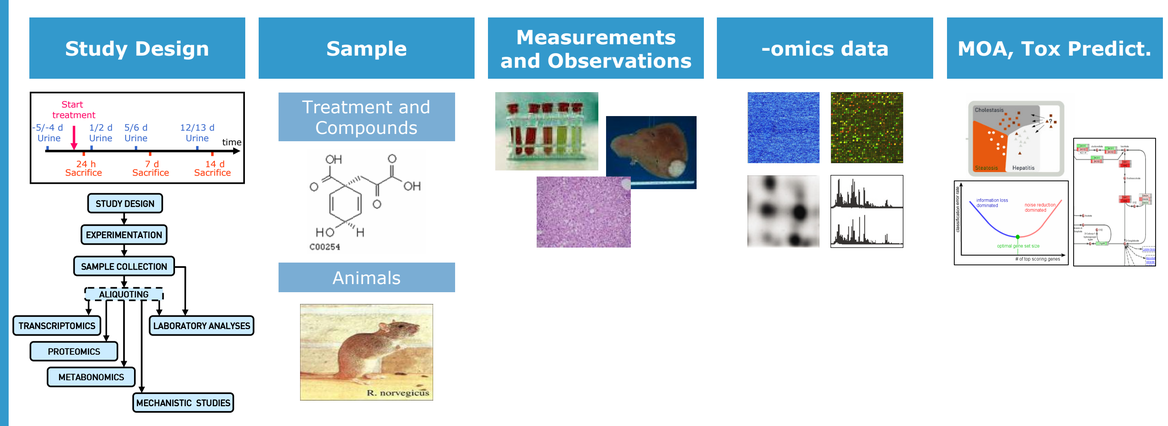
Toxicogenomics and Biomarker Discovery for Prediction of Long Term Toxicity

Hans Gmuender, Maria Wendt, Tobe Freeman and Jochen König*
Genedata AG, Maulbeerstrasse 46, CH-4052 Basel, Switzerland, *Jochen.Koenig@genedata.com

Toxicogenomics can be defined as a genome-wide approach to toxicology. The approach exploits functional genomics information, including gene, protein and metabolite data, to reveal the molecular basis of toxic reactions. Conventional toxicology data, for example histology, is used to guide a quantitative and unbiased search for biomarkers and to reveal the molecular mechanisms underlying toxicity. We present a case study of Genedata's toxicogenomics collaboration with the European InnoMed-PredTox consortium, a joint industry and academic initiative to promote development and improve drug safety.

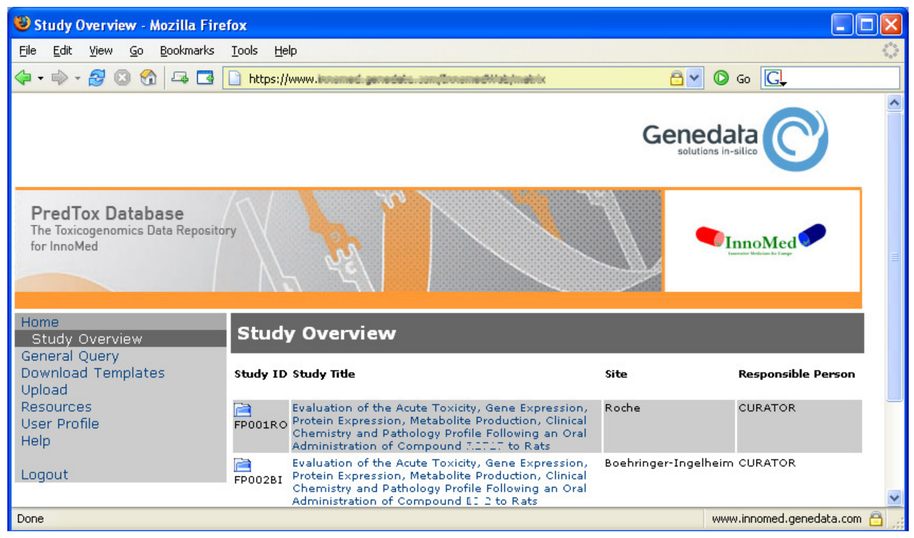
A | Systematic Approach

The InnoMed consortium unites 14 pharmaceutical companies, 3 universities and bioinformatics solution provider Genedata. From the onset of the three-year project a considerable investment was made into thorough design of the overall study. InnoMed-PredTox is the first initiative to systematically and comprehensively assess the response of a typical model organism on transcript, protein and metabolite level in the context of traditional toxicological endpoints (e.g.: histopathology, serum chemistry).



Data are generated and gathered according to standardized protocols across all sites and fed into a joint, relational database, developed by Genedata.

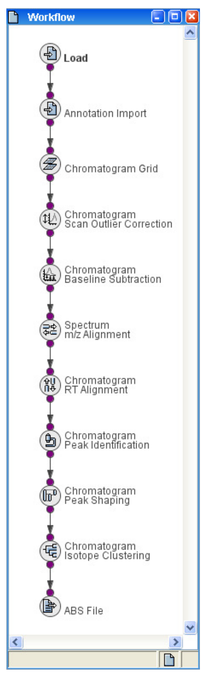
C | Data Organization

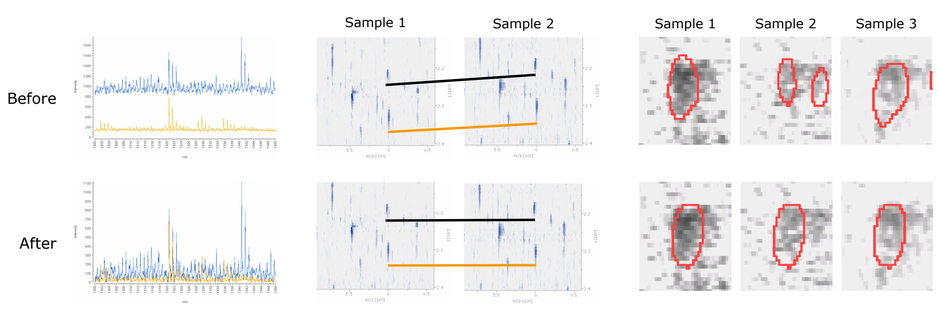


The database captures and stores the actual study design (e.g. study protocols, expected data) and the data that is generated throughout the project. This includes sample annotation details, measured data and platform-specific information (e.g. protocols, gene or protein annotations), and results from quality assessment and data analysis. Controlled vocabularies and consistency checks facilitate appropriate data curation.

Particularly powerful is the integration with sophisticated data analysis tools that can leverage the relationships of the database entries.

B | Data Quality Assurance





An automated workflow environment (left) performs standardized and reproducible processing of raw data from high-throughput/genome-wide molecular technologies, including GC-MS and LC-MS chromatograms (above), a number of 2D-gel platforms and high density microarrays (data not shown).

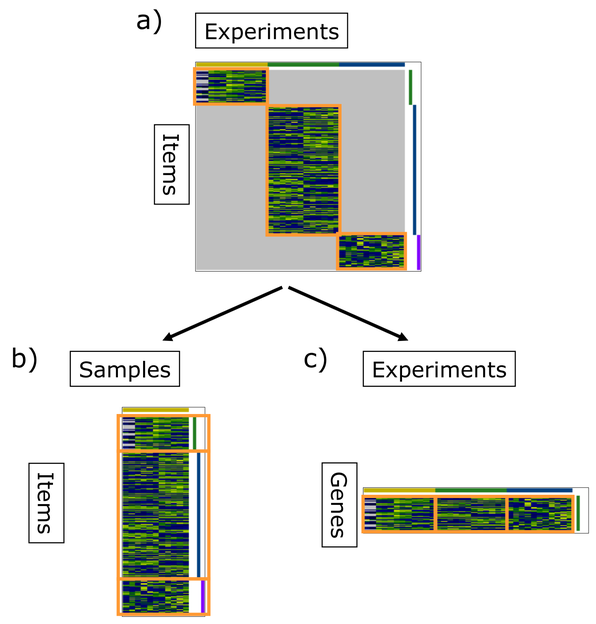
The MS workflow enables rapid, simultaneous processing of hundreds of data sets, a prerequisite for rigorous statistical comparison across several experimental conditions and samples. Continuous LC-MS chromatograms are shown before and after processing (above centre and right). Processing of MS data includes background subtraction, RT alignment and peak quantification.

Technology-specific workflows enable sensitive molecular profiling of proteomic, metabolomic and transcriptomic investigations. For data types not amenable to automated quality assessment (e.g. histopathology scores) peer-reviewing between project partners avoids potential site specific bias during integrated analysis.

D | Integrated Data Analysis

The key to the analysis of the data generated during the project is the integration of very different data types (e.g. classical toxicological endpoints, -omics data), typically collected independently (a).

The Genedata Expressionist® platform facilitates true integration of the data. The two most successful approaches are based on mapping to either a common sample (b) or joint biological name space, e.g. genes (c).



The sample based mapping enables analysis across the biological information layers without loss of valuable information. This "cross-omics" approach helps to find biomarkers characterizing a compound's effect without the bias of specific technologies. Mapping data on different molecules (e.g. transcript and protein expression) to a joint ontology (e.g. gene names) is particularly useful for the visualization of the findings within a broader biological context (e.g. pathways)

The resulting biomarkers are used for further mechanistic studies. Validation against larger data sets may justify their use for predicting the toxic potential of new uncharacterized compounds.

Scientifically validated enterprise IT platform

Genedata provides enterprise level computational biology solutions that are tailor made for toxicogenomic investigations.
Members of the Genedata team are based in Europe, the US and Asia, ensuring timely and expert service.

Advantage of Genedata Toxicogenomics Solution
Early detection of leads likely to fail for toxicity
Leverage of all available data (-omics and clinical)
Proven expertise in data analysis and IT

Members of the InnoMed-PredTox consortium

Altana, Bayer, Boehringer Ingelheim, Johnson & Johnson, Lilly S.A., Merck KGaA, Novartis, Novo Nordisk, Organon, Roche, Sanofi-Aventis, Schering AG, Serono, Servier, U Dublin, U Hacettepe (Turkey), U Würzburg, Genedata.

The InnoMed Project is supported by partial funding under the Sixth Research Framework Programme of the European Union (LSHB-CT-2005-518170)

Please contact us at expressionist@genedata.com or visit our website at www.genedata.com.