

# The Map Problem: A Comparison of Genetic and Sequence-Based Physical Maps

Andrew T. DeWan,<sup>1</sup> Antonio R. Parrado,<sup>1</sup> Tara C. Matise,<sup>2</sup> and Suzanne M. Leal<sup>1</sup>

<sup>1</sup>The Laboratory of Statistical Genetics, The Rockefeller University, New York, and <sup>2</sup>Department of Genetics, Rutgers University, Piscataway, NJ

The genetic order of autosomal genome-scan markers from Marshfield panels 9 and 10 were compared with their physical order, on the basis of the assembled nonredundant human genome sequence from the Human Genome Project–Santa Cruz (HGP-sc; October 2000 and April 2001 releases) and Celera (CEL; February 2001 release) databases. The genetic order of 96% of the markers on the Marshfield map for panel 10 is supported by a likelihood ratio of  $\geq 3$  (odds ratio of 1,000:1). Inconsistencies with the genetic panel 10 map were found for 5% and 2% of the markers in the CEL and HGP-sc sequences, respectively. These inconsistencies consisted of both positional and chromosomal-assignment disagreements. For the majority of these inconsistent markers, the genetic order was supported by a likelihood ratio of  $\geq 3$ , and the physical order in the other assembly matched the genetic order. The majority of the inconsistencies between the physical- and genetic-map order point to errors in the physical-map order. A Web site is made available that displays inconsistencies for genetic markers from Marshfield panels 9 and 10 between their genetic-map positions and sequence-based physical-map positions, as well as inconsistencies between their sequence-based physical position. This Web site also contains genetic-map distances, physical-map positions from the Celera and Human Genome Project sequence, and likelihood-ratio support for the genetic maps.

## Introduction

The genetic- and physical-map orders of markers are not without errors. Errors in maps can greatly affect the ability to map and isolate genes for complex and Mendelian traits (Risch and Giuffra 1992; Feakes et al. 1999; Göring and Terwilliger 2000). Inaccuracies in genetic maps can result from genotyping errors, as well as from the use of a limited number of informative meioses to generate maps. A higher confidence in genetic-map order can be obtained by creating maps using a likelihood-ratio criterion of  $\geq 3$ , as opposed to using a minimum-recombination map (Morton 1955). Errors in the order of markers on physical maps can be due to problems with assembly or to incorrect identification of marker positions. Even when the order of markers is known to be without error, accurate estimates of recombination fractions will play an important role in linkage and association studies (Clerget-Darpoux et al. 1986; Risch and Giuffra 1992; Goddard et al. 2000; Collins et al. 2001; Reich et al. 2001).

The accuracy of the order of genetic and sequence-based physical maps was evaluated. The genetic order

of autosomal genome-scan markers from Marshfield panels 9 and 10 were compared with their physical order, on the basis of the assembled nonredundant human genome sequence from the Human Genome Project–Santa Cruz (HGP-sc) (International Human Genome Sequence Consortium 2001) and Celera (CEL) (Venter et al. 2001) databases.

## Methods

### *Ordering Genetic Markers*

The Web-based MAP-O-MAT program (Matise and Gitlin 1999) was used to obtain the likelihood for multiple alternative marker orders. MAP-O-MAT implements many features of the CRI-MAP program (Lander and Green 1987) by use of genotype data from its own database, which is a compilation of marker genotypes for the Centre d'Etude du Polymorphisme Humain (CEPH), pedigrees from the Marshfield Medical Center for Medical Genetics, and the Foundation Jean Dausset CEPH databases. The order according to the Marshfield map was used as the base order (Broman et al. 1998). The marker orders were then permuted in three tuples, and the likelihood of the resulting orders was compared with the likelihood for the Marshfield map. The likelihood-ratio support for the Marshfield maps is the difference of the base-10 likelihoods for the Marshfield map and the second-most-likely order. In cases where the order could not be supported by a likelihood ratio of  $\geq 3$ ,

Received July 30, 2001; accepted for publication October 4, 2001; electronically published November 9, 2001.

Address for correspondence and reprints: Dr. Suzanne M. Leal, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399. E-mail: [suzanne@linkage.rockefeller.edu](mailto:suzanne@linkage.rockefeller.edu)

© 2002 by The American Society of Human Genetics. All rights reserved. 0002-9297/2002/7001-0011\$15.00

**Table 1****Descriptive Statistics of Marshfield Panels 9 and 10 Screening Sets**

PANEL	NO. OF MARKERS IN PANEL (No. [%] NOT SUPPORTED BY LR $\geq$ 3)	NO. (%) OF MARKERS IDENTIFIED IN			NO. (%) OF INCONSISTENCIES IN <sup>a</sup>	
		HGP-sc (No. [%] e-PCR) <sup>b</sup>	CEL <sup>c</sup>	HPG-sc and CEL	HGP-sc (No. [%] ICA)	CEL (No. [%] ICA)
9	366 (8 [2])	283 (77) (33 [9])	253 (69)	204 (56)	5 (2) (1 [0.4])	11 (4) (3 [1.2])
10	380 (17 [4])	296 (78) (42 [11])	265 (70)	210 (55)	6 (2) (2 [0.7])	14 (5) (3 [1.1])

NOTE.—HGP-sc refers to the October 2000 release.

<sup>a</sup> ICA = inconsistent chromosomal assignment (i.e., markers in which the chromosomal assignment was not consistent between the physical and genetic map).

<sup>b</sup> Markers that were placed on the HGP-sc assembled genome sequence through e-PCR.

<sup>c</sup> No markers could be placed on the CEL assembled genome sequence with e-PCR, because of the unavailability of the data.

additional markers from the region were incorporated into the map to increase the informativeness of flanking markers. This procedure was mainly beneficial for increasing the level of support for the position of the most telomeric markers by providing informative meioses on both flanking sides of the marker.

*Physical Positioning of Markers*

A search was carried out using both marker and primer name(s) in the HPG-sc and CEL databases of assembled draft data for all autosomal markers that comprise the Marshfield 9 and 10 panels. The HGP-sc contains 2.9 gigabases (Gb) in 6,094 nonredundant sequences and had an assembly date of October 7, 2000. The analysis was repeated for the HGP-sc data after all tracks for the April 2001 release became available. The CEL database is Celera Genomics' "Human Sequence D," which represents 2.9 Gb in 54,061 sequences, with a release date of February 2001. These data were obtained through an academic license for the Celera database.

To place additional markers on the physical map, assembled sequence data were downloaded from the Santa Cruz database (October 2000 and April 2001 releases) by chromosome, and electronic PCR (e-PCR) (Schuler 1998) was used to map the marker primers onto the assembled genome sequence by providing STS name, left and right PCR primers, and expected product length (Genome Database). To maximize the number of markers to be placed on the assembled sequence map, six runs were carried out for each marker permitting for differences in the number base pair mismatches ( $n = 0, 1, 2$ ) and allowed deviations in the product size ( $m = 50, 1,000$ ). The word length was set to its default value of 7. An additional 33 (9%) of the panel 9 markers and 42 (11%) of the panel 10 markers could be placed on the HGP-sc (October 2000 release) map by use of e-PCR. For HGP-sc (April 2001 release), an additional 30 (8%) of the panel 9 markers and 41 (11%) of the panel 10 markers could be placed on the sequence-based phys-

ical map using e-PCR. The position of all these markers was consistent with the genetic-map order. The use of too-stringent parameter values for e-PCR will limit the number of markers that can be placed on the assembled sequence. Although relaxation of the parameter stringency will enable placement of a higher percentage of markers, it also increases the false-positive placement rate. This procedure could not be performed on the CEL-assembled genomic sequence, because e-PCR was not available at this site and because the sequence data could not be obtained in a timely fashion.

There are differences in how HGP-sc and CEL sequences were generated and assembled. HGP-sc generated a series of overlapping clones that cover the entire genome and shotgun sequence of each clone. The genome sequence was reconstructed by assembling the fragments on the basis of sequence overlap and map and chromosomal position information on clones (International Human Genome Sequence Consortium 2001). CEL used a whole-genome shotgun-sequencing approach (Venter et al. 2001).

It is not possible to make direct comparisons of the positions of the markers in the HGP-sc and CEL databases but only of their relative order. This is because of the way gaps in the assembled sequence data are treated. Both HGP-sc and CEL insert strings of the letter "N" to represent gaps. HGP-sc generally uses strings of 100 "N"s to represent gaps, while CEL uses different lengths of "N"s to represent the estimated size of the gaps (Aach et al. 2001).

**Results**

The order of genetic markers on two commonly used 10-cM genome-scan panels, Marshfield 9 and 10, were compared with the physical order of these markers based on the assembled nonredundant human genome sequence from the HGP-sc and the private sector CEL databases. The results of the comparison of the genetic and physical maps can be found in tables 1 and 2.

**Table 2****Descriptive Statistics of Marshfield Panels 9 and 10 Screening Sets for HGP-sc (April 2001 Release)**

PANEL	NO. OF MARKERS IN PANEL (No. [%] NOT SUPPORTED BY LR $\geq$ 3)	NO. (%) OF MARKERS IDENTIFIED IN		NO. (%) OF INCONSISTENCIES HGP-sc (No. [%] ICA <sup>b</sup> )
		HGP-sc (No. [%] e-PCR) <sup>a</sup>	HGP-sc and CEL	
9	366 (8 [2])	324 (89) (30 [8])	224 (61)	5 (2) (1 [0.4])
10	380 (17 [4])	328 (86) (41 [11])	231 (61)	7 (2) (3 [0.7])

<sup>a</sup> Markers that were placed on the HGP-sc assembled genome sequence through e-PCR.

<sup>b</sup> ICA = inconsistent chromosomal assignment (i.e., markers in which the chromosomal assignment was not consistent between the physical and genetic map).

Marshfield panel 9 consists of 366 autosomal markers. Of these markers, 297 (81%) also appear in panel 10. The order of these markers was consistent in both panels. For panel 9, 283 (77%) and 253 (69%) markers could be found in the HGP-sc (October 2000 release) and CEL sequences, respectively. The genetic order was not supported by a likelihood ratio of  $\geq$ 3 for 8 of the 366 markers. There was an inconsistency in the order of 5 (2%) and 11 (4%) of the markers when the genetic order was compared with their order in HGP-sc and CEL, respectively. Of the 11 inconsistent CEL markers, 9 were located in the HGP-sc database, and 8 of these markers were consistent with the genetic order. Likewise, five of the five inconsistent HGP-sc markers were identified in the CEL database, and four of these five markers were consistent with the genetic order. In the HGP-sc April 2001 release, three of the five previously inconsistent markers were assigned to a position that was consistent with the genetic-map order. For three markers in the CEL database, their sequence-based chromosomal assignment did not match their genetic chromosomal assignment: marker D4S1652 was assigned to chromosome 5; D7S2195 was assigned to chromosome 12; and marker D22S445 was assigned to chromosome 1. Only D4S1652 could not be localized on the HGP-sc sequence; for markers D7S2195 and D22S445, their position in the HGP-sc sequence was consistent with the genetic order. For one marker in the HGP-sc database, sequence-based chromosomal assignment did not match the genetic chromosomal assignment: D10S1225 mapped to chromosome 14. For the April 2001 release, this marker was removed from the database; however, by use of e-PCR, it was possible to map this marker to chromosome 10 on the sequence-based physical map in a position which was consistent with the genetic-map order. In the CEL database, D10S1225 mapped to chromosome 10; however, its physical position did not replicate its genetic position. For the inconsistent markers, there was no evidence for linkage to any other chromosome, and the genetic chromosomal position is supported by a likelihood ratio of  $\geq$ 3 (except for marker D1S3721, whose physical-map position was inconsistent in CEL). The genetic location

for D1S3721 was supported by a likelihood ratio of 2.5 and by its physical position in HGP-sc (October 2000 and April 2001 releases).

Of the 366 autosomal Marshfield panel 9 markers, 324 (89%) could be localized on the HGP-sc (April 2001 release) sequence-based physical map (see table 2). Five (2%) of these markers were assigned to a position that was inconsistent with the genetic-map order. The physical chromosomal assignment of one of these five markers did not match its genetic chromosomal assignment. Three markers had previously been assigned a physical-map position in the HGP-sc (October 2000 release), which was consistent with the genetic-map order. One of the markers had a physical-map position in the HGP-sc October 2000 release that was also inconsistent with the genetic map. One marker, D8S373, was not localized in the HGP-sc October 2000 release and was physically mapped to chromosome 3 in the HGP-sc April 2001 release. None of the inconsistent markers were placed on the sequence-based physical map via e-PCR.

Of the 380 autosomal Marshfield panel 10 markers, 296 (78%) and 265 (70%) could be found in the HGP-sc (October 2000 release) and CEL sequences, respectively. Of these 380 markers, the genetic order of 17 markers could not be supported by a likelihood ratio of  $\geq$ 3. When the genetic order of the panel 10 markers was compared with their order in HGP-sc and CEL, there was an inconsistency for 6 (2%) and 14 (5%) of the markers, respectively. Only two of these markers were inconsistent in both HGP-sc and CEL. For 2 of 6 HGP-sc and 3 of 14 CEL markers, the physical and genetic chromosomal assignment did not agree. In both the HGP-sc and CEL databases, marker D20S159 was assigned to chromosome 2. In the HGP-sc April 2001 release, this marker remained mapped to chromosome 2. In the HGP-sc database (October 2000 release), marker D10S1225 was assigned to chromosome 14; however, through e-PCR, this marker could be mapped to chromosome 10 in a position that is consistent with the genetic-map order in the HGP-sc (April 2001 release) sequence data. D10S1225 was mapped to chromosome 10 in the CEL sequence data, but its position

was inconsistent with the genetic map. In the CEL database, marker D4S1652 was assigned to chromosome 5, and marker D7S2477 was assigned to chromosome 16. Of the 14 inconsistent CEL markers, 10 were present in the HGP-sc database and the position of 8 of these markers was consistent with the genetic order. Likewise, four of the six inconsistent HGP-sc markers were identified in the CEL database, and the positions of two of these four markers were consistent with the genetic order. Positions of two of the six inconsistent HGP-sc markers were modified in the April 2001 release, and their position is now consistent with the genetic-map order. An additional two of the six inconsistent markers had been removed from the April 2001 release but could be mapped, using e-PCR, to a location on the physical map that was consistent with the genetic-map order. Only one of six markers was inconsistent with the genetic chromosomal assignment in HGP-sc (October 2000 and April 2001 release) and in CEL. For the inconsistent markers, there was no evidence for linkage to any other chromosome, and the genetic position was supported by a likelihood ratio of  $\geq 3$ , except for marker D1S1627. The position of marker D1S1627 in HGP-sc (October 2000 release) was not consistent with either its CEL, HGP-sc (April 2001 release), or genetic-map position.

For the Marshfield panel 10 markers, it was discovered that one marker was incorrectly mapped: marker D11S4463 was placed as the most telomeric marker on 11q. This marker mapped between markers D11S4464 and D11S1304, with support of a likelihood ratio of  $\geq 3$ . The most telomeric chromosome 20 marker, D20S164, and its flanking marker—D20S451, on the Marshfield map—were shown to map with 11 cM between them. This appears to be an error, since the Marshfield database also reports that there are zero recombinant events between the two markers. Our analysis showed that the genetic-map distance between these two markers was 0 cM, and the sequence-based physical map suggests that marker D20S451 was the most telomeric marker, not D20S164.

Of the 380 autosomal Marshfield panel 10 markers, 328 (86%) markers could be found in the HGP-sc April 2001 release (see table 2). A total of 7 (2%) of 328 markers were not consistent with the genetic map. Three of these markers were assigned to a chromosome on the sequenced-based physical maps that was different from that assigned on the genetic maps. Five of these seven markers were previously consistent with the genetic-map order. Of the three markers whose sequence-based physical chromosomal assignment was inconsistent with the genetic map, one marker, D8S373, was not localized in the HPG-sc October 2000 release. Marker D20S159 was localized to chromosome 2 in both the October 2000 and April 2001 HGP-sc releases. Marker

D6S942 was mapped to a position that was consistent with the genetic map in the HGP-sc October 2000 release but is physically mapped to chromosome 20 in the April 2001 release. None of the inconsistent markers were placed on the sequence-based physical maps via e-PCR.

The number of markers from Marshfield panels 9 and 10 that were localized in the HGP-sc April 2001 release increased from the HPG-sc October 2000 release by 12% and 8%, respectively (see table 2). The number of markers with inconsistent chromosomal assignments and positions was approximately the same in the HGP-sc October 2000 and April 2001 releases (see tables 1 and 2).

A total of 204 panel 9 markers and 210 panel 10 markers could be identified in both the CEL and HGP-sc (October 2000) databases. For these markers, the physical order was consistent between CEL and HGP-sc, except for 12 panel 9 markers and 12 panel 10 markers. In the situation where the physical position in CEL was inconsistent with that in HGP-sc, either one or both of these physical maps were inconsistent with the genetic order.

The Marshfield panels 9 and 10 consist of dinucleotide, trinucleotide, and tetranucleotide short-tandem-repeat polymorphism (STRP) markers. Marshfield panel 9 consists of 12% dinucleotide, 13% trinucleotide, and 75% tetranucleotide markers. The percentages for Marshfield panel 10 were very similar, with 13% dinucleotide, 15% trinucleotide, and 72% tetranucleotide markers. For those markers in panel 9 whose physical-map positions in CEL and HGP-sc (October 2000 release) were inconsistent with the genetic order, 13% were dinucleotide, 7% were trinucleotide, and 80% were tetranucleotide markers. Likewise, for panel 10, 17% were dinucleotide, 11% were trinucleotide, and 72% were tetranucleotide markers. There was no statistically significant difference between the distribution of marker types for those markers contained in panels 9 and 10 and the inconsistent markers.

Markers whose location on the CEL and HGP-sc (October 2000 release) physical maps was inconsistent with their location on the genetic maps were evenly distributed throughout the chromosome (i.e., there was no preference for errors in physical mapping near the telomeres or centromeres). Genetic markers from panels 9 and 10 whose order in HGP-sc (October 2000 release) was inconsistent with the genetic order were located closer to gaps than were those markers whose order was consistent with the HGP-sc physical order. The distance from the closest left or right gap was measured for a total of nine markers from panels 9 and 10 (five markers from panel 9 and six markers from panel 10, of which two were on both panels) whose physical position in HGP-sc was inconsistent with the genetic order.

**Table 3**  
**Representative Data from Chromosome 1 Panel 10**

LOCUS	PRIMER	MEAN MALE/ FEMALE DISTANCE (cM)	DISTANCE BETWEEN (cM)	MARKER IN <sup>a</sup>			LIKELIHOOD RATIO <sup>b</sup>
				CEL (Feb. 2001)	HGP-sc (Oct. 2000)	HGP-sc (April 2001)	
D1S468	AFM280we5	4	4	3536348	3238564	3620013	...
D1S1612	GGAA3A07	16	12	7852129	7099412	8032125	...
D1S1597	GATA27E01	30	14	12382248	14608589	13915789	...
D1S3669	GATA29A05	37	7	15999514	18762108	18770940	...
D1S3720	ATA47D07	47	10	Absent	22213770	22187210	...
N/A	ATA79C10	63	16	Absent	Absent	Absent	...
D1S3721	GATA129H04	73	10	126052773 <sup>c</sup>	46522902	46112020	...
D1S2134	GATA72H07	76	3	38725439	53537868 <sup>d</sup>	53014166	...
D1S1596	GATA26G09	89	13	Absent	Absent	Absent	...
D1S1665	GATA61A06	102	13	69497578	83523413	83214185	...
D1S1728	GATA109	109	7	81053684	92067587	92600191	...
D1S551	GATA6A05	114	5	81953234	97047900	93582497	...
N/A	GATA124C08	129	15	97983894	109657112 <sup>d</sup>	110100436	...
N/A	GATA133A08	138	9	103722597	118730601	118356904	.02
D1S1627 <sup>e</sup>	ATA25E07	139	0	104403202	4202244 <sup>c</sup>	119184531 <sup>d</sup>	-.02
N/A	ATA42G12	139	1	104514269	Absent	Absent	...
D1S534	GATA12A07	152	13	107826394	132135880	127606003	...
D1S1653	GATA43A04	164	12	130097969	179441529	179415469	...
D1S1679	GGAA5F09	171	7	134642329	183733913	182287393	...
D1S1677	GGAA22G10	176	5	Absent	Absent	183790153	...
D1S1589	ATA4E02	192	16	149210694	196394156	195452379	...
D1S518	GATA7C01	202	10	159869348	216758316	210406121	...
D1S1660	GATA48B01	212	10	174286159	228229629	222162672	...
N/A	GATA124F08	226	14	Absent	233281756 <sup>d</sup>	234415260 <sup>d</sup>	...
D1S549	GATA4H09	240	14	194431323	252795726	246033727	...
D1S3462	ATA29C07	247	7	206458134	266903810	259585050	2.38
D1S235	AFM203yg9	255	8	210415223	270764150	273004084	...
D1S1594	GATA22D12	265	10	Absent	242015953 <sup>c</sup>	278364393	...
D1S1609	GATA50F11	275	10	220306831	279935369	267356033	...
D1S2682	AFMa272xc9	288	13	Absent	281702373	271290532	...

<sup>a</sup> Month and year of release are shown in parentheses. Absent indicates not present in database.

<sup>b</sup> Ellipsis indicates that order of markers is supported by a likelihood ratio of  $\geq 3$ . When a likelihood ratio is given, it is the difference in the log likelihood for the given order compared to the log likelihood when the order of the denoted marker is switched with the flanking marker that is displayed directly below.

<sup>c</sup> Physical order does not agree with genetic order.

<sup>d</sup> Markers assigned a physical position using e-PCR.

<sup>e</sup> The order shown is given by the Marshfield Genetic Map. There is slightly more support for the order where ATA42G12 and D1S1627 are transposed; however, the physical map suggests the transposed order is not correct.

The distance from the closest left or right gap was measured for an additional 45 markers chosen randomly from panels 9 and 10 (excluding markers from chromosomes 21 and 22) whose genetic-map position was consistent with the physical sequence-based order in HGP-sc. Chromosomes 21 and 22 were excluded from the analysis, since these chromosomes consist of finished sequence with few gaps. The difference in the distance from a gap for the 45 consistent and 9 inconsistent markers was significant ( $P = .002$ ), with the inconsistent markers being much closer to a gap. The 9 inconsistent markers were, on average, 8,723 bases from a gap (standard error [SE] = 3,405; range 808–27,704 bases from a gap), and the 45 consistent markers were,

on average, 73,465 bases from a gap (SE = 19,324; range 0–635,718 bases from a gap [value of 0 indicates that marker was within a gap]). This comparison was not made for CEL data, since it was not possible to elucidate the position of gaps from the CEL database.

Table 3 displays the markers from panel 10 on chromosome 1. To view marker data by chromosome for panels 9 and 10, see the Web site “Genetic and Sequence-Based Physical Maps for Genome Scan Markers.” This site displays the inconsistencies for Marshfield panel 9 and 10 markers between their genetic-map position and their sequence-based physical-map position, as well as inconsistencies between the sequence-based physical-map positions from data generated by Celera

and the Human Genome Project-Santa Cruz (October 2000 and April 2001 releases). This Web site also contains hotlinks, for each marker, to the Genome Database, as well as primer names, genetic-map distances, physical-map positions from the Celera and Human Genome Project sequences, and likelihood-ratio support for the genetic maps.

## Discussion

The majority of the inconsistencies between the physical and genetic-map order point to errors in the physical map order. The Marshfield panel 9 and 10 genetic maps have a high order of support, and the majority of the marker positions are supported by a likelihood ratio of  $\geq 3$  (table 1). In most cases where there is an inconsistency between the genetic and the physical marker order for one of the assembled sequence data sets, the genetic order is supported by the other physical map. Errors in marker order in the physical maps can be due to problems of assembly. For example, in HGP-sc (October 2000 release), it appears that a clone (AC037449.2) from chromosome 10 was assembled onto chromosome 14. Multiple chromosome 10 markers (D10S1225, D10S2278, SHGC-130658, stSG29297, and stSG43409) are located on this clone; however, there is no statistical support for genetic markers D10S2278 and D10S1225 mapping to chromosome 14. It should be noted that this clone is mapped to chromosome 10 in HGP-sc April 2001 release. Further problems in assembly can occur when markers are located near gaps in physical-sequence data. An additional cause of incorrect marker placement may be due to error in *in silico* mapping.

The genetic maps for Marshfield panels 9 and 10 were not generated using a likelihood-ratio criterion of  $\geq 3$  (Broman et al. 1998); however, in almost all cases, the order of the markers was supported by this criterion. Reasons for the failure to reach this level of support some regions of the genome may include the following: (1) the limited number of meioses (184; Yu et al. 2001) from the eight CEPH families genotyped for the majority of markers; (2) some markers may have a low heterozygosity and therefore even fewer informative meioses; and (3) a few markers were closely linked to each other (Terwilliger et al. 1992).

In the near future, as finished sequences become available and as gaps are filled and assembly problems solved for the human genome, it will become possible to know the position for genetic markers unequivocally. The problem that will remain is obtaining accurate estimates of recombination fractions between markers. It is not possible to interpolate between physical-map distances and the recombination fractions. It is known that there are jungles and deserts for the rates of recombination

(Buetow et al. 1991; Broman et al. 1998; Yu et al. 2001); however, very little is known about hot spots of recombination over small physical distances. In addition, the effects of sex (Weitkamp et al. 1973; Donis-Keller et al. 1987; Tanzi et al. 1992) and age of parent at conception (Weitkamp et al. 1973; Lange et al. 1975; Elston et al. 1976; Tanzi et al. 1992) on the rates of recombination are poorly understood. To understand the behavior of recombination events throughout the genome, additional informative meioses beyond the current commonly studied 184 meiotic events in the eight CEPH families would be required. An understanding of recombination is important for linkage and association studies. For linkage studies, inaccurate recombination fractions will lead to a loss of power and can bias the map position for the trait locus (Clerget-Darpoux et al. 1986; Risch and Giuffra 1992). In the case of association studies, accurate knowledge of the rates of recombination over small intervals will aid in the selection of the appropriate density of single-nucleotide polymorphism (SNP) markers. From our current knowledge, it appears that in regions of high rates of recombination, SNPs must be more closely spaced if linkage disequilibrium is to be detected (Goddard et al. 2000; Collins et al. 2001; Reich et al. 2001).

Although physical maps can greatly aid in confirmation of the order of genetic markers, physical maps are not without errors, and not all genetic markers have been localized on sequence-based physical maps. It is therefore useful not only to confirm the order of genetic markers, using information from sequence-based physical maps, but also to support the genetic-marker order, using a likelihood-ratio criterion of  $\geq 3$ . Likelihood-ratio support for a given order is most crucial when the sequence-based physical-marker order is inconsistent with a minimum-recombination genetic map or when a genetic marker cannot be localized on a sequence-based physical map.

## Acknowledgments

This work is supported by National Institutes of Health grants DC03594, HG01691, and HG00008.

## Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Celera, <http://www.celera.com/> (for the physical positions of STRP loci)

Foundation Jean Dausset CEPH, <http://www.cephb.fr/cephdb/dumps.html> (for genotype data from the CEPH pedigrees)

Genetic and Sequence-Based Physical Maps for Genome Scan Markers, <http://linkage.rockefeller.edu/maps/> (for a complete listing of panel 9 and 10 markers, primer names,

genetic map distances, physical map position from the Celera and Human Genome Project sequence (October 2000 and April 2001 releases), and likelihood-ratio support for the genetic maps)

Genome Database, <http://www.gdb.org> (for primer sequences used for e-PCR)

Human Genome Project Working Draft at UCSC, <http://genome.cse.ucsc.edu/> (for the physical positions of STRP loci and complete sequence data for e-PCR)

MAP-O-MAT, <http://compgen.rutgers.edu/mapomat/> (for ordering the STRP marker loci)

Marshfield Medical Center for Medical Genetics, <http://research.marshfieldclinic.org/genetics/> (for the order and genetic distances of STRP loci on panels 9 and 10, for the percentage of dinucleotide, trinucleotide and tetranucleotide markers on panels 9 and 10 and genotype data from the CEPH pedigrees)

## References

- Aach J, Bulyk ML, Church GM, Comanders J, Derti A, Shendure J (2001) Computational comparison of two draft sequences of the human genome. *Nature* 409:856–859
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861–869
- Buetow KH, Shiang R, Yang P, Nakamura Y, Lathrop GM, White R, Wasmuth JJ, Wood S, Berdahl LD, Leysens NJ, Ritty TM, Wise ME, Murray JC (1991) A detailed multipoint map of human chromosome 4 provides evidence for linkage heterogeneity and position-specific recombination rates. *Am J Hum Genet* 48:911–925
- Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986) Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 42:393–399
- Collins A, Ennis S, Taillon-Miller P, Kwok PY, Morton NE (2001) Allelic association with SNPs: metrics, populations, and the linkage disequilibrium map. *Hum Mutat* 17: 255–262
- Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES (1987) A genetic linkage map of the human genome. *Cell* 51:319–337
- Elston RC, Lange E, Namboodiri KK (1976) Age trends in human chiasma frequencies and recombination fractions. II. Method for analyzing recombination fractions and applications to the ABO:nail-patella linkage. *Am J Hum Genet* 28:69–76
- Feakes R, Sawcer S, Chataway J, Coraddu F, Broadley S, Gray J, Jones HB, Clayton D, Goodfellow PN, Compston A (1999) Exploring the dense mapping of a region of potential linkage in complex disease: an example in multiple sclerosis. *Genet Epidemiol* 17:51–63
- Goddard KA, Hopkins PJ, Hall JM, Witte JS (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am J Hum Genet* 66:216–234
- Görling HH, Terwilliger JD (2000) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am J Hum Genet* 66:1298–1309
- International Human Genome Sequence Consortium (2001) Initial sequence and analysis of the human genome. *Nature* 409:860–921
- Lander ES, Green P (1987) Construction of multi-locus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84: 2363–2367
- Lange K, Page BM, Elston RC (1975) Age trends in human chiasma frequencies and recombination fractions. I. Chiasma frequencies. *Am J Hum Genet* 27:410–418
- Matise TC, Gitlin JA (1999) MAP-O-MAT: marker-based linkage mapping on the World Wide Web. *Am J Hum Genet* 65:A2464
- Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* 411:199–204
- Risch N, Giuffra L (1992) Model misspecification and multipoint linkage analysis. *Hum Hered* 42:77–92
- Schuler GD (1998) Electronic PCR: bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol* 16:456–459
- Tanzi RE, Watkins PC, Stewart GD, Wexler NS, Gusella JF, Haines JL (1992) A genetic linkage map of human chromosome 21: analysis of recombination as a function of sex and age. *Am J Hum Genet* 50:551–558
- Terwilliger JD, Ding Y, Ott J (1992) On the relative importance of marker heterozygosity and intermarker distance in gene mapping. *Genomics* 13:951–956
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al (2001) The sequence of the human genome. *Science* 291:1304–1351
- Weitkamp LR, Van Rood JJ, Thorsby E, Bias W, Fotinio M, Lawler SD, Dausset J, Mayr WR, Bodmer J, Ward FE, Seignalet J, Payne R, Kissmeyer-Nielsen F, Gatti RA, Sachs JA, Lamm LU (1973) The relation of parental sex and age to recombination fraction in the HLA-A system. *Hum Hered* 23:197–205
- Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL (2001) Comparison of human genetic and sequence-based physical maps. *Nature* 409:951–953