

CHROMOSOME 21 AND DOWN SYNDROME: FROM GENOMICS TO PATHOPHYSIOLOGY

*Stylianos E. Antonarakis**, *Robert Lyle**, *Emmanouil T. Dermitzakis[‡]*,
*Alexandre Reymond** and *Samuel Deutsch**

Abstract | The sequence of chromosome 21 was a turning point for the understanding of Down syndrome. Comparative genomics is beginning to identify the functional components of the chromosome and that in turn will set the stage for the functional characterization of the sequences. Animal models combined with genome-wide analytical methods have proved indispensable for unravelling the mysteries of gene dosage imbalance.

ROBERTSONIAN TRANSLOCATIONS

Occurs when the long arms of two acrocentric chromosomes fuse at the centromere and the two short arms are lost. A non-robertsonian rearrangement is a chromosomal rearrangement other than a Robertsonian translocation.

**Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva, 1 rue Michel-Servet, 1211 Geneva, Switzerland.*

[‡]*Population and Comparative Genomics, The Wellcome Trust, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.*

Correspondence to S.E.A.

e-mail:

Stylianos.Antonarakis@medecine.unige.ch

doi:10.1038/nrg1448

Genomic aneuploidy, defined as an abnormal number of copies of a genomic region, is a common cause of human genetic disorders. Classically, the term aneuploidy was restricted to the presence of supernumerary copies of whole chromosomes (trisomy), or absence of chromosomes (monosomy), but we extend this definition to include deletions or duplications of subchromosomal regions.

Trisomy 21 is a model of all human disorders that are the result of supernumerary copies of a genomic region. In this review, we focus on **Down syndrome** (DS) and human chromosome 21 (HSA21) to show the effect that genomics has had on our understanding of the 'disorders of the genome'. We discuss the recent advances in genome sequencing, comparative genome analysis, functional genome exploration, use of model organisms and lessons from models of gene overexpression. We also discuss the consequences of genomic dosage imbalance owing to an extra copy of a genomic segment (trisomy).

These are exciting times for genomic disorders, such as trisomy 21, because we now have the necessary tools to understand how three copies of a functional genomic element result in abnormal phenotypes.

Trisomies

According to the size of the triplicated genomic region, trisomies can be divided into four categories: complete, or whole-chromosome, trisomies; partial trisomies;

microtrisomies and triplication of single genes or single functional genomic elements.

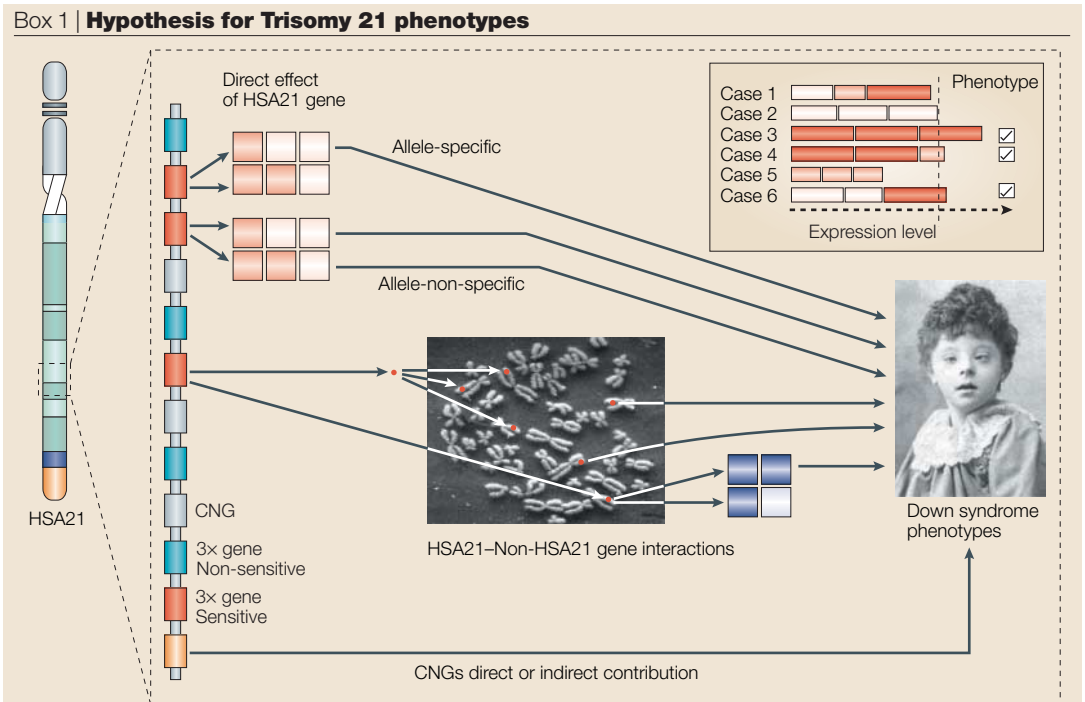
Whole-chromosome trisomies. Whole-chromosome trisomies that result from meiotic or mitotic non-disjunction events are common in humans; they account for ~0.3–0.5% of live births. Trisomy for HSA21, which results in Down syndrome and occurs at ~1 in 750 live births, is the most frequent event. Trisomies are often observed in a significant proportion of spontaneous abortions; for example, trisomy 16 is found in 1 out of 13, and trisomy 21 in 1 out of 43 such abortions¹.

Partial trisomies. Partial (or segmental) trisomies that involve a genomic region of more than one chromosomal band (usually larger than 5 Mb) are much less frequent than whole-chromosome trisomies. They usually result from abnormal meiosis and segregation in individuals with balanced chromosomal rearrangements. One in about 1,800 newborns have an unbalanced, non-robertsonian rearrangement and approximately half of these are partial trisomies. Unbalanced ROBERTSONIAN TRANSLOCATIONS with trisomies of the long arms of ACROCENTRIC CHROMOSOMES occur in 1 of about 14,000 newborns¹.

Microtrisomies. This type of trisomy is defined here as the partial trisomy of a genomic segment that is shorter

than 3–5 Mb and that is not detectable by routine high-quality cytogenetic analysis. It is also known as segmental duplication. The incidence of microtrisomies is, at present, unknown. Most are due to unequal crossovers in meiosis, mediated by the presence of interchromosomal duplicons or low copy repeats (LCRs; 10–100 Kb each). These duplicons, which make up ~5% of the human genome², promote unequal recombination events that

lead to microtrisomies and micromonosomies. Microduplications are seen, for example, in many cases of Charcot–Marie–Tooth disease, type 1A (CMT1A). This neurological disorder is caused by a ~1.4 Mb duplication of chromosome 17p12, the result of recurrent non-allelic homologous recombination between duplicons that flank the duplicated segment³. Another more recent example of a microtrisomy is that of SHFM3



We might consider that there are two categories of genes on human chromosome 21 (HSA21); those that are dosage sensitive (that is, three copies result in phenotypic effects; shown in red) and contribute to the phenotypes of Down syndrome (DS), and those that are not dosage sensitive (green) and therefore do not contribute to any of the phenotypes. This could be true for both protein-coding genes and non-coding RNA (ncRNA) genes (including microRNA genes).

The effect of some dosage-sensitive genes on the phenotypes might be allele specific. According to this hypothesis, certain combinations of alleles might contribute to the phenotype, whereas others might be silent. The effect of the combination of alleles could be either qualitative (alleles with amino-acid variation) or quantitative (alleles with variation in gene expression level). A threshold effect of total transcript/output or amount of protein could be envisaged in the following way: a phenotype is only present if the total transcript/protein level from the combination of the three alleles reaches a crucial amount. If the output from the combination of the three alleles does not reach the crucial amount, the associated phenotype is not manifested. The insert in the figure explains the hypothetical threshold effect of gene expression in DS phenotypes. The allelic combination in cases 3, 4 and 6 result in total expression levels beyond the threshold and in the appearance of the phenotype; by contrast, the allelic combination in cases 1, 2 and 5 are not sufficient for the necessary pathological overexpression.

The effect of the dosage sensitive genes could have a direct or indirect effect on the phenotype. The indirect effect might be due to the interaction of HSA21 genes or gene products with non-HSA21 genes or gene-products. This interaction could be allele-specific; only certain combinations of alleles of non-HSA21 genes (shown in light and dark blue) would contribute to susceptibility of specific phenotypes. The individual genetic background could therefore contribute to the phenotypic variation. So, the global dysregulation of the individual transcriptome or proteome could contribute to the DS phenotypes.

Finally, triplication of certain conserved functional non-genic sequences (CNGs — shown in grey) on HSA21 (or in the entire genome) might contribute to the DS phenotypes, as might the triplication of other functional but non-conserved genomic regions. The molecular mechanism of this type of locus dosage imbalance is unknown but it will probably be determined after the functional characterization of CNGs and other important genomic elements.

The picture of the trisomy 21 female was reproduced, with permission, from REF. 90 © (1998) Royal Society of Medicine Press.

ACROCENTRIC CHROMOSOMES
A chromosome, the centromere of which is located near one end.

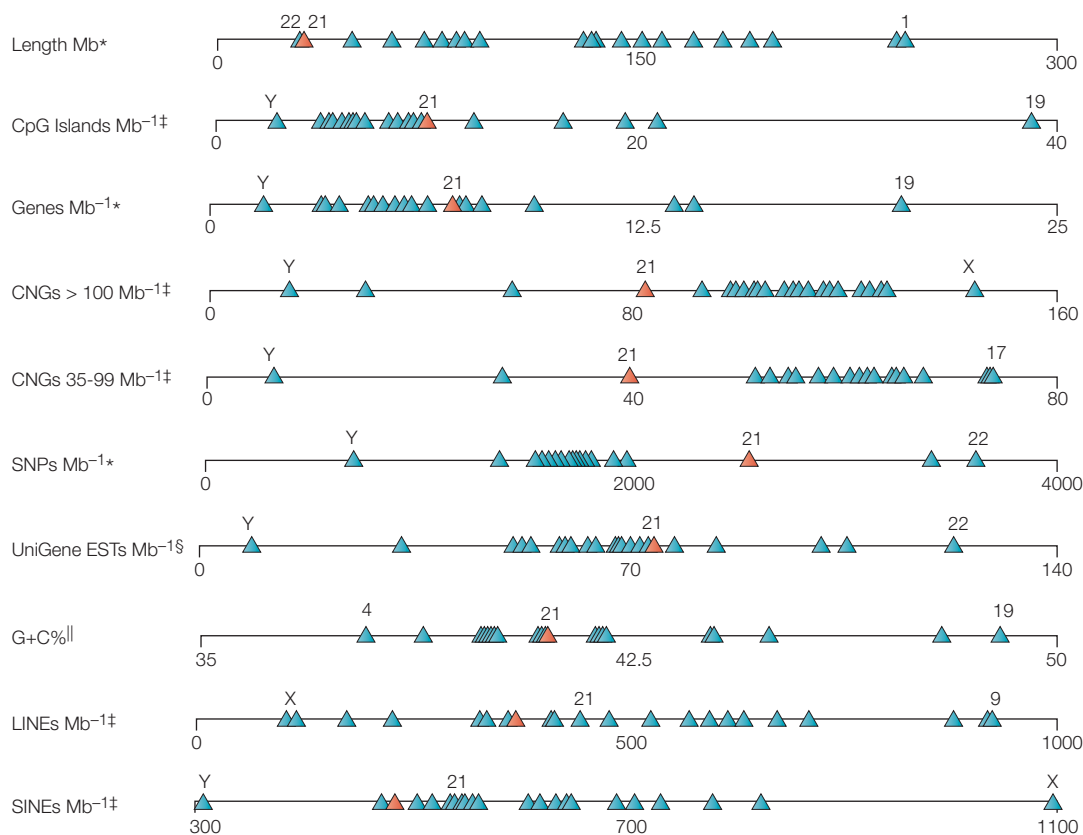


Figure 1 | Features of human chromosome 21 (HSA21). Each line describes a particular feature shown on the left. The position of HSA21 in each scale is shown as a red triangle. All other chromosomes are shown as blue triangles. The identity of chromosomes with extreme values per feature is also shown. The sources of information are: *ENSEMBL 34; †GALA 34; ‡NCBI 34; §REF. 20. The high SNP density is probably a biased estimate owing to the updated estimate provided by REF. 41. CNGs, conserved non-genic sequences; LINEs, long interspersed nuclear elements (such as L1 repeats) are retroelements present in over 100,000 copies in the mammalian genome; SINEs, short interspersed-repeat transposable elements.

(a form of ECTRODACTYLY), which is caused by a duplication of ~0.5 Mb on chromosome 10q24 (REF. 4). New pathogenic or polymorphic microtrisomies will probably be identified using diagnostic methods such as BAC ARRAY CGH (comparative genomic hybridization)⁵.

Single-gene duplications. Duplications of only one gene or one functional genomic element can also be pathogenic; for example, duplicated *PLP1* causes Pelizaeus–Mertzbacher disease, and duplicated *PMP22* causes CMT1A^{6,7}. The genomic duplications that include these genes also encompass additional coding sequences, but studies of transgenic mice that carry single gene duplications (either of *Plp1* or *Pmp22*) indicate that the abnormal phenotype is caused by duplication of these two genes^{8,9}.

Evidence from experiments using transgenic mice indicates that single gene dosage imbalance might cause abnormal phenotypes. For example, strain B6;SJL-Tg(EPO33)72Pt/J, with a single copy of human *EPO*.

Finally, the duplication of a non-genic, but functional, genomic element might also be pathogenic. A recent example of this comes from the duplication of ~30 Kb containing a conserved regulatory element that

maps ~1 Mb from the sonic hedgehog (*Shh*) locus and causes PREAXIAL POLYDACTYLY in the sasquatch (*Ssq*) mouse mutant¹⁰.

Genomic dosage imbalance frequently occurs in somatic cancer cells. Chromosomal instability of tumour cells, resulting in various trisomies and monosomies, is associated with tumour progression¹¹; this topic deserves a separate discussion and we do not return to it here. Below, we focus on constitutional trisomies, that is, those that are present in all cells, and in particular on trisomy 21 as it is the most frequent example and serves as a model for all constitutional trisomies.

Origin of the extra HSA21

The parental origin of the supernumerary HSA21 in trisomy 21 was determined using highly informative polymorphic markers in DNA from parents and DS offspring. DNA markers near the centromere indicate the stage of meiosis during which the segregation error occurred. Homozygosity for all polymorphic markers throughout 21q (the long arm of HSA21) is indicative of a mitotic, post-zygotic error. **Supplementary information S1** (table) lists the various origins of human trisomy 21.

ECTRODACTYLY

Limb malformations characterized by digit loss.

BAC ARRAY CGH

A microarray that contains DNA from bacterial artificial chromosomes, which is used in comparative genomic hybridization to determine copy number differences of DNA sequences.

PREAXIAL POLYDACTYLY

Polydactyly (additional fingers or toes) on the thumb side of the hand and the big-toe side of the foot.

Table 1 | **Characteristics of human chromosome 21**

	Property	Description	Comments	References
Physical size	21p	5–15 Mb		91
	21q	33.5 Mb		18
	% human genome	21q = ~1%		19
Genetic size	Sex average	61.7–67.3 cM		92,93
	Male meioses	47.3–54.3 cM		92,93
	Female meioses	76.4–80.1 cM		92,93
Recombination rate	Sex average	500 kb/cM		93
	Male meioses	420 kb/cM		93
	Female meioses	620 kb/cM		93
	Range	250–2125 kb/cM		93
Repeat content	40.06%			
	SINEs	10.84%		18
	LINEs	15.15%		18
	LTR elements	9.21%		18
	DNA transposons	2.4%		18
Paralogous regions	Intra HSA21	1.4%		18,19
	Interchromosomal	1.6%		18,19
Variability	Nucleotide diversity	5.19×10^{-4}		
	SNPs	105,334	Includes 335 coding synonymous, 355 coding non-synonymous, 27,433 intronic, 7,827 in UTRs	*
	Microsatellites	5,987	These are short sequence repeats, potentially polymorphic	18
Genes (protein coding)	Swiss-Prot	212 (including 21 KRTAP)		†
	Ensembl	261	(244 known and 17 novel)	§
	NCBI	263 (including 44 KRTAP)		
	Berlin list	284 (including 27 KRTAP)	240 confirmed, 26 predicted by ESTs, 18 computationally predicted	¶
	Denver analysis	364	(170 well conserved in mouse, 83 minimally conserved in mouse, 111 non-conserved)	25
Pseudogenes	Ensembl	31		§
	NCBI	49		
	Berlin list	83		¶
	Yale analysis	77 processed, 72 non-processed		94
Non-coding RNAs	miRNAs	5	E.T.D. & S.E.A., unpublished data	
	rRNA genes	Estimated ~40–50 in 21p (RNR4A and RNR4B)		95
Other	CNG content	2,978	>100 bp; >70 ungapped identity 35	#
	CpG islands	149	In blood cells: 103 non-methylated (Me); 31 fully Me; 7 composite Me; 3 monoallelic Me	96
	Mouse orthologous genomic regions	MMU16; 76.7–100 Mb MMU17; 30.6–31.7 Mb MMU10; 78.4–80.6 Mb		**

CNG, conserved non-genic; HSA, human chromosome; kRTAP, keratin-associated proteins; LINEs, long interspersed nuclear elements (such as L1 repeats) are retroelements present in over 100,000 copies in the mammalian genome; LTR, long terminal repeat; miRNA, micro RNAs; NCBI, National Center for Biotechnology Information; SINEs, short interspersed-repeat transposable elements. *<http://www.ncbi.nlm.nih.gov/SNP/>; †<http://au.expas.org/cgi-bin/lists?humchr21.txt>; §http://www.ensembl.org/Homo_sapiens/mapview?chr=21; ||<http://www.ncbi.nlm.nih.gov/mapview/map.cgi?ORG=hum&MAPS=idiogr.est,loc&LINKS=ON&VERBOSE=ON&CHR=21>; ¶http://chr21.molgen.mpg.de/chr21_catalog9.html; # <http://globin.cse.psu.edu/gala/>; **http://www.ensembl.org/Homo_sapiens/syntenview.

For an extensive discussion of the origin of aneuploidy in all human trisomies, and the link between recombination and non-disjunction, the reader is referred to the excellent recent review by Hassold and Hunt¹². The origin of translocation trisomy 21 is discussed in REFS 13,14.

Phenotypic variability of DS

There are two types of phenotypes that are observed in trisomy 21: those seen in every patient and those that occur only in a fraction of affected individuals (see **Supplementary information S2 (table)**^{15,16}. For example, cognitive impairment is present in all patients with

DS, whereas congenital heart defect occurs in ~40% and ATRIOVENTRICULAR CANAL in ~16% of patients. DUODENAL STENOSIS/ATRESIA, **Hirschsprung disease** and acute megakaryocytic leukemia occur 250-, 30- and 300-times more frequently, respectively, in patients with DS than in the general population. In addition, for any given phenotype there is considerable variability (severity) in expression. For example, the extent of cognitive impairment varies widely in individuals with DS¹⁷.

There are several working hypotheses that attempt to explain the phenotypic variability in trisomy 21 (or other trisomies; see BOX 1). The gene (or genomic functional unit) dosage imbalance is the main molecular mechanism that requires further investigation.

Genomic content of HSA21

The almost complete, high-quality sequence of 21q was published in May 2000 (REF 18). There are three cloning gaps on 21q, each of which is only 20–30 Kb long. There are, therefore, 4 contigs for 21q (from centromere to telomere): 28,602; 229; 1,378; 3,432 Kb for a total length of 33,642,989 nucleotides (NCBI Build 34). Only a region of 281 Kb has been sequenced from the short arm of HSA21, 21p. FIGURE 1 provides a comparison of chromosome 21 features with those of other chromosomes. The comparisons are based on the recent NCBI builds, and on published sequences^{19,20}.

HSA21 is among the smallest of human chromosomes and its 33.6 Mb-long arm represents ~1% of the total sequences obtained (3020.3 Mb of NCBI Build 34 version 1). TABLE 1 lists other characteristics of HSA21. The total number of genes (protein-coding and non-coding RNAs (ncRNAs)) on 21q has not yet been conclusively determined. A total of 225 genes was estimated when the initial sequence of 21q was published¹⁸, 35% of those are homologous to *Drosophila melanogaster*, 35% to *Caenorhabditis elegans* and 17% to *Saccharomyces cerevisiae* genes. Subsequent on-going analysis based on computational methods, EST sequencing, laboratory verification and comparative genome analysis, resulted in an estimated 261–364 protein-coding genes (TABLE 1, references wherein and REFS 18,21–28). These include potential transcripts that are supported by only one 'spliced' EST.

There are gene-rich and gene-poor regions on 21q, which correlate with the G+C content. The so-called L isochores (G+C <43%) contain few genes (1 per 300 Kb), whereas the H3 isochores in GIEMSA light chromosomal bands (G+C >48%) contain most of the genes (1 per 58 Kb, according to the first annotation).

More research is needed for the complete and correct genic annotation of HSA21. This is particularly true for single exon genes, ncRNAs and rare transcripts. An international collaborative effort to re-annotate HSA21 is in progress. Computational analysis has provided initial evidence for 5 microRNAs (miRNAs) on 21q, although their function and potential involvement in DS remain unknown.

Comparative sequence analysis of human and other genomes, particularly that of the mouse, has resulted in the discovery of novel genes and verification of putative transcripts^{25,29}. The availability of other mammalian

genome sequences will further increase our ability to recognize additional genes and validate initial gene predictions. The 33.5 Mb chimpanzee chromosome 22, which is homologous to HSA21, has recently been sequenced³⁰. There are 1.44% single nucleotide substitutions between the human and chimpanzee sequences, and nearly 68,000 insertions and deletions (indels); more than 99% of the indels are shorter than 300 bp. Remarkably, 86% of the 231 unambiguous coding sequences in both species show amino-acid differences. Indels within coding regions represent one of the main mechanisms that lead to protein diversity.

An interesting approach to improving the functional annotation of HSA21 involves studying transcriptional activity of the entire chromosome. In one study, oligonucleotide arrays that contain probes spaced, on average, 35 bp apart and that cover the entire 21q were used to estimate mRNA expression from 11 different human cancer cell lines³¹. As many as 9.7% of the probes showed positive hybridization signals in 5 of the 11 cell lines³¹, which indicate that the potentially transcribed genome is ~10-fold larger than the current genic annotation. This transcription potential could be due to additional unidentified genes, RNA transcripts without protein-coding capacity, alternative RNA isoforms of previously annotated genes, or 'illegitimate' non-functional transcription. Subsequent analysis of these data for chromosomes 21 and 22, showed that ~49% of the observed hybridization signal (transcription) was outside the known annotation³²; 65% of these transcripts were verified by reverse transcriptase PCR (RT-PCR). These results emphasize that the annotation of HSA21 sequence is far from complete.

Another study combined chromatin immunoprecipitation and high-density oligonucleotide arrays to map more than 300 binding sites of the three transcription factors, Sp1, cMyc, and p53 on HSA21 (REF 33). As expected, these sites clustered near 5' promoters of protein-coding genes and CPG ISLANDS but many were found in 3' ends of known protein-coding genes. Most of the binding sites identified are, in fact, associated with ncRNAs of as yet unknown function. Intriguingly, this study provided initial evidence for widespread antisense transcription, the functional significance of which remains unknown.

Functional analysis of the HSA21-encoded proteins is one of the research priorities for the understanding of DS. Analysis of Swiss-Prot-listed **HSA21-encoded proteins** with Interpro for protein families, domains and functional sites, gave 207 entries (see Online links box). Analysis of the same proteins using the **Gene Ontology** Annotation (see Online links box) shows that they are involved in 87 different biological processes, have 81 different molecular functions and are localized in 26 different cellular components. The most frequent molecular function is DNA binding and transcription factor activity (15 proteins); the most common cellular localizations are in the nucleus and the plasma membrane (19 and 15 proteins, respectively); and the most common biological process that they are involved in is signal transduction (11 proteins).

ATRIOVENTRICULAR CANAL
A complex congenital heart defect characterized by atrial septal defect, ventricular septal defect and abnormalities of the tricuspid and mitral valves.

DUODENAL STENOSIS/ATRESIA
Narrowing (stenosis) or complete obliteration of the duodenal lumen.

GIEMSA BANDS
Chromosome banding pattern produced by Giemsa staining.

CPG ISLANDS
A genomic region of about one kilobase that contains close to the theoretical, expected frequency of the CpG dinucleotide.

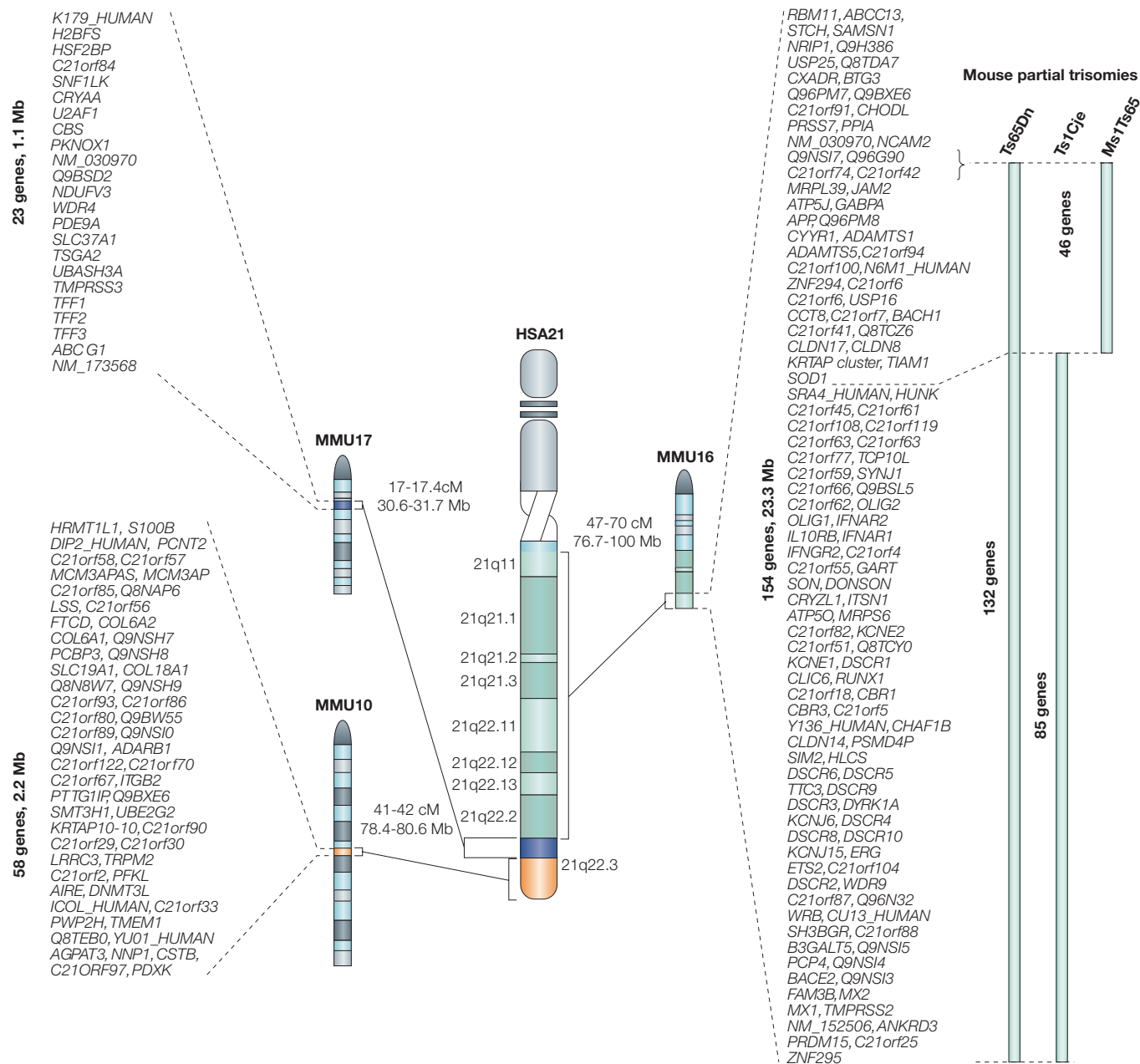


Figure 2 | **Regions of synteny between human chromosome 21 (HSA21) and mouse chromosomes (MMUs) 16, 17 and 10.** There are three partial trisomy mouse models of human trisomy 21, all trisomic for a portion of MMU16. The gene content of these partial trisomies is shown on the right. The list of gene names in Ts65Dn is from REF. 43. Sequence and gene data are taken from [ENSEMBL](#).

The completion of the sequencing of HSA21 (REF. 18) and of the mouse genome^{29,34} provided the first opportunity to compare DNA sequences of entire chromosomes of two mammalian species, so that conserved genomic elements that are likely to be functional can be identified. The comparison of the 33.5 Mb of HSA 21q with the orthologous mouse genomic regions on MMU16, MMU17 and MMU10 revealed 3,491 sequences of ≥ 100 bp and $\geq 70\%$ identity without gaps. Unexpectedly, only 1,229 of those corresponded to exons of previously known genes³⁵. Remarkably, the

remaining 2,262 conserved sequences mapped preferentially to the gene-poor regions of 21q. About 80% of these sequences are located in intergenic regions, with the remaining 20% in introns. Experimental, bioinformatic and evolutionary analysis strongly suggested that these sequences are not 'functionally' transcribed, and do not correspond to protein-coding genes^{35,36}. These sequences, which we call conserved non-genic (CNG), account for $\sim 1\%$ of the HSA21 sequence. The CNGs are highly conserved in mammals; from primates to monotremes, to marsupials³⁷⁻⁴⁰. The remarkable conservation, over

Table 2 | **Mouse models of Down syndrome***

Model	Genotype	Neurological phenotype
Segmental trisomy		
Ts16	Trisomy 16	Reduced brain size and some structural changes
Ts65Dn	Trisomic for <i>App-Znf295</i> (~132 known genes)	Learning and behavioural deficits 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 Failure of retrograde NGF signalling
Ts1Cje	Trisomic for <i>Znf295-Sod1</i> (~85 known genes)	Learning and behavioural deficits (less severe than in Ts65Dn) Cerebellar dysmorphology (a subset of Ts65Dn features)
Ms1Cje/Ts65Dn (~46 known genes)	Trisomic for <i>App-Sod1</i>	Learning deficits (less severe than in Ts1Cje) Cerebellar dysmorphology (a subset of Ts65Dn features)
C10-2	Chimeric mice with free HSA21	Learning and behaviour impairment
Single genes		
TgSod1	Transgenic for <i>SOD1</i>	Learning defects
TgPfkf1	Highly overexpressed cDNA	
TgS100 β	2–12 copies	Astrocytosis, neurite degeneration, slower to habituate to novelty, behavioural defects
TgApp	Low copy of YAC	Cognitive/behavioural defects
TgEts2	Highly overexpressed cDNA	
TgHmg14	2–6 copies	
TgDyrk1A	1–3 copies of YAC and highly overexpressed cDNA	Learning/memory defects
TgSim2	1–2 copies of BAC and highly overexpressed cDNA	Behavioural defects

*For references, see text and Refs 48,97,98. NGF, nerve growth factor.

more than 150 million years of evolution since the common mammalian ancestor, strongly indicates that most CNGs are functional, although their function is unknown.

Genomic variability

The variability of HSA21 might be partially responsible for the different DS phenotypes. It is therefore necessary to determine the common and rare DNA variants on this chromosome. Approximately 1.26% of HSA21 consists of short sequence repeats that might be polymorphic¹⁸. In addition, 105,334 probable SNPs have been identified (see Online links box).

The initial assessment of LINKAGE DISEQUILIBRIUM structure of polymorphic sites involved re-sequencing 20 haploid human genomes from different ethnic groups using tiled oligonucleotide arrays of the non-repetitive fraction of 21q (REF. 41). For this study, HSA21 from each individual was separated in somatic cell hybrids and haploid DNA hybridization probes were

generated by long-range PCR. A total of 24,386 SNPs were identified with the rare allele observed at least twice (in addition to 11,603 nucleotide variants that were seen only once). There were 4,135 inferred HAPLOTYPE BLOCKS with an average length per block of 7.83 Kb (REF. 41). Such studies describing the chromosome-wide linkage disequilibrium maps provide the genomic infrastructure for the determination of SNPs involved in gene expression variation, and in the predisposition to different variable phenotypic manifestations of DS and to common, complex phenotypes.

Mouse models and phenotypes

HSA21 is homologous to mouse chromosomal regions that map three different chromosomes. From 21cen to 21qter, about 23.2 Mb are homologous to MMU16, 1.1 Mb to MMU17, and 2.3 Mb to MMU10 (FIG. 2). Importantly, none of the existing mouse models perfectly mimic the chromosomal abnormality that is observed in DS.

LINKAGE DISEQUILIBRIUM
Refers to the fact that particular alleles at nearby sites can co-occur on the same haplotype more often than is expected by chance.

HAPLOTYPE BLOCKS
Long stretches (tens of megabases) along a chromosome that have low recombination rates (and relatively few haplotypes). Adjacent blocks are separated by recombination hot spots (short regions with high recombination rates).

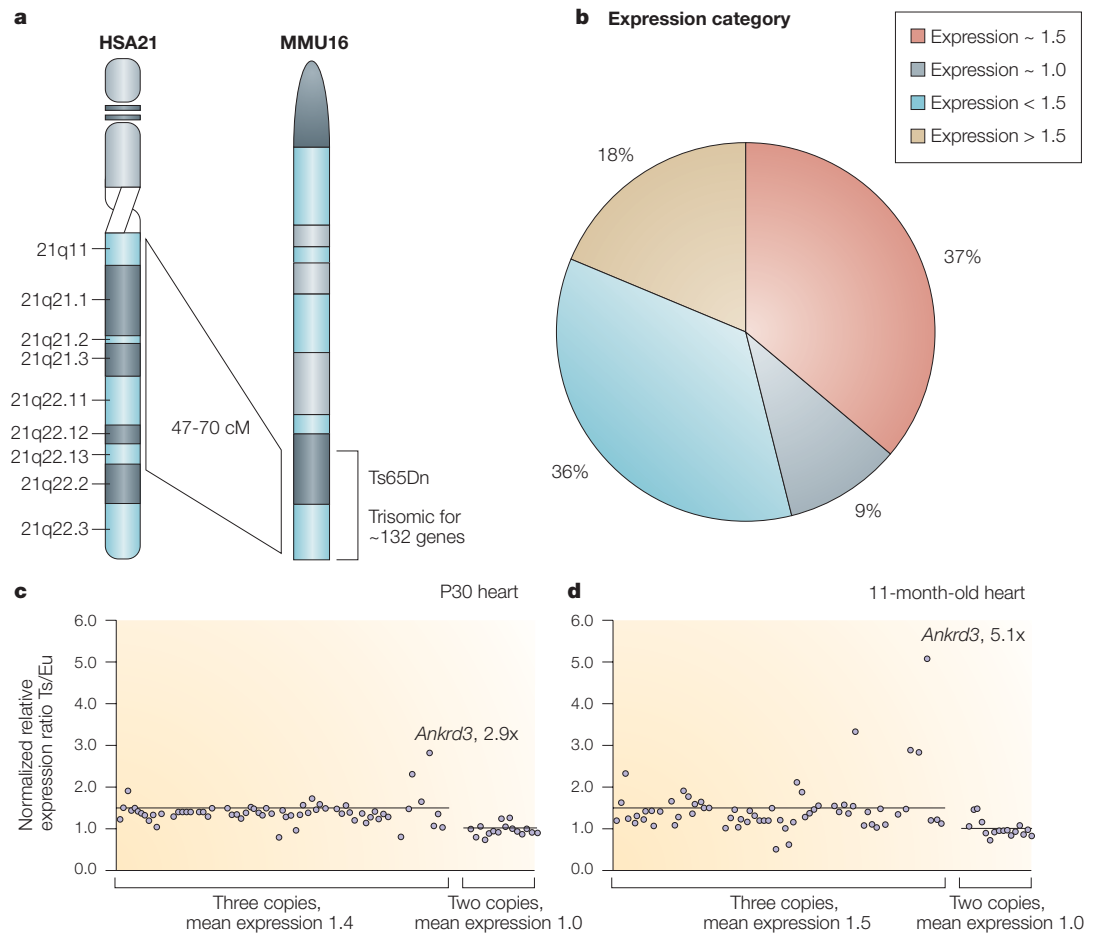


Figure 3 | Gene expression levels of trisomic genes in the Ts65Dn mouse model of Down syndrome. a | The Ts65Dn mouse is trisomic for a ~16 Mb region of mouse chromosome 16 containing ~132 genes. **b** | Summary of gene expression data of trisomic genes in Ts65Dn. Total data from 79 genes, 2 developmental stages and 6 tissues are shown. The percentages of genes with expression relative to the values 1.0 and 1.5 are also shown. **c** | Examples of expression data for the heart. The majority of genes present in 3 copies are expressed close to 1.5. A notable exception is the gene *Ankrd3*, which is highly overexpressed in both developmental stages P30, postnatal day 30; Ts, the trisomy state; Eu, the euploid state. Data are from REF. 60.

The best studied model is that of a partial (segmental) trisomy 16, named Ts65Dn (REF. 42). The trisomy in this mouse mutant extends for at least 23.3 Mb from *Mrpl39* to the *Znf295* genes⁴³ (predicted to contain ~132 genes that are homologous to HSA21). Detailed morphological and behavioural characterization of the Ts65Dn mouse model revealed several abnormal phenotypes that are similar to those seen in human trisomy 21 (REFS 44–51). Additional partial trisomies for MMU16 that extend from *Sod1* to *Znf295* (85 known genes) (Ts1Cje)⁵² and from *App* to *Sod1* (46 known genes) (Ms1Cje/Ts65Dn)⁵³ have been reported, and their phenotypic characterization is in progress. Neurological phenotypes in mouse models of DS are shown in TABLE 2. There is also a chimeric mouse model that has substantial numbers of cells that carry portions of HSA21 (REF. 54).

The contribution of triplicated individual genes to the mouse phenotype could be further evaluated by the sequential deletion of one copy of the gene from the partial trisomy mouse. This could be achieved by crossing

mice with a particular HSA21 gene knocked out with the partial trisomy mouse. One such example has been published recently for *Ifngr* and *Ifnar2* (REF. 55), which encode the interferon- δ and - α receptor 2, respectively. The resulting mouse with trisomy 16, but only 2 instead of 3 copies of these genes, showed significantly improved fetal growth and cortical neuron viability. A more direct approach involves creating transgenic mice that overexpress one gene that is orthologous to HSA21. Several such transgenic mice have been developed, following the first such mouse that carried *SOD1* transgenes⁵⁶ (TABLE 2). To achieve results that are biologically relevant to DS, the transgene (ideally the mouse gene) needs to be ‘physiologically’ regulated from its own regulatory elements and only one extra copy needs to be expressed. Ideally, a total transcript level (endogenous and from the transgene) should be ~1.5 times that of the normal expression level. One such example has been described for *Sim2*; the mice that carry three copies of *Sim2* are characterized by abnormal spatial exploration and social interactions, and reduced NOCICEPTION⁵⁷.

Multiplicity overexpression of a transgene from an exogenous promoter can also provide insights into gene function⁵⁸, although not necessarily into its role in trisomy. The crossing of these animal models into different genetic backgrounds might uncover contributions of modifier genes that affect certain DS phenotypes.

Gene expression in trisomy 21

Characterizing the spatial and temporal expression of HSA21 genes is an important step towards understanding how gene dosage imbalance affects DS phenotypes.

Since the discovery of trisomy 21 in 1959 (REF. 59), it has been hypothesized that the genes that are present in 3 copies are overexpressed 1.5-fold relative to the euploid state. This hypothesis has only recently been tested in the Ts65Dn partial trisomy mouse model of DS (REFS 43,60). In one study, a total of 78 genes present in 3 copies on MMU16 were tested by quantitative RT-PCR in 6 mouse tissues at 2 developmental stages (postnatal day 30 and 11 months old)⁶⁰. Surprisingly, only ~37% of genes are expressed at the theoretical value of 1.5-fold; ~45% of the genes are expressed at levels significantly lower than 1.5-fold, 9% are not significantly overexpressed and 18% had expression levels greater than 1.5-fold (FIG. 3). Similarly, a second study using cDNA arrays on nylon filters and quantitative RT-PCR to examine nine adult tissues found that nearly all triplicated genes had elevated transcript levels in most tissues where they were expressed, whereas a few showed downregulation, compensation or strong overexpression in a tissue-specific manner^{43,60}. These data not only provide candidate genes for contribution to DS phenotypes, but also highlight the complex regulation of gene expression related to genomic dosage imbalance.

The detailed and systematic gene expression profiles of HSA21 genes (and, for that matter, all human genes) should be one of the research priorities of the functional analysis of our genome. First attempts towards this goal have been recently published. A considerable number of mouse orthologues of HSA21 genes (~160) have been

studied by RNA *in situ* hybridization, by RT-PCR on adult and fetal mouse tissues, and by *in silico* mining of ESTs (REFS 61,62).

Initial attempts to study the global transcriptome differences between trisomy 21 and euploid state using microarrays or serial analysis of gene expression (SAGE) have also been published^{63–66} (TABLE 3). A meta-analysis of the existing data, and additional data collection, are needed to draw biologically meaningful conclusions and to develop further hypotheses about the aetiology of DS.

Finding candidate genes for DS phenotypes

Given the extent of sequencing data and the efforts to gather functional genomic data, how can we identify and prioritize candidate genes for their contribution to DS phenotypes? Several criteria could be used on the basis of present and future functional analysis of HSA21. The spatial and temporal pattern of expression is one criterion. Another is the potential functional classification of the predicted protein. Genes in which the expression varies widely between individuals might contribute to the variable phenotypes, whereas genes with no expression variation might underlie the phenotypes that are present in all individuals with DS. Abnormal phenotypes in transgenic mice with 1.5-fold overexpression would provide strong evidence for the role of such candidate genes.

The map position of a gene in a HSA21 interval that is associated with a given phenotype is another strong criterion. The study of rare patients with partial trisomy 21 defined the genomic regions that harbour genes associated with some DS phenotypes. For example, a region that is crucial for the heart defect (see [Supplementary information S2](#)) was identified in this way⁶⁷; and so searches for genes and genetic variation that contribute to this phenotype should focus on this region.

A number of investigators have described a 'DS critical region' that specifically contains genes that contribute to cognitive defects or other DS features. However, the definition of these regions has been

Table 3 | **Transcriptome studies of trisomy 21 (T21), or mouse model, versus euploid state**

Comparison*	Method	Results	References
Human amniocytes: T21 X3, normal X4; pooled RNAs	cDNA microarrays UniGene l Incyte; 8,466 human transcripts	187 upregulated (>2 SD); 85 downregulated (>2 SD)	63
Human astrocyte cell lines: T21 X4, normal X4; individual RNAs	Affymetrix microarrays U95Av2 or U133A; 18,462 human transcripts	679 differentially regulated (ANOVA $p < 0.05$)	64
Human frozen fetal cerebrum: T21 X4, normal X4; individual RNAs	Affymetrix microarrays U133A; 18,462 human transcripts	725 differentially regulated (ANOVA $p < 0.05$)	64
Mouse cerebellum: Ts65Dn X3, normal X3; individual and pooled RNAs	Affymetrix microarrays U74Av2; 12 488 mouse transcripts	29 with >2-fold consistent difference; 1,532 discriminating probes	65
Mouse brains: Ts65Dn X3, normal X7; pooled RNAs	152791 SAGE tags; 45,856 unique RNA tags	330 differentially expressed RNA tags ($p < 0.05$)	66

* X denotes the number of unrelated cases used in each experiment. SAGE, serial analysis of gene expression; SD, standard deviation.

NOICEPTION
The perception of pain.

SAGE
Serial analysis of gene expression: a method for comprehensive analysis of gene expression patterns using a short sequence tag (10–14bp) for each RNA molecule.

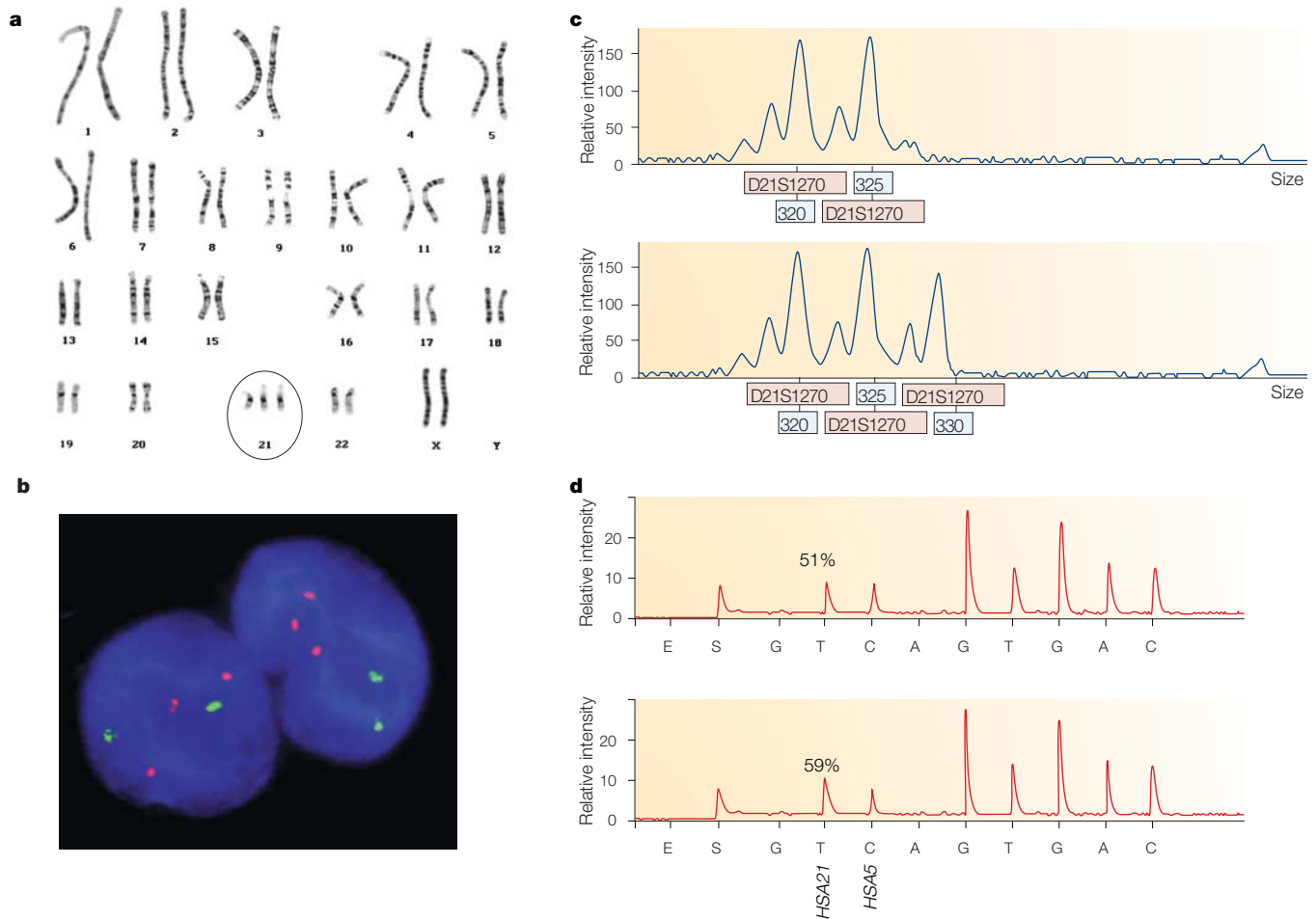


Figure 4 | **Trisomy 21 diagnostic methods: old and new.** **a** | G-banded karyotype of a trisomy 21 female, showing three copies of human chromosome 21 (HSA21). **b** | Fluorescent *in situ* hybridization (FISH) of interphase nuclei of a trisomy 21 fetus. In each cell there are two green spots (LSI 13 SpectrumGreen probe, Vysis) and three red spots (LSI 21 SpectrumOrange probe, Vysis) marking the 13q14 and 21q22.13-q22.2 chromosomal regions, respectively. **c** | Quantitative fluorescence PCR (QF-PCR) of marker D21S1270 of a trisomy 21 proband (bottom line) and his mother (top line). There are three different alleles in the patient DNA (alleles of sizes 320, 325 and 330 bp), but only two in the patient's mother (alleles 320 and 325 bp). **d** | Paralogous sequence quantification (PSQ). Two paralogous sequences, one mapping to HSA21 and the other to HSA5, which differ by a single nucleotide (in the example shown here, T on HSA21 and C on HSA5), are amplified and sequenced quantitatively (by PYROSEQUENCING) to determine the quantity of the variable nucleotide in a trisomy 21 proband (bottom line) and a parent (top line). Assuming similar amplification of both paralogous sequences, ratios of 0.5 for HSA21/HSA5, the two variable nucleotides in a normal control, and ratios of 1.5 in a trisomy 21 sample are expected. In this example, we observe 51% T and 49% C content in the parent, and 59% T and 41% C content in the trisomy 21 child.

controversial as there are patients with partial triplications outside this region who, nevertheless, manifest some features of DS (REFS 67–69). This issue should be re-examined now that the complete sequence of HSA21 and better diagnostic tools are available.

Finally, non-HSA21 transcripts with significant expression differences, as measured by microarray analysis in cells, tissues and organs with trisomy 21 versus normal organs, constitute a class of candidates that not only are likely to contribute to DS phenotypes, but could also be targets for potential therapeutic interventions.

Identifying and prioritizing CNGs that might contribute to DS phenotypes will pose a particular challenge. CNGs that are *cis*- or *trans*-regulators, or that harbour variation with functional consequences, are obvious candidates for dosage-related phenotypic abnormalities.

Diagnostic methods: old and new

Cytogenetic analysis of metaphase karyotypes remains the standard practice to identify not only trisomy 21, but also all other aneuploidies and balanced translocations. Over the past 10 years however, several other methods have been developed and used for the rapid detection of trisomy 21, either in fetal life or after birth (FIG. 4).

The most widely used is fluorescent *in situ* hybridization (FISH) of interphase nuclei, using HSA21-specific probes or whole-HSA21 PAINTING⁷⁰. An alternative method that is now widely used in some countries is quantitative fluorescence PCR (QF-PCR), in which DNA polymorphic markers (microsatellites) on HSA21 are used to determine the presence of three different alleles⁷⁰. This method relies on informative markers and the availability of parental DNA. Additional methods to measure copy number of DNA sequences include the multiple amplifiable probe hybridization (MAPH)⁷¹,

PYROSEQUENCING
A method for DNA sequencing, in which the inorganic pyrophosphate (PPi) that is released from a nucleoside triphosphate on DNA chain elongation is detected by a bioluminometric assay.

and multiplex probe ligation assay (MLPA)⁷². A recent method, termed paralogous sequence quantification (PSQ), uses paralogous sequences to quantify the HSA21 copy number⁷³. Finally, comparative genomic hybridization (CGH) on BAC chips can be used for the diagnosis of full trisomy or monosomy, and for partial (segmental) aneuploidies^{5,74}.

Examples of protein dosage imbalance

How does a supernumerary (third) copy of normal (wild-type) alleles result in an abnormal phenotype? This question has, for a long time, interested researchers in this area^{75–77}. Below, we provide some examples (mostly from model organisms) of potential molecular mechanisms that deal with this question.

Subunits of multimeric proteins. Many protein complexes consist of different subunits that are encoded by genes on different chromosomes. The stoichiometry of subunits is usually well controlled for the normal function of the complex. Disturbance of the normal stoichiometry of subunits, owing to a supernumerary copy of the gene that encodes one subunit, might result in protein complexes of abnormal composition and function. The nucleosome, for example, consists of a central core of eight histone proteins (two of each H2A, H2B, H3 and H4) with 146 bp of dsDNA coiled around it. The ratio of H2A and H2B over H3 and H4 is important for the fidelity of chromosomal replications and/or segregation in yeast⁷⁸.

Another example is that of nicotinic acetylcholine receptors (nAChRs), which are usually composed of 2 α - and 3 β -subunits. Experiments in *Xenopus laevis* oocytes and human embryonic kidney cells have shown that alteration of the stoichiometry of the subunits of nAChR results in alterations of functional properties of the receptors^{79,80}.

Transcription regulators. Three or more copies of genes that encode transcription regulators might result in abnormal expression of downstream target genes, which in turn could have phenotypic consequences. A characteristic example is provided by the mouse BAC transgenic experiments using the *clock* gene, which encodes a basic helix-loop-helix-PAS domain transcription factor, a component of the circadian pacemaker system. Normal mice have mean circadian periods of 23.48 hours, whereas transgenic mice harbouring just 1–2 supernumerary but normal copies of the *clock* transgene on a BAC had significantly shorter circadian periods with a mean of 22.89 hours⁸¹.

A further example is the protein product of the mouse *Bmi1* locus. *Bmi1* is a component of the chromatin-associated polycomb complex and is involved in maintaining the transcriptionally repressed state of genes; it modifies chromatin in a heritable way. Transgenic mice that overexpress *Bmi1* have dose-dependent skeletal anomalies, namely anterior transformation of vertebral identity⁸². In both examples, the phenotypic abnormalities are a consequence of the increased amount of a specific protein.

Cell surface receptors and ligands. Triplication of genes that encode receptors and ligands might also result in phenotypic differences. An illustrative example comes from *Notch* signalling in cell-fate determination in the epidermis of *D. melanogaster*. Cells that contained three copies of the wild-type *Notch*, which encodes a transmembrane receptor, always adopted an epidermal fate, whereas cells with two copies of *Notch* had a 50% chance of adopting neural or epidermal fate⁸³.

A good example of dosage effects of genes encoding ligands is that of erythropoietin (EPO), which is the primary protein that regulates mammalian erythropoiesis. Transgenic mice with an extra copy of EPO develop POLYCYTHAEMIA with HAEMATOCRIT of ~80% (normal ~45%)⁸⁴.

Transporter molecules. Transporters such as members of the transmembrane ABC transporter family could also have gene dosage effects. ABCA1 transporter affects intracellular cholesterol transport, and pathogenic mutations in this locus cause high-density lipoprotein deficiency type 1. BAC transgenic mice with an extra copy of ABCA1 have significantly increased cholesterol efflux in different tissues, and elevated high-density lipoprotein (HDL) cholesterol⁸⁵.

Cell adhesion molecules. Cell adhesion molecules might also cause dosage sensitive phenomena. For example, the rate of aggregation of synthetic vesicles carrying the neural cell adhesion molecule (NCAM) is proportional to the concentration of NCAM^{3,5}. Therefore, a 50% increase or decrease in NCAM concentration causes a 4-fold increase or 90% decrease in cellular adhesiveness, respectively⁸⁶.

Morphogens. Alterations in the concentration of morphogens can have considerable effects on development. One of the first examples to be described is *bicoid* (*bcd*) in *D. melanogaster*. Embryos derived from mothers that carry one, two, four or six copies of *bcd*-generated threshold concentrations of the homeodomain-containing BCD protein at progressively more posterior positions in these embryos⁸⁷. These threshold concentrations of a morphogen dictate distinct developmental outcomes as a function of distance from the source; in the case of BCD protein gradient, the outcome is the development of the anterior body pattern of the fly.

Regulatory elements. Duplication of regulatory elements could result in altered gene expression and specific pathological phenotypes. For example, tandem duplication (owing to a transgene insertion) of a regulatory element ~1 Mb upstream of the *Shh* locus in the *Ssq* mouse, causes ectopic *Shh* expression resulting in preaxial polydactyly¹⁰. Therefore, three copies of any functional element in the genome might be responsible for specific trisomy-related phenotypes.

Future directions of research

The dissection of the genomic infrastructure of HSA21, the appreciation of its genomic variability and the conservation of functional sequences in model organisms,

FISH

Fluorescent *in situ* hybridization is a method that uses fluorescent molecular DNA probes to visualize specific regions in chromosomes that hybridize to the probe.

CHROMOSOME PAINTING

Fluorescent *in situ* hybridization to chromosomes using a probe that represents a whole chromosome or a part of a chromosome.

POLYCYTHAEMIA

The increase of the red blood cell count, haemoglobin and the total red blood cell volume, accompanied by an increase in total blood volume.

HAEMATOCRIT

Percentage of red blood cells in blood.

provide new opportunities for understanding the function of the HSA21-encoded genes, and for explaining the molecular pathogenesis of different phenotypic manifestations of DS.

There are many important future research goals, which are listed below, but are by no means complete. Functional analysis of all HSA21 genes is a priority, particularly in the context of development (timing and cellular specificity of expression). Studying HSA21 orthologues in model organisms would facilitate such analysis. The results might lead to the identification of candidate genes for specific DS phenotypes and for gene dosage sensitivity.

Determining the variation of gene expression in different tissues is crucial. Genes with considerable allelic variation of expression in the population will probably be identified. These genes are probable candidates for allele-specific contribution to specific DS phenotypes. At the same time, genes without polymorphic variability of expression are probably those that contribute in an allele-independent manner to DS phenotypes. Furthermore, genes that show considerable overlap of the RNA expression output between normal and trisomy tissues or cells could be excluded as candidates. However, alterations in gene expression measured at the transcript level might not always accurately reflect alterations in protein levels⁸⁸. It is therefore necessary to either validate the differences at the protein level and/or study differences in the proteome.

Genomic variation in *cis* that is responsible for gene expression variation will need to be uncovered. This variation could then be used in association studies to detect genes that are responsible for certain DS phenotypes. Initial association studies have been published for the heart defects that are associated with DS (REF. 89).

Mouse models with trisomy of all HSA21 syntenic regions (that is, the MMU segments of chromosomes 16, 17 and 10) should be created and their phenotypes characterized in detail. Mice that are trisomic for single genes should also be generated as they provide valuable information on certain candidate genes. Another informative approach involves 'subtracting' the third allele of a gene from the trisomy mouse using appropriate crosses with heterozygous or homozygous mice that carry a deletion of that gene.

The function of the numerous CNGs on HSA21 will need to be determined. Identification of CNGs with potential gene regulatory function will reveal candidate genomic elements for dosage imbalance. Trisomy for some of them or their genomic variation might be related to certain DS phenotypes. CNGs that are not regulatory might also be involved in trisomy 21-linked phenotypes, owing to as yet unknown mechanisms.

Determination of the non-HSA21 (*trans*-acting) genes or non-coding sequences that predispose to DS phenotypes will also be important. Association, transmission disequilibrium and other tests could be used with genome-wide markers to define genomic regions that harbour variability, which contributes to the development of certain phenotypes. These genomic

regions might also contribute to similar phenotypes in non-trisomy 21 individuals.

The extent of partial trisomy 21 and correlation with the presence of certain phenotypic features of DS needs to be re-evaluated. A complete tiling path of BAC or oligonucleotide array CGH, or other methods, could be used to determine precisely the genomic trisomy 21, and therefore provide a comprehensive list of candidate genes (coding and non-coding) and conserved functional elements.

Relevant miRNAs and other non-coding RNAs and their targets need to be identified, and dosage sensitivity of the miRNAs and their potential involvement in certain DS phenotypes need to be evaluated.

Global gene expression differences between trisomy and euploidy states need to be evaluated in human cell lines and tissues, or in those of model organisms such as the mouse. Once defined, such differences could be used for diagnostic purposes, but they will be most important for the initial determination of biological processes that are dysfunctional in full or partial trisomy 21. The problem of variability of expression owing to polymorphisms could be avoided by studying tissues from discrepant identical twins (one twin with trisomy 21 and the other with a normal karyotype), or from clonal cell lines obtained from individuals with mosaic trisomy 21.

All functional genomic elements of HSA21 need to be identified. The exploration of the genomic sequences is far from complete. Multi-species comparative sequence analysis will provide initial evidence for additional protein-coding potential, ncRNAs and regulatory elements. Experimental analysis using whole chromosome approaches will hopefully provide convincing evidence for the diverse biological roles of a plethora of so far unknown functional regions.

Finally, the sequencing of the short arm of HSA21, or at least the short arm of an acrocentric chromosome awaits completion. A complete contig of these sequences might be difficult or impossible to obtain owing to the presence of multiple repeats in each short arm and the 'homogenization' of the short arms of the acrocentric chromosomes. It is likely, however, that important but as yet unknown functional elements (including genes) might be present on these short arms.

Therapy for trisomy 21?

Our understanding of the molecular pathogenesis of DS is remarkably poor; it is therefore unlikely that an effective therapeutic intervention will be found within the next decade. However, attempts towards treatment could be made in model organisms or in cultured cells. For example, reducing the total transcript levels of the dosage sensitive genes (perhaps by small interfering RNA (siRNA)) might alter cellular or whole organism phenotypes. Another option could involve pharmacological interference with the dysregulated metabolic pathways or key target molecules that are crucial for phenotypic characteristics.

Ultimately, the elucidation of the phenotypic consequences of gene dosage imbalance in DS might provide new opportunities for therapeutic interventions.

1. Hsu, L. Y. in *Genetic Disorders and the Fetus* (ed. Milunsky, A.) 179–248 (Johns Hopkins Univ. Press, Baltimore, 1998).
2. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
3. Pentao, L., Wise, C. A., Chinault, A. C., Patel, P. I. & Lupski, J. R. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nature Genet.* **2**, 292–300 (1992).
4. de Mollerat, X. J. *et al.* A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. *Hum. Mol. Genet.* **12**, 1959–1971 (2003).
5. Snijders, A. M. *et al.* Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genet.* **29**, 263–264 (2001).
6. Ellis, D. & Malcolm, S. Proteolipid protein gene dosage effect in Pelizaeus-Merzbacher disease. *Nature Genet.* **6**, 333–4 (1994).
7. Inoue, K. & Lupski, J. R. Molecular mechanisms for genomic disorders. *Annu. Rev. Genomics Hum. Genet.* **3**, 199–242 (2002).
8. Readhead, C., Schneider, A., Griffiths, I. & Nave, K. A. Premature arrest of myelin formation in transgenic mice with increased proteolipid protein gene dosage. *Neuron* **12**, 583–595 (1994).
9. Magyar, J. P. *et al.* Impaired differentiation of Schwann cells in transgenic mice with increased *PMP22* gene dosage. *J. Neurosci.* **16**, 5351–5360 (1996).
10. Lettice, L. A. *et al.* A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
11. Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nature Genet.* **34**, 369–376 (2003).
12. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Rev. Genet.* **2**, 280–291 (2001).
13. Antonarakis, S. E. 10 years of genomics, chromosome 21, and Down syndrome. *Genomics* **51**, 1–16 (1998).
14. Bandyopadhyay, R. *et al.* Parental origin and timing of *de novo* Robertsonian translocation formation. *Am. J. Hum. Genet.* **71**, 1456–1462 (2002).
15. Epstein, C. J. in *The Metabolic and Molecular Bases of Inherited Diseases* (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1223–1256 (McGraw-Hill, New York, 2001).
16. Roizen, N. J. & Patterson, D. Down's syndrome. *Lancet* **361**, 1281–1289 (2003).
17. Pennington, B. F., Moon, J., Edgin, J., Stedron, J. & Nadel, L. The neuropsychology of Down syndrome: evidence for hippocampal dysfunction. *Child Dev.* **74**, 75–93 (2003).
18. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
Landmark study that reports the high-quality nearly-complete nucleotide sequence of HSA21. Important for molecular studies on HSA21.
19. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
20. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
21. Raymond, A. *et al.* From PREDs and open reading frames to cDNA isolation: revisiting the human chromosome 21 transcription map. *Genomics* **78**, 46–54 (2001).
22. Raymond, A. *et al.* Nineteen additional unpredicted transcripts from human chromosome 21. *Genomics* **79**, 824–832 (2002).
23. Davison, M. *et al.* Evolutionary breakpoints on human chromosome 21. *Hum. Mol. Genet.* **10**, 99–106 (2001).
24. Gardiner, K. & Davison, M. The sequence of human chromosome 21 and implications for research into Down syndrome. *Genome Biol.* **1**, 1–9 (2000).
25. Gardiner, K., Fortna, A., Bechtel, L. & Davison, M. T. 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 (2003).
26. Pletcher, M. T., Wiltshire, T., Cabin, D. E., Villanueva, M. & Reeves, R. H. Use of comparative physical and sequence mapping to annotate mouse chromosome 16 and human chromosome 21. *Genomics* **74**, 45–54 (2001).
27. Rogers, M. A. *et al.* Hair keratin associated proteins: characterization of a second high sulfur KAP gene domain on human chromosome 21. *J. Invest. Dermatol.* **122**, 147–158 (2004).
28. Shibuya, K. *et al.* A cluster of 21 keratin-associated protein genes within introns of another gene on human chromosome 21q22.3. *Genomics* **83**, 679–693 (2004).
29. Mural, R. J. *et al.* A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* **296**, 1661–1671 (2002).
30. Watanabe, H. *et al.* DNA sequence and comparative analysis of chimpanzee chromosome 22. *Nature* **429**, 382–388 (2004).
31. Kapranov, P. *et al.* Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**, 916–919 (2002).
DNA chip technology used to study the transcriptional potential of the genome.
32. Kampa, D. *et al.* Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.* **14**, 331–342 (2004).
33. Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
DNA chip technology used to identify functional elements of the genome.
34. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
The sequence of the mouse genome provided the opportunity for comparative analysis to identify human functional elements.
35. Dermitzakis, E. T. *et al.* Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578–582 (2002).
36. Dermitzakis, E. T. *et al.* Comparison of human chromosome 21 conserved non-genic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. *Genome Res.* **14**, 852–859 (2004).
37. Dermitzakis, E. T. *et al.* Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302**, 1033–1035 (2003).
38. Kirkness, E. F. *et al.* The dog genome: survey sequencing and comparative analysis. *Science* **301**, 1898–1903 (2003).
39. Frazer, K. A. *et al.* Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.* **14**, 367–372 (2004).
40. Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
41. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
A whole chromosome linkage disequilibrium mapping that provides tools for association studies of human complex phenotypes.
42. Davison, M. T. *et al.* Segmental trisomy as a mouse model for Down syndrome. *Prog. Clin. Biol. Res.* **384**, 117–133 (1993).
43. Kahlem, P. *et al.* Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome. *Genome Res.* **14**, 1258–1267 (2004).
44. Reeves, R. H. *et al.* A mouse model for Down syndrome exhibits learning and behavioural deficits. *Nature Genet.* **11**, 177–184 (1995).
First phenotypic analysis of the most-used mouse model with partial trisomy 16.
45. Baxter, L. L., Moran, T. H., Richtsmeier, J. T., Troncoso, J. & Reeves, R. H. Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse. *Hum. Mol. Genet.* **9**, 195–202 (2000).
46. Costa, A. C., Walsh, K. & Davison, M. T. Motor dysfunction in a mouse model for Down syndrome. *Physiol. Behav.* **68**, 211–220 (1999).
47. Escorihuela, R. M. *et al.* Impaired short- and long-term memory in Ts65Dn mice, a model for Down syndrome. *Neurosci. Lett.* **247**, 171–174 (1998).
48. Dierssen, M. *et al.* Murine models for Down syndrome. *Physiol. Behav.* **73**, 859–871 (2001).
49. Richtsmeier, J. T., Baxter, L. L. & Reeves, R. H. Parallels of craniofacial maldevelopment in Down syndrome and Ts65Dn mice. *Dev. Dyn.* **217**, 137–145 (2000).
50. Delcroix, J. D. *et al.* Trafficking the NGF signal: implications for normal and degenerating neurons. *Prog. Brain Res.* **146**, 3–23 (2004).
51. Cooper, J. D. *et al.* Failed retrograde transport of NGF in a mouse model of Down's syndrome: reversal of cholinergic neurodegenerative phenotypes following NGF infusion. *Proc. Natl Acad. Sci. USA* **98**, 10439–10444 (2001).
52. Sago, H. *et al.* TsiCje, a partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioral abnormalities. *Proc. Natl Acad. Sci. USA* **95**, 6256–6261 (1998).
53. Sago, H. *et al.* Genetic dissection of region associated with behavioral abnormalities in mouse models for Down syndrome. *Pediatr. Res.* **48**, 606–613 (2000).
54. Shinohara, T. *et al.* Mice containing a human chromosome 21 model behavioral impairment and cardiac anomalies of Down's syndrome. *Hum. Mol. Genet.* **10**, 1163–1175 (2001).
55. Maroun, L. E., Heffernan, T. N. & Hallam, D. M. Partial IFN- α/β and IFN- γ receptor knockout trisomy 16 mouse fetuses show improved growth and cultured neuron viability. *J. Interferon Cytokine Res.* **20**, 197–203 (2000).
56. Epstein, C. J. *et al.* 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 (1987).
First description of single-gene transgenic mouse model to study the overexpression consequences of a HSA21 gene.
57. Chrast, R. *et al.* 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 (2000).
58. Sumarsono, S. H. *et al.* Down's syndrome-like skeletal abnormalities in *Ets2* transgenic mice. *Nature* **379**, 534–537 (1996).
59. Lejeune, J., Gautier, M. & Turpin, R. Études des chromosomes somatiques de neuf enfants mongoliens. *C. R. Acad. Sci.* **248**, 1721–1722 (1959).
A classic paper. The first description of trisomy 21.
60. Lyle, R., Gehrig, C., Neergaard-Henrichsen, C., Deutsch, S. & Antonarakis, S. E. Gene expression from the aneuploid chromosome in a trisomy mouse model of Down syndrome. *Genome Res.* **14**, 1268–1274 (2004).
61. Raymond, A. *et al.* Human chromosome 21 gene expression atlas in the mouse. *Nature* **420**, 582–586 (2002).
HSA21 gene expression atlas in the mouse.
62. Gitton, Y. *et al.* A gene expression map of human chromosome 21 orthologues in the mouse. *Nature* **420**, 586–590 (2002).
A gene expression map of HSA21 orthologues in the mouse.
63. FitzPatrick, D. R. *et al.* Transcriptome analysis of human autosomal trisomy. *Hum. Mol. Genet.* **11**, 3249–3256 (2002).
64. Mao, R., Zielke, C. L., Zielke, H. R. & Pevsner, J. Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain. *Genomics* **81**, 457–467 (2003).
65. Saran, N. G., Pletcher, M. T., Natale, J. E., Cheng, Y. & Reeves, R. H. Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. *Hum. Mol. Genet.* **12**, 2013–2019 (2003).
66. Chrast, R. *et al.* 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 (2000).
67. Barlow, G. M. *et al.* Down syndrome congenital heart disease: a narrowed region and a candidate gene. *Genet. Med.* **3**, 91–101 (2001).
68. Korenberg, J. R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl Acad. Sci. USA* **91**, 4997–5001 (1994).
69. Delabar, J. M. *et al.* Molecular mapping of twenty-four features of Down syndrome on chromosome 21. *Eur. J. Hum. Genet.* **1**, 114–124 (1993).
Systematic attempt to define chromosomal trisomic regions resulting in Down syndrome phenotypes.
70. Kuo, W. L. *et al.* Detection of aneuploidy involving chromosomes 13, 18, or 21, by fluorescence *in situ* hybridization (FISH) to interphase and metaphase amniocytes. *Am. J. Hum. Genet.* **49**, 112–119 (1991).
71. Armour, J. A., Sismari, C., Patsalis, P. C. & Cross, G. Measurement of locus copy number by hybridisation with amplifiable probes. *Nucleic Acids Res.* **28**, 605–609 (2000).
72. Slater, H. R. *et al.* Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J. Med. Genet.* **40**, 907–912 (2003).
73. Deutsch, S., Choudhury, U. & Antonarakis, S. E. Detection of trisomy 21 and other aneuploidies by paralogous gene quantification. *J. Med. Genet.* (in the press).
74. Ishkanian, A. S. *et al.* A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genet.* **36**, 299–303 (2004).
75. Epstein, C. J., Epstein, L. B., Weil, J. & Cox, D. R. Trisomy 21: mechanisms and models. *Ann. NY Acad. Sci.* **396**, 107–118 (1982).
76. Epstein, C. J. Mechanisms of the effects of aneuploidy in mammals. *Annu. Rev. Genet.* **22**, 51–75 (1988).
77. Epstein, C. J. *Consequences of Chromosome Imbalance: Principles, Mechanisms, and Models*, (Cambridge Univ. Press, New York, 1986).

78. Meeks-Wagner, D. & Hartwell, L. H. Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* **44**, 43–52 (1986).
79. Zwart, R. & Vijverberg, H. P. Four pharmacologically distinct subtypes of $\alpha 4\beta 2$ nicotinic acetylcholine receptor expressed in *Xenopus laevis* oocytes. *Mol. Pharmacol.* **54**, 1124–1131 (1998).
80. Nelson, M. E., Kuryatov, A., Choi, C. H., Zhou, Y. & Lindstrom, J. Alternate stoichiometries of $\alpha 4\beta 2$ nicotinic acetylcholine receptors. *Mol. Pharmacol.* **63**, 332–341 (2003).
81. Antoch, M. P. *et al.* Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. *Cell* **89**, 655–667 (1997).
82. Alkema, M. J., van der Lugt, N. M., Bobeldijk, R. C., Berns, A. & van Lohuizen, M. Transformation of axial skeleton due to overexpression of *bmi-1* in transgenic mice. *Nature* **374**, 724–727 (1995).
83. Heitzler, P. & Simpson, P. The choice of cell fate in the epidermis of *Drosophila*. *Cell* **64**, 1083–1092 (1991).
84. Semenza, G. L., Koury, S. T., Neifelt, M. K., Gearhart, J. D. & Antonarakis, S. E. Cell-type-specific and hypoxia-inducible expression of the human erythropoietin gene in transgenic mice. *Proc. Natl Acad. Sci. USA* **88**, 8725–8729 (1991).
85. Singaraja, R. R. *et al.* Human ABCA1 BAC transgenic mice show increased high density lipoprotein cholesterol and ApoA1-dependent efflux stimulated by an internal promoter containing liver X receptor response elements in intron 1. *J. Biol. Chem.* **276**, 33969–33979 (2001).
86. Hoffman, S. & Edelman, G. M. Kinetics of homophilic binding by embryonic and adult forms of the neural cell adhesion molecule. *Proc. Natl Acad. Sci. USA* **80**, 5762–5766 (1983).
87. Struhl, G., Struhl, K. & Macdonald, P. M. The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell* **57**, 1259–1273 (1989).
88. MacKay, V. L. *et al.* Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Mol. Cell. Proteomics* **3**, 478–489 (2004).
89. Baptista, M. J. *et al.* Heterotrissomy, a significant contributing factor to ventricular septal defect associated with Down syndrome? *Hum. Genet.* **107**, 476–482 (2000).
90. Ward, O. C., *John Langdon Down* (Royal Society of Medicine, London, 1998).
91. Wang, S. Y. *et al.* A high-resolution physical map of human chromosome 21p using yeast artificial chromosomes. *Genome Res.* **9**, 1059–1073 (1999).
92. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
93. Lynn, A. *et al.* Patterns of meiotic recombination on the long arm of human chromosome 21. *Genome Res.* **10**, 1319–1332 (2000).
94. Harrison, P. M. *et al.* Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res.* **12**, 272–280 (2002).
95. Worton, R. G. *et al.* Human ribosomal RNA genes: orientation of the tandem array and conservation of the 5' end. *Science* **239**, 64–68 (1988).
96. Yamada, Y. *et al.* A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* **14**, 247–266 (2004).
97. Antonarakis, S. E., Lyle, R., Deutsch, S. & Raymond, A. Chromosome 21: a small land of fascinating disorders with unknown pathophysiology. *Int. J. Dev. Biol.* **46**, 89–96 (2002).
98. Olson, L. *et al.* Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. *Dev. Dyn.* **230**, 581–589 (2004).

Acknowledgements

We thank H. Attar, E. Hafen, U. Schibler, K. Basler, A. Estreicher, M. Friedli, K. Casada, L. Curtis, D. Sutter, C. Borel, and C. Attanasio for their contributions in the preparation of the manuscript; S. Dahoun and D. Marelli for figures 4a,b, and F. Bena for figure 4c. We also thank R. Reeves and three anonymous reviewers for critical reading of the manuscript and numerous insightful suggestions; all members (past and present) of the Antonarakis laboratory for discussions, debates and experimental data; the Swiss National Science Foundation, NCCR 'Frontiers in Genetics', European Union/Swiss OFES, Lejeune and 'Childcare' Foundations for support. We also thank the patients and their families for the donation of their samples and the continuous inspiration for research.

Competing interests statement

The authors declare no competing financial interests.

 Online links

DATABASES

The following terms in this article are linked online to:

Entrez Gene: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene>
PLP1 | *PMP22* | *clock*
OMIM: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>
 Down syndrome | CMT1A | SHFM3 | Hirschsprung disease
Swiss-Prot: <http://www.expasy.ch/>

FURTHER INFORMATION

Interpro: <http://www.ebi.ac.uk/interpro/>
Gene Ontology Annotation: <http://www.ebi.ac.uk/GOA/>
Single Nucleotide Polymorphism: <http://www.ncbi.nlm.nih.gov/SNP/>

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (table)

Access to this links box is available online