

Genomic mismatch scanning: a new approach to genetic linkage mapping

Stanley F. Nelson¹, John H. McCusker², Mark A. Sander³, Yun Kee³, Paul Modrich⁴
& Patrick O. Brown^{1,2,3}

Genomic mismatch scanning (GMS) is a new method of genetic linkage analysis that does not require conventional polymorphic markers or gel electrophoresis. GMS is ideally suited to affected-relative-pair mapping. DNA fragments from all regions of identity-by-descent between two relatives are isolated based on their ability to form extensive mismatch-free hybrid molecules. The genomic origin of this selected pool of DNA fragments is then mapped in a single hybridization step. Here we demonstrate the practicality of GMS in a model organism, *Saccharomyces cerevisiae*. GMS is likely to be applicable to other organisms, including humans, and may be of particular value in mapping complex genetic traits.

¹Departments of Pediatrics and
²Biochemistry, and
³Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305, USA

⁴Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

Correspondence should be addressed to P.O.B.

Linkage analysis using polymorphic DNA markers has allowed many genes affecting human disease susceptibility to be mapped and several to be cloned¹. However, the usefulness of the discrete markers is limited by incomplete informativeness, their sparsity in some regions of the genome and, above all, the need to type a large set of individual polymorphic markers. With conventional mapping methods, the labour increases in proportion to the density of markers typed. In order to map genes responsible for complex genetic traits, including susceptibility to mental illnesses, coronary artery disease, hypertension, diabetes, obesity, asthma and cancer, it is likely that a dense set of polymorphic markers will need to be typed in a large number of individuals²⁻⁴. The task of scoring multiple discrete loci for a sufficiently large population by conventional methods is dauntingly expensive and tedious.

Genomic mismatch scanning (GMS) is a new, highly efficient method of genetic linkage mapping that does not require conventional polymorphic markers or prior map information. The method takes advantage of (i) the virtual certainty that single base pair mismatches will be present in sufficiently large heteroduplexes formed between allelic sequences that are not identical-by-descent (inherited from a common recent ancestor)⁵, and (ii) the ability of mismatch repair proteins to recognize single base pair mismatches in heteroduplexes. GMS directly maps regions of genetic identity-by-descent between two related individuals using only genomic DNA samples from the two individuals. No gel electrophoresis step is needed, and the entire genome is examined in a single hybridization step, at high resolution. In principle, the mapping power provided by GMS approaches the theoretical limit offered

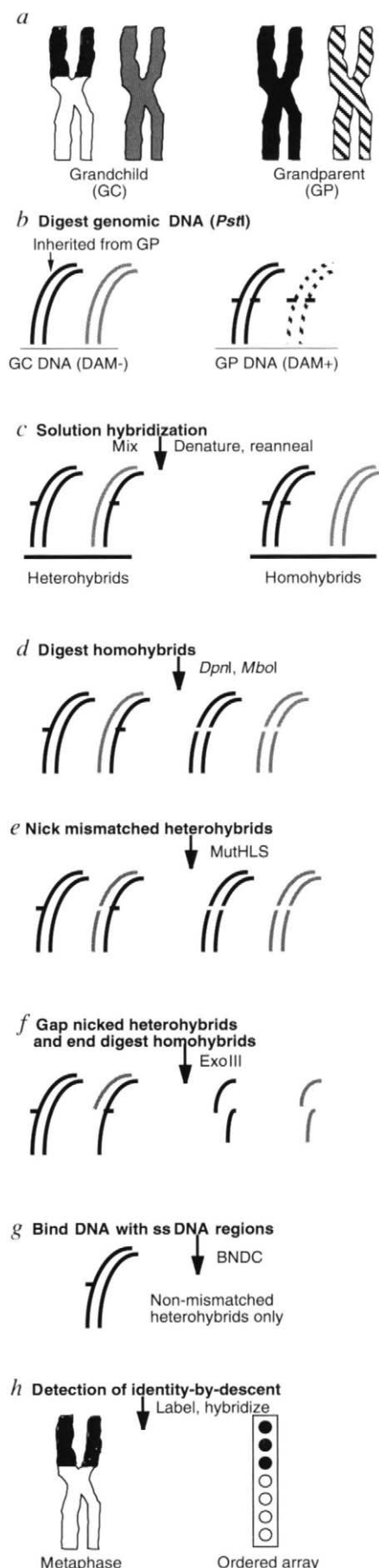
by an arbitrarily dense set of completely informative polymorphic markers⁶. The speed and mapping power of GMS may allow many new applications of linkage mapping, particularly to complex, common and multigenic traits in humans. We demonstrate in this report the ability of GMS to map regions of identity-by-descent between two relatives, using *S. cerevisiae* as a model organism.

The GMS procedure

GMS is designed to map all the regions of genetic identity-by-descent between two related individuals. In western European human populations, allelic sequences from unrelated individuals have single base pair differences on the order of one every 300 base pairs (bp)⁷. Thus, as the haploid human genome consists of about 3×10^9 bp, there are approximately 10^7 potential polymorphisms, of which only a small fraction are recognized by current techniques. GMS is designed to isolate all the regions of the genome that are identical-by-descent, based on the unique ability of DNA from these regions to form large "heterohybrid" DNA duplexes (one strand from each individual) that are free of mismatches⁵. This GMS-selected DNA can then be used as a hybridization probe onto an ordered array of DNA samples representing intervals of the genetic map, to map the regions of identity-by-descent.

Enrichment for identical-by-descent DNA is achieved in two steps: (i) heterohybrids are purified by using a combination of a restriction methylase and methylation-sensitive endonucleases, (ii) heterohybrids that contain mismatches are nicked by the *Escherichia coli* MutHLS enzyme system and subsequently eliminated from the pool, leaving only mismatch-free heterohybrids.

Fig. 1 Genomic mismatch scanning.



The procedure is illustrated using a hypothetical human grandchild(GC)-grandparent(GP) pair as an example (Fig. 1a). High molecular weight genomic DNA is prepared from the two related individuals. Both genomic DNAs are digested with the same restriction endonuclease (for example, *Pst*I) to yield fragments 2–20 kilobases (kb) in length with protruding 3' ends (Fig. 1b). The 3' protruding ends serve to protect these ends from digestion by exonuclease III (*Exo*III) in later steps. The DNA fragments need to be long enough so that most fragments are highly likely to contain both a GATC sequence and at least one base that differs between unrelated allelic fragments. However, the fragments must generally be small enough to be reannealed in solution. Numerous restriction enzymes can satisfy these criteria.

To differentiate the two DNAs, one individual's DNA (GP, in this example) is fully methylated at all GATC sites with the *E. coli* Dam methylase (DAM+). The other individual's DNA (GC) remains unmethylated (DAM-) at all GATC sites (Fig. 1b). The two restriction-digested genomic DNA samples are then mixed in equimolar ratios, denatured and allowed to reanneal (Fig. 1c). An adequate rate of reannealing of large fragments can be achieved in a genome as complex as the human genome using the formamide phenol emulsion reassociation technique (FPERT) which promotes rapid hybridization of allelic restriction fragments up to 20 kb^{8,9}. Approximately 50% of the reannealed fragments are heterohybrids and are, therefore, hemimethylated (methylated on only one strand) at each GATC sequence. "Homohybrids", formed from two strands of the same individual's DNA, are either methylated on both strands or neither (Fig. 1c). Digestion of the reannealed DNA with both *Dpn*I and *Mbo*I, which cut at fully methylated and unmethylated GATC sites, respectively, results in cleavage of the homohybrids to yield smaller duplexes with either blunt ends or 5' protruding ends, respectively (Fig. 1d). The heterohybrids are resistant to cleavage by both *Dpn*I and *Mbo*I. This step therefore distinguishes between homohybrids and heterohybrids. Other sets of methylases and methylase-sensitive restriction enzymes can be used to select for heterohybrids⁹, but the Dam methylase/*Dpn*I/*Mbo*I set is preferred because it requires only one methylase and results in heterohybrids with hemimethylated GATC sites, which is important in the next step.

Sensitive distinction between mismatch-free heterohybrids and those with base mismatches can be achieved using the methyl-directed mismatch repair system of *E. coli*¹⁰. Three proteins, MutL, MutS and MutH, are sufficient to recognize seven of the possible eight single base pair mismatches in duplex DNA and introduce a single strand nick on the unmethylated strand at GATC sites specifically in the mismatch-containing duplexes^{11,12}. The DNA fragments are incubated with MutL, MutS, and MutH (Fig. 1e). Only heterohybrids that are mismatch-free and, therefore, likely to be from regions of identity-by-descent, escape nicking.

All DNA molecules, except the mismatch-free heterohybrids, can be degraded with *Exo*III, a 3' to 5' exonuclease specific for double-stranded DNA (dsDNA). *Exo*III can initiate digestion at a nick and degrade one strand to yield a single-stranded DNA (ssDNA) gap^{13,14}. *Exo*III can also initiate at the blunt or 5' protruding ends produced by cleavage with *Dpn*I or *Mbo*I, respectively, to degrade the homohybrids to ssDNA. The heterohybrids

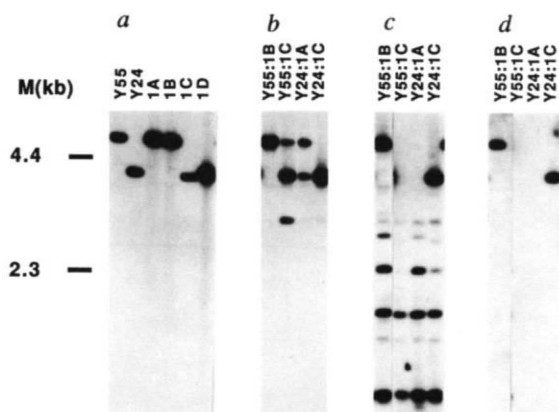


Fig. 2 FPERT reannealing of genomic yeast DNA, heterohybrid protection, and removal of homohybrids from heterohybrids. All of the panels were probed with URA3 DNA under standard conditions³¹. *a*, Southern blot of *Pst*I-digested genomic DNA (0.5 µg) from the parents and 4 spore clones from tetrad 1 showing the inheritance pattern of the URA3 gene. *b*, Four of the eight possible parent/spore clone pairs were hybridized by FPERT; one tenth of the 10 µg rehybridized DNA was loaded. *c*, FPERT reannealed genomic DNAs after digestion with *Mbo*I and *Dpn*I, one tenth was loaded. *d*, The DNA was then treated with *Exo*III and BNDC, one sixth of the remaining DNA was loaded. *c* and *d* were exposed to film for approximately three times as long as *a* and *b* to highlight the efficiency with which the *Mbo*I/*Dpn*I-digested homohybrids were eliminated by *Exo*III digestion and adsorption to BNDC.

have only 3' protruding ends, which resist degradation by *Exo*III. Thus, only DNA duplexes that are both heterohybrids and free of mismatches avoid digestion by *Exo*III (Fig. 1*f*). Following treatment with *Exo*III, all of the DNA molecules with ssDNA regions can be removed by adsorption to benzoylated naphthoylated DEAE cellulose (BNDC). BNDC, at high salt concentrations, retains ssDNA and does not bind dsDNA¹⁵. Thus, the full-length unaltered heterohybrids are purified from the other fragments (Fig. 1*g*).

The GMS procedure can collect information from every type of DNA polymorphism at a resolution as fine as the size of the initial restriction fragments. Single base pair differences and small insertions or deletions are efficiently detected by the mismatch repair proteins¹⁶. Moreover, allelic DNA strands that differ by large insertions or deletions form heterohybrids with extensive ssDNA regions and are efficiently removed by adsorption to BNDC (Fig. 2). We estimate that a ssDNA gap of approximately 100 bp is sufficient for a predominantly dsDNA molecule to be efficiently bound by BNDC (data not shown).

The recovered DNA fragments, representing regions of identity-by-descent, are labelled and used to probe an ordered array of DNAs representing the entire genome. Regions of hybridization indicate regions of identity-by-descent and the boundaries indicate sites of meiotic recombination (Fig. 1*h*).

Genomic mismatch scanning in yeast

Y55 and S288c are two independently isolated laboratory strains of the yeast *S. cerevisiae* that have base pair differences approximately every 100 bp as estimated from restriction fragment length polymorphism (RFLP) analysis (data not shown). Strains from the Y55 and S288c genetic

backgrounds, Y55-160 (HO *his4 leu2 cry1^r MAT α thr4 CAN1^s ura3 GAL2 LYS5*, called Y55), and YJM24 (ho *HIS4 LEU2 CRY1^s MAT α THR4 can1^r URA3 gal2 lys5*, called Y24), were mated by crossing haploid spores of Y55 with the haploid Y24. The resulting hybrid diploid was sporulated (induced to undergo meiosis), and multiple tetrads were dissected. We tested the ability of GMS to identify the map intervals that each spore clone inherited from each parent.

The DNA remaining after each step of the procedure was monitored by gel electrophoresis and Southern blotting. Filters were probed with selected cloned single copy DNA sequences to assess the efficiency and specificity of the selection. Reannealing of genomic yeast DNA proceeded efficiently, and heterohybrids were separated from homohybrids by hemimethylation protection (Fig. 2). Pairwise combinations of a parent and a spore clone were analyzed for identity-by-descent by GMS, using *Pst*I as the initial restriction enzyme. The parental DNAs were each methylated to completion. The genomic DNA samples were analysed by probing with a cloned gene, URA3, which reveals a *Pst*I RFLP (Fig. 2*a*). FPERT resulted in 30–60% recovery of denatured restriction fragments as duplex DNA, in 12–24 hour solution hybridizations (Fig. 2*b*). When the clones in a pair were genetically non-identical at URA3 and had different length *Pst*I fragments, as in Y55:1C and Y24:1A, only homohybrids migrated at the original allele positions (Fig. 2*b*), and these were completely digested by *Mbo*I or *Dpn*I (Fig. 2*c*). An unidentified faster-migrating band, which is apparent only in the Y55:1C pair, is presumed to represent heterohybrid DNA and is removed by subsequent steps. When the two clones in a pair were genetically identical at this locus (such as, Y55:1B and Y24:1C), about 50% of the duplexes after FPERT were heterohybrids and 50% were homohybrids, as assayed. In this case heterohybrids were not resolved from homohybrids on the gel, as they differ only in their methylation pattern (Fig. 2*b*). The homohybrids were efficiently digested by *Mbo*I and *Dpn*I to smaller fragments while the heterohybrids remained full-length (Fig. 2*c*). The smaller *Mbo*I/*Dpn*I-cleaved products were completely removed by the subsequent *Exo*III digestion followed by adsorption of ssDNA to BNDC, while the hemimethylated heterohybrid molecules (Y55:1B and Y24:1C) were undigested (Fig. 2*d*). Thus, heterohybrids can be efficiently purified from homohybrids that form during genomic reannealing.

In order to assay the complete GMS selection procedure, including Mut protein recognition and nicking of mismatched heterohybrids, four complete GMS-selected parent/spore clone pairs from tetrad 1 were analysed by Southern blotting. Multiple *Pst*I restriction fragments were monitored by hybridization with probes from four genetic loci. For every restriction fragment at each locus, DNA was recovered if, and only if, the two clones in the comparison were genetically identical at that locus, as determined by growth on test media (Fig. 3). The specificity of the final selection step was excellent, yielding a 20–100-fold enrichment of fragments from sites of identity-by-descent over a large range of fragment sizes from 700 bp to 9 kb (data not shown). The yield of DNA from the entire procedure was approximately 20–40% of the theoretical maximum. Most of the loss occurred at the FPERT hybridization step. The high proportion of DNA that survived the subsequent steps highlights the

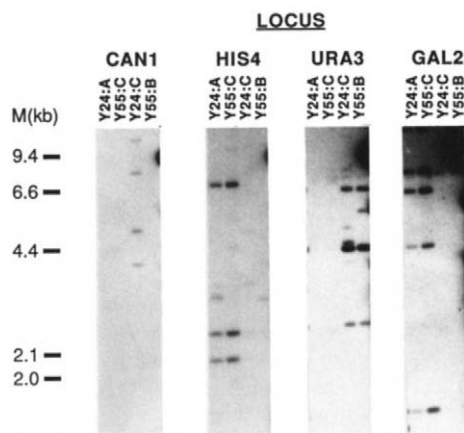


Fig. 3 Southern analysis of GMS selection at four loci. Four of the possible eight parent/spore clone pairs from tetrad 1 were subjected to complete GMS selection (see, Methodology). Heterohybrids were formed as indicated in Fig. 2c. One sixth of the 10 µg GMS-selected DNA was analyzed by Southern blotting. After transfer, the same filter was sequentially hybridized with radiolabelled probes prepared from each of 4 λ clones containing 15–20 kb inserts that span the auxotrophic or drug resistance markers, CAN1, URA3, HIS4, and GAL2 (ref. 19). The genotypes as determined by growth on test media are: Parental clones Y24 (can1⁺ URA3 HIS4 gal2), Y55 (CAN1⁺ ura3 his 4 GAL2); tetrad 1 spore clones 1A (CAN1⁺ ura3 HIS4 gal2), 1B (can1⁺ ura3 HIS4 gal2), 1C (can1⁺ URA3 his4 GAL2). Several *Pst*I restriction fragments were detected by the probes used at each locus. Several faint bands are detected with the HIS4 probe and represent crosshybridization of this probe with unlinked loci.

selectivity of the enzymatic steps. Furthermore, the recovery of linear dsDNA from BNDC was excellent, while virtually all ssDNA remained bound.

GMS selection is highly reproducible. Southern analysis of GMS-selected DNA has been repeated numerous times with parent/spore clone pairs from 5 independent tetrads, probing multiple loci with consistent results. *Hha*I, the only alternative to *Pst*I that we have tested for the initial restriction digestion, also worked well for GMS (data not shown).

An essential feature of GMS, as originally conceived, is the ability to scan the entire genome for identity-by-descent in a single hybridization step. To test this step we used an ordered array of 31 λ phage clones, representing most of yeast chromosome 5, as a sample hybridization target. Identity-by-descent maps of chromosome 5 were determined by GMS for pairwise combinations of Y55 and Y24 with each spore clone from tetrad 6 (Fig. 4). Unambiguous allele assignments could be determined from the intensity of the hybridization signal at each locus as compared to the signal obtained using total yeast DNA as a probe. The normalized hybridization signal was generally about 5–10-fold greater at genetically identical loci than at genetically non-identical loci, when selected DNA was hybridized to this ordered array of lambda clones (Fig. 4b). This can be compared with the 20–100-fold enrichment for genetically identical fragments, as determined by Southern analysis of the same GMS-selected samples (Fig. 3). The discrepancy, which may reflect crosshybridization from unlinked loci (for example, Fig. 3, His4 panel), is still under investigation.

Six λ clones in the initial array always gave positive and

5 clones always gave negative hybridization results. These clones thus provided no linkage information and have not been included in the array displayed in Fig. 4. Consistently negative clones may have resulted from a *Pst*I restriction fragment too long (>15 kb) to be reannealed efficiently. Consistently positive clones may have been due to the presence of one of the low level repetitive elements (for example, a δ element), significant cross hybridization to unlinked loci, a *Pst*I restriction fragment with no differences between the Y55 and the Y24 alleles, or a *Pst*I fragment without a GATC site. For example, seven of the 71 *Pst*I fragments from chromosome 3 contain no GATC sites¹⁷. These seven fragments include 2% of the chromosome. Use of a restriction enzyme that cuts less frequently than *Pst*I (such as, *Sac*I) would reduce the frequency of fragments without a GATC site.

While these uninformative clones created small gaps in the map, the presence of closely linked markers on either side allowed genotypes to be inferred with considerable confidence at the small gaps (~3–4 centiMorgans (cM)). The 52 kb (~18 cM) gap between clones 28 and 29 decreased the informativeness of this region. Suppressing cross-hybridization with excess unlabeled repetitive yeast DNA, spotting subclones from the uninformative regions, or using an alternative restriction enzyme for the initial cleavage may allow gaps of this kind to be eliminated, thereby decreasing the interval between effectively mapped loci.

The data from GMS-selected DNA hybridization to the filter array is summarized schematically in Fig. 5, indicating the resolution of the recombination breakpoints, as well as the gaps in the λ clone coverage. The genotypes assigned by GMS were in complete accord with the results of plate tests for URA3 and CAN1, and 7 RFLP analyses (Fig. 5, data not shown). The genetic length of chromosome 5 is approximately 210 cM so we expect on average 2 meiotic recombination breakpoints per spore clone¹⁸. Indeed, in our small sample there were 3, 2, 1 and 2 breakpoints on chromosome 5 in the 4 spore clones, for an average of 2. The reciprocal products of each recombination event were readily apparent in this analysis, and independent comparisons between a spore clone and each of the two parents gave the expected complementary result. Isolated strong hybridization signals in the midst of a series of negative signals, at clones 20 and 30 in the Y55:6B pairwise comparison and clone 30 in the Y24:6A comparison, may represent gene conversion events in the corresponding intervals (Fig. 4b and Fig. 5).

A set of cloned sequences providing nearly complete coverage of the yeast genome is available¹⁹. Thus, all chromosomes could be represented at about 5 cM resolution with about 800 clones, which can be arrayed on a single 8×12 cm filter, allowing the entire genome to be screened at once. This would be of considerable value even in yeast, which is renowned for its tractability in genetic linkage analysis. For instance, the mapping of multigenic and quantitative traits in yeast should be feasible with GMS.

The limited genetic resolution and the variability of the hybridization signal in yeast is due in part to two factors: (i) the molecular units on which GMS selection operates (the *Pst*I fragments) cover a relatively large genetic interval of approximately 1–3 cM (3–10 kb), and (ii) the target for detection (λ clone inserts) is about 7 cM (20 kb), based on a genome average in yeast of 2.92 kb per cM¹⁸. Therefore,

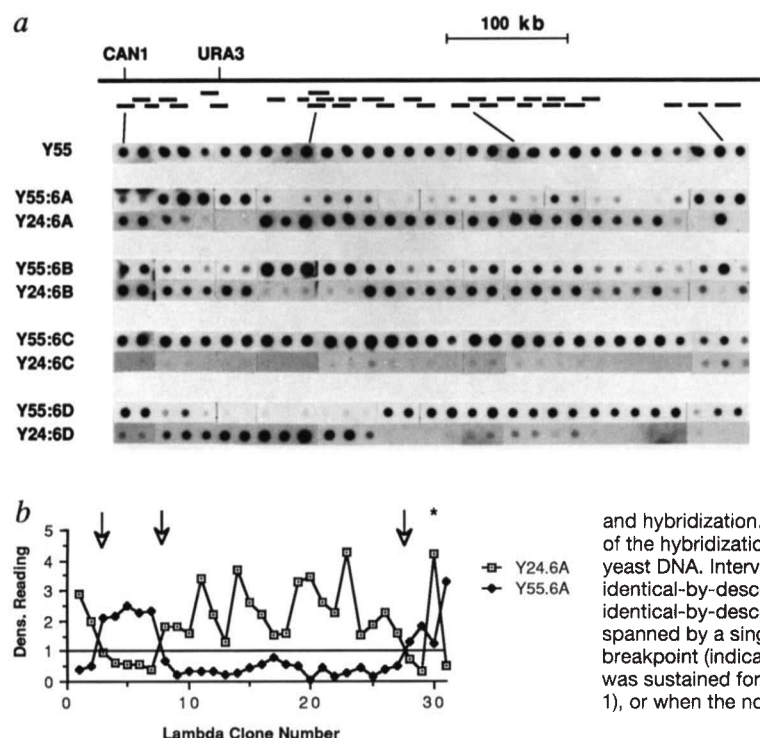


Fig. 4 Hybridization of GMS-selected DNA from parent/spore clone pairs onto an ordered array of cloned DNAs. Thirty-one λ clones represent 79% coverage of chromosome 5 (ref. 19). **a**, The intervals spanned by each λ clone are indicated by bars above the dot blot. Samples on the dot blot are ordered according to chromosome position. The map positions of CAN1 and URA3 are indicated. GMS-selected DNA from each pairwise comparison of a parental clone and a spore clone from tetrad 6 was labelled and used to probe the λ clone filter³¹. Hybridized filters were exposed to film for variable times to obtain comparable signal intensities. Row 1, Y55, shows results of hybridization using unselected total genomic DNA from clone Y55 as a probe, to control for variability in the amount of DNA immobilized at each dot. Rows 2–9 show results of hybridizations using GMS-selected DNA from the indicated parent:spore clone pair. Normalized densitometer reading of the hybridization signals from the Y24:6A and Y55:6A pairwise comparisons are graphed in **b**. The raw densitometry reading for each spot was divided by the reading obtained for that spot when total yeast DNA was used as a probe, to compensate for variability in labelling and hybridization. The resulting value was then divided by the densitometer reading of the hybridization signal from the GMS-selected probe at a spot containing total yeast DNA. Intervals yielding normalized values greater than 1 were scored as identical-by-descent, and intervals yielding values less than 1 were scored as non-identical-by-descent. However, because double recombination events in the interval spanned by a single λ clone (7 cM) are presumably rare, a meiotic recombination breakpoint (indicated by arrow) was inferred only when a significant change in signal was sustained for at least two consecutive clones (that is, either both > 1 or both < 1), or when the normalized signal had a value > 4 (asterisk).

artificial variations in signal intensity (due to the variable size of *Pst*I fragments, etc) are readily apparent, as most of the hybridization signal at each locus is contributed by one or a few *Pst*I fragments. It is likely that these artifactual differences will be averaged out in larger genomes in which many more restriction fragments are present per cM. For instance, in humans, 1 cM corresponds on average to about 1,000 kb, or about 200 *Pst*I restriction fragments.

GMS mapping in humans of affected relative pairs

It remains to be demonstrated that GMS can be applied to humans. The principal obstacles arise from the greater complexity and the high density of repetitive elements in the human genome compared to yeast. These features could interfere with the efficiency and specificity of the two hybridization steps. Preliminary investigations suggest that these factors should in fact pose minimal problems. First, solution hybridization using FPERT allows efficient reannealing of restriction fragments up to 15 kb even when they contain *Alu* elements (S.F.N. & P.O.B., unpublished results). Second, we and others have successfully used probes approaching the complexity of the complete human genome for *in situ* hybridization to metaphase chromosomes^{20,21} (S.F.N. & P.O.B., unpublished results). Our current efforts are therefore focused on applying GMS to human affected-relative-pair mapping.

The difficulty of collecting appropriate pedigrees can be a significant impediment to mapping genes that influence human traits. For medically significant traits, while affected individuals tend to present themselves for examination, other family members need to be traced and recruited. Moreover, individuals vital to pedigree analysis may be deceased. Therefore, it is generally much easier to collect many pairs (or small sets) of related individuals who share a trait of interest than to collect a few large, well-documented pedigrees in which the trait is segregating. In

general, much of the linkage information that can be obtained from a family with an unusual genetic trait is derived from the affected individuals^{2–4,6}. This is particularly true when the penetrance of the allele that confers the trait is low or age-dependent, as is frequently the case for disease-susceptibility genes. Thus, determining where on the genetic map a pair of affected relatives have inherited identical sequences from a common source, and disregarding other family members, is in principle a highly efficient and cost-effective strategy for extracting linkage information from a pedigree^{2–4,6}.

The identity-by-descent maps from multiple pairs of similarly-affected relatives can be combined and the resulting composite map searched for loci where genotypic concordance between affected relatives occurs more frequently than would be expected by chance (Fig. 6a,c). Statistical methods to determine appropriate thresholds for declaring linkage based on GMS data are currently under development²². Preliminary calculations indicate that a simple rare trait can be mapped to a unique interval in the human genome with as few as 10 grandparent-grandchild pairs or 6 first cousin pairs if a maximally informative method, like GMS, is used (Fig. 6a,b). With a sufficiently large number of affected relative pairs, such an analysis can reveal the positions of genes that account for even a slight fraction of the susceptibility to the trait when a trait is genetically heterogeneous (Fig. 6d). The number required is a function of many variables, including the genetic relationships of the affected pairs, the relative risk of the trait in individuals with a "susceptible" allele at the locus of interest compared with the source population, and the fraction of affected cases attributable to a particular locus (which depends on both genetic heterogeneity and frequency of phenocopies or misdiagnoses).

A critical determinant of the number of pairs required for affected-relative-pair mapping is the density and degree of polymorphism of the markers to which linkage is

sought^{2-4,6}. For practical purposes, the affected-relative-pair strategy requires a large number of highly-informative polymorphic markers, closely spaced throughout the genome. With conventional markers, with average polymorphism information content (PIC) < 0.6, additional family members generally need to be typed to insure appropriate allele assignment²³. By providing information from a dense set of completely informative markers, GMS should minimize the number of family members that need to be typed.

For human gene mapping, the most readily available ordered array at present is a preparation of metaphase chromosomes, which can be from any cytogenetically normal source. Fluorescent *in situ* hybridization using GMS-selected DNA as the probe should reveal the physical locations of the regions of identity-by-descent. Even though the probe mixture is extremely complex, specific hybridization can be achieved under conditions that suppress non-allelic cross hybridization due to repetitive elements^{20,21}. Complete ordered arrays of cloned or amplified DNA sequences representing specific map intervals, immobilized on a solid support, will soon become available²⁴⁻²⁶. These should allow automated collection and analysis of the GMS hybridization data, further enhancing its efficiency.

Advantages of GMS

GMS results in a great increase in the effective number of informative markers without a corresponding increase in the number of individual tests, since all the markers are screened in parallel in a single procedure. By requiring fewer total individuals for linkage mapping and many fewer discrete analyses, affected-relative-pair mapping with GMS could greatly reduce the cost and labour involved in developing the human genetic map. Furthermore, scanning the entire genome by means of a single hybridization to an ordered array, without the need for any gels, is a procedure that lends itself well to high volume automation. Because analysis of as few as two affected relatives in each family provides useful map information, the number of families that could potentially be used for mapping studies is greatly

increased. Therefore, GMS may allow large populations to be used for linkage mapping of genetically heterogeneous or quantitative traits in humans.

Other applications of GMS

Questions about meiotic recombination can also be addressed, since each pairwise analysis allows sites of meiotic recombination to be mapped. Grandparent-grandchild pairs are particularly informative in this regard, as their identity-by-descent maps specifically identify the sites of meiotic recombination in the corresponding parent. The relationship between the genetic and physical

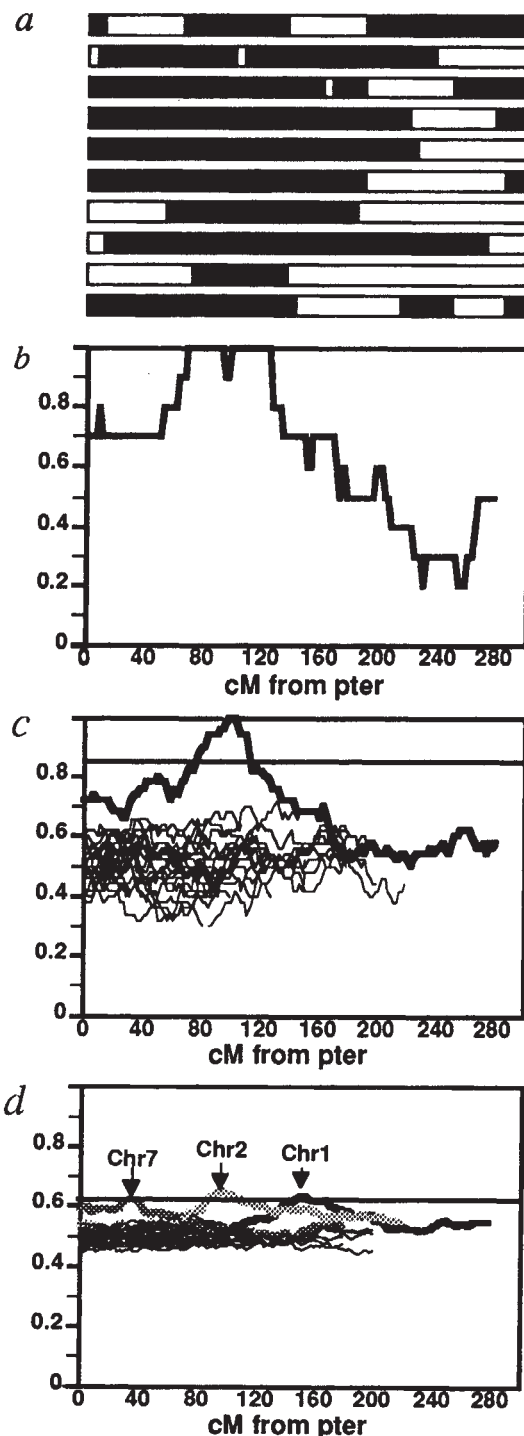


Fig. 6 Identity-by-descent mapping with affected-relative-pairs. The graphs illustrate some characteristics of the data expected from affected-relative-pair mapping by GMS. To generate these graphs, the genetic events that determine the location of identity-by-descent were simulated by a computer program (P. O. B., available upon request). Recombination was assumed to occur randomly, with a probability of 0.01 per 1 cM interval per meiosis. Interference and sex differences in recombination rates were neglected. In a-c, a rare dominant allele 100 cM from the short arm terminus (pter) of chromosome 1 was assumed to be responsible for all cases of a trait shared by the affected pairs. Each bar in a depicts one of 10 independent GP/GC pair simulations. Black indicates the regions of chromosome 1 that are identical-by-descent in each pair. In b-d, cumulative maps of the GP/GC pairs are plotted: the x-axis represents the genetic distance from the pter; the y-axis represents the fraction of pairs identical-by-descent at a given map location. b is derived from the 10 pairs shown in a. Only results for chromosome 1 are plotted. c shows the cumulative map derived from 50 GP/GC pairs with graphs for all 22 autosomes and the X chromosome superimposed. Chromosome 1 is in bold for clarity. Here and in d, the horizontal solid line indicates the 95% confidence level for linkage to the trait in question²². d, Identity-by-descent mapping of a complex trait. Here our model assumed a rare trait that can result independently with equal likelihood from dominant alleles at three different loci (160 cM from 1pter, 100 cM from 2pter, and 40 cM from 7pter), and 20% of cases of the trait are assumed to be unrelated to any of the three major genes. A map derived from 500 GP/GC pairs is shown. The curves for chromosomes 1, 2, and 7 are indicated by arrows. Statistical approaches specifically directed at genetically heterogeneous traits can result in power to detect linkage substantially greater than that reflected by the threshold indicated in d.

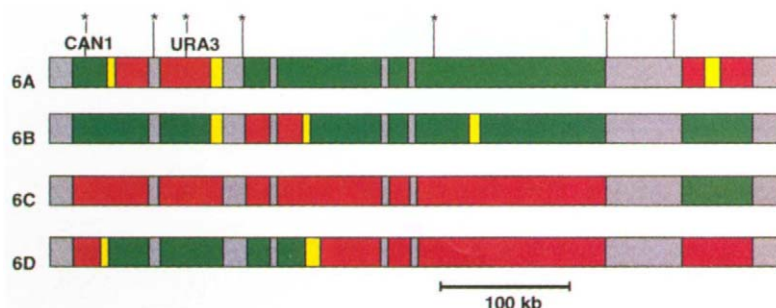


Fig. 5 Inheritance of chromosome 5 in spore clones from tetrad 6. Genotypes were determined according to empirical criteria described in Fig. 4 legend (using data from densitometric scanning of the dot blots shown). The information gathered by GMS from this single tetrad is presented. Where no informative λ clones were available to represent an interval, the chromosome is coloured grey. Red indicates intervals inherited from Y55. Green indicates intervals inherited from Y24. Yellow indicates an ambiguous genotype (see text). The positions of the CAN1 and URA3 genes, and 7 scored RFLPs are indicated.

map, the locations of sites of enhanced or diminished recombination, the effects of age, sex and other factors on the frequency and distribution of meiotic recombination events, and the relationship between recombination and non-dysjunction can be readily investigated by GMS.

GMS is in principle applicable to any species. As it does not rely on any previously-mapped genetic markers, one could use this method to develop simultaneously the genetic and physical maps of a genome for which little or no prior map information is available. The method may therefore be valuable in developing the genetics of economically-important plant or animal species. For instance, linkage of interesting traits to anonymous cloned DNA sequences could be determined by hybridization to unordered arrays of DNA clones. Evidence for genetic linkage among groups of anonymous clones themselves would be generated in parallel with data for linkage between a genetic trait and a group of clones, thus allowing the clones to be ordered into a genetic map.

Methodology

GMS conditions. High molecular weight genomic yeast DNAs were prepared in a standard manner²⁷. Restriction enzymes and Dam methylase (New England Biolabs) conditions were as recommended by the manufacturer. FPERT of parent/spore clone pairs was performed as follows: 5 μ g of *Pst*I-digested genomic parental DNA (Y55 or Y24) was methylated to completion with 10 units Dam methylase, 80 μ M S-adenosylmethionine, in a volume of 40 μ l in the supplied buffer. The methylated *Pst*I-digested parental DNA was mixed with 5 μ g *Pst*I-digested genomic DNA from a spore clone in 100 μ l volume, denatured by adding NaOH to 0.3 M for 15 min at RT, then neutralized to pH 8.0 by adding MOPS acid to 0.4 M. The final volume was adjusted to 400 μ l in 2M sodium thiocyanate, 10 mM tris, pH 8.0, 0.1 mM EDTA and 8% formamide²⁷. Water saturated phenol was added just until an emulsion formed, about 150 μ l. The samples were agitated continuously to maintain the emulsion for 12–24 h at RT. DNA was recovered by chloroform extraction twice

and ethanol precipitation. FPERT reannealed genomic DNAs (about 5 μ g) were digested with 5 U *Mbo*I and 10 U *Dpn*I in 100 mM NaCl, 50 mM tris-HCl, pH 7.9, 10 mM MgCl₂, 1 mM DTT for 2 h at 37 °C. After *Mbo*I/*Dpn*I digestion, the genomic DNA pairs were incubated with the MutH, MutL, and MutS proteins, purified as described^{13,28,29}. The DNA mixture (about 2.5 μ g) was incubated with 40 ng MutH, 2.5 μ g MutL, 5.2 μ g MutS, 2 mM ATP, in 20 mM KCl, 50 μ M Hepes(KOH), pH 8.0, 5 mM MgCl₂, 1 mM DTT, 50 μ g ml⁻¹ BSA for 30 min at 37 °C in a total volume of 360 μ l. The reaction was stopped by heating to 68 °C for 10 min, and the volume was increased to 400 μ l with 10 mM MgCl₂, 50 mM tris-HCl, pH 8.0. The sample was then incubated with 10 U *Exo*III for 10 min at 37 °C, phenol/chloroform extracted once, adjusted to 1 M NaCl, and mixed with moist BNDC (100 mg), previously equilibrated with 50 mM tris-HCl, pH 8.0, 1M NaCl. The mixture was agitated for 1–4 h at RT. The BNDC was pelleted at 14,000 rpm for 3 min, and the supernatant recovered, chloroform extracted and precipitated. The GMS-selected DNA was resuspended in 90 μ l 10 mM trisCl, pH 8.0, 1 mM EDTA.

Southern analysis. Each step of the GMS selection was analysed by DNA electrophoresis on a 0.7% agarose gel, transferring to nylon (Hybond-N, Amersham), UV crosslinking, and hybridizing with the indicated DNA probe. All probes were labelled by random primer incorporation of [α^{32} P]dCTP³⁰. Hybridizations were performed at 42 °C for about 24 h in 3 $\times$ SSC, 1% SDS, 50% formamide, 5 $\times$ Denhardt's solution, 8% dextran sulfate, and 100 μ g ml⁻¹ sheared salmon sperm DNA. Final washes were in 0.2 $\times$ SSC/0.1% SDS at 60 °C for 15 min, twice. Washed membranes were exposed to Kodak XAR film for variable lengths of time. For filters probed multiple times, probes were stripped with boiling 0.1 $\times$ SSC/0.1% SDS and agitated for 30 min.

λ filter arrays. λ clones were from a collection assembled by M. V. Olson¹⁹. Unique ID numbers for the 31 λ clones are in order from left to right in Fig. 4: 4722, 5917, 2340, 6662, 2183, 4057, 6150, 5884, 6977, 7150, 5961, 6224, 5083, 4049, 6793, 4581, 4678, 6592, 4742, 6019, 7108, 7088, 2650, 6052, 3955, 6116, 7084, 4394, 7221, 6197, 4783. Specific locus clones are: HIS4=4711, GAL2=6637, CAN1=4722, URA3=6150. Each hybridization dot contains approximately 50 ng denatured λ clone DNA, vacuum-blotted onto nylon.

Received 11 January; accepted 9 February 1993.

Acknowledgements

We thank Sue Klapholz, Karen Au, Taiyun Roe, Sam Chow, Joanna Mountain and John Mulligan for helpful discussions and comments on the manuscript. J.H.M. thanks R. W. Davis for support and encouragement. S. F. N. was supported by a Physician Scientist Award from NICHD. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute. This research was supported by the Howard Hughes Medical Institute and grants from the NIH.

1. Botstein, D., White, D. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. hum. Genet.* **32**, 314–331 (1980).
2. Risch, N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. hum. Genet.* **46**, 229–241 (1990).
3. Risch, N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. hum. Genet.* **46**, 242–253 (1990).
4. Bishop, D. T. & Williamson, J. A. The power of identity-by-state methods for linkage analysis. *Am. J. hum. Genet.* **46**, 254–265 (1990).
5. Sanda, A. I. & Ford, J. P. Genomic Analysis I: inheritance units and genetic selection in the rapid discovery of locus linked DNA markers. *Nucl. Acids Res.* **14**, 7265–7283 (1986).
6. Lander, E. S. & Botstein, D. Mapping Complex Genetic Traits in Humans: New Methods Using a Complete RFLP Linkage Map. *CSHSQB*. **51**, 49–62 (1986).
7. Cooper, D. N. *et al.* An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**, 201–205 (1985).
8. Kohn, D. E., Levinson, S. A. & Byers, M. J. Room temperature method for increasing the rate of DNA reassociation by many thousandfold: The phenol emulsion reassociation technique. *Biochemistry* **16**, 5329–5341 (1977).
9. Casna, N. J., Novack, D. F., Hsu, M. & Ford, J. P. Genomic analysis II: isolation of high molecular weight heteroduplex DNA following differential methylase protection and formamide-PERT hybridization. *Nucl. Acids Res.* **14**, 7285–7303 (1986).
10. Lahue, R. S., Au, K. G. & Modrich, P. DNA mismatch correction in a defined system. *Science*. **245**, 160–164 (1989).
11. Au, K. G., Welsh, K. & Modrich, P. Initiation of methyl-directed mismatch repair. *J. biol. Chem.* **267**, 12142–12148 (1992).
12. Su, S., Lahue, R. S., Au, K. G. & Modrich, P. Mismatch specificity of methyl-directed DNA mismatch correction in vitro. *J. biol. Chem.* **263**, 6829–6835 (1988).
13. Welsh, K. M., Lu, A., Clark, S. & Modrich, P. Isolation and characterization of the *Escherichia coli* mutH gene product. *J. biol. Chem.* **262**, 15624–15629 (1987).
14. Henikoff, S. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–360 (1984).
15. Sedat, J. W., Kelly, R. B. & Sinsheimer, R. L. Fractionation of nucleic acid on benzoylated-naphthoylated DEAE cellulose. *J. molec. Biol.* **26**, 537–540 (1967).
16. Learn, B. A. & Grafstrom, R. H. Methyl-directed repair of frameshift heteroduplexes in cell extracts from *Escherichia coli*. *J. Bact.* **171**, 6473–6481 (1989).
17. Oliver, S. G. *et al.* The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38–46 (1992).
18. Mortimer, R. K., Schild, D., Contopoulou, C. R. & Kans, J. A. Genetic map of *Saccharomyces cerevisiae*, edition 10. *Yeast* **5**, 321–403 (1989).
19. Riles, L. *et al.* Physical maps of the six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6-kilobase pairs. *Genetics* (in the press).
20. Meltzer, P. S., Guan, X., Burgess, A. & Trent, J. M. Rapid generation of region specific probes by chromosome microdissection and their application. *Nature Genet.* **1**, 24–28 (1992).
21. Pinkel, D. *et al.* Fluorescent in situ hybridization with human chromosome-specific libraries. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9138–9142 (1988).
22. Feingold, E., Brown, P. O. & Seigmund, D. Gaussian Models for Genetic Linkage Analysis Using Complete High Resolution Maps of Identity-by-Descend. *Am. J. hum. Genet.* (in the press).
23. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome. *Science* **258**, 67–86 (1992).
24. Chumakov, I. *et al.* Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**, 380–387 (1992).
25. Bellanne-Chantelot, C. *et al.* Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* **70**, 1059–1068 (1992).
26. Foote, S., Vollrath, D., Hilton, A. & Page, D. C. The Human Y Chromosome: Overlapping DNA clones Spanning the Euchromatic Region. *Science* **258**, 60–66 (1992).
27. Philippsen, P., Stotz, A. & Scherf, C. DNA of *Saccharomyces cerevisiae*. *Meth. Enzymol.* **194**, 169–182 (1991).
28. Grilley, M., Welsh, K. M., Su, S. S. & Modrich, P. Isolation and characterization of the *Escherichia coli* mutL gene product. *J. biol. Chem.* **264**, 1000–1004 (1989).
29. Su, S. S. & Modrich, P. E. coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5057–5061 (1986).
30. Feinberg, A. P. & Vogelstein, B. Addendum: a technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 266–267 (1983).
31. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).