EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

Exploratory Workshop Scheme

Standing Committees for
- Life, Earth and Environmental Sciences (LESC)
- Physical and Engineering Sciences (PESC)

ESF Exploratory Workshop on

Phenomics - Advancing High-Resolution Genome-Wide Phenotyping in Yeast

REPORT

Göteborg, Sweden, 2 - 4 March 2007

Convened by:
Anders Blomberg

Dept of Cell and Molecular Biology, Göteborg University

Co-sponsored by

GÖTEBORGS UNIVERSITET

CHALMERS

1. Executive Summary

This ESF supported exploratory workshop on the theme "Phenomics – advancing high-resolution genome-wide phenotyping in yeast" was hosted by the Göteborg University, Göteborg, Sweden. It was convened by Anders Blomberg at the Department of Cell and Molecular Biology, Göteborg University. In addition, highly valuable local administrative practical support was provided by PhD students and post-docs at the university and at Chalmers University of Technology. Besides sponsoring from ESF, local financial support was provided by Göteborg Mathematical Modelling Center and the Stochastic Centers at Chalmers which will cover travelling costs for some of the participants. The Chalmers Bioscience centre is also providing support for the Saturday night social event. The workshop programme started Friday lunch and ended Sunday lunch 2 – 4 March, 2007. The workshop was arranged and participants invited on a rather tight time-schedule (formal ESF letter reached the convenor early January 2007) to enable discussions and potentially consortia constellations regarding the EC framework 7 call (dead-line mid-April). The workshop was held at two different university venues: Friday at Wallenberg conference centre and Saturday plus Sunday at Ågrenska Villan. Most external participants (14 out of 17) stayed at a small family hotel called Villa Fridolin which is situated close to Ågrenska Villan.

Phenomics aims at bridging the genotype-to-phenotype gap on a genome-wide scale, an activity that is highly important in our mechanistic understanding of human diseases, industrial scale production of wine and beer, species evolution and population genetics. It is thus anticipated that the field will have far-reaching consequences on a wide array of central topics in biomedicine and biotechnology. The yeast species *Saccharomyces cerevisiae* (beer and baker´s yeast) was the first eukaryotic organism to have its genome sequence finished (published April, 1996). During the international effort to sequence yeast, Europe played a vital and essential role. The genome sequence was the starting point for the construction of large-scale strain collections where genes have been mutated (usually deleted) one by one. A gene-deletion collection for *S. cerevisiae* was completed in 2001, which identify the starting point for large-scale phenotypic screens using the genome-wide collections of strains – i.e. Phenomics.

The goal of the workshop was to address the three main activities in Phenomics, all of instrumental importance for the advancement of high-throughput and high-resolution genome-wide phenotyping: i) Construction of new yeast strain collections, ii) Technical advancement in high-throughput and high-resolution in vivo phenotyping, and iii) Database presentation, modelling and standardisation in phenome bioinformatics. This was also the titles of the three main sessions of the workshop. To achieve this goal it was essential to bring together leading European experts from multiple disciplines to present and discuss the current status of high-throughput yeast phenotyping. In total 27 participants from 10 European countries took part in this event. In addition, two non-European experts representing the main yeast database Saccharomyces Genome Database (SGD) at Stanford, USA, were also present during the workshop. This was essential to provide an international perspective to this European initiative, but also to discuss future joint efforts for presentation and analysis of genotype-phenotype information. It should also be pointed out that three European companies with highly relevant activities/products related to phenotyping participated on their own expense in the workshop.

The final programme in principle adhered to the initially proposed programme in the application to ESF. However, based on suggestions in the evaluation of our ESF application for an exploratory workshop, a number of adjustments were made. The evaluation pointed out that the procedures for high-throughput phenotyping should be extended to also incorporate other yeast species. The suggestion is adequate and timely since gene-deletion collection for the to *S. cerevisiae* very distantly related yeast species *Schizosaccharomyces pombe* is just about to be finished, an initiative now presented at the workshop by Per Sunnerhagen, Sweden. Genome-wide collections of gene-deletions in distantly related yeast species will be highly valuable in the identification and characterization of orthologous gene functions, including gene functions in humans. Another extension to the initial programme that widen the phenomics concept to include the genetic variability between and within species was the presentation by Gianni Liti, England (part of the Ed Louis group) of the sequencing of various strains of both *S. cerevisiae* and the closely related species *S. paradoxus*. This work is in progress and will be finished spring 2007. The strain specific sequence information will be a great resource for studies on the influence of single nucleotide polymorphism on gene function and phenotypic traits. It will furthermore have great impact on the way we study important mechanisms in population genetics as well as our description of the context dependence of gene function. It is truly a

major paradigm shift in our studies of complex genetic traits and these strains will surely be a great resource for the yeast community for quite some years. The great complexity of natural yeast population was also exemplified in the talk by Dorit Shuller, Portugal (part of the Margarida Casal group) who introduced studies on the genotypic and phenotypic characterisation of yeasts isolated from various European vineyards. The evaluators also pointed out that the phenomics field would benefit from interactions with the ongoing systems biology initiative. We are thus glad that Stefan Hohmann, Sweden, who is one of the main organisers of the international systems biology platform, could take part in the workshop during the first day and also present information to the workshop participants about local and international efforts that could be of relevance to phenomics (and vice versa). In addition, in the theoretical session on Phenome Bioinformatics Balazc Papp, England/Hungary, was invited to present work on the integration of systems biology based models of yeast metabolism in the prediction of the phenotypic outcome of genetic and environmental perturbations made to yeast. Thus, all aspects raised by the external evaluation and the LESC and PESC standing committees of ESF have been addressed in the final programme of the workshop.

Each presentation was strictly limited to 20 minutes to also allow for a good discussion in conjunction with the talk. It was a real pleasure to note the high level of engagement from all participants in these follow-up discussions throughout the workshop where technical details and concepts were vividly debated. Another really positive thing was that some participants were able to attend ”in the last minute” even though they had earlier indicated that it might be impossible to be part of the workshop. However, to encompass all talks we then had to make some last minute reallocation in the programme, the main change being that the round table discussions for all three sessions were moved to Sunday morning. Thus, on the last day a two hours long discussion summarising and extending the main topics concluded the workshop. Several new ideas and suggestions for future research projects were discussed. The ESF representative Constantinos Doukas presented the different ways that ESF could support this Phenomics initiative in the future. The presentations and the engagement in discussions from Constantinos Doukas during the whole workshop was highly appreciated by all participants. It was decided by the group to jointly apply for a Research Network Programme from ESF, coordinated by Anders Blomberg, with dead line late October 2007. It was strongly expressed by the participants that regular Phenomics meetings would be a highly valuable forum to form a strong international network and to further strengthen the European profile in this field.

2. Scientific content of the event

The following summary of the scientific content of the workshop is based on notes taken by Jonas Warringer and Anders Blomberg.

After a short welcome note from the convenor **Anders Blomberg (Göteborg University),** the workshop was opened by **Constantine Doukas** from **the Life, Earth and Environmental Science Committee**, at the **European Science Foundation, ESF**, who provided a brief introduction to the ESF and its mission to serve as a common scientific platform for its membership states and explore new directions for research at the European level. Established in 1974, ESF consists of 75 member organisations in 30 countries and extend beyond the European Union membership states to encompass e.g. Norway and Turkey. ESF intends to serve a complementary role to the European Research Council by focusing on member organisation coordination. It was made explicitly clear that funds for individual researchers, within a collaborative effort that has been embraced by ESF, is provided by the corresponding membership countries. Following Constantine Doukas introductory talk, representatives from two local financial contributors, **Stefan Hohmann** of the **Göteborg University, Science Faculty** and **Olle Nerman** of the **Chalmers Technical University**, presented the respective universtity’s yeast and phenomics related initiatives. Stefan Hohmann gave a brief overview of the yeast community in Gothenburg and over the platforms in Quantitative Biology and Chemical Biology that the Science Faculty supports. He also put Phenomics in the context of the international Yeast Systems Biology initiative. Olle Nerman provided an overview over the Gothenburg Mathematical Modelling Center and the Stochastic Centers which have allocated one third of its budget to biosciences, primarily Systems Biology. He also took the opportunity to advertise the Masters programs in Bioinformatics and Systems Biology that the Center supports.

Session I: Yeast Strain Collections

This session aimed at presenting various types of strain collections available for large-scale phenotypic screens. It covered presentations on the genetic variability found in natural isolates of yeast, man-made systematic strain collections as well as issues of how strain collections should be stored and distributed.

In the first talk of the yeast strain collection session **Dorit Schuller** from the **University of Minho** (Braga, Portugal) presented results concerning large scale biogeographical studies aimed at quantifying the biodiversity of natural yeast populations. In two separate studies, one focused on a set of vineyards in northern Portugal and one focused on vineyards from widely spaced locations throughout Europe, natural yeast samples (roughly 1,600) were isolated and characterised by a number of different methods. (e.g. genotyping by PCR amplification of micro-satellites). A surprising finding was that *S. cerevisiae* was not easy to isolate from the vineyards and was mainly found on rotten grapes. In both studies there were little correlation between geographic distance and genetic similarity between individual isolates. However, taken the genetic variability for a whole geographically isolated population into account it was clear that genotypes of closely situated populations clustered. The isolated natural populations were phenotyped with regards to different traits, some of them relevant to the wine-making industry like the aromatic profile, providing a link to industrial applications. It was discussed whether these natural isolates could be the base for a new natural *S. cerevisiae* strain collection. One issue to consider then is the usually low genotype/phenotype stability over time in the laboratory of these natural isolates.

Following-up on the topic on natural populations **Gianni Litti** from the **University of Nottingham** (Nottingham, England) presented a talk on a set of 72 wild *S. cerevisiae* and *Saccharomyces paradoxus* strains which are currently being sequenced to medium coverage (1-3x) in collaboration with The Sanger Institute. These strains were selected from a collection of 334 strains and comprised natural isolates from three continents, clinical isolates and commonly used laboratory strains. Preliminary results from this Saccharomyces sensu stricto survey were discussed, particularly in terms of links between SNP density and chromosomal location and subtelomeric dynamics. Data was presented that indicted that genotype diversity among the *S. paradoxus* strains were much greater than for *S. cerevisiae* which could be a reflection of the domestication of the latter species. It was proposed that that *S. paradoxus* in that respect more mimicked genetically the variation found in human populations and might then in the future constitute a good model system for populations studies. Also with regards to gamete compatibility and sequence divergence the relationship appeared to be linear, suggesting that there is no specific “line” where speciation occurs.

Turning to the topic of genetically manipulated collections of lab strains, **Anders Blomberg** of the **Göteborg University** (Gothenburg, Sweden), talked about a new collection of strains carrying temperature sensitive (TS) alleles corresponding to essential genes, which is currently being constructed by the Charlie Boone laboratory in Toronto, Canada. TS-alleles have been obtained from research laboratories through-out the world where ts-strains has been constructed in various *S. cerevisiae* strains, where after alleles have been PCR amplified and transformed into a common strain background (BY4741). This strain collection currently contains 742 ts alleles, corresponding to 467 genes, and the aim in this first round is to reach 500 genes in total. The main objective is to use this collection of essential genes in the large scale screening of double knockout arrays (SGA; Synthetic Genetic Arrays). However, each construction is also phenotyped as such, particularly with regards to the temperature response profile. It was shown that the specific pattern of temperature sensitivity of these strains in basal medium is distinct enough to allow functional groupings of genes.

Karl-Dieter Entian, the **Institute for Molecular Biosciences** (Frankfurt, Germany), continued the session with a talk on a collection of yeast strains which are engineered for use in drug discovery, specifically with regards to drugs that target human parasites. Traditional screening procedures, either *in vivo* of libraries of parasites or *in vitro* of specific targets, tend to produce a lot of false positives and very few candidates are of any real practical value. Therefore the TOP (Target Oriented Protein screening) methodology was developed. The aim is to find drugs that differentially target genes which have homologs in humans as well as pathogens. The homologous genes from both human and pathogen are introduced in yeast (the first targets have been homologues of essential yeast genes) and screening is performed to reveal substances which target only the pathogen. Combined with bioinformatics, termed TOP-SE (Screening Enhancement), the approach promises to be a potentially powerful drug screening tool. A successful case study was presented using the drug trimethoprim.

Turning from *S. cerevisiae* to *Schizosaccharomyces pombe* **Per Sunnerhagen** from **Göteborg University** (Gothenburg, Sweden) presented the fission yeast deletion collection which is currently being constructed partially by a South Korean consortium and partially by the laboratory of Paul Nurse (England). This collection has the benefit of being bar-coded, allowing also large scale competition assays to be performed. Some problems specific for the fission yeast, such as a minimal and unstable diploid state, lower frequency of homologous recombination and the existence of cryptic TS alleles in some of the strains (which are now being removed from the collection), were discussed. The lower rate of homologous recombination results in the need for longer homologous flanking regions (roughly 80 nt) than for *S. cerevisiae* to guide site-directed mutagenesis, and more time-consuming post-transformation strain verification. Overexpression assays were considered a viable methodology also for the fission yeast whereas the SGA technique in its current form was not. Currently 2,600 haploid strains are available whereas the diploid collections will be released after publication.

Joerg Hauf from **SRD – Scientific Research and Development, Euroscarf** (Frankfurt, Germany) finished the session with a talk on the collections housed and shipped by this company. Currently, their collections compromise 30,000 yeast strains in various strain backgrounds and about 2,000 plasmids. The strains represent both single deletions in various backgrounds as well as TAP (Tandem Affinity Purification)-tagged constructs and some smaller sub-collections. The strains are maintained using rigorous quality controls and verification of strains prior to shipping. This involves streaking for single colonies, replica plating on a routine set of drop-out plates to check for markers, and monitoring of the uniformity of growth. Strains are stored in glycerol stock and shipped on agar plats or on agar slopes within 24h of ordering. Specific issues concerning the collections were discussed, such as the question of suppressor mutations, strain stability and the increased evaporation associated with a scaling up from 96 to 384 well plates. The risk of more frequent contaminations were also discussed for the higher density formats.

Session II: In vivo High-Throughput Phenotyping

In this session presentations addressed how we can increase throughput in phenotype screens, and at the same time maintain, or even improve, resolution and precision. It mainly covered yeast phenotyping in vivo, like growth analysis in different formats and automated microscopy of GFP collections.

Carl Singer from **Singer Instruments Co Ltd.** (Roadwater, England) started the Phenotyping Session with a presentation of their flexible robotic system (RoToR) for agar-based phenotyping of yeast strains. The system is designed for speed as well as accuracy (currently achieving 5-10 µm precision). Speed is achieved by using disposable plastics, avoiding time consuming washings. A critical parameter is still the evenness of the surface, necessitating careful pouring of agar. The issue of planar and horizontal were discussed as well as the use of even bigger dishes or greater head pressure to resolve the problems. The system now works fine for the 1,536 pinhead format and they are aiming for good resolution at 6,144 colonies per plate.

Steve Oliver from **Manchester University** (Manchester, England) presented a talk addressing several important aspects of phenotyping. One part of the presentation focused on the tet-promotor system by which the expression level of a gene is controllable within certain limits by the addition of doxycyclin. It was put forward that the $tetO_2$ is better suited than $tetO_7$ in studies involving heterologous expression of human proteins since some human transcription factor(s) can bind and regulate the activity of $tetO_7$. A fundamental problem is that the repressible tet-promotor cannot be silenced completely even by addition of high concentrations of doxycyclin. Hence, for some essential genes the expression is still high enough at high concentrations of doxycyclin to allow unperturbed growth. An attempt to get around this problem is to combine the tet-promotor system with the degron system which adds an N-terminal degradation tag which, at higher temperatures, is recognised by the proteasome. Thus, these combined systems potentially allow for control of both expression and protein breakdown. For the expression of the *MSS4* gene the dual control system resulted in good control and no expression could be scored, while the problem of fully repressing *RAD53* could not be solved. It was concluded that there is a need for various, and even novel, systems for general gene expression control. Also mentioned was experiments aimed at phenotyping yeast knockouts in grape juice, addressing the lack of understanding of yeast ecology in its natural environment as well as attempting to widen phenotype space.

Jonas Warringer from the **Göteborg University** (Gothenburg, Sweden) continued the session with a talk that emphasized the necessity of extending the concept of fitness as it is mostly commonly applied in yeast phenotyping to encompass not only growth rate but also other features of growth. It was pointed out that many environments primarily affect growth lag (adaptation time) or growth efficiency rather than growth rate and that many gene-environment interactions only are detectable in physiological windows other than growth rate. In fact, certain gene-environment interactions were even reflected in non-standard growth dynamics such as multimodal growth rather than in defects in the classical growth variables.

Shifting the focus from the practical experimental aspects **Peter Gennemark** and **Olle Nerman** from the **Chalmers University of Technology** (Gothenburg, Sweden) talked about the statistical problems in high-throughput phenotyping. In particular they discussed certain systematic errors and other problems encountered when the measurement precision is high. It was shown that when the precision is increased systematic artefacts, such as spatial effects on the experimental plates, tend to increase in importance. This necessitates careful experiment design and planning, extensive randomisation and normalisation procedures, some of which were outlined in the talk. Simple measures, such as more repetitions in the control group (Wt) may also help alleviating the problem when it comes to final statistical tests.

Next, **Achim Kohler** from the **Norwegian University of Life Sciences** (Aas, Norway) presented a talk on vibrational spectroscopy for high-throughput phenotyping of cells. This technology which utilizes the absorbance of energy in the infra-red spectrum to look at molecular vibrations has the potential to resolve particular features of cells *in vivo*. Different regions of the spectra can be utilized to focus on lipids, secondary structures or polysaccharides in dried 10-30µl of yeast cultures. Data were presented that showed that the technique could clearly discriminate between *S. bayanus* and *S. cerevisiae* strains. However, the molecular information content is limited and it was discussed to what extent the technology has the power to resolve differences between mutants (which is currently being tested). The technique can also be used for micro-spectroscopy to look at single cells, but with a substantially lower throughput.

Sepp Kohlwein of the **University of Graz** (Graz, Austria) presented a talk on high-throughput microscopy and how to construct a simple and reliable highly automated system for yeast cell imaging. The talk focused on an approach using cells growing on agar in high-density format. This allows for utilisation of robotics where colonies are repeatedly re-pinned to allow cells to be actively growing at the stage of microscopic analysis. A critical points in this potentially high-throughput approach was the requirement for the cells to be distributed in a monolayer and thus it was optimised to generate small colonies. The system has been tested for roughly 600 GFP tagged proteins of *S. cerevisiae* and results/images are to be found at their local database YPL (http://ypl.uni-graz.at). A suggestion was to make systematic collections of both N- and C-terminal tagged proteins, since sometime problems have been encountered with the currently available C-terminally tagged GFP collection.

Mats Kvarnström from **Fraunhofer-Chalmers Centre for Applied Mathematics** and **Chalmers University of Technology** (Gothenburg, Sweden) gave a talk on image analysis for quantitative microscopy, particularly with regards to GFP-fluorescence labelled proteins in yeast cells. Some of the critical imaging steps were highlighted, such as finding and recognising the cells in bright field images, a task which is typically not accomplished simply by standard, “off-the-shelf” edge detection algorithms. The large variability in contour characteristics of individual cells is the main challenge here. However, by combining image gradient information (the rate of pixel change at each pixel) with constraints on the shape of the cells (typically convex), the performance is significantly increased. It is also possible to use information from several focal planes (so called z-stacks) to increase accuracy. During the discussion, the problem of overlapping cells was emphasized.

Ending the phenotyping session **Uros Petrovic** from **the Jozef Stefan Institute** (Ljubljana, Slovenia) presented a talk on the integration of genetic interaction (gene-gene and gene-chemical) and gene expression data by network analysis. The presentation also covered some technical notions on the use of the GenePix software for scoring colony growth on agar plates. In general there is little overlap between genetic interaction and expression data changes. However, by focusing on the master transcriptional regulator, for which the genetic buffering effects are less pronounced, order and predictability in combined networks of genetic and transcriptional data can be achieved. This was illustrated by an analysis of the TOR pathway for which a model predicted from large-scale data was largely consistent with published information.

Session III: Phenome Bioinformatics

High-throughput analysis generate massive amounts of information. This session presented and discussed how the phenotype data is to be stored and handled in databases, as well as analysed and integrated with other types of large-scale data. It also highlighted the connection of phenotyping to systems biology.

The Phenome Bioinformatics session was initiated by **Robert Stevens** from the **University of Manchester** (Manchester, England) with a talk where he presented the concept of ontological classification of genome features using OWL (the Ontology Web Language). The presentation outlined how to derive higher level information from different data sets and hinted at its potential in phenotype-based protein classifications. As an example, the talk focused on an ontological classification endeavour aimed at the detection of novel protein phosphatases, using protein domain data as input variables. It was shown that the ontological approach performed equally well as expert manual classification systems. The structure and requirements of an ontological approach was outlined, most specifically the requirement for digital input data.

Eurie Hong from the **Saccharomyces Genome Database (SGD)** housed at **Stanford University** (Stanford, USA) talked about phenotype analysis and presentation at SGD. The storage, accession and display of data was discussed as well as tools for data integration. It was emphasized that the source of the information is the scientific community and that SGD is dependent on input from researchers. Furthermore, the need for a controlled vocabulary and for standardizing phenotypic data sets was highlighted, especially since phenotypic data in itself covers such a wide area. Most phenotypes are conditional (e.g genetic background, growth temperature, media used) and it was discussed to what level that information should also be represented at the database. Interaction and integration with other model organism databases were also discussed and in this context the problems connected with the vastly different kinds of phenotypic information which is derived from different organisms were addressed.

Luciano Fernandez from **Göteborg University** (Gothenburg, Sweden) introduced the yeast phenotype database PROPHECY – PROfiling of PHEnotypic Characteristics in Yeast. This database is designed to manage and display high resolution growth curve data from genome-wide and sub-genome wide screens. The database has as one of its cornerstones that the phenotype data should be available and displayed at different levels of abstraction, i.e. anything from the full growth curves OD measurements to compact visual displays of gene-environment interactions via phenotype indexes (LPI; Logarithmic Phenotypic Index). The database also provides the opportunity to integrate external data resources and features (e.g. protein localisation, functional classification) and the use of a filtering tool which allows filtering on phenotypic response.

Connecting phenotyping to the growing field of Systems Biology **Balazs Papp** from **Manchester University** (Manchester, England) and the **Biological Research Centre** (Szeged, Hungary) presented a systematic effort to predict the outcome of genetic and environmental perturbations in yeast metabolism. This is particularly important as most perturbations has little or no phenotypic effect when tested experimentally, either due to the intrinsic buffering capacity of the cell or due to environment-specific phenotypic effects. A computational tool to predict synthetic lethality among all metabolic genes was presented. The tool is based on Flux Balance Analysis and thus uses the direction and stochiometry of the reactions as input variables to calculate the optimal growth rate during steady state growth. It has been previously shown that viability of single gene deletants can be predicted with 80-90% accuracy. For double deletions (synthetic lethals) that figure was 50%, which is substantially higher than the 0.6% expected by chance. The main limitation of the model is that it only predicts optimal network behaviour and does not track the dynamics of the system.

The final speaker of the session, and of the whole workshop, was **Blaz Zupan** from the **University of Ljubljana** (Ljubljana, Slovenia) who talked on the concept of Epistasis and Inference of Genetic Network Topology and presented the tool GenePath. This software utilizes input from genetic/phenotypic experiments to infer the topology of genetic networks and predict epistasis and parallelism in pathways. Suggestions for how to experimentally test a specific feature of the model is provided as an experiment planning tool. The utility of the tool in teaching as well as the role of network analysis in hypothesis formation and network planning was discussed.

3. Assessment of the results

On Sunday morning all participants assembled for a round-table discussion on key issues of the three main workshop topics (the three sessions) and desirable follow-up activities. All topics were vividly discussed and all participants took an active part.

Key-issues from the strain collection session:

- **Is the tet-promoter collection still a valuable collection that should be made?** It has earlier been proposed in a joint EC application involving many of the workshop participants that a strain collection with genes behind a controllable promoter would be valuable in a number of ways. As pointed out during the workshop by Steve Oliver it is not possible to fully shut-down transcription using the tet-system, even at high concentrations of doxycyclin. However, the good thing with doxycyclin as an external effector is that its impact on cell growth is marginal and thus hopefully only minor influence on cell physiology will result. No-one knew of progress in attempting to improve the tet-promoter via mutagenesis to make it more tightly regulated. It was concluded that tet-promoter optimisation would certainly be worth-while. It was also mentioned that the *MET25* promoter is under tighter control, however, the use of + or – methionine in the growth medium will of course have a more substantial influence on cell physiology. Apparently there is currently no "perfect" system since even degron-activated protein degradation was during the workshop shown to fail in certain cases.
- **Should collection of C/N terminal fusions of GFP be made?** It was put forward by Sepp Kohlwein during the workshop that a GFP collection where proteins are tagged independently in both the C- and N-terminal ends would be valuable. One reason for this is a rather high degree of false constructs (30% was put forward as a number) in the current GFP collection. Some of the other participants did not have that bad experience with the current C-terminally tagged collection, however, most had only analysed a rather limited number of strains. The problem of N-terminal tagging was discussed and the cre-lox system was proposed to be one good technical solution to systematic GFP-insertions. However, it was not really known to what extent the left-behind lox-p sites would influence translation.
- **Would a deletion collection for the closely related species *S. paradoxus* be of value?** It was pointed out that it would certainly be a more time-consuming knockout procedure because of less efficiency in homologous recombination. It was also not clear to some of the participants in what way such a collection would be worth doing. One argument put forward for its construction was in evolutionary studies on functional importance, and thus selection strength, that will be dependent on the genetic context. There were rather mixed views on this issue.
- **Is a Humanized (or in other ways substituted) yeast collection worth doing?** The humanized yeast is a collection that many still found to be of great interest. One problem in the constructions of these would be to have a good collection of cDNA clones. This is usually referred to as an orfeome collection. Marc Vidal at Harvard is in the process of making such a collection (at least 14,000 clones at the moment) but no-one knew if there are other similar large scale projects ongoing. At Harvard the Gateway system for sub-cloning is used, but none of the participants had any experience with this system. If genes from some parasite were to be introduced it was put forward by Karl-Dieter Entian that Trypanosoma would be a good first choice.
- **Should there be systematic and centralized storage of natural variants?** Of course the answer to this question is yes, but it was debated to what extent the 35 strains of *S. cerevisiae* that is now being sequenced will cover the majority of SNPs for this species. In other words, be a good representation of *S. cerevisiae*, or if more strains have to be included/sequenced. Ben Hitz strongly believed that this strain collection would only cover a minor part of the genetic variation. It was also pointed out that at SGD they have in the order of 600 genes isolated from *S. cerevisiae* strains that are not to be found in the sequenced S288c background (the normal laboratory strain). Thus, it is to be expected that also many of the natural isolates would contain new genes and not only SNP variants.

Key-issues from the in vivo phenotyping session:

- **How do we best enhance phenotype space?** Several times during the workshop the need for a better understanding of yeast ecology was addressed. It was believed that more knowledge on this topic would ensure a more intelligent selection of growth conditions to screen for genotype-phenotype links under relevant conditions. *S. cerevisiae* has been isolated from oak sap, so should we chemically analyse oak sap to find factors that would be a natural part of yeast selection? Or, is these biotopes just one of many where yeast has been selected in evolutionary time? The matter of feast and famine was also put forward and that we at present do not know if yeast has been selected under rather stable conditions or in a highly fluctuating environment - shifting cells between rapid growth and starvation. Phenotype space might be endless, but how do we maximize the environmental relevance to ensure the best possible coverage of gene function? These issues were debated. It was concluded that this is certainly an highly important issue and more information along these lines has to be gathered.
- **Should we use invasive or non-invasive techniques?** It was put forward by Uros Petrovic that transcriptional profiles on all deletion mutants should be made. This would of course be highly valuable data. However, some of the problems with invasive analysis (breaking cells open to be able to analyse the content) was discussed. It was debated if transcriptional profiles in that respect should be the premier data type, or if collecting information on proteom and the metabolom has equal weight. It was concluded that this is of course relevant activities, but that there is a strong need for standardisation in extract procedures etc since they will influence the end results. Maybe some strategies could be adopted in proteomics and metabolomics from the procedures for standardisation in the DNA microarray community?
- **How to handle cell-to-cell variation in microscopic studies?** This is certainly a hard issue since the quantitative value for individual cells can vary quite a bit. At this stage no-one knew exactly the under-laying factor for this but agreed on that this should be a prioritised issue.
- **How many features to extract from the microscopic images?** Procedures have been developed to extract hundreds of cellular features form images. It was debated to what extent many of these features would be of biological relevance. Another related problem is of course how to define specific internal structures. One solution to this could be to look not at the steady state situation but during dynamic changes since this will enable better classification if certain cellular structures are related.

Key-issues from the phenome bioinformatics session:

- **How to incorporate quantitative phenotypes in the ontology vocabulary?** Text-based description of phenotypes usually rely on classifying something as either or. It was put forward by Eurie Hong that one way to handle quantitative phenotypes is to have a general set of descriptors that one can link to each feature. However it was put forward that the problem is usually not including the numbers as such in the phentoypic description but to put the numbers/descriptors in some kind of context. The researcher generating the data should realize that the data themselves are usually not self-explanatory. The PROPHECY database provides quantitative phenotypic information at different levels. It was suggested that one informative display could be to visually show the data for a specific mutant in relation to the whole set of genome-wide data. It was concluded that this appear a good general suggestion that should be tested for many types of data.
- **How much experimental information should be provided at the centralized database?** Phenotypes are conditional. This means that in principle each and every phenotype should come with a rich description of conditions and procedures in the used assay. However, this is in many instances too much information for the general user of a database. It was discussed in what way the experimental details can be incorporated. Many put forward that there should be straightforward ways/links to get to the original publication.
- **Should experimental data be available at centralized databases before publication?** It was put forward by Olle Nerman that it would be valuable if one could have a system like the one practiced in Physics where data is deposited prior to publication. The question was raised

if that should in that case be the raw data or some processed form of the data? It was also discussed to what extent the now generated data will be of value in 20 years time.

- **Is federated databases the solution to data updating?** It was clearly stated by Graham Kemp that a federated database architecture does solve the updating problem. However, different architectures have different trade-offs. In the case of federated database systems, care must be taken to ensure that unspecific data requests do not place an unacceptable load on the databases participating in the fedaration. The data integration needs of the community should be examined carefully in deciding whether a centralised repository, or a federated solution, or some other architecture, provides the best solution for current needs.

Follow-up activites

The workshop ended with a discussion regarding different avenues for future joint activities and how to raise funding for these. The convenor had put together a compilation of the current EC calls that he could see would be potentially of interest to the workshop participants taken from “Health” and “Food” of the FP7 announcement. All participants agreed on that currently there were no apparent call(s) where this group of people could jointly apply. However, some calls were identified to be of value for specific subgroups and discussions continued during lunch about future collaborations.

The ESF representative Constantinos Doukas presented the different ways that ESF could support this Phenomics initiative in the future. A strong consensus in the group was that it would be valuable to meet on a more regular basis to continue discussions along the routes developed during this ESF workshop. It was decided by the group to jointly apply for a Research Network Programme from ESF with dead line for application late October 2007. The network programme would allow funding of annual workshops and educational efforts, like summer schools, in phenomics or in its subtopics. The ESF funded network programme could run for up to 5 years and would also encompass fellowship exchange which could be an excellent way of strengthening this European network and to enable joint collaboration on specific projects. The network programme could also open up for other groups in Europe to join this now established “working group” in Phenomics. The Network Programme application will be coordinated by the convenor Anders Blomberg.

4. Final programme

Friday 2 March 2007

Wallenberg Conference Hall, Medical Hill

12:00 – 13:00 *Welcome and lunch*

13:00 **Anders Blomberg**
Welcome

Words from our Sponsors

Constantinos Doukas (ESF Standing Committee for Life, Earth and Environmental Sciences)
Presentation of the European Science Foundation (ESF)

Stefan Hohmann (Science Faculty Scientific Advisory Board, GU)
Phenomics and the strategic plan for the Science Faculty

Olle Nerman (Chalmers Bioscience initiative and the Stochastic and Mathemacial Modelling Centre)
Phenomics and the strategic plan for Bioscience at Chalmers

Session I: Yeast Strain Collections

Dorit Schuller
Genetic variability of Saccharomyces cerevisiae natural isolates

Gianni Liti
A set of wild Saccharomyces strains for population genomics and phenomics studies

Anders Blomberg
Strains carrying temperature sensitive alleles of essential genes

Karl-Dieter Entian
Substituted Yeast Collections

Per Sunnerhagen
A deletion collection for Schizosaccharomyces pombe

Jörg Hauf
Handling and Distributing Strain Collections

17:00 *Transportation back to hotel*

18:30 *Dinner*

Pub with informal discussions

Saturday 3 March 2007

Venue Saturday – Ågrenska villan (close to hotel)

09.00 **Session II: In Vivo High-Troughput Phenotyping**

Carls Singer
High-density Agar Array Phenotyping

Steve Oliver
Expanding phenotype space by the use of controllable promoters and natural subtrates

Jonas Warringer
Microcultivation for High-Resolution Phenotyping

Achim Kohler
Vibrational spectroscopy for high-throughput phenotyping of cells: possibilities and limitations

Olle Nerman / Peter Gennemark
Statistical Stability and Experimental Precision of High-Throughput Phenotyping

Lunch

Sepp Kohlwein
High-Throughput Microscopy / Microscopy Automation

Mats Kvarnström
Image analysis for Quantitative Microscopy

Uros Petrovic
Combining genetic interaction and gene expression data

Break for coffee/tea

15.00 **Session III: Phenome Bioinformatics**

Robert Stevens
Phenotype Ontology - The Ontological Classification of a Genome's Protein Phosphatases

Eurie Hong
Phenome Databasing at SGD
How can one use software for the effective loading and integration of

Luciano Fernandez-Ricaud
PROPHECY – Phenotypic Profiling of Characterstics in Yeast

Balasz Papp
Predicting the outcome of genetic and environmental perturbations in yeast metabolism

Blaz Zupan
Epistasis and Inference of Genetic Network Topology

19:00 *Dinner*

Some social event

Sunday 4 March 2007

Venue Sunday – Ågrenska villan (close to hotel)

09.00 **Final "Round Table Discussion" on the three sessions**

Discussion about Future Applications / Actions

"The most clear outcome of the workshop is to establish a European Working Group on Yeast Phenomics. This working group will have as its first aim to apply for funding from relevant agencies. It will also discuss how to initiate training in high-throughput phenotyping at the European and international level."

Constantinos Doukas

Possible future ESF calls / funding

Presentation of EU calls related to phenotyping

Summing up and final discussion about future joint actions

Workshop closure

12:00 *Lunch and departure*

5. Statistical information on participants

There were in total 27 participants (including 4 local organisers) for the workshop whose country of origin included 10 European on 1 non-European countries.

	male	female	
Austria	1		
England	4		
Finland	1		
Germany	3		
Greece	1		
Hungary	1		
Norway	1		
Portugal		1	
Slovenia	2		
Sweden	9	1	(including 4 local organisers)
USA	1	1	
Total:	24	3	

6. Final list of Participants

Convenor:

1. **Anders BLOMBERG**
   Lundberg Laboratory
   Dept of Cell and Molecular Biology
   Göteborg University
   Medicinaregatan 9c
   413 90 Göteborg
   Sweden
   Tel: +46 317332589
   Email: anders.blomberg@gmm.gu.se

ESF Representative:

2. **Constantinos DOUKAS**
   Department of History, Geology and Palaeontology
   University of Athens
   Panepistimiopolis
   Post Office Zagrafou
   15784 Athens
   Greece
   Email: cdoukas@geol.uoa.gr

Local Organisers:

3. **Luciano FERNANDEZ-RICAUD**
   Department of Cell and Molecular Biology
   Göteborg University
   Medicinaregatan 9C
   405 30 Göteborg
   Sweden
   Email: gmmlf@gmm.gu.se

4. **Peter GENNEMARK**
   Department of Mathematical Statistics
   Chalmer University of Technology
   Chalmers campus
   412 96 Göteborg
   Sweden
   Email: peterg@cs.chalmers.se

5. **Katarina LOGG**
   Chalmers University of Technology
   Condensed Matter Physics
   Dept. of Applied Physics
   Chalmers campus
   412 96 Göteorg
   Sweden
   Email: katlogg@fy.chalmers.se

6. **Jonas WARRINGER**
   Göteborg University
   Medicinaregatan 9C
   405 30 Göteborg
   Sweden
   Email: jonas.warringer@gmm.gu.se

Participants:

7. **Karl-Dieter ENTIAN**
   Institute for Molecular Biosciences
   Max-von-Laue-Str. 9
   60438 Frankfurt
   Germany
   Email: entian@bio.uni-frankfurt.de

8. **Thomas GYLLING**
   Oy Growth Curves Ab Ltd.
   Ilmalankuja 2 M 17
   00240 Helsinki
   Finland
   Email: Sales@bioscreen.fi

9. **Joerg HAUF**
   SRD - Scientific Research and Development GmbH
   Köhlerweg 20
   61440 Oberursel
   Germany
   Email: hauf@srd-biotec.de

10. **Ben HITZ**
    Medical Center
    School of Medicine
    Stanford University
    CA 94305-5120 Stanford
    United States
    Email: hitz@genome.stanford.edu

11. **Stefan HOHMANN**
    Göteborg University
    Box 462
    405 30 Göteborg
    Sweden
    Email: hohmann@gmm.gu.se

12. **Eurie HONG**
    Medical Center
    Stanford University
    CA 94305-5120 Stanford
    United States
    Email: eurie@genome.stanford.edu

13. **Graham KEMP**
    Computer Science and Engineering
    Chalmers University of Technology
    Chalmers campus
    412 96 Göteborg
    Sweden
    Email: kemp@cs.chalmers.se

14. **Peter KOETTER**
    Institute for Molecular Biosciences
    Max-von-Laue-Str. 9
    60438 Frankfurt
    Germany
    Email: Koetter@em.uni-frankfurt.de

15. **Sepp KOHLWEIN**
University of Graz
Schubertstrasse 1S
8010 Graz
Austria
Email: sepp.kohlwein@uni-graz.at

16. **Mats KVARNSTRÖM**
Fraunhofer-Chalmers Research Centre
Chalmers Science Park
412 88 Göteborg
Sweden
Email: mats.kvarnstrom@fcc.chalmers.se

17. **Gianni LITI**
Queen's Medical Centre
University of Nottingham
Nottingham NG7 2UH
United Kingdom
Email: gianni.liti@nottingham.ac.uk

18. **Achim KOHLER**
Centre for Biospectroscopy and Data Modelling
Matforsk, Norwegian Food Research Institute and
CIGENE, Department of Mathematical Sciences
and Technology
Norwegian University of Life Sciences
1430 Ås, Norway
Email: achim.kohler@matforsk.no

19. **Olle NERMAN**
Chalmers University of Technology
Chalmers campus
412 96 Göteborg
Sweden
Email: nerman@math.chalmers.se

20. **Steve OLIVER**
University of Manchester
Oxford Road
Manchester M13 9PT
United Kingdom
Email: steve.oliver@manchester.ac.uk

21. **Balazs PAPP**
University of Manchester
Oxford Road
Manchester M13 9PT
United Kingdom
Email: Balazs.Papp@manchester.ac.uk

22. **Uros PETROVIC**
Biochemistry and Molecular Biology
Jozef Stefan InstituteJamova 39Jamova 39
1000 Ljubljana
Slovenia
Email: uros.petrovic@ijs.si

23. **Dorit SCHULLER**
Universidade do Minho
Campus de Gualtar
4710-057 Braga
Portugal
Email: dschuller@bio.uminho.pt

24. **Carl SINGER**
Singer Instruments Co.Ltd.
Roadwater
Somerset TA23 0RE
United Kingdom
Email: yeast@singerinst.co.uk

25. **Robert STEVENS**
University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom
Email: Robert.stevens@manchester.ac.uk

26. **Per SUNNERHAGEN**
Göteborg University
Medicinaregatan 9C
405 30 Göteborg
Sweden
Email: Per.Sunnerhagen@molbio.gu.se

27. **Blaz ZUPAN**
Faculty of Computer and Information Science
University of Ljubljana
Trzaska 25
1000 Ljubljana
Slovenia
Email: blaz.zupan@fri.uni-lj.si

Invited persons that could not come to the workshop:

Charlie BOONE, Terrence Donnelly Centre for Cellular and Biomolecular Research, Toronto, Canada
Mathias PETER, ETH Zürich, Switzerland
Stig OMHOLT, CIGENE, Norwegian University of Life Sciences, Aas, Norway
Nicholas LUSCOMBE, EMBL-European Bioinformatics Institute, Cambridge, England
Chris GLASBEY, Edingburg University, Scotland
Marc VIDAL, Harvard Medical School, USA
Joris WINDERICKX Katholic University Leuven, Belgium