A Comprehensive Human Linkage Map with Centimorgan Density

Cooperative Human Linkage Center (CHLC): Jeffrey C. Murray,* Kenneth H. Buetow, James L. Weber, Susan Ludwigsen, Titia Scherpbier-Heddema, Frank Manion, John Quillen, Val C. Sheffield, Sara Sunden, Geoffrey M. Duyk; Généthon: Jean Weissenbach, Gàbor Gyapay, Colette Dib, Jean Morrissette, G. Mark Lathrop, Alain Vignal; University of Utah: Ray White, Norisada Matsunami, Steven Gerken, Roberta Melis, Hans Albertsen, Rosemarie Plaetke, Shannon Odelberg; Yale University: David Ward; Centre d'Etude du Polymorphisme Humain (CEPH): Jean Dausset, Daniel Cohen, Howard Cann

In the last few years there have been rapid advances in developing genetic maps for humans, greatly enhancing our ability to localize and identify genes for inherited disorders. Through the collaborative efforts of three large groups generating microsatellite markers and the efforts of the 110 CEPH collaborators, a comprehensive human linkage map is presented here. It consists of 5840 loci, of which 970 are uniquely ordered, covering 4000 centimorgans on the sex-averaged map. Of these loci, 3617 are polymerase chain reaction–formatted short tandem repeat polymorphisms, and another 427 are genes. The map has markers at an average density of 0.7 centimorgan, providing a resource for ready transference to physical maps and achieving one of the first goals of the Human Genome Project—a comprehensive, high-density genetic map.

For the first time, humans have been presented with the capability of understanding their own genetic makeup and how it contributes to morbidity of the individual and the species. Rapid scientific advances have made this possible, and developments in molecular biology, genetics, and computing, coupled with a cooperative and interactive biomedical community, have accelerated the progress of investigation into human inherited disorders. A primary engine driving these advances has been the development and use of human genetic maps that allow the rapid positional assignment of an inherited disorder as a starting point for gene identification and characterization.

Linkage approaches to human disorders began with the successful identification of X chromosome linkage for color blindness and hemophilia by Bell and Haldane in the 1930s (*1*). A key obstacle to the routine performance of linkage analysis in humans, however, was the small number of progeny and outbred nature of human matings. In the 1950s, Morton developed the lod score (logarithm of the odds ratio for linkage) method, which overcame these difficulties and made the analysis of human linkage data practical (*2*). This statistical approach was successfully applied to the limited number of polymorphic blood group and serum protein markers to establish the first human linkage groups. In the early 1970s, advances in statistical algorithms by Elston and colleagues (*3*, *4*) were implemented by Ott in efficient computer-based analysis tools (*5*). Although these developments made human linkage analysis computationally powerful, it was still limited by marker availability.

Few linkages were identified until the application of DNA-based polymorphisms allowed for a sudden expansion in the number of available markers. Botstein and co-workers (6) proposed restriction fragment length polymorphisms (RFLPs) as the solution to the marker problem in humans less than 15 years ago, and these DNA-based variations, although first discovered in yeast, were rapidly identified in humans as an offshoot of work on hemoglobinopathies.

The 1980s saw the expansion of computer tools (*7*, *8*) to evaluate linkage for multiple markers and dramatic improvements followed in the numbers and types of DNA-based markers. The first markers often had low heterozygosity, which limited their use in families, but the popularization of variable number of tandem repeat (VNTR) minisatellites by Jeffreys, White, and Nakamura (*9*, *10*) and microsatellites of dinucleotide repeats by the laboratories of Weber and Weiringa provided markers of high information content (*11*, *12*).

Genome-wide human linkage maps first appeared from the Donis-Keller (*13*) and White laboratories in the late 1980s and were greatly aided by the availability of a set of 61 families of the CEPH (Centre d'Etude du Polymorphisme Humain). The CEPH contribution allowed investigators around the world to pool data from markers developed in individual laboratories but studied on a common set of families (*14*). These maps and markers have now been used as the initial steps in a number of positional cloning efforts for analysis of human diseases that began with the Huntington disease linkage to an RFLP marker in 1983 (*15*).

Currently, comprehensive genome-wide human linkage maps and markers are being developed by three large groups and the continued contributions of many individual investigators. The Généthon group is working on maps based on markers that contain the CA repeat motif, which is the most common short tandem repeat in the human genome (*16*, *17*). The Cooperative Human Linkage Center (CHLC) in the United States is constructing maps by means of tri- and tetranucleotide repeats which, although less frequent in number, are easier to genotype than CA repeats. The Utah group is developing maps based on di-, tri-, and tetranucleotide repeats. Chromosome-specific research groups associated with the National Institutes of Health (NIH)/CEPH consortium (*18*) and EUROGEM continue to contribute to genome-wide efforts, and recent syntheses have been published by Matise *et al.* (*19*) and the CHLC (*20*).

In parallel with the advances in genetic mapping have been similar advances in developing the physical map of humans. This is greatly facilitated by the use of polymerase chain reaction (PCR)–based markers, which are fundamental to the sequence-

J. C. Murray is in the Departments of Pediatrics and Biology and V. C. Sheffield and S. Sunden are in the Department of Pediatrics, University of Iowa, Iowa City, IA 52245, USA. K. H. Buetow, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, and J. Quillen are at the Fox Chase Cancer Center, Philadelphia, PA 19111–2412, USA. J. L. Weber is at the Marshfield Medical Research Foundation, Marshfield, WI 54449, USA. G. M. Duyk is in the Department of Genetics, Harvard Medical School, Boston, MA 02115–5716, USA. J. Weissenbach, G. Gyapay, C. Dib, and A. Vignal are at Généthon, Evry 91000, France. J. Morrissette is at the Centre Hospitalier de l'Université Laval, Quebec, Canada G1K 7P4. G. Mark Lathrop is at INSERM, Paris 75010, France. R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, R. Plaetke, and S. Odelberg are in the Eccles Institute for Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. D. Ward is in the Department of Genetics, Yale University, New Haven, CT 06510, USA. J. Dausset, D. Cohen, and H. Cann are with CEPH, Paris 75010, France.

*To whom correspondence should be addressed at the Department of Pediatrics and Biology, University of Iowa, 2613 JCP, Iowa City, IA 52242, USA.

tagged site (STS) approach to mapping. In STS mapping, specific primer pairs provide unique reference points on genetic and physical maps. Direct transfer between laboratories of reagents necessary for mapping is thus possible as information files (containing primer sequences) rather than as biological materials (*21*).

In this paper, and on the accompanying chart, we describe genetic maps that combine genotypic data generated over the last decade. The large number of PCR-based markers drawn from all classes of previous mapping efforts provides both a resource to tie maps to disorders and markers of historical interest as well as a mechanism to move quickly from a genetic assignment to a physical map or syntenic map of a model organism. These maps represent the culmination of work by hundreds of investigators worldwide.

Chart Construction

The genetic maps were based on genotypes generated from DNA samples obtained from the CEPH reference pedigree set. Because CEPH family cell lines and DNA are publicly available (Coriell Cell Repository, Camden, New Jersey), individual groups can add genotyping on their own markers in the future. Two collections of genotypic data were used. The first set (Integrated set, version 3.0; available from CHLC) consisted of genotype information compiled from published and unpublished sources. The primary sources of these data included Généthon, CHLC, the University of Utah, and the NIH/CEPH gene mapping consortium (*18*). These data sources are well-characterized collections of genotypes that have been stringently examined for data accuracy. Data quality was studied by searching for double recombinants in short genetic intervals, by detecting evidence for genetic map inflation when individual loci were sequentially removed from the maps, and by detecting heterogeneity in pairwise recombination. This data set was composed of 5150 markers.

The second set of genotypes used for mapping was that made publicly available through the CEPH linkage mapping collaboration. This set consisted of genetic markers from the consortium's version 7.0 database (release date May 1994) that were listed as public in the current release or that were public as a consequence of their presence in the previous release (version 6.0) of the database. This database consisted of 6625 markers. Because of overlaps in the data sets (Integrated version 3.0 and CEPH version 7.0), a total of 5840 loci were mapped. Marker content is presented by chromosome in Table 1. Additional descriptive data on markers are available electronically (see below), both through the Genome Data Base (GDB) and through CHLC.

Only markers genotyped on CEPH reference families are included in these maps. Loci with only physical assignments or genetically mapped only on non-CEPH families were not used. Nonetheless, this list is extensive. The CEPH database contains blood group markers and protein polymorphisms, as well as a variety of DNA-based markers including RFLPs and short tandem repeat polymorphisms (STRPs) (*14*). A recent emphasis on PCR-based markers is evident. These markers have been identified by a variety of methods including tandem repeat polymorphisms, single-strand conformational polymorphisms, denaturing gradient gel electrophoresis–based polymorphisms, and allele-specific oligonucleotides.

The maps presented here were constructed from the STRP loci of the Integrated version 3.0 data set by means of a 15-family subset of the CEPH reference panel. (Families used were 102, 884, 1331, 1332, 1333, 1341, 1344, 1346, 1347, 1349, 1362, 1408, 1413, 1416, and 1423). They were constructed by first establishing "meiotic bins" on each chromosome. The meiotic bins were defined by the loci within the version 2.0 CHLC skeletal maps. The skeletal maps had been constructed from an earlier version of the Integrated data set (version 2.0) that consisted of genotype data from the NIH/CEPH consortium, Généthon, CHLC, and other sources. The bins, on average, were 6.8 cM in size. For each meiotic event within the reference panel, the most likely parental chromosomal assignment of each allele (phase) of the genotypes was determined. On the basis of this phase assignment, the bin or bins in which each recombination event was likely to have occurred were identified. Finally, the recombinant haplotypes were identified for the subset of individuals in which recombination events were observed. Parental chromosomal phase assignments and likelihoods calculations were performed by means of the computer program package, CRIMAP.

We constructed the map presented here by first identifying a meiotic bin assignment for each locus in the version 3.0 Integrated data set. Only loci that could be uniquely assigned to a given bin were included. The criterion for assignment was that all other bins must be excluded with odds of 1000:1 or greater (≥lod 3). In the process of this binning, the parental chromosome of origin for each allele for each genotype was determined. The order of the loci in each interval was determined by permutation of all possible orders of the loci uniquely assigned to each bin. The order selected was the one that resulted in the minimum number of recombination events within the previously identified subset of individuals showing a recombination event for that interval (see above). A single recombination event between loci within the bin was used as the minimum criterion for determining order. This process was performed in two steps. First, only data with a high probability of phase determination were considered ($P >$ 0.95). Next, the phase of lower certainty was considered. In some instances, clusters of loci were not separated by recombination events. In these instances, the largest set of loci that could be uniquely ordered were selected. Map distances (male, female, and sex-averaged) for the complete collection of ordered loci over all bins were determined by means of maximum likelihood methods as implemented in CRIMAP (Table 2). The sex-averaged recombination maps are presented on the chart. The other map figures are available electronically. The absolute lengths of the genetic maps presented on the chart were scaled to the cytogenetic size of each chromosome. This was done to allow comparisons of genetic map density and marker distribution across chromosomes.

The maps presented here vary in length from those previously published based on integrated data sets. In some circumstances,

Table 1. Marker description for chart. Classification of markers used in map construction by chromosome assignment and type. A marker that is both a gene and an STRP is listed under gene only. Other markers are mostly RFLPs of anonymous DNA segments.

Chromosome	Marker*		
	Other	STRP	Gene
1	97	291	39
2	65	290	21
3	154	219	14
4	73	270	48
5	193	244	18
6	46	195	28
7	138	193	20
8	53	170	14
9	57	122	11
10	107	189	12
11	92	135	34
12	39	150	21
13	64	107	12
14	45	113	9
15	37	102	12
16	88	93	6
17	182	105	28
18	22	112	3
19	35	99	24
20	18	145	5
21	21	53	10
22	59	65	13
X	101	149	24
Y	10	6†	1
Totals	1796	3617	427

*The number of markers per chromosome shown here will be ≥ the numbers in the histogram on the chart because the boundaries were not included in the bins. †This includes one STRP that was included in the count for the X, and five that were mapped by somatic cell hybridization.

the chromosome's genetic length is substantially shorter than that previously described. For example, the sex-averaged length of chromosome 1 in this map is 356 cM, whereas it was 384 cM in a map based on mixed data (*20*). These shorter maps are due in part to error correction in the primary data, which reduced map inflation. Map inflation can occur as a result of errors arising from incorrect genotyping of markers or data entry errors. These mistakes will be incorrectly perceived as recombination events and thereby increase apparent genetic distance between markers. More significantly, the maps are commonly shorter due to differences in marker content. The decision to include only STRP loci for primary map construction results in the exclusion of centromeric and telomeric loci that are currently only marked by RFLPs.

Genetic locations within the meiotic bins defined by the skeletal map were also determined for the markers publicly available through the CEPH database. These assignments were made by selection of the most likely bin among all the bins on the chromosome to which the marker had been assigned. The likely locations for markers in both the Integrated and CEPH data sets are available electronically.

A histogram was compiled that displays the marker content of each chromosome in the map intervals. Bins defined by the skeletal map were combined to achieve a series of larger reference intervals with a target size, where possible within the skeletal map, of 10 cM. To represent the regions that extend beyond end points of the current map, terminal bins of 10 cM in size were created. Marker content of each combined bin (reference interval) was determined. Markers that could be identified as present in both the Integrated and CEPH sets were only counted once, as were multiple copies of the same marker within a data set. Markers were considered to be the same if they had the same probe name or the same gene name as specified by the Human Gene Mapping Nomenclature Committee as that identified through GDB. On the basis of information available for each marker, it was categorized as being a gene, an STRP, or other polymorphism [for example, RFLP or single-strand conformational polymorphism (SSCP)]. In a hierarchical manner, "gene" was given the highest priority in categorization. More specifically, if variation at a gene was determined on the basis of an STRP, the locus was counted as a gene. Similarly, an anonymous DNA segment was counted as an STRP if one or more of its characterizations within the data were based on an STRP. Because the actual intervals used to describe the reference points vary in size, the width of each bar was scaled (divided) by the map distance represented by the interval. This width of the bar, therefore, represents the average number of markers per centimorgan for that interval. Because each map has been scaled by the physical size of each chromosome, the width when compared between chromosomes gives an indication of the density of genetically mapped markers per physical unit. The X and Y maps presented special cases: The Y map shown is a composite of a genetic map of the pseudoautosomal region pairing Xp22.3 and Yp13, incorporating two loci. The remaining 14 markers assigned to the Y map were placed through the use of somatic cell hybrids and represented those that showed Y-specific inheritance and interindividual variation on the CEPH panel.

Fluorescence in situ hybridization (FISH) was used to localize several hundred yeast artificial chromosome (YAC) clones containing Généthon-based STRP markers to specific chromosomal bands. Total yeast DNA–containing YAC clones corresponding to polymorphic markers that were spaced an average of 10 cM apart on each chromosome were labeled with biotin by nick translation. The clones were cohybridized with a digoxigenin-labeled Alu sequence oligonucleotide (GM009) that generates an R-banded karyotype (*22*). A minimum of 10 metaphase spreads were analyzed for each probe to determine band assignment and to measure fractional chromosomal length relative to the terminus on the p arm (pter) (*23*). A subset of these FISH-mapped, STRP-positive YAC clones

Table 2. Marker description. Chromosome-specific data on the 970 loci used in primary map construction.

Chromosome	Number of loci	Sex-avg. map		Female map		Male map		Dinucleotide count	Other STRP count	Maximum gap (cM)
		Length (cM)	Density	Length (cM)	Density	Length (cM)	Density			
1	86	356.0	4.2	423.4	5.0	290.8	3.4	71	11	20.3
2	85	294.1	3.5	366.4	4.4	227.7	2.7	66	19	11.8
3	66	272.2	4.2	343.6	5.3	211.8	3.3	54	11	12.7
4	56	251.4	4.6	329.9	6.0	182.6	3.3	42	12	17.9
5	60	250.1	4.2	294.0	5.0	211.5	3.6	49	11	13.0
6	52	222.7	4.4	304.1	6.0	140.2	2.8	41	10	17.2
7	58	217.4	3.9	275.4	4.8	159.2	2.7	43	14	10.8
8	44	183.9	4.3	251.1	5.8	123.7	2.9	36	7	15.9
9	33	169.6	5.3	213.2	6.7	131.7	4.1	27	5	11.9
10	55	183.2	3.4	227.2	4.2	144.3	2.7	48	7	12.3
11	43	155.4	3.7	191.9	4.6	121.4	2.9	31	9	17.5
12	38	221.6	6.0	274.7	7.4	178.5	4.8	23	13	25.3
13	34	147.9	4.5	160.1	4.9	141.7	4.3	30	4	19.4
14	38	151.7	4.1	197.2	5.3	104.6	2.8	31	7	15.3
15	27	141.9	5.5	159.4	6.1	127.6	4.9	22	5	14.1
16	22	142.9	6.8	189.6	9.0	114.2	5.4	19	2	17.3
17	24	126.7	5.5	159.6	6.9	95.0	4.1	19	5	16.7
18	25	149.6	6.2	187.6	7.8	113.9	4.8	19	6	14.6
19	29	103.3	3.7	117.1	4.2	89.9	3.2	24	5	13.8
20	22	120.2	5.7	143.7	6.8	96.5	4.6	19	3	23.8
21	19	64.6	3.7	80.5	4.5	50.3	2.8	14	5	7.5
22	22	73.6	3.5	85.1	4.1	63.5	3.0	18	4	14.6
X,* Y	32			247.3	8.2			29	3	
Tot.	970	4000.0		5222.1		3120.6		775	178	
Avg.			4.2		5.5		3.3			15.6

*The X chromosome includes the region mapping to the pseudoautosomal region at Yp13.

is presented in the wall chart. Information concerning the map location of the remaining clones can be obtained electronically by accessing the CEPH YAC physical mapping database (Table 3).

Electronic Access

Public access databases are an especially important feature of the Human Genome Project. They are easy to use and facilitate rapid communication of new findings (well in advance of hard-copy publications) and can be updated efficiently. The enormous number of loci and amount of genotypic data available make it difficult to provide, in a single source, a comprehensive enumeration of data necessary for map construction. To accommodate the investigator's individual needs and interests, we have made primary data as well as a variety of different maps available electronically. In addition, a much broader description of individual markers is available through other on-line sources (Table 3). Information available in GDB, for example, includes data on identified human genes and anonymous markers where mapped; there is an extensive reference list, as well as descriptive materials on markers including heterozygosity, polymorphism type, map location, and data quality.

Additional electronic access to a variety of specific types of maps constructed in this project is available through CHLC, as shown in Table 3. These maps include scaffold maps (maps composed of high-heterozygosity, easy to use, PCR-based STRPs at interval spacings suitable for genome-wide linkage searches) and framework maps (maps that comprise all loci placed with odds of 1000:1 or greater). Additionally, the best map position and other positions from which loci could not be excluded are given for those loci without unique (1000:1) locations in the above maps. Primary data on markers developed specifically through the resources of CHLC are also described in more detail therein. Other resources of maps and markers for human and model systems are shown in Table 3 as well.

Map Integration

A particular strength of current human genetic linkage maps is that their construction has been based upon STSs (*21*), which allows for the immediate integration of linkage maps with the developing physical maps. A first-generation physical map of the human genome has been published (*24*) and uses as a backbone the Généthon linkage map. Physical reagents, consisting of YAC, P1, cosmid, or other clones, can be identified on a genome-wide basis through several commercial and genome center resources around the world, as well as through individual chromosome-organizing committees with knowledge of the availability of physical and genetic reagents on specified chromosomes. The STS approach also allows integration with data from subchromosomal mapping panels by means of somatic cell hybrids or from radiation hybrid mapping.

The maps shown here are aligned with standard karyotypes and FISH analysis. Integration of the genetic and cytogenetic maps is important for several reasons. First, these karyotypic reference points allow for molecular mapping of new chromosome rearrangements such as translocations or deletions to genetic markers. The karyotypic abnormalities, if associated with a clinical disorder, can also be powerful tools in positional cloning strategies, as was shown initially for Duchenne muscular dystrophy (*25*) and has been reviewed for other disorders (*26*). Second, inspection of the relationships between the fractional genetic map of a chromosome and the band into which a polymorphic marker maps can identify chromosomal regions that are either deficient in, or are hotspots for, recombination. For example, many of the human chromosomes exhibit a much higher than anticipated frequency of recombination near their telomeric regions; thus, some STRP-positive YAC clones that are 10 to 20% (25 to 40 cM) from the end of the genetic map are physically located in a terminal band that contains only a few percent of the chromosomal DNA. In contrast, near the centromeres, YAC clones separated by as few as 5 cM on the genetic map may be physically separated by 15 to 20% of the chromosomal length, indicating that little recombination is occurring is such regions. The acrocentric chromosomes also show anomolous relationships between the genetic and cytogenetic maps. Thus, *D13S292* (4 cM from pter) and *D15S122* (0 cM from pter) map to the q arms (q12 and q11.2-12, respectively) of chromosomes 13 and 15.

An additional level of integration is provided by the availability of very dense mouse linkage maps (*27*) with parallel mapping of mouse phenotypes. Mouse maps and their syntenic human components are available through the Jackson Laboratory's mouse genome database and their accompanying maps (Table 3). Many human genes now have comparative map localizations associated with mouse phenotypes, and this allows for rapid searches for animal homologs of human disorders that can greatly facilitate biological studies and candidate gene approaches to gene identification. A striking example of this is the association of the human Waardenburg syndrome and the mouse mutant splotch with the gene *Pax*-3 (*28*). In addition, maps of other organisms (including mammals) of both commercial and scientific interest are also being developed.

Genetic Maps and Genome-Wide Searches

The past decade has demonstrated the utility of increasingly dense and user-friendly genetic maps as one tool in the positional cloning strategy (*26*). Paralleling the increasing availability and accessibility of genetic markers has been an increase in the number of genetic disorders localized through a variety of genome-wide linkage searches as well as candidate gene–based strategies. Genetic maps provide the initial resource necessary to begin a genome-wide search for the locus of a Mendelian disorder. These maps also greatly assist loss-of-heterozygosity studies, searches for isodisomy or, potentially, studies mapping loci

Table 3. Information access for human genetic maps and allied resources. Major informatic sources for data and related information. FTP, file transfer protocol; WWW, World Wide Web.

Source	Type	Access
GDB	General human map and loci data	E-mail: help@gdb.org FTP: ftp.gdb.org WWW: http://gdbwww.gdb.org/
Généthon	Maps of linked CA repeat markers for humans	FTP: ftp.genethon.fr WWW: http://www.genethon.fr/
CHLC	Genotypes, marker data, linkage servers	E-mail: info-server@chlc.org FTP: ftp.chlc.org WWW: http://www.chlc.org/
Jackson Lab	Variety of data types for mouse mapping	E-mail: mgi-help@informatics.jax.org FTP: ftp.informatics.jax.org WWW: http://www.jax.org/
CEPH	YAC physical mapping data	E-mail: ceph-genethon-map@cephb.fr FTP: ceph-genethon-map.genethon.fr WWW: http://www.cephb.fr/bio/ceph-genethon-map.html
Whitehead Institute	Maps of linked CA repeats for mouse	E-mail: genome_database@genome.wi.mit.edu FTP: genome.wi.mit.edu WWW: http://www-genome.wi.mit.edu/

for more complex or quantitative traits. Maps presented here provide not only a resource of available markers but also graphically demonstrate the high density of PCR-based polymorphic markers, RFLPs, and genes within given chromosomal regions.

The 3617 STRP class markers alone provide about one marker every 1×10^6 base pairs (bp) throughout the genome. Once an initial localization is identified by means of a set of genome-wide maps, subsequent steps can be undertaken to increasingly narrow the region to be searched and eventually to select a specific interval on which studies of physical reagents, such as cloned DNA fragments, can be carried out. The density of the maps reported here is such that, on average, genetic localizations can be quickly used to identify critical recombinant events or small regions of allele loss that are within the range at which one can make an immediate transition to physical reagents (*29*).

Once the physical reagents are in hand, gene and mutation searches are possible with a number of strategies. The search for a gene can involve: (i) clone characterization for the presence of CpG islands or conserved sequences (*30*), (ii) studies of tissue-specific expression of sequences contained within the cloned sequences (*28*), (iii) direct searches for coding sequences by hybridization methods (*31*) or exon trapping (*32*), or (iv) direct large-scale sequencing and gene recognition–based searches of genomic sequences (*33*). Mutation search strategies include direct sequencing as well as chemical cleavage (*34*), denaturing gradient gels (*35*), and single-strand conformational variants (*36*). These strategies have already proven effective in several gene and mutation identification projects, and their effectiveness is likely to become more powerful and more readily available over time.

Despite the increasing availability of markers and strategies, considerable problems and difficulties still remain. First, markers mutate at measurable but variable frequencies both in the germ line and in the tissue culture lines maintained to provide a source of DNA (*37*). These aberrant genotypes may resemble cases of non-Mendelian inheritance, but in cases where non-Mendelian inheritance is suspected as an underlying disease mechanism (for example, isodisomy or loss of heterozygosity), careful distinctions need to be made. Next, the genetic maps themselves are not entirely complete, with gaps still present in some areas. Anchor points at the telomeres and centromeres are unavailable for many chromosomes. Genotyping errors continue to complicate genetic map construction as well, particularly efforts as comprehensive as this. CEPH-based consortium maps of chromosomes 1 (*38*), 2 (*39*), 9 (*40*), 10 (*41*), 13 (*42*), and 15 (*43*) have been published and use data that have undergone rigorous individual error checking. For the maps presented here, only a subset of markers from the consortium maps or genotyped in the primary contributing labs (Généthon, CHLC, Utah) underwent such explicit vetting. Thus, genotype errors resulting from lab error, sample mix-ups, or clerical errors are likely present in some of the binned markers and will, in some cases, result in incorrect assignments (*44*). Care needs to be applied in using these markers without additional checking.

Next, correlations between genetic and physical distance have not been explicitly defined throughout the genome. Variation in recombination rates per unit of physical distance may differ by a factor of 100 and seriously distort the perceived distance a given point might lie from a locus of interest (*45*, *46*). A far better understanding of this phenomenon is needed before quick translations to the use of physical resources can be made with high confidence. Physical reagents are not yet readily available for all stretches of the genome and obtaining them may require significant additional effort, including new library construction in some regions. Finally, the mere identification of the physical reagents that genetic data suggest have a high probability of containing a gene and mutation of interest will not necessarily lead to that gene's immediate identification. Novel genetic mechanisms, perhaps even more arcane than the trinucleotide repeats (*47*) and digenic inheritance (*48*) recently described, may be at work and may still require new insights and direct biological correlation before specific mutations can be identified. Genetic maps of the X present special challenges because they can only be generated in female meioses (where two X chromosomes can pair and recombine). The Y chromosome may be even more challenging because only two small portions of its length (the pseudoautosomal regions) pair with homologous regions in the X and generate meiotic recombination-based data. The accompanying wall chart contains a genetic map of X and the pseudoautosomal region based on male meioses; some markers have been placed by binning of somatic cell hybrids and some by linkage across the pseudoautosomal region. Such a map is tentative but provides a listing of markers useful in both studies of pseudoautosomal inheritance and for use in studying cross-generation, male-specific inheritance.

These maps also provide an opportunity to exploit physical maps and candidate genes in ways not possible with previous maps. Candidate genes may be mapped onto physical reagents like YACs or radiation hybrids. Through the use of low-heterozygosity, gene-based polymorphisms, a region that is very likely to contain an adjacent high-heterozygosity STRP and can serve as a surrogate for genetic studies of that candidate gene can be found. This is especially useful when studies are limited by small families or sample numbers, and a high-heterozygosity marker may provide the only opportunity to include or exclude a candidate gene. In a reciprocal fashion, mapping a disorder with a high-heterozygosity marker allows one to search its adjacent genetic and physical regions for the presence of a likely candidate gene and to screen newly identified candidates.

Finally, the high density of the map, as currently presented, affords a new scale of opportunities for those interested in the biology of human meiosis. Questions concerning the nature of sex-specific differences in recombination or of interference (regions where recombination is not additive) become increasingly practical to study as the density of the map increases along with the availability of markers. These problems can now be revisited in more powerful ways. The high density of markers has already made these studies feasible in some regions of the chromosome.

The application of these markers not only to Mendelian but also to complex, multifactorial or polygenic disorders presents a challenge that may now be met with this new set of maps at an unprecedented level of marker quality and density [see Lander and Schork (*49*)]. It is our hope that the maps identified here will provide resources not only for the identification of genes underlying human inherited disorders but also contribute to our ability to carry out studies of the biological basis of human inheritance patterns.

Table 4. Progress in human genetic maps. Progress in the density of available markers in human genetic maps is shown. Keats had partial maps for nine chromosomes; all others include the autosomes and X chromosome. Mixed refers to the use of classical, RFLP, and PCR-based markers in significant numbers. CRI and Généthon maps used data generated almost exclusively in their own laboratories. Others used data from multiple groups, mostly CEPH (except Keats).

Year	Group	Marker types	No. of loci	Resolution (cM)
1981	Keats (*63*)	Classical	53	16
1987	CRI (*13*)	RFLP	403	10
1992	Généthon (*16*)	CA	814	4.4
1992	NIH/CEPH (*18*)	Mixed	1416	3
1994	CHLC (*20*)	Mixed	1123	4.9
1994	Pittsburgh (*19*)	Mixed	655	6.2
1994	Généthon (*17*)	CA	2066	2.9
1994	This report	Mixed	5840	0.7*

*4.2 cM in the reference marker map.

Completion of the Genetic Map?

When will genetic maps be completed for humans? The NIH had set a goal of a 2- to 5-cM map by 1995 (*50*), and that goal has been achieved. However, a complete genetic map of humans might consist of a catalog of all polymorphic variations present; estimates of human heterozygosity (*51*, *52*) would suggest that polymorphic variants might occur every 100 to 300 bp on average, requiring a map density of markers every 0.0001 cM. The maps described here are below 1 cM in density (0.7 cM), and the progress to achieve this goal is shown in Table 4. Before resolution can increase by several orders of magnitude, major advances in variant identification must occur.

Genetic maps might be seen as complete when their usefulness is maximized for the currently available family, population, and technological resources. To this end, an increased density of markers beyond that presently available is desirable. For example, at the current average density, it becomes feasible to consider genome-wide linkage disequilibrium approaches to mapping, and a 1-cM interval is the distance at which disequilibrium may be observed between markers. The success of this approach with cystic fibrosis (*53*) and other common Mendelian disorders (*54*) suggests that it is useful for common as well as rare disorders (*55*) where founder effects have long been recognized. This disequilibrium may involve STRPs (*56*, *57*), so that despite their higher mutation rates they are a useful class of markers for association studies (*58*). Using disequilibrium in this manner would require testing several thousand markers on hundreds of individuals. This level of genotyping is available to only a few large laboratories, but advances in automation may change this. It also may become feasible to use pooled sample approaches and quantitation combined with linkage disequilibrium to identify alleles associated with a phenotype (*59*, *60*). Advances in obtaining large numbers of DNA samples efficiently through the use of cheek swabs (*61*) or newborn screening cards (*62*) make studying large numbers of samples also practical.

Addition of new genetic markers to maps at high resolution will become more complex because of an increased need for more data and the effect of errors. An investigator may fill a gap in the genetic map by making direct use of cloned reagents from the region of the gap to identify genetic markers. Sequencing strategies that identify markers of lower heterozygosity but higher density may be especially suited to this interface (*60*). However, until physical methods (for example, contigs and radiation hybrids) prove capable of high- and intermediate-level ordering, genetic maps will provide the backbone for loci placement. The future of genetic mapping will likely continue to include a requirement for new markers at higher density that will be used in novel ways to map common traits and quantitative loci. Very recent successes in mapping susceptibility genes for diabetes (*64*) and identifying the gene locus for one form of breast cancer (*65*) confirm the utility of dense linkage maps. Eventually, a detailed map of human genetic diversity will be developed down to the single–base pair level allowing, for the first time, a complete description not just of a composite human but of all humans.

Finally, it should be emphasized that the new opportunities and challenges for biological research provided by these dense maps also create an urgency to face the parallel challenges in the areas of ethics, law, and social policy. Our ability to distinguish individuals for forensic purposes, identify genetic predispositions for rare and common inherited disorders, and to characterize, if present, the underlying nature of the genetic components of normal trait variability such as for height, intelligence, sexual preference, or personality type, has never been greater. Although technically feasible, whether these maps should be used for these ends should be resolved after open dialogue to review those implications and devise policy to deal with the as yet unpredictable outcomes.

REFERENCES AND NOTES

1. J. Bell and F. R. S. Haldane, *Proc. R. Soc. London Ser. B* **123**, 119 (1937).
2. N. E. Morton, *Am. J. Hum. Genet.* **7**, 277 (1955).
3. R. C. Elston and J. Stewart, *Hum. Hered.* **21**, 523 (1971).
4. J. K. Haseman and R. C. Elston, *Behav. Genet.* **2**, 3 (1972).
5. J. Ott, *J. Hum. Genet.* **26**, 588 (1974).
6. D. Botstein, R. L. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
7. G. M. Lathrop, J.-M. Lalouel, C. Julier, J. Ott, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3443 (1984).
8. E. Lander and P. Green, *ibid.* **84**, 2363 (1987).
9. A. J. Jeffreys, V. Wilson, S. L. Thein, *Nature* **314**, 67 (1985).
10. Y. Nakamura *et al.*, *Science* **235**, 1616 (1987).
11. J. L. Weber and P. E. May, *Am. J. Hum. Genet.* **44**, 388 (1989).
12. J. Smeets *et al.*, *Hum. Genet.* **83**, 245 (1989).
13. H. Donis-Keller *et al.*, *Cell* **51**, 319 (1987).
14. J. Dausset *et al.*, *Genomics* **6**, 575 (1990).
15. J. F. Gusella *et al.*, *Nature* **306**, 234 (1983).
16. J. Weissenbach *et al.*, *ibid.* **359**, 794 (1992).
17. G. Gyapay *et al.*, *Nat. Genet.* **7**, 246 (1994).
18. NIH-CEPH Collaborative Mapping Group, *Science* **258**, 67 (1992).
19. T. C. Matise, M. Perlin, A. Chakravarti, *Nat. Genet.* **6**, 384 (1994).
20. K. H. Buetow *et al.*, *ibid.*, p. 391.
21. M. V. Olson, L. Hood, C. Cantor, D. Botstein, *Science* **245**, 1434 (1989).
22. A. G. Matera and D. C. Ward, *Hum. Mol. Genet.* **1**, 535 (1992).
23. P. Lichter *et al.*, *Science* **247**, 64 (1991).
24. D. Cohen, I. Chumakov, J. Weissenbach, *Nature* **366**, 698 (1993).
25. A. Monaco *et al.*, *ibid.* **316**, 842 (1985).
26. F. S. Collins, *Nat. Genet.* **1**, 3 (1992).
27. W. F. Dietrich *et al.*, *ibid.* **7**, 220 (1994).
28. G. Chalepakis, M. Goulding, A. Read, T. Strachan, P. Gruss, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3685 (1994).
29. M. Boehnke, *Am. J. Hum. Genet.* **55**, 379 (1994).
30. F. Antequera and A. Bird, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11995 (1993).
31. M. Lovett, J. Kere, L. M. Hinton, *ibid.* **88**, 9628 (1991).
32. G. M. Duyk, S. Kim, R. M. Myers, D. R. Cox, *ibid.* **87**, 8995 (1990).
33. E. C. Uberbacher and R. J. Mural, *ibid.* **88**, 11261 (1991).
34. R. G. H. Cotton, N. R. Rodrigues, R. D. Campbell, *ibid.* **85**, 4397 (1988).
35. R. M. Myers, S. G. Fischer, L. S. Lerman, T. Maniatis, *Nucleic Acids Res.* **13**, 3131 (1985).
36. M. Orita, H. Iwahana, H. Kanazawa, K. Kayashi, T. Sekiya, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2766 (1989).
37. I. Banchs *et al.*, *Hum. Mutat.* **3**, 365 (1994).
38. N. C. Dracopoli *et al.*, *Genomics* **9**, 686 (1991).
39. N. K. Spurr *et al.*, *ibid.* **14**, 1055 (1992).
40. J. Attwood *et al.*, *ibid.* **19**, 203 (1994).
41. R. L. White *et al.*, *ibid.* **6**, 393 (1990).
42. A. M. Bowcock *et al.*, *ibid.* **16**, 486 (1993).
43. A. M. Bowcock *et al.*, *ibid.* **14**, 833 (1992).
44. K. H. Buetow, *Am. J. Hum. Genet.* **49**, 985 (1991).
45. A. Chakravarti *et al.*, *ibid.* **36**, 1239 (1984).
46. L. B. Jorde *et al.*, *ibid.* **54**, 884 (1994).
47. J. D. Brook, *Nat. Genet.* **3**, 279 (1993).
48. K. Kajiwara, E. L. Berson, T. P. Dryja, *Science* **264**, 1604 (1994).
49. E. S. Lander and N. J. Schork, *Science* **265**, 2037 (1994).
50. F. Collins and D. Galas, *ibid.* **262**, 43 (1993).
51. D. N. Cooper *et al.*, *Hum. Genet.* **69**, 201 (1985).
52. J. C. Murray *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3486 (1984).
53. B. S. Kerem *et al.*, *Science* **245**, 1073 (1989).
54. J. Theilmann *et al.*, *J. Med. Genet.* **26**, 676 (1989).
55. J. Hastbacka *et al.*, *Nat. Genet.* **2**, 204 (1992).
56. C. B. Kunst and S. T. Warren, *Cell* **77**, 853 (1994).
57. G. Imbert, C. Kretz, K. Johnson, J.-L. Mandel, *Nat. Genet.* **4**, 72 (1993).
58. R. Sherrington *et al.*, *Am. J. Hum. Genet.* **49**, 966 (1991).
59. N. Arnheim, C. Strange, H. Erlich, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6970 (1985).
60. P. Kwok, C. Carlson, T. D. Yager, W. Ankener, D. A. Nickelson, *Genomics*, in press.
61. B. Richards *et al.*, *Hum. Mol. Genet.* **2**, 159 (1993).
62. C. Carducci, L. Elliul, I. Antonozzi, A. Pontecorvi, *BioTechniques* **13**, 735 (1992).
63. B. J. B. Keats, *Linkage and Chromosome Mapping in Man* (Univ. of Hawaii Press, Honolulu, 1981).
64. J. L. Davies *et al.*, *Nature* **371**, 130 (1994); L. Hashimoto *et al.*, *ibid.*, p. 160.
65. R. Wooster *et al.*, *Science* **265**, 2088 (1994).

66. We thank the CEPH collaborators for contributions of genotypes to the CEPH database. We are also grateful to the many families who contributed biological material to the CEPH. Contributors who assisted greatly in this map development include K. Wiles, D. Even, C. Franzen, C. Weichmann, M. Wise, N. Newkirk, A. McClain, T. Businga, G. Mattes, T. Rocklina, B. Morrisson, and J. Beck (University of Iowa); J. Gastier (Harvard University); D. David, S. Salzman, and M. Stephenson (Marshfield Medical Research Foundation); L. Ballard, E. Lawrence, M. Moore, X. Zhao, P. Holik, G. Staker, M. Robertson, P. Bradley, A. Tingey, T. Elsner, P. Cartwright, R. Sargent, B. Milner, D. Adamson, J. Knell, A. Marks, and D. Fuhrman (University of Utah); J. Menninger, J. Lieman, A. Banks, T. Desai, N. Macadi, and P. Bray-Ward (Yale University); and J. Gastier, N. Yandava, T. Brody, J. C. Pulido, and J. Ghazizadeh (Harvard Medical School).