Int. J. Dev. Biol. 46: 89-96 (2002)

Chromosome 21: a small land of fascinating disorders with unknown pathophysiology

STYLIANOS E. ANTONARAKIS*, ROBERT LYLE, SAMUEL DEUTSCH and ALEXANDRE REYMOND

Division of Medical Genetics, University of Geneva Medical School and University Hospitals, Geneva, Switzerland

ABSTRACT In the year 2000 we celebrated the sequencing of the entire long arm of human chromosome 21. This achievement now provides unprecedented opportunities to understand the molecular pathophysiology of trisomy 21, elucidate the mechanisms of all monogenic disorders of chromosome 21, and discover genes and functional sequence variations that predispose to common complex disorders. All of that requires the functional analysis of gene products in model organisms, and the determination of the sequence variation of this chromosome.

KEY WORDS: *Down Syndrome, HC21, molecular pathology, trisomy 21, single nucleotide polymorphism*

The Sequence of HC21 and the Gene Catalogue

On May 18, 2000, the (almost) entire sequence of 21q was published and became freely available (Hattori *et al.*, 2000). This landmark publication concluded research efforts of many investigators and laboratories that studied the infrastructure of HC21 over the last 15 years. Other historical landmarks of research on chromosome 21 include the first description of trisomy 21 in 1959 (Lejeune *et al.*, 1959), first cloning of a chromosome 21 gene in 1982 (Lieman-Hurwitz *et al.*, 1982), first disease-related gene mutation identified in 1989 (Kishimoto *et al.*, 1989), first linkage map in 1989 (Warren *et al.*, 1989), completion of the physical map in 1992 (Chumakov *et al.*, 1992).

The determination of sequence of 21q was only second to that of chromosome 22q (Dunham *et al.*, 1999). The completion of the sequence of chromosome 21 was an international event in which the leading contributions were made by laboratories in Japan and Germany. It is anticipated that the availability of the sequence will now accelerate and facilitate discoveries concerning the disease mechanisms of trisomy 21, the monogenic disorders and the susceptibility genes for common polygenic traits that map on this chromosome (Fig. 1).

The determination of the sequence was achieved by sequencing BAC and other clones by the shotgun method or by serial deletions using transposon insertions (Hattori *et al.*, 2000). The sequence is of high quality (error rate less than 1 in 10000 nucleotides) with only 3 cloning gaps on 21q, each of which is no more than 30-40 kb. One of these gaps was partially filled by the sequence generated by Celera Genomics (Venter *et al.*, 2001). There are therefore 4 contigs for 21q (from 21cen to 21qter) of lengths 28515, 219, 1378, and 3429 kb, for a total of 33,546,361 nucleotides. There are also 7 sequencing gaps in regions with long stretches of repeats in which it was impossible to know if the entire repeat was sequenced. A region of 281 kb from the short arm 21p has also been sequenced and contains just one gene (TPTE).

Only approximately 3% of the sequence encodes for proteins; in addition 1.3% encodes for short sequence repeats that may be polymorphic in human populations. About 10.8% of the sequence are SINE and 15.5% LINE repeats respectively; an additional 11.7% are other human repeats (total interspersed repeats 38% of the sequence). The remainder are gene flanking regions, introns and other DNA of unknown function.

Initial comparison of the genomic sequence to full- or partial-length cDNAs, and ESTs (Expressed Sequence Tag) from all species, together with the use of exon prediction programs revealed 127 known genes, 98 predicted genes (13 similar to known proteins, 17 anonymous ORFs featuring modular domains, and 68 anonymous transcription units with no similarity to known proteins) and 59 pseudogenes. One gene cluster the KAP (keratin associated proteins; counted as one among the 127 genes) consists of 16 genes and 2 pseudogenes ((Hattori *et al.*, 2000) and N Shimizu personal communication). The genes have been recognized by several criteria. Category 1 genes were those with identity or extensive homology to known genes; category 2 were those with similarity to cDNAs or ORFs (open reading frames) from other organisms; category 3 those with predicted amino acid similarity to a recognized protein domain or region; category 4 those identical to spliced ESTs or only predicted by exon recognition computer programs.

Abbreviations used in this paper: DS, Down Syndrome; HC21, human chromosome 21; cSNP, cDNA Single Nucleotide Polymorphism; SSRP, Short Sequence Repeat Polymorphism;

***Address correspondence to:** Dr. Stylianos E. Antonarakis. 1 rue Michel-Servet, 1211 Geneva, Switzerland. Fax: +41-22-702-5706.
e-mail: Stylianos.Antonarakis@medecine.unige.ch

0214-6282/2002/$25.00

Printed in Spain
www.ijdb.ehu.es

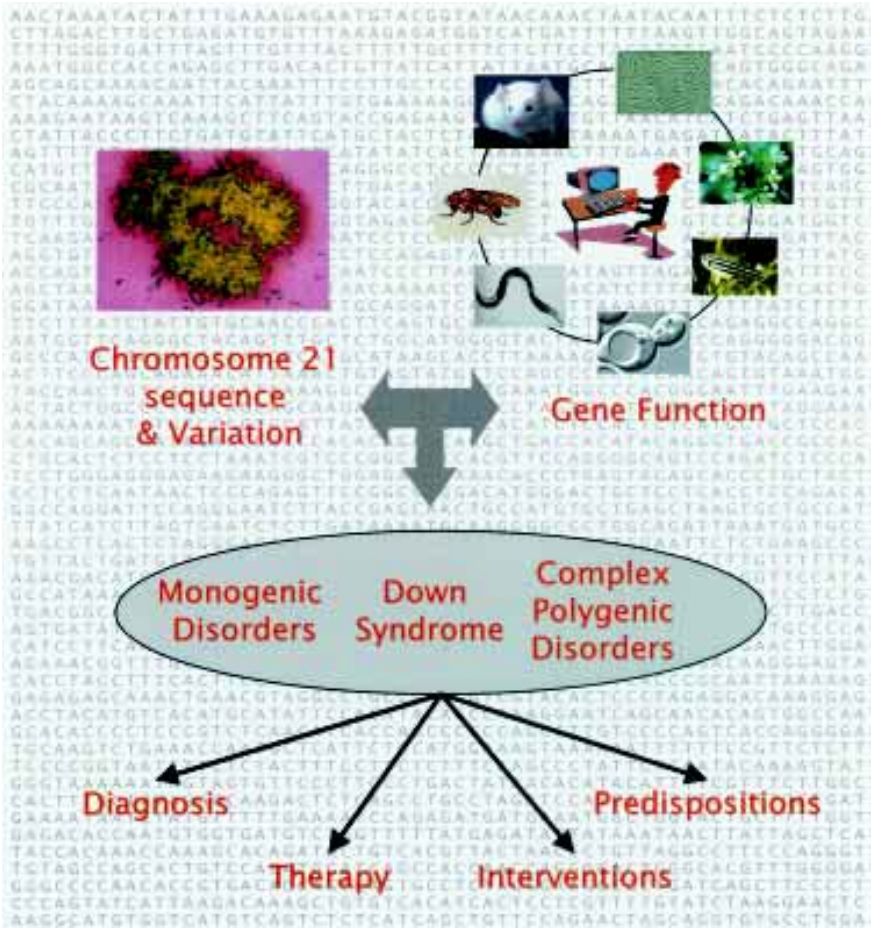


Fig. 1. The completion of the chromosome 21 sequence, the discovery of the normal variation of these sequences and the elucidation of the function of the proteins encoded by this chromosome will help in the following: *(i) the cloning and understanding of the pathophysiology of the monogenic disorders that map to HC21; (ii) the contribution of different HC21 genes to the phenotypes of Down syndrome; and (iii) the detection of functional variants associated with complex common polygenic disorders. All of these discoveries will then provide the basis for new diagnostic and therapeutic possibilities.*

The gene density is not uniform throughout the length of 21q. There exists a striking difference since the proximal half of 21q contains only 58, while the distal half contains 167 genes. Remarkably, in the proximal half there is a region of about 7 Mb containing only 7 genes. An intensive international effort is now directed towards the revision and update of the gene catalogue with the experimental confirmation and the discovery of previously unrecognized genes. This is possible because of the accumulation of additional ESTs, the sequence of the genome of other organisms, and wet lab experiments (e.g. RT-PCRs, 5' and 3' RACE, Northern blot analysis). For example, we showed that most of the C21orfs (chromosome 21 open reading frame) originally defined by matching spliced ESTs were correctly predicted, while many of the PREDs (gene Predictions), defined solely by computer prediction, do not correspond to genuine genes (Reymond *et al.,* 2001). Our study underscores the limit of *in silico* only gene prediction. The gene annotation of the chromosome is constantly updated and revised. At our most recent count we could recognize 151 known genes, and 79 predicted genes for a total of 230 genes. This number is likely to increase as more ESTs and sequences from model organisms become available. Gardiner and Davisson recently published an analysis of the gene categories according to their function (Gardiner & Davisson, 2000); for example there are 17 genes that encode transcription regulators, 8 kinases, 4 adhesion molecules etc. Among the HC21 genes, approximately 35% are homologous to genes in *Drosophila*, 35% to *C.Elegans*, and 18% to yeast *Saccharomyces cerevisiae*. Furthermore, the majority of human genes have homologous genes in the mouse. The completion of the sequence of the mouse genome would provide the opportunity to globally determine how many of the human genes are present in the mouse and vice versa.

HC21 is a chromosome with average gene density. The sequence of the entire human genome by the public (Lander *et al.*, 2001) and private projects (Venter *et al.*, 2001) revealed that the average gene density of the human genome is approximately 9 genes per Mb. The initial characterization of HC21 by the Celera reanalysis resulted in a content of approximately 8.7 genes per Mb (Venter *et al.*, 2001). This is in contrast to the same size but gene-rich HC22 that contained 17 genes per Mb. The gene catalogue of the (almost) fully sequenced chr22 contained 764 genes (Dunham *et al.*, 1999; de Souza *et al.*, 2000) (ratio HC22/HC21 of 3.3). Since both of these chromosomes contain about 2.88% of the human transcripts (Deloukas *et al.*, 1998), and their sequences are finished and of high quality, it follows that the minimum estimate for the number of human genes should be around 35,500.

The HC21 gene catalogue needs to be soon completed with the description of the full-length sequences of all genes, their alternative splicing variants, and their temporal and tissue distribution of expression. After this initial phase of characterization, a serious and long-term effort needs to be directed towards the elucidation of the function of each gene product.

The sequence of HC21 also provided knowledge on all the long repeats that may predispose to unequal crossing over and deletion/duplication syndromes (Ji *et al.*, 2000; Eichler *et al.,* 2000). In addition, comparison of the genotyping data of the members of the CEPH families over the last 15 years, with the nucleotide sequence, revealed the patterns of meiotic recombination on the different parts of the chromosome (Lynn *et al.*, 2000; Hattori *et al.*, 2000). The recombination rate varies considerably between genders across 21q; the greatest difference (eightfold) is in the pericentromeric region, with a rate of approximately 250 kb/cM in females and approximately 2125 kb/cM in males.

cSNPs and Other Sequence Variants

The sequence of a chromosome is certainly of great importance; of equal importance, however, is the documentation of the variability of the genomic sequence in different human populations. The most abundant class of sequence variants are the single nucleotide

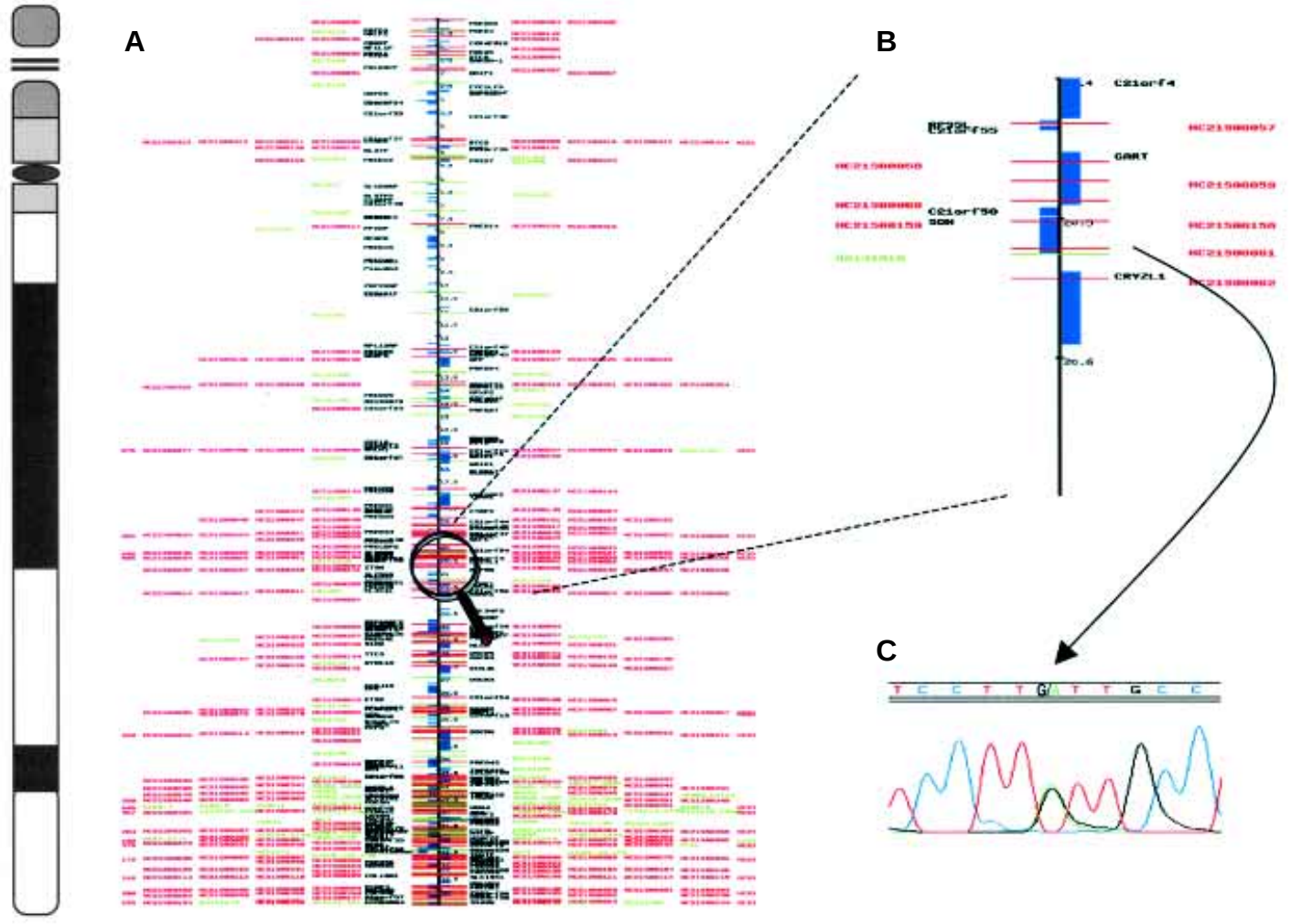


Fig. 2. Overview of the location of cSNPs and SSR (short sequence repeat polymorphisms) on human chromosome 21. *The ideogram of the chromosome is shown on the left.* **(A)** *A schematic map of the 33.5 Mb of the chromosome from the centromere (top) to the 21qter (bottom). Blue boxes represent the transcripts, and black letters their names; green lines and letters are the SSRs; red lines and letters are the cSNPs.* **(B)** *A magnification of 350 kb of A.* **(C)** *DNA sequence analysis of a cSNP shown in B. The polymorphic nucleotides are either G or A. The cSNP database is accessible at* http://cSNP.unige.ch/.

polymorphisms which occur in approximately every 500-1000 nucleotides between two randomly chosen chromosomes (Halushka *et al.*, 1999). The availability of the genomic sequence allows the identification of potential SNPs by comparing these sequences to those of the ESTs and other available sequences. Most of the SNPs identified in this manner are cSNP i.e. variability in cDNAs. These are of potentially significant value since they may be associated with quantitative or functional differences of the encoded proteins. A recent collection of such cSNPs has been published (Deutsch *et al.*, 2001) (Fig. 2). Other extensive collections of SNPs are also available in public databases (Altshuler *et al.*, 2000; Sachidanandam *et al.*, 2001)(http://snp.cshl.org/). The functional significance of SNPs needs to be experimentally determined as this "normal" variability may not be completely neutral. There are examples of SNPs in regulatory regions of genes that are associated with altered expression (Schork *et al.*, 2000). Furthermore, it has been recently found that exonic sequence variants with or without codon changes may alter splicing enhancers and are therefore associated with different degrees of exon skipping (Liu *et al.*, 2001). These "functional" variants may therefore be the basis of susceptibility to complex disorders.

In addition, a large number of potential short-sequence repeats (microsatellite) polymorphisms have been identified and their positions have been precisely mapped.

Together, all of these polymorphic markers could be used to map disease and other phenotypic trait loci, estimate the extend of linkage disequilibrium in different human populations (Reich *et al.*, 2001b), study the allelic contribution of genes to the phenotypes of Down syndrome (DS), and discover genes that confer susceptibility to complex disorders.

Down Syndrome and the HC21 Sequence

Patients with trisomy 21 and their families will almost certainly benefit greatly from the completion of the sequence of HC21. We know very little about the genes involved in the different phenotypes of DS and the molecular pathophysiology of these phenotypes. The

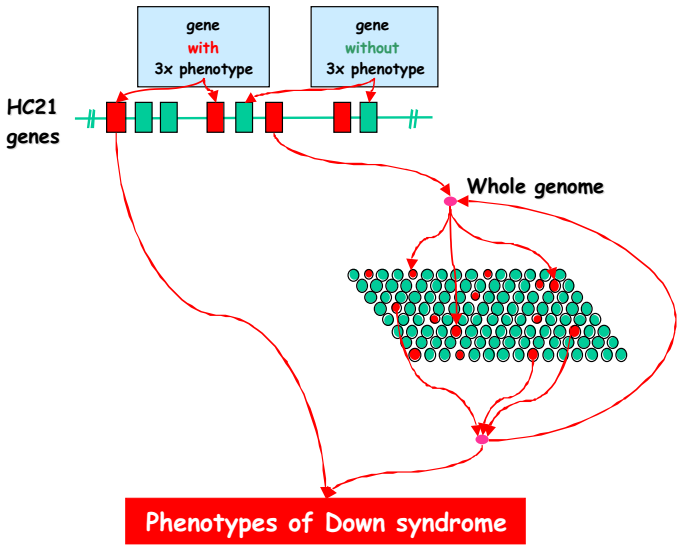


Fig. 3. Models for the pathogenesis of Down syndrome. *The genes of HC21 are depicted as green and red boxes. Some of the "red-type" genes contribute directly to the pathogenesis of DS, while other "red-type" genes contribute via a more general dysregulation of gene expression. "Green-type" genes do not contribute to any of the phenotypes of DS.*

current working hypotheses for the phenotypes of DS are (see also Fig. 3): (i) Some genes (but not all) contribute to the different phenotypes of DS; (ii) single genes may be involved in specific DS phenotypes; (iii) the effect of the expression of the 3 copies of these genes to the DS phenotypes could be either direct or indirect through the global dysregulation of gene expression; (iv) some phenotypes may be related to trisomy for specific alleles of HC21 genes or to the interaction of HC21 products with specific alleles of non-HC21 genes; (v) some DS phenotypes may be related to disturbances of gene regulation due to the presence of the extra chromosomal material and not specifically to particular genes. The notion of a minimal critical region (Delabar *et al.*, 1993), that harbors the majority of genes involved in the DS phenotypes, needs to be re-evaluated by the study of additional patients with partial trisomy 21, systematic analysis of the phenotypes, and diagnosis of the triplicated genomic regions using current methodology.

One important research goal is now to determine which of the HC21 genes are involved in the phenotypes of DS and which are not. This requires on one hand the discovery of the function of the HC21 encoded proteins and on the other the study of the effects of 1.5X overexpression of these genes. Both could be studied in knock-out (loss of function) or overexpression experiments of the homologous genes in model organisms.

In addition, the identification of all genes of HC21 provide the opportunity to study the level of their overexpression (or underexpression ?) in different cells, tissues and different stages of development. Furthermore, by using either microarray technologies (Brown & Botstein, 1999) or SAGE (serial analysis of gene expression)(Velculescu *et al.*, 1995) the global dysregulation of gene expression in trisomy 21 (or equivalent in model organisms) could be assessed. One such preliminary example of differences in gene expression by SAGE in partial trisomy 16 mouse brains versus normal littermates has been published (Chrast *et al.*, 2000). Finally, the effects of the overexpression of one gene using transfected cell lines or transgenic animals could be evaluated at the transcription level. All of the above studies could be also performed and validated at the protein level by qualitative and quantitative analysis of proteins (Yates, 2000; Pandey & Mann, 2000; MacBeath & Schreiber, 2000). Together these studies will provide a better understanding of the pathophysiology of DS and may reveal target molecules for therapeutic interventions of certain phenotypes.

Another benefit of the genomic sequence of HC21 will be in the accurate diagnosis and extent of HC21 partial deletions or duplications. Hybridizations of genomic DNA against microarrays of all the BACs and other clones used for the sequence could accurately reveal cryptic (partial) genomic aneuploidies (Antonarakis, 2001). The methodology has been recently worked out for different genomic regions and the current results are promising (Pinkel *et al.*, 1998; The BAC Research Consortium, 2001).

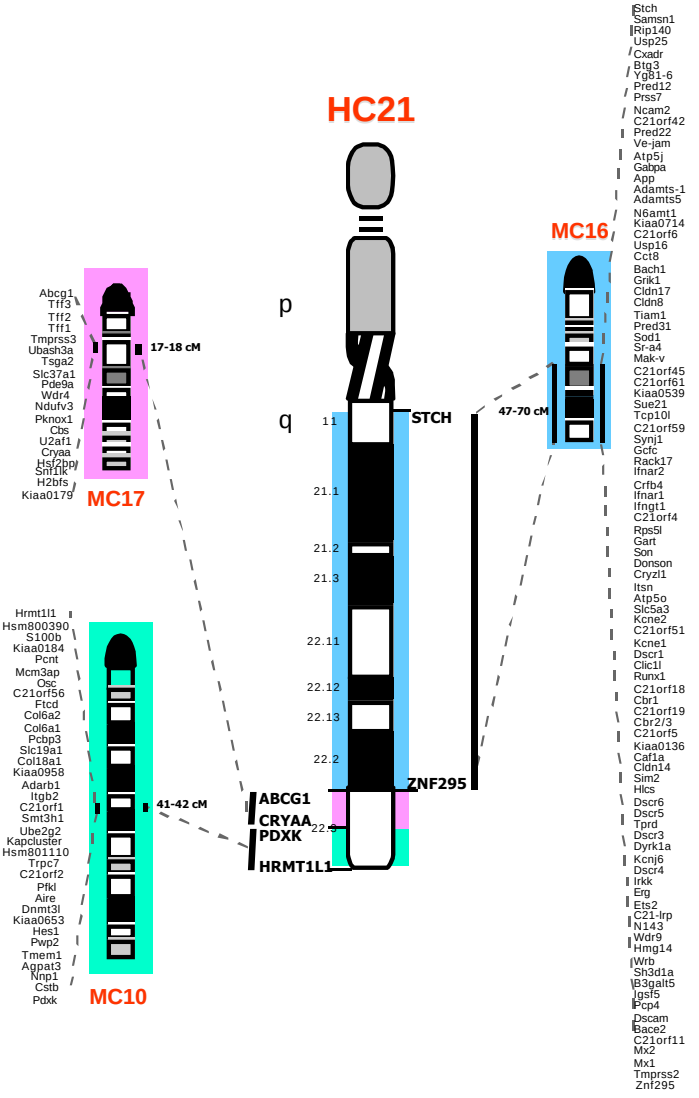
Mouse Models of Down Syndrome

The mouse is an adequate, albeit not perfect model to study the molecular pathophysiology of DS phenotypes. There are 3 chromosomal regions in the mouse that contain all the genes of human 21q (for review see (Dierssen *et al.*, 2001). From 21cen to 21qter, about 30 Mb are homologous to mouse chromosome 16, 2 Mb to mouse chromosome 17, and 2 Mb to mouse chromosome 10 (Fig. 4). A partial trisomy 16 mouse model Ts65Dn extending approximately from proximal to APP until ZNF295 (predicted to contain at least 86 HC21 homologous genes) has been generated. It shows several abnormal phenotypes (Reeves *et al.*, 1995; Baxter *et al.*, 2000; Costa *et al.*, 1999; Escorihuela *et al.*, 1998); additional partial

TABLE 1

MOUSE MODELS OF DS

Model	Genotype	Neurological phenotype
Segmental trisomy		
Ts16	Trisomy 16	Reduced brain size and some structural changes
Ts65Dn	Trisomic for *App-Znf295* (~86 known genes)	Learning and behavioural deficits, degeneration of basal forebrain cholinergic neurons Reduction of the cerebellar volume and granule cell density Reduced cell number and volume in the hippocampal dentate gyrus Reduction in excitatory (asymmetric) synapses in the temporal cortex at advanced ages Age-related degeneration of basal forebrain cholinergic neurons Astrocytic hypertrophy and increased astrocyte numbers
Ts1Cje	Trisomic for *Znf295-Sod1* (72 known genes)	Learning and behavioural deficits (less severe than in Ts65Dn)
Ms1Ts65	Trisomic for *App-Sod1* (~14 known genes)	Learning deficits (less severe than in Ts1Cje)
Single Genes		
TgSod1	Transgenic for human *SOD1*	Learning defects
TgPfkl	cDNA, highly overexpressed	-
TgS100ß	2-12 copies	Astrocytosis, neurite degeneration
TgApp	YAC, low copy	Cognitive/behavioural defects
TgEts2	cDNA, highly overexpressed	-
TgHmg14	2-6 copies	-
TgMnb	YAC, 1-3 copies, and cDNA, highly overexpressed	Learning/memory defects
TgSim2	BAC, 1-2 copies, and cDNA, highly overexpressed	Behavioural defects



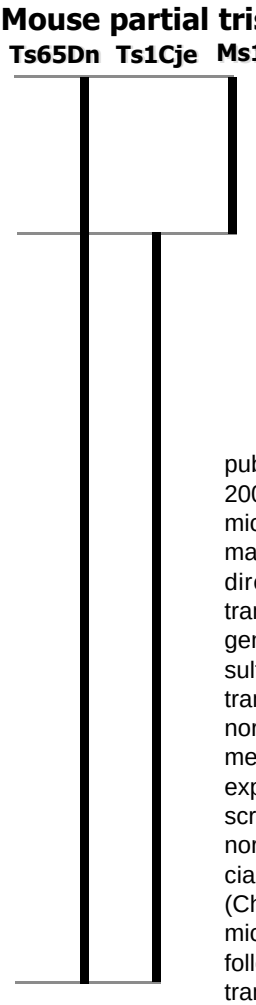


Fig. 4. Schematic representation of human chromosome 21 and the homologous chromosomal regions on mouse chromosomes 16, 17, and 10. *Gene order is based on the human sequences. The map order is not known for all mouse genes. The lines on the right depict the portions of mouse chromosome 16 triplicated in the partial trisomy 16 mouse models.*

trisomy 16s extending from SOD1 to ZNF295 (72 known genes) (Ts1Cje, (Sago *et al.*, 1998)) and from APP to SOD1 (14 known genes) (Ms1Ts65, (Sago *et al.*, 2000)) have been reported and their phenotypic characterization is in progress (table 1). Another approach has been to create a chimeric mouse model bearing substantial numbers of cells carrying portions of the human chromosome 21 (Shinohara *et al.*, 2001). The sequence of HC21 and of the mouse genome (www.celera.com) now permits the cloning of appropriate homologous regions of the mouse genome which in turn will facilitate the creation of additional partial trisomies for mouse chromosomes 17 and 10 by Cre-loxP-mediated somatic or meiotic recombination (Herault *et al.*, 1998; Zheng *et al.*, 2000). The breeding of these different mouse strains will eventually result in the creation and study of an "ultimate" mouse model with trisomy for the full mouse genome that is homologous to HC21. The contribution of the trisomy of individual genes in the mouse phenotype could further be evaluated by the sequential deletion of one copy of the gene from the partial trisomy mouse; this could be achieved by crossing knockout mice for a particular gene to the partial trisomy mouse. One such example has been recently published for the IFNAR gene (Maroun *et al.*, 2000). This is feasible since many knockout mice for HC21-homologous are available and many others could be created. Another more direct method is of course by creating transgenic mice with overexpression of one gene homologous to HC21. To achieve results of biological significance in DS, the transgene needs to be regulated from its normal promoter and other regulatory elements, and only one extra copy needs to be expressed. One such example has been described for the Sim2 gene, which shows abnormalities involving spatial exploration, social interactions and reduced nociception (Chrast *et al.*, 2000). Several other transgenic mice have been developed with HC21 genes following the first such mouse with Sod1 transgenes (Epstein *et al.*, 1987). Considerable knowledge could also gained by multicopy overexpression of the transgene (Sumarsono *et al.*, 1996). The breeding of these animal models to different mouse genetic backgrounds may uncover the contribution of the entire genome in the modification of certain DS phenotypes.

Monogenic Disorders on HC21; Great Progress but the Majority is Still Unknown

As it is estimated that HC21 contains approximately 1% of the human genes (Deloukas *et al.*, 1998), it is expected that the same percentage of the monogenic phenotypes map on this chromosome. Only a small number of those disorders have been mapped to the chromosome and the mutant gene associated with these disorders is known (a total of 19 gene-disorder combinations, table 2). The sequence of HC21 will therefore greatly enhance the search for genes causing monogenic disorders. The positional cloning of the elusive genes for mapped phenotypes could now be accelerated because: (i) the wealth of microsatellite polymorphisms and their known location helps to precisely map the boundaries of the genomic regions that contain the genes by

linkage or allelic association studies; (ii) Mutation searches of candidate genes are rapid since the known sequence facilitates the experimental protocols. Furthermore, the study of the phenotypes of mice with targeted disruption of each HC21 homologue gene, will provide candidate phenotype-gene associations in humans. Candidate disease genes will also be developed based on sequence homologies of HC21 genes with those of *Drosophila* (Adams *et al.,* 2000), *C. Elegans* (The *C.elegans* Sequencing Consortium, 1998), Yeast (Anonymous1997) and other organisms in which the functional analysis of the encoded proteins, and phenotypes associated with gain or loss of function could be experimentally obtained (see for example the large-scale loss of function experiments using RNA interference in the *C. Elegans* (Gonczy *et al.*, 2000; Fraser *et al.*, 2000) or gene disruption in yeast (Ross-Macdonald *et al.*, 1999)). During the sequencing of HC21, the genes for 4 monogenic disorders namely CSTB for progressive myoclonus epilepsy EPM1 (Pennacchio *et al.*, 1996) (OMIM 254800 and 601145), AIRE for autoimmune polyglandular disease APECED al.(Nagamine *et al.*, 1997) (The Finnish-German APECED Consortium.Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy, 1997)(OMIM 240300), AML1 for dominant familial thrombocytopenia with propensity to acute myelogenous leukaemia al.(Song *et al.*, 1999) (OMIM 601399), and COL18A1 for Knobloch syndrome (Sertie *et al.*, 2000) (OMIM 267750) have been identified. Eight months after the availability of the sequence the molecular mysteries for an additional 2 monogenic disorders have been solved: the gene responsible for recessive deafness DFNB10 and DFNB8 is the transmembrane protease TMPRSS3 (Scott *et al.*, 2001) (OMIM 605511, 605316, 601072) and that for DFNB29 is claudin 14 (Wilcox *et al.*, 2001) (OMIM 605608). It is anticipated that, given the availability of an adequate number of patients, the positional identification of the majority of the disease-causing gene mutations in HC21 monogenic phenotypes will be determined within no more than 6 months from the precise mapping of the gene linked to the phenotype.

Complex Phenotypes and Malignancies

Undoubtedly, genes that predispose to complex common phenotypes map to HC21 and their discovery will be a great challenge for the years to come. Complex, common phenotypes are disorders with a strong genetic predisposition due to the effect of many genes, and a substantial environmental component. At present 2 such potential predisposing genes have been mapped to HC21: one for bipolar disease (BPAD-21)(Straub *et al.*, 1994) and another for familial combined hyperlipidemia (FCHL-21)(Pajukanta *et al.*, 1999).

In addition, HC21 may contain genes involved in certain cancers. DS is associated with childhood leukemias, namely AML type M7 and transient neonatal leukemia (Zipursky *et al.*, 1999; Iselius *et al.*, 1990). Loss of heterozygosity of the 21q11-21 region has been found in about half of squamous non-small cell lung carcinoma. The critical genomic region of a potential tumor suppressor gene has been narrowed down to approximately 3 Mb al.(Groet *et al.*, 2000).

Epidemiologic studies recently revealed that DS individuals are "protected" from certain cancers since their incidence is lower from that of the general population (Hasle *et al.*, 2000).

The determination of the sequence variability of HC21 (SNPs and other polymorphisms) and its consequent functional significance, the inexpensive and rapid detection of SNPs, and the study of the extend of linkage disequilibrium in human populations (Chakravarti *et al.*, 1984; Reich *et al.*, 2001a) may result in the determination of the disease predisposing mutations/variants (Risch, 2000; Weiss & Terwilliger, 2000). This however is one of the most uncertain predictions and some success stories such as that

TABLE 2

GENES ON CHROMOSOME 21 ASSOCIATED WITH VARIOUS PHENOTYPES

#	Gene Name	Mapping Position	SwProt	MIM Code	Description	Disorder
1	APP	21q21.2	P05067	104760	Amyloid a4 protein (beta-APP).	AD1
2	SOD1	21q22.1	P00441	147450	Superoxide dismutase (Cu-Zn).	ALS1
3	PRKCBP2	21q22.1	Q13516	606386	Prot. Kinase C-binding protein, BHLHB1, OLIG2, RACK17	T(14;21) T-cell ALL
4	IFNGT1	21q22.1	P38484	147569	Interferon-gamma receptor beta chain, IFNGR2	Mycobacterial infection, familial disseminated
5	KCNE2	21q22.1	Q9Y6J6	603796	K channel, voltage-gated, Isk-related; MiRP1	Long QT syndrome
6	KCNE1	21q22.1	P15382	176261	Isk slow voltage-gated potassium channel protein.	Jarvell & Lange-Nielsen syndrome
7	RUNX1	21q22.1	Q01196	151385	Acute myeloid leukemia 1 protein (AML1), CBFA2	AML, t8;21; FPDMM (601399)
8	CLDN14	21q22.2	O95500	605608	Claudin-14	DFNB29
9	HLCS	21q22.2	P50747	253270	Biotin-protein ligase; Holocarboxylase synthetase; HCS	HLCS deficiency
10	KCNJ6	21q22.2	P48051	600854	G protein-activated K+channel, GIRK2, KCNJ7	Weaver mouse
11	TMPRSS3	21q22.3	P57727	601072	Transmembrane serine protease 3	DFNB10, DFNB8
12	CBS	21q22.3	P35520	236200	Cystathionine beta-synthase	Homocystinuria
13	CRYAA	21q22.3	P02489	123580	Alpha crystallin A chain (CRYA1)	AD cataract zonular cent; AR cataract
14	CSTB	21q22.3	P04080	601145	Cystatin B (Stefin B) (CST6)	EPM1
15	AIRE	21q22.3	O43918	240300	Autoimmune regulator	APECED
16	ITGB2	21q22.3	P05107	600065	Integrin beta-2 (CD18 antigen)	Leucocyte Adhesion Deficincy
17	COL18A1	21q22.3	P39060	120328	Collagen XVIII, alpha-1 chain	Knobloch syndrome, KNO (267750)
18	COL6A1	21q22.3	P12109	120220	Collagen VI, alpha-1 chain	Bethlem myopathy
19	COL6A2	21q22.3	P12110	120240	Collagen VI, alpha-2 chain	Bethlem myopathy
20	?	21q21		602097		USH1E; Usher syndrome type 1E
21	?	21q22.3		236100		HPE1; Holoprocencephaly
22	?	21q21				AML type M7 and Transient Neonatal Leukemia
23	?	21q22.3				BPAD-21; Bipolar affective disorder
24	?	21q22.3				FCHL-21; Familial Combined hyperlipidemia, one locus
25	?	21q11-21				NSCLC 1; Deleted in Non-small cell lung cancer

described for type I diabetes (Horikawa *et al*., 2000) and Crohn's disease (Hugot *et al*., 2001) are certainly needed to increase the level of enthusiasm of the investigators.

Functional Genomics of Chromosome 21

The sequence of HC21 provides now the infrastructure for real progress on the understanding of the molecular pathophysiology of monogenic disorders, predispositions to complex traits, functional variants, deleterious mutations and aneuploidies of chromosome 21. The next great challenge in the post-sequencing era will be to understand the function and the interactions of the proteins coded by our genes. HC21 is a good model in which to develop large scale functional genomic approaches as it contains a relatively small number of well characterized genes (Fig. 1). All available model organisms for which genomic sequence information is available could be used in different aspects of the functional characterization of the HC21 genes and molecular pathophysiology of its disorders.

Acknowledgements

We thank all other members of our laboratory, past and present, for ideas, debates and experiments (in particular HS Scott, R Chrast, M Lalioti and M Guipponi); numerous Foundations including the Swiss National, and Child Care Foundations for supporting our research; all collaborating laboratories and the entire chromosome 21 research community. The content of this article is biased toward the contributions of our laboratory. Due to page space restrictions, many important papers are not included in the reference list.

References

ANONYMOUS (1997). The yeast genome directory. *Nature* 387: 5.

ADAMS, M.D., CELNIKER, S.E., HOLT, R.A., EVANS, C.A., GOCAYNE, J.D., *et al.* (2000). The genome sequence of Drosophila melanogaster. *Science* 287: 2185-2195.

ALTSHULER, D., POLLARA, V.J., COWLES, C.R., VAN ETTEN, W.J., BALDWIN, J., LINTON, L. and LANDER, E.S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407: 513-516.

ANTONARAKIS, S.E. (2001). BACking up the promises. *Nat Genet* 27: 230-232.

BAXTER, L.L., MORAN, T.H., RICHTSMEIER, J.T., TRONCOSO, J. and REEVES, R.H. (2000). Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse. *Hum Mol Genet* 9: 195-202.

BROWN, P.O. and BOTSTEIN, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21: 33-37.

CHAKRAVARTI, A., BUETOW, K.H., ANTONARAKIS, S.E., WABER, P.G., BOEHM, C.D. and KAZAZIAN, H.H. (1984). Nonuniform recombination within the human beta-globin gene cluster. *Am J Hum Genet* 36: 1239-1258.

CHRAST, R., SCOTT, H.S., MADANI, R., HUBER, L., WOLFER, D.P., PRINZ, M., AGUZZI, A., LIPP, H.P. and ANTONARAKIS, S.E. (2000). Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome. *Hum Mol Genet* 9: 1853-1864.

CHRAST, R., SCOTT, H.S., PAPASAVVAS, M.P., ROSSIER, C., ANTONARAKIS, E.S., BARRAS, C., DAVISSON, M.T., SCHMIDT, C., ESTIVILL, X., DIERSSEN, M., PRITCHARD, M. and ANTONARAKIS, S.E. (2000). The mouse brain transcriptome by SAGE: differences in gene expression between P30 brains of the partial trisomy 16 mouse model of Down syndrome (Ts65Dn) and normals. *Genome Res* 10: 2006-2021.

CHUMAKOV, I., RIGAULT, P., GUILLOU, S., OUGEN, P., BILLAUT, A., GUASCONI, G., GERVY, P., LEGALL, I., SOULARUE, P. and GRINAS, L. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* 359: 380-387.

COSTA, A.C., WALSH, K. and DAVISSON, M.T. (1999). Motor dysfunction in a mouse model for Down syndrome. *Physiol Behav* 68: 211-220.

DE SOUZA, S.J., CAMARGO, A.A., BRIONES, M.R., COSTA, F.F., NAGAI, M.A., *et al.* (2000). Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc Natl Acad Sci USA* 97: 12690-12693.

DELABAR, J.M., THEOPHILE, D., RAHMANI, Z., CHETTOUH, Z., BLOUIN, J.L., PRIEUR, M., NOEL, B. and SINET, P.M. (1993). Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur J Hum Genet* 1: 114-124.

DELOUKAS, P., SCHULER, G.D., GYAPAY, G., BEASLEY, E.M., SODERLUND, C., *et al.* (1998). A physical map of 30,000 human genes. *Science* 282: 744-746.

DEUTSCH, S., ISELI, C., BUCHER, P., ANTONARAKIS, S.E. and SCOTT, H.S. (2001). A cSNP Map and Database for Human Chromosome 21. *Genome Res* 11: 300-307.

DIERSSEN, M., FILLAT, C., CRNIC, L., ARBONES, M., FLOREZ, J. and ESTIVILL, X. (2001). Murine models for Down syndrome. *Physiol Behav* 73: 859-871.

DUNHAM, I., SHIMIZU, N., ROE, B.A., CHISSOE, S., HUNT, A.R., COLLINS, J.E., *et al.* (1999). The DNA sequence of human chromosome 22 *Nature* 402: 489-495.

EPSTEIN, C.J., AVRAHAM, K.B., LOVETT, M., SMITH, S., ELROY-STEIN, O., ROTMAN, G., BRY, C. and GRONER, Y. (1987). Transgenic mice with increased Cu/Zn-superoxide dismutase activity: animal model of dosage effects in Down syndrome. *Proc Natl Acad Sci USA* 84: 8044-8048.

ESCORIHUELA, R.M., VALLINA, I.F., MARTINEZ-CUE, C., BAAMONDE, C., DIERSSEN, M., TOBENA, A., FLOREZ, J. and FERNANDEZ-TERUEL, A. (1998). Impaired short- and long-term memory in Ts65Dn mice, a model for Down syndrome. *Neurosci Lett* 247: 171-174.

FRASER, A.G., KAMATH, R.S., ZIPPERLEN, P., MARTINEZ-CAMPOS, M., SOHRMANN, M. and AHRINGER, J. (2000). Functional genomic analysis of C. elegans chromosome I by systematic RNA interference. *Nature* 408: 325-330.

GARDINER, K. and DAVISSON, M.T. (2000). The sequence of human chromosome 21 and implications for research in Down syndrome. *Genome Biology* 1: 2.1-2.9.

GONCZY, P., ECHEVERRI, G., OEGEMA, K., COULSON, A., JONES, S.J., COPLEY, R.R., DUPERON, J., OEGEMA, J., BREHM, M., CASSIN, E., HANNAK, E., KIRKHAM, M., PICHLER, S., FLOHRS, K., GOESSEN, A., LEIDEL, S., ALLEAUME, A.M., MARTIN, C., OZLU, N., BORK, P. and HYMAN, A.A. (2000). Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III. *Nature* 408: 331-336.

GROET, J., IVES, J.H., JONES, T.A., DANTON, M., FLOMEN, R.H., SHEER, D., HRASCAN, R., PAVELIC, K. and NIZETIC, D. (2000). Narrowing of the region of allelic loss in 21q11-21 in squamous non-small cell lung carcinoma and cloning of a novel ubiquitin-specific protease gene from the deleted segment. *Genes Chromosomes Cancer* 27: 153-161.

HALUSHKA, M.K., FAN, J.B., BENTLEY, K., HSIE, L., SHEN, N., WEDER, A., COOPER, R., LIPSHUTZ, R. and CHAKRAVARTI, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22: 239-247.

HASLE, H., CLEMMENSEN, I.H. and MIKKELSEN, M. (2000). Risks of leukaemia and solid tumours in individuals with Down syndrome. *Lancet* 355: 165-169.

HATTORI, M., FUJIYAMA, A., TAYLOR, T.D., WATANABE, H., YADA, T., PARK, H.S., *et al.* (2000). The DNA sequence of human chromosome 21. *Nature* 405: 311-319.

HERAULT, Y., RASSOULZADEGAN, M., CUZIN, F. and DUBOULE, D. (1998). Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nat Genet* 20: 381-384.

HORIKAWA, Y., ODA, N., COX, N.J., LI, X., ORHO-MELANDER, M., HARA, M., *et al.* (2000). Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26: 163-175.

HUGOT, J.P., CHAMAILLARD, M., ZOUALI, H., LESAGE, S., CEZARD, J.P., *et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411: 599-603.

ISELIUS, L., JACOBS, P. and MORTON, N. (1990). Leukaemia and transient leukaemia in Down syndrome. *Hum Genet* 85: 477-485.

JI, Y., EICHLER, E.E., SCHWARTZ, S. and NICHOLLS, R.D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res* 10: 597-610.

KISHIMOTO, T.K., O'CONNER, K. and SPRINGER, T.A. (1989). Leukocyte adhesion deficiency. Aberrant splicing of a conserved integrin sequence causes a moderate deficiency phenotype. *J Biol Chem* 264: 3588-3595.

LANDER, E.S., LINTON, L.M., BIRREN, B., NUSBAUM, C., ZODY, M.C., BALDWIN,

J., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

LEJEUNE, J., GAUTIER, M. and TURPIN, R. (1959). Etude des chromosomes somatiques de neuf enfants mongoliens. *C.R.Acad.Sci.* 248: 1721-1722.

LIEMAN-HURWITZ, J., DAFNI, N., LAVIE, V. and GRONER, Y. (1982). Human cytoplasmic superoxide dismutase cDNA clone: a probe for studying the molecular biology of Down syndrome. *Proc Natl Acad Sci USA* 79: 2808-2811.

LIU, H.X., CARTEGNI, L., ZHANG, M.Q. and KRAINER, A.R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet* 27: 55-58.

LYNN, A., KASHUK, C., PETERSEN, M.B., BAILEY, J.A., COX, D.R., ANTONARAKIS, S.E. and CHAKRAVARTI, A. (2000). Patterns of meiotic recombination on the long arm of human chromosome 21. *Genome Res* 10: 1319-1332.

MACBEATH, G. and SCHREIBER, S.L. (2000). Printing proteins as microarrays for high-throughput function determination. *Science* 289: 1760-1763.

MAROUN, L.E., HEFFERNAN, T.N. and HALLAM, D.M. (2000). Partial IFN-alpha/beta and IFN-gamma receptor knockout trisomy 16 mouse fetuses show improved growth and cultured neuron viability. *J Interferon Cytokine Res* 20: 197-203.

NAGAMINE, K., PETERSON, P., SCOTT, H.S., KUDOH, J., MINOSHIMA, S., HEINO, M., KROHN, K.J., LALIOTI, M.D., MULLIS, P.E., ANTONARAKIS, S.E., KAWASAKI, K., ASAKAWA, S., ITO, F. and SHIMIZU, N. (1997). Positional cloning of the APECED gene. *Nat Genet* 17: 393-398.

PAJUKANTA, P., TERWILLIGER, J.D., PEROLA, M., HIEKKALINNA, T., NUOTIO, I., ELLONEN, P., PARKKONEN, M., HARTIALA, J., YLITALO, K., PIHLAJAMAKI, J., PORKKA, K., LAAKSO, M., VIIKARI, J., EHNHOLM, C., TASKINEN, M.R. and PELTONEN, L. (1999). Genomewide scan for familial combined hyperlipidemia genes in finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am J Hum Genet* 64: 1453-1463.

PANDEY, A. and MANN, M. (2000). Proteomics to study genes and genomes. *Nature* 405: 837-846.

PENNACCHIO, L.A., LEHESJOKI, A.E., STONE, N.E., WILLOUR, V.L., VIRTANEVA, K., MIAO, J., D'AMATO, E., RAMIREZ, L., FAHAM, M., KOSKINIEMI, M., WARRINGTON, J.A., NORIO, R., DE LA CHAPELLE, A., COX, D.R. and MYERS, R.M. (1996). Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). *Science* 271: 1731-1734.

PINKEL, D., SEGRAVES, R., SUDAR, D., CLARK, S., POOLE, I., KOWBEL, D., COLLINS, C., KUO, W.L., CHEN, C., ZHAI, Y., DAIRKEE, S.H., LJUNG, B.M., GRAY, J.W. and ALBERTSON, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 20: 207-211.

REEVES, R.H., IRVING, N.G., MORAN, T.H., WOHN, A., KITT, C., SISODIA, S.S., SCHMIDT, C., BRONSON, R.T. and DAVISSON, M.T. (1995). A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nat.Genet.* 11: 177-184.

REICH, D.E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P.C., RICHTER, D.J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S.F., WARD, R. and LANDER, E.S. (2001a). Linkage disequilibrium in the human genome. *Nature* 411: 199-204.

REICH, D.E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P.C., RICHTER, D.J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S.F., WARD, R. and LANDER, E.S. (2001b). Linkage disequilibrium in the human genome. *Nature* 411: 199-204.

RISCH, N.J. (2000). Searching for genetic determinants in the new millennium. *Nature* 405: 847-856.

REYMOND, A., FRIEDLT, M., HENRICHSEN, C.N. et al. (2001). From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics* 78: 46-54.

ROSS-MACDONALD, P., COELHO, P.S., ROEMER, T., AGARWAL, S., KUMAR, A., JANSEN, R., CHEUNG, K.H., SHEEHAN, A., SYMONIATIS, D., UMANSKY, L., HEIDTMAN, M., NELSON, F.K., IWASAKI, H., HAGER, K., GERSTEIN, M., MILLER, P., ROEDER, G.S. and SNYDER, M. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402: 413-418.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S.C., KAKOL, J.M., STEIN, L.D., *et al.* (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.

SAGO, H., CARLSON, E.J., SMITH, D.J., KILBRIDGE, J., RUBIN, E.M., MOBLEY, W.C., EPSTEIN, C.J. and HUANG, T.T. (1998). Ts1Cje, a partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioral abnormalities. *Proc Natl Acad Sci USA* 95: 6256-6261.

SAGO, H., CARLSON, E.J., SMITH, D.J., RUBIN, E.M., CRNIC, L.S., HUANG, T.T. and EPSTEIN, C.J. (2000). Genetic dissection of region associated with behavioral abnormalities in mouse models for Down syndrome. *Pediatr Res* 48: 606-613.

SCHORK, N.J., FALLIN, D. and LANCHBURY, J.S. (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet* 58: 250-264.

SCOTT, H.S., KUDOH, J., WATTENHOFER, M., SHIBUYA, K., BERRY, A., CHRAST, R., *et al.* (2001). Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. *Nat Genet* 27: 59-63.

SERTIE, A.L., SOSSI, V., CAMARGO, A.A., ZATZ, M., BRAHE, C. and PASSOS-BUENO, M.R. (2000). Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum Mol Genet* 9: 2051-2058.

SHINOHARA, T., TOMIZUKA, K., MIYABARA, S., TAKEHARA, S., KAZUKI, Y., INOUE, J., KATOH, M., NAKANE, H., IINO, A., OHGUMA, A., IKEGAMI, S., INOKUCHI, K., ISHIDA, I., REEVES, R.H. and OSHIMURA, M. (2001). Mice containing a human chromosome 21 model behavioral impairment and cardiac anomalies of Down syndrome. *Hum Mol Genet* 10: 1163-1175.

SONG, W.J., SULLIVAN, M.G., LEGARE, R.D., HUTCHINGS, S., TAN, X., KUFRIN, D., *et al.* (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* 23: 166-175.

STRAUB, R.E., LEHNER, T., LUO, Y., LOTH, J.E., SHAO, W., SHARPE, L., ALEXANDER, J.R., DAS, K., SIMON, R. and FIEVE, R.R. (1994). A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. *Nat Genet* 8: 291-296.

SUMARSONO, S.H., WILSON, T.J., TYMMS, M.J., VENTER, D.J., CORRICK, C.M., KOLA, R., LAHOUD, M.H., PAPAS, T.S., SETH, A. and KOLA, I. (1996). Down's syndrome-like skeletal abnormalities in Ets2 transgenic mice. *Nature* 379: 534-537.

THE BAC RESEARCH CONSORTIUM (2001). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* 409: 953-958.

THE C.ELEGANS SEQUENCING CONSORTIUM (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* 282: 2012-2018.

THE FINNISH-GERMAN APECED CONSORTIUM.AUTOIMMUNE POLYENDOCRINOPATHY-CANDIDIASIS-ECTODERMAL DYSTROPHY (1997). An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat Genet* 17: 399-403.

VELCULESCU, V.E., ZHANG, L., VOGELSTEIN, B. and KINZLER, K.W. (1995). Serial analysis of gene expression. *Science* 270: 484-487.

VENTER, J.C., ADAMS, M.D., MYERS, E.W., LI, P.W., MURAL, R.J., SUTTON, G.G., *et al.* (2001). The sequence of the human genome. *Science* 291: 1304-1351.

WARREN, A.C., SLAUGENHAUPT, S.A., LEWIS, J.G., CHAKRAVARTI, A. and ANTONARAKIS, S.E. (1989). A genetic linkage map of 17 markers on human chromosome 21. *Genomics* 4: 579-591.

WEISS, K.M. and TERWILLIGER, J.D. (2000). How many diseases does it take to map a gene with SNPs?. *Nat Genet* 26: 151-157.

WILCOX, E.R., BURTON, Q.L., NAZ, S., RIAZUDDIN, S., SMITH, T.N., PLOPLIS, B., BELYATSEVA, I., BEN-YOSEF, T., LIBURD, N.A., MORELL, R.J., KACHAR, B., WU, D.K., GRIFFITH, A.J., RIAZUDDIN, S. and FRIEDMAN, T.B. (2001). Mutations in the gene encoding tight junction claudin-14 cause recessive deafness DFNB29. *Cell* 104: 165-172.

YATES, J.R. (2000). Mass spectrometry. From genomics to proteomics. *Trends Genet* 16: 5-8.

ZHENG, B., SAGE, M., SHEPPEARD, E.A., JURECIC, V. and BRADLEY, A. (2000). Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications. *Mol Cell Biol* 20: 648-655.

ZIPURSKY, A., BROWN, E.J., CHRISTENSEN, H. and DOYLE, J. (1999). Transient myeloproliferative disorder (transient leukemia) and hematologic manifestations of Down syndrome. *Clin Lab Med* 19: 157-67, vii.