Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 364 (2005) 90–98
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Human chromosome 21/Down syndrome gene function and pathway database

Oleksii Nikolaienko [a], Cao Nguyen [b], Linda S. Crinc [c], Krzysztof J. Cios [b,d,e,f], Katheleen Gardiner [a,d,*]

[a] *Eleanor Roosevelt Institute at the University of Denver, 1899 Gaylord Street, Denver, CO 80206, USA*
[b] *Department of Computer Science and Engineering, University of Colorado at Denver and the Health Sciences Center, 1200 Larimer Street, Denver, CO 80217, USA*
[c] *Department of Pediatrics, University of Colorado at Denver and the Health Sciences Center, Biomedical Research Tower RC1-North, Aurora, CO 80045, USA*
[d] *Department of Biochemistry and Molecular Genetics, University of Colorado at Denver and the Health Sciences Center, Biomedical Research Tower RC1-North, Aurora, CO 80045, USA*
[e] *Department of Computer Science, University of Colorado at Boulder, Boulder, CO 80309, USA*
[f] *Department of Preventive Medicine and Biometrics, School of Medicine, University of Colorado at Denver and the Health Sciences Center, 4200 E. 9th Avenue, Denver, CO 80262, USA*

Received 26 February 2005; received in revised form 4 June 2005; accepted 18 July 2005
Available online 28 November 2005
Received by O. Ohara

Abstract

Down syndrome, trisomy of human chromosome 21, is the most common genetic cause of intellectual disability. Correlating the increased expression, due to gene dosage, of the >300 genes encoded by chromosome 21 with specific phenotypic features is a goal that becomes more feasible with the increasing availability of large scale functional, expression and evolutionary data. These data are dispersed among diverse databases, and the variety of formats and locations, plus their often rapid growth, makes access and assimilation a daunting task. To aid the Down syndrome and chromosome 21 community, and researchers interested in the study of any chromosome 21 gene or ortholog, we are developing a comprehensive chromosome 21-specific database with the goals of (i) data consolidation, (ii) accuracy and completeness through expert curation, and (iii) facilitation of novel hypothesis generation. Here we describe the current status of data collection and the immediate future plans for this first human chromosome-specific database.

Keywords: Gene structure; Orthologs; Protein interactions; Sequence annotation; Mental retardation

1. Introduction

Down syndrome (DS) is due to an extra, i.e. third, copy of all or part of the long arm of human chromosome 21. It is the most common genetic cause of intellectual disability (ID), and is also associated with hypotonia, characteristic facial features and the early development of the neuropathology of Alzheimer's Disease (AD) (Hassold and Jacobs, 1984; Epstein, 1995). More variably, it is associated with an increased risk of leukemia, heart defects, abnormalities of the gut, immune system deficiencies, the early development of an AD-like dementia, and a reduced incidence of solid tumors. The severity of each of the phenotypic features is highly variable among individuals. For example, ID ranges from very mild (low normal) to severe (Tolmie, 1997). With an incidence of approximately 1 in 1000 live births, DS remains a significant medical and social problem, and one for which the development of therapeutics would be of wide benefit.

Gene dosage effects, i.e. the 50% increase in expression at the RNA level of trisomic genes, is the initiating cause of the DS

Abbreviations: DS, Down syndrome; AD, Alzheimer's Disease; ID, Intellectual disability; RT-PCR, Reverse transcription coupled polymerase chain reaction.

* Corresponding author. Eleanor Roosevelt Institute at the University of Denver, 1899 Gaylord Street, Denver, Colorado 80206, USA. Tel.: +1 303 336 5652; fax: +1 303 333 8423.
E-mail address: kgardine@du.edu (K. Gardiner).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.07.019

phenotype. Recent experiments using microarrays and quantitative RT-PCR of human DS and mouse model samples indicate that the majority of trisomic genes in the majority of tissues indeed show increased expression, although not always by precisely 50% (Mao et al., 2003; Amano et al., 2004; Kahlem et al., 2004; Lyle et al., 2004; Dauphinot et al., 2005). Complicating factors include differences both greater and less than 50% in comparison of some individual trisomy:control pairs and in some specific gene/tissue combinations. Such variations from strict gene dosage are not inconsistent with observations of expression differences among normal individuals, even genetically identical individuals, and the variability in the severity of the DS phenotype (Gardiner, 2004).

The key focus in DS research is the correlation of the increased expression, and therefore the level of protein activity, of specific chromosome 21 genes or sets of genes with specific phenotypic features, with the long-range goal of developing therapeutics to prevent or ameliorate the phenotype. This is a particularly difficult task because of the large number of genes that must be considered, essentially the entire ~350 genes contained with 21q. Effective consideration of therapeutics requires assimilation of as much information as possible for each gene. Currently this demands: (i) accurate descriptions of gene numbers, their structures and alternative splicing, (ii) descriptions of their functions, including targets and substrates, and the pathways or complexes in which they function, and (iii) tissue, cell line and developmental expression patterns of chromosome 21 genes and the correlations of these with expression of other genes both on and off chromosome 21.

Most facets of these data are not static and are not expected to be for some time to come. Gene lists need frequent updating and revisions, in part to keep up with novel entries in dbEST which add both new gene models and alternative splice variants of existing gene models. Gene lists also need curation. For example, Imanishi et al. (2004) recently reported 197 cDNAs mapping to chromosome 21. Comparison of this list with that of Gardiner et al. (2003) revealed 53 novel models, each of which was intronless, largely composed of repetitive sequences, and not conserved in mouse. It is useful to flag such cDNAs for their non-standard, ambiguous nature because hypotheses regarding such genes are likely to differ from those for highly conserved, functionally annotated protein coding genes. Information on gene function is also accumulating quickly. Individual gene reports from the last few years describe more than two dozen targets and four interactors of the chromosome 21 transcription factor, GABPA (reviewed in Rosmarin et al., 2004) and 12 substrates/interactors for the chromosome 21 kinase, DYRK1A (e.g. Woods et al., 2001; de Graaf et al., 2004; Sitz et al., 2004). Proteomics approaches are defining the compositions of cellular complexes, such as the nucleolus, the spliceosome, the SMAD signaling pathway (Scherl et al., 2002; Rappsilber et al., 2002; Colland et al., 2004), among others. Each report must be examined for the presence of chromosome 21 encoded proteins. By this means, C21orf66 and c21orf70, otherwise functionally unannotated, were found to be associated with the spliceosome. Such information can be used to predict, or speculate upon, the consequences of over-expression in trisomy.

Large scale projects involving model organisms are a rich source of information on chromosome 21 orthologous proteins. For example, Amsterdam et al. (2004) recently reported on embryonic lethal knockouts in zebra fish that included 5 chromosome 21 orthologs, among them the functionally unannotated c21orf59 and the "novel nuclear protein 1", NNP1. From *C. elegans* and yeast, the interactomes contain data on novel interactions of numerous chromosome 21 orthologs (Tong et al., 2004; Li et al., 2004). While phenotypes of knockouts do not imply the phenotype of over-expression and conservation from *C. elegans* to human of interaction networks cannot be assumed, awareness of such data provides fuel for prediction and testing of normal human function and consequences of over expression.

Important expression information includes not only the basics that can be obtained from dbEST and SAGE experiments found in the NCBI GEO database, but also that from literature reports of directed microarray experiments. A recent report on expression changes in normal aging showed that four chromosome 21 genes (cystathionine beta-synthetase; the transcription factor, OLIG1; the calcium sensor S100β; and the cholesterol transporter ABCG1) increased in expression, while the chromosome 21 transcription factor, ETS2, decreased in expression, when frontal cortex from adults <42 years was compared with those >73 years of age (Lu et al., 2004). In experiments comparing expression levels between human and chimpanzee cortex, the functionally unannotated, c21orf33, the receptor protein CXADR, and the collagen *COL6A1* genes were found to be expressed >5 fold higher in human (Caceres et al., 2003; Preuss et al., 2004; Khaitovich et al., 2004). Lastly, the wealth of microarray data suggests the potential usefulness of re-analysis. For example, Segal et al. (2004) described modules of correlated gene expression differences derived from microarray screenings of large numbers of normal and abnormal cell lines and tissues. Perhaps appropriate re-analysis of such data can predict additional correlations with and among sets of chromosome 21 genes and other cellular processes.

Data on normal function and expression can be used to predict the consequences of trisomy manifested as perturbations of biochemical, cellular and signaling pathways, to design additional mouse models of DS to explore regulation of the perturbations, and to aid in development of pharmacological agents to control these perturbations and manipulate mouse models. These data, of course, must be used in conjunction with detailed and consistent information on the cellular, developmental and cognitive phenotype of human DS, and its potential variation due allelic variation either on chromosome 21 or in the genomic background.

Data relevant to chromosome 21 and DS are dispersed among diverse databases. The variety of formats and locations of chromosome 21 gene information, plus their evolving nature, makes access and assimilation a time consuming and challenging process, yet it remains one that is critical for effective DS research. To aid the DS/chromosome 21 research community and researchers interested in individual chromosome 21 genes or pathways, we are developing a comprehensive chromosome 21-specific database with the goals of (i) data consolidation, (ii)

expert curation and (iii) hypothesis generation. Here we describe the current state of the database and the future plans.

2. Methods and database design

The database unifies molecular information on genes from human chromosome 21 and the orthologous regions on mouse chromosomes 16, 17 and 10. Initial gene lists for both organisms were those developed from genomic sequence annotation as described in Gardiner et al. (2003). These are updated based on review of new entries in dbEST, chromosome 21 genes in ENTREZ (http://www.nicb.nlm.nih.gov), and the recent H-Inv database (Imanishi et al., 2004; http://www.h-invitational.jp/). Sequence data were obtained from NCBI RefSeq, Unigene and dbEST, and SwissProt. Ortholog information was based on Homologene, plus BLASTP searches of chromosome 21 proteins versus chimpanzee, dog, chicken, zebra fish, *Drosophila*, *C* elegans and yeast, followed by reciprocal human searches. RNAi data were obtained from Flybase (http://www.flybase.org; Drysdale et al., 2005), Wormbase (http://wormbase.org release WS120 1/3/04) and the yeast database (Dolinski et al., 2004) (http://www.yeastgenome.org). Protein interaction data for *Drosophila*, *C. elegans* and *S. cerevisiae* were obtained from "The Grid" (http://biodata.mshri.on.ca/grid/servlet/Index); mammalian protein interaction data were obtained from the Database of Interacting Proteins (DIP: http://dip-doe-mbi.ucla.edu). Predictions of protein functions based on functions of their interacting proteins were derived using the Markov random field and Bayesian belief network of Deng et al. (2003). Data on post-translational modifications, substrates and targets were extracted from PubMed and, with the gracious assistance of Akhilesh Pandey, from HPRD (http://www.hprd.org).

The database was designed using a relational model, implemented in the open source MySQL database. A combination of HTML/Perl/Java languages were used to extract information from databases listed above. The high level model, in terms of the entity relationship (E-R) diagram is shown in Fig. 1 and detailed in Table 1.

3. Results and discussion

The database home page (http://chr21db.cudenver.edu) provides access to several search tools: basic and advanced searches, lists of gene tables, and genomic sequence annotation

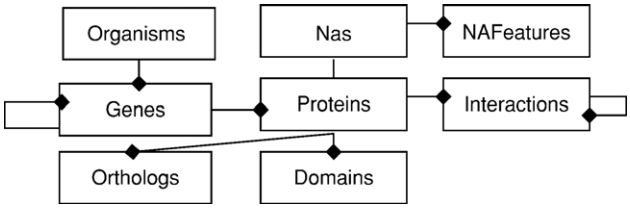

Fig. 1. Entity relationship diagram. Nas, Nucleic acid sequences; Naf, Nucleic acid features.

(Genotator) are currently available; mouse resources and curation information will be available by January 2006.

3.1. Simple searches (Fig. 2)

A general search provides complete lists of chromosome 21 genes and genes in orthologous mouse genomic regions (Fig. 2). Searches for individual genes can be done by gene name, symbol or accession number. This retrieves, for the human or the mouse gene, a list of aliases, cytogenetic map location and genomic position, plus links to the genes located immediately centromeric or telomeric and the mouse or human ortholog. A schematic showing the exon/intron organization is provided for up to ten cDNAs and/or spliced ESTs, with nucleic acid and protein sequence accession numbers. Exon boundaries are correlated with those of the encoded protein domains, facilitating identification of domain alterations caused by alternative splicing.

Literature references (PubMed) include those reporting the gene identification or structural or organizational features and those describing functional features (targets, substrates, interactions, phenotypes of mouse mutants).

The last section provides information on orthologs, accession numbers, comparisons of lengths and % similarity, plus schematics comparing functional domains. When available, this section also summarizes and provides links to results of large-scale experiments describing phenotypes from RNAi and knockouts.

3.2. Advanced searches (Fig. 3)

Three types of searches are available. A search on a gene name provides additional basic gene information: exon/intron structures and accession numbers of all cDNAs, spliced ESTs and encoded proteins. Exon size, nucleotide start and stop number and coding capacity are obtained by moving the cursor over the exon of interest (Fig. 3a). Clicking on the protein accession number provides links to protein sequence and amino acid characteristics (Fig. 3b) and domain composition.

A second search is for protein–protein interactions, which can be provided in table or graphical form. Names, descriptors and Genbank identification numbers are provided for each protein at each level of interaction; level 4 interactions (levels indicate the number of interaction steps from the chromosome 21 protein) for human synaptojanin, SYNJ1, are shown in Fig. 3c. Currently, human, rat and mouse proteins are not separated in the graphics. Level 2 interaction networks are provided for orthologous proteins from *Drosophila*, *C. elegans* and yeast, with links to the human ortholog for each interactor. Conservation in mammals of networks identified in simpler model organisms cannot be assumed, but is provided for use in hypothesis generation or as potential support of novel experimental data.

An option for protein function prediction is provided in the mammalian interaction search results. Applying the method of Deng et al. (2003), using the theory of Markov

Table 1
Details for entity relationship diagram

Attribute	Type	Description
Genes		
GeneID	bigint(10)	Gene ID, primary key
OrganismID	smallint(10)	Organism ID, link to Organisms
Name	varchar(32)	Gene name
Strand	char(1)	Strand, + or −
Chromosome	char(2)	Chromosome
Aliases	varchar(254)	Alternative gene names
Description	varchar(255)	GenBank description
ChrStart	bigint(10)	Nucleotide gene start in chromosome genomic sequence
ChrStop	bigint(10)	Nucleotide gene stop in chromosome genomic sequence
Contig	varchar(32)	Contig
OrthGeneID	bigint(10)	Gene's ortholog gene ID, link to Genes
Organisms		
OrganismID	smallint(10)	Organism ID, primary key
Organismname	varchar(127)	Organism name
Nucleic acid sequences (Nas)		
NAID	int(10)	Nucleotide acid ID, primary key
GeneID	bigint(10)	Gene ID, link to genes
AccNo	varchar(24)	GenBank accession number
NAType	varchar(4)	Nucleotide acid type: cDNA or EST
NALength	int(10)	Nucleotide acid length
CDSStart	int(10)	Coding start location
CDSStop	int(10)	Coding stop location
Nucleic acid features (Naf)		
FeatureID	bigint(20)	Feature ID, auto increment, primary key
NAID	int(10)	Nucleotide acid ID, link to Nas
FeatureName	varchar(4)	EXON, UTR5′ or UTR3′
RelChrStart	int(10)	Relational start location on this nucleotide acid
FeatureLength	varchar(4)	Feature length
Protein sequences		
SWProtID	varchar(32)	Swiss protein ID
ProtID	bigint(20)	Protein ID, primary key
GeneID	bigint(10)	Gene ID, link to genes
AccNo	varchar(24)	GenBank accession number
ProtLength	int(10)	Protein length, in amino acids
ProtStruct	text	Protein sequence
Description	varchar(255)	GenBank description of protein
NAAccNo	varchar(24)	GenBank nucleotide acid accession number, link to Nas
Protein domains		
DomainID	bigint(20)	Domain ID, auto increment, primary key
ProtID	bigint(20)	Protein ID, link to Proteins
Name	varchar(32)	Domain name
Start	int(10)	Start location within amino acid sequence
Length	int(10)	Domain length, in amino acids
EValue	float	Expected value
Ortholog proteins		
OrthID	bigint(20)	Ortholog id, auto increment, primary key
OrthProtID	bigint(20)	Ortholog protein ID, link to proteins
ProtID	bigint(20)	Protein ID, link to proteins
PercSimil	smallint(6)	Percent similarity
Protein–protein interactions		
InteractionID	bigint(20)	Interaction ID, auto increment, primary key
BaitID	bigint(20)	Bait protein ID, link to proteins
PreyID	bigint(20)	Prey protein ID, link to proteins
Experiment	varchar(255)	Interaction experiment

random fields, a protein's function is inferred from the functions of the components of its experimentally defined interaction network. Bayesian approaches are used to assign a probability to each functional prediction, indicating the level of confidence in the result. Deng et al. (2003) demonstrated the efficacy of this approach by analysis of the yeast interactome using a leave-one-out method. Clearly, all results must be considered as predictions and their use guided by biological intuition.

In the third type of search, GeneQuest, complex queries can be formulated. For example: list all chromosome 21 genes with orthologs in zebra fish, and the accession numbers and protein domain structures of each; list all proteins encoding pleckstrin domains; list all orthologs of Intersectin and describe the domain organization of each.

3.3. Gene tables

Basic gene tables published as supplementary material in Gardiner et al. (2003), and subsequently updated, are found here. These include lists of all chromosome 21 genes, all genes found in the orthologous mouse genomic regions, all genes with functional annotation, and genes that appear to not be conserved in mouse. Each table provides the gene name(s), accession number, chromosomal location, and a brief description.

3.4. Genomic sequence annotation

Graphical annotation of chromosome 21 and orthologous regions of mouse chromosomes 16, 17 and 10 are presented in 170 kb segments using Genotator software (Fortna and Gardiner, 2001). Results of standard gene finding analyses include coding exon predictions, spliced ESTs, CpG islands, GC content, repeat sequence identification, conservation in mouse/human, and gene structures/models. Color coding facilitates viewer evaluation of gene structures, based on open reading frame, repeat sequence content, relationship with known protein coding genes, presence in dbEST and antisense transcripts. An example is shown in Fig. 4. Results of expression analysis with genome tiling arrays and transcription factor binding sites determined by ChIP-Chip experiments are also provided (Kapranov et al., 2002; Cawley et al., 2004).

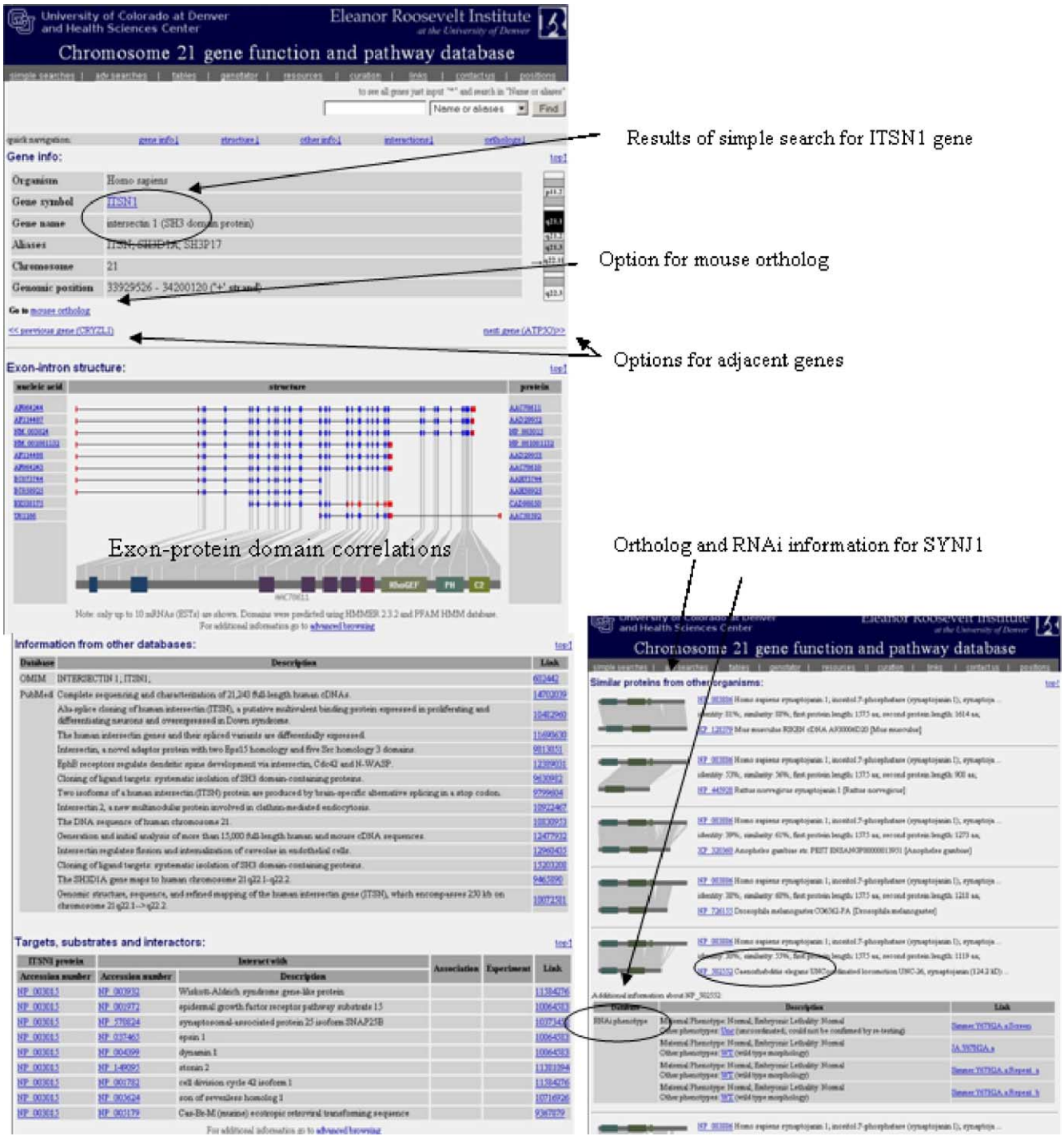

Fig. 2. Web site: simple search results. Results of a simple search on the gene name intersectin, ITSN (human) are shown, indicating options to link to the mouse ortholog, or to adjacent genes on the chromosome, and exon/intron structure correlated with protein domains, references to gene identification and interacting proteins. Orthologs and links to RNAi data are shown for the synaptojanin, SYNJ1, gene.

3.5. Additional features currently under development (completion scheduled for January 2006)

3.5.1. Curation

Automated data collection is the basic tool for creation of the database. Errors and incomplete information are expected, and may be especially relevant to identification of gene models, alternative splicing patterns, and features of experimental support. It is also a challenge to remain current on protein functional and expression information published in reports of both individual genes and large scale proteomics and microarray analysis. Because accurate and complete information is to be a critical and distinguishing feature of this database, manual curation will be carried out. An international team of curators who are experts in the aspects of Down syndrome, chromosome 21 gene functions, or mouse models is

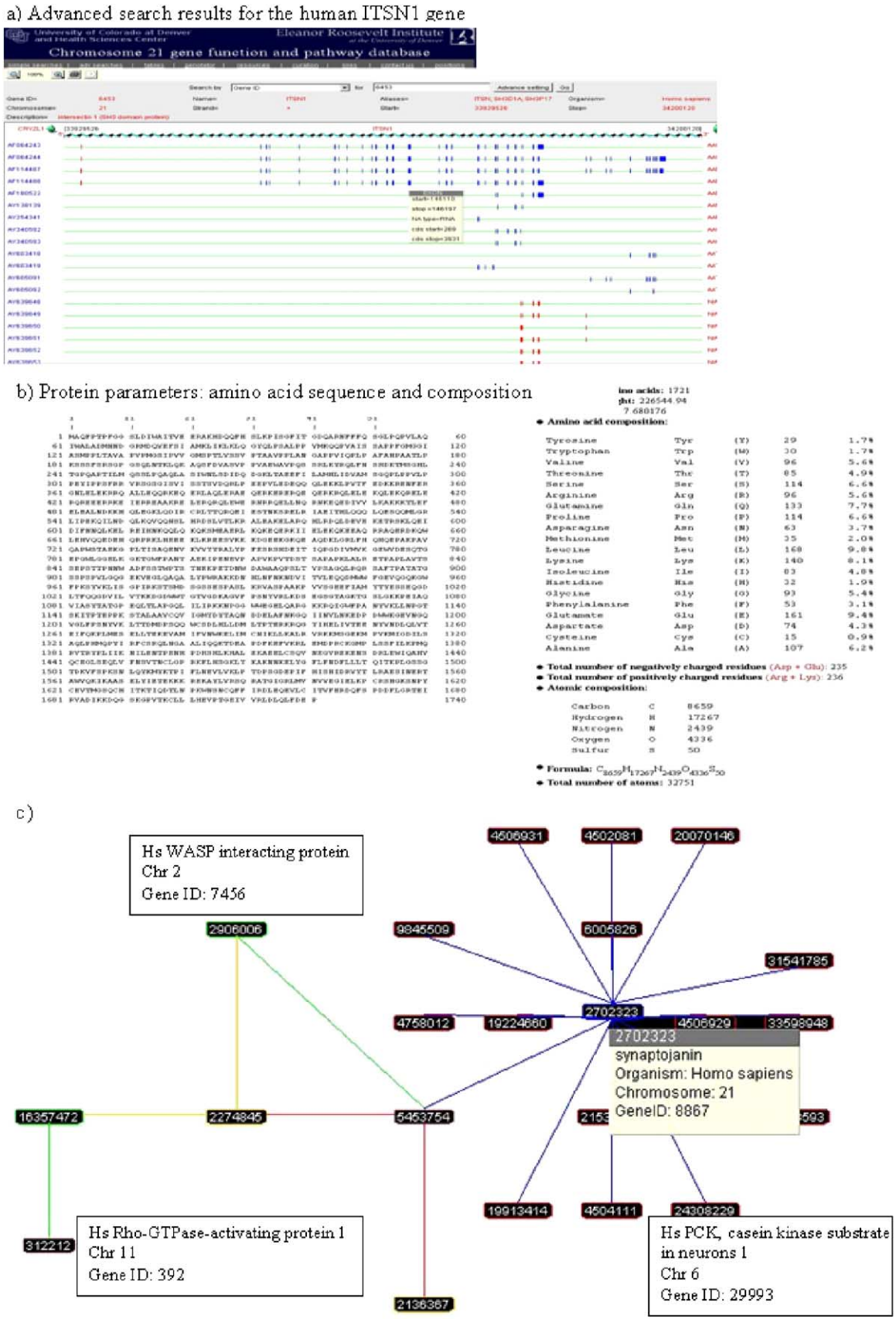

Fig. 3. Advanced search for a) gene structural features and b) protein features for the human intersectin gene, ITSN1; c) level 4 protein interactions for human synaptojanin 1, SYNJ1; gi numbers (boxed) are given for each interacting protein. Positioning the cursor over the gi number provides the gene name, organism, chromosome location and Gene ID number. Examples are shown.

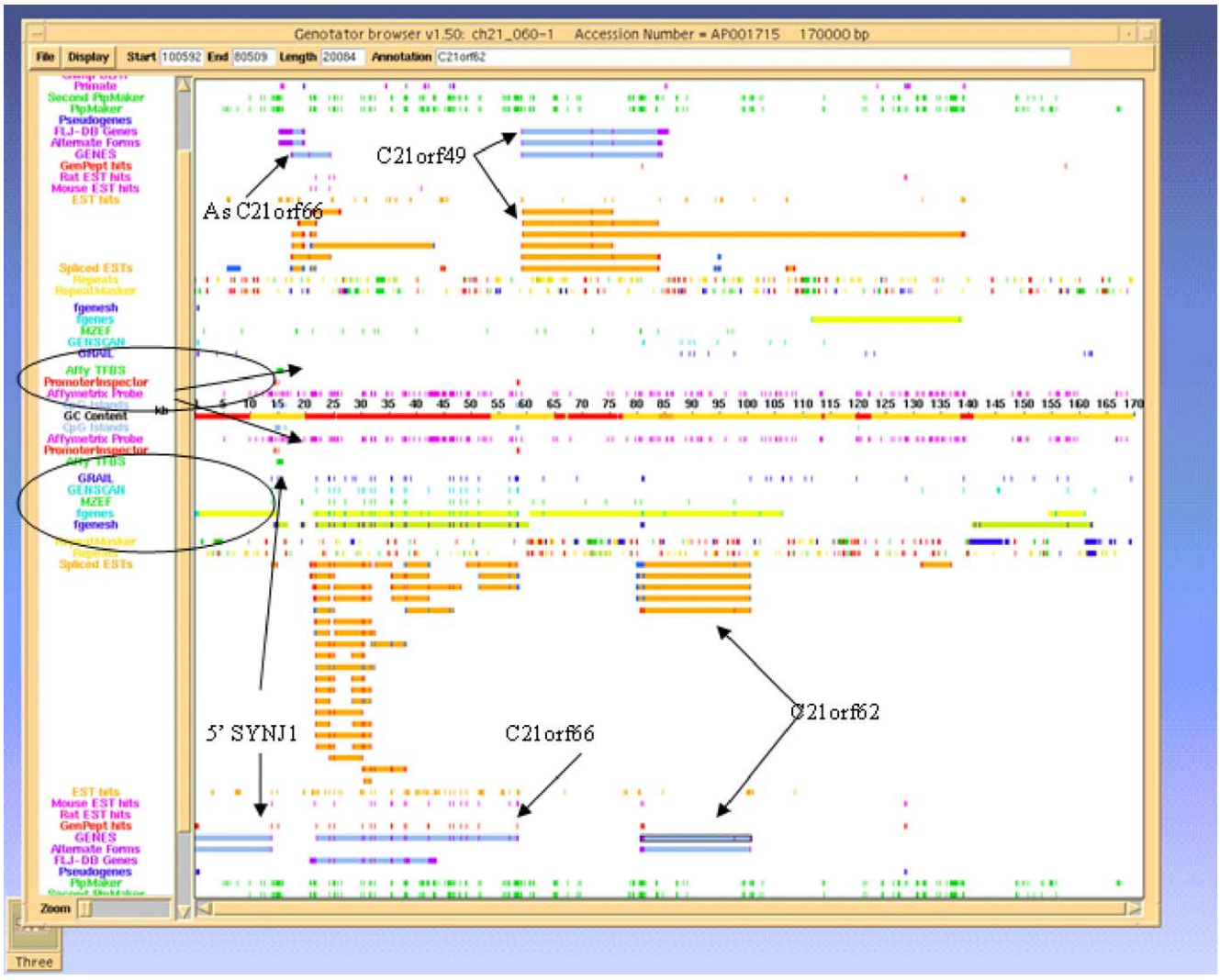

Fig. 4. Genomic sequence annotation for a segment of chromosome 21 spanning 170 kb. The DNA sequence is represented by the horizontal center line. Annotations are shown for the forward strand (centromere to telomere) above the line, and the reverse strand, below. Genes structures are indicated for the 5′ end of the *SYNJ1* gene, and for C21orf66 and c21orf63 on the reverse strand, and for a gene antisense to C21orf66 and for C21orf49 on the forward strand. Support for gene structures is given by spliced ESTs and, for c21orf66, by coding exon prediction (Grail, Genscan, Mzef, Fgenes and FgenesH). Other standard features include repeat sequence annotation and conservation in mouse (indicated by PipMaker). Results of expression analysis by genomic tiling arrays are indicated by "Affymetrix probe" (Kapranov et al., 2002) and transcription factor binding sites by ChIP-chip analysis by Affy TFBS (Cawley et al., 2004). Locations of CpG islands and predicted promoters (PromoterInspector) are indicated.

being established. Curators will review information in the database and the literature and post quarterly updates and corrections. Data from the Human Reference Protein Database, which is also manually curated by experts, will be incorporated regularly.

3.5.2. Resources

Mouse resources include embryonal stem (ES) cell lines containing gene-trap insertions (International Gene Trap Consortium http://www.igtc.ca). Lists of chromosome 21 orthologs for which ES cell lines have been identified and contacts for obtaining them will be maintained, searchable by gene name or alias. DS mouse models, knockouts, transgenics and segmental trisomies, are typically reported in the literature. To facilitate additional phenotypic analyses, information on all models will be provided.

3.5.3. Expression

This information will be available as a fourth option under Advanced Searches. Data will include: summaries of human and mouse gene expression based on data in dbEST/Unigene, links to the Mouse Gene Expression Database at the Jackson Laboratory (http://www.informatics.jax.org), and links to the BODYMAP human (http://bodymap.ims.u-tokyo.ac.jp). User friendly queries and effective formats for presenting SAGE and microarray data (Gene Expression Omnibus, NCBI) will be developed.

3.5.4. SNPs

We will add a new advanced search category for SNP annotation. For this, we will consider the usefulness of two databases: FESD, Functional Element SNPs Database (http://combio.kribb.re.kr/ksnp/resd), and SNPeffect, molecular

phenotype effects of non-synonymous coding region SNPs (http://snpeffect.vib.be). SNP data will be linked to protein domain compositions. SNP data will be valuable for future correlations of genotype with phenotypic variation.

3.5.5. Chimpanzee

For each chromosome 21 gene, tables will provide the percent identity of the chimpanzee protein, the amino acid number and domain location of sequence differences, predictions of structural and functional consequences, and expression level differences from microarray analyses. Searches by chromosome 21 gene name, accession number, alias, or Unigene ID will be possible.

3.6. Future developments

3.6.1. Pathways

Schematics of pathways that directly or indirectly involve or are impacted by chromosome 21 genes will be provided. These will come from databases such as the Kyoto Encyclopedia of Genes and Genomes (http://www.kegg.com) as well as being developed from functional and interaction data of chromosome 21 genes. Pathways will be searchable by pathway and chromosome 21 gene name.

3.6.2. Resources

Linking polymorphisms within specific genes to specific features (e.g. Kerstann et al., 2004) will require collection of DNA samples from individuals with DS and their families. References and links to collections, as they become available, will be provided, along with SNP data for chromosome 21 genes, correlated with protein sequence and functional consequences.

3.6.3. Expression

Tables of data from Serial Analysis of Gene Expression (SAGE) experiments will be generated to provide expression levels of chromosome 21 genes and to identify experiments with common sets of chromosome 21 genes. Links will be provided to the data in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo).

Tools will be developed for interpreting microarray data from a DS-relevant perspective. This will include re-analysis of microarray data to identify correlations of expression pattern changes among sets of chromosome 21 genes and among chromosome 21 genes and non-chromosome 21 genes. The goal will be the prediction of new pathway associations and gene networks relevant to chromosome 21 genes.

Acknowledgements

Database development has been supported by the National Down Syndrome Society (KG), the Coleman Institute for Intellectual Disabilities (LSC), the Colorado Institute of Technology (KC) and the Ministry of Education and Training – Vietnam (CN).

References

Amano, K., et al., 2004. Dosage-dependent over-expression of genes in the trisomic region of Ts1Cje mouse model for Down syndrome. Hum. Mol. Genet. 13, 1333–1340.

Amsterdam, A., Nissen, R.M., Sun, Z., Swindell, E.C., Farrington, S., Hopkins, N., 2004. Identification of 315 genes essential for early zebra fish development. Proc. Natl. Acad. Sci. U. S. A. 101, 12792–12797.

Caceres, M., et al., 2003. Elevated gene expression levels distinguish human from non-human primate brains. Proc. Natl. Acad. Sci. U. S. A. 100, 13030–13035.

Cawley, S., et al., 2004. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116, 499–509.

Colland, F., et al., 2004. Functional proteomics mapping of a human signaling pathway. Genome Res. 14, 1324–1332.

Dauphinot, L., et al., 2005. The cerebellar transcriptome during postnatal development of the Ts1Cje mouse, a segmental trisomy model for Down syndrome. Hum. Mol. Genet. 14, 373–384.

de Graaf, K., et al., 2004. Characterization of cyclin L2, a novel cyclin with an arginine/serine-rich domain: phosphorylation by DYRK1A and colocalization with splicing factors. J. Biol. Chem. 279, 4612–4624.

Deng, M., Zhang, K., Mehta, S., Chen, T., Sun, F., 2003. Prediction of protein function using protein–protein interaction data. J. Comput. Biol. 10, 947–960.

Dolinski, K., et al., "Saccharomyces Genome Database" http://www.yeastgenome.org/ (7/2004-9/2004).

Drysdale, R.A., Crosby, M., and The FlyBase Consortium, 2005. FlyBase: genes and gene models. Nucleic Acids Res. 33, D390–D395 (http://flybase.org/).

Epstein, C.J., 1995. Down syndrome (trisomy 21). In: Scriver, C.A., et al. (Ed.), Metabolic and Molecular Bases of Inherited Disease. McGraw Hill, New York, pp. 749–794.

Fortna, A., Gardiner, K., 2001. Genomic sequence analysis tools: a user's guide. Trends Genet. 17, 158–164.

Gardiner, K., 2004. Gene–dosage effects in Down syndrome and trisomic mouse models. Genome Biol. 5, 244–247.

Gardiner, K., Fortna, A., Bechtel, L., Davisson, M.T., 2003. Mouse models of Down syndrome: how useful can they be? Comparison of the gene content of human chromosome 21 with orthologous mouse genomic regions. Gene 318, 137–147.

Hassold, T.J., Jacobs, P.A., 1984. Trisomy in man. Annu. Rev. Genet. 18, 69–97.

Imanishi, T., et al., 2004. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. PLoS Biol. 2, e162.

Kahlem, P., et al., 2004. Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of Down syndrome. Genome Res. 14, 1258–1267.

Kapranov, P., et al., 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science 296, 916–919.

Kerstann, K.F., et al., 2004. Linkage disequilibrium mapping in trisomic populations: analytical approaches and an application to congenital heart defects in Down syndrome. Genet. Epidemiol. 27, 240–251.

Khaitovich, P., et al., 2004. Regional patterns of gene expression in human and chimpanzee brains. Genome Res. 14, 1462–1473.

Li, S., et al., 2004. A map of the interactome network of the metazoan *C. elegans*. Science 303, 540–543.

Lu, T., et al., 2004. Gene regulation and DNA damage in the ageing human brain. Nature 429, 883–891.

Lyle, R., Gehrig, C., Neergaard-Henrichsen, C., Deutsch, S., Antonarakis, S.E., 2004. Gene expression from the aneuploid chromosome in a trisomy mouse model of down syndrome. Genome Res. 14, 1268–1274.

Mao, R., Zielke, C.L., Zielke, H.R., Pevsner, J., 2003. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. Genomics 81, 457–467.

Preuss, T.M., Caceres, M., Oldham, M.C., Geschwind, D.H., 2004. Human brain evolution: insights from microarrays. Nat. Rev., Genet. 5, 850–860.

Rappsilber, J., Ryder, U., Lamond, A.I., Mann, M., 2002. Large-scale proteomic analysis of the human spliceosome. Genome Res. 12, 1231–1245.

Rosmarin, A.G., Resendes, K.K., Yang, Z., McMillan, J.N., Fleming, S.L., 2004. GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein-protein interactions. Blood Cells Mol. Dis. 32, 143–154.

Scherl, A., et al., 2002. Functional proteomic analysis of human nucleolus. Mol. Biol. Cell 13, 4100–4109.

Segal, E., Friedman, N., Koller, D., Regev, A., 2004. A module map showing conditional activity of expression modules in cancer. Nat. Genet. 36, 1090–1098.

Sitz, J.H., Tigges, M., Baumgartel, K., Khaspekov, L.G., Lutz, B., 2004. Dyrk1A potentiates steroid hormone-induced transcription via the chromatin remodeling factor Arip4. Mol. Cell. Biol. 24, 5821–5834.

Tolmie, J.L., 1997. Down syndrome and other autosomal trisomies. In: Rimoin, D., O'Connor, J.M., Pyeritz, R.E., Emergy, A.E.H. (Eds.), Principles and Practices of Medical Genetics, 3rd edn. W.B. Saunders, Livingstone, Scotland, UK, pp. 925–971.

Tong, A.H., et al., 2004. Global mapping of the yeast genetic interaction network. Science 303, 808–813.

Woods, Y.L., et al., 2001. The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bepsilon at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. Biochem. J. 355 (Pt 3), 609–615.