

S1 Table. Data sets for Gene Inference challenge. This table summarizes the training, validation and test data sets used for training and evaluation in this challenge.

Table 1. Data sets for the Gene inference challenge

Dataset	Size (genes \times samples)	Platform	Source
Current	$12,320 \times 12,000$	Affymetrix	GEO
Training	$12,320 \times 100,000$	Affymetrix	GEO
Testing	$970 \times 1,600$	L1000	GTE _x
Ground	$12,320 \times 1,600$	RNA-Seq	GTE _x

Competitors had access to the Training data set that they could use to develop their solutions. Their solutions were evaluated on the Testing (for provisional scoring) and Ground (for final scoring) data sets. The benchmark MLR method was initially trained on the Current data set.