



**Fig. 3. Overview of performance evaluation for Challenge 2: transcript isoform quantification.**

**(a)** Cartoon diagrams are used to explain different evaluation metrics under the ground truth given or not given.

**(b) - (e)** Overall evaluation results of eight quantification tools and seven protocols-platforms on real data with multiple replicates, cell mixing experiment, SIRV-set 4 data and simulation data. Box plots are employed to illustrate the five-number summary of evaluation results across various datasets, depicting the minimum, lower quartile, median, upper quartile, and maximum values. Bar plots are utilized to visualize the mean values of evaluation results across diverse datasets, with error bars indicating the standard deviation of evaluation metrics.

**(f) - (g)** Top 3 overall performance on quantification tools and protocols-platforms for each metric. Here, quantification tools showcase scores under common cDNA-ONT and cDNA-PacBio platforms across various evaluation metrics, with the top 3 performers highlighted for each metric. In addition, based on the average values of each metric across all quantification tools, scores for protocols-platforms are displayed, along with the top 3 performers for each metric. Blank spaces denote instances where the tool or protocols-platforms did not have participants submitting the corresponding quantitative results.

**(h)** Evaluation of quantification tools with respect to multiple transcript features, including the number of isoforms, number of exons, isoform length and a customized statistic K-value representing the complexity of exon-isoform structures. Here, the normalized MRD metric is used to evaluate performance of quantification tools on human cDNA-PacBio simulation data. Additionally, RSEM evaluation results with respect to transcript features based on human short-read simulation data are shown as a control.