

Pressemitteilung**Max-Planck-Institut für molekulare Genetik****Dr. Martin Ballaschk**

12.04.2021

<http://idw-online.de/de/news766538>Forschungsergebnisse, Wissenschaftliche Publikationen
Biologie, Medizin
überregional**Mehr als die Summe der Mutationen – 165 neue Krebsgene mit Hilfe maschinellen Lernens identifiziert**

Ein neuer Algorithmus sagt Gene vorher, die an der Entstehung von Krebs beteiligt sein können, deren DNA-Sequenz jedoch nicht zwangsläufig verändert ist. Ein Berliner Forschungsteam kombinierte unterschiedlichste Daten, analysierte sie mit „künstlicher Intelligenz“ und identifizierte so zahlreiche Krebsgene. Für die gezielte Krebstherapie in der personalisierten Medizin sowie für die Entwicklung von Biomarkern ergeben sich so neue Perspektiven.

Bei Krebs geraten Zellen außer Kontrolle. Sie vermehren sich uneingeschränkt, verdrängen dadurch andere Gewebe, zerstören Organe und beeinträchtigen so wichtige Lebensfunktionen. Die Ursache für dieses Wachstum sind in aller Regel angehäuften DNA-Veränderungen in Krebsgenen – also Mutationen in jenen Genen, die die Entwicklung der Zelle steuern. Doch einige Krebsarten entstehen bereits bei nur sehr wenigen mutierten Genen. In diesen Fällen führen andere, wenig verstandene Ursachen zu der Erkrankung.

Ein Forschungsteam vom Max-Planck-Institut für molekulare Genetik (MPIMG) in Berlin und dem Institut für Computational Biology des Helmholtz Zentrums München hat nun einen neuen Algorithmus entwickelt, der mit Hilfe von maschinellem Lernen 165 zuvor unbekannte Krebsgene identifizierte. Von diesen Genen sind längst nicht alle mutiert – offenbar kann bereits ihre Fehlregulation zu Krebs führen. Die neu entdeckten Gene stehen jedoch alle in engem Austausch mit bereits bekannten Krebsgenen. Überdies sind sie überlebenswichtig für Tumorzellen, wie sich im Zellkulturexperiment herausstellte.

Das auf den Namen „EMOGI“ (für Explainable Multi-Omics Graph Integration) getaufte Programm kann auch erklären, welche zellulären Zusammenhänge jedes der identifizierten Gene zu einem Krebsgen machen. Wie die Forschenden um Annalisa Marsico im Fachblatt Nature Machine Intelligence schildern, kombiniert das Programm dafür zehntausende Datensätze aus Patientenproben. Diese enthalten neben Sequenzdaten mit Mutationen auch Informationen über DNA-Methylierungen, die Aktivität einzelner Gene und Interaktionen von Proteinen, die an zellulären Signalwegen beteiligt sind. Ein Deep-Learning-Algorithmus erkennt in diesen Daten die Muster und molekularen Gesetzmäßigkeiten, die zu Krebs führen.

„Idealerweise haben wir irgendwann eine komplette Übersicht über alle Krebsgene, die bei verschiedenen Betroffenen einen unterschiedlichen Einfluss auf das Fortschreiten der Krebserkrankung haben können“, sagt Marsico, bis vor kurzem Leiterin einer Forschungsgruppe am MPIMG und nun am Helmholtz Zentrum München. „Das ist die Grundlage für die personalisierte Krebstherapie.“

Bei diesen neuartigen Heilverfahren werden anders als bei konventionellen Krebsbehandlungen wie der Chemotherapie die Medikamente genau auf die Krebsart zugeschnitten. „In der Klinik kann für die Erkrankten dann die jeweils beste Therapie gefunden werden – also die wirksamste Behandlung mit den wenigsten Nebenwirkungen“, erklärt die

Forscherin. „Wir könnten zudem anhand der molekularen Eigenschaften den Krebs schon frühzeitig erkennen.“

„Nur wenn wir die Ursachen kennen, können wir sie auch effektiv bekämpfen oder korrigieren. Daher ist es so wichtig, möglichst viele Mechanismen zu identifizieren, die Krebs auslösen können“, sagt Marsico.

„Bisher fokussierte sich die Forschung vor allem auf krankmachende Veränderungen in der Erbgutsequenz, also im Bauplan der Zelle“, sagt Roman Schulte-Sasse, Doktorand im Team von Marsico und Erstautor der Publikation. „Dabei hat sich in den letzten Jahren gezeigt, dass auch epigenetische Störungen oder fehlgesteuerte Genaktivität zu Krebs führen können.“

Aus diesem Grund führte das Forschungsteam Daten zusammen, die auch das Geschehen in der Zelle abbilden und nicht nur auf Fehler im Bauplan abzielen. Zunächst bestätigten die Wissenschaftlerinnen und Wissenschaftler: Veränderungen der DNA-Sequenz, also Mutationen bis hin zu Vervielfachung von Erbgutabschnitten, sind tatsächlich die Haupttreiber von Krebserkrankungen sind. In einem zweiten Schritt machten sie dann die Gen-Kandidaten ausfindig, die in einem größeren Kontext zum eigentlich krebsauslösenden Gen stehen.

„Wir haben zum Beispiel Gene gefunden, deren Sequenz bei Krebs meist unverändert bleibt, die jedoch trotzdem für den Tumor unverzichtbar sind, weil sie beispielsweise die Energiezufuhr regulieren“, sagt Schulte-Sasse. Diese Gene sind auf andere Weise außer Kontrolle geraten, etwa aufgrund von chemischen Erbgut-Modifikationen wie Methylierungen. Diese verändern die DNA-Informationen selbst nicht, bestimmen jedoch über deren Genaktivität. „Solche Gene sind vielversprechende Angriffspunkte für Medikamente – weil sie aber quasi im Schatten agieren, können wir sie nur mit komplexen Algorithmen finden“, sagt Schulte-Sasse.

Das neue Programm erweitert die Kandidatenliste der Krebsgene, die in den letzten Jahren auf 700 bis 1000 Einträge angewachsen ist, um ein beachtliches Stück. Den versteckten Genen sind die Forschenden erst durch die Kombination von bioinformatischen Analysemethoden und neuesten Algorithmen der künstlichen Intelligenz auf die Spur gekommen.

„Die Wechselwirkungen von Proteinen und Genen lassen sich als mathematisch als Netzwerk, als sogenannter Graph abbilden“, sagt Schulte-Sasse. „Es ist, als wollte man das Schienennetz der Eisenbahn erraten. Jeder Bahnhof entspricht einem Protein oder Gen und jede Wechselwirkung stellt eine Zugverbindung dar.“

Mit der Hilfe von Deep Learning – also jenen Algorithmen, die der künstlichen Intelligenz in den letzten Jahren zum Durchbruch verholfen haben – entdeckten die Forschenden auch jene Zugverbindungen, die bisher unentdeckt blieben. Schulte-Sasse ließ den Computer zehntausende verschiedener Netzkarten aus 16 unterschiedlichen Krebsarten analysieren, jede enthielt zwischen 12.000 und 19.000 Datenpunkte.

In den Daten sind zahlreiche weitere interessante Details versteckt. „Wir sehen Muster, die abhängig vom jeweiligen Krebs und Gewebe sind“, sagt Marsico. „Wir sehen das als Hinweis, dass Tumoren je nach Organ durch unterschiedliche molekulare Mechanismen ausgelöst werden.“

Das EMOGI-Programm ist nicht auf Krebserkrankungen festgelegt, betonen die Forschenden. Theoretisch kann es vielfältige biologische Daten miteinander integrieren und dort Muster finden, erklärt die Forscherin. „Nützlich könnte das vor allem für komplexe Krankheiten sein, für die vielseitige Daten erhoben werden und bei denen die Gene eine wichtige Rolle spielen. Neben Krebs sind dies auch komplexe Stoffwechselerkrankungen wie zum Beispiel Diabetes.“

wissenschaftliche Ansprechpartner:

Prof. Annalisa Marsico
Gruppenleiterin

Helmholtz Zentrum München
+49(0)89 3187-43073
annalisa.marsico@helmholtz-muenchen.de

Roman Schulte-Sasse
Wissenschaftler
Max-Planck-Institut für molekulare Genetik
+49 (30) 8413-1172
sasse@molgen.mpg.de

Originalpublikation:

Roman Schulte-Sasse, Stefan Budach, Denes Hnisz, and Annalisa Marsico (2021): „Integration of Multi-Omics Data with Graph Convolutional Networks to Identify New Cancer Genes and their Associated Molecular Mechanisms.“ Nature Machine Intelligence. DOI: 10.1038/s42256-021-00325-y

URL zur Pressemitteilung: <https://dx.doi.org/10.1038/s42256-021-00325-y> – Link zur Publikation

URL zur Pressemitteilung: <https://www.molgen.mpg.de/4385741/> – Link zur Pressemitteilung



Den Tumor mit intelligenter Datenanalyse „durchleuchten“ – ein neuer Algorithmus macht’s möglich.
Ella Maru Studio
MPI f. mol. Genet.