





## Why a declaration of cooperation and what is the scope

- Coordinating secure access to data from >1M genomes that are linked to health data, as well as pooling analytical capabilities, in compliance with the General Data Protection Regulation, is crucial to advance the understanding of genetic associations that cause or predispose diseases measures and to facilitate further development of personalised medicine.
- Citizens, researchers and health systems in the Union can benefit from the full potential of genomics
  to advance targeted health care interventions leading to better prevention, early diagnosis and
  treatment of diseases;
- Secure and authorised cross-border access to genomic and other health data in the Union will enable targeted research and innovation as well as efficient translation of that research into clinical settings and public health work, which can lead to more effective therapies for individual patients and improved preventive measures.
- Citizens needs are at the centre of data-driven healthcare ...;

Member States of the European Union, the European Economic Area (EEA) and the European Free Trade Association (EFTA)

2



### Aim of the initiative – MS commitment

- This cooperation, which will build upon existing initiatives in genomics and personalised medicine, will inter alia aim to:
- Define a voluntary coordination mechanism of national, regional and local public authorities to link ongoing genomic medicine initiatives and to steer the activities stemming from this declaration
- Ensure distributed, authorised and secure access to national and regional banks of genetic and other relevant data for the advancement of science and innovation, while taking appropriate measures to protect the privacy of individual data donors;
- Support the development of technical specifications for secure access and cross-border exchange
  of genomic datasets and facilitate interoperability of relevant registries and databases to support
  research; Promote the use of open standards and data management systems to ensure
  interoperability of genomic and other health data...;
- Define a governance model of cooperation, particularly concerning the terms and conditions for distributed access to genomic data

Vetenskapsrådet Malin Eklund 6





## Aim of the initiative – Signatories invite Commission to:

- Consider the shared vision enshrined in this declaration and endorsed by the signatories as it
  implements measures to support the digital transformation of health and care further to the Digital
  Single Market mid-term review, notably aiming at "supporting data infrastructure to advance research,
  diseases prevention and personalised health and care";
- Mobilise funds from the Horizon 2020 and Connecting Europe Facility programmes to support pilot actions, pooling data and resources across the Union, and demonstrate the benefits in advancing genomic medicine. Consider further support from future programmes;
- Regularly report on progress achieved regarding the implementation of the commitments outlined above, including updates on global developments concerning genomic medicine.



- Citizen's needs at the centre of data-driven innovation as active agents in their own health journey;
- Citizens can benefit from more precise and personalised treatments as well as a more participatory healthcare experience
- Researchers and health systems can benefit from the full potential of genomics to advance targeted healthcare interventions
- The union stays at the forefront of genomic and personalised medicine, and improves its scientific capabilities and industrial competitiveness



## Declaration on linking genomic databases across borders: "Towards access to 1 million Genomes in the EU by 2022"

- Signed on 10 April 2018 during Digital Day 2 by:
- Czech Republic, Cyprus, Estonia, Finland, Italy, Lithuania, Luxembourg, Malta, Portugal, Slovenia, Spain, Sweden and the UK
- A member states driven initiative supported by the European Commission





## Digital Health and Care







TRANSFORMATION OF HEALTH AND CARE IN THE DIGITAL SINGLE MARKET - Harnessing the potential of data to empower citizens and build a healthier society.

#### European health challenges

- Ageing population and chronic diseases putting pressure on health budgets
- Unequal quality and access to healthcare services
- Shortage of health professionals

#### Potential of digital applications and data to improve health

- Efficient and integrated healthcare systems
- Personalised health research, diagnosis and treatment
- Prevention and citizen-centred health services

#### What EU citizens expect...



To access their own health data (requiring interoperable and quality health data)



To share their health data (if privacy and security are ensured)



To provide feedback on quality of treatments

#### Support European Commission:



Secure access and exchange of health data





#### Ambition:

Citizens can securely access and share (e.g. with doctors or pharmacies) their health data anywhere in the EU.

- eHealth Digital Service Infrastructure will deliver initial cross-border services (patient summaries and ePrescriptions) and cooperation between participating countries will be strengthened.
- Proposals to extend scope of eHealth cross-border services to additional cases, e.g. full electronic health records.
- Recommended exchange format for interoperability of existing electronic health records in Europe.



Health data pooled for research and personalised medicine



#### Ambition:

Shared health resources (data, infrastructure, expertise...) allowing targeted and faster research, diagnosis and treatment.

#### Actions:

- Voluntary collaboration mechanisms for health research and clinical practice (starting with "one million genomes by 2022" target).
- Specifications for secure access and exchange of health data.
- Pilot actions on rare diseases, infectious diseases and impact data.



Digital tools and data for citizen empowerment and person-centred healthcare



#### Ambition:

Citizens can monitor their health, adapt their lifestyle and interact with their doctors and carers (receiving and providing feedback).

#### Actions:

- Facilitate supply of innovative digital-based solutions for health, also by SMEs, with common principles and certification. - Support demand uptake of innovative digital-based
- solutions for health, notably by healthcare authorities and providers, with exchange of practices and technical assistance.
- Mobilise more efficiently public funding for innovative digital-based solutions for health, including EU funding.







## Declaration on linking genomic databases across borders: "Towards access to 1 million Genomes in the EU by 2022"

EU countries agreed to cooperate in linking genomic data across borders



#### Signed by 20 EU Member States:

Austria, Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Finland, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Slovenia, Spain, Sweden and the UK.

**9 Observers**: Belgium, Denmark, France, Germany, Hungary, Ireland, Norway, Poland and Switzerland.









### The 1+ Million Genome initiative

- Federated framework that would allow secure and authorised cross-border access to genomic and other health data across the EU, supporting research, health care and prevention.
- To allow users to search and access the data through a userfriendly and effective data governance structure building on existing national and European initiatives.
- To ensure that citizens, researchers and health systems in Europe can benefit from the full potential of genomics to advance targeted health care interventions leading to better prevention, early diagnosis and treatment of diseases

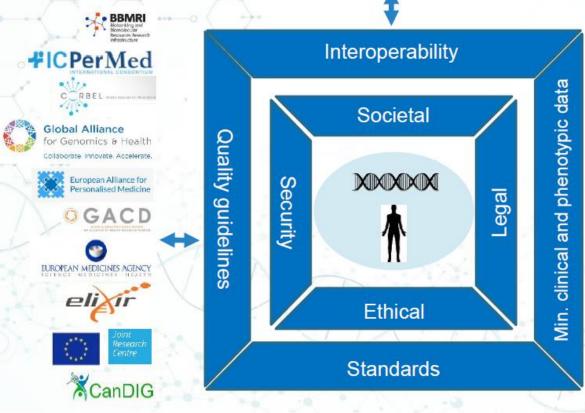
Member States of the European Union, the European Economic Area (EEA) and the European Free Trade Association (EFTA)





## Join forces!

















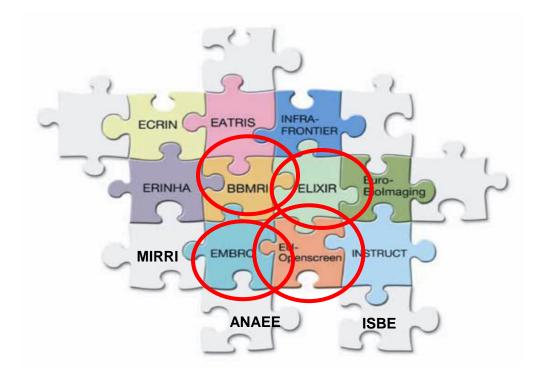






## **European Research Infrastructure Consortium (ERIC)**

- ERIC = European Research Infrastructure Consortium
- Aimed at meeting long-term needs of Europe's research communities across scientific areas.
- ESFRI are facilities, resources or services of a unique nature to conduct toplevel research activities in all fields.







BBMRI-ERIC is a research infrastructure of biobanks and biomolecular resources that provides expertise and services in order to facilitate the use of European sample collections and data for the benefit of human health.

#### **WE ARE** MEMBERS OF BBMRI-ERIC Austria Belgium **OBSERVERS OF** Bulgaria BBMRI-ERIC Czech Republic Switzerland Estonia Finland Cyprus Turkey France IARC Germany Greece Italy Latvia Malta Netherlands Norway Poland Sweden United Kingdom

ELIXIR's mission is to operate a sustainable European infrastructure for biological information, supporting life-science research and its translation to society, the bio-industries, environment



ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data







@ELIXIREurope



/company/elixir-europe





### Genomics-based National Initiative projects across the ELIXIR



- A-C-G-T Analysis of Czech Genomes for Theranostics (2018-ongoing)
- FarGen (Denmark) (2011-ongoing)
- France Médecine Génomique 2025 (2016-ongoing)
- Genome Denmark (2012 ongoing)
- Genome of the Netherlands (GoNL) / BBMRI-NL biobank 'omics' studies BIOS (2010-ongoing)
- Genomics England (2013-ongoing)
- FinnGen and the Sequencing Initiative Suomi, Finland (2015-ongoing)
- Eesti biopangas: Estonian Genome Project (2000-ongoing)
- The Scottish Genomes Partnership (2015 ongoing)
- 10. UK Biobank (2006 ongoing)

Public funding

Public private funding

- National Centre for Excellence in Research in Parkinson's Disease (Luxembourg) (2015-ongoing)
- National Center for Medical Genomics

   Czech national research infrastructure
   (2014 ongoing)
- Genomic Medicine Sweden (2018- ongoing)
- National contact point and network for rare diseases in Slovenia (2016-ongoing)
- Swiss Personalised Health Network (SPHN) (2017-ongoing)
- Welsh Genomics for Precision Medicine Strategy (2017-ongoing)
- Northern Ireland Genomic Medicine Centre (2017-pagoing)
- National Bionics Program, Hungary, (2018 ongoing)
- National Oncology Program, Hungary, (2015 ongoing)





#### The European Genome-phenome Archive (EGA; <a href="https://ega-archive.org/">https://ega-archive.org/</a>)

- The EGA is a resource for the permanent archival and sharing of controlledaccessgenetic and phenotypic human data, resulting from biomedical research projects. The central EGA (operated from EMBL-EBI, UK and CRG, Spain) hosts over 1,700 studies consisting of more than 4,000 datasets from in excess of 900 data providers
- The EGA includes important reference data collections for human genetics research, such as EC RD-Connect, EC BLUEPRINT, UK10K, UK Biobank, The Human Induced Pluripotent Stem Cells Initiative (HipSci), Wellcome Trust Case Control Consortium, and EU as well as nationally funded International Cancer Genome Consortium (ICGC) studies.
- The EGA is an ELIXIR Core Data Resource and is the recommended ELIXIR deposition
   Database for access controlled human data

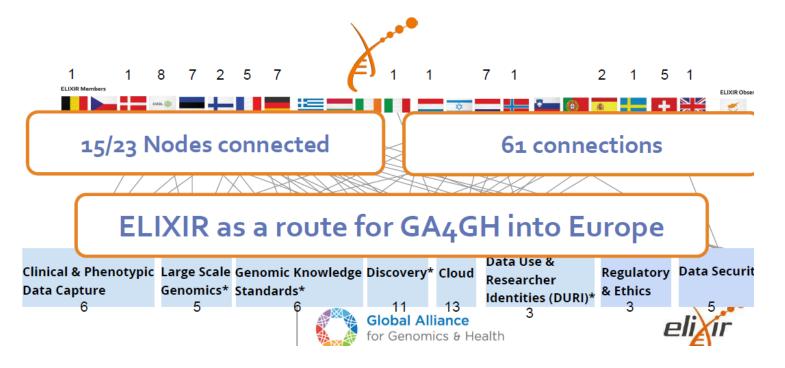




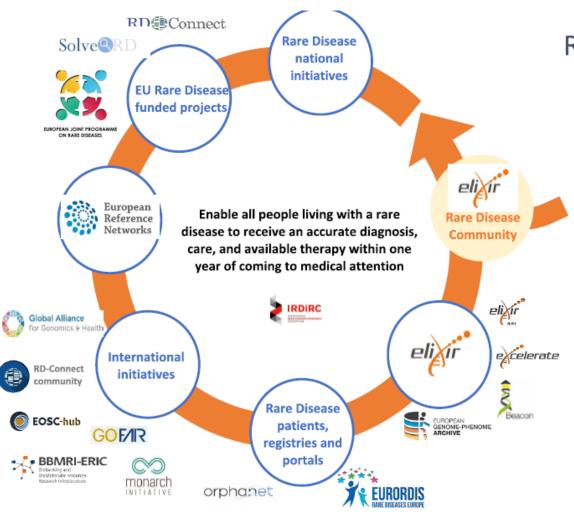


The Global Alliance for Genomics and Health aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic and health-related data sharing

#### Mapping ELIXIR::GA4GH Interactions







### Rare disease community

- Infrastructure for RD research:
  - Registry of Rare Disease data resources and analysis tools (https://rare-diseases.bio.tools)
  - Data sharing and data discovery
- Interoperability of RD resources:
  - Standards and ontologies
  - FAIR data services
- Training

### Biomedical ERICs & EU OPENSCREEN





EU-OPENSCREEN offers researchers from Europe and around the world open access to a uniquely broad range of high technologies and tools for the systematic screening of chemical substances for their biological effects. EU-OPENSCREEN integrates high-capacity screening platforms throughout Europe, which jointly use a rationally selected compound collection, comprising up to 140.000 commercial and proprietary compounds collected from European chemists.



### 2018 Questionnaire to participants

- A number of sequencing techniques are commonly used (Illumina platforms)
- Data quality and quality of samples are measured via a mix of standard-based and in-house solutions
- Standards in place for sequencing, alignment, quality control and storage
- In most cases there are standards for interpretation of DNA sequences
- Interoperability standards often deployed for data exchange
- Requirements to use specific methodologies apply to several activities (licensing of institutions, certification of professionals, external quality assessment)
- Phenotype data is collected via medical doctors, Human Phenotype Ontology (HPO) and clinical care teams
- Sequenced genomes, are linked to clinical and other types of data (e.g. proteomics and biobanks) by around 80% of respondents.
- Measuring benefits of WGS seem to still be in early phases of development
- Cost-effectiveness/sustainability studies of genomics in 50% of respondents



### **Current activities**

- 10 working groups link up to expertise/projects/initiatives identify gaps/needs/activities
- 3 use cases
- Mapping of current and future genomic initiatives and data
- Interim governance structure national mirror groups
- → Road map

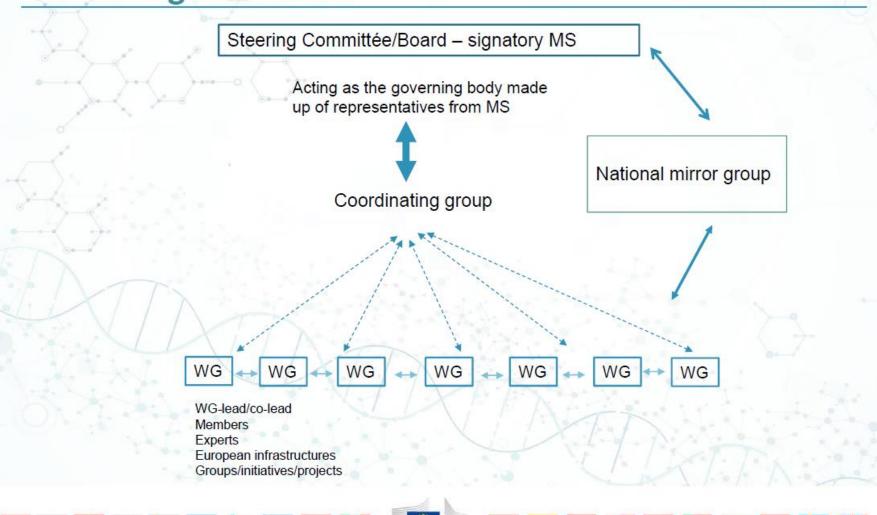








## Interim governance





## **Creation of working groups**

- WG1 Scope, stakeholders and governance
- WG2 Ethical, Legal, and Societal Issues (ELSI)
- WG3 Common standards and min. dataset for clinical and phenotypic data
- WG4 Good sequencing practice
- WG5 Federated, secure, interoperable and privacy-respecting framework and access governance
- WG6 Health economics and outcome research
- WG7 Involvement of the private sector.
- WG8 Use case Rare diseases
- WG9 Use case Cancer
- WG10 Use case Populations, Precision prevention, Pharmacogenomics





## WG1 Scope, stakeholders and governance



- Chair: Malin Eklund (Swedish Research Council, SE)
- Scope
  - Formulate goal and benefits of the initiative
  - Suggest governance structure
  - Identify opportunities for support from EU initiatives e.g. Horizon 2020, Structural Funds, initiatives for digital health
  - Suggest model of contribution from each participating MS
  - Examples/pilot studies for added value of genome sharing
  - Need for generating a collection of new data (genomic, phenotypic, exposomics)
  - Stakeholder engagement











## WG2 Ethical, Legal, and Societal Issues (ELSI) – Regina Becker (LUX)

Protecting genomic and clinical data

Harmonization of consent forms

How to store your data securely

How to anonymize your data

How to share data with all legal and ethical requirements

Application of European and local data protection regulations

















## WG3 Standards for capturing clinical and phenotypic data requirements

- Chair: Vacant
- Scope
  - Define the clinical and phenotypic data requirements necessary in order to make genetic data useful in each clinical use case
  - Mapping of available services and structures working on standards for structured clinical and phenotypic data collection
  - Provide guidelines for collection of clinical and phenotypic data
  - Bring forward best practise examples showing the added value of additional clinical information (including exposome data, life style data etc) linked to genetic data









## WG4 Good sequencing practice

- Chair: Ivo Gut (CNAG-CRG, ES)
- Scope
  - Publish guidelines for Good Genomics Practice, covering sample management, data generation and data analysis for the different use cases
  - Suggest a model for maintaining the guidelines ensure engagement of ongoing initiatives
  - Propose a meta data structure detailing the sample processing, sequencing and analysis work carried out
  - Define quality metrics for sequencing data and minimal thresholds

## WG5 Interoperability, data transfers, federated systems and data access governance

- Chair: Tommi Nyrönen / Ilkka Lappalainen (CSC, FI)
- Scope
  - Outline the infrastructure resource components needed to establish the framework that allows identification, selection, and access to the genetic data, and how these components shall interoperate
  - Access for all three key stakeholders need to be considered: healthcare, universities and commercial entities
  - · Identify and formulate benefits and rules of access for private stakeholders that will safeguard data protection rules and intellectual property rights
  - Suggest resources needed in order to establish the framework in time to make 1 million whole genomes accessible by 2022













## WG6 Health economics and outcome research

- Chair: Ilse Custers (ZonMW, NL)
- Scope
  - Based on health economic analysis:
     benefits of sharing genomic data that are linked to clinical data in order to more effectively prevent and treat diseases.
  - Benefits of longitudinal real time data
  - Contribution to sustainability challenge
  - Cost efficiency (gains in efficiency vs cost of sequencing and analysing) if possible in the context for the three use cases





## WG7 Involvement of the private sector

- Chair: vacant
- Scope
  - Examples of on-going initiatives and models for IP, contribution and access.
  - Analyse purpose, benefits and challenges and suggest a process/structure/model on how the initiative can work together with the private sector
  - Summarize ongoing similar initiatives where industry is involved. Describe the collaboration model used including IP rights











## WG8-10 "Rare diseases", "Cancer" & "Populations, Precision prevention, Pharmacogenomics"

- Chair WG8: Bruno Dallapiccola (Ospedale Bambino Gesù, IT)
- Chair WG9: Ruggero de Maria (Alleanza Contro il Cancro, IT) Dimitris Thanos (Biomedical Research Foundation Academy of Athens, GR)
- Chair WG10: Andres Metspalu (Estonian Genome Center, EE)
- Scope
  - Identify ongoing national/European pilot projects
  - Propose pilot projects / EU tool/database
  - Describe the value of European level data sharing;
  - Describe how shared data can be used within the specific disease area
  - Consider the needs of both end-users/patients, research, health care and industry including disease gene discovery, mechanism study, diagnostics, therapy, prevention and knowledge building









## Proposal on "Good Genomics Practice" (currently being discussed)

#### I. Gut&Bale

- Patient or research subjects should give consent for genomic analysis that allows for linkage to health (phenotypic) data, ideally this would also include longitudinal data linkage and re-contact and/or return of pertinent and additional looked-for findings.
- Sample acquisition should be consistent, documented and appropriate for a clinical, research or dual-use pipeline. Samples should ideally be stored in an easily retrievable facility such as a Biobank or Biosample Centre or a Biobank metwork.
- DNA extraction and preparation should be done using standardised and validated methods to ensure samples meet standards for integrity, quality, yield and volume.



- Whole genome sequencing should be carried out in an accredited facility clinical or research – with appropriate quality metrics. Ideally, the minimum standard of 30x read depth across 95% of the genome. Quality metrics should include measures of evenness of sequencing across genome, percentage of genome that reaches a minimum threshold for reliably calling variants, and the ratio of edits between paired reads.
- For tumour genomes, the read depth should be a minimum of 75X across 95% of the genome along with an appropriate quality germline sample. In addition, quality metrics for tumour genomes should include measures of callability of somatic mutations across the genome and paired-reads mapping to different chromosomes.



- WGS data should be analysed using a [standardised] validated, benchmarked and documented pipeline with – as far as possible - tools and file formats that are common and open-source;
  - WGS should be aligned to the latest reference genome standard, currently
     GRCh38 (but consideration given to cost-effective means of re-aligning certain older datasets)
  - Genome data should use formats compatible with BAM or VCF or that where possible conform to standards



- Sequence variation annotation should use standardised and documented methods.
   Where appropriate this should specify how the variants are called from the reference genome, and the method of variant annotation.
- Variant interpretation should use standardised guidelines (for both germline and somatic variation) in order to allow consistency in interpretation and data sharing.
   Specified relevant publicly available or shared variant databases (such as ClinVar) and gene-disease validity databases (such as PanelApp) should be utilised to aid interpretation. Publicly available variant databases, such as [ClinVar, HMGD plus others like the Human Variome Project



- Genome data should be stored in secure facilities. It is not necessary that raw data is made available across the EU 1 Million Genomes Initiative partnership.
- Variants together with phenotypes should be made available across the 1 M genome initiative in a federated system that allows two-sided queries, match making and integrative analyses.
- The tools and materials should be available in the languages used in national healthcare as well as opportunities use to make materials available in multiple languages – for example the approach uses in HPO.



## Spanish initiative

#### Representatives

Luis Serrano CRG Angel Carracedo FPGMX-USC

WG1 Scope, stakeholders and governance – Raquel Yotti, Gonzalo Arévalo

WG2 Ethical, Legal, and Societal Issues (ELSI) – Carmen Ayuso, Pilar Nicolas Jimenez

WG3 Common standards for capturing clinical and phenotypic data requirements –

#### Alfonso Valencia

WG4 Good Genomic Practice / development of standards for sequencing, annotation and clinical interpretation –**Ivo Gut** 

WG5 Interoperability, transfer between countries, local/federated system incl. systems development and deployment and data access governance –Ximo Dopazo

WG6 Health economics and outcome research- Alvaro Hidalgo

WG7 Involvement of the private sector: Not appointed

WG8 Use case Rare Diseases –Pablo Lapunzina

WG9 Use case Cancer – Nuria Lopez-Bigas

WG10 Use case, Common, Complex Diseases - Angel Carracedo, Marina Pollánç

National Mirror WGs being built



## Matrix Genomics State of Play

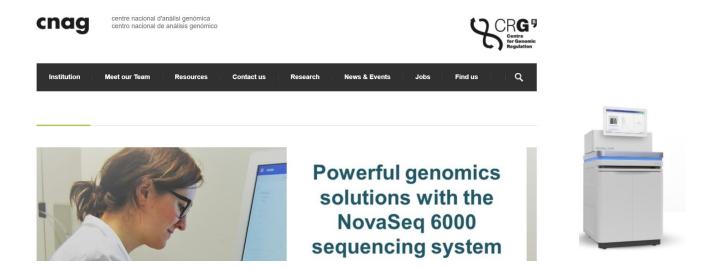


	Rare Diseases						
Name of project/initiative/infrastructure etc*.	Generated on what level: national, regional, local	Main purpose of data (research, clinical, both)	Linked clinical data (please describe)	WGS, WES or other (specify)	No of seq. available	Storage location (where those sequences are stored)**	Could be made accessible/sear chable
300 EXOMES TO ELUCIDATE RARE DISEASES - CNAG	National	Research	Brugada Syndrome	WES	30	FPGMX	
300 EXOMES TO ELUCIDATE RARE DISEASES - CNAG	National	Research	CMMR-D syndrome	WES		FPGMX	
300 EXOMES TO ELUCIDATE RARE DISEASES - CNAG	National	Research	spinocerebellar ataxia and spastic paraparesis	WES		FPGMX	
300 EXOMES TO ELUCIDATE RARE DISEASES - CNAG	National	Research	cerebral microangiopathy	WES		FPGMX	
ASC	International	both	TDAH/DI/TEA/TOC trios	WES	2.100	FPGMX	As soon as it was published
AES2016. PI1601057	National	Research	colorectal cancer	WES	120	FPGMX	As soon as it was published

				Cancer			
		Main purpose				Storage location	
Name of	Generated on what		Linked clinical			•	Could be made
project/initiative/inf	level: national,	(research,	data (please	WGS, WES or	No of seq.	,	accessible/sear
rastructure etc.*	regional, local	clinical, both)	describe)	other (specify)	available	stored)**	chable

Common, complex diseases (۲4 - Populations, Precision prevention, Pnarmacogenomics)							
						Storage location	
Name of	Generated on what	Main purpose of				(where those	Could be made
project/initiative/infr	level: national,	data (research,	Linked clinical data	WGS, WES or other		sequences are	accessible/search
astructure etc.*	regional, local	clinical, both)	(please describe)	(specify)	No of seq. available	stored)**	able





Current CNAG capacity 5.000 human genomes at 30x coverage. Planned a second NovaSeq by the end of this year which would double the capacity to 10.000.

FPGMX planning a NovaSeq by the end of this year

Intergation of the private sector (NIMGenetics ...)



#### "CONAMIGOS" PROJECT

(COHORTE NACIONAL ABIERTA MULTIPROPÓSITO DE INVESTIGACIÓN E INNOVACIÓN GLOBAL PARA EL OBJETIVO SALUD)

A highway for biomedical research in Spain

A Proposal for a Large National Cohort of Adults Residing in Spain

Timeline  o Progressive implementation (see diagram). o Follow-up (a minimum of 10 years after end of recruitment):  ✓ Continuous updates through repeated participant engage and sharing of participants' electronic health records (E  ✓ Updating health exposure information and physical	
examination every 4 years  ✓ Repeated gathering of biological samples every 4 years	
Recruitment centers  36 primary health care centers, covering all Autonomous Communities  Canting Pais Vasco Continue Pais Vasco Castille Valor Pais Vasco Castillo Valor Pais Vasco Castille Valor Pais Valor Pais Valor Pais Val	25.

Participant selection

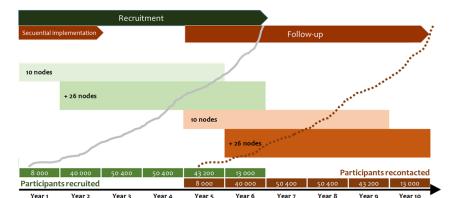
Adults (30-70 years) randomly selected from the population covered by the primary health recruitment centers + a sub-cohort of vulnerable populations (separate protocol amendments will be developed for this purpose) (Selection of the whole family? TBD: pros and cons)

8,44

Ceuta 0.08 1

Melilla 0,08 1

#### ConAmigos TIMELINE (First & second rounds)



## ciberesp isciii

\*population (in millions)

\*number of nodes

Marina Pollán **Centro Nacional de Epedimiología** 





# Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

Amit V. Khera<sup>1,2,3,4,5</sup>, Mark Chaffin<sup>4,5</sup>, Krishna G. Aragam<sup>1,2,3,4</sup>, Mary E. Haas<sup>4</sup>, Carolina Roselli<sup>4</sup>, Seung Hoan Choi<sup>4</sup>, Pradeep Natarajan<sup>2,3,4</sup>, Eric S. Lander<sup>4</sup>, Steven A. Lubitz<sup>2,3,4</sup>, Patrick T. Ellinor<sup>2,3,4</sup> and Sekar Kathiresan<sup>1,2,3,4</sup>

#### **WG10** Precision prevention

1% population analyzed for individual risk assesment for common traits

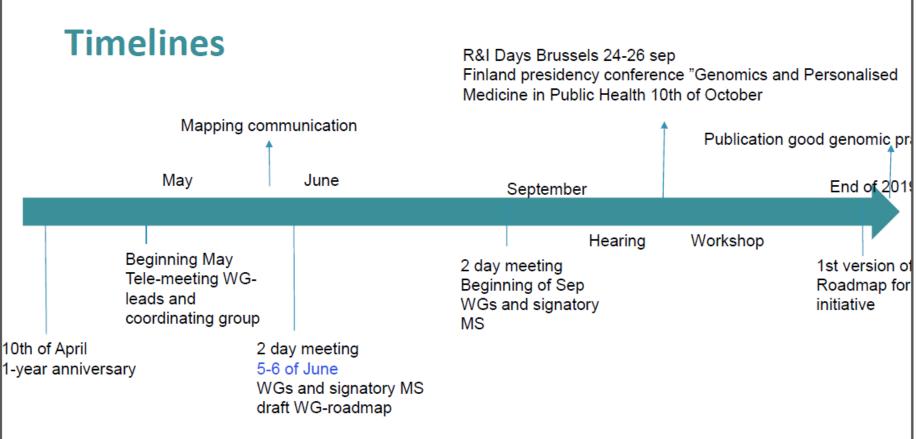
Combine WGS and genotyping data

Current genotyping capacity (40,000 GWAs/year)









- Analysis of mapping exercise
- · Identify current projects/initiatives/groups relevant for the topic
- Suggest process for how to connect and collaborate with relevant expertise
- Report on national/regional priorities for structural funds

- WGs: Formulate a suggestion for a road map on what need to be addressed and when for ex. road map for setting up interoperability framework and requirements to join
- → September meeting 1<sup>st</sup> draft roadmap for the initiative

