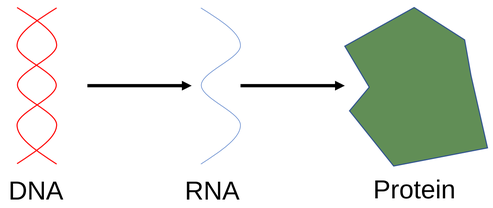
Improvements to single-cell mRNA sequencing

Yash Nagpal, Alex Chialastri, Siddharth S. Dey

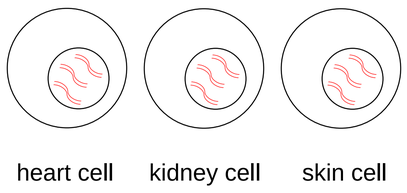
Department of Chemical Engineering, University of California, Santa Barbara

CHEMICAL ENGINEERING
UC SANTA BARBARA

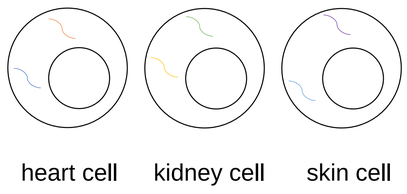
Introduction and Motives



Single-cell mRNA sequencing sheds light on the individual functionality of cells. RNA sequencing allows us to understand which genes are actively transcribed in cells, and single-cell sequencing allows us to identify rare cells. In our project, we made changes on the single-cell RNA sequencing technology (CEL-Seq2) to reduce the costs while maintaining sensitivity to uniquely identify a gene.



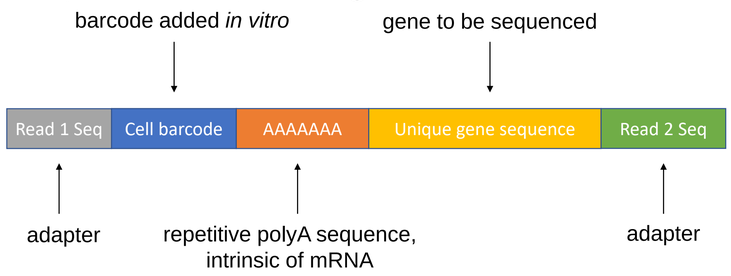
Model depicting the DNA code housed in the nucleus of each cell type is identical in a single organism.



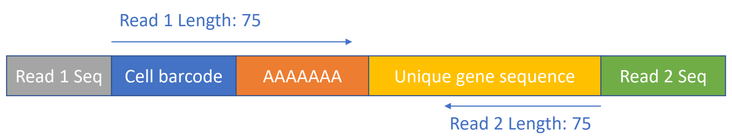
However, the mRNA produced in each cell is highly variable and understanding this variability gives us insights to the physiology of the cell.

Shortening of Sequencing Length

(I) Final molecule prior to sequencing



(II) Current methodology

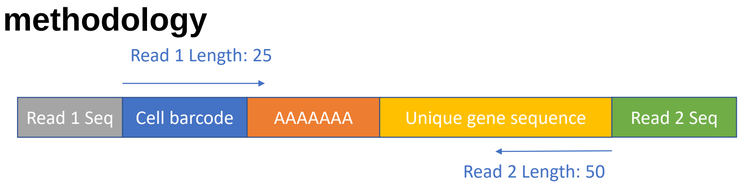


$3.20/cell

mRNA sequencing technology makes use of paired-end sequencing, in which the sequencing information is captured on both sides. This is conducted to obtain the cell specific barcode located on one end of the molecule and unique gene sequence from the other end.

The current protocol uses a 150 base pair kit, which sequences 75 bases on each side.

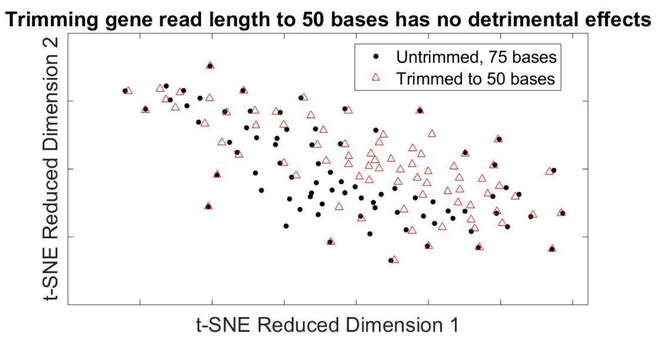
(III) Reduced methodology



$1.70/cell

In the reversed protocol, we make use of a 75 base pair kit, which uses 25 base pairs on the cell barcode and 50 on the unique gene sequence.

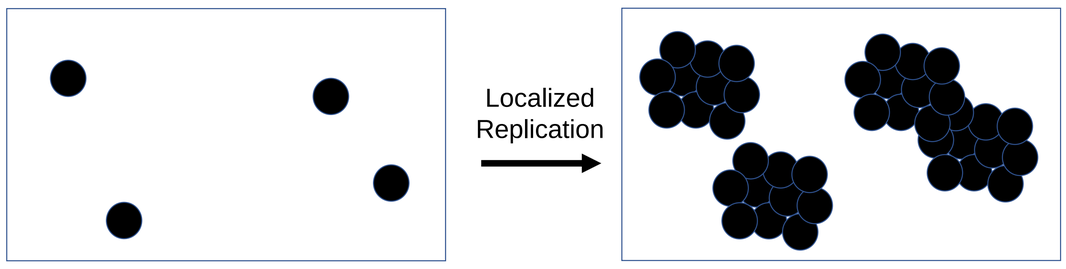
(IV) Gene sequence successfully reduced to 50 base pairs



New cells were created *in silico* by trimming Read 2 (containing the gene sequence) to 50 bases. Even with a reduced sequence, the same genes were uniquely identified and the two populations of cells are indistinguishable, as indicated by the homogeneity of the data. This demonstrates that lowering the gene read length to 50 will not affect the ability to uniquely identify a gene.

Reversal of Protocol

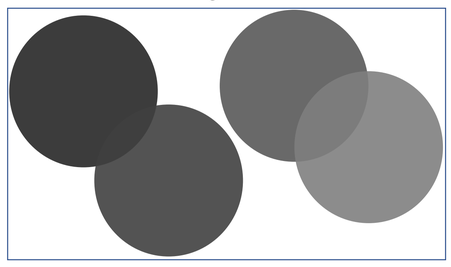
(I) Cluster Sequencing



Individual molecules randomly attach to a flow cell

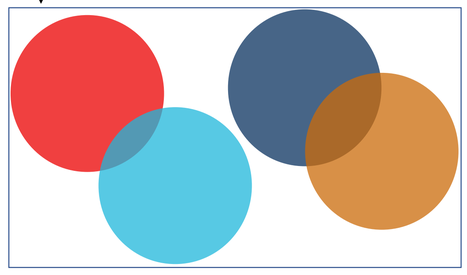
After repeated replications, the clusters of unique genes potentially could overlap. Read 1 is used to identify cluster location.

Original:



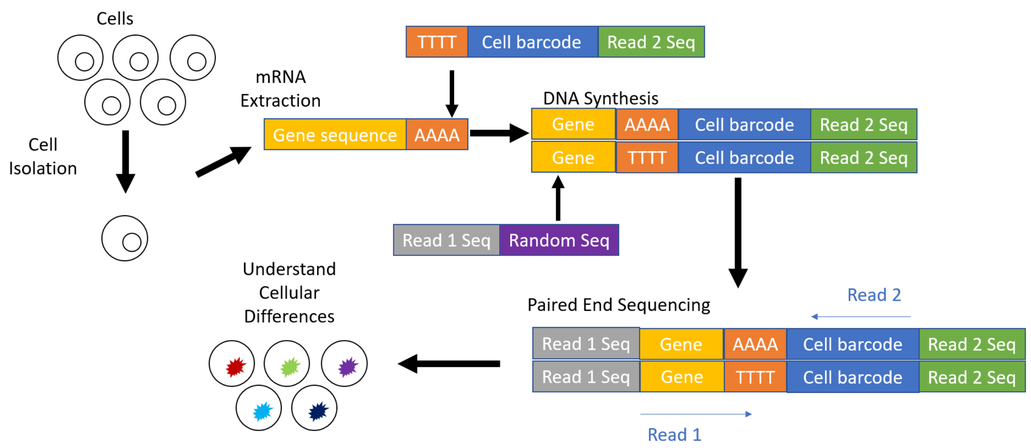
Homogeneity in read 1 leads to difficulties with cluster identification. To combat this, less molecules are added to the flow cell.

Revised:

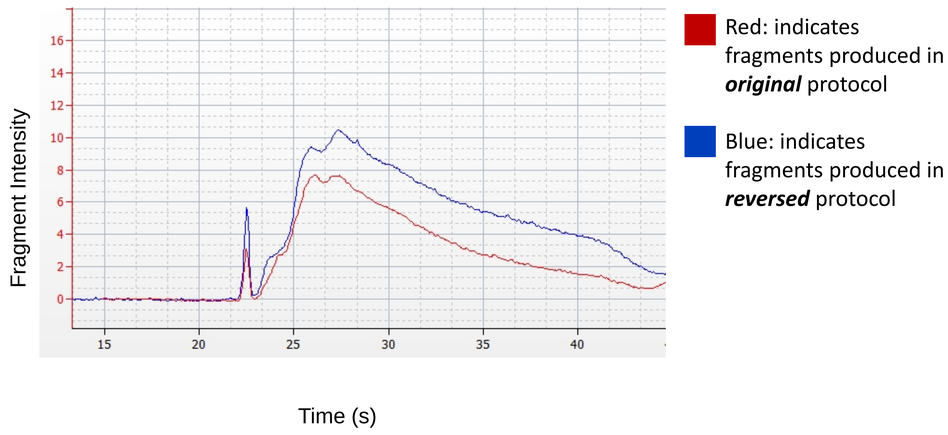


Read 1 now contains unique sequences, improving cluster identification over the original method.

(II) Implementation of revised protocol



(III) Test runs show new protocol to be effective



The checkpoint prior to sequencing shows that a similar amount of molecules was produced in the original and revised protocols. The data provides evidence that the procedure change was implemented successfully. Sequencing is needed to fully validate the method.

Conclusions

Original Cost: $3.20/cell

✓ Sequencing Less Nucleotides → $1.70/cell

✓ Reversing The Current Method → $1.36/cell (20%)

We were successful in developing a revised protocol for single-cell mRNA sequencing. This protocol, known as CEL-Seq3, shows a cost savings of 57.5%. By sequencing less nucleotides, we have already saved over $3,000 in sequencing costs.

Acknowledgments

I would like to thank my mentor: Alex Chialastri, our PI: Sid Dey, the Dey Group, Samantha Davis, Jin Kim, and the rest of the CSEP mentorship team for all of their unwavering support.