

Anonymisierung genetischer Daten in der Forschung Geht das überhaupt? Und was gilt rechtlich?

DuD 2025, Potsdam

Dr. Nuria Mackes, Neurowissenschaftlerin
Michael Weiß, LL.M. (Fordham), Legal Counsel

Agenda

1. Rechtliche Grundlagen der Anonymisierung genetischer Daten
2. Anonymisierung in der Praxis: Herausforderungen & Lösungen
3. Q & A

Roche auf einen Blick

Wer wir sind und was wir tun



128 Jahre

Gründet 1896 in Basel



Führend in F&E im Gesundheitswesen

mit Investitionen von CHF 13 Milliarden in 2024



3 Nobelpreise und **44 Prix Galien**,
seit 1974



CHF 60,5 Milliarden*

Umsatz der Roche Gruppe im Jahr 2024



44 Roche Medikamente & 90 Diagnostika**

auf WHO-Liste für unentbehrliche Medikamente & Diagnostika



>100.000

engagierte Mitarbeiter:innen weltweit



>26 Millionen Menschen wurden 2024 mit unseren Medikamenten behandelt



30 Milliarden Tests

wurden 2024 mit unseren Diagnostik-Produkten durchgeführt

*Soweit nicht anders angegeben, sind alle Wachstumsraten und Vergleiche zum Vorjahr zu konstanten Wechselkursen (CER; Durchschnittskurse 2022) berechnet und alle angegebenen Gesamtbeträge in CHF ausgewiesen.

** Medikamente und Tests, die entweder von Roche entwickelt oder erworben wurden.

Roche in Deutschland

18.260 Mitarbeitende arbeiten für Patient:innen auf der ganzen Welt¹



Standort	Mitarbeitende ²
Mannheim - Diagnostics, Pharma Entwicklung, Produktion, Logistik, Vertrieb und Marketing	8.710
Penzberg - Diagnostics, Pharma Forschung & Entwicklung, Produktion	7.710
Grenzach-Wyhlen - Pharma Vertrieb & Marketing, Qualitätssicherung	1.250
Ludwigsburg - Diagnostics Entwicklung & Produktion von Laborautomatisierungssystemen	480
Berlin & Potsdam - Diagnostics, Pharma Start-up-Förderung im Gesundheitswesen, Hauptstadtpräsenz - politischer Dialog	110
Gesamt	18.260

¹ Zahlen auf dieser Folie gerundet

² Headcount inkl. Auszubildende und Praktikant:innen, Stand 31.12.2024

1. Rechtliche Grundlagen der Anonymisierung genetischer Daten

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Was sind genetische Daten?

Art. 4 Nr. 13 DSGVO:

“[G]enetische Daten [sind] personenbezogene Daten zu den ererbten oder erworbenen genetischen Eigenschaften einer natürlichen Person, die eindeutige Informationen über die Physiologie oder die Gesundheit dieser natürlichen Person liefern und insbesondere aus der Analyse einer biologischen Probe der betreffenden natürlichen Person gewonnen wurden.”

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Spannungsfeld zwischen Forschung und Datenschutz



Das macht die Daten so wertvoll

- Medikamentenentwicklung
- Personalisierte Medizin
- Krebsbekämpfung
- Pharmakogenomik
- Untersuchung seltener Erkrankungen
- ...



Datenschutzrechtliche Herausforderungen

- Einzigartigkeit & Identifizierbarkeit
- Unveränderlichkeit & lebenslange Gültigkeit
- Verwandtschaftsbezug
- Sensitivität
- ...

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Risiko genetischer Daten am Beispiel 23andMe



Home News Sport Business Innovation Culture Arts Travel Earth Audio Video Live

23andMe: Profiles of 6.9 million people hacked

5 December 2023

Share Save

LILLY RAY NEWMAN SECURITY OCT 6, 2023 5:53 PM

23andMe User Data Stolen in Targeted Attack on Ashkenazi Jews

At least a million data points from 23andMe accounts appear to have been exposed on BreachForums. While the scale of the campaign is unknown, 23andMe says it's working to verify the data.

Genetic testing firm 23andMe admits hackers accessed DNA data of 7m users

US company says 'threat actor' responsible for security breach that affected nearly half of its 14m reported users

- Von 2015 bis 2024: 15 Polizeianfragen in den USA
- Data Breach in 2023: Daten und genetische Informationen von 7 Mio Kunden gestohlen und zum Verkauf angeboten
- Insolvenz 2025 und Verkauf an Regeneron Pharmaceuticals

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Anonymisierung vs. Pseudonymisierung



Rechtliche Grundlagen

Anonymisierung

- Personenbezogene Daten werden unumkehrbar so verändert, dass eine Identifizierung der betroffenen Person nicht mehr möglich ist.
- Daten fallen nicht mehr in den Anwendungsbereich der DSGVO.

Pseudonymisierung

- Identifizierung ist durch Hinzuziehung zusätzlicher Informationen möglich.
- Weiterhin personenbezogene Daten



Komplexität der Anonymisierung

- Hohe Dimensionalität
- Verfügbarkeit externer Referenzdatenbanken (z.B. 23andMe)
- Risiko der Re-Identifizierung durch Verknüpfung
- Vollständige Irreversibilität bei genetischen Daten kaum zu erreichen

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Die relative Anonymisierung als Lösung?

Relative Anonymisierung

- Keine absolute Anonymisierung, jedoch sind die Daten mangels Möglichkeit der Re-Identifizierung durch den jeweiligen Verantwortlichen für diesen anonymisiert.
- EuG hat diese Ansicht in der Rechtssache T-557/20 bestätigt, ebenso der Generalanwalt in seiner Stellungnahme in der anschließenden Revision (EuGH-Entscheidung steht noch aus).
- Herausforderung bei genetischen Daten:
 - Aufgrund der **extrem hohen Einzigartigkeit** genetischer Informationen ist es schwer, einen Zustand zu erreichen, in dem eine Re-identifizierung nicht möglich ist.
 - Selbst wenn der ursprüngliche Verantwortliche alle ihm bekannten direkten Bezüge kappt, können die **genetischen Daten selbst noch so einzigartig** sein, dass sie **quasi-identifizierend** wirken.

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Ansicht der Aufsichtsbehörden

Art. 29 Working Party¹:

Genetic data profiles are an example of personal data that can be at **risk of identification** if the sole technique used is the removal of the identity of the donor due to the **unique nature of certain profiles**. It has already been shown [...] that the combination of publicly available genetic resources [...] and the metadata about DNA donors [...] can **reveal the identity** of certain individuals [...].

EDPB²:

[...] the possibility to anonymise genetic data remains an **unresolved issue**. As yet, it remains open to be demonstrated whether any combination of technical and organisational means can be effectively employed to remove genetic information from the material scope of the GDPR. [...] It is **strongly advised that such genetic data is treated as personal data** [...].

DSK³:

In diesem Zusammenhang muss zudem berücksichtigt werden, dass eine wirksame Anonymisierung genetischer Daten in der Regel daran **scheitert**, dass [...] eine **Identifizierung möglich ist**.

¹ [WP216: Opinion 05/2014 on Anonymisation Techniques](#)

² [EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research \(2021\)](#)

³ [DSK Beschluss 15. Mai 2024: Positionspapier "Anforderungen an die Sekundärnutzung von genetischen Daten zu Forschungszwecken"](#)

Rechtliche Grundlagen der Anonymisierung genetischer Daten

Weitere Ansichten

GMDS¹:

Aktuellere Untersuchungen [2006-2013] haben ergeben, dass genetische Daten **nicht als anonym angesehen werden** können. Genetische Daten können daher maximal pseudonymisiert werden bzw. Bei entsprechender Anpassung bestenfalls als pseudonyme Daten angesehen werden.

Medizin Informatik Initiative²:

Wenn man Daten über das Erbgut einer Person hat, können die Daten grundsätzlich dieser Person eindeutig zugeordnet werden. Die Daten bilden eine Art genetischen Fingerabdruck. Es ist also grundsätzlich möglich, eine Person anhand ihrer genetischen Daten zu identifizieren. **Es hängt aber von Art und Umfang der genetischen Daten ab**, ob die Daten wirklich einen einzigartigen Fingerabdruck bilden.

¹ [Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e. V.: Praxishilfe zur Anonymisierung/Pseudonymisierung \(2024\)](#)

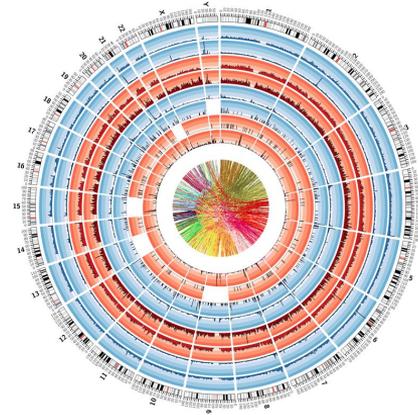
² [Häufige Fragen: Umgang mit genetischen Daten \(17.1\)](#)

2. Anonymisierung in der Praxis: Herausforderungen & Lösungen

Genetische Daten im Überblick



Genetische Daten sind vielfältig - das Risiko auch



Weniger komplex /
weniger identifizierend

Hochkomplex /
hoch identifizierend

Fokus: Analysemethoden

Spezifische Gen-Panels	SNP-Arrays (Genotypisierung)	Gesamtexomsequenzierung (WES)	Gesamtgenomsequenzierung (WGS)
<ul style="list-style-type: none"> ■ Sequenzierung einer ausgewählten Gruppe von Genen ■ Gezielte Diagnostik, Therapieentscheidungen 	<ul style="list-style-type: none"> ■ Analyse einer kleinen Auswahl von Einzelnukleotid-Polymorphismen (SNPs) an spezifischen Stellen im Genom ■ Abstammungsanalyse, Risikoprädiktion 	<ul style="list-style-type: none"> ■ Sequenzierung der kodierenden Regionen des Genoms (Exons, 1-2% des Genoms) ■ Kosteneffiziente Alternative zu WGS 	<ul style="list-style-type: none"> ■ Sequenzierung des nahezu gesamten Genoms einer Person ■ Forschung, Personalisierte Medizin

Geringer Informationsgehalt /
weniger identifizierend

Hoher Informationsgehalt/
hoch identifizierend

Fokus: Somatisch vs Keimbahn

Somatisch

- Entstehen postnatal in spezifischen Zellen oder Geweben
- Nur in betroffenem Gewebe (z.B. Tumor)
- Nicht vererbbar, kein direkter Zusammenhang mit Familie, mit der Zeit veränderlich



Geringe Stabilität /
weniger identifizierend

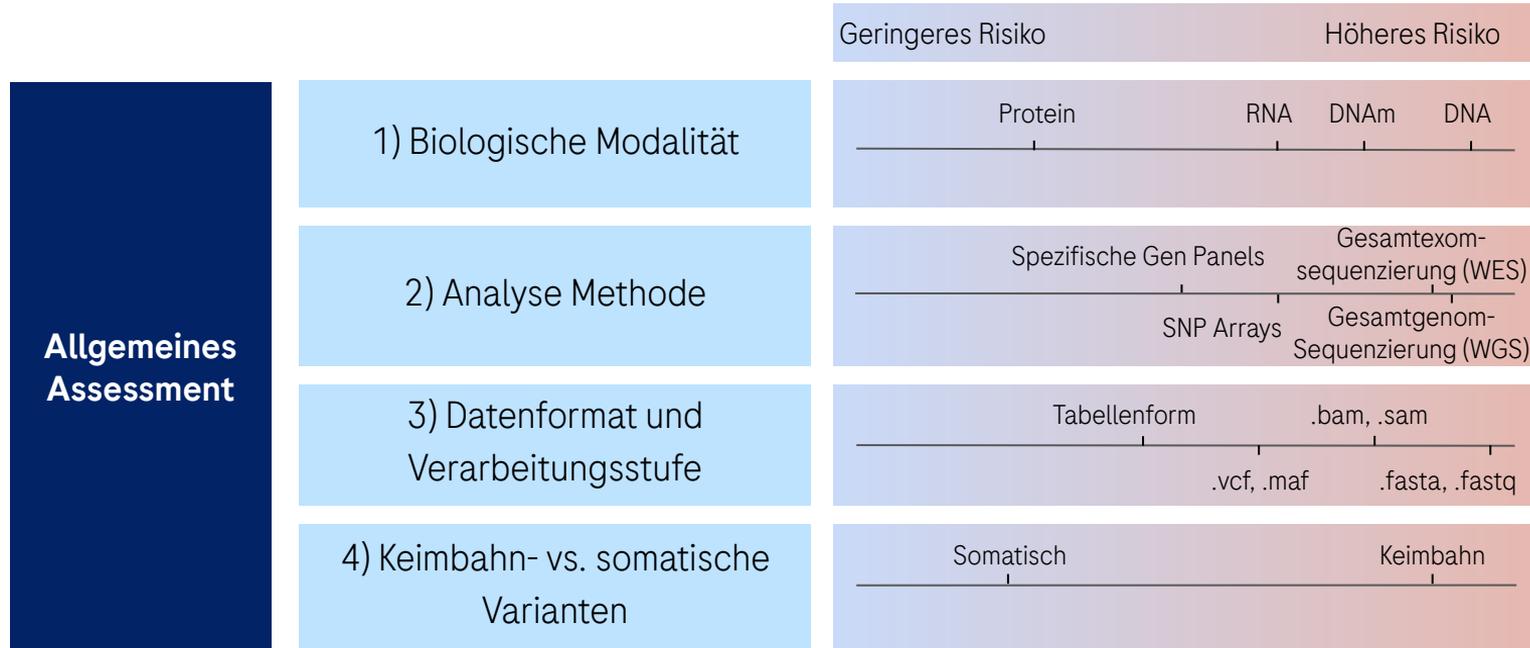
Keimbahn

- Entstehen in den Keimzellen der Eltern oder sehr früh in der Embryonalentwicklung
- In fast allen Körperzellen des Individuums vorhanden
- Informationen haben lebenslange Gültigkeit und betreffen die gesamte Familie.



Hohe Stabilität /
hoch identifizierend

9 Merkmale die das Risiko der Re-Identifizierung beeinflussen

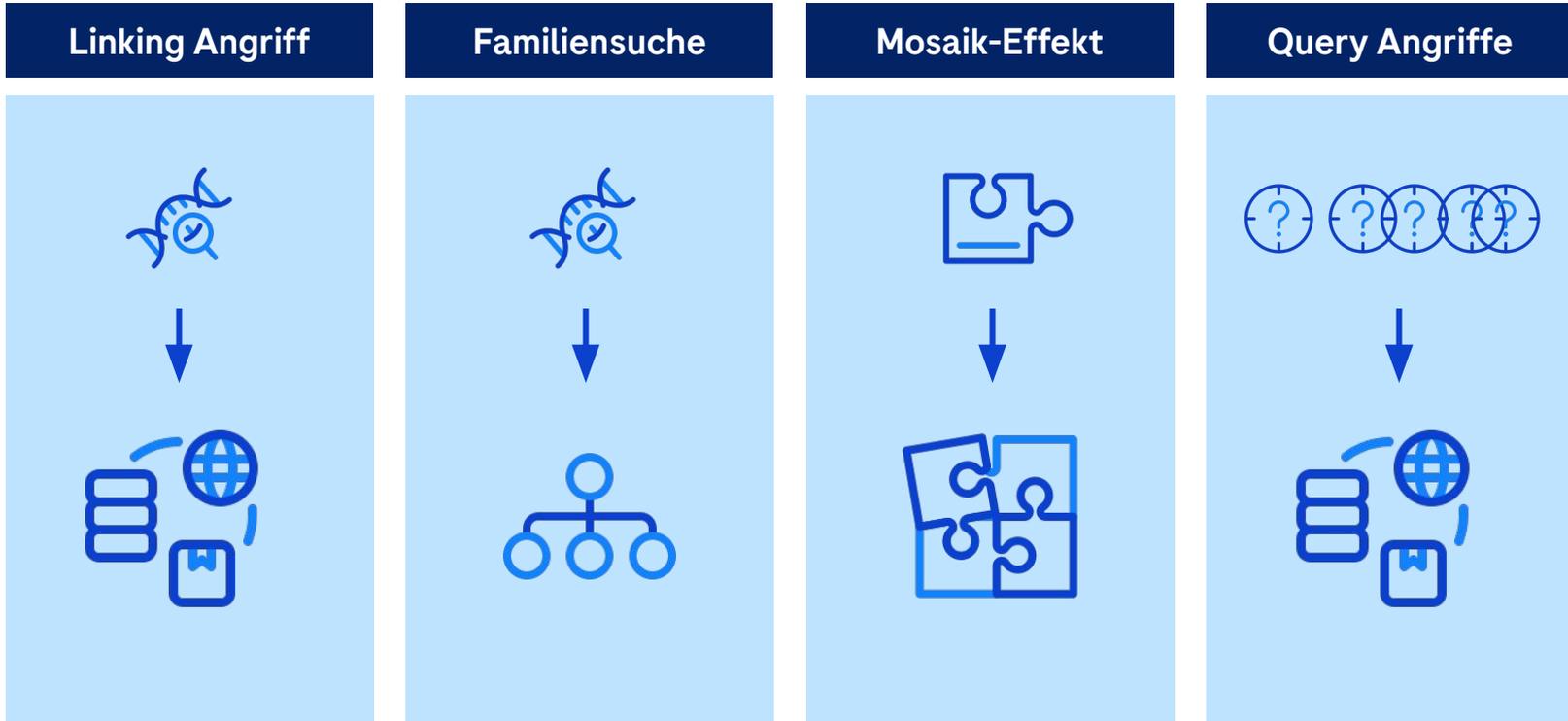


Thomas, M., Mackes, N., Preuss-Dodhy, A., Wieland, T., & Bundschuh, M. (2024). Assessing Privacy Vulnerabilities in Genetic Data Sets: Scoping Review. *JMIR bioinformatics and biotechnology*, 5, e54332. <https://doi.org/10.2196/54332>

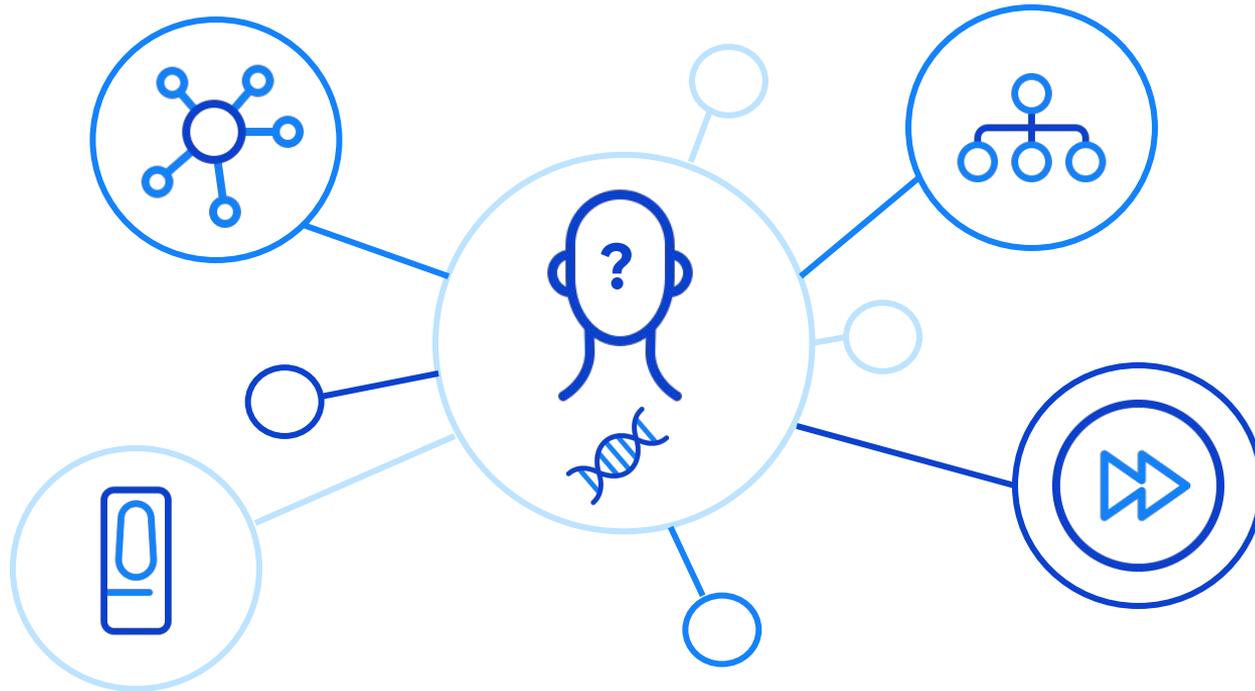
9 Merkmale die das Risiko der Re-Identifizierung beeinflussen

		Geringeres Risiko	Höheres Risiko
Komponenten mit hohem Risiko	5) Einzelnukleotid-Polymorphismen (SNPs)	1-10	30-100 >500 Assoziiert mit sensitiven Attributen / Auf häufig genutzten Arrays
	6) Mikrosatelliten / kurze Tandemwiederholungen (STRs)	1-5	5-10 10-15 >15
	7) Aggregierte Messwerte	p-Values	Allelfrequenzen Odds ratios
Komponenten mit geringem Risiko	8) Strukturelle Varianten	Keine Daten	
	9) Seltene Varianten	Keine Daten	

Wie funktioniert Re-identifizierung?



Die technischen Hürden der Anonymisierung im Detail



Anonymisierung genetischer Daten: Eine realistische Option?



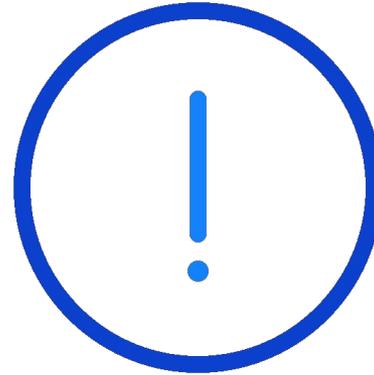
Für reichhaltige genetische Datensätze ist eine echte, zukunftssichere Anonymisierung technisch extrem schwer bis praktisch unmöglich.

Wenn nicht Anonymisierung, was dann? Der pragmatische Ansatz



3. Q & A

Q & A



Doing now what patients need next