

Machine-Learning-Methode entdeckt verborgene Muster in der DNA-Methylierung

In einer jüngst in Nature Communications veröffentlichten Studie stellen Forschende aus Berlin, Potsdam und Jena eine neue Methode zur Analyse des Epigenoms vor. Die Machine-Learning-Methode erkennt unterschiedlich methylierte DNA-Regionen ohne Probenkennzeichnung. So lassen sich bislang verborgene biologische Muster sowie neue Untergruppen von Zellen oder Erkrankungen identifizieren. Die Analyse von Blutzellen, Bauchspeicheldrüsenkrebs und Hirntumoren bestätigte bekannte biologische Zusammenhänge und deckte neue regulatorische Prozesse auf. Die metilene3-Methode eröffnet neue Möglichkeiten für ein besseres Verständnis von Krankheitsmechanismen und die Identifizierung potenzieller Biomarker.

Jena/Potsdam. Die Aktivität unserer Gene wird nicht nur durch die DNA-Sequenz bestimmt. Anheftungen kleiner chemischer Verbindungen, sogenannte Methylgruppen, beeinflussen, welche Gene aktiv sind und welche abgeschaltet bleiben. Die DNA-Methylierung ist damit ein zentraler Bestandteil des Epigenoms.

Veränderungen des Epigenoms spielen eine entscheidende Rolle bei der Entwicklung unseres Körpers, beeinflussen unseren Alterungsprozess und sind für zahlreiche Erkrankungen wie Krebs relevant. Um solche Veränderungen zu verstehen, suchen Forschende gezielt nach unterschiedlich methylierten DNA-Regionen (Differentially Methylated Regions, DMRs). Bisherige Verfahren setzen jedoch meist voraus, dass die untersuchten Proben bereits bekannten Gruppen zugeordnet sind – beispielsweise gesundem oder erkranktem Gewebe. Bei komplexen klinischen Datensätzen sind diese Informationen jedoch häufig unbekannt.

Neue Methode erkennt Unterschiede auch ohne bekannte Gruppen

Mit metilene3, einem neuen Software-Tool, haben Forschende nun ein Verfahren entwickelt, das diese Einschränkung überwindet. Die Software kann DNA-Methylierungsmuster sowohl zwischen bereits definierten Gruppen vergleichen (überwachter Modus) als auch Proben anhand ihrer epigenetischen Eigenschaften eigenständig clustern (unüberwachter Modus).

Bei letzterem Modus wird nach DMRs gesucht, ohne dass die Proben zuvor in Gruppen wie etwa „gesund“ oder „krank“ eingeteilt sein müssen. Die Software segmentiert das Genom eigenständig anhand der Methylierungssignale und gruppiert die Proben vollautomatisch. Durch diese automatische Klassifizierung lassen sich epigenetische Ähnlichkeiten und Entwicklungszusammenhänge zwischen den Proben sichtbar machen. So lassen sich bislang unbekannte Zelltypen oder Krankheitsuntergruppen erkennen und Regionen identifizieren, in denen die Proben sowohl bereits bekannten Erkrankungen ähneln als auch sich davon unterscheiden. Gleichzeitig bleiben die biologischen Unterschiede nachvollziehbar, da sich jede Gemeinsamkeit und jeder Unterschied auf bestimmte Methylierungsmuster zurückführen lassen.

Praxistest anhand realer Datensätze

Anschließend testeten die Forschenden die Methode an verschiedenen biologischen Datensätzen.

Bei menschlichen Blutzellen rekonstruierte metilene3 allein anhand ihrer DNA-Methylierung die bekannten Entwicklungswege verschiedener Immunzelltypen. Darüber hinaus identifizierte die Software regulatorische DNA-Bereiche, die mit Transkriptionsfaktoren verknüpft sind und die Identität dieser Zelltypen steuern. Die Ergebnisse belegen, dass sich mit dem Verfahren nicht nur Zellgruppen unterscheiden, sondern auch funktionell relevante regulatorische Elemente identifizieren lassen.

Auch bei Tumordaten erwies sich der Ansatz als äußerst leistungsfähig. In Datensätzen zu Glioblastomen identifizierte metilene3 verschiedene molekulare Untergruppen der Tumoren und erkannte sogar einzelne Proben mit ungewöhnlichen biologischen Eigenschaften.

In Gewebeproben von Bauchspeicheldrüsenkrebs konnte die Software zudem die schrittweise Entwicklung vom gesunden Gewebe über Krebsvorstufen bis hin zum Tumor nachvollziehen. Dabei entdeckten die Forschenden DNA-Regionen, in denen die Bindungsstellen der Transkriptionsfaktoren NF- κ B und NFAT besonders häufig zusammen vorkommen. Diese Regionen könnten eine wichtige Rolle bei der Entstehung von Bauchspeicheldrüsenkrebs spielen und ein wichtiger Ansatzpunkt für weitere Untersuchungen sein. Die biologische Hypothese muss jedoch nun experimentell überprüft werden.

„Die durch das Machine-Learning-Verfahren identifizierten Veränderungen der DNA-Methylierung im Krebs ermöglichen direkte Rückschlüsse auf die molekularen Störungen bei Transkriptionsfaktoren“, unterstreicht Prof. Helene Kretzmer vom Hasso-Plattner-Institut in Potsdam. Gerade bei medizinischen Fragestellungen seien die Anforderungen an die Erklärbarkeit von Vorhersagen besonders hoch.

„Unsere Methode eröffnet neue Möglichkeiten, bislang verborgene biologische Zusammenhänge und Therapieansätze aufzufinden“, ergänzt Prof. Steve Hoffmann vom Leibniz-Institut für Alternsforschung – Fritz-Lipmann-Institut (FLI) in Jena.

Neue Möglichkeiten für Forschung und Medizin

Mit metilene3 steht den Forschenden nun ein Werkzeug zur Verfügung, das die Analyse komplexer DNA-Methylierungsdaten deutlich erweitert. Insbesondere bei heterogenen Gewebeproben oder klinischen Datensätzen, in denen sich biologische Gruppen nicht immer im Vorfeld definieren lassen, kann die Methode dabei helfen, bislang verborgene, neue Zusammenhänge aufzudecken. Die Autoren der Studie sehen daher großes Potenzial der Software für die Erforschung von Alterungsprozessen, Krebs und anderen Erkrankungen sowie für die Identifizierung neuer Biomarker. Das Tool soll künftig auch für weitere Sequenzierungstechnologien und Einzelzellanalysen eingesetzt werden, um die Analyse epigenetischer Daten noch breiter nutzbar zu machen.

„Je besser wir die epigenetischen Muster verstehen, desto gezielter können wir die biologischen Prozesse dahinter entschlüsseln. Die Software bietet dafür eine wichtige Grundlage und könnte dabei helfen, aus großen Datensätzen neue Hypothesen für die biomedizinische Forschung abzuleiten“, sagt Kretzmer.

Publikation

metilene3: identifying DMRs across multiple conditions with auto-classification. Zhu Z, Bernhart SH, Jühling F, Kretzmer H, Hoffmann S. Nat Commun. 2026 Jul 4;17(1):5848. doi: 10.1038/s41467-026-74931-y.
<https://www.nature.com/articles/s41467-026-74931-y>

Der Quellcode des Programms ist frei über das metilene3 GitHub-Repository zugänglich. <https://github.com/zzhu1372/metilene3>

Hintergrundinformation

Das Leibniz-Institut für Alternsforschung – Fritz-Lipmann-Institut e.V. (FLI) in Jena ist eine von Bund und dem Freistaat Thüringen gemeinsam finanzierte Forschungseinrichtung in der Leibniz-Gemeinschaft. Am FLI wird international sichtbare Spitzenforschung zur Biologie des Alterns auf molekularer, zellulärer und systemischer Ebene betrieben. Wissenschaftlerinnen und Wissenschaftler aus rund 40 Ländern erforschen die Mechanismen des Alterns, um dessen Ursachen besser zu verstehen und Grundlagen für Strategien zu schaffen, die gesundes Altern fördern. Weitere Informationen: <https://www.leibniz-flj.de/de/>

Die Leibniz-Gemeinschaft verbindet 96 eigenständige Forschungseinrichtungen. Ihre Ausrichtung reicht von den Natur-, Ingenieur- und Umweltwissenschaften über die Wirtschafts-, Raum- und Sozialwissenschaften bis zu den Geisteswissenschaften.

Leibniz-Institute widmen sich gesellschaftlich, ökonomisch und ökologisch relevanten Fragen. Sie betreiben erkenntnis- und anwendungsorientierte Forschung, auch in den übergreifenden Leibniz-Forschungsverbünden, sind oder unterhalten wissenschaftliche Infrastrukturen und bieten forschungsbasierte Dienstleistungen an. Die Leibniz-Gemeinschaft setzt Schwerpunkte im Wissenstransfer, vor allem mit den Leibniz-Forschungsmuseen. Sie berät und informiert Politik, Wissenschaft, Wirtschaft und Öffentlichkeit.

Leibniz-Einrichtungen pflegen enge Kooperationen mit den Hochschulen – in Form der Leibniz-WissenschaftsCampi, mit der Industrie und anderen Partnern im In- und Ausland. Die Leibniz-Institute unterliegen einem transparenten und unabhängigen Begutachtungsverfahren. Aufgrund ihrer gesamtstaatlichen Bedeutung fördern Bund und Länder die Institute der Leibniz-Gemeinschaft gemeinsam. Die Leibniz-Institute beschäftigen rund 21.400 Personen, darunter 12.170 Wissenschaftlerinnen und Wissenschaftler. Der Gesamtetat der Institute liegt bei 2,3 Milliarden Euro. (<http://www.leibniz-gemeinschaft.de>).