**Table S3**

Performance of the unadjusted test, meta-analysis, and RUTH on the subset of TOPMed freeze 5 chromosome 20 variants that are also found in 1000G.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variant set** | **Genotype Format** | **HWE Test** | **Proportion of Significant Variants** | **Total Variant Count** |
| **P < 10-2** | **P < 10-3** | **P < 10-4** | **P < 10-5** | **P < 10-6** |
| **HQ Variants** | **raw GT** | **Unadjusted** | 0.890 | 0.842 | 0.800 | 0.766 | 0.736 | 16,924 |
| **raw GT** | **Meta-analysis** | 0.062 | 0.020 | 8.0x10-3 | 3.8x10-3 | 2.3x10-3 | 16,924 |
| **raw GT** | **RUTH-Score** | 0.145 | 0.046 | 0.016 | 6.3x10-3 | 2.8x10-3 | 16,924 |
| **GL** | **RUTH-Score** | 0.032 | 9.3x10-3 | 3.7x10-3 | 2.0x10-3 | 1.5x10-3 | 16,924 |
| **raw GT** | **RUTH-LRT** | 0.125 | 0.035 | 0.011 | 4.2x10-3 | 1.9x10-3 | 16,924 |
| **GL** | **RUTH-LRT** | 0.039 | 0.016 | 7.4x10-3 | 3.1x10-3 | 2.2x10-3 | 16,924 |
| **LQ Variants** | **raw GT** | **Unadjusted** | 0.762 | 0.728 | 0.702 | 0.683 | 0.667 | 10,513 |
| **raw GT** | **Meta-analysis** | 0.649 | 0.616 | 0.592 | 0.575 | 0.560 | 10,513 |
| **raw GT** | **RUTH-Score** | 0.727 | 0.693 | 0.673 | 0.656 | 0.640 | 10,513 |
| **GL** | **RUTH-Score** | 0.698 | 0.669 | 0.648 | 0.631 | 0.618 | 10,513 |
| **raw GT** | **RUTH-LRT** | 0.719 | 0.686 | 0.663 | 0.643 | 0.627 | 10,513 |
| **GL** | **RUTH-LRT** | 0.693 | 0.662 | 0.639 | 0.621 | 0.605 | 10,513 |

For HQ variants, GL-based HWE tests had much better control of false positives than GT-based tests. Conversely, for LQ variants, GT-based HWE tests had a slightly better true positive rate than GL-based tests. Overall, GL-based tests had the best performance when considering the tradeoff between false positives and true positives (Figure S5-6).