1 Supplementary Materials

1.1 The concordance and discordance rates among GenCall, GenoSNP and M^3 across the whole genome.

Table S1: The concordance and discordance rates of both homozygotes and heterozygotes among GenCall, GenoSNP and M^3

		M^3 (%)		
Algorithm		Major-Homo	Heter	Minor-Homo
GenCall (%)	Major-Homo	62.54	0.05	0.01
99.87	Heter	0.03	29.43	0.01
	Minor-Homo	0.01	0.01	7.90
GenoSNP (%)	Major-Homo	62.35	0.06	0.06
99.64	Heter	0.16	29.48	0.03
	Minor-Homo	0.04	0.01	7.80

Note: M^3: the modified finite mixture model; Major-Homo: major homozygote; Heter: heterzygote; Minor-Homo: minor homozygote.

Despite the overall high concordance rates (Table 1 in the main paper), there are many discrepancies among these three algorithms. In Table S1, the concordance for specific genotypes, i.e. major homozygote, heterozygote, and minor homozygote, is summarized when the null genotypes are excluded from the comparisons. We note that the major homozygote calls by GenCall are more frequently called heterozygotes by M^3. For example, 0.05% of genotypes called as major homozygote by GenCall are called heterozygote by M^3, but only 0.03% of genotypes called as major homozygote by M^3 are called heterozygote by GenCall. This is partially due to the fact that M^3 has the high call rate (99.64%). The opposite pattern is observed in the comparison between M^3 and GenoSNP, and this conclusion is consistent with one previous result in GenoSNP paper (Giannoulatou et al, 2008). The reason may be that the M^3 is largely a population-based strategy.

When the null genotypes are included in the analysis, the comparison results are summarized in Table S2. It can be seen that a certain proportion of null class by GenCall or GenoSNP are assigned to each of the three possible genotypes by M^3. For example, 0.98% and 0.38% of genotypes that cannot be called by GenCall and GenoSNP are respectively genotyped to major homozygote by M^3.

Table S2: The concordance and discordance rates of homozygotes, heterozygotes and null class among GenCall, GenoSNP and M^3

		M^3 (%)			
Algorithm		Major-Homo	Heter	Minor-Homo	Null
GenCall (%)	Major-Homo	61.26	0.04	0.01	0.14
97.95	Heter	0.03	28.83	0.01	0.08
	Minor-Homo	0.01	0.01	7.74	0.03
	Null	0.98	0.58	0.12	0.11
GenoSNP (%)	Major-Homo	60.70	0.06	0.06	0.13
98.69	Heter	0.16	29.18	0.03	0.11
	Minor-Homo	0.04	0.01	7.72	0.03
	Null	0.38	0.23	0.07	0.09

Note: M^3: the modified mixture model; Major-Hom: major homozygote; Heter: heterzygote; Minor-Hom: minor homozygote.

When we compare the calling results between different algorithms and what are recorded by the HapMap Project for the HapMap samples genotyped, there is the potential allele labeling issue on a SNP array. This is due to the well documented fact that the designated allele on a chip may not correspond to the allele that the probe is designed for. Rather, the probe may be detecting the variation on the complementary strand. For a given SNP whose allele frequency is far different from 50%, it is relatively straight forward to identify this labeling issue because if the observed allele frequency for an allele is 20%, whereas it is around 80% in the reference database, the result maybe likely due to a labeling problem. However, when the allele frequency is close to 50%, this issue may be more difficult to detect. If we do not consider this potential labeling problem, the estimated genotyping calling accuracy may be underestimated for all calling algorithms. Therefore, we implemented some strategies as detailed in the paper to avoid potential labeling issue.

Table S3: The summary of failed SNPs under different E cutoffs

	E (Error)	GenCall	GenoSNP	M^3
Num-SNPs	< 50	57148	58117	58110
	< 10	57805	60117	59464
	< 1	58405	61199	60702

Note: E: the error groups caused by the mis-assignment of the major allele, and this error is classified into three groups, that is, the E variable measuring the different values is less than 50, $E < 10$, and $E < 1$; Num-SNPs: the number of SNPs failed E thresholds.

The data set analyzed in this paper contains 3258 genotyped samples, most being African Americans or European Americans, with 141 out of the 3258 samples from 38 distinct HapMap samples. These HapMap samples were included in the genotyping set for quality control purpose. The genotypes of these 141 samples were called together with other samples.

When we examined the genotypes of these HapMap samples, the major homozygote frequencies are similar to the minor homozygote frequencies at some SNPs in these populations (i.e. the inferred major allele frequency is around 50%), so it is difficult to designate the major allele for these SNPs due to the potential labeling issue discussed above. Therefore, we introduced the following metric to remove these potentially problematic SNPs from our method comparisons.

$E_j = \| Homozygote_{j|HapMap} - Homozygote_{j|method} \|$

We set different thresholds for E so that the potentially problematic SNPs are not included in our comparisons. In real application, we varied the threshold at 50, 10, and 1 to exclude these SNPs in our calculations of concordance results for three genotyping algorithms with the HapMap data. Under three different cutoffs, the failed SNPs in terms of 3 criterions, GenCall, GenoSNP and M^3, are summarized in Table S3. In the main paper, we calculated the average call rate and accuracy by 3 criterions under each cutoff (Tables 2-3 in the main text) to compare the performance of all algorithms.

1.2 The effectiveness of prepocessing data, selection methods, posterior probability cutoffs, and Two-Stage calling in M^3.

In our empirical study, we tried a simple transformation procedure on chromosome 22 by taking the log transformation of the raw intensity (Giannoulatou et al, 2008), as the input for our allele calling algorithm. The comparison results are summarized in Table S4.

Table S4: The comparisons of call rates and concordance on HapMap samples between log intensity (M^3_{norm}) and raw intensity (M^3)

Criterion	E (Error)	Item	M^3_{norm} (%)	M^3 (%)
M^3_{norm}, M^3	< 50	Call Rate	99.72	99.76
		Accuracy	98.66	99.16
	< 10	Call Rate	96.73	99.77
		Accuracy	98.89	99.34
	< 1	Call Rate	99.72	99.77
		Accuracy	99.02	99.43

Note: M^3_{norm}: the modified mixture model for log intensity; M^3: the modified mixture model for raw intensity; The data is for chromosome 22; The call rate and accuracy are average values by two criterions (M^3_{norm} and M^3) under E < 50, 10, 1.

Under the same posterior probability cutoff (85%), the analysis of raw intensity data directly provides larger call rate and accuracy. We think this may be due to the fact that although normal transformation may be more in lines of model assumptions, the covariance matrix in the estimation procedure is hard to be positive definite compared to the covariance of raw intensity, and it may lower the posterior probabilities for some individuals. However, there may be other preprocessing or normalization approaches that can improve the calling rate and accuracy. This may be a topic worth further investigation. In addition to data transformation, we are exploring the incorporation of the 50mer probe sequence information in the reference SNP selection to improve calling accuracy in our ongoing work.

Table S5: The comparisons of call rates, concordance and HWE on HapMap samples in M^3

Criterion	E (Error)	Item	M^3_{APR} (%)	M^3_{MD} (%)	M^3_{CD} (%)
M^3_{APR}, M^3_{MD}, M^3_{CD}	< 50	Call Rate	99.73	99.75	99.76
		Accuracy	99.22	99.27	99.32
	< 10	Call Rate	99.74	99.76	99.77
		Accuracy	99.23	99.28	99.33
	< 1	Call Rate	99.74	99.76	99.77
		Accuracy	99.25	99.27	99.33
HWE	AA I	2005	400	440	440
	AA II	83	60	74	65
	EA I	867	650	674	673
	EA II	158	99	118	115

Note: M^3_{APR}: the modified mixture model using APR to select optimal reference SNP; M^3_{MD}: the modified mixture model using Maholanobis Distance to select reference SNP; M^3_{CD}: the modified mixture model using Cluster Distance to select reference SNP; Call Rate: the percentage of valid genotypes; Accuracy: the percentage of consistent genotype; Criterion: which algorithm is selected to count the different values between this algorithm and HapMap

project due to the mis-assignment of major allele. E: the error groups caused by the mis-assignment of the major allele, and this error is classified into three groups, that is, the E variable measuring the different values is less than 50, E < 10, and E < 1; The data is for chromosome 22.

The selection of the appropriate reference SNP is critical to M^3. In the second stage of M^3, we applied APR, Maholanobis Distance, and Cluster Distance in Step III to determine the final reference SNP. These metrics focus on different aspects of the clusters, so a rare SNP (testing SNP)'s more common allele may behave very much like the alleles of another SNP (reference SNP), although there is not enough information about the rare SNP cluster. In this case, a better calling result may be achieved when the testing and reference SNPs are jointly called. Among the three metrics, the comparison results, denoted by M^3_{APR}, M^3_{MD} and M^3_{CD}, are summarized in Table S5. In general, M^3_{CD} provides the best result in call rate and call accuracy.

Table S6: The comparisons of call rates and concordance on HapMap samples under two cutoffs in M^3

Criterion	E (Error)	Item	$M^3_{70\%}$ (%)	$M^3_{85\%}$ (%)
$M^3_{70\%}$, $M^3_{85\%}$	< 50	Call Rate	99.87	99.76
		Accuracy	99.28	99.19
	< 10	Call Rate	99.87	99.77
		Accuracy	99.45	99.37
	< 1	Call Rate	99.87	99.77
		Accuracy	99.53	99.45

Note: M^3: the modified mixture model; 70% and 85%: two cutoffs; The call rate and accuracy are average values by two criterions ($M^3_{70\%}$ and M^3) under E < 50, 10 and 1. This comparison table is for chromosome 22.

The metric posterior probability determines no calls in M^3. In an earlier analysis, we chose 70% posterior probability cutoff to genotype SNP arrays, and the relevant genotyping result is compared to that by GenoSNP using 85% cutoff. Due to the difference in the nature of the algorithms, it is not advisable to use the same threshold. The real data set considered in this paper contains 3258 individuals with 942,313 SNPs, and the posterior probability in M^3 is calculated for all individuals (3258 subjects) within each SNP, whereas the relative posterior probability in GenoSNP is calculated for all SNPs within each beadpool, the size of SNPs within one beadpool is much larger than that of individuals within every SNP (around 940,000 SNPs with 25 beadpools). Figure 3 (Giannoulatou et al, 2008) demonstrates this discrepancy between GenoSNP and the population based method. In general, the larger size

of the observations, the larger posterior probability is achieved. To be equivalent to compare GenoSNP and M^3, we should use a lower cutoff.

In a later analysis, we changed the posterior probability for M^3 from 70% to 85%, the same as that of GenoSNP for the whole genome to explore the internal connection between the posterior probability and the effect of M^3. We note that the cutoff of the posterior probability is not strongly related to our two-stage algorithm. In the first stage, a larger cutoff employed here will allow more SNPs to be recalled in the second stage, and an appropriate reference SNP will improve the genotyping quality of these poor SNPs and reduce the no-call rate in the second stage. The relevant results are given in Table S6. In brief, two cutoffs, 70% and 85%, only lead to 0.08%∼0.09% difference in calling accuracy in M^3.

Table S7: The comparisons of call rates and concordance on HapMap samples under two stages in M^3

Criterion		Item	GenCall (%)	GenoSNP (%)	M^3_{stepI} (%)	M^3_{stepII} (%)
GenCall, GenoSNP, M^3	Overall SNPs	Call Rate	96.56	99.14	99.35	99.76
		Accuracy	96.35	98.44	98.24	99.19
	MAF<0.1	Call Rate	95.90	99.00	99.03	99.69
		Accuracy	95.62	98.20	98.13	99.12
	MAF<0.05	Call Rate	96.04	98.87	98.63	99.64
		Accuracy	95.69	97.91	97.75	98.96
	MAF<0.01	Call Rate	95.38	98.82	97.87	99.56
		Accuracy	94.94	97.70	96.65	98.64

Note: M^3_{stepI}: the modified mixture model only using the first stage; M^3_{stepII}: the modified mixture model using the first and second stages; Call Rate and Accuracy: the average values of call rate and accuracy by three criterions (GenCall, GenoSNP and M^3) under three E cutoffs ($<$ 50, 10 and 1). This comparison table is for chromosome 22.

M^3 is a two-stage SNP calling strategy, so we calculated the concordance rate of two stages, separately, and the results are shown in Table S7. In brief, both call rate and accuracy were increased in the second stage of M^3 compared to that of the first stage. In particular, the second stage of M^3 greater increased call rate and accuracy for rare variants, compared to common SNPs.