Chloé Daul
Ehsan Dadashzadeh
Viola Renne
Pauline Theimer-Lienhard

18th October 2023

BIO-411 - Life Sciences engineering : genome to function

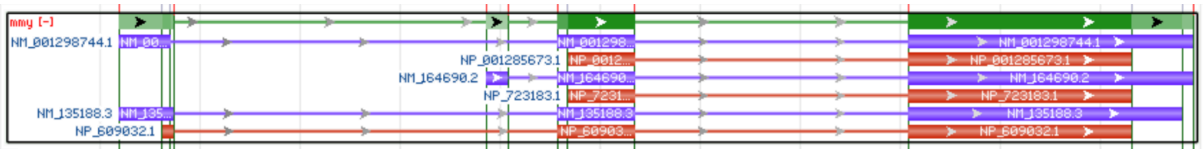
Project 1 - Week 3 & 4

1. Since the position where the Drosophila DNA sequences have been incorporated is unknown, it is necessary to sequence the whole genome. In addition, not knowing the exact position of the integrated sequence means that we don't have the primers for Sanger Sequencing. For these reasons, we would use the Next Generation Sequencing technology. To determine where and which sequences have been inserted, it is needed to align the reads to the reference human genome.

2. This type of phenomenon is described as a neomorph (gain-of-novel-function) mutation. The introduction of Drosophila DNA causes a novel gene production function: it does not increase the normal function but creates a different one (skin appears more insect-like). Since only one copy is sufficient to produce a phenotype, it is dominant.

3. By running the coding sequence on BLAST, we found that the inserted Drosophila sequence corresponds to the mummy gene (mmy). Looking at the FlyBase, we saw that it encodes for the UDP-N-acetylglucosamine diphosphorylase which is needed for chitin synthesis and regulates cuticle production (this explains the insect-like skin phenotype). It has three different isoforms (isoform A, isoform B and isoform D).

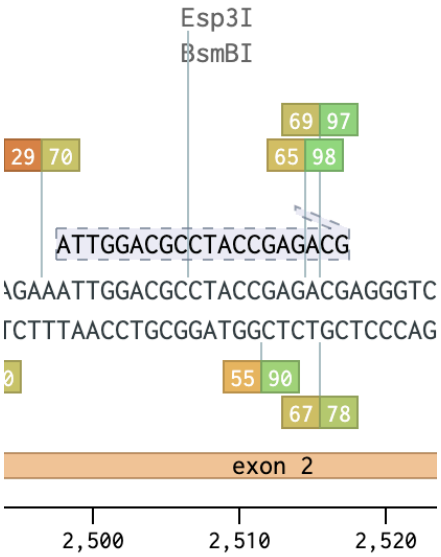
Drosophila melanogaster mummy (mmy), transcript variant A, mRNA	Drosophila mel…	2887	2887	100%	0.0	100.00%	2030	NM_135188.3
Drosophila melanogaster LD24639 full insert cDNA	Drosophila mel…	2887	2887	100%	0.0	100.00%	2036	AF145686.1
Synthetic construct clone BO32701 mmy-RA (mmy) gene, partial cds	synthetic constr…	2881	2881	99%	0.0	100.00%	1594	KX795808.1
Drosophila melanogaster mummy (mmy), transcript variant B, mRNA	Drosophila mel…	2774	2774	96%	0.0	100.00%	1921	NM_164690.2
Drosophila melanogaster mummy (mmy), transcript variant D, mRNA	Drosophila mel…	2774	2853	98%	0.0	100.00%	2064	NM_001298744.1



The gene products of this sequence are the mRNA (with three different isoforms) and the corresponding proteins. For targeting the protein, we need to engineer a monoclonal antibody that binds to all the three isoforms or create three different antibodies. For this reason, we decided to target the mRNA, using RNA interference. In this way, we just need to engineer a RNAi that binds to one of the exons that all the isoforms share. RNAi uses the RNA-induced silencing complex (RISC) to prevent the translation of the mRNA. Using RNAi, we can target sequences at different positions, not just in the 3'UTR of the mRNA. This technique has some advantages, with respect to genome editing, such as if we have off-target effects, we can stop delivering it or change sequence. In addition, it doesn't have immunity problems that Cas9 may have. However, there can be off-target effects if the

target sequence is present in other mRNAs. In addition, RNAi may change the cell functions and it may not work because of the 3D structure of the target mRNA. Also, it only has a short duration effect, so if you want to have a long-term effect, you should continuously deliver it. If we want to reduce the off-target effects, it is possible to use shRNA instead of siRNA, but shRNA has the disadvantage of needing to be introduced on plasmids or via transgenes. For this reason, we decided that siRNA is better, for its simplicity, since it can be directly introduced into cells.

4. To design the gRNA we considered that we want to target the coding region of the gene, since we want to introduce a premature stop codon. To increase the probability of a premature stop codon, we want to target a position at the start of the coding region. In addition, since the gene has three different isoforms, we also want to target an exon that is present in all the three products. Using NCBI (previous figure), we found that isoform A has all three exons, isoform B and isoform D have the second and third exons. So we decided to target the second or third exons, in order to target all the three isoforms. Using Benchling, we selected the following target: ATTGGACGCCTACCGAGACG. It is the gRNA proposed by Benchling with the highest off-target score, and still a high on-target score (higher than 60). We thought about optimizing the off-target score, because we don't want to target any gene in the human genome. The PAM of our gRNA is AGG. It is located on the second exon. It has an on-target score of 65.1 and an off-target score of 71.0 (using CFD score) and 98.9 (using Hsu et al., 2013). Comparing with eu.idtdna.com, we found that this gRNA has an off-target score of 90. All the off-target scores refer to the human genome. We designed only one gRNA because, even if the first cut doesn't introduce any indel, then the DNA sequence is not changed and the same gRNA will still target the same position and hopefully introduce an indel in a future cut.



A common risk of this technology is off-target effects. In our case, the risk is that the gRNA does not only target the Drosophila gene, but also a human gene. This is a possibility, since the guide RNA is a short sequence and it could have homology not only to the target gene (the Drosophila one), but also to human genes. For this reason, we have checked for off-target effects of our gRNA against the human genome.

If our strategy works as expected, the endonuclease Cas9 would cleave the double strand, producing a double-strand break (DSB), in the position targeted by the gRNA. The DSB would be repaired through NHEJ. The repair of DSBs by NHEJ is efficient but error-prone, so it would have high probability of adding or deleting a base (causing an indel) that would induce a frameshift and likely a premature stop codon, thus resulting in a truncated peptide (that hopefully is not functional) or in nonsense-mediated decay.

5. In addition to the gRNA, we need to introduce Cas9, which is necessary to form the Cas9-gRNA complex. The technology to deliver our complex would be AAV vectors, since a single administration is needed and it does not integrate into the genome.