MATERIAL SUPLEMENTARIO

Material utilizado para la optimización:

Para la optimización se simularon una serie de transcritos, teniendo en cuenta tan solo los aspectos mínimos necesarios que requerían las funciones analizadas para su funcionamiento. Estos aspectos eran los fragmentos que componían cada transcrito, la longitud de sus regiones UTR y de su ORF. Se hizo lo mismo con los fragmentos, creando archivos donde se incluía su anotación y su longitud.

Nombre	Fragmentos	3'UTR	5'UTR	ORF
t1	F1,F2,F3,F4	12	10	6
t2	F2,F1,F3,F4	10	12	8
t3	F2,F3,F1,F4	13	9	7
t4	F2,F3,F4,F1	15	7	5
t5	F2,F3,F4,F1	10	12	5
t6	F11,F2,F5,F3,	11	15	10
t7	F2,F11,F5,F3,	13	13	8
t8	F2,F5,F11,F3,	18	8	6
t9	F2,F5,F3,F11,	14	12	6
t10	F2,F5,F3,F10,	10	16	5
tref	F2,F5,F3,F10,	1	2	30
t11	F5,F3,F10,F1	10	65	20
t12	F2,F3,F10,F1	10	65	20
t13	F2,F5,F10,F1	10	67	20
t14	F2,F5,F3,F12	10	66	20
t15	F2,F5,F3,F10,	8	13	10

Nombre	Anotación	Longitud
F1	A3	4
F2	A3	6
F3	exon	4
F4	exon	8
F5	exon	6
F6	A3	5
F7	Intron	20
F8	exon	10
F9	exon	2

Figura1. Material utilizado durante la optimización A) Ejemplo de los transcritos simulados utilizados durante la optimización conteniendo nombre, fragmentos, longitud de las UTR y de las región codificante (ORF). B) Ejemplo de los fragmentos simulados utilizados durante la optimización conteniendo nombre, anotación y longitud

Errores en la detección de fragmentos.

En la figura 2 observamos 2 situaciones en las cuales el pipeline puede erróneamente clasificar como un solo fragmento exónico 2 exones o dividir un solo exón en 2. Estas situaciones generan problemas durante el análisis de los resultados, pues no podemos estar seguros de que los fragmentos exónicos se correspondan con verdaderos exones.

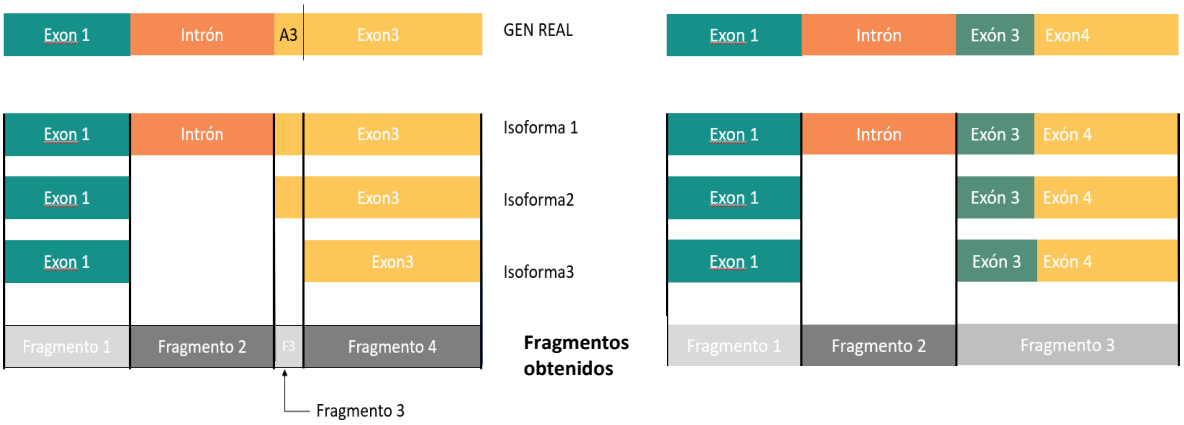

Figura 2. Casos especiales en la generación de fragmentos exónicos. Durante el mapeo de los transcritos contra el locus reconstruido para detectar los fragmentos (que deberían corresponderse con intrones o exones completos) pueden ocurrir casos especiales. A) Cuando hay un evento de splicing A3 o A5 en uno de los exones, en este caso el exón 3, el exón se divide en dos fragmentos, siendo uno de ellos el correspondiente a la secuencia comprendida entre los dos sitios de splicing alternativo. B) Cuando dos exones aparecen siempre juntos en todos los transcritos no se pueden distinguir, por lo que acaban uniéndose en el mismo fragmento exónico (exones 3 y 4 se unen en el fragmento 3).

Reconstrucción de secuencias mediante K-mers y grafos de Bruijn.

Para reconstruir la secuencia codificante original de la que provienen los transcritos se dividen las lecturas en K-mers, fragmentos deslizantes de determinado número de bases. Después se ordenan estos K-mers en grafos de Bruijn que intentan determinar que K-mer va antes que otro. En la figura 3 se observa el proceso de separación en K-mers de una secuencia y la posterior reconstrucción de esta gracias a los grafos de Bruijn. Cuando se relacionan los K-mers en el grafo puede ocurrir que existan varias posibilidades, dando lugar a ambigüedades que se representan como diferentes caminos en el grafo. Estas ambigüedades se intentan resolver, pero si no se consigue da lugar a más de una reconstrucción de la secuencia. Un ejemplo de estas ambigüedades la encontramos en la figura 3, donde debido al uso alternativo de los exones en los transcritos se generan dos posibles caminos en el grafo, siendo el naranja el camino correcto.

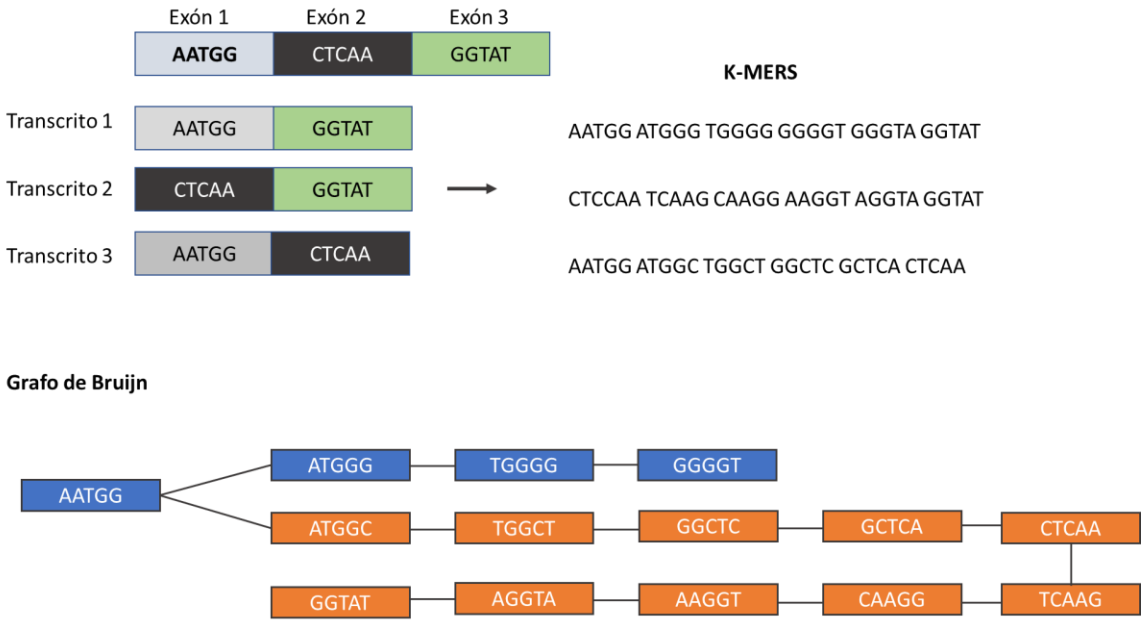

Figura 3. Ejemplo de reconstrucción de secuencia a partir de secuencias de transcritos. A partir de las secuencias de transcritos se obtiene todos los posibles K-mers, es decir, todas las subsecuencias de tamaño K. Posteriormente se relacionan los K-mers entre si: Si las primeras K-1 bases de un K-mer coinciden con las últimas K-1 bases de otra se entiende que están relacionadas. En la imagen observamos como debido al splicing en los transcritos el primer K-mer puede estar relacionado con dos K-mers distintos, por lo que se generan dos caminos alternativos, siendo el camino correcto el naranja.