Published in final edited form as:

Hum Hered. 2002 ; 54(1): 34–44. doi:10.1159/000066697.

Map Error Reduction: Using Genetic and Sequence-Based Physical Maps to Order Closely Linked Markers

Andrew T. DeWan^a, Antonio R. Parrado^a, Tara C. Matise^b, and Suzanne M. Leal^{a,c}

^aThe Laboratory of Statistical Genetics, The Rockefeller University, New York, N.Y.

^bDepartment of Genetics, Rutgers University, Piscataway, N.J.

^cDepartment of Molecular and Human Genetics, Baylor College of Medicine, Houston, Tex., USA

Abstract

The Marshfield comprehensive genetic maps are frequently used for linkage and association studies, however, for some regions of these maps the marker order has low level of likelihood ratio support. In order to investigate the level of statistical support and the accuracy of the genetic maps compared to sequence-based physical maps, two ~30 cM autosomal regions were selected. The first region was selected from chromosome 3 and consisted predominately of draft sequence. The second region was selected from chromosome 21 and consisted of finished sequence data. The physical order of these markers was based upon their position on Celera (CEL) and Human Genome Project-Santa Cruz (HGP-sc) sequence-based physical maps. The chromosome 3 and 21 regions contained 100 and 61 markers, respectively, on the Marshfield genetic map. The genetic and physical map order was consistent for 88.9 and 89.2% of the markers in the region on chromosome 3 and 21, respectively. Using a novel scoring criterion to assess inconsistent marker order between genetic and physical maps, it was determined that the physical order was likely the correct order for 3.3 and 7.1% of the markers in the chromosome 3 and 21 regions, respectively. To increase the accuracy of the order of markers selected for fine mapping a method is presented which combines information from genetic and sequence-based physical maps.

Keywords

Fine mapping; Comprehensive genetic maps; Sequence-based physical maps; Map errors

Introduction

Genetic maps play an important role in gene mapping. Errors in genetic maps can hinder the mapping and isolation of genes for complex and Mendelian traits [1–3]. The position of markers on commonly used genetic maps (e.g. Marshfield, CEPH, CHLC) are based upon a limited number of meioses [4] which can lead to incorrect order of marker loci and poor estimates of recombination fractions [5]. The problem of markers being incorrectly ordered is a particular problem for closely linked markers where there is a low order of support for

their relative map position [6]. This problem is most evident when genetic maps are used for fine mapping since closely spaced markers must be utilized. For linkage studies, inaccurate recombination fractions will lead to loss of power and can bias the map position for the trait locus [1, 7]. The effect of errors in genetic maps becomes increasingly serious as more marker loci are analyzed jointly [3].

In the near future the sequence-based order of markers will be known without error, however, this information will not elucidate the rates of recombination within specific regions. Recent efforts have been made to create integrated maps of chromosomes 21 [8] and 22 [8–10] which utilize information from various marker maps (i.e. sequence-based physical maps, genetic maps, radiation hybrid maps). This integration effort begins to answer more biological questions as to what influence recombination rates and will ultimately lead to more accurate genetic maps.

The majority of investigators choose their markers for fine mapping from published maps selecting those markers that permit the smallest genetic map distance between markers. The order of the markers which are selected for fine mapping are not always supported by a likelihood ratio criterion. A likelihood ratio of 3 is the standard criterion for generating framework genetic maps, however, the likelihood ratio support for comprehensive genetic maps can be much lower [6]. Selecting markers whose order is neither supported by a likelihood ratio criterion of 3, nor by the sequence-based physical map order can lead to misspecification of the order of markers in linkage analysis.

It was previously shown for Marshfield panels 9 and 10, which consist of genome scan markers spaced ~10 cM, that the genetic order for these markers not only had high statistical support but their order was also substantiated by the sequence-based physical maps [11]. This article investigates the frequency of errors in marker order for a commonly used comprehensive genetic map. The Marshfield genetic map [12] (<http://research.marshfield-clinic.org/genetics/>) is compared to two sequence-based physical maps Celera (CEL) [13] (<http://www.celera.com/>) and Human Genome Project-Santa Cruz (HGP-sc) [14] (<http://genome.cse.ucsc.edu/>). In addition, the level of likelihood ratio support is calculated for the Marshfield genetic map by comparing it to alternate orders. This comparison is made for two ~30 cM regions which were selected from a region that consists predominately of draft sequence (chromosome 3) and a second region that consists of finished sequence (chromosome 21). When there is an inaccuracy in the order of markers between genetic and physical maps, in most cases it is possible to elucidate whether the error occurs in the genetic or sequence-based physical map. To increase the accuracy of maps used for fine mapping, a method is demonstrated which combines information from genetic and sequence-based physical maps.

Methods

All short tandem repeat (STR) polymorphism markers in a ~30 cM region contained in the Marshfield Center for Medical Genetics genetic maps on chromosomes 3 and 21 were selected. Markers whose intermarker genetic distance was 0 cM were considered to be within the same *interval*. Markers within these *intervals* were reconfirmed to be completely

linked to each other (i.e. zero recombination events observed between these markers) using the web-based MAP-O-MAT program [15] (<http://compugen.rutgers.edu/mapomat/>) by comparing the Marshfield genetic order [12] to multiple alternative orders. For each *interval*, all alternative orders of the markers were equally likely. Since it is not usually possible to obtain a unique order for these markers through genetic mapping, these markers were ordered based upon their HGP-sc sequence-based physical map order.

In order to localize the markers that are within the two ~30 cM regions on chromosomes 3 (100 markers) and 21 (61 markers), a search was carried out using both marker and primer name(s) in the HGP-sc and CEL databases. The HGP-sc contains 2.9 Gb in 6,094 non-redundant sequences and had an assembly date of April 2001. 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.

In order to identify the physical position of additional markers not listed on the HGP-sc physical map, assembled sequence data was downloaded from the Santa Cruz database and electronic PCR (e-PCR) [13] was used to map the primers for these markers onto the assembled genome sequence by providing left and right PCR primers and expected product length. In order 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 of base pair mismatches ($n = 0, 1, 2$) and deviations in the product size ($m = 50, 1,000$ bp). The word length was set to its default value of 7. No additional chromosome 3 region markers and 4 (6.6%) of the chromosome 21 region markers could be placed on the sequence-based physical map using e-PCR. This procedure could not be performed on the CEL assembled genomic sequence as e-PCR is not available at this site and assembled sequence data could not be obtained.

It is not possible to make direct comparisons of the positions of the markers on the HGP-sc and CEL sequence-based physical maps, but only of their relative order. This is due to the way gaps in the assembled sequence data are treated. Both HGP-sc and CEL insert Ns to represent gaps. HGP-sc generally uses strings of 100 Ns to represent gaps, while CEL uses different lengths of Ns to represent the estimated size of the gaps [17].

When the Marshfield genetic map and the sequence-based physical map (CEL and HGP-sc) orders were inconsistent, the likelihood ratio was calculated by comparing the Marshfield genetic order [12] with the order obtained from the sequence-based physical maps. The likelihood ratios were obtained using the web-based MAP-O-MAT program [15]. MAP-O-MAT implements many features of the CRI-MAP program [18] using genotype data from its own database which is a compilation of marker genotypes for the Centre d'Etude du Polymorphisme Humain (CEPH) (<http://www.cephb.fr/cephdb/dumps.html>) pedigrees obtained from the Marshfield Center for Medical Genetics and the Foundation Jean Dausset CEPH databases.

In order to determine where errors occur and the most likely correct marker ordering, each marker was scored according to the criteria in table 1. The scoring criteria uses information

on position on physical and genetic maps and likelihood ratio support for a given genetic map position. It should be noted that the likelihood ratio was only calculated when the physical map(s) suggested an alternative marker ordering. When marker orders agreed between the physical and genetic maps, the likelihood ratio was not calculated and the marker was scored as genetic and physical maps consistent and therefore the order is probably correct. For each marker it was determined whether the genetic and physical maps were consistent (B), the genetic map was highly suggested (G), the physical map was highly suggested (P), inconclusive, but the genetic map was suggested (Ig), inconclusive, but the physical map was suggested (Ip), inconclusive (I) or position information was absent from both physical maps (A).

In order to select markers for fine mapping from the ~30 cM region on chromosome 3, a marker was selected from each *interval*. When there were multiple markers within an interval the following criteria was used to select one marker. First the position of the marker was corroborated by sequence-based physical position and secondly the marker in the *interval* with the highest heterozygosity obtained from GDB (<http://www.gdb.org/>) or MAP-O-MAT was selected. In order to verify the order of the markers, MAP-O-MAT was used to calculate the likelihood ratio by comparing the Marshfield map order [12] to all orders created by permuting marker orders in 4-tuples.

Results

The order of genetic markers in two ~30 cM regions was compared to their sequence-based physical map order. A marker scoring criteria was developed to elucidate where errors occur and establish the most likely correct marker order (table 1). For the 100 markers in the chromosome 3 region, 61 were localized on the CEL and 89 on the HGP-sc physical maps (table 2). Of these markers, 58 were found both on CEL and HGP-sc physical maps. The position of eight markers was inconsistent between the CEL physical map and the genetic map. All eight of these markers were localized on the HGP-sc physical map and the position of four of the markers was also inconsistent between the HGP-sc physical map and the genetic map. The HGP-sc physical map and the genetic map marker order were inconsistent for eight markers. Of these eight markers, six were also localized on the CEL physical map and four of these six markers were also inconsistent on the CEL physical map. For the eight inconsistent markers on the CEL physical map, four were supported by a likelihood ratio of 3 in favor of the genetic order. Likewise, the same was observed for HGP-sc, four of eight inconsistent markers were supported by a likelihood ratio of 3 in favor of the genetic order. There is one example where on both the CEL and HGP-sc physical maps the position for a marker (i.e. D3S3717) is different from that on the Marshfield genetic map, however, the genetic map order is supported by a likelihood ratio of 3 when compared to the sequence-based physical map order. Likelihood ratio data for the region on chromosome 3 can be found in table 3 along with primer name, map distances, physical map position on CEL and HGP-sc, score and heterozygosity.

For the 61 markers in the chromosome 21 region, 47 were localized on the CEL physical map, 60 on the HGP-sc physical map and 46 on both physical maps (table 2). Six and eight markers were inconsistent between the genetic map and the CEL and HGP-sc physical maps,

respectively. All of the six inconsistent CEL markers were localized on the HGP-sc physical map and five of these markers were also inconsistent with the HGP-sc physical map. The one marker (D21S1254) that was only inconsistent on the CEL physical map had a genetic position which was supported by a likelihood ratio of 3. Of the eight inconsistent HGP-sc markers, five were localized on the CEL physical map and all five were also inconsistent on the CEL physical map. One marker (D21S1245) that was localized only on the HGP-sc physical map had a likelihood ratio of 3 in support of the genetic order. Likelihood ratio data for the region on chromosome 21 can be found in table 4 along with primer name, map distances, physical map position on CEL and HGP-sc, score and heterozygosity.

Each marker was scored according to the criteria in table 1 and the results are summarized in table 5. Of the 100 chromosome 3 markers, 80 had a genetic and physical order which agreed between all maps. Eight markers were unable to be scored because their physical position was absent on both CEL and HGP-sc maps. Of the remaining 12 markers, five (41.7%) had an order highly suggested by the genetic map, while two (16.7%) had an order highly suggested by the physical map(s). Two of 12 (16.7%) markers were inconclusive but suggestive of the genetic map while 1/12 (8.3%) markers was inclusive but suggestive of the physical map. For two markers (16.7%) it was not possible to suggest whether the inconsistency was due to an error in the genetic or physical map and therefore, these markers were scored as being inconclusive. Similar results were found for chromosome 21 and can also be found in table 5.

To determine the percentage of markers where the genetic and physical maps were consistent, those markers which were scored inconclusive (I) or absent from both physical maps (A) were removed from the total number of markers. The number of markers where the genetic map was consistent (B) was compared to all scored categories (B, G, P, Ig, Ip) except inconclusive (I) and absent from both physical maps (A). It was determined that 88.9% (95% CI 80.5%, 94.5%) and 89.2% (95% CI 78.1%, 96.0%) of the markers in the region on chromosome 3 and 21, respectively, had positions that were consistent between the genetic and physical maps.

To determine the percentage of markers which may be incorrectly ordered on the genetic map, the total number of markers which could be scored as physical map highly suggested (P) or inconclusive but physical map suggested (Ip) was compared to all scored categories (B, G, P, Ig, Ip) except inconclusive (I) and absent from both physical maps (A). It was found that 3.3% (95% CI 0.6%, 9.4%) of the markers in the chromosome 3 region may be incorrectly ordered on the genetic map, while 7.1% (95% CI 2.0%, 17.2%) of the markers in the chromosome 21 region may be incorrectly ordered on the genetic map.

Fine Mapping

On chromosome 3 there were a total of 28 *intervals* (see Methods for description) within the ~30 cM region (table 6). A total of 28 markers were selected, one from each *interval*, with an average distance of 1.17 cM. The relative position of 17 of the markers was supported by a likelihood ratio of 3. The position of 12 markers was absent on the CEL physical map, one marker could not be localized on the HGP-sc physical map and none of the markers were absent on both CEL and HGP-sc physical maps. The genetic order of all markers was

confirmed by sequence-based physical map order on CEL, HGP-sc or both. One marker had a sequence position on CEL that was inconsistent with the genetic map, but was consistent between HGP-sc and the genetic map.

Discussion

The order of the markers on the genetic maps and the sequence-based physical maps for the most part are consistent, although a few inconsistencies were found between the genetic and physical maps. Genetic markers can be incorrectly ordered due to genotyping errors [19] and due to the limited number of meioses used to build the maps [4]. In physical maps errors can occur during assembly, *in-silico* mapping and annotation.

For eight markers that were inconsistent with one or both of the physical maps it was demonstrated that the order given by the Marshfield genetic map was not the order with the highest level of statistical support. This can occur when constructing genetic maps using a heuristic approach such as the one implemented in CRI-MAP [18], because the likelihood is not computed and compared for all possible marker orders. For these regions the best order was then considered to be the one supported by the sequence-based physical maps and scored as either physical map highly suggested (P) or inclusive, physical map suggested (Ip).

There is one example where the Marshfield genetic marker order for a marker position (D3S3717) was supported by a likelihood ratio of 5.4 when compared to the marker order given by both CEL and HGP-sc physical maps. In this circumstance it is difficult to conclude whether it is the genetic or sequence-based physical maps that are incorrect.

It should be noted that for the genetic maps it is usually not possible to distinguish the order of markers that have no recombination events between them. However, recombination events with other more distant markers may help to suggest a relative order for the completely linked markers, but the level of support for this order over transposed orders will be low.

The estimate of the map distances between markers on the genetic maps may be inaccurate due to the limited number of meioses used to generate the Marshfield maps (i.e. 184 for the majority of markers) [4]. The number of informative meioses for two markers will depend on the heterozygosity of these markers and on the number of individuals genotyped. For the chromosome 3 and 21 regions, based upon the Marshfield genotype data, the average number of informative meioses for adjacent marker pairs is 84.9 (95% CI 78.4, 91.4) and 68.2 (95% CI 58.9, 77.5), respectively.

Fine Mapping

Although for the most part the order of the markers on the genetic maps was correct, there were some instances where it was shown that the genetic order of the markers was incorrect or it was not possible to substantiate whether or not the genetic map order was correct. Therefore, it is suggested that information from genetic maps and physical maps can be combined to increase the prospect that the correct order of the markers is determined for fine mapping. An example of how this can be done is given for the ~30 cM region on chromosome 3 (table 6). Markers are selected in order to obtain the smallest non-zero

genetic distance between markers. Since multiple markers often map to the same *interval* (i.e. no recombination between a group of markers), a marker is selected from an interval based upon: (1) if its sequence-based physical map position agrees with the genetic map; (2) the heterozygosity of the marker; and (3) its position is supported by a likelihood ratio of >3 . In most instances, the sequence-based physical position will agree with the genetic map for all markers within an *interval*. If this is not the case, any markers whose position does not agree with the sequence-based physical map, should not be considered for inclusion on the fine map. E-PCR can be used to try to localize markers whose position is absent on sequenced-based physical maps. To rank the markers which are considered for selection within an *interval*, the highest priority is marker order agreement between genetic and physical maps and the second priority is highest level of marker heterozygosity. Once the markers have initially been selected for fine mapping, the likelihood ratio order of support for the genetic maps is calculated by comparing the published order with alternative orders. While it is usually not possible to obtain a likelihood ratio of >3 for the order of the markers for fine maps, reaching this level of support is critical when there is no sequence-based physical map information for a marker which will be incorporated into the fine map. If these criteria are not met, one or more markers should be substituted with other marker(s) from the same *interval*. In addition, marker substitutions may be made to aid in multiplexing panels.

For the region on chromosome 3, it was not always possible to select those markers which were both in the sequence-based physical map and had the highest heterozygosity, since the physical order of these markers was not always consistent with the best order suggested by the likelihood ratio. For example for marker D3S2321 (Het 0.94), its position on the genetic and physical map was consistent but this order had a likelihood ratio of -0.00003 , thus demonstrating that an alternative genetic map order had a slightly higher level of support (data not shown). Therefore D3S2321 was substituted with marker D3S1578 (Het 0.88) whose likelihood ratio support for its position on the genetic and physical map was >3 . Although marker D3S3522 was inconsistent with the CEL physical map, it was still selected for fine mapping, because the genetic order was supported by a likelihood ratio of >3 and the genetic map was consistent with the HGP-sc physical map. The other marker, D3S3658, in this *interval* was also inconsistent with the CEL physical map and although it was consistent with the HGP-sc physical map, it had a lower level of likelihood ratio support in addition to having a lower heterozygosity. For markers D3S3717 and D3S3588, the likelihood ratio support for their order is 0.0. Marker D3S3717 is the only marker within its *interval*. Marker D3S3588 has the highest heterozygosity and the other two markers within the same interval also have a likelihood ratio of 0.0 for this genetic map position. Although the likelihood ratio is 0.0 for markers D3S3717 and D3S3588, the CEL and HGP-sc physical maps confirm their genetic map position.

Once fine mapping has been completed, it may be beneficial to further refine the region for a disease locus by genotyping all available markers whose position is substantiated by the physical map within the disease interval (i.e. determined by haplotypes or support intervals). Even in the case where this strategy cannot reduce the genetic map region to which a disease locus maps, it can help to refine the physical region [20, 21]. For linkage disequilibrium mapping, all markers in a region whose order has been substantiated by sequenced-based physical maps can be utilized, whether or not a genetic distance separates the markers.

While it is recognized that the sequence-based physical maps continue to be updated and eventually the order of genetic markers will be known without error, there is still utility in comparing genetic and sequence-based physical maps at this stage. We have described a method that can be employed for evaluating the order of markers based on both sequence-based physical and genetic maps. This method will continue to be valuable to researchers as it is a straight-forward tool for integrating information from genetic and sequence-based physical maps which is necessary to generate dense maps of markers for fine-mapping trait loci.

The genetic and sequenced-based physical maps have a good agreement. There are regions where the genetic maps have low levels of statistical support, but in most instances the relative order was consistent with sequence-based physical maps. In the case where there was no support for a given marker (markers completely linked), the sequence-based physical maps provided information which was beneficial to order these markers. In addition, sequence-based physical maps are useful to verify the order of markers on genetic maps. This is especially true for closely spaced markers, such as those used in fine mapping.

Acknowledgements

This work is supported by National Institutes of Health grants DC03594, HG01691, HG00008 and MH59492.

References

1. Risch N, Giuffra L: Model misspecification and multipoint linkage analysis. *Hum Hered* 1992;42:77–92. [PubMed: 1555848]
2. Feakes R, Sawcer S, Chataway J, Coraddu F, Broadley S, Gray J, Jones HB, Clayton D, Goodfellow PN, Compston A: Exploring the dense mapping of a region of potential linkage in complex disease: An example in multiple sclerosis. *Genet Epidemiol* 1999;17:51–63. [PubMed: 10323184]
3. Göring HH, Terwilliger JD: Linkage analysis in the presence of errors III: Marker loci and their map as nuisance parameters. *Am J Hum Genet* 2000;66:1298–1309. [PubMed: 10731467]
4. Yu A, Zhao C, Fan Y, Jang W, Mungall AJ, Deloukas P, Olsen A, Doggett NA, Ghebranious N, Broman KW, Weber JL: Comparison of human genetic and sequence-based physical maps. *Nature* 2001;409:951–953. [PubMed: 11237020]
5. Terwilliger JD, Ding Y, Ott J: On the relative importance of marker heterozygosity and intermarker distance in gene mapping. *Genomics* 1992;13:951–956. [PubMed: 1505986]
6. Keats BJB, Sherman SL, Morton NE, Robson EB, Buetow KH, Cartwright PE, Chakravarti A, Francke U, Green PP, Ott J: Guidelines for human linkage maps: An international system for human linkage maps (ISLM 1990). *Genomics* 1991;9:557–560. [PubMed: 2032725]
7. Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J: Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 1986;42:393–399. [PubMed: 3741977]
8. Ke X, Tapper W, Collins A: LDB2000: sequence-based integrated maps of the human genome. *Bioinformatics* 2001;17:581–586. [PubMed: 11448876]
9. Tapper WJ, Morton NE, Dunham I, Ke X, Collins A: A sequence-based integrated map of chromosome 22. *Genome Res* 2001;11:1290–1295. [PubMed: 11435412]
10. Matisse TC, Porter CJ, Buyske S, Cuttichia AJ, Sulman EP, White PS: Systematic evaluation of map quality: Human chromosome 22. *Am J Hum Genet* 2002;70(6):1398–1410. [PubMed: 11992248]
11. DeWan AT, Parrado AR, Matisse TC, Leal SM: The map problem: A comparison of genetic and sequence-based physical maps. *Am J Hum Genet* 2002;70:101–107. [PubMed: 11706388]
12. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL: Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am J Hum Genet* 1998;63: 861–869. [PubMed: 9718341]

13. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, et al.: The sequence of the human genome. *Science* 2001; 291:1304–1351. [PubMed: 11181995]
14. International Human Genome Sequence Consortium: Initial sequence and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]
15. Matisse TC, Gitlin JA: MAP-O-MAT: Marker-based linkage mapping on the World Wide Web. *Am J Hum Genet* 1999;65:A2464.
16. Schuler GD: Electronic PCR: Bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol* 1998; 16:456–459. [PubMed: 9830153]
17. Aach J, Bulyk ML, Church GM, Comanders J, Derti A, Shendure J: Computational comparison of two draft sequences of the human genome. *Nature* 2001;409:856–859. [PubMed: 11237010]
18. Lander ES, Green P: Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 1987;84:2363–2367. [PubMed: 3470801]
19. Buetow KH: Influence of aberrant observations on high-resolution linkage analysis outcomes. *Am J Hum Genet* 1991;49:985–994. [PubMed: 1928104]
20. Kaplan N, Morris R: Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. *Genet Epidemiol* 2001;20:432–457. [PubMed: 11319784]
21. Rannala B, Reeve JP: High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* 2001;69:159–178. [PubMed: 11410841]

Table 1.

Scoring Criteria used to compare the order of genetic and physical maps

Score	Genetic and CEL consistent	Genetic and HGP-sc consistent	Genetic and CEL inconsistent	Genetic and HGP-sc inconsistent	LR 3^I	$0 < LR < 3^I$	LR = 0 ^I	LR < 0 ^I	Absent in CEL	Absent in HGP-sc
Genetic and Physical maps consistent ²										
B	×	×			-	-	-	-		
B	×				-	-	-	-		×
B		×			-	-	-	-	×	
Genetic map highly suggested										
G	×		×	×	×					
G		×	×	×	×					
Physical map highly suggested										
P			×	×			×	×		
Inconclusive, genetic map suggested										
Ig	×			×		×				
Ig		×	×	×		×				
Ig			×		×					×
Ig				×	×				×	
Inconclusive, physical map suggested										
Ip			×				×	×		×
Ip				×			×	×	×	
Inconclusive										
I	×			×				×		
I	×			×			×			
I			×			×				×
I				×		×			×	
I		×	×				×			
I		×	×	×			×			
I			×	×	×					
I			×	×		×				
I			×	×			×			×

Score	Genetic and CEL consistent	Genetic and HGP-sc consistent	Genetic and CEL inconsistent	Genetic and HGP-sc inconsistent	LR	3^I	$0 < LR < 3^I$	LR = 0^I	LR < 0^I	Absent in CEL	Absent in HGP-sc
I				X				X		X	
I			X	X				X			
Absent from both physical databases											
A										X	X

¹ Likelihood ratio calculated only when one (or both) physical position(s) was inconsistent with the genetic map. Likelihood ratio (LR) is calculated for the genetic position of marker compared to the sequence-based physical position of the marker. The case where $LR < 3$ does not occur in these datasets.

² Likelihood ratios are not calculated when the genetic and physical maps are consistent.

Table 2.

Summary of inconsistent markers

Chrom.	Markers	Markers located in database		Markers in CEL and HGP-sc	Markers inconsistent with genetic map		Inconsistent markers supported by LR		Inconsistent markers supported by LR	
		CEL	HGP-sc		CEL	HGP-sc	CEL	HGP-sc	CEL	HGP-sc
3	100	61	89(0) ²	58	8(4/8) ³	8(4/6) ³	4(1/4) ³	4(1/3) ³	4(3/4) ³	4(3/3) ³
21	61	47	60(4) ²	46	6(5/6) ³	10(5/6) ³	1(0/1) ³	1(0/0) ³	5(5/5) ³	9(5/6) ³

¹Markers in which their sequence-based physical map order disagrees with their genetic map order.

²Number in parentheses indicates number of markers physically mapped using e-PCR.

³Denominator = Number of markers inconsistent with genetic map localized in both physical maps; Numerator = number of markers from the denominator which were inconsistent with both physical maps.

Table 3.

All markers from Marshfield in chromosome 3 region

Locus	Primer	Map Pos. (cM) ¹	Dist btwn (cM)	CEL	HGP-sc	LR ²	Score ³	Het ⁴
D3S1293	AFM200za1	44.81	0	absent	27602015		B	0.80
D3S3038	GATA73D01	44.81	0	24065173	27624405		B	0.78
D3S1599	AFM301ze9	46.13	1.32	absent	28260067		B	0.38
D3S3659	AFMb331ye9	47.44	1.31	25046039	28706246		B	0.65
D3S3700	AFM328yc9	48.09	0.65	absent	29669182 ⁵	-0.0006	Ip	0.77
D3S3598	AFMa346xa9	49.18	1.09	25928410	29560884		B	0.57
D3S1567	AFM234ya11	49.18	0	26919639	30310315		B	0.48
D3S2336	AFM336zb5	49.18	0	absent	30613351		B	0.78
D3S2307	UT6471	49.18	0	absent	30805631		B	0.56
D3S2317	UT6375	49.18	0	absent	30821052		B	0.60
D3S2301	Mfd342	49.18	0	absent	30824658		B	0.65
D3S1583	AFM276vc9	49.18	0	absent	31230250		B	0.39
D3S2327	UT7684	49.18	0	absent	31248001		B	0.75
D3S2466	GGAA22H08	50.25	0	28350767	31687319		B	0.85
D3S2337	AFM015xd7	50.25	1.07	23514306 ⁵	32099804	12.5	G	0.72
D3S2335	AFM312yf5	50.94	0.69	28806987	32100125		B	0.64
D3S1266	AFM095xc1	52.6	1.66	30310680	33526796		B	0.74
D3S1283	AFM183yc3	52.6	0	31077374	34599152		B	0.71
D3S1609	AFM320wc9	55.11	2.51	32293044	35844204		B	0.64
D3S3547	AFMa108yc1	55.11	0	32508293	36077344		B	0.74
Unknown	GATA8A07	55.11	0	absent	absent		A	0.72
D3S3727	AFMa090zb1	56.07	0.96	absent	36590181		B	0.80
D3S3567	AFMa176wc1	56.69	0.62	33088568	36651351		B	0.70
D3S2432	GATA27C08	57.92	1.23	35428715	37815575		B	0.86
D3S1211	D3S1211	57.92	0	absent	absent		A	0.88
D3S1619	AFM350tf1	60.98	3.06	36910081	39351382		B	0.75
Unknown	GATA112E02	60.98	0	absent	absent		A	0.65
D3S3518	AFM331wh9	60.98	0	absent	absent		A	0.78
D3S1612	AFM339xh1	61.52	0.54	37397229	39763175		B	0.69
D3S3512	AFM207wb4	61.52	0	37401850	39767804		B	0.67
D3S1768	GATA8B05	61.52	0	37432076	39797981		B	0.76
D3S1277	AFM164we1	61.52	0	absent	39829404		B	0.82
D3S3718	AFMa070yg5	61.52	0	38968784	41359262		B	0.72
D3S1561	AFM220xg3	61.52	0	absent	41692941		B	0.65
D3S1611	AFM338xe5	61.52	0	39866779	42326339		B	0.66
D3S2417	ATA22D03	61.52	0	40230768	43035362		B	0.68
D3S3623	AFMb286yb1	61.52	0	40239973	43044481		B	0.76

Locus	Primer	Map Pos. (cM) ¹	Dist btwn (cM)	CEL	HGP-sc	LR ²	Score ³	Het ⁴
D3S2411	ATA15H02	62.05	0.53	39157919 ⁵	41548221 ⁵	-0.0004	P	0.38
D3S1298	AFM220yb8	62.05	0	absent	43649892		B	0.88
D3S3639	AFMb314yh1	63.12	1.07	absent	43989165		B	0.63
D3S1260	AFM038xc1	63.12	0	41208761	44001215		B	0.66
D3S3605	AFMb020yd9	63.12	0	42040125	44339455		B	0.73
D3S3593	AFMa312ze9	63.12	0	42075907	absent		B	0.73
D3S3527	AFMa073yf5	63.12	0	42271240	absent		B	0.85
D3S3572	AFMa190yf5	63.12	0	absent	absent		A	0.75
D3S3521	AFM345yg5	63.12	0	absent	absent		A	0.83
D3S3522	AFM349ta5	65.26	2.14	47250417 ⁵	45829520	4.63	G	0.58
D3S3658	AFMb330ze5	65.26	0	47363250 ⁵	45941878	3.17	G	0.56
D3S2407	ACT2E06	67.94	2.68	absent	46496315		B	0.74
D3S2319	UT6506	67.94	0	absent	47536257		B	0.50
D3S3564	AFMa162wd9	67.94	0	45006574	47564920		B	0.80
D3S3685	AFMc006yh1	67.94	0	45056735	47615089		B	0.89
D3S3687	AFMc016yd9	67.94	0	45207568	47766045		B	0.40
D3S3559	AFMa139wf9	67.94	0	45276358	47834751		B	0.85
D3S2304	UT5662	67.94	0	45402990	47908033		B	0.68
D3S3563	AFMa154wb9	68.47	0.53	44829581 ⁵	47402791 ⁵	0.04	I	0.76
D3S3678	AFMb362yc1	68.47	0	46010664	48640801		B	0.75
D3S3647	AFMb320vc1	68.47	0	46160381	48774965		B	0.73
D3S3597	AFMa344zg5	68.47	0	absent	49243464		B	0.74
D3S3624	AFMb291zc9	68.47	0	47963511	50064882		B	0.73
D3S3582	AFMa231xe9	69.19	0.72	absent	50789014		B	0.69
D3S1767	GATA7A01	69.9	0.71	48749225	52329896		B	0.76
D3S3640	AFMb315ze9	70.61	0.71	51177968	19163897 ⁵	3.08	G	0.65
D3S3729	AFMa100ze1	70.61	0	51219341	19229847 ⁵	3.08	G	0.45
D3S2420	ATA25A07	70.61	0	absent	19246148 ⁵	3.08	Ig	0.77
D3S1581	AFM273ve9	70.61	0	absent	53207022		B	0.88
D3S3560	AFMa141zc5	70.61	0	absent	53659754		B	0.66
D3S2409	ATA10H11	70.61	0	absent	54192690		B	0.75
D3S3629	AFMb296zc5	70.61	0	53254065	54761522		B	0.79
D3S2456	GATA63E04	70.61	0	absent	54938724		B	0.75
D3S3667	AFMb348za9	70.61	0	absent	55463559		B	0.69
D3S1568	AFM238yc5	70.61	0	53924139 ⁶	55911856		B	0.87
D3S1621	AFMa133wh1	70.61	0	53836240	56000142		B	0.73
D3S3615	AFMb049zf9	70.61	0	absent	56045550		B	0.63
D3S1235	Mfd93	70.61	0	54705315	absent		B	0.44

Locus	Primer	Map Pos. (cM) ¹	Dist btwn (cM)	CEL	HGP-sc	LR ²	Score ³	Het ⁴
D3S3026	GAAT11D03	70.61	0	55547842	57055862		B	0.71
D3S3688	AFMc017wg9	70.61	0	absent	57090052		B	0.61
D3S3561	AFMa142yf5	70.61	0	55984112	57486920		B	0.74
D3S3648	AFM238xb10	70.61	0	56015975	57518831		B	0.25
D3S2321	UT6992	70.61	0	56726726	58293363		B	0.94
D3S1578	AFM268wg9	70.61	0	absent	59028515		B	0.88
D3S1588	AFM287yd9	70.61	0	57720531	59389992		B	0.74
D3S3604	AFMb008yh5	70.61	0	absent	absent		A	0.66
D3S1573	AFM126xg7	70.61	0	absent	absent		A	0.79
D3S1289	AFM198yf2	71.41	0.8	58098072	59706713		B	0.82
D3S3672	AFMb352xe1	72.21	0.8	58323953 ⁶	59857625		B	0.67
D3S3666	AFMb347zg9	72.21	0	absent	59889997		B	0.85
D3S1582	AFM274yd5	72.21	0	58303689	59933430		B	0.80
D3S3660	AFMb338zg9	72.21	0	absent	60165621		B	0.73
D3S1613	AFM340xf1	72.21	0	58554230	60198283		B	0.83
D3S3719	AFMa071yd1	72.21	0	59045029	60756293		B	0.80
D3S1606	AFM318wc5	74.35	2.14	absent	60965526		B	0.79
D3S2408	ATA10D01	74.35	0	59304647	61031477		B	0.69
D3S3717	AFMa066wd5	74.35	0	58775661 ⁵	60476808 ⁵	5.43	I	0.78
D3S3588	AFMa275zh1	74.88	0.53	59672322	61428438		B	0.65
D3S3721	AFMa072yb9	75.41	0.53	59474379 ⁵	61283531 ⁵	-0.0008	P	0.71
D3S3048	GATA87B02	75.41	0	59724068	61480238		B	0.62
D3S3621	AFMb283ya5	75.41	0	absent	61665377		B	0.77
D3S3724	AFMa082wh1	75.41	0	61096939 ⁵	61721079	2.56	Ig	0.72
D3S3616	AFMb059wf1	76.48	1.07	60908605	62673287		B	0.85

¹Sex-average Kosambi cM map position for males and females from the Marshfield genetic map.

²Likelihood ratio (LR) for genetic position of marker compared to the sequence-based physical position of the marker. Positive LR indicates genetic order is favored, negative LR supports sequence-based physical order.

³Scoring criteria (see table 1 and methods for description of scoring).

⁴Heterozygosity: Average heterozygosity for chromosome 3 region markers 0.71 (95% CI 0.68, 0.73).

⁵Physical order does not agree with the genetic order.

⁶CEL physical order does not agree with HGP-sc physical order, but both CEL and HGP-sc independently agree with the genetic order (the markers map to the same *interval*). Note: markers with zero recombination between them and no order had greater statistical support were ordered according to HGP-sc.

Table 4.

All markers from Marshfield in chromosome 21 region

Locus	Primer	Map pos. (cM) ¹	Dist btwn (cM)	CEL	HGP-sc	LR ²	Score ³	Het ⁴
D21S1901	AFMb341wd9	27.4	0	16018402	27704291		B	0.52
Unknown	GATA148F04	27.4	0	16228325	absent		B	0.67
D21S1270	GATA8G04	27.4	0	16878224	28564740		B	0.86
D21S263	AFM211zg9	27.4	0	17394090	29079907		B	0.75
D21S1908	AFMc016yg9	27.4	0	17415232	29100929		B	0.74
D21S1239	UT489	28.48	1.08	17133744 ⁵	28820107 ⁵	0.83	I	0.67
D21S1909	AFM288yg5	28.48	0	17705529	29391224		B	0.86
D21S1280	UT706	28.48	0	absent	29440664		B	0.86
Unknown	ATA42C09	28.48	0	17906774	29592534		B	0.75
D21S1245	UT762	28.48	0	absent	39268147 ⁵	6.05	Ig	0.96
D21S261	AFM193xf10	29.05	0.57	18455703	30116607		B	0.51
D21S262	AFM198tc5	29.4	0.35	absent	30674114		B	0.67
D21S1413	UT7582	29.78	0.38	absent	30705823		B	0.88
D21S1888	AFMa218wb1	31.26	1.48	18024293 ⁵	29712260 ⁵	-0.0058	P	0.86
D21S1898	AFMb308xe5	31.26	0	19771545	31467138		B	0.64
Unknown	IFNAR	31.26	0	absent	31570792		B	0.75
D21S1910	AFM289xh1	31.26	0	20122200	31807663		B	0.94
D21S1254	AFM276za5	31.26	0	22742542 ⁵	31933259	7.79	G	0.69
Unknown	GATA45C03	31.26	0	20569721	32233575		B	0.69
D21S1920	AFMa086yf9	31.99	0.73	20890951	32555049		B	0.74
D21S1895	AFMb280xd9	33.84	1.85	21544582	33208934		B	0.82
D21S1921	AFMa099wc9	33.84	0	21615735	33280018		B	0.81
D21S1894	AFMb277xc9	35.45	1.61	21848169	33510135		B	0.64
D21S1252	AFM261zg1	35.45	0	23359418 ⁵	34684591 ⁵	0.00	I	0.80
D21S211	Mfd163	36.77	1.32	22229591	33884352		B	0.53
D21S1440	ATA27F01	36.77	0	24677470 ⁵	35999265 ⁵	0.00013	I	0.74
D21S270	AFM031xc5	38.08	1.31	absent	35663564 ⁵	0.0032	I	0.86
D21S167	D21S167	38.65	0.57	absent	35053421		B	0.82
D21S267	AFM238wc3	38.65	0	23999211	35325119		B	0.88
D21S1900	AFMb335ze5	38.65	0	24005334	35331279		B	0.63
D21S1919	AFMa085za5	38.65	0	24096580	35422472		B	0.82
D21S1238	UT409	38.65	0	24799430	36121190		B	0.67
D21S1917	AFMa066yb1	38.65	0	24878383	36147271		B	0.63
Unknown	GATA29D01	38.65	0	24895597	36217382		B	0.82
D21S259	AFM016xe5	38.65	0	24984890	36306719		B	0.80
D21S1883	AFMa114yb9	38.65	0	25034216	36356487		B	0.73

Locus	Primer	Map pos. (cM) ¹	Dist btwn (cM)	CEL	HGP-sc	LR ²	Score ³	Het ⁴
D21S1408	UT5040	38.65	0	26586561 ⁵	37906144 ⁵	-0.53	P	0.54
D21S1255	AFM283xb9	39.22	0.57	25330029	36652523		B	0.80
D21S1891	AFMb024yh5	39.22	0	25956217	37279103		B	0.75
D21S1246	UT973	40.43	1.21	absent	37725944 ⁶		B	0.75
D21S168	D21S168	40.45	0.02	absent	37799608 ⁶		B	0.76
D21S1412	UT6930	40.47	0.02	absent	37605125 ^{5,6}	-0.0052	Ip	0.80
D21S268	AFM260ze9	40.49	0.02	26115382	37436973 ⁵	-0.0052	I	0.88
D21S2055	GATA188F04	40.49	0	absent	38045648		B	0.88
D21S1893	AFM206zc7	43.67	3.18	26891498	38210959		B	0.76
D21S1889	AFMa230yd1	45.26	1.59	27360199	38679373		B	0.83
D21S231	D21S231	45.55	0.29	absent	38865912 ⁶		B	0.74
D21S1906	AFMc013wd1	45.87	0.32	27663608	38983274		B	0.61
D21S1887	AFMa137yf5	45.87	0	27749710	39069857		B	0.62
D21S266	AFM234xg9	45.87	0	28218750	39539111		B	0.60
D21S1260	AFMa152wd1	46.71	0.84	28330918	39650648		B	0.74
D21S1411	UT1355	51.49	4.78	29658559	40989267		B	0.93
D21S112	D21S112	51.49	0	absent	44115802 ⁵	-0.62	Ip	0.95
D21S1890	AFMb008ze5	52.5	1.01	30313775	41638779		B	0.88
D21S1885	AFMa133xc9	52.5	0	30325725	41650685		B	0.57
D21S1259	AFM326tf1	52.5	0	30798321	42114523		B	0.67
D21S1912	AFM071xa1	53.05	0.55	31051830	42368416		B	0.80
D21S1903	AFMb347zc1	53.64	0.59	31591937	42931710		B	0.83
D21S1897	AFMb298xd5	54.22	0.58	absent	43322776		B	0.74
D21S2058	21QTEL14	57.77	3.55	33403177	44755784		B	0.30
D21S1446	GATA70B08	57.77	0	33479005	44825305		B	0.68

¹Sex-average Kosambi cM map position for males and females from the Marshfield genetic map.

²Likelihood ratio (LR) for genetic position of marker compared to the sequence-based physical position of the marker. Positive LR indicates genetic order is favored, negative LR supports sequence-based physical order.

³Scoring criteria (see table 1 and methods for description of scoring).

⁴Heterozygosity: Average heterozygosity for chromosome 21 region markers 0.74 (95% CI 0.71, 0.78).

⁵Physical order does not agree with the genetic order.

⁶Marker assigned a physical position using e-PCR.

Table 5.

Summary of scored markers

Chrom. No.	Score ^I							Total
	B	G	P	Ig	Ip	I	A	
3	80 (80.0%)	5 (5.0%)	2 (2.0%)	2 (2.0%)	1 (1.0%)	2 (2.0%)	8 (8.0%)	100
21	50 (82.0%)	1 (1.6%)	2 (3.3%)	1 (1.6%)	2 (3.3%)	5 (8.2%)	0 (0.0%)	61

^I B = Genetic *and* physical maps consistent; G = genetic map highly suggested; P = physical map highly suggested; Ig = inconclusive, genetic map suggested; Ip = inconclusive, physical map suggested; I = inconclusive; A = absent from both physical databases. See table 1 and methods for additional description of scoring.

Table 6.

Selected fine-mapping markers in chromosome 3 region

Locus	Primer	Map Dist ¹	CEL	HGP-sc	LR ²	Het
D3S1293	AFM200za1	44.81	absent	27602015		0.80
D3S1599	AFM301ze9	46.13	absent	28260067	2.89 ³	0.38
D3S3659	AFMb331ye9	47.44	25046039	28706246	2.89 ³	0.65
D3S3700	AFM328yc9	48.09	absent	29669182	2.91 ³	0.77
D3S2336	AFM336zb5	49.18	absent	30613351		0.78
D3S2466	GGAA22H08	50.25	28350767	31687319		0.85
D3S2335	AFM312yf5	50.94	28806987	32100125		0.64
D3S1266	AFM095xc1	52.6	30310680	33526796		0.74
D3S3547	AFMa108yc1	55.11	32508293	36077344	2.33 ⁴	0.74
D3S3727	AFMa090zb1	56.07	absent	36590181	2.33 ⁴	0.80
D3S3567	AFMa176wc1	56.69	33088568	36651351	2.33 ⁴	0.70
D3S2432	GATA27C08	57.92	35428715	37815575		0.86
D3S1619	AFM350tf1	60.98	36910081	39351382		0.75
D3S1277	AFM164we1	61.52	absent	39829404		0.82
D3S1298	AFM220yb8	62.05	absent	43649892		0.88
D3S3527	AFMa073yf5	63.12	42271240	absent		0.85
D3S3522	AFM349ta5	65.26	47250417 ⁷	45829520		0.58
D3S3685	AFMc006yh1	67.94	45056735	47615089		0.89
D3S3597	AFMa344zg5	68.47	absent	49243464	2.70 ⁵	0.74
D3S3582	AFMa231xe9	69.19	absent	50789014	2.70 ⁵	0.69
D3S1767	GATA7A01	69.9	48749225	52329896	2.70 ⁵	0.76
D3S1578	AFM268wg5	70.61	absent	59028515		0.88
D3S1289	AFM198yf2	71.41	58098072	59706713		0.82
D3S3666	AFMb347zg9	72.21	absent	59889997		0.85
D3S3717	AFMa066wd5	74.35	58775661	60476808	0.00 ⁶	0.78
D3S3588	AFMa275zh1	74.88	59672322	61428438	0.00 ⁶	0.65
D3S3621	AFMb283ya5	75.41	absent	61665377		0.77
D3S3616	AFMb059wf1	76.48	60908605	62673287		0.85

¹ Sex-average Kosambi cM map position for males and females from the Marshfield genetic map.

² Minimum Likelihood Ratio for given a marker position from all possible marker orders (see additional footnotes for all possible orders with LR < 3). If blank, LR > 3.

³ Order: D3S3659 D3S3700 D3S1599 (LR = 2.91), Order: D3S3659 D3S1599 D3S3700 (LR = 2.89).

⁴ Order: D3S3567 D3S3547 D3S3727 (LR = 2.52), Order: D3S3547 D3S3567 D3S3727 (LR = 2.51), Order: D3S3727 D3S3547 D3S3567 (LR = 2.33), Order: D3S3727 D3S3567 D3S3547 (LR = 2.33).

⁵Order: D3S3582 D3S3597 D3S1767 (LR = 2.99), Order: D3S3582 D3S1767 D3S3597 (LR = 2.99), Order D3S3597 D3S1767 D3S3582 (LR = 2.70), Order: D3S1767 D3S3597 D3S3582 (LR = 2.70).

⁶Order: D3S3588 D3S3717 (LR = 0.00).

⁷Physical order does not agree with genetic order.