Transformando el Sistema Público de Salud mediante el uso de los datos genómicos del paciente: perspectivas y desafíos

Joaquín Dopazo
Director Área de Bioinformática
Fundación Progreso y Salud
Functional Genomics Node, (INB-ELIXIR-es),
Bioinformatics in Rare Diseases (BiER-CIBERER)
Sevilla

@xdopazo, @ClinicalBioinfo
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.
Personalized health care and the transition to precision medicine

Intuitive Medicine
- Based on trial and error

Empirical Medicine
- Identification of probabilistic patterns

Precision Medicine
- Decisions and actions based on knowledge

Degree of personalization

Today

Tomorrow
Empirical medicine
Phase I: generation of knowledge

Patient Variants

Clinical genomic studies enable knowledge generation

Genomic variants (biomarkers) can be quickly associated to precise diagnosis or therapy outcomes

Initially the system will need much feedback: Knowledge generation phase.
Precision medicine. Phase II: using the knowledge database

1) Genomic sequencing
2) Database of biomarkers
3) Therapy prediction

Other factors (risk, cost, etc.) + Hints on possible prescriptions
An historical perspective: International projects (ICGC) and the local genomic initiatives in Spain

- EnoD (CIBERER) 2016
- CIBERER 2012
- RareGenomics 2018
- NaGen 2016
- MedPerCan, URDcat 2017
- CLL 2009-2014
- MGP 2011

We were there 8 years ago
Current solutions for managing genomic data of patients

• No scalability
• Expensive
• Lack of experts
• Inequity

Commercial software

Bioinformatician and biostatistician teams
"Experts from hundreds of institutions interpret variants and flag them with the appropriate level of pathogenicity … feed the variant knowledge base"!!!!!!

"Our database of 200,000 patients tested"!!!!!

We are paying for the use of software of companies that collect for free genomic data and information generated by our experts (their customers), which in turn increase the value of the service offered by the companies.

Our genomic data are externalized with no value for the health system.
Solutions for managing the genomic data of the patient

- Expensive (pay per use)
- GDPR non compliant
- Inequity

Bioinformatician and biostatistician teams

- No scalability
- Expensive
- Lack of experts
- Inequity

Our solution: corporative software

- Equity
- Scalability
- Affordable
- End user: clinician
- GDPR compliant
Use of genomic data in the public health system requires sustainability

User: clinician (not bioinformatician)

- Tools for end users, which involves **hiding the complexity** of the analysis to the clinician
- A solution for the management of genomic data must be **integrated** the same way other analyses of the health system are.

- **Genomic** data must be **stored** in the system, **linked** to clinical data the same way that other data are for further potential prospective clinical studies
Complexity of the general diagnosis protocol (rare diseases)

- **Known genes**
- **Suspected diagnosis**
- **Unexpected findings:**
  - Pharmacogenomics
  - Actionable diseases
  - Reproductive risk

**Disease panel:**
- Diagnostic variants
- Genes

**VUS in Known genes found**
- yes: VUS in Known genes prioritization successful
- no: VUS everywhere found

**VUS prioritization successful**
- yes: VUS everywhere found
- no: VUS prioritization successful

**Sample QC**
- yes: Report positive diagnosis
- no: Report negative diagnosis

**Variants found**
- yes: Expert validation
- no: yes

**50-70% < 1 min.**
- yes: yes
- no: yes
Our approach: hiding the complexity
Decision support systems democratize the use of complex (genomic) data

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

- Sequencing Unit
- eHR
- Clinical research
- Knowledge

1. Corporative
2. ?
3. 3
4. 4
5. 5
6. 6
7. 7
8. 8
Front end: Personalized Medicine Module (MMP)

Sample selection

Variant prioritization

Selection of variants for the report

Report generation (sent to the eHR)
Currently, the fastest and more powerful genomic database engine in the world. Used in the GEL for genomic data management.

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

In collaboration with Genomics England.
We share the backend with the 100,000 genomes project

CellBase, the Knowledge base and OpenCGA, the genomic data management engine are projects initiated in our group by 2010. Now are the backbone of the GEL

Ignacio Medina, Head of Computational Biology Lab HPC Service, University of Cambridge, and Head of Bioinformatics Databases at Genomics England has been building many of the applications that sit on top of MongoDB. He said:

“MongoDB is performing beautifully for us. From the beginning of the project it’s been fantastic for our developers to iterate quickly. Now that the 100,000 Genomes Project is running at scale, MongoDB is also helping us extend our experience on to the scientists and clinicians who access our data, making it easier and faster for them to find critical insights in the data.”

Two of the important projects also utilising MongoDB are Cellbase and OpenCGA (Computational Genomics Analysis). Cellbase is a data warehouse and open API that stores reference genomic data from public resources such as Ensembl, Clinvar, and Uniprot. By relying on MongoDB, Cellbase can typically run sophisticated queries in an average of 40 milliseconds or less, and complex aggregations in less than one second – down from six hours using previous filesystem-based querying and storage. Importantly, it can annotate about 20,000 variants per second, making it compatible with whole genome sequencing data throughput requirements, while also returning a rich set of annotations that helps scientists better understand the data.
Personalized Medicine in cancer

Current use of biomarkers

1\textsuperscript{st} line

Biomarker 1 → Therapy 1

2\textsuperscript{nd} line

Biomarker 2 → Therapy 2
Biomarker 3 → Therapy 3

3\textsuperscript{rd} line ......

Prospective healthcare

Enhanced use of biomarkers

Patient genomic data analysis allows one-step association of biomarkers with therapies and enables the detection of new actionable biomarkers, or clinical trials compatible with patients saving time and cost and increasing treatment success.
## Niveles de evidencia para la recomendación de tratamiento en cáncer

<table>
<thead>
<tr>
<th>Nivel</th>
<th>Descripción</th>
<th>Implicaciones terapéuticas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Biomarcador de tratamiento estándar predictivo de respuesta a un fármaco indicado para este cáncer</td>
<td>estándar</td>
</tr>
<tr>
<td>2A</td>
<td>Biomarcador predictivo de respuesta a tratamiento con un fármaco indicado para una mutación de este cáncer*</td>
<td>*Incluye biomarcadores de ensayos basket</td>
</tr>
<tr>
<td>2B</td>
<td>Biomarcador de tratamiento estándar predictivo de respuesta a un fármaco indicado para otro cáncer</td>
<td></td>
</tr>
<tr>
<td>3A</td>
<td>Biomarcador con evidencia clínica convincente de respuesta de este cáncer a un fármaco que no es indicación terapéutica</td>
<td>Terreno de investigación terapéutica. Posibilidad de reclutamiento en ensayos clínicos</td>
</tr>
<tr>
<td>3B</td>
<td>Biomarcador con evidencia clínica convincente de respuesta de otro cáncer a un fármaco que no es indicación terapéutica</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Biomarcador con evidencia biológica convincente de respuesta de algún cáncer a un fármaco que no es indicación terapéutica</td>
<td>Implicaciones terapéuticas hipotéticas</td>
</tr>
<tr>
<td>R1</td>
<td>Biomarcador de tratamiento estándar predictivo de resistencia a tratamiento con un fármaco indicado para este cáncer</td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>Biomarcador no estándar con evidencia clínica convincente de ser predictivo de resistencia a tratamiento con un fármaco</td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>Biomarcador no estándar con evidencia biológica convincente de ser predictivo de resistencia a tratamiento con un fármaco</td>
<td></td>
</tr>
</tbody>
</table>

*Adaptado de: Chakravarty et al. JCO precision oncology 2017*
However, precision diagnostic is only the beginning to implement personalized medicine in the health system

• Precision diagnostic using genomic data can be carried out everywhere (only a sequencer and the appropriate software is required)
• Genomic data generation in a disconnected health system generates silos of data at different hospitals or even departments, which limits the sample size for clinical studies
Actually, genomic initiatives are just clinical studies but not Personalized Medicine yet

- Each study requires a specific genomic and clinical data collection into an external database
- Serious security concerns (genomic + clinical data outside the hospital)
- 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), PERIS, RAREgenomics, etc.
The real implementation of Personalized Medicine is facilitated by a model that integrates genomic data and universal EHR

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

Health systems with no universal EHR will suffer extra problems in implementing a real practice of personalized medicine. There are solutions not exempt of complications:

- External repositories (GDPR, encryption)
- Federated repositories (security breaches)
The population health database
Possibly the largest database ever created with detailed clinical data, storing information on 12,083,681 patients since 2001
Lessons learned from MGP: the importance of local variability

We discovered some 12,000 “Spanish” polymorphisms not present in other databases. The filtering efficiency enormously increases using local population data.
The CSVS is a crowdsourcing project

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

http://csvs.babelomics.org/

Allelic population frequencies obtained from 1,600 exomes are currently available in CSVS
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 only 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 anonymized genomic data but not to eHR.
Future vision involves **big data integration**: Genomic data are especially relevant for discovering the genetic determinants of diseases, but not the only useful **big data**

- Other **big data** are being collected (medical image, digital pathology, wearable devices, etc.)
- Microbiota in the future (CR cancer screening)
- Clinical data in the BPS will be **dynamically** associated to different **big data**
- The whole health system becomes a enormous potential **prospective clinical study**
- Immense possibility for data **reusability**
- Growing genomic DB with increasing study possibilities
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

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