

USO DE REDES NEURONALES EN EDICIÓN GENÉTICA

Daniel Fraile Belmonte - 721525

Pablo Noel Carreras Aguerri - 718743





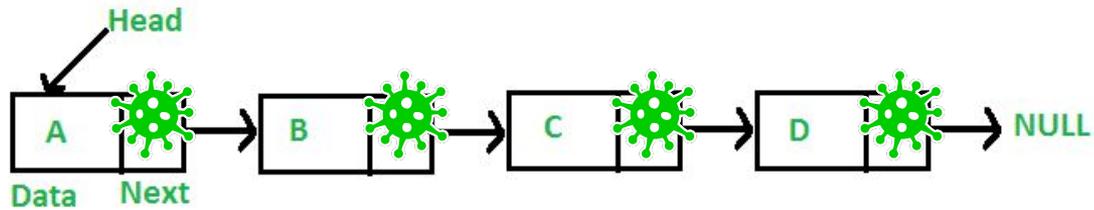
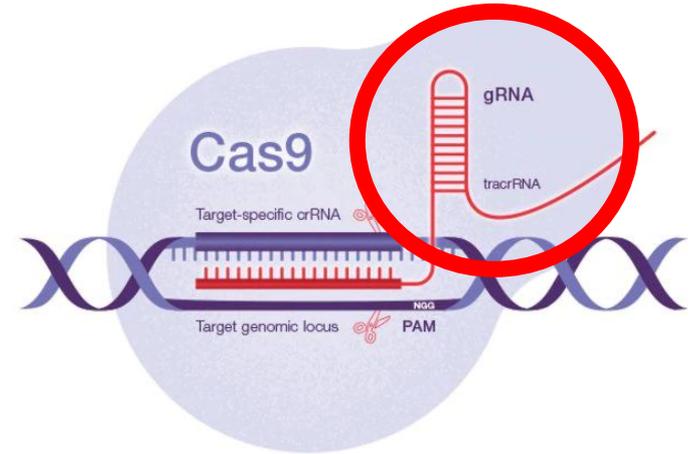
A CONTINUACIÓN:

- 1. CRISPR-cas9**
- 2. Gene Drive**
- 3. Cas9 vs Cpf1**
- 4. Redes neuronales en CRISPR-cpf1**



CRISPR

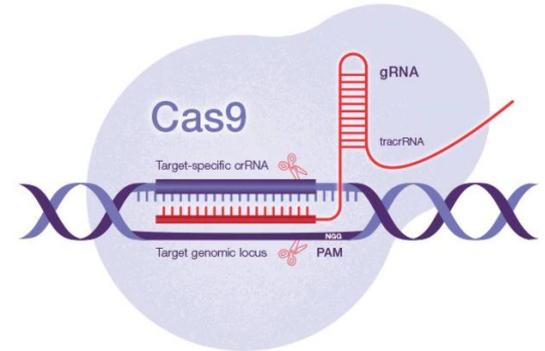
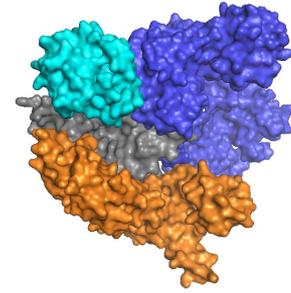
- Repeticiones Palindrómicas Cortas Agrupadas y Regularmente interespaciadas
- Palindrómicas
- Interspaciadas
- Hay una secuencia denominada “Líder”
- Secuencias similares a las de los virus





Cas9

- Enzimas de las bacterias
- Localizan y eliminan el ADN malicioso introducido por los virus
- No solo eliminan, lo añaden a la lista de virus (PUSH)



CRISPR-cas9



- Más allá del organismo, años después de que se descubriera, se publica su uso como herramienta para la modificación del ADN
- Se diseña una molécula de ARN guía que se inserta en una célula cas9
- No solo permite eliminar sino también sustituir

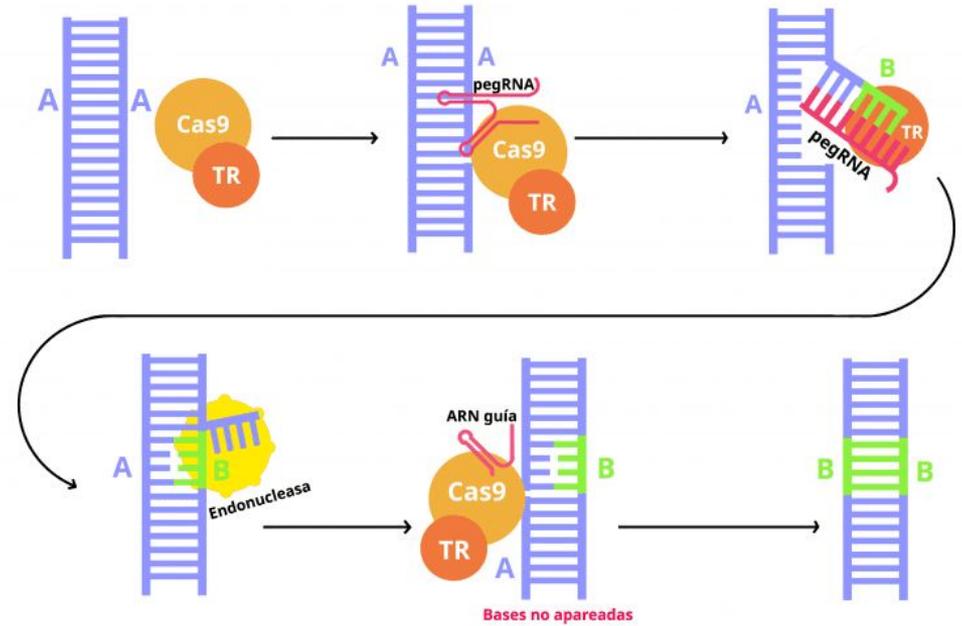


Objetivo y Futuro

- Encontrar fragmentos del ADN que se repitan el número posible de veces y que sean fácilmente sustituibles por el ADN deseado
- Mejorar la eficacia
- Evitar efectos off-target
- Como cas9 es de origen bacteriano, hay anticuerpos que la rechazan
- Hay alternativas más complicadas, pero esta sigue siendo muy popular

Ejemplo: Prime Editing

- Es similar, usa cas9 pero hace un corte simple, no doble.
- Es eficaz en un mayor número de células
- Produce menos errores
- Potencial para usarse en el 90% de variantes genéticas humanas que causan patologías)

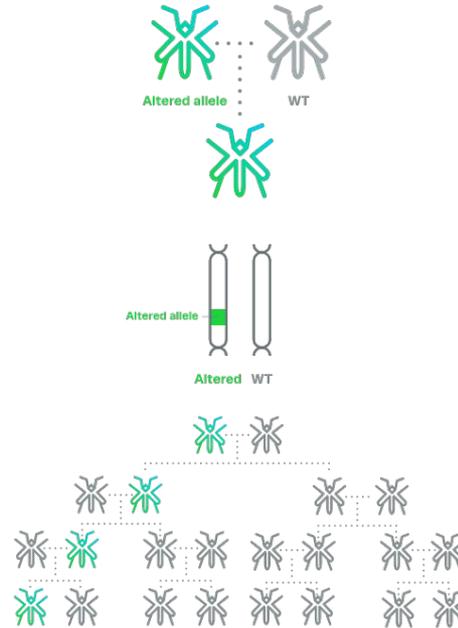


Gene Drive

Permite que los genes introducidos por la edición sean dominantes y tengan una tasa del 100% de probabilidad de transmitirse a descendientes

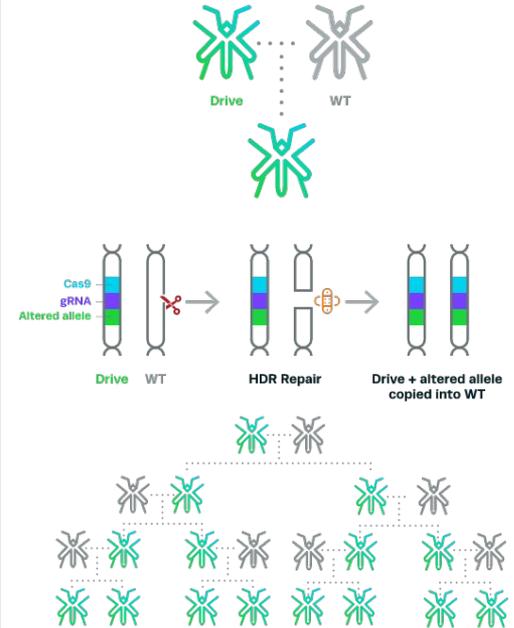
Además de modificar, también aporta el método de prevención

Altered gene spread by **normal inheritance**



50% chance of passing altered gene via normal inheritance

Altered gene spread by **gene drive**



>50% chance of passing altered gene via gene drive



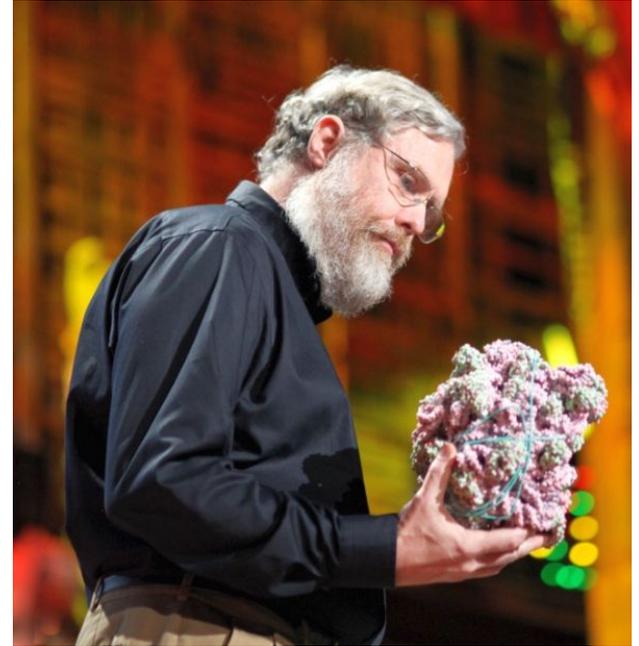
Casos

- Eliminar la transmisión de la malaria por parte de los mosquitos
- Erradicar especies no nativas (existen estudios sobre la erradicación de ratas, conejos y similares en islas por esterilización)
- Reintroducir la debilidad ante pesticidas de especies de insectos que han desarrollado inmunidad
- No hay muchos casos de aplicaciones pero sí planes a futuro

Mutaciones

- No solo se pueden evitar enfermedades
- Las 10 variantes genéticas “ventajosas y sin efectos secundarios” que George Church quiere meter en nuestros hijos con CRISPR

<https://www.xataka.com/medicina-y-salud/10-variantes-geneticas-ventajosas-efectos-secundarios-que-george-church-quiere-meter-nuestros-hijos-crispr>





Mutaciones

1. **MSTN**: común en atletas profesionales, y en mutaciones más raras de esta facilita el crecimiento de la masa muscular en hasta un 100%
2. **Gen del Crecimiento**: no siempre implica la ganancia de altura, pero si que reduce el riesgo de padecer cáncer
3. **CCR5**: resistencia al VIH
4. **PCSK9**: distintas mutaciones protegen ante enfermedades cardiovasculares hasta un 88%



Mutaciones

5. LRP5: mutación encontrada en una familia cuyos miembros nunca se habían roto un hueso y para operarlos se necesitaba un taladro quirúrgico especial
6. A673T: reduce el riesgo de contraer alzheimer hasta en un 80%
7. SCN9: insensibilidad congénita al dolor (no total). Se teme que pueda afectar a la empatía por el dolor ajeno.



Mutaciones

8. IFIH1: reducir su tamaño puede crear un efecto protector frente a la diabetes tipo 1
9. Transportador de Zinc SLC30A8: reducir el riesgo de diabetes tipo 2 un 65% reduciendo el tiempo en que este tarda en convertirse en ARN mensajero
10. ABCC11: Reduce de forma bastante importante el olor corporal



Cas9 vs Cpf1

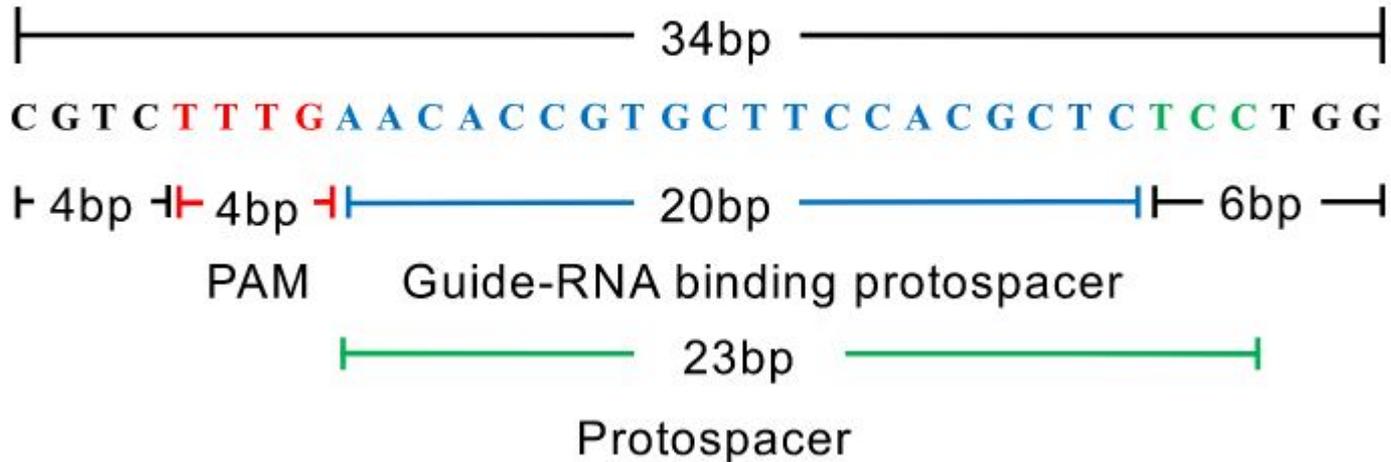
- La especificidad de CRISPR depende del motivo adyacente de protoespaciador (PAM).
- El PAM es una secuencia corta de ADN que tiene que aparecer tras el fragmento objetivo para que el corte se produzca.
- La secuencia PAM depende de la endonucleasa utilizada.
- Es importante tener disponibles otras endonucleasas porque, entre otras cosas, incrementan el número de PAMs distintas disponibles y por lo tanto el de secuencias objetivo.
- En la bacteria el PAM sirve para distinguir su propio ADN y el del virus.



Cas9 vs Cpf1

- Cas9: la secuencia PAM limitante es un trinucleótido 5'-NGG-3'
- Cpf1: la secuencia PAM limitante es 5'-TTTN-3'

CRISPR AsCpf1 Target Sequence



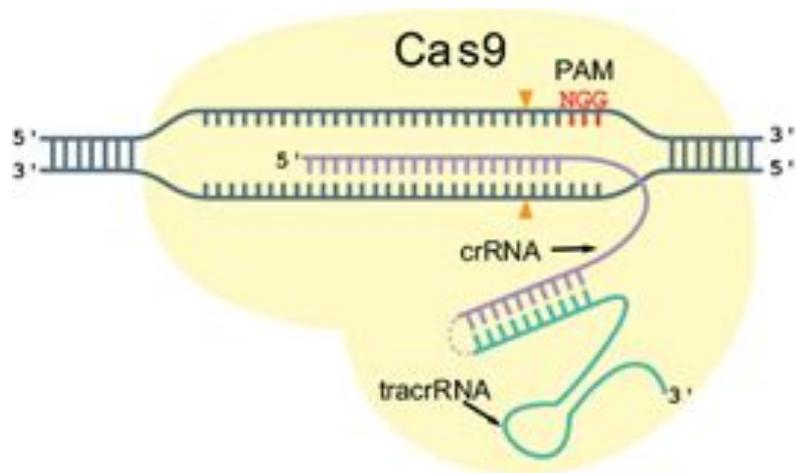


Cas9 vs Cpf1

El sistema CRISPR-Cas se divide en dos clases. Cas9 pertenece a la primera y Cpf1 a la segunda.

Las endonucleasas de primera clase requieren crRNA y tracrRNA mientras que las de la segunda solo necesitan el crRNA, por lo que son más sencillas.

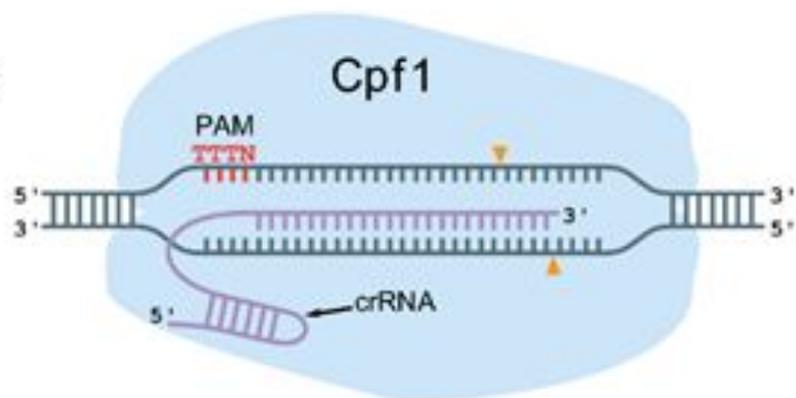
Además Cpf1 produce un corte escalonado en lugar del recto de Cas9. Este corte facilita la inserción posterior de ADN.



DSB



VS

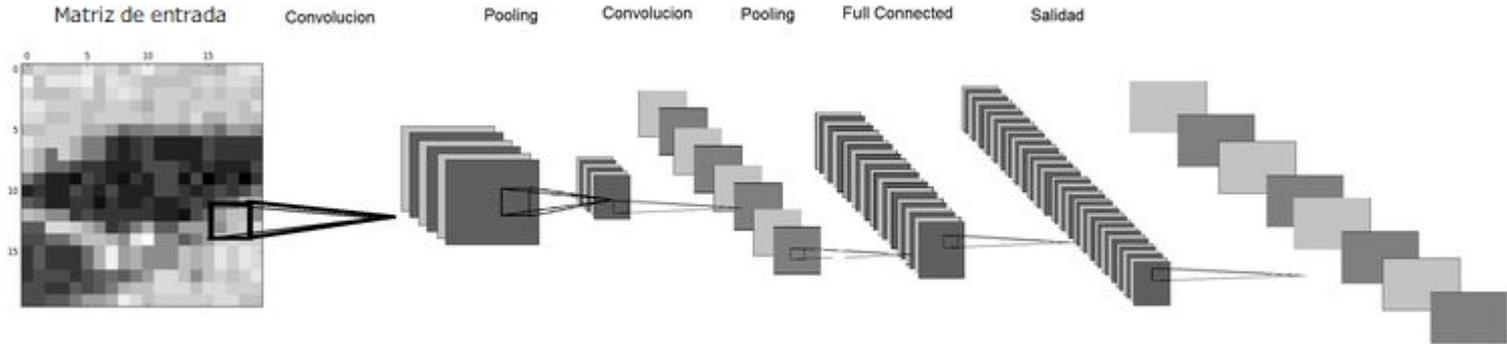


DSB



Redes Neuronales en la predicción de los efectos de un RNA guía

- ¿Por qué?: optimización
- ¿Como?: redes neuronales convolucionales



¿Por qué?

- Para conocer los resultados aplicar una edición genética sobre una célula con exactitud se necesita probar los efectos experimentalmente
- **Objetivo:** ahorrar el tiempo y recursos que se invierten en experimentación mediante un análisis previo





Redes Neuronales Convolucionales

CONVOLUCIONES

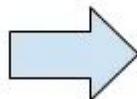
- De una matriz de entrada se obtiene otra imagen adicional por cada neurona de la capa de convolución resultado de aplicar el kernel a la imagen
- Se aplica una función de activación (RELU en este caso)

MATRIZ

		0,6	0,6		
	0,6			0,6	
	0,6	0,6	0,6	0,6	
	0,6			0,6	

KERNEL

1	0	-1
2	0	-2
1	0	-1



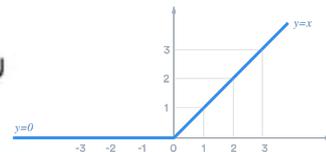
CONVOLUCION DEL KERNEL

-1,2	-0,6	0,6	1,2
-1,2	0,6	-0,6	1,2
-1,2	1,2	-1,2	1,2
-0,6	1,2	-1,2	0,6



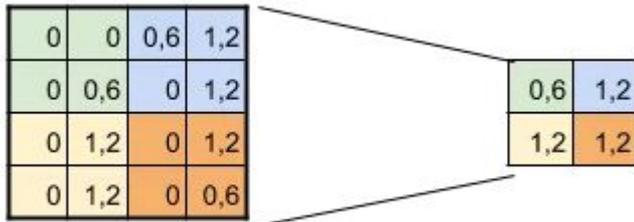
APLICO RELU

0	0	0,6	1,2
0	0,6	0	1,2
0	1,2	0	1,2
0	1,2	0	0,6

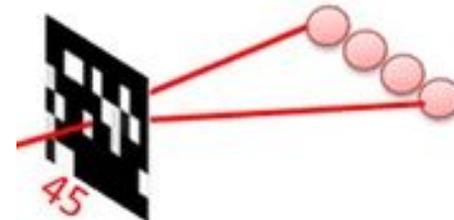


Redes Neuronales Convolucionales

POOLING Y CONEXIÓN CON UNA RED NEURONAL TRADICIONAL

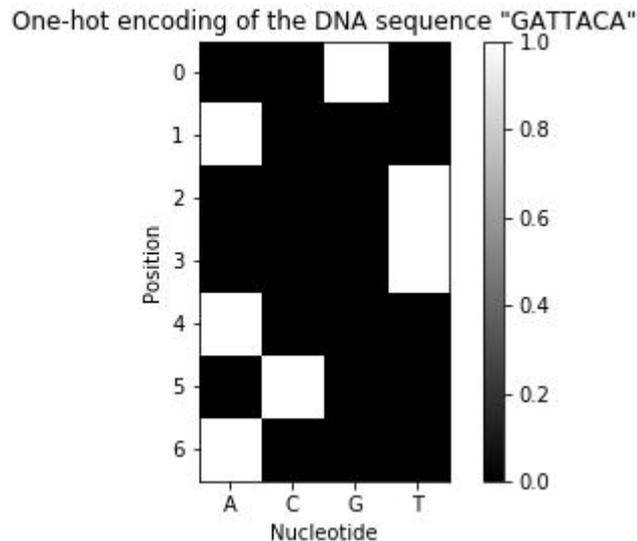


- Pooling: Se reduce el tamaño de la matriz obtenida al mismo tiempo que continúa guardando la información más relevante
- A cada matriz resultante se le hace corresponder un nuevo vector de neuronas que serán la entrada de la red tradicional



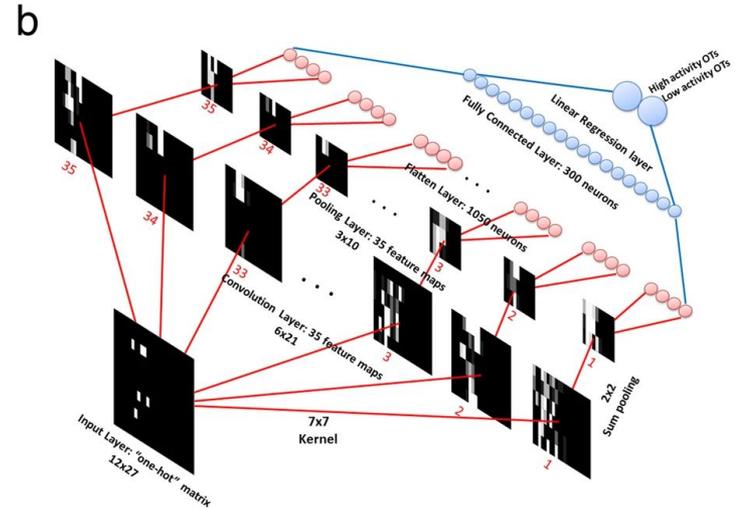
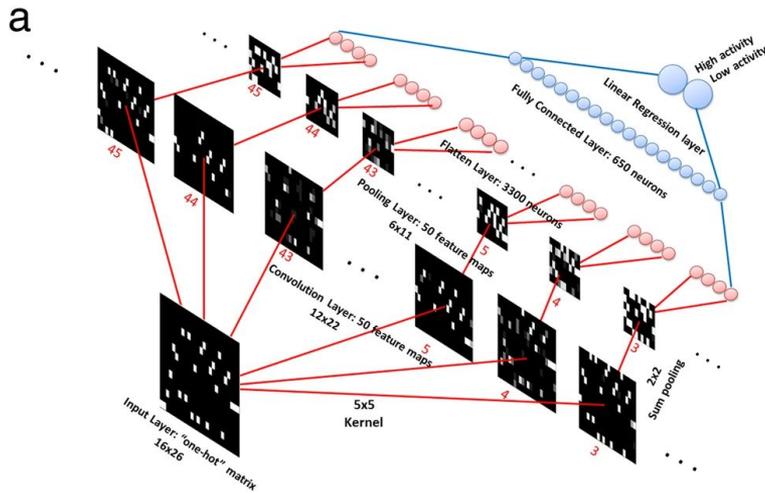
Transformación secuencia-matriz

- Se crea una nueva matriz en la que:
 - La línea indica la posición en la secuencia
 - La columna guarda la posición dentro de un alfabeto de nucleótidos (ej: A,C,G,T). 1 si el nucleótido en esa posición es el correspondiente a dicha columna, 0 en el caso contrario
- En nuestro caso: este “alfabeto” lo componen los 16 posibles pares entre A,C,G,T. De esta manera, una secuencia de 27 bp se convierte en una matriz 16x26:
 - GATTA = GA-AT-TT-TA



2 Redes diferentes:

- On-target CNN: éxito en la modificación de la secuencia "Diana" (a)
- Off-target CNN: mutaciones "fuera de diana" (b)



On-target CNN

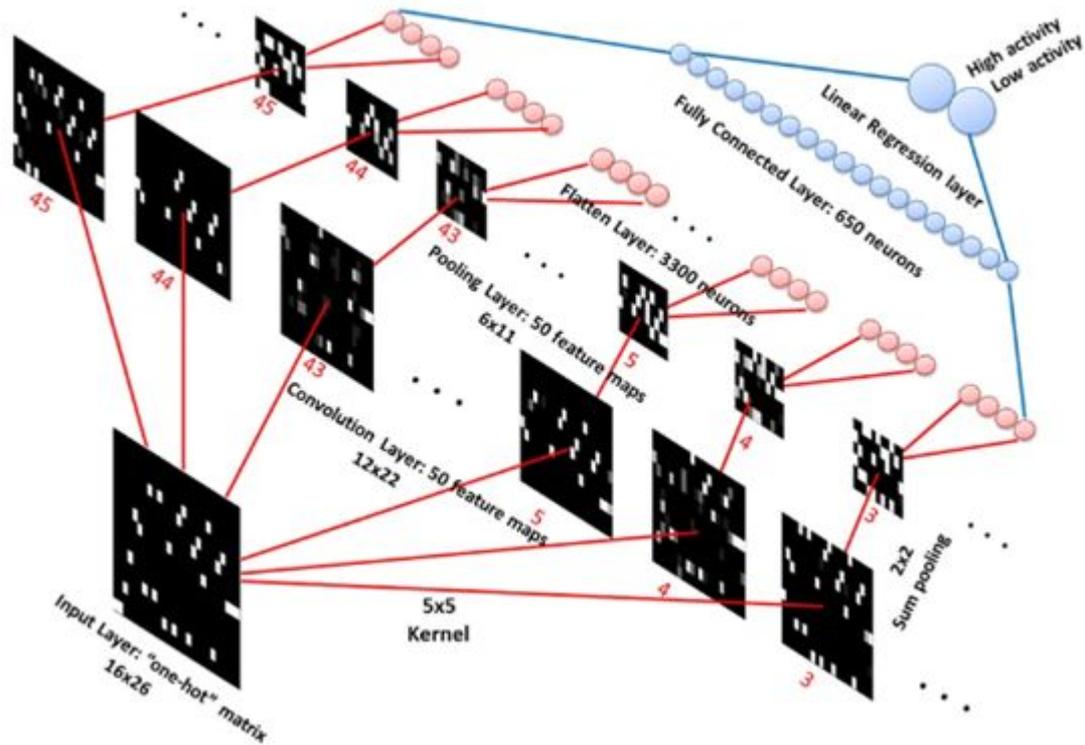
Input: matriz one-hot 16x26

Capa convolucional ejecuta 50 convoluciones con filtro 5x5, produce 50 feature maps 12x22.

Capa de pooling hace pooling espacial 2x2 utilizando la suma y produce 50 feature maps nuevos de tamaño 6x11.

Capa de flatten de 3300 neuronas reescala a un vector unidimensional.

Fully connected layer de 650 neuronas recibe la salida de la anterior y la introduce en una capa de regresión lineal que asigna una puntuación.



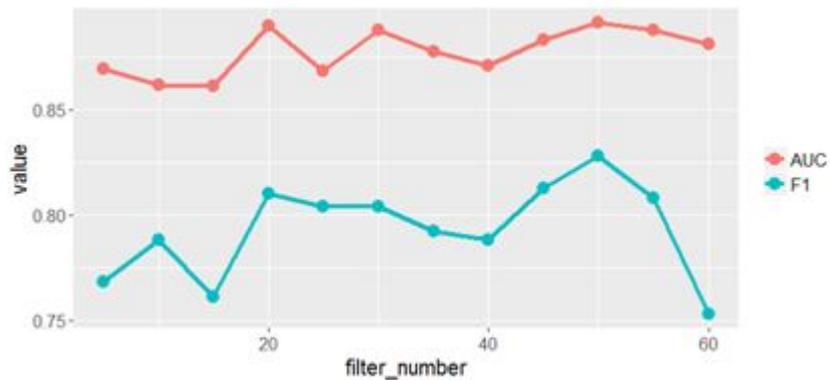


On-target CNN

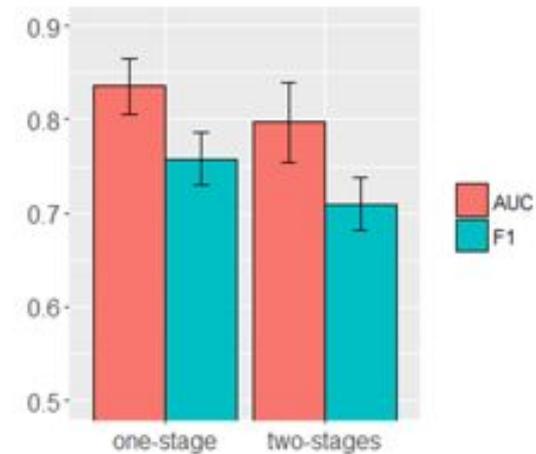
Para llegar a los resultados anteriores se estudiaron tres aspectos fundamentalmente: tamaño de kernel, número de feature maps y número de capas. Para ello se tuvieron en cuenta la F1 score y el AUC

- **Tamaño de kernel:** se estudiaron distintos tamaños manteniendo fijos los otros dos atributos, 5x5 fue el tamaño que mejores resultados arrojó.
- **Feature maps:** se mantuvieron fijos el tamaño de kernel y número de capas. Usando 50 feature maps se obtuvieron los mejores resultados.
- **Número de capas:** se comparó el rendimiento de dos escenarios, el primero constaba de una capa convolucional y una de pooling y el segundo de dos de cada. El rendimiento no mejoró al incrementar el número de capas.

On-target CNN



Número de feature maps



Número de capas

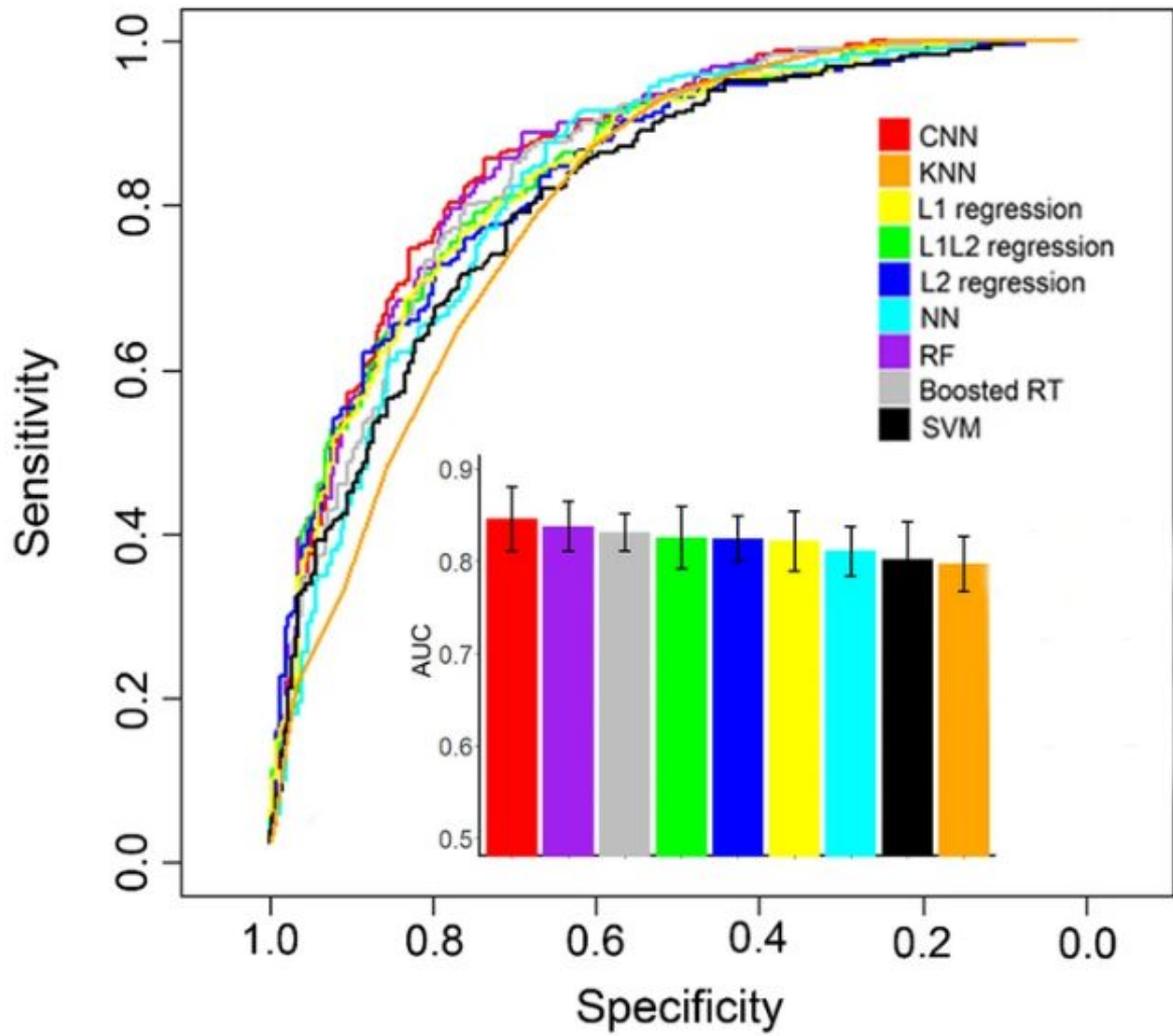


On-target CNN

Para la comprobación de resultados se utilizó validación cruzada con 5-fold. Además se comparó el rendimiento con el de otras técnicas de machine learning, concretamente:

- Red neuronal.
- K vecinos más próximos.
- Máquinas de soporte vectorial.
- Bosques aleatorios.
- Distintos tipos de regresión lineal regularizada .
- Árbol de regresión con potenciación del gradiente.

Adicionalmente se contrastaron los resultados con los obtenidos con CINDEL, una herramienta previamente existente con el mismo objetivo que DeepCpf1.



Off-target CNN

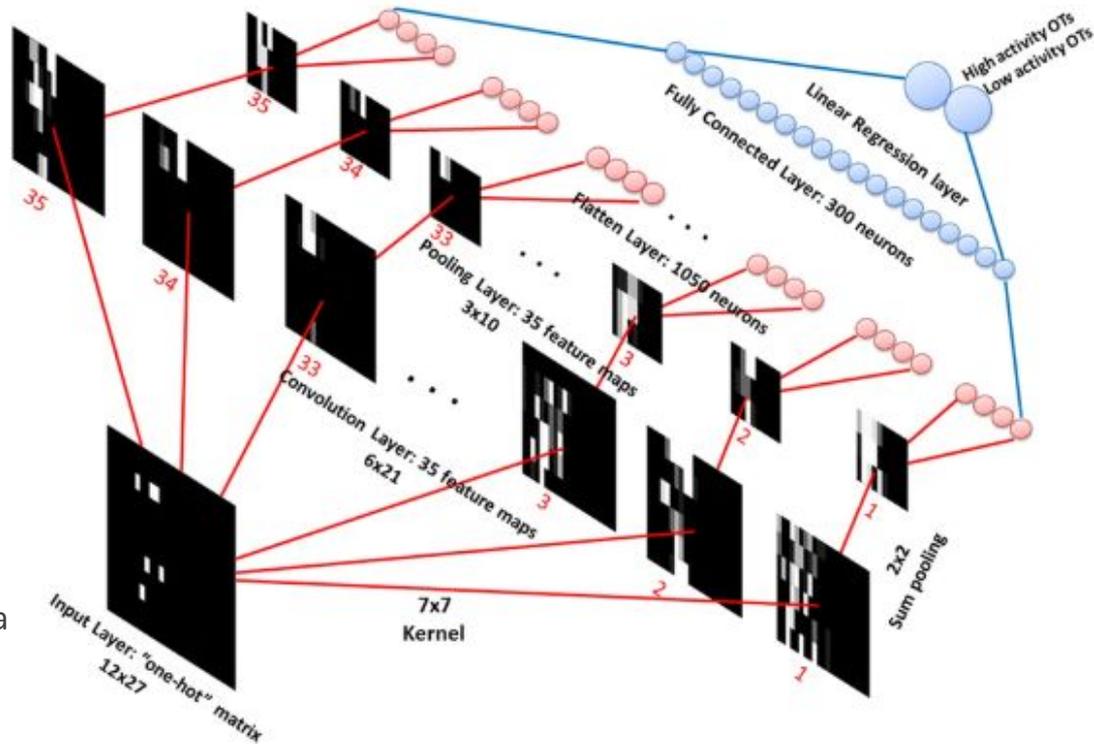
Input: matriz one-hot 12x27

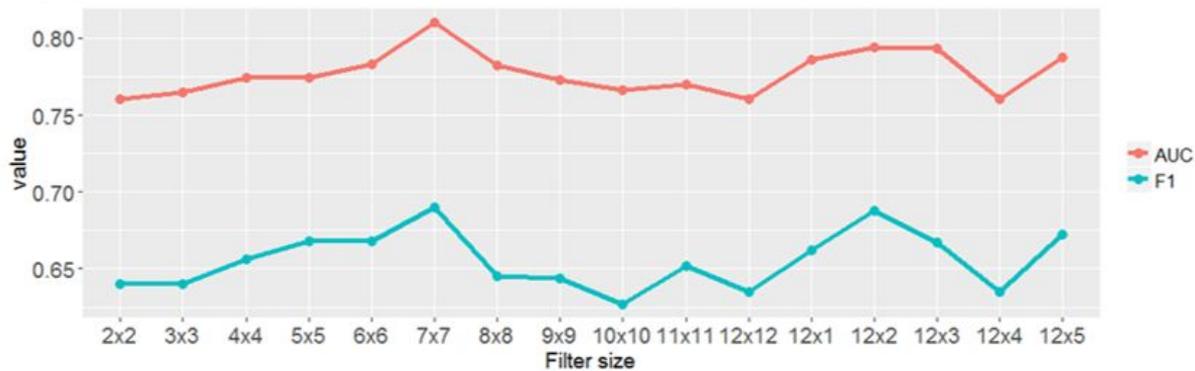
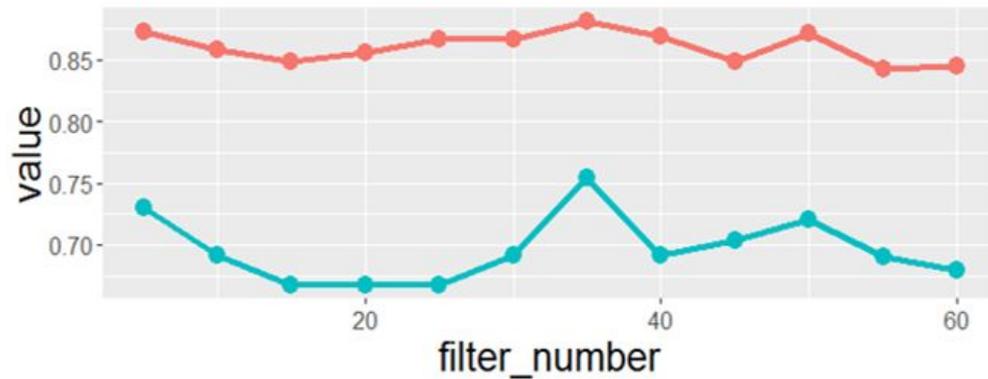
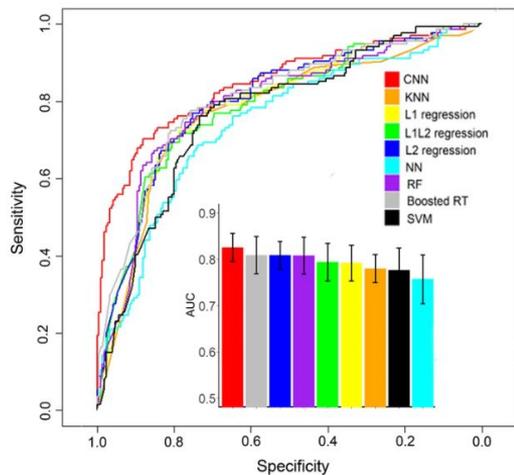
Capa convolucional ejecuta 35 convoluciones con filtro 7x7, produce 35 feature maps 6x21.

Capa de pooling hace pooling espacial 2x2 utilizando la suma y produce 35 feature maps nuevos de tamaño 3x10.

Capa de flatten de 1050 neuronas reescala a un vector unidimensional.

Fully connected layer de 300 neuronas recibe la salida de la anterior y la introduce en una capa de regresión lineal que asigna una puntuación.







Referencias

Información general sobre CRISPR-cas9:

- Explicación general y simple del CRISPR-cas9:
<https://ghr.nlm.nih.gov/primer/genomicresearch/genomeediting>
- Explicación general de un blog en español de divulgación científica:
<https://www.dciencia.es/que-es-la-tecnologia-crispr-cas9/> ,
<https://www.dciencia.es/novedades-sobre-crispr/>
- Más divulgación que menciona bastantes cosas, habla también de alternativas y el futuro de la técnica:
<https://www.investigacionyciencia.es/noticias/las-5-preguntas-ms-importantes-sobre-crispr-cas9-17711>
- Artículo sobre el cas9 menos divulgatorio que introduce la alternativa “Prime Editing”:
<https://www.rafer.es/innovacion-laboratorio-clinico/crispr-cas9-y-prime-editing/>
- Prime Editing:
https://genotipia.com/genetica_medica_news/prime-editing-nueva-herramienta-de-edicion-del-genoma/
- Artículo sobre mosquitos y gene drive:
<https://malariajournal.biomedcentral.com/articles/10.1186/s12936-019-2978-5>
- Más información sobre gene drive y la malaria: <https://www.synthego.com/blog/gene-drive-crispr>
- Diferencias entre cas9 y cpf1 y su aplicación en hongos:
<https://fungalbiolbiotech.biomedcentral.com/articles/10.1186/s40694-019-0069-6>



Referencias

Información sobre el CRISPR Cpf1 y redes neuronales:

- Introducción de CRISPR y comparación cas1 y cpf1:
<https://cellandbioscience.biomedcentral.com/articles/10.1186/s13578-019-0298-7>
- Uso de redes neuronales en CRISPR-cpf1:
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2939-6>
- Uno más concreto sobre lo mismo que el anterior:
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-3395-z>
- Explicación de uso de deep learning en cpf1 y comparación con otros métodos
<https://www.nature.com/articles/nbt.4061>

Otros:

- Introducción a redes neuronales convolucionales:
<https://www.aprendemachinelearning.com/como-funcionan-las-convolutional-neural-networks-vision-por-ordenador/>



FIN