**Table S5**

The effect of high vs. low quality subpopulation classification on meta-analysis in simulated samples.

|  |  |  |  |
| --- | --- | --- | --- |
| Grouping | Depth | Theta | Proportion of significant variants |
| P < 10-6 | P < 10-5 | P < 10-4 | P < 10-3 | P < 0.01 |
| True ancestry labels | 5 | -0.05 | 0.0073 | 0.0125 | 0.0235 | 0.05 | 0.1145 |
| 0 | 0.0147 | 0.0388 | 0.0919 | 0.1955 | 0.3519 |
| 30 | -0.05 | 0.0139 | 0.04 | 0.1048 | 0.2389 | 0.4594 |
| 0 | 0 | 0 | 0.0001 | 0.0016 | 0.0127 |
| k-means (3 groups) | 5 | -0.05 | 0.1201 | 0.149 | 0.19 | 0.2509 | 0.3513 |
| 0 | 0.2907 | 0.3496 | 0.4195 | 0.4977 | 0.5826 |
| 30 | -0.05 | 0.0919 | 0.1122 | 0.1447 | 0.2017 | 0.3097 |
| 0 | 0.2183 | 0.2553 | 0.3054 | 0.3734 | 0.4747 |

We simulated 50,000 variants in 5,000 samples arising from 5 distinct subpopulations (1,000 samples each), at low (5x) and high (30x) depth, with no deviation from HWE (θ = 0) and moderate excess heterozygosity (θ = -0.05). We used one of two different groupings for our samples: for high-quality labels, we used the original true ancestry labels from which we simulated our data; for low-quality labels, we ran k-means classification on the first 2 principal components of genetic variation for all our samples to generate 3 groups. We meta-analyzed all data sets using Stouffer’s method. Type I error rates for low-depth samples were greatly inflated. For high-depth samples, when we used the true ancestry labels, Type I errors were well-controlled, with reasonable power to discover deviations from HWE, while when we used the crude k-means labels, Type I errors were greatly inflated, with surprisingly less power to discover deviations from HWE at less stringent P-value thresholds. These results highlight the importance of high-quality subpopulation classification for meta-analysis.