https://gen-ethisches-netzwerk.de/node/2497

Gen-ethischer Informationsdienst

Das gläserne Genom

Genomforschung trifft Internet

AutorIn
Alexander von Schwerin

Wenn die Genomforschung auf das Internet trifft, ist eine neue Dimension erreicht. Das zeigt eine Studie über mangelnde Datensicherheit von frei zugänglichen Datenbanken im Internet. Ein Einblick in die neuen Gefahren der Informations-Kombinatorik.

Biobanken stehen für die vollmundigen Versprechen der Genomforschung. *Open Source* steht für ein Zeitalter voller neuer Entfaltungsmöglichkeiten, für das schon lange nicht mehr nur Internet-Enthusiasten eintreten. Wenn Biobanken als Open Source im Internet zugänglich gemacht werden, dann kann das auf Kosten derjenigen gehen, deren DNA-Daten dort für die Forschung freigeschaltet werden. Alles ist sicher! - so beruhigen die Administratoren der riesigen genetischen Datenbanken, die über die Jahre im Internet entstanden sind. Eine kürzlich im Wissenschaftsmagazin Science veröffentlichte Studie zeigt das Gegenteil.[1] Die Einfachheit, mit der es WissenschaftlerInnen am *Whitehead Institute for Biomedical Research* in Cambridge, Massachusetts, in Zusammenarbeit mit anderen Forschungsstellen in den USA und Israel gelang, die Herkunft von anonym in Datenbanken hinterlegten DNA-Sequenzen zu recherchieren und die Identität der dazugehörigen Personen zu lüften, ist alarmierend.

Geniale Genealogie

Ausgangspunkt der spitzfindigen Recherche war die Datenbank des 1.000-Genome-Projekts. Erst kürzlich vermeldeten die Initiatoren, dass das erste große Etappenziel des Projekts erreicht sei. Über 1.000 menschliche Genome seien komplett sequenziert und die entsprechenden Daten frei zugänglich ins Internet gestellt worden. Es ist ein verständliches Anliegen der Spender und Spenderinnen, dass die Genome anonymisiert veröffentlicht werden und dass niemand ohne weiteres Rückschlüsse von den Genomdaten auf ihre Person ziehen kann. Ein Team um den Humangenetiker und Bioinformatiker Yaniv Erlich, das Analysemethoden für DNA-Sequenzen entwickelt, wurde neugierig, ob man mit speziellen Analysemethoden nicht doch die Anonymität der Daten knacken könnte. Ausgangspunkt der Überlegungen war das in den USA boomende Hobby, die Familiengenealogie zu erforschen. Genealogische Websites haben Konjunktur. Privatleute hinterlegen dort die Informationen zur eigenen Person und ihrer Familie, um auf diese Weise Licht in die Verwandtschaftsverbindungen ihrer Familie zu bringen. In einem Land, das durch Einwanderung und Sklaverei entstanden ist und bis heute von Immigration lebt, ist das Bedürfnis, irgendwann die verschütteten eigenen Wurzeln zu erforschen, nicht erstaunlich. Die private Verwandtschaftsforschung bewegt sich auf dem neuesten Stand der technischen Entwicklung. Zum Einsatz kommt etwa auch der Vergleich von Genom-Daten. Zunutze machen sich dabei die Anbieter der Genealogie-Websites eine auch in der Forensik bekannte Methode: die Zuordnung von DNA-Spuren durch ihre Identifikation anhand von kurzen, sich wiederholenden DNA-Sequenzen (*short tandem repeats*). Bei der polizeilichen Ermittlung

werden diese Mini-Sequenzen inzwischen als Beweismittel benutzt, das als sicherer als eine Überführung per Fingerabdruck gilt. Anders als bei der polizeilichen Personenidentifikation geht es den Hobby-Genealogen darum, Personen zu finden, deren DNA-Mini-Sequenzen ihren eigenen ähnlich sind und die deshalb mit ihnen verwandt sein könnten. Was haben also die WissenschaftlerInnen am Whitehead-Institut gemacht? Sie haben die familien-spezifischen DNA-Mini-Sequenzen von zufällig ausgewählten Genomen in der Datenbank des 1.000-Genome-Projekts analysiert. Mit dieser Information ermittelten sie auf genealogischen Websites den Namen von Familien, bei denen diese Sequenzen vererbt werden. Auf diese Weise war der Kreis der in Frage kommenden Personen bereits sehr eingeschränkt. Durch die Kombination mit ebenfalls auf der 1.000-Genome-Datenbank hinterlegten Information über Alter und regionale Herkunft der Probanden sowie andere Details gelang es den Whitehead-WissenschaftlerInnen schließlich mehr oder weniger mühelos, die Probanden zu identifizieren.

Lapidar und naiv

Die Veröffentlichung des Coups hat für Aufsehen gesorgt. Vor der Veröffentlichung informierten die Whitehead-WissenschaftlerInnen Eric D. Green, den Direktor des *National Human Genome Research Institute* der *National Institutes of Health* (NIH). Green gab sich überrascht, wie einfach die Identität von Probanden im Internet gelüftet werden kann. Die NIH entfernte darauf hin die Altersangaben in ihren öffentlich zugänglichen Datenbanken, um einen Missbrauch zumindest zu erschweren. Darüber hinaus erklärte Green, dass die NIH derzeit keine Lösung für die neu aufgeworfenen Fragen zum Datenschutz habe. So neu waren die Fragen jedoch nicht. Immerhin war den Verantwortlichen des 1.000-Genome-Projekt bewusst, dass es nicht ratsam war, die Anonymität der gespeicherten DNA-Daten gegenüber den Probanden auch rechtlich zu garantieren. Zwar hat man bei jeder Gelegenheit beteuert, die Anonymität sei gesichert und es sei kaum denkbar, dass jemand sich an die Daten heranmacht. Das von den Probanden unterzeichnete Teilnahmeformular beinhaltete dagegen die Klausel, dass die ForscherInnen die Anonymität der Daten nicht garantieren können. Die Reaktion der Universität Utah, die sich insbesondere an der Rekrutierung von Probanden für das 1.000-Genome-Projekt beteiligt hat, war noch lapidarer. Der für Forschungsethik zuständige Vizepräsident Jeffrey R. Botkin meinte, dass bislang kein Datenmissbrauchsfall bekannt geworden sei, obwohl DNA-Daten von Hunderttausenden bereits in der einen oder anderen Form frei im Netz zugänglich seien. Er könne sich auch nicht denken, warum es interessant sein könnte, an diese Informationen zu kommen. Hinter der zur Schau getragenen Naivität steckt natürlich Kalkül. Klar ist, dass die Forschungscommunity vor allem eins befürchtet: Restriktionen. Forderungen, DNA-Daten besser zu schützen und unter Verschluss zu halten, treffen auf wenig Gegenliebe. Selbst die Forderung, Probanden müssten ausdrücklich über die Datenschutz-Risiken aufgeklärt werden, wird abgelehnt. Man fürchtet offensichtlich, dass die inzwischen wie geschmiert laufende Maschinerie der Genomforschung ins Stottern kommt. Denn ihre empfindliche Stelle ist das Vertrauen der PatientInnen und ProbandInnen. Dabei machen sich viele keine Illusionen mehr. BioethikerInnen gehen bereits in die Offensive: Privatheit um jeden Preis schützen und genetische Daten effektiv anonymisieren zu wollen, sei keine realistische Option.

Bloß eine Lücke im System?

Die neue Studie der Whitehead-WissenschaftlerInnen ist nicht die erste ihrer Art. Der Biometriker Eric Schadt vom *Mount Sinai Hospital* in New York City hat bereits plausibel argumentiert, dass man eine Person identifizieren kann, wenn ihre DNA in zwei Datenbanken vorhanden ist. In der neuen Studie handelt es sich bei den identifizierbaren Personen zwar ausschließlich um Männer, da die Genealogie-Websites nur die Informationen von Y-Chromosomen benutzen; sie beweist aber, dass Individuen selbst dann identifiziert werden können, wenn lediglich ein entfernter Verwandter sein Erbgut öffentlich zugänglich macht. „Wir zeigen, dass Sie identifiziert werden können, wenn beispielsweise Ihr Onkel Dave seine DNA an eine Datenbank für Familienforschung geschickt hat“, sagt eine Mitautorin der Studie.[2] Tatsächlich könne man sogar identifiziert werden, wenn nicht Onkel Dave, sondern Patrick, der unbekannte Cousin vierten Grades, seine DNA online gestellt habe, solange er über die väterliche Linie verwandt sei. Auch wenn die Situation in den USA eine spezielle ist, weil in Kontinentaleuropa der Genealogie-Hype noch nicht angekommen ist - das

Exempel, das die Whitehead-WissenschaftlerInnen statuiert haben, macht deutlich, welcher Art die zukünftigen Probleme sein werden. Nicht die einzelne Biobank ist das Problem. Die Probleme liegen in der Vielfalt: Die Datensicherheit mag für jede einzelne Datenbank formal korrekt geklärt sein. In der Kombination verschiedener Datenquellen entsteht aber eine neue datenschutzrechtlich prekäre Situation. Das Gen-ethische Netzwerk hat in seiner Kampagne gegen die Ausweitung und Vernetzung von DNA-Datenbanken nationaler Polizeibehörden auf dieses Problem aufmerksam gemacht. Denn wie will man den Austausch und die Kombination der verschiedenen Datenquellen kontrollieren? „Getreu dem Creative-Commons-Gedanken“, wie man bitter anmerken kann, werden all diese Daten kostenfrei im Internet zur Verfügung gestellt.3 Das Interesse der biomedizinischen Forschung ist allerdings viel konkreter. In der heutigen Genomforschung und Biomedizin ist kein einzelnes Team in der Lage, die Vielzahl mit modernen Mitteln erhobener Daten auf jede erdenkliche Weise zu analysieren. Nur wenn der Datenberg ohne Zugangsbeschränkungen bereitsteht, lässt er sich effizient aufbereiten. Erst im Dezember letzten Jahres hat die englische Gesundheitsverwaltung ein neues Megaprojekt in Sachen Datenerfassung angekündigt: Die Genomdaten von 100.000 PatientInnen sollen gespeichert werden. Das klingt nach unwahrscheinlich viel; es handelt sich aber nur um das Pilot-Projekt für die genom-gestützte Gesundheitsversorgung der Zukunft (siehe Kurz notiert, Seite 32). Wie man hört, soll der Zugang zu diesen Daten reglementiert werden. Die Missbrauchsgefahren durch Informations-Kombinatorik können auf diese Weise aber nicht gebannt werden. Die britische NGO Gene Watch befürchtet, dass der Missbrauch bei einer solchen bevölkerungsweiten Datenbank, die sich einfach mit der bei der Polizei gebräuchlichen Methode des DNA-Fingerprintings durchsuchen ließe, vorprogrammiert ist.4 Auch unabhängig von Open Source bleibt also die Frage, wer Zugang zu all den neuen Datenquellen hat oder sich verschaffen kann.

- 1Melissa Gymrek et al.: Identifying Personal Genomes by Surname Inference, in: Science 339 (2013), S. 321-324, DOI: 10.1126/science.1229566.
- 2Die Welt online, 17.01.13: Anonyme Erbgut-Spender sind identifizierbar.
- 3Matthias Gräbner: Neue Datenschutzprobleme in der Genomik 18.01.13, www.heise.de/tp/artikel/38/38384/1.html.
- 4Helen Wallace: A DNA Database in the NHS: The End of Privacy?, 12.12.12, www.publicservice.co.uk.

Informationen zur Veröffentlichung

Erschienen in:
GID Ausgabe 216 vom Februar 2013
Seite 38 - 39