

From reads to disease variants

Introduction to openCGA and IVA for variant prioritization

Javier Pérez Florido

javier.perez.florido.sspa@juntadeandalucia.es

<http://www.clinbioinfospa.es>

Clinical Bioinformatics Area

Fundación Progreso y Salud / Servicio Andaluz de Salud
Seville

Overview

- Introduction
- *Big Data* in Genomics
- OpenCB: Open source initiative for Computational Biology
- A case study: Personalized Medicine Module (MMP)

Overview

- **Introduction**
- *Big Data* in Genomics
- OpenCB: Open source initiative for Computational Biology
- A case study: Personalized Medicine Module (MMP)

The Clinical Bioinformatics Area

Clinical Bioinformatics Area The Area Tools Publications Courses News

Welcome

The Clinical Bioinformatics Area has opened a position for a Web Programmer.

Submitted by jdopazo on Tue, 03/20/2018 - 09:54



The Clinical Bioinformatics Area from the Fundación Progreso y Salud (FPS) has been conceived as a fundamental piece of the Personalized Medicine plan of the Andalusian community, with the mission of facilitating and providing the tools for the inclusion of the genomic data of the patient in the electronic health record.

This Area has the dual aim of developing innovative algorithms and methods for the analysis of genomic data of

<http://www.clinbioinfospa.es/>

Fundación Progreso y Salud
CONSEJERÍA DE SALUD

Tools:

Other tools

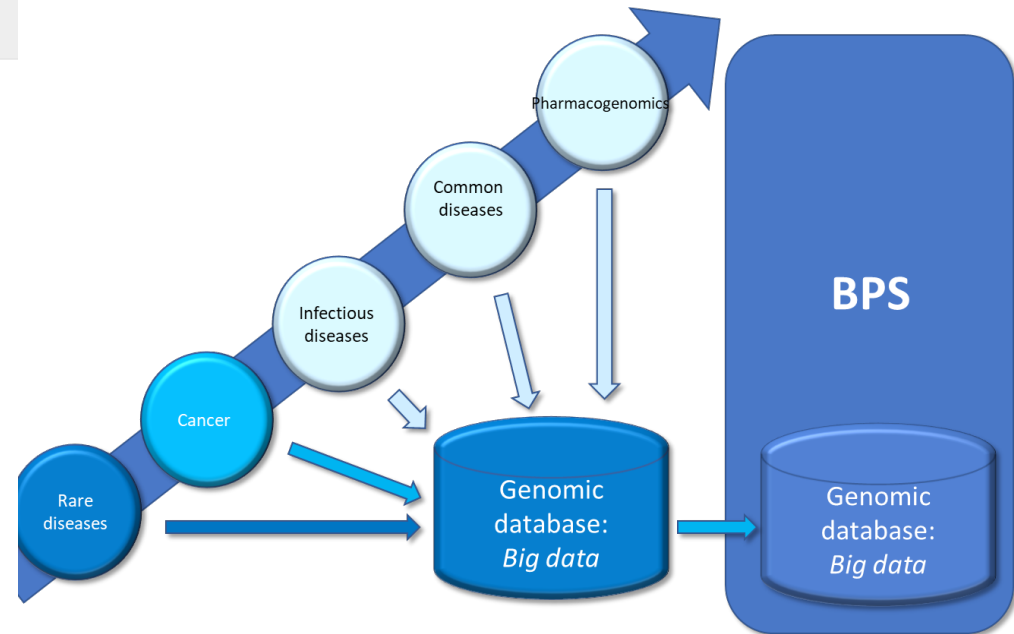
Tweets by @ClinicalBioInfo

Clinical BioInfo Retweeted

Joaquín Dopazo
@xdopazo
clinbioinfospa.es/content/becca-r...

Beca remunerada para...
Formación práctica y teó...
clinbioinfospa.es

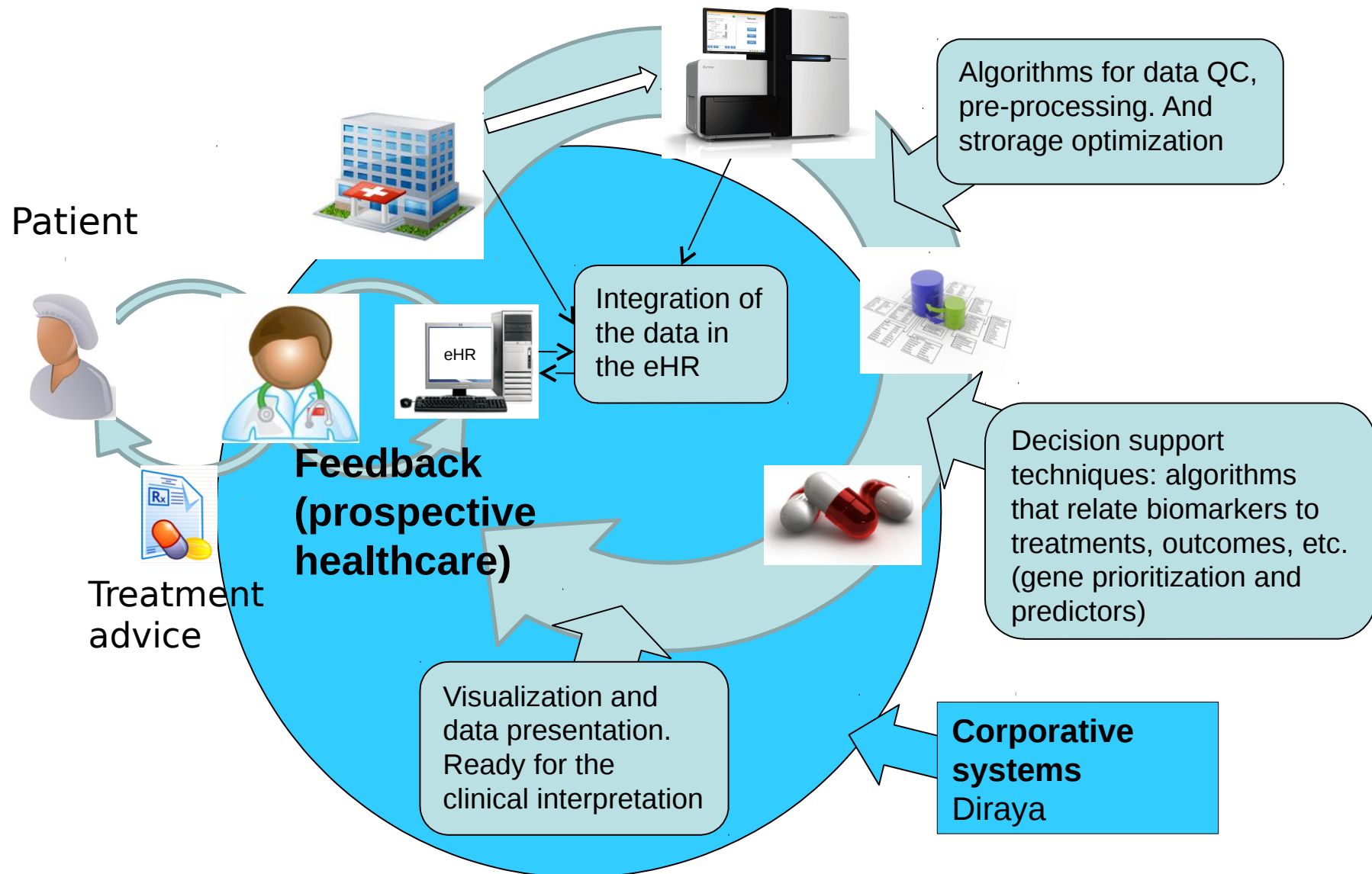
Apr 30, 2018



The **Bioinformatics Area**, created in June 2016 in the Fundación Progreso y Salud, has as main goal supporting the Program of Personalized Medicine of the Andalusian Community by facilitating the use of genomic data for precision diagnostic and treatment recommendation, implementing a prospective health care functionality in the public health system.

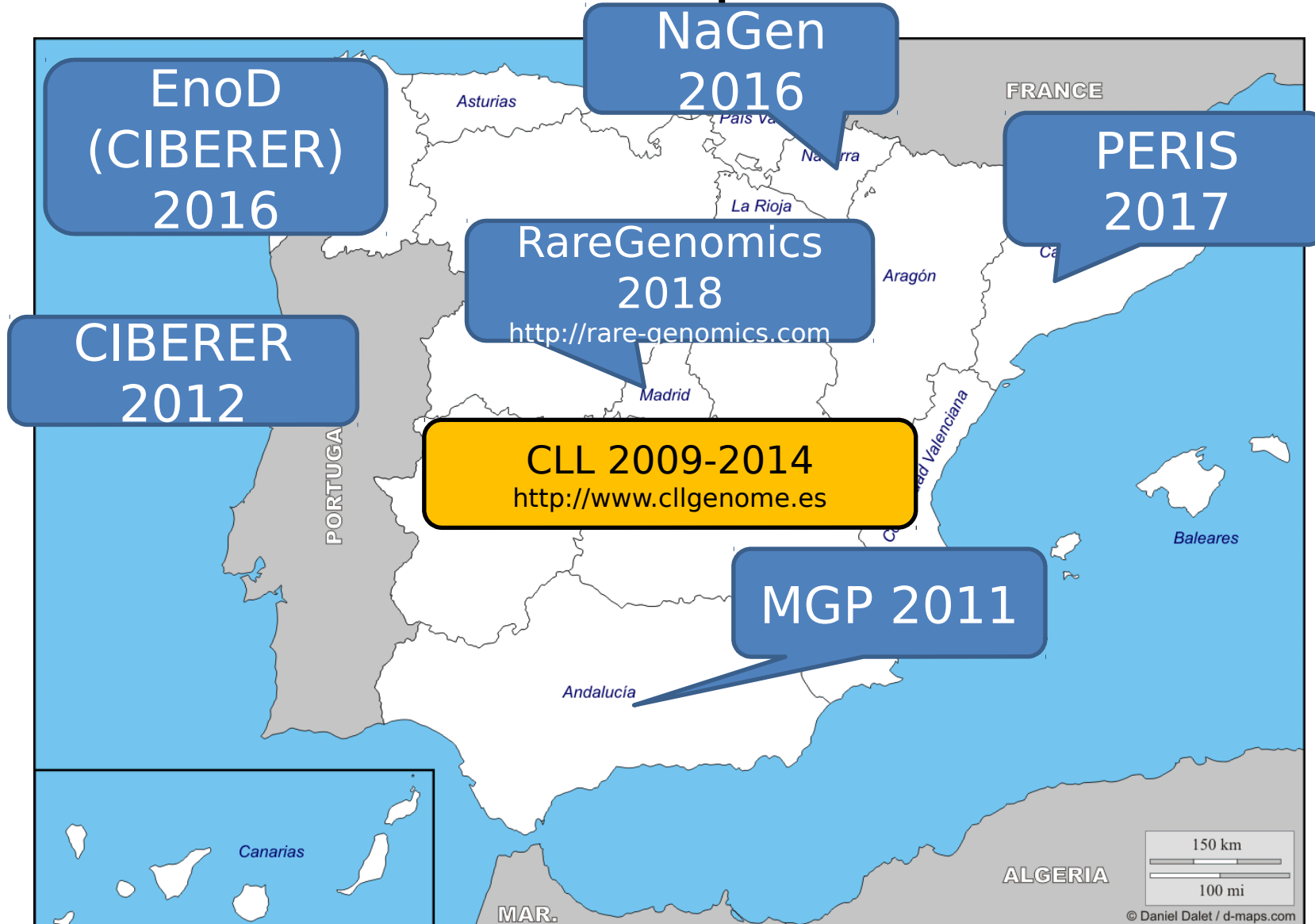
Introduction

Bioinformatics for personalized medicine within the Andalusian health system



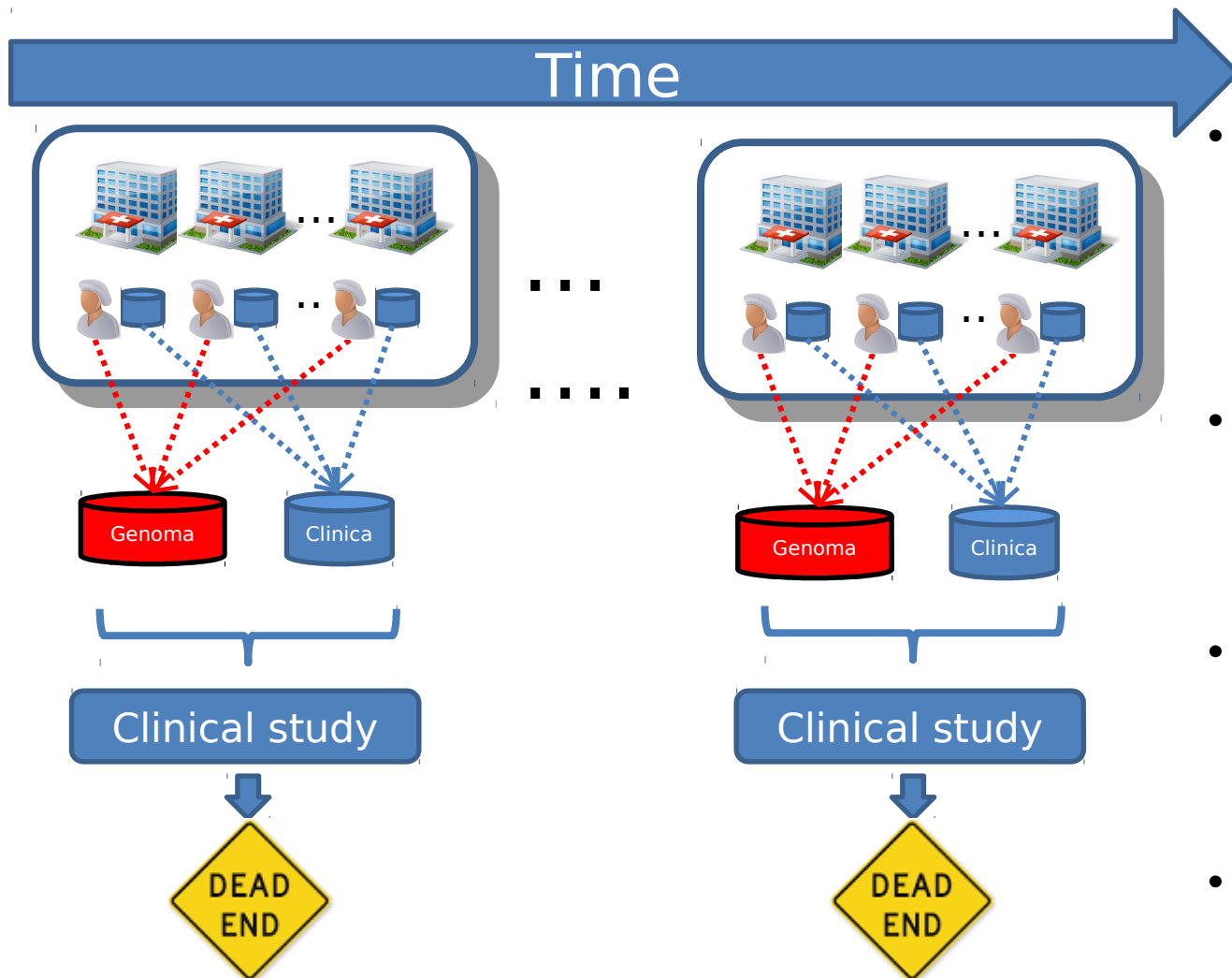
Introduction

Genomic initiatives in Spain



Introduction

Genomic initiatives without universal eHR are just clinical studies but not Personalized Medicine



- Each study requires of a specific genomic and clinical data collection into an external database
- Static clinical data (e.g. if a control becomes a case the external DB will not be updated)
- Limited genomic data reuse for purposes different from the original study
- Model of GEL (100,000 genomes) and other initiatives.

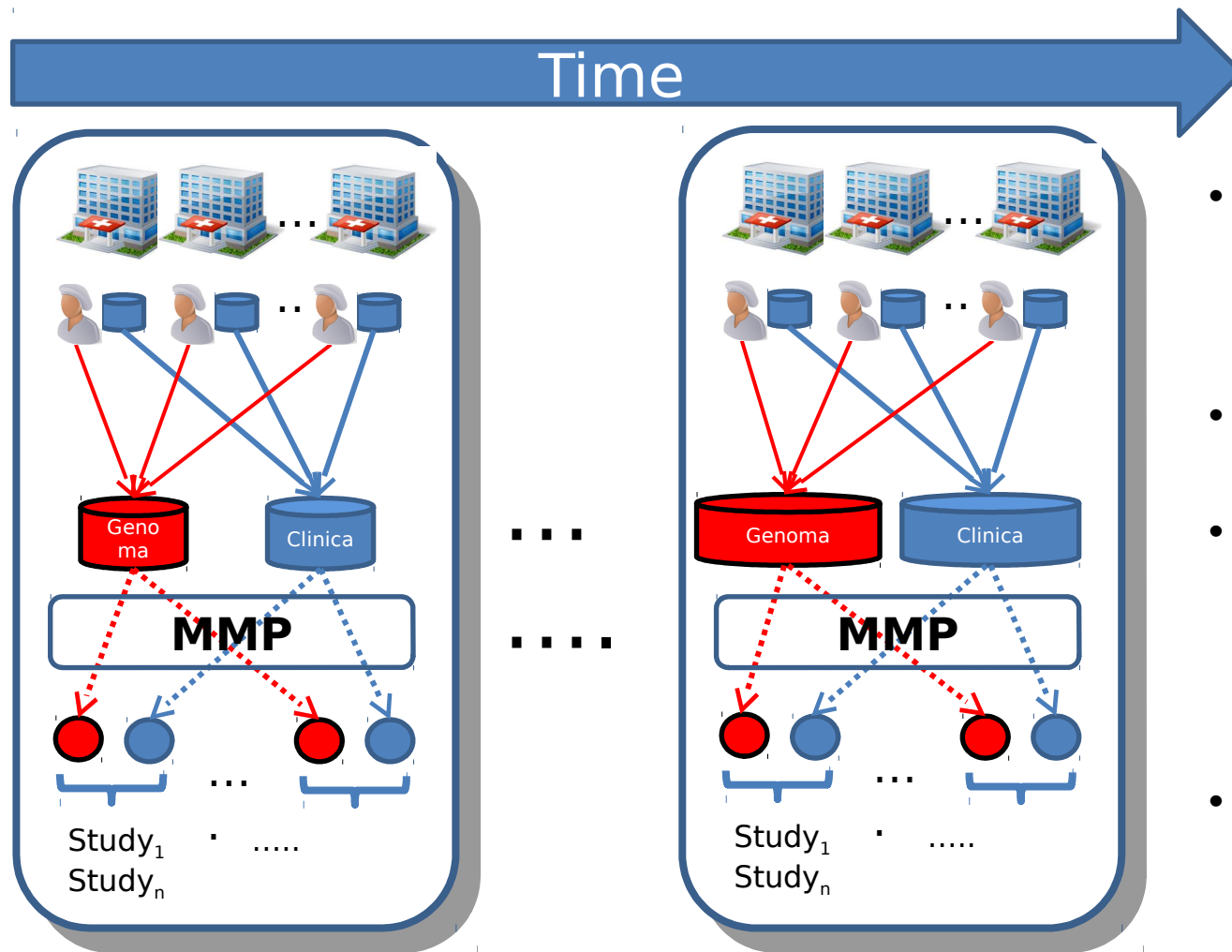
The value of genomic data for the health system



8

Introduction

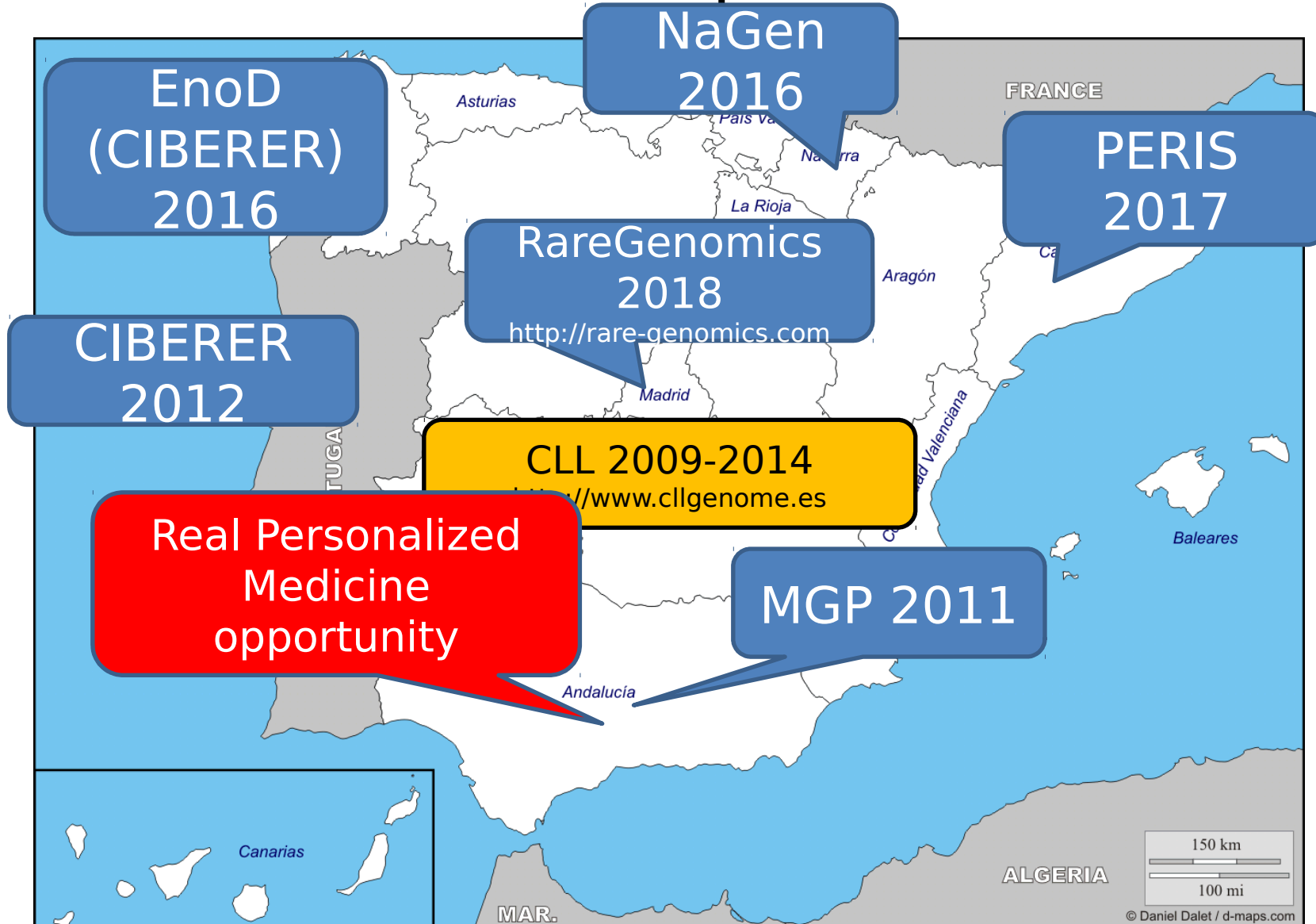
Personalized medicine requires a model that integrates genomic data and universal eHR



- The whole health system becomes a enormous potential prospective study
- Clinical data dynamically associated to patients
- Possibility of many clinical studies by reanalyzing genomic data under diverse perspectives (with no extra investment)
- Growing genomic DB with increasing study possibilities

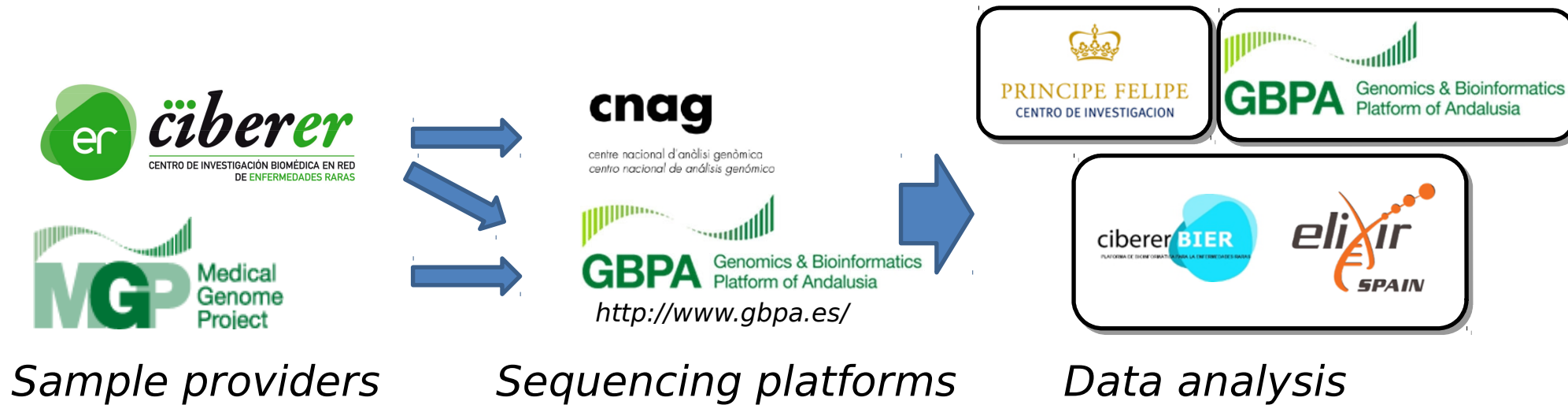
Introduction

Genomic initiatives in Spain



Introduction

Previous initiatives: MGP and CIBERER initiatives to sequence rare disease patient exomes



Diseases with

- Unknown causal genes
- No mutations in known genes

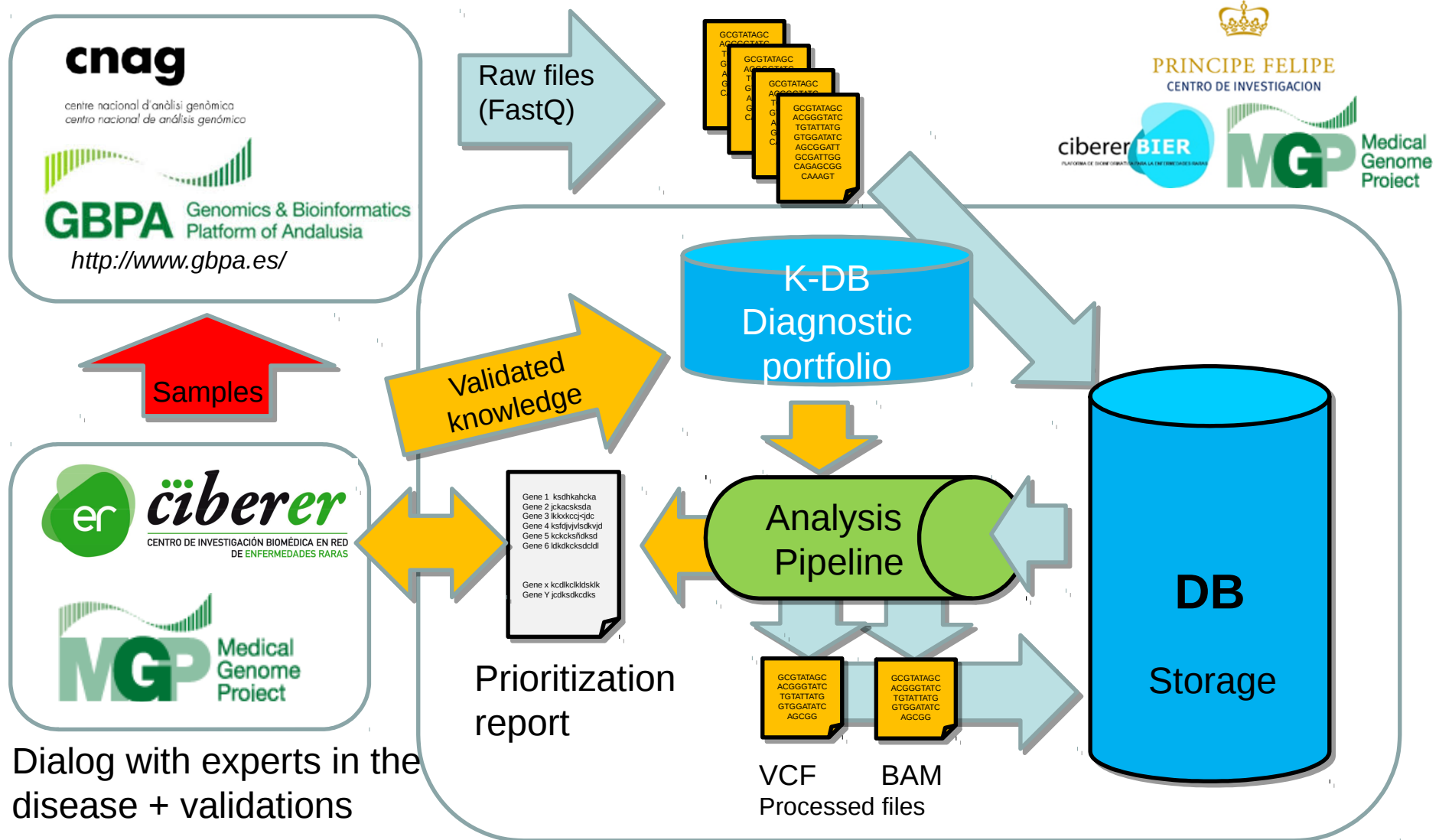
Search for:

- New disease genes
- Susceptibility genes

A total of 1044 exomes of 300 healthy controls and patients of more than 30 diseases were sequenced between 2012 and 2013.

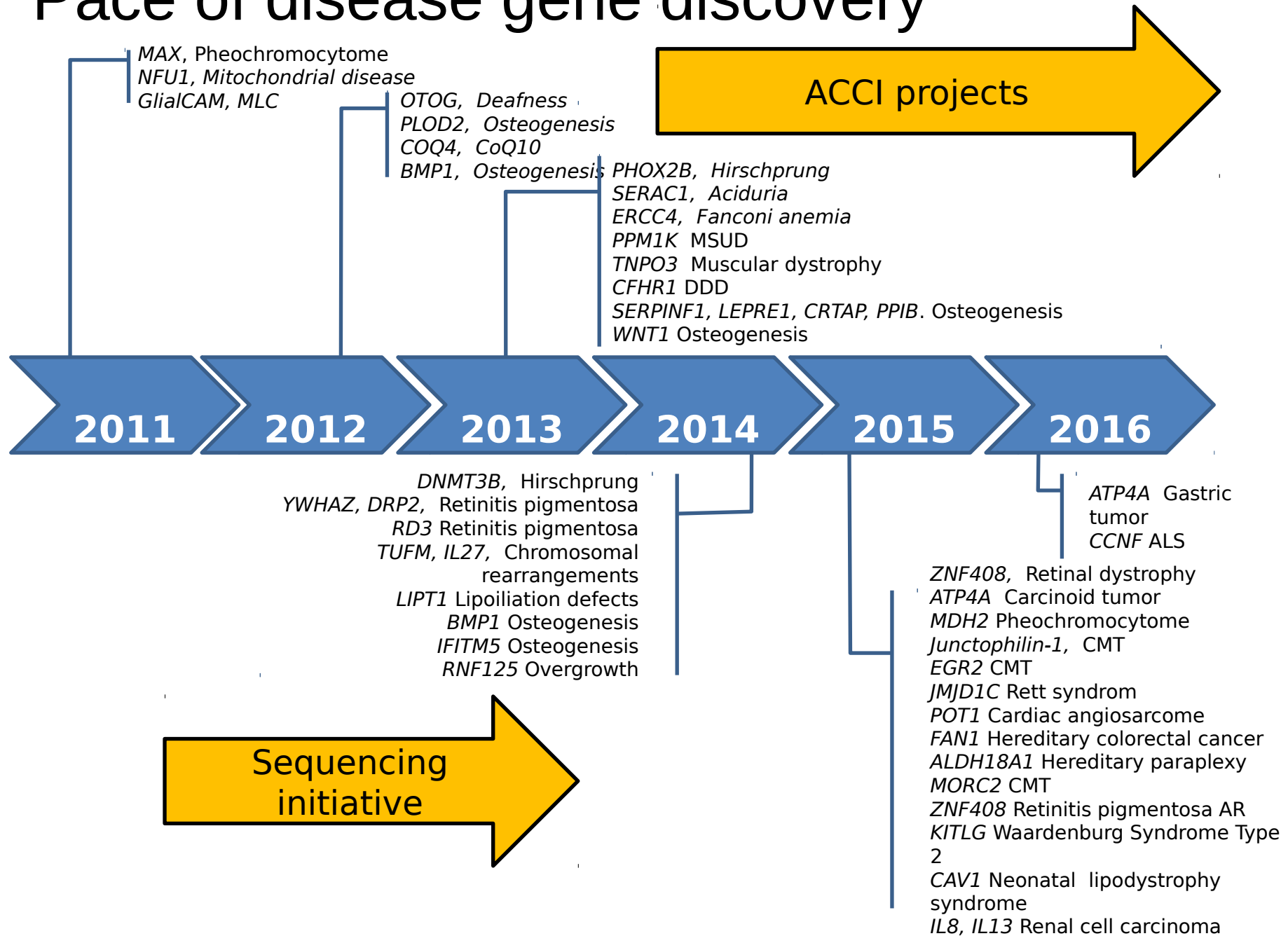
Introduction

Data analysis and the cycle of knowledge generation



Introduction

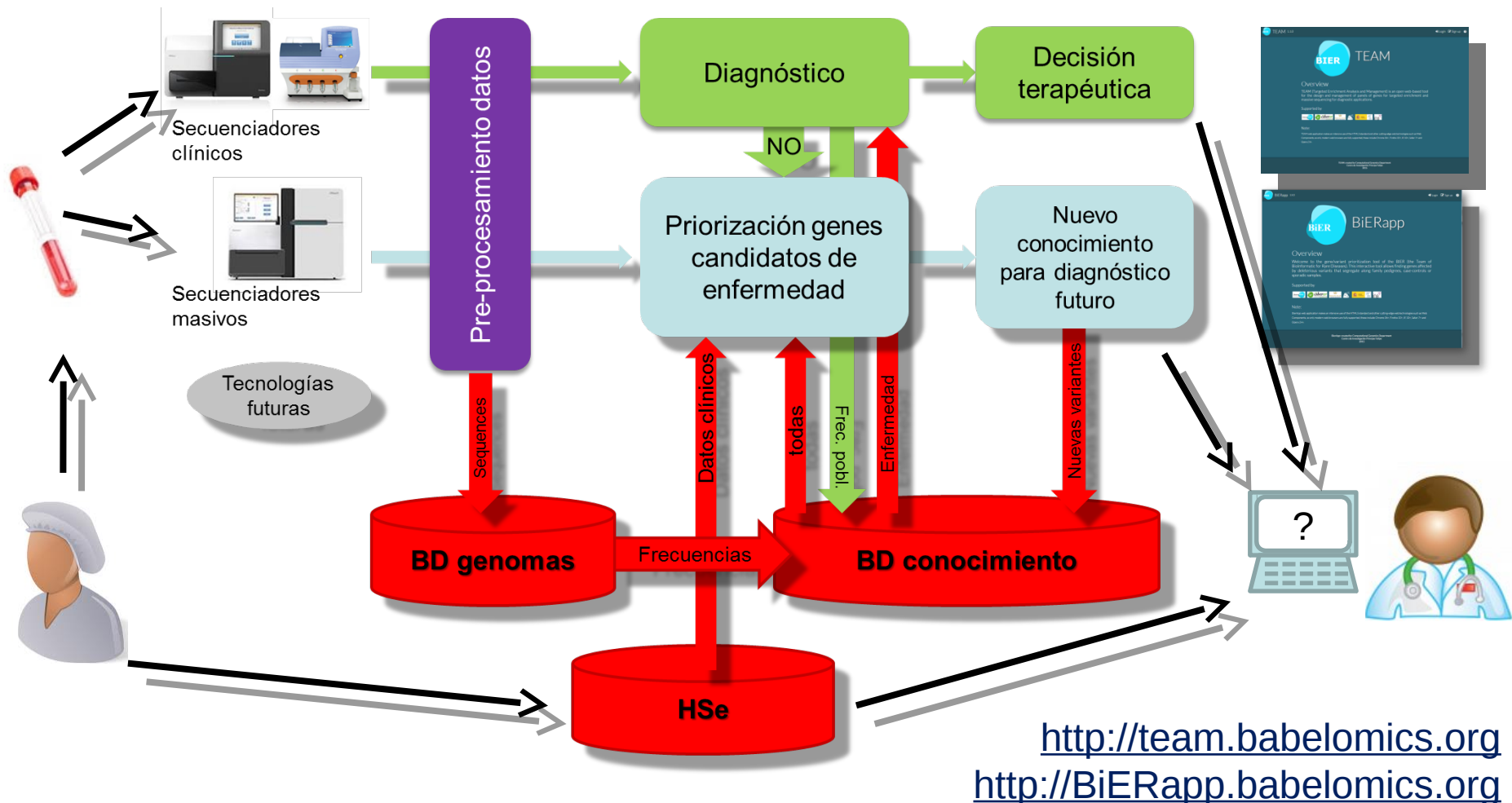
Pace of disease gene discovery



Introduction

CIBERER initiative for diagnosis and biomarker discovery using massive sequencing

Ongoing (second phase) Project with the collaboration of seven hospitals: La Paz, FJD, Ramón y Cajal, CBM (Madrid), Virgen del Rocío (Sevilla), Hospital del Mar (Barcelona), HU La Fe (Valencia)., **within the context of CIBERER**



Introduction

Lessons learned: the importance of local variability in the prioritization process

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

MOLECULAR BIOLOGY AND EVOLUTION

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Oxford Journals > Medicine & Health & Science & Mathematics > Molecular Biology and Evolution > Volume 33, Issue 5 > Pp. 1205-1218.

267 Spanish Exomes Reveal Population-Specific Differences in Disease-Related Genetic Variation

Joaquín Dopazo^{1,2,3,4}, Alicia Amadoz¹, Marta Bleda^{1,3}, Luz García-Alonso¹, Alejandro Alemán^{1,3}, Francisco García-García¹, Juan A. Rodríguez⁵, Josephine T. Daub⁵, Gerard Muntané⁵, Antonio Rueda², Alicia Vela-Boza², Francisco J. López-Domingo², Javier P. Florido², Pablo Arce², Macarena Ruiz-Ferrer^{2,6,7}, Cristina Méndez-Vidal^{6,7}, Todd E. Arnold^{1,8}, Olivia Spleiss⁹, Miguel Alvarez-Tejado¹⁰, Arcadi Navarro^{11,12,13}, Shomi S. Bhattacharya^{2,14}, Salud Borrego^{6,7}, Javier Santoyo-López^{1,2} and Guillermo Antiñolo^{2,6,7}

+ Author Affiliations

*Corresponding author: E-mail: jdopazo@cipf.es; gantinolo@us.es.

Abstract

« Previous | Next Article »
Table of Contents

This Article

Mol Biol Evol (2016) 33 (5): 1205-1218.
doi: 10.1093/molbev/msw005
First published online: January 13, 2016

This article is Open Access
Abstract **Free**
» Full Text (HTML) **Free**
Full Text (PDF) **Free**
Supplementary Data

All Versions of this Article:
msw005v1
msw005v2
msw005v3
33/5/1205 most recent

Search this journal:
Advanced »

Current Issue

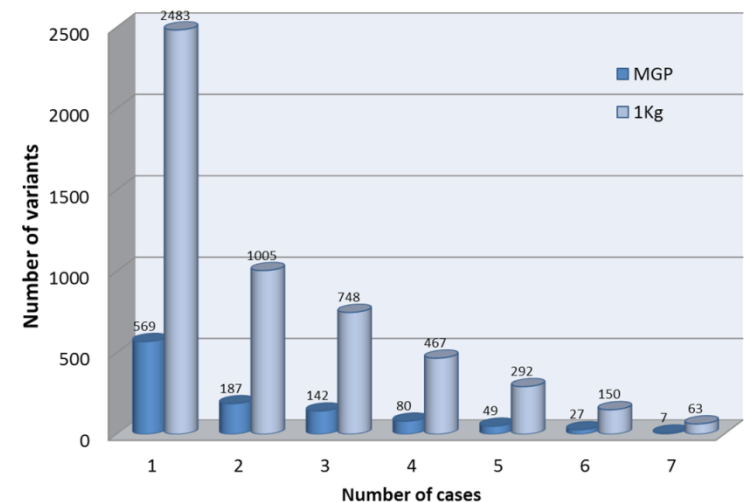
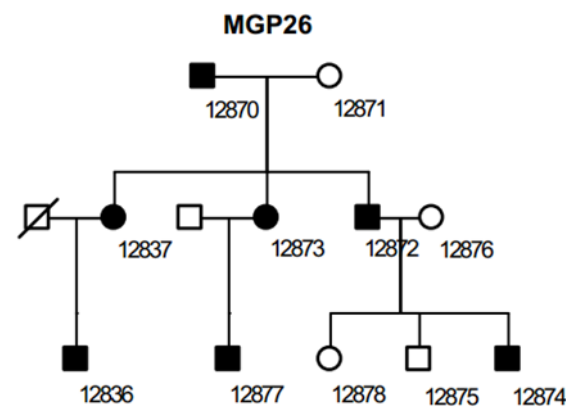
November 2016 33 (11)

MOLECULAR BIOLOGY AND EVOLUTION

Alert me to new issues

Editors

We discovered some 12,000 “spanish” polymorphisms not present in other databases. The filtering efficiency enormously increases using local population data



Introduction

The CSVS is a crowdsourcing project

Collaborative Spanish Variant Server

BIER CSVS

Start

Overview

Welcome to the Collaborative Spanish Variant Server. CSVS was created to provide information about the variability of the Spanish population to the scientific/medical community. It is useful for filtering polymorphisms and local variations in the process of prioritizing candidate disease genes. CSVS currently stores information on 1582 unrelated Spanish individuals. We accept submissions from WES or WGS. See the protocol for sending samples.

Supported by

Note:

CSVs web application makes an intensive use of the HTML5 standard and other cutting-edge web technologies such as Web Components, so only modern web browsers are fully supported, these include Chrome 36+, Firefox 32+, IE 10+, Safari 7+ and Opera 24+.

CSVs created by Clinical Bioinformatics Area
Fundación Príncipe y Salud
2013-2017

ID	Gene	RET	rs12267460	575	2	1	0	0.997	0.003	0.003	0.350	0.400	0.270	0.330	0.400	0.263	0.248	0.300	0.482	0.285
10_43572832	G-A	RET	rs12267460	508	52	18	0	0.924	0.076	0.076	0.350	0.400	0.270	0.330	0.400	0.410	0.438	0.410	0.305	0.410
10_43595808	C-T	RET	.	575	3	0	0	0.997	0.003	0.003	.	.	.	.	.	.	.	.	.	.
10_43595818	C-T	RET	.	575	3	0	0	0.997	0.003	0.003	.	.	.	.	.	.	.	.	.	.
10_43595837	T-T	RET	.	474	71	33	0	0.881	0.119	0.119	.	.	.	.	.	.	.	.	.	.
10_43595837	T-T	RET	.	474	71	33	0	0.881	0.119	0.119	.	.	.	.	.	.	.	.	.	.
10_43595968	A-G	RET	rs1808058	348	74	156	0	0.666	0.334	0.334	0.280	0.220	0.470	0.090	0.300	0.246	0.251	0.306	0.473	0.285
10_43596003	G-A	RET	.	577	1	0	0	0.999	0.001	0.001	.	.	.	.	.	.	.	.	.	.
10_43596179	G-A	RET	rs12433351	280	155	36	7	0.801	0.199	0.199	0.180	0.260	0.130	0.070	0.260	0.185	0.280	0.288	0.116	0.116

<http://csvs.babelomics.org/>

Allelic population frequencies obtained from about 2,000 samples (exomes and genomes) available in CSVS

Scenario: Sequencing projects of healthy population are expensive and funding bodies are reluctant to fund them

CSVS Aim: To offer increasingly accurate information on variant frequencies characteristic of Spanish population.

CSVS Main use: Frequency-based filtering of candidate variants

Main data source: Sequencing projects of individual researchers (CIBERER and others)

Problem: Most of the contributions correspond to patient exomes

Idea: Patients of disease A can be considered healthy **pseudo-controls** for disease B (providing no common genetic background exist between A and B)

Beacon: CSVS has a Beacon server (Beacon network is search engine across beacons which enables global discovery of genetic mutations)

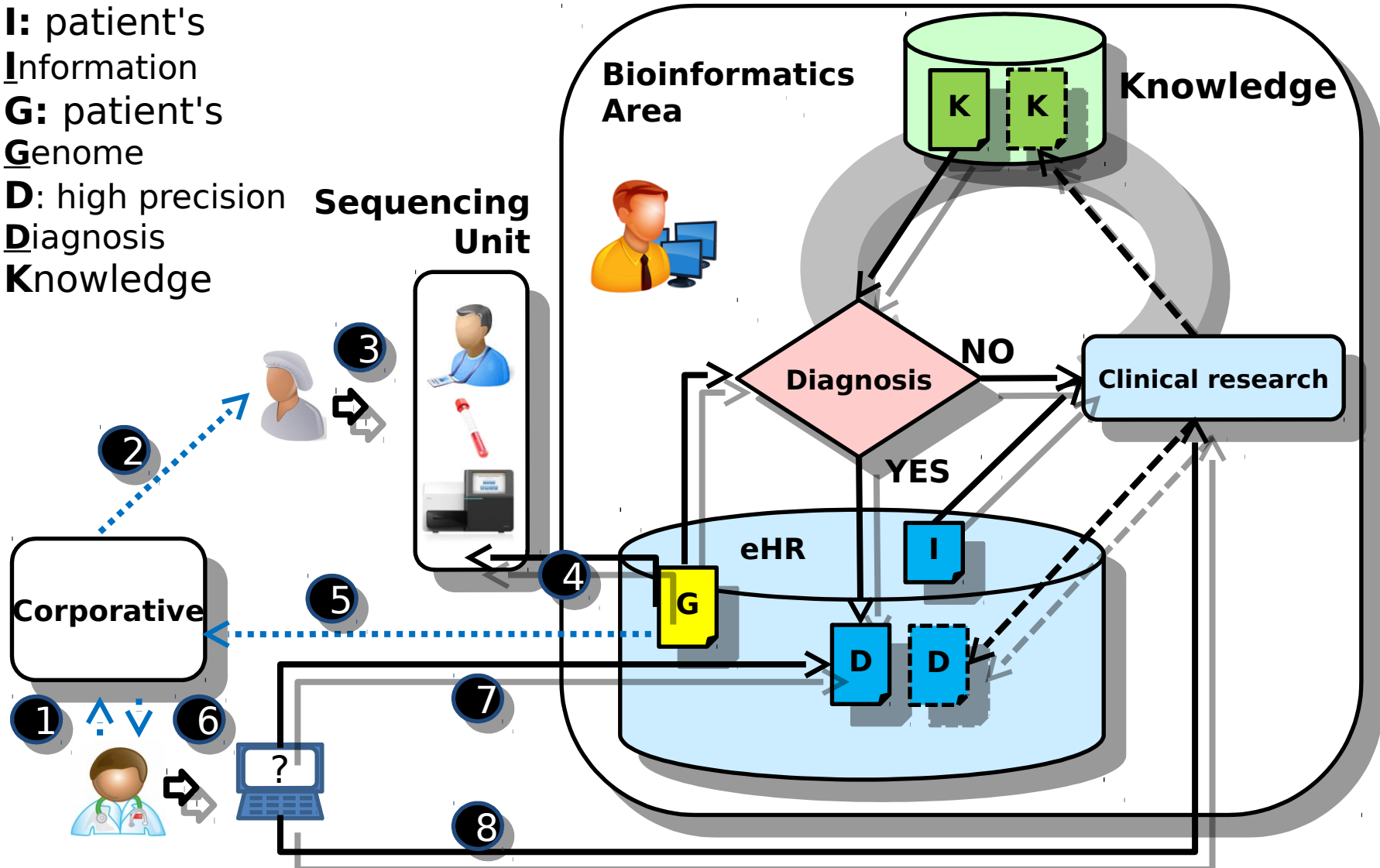
URL: <https://beacon-network.org>

Introduction

MMP for diagnosis and clinical research within the Andalusian health system

I: patient's
Information
G: patient's
Genome
D: high precision
Diagnosis
K: Knowledge

Sequencing Unit



Overview

- Introduction
- ***Big Data* in Genomics**
- OpenCB: Open source initiative for Computational Biology
- A case study: Personalized Medicine Module (MMP)

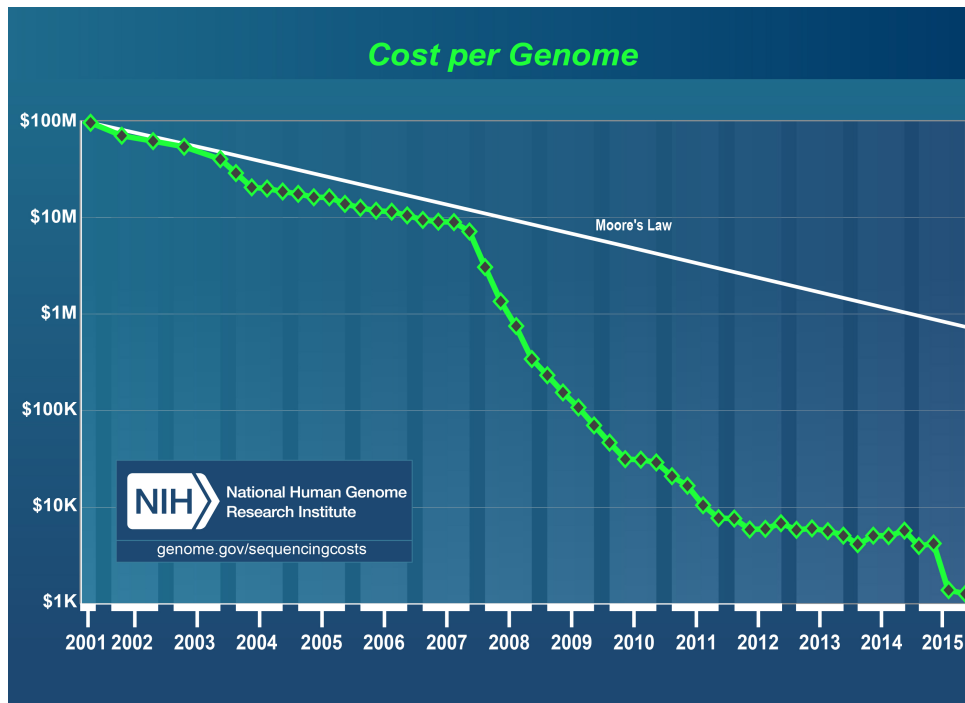
Big data in Genomics

A new scenario in biology

NGS experiments *have increased data size by more than 5000x* compared to microarrays.

Challenges:

- **Sequencing cost** keeps falling while throughput keeps increasing. Today a whole genome can be sequenced by less than **\$1000**
- **Variant analysis**: re-sequencing projects such as Whole Genome Sequencing (WGS) aims to find genomic variants and genes involved in phenotypes and diseases



A single HiSeq X Ten System can sequence
~20,000 human genomes a year

Human whole-genome sequencing power

Maximum throughput for population- and production-scale genomics.

HAVE QUESTIONS ABOUT THE HISEQ X SEQUENCING SYSTEMS? [CONTACT A SALES REPRESENTATIVE »](#)

The \$1000 Genome Is Here

Discover how HiSeq X Ten breaks the \$1000 genome barrier for human whole-genome sequencing.
[Learn more »](#)

New HiSeq X Five System Available

The HiSeq X Five is a set of five ultra-high-throughput sequencers for production-scale human whole-genome sequencing.
[Learn more »](#)

Big data in Genomics

A new scenario in biology

NGS experiments *have increased data size by more than 5000x* compared to microarrays.

Challenges:

- **Sequencing cost** keeps falling while throughput keeps increasing. Today a whole genome can be sequenced by less than **\$1000**
- **Variant analysis**: re-sequencing projects such as Whole Genome Sequencing (WGS) aims to find genomic variants and genes involved in phenotypes and diseases



- Illumina NGS sequencer series:
 - **HiSeq 2500** provides high-quality 2x125bp: 50-1000Gb in 1-6 days, 90.2% bases above Q30. One human genome at ~60x coverage
 - **HiSeq 4000** provides high-quality 2x150bp: 125-1500Gb in 1-4 days, >75% bases above Q30. Up to 12 human genomes at ~40x coverage
- **Each sample** produces a **FASTQ** file ~**1TB** size containing ~**1-2B** reads
- **Illumina X Ten**: Consists of 10 ultra-high-throughput HiSeq X sequencers. First \$1000 human genome sequencer, it can sequence up to 20,000 genomes per year

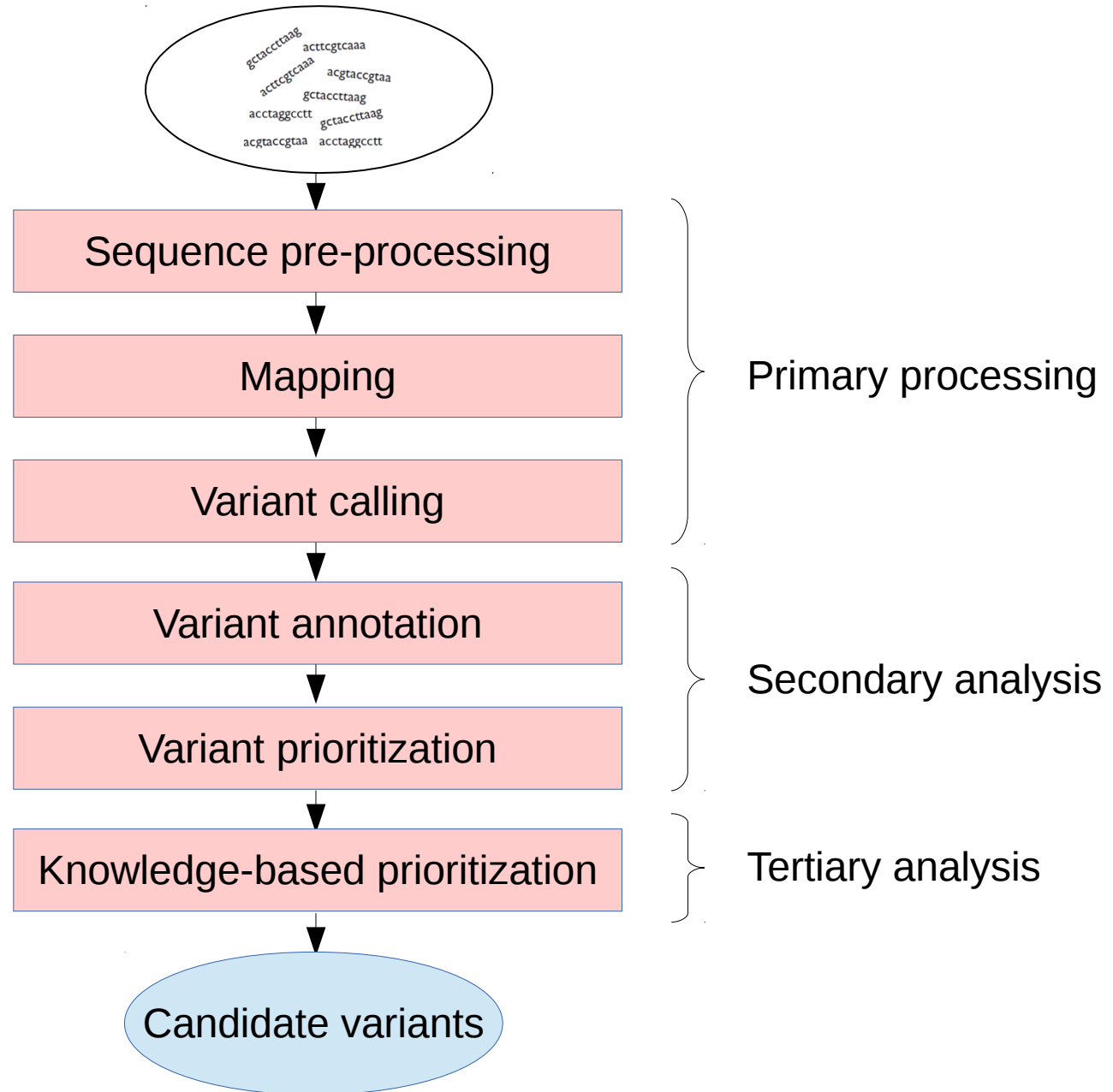


Real flexibility.
Real throughput.
Real data quality.

The HiSeq 2500 is ready for any application, any sample size—today.

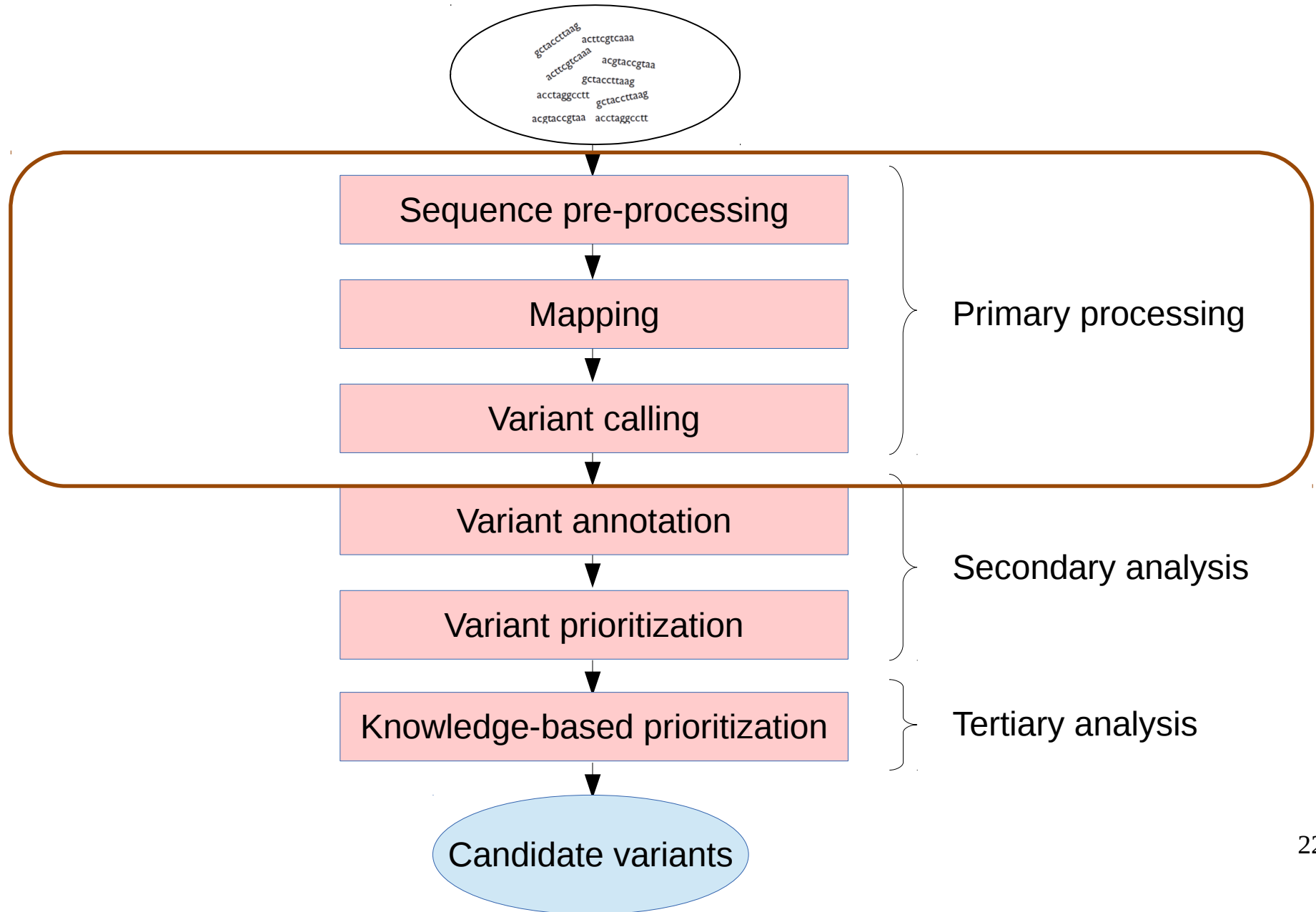
Big data in Genomics

Standard NGS variant data analysis



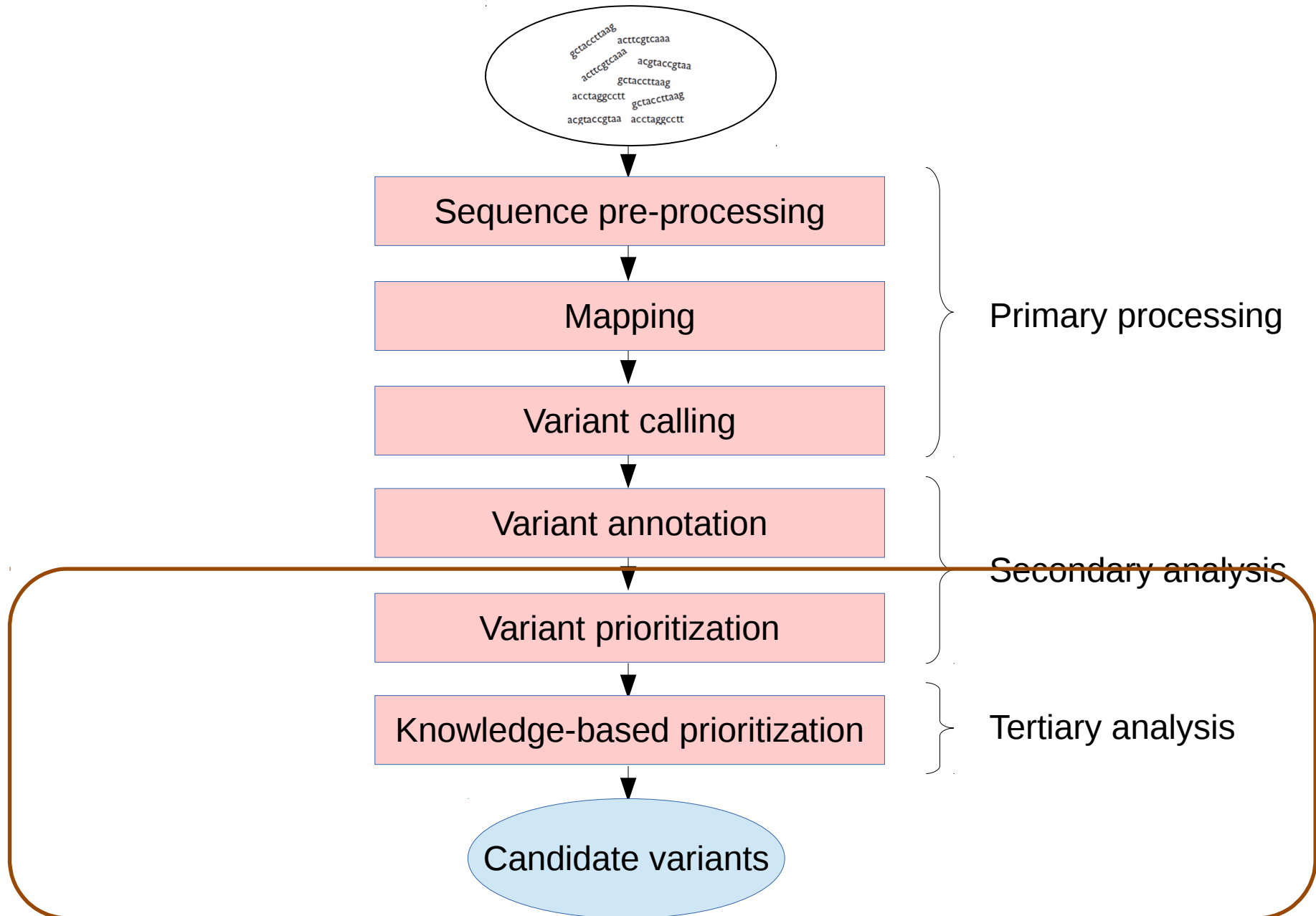
Big data in Genomics

Standard NGS variant data analysis



Big data in Genomics

Standard NGS variant data analysis



Big data in Genomics

Standard NGS variant data analysis

CAUTION!

On average, each *normal* person is found to carry in the **exome**:

~11,000 **synonymous** variants

~11,000 **non-synonymous** variants

250 to 300 loss-of-function variants in annotated genes

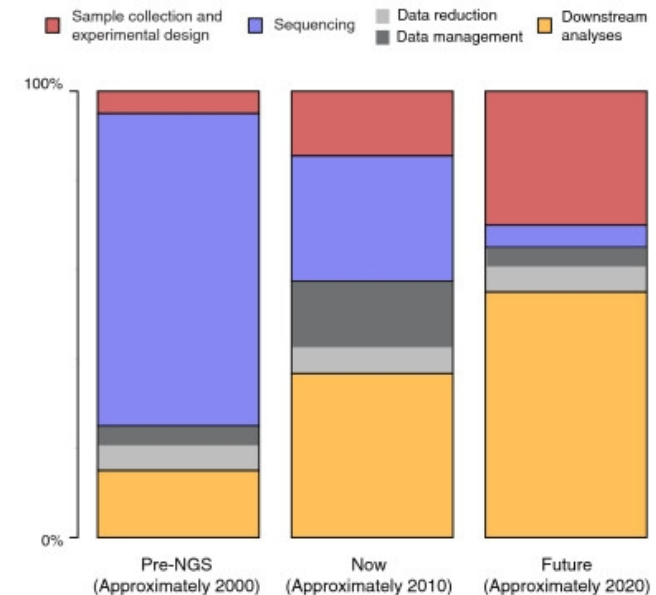
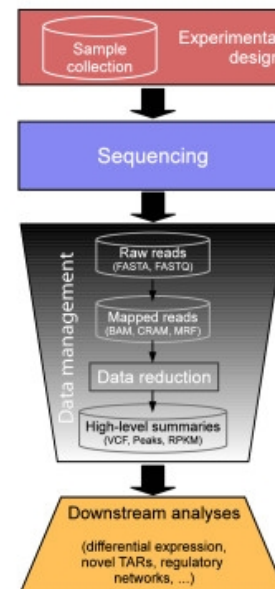
50 to 100 variants previously implicated in **inherited disorders**

1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. **Nature**. 2010 Oct 28;467(7319):1061-73. PubMed PMID: 20981092



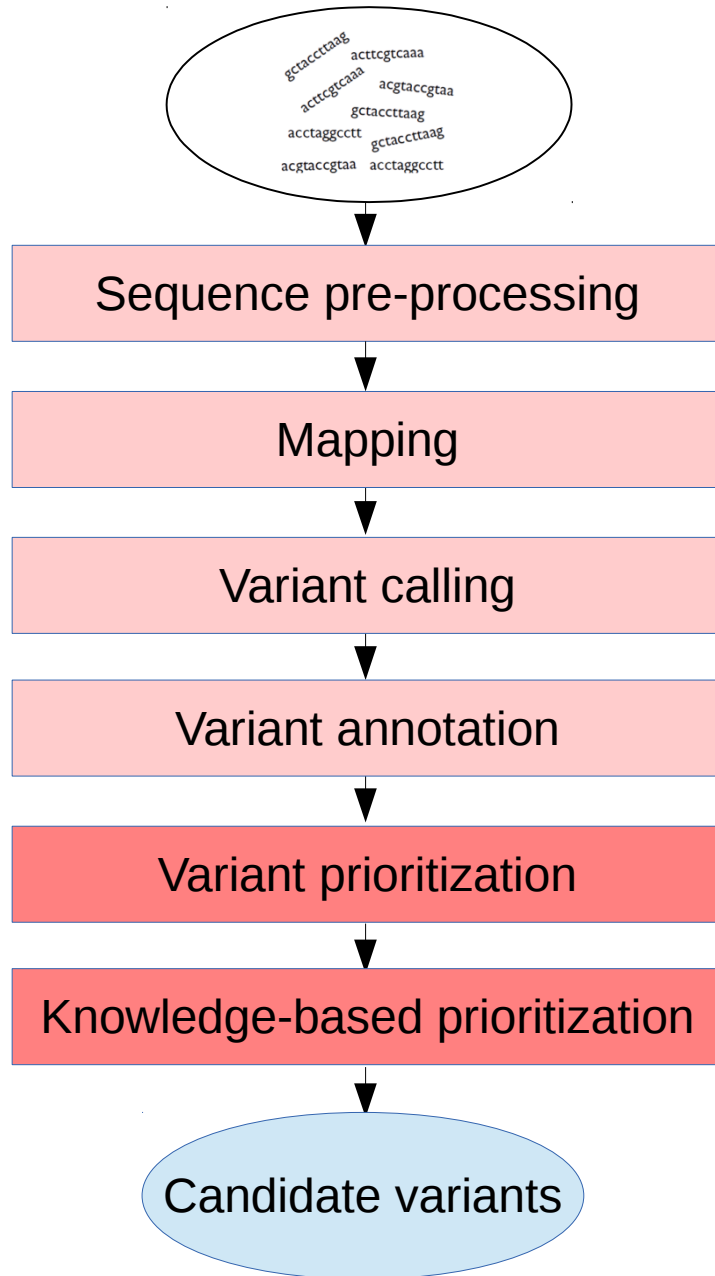
Still a challenge → **PRIORITIZATION**

- A **whole genome** can carry 3.5 million variants on average
- After annotating there will be **hundreds** of **deleterious** variants

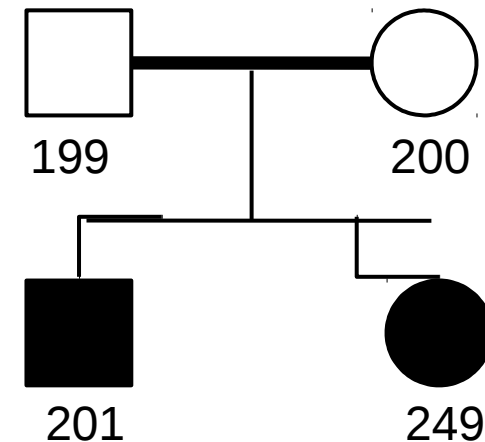


Big data in Genomics

Standard NGS variant data analysis

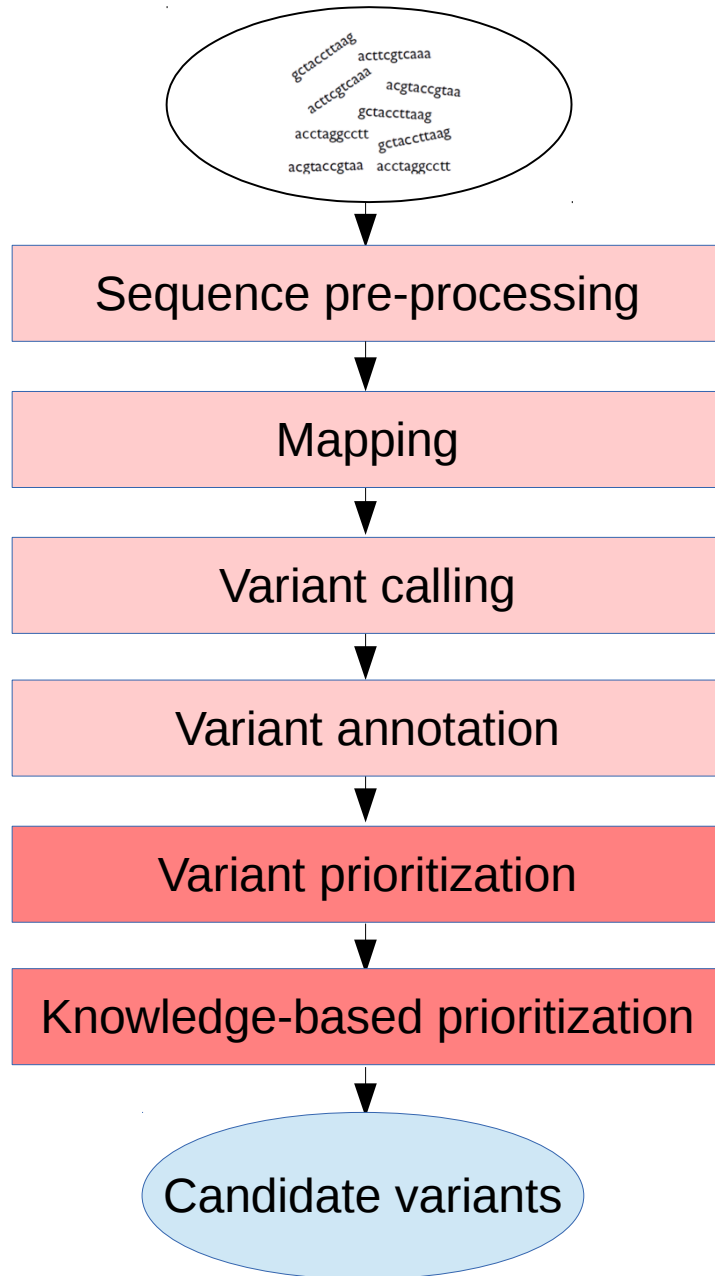


- Variant level: population frequencies, conservation scores, protein substitution scores....
- Experimental design level: case/control, Family, Trios, Disease Panels



Big data in Genomics

Standard NGS variant data analysis



- Variant level: population frequencies, conservation scores, protein substitution scores....
- Experimental design level: case/control, Family, Trios, Disease Panels

MOLECULAR BIOLOGY AND EVOLUTION

Issues Advance articles Submit Purchase Alerts About All Molecular B



Volume 33, Issue 5
May 2016

Article Contents

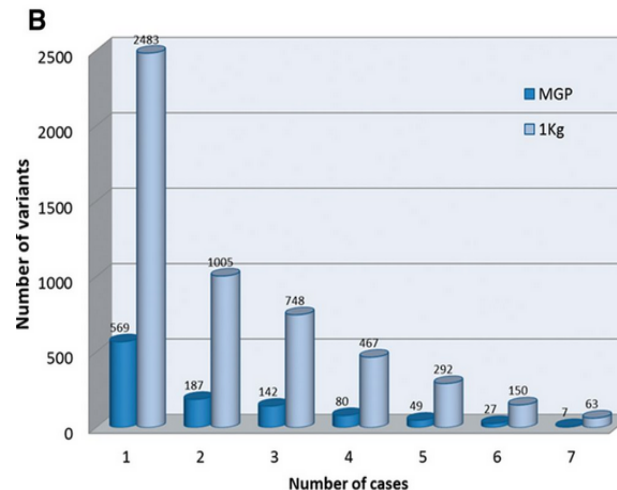
Abstract
Introduction
..

267 Spanish Exomes Reveal Population-Specific Differences in Disease-Related Genetic Variation

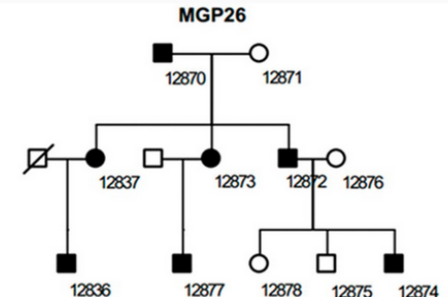
Joaquín Dopazo, Alicia Amadoz, Marta Bleda, Luz García-Alonso, Alejandro Alemán, Francisco García-García, Juan A. Rodríguez, Josephine T. Daub, Gerard Muntané, Antonio Rueda Alicia Vela-Boza, Francisco J. López-Domingo, Javier P. Florido, Pablo Arce, Macarena Ruiz-Ferrer, Cristina Méndez-Vidal, Todd E. Arnold, Olivia Spleiss, Miguel Alvarez-Tejado, Arcadi Navarro, Shomi S. Bhattacharya, Salud Borrego, Javier Santoyo-López, Guillermo Antiñolo

Molecular Biology and Evolution, Volume 33, Issue 5, 1 May 2016, Pages 1205–1218,
<https://doi.org/10.1093/molbev/msw005>
Published: 13 January 2016

Views PDF Cite Permissions Share



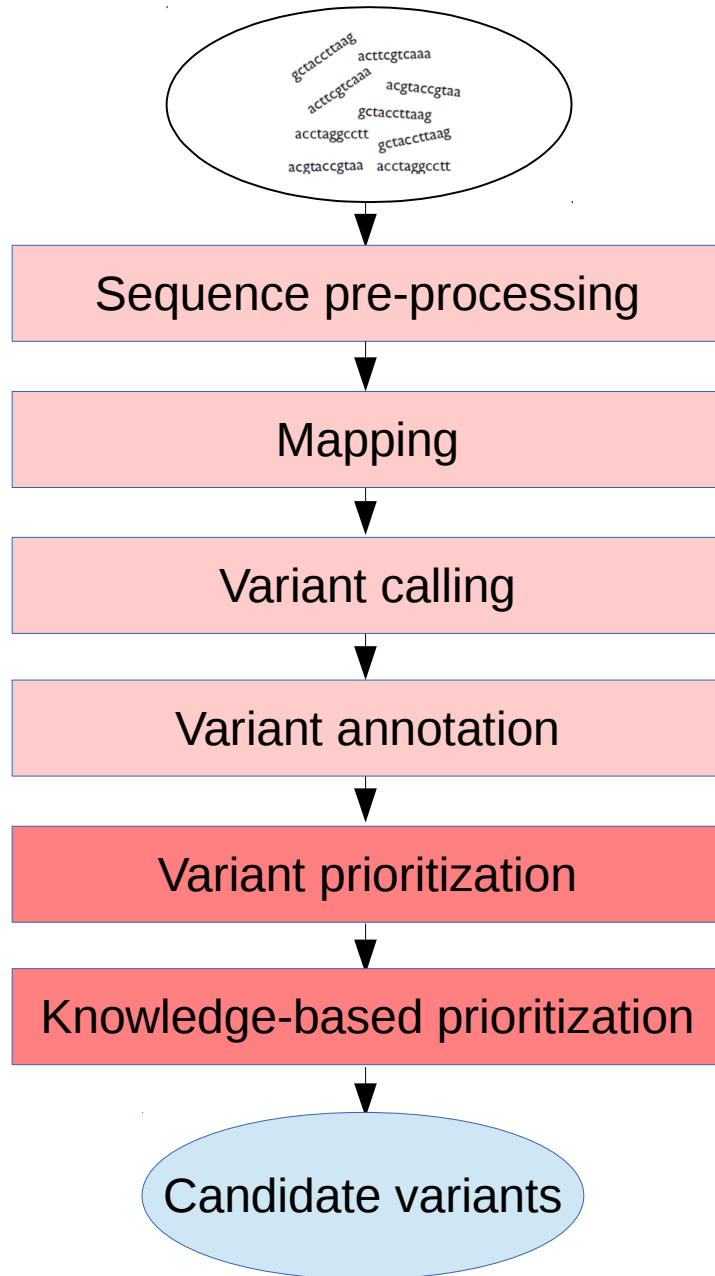
A



The filtering efficiency of the local population can be between 5 and 10 times those of a general database, such as the 1000 genomes

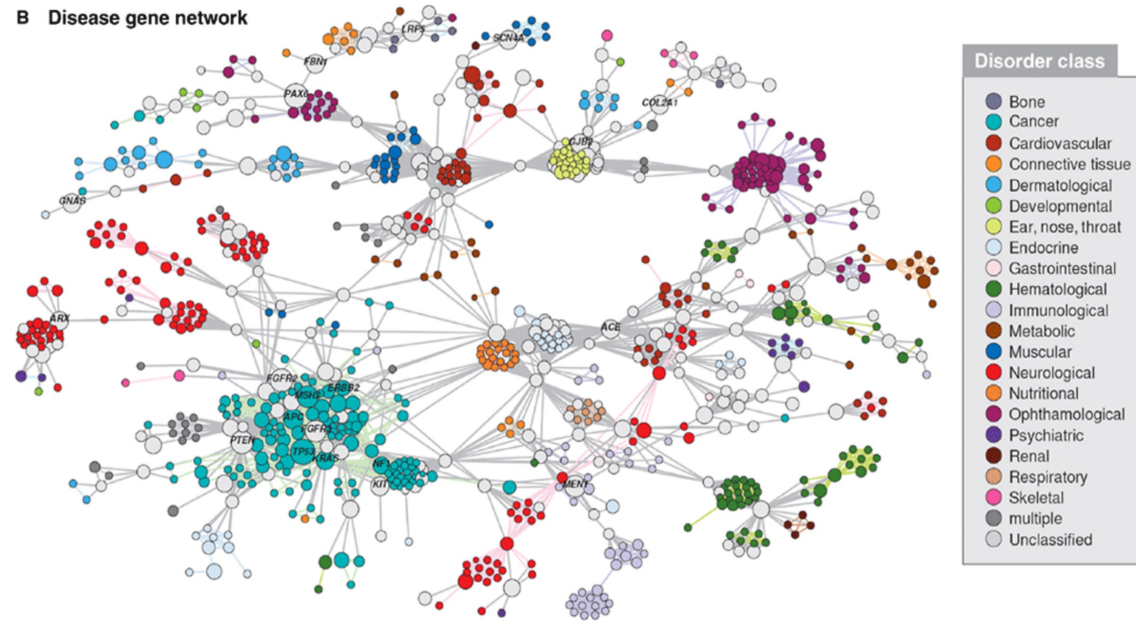
Big data in Genomics

Standard NGS variant data analysis



- **Functional (system) level: gene set, network analysis, pathway analysis, burden tests....**

B Disease gene network



Big data in Genomics

Genomic Variant Dataset, big and complex

Logical view of genomic variant dataset, data come from **different VCF files**.

Hundreds of millions of mutations, some meta data needed: **Variant annotation**

- Clinical info
- Consequence types
- Conservation scores
- Population frequencies
- ...

Genomic Variants

		Samples					
							
		var 1	28	32	29	28	35
		var 1	16,12	16,17	18,14	12,14	16,14
var_1	A/T	A/A	A/T	T/T	A/A	A/T	14
var_2	C/C	C/G	C/C	C/G	C/C	G/G	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
..	..	..	..	..	..	..	
var_n	..	..	..	..	..	..	

Clinical data: Sample annotation

- Phenotype
- Family and population pedigree
- Clinical variables
- ...

Heterogeneous data analysis and algorithms, different technologies and solutions required:

- Search and filter using data and meta data
- Data mining, correlation
- Statistic tests
- Machine learning
- Interactive analysis
- Network-based analysis
- Visualization
- Encryption
- ...

Applications:

- Personalized medicine
- Clinical Genomics
- ...

Different layers of information:

- Genotype for samples
- Allele counts
- Quality scores
- Phase information
- ...

Big data in Genomics

Some Big Data current projects

- **NIHR BRIDGE**: 10,000 whole genomes from rare diseases, ~1-2PB of data expected
- **Genomics England (GEL)**: is sequencing **100,000** whole genomes from UK, several rare diseases and cancers being studied, data estimation: ~20PB of BAM and **~400TB of compressed VCF** data are expected! About 100 whole genomes/day, ~5-10TB/day
- **International Cancer Genome Consortium (ICGC)**: store more than 10,000 sequenced cancers, few PB of data
- **NAGEN**: 1000 whole genomes from rare diseases and cancer
- **France Médecine Génétique 2025**: 10.000 WGS correspondign to 20.000 patients with rare diseases and their families, and 50.000 patients with metastatic or refractory cancers
- **Genome of the Netherlands Consortium (GoNL)**: GoNL is interested in genetic variation in the Dutch Population.
- **Andalusian program for Personalized Medicine**: 4500 samples per year (targeted sequencing, rare diseases)
-
- Danish National Strategy of Personalized Medicine (2017-2020)
- Qatar Genome Programme (QGP)
-

Big data in Genomics

Some Big Data current projects

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 532 > Issue 7600 > News > Article

NATURE | NEWS

AstraZeneca launches project to sequence 2 million genomes

Drug company aims to pool genomic and medical data in hunt for rare genetic sequences associated with disease.

Heidi Ledford

22 April 2016

[PDF](#) [Rights & Permissions](#)

One of the world's largest pharmaceutical companies has launched a massive effort to compile genome sequences and health records from two million people over the next decade. In doing so, AstraZeneca and its collaborators hope to unearth rare genetic sequences that are associated with disease and with responses to treatment.

It's an unprecedented number of participants for this type of study, says Ruth March, vice-president and head of personalized health care and biomarkers at AstraZeneca, which is headquartered in London. "That's necessary because we're going to be looking for very rare differences among individuals."



European Commission > Strategy > Digital Single Market > Policies >

Digital Single Market

POLICY

European '1+ Million Genomes' Initiative

Declaration for delivering cross-border access to genomic database



1 million **genomes accessible** in the EU by 2022



Linking access to existing and future genomic database across the EU



Providing a sufficient scale for **new clinically impactful** associations in research



Big data in Genomics

Big data analysis challenges

- **Data Analysis and visualization:** Real-time and Interactive graphical data analysis and visualization is needed.
- **Data mining:** Complex queries, aggregations, correlations, ...
- **Security:** sometimes data access require authentication, authorization, *encryption*, ...
- **Performance and scalability:** software must be high-performance and scalable
- **Data Integration:** different types of data such as variation, expression, ChIP, ...
- **Share and collaboration:** data models to ease the collaboration among different groups. Avoid moving data.
- **Knowledge base and sample annotations:** many of the visual analytic tools need genome and sample annotations



Do current bioinformatic tools solve these problems?

Big data in Genomics

Current status of *big data* tools in bioinformatics

- Many bioinformatic tools are great but, in general, not designed and implemented for processing and analyzing *big data*.
- Tools usually don't exploit the parallelism of modern hardware and current high-performance and scalable technologies. Poor performance and scalability.
- We need to develop new generation of software and methodologies to:
 - Improve performance and scalability of analysis
 - Store data efficiently and secured to be queried and visualized
- **Bioinformaticians have new challenges and requirements, given the volume, complexity, heterogeneity and nature of data.**



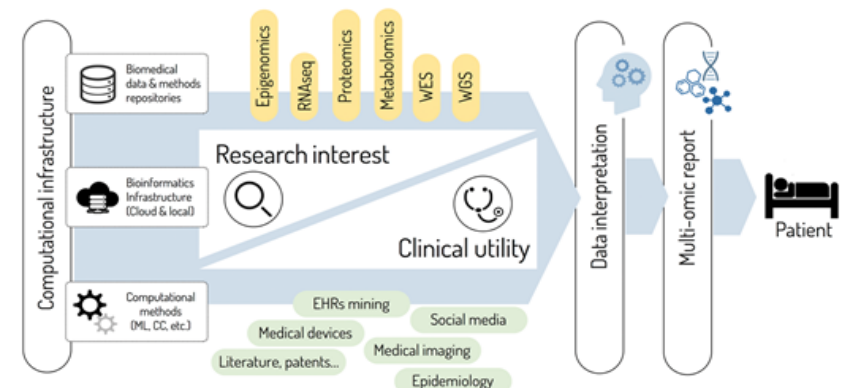
Briefings in Bioinformatics, 2017, 1–15

doi: 10.1093/bib/bbx144
Paper

Precision medicine needs pioneering clinical bioinformaticians

Gonzalo Gómez-López, Joaquín Dopazo, Juan C. Cigudosa, Alfonso Valencia and Fátima Al-Shahrour

Corresponding author: Fátima Al-Shahrour, Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro, 3. ES-28029 Madrid, Spain. Tel: +34 91 732 80 00; E-mail: falshahrour@cnio.es



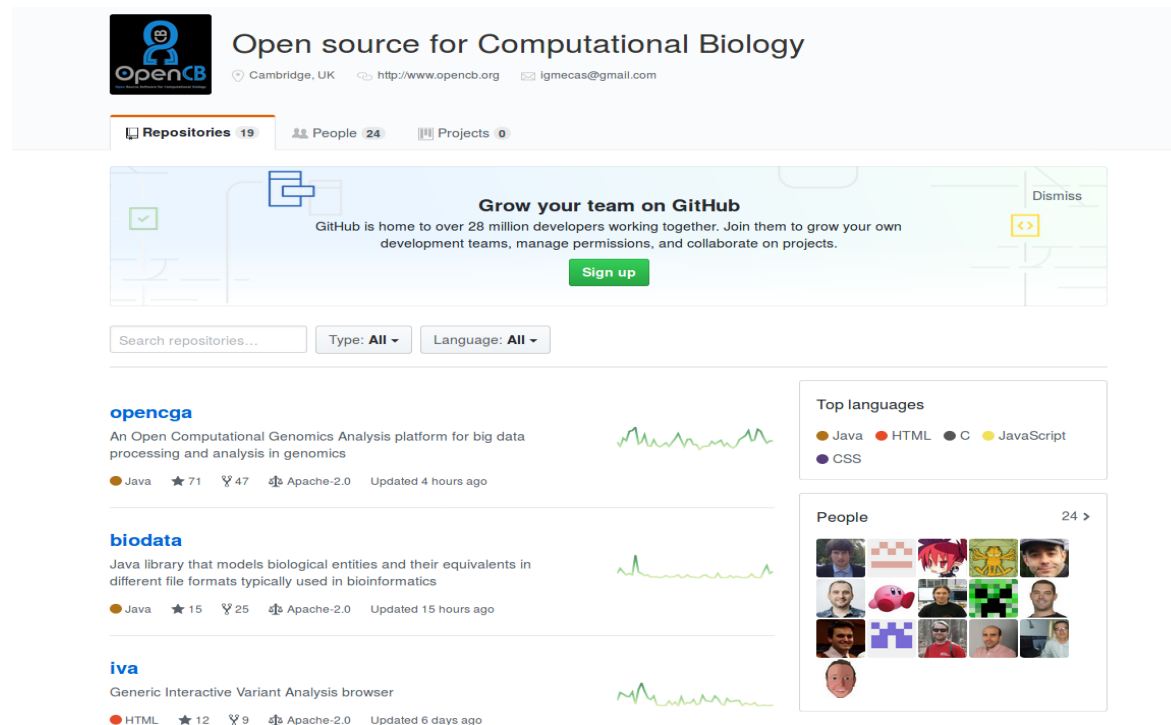
Overview

- Introduction
- Big Data in Genomics
- **OpenCB: Open source initiative for Computational Biology**
- A case study: Personalized Medicine Module (MMP)

OpenCB

Open source initiative for Computational Biology

- OpenCB is a collaborative project that **aims** to design and develop high-performance and scalable solutions for genomic big data analysis using **most modern computing technologies**.
- OpenCB is a collaborative project with more than 15 active developers and data analysts and more than 12 repositories (<http://www.opencb.org>)
- No one computing programming language oriented: BioPerl, BioPython, Bioconductor, ... Good software solutions may use different languages and technologies to solve different problems and use cases
- So far, is where all the software we develop is being released. About 15 active committers. Available as open-source at GitHub <https://github.com/opencb>



The screenshot shows the OpenCB GitHub repository page. At the top, the OpenCB logo is displayed next to the text "Open source for Computational Biology". Below this, the location "Cambridge, UK" and contact information "http://www.opencb.org" and "lgmecas@gmail.com" are listed. The page features a navigation bar with "Repositories 19", "People 24", and "Projects 0". A prominent banner for GitHub states "Grow your team on GitHub" and includes a "Sign up" button. Below the banner, there are search filters for "Search repositories...", "Type: All", and "Language: All". The main content area lists three repositories: "opencga" (An Open Computational Genomics Analysis platform for big data processing and analysis in genomics), "biodata" (Java library that models biological entities and their equivalents in different file formats typically used in bioinformatics), and "iva" (Generic Interactive Variant Analysis browser). Each repository entry includes its language (Java, HTML, or CSS), star count, fork count, license (Apache-2.0), and update time. On the right side, there are sections for "Top languages" (Java, HTML, C, JavaScript, CSS) and "People" (24 >), which displays a grid of developer avatars.

OpenCB

Some relevant projects

- **biodata** (<https://github.com/opencb/biodata>) and **ga4gh** (<https://github.com/opencb/ga4gh>)
 - Contain all data models (Variant, Alignment...) parsers and converters (*avro*, *protobuf*) for all OpenCB projects
- **CellBase** (<https://github.com/opencb/cellbase>)
 - A NoSQL database the integrates the most relevant biological information about genomic features and proteins, gene expression regulation, etc. **The biological knowledge-base for OpenCB project** . A **Variant Annotation** tool implemented.
 - A high-performance NoSQL implementation, a CLI and web services implemented
- **HPG BigData** (<https://github.com/opencb/hpg-bigdata>)
 - Hadoop-based implementation of data converters (*avro*, *parquet*) and bioinformatic tools (ie. *samtools*)
 - Simple indexing for HBase, Hive and Impala developed
 - C code embedded using JNI to speed-up processing
- **OpenCGA** (<https://github.com/opencb/opencga>)
 - Integrates most of the OpenCB projects to provide a scalable and high-performance platform for the analysis of HT genomic data
 - **OpenCGA Catalog** provides an authenticated environment, files and sample annotations, system audit, ...
 - **OpenCGA Storage** is a plugin-oriented framework that allows to index hundreds of millions of variants for thousands of samples in different storage engines. Stats and annotation implemented..
- **Interactive Variant analysis (IVA)**
 - Web-based variant analysis tool
- **Genome Maps** (<https://github.com/opencb/genome-maps>)
 - A web-based NGS and genome browser: <http://genomemaps.org/>
- Many other related projects for big data analysis and visualization, check <https://github.com/opencb>

OpenCB: CellBase

An integrative database and RESTful Web Service API

- **CellBase** is a comprehensive integrative NoSQL database and a *RESTful Web Service API*, designed to provide a **high-performance and scalable** solution.
- It integrates the most relevant biological information about genomic features and proteins, gene expression regulation, functional annotation, genomic variation...Currently contains more than 2TB of data
- Used by EMBL-EBI, ICGC, GEL, MMP among others

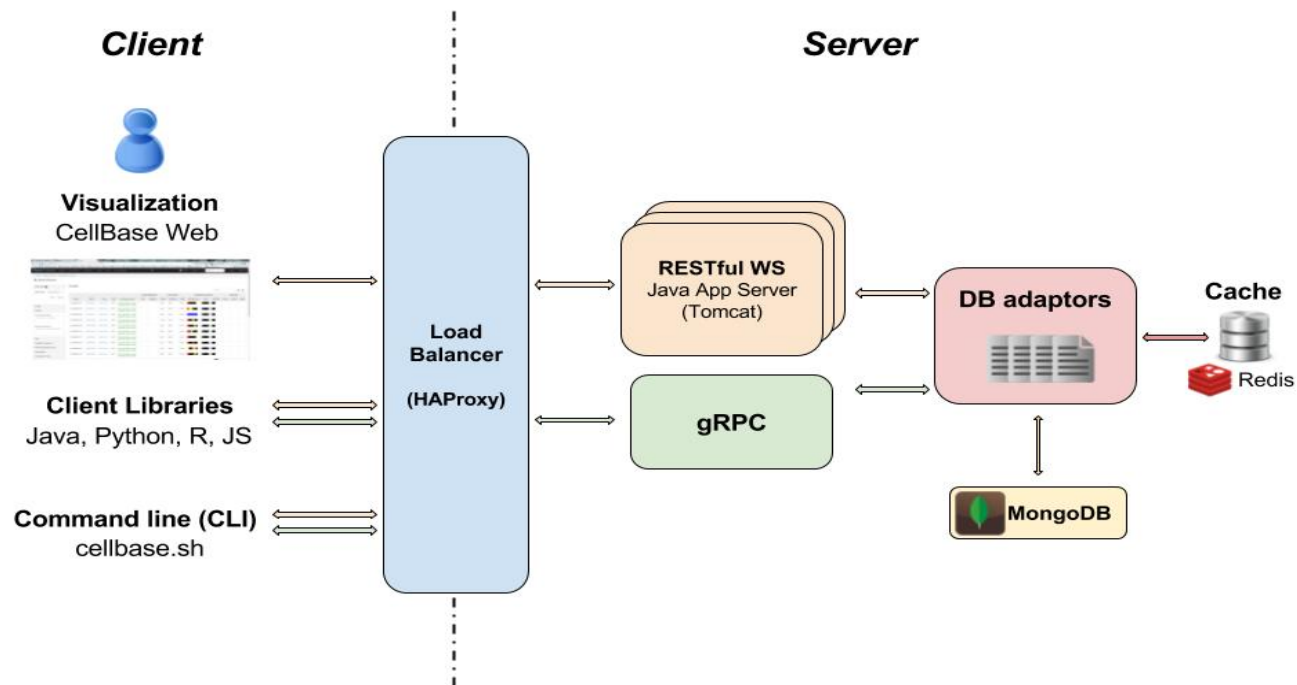
Category	Data source	Version/Date	
		CellBase v3 (March 2015)	CellBase v4 (June 2016)
Core	Ensembl Core	v79	v82
Protein	UniProt	March 2015	Release 2015_10
	InterPro	v50	v54
	Polyphen2/Sift	Ensembl 79	Ensembl v82
Variation	Ensembl Variation	v79	v82
	1000 genomes project		Phase 3 2016-05
	ExAC		0.3.1
	GoNL		Release 5
	UK10K		2016-05
	ESP		2016-05
Regulatory	Ensembl Regulatory	v79	v82
Conservation	PhastCons		June 2016
	PhyloP		June 2016
	GERP++		June 2016

Clinical	ClinVar	March 2015	2016-12
	COSMIC	v71	v79
	HPO		2015-11
	DisGeNET		Version 3.0
Biological Networks	Reactome	v51	June 2016
	IntAct	March 2015	June 2016
Others	DGIdb		2.0
	Gene Expression Atlas		June 2016
	CADD		v1.3

- Official Domain and swagger: <http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/>
- Current version: v4.7.1
- Data sources for homo sapiens:
<http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest/v4/meta/hsapiens/versions?assembly=grch37&limit=-1&skip=-1&skipCount=false&count=false&Output%20format=json>
- Project: <https://github.com/opencb/cellbase> Wiki: <http://docs.opencb.org/display/cellbase/>

OpenCB: CellBase

Architecture



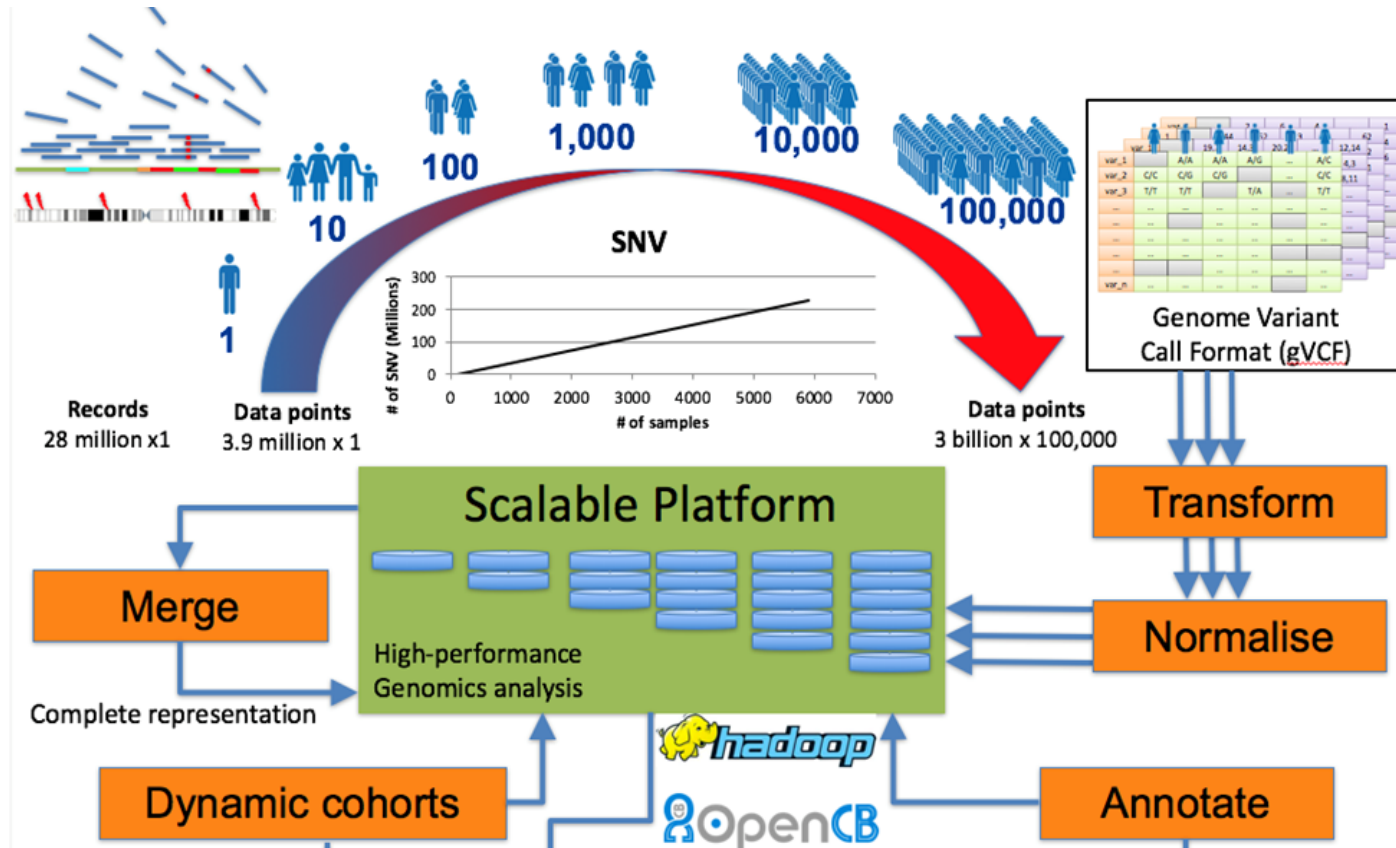
- **Example of variant annotation:**

- <http://bioinfo.hpc.cam.ac.uk/cellbase/webservices/rest/v4/hsapiens/genomic/variant/19:45411941:T:C/annotation>

OpenCB: OpenCGA

Overview and goals

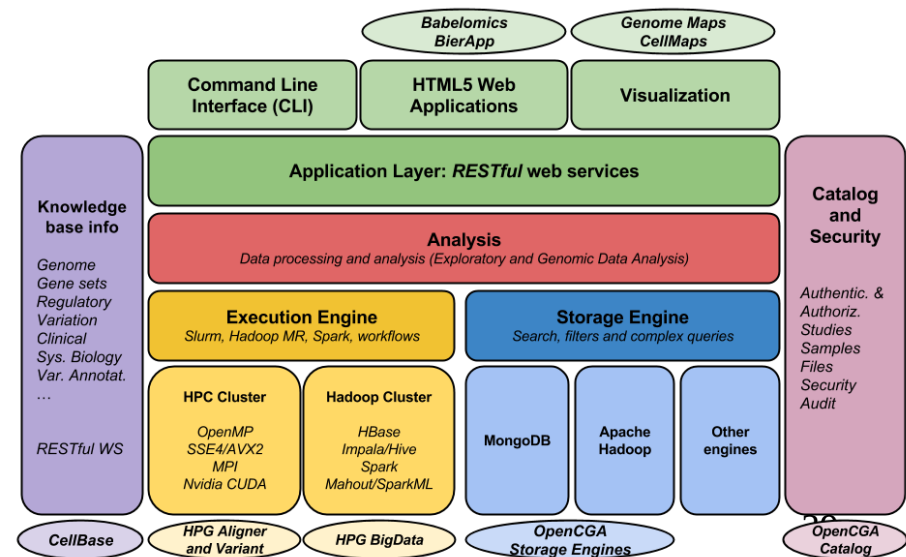
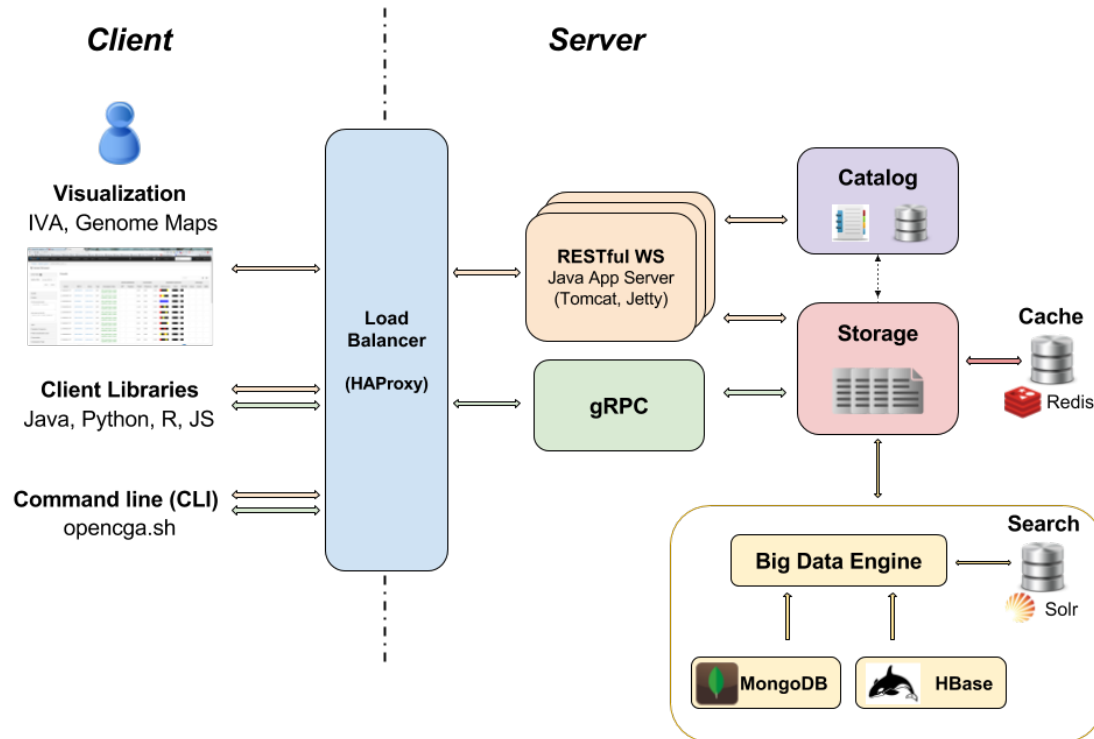
- Open-source Computational Genomics Analysis (**OpenCGA**) aims to provide to researchers and clinicians a **high performance and scalable solution** for genomic big data processing and analysis
- **OpenCGA** is built on OpenCB
- Project: <https://github.com/opencb/opencga> ; Doc: <http://docs.opencb.org/display/opencga>
- Currently, the fastest and more powerful genomic engine in the world. Used in Genomics England (GEL) and MMP



OpenCB: OpenCGA

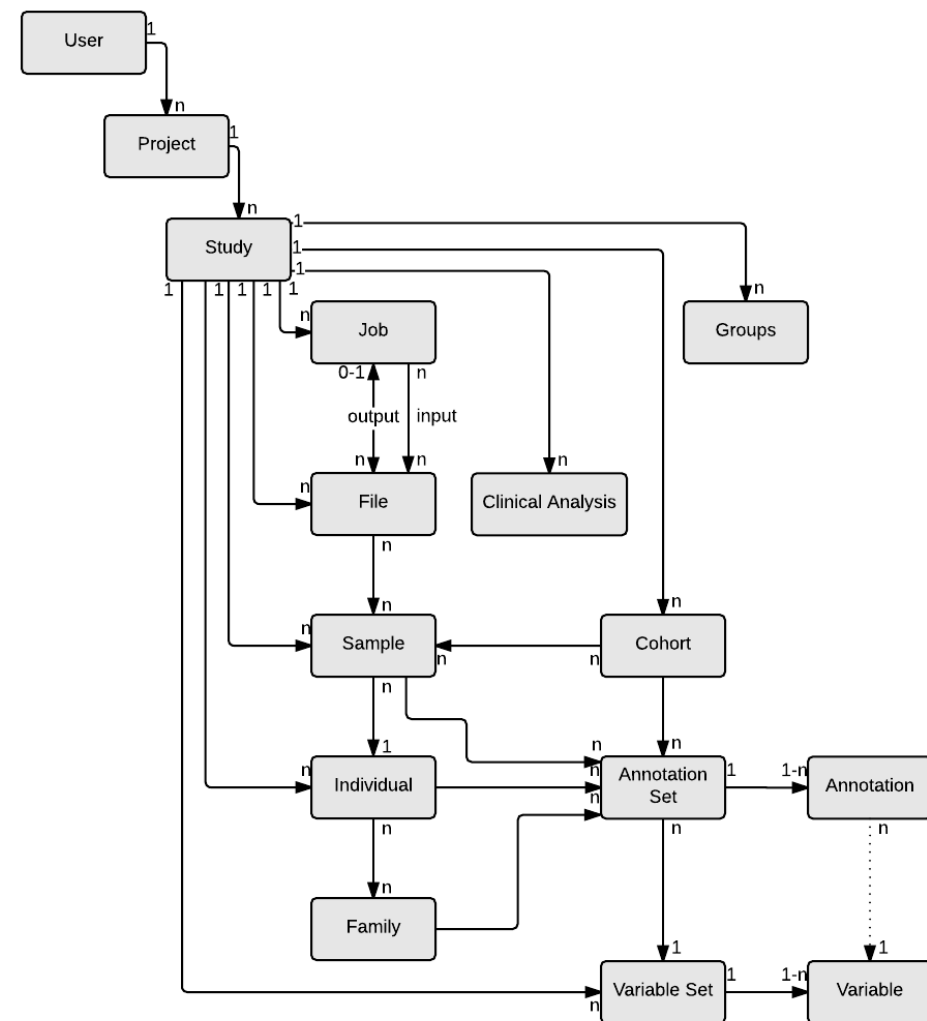
Overview and goals

- Main features:
 - High-performance and scalable variant storage and index that allow to load and merge VCF/gVCF files
 - Annotate and calculate statistics for all the variants
 - Client libraries developed in Java, Python, R and Javascript
 - Clinical interpretation analysis of samples and families
 - Integrated Catalog keeps track of users, files, jobs, clinical data...
 - Interactive web-based data mining tool based on IVA



OpenCB: OpenCGA Catalog, a *metadata* database

- **OpenCGA Catalog** provides a user authentication and authorization environment, implements *sample annotation* database, tracks all files and job, audit, ...
- **Authentication & authorization**: *roles* definition, detailed ACL system implemented to share sample, file, jobs, ...
- **File and jobs**: file and jobs are tracked, stats calculated
- **Sample annotations** is one of the main features:
 - Allow complex queries and aggregations
 - Allow to detect bias and other problems with the data
 - Cohort definitions
- **Audit**: all actions (login, data indexing, ...) and queries are audited
- Allow to the different big data storage engines to perform optimizations
- Implements a job launcher and a execution monitor
- Data models at <https://github.com/opencb/opencga/tree/develop/opencga-catalog/src/main/java/org/opencb/opencga/catalog/models>

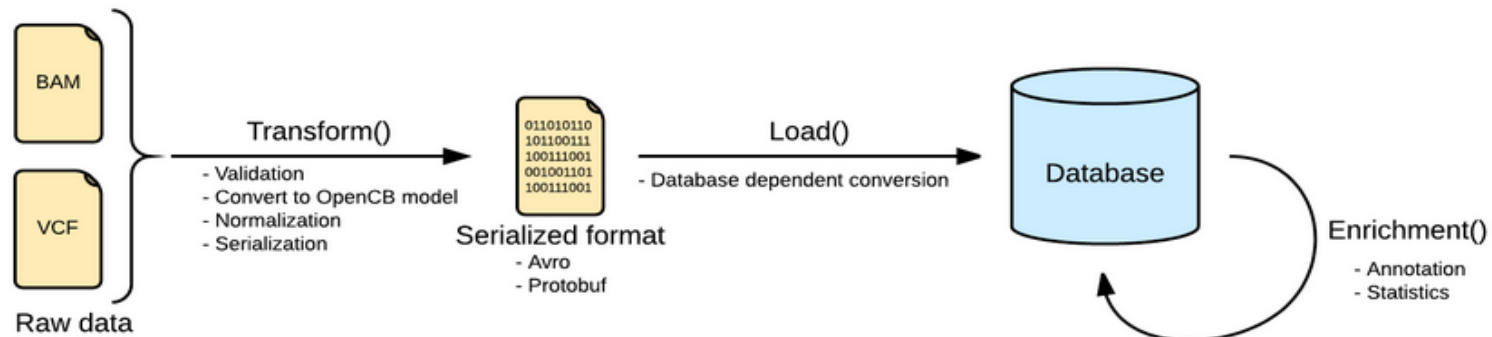


OpenCB: OpenCGA

Storage Engines

OpenCGA Storage provides a *pluggable* Java framework for storing and querying alignment and variant data

- Two default implementations: **MongoDB** and **Hadoop** for huge performance and scalability ~ hundreds of thousands of samples



Variant Index Pipeline

1. Transform

- Validation: VCF files are read using the library HTSJDK → validation
- Files are converted to Biodata models. The metadata is stored into a file serializing in json a single instance of the biodata model “VariantSource” (header and some general stats). The real variants data is serialized in file (avro) with a set of variant records described as the biodata model “Variant”
- Normalization → unify the variants representation, since the VCF specification allows multiple ways of referring to a variant and some ambiguities.

2. Load

- Variants are loaded into the database and merged with the existent ones.

3. Enrichment

- Variant annotation (from Cellbase or read from local files provided by the user)
- Stats calculation useful for filtering variants

OpenCB: OpenCGA

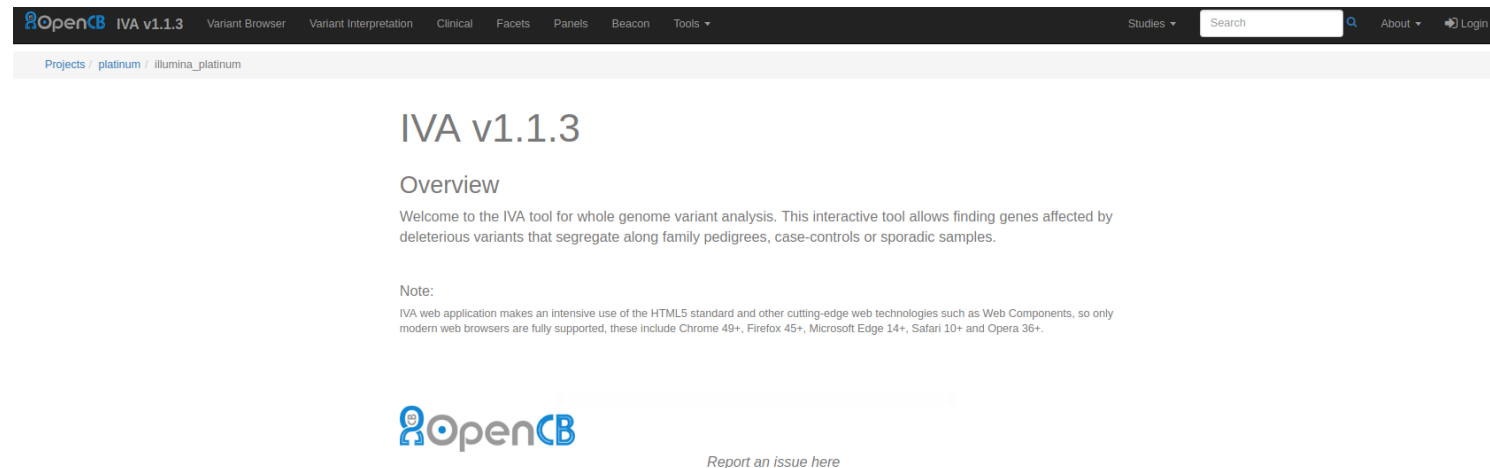
User interfaces: REST, command lines and client libs

- **RESTful API:**
 - more than 160 web services developed for **Catalog** (e.g. search *samples* or *files*) and **Analysis** (e.g. query *variants* or fetch *alignment reads*)
 - Other features: authentication and authorization, Swagger documentation (<http://bioinfo.hpc.cam.ac.uk/hgva/webservices/>), ...
- ***opencga.sh*** command line: you can query remote REST or gRPC services after authentication
- **Client libs:** allow to easily query REST web services. At the moment Java, Python, R and Javascript are fully developed and with similar level of functionality (e.g. authentication, parallel fetches, error support, ...).
- **Current version:** v1.4.1
 - Roadmap: <http://docs.opencb.org/display/opencga/Roadmap>

OpenCB: Interactive Variant Analysis (IVA)

An interactive web-based variant analysis suite

- A collaborative project to implement a rich and interactive web-based variant analysis tool on top of **OpenCGA**, currently supports:
 - Variant Browser
 - Prioritization
 - Clinical Analysis
 - Panels
 - Tools
 - ...



OpenCB: Interactive Variant Analysis (IVA)

An interactive web-based variant analysis suite

Projects / test@enod_solr / study_solr

Variant Interpreter

Search

Clear

No filters selected

Filters

Study and Cohorts

Samples

Select Sample:

Search name...

+

Select Sample Genotypes:

No samples selected

Sample Genotypes Query Options

☒ Approximate Count

☐ Select all multi-allelic variants

Genomic

Population Frequency

Deleteriousness

Conservation

Consequence Type

Gene Ontology

Phenotype-Disease

VCF Metrics

Table Result

Genome Browser (Beta)

Showing 1-10 of 139,840 variants

Save Interpretation...

Download

Share

	Variant	SNP Id	Genes	Type	Consequence Type	Deleteriousness			Conservation			Population Frequencies			Phenotype	
						SIFT	Polyphen	CADD	PhyloP	PhastCons	GERP	1000 Genomes	gnomAD Genomes	ESP6500	Clinvar	Cosmic
☐	2:179575511 C/T	rs72648998	TTN,TTN-AS1	SNV	missense_variant	-	-	23.50	0.655	0.925	5.300				Cardiovascular phenotype	carcinoma
☐	2:179576596 A/T	rs2742331	TTN,TTN-AS1	SNV	intron_variant regulatory_region_variant	-	-	0.44	-2.157	0.000	-3.230				-	-
☐	2:179578109 -/ACAAA	rs71393436	TTN,TTN-AS1	INDEL	intron_variant regulatory_region_variant	-	-	-	0.528	0.093	1.260				-	-
☐	2:179578730 G/A	rs2562839	TTN,TTN-AS1	SNV	synonymous_variant	-	-	15.67	-0.254	0.990	1.840				Cardiovascular phenotype	haematopoietic neoplas
☐	2:179579093 T/C	rs12693164	TTN,TTN-AS1	SNV	missense_variant	-	-	0.06	-2.240	0.345	2.250				cardiovascular phenotype Cardiovascular phenotype	haematopoietic neoplas
☐	2:179579212 T/C	rs2562838	TTN,TTN-AS1	SNV	synonymous_variant	-	-	2.39	-0.119	0.862	4.180				cardiovascular phenotype Cardiovascular phenotype	haematopoietic neoplas
☐	2:179579822 T/A	rs2562836	TTN,TTN-AS1	SNV	synonymous_variant	-	-	12.76	-0.257	0.790	-4.710				Cardiovascular phenotype cardiovascular phenotype	haematopoietic neoplas
☐	2:179580093 A/C	rs12622914	TTN,TTN-AS1	SNV	intron_variant 2KB_downstream_variant regulatory_region_variant	-	-	3.74	-0.213	0.121	-3.450				-	-
☐	2:179580583 T/C	rs2562834	TTN,TTN-AS1	SNV	intron_variant 2KB_downstream_variant regulatory_region_variant	-	-	4.44	0.533	0.079	-0.892				-	-
☐	2:179582162 C/T	rs62178977	TTN,TTN-AS1	SNV	intron_variant regulatory_region_variant	-	-	0.00	-0.302	0.001	-7.590				-	-

Showing 1 to 10 of 139840 rows

10 rows per page

1

2

3

4

5

...

13984

Other OpenCB projects: HGVA

The Human Genomic Variation Archive

- Integrates variants from the main reference human projects
 - Adds valuable information such as variant annotation: consequence types, population frequencies, protein effect predictions, variant-associated phenotype, etc
- Main features:
 - Annotation performed using CellBase
 - Population frequencies calculated, including populations and super-populations
 - Data is indexed using OpenCGA
 - Rich interactive web-based data mining tool based on IVA

Project name (alias)	Studies		HGVA Version (date)	
	Name	Alias	v1 (Dec. 2016)	v2 (Jan. 2018)
Reference GRCh37 (reference_grch37)	1000 Genomes Project GRCh37	1kG_phase3	Phase 3 2016-05	Phase 3 2016-05
	Exome Sequencing Project (ESP6500)	ESP6500	2016-05	2016-05
	Exome Aggregation Consortium (ExAC)	EXAC	0.3.1 2016-05	0.3.1 2016-05
	Genome of the Netherlands (GoNL)	GONL	Release 5 2016-05	Release 5 2016-05
	UK10K Project	UK10k	2016-05	2016-05
	DiscovEHR	DISCOVEHR	-	
	Genome Aggregation Database (gnomAD Exomes)	GNOMAD_EXOMES	-	
	Genome Aggregation Database (gnomAD Genomes)	GNOMAD_GENOMES	-	
	Spanish Medical Genome Project (MGP)	MGP	2016-12	2016-12

Reference GRCh38 (reference_grch38)	1000 Genomes Project GRCh38	1kG_phase3	Phase 3 2016-10	Phase 3 2016-10
	ESP6500	ESP6500	-	
	UK10K Project (*)	UK10K	-	
	DiscovEHR (*)	DISCOVEHR	-	
	Genome Aggregation Database (gnomAD Exomes) (*)	GNOMAD_EXOMES	-	
Cancer GRCh37 (cancer_grch37)	Genome Aggregation Database (gnomAD Genomes) (*)	GNOMAD_GENOMES	-	
	QIMR Berghofer Melanoma	QIMR_Berghofer_Melanoma	2016-12	2016-12
	Chronic Myeloid Leukemia - Russian Academy of Medical Sciences	RAMS_CML	2016-12	2016-12
Platinum (platinum)	Illumina Platinum	illumina_platinum	2015-08	2015-08

(*) Liftover carried out by Genomics England (GEL)

Other OpenCB projects: HGVA

The Human Genomic Variation Archive

OpenCB HGVA v2.0.0 Variant Browser Facets (New) Beacon Studies Search About

Projects / reference_grch37 / UK10K

Variant Browser

Search

Clear No filters selected Filters

Table Result Summary (Beta)

Showing 1-10 of 46,665,970 variants

Download Share

Variant	SNP Id	Genes	Type	Consequence Type	Deleteriousness			Conservation			Population Frequencies			Phenotype	
					SIFT	Polyphen	CADD	PhyloP	PhastCons	GERP	1000 Genomes	gnomAD Genomes	ESP6500	Clinvar	Co
2:179588838 G/A	rs202089818	TTN,RP11-171I2.1,TTN-AS1	SNV	synonymous_variant	-	-	14.15	0.655	0.993	4.240	■■■■■■■■	■■■■■■■■	■■■■	Dilated cardiomyopathy 1G limb-girdle muscular dystrophy, type 2j	-
2:179588874 T/C	rs375874660	TTN,RP11-171I2.1,TTN-AS1	SNV	splice_region_variant	-	-	9.43	0.533	0.721	-3.180	■■■■■■■■	■■■■■■■■	■■■■	-	-
2:179588908 A/C	rs72648963	TTN,RP11-171I2.1,TTN-AS1	SNV	intron_variant non_coding_transcript_exon_variant regulatory_region_variant	-	-	7.36	-0.256	0.727	-1.160	■■■■■■■■	■■■■■■■■	■■■■	-	-
2:179588996 C/T	rs72648962	TTN,RP11-171I2.1,TTN-AS1	SNV	missense_variant	-	-	20.80	0.655	0.477	6.020	■■■■■■■■	■■■■■■■■	■■■■	dilated cardiomyopathy 1g Cardiovascular phenotype Limb-girdle muscular dystrophy, type 2J	-

Genomic

Population Frequency

Deleteriousness

Conservation

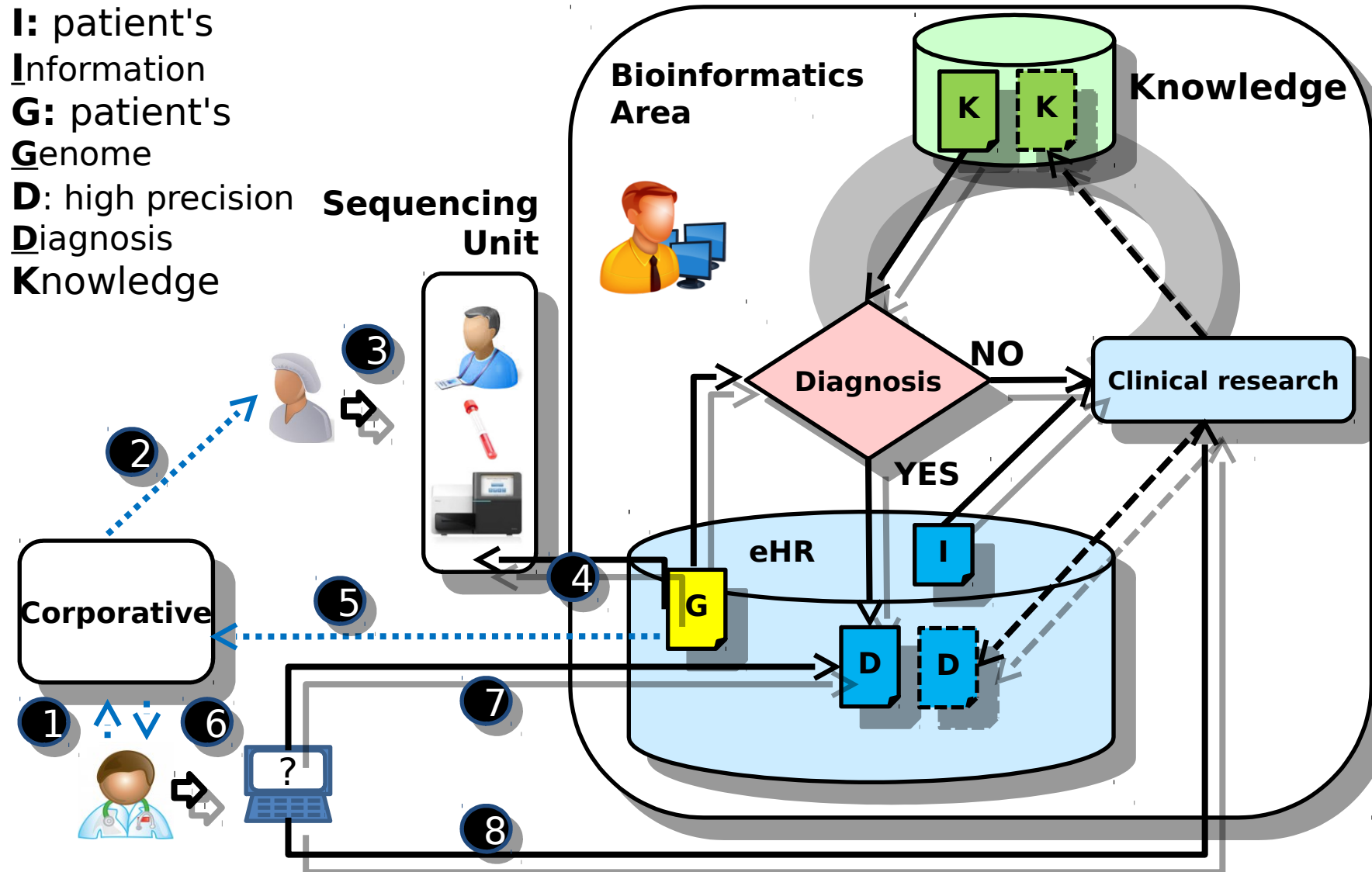
- Published at NAR, 2017
- <http://hgva.opencb.org>

Overview

- Introduction
- Big Data in Genomics
- OpenCB: Open source initiative for Computational Biology
- **A case study: Personalized Medicine Module (MMP)**

Big data in Genomics

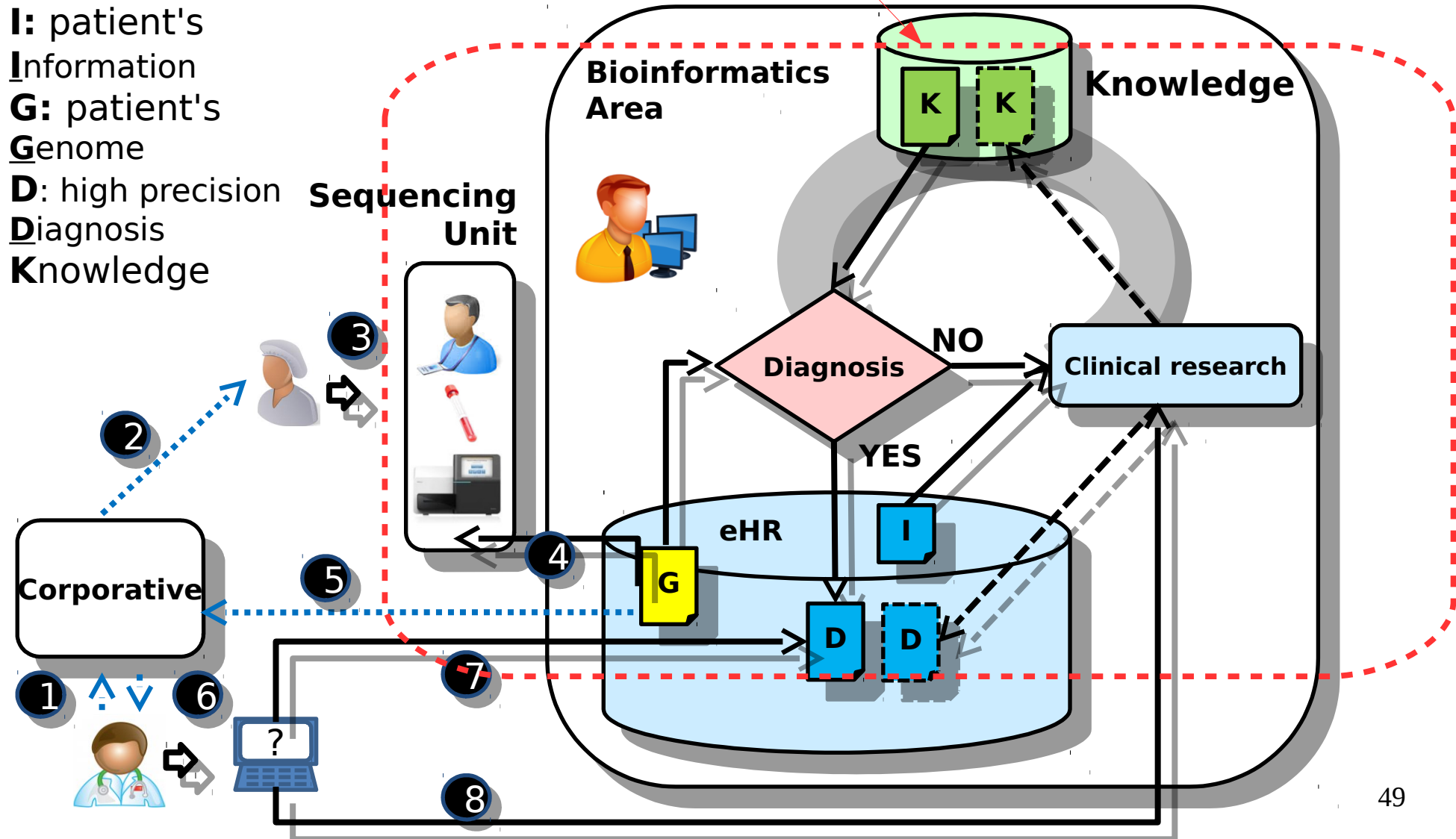
MMP for diagnosis and clinical research within the Andalusian health system



Big data in Genomics

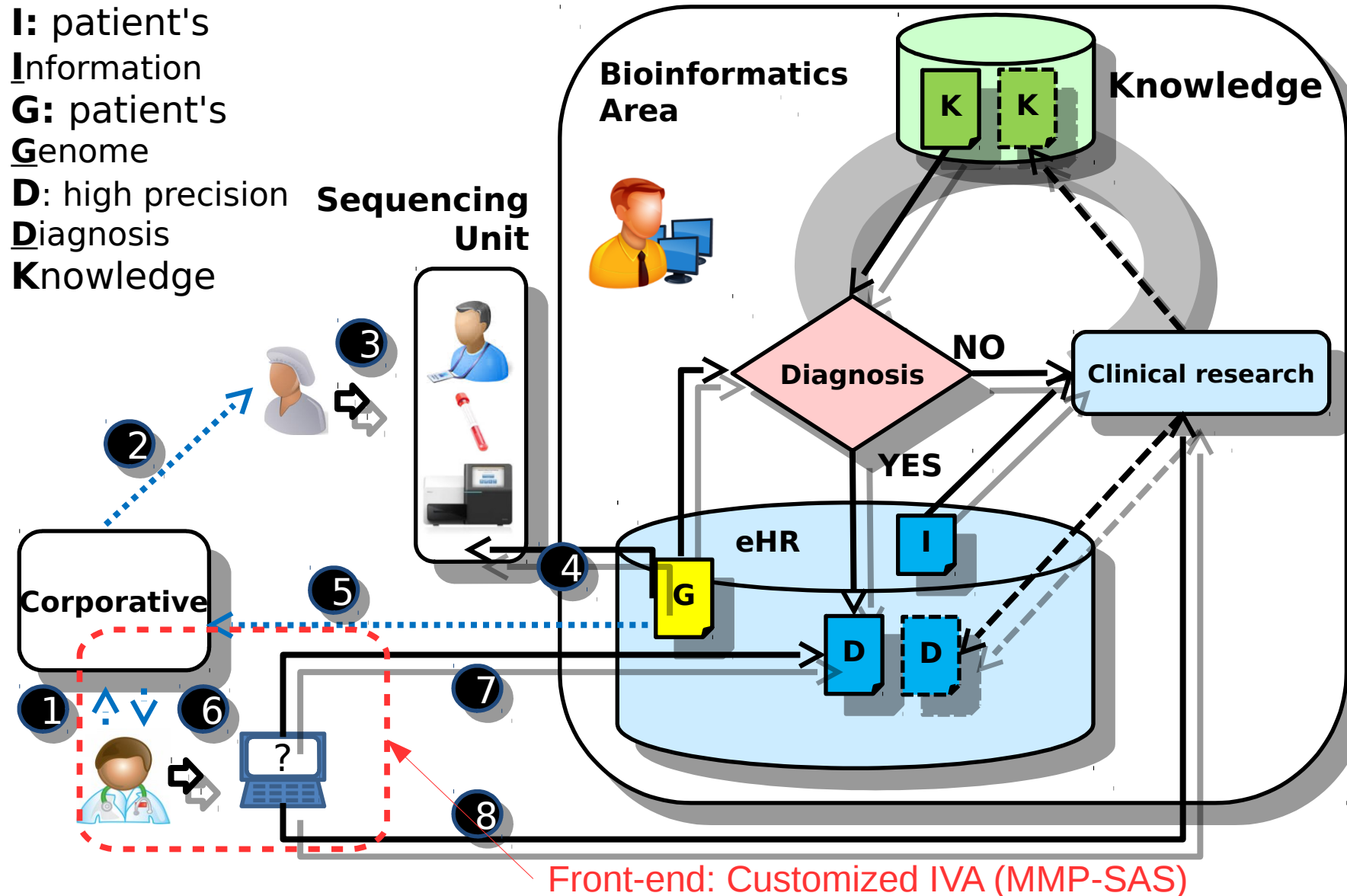
MMP for diagnosis and clinical research within the Andalusian health system

Back-end: OpenCGA



Big data in Genomics

MMP for diagnosis and clinical research within the Andalusian health system



Front-end: MMP-SAS

IVA v1.1.9

Overview

Welcome to the MMP-SAS tool for whole genome variant analysis. This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees, case-controls or sporadic samples.

Supported by:



Servicio Andaluz de Salud
CONSEJERÍA DE SALUD



Clinical Bioinformatics Area



Fundación Progreso y Salud
CONSEJERÍA DE SALUD



Note:

MMP-SAS web application makes an intensive use of the HTML5 standard and other cutting-edge web technologies such as Web Components, so only modern web browsers are fully supported, these include Chrome 49+, Firefox 45+, Microsoft Edge 14+, Safari 10+ and Opera 36+.

[Report an issue here](#)

Front-end: MMP-SAS

Name

 mmp_admin

Password



Sign In

Front-end: MMP-SAS

 MMP-SAS v1.1.9 [Start](#)

 test ▾ [Studies ▾](#) [About ▾](#) [Logout](#)

Projects / [mmp_admin@misc_grch37](#) / miscellaneous Clinical Analysis

Patient Information

Patient ID

Chromosomal Gender

Affectation Status

Life Status

Year of Birth

Year of Birth must be prior to year of Test

Birth Place

Ethnicity

Parental Consanguinity

- ☒ no
☐ yes

Front-end: MMP-SAS

Upload and check file

Type of sample

Select...

Cell line

- ☒ constitutive (germline)
☐ somatic

Hospital

Name

Physician

Name Surname

Year of Test

Select...

Age of Test

Age of Test number

NGS Platform

Select...

Technical data

Additional Information

VCF or gVCF file

...

Clear form

Upload and check sample



Front-end: MMP-SAS

CBA

MMP-SAS v1.1.9

Start

test

Studies

About

Logout

Projects / mmp_admin@misc_grch37 / miscellaneous

Clinical Analysis

Upload

Analysis

Prioritization

Report

Search

Clear

No filters selected

Filters

Sample

Sample name

HG01879, HG01880, HG01881...

By Date

All

Recently Uploaded

Date (Year-Month-Day)

2018

Any

Any

Patient Filters

Individual ID

Smith, Grant ...

HPO

HP:000145

Diagnosis

Smith, Grant ...

Sample Results

Showing 1-4 of 4 samples

☐	Sample	Individual ID	Date	Status	Sex	Diagnosis	HPO	Father	Mother	Cell Line	QC
☐	prueba	43-215	2018-0-08	READY	MALE	-	Vitreoretinal degeneration	-	-	Germline	Download
☐	3_38	3_38	2018-3-13	READY	FEMALE	-	Subretinal deposits	-	-	Germline	Download
☐	33_12550	Fake_patient	2018-4-02	READY	FEMALE	-	Breast carcinoma	-	-	Germline	Download
☐	68_24021	fake_patient2	2018-4-02	READY	FEMALE	-	Mucinous gastric carcinoma	-	-	Germline	Download

Showing 1 to 4 of 4 rows

Analysis Type

Single Analysis

Duo Analysis

Trio Analysis

Family Analysis

Clinical Analysis

Delete

Filters

Analysis ID	Description	Samples	Analysis Type	Definition	Date	Subject	Disease	Action
No matching records found								

55

Front-end: MMP-SAS

Projects / enod@enod_grch37 / agilent_sureselect_v3_p1_1

Clinical Analysis

Upload Analysis Prioritization Report

Search

Clear

No filters selected

Filters

Sample

Sample name

HG01879, HG01880, HG01881...

By Date

☒ All

☐ Recently Uploaded

☐ Date (Year-Month-Day)

2018 - Any - Any

Patient Filters

Individual ID

Smith, Grant ...

HPO

HP:000145

Diagnosis

Smith, Grant ...

Sample Results

Showing 1-3 of 3 samples

☒	Sample	Individual ID	Date	Status	Sex	Diagnosis	HPO	Father	Mother	Cell Line
☒	0044-018-COHO_1	0044-018-COHO_1	2018-0-15	READY	MALE	Unspecified intellectual disabilities	Intellectual disability	-	-	Germline
☒	0044-018-COHO_2	0044-018-COHO_2	2018-0-15	READY	MALE	-	-	-	-	Germline
☒	0044-018-COHO_3	0044-018-COHO_3	2018-0-15	READY	FEMALE	-	-	-	-	Germline

Showing 1 to 3 of 3 rows

Analysis Type



Clinical Analysis

There are unsaved analyses!

Delete

	Analysis ID	Description	Samples	Analysis Type	Definition	Date	Subject	Disease	Action
☐	AN-4163		0044-018-COHO_1,0044-018-COHO_2,0044-018-COHO_3	Trio	Edit	7/05/2018	0044-018-COHO_1	F79	Save

Front-end: MMP-SAS

Family Editor ✕

Name

FA-6031

Description

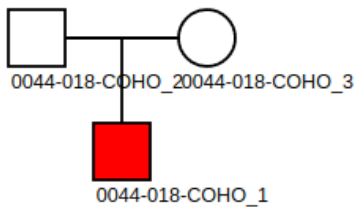
Diseases

⊞ Add custom disease

Unspecified intellectual disabilities(F79)

Intellectual disability(HP:0001249)

Sample ID	Father	Mother	Phenotypes	Deceased	Parental Consanguinity	Sex
0044-018-COHO_1 (0044-018-COHO_1)	0044-018-COHO_2 ▾	0044-018-COHO_3 ▾	☐ F79 ☒ HP:0001249	☐	☐	MALE
0044-018-COHO_2 (0044-018-COHO_2)	Select a sample... ▾	Select a sample... ▾	☐ F79 ☐ HP:0001249	☐	☐	MALE
0044-018-COHO_3 (0044-018-COHO_3)	Select a sample... ▾	Select a sample... ▾	☐ F79 ☐ HP:0001249	☐	☐	FEMALE



Front-end: MMP-SAS

Projects / enod@enod_grch37 / agilent_sureselect_v3_p1_1

Clinical Analysis

 Upload  Analysis  Prioritization  Report

Clinical Analysis AN-3622

Select Recent Analysis:

☒ AN-3622

Select analysis: [Browse...](#)

AN-3622 Clinical Analysis Summary

Clinical Analysis

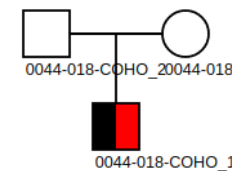
Name: AN-3622
Disease: Unspecified intellectual disabilities (F79)
Type: TRIO
Date: 15 Jan 18

Sample

Name: 0044-018-COHO_1
Somatic: false
Phenotypes: Unspecified intellectual disabilities (F79)
Intellectual disability (HP:0001249)

Subject

Name: 0044-018-COHO_1
Sex (karyotype): MALE (XY)
Date of Birth: 1920 (UNKNOWN)
Pedigree:



Phenotypes: Unspecified intellectual disabilities (F79)
Intellectual disability (HP:0001249)

Interpretation Algorithms

Several prioritization and interpretation algorithms are available, you can choose an **interactive** tool:

☒ Interactive Prioritization (based on [TEAM](#) paper)

[OK](#)

Front-end: MMP-SAS

Projects / enod@enod_grch37 / agilent_sureselect_v3_p1_1

Clinical Analysis

Upload

Analysis

Prioritization

Report

Prioritization

Search

Clear

genotype

Filters

Table Result

Genome Browser (Beta)

Back to Clinical Analysis Selector

Save Interpretation...

Study and Cohorts

Samples

Select Sample Genotypes:

Inheritance Mode: Autosomal Dominant

Sample

0/0

0/1

1/1

J.

0044-018-COHO_1

☐

☒

☐

☐

0044-018-COHO_2

☒

☐

☐

☐

0044-018-COHO_3

☒

☐

☐

☐

Any allele in multi-allelic variants

Cohort Stats (MAF)

Project not found

Studies Filter

In (AND)

☒ agilent_sureselect_v3_p1_1

Genomic

Population Frequency

Deleteriousness

Conservation

Consequence Type

Showing 1-10 of 1,810 variants

Download

Share

	Variant	SNP Id	Genes	Type	Samples			Consequence Type	Deleteriousness			Conservation			Population Frequencies		
					0044-018-COHO_1	0044-018-COHO_2	0044-018-COHO_3		SIFT	Polyphen	CADD	PhyloP	PhastCons	GERP	1000 Genomes	gnomAD Genomes	ESP650
☐	1:1509825 A/G	rs6666293	SSU72,AL645728.1	SNV	A / G	A / A	A / A	intron_variant 2KB_upstream_variant regulatory_region_variant TF_binding_site_variant	-	-	5.09	-2.220	0.049	-4.570			
☐	1:5940092 CA/-	rs140612295	NPHP4	INDEL	CA / -	CA / CA	CA / CA	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	0.113	0.005	-0.338			
☐	1:11156046 T/-	rs374028571	EXOSC10,RP4-635E18.6	INDEL	T / -	T / T	T / T	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	-1.193	0.108	1.070			
☐	1:17570466 T/C	rs2293921	PADI1	SNV	T / C	T / T	T / T	intron_variant regulatory_region_variant	-	-	3.54	-0.282	0.003	-0.769			
☐	1:16974869 G/C	rs4570422	CROCCP2,MST1P2	SNV	G / C	G / G	G / G	non_coding_transcript_exon_variant regulatory_region_variant	-	-	6.23	0.000	0.000	0.000			
☐	1:19611487 A/T	rs59348345	AKR7A3	SNV	A / T	A / A	A / A	intron_variant	-	-	0.39	-1.496	0.017	-1.350			
☐	1:19612477 C/T	rs2231198	AKR7A3	SNV	C / T	C / C	C / C	missense_variant	tolerated	benign	18.62	0.194	0.263	3.040			
☐	1:25256444 G/A	rs71514255	RUNX3	SNV	G / A	G / G	G / G	intron_variant 2KB_upstream_variant regulatory_region_variant	-	-	12.75	0.186	0.370	2.020			
☐	1:29037241 -/T	rs11415888	GMEB1	INDEL	- / T	- / -	- / -	intron_variant regulatory_region_variant	-	-	-	0.650	0.067	0.875			
☐	1:36884530 G/A	rs72663495	OSCP1,SNORA63	SNV	G / A	G / G	G / G	intron_variant 2KB_downstream_variant 2KB_upstream_variant	-	-	4.83	-0.965	0.086	0.613			

Showing 1 to 10 of 1810 rows

10 rows per page

1

2

3

4

5

...

181

59

Front-end: MMP-SAS

0044-018-COHO_2

☒☐☐☐

0044-018-COHO_3

☒☐☐☐

☐ Any allele in multi-allelic variants

Cohort Stats (MAF)

Project not found

Studies Filter

In (AND)

☒ *agilent_sureselect_v3_p1_1*

Genomic

Population Frequency

Deleteriousness

Conservation

Consequence Type

Gene Ontology

Phenotype-Disease

VCF Metrics

☐	1:5940092 CA/-	rs140612295	NPHP4	INDEL	CA / -	CA / CA	CA / CA	regulatory_region_variant TF_binding_site_variant											
☐	1:11156046 T/-	rs374028571	EXOSC10,RP4-635E18.6	INDEL	T / -	T / T	T / T	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	0.113	0.005	-0.338					
☐	1:17570466 T/C	rs2293921	PADI1	SNV	T / C	T / T	T / T	intron_variant regulatory_region_variant	-	-	3.54	-0.282	0.003	-0.769					
☐	1:16974869 G/C	rs4570422	CROCCP2,MST1P2	SNV	G / C	G / G	G / G	non_coding_transcript_exon_variant regulatory_region_variant	-	-	6.23	0.000	0.000	0.000					
☐	1:19611487 A/T	rs59348345	AKR7A3	SNV	A / T	A / A	A / A	intron_variant	-	-	0.39	-1.496	0.017	-1.350					
☐	1:19612477 C/T	rs2231198	AKR7A3	SNV	C / T	C / C	C / C	missense_variant	tolerated	benign	18.62	0.194	0.263	3.040					
☐	1:25256444 G/A	rs71514255	RUNX3	SNV	G / A	G / G	G / G	intron_variant 2KB_upstream_variant regulatory_region_variant	-	-	12.75	0.186	0.370	2.020					
☐	1:29037241 -/T	rs11415888	GMEB1	INDEL	- / T	- / -	- / -	intron_variant regulatory_region_variant	-	-	-	0.650	0.067	0.875					
☐	1:36884530 G/A	rs72663495	OSCP1,SNORA63	SNV	G / A	G / G	G / G	intron_variant 2KB_downstream_variant 2KB_upstream_variant	-	-	4.83	-0.965	0.086	0.613					

Showing 1 to 10 of 1810 rows

10 rows per page

<

1

2

3

4

5

...

181

>

Variant: 1:19612477:C:T

Advanced Annotation

File Metrics

Beacon Network

Summary

Consequence Types (2)

Population Frequencies (70)

Variant Trait Association ()

Gene Trait Association ()

ID

rs2231198

HGVS

ENST00000361640(ENSG00000162482):c.412G>A

Alleles

C/T

Location

1:19612477

Most Severe Consequence

missense_variant (Gene : AKR7A3, Transcript : ENST00000361640)

Type

Most Severe Deleterious Score

Sift tolerated (Gene:AKR7A3, Transcript: ENST00000361640)
Polyphen benign (Gene:AKR7A3, Transcript: ENST00000361640)
CADD Scaled 18.62

Front-end: MMP-SAS

Genomic

Population Frequency

Deleteriousness

Protein Substitution Score

SIFT

Polyphen

Logical Operator

CADD

Raw

Scaled

Conservation

Conservation Score

PhyloP

PhastCons

Gerp

Logical Operator

Consequence Type

Gene Ontology

Phenotype-Disease

VCF Metrics

☐	1:19612477 C/T	rs2231198	AKR7A3	SNV	C / T	C / C	C / C	missense_variant	tolerated	benign	18.62	0.194	0.263	3.040			
☐	1:25256444 G/A	rs71514255	RUNX3	SNV	G / A	G / G	G / G	intron_variant 2KB_upstream_variant regulatory_region_variant	-	-	12.75	0.186	0.370	2.020			
☐	1:29037241 -T	rs11415888	GMEB1	INDEL	- / T	- / -	- / -	intron_variant regulatory_region_variant	-	-	-	0.650	0.067	0.875			
☐	1:36884530 G/A	rs72663495	OSCP1,SNORA63	SNV	G / A	G / G	G / G	intron_variant 2KB_downstream_variant 2KB_upstream_variant	-	-	4.83	-0.965	0.086	0.613			

Showing 1 to 10 of 1810 rows

10 rows per page

1

2

3

4

5

...

181

Variant: 1:19612477:C:T

Advanced Annotation

File Metrics

Beacon Network

Summary

ID

rs2231198

HGVS

ENST00000361640(ENSG00000162482):c.412G>A

Alleles

C/T

Location

1:19612477

Most Severe Consequence Type

missense_variant (Gene : AKR7A3, Transcript : ENST00000361640)

Most Severe Deleterious Score

Sift tolerated (Gene:AKR7A3, Transcript: ENST00000361640)
Polyphen benign (Gene:AKR7A3, Transcript: ENST00000361640)
CADD Scaled 18.62

Front-end: MMP-SAS

Consequence Type

Select SO terms

Loss-of-Function (LoF) terms:
☐ LoF terms

Consequence Type terms:
☐ Intergenic
 ☐ upstream_gene_variant (SO:0001631)
 ☐ 2KB_upstream_variant (SO:0001636)
 ☐ downstream_gene_variant (SO:0001637)
 ☐ 2KB_downstream_variant (SO:0002020)
 ☐ intergenic_variant (SO:0001628)
☐ Regulatory
 ☐ mature_miRNA_variant (SO:0001620)
 ☐ regulatory_region_ablation (SO:0001638)
 ☐ regulatory_region_amplification (SO:0001639)
 ☐ regulatory_region_variant (SO:0001640)
 ☐ TF_binding_site_variant (SO:0001788)
 ☐ TFBS_ablation (SO:0001895)
 ☐ TFBS_amplification (SO:0001892)
☐ Coding
 ☐ coding_sequence_variant (SO:0001515)
 ☐ feature_elongation (SO:0001907)
 ☐ feature_truncation (SO:0001906)
 ☐ frameshift_variant (SO:0001589)
 ☐ incomplete_terminal_codon_variant (SO:0001822)
 ☐ inframe_deletion (SO:0001821)
 ☐ inframe_insertion (SO:0001821)
 ☐ missense_variant (SO:0001583)
 ☐ NMD_transcript_variant (SO:0001621)
 ☐ protein_altering_variant (SO:0001818)
 ☐ synonymous_variant (SO:0001819)
 ☐ start_lost (SO:0002012)
 ☐ stop_gained (SO:0001587)
 ☐ stop_lost (SO:0001578)
 ☐ stop_retained_variant (SO:0001567)
☐ Non-coding
 ☐ 3_prime_UTR_variant (SO:0001624)
 ☐ 5_prime_UTR_variant (SO:0001623)
 ☐ intron_variant (SO:0001627)
 ☐ non_coding_transcript_exon_variant (SO:0001628)
☐ Splice
 ☐ splice_acceptor_variant (SO:0001574)
 ☐ splice_donor_variant (SO:0001575)
 ☐ splice_region_variant (SO:0001630)
☐ transcript_ablation (SO:0001893)
☐ transcript_amplification (SO:0001889)

☐	1:1509825 A/G	rs6666293	SSU72,AL645728.1	SNV	A / G	A / A	A / A	intron_variant 2KB_upstream_variant regulatory_region_variant TF_binding_site_variant	-	-	5.09	-2.220	0.049	-4.570			
☐	1:5940092 CA/-	rs140612295	NPHP4	INDEL	CA / -	CA / CA	CA / CA	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	0.113	0.005	-0.338			
☐	1:11156046 T/-	rs374028571	EXOSC10,RP4-635E18.6	INDEL	T / -	T / T	T / T	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	-1.193	0.108	1.070			
☐	1:17570466 T/C	rs2293921	PADI1	SNV	T / C	T / T	T / T	intron_variant regulatory_region_variant	-	-	3.54	-0.282	0.003	-0.769			
☐	1:16974869 G/C	rs4570422	CROCCP2,MST1P2	SNV	G / C	G / G	G / G	non_coding_transcript_exon_variant regulatory_region_variant	-	-	6.23	0.000	0.000	0.000			
☐	1:19611487 A/T	rs59348345	AKR7A3	SNV	A / T	A / A	A / A	intron_variant	-	-	0.39	-1.496	0.017	-1.350			
☒	1:19612477 C/T	rs2231198	AKR7A3	SNV	C / T	C / C	C / C	missense_variant	tolerated	benign	18.62	0.194	0.263	3.040			
☐	1:25256444 G/A	rs71514255	RUNX3	SNV	G / A	G / G	G / G	intron_variant 2KB_upstream_variant regulatory_region_variant	-	-	12.75	0.186	0.370	2.020			
☐	1:29037241 -/T	rs11415888	GMEB1	INDEL	- / T	- / -	- / -	intron_variant regulatory_region_variant	-	-	-	0.650	0.067	0.875			
☐	1:36884530 G/A	rs72663495	OSCP1,SNORA63	SNV	G / A	G / G	G / G	intron_variant 2KB_downstream_variant 2KB_upstream_variant	-	-	4.83	-0.965	0.086	0.613			

Showing 1 to 10 of 1810 rows 10 rows per page < 1 2 3 4 5 ... 181 >

Variant: 1:19612477:C:T

Advanced Annotation

File Metrics

Beacon Network

Summary

Consequence Types (2)

Population Frequencies (70)

Variant Trait Association ()

Gene Trait Association ()

ID

rs2231198

HGVIS

ENST00000361640(ENSG00000162482):c.412G>A

Alleles

C/T

Location

1:19612477

Most Severe Consequence Type

missense_variant (Gene: AKR7A3, Transcript: ENST00000361640)

Most Severe Deleterious Score

Sift tolerated (Gene:AKR7A3, Transcript: ENST00000361640)
Polyphen benign (Gene:AKR7A3, Transcript: ENST00000361640)
CADD Scaled 18.62

Consequence type: the location or effect of a sequence variation on a transcript

Front-end: MMP-SAS

Gene Ontology

GO Accessions

GO:0000145

Add GO Term Q

Phenotype-Disease

HPO Accessions

HP:0000001, HP:3000079

Add HPO Term Q

ClinVar Accessions

RCV000058226

Full-text search on HPO, ClinVar, protein domains or keywords. Some OMIM and Orphanet IDs are also supported

Full-text search, e.g. *melanoma*

VCF Metrics

VCF FILTER

PASS

PASS

QualByDepth

FisherStrand

RMSMappingQuality

MappingqualityRankSumTest

ReadPosRankSumTest

StrandOddsRatio

								2KB_upstream_variant regulatory_region_variant TF_binding_site_variant											
☐	1:5940092 CA/-	rs140612295	NPHP4	INDEL	CA / -	CA / CA	CA / CA	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	0.113	0.005	-0.338					
☐	1:11156046 T/-	rs374028571	EXOSC10,RP4-635E18.6	INDEL	T / -	T / T	T / T	intron_variant NMD_transcript_variant regulatory_region_variant	-	-	-	-1.193	0.108	1.070					
☐	1:17570466 T/C	rs2293921	PADI1	SNV	T / C	T / T	T / T	intron_variant regulatory_region_variant	-	-	3.54	-0.282	0.003	-0.769					
☐	1:16974869 G/C	rs4570422	CROCCP2,MST1P2	SNV	G / C	G / G	G / G	non_coding_transcript_exon_variant regulatory_region_variant	-	-	6.23	0.000	0.000	0.000					
☐	1:19611487 A/T	rs59348345	AKR7A3	SNV	A / T	A / A	A / A	intron_variant	-	-	0.39	-1.496	0.017	-1.350					
☐	1:19612477 C/T	rs2231198	AKR7A3	SNV	C / T	C / C	C / C	missense_variant	tolerated	benign	18.62	0.194	0.263	3.040					
☐	1:25256444 G/A	rs71514255	RUNX3	SNV	G / A	G / G	G / G	intron_variant 2KB_upstream_variant regulatory_region_variant	-	-	12.75	0.186	0.370	2.020					
☐	1:29037241 -T	rs11415888	GMEB1	INDEL	- / T	- / -	- / -	intron_variant regulatory_region_variant	-	-	-	0.650	0.067	0.875					
☐	1:36884530 G/A	rs72663495	OSCP1,SNORA63	SNV	G / A	G / G	G / G	intron_variant 2KB_downstream_variant 2KB_upstream_variant	-	-	4.83	-0.965	0.086	0.613					

Showing 1 to 10 of 1810 rows

10 rows per page

<

1

2

3

4

5

...

181

>

Variant: 1:19612477:C:T

Advanced Annotation

File Metrics

Beacon Network

Summary

Consequence Types (2)

Population Frequencies (70)

Variant Trait Association (0)

Gene Trait Association (0)

ID

rs2231198

HGVS

ENST00000361640(ENSG00000162482):c.412G>A

Aleles

C/T

Location

1:19612477

Most Severe

missense_variant (Gene : AKR7A3, Transcript : ENST00000361640)

Consequence

Type

Most Severe

Sift

tolerated (Gene:AKR7A3, Transcript: ENST00000361640)

Deleterious

Polyphen

benign (Gene:AKR7A3, Transcript: ENST00000361640)

Score

CADD Scaled

18.62

Front-end: MMP-SAS

OpenC IVA v1.1.3

Variant Browser

Variant Interpretation

Clinical

Facets

Panels

Beacon

Tools

test

Studies

Search

About

Logout

Approximate Count

Select all multi-allelic variants

Genomic

Population Frequency

Deleteriousness

Conservation

Consequence Type

Gene Ontology

Phenotype-Disease

VCF Metrics

☐	10:81318578 T/C	rs17880662	SFTPA2	SNV	T / C	T / T	T / T	intron_variant 2KB_downstream_variant regulatory_region_variant	-	-	0.35	-1.578	0.000	-1
☐	10:82331437 C/T	rs45526031	SH2D4B	SNV	C / T	C / C	C / C	intron_variant 2KB_downstream_variant regulatory_region_variant	-	-	2.27	-0.768	0.002	-5
☐	10:85899115 A/G	-	RP11-338I21.1,GHITM	SNV	A / G	A / A	A / A	non_coding_transcript_exon_variant 2KB_upstream_variant regulatory_region_variant TF_binding_site_variant	-	-	8.05	0.496	0.068	1
☐	10:89118125 C/T	rs77153116	LINC00863,NUTM2D	SNV	C / T	C / C	C / C	missense_variant	deleterious	possibly damaging	18.27	0.175	0.040	0

Showing 61 to 70 of 1517 rows

10 rows per page

<

1

...

6

7

8

...

152

>

Variant: 10:89118125:C:T

Advanced Annotation

File Metrics

Beacon Network

Summary

Consequence Types (4)

Population Frequencies (66)

Variant Trait Association (8)

Gene Trait Association (0)

Clinvar

No ClinVar data available

Cosmic

Mutation Id	Primary Site	Site Subtype	Primary Histology	Histology Subtype	Sample Source	Tumour Origin	Gene Name	Mutation Somatic Status
COSM3752025	thyroid		other	neoplasm			FAM22D_ENST00000381697	Confirmed somatic variant
COSM3752024	breast		carcinoma	ER-positive carcinoma	surgery fresh/frozen		NUTM2A	Confirmed somatic variant
COSM3752024	thyroid		other	neoplasm			NUTM2A	Confirmed somatic variant
COSM3752025	pancreas		carcinoma	ductal carcinoma			FAM22D_ENST00000381697	Confirmed somatic variant
COSM3752024	large intestine	colon	carcinoma	adenocarcinoma			NUTM2A	Confirmed somatic variant
COSM3752025	breast		carcinoma	ER-positive carcinoma	surgery fresh/frozen		FAM22D_ENST00000381697	Confirmed somatic variant
COSM3752024	pancreas		carcinoma	ductal carcinoma			NUTM2A	Confirmed somatic variant
COSM3752025	large intestine	colon	carcinoma	adenocarcinoma			FAM22D_ENST00000381697	Confirmed somatic variant

65

Front-end: MMP-SAS

MMP-SAS v1.1.9
Start

test
Studies
About
Logout

Clinical Analysis

Upload
Analysis
Prioritization
Report

Search

Clear
genotype = 0/1,1/1
gene
Consequence Types

Table Result
Genome Browser (Beta)

Back to Clinical Analysis Selector
Save Interpretation...

Study

Samples
Select Sample Genotypes:
Inheritance Mode:
Select...
Sample
0/0
0/1
1/1
.J.

Genomic

Chromosomal Location
3:444-55555,1:1-100000

Feature IDs (gene, SNPs, ...)
Search for Gene Symt
BRCA2,ENSG00000139618,ENST00000544455,rs28897700

Showing 1-10 of 16 variants

	Variant	SNP Id	Genes	Type	Samples	Consequence Type	Deleteriousness			Conservation			Population Freq	
							SIFT	Polyphen	CADD	PhyloP	PhastCons	GERP	1000 Genomes	gnomAD Genomes
☒	16:68853296 C/G	COSM159879	CDH1,RP11-354M1.2,FTLP14	SNV	C / G	missense_variant	deleterious	probably damaging	23.70	0.555	0.190	4.570	-	-
☐	16:68771372 C/T	rs3743674	CDH1	SNV	T / T	splice_region_variant	-	-	10.26	-1.988	0.005	-1.510		
☐	16:68771419 -/GCCCCAGCCCCGT	rs147838237	CDH1,RNA5SP429	INDEL	GCC... / GCC...	intron_variant NMD_transcript_variant regulatory_region_variant TF_binding_site_variant	-	-	-	-0.339	0.009	-0.555		
☐	16:68857441 T/C	rs1801552	RP11-354M1.2,CDH1,FTLP14	SNV	T / C	synonymous_variant	-	-	6.21	-0.213	0.000	2.580		
☐	16:68867612 T/-	rs369254048	CDH1	INDEL	T / -	3_prime_UTR_variant NMD_transcript_variant 2KB downstream variant	-	-	-	0.533	0.003	-6.620		

Front-end: MMP-SAS

 MMP-SAS v1.1.9 [Start](#)

 test ▾ [Studies ▾](#) [About ▾](#) [Logout](#)

 Clinical Analysis[Upload](#) [Analysis](#) [Prioritization](#) [Report](#) [Prioritization](#)[Search](#)[Clear](#) [genotype = \[redacted\];0/1,1/1](#) [gene ▾](#) [Consequence Types ▾](#)[Table Result](#) [Genome Browser \(Beta\)](#)[Back to Clinical Analysis Selector](#) [Save Interpretation...](#)

Study

Samples

Select Sample Genotypes:

Inheritance Mode: [Select...](#)

Sample	0/0	0/1	1/1	./.
[redacted] 1	☐	☒	☒	☐

Genomic

Chromosomal Location

3:444-55555,1:1-100000

Feature IDs (gene, SNPs, ...)

Save Interpretation

Positive Diagnosis

ID

my_interpretation

Name

sample_interpretation

Description

Description of this interpretation ...

Comment

Add a comment ...

[View](#) [Save](#)

Front-end: MMP-SAS

CBA

MMP-SAS v1.1.9

Start

test

Studies

A

Clinical Analysis

Upload

Analysis

Prioritization

Report

Clinical Analysis AN-5046

Select analysis:

Browse...

Clinical Interpretation

Select Interpretation:

3182_68_interpretation

Interpretation Analysis: my_interpretation

Summary

Sample

Name:
Somatic:
Phenotypes:

false

Mucinous gastric carcinoma (HP:0031498)

Analysis

Name (date):
Type:
Analyst:
Description:

my_interpretation (2 May 18)

SINGLE

Subject

Name:
Sex (karyotype):
Date of Birth:
Parental Consanguinity:
Phenotypes:

fake_patient2

FEMALE (XX)

1982 (ALIVE)

false

Mucinous gastric carcinoma (HP:0031498)

Family

Name:
Phenotypes:
Pedigree:

Software

Name:
Version:
Dependencies:

TEAM ([website](#))

2.0 ([commit](#))

CellBase (v4.5.0)

Variant Filters

genotype:
alternate_frequency:
protein_substitution:
annot-ct:

studies:

/1, 1/1

1kG_phase3:ALL<0.01, MGP:ALL<0.01

sift==deleterious, polyphen==probably damaging

coding_sequence_variant, feature_elongation,

feature_truncation, frameshift_variant,

incomplete_terminal_codon_variant, inframe_deletion,

inframe_insertion, missense_variant,

NMD_transcript_variant, protein_altering_variant,

synonymous_variant, start_lost, stop_gained, stop_lost,

stop_retained_variant

mmp_admin@misc_grch37:miscellaneous

Reported Variants

Showing 1-1 of 1 variants

Create Report

Front-end: MMP-SAS

CBA

MMP-SAS v1.1.9

Start

test

Studies

About

Logout

Projects / mmp_admin@misc_grch37 / miscellaneous

Clinical Analysis

Upload

Analysis

Prioritization

Report

Personal Information

Service petitioner:

Unidad de Oncología

Department:

City:

Patient:

NUHSA:

Family Number:

Analysis type:

Entry date:

Birth Year:

Relationship:

Physician one:

Physician two:

Analysis

Interpretation

Other Information

Save report

Export report



Solicitante:

Unidad de Oncología

Departamento / Unidad:

Sección de Genética y Genómica

Centro:

HUVR

Ciudad:

UGC Medicina Maternofetal, Genética y Reproducción

Paciente:

HU virgen del Rocío

NUHSA:

Avda. Manuel Siurot s/n, 41013 Sevilla

Nº ADN:

Nº de familia:

Tipo de Análisis:

Tipo de Muestra:

Sangre

Fecha de entrada:

Año de Nacimiento:

Sexo:

FEMENINO

Parentesco:

Estatus:

AFECTO

Fecha de informe:

5/7/2018, 4:31:14 PM

Motivo de la consulta

Cáncer gástrico

Informe de Análisis Genético Molecular de Enfermedad de Cáncer gástrico

Información de Análisis

ID Análisis:

3182

Descripción:

analysis

Información de Priorización

Resultado

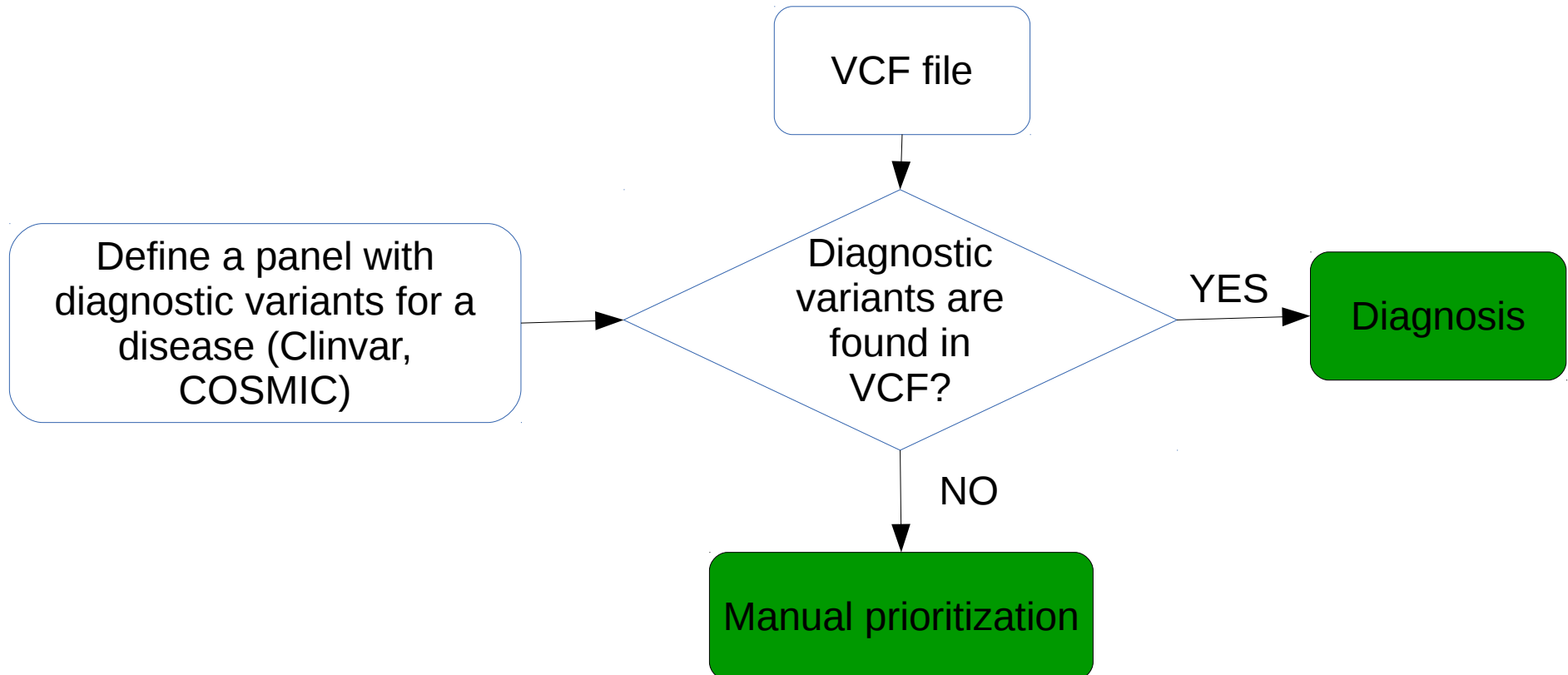
Variante	ID SNP	Genes	Tipo	Tipo de consecuencia
16:68853296:C:G	16:68853296:C:G	CDH1,RP11-354M1.2,FTLP14,	SNV	missense_variant

Resumen

presenta la mutación 16:68853296:C:G del gen CDH1,RP11-354M1.2,FTLP14, en Heterocigosis

IVA dev version (available soon)

- Gene panel component
 - Design and management of panels of genes
 - Goal: Diagnostic purposes



IVA dev version (available soon)

- Gene panel component
 - Essential for defining (informed consent is necessary):
 - **Pharmacogenetic variants:** variability in genes implicated in drug response can modulate treatment efficacy or predispose to adverse drug reactions
 - **Secondary findings:** reporting of known pathogenic or expected pathogenic variants in the 56 ACMG genes even when unrelated to the primary medical reason for testing

© American College of Medical Genetics and Genomics

ACMG STATEMENT | **Genetics
inMedicine**

Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics

Sarah S. Kalia, ScM¹, Kathy Adelman²,
Christine Eng, MD⁶, James P. Evans, MD, PhD³,
Teri E. Klein, PhD¹⁰, Bruce R. Korf, MD, PhD⁴,
C. Sue Richards, PhD¹⁴, Christopher N. Vango⁵,
David T. Miller, MD, PhD¹⁸; on behalf of the /

**Genetics
inMedicine** | **LETTER TO THE EDITOR**

© American College of Medical Genetics and Genomics

ACMG secondary findings 2.0

To the Editor: The authors of the recent American College of Medical Genetics and Genomics secondary findings

leads to an inconsistency in that many EP novel predicted loss of function variants would score as “likely pathogenic” in the new pathogenicity criteria, while it is unclear whether expected pathogenic missense variants should be reported as secondary findings. The second issue is that an explicit decision should be made as to the pathogenicity threshold

IVA dev version (available soon)

- Variant classification according to ACMG Standards:
 - Pathogenic, likely pathogenic, benign, likely benign, uncertain significance

© American College of Medical Genetics and Genomics **ACMG STANDARDS AND GUIDELINES** | **Genetics
inMedicine**

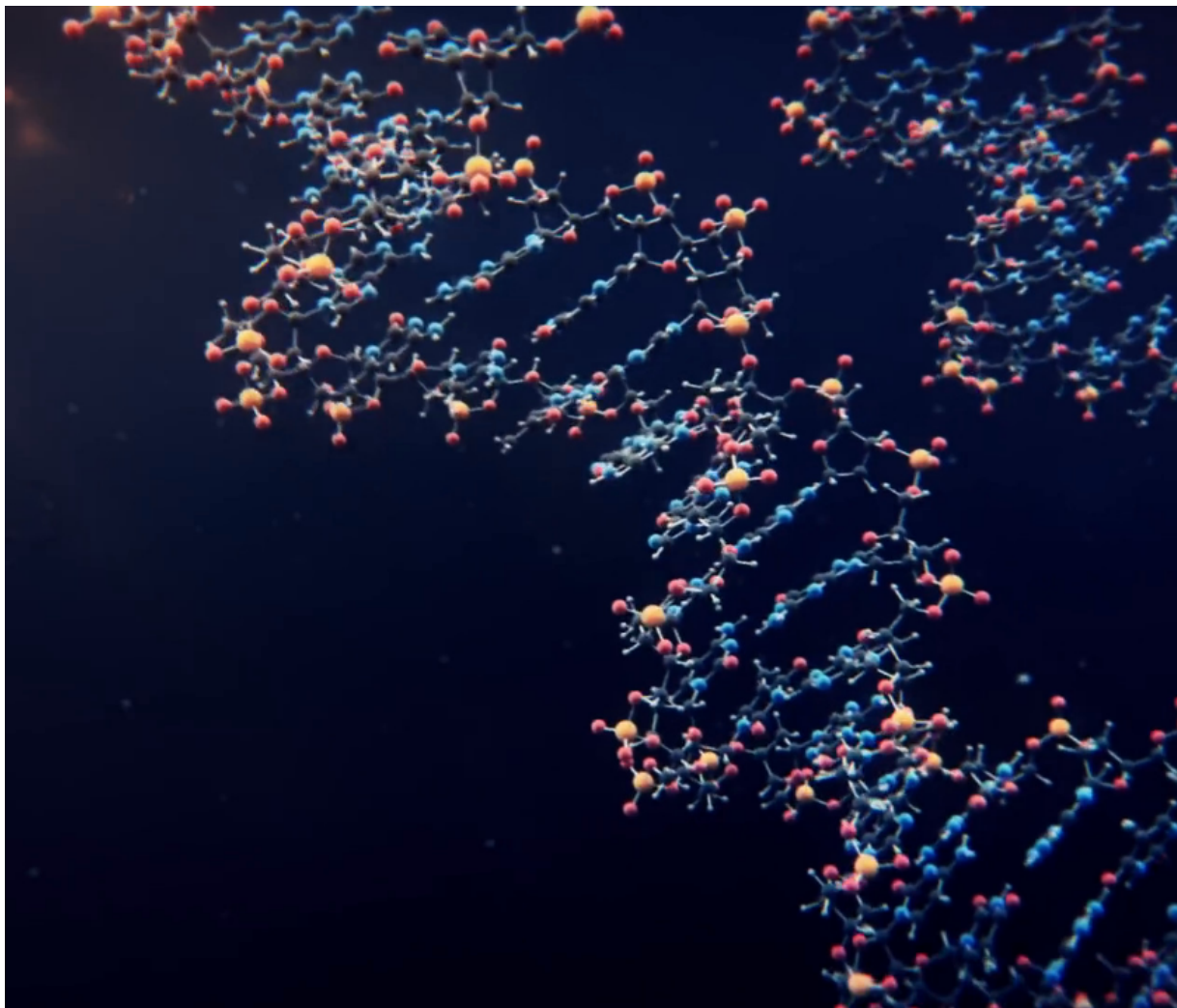
Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology

Sue Richards, PhD¹, Nazneen Aziz, PhD^{2,16}, Sherri Bale, PhD³, David Bick, MD⁴, Soma Das, PhD⁵, Julie Gastier-Foster, PhD^{6,7,8}, Wayne W. Grody, MD, PhD^{9,10,11}, Madhuri Hegde, PhD¹², Elaine Lyon, PhD¹³, Elaine Spector, PhD¹⁴, Karl Voelkerding, MD¹³ and Heidi L. Rehm, PhD¹⁵; on behalf of the ACMG Laboratory Quality Assurance Committee

IVA dev version (available soon)

- Cohort information
 - Existence of a given variant in a cohort
- Federation of applications
 - A single installation can query other installations for variants, additional information, stats, etc (more powerful than Beacon)
- User annotation of variants:
 - Technological artifacts
 - Novel and diagnostic variants
 -
 -
 -

IVA dev version (available soon)



Welcome to the MMP-SAS tool for whole genome variant analysis. This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees, case-controls or sporadic samples.

[SIGN IN](#)

IVA dev version (available soon)

MMPStudy

BACK

SAMPLES

ANALYSES

Search...

Created after

mm / dd / yyyy

Created after

mm / dd / yyyy

Sort by

Creation date

Sort order

Descending

CREATE ANALYSIS

ID	Individual ID	Creation date	Type
211_25209	-	5/17/2019	GERMINAL
244_10651	-	5/17/2019	GERMINAL
161_14654	-	5/17/2019	GERMINAL
198_13437	-	5/17/2019	GERMINAL

IVA dev version (available soon)

MMP

Analysis

×

Analysis: 5da43804b7bf434a9f200b18

Name

Huntington's disease #12

Description

Confirm Huntington's diagnose for patient #12.

Created on

October 24, 2019

Priority

Normal

Created by

test

Cases

Controls

DISCARD CHANGES

BACK TO ANALYSES

SAVE

CLOSE

IVA dev version (available soon)

MMP

Summary

SAMPLES

GENOMIC

CONSEQUENCES

FREQUENCY

DELETTERIOUSNESS

CONSERVATION

CLINICAL

VCF METRICS

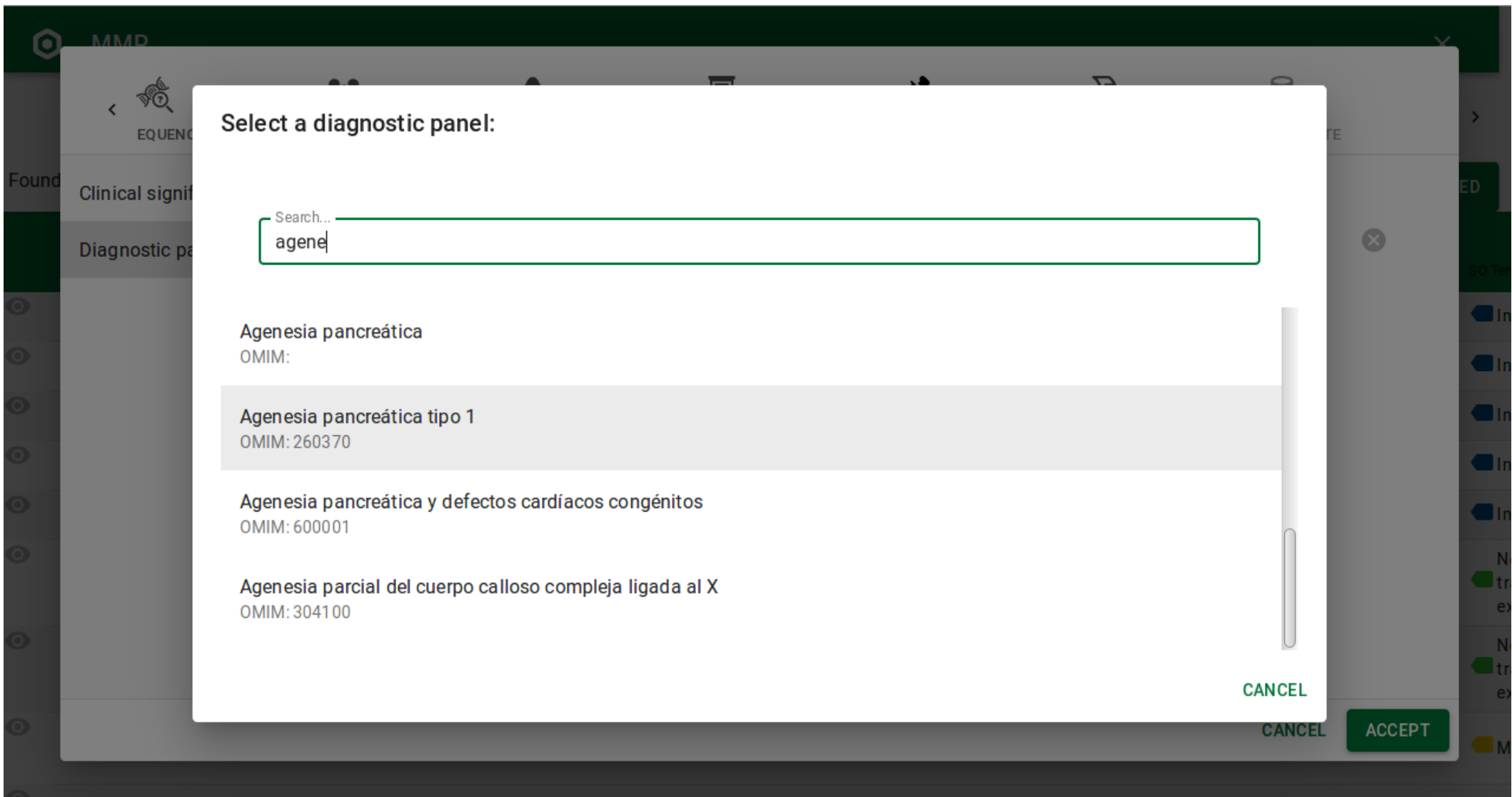
READ & WRITE

Use RefSeq

0 SELECTED

Samples												Conservation					Population frequencies				
ID	RefSeq	Type	ClinVar	Gene	Transcript	HGVS	SO Term	Exon	Sift	Polyphen	CADD	GERP	PhyloP	PhastCons	Appearances	1000 genomes	gnomAD genomes	gnomAD exomes	ESP6500	MGP	
1.977330.T:C	rs2799066	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.1178-6T>C	Splice region		0.10	-2.770	-0.854	0.003	2/2						
1.981931.A:G	rs2465128	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.3066A>G	Synonymous	18/36	0.00	-7.780	-1.982	0.000	2/2						
1.982994.T:C	rs10267	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.3558T>C	Synonymous	21/36	0.04	-0.320	-0.460	0.401	2/2						
1.984302.T:C	rs9442391	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.4161T>C	Synonymous	24/36	0.00	-7.400	-3.355	0.000	2/2						
1.986732.G:A	rs144245019	-/-	SNV	uncertain significance	AGRN	ENSG00000188157	ENST00000379370	c.5353G>A	Missense	31/36	0.51,tolerated	0.00,benign	2.21	-2.070	-1.333	0.001	1/2				
1.987200.C:T	rs9803031	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.5651+5C>T	Splice region		5.35	-6.110	-1.225	0.021	2/2	EAS: unknown					
1.990280.C:T	rs4275402	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.6057C>T	Synonymous	36/36	12.10	-6.220	-0.258	0.129	2/2						
1.1163073.C:A	rs1163073	-/-	SNV	benign	AGRN	ENSG00000188157	ENST00000379370	c.218C>A	Missense	1/1	0.18,tolerated	0.11,benign	10.28	-2.270	-0.280	0.711	1/2				

IVA dev version (available soon)



IVA dev version (available soon)

The screenshot shows the IVA dev version interface with the FREQUENCY tab selected. The interface includes a top navigation bar with icons for SEQUENCES, FREQUENCY, DELETTERIOUSNESS, CONSERVATION, CLINICAL, VCF METRICS, and READ & WRITE. A left sidebar lists data sources: All sources, 1000 genomes, gnomAD genomes, gnomAD exomes, ESP6500, and MGP. The main table displays frequency settings for each source, with columns for the source name, a frequency type (min), and a maximum value (0.1). The table is currently empty, showing only the header row for 'All sources'.

Source	Frequency	Max Value
All sources	min	0.1
1000 genomes	min	0.1
gnomAD genomes	min	0.1
gnomAD exomes	min	0.1
ESP6500	min	0.1
MGP	min	0.1

CANCEL **ACCEPT**

IVA dev version (available soon)

The screenshot shows a web application interface for user annotation. At the top, there is a dark green header with the 'MMP' logo and a close button. Below the header, a grey bar displays the variant ID 'Variant: 14:88417096:A:-'. The main content area is a white box titled 'User annotation'. It contains three dropdown menus: 'Visibility' set to 'Public', 'Scope' set to 'Study', and 'Category' set to 'Artifact'. Below these is a large text area for a comment, which contains the text: 'Esta variante ha sido encontrada en muchas de las muestras secuenciadas con la misma tecnología, por lo que sospechamos que es un artefacto. Revisar.' The text is underlined in red. At the bottom right of the white box are 'CANCEL' and 'OK' buttons. At the bottom right of the grey bar are 'SELECT' and 'CLOSE' buttons.

Variant: 14:88417096:A:-

User annotation

Visibility: Public

Scope: Study

Category: Artifact

Comment:

Esta variante ha sido encontrada en muchas de las muestras secuenciadas con la misma tecnología, por lo que sospechamos que es un artefacto. Revisar.

CANCEL OK

SELECT CLOSE

IVA dev version (available soon)

 MMP

Variant: 14:88417096:A:-

SUMMARYSAMPLESUSER ANNOTATIONSTRANSSCRIPTS

+

User	Visibility	Scope	Comment	
mmp_admin	PUBLIC	Study	ARTIFACT	Esta variante ha sido encontrada en muchas de las muestras secuenciadas con la misma tecnología, por lo que sospechamos que es un artefacto. Revisar.
mmp_admin	PRIVATE	Study	BENIGN	La variante ha sido descrita como benigna y común en la literatura.

SELECTCLOSE

Acknowledgements

- **Ignacio Medina** (*Genomics England, University of Cambridge, OpenCB Team leader*)
- Jacobo Coll (*Genomics England*)
- Javier Lopez (*Genomics England*)
- Pedro Furio (*Genomics England*)
- Joaquin Tarraga (*University of Cambridge*)
- Antonio Rueda (*Genomics England*)

Clinical Bioinformatics Area Fundación Progreso y Salud, Sevilla, Spain, and...

...the INB-ELIXIR-ES, National Institute of Bioinformatics
and the BiER (CIBERER Network of Centers for Research in Rare Diseases)



Follow us on
twitter

@xdopazo

@ClinicalBioinfo



"Una manera de hacer Europa"

<https://www.slideshare.net/xdopazo/>

