

Gemeinsame Pressemitteilung des IPK Leibniz-Institutes und des
Forschungszentrums Jülich

KI entschlüsselt die Sprache der Gene: Neuer Blick in die „Steuerzentrale“ der Pflanzen

Gatersleben, 16.06.2026 Ein internationales Forschungsteam unter Führung des IPK Leibniz-Instituts und des Forschungszentrums Jülich hat mit Hilfe künstlicher Intelligenz (KI) ein Modell entwickelt, das vorhersagt, an welchen Stellen regulatorische Proteine an die pflanzliche DNA andocken, um anschließend Gene ein- und auszuschalten. Das Modell, das mit umfangreichen Genomdaten der Modellpflanze *Arabidopsis thaliana* trainiert wurde, lässt sich auch auf Nutzpflanzen übertragen. Damit lässt sich insgesamt besser verstehen, wie die genetische Variation die Leistungsfähigkeit von Nutzpflanzen beeinflusst. Die Studie wurde im Fachmagazin „Nature Communications“ veröffentlicht.

Wenn vom Erbgut die Rede ist, denken viele Menschen an Gene. Doch Gene allein erklären nicht, warum Pflanzen unterschiedlich wachsen oder auf Umweltreize reagieren. Tatsächlich enthält die DNA auch viele Abschnitte, die wie Schalter oder Regler wirken. Ein besonders wichtiges regulatorisches Element sind Transkriptionsfaktoren. Diese binden an die DNA und bestimmen, wann ein Gen aktiv wird und wie stark es exprimiert wird. Es geht also darum, ob und wie oft ein Gen abgelesen wird, sodass seine Information in RNA oder Protein umgesetzt werden kann und physiologisch aktiv wird.

Man kann sich das wie ein Haus vorstellen: Die Gene sind die Räume. Die regulatorischen Elemente sind Lichtschalter, Thermostate und Sicherungskästen. Wer verstehen will, wie das Haus funktioniert, muss nicht nur die Räume kennen, sondern auch die Technik in diesen. Das Forschungsteam wollte die entsprechenden Mechanismen anhand der umfangreichen Daten der Modellpflanze *Arabidopsis thaliana* besser verstehen.

Dafür trainierten die Forscherinnen und Forscher ein Deep-Learning-Modell anhand von Hunderten experimenteller DNA-Bindungsdatensätzen und brachten ihm bei, die Bindungsmuster von insgesamt 46 Transkriptionsfaktor-Familien gleichzeitig zu erkennen. Dieses „Multi-Label“-Konzept unterscheidet sich von früheren Ansätzen, bei denen meist für jeden einzelnen Faktor ein eigenes Modell erstellt wurde, was sich auf das gesamte Genom nur schwer übertragen ließ. Anschließend prüfte das Team, ob das Modell Bindungsstellen korrekt lokalisieren und neue regulatorische Zusammenhänge aufdecken konnte.

„Unsere Ergebnisse zeigen, dass Transkriptionsfaktoren nicht nur einzelne DNA-Motive lesen. Entscheidend sind der Kontext und die Art und Weise, wie die Signale zueinander in Beziehung stehen“, erläutert Fritz Forban Peleke, Erstautor der Studie. Das ist wie bei einem Text. Erst die Reihenfolge und der Zusammenhang mehrerer Wörter ergeben einen verständlichen Text. Ähnlich verhält es sich mit der DNA. Nicht einzelne Bausteine bestimmen die Funktion, sondern ihr Zusammenspiel.

Die KI analysierte jedoch auch die Bindungsmuster der Transkriptionsfaktoren und ordnete die Gene der Pflanze verschiedenen Gruppen zu. Dabei zeigte sich, dass sich Tausende Gene auf nur 14 große regulatorische Cluster zurückführen lassen. Dabei korrelieren mehrere

Pressemitteilung

Wissenschaftlicher Kontakt

Dr. Jędrzej Jakub Szymanski
Tel.: +49 39482 5753
szymanski@ipk-gatersleben.de

Medienkontakt

Christian Schafmeister
Tel.: +49 39482 5461
schafmeister@ipk-gatersleben.de

Anna Tipping
Tel.: +49 2461 61 5281
a.tipping@fz-juelich.de

davon miteinander und weisen gemeinsame biologische Funktionen auf. „Obwohl Pflanzen Tausende Gene besitzen, entstehen viele ihrer Funktionen ganz offensichtlich aus einer überraschend kleinen Zahl regulatorischer Muster“, sagt Fritz Forban Peleke.

Das Forschungsteam untersuchte zudem mehr als 7.000 DNA-Varianten, die zuvor in genomweiten Studien mit Merkmalen wie Blütezeit, Krankheitsresistenz und Wachstum der Keimlinge in Verbindung gebracht worden waren. Bei etwa jeder fünften dieser Varianten wurde eine Veränderung der Bindung von Transkriptionsfaktoren vorhergesagt. „Wir können nun abschätzen, wie eine einzelne Veränderung in einem regulatorischen DNA-Abschnitt die Genaktivität und damit ein wichtiges Pflanzenmerkmal beeinflusst“, erklärt Dr. Jędrzej Szymański, Leiter der Forschungsgruppe „Netzwerkanalyse und Modellierung“ am IPK und der Forschungsgruppe „Omics-Daten“ am Forschungszentrum Jülich. „Dies bietet die Möglichkeit, von einem statistischen Zusammenhang zu einem plausiblen molekularen Mechanismus zu gelangen.“

Besonders interessant war ein Beispiel zum Blühzeitpunkt. Die KI prognostizierte, dass eine einzelne Veränderung in einem regulatorischen DNA-Abschnitt die Bindung mehrerer Transkriptionsfaktoren beeinflusst. Genau solche Veränderungen können dazu beitragen, dass Pflanzen früher oder später blühen. Die Vorhersagen der KI wurden anschließend experimentell überprüft und bestätigt.

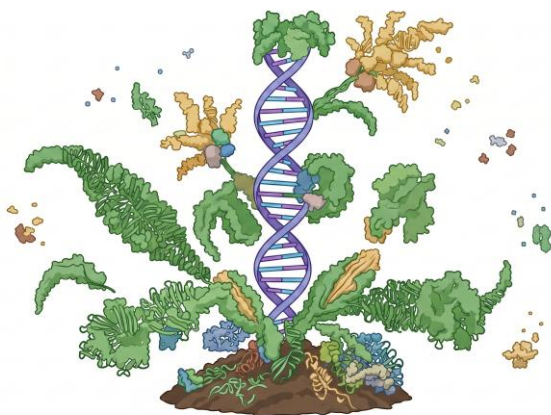
Obwohl das Modell mit Daten aus der Modelnpflanze *Arabidopsis* trainiert wurde, konnte es auch erfolgreich auf Mais angewendet werden. Dort half die KI, Transkriptionsfaktoren zu identifizieren, die auf Hitzestress reagieren. Besonders bedeutsam waren bereits bekannte Regulatoren von Hitzereaktionen. Dies verdeutlicht das Potenzial des Ansatzes für die Pflanzenforschung, auch bei Arten, für die bisher weniger molekulare Bindungsdaten vorliegen.

Originalpublikation:

Peleke *et al.* (2026): Genome-wide modelling of plant transcription factor binding captures regulatory variants associated with phenotypic traits. Nature Communications.

DOI: [10.1038/s41467-026-73634-8](https://doi.org/10.1038/s41467-026-73634-8)

Grafik:



Neue Deep-Learning-Modelle sagen voraus, wie komplexe DNA-Protein-Wechselwirkungen Pflanzengene aktivieren und deaktivieren, und zeigen damit auf, wie genetische Variationen Wildarten und Kulturpflanzen helfen, sich an Umweltbelastungen anzupassen.