

Pressemitteilung

Friedrich-Schiller-Universität Jena

Sebastian Hollstein

23.11.2020

<http://idw-online.de/de/news758436>

Forschungsergebnisse, Wissenschaftliche Publikationen
Biologie, Chemie, Informationstechnik, Medizin
überregional



Durch maschinelles Lernen Stoffklassen erkennen

Bioinformatiker der Friedrich-Schiller-Universität Jena haben gemeinsam mit Kollegen aus Finnland und den USA eine weltweit einmalige Methode entwickelt, bei der alle Metaboliten in einer Probe berücksichtigt werden können und sich somit der Erkenntnisgewinn bei der Untersuchung solcher Moleküle erheblich vergrößert. Über seinen Erfolg berichtet das Team aktuell im renommierten Fachjournal „Nature Biotechnology“.

Alles, was lebt, hat Metabolite, produziert Metabolite und verbraucht Metabolite. Diese Moleküle gehen als Zwischen- und Endprodukte aus chemischen Prozessen innerhalb des Stoffwechsels eines Organismus hervor. Damit haben sie nicht nur eine enorme Bedeutung für unser Leben, sondern sie liefern auch wertvolle Informationen über den Zustand eines Lebewesens oder einer Umgebung. So lassen sich anhand von Metaboliten beispielsweise Krankheiten erkennen oder – im Bereich der Umwelttechnologie – Trinkwasserproben untersuchen. Doch die Diversität dieser chemischen Verbindungen bereitet der Wissenschaft Schwierigkeiten. Denn bisher sind nur vergleichsweise wenige Moleküle bekannt und definiert. Wird eine Probe im Labor analysiert, so kann bislang nur ein relativ kleiner Teil davon wirklich identifiziert werden – der Großteil an Molekülen bleibt unbekannt.

Bioinformatiker der Friedrich-Schiller-Universität Jena haben nun gemeinsam mit Kollegen aus Finnland und den USA eine weltweit einmalige Methode entwickelt, bei der alle Metaboliten in einer Probe berücksichtigt werden können und sich somit der Erkenntnisgewinn bei der Untersuchung solcher Moleküle erheblich vergrößert. Über seinen Erfolg berichtet das Team aktuell im renommierten Fachjournal „Nature Biotechnology“.

Struktureigenschaften lernen, erkennen und zuordnen

„Bei der Massenspektrometrie, eine der meistgenutzten experimentellen Methoden zur Analyse von Metaboliten, werden nur die Moleküle identifiziert, die durch den Abgleich mit einer Datenbank eindeutig zugeordnet werden können. Alle anderen, bisher unbekanntes Moleküle, die in der Probe enthalten sind, liefern keine Informationen“, erklärt Prof. Dr. Sebastian Böcker von der Universität Jena. „Mit unserem neu entwickelten Verfahren namens CANOPUS entlocken wir allerdings auch den unidentifizierten Metaboliten in einer Probe wertvolle Erkenntnisse, da wir sie bereits bekannten Stoffklassen zuordnen können.“

CANOPUS funktioniert in zwei Phasen: Zunächst erzeugt das Verfahren, aus dem mittels Massenspektrometrie gemessenen Fragmentierungsspektrum, einen sogenannten molekularen Fingerabdruck. Dieser beinhaltet Informationen über die Struktureigenschaften des gemessenen Moleküls. In einem zweiten Schritt ordnet das System den Metaboliten mithilfe des Fingerabdrucks einer bestimmten Stoffklasse zu, ohne diesen dafür identifizieren zu müssen.

Das System lernt selbst

„Maschinelle Lernverfahren benötigen in der Regel große Datenmengen, um trainiert zu werden. Unser zweistufiges Verfahren hingegen ermöglicht es, im ersten Schritt auf einer vergleichsweise kleinen Datenmenge von zehntausenden Fragmentierungsspektren zu trainieren, um dann im zweiten Schritt aus Millionen von Strukturen die charakteristischen Struktureigenschaften zu bestimmen, die für eine Stoffklasse signifikant sind“, erklärt Dr. Kai Dührkop von der Universität Jena. Das System detektiert also diese Struktureigenschaften bei einem unbekanntem Molekül innerhalb einer Probe und ordnet es dann einer bestimmten Stoffklasse zu. „Allein diese Information reicht bereits aus, um viele wichtige Fragestellungen zu beantworten“, betont Böcker. „Die eindeutige Identifikation eines Metabolits wäre weitaus aufwendiger und ist häufig überhaupt nicht notwendig.“ Insgesamt liege dem CANOPUS-Verfahren ein tiefes neuronales Netz von rund 2.500 Verbindungsklassen zugrunde.

Mit ihrer Methode haben die Jenaer Bioinformatiker beispielsweise die Darmflora von Mäusen verglichen, bei denen eine Versuchsgruppe mit Antibiotika behandelt worden war. Die Untersuchungen zeigen, welche Metabolite die Maus und ihre Darmflora produzieren. Solche Forschungsergebnisse können wichtige Erkenntnisse über das menschliche Verdauungs- und Stoffwechselsystem ermöglichen. Durch zwei weitere Anwendungsbeispiele, die sie in ihrer Studie ausführen, zeigen die Jenaer Wissenschaftler die Funktionalität und Aussagekraft des CANOPUS-Verfahren.

Jenaer Molekül-Suchmaschine millionenfach genutzt

Mit der neue Methode erweitern die Jenaer Bioinformatiker die Möglichkeiten der Suchmaschine für molekulare Strukturen „CSI:FingerID“, die sie der internationalen Forschungsgemeinschaft seit rund fünf Jahren zur Verfügung stellen. Weltweit nutzen Forscher dieses Angebot inzwischen tausende Male täglich, um ein Massenspektrum aus einer Probe mit verschiedenen Online-Datenbanken abzugleichen und so einen Metaboliten genauer bestimmen zu können. „Wir nähern uns der einhundertmillionsten Anfrage und sind uns sicher, dass das CANOPUS-Angebot die Nutzerzahlen weiter steigen lassen wird“, sagt Sebastian Böcker.

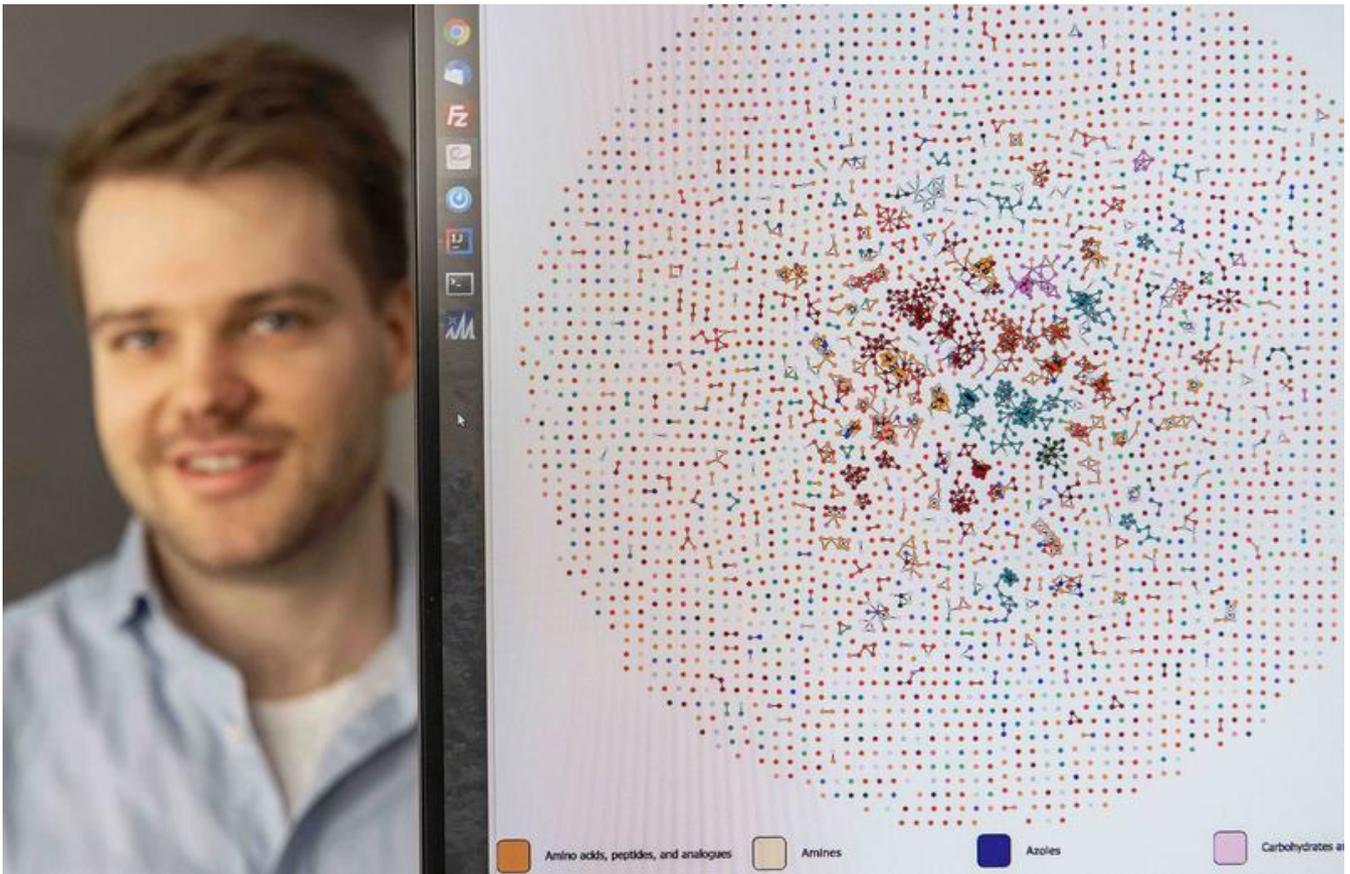
Das neue Verfahren stärkt die Metabolomik, also die Erforschung dieser omnipräsenten Moleküle, und fördert ihr Potenzial in vielen Bereichen, etwa in der Pharmazie. Viele bereits lange Zeit verwendete Arzneiwirkstoffe sind Metabolite, beispielsweise das bekannte Penicillin – weitere könnten mit ihrer Hilfe entwickelt werden.

wissenschaftliche Ansprechpartner:

Prof. Dr. Sebastian Böcker
Institut für Informatik der Friedrich-Schiller-Universität Jena
Ernst-Abbe-Platz 2, 07743 Jena
Tel.: 03641 / 946450
E-Mail: [sebastian.boecker\[at\]uni-jena.de](mailto:sebastian.boecker[at]uni-jena.de)

Originalpublikation:

K. Dührkop, L.-F. Nothias, M. Fleischauer., R. Reher, M. Ludwig, M. A. Hoffmann, D. Petras, W. H. Gerwick, J. Rousu, P. C. Dorrestein, S. Böcker: Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra, Nature Biotechnology, 2020: <https://doi.org/10.1038/s41587-020-0740-8>



Dr. Kai Dührkop von der Universität Jena präsentiert die Visualisierung eines gemessenen Datensatzes mit der Software CANOPUS.

Foto: Jens Meyer/Uni Jena