Opinion
After 'completion': the changing face of human chromosomes 21 and 22
Todd Duane Taylor

Address: RIKEN Genomic Sciences Center, Human Genome Research Group, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. E-mail: taylor@gsc.riken.jp

Abstract

In the four years since the publication of the first two 'complete' human chromosome sequences the type of research being done on each has shifted subtly, reflecting the impact of genomic data on biological science in general. There is now considerably more gene-expression evidence to support predicted genes, and the annotation of functions for previously unknown genes, including those implicated in disease, is gradually improving.

More than four years have now passed since Dunham *et al.* [1] published 'The DNA sequence of human chromosome 22', in December 1999. This was the first 'essentially' complete human chromosome sequence to be finished. A few months later, in May 2000, Hattori *et al.* [2] published 'The DNA sequence of human chromosome 21'. At that time it seemed as though a rapid succession of completed chromosomes and their publications were to follow (perhaps in reverse numerical order, reflecting chromosomal size), but it wasn't until almost two years later, in December 2001, that the completion of chromosome 20 was announced [3]. Since then, a few more of the remaining chromosomes (successively 14, Y, 7, 6, 13, 19, 9 and 10) have been published, but we are still waiting on the rest, hopefully all of which will appear by the end of this year. With the announcement of the 'completion' of the entire human genome in April 2003, it's just a matter of time.

As the first two chromosome sequences have been complete for a relatively long time (in comparison to the rest of the chromosomes), now seems an appropriate time to take a look at how research on these chromosomes, and how genomic research in general, has been affected. How can we measure the impact of the completion and publication of the first two finished chromosomes? By counting the number of times each chromosome paper has been cited? By detecting an increase in the number of publications related to each chromosome? By noticing a shift in the types of research being carried out on each chromosome? By seeing an increase in the gene count, or a decrease in the number of unidentified disease genes? This article takes a brief look at these measures and more, concluding that the overall number of genes on chromosomes 21 and 22 has not changed much since the initial annotation of these chromosomes, but experimental verifications have increased the number of confirmed genes. Furthermore, the availability of the entire chromosome sequences seems to have facilitated the localization of some disease loci on chromosomes 21 and 22.

Numbers of citations

According to the ISI Web of Knowledge [4] (as of October 31 2003), among 3,001 articles and reviews (keywords 'human genome') written from 1999 to 2002, the first two chromosome completion papers were among the top ten most-cited. As expected, the two papers published in 2001 reporting the human genome draft sequence, by the International Human Genome Sequencing Consortium [5] and Celera Genomics [6], were the most-cited, with 2,666 and 2,058 citations, respectively. Following in third place was the chromosome 22 paper [1], and in seventh the chromosome 21 paper [2] (on which I am an author), with 558 and 405 citations, respectively. Only 23 other papers describing large-scale studies, in areas such as single-nucleotide polymorphisms

(SNPs), linkage disequilibrium, microarray analysis of gene expression, and transposable elements, were cited 100 times or more.

The types of articles that cited the first two chromosome publications covered a range of research areas, with the majority being comparative genomics, comparative mapping, gene discovery, haplotype analysis, genomic organization, and chromosome-wide gene expression analysis. Clearly, the availability of whole, 'completely' finished chromosomes made possible some of these new broad-scale types of research. For example, when doing comparative genomics to try and identify conserved regions that may contain regulatory elements, it is essential that both of the sequences that are being compared be as complete as possible, in order to minimize the false-negative rate. While the syntenic regions of these two chromosomes in other species are not necessarily finished to the same high quality, for example for mouse, rat and chicken, they are available at various levels of draft from whole-genome shotgun assemblies. Fortunately, in the case of human chromosome 21, the equivalent chromosome in chimpanzee, chromosome 22, is now available in high-quality finished form [7], and the same is being done for regions similar to human chromosome 22.

The number of chromosome-related publications

If we look at the number of publications in PubMed [8] using the search criteria 'human chromosome 21 OR human chromosome 22', the average number of articles per year for both chromosomes begins to level off in 1990 (106 for chromosome 21 and 83 for chromosome 22), several years before the sequence publications (Figure 1a). On the basis of this information, the publications of the first two chromosome sequences had no effect on the number of chromosome-related papers published per year. If the number of publications per chromosome is weighted by chromosome size (Figure 1b), chromosomes 21 and 22 (as well as chromosomes 17 and 19) appear to be very 'high impact' chromosomes. In the case of chromosome 21, this effect could be due to the special interest in Down syndrome (trisomy 21). If the number of publications per chromosome is weighted by the number of genes on the chromosome (Figure 1c), chromosome 21 appears to be very significant, followed closely by chromosomes 13, 18 and 22. This observation may be due to the relatively small size of these chromosomes and low numbers of genes in comparison with the other chromosomes.

It might have been expected that the number of chromosome-related papers would increase after the original publication of the first chromosome sequences, but instead we see a shift in the type of research that is being conducted. Whereas before their publication the research emphasis was on mapping and novel gene discovery, after their publication
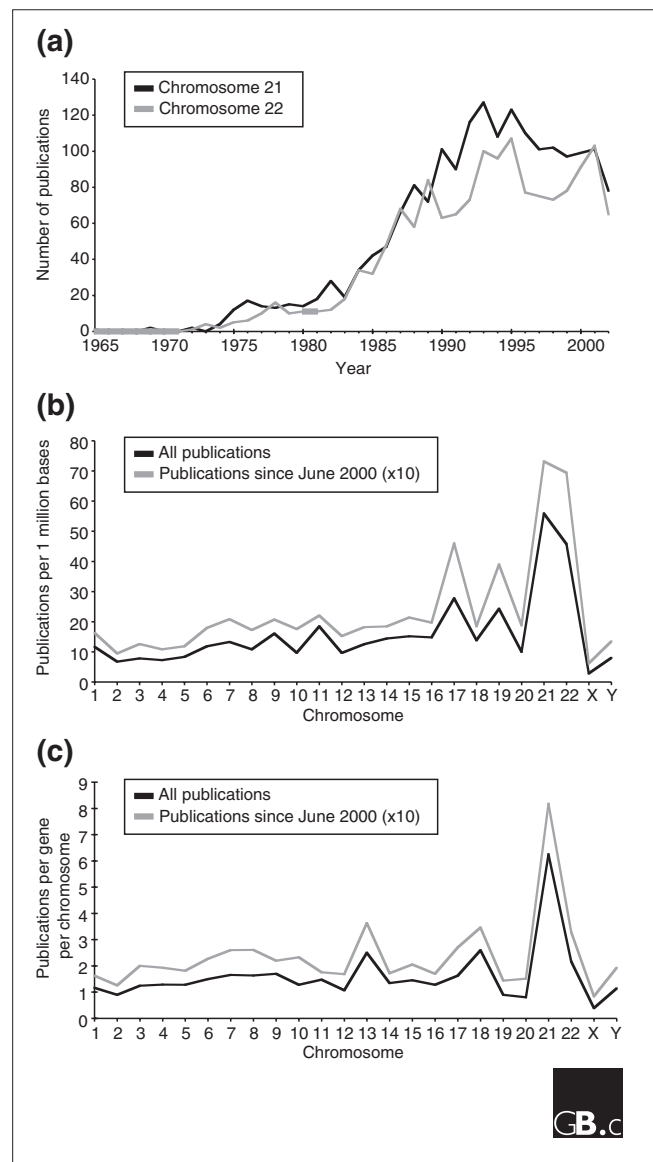

Figure 1
The publication history of chromosomes 21 and 22. **(a)** The number of articles in PubMed [8] that include 'chromosome 21' or 'chromosome 22' and 'human' in their title or abstract. **(b)** Chromosome publications weighted by sequence length. The number of publications since June 2000 has been multiplied by a factor of 10 to make the trend more visible. **(c)** Chromosome publications weighted by gene content per chromosome. The number of publications since June 2000 has been multiplied by a factor of 10 to make the trend more visible.

the emphasis turned to comparative analysis (for example, between mouse and human, as by Pletcher *et al.* [9]), haplotype analysis (for example, by Dawson *et al.* [10]) and whole-chromosome transcription analysis (for example, by Rinn *et al.* [11]). Hence, the availability of essentially complete, high-quality sequence is ushering in a whole new era of genomic research. Individual scientists generally no longer have to worry about the tedious tasks of mapping, sequencing and

gene identification, but can instead focus their efforts on finer details of their research, such as functional and regulatory analyses.

Other reasons for the leveling off in publication numbers could be that the number of researchers interested in these two chromosomes, and the amount of funding available for studying them, has not changed in recent years. And, because of the International Human Genome Sequencing Consortium's adherence to the 'Bermuda rules' [12], researchers around the world were able to access the sequence as it was being produced: they didn't have to wait until the chromosomes (or worse yet, the whole genome!) were published to utilize it. If this policy had not been implemented, we might have seen a spike in the number of chromosome-related publications upon publication and release of the sequence, assuming that researchers were eager to make use of it.

The number of genes

Another measure of the significance of the publications of the first full chromosome sequences might be the number of genes that have been identified since the original publications. When the sequences of chromosomes 21 and 22 were first published, it is safe to assume that the papers' authors did not believe that they had identified all of the genes on these chromosomes. They (we) knew that, upon release of the data, other scientists would identify more genes, and that new information would become available to help verify and append the initial annotations - and this is exactly what has taken place over the past four years. If we look at the number of genes (total non-pseudogenes) for each chromosome at the time of publication and compare it to the most recently available counts (Table 1), we can see that overall the gene

numbers have not risen that dramatically - an indication that the initial gene identification was done very well. In the case of chromosome 21 there is quite a jump in number of genes, but this is mainly due to the annotation of two keratin-associated protein gene clusters, one of which was only counted as a single gene in the original analysis.

We can also see that for both chromosomes the number of genes in the 'known' category has dramatically increased, while the number of 'novel' and 'putative' genes has generally decreased (Table 1). This re-categorization is due in part to the number of experimental verifications that have since been carried out on the predicted genes, and in part to the significant increase in number of full-length cDNAs and expressed sequence tags (ESTs) that have recently been deposited in the public databases. Many more human genes are now covered by at least one of these valuable mRNA resources than when the chromosomes were first annotated; four years ago mRNA data were much scarcer, and many gene models were based on partial EST evidence or solely on *in silico* gene-prediction analysis. At that time, for each chromosome only one representative model was annotated per gene; because of all the new mRNA data, however, roughly 30-40% of genes now have multiple transcripts annotated. And, also because of the new mRNA data, most annotators now agree that, in order to keep the number of false-positive gene models to a minimum, computer-only gene predictions should not become part of the annotation set until they are experimentally verified. Another noticeable change that can be seen in Table 1 is the near doubling in the number of pseudogenes for both chromosomes. This jump is due to several factors, including the increase in mRNA data, the completion of the rest of the human genome and subsequent improvement of annotation elsewhere within the genome, and the development of standards on how to define pseudogenes.

Table 1

Original and current gene catalogs for human chromosomes 21 and 22

Gene category	Chromosome 21		Chromosome 22	
	May 2000 [2]	February 2004 [16]	December 1999 [1]	January 2003 [13]
Known	127	241	247	393
Novel coding sequences	13	1	150	153
Novel transcripts	17	25	0	0
Putative	68	17	148	31
Total non-pseudogenes	225	284	545	577
Pseudogenes	59	98	134	234
Total	284	382	679	811
Immunoglobulin segments (all types)*				125
Grand total				936

*Chromosome 22 contains the human immunoglobulin lambda gene locus, which is a very complex region consisting of active and inactive (pseudogene) immunoglobulin gene segments.

Annotation of genomes is an evolving process that improves with time as additional experiments, tools, and resources become available. In the same way as Collins *et al.* [13] have now published the first follow-up to their initial annotation of chromosome 22, annotation of the other chromosomes is sure to improve over time (of course, for those chromosomes that are not yet published we should expect that the first-pass annotation is up to current standards). Researchers from the human genome community are working together to standardize gene and genome annotation. In March 2002, the first Human Annotation Workshop (HAWK) [14] was held at The Wellcome Trust Sanger Institute, bringing together scientists from most of the public sequencing centers, various databases, and the Human Genome Organization (HUGO) Nomenclature Committee (HGNC) [15]. The goals of the workshop were to establish communication between the groups involved in annotation, to standardize the way annotation is done across the human genome, and to exchange information, all with the aim of producing the highest standards of manual curation for the human genome. It should be noted that the HGNC has the daunting task of assigning unique identifiers, or gene symbols, to each gene in the human genome, thus reducing the amount of confusion often associated with multiple and non-unique gene names.

The number of disorders characterized

If we look at the number of human diseases and disorders (26 and 62, respectively) that have been mapped to chromosomes 21 and 22 (see Table 2, 3 and 4), we find that 3 (12%) and 12 (19%), respectively, were not mapped to the chromosomes until after January 2000. Thus, it appears that the availability of the entire chromosome sequences was necessary for locating some disease loci. Even now that all of these disorders have been mapped to their respective chromosomes, determining the exact location of the disease locus, the full-length cDNA product, and the mutation(s) that correlates phenotype and genotype remains a challenge.

In the case of chromosome 21, 6 (23%, including Down syndrome) out of 26 disorders do not have any conclusive mutation identified, and 4 disorders (15%) do not yet have any specific sequence location. And, for chromosome 22, an amazing 30 (48%) out of 62 disorders do not have any conclusive mutation identified, and 14 disorders (23%) do not yet have any specific sequence location; but several of the disease loci on chromosome 22 are involved in chromosomal rearrangement disorders, which are difficult to pinpoint, such as chronic myeloid leukemia. Two of the biggest barriers to identifying disease-gene locations and mutations are the lack of patient (and family) samples and complexity of the disease, particularly in multi-gene disorders such as Down syndrome or heterogeneous disorders such as schizophrenia and Alzheimer's disease. By having the full human genome sequence available, investigators need only to concentrate on matching disease phenotypes with genes from the current annotation, rather than having to identifying the genes themselves.

In Tables 3 and 4, all the currently mapped disorders for both chromosomes 21 and 22 are listed, along with information about how many related publications there have been for each disorder in total and since the publication of the respective chromosome sequence. While some genes, such as the amyloid β a4 precursor protein (*APP*; 132 publications in total and 39 since [2]) on chromosome 21, have been of research interest for a long time, other genes, such as *CHEK2* (Li-Fraumeni syndrome; 24 publications in total and 20 since [1]) on chromosome 22, have come into focus since the publication of the chromosome sequence. Some of the genes, such as *OGS2* (Opitz G syndrome, type II; 60 publications in total), have had no significant publications since the chromosome sequence was published, and other, such as *TMPRSS3* (childhood-onset neurosensory deafness; four publications in total), have only had papers published since the chromosome-sequence papers. Of course, some disorders, such as Down syndrome on chromosome 21 and DiGeorge syndrome on chromosome 22, continue to be the focus of much

Table 2

The history of disease-related loci currently mapped to chromosomes 21 and 22

	Chromosome 21 (26 disorders)			Chromosome 22 (62 disorders)		
	To December 1999	From January 2000	Pending	To December 1999	From January 2000	Pending
Mapped to chromosome	23 (88%)	3 (12%)	0 (0%)	50 (81%)	12 (19%)	0 (0%)
Mutation identified	15 (58%)	5 (19%)	6 (23%)	23 (37%)	9 (15%)	30 (48%)
First gene sequence published	22 (85%)	0 (0%)	4 (15%)	42 (68%)	6 (10%)	14 (23%)
Full-length cDNA published	20 (77%)	2 (8%)	4 (15%)	39 (63%)	7 (11%)	16 (26%)

There are currently 26 disorders mapped to chromosome 21, and 62 disorders mapped to chromosome 22. Numbers indicate the date by which the loci were first mapped to the chromosomes, when the first phenotype-related mutations were identified, when the first gene sequences were published, and when the first full-length cDNAs were published. Since exact date information was not available, a cutoff of January 2000 was used. Data were obtained from the NCBI resources OMIM [17], GenBank [18] and LocusLink [19].

Table 3

Disease-related loci currently mapped to chromosome 21

Gene symbol	Total publications	Publications since [2]	OMIM number	Disorder
APP	132	**39** *(30%)*	104760	Cerebroarterial, Dutch type amyloidosis; APP-related Alzheimer disease-1; chronic schizophrenia
CBS	132	**14** (11%)	236200	B6-responsive and nonresponsive comocystinuria; hyperhomocysteinemic thrombosis
SOD1	129	**27** *(21%)*	147450	Amytrophic lateral sclerosis due to SOD1 deficiency
ITGB2	68	5 (7%)	600065	Leukocyte adhesion deficiency
DCR	44	5 (11%)	190685	Down syndrome
HPE1	40	5 (13%)	236100	Holoprosencephaly-1
RUNX1	28	6 *(21%)*	151385	Acute myeloid leukemia; familial platelet disorder with associated myeloid malignancy
COL6A1	28	4 (14%)	120220	Bethlem myopathy
HLCS	28	2 (7%)	253270	Biotin-responsive and biotin-unresponsive multiple carboxylase deficiency
AIRE	23	**10** *(43%)*	607358	Autoimmune polyglandular disease, type I
PFKL	20	0 (0%)	171860	Hemolytic anemia due to phosphofructokinase deficiency
CRYAA	19	3 (16%)	123580	Cataract, zonular central nuclear, autosomal dominant; cataract, congenital progressive, autosomal recessive; cataract, autosomal dominant nuclear
KCNE1	18	2 (11%)	176261	Jervell and Lange-Nielsen syndrome; long QT syndrome-5
CSTB	17	2 (12%)	601145	Progressive myoclonic epilepsy 1
COL18A1	14	6 *(43%)*	120328	Knobloch syndrome
DYRK1A	10	2 (20%)	600855	Possible triplicate state responsible for mental defect in Down syndrome
COL6A2	9	4 *(44%)*	120240	Bethlem myopathy; Ullrich scleroatonic muscular dystrophy
IFNGR2	8	0 (0%)	147569	Atypical, familial disseminated mycobacterial infection
TAM	6	0 (0%)	159595	Transient leukemia of Down syndrome
PRSS7	5	1 (20%)	606635	Enterokinase deficiency
TMPRSS3	4	4 *(100%)*	605511	Autosomal recessive childhood onset deafness 8; congenital autosomal recessive deafness 10
SLC5A3	4	0 (0%)	600444	{Possible role in Down syndrome}
FTCD	3	1 *(33%)*	606806	Glutamate formiminotransferase deficiency
CLDN14	2	1 *(50%)*	605608	Autosomal recessive deafness 29
KCNE2	1	0 (0%)	603796	Long QT syndrome-6
USH1E	1	0 (0%)	602097	Usher syndrome, type 1E

The total number of relevant papers and the number of relevant papers since the chromosome 21 publication [2] are listed. A bold number indicates that there were ten or more post-chromosome locus-related publications; an italic percentage indicates that 25% or more of all locus-related publications appeared after the chromosome sequence was published. Curly brackets indicate examples of mutations that lead to universal susceptibility to a specific infection (diphtheria or polio), to frequent resistance to a specific infection (vivax malaria), protection from nicotine addiction, or other susceptibilities. Data were obtained from the NCBI resources OMIM [17] and PubMed [8].

research, even though the exact gene (or genes) that cause these syndromes has not yet been pinned down.

In summary, depending on how impact is measured, the publications of the first two finished human chromosomes may or may not appear to be significant, although one would have a hard time arguing against significance. From the analysis here, each chromosome has had various influences on the research that is being done on that particular chromosome and in other areas of biological research. Although the authors may not have done everything the 'right' way the first time around, they certainly set a standard for how other chromosomes should be finished, annotated and maintained over time. Given that both of the first-finished chromosomes are relatively small (each representing about 1.5% of the entire genome) they have subsequently become

Table 4

Disease-related loci currently mapped to chromosome 22

Gene symbol	Total publications	Publications since [1]	OMIM number	Disorder
DGCR	91	**18** (20%)	188400	DiGeorge syndrome; velocardiofacial syndrome
DIA1	80	5 (6%)	250800	Methemoglobinemia, type I; methemoglobinemia, type II
CYP2D	75	7 (9%)	124030	Debrisoquine sensitivity; {possible susceptibility to Parkinsonism}
BCR	71	**18** (25%)	151410	Chronic myeloid leukemia; acute lymphocytic leukemia
OGS2	60	0 (0%)	145410	Opitz G syndrome, type II
NF2	56	**22** (39%)	607379	Neurofibromatosis, type 2; NF2-related somatic meningioma; schwannomatosis
TCN2	53	3 (6%)	275350	Transcobalamin II deficiency
CTLO	51	9 (18%)	218040	Possibly Costello syndrome
ARSA	51	5 (10%)	607574	Metachromatic leukodystrophy
COMT	48	**13** (27%)	116790	{Possible susceptibility to schizophrenia}
CECR	48	1 (2%)	115470	Cat eye syndrome
PDGFB	39	2 (5%)	190040	SIS-related meningioma; dermatofibrosarcoma protuberans; giant-cell fibroblastoma
EWSR1	38	1 (3%)	133450	Ewing sarcoma; neuroepithelioma; extraskeletal myxoid chondrosarcoma
CHEK2	24	**20** (83%)	604373	Li-Fraumeni syndrome; somatic osteosarcoma; {possible susceptibility to breast cancer}; familial prostate cancer; {possible susceptibility to breast and colorectal cancer}
SOX10	24	**10** (42%)	602229	Waardenburg-Shah syndrome; Yemenite deaf-blind hypopigmentation syndrome; neurologic variant Waardenburg-Shah syndrome
NAGA	23	1 (4%)	104170	Schindler disease; Kanzaki disease; mild NAGA deficiency
ACF	21	1 (5%)	125520	Cayler cardiofacial syndrome
MYH9	20	**12** (60%)	160775	May-Hegglin anomaly; Fechtner syndrome; Sebastian syndrome; autosomal dominant deafness 17; Epstein syndrome
EP300	16	8 (50%)	602700	Colorectal cancer
TIMP3	16	6 (38%)	188826	Sorsby fundus dystrophy
SMARCB1	16	3 (19%)	601607	Rhabdoid tumors; rhabdoid familial predisposition syndrome
CTHM	16	2 (13%)	217095	? Conotruncal cardiac anomalies
SCZD4	15	2 (13%)	600850	{Possible susceptibility to schizophrenia}
ADSL	15	2 (13%)	608222	Adenylosuccinase deficiency
NEFH	15	0 (0%)	162230	{Possible susceptibility to amyotrophic lateral sclerosis}
SLC5A1	15	0 (0%)	182380	Glucose/galactose malabsorption
CRYBB2	12	2 (17%)	123620	Cataract, cerulean, type 2; cataract, sutural, with punctate and cerulean opacities
GGT1	12	0 (0%)	231950	Glutathioninuria
A4GALT	11	5 (45%)	607922	[Blood group, P system]
HMOX1	11	5 (45%)	141250	Heme oxygenase-1 deficiency
CSF2RB	11	1 (9%)	138981	Pulmonary alveolar proteinosis
ACR	11	0 (0%)	102480	{Possible male infertility due to acrosin deficiency}
MLC1	10	8 (80%)	605908	Megalencephalic leukoencephalopathy with subcortical cysts
SCA10	10	5 (50%)	603516	Spinocerebellar ataxia-10
FBLN1	10	2 (20%)	135820	Synpolydactyly, 3/3'4, associated with metacarpal and metatarsalsynostoses
GP1BB	9	1 (11%)	138720	Bernard-Soulier syndrome, type B; isolated giant platelet disorder
HCF2	9	0 (0%)	142360	Thrombophilia due to heparin cofactor II deficiency
RAC2	8	6 (75%)	602049	Neutrophil immunodeficiency syndrome

Table 4 *(continued)*

Gene symbol	Total publications	Publications since [1]	OMIM number	Disorder
PRODH	8	5 (63%)	606810	Hyperprolinemia, type I; {possible susceptibility to schizophrenia}
GGT2	8	0 (0%)	137181	[Familial high serum gamma-glutamyltransferase]
HIRA	8	0 (0%)	600237	{Possible role in CATCH22}
IGLL1	7	1 (14%)	146770	Autosomal agammaglobulinemia recessive
APOL1	6	5 (83%)	603743	{Possible susceptibility to schizophrenia}
XBP1	6	4 (67%)	194355	{Possible susceptibility to schizophrenia}
SCO2	6	4 (67%)	604272	Fatal infantile cardioencephalomyopathy due to cytochrome coxidase deficiency
ECGF1	6	2 (33%)	131222	Myoneurogastrointestinal encephalomyopathy syndrome
PVALB	6	1 (17%)	168890	{Possible role in DiGeorge syndrome}
APOL2	4	4 (100%)	607252	{Possible susceptibility to schizophrenia}
APOL4	4	4 (100%)	607254	{Possible susceptibility to schizophrenia}
MKL1	4	4 (100%)	606078	Acute megakaryoblastic leukemia
LARGE	4	3 (75%)	603590	{Possibly meningioma}
HPS4	4	3 (75%)	606682	Hermansky-Pudlak syndrome
SYN3	4	1 (25%)	602705	{Possible relation to schizophrenia}
KAZA	3	3 (100%)	608207	{Possible susceptibility to Kala-azar}
PSAP2	3	2 (67%)	606230	Chromosome 22q13.3 deletion syndrome
UFD1L	3	1 (33%)	601754	{Possible role in CATCH22}
MN1	3	0 (0%)	156100	Meningioma
ALG12	2	2 (100%)	607144	Congenital disorder of glycosylation, type Ig
FPEVF	2	0 (0%)	604364	Partial epilepsy with variable foci
ARVCF	2	0 (0%)	602269	{Possible role in VCFS}
DFNB40	1	1 (100%)	608264	Autosomal recessive deafness 40
ITS	1	1 (100%)	606960	Insulinoma

The total number of relevant papers and the number of relevant papers since the chromosome 22 publication [1] are listed. As in Table 3, a bold number indicates that there were ten or more post-chromosome locus-related publications; an italic percentage indicates that 25% or more of all locus-related publications appeared after the chromosome sequence was published. Curly brackets indicate examples of mutations that lead to universal susceptibility to a specific infection (diphtheria or polio), to frequent resistance to a specific infection (vivax malaria), protection from nicotine addiction, or other susceptibilities; certain 'nondiseases', mainly genetic variations that lead to apparently abnormal laboratory test values (such as dysalbuminemic euthyroidal hyperthyroxinemia), are included in square brackets. Data were obtained from the NCBI resources OMIM [17] and PubMed [8].

the test subjects of many other types of research, such as whole-chromosome gene-expression [11] and haplotype analysis [10]. Given the continuing number of publications that come out each year related to chromosomes 21 and 22, there should be no doubt that the availability of these complete sequences has had a lasting influence on many areas of research.

References

1. Dunham I, Shimizu N, Roe BA, Chissoe S, Hunt AR, Collins JE, Bruskiewich R, Beare DM, Clamp M, Smink LJ, *et al.*: **The DNA sequence of human chromosome 22.** *Nature* 1999, **402**:489-495.
2. Hattori M, Fujiyama A, Taylor TD, Watanabe H, Yada T, Park HS, Toyoda A, Ishii K, Totoki Y, Choi DK, *et al.*: **The DNA sequence of human chromosome 21.** *Nature* 2000, **405**:311-319.
3. Deloukas P, Matthews LH, Ashurst J, Burton J, Gilbert JG, Jones M, Stavrides G, Almeida JP, Babbage AK, Bagguley CL, *et al.*: **The DNA sequence and comparative analysis of human chromosome 20.** *Nature* 2001, **414**:865-871.
4. **ISI Web of Knowledge** [http://isiwebofknowledge.com/]
5. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
6. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al.*: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
7. Watanabe H, Fujiyama A, Hattori M, Taylor TD, Toyoda A, Kuroki Y, Noguchi H, BenKahla A, Lehrach H, Sudbrak R, *et al.*: **DNA sequence and comparative analysis of chimpanzee chromosome 22.** *Nature* 2004, **429**:382-388.
8. **NCBI Entrez PubMed** [http://www.ncbi.nih.gov/entrez/]
9. Pletcher MT, Wiltshire T, Cabin DE, Villanueva M, Reeves RH: **Use of comparative physical and sequence mapping to annotate mouse chromosome 16 and human chromosome 21.** *Genomics* 2001, **74**:45-54.

10. Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, *et al.*: **A first-generation linkage disequilibrium map of human chromosome 22.** *Nature* 2002, **418:**544-548.
11. Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, *et al.*: **The transcriptional activity of human chromosome 22.** *Genes Dev* 2003, **17:**529-540.
12. **Policies on release of human genomic sequence data: Bermuda-quality sequence** [http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml]
13. Collins JE, Goward ME, Cole CG, Smink LJ, Huckle EJ, Knowles S, Bye JM, Beare DM, Dunham I: **Reevaluating human gene annotation: a second-generation analysis of chromosome 22.** *Genome Res* 2003, **13:**27-36.
14. **Human annotation workshop** [http://www.sanger.ac.uk/Info/workshops/hawk1/]
15. **HUGO gene nomenclature committee** [http://www.gene.ucl.ac.uk/nomenclature/]
16. **Max-Planck-Institute for Molecular Genetics - human chromosome 21** [http://chr21.molgen.mpg.de/]
17. **OMIM - Online Mendelian Inheritance in Man** [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM]
18. **GenBank** [http://www.ncbi.nlm.nih.gov/Genbank/]
19. **LocusLink** [http://www.ncbi.nlm.nih.gov/LocusLink/]