DSCAM Isoforms in Single Drosophila Neurons

VOLLMERS LAB

Roger Volden[1], Ashley Byrne[2], Charles Cole[3], Christopher Vollmers[3]

[1]Department of Chemistry and Biochemistry, [2]Department of MCD Biology,
[3]Department of Biomolecular Engineering at UC Santa Cruz

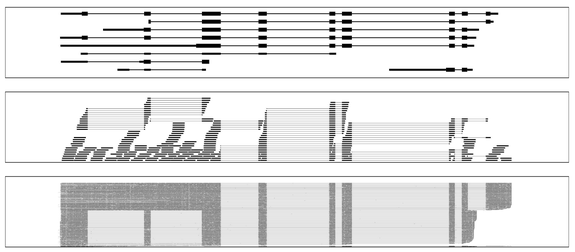
Baskin Engineering UC SANTA CRUZ

Abstract

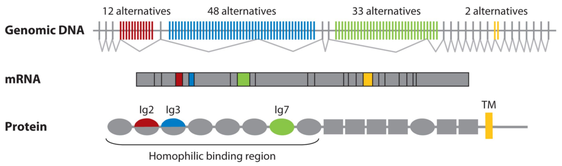
Down syndrome cell adhesion molecules (DSCAM) are encoded by the DSCAM gene, which can adopt upwards of 38000 isoforms in *Drosophila Melanogaster* (*Drosophila*). Different isoforms are proposed to have different physiological functions. One of these proposed functions applies directly to the tiling of neurons by mediating self-avoidance at a molecular level. Because of its importance in understanding neural circuit assembly, it is crucial to understand DSCAM's isoforms at the sequence level. I looked at *Drosophila* neuron transcripts to see what kind of interesting data can be gleaned from long-read MinION data. This involved a fairly simple bioinformatics pipeline that resulted in visualizing the reads using Matplotlib. Because DSCAM is part of the Ig superfamily, the Vollmers lab is well equipped to study this particular gene.

Introduction

The Vollmers lab does single cell RNA sequencing because cell populations, especially neurons, tend to be quite heterogeneous. Each individual neuron is going to have different surface markers, and we want to see all of the transcripts that cell is making. Unlike most studies that use Illumina technologies, our lab uses Oxford Nanopore Technology's MinION for long reads. This allows us to easily distinguish isoforms, which is difficult and inaccurate using short read technologies (shown below).



On top is a gene model for CD37, which shows known isoforms. The middle panel is Illumina data, where it is difficult to see how many isoforms there are because of trailing reads. In the bottom panel, however, where we see the nanopore reads, it is clear that there are only 2 isoforms of CD37 getting expressed by this single cell.



A gene that takes alternative splicing to the next level is DSCAM (Down syndrome cell adhesion molecule, gene model shown above). DSCAM is able to form about 38000 different transcripts from a single gene. DSCAM is important for neuronal development because homophilic repulsion is determined by DSCAM complementarity. If two cells have complementary DSCAM isoforms, then they will be able to bind to each other, which could result in synapse formation. Conversely, if their DSCAM isoforms are incompatible with each other, then they will not be able to bind to each other, which results in neuron tiling. The Quake lab at Stanford is particularly interested in this gene, so they sent us their *Drosophila* cDNA for us to analyze.

Methods

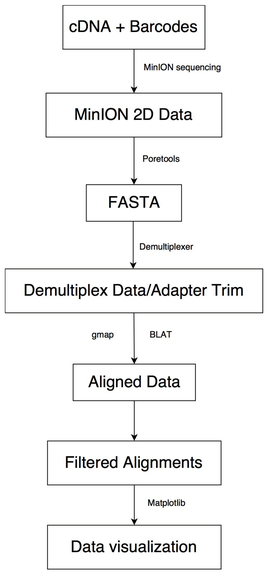


Figure 1. This is the general flow of the bioinformatics pipeline. We first received cDNA from the Quake lab, barcoded it, pooled, and sequenced it on a MinION. We got FAST5 files from the sequencing, which needed to be made into FASTA and FASTQ files for analysis. These data were demultiplexed into the original six cells, because each cell was uniquely barcoded in the library prep. Reads were aligned with BLAT at first, and then gmap for better results. This was parallelized on the UCSC Hummingbird server. From there, we took our aligned data and did some filtering based on read length and quality. I wrote a program that would take the alignment files and plot them against the reference using Matplotlib.

Figure 2. The knapsack problem. In computer science, there is something called the knapsack problem, which is a space optimization problem. A summary is that you want to fit as much into your allotted space as possible without going over your limit. This concept was applied to how reads were plotted in the figures below. By using a boolean flag on each read, we can optimize how we use the figure space to show all of the reads effectively.

We will be using our lab's own software, MandalorION, for future analyses. These analyses include isoform quantification and estimation of gene expression. The dataset will be augmented with 1D reads to get more read depth.

Results

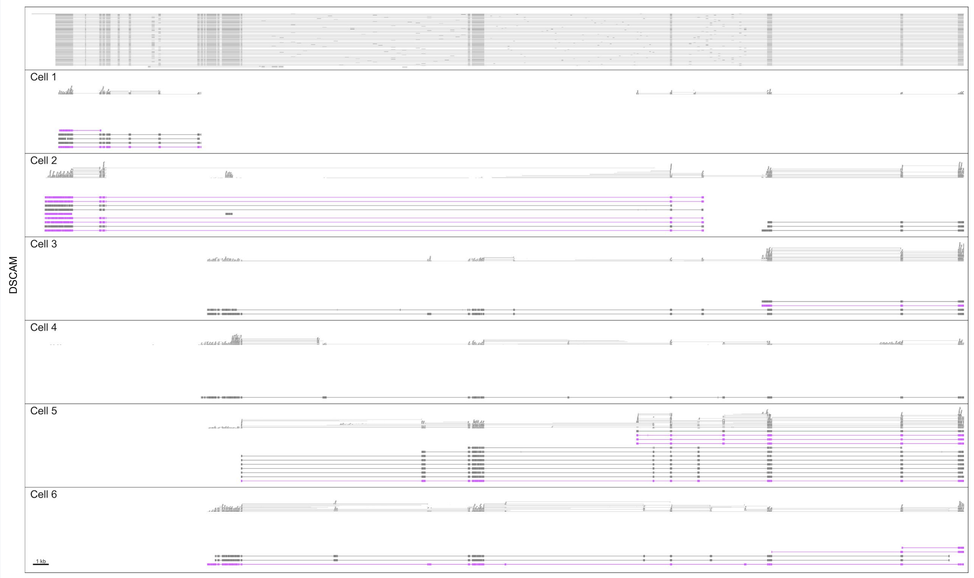


Figure 3. This shows how the reads from each of the six cells align to the *Drosophila* reference. The very top is a gene model, which shows the known isoforms of DSCAM. Each cell shows Illumina reads on top, and then nanopore reads below the Illumina reads. Purple reads are 5' to 3', whereas grey reads have unresolved directionality. **Cell 1.** Nanopore sequencing only found one isoform, but Illumina reads had many hits upstream. **Cell 2.** Two distinct isoforms were found by both Illumina and nanopore. **Cell 3.** Two distinct isoforms found, only resolved by nanopore reads. **Cell 4.** One isoform resolved. **Cell 5.** Two isoforms found by nanopore, somewhat by Illumina. **Cell 6.** Two possible isoforms found in this cell, but the topmost nanopore read could be truncated, giving the appearance of another isoform. Some cells lack the constant region, which is concerning.

Figure 4. This shows reads from the same six cells (cells without aligned reads omitted) aligning to fruitless. Only the first two cells had reads that aligned, and they seem to be expressing different isoforms, one of which is not seen in the gene model.

Discussion and Future Directions

Overall, we were able to see that each neuron expresses at least one DSCAM isoform, which is expected. It was previously seen that neurons would express ~50 different DSCAM isoforms, but the most we saw in our exploration was 2. This makes sense with each cell technically having two copies of the gene, meaning its epigenetic state could be different for each copy. It would be better to target DSCAM specifically instead of poly-A targeting transcripts because the read depth is not good enough to see all of the DSCAM isoforms expressed by these single cells. Our collaborators want us to look at different genes (rox1/2 and fruitless), but these genes were not expressed in all of the cells. Fruitless is the most interesting gene of these in terms of splicing.

Above is the gene model for fruitless. Fruitless is interesting because depending on the way it is spliced, it affects the ability of that individual fly to court a mate. In *Drosophila*, males only court females. Females will never try to court males or other females. Similarly, males will only try to court females, but not other males. The way the male Fruitless gene is spliced leads to a molecular cascade which makes the male perform an elaborate sequence of actions to try to court a female. What's interesting is that if a female fly has a Fruitless gene that is spliced like a male, that female will try to court another female. Normally, the female isoform of Fruitless doesn't do anything, but when given the male isoform for Fruitless, it will suddenly court females. Similarly, we can introduce a female isoform of Fruitless into a male and it will not try to court a female.

Acknowledgements

A huge thanks to the Vollmers lab, particularly Charles, Ashley and Chris. Charles assisted in parallelizing the pipeline, which made everything much more efficient. Ashley made the libraries and actually sequenced the cDNA. Chris helped every step of the way on everything. Another thanks to the Quake lab at Stanford, particularly Felix Horns.

References

1. Ahrens, J. H. & Finke, G. Merging and Sorting Applied to the Zero-One Knapsack Problem. *Operations Research* 1099–1109 (1975).
2. Breitbart, R., Andreadis, A. & Nadal-Ginard, B. Alternative Splicing: A Ubiquitous Mechanism for the Generation of Multiple Protein Isoforms from Single Genes. *Annual Review of Biochemistry* **56,** 467–495 (1987).
3. Hattori, D., Millard, S. S., Wojtowicz, W. M. & Zipursky, S. L. Dscam-Mediated Cell Recognition Regulates Neural Circuit Formation. *Annu Rev Cell Dev Biol* **24,** 597–620 (2008).
4. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12,** 656–664 (2002).
5. Neville, M. C. *et al.* Male-Specific Fruitless Isoforms Target Neurodevelopmental Genes to Specify a Sexually Dimorphic Nervous System. *Curr Biol* **24,** 229–241 (2014).
6. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463,** 457–463 (2010).