# Deliverable D7.1

<table>
<thead>
<tr>
<th><strong>Project Title:</strong></th>
<th>Building data bridges between biological and medical infrastructures in Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Acronym:</strong></td>
<td>BioMedBridges</td>
</tr>
<tr>
<td><strong>Grant agreement no.:</strong></td>
<td>284209</td>
</tr>
<tr>
<td><strong>Research Infrastructures, FP7 Capacities Specific Programme; [INFRA-2011-2.3.2.] &quot;Implementation of common solutions for a cluster of ESFRI infrastructures in the field of &quot;Life sciences&quot;</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Deliverable title:</strong></td>
<td>Human-mouse ontology mapping workshop</td>
</tr>
<tr>
<td><strong>WP No.</strong></td>
<td>7</td>
</tr>
<tr>
<td><strong>Lead Beneficiary:</strong></td>
<td>11: HMGU</td>
</tr>
<tr>
<td><strong>WP Title</strong></td>
<td>Technical integration</td>
</tr>
<tr>
<td><strong>Contractual delivery date:</strong></td>
<td>31 December 2013</td>
</tr>
<tr>
<td><strong>Actual delivery date:</strong></td>
<td>20 December 2013</td>
</tr>
<tr>
<td><strong>WP leader:</strong></td>
<td>Michael Raess</td>
</tr>
<tr>
<td><strong>Partner(s) contributing to this deliverable:</strong></td>
<td>1: EMBL, 11: HMGU, 20: CIRMMP</td>
</tr>
</tbody>
</table>

*Authors: Philipp Gormanns, Nathalie Conte, Helen Parkinson, Michael Raess*
Contents

1 Executive summary ........................................................................................................... 3
2 Project objectives ............................................................................................................... 3
3 Detailed report on the deliverable ....................................................................................... 4
  3.1 Background .................................................................................................................. 4
     3.1.1 BioMedBridges WP7 - Use case: PhenoBridge ......................................................... 4
     3.1.2 Objectives of the Ontology Mapping Workshop ...................................................... 5
     3.1.3 Planning activities .................................................................................................. 6
3 Outcomes of the workshop ................................................................................................ 7
  3.2.1 Collection of PhenoBridge related scientific questions and use case development ............ 7
  3.2.2 Towards an improved diabetes and obesity ontology: Ontologies and terminologies .......... 8
3 Future work ....................................................................................................................... 11
  3.3.1 Ontology improvements ............................................................................................ 11
  3.3.2 Genome bridging ...................................................................................................... 11
4 Delivery and schedule ....................................................................................................... 12
5 Adjustments made ............................................................................................................. 12
8 Background information .................................................................................................... 12
9 Appendices ....................................................................................................................... 14
  Appendix 1: Workshop attendees ...................................................................................... 14
  Appendix 2: Workshop briefing materials ......................................................................... 15
     9.3.1 Outline .................................................................................................................. 15
     9.3.2 Factsheet ............................................................................................................... 17
  Appendix 3: Workshop agenda .......................................................................................... 18
  Appendix 4: Workshop summary ....................................................................................... 25
  Appendix 5: Clinical term mapping .................................................................................... 28

Supplemental material (online)

1. Manual Phenotypic terms mapping between Human and Mouse in Diabetes and Obesity fields
2. Computational Phenotypic terms mapping between Human and Mouse in Diabetes and Obesity fields
1 Executive summary

This deliverable report describes the planning, delivery and outcomes of a workshop in which blockers and solutions to human and mouse semantic data integration in the domain of diabetes and obesity were investigated. Domain experts in ontology (or terminology) creation and use, and experts in diabetes research in human and mouse were invited to address semantic data integration challenges and to design an improved ontology to support the activities of WP7. Preparatory work reviewing relevant domain ontologies and their uses in annotation and computational and available datasets allowed the workshop participants to focus on the identification of the blockers to translation between human and mouse, and to design an improved ontological solution. Improvements have been provided to the community resources in question and future work will involve further expanding and refining the ontologies, their use in computation and publication of this work. All partners of WP7 were invited to the workshop and WP3 and 4 were also represented.

2 Project objectives

With this deliverable, the project has reached or the deliverable has contributed to the following objectives:

<table>
<thead>
<tr>
<th>No.</th>
<th>Objective</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identify and develop a set of annotations, necessary terminologies, and mappings between terminologies for human and mouse models of diabetes and obesity</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Identify and group related interacting parameters in human and mouse which are involved in the development of clinical and molecular phenotypes</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Formalise rules for phenotypic annotation in human and mouse to work towards automation of phenotypic discovery and develop a related prototype service</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>
3 Detailed report on the deliverable

3.1 Background

3.1.1 BioMedBridges WP7 - Use case: PhenoBridge

WP7 PhenoBridge aims to tackle the challenge of connecting the different ontological phenotypic annotations of mouse and human. Bridging of this gap is required for a whole new interspecies analysis of disease datasets, which will lead towards novel disease candidates, pathways and therapies and support translational research. As a pilot project, PhenoBridge focuses on diabetes and obesity phenotypic data but is relevant and extensible for other areas, for example rare disease, which is known to inform our understanding of common disease and for which mouse models provide a rich phenotypic dataset and a functional tool.

Diabetes and obesity are complex diseases that represent a major international public health threat. To date, the genetic factors that influence their susceptibility haven’t been identified. A better understanding of the genetic factors underlying these complex conditions would greatly help the diagnosis, treatment, and prevention. Extensive GWAS studies have identified loci and phenotypes used as diagnostic and prognostic indicators, and more recently metabolomic studies are being used to explore profiles e.g. in urine of obese and non-obese patients and similar technology is used in mice in the form of metabolic cages. The importance of murine genetics has been clearly shown in the field of energy balance and obesity. In recent years, a large number of new genetically modified animal models including transgenic, generalized and/or tissue-specific knockout mice have been engineered for the study of diabetes and obesity.

Currently, data integration between mouse model and human studies is hindered by fundamental differences in the terminologies used by each respective community to describe the same phenotypes. PhenoBridge aims to bridge this gap in order to open the systematic use of extensive mouse phenotype data resources by researchers. In PhenoBridge we leverage the existing public domain and partner datasets, the well understood orthology
between human and mouse, and semantic tools with which to align data and ontologies, alignment of ontologies e.g. Mouse Phenotype (MP) and Human Phenotype (HPO). This deliverable describes the PhenoBridge workshop, which explored the use cases from this community and the semantic requirements for these. The workshop invitees (see Appendices 7.1) were selected to include ontology experts from human and mouse, clinical and mouse researchers and preparatory work was performed to assess the utility of the available ontologies in the domain for the human translational research community. The workshop had representation from WP7 and also from WP3 and 4, which are delivering technical solutions such as GUIs, APIs that form the BioMedBridges toolkit (WP4) and using semantic and content standards from WP3. Critically the workshop was designed to engage the wider community and international experts and projects delivering ontologies and data critical to WP7 were invited.

3.1.2 Objectives of the Ontology Mapping Workshop

To fulfil the objectives of the PhenoBridge use case, several questions needed to be addressed:

— The identification of key questions and current opinions in the research fields of diabetes and obesity. This input is required for the formalization of use cases essential for a successful PhenoBridge.

— The identification of the possibilities and constraints present in current mouse/human data integration approaches.

— The identification of key resources, key datasets, and key methodological approaches (in particular in the area of ontologies) to form PhenoBridges and allow interoperability of phenotypic data of mouse and human in the fields of diabetes and obesity.

The Ontology Mapping Workshop described in this report was organised to receive input on these questions by leading experts in the fields of diabetes and obesity as well as ontology development and bioinformatics, and together identify innovative approaches for building PhenoBridges.
3.1.3 Planning activities

The Ontology Mapping Workshop was planned and organized during four telephone conferences (24 July 2012, 24 October 2012, 16 January 2013, 18 February 2013) and two dedicated work package meetings that took place at Munich Airport on 16 November 2012 and at the BioMedBridges Annual General Meeting on 11 March 2013 in Düsseldorf. Members of the Scientific Advisory Board of BioMedBridges, who joined the WP7 Meeting during the AGM, provided important input.

To leave as much room as possible for practical exercises and discussions participants were provided with briefing materials (see Appendices 7.2) and to focus the thematic presentations on the workshop goals guiding questions were provided (also beforehand). This enabled us to address the problems of data and semantic integration during the workshop.

Guiding questions for the diabetes and obesity experts:

— What is the (or could be the potential) value of mouse models in the research on diabetes and obesity?

— Which research questions in your field would profit from being informed by both, mouse and human data?

— Have you already worked on a research project that combines mouse and human data? What was the benefit of this approach, where were the practical pitfalls and shortcomings?

Guiding questions for the ontology and bioinformatics experts:

— Describe approaches from your field that would be useful for the PhenoBridge task of mapping mouse and human phenotypes?

— Where are the major roadblocks and pitfalls? What were the limitations of previous attempts to bridge mouse and human data?

— Which tools are already available, which ones should be built?

The general focus of the workshop lay on the practical exercises to map mouse and human ontology terms for obesity and diabetes, to identify key resources and databases, to identify the existing and potential assays with relevance for diabetes and obesity research in the large-scale global mouse
phenotyping initiatives, and to assess the representation of the disease progression in the existing disease ontologies.

3.2 Outcomes of the workshop

3.2.1 Collection of PhenoBridge related scientific questions and use case development

Before the workshop we briefed invited experts to provide insights into their research field and to think about open research questions related to our Use Case PhenoBridge (see guiding questions, section 3.1.3). Specifically:

— Which genes within human GWAS loci account for the risk of susceptibility to diabetes and obesity?

— Can PhenoBridge provide applications to support the identification of conserved genomic structural variants in mouse?

— What is the degree of conservation between mouse and human?

— Life style resistant diabetes, how could it be modelled in mice?

— Do the differences between disease phenotypes in mice and humans give insights into fundamental biology?

— Are there fundamental differences in tissue biology between mice and humans?

Clearly some of these are open research questions, but they informed the workshop discussion and we focussed on those in scope for follow up at the workshop.

3.2.1.1 Use case 1: Finding mouse models and annotation genes

Mouse data is not easily accessible for the human research community. Human diseases may present sub-phenotypes in the mouse and translational researchers can’t directly work with this mouse data without going through a complex mapping process. Some disease phenotypes do not translate to the mouse, for example, secondary complications of Type II diabetes typically do not present in mouse models. To help with ontology mappings and being able to find a suitable mouse model corresponding to a specific condition, we would
need to decompose disease in phenotypic terms and then find the relevant mouse model corresponding to these phenotypic terms. In addition understanding mappings between human and mouse diagnostic and prognostic assays and their clinical context is useful, as the mouse assays are often different due to challenges of working within animal handling guidelines and the biology of the mouse. For example the Intra-peritoneal glucose tolerance test (IPGTT) in mouse which assays the rate at which glucose is cleared from the blood is analogous to the clinical 'oral glucose tolerance test' but sampling times, values etc. are not easily interpretable by non-expert mouse biologists.

### 3.2.1.2 Use case 1: Bridging genomes

In both Mouse and Human Diabetes and Obesity research community, the systematic mapping between syntenic regions allows discovery of functional conservation and would help to validate/prioritize candidate disease genes/region. We will therefore in WP7 explore mapping approaches using existing resources such as Ensembl application programme interface\(^1\) to access syntenic mapping, variation and phenotypes in both species and also integrating International Laboratory mouse resource databases of MGI\(^2\) and IMPC\(^3\) providing integrated genetic and functional/phenotypic data.

This tool would allow retrieving syntenic regions, associated gene, variation, regulatory and phenotypic annotations in both species to help identification of new disease causing genes.

### 3.2.2 Towards an improved diabetes and obesity ontology:

**Ontologies and terminologies**

The human disease ontology (DO), Human Phenotype Ontology OMIM, Experimental Factor Ontology (which imports the Orphanet Rare Genetic Disease Classification) were manually reviewed in terms of their Type 2 diabetes and Type 1 diabetes representation. It became apparent that diabetes and its subphenotypes and related assays are only poorly described in open ontologies on the human side. Both the rare and the common forms of

---

1. [http://www.ensembl.org/info/docs/api/index.html](http://www.ensembl.org/info/docs/api/index.html)
the disease were included, as the researchers represented at the meeting were interested in how rare disease informs common disease (Blair et al.⁴) and this is an area growing in importance. The outcomes of the ontology review performed during the workshop and in preparation are summarised below:

- Lack of a merged rare and common disease ontology was highlighted as a challenge for the translational clinical community, and as a blocker for querying diseases between human and mouse in query interfaces
- Phenotypic terms describing Type 2 diabetes such as age of onset were absent
- Progression information on disease in the human and mouse was absent
- Terms used to describe assays were different in human and mouse and while the MP ontology contains these they are absent from the HP for reasons of scope. These were deemed useful to map and the IMPC protocols for Clinical Blood Chemistry⁵, and IPGTT⁶ were explored as a useful start point for these
- A set of common free text queries from clinical users (acquired from colleagues in Metabolic Medicine) were manually mapped to MP terms and while 1:1 mappings were found often the MP term was not used by mouse annotators as it was too granular, or not granular enough, so numbers of annotations do not follow the mappings of the ontology (see Appendix 7.5)

The primary workshop activity was to produce a merged component of the human disease ontology DO, human phenotype HP and Orphanet rare disease classification. During the merging and extension process several terms were added to the ontology and defined by the experts in the workshop. A follow up activity was identified based on previous text mining of the Diabetes literature, which generated a list of diabetes related terms comprising etiology terms, secondary complications, diagnostic terms and additional phenotype terms. These were provided to a clinical pathologist with experience of human and mouse data for review and categorisation, and will

⁵ https://www.mousephenotype.org/impress/protocol/151/
⁶ https://www.mousephenotype.org/impress/protocol/87/
be suggested as new terms for the disease ontology and/or human phenotype ontology once reviewed. Figure 1 shows a graphical representation of the modified ontology that presents a pattern for integration of the HPO and DO ontologies. Table 1 describes categories and assays that are relevant for diabetes.

**Figure 1** The revised DO/HPO merge with new terms produced during the workshop

**Table 1** Diabetes categories and assays described by the experts

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Prediabetic</th>
<th>Diabetic</th>
<th>Late consequences</th>
<th>Very late consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms: frequent</td>
<td>Intermittent high blood glucose levels</td>
<td>Constant high blood glucose levels</td>
<td>Long duration elevated blood glucose</td>
<td>Long duration elevated blood glucose</td>
</tr>
<tr>
<td>Symptoms: Less frequent</td>
<td>Headache</td>
<td>Polydipsy (thirsty)</td>
<td>Bacterial infections (bladder, kidneys)</td>
<td>Microangiopathy: Ulcer of feet and legs, retinopathy, kidney failure</td>
</tr>
<tr>
<td>Symptoms: rare</td>
<td>Polydipsy (thirst)</td>
<td>Polyphagia, ketoacidosis (T1D)</td>
<td>Proteinuria (starting impairment of kidney function)</td>
<td>Progressive polyneuropathy (autonomic nervous system)</td>
</tr>
<tr>
<td>Symptoms: rare</td>
<td>Sleepiness (somnolence)</td>
<td>Microangiopathy (ulcer of legs)</td>
<td>Polyneuropathy (impaired nerve function)</td>
<td>Systemic bacteriaemia</td>
</tr>
<tr>
<td>Symptoms: rare</td>
<td>Bacterial infections</td>
<td></td>
<td></td>
<td>Stroke</td>
</tr>
</tbody>
</table>
### 3.3 Future work

Future work is split into two themes, phenotype and genotype related. Both will be required to deliver functional tools later in this work package.

#### 3.3.1 Ontology improvements

1. Summarize and compile changes that were made to the ontologies during the workshop.

2. Categorize disease terms from textmining approach.

3. Incorporation of disease-phenotype associations into Phenotype commons.

4. Incorporation of identified phenotypes into HPO.

5. Compare Phenodigm performance with both ontologies (old and new).

6. Align and map human and mouse assays relevant in this domain e.g. IMPC, GWAS, Ensembl Phenotypes, MPD and MP.

#### 3.3.2 Genome bridging

1. Mapping syntenic regions with chromosomal coordinates using Ensembl API.

2. Generate output format containing species alignment, conservation, variate information and their consequences.
4 Delivery and schedule

The delivery is delayed: ☐ Yes ☑ No

5 Adjustments made

No adjustments were made.

8 Background information

This deliverable relates to WP 7; background information on this WP as originally indicated in the description of work (DoW) is included below.

WP 7 Title: PhenoBridge-crossing the species bridge between mouse and human

Lead: Michael Raess (HMGU)
Participants: EMBL, HMGU, MUG, CIRMMP

This demonstrator project tackles a major challenge related to the available mouse phenotype and human clinical data: different ontological phenotype descriptions hinder researchers from both sides to cross the species bridge between mouse models and human. We will make use of comparable diabetes and obesity-related large-scale datasets in mouse and human provided by Infrafrontier, BBMRI and ELIXIR.

To achieve integration at the level of phenotypes in these species interaction with the wider community is required, several resources are currently in use by different resources and we require expert input to describe the phenotypes, but also to formalize the phenotypic descriptions. In order to make maximum use of existing terminologies (mouse phenotype, human phenotype ontology etc) we will work with these communities to map between existing terms, provide new terms and also to annotate our datasets.

<table>
<thead>
<tr>
<th>Work package number</th>
<th>WP7</th>
<th>Start date or starting event:</th>
<th>month 13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work package title</td>
<td>PhenoBridge-crossing the species bridge between mouse and human</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity Type</td>
<td>RTD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant number</td>
<td>1:EMBL 20:CRIMMP</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11:HMGU</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12:MUG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person-months per participant</td>
<td>21</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

### Objectives

1. Identify and develop a set of annotations, necessary terminologies, and mappings between terminologies for human and mouse models of diabetes and obesity
2. Identify and group related interacting parameters in human and mouse which are involved in the development of clinical and molecular phenotypes
3. Formalise rules for phenotypic annotation in human and mouse to work towards automation of phenotypic discovery and develop a related prototype service.

### Description of work and role of participants

**Task 1:**
Analysis of existing mouse phenotype and human disease ontology terms, leading to submission of proposals for new terms to gather a comprehensive set of terms to describe diabetes and obesity phenotypes in mouse and human. As a first step, potentially relevant phenotypic parameters available from mouse high-throughput screening and in-depth phenotyping studies as well as from human studies across technologies such as gene expression and GWAS studies from the BioSD and metabolome profiles will be listed. For analysed parameters associated with diabetic/obese conditions that are not yet described appropriately in ontology terms, new terms have to be defined. A many to many mapping of phenotypic or diagnostic parameters onto ontology terms will be developed using the clinical expertise of scientists coming from the mouse or human diabetes and obesity fields. (EMBL-EBI, HMGU, MUG, CIRMMP).

**Task 2:**
Mapping of mouse and human phenotypes of diabetes and obesity. Based on the ontology terms developed in Task 1, observation patterns will be defined that describe clinical and molecular characteristics of diabetes and obesity phenotypes in mouse and human (EMBL-EBI, HMGU, MUG, CIRMMP). The identification of rules and criteria for identifying diabetes and obesity phenotypes (and how they map in mice and humans) will lead to a prototype for an automated procedure to identify phenotype matches across mouse and human (EMBL-EBI, HMGU, MUG, CIRMMP).
9 Appendices

1. Workshop attendees
2. Workshop briefing materials
3. Workshop agenda
4. Workshop summary
5. Clinical term mapping

Appendix 1: Workshop attendees

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Workshop role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elizabeth Bentley</td>
<td>MRC, Harwell, UK</td>
<td>Diabetes Expert</td>
</tr>
<tr>
<td>Nathalie Conte</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Mario Falchi</td>
<td>Imperial, London, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Philipp Gormanns</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Martin Hrabé de Angelis</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Gautier Koscielny</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Christoph Lengger</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Terry Meehan</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Chris Mungall</td>
<td>Lawrence Berkeley Lab, Berkeley, US</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Frauke Neff</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Helen Parkinson</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Michael Raess</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Damian Smedley</td>
<td>Sanger Institute, Hinxton, UK</td>
<td>Ontology expert</td>
</tr>
<tr>
<td>Cynthia Smith</td>
<td>MGI, Bar Harbor, US</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Harald Staiger</td>
<td>University Tübingen, Germany</td>
<td>Ontology expert</td>
</tr>
<tr>
<td>Leonardo Tenori</td>
<td>CERM, Florence, Italy</td>
<td>Ontology expert</td>
</tr>
<tr>
<td>Hendrik Tiedemann</td>
<td>Helmholtz, Munich, Germany</td>
<td>Diabetes Expert</td>
</tr>
</tbody>
</table>

BioMedBridges Deliverable D7.1
Appendix 2: Workshop briefing materials

9.3.1 Outline

Bridging the gap between mouse and human phenotypes in the field of diabetes and obesity

Ontology Mapping Workshop

- Date: 15-16 April 2013
- Venue: UK; probably London

Abstract

The PhenoBridge use case is part of BioMedBridges, a joint effort of ten biomedical sciences research infrastructures on the roadmap of the European Strategy Forum for Research Infrastructures (ESFRI). PhenoBridge aims to tackle the challenge of connecting the different ontological phenotypic annotations of mouse and human. Bridging this gap is mandatory for a whole new inter-species analysis of disease datasets, which will lead towards novel disease candidates, pathways and therapies. As a pilot project PhenoBridge focuses on diabetes and obesity.

The workshop aims to identify key questions and current opinions in those research fields. This input will strongly support the formalization of use cases essential for a successful PhenoBridge. Moreover, an investigation of barriers present in current mouse/human data integration approaches will be investigated. Especially, ontologies and their potential as a PhenoBridge will be of main interest as well as existing key datasets and resources.

On the first day of the workshop we will introduce the project and analyze the current situation regarding the interoperability of mouse and human phenotype data in diabetes and obesity. We will address the main problems and roadblocks and the expectations from the clinical community. On the second day solutions for most critical problems will be framed and their realization in the PhenoBridge use case will be discussed.

Outline

Day 1

- Introduction to BioMedBridges and WP7 and to resources available from the involved ESFRI projects
- Current research questions / methods & data / bioinformatics tools used in diabetes and obesity research
- Challenges, roadblocks and mouse model expectations in the clinical community
- Approaches for ontology based human-mouse data integration
- Overview of relevant and crucial resources
Day 2
- Focus on solutions
- Gap analysis
- Prioritization of approaches / tools provided by the PhenoBridge use case
9.3.2 Factsheet

**What is PhenoBridge?**

PhenoBridge aims to tackle the challenge of connecting the different ontological phenotypic annotations of mouse and human. Bridging this gap is mandatory for the whole new winter species analysis of disease datasets, which will lead towards novel disease candidates, pathways and therapies. As part of project PhenoBridge focuses on Diabetes and Obesity.

**Why Bridging human and mouse phenotypes in the fields of type 2 diabetes and obesity?**

Diabetes and Obesity are complex diseases that represent a major international public health threat. To date, the genetic factors that influence their susceptibility haven’t been identified. A better understanding of the genetic factors underlying these complex conditions could greatly help the diagnosis, treatment, and prevention. The importance of murine genetics has been clearly shown in the field of energy balance and obesity. In recent years, a large number of new genetically modified animal models, including transgenic, generalized and/or tissue-specific knockout mice have been engineered for the study of Diabetes and Obesity. Currently, data integration between mouse model and human studies is hindered by fundamental differences in the terminologies used by each respective community to describe the same phenotypes. PhenoBridge aims to bridge this gap in order to open the systematic use of extensive mouse phenotype data resources by researchers.

**Who’s in the PhenoBridge team?**

The PhenoBridge team is a group of specialists from European bioinformatics institutes (EMBL-EBI, UK), Helmholtz Zentrum München, HMGU, (Germany), the Centro de Risonanz BufferedReader (CERM, Italy) and the Medical University of Graz (MUG, Austria). We are biologists, ontologists, bioinformaticians, software engineers and data integration experts.

**What does the PhenoBridge team do?**

Check existing ontologies for mouse and human data for their coverage and granularity.

Improve/Enhance existing ontologies and develop new ones where needed.

Formalize rules for phenotypic annotation in human and mouse to work towards automation of phenotypic discovery and develop a related prototype service.

Apply these rules to a subset of mouse model and human studies to demonstrate the effectiveness of mouse model/human data integration.

**Find more information**

PhenoBridge: [http://www.biomedbridges.eu/workpackages/wp790+](http://www.biomedbridges.eu/workpackages/wp790+)

BioMedBridges: [http://www.biomedbridges.eu](http://www.biomedbridges.eu)

---

BioMedBridges Deliverable D7.1
Appendix 3: Workshop agenda

AGENDA

BioMedBridges PhenoBridge
Ontology Mapping Workshop

15-16 April 2013

Location
EBI, Hinxton
Day 1 Monday, 15 April morning session
Chair: Martin Hrabé de Angelis

Workshop start
09:30 to 10:30

Introduction of participants
Introduction of workshop goals

Diabetes and Obesity
10:00 to 11:30

Speakers: (20 minutes each)
Harald Staiger
  o Translational Diabetology: what can we learn from mice
Mario Falchi
  o Integrative genomic approaches to human disease
Liz Bentley
  o Mouse Models of Diabetes and Obesity

Guiding questions:

What is the (or could be the potential) value of mouse models in the research on diabetes and obesity?
Which research questions in your field would profit from being informed by both, mouse and human data?
Have you already worked on a research project that combines mouse and human data? What was the benefit of this approach, where were the practical pitfalls and shortcomings?
Ontology and Tools

11:30 to 13:00

Speakers: (20 minutes each)

Cynthia Smith
- The Mammalian Phenotype Ontology: a tool for unifying experimental and high-throughput screen phenotyping data

Terry Meehan (10 minutes)
- Overview on mouse resources for Diabetes and Obesity research

Chris Mungall
- Mapping Phenotype Ontologies

Damian Smedley
- PhenoDigm – animal models for humans disease through cross-species phenotype mapping

Guiding questions:

Can you describe approaches from your field that would be useful for the PhenoBridge task of mapping mouse and human phenotypes?

Where are the major roadblocks and pitfalls? What were the limitations of previous attempts to bridge mouse and human data?

Which tools are already available, which ones should be built?

General discussion: (20 minutes)

Preparation of mapping exercise and brainstorming session.

Focus question: How can we help your research?

Lunch

13:00 to 14:00
Day 1 Monday, 15 April afternoon session

Chair: Terry Meehan

Mapping together
14:00 to 16:00

Speakers: (10 minutes each)
- Helen Parkinson
  - Textmining for Type 2 Diabetes
- Nathalie Conte
  - Mapping of clinical terms with MP and HPO

Practical mapping exercise: (All participants, 90 minutes)
Bringing together ontology terms of mouse and human in diabetes and obesity.

Coffee Break
16:00 to 16:30
Development of small Phenobridge use cases

16:30 to 18:30

Brainstorming in small groups: (All participants, 60 minutes)
Small phenobridge use cases shall be developed based on the previous discussions and mapping exercise.

General discussion: (All participants, 60 minutes)
How to go from here? Representation of diabetes and obesity in the existing technologies for mice and humans?

Issues to consider:

How to represent relevant environmental factors (e.g. dietary challenges, exercise) in the data/ontologies?

How to represent the time and progression of the disease in the data/ontologies?

End of Day 1, Transport to Cambridge

19:00

Dinner in Cambridge

20:00
Day 2 Tuesday, 16 April
Chair: Helen Parkinson

Summary of first day
09:30 to 10:00

Towards an approach for building phenobridges
10:00 to 12:45

Open discussion and brainstorming (All participants)
This should be an open discussion and brainstorming on how to proceed in the PhenoBridge work package. The actual content will depend a lot on the discussions of day 1. However the following points should be investigated:

What are the scientific questions on diabetes and obesity that will profit from bridging together mouse and human data?

Moderator: Harald Staiger & Mario Falchi

Representations of diabetes and obesity in the existing ontologies for mice and humans?

a) Do we have the right terms? What should be added?
b) Do the mouse and human terms map? What should be modified?
c) Are environmental factors and disease timing and progression sufficiently captured?

Moderator: Cynthia Smith & Liz Bentley

Mapping mouse and human phenotypes in diabetes and obesity.

a) What are the most promising approaches?
b) Which tools should be (can be) built to help the researchers in the field?

Moderator: Chris Mungall & Damian Smedley
Alternate approaches: A “precision medicine” approach for classifying diabetes and obesity.

Moderator: Hendrik Tiedemann & Philipp Gormanns

Summary of workshop
12:45

Lunch and end of workshop
13:00
Appendix 4: Workshop summary

BioMedBridges Ontology Mapping Workshop, April 15–April 16, 2013, Hinxton UK

Background of the workshop

The ontology mapping workshop emerged from the BioMedBridges WP7 Use Case „PhenoBridge”, which aims to tackle the challenge of connecting the different ontological phenotypic annotations of mouse and human. Bridging this gap is mandatory for a whole new inter-species analysis of disease datasets, which will lead towards novel disease candidates, pathways and therapies. As a pilot project PhenoBridge focuses on diabetes and obesity.

Workshop participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
<th>Workshop role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elizabeth Bentley</td>
<td>MRC, Harwell, UK</td>
<td>Diabetes Expert</td>
</tr>
<tr>
<td>Nathalie Conte</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Mario Faichi</td>
<td>Imperial, London, UK</td>
<td>Diabetes Expert</td>
</tr>
<tr>
<td>Philipp Gormanns</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Martin Hrabé de Angelis</td>
<td>Helmholtz, Munich, Germany</td>
<td>Coordinator</td>
</tr>
<tr>
<td>Gautier Koscielny</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Christoph Lengerger</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Terry Meehan</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Chris Mungall</td>
<td>Lawrence Berkeley Lab, Berkeley, US</td>
<td>ontology expert</td>
</tr>
<tr>
<td>Frauke Neff</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Heien Parkinson</td>
<td>EBI, Hinxton, UK</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Michael Raess</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Damian Smedley</td>
<td>Sanger Institute, Hinxton, UK</td>
<td>ontology expert</td>
</tr>
<tr>
<td>Cynthia Smith</td>
<td>MGI, Bar Harbor, US</td>
<td>ontology expert</td>
</tr>
<tr>
<td>Harald Staiger</td>
<td>University Tubingen, Germany</td>
<td>Diabetes Expert</td>
</tr>
<tr>
<td>Leonardo Tenori</td>
<td>CERM, Florence, Italy</td>
<td>WP 7 member</td>
</tr>
<tr>
<td>Hendrik Tiedemann</td>
<td>Helmholtz, Munich, Germany</td>
<td>WP 7 member</td>
</tr>
</tbody>
</table>

Topics addressed in this workshop

- Scientific underpinning: Which data bridges are required?
- Bioinformatic underpinning: How to best build these data bridges?
- What are the scientific questions on diabetes and obesity that will profit from bridging together mouse and human data?
- How to represent relevant environmental factors (e.g. dietary challenges, exercise) in the data / ontologies?
- How represent the time and progression of the disease in the data/ontologies?
Workshop activities

- Expert talks to provide scientific uses as well as possible methodological approaches for "PhenoBridge"
- Review and extension of existing ontologies with diabetic and obesity terms provided by the experts
- Practical Ontology mapping session
- Identification of PhenoBridge Use Cases

Results

- Diabetes and Obesity related phenotype improvement in terms of frequency and progression information.
- Review and Suggestions for an Extension of the Human Phenotype Ontology
- Suggestion to the Disease Ontology and a diabetes subset of the Orphanet rare disease classification.
- Identification of diabetes relevant IMPC assays/parameters sets.
- Collection of relevant scientific questions from the different communities
  - Life style resistant diabetes, how could it be modelled in mice?
  - Which genes within human GWAS loci account for the risk?
  - What degree of conservation if conservation between mouse and human?
  - Do the differences between disease phenotypes in mice and humans give insights into fundamental biology?
  - Fundamental differences in tissue biology between mice and humans?
  - Phenobridge applications to support the identification of conserved genomic structural variants in mouse?

- Ad hoc use cases derived from the questions in the workshop
  - Finding mouse models and annotating genes.
    To help with terminology mappings and being able to find the right mouse model corresponding to a specific condition, we would need to decompose disease in phenotypic terms, then find the relevant mouse model corresponding to these phenotypic terms. In addition, adding environmental factors mappings would be extremely valuable.
  - Bridging genomes
    In both Mouse and Human Diabetes and Obesity research community, the systematic mapping between human-mouse syntenic regions would allow to uncover functional conservation and to validate/prioritize candidate disease genes/region.
    We propose to build an automatic mapping tool to retrieve syntenic regions, associated gene, variation, regulatory and phenotypic annotations in both species.
Immediate next steps

- Incorporate new phenotype terms into the modified HPO file from the workshop including text mining terms from Disease PhenolView.
- Review determined HPO changes with Peter Robinson.
- Compare PhenoDigm performance with improved ontologies.
- Structure outcomes from assay mapping session.
- Define action items/working tasks for the WP7 identified Use Cases “Finding mouse models and annotation” and “Bridging the genomes”.

List of relevant resources related to the workshop

<table>
<thead>
<tr>
<th>Resource</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE - Array Express</td>
<td><a href="http://www.ebi.ac.uk/arrayexpress/">http://www.ebi.ac.uk/arrayexpress/</a></td>
</tr>
<tr>
<td>BioMedBridges</td>
<td><a href="http://www.biomedbridges.eu/">http://www.biomedbridges.eu/</a></td>
</tr>
<tr>
<td>DO - Disease Ontology</td>
<td><a href="http://disease-ontology.org/">http://disease-ontology.org/</a></td>
</tr>
<tr>
<td>EUMODIC - European Mouse Disease Clinic (Pilot for the INFRAFRONTIER mouse clinics and the IMPC)</td>
<td><a href="http://www.eumodic.org/">http://www.eumodic.org/</a></td>
</tr>
<tr>
<td>EuroPhenome - Mouse phenotyping data repository</td>
<td><a href="http://www.europhenome.org/">http://www.europhenome.org/</a></td>
</tr>
<tr>
<td>GXA - Gene Expression Atlas</td>
<td><a href="http://www.ebi.ac.uk/gxa/">http://www.ebi.ac.uk/gxa/</a></td>
</tr>
<tr>
<td>HPO - Human Phenotype Ontology</td>
<td><a href="http://www.human-phenotype-ontology.org/">http://www.human-phenotype-ontology.org/</a></td>
</tr>
<tr>
<td>IMPC - International Mouse Phenotyping Consortium</td>
<td><a href="https://www.mousephenotype.org/">https://www.mousephenotype.org/</a></td>
</tr>
<tr>
<td>INFRAFRONTIER - European infrastructure for phenotyping and archiving of mouse models</td>
<td><a href="http://www.infraf%E5%B6%B6ient.eu/">http://www.infraf嶶ient.eu/</a></td>
</tr>
<tr>
<td>MGI - Mouse Genome Informatics</td>
<td><a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></td>
</tr>
<tr>
<td>MP - Mammalian Phenotype Ontology</td>
<td><a href="http://www.informatics.jax.org/searches/MP_form.shtml">http://www.informatics.jax.org/searches/MP_form.shtml</a></td>
</tr>
<tr>
<td>OMIM - Online Mendelian Inheritance in Man</td>
<td><a href="http://www.omim.org">http://www.omim.org</a></td>
</tr>
<tr>
<td>PhenoDigm - Phenotype comparisons for Disease and Gene Models</td>
<td><a href="http://www.sanger.ac.uk/resources/databases/phenodigm/">http://www.sanger.ac.uk/resources/databases/phenodigm/</a></td>
</tr>
</tbody>
</table>
Appendix 5: Clinical term mapping

We obtained a list of clinical terms used in diagnosis from clinical obesity specialist (Pr Sadaf Farooqi, Addenbrookes hospitai, Cambridge, UK).

- **Body composition**
  - Increased adiposity
  - Increased fat mass
  - Weight gain (standard versus high fat diet)
  - Obesity

- **Energy homeostasis**
  - Food intake, hyperphagia, anorexia, hypophagia, aphagia
  - Basal metabolic rate, oxygen consumption, physical activity, heart rate, heart rate variability, total energy expenditure

- **Behavioural phenotype**
  - Food-seeking behaviour, food preference, sweet preference

- **Metabolic**
  - Diabetes, Hyperinsulinemia, insulin resistance, hyperglycaemia, glucose intolerance, hyperlipidemia, hypercholesterolemia, dyslipidemia, steatosis, fatty liver disease

- **Hypertension, cardiomegaly**

We wanted to map these terms to Human and mouse ontologies. We used both manual and computational mappings using the OWLSim software algorithm for cross-species phenotype comparisons using MP and HPO ontologies.

The mappings are available as supplemental material here:
- Manual Phenotypic terms mapping between Human and Mouse in Diabetes and Obesity fields
- Computational Phenotypic terms mapping between Human and Mouse in Diabetes and Obesity fields

There was an overall good correspondence between terms and some challenges were identified which were discussed with the specialists during the workshop.

Challenges:
- Incomplete mappings between HPO and MP
- Different descriptions in human and mouse domain
  - i.e.: eating behaviour.
- Semantic representation of assay used to assess metabolic disorder
- Additional complexity: ageing, environmental factors and genetic variability

Following this mapping, new terms were added in MP ontology to allow better mappings between both species also some gaps in HPO were flagged and reported to HPO developers.