



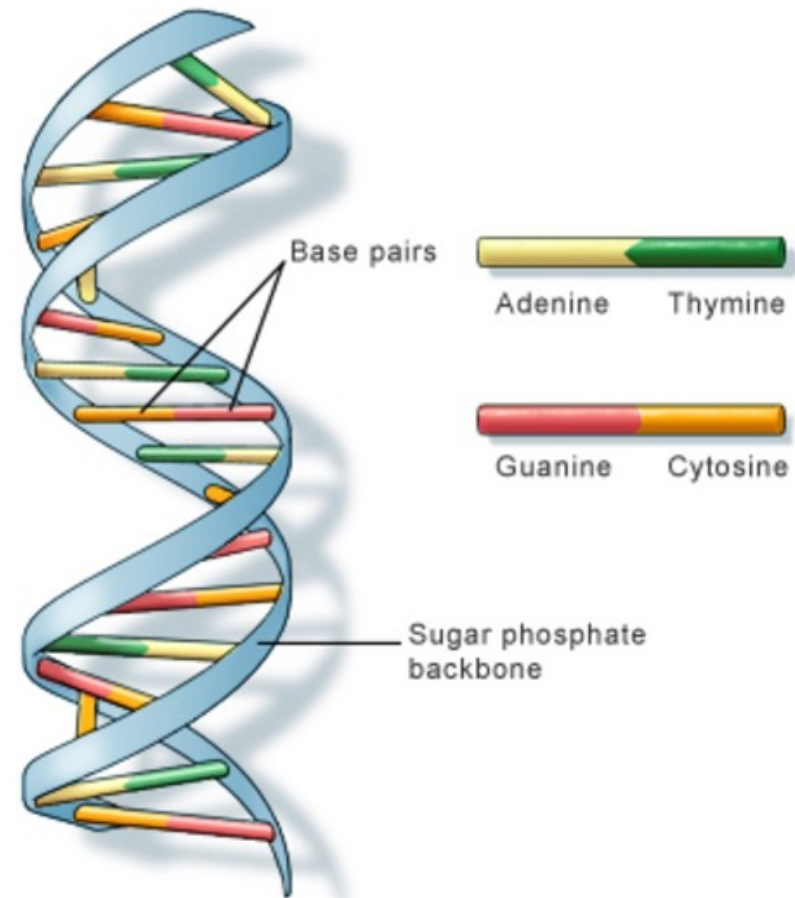
Central European Institute of Technology
BRNO | CZECH REPUBLIC



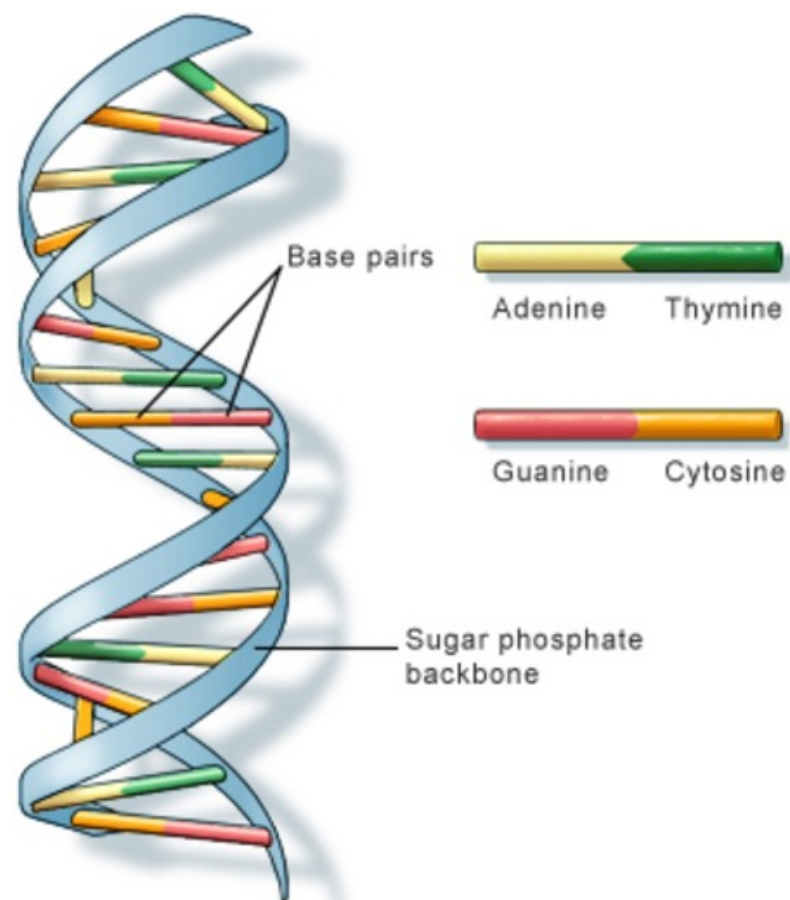
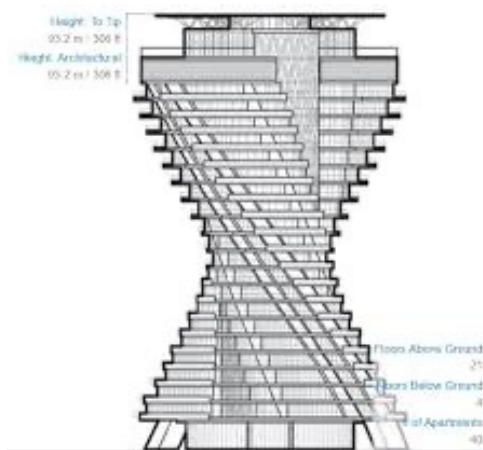
Sequencing Data Management: Specifics, Applications, and European Initiatives

ELIXIR CZ Annual Conference 2024: Data management

Vojtech Bystry
vojtech.bystry@ceitec.muni.cz

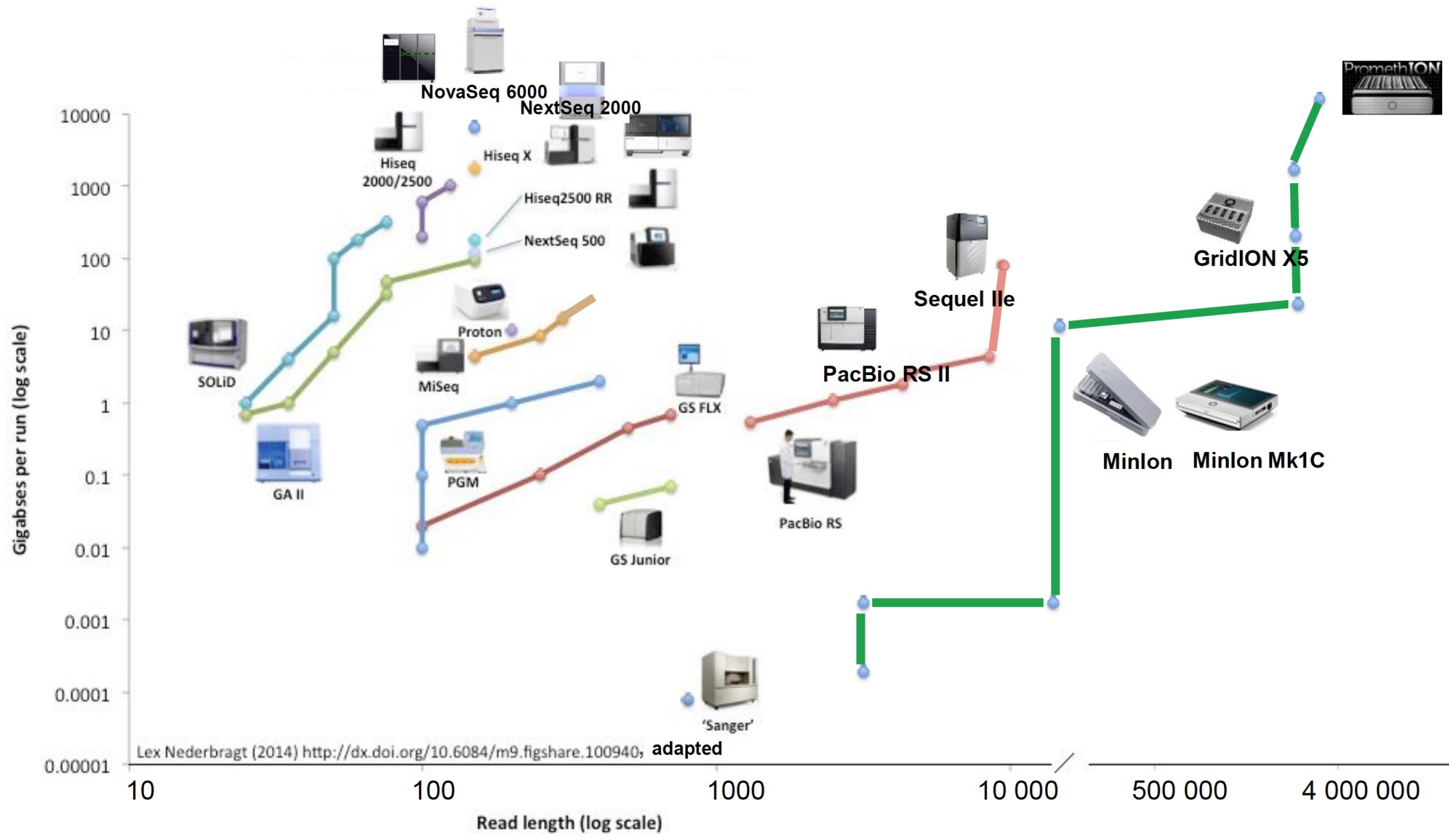


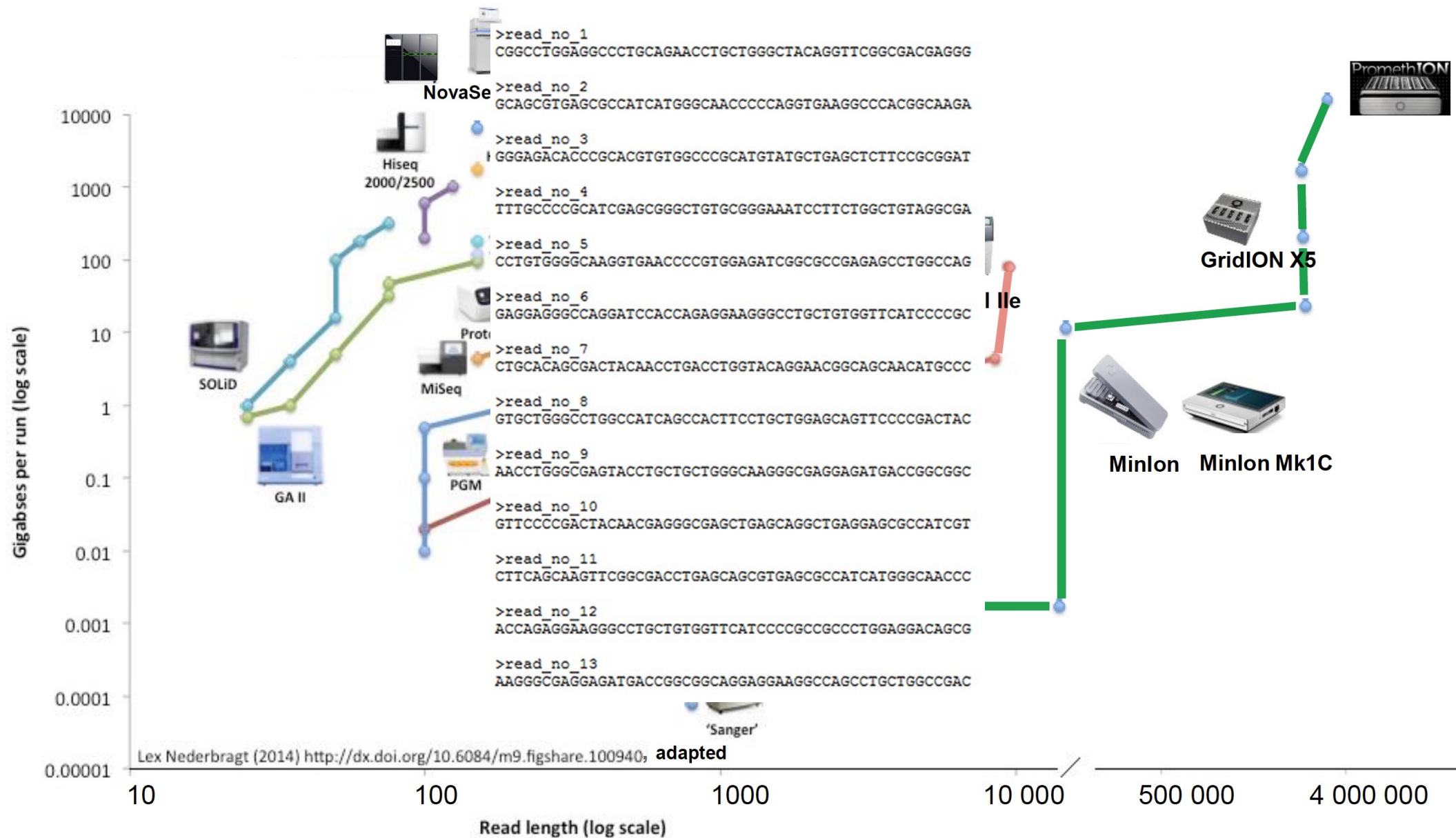
U.S. National Library of Medicine



U.S. National Library of Medicine

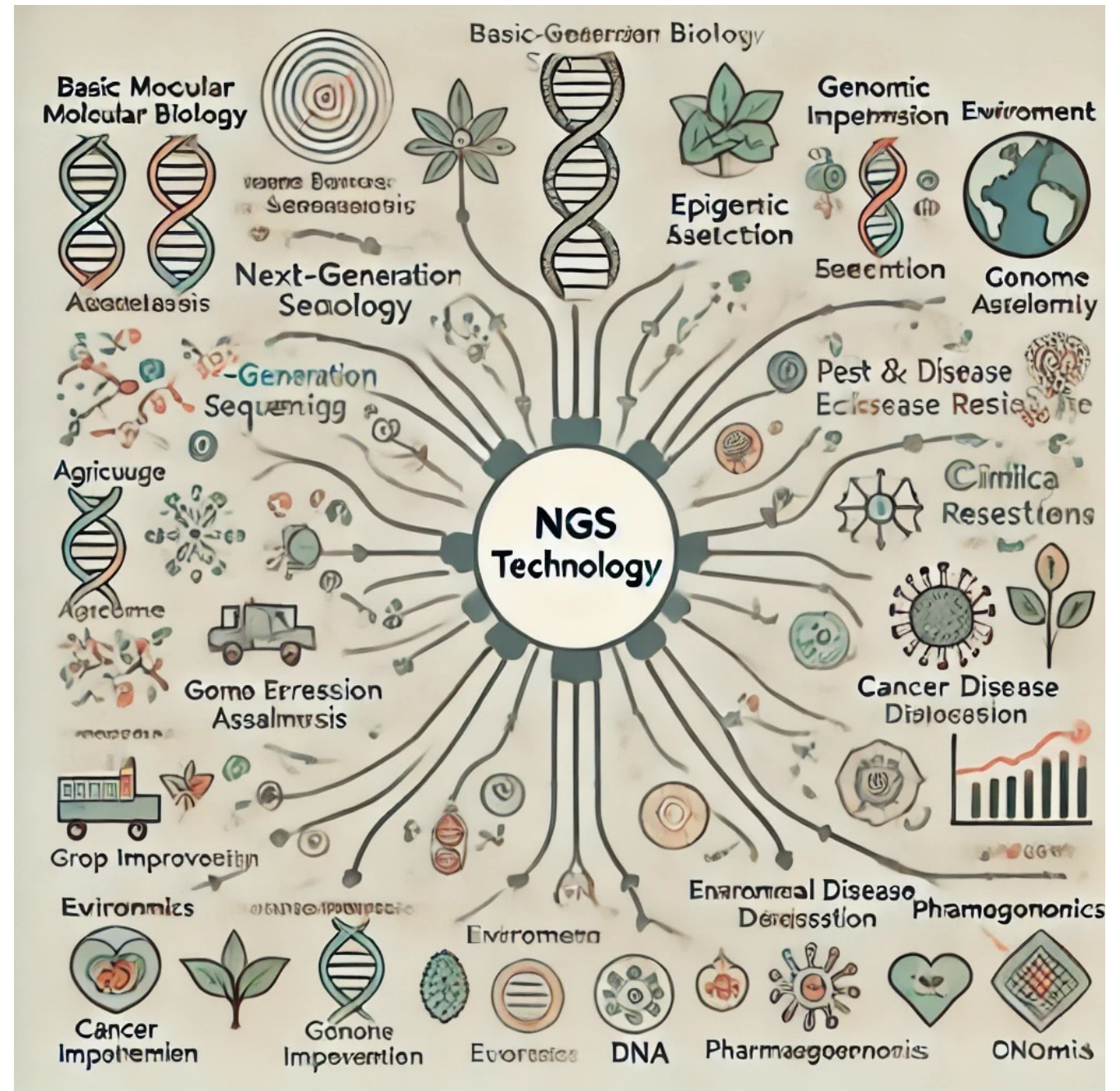






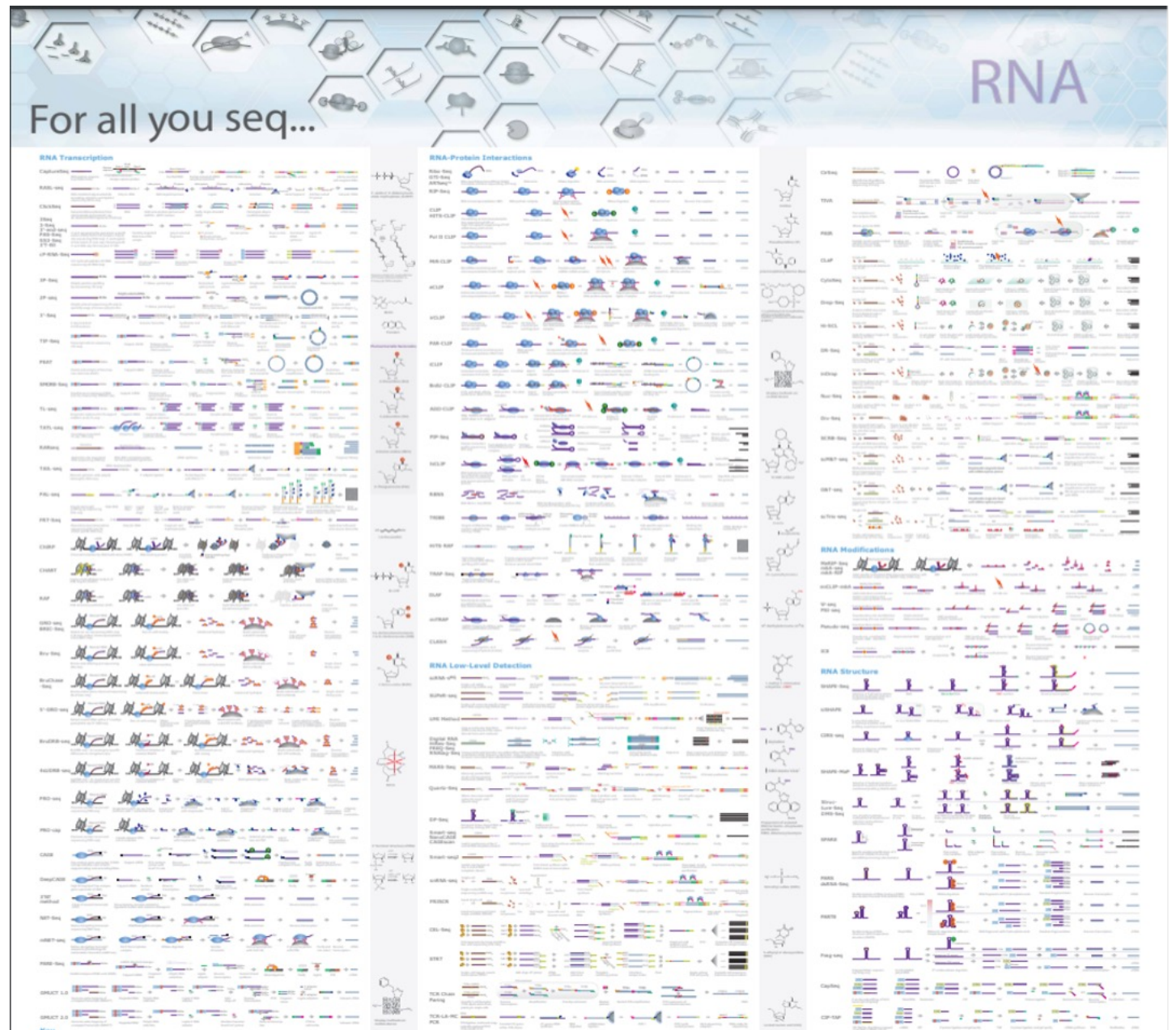
Broad domain application

- Basic Molecular Biology
- Environmental Studies
 - Agriculture
 - Archeology
 - Human Health



Broad experimental design

- Genomic variants
- Expression profiling
- Methylation profiling
- Species profiling



Broad experimental design

- Genomic variants
- Expression profiling
- Methylation profiling
- Species profiling





Broad institute



EMBL/EBI

Broad institute
NCBI



EMBL/EBI
Ensembl

Broad institute

NCBI

HG19, HG38 – chr1



EMBL/EBI

Ensembl

GRCh37, GRCh38 - 1

Broad institute

NCBI

HG19, HG38 – chr1

SRA

Sequence Read Archive

GenBank



EMBL/EBI

Ensembl

GRCh37, GRCh38 - 1



Broad institute

NCBI

HG19, HG38 – chr1

SRA

Sequence Read Archive

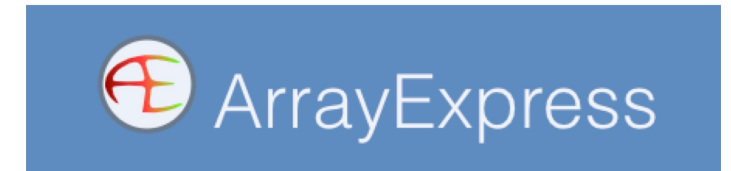
GenBank



EMBL/EBI

Ensembl

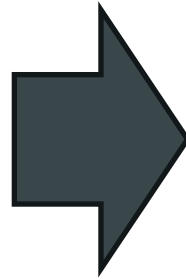
GRCh37, GRCh38 - 1



Human Genomic Data



Human
Genomic
Data



Vojtech Bystry
vojtech.bystry@ceitec.muni.cz

Technical Report | Published: 19 October 2014

Comprehensive variation discovery in single human genomes

[Neil I Weisenfeld](#), [Shuangye Yin](#), [Ted Sharpe](#), [Bayo Lau](#), [Ryan Hegarty](#), [Laurie Holmes](#), [Brian Sogoloff](#), [Diana Tabbaa](#), [Louise Williams](#), [Carsten Russ](#), [Chad Nusbaum](#), [Eric S Lander](#), [Iain MacCallum](#) & [David B Jaffe](#) 

[Nature Genetics](#) **46**, 1350–1355 (2014) | [Cite this article](#)

9833 Accesses | **135** Citations | **49** Altmetric | [Metrics](#)

JOURNAL ARTICLE

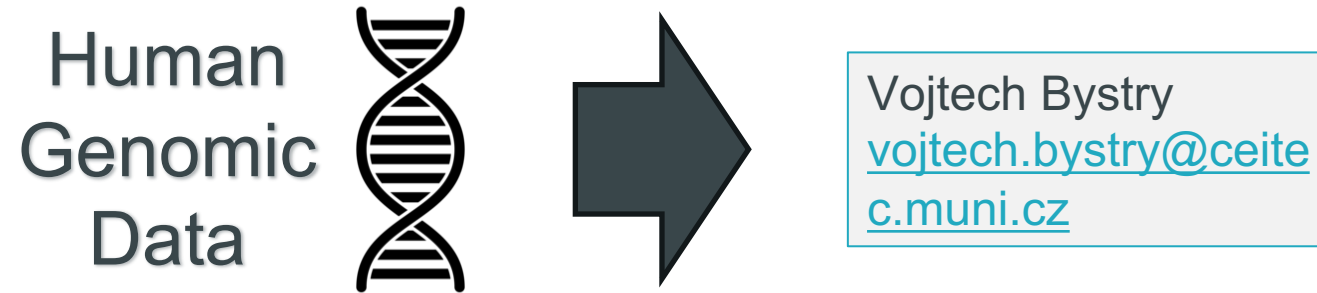
Re-identification of individuals in genomic data-sharing beacons via allele inference

[Nora von Thenen](#), [Erman Ayday](#) , [A Ercument Cicek](#) 

Bioinformatics, Volume 35, Issue 3, February 2019, Pages 365–371,

<https://doi.org/10.1093/bioinformatics/bty643>

Published: 20 July 2018 **Article history** ▼



[Am J Hum Genet](#). 2009 Feb 13; 84(2): 210–223.
doi: [10.1016/j.ajhg.2009.01.005](https://doi.org/10.1016/j.ajhg.2009.01.005)

PMCID: PMC2668004
PMID: [19200528](https://pubmed.ncbi.nlm.nih.gov/19200528/)

A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals

[Brian L. Browning](#)^{1,*} and [Sharon R. Browning](#)¹

► [Author information](#) ► [Article notes](#) ► [Copyright and License information](#) [PMC Disclaimer](#)

HOME > SCIENCE > VOL. 339, NO. 6117 > IDENTIFYING PERSONAL GENOMES BY SURNAME INFERENCE

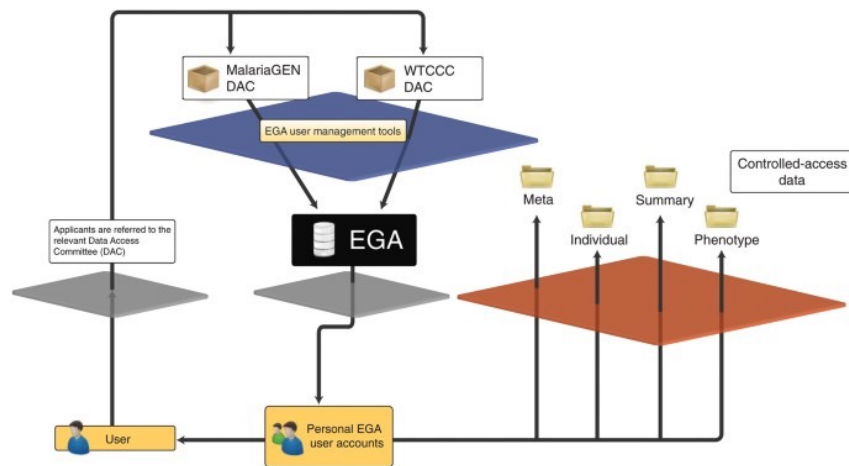
 | **REPORT**

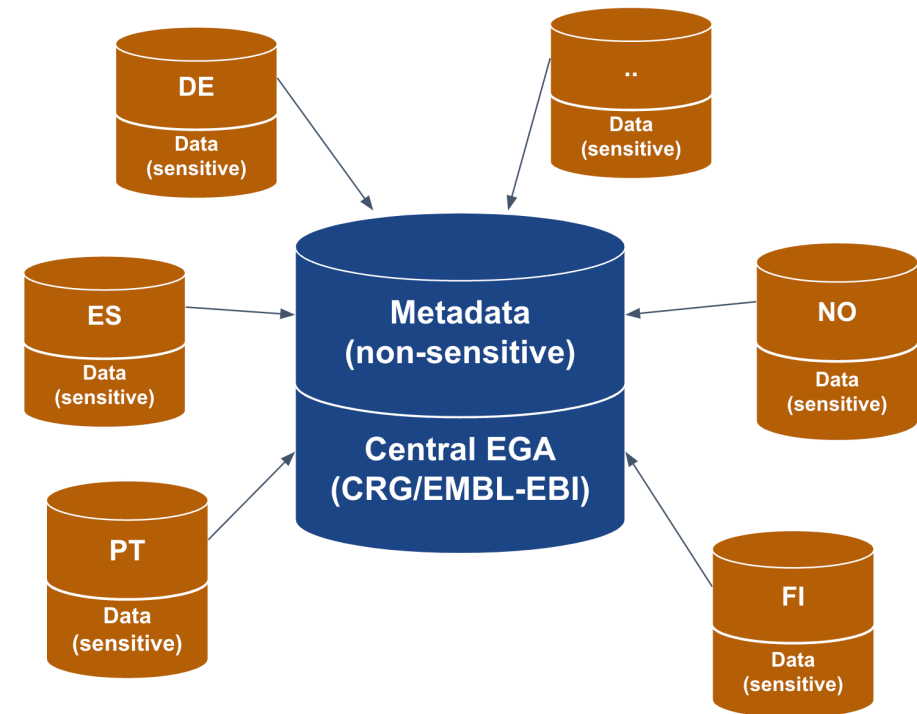
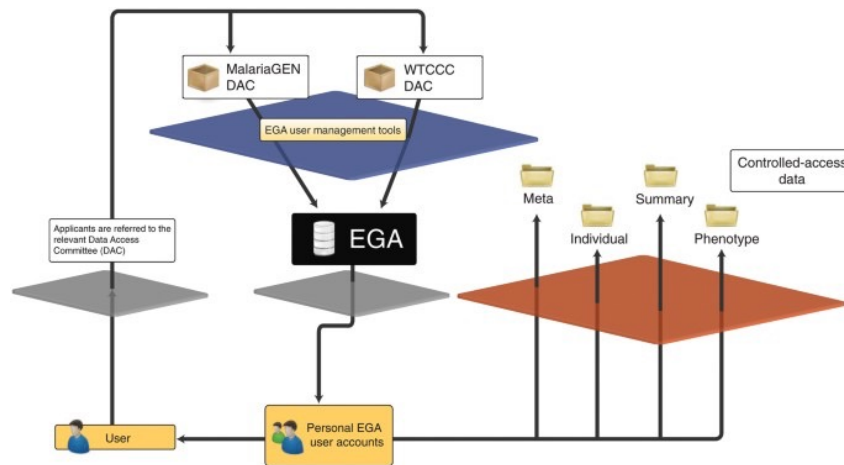
      

Identifying Personal Genomes by Surname Inference

[MELISSA GYMREK](#), [AMY L. MCGUIRE](#), [DAVID GOLAN](#), [ERAN HALPERIN](#), AND [YANIV ERLICH](#) [Authors Info & Affiliations](#)

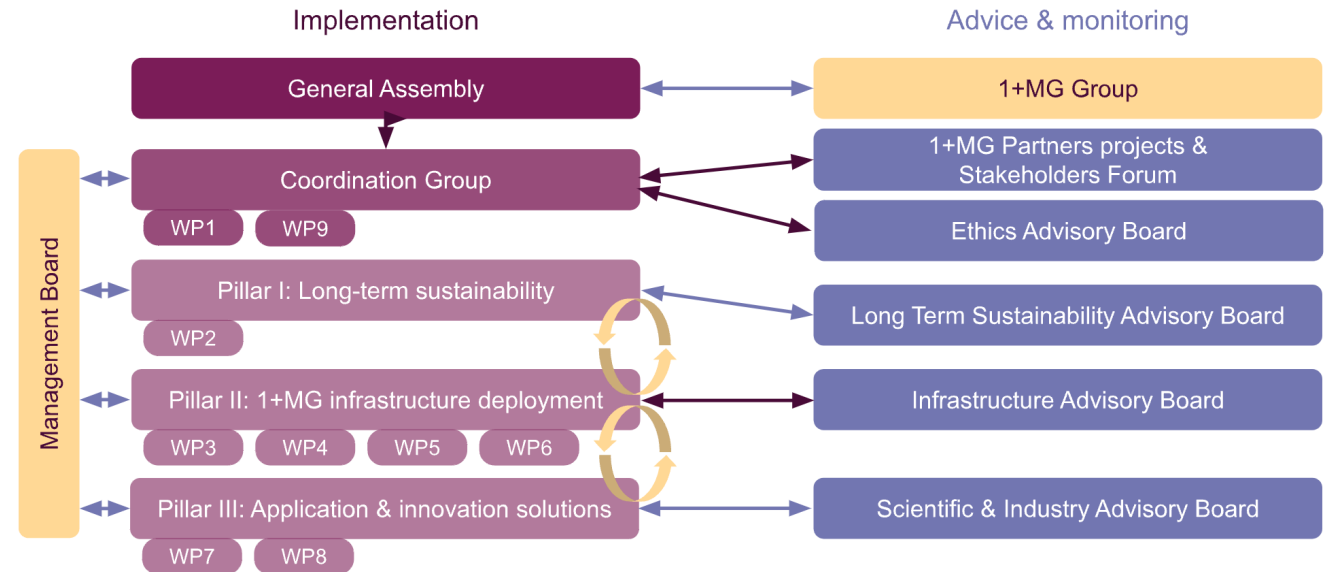






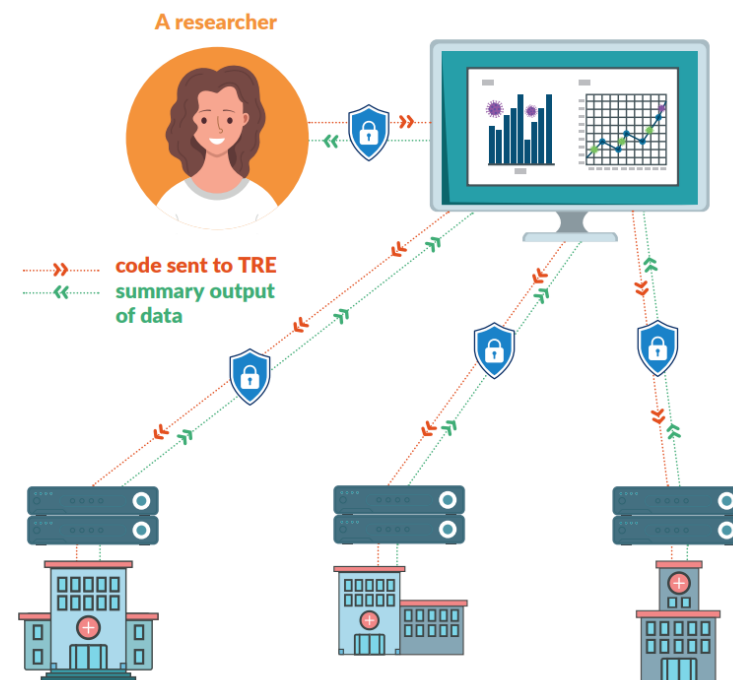
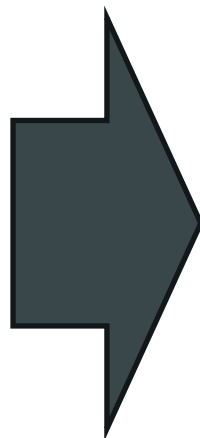


European Genomic Data Infrastructure





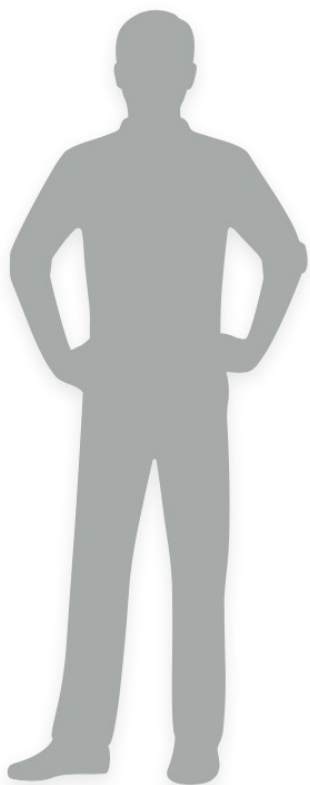
Data processing



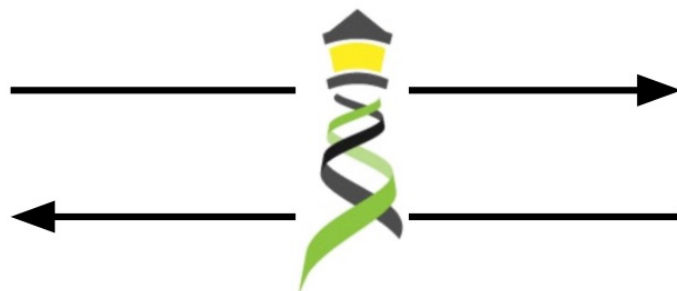
Product	Outline	Prod	PO	Function
Sensitive Data Archive	Securely stores data	✓	Sweden	
Life Science AAI	Provides a federated Identity	✓	Czechia	
REMS	Allows data access applications and decisions	✓	Finland	
Beacon	Genetic and phenotypic data discovery	✓	Spain	
Beacon Network	Federated network of Beacons	✓	Spain	
Synthetic Data	Artificial anonymous data		Finland	
htsget	Secure genetic data distribution		Sweden	
Containerised Computation	Computation via virtualised portable software packages	✓	Czechia	
Federated Computation	Interoperable distributed workflows			
Packaging and Deployment	Packaging and deployment of the starter kit		Spain	
User Portal – Data Catalogue	European level catalogue of data within nodes		Netherlands	
User Portal – Access management	European level application / access management		Luxembourg	

✓ - Product **in production** in another project (NB: Beacon V1 / ELIXIR Beacon Network)

European level



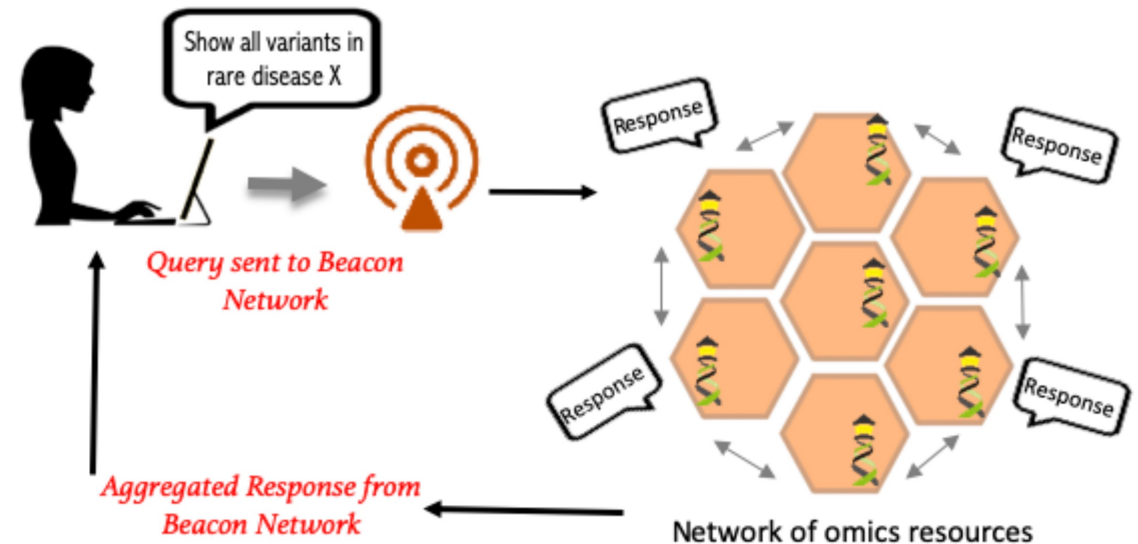
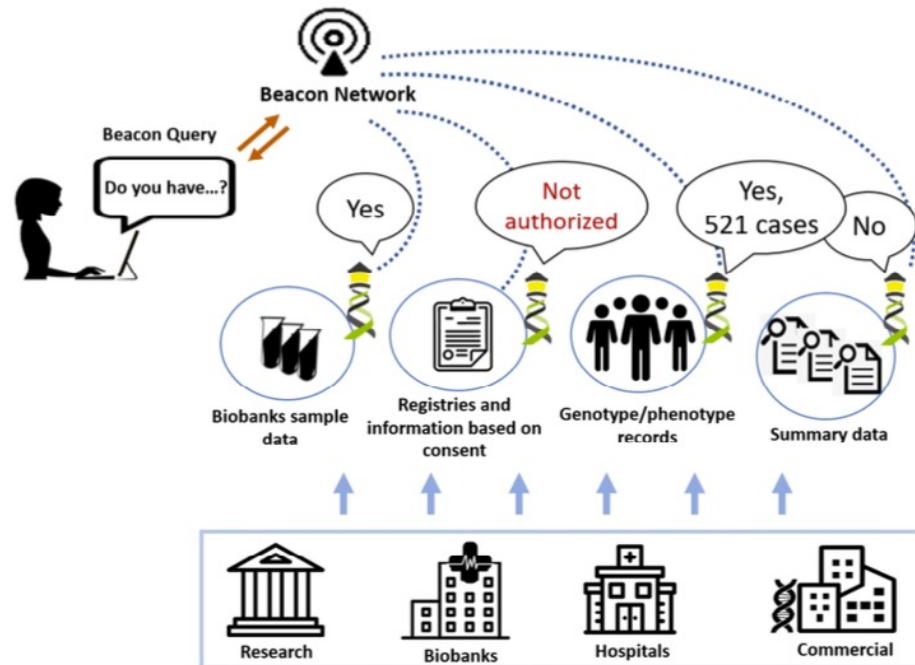
“Do you have a ‘C’ at
chromosome 13 at
position 32,936,732?”



“Yes” (or “no”)

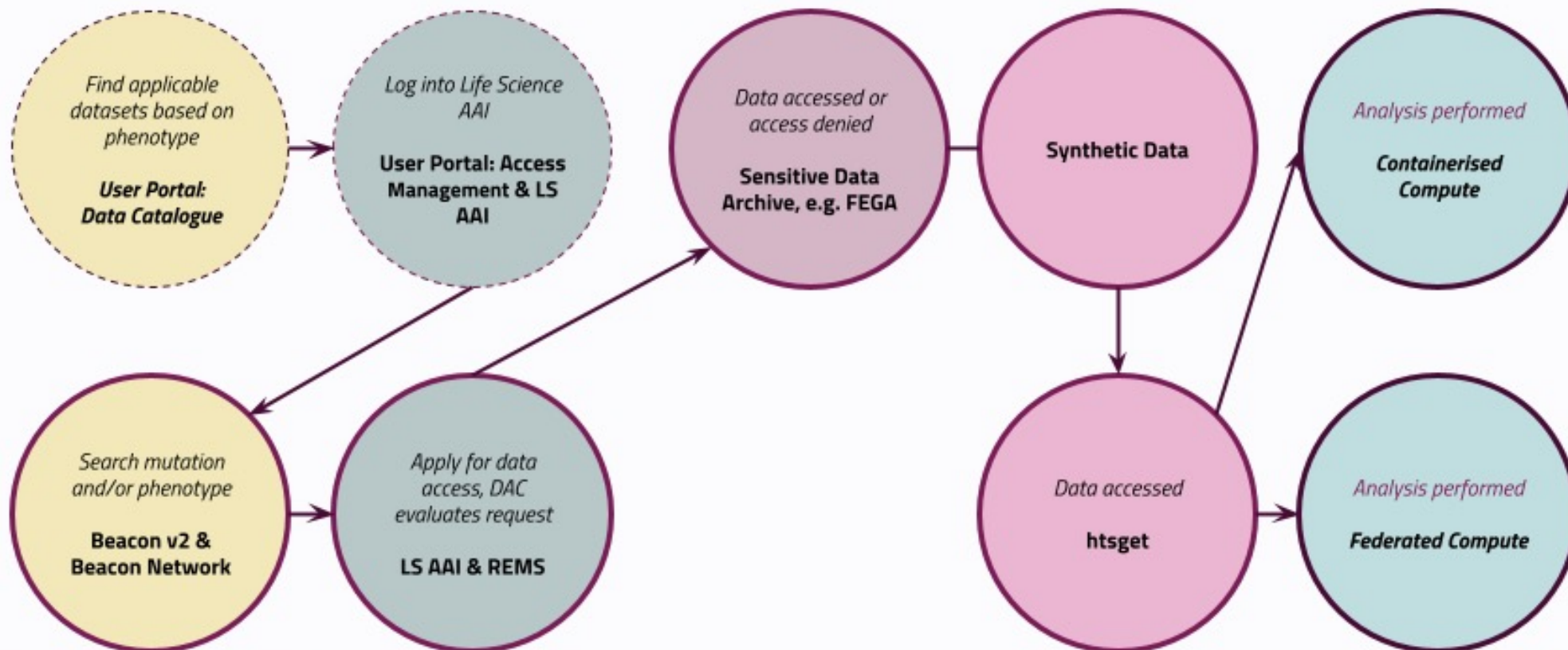


Beacon Network





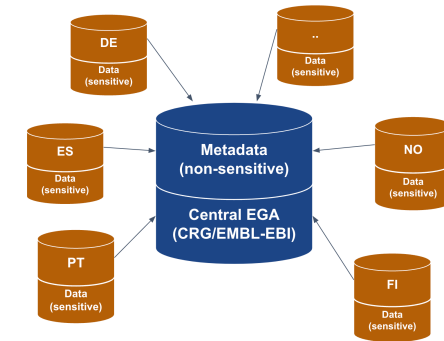
STARTER KIT



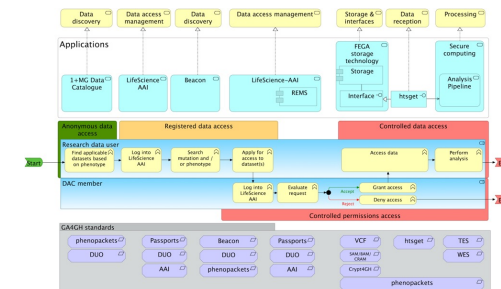
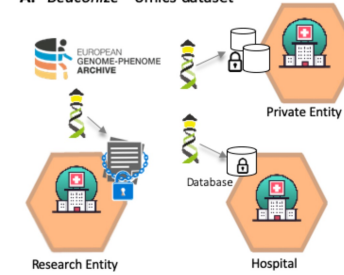
- FEAGA

- Beacon Network

- GDI (Starter kit)



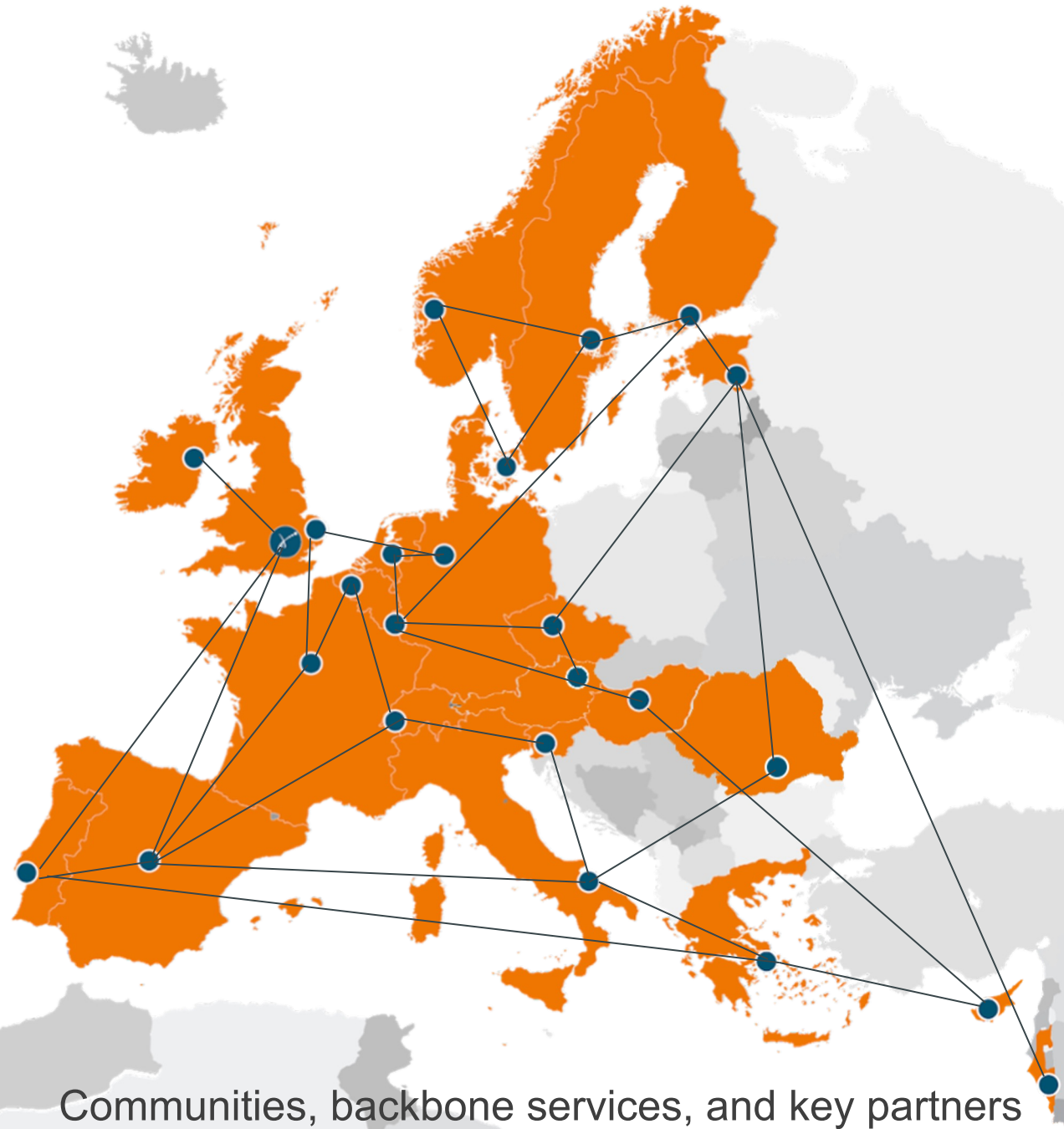
A. "Beaconize"-omics dataset



ELIXIR Human Genomic data



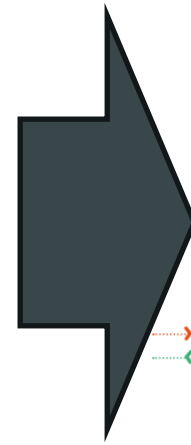
Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.



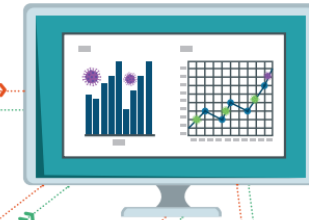
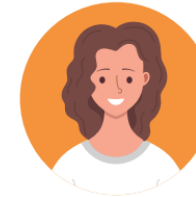
ELIXIR Human Genomic data



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.



A researcher



code sent to TRE
summary output
of data



Communities, backbone services, and key partners