Review

The sequence of human chromosome 21 and implications for research into Down syndrome
Katheleen Gardiner* and Muriel Davisson†

Addresses: *Eleanor Roosevelt Institute, Gaylord Street, Denver, Colorado 80206-1210, USA. †The Jackson Laboratory, Main Street, Bar Harbor, Maine 04609-1500, USA.

Correspondence: Katheleen Gardiner. E-mail: gardiner@eri.uchsc.edu

Abstract

The recent completion of the DNA sequence of human chromosome 21 has provided the first look at the 225 genes that are candidates for involvement in Down syndrome (trisomy 21). A broad functional classification of these genes, their expression data and evolutionary conservation, and comparison with the gene content of the major mouse models of Down syndrome, suggest how the chromosome sequence may help in understanding the complex Down syndrome phenotype.

Down syndrome (DS), affecting one in 700 live births, is the most common genetic cause of mental retardation [1]. The phenotype of DS is complex and variable in severity among individuals; it includes mental retardation and cognitive deficits, heart defects, hypotonia, motor dysfunction, immune system deficiencies, an increased risk of leukemia, and development of the pathology of Alzheimer's disease [2]. Most commonly, DS is due to the presence of an extra copy of a complete chromosome 21 and it is assumed that the DS phenotypic features are a direct consequence of the overexpression of some number of genes contained within 21q (21p is largely made up of ribosomal RNA genes and other repeat sequences). Recently, the essentially complete sequence of 21q - 33.5 Mb - was finished, and 225 genes were identified by the application of a variety of experimental and computer-based approaches [3]. The availability of this massive amount of new data has immediate importance to DS research. This review discusses the following issues: the reliability of gene identification; what is known or can be inferred about the biological function of the 225 identified genes; expression patterns of the novel genes; evolutionary conservation of, in particular, those genes lacking functional associations; inferences about the gene content of the major mouse models of DS and therefore the causes of the phenotypic differences among them; and reasonable next steps towards the goal of understanding the gene-phenotype relationships in DS. Throughout the following discussions, references to numbers and kinds of genes and additional analyses of 21q gene content are based on the data presented in [3].

Gene number
Two hundred and twenty-five is a surprisingly small number for the complete gene content of approximately 1% of the human genome. It is significantly less than 1% of the 50,000-100,000 genes previously estimated in total for the human genome (see also [4]) and it is significantly less than the 545 genes identified on chromosome 22 in approximately the same amount of DNA [5]. Previous data from the mapping of expressed sequence tags (ESTs) and genes, and efforts at cDNA selection, have consistently suggested that chromosome 21 was relatively gene-poor overall, and extremely so in some regions [6,7]. It could also be predicted that chromosome 21 would have fewer genes than chromosome 22. Approximately half of chromosome 21 is a large dark band when stained with Giemsa, and such bands are known to be gene-poor, while chromosome 22 is almost entirely comprised of gene-rich R bands [8,9]. In addition, trisomy 21 is compatible with life, while trisomy 22 is not [1]. Chromosome 21, therefore, was expected to be relatively

gene-poor. Its extreme paucity of genes, however, justifies further consideration. In particular, are there consistent errors or weaknesses in gene-finding techniques that could have missed a significant proportion of genes? To see where errors may have accumulated, it is worth reviewing the gene identification methods.

Genes were identified on the basis of the following types of data: identities or similarities to known proteins; identities to spliced ESTs; and patterns of consistent coding-exon prediction. First, protein matches identified genes that were identical or similar to known genes, and also found pseudogenes. With some minor corrections, all 107 genes associated with complete cDNAs that had been mapped previously to chromosome 21 (listed by Swiss-PROT [10] in March 2000) were found. In addition, within 21q, 52 protein matches were classed as pseudogenes on the basis of a lack of introns and, most importantly, on the presence of multiple in-frame stop codons. Given the inability of transcripts from these genes to produce a complete protein, it is unlikely that any pseudogenes were incorrectly classified. Secondly, for EST matches, only those that showed evidence of splicing were used - that is, those that were non-contiguous with genomic sequence, showed consensus splice sites, and represented essentially perfect matches (>95% identity) to the genomic sequence. This eliminates many of the artifacts common to cDNA libraries. A survey of the EST database [11] for fifty of the known chromosome 21 genes found that forty-three were present as spliced ESTs, six were present only as unspliced ESTs (five of these were intronless genes), and one was not present in dbEST.

Finally, the criterion of consistent exon prediction required that two of the three coding-exon prediction programs (Grail, Genscan and MZEF) agreed on the location of an exon, and that a minimum of three consistent exons were found within <60 kb, with introns <30 kb. It is noteworthy that the coding regions of intronless genes were well predicted but only as single exons. Such exons tend to be very large - greater than a kilobase (kb) in length - in contrast to typical coding exons that average 100-150 base pairs (bp). After making exceptions for, and including, large single-coding-exon genes, by these criteria, all but one of the 107 known genes could be identified by exon prediction. This included very large genes, such as *DSCAM*, which spans >800 kb, and *GRIK1*, which spans >400 kb, both of which were well predicted through at least some of their coding regions.

The important conclusion here is that each of the 107 genes previously known to map to chromosome 21 would have been identified, in the absence of protein similarities, by the criteria of EST matches plus exon prediction. These criteria do not, in most cases, define a complete gene structure, but they do successfully indicate the presence of a gene. Thus, unless novel genes have very different characteristics, it is reasonable to expect a similarly high level of

success in their identification. Using these criteria a further 118 genes were identified.

What is likely to have been missed? First, there are gaps in the sequence of 21q. They are few (three) and small (<50 kb each), however, and therefore cannot harbor large numbers of genes. Second, genes that would not be identified would have to possess the following features: no similarity to any known protein; consistently very large introns (>30-60 kb), so that patterns of predicted exons would not be scored; and long intronless 3′ untranslated regions (UTRs) or restricted and/or low expression levels, so that no spliced EST is present in dbEST. It is certainly possible that some number of genes with such characteristics exist; that they represent a significant proportion of chromosome 21 genes is unlikely, however. The distal one third of 21q is the most gene-rich (and GC-rich); but intergenic distances here are not large enough to accommodate additional genes with uniformly large introns. So, unless coding exons in these genes are for some reason not recognized, such genes would be scored on the basis of patterns of predicted exons. The proximal two thirds, in contrast, is uniformly AT-rich and does have large segments lacking gene features; indeed, there is one segment of approximately 7 Mb that harbors only seven genes. Here there is room for numerous genes that have large introns and restricted expression. One argument against this is a biological one: an individual who is monosomic for this region has only mild phenotypic abnormalities [12]. A second argument is a general scarcity of any consistent exon prediction in the region, regardless of 'intron' size. If there are many coding exons within this region, they must also be largely unrecognized by prediction programs. Together, these data suggest that the total of 225 genes is likely to be reliable: false negatives should be few.

But what about the possibility of false positives? Genes with complete protein or cDNA sequences identical or highly similar to known genes (these are the class 1 and class 2 genes in [3]) are unambiguous. Gene models (classes 3 and 4), however, are still open to further investigation and interpretation. For example, some investigators will choose to disregard a specific match to a protein domain if the similarity is weak. How many exons to include in a model, and whether an EST should be included will also sometimes be debatable. Thus, details in the gene catalog of 21q should be considered provisional. Investigators should review the basis for specific gene predictions of interest (available at [13]).

The nature of chromosome 21 genes

DS can be considered as a contiguous gene syndrome, with almost the entirety of 21q the relevant region. The segment of 21q22.2 that is referred to as the Down syndrome chromosomal region (DSCR) was defined to contain genes relevant to aspects of the DS phenotype on the basis of the phenotypes of several cases of partial trisomy 21 [14,15]. Data

using a larger number of partial trisomy cases showed that only the most centromeric region of 21q could be excluded from containing relevant genes, in particular for mental retardation [16]. It is assumed that overexpression of chromosome 21 genes, as a result of their presence in an extra copy, causes the DS phenotype. Are all chromosome 21 genes overexpressed? Can overexpression of some genes be tolerated with no phenotypic effect? How many genes are overexpressed and relevant? Currently, there are no answers to these questions. It is, however, worth considering what is known about the function of chromosome 21 genes.

Table 1 lists the 122 genes for which some functional association can be inferred. Functional inferences are based on partial or complete similarities of the chromosome 21 genes or gene models to proteins or protein domains for which experimental data has demonstrated a specific function. For example, *ZNF295* is a gene model with an open reading frame that contains zinc finger domains. Some zinc finger proteins have been shown to be transcription factors, so *ZBF295* is classed as such. In general, genes are classified as broadly as possible. For example *ITGB2*, is classed only as a cell adhesion molecule, although because it has been studied essentially only in lymphocytes, it is regarded as an immune system gene [17]. Future studies may well reveal functions other than those that have been observed, so it is as well to speculate about the functions of genes as broadly as possible.

Every biologist will bring their own expertise to bear in deciding which of the genes in Table 1 are of greatest potential relevance to the DS phenotype. Transcription factors are attractive candidates because imbalance of one component of a transcription factor complex may alter the effectiveness of the activation or repression of transcription of target genes. Genes within the ubiquitin pathway may alter rates of target protein degradation. Cell adhesion was long ago postulated, with intriguing preliminary data [18], to play a role in altering rates and extents of cell migration during development. Overexpression of one potassium channel gene has been shown to disregulate expression of other channel genes, affecting neuronal network excitability [19]. If mental retardation and cognitive deficits are the primary focus of study, almost any of the categories in Table 1 could be relevant, such is the extent of our current understanding of the complex developmental processes leading to these conditions.

Expression data
Only about half the 225 chromosome 21 genes have any functional association, and some of these are particularly weak - for example, the presence of a transmembrane domain is not very definitive. In some cases, the lack of protein or functional domain data may be due to the lack of complete coding sequence information. While awaiting the generation of complete cDNA sequences (which may be laborious to obtain), and even for further analysis of complete cDNAs lacking functional associations, expression patterns may help in prioritizing genes for further study. Of the novel genes with incomplete cDNAs, thirty-eight are represented by ESTs from Soares or CGAP cDNA libraries [20]. Of these, only seven would be classed as ubiquitous in expression - that is, present in dbEST with more than 30 entries from numerous tissues. Twenty-six ESTs are each associated with fewer than five dbEST entries. Five of these ESTs are seen only in testes/prostate and three are seen only in fetal sources. While there are features of dbEST construction that can produce artifactual pictures of expression patterns, these data suggest that the novel genes within 21q may be largely of limited expression. In some cases at least, this is consistent with the failure to identify these genes previously.

For relevance to mental retardation and cognitive deficits, genes with brain-specific expression, such as *PCP4* [21], are of interest. Equally interesting are examples of brain-specific alternative processing, as is seen with *Intersectin* and *DSCAM* [22,23]. In an analysis of a number of novel gene models, alternative processing, some of it brain-specific, was observed in the majority of cases [24]. It is unlikely that even most known genes have been examined thoroughly for instances of multiple transcripts. Because these may alter protein sequences and therefore function, their role in DS may be relevant.

Evolutionary conservation
Model organisms will provide the basis for functional studies of the known and novel chromosome 21 genes. The genomes of *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila* [25-27] have been completely sequenced, and thus the complete set of proteins of each of these organisms is known. Annotation of the *Drosophila* genome identified approximately 13,500 genes. Comparison of the translations of all annotated chromosome 21 genes with the *Drosophila* set identified 23 chromosome 21 gene products with similarity to a *Drosophila* protein over the complete length. Many of these similarities involve basic biochemical/biological functions and include such proteins as SOD1 (superoxide dismutase), GART (a purine biosynthesis enzyme), CBS (cystathionine beta-synthetase), and those involved in RNA splicing and the ubiquitin pathway. A further set of 31 genes showed excellent informative matches but only over a domain or subregion of the human protein. Previously known homologs include *MNB* (minibrain) and *SIM2* (single-minded). Perhaps most interesting in both sets are those genes for which there is little or no functional data. Table 2 lists some of the known and novel chromosome 21 genes with partial and complete similarities in *Drosophila*. Among the novel genes, identities at the amino-acid level range as high as 64% (*c21orf19*) and over as many as 1,600 residues (*c21orf5*). Additional details remain to be resolved; for example, in several cases the lengths of the human and *Drosophila* proteins are significantly different. Correcting

Table 1

Chromosome 21 functional gene categories

Functional categories	Number of genes	Functional assignments
Transcription factors, regulators, and modulators	17	GABPA, BACH1, RUNX1, SIM2, ERG, ETS2 (transcription factors); ZNF294, ZNF295, Pred65, *ZNF298, APECED (zinc fingers); KIAA0136 (leucine zipper); GCFC (GC-rich binding protein); SON (DNA binding domain); PKNOX1 (homeobox); HSF2BP (heat shock transcription factor binding protein); NRIP1 (modulator of transcriptional activation by estrogen)
Chromatin structure	4	H2BFS (histone 2B), HMG14 (high mobility group), CHAF1B (chromatin assembly factor), PCNT (pericentrin, an integral component of the pericentriolar matrix of the centrosome)
Proteases and protease inhibitors	6	BACE (beta-site APP cleaving enzyme); TMPRSS2, TMPRSS3 (transmembrane serine proteases); ADAMTS1, ADAMTS5 (metalloproteinases); CSTB (protease inhibitor)
Ubiquitin pathway	4	USP25, USP16 (ubiquitin proteases); UBE2G2 (ubiquitin conjugating enzyme); SMT3A (ubiquitin-like)
Interferons and immune response	9	IFNAR1, IFNAR2, IL10RB, IFNGR2 (receptors/auxilliary factors); MX1, MX2 (interferon-induced); CCT8 (T-complex subunit), TIAM1 (T-lymphoma invasion and metastasis inducing protein), TCP10L (T-complex protein 10 like)
Kinases	8	ENK (enterokinase); MAKV, MNB, KID2 (serine/threonine); PHK (pyridoxal kinase), PFKL (phosphofructokinase); *ANKRD3 (ankyrin-like with kinase domains); PRKCBP2 (protein kinase C binding protein)
Phosphatases	2	SYNJ1 (polyphosphinositide phosphatase); PDE9A (cyclicphosphodiesterase)
RNA processing	5	rA4 (SR protein), U2AF35 (splicing factor), RED1 (editase), PCBP3 (poly(C)-binding protein); *RBM11 (RNA-binding motif)
Adhesion molecules	4	NCAM2 (neural cell), DSCAM; ITGB2 (lymphocyte); c21orf43 (similar to endothelial tight junction molecule)
Channels	7	GRIK1 (glutamate receptor, calcium channel); KCNE1, KCNE2, KNCJ6, KCNJ15 (potassium); *CLIC1l (chloride); TRPC7 (calcium)
Receptors	5	CXADR (Coxsackie and adenovirus); Claudins 8, 14, 17 (Claustridia); Pred12 (mannose)
Transporters	2	SLC5A3 (Na-myoinositol); ABCG1 (ATP-binding cassette)
Energy metabolism	4	ATP50 (ATP synthase oligomycin-sensitivity conferral protein); ATP5A (ATPase-coupling factor 6); NDUFV3 (NADH-ubiquinone oxoreductase subunit precursor); CRYZL1 (quinone oxidoreductase)
Structural	4	CRYA (lens protein); COL18, COL6A1, COL6A2 (collagens)
Methyl transferases	3	DNMT3L (cytosine methyl transferase), HRMT1l1 (protein arginine methyl transferase); Pred28 (AF139682) (N6-DNA methyltransferase)
SH3 domain	3	ITSN, SH3BGR, UBASH3A
One carbon metabolism	4	GART (purine biosynthesis), CBS (cystathionine-β-synthetase), FTCD (formiminotransferase cyclodeaminase), SLC19A1 (reduced folate carrier)
Oxygen metabolism	3	SOD1 (superoxide dismutase); CBR1, CBR3 (carbonyl reductases)
Miscellaneous	28	HLCS (holocarboxylase synthase); LSS (lanosterol synthetase); B3GALT5 (galactosyl transferase); *AGPAT3 (acyltransferase); STCH (microsomal stress protein); ANA/BTG3 (cell cycle control); MCM3 (DNA replication associated factor); APP (Alzheimer's amyloid precursor); WDR4, WDR9 (WD repeat containing proteins); TFF1, 2, 3 (trefoil proteins); UMODL1 (uromodulin); *Pred5 (lipase); *Pred3 (keratinocyte growth factor); KIAA0653, *IgSF5 (Ig domain); TMEM1, *Pred44 (transmembrane domains); TRPD (tetratricopeptide repeat containing); S100b (Ca binding); PWP2 (periodic tryptophan protein); DSCR1 (proline rich); DSCR2 (leucine rich); WRB (tryptophan rich protein); Pred22 (tRNA synthetase); SCL37A1 (glycerol phosphate permease)

In the table, 122 genes are assigned. The majority have complete or presumed complete cDNA sequences. Functional assignments have been based either on literature reports of direct experiment or on inferences from similarities to other proteins. Genes where models are incomplete (*) contain domains that suggest a function. Functional categories were chosen to be broadly descriptive; each gene appears in only one category.

Table 2

Similarities between selected human and *Drosophila* gene products

| Gene	Size (amino acids)		% ID*	%Sim*	E value	Length of similarity (amino acids)
	Human	*Drosophila*				
SOD1	154	153	61	73	10^{-47}	152
GART	1,010	(1,747)	46	63	10^{-180}	995
CRYAA	173	187	38	56	10^{-25}	154
UBE2G2	165	167	78	88	10^{-70}	165
DSCR3	297	295	49	69	10^{-72}	281
KIAA0958	428	490	33	49	10^{-50}	375
c21orf4	158	113	37	62	10^{-15}	94
c21orf19	439	295	64	77	10^{-94}	291
SIM2	667	634	70	79	10^{-125}	353
MNB	763	722	44	61	10^{-69}	314
APP	770	816	24	37	10^{-38}	497
CHAF1B	559	747	47	64	10^{-98}	381
KIAA0179	740`	687	26	42	10^{-17}	272
KIAA0539	2,300	2,029	26	42	10^{-28}	409
DSCR1	197	292	43	67	10^{-33}	153
c21orf2	256	454	53	67	10^{-33}	146
c21orf5	2,298	2,599	29	47	10^{-115}	1,082
			+ 40	58	10^{-96}	533

* The number of amino acids over which the % identity (%ID) and the % similarity (%Sim) was calculated. The *E* value is the expectation value, an indication of the probability of finding this level of similarity by chance.

these differences, if it is necessary, may strengthen the similarity data. In addition, defining complete cDNAs may reveal new homologies not discernible with partial gene models. Determining the phenotypes of mutants in the *Drosophila* genes is likely to shed light on the function of the homologous human genes.

Mouse models

Regions of human chromosome 21 are conserved within segments of three mouse chromosomes. The centromere-proximal region of chromosome 21 through the *MX* genes is homologous with the telomeric region of mouse chromosome 16 (Figure 1). The next approximately 2 Mb segment of chromosome 21 is homologous with the centromere-proximal region of mouse chromosome 17, and the telomeric 2 Mb of chromosome 21 is homologous with an internal segment of mouse chromosome 10. On the basis of current data, the order of chromosome 21 homologues in the mouse chromosome 16 and 10 segments appears to be completely conserved, although the boundaries of these regions are still approximate [28,29]. For example, the most centromere-proximal gene on chromosome 21 verified to map to mouse chromosome 16 is *STCH*. There are seven genes proximal to this that should be mapped in mouse. Similarly, although it is known that *Mx* maps to mouse chromosome 16 and *Tff3*, *Cbs* and *Crya* map to mouse chromosome 17, there are 11 genes between and among these that are of unknown map location in mouse. Lastly, *PDXK* is the most proximal chromosome 21 gene mapped to mouse chromosome 10 [28]. Genes in this region are relatively small, however, and additional chromosome 21 genes may be located on mouse chromosome 10 between *Pdxk* and the adjacent region homologous with human chromosome 19. Defining the endpoints of these homologous regions is critical for evaluating gene-phenotype correlations within existing mouse models and for designing new ones.
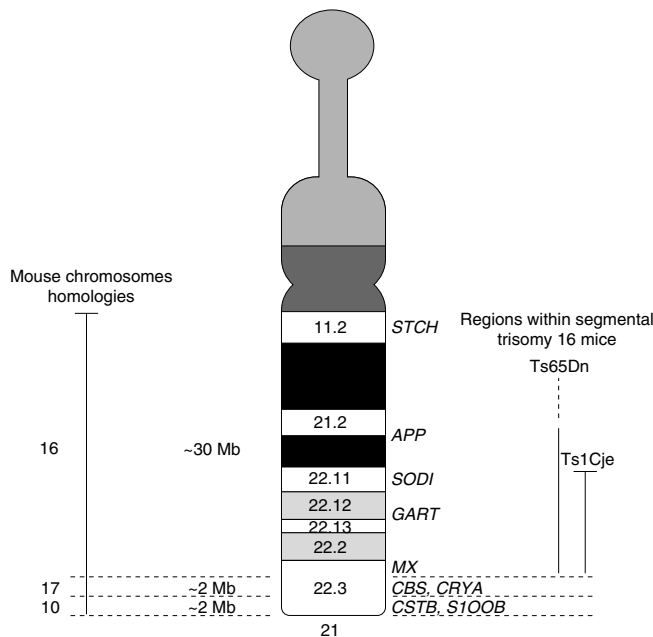
Figure I
The regions of human chromosome 21 that are syntenic with mouse chromosomes are indicated on the left; those that are trisomic in the major mouse models are indicated on the right.

Currently, the best mouse models of DS are the mouse chromosome 16 segmental trisomies, Ts65Dn and Ts1Cje. Ts65Dn is trisomic for the region spanning an undefined distance proximal to *App* through *Mx* to presumably the telomere of chromosome 16. The phenotype of Ts65Dn includes working memory impairment and long term memory deficits; delayed development and lower body weight; motor dysfunction; decreased responsiveness to pain; hyperactivity; and decreased ability to inhibit behavior (reviewed in [30,31]; see also [32,33]). Particularly interesting are observations of age-related loss of cholinergic neurons, decreased numbers of asymmetric synapses in the temporal cortex, abnormalities in neuron number in hippocampal regions, and deficiencies of beta-noradrenergic transmission within the hippocampus and cerebral cortex [34-38]. Some of these deficits have been observed in DS; others suggest new avenues of investigation. Knowing which genes cannot be responsible for the phenotype can be helpful. Table 3a lists the 32 genes found centromeric to the Alzheimer's-associated gene APP on chromosome 21. On the basis of current comparative mapping data, most of these may be present in only two copies in Ts65Dn and therefore would not contribute to its phenotypic features. The Ts1Cje mouse is a more recent model, and is trisomic for the region of mouse chromosome 16 from *Sod1* through *Mx* (and again presumably to the telomere). While it has not yet been studied so thoroughly as Ts65Dn, there are phenotypic differences between the two mice. In contrast to Ts65Dn, Ts1Cje shows

hypoactivity, no loss of cholinergic neurons, and no deficits in the visible platform part of the water maze tests (which tests only memory and not the ability to make spatial correlations) [39]. Table 3b lists 27 genes that are expected to be trisomic in the Ts65Dn but only disomic in the Ts1Cje, based on the genetic map [29]. It is tempting to conclude that these genes must account for the phenotypic differences, but it must be kept in mind that the two mouse strains have been produced on different genetic backgrounds, which may have phenotypic consequences.

Segmental trisomies for the regions of chromosome 21 homologous with mouse chromosomes 17 and 10 do not exist. If *Mx* is the most telomeric gene on mouse chromosome 16 and *Pdxk* is the most centromeric on mouse chromosome 10, there are 33 genes within the approximately 2.2 Mb of the mouse chromosome 17 region (Table 4) and 50 genes within the approximately 2.9 Mb of the mouse chromosome 10 region. Adding the maximum of 32 genes not trisomic in the Ts65Dn, half of the chromosome 21 homologous genes are not trisomic in Ts65Dn. The phenotypic consequences of these genes must be assessed in some fashion, because the Ts65Dn lacks some features of DS. Constructing single-gene transgenic mice expressing each of these and then combining each with the Ts65Dn by breeding would be laborious and probably of limited success. An alternative is to generate additional segmental trisomies using the Cre-*lox* system [40].

From genes to functions

Analysis of the complete sequence of chromosome 21 has provided the first look at all candidate DS genes. The next steps require verifying and refining the predicted, incomplete gene models, defining new models as necessary, and isolating complete cDNAs for each gene. With complete coding sequences, protein sequences can be examined for motifs, domains, and biochemical characteristics that may suggest function. The most challenging problem will then be determining the functions of these genes and the other 'known' genes. While it is tempting to focus on genes whose protein characteristics suggest a hypothesis for relevance to some aspect of DS, the more than 100 genes distributed throughout the chromosome that have no functional association are too large a dataset to ignore. For these and other genes on 21q, detailed expression analysis may be informative. Demonstration that a gene shows increased expression in the trisomic state by northern blot or RT-PCR analysis, followed by RNA tissue *in situ* hybridization to define specific cell types, brain regions and developmental stages of expression, may help in selecting genes of greater or lesser interest.

The most direct assessment of function will require mutation or overexpression of individual genes or sets of genes. For these experiments, the 'complete' protein databases for *S. cerevisiae*, *C. elegans* and *Drosophila* will provide homologous genes

Table 3

Human chromosome 21 centromere-proximal genes

(a) Genes proximal to APP		(b) Genes from APP to SOD1	
Gene	Classification	Gene	Classification
Pred 65	Zn finger	APP	
Pred 3	Keratinocyte growth factor	Pred24	
Pred 4	Similar to KIAA1074	ADAMTS1	Metalloproteinase
orf15	EST	ADAMTS5	Metalloproteinase
Pred5	Lipase	Pred25	Exon
RBM11	RNA binding	Pred26	Exon
Pred6	Exon	orf23	EST
STCH	Stress protein	Pred27	Exon
SAMSN-1	Similar to KIAA0790	Pred28	Methyltransferase
NRIP1	Nuclear factor	ZNF294	Zinc finger
USP28	Ubiquitin protease	orf6	EST
orf34	EST	USP16	Ubiquitin protease
orf35	EST	CCT8	T-complex subunit
orf36	EST	orf7	Exon
orf37	EST	Bach1	Transcription factor
CXADR	Viral receptor	orf12	EST
BTG3	Cell cycle control	orf8	EST
YG81	CDNA	GRIK1	Glutamate receptor
orf39	EST	Orf41	EST
Pred12	Mannose receptor	orf9	EST
PRSS7	Enterokinase	CLDN17	Claudin receptor
orf40	EST	CLDN8	Claudin receptor
NCAM2	Neural adhesion	Pred29	Exon
Pred15	Exon	Pred30	Exon
Pred16	EST	TIAM1	Lymphoma metastasis
orf53	EST	Pred 31	Exon
orf42	EST	SOD1	
Pred21	EST		
Red22	tRNA synthetase		
orf43	Junction adhesion		
ATP5A	ATPase factor		
GABPA	Transcription factor		

(a) Mouse homologs possibly disomic in both Ts65Dn and Ts1Cje. **(b)** Mouse homologs expected to be trisomic in Ts65Dn but disomic in Ts1Cje.

that can be analyzed in more tractable systems. The increasing complexity of the zebrafish EST database will add another model organism system of increasing utility. Issues remain with all model organisms, however, of verifying correct gene structures, identifying orthologous genes versus merely homologous genes, and interpreting mutation and knockout data in one system versus overexpression in another. The ultimate model organism, of course, will remain the mouse. Multiple genes can be 'added' to the Ts65Dn using transgenics carrying bacterial chromosomes

Table 4

Human genes with homologues mapping within the 2.2 Mb maximum mouse chromosome 17 homologous region

Gene	Functional class	Chromosome 21 ORFs	Exon model
		orf 20	
		orf 21	
		orf 22	
ANKRD3	Ankyrin kinase		
ZNF298	Zinc finger	orf 25	
ZNF295	Zinc finger		
UMODL1	Uromodulin		Pred46
ABCG1*	ATP-binding cassette transporter		
TFF3 *	Intestinal trefoil		
TFF2	Spasmolytic peptide		
TFF1	Estrogen-induced		
TMPRSS3	Membrane serine protease		
UBASH3A	SH3 domain		
TSGA2*	Testis-specific		
SLC37A1	Glycerol 3-phosphate permease		
PDE9A	Cyclic phosphodiesterase		
WDR4	WD repeats		
NDUFV3	NADH-ubiquinone oxoreductase subunit		
PKNOX1*	Homeobox		
CBS*	Cystathionine β synthetase		
U2AF1	Splicing factor		
CYRA *	Alpha-crystallin		
HSF2BP	Heat shock transcription factor binding		
			Pred47
			Pred48
SNF1LK*	KID2 kinase		
			Pred49
			Pred50
			Pred51
H2BFS	Histone		
KIAAD179			

Genes are listed in order from centromere to telomere on chromosome 21. * Genes verified as mapping to mouse chromosome 17.

(BACs), to look for enhanced DS-relevant phenotypes. The human sequence will be useful here in ensuring that clones are extensive enough to contain appropriate regulatory regions. Single-gene knockouts can also be 'subtracted' from the Ts65Dn mouse model, to search for amelioration of phenotype. With good biological intuition and luck, it may not be necessary to understand all of the genes within chromosome 21 before promising candidates are identified and the design of potential therapeutics can begin.

References
1. Hassold TJ, Jacobs PA: **Trisomy in man**. *Annu Rev Genet* 1984, **18**:69-97.
2. Epstein CJ: **Down syndrome (trisomy 21)**. In *Metabolic and Molecular Bases of Inherited Disease*. Edited by Scriver CA *et al.* New York: McGraw Hill; 1995:749-794.
3. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park H-S, Toyoda A, Ishii K, Totoki Y, Choi DK, *et al.*: **The DNA sequence of human chromosome 21.** *Nature* 2000, **405**:311-319.
4. Pennisi E: **And the gene number is ...?** *Science* 2000, **288**:1146-1147.
5. Dunham I, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, Ainscough R, Almeida JP, Babbage A, *et al.*: **The DNA sequence of human chromosome 22.** *Nature* 1999, **402**:489-495.
6. Gardiner K: **Clonability and gene distribution on human chromosome 21: reflections of junk DNA content?** *Gene* 1997, **205**:39-46.
7. Xu h, Wei H, Tassone F, Graw S, Gardiner K, Weissman SM: **A search for genes from the dark band regions of human chromosome 21.** *Genomics* 1995, **27**:1-8.
8. Francke U: **Digitized and differentially shaded human chromosome ideograms for genomic applications.** *Cytogenet Cell Genet* 1994, **65**:206-219.
9. Saccone S, Caccio S, Kusuda J, Andreozzi L, Bernardi G: **Identification of the gene-richest bands in human chromosomes.** *Gene* 1996, **174**:85-94.
10. **Swiss-PROT Index of protein sequence entries encoded on human chromosome 21** [http://expasy.cbr.nrc.ca/cgi-bin/lists?humchr21.txt].
11. **DbEST** [http://www.ncbi.nlm.nih.gov/dbEST/index.html].
12. Korenberg JR, Kalousek DK, Anneren G, Pulst S-M, Hall JG, Epstein CJ, Cox DR: **Deletion of chromosome 21 and normal intelligence: molecular definition of the lesion.** *Hum Genet* 1991, **87**:112-118.
13. **Eleanor Roosevelt Institute** [http://www-eri.uchsc.edu/].
14. Rahmani Z, Blouin JL, Creau-Goldberg N, Watkins PC, Mattei JR, Poissonnier M, Prieur M, Chettouh Z, Nicole A, Aurias A, *et al.*: **Critical role of the D21S55 region on chromosome 21 in the pathogenesis of Down syndrome.** *Proc Natl Acad Sci USA* 1989, **86**:5958-5962.
15. Delabar JM, Theophile D, Rahmani Z, Chettouh Z, Blouin JL, Prieur M, Noel B, Sinet PM: **Molecular mapping of twenty-four features of Down syndrome on chromosome 21.** *Eur J Hum Genet* 1993, **1**:114-124.
16. Korenberg JR, Chen X-N, Schipper R, Sun Z, Gonsky R, Gerwehr S, Carpenter N, Daumer C, Dignan P, Disteche C *et al.*: **Down syndrome phenotypes: the consequences of chromosomal imbalance.** *Proc Natl Acad Sci USA* 1994, **91**:4997-5001.
17. Marlin SD, Morton CC, Anderson DC, Springer TA: **LFA-1 immunodeficiency disease: definition of the genetic defect and chromosomal mapping of alpha and beta subunits of the lymphocyte function-associated antigen 1 (LFA-1) by complementation in hybrid cells.** *J Exp Med* 1986, **164**:855-867.
18. Kurnit DM, Aldridge JF, Matsuoka R, Matthysse S: **Increased adhesiveness of trisomy 21 cells and atrioventricular canal malformations in Down syndrome: a stochastic model.** *Am J Med Genet* 1985, **20**:385-399.
19. Sutherland ML, Williams SH, Abedi R, Overbeek PA, Pfaffinger PJ, Noebels JL: **Overexpression of a Shaker-type potassium channel in mammalian central nervous system dysregulates native potassium channel gene expression.** *Proc Natl Acad Sci USA* 1999, **96**:2451-2455.

20. Bonaldo MF, Lennon G, Soares MB: **Normalization and subtraction: two approaches to facilitate gene discovery.** *Genome Res* 1996, **6:**791-806.

21. Ziai MR, Sangameswaran L, Hempstead JL, Danho W, Morgan JI: **An immunochemical analysis of the distribution of a brain-specific polypeptide, PEP-19.** *J Neurochem* 1988, **51:**1771-1776.

22. Guipponi M, Scott HS, Chen H, Schebesta A, Rossier C, Antonarakis SE: **Two isoforms of a human intersectin (ITSN) protein are produced by brain-specific alternative splicing in a stop codon.** *Genomics* 1998, **53:**369-376.

23. Yamakawa K, Huot YK, Haendelt MA, Hubert R, Chen XN, Lyons GE, Korenberg JR: **DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system.** *Hum Mol Genet* 1998, **7:**227-237.

24. Slavov D, Hattori M, Sakaki Y, Rosenthal A, Shimizu N, Minoshima S, Kudoh J, Yaspo M, Ramser J, Reinhardt *et al.*: **Criteria for gene identification and features of genome organization: analysis of 6.5 Mb of DNA sequence from human chromosome 21.** *Gene* 2000, **247:**215-232.

25. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M *et al.*: **Life with 6000 Genes.** *Science* 1996, **274:**546-567.

26. The *C. elegans* Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012-2018.

27. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PC, Scherer SE, Li PW, Hoskins RA, Galle RF *et al.*: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287:**2185-2195.

28. Wiltshire T, Pletcher M, Cole SE, Villaneuva M, Birren B, Lehoczky J, Dewar K, Reeves RH: **Perfect conserved linkage across the entire mouse chromosome 10 region homologous to human chromosome 21.** *Genome Res* 1999, **9:**1214-1222

29. **Mouse Genome Database** [http://www.informatics.jax.org]

30. Davisson MT, Costa ACS: **Mouse models of Down syndrome.** In *Mouse Models in the Study of Genetic Neurological Disorders*, Volume 9 of *Advances in Neurochemistry*. Edited by Popko B. New York: Plenum Press; 1999:297-327.

31. Crnic LS, Pennington BF: **Down syndrome: neuropsychology and animal models.** *Progr Infancy Res* 2000, **1:**69-111.

32. Costa ACS, Walsh K, Davisson MT: **Motor dysfunction in a mouse model for Down syndrome.** *Physiol Behav* 1999, **68:**211-220.

33. Martínez-Cué C, Baamonde C, Lumbreras MA, Villina IF, Dierssen M, Florez J: **A murine model for Down syndrome shows reduced responsiveness to pain.** *NeuroReport* 1999, **10:**1119-1122.

34. Holtzman DM, Santucci D, Kilbridge J, Chua-Couzens J, Fontana DJ, Daniels SE, Johnson RM, Chen K, Sun Y, Carlson E *et al.*: **Developmental abnormalities and age-related neurodegeneration in a mouse model of Down syndrome.** *Proc Natl Acad Sci USA* 1996, **93:**13333-13338.

35. Granholm AC, Sanders LA, Crnic LS: **Loss of cholinergic phenotype in basal forebrain coincides with cognitive decline in a mouse model of Down's syndrome.** *Exp Neurol* 2000, **161:**647-663.

36. Kurt MA, Davies DC, Kidd M, Kierssen M, Flórez J: **Synaptic deficit in the temporal cortex of partial trisomy 16 (Ts65Dn) mice.** *Brain Res* 2000, **858:**191-197.

37. Insausti AM, Megias M, Crespo D, Cruz-Orive LM, Dierssen M, Vallina TF, Insausti R, Florez J, Vallina TF: **Hippocampal volume and neuronal number in Ts65Dn mice: a murine model of Down syndrome.** *Neurosci Lett* 1998, **253:**175-178.

38. Dierssen M, Vallina IF, Baamonde C, García-Calatayud S, Lumbreras MA, Flórez J: **Alterations of central noradregenergic transmission in Ts65Dn mouse, a model for Down syndrome.** *Brain Res* 1997, **749:**238-244.

39. Sago H, Carlson EJ, Smith DJ, Kilbridge J, Rubin EM, Mobley WC, Epstein CJ, Huang TT: **Ts1Cje, a partial trisomy 16 mouse model for Down syndrome, exhibits leaning and behavioral abnormalities.** *Proc Natl Acad Sci USA* 1998, **95:**6256-6261.

40. Zheng B, Sage M, Sheppeard EA, Bradley A: **Engineering mouse chromosomes with Cre-LoxP: range, efficiency, and somatic applications.** *Mol Cell Biol* 2000, **20:**648-655.