Published in final edited form as:

Genomics. 2001 April 15; 73(2): 171–178. doi:10.1006/geno.2000.6485.

09/15: Comparative Genomics of a Conserved Chromosomal Region associated with a Complex Human Phenotype

Claudia Kappen and J. Michael Salbaum

Center for Human Molecular Genetics and Department of Cell Biology and Anatomy Munroe-Meyer Institute, ZIP 5455 University of Nebraska Medical Center Omaha, NE 68198-5455

Abstract

Three genes that encode related Ig-superfamily molecules have recently been mapped to human chromosome 15 in the region q22.3-23, and to the syntenic region on mouse chromosome 9. These genes presumably derived from gene duplications and they are highly similar to Deleted in Colorectal Cancer (DCC), which functions as an axon guidance molecule during development of the nervous system. In order to find out whether additional genes of this class were present in a chromosomal cluster, we produced a comparative physical map within the region of synteny between mouse chromosome 9 and human chromosome 15. This interval overlaps the critical region for the fourth genetic locus for Bardet-Biedl Syndrome (BBS4) in humans. Bardet-Biedl Syndrome (OMIM 600374) is characterized by poly/syn/brachydactyly, retinal degeneration, hypogonadism, mental retardation, obesity, diabetes, and kidney abnormalities. A detailed map of this locus will help to identify candidate genes for this disorder.

Introduction

Recently, three genes that encode molecules of the Ig-superfamily have been mapped to human chromosome 15 in the region q22.3-23. The Neogenin gene was first identified in chicken as a gene expressed during early neural development (Vielmetter *et al.*, 1994). Its human counterpart encodes a protein with 86% identity, and was mapped to 15q22.3-23 by fluorescent in situ hybridization (FISH) (Vielmetter *et al.*, 1997). Punc (putative neural cell adhesion molecule) was first isolated in the mouse as a possible target gene for regulation by the LIM-domain-homeodomain transcription factor Isl-1 (Salbaum, 1998). It is also highly expressed in the developing nervous system and in the developing limbs, but sharply downregulated in midgestation. It was mapped by FISH to mouse chromosome 9 (band D-E1), and human chromosome 15 (Salbaum, 1999). Moreover, the human Punc gene sequence contains the STS marker WI-14920 (Salbaum, 1999), previously mapped to human chromosome 15 (Hudson *et al.*, 1995). Adjacent to Punc, another member of the same Ig-family subclass was found, Nope (neighbor of Punc/E11) (Salbaum and Kappen, 2000). During development, Nope is highly expressed in muscle and also in brain at later stages of development. Radiation hybrid mapping in the mouse placed Nope between Neogenin and Punc, thus suggesting the presence of a gene cluster of related genes on mouse chromosome 9 (Salbaum and Kappen, 2000). The human Nope cDNA contains STS markers WI-16786 and WI-18508 with known location on human chromosome 15 in the interval D15SS159 - D15S125 and D15S125 - D15S216, respectively (GeneMap'99).

Tel: 402-559-2869 FAX: 402-559-4001 ckappen@unmc.edu, msalbaum@unmc.edu

The BNNO gene is equivalent to the DPP8 gene, published by Abbott *et al.*, Eur. J. Biochem. 267, 6140-6150 (2000) while this paper was in review.

The distal portion of the long arm of human chromosome 15 carries the locus for the fourth genetic form of Bardet-Biedl Syndrome. This syndrome (OMIM 600374) is characterized by poly/syn/brachydactyly, retinal degeneration, hypogonadism, mental retardation, obesity, diabetes, and kidney abnormalities. Recently, the critical region for BBS4 has been narrowed to the interval D15S125 to D15S114 on human chromosome 15 (Carmi *et al.*, 1995). Genes that map within this region, such as Neogenin, Nope and Punc, therefore present potential candidate genes for the disorder. We here present a physical map for 14 genes and 33 anonymous markers in this region, which is syntenic with mouse chromosome 9.

Materials and Methods

Mouse sequences

Mouse BAC clones were obtained by screening a 129/J BAC library (Genome Systems, St. Louis, MO) starting from an initial contig of two oriented clones (BAC296 and BAC331) that contain both the Nope and the Punc gene (Salbaum and Kappen, 2000). From these two BAC clones, approximately 30 kb of fragmented genomic sequence were accumulated during various cloning projects. Further sequence information (approximately 200 kb) was obtained from end sequences of clones derived from sublibraries prepared from both BAC331 and BAC296, where reads of an average of 500bp were obtained from a total of 400 clones (Salbaum, unpublished). Searches of the EST database and subsequent EST walking were used to identify the novel mouse genes Glob (starting from the end sequence of BAC871), Bnno (starting from subclone sequences of BAC296) and Dop (starting from subclone sequences of BAC331).

Nomenclature

The nomenclature committees for the Human and Mouse Genomes assigned the following approved symbols, respectively: ORF13 (mouse) and C15orf7 (human) for DOP, ORF14 (mouse) and C15orf8 (human) for GLOB, ORF15 (mouse) and C15orf9 (human) for BNNO, ORF16 (mouse) and C15orf10 (human) for the previously published Nope gene (Salbaum and Kappen, 2000).

Establishment of a BAC clone contig

Sequences from the distal ends of BAC clones 296 and 331 were used to develop PCR assays to screen the BAC library for overlapping clones. For a 5'-walk from BAC296, we used primers 5'-CTGTGTCTCAATCTTGAACAAACACA-3' and 5'-GGAAGAGAGACAGTAAACATTTTCGT-3' yielding a PCR amplicon of 266bp, which identified BAC870 and BAC871 from the library. A 3'-walk from BAC331 was carried out with primers 5'-CTCCCTTCCTTCCTGATCGTTTTC-3' and 5'-CGGCTCTCAAGCACTGCAGATTTTG-3' for a PCR amplicon of 111bp, thereby marking BAC868. To screen the library for Neogenin, we used the primers 5'-TCAAGCAGTTGACACTTGACTGTG-3' and 5'-TAATCTCACAGTGATGAGAGGAGA-3' (derived from the mouse Neogenin cDNA, Genbank Acc. No. Y09535, position 5107 to 5367), which gave a PCR amplicon of 260bp and identified BAC287 and BAC286. BAC286 tested positive in PCR assays derived from end sequences of BAC870 and BAC871 (using the primers 5'-TCTAGCTCTGGCGCTGTGACTTGT-3' and 5'-ACAGTTCTCTCCTGCAACTCTCCA-3' for BAC870 as well as 5'-CCTTGACTAAGTTCCTAGCCTCA-3' and 5'-AAGTGCTATGGGCAAGACCAAT-3' for BAC871, respectively, thereby establishing a contig for the mouse genomic region. PCR conditions for all assays were 94°C for 3 min, 33 cycles of 93°C for 30 sec, 58°C for 30 sec,

72°C for 35 sec. Reactions were finished for 5 min at 72°C and analyzed by electrophoresis on 2% agarose gels.

Sequence Analysis, Databases and Software

Database information was used as publicly available up to December 20, 2000. Software used for sequence analysis was either MacMolly (SoftGene, Berlin, Germany) or the BLAST suite on the NCBI website (<http://www.ncbi.nlm.nih.gov/>). For sequence interpretation of BACs, we used GenScan (<http://genes.mit.edu/GENSCAN.html>), and SMART (<http://smart.embl-heidelberg.de/>) was used to detect domains in predicted protein sequences (Burge and Karlin, 1997; Schultz *et al.*, 1998).

Human BAC clone sequences were obtained from Genbank (htgs) by searching with cDNA sequences for Neogenin, Nope and Punc. To search for possible coding sequences on human BACs, each BAC was subjected to GenScan analysis, and predicted open reading frames were used to screen for similarities to expressed genes. This was done using Tblastx in BLAST/NCBI against Genbank (nr).

Furthermore, all STS entries in the G4 GeneMap'99 database within the interval D15S159 to D15S114 were compared against Genbank (htgs) to identify their location on known human BACs. A total of 336 STS markers (bins 481 through 485) were assayed (the complete dataset is available from the authors upon request), of which 37 could be localized to two BAC contigs that contained the Neogenin and the Nope and Punc genes, respectively. Additional STS markers were found by screening dbSTS with the assembled contig sequences. STS sequences were screened against Genbank (nr) to find identities to known sequences, and screening against Genbank (dbEST) was used to identify expressed sequences. EST clusters were constructed using reiterative dbEST searches and entries in the Unigene (human) database. From overlapping ESTs, hypothetical cDNAs were constructed, and encoded protein products were predicted using MacMolly (SoftGene, Berlin, Germany). Similarities to known families of molecules were established using blastp/NCBI on Genbank (nr) and Swissprot, as well as SMART domain prediction.

Ordering of unordered pieces of BAC sequences was done manually in MacMolly following comparison of the cDNA sequence to BAC sequences in both orientations. Similarity plots were produced with MacMolly. Unassigned fragments of BACs were then used in BLAST searches against Genbank (nr, dbEST and htgs) to identify further potential matches, which were then mapped onto existing BAC sequence information.

Gene expression profiles for each of the identified genes were determined by searching cDNA sequences against dbEST and compiling the number of independent EST's for each tissue.

Results and Discussion

Previous Radiation Hybrid mapping results had documented close linkage of Neogenin, Nope and Punc on mouse chromosome 9 (Salbaum and Kappen, 2000). These three genes all encode proteins of the immunoglobulin (Ig) superfamily with a signature of four Ig domains. The close sequence relationships between these genes raised the possibility of the presence of an Ig superfamily gene cluster on this region of mouse chromosome 9. With our interest in function and evolution of these three genes, we sought to establish a physical map of the region containing these genes in the mouse. To that end, we initiated a Bacterial Artificial Clone (BAC) walk starting from a contig of two oriented and overlapping BAC clones (BAC296 and BAC331) containing the genes for Nope and Punc. At the same time, we isolated BAC clones containing the Neogenin gene. We developed PCR assays from end

sequences of the resulting BACs. Two PCR assays derived from end sequences of BAC clones that extended our starting contig in the direction upstream of the Nope gene were positive on one of the BAC clones containing the Neogenin gene. We therefore concluded that we had closed the gap between these genes and established physical linkage. Figure 1 shows the structure of the murine BAC clone contig.

We generated sublibraries from the two initial BAC clones BAC296 and BAC331 and have begun to produce sequencing data from these clones. In this process, we have accumulated about 200kb of genomic sequence data from this locus. When we compared these sequences to the database of expressed sequence tags dbEST, there was evidence for at least three other, novel genes in the region in addition to the genes for Nope and Punc.

Combining database searches and walks with human ESTs, we determined that a gene was located on BAC331 that we termed Dop (for Downstream of Punc). The putative transcript of 2.6kb includes the previously isolated partial cDNA AK000516, contains an open reading frame of 831bp, and SMART analysis of the encoded protein of 277 amino acids predicted a transmembrane domain. On BAC296, our data indicated the presence of a gene that we termed Bnno (for Between Neogenin and Nope). Including information from the previously isolated partial cDNA AK000290, we assembled a putative transcript of 5.3kb. An open reading frame of 2646bp can encode a protein of 882 amino acids with high similarity to the DPP_{IV}/CD26 family of type II transmembrane proteases (De Meester *et al.*, 1999). The highest degree of similarity was found in the region of CD26 that encompasses the catalytic domain, whereas the protein-interaction domains were conserved to a lesser extent. An EST walk initiated from an end sequence of BAC871 resulted in a third novel gene in this region that we termed Glob (for Gene left of Bnno). With a putative transcript size of 2.7kb and an open reading frame of 1482bp, the Glob gene can encode a protein of 494 amino acids that, according to SMART domain analysis, exhibits similarity to a protein interaction domain of von Willebrand factor A (VWA domain) and to the Drosophila protein CG4785 (Genbank Acc. # AAF15404). Through this analysis, it became clear that two of the novel genes were located between Neogenin and Nope, ruling out the possibility of a contiguous Ig superfamily gene cluster.

The presence of these novel genes raised the question about the structure of this genomic locus. Since our own mouse DNA sequence information was of fragmentary nature, we proceeded to utilize information from the human genome project as available in public databases. By analyzing human sequence information in conjunction with mapping and marker information available from GeneMap'99, we were able to generate an extended map of the syntenic human area that encompasses a total of 14 genes.

For the human locus, we identified BACs containing the NEO, NOPE and PUNC genes through searching the htgs division of Genbank with the human and mouse cDNA sequences. These BACs were all working draft sequence consisting of unordered and unoriented sequence fragments, except for BAC AC013553, which consisted of one piece by August 28, 2000 (sequence update December 13, 2000). GenScan searches on those BACs and subsequent Genbank and Unigene searches identified sequences that represented expressed genes. The overlap of BACs was confirmed by screening whole BACs against each other or by searching for the location of pieces of one BAC on another (BLAST2 and MacMolly). In this way, overlapping BACs were found for NEO and adjacent genes, and for NOPE, PUNC and adjacent genes. In addition, by mapping STS markers onto BACs using Genbank (htgs), we were able to assign additional markers to each BAC and derive the likely order of markers and genes in this genomic region.

Only a fraction of markers for this region (according to GeneMap'99) could actually be assigned to the two BAC contigs. In addition, sequences derived from the same gene appear as independent entries in distinct bins of GeneMap'99, indicating the limit of resolution of previous mapping results. Although we could not order and orient all BAC fragments based on the available information, we did not find evidence for the existence of additional genes between stSG58291 and AW194628, and D15S1026 and UT5583, the most distant markers on the respective BAC contig. Our mapping efforts were facilitated by the fact that the majority of markers in GeneMap'99 correspond to expressed genes. This allowed us to perform EST walks and map additional EST's for each transcription unit, thereby strengthening the original assignment.

Independently, we screened Genbank (htgs) with the sequence for each of the markers from the Radiation Hybrid map of Chromosome 15 (Whitehead Institute) between D15S159 and D15S114, some of which coincide with those listed in GeneMap'99. However, only some of the D15S markers known to localize in this region could be found on our sequence assemblies. These searches were done with MacMolly as it allows comparisons with low complexity sequences, such as microsatellite sequences. The combined results of these searches are compiled in Table 1.

To gain closer insight into the gene structures in this locus, we performed extensive sequence comparisons. First, the human BAC working draft sequences were compared to each other using BLAST2 or MacMolly software to establish sequence contigs of overlapping fragments. Second, a collection of cDNA sequences representing the genes in this region was established by comparing the human BAC sequences to the Genbank (nr) and dbEST databases. Our collection contained full-length human cDNAs where available, full-length mouse cDNAs where human information was lacking, and putative human transcript sequences assembled from EST walks. Third, relationships between each cDNA or transcript sequence and the human BAC working draft sequence fragments and other available genomic information (Lorenzo *et al.*, 1999; Nakamura *et al.*, 1999) were established by extensive pairwise sequence comparisons. In this fashion, we were able to establish six sequence assemblies (of which two are actual sequence contigs) that represent approximately 600kb and are shown in Figure 2.

We did not find human BAC sequences that span the gap between the HCN4/NEO assembly and the assembly beginning with the IRLB gene. Except for the HCN4/NEO-assembly, the order of sequence assemblies was derived from the BAC contigs in Table 1. However, the relative orientation of individual assemblies towards each other can not be determined from the currently available sequence information. Sequence assemblies contain gaps at locations where cDNA or transcript sequence information bridges two working draft fragments that do not show overlap. These gaps typically occur in introns, and an assembly size therefore represents a minimal size. In the four cases where an assembly contains more than one gene, we were able to deduce the transcriptional orientation of the genes within one sequence assembly. Accordingly, the HCN4 and NEO genes, as well as the IRLB, SLC24A1, GLOB, BIND and BNNO genes are all transcribed in opposite directions, whereas the NOPE and PUNC genes, as well as the CLPX, PDCD7 and RIS genes, are transcribed in the same direction.

Seven results arise from these data: (i) We have established comparative annotation maps for a portion of the region syntenic between mouse chromosome 9 and human chromosome 15. This was done using first pass mouse sequences and working draft human sequences, STS marker information and EST sequences. Thus, even without optimum quality sequence data, physical mapping of existing and novel genes was accomplished. (ii) We identified the three novel genes GLOB, BNNO, and DOP in the 09/15 syteny region, which exist in both

mouse and human. (iii) Except for NEO, no additional Ig-superfamily genes with relation to DCC were found in the vicinity of PUNC and NOPE, indicating that the duplication of these latter two genes was unique. (iv) For the mouse locus, we physically linked Neogenin in a BAC contig to Nope and Punc. We were unable to close the gap between the Neogenin-containing and the Nope/Punc-containing BAC contigs on the basis of the existing human BAC sequences. This may indicate larger distances within the human synteny region or the presence of additional genes between both BAC contigs. Alternatively, the gap may reflect the presence of repetitive sequences (Kappen and Salbaum, unpublished) which confound sequence comparisons. (v) Our map for the locus around NOPE and PUNC includes markers previously localized to GeneMap'99 bins 481 and 482. It is likely that the overlapping BACs we identified contain all the genes present in this region. (vi) The 09/15 syntenic region contains 14 distinct genes that are (except for Punc, Nope and Neogenin) unrelated in sequence and structure, but their order is highly conserved between human and mouse. (vii) The locus overlaps with the region defined as the locus for Bardet-Biedl Syndrome 4. One or more of the genes residing in this genomic region may therefore be a candidate gene for BBS4.

In order to investigate whether the genes we mapped in the 09/15 syntenic region could be involved in Bardet-Biedl Syndrome, we analyzed their expression. The mouse Punc gene is highly expressed in the developing nervous system and the limbs, and sharply downregulated at midgestation (Salbaum, 1998). In the adult mouse, expression is found in testis. Nope expression in developing mouse embryos is initially highly restricted to muscle at E10.5. After day E13.5, it is also found in the developing nervous system. Nope can also be detected in the developing testis at E15.5 (Salbaum, unpublished). In the adult, Nope expression is found in muscle and in brain (Salbaum and Kappen, 2000), and strongly in the retina (Salbaum, unpublished). For the other genes, there is little information available about their expression during embryogenesis. However, many ESTs are available which we used to compile EST occurrence in different adult tissues. Such data are highly dependent on the composition of the EST databases and libraries, and therefore the resulting profiles are a qualitative rather than a quantitative reflection of gene expression in different tissues.

From the expression profiles depicted in Figure 3, it is evident that each gene is expressed with a unique tissue distribution. These data indicate that specific regulatory elements must be present for each gene. Secondly, the developmental timing of expression (where known) is also different for a number of these genes, indicating that specific temporal regulatory mechanisms must exist. It will be particularly important to identify the relevant regulatory elements and understand how each element, or cooperating elements, controls one gene specifically. For example, preliminary data show that the regulatory element responsible for expression of Punc in the developing limb is located in exon 20 and part of the 3'-untranslated region of the Nope gene (Salbaum and Kappen, unpublished). These experiments also established that the element responsible for nervous system expression of Punc is not located within 17kb upstream or 5kb downstream of the transcription start site of Punc. The findings suggest that, while located at a distance, the action of regulatory elements and their enhancer function is restricted to specific genes by yet unidentified mechanisms. The physical mapping of all genes within this interval will now allow us to understand how closely spaced genes are differentially regulated, and how enhancer activities are insulated from each other. Interestingly, a phenotype resembling aspects of Bardet-Biedl Syndrome was found in mice that carry a targeted disruption of the *Ovo1* gene, which encodes a Zinc-finger transcription factor (Dai *et al.*, 1998). Assuming that genes on different Bardet-Biedl loci act in the same developmental and cellular pathways, it is conceivable that genes in the 09/15 synteny region could be regulated by this transcription factor.

Our comparative map includes markers and genes that have previously been assigned within the region that spans between D15S159 and D15S114. The BBS4 locus was mapped between D15S125 and D15S114 (Carmi *et al.*, 1995), and based upon one family, the interval D15S131 to D15S114 has been suggested as the shortest critical interval for BBS4 (Bruford *et al.*, 1997). Several genes have been evaluated specifically as potential candidates for BBS4, including FEM1B, ISLR and MYO9A. FEM1B was excluded as a candidate gene on the basis of its map position between D15S125 and D15S216 (Ventura-Holman *et al.*, 2000), outside of the D15S131 to D15S114 interval. Sequencing of the MYO9A (Gorman *et al.*, 1999) and ISLR (Nagasawa *et al.*, 1999) genes did not reveal mutations in BBS patients. Very recently, Katsanis *et al.* (2000) and Slavotinek *et al.* (2000) reported mutations in Bardet-Biedl patients in the MKKS gene, suggesting that it could be the locus for BBS6. MKKS (McKusick-Kaufman Syndrome; OMIM 236700) clinically overlaps with Bardet-Biedl syndrome (Schaap *et al.*, 1998). The MKKS gene encodes a protein with similarity to chaperonins and implicates impaired protein processing as a cause of McKusick-Kaufman syndrome (Stone *et al.*, 2000) and in the sixth genetic form of Bardet-Biedl Syndrome. Interestingly, the CLPX gene in the 09/15 synteny region also encodes a chaperonin, and by analogy, could be a candidate for BBS4. Alternatively, the different genetic forms of BBS may affect different types of proteins a unique molecular or cellular pathway. The 09/15 syntenic region thus may encode the functional effector molecules in tissues affected in Bardet-Biedl Syndrome.

Acknowledgments

We wish to thank Jolene Andrews for expert technical help with mouse BACs and the generation and analysis of BAC subclone libraries.

References

- Bruford EA, Riise R, Teague PW, Porter K, Thomson KL, Moore AT, Jay M, Warburg M, Schinzel A, Tommerup N, Tornqvist K, Rosenberg T, Patton M, Mansfield DC, Wright AF. Linkage mapping in 29 Bardet-Biedl syndrome families confirms loci in chromosomal regions 11q13, 15q22.3-q23, and 16q21. *Genomics*. 1997; 41:93–99. [PubMed: 9126487]
- Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 1997; 268:78–94. [PubMed: 9149143]
- Carmi R, Rokhlina T, Kwitek-Black AE, Elbedour K, Nishimura D, Stone EM, Sheffield VC. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. *Hum. Mol. Genet.* 1995; 4:9–13. [PubMed: 7711739]
- Dai X, Schonbaum C, Degenstein L, Bai W, Mahowald A, Fuchs E. The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev.* 1998; 12:3452–3463. [PubMed: 9808631]
- De Meester I, Korom S, Van Damme J, Scharpe S. CD26, let it cut or cut it down. *Immunology Today*. 1999; 20:367–375. [PubMed: 10431157]
- Gorman SW, Haider NB, Grieshammer U, Swiderski RE, Kim E, Welch JW, Searby C, Leng S, Carmi R, Sheffield VC, Duhl DM. The cloning and developmental expression of unconventional myosin IXA (MYO9A) a gene in the Bardet-Biedl syndrome (BBS4) region at chromosome 15q22-q23. *Genomics*. 1999; 59:150–160. [PubMed: 10409426]
- Hudson TJ, Stein LD, Gerety SS, Ma J, Castle AB, Silva J, Slonim DK, Baptista R, Kruglyak L, Xu SH, *et al.* An STS-based map of the human genome. *Science*. 1995; 270:1945–1954. [PubMed: 8533086] , with supplementary data from the Whitehead Institute / MIT Center for Genome Research, Human Genetic Mapping Project, Data Release 12 (July 1997).
- Katsanis N, Beales PL, Woods MO, Lewis RA, Green JS, Parfrey PS, Ansley SJ, Davidson WS, Lupski JR. Mutations in MKKS cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nat. Genet.* 2000; 26:67–70. [PubMed: 10973251]

- Lorenzo P, Aman P, Sommarin Y, Heinegard D. The human CILP gene: exon/intron organization and chromosomal mapping. *Matrix Biol.* 1999; 18:445–454. [PubMed: 10601732]
- Nagasawa A, Kudoh J, Noda S, Mashima Y, Whright A, Oguchi Y, Shimizu N. Human and mouse ISLR (immunoglobulin superfamily containing leucine-rich repeat) genes: genomic structure and tissue expression. *Genomics.* 1999; 61:37–43. [PubMed: 10512678]
- Nakamura I, Okawa A, Ikegawa S, Takaoka K, Nakamura Y. Genomic organization, mapping, and polymorphisms of the gene encoding human cartilage intermediate layer protein (CILP). *J. Hum. Genet.* 1999; 44:203–205. [PubMed: 10319588]
- Salbaum JM. Punc, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system. *Mech Dev.* 1998; 71:201–204. [PubMed: 9507132]
- Salbaum JM. Genomic structure and chromosomal localization of the mouse gene Punc. *Mamm. Genome.* 1999; 10:107–111. [PubMed: 9922388]
- Salbaum JM, Kappen C. Cloning and Expression of Nope, a new Mouse Gene of the Immunoglobulin Superfamily related to Guidance Receptors. *Genomics.* 2000; 64:15–23. [PubMed: 10708514]
- Schaap C, tenTusscher MP, Schrandt JJ, Kuijten RH, Schrandt-Stumpel CT. Phenotypic overlap between McKusick-Kaufman syndrome and Bardet-Biedl syndromes: are they related? *Eur. J. Pediatr.* 1998; 157:170–171. [PubMed: 9504798]
- Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: Identification of signalling domains. *Proc. Natl. Acad. Sci. USA.* 1998; 95:5857–5864. [PubMed: 9600884]
- Slatovinec AM, Stone EM, Mykytyn K, Heckenlively JR, Green JS, Heon E, Musarella MA, Parfrey PS, Sheffield VC, Biesecker LG. Mutations in MKKS cause Bardet-Biedl Syndrome. *Nat. Genet.* 2000; 26:15–16. [PubMed: 10973238]
- Stone DL, Slavotinek A, Bouffard GG, Banerjee-Basu S, Baxevanis AD, Barr M, Biesecker LG. Mutation of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nat. Genet.* 2000; 25:79–82. [PubMed: 10802661]
- Ventura-Holman T, Haider NB, Maher JF. The human FEM1B gene maps to chromosome 15q22 and is excluded as the gene for Bardet-Biedl syndrome, type 4. *Am. Med. Sci.* 2000; 319:268–270.
- Vielmetter J, Chen XN, Miskevich F, Lane RP, Yamakawa K, Korenberg JR, Dreyer WJ. Molecular characterization of human neogenin, a DCC-related protein, and the mapping of its gene (NEO1) to chromosomal position 15q22.3-q23. *Genomics.* 1997; 41:414–421. [PubMed: 9169140]
- Vielmetter J, Kayyem JF, Roman JM, Dreyer WJ. Neogenin, an avian cell surface protein expressed during terminal neuronal differentiation, is closely related to the human tumor suppressor molecule deleted in colorectal cancer. *J Cell Biol.* 1994; 127:2009–2020. [PubMed: 7806578]

Figure 1. A BAC clone contig linking the genes for Neogenin, Nope and Punc

The starting contig contained two overlapping BAC clones BAC296 and BAC331, which harbor the Nope and Punc genes, respectively. PCR assays were developed from BAC end sequences and used to screen a mouse 129/J BAC library. Screening with the assay from BAC296 yielded BAC871 (as well as two other isolates BAC869 and BAC870, not shown), and screening with the assay from BAC331 resulted in two isolates, BAC868 and BAC867 (not shown). Screening the library with a PCR assay derived from the mouse Neogenin cDNA yielded two overlapping clones in BAC267 and BAC268. A physical contig could be established when PCR assays derived from end sequences of BAC870 and BAC871 were positive on DNA from the Neogenin BAC clone BAC286. Clones are shown schematically and are not drawn to scale. Red arrows indicate the approximate location of genes, whereas black arrows represent PCR assays.

Figure 2. Construction of sequence assemblies from overlapping BACs with order and orientation of coding sequences

From overlapping BACs (designations on the left), sequence assemblies were constructed (black vertical lines). Matches between cDNA and genomic sequences were plotted in 2-dimensional plots displaying the cDNA in perpendicular direction (colored lines). Gene names and potential function are given in the same colors. The direction of transcription for each cDNA is indicated with a black arrow on the left side. Open triangles mark the positions of gaps in the sequence assemblies, which could not be filled unambiguously. Not all fragments from BACs could be assembled; unassigned fragments from each BAC may

represent the intervening or intron sequences. The sequence assemblies are displayed according to their minimum sizes.

Figure 3. Gene expression profiles for the 09/15 synteny region in human tissues
For each gene, all ESTs found were tabulated according to tissue origin (hits in pooled libraries were excluded). The tissues most frequently found for all genes were selected for graphic representation and are listed alphabetically. The size of each color block area is proportional to the number of ESTs found. No attempt was made to normalize to the total number of ESTs for each gene. Colors correspond to those in Figure 2. The total number of ESTs found for different genes was variable, and the expression profiles are quite different from each other.

Table 1

Mapping of STS and EST markers onto overlapping human BACs

Marker	GeneMap'99	STS-Name	Genbank Acc.#	Identity	assembly: site	BAC
D15S114	265.55 >					
		D15S1026	Z53932		hcn/heo: 6759	AC009660
			AJ238850	HCN4	hcn/heo: 12324	AC009660
		D15S215	Z24647		hcn/heo: 39899	AC009660
	258.51	stSG47652	AA224594		hcn/heo: 60856	AC009660
			U72391	NEOGENIN	hcn/heo: 74412	AC022320
	258.61	Cda0sh01	AA009813		hcn/heo: 75980	AC009660
		WL-6892	G06023		hcn/heo: 75980	AC009660
		SHGC-19068	G31056		hcn/heo: 97011	AC009660
		stSG48277	AA358159		hcn/heo: 108213	AC009660
		SHGC-13080	G17863		hcn/heo: 146924	AC022320
		UT5583	L18126		hcn/heo: 218544	AC022320
	252.15/247.3	WL-15859	R52522			AC022320-21
D15S124	251.04 >					
D15S131						
D15S216	243.25 >					
		stSG8039	AA470795			AC011939-14
			N54726			AC011939-14
		TIGR-A003P32	G25284			AC011939-13
			AW194628	IRL-B	irlb-bnmo: 6694	AC027220
	231.16	sts-R98411	R98411	SLC24A1	irlb-bnmo: 12772	AC027220
	231.26	A009S19	G32806	SLC24A1	irlb-bnmo: 12924	AC027220
	231.16/223.9	A002H06	G26215			AC027220
	231.16	stSG15823	R83880	GLOB	irlb-bnmo: 70856	AC027220
	231.26	stSG4803	AA010532	GLOB	irlb-bnmo: 77017	AC027220
	231.26/223.2	WL-15305	AA025364		irlb-bnmo: 77434	AC027220
		SHGC-14524	G13882		irlb-bnmo: 78730	AC027220
	230.25/223.2	stSG54785	AA130811	b-IND1/HSPC121	irlb-bnmo: 79422	AC027220
	231.26	stSG48116	N54387			AC027220-17
						AC011846-3

Marker	GeneMap'99	STS-Name	Genbank Acc.#	Identity	assembly: site	BAC
	231.16	stSG30183	AA157110	BNNO	irlb-bmno: 199274	AC011846
	236.22	sts-N29468	AA236898		irlb-bmno: 204967	AC068507
	/225.6	SGC31814	H56172	BNNO	irlb-bmno: 207364	AC068507
	231.16/225.6	stSG10196	AA128184	BNNO	irlb-bmno: 207647	AC011846
	231.16	WI-19652	AA128184	BNNO	irlb-bmno: 207647	AC011846
		SHGC-84628	G51992		nope/punc: 11035	AC068507
		A006137	G32363		nope/punc: 44093	AC068507
	231.16/230.25	A007C09	AA706095	NOPE	nope/punc: 45618	AC068507
	230.25/228.4	WI-16786	R69087	NOPE	nope/punc: 45626	AC068507
	231.26/223.8	WI-18508	G22216	NOPE	nope/punc: 45684	AC068507
	231.16	A006115	AA086299/WI-21066		nope/punc: 48195	AC068507
	231.26	WI-21066	F09441		nope/punc: 46240	AC068507
		SHGC-86276	G53588		nope/punc: 52487	AC068507
	230.25/228.4	WI-14920	G23164	PUNC	nope/punc: 77254	AC068507
	231.16	stSG53788	AA903499		nope/punc: 80912	AC068507
			AK000516	DOP	dop: 2512	AC068507
	231.26	WI-12537	AA723959	DOP	dop: 30821	AC068507
			AB022430	CILP	clp: 22190	AC068573
	236.54	A009P37	AJ006267	CLPX1	clpx: 5110	AC068573
	231.16	stSG10346	R27323		clpx: 36003	AC068213
	231.16	A004Z39	G26521		clpx: 37042	AC068213
	236.5	sts-H79322	H79322		clpx: 37073	AC068213
	231.16	stSG42450	N71334		clpx: 37095	AC068573
		SHGC-2918	G13960		clpx: 38415	AC068573
			NM_016688	PDCD7	clpx: 51490	AC068573
			AA912159		clpx: 58291	AC068573
	231.16	SGC30454	AA350297		clpx: 67934	AC068573
	231.16	stSG53115	AA812905		clpx: 92310	AC068573
			AF233588	RIS	clpx: 117363	AC068573
	226.54	stSG43303	AI000450	RIS	clpx: 131929	AC068573
	226.54	Cda08f05	H22943	RIS	clpx: 131934	AC068573

D15S125 231.16 >

Marker	GeneMap'99	STS-Name	Genbank Acc.#	Identity	assembly: site	BAC
D15S159	227.55	stSG30156	AA156475	RIS	clpx: 131947	AC013553
	227.55	stSG58291	N90523		clpx: 182317	AC013553
	224.23 >					

Starting from GeneMap'99, all markers between D15S159 and D15S114 were mapped to human BAC sequences (Genbank higs division). Sequences located on more than one BAC were used to assemble overlapping BACs into two contigs, the first spanning from D15S1026 to WI-15859, and the second from stSG8039 to stSG58291 (listed in correlation with the mouse locus map). Using cDNA information, unordered and unoriented pieces from each BAC were assembled into sequence contigs where possible. These assemblies were then used to order the STS and EST markers along the two major BAC contigs. Based upon GenScan predictions, the existence of other genes within these contigs is unlikely, however, it is possible that additional genes may be located between the two major BAC contigs. Physical positions were taken from GeneMap'99; in cases of discrepancies, both positions are listed. For reference, the position of microsatellite markers is given in the first two columns according to their positions as listed in GeneMap'99.