

## Wie wirken sich seltene Erbgut-Varianten auf die Gesundheit aus? KI ermöglicht präzisere Vorhersagen

**Ob wir Veranlagungen für bestimmte Krankheiten haben, hängt im hohen Maße von den unzähligen Varianten in unserem Erbgut ab. Doch insbesondere bei Erbgut-Varianten, die in der Bevölkerung nur selten auftreten, ist der Einfluss auf die Ausprägung bestimmter krankhafter Merkmale bislang nur schwer zu ermitteln.**

Forschende aus dem Deutschen Krebsforschungszentrum (DKFZ), dem Europäischen Molekular Biologie Laboratorium (EMBL) und von der TU München stellen nun einen auf Deep Learning basierenden Algorithmus vor, der auch die Auswirkungen seltener Erbgut-Varianten vorhersagen kann. Das Verfahren ermöglicht, Personen mit hohen Erkrankungsrisiken präziser herauszufiltern, und erleichtert die Identifizierung von Genen, die an der Entstehung von Krankheiten beteiligt sind.

Das Erbgut jedes Menschen unterscheidet sich von dem seiner Mitmenschen in vielen Millionen Einzelbausteinen. Diese Unterschiede im Erbgut werden als Varianten bezeichnet. Zahlreiche dieser Varianten stehen mit besonderen biologischen Merkmalen und mit Erkrankungen in Zusammenhang. Solche Zusammenhänge werden in der Regel über so genannte genomweite Assoziationsstudien ermittelt.

Doch der Einfluss seltener Varianten, die mit einer Häufigkeit von nur 0,1 Prozent oder weniger in der Bevölkerung auftreten, geht bei den Assoziationsstudien rein statistisch oft unter. „Gerade die seltenen Varianten haben aber vielfach einen deutlich größeren Einfluss auf die Ausprägung eines biologischen Merkmals oder einer Erkrankung“, sagt Brian Clarke, einer der Erstautoren der Studie. „Sie können daher helfen, diejenigen Gene zu identifizieren, die bei der Entstehung einer Erkrankung eine Rolle spielen und die uns dann den Weg zu neuen Therapieansätzen weisen können“, ergänzt Eva Holtkamp, ebenfalls Erstautorin.

Um die Effekte seltener Varianten besser vorherzusagen, entwickelten die Teams um Oliver Stegle und Brian Clarke am DKFZ und EMBL sowie um Julien Gagneur von der TU München nun ein auf maschinellem Lernen basierendes Risikobewertungs-Tool. „DeepRVAT“ (rare variant association testing), wie die Forschenden die Methode benennen, setzt erstmals künstliche Intelligenz (KI) in genomischen Assoziationsstudien zur Entschlüsselung seltener Erbgut-Varianten ein.

Das Modell wurde zunächst an den Sequenzdaten (Exom-Sequenzen) von 161.000 Personen aus der UK Biobank trainiert. Zusätzlich speisten die Forscher Information zu genetisch beeinflussten biologischen Merkmalen der einzelnen Personen, z.B. Blutbild-Werte, sowie zu den an den Merkmalen beteiligten Genen ein. Die zum Training verwendeten Sequenzen umfassten rund 13 Millionen Varianten. Zu jeder davon liegen detaillierte so genannte „Annotationen“ vor, quantitative Angaben zu den möglichen Auswirkungen, die die jeweilige Variante auf zelluläre Prozesse oder auf die Proteinstruktur haben kann. Diese Annotationen waren ebenfalls zentraler Bestandteil des Trainings.

Nach dem Training ist DeepRVAT in der Lage, für einzelne Personen vorherzusagen, welche Gene durch seltene Erbgutvarianten in ihrer Funktion beeinträchtigt sind. Dazu errechnet es anhand individueller Varianten und deren Annotationen einen numerischen Wert, der das Ausmaß der Beeinträchtigung eines Gens und dessen mögliche Auswirkungen auf die Gesundheit beschreibt.

Die Forscher validierten DeepRVAT ebenfalls an Genomdaten aus der UK Biobank. Für 34 getestete Merkmale wie etwa krankheitsrelevante Blutwerte fand das Testsystem 352 Assoziationen zu beteiligten Genen und übertraf damit alle vorhandenen Modelle bei weitem. Die mit DeepRVAT erzielten Ergebnisse erwiesen sich außerdem als sehr robust und besser in unabhängigen Daten replizierbar als die Resultate alternativer Ansätze.

Eine wichtige weitere Einsatzmöglichkeit von DeepRVAT ist die Abschätzung der genetischen Veranlagung für bestimmte Krankheiten. Die Forschenden kombinierten DeepRVAT dazu mit einem polygenetischen Risiko-Scoring auf der Grundlage der häufigeren Erbgut-Varianten. Damit ließ sich die Treffsicherheit der Vorhersagen erheblich verbessern, insbesondere für Hochrisikovarianten. Zusätzlich stellte sich heraus, dass DeepRVAT für zahlreiche Krankheitsbilder – darunter etwa verschiedene Herz-Kreislauferkrankungen, Krebsarten, Stoffwechsel- und neurologische Erkrankungen – genetische Zusammenhänge erkannte, die mit bereits vorhandenen Tests nicht gefunden worden waren.

„DeepRVAT hat das Potenzial, die personalisierte Medizin wesentlich voranzubringen. Unser Verfahren funktioniert unabhängig von der Art des untersuchten Merkmals und kann flexibel mit anderen Testsystemen kombiniert werden“, sagt der Physiker und Datenwissenschaftler Oliver Stegle. Sein Team will das Risikobewertungs-Tool nun so schnell wie möglich in großen Studien weiter erproben und in die Anwendung bringen. In der Diskussion sind die Datenwissenschaftler beispielsweise bereits mit den Organisatoren von INFORM. Diese Studie hat das Ziel, für krebserkrankte Kindern, die einen Rückfall erleiden, auf der Basis von Genomdaten eine individuell zugeschnittene Behandlung zu identifizieren. Hier könnte DeepRVAT helfen, die genetischen Grundlagen bestimmter Krebserkrankung der Kinder aufzudecken.

„Für besonders vielversprechend halte ich das Potenzial von DeepRVAT für die Untersuchung seltener Krankheiten. Eine der größten Herausforderungen in der Forschung zu seltenen Krankheiten ist der Mangel an groß angelegten, systematischen Daten. Mit der Hilfe von KI und den halben Million Exomen in der UK Biobank können wir nun treffsicher vorhersagen, welche Varianten die Genfunktion am stärksten beeinträchtigen“, sagt Julien Gagneur von der TU München.

DeepRVAT soll in Zukunft in die Infrastruktur des Deutschen Humangenom-Phänom-Archivs (GHGA) integriert werden, um Anwendungen in der Diagnostik und Grundlagenforschung zu erleichtern. Ein Vorteil von DeepRVAT ist hierbei auch, dass die Methode deutlich weniger Rechenleistung benötigt als vergleichbare Modelle. DeepRVAT steht als nutzerfreundliches Softwarepaket zur Verfügung, das entweder mit den prätrainierten Risikobewertungsmodellen genutzt oder für spezielle Fragestellungen mit eigenen Datensätzen trainiert werden kann.

### Info:

Das Deutsche Krebsforschungszentrum (DKFZ) ist mit mehr als 3.000 Mitarbeiterinnen und Mitarbeitern die größte biomedizinische Forschungseinrichtung in Deutschland. Wissenschaftlerinnen und Wissenschaftler erforschen im DKFZ, wie Krebs entsteht, erfassen Krebsrisikofaktoren und suchen nach neuen Strategien, die verhindern, dass Menschen an Krebs erkranken. Sie entwickeln neue Methoden, mit denen Tumoren präziser diagnostiziert und Krebspatienten erfolgreicher behandelt werden können. Beim Krebsinformationsdienst (KID) des DKFZ erhalten Betroffene, Interessierte und Fachkreise individuelle Antworten auf alle Fragen zum Thema Krebs.

Um vielversprechende Ansätze aus der Krebsforschung in die Klinik zu übertragen und so die Chancen von Patientinnen und Patienten zu verbessern, betreibt das DKFZ gemeinsam mit exzellenten Universitätskliniken und Forschungseinrichtungen in ganz Deutschland Translationszentren:

- Nationales Centrum für Tumorerkrankungen (NCT, 6 Standorte)
- Deutsches Konsortium für Translationale Krebsforschung (DKTK, 8 Standorte)
- Hopp-Kindertumorzentrum (KITZ) Heidelberg
- Helmholtz-Institut für translationale Onkologie (HI-TRON) Mainz – ein Helmholtz-Institut des DKFZ
- DKFZ-Hector Krebsinstitut an der Universitätsmedizin Mannheim
- Nationales Krebspräventionszentrum (gemeinsam mit der Deutschen Krebshilfe)

Das DKFZ wird zu 90 Prozent vom Bundesministerium für Bildung und Forschung und zu 10 Prozent vom Land Baden-Württemberg finanziert und ist Mitglied in der Helmholtz-Gemeinschaft Deutscher Forschungszentren.

### Publikation:

Brian Clarke, Eva Holtkamp, Hakime Öztürk, Marcel Mück, Magnus Wahlberg, Kayla Meyer, Felix Munzlinger, Felix Brechtmann, Florian R. Hölzlwimmer, Jonas Lindner, Zhifen Chen, Julien Gagneur, Oliver Stegle: Integration of Variant annotations using deep set networks boosts rare variant testing. Nature Genetics 2024, DOI: <https://www.nature.com/articles/s41588-024-01919-z>

---

## Pressemitteilung

25.09.2024

Quelle: Deutsches Krebsforschungszentrum (DKFZ)

---

## Weitere Informationen

- ▶ [Deutsches Krebsforschungszentrum \(DKFZ\)](#)