Assignment 2

Simon Lee, Daniel Cavaleri, Daniel Molinuevo, Lucille Niederhauser

You sneak in to try EPFL's top-secret teletransportation device and inadvertently have your genome spliced with a stray *Drosophila melanogaster* that entered the machine with you (you should have known the risks, this was the plot device for both the 1958 and 1986 movies 'The Fly').
Unlike in the films, it fortunately appears only some of your epidermal cells have had their genomes infused with *Drosophila* DNA sequences. You are able to culture some of these epidermal cells and first need to find out how much *Drosophila* DNA has been introduced into your genome.

1. What genome sequencing technology will you use?
We need to sequence our whole genome from at least one cell, therefore we have to use Next generation sequencing. Sanger sequencing would take at least millions of cycles!

2. You discover that you have been relatively lucky, only a single stretch of *Drosophila* DNA has been introduced into one of your chromosomes but not into the other homologous chromosome. Nonetheless, a single copy of this sequence is causing your skin to appear more insect-like. In genetic terms how would you describe this phenomena?
Since we are seeing visible effects of our skin appearing more insect-like, it seems that the new gene is replacing the function that is analogous from what our genome should be performing. In genetic terms, this means that the Drosophila's gene has a dominant effect on our own gene. We can see this as a gain of function (GoF) like mutation where, instead of having a mutation on one gene, we are inserting a new one that has similar effects.

3. You identify the inserted *Drosophila* sequence (attached file, coding sequences in pink, non-coding sequences in blue). You decide to inactivate the gene product (or products) of this sequence using genome editing technology. You must first design a guide RNA (or gRNAs) to target this sequence. Describe how you would do this and the sequence (or sequences) you would choose [only the target sequence(s) is required, not the full gRNA sequence].
Since we are working with a GoF-like mutation, we want to inactivate the new gene by creating any type of loss of function (we do not care how). Therefore, it is quite likely that, if we remove enough parts of its coding sequences we will successfully inhibit the gene expression.

After running BLAST on our sequence, it shows that we have many exons, which suggests that we will probably need at least 2 gRNAs to cut a long enough sequence that disrupts most of these exons. Using 2 gRNA would allow us to cause double strand breaks at two distant points of the gene and remove all the content between them after going through a DNA repair.

When it came to designing our gRNAs, we looked for possible gDNA sequences in the endogenous DNA sequence by using Benchling's gDNA design which looks for off-targets in the human genome. This is because the Drosophila gene has been introduced into our genome, therefore the off-targets could occur within the Human genome (not within the Drosophila's!)

We also made sure that a protospacer adjacent motif (PAM) sequence is encoded right after the targeted DNA region. PAM sequence is a short DNA sequence required for a Cas9 to perform a cleavage (often 3-4 nucleotides downstream from the cut site).

We have found two different promising sequences fulfilling these criteria. The sequences start at the following positions:

- 3588 (+ strand): 77.5 on-target score and 96.6 off-target score.
- 8903 (+ strand): 76.1 on-target score and 96.5 off-target score.

We prefer to compensate for a lower on-target score by increasing the amount of CRISPR-Cas9 we introduce rather than risking possible off-target effects.

The following contains the target sequences (PAM in bold):

- GGATGCCGCCTACAATACGG**TGG**
- TTACGAGCATGTGATCGACG**AGG**

Therefore our gRNA would code for their corresponding complementary sequences (not including the PAM).

Describe a common risk with genome editing technology and the steps you took to avoid this with your design.
The main risk associated with genome editing technology are the possible off-target effects. In the context of our problem, adding CRISPR-Cas9 to target the Drosophila's gene might mistakenly target another gene of our own genome which could have horrible consequences.

This is why, when selecting sequences, we prioritized those that successfully carried out the task, whilst having the lowest possible chance of off-target effects.

If your strategy works as expected what would be the ideal edit you could achieve and how would this affect gene expression?

Our strategy is to induce two double-strand breaks that when repaired by non-homologous end joining will join around the cutting sites. As a result only the first 3 exons of the endogenous gene would remain intact. We believe that this will be sufficient enough to generate a loss of function on our target gene. This could be tested by first performing genome editing in the cultivated epithelial cells and see if there is any effect. If not sufficient, we would have to look for sequences that induce a bigger cut or that prioritize the loss of these exons.

4. With your strategy designed, you are now ready to genome edit yourself. In addition to the guide RNA(s) you designed, what other factor(s) will you have to introduce into your cells for the gene editing to work?

For Cas9 to work we will need to introduce the following:

- gRNA, which includes the complementary of the target sequence and the scaffold sequence.
- Cas9 Proteins.

Other options are available, like introducing a plasmid with the Cas9 + gRNA sequences or introducing a mRNA that codes for the Cas9 protein. As will be explained later, we decided that using proteins directly seems like the option that takes less risks.

What gene delivery technology could be suitable to introduce these factors into your cells?

The main characteristic of our delivery method is that it has to be safe since we are introducing it in our own genome! To this end, we considered that lipid nanoparticles seemed like a good option, but the main caveat is that this methodology is not very efficient [1]. This may cause issues since we need to deliver CRISPR-Cas9 in many cells (all skin cells that appear insect-like). We therefore thought that adeno-associated viruses (AAV) might be a better method. This method is also safe and has a high delivery efficiency [1]. However, this method also contains a caveat which is that it is only possible to use them as a delivery method if we decide to deliver Cas9 and the gRNA as a plasmid [1]. We feel like this might not be the best option as the human cells would have to produce Cas9 proteins and it seems there is a higher chance that something could go wrong in the process, simply because more steps are involved. We therefore think that lipid nanoparticles would be the best option. Using this method allows us to directly deliver Cas9 proteins and gRNAs [1], meaning that we remove the risk that something could go wrong during plasmid translation and protein production. We can overcome the low delivery efficiency issue by simply injecting more lipid nanoparticles in each cell which should not be an issue since they are easy to produce.

References:

[1] Liu C, Zhang L, Liu H, Cheng K. Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications. J Control Release. 2017 Nov 28;266:17-26.