TOWARDS AN INTEGRATED INTERACTIVE DATABASE FOR THE SEARCH OF STRATIFICATION BIOMARKERS IN ALKAPTONURIA

Vittoria Cicaloni

Siena, 12 Aprile 2017
PRECISION MEDICINE (PM)

- Approach for prevention/diagnosis/treatment of disease based on individual variability
- Final aim: suitable therapy

[1. ec.europa.eu/research/health/biomarkers-for-patient-stratification.en, 2010
Precision Medicine Iniziative (PMI)

“Tonight I’m launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes.

And to give us all access to the personalized information we need to keep ourselves and our families healthier.”

President Barack Obama
2015 State of the Union Address | January 20, 2015

“Most medical treatments have been designed for the ‘average patient’ ... treatments can be very successful for some patients but not for others.”

President Barack Obama
January 30, 2015
Precision Medicine Initiative (PMI)

- A national cohort of one million or more U.S. participants

- Allocation of $215 million in fiscal year 2016 to support the Initiative

- The cohort reflects the diversity of the U.S. population: diverse social, racial/ethnic, populations living in a variety of geographies, social environments, and economic circumstances, and from all age groups and health statuses.

- Common diseases such as diabetes, heart disease, Alzheimer’s, obesity, and mental illnesses like depression, bipolar disorder, and schizophrenia, as well as rare diseases.
Prevent Health risk

Personalized Health(care) planner
GPS to health

Route 1
= Default → First signs of disease risk

Route 2
→ Alternative route
1: Patient stratification

Principle of Personalized/Precision/Targeted Medicine

- Drug toxic but beneficial
- Drug NOT toxic and NOT beneficial
- Same diagnosis, same prescription
- Drug toxic but NOT beneficial
- Drug NOT toxic and beneficial

2: Biomarkers discovery

Biological characteristic (molecular, anatomic, physiologic, or biochemical) which can be measured and evaluated objectively. It acts as indicator of an ordinary or a pathogenic biological process.
Precision Medicine and rare diseases

RARE DISEASES BY THE NUMBERS

A disease is defined as orphan in the U.S. when it affects fewer than 200,000 people.

There are approximately 7,000 types of rare diseases and disorders.

95% of rare diseases have no FDA-approved drug treatment.

80% of rare diseases are genetic in origin.

Approximately 50% of those affected by rare diseases are children.

30% of children with a rare disease will not live to see their fifth birthday.

8: Average number of physicians visits before diagnosis

3: Average number of misdiagnoses

7+ years: Average time until diagnosis

SOURCES: National Organization for Rare Diseases, Global Genes Project

Data management in rare diseases

Statistic analysis
ALKAPTONURIA (AKU)

- Ultra-rare disease (950 patients in the world)
- Defect in Homogentisate 1,2-dioxygenase (HGD) that produces Homogentisic Acid (HGA) accumulation
- Multisystemic disorder
- No approved biomarkers and drugs
WORKFLOW:

Collect AKU patients data → Database → Analyze data → Search networks
DATA COLLECTION:

**English and Slovak patient associations**
- 15 UK patients
- 25 SK patients

**aimAKU (http://www.aimaku.it/):**
- 70 IT patients
WORKFLOW:

1. Collect AKU patients data
2. Analyze data
3. Search networks
4. Database
Collaboration:

UNIVERSITÀ DEGLI STUDI DI SIENA
Dipartimento di Ingegneria dell’informazione e Scienze matematiche
Prof.ssa Monica Bianchini

Dott. A. Rossi

Dott. M. Zazzeri

Dott. A. Zugarini
1) PATIENTS AND MUTATIONS
2) BIOMARKERS
3) BLOOD ANALYSIS
4) LIFESTYLE
5) CONCOMITANT DISEASES
6) HEALTH STATUS
7) DRUGS
8) PLASMA ANALYSIS
9) URINE ANALYSIS
10) HISTOPATHOLOGY
11) HETEROZYGOUS

DATABASE
DATABASE:

http://www.sbl.unisi.it/aprecisekure/

- Relational database
- MySQL/PHP

What is Alkaptonuria?

Alkaptonuria (AKU) is a rare autosomal recessive disorder resulting from homogentisate 1,2-dioxygenase (HGD) deficiency, an enzyme involved in Tyrosine and Phenylalanine metabolism. This defect leads to homogentisic acid (HGA) accumulation which causes discoloration of bone (a process called “ochronosis”) and induces early-onset osteoarthitis. AKU is a painful and degenerative disease and has also been linked to heart disease and kidney stones.

Precision Medicine in AKU

Precision medicine (PM) is an emerging approach for the study of diseases based on individual variability in genes, environment, and lifestyle to customize different illness courses. For PM application to rare diseases like AKU, collecting as much information as possible about each patient, not overlooking apparently ordinary factors, is crucial. Here, genetic, proteic, biochemical, histopathologic and clinical data about AKU patients are organized as database and freely available. Indeed, researchers, clinicians and patients can easily access the information, as well as being able to insert new data or updating entries. Furthermore, this integral database aims to the discovery of new Alkaptonuria biomarkers, thanks to a refreshable correlation system.
## DATABASE:

[http://www.sbl.unisi.it/aprecisekure/]

### Stress Ox Inflammation and Amyloidosis Biomarkers

<table>
<thead>
<tr>
<th></th>
<th>saa</th>
<th>serum_hga</th>
<th>hga_hplc</th>
<th>catd</th>
<th>il-6</th>
<th>il-1beta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>62.61</td>
<td>4.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>67.2</td>
<td>3.88</td>
<td></td>
<td></td>
<td></td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>84.5</td>
<td>3.92</td>
<td></td>
<td></td>
<td></td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>185.17</td>
<td>4.18</td>
<td></td>
<td></td>
<td></td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>100.94</td>
<td>4.34</td>
<td></td>
<td></td>
<td></td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>64.04</td>
<td>4.18</td>
<td></td>
<td></td>
<td></td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>8.72</td>
<td>21.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.57</td>
<td>28.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Serum Amyloid A (μg/mL (ELISA))**
WORKFLOW:

Collect AKU patients data

Analyse data

Search networks

Database
DATA ANALYSIS:

PEARSON’S CORRELATION and P-VALUE
## Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>age</th>
<th>saa</th>
<th>serum_hga</th>
<th>hga_hplc</th>
<th>catd</th>
<th>il-6</th>
<th>il-1beta</th>
<th>il-1ra</th>
<th>tn_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>ldl_cholesterol</td>
<td>0.234</td>
<td>-0.042</td>
<td>0</td>
<td>0.117</td>
<td>-0.07</td>
<td>-0.216</td>
<td>-0.056</td>
<td>-0.016</td>
<td></td>
</tr>
<tr>
<td>alkaline_phosphatase</td>
<td>0.197</td>
<td>0.109</td>
<td>0</td>
<td>-0.164</td>
<td>0.132</td>
<td>-0.155</td>
<td>0.182</td>
<td>0.101</td>
<td></td>
</tr>
<tr>
<td>cystatin</td>
<td>-0.017</td>
<td>-0.362</td>
<td>0</td>
<td>-0.315</td>
<td>-0.011</td>
<td>-0.007</td>
<td>0.144</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>smoker/cigarettes_a_day</td>
<td>0.056</td>
<td>-0.115</td>
<td>0</td>
<td>0.096</td>
<td>-0.085</td>
<td>-0.104</td>
<td>0.067</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td>alcohol_units_weekly</td>
<td>-0.163</td>
<td>0.119</td>
<td>0</td>
<td>0.411</td>
<td>-0.031</td>
<td>-0.06</td>
<td>-0.018</td>
<td>-0.10</td>
<td></td>
</tr>
</tbody>
</table>
Correlation Matrix

- Overview about all possible numeric data correlations
- Discover interconnections
- Alkaptonuria and transversal studies
- Monitoring on the correlations evolution thanks to a refreshable system
WORKFLOW:

1. Collect AKU patients data
2. Analyze data
3. Search networks
4. Database
SEARCH NETWORKS:

- STITCH (http://stitch.embl.de/)
- KEGG (http://www.genome.jp/kegg)
Cathepsin D – Cystatin C

INVERSE STATISTICAL CORRELATION

AKU → Kidney stones → Cystatin C

Ordinary values:
Female: 0.52 to 0.90 mg/L
Male: 0.56 to 0.98 mg/L

F: 2/11 AKU patients = 0.90 mg/L
M: 1/29 AKU patients > 0.98 mg/L

3. Khalkhali-Ellis Z. Biology and Medicine, 2014.]
CLINICAL INTERPRETATION:

Arthritis

GFR marker

↑CatD

CysC

DEGRADATION

Correlations:

- Serum Amyloid A (SAA)-Age

**STATISTIC INTERPRETATION:**

<table>
<thead>
<tr>
<th>SAA-Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s index</td>
</tr>
<tr>
<td>P-test</td>
</tr>
</tbody>
</table>

- 47/64 AKU patients SAA > 5μg/mL
- Accumulation
CRP (Immunoturbidimetric)-CRP (ELISA)

STATISTIC INTERPRETATION:

<table>
<thead>
<tr>
<th>CRP (immunot.)-CRP (ELISA)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson’s index</td>
<td>0.67</td>
</tr>
<tr>
<td>P-test</td>
<td>0.000</td>
</tr>
</tbody>
</table>

✓ Kidney function marker
✓ Third National Health and Nutrition Examination Survey
Conclusion:

- Database development based on patients data
- Knowledge network building in order to develop an AKU-dedicated Precision Medicine Ecosystem
- Enhancement of new scientific knowledge related to AKU disease
Integration of clinical data with “omics” data for a completed Precision Medicine Approach

Reference: Craig E. Wheelock, ERS journal, 2013
Omics Technology: refers to a field of study in biology ending in -omics.

Genomics:

- Genomics is a branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes

- Genome is a complete set of DNA within a single cell

- Genetics refers to the study of genes and their roles in inheritance

- Genes direct the production of proteins. If a cell's DNA is mutated an abnormal protein may be produced
Proteomics:

✓ Proteomics is the large-scale study of proteomes.

✓ Proteome is a set of proteins produced in an organism, system or organ.

✓ The proteome is not constant: it differs from cell to cell and changes over time.

✓ Proteomics is used to investigate when and where proteins are expressed, rates of protein production, degradation and steady-state abundance.
Proteins:

- Proteins are made up of hundreds or thousands of smaller units called **amino acids**, which are attached to one another in long chains.

- **Protein primary structure** is the linear sequence of amino acids in a peptide or protein.

- **Protein secondary structure** is the three dimensional form of local segments of proteins. The two most common secondary structural elements are alpha helices and beta sheets.

- **Protein tertiary structure** is the three dimensional shape of a protein.

- **Protein quaternary structure** is the number and arrangement of multiple folded protein subunits in a multi-subunit complex.
What are proteins and what do they do?

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibody</td>
<td>Antibodies bind to specific foreign particles, such as viruses and bacteria, to help protect the body.</td>
</tr>
<tr>
<td>Enzyme</td>
<td>Enzymes carry out almost all of the thousands of chemical reactions that take place in cells. They also assist with the formation of new molecules by reading the genetic information stored in DNA.</td>
</tr>
<tr>
<td>Messenger</td>
<td>Messenger proteins, such as some types of hormones, transmit signals to coordinate biological processes between different cells, tissues, and organs.</td>
</tr>
<tr>
<td>Structural component</td>
<td>These proteins provide structure and support for cells. On a larger scale, they also allow the body to move.</td>
</tr>
<tr>
<td>Transport/storage</td>
<td>These proteins bind and carry atoms and small molecules within cells and throughout the body.</td>
</tr>
</tbody>
</table>
Homogentisate 1,2-dioxygenase (HGD)

- Enzyme involved in the metabolic pathway of tyrosine and phenilalanine degradation
- Hexameric structure arranged as dimer of trimers
- Dioxygenase O2 dependent having Fe$^{2+}$ ion as cofactor
- A little error in HGD amino acidic code could lead to misfolding of the enzyme and loss of its activity (Alkaptonuria).
WHAT ABOUT THE STRUCTURE???
AKU is a genetic disease

In biology, a mutation is the permanent alteration of the nucleotide sequence of the genome of an organism. Mutations result from errors during DNA replication.

**Missense mutations:** refer to a change in one amino acid in a protein, arising from a point mutation in a single nucleotide. In AKU there are more than 100 different missense mutations affecting HGD (Homogentisate 1,2- dioxygenase) protein which cause disease.
Structure characterization of missense mutations

- Core
- Active site
- Accessible solvent surface
- Protomer Interface

MOLE 2.0
Rapid and fully automated location and characterization of pores

PyMOL
Surfaceres
MD Trajectory

ProCoCoA (Protein Core Composition Analyzer)
http://www.sbl.unisi.it/prococoa/
Molecular Dynamic of HGD protein

- Fe$^{2+}$ parameterization in AMBER99sb force field
- GROMACS Molecular Minimization
- 100 ns of Molecular Dynamic simulation (MD)
- Analysis of MD trajectory
Molecular Dynamics (MD)

Molecular dynamics (MD) is a computer simulation method for studying the physical movements of atoms and molecules, and is thus a type of N-body simulation.

The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamical evolution of the system.

The trajectories of atoms and molecules are determined by numerically solving Newton's equations of motion for a system of interacting particles, where forces between the particles and their potential energies are calculated using interatomic potentials or molecular mechanics force fields.
MD is generally used for:

Determining where drug molecules bind, and how they exert their effects

We used simulations to determine where this molecule binds to its receptor, and how it changes the binding strength of molecules that bind elsewhere.

Dror et al., Nature 2013
Determining functional mechanisms of proteins

Simulation started from active structure vs. Inactive structure

- We performed simulations in which a receptor transitions spontaneously from its active structure to its inactive structure.
- We used these to describe the mechanism by which drugs binding to one end of the receptor cause the other end of the receptor to change shape (activate).

Rosenbaum et al., Nature 2010; Dror et al., PNAS 2011
Protein folding is the physical process by which a protein chain acquires its native 3-dimensional structure, a conformation that is usually biologically functional, in an expeditious and reproducible manner. It is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from random coil. Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids.
Red zones represent *HGD* protein amminoacids sensitive to substitution (missense mutation) which cause AKU.
Metabolomics:

- Metabolomics is the scientific study of chemical processes involving metabolites.

- Metabolome refers to the complete set of small-molecule metabolites.

- Metabolites are the intermediates and products of metabolism, they are usually defined as molecules less than 1 kDa in size.

- The metabolome is dynamic.
Metabolites are influenced by both genetic and environmental factors.

Metabolic profile is highly dynamic compared with genes and protein levels.

Metabolomics best represents the molecular phenotype.

Urine and Nuclear Magnetic Resonance (NMR) are the bio-fluid and technique of choice for screening.

Urine samples can be easily collected and transported.

Metabolomics data analysis

- Multivariate analysis (Metaboanalyst, R software, XCMS)
- Identification of potential prognostic or predictive biomarkers
- Enhancement of AKU knowledge network
- Deeper understanding of biological pathways

GOALS OF DIFFERENT RESEARCH LINES:

- Patient Stratification
- As AKU is a chronic disease, it may be interesting to monitor the clinical conditions over time
- Model of precision medicine for other inborn errors of metabolism.
Acknowledgments

Dr. Bernini Andrea

Dr. ssa Spiga Ottavia

Prof.ssa Santucci Annalisa

Galderisi Silvia PhD. Student

Cicaloni Vittoria PhD. Student

Trezza Alfonso PhD. Student

Cevenini Lorenzo Graduated Student

Mecacci Leonardo Graduated Student
“The study of a rare disease: butterfly collecting or an entretè to understanding common conditions? “
(K. Talbot, Pract Neurol. 2007)

Thank you for your attention