We model gene regulation using a Random Forest machine learning approach on the 758 single-cell expression profiles and 4,924 highly variable genes with 208 transcription factors(TFs).

Our Random Forest model allows for non-linear dependencies of target genes on causal transcription factors. Each single-cell expression profile is treated as a steady-state condition, thus our model learns a function that maps expression values of potentially all TFs, to the expression value of each target gene.

To address drop-out effects and other noise in single-cell data, we merged the expression of consecutive cells to generate pseudo-cells. Therefore, we subdivide the 758 single-cell expression profiles into bins and take the median of the expression value of each gene in each bin or pseudo-cell.

The number of cells within a pseudo-cell is called "bin size", which is tuned based on Out-of-Bag errors on the training set. The Random Forest uses bootstrap aggregation, where each new tree is trained on a bootstrap sample of the training data. The Out-of-Bag error is estimated as the average error for each training data point *pi* evaluated on predictions from trees that do not include *pi* in their corresponding bootstrap sample.

The final optimum bin size found is 12, hence our steady-state model inferencer is trained on 64 pseudo-cells.

Finally, the Random Forest model ranks TFs based on their influence on target gene expression, generating a predicted GRN based on TF causality.

To refine these TF–target predictions, we retained the highest-confidence edges, specifically, the top-10 transcription factors for each gene target according to the score, resulting into 49,240 edges. [Please say where the code is]