

Joint press release from the IPK Leibniz Institute and
Forschungszentrum Jülich

AI decodes the language of genes: A new look inside the “control centre” of plants

Gatersleben, 16.06.2026 **An international research team led by the IPK Leibniz Institute and Forschungszentrum Jülich has developed an artificial intelligence (AI) model that predicts where regulatory proteins bind to plant DNA in order to switch genes on and off. Trained entirely on the rich genomic data available for the model plant *Arabidopsis thaliana*, the model can also be successfully transferred to crops. This opens up new opportunities to better understand how genetic variation influences crop performance. The study is published in “Nature Communications.”**

When people talk about the genome, many think of genes. But genes alone do not explain why plants grow differently or react to environmental stimuli. In fact, DNA also contains many sections that act like switches or regulators. Particularly important regulatory elements are so-called transcription factors. These bind to DNA and determine when a gene becomes active and how strong it works.

A useful picture is a house: the genes are the rooms, while the regulatory regions are the light switches, thermostats and fuse boxes. To understand how the house works, you need to know not just the rooms but the wiring behind the walls. The IPK team set out to map that wiring using massive data resources available for the “lab rat” of plant science - *Arabidopsis thaliana*.

To do so, the researchers trained a deep learning model on hundreds of experimental DNA-binding datasets, teaching it to recognise the binding patterns of 46 transcription factor families at once. The previous methods, which usually involved making a separate model for each factor and didn't work well across a genome, are being replaced by this "multi-label" design. The team then tested whether the model could correctly locate binding sites that had not been shown, and uncover new regulatory relationships.

“Our results indicate that transcription factors don't simply read isolated DNA motifs. What matters is the surrounding sequence and the way these signals are arranged together,” says Fritz Forbang Peleke, first author of the study. The analogy is language: individual words carry little meaning until their order and context form a sentence. In DNA, too, function emerges from how regulatory elements combine - a kind of regulatory grammar - rather than from single building blocks alone.

Using these predicted binding patterns, the model sorted *Arabidopsis* genes into groups based on their likely regulation. Strikingly, thousands of genes fell into just 14 broad regulatory clusters, several of which lined up with shared biological functions and coordinated gene activity. “Plants carry thousands of genes, yet many of their functions appear to arise from a surprisingly small set of recurring regulatory patterns,” Peleke says.

Press Release

Scientific Contact

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

Media Contact

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

Anna Tipping

Phone: +49 2461 61 5281
a.tipping@fz-juelich.de

The team also examined more than 7,000 DNA variants previously linked in genome-wide studies to traits such as flowering time, disease resistance, and seedling growth. About one in five of these variants was predicted to alter transcription factor binding. “We can now estimate how a single change in a regulatory stretch of DNA alters gene activity and, in turn, an important plant trait,” explains Dr. Jędrzej Szymański, head of the Network Analysis and Modelling research group at the IPK and of the Omics Data research group at the Forschungszentrum Jülich. “This gives researchers a way to move from a statistical association to a plausible molecular mechanism. „One example of flowering time proved especially telling. The model predicted that a single base change in a regulatory region would simultaneously affect the binding of several transcription factors - the kind of change that can nudge a plant to flower earlier or later. The prediction was then confirmed experimentally using a high-throughput reporter assay.

Although trained only on *Arabidopsis*, the model could be applied to the distantly related crop maize, where it helped annotate which transcription factors respond to heat stress. Known heat-stress regulators, including heat shock factors, stood out as particularly important, illustrating how the approach could support crop research in species where binding data remain scarce.

Original publication:

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)

Graphic:



New deep learning models predict how complex DNA-protein interactions turn plant genes on and off, revealing how genetic variations help wild species and crops adapt to environmental stress.