Künstliche Intelligenz hilft beim Knacken des genetischen Codes

Die künstliche Intelligenz hat längst auch Einzug in den Forschungsalltag gehalten, Bild: Pixabay, CC0

Forscher:innen vom Institut für molekulare Pathologie (IMP) in Wien ist es mit Hilfe von KI gelungen, den Code von Enhancern zu knacken. Diese DNA-Elemente sind für die Regulation der Gene wichtig und Teil des „zweiten genetischen Codes".

Jeder Organismus ist ein einzigartiges Gebilde aus verschiedensten Zelltypen mit einzigartiger Identität und Funktion. Jede Zelle besitzt zwar dieselbe DNA und somit den gleichen genetischen Code, aber um unterschiedliche Zellidentitäten zu bilden, werden je nach Zelltyp verschiedene Gene abgelesen und andere nicht. Nur durch eine präzise Steuerung der Genaktivität kann eine Zelle im Laufe ihrer Entwicklung einen bestimmten Lebensweg einschlagen und beispielsweise zu einer Nerven-, Muskel-, Haut- oder Blutzelle werden.

Ein- und Ausschalter von Genen

Gene sind also nicht immer aktiv – ein Grund dafür ist die Aktivität von speziellen Kontrollelementen der DNA, die das Ablesen der Gene entweder fördern oder unterdrücken können. Dabei wirken so genannte Enhancer positiv, Silencer negativ. Ihre genregulierende Funktion üben sie aus, indem sie das Binden von Proteinen, welche die DNA ablesen, entweder fördern oder unterdrücken. Die regulatorischen Elemente der DNA stellen einen „zweiten genetischen Code" dar, der die Aktivität von Genen regelt.

Enhancer schwer zu identifizieren

Enhancer stellen regulatorische DNA-Sequenzen dar, die Gene einschalten oder deren Ableserate erhöhen können – sie sind quasi „Ableseverstärker". Obwohl diese DNA-Elemente seit den 1980er-Jahren bekannt sind, ist eine genaue Voraussage ihrer Aktivität in einem spezifischen Gewebe zu einem gewissen Zeitpunkt der Entwicklung bis heute schwierig. Um das Ablesen eines Gens kontrollieren zu können, ist die Enhancer-Sequenz selbst wichtig, aber auch umliegende DNA-Abschnitte können dabei eine Rolle spielen. Die Tatsache, dass Enhancer auch sehr weit weg von dem Gen liegen können, welches sie regulieren, stellt eine zusätzliche Herausforderung bei deren Untersuchung dar. Konkrete Methoden, um Enhancer-Elemente in der DNA experimentell zu identifizieren, gibt es erst seit dem letzten Jahrzehnt.

KI-unterstützter Ansatz zur Erkennung von Enhancern

Wissenschaftler:innen rund um Alexander Stark vom IMP machten sich nun die Fortschritte bei der Erfassung und Analyse des Genoms (Genomik) sowie die künstliche Intelligenz zunutze, um die Aktivität von Enhancern aus ihrer DNA-Sequenz zu entschlüsseln. Die Forschenden adressierten mit diesem Ansatz drei lange gehegte Ziele, die bis dato in weiter Ferne schienen: Die Aktivität von Enhancern aus ihrer DNA-Sequenz vorauszusagen, die Folgen von Mutationen in Enhancern vorauszusagen sowie selbst gewebespezifische Enhancer zu designen. Oder, anders gesagt: Stark und sein Team wollten den „zweiten genetischen Code" für die Genregulation lesen, verstehen und dann auch selbst schreiben.

Gemeinsam mit Kolleg:innen vom EMBL in Heidelberg entwickelten die Wiener Forscher:innen dafür ein leistungsfähiges Deep Learning Modell, bei dem maschinelles Lernen durch künstliche neuronale Netze und große Datenmengen (Big Data) ermöglicht wird. Stark und seine Kolleg:innen trainierten ihr KI-basiertes Deep learning System zunächst mit genomweiten DNA-Sequenzen und entsprechenden Daten der DNA-Zugänglichkeit, allesamt aus Studien in der Fruchtfliege **Drosophila melanogaster.** Die Zugänglichkeit von DNA stellt ein relativ leicht messbares Vorhersagekriterium für Enhancer dar. Für die KI war es dann möglich, ein Muster für Enhancer in den eingespielten Daten zu erkennen.

Basierend darauf wurde dann ein weiteres KI-Modell mit Ergebnissen aus Studien zur Enhancer-Aktivität trainiert, um schlussendlich das erste Modell zu verfeinern. Dem daraus resultierenden KI-Tool gelang es schließlich, direkt anhand der DNA-Sequenz gewebespezifische Enhancer im zentralen Nervensystem, einem Teilbereich des Gehirns, der Epidermis, dem Darm und in Muskeln von Fruchtfliegenembryonen vorauszusagen.

Vom KI-gestützen Computer-Modell wieder zurück ins Labor

Danach ging es für die Wissenschaftler:innen vom Computer wieder zurück ins Labor. Die vom KI-Tool rechnerisch entwickelten Enhancer-Sequenzen, die in verschiedenem Gewebe aktiv sein sollten, wurden mithilfe etablierter molekularbiologischer Methoden hergestellt. Das Team um Stark testete daraufhin die Aktivität von insgesamt 40 solcher synthetischer Enhancer in lebenden Embryos von Fruchtfliegen. Und der innovative Ansatz der Forscher:innen entpuppte sich als äußerst erfolgreich, denn die synthetischen Enhancer aktivierten tatsächlich Gene in den Zielgeweben.

Bernardo de Almeida, Erstautor der im Fachjournal "Nature" veröffentlichten Studie, dazu: "Die Möglichkeit, künstliche Enhancer mit spezifischen Eigenschaften zu bauen, eröffnet ungeahnte Möglichkeiten für die gezielte Steuerung von Genexpression".

Die Forschenden sehen zukünftige Anwendungen in der synthetischen Biologie oder Gentherapie, denn dort ist die präzise Gestaltung und Manipulation der Genaktivität eine wichtige Voraussetzung.

CS, 26.01.2024

Quellenangaben

Originalpublikation:

BP de Almeida, Schaub C., Pagani M., Secchia S., Furlong EEM and Stark A.:. Targeted design of synthetic enhancers for selected tissues in the **Drosophila** embryo. Nature (2023). DOI: 10.1038/s41586-023-06905-9

Referenz:

Pressemeldung des IMP vom 12.12.2023