The experience of Andalusia in eHealth and big data

Big data in health - IMI's HARMONY project, European Parliament, Brussels, 19 June 2018

Joaquín Dopazo

Área de Bioinformática, Fundación Progreso y Salud, Nodo de Genómica Funcional, (INB-ELIXIR-es), Bioinformática de ER (BiER-CIBERER), CDCA, Hospital Virgen del Rocío, Sevilla, Spain

http://www.clinbioinfosspa.eshttp://www.babelomics.org@xdopazo, @ClinicalBioinfo



Fundación Progreso y Salud CONSEJERÍA DE SALUD

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge: prospective healthcare
- Data integration
- GDPR compliance



The clinical bioinformatics area



Rare diseases Cancer **Bioinformatics** Common diseases Infectious diseases

http://www.clinbioinfosspa.es/

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 <u>genomic data</u> for <u>precision diagnostic</u> and <u>treatment</u> recommendation, implementing a <u>prospective health care</u> functionality in the public health system.

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge: prospective healthcare
- Data integration
- GDPR compliance



Data analysis and the sustainability of the cycle of knowledge generation



Sustainability requires tools for end users, which involves hiding the complexity of the analysis

- A solution for the management of genomic data must be integrated the same way that other analyses of the health system.
- Genomic data are stored in the system, linked to clinical data the same way that other data are (medical image, digital pathology –under implementation-, etc.) for further potential clinical studies



Our approach: hiding the complexity



Personalized Medicine in cancer





Enhanced use of biomarkers

Patient genomic data analysis allows one-step association of biomarkers with therapies and enables the <u>detection of new actionable</u> <u>biomarkers</u>, or <u>clinical trials</u> compatible with patients saving time and cost and increasing treatment success

Front end: Personalized Medicine Module (MMP)

					_
CBA IVA-ACCI v0.9.0	Start		📸 acci_admin = Studies = About = 🕪 Logout	○ CBA NOLADO VESE Int	(+ topot
Projects / acci_admin@other_grch37 /	other_acci			P19603 / 403.489m@1044.40707 / 6844.400	
Clinical Analysis				d. Clinical Analysis	
	A linkast	T Prioritization		4/1000 A/2000 - 1000000	
Q Search					- 18
Samola 0	Cicar No filters selected		▼ Filters -		- 8
United and U	Sample Results			Q Security Bit Date tabled T	101 F
Sample name	Cample Results			Netv	_ 11
Hou 1673, HOU 1660, HOU 1661	Showing 1-6 of 6 samples		0 0	El Tala Asanti El Concrete Bonnon Bonno	
By Date ® Al	Sample In	Individual ID Date Status Sex Diagnosis	Father Mother Cell Line	beergees beerge	Char
 Recently Uploaded Date (Year-Month-Day) 	U ARSS-1 P	PaterPamilas 2017-9-31 READY MALE -	Germine	Searcano. •	110000
2017 ¥ - Any ¥ - Any ¥	© \$1937-4 F	Filus 2017-8-31 READY MALE Other juvenile osteochondrosi	Gernine	Beden Spanje Conservation () Lagand BER	
Patient Filters	D1-RyC_S1_L001_R1_001.faste_mapped_to_hg19 N	MguelAngel 2017-10-02 READY FEMALE Disorders of vestibular function	Germine	Striph of 01 of of a server server terms Tax model and a company by STT Pagear GAD Plays making cetter tell Servers D	xw:
Individual ID	Sample1 S	SAI 2017-10-03 READY MALE IV. Endocrine, nutritional and r	vetabolic diseases Gerntine	A0041 0 0 0 0	-
Smith, Grant HPO	Showing 1 to 5 of 5 rows			54374 0 0 0 0	
HP:030145					•
Smith, Grant	Analysis Type	Sample col	oction	Genoret:	
		Sample Ser		Possiblen Processor	_
	Analysis Analysis Analysis Analysis	•		Variant prioritization	_
	Clinical Analysia				_
	Clinical Analysis				
	x Delete		0 III -		_
	Analysis ID Samples A	Analysis Type Definition Date Subject	t Disease Action	0 4470007 #0000F A000FF440214 260 45 45 MILLION 45 45 MILLIONAUMAN 6.22 4.22 6.02 4.45 MILLION 4	
		No matching records found			_
					- 1 C
CILL MARCELORS AND INCOME.			🗿 der i faire i Mairi Dringeri	© CRA 100-0000 - 10000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000 - 1000	1- 0+L004
				Interpretation Analysis	
Obarray				manpresentes i recompanie	
Sample	tatjur	Family		Obarray	
0.001	Nue Martine	Same			
Party	PHE 200	20 Million and a second s			
Contraction of the second seco	a conservations and a conservation of the cons	Prises.		The state of the s	
	Lonarystry	• Contraction Contraction		Trans the Transmission Transmission Transmission Transmission	Circuit Insiyin
	Descipes managements and			Variants for	
	Remain should	AND CARE MANY OF STORE 1			
				the mean ant	
Padeds -	intern ben	Parantinan			
User (I	Certi) Tellin(2 see	ter PON some ter		tyme 19.5 tourn 2010 and an and branch any collocated any collocat	
17-2) Bolly 27	t n/T many	Not present		Anaja Bagawana Ba	
ute 1 Devisi	Report	t deneration	tte.	state: so,enorgene_person,eno	
Una ministra		gonoradion			
				Reported Variants	
Danodad	(Sent to	O THE EHR)		2 met hav 0	
Neponeo				Deriv	
Branks Hid Ital	25		E Einde Report O	mean Gass Nas Castanan Pastana Materia Declarate Declarate Declarate Material Statement	
		Lignaly.		* 21046415 2017 DV - 0 0 0 0 0 - 5444 78 - 2405150	
Periori	Denn har GenerAmatalan Penduka Mendesankelan	Advertantian Philosophy Manadembry Philosophy Comp	Another Internation	and the second sec	
- Charles	• •	Carly 19	Covera D		
		18.	*	During 18141 max	
					-
P	Fundación Proc	areso v Salud			
	Fundación Prog	greso y Salud			

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge: prospective healthcare
- Data integration
- GDPR compliance



Backend: OpenCGA, a scalable storage and genomic data management platform



In collaboration with Genomics England (GEL)

Currently, the fastest and more powerful genomic database engine in the world. Used in the GEL for genomic data management



Extensive capabilities to query across genotype and phenotype relationships

https://github.com/opencb/opencga

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge: prospective healthcare
- Data integration
- GDPR compliance



Prospective healthcare is facilitated by a model that integrates genomic data and universal EHR



- Revolutionary concept: the whole health system becomes an enormous potential prospective clinical study
- <u>Clinical</u> data <u>dynamically</u> associated to <u>genomic</u> data
- Possibility of many clinical studies by <u>reanalyzing genomic data</u> under diverse perspectives (with no extra investment)
- Growing genomic DB with increasing study possibilities

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge: prospective healthcare
- Data integration
- GDPR compliance



Future vision involves *big data* integration: Genomic data are especially relevant but not the only useful *big data*



- Other *big data* are being collected (medical image, digital pathology, wearable devices, etc.)
- <u>Clinical</u> data <u>dynamically</u> associated to different <u>big data</u>
- The whole health system becomes a enormous potential prospective clinical study
- Immense possibility for data reusability
- Growing genomic DB with increasing study possibilities

- Genomic big data for personalized medicine
- Bioinformatics
- Sustainability
- Scalability
- Generation of knowledge
- Data integration
- GDPR compliance



GDPR compliance

The system has been designed in a way that is compliant with EU and Spanish General Data Protection Regulation

- Clinicians requesting for a genomic diagnostic have access to eHR and get the result of the test.
- Geneticists have access to eHR and can query the genomic data (but never extract them)
- IT have access to deidentified genomic data and no to eHR.



Genomic and clinical data within the health system enable Personalized Medicine

- Database of patients with prospective clinical information. Patients sequenced:
 - Will have different responses to treatments in the future
 - Can have other diseases in the future
 - Dynamic diagnostic of undiagnosed patients as knowledge databases update
 - Dynamic assignment of treatments for patients without therapeutic options as knowledge databases update
- Preventive medicine:
 - Dynamic discovery of pharmacogenomics relevant variants in sequenced individuals
 - Dynamic discovery of new risk variants in sequenced individuals
 - Dynamic discovery of reproductive risk variants
- Health system as a prospective genomic study for clinical knowledge generation:
 - Prospective discovery of new biomarkers of response to drugs, therapies, prognostic, etc.
 - The pool of disease or risk variants is limited and could be surveyed soon

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

earning

sonalized



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







•••