The American Journal of Human Genetics, Volume *99*

Supplemental Data

Robust Inference of Identity by Descent from Exome-Sequencing Data

Wenqing Fu, Sharon R. Browning, Brian L. Browning, and Joshua M. Akey

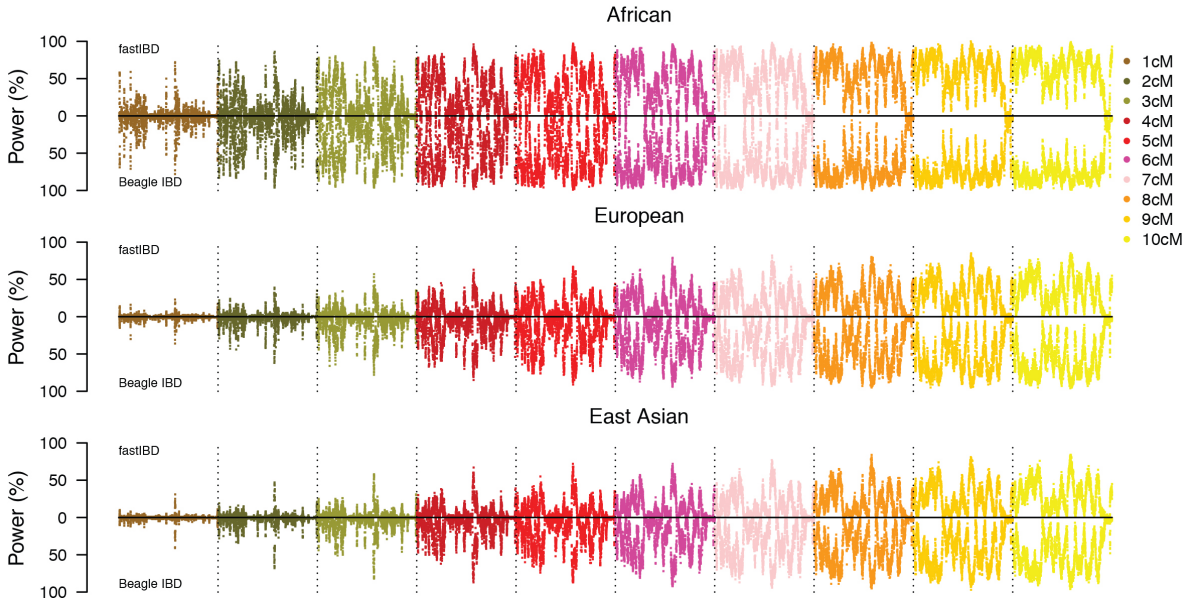

Figure S1 The locus specific power to detect IBD segments ranging in size from 1 cM to 10 cM in exome sequencing data is heterogeneous across chromosome 1. The locus specific power was evaluated by fastIBD with *fastibdthreshold* of 10^{-10} for ten independent runs and Beagle IBD with *ibd2nonibd* = 0.01 and *nonibd2ibd* = 0.0001 for five independent runs, respectively.

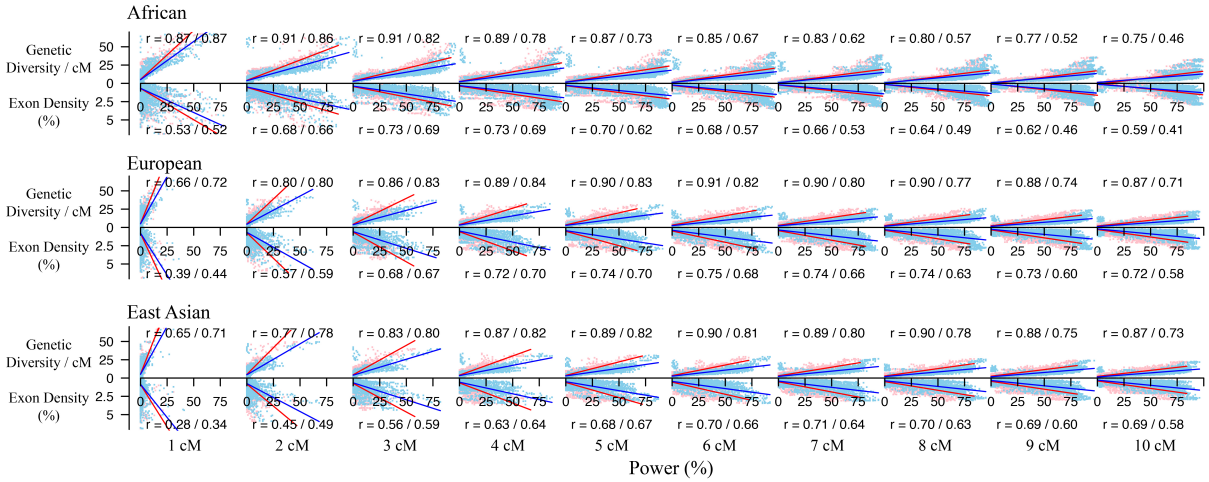

Figure S2 The locus specific power to detect IBD segments in exome sequencing data is positively correlated with exon density and exonic genetic diversity in the corresponding regions (Pearson's correlation test; $P < 10^{-15}$). The locus specific power to detect IBD segments ranging in size from 1 cM to 10 cM was evaluated by fastIBD with *fastibdthreshold* of 10^{-10} for ten independent runs (in red) and by Beagle IBD with *ibd2nonibd* = 0.01 and *nonibd2ibd* = 0.0001 (in blue) for five independent runs, respectively.

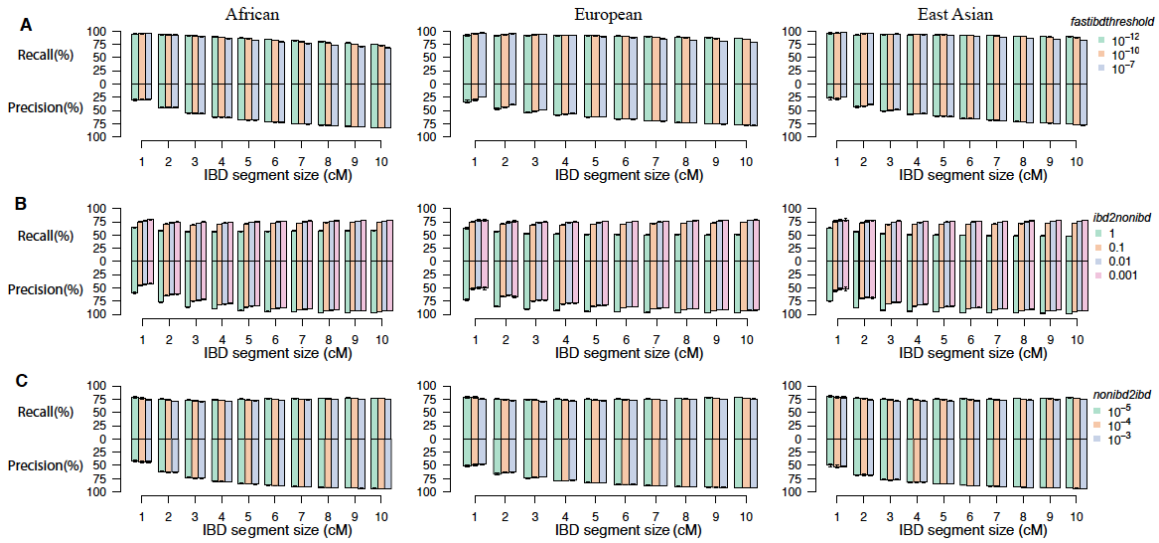

Figure S3 The average locus specific precision and recall (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM in exome sequencing data by Beagle fastIBD and Beagle IBD under different parameter settings, i.e., (**A**) Beagle fastIBD, *fastibdthreshold* = 10^{-12}, 10^{-10} and 10^{-7} (**B**) Beagle IBD, ibd2nonibd = 1, 0.1, 0.001, and 10^{-4}, and *nonibd2ibd* = 0.0001. (**C**) Beagle IBD, *nonibd2ibd* = 10^{-5}, 0.0001, and 0.001, and *ibd2nonibd* = 0.01.

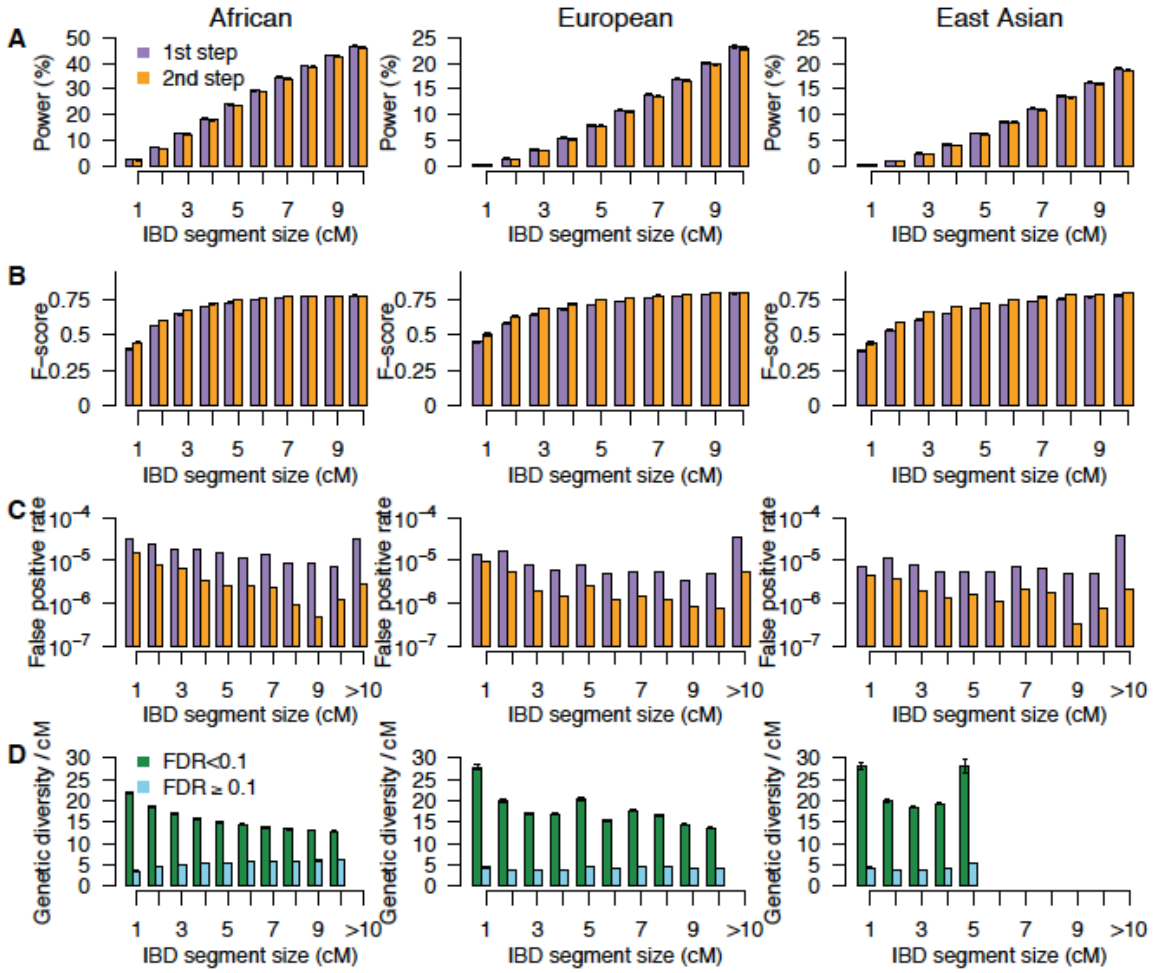

Figure S4 Performance of the exome-based IBD detection method ExIBD. (**A**) The average locus specific power (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM after the 1st step (identification) and 2nd steps (refinement), respectively. (**B**) The average locus specific F-score measuring the tradeoff between precision and recall (with 95% confidence interval) after the 1st (identification) and 2nd steps (refinement), respectively. (**C**) The false positive rate to detect IBD segments with different size intervals (e.g. 0.5~1.5 cM, 1.5~ 2.5 cM, …, 9.5~10.5 cM, and ≥10.5 cM) after the 1st step (identification) and the 2nd steps (refinement), respectively. (**D**) Genomic regions that can be used to detect IBD segments in exome sequencing data (FDR < 0.1) have significantly higher exonic genetic diversity than those cannot (FDR ≥ 0.1).

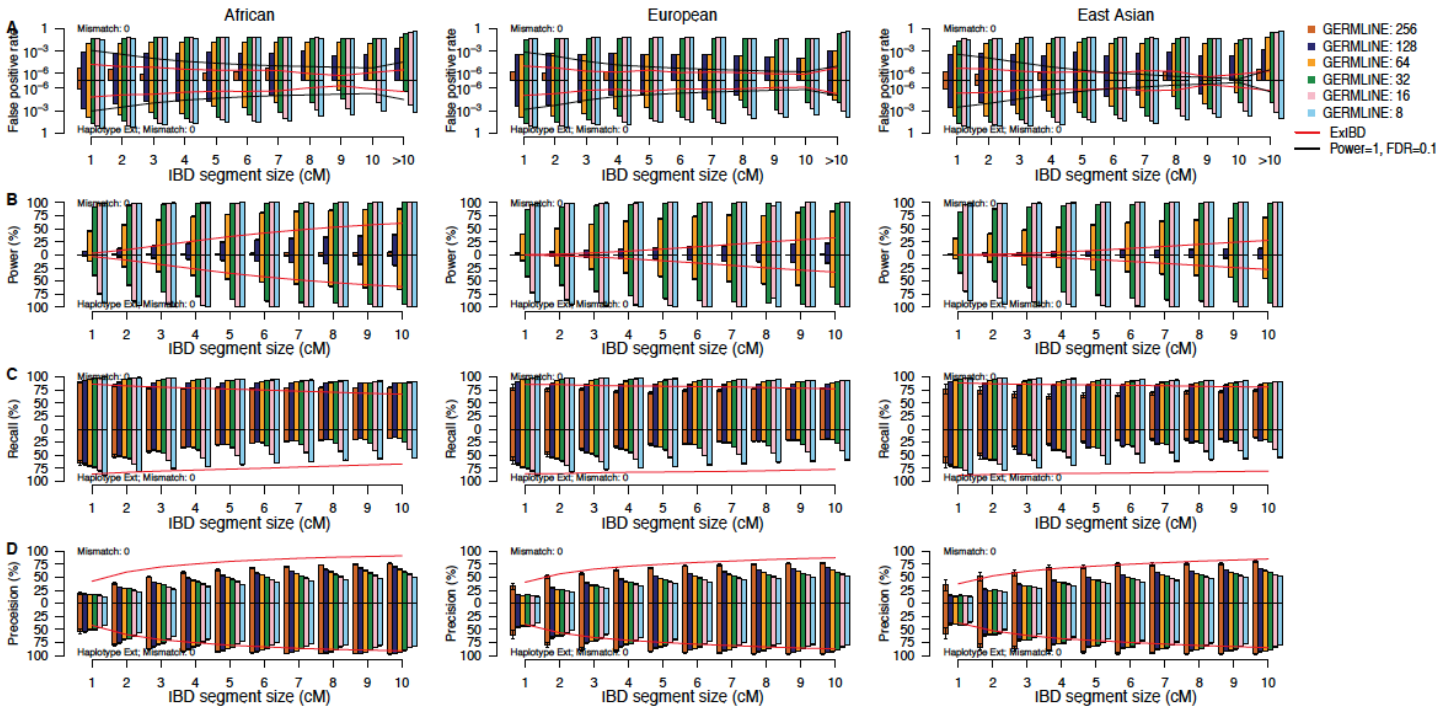

Figure S5 Performance of GERMLINE to detect IBD segments in exome sequencing data. GRMLINE ran by setting the size of slice from 256 to 8, no mismatching allowed for a slice, and turn-on (lower) or turn-off (upper) of haplotype extension. (**A**) The false positive rate to detect IBD segments with different size intervals (e.g. 0.5~1.5 cM, 1.5~2.5 cM, …, 9.5~10.5 cM, and ≥10.5 cM), (**B**) The average locus specific power (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM, (**C**) The average locus specific recall (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM, and (**D**) The average locus specific precision (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM. For comparison, the performance of ExIBD under the default setting (*fastibdthreshold* = 10^{-10}, *ibd2nonibd* = 0.01 and *nonibd2ibd* = 0.0001) was shown in red line. A critical line, any false positive rates above which can not control FDR < 0.1 even assuming the detection power is as high as 100%, was shown in black line.

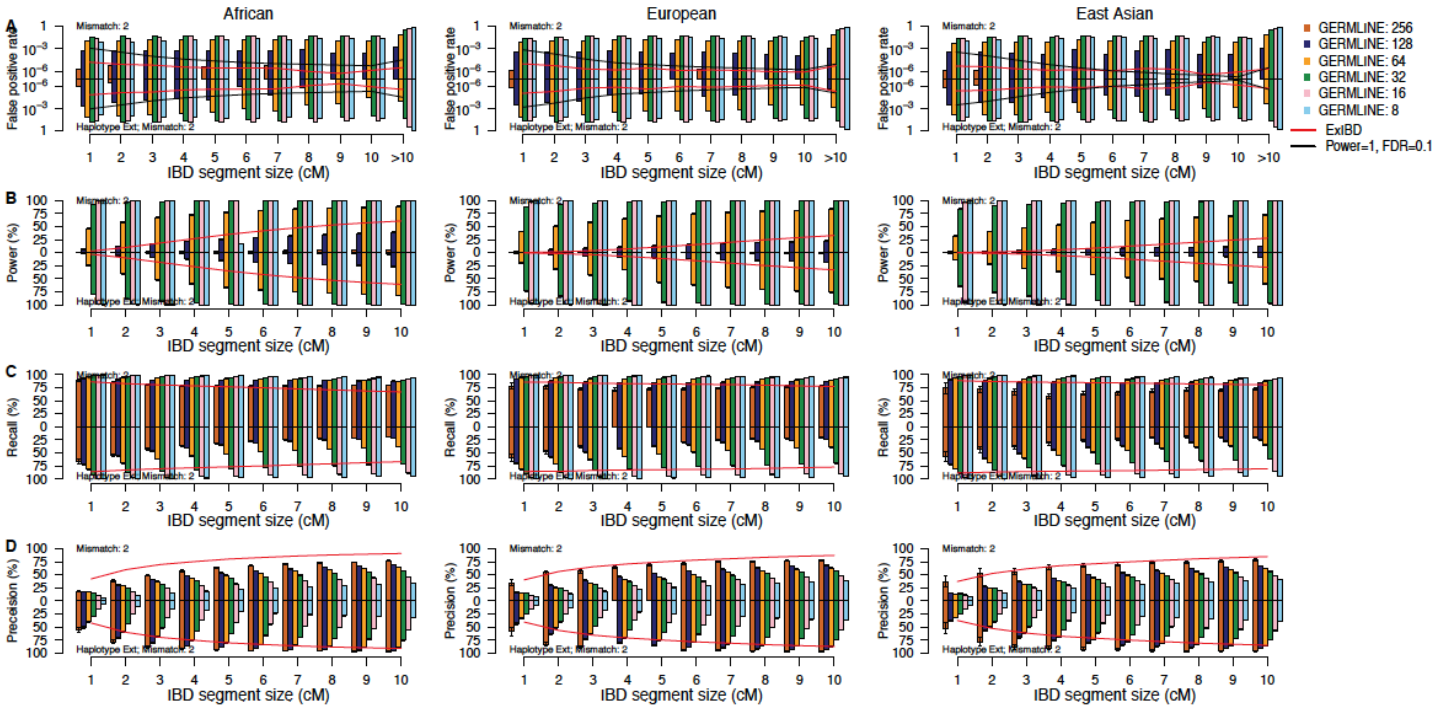

Figure S6 Performance of GERMLINE to detect IBD segments in exome sequencing data. GRMLINE ran by setting the size of slice from 256 to 8, the maximum number of mismatching homozygous markers for a slice as 2, and turn-on (lower) or turn-off (upper) of haplotype extension. (**A**) The false positive rate to detect IBD segments with different size intervals (e.g. 0.5~1.5 cM, 1.5~ 2.5 cM, …, 9.5~10.5 cM, and ≥10.5 cM), (**B**) The average locus specific power (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM, (**C**) The average locus specific recall (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM, and (**D**) The average locus specific precision (with 95% confidence interval) to detect IBD segments ranging in size from 1 cM to 10 cM. For comparison, the performance of ExIBD under the default setting (*fastibdthreshold* = 10^{-10}, *ibd2nonibd* = 0.01 and *nonibd2ibd* = 0.0001) was shown in red line. A critical line, any false positive rates above which can not control FDR < 0.1 even assuming the detection power is as high as 100%, was shown in black line.

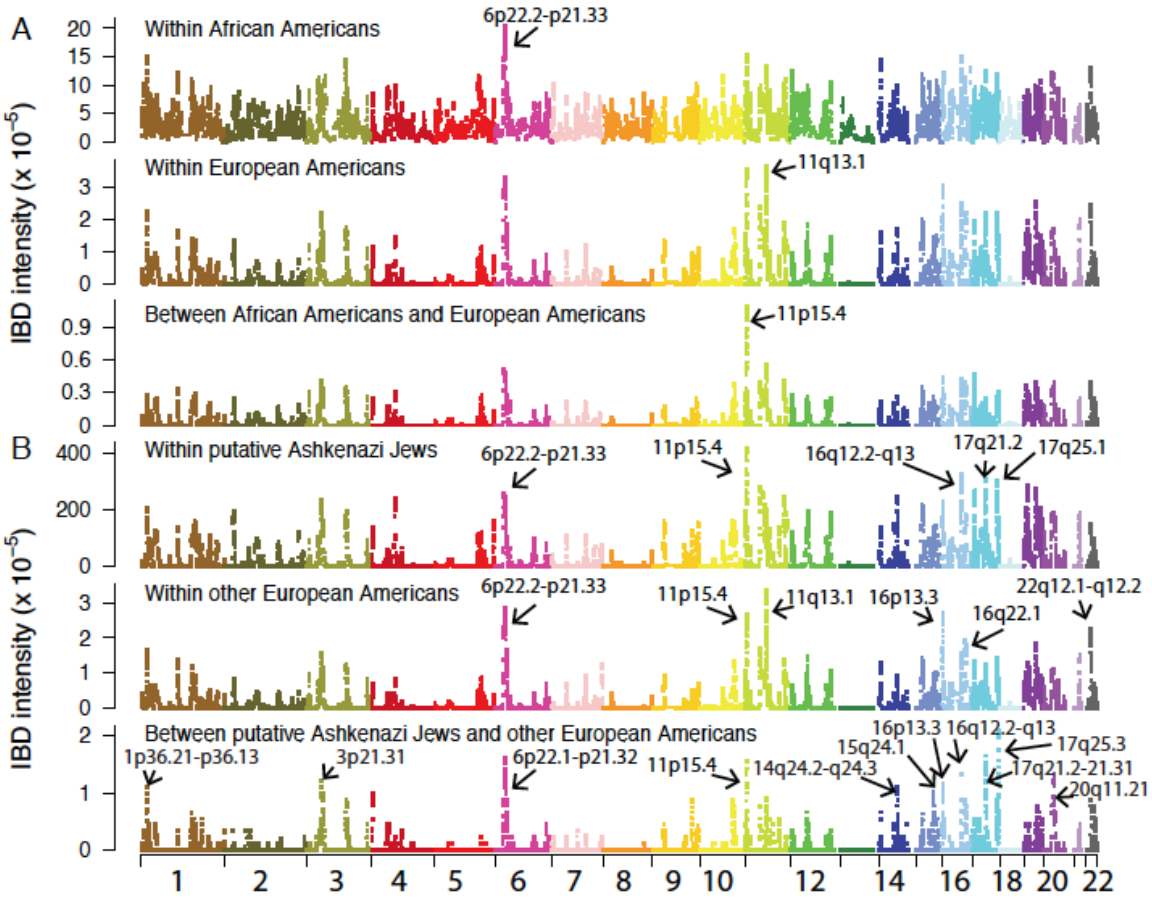

Figure S7 The physical distribution of IBD sharing within and between U.S. populations. (**A**) IBD intensity within European Americans, within African Americans, and between European Americans and African Americans, respectively; (**B**) IBD intensity within putative Ashkenazi Jews, within other European Americans, and between putative Ashkenazi Jews and other European Americans. Regions with significantly high degree of IBD sharing were highlighted by the arrows (Z-test; $P < 10^{-8}$ after adjusting for exonic genetic diversity by linear regression).

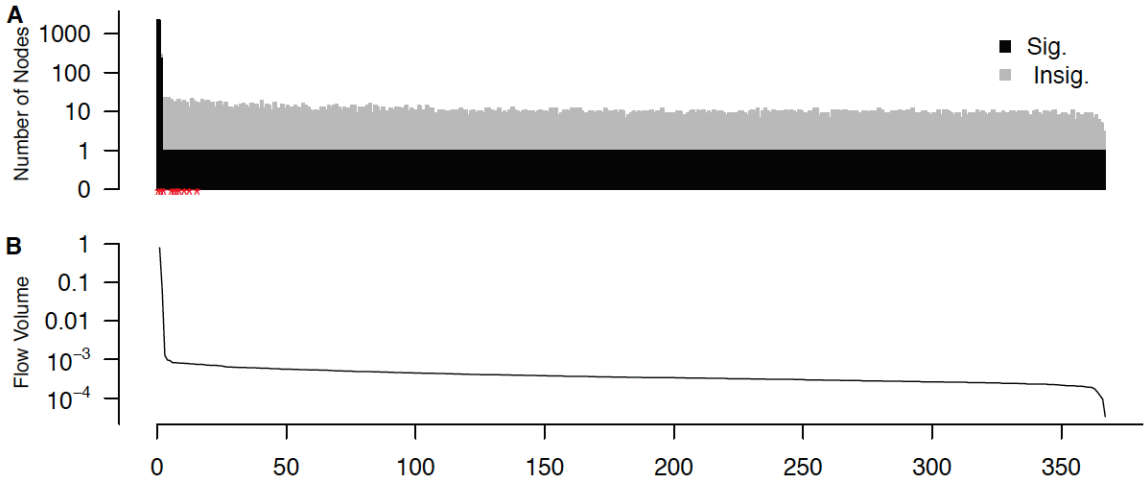

Figure S8 Characteristics of the 367 clusters identified in the IBD network by conf-infomap. (**A**) The number of nodes, including significantly associated nodes and insignificantly associated nodes, for each cluster. Ten significant clusters were highlighted in red. (**B**) The flow volume (i.e., the fraction of time that a random walker spends within a cluster) for each cluster.

Table S1. The rate of true IBD segments with different segment size interval ($\hat{T}$) for three major continental populations. The rate of true IBD segments with different size interval is determined by the IBD discovery rate ($\hat{D}$), the false positive rate ($\hat{F}$) and the power ($\hat{P}$). The IBD discovery rate was estimated based on HapMap Phase 3 genotyping data. The power and false positive rate was evaluated by simulations on HapMap Phase 3 genotyping data. We also compared the segment overlap between the detected and artificial IBD segments as measured by precision and recall. All the estimates were based on the combination of ten independent runs of fastIBD.

IBD size interval (cM)	$\hat{T}$	$\hat{D}$	$\hat{F}$	$\hat{P}$	Recall	Precision
African						
[0.5, 1.5)	0.0104	7.32×10^{-3}	5.00×10^{-4}	0.65	0.88	0.89
[1.5, 2.5)	3.09×10^{-3}	2.84×10^{-3}	1.21×10^{-5}	0.92	0.79	0.94
[2.5, 3.5)	8.57×10^{-4}	8.37×10^{-4}	3.08×10^{-6}	0.97	0.72	0.96
[3.5, 4.5)	3.55×10^{-4}	3.51×10^{-4}	2.50×10^{-6}	0.98	0.66	0.97
[4.5, 5.5)	2.02×10^{-4}	2.01×10^{-4}	1.41×10^{-6}	0.99	0.62	0.98
[5.5, 6.5)	1.21×10^{-4}	1.21×10^{-4}	4.65×10^{-7}	0.99	0.60	0.98
[6.5, 7.5)	8.74×10^{-5}	8.73×10^{-5}	4.93×10^{-7}	0.99	0.55	0.98
[7.5, 8.5)	7.11×10^{-5}	7.05×10^{-5}	0	0.99	0.53	0.99
[8.5, 9.5)	5.64×10^{-5}	5.58×10^{-5}	0	0.99	0.50	0.99
[9.5, 10.5)	4.66×10^{-5}	4.63×10^{-5}	0	0.99	0.49	0.99
≥10.5	$2.96\times10^{-4*}$	2.94×10^{-4}	0	-	-	-
European						
[0.5, 1.5)	6.30×10^{-3}	1.68×10^{-3}	2.12×10^{-4}	0.23	0.94	0.82
[1.5, 2.5)	1.77×10^{-3}	1.25×10^{-3}	2.63×10^{-5}	0.69	0.88	0.91
[2.5, 3.5)	3.58×10^{-4}	3.15×10^{-4}	3.37×10^{-6}	0.87	0.81	0.94
[3.5, 4.5)	1.13×10^{-4}	1.07×10^{-4}	8.99×10^{-7}	0.94	0.76	0.95
[4.5, 5.5)	6.74×10^{-5}	6.65×10^{-5}	1.72×10^{-6}	0.96	0.73	0.96
[5.5, 6.5)	4.17×10^{-5}	4.07×10^{-5}	0	0.97	0.68	0.97
[6.5, 7.5)	2.97×10^{-5}	2.91×10^{-5}	0	0.98	0.65	0.97
[7.5, 8.5)	2.29×10^{-5}	2.25×10^{-5}	0	0.98	0.61	0.98
[8.5, 9.5)	1.68×10^{-5}	1.65×10^{-5}	0	0.98	0.59	0.98
[9.5, 10.5)	1.42×10^{-5}	1.40×10^{-5}	0	0.98	0.56	0.98
≥10.5	$8.00\times10^{-5*}$	7.88×10^{-5}	0	-	-	-
East Asian						
[0.5, 1.5)	3.16×10^{-3}	7.22×10^{-4}	1.35×10^{-4}	0.18	0.94	0.79
[1.5, 2.5)	1.00×10^{-3}	6.44×10^{-4}	3.59×10^{-5}	0.61	0.90	0.88
[2.5, 3.5)	2.46×10^{-4}	2.10×10^{-4}	4.96×10^{-6}	0.83	0.84	0.92
[3.5, 4.5)	6.41×10^{-5}	6.03×10^{-5}	8.71×10^{-7}	0.93	0.79	0.94
[4.5, 5.5)	2.30×10^{-5}	2.19×10^{-5}	0	0.95	0.75	0.95

[5.5, 6.5)	9.69×10^{-6}	9.47×10^{-6}	0	0.98	0.71	0.96
[6.5, 7.5)	6.08×10^{-6}	5.95×10^{-6}	0	0.98	0.67	0.97
[7.5, 8.5)	3.39×10^{-6}	3.34×10^{-6}	0	0.98	0.64	0.97
[8.5, 9.5)	2.53×10^{-6}	2.49×10^{-6}	0	0.98	0.62	0.97
[9.5, 10.5)	1.51×10^{-6}	1.48×10^{-6}	0	0.98	0.60	0.98
≥10.5	$2.66 \times 10^{-5*}$	2.62×10^{-5}	0	-	-	-

* We assumed the power to detect the IBD segments with size no less than 10.5 cM was same to the power to detect the IBD segments of 10 cM.

Table S2. The locus specific FDR to detect IBD segments with different size intervals in exome sequencing data for African, European and East Asian populations, respectively (shown in a separate table).

Movie S1 The change of population structure in 6,515 U.S. individuals by a dynamic IBD network. The dynamic network illustrated the patterns of IBD sharing with different size categories (i.e., [0.5, 1.5), [1.5, 2.5), [2.5, 3.5), [3.5, 4.5), and ≥ 4.5 cM). African Americans, putative Ashkenazi Jews, and other European Americans were highlighted in blue, in brown, and in orange, respectively (shown in a separate file).