A semantic collaborative system for the management of translational research projects

Matteo Gabetta, Giuseppe Milani, Cristiana Larizza, Valentina Favalli, Eloisa Arbustini, Riccardo Bellazzi
Outline

- The INHERITANCE project
- Biomedical Informatics Tools
- Semantic Wiki
  - Technologies
  - Organizational Data Management
  - Scientific Data Management
    - NLP
    - Literature Mining
- Conclusions
The INHERITANCE project

Cardiomyopathies:

"primary myocardial disorders of unknown cause"

4 main subtypes:

- Hypertrophic (HCM)
- Dilated (DCM)
- Restrictive (RCM)
- Arrhythmogenic Right Ventricular (ARVC)
Cardiomyopathies:

"primary myocardial disorders of unknown cause"

4 main subtypes:
- Hypertrophic (HCM)
- Dilated (DCM)
- Restrictive (RCM)
- Arrhythmogenic Right Ventricular (ARVC)
Dilated Cardiomyopathy:

"[...] myocardial disorder characterized by the presence of left ventricular dilatation and systolic impairment, in the absence of abnormal loading conditions (e.g. hypertension, valve disease) or coronary artery disease sufficient to cause global systolic dysfunction." *

Dilated Cardiomyopathy:

"[...] myocardial disorder characterized by the presence of left ventricular dilatation and systolic impairment, in the absence of abnormal loading conditions (e.g. hypertension, valve disease) or coronary artery disease sufficient to cause global systolic dysfunction." *

➢ **20 disease-causing genes** (to date)

Dilated Cardiomyopathy:

The INHERITANCE project

INtegrated HEart Research IN TrANslational genetics of Cardiomyopathies in Europe

- 3-year health research project
- European Commission Funding Program 7
- 11 European centers
The INHERITANCE project

INtegrated HEart Research In TrANslational genetics of Cardiomyopathies in Europe

Translational strategy:

**Disease-specific features**
*(red flags)*

**Biological features**
*(genetic or metabolic pathways)*
The INHERITANCE project

INtegrated HEart Research In TrANslational genetics of Cardiomyopathies in Europe

6 research areas:
  - Clinical Cardiogenetics
  - -omics
  - Animal Studies
  - Structural Studies
  - Treatments
  - Biomedical Informatics
The INHERITANCE project

INtegrated HEart Research In TrANslational genetics of Cardiomyopathies in Europe

6 research areas:

- Clinical Cardiogenetics
- -omics
- Animal Studies
- Structural Studies
- Treatments
- Biomedical Informatics
Biomedical Informatics Tools

- Data Warehouse
- Automated Literature Analysis
- Case-Based Reasoning
- Literature-Based Gene Prioritization
- Semantic Wiki
Biomedical Informatics Tools

- Data Warehouse
- Automated Literature Analysis
- Case-Based Reasoning
- Literature-Based Gene Prioritization
- Semantic Wiki
Biomedical Informatics Tools

- Data Warehouse
- **Automated Literature Analysis**
- Case-Based Reasoning
- Literature-Based Gene Prioritization
- Semantic Wiki
Biomedical Informatics Tools

- Data Warehouse
- Automated Literature Analysis
- **Case-Based Reasoning**
- Literature-Based Gene Prioritization
- Semantic Wiki
Biomedical Informatics Tools

- Data Warehouse
- Automated Literature Analysis
- Case-Based Reasoning
- Literature-Based Gene Prioritization
- Semantic Wiki
Biomedical Informatics Tools

- Data Warehouse
- Automated Literature Analysis
- Case-Based Reasoning
- Literature-Based Gene Prioritization
- Semantic Wiki
Semantic Wiki

- Track project activities
- Share ideas
- Share data
- Exchange information between investigators
- Manage scientific research products

ORGANIZATIONAL ASPECTS

+ SCIENTIFIC KNOWLEDGE
Semantic Wiki

- Free web-based wiki software
- Wikimedia Foundation / Wikipedia
- Extensibility

- MediaWiki extension
- Semantic data
- Semantic search
- Data export (e.g. RDF)
Semantic Wiki

- Open-source framework for NLP
- Libraries of Text Mining tools
- API’s for tools development

RelFinder
- Querying tool
- Graphical relation browser

Entrez Utilities Web Service
- Pubmed access
- Web service + APIs
- SOAP protocol
Semantic MediaWiki

Building blocks:

- **Categories** ➔ data model in the Wiki
- **Templates** ➔ define content of Categories
- **Forms**
Organizational Aspects

Categories:
- Person
- Organization
- Meeting
- Work Package

Semantics Wiki pages:
- Person
- Organization
- Meeting
- Work Package

RDF triplestore:
- has leader
- is organized by
Organizational Aspects

Queries:
- Built-in tool (inline queries)
- RDF export ➔ SPARQL
- RelFinder
Organizational Aspects

Queries:

- **Built-in tool (inline queries)**
- RDF export ➔ SPARQL
- RelFinder

Semantic search

Query

```plaintext
[[Category:Person]]
```

[Add sorting condition]

Format as: [Table]
Organizational Aspects

Queries:
- Built-in tool (inline queries)
- RDF export $\rightarrow$ SPARQL
- RelFinder

Export pages to RDF

This page allows you to obtain data from a page in RDF format. To export pages, enter the titles.
Organizational Aspects

Queries:
- Built-in tool (inline queries)
- RDF export ➔ SPARQL
- RelFinder
Organizational Aspects

Queries:
- Built-in tool (inline queries)
- RDF export → SPARQL
- RelFinder
## Organizational Aspects

### Summary Page

<table>
<thead>
<tr>
<th>Organization</th>
<th>Partner number</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRCCS Policlinico San Matteo</td>
<td>1</td>
</tr>
<tr>
<td>Heart Hospital University College London</td>
<td>2</td>
</tr>
<tr>
<td>University Hospital of Umeå</td>
<td>3</td>
</tr>
<tr>
<td>University of Heidelberg</td>
<td>4</td>
</tr>
<tr>
<td>Hôpital Pitié Salpêtrière</td>
<td>5</td>
</tr>
<tr>
<td>Health in Code Hospital Marítimo de Oza As Xubias</td>
<td>6</td>
</tr>
<tr>
<td>Sanger Building Biochemistry Department</td>
<td>7</td>
</tr>
<tr>
<td>Skjøby University Hospital</td>
<td>8</td>
</tr>
<tr>
<td>Academisch Medisch Centrum</td>
<td>9</td>
</tr>
<tr>
<td>Department of Computer Science and Systems, University of Pavia</td>
<td>10</td>
</tr>
<tr>
<td>Department of Biomolecular and Biotechnology Science Milan</td>
<td>11</td>
</tr>
</tbody>
</table>
Organizational Aspects

➢ Summary Page

Work Package

- WP 1
- WP 6
- WP 9
- WP 10
- WP 11
- WP 12
- WP 2
- WP 3
- WP 4
- WP 5
- WP 7
- WP 8

Timeline:

2005 2006 2007 2008 2009 2010 2011
Organizational Aspects

➢ Summary Page

Meeting

- Kick-off meeting
- 2nd meeting
- 3rd meeting
- 4th meeting
- 5th meeting
- 6th meeting
- 7th meeting
- 8th meeting
- 9th meeting

Timeline © SNIIE
Categories:
- Gene
- Protein
- Dilated Cardiomyopathy Document

Scientific Knowledge

Categories:
- Gene
- Protein
- Dilated Cardiomyopathy Document

Diagram:
- Documents
- Semantic Wiki pages
- RDF triplestore
- NLP
  - concepts
- documents
- categories

Graph:
- Protein
- Gene
- DCM Document
- has protein
- has gene
Natural Language Processing

- GATE
- accessed via servlet
- .txt, .rtf, MS Word
- API plugins + purposely developed plugins
- GeneExtractor (NCBI Gene)
- ProteinExtractor (Uniprot / Swiss-Prot)
Natural Language Processing

Documents
- Sectionizer
- Tokenizer
- POS Tagger
- Chunker

NCBI Gene
GeneExtractor

UNIPROT SwissProt
ProteinExtractor

Annotated Documents

Semantic WIKI
Natural Language Processing

Select a file

File uploaded, wait for annotations, this could take some seconds
Post-natal myogenic and adipogenic developmental Defects and metabolic impairment upon loss of A-type lamins

A-type lamins are a major component of the nuclear lamina. Mutations in the $\text{LMNA}$ gene, which encodes the A-type lamins A and C, cause a set of phenotypically diverse diseases collectively called laminopathies. While adult $\text{LMNA}$ null mice show various symptoms typically associated with laminopathies, the effect of loss of lamin A/C on early postnatal development is poorly understood. Here we developed a novel $\text{LMNA}$ null mouse (LMNAGT/-) based on genetrap technology and analyzed its early post-natal development. We detect $\text{LMNA}$ transcripts in heart, the outflow tract, dorsal aorta, liver and somites during early embryonic development. Loss of A-type lamins results in severe growth retardation and developmental defects of the heart, including impaired myocyte hypertrophy, skeletal muscle hypertrophy, decreased amounts of subcutaneous adipose tissue and impaired ex vivo adipogenic differentiation. These defects cause death at 2 to 3 weeks post partum associated with muscle weakness and metabolic complications, but without the occurrence of dilated cardiomyopathy or an obvious progeroid phenotype. Our results indicate that defective early postnatal development critically contributes to the disease phenotypes in adult laminopathies.
Post-natal myogenic and adipogenic developmental Defects and metabolic impairment upon loss of A-type lamins

A-type lamins are a major component of the nuclear lamina. Mutations in the [[LMNA]] gene, which encodes the A-type lamins A and C, cause a set of phenotypically diverse diseases collectively called laminopathies. While adult [[LMNA]] null mice show various symptoms typically associated with laminopathies, the effect of loss of lamin A/C on early postnatal development is poorly understood. Here we developed a novel [[LMNA]] null mouse (LMNAGT/-) based on genetrap technology and analyzed its early post-natal development. We detect [[LMNA]] transcripts in heart, the outflow tract, dorsal aorta, liver and somites during early embryonic development. Loss of A-type lamins results in severe growth retardation and developmental defects of the heart, including impaired myocyte hypertrophy, skeletal muscle hypotrophy, decreased amounts of subcutaneous adipose tissue and impaired ex vivo adipogenic differentiation. These defects cause death at 2 to 3 weeks post partum associated with muscle weakness and metabolic complications, but without the occurrence of dilated cardiomyopathy or an obvious progeroid phenotype. Our results indicate that defective early postnatal development critically contributes to the disease phenotypes in adult laminopathies.
Post-natal myogenic and adipogenic developmental defects and metabolic impairment upon loss of A-type lamins

A-type lamins are a major component of the nuclear lamina. Mutations in the LMNA gene, which encodes the A-type lamins A and C, cause a set of phenotypically diverse diseases collectively called laminopathies. While adult LMNA null mice show various symptoms typically associated with laminopathies, the effect of loss of lamin A/C on early postnatal development is poorly understood. Here we developed a novel LMNA null mouse (LMNAGT-/-) based on genetrap technology and analyzed its early postnatal development. We detect LMNA transcripts in heart, the outflow tract, dorsal aorta, liver and somites during early embryonic development. Loss of A-type lamins results in severe growth retardation and developmental defects of the heart, including impaired myocyte hypertrophy, skeletal muscle hypotrophy, decreased amounts of subcutaneous adipose tissue and impaired ex vivo adipogenic differentiation. These defects cause death at 2 to 3 weeks post partum associated with muscle weakness and metabolic complications, but without the occurrence of dilated cardiomyopathy or an obvious progeroid phenotype. Our results indicate that defective early postnatal development critically contributes to the disease phenotypes in adult laminopathies.
### Art2-abs

| Has Author(s) | Angelo Nuzzo  
| Has Gene(s)   | LMNA, LMNA, LMNA, LMNA  
| Has Protein(s)  |

Post-natal myogenic and adipogenic A-type lamins are a major component collectively called laminopathies. While poorly understood. Here we develop in heart, the outflow tract, dorsal aorta, heart, including impaired myocyte hypertrophy defects cause death at 2 to 3 weeks progeroid phenotype. Our results indicate **defects and metabolic impairment upon knockdown of the nuclear lamina. Mutations in the **LMNA** gene, which are a consequence of loss of lamin A/C on early postnatal development and impaired ex vivo adipogenic differentiation. These defects cause death at 2 to 3 weeks postnatal development**. We detect **LMNA** transcripts in heart, the outflow tract, dorsal aorta, liver and somites during early embryonic development.

**Category:** Dilated Cardiomyopathy

**Facts about Art2-abs**

| Has Gene | LMNA |

RDF feed ☰
Relevant Literature

- NCBI E-utils
- for Genes and Proteins pages
- 5 most recent articles in Pubmed
- Gene/Protein + “Dilated Cardiomyopathy” (or synonyms)
- retrieved “on the fly”
- link to Pubmed
Relevant Literature

**LMNA**

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Ncbi</th>
</tr>
</thead>
</table>

Nicola Carboni, Claudia Sardu, Eleonora Cocco, Giovanni Marrosu, Rosa C Manzi, Vincenzo Nissardi, Franco Isola, Anna Mateddu, Elisabetta Soila, Maria A Maioli, Valentina Oppo, Rachela Piras, Giancarlo Coghe, Carlo Lai, Maria G Marrosu

**Cardiac involvement in patients with lamin A/C gene mutations: a cohort observation.**
[PubMed.22806367](https://www.ncbi.nlm.nih.gov/pubmed/22806367) [WorldCat.org] [DOI] (I p)


**Rapamycin reverses elevated mTORC1 signaling in lamin A/C-deficient mice, rescues cardiac and skeletal muscle function, and extends survival.**
Sci Transl Med: 2012, 4(144):144ra103
[PubMed.22837538](https://www.ncbi.nlm.nih.gov/pubmed/22837538) [WorldCat.org] [DOI] (I p)

Jason C Choi, Antoine Muchir, Wei Wu, Shinichi Iwata, Shinichi Homma, John P Morrow, Howard J Worman

**Temsirolimus activates autophagy and ameliorates cardiomyopathy caused by lamin A/C gene mutation.**
[PubMed.22837537](https://www.ncbi.nlm.nih.gov/pubmed/22837537) [WorldCat.org] [DOI] (I p)

Antoine Muchir, Wei Wu, Jason C Choi, Shinichi Iwata, John Morrow, Shinichi Homma, Howard J Worman

**Abnormal p38α mitogen-activated protein kinase signaling in dilated cardiomyopathy caused by lamin A/C gene mutation.**
[PubMed.22773734](https://www.ncbi.nlm.nih.gov/pubmed/22773734) [WorldCat.org] [DOI] (I p)
Cardiac involvement in patients with lamin A/C gene mutations: a cohort observation.


Neuromuscular Unit, Multiple Sciences Center, Department of Cardiovascular and Neurological Sciences, University of Cagliari, Cagliari, Italy. nikola.carboni@iscali.it

Abstract

INTRODUCTION: LMNA gene mutations are associated with cardiac and skeletal muscle alterations.

METHODS: A cohort of 21 mutated individuals was assessed with clinical and instrumental investigations over the years.

RESULTS: The median observation period was 8 years. Cardiac compromise was detected in 16 patients. Bradycardia/rhythms were the most frequent manifestations, followed by supraventricular arrhythmias. Two individuals suffered from nonsustained and 1 from sustained ventricular tachyarrhythmias. Dilated cardiomyopathy was detected in 3 patients. Evaluation of the frequencies of the clinical expressions showed a high probability of suffering from analogue heart compromise in study subjects bearing the same LMNA gene mutation.

CONCLUSIONS: Cardiac involvement represents a very common phenotypic expression of LMNA gene mutation. Subjects sharing common genetic background seem to suffer from analogue pattern of cardiac manifestation.
In conclusion...

- Collaborative Wiki System + Semantic features
- Organizational + Scientific data management
- NLP
- Literature retrieval
- Different query strategies
Future Developments

- Improve scientific knowledge management
  - New Text Mining pipelines ➔ New concepts
  - Link to new databases
- Evaluate usage of INHERITANCE partners
- Integration with other systems
A semantic collaborative system for the management of translational research projects

THANKS FOR YOUR ATTENTION!