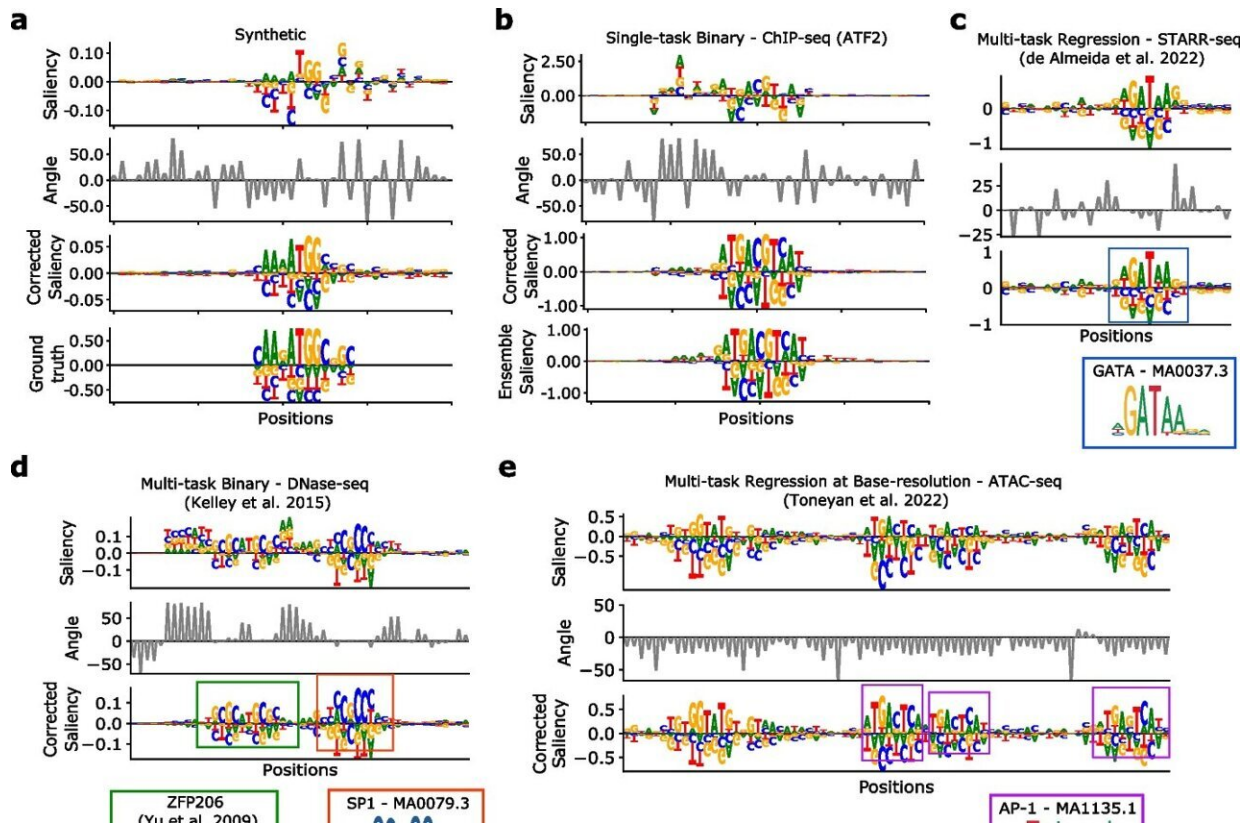


# The digital dark matter clouding AI in genome analysis

June 5 2023, by Luis Sandoval



Visualizing the gradient correction. Sequence logo of the uncorrected saliency map (top row), gradient angles at each position (second row), and corrected saliency map (third row) for a patch from representative test sequences. a, b CNN-deep-relu trained to make binary predictions on a synthetic data and b ChIP-seq data for ATF2 protein in GM12878. The sequence logo of ground truth is shown for CNN-deep-exp for a synthetic data. b An ensemble average saliency map is shown in lieu of ground truth (bottom row). c–e A similar plot is made for a c DeepSTARR model trained to predict enhancer activity via STARR-

seq data, d Basset model trained to make binary predictions of chromatin accessibility sites via DNase-seq data, and e CNN model trained to predict base-resolution read-coverage values from ATAC-seq data in PC-3 cell line. c–e A colored box and a corresponding sequence logo of a known motif from JASPAR (with a corresponding ID) are shown for comparison. Credit: *Genome Biology* (2023). DOI: 10.1186/s13059-023-02956-3

Artificial intelligence has entered our daily lives. First, it was ChatGPT. Now, it's AI-generated pizza and beer commercials. While we can't trust AI to be perfect, it turns out that sometimes we can't trust ourselves with AI either.

Cold Spring Harbor Laboratory (CSHL) Assistant Professor Peter Koo has found that scientists using popular computational tools to interpret AI predictions are picking up too much "noise," or extra information, when analyzing DNA. And he's found a way to fix this. Now, with just a couple new lines of code, scientists can get more reliable explanations out of powerful AIs known as [deep neural networks](#). That means they can continue chasing down genuine DNA features. Those features might just signal the next breakthrough in health and medicine. But scientists won't see the signals if they're drowned out by too much noise.

So, what causes the meddlesome noise? It's a mysterious and invisible source like digital "[dark matter](#)." Physicists and astronomers believe most of the universe is filled with dark matter, a material that exerts gravitational effects but that no one has yet seen. Similarly, Koo and his team discovered the data that AI is being trained on lacks critical information, leading to significant blind spots. Even worse, those blind spots get factored in when interpreting [AI predictions of DNA function](#). The study is published in the journal *Genome Biology*.

Koo says, "The deep neural network is incorporating this random behavior because it learns a function everywhere. But DNA is only in a small subspace of that. And it introduces a lot of noise. And so we show that this problem actually does introduce a lot of noise across a wide variety of prominent AI models."

The digital dark matter is a result of scientists borrowing computational techniques from computer vision AI. DNA data, unlike images, is confined to a combination of four nucleotide letters: A, C, G, T. But [image data](#) in the form of pixels can be long and continuous. In other words, we're feeding AI an input it doesn't know how to handle properly.

By applying Koo's computational correction, scientists can interpret AI's DNA analyses more accurately.

Koo says, "We end up seeing sites that become much more crisp and clean, and there is less spurious noise in other regions. One-off nucleotides that are deemed to be very important all of a sudden disappear."

Koo believes noise disturbance affects more than AI-powered DNA analyzers. He thinks it's a widespread affliction among computational processes involving similar types of data. Remember, dark matter is everywhere. Thankfully, Koo's new tool can help bring scientists out of the darkness and into the light.

**More information:** Antonio Majdandzic et al, Correcting gradient-based interpretations of deep neural networks for genomics, *Genome Biology* (2023). [DOI: 10.1186/s13059-023-02956-3](https://doi.org/10.1186/s13059-023-02956-3)

Provided by Cold Spring Harbor Laboratory

Citation: The digital dark matter clouding AI in genome analysis (2023, June 5) retrieved 25 April 2024 from <https://phys.org/news/2023-06-digital-dark-clouding-ai-genome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.