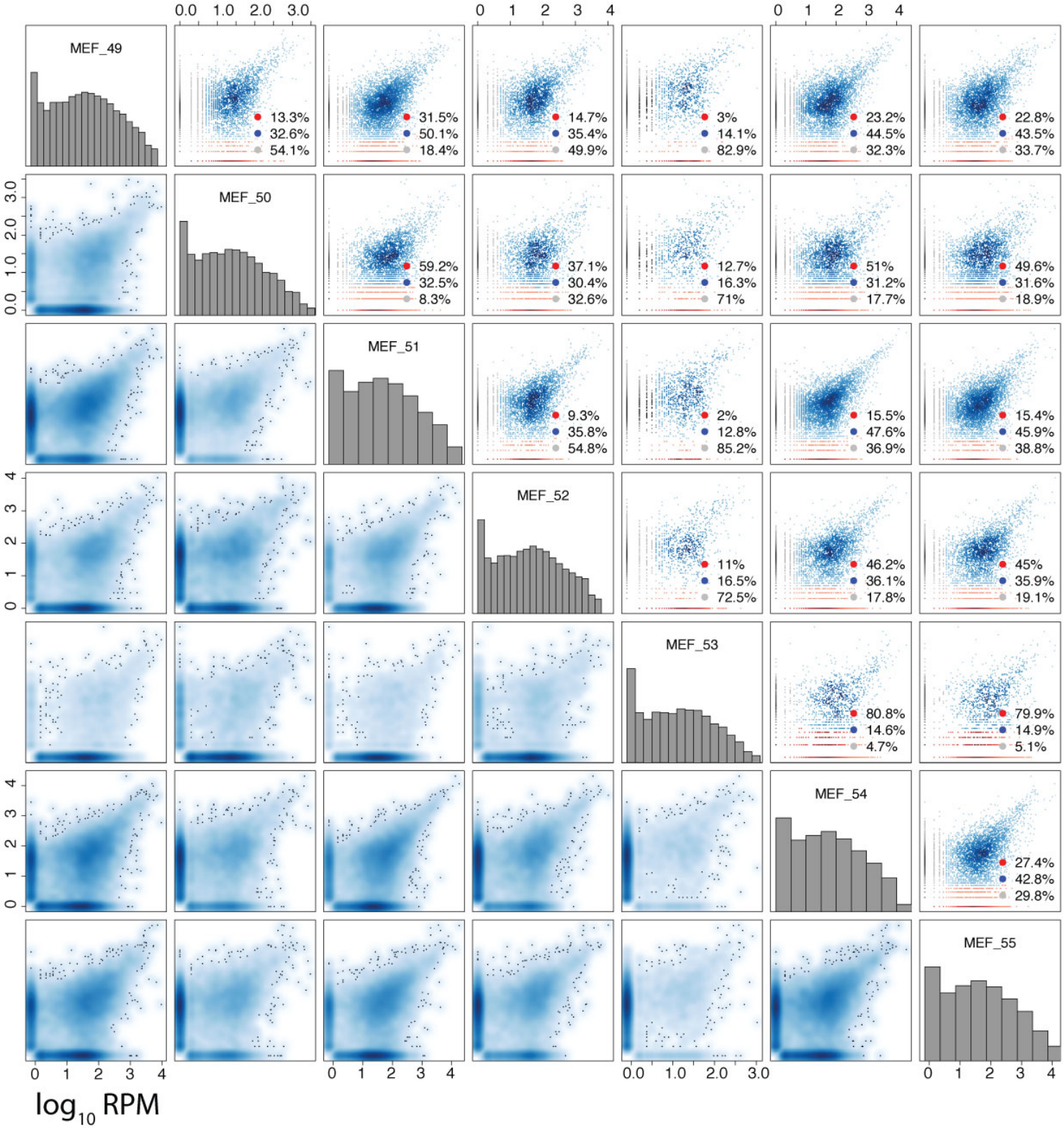
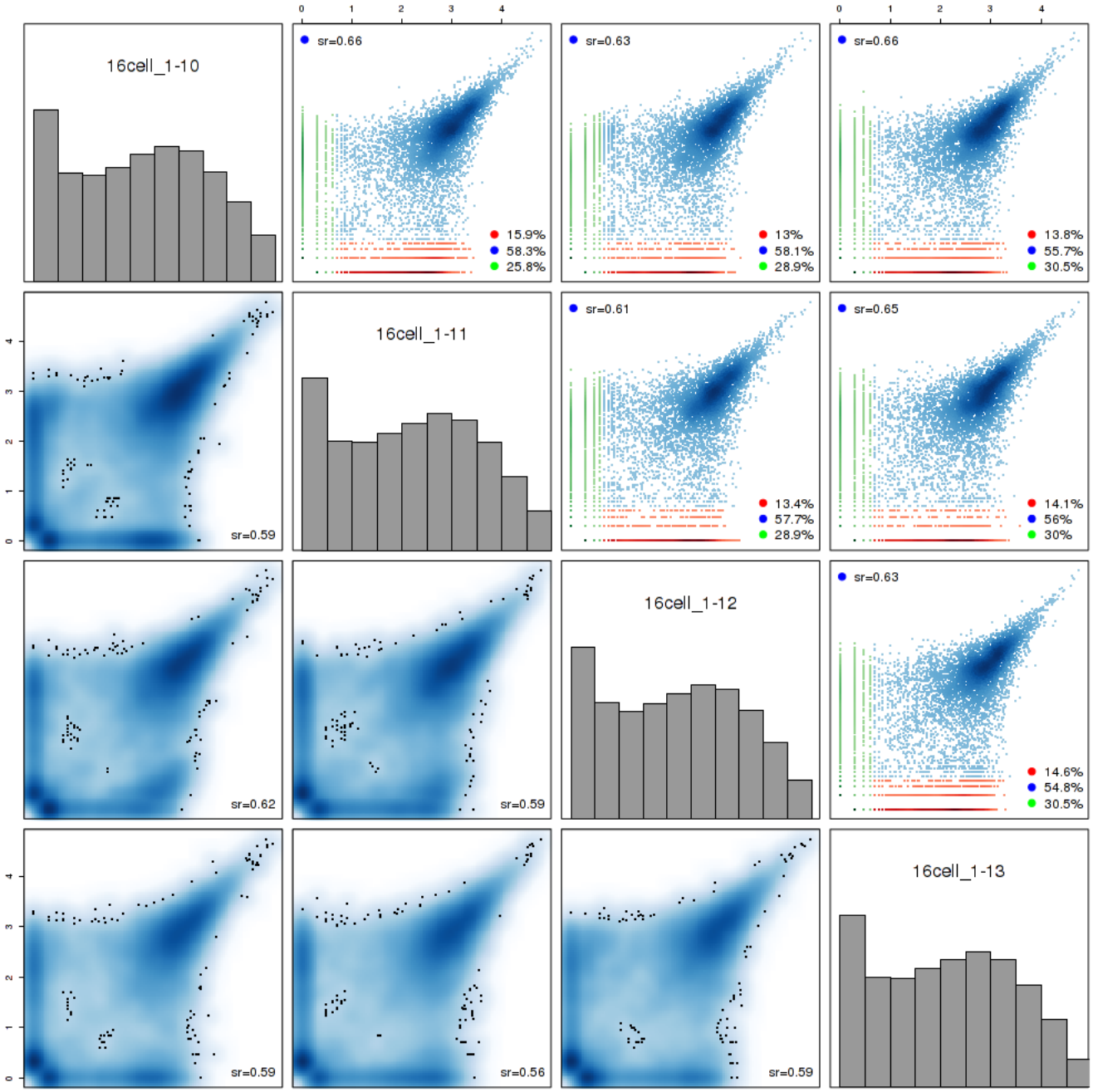
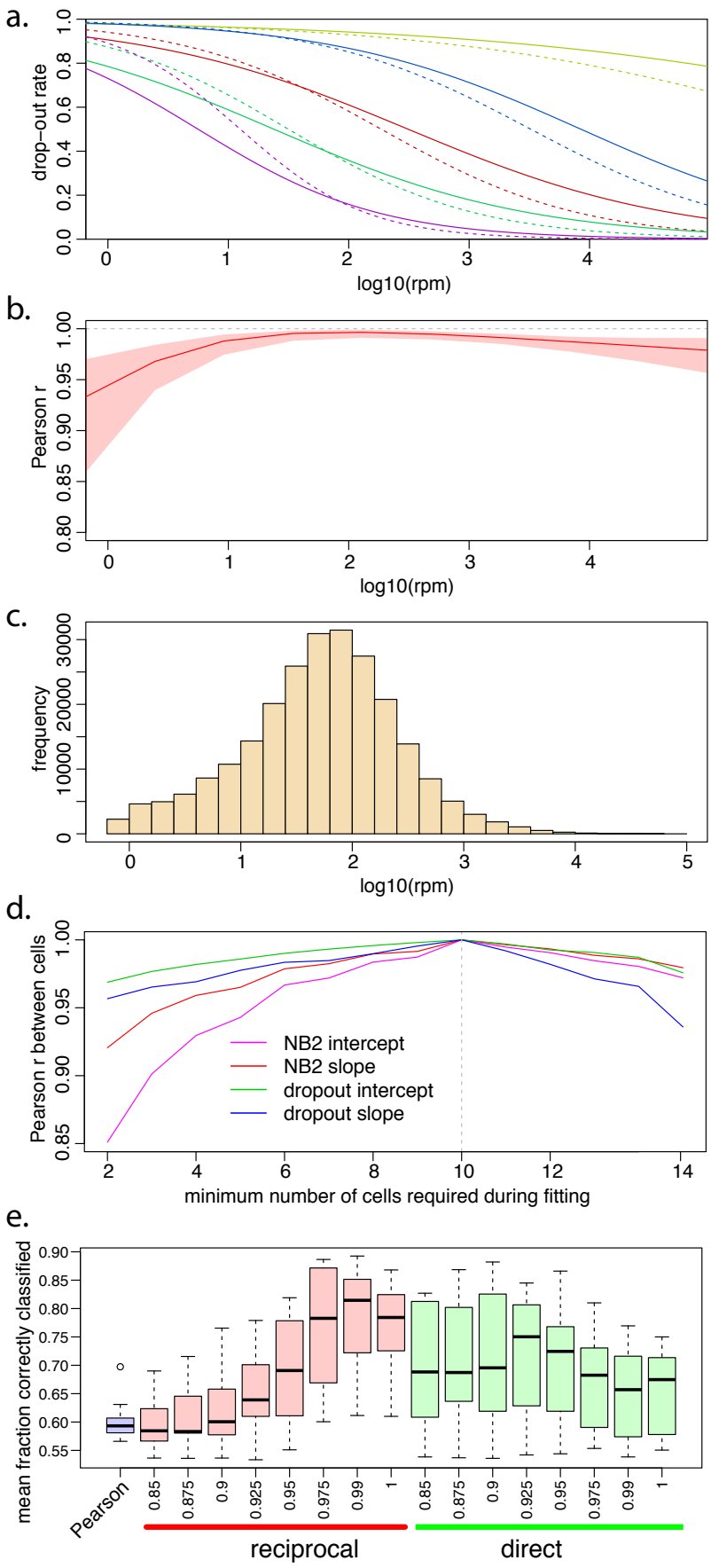


Supplementary Figure 1. Pair-wise comparison of MEF cells.
The matrix shows pair-wise comparisons between first seven MEF single-cell measurements. The x and y axis of each plot show $\log_{10}$ RPM estimates in the cell corresponding to a given column and row respectively. The set of smoothed scatter plots on the lower left shows the overall correspondence between the transcript abundances estimated in two given cells. The upper right corner shows three-component mixture model, separating genes that "drop-out" in one of the cells (green component shows drop-out events in the column cell, red component shows drop-out events in the row cell). The correlated component is shown in blue. The percent of genes within each component is shown in the legend. Some cells, such as MEF_53 show consistently higher drop-out rates.

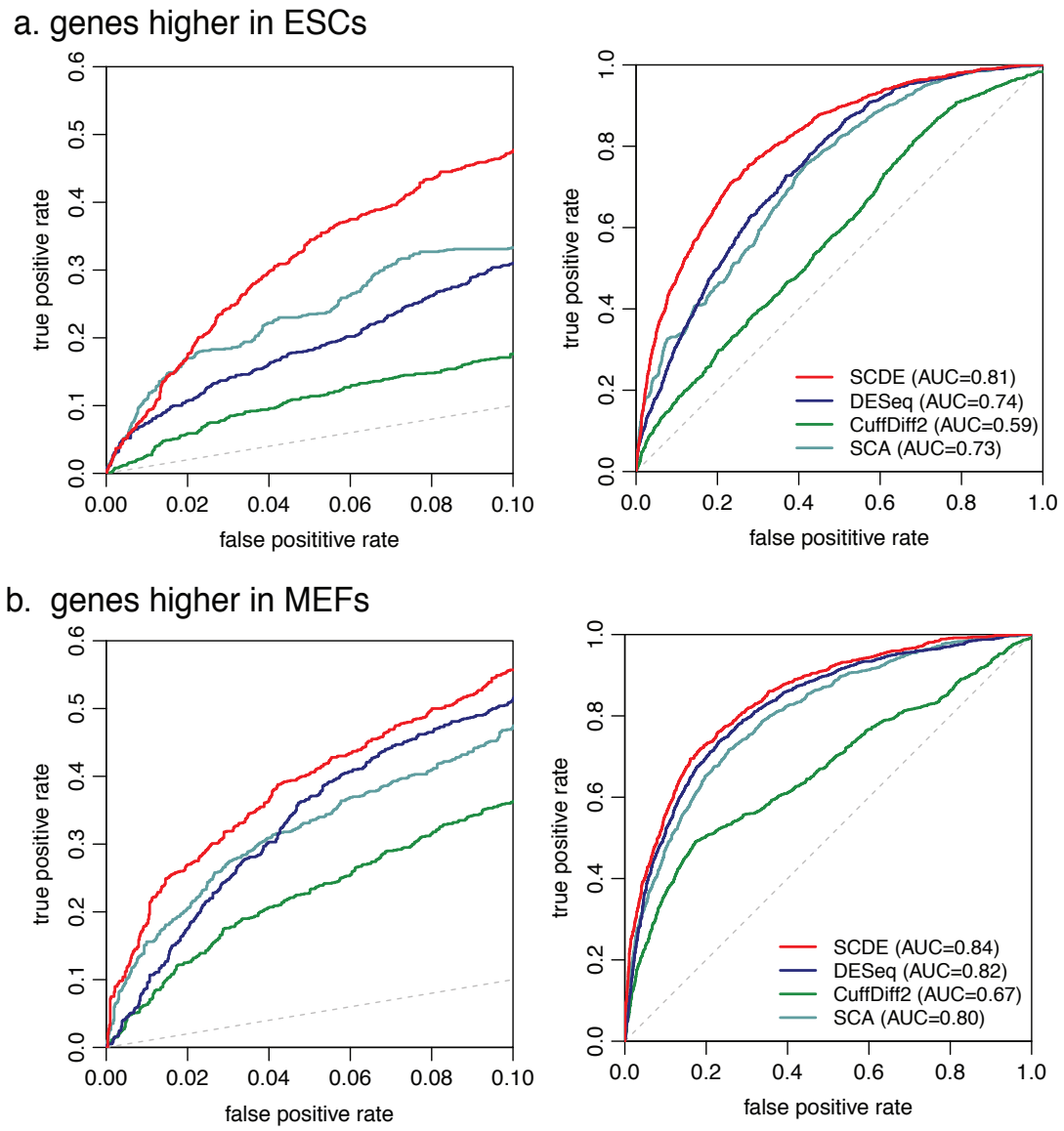


Supplementary Figure 2. Pair-wise comparison of cells from 16-cell mouse embryo measured by Deng *et al*. Similar to Supp. Figure 1, the matrix shows pair-wise comparisons between first seven four cells from 16-cell embryo measured by Deng *et al.* Despite the reduced noise of the newer protocol, the drop-out events are notable, and are more common in some cells then others (e.g. cell 1-13 shows ~30% drop-out most comparisons, whereas cell 1-12 shows ~15%).

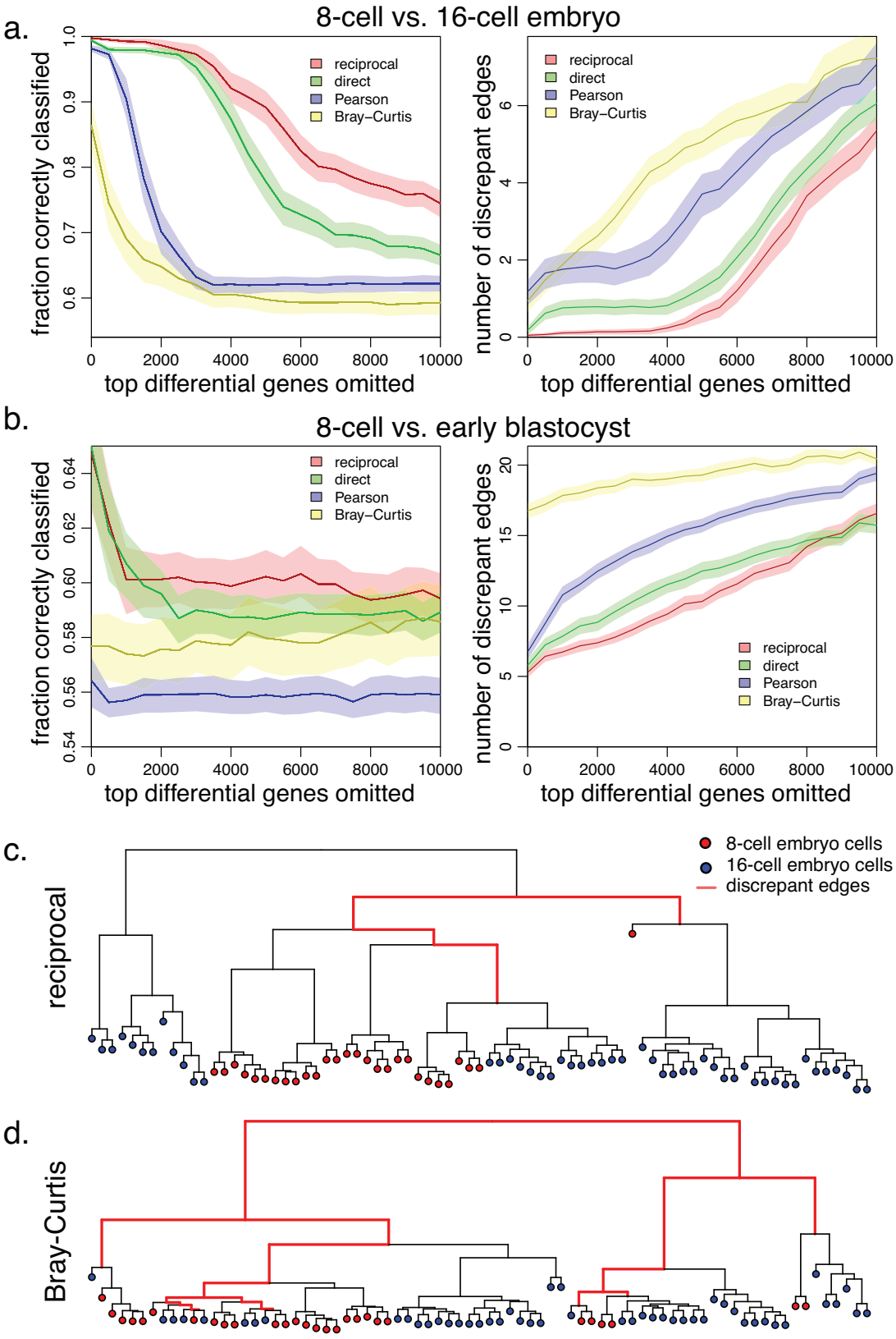


Supplementary Figure 3. Robustness and accuracy of the model fitting.
a. Examples of predicted (slid lines) and true (dashed) drop-out dependencies obtained on simulated data (see simulation description below). **b.** Correlation of predicted drop-out rates is shown as a function of expression magnitude based on 100 random simulations (95% confidence band shown). Correlation remains relatively high (r~0.9) even for genes expressed at low magnitudes. **c.** Distribution of expression magnitudes illustrates that relatively few genes would be impacted by slight decrease in correlation at the extreme expression magnitudes. **d.** Dependency of the model parameters on the selection criteria for "reliable" genes whose median expression magnitude estimates are used for fitting the error models. The plot shows correlation between the model parameters obtained using the default estimate (obtained requiring at least 10 non-failed measurements of a gene) and other values of the minimum number of non-failed measurements (cells) per gene. **e.** Performance of the proposed direct and reciprocal adjusted similarity measures for different values of stabilization parameter *k*. The boxplots show mean fraction of cells classified correctly is shown is shown for the Pearson measure (blue), direct (red) and reciprocal (green) adjusted measures for different values of *k* (x-axis). Boxplots illustrate variability of the performance based on 50 random single-cell dataset simulations. In each simulation, 20 cells of each type were sampled.

Simulation of single-cell data. In each iteration, we simulated 50 single-cell measurements, using a set of 5000 genes. The expression magnitudes were drawn from the empirical distribution of the MEF expression magnitudes in Islam *et al.* (across all of the cells). The library size was chosen for each simulated cell from a uniform distribution U(0.4-3 million reads), with the negative binomial overdispersion parameter (size) chosen from a uniform distribution U(0.3, 2). The drop-out dependency on the expression magnitudes were simulated using logistic function (intercept ~ N(1.5,0.5), slope ~N(3,1)) to approximately resemble dependencies observed in the Islam *et al.* data. For a-c. 500 (10%) differentially expressed genes were simulated. For classification measurements shown in e., a smaller number (50) of differentially expressed genes were used to pose a more challenging classification problem.



Supplementary Figure 4. Performance in detecting genes significantly upregulated in ES or MEF cells. Similar to the Figure 2c of the main manuscript, the ROC curves are used to assess the performance of the differential expression analysis methods. In this case, the genes were ranked based on the significance of their up-regulation in ES (**a.**) or MEF (**b.**) cells. Most methods perform better in detecting genes that are significantly higher in MEFs, which is likely related to much higher overall RNA-seq signals detected in MEF cells (as opposed to ES cells with relatively low RNA content and fewer detected reads).



Supplementary Figure 5. Clustering performance on early embryo cells measured by Deng *et al.*
a. Separating cells derived from 8-cell and 16-cell embryos. The left plot shows fraction of cells correctly classified by the top-level clustering split as a function of the number of top differentially expressed genes removed (as in Figure 2d in the main manuscript). The right plot shows another measure of clustering performance – number of discrepant edges (lower number corresponds to better performance, see figures c,d below). 10 cells of each type were randomly sampled in each iteration. By both measures, the reciprocal distance measure shows better performance.
b. Separating cells derived from 8-cell embryos and early blastocysts. This is a more challenging contrast, and the top-level split identified by all of the examined methods does poorly in separating the two types of cells. However the "number of discrepant edges" (right plot) demonstrates that the drop-out weighed distance measures (reciprocal and direct) tend to group 8-cell embryo cells in one part of the resulting tree.
c,d. Lower number of discrepant edges corresponds to better classification performance. All measured 8-cell and 16-cell embryo cells are clustered using reciprocal-weighted distance (c.) or Bray-Curtis distance (d.). The discrepant edges (highlighted in red) mark all branches of the tree that contain more than one type of cells within them. Here, a clustering derived using reciprocal-weighted distance (c.) groups most of the 8-cell embryo cells into one branch and the resulting number of discrepant edges is small. By contrast, clustering derived using Bryan-Curtis distance distributes 8-cell embryo cells more evenly, resulting in many more mixed branches, and a higher number of discrepant edges.