The Relationship between Imputation Error and

Statistical Power in Genetic Association Studies

in Diverse Populations
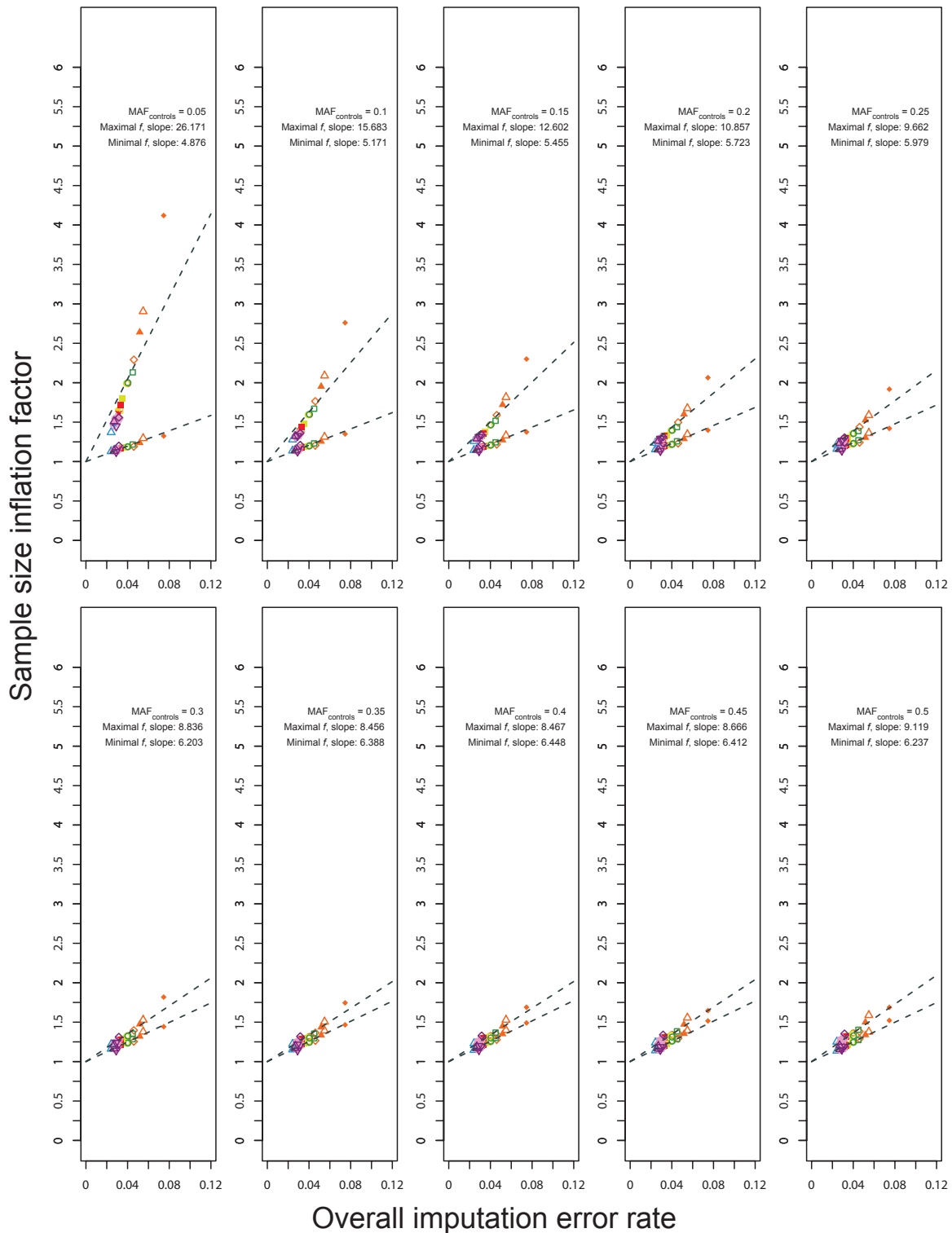
Lucy Huang, Chaolong Wang, and Noah A. Rosenberg

Figure S1: Maximal and minimal sample-size inflation factors as functions of the overall imputation error rate, for an imputed disease locus with true minor allele frequency fixed in controls, excluding the San and Mbuti Pygmy populations. Each plot has $MAF_{controls}$ fixed at a different value. Population symbols are the same as in Figure 4, and two data points appear for each population, a maximum and a minimum. Best-fit linear regression lines for the maxima and minima, forced through the point $(0,1)$, indicate the increase in the inflation factor with increasing imputation error rate.
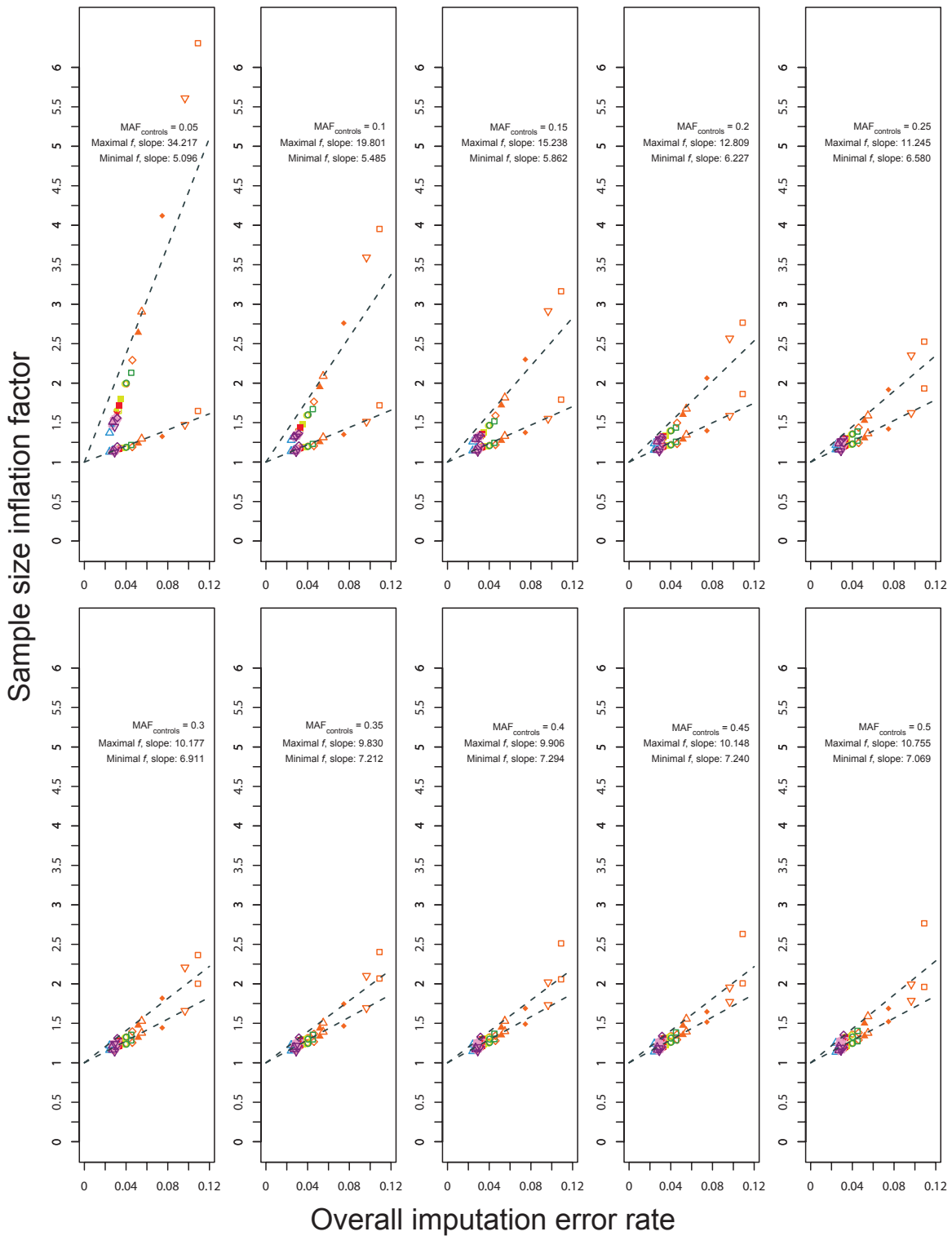
Figure S2: Maximal and minimal sample-size inflation factors as functions of the overall imputation error rate, for an imputed disease locus with true minor allele frequency fixed in controls, considering all 29 populations. Each plot has $\text{MAF}_{controls}$ fixed at a different value. Population symbols are the same as in Figure 4, and two data points appear for each population, a maximum and a minimum. Best-fit linear regression lines for the maxima and minima, forced through the point $(0,1)$, indicate the increase in the inflation factor with increasing imputation error rate. The plot for $\text{MAF}_{controls} = 0.3$ also appears in Figure 4.

Table S1: Genotype misclassification error rates ϵ_{ij} in each of 29 populations. Each column shows a particular error rate ϵ_{ij}, where ϵ_{ij} represents the probability that genotype i is imputed as genotype j (1, minor allele homozygote; 2, heterozygote; 3, major allele homozygote). For each population, the greatest of the six error rates is shown in bold. The values of ϵ_{ij} are identical to those plotted in Figure 1. The final column represents the overall imputation error rate, a weighted sum of the six ϵ_{ij} with the weights determined by the frequencies of the three categories of genotype.

Population	ϵ_{12}	ϵ_{13}	ϵ_{21}	ϵ_{23}	ϵ_{31}	ϵ_{32}	Overall imputation error rate
San	**0.236**	0.082	0.101	0.180	0.037	0.104	0.109
Mbuti Pygmy	0.159	0.059	0.067	**0.161**	0.034	0.103	0.096
Biaka Pygmy	**0.134**	0.038	0.045	0.111	0.026	0.086	0.075
Bantu (Kenya)	**0.101**	0.035	0.028	0.073	0.021	0.053	0.052
Bantu (S. Africa)	**0.108**	0.049	0.027	0.066	0.023	0.063	0.055
Yoruba	**0.073**	0.030	0.021	0.050	0.023	0.034	0.040
Mandenka	**0.080**	0.025	0.027	0.065	0.022	0.042	0.046
Mozabite	**0.073**	0.027	0.019	0.045	0.023	0.035	0.039
Bedouin	**0.057**	0.038	0.020	0.040	0.019	0.028	0.035
Palestinian	**0.057**	0.033	0.017	0.040	0.020	0.023	0.033
Druze	**0.047**	0.028	0.014	0.032	0.020	0.024	0.031
Basque	0.034	**0.037**	0.009	0.022	0.021	0.018	0.027
Russian	**0.049**	0.027	0.011	0.028	0.018	0.013	0.024
Adygei	**0.038**	0.022	0.011	0.027	0.021	0.017	0.027
Balochi	**0.046**	0.036	0.015	0.038	0.020	0.021	0.031
Kalash	**0.044**	0.036	0.015	0.028	0.021	0.022	0.031
Burusho	**0.060**	0.032	0.018	0.033	0.021	0.026	0.033
Uygur	**0.048**	0.021	0.012	0.022	0.024	0.020	0.029
Yakut	**0.050**	0.034	0.015	0.027	0.025	0.019	0.032
Mongola	**0.064**	0.025	0.011	0.029	0.023	0.020	0.029
Daur	**0.057**	0.026	0.011	0.025	0.021	0.020	0.028
Yi	**0.057**	0.028	0.009	0.023	0.024	0.021	0.029
Cambodian	**0.050**	0.016	0.013	0.032	0.023	0.019	0.030
Lahu	**0.045**	0.025	0.013	0.021	0.022	0.017	0.027
Melanesian	**0.059**	0.030	0.025	0.041	0.024	0.035	0.040
Papuan	**0.068**	0.037	0.036	0.047	0.025	0.038	0.045
Pima	**0.043**	0.012	0.011	0.019	0.029	0.017	0.029
Maya	**0.032**	0.027	0.016	0.027	0.019	0.018	0.027
Colombian	**0.054**	0.051	0.013	0.031	0.023	0.020	0.032