1 SUPPLEMENTARY MATERIAL

1.1 S.1 Emission probabilities with missing and genotyping errors

We should also take into account the effect of missing genotypes and genotyping errors. We use a random variable T to indicate whether there are typing errors at a certain locus, $T = 0$ indicates that there are no typing errors and $T = 1$ otherwise. Similarly, we use M to indicate whether there are missing genotypes. We introduce a new IBS state $G = 3$ to indicate the occurrence of missing genotypes, we have the following probability:

$$P(G|I,T,M=1) = \begin{cases} 1 & \text{if } G=3 \\ 0 & \text{if } G=0,1,2 \end{cases}$$

Assume that in the case of typing errors, two alleles will be reported at an equal probability, we have:

$$P(G|I,T=1,M=0) = \begin{cases} (1/2)^2 & \text{if } G=0 \\ 2\cdot(1/2)^2 & \text{if } G=1 \\ (1/2)^2 & \text{if } G=2 \\ 0 & \text{if } G=3 \end{cases}$$

If there are no missing genotypes and no typing errors then,

$$P(G|I,T=0,M=0) = \begin{cases} \text{Table 1 (main text)} & \text{if } G=0,1,2 \\ 0 & \text{if } G=3 \end{cases}$$

We derive the marginal probability by summing over the prior probabilities of missing genotypes and typing errors,

$$P(G|I) = \sum_{i=0}^{1}\sum_{j=0}^{1}(P(G|I,T=i,M=j)P(T=i)P(M=j)).$$

The prior probabilities are estimated from the data.

1.2 S.2 Variables used in the text

Table 1. Notations.

$p^{a,b}$	A path on a descent graph which links allele a and allele b
$\theta_{a,b}(h)$	Inheritance generating function for 2 alleles a and b
$\Phi(p)$	Probability of the occurrence of an inheritance path p
$\phi(p,i,i+1)$	Probability that p remains unchanged from locus i to $i+1$
$\psi(i,i+1)$	$P(a_{i+1} \overset{\text{ibd}}{=} b_{i+1} \mid a_i \overset{\text{ibd}}{=} b_i)$

1.3 S.3 Typical errors in inferred IBD regions

Fig. 1 shows some typical errors in inferred IBD regions. False positives are usually caused by long segments of background IBD sharing as shown in Fig. 1(a). Typing errors can also introduce false breakpoints (Fig. 1(c)) in the inferred IBD regions. On the other hand, missed IBD regions (Fig. 1(b)) are usually short and they are falsely inferred to be background sharing. Another common error happens in the centromere region around 15Mb (Fig. 1(a)), where the marker density is low. Breakpoints are easily misdetected in this area.

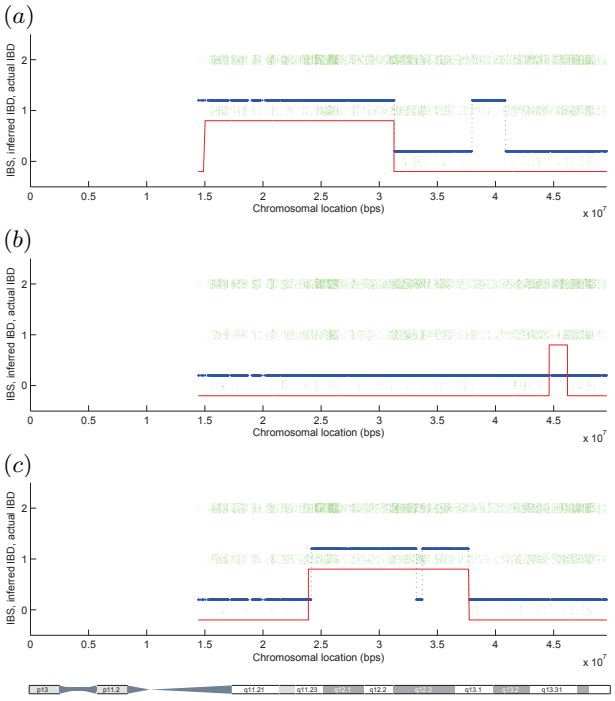

Fig. 1. Typical errors in inferred IBD regions. The dotted bar indicates the density of markers of IBS number 0, 1 and 2. The bold line is the inferred IBD sharing number, the slim line is the actual IBD sharing number.