BioMedBridges and COSMOS

Joint statement on the need for data standardisation in metabolomics

COSMOS[1] (COordination of Standards in MetabOlomicS) is a consortium of leading European groups in metabolomics which aims at developing and improving common standards to describe, exchange and query both NMR metabolomics data and experimental metadata (e.g. source of study materials, technology and measurement types, sample-to-data relationships). This will guarantee that the NMR-metabolomics data will be stored, distributed, and managed according to well-established standards at the European level.

Specifically, COSMOS aims at ensuring that metabolomics data will be:

1. encoded in open standards to allow barrier-free and wide-spread analysis
2. tagged with a community-agreed, complete set of metadata (minimum information standard)
3. supported by a communally developed set of open source data management and capturing tools
4. disseminated in open-access databases adhering to the above standards
5. supported by vendors and publishers, who require deposition upon publication
6. properly interfaced with data in other biomedical and life-science e-infrastructures.

[1] http://www.cosmos-fp7.eu/

BioMedBridges[2] is a joint effort of twelve biomedical sciences research infrastructures on the ESFRI roadmap. Together, the project partners will develop the shared e-infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains and thus strengthen biomedical resources in Europe. The main aim of BioMedBridges is to build "bridges" between different sources and types of data.

Metabolomics is one of the latest -omics sciences and there is a growing interest for its application in prognosis, diagnosis, patient stratification and personalized medicine. For this reason, biomedical research infrastructures are strongly interested in the development of this technology. Because metabolomics is multidisciplinary in nature, it could rapidly become an "experimental bridge" between the different communities served by the research infrastructures. To achieve this, a strong requirement is that any methodology deployed can be applied in the various biomedical contexts with little or no modification. In this sense, the COSMOS project represents an extremely successful ongoing effort, and BioMedBridges is instrumental in providing knowledge on how to harmonize data and metadata produced by users of the biomedical research infrastructures.

Regarding the individual infrastructures involved in BioMedBridges, metabolomics can be considered extremely relevant to, in particular, BBMRI-ERIC, ELIXIR, Euro-BioImaging, EU-OPENSCREEN, INSTRUCT, EATRIS, ECRIN and EMBRC.

The robustness of a metabolomic study relies on the availability of large sets of high quality samples. This creates an obvious link between metabolomics and biobanks: biobanks are sources of samples and associated data; metabolomics may provide a sensitive tool to assess degradation phenomena as intactness of the metabolome is a good indicator of the quality of stored materials. In terms of analytical platforms, NMR-based metabolomics may be preferred to mass spectrometry for this type of evaluation studies because it requires only minimal if no sample handling and is highly reproducible and fast, although able to detect only the most concentrated metabolites, i.e. only a few tens of molecules. The use of NMR profiles as further data to be associated to stored biological samples to assess their quality and to evaluate the impact of pre-analytical treatments can be

[2] http://www.biomedbridges.eu/

an important added value for biobanks. Metabolomics may also provide a direct approach to monitoring the performance of different storage conditions on the molecular profiles of different types of samples. Nevertheless, systematic studies on large sets of samples are still lacking and, in particular, no precise data exist on the "shelf-life time" of samples stored in biobanks. Metabolomics may become an efficient means to fill this gap. Emphasis should be given to the fact that successful molecular analysis not only depends on the quality of the clinical data but also on the availability of the information on sample history: collection and handling times and temperatures should be carefully annotated and become part of the associated dataset. Inclusion of metabolomic profiles in the biobank databases would be a useful addition to assess sample quality and history. To take up advantage of the above opportunities for the development of metabolomics as a crucial technique to flank the development of standard operating procedures and quality control in biobanks, an Expert Center for Metabolomics[3] (EXCEMET) has been formally established. EXCEMET proposes itself as a reference infrastructure for biobanks and has been described as a model of a BBMRI-ERIC Expert Center[4] . EXCEMET has been established as a not-for-profit public-private-partnership based on a consortium agreement between participants from academia and industry. The involvement of the Medical University of Graz, with a research unit lead by Professor Kurt Zatloukal, who coordinated the preparatory phase of the European biobanking and biomolecular research infrastructure (BBMRI) during the 7th EU framework programme, testifies the close links that have been established between the metabolomics and the biobanking community. Kurt Zatloukal is also member of the Advisory Board of COSMOS.

BioMedBridges is providing a framework for the harmonization of metadata. This will involve ELIXIR (the European Life-sciences Infrastructure for biological Information) as the main submission hub for public release of data generated at the research infrastructures. ELIXIR is an infrastructure that will allow life science laboratories across Europe to share and store their research data as part of an

[3] http://www.excemet.org/

[4] Gert-Jan B van Ommen et al. (2014) BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based Expert Centres. European Journal of Human Genetics doi:10.1038/ejhg.2014.235

organised network. Its goal is to bring together Europe's laboratories and data centres to help coordinate the collection, quality control and storage of large amounts of biological data produced by life science experiments. ELIXIR aims to ensure that biological data is integrated into a coordinated system in which all parts of the scientific community can access existing research easily. MetaboLights[5], the only public repository for metabolomics data in Europe, is located at the European Bioinformatics Institute (EMBL-EBI) which is linked to ELIXIR as one of its Nodes.

Beside biological fluids, metabolomic profiles can be acquired on tumour biopsies. Metabolomics on biopsies, in general, allows for a more detailed fingerprinting of the tumour metabolism, while the systemic biofluids allow for the study of the tumour–host interactions. An integration of this information with in vivo imaging data, especially NMR-based metabolomics with NMR-based imaging, could improve the translation of basic science to the clinical practice. These two worlds are going to communicate even more in the future, so there is a need for adopting common standards and ontologies in order to make the analysis of metabolomics data acquired with either technique faster and simpler and to enhance the interoperability of both platforms. Euro-BioImaging is a pan-European infrastructure whose mission is to build a distributed imaging infrastructure across Europe that will provide open access to innovative biological and medical imaging technologies for European researchers and it could play a crucial role in the interoperability of imaging and metabolomics.

Toxicology is another expanding field of research for metabolomics. Profiles of urine and serum of patients that have been treated with drugs in clinical trials may contribute to the prediction of efficacy and/or toxicity of the treatment, and provide hints about the underlying biochemical mechanisms. EU-OPENSCREEN would thus highly benefit from the development of consistent procedures to collect, store, interpret and report metabolomic data. Because of its conceptual and technical overlap with many aspects of pharmaceutical research, metabolomics is finding applications that span almost the full length of the drug discovery and development pipeline. Metabolomics can be used to facilitate lead compound

[5] http://www.ebi.ac.uk/metabolights/

discovery, to improve biomarker identification (for monitoring disease status and drug efficacy), to monitor drug metabolism and toxicity, to facilitate clinical trial testing and to improve post-approval drug monitoring. At the two extreme points of the development pipeline, metabolomics could complement both preclinical studies, monitoring the systemic effects of the drug candidates on treated mouse models, and phase 4 postmarketing surveillance, helping to clarify the molecular mechanisms of adverse effects onset. Metabolomics potentially offers drug researchers and drug regulators an effective, inexpensive route to addressing many of the riskier or more expensive issues associated with the discovery, development and monitoring of drugs.

Metabolomics on intact cells or on cell extracts can help structural biologists to understand how structural changes in the three-dimensional organization of a protein at the atomic level dynamically affect its function, and how this function is reflected on variations in the metabolism of the cell. In this respect, INSTRUCT will benefit from new methods to characterize the cellular metabolome in a time-dependent manner and as a function of external stimuli. Parallel studies of protein structures in cells and cell metabolism may open new exciting perspectives for the mechanistic systems biology approach: the atomic view can be complemented by metabolomics as a tool to characterize and quantify small molecules in a biological sample, and the two views will help to connect molecular events at the cellular level to those occurring at the systemic level.

Personalized medicine is a topic of interest to various infrastructures, especially for EATRIS and ECRIN, in particular regarding the effects of dietary intake on the individual metabolome. There are sound indications that the metabolomic assessment of controlled dietary interventions may result in better evaluation of a research subject’s diet than traditional observational data. These aspects can be readily extended to the analysis of patients’ response to pharmaceutical and/or surgical treatments. This warrants the deployment of metabolomics as a tool to flank clinical trials and translational medicine efforts to obtain deeper insight as well as better predictive power for the outcome of treatments. Notably, this is a different area of application of metabolomics with respect to its relatively common involvement in the discovery/validation of biomarkers.

Closely related to the above concepts is the application of metabolomics within the context of translational studies on mouse models of diseases, again ranging from the identification and validation of biomarkers to the analysis of the response to diet and pharmacological treatment. Within this frame, some related applications may become interesting also for INFRAFRONTIER and MIRRI.

The genetic, taxonomic and ecosystem diversities in the marine environment far outweigh those of the terrestrial environment and as access to marine biological resources improves, marine biodiversity is increasingly becoming a focus for diverse fields of fundamental and applied research. The EMBRC infrastructure promotes the use of state-of-the-art metabolomic techniques in the study of emerging-model and non-model marine organisms in order to develop a more comprehensive understanding of the novel biology of these organisms and impacts of environmental change on this diversity. Marine metabolomics is also of significant interest for applications in marine bioprospecting and marine biotechnology. In this context, EMBRC fully supports the COSMOS initiative for development of common standards to collect, manage, and interpret metabolomic data.

Systems biology, defined as the computational and mathematical modeling of complex biological systems requires significantly more multidisciplinary facilities than are typically present in any single institute. For this reason ISBE (Infrastructure for Systems Biology–Europe), was created as a large-scale

European research infrastructure designed to tackle the grand challenge of developing a systematic understanding of complex biological processes in living organisms. ISBE recognizes that metabolomics is a key technology for systems biology: metabolites are system variables that can be used to model the dynamics of biological systems.

In light of what is stated in the present document, and in order to make more effective use of metabolomics in biomedical research, BioMedBridges, on behalf of the research infrastructures involved in the project, expresses strong interest in metabolomics and will help ease its adoption in the biomedical community. COSMOS and BioMedBridges agree on the emerging role of this technology in

biomedical research and stress the need for further efforts in developing common standards and procedures to harmonize both experimental procedures and the data and metadata produced in metabolomics.