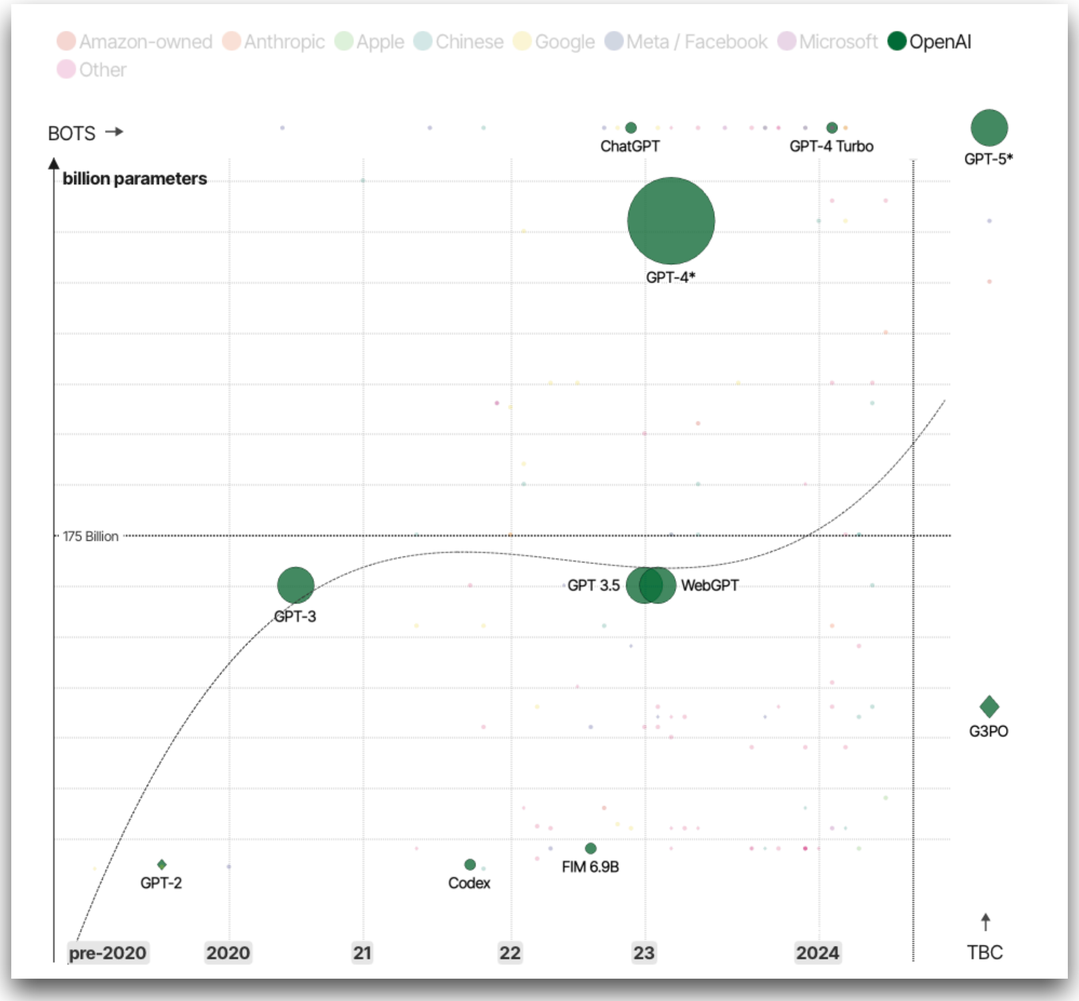
SHAP zero explains biological sequence models with near-zero marginal cost for future queries

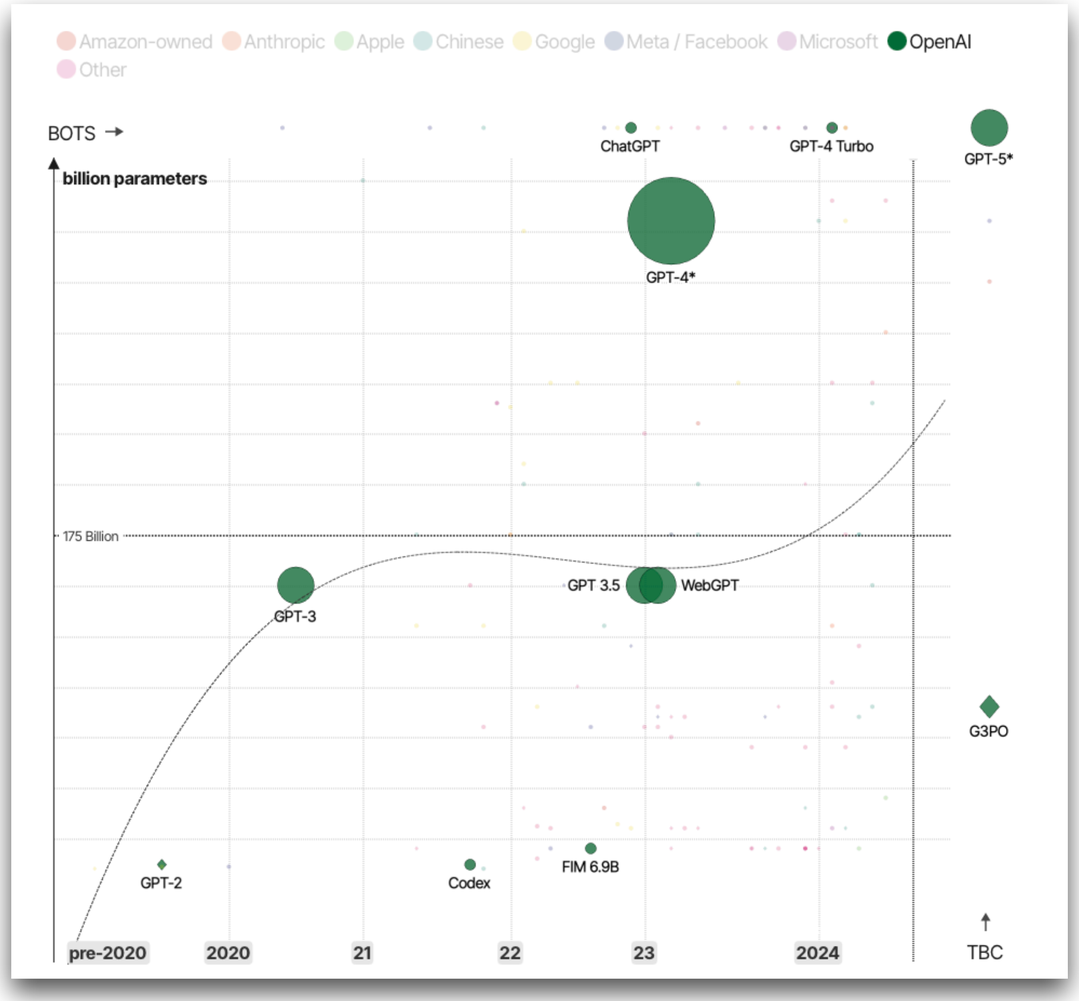
Darin Tsui

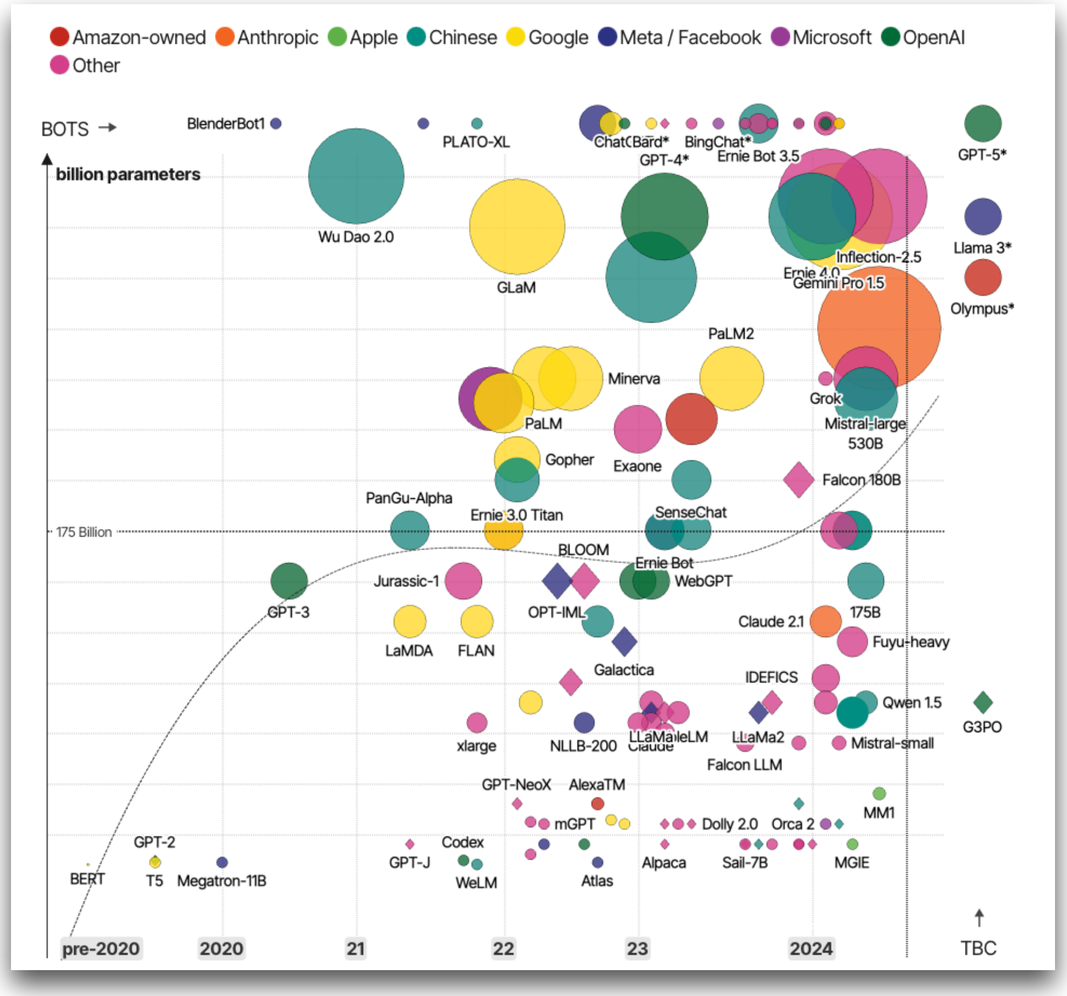
the rise and rise of AI



https://informationisbeautiful.net

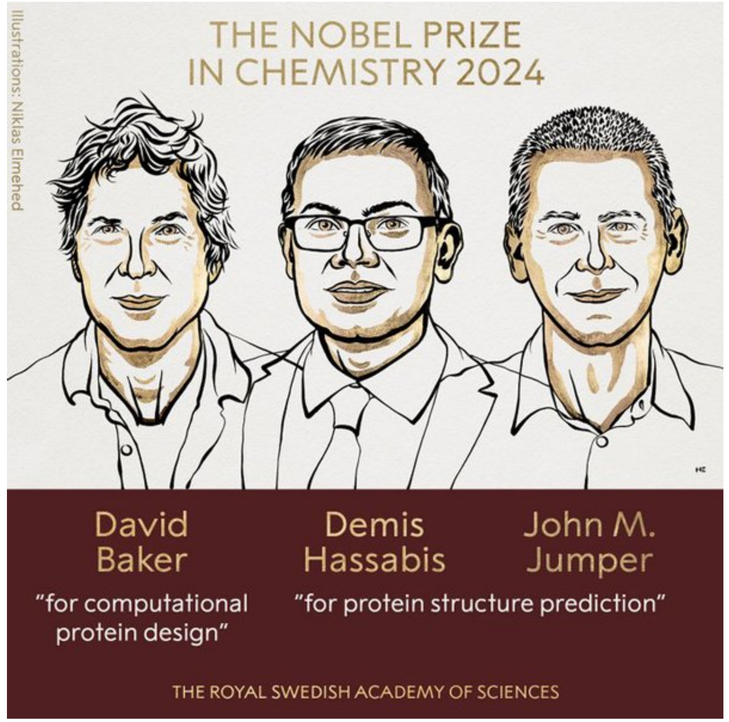
the rise and rise of AI



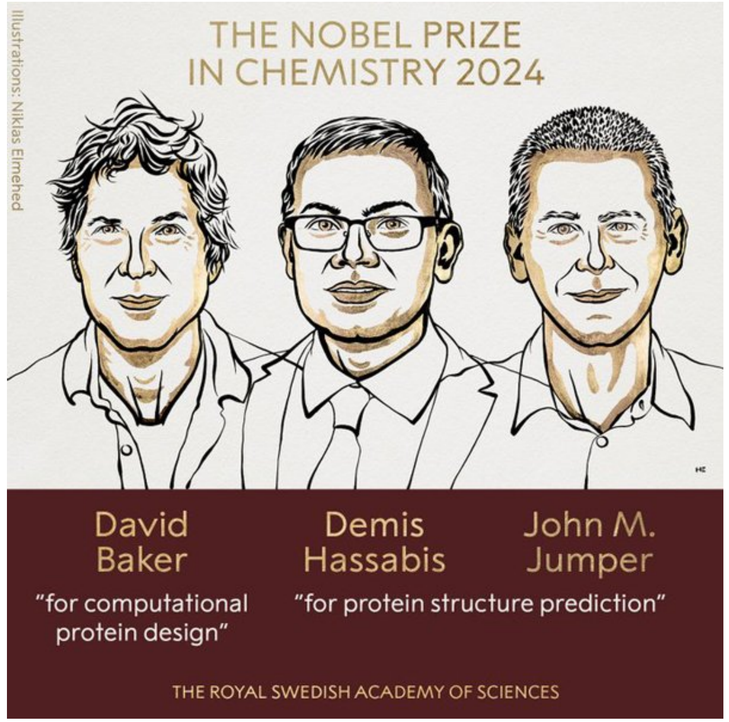


https://informationisbeautiful.net

AI breakthroughs in science



AI breakthroughs in science



JUNE 25, 2024//ANNOUNCEMENT

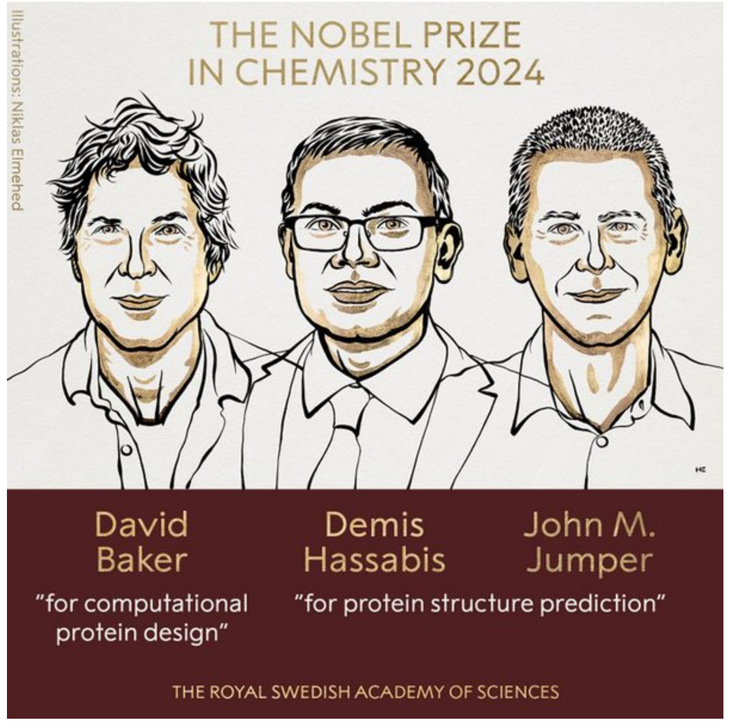
ESM3: Simulating 500 million years of evolution with a language model

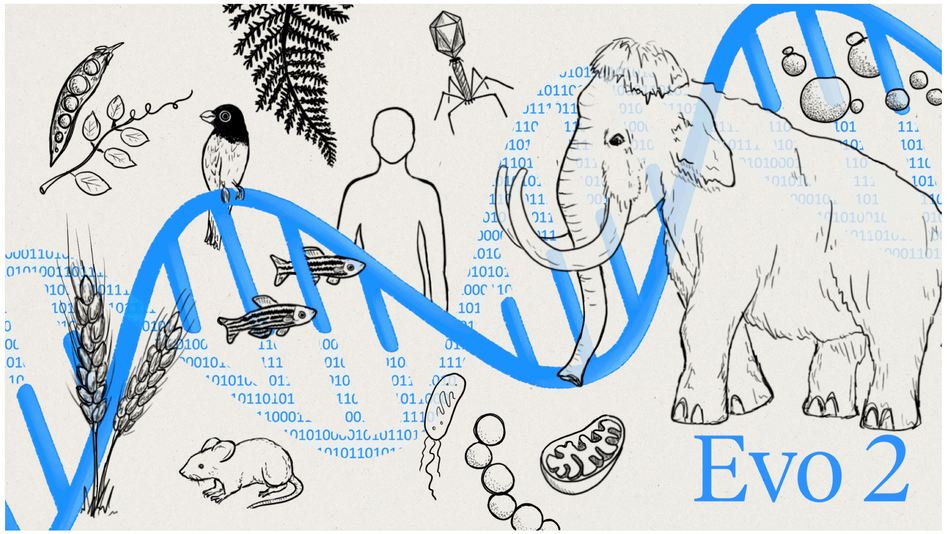
Preview our paper ↗

Trained on **2.78 billion** natural proteins

96 billion parameters

AI breakthroughs in science





Trained on **9.3 trillion** nucleotides
40 billion parameters

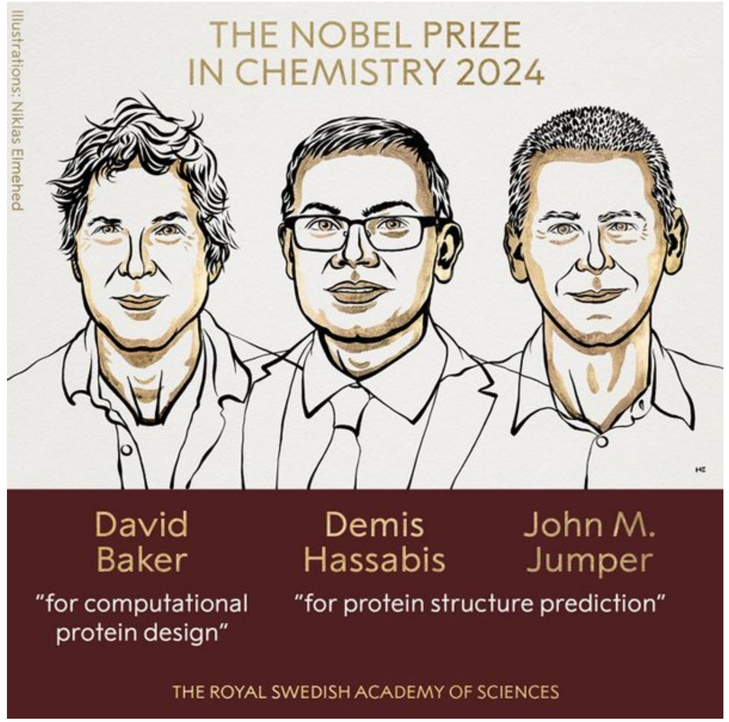
JUNE 25, 2024//ANNOUNCEMENT

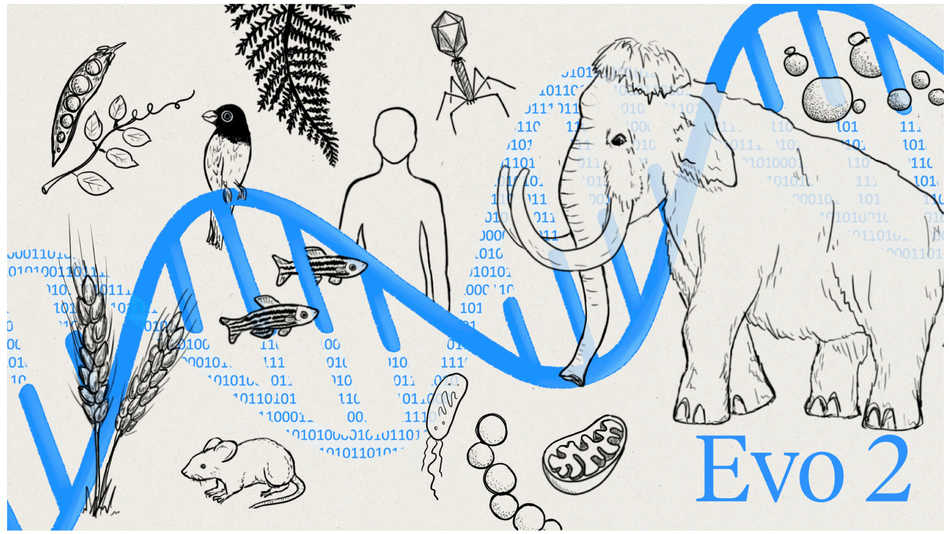
ESM3: Simulating 500 million years of evolution with a language model

Preview our paper

Trained on **2.78 billion** natural proteins
96 billion parameters

AI breakthroughs in science





Trained on **9.3 trillion** nucleotides
40 billion parameters

JUNE 25, 2024//ANNOUNCEMENT

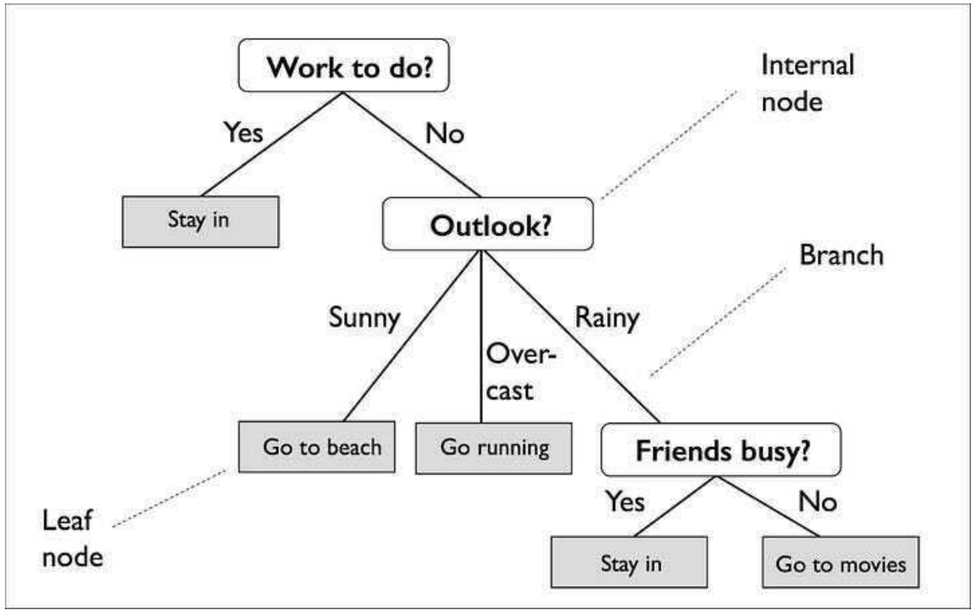
ESM3: Simulating 500 million years of evolution with a language model

Preview our paper ↗

Trained on **2.78 billion** natural proteins
96 billion parameters

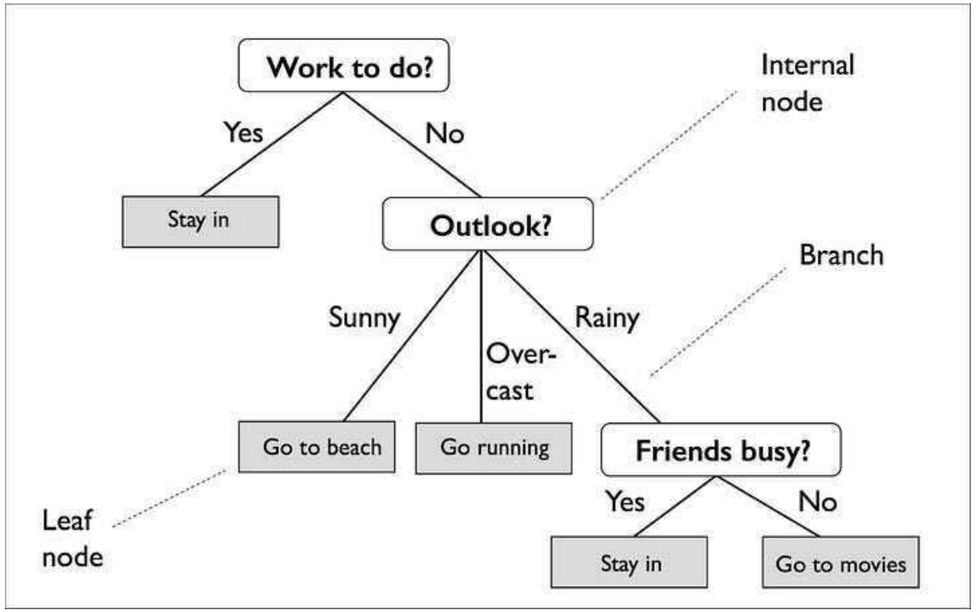
But does our power to **explain** them also **scale**?

interpretable machine learning models



dev.to, shap

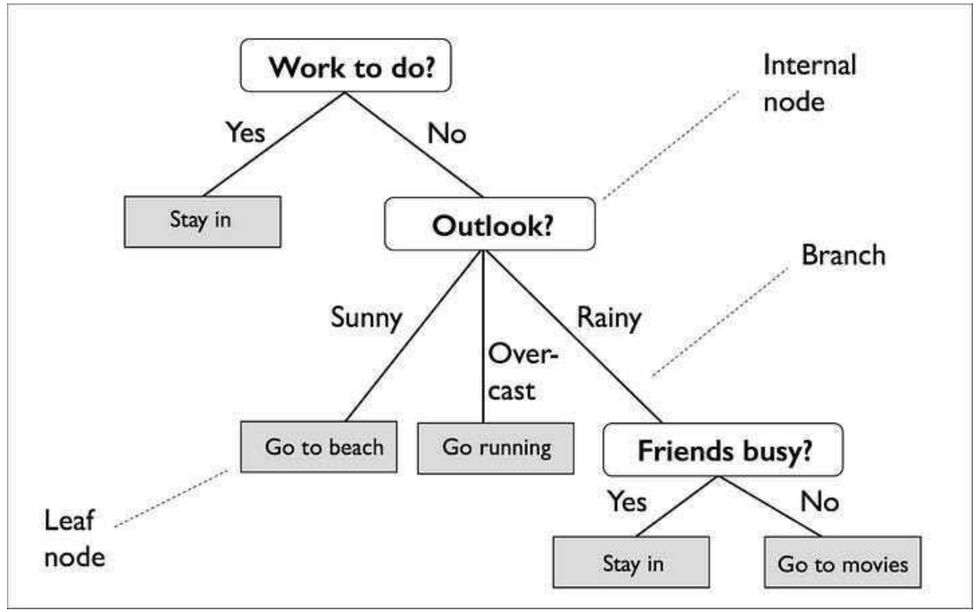
interpretable machine learning models



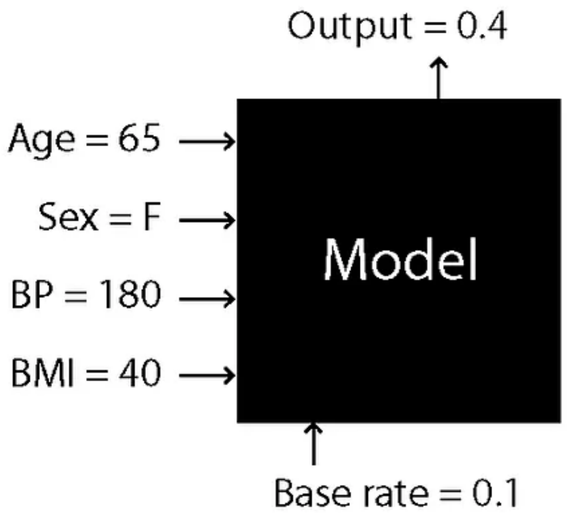
$$y = 3x_1 - x_2 + 2x_3$$

dev.to, shap

interpretable machine learning models

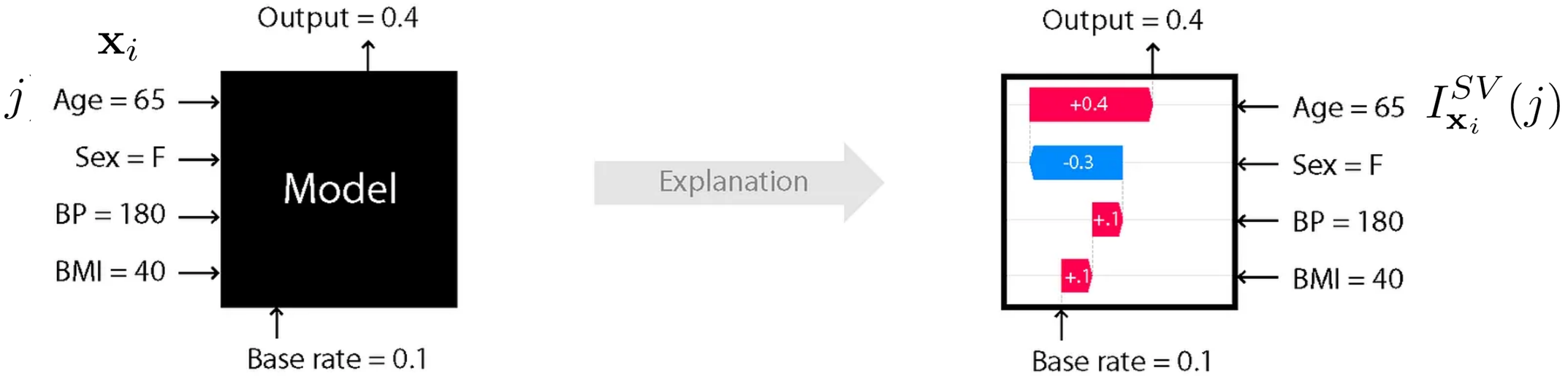


$$y = 3x_1 - x_2 + 2x_3$$



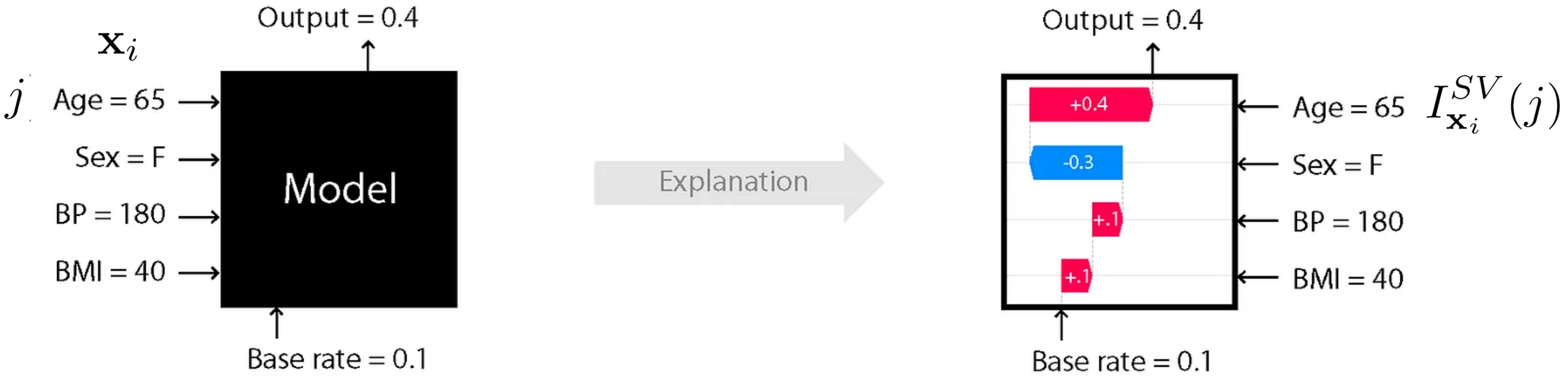
dev.to, shap

Shapley values



$$I^{SV}_{\mathbf{x}_i}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

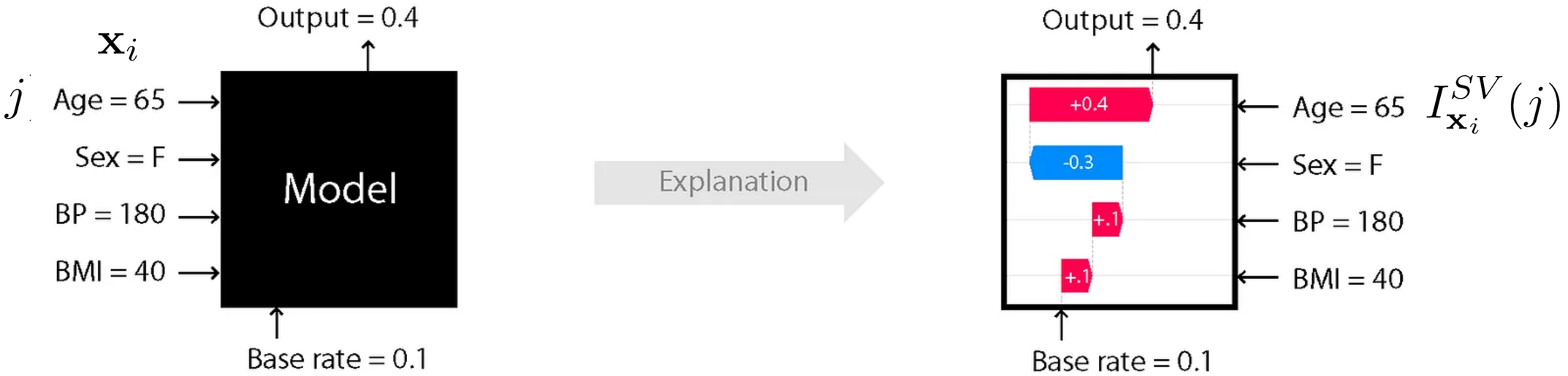
Shapley values



$$I^{SV}_{\mathbf{x}_i}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

$$\begin{aligned} I^{SV}_{\mathbf{x}_i}(j = Age) = & \frac{1}{4} \left[v_{\{Age\}}(\mathbf{x}_i) - v_{\emptyset}(\mathbf{x}_i) \right] \\ & + \frac{1}{12} \left[v_{\{Age,Sex\}}(\mathbf{x}_i) - v_{\{Sex\}}(\mathbf{x}_i) \right] \\ & \cdots \\ & + \frac{1}{4} \left[v_{\{Age,Sex,BP,BMI\}}(\mathbf{x}_i) - v_{\{Sex,BP,BMI\}}(\mathbf{x}_i) \right] \end{aligned}$$

Shapley values



$$I^{SV}_{\mathbf{x}_i}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

$$I^{SV}_{\mathbf{x}_i}(j = Age) = \frac{1}{4} \left[v_{\{Age\}}(\mathbf{x}_i) - v_{\emptyset}(\mathbf{x}_i) \right]$$

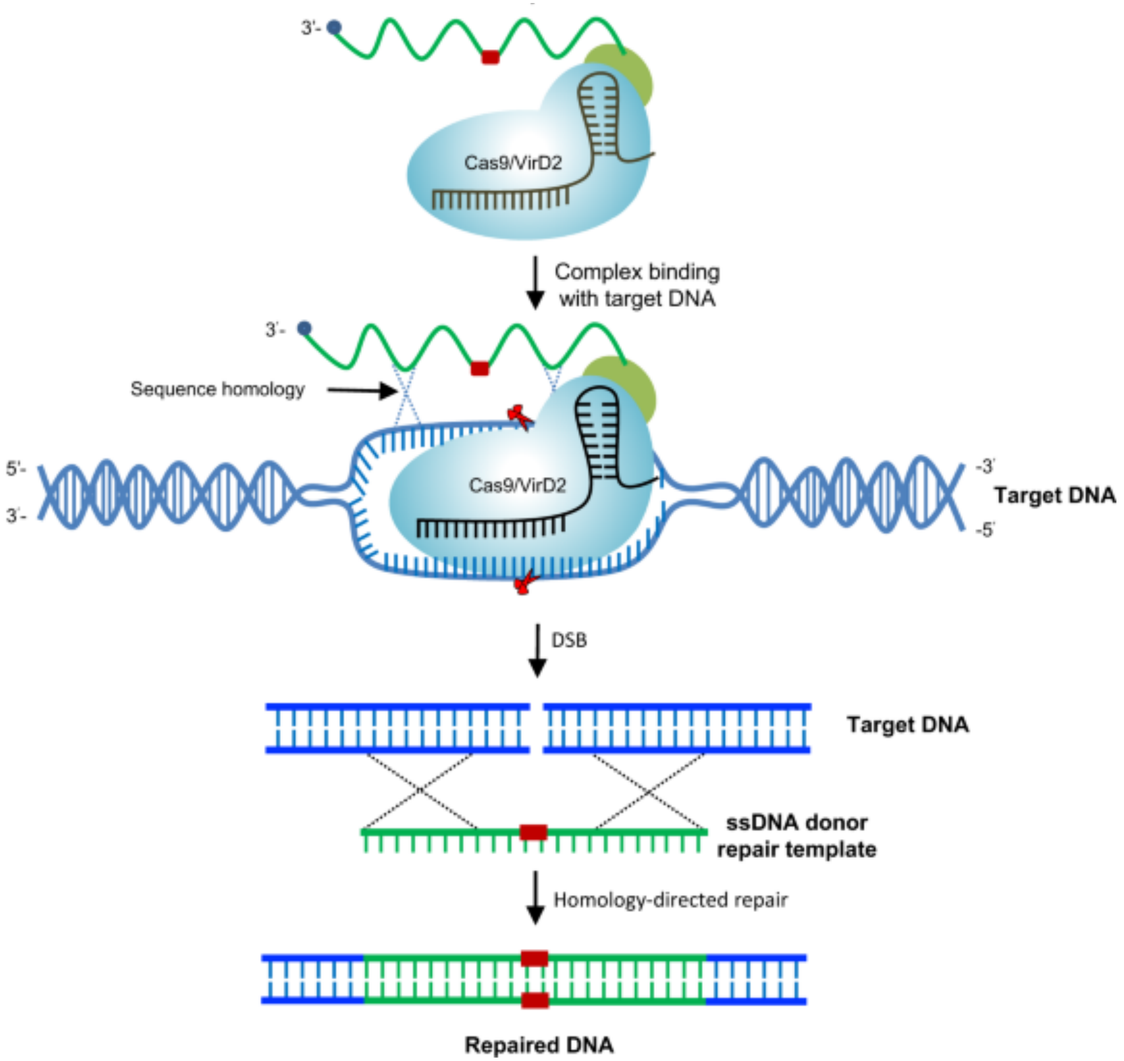
$$+ \frac{1}{1}$$

$$\dots$$

But how about explaining **many** sequences?

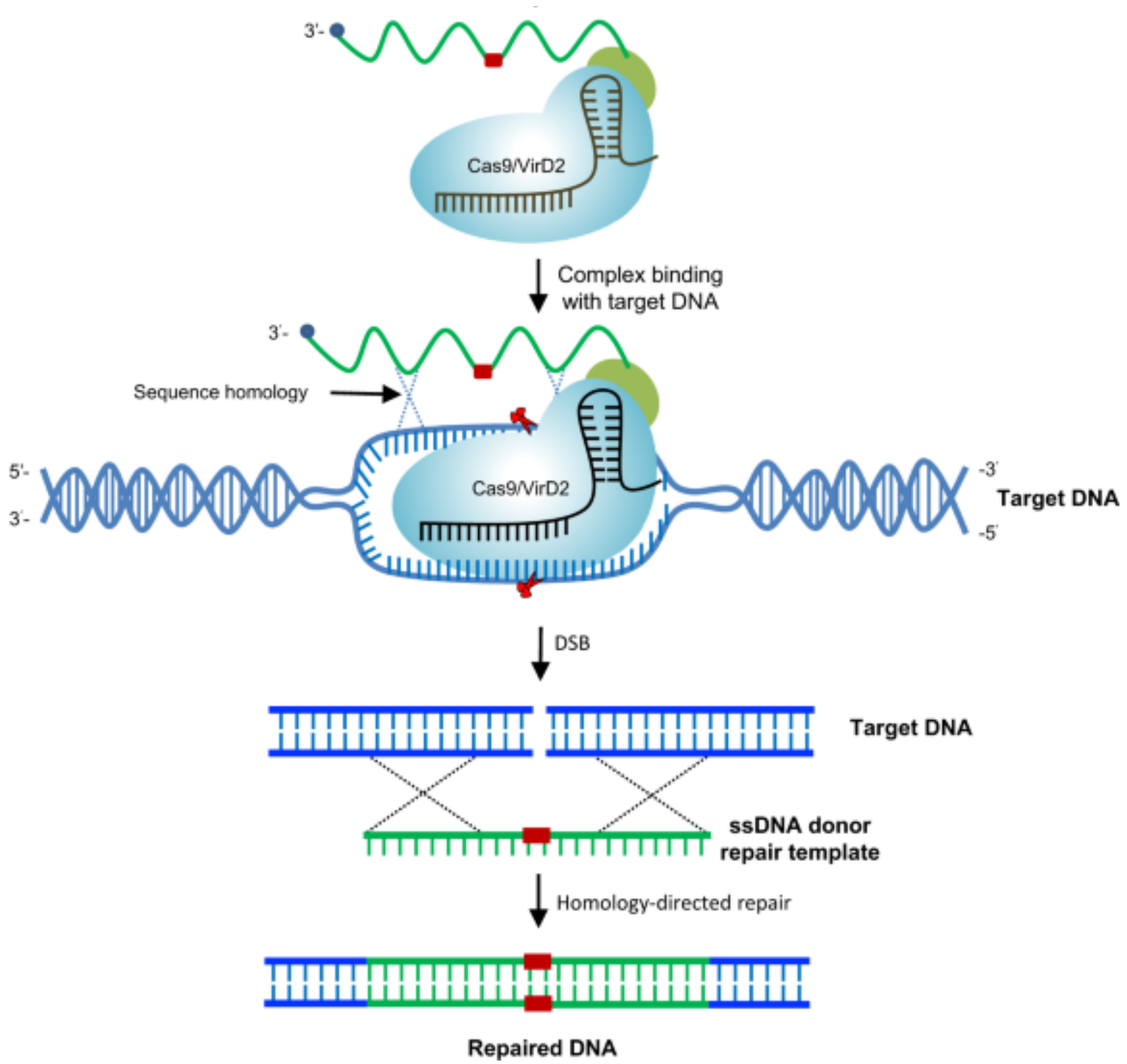
$$+ \frac{1}{4} \left[v_{\{Age,Sex,BP,BMI\}}(\mathbf{x}_i) - v_{\{Sex,BP,BMI\}}(\mathbf{x}_i) \right]$$

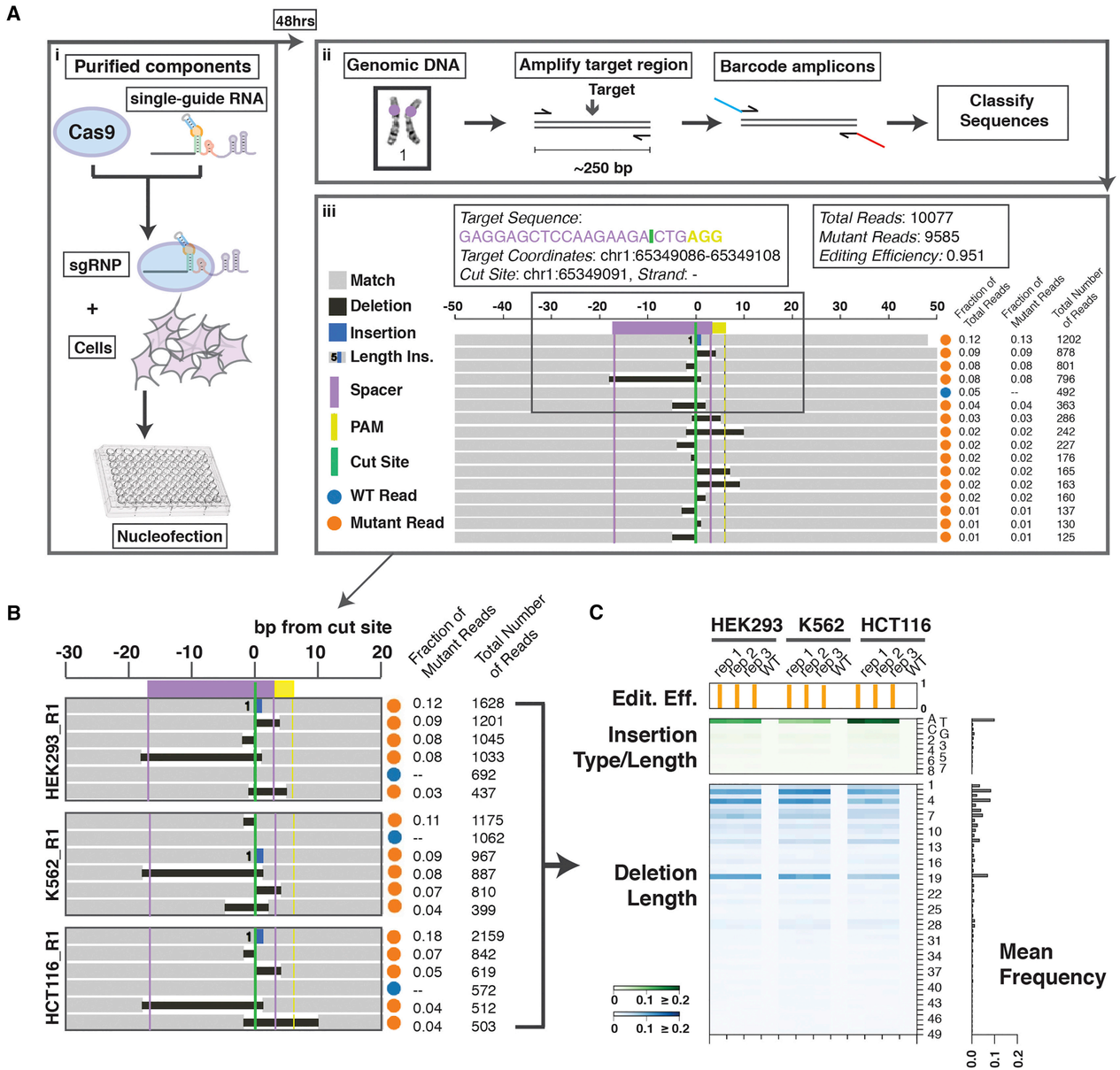
we want to explain biological models over <u>many</u> sequences



Ali *et. al*, *Communications Biology* (2020), van Overbeek *et. al*, *Molecular Cell* (2016)

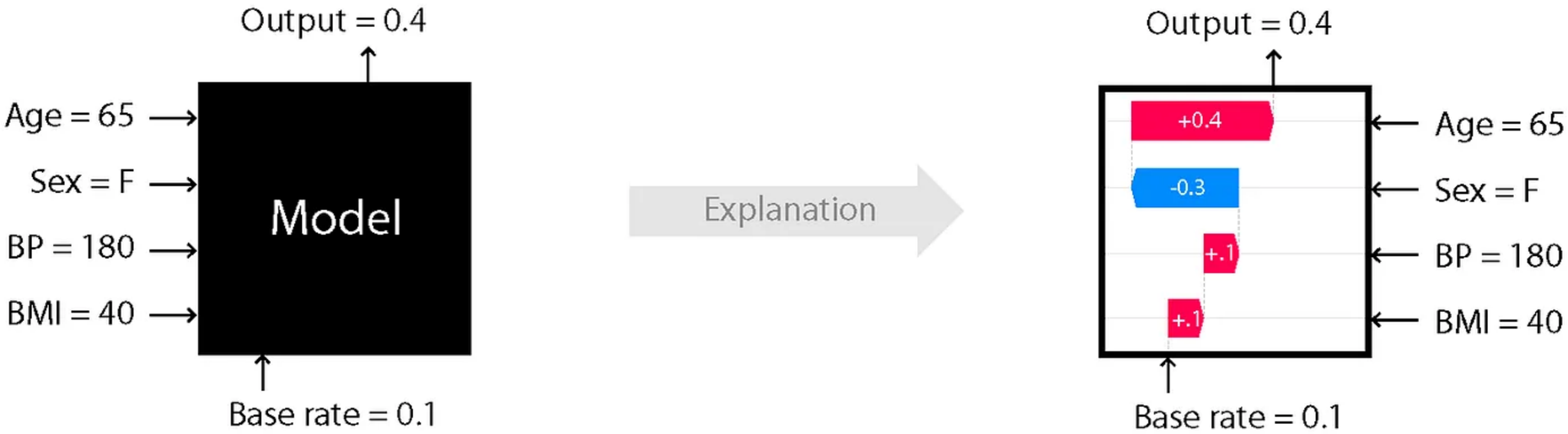
we want to explain biological models over <u>many</u> sequences





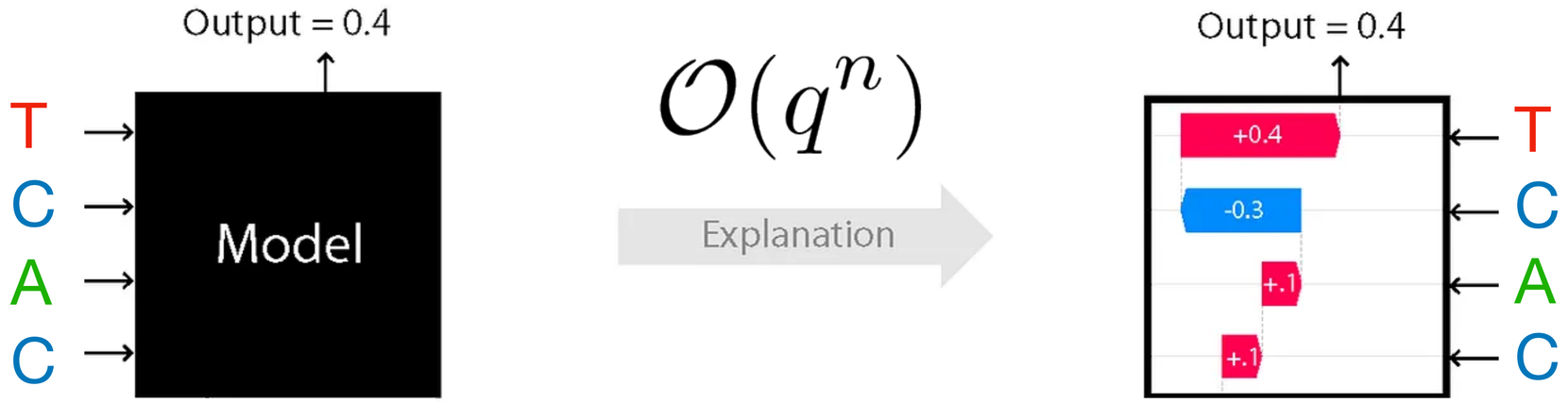
Ali *et. al*, *Communications Biology* (2020), van Overbeek *et. al*, *Molecular Cell* (2016)

we want to explain biological models over <u>many</u> sequences



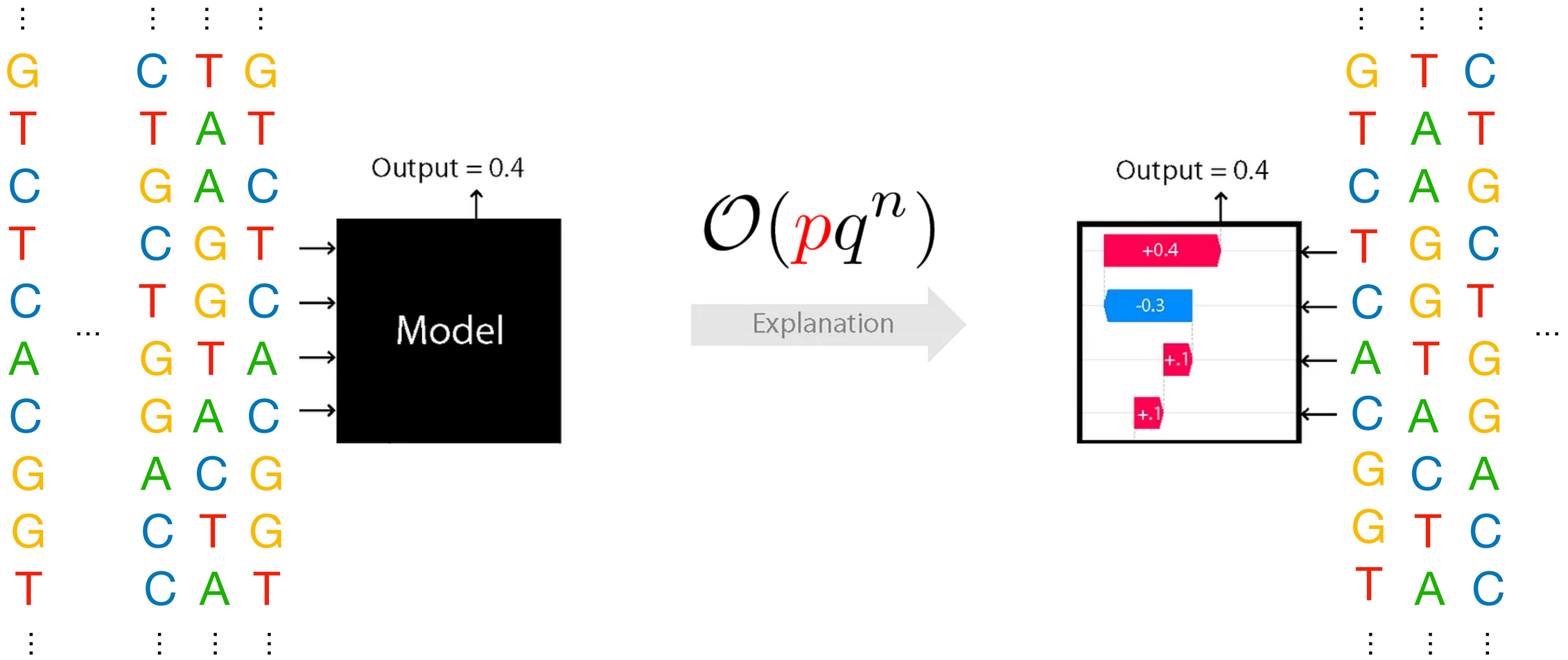
$$I^{SV}_{\mathbf{x}_i}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

we want to explain biological models over <u>many</u> sequences



$$I^{SV}_{\mathbf{x}_i}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

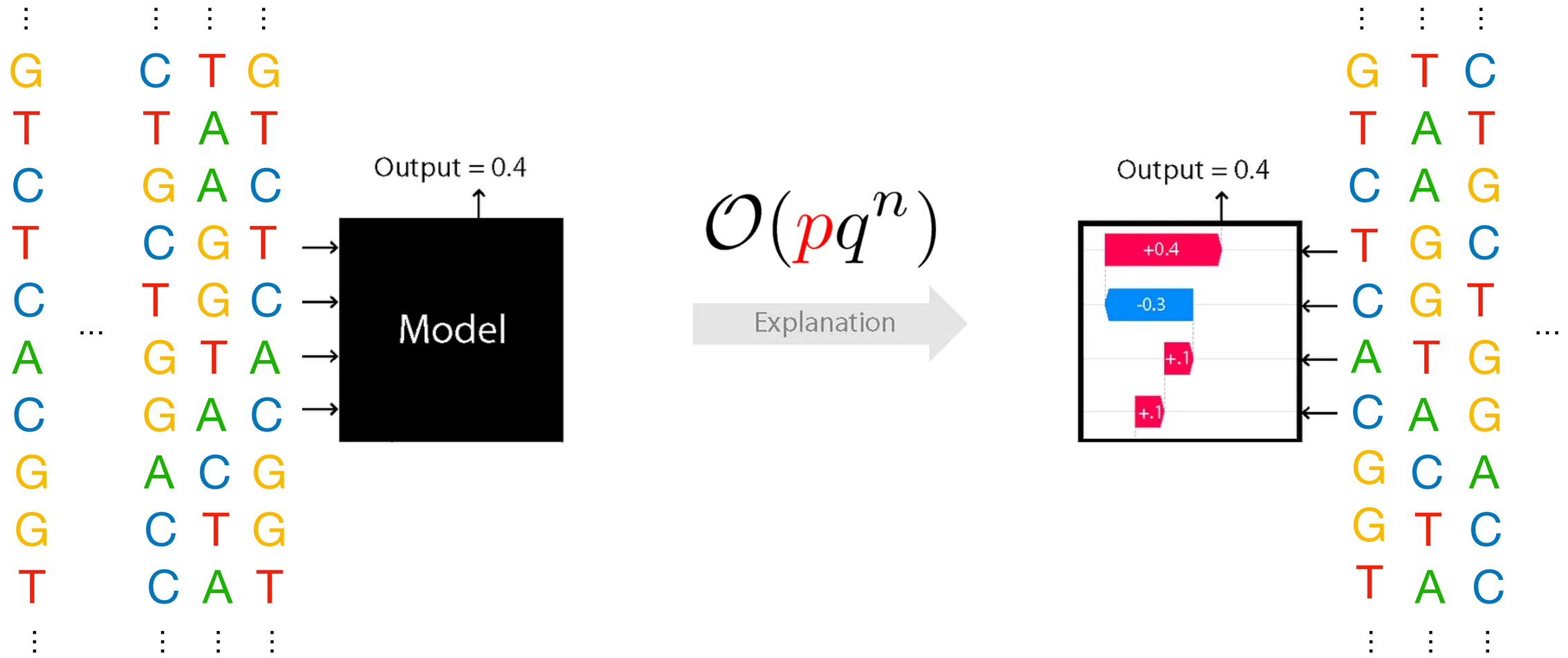
we want to explain biological models over <u>many</u> sequences



$$I_{\mathbf{x}_i}^{SV}(j) = \sum_{T \subseteq D \setminus \{j\}} \frac{|T|!\,(|D| - |T| - 1)!}{|D|!} \left[v_{T \cup \{j\}}(\mathbf{x}_i) - v_T(\mathbf{x}_i) \right]$$

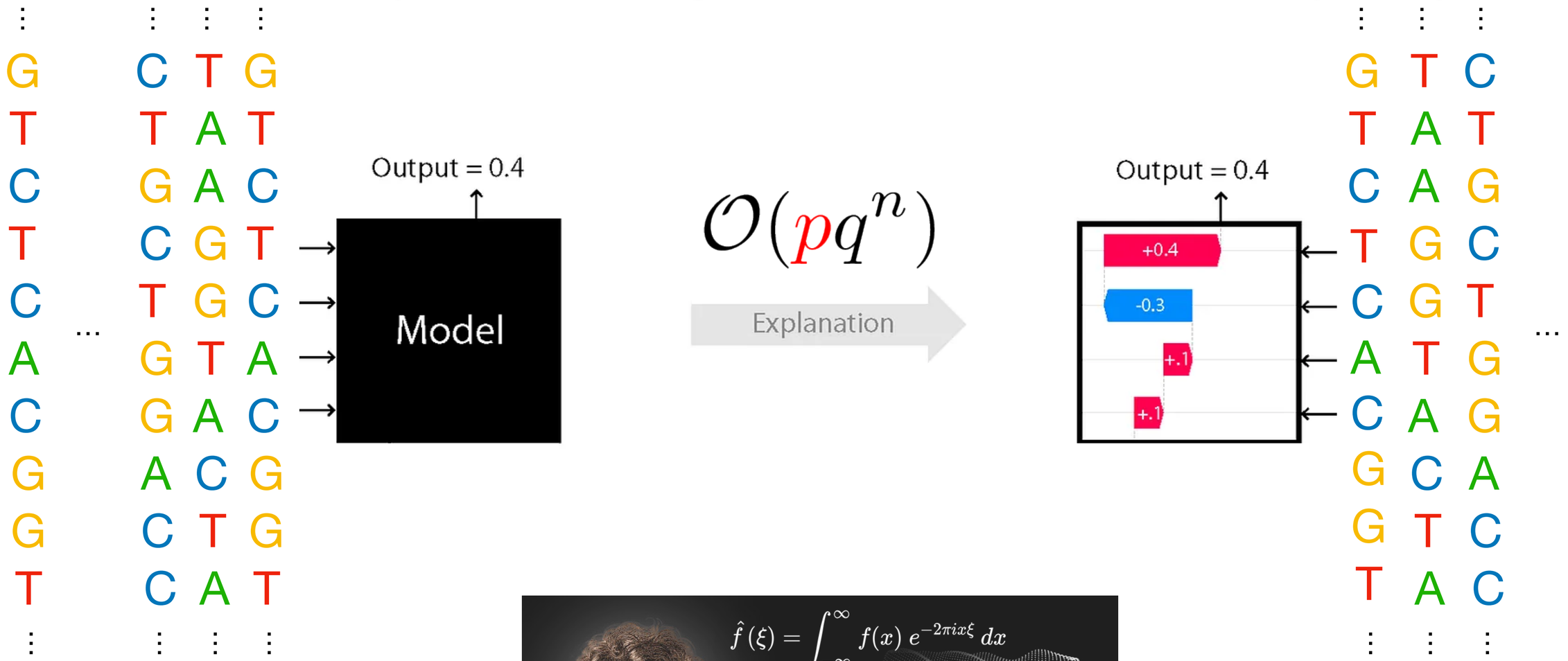
shap

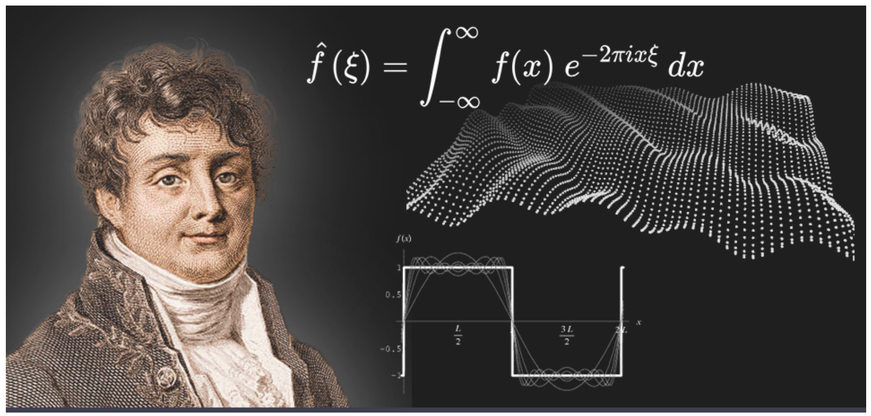
we want to explain biological models over <u>many</u> sequences



Explaining 1038 sequences
Faith-Shap ~ 81 days!

we want to explain biological models over <u>many</u> sequences

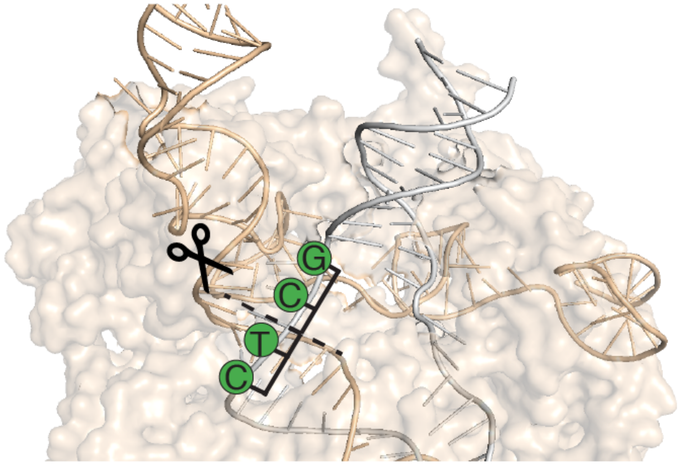




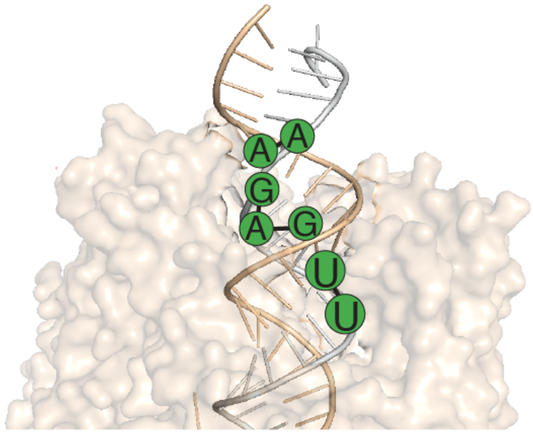
shap

biological sequence models are fundamentally sparse

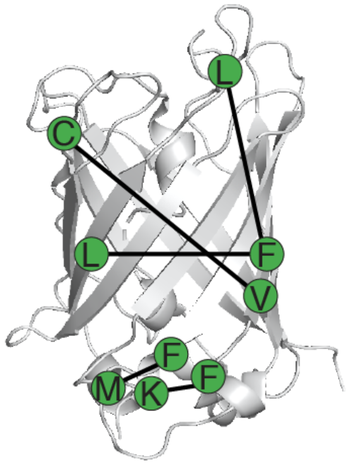
a Interactions in DNA repair



b Interactions in guide RNA binding



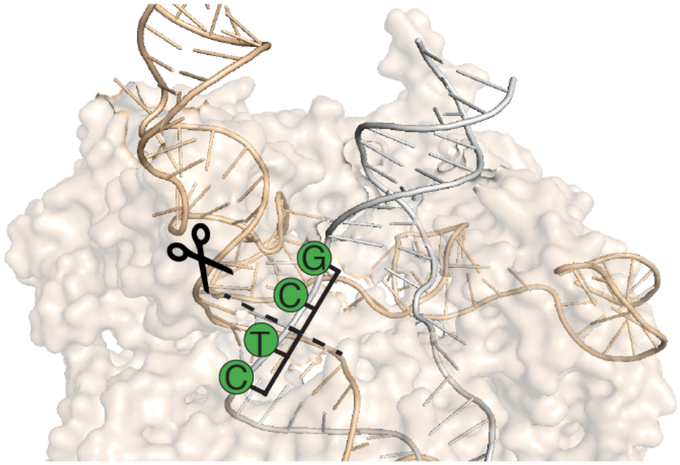
c Interactions in protein function



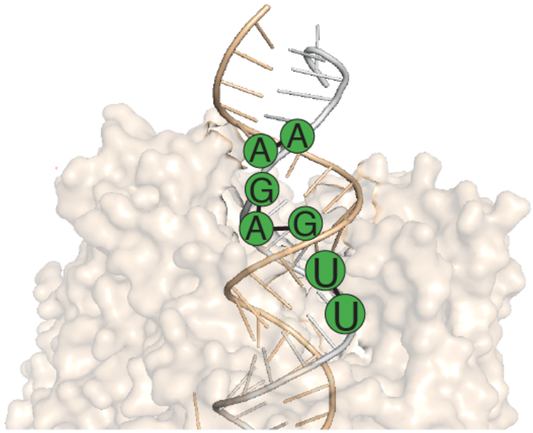
Kang, **Tsui**, Erginbas, Butler, Aghazadeh, Ramchandran, *under review* (2025), **Tsui**, Aghazadeh, ICLR GEM Bio (2024)

biological sequence models are fundamentally sparse

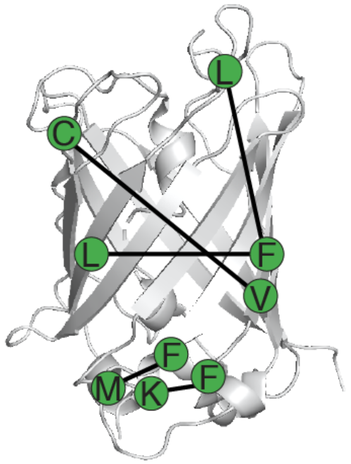
a Interactions in DNA repair

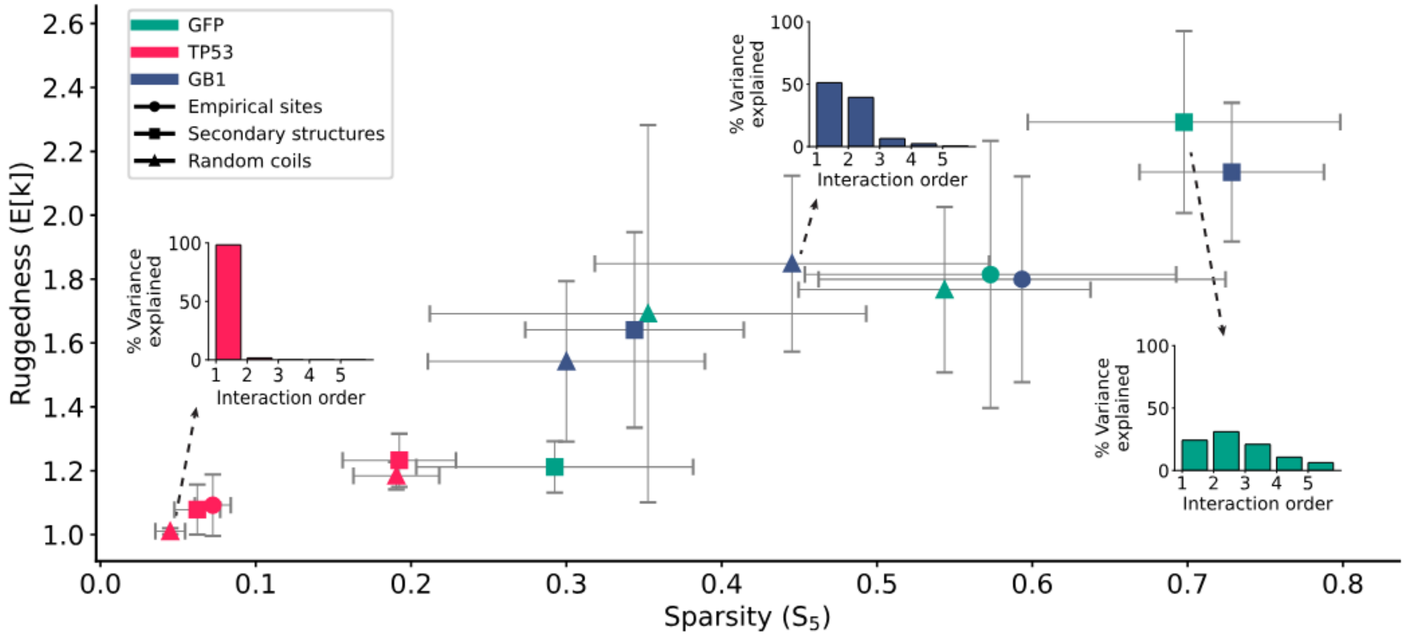


b Interactions in guide RNA binding



c Interactions in protein function





Kang, **Tsui**, Erginbas, Butler, Aghazadeh, Ramchandran, *under review* (2025), **Tsui**, Aghazadeh, ICLR GEM Bio (2024)

biological sequence models as *q*-ary functions

$$f : \mathbb{Z}_q^n$$

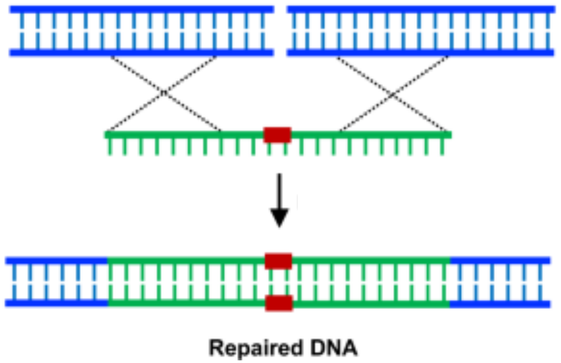
$$f([x_1, x_2, \ldots, x_n]) \rightarrow y$$

$$x_i \in \{0, 1, \ldots, q-1\}$$

biological sequence models as *q*-ary functions

$$f : \mathbb{Z}_4^n$$

$$f([x_1, x_2, \ldots, x_n]) \rightarrow$$



$$x_i \in \{A, C, G, T\}$$

Ali *et. al*, *Communications Biology* (2020)

biological sequence models as *q*-ary functions

$$f : \mathbb{Z}_{20}^{n}$$

$$f([x_1, x_2, \ldots, x_n]) \rightarrow$$

$$x_i \in \{P, L, \ldots, Y\}$$

biological sequence models as *q*-ary functions

$$f : \mathbb{Z}_{20}^{n}$$

$$f([x_1, x_2, \ldots, x_n]) \rightarrow$$

$$x_i \in \{P, L, \ldots, Y\}$$

$$f(\mathbf{m}) = \sum_{\mathbf{y} \in \mathbb{Z}_q^n} F[\mathbf{y}] \omega^{\langle \mathbf{m}, \mathbf{y} \rangle} \quad \omega = e^{\frac{2\pi j}{q}}$$

SHAP zero amortizes Shapley explanations

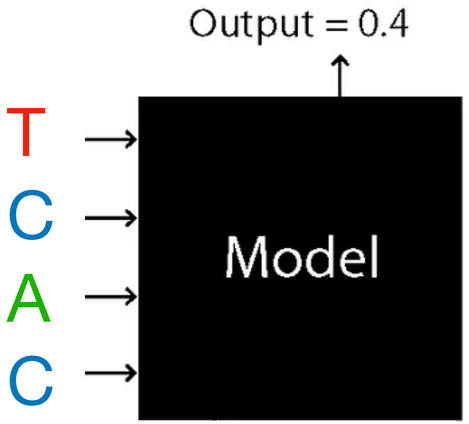
Algorithm. SHAP zero in three steps:

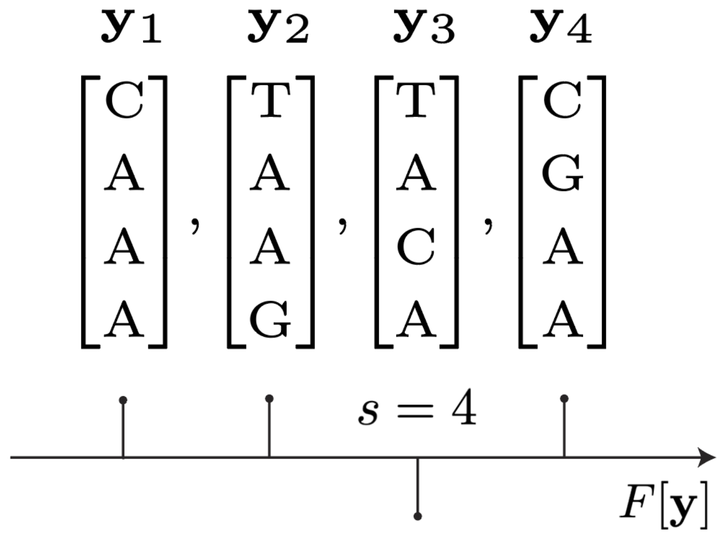
- **Step 1**: Pay a **one-time cost** to sketch the model globally via the Fourier transform with **sample complexity** $\mathcal{O}(sn^2)$ and **computational complexity** $\mathcal{O}(sn^3)$

For each future sequence:

- **Step 2**: Map global sketch to the local Möbius transform to isolate interacting features with complexity $\mathcal{O}(s^2(2q)^\ell)$
- **Step 3**: Map Möbius transform to Shapley values and interactions with complexity $\mathcal{O}(s^2(2q)^\ell)$

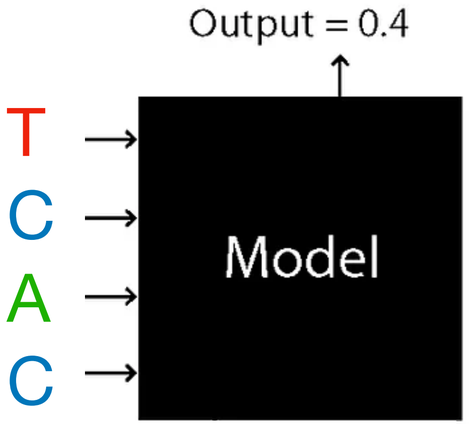
step 1: pay a one-time cost to sketch the model globally





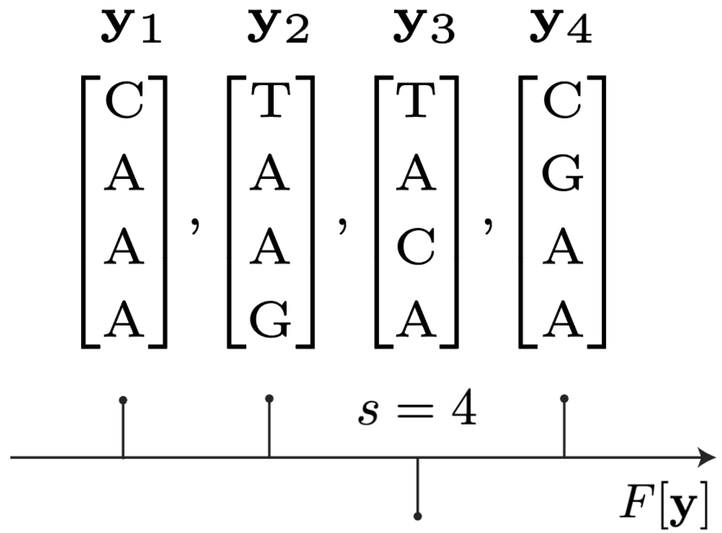
Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 1: pay a one-time cost to sketch the model globally



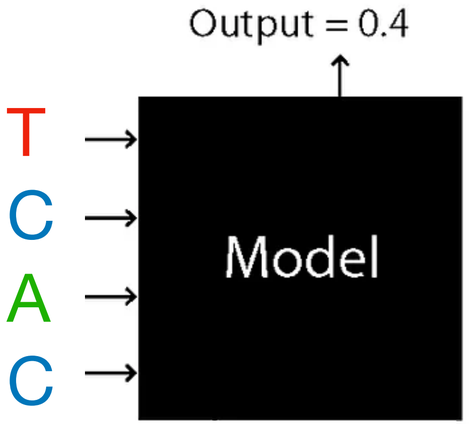
Subsampling

$$f(\mathrm{AAAA}) \quad f(\mathrm{AAAA})$$
$$f(\mathrm{ACAA}) \quad f(\mathrm{AAAC})$$
$$\ldots \qquad\qquad \ldots$$
$$f(\mathrm{GGAA}) \quad f(\mathrm{AAGG})$$



Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 1: pay a one-time cost to sketch the model globally

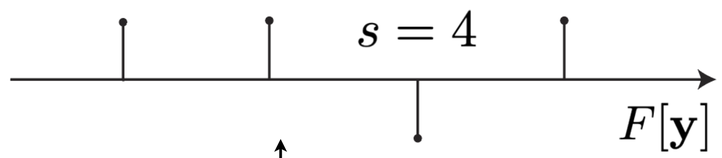


Subsampling

$$f(\mathrm{AAAA}) \quad f(\mathrm{AAAA})$$
$$f(\mathrm{ACAA}) \quad f(\mathrm{AAAC})$$
$$\ldots \qquad\qquad \ldots$$
$$f(\mathrm{GGAA}) \quad f(\mathrm{AAGG})$$

Aliasing

$$\mathbf{y}_1 \quad \mathbf{y}_2 \quad \mathbf{y}_3 \quad \mathbf{y}_4$$
$$\begin{bmatrix}\mathrm{C}\\ \mathrm{A}\\ \mathrm{A}\\ \mathrm{A}\end{bmatrix}, \begin{bmatrix}\mathrm{T}\\ \mathrm{A}\\ \mathrm{A}\\ \mathrm{G}\end{bmatrix}, \begin{bmatrix}\mathrm{T}\\ \mathrm{A}\\ \mathrm{C}\\ \mathrm{A}\end{bmatrix}, \begin{bmatrix}\mathrm{C}\\ \mathrm{G}\\ \mathrm{A}\\ \mathrm{A}\end{bmatrix}$$



Peeling

$$U_1[\mathrm{AA}] = F[\mathbf{y}_1] + F[\mathbf{y}_4]$$
$$U_1[\mathrm{AT}] = F[\mathbf{y}_2]$$
$$U_1[\mathrm{CA}] = F[\mathbf{y}_3]$$

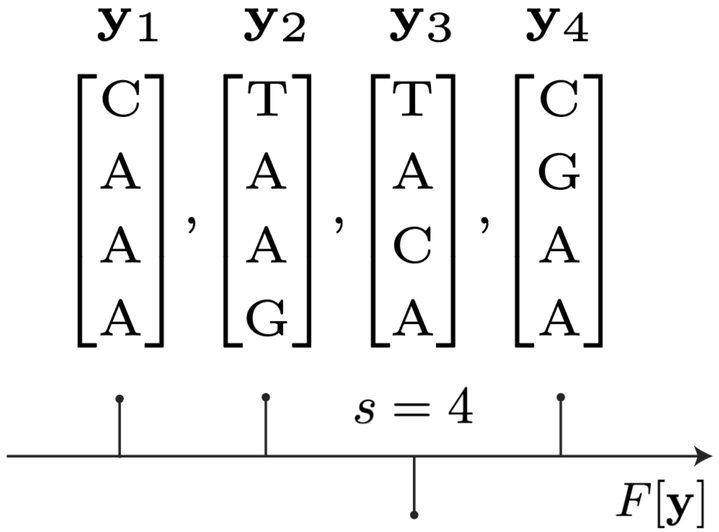
Singletons

$$U_2[\mathrm{GA}] = F[\mathbf{y}_2] + F[\mathbf{y}_3]$$
$$U_2[\mathrm{CA}] = F[\mathbf{y}_1]$$
$$U_2[\mathrm{CG}] = F[\mathbf{y}_4]$$

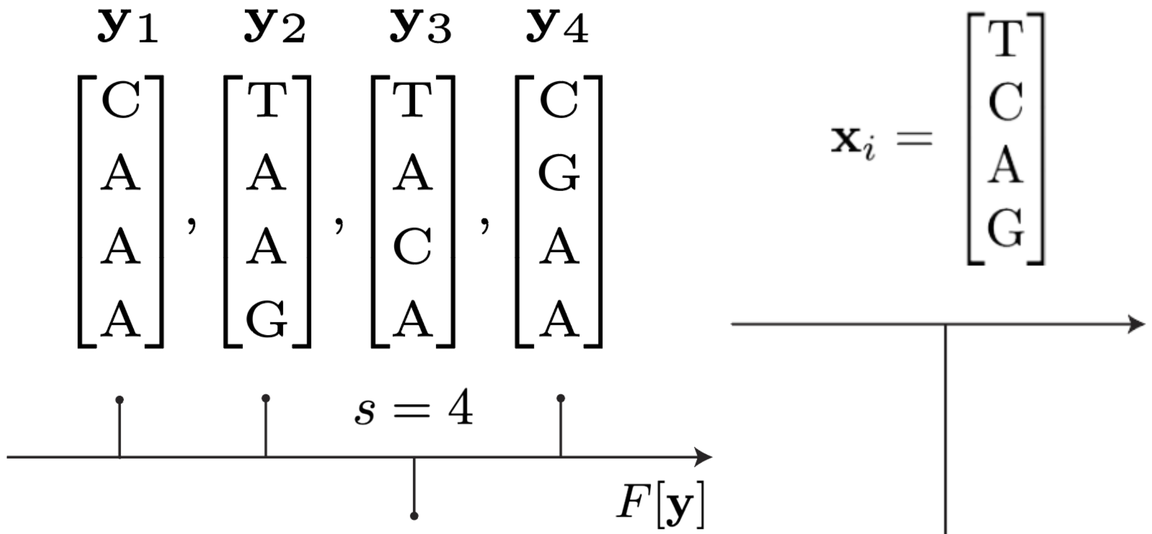
Multi-ton

Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 2: map sketch to local Möbius transform



step 2: map sketch to local Möbius transform

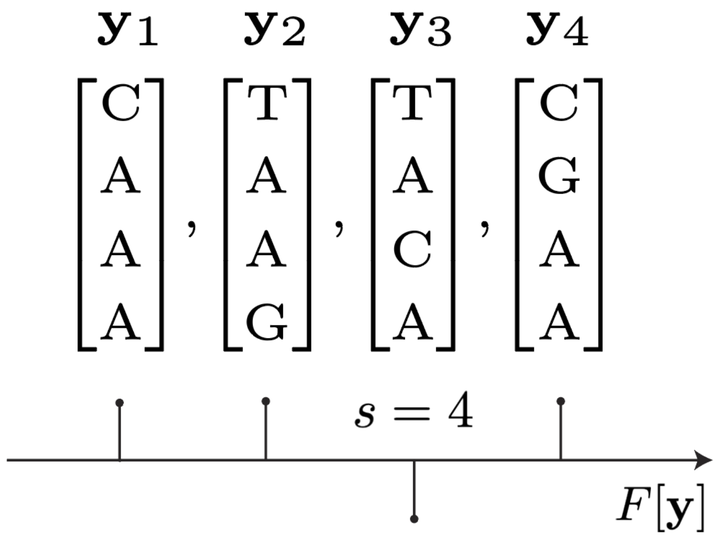


Future Queries

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 2: map sketch to local Möbius transform

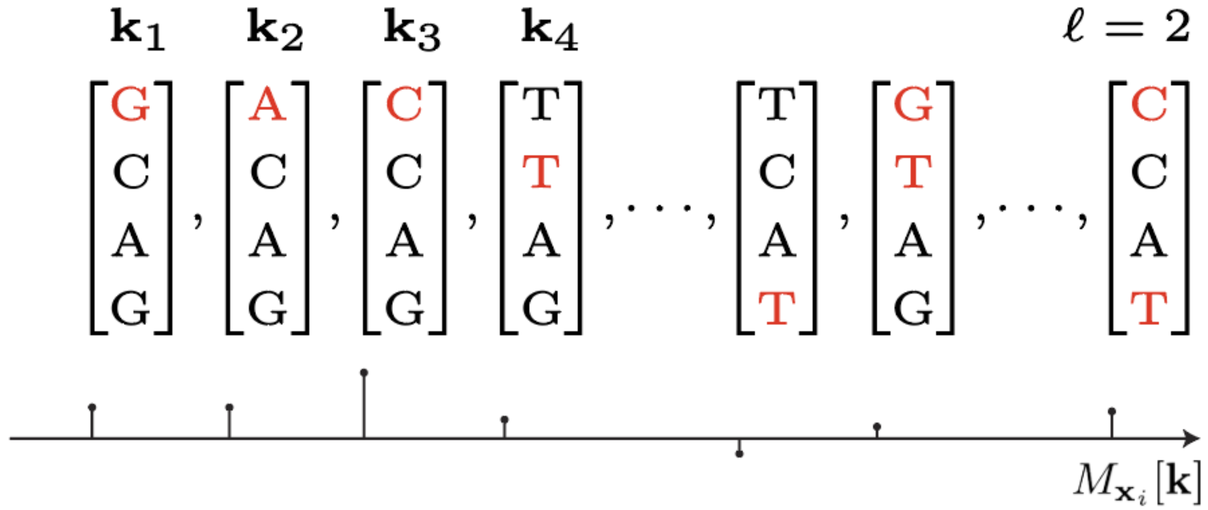


$$\mathbf{x}_i = \begin{bmatrix} \mathrm{T} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix}$$

Future Queries

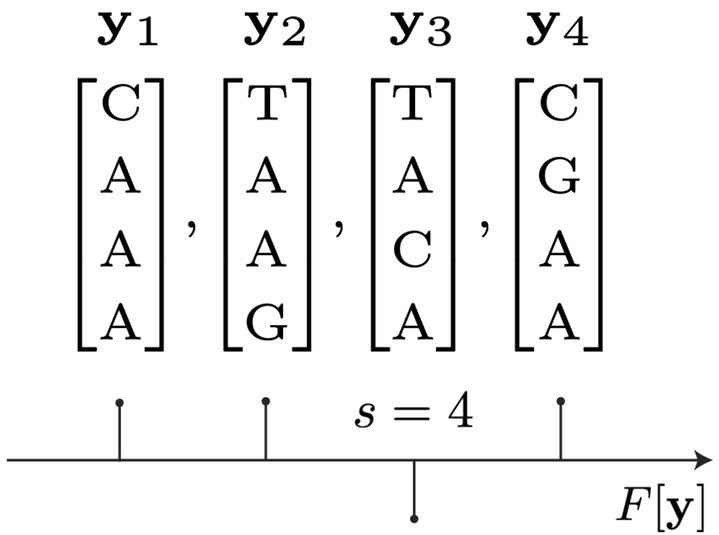
$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

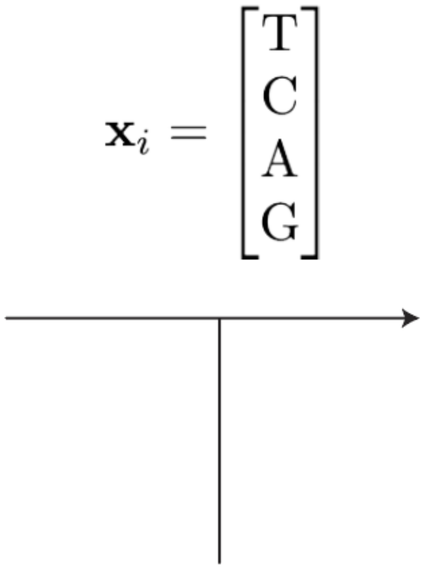
Fourier to Mobius $F \to M_{\mathbf{x}_i}$



Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 3: map Möbius transform to SHAP values and interactions

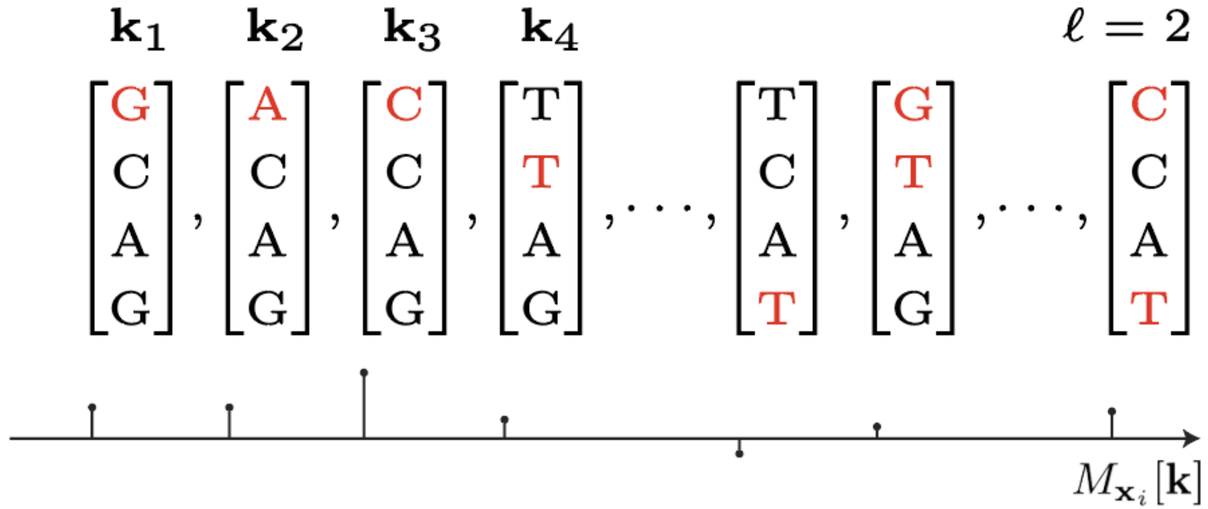




Future Queries

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

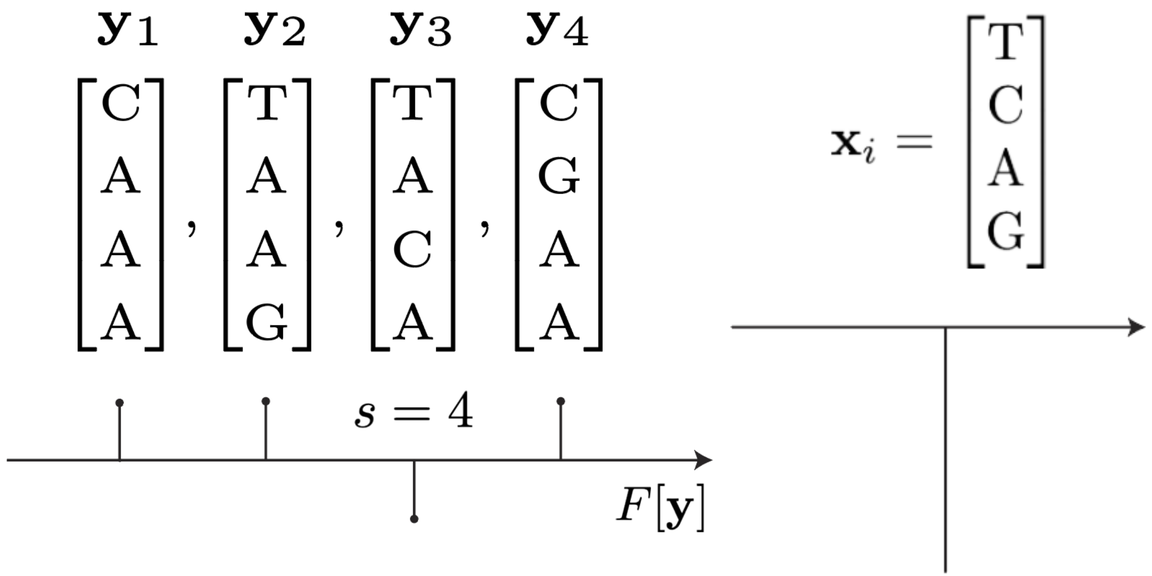
Fourier to Mobius $F \to M_{\mathbf{x}_i}$



Möbius to SHAP values $M_{\mathbf{x}_i} \to I^{SV}_{\mathbf{x}_i}$

$I^{SV}_{\mathbf{x}_i}(\mathrm{T@1})$

step 3: map Möbius transform to SHAP values and interactions

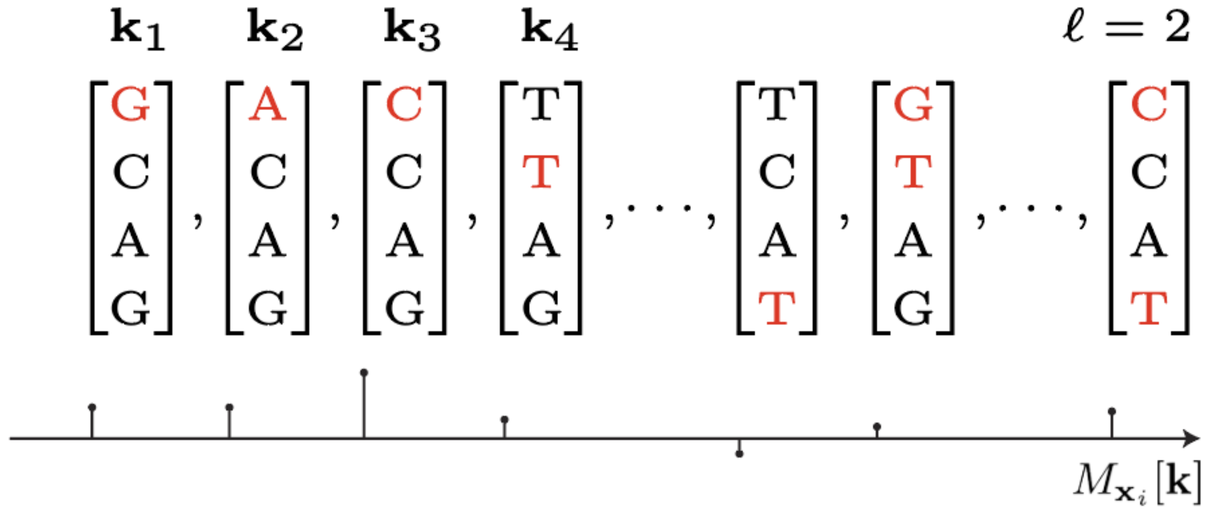


$$\mathbf{x}_i = \begin{bmatrix} \mathrm{T} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix}$$

Future Queries

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

Fourier to Mobius $F \to M_{\mathbf{x}_i}$

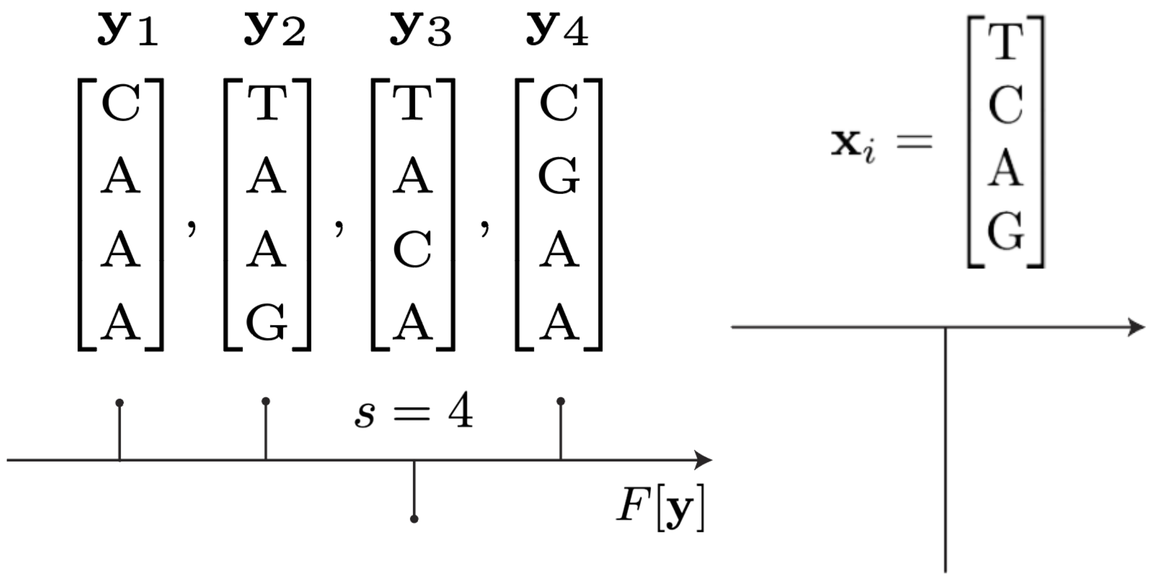


Möbius to SHAP values $M_{\mathbf{x}_i} \to I^{SV}_{\mathbf{x}_i}$

$$I^{SV}_{\mathbf{x}_i}(\mathrm{T@1}) = -\frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{G} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{A} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \cdots - \frac{1}{16} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{C} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{T} \end{bmatrix}$$

Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 3: map Möbius transform to SHAP values and interactions

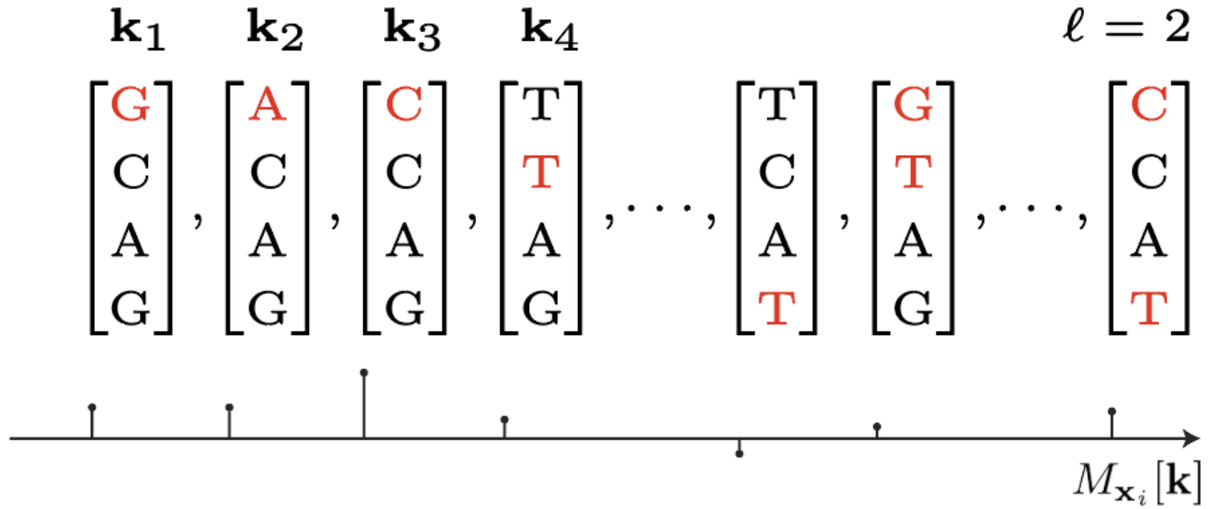


$$\mathbf{x}_i = \begin{bmatrix} \mathrm{T} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix}$$

Future Queries

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

Fourier to Mobius $F \to M_{\mathbf{x}_i}$



Möbius to SHAP values $M_{\mathbf{x}_i} \to I^{SV}_{\mathbf{x}_i}$

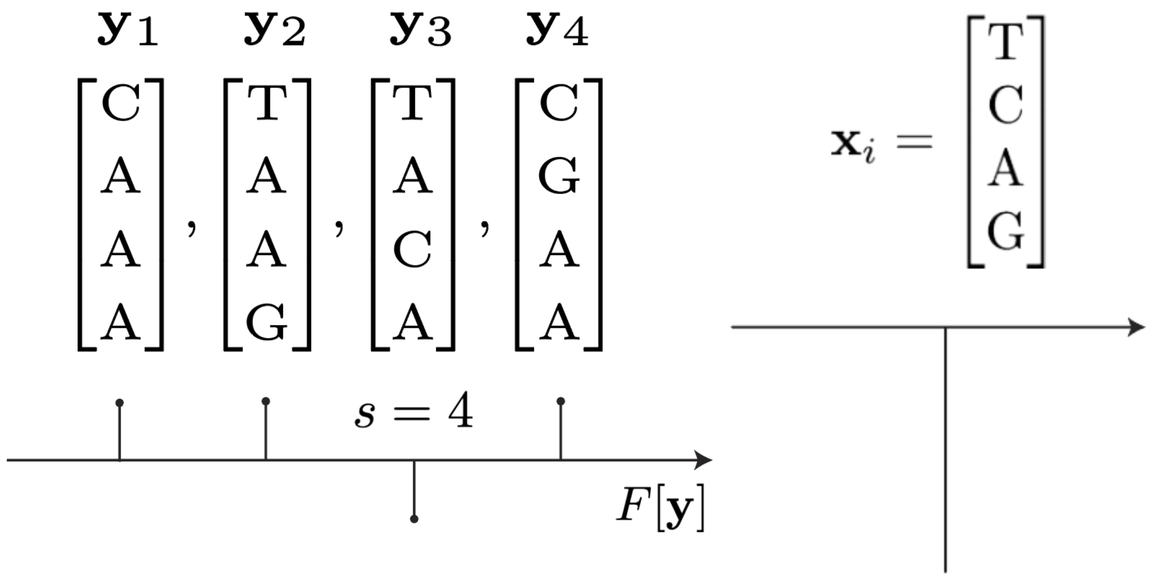
$$I^{SV}_{\mathbf{x}_i}(\mathrm{T@1}) = -\frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{G} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{A} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \cdots - \frac{1}{16} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{C} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{T} \end{bmatrix}$$

Möbius to Shapley interactions $M_{\mathbf{x}_i} \to I^{FSI}_{\mathbf{x}_i}$

$$I^{FSI}_{\mathbf{x}_i}(\mathrm{TC@}\{1,2\})$$

Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

step 3: map Möbius transform to SHAP values and interactions

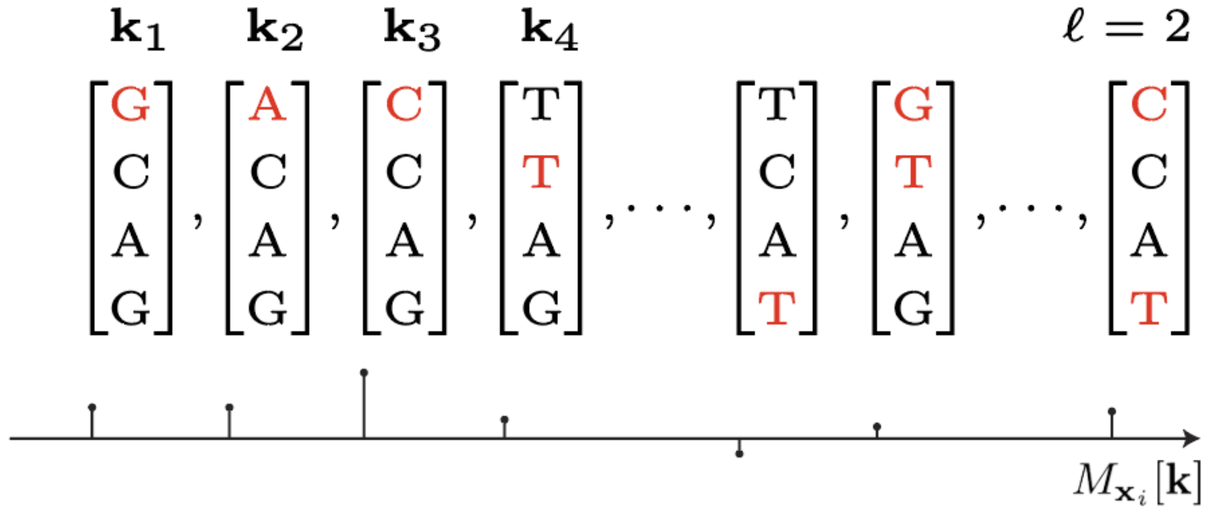


$$\mathbf{x}_i = \begin{bmatrix} \mathrm{T} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix}$$

Future Queries

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix}$$

Fourier to Mobius $F \to M_{\mathbf{x}_i}$



Möbius to SHAP values $M_{\mathbf{x}_i} \to I^{SV}_{\mathbf{x}_i}$

$$I^{SV}_{\mathbf{x}_i}(\mathrm{T@1}) = -\frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{G} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \frac{1}{4} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{A} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} - \cdots - \frac{1}{16} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{C} \\ \mathrm{C} \\ \mathrm{A} \\ \mathrm{T} \end{bmatrix}$$

Möbius to Shapley interactions $M_{\mathbf{x}_i} \to I^{FSI}_{\mathbf{x}_i}$

$$I^{FSI}_{\mathbf{x}_i}(\mathrm{TC@}\{1,2\}) = \frac{1}{16} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{G} \\ \mathrm{T} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix} + \cdots + \frac{1}{16} M_{\mathbf{x}_i} \begin{bmatrix} \mathrm{C} \\ \mathrm{G} \\ \mathrm{A} \\ \mathrm{G} \end{bmatrix}$$

Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

SHAP zero amortizes Shapley explanations

Algorithm. SHAP zero in three steps:

- **Step 1**: Pay a **one-time cost** to sketch the model globally via the Fourier transform with **sample complexity** $\mathcal{O}(sn^2)$ and **computational complexity** $\mathcal{O}(sn^3)$

For each future sequence:

- **Step 2**: Map global sketch to the local Möbius transform to isolate interacting features with complexity $\mathcal{O}(s^2(2q)^\ell)$
- **Step 3**: Map Möbius transform to Shapley values and interactions with complexity $\mathcal{O}(s^2(2q)^\ell)$

SHAP zero amortizes Shapley explanations

Algorithm. SHAP zero in three steps:

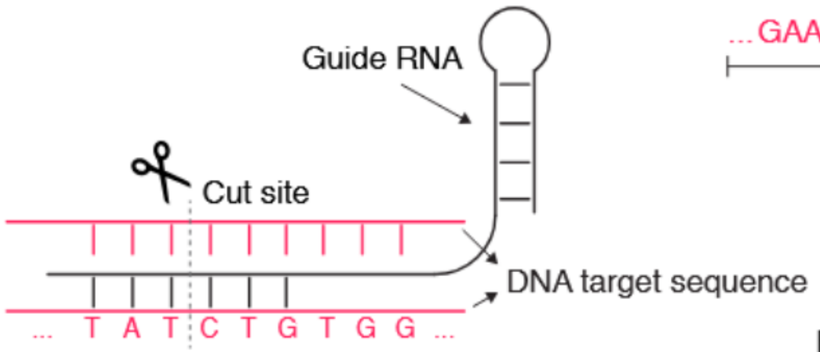
- **Step 1**: Pay a **one-time cost** to sketch the model globally via the Fourier transform with **sample complexity** $\mathcal{O}(sn^2)$ and **computational complexity** $\mathcal{O}(sn^3)$

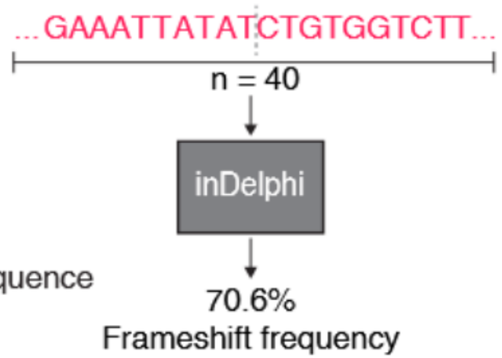
For each future sequence:

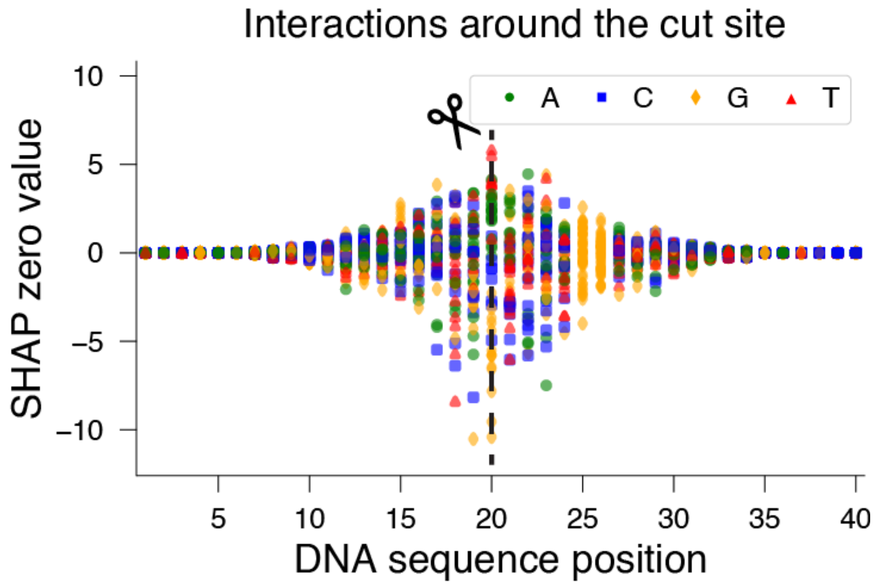
- **Step 2**: Map global sketch to the local Möbius transform to isolate interacting features with complexity $\mathcal{O}(s^2(2q)^\ell)$
- **Step 3**: Map Möbius transform to Shapley values and interactions with complexity $\mathcal{O}(s^2(2q)^\ell)$

After Step 1, SHAP zero scales **essentially** free!

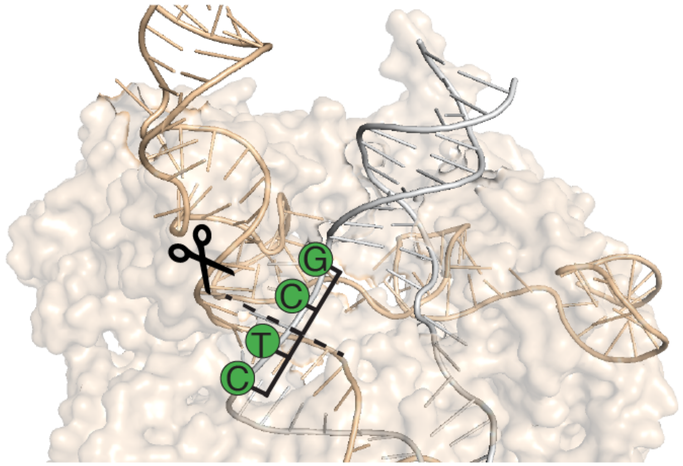
SHAP zero uncovers DNA repair interactions at scale

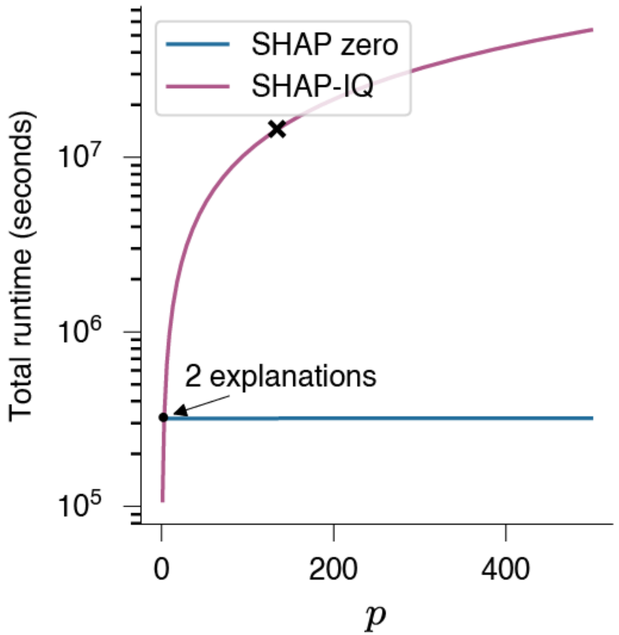






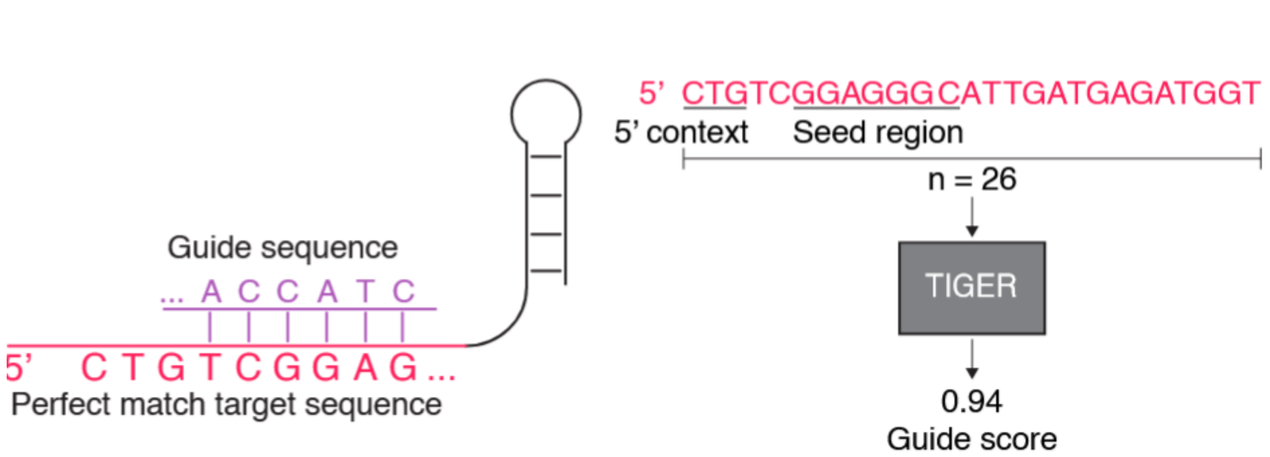
Interactions in DNA repair

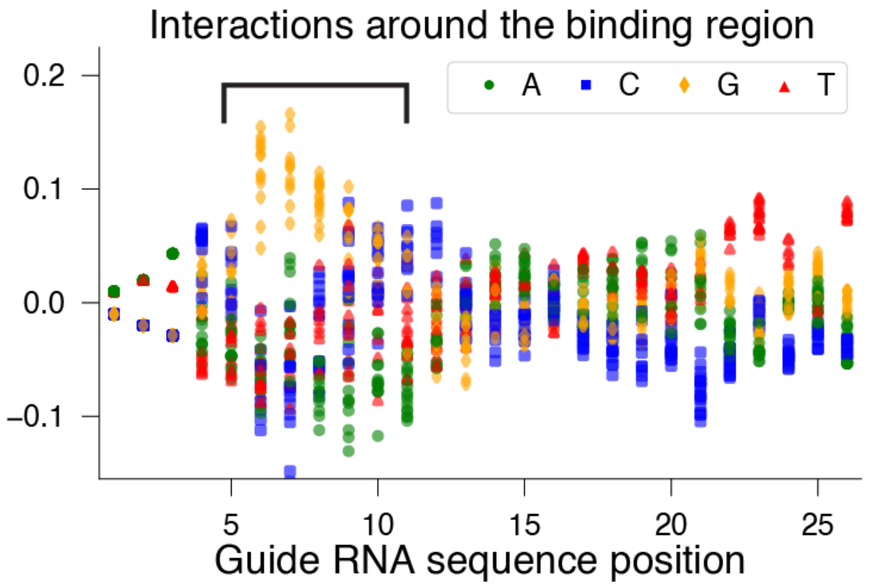




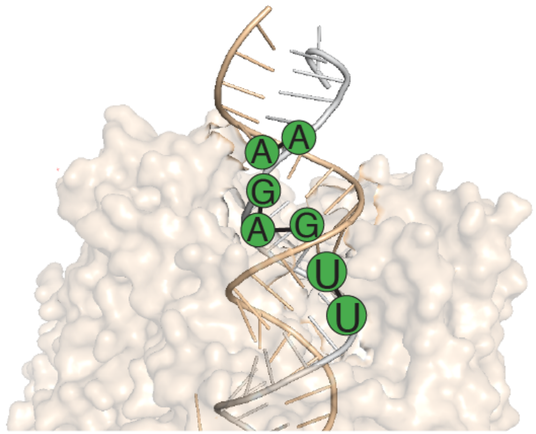
Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

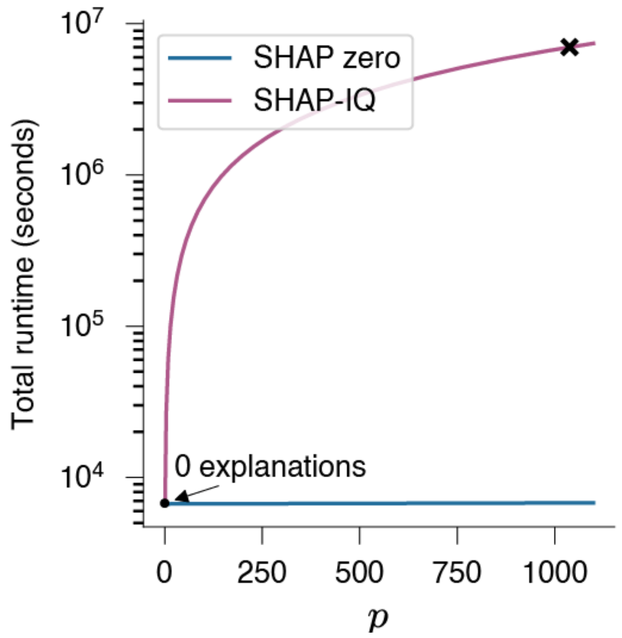
SHAP zero uncovers guide RNA binding interactions at scale





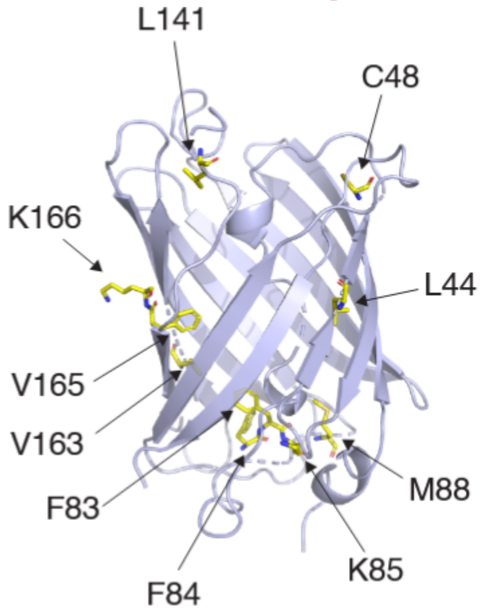
Interactions in guide RNA binding

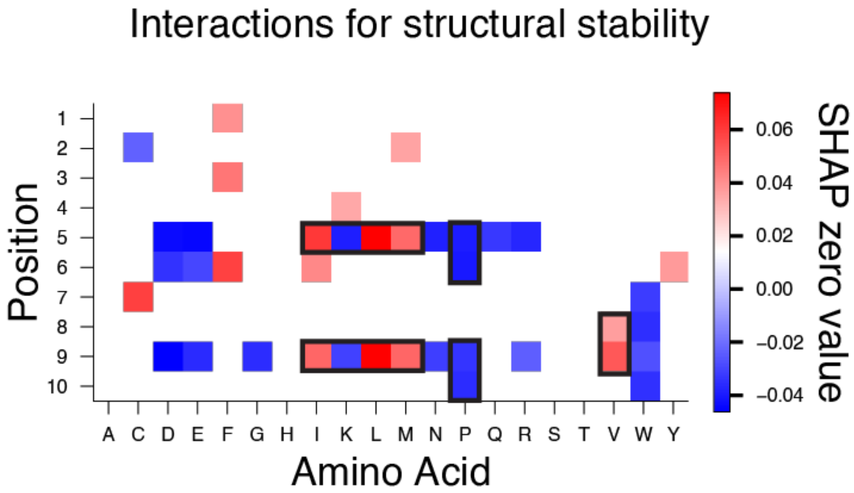


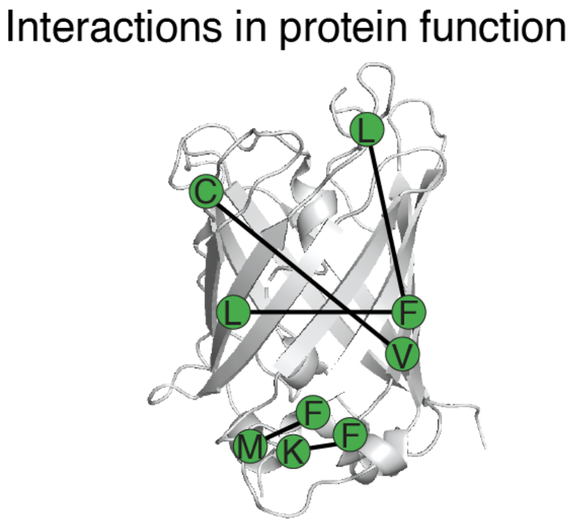


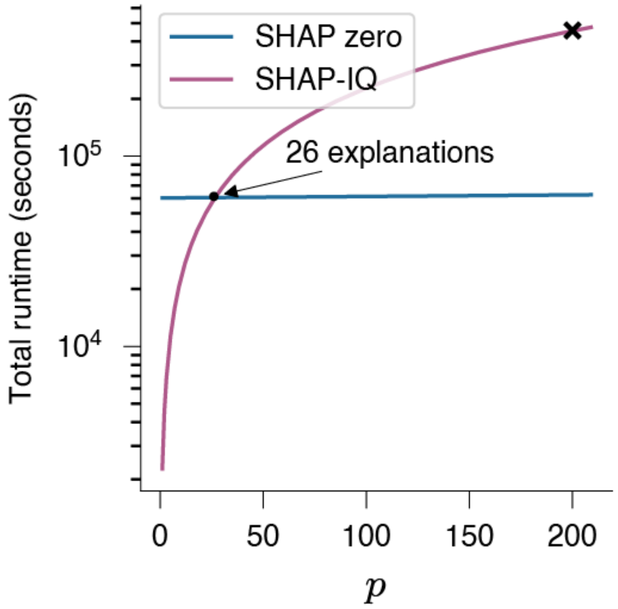
Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

SHAP zero uncovers protein function interactions at scale









Tsui, Musharaf, Erginbas, Kang, Aghazadeh, *NeurIPS* (2025)

promoting open-source algorithms in biology

SHAP zero: Explaining Biological Sequence Models

pypi package 0.0.4 | license MIT | status beta | python 3.10 | 3.11 | 3.12 | code style black | last commit september

SHAP zero is a Python package that enables the amortized computation of Shapley values and interactions. It does this by paying a one-time cost to sketch the model's Fourier transform. After this one-time cost, SHAP zero enables **near-zero marginal cost** for future query sequences by mapping the Fourier transform to Shapley values and interactions.

pip install shapzero

https://github.com/amirgroup-codes/shap-zero

thank you!

Paper

Code