Correction to Protein Identification Using Customized Protein Sequence Databases Derived from RNA-Seq Data

Xiaojing Wang, Robbert J. C. Slebos, Dong Wang, Patrick J. Halvey, David L. Tabb, Daniel C. Liebler, and Bing Zhang*

Due to a bug in the code for data processing, all RPKM values in this paper were erroneously multiplied by 10. All "RPKM20" mentioned in the text, tables, and figures should be "RPKM2". On page 1013, the "RPKM greater than 20" should be "RPKM greater than 2". On page 1011, the "RPKM value greater than 6.16" should be "RPKM value greater than 0.616". In Figure 3, "RPKM > 20" should be "RPKM > 2". Moreover, the *x*-axis in the original Figure 1 should be labeled as in the Figure 1 below. All other results and conclusions are not affected by this mistake. The authors apologize for this error and for any confusion this may have caused.
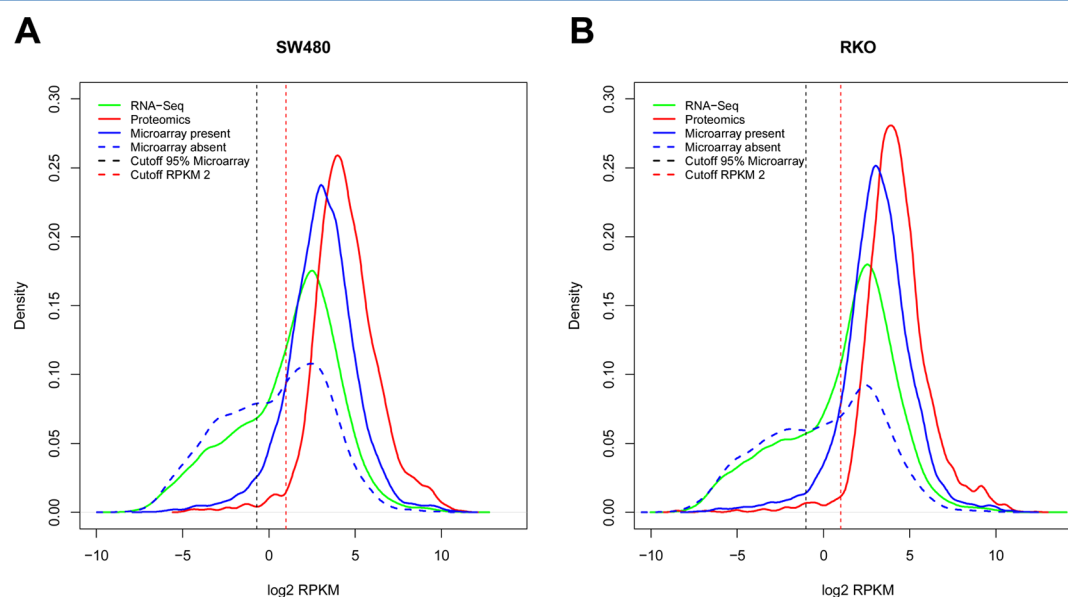

Figure 1. The distribution of log2RPKM for three technologies in SW480 and RKO. The green line reprsents the distribution for all transcripts detectd by RNA-Seq. The blue lines represent microarray data, solid for transcripts with a present call and dashed for absent call. The red line is the distribution for all proteins detected by shotgun proteomics using regular database searching. The black vertical dashed line corresponds to the 95% quantile of microarry present call and the red vertical dashed line corresponds to RPKM2.