Data collection and Management in PM Research

Alfonso Valencia
ICREA Research Professor
Director, Life Sciences Department BSC
Director, Spanish National Bioinformatics Institute (INB-ISCIII)
Head of the Spanish Node of ELIXIR (ELIXIR-ES)

valencia@bsc.es
@alfons_valencia

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

INB
Instituto Nacional de Bioinformática

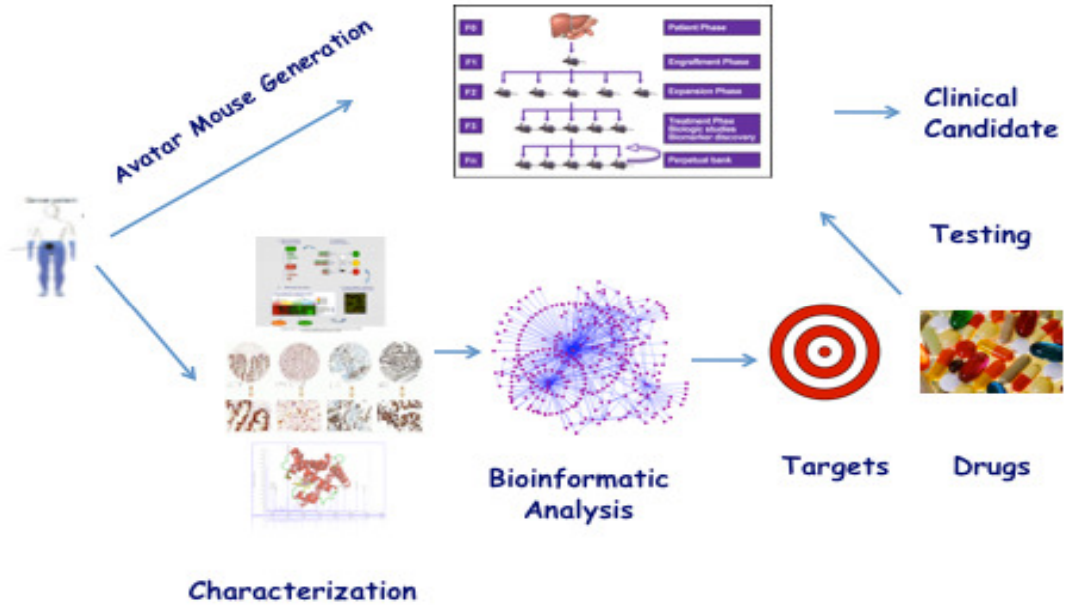
elixir SPAIN

"INNOVATIVE CONCEPTS ON DATA GENERATION AND USE FOR PERSONALISED MEDICINE RESEARCH"
ICPerMed- First RESEARCH WORKSHOP
Milan, Palazzo Lombardia, 26-27th June 2017

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ICPerMed
INTERNATIONAL CONSORTIUM

Spanish National Cancer Research Center
CNIO co-clinical cancer initiative



30% success ..

N	Tumor	#Muts	Relevant somatic muts	#CNV	Relevant CNV	Targets	Xenograft	Personalized treatment
1	High grade pancreatic neuroendocrine tumor	62	ARID1A, MYO1B, JAKMIP2, JARID2 PIK3C2A, PIK3CA, SSTR2, DDR2, TP53	6	GNG11	PIK3CA DDR2	Yes	Dasatinib
2	Glioblastoma	63	EPHA3, NF1, PTPN11 , FAS , CDKN2A	23	0	NF1	Yes	Erlotinib + everolimus
3	Intestinal neuroendocrine tumor	5	CREB3L3, ITPR2, MYO5B	0	NA	CREB3L3	Yes	Metformine
4	Pancreatic adenocarcinoma	38	KRAS, UBA1, FAM83H, SMAD4, SLC15A2, PIWIL3, SLC3A2, SLC22A17, TP53.	10	0	KRAS	No	No
5	Uveal melanoma	5	GNA11, TAOK3	0	-	GNA11	Yes	PKC inhibitor. PI3k/mtor inh
6	Colon cancer	71	APC, DICER 1, TP53, CHEK1, SOS1	36	0	CHEK1	Yes	Not used yet
7	Melanoma	952	BRCA1, EZH2, FGFR2, FN1, IGF1R, KDR, KRAS, MET, MPL, PRKCB, PIK3C2G, PTK2B.	0	-	FGFR2,IGF1R, PIK3C2G, MET, BRCA1	No	Not used yet
8	Melanoma	29	BAI3, DNAH5, MDN1, NRAS	2	SKT19	NRAS	No	MEK inh
9	Glioblastoma	64	MLLT10, PBRM1	26	EGFR, CDKN2A, ERRFI1	EGFR	No	Erlotinib
10	Pancreatic adenocarcinoma	21	KRAS, XPC, P53	0	-	KRAS, XPC	No	Mitomycin C
11								Not used yet
12								-
13	Pancreatic adenocarcinoma	No results	-	No results	-	-	No	-

Garralda et al., *Clin. Cancer Res.* ***2014***

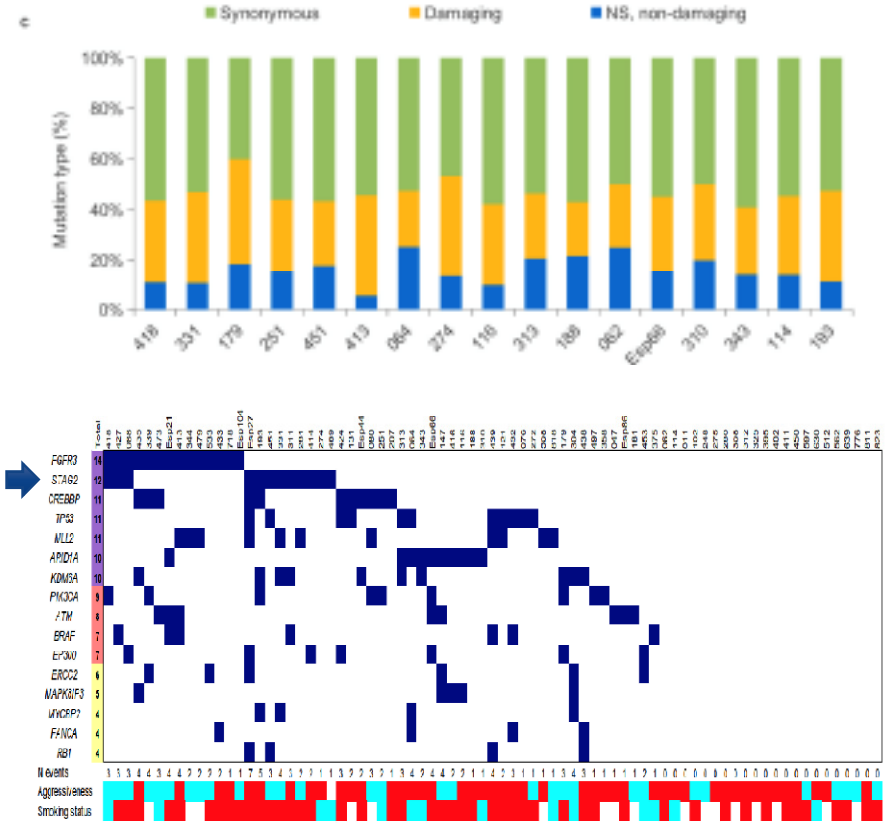
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Bladder Cancer (with F. Real CNIO)

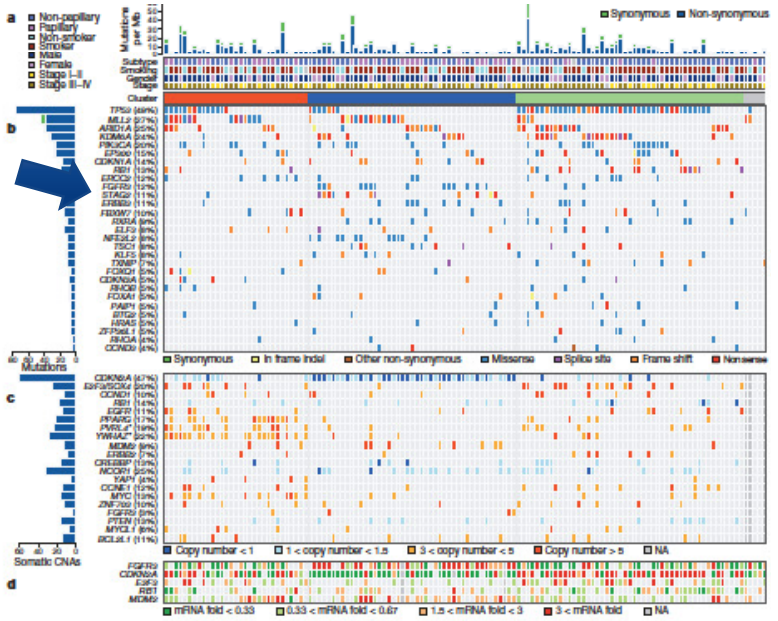
Recurrent inactivation of *STAG2* in bladder cancer is not associated with aneuploidy

Cristina Balbás-Martínez[1], Ana Sagrera[1,19], Enrique Carrillo-de-Santa-Pau[1,19], Julie Earl[1,2], Mirari Márquez[3], Miguel Vazquez[4], Eleonora Lapi[1], Francesc Castro-Giner[5], Sergi Beltran[5], Mònica Bayés[5], Alfredo Carrato[2], Juan C Cigudosa[6], Orlando Domínguez[7], Marta Gut[5], Jesús Herranz[3], Núria Juanpere[8], Manolis Kogevinas[9–12], Xavier Langa[1], Elena López-Knowles[10], José A Lorente[13], Josep Lloreta[8,14], David G Pisano[15], Laia Richart[1], Daniel Rico[4], Rocío N Salgado[6], Adonina Tardón[16], Stephen Chanock[17], Simon Heath[5], Alfonso Valencia[4], Ana Losada[18], Ivo Gut[5], Núria Malats[3] & Francisco X Real[1,14]

VOLUME 45 | NUMBER 12 | DECEMBER 2013 NATURE GENETICS



Enrique Carrillo



Comprehensive molecular characterization of urothelial bladder carcinoma

The Cancer Genome Atlas Research Network*

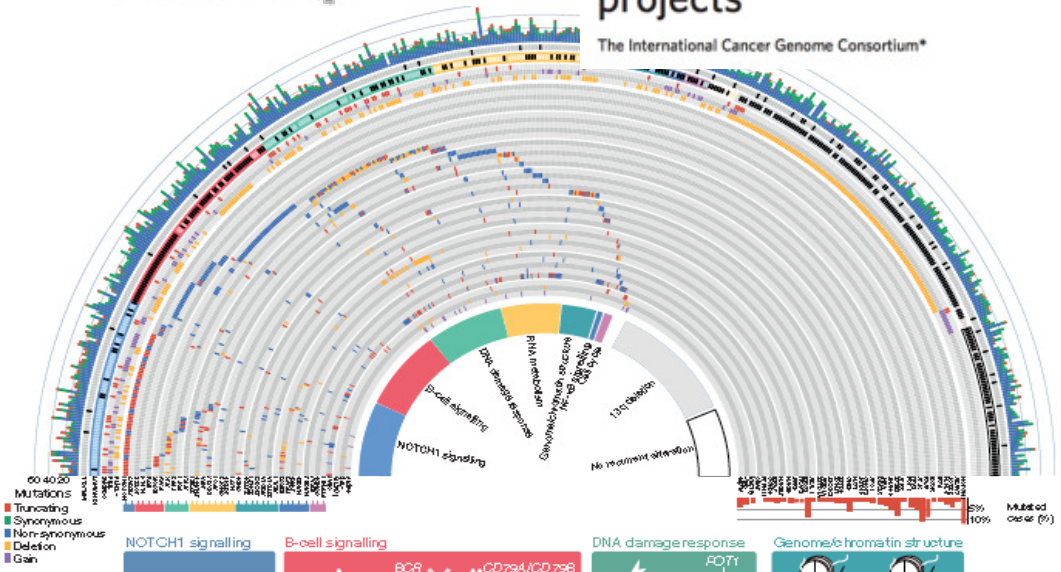
Vol 464|15 April 2010|doi:10.1038/nature08987 nature

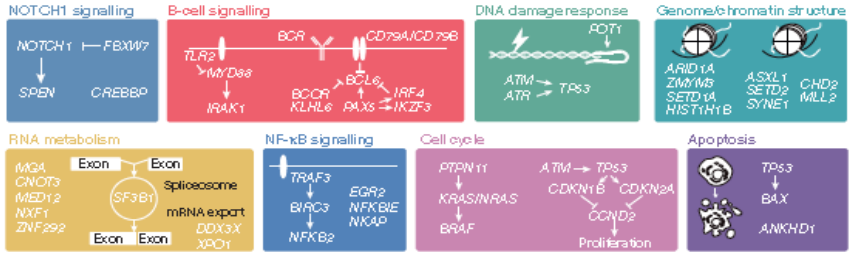
International Cancer Genome Consortium

PERSPECTIVES

International network of cancer genome projects

The International Cancer Genome Consortium*





Non-coding recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente[1], Silvia Beà[2], Rafael Valdés Mas[1], Neus Villamor[3], Jesús Gutiérrez Abril[1], José I. Martín Subero[4], Marta Munar[5], Carlota Rubio Pérez[6], Pedro Jares[7], Marta Aymerich[3], Tycho Baumann[8], Renée Beekman[2], Laura Belver[9], Anna Carrio[3], Giancarlo Castellano[2], Guillem Clot[2], Enrique Colado[10], Dolors Colomer[3], Dolors Costa[3], Julio Delgado[8], Anna Enjuanes[7], Xavier Estivill[11], Adolfo A. Ferrando[9], Josep L. Gelpi[6], Blanca González[3], Santiago González[5], Marcos González[12], Marta Gut[13], Jesús M. Hernández Rivas[12], Mónica López Guerra[3], David Martín García[2], Alba Navarro[2], Pilar Nicolás[14], Modesto Orozco[5], Ángel R. Payer[10], Magda Pinyol[7], David G. Pisano[15], Diana A. Puente[1], Ana C. Queirós[4], Víctor Quesada[1], Carlos M. Romeo Casabona[14], Cristina Royo[2], Romina Royo[5], María Rozman[3], Nuria Russiñol[2], Itziar Salaverría[2], Kostas Stamatopoulos[16], Hendrik G. Stunnenberg[17], David Tamborero[6], María J. Terol[18], Alfonso Valencia[15], Nuria López Bigas[6], David Torrents[5], Ivo Gut[13], Armando López Guillermo[8], Carlos López Otín[1]§ & Elías Campo[3]§

nature International weekly journal of science

nature International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For
Archive > Volume 475 > Issue 7354 > Letters > Article
NATURE | LETTER OPEN
« previous article next article »

Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia

Xose S. Puente, Magda Pinyol, Víctor Quesada, Laura Conde, Gonzalo R. Ordóñez, Neus Villamor, Georgia Escaramis, Pedro Jares, Silvia Beà, Marcos González-Díaz, Laia Bassaganyas, Tycho Baumann, Manel Juan, Mónica López-Guerra, Dolors Colomer, José M. C. Tubio, Cristina López, Alba Navarro, Cristian Tornador, Marta Aymerich, María Rozman, Jesús M. Hernández, Diana A. Puente, José M. P. Freije, Gloria Velasco, Ana Gutiérrez-Fernández, Dolors Costa, Anna Carrió, Sara Guijarro, Anna Enjuanes, Lluís Hernández, Jordi Yagüe, Pilar Nicolás, Carlos M. Romeo-Casabona, Heinz Himmelbauer, Ester Castillo, Juliane C. Dohm, Silvia de Sanjosé, Miguel A. Piris, Enrique de Alava, Jesús San Miguel, Romina Royo, Josep L. Gelpi, David Torrents, Modesto Orozco, David G. Pisano, Alfonso Valencia, Roderic Guigó, Mónica Bayés, Simon Heath, Marta Gut, Peter Klatt, John Marshall, Keiran Raine, Lucy A. Stebbings, P. Andrew Futreal, Michael R. Stratton, Peter J. Campbell, Ivo Gut, Armando López-Guillermo, Xavier Estivill, Emili Montserrat, Carlos López-Otín & Elías Campo – Show fewer authors
Affiliations | Contributions | Corresponding authors

Nature **475**, 101–105 (07 July 2011) | doi:10.1038/nature10113
Received 16 November 2010 | Accepted 06 April 2011 | Published online 05 June 2011

LETTERS
nature genetics
VOLUME 44 | NUMBER 11 | NOVEMBER 2012 NATURE GENETICS

Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia

Marta Kulis[1,11], Simon Heath[2,11], Marina Bibikova[3,11], Ana C Queirós[4,11], Alba Navarro[1], Guillem Clot[1], Alejandra Martínez-Trillos[5], Giancarlo Castellano[1], Isabelle Brun-Heath[2], Magda Pinyol[6], Sergio Barberán-Soler[7], Panagiotis Papasaikas[7], Pedro Jares[1], Silvia Beà[1], Daniel Rico[8], Simone Ecker[8], Miriam Rubio[8], Romina Royo[9], Vincent Ho[3], Brandy Klotzle[3], Lluís Hernández[1], Laura Conde[1], Mónica López-Guerra[1], Dolors Colomer[1], Neus Villamor[1], Marta Aymerich[1], María Rozman[1], Mónica Bayes[2], Marta Gut[2], Josep I. Gelpi[9], Modesto Orozco[9], Jian-Bing Fan[3], Víctor Quesada[10], Xose S Puente[10], David G Pisano[6], Alfonso Valencia[8], Armando López-Guillermo[5], Ivo Gut[2], Carlos López-Otín[10], Elías Campo[1] & José I Martín-Subero[4]

nature genetics
nature.com ▸ journal home ▸ archive ▸ issue ▸ letter ▸ full text
NATURE GENETICS | LETTER

Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia

Víctor Quesada, Laura Conde, Neus Villamor, Gonzalo R Ordóñez, Pedro Jares, Laia Bassaganyas, Andrew J Ramsay, Silvia Beà, Magda Pinyol, Alejandra Martínez-Trillos, Mónica López-Guerra, Dolors Colomer, Alba Navarro, Tycho Baumann, Marta Aymerich, María Rozman, Julio Delgado, Eva Giné, Jesús M Hernández, Marcos González-Díaz, Diana A Puente, Gloria Velasco, José M P Freije, José M C Tubío, Romina Royo, Josep L Gelpí, Modesto Orozco, David G Pisano, Jorge Zamora, Miguel Vázquez, Alfonso Valencia, Heinz Himmelbauer, Mónica Bayés, Simon Heath, Marta Gut, Ivo Gut, Xavier Estivill, Armando López-Guillermo, Xose S Puente, Elías Campo & Carlos López-Otín
– Show fewer authors
Affiliations | Contributions | Corresponding authors

Nature Genetics **44**, 47–52 (2012) | doi:10.1038/ng.1032
Received 12 July 2011 | Accepted 10 November 2011 | Published online 11 December 2011

CSH PRESS GENOME RESEARCH

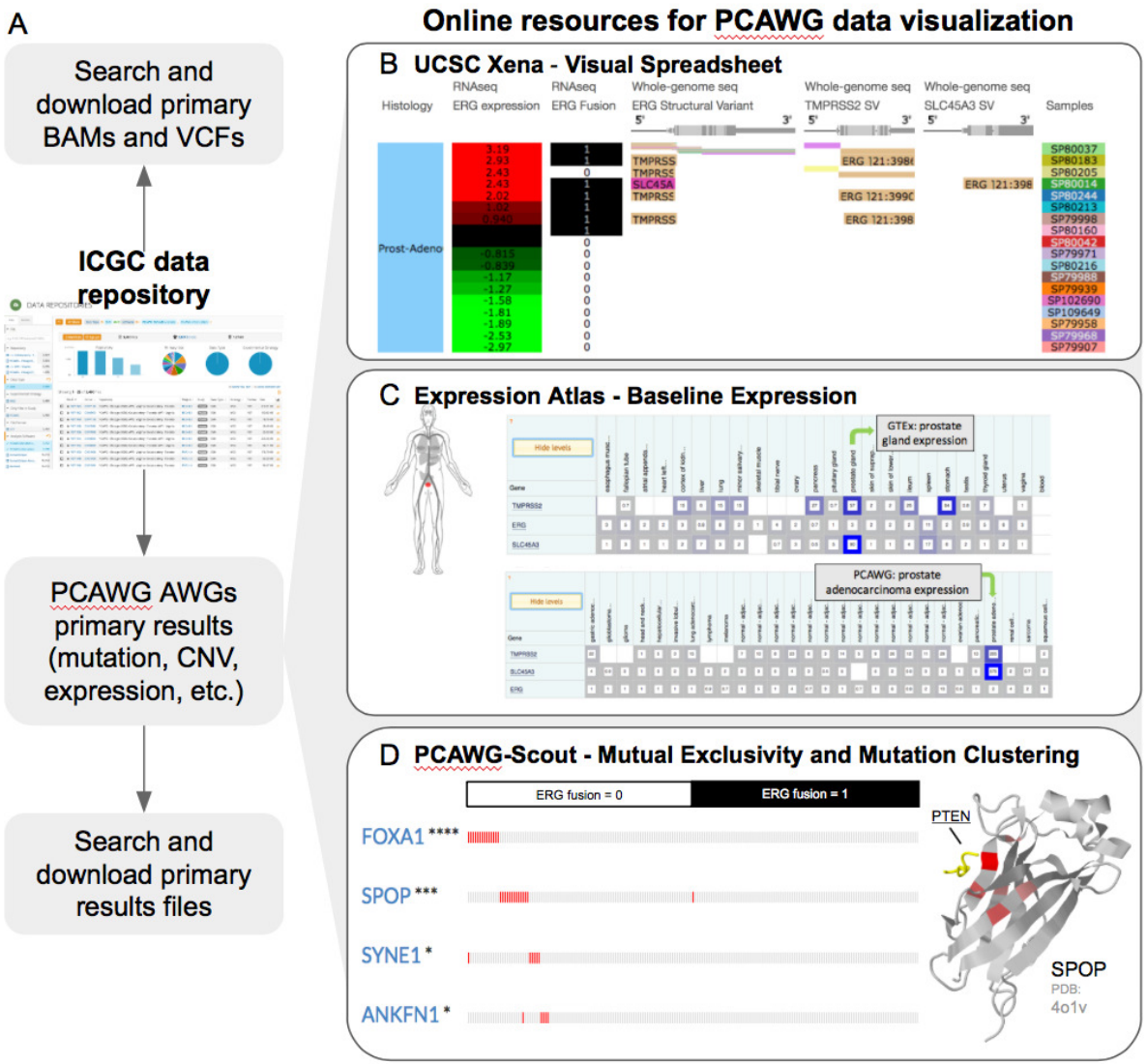
Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia

Pedro G. Ferreira, Pedro Jares, Daniel Rico, et al.

Genome Res. published online November 21, 2013
Access the most recent version at doi:10.1101/gr.152132.112

CLL (~700 TB): 500 patients. 150 WGS PanCancer PCAWG

2,834 donors – 48 tumor types (WGS)



ICGC

PCAWG PanCancer Analysis OF WHOLE GENOMES

Rbbt

Ruby bioinformatics toolbox

Miguel Vazquez

OPTIMISTIC VIEW

ASCO 2016: Precision Medicine Yields Better Outcomes for Patients in Phase I Clinical Trials

By The ASCO Post
Posted: 5/18/2016 5:10:00 PM

Key Points

- In treatment arms employing precision medicine, the tumor shrinkage rate was 30.6% compared to 4.9% in those that did not.
- Patients in precision medicine arms also had a longer progression-free survival compared to other arms (median = 5.7 vs 2.95 months).
- Results were similar in a subanalysis that included 57 trials of targeted therapies.

The Economist

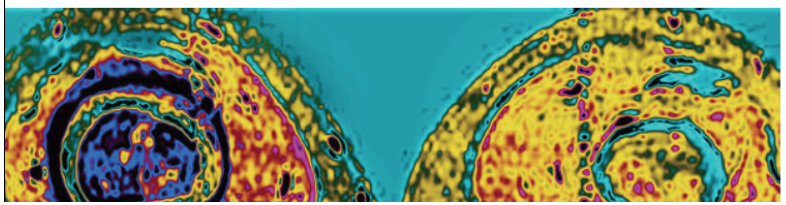
Cancer treatment

The personalisation of cancer treatments is leading to better outcomes for patients

All latest updates

It will also pave the way to cures

Jun 4th 2016 | CHICAGO | Science and technology | Timekeeper | Tweet



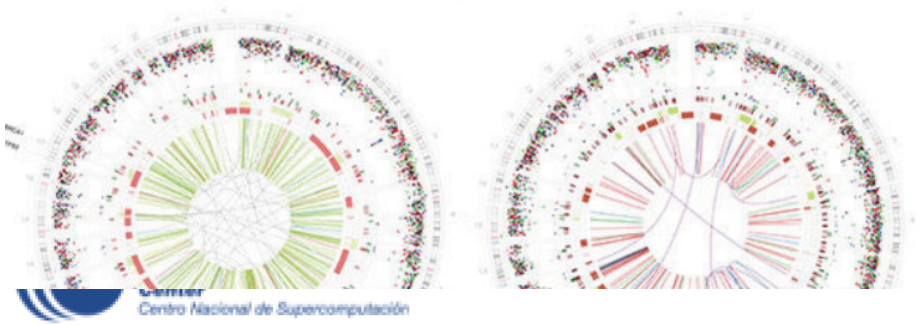
Genes and disease

Encore une fois

The Economist

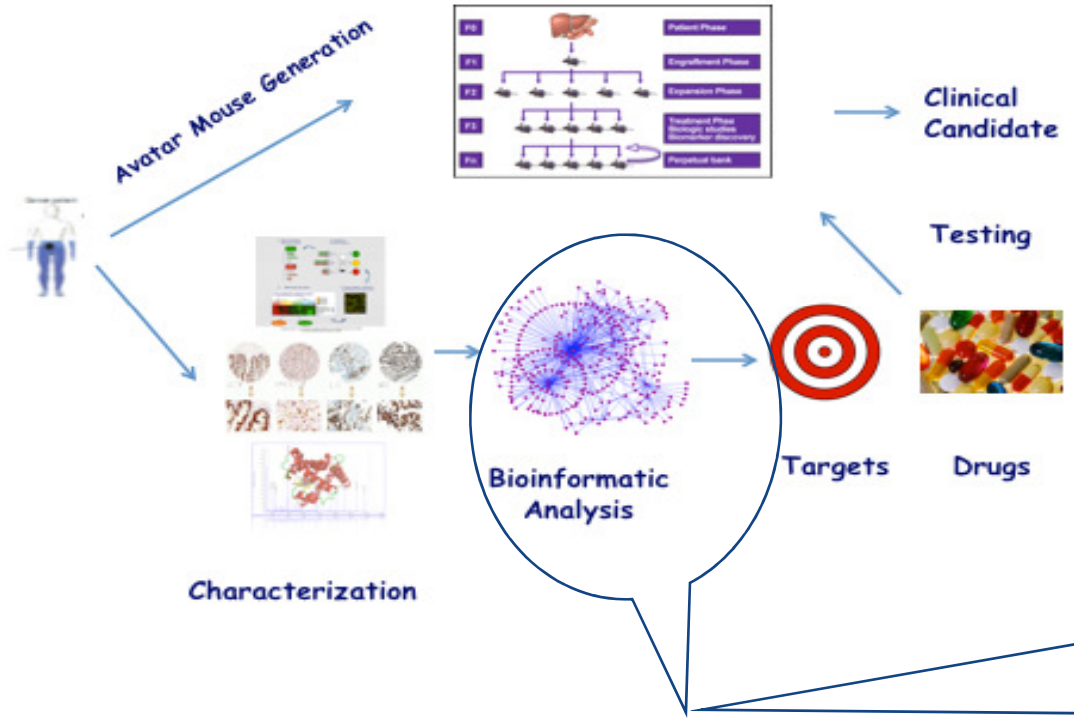
The genomic era arrives. And this time it's probably real.

May 7th 2016 | From the print edition | Timekeeper | Tweet



Centro Nacional de Supercomputación

Spanish National Cancer Research Center
CNIO co-clinical cancer initiative



N	Tumor	#Muts	Relevant somatic muts	#CNV	Relevant CNV	Targets	Xenograft	Personalized treatment
1	High grade pancreatic neuroendocrine tumor	62	ARID1A, MYO1B, JAKMIP2, JARID2 PIK3C2A, PIK3CA, SSTR2, DDR2, TP53	6	GNG11	PIK3CA DDR2	Yes	Dasatinib
2	Glioblastoma	63	EPHA3, NF1, PTPN11 , FAS , CDKN2A	23	0	NF1	Yes	Erlotinib + everolimus
3	Intestinal neuroendocrine tumor	5	CREB3L3, ITPR2, MYO5B	0	NA	CREB3L3	Yes	Metformine
4	Pancreatic adenocarcinoma	38	KRAS, UBA1, FAM83H, SMAD4, SLC15A2, PIWIL3, SLC3A2, SLC22A17, TP53.	10	0	KRAS	No	No
5	Uveal melanoma	5	GNA11, TAOK3	0	-	GNA11	Yes	PKC inhibitor. PI3k/mtor inh
6	Colon cancer	71	APC, DICER 1, TP53, CHEK1, SOS1	36	0	CHEK1	Yes	Not used yet
7	Melanoma	952	BRCA1, EZH2, FGFR2, FN1, IGF1R, KDR, KRAS, MET, MPL, PRKCB, PIK3C2G, PTK2B.	0	-	FGFR2,IGF1R, PIK3C2G, MET, BRCA1	No	Not used yet
8	Melanoma	29	BAI3, DNAH5, MDN1, NRAS	2	SKT19	NRAS	No	MEK inh
9	Glioblastoma	64	MLLT10, PBRM1	26	EGFR, CDKN2A, ERRFI1	EGFR	No	Erlotinib
10	Pancreatic adenocarcinoma	21	KRAS, XPC, P53	0	-	KRAS, XPC	No	Mitomycin C
11	Pancreatic adenocarcinoma	21	KRAS, XPC, P53	0	-	KRAS, XPC	No	Not used yet
12	Pancreatic adenocarcinoma	21	KRAS, XPC, P53	0	-	KRAS, XPC	No	-
13	Pancreatic adenocarcinoma	No results	-	No results	-	-	No	-

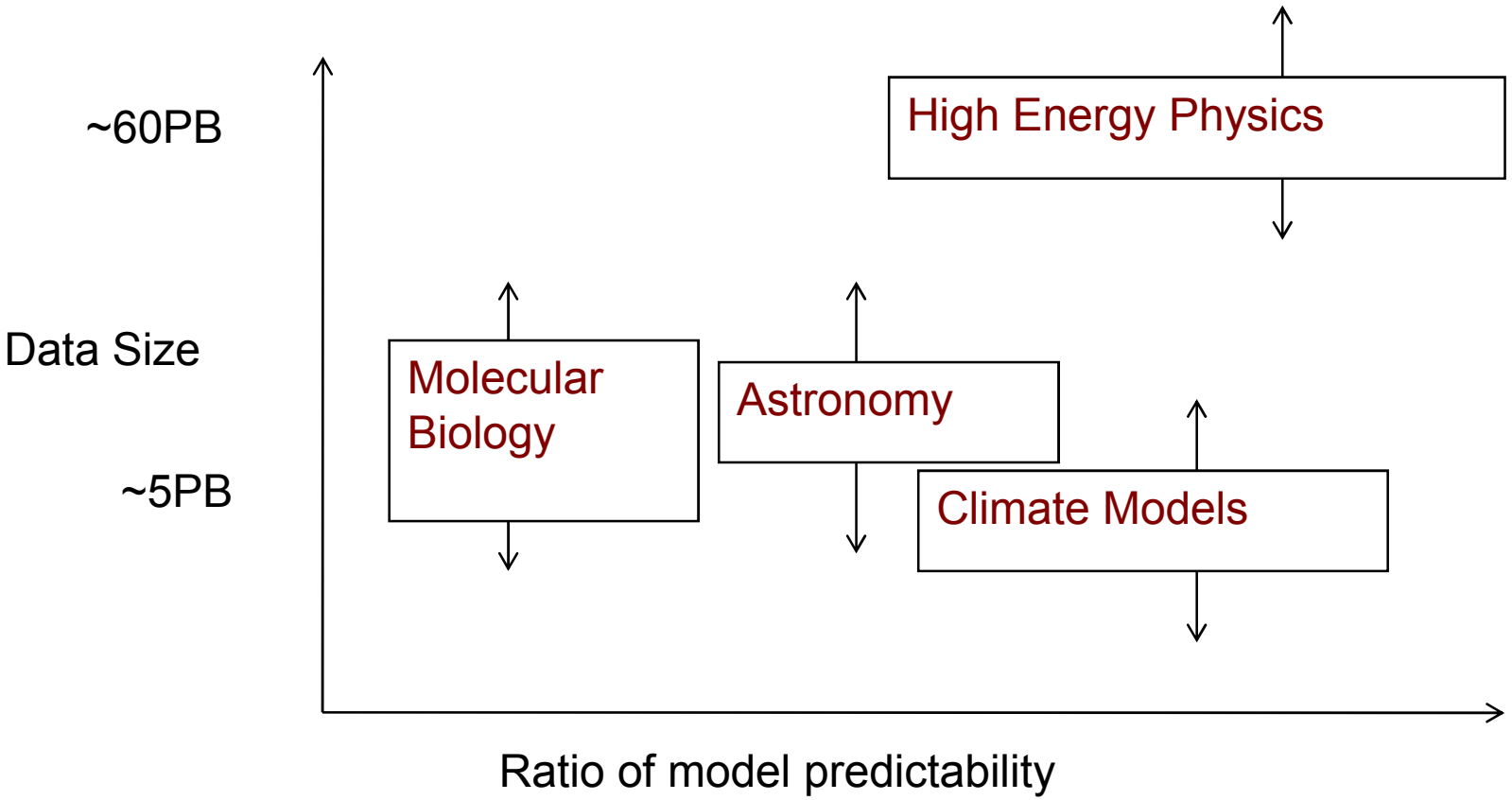
30% success ..

Bioinformatiics is the limiting factor

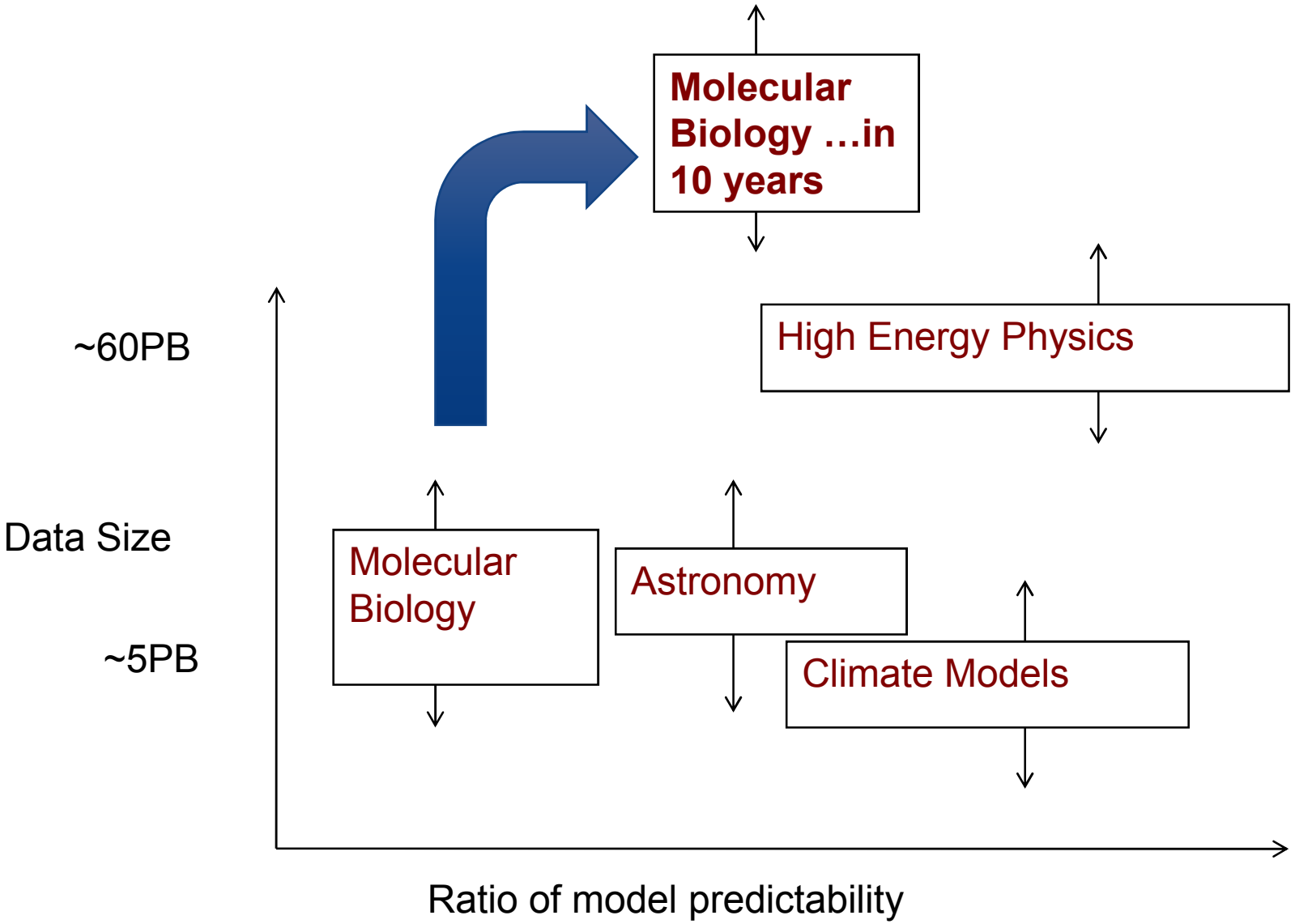
Garralda et al., *Clin. Cancer Res.* ***2014***

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Modern Biology is Big Data

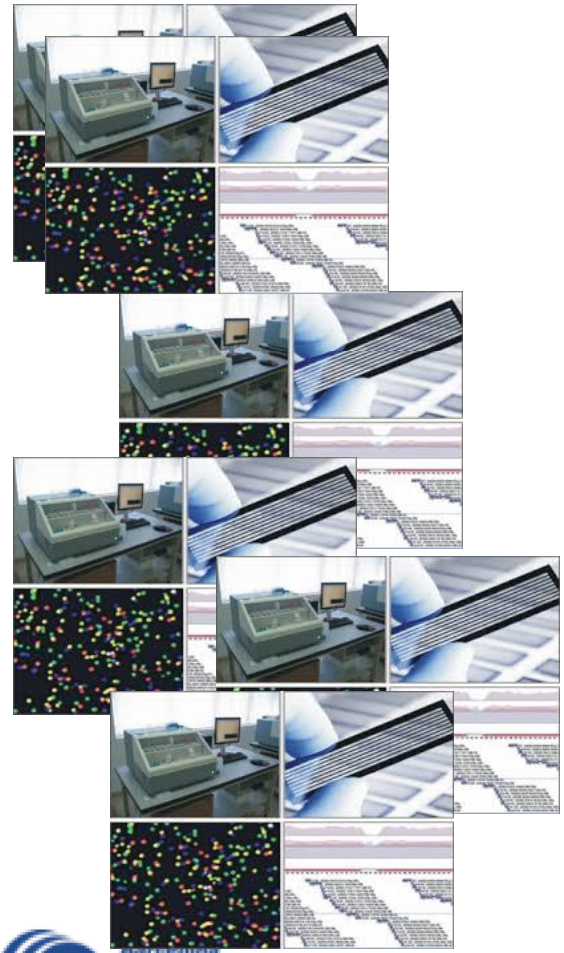


From Ewan Birney EBI



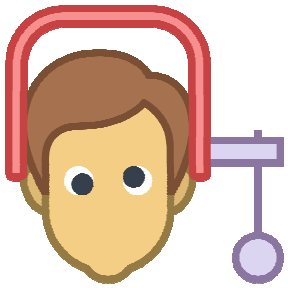
From Ewan Birney EBI

Heterogeneous distributed HT capacity



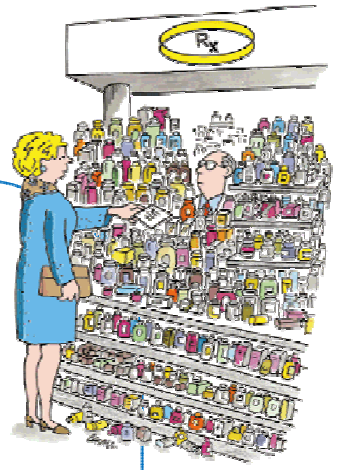
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

A few Bioinformaticians



Millions of users (biologists and biomeds)

Personalized Medicine: present & future



RNAseq
Epigenomics
HiC
Arrays
Exomes
Complete genomes
Metagenomes
Single Cells

Computational Infrastructure

Biomedical repositories of Data and Methods

Bioinformatics Infrastructure

Computational Methods (ML, Cognitive Computing)

Research Interest

Practical utility in the clinic

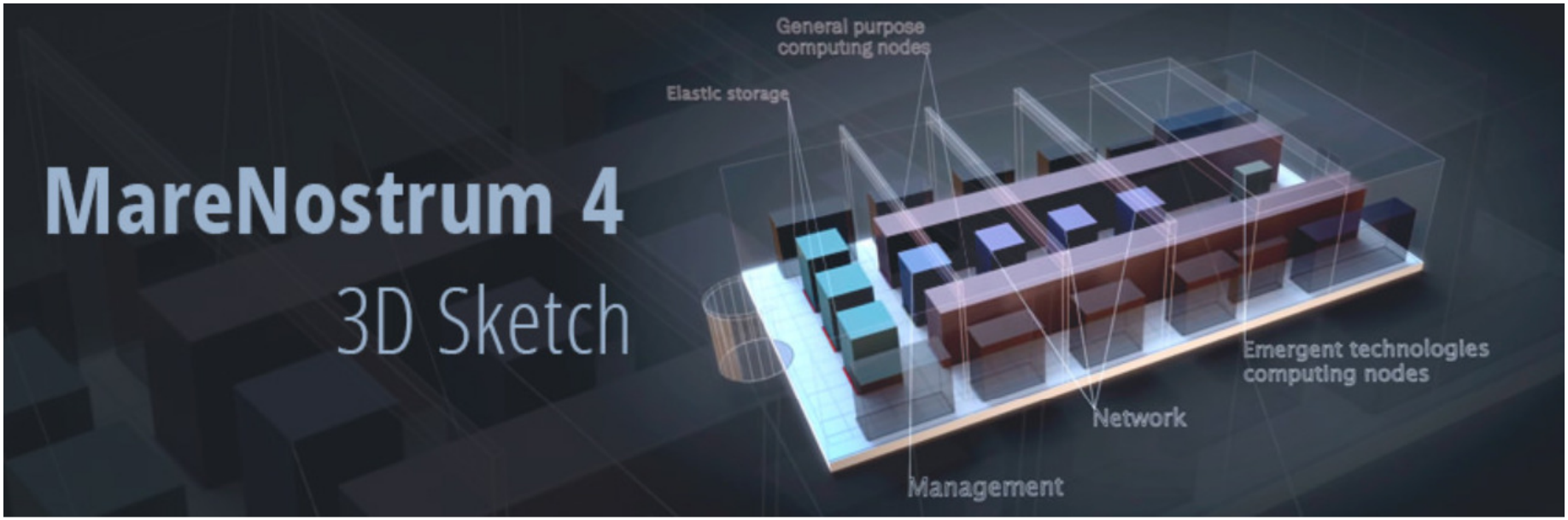
Interpretation

Mining EHR
Social Media
Medical Imaging
Other Health Data (papers, patents, etc)
Medical devices
Epidemiology

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

BSC Life Sciences Department

The MareNostrum 4 Supercomputer



Technologies from IBM, Lenovo, Intel and Fujitsu

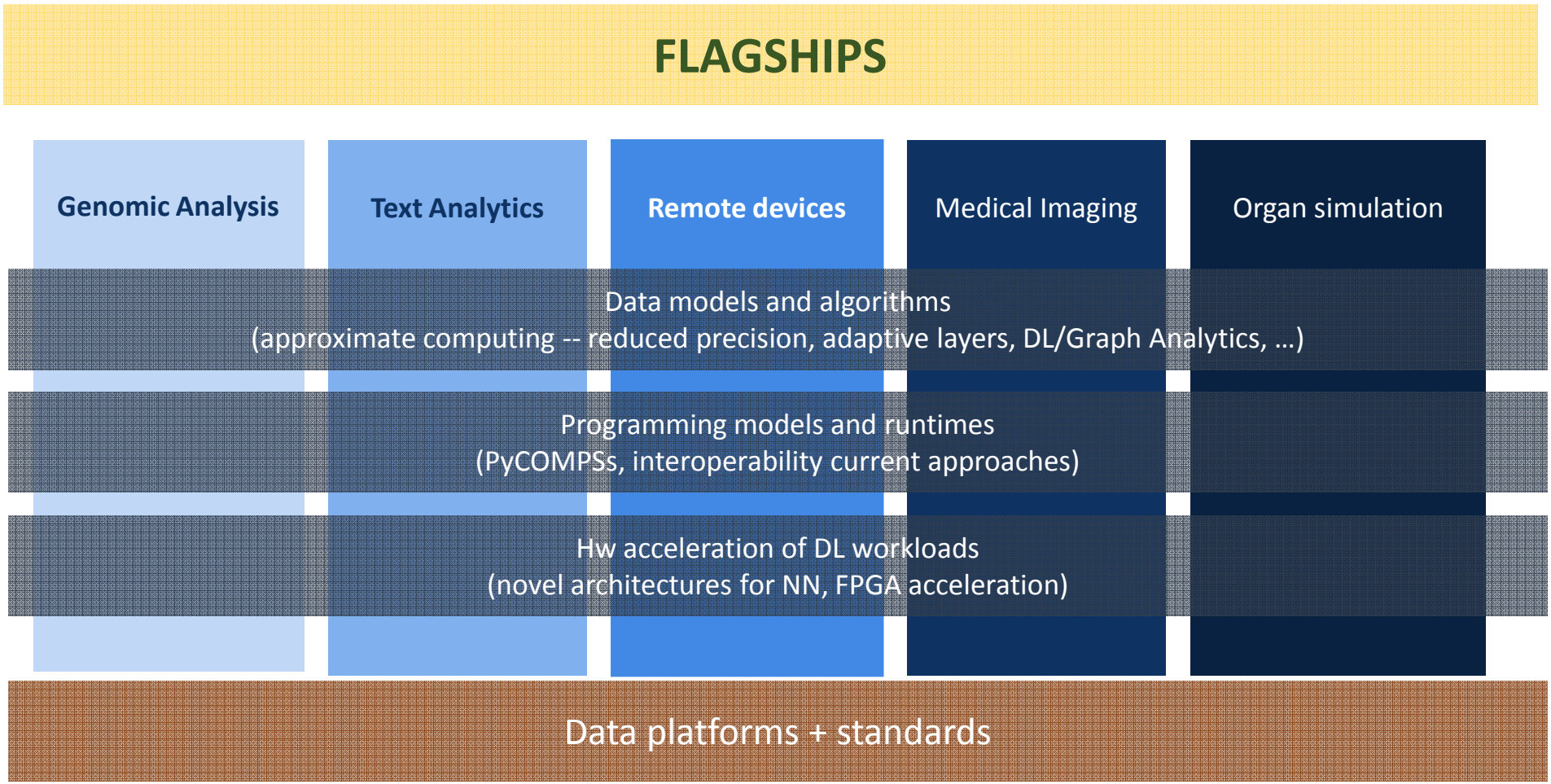
Over 13.7 Petaflops per second

→ 12 times more powerful than MN3

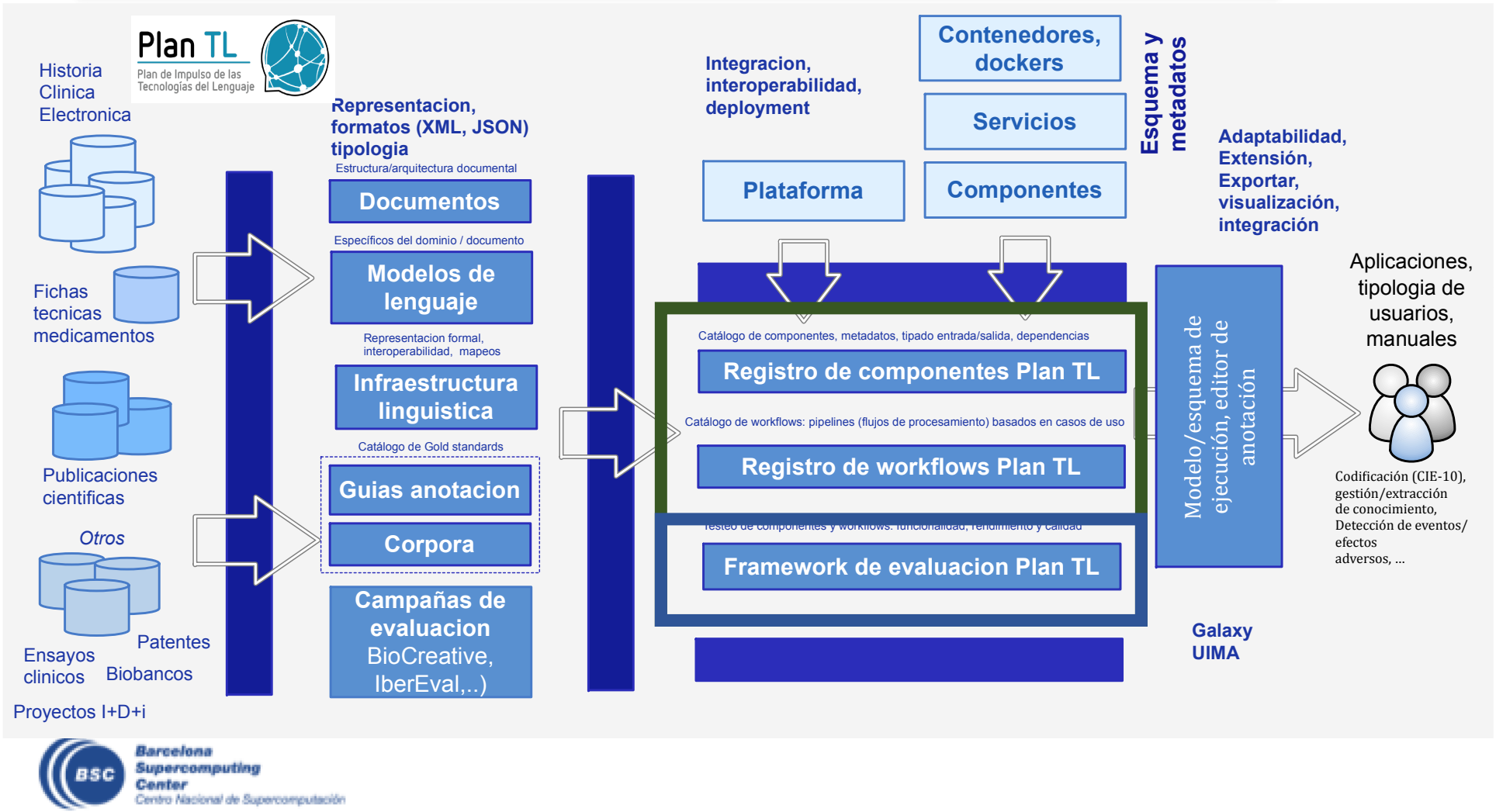
→ Over 10 Petabytes disk storage

13th in the world 3rd in Europe

BSC strategy for Personalized Medicine



ASPECTOS PLAN TL - SALUD



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ELIXIR: European infrastructure for biological information

- *Data*
- *Interoperability*
- *Tools*
- *Compute*
- *Training*


Marine metagenomics


Crop and forest plants


Human data


Rare diseases

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

www.elixir-europe.org

@ELIXIREurope

A network of National Nodes

- ELIXIR Nodes are funded nationally
- ELIXIR Nodes build on national strengths and priorities
- ELIXIR Nodes provides a national framework for long-term resource management


ELIXIR: *The Spanish Institute of Bioinformatics*

About INB

Collaborating organizations

INB core areas

Recent projects

INB & ELIXIR

www.inab.org

Contact

ELIXIR-EXCELERATE

EU funded project to:

- Accelerate the implementation of ELIXIR across Nodes
- Develop and connect resources and services (from Nodes)
- Build bioinformatics capacity across Europe (Capacity Building and Training)

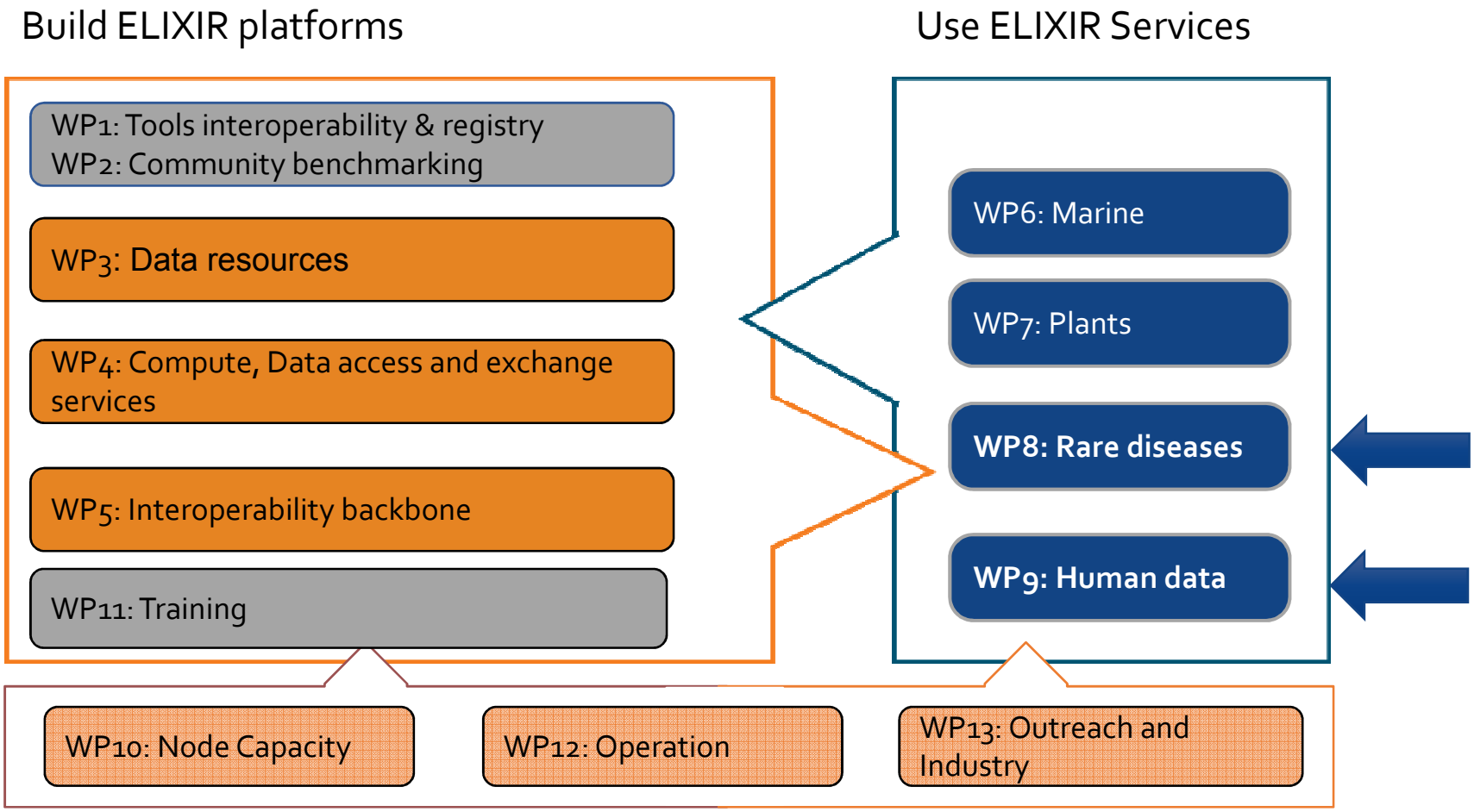
Basic facts:

- €19 million,
- 4-year project
- 41 partners across 17 countries
- Kick-off meeting: December 2015

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

excelerate

EXCELERATE Structure



Barcelona Supercomputing Center
Centro Nacional de Supercomputación

excelerate

Mission

- To construct and operate a **sustainable infrastructure for Human Genomics** and Translational data in Europe to support life science research and its translation to medicine

Vision

- To facilitate **discoverability, access, sharing and analysis** of genomics data, including rare disease, linked to other data types, at scale (**4-5M participants**)

- To **demonstrate how use of infrastructure** can impact translation of genomics research into medicine

ELIXIR Human genomics portfolio 2017

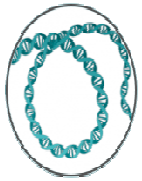
Generate

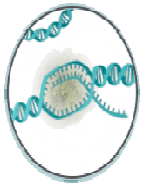
Diverse data from diverse providers

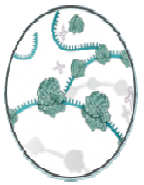
Archive, discover, manage

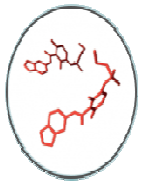
Store, access and share data of multiple types and origin

Analyse









Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EUROPEAN GENOME-PHENOME ARCHIVE

LOCAL EGA

Beacon

elixir AAI

RD-Connect

- Remote real-time visualization of human rare disease genomics data (RD-Connect) stored at the EGA
- Interpretation of phenotypic and genotypic variation for rare diseases in terms of biological pathways

excelerate

European Genome-phenome Archive

Controlled access for human genomic data

Data submitters: >600
(Universities, research institutes, consortia, pharma …)

International Cancer Genome Consortium

IRDiRC
INTERNATIONAL RARE DISEASES RESEARCH CONSORTIUM

wellcome trust sanger institute

UK 10K

Data requesters:
> 7,100 (worldwide)
Controlled access:
Data Access Committees

www.ega-archive.org

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EUROPEAN GENOME-PHENOME ARCHIVE

EMBL-EBI

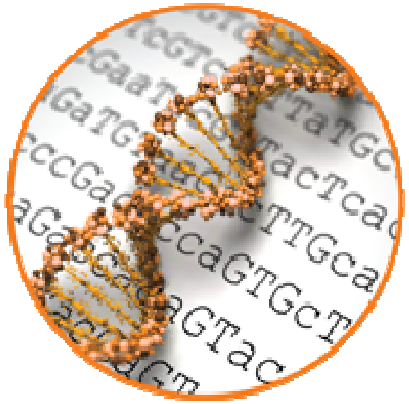
CRG Centre for Genomic Regulation

INB
Instituto Nacional de Bioinformática

Use Case: Human data

Mission: develop a long-term management policy for controlled-access human data and provide tools for data discovery and access

EGA as the central ELIXIR service for secure sharing of -omics data



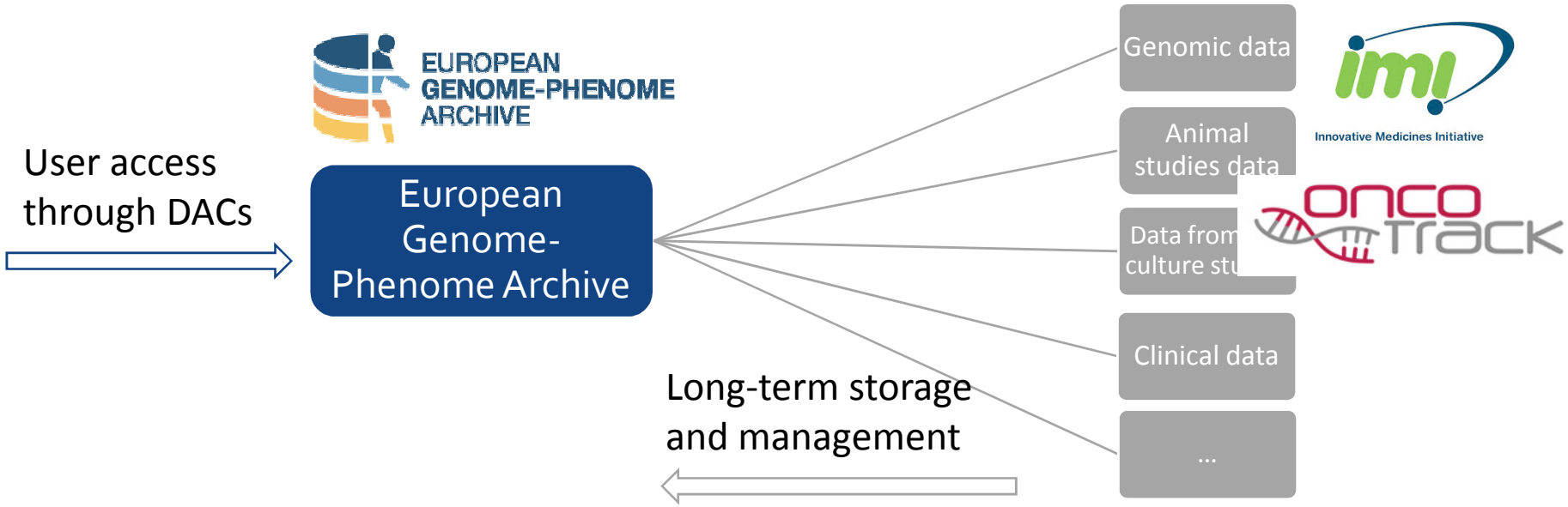
Projects:

- Connect **TRaIT platform (NL) with EGA**
- **IMI OncoTrack –** explore long-term storage and data-discovery for OncoTrack's data
- **Beacons –**discovery service for genomics data
- **Local EGA** - Stores sensitive data locally Integrates metadata globally in EGA (part of ELIXIR Data Nodes network)

Arcadi Navarro, ES

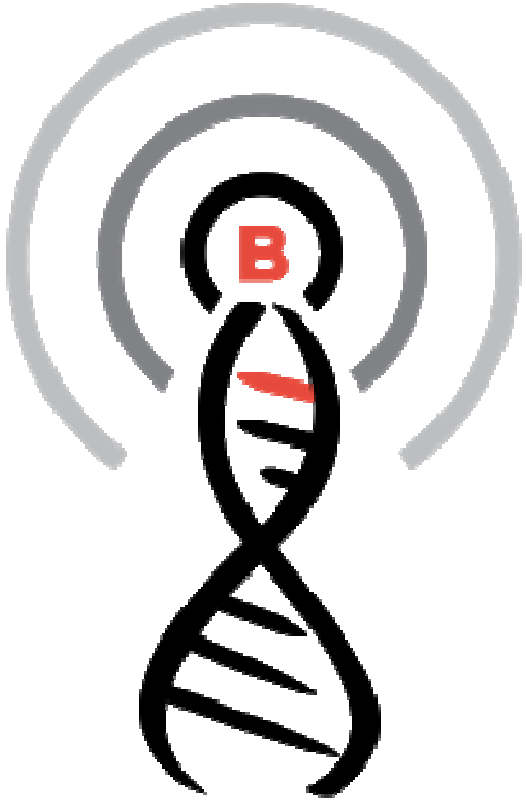
Accessible data: ELIXIR – IMI OncoTrack

- Scoping study to understand long-term knowledge management requirements
  - Data storage, Meta data mappings, Data governance - Consent and Access management

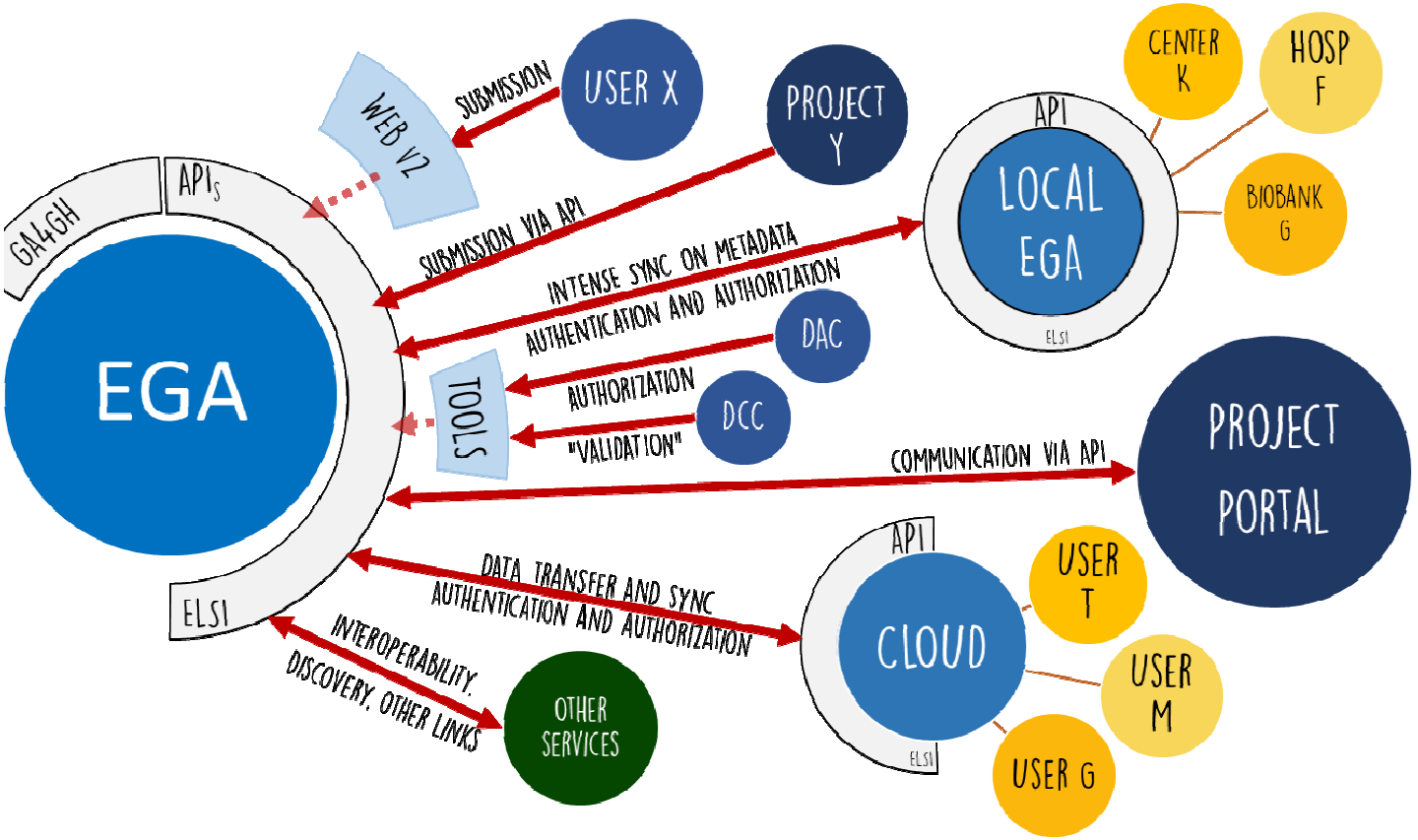


ELIXIR - GA4GH Beacon project

- **GA4GH Beacons are a discovery service:**
  - which datasets include genomes with allele of interest
- **ELIXIR pilot project with partners from the Netherlands, Sweden, Finland and Spain.**
- **Complies to the GA4GH Beacon project standard and security working group policies.**
- **Three objectives:**
  - Provide ELIXIR reference implementation on GA4GH Beacon with 3 authorization levels.
  - Provide ELIXIR standards for data types
  - Provide an example on capacity and expertise build across ELIXIR Nodes to integrate national resources as part of a joint service interface.



Perspectives New Application INB/ELIXIR-ES



EUROPEAN GENOME-PHENOME ARCHIVE

Global Alliance for Genomics & Health

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

elixir SPAIN

openMINTED

IPCEI

Thanks

Alfonso Valencia
ICREA Research Professor
Director, Life Sciences Department BSC
Director, Spanish National Bioinformatics Institute (INB-ISCIII)
Head of the Spanish Node of ELIXIR (ELIXIR-ES)

valencia@bsc.es
@alfons_valencia