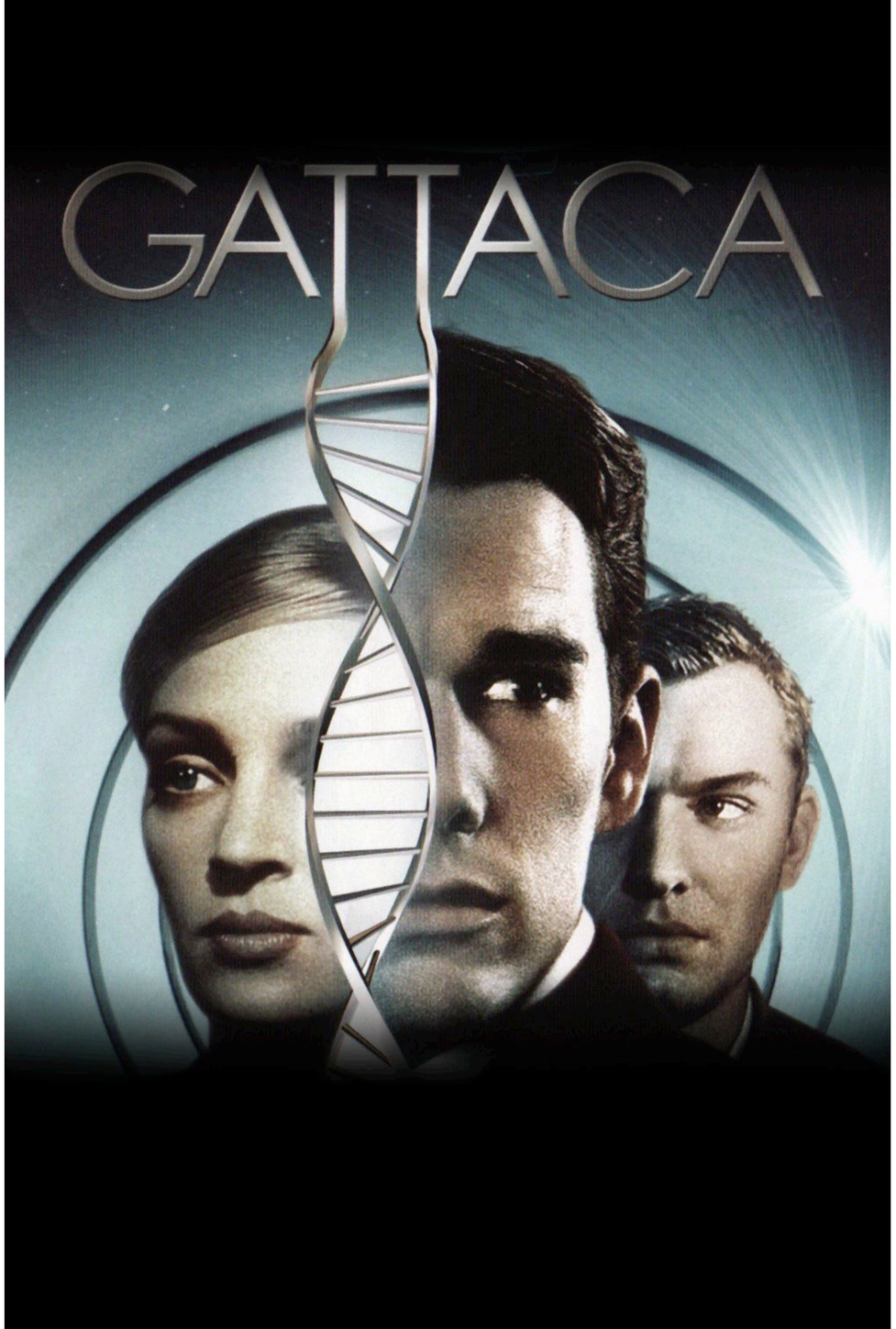
# Genomes Privacy Laws Society




## Technical, legal and ethics aspects of genomic data sharing

Michael Baudis | DSI Ethics | 2022-01-12

# Genome screening at the core of “Personalised Health”

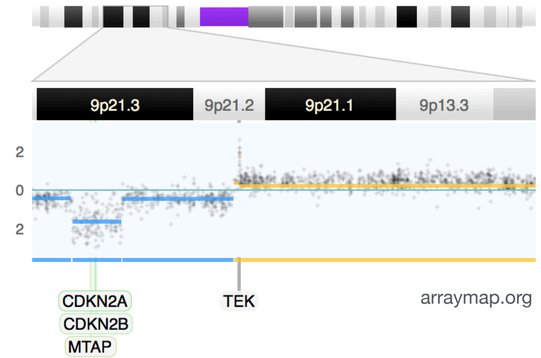
- **Genome analyses** (including transcriptome, metagenomics) are core technologies for Personalised Health™ applications
- The unexpectedly large amount of **sequence variants** in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of **reference repositories**
- **Standardized data formats** and **exchange protocols** are needed to connect these resources throughout the world, for reciprocal, international **data sharing** and **biocuration** efforts
- Our work @ UZH:
  - ***cancer*** genome repositories
  - biocuration
  - protocols & formats

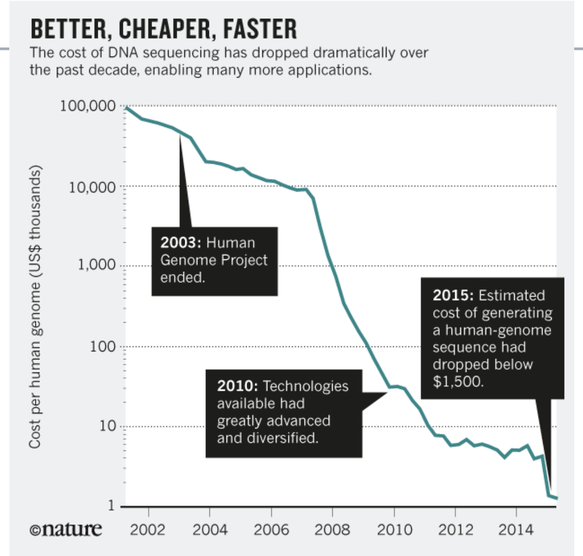
arrayMap

progenetix

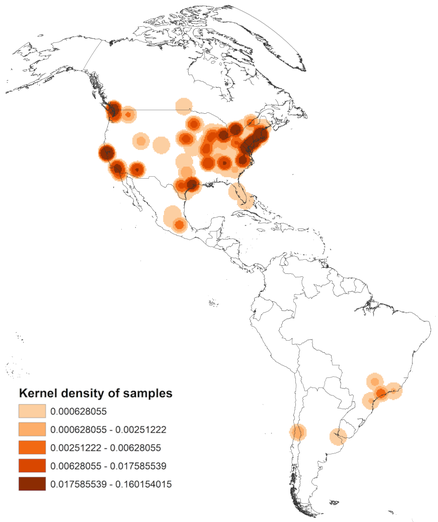
elixir

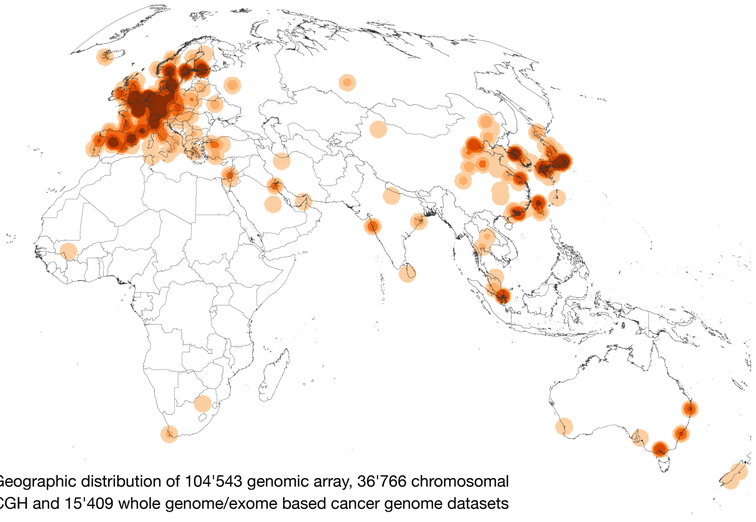
Global Alliance for Genomics & Health




The future of DNA sequencing. Eric D. Green, Edward M. Rubin & Maynard V. Olson. Nature; 11 October 2017 (News & Views)

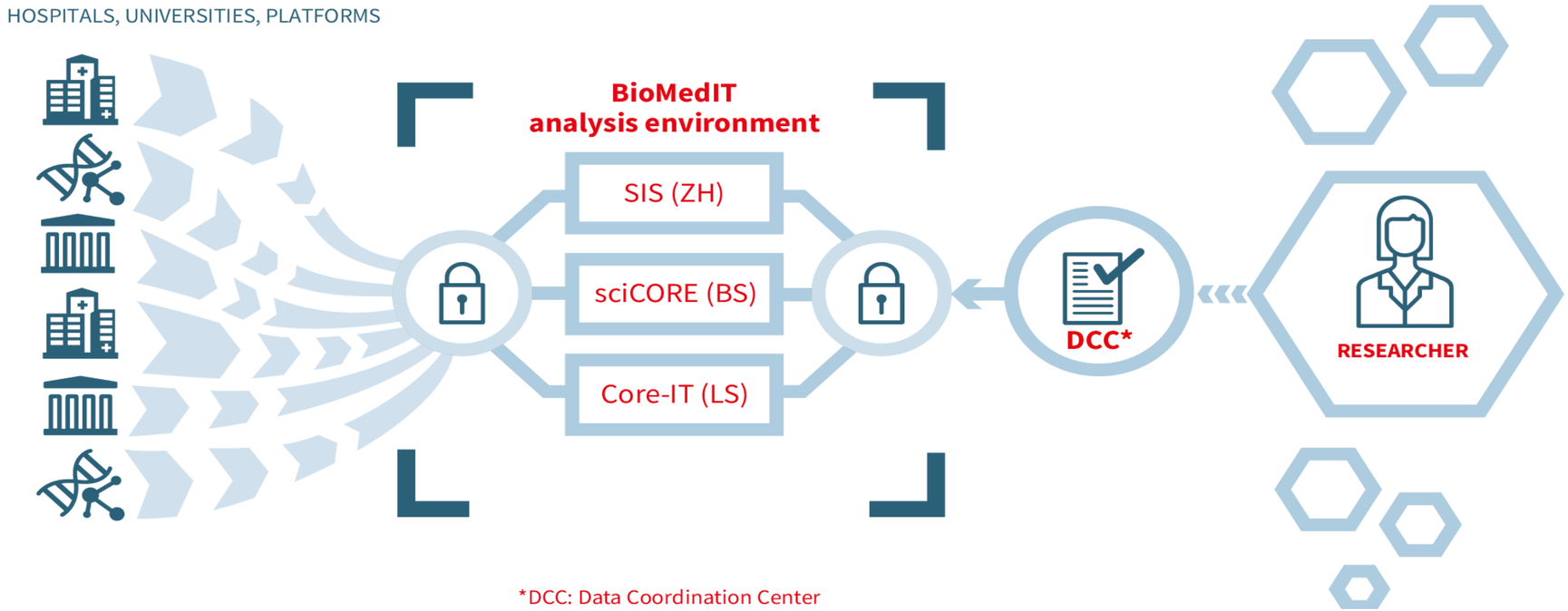


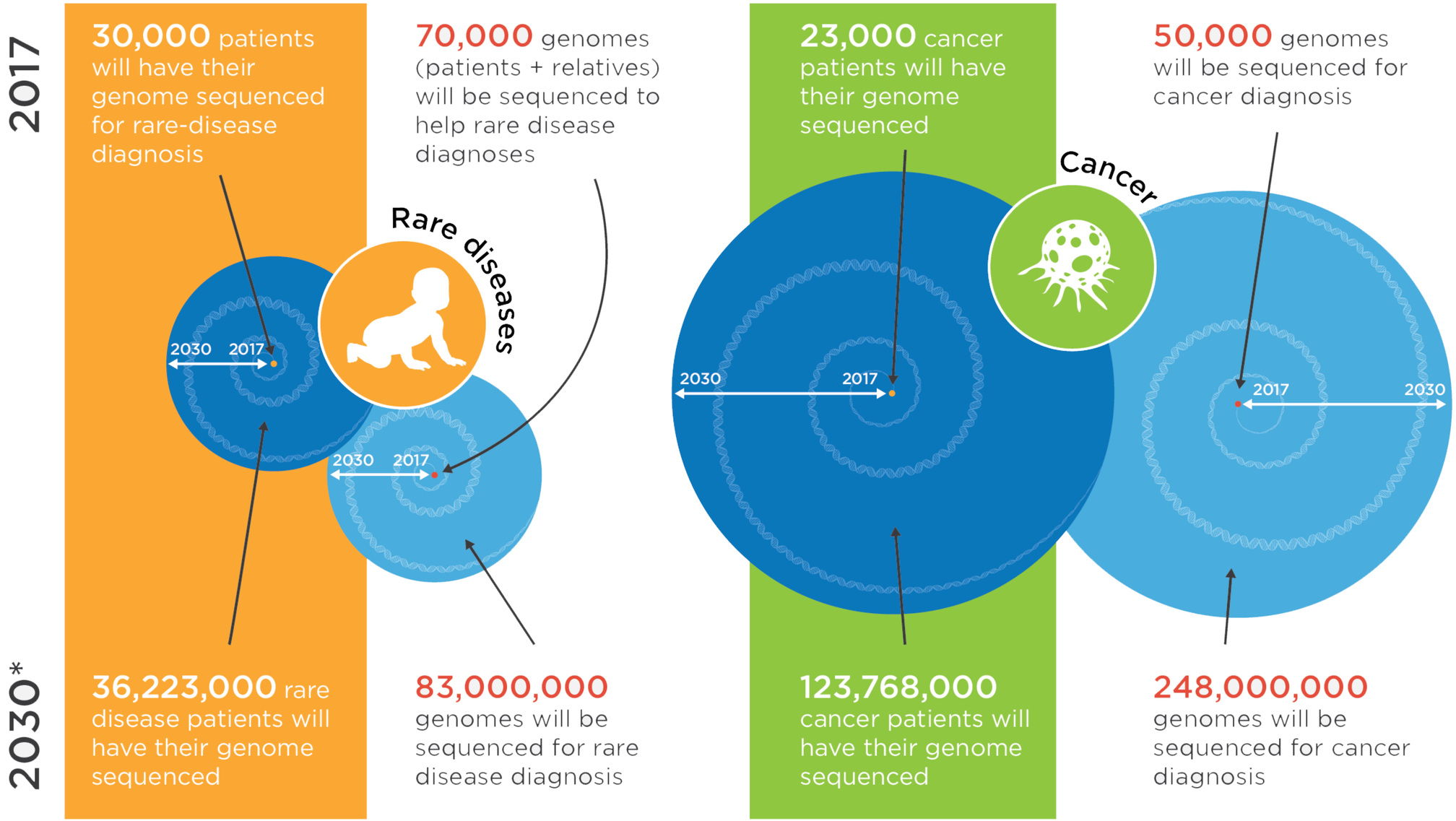
Geographic distribution of 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets

# The BioMedIT network

BioMedIT provides researchers with access to a secure and protected computing environment for analysis of sensitive data without compromising data privacy




*DCC: Data Coordination Center




* Projected figures, based on current data and known status of genomics initiatives worldwide.

# Limited Population Diversity in Cancer Studies

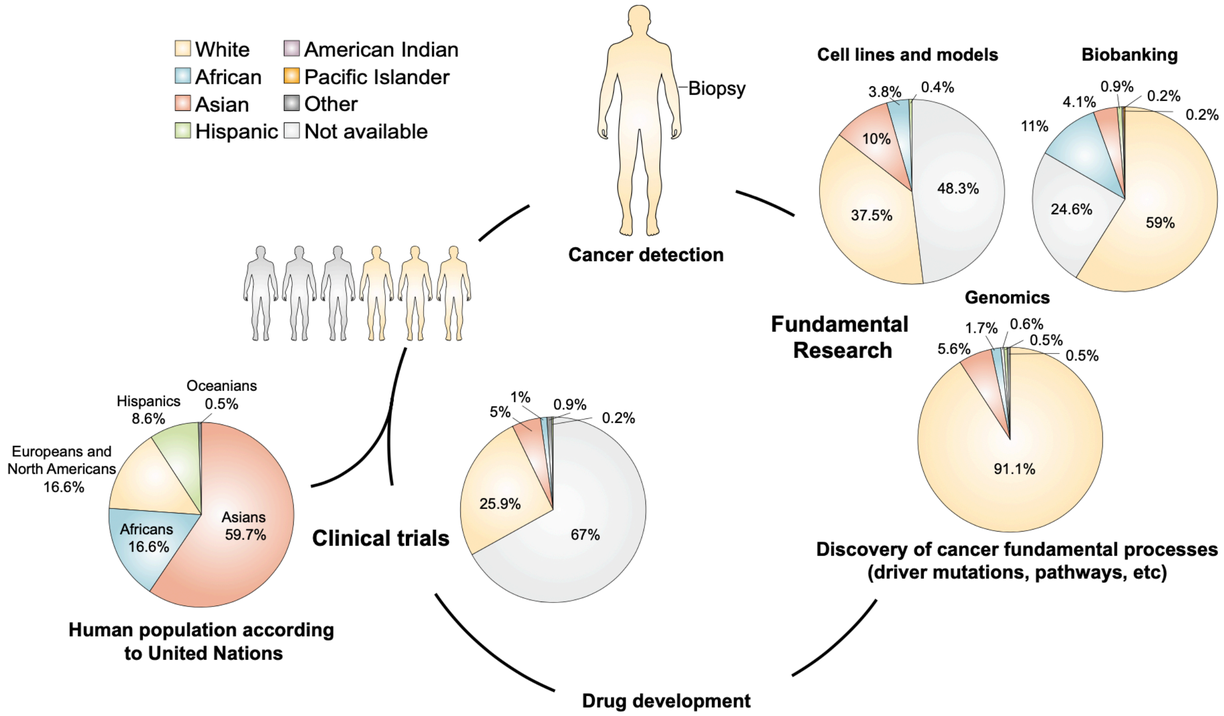


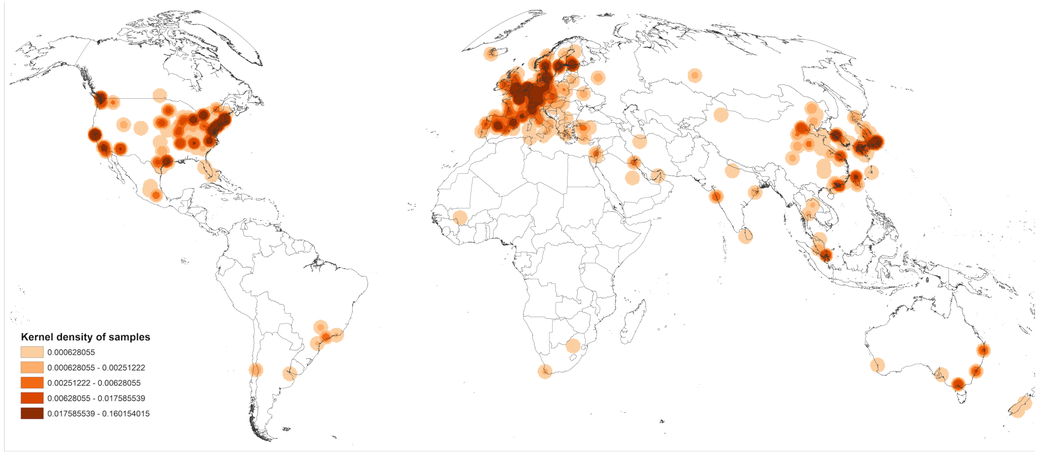

**Figure 1.** Racial/Ethnic disparities in cancer research. Racial/ethnic inclusion was studied in several aspects of oncological research, from cell lines and patient-derived xenografts to biobanking, genomics and clinical trials.

Guerrero S, López-Cortés A, Indacochea A, et al. Analysis of Racial/Ethnic Representation in Select Basic and Applied Cancer Research Studies. *Sci Rep*. 2018;8(1):13978.

## Publication Landscape of Cancer CNV Profiling

Publication statistics for cancer genome screening studies. The graphic shows our as- sessment of publications reporting whole-genome screening of cancer samples, using molecular detection methods (chromosomal CGH, genomic array technologies, whole exome and genome sequencing).

For the years 1993-2018, we found 3'229 publications reporting 174'530 individual samples in single series from 1 to more than 1000 samples. Y-axis and size of the dots correspond to the sample number; the color codes indicate the technology used.




progenetix

# The vision: Federation of data

# Enabling genomic data sharing for the benefit of human health

The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework

Genomic Data Toolkit →

Regulatory & Ethics Toolkit →

Data Security Toolkit →

VIEW OUR LEADERSHIP

MORE ABOUT US

BECOME A MEMBER

# The Global Alliance for Genomics and Health

## Making genomic data accessible for research and health

- January 2013 - 50 participants from eight countries
- June 2013 - White Paper, over next year signed by 70 “founding” member institutions (e.g. SIB, UZH)
- March 2014 - Working group meeting in Hinxton & 1st plenary in London
- October 2014 - Plenary meeting, San Diego; interaction with ASHG meeting
- June 2015 - 3rd Plenary meeting, Leiden
- September 2015 - GA4GH at ASHG, Baltimore
- October 2015 - DWG / New York Genome Centre
- April 2016 - Global Workshop @ ICHG 2016, Kyoto
- October 2016 - 4th Plenary Meeting, Vancouver
- May 2017 - Strategy retreat, Hinxton
- October 2017 - 5th plenary, Orlando
- May 2018 - Vancouver
- October 2018 - 6th plenary, Basel
- May 2019 - GA4GH Connect, Hinxton
- October 2019 - 7th Plenary, Boston
- October 2020 - Virtual Plenary, June 2021 - Virtual Connect ...
- October 2021 - Virtual Plenary ...


GENOMICS

*A federated ecosystem for sharing genomic, clinical data*

Silos of genome data collection are being transformed into seamlessly connected, independent systems

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

Global Alliance for Genomics & Health

# GA4GH API promotes sharing


GENOMICS

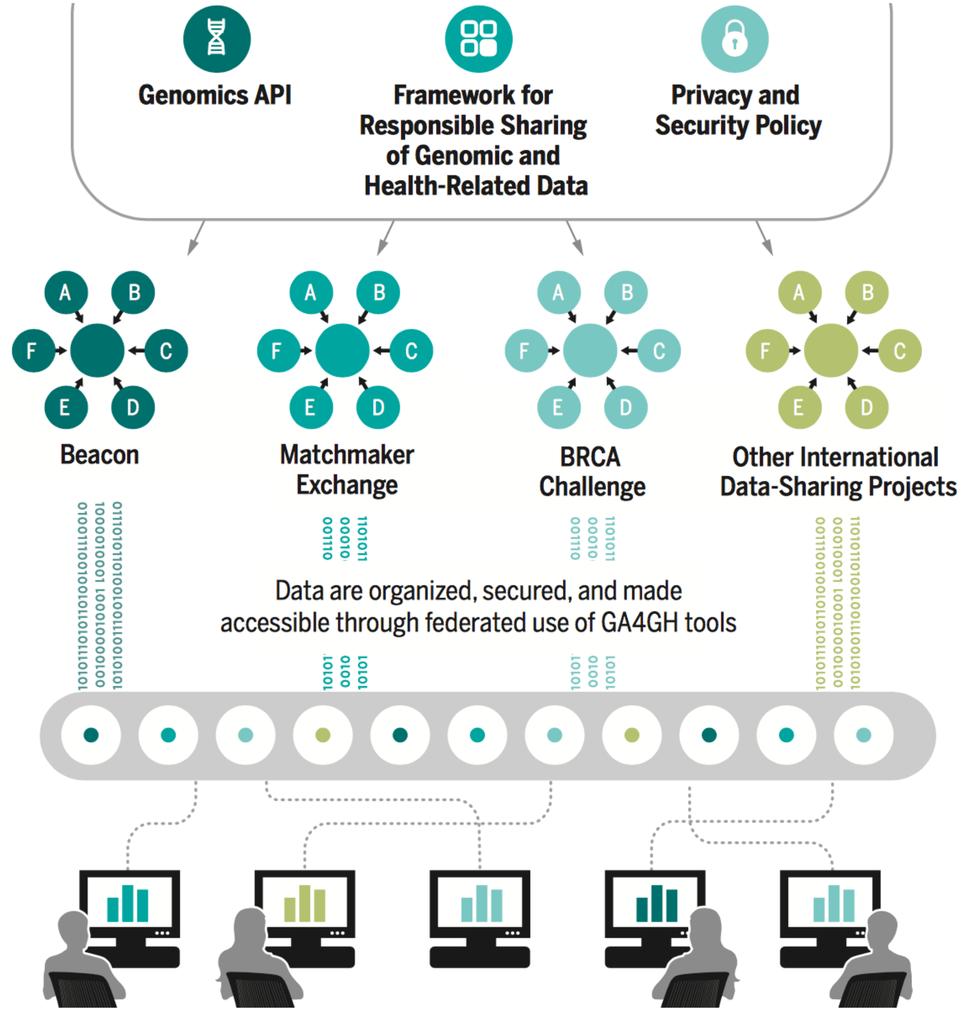
## *A federated ecosystem for sharing genomic, clinical data*

Silos of genome data collection are being transformed into seamlessly connected, independent systems

**The Global Alliance for Genomics and Health***

**SCIENCE** 10 JUNE 2016 • VOL 352 ISSUE 6291

**A federated data ecosystem.** To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.




Global Alliance for Genomics & Health


GENOMICS

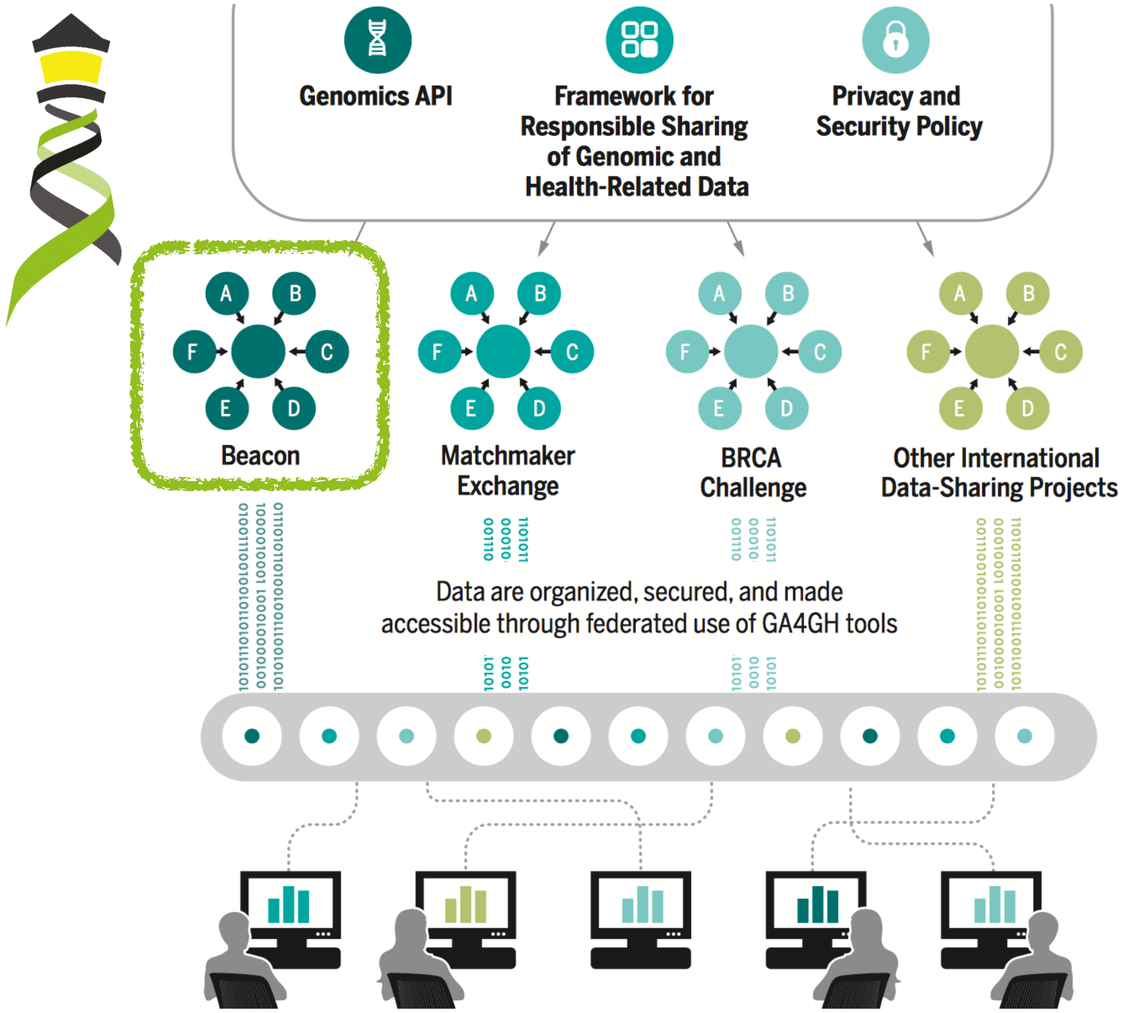
# *A federated ecosystem for sharing genomic, clinical data*

Silos of genome data collection are being transformed into seamlessly connected, independent systems

**The Global Alliance for Genomics and Health***

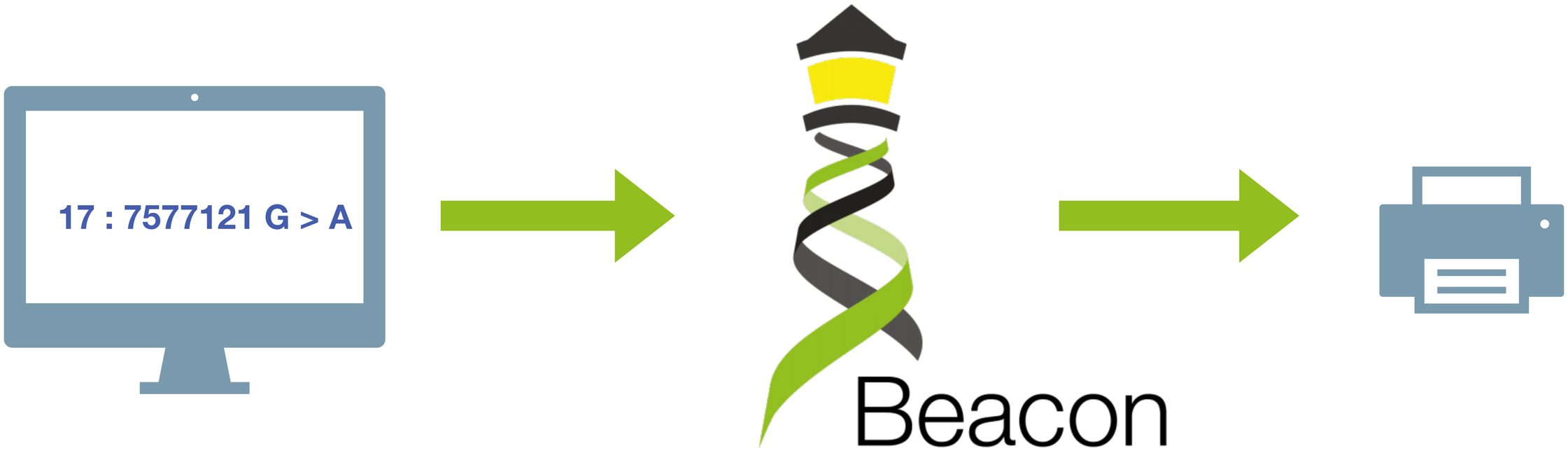
**SCIENCE** 10 JUNE 2016 • VOL 352 ISSUE 6291

**A federated data ecosystem.** To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.




DNASTACK

Global Alliance for Genomics & Health




A ***Beacon*** answers a query for a specific genome variant against individual or aggregate genome collections

**YES** | **NO** | **\0**


Have you seen this variant? It came up in my patient and we don't know if this is a common SNP or worth following up.

A Beacon network federates *genome variant queries* across databases that support the ***Beacon API***

Here: The variant has been found in **few** resources, and those are from **disease** specific **collections**.

# Making Beacons Biomedical - Beacon v2

- Scoping queries through "biodata" parameters
- Extending the queries towards clinically ubiquitous variant formats
  - cytogenetic annotations, named variants, variant effects
- Beacon queries as entry for **data delivery**
  - Beacon v2 permissive to respond with variety of data types
    - Phenopackets, biosample data, cohort information ...
  - handover to stream and download using htsget, VCF, EHRs
- Interacting with EHR standards
  - FHIR translations for queries and handover ...
- Beacons as part of local, secure environments
- Authentication to enable non-aggregate, patient derived datasets
  - ELIXIR AAI with compatibility to other providers (OAuth...)


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

# Beacon *v2* API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "**genomics API**".

# Making Beacons Biomedical - Beacon v2

- Scoping queries through "biodata" parameters
- Extending the queries towards clinically ubiquitous variant formats
  - cytogenetic annotations, named variants, variant effects
- Beacon queries as entry for **data delivery**
  - Beacon v2 permissive to respond with variety of data types
    - Phenopackets, biosample data, cohort information ...
  - handover to stream and download using htsget, VCF, EHRs
- Interacting with EHR standards
  - FHIR translations for queries and handover ...
- Beacons as part of local, secure environments
- Authentication to enable non-aggregate, patient derived datasets
  - ELIXIR AAI with compatibility to other providers (OAuth...)

**Definitely breaks the "Relative Security by Design" Concept!**

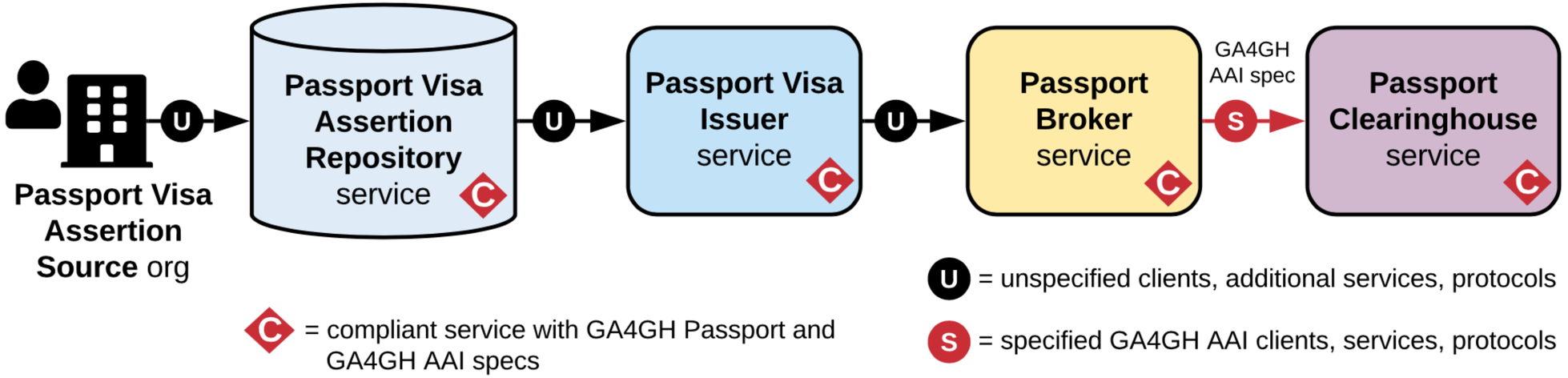
**Mitigation by tailored implementation and security practices**

Global Alliance for Genomics & Health

# GA4GH Passports

Global Alliance for Genomics & Health

## Communicating a user's data access authorizations



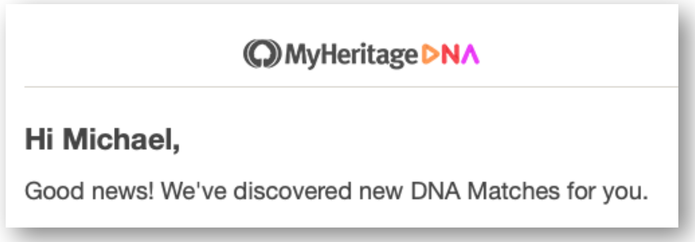

www.ga4gh.org/ga4gh-passports/

- format to communicate a user's data access authorizations based on either their role (e.g. researcher), affiliation, or access status
- works together with the GA4GH Authentication and Authorization Infrastructure (AAI) OpenID Connect Profile to streamline researchers' data access over federated data access protocols
- both standards approved in Dec 2019 with early implementation by Google Cloud services and ELIXIR

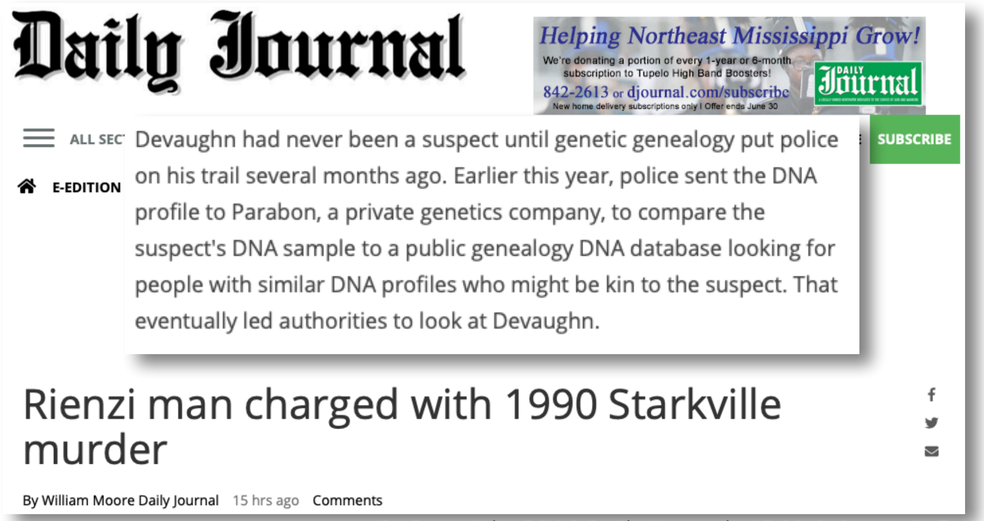
Google Cloud

elixir

# Long-Range Familial Searches




- Commercial, "Direct to Customer" DNA analyses are provided through independend sites and such affiliated to genealogy services (MyHeritage, Ancestry.com, 23andMe...)
- Genealogy sites identify individuals with matching haplotype blocks & provide a prediction about degree of genetic relation
- Law enforcement agencies (and who else?!) can send individual SNP profiles (e.g. recovered from evidence many years after a crime) using a *Jane Doe* identity, to identify relatives of the suspect - **long range familial search**



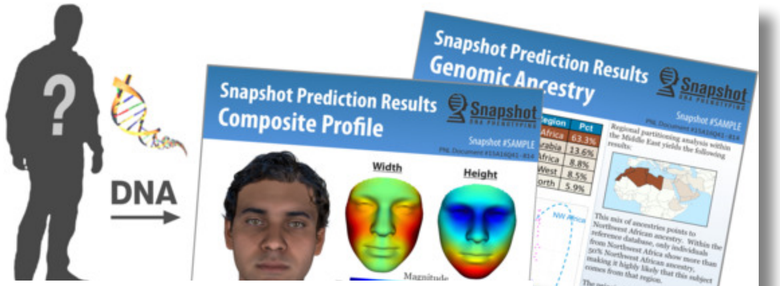


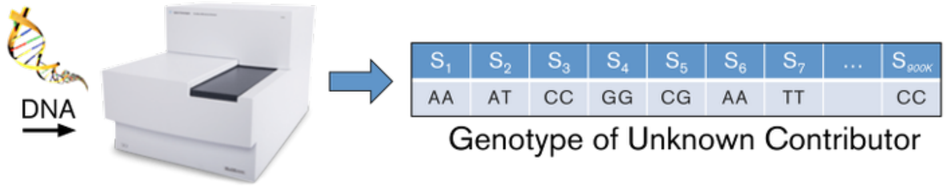

# Phenotyping from DNA

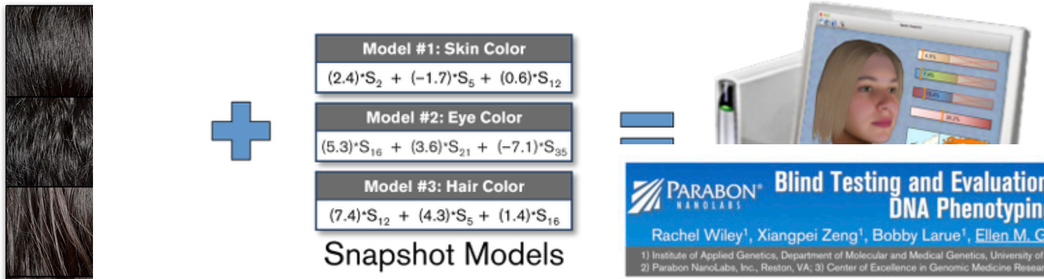
## From DNA to "Wanted" Posters?

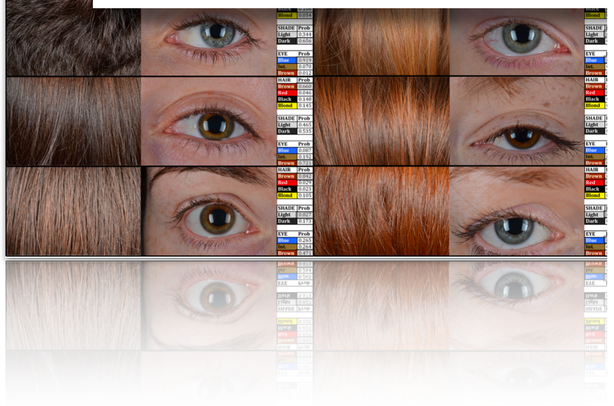
- association of genomic variants with phenotypic data collection
- while hair, eye color are easy targets not useful for relevant phenotypic features especially if large environmental component
- huge biases based on input/collection data
- Belgium and Germany do not allow forensic DNA phenotyping
- Switzerland: Bundesrat decision on 2020-12-04 to allow phenotyping for law enforcement purposes

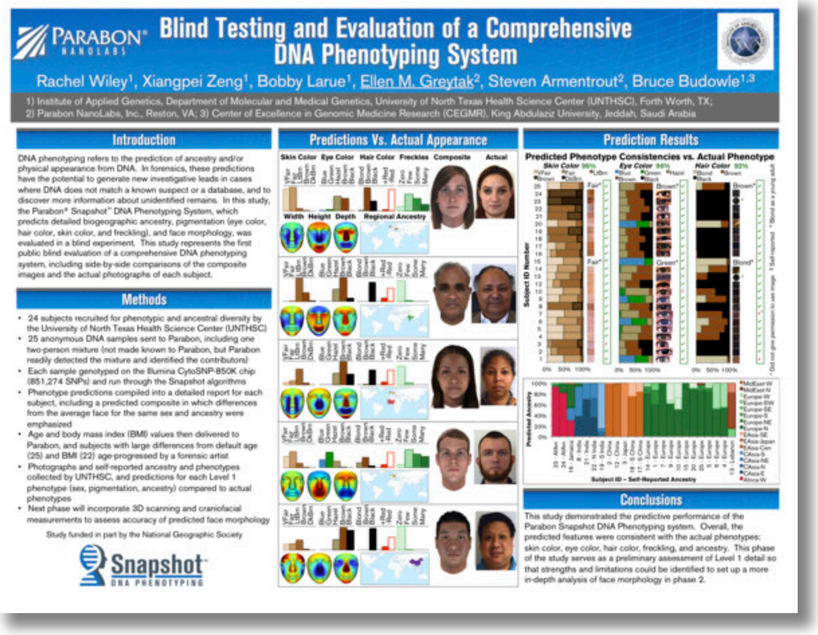
**Paragon Nanolabs Inc.**
**The Snapshot DNA Phenotyping Service**






"When the New York Times ran an informal test of the Parabon system with one of its reporters, it failed badly." (ACLU.org)

https://snapshot.parabon-nanolabs.com/phenotyping

# Federal Act on the Use of DNA Profiles in Criminal Proceedings and for Identifying Unidentified or Missing Person, DNA Profiles Act

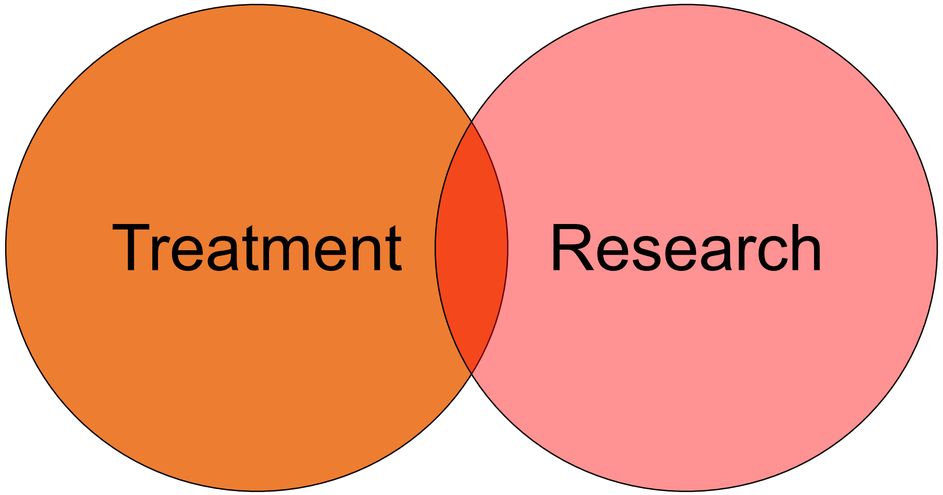
## An Area in Transition...

- Currently: «Genetic Fingerprint»
- Future: Will it be allowed to take a deeper look and how far can genetic data be used to determine the characteristics of an unknown perpetrator (colour of hair and eyes, height, ethnicity, etc.)
- Switzerland: Bundesrat decision on 2020-12-04 to allow phenotyping for law enforcement purposes

Content provided by Julian Mausbach | Oberassistent für Strafrecht und Strafprozessrecht

# Law's View on Modern Medicine

**HGTA and Ordonnances**




**HRA and Ordonnances**

- How do we handle the growing overlap area?

➡ unclear; current legislative movement:
HRA will relate more to HGTA in the future

**HGTA : Federal Act on Human Genetic Testing**

**HRA: Human Research Act**

Content provided by Julian Mausbach | Oberassistent für Strafrecht und Strafprozessrecht

# HGTA : Federal Act on Human Genetic Testing

| HGTA new (probably 2021) | medical field | outside the medical field | |
|---|---|---|---|
| Investigated characteristics | medical relevant | especially protective values characteristics | other characteristics |
| General Requirements | Non-discrimination, information and consent, right to information, right not to know, avoidance of surplus information, protection of samples and genetic data, Circulation concerning public advertising, state of science and technology, penal provisions | | |
| Initiation | Physician | Health professional (controlled taking of samples) | Consumer **(DTC)** |
| Persons concerned | Persons with **and** without capacity of judgement, pregnant woman (PND) | ONLY persons with Capacity of judgement | ONLY persons with Capacity of judgement |
| Communication of surplus information | as a rule according to decision of the person concerned | Not allowed | Not allowed |
| Laboratory | subject to authorization (cyto and molecular genetic studies) | subject to authorization (cyto and molecular genetic studies) | not subject to authorisation |
| Employers and Insurance institutions | Studies and Recovery of Results / Data only in regulated exceptional cases | Prohibition to carry out investigations and the Recovery of Results / Data | Prohibition to carry out investigations and the Recovery of Results / Data |

# Is Genomic Data Special?

# Health Related Data & Privacy

## Considerations when evaluating risks of data sharing

- Is the genetic condition outwardly visible?
- How severe is it? (serious disease, penetrance, age of onset)
- Is it associated with what could be considered to be stigmatizing health information (e.g., associated with mental health, reproductive care, disability)?
- Is it familial (i.e., potential carrier status/reproductive implications for family/relatives)?
- Does it provide information about the likely geographical location of individuals?
- Does it provide information about ethnicity that may be considered potentially stigmatizing information?

Sharing health-related data: a privacy test?

Stephanie OM Dyke[1], Edward S Dove[2] and Bartha M Knoppers[1]

Greater sharing of potentially sensitive data raises important ethical, legal and social issues (ELSI), which risk hindering and even preventing useful data sharing if not properly addressed. One such important issue is respecting the privacy-related interests of individuals whose data are used in genomic research and clinical care. As part of the Global Alliance for Genomics and Health (GA4GH), we examined the ELSI status of health-related data that are typically considered 'sensitive' in international policy and data protection laws. We propose that 'tiered protection' of such data could be implemented in contexts such as that of the GA4GH Beacon Project to facilitate responsible data sharing. To this end, we discuss a Data Sharing Privacy Test developed to distinguish degrees of sensitivity within categories of data recognised as 'sensitive'. Based on this, we propose guidance for determining the level of protection when sharing genomic and health-related data for the Beacon Project and in other international data sharing initiatives.

*npj Genomic Medicine* (2016) **1,** 16024; doi:10.1038/npjgenmed.2016.24; published online 17 August 2016

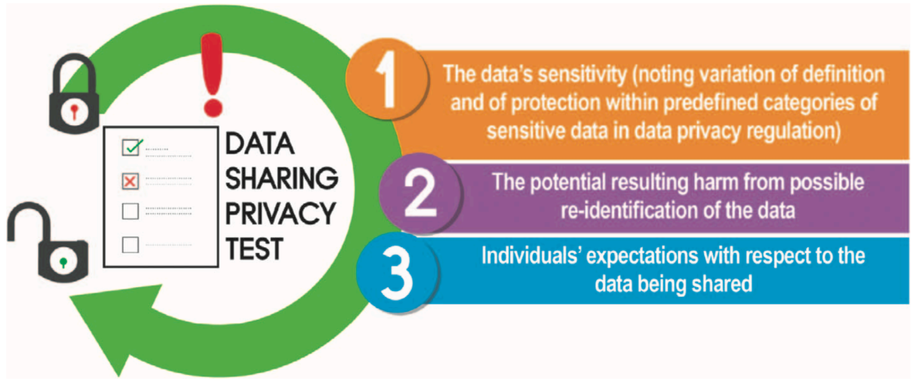



**Figure 1.** The three steps of a Data Sharing Privacy Test to distinguish degrees of data sensitivity within categories of data recognised as 'sensitive'.

John Yuyi, NYT 2018-02-09

## Welcome to *openSNP*

***openSNP*** **lets customers of direct-to-customer genetic tests publish their test results, find others with similar genetic**

For Genotyping Users

**Upload Your Genotyping File**

Upload your raw genotyping

Phenotypes are the

*openSNP* gets the latest

Home | Family tree | Discoveries | DNA | Research

MyHeritage DNA

Valentine's Day
**DNA SALE**

Only **59€** ~~89€~~ per kit
When ordering 2+ kits

Order now

Shipping not included
Ends February 14th

23andMe

Welcome to you

saliva collection kit

Find out what your DNA says about you and your family.

- See how your DNA breaks out across 31 populations worldwide
- Discover DNA relatives from around the

ancestry

SUBSCRIBE SIGN IN

**THE AVERAGE BRITISH PERSON'S DNA IS ONLY 36% BRITISH**

GROW YOUR TREE

Find your ancestors in

ancestryDNA

Discover

“We’re an information economy. They teach you that in school. What they don’t tell you is that it’s impossible to move, to live, to operate at any level without leaving traces, bits, seemingly meaningless fragments of personal information. Fragments that can be retrieved, amplified . . .”

**–William Gibson in "Johnny Mnemonic" (1986)**

# Improving Data Privacy but Empowering Beneficial Use

## Intersecting Areas of Development

- Make genomic (and functional) data "obfuscated" for malicious use
  - e.g. spiking / randomization of variants in "not-disease" loci
- access protection with defined user access using standardized protocols for users' roles and permissions, in contrast to individual per user, per dataset access requests over data access committees (DACs)
  - digital "differential" consent using e.g. data use ontologies
- intentional and unintentional (!) data providers have to be protected from abuse by legal regulations - though thin line regarding "overzealous" use by law enforcement
- alternative solution for active consent
  - encrypted wide-area networking solutions with managed access control (e.g. SPHN's BiomedIT) and limited access to anonymized data (e.g. using the Beacon protocol with "handover" scenarios)
  - (genomic) data ownership by the individual "data donors, together with strong privacy protection by law

# Modernizing Patient Consent

forward looking, transparent and technically feasible regulations for enabling access to research material and data while empowering ***patients***

## **Generalkonsent:** Eine einheitliche Vorlage soll schweizweite Forschung erleichtern

| Art des Forschungs-materials / Personenbezug | Biologisches Material und genetische Daten | Nicht-genetische Daten |
|---|---|---|
| **Unverschlüsselt (identifizierend)** | Information + Einwilligung in jedes einzelne Forschungsprojekt | Information über Weiterverwendung für zukünftige noch unbestimmte Forschungsprojekte + Generalkonsent für Forschungszwecke |
| **Verschlüsselt** | Information über Weiterverwendung für zukünftige noch unbestimmte Forschungs-projekte + Generalkonsent für Forschungs-zwecke | Information über Weiterverwendung für zukünftige noch unbestimmte Forschungs-projekte + über Möglichkeit Weiterverwen-dung abzulehnen > Widerspruchsrecht |
| **Anonymisiert** | **Genetische Daten:** Information über Weiterverwendung für zukünftige noch unbestimmte Forschungszwecke + über Möglichkeit Weiterverwendung abzulehnen > Widerspruchsrecht<br>**Proben:** Information zur Anonymisierung > Widerspruchsrecht | Ausserhalb des Geltungsbereichs des HFG |

SAMWASSM

Switzerland: Definition of a unified "Generalkonsent", to provide a single framework to manage permissions for access to patient derived material and related data

## Consent Codes: Upholding Standard Data Use Conditions

**Stephanie O. M. Dyke[1]*, Anthony A. Philippakis[2], Jordi Rambla De Argila[3,4], Dina N. Paltoo[5], Erin S. Luetkemeier[5], Bartha M. Knoppers[1], Anthony J. Brookes[6], J. Dylan Spalding[7], Mark Thompson[8], Marco Roos[8], Kym M. Boycott[9], Michael Brudno[10,11], Matthew Hurles[12], Heidi L. Rehm[2,13], Andreas Matern[14], Marc Fiume[15], Stephen T. Sherry[16]**

Global Alliance for Genomics & Health

Consent Codes

| Name | Abbreviation | Description |
|---|---|---|
| **Primary Categories** ($I^{ry}$) | | |
| no restrictions | NRES | No restrictions on data use. |
| general research use and clinical care | GRU(CC) | For health/medical/biomedical purposes and other biological research, including the study of population origins or ancestry. |
| health/medical/biomedical research and clinical care | HMB(CC) | Use of the data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry. |
| disease-specific research and clinical care | DS-[XX](CC) | Use of the data must be related to [disease]. |
| population origins/ancestry research | POA | Use of the data is limited to the study of population origins or ancestry. |
| **Secondary Categories** ($II^{ry}$) (can be one or more extra conditions, in addition to $I^{ry}$ category) | | |
| other research-specific restrictions | RS-[XX] | Use of the data is limited to studies of [research type] (e.g., pediatric research). |
| research use only | RUO | Use of data is limited to research purposes (e.g., does not include its use in clinical care). |
| no "general methods" research | NMDS | Use of the data includes methods development research (e.g., development of software or algorithms) ONLY within the bounds of other data use limitations. |
| genetic studies only | GSO | Use of the data is limited to genetic studies only (i.e., no research using only the phenotype data). |
| **Requirements** | | |
| not-for-profit use only | NPU | Use of the data is limited to not-for-profit organizations. |
| publication required | PUB | Requestor agrees to make results of studies using the data available to the larger scientific community. |
| collaboration required | COL-[XX] | Requestor must agree to collaboration with the primary study investigator(s). |
| return data to database/resource | RTN | Requestor must return derived/enriched data to the database/resource. |
| ethics approval required | IRB | Requestor must provide documentation of local IRB/REC approval. |
| geographical restrictions | GS-[XX] | Use of the data is limited to within [geographic region]. |
| publication moratorium/embargo | MOR-[XX] | Requestor agrees not to publish results of studies until [date]. |
| time limits on use | TS-[XX] | Use of data is approved for [x months]. |
| user-specific restrictions | US | Use of data is limited to use by approved users. |
| project-specific restrictions | PS | Use of data is limited to use within an approved project. |
| institution-specific restrictions | IS | Use of data is limited to use within an approved institution. |

SOM Dyke, *et al*. Consent Codes: Upholding Standard Data Use Conditions. *PLoS Genetics* 12(1): e1005772. http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005772

Contact: Dr. Stephanie Dyke (stephanie.dyke@mcgill.ca)

# Data Ownership

- Within Switzerland, there is no coherent approach on ownership of data as such (but academic discussion is ongoing, if that is needed).
- Restrictions of usage and disclosure of data other than personal data mainly stem from contractual relationships.
- In the field of research this leads mostly to a data ownership by the research institution.

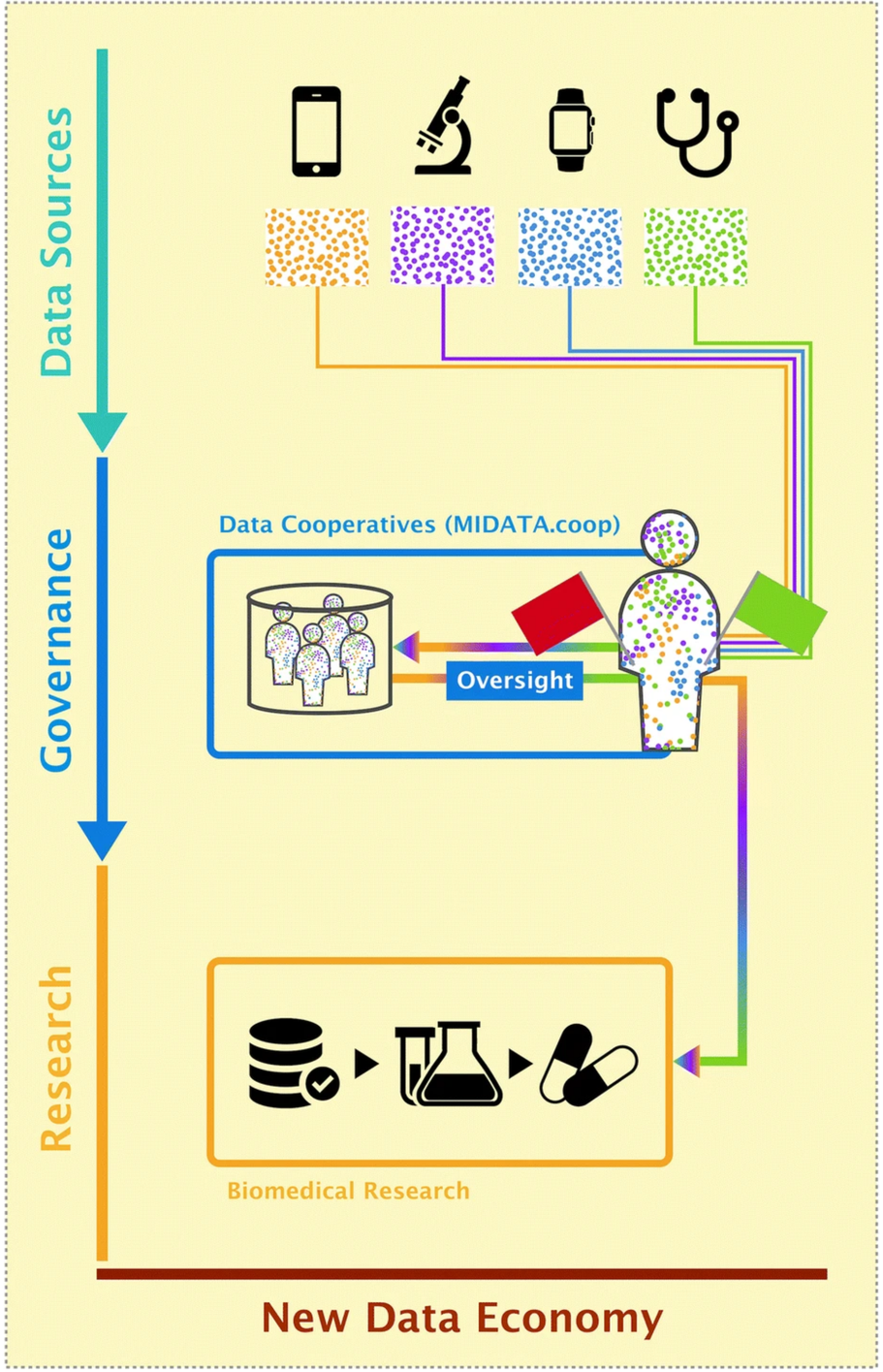
Of course the restrictions of the different acts that are in the field need to be respected (procuring data lawfully, consent for further use, etc.)

Content provided by Julian Mausbach | Oberassistent für Strafrecht und Strafprozessrecht

# Power to the People?!

## Individuals as Owners & Managers of their Data

- (genomic) data ownership by the individual "data donors"
- supported by technological frameworks for data management and arbitration
- one vision here are "data cooperatives"
- need strong support from policy makers and financial sustainability support




Citizens aggregate data from different sources and make them available for research through data cooperatives. Cooperatives offer oversight mechanisms to filter data access requests and tools for the democratic governance of the data.
Blasimme, A., Vayena, E. & Hafen, E. **Democratizing Health Research Through Data Cooperatives**. *Philos. Technol.* 31, 473–479 (2018). https://doi.org/10.1007/s13347-018-0320-8

graphic credit Manuel Schneider

Generalkonsent

BENEFIT

BLOCKCHAIN

HEALTH

PRIVACY

SECURITY

CONSENT

ACCESS

Right to Research

HACKERS

LAWS

Genetic
Information
Nondiscrimination
Act

Health
Insurance
Portability and
Accountability
Act

SAFETY

CRYPTOGRAPHY

# Genomic Data & Privacy - Key Areas

- **Re-identification**
  - identification of an individual based on sets of genomic variants they (or close relatives) carry - so one needs some genome data first
  - information to be gained is circumstantial (e.g. their genome is in a particular disease related dataset)
  - currently only risk with some practical use (e.g. **long-range familial attacks**)
- **Genotype-to-Phenotype (G2P) attacks**
  - determination of some disease risk or phenotypic features from a genome itself
  - needs access to genome data which is illegal in many jurisdictions (but technically more & more feasible)
  - real-world use cases are limited but abuse through wrong perception of utility
- **Genomic Determinism**
  - assignment of individual abilities and personal development trajectories from genomic profiling
  - topic of (some good, most bad) SciFi
  - but: **Wehret den Anfängen**!

# Genomic Data & Privacy - Some Take-Home Messages

- Many clinical and research applications in genomics **need vast numbers of genomes** to evaluate e.g. genotype-phenotype relationships
- Such data cannot simply be provided by a few reference data curation resources - and those again rely on multitudes of original data resources > **federated data access** + **data curation**
- Genomic data is considered to potentially expose unwilling individuals through **re-identification**/de-anonymization but also through direct information (genotype -> phenotype/disease)
- Legislative bodies and law enforcement have varying and *curious* approaches to "genomic privacy", with a mix of de-legalizing genomic data generation (e.g. in Switzerland) or strictly limiting its use while also using "eminent domain" to co-opt such data for criminal persecution in a possibly extending set of use cases

# Technical, legal and ethics aspects of genomic data sharing

## DSI Proposal Summary

- The area of "personalized" or "precision" health relies on the use of molecular characteristics - e.g. inherited genomic variations or mutations in a tumor genome - to tailor individual recommendations in such diverse areas as nutrition, pharmaceutic selection and dosing or use of preventive screening, up to the tailoring of antineoplastic therapies based on the molecular profiling of a patients cancer cells. While it has been widely recognized that many of these applications can provide personal health benefits and have a positive impact on society at large, large questions remain regarding the secondary use of genomic data, the potential dangers arising from collection, storage and sharing of genomic information as well as the associated legal regulations.

- The need to have access to an ever increasing set of genomic data for biomedical research and treatment decisions is being addressed on an international level through such efforts as the Global Alliance for Genomics and Health **GA4GH** - of which **UZH is a founding member** - or European B1MG initiative. The Swiss Personalized Health Network **SPHN** provides the technical and and logistic backbone for the exchange of genomic and clinically derived information for research applications and clinical re-use, while not directly generating data itself.

- Issues related to the generation, handling and application of personal genomic information span a large set of academic specialities, from **medicine**, **life science** research, **computational science** and **data security** to **ethics**, **law**, **psychology** as well as **communication** and **media sciences**. The topic of "**Technical, legal and ethics aspects of genomic data sharing**" is a natural fit for the Digital Society Initiative, offering participation and networking opportunities for a wide range of potential stakeholders from different faculties to develop projects in an area with **scientific relevance** and large **public interest**.


University of Zurich[UZH]

Global Alliance
for Genomics & Health


Prof. Dr. Michael Baudis
Institute of Molecular Life Sciences
University of Zurich
**SIB** | Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

*progenetix.org*
*info.baudisgroup.org*
*sib.swiss/baudis-michael*
*imls.uzh.ch/en/research/baudis*
*beacon-project.io*
*schemablocks.org*