The 1+Million Genome

Ivo Glynne Gut, CNAG-CRG, Barcelona

1/12/2021
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
Stakeholder Engagement

ELSI guidelines & data governance

Standards and quality guidelines

Federated secure cross-border technical infrastructure

Personalised medicine delivery, scientific and societal impact

Sustainability and national coordination

European reference genome (GoE)

Infectious diseases COVID-19

Common complex diseases

Industry involvement

Rare Diseases

Cancer

‘1+MG-Proof’ Trust Framework
Stakeholder engagement

Citizens engagement and public trust
Drive engagement of patients and citizens in B1MG and NMGs to build public trust in genomic-based health strategies.

Partner Projects + Working Groups
Selection of specific partners to find synergies and offer different capabilities alongside stakeholder experts involved in the 1+MG Working Groups.

Stakeholder Forum
Wider stakeholder community of the 1+MG initiative facilitated by B1MG (WP1) via Stakeholder Portal (WP6).
Stakeholder Engagement

ELSI guidelines & data governance

Standards and quality guidelines

Federated secure cross-border technical infrastructure

Personalised medicine delivery, scientific and societal impact

Sustainability and national coordination

‘1+MG-Proof’ Trust Framework

European reference genome (GoE)

Infectious diseases COVID-19

Common complex diseases

Industry involvement

Rare Diseases

Cancer
CNAG’s germline plLC workflow

Germline sample

LAB W

WGS with Illumina

FASTQ files

Parent's data

Merge Illumina FASTQ data

ONT data

Analysis Pipeline

Build a gold set of variants (small and large vars)

Gold set

Variant calling

Calculate QC metrics

Compare QC metrics across labs

Variant calling

Compare participant's VCF vs gold set

Performance reports for labs
WP3 Quality Metrics for Sequencing

Technical document for specialists
Living Document
Vendor agnostic recommendations
Next update after the conclusion of the benchmark and the ILC
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
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:
  • And associated presentation: https://bit.ly/3aSy0sQ
  • Feedback ongoing
# Cancer GPAP - User Interface

![User Interface Screenshot]

## Gene Name: ARID1A
- **Transcript ID:** ENST00000324856
- **Effect Impact:** HIGH
- **Consequence:** frameshift_variant
- **Feature Type:** transcript
- **HGVS Coding:** c.908_909delGC
- **Human Splicing Finder:**
- **Amino Acid Change:** p.Ser303IlefsTer96
- **Amino Acid Length:** 2285
- **Exon Rank:** 1/20
- **CDS Position:** 908/6858
- **Transcript BioType:** protein_coding

## Gene Name: ARID1A
- **Transcript ID:** ENST00000457599
- **Effect Impact:** HIGH
- **Consequence:** frameshift_variant
- **Feature Type:** transcript
- **HGVS Coding:** c.908_909delGC
- **Human Splicing Finder:**
- **Amino Acid Change:** p.Ser303IlefsTer96
- **Amino Acid Length:** 2068
- **Exon Rank:** 1/20
- **CDS Position:** 908/6207
- **Transcript BioType:** protein_coding

## Gene Name: RPS9-968P14.2
- **Transcript ID:** ENST00000569378
- **Effect Impact:** MODIFIER
- **Consequence:** upstream_gene_variant
- **Feature Type:** transcript
- **HGVS Coding:**
- **Human Splicing Finder:**
- **Amino Acid Change:**
- **Amino Acid Length:**
- **Exon Rank:**
- **CDS Position:**
- **Transcript BioType:** antisense

## Variants List

<table>
<thead>
<tr>
<th>Chr</th>
<th>Position</th>
<th>dbSNP</th>
<th>COSMIC</th>
<th>Ref</th>
<th>Alt</th>
<th>Candidate</th>
<th>Type</th>
<th>INDEL</th>
<th>Gene</th>
<th>Effect Impact</th>
<th>ClinVar</th>
<th>CADD</th>
<th>SIFT</th>
<th>PP2</th>
<th>MT</th>
<th>ExAC</th>
<th>1000GP</th>
<th>gnomAD AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27023801</td>
<td>.</td>
<td>.</td>
<td>AGC</td>
<td>A</td>
<td>0</td>
<td>TAG</td>
<td>AGC/A</td>
<td>S</td>
<td>ARID1A RP5-968P14.2</td>
<td>predicted driver: tier 1</td>
<td>HIGH</td>
<td>&lt; 20</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0.000004</td>
</tr>
<tr>
<td>7</td>
<td>5569230</td>
<td>.</td>
<td>.</td>
<td>C</td>
<td>T</td>
<td>0</td>
<td>TAG</td>
<td>C/T</td>
<td>S</td>
<td>ACTB AC006483.1</td>
<td>predicted driver: tier 1</td>
<td>MODERATE</td>
<td>&lt; 20</td>
<td>B</td>
<td>D</td>
<td>MT</td>
<td>ExAC</td>
<td>1000GP</td>
</tr>
<tr>
<td>10</td>
<td>8100490</td>
<td>.</td>
<td>.</td>
<td>C</td>
<td>T</td>
<td>0</td>
<td>TAG</td>
<td>C/T</td>
<td>S</td>
<td>GATA3</td>
<td>predicted driver: tier 1</td>
<td>MODERATE</td>
<td>&lt; 20</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>ExAC</td>
<td>1000GP</td>
</tr>
</tbody>
</table>
Where are you on your journey?

Governance and strategy
Investment and economic model
Ethics, legislation and policy
Public awareness and acceptance
Workforce skills and organisation
Clinical organisation, infrastructure and tools
Clinical genomics guidelines and infrastructure
Data management, standards and infrastructure
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 through the 1+MG trust framework.
B1MG has received funding from the European Union’s Horizon 2020 Research and Innovation programme under grant agreement No 951724