Large-Scale Integration of Human Genetic and Physical Maps

Caroline M. Nievergelt, Douglas W. Smith, J. Bradley Kohlenberg, et al.

Genome Res. 2004 14: 1199-1205
Access the most recent version at doi:10.1101/gr.1475304

Resource

Large-Scale Integration of Human Genetic and Physical Maps

Caroline M. Nievergelt,[1] Douglas W. Smith,[1,2] J. Bradley Kohlenberg,[1] and Nicholas J. Schork[1,3]

[1]*Polymorphism Research Laboratory, Department of Psychiatry and* [2]*Department of Biology, University of California at San Diego, La Jolla, California 92093-0603, USA*

Genetic maps are used routinely in family-based linkage studies to identify the rough location of genes that influence human traits and diseases. Unlike physical maps, genetic maps are based on the amount of recombination occurring between adjacent loci rather than the actual number of bases separating them. Genetic maps are constructed by statistically characterizing the number of crossovers observed in parental meioses leading to the transmission of alleles to their offspring. Considerations such as the number of meioses observed, the heterozygosity and physical distance between the loci studied, and the statistical methods used can impact the construction and reliability of a genetic map. As is well known, poorly constructed genetic maps can have adverse effects on linkage mapping studies. With the availability of sequence-based maps, as well as genetic maps generated by different researchers (such as those generated by the Marshfield and deCODE groups), one can investigate the compatibility and properties of different maps. We have integrated information from the most current human genome sequence data (UCSC genome assembly Human July 2003) as well as 8399 microsatellite markers used in the Marshfield and deCODE maps to reconcile the these maps. Our efforts resulted in updated sex-specific genetic maps.

[Supplemental material is available online at www.genome.org and http://elcapitan.ucsd.edu/hyper/.]

The use of genetic maps for identifying genes that influence human traits and diseases has a long and illustrious history in the medical and biological sciences. Indeed, as pointed out in a recent historical perspective on human biological research by Lander and Weinberg (2000), many advances in the identification of disease-predisposing genes for traits and diseases of all sorts have paralleled advances in the development of genetic and physical maps. Unlike physical maps of genomic positions or loci—which attempt merely to tally the number of bases separating adjacent loci—genetic maps seek to characterize the frequency of meiotic recombination occurring between loci. As such, the construction of genetic maps is largely a statistical undertaking, whereby individuals within families are genotyped at many loci and the patterns associated with the transmission of alleles are recorded to determine how often alleles or variants at adjacent loci are either coinherited or shuffled as a result of recombination in relevant parental meioses (see, e.g., Ott 1999).

The reconciliation of locus positions dictated by genetic and physical maps is not trivial, unfortunately. Large-scale sequencing and the ordering of loci from multiple sequence reads can result in errors of all sorts, most simply caused by sequencing errors and missing sequence or "gaps" between the sequences (see, e.g., Lander et al. 2001; Venter et al. 2001). These errors are often corrected over time, resulting in periodic updates to available human genome sequence information. There are also issues that compromise the reliability of genetic maps. As is well recognized and emphasized in the description of the recent construction of the genetic map by Kong et al. (2002), factors such as the number of meioses studied and the informativity of the DNA markers used, can impact the reliability of a genetic map. In addition, it is also known that the construction and ultimate use of genetic maps can be compromised because of missing data and simple genotyping errors (Goring and Terwilliger 2000).

As a result of problems associated with the construction of human physical and genetic maps, differences with respect to the positions of loci exist not only between physical and genetic maps, but also among different physical and genetic maps. Because genetic maps play a crucial role in pedigree-based meiotic (or "linkage") mapping gene discovery strategies, it is important to consider their reconcilability with physical maps. In addition, it has been well documented that misspecification of genetic maps can have negative effects on linkage analyses (see, e.g., Daw et al. 2000; Goring and Terwilliger 2000; Hackett and Broadfoot 2003). Three types of genetic map misspecification, listed from the most severe forms of misspecification to the least severe, plague linkage analyses, especially in the context of "multipoint" analyses that depend crucially on reliable DNA marker positions: (1) Loci are erroneously assumed to be on a particular chromosome. (2) Loci are out of order but on the correct chromosome. (3) Loci are on the correct chromosome and in the correct order but their interlocus distances, in terms of the amount of recombination that occurs between them, are misspecified. It must be emphasized that when reconciling genetic and physical maps, there are likely to be errors in each—despite the efforts of those who created them. Thus, incompatibilities found across the maps really only point out areas of the genome that call for re-evaluation of both types of map.

Some researchers have considered the quality and reconcilability of genetic and physical maps (see http://cedar.genetics.soton.ac.uk/public_html/LDB2000), but most of these efforts have focused on a single chromosome or genomic region (see, e.g., Tapper et al. 2001; Matise et al. 2002), or were pursued prior to recent updates of the draft human genome sequence and the availability of the comprehensive genetic map constructed by Kong et al. (2002; see, e.g., Yu et al. 2001; DeWan et al. 2002). In this paper, we extend a recent study by Matise et al. (2002) that sought to evaluate the comparability of genetic and physical

[3]Corresponding author.
E-MAIL nschork@ucsd.edu; FAX (858) 822-2113.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.1475304. Article published online before print in May 2004.

maps involving a subset of routinely used genetic markers on Chromosome 22, by considering the entire set of 8325 microsatellite (STRPs) markers examined by Broman et al. (1998; referred to as the "Marshfield" map and markers), the 5136 microsatellites examined in a recently developed genetic map by Kong et al. (2002; referred to as the "deCODE" map and markers), and the most recent releases of the sequence of the Human genome (UCSC genome assembly Human July 2003, http://genome.ucsc.edu/; and the Celera database, http://myscience.appliedbiosystems.com/; Venter et al. 2001). We document ambiguities and differences in the physical and genetic maps and comment on the problems these ambiguities might create when the maps are used for different purposes.

Most importantly, we have developed a comprehensive and integrated sex-specific genetic map that could be used in, for example, multipoint linkage analyses or help focus further sequencing and marker ordering efforts. Our map includes all of the Marshfield and deCODE markers and is based primarily on the genetic positions of loci in deCODE map because of the large number of meioses studied to develop that map. The high-resolution deCODE map is based on 1257 meiotic events in two-generation pedigrees and is the most accurate genetic map available to date. However, deCODE only includes 5136 microsatellites, whereas the older, less accurate Marshfield map, which was calculated based on only 188 meioses in three generation pedigrees, includes 8325 microsatellite markers. Our map uses up-to date physical sequence information to interpolate the positions of markers not included in the deCODE map, but whose position can be estimated relative to markers on that map. A method for querying markers in this map using a simple Excel spreadsheet query macro is available as Supplemental material from the authors' Web site (http://elcapitan.ucsd.edu/hyper/).

RESULTS

Chromosome Positions

Of the 7737 Marshfield markers with available physical positions (UCSC July 2003), 45 were found to have chromosomal assignments that did not match the assignment provided by the physical map (Table 1). Of these 45 markers, 28 are "Utah" markers, which were typed on only a few meioses (four of the eight CEPH families used to form the Marshfield genetic maps). Only one of these 45 markers was considered in the construction of the deCODE map. In addition, only three markers are used in the routine linkage mapping panels provided by Marshfield Clinic. Three markers (PLA2, GC, and FB7F11) have been mapped to the chromosome assigned by the physical map via radiation hybrids, suggesting that the initial chromosomal positions provided by the Marshfield map were likely wrong. Finally, some of these markers have chromosomal assignments based on the physical map that have changed as the physical map has been updated (Table 1, cf. columns 9–11), suggesting a lack of confidence in their location as dictated by available sequence information. For comparison purposes, we also provide positional information obtained from the Celera database for these 45 markers. For a subset of these markers, it is quite possible that the sequence-based position is simply wrong as well.

Marker Order

To assess marker order, we considered the number of markers implicated in a "block" of markers where there is misspecification. To clarify, consider a string of markers whose correct order is 1–2–3–4–5–6–7–8–9–10–11–12, but whose assumed order is 1–2–3–9–7–4–5–6–8–10–11–12: the block of markers 9–7–4–5–6–8 is "order misspecified." This type of order misspecification captures entire blocks, such that if, say, multipoint linkage methods were used to extract identity-by-descent allele-sharing information among relative pairs, the calculations would be wrong even though some of the markers in the block are in the correct order (i.e., markers 4–5–6 are not out of order but do sit in a block that contains order misspecification). Table 2 describes the percentages of markers examined from the Marshfield and deCODE genetic maps exhibiting block order problems, the average and maximum block length on a chromosome-by-chromosome basis, and the percentage of markers that are misplaced in their position by >2 cM and >5 cM, respectively. These parameters were calculated by ordering the markers according to their physical position and determining all markers that are out of order compared with their Marshfield and deCODE map locations, respectively. For these markers, "true" map positions were interpolated by using the physical and genetic map position of the markers preceding and following a block with misspecified markers, analogous to the way described in the Methods section for the comprehensive genetic map. This resulted in especially large values for Chromosomes 16 and 17 for the Marshfield map and for Chromosomes 5 and 7 for the deCODE map, respectively. These large values are partly caused by a few markers with relatively large disagreement between their physical and genetic map positions, which, in accordance with our definition of blocks, will result in a large maximum block length. Furthermore, because we used markers that were adjacent to the blocks as a basis for the estimation of the "true" genetic position of the markers in these blocks, a large block length negatively influences the accuracy of the estimated marker positions as well as the number of markers misplaced at a certain distance.

In general, markers whose order is misspecified on the genetic maps that reside in a small region of the genome are likely to arise from difficulties in resolving recombination events between markers with a finite number of meioses. Although it is clear that the deCODE map contains fewer order problems than the earlier Marshfield map, this is more than likely because Kong et al. (2002) had the luxury of leveraging human genome sequence information in the construction of their map, as well as having many more meiotic events. However, discrepancies in order between the deCODE map and the physical map do exist, and this probably reflects updates and changes to the physical map that occurred after Kong et al. (2002) constructed their map. In addition, as with markers on genetic maps with different chromosome positions from those provided by the physical map, there is the possibility that the physical map positions are incorrect as well and that the genetic orderings are correct in one, the other, or both of the genetic maps.

Interlocus Distances

To assess potential differences in the amount of recombination thought to occur between loci, we compared interlocus recombination fractions between marker loci across the Marshfield and deCODE maps. This comparison involved 4961 markers whose relative positions on all three maps could be obtained. Markers were first ordered according to their physical position, then the absolute genetic distance between adjacent markers was calculated for both the Marshfield and deCode maps. Differences in these interlocus distances between the two maps are shown in Figure 1. It can be seen that although there are some adjacent loci for which large discrepancies (i.e., >5 cM difference) in the amount of recombination exist, they are relatively few (<3%). In fact, 59% of the interlocus distances between the two maps are between 1 and 2 cM. This is owing to the fact that we studied such a large number of markers with a density large enough to preclude big gaps between our markers whose relative positions could be reconciled with the physical map. Differences in recom-

Table 1. Microsatellite Markers ($N = 45$) Whose Sequence Positions (UCSC Assembly Human July 2003/NCBI Build 34) Do Not Correspond to the Marshfield Chromosomal Position (http://research.marshfieldclinic.org/genetics/)

Marker name	UCSC ID	UCSC Chromosome	Marshfield chromosome	Other genetic maps	Yac map chromosome	RH maps chromosome	Marshfield screening	UCSC 11/02	UCSC 06/02	UCSC 04/02	CELERA 03/03
UT5086	7782	1	9	No	No	No	No	1	1	1	No
PLA2	7128	1	12	No	No	Gb4, G3, Wi:1	No	1	1	1	1 CelAnno
GC/SGC31591	5929	2	4	No	No	Wi:2	No	2	2	2	2RefSeq
UT710	6472	2	7	No	No	No	No	2	2	2	No
UT2548	6998	2	11	No	No	No	No	2	2	2	No
UT1307	7844	2	20	No	No	No	10	2	2	2	2 RefSeq
AFMA101XG5	842	3	2	Gen:2	No	No	No	3	3	3	No
GATA43F06	5557	3	2	No	Wi:3	No	No	3	3	3	No
UT885	6629	3	8	No	No	No	No	3	3	3	No
UT5819	6876	3	10	No	No	No	No	3	3	3	No
ATA4F06	6032	4	5	No	No	No	4–9	4	No	4	No
UT1531	6710	4	9	No	No	No	No	4	4	4	No
UT2361	6715	5	9	No	No	No	No	5	5	5	No
UT5140	7139	5	12	No	No	No	No	5	5	5	No
UT6886	7407	5	15	No	No	No	No	5	5	5	No
AFM023TG5	4343	5	17	Gen:17	No	No	No	5	5	5	17 RefSeq
ATA1H07	7502	5	17	No	No	No	No	5	5	5	5 CelAnno
GATA124A11	7816	6	20	No	No	No	No	6	No	No	6 CelAnno
GATA116E08	7893	8	21	No	No	No	6	8	8	8	No
AFM123yc5	328	8	1	Gen:1	No	No	No	8	No	8	No
UT924	7349	10	14	deC:14	No	No	No	10	10	10	No
ATA37G10	6499	11	8	No	No	No	No	11	11	11	No
UT1607	6989	12	11	No	No	No	No	12	12	12	No
UT935	7414	12	15	No	No	No	No	12	12	12	No
UT1888	7327	16	14	No	No	No	No	16	16	22	1 RefSeq
UT1598	7551	16	17	No	No	No	No	16	16	16	No
UT1025	7911	16	21	No	No	No	No	16	No	No	No
UT556	7920	16	21	No	No	No	No	16	No	2, 13	No
GAAT1A12	6043	16	5	No	Wi:5	No	No	16 (2×)	16	5	No
UT18	5621	17	2	No	No	No	No	17	17	17	No
UT5182	6611	18	8	No	No	No	No	18	18	18	18 RefSeq
FB7F11	6780	18	10	No	No	Gb4, Wi:18	No	18	18	18	No
PI	7319	18	14	No	No	No	No	18	18	18	No
UT7368	6476	19	7	No	No	No	No	19	19	19	19 CelAnno
UT5146	5641	21	2	No	No	No	No	11	1,2	1	No
GATA115F05	6519	21	8	No	No	No	No	21	21	21	21 RefSeq
UT5116III	5638	22	2	No	No	No	No	22	22	14,22	No
UT597	7339	22	14	No	No	No	No	22	14,22	14,22	No
UT6047	7341	22	14	No	No	No	No	22	No	2,14,22	No
UT7691	5657	X	2	No	No	No	No	X	No	No	2 CelAnno
UT764	7635	X	9	No	No	No	No	X	X	X	1 RefSeq
UT832	7020	X	11	No	No	No	No	X	No	No	X CelAnno
UT6574	7156	X	12	No	No	No	No	X	X	X	No
ATA5E03	5293	Y	1	No	No	No	7	Y	Y	Y	No
IFNAR	7908	Y	21	No	No	No	No	Y	Y	1,Y	No

A comparison to chromosomal locations from additional sources is shown.
Key: 7737 markers are included in the comparison. Available positional information from genetic maps (Genethon, DeCODE), Yac map (Whitehead Institute), and radiation hybridization (RH) maps (NCBI GeneMap99 GB4 and G3, Whitehead Institute, and Stanford TNG RH maps) is provided to help determine which source of information is erroneous. In addition, markers included in the commonly used Marshfield screening sets (1–12) are indicated. Chromosomal positions are also compared with older UCSC assemblies and with the sequence-based Celera database (http://myscience.appliedbiosystems.com/). Markers listed on the Y-chromosome are in the pseudoautosomal region of the Y-chromosome.

bination rates between the same loci in the two genetic maps could be the result of errors in positioning the markers, sampling variation due to the finite number of meioses used to construct each map, or inherent differences in recombination rates in the populations used to construct the maps.

Comprehensive Genetic Map

From our efforts at examining the reconcilability and comparability of the Marshfield and deCODE genetic maps along with the latest versions of the available human genome sequence, we developed a comprehensive and integrated map (see Methods). An extract is shown in Table 3. The comprehensive genetic map includes a total of 8399 markers with information on the physical position (UCSC assembly July 2003), and Marshfield and deCODE recombination rates, as well as interpolated deCODE values for 2838 markers. A summary is shown in Table 4. This map is available for downloading with macros to facilitate marker searches from http://elcapitan.ucsd.edu/hyper/.

Table 2. Number and Percentage of Microsatellites, Block Length, and Percentage of Markers Misplaced for the 7634 Marshfield and 5093 deCODE Markers With Unambiguously Determined Physical Positions Whose Marker Order (See Text) Does Not Correspond to the Order Based on the Sequencing Data (UCSC Assembly July 2003; NCBI Build 34)

	Marshfield						deCODE					
	Markers misspecified		Block length		Misplaced		Markers misspecified		Block length		Misplaced	
Chromosome	No.	(%)	Mean	Max.	% > 2 cM	%>5 cM	(no.)	(%)	Mean	Max.	%>2 cM	%>5 cM
1	231	36.61	2.27	10.49	4.75	0.16	4	0.88	1.19	2.25	0.00	0.00
2	254	41.17	2.97	17.13	4.21	0.81	18	4.43	4.57	10.67	0.25	0.00
3	196	37.98	2.03	7.87	4.07	0.58	46	12.50	1.18	2.87	0.00	0.00
4	199	45.43	2.97	8.03	5.71	0.23	5	1.67	1.28	1.72	0.00	0.00
5	209	45.83	2.39	7.13	4.82	0.00	119	35.95	14.95	28.44	12.08	0.30
6	192	43.34	2.50	8.32	6.55	0.23	0	0.00	0.00	0.00	0.00	0.00
7	219	54.48	3.13	10.71	13.68	3.48	68	27.98	11.15	36.23	15.23	7.82
8	179	46.74	2.98	23.54	3.39	0.26	17	6.88	2.47	4.06	1.21	0.00
9	147	52.13	4.39	11.14	10.28	2.48	6	3.16	1.54	2.98	0.00	0.00
10	211	51.72	2.92	8.87	6.86	1.23	2	0.79	2.39	2.39	0.00	0.00
11	199	51.69	2.98	10.67	3.64	0.78	4	1.54	2.61	3.93	0.00	0.00
12	168	45.53	3.04	11.51	7.32	1.08	2	0.84	0.94	0.94	0.00	0.00
13	80	32.79	2.56	6.99	3.28	0.82	2	1.14	3.00	3.00	0.00	0.00
14	96	39.02	2.97	10.71	5.28	2.03	0	0.00	0.00	0.00	0.00	0.00
15	78	38.42	2.92	6.71	5.42	0.00	0	0.00	0.00	0.00	0.00	0.00
16	149	59.13	3.54	26.55	16.27	11.11	8	5.33	2.00	3.17	0.00	0.00
17	204	71.33	6.06	27.18	26.92	11.89	4	2.23	0.44	0.52	0.00	0.00
18	88	41.90	2.83	9.63	6.67	2.38	0	0.00	0.00	0.00	0.00	0.00
19	83	39.52	2.59	7.94	2.38	0.48	0	0.00	0.00	0.00	0.00	0.00
20	86	41.55	4.04	11.59	6.28	1.93	2	1.42	0.28	0.28	0.00	0.00
21	69	65.09	3.32	6.28	12.26	2.83	3	4.48	3.39	4.86	2.99	2.99
22	51	53.13	3.52	10.56	10.42	1.04	0	0.00	0.00	0.00	0.00	0.00
X	79	32.38	4.52	8.77	11.07	2.87	10	5.68	3.20	8.32	2.84	2.84
Total	3467	45.42	3.19	11.67	7.22	1.77	320	6.28	2.46	5.07	1.73	0.53

DISCUSSION

The study of recombination rates, as well as the exploitation of recombination rates in human gene-mapping studies, have received a great deal of attention, especially since the publication of the draft of human genome sequence and the announcement of the human Haplotype Map Initiative (see, e.g., Dawson et al. 2002; Gabriel et al. 2002). With the current availability of multiple physical and genetic maps, it behooves researchers to consider their reconcilability, if for no other reason than to have confidence in their correctness. Our assessment of the complete set of markers used in the construction of the genetic maps by Broman et al. (1998) and Kong et al. (2002) in conjunction with an updated physical map suggests that many discrepancies in maps exist, raising the question of which map might be correct. Although it is quite likely that the genetic map constructed by Kong et al. (2002) is more reliable then the initial Broman et al. (1998) map, given the use of a larger number of meioses in the construction of that map and the fact that Kong et al (2002) pursued the construction of their map after the publication of the draft human genome sequence, there still may be problems with all maps. The published human genome sequence is somewhat "fluid" given constant updating, and this is consistent with the existence of discrepancies involving the Kong et al. (2002) map compared with the physical map, as we have shown.

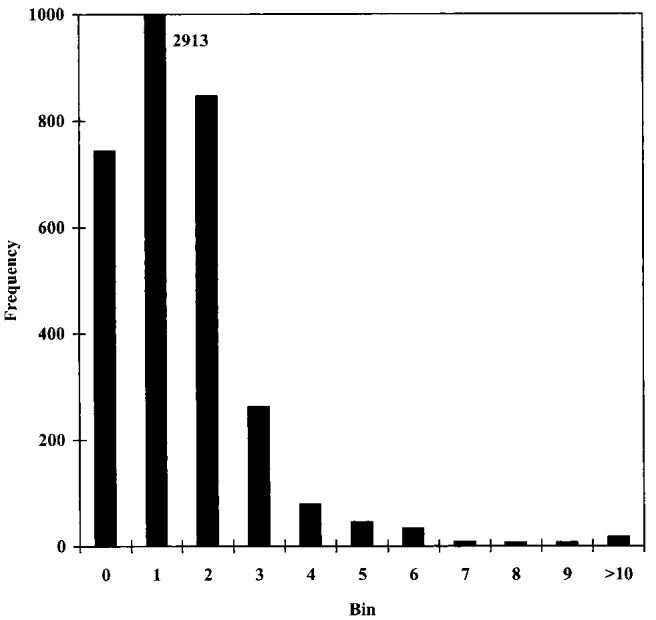

Figure 1 Distribution of the differences in interlocus distances between 4961 markers quoted both in the initial Marshfield map (Broman et al. 1998) and also computed in the newer deCODE map (Kong et al. 2002). Note: Bin >10 includes interlocus distance differences up to 36 cM.

Map misspecification can have serious negative consequences on gene-mapping studies. Incorrect physical maps can complicate late stages of positional cloning efforts, whereas incorrect genetic maps can complicate initial linkage analyses. Our analysis involved all available markers in the Marshfield (Broman et al. 1998) and deCODE maps (Kong et al. 2002) and not just the few hundred or so that researchers might use in an initial linkage-based gene-mapping project. Thus, our findings relating to marker order involved many more markers (and hence many more opportunities for marker order misspecification) than is likely to arise in a routine linkage analysis. Despite this, we have found evidence for map problems with previous linkage analyses using specific sets of DNA markers. For example, in our reassess-

Table 3. Extract From the "Comprehensive Genetic Map" Available on the Author's Web Page, Showing UCSC Physical Positions and Average (av.), Female (f), and Male (m) Recombination Rates (in Centimorgans) From the deCODE (deC) and Marshfield (M) Genetic Maps

UCSC chr	Marker	Primer	Phys. pos.	UCSC ID	deC av.	deC f	deC m	M av.	M f	M m	Remark	deC Information	Dupl. Marker	Dupl. Primer
1	ATA9B08	no	10697084	5299	18.11	20.41	15.81	23.35	23.37	23.19		deC interpol.		
1	UT5086	D19S724	11174025	7782	19.80	22.10	17.51	45.48	49.79	41.63	Marshfield Chr19	deC interpol.		
1	UT491	D1S1151	11174106	5431	19.80	22.10	17.51	24.68	30.54	23.19		deC interpol.		
1	AFMA224WG9	D1S2667	11196506	21	19.88	22.18	17.59	24.68	26.18	23.19		deC original		
1	AFMA153XE9	D1S2644	18459547	34	35.56	43.24	27.88	43.72	48.44	39.23		deC original		
1	ATA43C09	D1S3726	18575541	5287	36.74	44.81	28.67	45.33	50.41	40.52		deC original		
1	AFM296ZC9	D1S483	18701543	39	36.92	44.81	29.03	45.33	50.41	40.52	score 750	deC original		
1	GGAT2A07	D1S552	18736435	5379	37.30	45.57	29.03	45.33	50.41	40.52		deC original		
1	UT497	D1S378	18983812	5434	37.38	45.73	29.03	48.53	55.58	41.76		deC interpol.		
1	AFMA162ZC9	D1S2647	19297590	36	37.48	45.93	29.03	45.33	50.41	40.52		deC original		
1	AFM078YG5	D1S199	19426467	35	37.48	45.93	29.03	45.33	50.41	40.52		deC original		
1	AFM343ZA9	D1S2843	19979855	37	38.75	47.71	29.78	46.61	51.70	41.76		deC interpol.		
1	AFMB040YB1	D1S2732	20105781	41	39.04	48.12	29.95	48.53	55.58	41.76		deC original		
1	ATA47D07	no	20381259	5288	39.31	48.40	30.22	46.61	51.70	41.76		deC interpol.		
1	AFM290VB9	D1S478	21069079	38	40.00	49.11	30.88	48.53	55.58	41.76		deC original		
1	MFD293	D1S334	21253109	5394	40.20	49.51	30.88	48.53	55.58	41.76		deC original		
1	AFM303TG1	D1S2828	21411895	40	40.64	50.02	31.26	48.53	55.58	41.76		deC original		
1	GGAA30B06	no	21449412	5375	40.66	50.06	31.26	48.53	55.58	41.76		deC interpol.		
1	AFMB016YB5	D1S2725	21599275	42	40.74	50.23	31.26	48.53	55.58	41.76		deC original		
1	AFM193XB12	D1S2702	22017886	43	40.98	50.71	31.26	49.07	56.66	41.76		deC interpol.		
1	AFMa134vb1	D1S512	22057501	no	41.01	50.75	31.26	48.53	55.58	41.76	placed by BLAT	deC interpol.		
1	UT5123	D1S1539	22190088	5438	41.08	50.91	31.26	49.07	56.66	41.76		deC interpol.		
2	AFMA130WG1	D2S2163	37910179	546	62.04	73.14	50.95	59.36	65.66	53.02		deC orginal		
2	AFM267ZC9	D2S177	38001320	547	62.57	74.19	50.95	59.36	65.66	53.02		deC orginal		
2	ATA21D11	D2S1346	38241644	5492	63.10	75.25	50.95	59.36	65.66	53.02		deC orginal	ATA23F12	D2S1348
2	GATA190D04	no	38402558	5537	63.36	75.57	51.13	59.36	64.09	53.02		deC interol.		
2	ATA23F12	D2S1348	no	5492	no	no	no	59.36	65.66	53.02		no phys. pos.	ATA21D11	D2S1346
2	GATA194B06	no	no	no	no	no	no	61.66	70.24	53.19		no phys. pos.		
2	ATA47C04	no	no	no	no	no	no	67.58	77.51	57.65		no phys. pos.	ATA36B01	Unknown
2	ATA76C08	no	no	no	no	no	no	72.79	85.75	59.96		no phys. pos.		

Missing genetic positions for the deCODE map were interpolated (see Methods).

ment of linkage analyses of autism using data available in the public domain (data available from http://www.agre.org), we identified a marker (UT1307) that was initially thought to be on Chromosome 20 but has since been physically mapped to Chromosome 2 (C. Mathews, C. Nievergelt, and N. Schork, in prep.). Subsequent genetic analyses of markers on Chromosomes 2 and 20 showed that the marker in question, indeed, belonged to a linkage group on Chromosome 2. Although our reanalyses of linkages to autism on Chromosomes 2 and 20 did not result in evidence for linkage, the potential was there for initial reports to have not found evidence for linkage because the multipoint procedures they used were, in some sense, infected by a DNA marker locus whose alleles must have provided erroneous recombination patterns with neighboring marker allele patterns. We also identified some differences in interlocus distances between markers used in that map and interlocus distances computed by us.

One easy way potentially to verify that the genetic map used for a particular linkage analysis study is reliable is to assess the linkages between the markers empirically using the family data at hand (i.e., merely assess linkages among the markers using programs such as CRI-MAP or ASPEX rather than linkages between the markers and a potential trait-influencing locus). One can then compare the assumed map with that derived from the data. Although published genetic maps such as the Marshfield map with 188 meioses in three generation pedigrees and the deCODE map with 1257 meioses in two-generation families might be based on more meiotic events than those available in a given linkage study, such an exercise is well worth the effort as it can reveal other items in addition to incorrect map positions such as genotyping errors.

In pursuing empirical studies of map reliability with a given data set, it may be worthwhile to consider the use of sex-specific recombination rates. It is well known that males and females differ greatly in recombination rates (Broman et al. 1998; Kong et al. 2002). Although the construction of sex-specific maps may be less reliable than constructing a sex-averaged map, because the number of meioses in a given data set will be less (usually half) for either sex in isolation, the proper implementation of multipoint linkage analysis requires correct specification of recombination rates (Daw et al. 2000; see also Schork and Greenwood 2004). In this light, it is rather awkward that many of the most widely used programs for human linkage analysis such as GENEHUNTER (Kruglyak et al. 1996), SOLAR (Almasy and Blangero 1998), and MERLIN (Abecasis et al. 2002) do not accommodate sex-specific maps in their implementation of multipoint linkage analysis.

Discrepancies between genetic and physical maps will likely diminish as more and more polymorphic loci are mapped. In addition, further genetic linkage analyses of DNA markers (e.g., those used in genetic mapping studies) will provide greater con-

Table 4. Summary Information of the "Comprehensive Genetic Map"

Chromosome	Physical distance spanned (kb)	Marshfield			deCODE original and interpolated			Gaps where markers were interpolated (cM) and (kb)			
		No. of markers	Distance spanned (cM)	Average spacing (cM)	No. of markers	Distance spanned (cM)	Average spacing (cM)	Mean	Max.	Mean	Max.
1	245,963	687	289.66	0.42	667	276.40	0.42	0.87	4.97	969	23,665
2	243,243	657	269.07	0.41	635	266.79	0.42	1.03	4.23	1087	9342
3	199,130	549	231.27	0.42	548	221.83	0.40	0.96	5.03	859	3975
4	191,136	466	211.65	0.46	456	204.74	0.45	1.03	6.5	983	4185
5	180,458	493	197.54	0.40	479	205.69	0.43	1.13	7.47	1016	5204
6	170,695	480	193.14	0.40	465	189.60	0.41	1.07	6.43	961	6232
7	157,015	440	181.97	0.41	428	187.09	0.43	1.93	36.23	1042	4686
8	146,149	408	194.32	0.48	399	166.08	0.42	1.19	6.19	1117	5106
9	136,236	309	168.98	0.55	298	160.01	0.54	1.12	5.45	1054	5177
10	134,677	440	173.13	0.39	420	179.12	0.42	1.14	5.11	868	3677
11	133,704	404	147.77	0.37	398	152.45	0.38	0.95	7.1	841	4590
12	131,939	404	170.66	0.42	391	171.96	0.44	1.09	4.5	928	6007
13	112,931	255	114.98	0.45	250	129.52	0.52	1.25	5.05	896	2698
14	103,950	264	138.18	0.50	251	124.81	0.48	1.13	5.87	763	2183
15	100,070	213	122.14	0.58	203	133.61	0.66	1.55	8.29	1089	2972
16	89,242	276	134.12	0.49	266	130.00	0.49	1.37	7.3	1009	15,264
17	81,287	313	126.46	0.41	297	137.99	0.46	1.16	5.75	716	4534
18	75,963	231	126.00	0.55	218	121.65	0.56	1.17	4.12	830	4231
19	63,727	237	105.02	0.45	218	109.73	0.51	1.42	7.58	876	10,696
20	59,563	228	101.22	0.45	220	98.63	0.45	0.99	3.97	722	5060
21	46,894	114	61.40	0.54	114	78.48	0.66	1.84	7.42	829	1722
22	48,058	107	142.83	1.35	103	70.60	0.69	1.82	8.88	1101	3616
X	153,510	271	206.74	0.77	261	190.67	0.73	1.68	6.51	1352	5345
Total	3,005,542	8246	3808	0.51	7985	3707	0.49	1.26	7.39	953	6094

The distances spanned by the physical, Marshfield, and deCode map (including markers with interpolated recombination values), number of markers and mean marker spacing for the genetic maps, and mean and maximum gaps between original deCODE recombination rates where markers were interpolated are shown for all chromosomes.

fidence in marker order and interlocus distances. The genetic map that resulted from our own efforts encompasses physical/sequence information to interpolate positions of markers that are not available on either of the Marshfield or deCODE maps and as such is likely to be more comprehensive and accurate than either of them alone. Also, as the Haplotype Map Initiative unfolds, greater insight into recombination hot spots will emerge that can be further reconciled with different sorts of maps.

One final note on the issue of genetic and physical maps concerns their ubiquity. It is assumed that genetic maps capture recombination rates that are somewhat universal in that chromosomes are organized and recombined in roughly the same way among individuals. This assumed ubiquity is, in fact, what motivates researchers pursuing linkage analyses to use publicly available genetic maps in the first place. In addition, the potential ubiquity of the "block" structure of the human genome, which is motivating the human Haplotype Map Initiative, also assumes that, for example, genomic recombination hot spots, chromosomal sites for heavy gene conversion, and mutation hot spots are universal (see, e.g., Phillips et al. 2003; T. Greenwood, B. Rana, and N. Schork, in prep.). However, there is growing evidence that there exists very large variation in recombination rates among individuals during human meiosis (Cullen et al. 2002; Lynn et al. 2002). Because this individual variation is not really captured by simple specification of interlocus recombination fractions in genetic maps (which represent only "average" rates of recombination), it is an open question as to the degree to which such individual variation influences not only genetic map construction, but also the use of genetic maps in multipoint linkage analysis contexts.

METHODS

Integration of Genetic and Physical Map Information

To construct an "integrated map" including the newest physical positions for microsatellite markers used in the Marshfield (Broman et al. 1998; http://research.marshfieldclinic.org/genetics/) and deCODE maps (Kong et al. 2002), as well as sex-averaged and sex-specific recombination rates calculated in the two studies, we used information on physical positions from the University of California–Santa Cruz genome browser (July 2003 freeze; http://genome.ucsc.edu/). Specifically, we used the 7879 physically positioned STS markers from UCSC that had a Marshfield and/or deCODE genetic map position. To compare markers from UCSC with deCODE and Marshfield, UCSC names were replaced with marker and primer names (i.e., D-numbers) used by the genetic maps (by looking them up on the genome database, UniSTS, etc.). The UCSC map was then complemented with sex-specific recombination rates from the two genetic maps. Finally, 42 markers (corresponding to 63 entries) with multiple physical positions (i.e., UCSC score <1000) and 45 markers whose chromosomal position on the UCSC map did not correspond to the position assigned by the genetic maps (see Table 1) were excluded. The resulting unambiguous "integrated map" with 7770 markers was used for all the analyses presented in this paper.

Constructing a New "Comprehensive Genetic Map"

To construct a "comprehensive genetic map" based on deCODE's genetic positions, we first ordered the 7770 markers included in the integrated map according to their UCSC physical positions and compared the marker order with its order on the deCODE map. Then 65 markers with order problems were excluded. Second, we estimated sex-specific and sex-averaged genetic positions (i.e., estimated deCODE positions) for the 2677 markers not used

in the deCODE map through linear interpolation between physical positions and genetic positions of adjacent deCODE markers, using the 5028 deCODE markers with corresponding orders between the physical and genetic maps. For example, a marker at a physical position of 50,000 bp would be placed two-thirds between two markers with original deCODE positions of 2 cM at 10,000 bp and 5 cM at 70,000 bp, respectively, at an estimated genetic position of 4 cM. The use of linear interpolation of this sort is problematic for large interlocus distances, but not likely to be a huge problem for loci with small interlocus distances. Third, we included 88 markers with ambiguous physical positions (109 entries, see above) and 91 Marshfield and deCODE markers not placed on the UCSC July 2003 assembly, but whose physical positions could be found using BLAT (NCBI Entrez was used to find the relevant nucleotide sequence for a given STS, using annotation either from the nucleotide database or from the UniSTS database, and a FASTA-formatted version of the STS-containing sequence was then run at the BLAT facility of the UCSC Genome Center, to determine the chromosome and chromosome position of the marker in the UCSC coordinate system). Thus, 161 missing deCODE values of these markers were estimated as described above. Finally, 18 deCODE and 224 Marshfield markers without physical positions, as well as 187 cryptic duplicates from the Marshfield site, were added, resulting in a map with 8399 searchable marker entries. This comprehensive genetic map (see Table 3) is available as an Excel worksheet from the authors' Web site (http://elcapitan.ucsd.edu/hyper/), and includes a Macro for convenient Marker and Primer name look-up.

ACKNOWLEDGMENTS

This work was supported in part by the following NIH grants: the NHLBI Family Blood Pressure Program (FBPP; HL64777-01); the NHLBI hypertension SCOR program (HL54998); the NIH Pharmacogenetics Network (HL69758-01); and The Consortium on the Genetics of Schizophrenia (COGS; MH06557-01A1). The authors thank Tiffany Greenwood, John Kelsoe, and Daniel O'Connor for critical discussions of this work.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. 2002. MERLIN—Rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30:** 97–101.

Almasy, L. and Blangero, J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62:** 1198–1211.

Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, J.L. 1998. Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63:** 861–869.

Cullen, M., Perfetto, S.P., Klitz, W., Nelson, G., and Carrington, M. 2002. High-resolution patterns of meiotic recombination across the human major histocompatibility complex. *Am. J. Hum. Genet.* **71:** 759–776.

Daw, E.W., Thompson, E.A., and Wijsman, E.M. 2000. Bias in multipoint linkage analysis arising from map misspecification. *Genet. Epidemiol.* **19:** 366–380.

Dawson, E., Abecasis, G.R., Bumpstead, S., Chen, Y., Hunt, S., Beare, D.M., Pabial, J., Dibling, T., Tinsley, E., Kirby, S., et al. 2002. A first-generation linkage disequilibrium map of human chromosome 22. *Nature* **418:** 544–548.

DeWan, A.T., Parrado, A.R., Matise, T.C., and Leal, S.M. 2002. The map problem: A comparison of genetic and sequence-based physical maps. *Am. J. Hum. Genet.* **70:** 101–107.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. 2002. The structure of haplotype blocks in the human genome. *Science* **296:** 2225–2229.

Goring, H.H. and Terwilliger, J.D. 2000. Linkage analysis in the presence of errors III: Marker loci and their map as nuisance parameters. *Am. J. Hum. Genet.* **66:** 1298–1309.

Hackett, C.A. and Broadfoot, L.B. 2003. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90:** 33–38.

Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., et al. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* **31:** 241–247.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. 1996. Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.* **58:** 1347–1363.

Lander, E.S. and Weinberg, R.A. 2000. Genomics: Journey to the center of biology. *Science* **287:** 1777–1782.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Lynn, A., Koehler, K.E., Judis, L., Chan, E.R., Cherry, J.P., Schwartz, S., Seftel, A., Hunt, P.A., and Hassold, T.J. 2002. Covariation of synaptonemal complex length and mammalian meiotic exchange rates. *Science* **296:** 2222–2225.

Matise, T.C., Porter, C.J., Buyske, S., Cuttichia, A.J., Sulman, E.P., and White, P.S. 2002. Systematic evaluation of map quality: Human chromosome 22. *Am. J. Hum. Genet.* **70:** 1398–1410.

Ott, J. 1999. *Analysis of human genetic linkage.* The Johns Hopkins University Press, Baltimore, MD.

Phillips, M.S., Lawrence, R., Sachidanandam, R., Morris, A.P., Balding, D.J., Donaldson, M.A., Studebaker, J.F., Ankener, W.M., Alfisi, S.V., Kuo, F.S., et al. 2003. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nat. Genet.* **33:** 382–387.

Schork, N.J. and Greenwood, T.A. 2004. Inherent bias toward the null hypothesis in conventional multipoint nonparametric linkage analysis. *Am. J. Hum. Genet.* **74:** 306–316.

Tapper, W.J., Morton, N.E., Dunham, I., Ke, X., and Collins, A. 2001. A sequence-based integrated map of chromosome 22. *Genome Res.* **11:** 1290–1295.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291:** 1304–1351.

Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W., et al. 2001. Comparison of human genetic and sequence-based physical maps. *Nature* **409:** 951–953.

WEB SITE REFERENCES

http://cedar.genetics.soton.ac.uk/public_html/LDB2000; The New Genetic Location Database.

http://elcapitan.ucsd.edu/hyper,/; NIH hypertension PPG home page.

http://genome.ucsc.edu/; UCSC Genome Bioinformatics Home.

http://myscience.appliedbiosystems.com/; Celera Discovery System.

http://research.marshfieldclinic.org/genetics/; Welcome to the Center for Medical Genetics.

http://www.agre.org; Autism Genetic Resource Exchange.

Received April 28, 2003; accepted in revised form January 14, 2004.