Linkage Analysis without Defined Pedigrees.

Supplementary Material.

Aaron G. Day-Williams[1,a], John Blangero[2], Thomas D. Dyer[2],
Kenneth Lange[1,3], Eric M. Sobel[1,*]

[1]Department of Human Genetics, David Geffen School of Medicine,
University of California Los Angeles, Los Angeles, CA 90095-7088
[2]Department of Genetics, Southwest Foundation for Biomedical Research,
San Antonio, TX 78245-0549
[3]Department of Biomathematics, David Geffen School of Medicine,
University of California Los Angeles, Los Angeles, CA 90095-1766
[a]Current address: Wellcome Trust Sanger Institute,
Hinxton, Cambridge, UK CB10 1HH
[*]Corresponding author: esobel@ucla.edu & 1.310.825.1111

January 24, 2012

1 Results

1.1 Global Kinship Coefficient Estimation

Summaries of the global kinship coefficient estimation results are presented in Supplementary Tables 1, 4, 7, 10, 13, 16, 19, and 22 for selected pairs of individuals with a spectrum of true global kinship coefficients. The estimated means are always close to the true values. As the density of SNPs increases, the means become more accurate and the standard deviations decrease.

To visualize if the spread of estimates overlapped between different relationships, we graphed the global kinship coefficient estimates based on 500K SNP data for selected pairs in Supplementary Figures 1, 2, 3, 4, 5, and 6, and in Figure 3 of the main article. These graphs show that even with the 500K SNP data it is difficult to distinguish between pairs with global kinship coefficients below 0.007 and unrelated pairs.

To formally test our global kinship coefficient estimates for bias, we conducted the Kolmogorov-Smirnov (KS) test comparing the distribution of the estimates against a normal distribution with mean equal to the estimated mean and standard deviation equal to the standard deviation of the estimates. The results of the KS tests are shown in Supplementary Tables 3, 6, 9, 12, 15, 18, 21, and 24. In all tests but one, we are unable to reject the null hypothesis that the values are normally distributed and the estimation procedure is unbiased. Out of the 32 tests conducted, only for our most distant relationship tested with the 200K SNP density does the KS p-value fall below 0.05.

We compared our estimates to the estimates obtained by the genome option of the software package PLINK, which is the most wide used package for estimating global IBD from genome-wide chip data[**?**]. Supplementary Tables 2, 5, 8, 11, 14, 17, 20 and 23 show the comparison of the estimates obtain from PLINK and our method. The tables show that our much simpler method-of-moments algorithm produces estimates as good as PLINK's in all instances, and in certain instances produces better results.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0	-0.024640	-0.01521	-0.0001802	0.01485	0.021090	0.00751
100K	0	-0.006876	-0.00540	-0.0001257	0.00515	0.008267	0.00264
200K	0	-0.004975	-0.00388	0.0000906	0.00406	0.006254	0.00199
500K	0	-0.004236	-0.00250	0.0000428	0.00258	0.004536	0.00127

Supplementary Table 1: Global kinship coefficient estimation results for the pair 1 and 2 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0	0	0.00514	0.0312	0.00672
Our 10K	0	0	0.00289	0.02109	0.00442
PLINK 100K	0	0	0.00840	0.0172	0.00242
Our 100K	0	0	0.00098	0.00827	0.00150
PLINK 200K	0	0.0639	0.068378	0.07265	0.00136
Our 200K	0	0	0.00083	0.00625	0.00120
PLINK 500K	0	0	0.00086	0.00535	0.00111
Our 500K	0	0	0.00053	0.00454	0.00075

Supplementary Table 2: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 1 and 2. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.02204640	0.9682626
100K	0.02459163	0.9229316
200K	0.02233089	0.9643697
500K	0.02090449	0.9812610

Supplementary Table 3: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 1 and 2 for all four SNP densities.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.25	0.1793	0.20160	0.2478	0.29400	0.3135	0.02310
100K	0.25	0.1949	0.20920	0.2485	0.28789	0.3109	0.01967
200K	0.25	0.1976	0.21280	0.2503	0.28772	0.2973	0.01873
500K	0.25	0.2139	0.22050	0.2502	0.27995	0.3007	0.01486

Supplementary Table 4: Global kinship coefficient estimation results for the pair 3 and 4 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.25	0.18295	0.24790	0.3069	0.02241
Our 10K	0.25	0.17932	0.24778	0.31354	0.02310
PLINK 100K	0.25	0.2306	0.25437	0.275	0.00686
Our 100K	0.25	0.19491	0.24855	0.31086	0.01967
PLINK 200K	0.25	0.23805	0.28398	0.3261	0.01628
Our 200K	0.25	0.19760	0.25026	0.29733	0.01873
PLINK 500K	0.25	0.21205	0.25022	0.3008	0.01482
Our 500K	0.25	0.21393	0.25022	0.30065	0.01486

Supplementary Table 5: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 3 and 4. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.03096577	0.7238213
100K	0.02664091	0.8699132
200K	0.02375163	0.9405131
500K	0.02586740	0.8922024

Supplementary Table 6: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 3 and 4 for all four SNP densities.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.08004	0.09047	0.1236	0.15664	0.1739	0.01654
100K	0.125	0.08556	0.09705	0.1238	0.15047	0.1594	0.01335
200K	0.125	0.09049	0.09948	0.1249	0.15033	0.1645	0.01271
500K	0.125	0.09986	0.10635	0.1254	0.14442	0.1519	0.00952

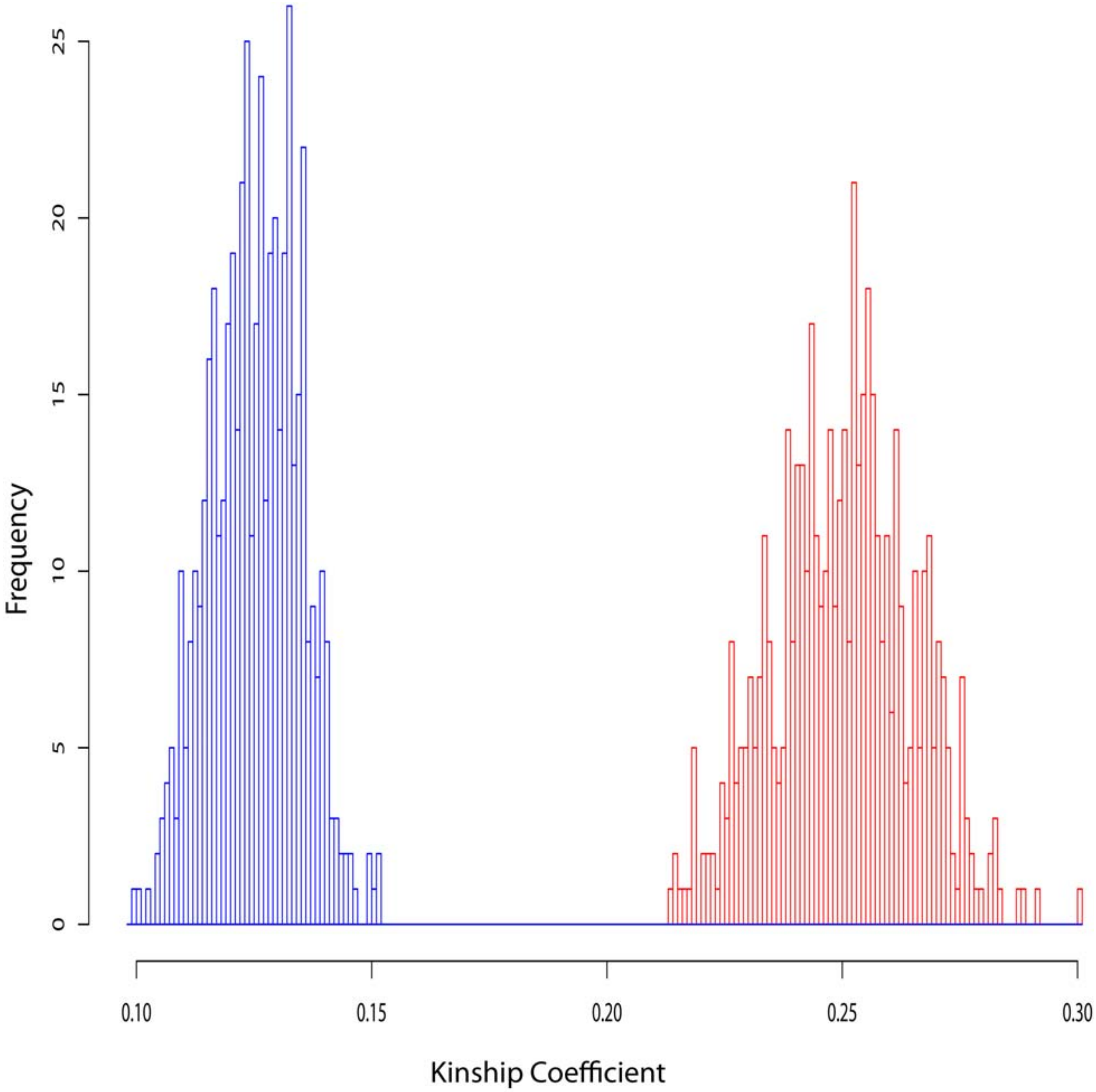
Supplementary Table 7: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.125	0.0835	0.12594	0.1734	0.01576
Our 10K	0.125	0.08004	0.12355	0.17392	0.01654
PLINK 100K	0.125	0.11515	0.13148	0.14375	0.00460
Our 100K	0.125	0.08556	0.12376	0.15936	0.01335
PLINK 200K	0.125	0.14465	0.17553	0.20865	0.01097
Our 200K	0.125	0.09049	0.12490	0.16451	0.01271
PLINK 500K	0.125	0.0995	0.12582	0.1523	0.00951
Our 500K	0.125	0.09986	0.12538	0.15192	0.00952

Supplementary Table 8: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 4 and 7. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.01885266	0.9942550
100K	0.02859544	0.8082093
200K	0.03205186	0.6832826
500K	0.04013574	0.3974943

Supplementary Table 9: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 4 and 7 for all four SNP densities.



Supplementary Figure 1: Distributions of global kinship coefficient estimates using 500K SNPs for pair 3 and 4 (red, true global kinship coefficient 0.25) versus pair 4 and 7 (blue, true global kinship coefficient 0.125).

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.0625	0.00661	0.02701	0.06214	0.09728	0.1153	0.01757
100K	0.0625	0.02548	0.03093	0.06190	0.09288	0.1109	0.01549
200K	0.0625	0.01616	0.03239	0.06230	0.09221	0.1082	0.01495
500K	0.0625	0.03160	0.03869	0.06268	0.08667	0.0974	0.01200

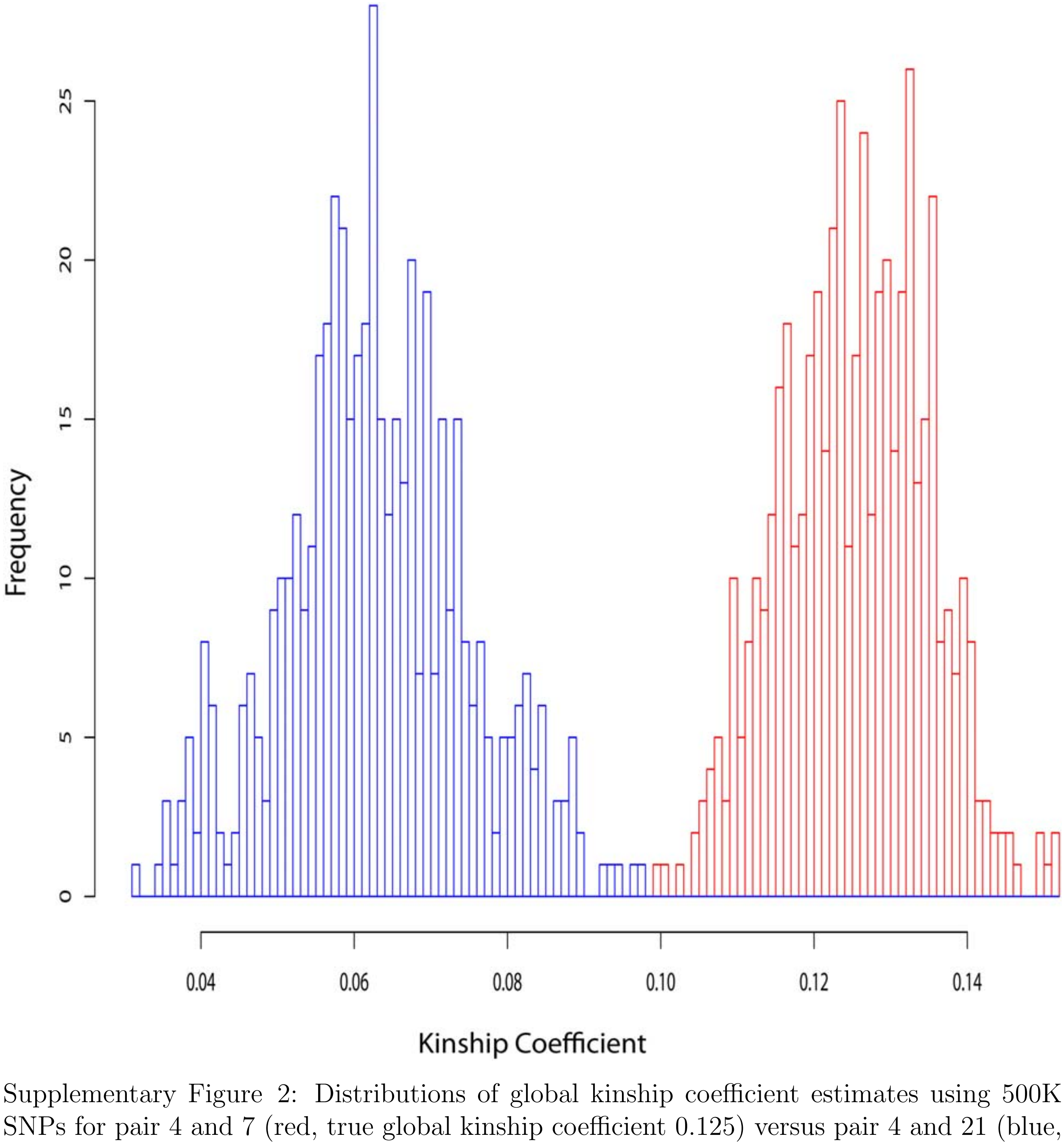
Supplementary Table 10: Global kinship coefficient estimation results for the pair 4 and 21 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.0625	0.0188	0.06548	0.1203	0.01735
Our 10K	0.0625	0.00661	0.06214	0.11526	0.01757
PLINK 100K	0.0625	0.05255	0.07015	0.08765	0.00518
Our 100K	0.0625	0.02548	0.06190	0.11094	0.01549
PLINK 200K	0.0625	0.08195	0.12133	0.16025	0.01281
Our 200K	0.0625	0.01616	0.06230	0.10822	0.01495
PLINK 500K	0.0625	0.0332	0.06327	0.09705	0.01197
Our 500K	0.0625	0.03160	0.06268	0.09739	0.01200

Supplementary Table 11: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 4 and 21. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.03375575	0.6190896
100K	0.03792196	0.4684229
200K	0.03472519	0.5828407
500K	0.04334377	0.3045118

Supplementary Table 12: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 4 and 21 for all four SNP densities.



Supplementary Figure 2: Distributions of global kinship coefficient estimates using 500K SNPs for pair 4 and 7 (red, true global kinship coefficient 0.125) versus pair 4 and 21 (blue, true global kinship coefficient 0.0625).

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.03125	-0.00035	0.00660	0.03154	0.05648	0.07819	0.01247
100K	0.03125	0.00413	0.01153	0.03113	0.05073	0.06088	0.00980
200K	0.03125	0.00983	0.01293	0.03073	0.04854	0.06285	0.00890
500K	0.03125	0.01424	0.01698	0.03127	0.04556	0.05073	0.00715

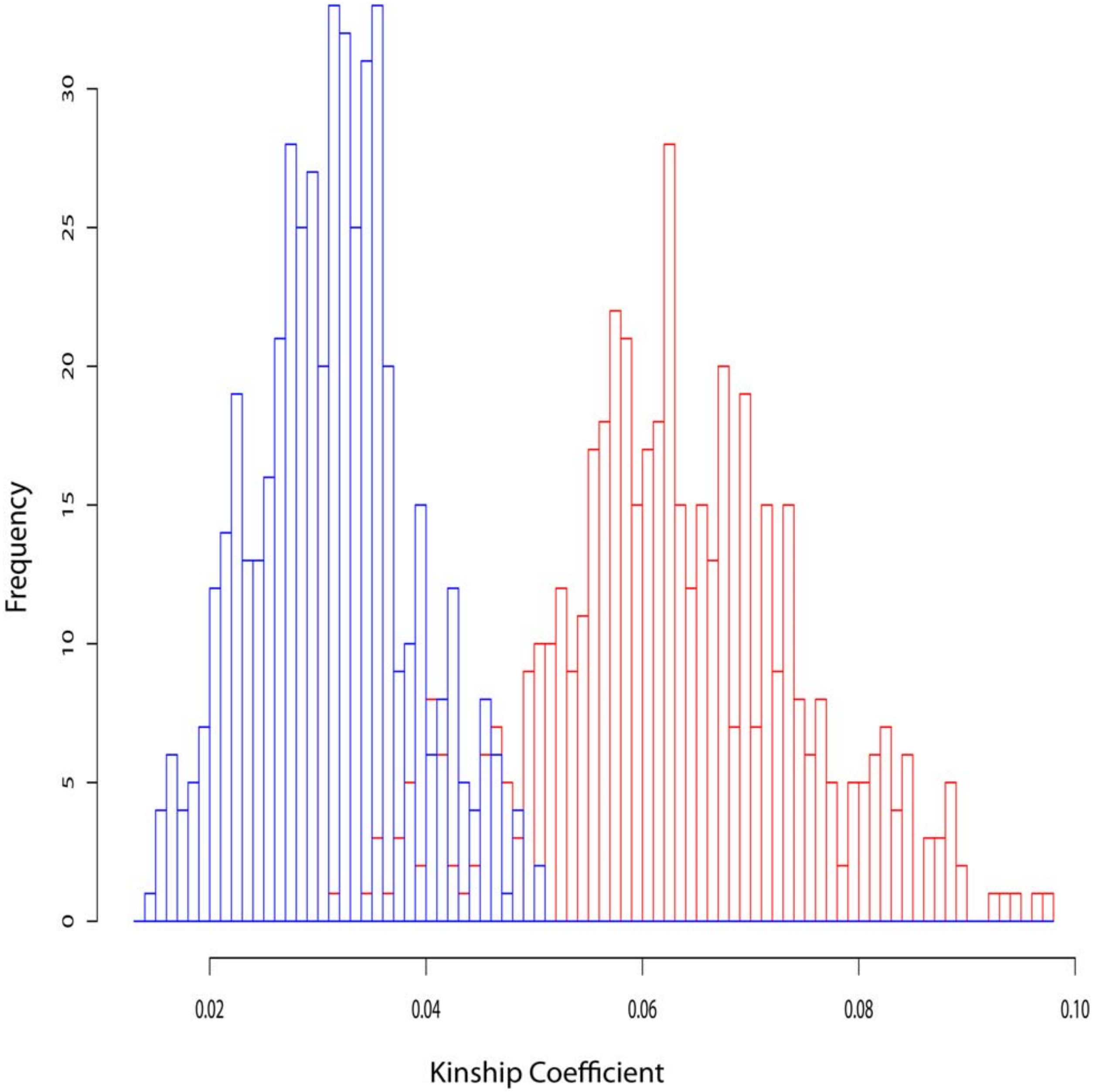
Supplementary Table 13: Global kinship coefficient estimation results for the pair 4 and 19 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.03125	0	0.03514	0.07175	0.01290
Our 10K	0.03125	0	0.03154	0.07819	0.01247
PLINK 100K	0.03125	0.02915	0.03899	0.0496	0.00367
Our 100K	0.03125	0.00413	0.03113	0.06087	0.00980
PLINK 200K	0.03125	0.0739	0.09407	0.1222	0.00767
Our 200K	0.03125	0.00983	0.03073	0.06285	0.00890
PLINK 500K	0.03125	0.01535	0.03188	0.05060	0.00709
Our 500K	0.03125	0.01424	0.03127	0.05073	0.00715

Supplementary Table 14: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 4 and 19. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.03656664	0.5157035
100K	0.04312246	0.3103269
200K	0.04747099	0.2098196
500K	0.03630494	0.5263683

Supplementary Table 15: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 4 and 19 for all four SNP densities.



Supplementary Figure 3: Distributions of global kinship coefficient estimates using 500K SNPs for pair 4 and 21 (red, true global kinship coefficient 0.0625) versus pair 4 and 19 (blue, true global kinship coefficient 0.03125).

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.015625	-0.01541	-0.00505	0.01503	0.03512	0.04567	0.01004
100K	0.015625	0.000055	0.00263	0.01559	0.02855	0.04016	0.00648
200K	0.015625	0.00165	0.00349	0.01515	0.02681	0.03488	0.00583
500K	0.015625	0.00254	0.00631	0.01579	0.02527	0.03324	0.00474

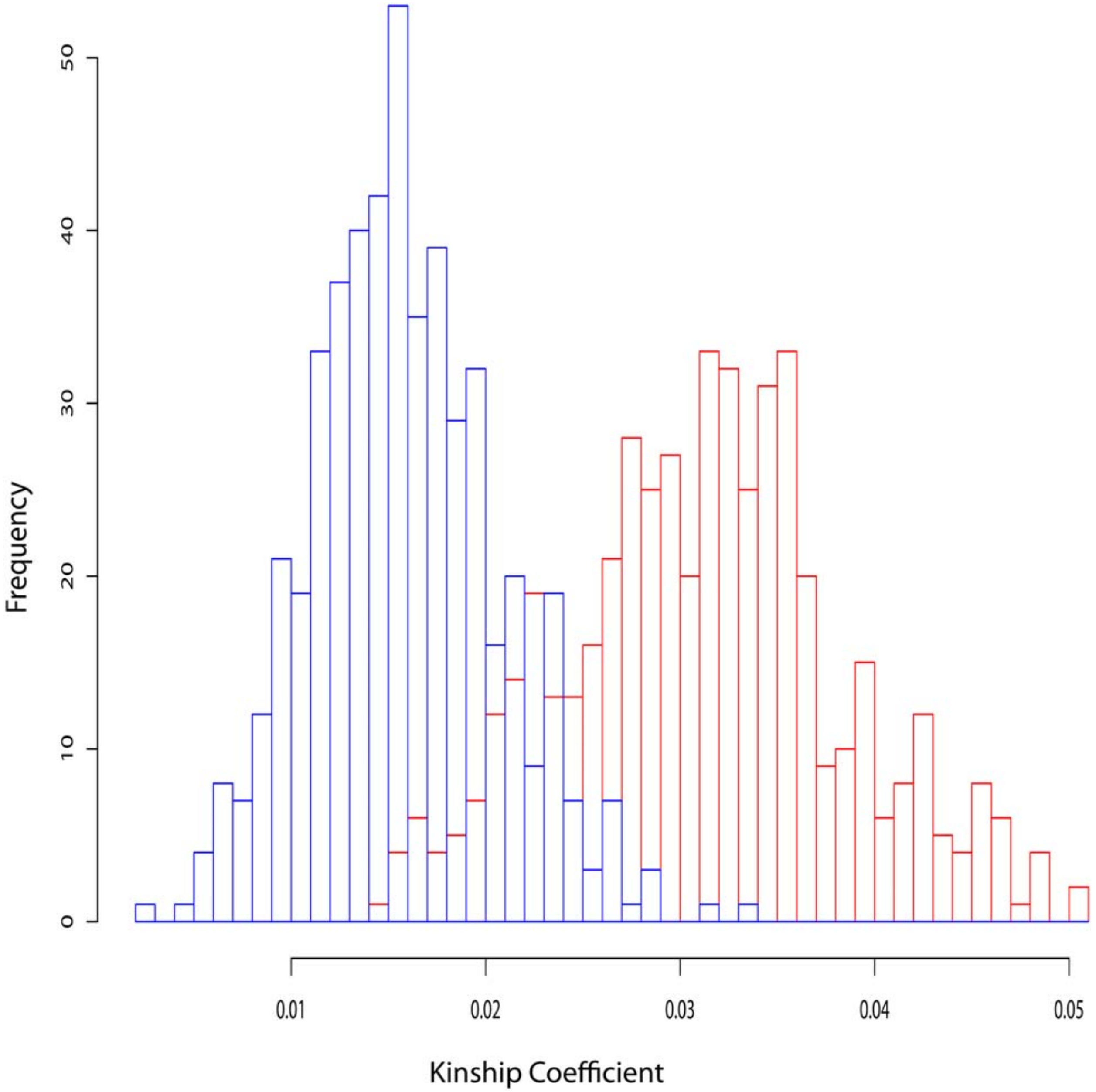
Supplementary Table 16: Global kinship coefficient estimation results for the pair 7 and 21 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.015625	0	0.01817	0.05445	0.01094
Our 10K	0.015625	0	0.01522	0.04567	0.00970
PLINK 100K	0.015625	0.01485	0.02362	0.03335	0.00305
Our 100K	0.015625	0.00005	0.01559	0.04016	0.00648
PLINK 200K	0.015625	0.06965	0.08061	0.09044	0.00491
Our 200K	0.015625	0.00165	0.01515	0.03488	0.00583
PLINK 500K	0.015625	0.00425	0.01630	0.03265	0.00474
Our 500K	0.015625	0.00254	0.01579	0.03324	0.00474

Supplementary Table 17: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 7 and 21. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.04249858	0.3271328
100K	0.04110611	0.3668298
200K	0.03798183	0.4663846
500K	0.04268127	0.3221485

Supplementary Table 18: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 7 and 21 for all four SNP densities.



Supplementary Figure 4: Distributions of global kinship coefficient estimates using 500K SNPs for pair 4 and 19 (red, true global kinship coefficient 0.03125) versus pair 7 and 21 (blue, true global kinship coefficient 0.015625).

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.0078125	-0.01725	-0.01032	0.007738	0.02580	0.03445	0.00903
100K	0.0078125	-0.00316	-0.00218	0.007441	0.01706	0.02290	0.00481
200K	0.0078125	-0.00356	-0.00119	0.007413	0.01602	0.02254	0.00430
500K	0.0078125	0.000470	0.00109	0.007839	0.01458	0.02185	0.00337

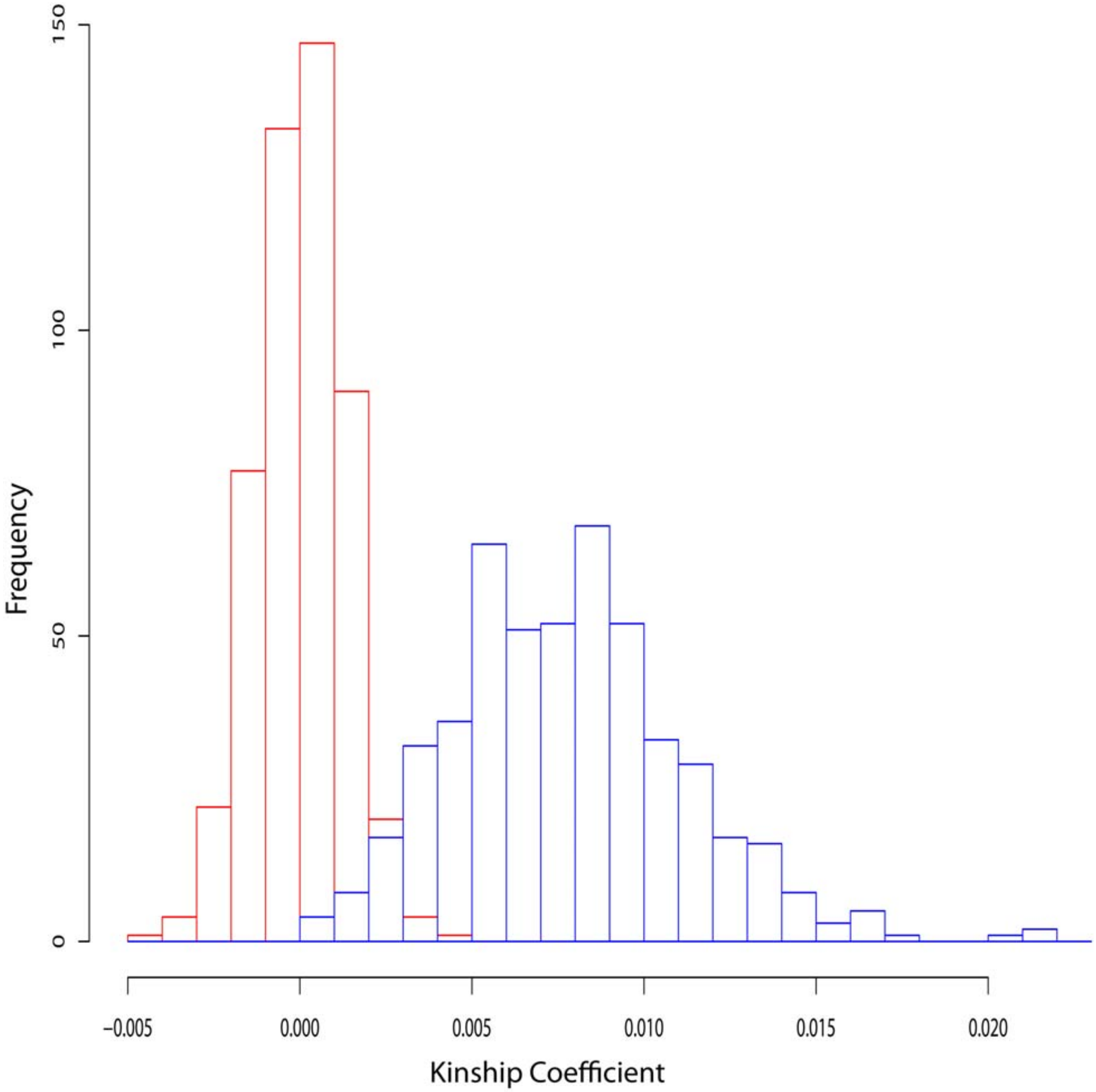
Supplementary Table 19: Global kinship coefficient estimation results for the pair 14 and 21 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.0078125	0	0.01167	0.04225	0.00924
Our 10K	0.0078125	0	0.00865	0.03445	0.00768
PLINK 100K	0.0078125	0.0084	0.0160103	0.02495	0.00273
Our 100K	0.0078125	0	0.00748	0.02290	0.00474
PLINK 200K	0.0078125	0.0669	0.07408	0.0863	0.00354
Our 200K	0.0078125	0	0.00744	0.02254	0.00426
PLINK 500K	0.0078125	0	0.008478	0.02335	0.00338
Our 500K	0.0078125	0.00047	0.007839	0.02185	0.00337

Supplementary Table 20: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 14 and 21. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.02704439	0.8579572
100K	0.05384713	0.1100817
200K	0.03973323	0.4088592
500K	0.03955475	0.4145268

Supplementary Table 21: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 14 and 21 for all four SNP densities.



Supplementary Figure 5: Distributions of global kinship coefficient estimates using 500K SNPs for unrelated pair 1 and 2 (red, true global kinship coefficient 0.0) versus distantly related pair 14 and 21 (blue, true global kinship coefficient 0.0078125).

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.003906	-0.01740	-0.01219	0.003557	0.01930	0.03047	0.00787
100K	0.003906	-0.00649	-0.00414	0.003770	0.01168	0.02326	0.00396
200K	0.003906	-0.00410	-0.00301	0.003731	0.01047	0.01785	0.00337
500K	0.003906	-0.00239	-0.00113	0.004023	0.00917	0.01542	0.00257

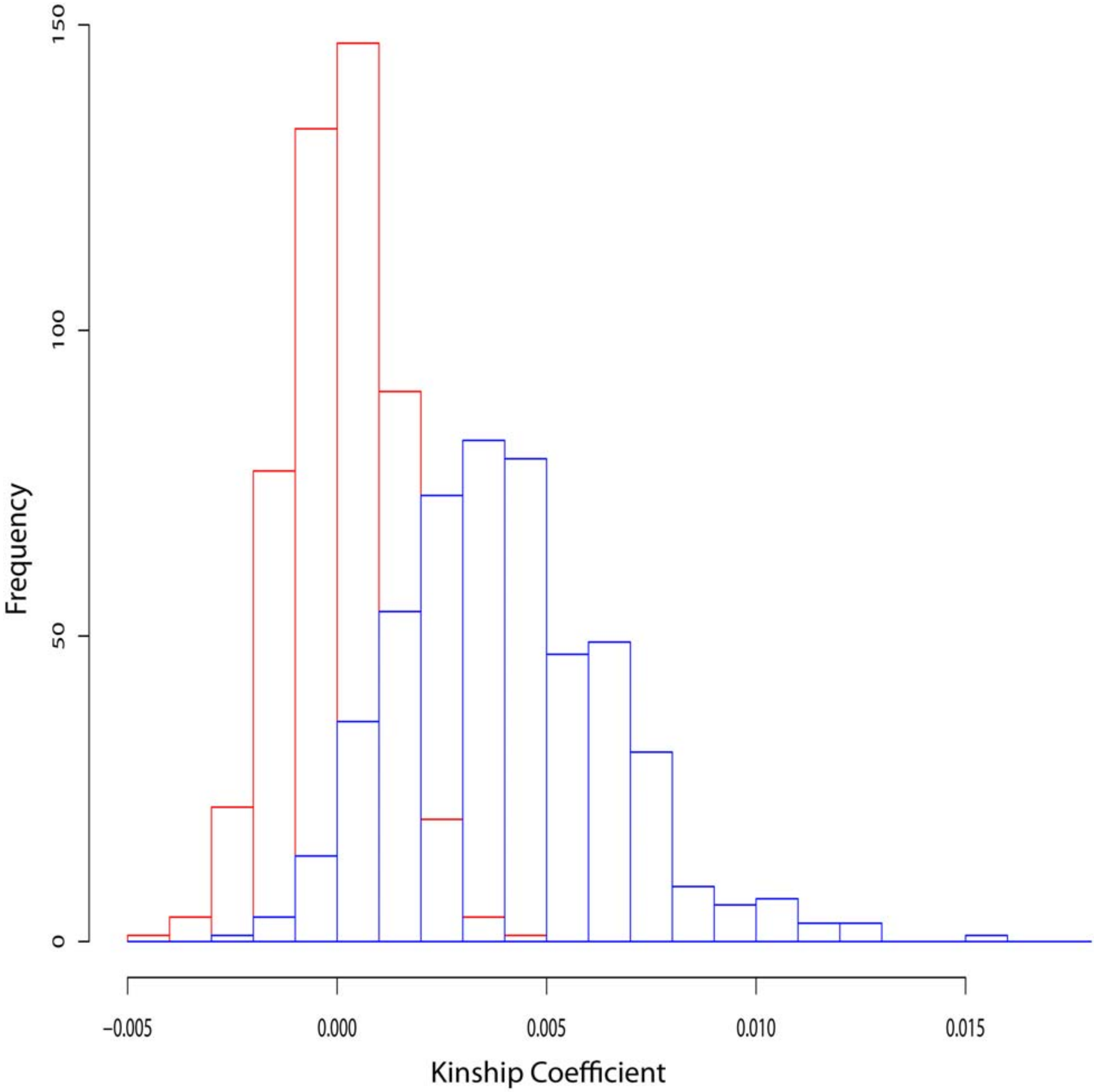
Supplementary Table 22: Global kinship coefficient estimation results for the pair 19 and 21 for all four SNP densities. (SD = Standard Deviation.)

SNPS	True	Min	Mean	Max	SD
PLINK 10K	0.003906	0	0.008226	0.0335	0.00796
Our 10K	0.003906	0	0.00514	0.03047	0.00587
PLINK 100K	0.003906	0	0.01204	0.025	0.00249
Our 100K	0.003906	0	0.00406	0.02326	0.00354
PLINK 200K	0.003906	0.06545	0.07094	0.08255	0.00260
Our 200K	0.003906	0	0.00387	0.01785	0.00316
PLINK 500K	0.003906	0	0.00460	0.0141	0.00267
Our 500K	0.003906	0	0.00405	0.01542	0.00252

Supplementary Table 23: Global kinship coefficient estimation comparision between PLINK and our estimates for the pair 19 and 21. The PLINK estimates were calculated as $\widehat{\pi}/2$. We truncated our estimates at zero to accurately compare our estimates to PLINK. (SD = Standard Deviation)

SNPs	D	P-value
10K	0.03491145	0.5759316
100K	0.03793606	0.4679426
200K	0.06698928	0.0224960
500K	0.05490507	0.0987158

Supplementary Table 24: Kolmogorov-Smirnov test of normality for global kinship coefficient estimates for the pair 19 and 21 for all four SNP densities.



Supplementary Figure 6: Distributions of global kinship coefficient estimates using 500K SNPs for unrelated pair 1 and 2 (red, true global kinship coefficient 0.0) versus distantly related pair 19 and 21 (blue, true global kinship coefficient 0.003906).

We next tested how allele frequency misspecification effects our global kinship coefficient estimates. We recomputed the estimates for the pair 4 and 7 assuming major allele frequencies that were 1% larger than their true values for either 25%, 50%, or 100% of the SNPs. The results are in Supplementary Tables 25, 26, and 27, which can be compared to Supplementary Table 7 where the correct allele frequencies were used. These tables show that a systematic over estimate of the major allele frequencies results in a decrease in the global kinship coefficient estimates. Similarly, a systematic under estimate of the major allele frequencies results in an increase in the global kinship coefficient estimates, as seen in Supplementary Tables 28, 29, and 30. In all these results, the standard deviations do not change substantially, only the means. These changes can be explained by examining equation (2) of the main article. By systematically increasing (decreasing) the major allele frequencies the second term in both the numerator and the denominator of equation (2), the sum of squares of the allele frequencies, is increased (decreased). This term represents the expected number of IBS matches at each locus. Since more (less) IBS matches are now expected, the global kinship estimate will decrease (increase).

SNPs	True	Min	−2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.07307	0.08492	0.11819	0.15145	0.16682	0.0166
100K	0.125	0.07577	0.08838	0.11538	0.14238	0.15171	0.0135
200K	0.125	0.08326	0.09202	0.11769	0.14335	0.15699	0.0128
500K	0.125	0.09157	0.09660	0.11676	0.13690	0.14299	0.0101

Supplementary Table 25: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% greater than the true values at 25% of the SNPs.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.06583	0.07941	0.11291	0.14639	0.16256	0.0167
100K	0.125	0.06730	0.07975	0.10705	0.13433	0.14345	0.0136
200K	0.125	0.07336	0.08132	0.10860	0.13588	0.14757	0.0136
500K	0.125	0.08276	0.08951	0.10896	0.12840	0.13583	0.0097

Supplementary Table 26: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% greater than the true values at 50% of the SNPs.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.05546	0.06875	0.10234	0.13592	0.15205	0.0168
100K	0.125	0.05016	0.06269	0.09038	0.11806	0.12668	0.0138
200K	0.125	0.05290	0.06217	0.08871	0.11525	0.12973	0.0133
500K	0.125	0.06579	0.07273	0.09246	0.11218	0.11968	0.0099

Supplementary Table 27: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% greater than the true values at 100% of the SNPs.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.08557	0.09628	0.12910	0.16192	0.17917	0.0164
100K	0.125	0.09296	0.10562	0.13221	0.15879	0.16888	0.0133
200K	0.125	0.10158	0.11003	0.13517	0.16031	0.17376	0.0126
500K	0.125	0.10885	0.11561	0.13398	0.15234	0.16069	0.0092

Supplementary Table 28: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% less than the true values at 25% of the SNPs.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.09111	0.10207	0.13455	0.16703	0.18386	0.0162
100K	0.125	0.10183	0.11429	0.14062	0.16694	0.17743	0.0134
200K	0.125	0.10987	0.11811	0.14304	0.16796	0.18073	0.0125
500K	0.125	0.11749	0.12370	0.14239	0.16107	0.16882	0.0093

Supplementary Table 29: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% less than the true values at 50% of the SNPs.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.10035	0.11351	0.14562	0.17772	0.19480	0.0161
100K	0.125	0.11842	0.13170	0.15752	0.18334	0.19294	0.0129
200K	0.125	0.12660	0.13430	0.15878	0.18326	0.19586	0.0122
500K	0.125	0.13515	0.14116	0.15946	0.17776	0.18546	0.0092

Supplementary Table 30: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies used in the analysis 1% less than the true values at 100% of the SNPs.

A more realistic error model is that the estimated major allele frequencies are taken from a normal distribution with mean equal to the true value and some small standard deviation. We recomputed our global kinship coefficient estimates under this error model with standard deviations for each locus set to either 2.5%, 5%, 7.5%, or 10% of the major allele frequency. The results are in Supplementary Tables 31, 32, 33, and 34 respectively, which can again be compared to Supplementary Table 7 where the correct allele frequencies were used. Unfortunately, even this error model did not result in an even distribution of high and low allele misspecifications. The SNP chips are weighted toward SNPs with high major allele frequencies and since even after applying the error model these frequencies need to be below 1.0, the analyzed distribution was skewed relative to the true frequencies. This accounts for the shift towards underestimating the relatedness of the individuals in these results. The 10K and 500K chips appear to be more robust against allele misspecifications than the 100K and 200K chips. This might be related to the criteria employed in selecting the SNPs to include on the chips.

Clearly considerable effort should be employed to ensure good, unbiased estimation of allele frequencies when estimating global kinship coefficients. This is common sense, since allele frequencies play such an important role in the estimation procedure.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.07805	0.09096	0.12367	0.15637	0.17435	0.0163
100K	0.125	0.07925	0.09238	0.11934	0.14628	0.15470	0.0135
200K	0.125	0.06907	0.07950	0.10539	0.13127	0.14538	0.0129
500K	0.125	0.09810	0.10513	0.12412	0.14313	0.15066	0.0095

Supplementary Table 31: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies drawn from a normal distribution with mean = true frequency, and standard deviation = 0.025 × mean.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.07151	0.08962	0.12245	0.15527	0.17236	0.0164
100K	0.125	0.06598	0.08119	0.10846	0.13572	0.14514	0.0136
200K	0.125	0.04417	0.05395	0.08013	0.10631	0.11409	0.0131
500K	0.125	0.08877	0.09485	0.11412	0.13284	0.14232	0.0096

Supplementary Table 32: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies drawn from a normal distribution with mean = true frequency, and standard deviation = 0.05 × mean.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.07009	0.08464	0.11745	0.15025	0.16682	0.0164
100K	0.125	0.05182	0.06487	0.09264	0..1204	0.12901	0.0139
200K	0.125	0.01709	0.02645	0.05385	0.08123	0.09608	0.0138
500K	0.125	0.07242	0.07937	0.09881	0.11825	0.12637	0.0097

Supplementary Table 33: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies drawn from a normal distribution with mean = true frequency, and standard deviation = 0.075 × mean.

SNPs	True	Min	–2 SD	Mean	+2 SD	Max	SD
10K	0.125	0.06697	0.08416	0.11711	0.15005	0.16554	0.0165
100K	0.125	0.03205	0.04576	0.07400	0.10222	0.11135	0.0141
200K	0.125	-0.01301	-0.00294	0.02523	0.05339	0.06977	0.0141
500K	0.125	0.05347	0.06081	0.08077	0.10073	0.10837	0.0100

Supplementary Table 34: Global kinship coefficient estimation results for the pair 4 and 7 for all four SNP densities, and with the major allele frequencies drawn from a normal distribution with mean = true frequency, and standard deviation = 0.1 × mean.

1.2 Local Kinship Coefficient Estimation

The penalized optimization technique we propose for estimating the conditional kinship coefficient for each SNP on a chip has two penalty terms, λ_1 and λ_2. The λ_1 parameter is to penalize the point estimate at a SNP for being different than the chromosome specific theoretical kinship coefficient. The λ_2 parameter is to penalize neighboring SNPs for belonging to different conditional kinship coefficient sets. The first step for our method is to find the optimal λ_1, λ_2 combination that minimizes the error between the true conditional kinship coefficient and our estimated conditional kinship coefficient in a variety of IBD sharing configurations and relationships. We chose four relationships from the pedigree structure (Figure 1 in main manuscript) that represented a wide range of relationships and would exhibit complex patterns of IBD sharing along the chromosomes. For each of these relationships we analyzed the first 100 replicates of the 500 replicates analyzed for the theoretical kinship estimation above, and performed a grid search of all combinations of λ_1 and λ_2 for λ_1 from 0-2 in increments of 0.2 and for λ_2 from 0 - 110 in increments of 5. For each replicate, for each SNP, for each pair of λ_1 and λ_2, we compared our estimated conditional coefficient to the true conditional coefficient. The true conditional coefficient was determined by performing gene-dropping in the program Mendel with its option to uniquely label all founder alleles so IBD sharing can be calculated exactly. For each replicate we calculated the average absolute difference (aad) for a chromosome as defined by equation 1, where m is the number of loci analyzed, z_i is the calculated estimate and t_i is the true value. We wanted to find the λ_1, λ_2 combination that minimized the aad over the 100 replicates. We chose to use chromosome 21 of the 200K chip to perform the grid search to find the optimal λ_1, λ_2 combination. The 4 relationships we tested were individual pairs 3-4, 4 -7, 4-19 and 19-21. The hope was to find a single λ_1, λ_2 combination that works well across all relationships and all chromosomes. The optimal λ_1 and λ_2 penalties discovered by the optimization search are seen in Supplementary Table 35 for all relationships. The optimal λ_1, λ_2 combination for all pairs are all very close. In order to simplify the calculation and increase the generality of the technique we would like a single λ_1, λ_2 combination.

$$aad = \frac{\sum_{i=1}^{m} |t_i - z_i|}{m} \tag{1}$$

Individual Pair	Theoretical Kinship	Lambd1	Lamda2
3-4	0.25	0	90
4-7	0.125	0.2	110
4-19	0.03125	0	100
19-21	0.003906	0.6	90

Supplementary Table 35: Optimal Lambda Penalty Values for 4 relationship types for Chr 21 on 200K chip

We wish to show that for all pairs of individuals, all SNP densities, and all chromosomes, using $\lambda_1 = 0$ and $\lambda_2 = 100$ (the optimal penalties for individual pair 4-19) works nearly as well as using the optimal λ_1 and λ_2. If so, then the clear benefit of using a fixed set of penalties for all data sets outweighs the small improvement obtained by using individualized penalties. We first compare the distribution of *aad* for these two sets of λ values, the optimal λ_1, λ_2 and $\lambda_1 = 0$, $\lambda_2 = 100$, for relationship pairs 3-4, 4-7 and 19-21 for Chr 21 on the 200K chip. The results are in Supplementary Tables 36, 37, and 38.

λ_1	λ_2	Min	–2 SD	Median	Mean	+2 SD	Max
0	90	0	-0.006881624	0.005754277	0.00761213	0.02210589	0.0403577
0	100	0	-0.007394278	0.005987558	0.007966096	0.02332647	0.0403577

Supplementary Table 36: Distributions for 500 simulated pedigrees of the *aad* between the optimal λ_1, λ_2 and general λ_1, λ_2 using pair 3 and 4 (true global kinship coefficient 0.25), chromosome 21, and the 200K SNP density.

λ_1	λ_2	Min	–2 SD	Median	Mean	+2 SD	Max
0.2	110	0	-0.005072853	0.002021773	0.003388336	0.01184952	0.03693624
0	100	0	-0.005494397	0.001944012	0.003396734	0.01228786	0.04066874

Supplementary Table 37: Distributions for 500 simulated pedigrees of the *aad* between the optimal λ_1, λ_2 and general λ_1, λ_2 using pair 4 and 7 (true global kinship coefficient 0.125), chromosome 21, and the 200K SNP density.

λ_1	λ_2	Min	–2 SD	Median	Mean	+2 SD	Max
0.6	90	0	-0.005460422	0	0.0006533437	0.00676711	0.03569207
0	100	0	-0.004834165	0	0.0006695179	0.0061732	0.03273717

Supplementary Table 38: Distributions for 500 simulated pedigrees of the *aad* between the optimal λ_1, λ_2 and general λ_1, λ_2 using pair 19 and 21 (true global kinship coefficient 0.003906), chromosome 21, and the 200K SNP density.

So far, we have seen that fixing $\lambda_1 = 0$ and $\lambda_2 = 100$ allows the local kinship coefficient estimation procedure to work well for pairs of individuals with differing degrees of relationship, using the 200K SNP density, on chromosome 21. We next determine how well this static choice of λ_1 and λ_2 works for other SNP densities. We computed the distributions over 500 simulated pedigrees of the average absolute differences for each of the four SNP densities and for several relative pairs. The results are in Supplementary Tables 39, 40, 41, and 42. In all cases the accuracy of the local kinship coefficient estimation increases with increasing SNP density. It appears that 10K SNP density is not sufficient for good local kinship coefficient estimation, while it was for global estimates. This is not a surprising result. More importantly, with 500K SNP density, the local estimates are very good.

SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.0536713	0.05151515	0.06469091	0.1830531	0.2621212
100K	0	-0.01104559	0.007122321	0.01020413	0.03145384	0.07239937
200K	0	-0.007394278	0.005987558	0.007966096	0.02332647	0.0403577
500K	0	-0.003425286	0.003202436	0.004283354	0.01199199	0.03016940

Supplementary Table 39: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 3 and 4 (true global kinship coefficient 0.25) and chromosome 21 across chip types.

To visualize how good these local kinship coefficient estimates are, we selected the simulated pedigree at the mean of the distributions of the average absolute differences for the 500K SNP density and graphed the true and estimated local kinship coefficients for all 7143 SNPs listed on chromosome 21. The results are in Supplementary Figures 7, 8, 9, and 10. In each of these results at most 58 of the 7143 SNPs listed on chromosome 21 were assigned an

SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.06276659	0.01666667	0.03853333	0.1398333	0.2772727
100K	0	-0.007246174	0.002483011	0.004853372	0.01695292	0.0343701
200K	0	-0.005494397	0.001944012	0.003396734	0.01228786	0.04066874
500K	0	-0.002654649	0.001259975	0.002019950	0.006694548	0.01462971

Supplementary Table 40: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 4 and 7 (true global kinship coefficient 0.125) and chromosome 21 across chip types.

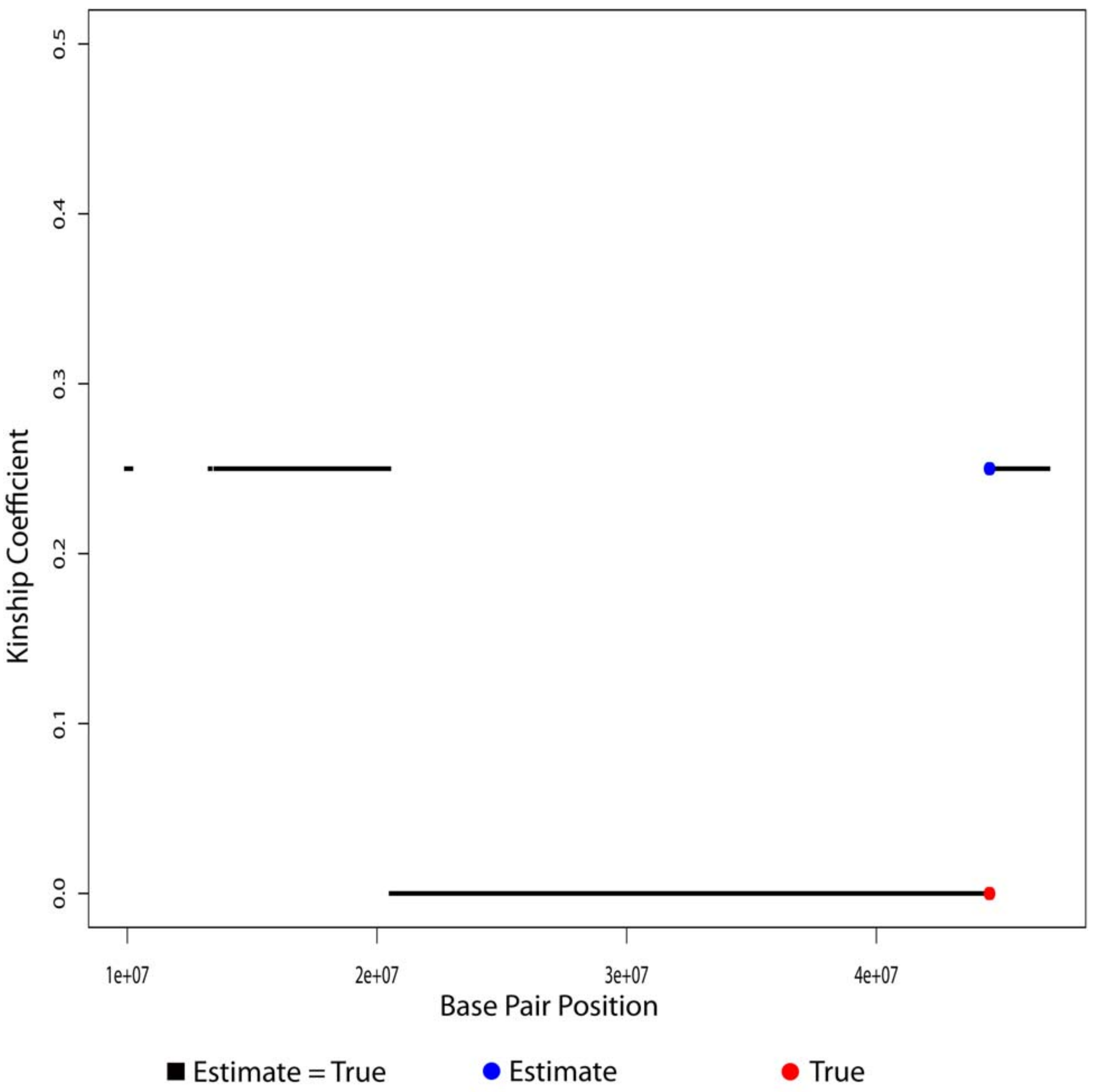
SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.04731742	0	0.01327273	0.07386288	0.25
100K	0	-0.007073608	0	0.002178254	0.01143012	0.03345531
200K	0	-0.005261046	0	0.001646967	0.008554981	0.02566096
500K	0	-0.002934514	0	0.0009808204	0.004896155	0.01431471

Supplementary Table 41: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 4 and 19 (true global kinship coefficient 0.03125) and chromosome 21 across chip types.

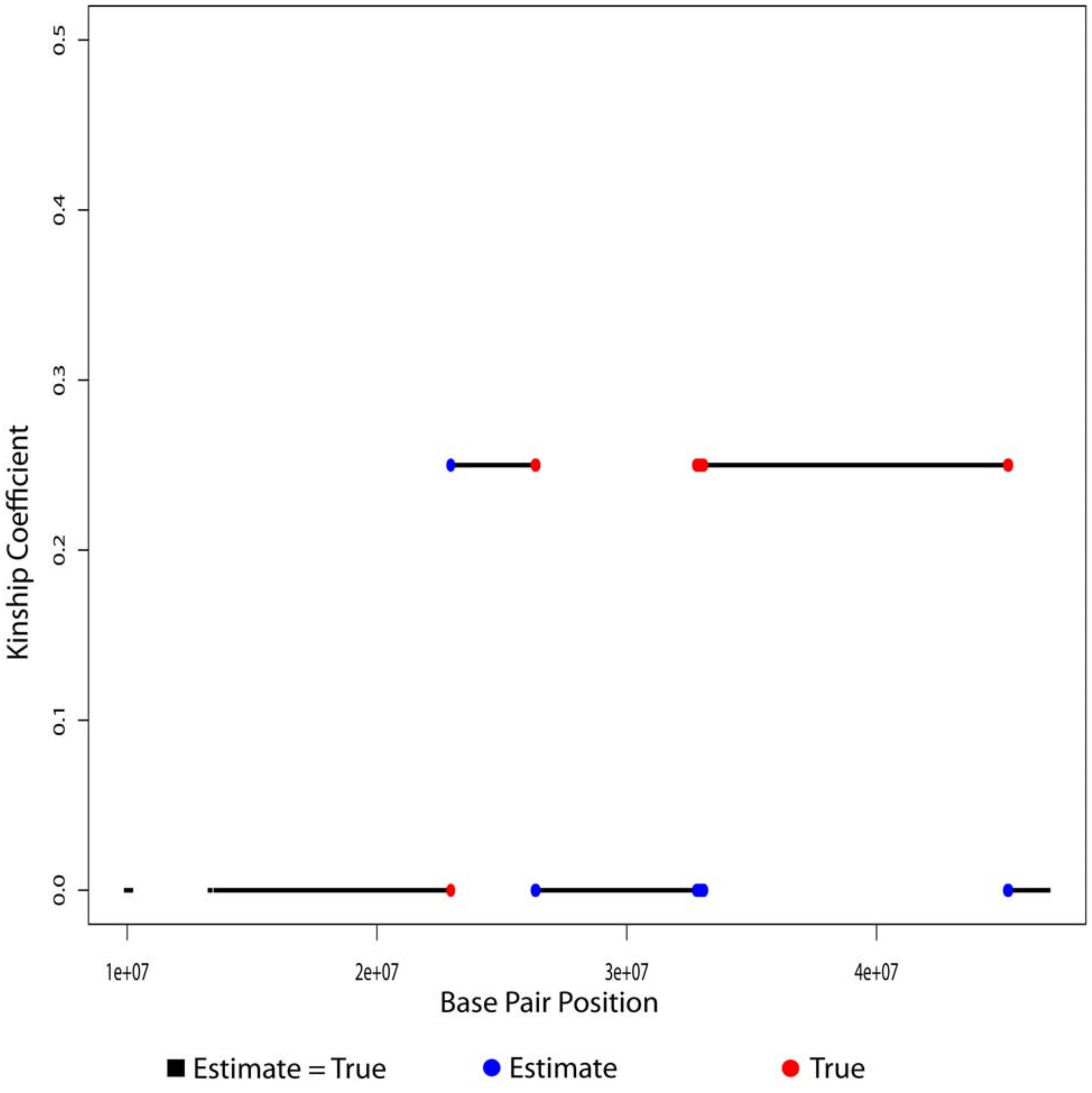
SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.02441358	0	0.002521212	0.02945601	0.1727273
100K	0	-0.003054624	0	0.0003818610	0.003818346	0.01450601
200K	0	-0.004834165	0	0.0006695179	0.0061732	0.03273717
500K	0	-0.001808482	0	0.0002246255	0.002257733	0.009799804

Supplementary Table 42: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 19 and 21 (true global kinship coefficient 0.003906) and chromosome 21 across chip types.

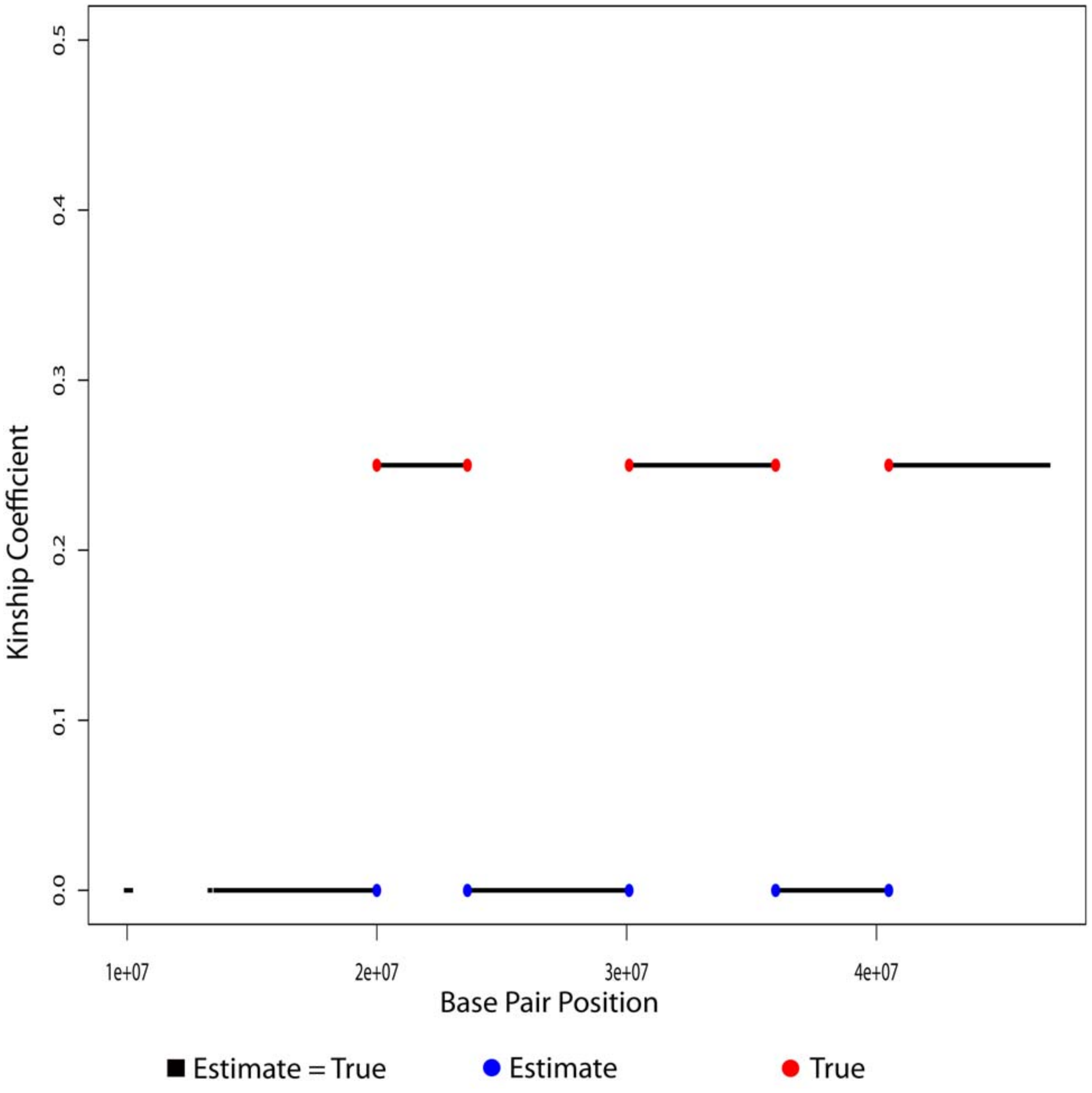
incorrect local kinship coefficient. Misassignments usually occur at IBD block boundaries. Errors tend to extend one block at the expense of a neighboring block.



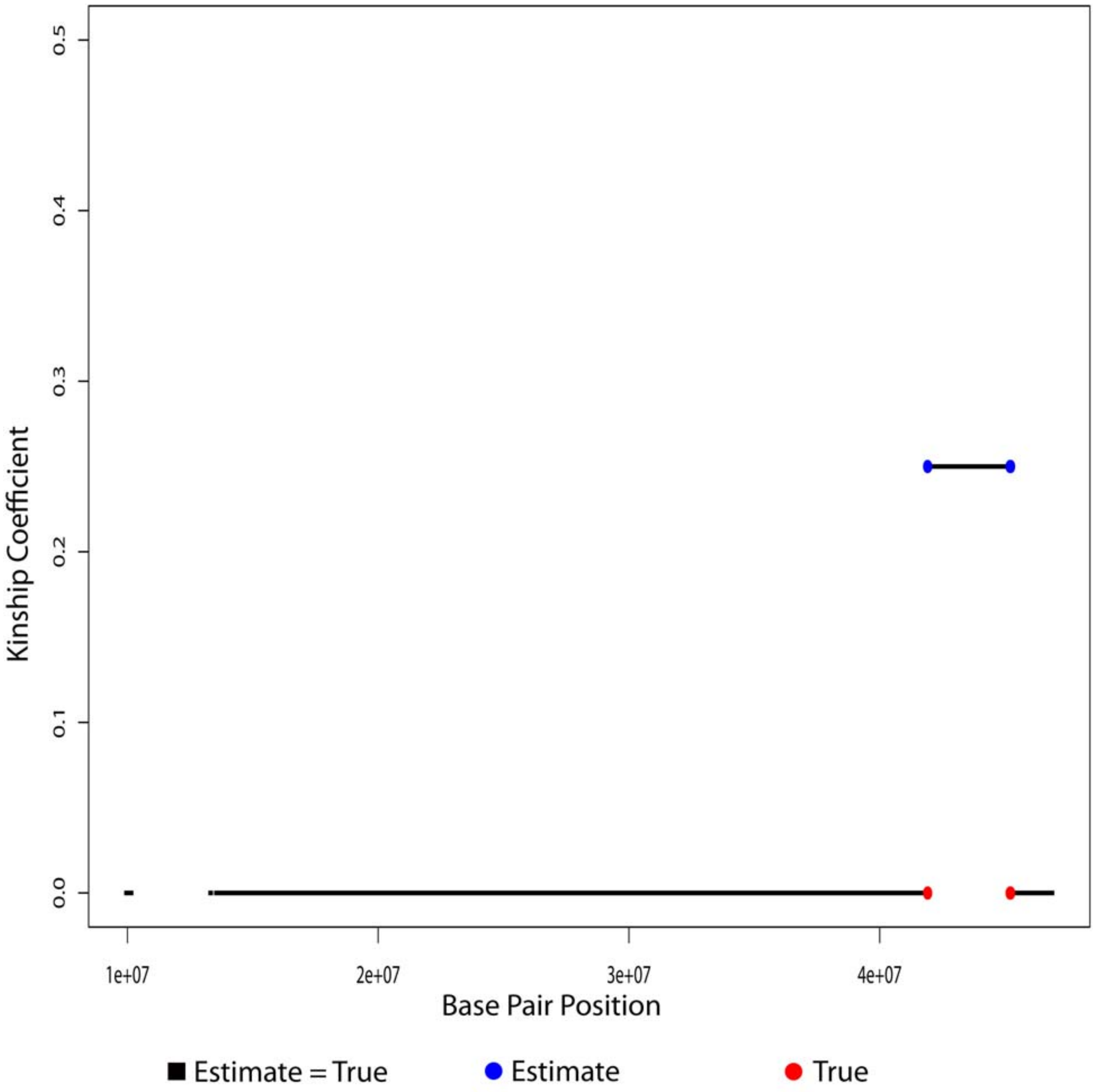
Supplementary Figure 7: True and estimated local kinship coefficients on chromosome 21 for a typical replicate of the 500K SNP set using the pair 3 and 4 (true global kinship coefficient 0.25). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 7143 SNPs depicted, 12 were assigned incorrect local kinship coefficients.



Supplementary Figure 8: True and estimated local kinship coefficients on chromosome 21 for a typical replicate of the 500K SNP set using the pair 4 and 7 (true global kinship coefficient 0.125). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 7143 SNPs depicted, 58 were assigned incorrect local kinship coefficients.



Supplementary Figure 9: True and estimated local kinship coefficients on chromosome 21 for a typical replicate of the 500K SNP set using the pair 4 and 19 (true global kinship coefficient 0.03125). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 7143 SNPs depicted, 29 were assigned incorrect local kinship coefficients.



Supplementary Figure 10: True and estimated local kinship coefficients on chromosome 21 for a typical replicate of the 500K SNP set using the pair 19 and 21 (true global kinship coefficient 0.003906). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 7143 SNPs depicted, 7 were assigned incorrect local kinship coefficients.

The final step in testing our local kinship coefficient estimation procedure is to show that it works well on long chromosomes, not only short chromosomes such 21. Using chromosome 1 data, we computed the distributions over 500 simulated pedigrees of the average absolute differences for each of the four SNP densities and for several relative pairs. The results are in Supplementary Tables 43, 44, 45, and 46.

SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0.001636126	0.003193252	0.05971859	0.06309293	0.1229926	0.1717932
100K	0.001304773	-0.004157297	0.01606502	0.01779450	0.03974629	0.07081113
200K	0.000415999	7.627588e-05	0.005207691	0.005668557	0.01126084	0.01731788
500K	0.0007439369	0.0005636464	0.002963349	0.003139079	0.005714512	0.008877647

Supplementary Table 43: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 3 and 4 (true global kinship coefficient 0.25) and chromosome 1 across chip types.

SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.01480084	0.03141361	0.03661453	0.0880299	0.1485602
100K	0.0002174622	-0.006724771	0.004756986	0.007163097	0.02105096	0.05058715
200K	0.0002002958	-0.0005850755	0.002141625	0.002440558	0.005466192	0.008412425
500K	8.679264e-05	-7.86782e-05	0.001419680	0.001548703	0.003176084	0.005058771

Supplementary Table 44: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 4 and 7 (true global kinship coefficient 0.125) and chromosome 1 across chip types.

SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.02183667	0.009162304	0.01679712	0.05543091	0.09554974
100K	0	-0.005713569	0.002215396	0.003363651	0.01244087	0.04093726
200K	0	-0.001525074	0.0009244423	0.001339178	0.00420343	0.007580426
500K	0	-0.0006881066	0.0005672519	0.0007883252	0.002264757	0.004662005

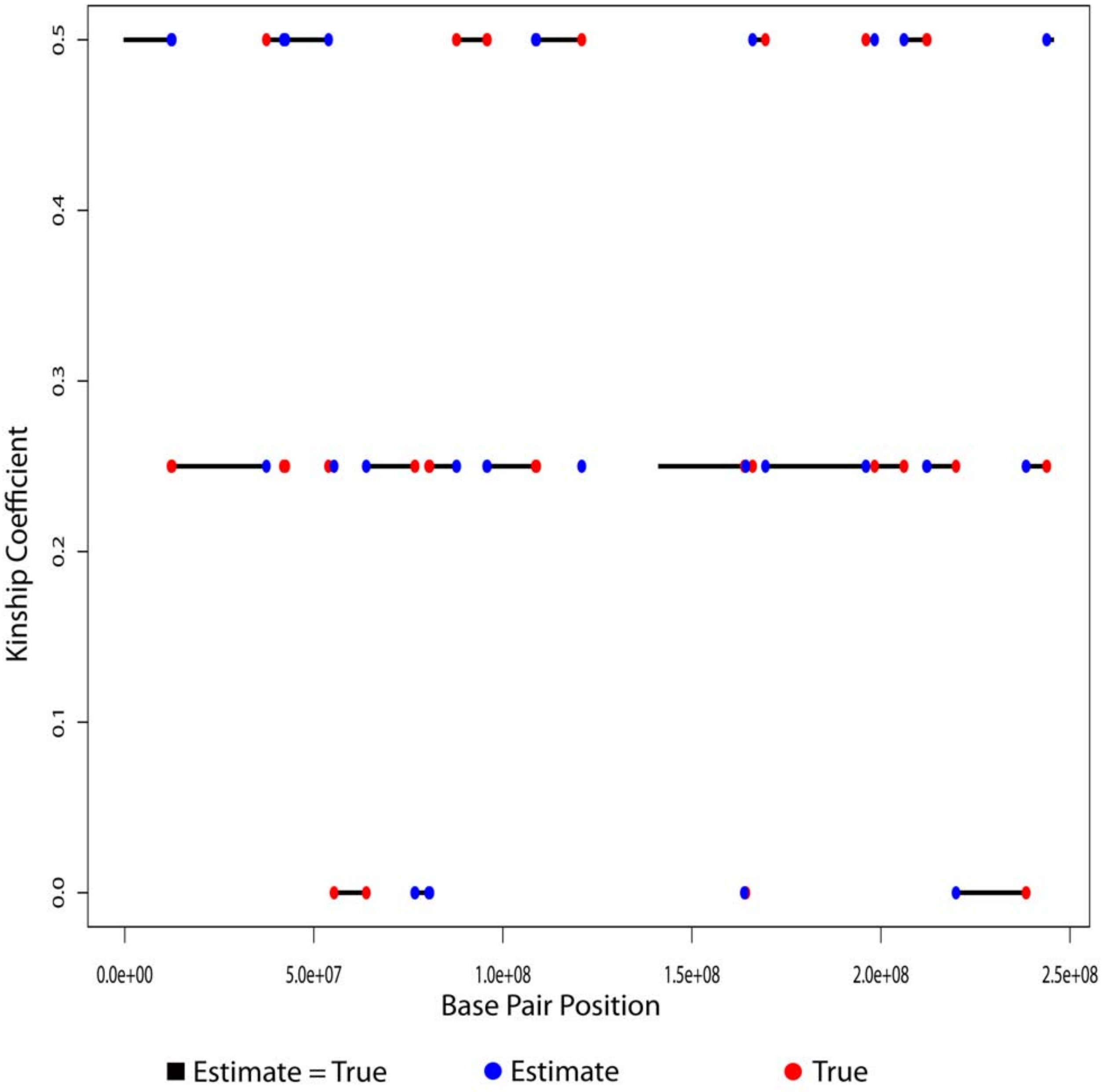
Supplementary Table 45: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 4 and 19 (true global kinship coefficient 0.03125) and chromosome 1 across chip types.

To visualize how good these local kinship coefficient estimates are, we again selected the simulated pedigree at the mean of the distributions of the average absolute differences for the

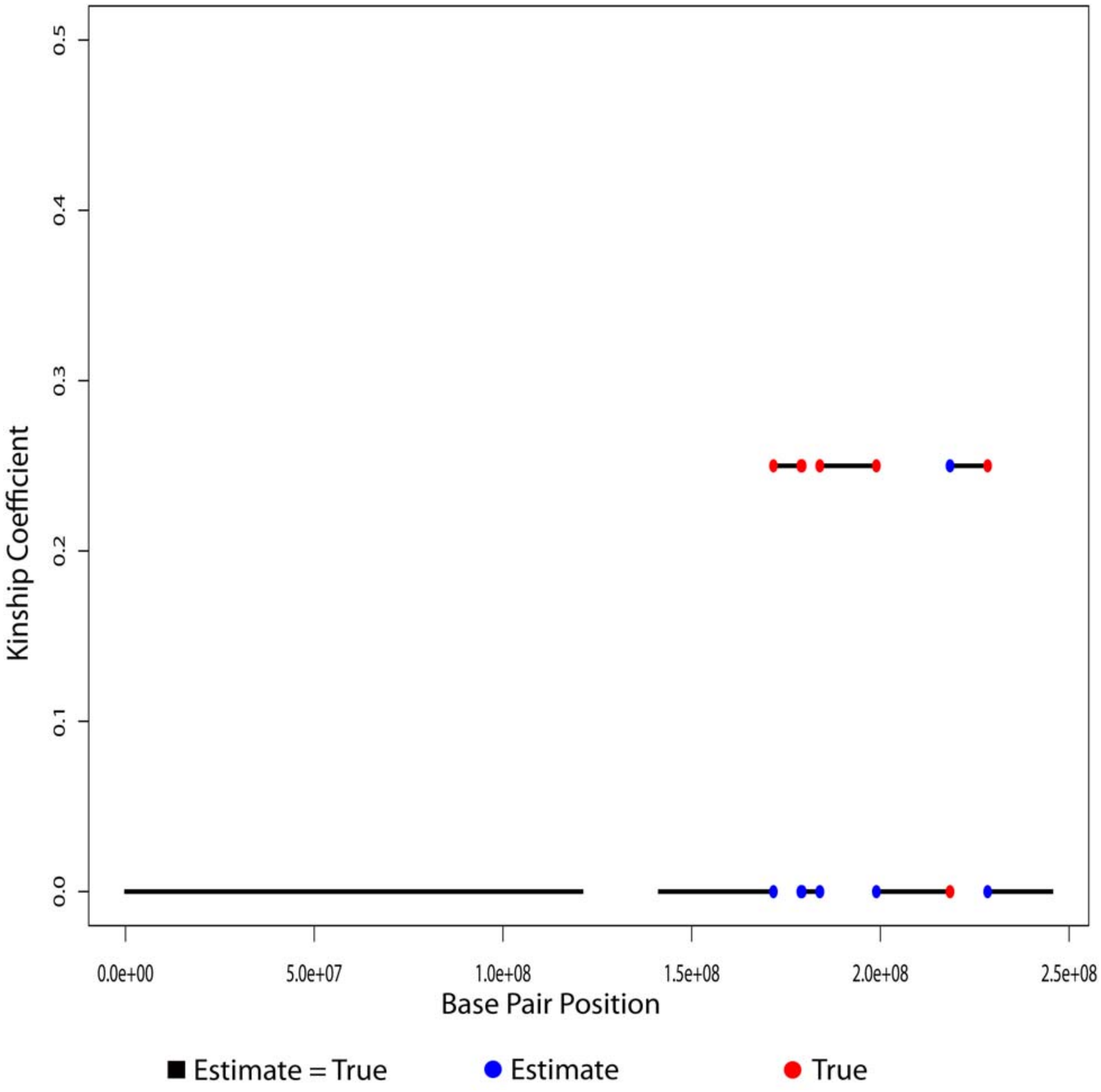
SNPs	Min	–2 SD	Median	Mean	+2 SD	Max
10K	0	-0.01454960	0	0.003329188	0.02120798	0.07591623
100K	0	-0.003431840	0	0.001422638	0.006277115	0.03463086
200K	0	-0.001359265	0	0.0004023207	0.002163906	0.006224578
500K	0	-0.0006511118	0	0.0002259212	0.001102954	0.003577097

Supplementary Table 46: Distributions for 500 simulated pedigrees of the *aad* for general λ_1, λ_2, for pair 19 and 21 (true global kinship coefficient 0.003906) and chromosome 1 across chip types.

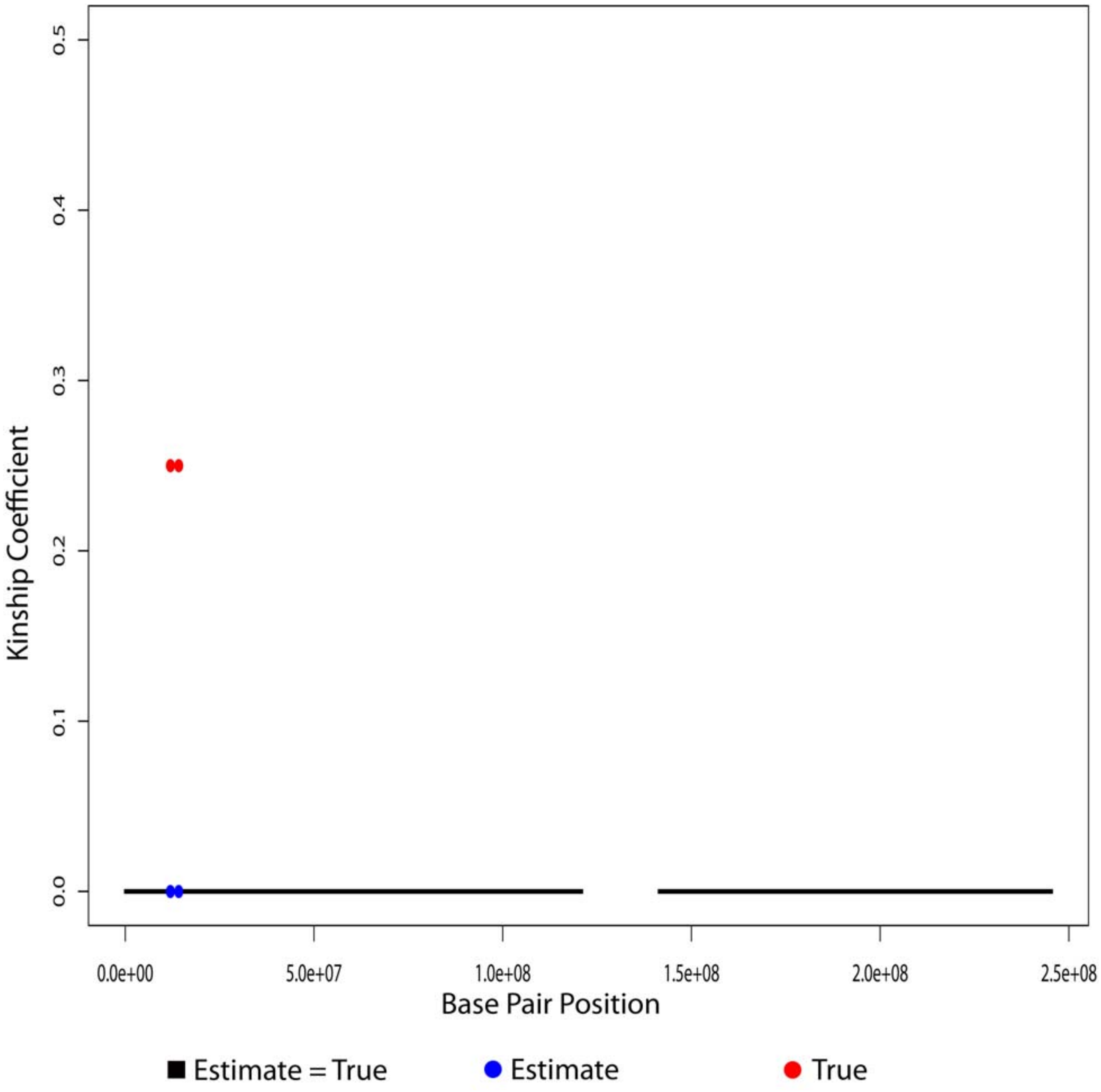
500K SNP density and graphed the true and estimated local kinship coefficients for all 40,326 SNPs listed on chromosome 1. The results are in Supplementary Figures 11, 12, and 13, and in Figure 6 from the main article. These results show that our local kinship coefficient estimation procedure captures even complicated IBD structure by using the information inherent to dense SNP genotypes.



Supplementary Figure 11: True and estimated local kinship coefficients on chromosome 1 for a typical replicate of the 500K SNP set using the pair 3 and 4 (true global kinship coefficient 0.25). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 40,326 SNPs depicted, 506 were assigned incorrect local kinship coefficients.



Supplementary Figure 12: True and estimated local kinship coefficients on chromosome 1 for a typical replicate of the 500K SNP set using the pair 4 and 19 (true global kinship coefficient 0.03125). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 40,326 SNPs depicted, 127 were assigned incorrect local kinship coefficients.



Supplementary Figure 13: True and estimated local kinship coefficients on chromosome 1 for a typical replicate of the 500K SNP set using the pair 19 and 21 (true global kinship coefficient 0.003906). Each SNP's local kinship coefficient is shown with a small black square when the true and estimated values are identical. When they differ, the true value is shown as a red dot and the estimated as a blue dot. Of the 40,326 SNPs depicted, 36 were assigned incorrect local kinship coefficients.

We have demonstrated that with dense SNP genotypes our penalized optimization method of estimating local kinship coefficients is accurate and captures even complex patterns of IBD sharing for both large and small chromosomes, and across the entire spectrum of genetic relatedness for which our global kinship coefficient estimator works well.

1.3 Pedigree Construction

To test our pedigree construction algorithm, we again used simulated replicates of the pedigree in Figure 1 of the main article. In data set 1 we genotyped all individuals in two replicated pedigrees. In data set 2 we genotyped everyone except individuals 7–12 in two replicated pedigrees. We then clustered the individuals based on their estimated global kinship coefficients, using several cutoff values. The results are in Supplementary Table 47. The table shows that all but one test accurately reconstructed the expected pedigree structures. Moreover, individuals originating in different pedigrees were never clustered together. Since the true global kinship coefficient for a grandparent-grandchild pair is 0.125, it is not surprising that using this exact value as the cutoff will cause problems when individuals 7–12 are missing and thus many grandparent-grandchild pairs are the closest relationships linking people, as in data set 2. For example, individuals 6 and 15 are a grandparent-grandchild pair, so with this cutoff, if the global kinship coefficient estimate is slightly undervalued, they will not be clustered, as occurred. However, the husband of 6, individual 4, was clustered with 15. When the cutoff is lowered slightly, for example to 0.1, all links are included in the clustering algorithm even in data set 2, and the complete pedigree structure is imputed.

Cutoff Value	Data Set	Expected Ped Count	Formed Ped Count	Formed Pedigrees
0.20	1	2	2	[1-1 to 1-21], [2-1 to 2-21]
0.20	2	12	12	[1-1,1-2,1-3,1-4], [1-5], [1-6], [1-13,1-16,1-19], [1-14,1-17,1-20], [1-15,1-18,1-21], [2-1,2-2,2-3,2-4], [2-5], [2-6], [2-13,2-16,2-19], [2-14,2-17,2-20], [2-15,2-18,2-21]
0.125	1	2	2	[1-1 to 1-21], [2-1 to 2-21]
0.125	2	2	7	[1-1,1-2,1-3,1-4,1-14,1-15, 1-17,1-18,1-20,1-21], [1-6], [1-5,1-13,1-16,1-19], [2-1,2-2,2-3,2-4], [2-5,2-13,2-14,2-16,2-17,2-19,2-20], [2-15,2-18,2-21], [2-6]
0.10	1	2	2	[1-1 to 1-21], [2-1 to 2-21]
0.10	2	2	2	[1-1 to 1-21], [2-1 to 2-21]

Supplementary Table 47: Pedigree construction results from two simulated replicates of the pedigree in Figure 1 of the main article. In data set 1 everyone is genotyped. In data set 2 individuals 7–12 are not genotyped in either pedigree. The “1-” prefix indicates an individual is from the first simulated pedigree and “2-” indicates the second. All individuals listed within a set of square brackets were clustered together.