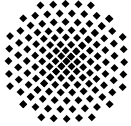


# Integration of Transcriptome and Proteome data of Human Pathogens using a Data Warehouse and Attempts to describe Complex Infection Processes

Nicole Hauser<sup>1</sup>, Corinna Schmitt<sup>1</sup>, Herwig Brunner<sup>1</sup>, Kurt Fellenberg<sup>2</sup> and Steffen Rupp<sup>1</sup>



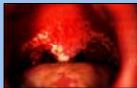
<sup>1</sup>Fraunhofer-Institut für Grenzflächen- und Bioverfahrenstechnik (IGB), Molecular Biotechnology, Stuttgart  
<sup>2</sup>Deutsches Krebsforschungszentrum, Functional Genome Analysis, Heidelberg

Correspondence: Dr. Nicole Hauser, Nobelstr. 12, 70569 Stuttgart, Tel.: +49 (0) 711/ 970 - 4044, nicole.hauser@igb.fraunhofer.de



## Fungal Pathogens

**Introduction:** In recent years, infections with eukaryotic pathogens have become a major problem in the industrialized countries. In particular immuno-compromised individuals, as caused for instance by chemotherapies, organ transplantations or AIDS, may acquire superficial and eventually severe and life-threatening systemic infections.

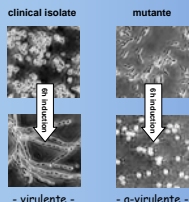


Typical clinical syndromes: chronic mucocutaneous candidiasis of a child with a cellular immune defect (left panel) oral thrush, superficial mucosal infection of an AIDS patient (right panel)

Only a few compounds exist that serve as active substances and resistance often occurs. Comprehensive knowledge about virulence and resistance mechanisms on the pathogen side, reactions on the host side as well as the overall interaction of pathogen and host is prerequisite to the development of new drugs and therapeutics. We have developed an *in vitro* reconstructed three-dimensional tissue model of various organs like gut or skin (see Infection Model, upper box right). This model system is used to investigate infection processes and host-pathogen interactions on transcriptome and proteome level. Results from such assays are analyzed using MCHIPS (see Data Analysis, lower box right).

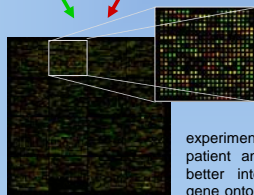
This data warehouse platform combines data from various origins and integrates not only experimental and clinical descriptions but also other parameters and information such as gene annotations. Our recent developments of integrating annotations from the GO-consortium demonstrate the expedience of computer-based statistical analysis of annotations for a systemic description of such complex processes like pathogen infection.

The pathogenesis of fungal infections is mainly determined by the status of the host. However, specific factors exist, which allow the fungi to colonize, survive and eventually infect the host. Among these factors, there is the ability to morphologically switch to hyphal growth. Mutants lacking the yeast-to-hyphal transition show strongly reduced virulence *in vivo*.



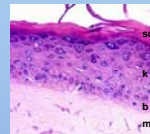
Here two strains, a virulent clinical isolate (left) and an a-virulent strain defective for hyphal growth and thus for virulence (mutante, right panel) are compared. cDNA from both strains grown under identical conditions was generated by reverse transcription of total RNA extracted from each strain and analysed on CGA.

Candida Genomic Array



**Transcriptional profiling** A DNA-Array, containing virtually all of the predicted ORFs (~7200) has been established and is used for whole genome comparisons of various strains and conditions. Analysis is done using M-CHiPS, which allows to integrate experimental parameters and genotypes, as well as patient and therapy associated information. For better interpretation, significant associations of gene ontology terms with gene clusters have been implemented.

## Infection Model



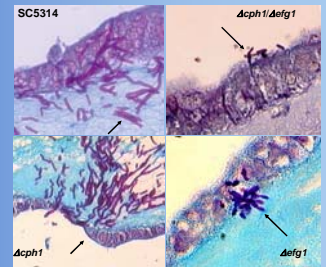
### *In vitro* reconstructed human epidermis

Three dimensional *in vitro* reconstituted model tissues were established as test systems representing human skin (left; **sc** stratum corneum, **k** differentiating keratinocytes, **b** basal lamina, **m** collagen matrix containing dermal fibroblasts ) or human gut (down) equivalents. These model tissues can be used to monitor the different steps of fungal infection.

With these assay systems we have a means to assess virulence of various mutant strains without need for animal testing. Moreover they facilitate drug screening procedures.

### Invasion into intestinal model

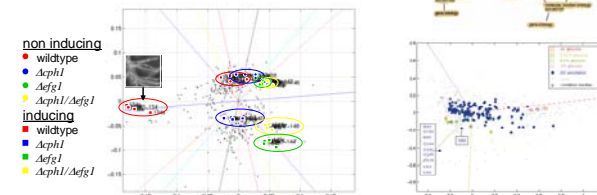
The infection assay was performed using immobilised *C. albicans* strains that were placed on top of the tissue and incubated for 18h. Only the strain  $\Delta$ cpH1 $\Delta$ efg1 is not able to infect the tissue.  $\Delta$ efg1 shows reduced invasion but is able to penetrate the epithelial layer in a pseudohyphal growth form.



## Integration of GO-Annotations

The Gene Ontology Consortium ([www.geneontology.org](http://www.geneontology.org)) defines GO as an international standard to annotate genes. GO has a hierarchical structure starting with top-level ontologies for molecular functions, biological processes and cellular components.

In order to understand the biological processes present in a large dataset of genes we make use of the annotations and structure of the GOs.



To facilitate interpretation of complex microarray data, we integrated GO-annotations into Correspondence Analysis 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 functional interpretation of clusters of genes or experimental conditions.

M-CHiPS is a microarray data warehouse rather than a mere repository. It integrates a software platform for microarray data analysis and can also be used as a LIMS (Laboratory information management systems) for DNA chip experiments. All information, including experimental and clinical data is held in a format ready for statistical analysis.

The Database currently holds more than 6,500 experiments from different organisms:

**M-CHiPS** Multi-Conditional Hybridisation Intensity Processing Software

PostgreSQL Database  
Database Browser  
Selection of subsets

dfz

[www.mchips.org](http://www.mchips.org)

*Man* (cancer-specific)  
*Mouse*  
*Drosophila melanogaster*  
*Arabidopsis thaliana*  
*Saccharomyces cerevisiae*  
*Candida albicans*  
*Bacillus subtilis*  
*Trypanosoma brucei*  
*Neurospora crassa*  
*Pseudomonas putida*  
*Rhizobium meliloti*

A Matlab-based microarray data mining tool including

- Normalisation
- Clustering, Correspondence analysis
- Quality filtering
- Automated analysis of annotations

## Transcriptome and Proteome

## Data Analysis