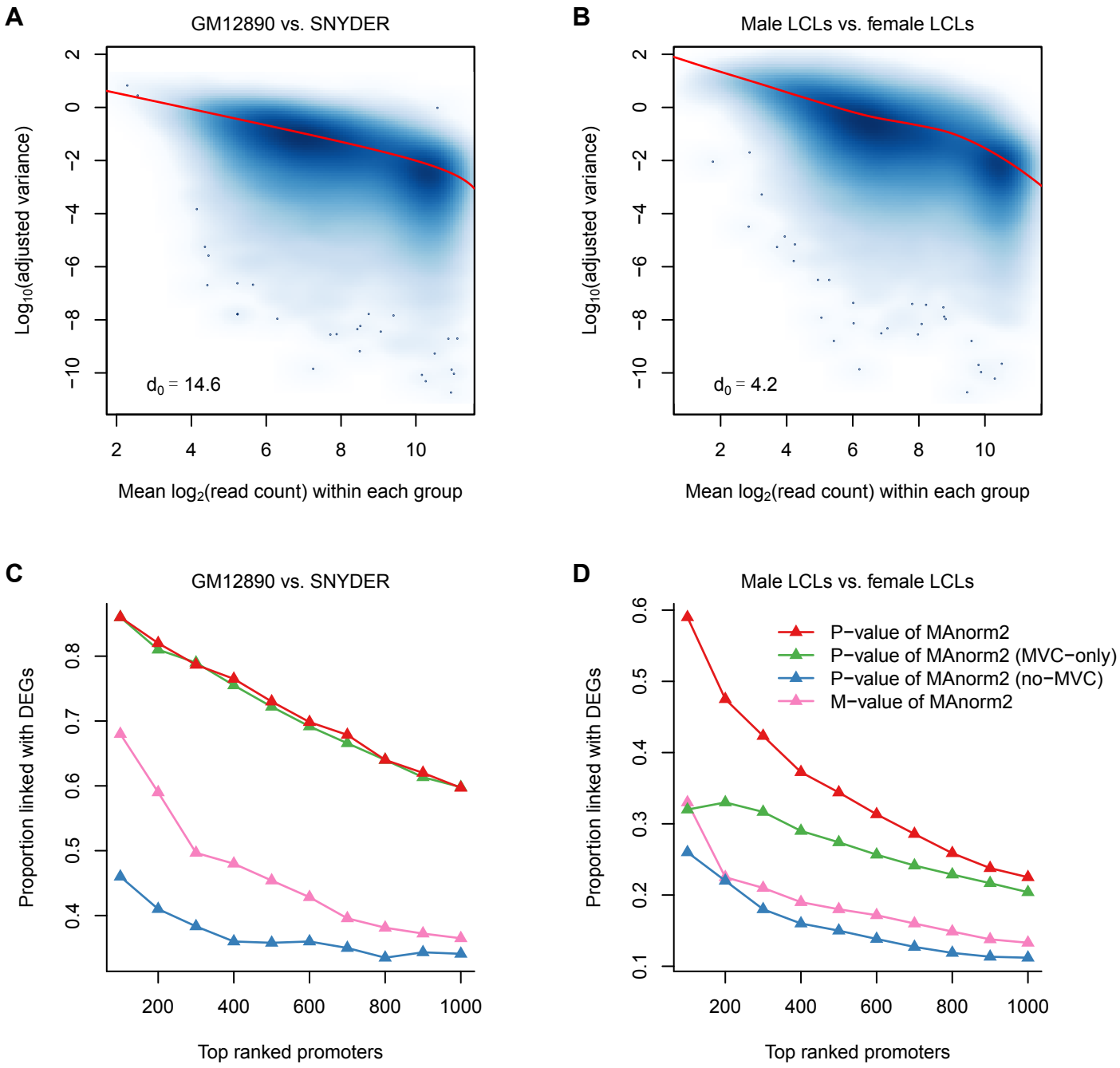
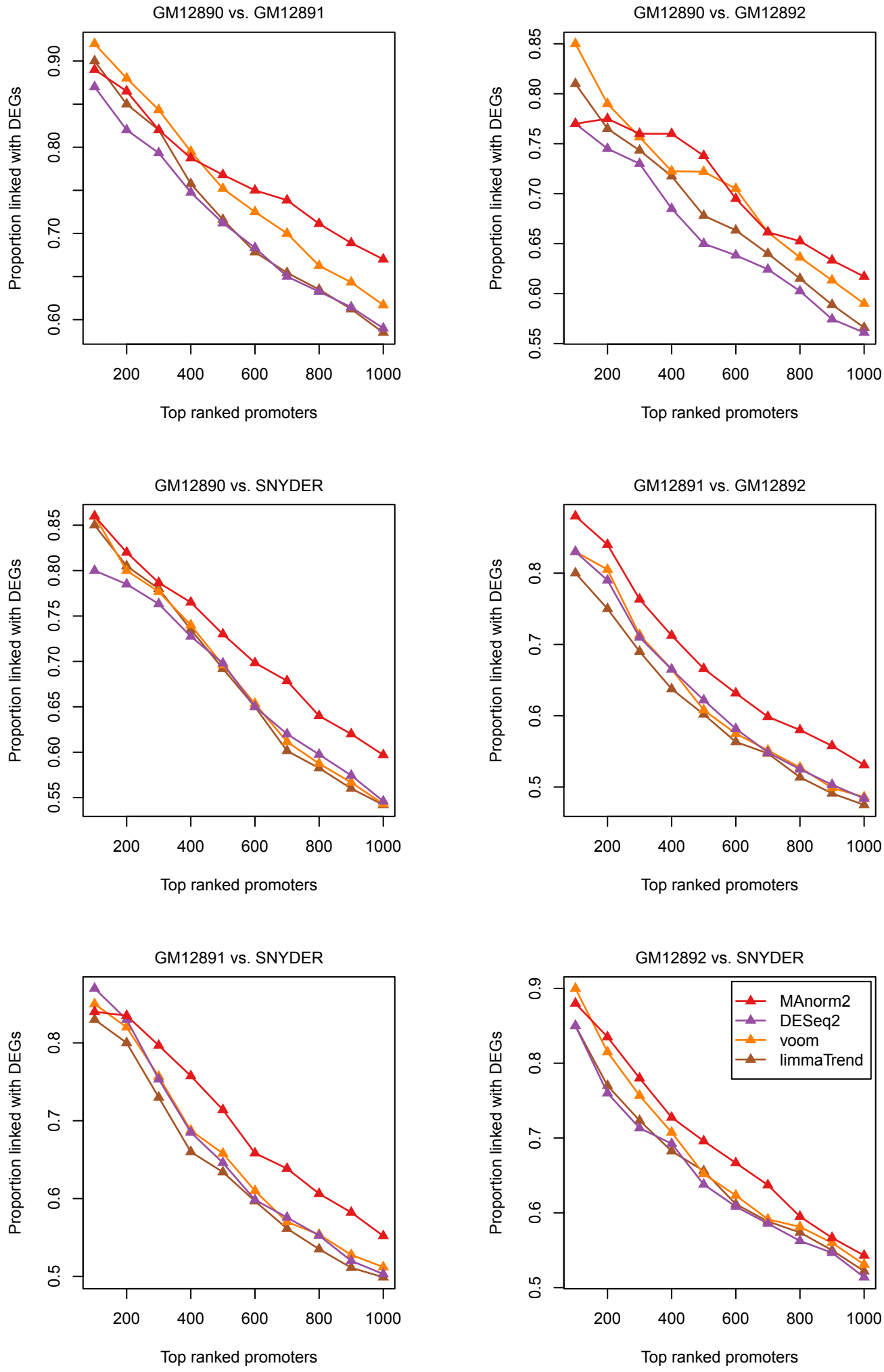


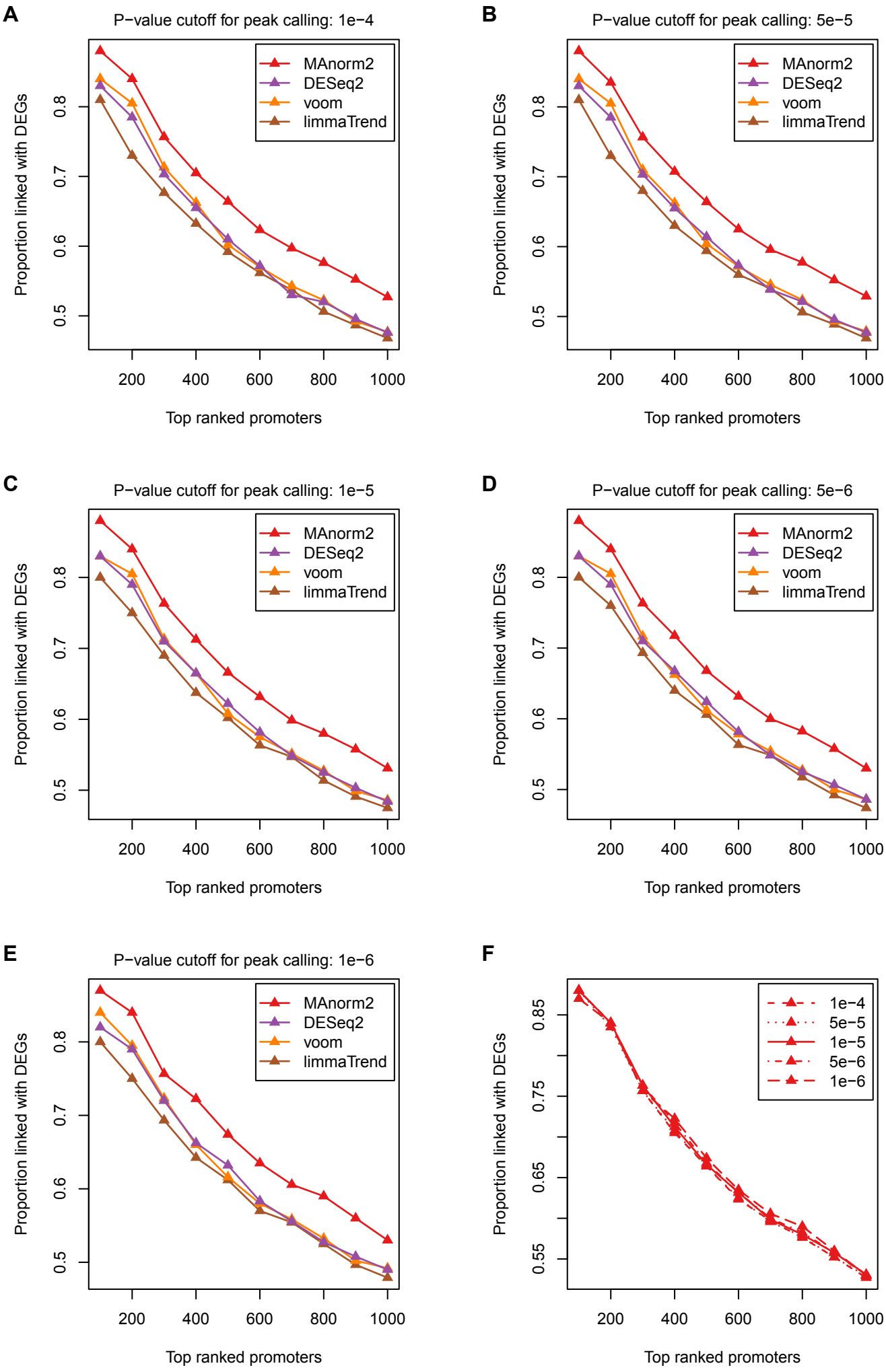
Supplemental Figure S1

Supplemental Figure S1. Being adaptive to the complexity of variance structure across different scenarios. (A) Scatter plot showing the mean-variance trend associated with the comparison of H3K4me3 levels between GM12890 and SNYDER lymphoblastoid cell lines (LCLs). Red line depicts the fitted mean-variance curve (MVC), and d_0 gives the estimated number of prior degrees of freedom. **(B)** Scatter plot for a comparison of H3K4me3 levels between two male LCLs (GM12891 and SNYDER) and two female LCLs (GM12890 and GM12892). Note that both comparisons are two-versus-two and, thus, the dispersion level of observed mean-variance pairs around MVC is comparable between the two cases. **(C, D)** For each differential analysis, the proportion of true discoveries among top ranked genomic intervals at gene promoters is plotted against the number of top ranked intervals. Here true discoveries are defined as intervals that are linked with differentially expressed genes (DEGs), which are identified by applying DESeq2 to the corresponding RNA-seq data with a *p*-value cutoff of 0.01.



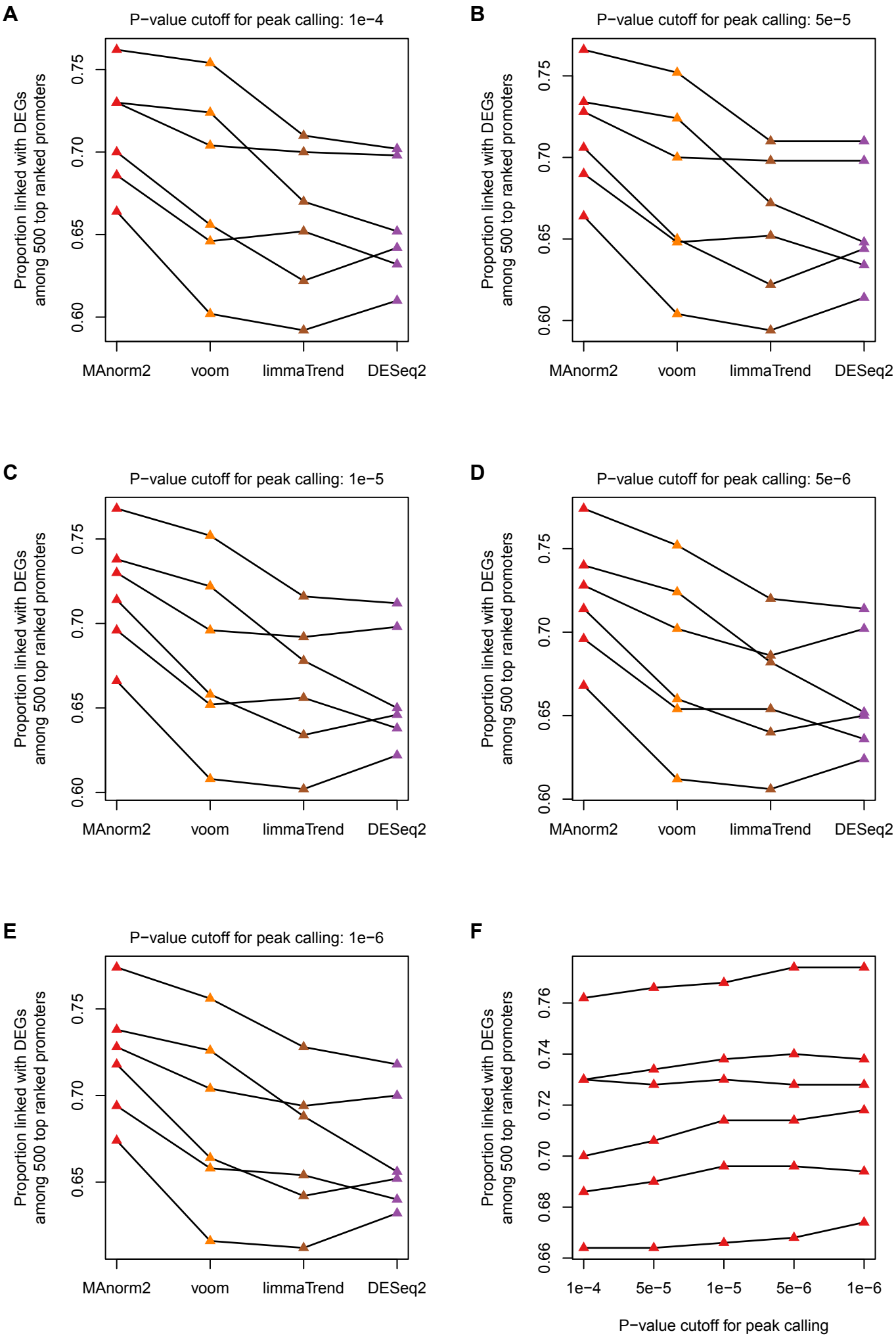
Supplemental Figure S2

Supplemental Figure S2. Comparing the performance of MAnorm2 with that of limma-trend, voom and DESeq2 in all pairwise comparisons of H3K4me3 levels among GM12890, GM12891, GM12892 and SNYDER LCLs. Each plot corresponds to an individual comparison between two LCLs.



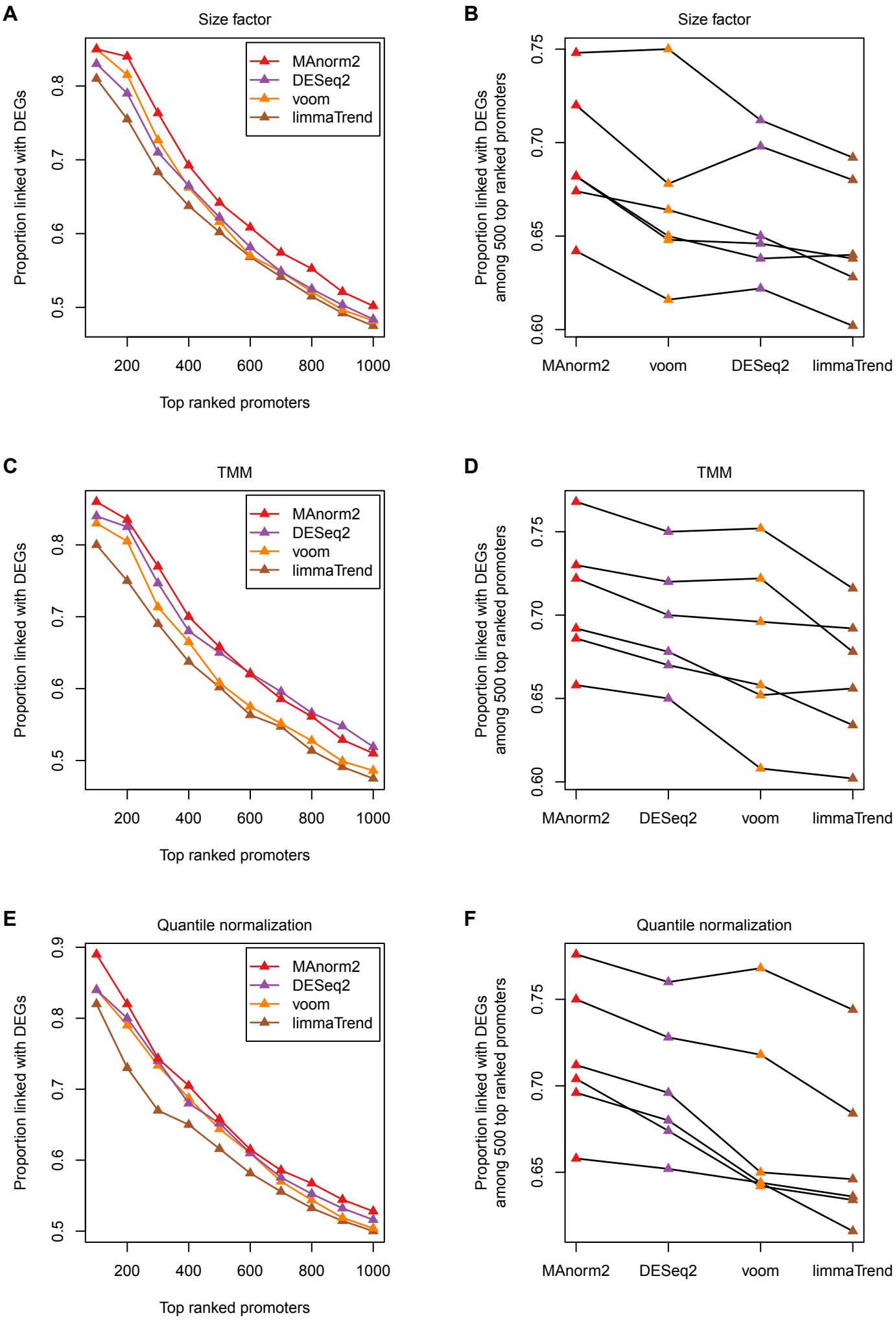
Supplemental Figure S3

Supplemental Figure S3. Method comparison in the differential analysis of H3K4me3 ChIP-seq data between GM12891 and GM12892, with various cutoffs for peak calling. (A-E) Each plot corresponds to a specific *p*-value cutoff for peak calling (performed by MACS 1.4, whose default *p*-value cutoff is 1e-5). **(F)** Comparing the performance of MAnorm2 across different *p*-value cutoffs for peak calling.



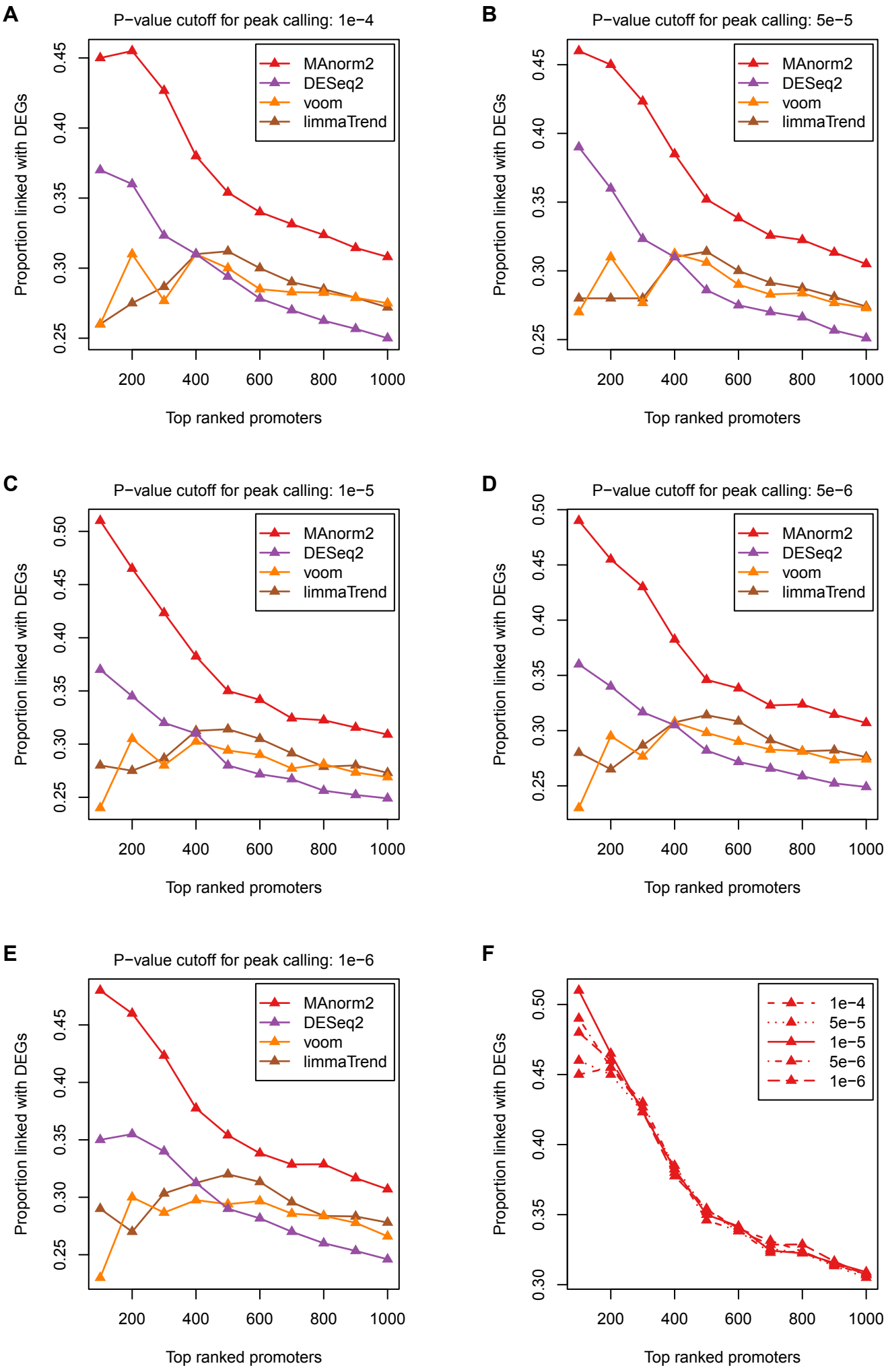
Supplemental Figure S4

Supplemental Figure S4. Method comparison in all the differential analyses of H3K4me3 ChIP-seq data between a pair of LCLs, with various cutoffs for peak calling. (A-E) Each plot corresponds to a specific *p*-value cutoff for peak calling (performed by MACS 1.4, whose default *p*-value cutoff is 1e-5). Each line in each plot corresponds to an individual differential analysis between a pair of LCLs. **(F)** Comparing the performance of MAnorm2 across different *p*-value cutoffs for peak calling.



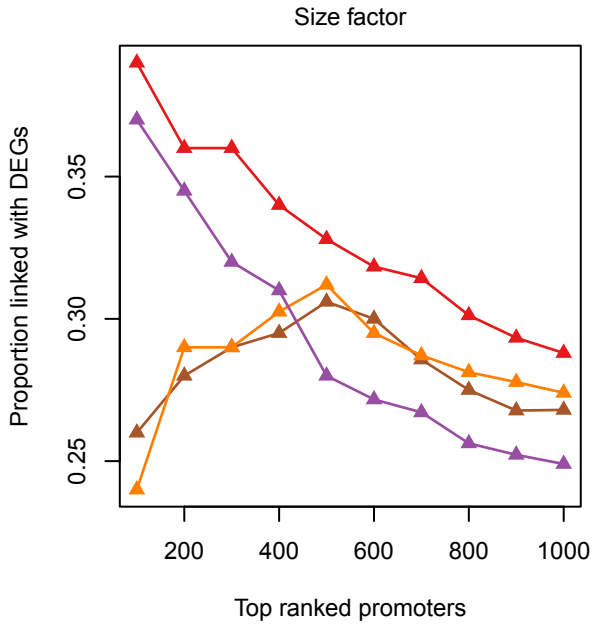
Supplemental Figure S5

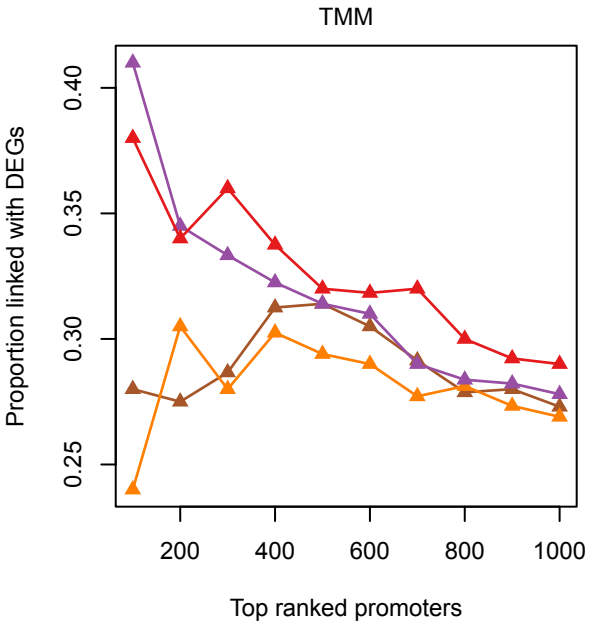
Supplemental Figure S5. In the pairwise comparisons of H3K4me3 levels among GM12890, GM12891, GM12892 and SNYDER, comparing the performance of MAnorm2 with that of limma-trend, voom and DESeq2 with unifying their normalization methods to each of size factor, TMM and quantile normalization methods. (A, B) Comparing the performance of the four methods with unifying their normalization methods to the size factor method. (A) is about the differential analysis between GM12891 and GM12892; (B) is about all pairwise comparisons among the four LCLs. **(C, D)** Similar to (A) and (B), except that the TMM method is used. **(E, F)** Similar to (A) and (B), except that the quantile normalization method is used. Note that differential analysis methods in (B), (D) and (F) have been sorted by the average DEG proportion (among 500 top ranked promoters) across all the pairwise comparisons.

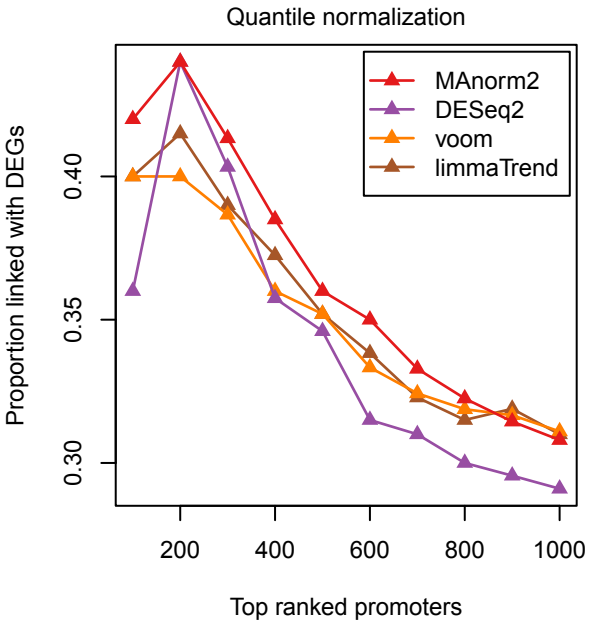


Supplemental Figure S6

Supplemental Figure S6. Method comparison in the differential analysis of H3K27ac ChIP-seq data between three LCLs and three chronic lymphocytic leukemia (CLL) cell lines, with various cutoffs for peak calling. (A-E) Each plot corresponds to a specific *p*-value cutoff for peak calling (performed by MACS 1.4, whose default *p*-value cutoff is 1e-5). **(F)** Comparing the performance of MAnorm2 across different *p*-value cutoffs for peak calling.

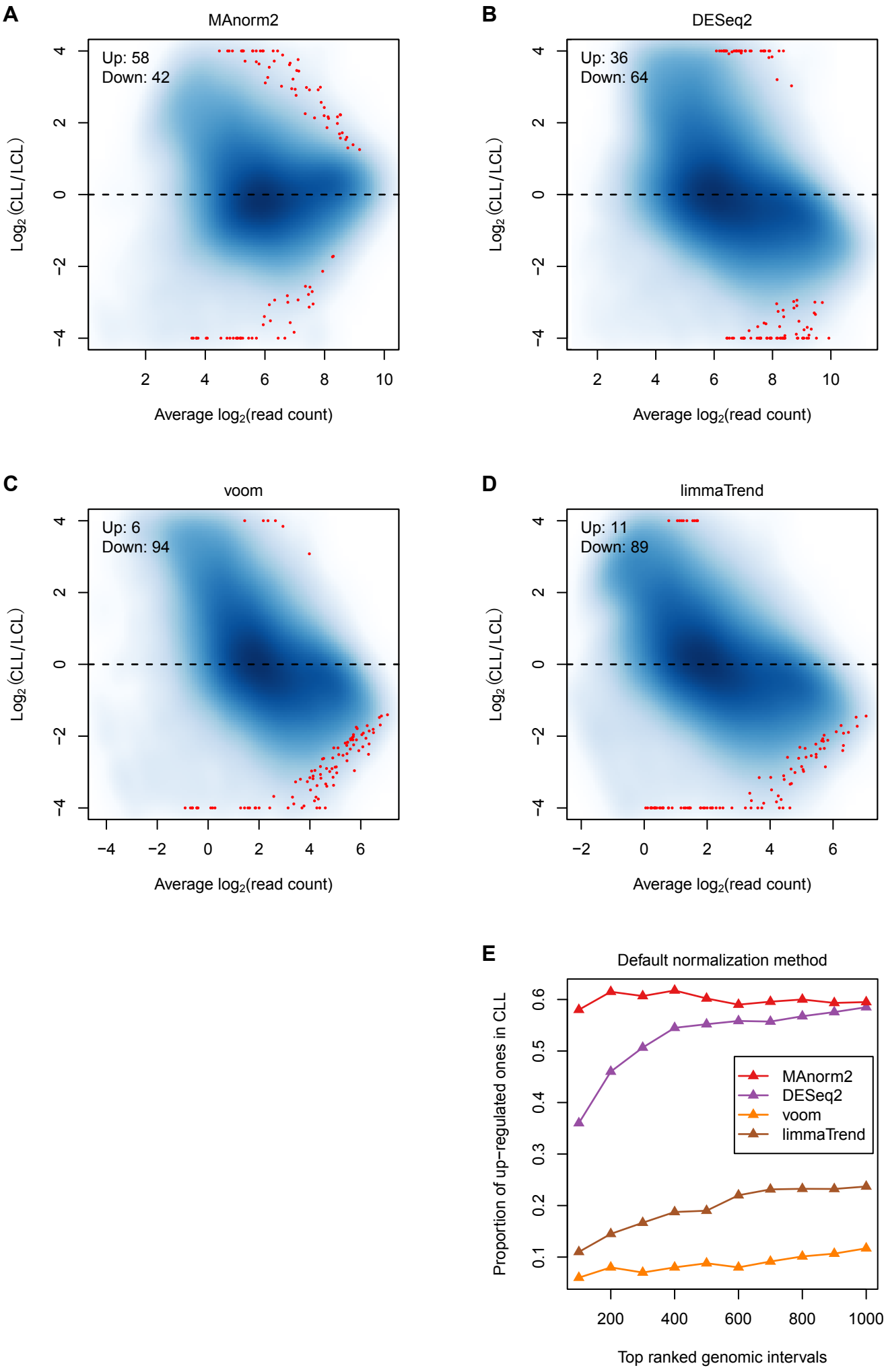






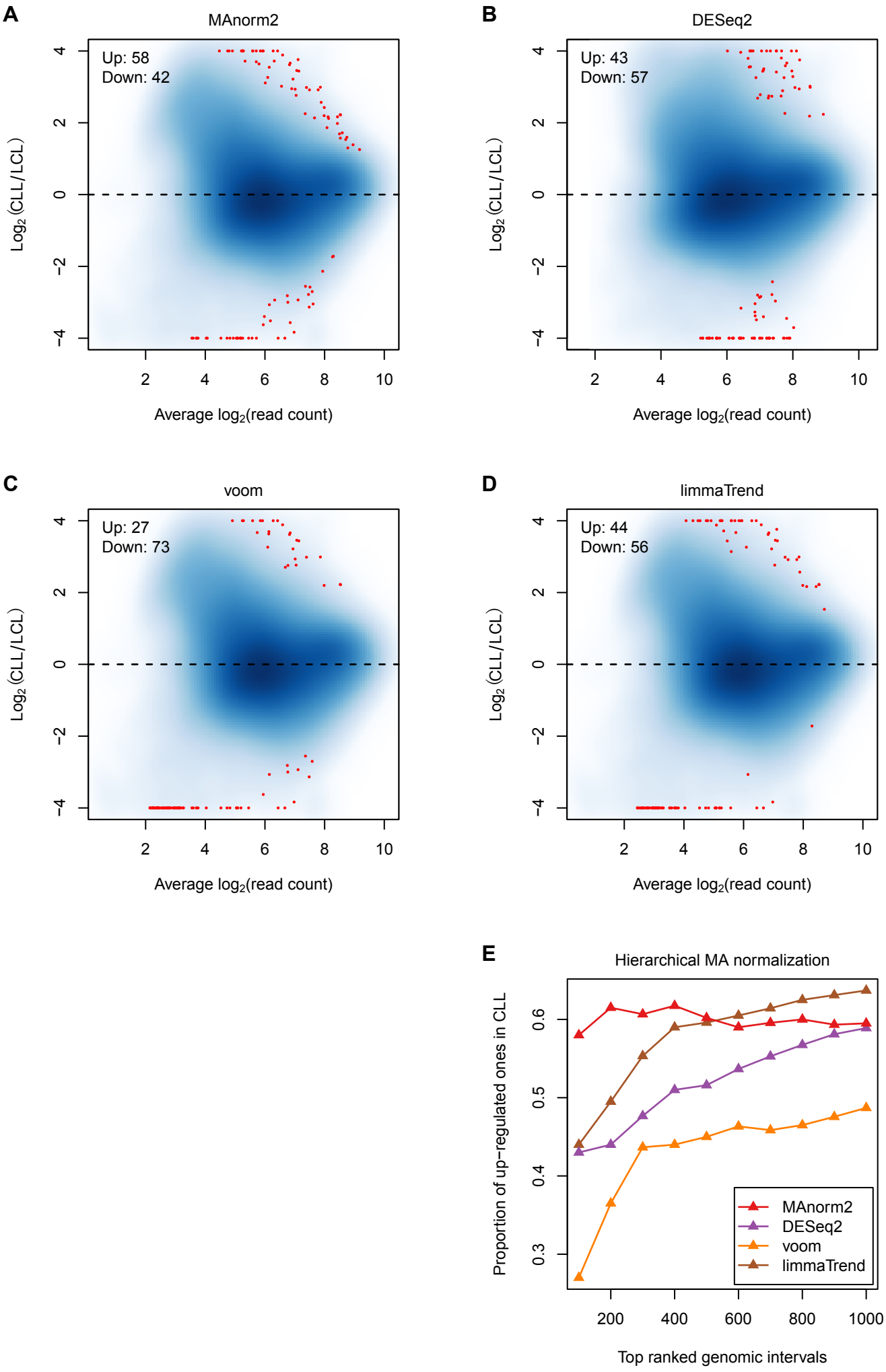
Supplemental Figure S7

Supplemental Figure S7. In the differential analysis of H3K27ac ChIP-seq data between LCLs and CLL cell lines, comparing the performance of MAnorm2 with that of limma-trend, voom and DESeq2 with unifying their normalization methods to each of size factor, TMM and quantile normalization methods.



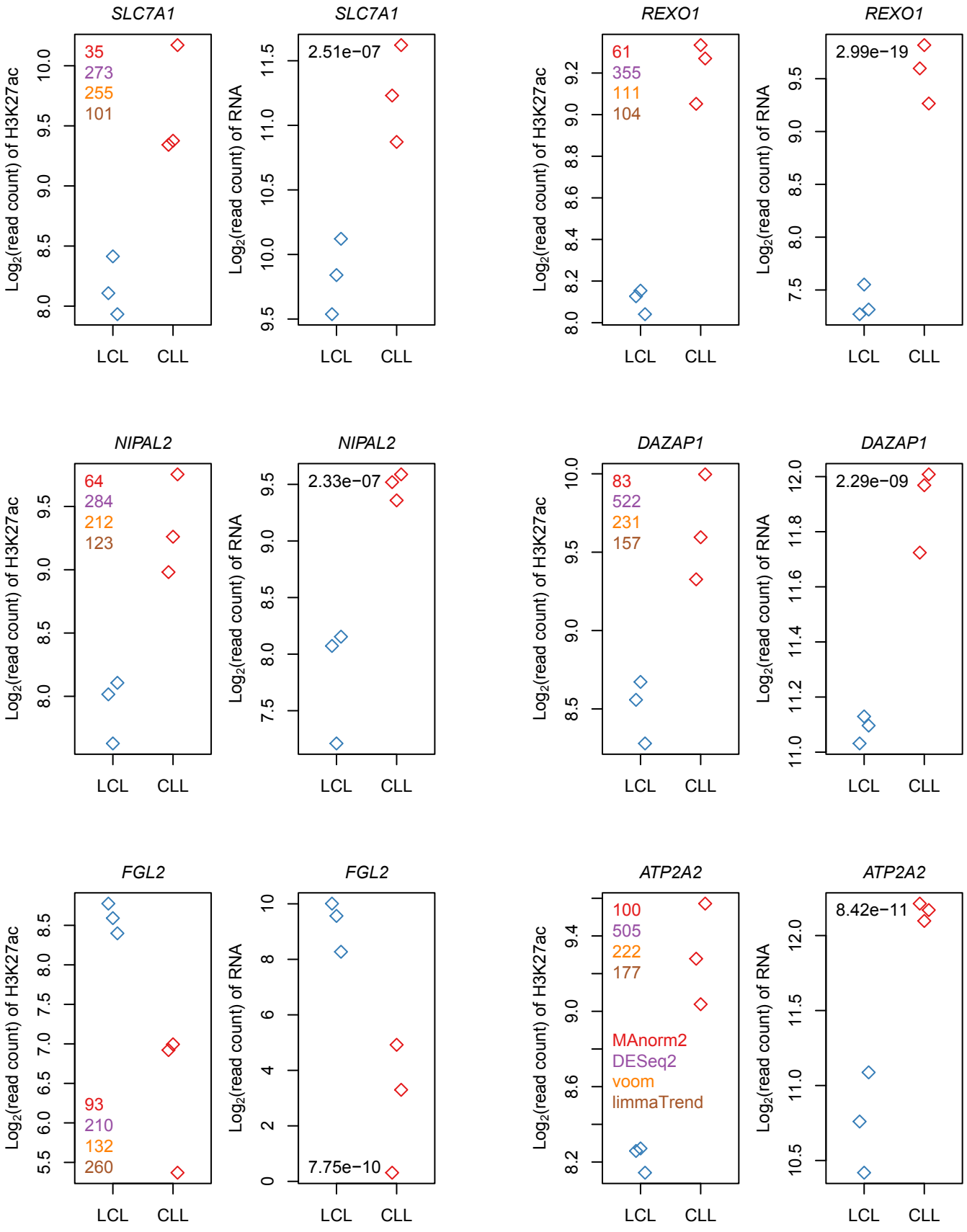
Supplemental Figure S8

Supplemental Figure S8. MA scatter plots for MAnorm2, DESeq2, voom and limma-trend with their respective default normalization methods, in the differential analysis of H3K27ac ChIP-seq data between LCLs and CLL cell lines. **(A-D)** Each MA scatter plot corresponds to a differential analysis method. In each plot, the 100 top ranked differential genomic intervals are marked by red points, among which the numbers of down- and up-regulated ones (in the CLL group) are given in the top-left corner. **(E)** For each differential analysis method, the proportion of up-regulated genomic intervals (in the CLL group) among top ranked ones is plotted against the number of top ranked intervals.



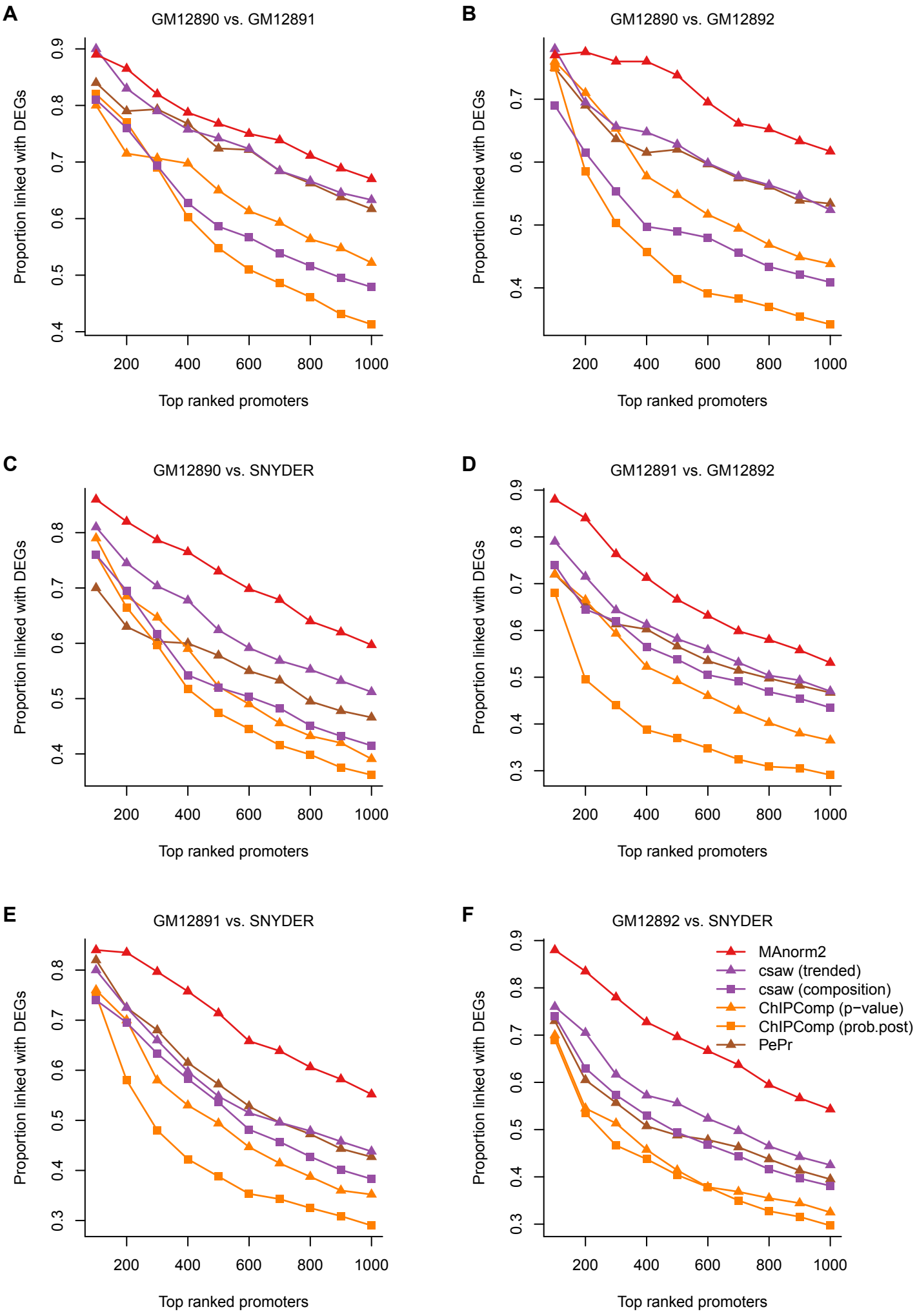
Supplemental Figure S9

Supplemental Figure S9. MA scatter plots for MAnorm2, DESeq2, voom and limma-trend with the hierarchical MA normalization method, in the differential analysis of H3K27ac ChIP-seq data between LCLs and CLL cell lines.



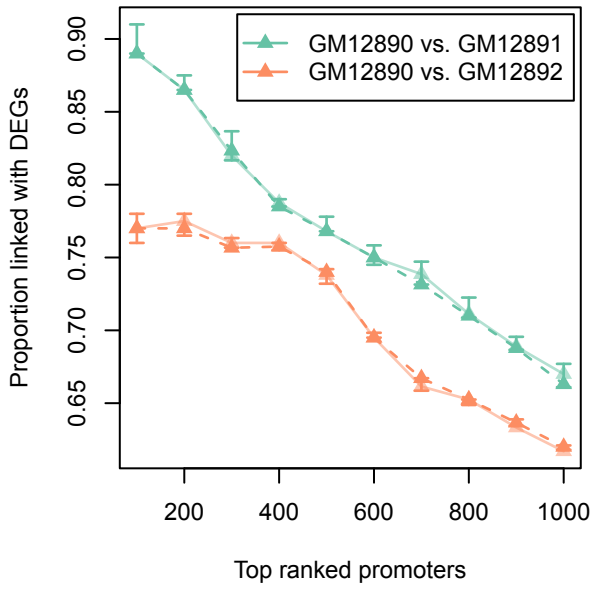
Supplemental Figure S10

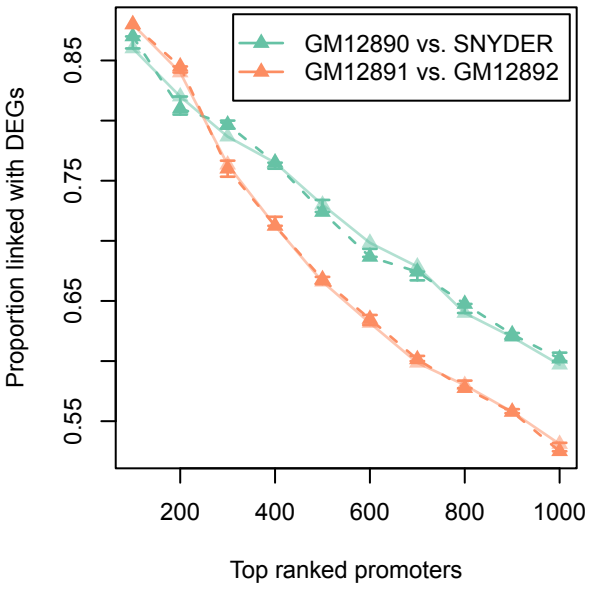
Supplemental Figure S10. Example promoter intervals that are linked with DEGs between the three LCLs and the three CLL cell lines. For each example promoter interval, we give normalized H3K27ac ChIP-seq signals (derived by hierarchical MA normalization) for the six cell lines as well as the ranks among all promoter intervals derived by different methods (applied to the hierarchical MA normalization results as well). For each associated gene, we give normalized RNA-seq signals (derived by DESeq2) as well as the DESeq2 *p*-value for assessing the significance of differential expression between the two conditions.

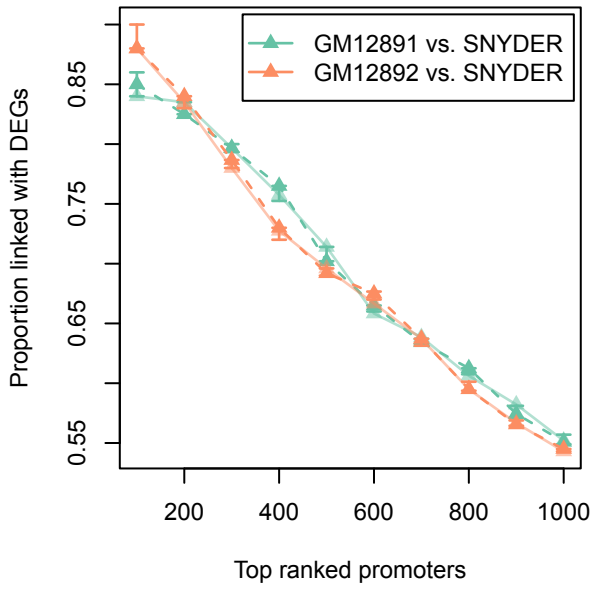


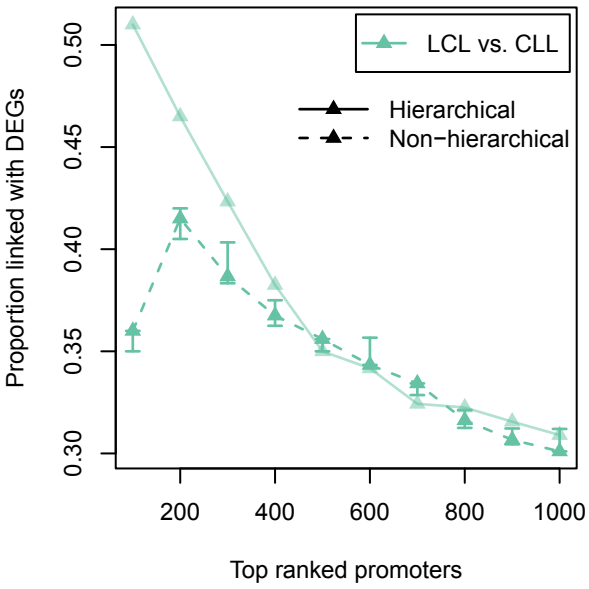
Supplemental Figure S11

Supplemental Figure S11. Comparing MAnorm2 with other tools for differential ChIP-seq analysis in the pairwise comparisons of H3K4me3 levels among GM12890, GM12891, GM12892 and SNYDER. Each plot corresponds to an individual comparison between two LCLs.



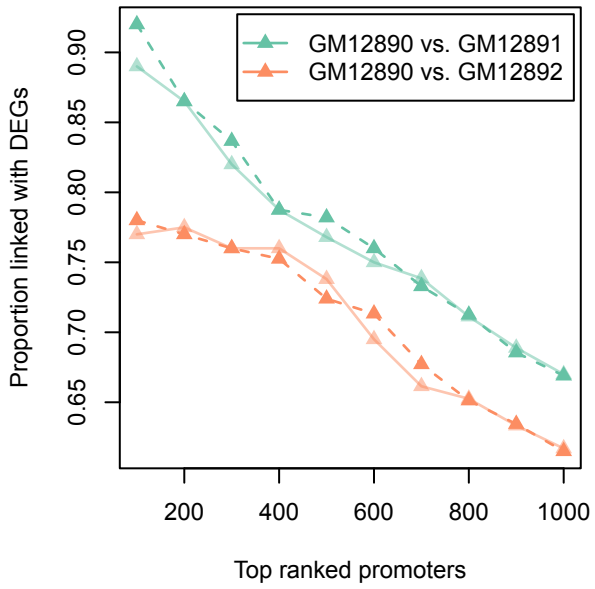


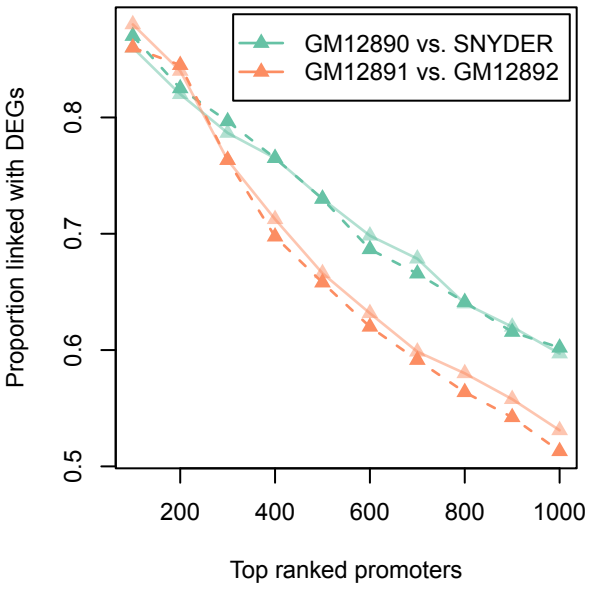


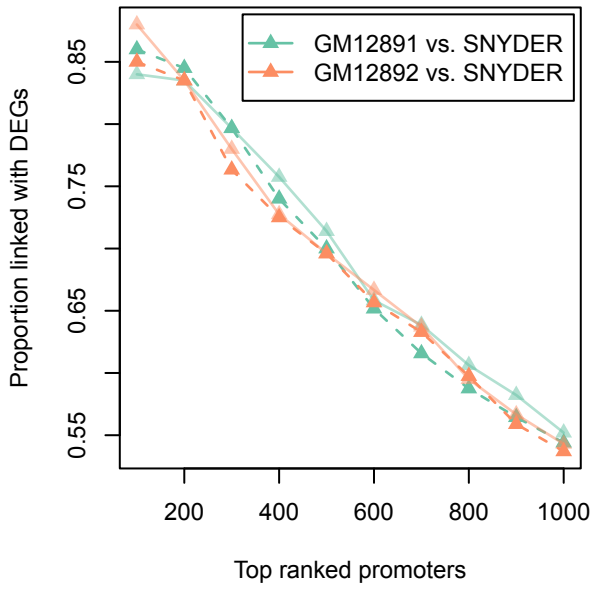


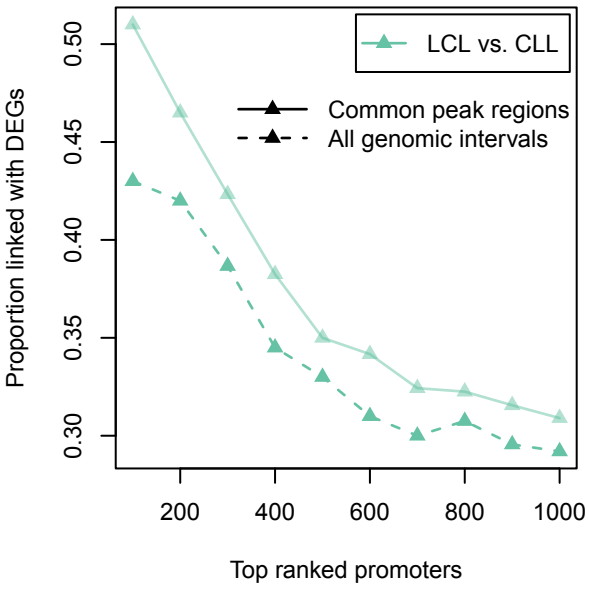
Supplemental Figure S12

Supplemental Figure S12. Comparing the performance of hierarchical and non-hierarchical MA normalization in the pairwise comparisons of H3K4me3 levels among LCLs and the differential analysis of H3K27ac ChIP-seq data between LCLs and CLL cell lines. For each differential analysis, MAnorm2 has been applied to normalization results that are separately derived by each of the two methods. Each curve for the non-hierarchical method results from an application of it to the corresponding ChIP-seq samples with default baseline selection (see Methods in the main text), and the associated error bars give the variations across all possible baselines. For the 60 proportions of true discoveries associated with the pairwise comparisons among LCLs, hierarchical and non-hierarchical methods (with default baseline selection) outperform the other in 22 and 23 cases, respectively (the remaining 15 cases are ties).



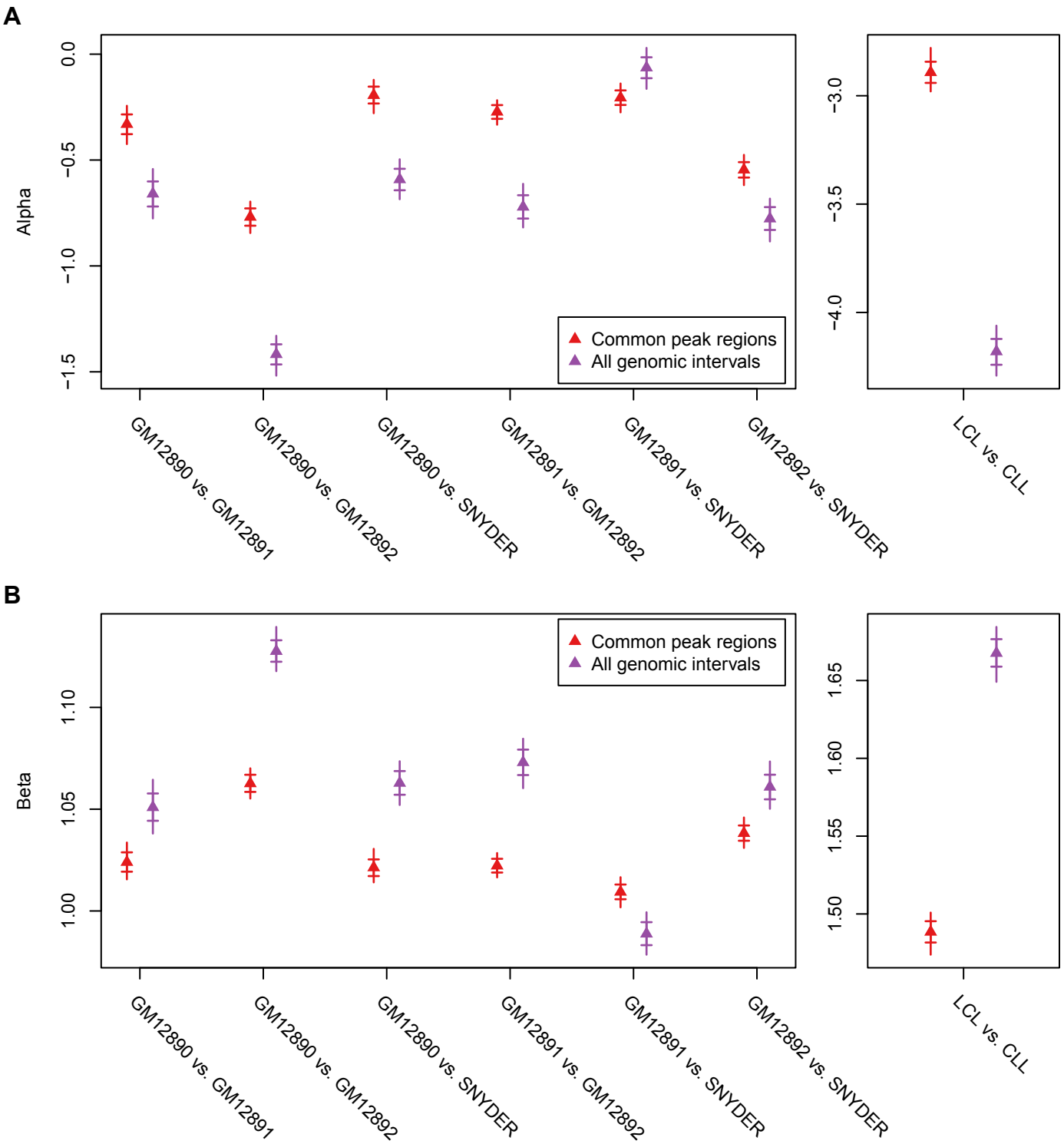






Supplemental Figure S13

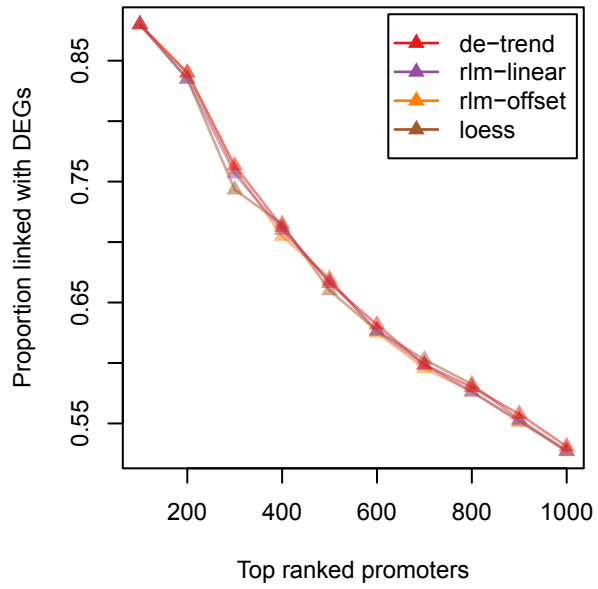
Supplemental Figure S13. Comparing hierarchical MA normalization with a variant of it that deduces the linear transformations for normalization based on all genomic intervals rather than common peak regions. For each differential analysis, MAnorm2 has been applied to normalization results that are separately derived by each of the two methods. For the 60 proportions of true discoveries associated with the pairwise comparisons among LCLs, the original method and the variant outperform the other in 33 and 18 cases, respectively (the remaining 9 cases are ties).



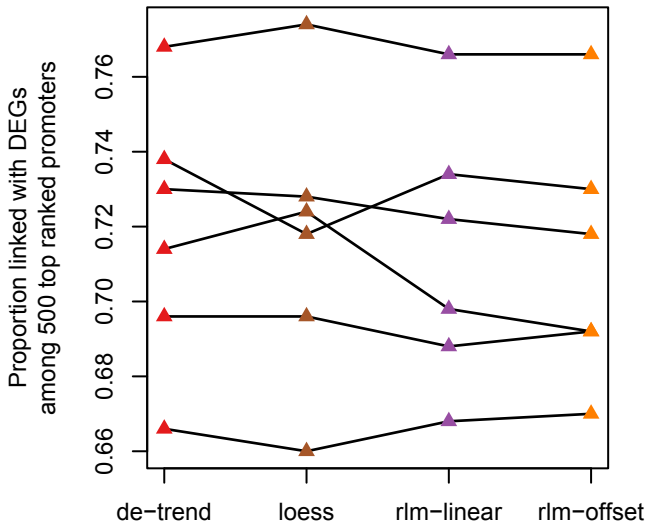
Supplemental Figure S14

Supplemental Figure S14. Comparing the uncertainty of coefficient estimates between hierarchical MA normalization and the variant of it that deduces coefficient estimates based on all genomic intervals. Results shown here are about **(A)** intercept and **(B)** slope estimates for between-group normalizations. Here, for each between-group normalization, the baseline group is selected such that the slope estimate deduced from common peak regions is always larger than 1. For each coefficient estimate (denoted by a triangle symbol), the associated uncertainty is assessed by the bootstrap method with 10,000 times of resampling, and we show here the variation across all resamplings as well as the 95% bootstrap percentile interval.

A



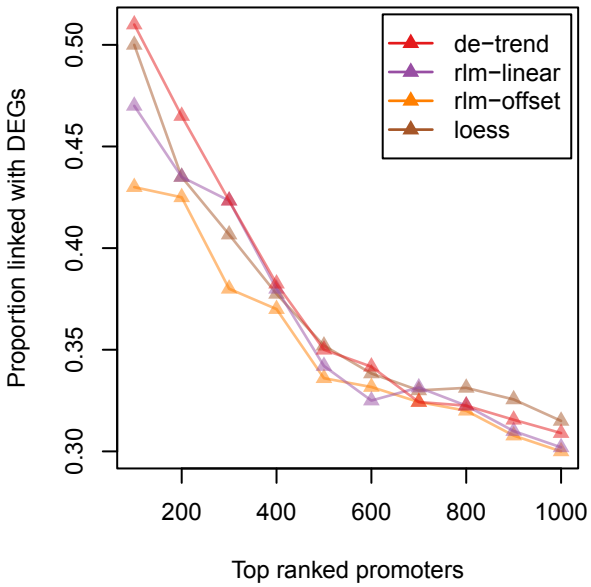
B



C

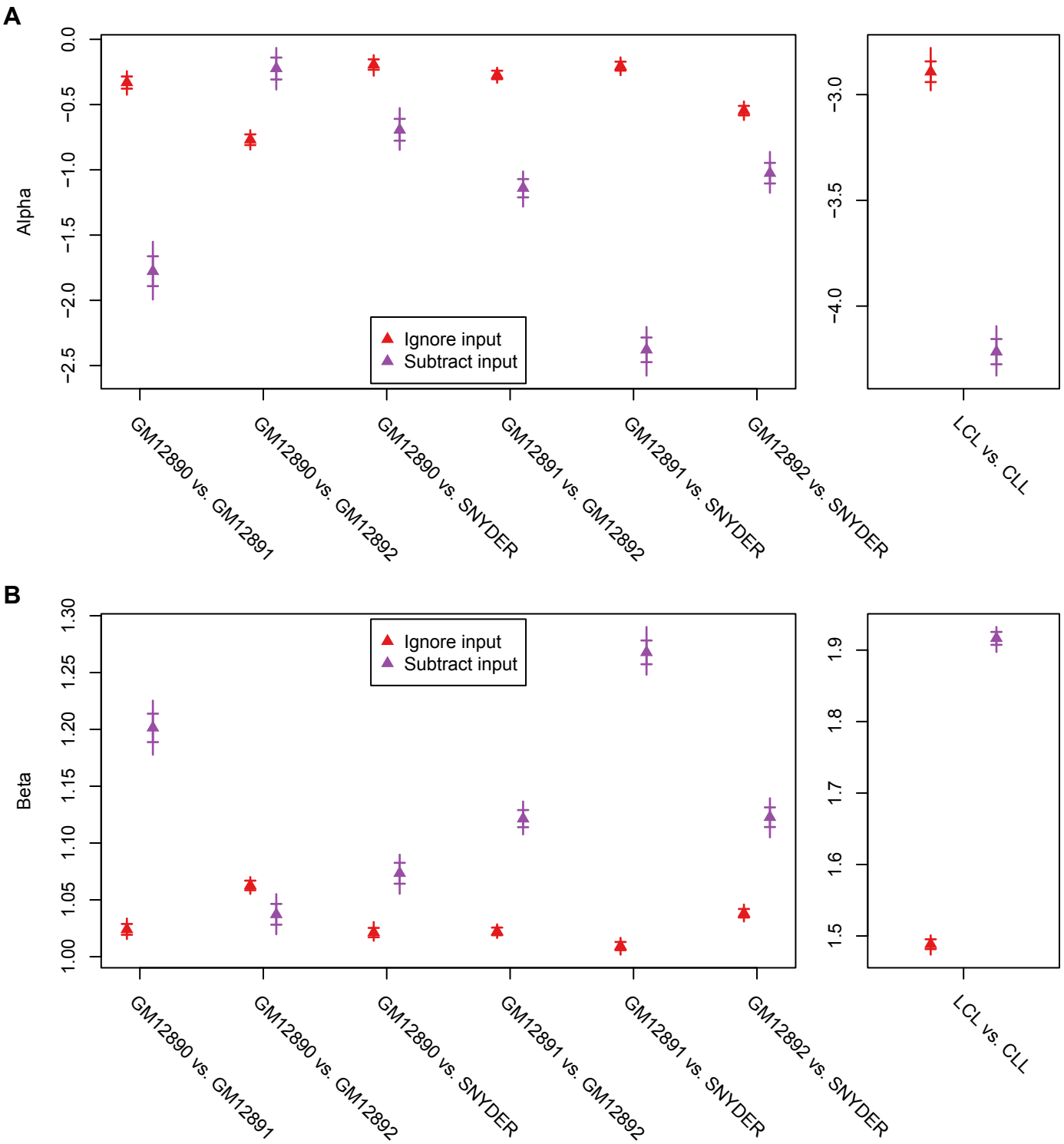
	Win	Lose	Tie
rlm-linear	8	46	6
rlm-offset	11	41	8
loess	24	29	7

D



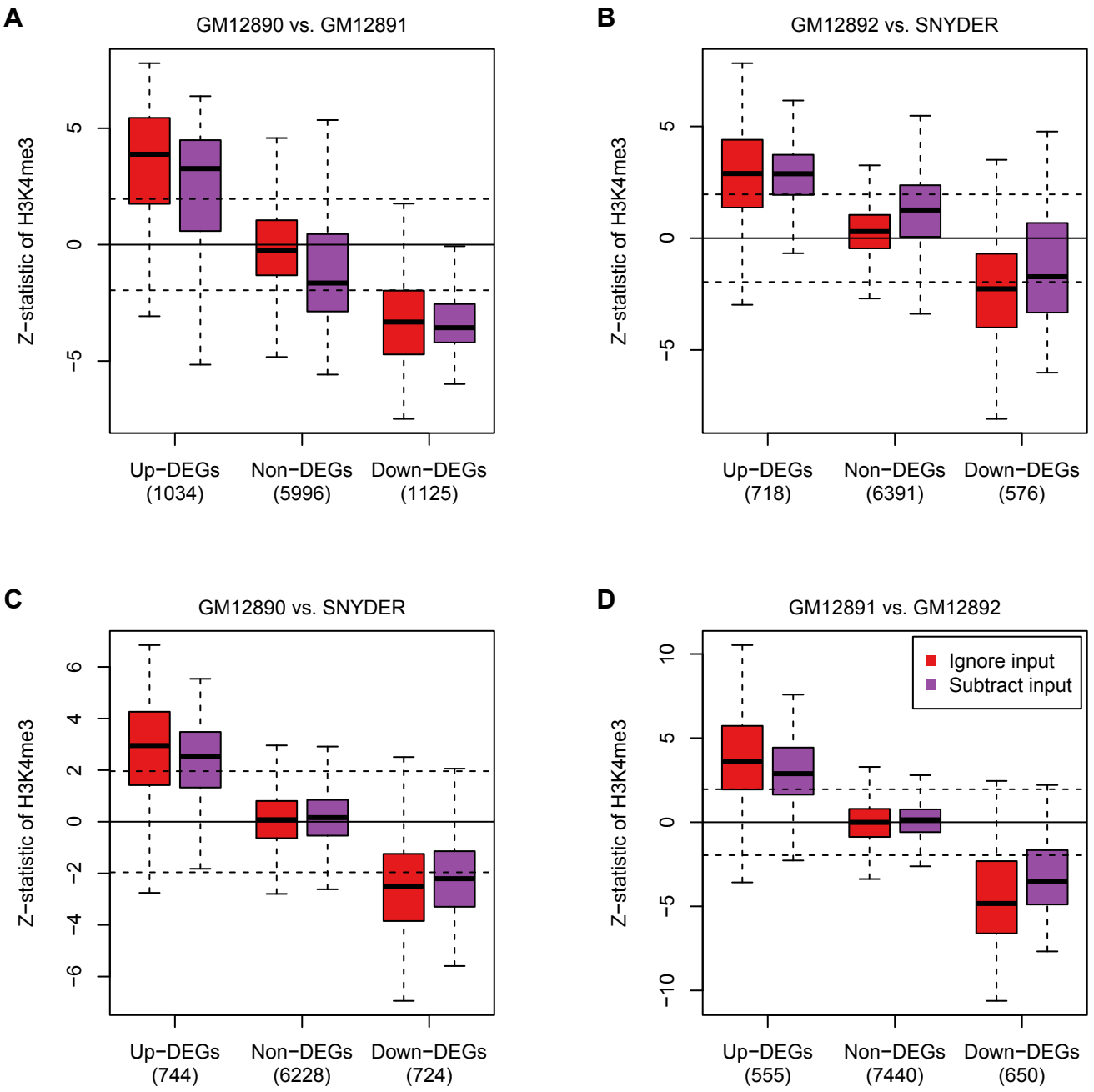
Supplemental Figure S15

Supplemental Figure S15. Comparing hierarchical MA normalization with two variants of it that fit M-A trend by using LOESS (local polynomial regression) and robust linear regression respectively. There are two approaches for integrating the latter variant into the hierarchical normalization framework, which are referred to as rlm-offset and rlm-linear (Supplemental Note S3). The original method is referred to as de-trend. **(A)** Method comparison in the differential analysis of H3K4me3 ChIP-seq data between GM12891 and GM12892. **(B)** Method comparison in all pairwise comparisons of H3K4me3 levels among GM12890, GM12891, GM12892 and SNYDER. Each line corresponds to an individual comparison between two LCLs, and the methods are sorted by the average true discovery proportion (among 500 top ranked promoter intervals) across all comparisons. **(C)** Summary statistics regarding the relative performance of each variant compared to the original method, for the 60 proportions of true discoveries associated with the pairwise comparisons among LCLs. **(D)** Method comparison in the differential analysis of H3K27ac ChIP-seq data between LCLs and CLL cell lines.



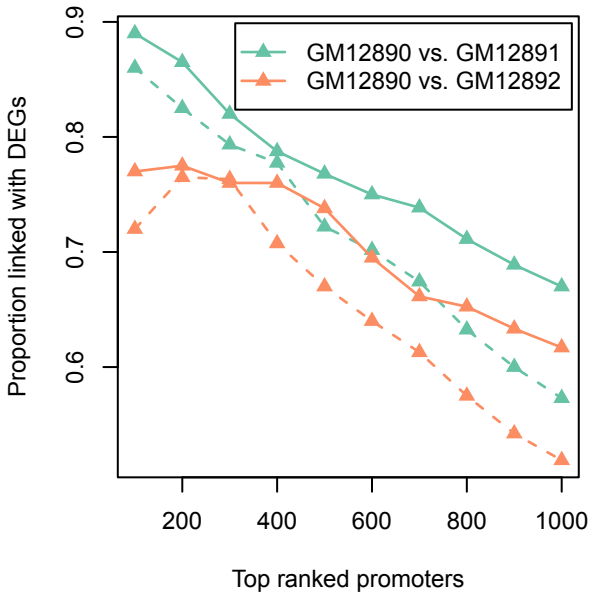
Supplemental Figure S16

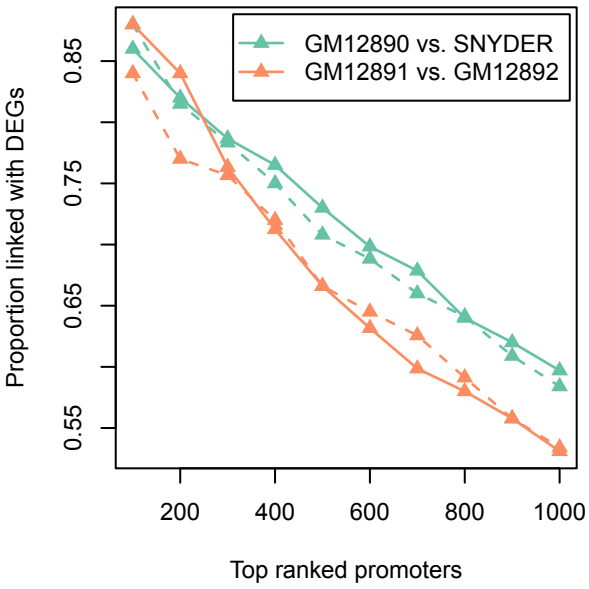
Supplemental Figure S16. Assessing the uncertainty of coefficient estimates for normalization after subtracting read counts from input samples. Results shown here are about **(A)** intercept and **(B)** slope estimates for between-group normalizations, for which baseline groups have been selected such that all slope estimates are larger than 1 (note that we still deduce the coefficient estimates based on common peak regions). For each coefficient estimate (denoted by a triangle symbol), the associated uncertainty is assessed by the bootstrap method with 10,000 times of resampling, and we show here the variation across all resamplings as well as the 95% bootstrap percentile interval.

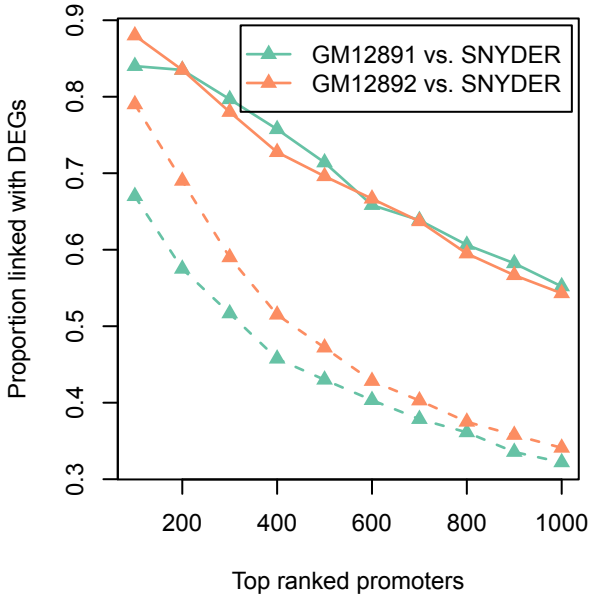


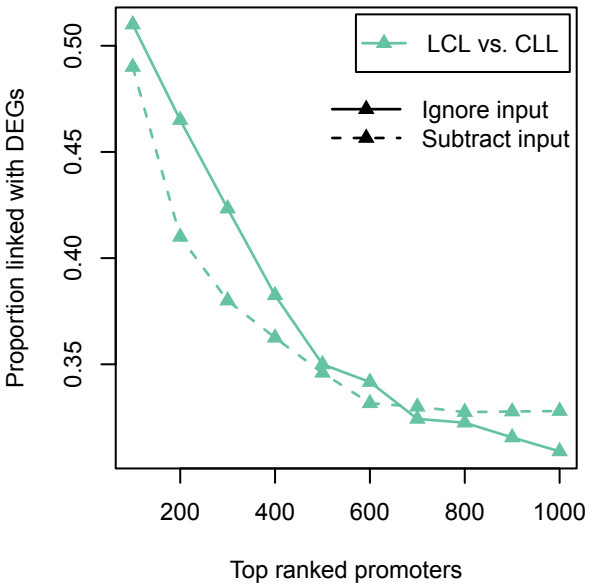
Supplemental Figure S17

Supplemental Figure S17. Assessing the specificity and sensitivity for identifying differential ChIP-seq signals after subtracting read counts from input samples. (A-D) Box plots for *z*-statistic equivalents of the *p*-values assigned to the promoter regions of DEGs and non-DEGs. Dotted lines correspond to a two-tailed *p*-value of 0.05. Here non-DEGs are defined as those genes that have a DESeq2 *p*-value larger than 0.5 and a fold change less than 2.









Supplemental Figure S18

Supplemental Figure S18. Assessing the rankings of promoter intervals after subtracting read counts from input samples. Only in a few cases has the subtraction of input read counts led to an improvement in true discovery proportion.