The 1+Million Genome

Ivo Glynne Gut, CNAG-CRG, Barcelona

1/12/2021

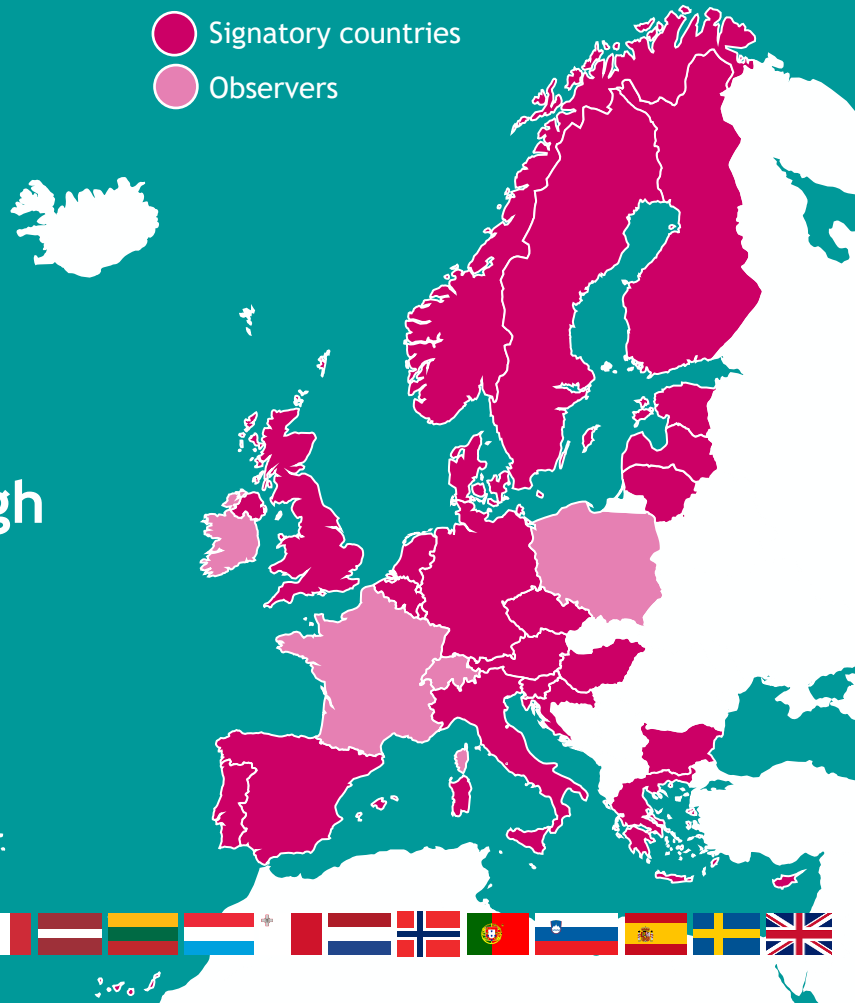
cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG
Centre for Genomic Regulation

Why do we need secure cross border access to genomic data that is generated at a national level?

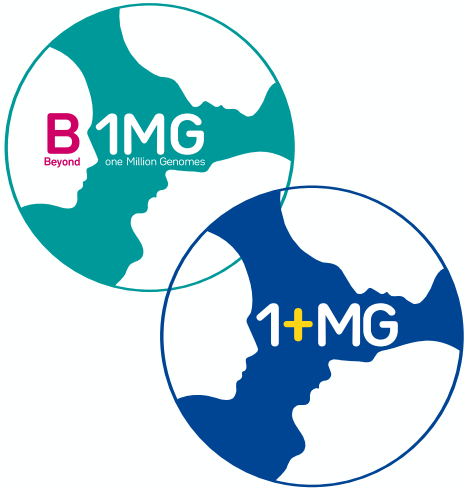
1+MG Declaration of cooperation: 2018

-24 signatory countries
-4 observers (France, Ireland, Poland, Switzerland)



Final Goal:

Cross-border access to 1+Million high quality whole genome sequence datasets as reference (by 2022)



- Beyond One Million Genomes (B1MG) aims to create legal guidance, best practices and recommendations to create infrastructure to enable the commitment of 22 European Member States and Norway to give cross-border access to one million sequenced genomes by 2022 **(1+ Million Genomes Initiative)**

Accessing genomic data at scale across borders

Long-term strategy: cross-border access to genomic data, implementation of genomics-based health

1+MG Group, National Mirror Groups and Thematic Working Groups

Use Cases Working Groups: cancer, infectious diseases, rare diseases, common complex diseases, industry

Genome of Europe (GoE)

Design and testing

Maturity Model

ELSI recommendations and toolkits

Technical recommendations and guidelines

1+MG dashboard of genomic data sets

1+MG trust framework

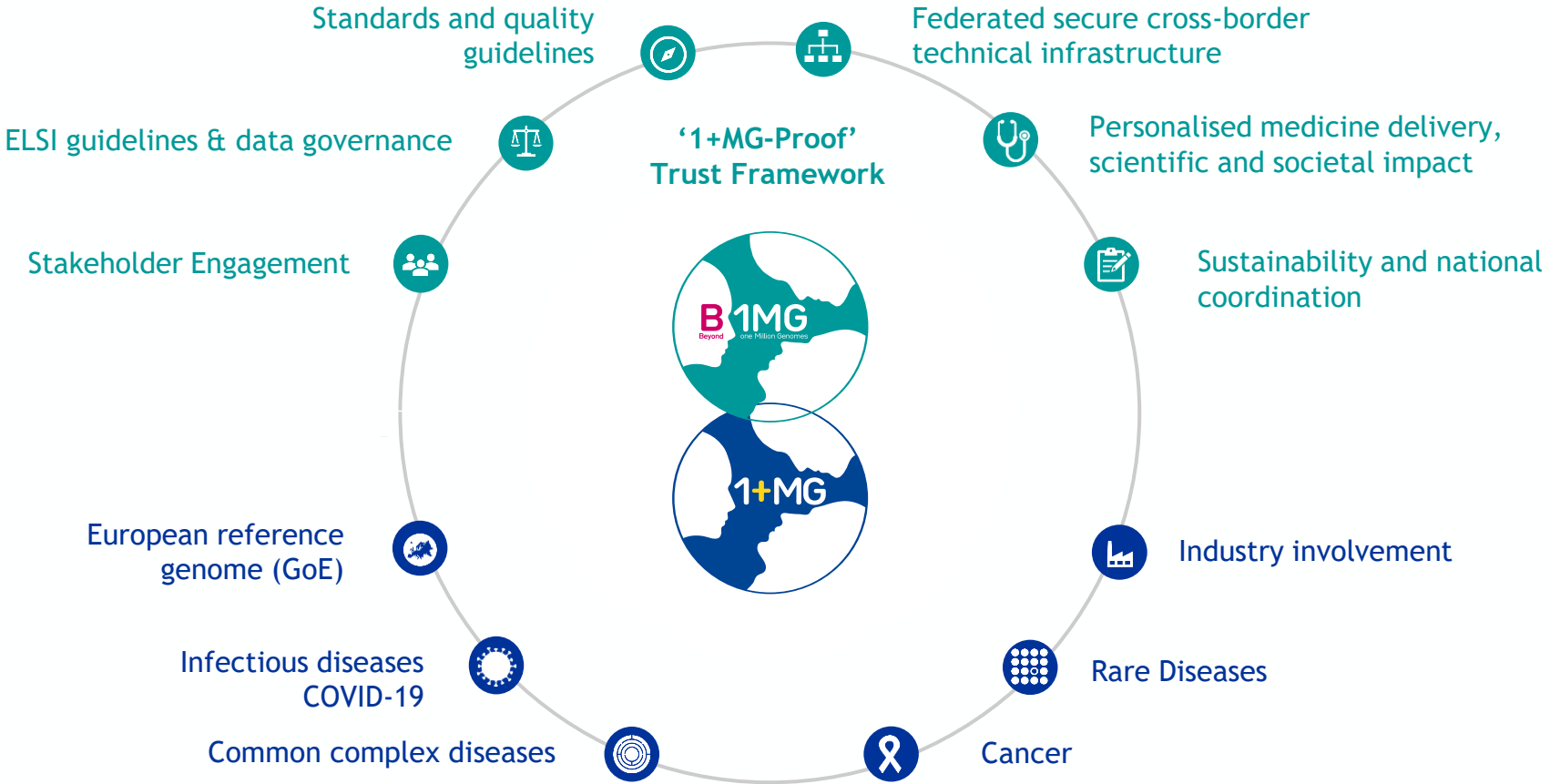
- **ELSI**
- **Data and Quality**
- **Infrastructure**
- **Maturity model**

Scale up and sustainability

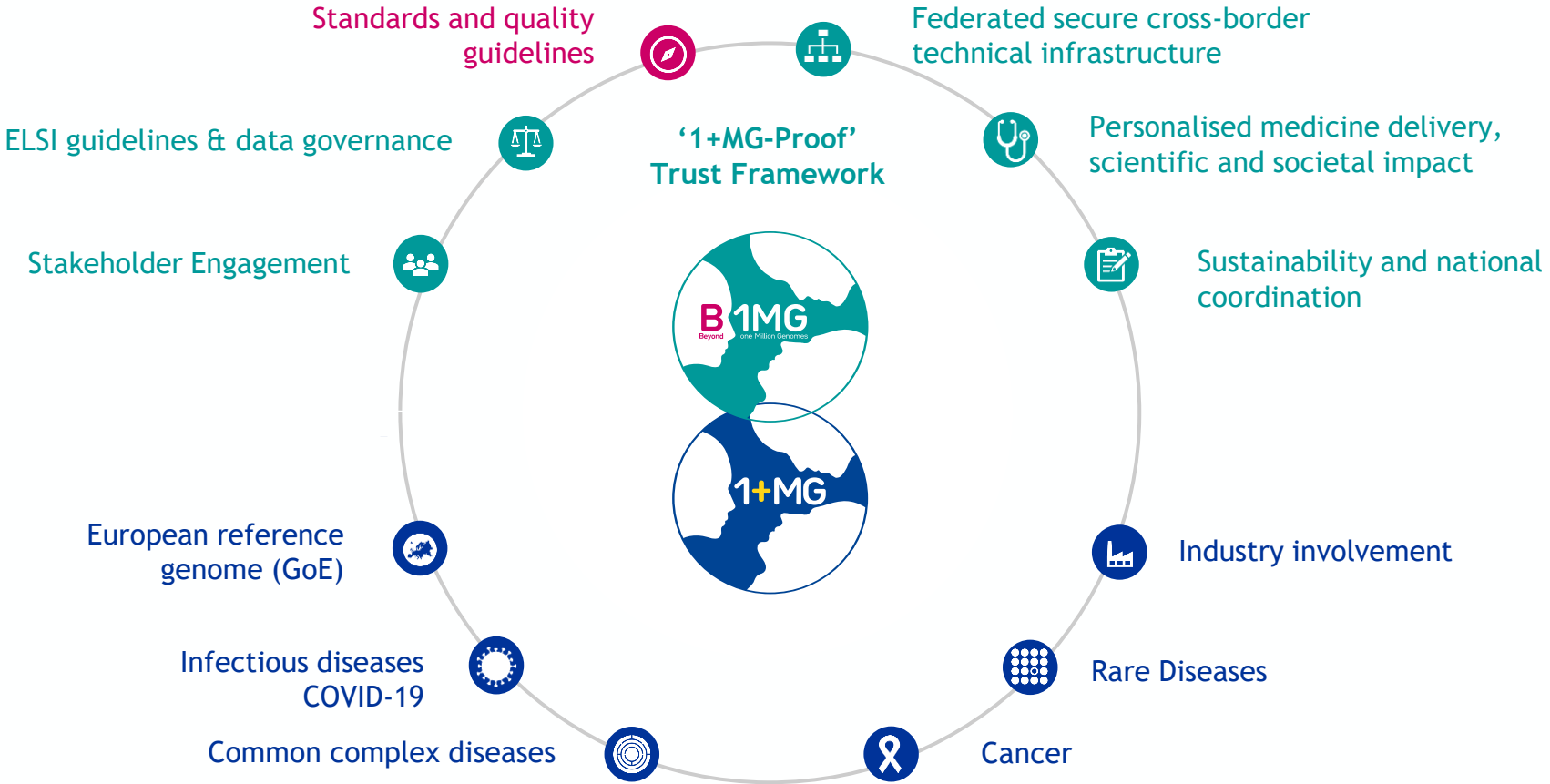
Sustainable cross-border access to genomic health data

Genomic Data Infrastructure sustained among

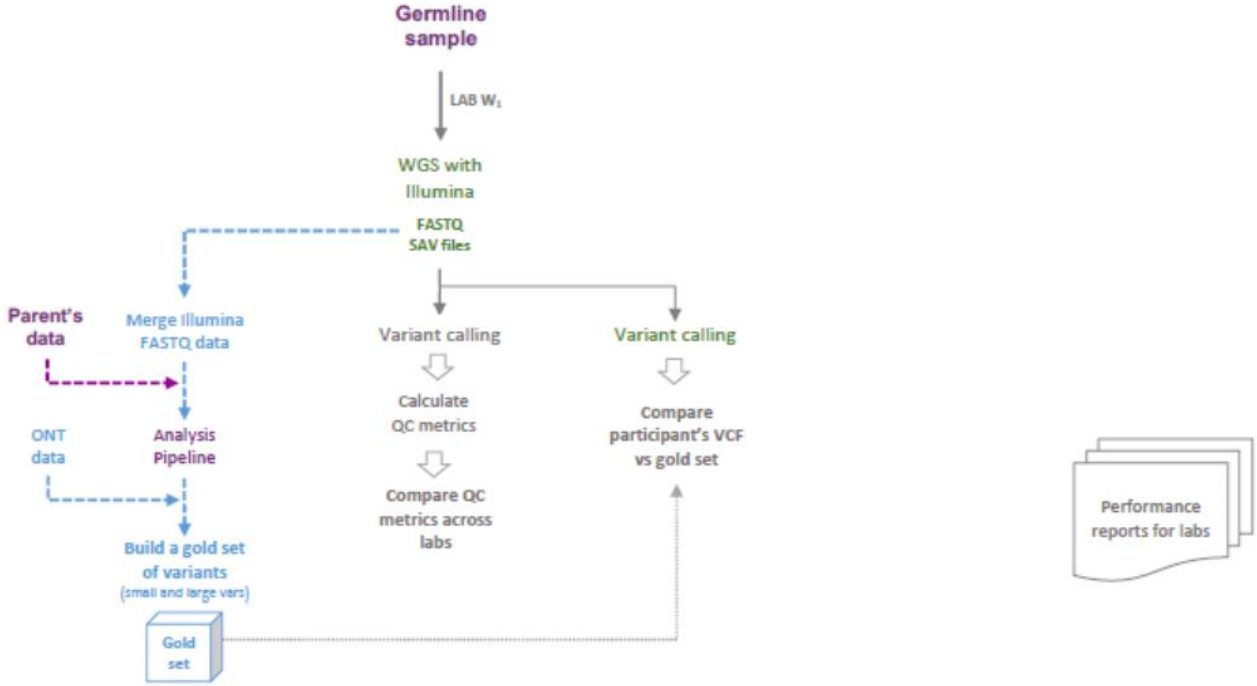
- European Health Data Space, European Open Science Cloud, Digital Europe
- National infrastructures & genome-based health programmes


Standards and quality guidelines
Federated secure cross-border technical infrastructure
ELSI guidelines & data governance
‘1+MG-Proof’ Trust Framework
Personalised medicine delivery, scientific and societal impact
Stakeholder Engagement
Sustainability and national coordination
B 1MG Beyond one Million Genomes
1+MG
European reference genome (GoE)
Industry involvement
Infectious diseases COVID-19
Rare Diseases
Common complex diseases
Cancer

Stakeholder engagement


Standards and quality guidelines
Federated secure cross-border technical infrastructure
ELSI guidelines & data governance
‘1+MG-Proof’ Trust Framework
Personalised medicine delivery, scientific and societal impact
Stakeholder Engagement
Sustainability and national coordination
B 1MG Beyond one Million Genomes
1+MG
European reference genome (GoE)
Industry involvement
Infectious diseases COVID-19
Rare Diseases
Common complex diseases
Cancer

CNAG's germline plLC workflow



WP3 Quality Metrics for Sequencing

cnag

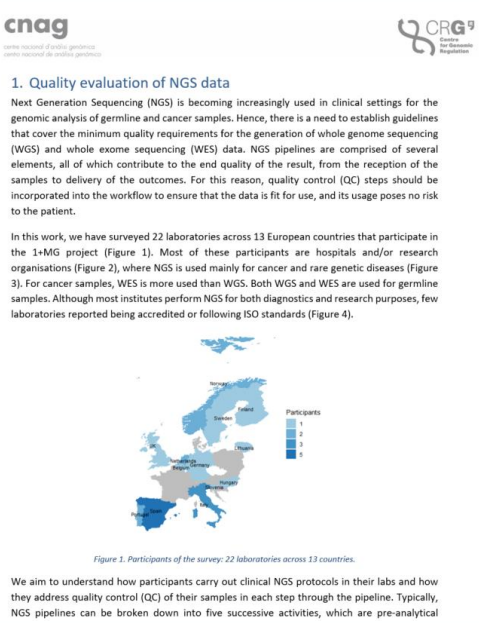
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG Centre for Genomic Regulation

Quality metrics for sequencing

B1MG – Deliverable 3.1

January 2021



cnag

centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

CRG Centre for Genomic Regulation

1. Quality evaluation of NGS data

Next Generation Sequencing (NGS) is becoming increasingly used in clinical settings for the genomic analysis of germline and cancer samples. Hence, there is a need to establish guidelines that cover the minimum quality requirements for the generation of whole genome sequencing (WGS) and whole exome sequencing (WES) data. NGS pipelines are comprised of several elements, all of which contribute to the end quality of the result, from the reception of the samples to delivery of the outcomes. For this reason, quality control (QC) steps should be incorporated into the workflow to ensure that the data is fit for use, and its usage poses no risk to the patient.

In this work, we have surveyed 22 laboratories across 13 European countries that participate in the 1+MG project (Figure 1). Most of these participants are hospitals and/or research organisations (Figure 2), where NGS is used mainly for cancer and rare genetic diseases (Figure 3). For cancer samples, WES is more used than WGS. Both WGS and WES are used for germline samples. Although most institutes perform NGS for both diagnostics and research purposes, few laboratories reported being accredited or following ISO standards (Figure 4).

Figure 1. Participants of the survey: 22 laboratories across 13 countries.

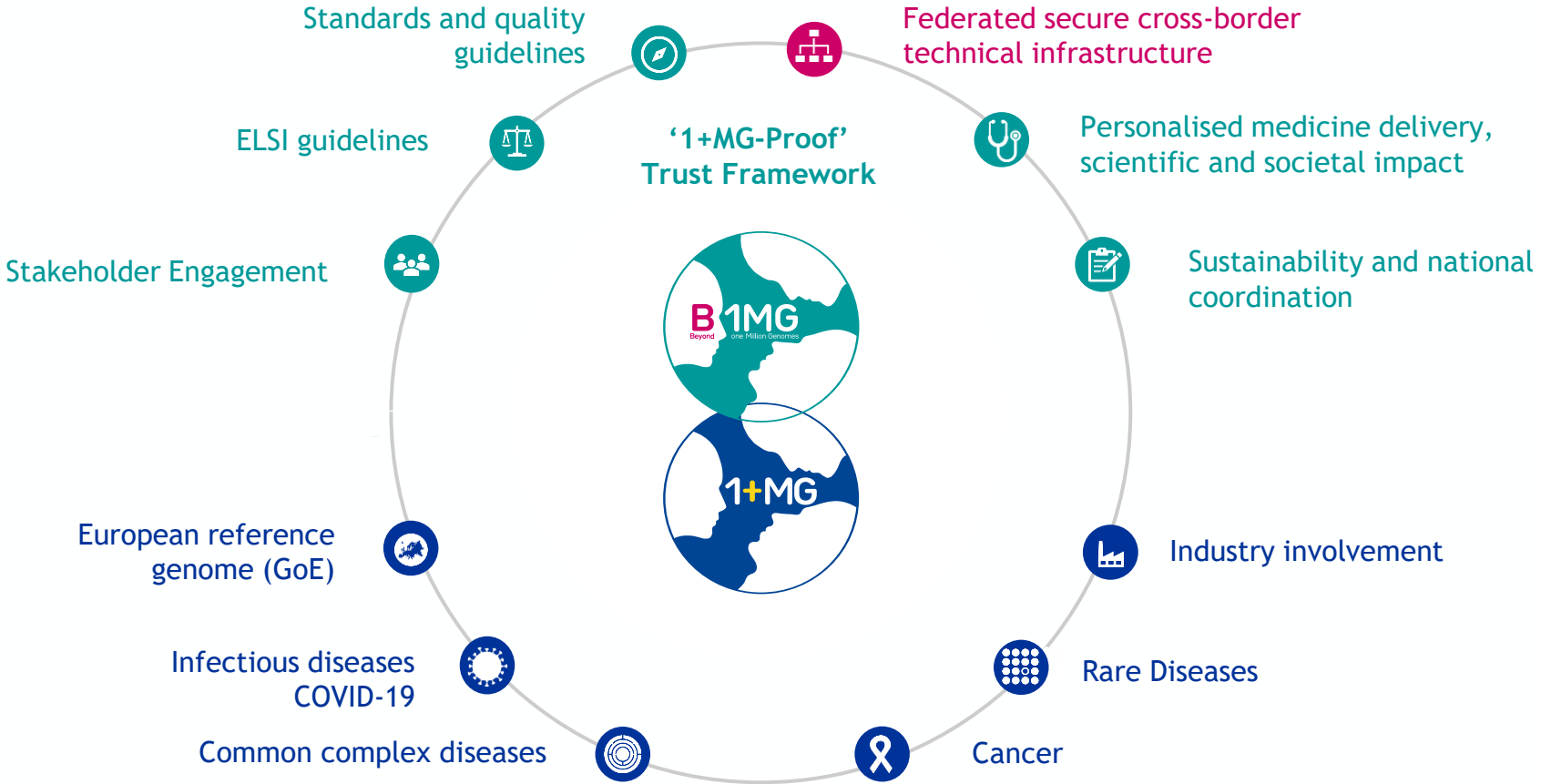
We aim to understand how participants carry out clinical NGS protocols in their labs and how they address quality control (QC) of their samples in each step through the pipeline. Typically, NGS pipelines can be broken down into five successive activities, which are pre-analytical

Technical document for specialists

Living Document

Vendor agnostic recommendations

Next update after the conclusion of the benchmark and the ILC


Standards and quality guidelines
Federated secure cross-border technical infrastructure
ELSI guidelines
‘1+MG-Proof’ Trust Framework
Personalised medicine delivery, scientific and societal impact
Stakeholder Engagement
Sustainability and national coordination
B 1MG
Beyond one Million Genomes
1+MG
European reference genome (GoE)
Industry involvement
Infectious diseases COVID-19
Rare Diseases
Common complex diseases
Cancer

Adapted from Bruno Dallapiccola
"Use Case Needs WG8 – Rare diseases"
B1MG Synthetic Genomes WS – Dec 7th 2020

Rare Disease Use Case

Exemplary scenarios tackled by WES/WGS:

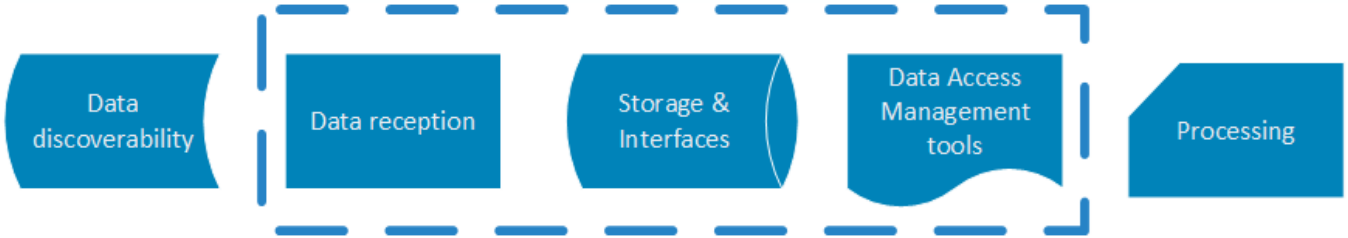
1. Undiagnosed patients waiting for clinical and molecular classification.

2. Patients affected by a known genetic disorder not solved by the disease genes' panel analysis (genetic heterogeneity).

3. Patients affected by a known likely genetic disorder awaiting for the identification of the molecular make-up.

1+MG Proof of Concept objectives

- Define a set of standards, services, and components that can support the five 1+MG Infrastructure functionalities and demonstrate these in action for one of the WG use cases – in this case Rare Disease (WG8)
- Demonstrate the use case from the viewpoint of 2 actors:
  1. Researcher Clinician
  2. Data Access Committee
- All data within the PoC is synthetic data based on open-access 1000 Genomes data



Synthetic Data

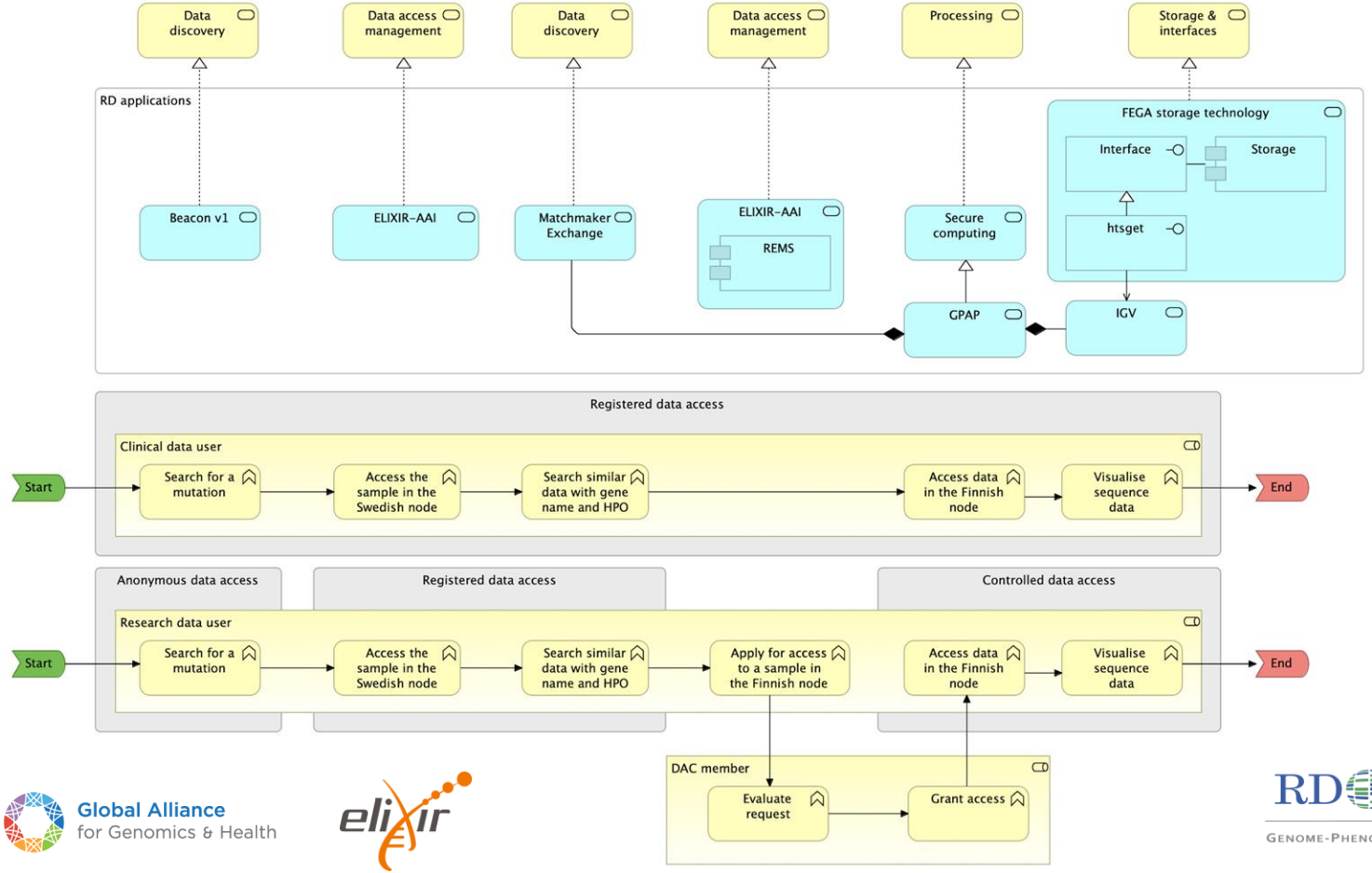
- Rare Disease dataset generated based on known deleterious variants and associated disease phenotypes

- 6 Trios:
  1. Congenital myasthenic syndrome
  2. Macular dystrophy
  3. Muscular dystrophy

  1. Mitochondrial disorder
  2. Breast cancer
  3. Congenital myasthenic syndrome

- Each trio has:
  - Clinical and phenotypic information (ORDO, HPO, OMIM)
  - Pedigree (PED)
  - Files: Phenopackets, pair of FASTQs, BAM and index, 25 chromosome gVCF files plus indexes per individual

PoC Workflow

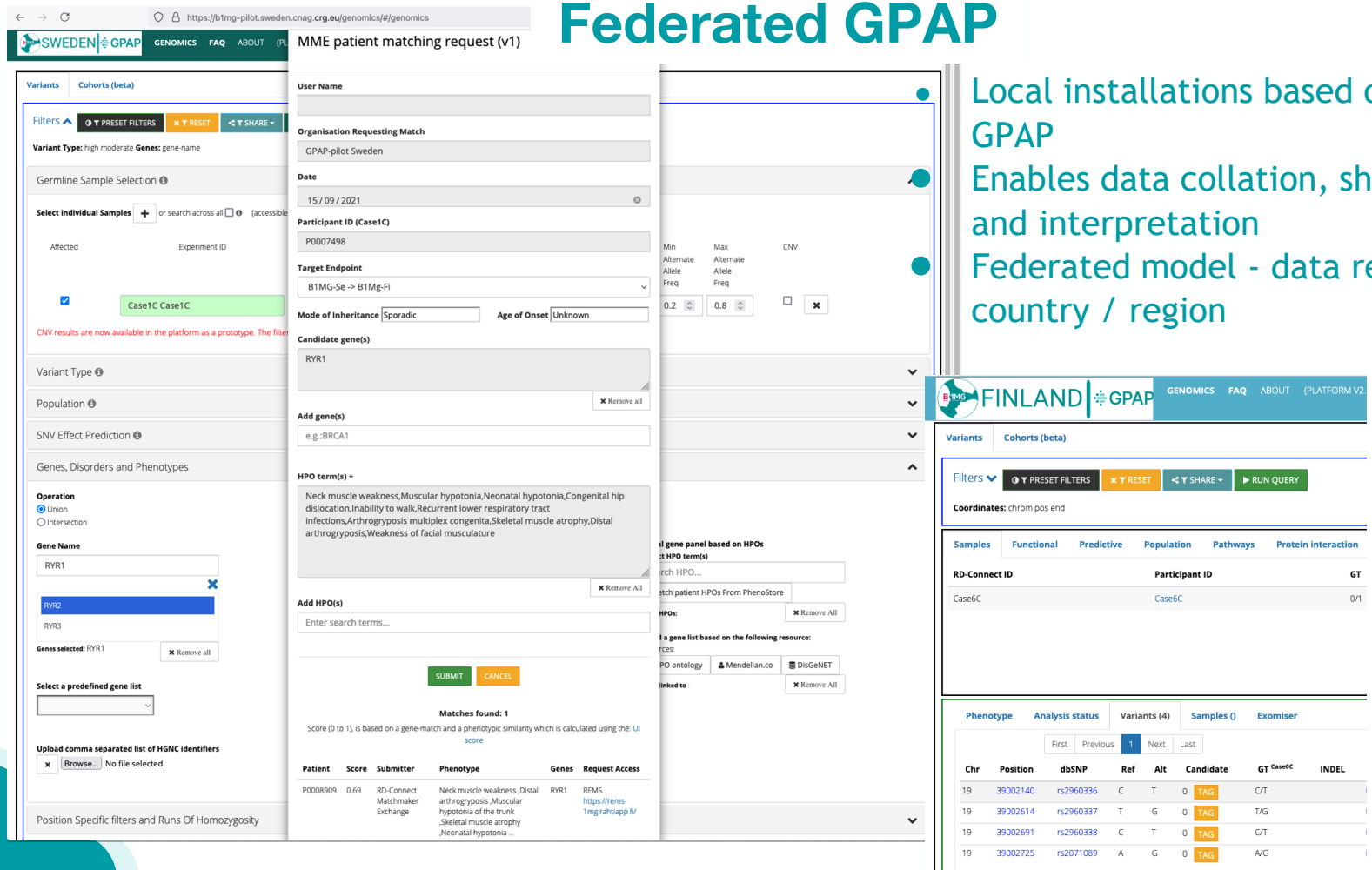


Global Alliance for Genomics & Health

elixir

RD Connect
GENOME-PHENOME ANALYSIS PLATFORM

Federated GPAP



Local installations based on RD-Connect GPAP

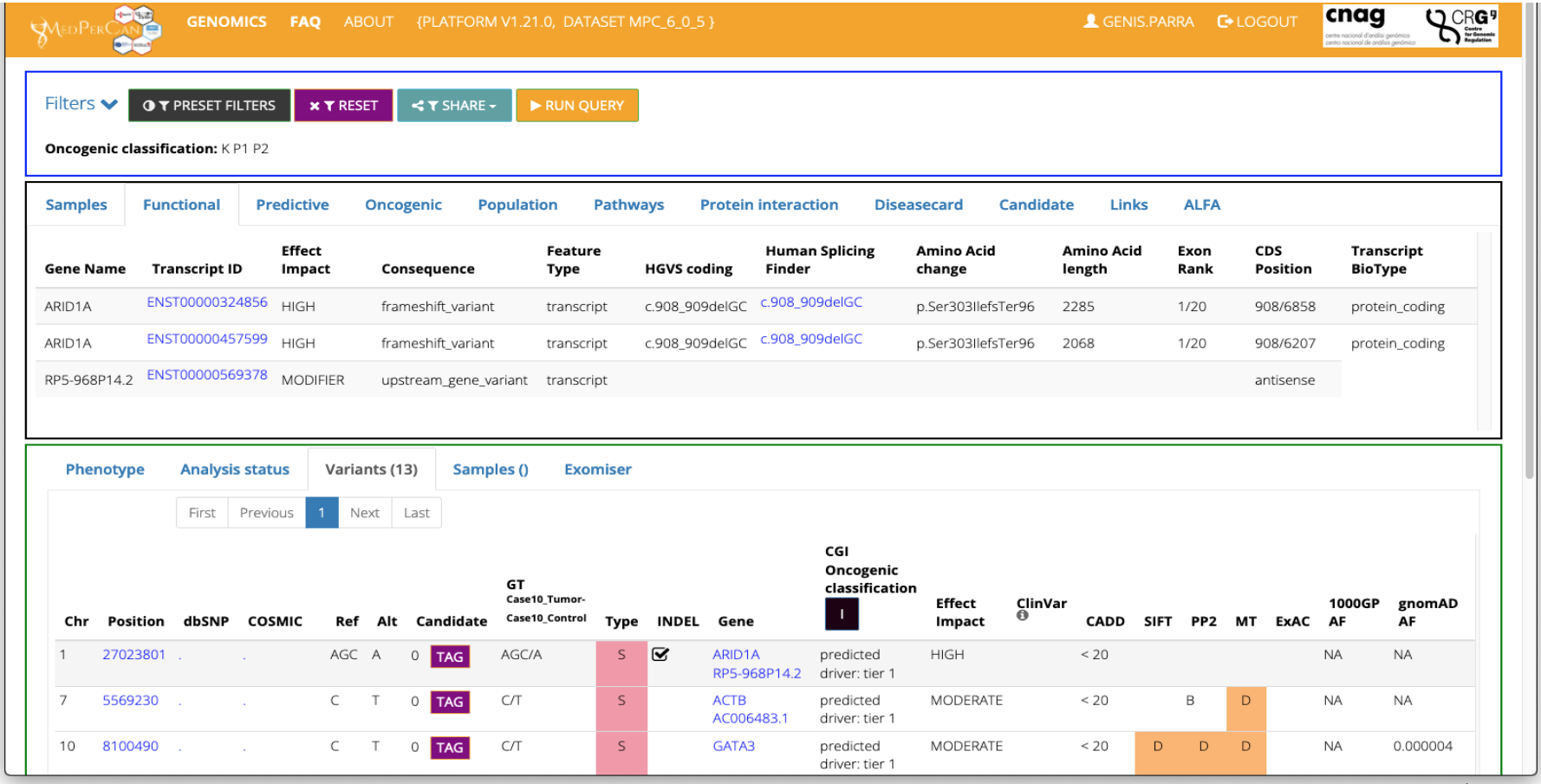
Enables data collation, sharing, analysis, and interpretation

Federated model - data remains in the country / region

WP4: Infrastructure

- Proof of Concept video shared:
  - https://bit.ly/3jd22MA
  - And associated presentation: https://bit.ly/3aSy0sQ
  - Feedback ongoing

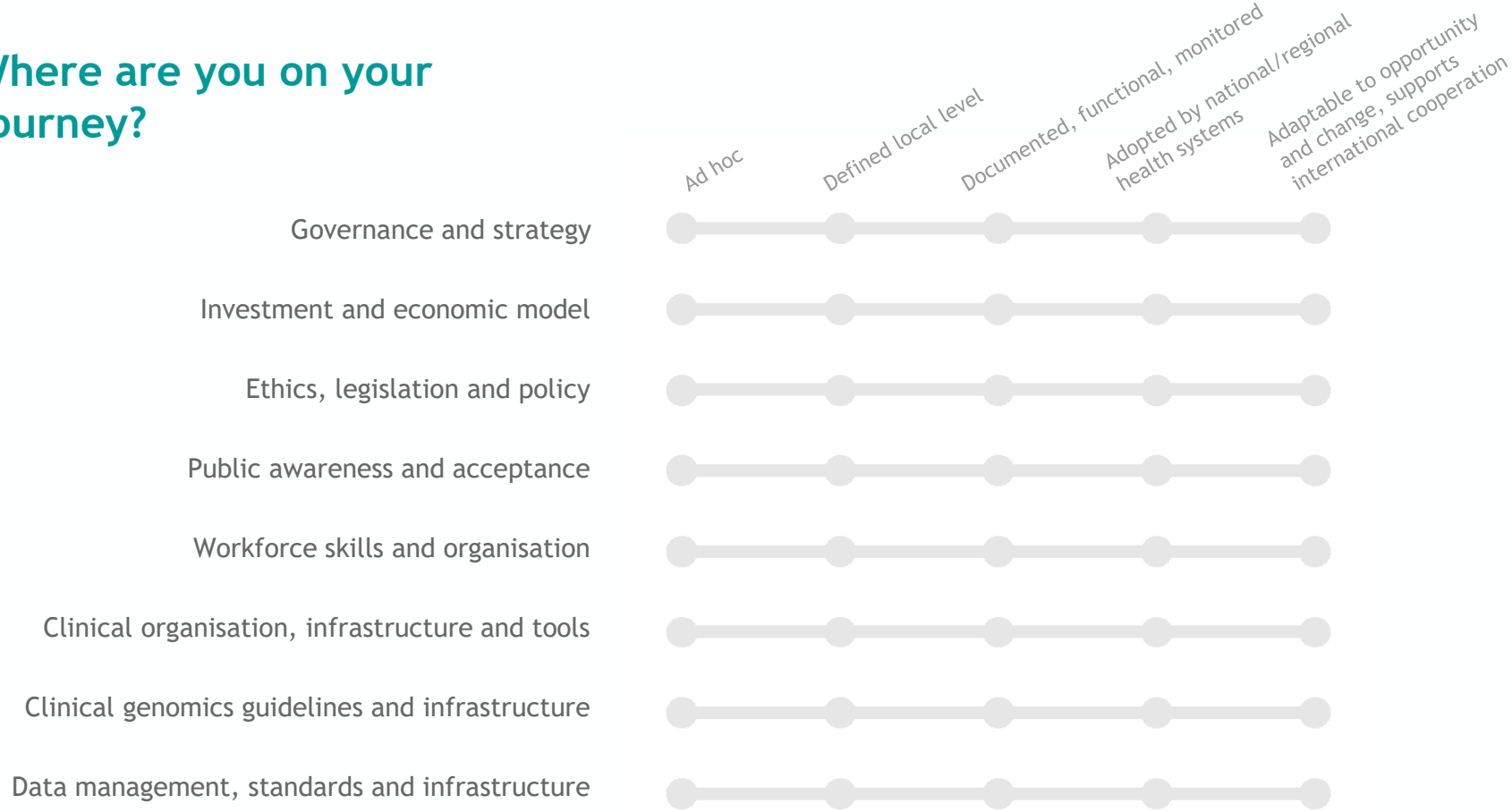
Cancer GPAP - User Interface



Gene Name	Transcript ID	Effect Impact	Consequence	Feature Type	HGVS coding	Human Splicing Finder	Amino Acid change	Amino Acid length	Exon Rank	CDS Position	Transcript BioType
ARID1A	ENST00000324856	HIGH	frameshift_variant	transcript	c.908_909delGC	c.908_909delGC	p.Ser303IlefsTer96	2285	1/20	908/6858	protein_coding
ARID1A	ENST00000457599	HIGH	frameshift_variant	transcript	c.908_909delGC	c.908_909delGC	p.Ser303IlefsTer96	2068	1/20	908/6207	protein_coding
RP5-968P14.2	ENST00000569378	MODIFIER	upstream_gene_variant	transcript						antisense	

Chr	Position	dbSNP	COSMIC	Ref	Alt	Candidate	GT Case10_Tumor-Case10_Control	Type	INDEL	Gene	CGI Oncogenic classification I	Effect Impact	ClinVar	CADD	SIFT	PP2	MT	ExAC	1000GP AF	gnomAD AF
1	27023801	.	.	AGC	A	0 TAG	AGC/A	S	☑	ARID1A RP5-968P14.2	predicted driver: tier 1	HIGH		< 20					NA	NA
7	5569230	.	.	C	T	0 TAG	C/T	S		ACTB AC006483.1	predicted driver: tier 1	MODERATE		< 20		B	D		NA	NA
10	8100490	.	.	C	T	0 TAG	C/T	S		GATA3	predicted driver: tier 1	MODERATE		< 20	D	D	D		NA	0.000004

Where are you on your journey?

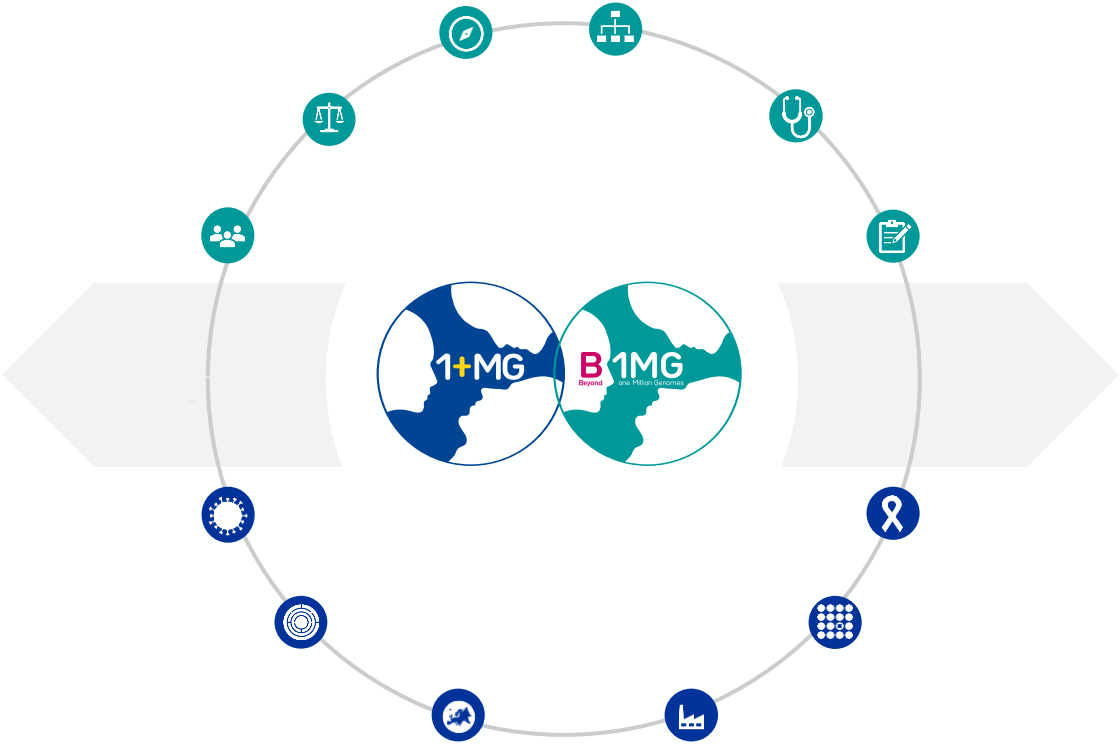


Advantages of collaboration within Europe

- Early alignment and discussion prevents silos/duplication of efforts
- 1+MG trust framework - agreed upon recommendations and guidelines across key domains:
  - ✔ ELSI
  - ✔ Data Standards
  - ✔ Data Quality
  - ✔ Technical infrastructure
- Capacity building
  - Country visits
  - B1MG Maturity Level Model

Coupled with a sustainable, long-term initiative

Personalised medicine data can be accessed **across Europe successfully**



1+MG trust framework

B1MG has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 951724