# pKWmEB: Integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study

Wen-Long Ren[1, §], Yang-Jun Wen[1, §], Jim M. Dunwell[2], Yuan-Ming Zhang[1,*]

1 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China / Statistical Genomics Lab, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

2 School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, United Kingdom

[§]: These authors contributed equally to this work.

[*] **Correspondence:** Dr. Yuan-Ming Zhang, College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China. E-mail: soyzhang@njau.edu.cn / College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China. E-mail: soyzhang@mail.hzau.edu.cn

**Table S1. Comparison of statistical powers and false positive rates (FPR) among pKWmEB, KWmEB, KWsBC, GEMMA and mrMLM methods in the first simulation experiment**

| QTN | True parameter | | | | Power (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | KWsBC | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 95.4 | 93.7 | 89.9 | 72.8 | 94.8 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 66.7 | 60.4 | 6.0 | 22.1 | 63.9 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 52.5 | 53.1 | 53.0 | 24.2 | 49.3 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 96.3 | 95.4 | 98.8 | 89.9 | 97.3 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 40.3 | 37.0 | 88.0 | 35.6 | 42.7 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 67.7 | 64.0 | 28.2 | 31.4 | 63.5 |
| Average | | | | | 69.8 | 67.3 | 60.7 | 46.0 | 68.6 |
| False positive rate (FPR, %) | | | | | 3.18E-2 | 3.38E-2 | 4.74E-1 | 3.25E-2 | 1.68E-2 |

KWmEB: integration of Kruskal-Wallis test with empirical Bayes without polygenic background control; KWsBC: Kruskal-Wallis test with Bonferroni correction; GEMMA: genome-wide efficient mixed model association; mrMLM: multi-locus random-SNP-effect mixed linear model. The same is true for the later tables.

**Table S2. Comparison of statistical powers and false positive rates (FPR) among pKWmEB, KWmEB, KWsBC, GEMMA and mrMLM methods in the second simulation experiment**

| QTN | True parameter | | | | Power (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | KWsBC | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.6171 | 10 | 94.8 | 94.2 | 86.5 | 71.7 | 95.6 |
| 2 | 1 | 11655607 | 1.1435 | 5 | 71.0 | 69.6 | 1.3 | 22.1 | 65.9 |
| 3 | 2 | 5066968 | 1.1435 | 5 | 44.9 | 43.7 | 40.4 | 8.5 | 39.7 |
| 4 | 2 | 5134228 | 1.9806 | 15 | 98.5 | 99.2 | 100.0 | 98.8 | 98.8 |
| 5 | 2 | 5464675 | 1.1435 | 5 | 30.4 | 25.1 | 77.6 | 16.6 | 29.6 |
| 6 | 2 | 6137189 | 1.1435 | 5 | 75.0 | 74.2 | 47.4 | 37.3 | 76.1 |
| Average | | | | | 69.1 | 67.7 | 58.9 | 42.5 | 67.6 |
| False positive rate (FPR, %) | | | | | 3.97E-2 | 4.07E-2 | 6.01E-1 | 1.66E-2 | 2.10E-2 |

**Table S3. Comparison of statistical powers and false positive rates (FPR) among pKWmEB, KWmEB, KWsBC, GEMMA and mrMLM methods in the third simulation experiment**

| QTN | True parameter | | | | Power (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | KWsBC | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.7310 | 10 | 93.1 | 92.8 | 90.3 | 72.9 | 93.4 |
| 2 | 1 | 11655607 | 1.2240 | 5 | 57.4 | 51.6 | 1.8 | 15.4 | 51.3 |
| 3 | 2 | 5066968 | 1.2240 | 5 | 46.8 | 45.7 | 36.8 | 16.0 | 40.9 |
| 4 | 2 | 5134228 | 2.1201 | 15 | 98.3 | 99.1 | 99.7 | 93.8 | 98.3 |
| 5 | 2 | 5464675 | 1.2240 | 5 | 23.4 | 18.9 | 91.7 | 19.7 | 24.6 |
| 6 | 2 | 6137189 | 1.2240 | 5 | 52.5 | 51.4 | 9.2 | 16.7 | 44.9 |
| Average | | | | | 61.9 | 59.9 | 54.9 | 39.1 | 58.9 |
| False positive rate (FPR, %) | | | | | 4.81E-2 | 5.02E-2 | 6.25E-1 | 2.53E-2 | 2.65E-2 |

**Table S4. Comparison of statistical powers and false positive rates (FPR) among pKWmEB, KWmEB, KWsBC, GEMMA and mrMLM methods in the fourth simulation experiment**

| QTN | True parameter | | | | Power (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | KWsBC | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 92.6 | 90.1 | 100.0 | 79.2 | 93.5 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 78.9 | 73.6 | 17.7 | 32.7 | 79.5 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 63.2 | 64.8 | 91.2 | 34.3 | 64.3 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 93.1 | 90.3 | 100.0 | 87.3 | 93.7 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 54.6 | 52.6 | 100.0 | 45.1 | 60.6 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 74.5 | 74.9 | 71.4 | 44.9 | 78.4 |
| Average | | | | | 76.2 | 74.4 | 80.1 | 53.9 | 78.3 |
| False positive rate (FPR, %) | | | | | 2.61E-2 | 3.26E-2 | 8.81E-1 | 4.21E-2 | 2.49E-2 |

**Table S5. Comparison of statistical powers and false positive rates (FPR) among pKWmEB, KWmEB, KWsBC, GEMMA and mrMLM methods in the fifth simulation experiment**

| QTN | True parameter | | | | Power (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | KWsBC | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 92.2 | 90.1 | 91.2 | 70.8 | 94.2 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 66.7 | 61.9 | 4.9 | 20.1 | 64.0 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 52.2 | 50.9 | 51.7 | 18.6 | 46.4 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 94.6 | 95.4 | 99.3 | 89.1 | 95.7 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 39.9 | 37.2 | 90.0 | 35.4 | 42.9 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 66.6 | 66.0 | 28.4 | 30.6 | 64.6 |
| Average | | | | | 68.7 | 66.9 | 60.9 | 44.1 | 68.0 |
| False positive rate (FPR, %) | | | | | 3.23E-2 | 3.50E-2 | 4.84E-1 | 2.83E-2 | 1.78E-2 |

**Table S6. Comparison of mean squared errors (MSE) for each QTN effect among pKWmEB, KWmEB, GEMMA and mrMLM methods in the first simulation experiment**

| QTN | True parameter | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 0.0891 | 0.0945 | 0.2748 | 0.0701 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 0.0564 | 0.0608 | 0.6408 | 0.0674 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 0.0851 | 0.0869 | 0.7045 | 0.1423 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 0.1198 | 0.1228 | 0.1766 | 0.0867 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 0.0790 | 0.0806 | 0.9127 | 0.1318 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 0.0490 | 0.0496 | 0.5705 | 0.0659 |
| Average | | | | | 0.0797 | 0.0825 | 0.5467 | 0.0940 |

**Table S7. Comparison of mean squared errors (MSE) for each QTN effect among pKWmEB, KWmEB, GEMMA and mrMLM methods in the second simulation experiment**

| QTN | True parameter | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.6171 | 10 | 0.1210 | 0.1252 | 0.2108 | 0.0692 |
| 2 | 1 | 11655607 | 1.1435 | 5 | 0.0727 | 0.0722 | 0.6968 | 0.0484 |
| 3 | 2 | 5066968 | 1.1435 | 5 | 0.0636 | 0.0628 | 0.6659 | 0.0973 |
| 4 | 2 | 5134228 | 1.9806 | 15 | 0.1183 | 0.1167 | 0.3080 | 0.1177 |
| 5 | 2 | 5464675 | 1.1435 | 5 | 0.0559 | 0.0602 | 0.9415 | 0.0924 |
| 6 | 2 | 6137189 | 1.1435 | 5 | 0.0607 | 0.0598 | 0.5917 | 0.0861 |
| Average | | | | | 0.0820 | 0.0828 | 0.5691 | 0.0852 |

**Table S8. Comparison of mean squared errors (MSE) for each QTN effect among pKWmEB, KWmEB, GEMMA and mrMLM methods in the third simulation experiment**

| QTN | True parameter | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.7310 | 10 | 0.1377 | 0.1329 | 0.2406 | 0.0886 |
| 2 | 1 | 11655607 | 1.2240 | 5 | 0.0748 | 0.0804 | 0.7373 | 0.0550 |
| 3 | 2 | 5066968 | 1.2240 | 5 | 0.0863 | 0.0859 | 0.7219 | 0.1497 |
| 4 | 2 | 5134228 | 2.1201 | 15 | 0.1693 | 0.1723 | 0.2132 | 0.1114 |
| 5 | 2 | 5464675 | 1.2240 | 5 | 0.0710 | 0.0666 | 1.0249 | 0.1590 |
| 6 | 2 | 6137189 | 1.2240 | 5 | 0.0919 | 0.0892 | 0.6614 | 0.0507 |
| Average | | | | | 0.1052 | 0.1046 | 0.5999 | 0.1024 |

**Table S9. Comparison of mean squared errors (MSE) for each QTN effect among pKWmEB, KWmEB, GEMMA and mrMLM methods in the fourth simulation experiment**

| QTN | True parameter | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 0.0770 | 0.0762 | 0.1990 | 0.0618 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 0.0519 | 0.0522 | 0.4231 | 0.0446 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 0.0441 | 0.0472 | 0.4015 | 0.0632 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 0.0987 | 0.1051 | 0.1831 | 0.0708 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 0.0489 | 0.0419 | 0.7207 | 0.0748 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 0.0553 | 0.0535 | 0.4329 | 0.0519 |
| Average | | | | | 0.0627 | 0.0627 | 0.3933 | 0.0612 |

**Table S10. Comparison of mean squared errors (MSE) for each QTN effect among pKWmEB, KWmEB, GEMMA and mrMLM methods in the fifth simulation experiment**

| QTN | True parameter | | | | MSE | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 0.0865 | 0.0899 | 0.2596 | 0.0699 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 0.0569 | 0.0608 | 0.6509 | 0.0682 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 0.0639 | 0.0644 | 0.6296 | 0.1140 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 0.1204 | 0.1227 | 0.2001 | 0.0875 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 0.0743 | 0.0696 | 0.9173 | 0.1504 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 0.0553 | 0.0524 | 0.5518 | 0.0702 |
| Average | | | | | 0.0762 | 0.0766 | 0.5349 | 0.0934 |

**Table S11. Average estimate of each QTN effect across 1000 replicates using pKWmEB, KWmEB, GEMMA and mrMLM methods in the first simulation experiment**

| QTN | True parameter | | | | Effect | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 1.3877 | 1.3792 | 1.9474 | 1.5412 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 1.0379 | 1.0401 | 1.8344 | 1.2083 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 1.1519 | 1.1601 | 1.8674 | 1.3291 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 1.6662 | 1.6635 | 2.1589 | 1.7922 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 1.1671 | 1.1636 | 1.9861 | 1.3178 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 1.0276 | 1.0302 | 1.7837 | 1.2072 |

**Table S12. Average estimate of each QTN effect across 1000 replicates using pKWmEB, KWmEB, GEMMA and mrMLM methods in the second simulation experiment**

| QTN | True parameter | | | | Effect | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.6171 | 10 | 1.4201 | 1.4164 | 2.0249 | 1.5845 |
| 2 | 1 | 11655607 | 1.1435 | 5 | 1.0548 | 1.0500 | 1.9674 | 1.2283 |
| 3 | 2 | 5066968 | 1.1435 | 5 | 1.1420 | 1.1536 | 1.9481 | 1.3408 |
| 4 | 2 | 5134228 | 1.9806 | 15 | 1.9583 | 1.9586 | 2.4673 | 2.0872 |
| 5 | 2 | 5464675 | 1.1435 | 5 | 1.1410 | 1.1307 | 2.1039 | 1.3282 |
| 6 | 2 | 6137189 | 1.1435 | 5 | 1.1397 | 1.1406 | 1.8970 | 1.3339 |

**Table S13. Average estimate of each QTN effect across 1000 replicates using pKWmEB, KWmEB, GEMMA and mrMLM methods in the third simulation experiment**

| QTN | True parameter | | | | Effect | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.7310 | 10 | 1.5624 | 1.5650 | 2.1624 | 1.7479 |
| 2 | 1 | 11655607 | 1.2240 | 5 | 1.1000 | 1.0941 | 2.0730 | 1.3319 |
| 3 | 2 | 5066968 | 1.2240 | 5 | 1.2908 | 1.3039 | 2.0657 | 1.5137 |
| 4 | 2 | 5134228 | 2.1201 | 15 | 1.9290 | 1.9399 | 2.5026 | 2.1022 |
| 5 | 2 | 5464675 | 1.2240 | 5 | 1.2661 | 1.2792 | 2.2239 | 1.5209 |
| 6 | 2 | 6137189 | 1.2240 | 5 | 1.0515 | 1.0505 | 2.0237 | 1.2886 |

**Table S14. Average estimate of each QTN effect across 1000 replicates using pKWmEB, KWmEB, GEMMA and mrMLM methods in the fourth simulation experiment**

| QTN | True parameter | | | | Effect | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 1.3674 | 1.3458 | 1.8476 | 1.4850 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 1.0055 | 1.0050 | 1.6763 | 1.0904 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 1.0350 | 1.0207 | 1.6569 | 1.1464 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 1.6814 | 1.6557 | 2.1617 | 1.7634 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 1.0618 | 1.0326 | 1.8724 | 1.1571 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 1.0128 | 0.9961 | 1.6694 | 1.1134 |

**Table S15. Average estimate of each QTN effect across 1000 replicates using pKWmEB, KWmEB, GEMMA and mrMLM methods in the fifth simulation experiment**

| QTN | True parameter | | | | Effect | | | |
|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | pKWmEB | KWmEB | GEMMA | mrMLM |
| 1 | 1 | 11298364 | 1.4762 | 10 | 1.3811 | 1.3832 | 1.9327 | 1.5278 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 1.0157 | 1.0206 | 1.8340 | 1.2004 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 1.1153 | 1.1151 | 1.8255 | 1.2878 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 1.6818 | 1.6808 | 2.1814 | 1.8129 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 1.1606 | 1.1466 | 1.9850 | 1.3381 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 1.0257 | 1.0313 | 1.7730 | 1.2113 |

**Table S16. Paired *t* tests and their P-values for statistical power and mean squared error (MSE) between pKWmEB and other methods in the second simulation experiment**

| Case | | pKWmEB & KWmEB | pKWmEB & KWsBC | pKWmEB & GEMMA | pKWmEB & mrMLM |
|---|---|---|---|---|---|
| Power | t-value | 1.73 | 0.66 | 3.62 | 1.25 |
| | P-value | 0.1447 | 0.5411 | 0.0152* | 0.2675 |
| MSE | t-value | -0.71 | - | -4.03 | -0.22 |
| | P-value | 0.5105 | - | 0.0101* | 0.8369 |

* and **: significances at the 0.05 and 0.01 levels, respectively.

**Table S17. Paired *t* tests and their P-values for statistical power and mean squared error (MSE) between pKWmEB and other methods in the third simulation experiment**

| Case | | pKWmEB & KWmEB | pKWmEB & KWsBC | pKWmEB & GEMMA | pKWmEB & mrMLM |
|---|---|---|---|---|---|
| Power | t-value | 1.91 | 0.39 | 3.46 | 1.89 |
| | P-value | 0.1150 | 0.7095 | 0.0181* | 0.1179 |
| MSE | t-value | 0.36 | - | -3.44 | 0.11 |
| | P-value | 0.7334 | - | 0.0185* | 0.9180 |

**Table S18. Paired t tests and their P-values for statistical power and mean squared error (MSE) between pKWmEB and other methods in the fourth simulation experiment**

| Case | | pKWmEB & KWmEB | pKWmEB & KWsBC | pKWmEB & GEMMA | pKWmEB & mrMLM |
|---|---|---|---|---|---|
| Power | t-value | 1.76 | -0.26 | 3.54 | -2.38 |
| | P-value | 0.1384 | 0.8033 | 0.0166* | 0.0635 |
| MSE | t-value | -0.02 | - | -3.82 | 0.18 |
| | P-value | 0.9863 | - | 0.0124* | 0.8676 |

**Table S19. Paired t tests and their P-values for statistical power and mean squared error (MSE) between pKWmEB and other methods in the fifth simulation experiment**

| Case | | pKWmEB & KWmEB | pKWmEB & KWsBC | pKWmEB & GEMMA | pKWmEB & mrMLM |
|---|---|---|---|---|---|
| Power | t-value | 2.28 | 0.49 | 3.51 | 0.54 |
| | P-value | 0.0715 | 0.6423 | 0.0171* | 0.6142 |
| MSE | t-value | -0.29 | - | -3.95 | -1.03 |
| | P-value | 0.7832 | - | 0.0108* | 0.3483 |

**Table S20. Effect of heterozygote genotypes (5%, 15% and 25%) on statistical powers, mean squared errors and false positive rate (FPR) using pKWmEB in the first simulation experiment**

| QTN | True parameter | | | | 5% | | | 15% | | | 25% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | Power (%) | Effect | MSE | Power (%) | Effect | MSE | Power (%) | Effect | MSE |
| 1 | 1 | 11298364 | 1.4762 | 10 | 97.9 | 1.4226 | 0.0790 | 98.4 | 1.4296 | 0.0789 | 98.3 | 1.4165 | 0.0793 |
| 2 | 1 | 11655607 | 1.0438 | 5 | 67.6 | 1.0208 | 0.0510 | 65.9 | 1.0182 | 0.0517 | 67.6 | 1.0148 | 0.0507 |
| 3 | 2 | 5066968 | 1.0438 | 5 | 57.7 | 1.1455 | 0.0658 | 57.5 | 1.1560 | 0.0780 | 66.0 | 1.1385 | 0.0701 |
| 4 | 2 | 5134228 | 1.8080 | 15 | 99.9 | 1.7344 | 0.0874 | 99.8 | 1.7374 | 0.0870 | 99.7 | 1.7386 | 0.0851 |
| 5 | 2 | 5464675 | 1.0438 | 5 | 51.7 | 1.1515 | 0.0632 | 47.1 | 1.1503 | 0.0642 | 61.0 | 1.1309 | 0.0597 |
| 6 | 2 | 6137189 | 1.0438 | 5 | 66.8 | 1.0481 | 0.0463 | 60.8 | 1.0502 | 0.0469 | 65.5 | 1.0508 | 0.0472 |
| Average | | | | | 73.6 | | 0.0655 | 71.6 | | 0.0678 | 76.4 | | 0.0654 |
| False positive rate (FPR, %) | | | | | 2.80E-3 | | | 3.15E-3 | | | 1.76E-3 | | |

**Table S21. Paired t tests and their P-values for statistical power and mean squared error (MSE) between no heterozygote and having heterozygote (5%, 15% and 25%) in the first simulation experiment**

| Case | | 5% heterozygote | 15% heterozygote | 25% heterozygote |
|---|---|---|---|---|
| power | t-value | -2.16 | -0.88 | -1.84 |
| | P-value | 0.0831 | 0.4211 | 0.1259 |
| MSE | t-value | 3.23 | 2.63 | 2.99 |
| | P-value | 0.0232* | 0.0465* | 0.0303* |

**Table S22. Effect of mean and median as genotypic transformation thresholds on statistical powers, mean squared errors and false positive rate in the first simulation experiment**

| QTN | True parameter | | | | | Mean | | | | Median | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Chr. | Posi (bp) | Effect | $r^2$ (%) | | Power (%) | Effect | MSE | | Power (%) | Effect | MSE |
| 1 | 1 | 11298364 | 1.4762 | 10 | | 95.4 | 1.3877 | 0.0891 | | 94.5 | 1.4134 | 0.0832 |
| 2 | 1 | 11655607 | 1.0438 | 5 | | 66.7 | 1.0379 | 0.0564 | | 61.4 | 1.0536 | 0.0562 |
| 3 | 2 | 5066968 | 1.0438 | 5 | | 52.5 | 1.1519 | 0.0851 | | 51.1 | 1.1751 | 0.0816 |
| 4 | 2 | 5134228 | 1.8080 | 15 | | 96.3 | 1.6662 | 0.1198 | | 95.3 | 1.6937 | 0.1104 |
| 5 | 2 | 5464675 | 1.0438 | 5 | | 40.3 | 1.1671 | 0.0790 | | 36.3 | 1.1854 | 0.0827 |
| 6 | 2 | 6137189 | 1.0438 | 5 | | 67.7 | 1.0276 | 0.0490 | | 65.6 | 1.0429 | 0.0503 |
| Average | | | | | | 69.8 | | 0.0797 | | 67.4 | | 0.0774 |
| False positive rate (FPR, %) | | | | | | 3.18E-2 | | | | 2.40E-2 | | |

**Table S23. The numbers of SNPs significantly associated with four flowering time related traits in *Arabidopsis thaliana* and the number of genes around these SNPs identified by pKWmEB, KWmEB, mrMLM and GEMMA methods**

| Trait | No. of SNPs significantly associated with the trait | | | | No. of the known genes detected | | | |
|---|---|---|---|---|---|---|---|---|
| | pKWmEB | KWmEB | mrMLM | GEMMA | pKWmEB | KWmEB | mrMLM | GEMMA |
| FLC | 17 | 23 | 21 | 1 | 7 | 5 | 2 | 0 |
| FRI | 15 | 15 | 8 | 34 | 4 | 3 | 3 | 4 |
| FT Field | 28 | 19 | 4 | 17 | 6 | 3 | 0 | 0 |
| FT GH | 20 | 20 | 23 | 1 | 6 | 5 | 5 | 1 |
| Total | 80 | 77 | 56 | 53 | 23 | 16 | 10 | 5 |

FLC: FLC gene expression; FRI: FRI gene expression; FT Field: Days to flowering of plants grown in the field; FT GH: days to flowering growth in greenhouse

**Table S24. GWAS for four flowering time related traits in *Arabidopsis thaliana* using pKWmEB, KWmEB, mrMLM and GEMMA methods**

| Trait | Gene[§] | | | pKWmEB | | | KWmEB | | | mrMLM | | | GEMMA | | | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ID | Chr | Position (bp) | LOD | Effect | $r^2$(%) | LOD | Effect | $r^2$(%) | LOD | Effect | $r^2$(%) | P-value | Effect | $r^2$(%) | |
| FLC | ATBZIP27 | 2 | 7720804 | 4.54 | -0.185 | 3.08 | | | | | | | | | | Jaeger *et al.*, 2013 |
| | VGD1 | 2 | 19344951 | 9.42 | -0.307 | 6.23 | 4.87 | -0.200 | 3.37 | 5.11 | -0.219 | 3.12 | | | | Jiang *et al.*, 2005 |
| | ATMYB65 | 3 | 3617387 | 4.75 | -0.232 | 2.03 | 5.58 | -0.224 | 2.43 | | | | | | | Achard et al., 2004 |
| | ETC3 | 4 | 454422/<br>454542 | 6.82 | -0.169 | 6.46 | 4.40 | -0.148 | 2.34 | 9.37 | -0.241 | 4.85 | | | | Tominaga-Wada *et al.*, 2013 |
| | DFL2 | 4 | 1487946 | 8.34 | 0.279 | 5.05 | | | | | | | | | | |
| | AT4G33280 | 4 | 16056698 | | | | 5.35 | -0.170 | 2.98 | | | | | | | |
| | APETALA2 | 4 | 17392527 | 3.05 | -0.183 | 2.03 | | | | | | | | | | Huang *et al.*, 2016 |
| | SAG12 | 5 | 18623805/<br>186228833 | 13.99 | -0.314 | 9.22 | 9.58 | -0.220 | 5.81 | | | | | | | |
| FRI | ARPN | 2 | 844399 | 6.99 | -0.120 | 3.41 | 8.76 | -0.134 | 4.78 | | | | | | | Chae and Lord, 2011 |
| | CYP77A6 | 3 | 3297055 | | | | | | | 5.34 | -0.180 | 7.86 | | | | Li-Beisson *et al.*, 2009 |
| | KLCR2 | 3 | 10368451 | 4.35 | 0.075 | 1.86 | | | | | | | | | | |
| | CRP | 4 | 224997 | | | | | | | | | | 4.06E-8 | 0.321 | 2.93 | Imura *et al.*, 2012 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FLA | 4 | 268990 | 6.01 | 0.144 | 31.58 | 9.07 | 0.174 | 30.65 | 14.27 | 0.356 | 29.20 | 1.39E-16 | 0.431 | 6.31 | Sanchez-Bermejo and Balasubramanian , 2016 |
| | AT4G00690/ sim to ESD4 | 4 | 268990 | 6.01 | 0.144 | 31.58 | 9.07 | 0.174 | 30.65 | 14.27 | 0.356 | 29.20 | 1.39E-16 | 0.431 | 6.31 | Murtas *et al.*, 2003 |
| | MGP4 | 4 | 497146 | | | | | | | | | | 2.38E-8 | -0.402 | 3.14 | |
| FT Field | KNAT6 | 1 | 8312819/<br>8313345 | 4.24 | 0.059 | 0.33 | 9.28 | 0.023 | 0.47 | | | | | | | Tabata *et al.*, 2010 |
| | ATCOL3 | 2 | 10574932 | 3.23 | 0.098 | 1.38 | | | | | | | | | | Izawa et al., 2003 |
| | ATCYP71 | 3 | 16161334 | 4.04 | 0.099 | 1.35 | | | | | | | | | | |
| | LD | 4 | 1135406 | | | | 3.98 | -0.097 | 1.54 | | | | | | | Jin *et al.*, 2008 |
| | TSF | 4 | 10987195 | | | | 13.89 | 0.223 | 9.53 | | | | | | | Ando *et al.*, 2013 |
| | ATCOL1 | 5 | 5195272 | 5.34 | -0.132 | 2.87 | | | | | | | | | | Simon *et al.*, 2015 |
| | ANAC089 | 5 | 7372523 | 3.96 | 0.122 | 1.86 | | | | | | | | | | Li *et al.*, 2010 |
| | ATTIP49A | 5 | 7372523 | 3.96 | 0.122 | 1.86 | | | | | | | | | | Holt *et al.*, 2002 |
| FT GH | ATARP4 | 1 | 6371569 | 3.54 | -0.186 | 9.18 | 11.51 | -0.146 | 6.95 | 15.73 | -0.188 | 13.19 | 9.09E-8 | -0.297 | 7.87 | Kandasamy *et al.*, 2005 |
| | PRK2 | 2 | 2916675 | 4.90 | 0.062 | 0.92 | | | | | | | | | | Zhao *et al.*, 2013 |
| | UBP13 | 3 | 3774664 | | | | | | | 7.90 | -0.090 | 4.55 | | | | Zhang *et al.*, 2013 |
| | ATGATA18 | 3 | 18922377/<br>18928630 | 8.47 | -0.101 | 1.90 | 4.38 | -0.061 | 1.43 | 3.57 | -0.113 | 3.69 | | | | |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | PIP5K4 | 3 | 21079518 | 6.31 | -0.080 | 1.38 | 5.39 | -0.075 | 2.51 | | | | | | | Ischebeck *et al.*, 2008 |
| | FLA | 4 | 260906 | | | | 7.16 | -0.094 | 3.74 | | | | | | | Sanchez-Bermejo and Balasubramanian, 2016 |
| | ETC3 | 4 | 458226/<br>468387 | 6.61 | -0.149 | 2.03 | 5.22 | -0.065 | 2.33 | 3.34 | -0.082 | 2.73 | | | | Tominaga-Wada *et al.*, 2013 |
| | AT4G03070 | 4 | 1368585 | | | | | | | 3.20 | 0.067 | 2.15 | | | | |
| | SPL7 | 5 | 6289819 | 5.67 | -0.062 | 0.97 | | | | | | | | | | |

§: All the genes had been identified by previously biological studies and the related references were listed in the last column.

## Reference

Achard P, Herr A, Baulcombe DC, Harberd NP (2004). Modulation of floral development by a gibberellin-regulated microRNA. *Development* **131**: 3357-3365.

Ando E, Ohnishi M, Wang Y, Matsushita T, Watanabe A, Hayashi Y *et al.* (2013). TWIN SISTER OF FT, GIGANTEA, and CONSTANS have a positive but indirect effect on blue light-induced stomatal opening in Arabidopsis. *Plant Physiol.* **162**: 1529-1538.

Chae K, Lord EM (2011). Pollen tube growth and guidance: roles of small, secreted proteins. *Ann. Bot.* **108**: 627-636.

Holt BF, Boyes DC, Ellerström M, Siefers N, Wiig A, Kauffman S *et al.* (2002). An evolutionarily conserved mediator of plant disease resistance gene function is required for normal Arabidopsis development. *Dev. Cell* **2**: 807-817.

Huang Z, Shi T, Zheng B, Yumul RE, Liu X, You C *et al.* (2016). APETALA2 antagonizes the transcriptional activity of AGAMOUS in regulating floral stem cells in Arabidopsis thaliana. *New Phytol.* DOI: 10.1111/nph.14151.

Imura Y, Kobayashi Y, Yamamoto S, Furutani M, Tasaka M, Abe M *et al*. (2012) CRYPTIC PRECOCIOUS/MED12 is a novel flowering regulator with multiple target steps in Arabidopsis. *Plant Cell Physiol*. **53**: 287-303.

Ischebeck T, Stenzel I, Heilmann I (2008). Type B phosphatidylinositol-4-phosphate 5-kinases mediate Arabidopsis and Nicotiana tabacum pollen tube growth by regulating apical pectin secretion. *Plant Cell* **20**: 3312-3320.

Izawa T, Takahashi Y, Yano M (2003). Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and Arabidopsis. *Curr. Opin. Plant. Biol.* **6**: 113-120.

Jaeger KE, Pullen N, Lamzin S, Morris RJ, Wigge PA (2013). Interlocking feedback loops govern the dynamic behavior of the floral transition in Arabidopsis. *Plant Cell* **25**: 820-833.

Jiang L, Yang SL, Xie LF, Puah CS, Zhang XQ, Yang WC *et al*. (2005). VANGUARD1 encodes a pectin methylesterase that enhances pollen tube growth in the Arabidopsis style and transmitting tract. *Plant Cell* **17**: 584-596.

Jin JB, Jin YH, Lee JY, Miura KJ, Yoo CY, Kim WY *et al.* (2008). The SUMO E3 ligase, AtSIZ1, regulates flowering by controlling a salicylic acid-mediated floral promotion pathway and through affecting on FLC chromatin structure. *Plant J.* **53**: 530-540.

Kandasamy MK, Deal RB, McKinney EC, Meagher RB (2005). Silencing the nuclear actin-related protein AtARP4 in Arabidopsis has multiple effects on plant development, including early flowering and delayed floral senescence. *Plant J.* **41**: 845-858.

Li J, Zhang J, Wang X, Chen J (2010). A membrane-tethered transcription factor ANAC089 negatively regulates floral initiation in Arabidopsis thaliana. *Sci. China Life Sci.* **53**: 1299-1306.

Li-Beisson Y, Pollard M, Sauveplane V, Pinot F, Ohlrogge J, Beisson F (2009). Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. *Proc. Natl. Acad. Sci. USA.* **106**: 22008-22013.

Murtas G, Reeves PH, Fu YF, Bancroft I, Dean C, Coupland G (2003). A nuclear protease required for flowering-time regulation in Arabidopsis reduces the abundance of SUMO conjugates. *Plant Cell* **15**: 2308-2319.

Sanchez-Bermejo E, Balasubramanian S (2016). Natural variation involving deletion alleles of FRIGIDA modulate temperature-sensitive flowering responses in Arabidopsis thaliana. *Plant Cell Environ*. **39**: 1353-1365.

Simon S, Rühl M, de Montaigu A, Wötzel S, Coupland G (2015). Evolution of CONSTANS regulation and function after gene duplication produced a photoperiodic flowering switch in the Brassicaceae. *Mol. Bio. Evol.* **32**: 2284-2301.

Tabata R, Ikezaki M, Fujibe T, Aida M, Tian CE, Ueno Y *et al.* (2010). Arabidopsis auxin response factor6 and 8 regulate jasmonic acid biosynthesis and floral organ development via repression of class 1 KNOX genes. *Plant Cell Physiol.* **51**: 164-175.

Tominaga-Wada R, Nukumizu Y, Wada T (2013). Flowering is delayed by mutations in homologous genes CAPRICE and TRYPTICHON in the early flowering Arabidopsis cpl3 mutant. *Plant Physiol*. **170**: 1466-1468.

Zhang X, Zhou Y, Ding L, Wu Z, Liu R, Meyerowitz EM (2013). Transcription repressor HANABA TARANU controls flower development by integrating the actions of multiple hormones,floral organ specification genes, and GATA3 family genes in Arabidopsis. *Plant Cell* **25**: 83-101.

Zhao XY, Wang Q, Li S, Ge FR, Zhou LZ, McCormick S *et al.* (2013). The juxtamembrane and carboxy-terminal domains of Arabidopsis PRK2 are critical for ROP-induced growth in pollen tubes. *J. Exp. Bot.* **64**: 5599-5610.

**Table S25. Linkage disequilibrium score regression test for population structures of four flowering time related traits in *Arabidopsis thaliana***

| Trait | Intercept | t value | P value | Population structure in the model |
|---|---|---|---|---|
| FLC | -2.44 | -1.79 | 3.76E-2 | Yes |
| FRI | -3.84 | -2.90 | 2.12E-3 | Yes |
| FT Field | -3.75 | -2.26 | 1.25E-2 | Yes |
| FT GH | 10.85 | 9.26 | 5.31E-17 | Yes |

## Dataset S1

In the three simulation experiments (I, II and III), all the SNP genotypes were derived from the 216130 SNPs in Atwell *et al.* (2010). All the SNPs between 11226256 and 12038776 bp on Chr. 1, between 5045828 and 6412875 bp on Chr. 2, between 1916588 and 3196442 bp on Chr. 3, between 2232796 and 3143893 bp on Chr. 4, and between 19999868 and 21039406 bp on Chr. 5 were used to conduct simulation studies. All the three simulation experiments were based on the same genotypic dataset, which is composed of 10,000 SNPs (row) and 199 individuals (column).

In the three simulation experiments (I, II and III), all the phenotypic values were simulated based on QTN effects, genetic background and residual error. All the phenotypic values were listed in the **y1.csv**, **y2.csv** and **y3.csv** files for the **first**, **second** and **third** simulation experiments, respectively. In each file, the *i*th replication consists of the [(*i*-1)*199+1]th to [*i**199]th rows.

All the above datasets were included in **Dataset S1** and in the dryad link. The related information for the dryad link was as follows:

**Data package title**: Data from: pLARmEB: Integration of least angle regression with empirical Bayes for multi-locus genome-wide association studies

**DOI**: doi:10.5061/dryad.sk652

**Data files**: data

## Dataset S2

In the fourth and fifth simulation experiments, all the SNP genotypes were the same as those in Dataset S1. The phenotypic values in the fourth (log-normal distribution for residual error) and fifth (logistic distribution for residual error) simulation experiments were listed in the first and second columns of Dataset S2.csv, respectively. In each column of phenotypic values, the *i*th replication consists of the [(*i*-1)*199+1]th to [*i**199]th rows.

**Figure S1. Running times (hours) using pKWmEB, KWmEB, KWsBC, mrMLM and GEMMA methods in the first simulation experiment**

**Figure S2. The Venn Diagram for the numbers of previously reported genes for four flowering time related traits in *Arabidopsis thaliana* using pKWmEB, KWmEB, mrMLM and GEMMA. (a) FLC, (b) FRI, (c) FT Field, and (d) FT GH.**

**Figure S3. Effect of mean and median as the critical values of genotypic transformation on statistical powers (a), mean squared errors (b) and false positive rate (c) in the first simulation experiment**