Human Disease Genes and Their Cloned Mouse Orthologs: Exploration of the FANTOM2 cDNA Sequence Data Set

Lynn M. Schriml,[1,11] David P. Hill,[2] Judith A. Blake,[2] Hidemasa Bono,[3] Anthony Wynshaw-Boris,[4] William J. Pavan,[5] Brian Z. Ring,[6] Kirk Beisel,[7] Mitsutoshi Setou,[8] RIKEN GER Group[3] and GSL Members,[9,10] and Yasushi Okazaki[3,9]

[1]*National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland 20894, USA;* [2]*The Jackson Laboratory, Bar Harbor, Maine 04609, USA;* [3]*Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan;* [4]*Departments of Pediatrics and Medicine, University of California, San Diego School of Medicine, San Diego, California 92093, USA;* [5]*National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892-4472, USA;* [6]*Applied Genomics, Inc., Sunnyvale, California 94085, USA;* [7]*Department of Genetics, Boys Town National Research Hospital, Omaha, Nebraska 68131, USA;* [8]*Graduate School of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan;* [9]*Genome Science Laboratory, RIKEN, Hirosawa, Wako, Saitama 351-0198, Japan*

The FANTOM2 cDNA sequence data set is an excellent model to demonstrate the power of large-scale cDNA sequencing, with the goal of providing a full-length transcript sequence for each mouse gene. This data set enhances the use of the mouse as a model for human disease. Here we identify mouse cDNA sequences in the FANTOM2 data set for a set of 67 human disease genes that as of May 2002 had no corresponding mouse cDNA annotated in the Mouse Genome Informatics (MGI) database. These 67 human disease genes include genes related to neurological and eye disorders and cancer. We also present a list of the human disease genes and their cloned mouse orthologs found in two public databases, LocusLink and MGI. Allelic variant and gene functional information available in MGI provides additional information relative to these mouse models, whereas computed sequence-based connections at NCBI support facile navigation through multiple genomes.

[Supplemental material is available online at www.genome.org.]

The mouse has been used as a genetic system for more than 100 years and provides a rich resource of genetic mutations and inbred strains for biomedical research (Beck et al. 2000; Denny and Justice 2000). We can now couple mutational and phenotypic analysis in the mouse with conserved sequence and positional analysis to identify molecular sequences of candidate disease genes in humans. Conversely, it is also possible, through conserved sequence and positional information, to identify candidate mouse orthologs and their phenotypic variants that can be used as models for human disease. Once candidate orthologous genes have been identified in the mouse, the power of targeted mutagenesis can be used to create precise mouse models for studying human disease. In the past, identification of candidate disease genes in the mouse often required months of work using cloning strategies based on conserved synteny to physically isolate genes and then sequence them.

A view of the complete set on mouse transcripts, the transcriptome, is emerging as new mouse sequence data from whole genome annotation projects (ENSEMBL, http://www.ensembl.org/; and NCBI, http://ncbi.nlm.nih.gov) and full-length cDNA sequencing projects (Mammalian Gene Collection, MGC, http://mgc.nci.nih.gov/; and RIKEN's Mouse Encyclopedia, http://www.gsc.riken.go.jp/e/FANTOM/) are integrated and analyzed. The view of the mouse transcriptome from the FANTOM2 data set is an excellent resource to explore the power of large-scale cDNA sequencing to enhance the use of the mouse as a model for human disease. (The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase II Team 2002).

Here we present an analysis of the FANTOM2 cDNA data set (as of May 2002) wherein we looked for novel mouse cDNAs representing mouse orthologs of human disease genes. Using the publicly available human and mouse gene, sequence, and orthology data from the NCBI LocusLink project (http://www.ncbi.nlm.nih.gov/LocusLink/; Pruitt and Maglott 2001) and the Mouse Genome Informatics (MGI) Database (http://www.informatics.jax.org/), we assembled a list of cloned human disease genes represented in LocusLink and their cloned mouse orthologs represented in MGI (The FANTOM Consortium and the RIKEN Genome Exploration Re-

[10]**Yoshihide Hayashizaki, Takahiro Arakawa, Piero Carninci, and Jun Kawai.**
[11]**Corresponding author.**
E-MAIL schriml@ncbi.nlm.nih.gov; FAX (301) 480-2288.

Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.979503.

13:1496–1500 ©2003 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/03 $5.00; www.genome.org

search Group Phase II Team 2002). A similar strategy was used in the examination of 288 cloned human disease genes in the *Drosophila* genome sequence paper (Rubin et al. 2000) and the first FANTOM consortium paper (Kawai et al. 2001).

The goal of this study was to find novel full-length mouse cDNAs in the FANTOM2 cDNA data set that represent orthologs of human disease genes. In addition, we generated a comprehensive list (as of May 2002) of the cloned human disease genes and their annotated mouse orthologs in MGI. Here, we present the set of human disease genes, their orthologs, and the mouse cDNA clones we have identified and discuss the information that is presently available for their study.

RESULTS

Analysis of the FANTOM2 Set: Identifying Orthologs

Examining the annotated gene records at MGI and LocusLink, we found that of the 1022 human disease genes, 921 had a cloned ortholog in MGI and 101 did not. Then examining only those human disease genes included in the BLAST analysis (*N* = 993; Supplemental Table 1, available online at www.genome.org), we found that of the set of human disease genes that had no annotated mouse orthologous cDNA as of May 2002 (*N* = 80), 84% had a probable ortholog in the FANTOM2 cDNA data set (Supplemental Table 3). These include a variety of disease genes including cancer-related genes, for example, *NUMA1*, which is associated with acute promyelocytic leukemia, and *SDHD*, mutations of which have been linked to hereditary paraganglioma. Additionally, this set includes examples of neurological disease genes such as *USH3A*, which is associated with the phenotype of Usher syndrome, type 3, and *IL1RAPL1*, which is associated with type 1 X-linked nonspecific mental retardation. Examples of human diseases affecting the eye in this set include *PRPF8*, which is a candidate for the autosomal dominant form of retinitis pigmentosa, and *NYX*, mutations of which have been shown to cause X-linked complete congenital stationary night blindness (*CSNB1*). The lists of cloned human disease genes and the results of the BLAST analysis for the subset for which we found curated orthologs are presented in Supplemental Table 2. For completeness, the human disease genes for which the Fantom2 cDNA set did not yield significant BLAST results are listed in Supplemental Table 4 (set with curated orthologs) and Supplemental Table 5 (set with no curated orthologs).

A Second Look: Re-evaluation of the No Ortholog Human Disease Set

We re-examined the set of 101 cloned human disease genes used in this study just prior to submission of this paper (September 25, 2002) to determine if mouse orthologs had been recently annotated for these genes. By looking at the Human–Mouse Homology data set at MGI (ftp://ftp.informatics.jax.org/pub/informatics/reports/GDB_Accession.rpt), we could identify newly annotated mouse genes and determine if the mouse records contained sequences other than partial mRNAs or ESTs in the MGI database. For this set, we determined if there were highly similar proteins (partial or complete) in Entrez at NCBI. The Entrez protein database includes records from GenBank translations, RefSeq (both curated and model, XP_000000), and SWISS-PROT. To emulate general user access, we looked in BLink for the mouse protein with the highest BLAST score. We followed BLink (e.g., http://www.ncbi.nih.gov/cgi-bin/Entrez/blink?pid=4557225&all=1) links on the LocusLink page for the human gene to precomputed protein neighbors. Clicking on the SCORE of the BLAST2 results, we recorded the percent identity of the longest aligned fragment. Keeping in mind that on average, mouse and human proteins have been found to exhibit 86.4% (SD = 12.3) identity, with a range of 41.1% to 100% (Makalowski et al. 1996), the percent identity of the hit can suggest whether the protein identified (Table 1) was the ortholog of the human protein.

We included in our results those proteins that shared sequence identity >65% over the entire length of the alignable region. Because we only used percent identity to identify candidate orthologs, some of the protein accessions listed in Table 1, such as *KRT1*, may be from paralogs. We also looked for mouse models via text queries using the gene name for the human disease gene (e.g., XP_134985 for *BBS4* and XP_130099 for *NUP214*). Protein sequence similarity for *BBS4* and *NUP214*, the mouse proteins XP_134985 and XP_130099, was determined by BLAST2. Using both methods, we identified highly related mouse proteins for 37 of the 101 human

Table 1. Highly Related Mouse Proteins to Human Disease Genes in No Ortholog Set Identified Via Blink

Human LocusID	Human symbol	Mouse protein	% ID
275[a]	*AMT*	XP_147096	88
585	*BBS4*	XP_134985	88
1203	*CLN5*	XP_127882, AAH25487	77
1540	*CYLD*	XP_134376	94
2162	*F13A1*	XP_138580	70
2799	*GNS*	XP_125895	93
3030	*HADHA*	XP_131963	86
3032	*HADHB*	NP_663533, AAH05585	91
3848	*KRT1*	BAB31776	75[b]
3850	*KRT3*	XP_128174	68
4026	*LPP*	XP_155911	83
4247	*MGAT2*	NP_666147, AAH10583	88
4306	*NR3C2*	XP_146468	88
4552	*MTRR*	XP_127460	78
4719	*NDUFS1*	NP_663493, AAH06660	93
4926	*NUMA1*	AAH04667	86
5190	*PEX6*	NP_663463, AAH03424	87
5205	*ATP8B1*	XP_129012	95
5378	*PMS1*	XP_129877, NP_705784	74
6392	*SDHD*	XP_134803, BAB29086	82
6906	*SERPINA7*	XP_111955	73
7401	*USH3A*	NP_700433, AAM88775	85
8021	*NUP214*	XP_130099	80
8030[a]	*D10S170*	XP_125664	99
8050	*PDX1*	XP_130563, XP_207088	85
8086	*AAAS*	NP_700465	93
8301	*PICALM*	NP_666306, AAH11470	95
8540	*AGPS*	XP_130294	91
8659	*ALDH4A1*	XP_204153	91
9825	*SPATA2*	XP_111839	85
11141	*IL1RAPL1*	XP_141905	98
23230	*CHAC*	XP_129223, XP_195100	84, 82
26119	*ARH*	NP_663529, AAH21467	87
54982[a]	*CLN6*	XP_134858	90
65125	*PRKWNK1*	XP_132838	95
80207	*OPA3*	XP_133224	85
114548	*CIAS1*	NP_665826, AAL90874	82

[a]Human LocusID in the No Homolog set that did not have a match in the FANTOM2 data set.
[b]Possible paralog.

genes in the No Ortholog Set. Of these, three (*AMT*, *D10S170*, *CLN6*) had no hits in the FANTOM2 data set (Table 1). We found that 23 genes in this No Ortholog Set now have an associated orthologous mouse gene with cDNA sequence data available in MGI. This number will continue to grow as additional data are entered into MGI.

Allelic Variants in Mouse

To provide additional information relative to this set of mouse models for human disease, we assessed the availability of allelic variants data at MGI. For genes that had no curated alleles, we used the gene-to-reference links in MGI to manually scan abstracts to determine if authors had reported allelic variants of each gene. We found 594 of the mouse orthologs (63%) to have reported allelic variants. Of the genes that have reported allelic variants, 428 have variants that are already described in MGI. An additional 165 genes have an allelic variant that was reported in the literature. Given that our literature search was based on a visual scan of titles and abstracts, these alleles will require further, in-depth investigation. MGI curators can now focus on curation of alleles of these genes in the MGI database. The results of this analysis are shown in Supplemental Table 6.

It is important to note that in this study no emphasis was placed on the phenotypic characterization of alleles to determine their validity in modeling human disease. One could think of alleles as falling into four broad categories: (1) natural variants that were used in initial mapping and characterization of genes, (2) natural or induced mutations that were isolated on the basis of a noticeable phenotype, (3) engineered transgenic animals that result in a neomorphic or hypermorphic mutation, or (4) engineered knockout animals that represent a null or hypomorphic mutation. Mutants from the first category will likely not be very useful as disease models, because they were usually originally identified as isoenzyme variants or simple restriction-fragment-length polymorphisms. One example of these types of alleles is the electrophoretic variants of *Gpi1* (DeLorenzo and Ruddle 1969; Padua et al. 1978; Charles and Lee 1980). Mutants from the second category may be useful as disease models, particularly if the original phenotype was noted to mimic a human disease. For example, a spontaneous mutation in *Crygs* (MGI: 2181679) serves as a model for cataracts (Bu et al. 2002). Mutants from the last two categories are often engineered for the express purpose of studying a model for disease. As gene-targeting technologies continue to expand and improve, better and better models for human genetic diseases will become available. This initial list of genes provides a useful starting point for future analyses.

Functional Analysis of Human Disease Orthologs: Gene Ontology

To assess the functional knowledge base about the mouse orthologs of human disease genes, we mined the MGI database for Gene Ontology annotations. The Gene Ontology (GO) is a set of three structured vocabularies that describe the biochemical function, subcellular localization, or global biological process with which a gene product is associated (The Gene Ontology Consortium 2001). MGI curates genes to GO terms using a combination of electronic algorithms and annotation from the literature (Hill et al. 2001). Of the human disease gene orthologs represented in MGI as of August 2002, 811 (86%) had at least one meaningful GO annotation. Of the remainder, four genes were annotated to unknown in all three categories. An unknown annotation in the MGI database means that an MGI curator has examined the gene and based on literature curated at MGI, no functional information is available about the mouse gene. The remaining 125 genes have no MGI annotation, meaning that they have not yet been analyzed by MGI curators, nor have they been assigned a GO annotation via MGI's electronic methods. As in our allele analysis above, this set now provides rich material for further curation. These important genes are being targeted for curation at MGI. A breakdown of the annotations by ontology is shown in Table 2. For a complete gene-by-gene breakdown, see Supplemental Table 7.

DISCUSSION

In this study we examined data for more than 1000 human disease genes and their mouse orthologs. From the FANTOM2 cDNA clone data set, we identified 67 cDNAs representing mouse orthologs to human disease genes, for which no full-length cDNA previously existed at MGI. This information will be useful to mouse geneticists and other researchers investigating the genetic basis of human disease. In this study, we found that 90% (921/1022) of the human disease genes identified in the initial data set were represented at that time in MGI or LocusLink. This underscores the power of cocuration of mouse and human genes between LocusLink and MGI. Of the remaining 10% (101), 80 were represented in the protein BLAST database and analysis at RIKEN. Of these 80 human proteins, 84% ($N = 67$) shared significant sequence similarity to one or more proteins encoded by cDNAs in the FANTOM2 clone set, thus demonstrating the power of a large-scale sequencing project like the Mouse Gene Encyclopedia project to increase the representation of novel mouse cDNAs in the public databases.

In our re-examination of the set of 101 human disease genes (the No Ortholog Set), we identified related protein sequences (some partial) via BLink. We identified 37 candidate orthologs. This result includes candidate orthologs for three human disease genes that had no hits in the FANTOM2 data set analysis. Additionally, querying MGI, we found that several more highly similar mouse sequences have been characterized since May 2002 as genes with sequences (23 genes) in MGI, thus demonstrating the power of using multiple lines of analysis to mine the wealth of data in the public domain. To complete this analysis of identifying mouse orthologs to these human disease genes and storing these data in the public databases, we are evaluating further evidence of orthology to

Table 2. Breakdown of the Number of GO Annotations for Mouse Orthologs of Human Disease Genes by GO Vocabulary

Number of genes with	Molecular function	Biological process	Cellular component
1 annotation	314	219	321
2 annotations	186	234	179
3 annotations	78	139	94
4 annotations	13	71	39
>4 annotations	0	79	27
Unknown	10	10	12
No annotations	340	188	268

The columns represent each GO vocabulary, and the rows represent the number of genes annotated to the vocabulary.

these human disease genes beyond sequence similarity. We are examining shared synteny data, accessible publicly via MGI's Mammalian Homology and Comparative Maps page (http://www.informatics.jax.org/menus/homology_menu.shtml), NCBI's Map Viewer, and NCBI's Human–Mouse Homology Map. Further examination of highly similar proteins is necessary to differentiate between orthology and paralogy, because different proteins within gene families can share a high degree of sequence similarity.

The increasing number of orthologous relationships (mRNA and gene models) between mouse sequences and human disease genes, as seen in our re-evaluation of the May data set and requerying of LocusLink in September 2002, is the result of the continued daily curation efforts of the mouse and human research communities, at NCBI, at MGI, and externally. It illustrates that our view of the transcriptome is still highly dynamic and becomes more and more complete as data are integrated, and that re-evaluation continues to result in additional orthology relationships being identified.

The methods used in this study include a few of the possible approaches for increasing the power of the mouse as a model for human disease. Additional resources to explore include, for example, using LocusLink as a gateway to other NCBI resources. One can take advantage of the large body of computational analyses that are available at the nucleotide level, in HomoloGene (http://www.ncbi.nlm.nih.gov/HomoloGene/) or the Human-Mouse Homology Map (http://www.ncbi.nlm.nih.gov/Homology/), or at the protein level using BLink's display of protein relationships within or between taxa. As well, information related to human diseases and their mouse models can be explored at OMIM's Gene Map (http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=CYP1&start=-2), OMIM's Morbid Map (http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmorbid/), or GeneRIFs attached to records in LocusLink. Computational analyses of the human and mouse genomic sequences and their predicted transcripts is another powerful approach.

METHODS

Identification of Candidate Genes

We first identified a list of 1022 cloned human disease genes in NCBI's May 03, 2002 release of LocusLink using the query "disease_known AND has_seq." Of the 1022 cloned human disease genes, we excluded 29 genes from the BLAST analysis because they did not encode proteins. These 29 gene records included only either partial mRNA or ESTs at the time of the analysis and are indicated in Supplemental Tables 3 and 4. Therefore, the BLAST set included 993 protein sequences. We downloaded this list from NCBI and created a BLAST-able database at RIKEN of the protein sequences for each human gene. For each human disease protein sequence, we identified those FANTOM2 cDNAs that shared a high sequence similarity from a TBLASTN (Altschul et al. 1997) analysis at RIKEN with the FANTOM2 set as our database and the human protein set as our query, with a minimum threshold for a match of e-50.

Looking at each human disease gene manually in LocusLink, we then determined whether a cloned mouse orthologous cDNA had been identified in the MGI curated human–mouse orthology data set (http://www.informatics.jax.org). Additionally, beginning with the data and links provided in LocusLink, we also looked in LocusLink for the mouse gene (based on similar gene name or symbol), in OMIM (Online Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) for listings of mouse orthologs under the Cloning category, and in NCBI's HomoloGene resource (http://www.ncbi.nlm.nih.gov/HomoloGene/) for the curated human and mouse homology data from MGI. We defined a human disease locus as having an existing mouse orthologous sequence only when we could identify a nucleic acid sequence (that was not partial or an EST) associated with the curated mouse gene orthology record in MGI. Our intention was to identify those human disease genes for which an orthologous mouse full-length cDNA clone was not yet annotated in MGI. The scope of this study was limited to the human and mouse curated orthology data sets found in MGI.

ACKNOWLEDGMENTS

We thank Takeya Kasukawa for technical assistance on the analysis and Monica McAndrews-Hill, Janan Eppig, and Donna Maglott for stimulating discussions and critical help in editing. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This project has been supported in part with Federal funds from NIH/NICHD grant HD33745 for the Gene Expression Database, NIH/NHGRI grant HG002273 to the Gene Ontology Project, and NIH/NHGRI grant P40 HG00330 for the Mouse Genome Database.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Beck, J.A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J.T., Festing, M.F.W., and Fisher, E.M.C. 2000. Genealogies of mouse inbred strains. *Nat. Genet.* **24:** 23–25.

Bu, L., Yan, S., Jin, M., Jin, Y., Yu, C., Xiao, S., Xie, Q., Hu, L., Xie, Y., Solitang, Y., et al. 2002. The γS-crystallin gene is mutated in autosomal recessive cataract in mouse. *Genomics* **80:** 38–44.

Charles, D.J. and Lee, C.Y. 1980. Biochemical and immunological characterization of genetic variants of phosphoglucose isomerase from mouse. *Biochem. Genet.* **18:** 153–169.

DeLorenzo, R.J. and Ruddle, F.H. 1969. Genetic control of two electrophoretic variants of glucosephosphate isomerase in the mouse (*Mus musculus*). *Biochem. Genet.* **3:** 151–162.

Denny, P. and Justice, M.J. 2000. Mouse as the measure of man? *Trends Genet.* **16:** 283–287.

The FANTOM Consortium and The RIKEN Genome Exploration Research Group Phase I and II Team. 2002. Analysis of the mouse transcriptome based upon functional annotation of 60,770 full length cDNAs. *Nature* **420:** 563–573.

The Gene Ontology Consortium. 2001. Creating the gene ontology resource: Design and implementation. *Genome Res.* **11:** 1425–1433.

Hill, D.P., Davis, A.P., Richardson, J.E., Corradi, J.P., Ringwald, M., Eppig, J.T., and Blake, J.A. 2001. Biological annotation of mammalian systems: Implementing gene ontologies in mouse genome informatics. *Genomics* **74:** 121–128.

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., et al. 2001. Functional annotation of a full-length mouse cDNA collection. *Nature* **409:** 685–690.

Makalowski, W., Zhang, J., and Boguski, M.S. 1996. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res.* **6:** 846–857.

Padua, R.A., Bulfield, G., and Peters, J. 1978. Biochemical genetics of a new glucosephosphate isomerase allele (Gpi-1c) from wild mice. *Biochem. Genet.* **16:** 127–143.

Pruitt, K.D. and Maglott, D.R. 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29:** 137–140.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., et al. 2000. Comparative genomics of the eukaryotes. *Science* **287:** 2204–2215.

WEB SITE REFERENCES

http://ftp.informatics.jax.org/pub/informatics/reports/GDB_Accession.rpt; Human/Mouse Homology data set at MGI.
http://mgc.nci.nih.gov/; Mammalian Gene Collection.
http://www.ensembl.org/; ENSEMBL.
http://www.gsc.riken.go.jp/e/FANTOM/; RIKEN's Mouse Encyclopedia.
http://www.informatics.jax.org/; The Mouse Genome Informatics (MGI) Database.
http://www.informatics.jax.org/menus/homology_menu.shtml; MGI's Mammalian Homology and Comparative Maps.
http://www.ncbi.nih.gov/cgi-bin/Entrez/blink?pid=4557225&all=1; BLink links on LocusLink page to precomputed protein neighbors.
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM/; Online Mendelian Inheritance in Man (OMIM) database.
http://www.ncbi.nlm.nih.gov/HomoloGene/; NCBI's HomoloGene Home page.
http://www.ncbi.nlm.nih.gov/Homology; NCBI's Human–Mouse Homology Map.
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmap?chromosome=CYP1&start=-2; OMIM's Gene Map.
http://www.ncbi.nlm.nih.gov/htbin-post/Omim/getmorbid; OMIM's Morbid Map.
http://www.ncbi.nlm.nih.gov/LocusLink/; NCBI's LocusLink Home page.

Received November 12, 2002; accepted in revised form January 28, 2003.