

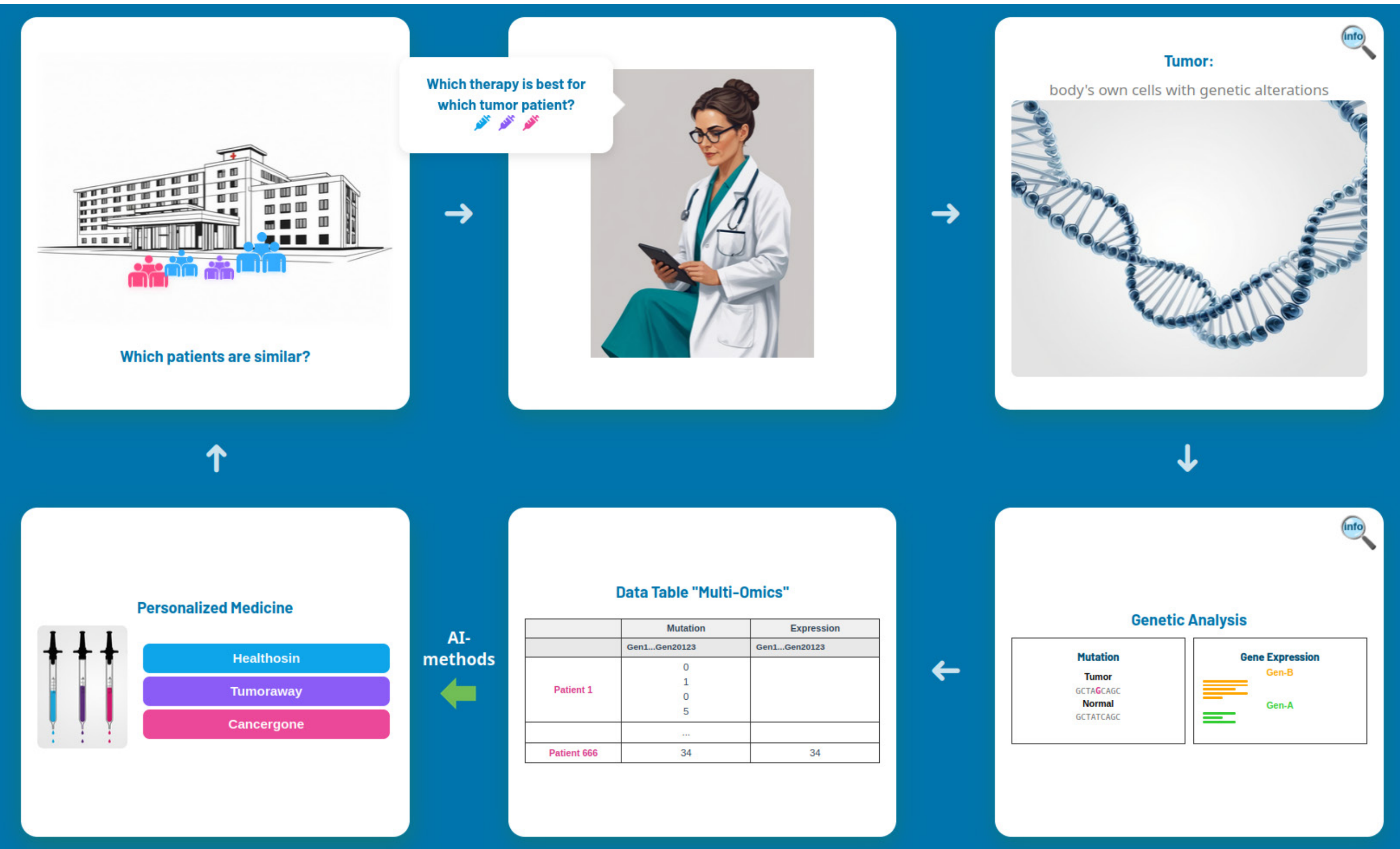
DIMENSIONSREDUKTION FÜR PERSONALISIERTE MEDIZIN

ALICIA DIEM

FUNKTION

Dieser Demonstrator zeigt verschiedene Methoden der Dimensionsreduktion – ein wichtiges Werkzeug in der modernen Medizin und KI. Anhand von synthetischen Datensätzen können Nutzer auf spielerische Weise lernen, wie drei dieser zentralen Methoden der Dimensionsreduktion in der personalisierten Medizin eingesetzt werden.

Der Demonstrator richtet sich an ein breites Publikum: Auch ohne Vorkenntnisse können Nutzer die Methoden intuitiv erkunden. Für Nutzer mit einem Hintergrund in Mathematik oder maschinellem Lernen bietet er zudem die Möglichkeit, das Verständnis zu vertiefen und die verschiedenen Ansätze miteinander zu vergleichen.



Bildunterschrift hinzufügen Beschreibung des Bildes

ANWENDUNG

Zu Beginn kann der Nutzer seine Sprache (Deutsch oder Englisch) auswählen. Ein dynamisches Highscore-System gibt einen ersten Eindruck von den Gamification-Elementen. Die Anwendung ist ein interaktives Erlebnis, das auf einem Szenario aus der Krebsforschung basiert. Da Krebs die wohl komplexeste und individuellste Erkrankung ist, helfen computergestützte Methoden dabei, die riesigen Datenmengen, die für jeden Patienten anfallen, zu verstehen. Sowohl Nutzer als auch Forscher zielen darauf ab, Patienten mit ähnlicher Tumorbiologie in Gruppen einzuteilen. Dadurch lässt sich die beste Behandlung für jede Untergruppe zuordnen, ein Schlüssel zur personalisierten Medizin.

Im „Gaming-Menü“ können die Nutzer die drei Dimensionsreduktionsmethoden – PCA, UMAP und Autoencoder – auf synthetische Datensätze anwenden. Sie wählen zunächst einen Datensatz und dann eine Methode. Interaktive Elemente führen sie durch die wichtigsten Schritte und zeigen, wie jede Methode die Daten reduziert.

Am Ende wird das Krebszenario noch einmal in einen breiteren Kontext gestellt: Eine präzisere und stabilere Repräsentationsmethode verbessert die Zuweisung der richtigen Behandlung. Das erhöht die Überlebensraten, reduziert unnötige Behandlungen und senkt die Kosten.

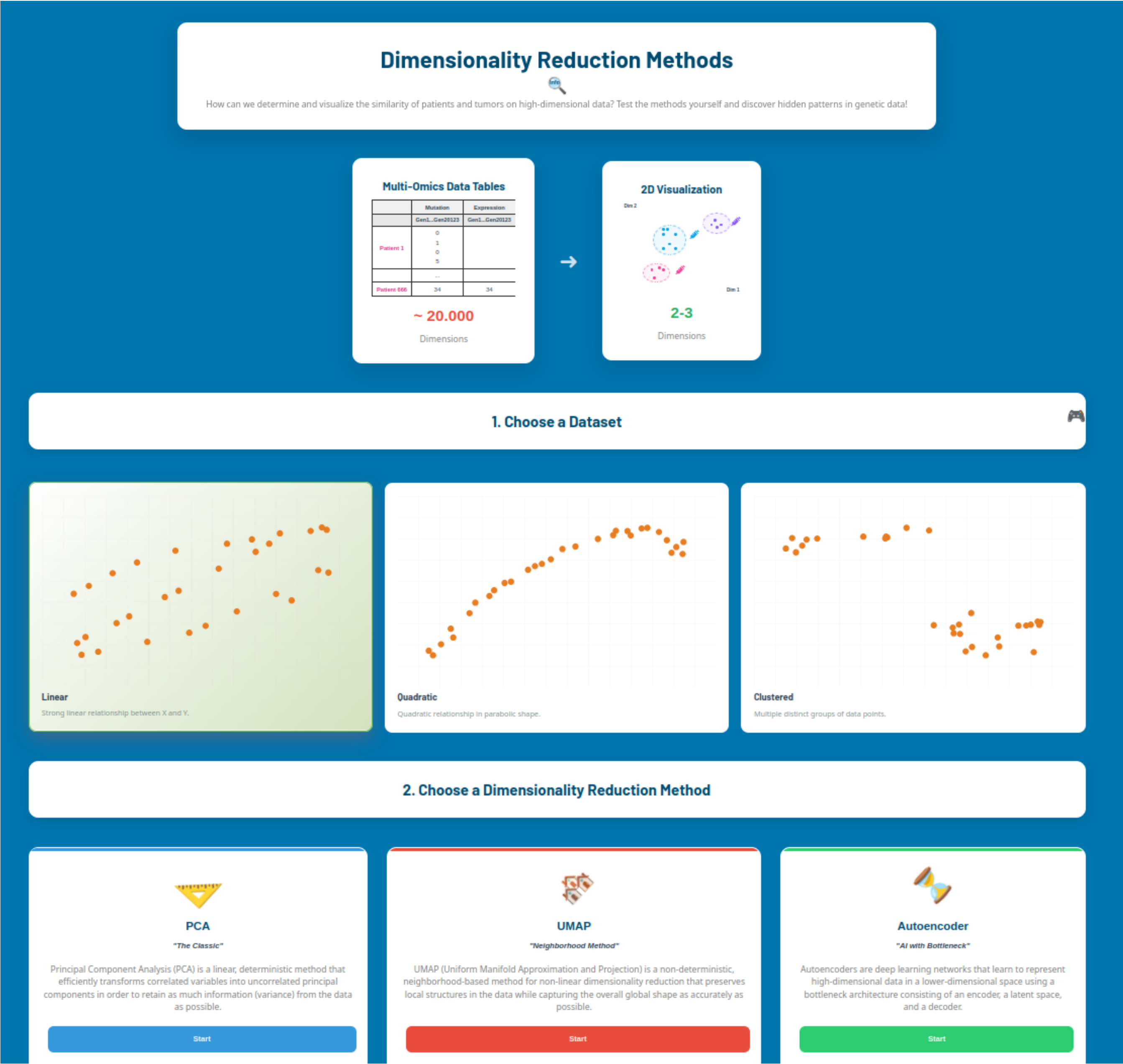
LITERATURANGABEN

AUTOENCODIX <https://github.com/jan-forest/autoencodix>
MIRACLE <https://scads.ai/research/applied-ai-and-big-data/life-science-and-medicine/projects/miracle/>
[1] Joas, M. J., Jurenaite, N., Prašćević, D., Scherf, N., & Ewald, J. (2026). AUTOENCODIX: a generalized and versatile framework to train and evaluate autoencoders for biological representation learning and beyond. Nature Computational Science, 6(1), 96-108.
[2] Selby, D. A., Sprang, M., Ewald, J., & Vollmer, S. J. (2025). Beyond the black box with biologically informed neural networks. Nature Reviews Genetics, 26(6), 371-372.

METHODEN UND TECHNOLOGIE

Moderne Technologien in der Medizin (wie Sequenzierungs- und molekulare Profiling-Technologien) erzeugen große Datensätze, sogenannte „Omics-Daten“, bei denen Tausende genetischer Merkmale pro Patient oder Zelle erfasst werden können. Diese hochdimensionalen Daten bieten tiefgreifende Einblicke in Krankheiten wie Krebs, sind aber auch komplex und schwer zu analysieren. Um aussagekräftige Erkenntnisse aus diesen Daten zu gewinnen benötigt man computergestützte Methoden, insbesondere KI-basierte Ansätze. Dimensionsreduktion hilft dabei, diese hochdimensionalen Daten in einer übersichtlicheren Form ohne nennenswerten Informationsverlust darzustellen. In diesem Demonstrator werden drei solcher Methoden mit steigender Komplexität vorgestellt, die sich spielerisch erkunden lassen:

1. PCA (Principal Component Analysis): Ein linearer Ansatz, der Daten auf Richtungen mit höchster Varianz projiziert. Effektiv bei einfachen, linearen Zusammenhängen, aber begrenzt bei komplexen biologischen Daten.
2. UMAP (Uniform Manifold Approximation and Projection): Eine nichtlineare Methode, die komplexe Strukturen gut erhält. Allerdings kann sie bei wiederholten Durchläufen unterschiedliche Ergebnisse liefern, da sie stochastisch arbeitet.
3. Autoencoder: Ein neuronales Netzwerk mit einer engen Zwischenarchitektur, das hochdimensionale Daten in kompakte, deterministische Darstellungen umwandelt. Im Gegensatz zu PCA und UMAP erzeugt es immer dieselben Ergebnisse und ist daher besonders geeignet für moderne Anwendungen in der Biomedizin. Bei ScaDS.AI haben wir das Open-Source-Framework AUTOENCODIX entwickelt, um Autoencoder für biologische und andere Daten zugänglich zu machen.



Bildunterschrift hinzufügen Beschreibung des Bildes



Zentrum für skalierbare
Datenanalyse und
künstliche Intelligenz

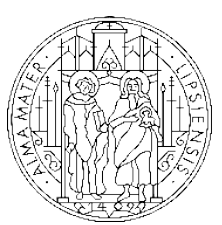
KONTAKT
Alicia Diem
ad69bevi@studserv.uni-leipzig.de

GEFÖRDERT VOM



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

PARTNER



UNIVERSITÄT
LEIPZIG



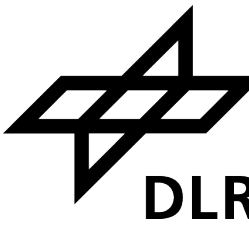
TECHNISCHE
UNIVERSITÄT
DRESDEN



MAX-PLANCK-INSTITUT
FÜR KOGNITIONS- UND NEUROWISSENSCHAFTEN



HELMHOLTZ
Zentrum für Umweltforschung



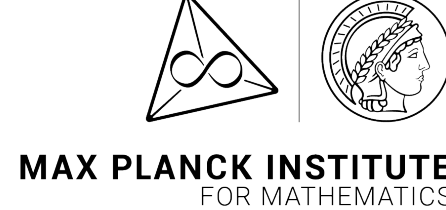
InFAI
Institut für Angewandte Informatik



Fraunhofer
IAIS



Fraunhofer
IZI



Leibniz-Institut
für ökologische
Raumentwicklung