Human Molecular Genetics, 2004, Vol. 13, No. 17 **1943–1949**
doi:10.1093/hmg/ddh202
Advance Access published on July 6, 2004

Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping

The International Multiple Sclerosis Genetics Consortium*

Received March 19, 2004; Revised and Accepted June 22, 2004

To explore the potential value of recently developed high-density linkage mapping methods in the analysis of complex disease we have regenotyped five nuclear families first studied in the 1996 UK multiple sclerosis linkage genome screen, using Applied Biosystems high-density microsatellite linkage mapping set, the Illumina BeadArray™ linkage mapping panel (version 3) and the Affymetrix GeneChip® Human Mapping 10K array. We found that genotyping success, information extraction and genotyping accuracy were improved with all systems. These improvements were particularly marked with the SNP-based methods (Illumina and Affymetrix), with little difference between these. The extent of additional information extracted is considerable, indicating that reanalysis of existing multiplex families using these newer systems would substantially increase power.

INTRODUCTION

Linkage analysis has been extremely successful in mapping the genes responsible for monogenic (Mendelian) traits but comparatively disappointing when applied to complex diseases, such as multiple sclerosis, where the correlation between genotype and phenotype is reduced (1). To date 10 non-parametric whole genome linkage screens have been performed in the disease (2–11) but none has succeeded in identifying any region of statistically unequivocal linkage, although nearly all have identified more regions of potential linkage than would be expected under the null hypothesis. Meta-analysis of these data has been published recently (12), revealing several important properties. Firstly, together these families have the power to demonstrate linkage with genome-wide significance in regions containing susceptibility genes of modest effect. Linkage is confidently identified at the major histocompatibility complex (MHC) on chromosome 6p21, a region known to contain at least one susceptibility gene (13). Secondly, the degree of sharing outside the MHC is confirmed as significantly greater than expected by chance alone, indicating that additional loci exist mapping outside the MHC. Thirdly, and perhaps most importantly, the meta-analysis demonstrated significant weaknesses in the published studies. The mean information extraction across the genome was only 44%, indicating that substantial power lies unused within existing families. The only region where linkage achieved genome-wide significance coincided with that having the greatest information extraction (68%). Although this is not surprising, given the tendency to include additional markers or human leukocyte antigen (HLA) typing data within the analysis of the MHC, it further illustrates the potential dividend from increased typing of existing families outside the MHC region.

Non-parametric linkage analysis relies on the fact that regions of the genome in linkage with disease alleles will be inherited by co-affected relatives more often than expected by chance alone. Testing for linkage in this manner is thus performed by observing the actual extent of allele sharing amongst affected relatives and comparing this with the expected sharing under Mendelian principles. The apparent simplicity of this process is deceptive because, in practice, many factors limit the extent to which sharing can reliably be determined, the most critical being the marker map employed and the extent of missing information. The existing multiple sclerosis linkage screens have employed low-density ($\approx$10 cM) microsatellite maps, which have hitherto been considered to represent the optimum balance between cost and information extraction (14). Missing data have further limited these studies with DNA from founders and unaffected relatives being available in only a proportion of families and these samples only being systematically genotyped in a minority of the studies. As allele sharing can be determined only when both affected relatives in a pair have been successfully genotyped, it follows that the rate at which genotyping fails is particularly crucial. Assuming a modest genotyping failure rate of n%, it follows that sharing will not be determined in almost $2n$% of pairs. In the original 1996 UK screen (2), for example, the genotyping success rate was 92%, indicating that on average sharing could not be

*To whom correspondence should be addressed: Stephen J. Sawcer. Tel: +44 1223217091; Fax: +44 1223336941; Email: sjs1016@mole.bio.cam.ac.uk. Contributing Consortium Members are listed in the Appendix.

determined in almost 16% of sib pairs. It follows that retyping families from these previously published screens using more reliable markers, capable of providing higher rates of genotyping success, would enable additional information to be extracted and thereby increase the power to detect linkage.

Each of the published screens employed multipoint analysis in order to estimate sharing across the whole genome rather than limiting analysis to individual markers. Although these methods maximize the amount of information that can be extracted, they are inherently dependent on the accuracy of the genetic map employed. These maps have substantially improved in recent years (15), bringing into question the reliability of the results available from the published screens. For example, in the original 1996 UK study the marker D3S11 was mapped to chromosome 3 and was analyzed as such, but it has now been determined that in fact this marker lies on chromosome 2. The marker D3S11 is not alone in finding itself relocated with improved knowledge of the genome.

The accuracy of genotyping performed in the published screens must also be considered. Although each employed the optimum affordable technology of the time, developments in the human genome project have progressed at such a pace that by today's standards, many of these data would be considered unreliable. Several authors have shown that previously employed typing methods introduce genotyping error rates of 0.5–1%, even for well-behaved microsatellites (16), with much higher error rates expected for less robust markers (17). Although modest and not previously attracting much attention, Abecasis *et al.* (18) have shown that error rates of just 1% can reduce observed LOD scores by as much as 50%. This massive effect on power suggests that increasing map density by typing more microsatellites using the same typing system is unlikely to be beneficial unless much greater accuracy can also be assured. Carefully designed and optimized microsatellite marker sets help in achieving this standard, to some extent, and are available at considerably higher density than has been employed in the published linkage genome screens.

The greater ease with which single nucleotide polymorphisms (SNPs) can be genotyped accurately, reliably and cheaply has prompted the development of several high-density SNP linkage mapping sets. In order to test the extent to which these improved reagents might enable us to capture the unused information available in our existing families, we have regenotyped five nuclear families originally included in the 1996 UK screen (2) using the Applied Biosystems high-density linkage mapping microsatellite marker set (HDLMS), the Illumina BeadArray™ linkage mapping panel version 3 (Illumina) and the Affymetrix GeneChip® Human Mapping 10K array (Affymetrix).

Table 1. The number of markers genotyped on each chromosome

Chromosome	UK Original	HDLMS	Illumina	Affymetrix
1	20	63	385	875
2	20	58	394	952
3	15	55	364	795
4	15	46	276	796
5	19	43	276	752
6	20	45	296	739
7	17	38	271	578
8	17	34	212	552
9	14	39	176	531
10	11	37	207	605
11	15	36	186	632
12	16	35	231	524
13	9	25	169	491
14	16	28	173	401
15	8	26	166	327
16	15	29	163	260
17	22	27	132	181
18	11	27	131	331
19	16	22	112	93
20	10	25	118	221
21	8	12	87	195
22	10	13	86	82
X	14	48	152[a]	310
Total	338	811	4763	11 223

[a]Includes 26 from the pseudoautosomal regions of X.

Table 2. The percentage of successfully scored genotypes by chromosome

Chromosome	UK Original	HDLMS	Illumina	Affymetrix
1	94.0	97.2	99.8	93.5
2	92.8	95.7	99.8	94.4
3	89.3	97.0	100.0	93.9
4	91.0	97.9	99.9	93.3
5	79.2	98.6	99.4	94.7
6	91.5	98.8	99.4	93.8
7	100.0	98.0	99.0	95.1
8	95.3	94.4	99.3	94.9
9	91.4	89.4	99.7	93.8
10	90.9	97.2	99.6	93.2
11	98.7	96.4	99.2	94.9
12	95.0	97.0	99.8	93.7
13	97.2	97.2	99.9	93.9
14	97.8	85.5	99.7	92.9
15	98.1	89.8	99.9	95.1
16	98.3	91.7	99.9	93.3
17	87.5	95.4	99.9	94.0
18	97.7	96.1	99.8	93.8
19	93.4	96.1	99.6	95.1
20	96.0	96.4	99.8	96.0
21	96.2	88.8	99.8	93.6
22	92.0	91.9	99.9	93.6
X	99.6	96.9	100.0	90.0
Total	93.5	95.57	99.69	93.95

RESULTS

Table 1 shows the number of markers employed by each screening system on each chromosome. Table 2 shows successfully scored genotypes (%) recorded for each chromosome across the five families. Even though the particular five families employed in this study had a higher than average genotyping success rate in the original UK screen, each of the more recently developed mapping sets achieved a still higher rate of success. A total of 15 501 (95.57%) genotypes were called with HDLMS, 94 960 (99.69%) with Illumina and 210 887 (93.95%) with Affymetrix. The improved genotyping success rate was particularly noticeable for the Illumina system, where only 300 (0.31%) genotypes were missed.

In the 'with parents' analysis (see Materials and Methods), 16 Mendelian inconsistencies were seen with the Illumina system and 119 with the Affymetrix system. No Mendelian inconsistencies were seen in the HDLMS data as inheritance consistency was used in the genotyping process. MERLIN identified improbable genotypes in each data set—19 (0.117%) with HDLMS, 19 (0.020%) with Illumina and 52 (0.023%) with Affymetrix. The low heterozygosity of SNPs means that only ~37% of genotyping errors will declare themselves as Mendelian inconsistencies (19) indicating approximate genotyping error rates of 0.045% for Illumina and 0.143% for Affymetrix. These rates are substantially lower than the ~1% rates anticipated for the published screens, with an important proportion of these being detectable as Mendelian errors or improbable genotypes. The rate of improbable genotypes detected by the MERLIN program is considerably higher for HDLMS than for the SNP-based systems (even allowing for the higher heterozygosity of these markers), suggesting that the underlying, but latent, rate of genotyping errors is higher with microsatellites. In the 'without parents' analysis (an analysis where the parental genotypes were ignored in order to simulate the situation where parental DNA is missing, see Materials and Methods) the lack of any information concerning parent of origin, reduces the power to identify genotyping errors as improbable meiotic events. As a result rather fewer such genotypes were detected by the MERLIN program—1 (0.003%) with HDLMS, 6 (0.003%) with Illumina and 22 (0.005%) with Affymetrix. Approximately 20% of these genotypes had been identified as Mendelian inconsistencies, and a further 40% as improbable, in the 'with parents' analysis. In summary, both SNP-based systems provide considerably higher levels of genotyping accuracy than can have been achieved in the currently published whole genome scans. These more sophisticated methods are able to identify almost 8% of genotyping errors even when parental DNA is not available.

Tables 3 and 4 list the mean information extraction achieved on each chromosome. We also explored the proportion of each chromosome where the sharing could be determined with confidence. At fully informative markers the degree of allele sharing in a sib pair will be known with certainty (two alleles, one allele or no alleles shared), however, where genotypes are missing or uninformative the number of alleles shared identical by descent (IBD) will be uncertain. In this situation, MERLIN uses data from flanking markers together with provided allele frequencies in order to estimate the probability of each possible degree of allele sharing. We determined the proportion of each chromosome where one of the sharing possibilities had a probability of >95% and present these results in Tables 5 and 6.

These results indicate that the SNP-based systems are able to increase the information extracted by almost 50% in those families with parents and by nearly 75% in those families without parents. The greater number of markers on the Affymetrix chip results in a greater information extraction with this system, particularly when founder genotypes are unavailable. Equivalent results were obtained with the GENEHUNTER program (data not shown).

Table 3. The mean information extraction (%) on each chromosome in the 'with parents' analysis

Chromosome	UK Original	HDLMS	Illumina	Affymetrix
1	65.1	85.5	93.7	92.7
2	60.4	83.2	93.5	94.0
3	50.8	85.2	93.5	94.3
4	67.2	84.8	91.5	94.7
5	57.1	84.7	94.3	94.2
6	59.8	86.0	94.3	95.6
7	63.7	84.0	92.2	93.2
8	67.0	80.6	91.8	93.0
9	57.3	84.1	88.3	87.9
10	49.6	82.5	92.1	94.2
11	68.1	85.6	90.5	94.2
12	69.9	85.2	91.9	92.6
13	61.1	83.9	92.4	93.4
14	73.6	79.5	91.8	92.0
15	58.0	78.3	93.1	95.2
16	68.3	82.3	92.3	88.9
17	67.3	83.8	90.6	85.2
18	59.7	82.7	90.0	92.9
19	70.8	82.7	91.2	74.0
20	55.7	80.9	91.7	84.6
21	61.4	62.4	92.9	91.3
22	72.8	72.7	89.4	85.9
X	49.2	81.0	62.6	86.5
Total	61.7	82.8	90.7	91.5

Table 4. The mean information extraction (%) on each chromosome in the 'without parents' analysis

Chromosome	UK Original	HDLMS	Illumina	Affymetrix
1	30.7	47.1	53.1	56.1
2	32.1	52.4	61.5	63.3
3	21.9	45.4	55.1	57.3
4	32.1	50.3	50.8	58.1
5	26.4	44.6	51.7	55.3
6	28.5	52.1	56.8	62.1
7	31.6	49.0	55.8	58.4
8	30.2	42.1	52.9	56.8
9	33.8	56.7	56.5	59.0
10	25.8	48.6	50.4	57.4
11	46.0	57.9	59.2	68.1
12	42.7	49.1	55.5	61.3
13	38.5	57.0	65.3	69.7
14	52.5	57.6	67.7	65.3
15	35.7	49.4	59.2	64.2
16	35.8	45.3	55.3	54.7
17	35.8	44.0	48.4	40.2
18	21.3	40.7	45.1	54.3
19	41.4	50.2	54.9	41.7
20	25.2	42.7	42.4	47.8
21	30.1	37.5	49.2	53.7
22	27.4	39.2	54.2	45.4
X	30.2	70.0	46.3	72.9
Total	32.4	49.7	54.4	58.4

DISCUSSION

We have shown that high-density SNP linkage mapping systems (in particular the Illumina BeadArray™ linkage mapping panel version 3 and Affymetrix GeneChip® Human Mapping 10 K array) provide significantly improved

Table 5. The mean proportion (%) of each chromosome where the IBD sharing could be determined with >95% confidence in the 'with parents' analysis

Chromosome	UK Original	HDLMS	Illumina	Affymetrix
1	32.3	73.9	87.9	84.9
2	24.6	69.7	88.0	86.7
3	15.9	71.8	86.5	88.5
4	31.5	72.3	84.0	89.2
5	28.1	72.1	89.4	88.1
6	29.4	76.3	89.5	90.9
7	29.4	71.8	83.5	86.2
8	32.3	62.0	85.0	86.8
9	23.9	73.8	79.2	79.8
10	10.7	68.9	87.2	88.8
11	31.6	73.0	82.8	87.0
12	34.9	71.4	83.7	84.3
13	30.1	69.0	85.0	86.7
14	53.6	61.6	84.5	84.5
15	15.7	55.1	87.4	91.9
16	46.6	67.6	84.1	76.0
17	44.4	70.1	83.8	68.8
18	21.0	67.2	83.1	85.9
19	40.2	65.8	82.8	51.4
20	19.7	62.1	85.7	71.7
21	34.7	32.4	86.6	79.0
22	43.7	42.4	80.0	69.7
X	39.1	79.5	55.3	84.2
Total	30.0	68.7	83.9	83.7

Table 6. The mean proportion (%) of each chromosome where the IBD sharing could be determined with >95% confidence in the 'without parents' analysis

Chromosome	UK Original	HD-LMS	Illumina	Affymetrix
1	6.4	39.9	57.1	59.4
2	5.1	48.4	66.4	69.0
3	4.1	37.5	57.2	66.4
4	6.6	45.6	47.7	63.8
5	3.5	38.8	59.8	70.4
6	8.0	54.4	64.5	73.1
7	5.9	37.6	56.1	59.8
8	5.1	30.0	52.3	64.5
9	7.4	43.4	47.0	55.6
10	1.8	44.1	36.4	65.3
11	13.8	49.5	58.6	78.2
12	14.9	36.6	48.3	61.3
13	10.2	43.8	63.8	74.7
14	30.2	40.5	58.9	58.1
15	4.6	34.5	49.8	59.4
16	16.0	37.3	56.0	49.5
17	21.6	29.8	43.2	32.6
18	3.1	24.7	30.2	57.4
19	19.3	36.8	47.0	27.2
20	4.3	31.5	29.4	42.8
21	9.4	23.3	36.5	49.4
22	13.4	25.5	48.9	28.2
X	17.9	59.3	36.3	65.9
Total	9.3	40.4	51.5	60.3

levels of information extraction with extremely high fidelity. Application of these systems to cohorts of multiply affected families previously typed with low-density microsatellites could nearly double the level of information extraction and reduce genotyping errors by perhaps as much as an order of magnitude. These results indicate that significant additional power can be obtained readily by revisiting previously typed complex disease family cohorts using these high throughput SNP technologies, or similar systems with equivalent characteristics. The marker densities employed in the systems studied represent a new optimal balance between cost and effort. Although marker density could be increased further the additional linkage information extracted per marker typed will of course fall and thus although denser SNP maps could achieve even greater rates of information extraction the benefit/cost ratio would be expected to diminish, unless genotyping costs also reduce.

It is well recognized that the power of linkage analysis is rapidly attenuated as the underlying genetic complexity increases (1). Furthermore classical parametric linkage analysis is difficult to apply, when multiple interacting loci are likely to be involved, no clear mode of inheritance can be discerned and large extended pedigrees are extremely unusual. Non-parametric methods rely on the fact that regions of the genome in linkage with disease alleles will be inherited by co-affected relatives more often than expected by chance. Therefore, they avoid many of the practical limitations imposed in the application of parametric linkage analysis to complex disease. However, the statistical limitations of linkage analysis cannot easily be escaped. Although the existing linkage genome screens in multiple sclerosis have failed to identify any region of statistically unequivocal linkage, each has observed excess peaks of potential linkage and the IBD allele sharing observed in these regions provides a guide to the likely size of genetic effects attributable to individual multiple sclerosis susceptibility loci [as measured in terms of λ_s, the sibling/population risk ratio (20)]. Most of these estimates range between λ_s values of 1.1 and 1.9.

Power calculations show that a sample size of 500–1000 affected sib pairs with high information extraction would have >80% power to identify loci with λ_s of 1.3 or greater, at LOD scores of genome-wide significance (1). Such a study would provide the best means of prioritizing genomic regions for investigation with the haplotype map, which currently cannot economically be applied systematically to the whole genome (although ultimately this may become affordable as costs reduce). Progress in determining the genetic basis of Crohns disease illustrates the value of a statistically robust linkage peak. Here, identification of linkage on chromosome 16 through combining linkage screening family sets (21) enabled efforts to concentrate on a genuinely relevant region with the subsequent identification of a susceptibility gene (22,23)

Our observations are consistent with those made by others comparing the performance of these same methods in monogenic and complex diseases (24–26). Although our study involves only a small number of samples, and is therefore subject to a potentially important degree of sampling variance, a large number of independent chromosomes have been studied and each provides an independent assessment. It remains to be determined whether the extremely encouraging levels of information extraction and genotyping accuracy achieved in this study could be maintained across a larger study involving many thousands of samples.

The limited number of families considered in this experiment do not provide sufficient data to enable us to identify

markers with genotype frequencies deviating from Hardy Weinberg equilibrium or to test for the presence of disequilibrium between linked markers. Both of these phenomena could be tested in a completed screen and would need to be considered. As the average separation of markers in these panels is substantially larger than the expected average extent of significant linkage disequilibrium it is unlikely that substantial linkage disequilibrium will exist between more than a few markers, and therefore it should be possible to allow for these effects without significantly reducing the power of any screen performed.

The results from this experiment indicate that the two SNP linkage systems under consideration are remarkably similar in performance—the Affymetrix system extracting marginally more information (particularly where parental genotypes were unavailable) and the Illumina system showing more complete genotyping and being minimally more accurate (at least in this data set). Other differences may therefore influence the preferred system such as the amount of DNA required to analyze a sample (250 ng for Affymetrix and 2 μg for Illumina). Ultimately, cost and availability may prove decisive. Markers included on the Affymetrix chip are randomly distributed across the genome, being determined by the location of suitable *Xba* restriction sites. As a consequence, the distribution of Affymetrix markers is more variable than in the Illumina set where markers are specifically chosen to be as uniformly distributed as possible. For some chromosomes, the density of Affymetrix markers is actually lower than for the Illumina set, whereas on others it is substantially higher. Both companies are developing their respective products, which are expected to improve in the coming years.

Given that results from the published linkage genome screens are based on significantly incomplete and inaccurate data analyzed using outdated genetic maps, the potential benefits of retyping existing family resources is clear. We have shown that high-density SNP linkage mapping sets have the ability to extract significantly more information and achieve higher rates of genotyping success and accuracy.

MATERIALS AND METHODS

Families

The five nuclear families considered in this study were selected from those included in the original 1996 UK screen (2). Each family consists of an affected sibling pair and both parents, providing a total of 20 individuals. The sib pairs include three sister–brother pairs, one sister–sister pair and one brother–brother pair. Each individual gave informed consent for genetic analysis and provided a venous blood sample from which DNA was extracted by standard means (more details are given in 2).

Genetic map

The analysis of each data set was based on the deCODE genetic map (15) with sex averaged recombination distances used throughout. In reanalysing the original UK data those markers ($n = 53$) not included in the deCODE map were either placed midway between flanking markers according to the map employed in the original 1996 analysis or, in the case of telomeric markers, placed according to their physical position relative to identifiable flanking deCODE markers. Comparing marker order according to the deCODE map with that originally employed in the 1996 screen revealed seven markers that had been reordered on their respective chromosomes. In all but one instance this was simply reordering with respect to an extremely close flanking marker. For the HDLMS set, all but seven markers had coordinates on the deCODE map. These seven were placed according to the same rules described above employing the map provided by Applied Biosystems. One hundred and fifty five of 4763 Illumina markers are not included in the deCODE map and were positioned using the genetic map developed in-house by Illumina (27) combined with estimates made by identifying flanking markers, that are included in the deCODE map. All of the 11 223 markers with unique positions from the Affymetrix chip are included on the deCODE map.

Statistical analysis

Multipoint linkage analysis was performed at 1 cM intervals along the entire length of each chromosome using the MERLIN program developed by Abecasis *et al.* (19) As each data set (UK original, HDLMS, Illumina and Affymetrix) was analyzed using the same map, the amount of data generated on each chromosome is identical. Two analyses were performed for each data set, the first including parental genotypes and the second ignoring these to simulate the situation in which parental DNA is unavailable. Although the MERLIN program automatically ignores genotypes that are inconsistent with Mendelian transmission, it should be noted that such genotyping errors can only be revealed in the 'with parents' analysis. It therefore follows that such genotypes will have been included in the 'without parents' analyses. This scenario more accurately reflects the real situation where such genotypes cannot be identified in a lone sib pair which can never have more than four alleles apparently distinct by descent.

In each of the three newly generated data sets (HDLMS, Illumina and Affymetrix) genotypes that were considered improbable using default settings of the MERLIN error option were zeroed out using the MERLIN wipe function (19). This error reduction process was applied independently in both 'with parents' and 'without parents' analysis, but was not applied to the UK original genotypes to preserve this data set as previously analyzed. The 'wiped' genotypes were analyzed in order to estimate the IBD sharing probabilities in each family and the information extraction across all families. 'Information extraction' being the proportion of available inheritance information that can be determined using the available genotypes, sometimes also referred to as information content.

In our reanalysis of the original UK data, allele frequencies estimated from the full set of families typed at the time, were reused. However, given the small number of families considered in the present study, we did not have sufficient data to reliably estimate allele frequencies for analysis of the newly generated data sets. Alternative sources of allele frequencies were therefore used. For the HDLMS markers, we used allele frequency estimates generated from the analysis

of pooled UK DNA samples genotyped as part of our previously reported low-density screen for linkage disequilibrium (28). These estimates were not corrected for length dependent amplification or stutter band artefacts and are thus crude. However, we reasoned that any bias induced by underestimating the frequency of longer alleles would be at least partially balanced by overestimates in the frequency of shorter alleles, which across all markers are no more likely to appear in families. For the Illumina and Affymetrix markers allele frequencies were provided by the respective companies, based in each case on the typing of unrelated Caucasian individuals.

Determination of the 'with parents' estimates of information extraction was repeated using the GENEHUNTER program version 2.1 (29).

Markers

UK original. The 358 markers in this data set include the 311 markers reported in the 1996 screen (2), together with HLA typing data and additional markers subsequently added to this screen (30).

HDLMS. Applied Biosystems high-density linkage mapping set consists of 811 microsatellite markers designed to provide an approximately uniform 5 cM map across the genome. These markers were typed in Cambridge UK using capillary electrophoresis on an Applied Biosystems 3700 DNA analyzer. Each marker was amplified by PCR using TrueAllele Premix (Applied Biosystems) according to the manufacturers recommended conditions on 9700 thermalcyclers (Applied Biosystems). In order to ensure the highest level of genotyping reliability and accuracy, electrophoresis was performed without multiplexing, products being diluted to prevent signal saturation. Fragment sizing was performed using Applied Biosystems high-density internal lane size standard (ROX 400HD), the latest sizing algorithm and GENESCAN software. Sized fragments were genotyped in a semi-automated manner using GENOTYPER software. Mendelian inheritance was employed during this process making it difficult to assess genotyping accuracy. These markers provide an $\sim$5 cM map.

Illumina. The oligonucleotide pools provided by Illumina produce genotyping results for 4937 SNPs, 174 of which are not currently mapped to a unique position and therefore could not be employed in the analysis. The remaining 4763 SNPs constitute the Illumina linkage mapping panel (version 3) and analysis was based on these. In Caucasians, these markers have been shown to have an average minor allele frequency (MAF) of 39%, with no marker having a MAF of $<$10%, and a mean heterozygosity of 0.46 (27). The linkage panel includes 26 SNPs from the pseudoautosomal regions of the X chromosome, 19 from the short arm (a roughly 2 cM interval) and seven from the long arm (a roughly 1 cM interval). Because no currently available multipoint linkage program is able to integrate data from X-linked and pseudoautosomal markers in a single analysis, each pseudoautosomal region was analyzed separately as though it were an independent autosomal chromosome. Results from these analyses were then combined with those from the standard X-linked markers so that the final length of the X chromosome was the same in the Illumina analysis as in each other data set. Results from analysis of the pseudoautosomal markers replaced those inferred for these regions from the X-linked analysis. The Illumina markers were typed at the Broad Institute according to the manufactures standard conditions.

Affymetrix. The Affymetrix GeneChip® Human Mapping 10 K array includes 11 560 SNP markers of which all but 337 have unique positions in the genome. These 337 could not be employed in the analysis, which was therefore based on the remaining 11 223 SNPs. In an analysis of Caucasian individuals, these markers have been shown to have a mean heterozygosity of 37% (company data). Affymetrix chip genotyping was performed by Medical Research Council (MRC) geneservice at Hinxton in Cambridge UK, all results were obtained from single pass experiments with a single sample hybridized to a single array, and no replication or averaging of multiple readings.

ACKNOWLEDGEMENTS

We would like to thank the representatives and technical staff from each company for their patience and support, in particular Helen Straw from Applied Biosystems, Sandy McBean and Sarah Shaw Murray from Illumina and Steve Picton and Claire McDonnell from Affymetrix. This work was supported by the International Multiple Sclerosis Genetics Consortium and a National Multiple Sclerosis Society Collabrative MS Research Center Award.

REFERENCES

1. Risch, N.J. (2000) Searching for genetic determinants in the new millennium. *Nature*, **405**, 847–856.
2. Sawcer, S., Jones, H.B., Feakes, R., Gray, J., Smaldon, N., Chataway, J., Robertson, N., Clayton, D., Goodfellow, P.N. and Compston, A. (1996) A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. *Nat. Genet.*, **13**, 464–468.
3. Haines, J.L., Ter Minassian, M., Bazyk, A., Gusella, J.F., Kim, D.J., Terwedow, H., Pericak-Vance, M.A., Rimmler, J.B., Haynes, C.S., Roses, A.D. *et al.* (1996) A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. *Nat. Genet.*, **13**, 469–471.
4. Ebers, G.C., Kukay, K., Bulman, D.E., Sadovnick, A.D., Rice, G., Anderson, C., Armstrong, H., Cousin, K., Bell, R.B., Hader, W. *et al.* (1996) A full genome search in multiple sclerosis. *Nat. Genet.*, **13**, 472–476.
5. Kuokkanen, S., Gschwend, M., Rioux, J.D., Daly, M.J., Terwilliger, J.D., Tienari, P.J., Wikstrom, J., Palo, J., Stein, L.D., Hudson, T.J. *et al.* (1997) Genomewide scan of multiple sclerosis in Finnish multiplex families. *Am. J. Hum. Genet.*, **61**, 1379–1387.
6. Broadley, S., Sawcer, S., D'Alfonso, S., Hensiek, A., Coraddu, F., Gray, J., Roxburgh, R., Clayton, D., Buttinelli, C., Quattrone, A. *et al.* (2001) A genome screen for multiple sclerosis in Italian families. *Genes Immun.*, **2**, 205–210.
7. Coraddu, F., Sawcer, S., D'Alfonso, S., Lai, M., Hensiek, A., Solla, E., Broadley, S., Mancosu, C., Pugliatti, M., Marrosu, M.G. and Compston, A. (2001) A genome screen for multiple sclerosis in Sardinian multiplex families. *Eur. J. Hum. Genet.*, **9**, 621–626.
8. Akesson, E., Oturai, A., Berg, J., Fredrikson, S., Andersen, O., Harbo, H.F., Laaksonen, M., Myhr, K.M., Nyland, H.I., Ryder, L.P. *et al.* (2002)

A genome-wide screen for linkage in Nordic sib-pairs with multiple sclerosis. *Genes Immun.*, **3**, 279–285.
9. Ban, M., Stewart, G.J., Bennetts, B.H., Heard, R., Simmons, R., Maranian, M., Compston, A. and Sawcer, S.J. (2002) A genome screen for linkage in Australian sibling-pairs with multiple sclerosis. *Genes Immun.*, **3**, 464–469.
10. Eraksoy, M., Kurtuncu, M., Akman-Demir, G., Kilinc, M., Gedizlioglu, M., Mirza, M., Anlar, O., Kutlu, C., Demirkiran, M., Idrisoglu, H.A. *et al.* (2003) A whole genome screen for linkage in Turkish multiple sclerosis. *J. Neuroimmunol.*, **143**, 17–24.
11. Hensiek, A.E., Roxburgh, R., Smilie, B., Coraddu, F., Akesson, E., Holmans, P., Sawcer, S.J. and Compston, D.A. (2003) Updated results of the United Kingdom linkage-based genome screen in multiple sclerosis. *J. Neuroimmunol.*, **143**, 25–30.
12. GAMES and the Transatlantic Multiple Sclerosis Cooperative (2003) A meta-analysis of whole genome linkage screens in multiple sclerosis. *J. Neuroimmunol.*, **143**, 39–46.
13. Fogdell-Hahn, A., Ligers, A., Gronning, M., Hillert, J. and Olerup, O. (2000) Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. *Tissue Antigens*, **55**, 140–148.
14. Hauser, E.R., Boehnke, M., Guo, S.W. and Risch, N. (1996) Affected-sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet. Epidemiol.*, **13**, 117–137.
15. 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.
16. Lathrop, G.M., Hooper, A.B., Huntsman, J.W. and Ward, R.H. (1983) Evaluating pedigree data. I. The estimation of pedigree error in the presence of marker mistyping. *Am. J. Hum. Genet.*, **35**, 241–262.
17. Brzustowicz, L.M., Merette, C., Xie, X., Townsend, L., Gilliam, T.C. and Ott, J. (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am. J. Hum. Genet.*, **53**, 1137–1145.
18. Abecasis, G.R., Cherny, S.S. and Cardon, L.R. (2001) The impact of genotyping error on family-based analysis of quantitative traits. *Eur. J. Hum. Genet.*, **9**, 130–134.
19. 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.
20. Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.*, **46**, 222–228.
21. Cavanaugh, J. (2001) International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *Am. J. Hum. Genet.*, **68**, 1165–1171.
22. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M. *et al.* (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411**, 599–603.
23. Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H. *et al.* (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, **411**, 603–606.
24. Tsai, Y.Y., Pugh, E.W., Boyce, P., Doheny, K.F., Fan, Y.T., Scott, A.F., St Hansen, M., Oliphant, A., Loi, H., Mei, R. and Puck, J.M. (2003) Replication of linkage on 2p for chronic mucocutaneous candidiasis and thyroid disease using two different high-density SNP genome scan technologies. *Am. J. Hum. Genet.*, **73** (suppl. poster) 1911, 495.
25. Middleton, F.A., Pato, M.T., Gentile, K.L., Morley, C.P., Zhao, X., Eisener, A.F., Brown, A., Petryshen, T.L., Kirby, A.N., Medeiros, H. *et al.* (2004) Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am. J. Hum. Genet.*, **74**, 886–897.
26. John, S., Shephard, N., Liu, G., Zeggini, E., Cao, M., Chen, W., Vasavda, N., Mills, T., Barton, A., Hinks, A. *et al.* (2004) Whole-genome scan, in a complex disease, using 11 245 single-nucleotide polymorphisms: comparison with microsatellites. *Am. J. Hum. Genet.*, **75**, 54–64.
27. Shaw, S. (2003) High density SNP linkage panel for human genetic studies. *Am. J. Hum. Genet.*, **73** (suppl. poster) 1813, 479.
28. Sawcer, S., Maranian, M., Setakis, E., Curwen, V., Akesson, E., Hensiek, A., Coraddu, F., Roxburgh, R., Sawcer, D., Gray, J. *et al.* (2002) A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain*, **125**, 1337–1347.
29. 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.
30. Chataway, J., Feakes, R., Coraddu, F., Gray, J., Deans, J., Fraser, M., Robertson, N., Broadley, S., Jones, H., Clayton, D. *et al.* (1998) The genetics of multiple sclerosis: principles, background and updated results of the United Kingdom systematic genome screen. *Brain*, **121**, 1869–1887.

APPENDIX

Stephen J. Sawcer[1], Mel Maranian[1], Sara Singlehurst[1], TaiWai Yeo[1], Alastair Compston[1], Mark J. Daly[2], Philip L. De Jager[2,3,4], Stacey Gabriel[2], David A. Hafler[2,3,4,5], Adrian J. Ivinson[4,5], Eric S. Lander[2,4], John D. Rioux[2,3,4], Emily Walsh[2], Simon G. Gregory[6], Silke Schmidt[6], Margaret A. Pericak-Vance[6], Lisa Barcellos[7,8], Stephen L. Hauser[7], Jorge R. Oksenberg[7], Shannon J. Kenealy[9] and Jonathan L. Haines[9], on behalf of The International Multiple Sclerosis Genetics Consortium

[1]University of Cambridge Neurology Unit, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK
[2]The Eli and Edythe L. Broad Institute of Harvard and MIT, Cambridge, MA 02139-1561, USA
[3]Center for Neurologic Diseases, Brigham and Women's Hospital, Boston, MA 02115, USA
[4]Harvard Medical School, Boston, MA 02115, USA
[5]Harvard Center for Neurodegeneration and Repair, Harvard Medical School, Boston, MA 02115, USA
[6]Duke University Medical Center, Center for Human Genetics, PO Box 3468, Durham, NC 27710, USA
[7]Department of Neurology, School of Medicine, University of California San Francisco, San Francisco, CA 94143-0435, USA
[8]Division of Epidemiology, School of Public Health, University of California at Berkeley, Berkeley, CA 94720-7360, USA
[9]Center for Human Genetics Research, 519 Light Hall, Vanderbilt University Medical Center, Nashville, TN 37232-0700, USA