**Challenge 2 Evaluation** a **Predictions Ground truth** Entry Entry Available before Entry submission 7 submission 4 submission submission 1 SIRV Set 4 Entry Entry submission 5 Available after submission 2 submission Entry Entry Entry ground truth of simulated submission 6 submission 3 submission N data Samples: WTC11 + SIRV4, H1\_mix + SIRV4, human\_simulation, mouse\_simulation Ground truth given No ground truth H1 mix, WTC11 human simulation, mouse simulation SIRV4 in each sample Multiple replicates SIRV4 (Log fold change of SIRV  $\approx$  0) Irreproducibility Consistency Single sample Two conditions Correlation Precision Abundance Recovery Rate (ARR) Recall Single sample Median Relative Difference (MRD) Accuracy Resolution Entropy (RE) Normalized Root Mean Square Error (NRMSE) F1-score ROC and AUC b H1 mix: mix of H1 and endodermal cell at r1:r2 ratio  $r1*TPM_{H1} + r2*TPM_{endodermal cell} \approx expected TPM_{H1\_mix}$ Submission N Submission 2 Submission 3 Submission 1 Initial submission: Initial submission: Initial submission: Initial submission: observed  $\mathsf{TPM}_{\mathtt{H1\ mix}}$ observed TPM<sub>H1\_mix</sub> observed TPM<sub>H1 mix</sub> observed TPM<sub>H1 mix</sub> Final submission: Final submission: Final submission: Final submission: TPM<sub>H1</sub>, TPM<sub>endodermal cell</sub> TPM<sub>H1</sub>, TPM<sub>endodermal cell</sub> TPM<sub>H1</sub>, TPM<sub>endodermal cell</sub> TPM<sub>H1</sub>, TPM<sub>endodermal cell</sub> Gene or isoform level quantification mix - expected TPM<sub>H1</sub> Compare No difference between observed  ${\rm TPM}_{{\rm H1\_mix}}$ observed and expected and expected  $TPM_{_{\rm H1\ mix}}$ 0 Correlation observed TPM<sub>H1</sub> **MRD** NRMSE Sub 1 Sub 2 Sub 3 Sub 4