Identity by Descent Between Distant Relatives: Detection and Applications

Sharon R. Browning[1] and Brian L. Browning[2]

[1] Department of Statistics, University of Washington, Seattle, Washington 98195;
email: sguy@uw.edu

[2] Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington 98195

Annu. Rev. Genet. 2012. 46:617–33

First published online as a Review in Advance on September 17, 2012

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi:
10.1146/annurev-genet-110711-155534

0066-4197/12/1201-0617$20.00

Keywords

cryptic relatedness, coancestry, IBD mapping, haplotype phase inference, effective population size

Abstract

Short segments of identity by descent (IBD) between individuals with no known relationship can be detected using genome-wide single nucleotide polymorphism data and recently developed statistical methodology. Emerging applications for the detected IBD segments include IBD mapping, haplotype phase inference, genotype imputation, and inference of population structure. In this review, we explain the principles behind methods for IBD segment detection, describe recently developed methods, discuss approaches to comparing methods, and give an overview of applications.

Identity by descent (IBD): two alleles or haplotypes that are identical and are inherited from a shared ancestor

IBD segment: a continuous segment over which two haplotypes are identical by descent

INTRODUCTION

Two haplotypes are identical by descent if they share the same alleles inherited from a common ancestor. Identity by descent (IBD) can be considered on various timescales. In this review, we focus on IBD that is due to recent common ancestry. The common ancestry may be known, such as between cousins (pedigree IBD), or the relationship may be unknown, as is likely for tenth cousins. The limits on recent common ancestry are defined by the resolution of IBD segment detection for a data set. If one can detect most IBD segments resulting from common ancestry 25 generations ago, then recent effectively means shared ancestry within the past 25 generations. The idea of a neutral coalescent in population genetics theory implies that all individuals have common ancestry in the distant past, but this is not a focus of this review. Such distant common ancestry is associated with many mutation and recombination events, and so shared ancestry is rarely evident from sequence similarity. In this review, we start with the principles behind methods for inferring segmental IBD sharing and then describe specific methods. We discuss approaches for comparing different methods. We then describe a variety of applications that make use of detected IBD segments. We focus on humans; however, the concepts described in this review also apply to other diploid species. We also focus on IBD between individuals rather than on IBD within individuals [homozygosity by descent (HBD)]. HBD has important application in mapping recessive traits and is estimated using similar methods.

PRINCIPLES BEHIND INFERENCE OF SEGMENTAL IDENTITY-BY-DESCENT SHARING

The key idea behind IBD segment detection is haplotype frequency. If the frequency of a shared haplotype is very small, the haplotype is unlikely to be observed twice in independently sampled individuals, so one can infer the presence of an IBD segment. This criterion can be applied in several ways. The first is length of sharing, which is a proxy for frequency. If two densely genotyped haplotypes are identical at all or most (allowing for some genotyping error) assayed alleles over a very large segment of a chromosome, then the haplotypes are likely to be identical by descent across the whole segment. The second is direct use of haplotype frequency: Shared haplotypes with estimated frequency below some threshold are determined to be identical by descent. The third makes use of a population genetics model to infer probability of IBD. Given the frequency of the shared haplotype and a probability model for the IBD process along the chromosome, one can estimate the probability that the individuals are identical by descent at any position on the segment.

The problem of inferring IBD segments is made more difficult because we do not usually have data at the haplotype level, but instead the data are unphased genotypes. Haplotypes can be estimated from genotype data. Thus, many methods for inferring IBD involve haplotype estimation and may allow for errors in the haplotype phase.

Patterns of Identity-by-Descent Sharing

Closely related individuals have a high proportion of IBD, and their IBD is arranged in long continuous segments. For example, half-siblings (see **Figure 1**) have, on average, IBD sharing across one half of the genome, and the average IBD segment length is 50 centiMorgans (cM) or approximately 50 megabases (Mb). As the relationship becomes more distant, the average proportion of the genome shared identical by descent decreases exponentially, but the average length of an IBD segment, when IBD sharing occurs, remains relatively long. For a pair of individuals separated by m meioses, the average proportion of the genome shared identical by descent (the sum of lengths of IBD segments divided by the length of the genome) owing to that shared ancestry is $2^{-(m-1)}$. In contrast, the lengths of IBD segments owing to the

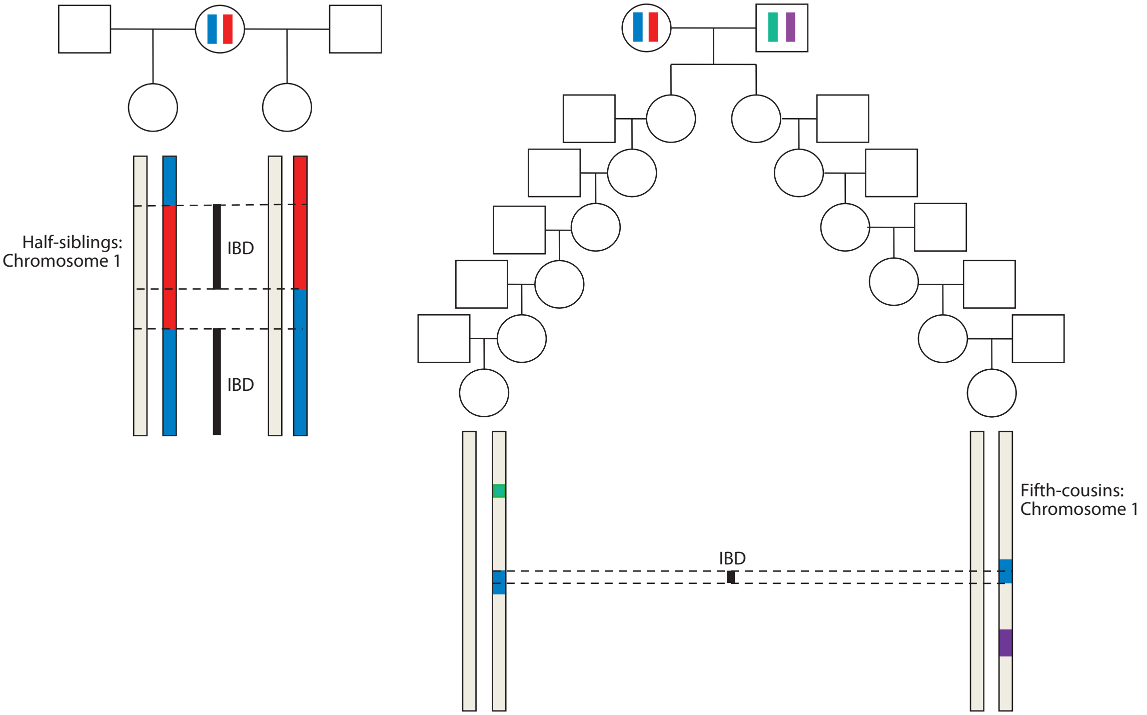

Figure 1

Identity by descent (IBD) on chromosome 1 for half-siblings and fifth cousins. Chromosome 1 is approximately 250 Mb long and has a genetic length of approximately 280 cM. The common ancestors' copies of chromosome 1 are shown in various colors, and tan represents all other haplotypes. Regions of IBD are shown with black bars.

shared ancestry are approximately exponentially distributed with a mean of 100 m^{-1} cM. Thus, for example, fifth cousins (see **Figure 1**) are separated by 12 meioses. On average, 0.05% of their genome, or approximately 1.5 cM (~1.5 Mb), is identical by descent through their great-great-great-great grandmother. If they are full fifth cousins, they may also have IBD sharing through their great-great-great-great grandfather, which doubles the expected IBD proportion to 0.1% of their genome, or approximately 3 cM (~3 Mb). However, fifth cousins usually have no detectable IBD sharing, and when they do have IBD sharing it is usually composed of a single IBD segment with a mean length of 8.3 cM (~8 Mb). Extrapolating further, individuals who have shared ancestry through a certain common ancestor 25 generations ago (with 50 meioses of separation) almost always share none of their genome identical by descent through that ancestor, but if they do have an IBD segment through that ancestor it will have a mean length of 2 cM (~2 Mb).

Any given pair of individuals is related through many common ancestors. For a pair of individuals on different continents, the relationships may be too distant to result in detectable IBD sharing. However, pairs of individuals from the same geographic region may have many recent common ancestors. Such individuals may, however, have only one or two detectable IBD segments, as many of the relationships have not resulted in any IBD sharing.

In a data set with N unrelated individuals, there are $N(N\text{-}1)/2$ pairs of individuals. Although any given pair has very little IBD sharing, the total amount of IBD sharing in the sample, and the total amount per individual, can be large.

Linkage disequilibrium (LD): correlation between genetic markers; usually due to proximal location on a chromosome

Data Quality

IBD segment detection is critically dependent on genotype data accuracy. Although most methods make some allowance for the occasional genotype error, such errors significantly reduce the power to detect an IBD segment (42). Therefore, stringent data quality control should be applied to remove lower-quality variants. For some data with slightly higher rates of genotype error, we recommend first recalling the genotypes using a genotype calling method that utilizes linkage disequilibrium (LD) to significantly improve genotype accuracy and to detect and remove difficult-to-call variants (11).

METHODS FOR IDENTITY-BY-DESCENT SEGMENT DETECTION

Length Threshold on Identity-by-Descent-Consistent Genotypes

Perhaps the simplest approach to IBD segment detection is to look at the genotypes of a pair of individuals (in a later section we review methods for inferring IBD regions shared by multiple individuals) and find very long segments that are consistent with IBD (42, 44, 47, 49, 59, 60). Genotypes are consistent with IBD if they have at least one allele in common. For diallelic markers such as single nucleotide polymorphisms (SNPs), a pair of genotypes is consistent with IBD unless the genotypes are discordant homozygotes. For example, the genotype pair AA/AT is consistent with IBD, but the genotype pair AA/TT is not.

For common SNPs with minor allele frequency of approximately 50%, only 12.5% of SNPs in a non-IBD region can be expected to be discordant homozygotes. For less common SNPs, the proportion of discordant homozygotes in a non-IBD region decreases further. For example, for an SNP with a minor allele frequency of 10%, the frequency of discordant homozygotes is only 1.6%. Thus, a large number of IBD-consistent SNPs are needed to be confident that a continuous IBD segment is present. Furthermore, the problem is exacerbated by LD. Discordant homozygotes tend to cluster along the genome because of LD correlation, as do genotypes consistent with IBD. For example, when two markers are in perfect LD ($r^2 = 1$), concordant homozygotes at one position are always accompanied by concordant homozygotes at the other position. This correlation increases the length of IBD-consistent SNPs required to infer recent IBD.

The length threshold used depends on the desired tradeoff between false-negative and false-positive results. Thus, for example, Miyazawa et al. (47) used a threshold of 3 cM, which gave almost equal weight to false positives and false negatives when using IBD to map the gene for a Mendelian disorder in a large family. In contrast, Kong et al. (42) used a threshold of 10 cM to achieve a very low false-positive rate when using IBD to infer haplotype phase.

Length Threshold on Identity-by-Descent-Consistent Haplotypes

If haplotype phase is available, a length threshold can be applied to IBD-consistent haplotypes. In studies in which close relatives are genotyped, haplotype phase is known at most markers because of Mendelian constraints. Shared segments may be found by direct comparison of haplotypes, and if these segments are long enough they can provide evidence of IBD sharing. For example, Houwen et al. (35) used genotypes on parents and siblings to phase three cases and then identified a long region on which all three shared a haplotype.

GERMLINE (31) extended this approach to unrelated individuals that have been phased with standard methods (16). The algorithm uses a sliding window with a dictionary (hash table) to find matching haplotypes, and it allows for haplotype phase error. GERMLINE is very computationally efficient. It can be applied to data sets with tens of thousands of individuals.

Probabilistic Framework: Without Linkage Disequilibrium

Probabilistic methods rely on a model for IBD status. The IBD status for a pair of individuals can take several forms. It can be binary (presence/absence of IBD); it can take values 0, 1, or 2 (number of pairs of haplotypes shared identical by descent) (52); or it can take multiple values reflecting Jacquard's nine IBD states for unphased data or 15 IBD states for phased data (7, 39, 63). For computational reasons, it is convenient to model the process of change of IBD status with a Markov model. The modeling is always performed on the genetic distance scale (i.e., centiMorgans) so that variation in the rate of recombination is already accounted for and does not need to be directly considered in the model. Transition rates may be assumed a priori or may be estimated from the data. PLINK (52) first estimates the genome-wide IBD sharing proportion for a pair of individuals by means of identity by state and then determines the smallest number of meiosis, *m*, that could, on average, give rise to that level of IBD. Assuming that all the IBD for the pair comes from a single shared ancestor *m*/2 generations ago yields transition probabilities.

When LD is not considered, the model for the data is a hidden Markov model, with the IBD status being the hidden state and the observed genotypes being independent given the IBD status. Hardy-Weinberg equilibrium is typically assumed so that allele frequencies can be easily transformed into genotype probabilities. For example, if the observed genotypes in a pair of individuals at a locus are AA and AT and the frequency of the A and T alleles are f_A and f_T, respectively, the probability of the genotypes is $f_A^2 \times 2 f_A f_T$ if the IBD status is 0 (no IBD), $f_A^2 f_T$ if the IBD status is 1 (one pair of haplotypes shared identical by descent; this pair must be A alleles for these data), and zero if the IBD status is 2 (two pairs of haplotypes shared identical by descent, which is inconsistent with the data). One can also incorporate probabilities of genotype error into these calculations (7, 48).

The PLINK (52) documentation recommends thinning markers to remove markers in LD before applying this type of method. In contrast, IBD_Haplo (7) has been used without thinning markers, which does result in some false-positive IBD. Thinning markers reduces power to detect IBD, so a trade-off of power versus false-positive rate must be made.

Probabilistic Framework: With Linkage Disequilibrium

Probabilistic methods that incorporate LD also use a hidden Markov model framework, but the modeling is complicated by the incorporation of LD. Existing methods use a binary IBD status (5, 15, 24), a ternary IBD status (0, 1, or 2 pairs of haplotypes identical by descent) (5), or Jacquard's nine condensed IBD states (32). Genotype error probabilities can be incorporated in the same manner as for the probabilistic framework without LD (5, 32).

RELATE (5) conditions the probability of a genotype on both the IBD status and the genotype of a nearby marker in greatest LD with the marker under consideration. After removing SNPs with low minor allele frequency and SNPs in near-perfect LD, the method obtains substantially less false-positive IBD than when ignoring LD (5).

IBDLD (32) extends RELATE's approach by using ridge regression to condition on twenty neighboring SNPs. Using simulated data based on HapMap Northern European (CEU) data, with SNPs with low minor allele frequency removed, IBDLD's approach was significantly more accurate than conditioning on the single best SNP (32).

Browning (13) extended the BEAGLE hidden Markov model for LD (12, 14) to include binary IBD status for pairs of haplotypes. The combined IBD and LD model is a hidden Markov model. BEAGLE IBD (15) extends this approach from pairs of haplotypes to pairs of individuals with unphased genotypes. When applied to Northern European Affymetrix 500K SNP data, the method was more powerful and produced fewer false-positive results than a probabilistic method without LD (PLINK; after thinning SNPs to remove LD) (15). However, BEAGLE IBD is computationally intensive and hence is not suitable for genome-wide analyses with thousands of individuals.

Shared Haplotype Frequency Below a Threshold

Because the probabilistic IBD segment detection method BEAGLE IBD (15) is computationally expensive, a new method called BEAGLE fastIBD was proposed (10). BEAGLE fastIBD utilizes the BEAGLE LD model but does not employ a full probabilistic framework. The IBD status process along the chromosome is not modeled, and posterior probabilities of IBD are not calculated. Instead, this model estimates the frequency of a shared haplotype. If two individuals share a haplotype with an estimated frequency below a user-specified threshold, the shared haplotype is estimated to be IBD. This approach is similar to GERMLINE's approach but with thresholding on haplotype frequency rather than haplotype length. As with GERMLINE, BEAGLE fastIBD considers windows of markers and utilizes a hash table in each window to greatly improve computational efficiency. The hash table enables fast identification of all individuals who share a haplotype fragment. The fastIBD algorithm uses multiple estimates of an individual's haplotypes to increase robustness to phase error. The fastIBD method (when properly calibrated) was shown to have similar false-positive rates and power to BEAGLE IBD but with much less computational cost (10).

Methods for Multiple Individuals

In pairwise analysis, some short segments are missed that can be found in multiple-individual analysis by using the information from other individuals identical by descent with the pair. A simple approach to analysis of multiple individuals is to look for identity by state (alleles consistent with IBD) across the set of individuals, either at the haplotypic (35) or genotypic (33, 44, 60) level.

The MCMC_IBDfinder method (48) extends the probabilistic approach without LD to consideration of more than two individuals simultaneously. The method is very computationally intensive and is not suitable for genome-wide analysis of hundreds or thousands of individuals (48). The largest data set analyzed by Moltke et al. (48) with this method consisted of 15 individuals on a single chromosome.

When the individuals of interest have known relationships (i.e., form an extended pedigree), ALADIN (4) obtains probabilities of IBD configurations by approximating the multi-individual IBD status as a first-order Markov chain, with transition probabilities determined by the pedigree structure. Limited LD is incorporated through a cluster approach (1, 8) in which LD is allowed within haplotype blocks but not between haplotype blocks. The method allows for linkage calculations in large pedigrees that would otherwise be intractable.

APPROACHES FOR COMPARING IDENTITY-BY-DESCENT SEGMENT DETECTION METHODS

In order to compare methods for IBD segment detection, three factors need to be considered: power, false-positive rate, and computation time. Power is the most obvious metric; however, it should only be considered in the context of the false-positive rate. Power and false-positive metrics measure the ability of a method to correctly identify the presence of an IBD segment and to avoid inferring an IBD segment when none exists as well as the ability to correctly determine the endpoints of an inferred segment. Even when there is high confidence that two individuals share a haplotype identically by descent in a region, there is often ambiguity regarding the precise endpoints for the shared haplotype.

Many methods have tuning parameters (such as the minimum length of an IBD segment or a threshold on posterior probability of IBD).

Adjusting the tuning parameters can increase power at the cost of increased false-positive rates. Computation time is also significant, as many methods are simply too computationally intensive for application to large data sets.

Assessment of Power

The power of IBD detection can be considered at a segment level, as the proportion of true IBD segments that are at least partially included in an estimated IBD segment for the same pair of individuals. Alternatively, IBD detection power can be considered at the marker level, as the proportion of markers in a true IBD segment that are contained within the estimated IBD segments for the same pair of individuals. The second definition includes consideration of whether the IBD segment lengths are underestimated.

To assess power of a method for detecting IBD segments, one needs to have some data with known IBD segments. One approach is to simulate data with information on genealogy and IBD sharing from the most recent generations recorded. However, it is difficult to generate realistic simulated data. Coalescent approaches usually allow for input of a model for the past demography of a population, including past growth, bottlenecks, founder effects, recent expansion, etc. Such models are usually estimated from allele frequency spectrum data. Previously, SNP array data were used to fit these models, which allows for relatively accurate estimation of ancient effective population sizes, but poor estimation of recent effective population size. Models fit from SNP array data can give current effective population sizes in the tens of thousands. In contrast, allowing for recent super-exponential population growth and utilizing sequence data gives an estimated effective population size for Europe of over one million (20). Data simulated from models with current effective population sizes in the tens of thousands are typically realistic in terms of LD patterns and allele frequencies for common variants, but do not properly model rare variants (20). Similarly, such data are not realistic for recent IBD because population size is directly related to the amount of IBD expected. The expected amount of IBD is inversely proportional to the recent effective population size (18). Thus, simulations with a current effective size less than 100,000 have much more IBD than is seen in real data from large outbred populations.

Another approach to comparing IBD detection methods is to measure the amount of detected IBD in a sample while also obtaining a measure of error rate (see below). The detection rate is not the same as power, as it includes false-positive detection, but it can be used instead of power to compare methods. If one method has both a higher rate of detection and a lower rate of error than another method, then it is superior. Alternatively, if methods are tuned to equalize the error rate, then the detection rate can be used to compare the methods.

Power is a function of many factors, including detection method, genotyping platform, and length of the underlying IBD segments. All methods can find most 10-cM segments using typical genome-wide SNP array data. With million-SNP array data, some methods have reasonably high power to find segments of size 1 cM (S. Browning & B. Browning, unpublished data). Resolution will improve further with sequence data.

Assessment of False-Positive Rate

As with power, false-positive IBD can be considered at a segment level, i.e., as the proportion of estimated IBD segments that are completely wrong. Or false-positive IBD can be considered at the marker level, i.e., as the proportion of markers in an estimated IBD segment that are not contained within a true IBD segment. The second definition includes consideration of whether the lengths of true IBD segments are overestimated. Use of real data to determine the false-positive rate is problematic because there is typically some level of latent IBD in a population, making it difficult to know whether detected IBD is a false positive or not.

IBD mapping: the use of detected IBD segments to determine which regions of the genome are likely to harbor disease-susceptibility variants

In appropriately simulated data (see above), the true IBD status is known. One can obtain a list of true IBD segments of a specified length and check what proportion of estimated IBD segments or markers fall on a true IBD segment.

As an alternative approach to assessing false-positive rates, one can mask a subset of markers, infer IBD, and check whether the genotypes at the masked SNPs are compatible with IBD. One can divide the rate of discordant homozygotes at masked markers in an inferred IBD segment by the rate of discordant homozygotes at the same markers in random pairs of individuals. This normalized discordance gives a useful measure for comparing methods, even though it does not directly give the rate at which inferred IBD is incorrect. Only a small proportion of non-IBD SNPs are discordant homozygotes (as described earlier). However, because this approach can be applied to data from unrelated individuals, very large data sets can be analyzed, resulting in very precise estimates.

Computation Time

All the pairwise IBD detection methods described above (but perhaps not the multi-individual IBD detection methods) are suitable for small-scale analyses, such as those involving hundreds of individuals (tens of thousands of pairs of individuals) genome-wide or those involving larger numbers of individuals on small genomic regions. Several methods have been applied to much larger data sets, such as analyses of thousands of individuals (millions of pairs of individuals) genome-wide, or have been seen to be fast enough for such application. These include PLINK (15, 52), GERMLINE (31), BEAGLE fastIBD (10), and Kong's method (42). These analyses generally require the use of a cluster rather than a single desktop computer. With GERMLINE, the computation bottleneck is the prephasing of the data. For example, to phase 5,000 individuals across 1,000,000 SNPs takes approximately 45 days computing time with BEAGLE on a 2.4 GHz computer, whereas the analysis of the phased data with GERMLINE takes less than 1 h. A single run of BEAGLE fastIBD (ten runs are recommended) takes approximately the same time as phasing the data with BEAGLE (e.g., 45 days for 5,000 individuals on 1,000,000 SNPs). With a cluster, analyses can be parallelized by chromosome or chromosomal segment.

RELATE (5), IBDLD (32), and IBD_Haplo (7) scale linearly in the number of pairs and in the number of markers. Extrapolating from reported computing times (5, 7, 32), the time to analyze all pairs from 5,000 individuals across 1,000,000 SNPs with one of these methods would be on the order of 10,000 days. Thus, these probabilistic methods are best suited to somewhat smaller data sets.

GERMLINE and BEAGLE fastIBD use an efficient dictionary (hashing) approach to finding shared haplotype segments. Whereas most other methods scale linearly (or worse than linearly) in the number of pairs of individuals (i.e., quadratically or worse in the number of individuals), these methods scale better than linearly in the number of pairs. Excluding phasing time, GERMLINE is approximately linear in the number of individuals (31), whereas phasing time depends on the algorithm used. The computing time for BEAGLE fastIBD is dominated by the phasing algorithm, which scales a little better than quadratically in the number of individuals (16).

Memory constraints are not usually an issue for IBD detection. Long chromosomes can be broken into smaller overlapping pieces for analysis. The use of hash tables or pair-by-pair analysis limits the need for extremely large amounts of memory.

APPLICATIONS

Identity-by-Descent Mapping

As illustrated in **Figure 2**, IBD mapping (the use of recent IBD segments for gene mapping) falls on a continuum between linkage mapping with family data and association analysis. With IBD mapping, it is possible to apply a form of linkage analysis in individuals without known pedigree information or in pedigrees that are too large for standard linkage calculations.

Identity-by-Descent Mapping: In Population Samples

Within a small isolated population, all individuals tend to be related to a moderately high extent, although the exact relationships may not be known. This relatedness can be exploited through IBD inference. As an early example of this approach, three cases of benign recurrent intrahepatic cholestasis were ascertained from an isolated fishing community of several thousand individuals in the Netherlands. Parents and siblings were also collected and were used to determine haplotype phase in each case. Haplotypes were then aligned and a 20-cM haplotype was found on chromosome 18 that was shared by five of the six haplotypes in the three cases (35).

IBD mapping was used to follow up on the results of a genome-wide association study (GWAS) of plasma plant sterol (PPS) levels, a surrogate measure of cholesterol absorption from the intestine, in individuals from the Micronesian island of Kosrae. IBD analysis with GERMLINE revealed a 526-kb haplotype shared by 44 individuals that was highly associated with PPS levels. All but one of the carriers of this haplotype were from the same village. Sequencing of two genes in this subregion revealed a putative causal missense variant (41). This variant was in addition to a probable causal nonsense mutation located nearby, which was not carried by the 44 identical-by-descent individuals. An advantage of the IBD approach over standard GWAS analysis, as shown in this example, is that the patterns of IBD sharing can make it possible to disentangle multiple causal variants within an associated region. Moreover, the IBD approach delineates more precisely the region that should be sequenced and indicates exactly which individuals are carriers, and this is helpful in determining which individuals to sequence.

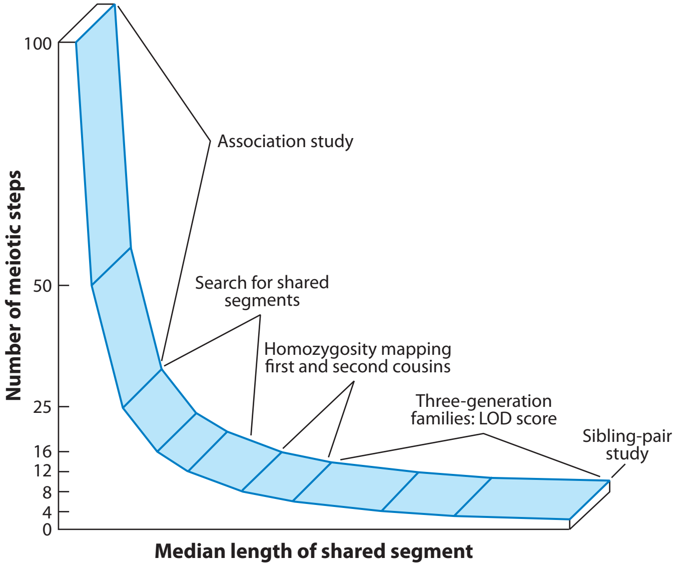

Figure 2

The median lengths of chromosomal segments inherited identical by descent around a disease gene by two individuals separated from a common ancestor by differing numbers of meiotic steps (35). Note the inverse relationship between median length of shared segment and number of meiotic steps. The original illustration was prepared before the days of genome-wide single nucleotide polymorphism (SNP) data. Today, with dense SNP data, the applicability of association studies ranges up to thousands of meiotic steps. Also, with dense SNP data, the range of applicability of identity-by-descent mapping (searching for shared segments) is extended to at least 50 meiotic steps (25 generations to common ancestor). Adapted by permission from Macmillan Publishers Ltd: Nature Genetics.

Browning & Thompson (18) performed a simulation study to assess the power of IBD mapping in population samples, and compared this to the power of standard single-marker association testing. They showed that IBD mapping has higher power than association testing when multiple rare variants within a gene contribute to disease susceptibility. They also developed multiple-testing adjustment guidelines for population-based IBD mapping. The multiple-testing adjustment for IBD mapping is less than for genome-wide single-marker association testing but more than for family-based linkage analysis. The actual adjustment depends on the resolution of IBD detection. In IBD mapping analysis of type I diabetes data from the Wellcome Trust Case Control Consortium study (64), with IBD detected using BEAGLE fastIBD, a p-value threshold of 6×10^{-6} gave a genome-wide significance level of 5%.

The DASH method clusters haplotypes by inferred IBD status and tests for association with disease status (30). Several new associations were found using this approach, both in data from an outbred population (the United Kingdom) and in data from a founder population (Kosrae).

Identity-by-Descent Mapping: Within and Between Families

Full linkage analysis of large families with genome-wide SNP data is only possible with Markov chain Monte Carlo methods (65), which are very computationally intensive and can be unreliable when multiple generations of individuals at the top of the pedigree are not genotyped (7, 62). Pedigree-free methods of IBD detection, such as those described in this review, can be used to detect IBD sharing even in those circumstances where there is a known pedigree, which can then be used in calculating nonparametric (4, 21, 60, 61) or parametric (4, 26) linkage statistics.

Moreover, in some cases it is possible to utilize IBD information from unknown relationships across families (26). Ramensky et al. (53) used IBD to analyze 15 families with Tourette syndrome. Linkage analysis within families had identified genome-wide significant linkage. IBD analysis across families with BEAGLE fastIBD supported the linkage signals and narrowed the peaks significantly.

As a further example of cross-family IBD mapping, we report here a previously unpublished IBD mapping analysis. Gray platelet syndrome (GPS) is a very rare inherited bleeding disorder found in multiple populations. Two Native American families, each with cases of GPS, had no known relationship, although they were part of the same settlement that has limited outbreeding (23, 40). One family contained three cases (two siblings and their first cousin once removed), and the other family contained two cases (siblings). Linkage analysis of other families from around the world with GPS had implicated chromosome 3p21 (29). More recently, sequencing revealed that the genetic causal factor for the disease was compound heterozygosity in the *NBEAL2* (*neurobeachin-like* 2) gene (3, 28, 40). A causal variant was shared between the two Native American families: In one family, the three cases were recessive for this variant, whereas in the other family the two cases carried a second variant in the gene (40).

We used BEAGLE IBD (BEAGLE version 3.2) to detect segments of IBD and HBD between individuals genotyped with Affymetrix 500K SNP data. The analyzed individuals included two cases from each family (one of the siblings and the first cousin once removed from the three-case family and both cases from the two-case family), parents (who were unaffected), and several unaffected siblings. BEAGLECALL (11) was used to call the genotypes while utilizing LD information to obtain highly accurate genotypes, which improve IBD analysis. BEAGLE IBD needs at least a moderately sized sample to obtain accurate estimates of haplotype frequencies, so we augmented the sample with data from the Hapmap3 MEX samples (individuals of Mexican ancestry in Los Angeles, CA) (37a). We found two regions with IBD sharing between cases in both families. One of these was a 7.7-Mb region on chromosome 3, and it encompassed the *NBEAL2* gene. The other was on chromosome 18. The two cases from the three-case family were both HBD across the IBD region on chromosome 3; these cases are homozygous for the causal variant in *NBEAL2*.

The earlier linkage analysis of six worldwide families had implicated a 9.3-Mb region on chromosome 3 (29), which covered *NBEAL2*. It is notable that with just two distantly related families with four cases, a smaller region was obtained than with six worldwide families with 14 cases. Moreover, utilization of both sources of information would have further narrowed the implicated region.

Finding Disease-Causing Variants in Familial Sequence Data

In exome or whole-genome sequence data one looks for putatively functional (e.g., protein altering and/or extremely rare) variants that are shared by cases or for genes for which a significantly large number of cases have such a variant. Early exome sequence analyses of Mendelian traits soon found that this task was not as easy as first thought because the exome (or genome) is large, meaning that cases may all have putatively functional variants in several genes. IBD can be extremely useful in such cases. If one assumes that all (or most) cases in a single family have the same causal variant inherited identical by descent from a common ancestor, then one can narrow down the part of the genome in which to search by looking for regions in which most cases within each family are identical by descent (2, 43, 54, 55, 57). IBD can also be used to determine haplotype phase and to detect sequencing errors that might otherwise lead to false-positive results (54).

Heritability Estimation

Yang et al. (66) showed that estimates of pairwise genome-wide IBD sharing (the proportion of the genome shared identical by descent by a pair of individuals) can be used to estimate narrow-sense (additive) heritability. Such heritability estimates, based on distantly related individuals, reduce problems of confounding caused by shared environments that bias estimates of heritability based on close relatives, such as twins or siblings. However, the estimates can be biased because of population structure (17, 27). Although Yang et al. used identity-by-state-based methods, in their estimation it is also possible to use detected IBD segments. In outbred populations with SNP array data, the amount of detectable IBD is not sufficient to obtain estimates with useful precision (the standard errors of the heritability estimates are very high; S. Browning & B. Browning, unpublished data). However, in small founder populations the amount of IBD sharing is much higher, so this is a useful approach (51, 67).

Price et al. (51) used IBD within and between families from Iceland to investigate heritability of gene-expression traits. Zuk et al. (67) showed that heritability estimated from close relatives can contain nonadditive components, thus overestimating narrow-sense heritability, which may partially explain the phenomenon of missing heritability (45). This overestimation problem is circumvented when heritability is estimated using IBD segments in distantly related individuals.

Distinguishing Between Recurrent Mutation and Shared Ancestry

When one finds several individuals in a population carrying a previously unknown deleterious variant, a natural question to ask is whether the individuals share the variant because of shared ancestry or whether the multiple copies of the variant are due to independent mutation events. Hansen et al. (33) found a novel BRCA1 mutation in 13 apparently unrelated Greenlandic Inuit with familial breast or ovarian cancer. Genotypes in a 4.5-Mb region around the variant were consistent with IBD sharing. In contrast, another variant carried by one Greenlandic Inuit familial case was shared IBD with carriers from a Danish study (34). It was concluded that this variant in the Greenlandic population is likely of Danish origin.

Girirajan et al. (25) identified a 520-kb deletion associated with childhood developmental delay. IBD analysis with BEAGLE IBD of 17 unrelated individuals with the deletion found only one pair of individuals identical by descent. This suggested that the deletion was recurrent, which was consistent with other lines of evidence.

Inference of Haplotype Phase and Imputation

When two individuals are known to share a haplotype identical by descent, their haplotype phase is determined at all SNP genotypes, except where both individuals are heterozygous. This fact is utilized in the phasing of parent-offspring trios, which is significantly more accurate than phasing of unrelated individuals (9, 46). Once haplotype phase is determined,

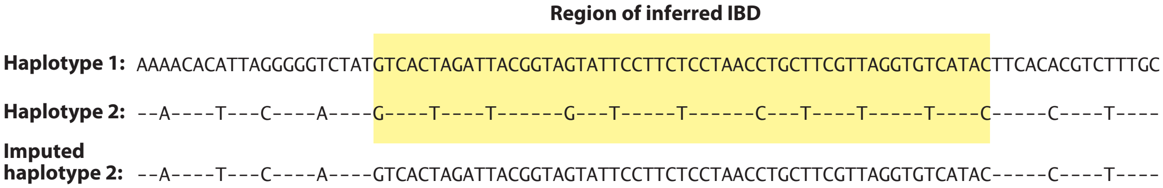
Region of inferred IBD
Haplotype 1: AAAACACATTAGGGGGTCTATGTCACTAGATTACGGTAGTATTCCTTCTCCTAACCTGCTTCGTTAGGTGTCATACTTCACACGTCTTTGC
Haplotype 2: --A----T---C----A----G----T----T------G---T-----T------C---T----T-----T----C-----C----T----
Imputed haplotype 2: --A----T---C----A----GTCACTAGATTACGGTAGTATTCCTTCTCCTAACCTGCTTCGTTAGGTGTCATAC-----C----T----

Figure 3

Imputation using identity-by-descent (IBD) haplotypes. Haplotype 1 is sequenced (only positions varying in the population are shown), whereas haplotype 2 is genotyped at selected sites (positions at which alleles are not assayed are denoted with dashes). Haplotypes 1 and 2 have been inferred to be IBD across the region within the yellow box using the variant positions in common (this number of positions would not normally be sufficient to infer segmental IBD sharing; the figure is conceptual only). Haplotype 2 can then be inferred to copy the same variants as haplotype 1 across this region (shown in imputed haplotype 2).

missing genotypes can also be imputed accurately. For example, if haplotype 1 is sequenced and haplotype 2 is sparsely genotyped and known to be identical by descent with haplotype 1 over a certain region (see **Figure 3**), we can infer that haplotype 2 carries the same variants as haplotype 1 over the region.

Kong et al. (42) demonstrated the feasibility of using IBD for phasing and imputation in the Icelandic population. A large proportion (10%) of the population was genotyped, and the amount of IBD sharing was high because of the small population size (320,000). As a result, a high proportion of genotypes could be accurately phased (over 90% of heterozygotes were phased, and error rates were estimated to be less than 0.1% in the three regions examined). The potential for further developments of this approach for large populations are discussed in Reference 16.

Setty et al. (56) applied IBD-based imputation to human leukocyte antigen (HLA) typing in 182 European CEU samples. This approach shows some promise; however, the low sample size caused lower yield and accuracy of IBD detection, and hence of imputation, than could have been achieved in a larger data set.

Inference of Relationships and Population Structure

Inferred segments of IBD provide information about relationships between individuals. Pairs of individuals sharing no detected IBD segments, or only a small number of short segments, are only distantly related. In contrast, pairs sharing a large number of large segments are closely related. Browning & Browning (10) show that, for close relationships up to fifth cousins, the proportion of genome covered by detected IBD segments is a much more accurate estimator of the true proportion of the genome shared identical by descent than usual identity-by-state methods. Huff et al. (36) go further and use the observed pattern of IBD sharing, rather than just the total quantity of IBD sharing, to infer relationships. They show that the inferred relationships are usually accurate to within a degree of relatedness (one meiotic step) for up to seventh degree relatives (e.g., third cousins).

Although more distant relationships cannot be estimated accurately using IBD sharing alone, IBD sharing across many pairs of individuals within or between populations can provide useful information on population structure. Browning & Browning (10) found increased IBD sharing within pairs of Welsh individuals, compared with pairs of individuals from other regions of the United Kingdom, as would be expected given the relative geographic isolation of Wales. Soi et al. (58) used GERMLINE to infer IBD tracts between individuals in hunting-gathering populations and neighboring agriculturalist and pastoralist populations. Palamara & Pe'er (50) used GERMLINE and additional modeling to estimate the demographic history for the past two millennia for two populations (Ashkenazi Jewish and Maasai).

Inferring Signals of Natural Selection

Recent or ongoing natural selection will lead to an increase in segmental IBD sharing (6). Albrechtsen et al. (6) analyzed HapMap phase 3 CEU data with RELATE and found significantly elevated levels of IBD in the HLA region on chromosome 6 and over the chromosome 8p23 inversion. Both regions have previously shown signatures of selection using alternative methods (22, 37). Whereas the mean posterior probability of IBD sharing in this analysis was 0.002 on average across the genome, in the HLA region it was 0.06, and in the chromosome 8 region it was 0.014. Cai et al. (19) also looked for signals of selection using IBD segments in HapMap data. They analyzed HapMap 2 data, and in each population they examined 45 individuals and looked for segments compatible with sharing by all 45 individuals. The 20 longest such regions (in terms of number of SNPs) were noted. Sixteen of these have already been shown to be regions of selection using other methods. The regions found were different from those of Albrechtsen et al., which is not surprising given that the statistics used are quite different.

One factor to bear in mind in these analyses is that the presence of extended and strong LD in a region enhances the ability to detect IBD and may also lead to some false-positive IBD, depending on how well the method adjusts for LD. Extended LD is in itself a signature of natural selection. Thus, excess IBD may be a proxy for extended LD as a signature of selection rather than providing fully complementary information for finding signatures of selection.

DISCUSSION

In this review, we have surveyed the existing methods for IBD segment detection. Methods for IBD detection are based on the frequency of shared haplotypes, but there are differences in approach between length-based methods, direct frequency-based methods, and probabilistic methods. Probabilistic approaches have the greatest potential to yield high power with low false-positive rates because they can account for all aspects of the underlying data generation processes. However, it is difficult to incorporate LD into these methods without causing slow computation times.

It is important to compare methods in terms of power, false-positive rate, and computation time. Assessment of false-positive rates is particularly challenging but extremely important. Comparison of methods in terms of power without controlling false-positive rate is meaningless, as false-positive rates can differ substantially between methods. Many methods have tuning parameters that allow a trade-off between false-positive rate and power.

We also described applications of IBD segment detection. These include IBD mapping, distinguishing between shared ancestry and recurrent mutation, haplotype phasing and imputation, inferring relationships and population structure, and detecting signatures of selection. The diversity of applications reflects the centrality of IBD to genetics. Many of these applications are only just emerging, and it is not yet clear how much impact they will have. IBD mapping in outbred populations has not yet proved its potential but may do so, particularly as methods for detecting multi-individual IBD become available and as linkage statistics are developed for better utilizing multi-individual IBD information in population samples. IBD-based haplotype phasing and imputation has proved its worth in founder populations but is not yet applicable to outbred populations. Improvements in IBD detection and in utilization of the IBD in determining haplotype phase are being developed.

The field of IBD segment detection has developed rapidly in response to increased density of genetic data along chromosomes. Full sequence data will soon be readily available, and much greater resolution of IBD detection is to be expected in such data. Methods for IBD detection will need to change for sequence data. Many existing IBD detection methods are designed specifically for common variants; however, rare variants are highly informative for IBD and should be utilized. Sequence data is noisy, with higher rates of error than SNP array data, so these errors will need to be accommodated. Finally, IBD detection on sequence data will need to account for very recent mutations that disrupt sequence identity in individuals with recent common ancestry.

SUMMARY POINTS

1. Recent IBD is IBD resulting from shared ancestry within the past 25 generations or so. This IBD is in the form of fairly large (approximately 2 Mb or larger) segments, which are often detectable with a dense SNP array or sequence data.
2. There are several types of methods for detecting IBD segments, including those based on the length of segment consistent with IBD sharing and probabilistic methods that determine a posterior probability of IBD. The underlying principle for all types of detection algorithms is that shared haplotypes of very low frequency are likely to be IBD.
3. When comparing methods for IBD segment detection, it is important to consider power to detect IBD segments, false-positive rates, and computational time for a range of sample sizes.
4. Applications of detected IBD segments are diverse and include the identification of disease-susceptibility genes (IBD mapping), improved accuracy in haplotype phasing and imputation, population genetic inference, and finding signals of recent natural selection.

FUTURE ISSUES

1. The use of IBD segments for significantly improving haplotype phase and genotype imputation accuracy has been demonstrated in founder populations. It is difficult to extend the existing methods to outbred populations; improved IBD segment detection methods coupled with more sophisticated ways to utilize IBD segments in the haplotype phasing are needed.
2. Most existing methods for IBD detection consider only pairs of individuals. Existing methods for considering multiple individuals simultaneously are either too slow for large data sets or are fairly ad hoc with less than optimal characteristics. Consideration of multiple individuals simultaneously will be important for improving the performance of various applications, particularly IBD mapping and haplotype phasing.
3. IBD detection in full sequence data will require new IBD detection methods that make use of rare variants and account for sequencing error rates and very recent mutations.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Jorge DiPaola for providing the Native American gray platelet syndrome data. This work was supported by National Institutes of Health (NIH) awards R01HG005701, R01HG004960, and R01GM057091.

LITERATURE CITED

1. Abecasis GR, Wigginton JE. 2005. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am. J. Hum. Genet.* 77:754–67
2. Akula N, Detera-Wadleigh S, Shugart Y, Nalls M, Steele J, McMahon FJ. 2011. Identity-by-descent filtering as a tool for the identification of disease alleles in exome sequence data from distant relatives. *BMC Proc.* 5(Suppl. 9):S76
3. Albers CA, Cvejic A, Favier R, Bouwmans EE, Alessi MC, et al. 2011. Exome sequencing identifies NBEAL2 as the causative gene for gray platelet syndrome. *Nat. Genet.* 43:735–37
4. Albers CA, Stankovich J, Thomson R, Bahlo M, Kappen HJ. 2008. Multipoint approximations of identity-by-descent probabilities for accurate linkage analysis of distantly related individuals. *Am. J. Hum. Genet.* 82:607–22
5. Albrechtsen A, Korneliussen TS, Moltle I, Hansen TV, Nielsen FC, Nielsen R. 2009. Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genet. Epidemiol.* 33:266–74
6. Albrechtsen A, Moltke I, Nielsen R. 2010. Natural selection and the distribution of identity-by-descent in the human genome. *Genetics* 186:295–308
7. Brown MD, Glazner CG, Zheng C, Thompson EA. 2012. Inferring coancestry in population samples in the presence of linkage disequilibrium. *Genetics* 190:1447–60
8. Browning BL, Brashear DL, Butler AA, Cyr DD, Harris EC, et al. 2004. Linkage analysis using single nucleotide polymorphisms. *Hum. Hered.* 57:220–27
9. Browning BL, Browning SR. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84:210–23
10. Browning BL, Browning SR. 2011. A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* 88:173–82
11. Browning BL, Yu Z. 2009. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* 85:847–61
12. Browning SR. 2006. Multilocus association mapping using variable-length Markov chains. *Am. J. Hum. Genet.* 78:903–13
13. Browning SR. 2008. Estimation of pairwise identity by descent from dense genetic marker data in a population sample of haplotypes. *Genetics* 178:2123–32
14. Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–97
15. Browning SR, Browning BL. 2010. High-resolution detection of identity by descent in unrelated individuals. *Am. J. Hum. Genet.* 86:526–39
16. Browning SR, Browning BL. 2011. Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* 12:703–14
17. Browning SR, Browning BL. 2011. Population structure can inflate SNP-based heritability estimates. *Am. J. Hum. Genet.* 89:191–93; author reply 193–95
18. Browning SR, Thompson EA. 2012. Detecting rare variant associations by identity-by-descent mapping in case-control studies. *Genetics* 190:1521–31
19. Cai Z, Camp NJ, Cannon-Albright L, Thomas A. 2011. Identification of regions of positive selection using Shared Genomic Segment analysis. *Eur. J. Hum. Genet.* 19:667–71
20. Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, et al. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* 1:131
21. Day-Williams AG, Blangero J, Dyer TD, Lange K, Sobel EM. 2011. Linkage analysis without defined pedigrees. *Genet. Epidemiol.* 35:360–70
22. Deng L, Zhang Y, Kang J, Liu T, Zhao H, et al. 2008. An unusual haplotype structure on human chromosome 8p23 derived from the inversion polymorphism. *Hum. Mutat.* 29:1209–16
23. Fabbro S, Kahr WH, Hinckley J, Wang K, Moseley J, et al. 2011. Homozygosity mapping with SNP arrays confirms 3p21 as a recessive locus for gray platelet syndrome and narrows the interval significantly. *Blood* 117:3430–34

24. Genovese G, Leibon G, Pollak MR, Rockmore DN. 2010. Improved IBD detection using incomplete haplotype information. *BMC Genet.* 11:58

25. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, et al. 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* 42:203–9

26. Glazner C, Thompson EA. 2012. Improving pedigree-based linkage analysis by estimating coancestry among families. *Stat. Appl. Genet. Mol. Biol.* 11(Issue 2):Article 11

27. Goddard ME, Lee SH, Yang J, Wray NR, Visscher PM. 2011. Response to Browning and Browning. *Am. J. Hum. Genet.* 89:193–95

28. Gunay-Aygun M, Falik-Zaccai TC, Vilboux T, Zivony-Elboum Y, Gumruk F, et al. 2011. NBEAL2 is mutated in gray platelet syndrome and is required for biogenesis of platelet alpha-granules. *Nat. Genet.* 43:732–34

29. Gunay-Aygun M, Zivony-Elboum Y, Gumruk F, Geiger D, Cetin M, et al. 2010. Gray platelet syndrome: natural history of a large patient cohort and locus assignment to chromosome 3p. *Blood* 116:4990–5001

30. Gusev A, Kenny EE, Lowe JK, Salit J, Saxena R, et al. 2011. DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. *Am. J. Hum. Genet.* 88:706–17

31. Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, et al. 2009. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 19:318–26

32. Han L, Abney M. 2011. Identity by descent estimation with dense genome-wide genotype data. *Genet. Epidemiol.* 35:557–67

33. Hansen TV, Ejlertsen B, Albrechtsen A, Bergsten E, Bjerregaard P, et al. 2009. A common Greenlandic Inuit BRCA1 RING domain founder mutation. *Breast Cancer Res. Treat.* 115:69–76

34. Hansen TV, Jonson L, Albrechtsen A, Steffensen AY, Bergsten E, et al. 2010. Identification of a novel BRCA1 nucleotide 4803delCC/c.4684delCC mutation and a nucleotide 249T>A/c.130T>A (p.Cys44Ser) mutation in two Greenlandic Inuit families: implications for genetic screening of Greenlandic Inuit families with high risk for breast and/or ovarian cancer. *Breast Cancer Res. Treat.* 124:259–64

35. Houwen RH, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, et al. 1994. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat. Genet.* 8:380–86

36. Huff CD, Witherspoon DJ, Simonson TS, Xing J, Watkins WS, et al. 2011. Maximum-likelihood estimation of recent shared ancestry (ERSA). *Genome Res.* 21:768–74

37. Hughes AL, Yeager M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. *Annu. Rev. Genet.* 32:415–35

38. Int. HapMap 3 Consort. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58

39. Jacquard A. 1974. *The Genetic Structure of Populations*. New York: Springer

40. Kahr WH, Hinckley J, Li L, Schwertz H, Christensen H, et al. 2011. Mutations in NBEAL2, encoding a BEACH protein, cause gray platelet syndrome. *Nat. Genet.* 43:738–40

41. Kenny EE, Gusev A, Riegel K, Lutjohann D, Lowe JK, et al. 2009. Systematic haplotype analysis resolves a complex plasma plant sterol locus on the Micronesian Island of Kosrae. *Proc. Natl. Acad. Sci. USA* 106:13886–91

42. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, et al. 2008. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* 40:1068–75

43. Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U, et al. 2010. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat. Genet.* 42:827–29

44. Leibon G, Rockmore DN, Pollak MR. 2008. A SNP streak model for the identification of genetic regions identical-by-descent. *Stat. Appl. Genet. Mol. Biol.* 7(Issue 1):Article 16

45. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461:747–53

46. Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, et al. 2006. A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.* 78:437–50

47. Miyazawa H, Kato M, Awata T, Kohda M, Iwasa H, et al. 2007. Homozygosity haplotype allows a genomewide search for the autosomal segments shared among patients. *Am. J. Hum. Genet.* 80:1090–102

48. Moltke I, Albrechtsen A, Hansen TV, Nielsen FC, Nielsen R. 2011. A method for detecting IBD regions simultaneously in multiple individuals–with applications to disease genetics. *Genome Res.* 21:1168–80

49. Nelson S, Merriman B, Chen Z, Ogdie M, Stone J, Strom S. 2006. *Applications of pedigree-free identity-by-descent mapping to localizing disease genes. Abstr. 1530*. Presented at Annu. Meet. The Am. Soc. Hum. Genet., Oct. 11, New Orleans, LA. **http://www.ashg.org/genetics/ashg06s/f20776.htm**

50. Palamara P, Pe'er I. 2011. *Length distributions of identity by descent reveal fine-scale demographic history*. Presented at 12th Int. Congr. Hum. Genet./61st Annu. Meet. The Am. Soc. Hum. Genet., Oct. 13, Montreal, Can.

51. Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K. 2011. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* 7:e1001317

52. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–75

53. Ramensky V, Yu D, Service S, Matthews C, Heutink P, et al. 2011. *From linkage to sequencing: Using cross-family IBD sharing to refine susceptibility loci in Tourette Syndrome multi-generational families*. Presented at 12th Int. Congr. Hum. Genet. /61st Annu. Meet. The Am. Soc. Hum. Genet., Oct. 13, Montreal, Can.

54. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–39

55. Rodelsperger C, Krawitz P, Bauer S, Hecht J, Bigham AW, et al. 2011. Identity-by-descent filtering of exome sequence data for disease-gene identification in autosomal recessive disorders. *Bioinformatics* 27:829–36

56. Setty MN, Gusev A, Pe'er I. 2011. HLA type inference via haplotypes identical by descent. *J. Comput. Biol.* 18:483–93

57. Smith KR, Bromhead CJ, Hildebrand MS, Shearer AE, Lockhart PJ, et al. 2011. Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. *Genome Biol.* 12:R85

58. Soi S, Scheinfeldt L, Lambert C, Hirbo J, Ranciaro A, et al. 2011. *Demographic histories of African hunting-gathering populations inferred from genome-wide SNP variation*. Presented at 12th Int. Congr. Hum. Genet. /61st Annu. Meet. The Am. Soc. Hum. Genet., Oct. 13, Montreal, Can.

59. Thomas A. 2010. Assessment of SNP streak statistics using gene drop simulation with linkage disequilibrium. *Genet. Epidemiol.* 34:119–24

60. Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA. 2008. Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. *Ann. Hum. Genet.* 72:279–87

61. Thomas A, Skolnick MH, Lewis CM. 1994. Genomic mismatch scanning in pedigrees. *IMA J. Math. Appl. Med. Biol.* 11:1–16

62. Thompson EA. 2000. *Statistical Inferences from Genetic Data on Pedigrees*. Beachwood, OH: Inst. Math. Stat.

63. Thompson EA. 2008. The IBD process along four chromosomes. *Theor. Popul. Biol.* 73:369–73

64. Wellcome Trust Case Control Consort. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78

65. Wijsman EM, Rothstein JH, Thompson EA. 2006. Multipoint linkage analysis with many multiallelic or dense diallelic markers: Markov chain-Monte Carlo provides practical approaches for genome scans on general pedigrees. *Am. J. Hum. Genet.* 79:846–58

66. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–69

67. Zuk O, Hechter E, Sunyaev SR, Lander ES. 2012. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* 109:1193–98

Annual Review of Genetics

Volume 46, 2012

Contents

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at http://genet.annualreviews.org/errata.shtml