A



B



**S2 Figure. Statistical analysis of method performances.**

Fisher exact test was used for pairwise comparison of methods. The color coding indicates negative logarithm of p values that range from 0 (for p value 1) to 16 (for p value 10-16). All p values below 10-16 were set to 10-16. (A) Comparison of all the data, and (B) variants that all the methods predicted.