BMI 826

Advanced Data Analysis

Karl Broman

Biostatistics & Medical Informatics, UW–Madison

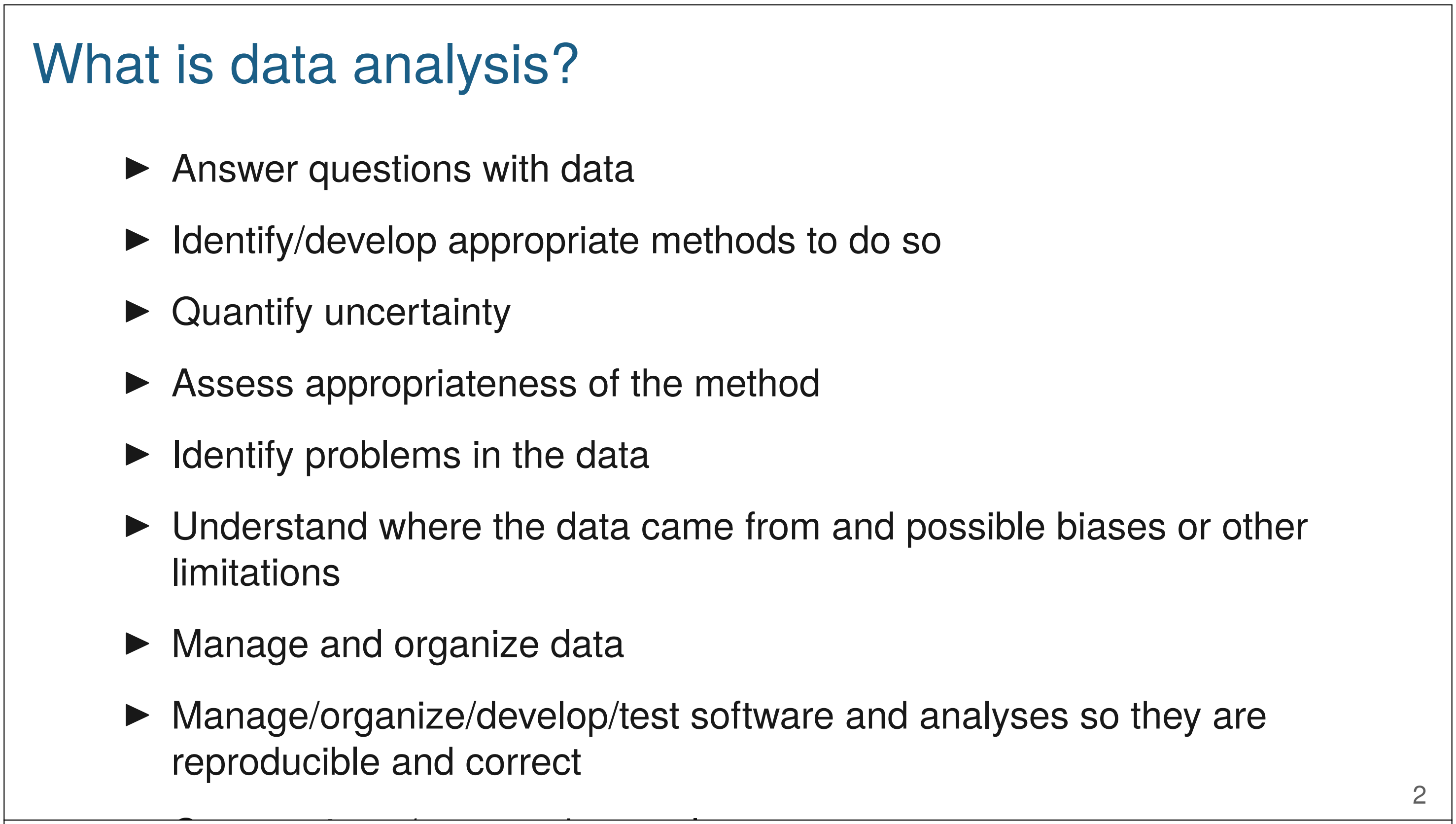
kbroman.org
github.com/kbroman
@kwbroman

Course web: kbroman.org/AdvData

This is the introductory lecture for a special topics course at UW–Madison on advanced data analysis. It might be better called "The craft of data analysis."

Data analysis involves both a set of skills and a way of thinking about and looking at data. It is critical to consider the scientific context of the data, and to focus on the scientific questions for which data were gathered.

And so the course has three streams: first, a set of case studies chosen to illustrate important lessons; second, a set of tutorials to introduce certain skills useful for developing and organizing reproducible data analyses; and third, homework assignments that involve direct, guided analyses of data.



What is data analysis?

- Answer questions with data
- Identify/develop appropriate methods to do so
- Quantify uncertainty
- Assess appropriateness of the method
- Identify problems in the data
- Understand where the data came from and possible biases or other limitations
- Manage and organize data
- Manage/organize/develop/test software and analyses so they are reproducible and correct

What is data analysis? It is a lot of things. First, of course, it is an effort to use data to answer questions. But also it involves identifying appropriate methods to answer questions with data, or developing new methods if such methods don't exist. And I would assert that it is always important to attempt to quantify the uncertainty in our answers.

We also need to be able to assess the appropriateness of methods, to identify problems in the data, and to understand where the data come from and any biases and other limitations in the data.

Further, good data analysis includes methods for managing and organizing data, and for managing, organizing, developing, and testing software and data analyses so that they are reproducible and correct.

Finally, data analysis includes the communication and presentation of the results, in a form appropriate to the audience of the work.

Important principles

1. You'll never know all the methods
2. Focus on the question and data, not the method
3. "Because you can" is not a good reason to do something

Here are some key principles that I live by.

Courses focused on methods will always be incomplete and can quickly become outdated. And so that's led me, in this course, to focus on my general approach data analysis, and the various lessons and principles I've acquired over time.

This course

- Data analysis projects
- Tools for organizing analyses so that they are reproducible
- Stories of data analysis projects, with lessons

Learning in this course will come from three streams of effort.

First, homework assignments that give direct experience in data analysis.

Second, explicit direct instruction in tools for organizing data analyses so that they are reproducible.

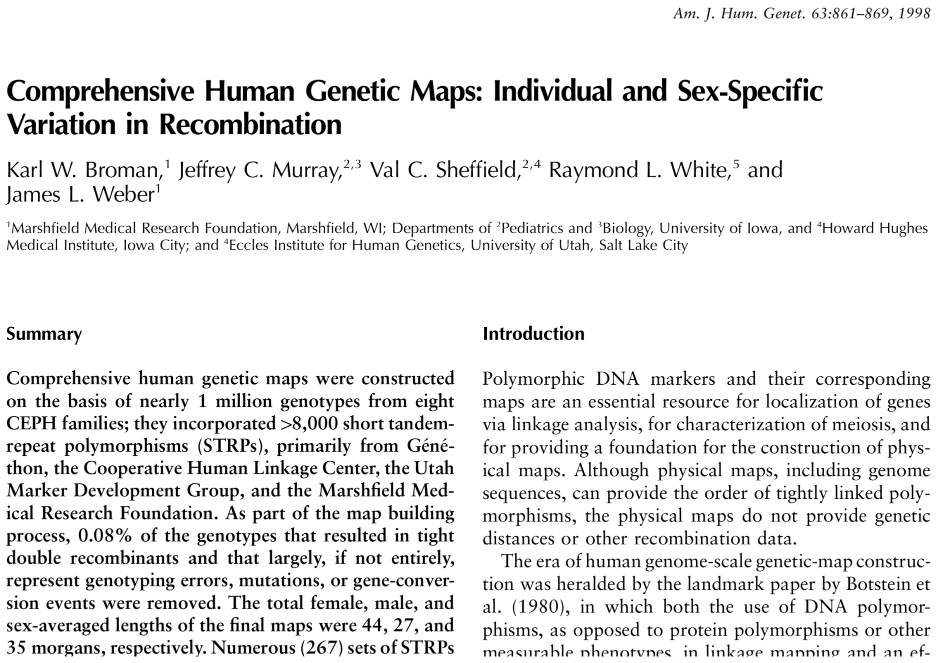
And finally, stories of past data analysis projects, with lessons I've learned.

Lesson 1

Follow up artifacts

They might be the most interesting results

Today, I'll provide a specific case study, with the key lesson being, "Follow up artifacts, as they might be the most interesting of your results."



Am. J. Hum. Genet. 63:861–869, 1998

Comprehensive Human Genetic Maps: Individual and Sex-Specific Variation in Recombination

Karl W. Broman,[1] Jeffrey C. Murray,[2,3] Val C. Sheffield,[2,4] Raymond L. White,[5] and James L. Weber[1]

[1]Marshfield Medical Research Foundation, Marshfield, WI; Departments of [2]Pediatrics and [3]Biology, University of Iowa, and [4]Howard Hughes Medical Institute, Iowa City; and [4]Eccles Institute for Human Genetics, University of Utah, Salt Lake City

Summary

Comprehensive human genetic maps were constructed on the basis of nearly 1 million genotypes from eight CEPH families; they incorporated >8,000 short tandem-repeat polymorphisms (STRPs), primarily from Généthon, the Cooperative Human Linkage Center, the Utah Marker Development Group, and the Marshfield Medical Research Foundation. As part of the map building process, 0.08% of the genotypes that resulted in tight double recombinants and that largely, if not entirely, represent genotyping errors, mutations, or gene-conversion events were removed. The total female, male, and sex-averaged lengths of the final maps were 44, 27, and 35 morgans, respectively. Numerous (267) sets of STRPs

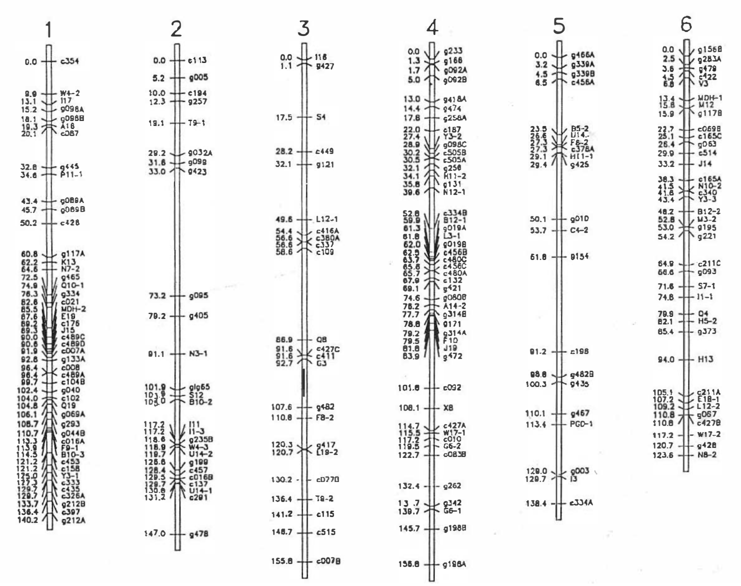
Introduction

Polymorphic DNA markers and their corresponding maps are an essential resource for localization of genes via linkage analysis, for characterization of meiosis, and for providing a foundation for the construction of physical maps. Although physical maps, including genome sequences, can provide the order of tightly linked polymorphisms, the physical maps do not provide genetic distances or other recombination data.

The era of human genome-scale genetic-map construction was heralded by the landmark paper by Botstein et al. (1980), in which both the use of DNA polymorphisms, as opposed to protein polymorphisms or other measurable phenotypes, in linkage mapping and an ef-

After I finished my PhD, I did a postdoc with a geneticist, Jim Weber, at the Marshfield Clinic. My central project was to develop new human genetic maps.

Eucalypt genetic map



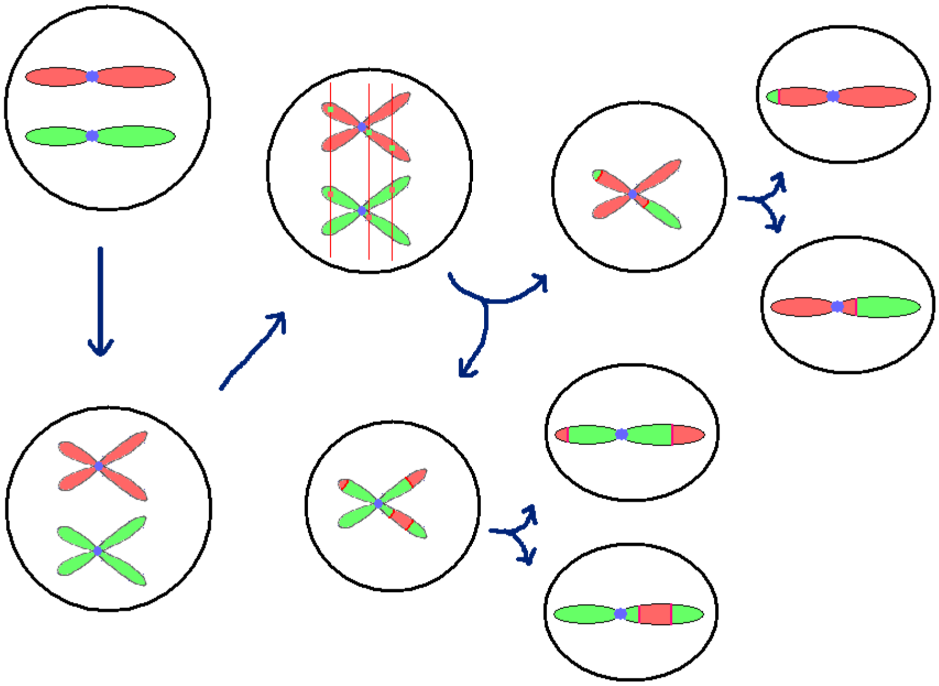
Byrne et al., Theor Appl Genet 91:869–875, 1995

A genetic map specifies the order of a set of markers along chromosomes.

This is part of a genetic map for eucalyptus trees. It is the first map that I had looked at in detail.

The original genetic maps were for observeable mutations, in Drosophila (fruit flies). Later markers were more directly DNA-based, and really chosen due to the convenience of measurement.

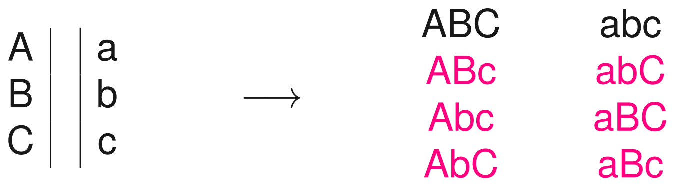
Meiosis



Distances on a genetic map are according to recombination at meiosis. Meiosis is the cell division process that produces sperm and egg cells. DNA duplicates, and then homologous chromosomes find each other and become intimately associated with each other and then actually exchange material at locations called chiasmata. Two cell divisions later you have gametes with one copy of each chromosome, which will generally be mosaics of the original chromosomes, with the points of exchange called crossovers.

Distance on a genetic map is measured by the frequency of crossovers. Two points are d cM apart if there is an average of d crossovers in the interval per 100 meiotic products.

Ordering markers



Marker orders: A–B–C A–C–B B–A–C

With M markers, there are M!/2 possible orderings.

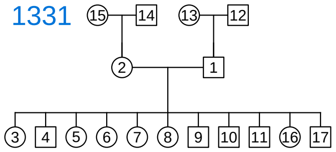
For $M = 100$, $M!/2 \approx 10^{157}$

We can use this sort of marker information to order markers along chromosomes. Consider a case where the parent cell has ABC on one chromosome and abc on the other chromosome. There are 8 possible daughter cells. The least frequent of them will indicate which of the three possible marker orders is the true one.

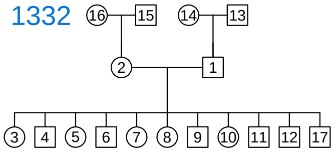
With M markers, there are M!/2 possible orderings, which is too many to consider exhaustively, so we need heuristic methods to try to find the good ones.

CEPH pedigrees

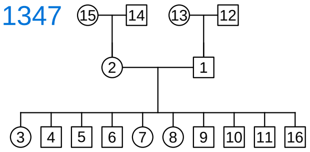
1331



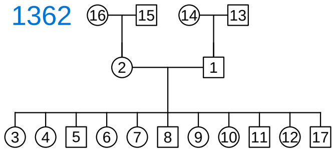
1332



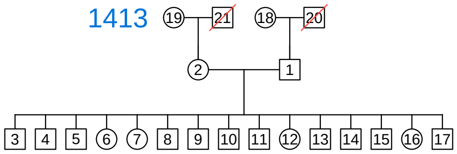
1347



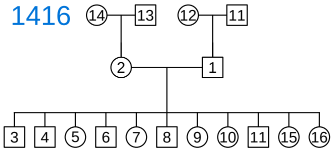
1362



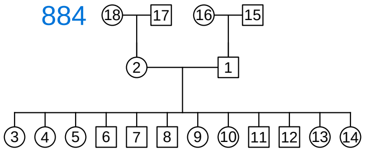
1413



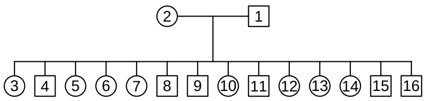
1416



884



102



In my postdoc, I focused on data on a set of large 8 human families. A mother/father pair with 10-15 offspring. Most of the families also included data on the grandparents.

Marshfield genetic maps: Tasks

- Assemble data
- Understand marker names

  AFM, UT, CHLC (GATA etc.), Mfd, D*S*
- Identify cryptic duplicates
- Order markers and identify genotyping errors

  Removed 764 / 969,425 genotypes

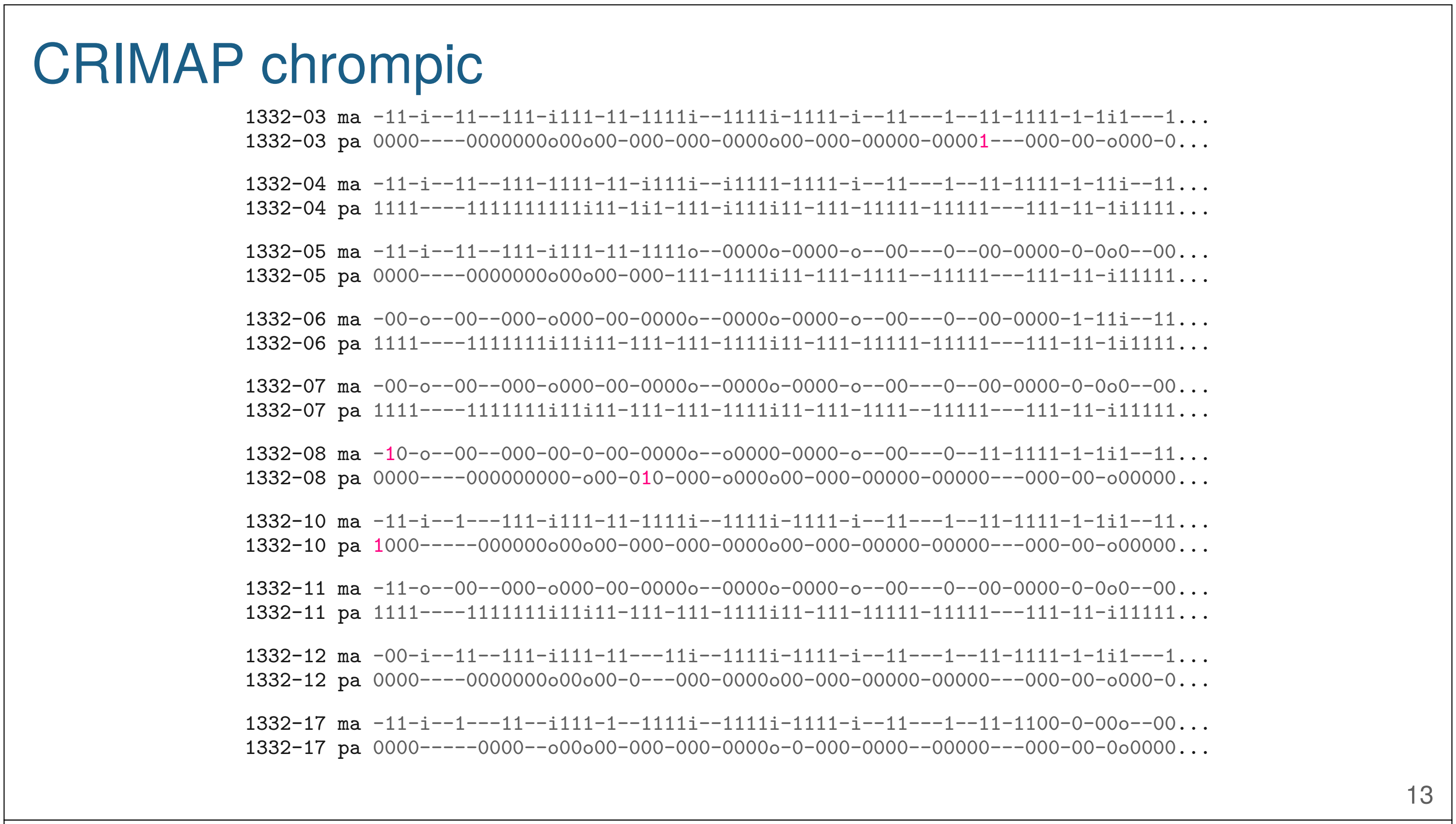
The main tasks were to assemble genotype data from multiple sources, get an understanding of the different marker names, figure out which markers were actually duplicates even though they didn't look like duplicates, and the mostly to determine the order of the markers while simultaneously identifying genotyping errors. In the end I removed 765 out of nearly one million genotypes as likely errors.

CRIMAP chrompic

```
1332-03 ma -11-i--11--111-i111-11-1111i--1111i-1111-i--11---1--11-1111-1-1i1---1...
1332-03 pa 0000----0000000o00o00-000-000-0000o00-000-00000-00001---000-00-o000-0...

1332-04 ma -11-i--11--111-1111-11-i111i--i1111-1111-i--11---1--11-1111-1-11i--11...
1332-04 pa 1111----1111111111i11-1i1-111-i111i11-111-11111-11111---111-11-1i1111...

1332-05 ma -11-i--11--111-i111-11-1111o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-05 pa 0000----0000000o00o00-000-111-1111i11-111-1111--11111---111-11-i11111...

1332-06 ma -00-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-1-11i--11...
1332-06 pa 1111----1111111i11i11-111-111-1111i11-111-11111-11111---111-11-1i1111...

1332-07 ma -00-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-07 pa 1111----1111111i11i11-111-111-1111i11-111-1111--11111---111-11-i11111...

1332-08 ma -10-o--00--000-00-0-00-0000o--o0000-0000-o--00---0--11-1111-1-1i1--11...
1332-08 pa 0000----000000000-o00-010-000-o000o00-000-00000-00000---000-00-o00000...

1332-10 ma -11-i--1---111-i111-11-1111i--1111i-1111-i--11---1--11-1111-1-1i1--11...
1332-10 pa 1000-----000000o00o00-000-000-0000o00-000-00000-00000---000-00-o00000...

1332-11 ma -11-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-11 pa 1111----1111111i11i11-111-111-1111i11-111-11111-11111---111-11-i11111...

1332-12 ma -00-i--11--111-i111-11---11i--1111i-1111-i--11---1--11-1111-1-1i1---1...
1332-12 pa 0000----0000000o00o00-0---000-0000o00-000-00000-00000---000-00-o000-0...

1332-17 ma -11-i--1---11--i111-1--1111i--1111i-1111-i--11---1--11-1100-0-00o--00...
1332-17 pa 0000-----0000--o00o00-000-000-0000o-0-000-0000--00000---000-00-0o0000...
```

I spent a lot of time looking at output like this, from the software CRIMAP which I had used to order and Q/C the data.

Each row is one chromosome (“ma” for maternal and “pa” for paternal). The 1’s and 0’s indicate grandfather’s and grandmother’s DNA; the dashes mean indeterminate. The i’s and o’s mean the grandparents’ genotypes weren’t informative and so grandparental origin was determined based on surrounding markers.



CRIMAP chrompic

```
1332-03 ma -11-i--11--111-i111-11-1111i--1111i-1111-i--11---1--11-1111-1-1i1---1...
1332-03 pa 0000----0000000o00o00-000-000-0000o00-000-00000-00001---000-00-o000-0...

1332-04 ma -11-i--11--111-1111-11-i111i--i1111-1111-i--11---1--11-1111-1-11i--11...
1332-04 pa 1111----1111111111i11-1i1-111-i111i11-111-11111-11111---111-11-1i1111...

1332-05 ma -11-i--11--111-i111-11-1111o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-05 pa 0000----0000000o00o00-000-111-1111i11-111-1111--11111---111-11-i11111...

1332-06 ma -00-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-1-11i--11...
1332-06 pa 1111----1111111i11i11-111-111-1111i11-111-11111-11111---111-11-1i1111...

1332-07 ma -00-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-07 pa 1111----1111111i11i11-111-111-1111i11-111-1111--11111---111-11-i11111...

1332-08 ma -10-o--00--000-00-0-00-0000o--o0000-0000-o--00---0--11-1111-1-1i1--11...
1332-08 pa 0000----000000000-o00-010-000-o000o00-000-00000-00000---000-00-o00000...

1332-10 ma -11-i--1---111-i111-11-1111i--1111i-1111-i--11---1--11-1111-1-1i1--11...
1332-10 pa 1000-----000000o00o00-000-000-0000o00-000-00000-00000---000-00-o00000...

1332-11 ma -11-o--00--000-o000-00-0000o--0000o-0000-o--00---0--00-0000-0-0o0--00...
1332-11 pa 1111----1111111i11i11-111-111-1111i11-111-11111-11111---111-11-i11111...

1332-12 ma -00-i--11--111-i111-11---11i--1111i-1111-i--11---1--11-1111-1-1i1---1...
1332-12 pa 0000----0000000o00o00-0---000-0000o00-000-00000-00000---000-00-o000-0...

1332-17 ma -11-i--1---11--i111-1--1111i--1111i-1111-i--11---1--11-1100-0-00o--00...
1332-17 pa 0000-----0000--o00o00-000-000-0000o-0-000-0000--00000---000-00-0o0000...
```

I spent most of my time hunting for misplaced 1's amidst surrounding 0's, or misplaced 0's amidst surrounding 1's, and then trying to decide if it was an error that should be deleted, or if the marker should maybe be moved somewhere else.

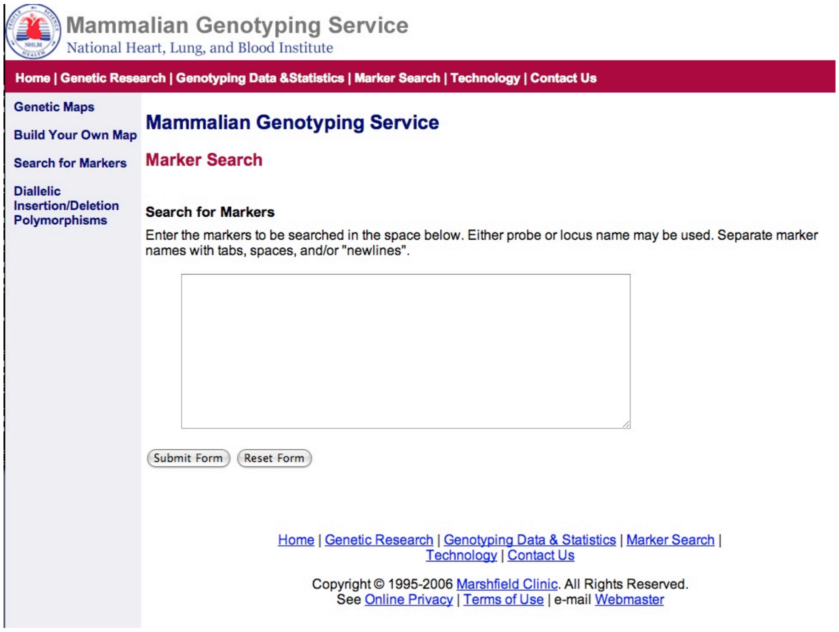
Top of chr 22

```
 Marker       Dnumber     sex-ave(cM)       female(cM)       male(cM)

1 ATA2G02     Unknown             0.00             0.00             0.00
                          1.79             0.00             2.60
2 GATA198B05  Unknown             1.79             0.00             2.60
                          2.27             3.32             0.00
3 AFM217xf4   D22S420             4.06             3.32             2.60
                          4.26             4.51             5.42
4 AFM288we5   D22S427             8.32             7.83             8.02
                          5.25             7.52             3.00
5 265yf5      D22S425            13.57            15.35            11.02
                          0.03             0.00             0.65
6 GGAA10F06   D22S686            13.60            15.35            11.67
                          0.84             0.00             0.82
7 AFMa037zd1  D22S539            14.44            15.35            12.49
                          0.00             0.00             0.00
8 AFM292va9   D22S446            14.44            15.35            12.49
                          3.27             5.91             0.00
9 Mfd51       D22S257            17.71            21.26            12.49
```

The result of the work was a set of plain text files with the marker positions on each chromosome.

Perhaps surprising is that much of the positive feedback I got about the work was really about the ease of use of these plain-text files. Most other genetic maps were distributed as images rather than providing the direct data.

Marker search


Mammalian Genotyping Service
National Heart, Lung, and Blood Institute
Home | Genetic Research | Genotyping Data &Statistics | Marker Search | Technology | Contact Us
Genetic Maps
Build Your Own Map
Search for Markers
Diallelic Insertion/Deletion Polymorphisms
Submit Form
Reset Form
Home | Genetic Research | Genotyping Data & Statistics | Marker Search | Technology | Contact Us

Mammalian Genotyping Service

Marker Search

Search for Markers

Enter the markers to be searched in the space below. Either probe or locus name may be used. Separate marker names with tabs, spaces, and/or "newlines".

Copyright © 1995-2006 Marshfield Clinic. All Rights Reserved.
See Online Privacy | Terms of Use | e-mail Webmaster

I also created a simple perl script and web form, where you could paste in a set of marker names, and it would pull out the locations of just those markers.

This was hardly any work for me, but it was hugely useful to the community. And probably the most important thing I learned from this project is that it's these little things that can have the biggest impact.

10th worst graph

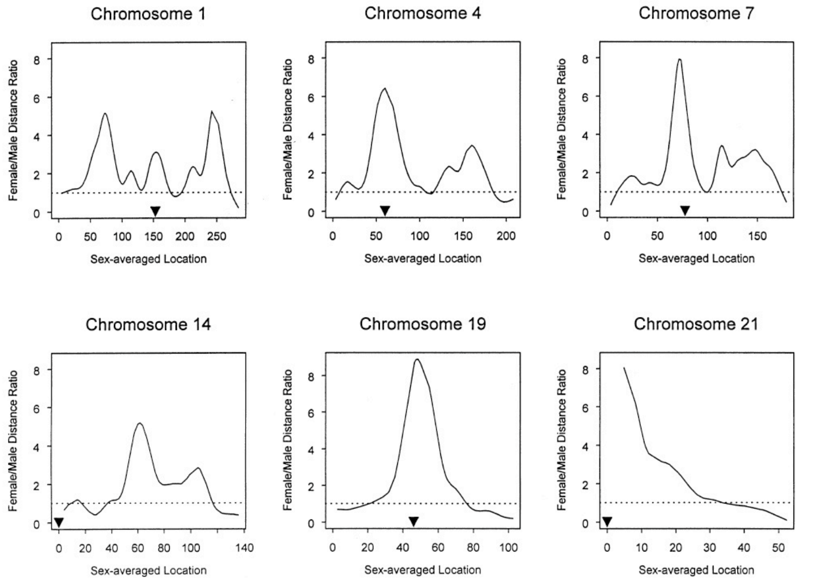


Figure 1 Plots of the female:male genetic-distance ratio against sex-averaged genetic location (in cM) along six selected chromosomes. Approximate locations of the centromeres are indicated by the triangles. The dashed lines correspond to equal female and male distances.

Broman et al., Am J Hum Genet 63:861–869, 1998

My interest in this project was not so much in the genetic maps as in what they tell us about the recombination process at meiosis.

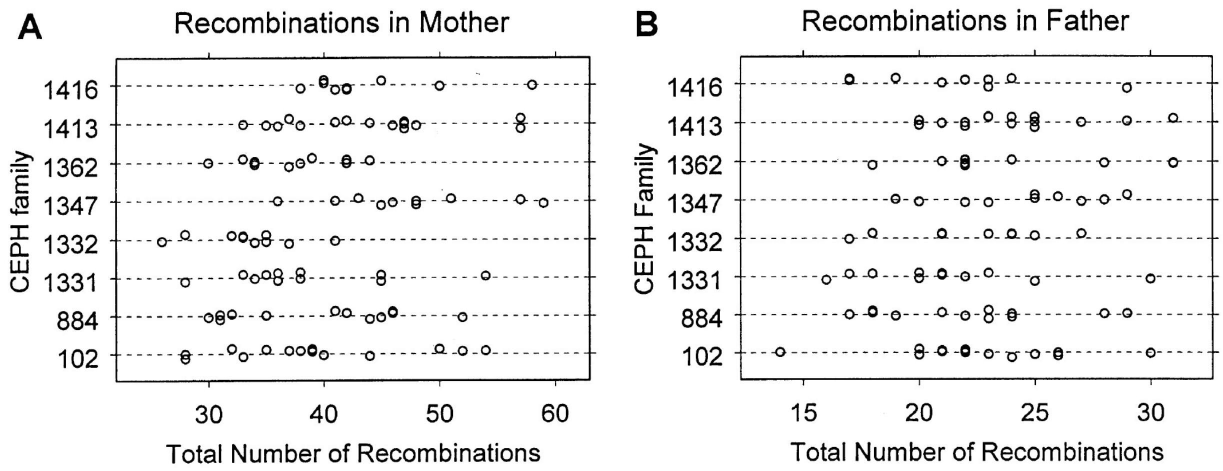
This graph shows the relative rates of recombination in females vs males. Female recombination is generally higher, but it varies a great deal between and along chromosomes, and at the ends of the chromosome, males tend to have higher recombination.

I call this the “10th worst graph” because I had included it on a web page of the “top ten worst graphs in the scientific literature.” The problem here is that I’m plotting a ratio, and it over-emphasizes where female recombination is greater than male (which stretches from 1 to infinity) and under-emphasizes where male recombination is greater than female (sandwiched between 0 and 1).

This plot should have been on a log scale, and really whenever I submit a paper I do a quick search for the word “which” and see if I should change it to “that,” and then I look at the plots and see if they would be better on the log scale.

When measurements span multiple orders of magnitude, you should probably take logs. And for ratios, you almost surely want to take logs. And I recommend log base 2, because you can multiply by 2 easily (2, 4, 8, 16, ...) and because the values are closer together than

Total no. crossovers

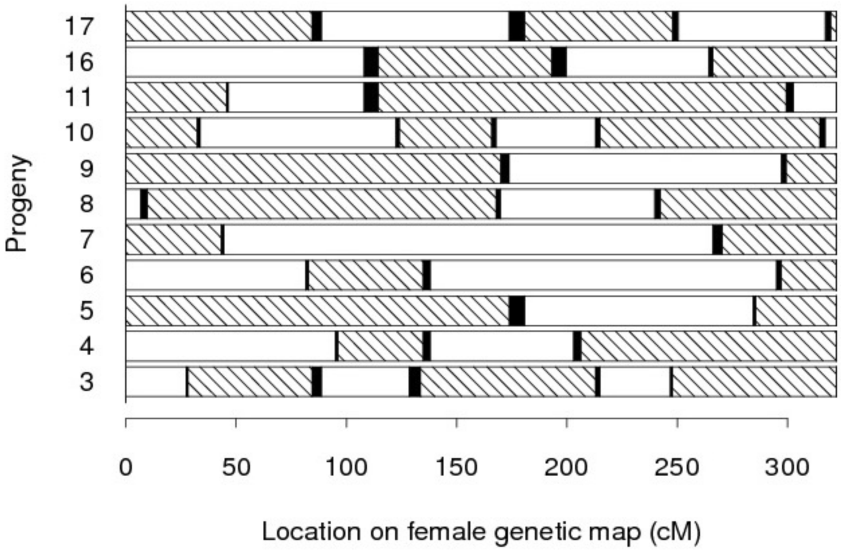


Broman et al., Am J Hum Genet 63:861–869, 1998

Another interesting thing we learned was about individual variation in recombination rates. If you count up the total number of recombination events in each egg and in each sperm that went on to be the children in these families, you notice that there is remarkable variation among women in their recombination rate, but little variation among the men.

Note also here the huge difference in the overall rate between mothers and fathers: an egg has an average of about 40 crossovers, where a sperm cell has more like 23.

Crossover locations



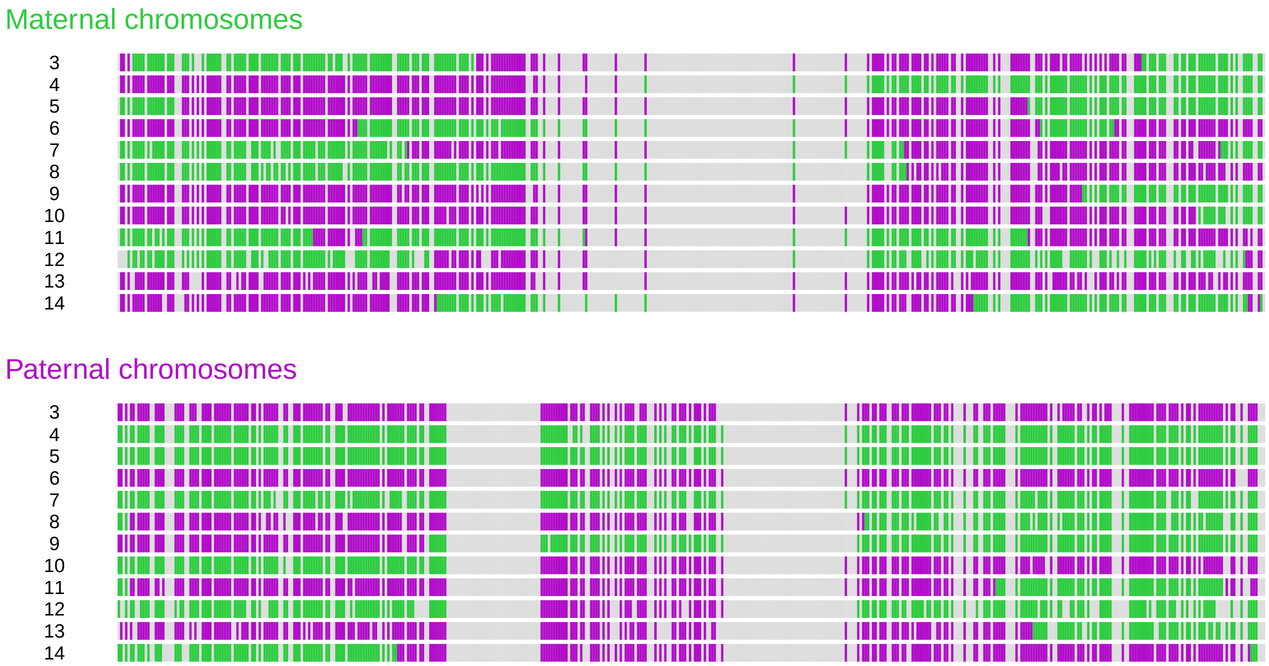
Broman and Weber, Am J Hum Genet 66:1911–1926, 2000

What I was really interested in was crossover interference: the tendency of the crossovers to not be too close together on chromosomes. The open and hatched segments here are the grandmother's and grandfather's DNA, and the black bars are the intervals in which crossovers occurred. We can't determine the crossover locations exactly, the data are "interval censored."

So the next thing I was going to look at was this dependence in crossover locations.

In thinking about how best to handle that interval censoring, I was reminded of some cases where there were really large, non-informative intervals.

Family 884, chr 6



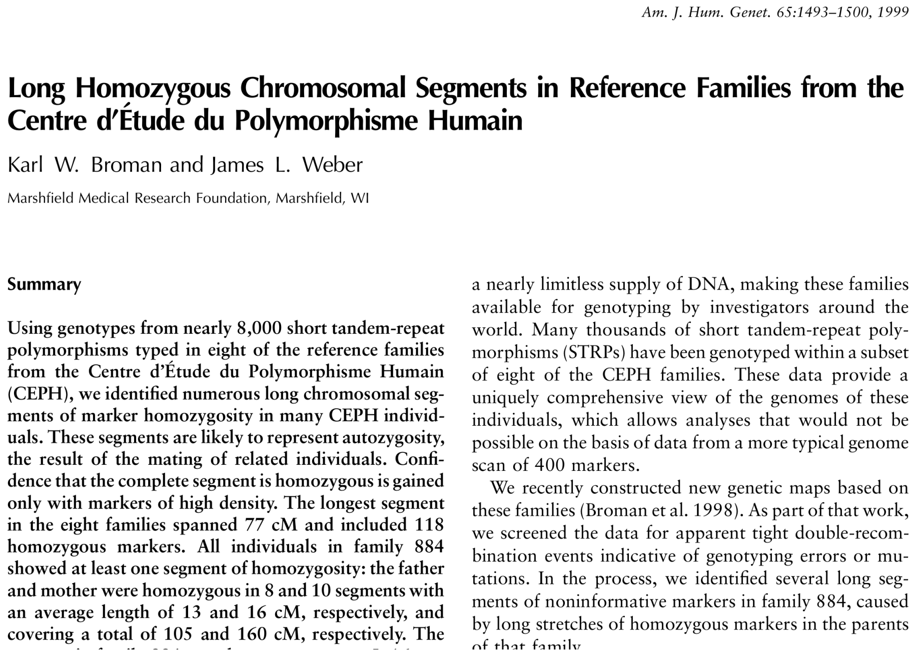
Maternal chromosomes

Paternal chromosomes

These are the maternal and paternal chromosomes 6 in one family. The purple and green are the grandmother and grandfather DNA; the gray is indeterminate.

Note the big chunk of gray on the maternal chromosomes, and the two big chunks on the paternal chromosomes. What could be causing that?

We basically have long stretches of markers where the mother or father is homozygous, where the haplotypes they got from their parents were the same.



Am. J. Hum. Genet. 65:1493–1500, 1999

Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain

Karl W. Broman and James L. Weber

Marshfield Medical Research Foundation, Marshfield, WI

Summary

Using genotypes from nearly 8,000 short tandem-repeat polymorphisms typed in eight of the reference families from the Centre d'Étude du Polymorphisme Humain (CEPH), we identified numerous long chromosomal segments of marker homozygosity in many CEPH individuals. These segments are likely to represent autozygosity, the result of the mating of related individuals. Confidence that the complete segment is homozygous is gained only with markers of high density. The longest segment in the eight families spanned 77 cM and included 118 homozygous markers. All individuals in family 884 showed at least one segment of homozygosity: the father and mother were homozygous in 8 and 10 segments with an average length of 13 and 16 cM, respectively, and covering a total of 105 and 160 cM, respectively. The

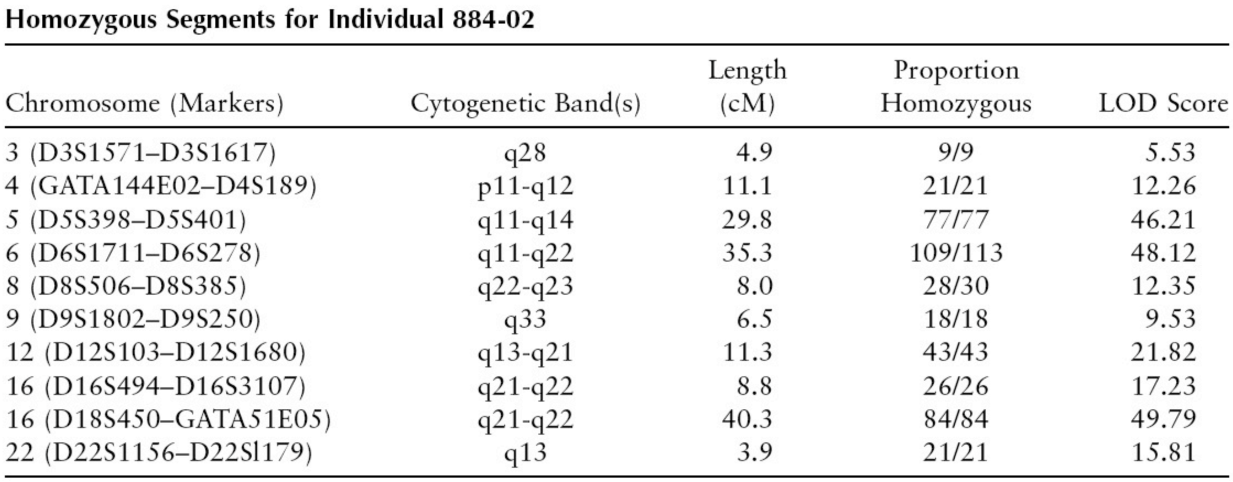
a nearly limitless supply of DNA, making these families available for genotyping by investigators around the world. Many thousands of short tandem-repeat polymorphisms (STRPs) have been genotyped within a subset of eight of the CEPH families. These data provide a uniquely comprehensive view of the genomes of these individuals, which allows analyses that would not be possible on the basis of data from a more typical genome scan of 400 markers.

We recently constructed new genetic maps based on these families (Broman et al. 1998). As part of that work, we screened the data for apparent tight double-recombination events indicative of genotyping errors or mutations. In the process, we identified several long segments of noninformative markers in family 884, caused by long stretches of homozygous markers in the parents of that family

This is what is called "autozygosity." Where an individual is homozygous for a region because their parents are related, perhaps distantly, and they received two copies of the same bit of DNA, inherit from some ancestor, twice.

Have realized what I was looking at, I went and looked for all other such regions, and found bunches of them in these families.

Autozygosity

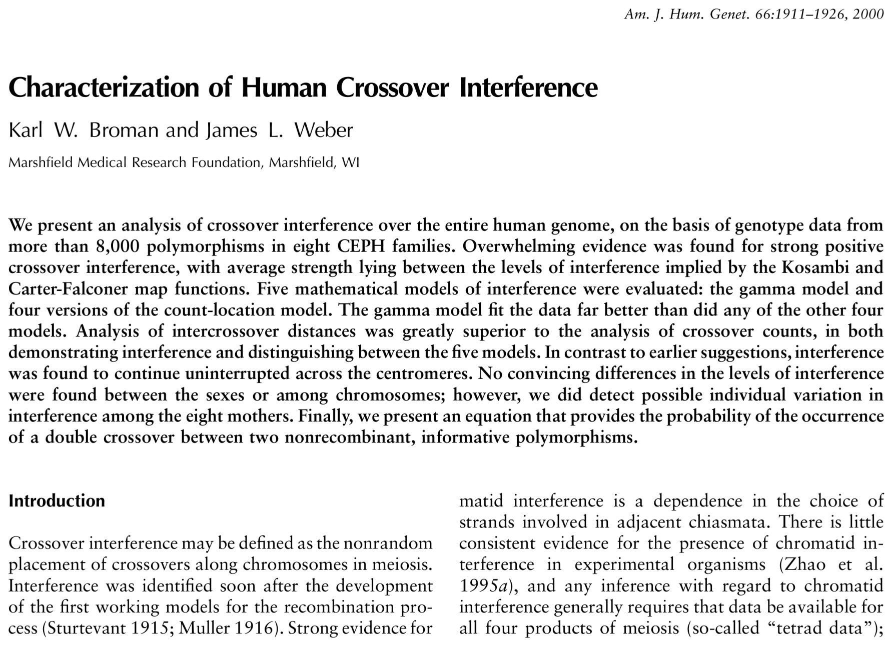


Homozygous Segments for Individual 884-02

Chromosome (Markers)	Cytogenetic Band(s)	Length (cM)	Proportion Homozygous	LOD Score
3 (D3S1571–D3S1617)	q28	4.9	9/9	5.53
4 (GATA144E02–D4S189)	p11-q12	11.1	21/21	12.26
5 (D5S398–D5S401)	q11-q14	29.8	77/77	46.21
6 (D6S1711–D6S278)	q11-q22	35.3	109/113	48.12
8 (D8S506–D8S385)	q22-q23	8.0	28/30	12.35
9 (D9S1802–D9S250)	q33	6.5	18/18	9.53
12 (D12S103–D12S1680)	q13-q21	11.3	43/43	21.82
16 (D16S494–D16S3107)	q21-q22	8.8	26/26	17.23
16 (D18S450–GATA51E05)	q21-q22	40.3	84/84	49.79
22 (D22S1156–D22Sl179)	q13	3.9	21/21	15.81

Broman and Weber, Am J Hum Genet 65:1493–1500, 1999

For example, here's a single individual who has autozygous regions of various sizes on ten different chromosomes.



Am. J. Hum. Genet. 66:1911–1926, 2000

Characterization of Human Crossover Interference

Karl W. Broman and James L. Weber

Marshfield Medical Research Foundation, Marshfield, WI

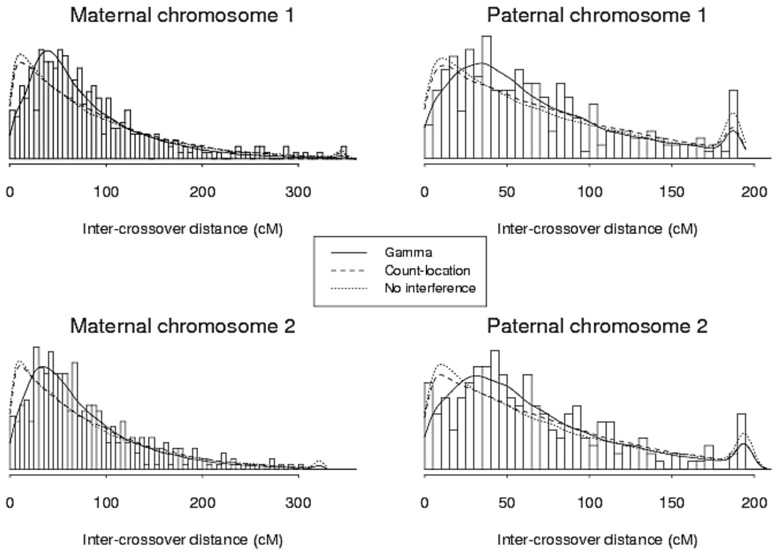
We present an analysis of crossover interference over the entire human genome, on the basis of genotype data from more than 8,000 polymorphisms in eight CEPH families. Overwhelming evidence was found for strong positive crossover interference, with average strength lying between the levels of interference implied by the Kosambi and Carter-Falconer map functions. Five mathematical models of interference were evaluated: the gamma model and four versions of the count-location model. The gamma model fit the data far better than did any of the other four models. Analysis of intercrossover distances was greatly superior to the analysis of crossover counts, in both demonstrating interference and distinguishing between the five models. In contrast to earlier suggestions, interference was found to continue uninterrupted across the centromeres. No convincing differences in the levels of interference were found between the sexes or among chromosomes; however, we did detect possible individual variation in interference among the eight mothers. Finally, we present an equation that provides the probability of the occurrence of a double crossover between two nonrecombinant, informative polymorphisms.

Introduction

Crossover interference may be defined as the nonrandom placement of crossovers along chromosomes in meiosis. Interference was identified soon after the development of the first working models for the recombination process (Sturtevant 1915; Muller 1916). Strong evidence for matid interference is a dependence in the choice of strands involved in adjacent chiasmata. There is little consistent evidence for the presence of chromatid interference in experimental organisms (Zhao et al. 1995*a*), and any inference with regard to chromatid interference generally requires that data be available for all four products of meiosis (so-called "tetrad data");

I did then get to my analysis of crossover interference (the tendency of crossovers to not be too close together).

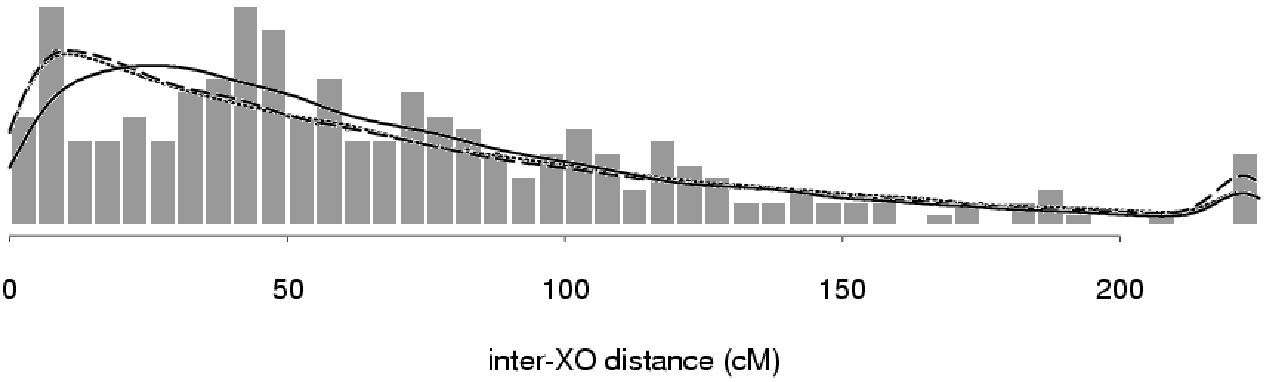
Crossover interference



Broman and Weber, Am J Hum Genet 66:1911–1926, 2000

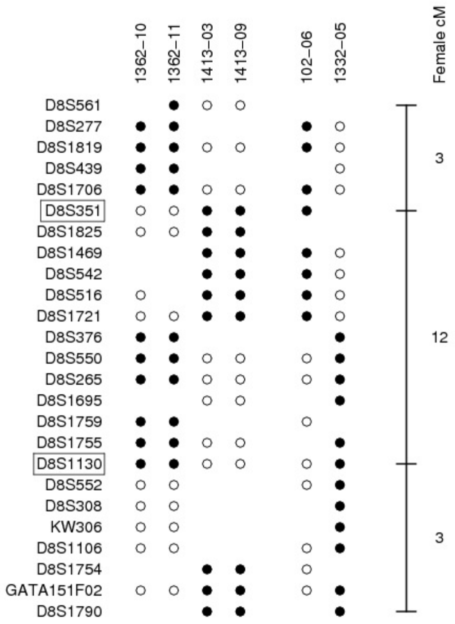
A main part of the result concerned fitting different models to the inter-crossover distance data. One model fit much better than others.

Maternal chr 8



But on one particular chromosome (maternal chromosome 8), my favorite model really didn’t fit well at all.

Apparent triple XOs



Broman et al., In: *Science and Statistics: A Festschrift for Terry Speed*, 2003

I could have just left it at that, but I was curious about what was going on, and in studying the problem, I found that there were two families that showed an apparent triple-crossover event in a small region. This really shouldn't happen.

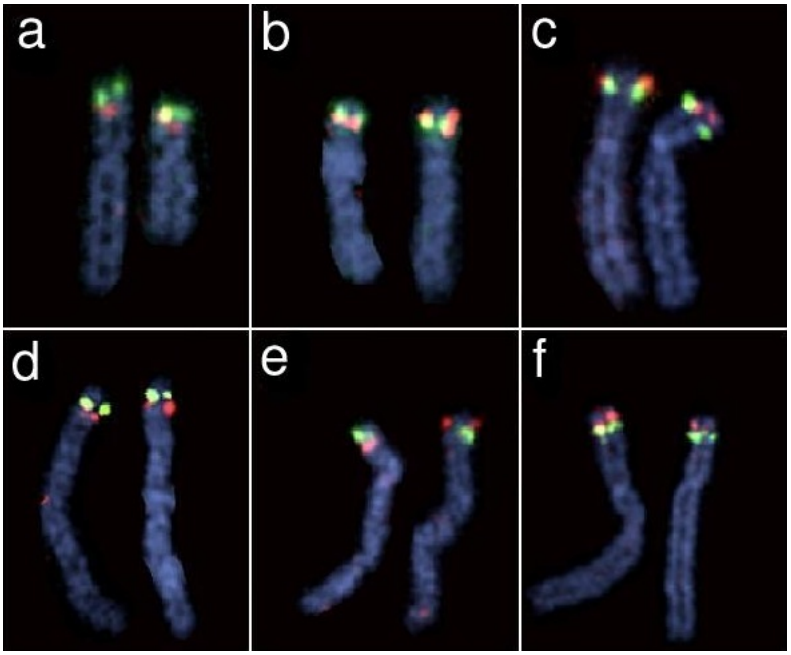
My initial reaction was that I had the marker order messed up; if I were to invert this region, the triple crossovers would become single crossovers.

But there were other families that showed a crossover in the region. If I invert the region, these single crossovers will become triple crossovers.

So then I thought: suppose the region is inverted in these two families but not in the other families? This was a pretty crazy idea, because the region is quite large (12 cM, which turned out to be about 5 Mbp), and we would need individuals to be homozygous for each of the two orientations to have recombination occur.

So a crazy idea: a very long inversion polymorphism where the two orientations were each reasonably common.

Chr 8p inversion



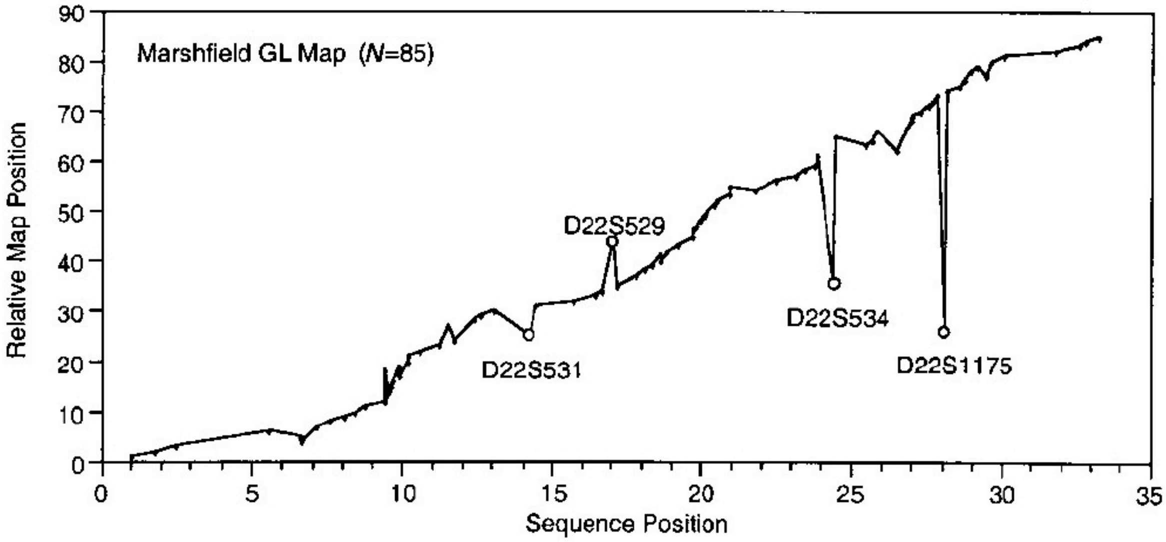
Broman et al., In: *Science and Statistics: A Festschrift for Terry Speed*, 2003

I posed the hypothesis to my postdoc advisor, who talked to a friend whose lab had the ability to investigate this sort of thing, and sure enough, we had discovered the largest common inversion polymorphism in the human genome.

This picture shows chromosome 8 with the green and red lighting up the two ends of the region. On the left, green is above red on both chromosomes. On the right, red is above green on both chromosomes, and in the middle green is above red on one chromosome and red is above green on the other.

So this is the best possible example of the importance of following up artifacts. Lack of model fit for a particular chromosome led me to investigate the cause of the problem, which led me to postulate this idea of an inversion polymorphism, which really seemed kind of crazy at the time. But it turned out to be real, and it's the coolest thing I've discovered in all my work as a data scientist.

Comparison to sequence

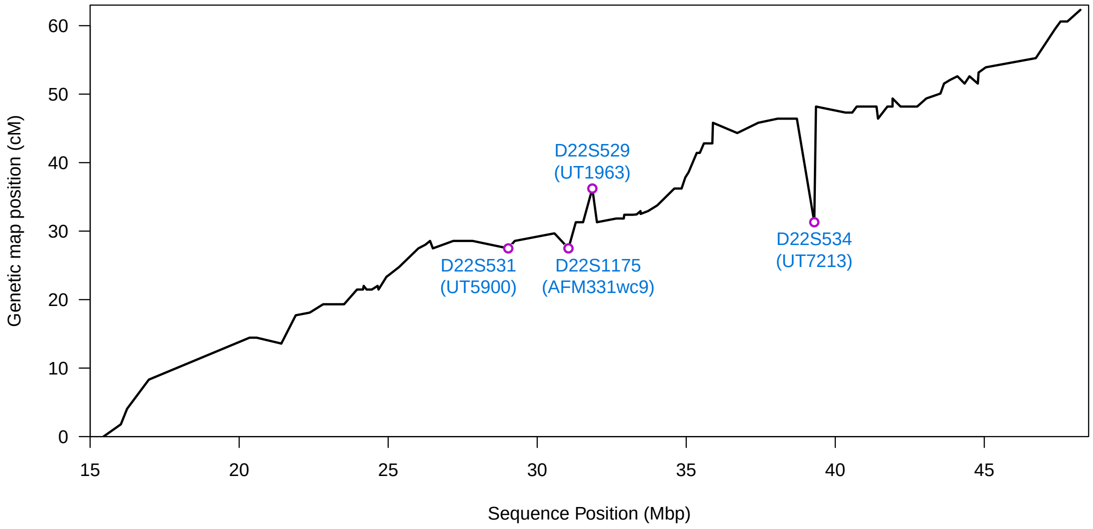


Matise et al., Am J Hum Genet 70:1398–1410, 2002

A few years later, when the human genome sequence was available, there was a paper that investigated the order of markers in my maps. I had a few rather embarrassing errors, mostly due to markers whose locations were poorly resolved as they had only been genotyped in a subset of the families.

I didn't provide any measures of uncertainty, and it would have been good to at least flag the markers that were especially uncertain.

Comparison to sequence



Funny thing though; it turned out that 10 years later, the worst of those problems were seen to be not a problem with my genetic map, but rather with the initial human genome sequence. I still made some mistakes, but the biggest mistake on that previous slide was likely in the initial draft of the human genome sequence.

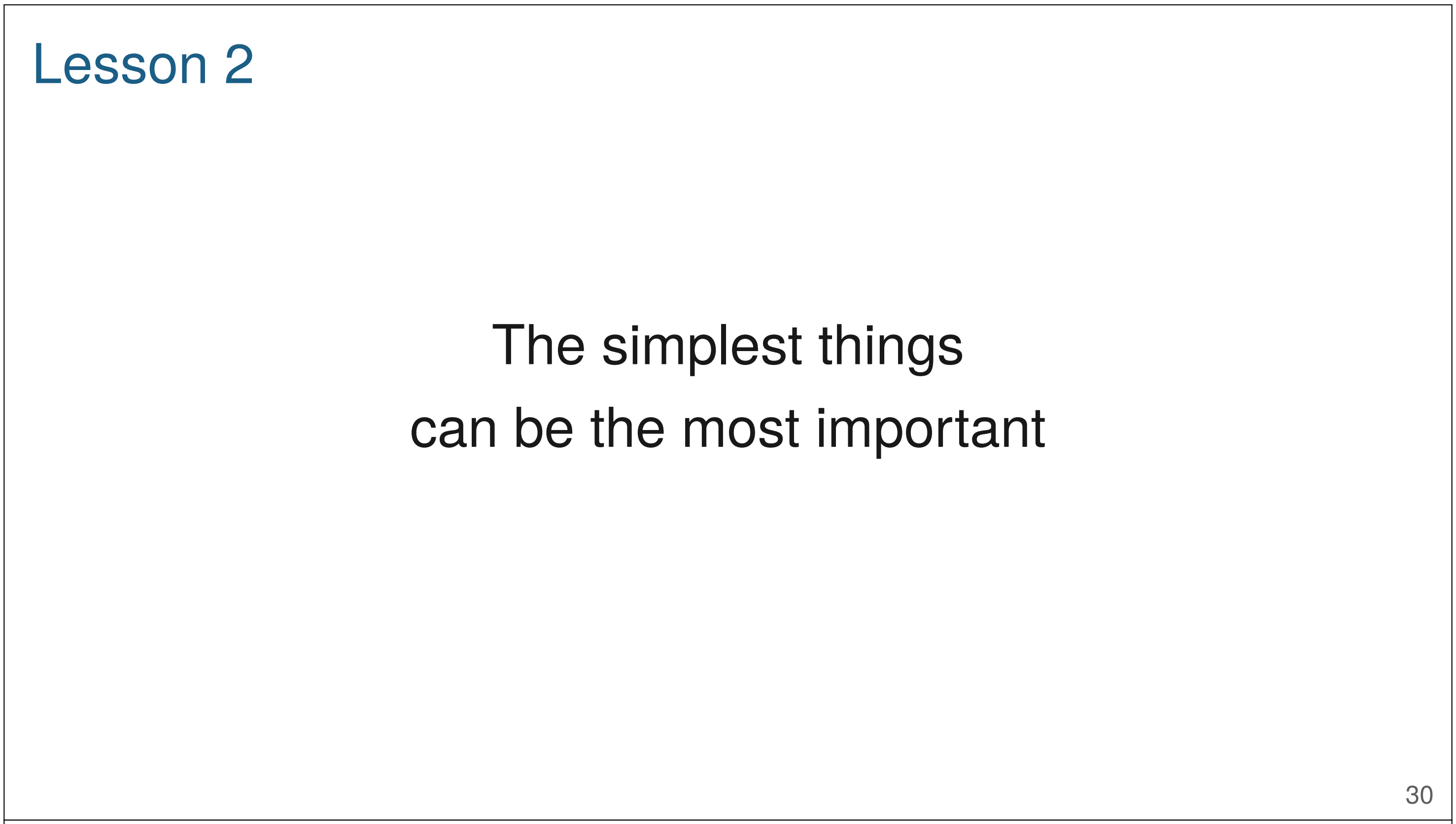
Lesson 1

Follow up artifacts

They might be the most interesting results

Two key lessons here.

First, follow up artifacts, as they can be the most important findings. Here, the autozygosity, and then the large common inversion.



Lesson 2

The simplest things
can be the most important

Second, the simplest things can be the most important. In this project, most of the positive feedback I got were due to way in which I provided the results, plus that I took a little time to create a web form for searching for markers.