76

Table 1. (Continued)

Marker name	Accession number(s)	Primer sequences	Temp	Gene with closest matching sequence homology	Gene symbol	p-value of match	Human location	Canine location	Adjacent marker	LOD score
UMN44A12	BI421068	AGGCAGCACAAAGACCAAAAC CACAGAAATGGCAACAGACCAG	58	Mitogen-activated protein kinase kinase	MAP2K4	3.E-47	17p11.2	UNLINKED	—	—
UMN45G01	BI398123 BQ426188	AAAAGCTAAAGGTAACCCCAAC TGGGAAACAGTGATAGTAGAG	56	Ferritin heavy chain	FTH1	e-159	11q13	UNLINKED	—	—
UMN45H05	BI398130 BQ426189	CAAACAAAAATCCAAACAGC TTCTACTTCCCTCTCTCTCC	56	Vacuolar protein sorting factor 4A	VPS4A	e-159	16q23.1	UNLINKED	—	—
UMN51A12	BI398199	GAACGTGCAAGTCTACTG	56	Cold inducible RNA-binding protein	CIRBP	e-115	19p13.3	UNLINKED	—	—
UMN51E12	BQ426194 BI398230	GTGCCCAATTCACCTTCTC TGAGGGCCGAGATGCGGTG	54	EH domain-binding mitotic phosphoprotein	EPSIN	5.E-19	19q13.43	UNLINKED	—	—
UMN51H02	BI398243	TGGTGGCAGGTGGTGGCGAG GGATGCTGACAAAAGGAGAC	58	Calcium channel, voltage-dependent, gamma subunit 3	CACNG3	0.E + 00	16p12-16p13.1	UNLINKED	—	—
UMN52H03	BI405236 BQ426199	TTGGAGGGTAGGTTCAATG CCCAGCTACCCTCTCTATTTC ATAACTCTCGCACCCCCATC	58	Dynalectin 2 [p50]	DCTN2	e-155	12q13.2-13q13.3	UNLINKED	—	—

a-synteny break within larger CS b-synteny break within larger CS supported by ZOO-FISH not previously detected on RH panel d-marker that made a previously reported RH singleton into a CS e- This EST also had highly significant matches (less than e^{-100}) to ANT2 and ANT2-like proteins on HSA 7 and 2 respectively. It is highly unlikely that a gene on the human X Chromosome would be on a canine autosome. It is more likely that the most significant human match to this EST is a paralog

Of the 97 ESTs with significant lod scores, 92 (95%) were greater than 10.0 and allowed the positioning of each new marker along its chromosome with a high level of statistical support, while five had lod score values of 8.4 to 10, which were only sufficient to place that marker within a chromosome. The 97 assigned ESTs mapped to 31 different chromosomes and were distributed from one to nine new loci per chromosome (Table 1). Nine canine chromosomes did not acquire new EST markers. Most of these are short chromosomes (CFA29, 33, 34, 36, 37, 38 and Y) that present a lower probability of linking randomly generated markers. EST RH data presented here represent a 30% increase in mapped gene-based markers with value in comparative mapping.

Comparative mapping. Table 1 indicates the map location of the most likely human ortholog of each canine EST. RH data were merged with and positioned on the 1,500 marker canine RH map comprising 320 gene-based markers (Breen et al. 2001). The updated number of gene-based markers on each canine chromosome and current conclusions regarding human-canine synteny for each canine chromosome are reported in Table 2.

Previous ZOO-FISH and RH mapping data identified 71 and 64 evolutionarily conserved segments (CS) respectively between humans and dogs (Breen et al. 2001). The placement of one to five of the canine brain ESTs strengthened 41 of the CS previously identified by RH mapping. The current work has also increased the number of CS identified by RH mapping from 64 to 72. In particular, ten RH singletons identified on CFA 3, 5, 13, 16, 18, 19, 25, 31, and 38 are now supported by two or more gene markers to form CS (Table 2). Thus, 58 of the 72 CS identified by RH mapping are now comprised of at least 2 loci, while 14 remain as RH singletons.

The RH map assignment of the 97 novel ESTs created eight new putative CS represented by singletons on CFA 1, 15, 16, 18, 25, 28, and 32 (Table 2). To confirm these findings, these ESTs were repeatedly searched against multiple databases (Genbank and Ensembl) to determine the most likely human orthologs. Furthermore, amplicons of all new EST markers that mapped to a previously unidentified CS were sequenced to confirm that the PCR product was indeed that expected from the initial primer design. Of the eight putative CS identified in this study by RH mapping, three were previously identified by ZOO-FISH on CFA01, CFA15, and CFA32, but not detected by RH mapping. The other five singletons, on CFA01, 15, 16, 18, 25, and 28, may constitute additional breakpoints. Of these five RH singletons, those located on CFA15, 16, and 25,

Table 2. New canine EST assignments and synteny results by chromosome. RH mapping data for the ESTs mapped in this work is provided as well as the updated dog/human comparative mapping information

GFA	Gene ^a content	New ESTs ^b mapped	Updated gene content	ZOO-FISH ^a conserved segment (CS)	RH ^a CS	New RH ^b CS	Updated RH CS	New CS
1	14	9	23	4	3	1	4	HSA18
2	14	3	17	4	3	-	3	
3	5	4	9	3	3	-	3	
4	12	4	16	3	2	-	2	
5	5	2	7	4	3	-	3	
6	10	0	10	3	3	-	3	
7	10	1	11	2	2	-	2	
8	11	5	16	1	1	-	1	
9	26	6	32	2	2	-	2	
10	4	2	6	3	2	-	2	
11	6	0	6	2	2	-	2	
12	13	1	14	1	1	-	1	
13	5	2	7	2	2	-	2	
14	8	0	8	1	1	-	1	
15	5	3	8	3	2	2	4	HSA01, HSA14 HSA04
16	4	3	7	2	2	1	3	
17	10	1	11	2	2	-	2	
18	11	8	19	2	2	1	3	HSA19
19	2	1	3	2	2	-	2	
20	12	5	17	2	2	-	2	
21	13	3	16	1	1	-	1	
22	6	1	7	1	1	-	1	
23	2	3	5	1	1	-	1	
24	4	2	6	1	1	-	1	
25	5	5	10	3	3	1	4	HSA04
26	7	3	10	2	2	-	2	
27	14	3	17	1	1	-	1	
28	6	3	9	1	1	1	2	HSA09
29	2	0	2	1	1	-	1	
30	4	1	5	1	1	-	1	
31	3	2	5	2	2	-	2	
32	0	1	1	1	0	1	1	HSA04
33	3	0	3	1	1	-	1	
34	2	0	2	1	1	-	1	
35	2	1	3	1	1	-	1	
36	0	0	0	1	0	-	0	
37	3	0	3	1	1	-	1	
38	1	1	2	1	1	-	1	
X	8	3	11	1	1	-	1	
Y	1	0	1	0	1	-	1	
Signed+unli	47	17	64	/	/	/	/	
Total	320	109	429	71	64	8	72	

^a Previous data are from Breen et al (2001).^b Data presented from this work.

which identify a segment from HSA14 within HSA1, a segment of HSA4 within HSA8, and a segment of HSA4 within HSA2, respectively, have been confirmed by the RH mapping of a second locus (Guyon et al., pers. com.). Therefore, only two of the new putative CS remain as RH singletons undetected by ZOO-FISH and not yet confirmed by additional RH-mapped gene markers. They should be considered as tentative until confirmed with more mapped markers.

Discussion

The density of markers on the canine maps is now increasing rapidly; however, there is still an insufficient number of gene-based markers on the RH map to establish the major CS, identify new or potential CS, or allow the efficient prediction of candidate genes for mapped canine traits by comparative mapping approaches. Through characterization of the first 2700 sequences from the canine brain cDNA library, we have identified a variety of brain-specific and ubiquitously expressed genes that map to positions throughout the genome. Of 109 ESTs mapped on the RH panel, 97 could be placed on chromosomes or RH groups, while 12 remained unlinked under our criteria.

The addition of these markers to the RH map allowed us to strengthen the dog/human comparative data, to increase the marker density of the existent CS, to confirm CS previously represented only by RH singletons, and to identify eight new or putative CS between human and dog. Three of these eight new CS were previously detected with ZOO-FISH, while three other CS have been confirmed with other RH-mapped genes. A recent report that placed 52 genes on the canine genetic linkage map concluded that most areas of CS between humans and dogs have already been identified (Parker et al. 2001). Although we agree that placement of more gene markers is needed to address likely alterations in gene order between humans and dogs, we feel our current report clearly demonstrates the need for placing additional markers on the RH map to identify possible microchromosomal rearrangements that ZOO-FISH may not detect, and that meiotic linkage has not yet been detected. Such investigations are critical to the search for and positional cloning of disease loci in dogs.

Acknowledgments

The authors thank Drs. Ned Patterson and Jane Armstrong for useful discussions on canine genetics and veterinary medicine and Ms. Lisa Herron for

performing the phred analysis. This project was supported by the American Kennel Club Canine Health Foundation Grant 2042. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the views of the Foundation.

References

1. Acland GM, Ray K, Mellersh CS, Gu W, Langston AA et al. (1998) Linkage analysis and comparative mapping of canine progressive rod-cone degeneration (prcd) establishes potential locus homology with retinitis pigmentosa (RP17) in humans. *Proc Nat Acad Sci USA* 95, 3048–3053
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215, 403–410
3. Breen M, Thomas R, Binns MM, Carter NP, Langford CF (1999) Reciprocal chromosome painting reveals detailed regions of conserved synteny between the karyotypes of the domestic dog (*Canis familiaris*) and human. *Genomics* 61, 145–155
4. Breen M, Jouquand S, Renier C, Mellersh CS, Hitte C et al. (2001) Chromosome-specific single-locus FISH probes allow anchorage of an 1800-marker integrated radiation-hybrid/linkage map of the domestic dog genome to all chromosomes. *Genome Res* 11, 1784–1795
5. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8, 175–185
6. Jonasdottir TJ, Mellersh CS, Moe L, Hebbebo R, Gamlem H et al. (2000) Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. *Proc Nat Acad Sci USA* 97, 4132–4137
7. Lin L, Faraco J, Li R, Kadotani H, Rogers W et al. (1999) The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* 98, 365–376
8. Matise TC, Perlin M, Chakravarthi A (1994) Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map. *Nat Gene* 4, 384–390
9. Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K et al. (1997) A linkage map of the canine genome. *Genomics* 46, 326–336
10. Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C et al. (2000) An integrated linkage-radiation hybrid map of the canine genome. *Mamm Genome* 11, 120–130
11. Neff MW, Broman KW, Mellersh CS, Ray K, Acland GM et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics* 151, 803–820
12. Ostrander EA, Giniger E (1997) Semper fidelis: what man's best friend can teach us about human biology and disease. *Am J Hum Gene* 61, 475–480
13. Parker HG, Yuhua X, Mellersh CS, Khan S, Shibuya H et al. (2001) Meiotic linkage mapping of 52 genes onto

- the canine map does not identify significant levels of microrearrangement. *Mamm Genome* 12, 713–718
14. Patterson DF (1980) *Current Veterinary Therapy VII*. (Philadelphia: Saunders)
 15. Patterson DF (2000) *Canine genetic disease information system; a computerized knowledge base of genetic diseases in the dog*. (St. Louis, M.: Mosby-Harcourt)
 16. Priat C, Hitte C, Vignaux F, Renier C, Jiang Z et al. (1998) A whole genome radiation hybrid map of the dog genome. *Genomics* 54, 361–378
 17. Roberts MC, Mickelson JR, Patterson EE, Nelson TE, Armstrong PJ et al. (2001) Autosomal dominant canine malignant hyperthermia is caused by a mutation in the gene encoding the skeletal muscle calcium release channel (RYR1). *Anesthesiology* 95, 716–725
 18. van De Sluis B, Rothuizen J, Pearson PL, van Oost BA, Wijmenga C (2002) Identification of a new copper metabolism gene by positional cloning in a purebred dog population. *Hum Mol Genet* 11, 165–173
 19. Vignaux F, Hitte C, Priat C, Chuat JC, Andre C et al. (1999) Construction and optimization of a dog whole-genome radiation hybrid panel. *Mamm Genome* 10, 888–894
 20. Wilcox B, Walkowicz C (1995) *Atlas of dog breeds of the world*. (Neptune City, NJ.: T.F.H. Publications)
 21. Yang F, O'Brien PCM, Milne BS, Graphodatsky AS, Solanky N et al. (1999) A complete comparative chromosome map for the dog, red fox, and human and its integration with canine genetic maps. *Genomics* 62, 189–202