

# Holistic approach to genomics of human-pathogenic fungi:

## Data warehouse for integration of data on transcriptome, proteome and metabolome of *C. albicans* and *A. fumigatus*

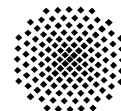
### Data Analysis and Interpretation



C. Schmitt (1), D. Albrecht (2), R. Guthke (2), A.A. Brakhage (2), S. Rupp (1), N.C. Hauser (1)

(1) Fraunhofer Institute for Interfacial Engineering and Biotechnology IGB, Nobelstr. 12, D-70569 Stuttgart

(2) Leibniz-Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (HKI), Beutenbergstr. 11a, D-07745 Jena



#### Motivation

- The proposed project aims to support a multi-disciplinary approach to characterize and quantify the infection process of human-pathogenic fungi.
- Host-pathogen interaction investigated on transcriptome, proteome and metabolome level within several projects of the SPP1160 could be stored and analyzed.
- The experimental data are deposited together with microbiological and biochemical data for functional genome analysis of both, pathogenic fungi and the host cells.
- A common platform for all members of the SPP1160-programme was provided in the first funding period in order to support genome-wide analyses and interpretations of the generated data.
- The second funding period will focus on the integrated analysis combining several types of data from all projects of the SPP1160 to support a comprehensive understanding of host-pathogen interaction in a systems biology attempt

#### Data Warehouse

##### DOM-CHiPS Central Portal ([www.domchips.de](http://www.domchips.de)):

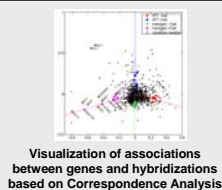
- Access to components A and B

##### Component A:

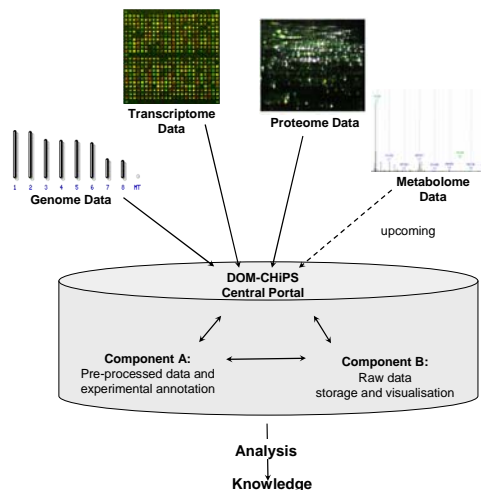
- including all preprocessed (secondary) experimental data annotations and analysis results
- Based on M-CHiPS, a microarray data warehouse developed and hosted at the DKFZ Heidelberg.

##### Component B:

- stores image based raw (primary) experimental data including annotations with features for primary data visualization
- stores genomic data from NCBI GenBank and UniProt as basis for experimental data
- Based on PROTECS (Decodon GmbH, Greifswald) and hosted at the HKI Jena.



Visualization of primary transcriptomic and proteomic data



#### Results

##### Matlab – R - Interface

- This Interface allows to apply and/or write any R (bioconductor) package or MatLab function on top of the existing functionality.
- The biologically interested user can easily switch between the two software tools.
- The statisticians/bioinformaticians can use either software package to write new functions.

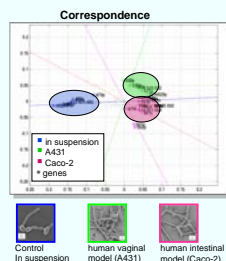
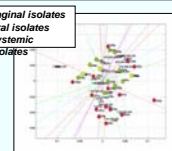


##### Visualization of associations

- Correspondence Analysis (CA) has been proven useful to visualize and elucidate associations among genes and the investigated conditions (1).
- CA can implicate experimental (numerical) data and detailed annotations of the experiment (e.g. descriptions of the biological background or clinical data) (2).
- To facilitate interpretation, Gene Ontology (GO) terms can be integrated into CA to display genes, experimental conditions and gene-annotations in a single plot. The position of the annotations in these plots can be directly used for the functional interpretation of clusters of genes or experimental conditions (3).

##### Clustering Clinical Isolates

- Clinical isolates (vaginal, oral and systemic) was grown in YPD (6h, 30 °C) and analyzed by transcriptional profiling.
- Projection by CA showed no distinct clusters related to the origin of isolation i.e. host niche.



##### Host – Pathogen - Interaction

In order to investigate epithelium-specific adaptation in *C. albicans* cells were grown on different human tissue models (vaginal, intestinal) and inert surfaces. Transcript profiling and CA revealed a set of genes specifically expressed upon contact with the respective surface, indicating that *C. albicans* is able to recognize and react specifically to distinct surfaces (4). Among these a set of potential cell wall genes could be identified which were further characterized (details see application Rupp).

#### Cooperations

- Hans Knöll Institute HKI (Brakhage, Guthke, Hube)
- Universität Würzburg (Morschhäuser, Einsele/Löffler)
- German Cancer Research Center (DKFZ)
- Universität Stuttgart (Dippon)
- Heinrich-Heine-Universität (Ernst)
- Universität Tübingen (Schröppel)

#### Acknowledgement

- This work is funded by DFG – SPP1160 „Colonisation and Infection by human-pathogenic fungi“ (project number HA5272/1-1)
- Special thanks to Kurt Fellenberg and Amin Moghaddas Gholami, DKFZ Heidelberg
- Thanks to Stefan Winter and Jürgen Dippon, Institute of stochastics and applications, Universität Stuttgart for mathematical input and discussions
- Thanks to all members for providing data and helpful discussions on improvements of data analysis

#### Outlook

- Advanced search and visualisation functionalities
- Include other data sources like CADRE, e-fungi, CGD
- Support of additional image analysis software
- Support of gel free proteomics
- Modelling of relations between transcriptome and proteome
- Support of metabolomics
- Analysis of host reaction after infection
- storage of image and movie data
- Collection of experimental data
- Use of international standards for data storage and representation
- Advanced methods for data pre-processing and analysis
- Workshop on „Transcriptome and Proteome Data Analysis and Data Warehousing“ will be held 12./13. June 2008 in Stuttgart. Invited speakers include Steve Oliver, Manchester; Rainer Spang, Regensburg; Ulrich Mansmann, MPI München; Michael Glocker, Rostock; Matthias Berth, Greifswald

#### References:

- Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD & Vingron M. (2001) Correspondence analysis applied to microarray data. *Proc Natl Acad Sci USA* 98: 10781-10786.
- Fellenberg K, Busold CH, Witt O, Bauer A, Beckmann B, Hauser NC, Frohme M, Winter S, Dippon J & Hoheisel JD. (2006) Systematic interpretation of microarray data using experiment annotations. *BMC Genomics* 7, 319
- Busold CH, Winter S, Hauser NC, Bauer A, Dippon J, Hoheisel JD & Fellenberg K. (2005): Integration of GO annotations in Correspondence Analysis: facilitating the interpretation of microarray data. *Bioinformatics* 21: 2424-2429.
- Sohn K, Senyürek I, Fertey J, Königsdorfer A, Joffroy C, Hauser NC, Zelt G, Brunner H & Rupp S. (2006). *FEMS Yeast Research* 6: 1085-93.

Work was presented at:

- Hauser N, Schmitt C, Brunner H, Fellenberg K, Rupp S: Integration of Transcriptome and Proteome data of Human Pathogens using a Data Warehouse and Attempts to describe Complex Infection Processes Foundations of Systems Biology FOSBE, 9.-12. 9. 2007, Stuttgart
- Moghaddas Gholami A, Schmitt C, Visvanathan M, Hauser N, Rupp S, Hoheisel J, Fellenberg K: M-CHIPS: Universal platform for Microarray Data Interpretation. German Conference on Bioinformatics 26.-28. 9.2007, Potsdam
- Moghaddas Gholami A, Schmitt C, Visvanathan M, Hauser N, Rupp S, Hoheisel J, Fellenberg K: M-CHIPS: Universal platform for Microarray Data Interpretation. DECHEMA Statusseminar Chipstechnologien 31.1.-1.2.2008, Frankfurt