

Anfälligkeit von anonymen Gen-Datenbanken für Datenschutzverletzungen

Eine neue Studie zeigt, dass anonyme genetische Datenbanken anfällig sind für Identitätsdiebstahl und Datenschutzverletzungen. Forschende warnen vor den Folgen.



Eine Studie hat Bedenken ausgelöst, dass eine Art von genetischer Datenbank, die bei Forschern zunehmend beliebt ist, ausgenutzt werden könnte, um die Identität der Teilnehmer offenzulegen oder private Gesundheitsinformationen mit ihren öffentlichen genetischen Profilen zu verknüpfen.

Einzelzell-Datensätze können Informationen zur Genexpression in Millionen von Zellen enthalten, die von Tausenden von Personen gesammelt wurden. Diese Daten sind oft frei zugänglich und bieten eine wertvolle Ressource für Forscher, die die Auswirkungen von Krankheiten auf zellulärer Ebene untersuchen. Die Daten sollen anonymisiert sein, doch

eine am 2. Oktober in der Zeitschrift Cell veröffentlichte Studie¹ zeigt, wie genetische Daten aus einer Studie „ausgenutzt werden können, um private Informationen über Individuen in einer anderen Studie aufzudecken“, schreiben die Autoren.

Die Ergebnisse heben die Schwierigkeit hervor, die Interessen der Forscher mit der Privatsphäre der Spender in Einklang zu bringen. „Unsere Genome sind sehr identifizierend. Sie können viel über uns, unsere Merkmale und unsere Krankheitsanfälligkeiten aussagen“, sagt die Mitautorin der Studie, Gamze Gürsoy, Bioinformatik-Forscherin an der Columbia University in New York City. „Man kann seine Kreditkartennummer ändern, wenn sie in die Öffentlichkeit gerät, aber man kann sein Genom nicht ändern.“

Sensible Daten

Bedenken bezüglich der Privatsphäre in genetischen Datensätzen wurden bereits geäußert, konzentrierten sich jedoch hauptsächlich auf „Massendaten“ genetischer Profile. Diese enthalten Informationen zur Genaktivität, die über eine große Zellpopulation hinweg gemittelt sind und nicht über individuelle Zellen.

Früher dachte man, dass Einzelzell-Datensätze nicht so anfällig für Datenschutzverletzungen wären, wegen des Niveaus an „Rauschen“, oder Variation in der Genexpression, zwischen den verschiedenen Zellen. Doch Gürsoy und ihr Team konnten nachweisen, dass dem nicht so ist.

Das Team untersuchte drei öffentlich verfügbare Einzelzell-Datensätze, die Blutkörperchen von Menschen mit Lupus, einer chronischen Autoimmunerkrankung, enthielten. Die Forscher fanden heraus, dass sie die Daten zur Genexpression nutzen konnten, um die Struktur des Genoms einer Person vorherzusagen, indem sie diese Werte mit Informationen über expression quantitative trait loci (eQTLs) kombinierten. Die Details der eQTLs – Variationen im Chromosom, die mit der

Genexpression korrelieren – sind ebenfalls öffentlich zugänglich in Einzelzell-Datensätzen.

Um die Zuverlässigkeit ihrer Arbeit zu testen, überprüften die Forscher ihre Genome-Vorhersagen anhand einer Genomdatenbank, die den verwendeten Zellen entsprach. Sie konnten die meisten Datensätze mit dem entsprechenden Genom verknüpfen, mit einer Genauigkeitsrate von über 80 %.

Im Gegensatz zu den Daten zur Genexpression und eQTLs können vollständige Genomdatenbanken normalerweise nur von Wissenschaftlern eingesehen werden, um die identifizierenden Informationen der Spender zu schützen. Die Forscher weisen jedoch darauf hin, dass die Genomdaten eines Teilnehmers woanders öffentlich verfügbar sein könnten. Zum Beispiel könnten sie diese auf einer Genealogie-Website hochgeladen haben, auf der Nutzer DNA-Proben einsenden, um mehr über ihre Abstammung zu erfahren. In diesem Fall könnte ein Angreifer eine Person identifizieren, deren Zellen sich in einem Einzelzell-Datensatz befinden, indem er ihr Genom analysiert. Dies könnte persönliche Daten aufdecken, die mit einem sensiblen Merkmal wie einer psychiatrischen Störung in Zusammenhang stehen, da Forschungsteilnehmer oft ausgewählt werden, um die Biologie dieser komplexen Bedingungen zu untersuchen.

Datenschutzverletzungen wie diese könnten reale Konsequenzen haben, wie z.B. Diskriminierung am Arbeitsplatz, sagt Gürsoy. Sie fügt hinzu, dass Leaks sogar Auswirkungen auf zukünftige Generationen haben könnten, da genetische Merkmale an Nachkommen weitergegeben werden können. „Alles, was über uns bekannt wird, wird durch Generationen weitergetragen“, sagt sie.

Bradley Malin, der im Bereich des großangelegten Genom-Datenaustauschs an der Vanderbilt University in Nashville, Tennessee, forscht, beschreibt die Studie als eine „neuartige Erweiterung und Beitrag zur Literatur“. Er fügt hinzu, dass

zukünftige Forschungen untersuchen könnten, ob Genomdaten auch in größeren Datensätzen, die Proben von Tausenden oder Millionen von Menschen enthalten, verknüpft werden könnten.

Wettbewerb Interessen

Wissenschaftler sind sich unsicher, wie man die Datenschutzbedenken am besten angehen kann. „Es gibt den Wunsch, die Privatsphäre des Einzelnen zu schützen, aber auch den Wunsch, die medizinische Forschung kollektiv voranzutreiben, und diese stehen leider im Widerspruch zueinander“, sagt Mark Gerstein, der an der Yale University in New Haven, Connecticut, Medizindatenwissenschaften erforscht. Die einfachste Lösung wäre, den Zugang zu genetischen Daten zu erschweren, aber das würde die Forschung negativ beeinflussen, sagt er. „Wir müssen große Mengen an Informationen teilen und aggregieren“, erklärt er. „Wenn wir alles blockieren und privater machen, behindert das wirklich den gesamten Prozess.“

In ihrer Studie fordern Gürsoy und ihre Kollegen eine größere Transparenz über die Risiken für Teilnehmer, die ihre Genomdaten teilen, und schlagen vor, dass Forscher sicherstellen sollten, dass die Spender der Weitergabe ihrer Daten zustimmen. Ein weiterer möglicher Weg könnte die Verschlüsselung persönlicher Daten sein, wenn sie Teil einer öffentlichen Datenbank sind. Die Autoren erkennen an, dass dies den Prozess des Erstellens und Wartens von Datensätzen komplizieren würde, sind jedoch der Meinung, dass es helfen könnte, die Privatsphäre der Teilnehmer zu schützen.

1. Walker, C. R. et al. Cell
<https://doi.org/10.1016/j.cell.2024.09.012> (2024).

Artikel
Google Scholar

Referenzen herunterladen

Details

Besuchen Sie uns auf: natur.wiki