

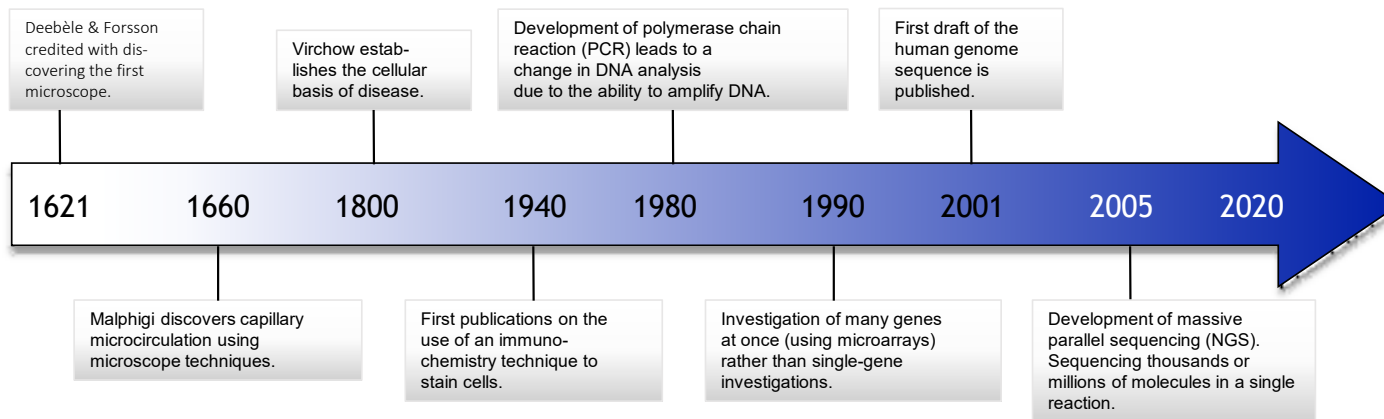
# Standards: Towards data sharing

Eivind Hovig

University of Oslo & Oslo University Hospital

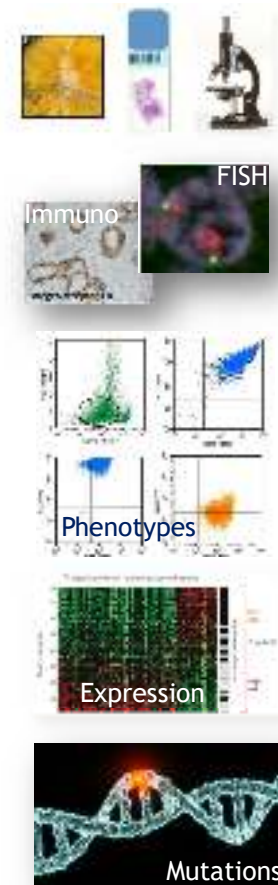


# Modern pathology

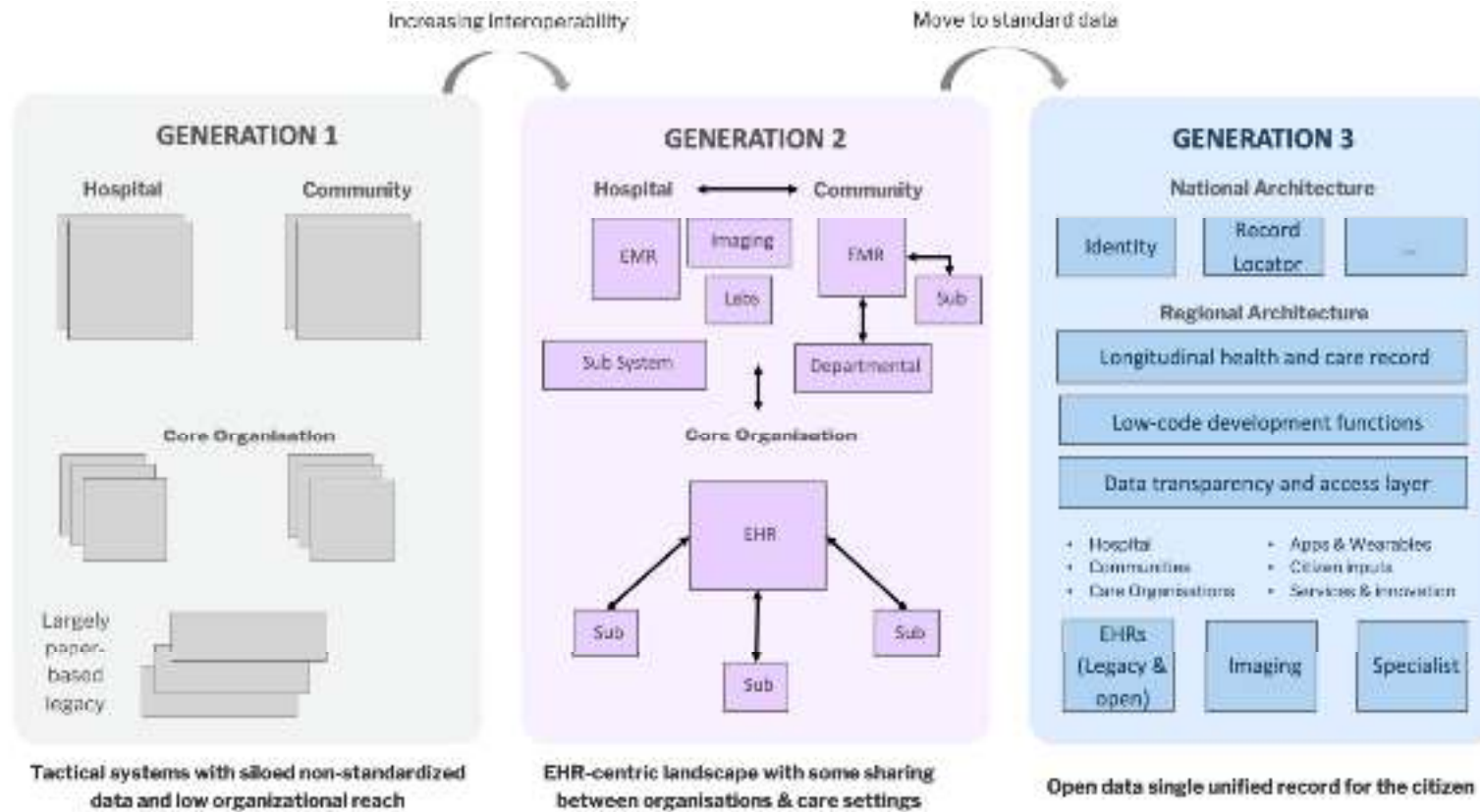


Modern pathology has become multidisciplinary, integrating

- Morphology
- Histology
- Phenotypic data
- Molecular data



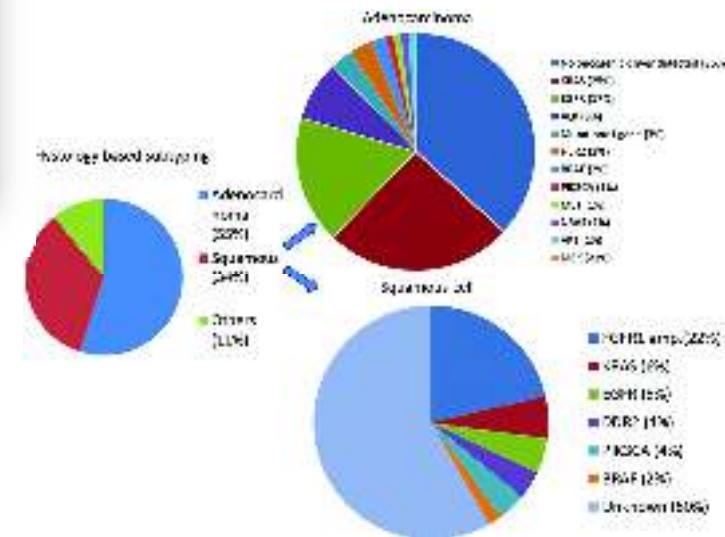
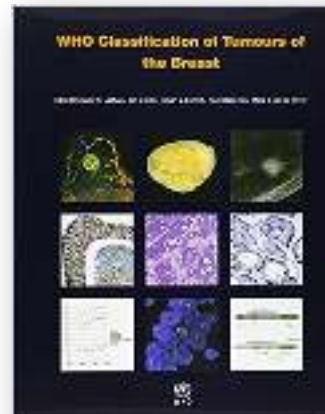
# Evolution of the health and care landscape



Source: Professor Rachel Duncombe NHS Digital Academy

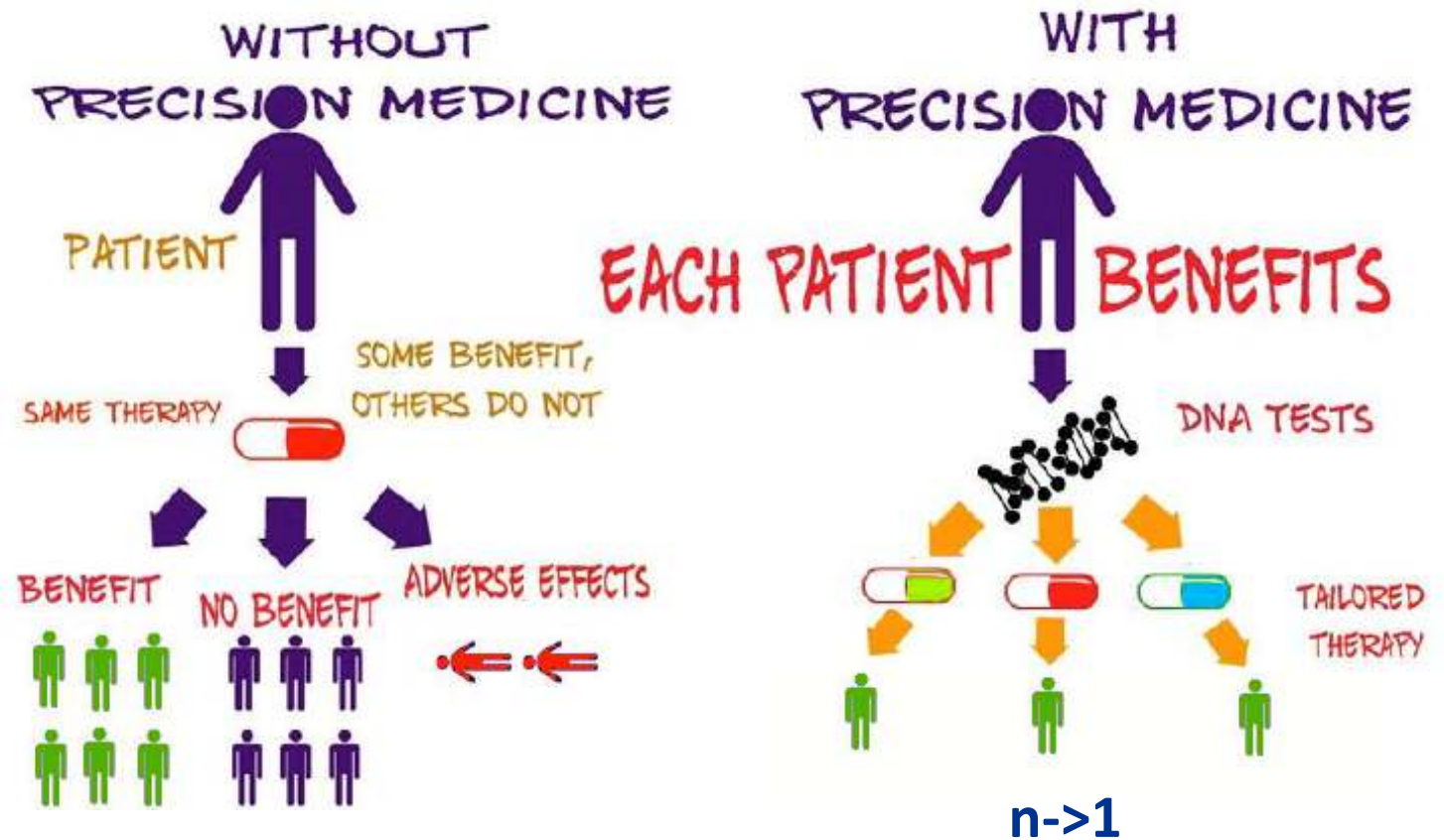
# Increasing number of biomarkers

- **Diagnostic biomarkers**
  - Gene fusions, *EWSR*, *SSX*, *BCR-ABL*
  - *KIT*.... mutations
- **Prognostic biomarkers**
  - *FLT3*, *TP53*, *KRAS*.... mutations
  - *EGFR* amplifications
- **Predictive biomarkers**
  - *EGFR*, *BRAF*, *FGFR*, *KIT*.... mutations
  - *ALK*, *RET*, *NTRK* fusions
  - *ERBB2* amplification
  - MSI, mutational burden, BRACAness



There is a need for advanced diagnostics to cope with the increasing number of new biomarkers, and to better identify cancer patients into clinical trials

# Precision medicine concept



Adapted from USA Today

# Need: expanding population when $n \rightarrow 1$ ...

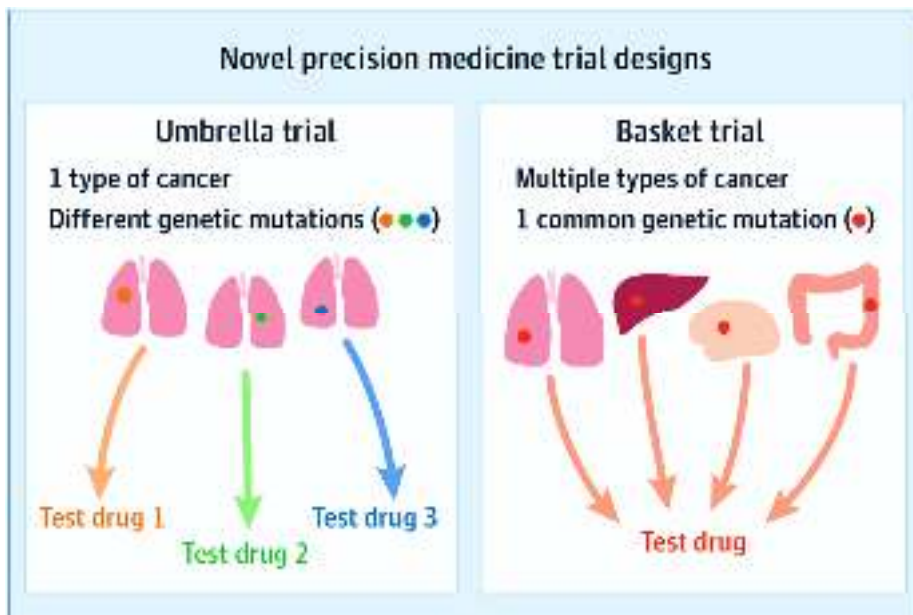
With increasing search space, expansion is often required

But how?

- Federation of data for increased power for detection ....
- Institutions and countries must collaborate to merge data...

These are new challenges that require new solutions...

# Study-design: combined umbrella-basket Simon two-stage model

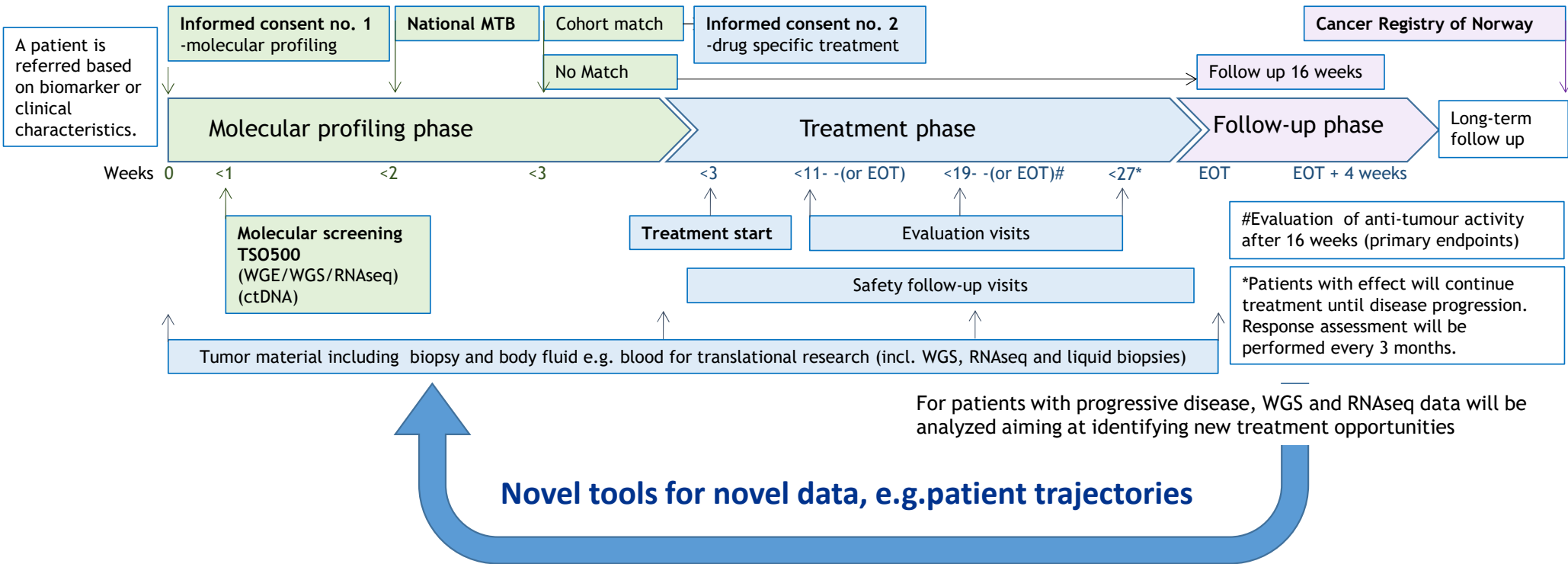


## A combination

- Different cancers
  - Different genetic mutations or other disease-modifying changes
- ↓
- Treatment to fit the individual patient disease and tumor profile
- ↓
- A portfolio of registered drugs

By West et.al JAMA Oncology 2017

# IMPRESS-Norway: Study Outline





## DRUP in the Netherlands and similar studies in US, Canada, the Nordics and other European countries ongoing or to start



### ProTarget

A Danish Nationwide Clinical Trial on Targeted Anti-Cancer Treatment based on Molecular Profiling

FIN-DRUP  
DRUG REDISCOVERY PROGRAM IN FINLAND



## Canadian Profiling and Targeted Agent Utilization Trial (CAPTUR)

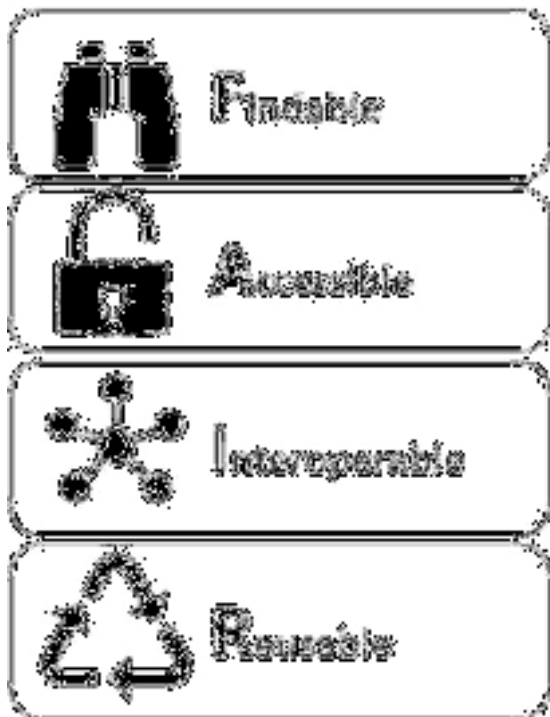


# How to realize the potential?

Large datasets of relevance available on the fly:

- Datasets consented and legally accessible
- Legal access clarified and machine readable
- Metadata machine readable

# FAIR data and tools



Data and materials enriched with metadata assigned with a unique identifier

Data and metadata stored in a trusted repository with an open and free protocol. Accessible by machines and humans

Using vocabularies and public domain ontologies, the metadata can be referenced and linked

Additional documentation and protocols describing the acquisition of the data, licensed with a detailed provenance

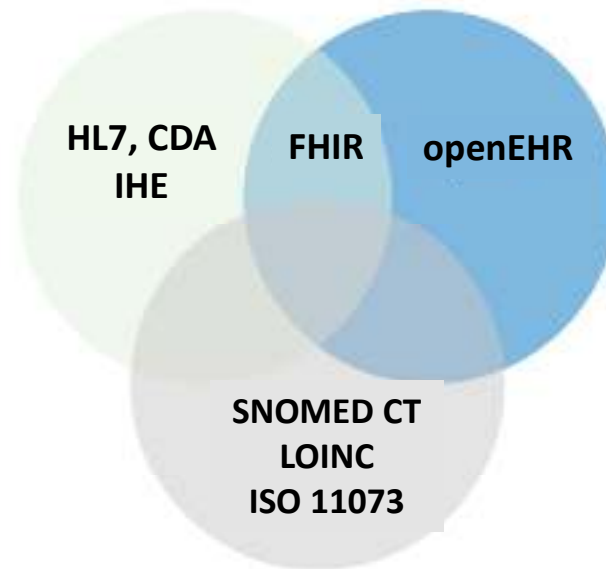
# Standards are interoperable: Clinical data

**Ontologies, structured vocabularies (ODHSI OMOP and others)**

**Patient-Reported Outcomes Measurement Information Systems (RedCap and others)**

**LOINC** = Logical Observation Identifiers Names and Codes  
**OpenEHR** = Open Electronic Health Records  
**(LOINC)** = Fast Health Interoperable Resources  
**HL7** = Health Level 7  
**CDA** = Clinical Document Architecture

Exchange and workflow



Clinical models and persistent data

Clinical terminology

Source: Professor Rachel Duncombe NHS Digital Academy

# **Sensitive human data are made interoperable**

**Software can be brought to data – container technology**

**Systems are available for federated analysis**

**Trusted research environments (TRE)**

**Federated EGA – sensitive data made discoverable and  
accessible**

**Compute resources available when required**

**AI tools available when required**

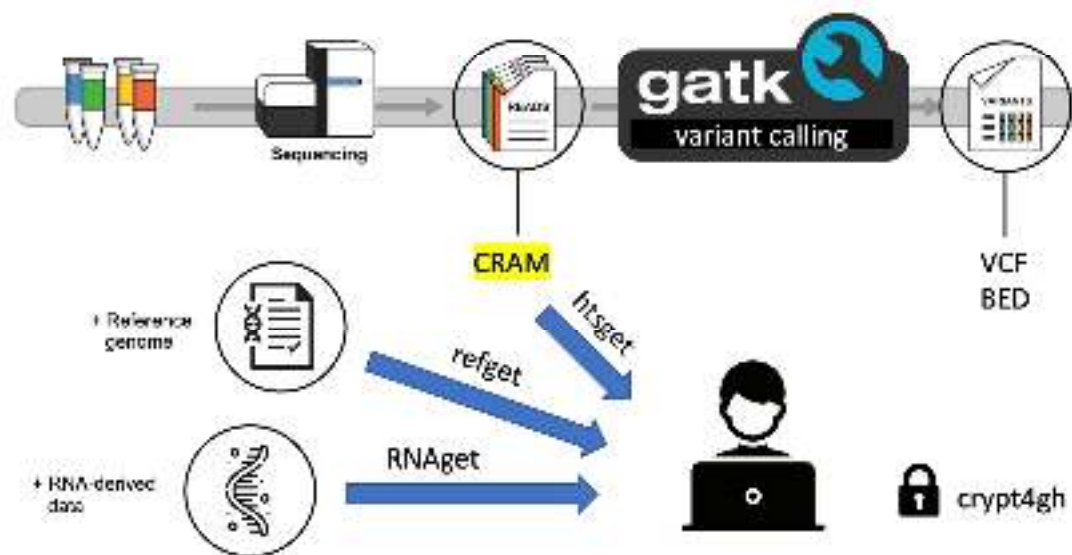


# Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

## Large-Scale Genomics

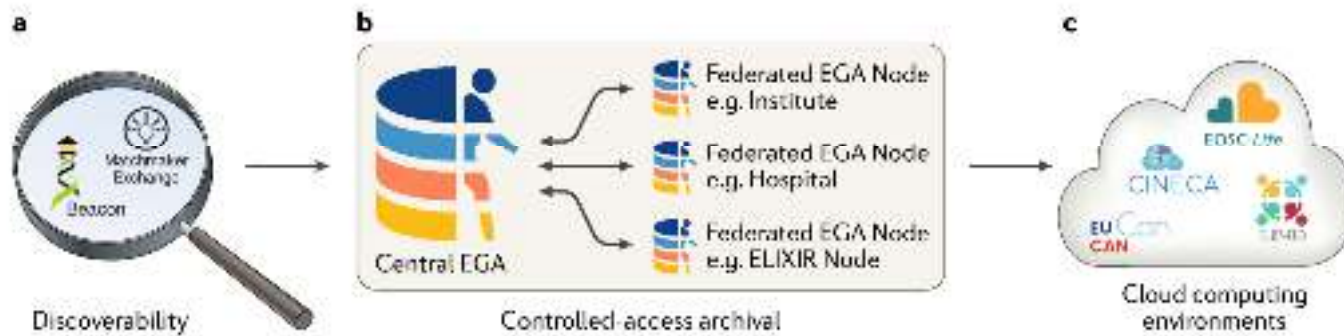
Standardised file formats and remote access protocols for storing, compressing, encrypting, querying and sharing genomic data at scale.





# Developing standards and driving use

Driving interoperability using GA4GH standards



Beacon  
Phenopackets  
Authentication and  
Authorization  
Infrastructure (AAI)

Data Use Ontology  
Read file formats (SAM/BAM,  
VCF/BCF, CRAM/CRYPT4GH)  
AAI  
htget  
refget

Tools Registry Service  
Workflow Execution  
Service  
Data Registry Service  
Task Execution  
Service



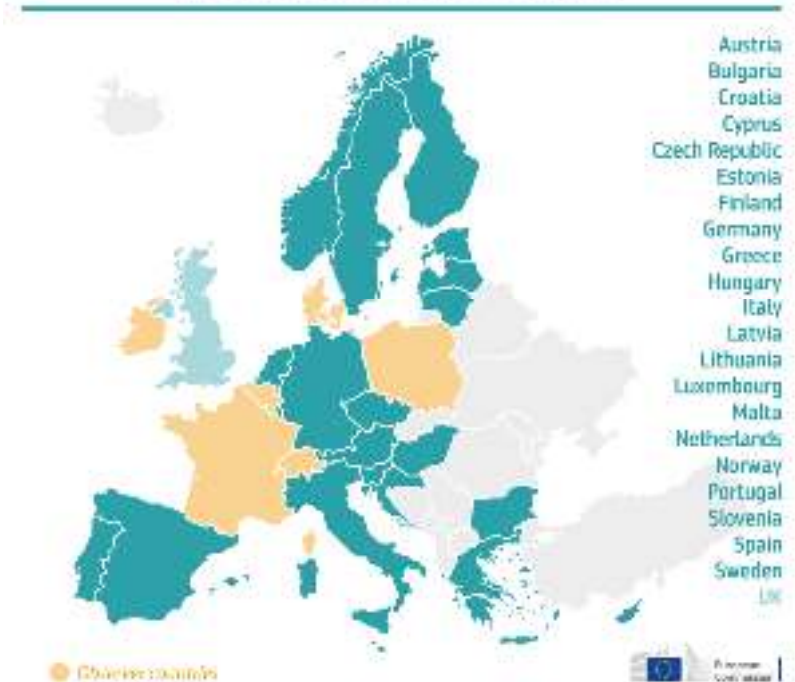
# 1+MG Declaration of cooperation - April 2018



DECLARATION OF COOPERATION

*Towards access to at least 1 million sequenced genomes in the European Union by 2022*

EU countries agreed to cooperate in linking genomic data across borders



**22 countries have now signed; 6 are observers**

<https://ec.europa.eu/digital-single-market/en/european-1-million-genomes-initiative>



# Some ongoing efforts to help realize cancer sharing



## DIGICORE

DIGICORE is a pan-European research network built to accelerate the implementation of precision oncology in Europe.

DIGICORE promotes and equips cancer centres in their use of routine electronic health records (EHR) and molecular diagnostic information (MDX) for trial automation, real world outcomes research, digital diagnostics and care quality management.

## Genomic Data Infrastructure (GDI) DIGITAL-2021-CLOUD-AI-01-FEI-DS-GENOMICS

- 40 M Euros (20M of this is in kind)
- 54 partners
- 21 countries
- National anchoring in each country

# Connecting to the European Health Data Space

