

Plant STARR-seq: von der DNA-Sequenz zur Funktion pflanzlicher Genregulation

Julia Macholl, Sebastian Triesch & Tobias Jores

Article - Version of Record

Suggested Citation:

Macholl, J., Triesch, S., & Jores, T. (2026). Plant STARR-seq: von der DNA-Sequenz zur Funktion pflanzlicher Genregulation. *Biospektrum*, 32(1), 57–60. <https://doi.org/10.1007/s12268-026-2644-8>

Wissen, wo das Wissen ist.



UNIVERSITÄTS-UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20260309-141306-8>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>

Hochdurchsatzscreening

Plant STARR-seq: von der DNA-Sequenz zur Funktion pflanzlicher Genregulation

JULIA MACHOLL, SEBASTIAN TRIESCH, TOBIAS JORES

INSTITUT FÜR SYNTHETISCHE BIOLOGIE UND EXZELLENZCLUSTER FÜR PFLANZENWISSENSCHAFTEN (CEPLAS), UNIVERSITÄT DÜSSELDORF

***Cis*-regulatory elements are promising targets for genome editing in crop improvement. However, our understanding of their activity is limited. Plant STARR-seq enables the quantitative, systematic, high-throughput measurement of regulatory activity. In combination with machine learning, this approach opens up new possibilities for the analysis, prediction, and design of regulatory DNA sequences.**

DOI: 10.1007/s12268-026-2644-8

© The Author(s) 2026

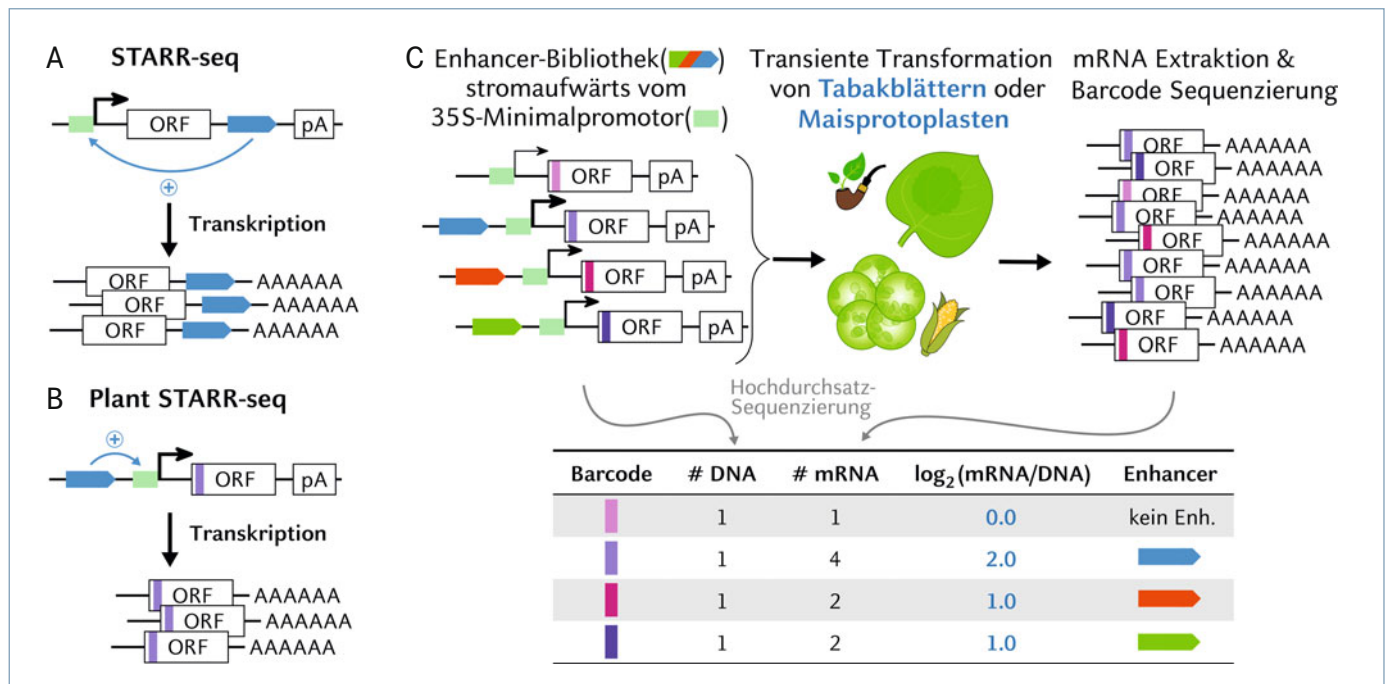
■ Damit Pflanzen sich entwickeln und flexibel an Umweltbedingungen anpassen können, ist die genaue Kontrolle ihrer Genexpression essenziell. Diese wird durch *cis*-regulatorische Elemente gesteuert – wie

Kernpromotoren, Enhancer, Silencer und Insulatoren [1]. Bereits kleine Unterschiede in der Genexpression können phänotypische Unterschiede verursachen oder Entwicklungsprozesse beeinflussen, weshalb die

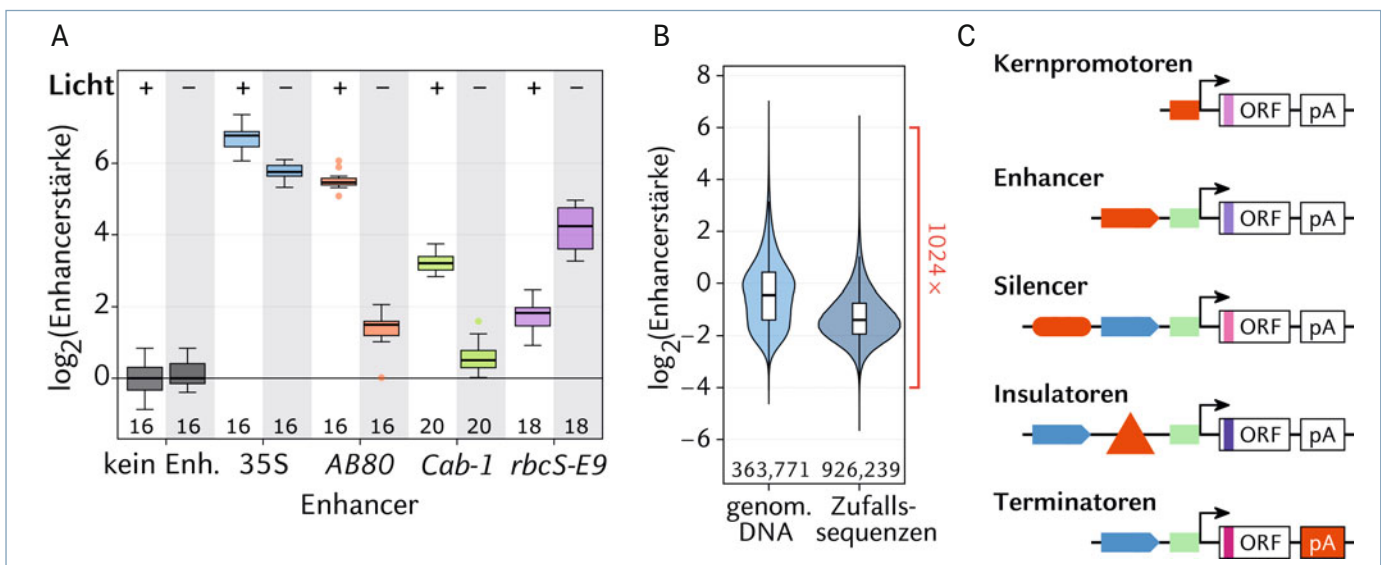
Aktivität *cis*-regulatorischer Elemente quantitativ fein abgestimmt und kontextabhängig sein muss. Daher beruhen sowohl evolutionäre Anpassungen als auch die Domestizierung von Nutzpflanzen häufig auf Veränderungen *cis*-regulatorischer Elemente [2]. Folglich sind sie attraktive Ziele für die gezielte Genomeditierung, um ertragreichere und widerstandsfähigere Nutzpflanzen zu entwickeln, die einen Beitrag zur Ernährungssicherheit leisten können. Um *cis*-regulatorische Elemente gezielt verbessern zu können, ist es jedoch notwendig, ein tiefergehendes Verständnis von deren Funktion und Aktivität zu entwickeln.

Das Prinzip von Plant STARR-seq

Die Aktivität *cis*-regulatorischer Elemente wurde lange Zeit mithilfe von klassischen Reporterassays bestimmt. Diese Ansätze



▲ **Abb. 1:** Prinzip von Plant STARR-seq zur Bestimmung der Aktivität *cis*-regulatorischer Elemente im Hochdurchsatz. **A, B,** Vergleich des ursprünglichen STARR-seq Assays (A) und der pflanzenspezifischen Anpassung, Plant STARR-seq (B). Während bei Ersterem die Enhancerkandidaten (blau) in die 3'-UTR eines Reportergens kloniert werden, werden sie bei Plant STARR-seq direkt stromaufwärts vom 35S-Minimalpromotor (grün) platziert und mit einem Reportergen mit einem DNA-Barcode (violett) kombiniert. **C,** Die Häufigkeit der Barcodes in DNA und mRNA transformierter Pflanzenzellen wird mittels Hochdurchsatz-Sequenzierung bestimmt. Das Verhältnis der Häufigkeit in der mRNA und DNA (mRNA/DNA) dient als Maß für die Aktivität des zugehörigen Enhancers.



▲ **Abb. 2:** Anwendungen von Plant STARR-seq. **A**, Quantitative Messung von Enhancer-Aktivitäten unter spezifischen Umweltbedingungen. Daten aus [10]. **B**, Erfassung der Aktivität von bis zu einer Million *cis*-regulatorischer Elemente über einen großen Dynamikbereich (unveröffentlichte Daten). **C**, Durch Veränderung der Reporter-Konstrukte lässt sich Plant STARR-seq auf weitere *cis*-regulatorische Elemente (rot markiert) anwenden.

erlauben jedoch nur die Analyse einzelner Sequenzen und erschweren somit den systematischen Vergleich der regulatorischen Aktivität von mehreren Sequenzen oder Sequenzvarianten. Als Weiterentwicklung dazu wurde mit *Self-transcribing Active Regulatory Region Sequencing* (STARR-seq) ein Hochdurchsatzverfahren zur Messung von Enhanceraktivität in *Drosophila* entwickelt [3]. Für diese Methode werden Enhancerkandidaten in die 3'-untranslatierte Region (3'-UTR) eines Reportergens kloniert (**Abb. 1A**). Testsequenzen mit Enhanceraktivität erhöhen die Expression der Reporter-mRNA, in die sie integriert sind. Die mittels Hochdurchsatzsequenzierung bestimmte Häufigkeit der zu einer Testsequenz gehörenden Reporter-mRNA dient als Maß für deren Enhancerstärke.

Klassisches STARR-seq ist für Pflanzen nicht direkt übertragbar, da Pflanzenenhancer in transkribierten Regionen (und damit auch in der 3'-UTR) oft keine oder nur geringe Aktivität zeigen [4, 5]. Daher wurde Plant STARR-seq als angepasste Methode für Pflanzen entwickelt, bei der die Testsequenzen nicht im transkribierten Bereich liegen (**Abb. 1B**). Für Plant STARR-seq wird eine Bibliothek von Enhancerkandidaten direkt stromaufwärts eines 35S-Minimalpromotors platziert und mit einem Reportergen kombiniert, das einen kurzen DNA-Barcode enthält (**Abb. 1C**). Der Barcode ermöglicht die eindeutige Zuordnung der Reporter-mRNA zu dem zugehörigen Enhancerkandidaten. Die

Bibliothek wird anschließend zur transienten Transformation von Pflanzenzellen, beispielsweise Tabakblättern oder Maisprotoplasten, verwendet. Nach ein bis zwei Tagen wird daraus die Reporter-mRNA extrahiert. Mittels Hochdurchsatzsequenzierung werden schließlich die extrahierte mRNA sowie die DNA der Reportergenbibliothek sequenziert. Da Enhancer die mRNA-Transkription erhöhen, gibt das Verhältnis der Häufigkeit von mRNA zu DNA direkt Aufschluss über die Aktivität des jeweiligen Enhancerkandidaten.

Messung regulatorischer Aktivität mit Plant STARR-seq

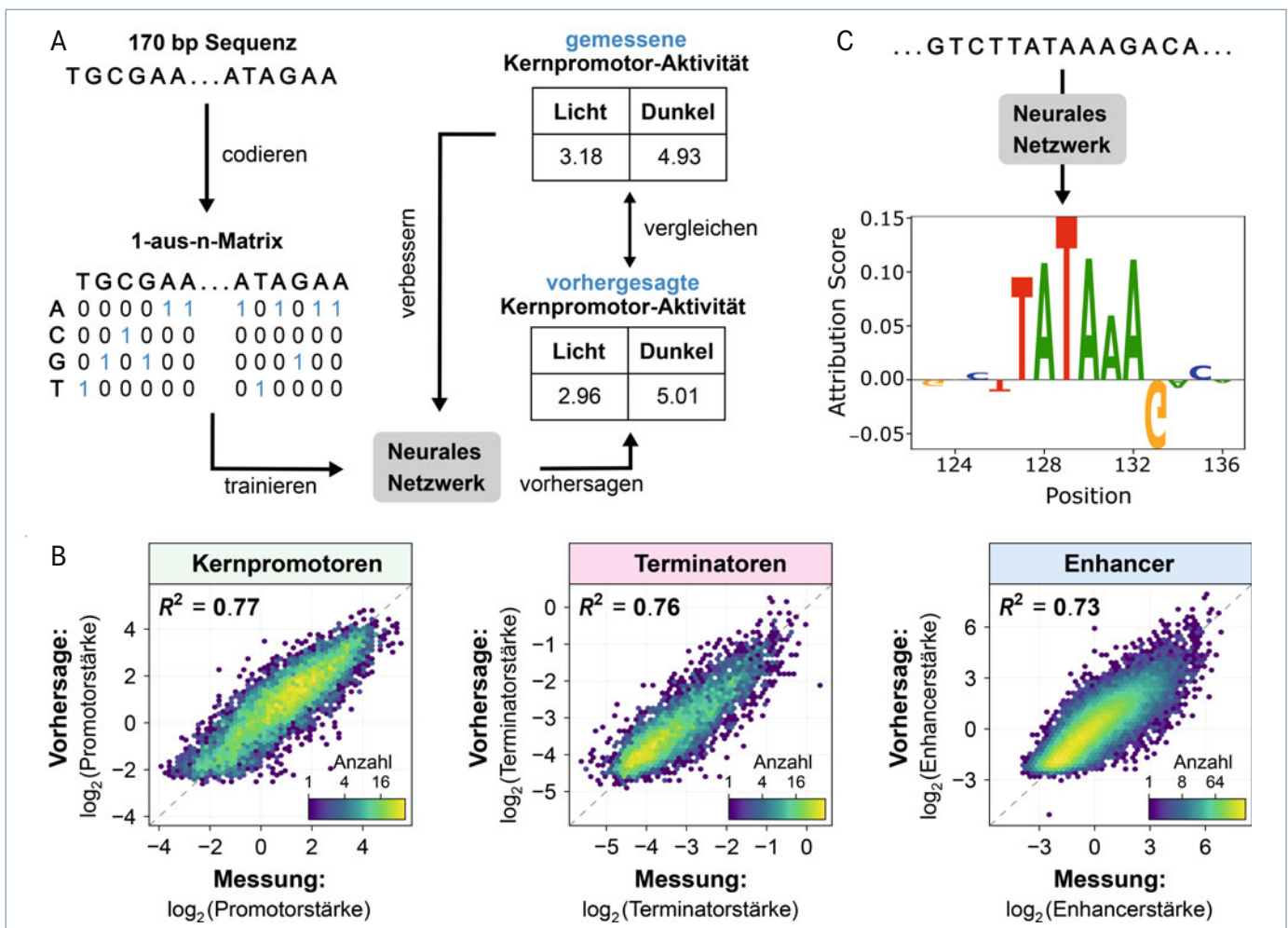
Mit Plant STARR-seq können unterschiedliche Aspekte der Aktivität von *cis*-regulatorischen Elementen systematisch untersucht werden. So ermöglicht die Methode die quantitative Bestimmung der Aktivität von Enhancern unter verschiedenen Umweltbedingungen (**Abb. 2A**) und in unterschiedlichen Spezies, wie etwa Tabak und Mais. Durch den Einsatz von Hochdurchsatzsequenzierung können mit Plant STARR-seq sehr große Bibliotheken mit bis zu einer Million Testsequenzen in einem einzelnen Experiment untersucht werden (**Abb. 2B**). Die daraus resultierenden Aktivitätswerte decken dabei einen großen Dynamikbereich ab, der selbst geringe Unterschiede in der regulatorischen Aktivität sichtbar macht.

Über Enhancer hinaus kann Plant STARR-seq auch genutzt werden, um weitere *cis*-

regulatorische Elemente zu untersuchen (**Abb. 2C**). Beispielsweise wurde die Stärke von über 75.000 Kernpromotoren aus *Arabidopsis*, *Sorghum* und Mais bestimmt [6]. Dabei wurden mehrere Kernpromotorelemente, darunter die TATA-Box, charakterisiert, die einen entscheidenden Einfluss auf die Promotorstärke haben. Neben *cis*-regulatorischen Elementen, die die Transkription initiieren oder verstärken, lassen sich mit Plant STARR-seq auch Elemente wie Insulatoren und Silencer untersuchen, die die Transkription begrenzen oder unterdrücken. In einer aktuellen Studie wurden bekannte Insulatoren in kleinere Fragmente zerlegt und systematisch getestet [7]. Viele dieser Fragmente blockierten die Wirkung von Enhancern und zeigten dementsprechend Insulator-Aktivität. Stromaufwärts von schwachen Enhancern wirkten die gleichen Fragmente jedoch wie Silencer, was die Kontextabhängigkeit regulatorischer Elemente verdeutlicht. Selbst Elemente, die nicht direkt die Transkriptionsregulation beeinflussen, können mittels Plant STARR-seq untersucht werden. So wurde der Einfluss nahezu aller Terminatoren aus *Arabidopsis* und Mais auf mRNA-Stabilität und Prozessierung ermittelt [8].

Vorhersage von regulatorischer Aktivität mit maschinellem Lernen

Da mit Plant STARR-seq die regulatorische Aktivität von sehr vielen DNA-Sequenzen direkt und quantitativ bestimmt werden



▲ **Abb. 3:** Nutzung von Plant STARR-seq-Daten für maschinelles Lernen. **A,** Auf Basis von Plant STARR-seq Daten können Computermodelle trainiert werden, um die Aktivität regulatorischer Sequenzen vorherzusagen. **B,** Vorhersage der Aktivität verschiedener *cis*-regulatorischer Elemente. Daten aus [6, 8] sowie unveröffentlicht. **C,** Die Interpretation der Modellvorhersagen ermöglicht die Identifikation von Sequenzmerkmalen (wie z. B. einer TATA-Box), die für die regulatorische Aktivität relevant sind.

kann, eignen sich die gewonnenen Daten besonders gut für maschinelles Lernen. Im Gegensatz zu vielen Methoden der funktionellen Genomik, die auf binärem Peak-Calling beruhen, erzeugt Plant STARR-seq kontinuierliche Aktivitätsmessungen und formuliert die Lernaufgabe damit als Regressionsproblem [9]. Diese kontinuierlichen Werte bewahren Informationen über Aktivitätsstärke und -dynamik, was wichtig ist, um auch subtile regulatorische Effekte zu modellieren.

Für Plant STARR-seq-Modelle bestehen die Eingabedaten aus Sequenzen fester Länge (170 bp in unserem Fall; **Abb. 3A**), die sich einfach im 1-aus-n-Code (One-Hot-Encoding) mathematisch darstellen lassen. Dabei wird jede Base in einen vierstelligen Vektor umgewandelt (z. B.: [1, 0, 0, 0] für A oder [0, 1, 0, 0] für C), sodass für jede Sequenz eine 170 x 4-Matrix entsteht. Diese Matrizen

können anschließend als Eingabe für eine Vielzahl von maschinellen Lernmodellen verwendet werden, von künstlichen neuronalen Netzwerken bis hin zu komplexeren Architekturen. Mit solchen Modellen kann die Aktivität von Kernpromotoren, Terminatoren und Enhancern mit hoher Genauigkeit vorhergesagt werden (**Abb. 3B**).

Entgegen einer weit verbreiteten Annahme fungieren diese Modelle dabei nicht als reine „Black Boxes“. Mithilfe von Feature-Attributionsmethoden lässt sich nachvollziehen, welche Sequenzmerkmale die Vorhersagen der Modelle beeinflussen (**Abb. 3C**). Bei Enhancern hoben diese Analysen vor allem Transkriptionsfaktor-Bindemotive hervor, während sich die Modelle bei Kernpromotoren stark auf das Vorhandensein und die Positionierung von TATA-Boxen stützten. Solche Analysen sind sehr hilfreich, um funktionelle DNA-Motive in *cis*-regulatori-

schen Elementen zu identifizieren oder die Auswirkungen von gezielten Mutationen vorherzusagen.

Zukunftsaussichten

Angesichts der zentralen Rolle der Genregulation für viele biologische Prozesse ist ihr Verständnis von grundlegender Bedeutung für die Pflanzenforschung und -züchtung.

Plant STARR-seq beschleunigt durch Hochdurchsatzansätze die Analyse *cis*-regulatorischer Elemente erheblich. Neben einzelnen Elementen können auch Wechselwirkungen zwischen mehreren Elementen untersucht werden, die aufgrund ihrer kombinatorischen Komplexität bislang nur schwer zugänglich waren. Die dabei gewonnenen Daten können zur Erstellung von Computermodellen genutzt werden, um die regulatorische Aktivität von ungetesteten Sequenzen

vorherzusagen. Auf diese Weise leistet Plant STARR-seq einen direkten Beitrag zur Entwicklung ertragreicherer und resilienterer Nutzpflanzen, sei es durch klassische Züchtung oder durch die gezielte Veränderung regulatorischer DNA. ■

Literatur

- [1] Schmitz RJ, Grotewold E, Stam M (2022) Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell* 34: 718–741. <https://doi.org/10.1093/plcell/koab281>
- [2] Li X, Schmitz RJ (2025) Cis-regulatory dynamics in plant domestication. *Trends Genet* 41:984–994. <https://doi.org/10.1016/j.tig.2025.02.005>
- [3] Arnold CD, Gerlach D, Stelzer C et al. (2013) Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* 339: 1074–1077. <https://doi.org/10.1126/science.1232542>
- [4] Jores T, Tonnie J, Dorrity MW et al. (2020) Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. *Plant Cell* 32:2120–2131. <https://doi.org/10.1105/tpc.20.00155>
- [5] Voichek Y, Hristova G, Mollá-Morales A et al. (2024) Widespread position-dependent transcriptional regulatory sequences in plants. *Nat Genet* 56: 2238–2246. <https://doi.org/10.1038/s41588-024-01907-3>
- [6] Jores T, Tonnie J, Wrightsman T et al. (2021) Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. *Nat Plants* 7: 842–855. <https://doi.org/10.1038/s41477-021-00932-y>
- [7] Jores T, Mueth NA, Tonnie J et al. (2025) Small DNA elements can act as both insulators and silencers in plants. *Plant Cell* 37: koaf084. <https://doi.org/10.1093/plcell/koaf084>
- [8] Gorjifard S, Jores T, Tonnie J et al. (2024) Arabidopsis and maize terminator strength is determined by GC content, polyadenylation motifs and cleavage probability. *Nat Commun* 15: 5868. <https://doi.org/10.1038/s41467-024-50174-7>
- [9] Barshai M, Tripto E, Orenstein Y (2020) Identifying Regulatory Elements via Deep Learning. *Annu Rev Biomed Data Sci* 3: 315–338. <https://doi.org/10.1146/annurev-bio-datasci-022020-021940>
- [10] Jores T, Tonnie J, Mueth NA et al. (2024) Plant enhancers exhibit both cooperative and additive interactions among their functional elements. *Plant Cell* 36: 2570–2586. <https://doi.org/10.1093/plcell/koae088>

Funding: Open Access funding enabled and organized by Projekt DEAL.

Open Access: This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Korrespondenzadresse:

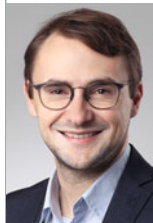
Dr. Tobias Jores
Institut für Synthetische Biologie
Mathematisch-Naturwissenschaftliche Fakultät
Universität Düsseldorf
Universitätsstraße 1
D-40225 Düsseldorf
tobias.jores@hhu.de
www.synthetic-biology.hhu.de/research/dr-jores-group

AUTORINNEN UND AUTOREN



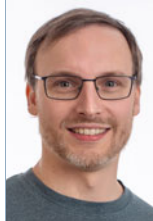
Julia Macholl

2016–2023 Studium der Molekularbiologie an der Universität Bielefeld. Seit 2024 Doktorandin am Institut für Synthetische Biologie der Universität Düsseldorf bei Dr. T. Jores.



Sebastian Triesch

2013–2019 Studium der Biologie, Chemie und Biochemie an den Universitäten Wuppertal und Düsseldorf. 2019–2024 Promotion in Biochemie der Pflanzen im Institut von Prof. Dr. A. Weber an der Universität Düsseldorf. 2024–2025 Bioinformatiker in der NGS Core Unit des Universitätsklinikums Düsseldorf. Seit 2025 PostDoc an der Universität Düsseldorf in der Arbeitsgruppe von Dr. T. Jores.



Tobias Jores

2008–2013 Studium der Biochemie an der Universität Tübingen; dort bis 2018 Promotion in Biochemie im Labor von Prof. Dr. D. Rapaport. 2019–2023 PostDoc an der University of Washington in Seattle, WA, USA, in den Laboren von Prof. Dr. S. Fields und Prof. Dr. C. Queitsch. Seit 2024 Emmy-Noether-Gruppenleiter im Institut für Synthetische Biologie der Universität Düsseldorf.