

1 Running Title: Find IBD Shared Haplotypes Rapidly

2

3

4

5

6

7

8

9

10 **A fast and accurate method for detection of IBD shared haplotypes in genome-wide SNP**

11 **data**

12

13

14 Douglas W. Bjelland<sup>1</sup>, Uday Lingala<sup>1</sup>, Piyush S. Patel<sup>1</sup>, Matt Jones<sup>2</sup>, and Matthew C. Keller<sup>1,2,\*</sup>

15

16 <sup>1</sup>Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, CO, 80303

17 <sup>2</sup>Department of Psychology & Neuroscience, University of Colorado at Boulder, Boulder, CO,

18 80301

19 \*Corresponding author: [matthew.c.keller@gmail.com](mailto:matthew.c.keller@gmail.com)

20

21 Key words: Identical by descent, shared haplotype

22

## 23 **Abstract**

24 Identical by descent (IBD) segments are used to understand a number of fundamental issues in  
25 genetics. IBD segments are typically detected using long stretches of identical alleles between  
26 haplotypes in whole-genome SNP data. Phase or SNP call errors in genomic data can degrade  
27 accuracy of IBD detection and lead to false positive calls, false negative calls, and under- or  
28 overextension of true IBD segments. Furthermore, the number of comparisons increases  
29 quadratically with sample size, requiring high computational efficiency. We developed a new  
30 IBD segment detection program, FISHR (Find IBD Shared Haplotypes Rapidly), in an attempt to  
31 accurately detect IBD segments and to better estimate their endpoints using an algorithm that is  
32 fast enough to be deployed on the very large whole-genome SNP datasets. We compared the  
33 performance of FISHR to three leading IBD segment detection programs: GERMLINE,  
34 refinedIBD, and HaploScore. Using simulated and real genomic sequence data, we show that  
35 FISHR is slightly more accurate than all programs at detecting long (>3 cM) IBD segments but  
36 slightly less accurate than refinedIBD at detecting short (~1 cM) IBD segments. Moreover,  
37 FISHR outperforms all programs in determining the true endpoints of IBD segments, which is  
38 important for several reasons. FISHR takes two to four times longer than GERMLINE to run,  
39 whereas both GERMLINE and FISHR were orders of magnitude faster than refinedIBD and  
40 HaploScore. Overall, FISHR provides accurate IBD detection in unrelated individuals and is  
41 computationally efficient enough to be utilized on large SNP datasets > 20,000 individuals.

42

## 43 **Introduction**

44 Two haplotypes (homologous chromosomal segments of DNA) can be defined as being identical  
45 by descent (IBD) if they descend from a common ancestor without either haplotype experiencing  
46 an intervening recombination (Powell et al. 2010). Using this definition, IBD haplotypes are  
47 identical at all measured and unmeasured genetic polymorphisms except at sites harboring  
48 (typically very rare) mutations that arose on either haplotype since the last common ancestor.  
49 The probability of two individuals co-inheriting an IBD haplotype from a common ancestor at a  
50 given location is a function of the number of generations ( $g$ ) since the common ancestor:  
51  $P(\text{IBD} | g) = 2^{1-2g}$ . Thus, siblings ( $g=1$ ) have a 0.5 probability of sharing a segment IBD from  
52 one of their common ancestors (one parent) at a given genomic location, cousins ( $g=2$ ) have a  
53 0.125 probability, second cousins ( $g=3$ ) a .03125 probability, and so forth. Although this  
54 probability drops off rapidly as a function of generations since the common ancestor, when  
55 haplotypes are shared IBD, they can be quite long, even for distantly related pairs of individuals.  
56 Under Haldane's (1919) model of recombination, the length of IBD haplotypes shared between  
57 two individuals is exponentially distributed with mean  $100/2g$  centiMorgans (cM). Thus,  
58 although a pair of individuals sharing a common ancestor 15 generations ago is highly unlikely  
59 to share any IBD haplotypes from that ancestor, when they do, the expected length of the  
60 segment is  $\sim 3.3$  cM. Given that the probability of two random individuals sharing at least one  
61 common ancestor within 15 generations is  $\sim 1$  in even large, randomly mating populations (Keller  
62 et al. 2011), a large number of IBD shared haplotypes around this length exist in any group of  
63 'unrelated' individuals of the same population.

64

65 IBD shared haplotypes in samples with no known pedigree relatedness have been used for  
66 genotype imputation (Kong et al. 2008; Setty et al. 2011), IBD mapping (Vacic et al. 2014),  
67 heritability estimation (Browning and Browning 2013), phase inference (Kong et al. 2008), and  
68 inference of population structure (Palamara et al. 2012; Soi et al. 2011). Such IBD shared  
69 haplotypes are typically inferred from long stretches of identical alleles in phased, whole-  
70 genome single nucleotide polymorphism (SNP) arrays, but accurate and efficient IBD detection  
71 from such data is difficult for several reasons. First, phase and SNP-call errors can split long IBD  
72 segments into two or more shorter segments or lead to artificial truncation of IBD segments.  
73 Such splitting and truncating of IBD segments can lead to failure to detect a segment altogether,  
74 due to the segment being shorter than a prespecified length threshold or due to the fact that  
75 shorter segments have lower posterior probabilities of being IBD, depending on the IBD  
76 detection algorithm. Thus, errors in SNP calling and phasing inflate false negative (miss) rates of  
77 IBD detection. Second, the sheer number of comparisons that must be made at each site (four  
78 comparisons between each pair of diploid individuals leads to a number of comparisons ~twice  
79 the squared sample size), combined with the low base rate of true IBD segments between pairs of  
80 unrelated individuals, means that a substantial fraction of called IBD segments can be false  
81 positives. Similar to the case of false negatives, a false positive can be due to either an entire  
82 called segment not being IBD or to a called segment being overextended in one or both  
83 directions. Finally, because of the computational complexity of IBD detection, algorithms that  
84 sacrifice speed for accuracy can be unusable on the large sample sizes (e.g., >50,000) currently  
85 being accumulated (e.g., Schizophrenia Working Group of Psychiatric Genomics Consortium  
86 2014; Sudlow et al. 2015). In a sample of 50,000 individuals, nearly 5 billion comparisons must  
87 be made per site. Thus, successful IBD detection programs must simultaneously meet a number

88 of goals—computational efficiency, low false positive rates, low false negative rates, and  
89 accurate detection of IBD segment endpoints—that typically trade off with one another.  
90  
91 Several programs have been developed to discover IBD segments in SNP datasets when  
92 expected pedigree relatedness is low. GERMLINE (Gusev et al. 2009), often considered the  
93 benchmark IBD discovery program, is computationally efficient and therefore usable on very  
94 large samples, but the literature has indicated that its accuracy is lower than more recently  
95 developed programs. Because GERMLINE is fast and can be run in a way that leads to few  
96 false-negative calls at the expense of many false-positive calls, two newer IBD detection  
97 programs that reportedly outperform GERMLINE in accuracy, refined IBD (rIBD; Browning  
98 and Browning 2011) and HaploScore (Durand et al. 2014), use GERMLINE to detect candidate  
99 IBD segments. These candidate IBD segments are found using GERMLINE parameters that are  
100 optimized for each program. They are then post-processed, by extending, removing, or slicing  
101 the candidate segments in the hope of providing more accurate detection of IBD segments. rIBD  
102 uses a probabilistic hidden Markov model to give each candidate IBD segment obtained from  
103 GERMLINE a posterior LOD score as to whether it is truly IBD or not. rIBD has a lower false-  
104 positive rate than GERMLINE with only a modest increase in the false-negative rate, but it is  
105 computationally intensive and therefore has a very long runtime for large datasets. HaploScore  
106 uses information on the switch error rate and the SNP error rate to give a posterior probability of  
107 whether each candidate segment from GERMLINE is truly IBD or not.  
108  
109 The current paper describes a new program, FISHR (Find IBD Shared Haplotypes Rapidly), we  
110 developed to have a computational efficiency comparable to GERMLINE with accuracy as good

111 as or better than rIBD or HaploScore. Importantly, because we had observed that existing  
112 programs tend either to over-extend true IBD segments or to split true IBD segments into  
113 multiple smaller ones, one of our central goals was to develop an algorithm that accurately  
114 determines the endpoints and hence the true lengths of IBD segments. This is important because  
115 bias in estimating the true length of IBD segments can lead to under- or over-estimates of  
116 heritability using IBD haplotypes, and inaccurate endpoint estimates can lead to decreased  
117 accuracy of imputation, phasing, and mapping near endpoints. As with rIBD and HaploScore,  
118 FISHR obtains candidate IBD segments by using GERMLINE. Segments can then be stitched  
119 together if separated by a small number of SNPs. After this, the number of “implied errors”  
120 (IE)—likely SNP call or phase errors—throughout the segment are counted, and the segment can  
121 then be shortened or removed entirely based on the number and location of the of IEs (see  
122 *Methods*). To analyze the programs, we compare the runtimes and offer extrapolated estimates  
123 for running them on large, whole-genome datasets. We then compare the positive predictive  
124 value (PPV, the proportion of called segments that are truly IBD) and sensitivity (the proportion  
125 of true IBD segments that are called) across a range of tuning parameters to explore the PPV-  
126 sensitivity trade-off for each program. We also compare the bias, precision, and accuracy of  
127 endpoint detection of truly IBD segments across programs and explain how these are related to  
128 PPV and sensitivity depending on how these metrics are defined. Much of the apparent  
129 discrepancy in comparisons of IBD detection programs that exist in the literature can be  
130 explained by how researchers have decided how over- and underextensions of called segments  
131 affect PPV and sensitivity.

132

## 133 **Results**

134 *Comparison of run times*

135 Figure 1 presents the  $\log_2$  runtimes of the four programs as a function of sample size for five  
136 sample sizes. We calculated runtimes based on the optimal parameters found for each of the  
137 programs as described below. Runtimes were averaged from three separate simulated  
138 subchromosomes that were on average 16 cM long and contained 1,185 SNPs each (see  
139 *Methods*). Because GERMLINE is used as a first step for FISHR, HaploScore and rIBD, the  
140 runtimes for those programs include the time it took GERMLINE to find the candidate segments  
141 as well. Whereas GERMLINE is run internally for rIBD, FISHR and HaploScore require  
142 GERMLINE to be run separately and with user-specified parameters. Thus, in the present  
143 manuscript, we used three different sets of GERMLINE parameters: those that optimized  
144 accuracy for GERMLINE when reporting GERMLINE results, those that did so for FISHR for  
145 FISHR results, and those that did so for HaploScore for HaploScore results. For this reason, the  
146 runtimes presented in Supplemental Table 1 show different runtimes for GERMLINE when run  
147 by itself than when used as a precursor program.

148  
149 GERMLINE was the fastest program to run at any of the sample sizes, with FISHR doubling to  
150 quadrupling its runtime at all sample sizes. Most of the increase in runtime for FISHR compared  
151 to GERMLINE was caused by using a smaller minimum cM threshold for the initial  
152 GERMLINE segment discovery, which is necessary in order for FISHR to stitch together any  
153 segments that GERMLINE splits apart. Both HaploScore and rIBD had runtimes hundreds to  
154 thousands of times longer than FISHR, with this ratio increasing with larger sample sizes for  
155 rIBD. To gauge how the programs performed on a realistic, large SNP dataset, we also calculated  
156 runtime on a sample of 17,093 individuals aggregated from four datasets (the Atherosclerosis

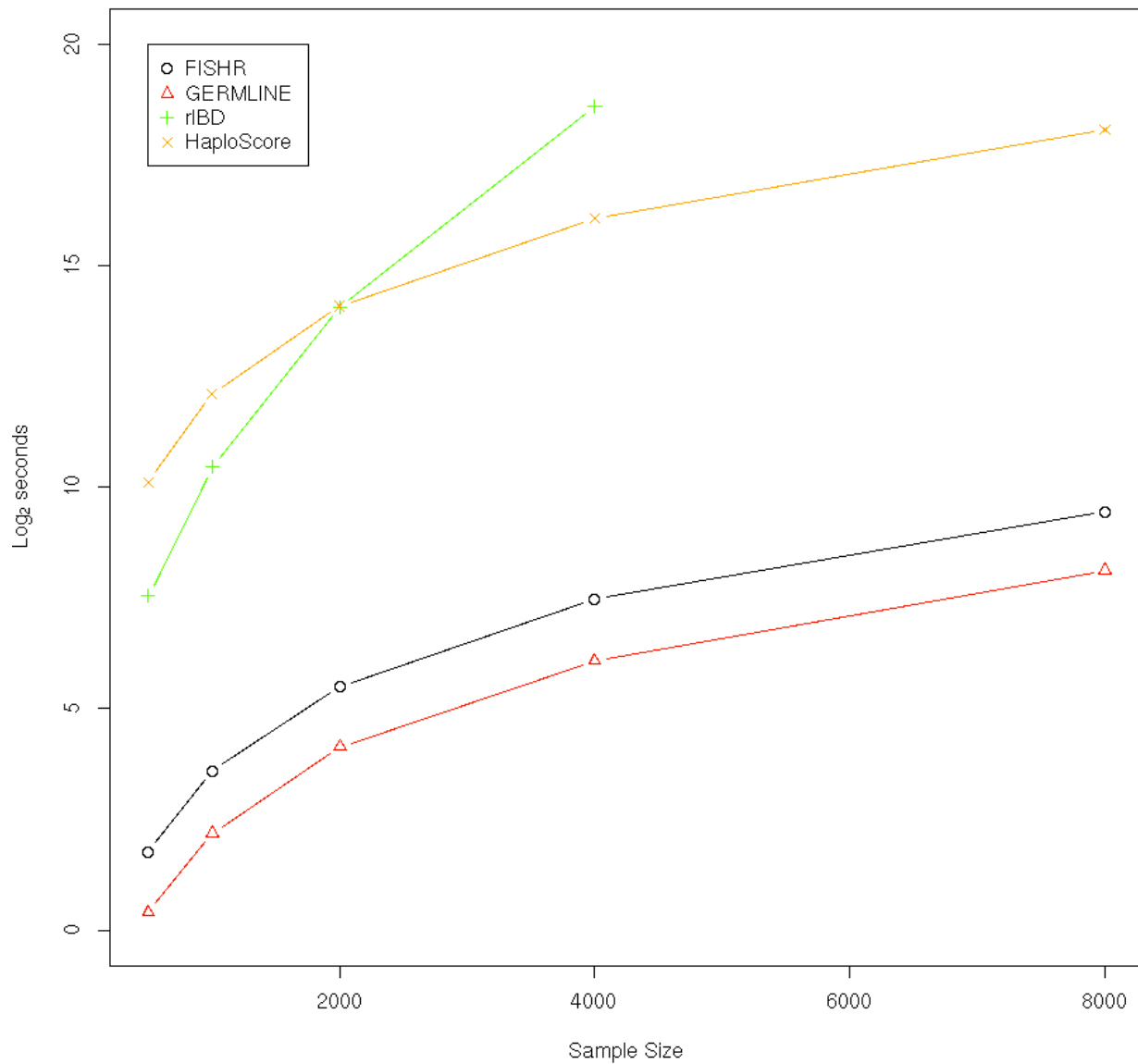
157 Risk in Communities cohort, the Coronary Artery Risk Development in Young Adults study, the  
158 controls from the Molecular Genetics of Schizophrenia study, and the GENEVA Genes and  
159 Environment Initiative in Type 2 Diabetes study; dbGap accessions phs000280.v2.p1,  
160 phs000285.v3.p2, phs000167.v1.p1, and phs000091.v2.p1, respectively) from the NIH Genotype  
161 and Phenotype database. Because IBD detection is typically done in parallel for each  
162 subchromosome arm, we analyzed the longest chromosome arm, 5q, which contained 19,772  
163 SNPs on the Affy 6.0 SNP array. When the threshold for segment length was set to 1 cM,  
164 GERMLINE took about 1.5 days to run, FISHR took about 6.5 days (including 5 days, 16 hours  
165 for GERMLINE initial candidate segment discovery), whereas both rIBD and HaploScore ran  
166 for nearly two months before the server required maintenance and the processes were stopped.  
167 From extrapolations of the runtimes on simulated data (Figure 1), we predict that HaploScore  
168 would have finished running in just over two months and rIBD would have required over a year  
169 to finish.

170

171 **Figure 1.** Runtime in  $\log_2$  seconds for FISHR, GERMLINE, HaploScore, and rIBD at sample  
172 sizes of 500, 1,000, 2,000, 4,000, and 8,000 averaged from three 16-cM simulated chromosomal  
173 segments consisting of 1,185 SNPs each. rIBD with a sample size of 8,000 ran for one month  
174 ( $\sim 2^{21}$  sec) before the server required maintenance and was shut down.

175





176

177

178 *PPV and sensitivity in simulated data*

179 PPV and sensitivity are the most common metrics in this literature for comparing the accuracies

180 of the programs, and so we focus on these for commensurability. An inherent tradeoff exists

181 between the two metrics: conservative calling algorithms that call fewer IBD segments tend to

182 have relatively high PPVs and low sensitivities, whereas more liberal calling algorithms that call

183 more IBD segments tend to have relatively high sensitivities and low PPVs. Figure 2 illustrates

184 how we defined PPV and sensitivity depending on the degree to which called segments over- or  
185 underextend the endpoints of true IBD segments. For PPV, we first calculated the total length of  
186 overlap between each called segment and any corresponding true IBD segment(s) and divided  
187 this overlap by the length of each called segment. Thus, this proportion was 1 for the called  
188 segment in Figure 2A and for both of the called segments in Figure 2D,  $< 1$  for the called  
189 segments in Figures 2B, 2C, and 2E, and 0 for called segments that did not overlap any true IBD  
190 segments. When a single called segment overlapped multiple true IBD segments (Figure 2E),  
191 overlap was defined as the sum of the overlapping lengths. PPV was then calculated as the  
192 average of these proportions across all called segments weighted by their length in basepairs.  
193 Similarly, for sensitivity, we calculated the length of total overlap between each true IBD  
194 segment and any corresponding called segment(s) and divided this overlap by the length of the  
195 true IBD segment. When multiple called segments split up a single true IBD segment (Figure  
196 2D), overlap was again calculated as the sum of the overlapping lengths. Thus, these proportions  
197 were  $< 1$  for Figures 2A, 2C, and 2D but 1 for Figure 2B and for both true segments in Figure 2E.  
198 We defined sensitivity as the average of these proportions weighted by base pair length across all  
199 true IBD segments. Alternative definitions of these metrics are possible. For example,  
200 proportions greater than a threshold (.5) have been treated as true positives and those less than .5  
201 as false positives for calculating PPV (Browning and Browning 2011). We prefer our definitions  
202 because they result in PPV and sensitivity being continuous functions, rather than step functions,  
203 of the degree of over- or underextension, respectively.

204

205 To estimate the accuracies of the programs, we used perfectly matching phased haplotypes from  
206 simulated, dense sequence data with no phase or call errors to define the endpoints of true IBD

207 segments (see *Methods*). We then called segments by applying each of the programs to a subset  
208 of the sequenced variants designed to mimic phased SNP array data, with realistic linkage  
209 disequilibrium (LD) patterns, allele frequencies, SNP densities, and levels of SNP-call and phase  
210 errors. Figure 3 displays PPV and sensitivity where both called and true IBD segments had  
211 minimum lengths of 3 cM (Figure 3A) or 1 cM (Figure 3B). For each program, we varied  
212 thresholds to produce a spectrum of conservative to liberal segment calling. In particular, we  
213 varied the *moving average* threshold for FISHR, the minimum *LOD score* for rIBD, and the *bits*  
214 argument for GERMLINE and HaploScore. At 3 cM minimum segment lengths, FISHR  
215 outperformed every other program with a higher PPV for any given sensitivity or, alternatively, a  
216 higher sensitivity for any given PPV. At 1cM minimum threshold lengths, FISHR and rIBD  
217 performed similarly and outperformed both GERMLINE and HaploScore.

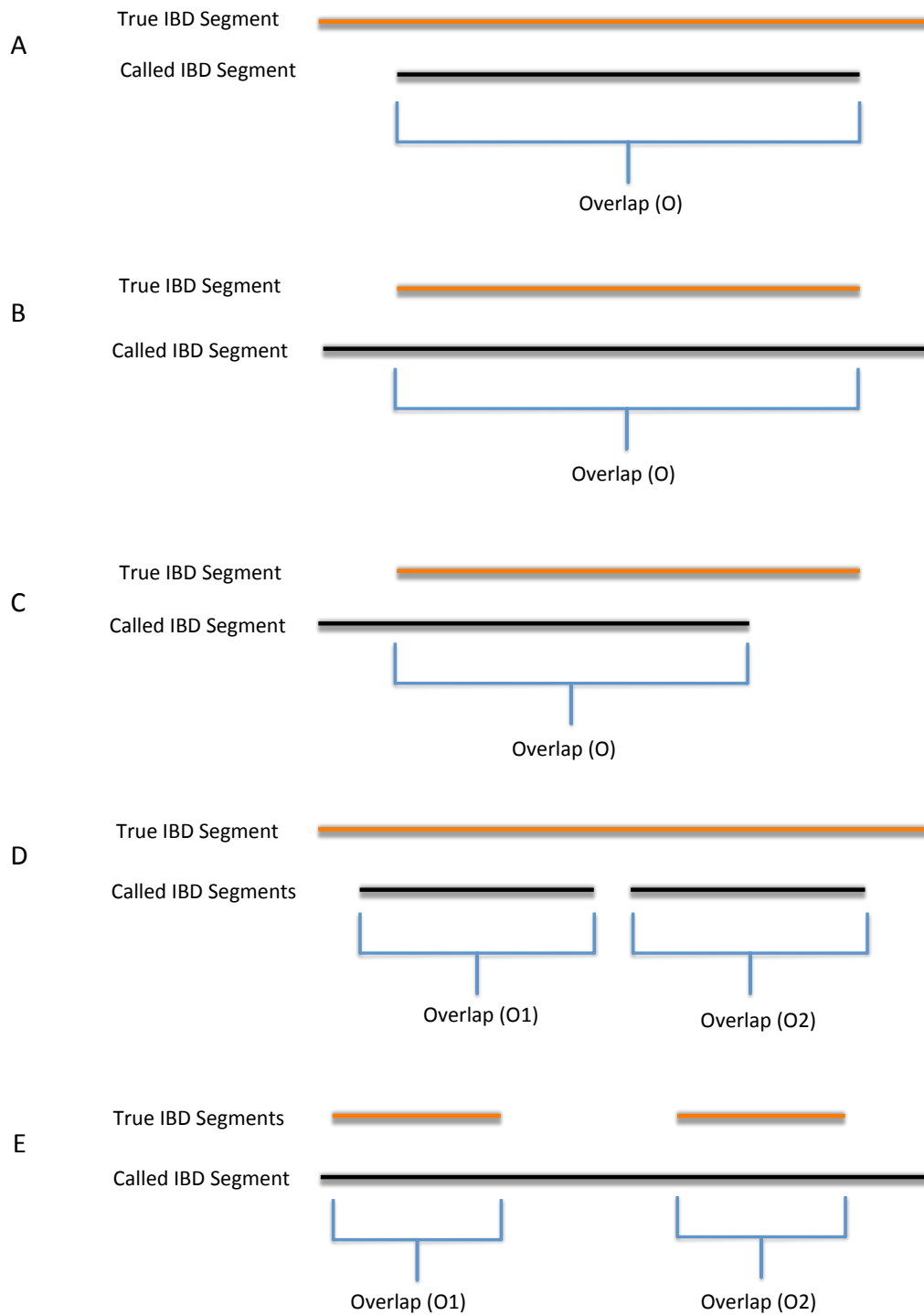
218  
219 By using the same minimum-length thresholds (e.g., 3 cM) for both the called and true IBD  
220 segments, the results displayed in Figure 3 are highly sensitive to the accuracy of the endpoints  
221 of the called segments, as well as to truncation and splitting errors. For example, all sensitivity  
222 estimates of rIBD in Figure 3A are less than 0.3, below those of other programs and below those  
223 reported in the manuscript introducing rIBD (Browning and Browning 2011). As we demonstrate  
224 below, this is because rIBD tends to split true IBD segments into multiple, smaller called  
225 segments; when these called segments are shorter than the threshold (e.g., 3 cM), they are  
226 dropped for the purposes of calculating sensitivity, and therefore most true IBD segments  $> 3$  cM  
227 appear to be missed. Because the endpoints of segments called by GERMLINE and especially  
228 FISHR are more accurate (see below), the performances of these programs are not degraded to  
229 the same extent. An alternative definition of PPV that is less affected by such truncation/split

230 errors is to compare all called segments greater than a length threshold (3 or 1 cM) to all true  
231 IBD segments that are at least half that length (1.5 or 0.5 cM, respectively). Similarly, sensitivity  
232 can be computed by comparing all true IBD segments greater than 3 or 1 cM to all called  
233 segments greater than 1.5 or 0.5 cM, respectively. Figure 4 shows PPV and sensitivity calculated  
234 in this way. The performance of all programs improved but the improvement was greater for  
235 programs that are inaccurate at endpoint estimation (rIBD and HaploScore) than for programs  
236 that are more accurate at endpoint estimation (GERMLINE and especially FISHR; see results on  
237 endpoint accuracy below). At 3 cM minimum called (PPV) and true IBD (sensitivity) segment  
238 lengths, FISHR performed slightly better than GERMLINE or rIBD, whereas at 1 cM minimum  
239 thresholds, rIBD outperformed FISHR. Because rIBD uses a posterior probability instead of a  
240 minimum cM length threshold to call segments, Figure 4 also shows rIBD results when no  
241 minimum length was used in calculating sensitivity and when much smaller true IBD lengths  
242 (0.5 cM for Figure 4A and 0.25 cM for Figure 4B) were used for calculating PPV. The  
243 sensitivity values for these instances of rIBD were improved and show rIBD to be superior to all  
244 other programs with respect to IBD detection accuracy. However, as demonstrated above, these  
245 conclusions rest upon arbitrary decisions on how PPV and sensitivity are defined. Moreover, as  
246 demonstrated below, the improved sensitivity of rIBD when there was no minimum length of  
247 called segments occurred because rIBD often splits long, true IBD segments into multiple, short  
248 called segments, which were sometimes dropped when a length threshold was used in calculating  
249 sensitivity.

250

251 **Figure 2.** Method for calculating PPV and sensitivity from the called IBD segments and the  
252 known true IBD segment from an (A) underextended call, (B) overextend call, (C) off-center

253 call, (D) situation where two called segments occur within a single true IBD, and (E) situation  
254 where one called segments occurs within two true IBD segments. For each called segment, we  
255 divided the length of the overlap with the true segment (O) or sum of the overlaps ( $O_1+O_2$ ) by  
256 the length of the called segment. PPV was the average of these proportions across all called  
257 segments, weighted by base pair length. To determine sensitivity, for each true segment, we  
258 divided the length of overlap (O) or sum of the overlaps ( $O_1+O_2$ ) by the length of true IBD  
259 segment. Sensitivity was the average of these proportions across all true IBD segments, weighted  
260 by base pair length. When two called segments overlapped one true IBD segment (D), two  
261 proportions contributed to PPV (one for each of the called segments) but one proportion to  
262 sensitivity. Conversely, when one called segment overlapped two true IBD segments (E), one  
263 proportion contributed to PPV and two to sensitivity.  
264



265

266

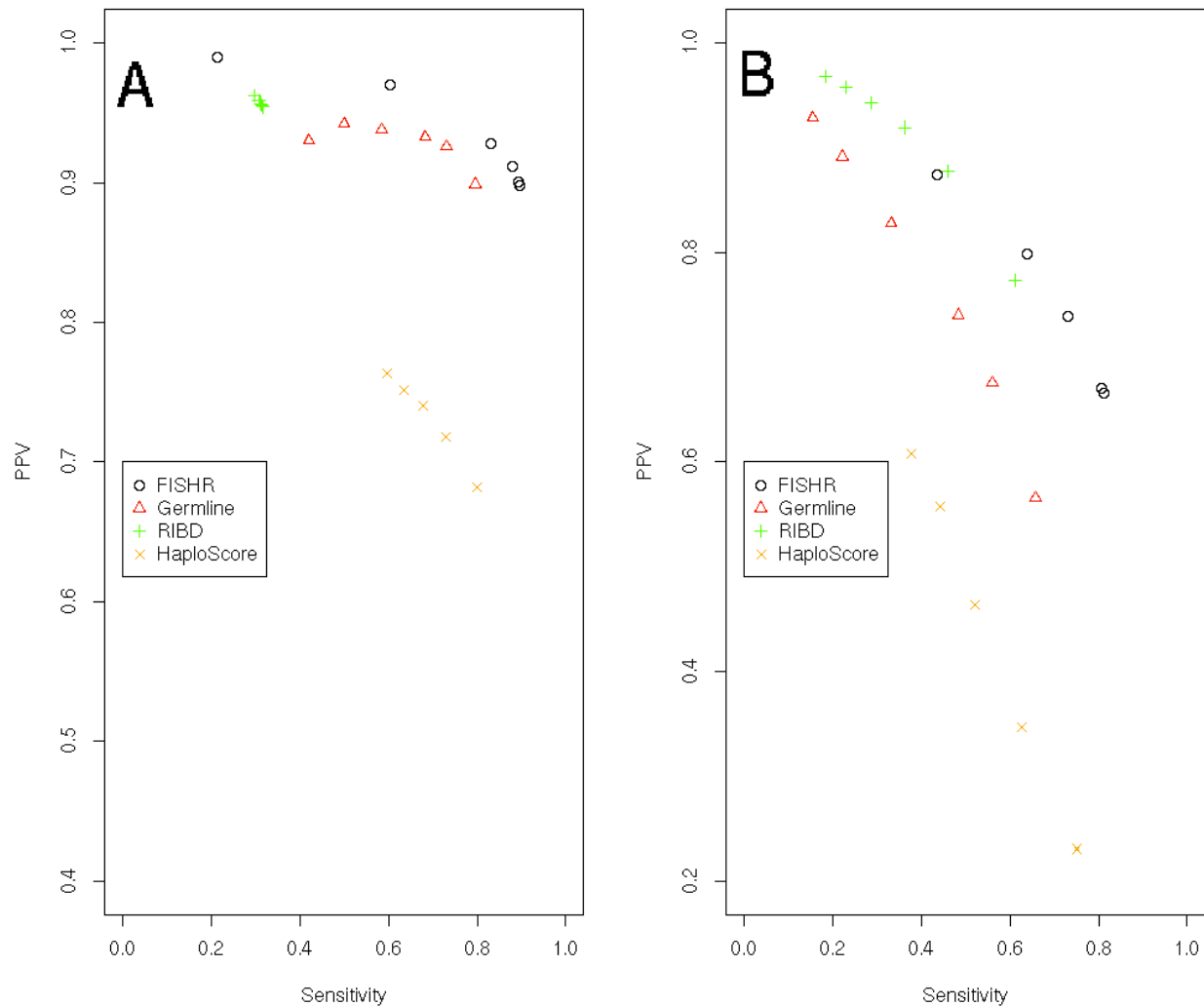
267 **Figure 3.** PPV-Sensitivity plots for FISHR (o), GERMLINE ( $\Delta$ ), rIBD (+), and HaploScore (x)

268 when calculated using a minimum of 3 cM for called IBD and a minimum of 3 cM for true IBD

269 (A) and when using a minimum of 1 cM for called IBD and a minimum of 1 cM for true IBD

270 (B).

271



272

273

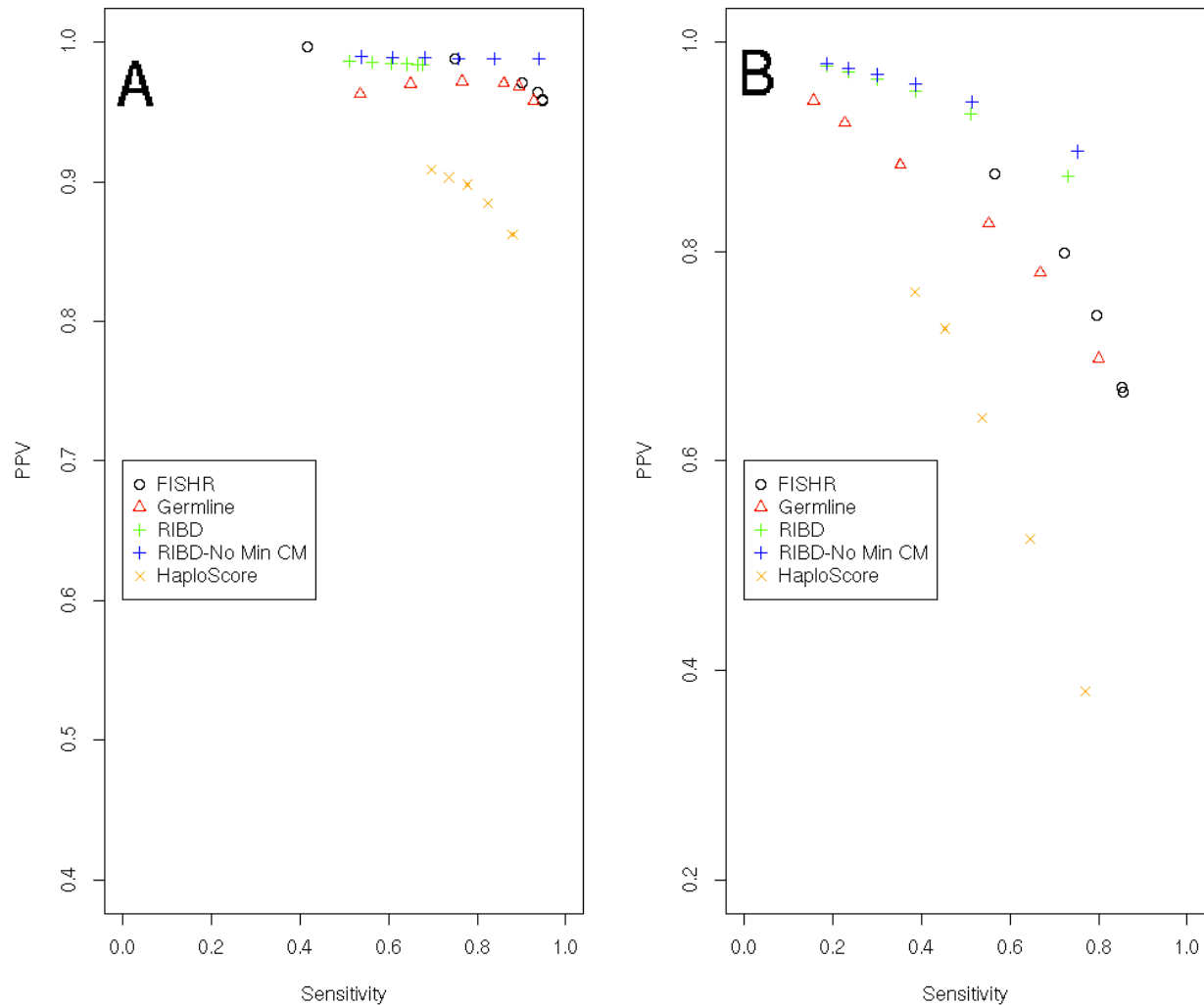
274 **Figure 4.** PPV-Sensitivity plots for FISHR (o), GERMLINE (Δ), rIBD (+), and HaploScore (x)

275 when calculated using a minimum of 3 cM for called IBD and a minimum of 1.5 cM for true

276 IBD (A) and when using a minimum of 1 cM for called IBD and a minimum of 0.5 cM for true

277 IBD (B). Additional measures are present for rIBD (+) using a minimum true IBD length of 0.5

278 cM for PPV and no minimum called cM length for sensitivity (A) and a minimum true IBD  
279 length of 0.25 cM for PPV and no minimum called cM length for sensitivity (B).  
280



281

282

### 283 *Accuracy of called segment endpoints in simulated data*

284 As noted above, the differences between the results in Figures 3 and 4 correspond to how

285 accurately the endpoints were estimated by each program. To quantify accuracy of endpoint

286 estimation, we first found optimal parameters for each program by searching through

287 combinations of the various input parameters, choosing those that maximized the sum of PPV



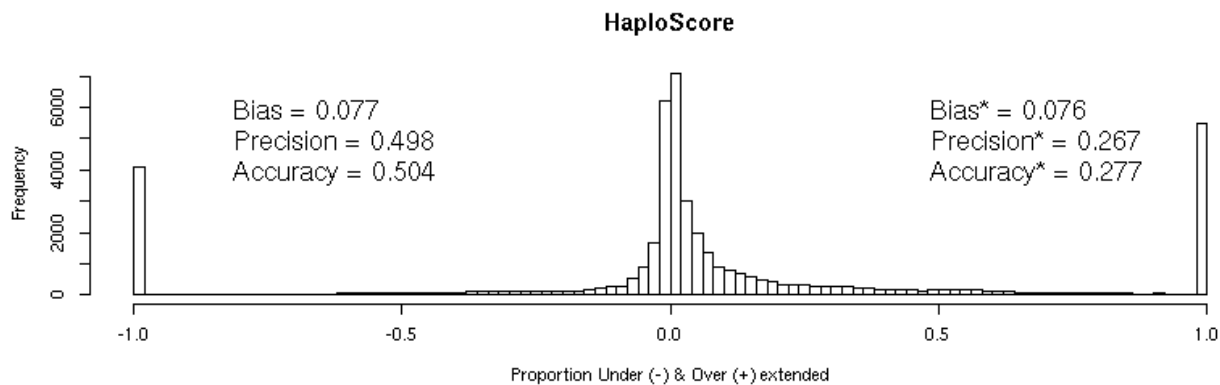
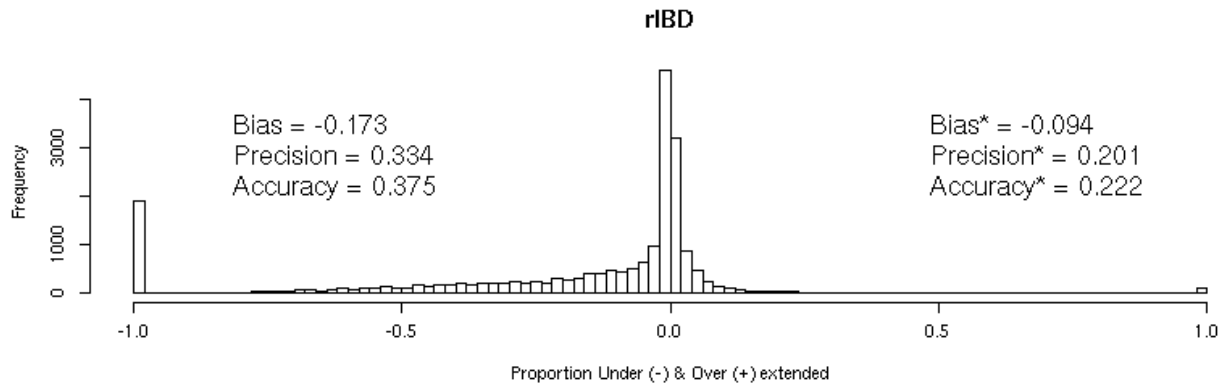
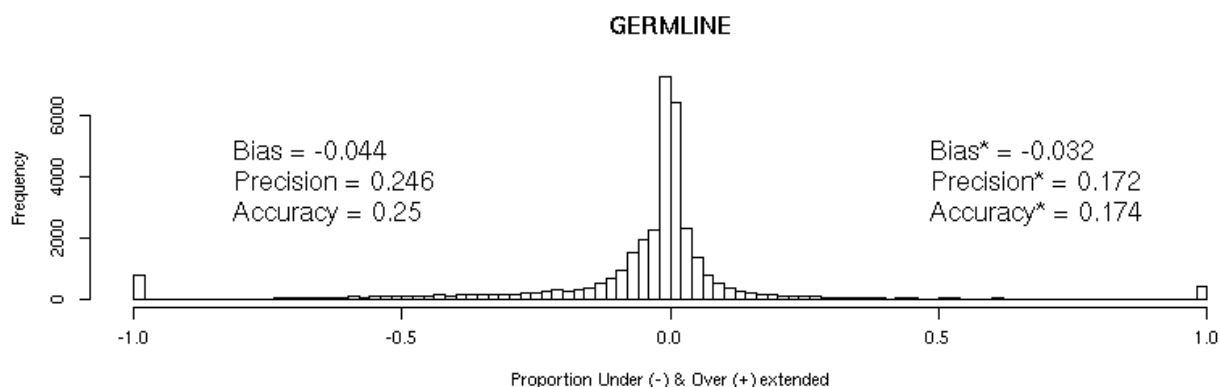
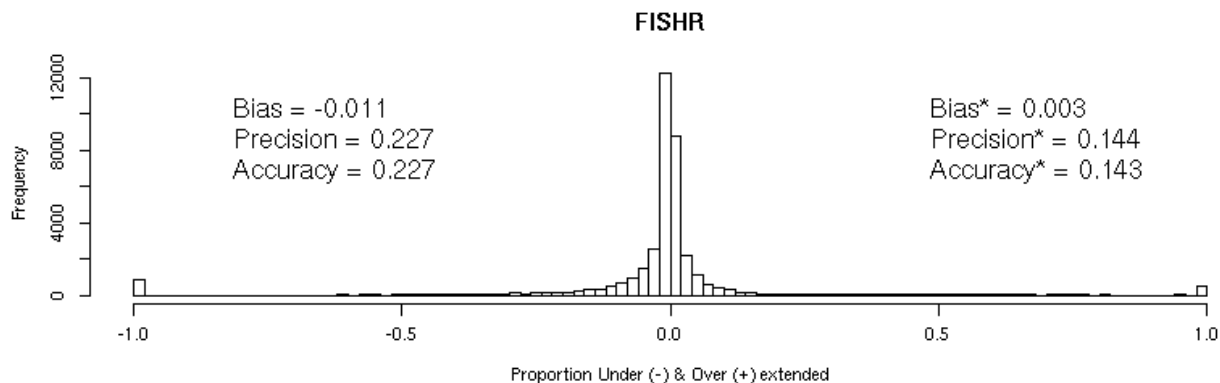
288 and sensitivity. Using these parameters, we divided the length of over- or underextension of each  
289 called segment endpoint by the length of the corresponding true IBD segment. Figure 5 shows  
290 the distribution of these proportions—the degree to which each endpoint was over- or  
291 underextended—when called segments had minimum length of 3 cM and true IBD segments had  
292 minimum length of 1.5 cM (results for 1 cM called and .5 cM true thresholds are shown in  
293 Supplemental Figure S1). It should be noted that using a 3 cM threshold for called and 1.5 cM  
294 for true IBD segments was the optimal scenario for all programs (Supplemental Figure S2). Any  
295 called segment that had no corresponding true IBD segment (false positive) was given an  
296 arbitrary value of 1 and any truly IBD segment with no corresponding called segment (false  
297 negative) was given a value of -1. The text to the left of each histogram shows the bias (defined  
298 as the mean proportion), precision (defined as the standard deviation of the proportion), and  
299 accuracy (defined as the standard deviation from 0 rather than from the mean proportion) when  
300 the false positive and false negative calls were included. Accuracy provides an estimate of how  
301 accurate the called segments are compared to perfect calls with no under- or overextension, and  
302 incorporates information on both bias and precision ( $\text{accuracy}^2 = \text{bias}^2 + \text{precision}^2$ ). FISHR had  
303 the most accurate (0.227) endpoints and was the most precise (0.227) of all algorithms. FISHR  
304 also showed very little bias (-0.011) with respect to under- or overextending calls. HaploScore  
305 (bias = 0.077) tended to overextend segments, whereas GERMLINE (bias = -0.044) and to a  
306 greater extent rIBD (bias = -0.177) tended to call segments that were shorter than the true IBD  
307 segments. rIBD also tends to miss truly IBD segments at a much higher rate than either FISHR  
308 or GERMLINE while HaploScore tends to both miss true IBD segments and call segments which  
309 are not IBD, as shown by the large values at -1 and 1, respectively. These conclusions remained

310 unchanged when we excluded false positive and false negative calls (reported on the right side of  
311 histograms in Figure 5).

312

313 **Figure 5.** Histograms displaying the distributions of the proportional under- and overextension  
314 for each called IBD segment for FISHR, GERMLINE, rIBD, and HaploScore, with the bias,  
315 precision, and accuracy observed for each program. Results were found using a minimum of 3  
316 cM for called segments and 1.5 cM for true IBD segments. All called segments with no  
317 corresponding true IBD segments (the entire segment was overextended) were classified as 1,  
318 and all true segments with no corresponding called segments (the entire “called” segment was  
319 underextended) were classified as -1. Results listed on the left sides on the histograms include  
320 these false positive and false negative calls while the results listed on the right sides of  
321 histograms marked with a \* only included the called segments which had a corresponding true  
322 IBD segment.

323



325

326 *Accuracy of called segment endpoints in real data*

327 All previous results used simulated data where the true IBD segment endpoints were known  
328 within a small margin of error. To determine how well the programs detect IBD segment  
329 endpoints in real data, we obtained data from 1,872 unrelated individuals from the UK10K  
330 dataset (The UK10K Consortium 2015), who were whole-genome sequenced at over 28 million  
331 markers. We extracted markers in the Illumina 650K SNP panel, re-phased them using  
332 SHAPEIT2 (Delaneau et al. 2012), and called segments from each of the four programs on this  
333 SNP dataset (see *Methods*). All remaining markers were retained as a holdout sample to calculate  
334 opposite homozygosity (OH) in and around regions where segments were called by each  
335 program. OH (e.g., an A-A genotype in one individual and a C-C genotype in the other) at  
336 masked markers within and around the called segments can be used to estimate the programs'  
337 rates of false-positive and false-negative calls and to infer where called segments over- or  
338 underextended true IBD segments (Browning and Browning 2012). Even when the underlying  
339 haplotypes are truly IBD, sporadic mismatching alleles within a called segment can occur due to  
340 SNP errors, and a string of such mismatches can occur due to one or more phase errors.  
341 However, phase errors cannot cause OH at true IBD locations; only the rare event of SNP call  
342 errors changing a heterozygous SNP to the opposite homozygous call can cause (very low levels  
343 of) sporadic false OH in the data. Therefore, locations where the rate of OH in holdout markers  
344 is high *within* the boundaries of called segments suggest regions of false positive calls (typically  
345 overextended segments), whereas locations where the rate of OH is low *outside* the boundaries  
346 of called segments suggest regions of false negative calls (typically underextended calls).

347

348 Figure 6 shows an example of a region where all four programs called a segment between two  
349 individuals and the locations where OH occurred in the holdout sequence data. To compare these  
350 instances of OH to the rate of OH expected in a pair of non-IBD segments, we also show the  
351 locations of OH at all holdout markers between a pair of randomly selected individuals at this  
352 location. Given the highly discrepant rate of OH between the focal pair and the rate of OH  
353 between the random pair, it is safe to assume that a true IBD segment existed between the focal  
354 individuals at this region, and the endpoints of this true IBD segment can be roughly inferred  
355 from where the OH rates between the focal individuals increase in the holdout sequence data.  
356 The results depict a fairly typical example in which rIBD apparently broke up a long true IBD  
357 segment into multiple short called segments. FISHR, GERMLINE, and HaploScore appear to  
358 have done better in this example at discovering one long true IBD segment, with the main  
359 differences between programs being where the endpoints were estimated. Supplemental Figures  
360 S3-S22 display an additional 20 similar examples chosen at random from among 5 FISHR called  
361 segments, 5 rIBD called segments, 5 HaploScore called segments, and 5 GERMLINE called  
362 segments.

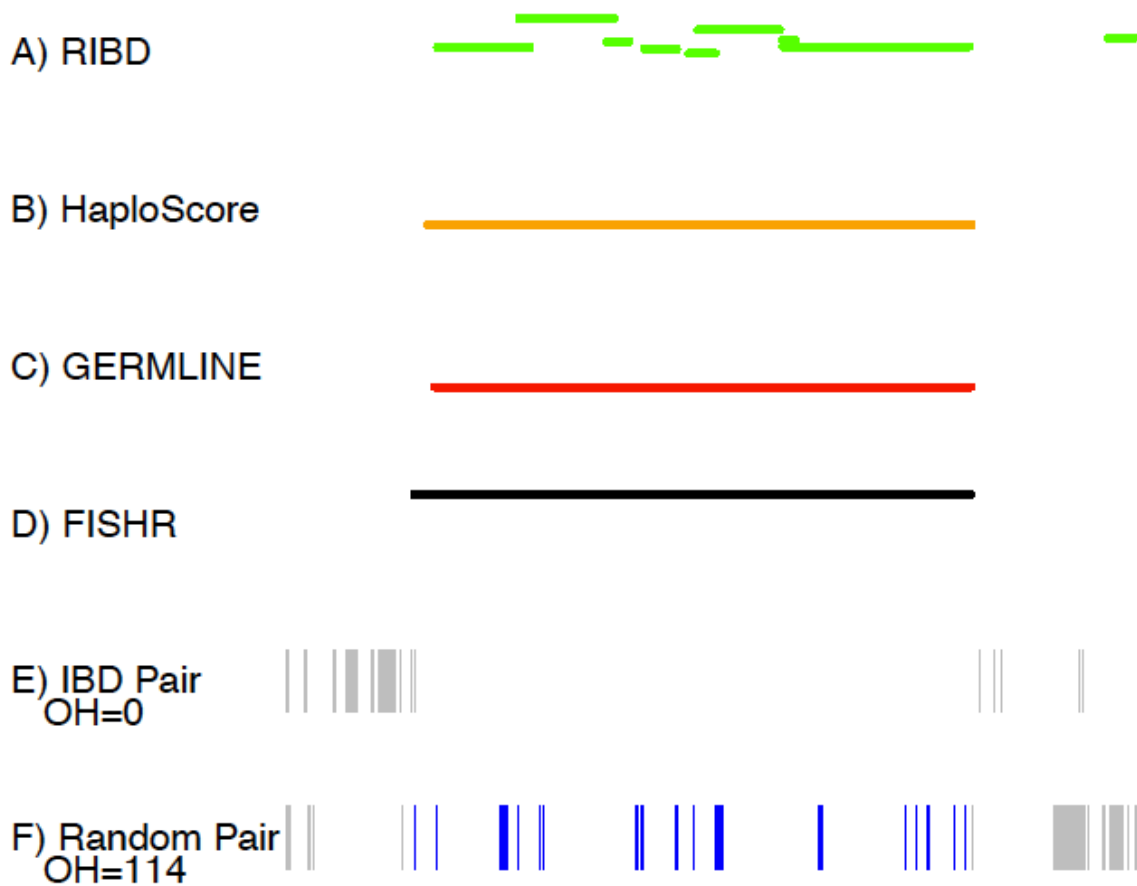
363  
364 To quantify the accuracy of the called segment endpoints for each program in this real dataset,  
365 we calculated the proportion of OH (POH) of holdout markers in 4 quarters of each called  
366 segment from the UK10K data, as well as two regions of the same base-pair length upstream and  
367 downstream from the called segment. We then calculated the average POH of the four quarters  
368 and two quarter-length flanking regions for each called segment. These results are presented in  
369 Figure 7 and corroborate our earlier conclusions about endpoint accuracy of the four programs in  
370 the simulated data (Figure 5). Figure 7A displays the four quarters of the called segment and the

371 flanking regions, whereas the Figure 7B displays only the first through fourth quarters within the  
372 called segments on an expanded scale. Figure 8 illustrates how these POH profiles should appear  
373 for programs that estimate endpoints perfectly, tend to underextend them, tend to overextend, or  
374 both. Of the four programs, the POH profile of FISHR was the most similar to the profile  
375 expected when the estimated endpoints of the called segments are perfect (Figure 8A); FISHR  
376 had levels of POH in the two flanking regions (“downstream” and “upstream”) very close to that  
377 between pairs of random individuals, indicating very little under-extension, and it had ~0 POH in  
378 quarters 1 through 4, indicating very little overextension. rIBD was very precise at finding  
379 segments that were truly IBD (~0 POH in quarters 1 through 4), but as predicted, it tended to  
380 under-extend the IBD segments much more than any of the other programs (low POH in the  
381 flanking regions). On the other hand, HaploScore tended to overextend true IBD segments, as  
382 indicated by its higher POH in the first and fourth quarters. GERMLINE tended to both  
383 overextend called segments and under-extend them, especially at the beginning of called  
384 segments. Supplementary Figure S23 illustrates the same POH analysis but instead uses a  
385 minimum of 1 cM for called IBD segments.

386

387 **Figure 6.** An example of called IBD segments between two individuals in the UK10K dataset,  
388 from (A) rIBD, (B) HaploScore, (C) GERMLINE, and (D) FISHR, with (E) opposite  
389 homozygous SNPs (OH) occurring for that pair of individuals in and surrounding the FISHR  
390 called IBD segment with the number of OH within the called segment listed, and (F) OH  
391 occurring in a random pair of individuals at the same location of the called IBD segment with the  
392 number of OH listed. (Note that rIBD can call two individuals as IBD 2 at some locations, i.e.  
393 sharing two IBD haplotypes; hence the overlapping segments shown for that program.)

394

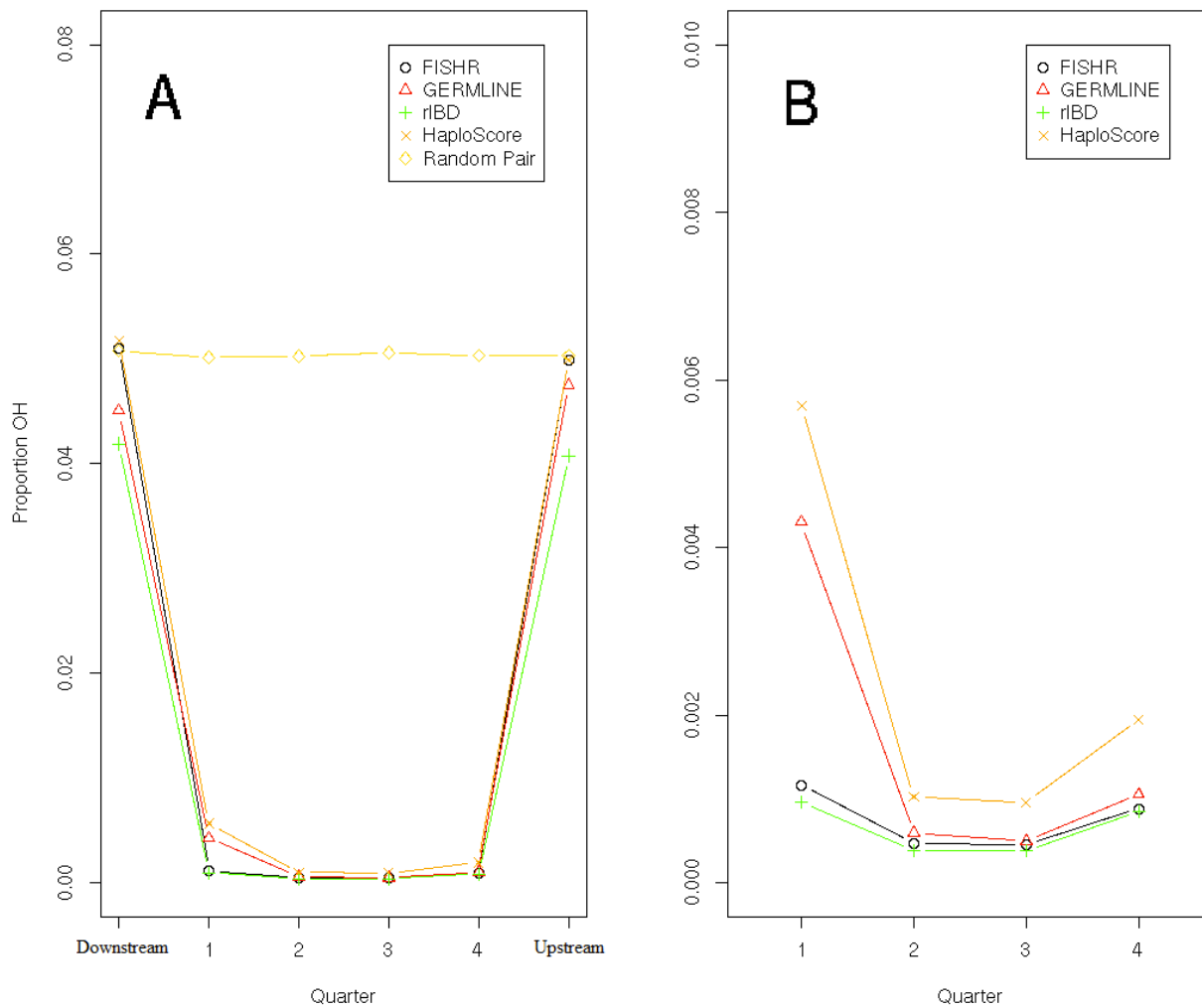


395

396

397 **Figure 7.** Results of the analysis of proportion of opposite homozygosity (OH) in (A) four  
398 quarters of called IBD segment and the two flanking regions and in (B) just the four quarters of  
399 the called IBD segments for FISHR (o), GERMLINE ( $\Delta$ ), rIBD (+), HaploScore (x), and random  
400 individuals at the same location of called IBD ( $\diamond$ ) where called IBD segments were a minimum  
401 of 3 cM.

402



403

404

405 **Figure 8.** Examples that summarize the proportion of opposite homozygosity (POH) calculated

406 from 4 quarters within and the two flanking regions around each called IBD segment, with the

407 POH for the called segment in black and the POH of segments from random individuals at the

408 same location in the genome presented in red. A program that makes every IBD call perfectly

409 from perfectly genotyped data (A) would have no OH in quarters one through four and the same

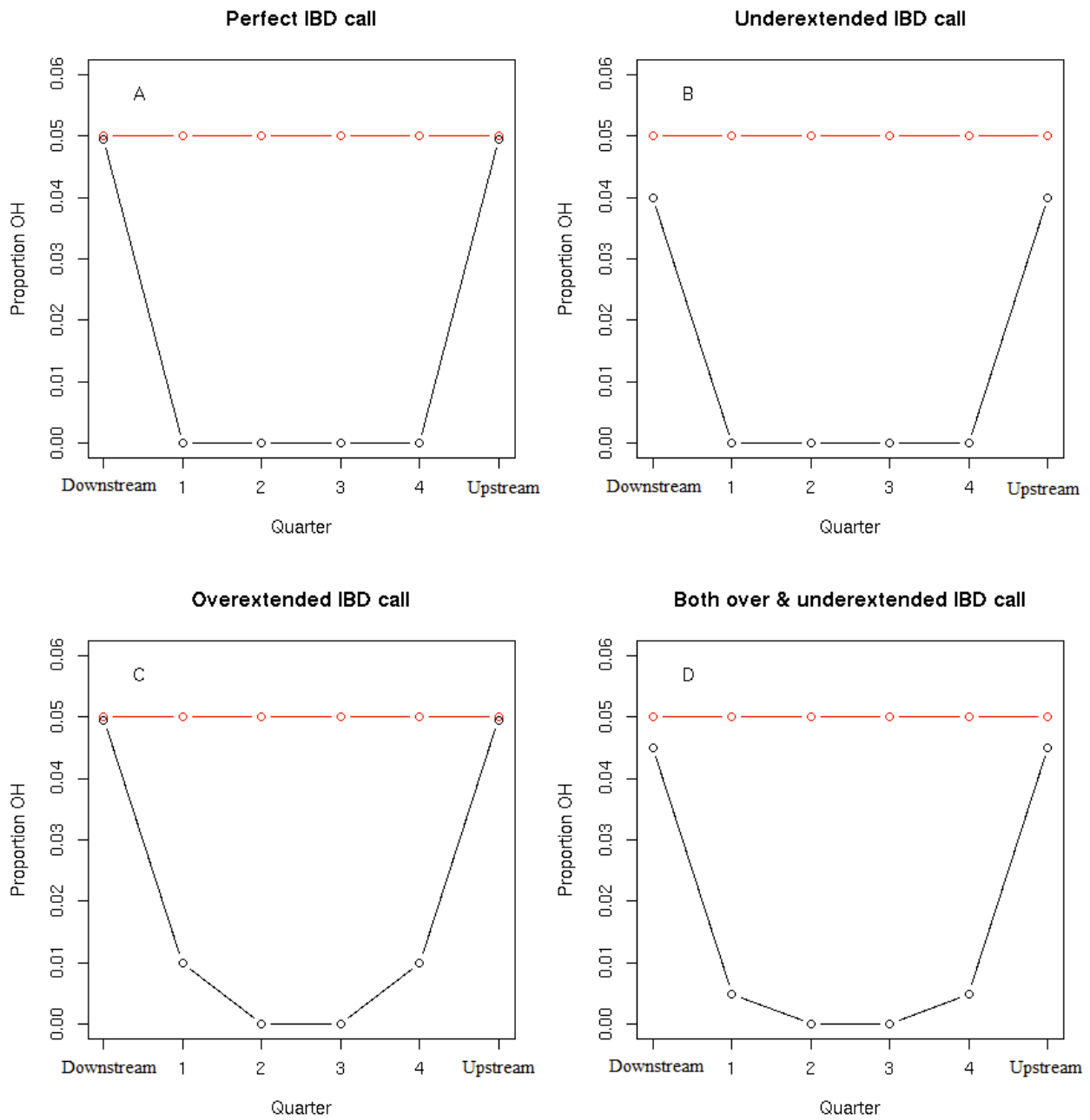
410 POH as random segments in the flanking regions. A program that underextends calls (B) would

411 have no OH in quarters one through four and lower POH in the flanking regions than random

412 segments. A program that overextends calls (C) would have positive POH in the first and fourth



413 quarters and the same POH in flanking regions as random segments. A program that both under-  
414 and overextends IBD calls (D) would display both increased POH in quarters one and four and  
415 decreased POH in the flanking regions.  
416



417

418

## 419 **Discussion**

420 We developed FISHR as an alternative method to detect segments of the genome shared IBD  
421 between pairs of individuals in a sample measured on genome-wide SNP data. Our goal was to  
422 develop a program that would be fast enough to be utilized with very large SNP datasets and be  
423 more accurate than existing programs at detecting IBD segments and their true endpoints. As we  
424 demonstrated using simulated data where true IBD status was known, FISHR performs as well or  
425 better than all competitor programs in terms of PPV and sensitivity for detecting long IBD  
426 segments, while slightly worse than rIBD but better than GERMLINE and HaploScore at  
427 detecting short IBD segments. Furthermore, as we demonstrated in both simulated and real data,  
428 FISHR is substantially more accurate than any existing program at estimating the correct  
429 endpoints of IBD segments. Accurately estimating these endpoints is important for several  
430 reasons. First, the length of IBD segments is relevant to many parameters of interest in  
431 population genetics (time to recent common ancestor, effective population size, population  
432 bottlenecks, etc.); systematic biases in estimating these lengths can lead to incorrect conclusions  
433 regarding these and other parameters. Second, phasing and imputation (Kong et al. 2008) based  
434 on IBD segments can be affected by the accuracy of the endpoints, with under- and  
435 overextensions of IBD segments causing regions to be incorrectly imputed or phased. Finally, in  
436 calculating genome-wide relatedness using IBD segments (Browning and Browning 2013),  
437 programs that tend to overextend IBD calls will lead to systematically inflated relatedness, and  
438 those that tend to underextend IBD calls to deflated relatedness.

439

440 Despite the computationally efficient, deterministic algorithm FISHR uses to call candidate  
441 segments (see *Methods*), FISHR remains surprisingly accurate. It is fast enough to be used on

442 very large SNP datasets (e.g., 20,000-50,000 individuals), running two to five times slower than  
443 GERMLINE but running over a thousand times faster than rIBD and HaploScore at large sample  
444 sizes. One practical downside of FISHR is that it requires much more RAM than its competitors.  
445 This is because FISHR attempts to stitch together long called segments that are separated by a  
446 small number of SNPs, which may represent erroneously split IBD segments (although FISHR  
447 may subsequently break up some of these consolidated segments if the data suggests the full  
448 segment is not IBD). To accomplish this, FISHR must pull all the candidate segments from  
449 GERMLINE into RAM to sort them, making its memory overhead high compared to programs  
450 such as GERMLINE that simply stream data. However, given that the price of RAM is  
451 plummeting, and that the RAM capacity of many high-performance computers (e.g., 1 Tb) is  
452 already sufficiently large for FISHR to be applied on samples of ~100,000, we do not see this as  
453 a major impediment to using the program. Nevertheless, we have developed a version of FISHR  
454 (accessed using the `-low_ram` flag) that uses a negligible amount of RAM at the cost of failing to  
455 stitch together called segments that are erroneously split. The accuracy of this version of FISHR  
456 is only slightly degraded compared to the default version.

457

458 Another limitation of FISHR vis-à-vis rIBD is that, using the approach we presented here, it  
459 cannot call regions that are greater than IBD 1 – i.e., where more than one IBD segment exists at  
460 the same location between individuals. For example, ~25% of regions between siblings are  
461 expected to be IBD 2, meaning both haplotypes are IBD. FISHR (as well as GERMLINE) would  
462 call these regions as IBD 1, whereas rIBD can call these regions as IBD 2 (or greater). We have  
463 incorporated a method for detecting such multi-IBD states into FISHR (by post-processing  
464 GERMLINE segments found using the `-haploid` flag), but because such IBD 2+ situations are

465 extremely rare among unrelated individuals (occurring at a rate proportional to the square of  
466 relatedness, or  $\sim 0.0001$  for IBD 2 vs. 0.01 for IBD 1 in typical datasets of nominally unrelated  
467 individuals), the benefit of these additional called segments did not outweigh the cost in missing  
468 truly IBD segments incurred by post-processing data called using *-haploid* in GERMLINE.  
469 Nevertheless, the standard version's limitation to detecting IBD 1 must be kept in mind when  
470 working with highly related samples.

471

## 472 **Conclusion**

473 With increasingly large whole-genome SNP datasets being accumulated, it is important to have a  
474 method for detecting IBD segments that is both accurate and efficient. We introduced a program,  
475 FISHR, that accomplishes both, and that is particularly accurate at determination of the correct  
476 endpoints of IBD segments. We demonstrated these properties using simulations, and confirmed  
477 these conclusions using a novel approach on real sequence data from the UK10K project. Due to  
478 the number of pairwise comparisons that must be made in IBD detection, computationally  
479 intensive programs such as rIBD and HaploScore cannot be easily run on datasets of more than  
480  $\sim 10,000$  individuals. FISHR is a more accurate alternative to GERMLINE as an IBD detection  
481 program on large datasets, with only a modest increase in runtime.

482

## 483 **Methods**

484 *Description of the FISHR algorithm*

485 FISHR is written in C++ and is available freely for download at

486 [http://matthewckeller.com/html/program\\_code.html](http://matthewckeller.com/html/program_code.html). FISHR utilizes GERMLINE (described in

487 detail by Gusev et al. 2009), as an initial screen to quickly detect candidate segments. In

488 particular, in the results presented here, we used the *-h\_extend* method in GERMLINE, which  
489 incorporates information on phased mismatches and which we found to be the most accurate of  
490 the three alternative methods (*-h\_extend*, *-w\_extend*, and *-haploid*) GERMLINE uses. FISHR  
491 then further refines the called segments as follows. First, because two long IBD calls that are  
492 separated by a short distance may actually be a single contiguous IBD segment that was  
493 artificially broken apart in GERMLINE due to phase or SNP call errors, FISHR stitches together  
494 segments separated by a user-defined number of SNPs (*-gap*). Next, FISHR finds the locations of  
495 IEs for all called segments. To do this, FISHR finds the longest exact match between either of  
496 the two phased haplotypes of the first person and either of the two phased haplotypes of the  
497 second person (a total of four possible combinations), starting at the first SNP of the called  
498 segment. An IE occurs at the first mismatching SNP after the exact match ends. FISHR then  
499 finds the next longest exact match between any of the four possible combinations of phased  
500 haplotypes, starting from the SNP following the previous IE, and extends until the next  
501 mismatching SNP is encountered. This process is continued until the end of the called segment.  
502  
503 IEs represent locations along a candidate segment that are potentially inconsistent with IBD  
504 inheritance. Some IEs are expected by chance due to SNP and phase errors even in truly IBD  
505 segments. However, too many IEs within a particular region are a likely signal that the segment  
506 is not IBD in that area and that the segment should be truncated (if near an endpoint of the  
507 segment) or split into two (if in the middle of the segment). To determine such called segment  
508 endpoints, FISHR calculates a moving average (MA) of IEs centered at each SNP within a user-  
509 defined window (using the *-window* flag) of SNPs, as outlined in Figure 9. FISHR then starts at  
510 the center of the called IBD segment and moves towards each endpoint until it reaches the first

511 SNP with a MA value greater than the user-defined maximum (*-emp\_ma\_threshold*), as shown in  
512 Figure 10. These points signal the endpoints of a called segment. Note that in addition to  
513 trimming the segment ends, this process can split a GERMLINE candidate segment into two or  
514 more shorter segments. Moreover, if the flag *-count\_gap\_errors* is set to TRUE, as it is by  
515 default, segments that had been stitched together from the first step can be broken up again at this  
516 stage if enough IEs are clustered near the gap. Because segments that are too short, in terms of  
517 either number of SNPs or cM distance, are increasingly likely to be false positives, FISHR then  
518 drops segments shorter than user-defined thresholds of both SNP and cM length (using  
519 the *-min\_snp* and *-min\_cm* flags, respectively). The final process FISHR performs is to calculate  
520 the total proportion of SNPs that are IEs (PIE) within each segment. Too many IEs scattered  
521 across the entire length of a segment are a signal that the whole segment is unlikely to be IBD.  
522 Thus, if the PIE of a segment is greater than the value supplied in the *-emp\_pie\_threshold*  
523 argument, the segment is dropped.

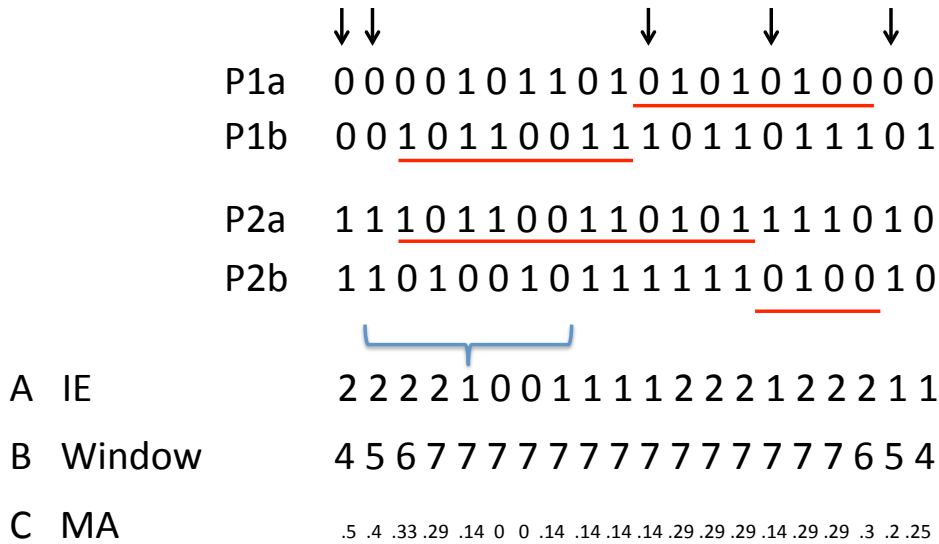
524  
525 Because values of PIE and MA depend on the quality of SNP calls and phasing in the data at  
526 hand, the thresholds for these values require careful consideration by users. The approach we  
527 recommend and that we used here was to identify long stretches (>8 cM) of the genome where  
528 no opposite homozygotes occurred between pairs of individuals (this can be accomplished using  
529 GERMLINE *-w\_extend* flag without the *-h\_extend* flag). Because information on phase was not  
530 used in calling these segments, they are not biased to be in regions where phasing is more  
531 difficult. We then found the distribution of the PIE and maximum MA values calculated from the  
532 middlemost 50% of these segments, which can be assumed with high confidence to be truly IBD.  
533 We compared these distributions to distributions of PIE and maximum MA values calculated

534 from segments matched in location to the likely IBD segments but that were between random  
535 pairs of individuals. The PPV and sensitivity that will result from any choices of PIE and MA  
536 thresholds can be estimated by how well those thresholds separate these distributions, and thus  
537 thresholds can be chosen that lead to a desired PPV-sensitivity combination. We have supplied a  
538 utility (*gl\_parameter\_finder*) for accomplishing this step along with the FISHR download.

539

540 **Figure 9.** Calculating the moving average (MA) of implied errors (IE) of a potential IBD call  
541 between two individuals, P1 & P2. The red underlined segments indicate the called haplotypes,  
542 and the arrows designate where IEs occur in the call. Using a moving window size of 7, line A  
543 displays the number of IEs within the window for each given SNP, line B displays the window  
544 size (which is truncated at each end of the “chromosome”), and line C displays the MA for each  
545 SNP.

546



547

548

549 **Figure 10.** An example how FISHR utilizes the moving average of implied errors (MA)

550 calculated for each SNP to determine endpoints of a called IBD segment. In this case the

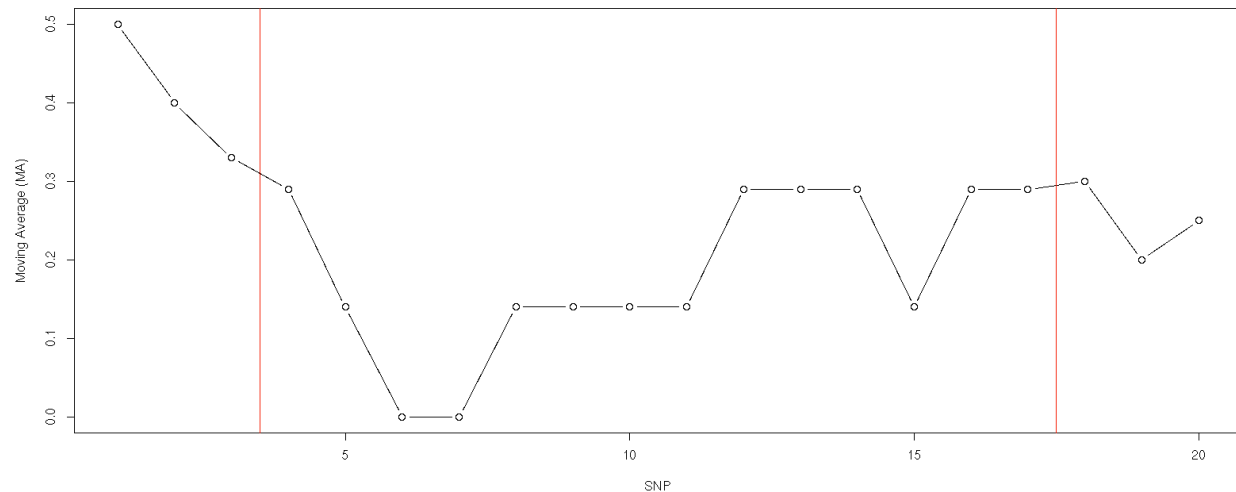
551 maximum allowed MA is 0.3. The red vertical line denotes the location of where the MA

552 increases above the 0.3 threshold, leaving the called segment to consist of the SNPs between the

553 red lines.

554





555

556

### 557 *Simulated Sequence and SNP Data*

558 We simulated genotypic data using the sequence simulator HAPGEN2 (Su et al. 2011), which  
559 simulated haplotypes by conditioning on a reference set of population haplotypes (here, the 1000  
560 Genomes Project (Clarke et al. 2012) European ancestry (CEU) haplotypes of chromosome 15)  
561 and created a new population by combining haplotypes according to a fine-scaled recombination  
562 rate map (from deCODE; Kong et al. 2010). Here, we defined the effective population sizes as  
563 11,418 and the sample size (defined as “controls” in HAPGEN2) as 28,000. For computational  
564 efficiency, we created 13 independent datasets of 1,000 individuals each and averaged all results  
565 across these 13 replicates. The data had LD, haplotype diversity, and allele frequency  
566 distributions that mimic those in the initial set of haplotypes.

567

568 We used the perfectly phased, simulated sequence data with no errors obtained from HAPGEN2  
569 to obtain “true IBD segments.” Because no program exists to our knowledge that tracks IBD  
570 status between pairs of haplotypes, we defined true IBD segments as perfectly matching  
571 haplotypes that spanned a desired cM threshold (0.25, 0.5, 1.5, or 3, depending on the analysis).

572 To increase computational efficiency and to ensure that rare mutations that arose on a haplotype  
573 since the common ancestor did not cause a true IBD segment to be missed, we pruned this  
574 sequence data to have  $MAF > .05$ , resulting in a density of  $\sim 1$  variant per 1000 base pairs. To  
575 create data that mimicked post-quality-control SNP data on existing platforms, we then extracted  
576 SNPs pseudo-randomly such that the MAF distribution was about uniform and the density of  
577 SNPs was one per 6,750 base pairs (corresponding to  $\sim 400,000$  SNPs genomewide). To simulate  
578 SNP call errors, we randomly changed one allele to its alternative allele at a rate of 0.2%, in the  
579 middle of what has been found empirically for SNP calls (Steemers and Gunderson 2007; Teo et  
580 al. 2007; Korn et al. 2008; Hong et al. 2012). Finally, we unphased the SNP data and rephased it  
581 using SHAPEIT2 (Delaneau et al. 2012).

582

### 583 *Real Sequence Data*

584 We also compared performance of the IBD detection algorithms using the UK10K ALSPAC  
585 sequence data on 1,872 unrelated individuals (The UK10K Consortium 2015). In this data, we  
586 utilized 4 subchromosomes (5q, 9q, 14q, and 20q) and removed markers with less than a 1%  
587 MAF, markers in violation of Hardy-Weinberg equilibrium with p-values of less than 0.0001,  
588 and markers that contained missing data for any individuals. We then extracted SNPs that were  
589 on the Illumina 650K SNP panel (21,802 markers for subchromosome 5q, 13,716 markers for  
590 subchromosome 9q, 16,199 markers for subchromosome 14q, and 6,307 markers for  
591 subchromosome 20q) and phased this data using SHAPEIT2 for calling segments using each  
592 program. We retained the remaining markers not in the SNP data (an average of one marker per  
593 3,000 base pairs) as a holdout sample to calculate the proportion of opposite-homozygote SNPs  
594 within called segments.

595

596 *Running the four IBD detection programs*

597 We ran FISHR, GERMLINE, rIBD, and HaploScore on the simulated SNP data that was phased  
598 using SHAPEIT2, varying input parameters to determine the optimal parameters for discovering  
599 IBD segments with minimum lengths both of 1 and 3 cM for each program (optimal parameters  
600 for finding IBD with a minimum of 3 cM bolded). For FISHR, we varied the candidate segment  
601 detection parameters (using GERMLINE) *-h\_extend* vs. *-w\_extend*, *-bits* (30, 45, **60**, 75, or 90),  
602 *-err\_het* (0, **1**, 2, 3), *-err\_hom* (0, **1**, 2, 3), as well as the FISHR-specific parameters *-gap* (0, 1, or  
603 **30**), *-count\_gap\_errors* (**TRUE** or FALSE), *-emp\_ma\_threshold* (0.025, **0.045**, 0.065, 0.085),  
604 and *-emp\_pie\_threshold* (0.005, **0.015**, 0.025). For GERMLINE, we compared both  
605 the *-h\_extend* vs. *-w\_extend* options and varied *-bits* (30, 45, **60**, 90, 120, 150), *-err\_het* (0, **1**, 2,  
606 3), and *-err\_hom* (0, **1**, 2, 3). For rIBD, we varied *-ibd lod* (**1**, 2, 3, 4, 5, 6), *-overlap* (100, **157**,  
607 200), *-window* (7,500, **10,000**, 12,500), *-scale* (2.5, 3, **3.16**, 3.5), and *-trim* (11, **16**, 21). Finally,  
608 for HaploScore, we varied the candidate segment detection parameters (using GERMLINE)  
609 of *-h\_extend* vs. *-w\_extend*, *-bits* (**30**, 45, 60, 90, 120), *-err\_het* (0, **1**, 2, 3), and *-err\_hom* (0, **1**,  
610 2, 3), and then varied the *-switch\_error* (**0.0005**, 0.001, 0.0015, 0.01), *-snp\_error* (**0.0006**,  
611 0.00125, 0.0025, 0.01), and HaploScore thresholds (**1**, 3, 5, 7, 9, 11, 13, 15) in HaploScore. For  
612 each program, we plotted the PPV and sensitivity, as shown in Figure 4, and the combination  
613 closest to perfect performance (Sensitivity=1 and PPV=1) was kept as the optimal for that  
614 specific program. The exact command lines used for each program with these optimal parameters  
615 are included in Supplemental Table S2. For Figures 3 and 4, we kept constant all the optimal  
616 parameters for each program other than the parameter that most influenced the PPV-sensitivity

617 tradeoff. In particular, we varied *-emp\_ma\_threshold* for FISHR, *-ibd lod* for rIBD, and the *-bits*  
618 argument for GERMLINE and HaploScore.  
619

## 620 **Data Access**

621 MGS dataset(s) used for the analyses described in this manuscript were obtained from dbGaP  
622 found at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP study accession numbers phs000167.  
623 This dataset was provided by Alan R. Sanders, M.D. CARDIA dataset(s) used for the analyses  
624 described in this manuscript were obtained from dbGaP found at  
625 <http://www.ncbi.nlm.nih.gov/gap> through dbGaP study accession numbers phs000309. The  
626 ARIC datasets used for the analyses described in this manuscript were obtained from dbGaP  
627 found at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP study accession numbers phs000090.  
628 The GENEVA datasets used for the analyses described in this manuscript were obtained from  
629 dbGaP found at <http://www.ncbi.nlm.nih.gov/gap> through dbGaP study accession numbers  
630 phs000091. Simulated data, scripts to evaluate IBD detection, and FISHR can be downloaded  
631 from our personal website, <http://matthewckeller.com/html/>.

632

## 633 **Acknowledgements**

634 The authors thank Nathan Lapinski, Teresa deCandia, and Rasool Tahmasbi for their help in  
635 coding, ideas, and writing.

636

637 The Atherosclerosis Risk in Communities Study was carried out as a collaborative study  
638 supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C,  
639 HHSN268201100006C, HHSN268201100007C, HHSN268201100008C,  
640 HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and  
641 HHSN268201100012C). The authors thank the staff and participants of the ARIC study for their

642 important contributions. Funding for GENEVA was provided by National Human Genome  
643 Research Institute grant U01HG004402 (E. Boerwinkle).

644

645 Funding support for the GWAS of Gene and Environment Initiatives in Type 2 Diabetes was  
646 provided through the NIH Genes, Environment and Health Initiative [GEI] (U01HG004399).

647 The human subjects participating in the GWAS derive from The Nurses' Health Study and

648 Health Professionals' Follow-up Study and these studies are supported by National Institutes of

649 Health grants CA87969, CA55075, and DK58845. Assistance with phenotype harmonization and

650 genotype cleaning, as well as with general study coordination, was provided by the Gene

651 Environment Association Studies, GENEVA Coordinating Center (U01 HG004446). Assistance

652 with data cleaning was provided by the National Center for Biotechnology Information. Funding

653 support for genotyping, which was performed at the Broad Institute of MIT and Harvard, was

654 provided by the NIH GEI (U01HG004424).

655

656 *Author contributions:* M.C.K. and M.J. conceived the concept for the program. U.L. and P.S.P.

657 coded the program. D.W.B and M.C.K. analyzed and compared results from the programs and

658 wrote the manuscript.

659

## 660 **Disclosure Declaration**

661 The authors have no financial disclosure to declare.

662

## 663 **References**

- 664 Browning BL, Browning SR. 2011. A fast, powerful method for detecting identity by  
665 descent. *The American Journal of Human Genetics* **88**: 173-182.
- 666 Browning SR, Browning BL. 2012. Identity by descent between distant relatives: detection  
667 and applications. *Annual review of genetics* **46**: 617-633.
- 668 Browning SR, Browning BL. 2013. Identity-by-descent-based heritability analysis in the  
669 Northern Finland Birth Cohort. *Human genetics* **132**: 129-138.
- 670 Clarke L, Zheng-Bradley X, Smith R, Kulesha E, Xiao C, Toneva I, Vaughan B, Preuss D,  
671 Leinonen R, Shumway M. 2012. The 1000 Genomes Project: data management and  
672 community access. *Nature methods* **9**: 459-462.
- 673 Delaneau O, Marchini J, Zagury J-F. 2012. A linear complexity phasing method for  
674 thousands of genomes. *Nature methods* **9**: 179-181.
- 675 Durand EY, Eriksson N, McLean CY. 2014. Reducing pervasive false-positive identical-by-  
676 descent segments detected by large-scale pedigree analysis. *Molecular biology and*  
677 *evolution*: msu151.
- 678 Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Pe'er I. 2009.  
679 Whole population, genome-wide mapping of hidden relatedness. *Genome research*  
680 **19**: 318-326.
- 681 Haldane J. 1919. The combination of linkage values and the calculation of distances  
682 between the loci of linked factors. *J Genet* **8**: 299-309.
- 683 Hong H, Xu L, Liu J, Jones WD, Su Z, Ning B, Perkins R, Ge W, Miclaus K, Zhang L. 2012.  
684 Technical reproducibility of genotyping SNP arrays used in genome-wide  
685 association studies. *PLoS One* **7**: e44483.

- 686 Keller MC, Visscher PM, Goddard ME. 2011. Quantification of inbreeding due to distant  
687 ancestors and its detection using dense single nucleotide polymorphism data.  
688 *Genetics* **189**: 237-249.
- 689 Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI,  
690 Ingason A, Steinberg S, Rafnar T. 2008. Detection of sharing by descent, long-range  
691 phasing and haplotype imputation. *Nature genetics* **40**: 1068-1075.
- 692 Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Walters  
693 GB, Jonasdottir A, Gylfason A, Kristinsson KT. 2010. Fine-scale recombination rate  
694 differences between sexes, populations and individuals. *Nature* **467**: 1099-1103.
- 695 Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J,  
696 Collins PJ, Darvishi K. 2008. Integrated genotype calling and association analysis of  
697 SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* **40**:  
698 1253-1260.
- 699 Palamara PF, Lencz T, Darvasi A, Pe'er I. 2012. Length distributions of identity by descent  
700 reveal fine-scale demographic history. *The American Journal of Human Genetics* **91**:  
701 809-822.
- 702 Powell JE, Visscher PM, Goddard ME. 2010. Reconciling the analysis of IBD and IBS in  
703 complex trait studies. *Nat Rev Genet* **11**: 800-805.
- 704 Schizophrenia Working Group of Psychiatric Genomics Consortium, Ripke S, Neal BM et al.  
705 2014. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*  
706 **511**: 421-427.
- 707 Setty MN, Gusev A, Pe'er I. 2011. HLA type inference via haplotypes identical by descent.  
708 *Journal of Computational Biology* **18**: 483-493.



709 Soi S, Scheinfeldt L, Lambert C, Hirbo J, Ranciaro A, Thompson S, Bodo J, Froment A,  
710 Ibrahim M, Juma A. 2011. Demographic histories of African hunting-gathering  
711 populations inferred from genome-wide SNP variation. In *International Congress of*  
712 *Human Genetics/American Society of Human Genetics meeting Montreal, Canada.*

713 Steemers FJ, Gunderson KL. 2007. Whole genome genotyping technologies on the  
714 BeadArray™ platform. *Biotechnology journal* **2**: 41-49.

715 Su Z, Marchini J, Donnelly P. 2011. HAPGEN2: simulation of multiple disease SNPs.  
716 *Bioinformatics* **27**: 2304-2305.

717 Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J,  
718 Landray M. 2015. UK Biobank: an Open Access resource for identifying the causes of  
719 a wide range of complex diseases of middle and old age. *PLoS medicine* **12**: 1-10.

720 Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, Clark TG. 2007. A  
721 genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* **23**:  
722 2741-2746.

723 The UK10K Consortium 2015. The UK10K project identifies rare variants in health and  
724 disease. *Nature* **526**: 82-90.

725 Vacic V, Ozelius LJ, Clark LN, Bar-Shira A, Gana-Weisz M, Gurevich T, Gusev A, Kedmi M,  
726 Kenny EE, Liu X. 2014. Genome-wide mapping of IBD segments in an Ashkenazi PD  
727 cohort identifies associated haplotypes. *Human molecular genetics* **23**: 4693-4702.

728

729