RUHR-UNIVERSITÄT BOCHUM

Faculty of Medicine
Medizinisches Proteom-Center[¥]
Workgroup Bioinformatics / Biostatistics*
Workgroup Clinical Proteomics[†]
University Hospital of Essen[‡]
Kairos GmbH[Ħ]

MEDIZINISCHES **PROTEOM** CENTER

Combining data from several 'OMICS' platforms

Michael Kohl*, Hagen Meckel*, Martin Trippler[‡], Dominik A. Megger[†], Maike Ahrens*, Thilo Bracht[†], Frank Weber[‡], Andreas-Claudius Hoffmann[‡], Hideo A. Baba[‡], Barbara Sitek[†], Jörg F. Schlaak[‡], Helmut E. Meyer[¥], Christian Stephan*[Ħ], Martin Eisenacher*

Introduction

Prevention and personalized medicine are key issues of contemporary medical research. Multi-OMICS approaches aim at measuring the dynamics of the most important biomolecules (i.e. genes, mRNAs, proteins and metabolites) in order to gain better understanding of the complex regulation of a cell. In the medical context, such efforts are promising for the discovery of novel biomarkers and the development of new drug targets. However, processing and interpretation of multi-OMICS data is usually challenging and requires a structured workflow.

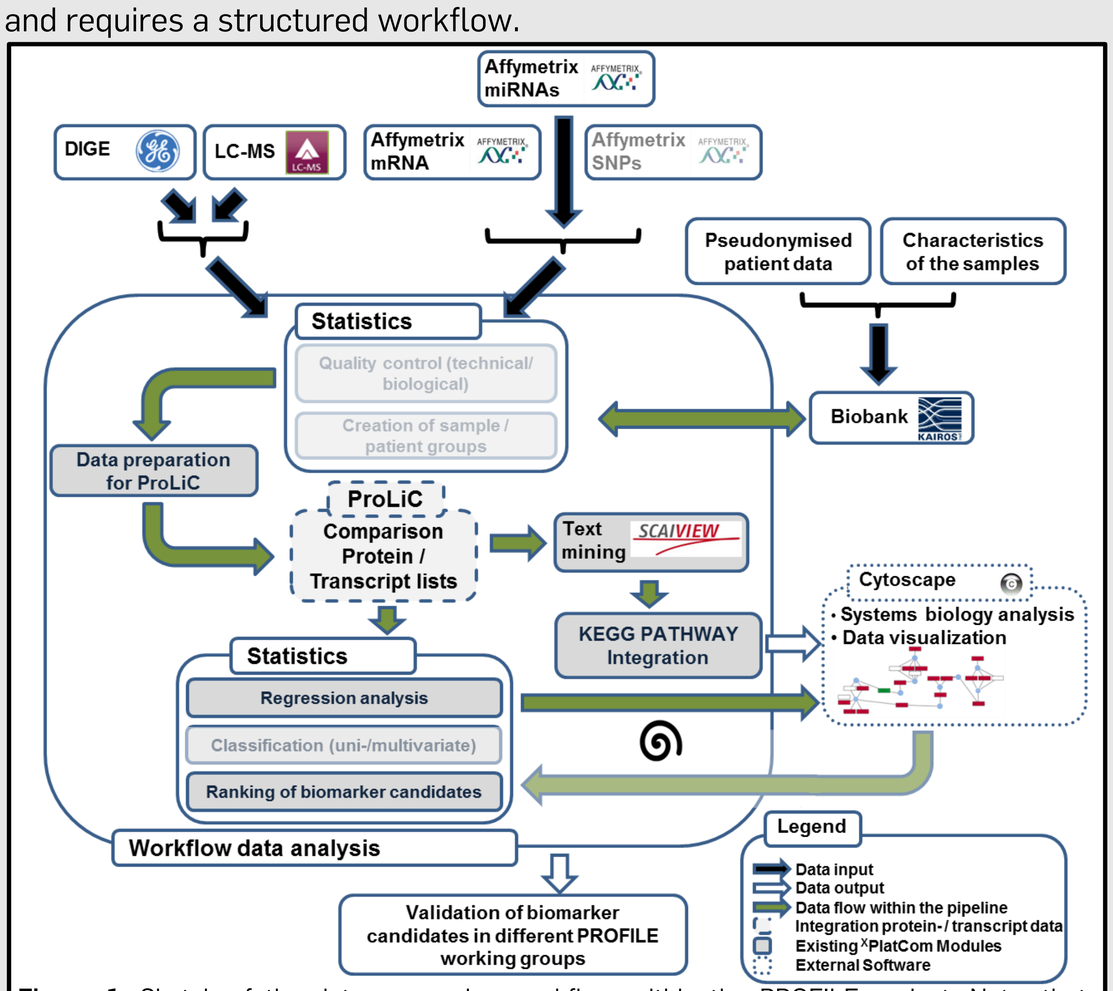


Figure 1: Sketch of the data processing workflow within the PROFILE project. Note, that transparent parts of the sketch indicate data processing steps, which will be implemented within XPlatCom in the near future.

Methods

Here, we present such a scheme for the processing of Proteomics and Transcriptomics (mRNA and miRNA) data (Fig.1)[1]. The workflow comprises several steps of data conversion, quality control, data comparison, text mining and statistical analyses. Additionally, a software named CrossPlatformCommander (XPlatCom) is sketched, which facilitates several steps of the proposed workflow in a semi-automatic manner (Fig 1., grey highlighted boxes). The performance of the workflow is shown using a hepatocellular carcinoma data set, obtained from the multi-OMICS project named PROFILE[2]. Samples from both tumor and healthy tissue of six HCC patients was used to carry out both a *DIfference Gel Electrophoresis* (DIGE)[3] and a label free (LC-MS) Proteomics experiment. The dataset was complemented with the results from an Transcriptomics experiment using the same samples.

This poster focuses on data comparison and the correlation between corresponding protein and gene expression profiles. Further details of the workflow are presented elsewhere[1].

Data comparison: XPlatCom takes advantage from the ProliC algorithm, which supports computing of overlaps (OL) and complements for the Proteomics and Transcriptomics experiments. To this end, ProLiC compares the measured peptide sets of the DIGE and LC-MS data sets (Fig.2, a). Proteins with similar peptide sets are considered as members of the same so-called biomolecule group (BG). Finally, measured genes are integrated into the BGs via gene name comparison.

Correlation of protein and gene expression profiles: For each overlap, regression analyses are performed both *globally* (each dot of the scatter calculated from the mean protein/transcript fold changes for a *certain biomolecule group*) and on the *BG level* (each dot represents protein/transcript fold changes of a *certain sample (i.e. a patient)*). Prior to calculation, Box-Cox transformation is performed to ensure normality distribution of the fold change values.

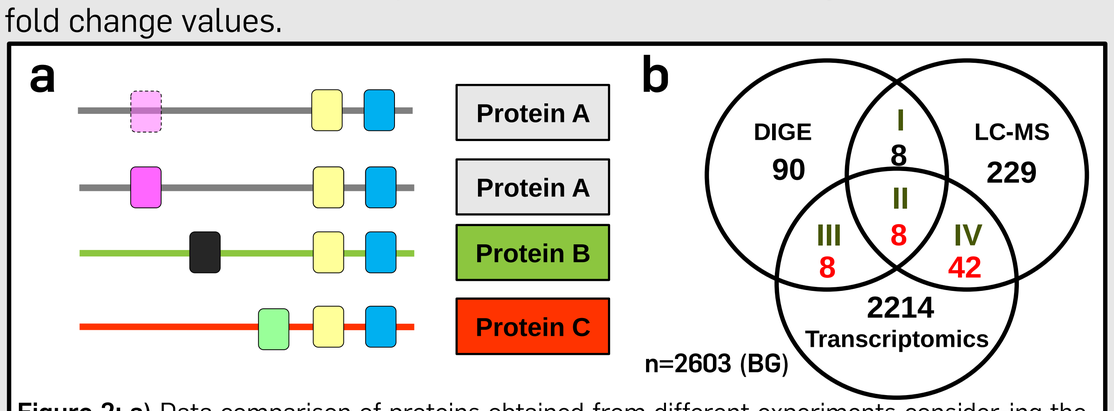


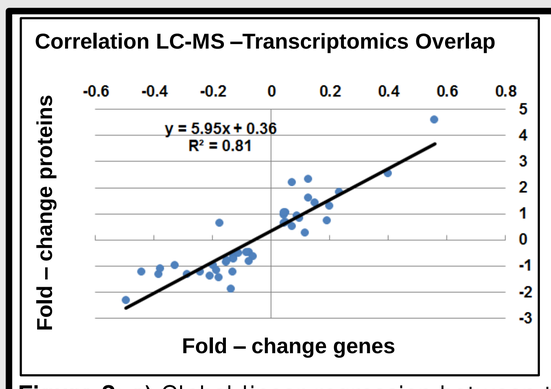
Figure 2: a) Data comparison of proteins obtained from different experiments consider-ing the measured peptide identifications. **b)** Venn diagram showing the calculated BGs within the overlaps and complements of the HCC experiments. Overlaps are marked with green roman numerals.

Results

Data comparison yields 2603 significantly regulated biomolecule groups in total. 66 BGs are found regulated in at least two of the experiments. 8 BGs are regulated significantly in each Proteomics and the Transcriptomics experiment.

Regression analysis shows good overall correlation for both OL IV (Fig. 3, b; R^2=0.81) and OL II (R^2=0.83, data not shown). However, there is almost no correlation between protein and transcript regulation for OL III (R^2=0.14, data not shown).

On the biomolecule group level 19% (OL IV) and 12.5% (OL II) of the BGs show R^2 values greater than 0.7. Concerning OL III, correlation is again weak (only 12.5% of the BGs with R^2 greater than 0.5)



	OL II	OL III	OL IV
Number BG	8	8	42
Number BG with R^2 > 0.7	1	0	8
Number BG with R^2 > 0.5	2	1	16
Number BG with p-value < 0.05	1	0	10

OL II: DIGE – LC-MS – Transcriptomics Overlap
OL III: DIGE – Transcriptomics Overlap
OL IV: LC-MS – Transcriptomics Overlap

Figure 3: a) Global linear regression between the Box–Cox-transformed transcript and protein fold changes given for OL VI **b)** BG numbers for OL II, III and IV calculated for the R^2 thresholds 0.7 and 0.5 and for the 0.05 threshold for the p-value of the F statistics.

Conclusion and Outlook

Our results confirm previous findings that show a rather weak correlation between mRNA and protein expression ratio[4]. It has been shown that correlation is related to pathways and functional categories[4].Therefore, XPlatCom links the results of the regression analysis to pathway information (KEGG Pathway module, Fig.1) and -in the near future- to GO categories. Moreover, weak correlation is at least partially caused by post-transcriptional regulations that modify the gene – protein interaction. Therefore, current XPlatCom extension efforts aim at integration of corresponding and already measured expression levels of miRNA, which is maybe the most important regulatory mechanism of mRNA expression. Currently, partial least square (PLS) regression[5] is being implemented within XPlatCom in order to assess the relationship between mRNA and miRNA expression profiles. PLS seems appropriate to cope both with abounding collinearities between the miRNAs (covariates) and the usually rather small sample sizes.

References
1. Kohl, M. et al. A practical data processing workflow for multi-OMICS projects, Biochim Biophys Acta (2013).
2. http://www.profile-project.de/
3. http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/2d_electrophoresis
4. Nie, L. et al. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications, Crit. Rev. Biotechnol. 27 (2007).
5. Li, X. et al. Modeling microRNA-mRNA Interactions Using PLS Regression in Human Colon Cancer, Bmc Medical Genomics 4 (2011).

PROFILE
Ministerium für Wissenschaft und Forschung des Landes Nordrhein-Westfalen
Ziel2.NRW Regionale Wettbewerbsfähigkeit und Beschäftigung
PURE

This work was funded by the BioNRW.PROFILE, Förderkennzeichen 780 005-1006-0050 and also by P.U.R.E. (Protein Unit for Research in Europe), a project of North Rhine–Westphalia (Germany).