

## 1. The organization and data content of GermOnline 4.0

GermOnline 4.0 is a cross-species database portal focusing on high-throughput expression data relevant for germline development, the meiotic cell cycle and mitosis in healthy versus malignant cells. The database is available for free via [www.germonline.org/](http://www.germonline.org/). Microarray data are annotated using the Multiomics Information Management and Annotation System (MIMAS 3.0).

The project is now organized as a genomics gateway which provides access to information produced with 3'-UTR high-density oligonucleotide microarrays (GeneChips), the output of genome-wide protein-DNA binding assays (ChIP-Chip) and protein-protein interaction studies in budding yeast. Furthermore, 3'-UTR GeneChips expression data are available for rat, mouse and human.

In addition to the database that holds 3'-UTR GeneChip data the gateway provides access to the prototype *Saccharomyces Genomics Viewer* (SGV). It was developed to facilitate the interpretation of highly complex data – shown as false-color heatmaps – from experiments with high-density oligonucleotide tiling microarrays that cover the entire budding yeast genome.

## 2. Simple query

The current release of GermOnline provides a graphical display of selected studies of *S. cerevisiae*, *R. norvegicus*, *M. musculus* and *H. sapiens*. First select the species in the popup menu and then use query terms such as the gene symbol (Mouse Spo11), a database identification number (MGI:1349669) or the Ensembl (ENSMUSG00000005883) identifier.

Search

- Arabidopsis thaliana
- Caenorhabditis elegans
- Danio rerio
- Drosophila melanogaster
- Homo sapiens
- Macaca mulatta
- Mus musculus
- Pan troglodytes
- Rattus norvegicus
- Saccharomyces cerevisiae**
- Schizosaccharomyces pombe

**Fig. 1: A simple query.** Select the species, enter the query term into the search field and press the Go button.

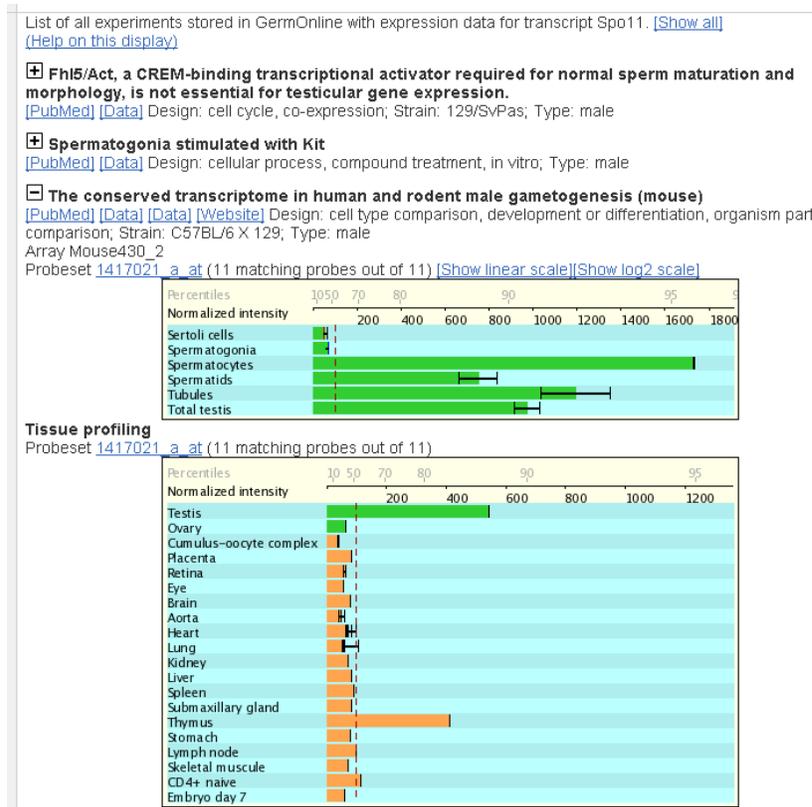
Depending on the species the report page contains basic information about the gene and its transcripts including a graphical display of the exon/intron structure, information on external reference IDs associated with the transcript, Gene Ontology annotation and data on protein domains. To access microarray data scroll down to the section to see referenced studies. Select all datasets for simultaneous display or open and collapse individual displays.

List of all experiments stored in GermOnline with expression data for transcript Spo11. [\[Show all\]](#)  
[\[Help on this display\]](#)

- Fhl5/Act, a CREM-binding transcriptional activator required for normal sperm maturation and morphology, is not essential for testicular gene expression.**  
[\[PubMed\]](#) [\[Data\]](#) Design: cell cycle, co-expression; Strain: 129/SvPas; Type: male
- Spermatogonia stimulated with Kit**  
[\[PubMed\]](#) [\[Data\]](#) Design: cellular process, compound treatment, in vitro; Type: male
- The conserved transcriptome in human and rodent male gametogenesis (mouse)**  
[\[PubMed\]](#) [\[Data\]](#) [\[Data\]](#) [\[Website\]](#) Design: cell type comparison, development or differentiation, organism part comparison; Strain: C57BL/6 X 129; Type: male
- Identifying genes important for spermatogonial stem cell self-renewal and survival**  
[\[PubMed\]](#) [\[Data\]](#) Design: is expressed, time series; Strain: C57BL/6 X 129; Type: male
- Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated**  
[\[PubMed\]](#) [\[Data\]](#) Design: development or differentiation, unknown experiment; Strain: CD-1; Type: male
- Postmeiotic sex chromatin in the male germline of mice**  
[\[PubMed\]](#) [\[Data\]](#) Design: development or differentiation; Strain: C57BL/6 X 129; Type: male
- Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination**  
[\[PubMed\]](#) [\[Data\]](#) Design: development or differentiation, is expressed, sex; Types: female, male
- Profiling gene expression during the differentiation and development of the murine embryonic gonad**  
[\[PubMed\]](#) [\[Data\]](#) Design: development or differentiation, sex; Strain: C57BL/6 X 129; Types: female, male
- Transcript profiling during preimplantation mouse development**  
[\[PubMed\]](#) [\[Data\]](#) Design: array platform variation, cell type comparison, development or differentiation, organism part comparison; Strain: CF-1 x B6D2F1/J; Types: female, unknown
- Major molecular differences between mammalian sexes are involved in drug metabolism and renal function (mouse)**  
[\[PubMed\]](#) [\[Data\]](#) Design: organism part comparison, sex; Strain: Swiss Webster; Types: female, male
- Essential role of Plzf in maintenance of spermatogonial stem cells**  
[\[PubMed\]](#) [\[Data\]](#) Design: genetic modification; Strain: C57BL/6J; Type: male
- Defining the spermatogonial stem cell (mouse)**  
[\[PubMed\]](#) [\[Data\]](#) Design: organism part comparison, species; Strain: C57BL/6; Type: male
- A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets**  
[\[PubMed\]](#) [\[Data\]](#) Design: development or differentiation, organism part comparison, time series; Strain: C57BL/6 X 129; Type: male

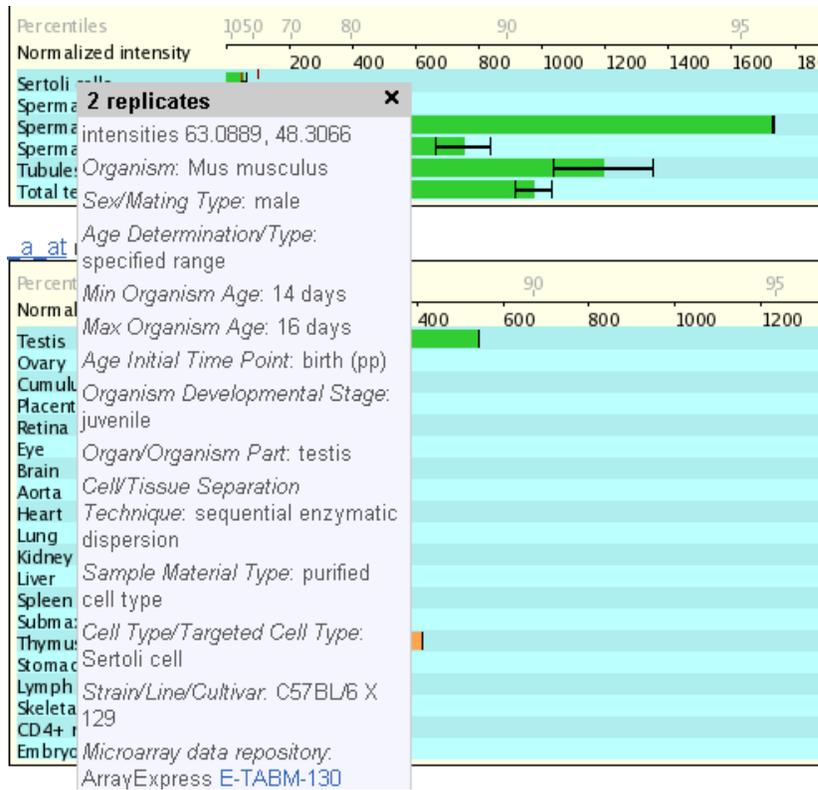
**Fig. 2: Expression data display.** The title of selected studies is shown and external [Data] links to PubMed and raw data files stored at public repositories such as GEO (NCBI) or ArrayExpress (EBI) are given. Additional information on experimental design and strain backgrounds is displayed.

To select an individual display click on the [+] button and view linear or log2-transformed data by following the links [Show linear scale] or [Show log2 scale].



**Fig. 3: Histogram display of a typical study.** Samples are given and expression data are indicated in the horizontal scale. The percentile expression signals are indicated. Testicular samples are shown in green, somatic controls are shown in orange.

To know more about the samples and the expression values click on the sample name. You will see a popup window that contains information which was entered into MIMAS during the experimental annotation process. It includes array signal intensities for replicates and information about the sample source and the preparation procedure used to prepare it.



**Fig. 4: Sample annotation.** Click on the sample name to call up information entered via MIMAS. A link at the bottom leads the user to the raw data source.

### 3. Moving from one species to another within the report page

A useful feature within the report pages enables you to access corresponding data via ortholog links. The information is provided by Ensembl's own ortholog identification project and by OMA (Orthologous Matrix Browser, <http://www.cbrg.ethz.ch/oma>). We have used these data to link report pages from orthologs represented in GermOnline with each other. This new feature enables users to move from one locus (and its expression data) to its (likely) ortholog within one mouse click.

Orthologue Prediction		The following gene(s) have been identified as putative orthologues:	
Species	Type Orthology	Gene identifier	
<i>Caenorhabditis elegans</i>	1-to-1 non-ortholog <a href="#">OMA113948</a>	<a href="#">T05E11.4</a> (spo-11) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	spo-11 encodes an ortholog of Spo11p in <i>S. cerevisiae</i> that is required for meiotic DNA recombination, and (in conjunction with CHK-2 and MRE-11) for the localization of RAD-51 to foci in late zygotene/early pachytene nuclei. Like its yeast ortholog, SPO-11 is related to a subunit of archaeobacterial topoisomerase VI, and probably causes double-stranded DNA breaks through a topoisomerase-like transesterase mechanism. However, unlike Spo11p, SPO-11 is dispensable for homologous synapsis and formation of the synaptonemal complex. SPO-11 is active even in syp-1 mutants where its double-stranded breaks cannot be correctly processed. spo-11 is transcribed in the germline. [Source: WormBase]
<i>Danio rerio</i>	1-to-1 non-ortholog <a href="#">OMA049234</a>	<a href="#">ENSDARG00000013837</a> (zgc:77876) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	hypothetical protein LOC402982 [Source: RefSeq_peptide (NP_991245)]
<i>Drosophila melanogaster</i>	1-to-1 non-ortholog <a href="#">OMA072105</a>	<a href="#">FBgn0002716</a> (mei-W68) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	No description
<i>Homo sapiens</i>	1-to-1 non-ortholog <a href="#">OMA049234</a>	<a href="#">ENSG00000054796</a> (SPO11) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	Meiotic recombination protein SPO11 (Cancer/testis antigen 35) (CT35). [Source: Uniprot/SWISSPROT, Acc:Q9Y5K1]
<i>Macaca mulatta</i>	1-to-1 non-ortholog <a href="#">OMA049234</a>	<a href="#">ENSMMUG00000000248</a> (SPO11) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	Meiotic recombination protein SPO11 (Cancer/testis antigen 35) (CT35). [Source: Uniprot/SWISSPROT, Acc:Q9Y5K1]
<i>Pan troglodytes</i>	1-to-1 non-ortholog <a href="#">OMA049234</a>	<a href="#">ENSPTRG00000013657</a> (SPO11) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	Meiotic recombination protein SPO11 (Cancer/testis antigen 35) (CT35). [Source: Uniprot/SWISSPROT, Acc:Q9Y5K1]
<i>Rattus norvegicus</i>	1-to-1 undetermined	<a href="#">ENSRNOG00000006732</a> (NP_001102434.1) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	sporulation protein, meiosis-specific, SPO11 homolog [Source: RefSeq_peptide (NP_001102434)]
<i>Saccharomyces cerevisiae</i>	1-to-1 undetermined	<a href="#">YHL022C</a> (SPO11) <a href="#">[MultiContigView]</a> <a href="#">[Align]</a>	Meiosis-specific protein SPO11 (Sporulation-specific protein 11). [Source: Uniprot/SWISSPROT, Acc:P23179]

**Fig. 5: Cross-linking report pages via orthologs.** Information on protein similarity is used to link report pages. Users can click on the gene identifier to call up the corresponding report page from another species.

#### 4. Protein-protein interaction data

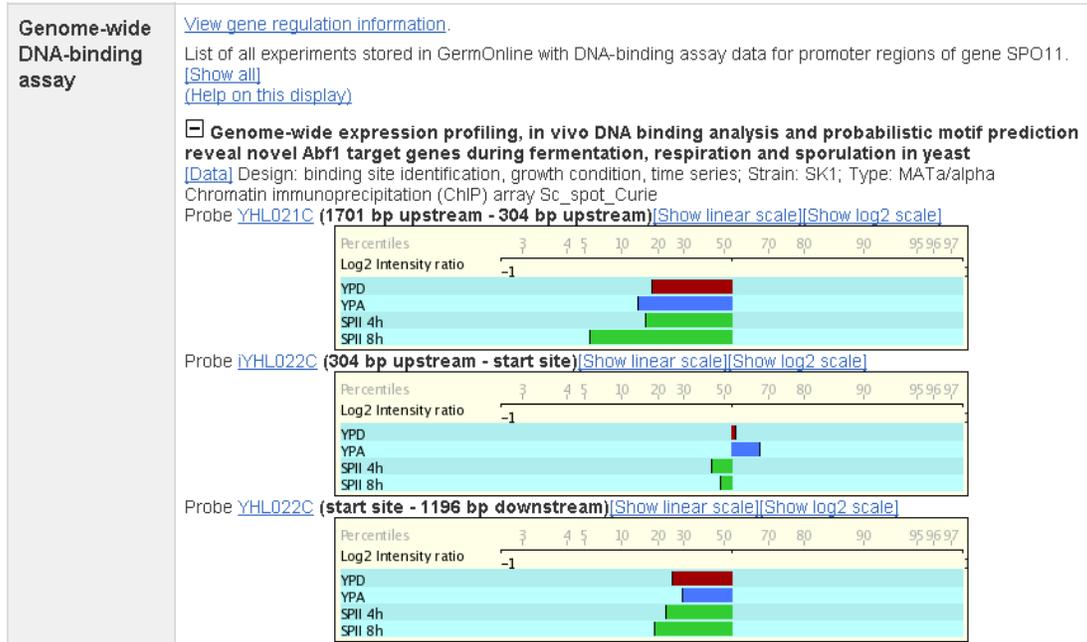
Whenever available, protein network information is organized into three classes that comprise nine categories: genetic (Dosage Lethality, Phenotypic Enhancement, Synthetic Lethality), *in vivo* (Affinity Capture-MS, Affinity Capture-Western, Co-purification, Two-hybrid) and *in vitro* (Biochemical Activity, Reconstituted Complex). Users can select each category individually or choose the option to display all interactions at once. This approach keeps the display simple and easy to read (Figure 2). For each interacting factor the gene name and annotation information as well as the references including a hyperlink to the respective PubMed entries are displayed. The gene names are hyperlinked with their corresponding report pages within GermOnline. Interaction data sources are provided by the BioGRID, and IntAct projects.

Interactions	Interaction data from <a href="#">BioGRID</a> and <a href="#">IntAct</a>	
<input type="checkbox"/>	Genetic: Dosage Rescue (2 interactions)	
<input type="checkbox"/>	Genetic: Phenotypic Enhancement (13 interactions)	
<input type="checkbox"/>	Genetic: Phenotypic Suppression (9 interactions)	
<input type="checkbox"/>	Genetic: Synthetic Lethality (1 interaction)	
<input type="checkbox"/>	Genetic: Synthetic Rescue (10 interactions)	
<input type="checkbox"/>	In vitro: Biochemical Activity (2 interactions)	
<input checked="" type="checkbox"/>	In vivo: 2 hybrid (3 interactions)	
<input type="checkbox"/>	In vivo: Affinity Capture-Western (2 interactions)	
<input type="checkbox"/>	In vivo: Co-localization (1 interaction)	
<input type="checkbox"/>	In vivo: two hybrid array (1 interaction)	
<input type="checkbox"/>	All (44 interactions)	
	YGL213C <a href="#">YGL213C</a> (SK18) Antiviral protein SK18 (Superkiller protein 8). [Source:Uniprot/SWISSPROT;Acc:G02793]	Dosage Rescue: <a href="#">Arora C (2004)</a> 2 hybrid: <a href="#">Arora C (2004)</a> 2 hybrid: <a href="#">Cheng Z (2004)</a> 2 hybrid: <a href="#">Uetz et al. (2000)</a>
	YHR157W <a href="#">YHR157W</a> (REC104) Meiotic recombination protein REC104. [Source:Uniprot/SWISSPROT;Acc:P33323]	Dosage Rescue: <a href="#">Jiao K (2003)</a> 2 hybrid: <a href="#">Arora C (2004)</a> 2 hybrid: <a href="#">Maleki S (2007)</a>
	YLR329W <a href="#">YLR329W</a> (REC102) Meiotic recombination protein REC102. [Source:Uniprot/SWISSPROT;Acc:G02721]	Phenotypic Enhancement: <a href="#">Kee K (2002)</a> Synthetic Rescue: <a href="#">Jiao K (2003)</a> 2 hybrid: <a href="#">Maleki S (2007)</a> Affinity Capture-Western: <a href="#">Jiao K (2003)</a> Affinity Capture-Western: <a href="#">Kee K (2002)</a>

**Fig. 6: Protein network data.** Tick a box to call up the interacting proteins. The display provides a link to the corresponding report page in GermOnline and external PubMed references.

## 5. Protein-DNA binding data (*S. cerevisiae* only)

The database provides a display of Chromatin Immunoprecipitation and Chip (ChIP-Chip) data for budding yeast. As a prototype entry a study is included that identified the yeast target promoters of Abf1, a general DNA binding regulator in vegetatively growing versus sporulating cells. Note that the data were obtained with microarrays containing spotted PCR fragments covering all known yeast intergenic regions and open reading frames. Data are displayed on a log<sub>2</sub> scale with a full reference and a link to the data repository and the corresponding PubMed entry. To facilitate the interpretation of *in vivo* binding data that are show as fold-enrichment of a bound DNA fragment, the corresponding data obtained with upstream and downstream fragments are also displayed. The title of the corresponding study is displayed. Click on the [+] button to display the data. More information is available via the “Help on this display” link. Note that corresponding microarray expression data are available for this study in the microarray data section of the report page.



**Fig. 7: In vivo binding of Abf1.** Color-coded bars show log intensity ratios for Abf1 binding in YPD (red), YPA (blue) and SPII (4h and 8h into sporulation, green).

## 6. External links

Useful external information sources grouped into Expression-, Protein and Literature categories are indicated in a species-specific manner within the report page.

a Quick Links	
Expression	<b>GNF SymAtlas:</b> <a href="#">[View expression pattern]</a>
Protein	<b>BioGRID:</b> <a href="#">[List biological interactions]</a>
	<b>UniProtKB:</b> <a href="#">[View protein annotation]</a>
	<b>Swiss-Model:</b> <a href="#">[View protein structure]</a>
Literature	<b>PubMed:</b> <a href="#">[List articles]</a>
	<b>Google Scholar:</b> All articles <a href="#">[All species]</a> <a href="#">[This species only]</a> Recent articles <a href="#">[All species]</a> <a href="#">[This species only]</a>

b Quick Links	
Gene	<b>SGD:</b> <a href="#">[Gene information]</a>
	<b>GeneDB:</b> <a href="#">[Gene information]</a>
	<b>AGD:</b> <a href="#">[View Ashbya gossypii homologs]</a>
	<b>YGOB:</b> <a href="#">[Gene order conservation]</a>
Protein	<b>BioGRID:</b> <a href="#">[List biological interactions]</a>
	<b>PeptideAtlas:</b> <a href="#">[Peptide identification]</a>
	<b>UniProtKB:</b> <a href="#">[View protein annotation]</a>
	<b>Swiss-Model:</b> <a href="#">[View protein structure]</a>
Literature	<b>PubMed:</b> <a href="#">[List articles]</a>
	<b>Google Scholar:</b> All articles <a href="#">[All species]</a> <a href="#">[This species only]</a> Recent articles <a href="#">[All species]</a> <a href="#">[This species only]</a>

**Fig. 8: External links.** Panel a shows an example for mouse Spo11. Panel b gives external links available for yeast Spo11.

## 7. Advanced query

To call up lists of genes that correspond to publications first select *M. musculus* or *S. cerevisiae* to access PubMed abstracts.

Search

**Query functional genomics & expression data**

**Mus musculus**  
The core meiotic transcriptome in mammals: 9280 mouse genes differentially expressed in germ cells.

**Saccharomyces cerevisiae**  
A wide range of functional genomics and expression experiments related to the cell cycle and gametogenesis.

**Fig. 8: Advanced query form.** Select the species to get to the abstracts of the publications relevant for gene lists of interest.

Next, click on the button “Query form”, select “GermOnline” as the database and choose a dataset (Mmusculus or Scerevisiae). Subsequently, tick appropriate filter criteria (ID List, High-quality genes or High-quality transcripts) and identify gene lists (it is highly recommended to read the papers before you pull out gene lists) using criteria such as “Genes

New Count Results URL XML Perl Help

Dataset

**Filters**

High-quality genes : Only  
High-quality transcripts : Only  
Genes Differentially Expressed in Testis (DET) : ME (meiotic cluster)

**Attributes**

Gene stable id  
Gene primary acc  
Gene display label

Please restrict your query using criteria below

FILTERS

ID LIST

High-quality genes  Only  Excluded

High-quality transcripts  Only  Excluded

The conserved transcriptome in human and rodent male gametogenesis [Abstract](#)

Genes Differentially Expressed in Testis (DET)

Conserved expression pattern in mouse and rat (CDET)

Number of tissues (in addition to testis) where expression is detected

min

max

**Fig. 9: Advanced query form filter criteria.** Select filter criteria and experimental features to define gene lists.

Differentially Expressed in Testis (DET)”, “Conserved and Differentially Expressed in Testis (CDET)” or the number and type of somatic tissues where expression is detected.

Next, click on the “Attributes” link to display a form that enables users to define which information they want to export.

**Fig. 9: Attributes for data export definition.** Select transcript features, annotation data and expression data to be exported.

Finally, click the “Results” button in the top navigation bar and define the file type you want to export the data using the popup menus. Press the “Go” button. Use appropriate software to save the file.

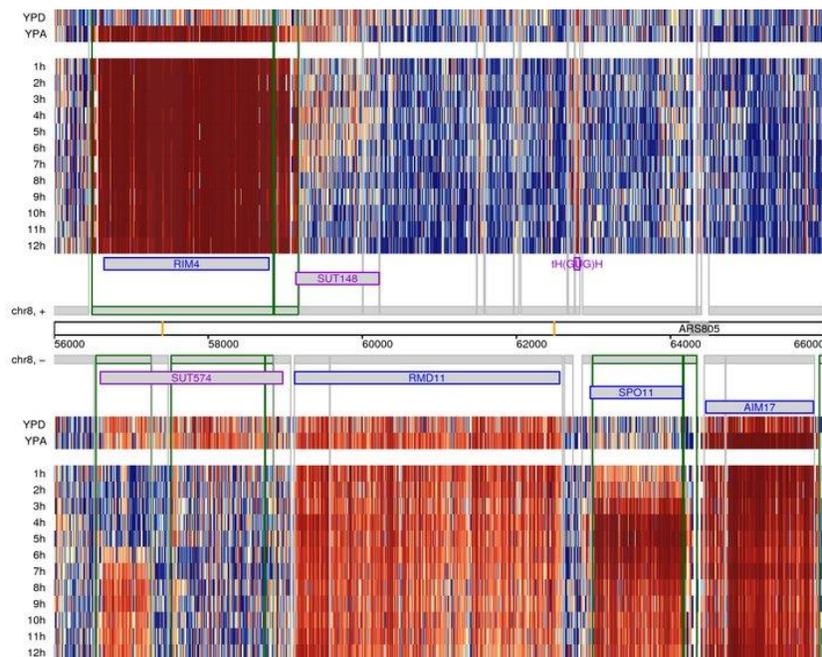
Gene stable id	Gene primary acc	Gene display label
ENSMUSG000000063889	MGI:88495	Crem
ENSMUSG000000063889	MGI:88495	Crem
ENSMUSG000000063889	MGI:88495	Crem
ENSMUSG00000050945	MGI:2444919	Zfp436
ENSMUSG00000039540	MGI:1918151	4921524L21Rik
ENSMUSG00000039540	MGI:1918151	4921524L21Rik
ENSMUSG00000061802	MGI:1922184	Armc4
ENSMUSG00000050996	MGI:1347086	Cetn1
ENSMUSG00000036743	MGI:1920927	Prsmab
ENSMUSG00000014232	MGI:1924029	Clusp1

**Fig. 10: Results and export file type.** Popup menus enable the user to define to export file type.

## 8. Using the Saccharomyces Genomics Viewer

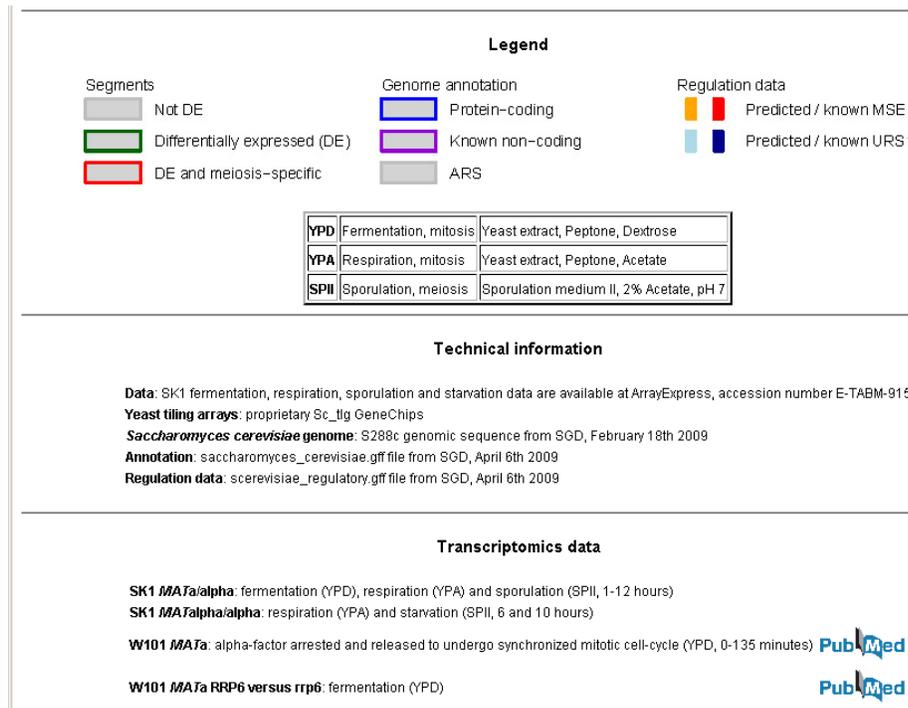
The graphical display implemented in GermOnline is not suitable for displaying multi-sample tiling array datasets. We have therefore decided to create a separate prototype solution for budding yeast tiling data that is integrated into the GermOnline gateway. To access the Saccharomyces Genomics Viewer (SGV) click on the logo in the top left corner of the GermOnline welcome page.

In the query form first define the type of query (coding transcript, non-coding transcript, segment ID, chromosomal coordinate, ARS) and then type a yeast systematic name (YHR124W), standard name (*NDT80*), SGD ID (SGD:S000001166) for coding transcripts, or the names of non-coding transcripts (SUT587, MUT686), segment IDs (8.+356585.358433), chromosomal coordinates (8:356585) or ARS numbers (ARS1215). The report page is organized into four different datasets obtained with different yeast strains.



**Fig. 9: SGV data display.** An image of normalized tiling microarray expression data at the single oligonucleotide probe level is shown via a false-color heatmap (red and blue indicate strong and weak signals, respectively). The samples are indicated for fermenting (YPD), respiring (YPA) and sporulating cells (1-12h in SPII medium). Protein-coding genes, ncRNAs and ARS elements are shown as blue, purple and gray boxes, respectively. The chromosome number and coordinates are given.

A detailed legend at the bottom of the report page explains the display for each study and provides PubMed links to the studies whenever they are available. It also includes log<sub>2</sub> data scales.



**Fig. 10: SGV legend.** Detailed information on the report page content is given. Published studies are linked to PubMed entries.

## 9. Making a contribution to GermOnline

You are invited to contact Michael Primig (michael.primig@inserm.fr) if you wish to contribute raw CEL data files. Please note that the database only displays data obtained with high-density oligonucleotide microarrays (GeneChips). Be aware that GermOnline is not meant to serve as a repository, so any submission to the database does not substitute for raw data submission to a certified public repository such as NCBI's GEO or EBI's ArrayExpress which is mandatory according to the MIAME standard.

## **10. Citing the database update paper**

Our database provides a free service to the community. It has become increasingly difficult to obtain sustainable funding for database projects but being able to publish an update paper in a well-cited journal does help. However, editors will (among other criteria) verify the citation rate of previously published papers. So, if you use GermOnline and you find it helpful, please cite the most recent update paper as indicated in the welcome page's footer. Citing the paper will facilitates our attempts to publish papers in journals with a respectable impact factor – this is important for our efforts to secure funding for GermOnline which is after all a service for the community and not a research project.