Development and Application of a Database for Research on **G**enes **i**nvolved in **S**enescence, **A**poptosis and **O**xidative Stress

Edith Hofer

Doctoral Thesis

Graz University of Technology
Institute for Genomics and Bioinformatics
Petersgasse 14, 8010 Graz, Austria

Graz, June 2011

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ……. ……..
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……. ……..
date (signature)

Abstract

A major goal of ageing research is the identification of genes which play a role in the process of ageing. This task is facilitated by the microarray technology, which enables genome-wide gene expression profiling. Bioinformatics methods are used to analyse microarray gene expression profiles and to determine gene functions and gene regulators by combining expression data with other types of biological data, e.g. annotations or genomic data.

In order to manage ageing-related gene data GiSAO.db has been developed. GiSAO.db is a web-based database system for storage and retrieval of information concerning **G**enes **i**nvolved in **S**enescence, **A**poptosis and **O**xidative stress. The database stores microarray gene expression data, annotation data, ortholog data and data of follow-up experiments. The web application is based on the Java EE framework and offers an intuitive user interface which simplifies data input and access. A mature authentication and authorization system is used to administrate user access rights.

Integrative data analysis has been performed on selected gene expression profiles stored in GiSAO.db to investigate mechanisms of ageing.

Gene expression patterns of various human ageing models have been compared to determine genes and pathways which are associated with oxidative stress induced senescence or other forms of senescence.

To identify regulators, functions and interactions of genes which are involved in UVB induced senescence, promoter analysis, pathway analysis and correlation analysis were conducted on corresponding gene expression profiles.

In conclusion, GiSAO.db assists researchers in the management of various types of data which are used to study genetic processes of ageing. Integrative data analysis has been performed to determine ageing-associated genes as well as pathways, potential regulators and interactions of these genes.

Keywords: ageing, microarrays, database, integrative data analysis

Publications

This thesis was based on following publications, as well as upon unpublished work:

Papers

Laschober G, Ruli D, **Hofer E**, Muck C, Carmona-Gutierrez D, Ring J, Hutter E, Ruckenstuhl C, Micutkova L, Brunauer R, Jamnig A, Trimmel D, Herndler-Brandstetter D, Sampson N, Breitenbach M, Fröhlich KU, Grubeck-Loebenstein B, Berger P, Wieser M, Grillari-Voglauer R, Thallinger GG, Grillari J, Trajanoski Z, Madeo F, Lepperdinger G, Jansen-Dürr P. **Identification of evolutionarily conserved genetic regulators of cellular aging**. Aging Cell 2010, 9:1084-1097. PMID: 20883526

Hofer E, Laschober GT, Hackl M, Thallinger GG, Lepperdinger G, Grillari J, Jansen-Dürr P, Trajanoski Z. **GiSAO.db: a database for ageing research**. BMC Genomics 2011, 12:262. PMID: 21609420

Contents

1 Introduction . . . 1

1.1 Ageing . . . 1
1.2 Experimental methods . . . 2
1.3 Ageing databases . . . 3
1.4 Integrative data analysis . . . 3
1.5 Objectives . . . 5

2 Results . . . 6

2.1 Development of GiSAO.db . . . 6
2.2 Integrative data analysis . . . 16
2.2.1 Comparison of gene expression profiles of various ageing models . . . 16
2.2.2 Analysis of UVB induced senescent gene expression profiles . . . 23

3 Discussion . . . 37

3.1 Development of GiSAO.db . . . 37
3.2 Integrative data analysis . . . 40
3.2.1 Comparison of gene expression profiles of various ageing models . . . 40
3.2.2 Analysis of UVB induced senescent gene expression profiles . . . 41
3.3 Conclusion . . . 43

4 Methods . . . 44

4.1 Software technology . . . 44
4.1.1 Java Enterprise Edition (Java EE) . . . 44
4.1.2 Relational database management system (RDBMS) . . . 52
4.1.3 Model driven architecture (MDA) . . . 54
4.1.4 Web page presentation . . . 56
4.1.5 Authentication and authorization . . . 57
4.1.6 Development tools . . . 58
4.2 Integrative data analysis . . . 60
4.2.1 Statistical analysis . . . 60
4.2.2 Pathway analysis . . . 61

4.2.3 Cluster analysis 62
4.2.4 Promoter analysis 64
4.2.5 Gene co-expression network 68

A Bibliography 70

B Glossary 84

C Acknowledgments 88

D Supplementary Information 89

E Publications 135

Chapter 1

Introduction

1.1 Ageing

Life expectancy in industrial countries has increased steadily since the middle of the 19th century [Wilmoth, 2000]. For this reason, ageing-related diseases, such as cancer or heart diseases, are nowadays the most common causes of early death.

In order to find cures for these diseases and prevent them as well as remedy premature symptoms of ageing like skin changes, abdominal obesity or decreased memory, it is necessary to investigate the causes and mechanisms of ageing.

Biological ageing is the result of accumulated cell damage [Adams and White, 2004]. If the damage remains permanently and is not corrected by DNA repair systems, apoptosis or cellular senescence occur. Apoptosis is the process of programmed cell death. Apoptotic cells die in a controlled way and are absorbed by phagotic cells afterwards [Reed, 2000]. Cellular senescence is a mechanism which prevents cells from proliferating, meaning that they are not able to divide anymore [Campisi and d'Adda di Fagagna, 2007]. The cells arrest their growth permanently and do not complete the cell cycle, although their metabolism stays active.

Both, apoptosis and cellular senescence, are assumed to lead to phenotypes which are typical for ageing and ageing associated diseases [Campisi, 2003].

Aside from somatic mutations and mitochondrial changes, one of the perpetrators of cellular damage is oxidative stress [Adams and White, 2004] [Harman, 1956]. Oxidative stress is an unbalanced proportion of oxidants to antioxidants [Sies, 1997]. Oxidants, e.g. free radicals or peroxides, harm cell components, whereas antioxidants, like antioxidant enzymes and vitamins, protect a cell against damage [Sies, 1993]. Endogenous causes for oxidative stress are aerobic metabolic

activities or infections [Ames et al., 1993]. Additionally, there are environmental oxidative stress causes such as different types of radiation (e.g. UVB light, ultrasound and microwave), cigarette smoke and dietary elements (e.g. salt or iron).
Although it is known that oxidative stress triggers premature senescence and apoptosis by causing damage to cellular components and activating corresponding signaling pathways (Fig. 1.1), the causal relationships as well as the underlying processes have not yet been revealed [Finkel and Holbrook, 2000].

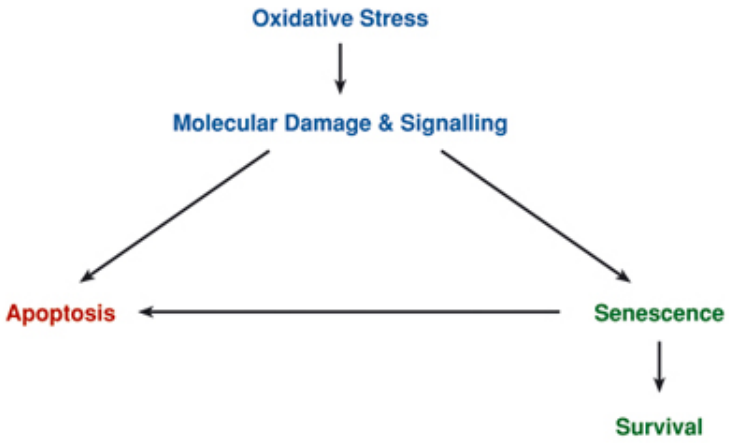

Figure 1.1: The relationship between oxidative stress, senescence and apoptosis (from [NRN, 2010]).

In order to understand mechanisms of ageing, it is essential to identify and study genes and miRNAs which are involved in the ageing process. MiRNAs are short non-protein-coding RNA molecules, which regulate the expression of genes. One way to determine these genes and miRNAs is by studying their expression patterns, as it is well known that expression patterns of organisms change with age [Weinert and Timiras, 2003][Bates et al., 2009][Grillari and Grillari-Voglauer, 2010][Chen et al., 2010].

1.2 Experimental methods

Gene expression is the process in which first DNA is transcribed to RNA, which is then transformed to mRNA and translated to proteins [Lodish et al., 2007]. Therefore, by measuring the amount of mRNA, it can be determined which genes are active in the examined cell.
For the determination of all active genes in a cell in a certain state (e.g. diseased or healthy) at a certain time point, it is necessary to measure the gene expression of the whole genome at once.

Microarrays are well suited for this task as they are a high-throughput method for the determination of the expression of tens of thousands of genes in parallel [Schena, 2000].
Results of microarray experiments can be validated by applying low-throughput methods for measuring the expression of genes, e.g. qPCR or Northern blot [Chuaqui et al., 2002]. Additionally, microarray experiments can be confirmed with protein measuring methods like Western blot.
A miRNA is transcribed from DNA and prevents the translation of a specific target mRNA by binding to it [Lodish et al., 2007]. MiRNA expression can be determined by quantifying the miRNA amount using the techniques applied for genes, e.g. microarrays.

1.3 Ageing databases

Further processing and analysis of data from genes involved in ageing requires fast and simple access to the information thereby generated and results of the experiments. Databases are very well suited to the storage of such data as they provide an organized way of storing and managing large data quantities.
Several public databases containing ageing specific gene information are available. The *Human Aging Genomic Resources* (HAGR) provides data about genes associated with ageing in humans and in a large number of other species [de Magalhães et al., 2009a]. For each ageing-related gene, the promoter sequences, location, proteins and protein interactions, orthologs and literature references are listed.
The University of Washington hosts the *Aging Gene Database* which stores information about genes that have been examined for their role in ageing and in age-related neurological diseases [genesDB, 2010]. Function, orthologs and references of a gene can be retrieved from this database.
The Gene Aging Nexus (GAN) [Pan et al., 2007] and the *Atlas of Gene Expression in Mouse Aging Project* (AGEMAP) [Zahn et al., 2007] are the only repositories containing ageing-related gene expression data, though solely mouse data are stored in AGEMAP.
GAN is a repository for microarray data obtained from ageing experiments providing gene annotation, data visualization, differential expression analysis and co-expression analysis [Pan et al., 2007].

1.4 Integrative data analysis

Microarray experiments generate large amounts of data and in order to extract relevant information for biological interpretation, bioinformatics methods are used [Keith, 2008a]. Gene expres-

sion data are integrated with genomic data and annotations to determine and visualize genetic processes, gene functions and associations between genes. Usually, the analysis involves two steps. First, a small number of genes is selected, based on differential expression or clustering methods, which identify co-expressed genes. Next, different kinds of analyses are performed to obtain comprehensive knowledge about the genes of interest. Annotation data like Gene Ontology (GO) terms [Ashburner et al., 2000] provide known gene functions. Pathway analysis is used to retrieve information about the biological context of a gene which may help to determine its functionality. Transcription factor binding site prediction in promoter sequences of genes is a useful method to detect regulators of gene expression. Additionally, correlation networks visualize potential gene interactions [Friedman et al., 2000].

1.5 Objectives

The aim of this thesis is to develop and implement GiSAO.db, a web-based storage and retrieval system for ageing-related microarray gene expression data.
Additionally, datasets from several model systems should be integrated and analyzed in order to facilitate biological interpretation.

The specific objectives are:

- GiSAO.db
  - develop the database schema to enable consistent storage of gene and miRNA expression data obtained from microarray experiments
  - integrate annotation data, ortholog data and data of follow-up experiments
  - provide simple and intuitive functions for data access
  - implement data input, update and export functionalities
  - design a user friendly web interface.
- Integrative data analysis
  - import datasets
  - perform statistical analysis
  - carry out pathway analysis
  - conduct cluster analysis
  - perform promoter analysis
  - construct gene interaction networks

Chapter 2

Results

2.1 Development of GiSAO.db

GiSAO.db is a web-based database system for storing and retrieving data concerning **g**enes **i**nvolved in **s**enescence, **a**poptosis and **o**xidative stress.
The database contains normalized mRNA expression values obtained from Affymetrix [Affymetrix, 2010] microarray experiments and miRNA expression data from Exiqon [Exiqon, 2010] microarray experiments as well as gene annotation data, ortholog data and data of follow-up experiments (Fig. 2.1). These data can be accessed through various search functions and genes or miRNAs of interest can be collected in favourite lists. GiSAO.db further offers comprehensive functions for data upload, update and export. Additionally, the data are complemented by links to external databases, e.g. RefSeq [RefSeq, 2010], and integrated KEGG pathways [Ogata et al., 1999].

GiSAO.db is a Java EE 5 web application which is presently deployed on a JBoss application server, whereas its application data are persisted in an Oracle RDBMS.
The business logic of the web application was developed using the model driven architecture approach. Therefore, the base frame of the business tier was generated with the AndroMDA EJB3 cartridge. The application logic was implemented in session beans, whereas message driven beans were used for uploading data in order to increase the responsiveness of the application. The data model of GiSAO.db was developed using EJB3 entities. Java Server Pages were used together with the Struts 1 framework to manage the data presentation in the web tier. The layout of the web pages was designed using cascading style sheets (CSS) and Web 2.0 functionality has been established by integrating AJAX features using the script.aculo.us [script.aculo.us, 2010]

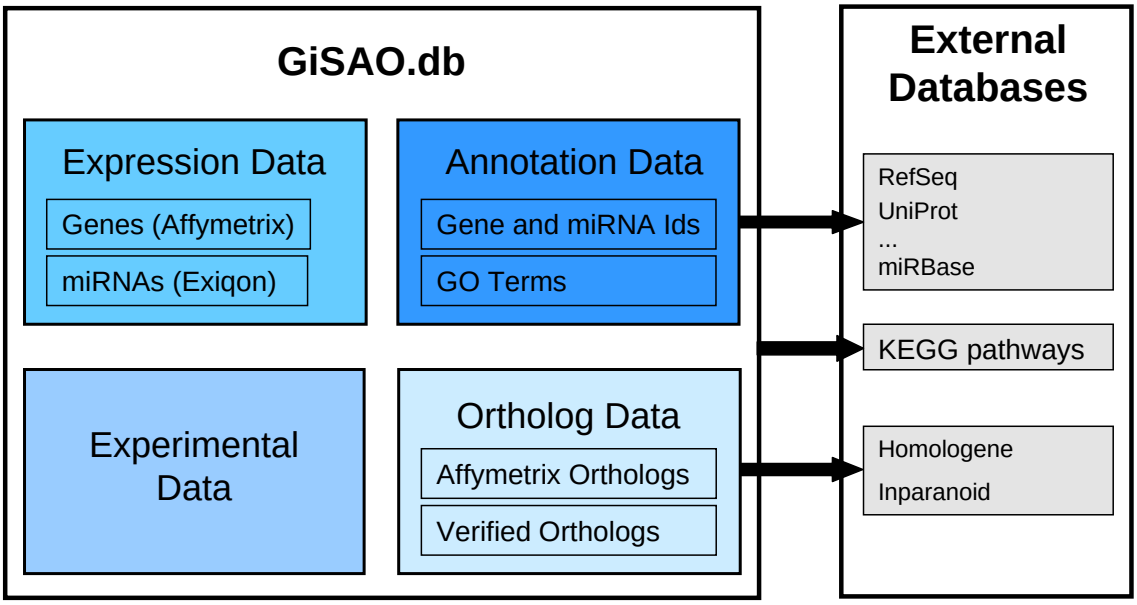

Figure 2.1: GiSAO.db stores four different types of data: gene and miRNA expression data obtained from either Affymetrix or Exiqon microarray experiments, annotations for genes and miRNAs including gene ontology (GO) terms [Ashburner et al., 2000], ortholog data provided by Affymetrix as well as manually entered verified ortholog data and experimental data of follow-up experiments. To amend the data, GiSAO.db also provides links to external databases and integrated KEGG pathways.

Javascript library. An integrated authentication and authorization system is used to manage user data and control data access.

The initial version of GiSAO.db has been deployed in 2007 [Pitzl, 2007] [Lepperdinger et al., 2008] and has been continuously developed further since then.

Database schema

The core of the GiSAO.db database is the table which represents a spot on a microarray (Fig. 2.2). As GiSAO.db is capable of storing data of two different kinds of microarrays - Affymetrix mRNA microarrays and Exiqon miRNA microarrays - a spot on the microarray is identified by the Affymetrix probe set Id or the Exiqon probe Id. The identifier of the microarray spot is linked to its corresponding array which is specified by name and type. As probe (set) Ids are not very meaningful and hard to memorize, annotation data consisting of gene or miRNA identifiers that match the mRNA or miRNA sequence(s) applied on a spot, are available for each spot on a microarray.

Moreover, expression values obtained from microarray experiments, which represent the main domain of the database, are linked to the corresponding microarray spot identifier. Other gene or

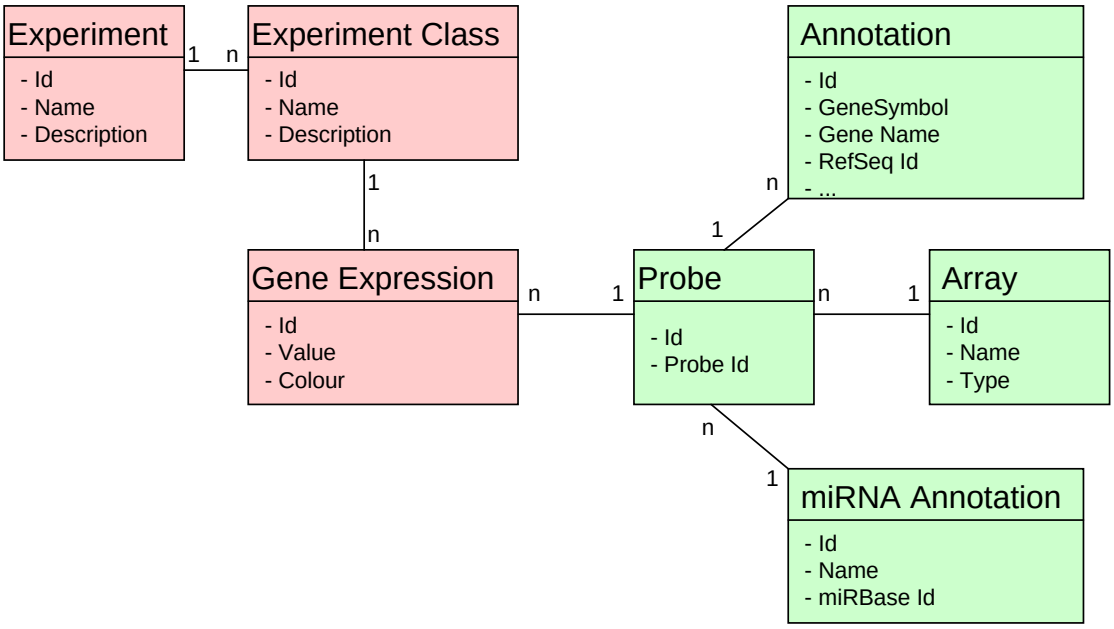

Figure 2.2: Database core schema for gene information storage. A spot on the microarray is identified by its probe (set) Id. The probe (set) Id is referenced to its associated microarray and annotation is provided as additional information so that the user can identify the gene(s) that match a probe. Gene and miRNA specific data, such as expression data, are linked to the corresponding microarray probe (set) Ids.

miRNA specific data in GiSAO.db, such as experimental data, favourite lists or ortholog data, are also stored based on probe (set) Ids.

Expression data

GiSAO.db supports storage and retrieval of normalized expression data obtained from microarray experiments. The database is capable of storing gene expression values of Affymetrix one-colour microarrays and miRNA expression values of Exiqon two-colour microarrays. For the latter type, the expression values of each channel are stored separately in the database, but only the expression ratios of the two channels are shown as output. The result of a microarray experiment is displayed using colour coded boxes representing expression values and ratios (Fig. 2.3). Expression values of highly regulated genes are displayed in a red colour tone, while expression values of lowly regulated genes are presented in a yellow tone. Expression ratios of up-regulated miRNAs are represented by reddish boxes and expression ratios of down-regulated miRNAs are displayed in green shades. The colour presentation facilitates the determination of highly expressed genes, up- or down-regulated miRNAs and the comparison of expression values and ratios of different microarrays. Expression values and ratios may be output in logarithmic or decimal scale. Further-

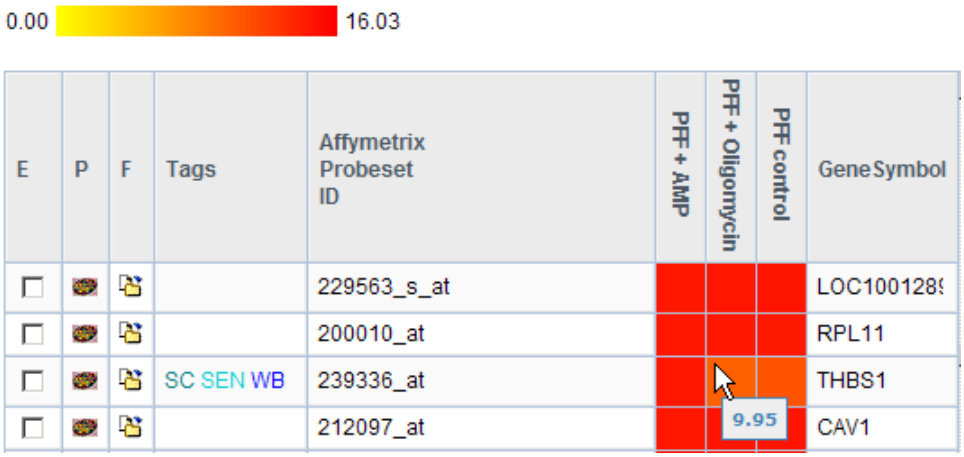

Figure 2.3: Representation of gene expression values. To facilitate the identification of prominent expression values and comparison between values of a gene of different arrays, GiSAO.db displays a gene's corresponding expression values by colour coded boxes. High expression values are displayed in reddish colour tones while low expression values are displayed in yellow shades.

more, a cut-off can be defined to display only genes or miRNAs with expression values or ratios exceeding the specified threshold.

Annotation data

GiSAO.db provides annotation data, which are essentially gene identifiers and gene ontology terms, to add a meaning to the identifiers of each spot on a microarray. For each Affymetrix probe set Id the corresponding gene name, gene symbol, RefSeq Id [RefSeq, 2010], EntrezGene Id [EntrezGene, 2010], UniProt Id [UniProt, 2010], UniGene Id [UniGene, 2011], SGD Id [SGD, 2010], MGI Id [MGI, 2011], FlyBase Id [FlyBase, 2011], RGD Id [RGD, 2010] and AGI Id [TAIR, 2011] are available in the database. These identifiers are obtained together with GO terms from Affymetrix for each spot of each microarray chip.
For each Exiqon probe Id of a miRNA microarray, the miRNA name and the miRBase Id [Griffiths-Jones et al., 2006] are stored as annotations in GiSAO.db.

Ortholog data

In order to enhance the data, GiSAO.db also handles the storage of two types of ortholog data:

- **Affymetrix orthologs** are lists containing the probe sets of an Affymetrix array and the corresponding orthologous probe sets on Affymetrix arrays of other organisms. These orthologs are provided by Affymetrix and have been integrated into GiSAO.db.
- **Verified orthologs** are orthologs which have been discovered by the user e.g. by literature

research or in external databases. They are entered manually including the reference or source.

To search for orthologs, the gene symbol or Affymetrix probe set Id as well as the organism have to be specified. As a result, corresponding Affymetrix orthologs and verified orthologs are listed (Fig. 2.4). Additionally, links to the respective orthologs in the external ortholog databases HomoloGene [HomoloGene, 2010] and InParanoid [O'Brien et al., 2005] are provided.

(a)

Affymetrix Ortholog Data

Homo sapiens (human)		Caenorhabditis elegans (nematode)		Drosophila melanogaster (fruit fly)		Mus musculus (mouse)	
Human Genome U133 Plus 2.0 Array		C. elegans Genome Array		Drosophila_2 Array		Mouse Genome 430 2.0 Array	
Gene Symbol	Array Id	Gene Symbol	Array Id	Gene Symbol	Array Id	Gene Symbol	Array Id
NPC1	202679_at					Npc1	1423086_at
NPC1	202679_at			NPC1	1640496_at		
NPC1	202679_at						
NPC1	202679_at	ncr-1	187969_s_at				
NPC1	202679_at	WBGene00003561	187969_s_at				
NPC1	217584_at					Npc1	1423086_at
NPC1	217584_at			NPC1	1640496_at		
NPC1	217584_at						
NPC1	217584_at	ncr-1	187969_s_at				
NPC1	217584_at	WBGene00003561	187969_s_at				

(b)

Verified Ortholog Data

Organism	Gene Symbol	ID	ID Type	Comment	Reference/Source	User
Homo sapiens (human)	NPC1	NPC1	Gene Symbol			
Saccharomyces cerevisiae	NCR1	S000005927	SGD accession number		Inparanoid	lepperdinger

(c)

Ortholog Databases

Database	Gene ID	
HomoloGene	NPC1	
InParanoid	ENSG00000141458	

Figure 2.4: Ortholog search result. First, orthologs provided by Affymetrix are listed **(a)**. Beneath manually entered orthologs are displayed **(b)** and finally corresponding links to HomoloGene and InParanoid are provided **(c)**.

Experimental data

Based on gene expression patterns obtained from microarray experiments, genes or miRNAs of interest are identified. On these selected genes or miRNAs, follow-up experiments are performed to confirm the microarray results or carry out subsequent investigations. GiSAO.db is capable of storage and retrieval of data about those experiments. An experiment is specified by properties such as type (e.g. qPCR, Western blot), classification (e.g. senescence, inflammation), organism and cell type. All genes that were investigated are assigned to the respective experiment. In case of antibody or primer usage, they can be linked to the corresponding gene. Furthermore, it is possible to enter references or upload files, e.g. experiment protocols or result files, and attach them to one of the following entities: an experiment, a gene or miRNA, an antibody or a primer. To query experimental data a flexible search function which accepts organism, experiment classification and experiment type as parameters has been implemented. The result is a list of genes on which experiments with the specified properties were carried out.
To get a quick overview of the kinds of experiments performed on a gene or miRNA, tags, which are essentially shortcuts describing the experiment type, the experiment classification and the organism, have been introduced to GiSAO.db. On the web site, these tags are located next to the Affymetrix probe set Id or Exiqon Id in various gene lists, e.g. favourite gene lists (Fig. 2.5). Thereby, the user sees at first glance all the types, classifications and organisms of the experi-

P	C	O	T	ID	Gene Symbol	Comment	User	Time Stamp	
			HS OXS NB	1552957_at	LOC200383	---	hofer	2009-01-29 11:30:58.0	
			HS SEN QP	65521_at	C11orf65	---	hofer	2009-01-29 11:30:59.0	

Figure 2.5: Tags displayed in gene lists next to Affymetrix probe set Ids. Dark green tags denote organisms, turquoise tags are experiment classifications, and dark blue tags denote experiment types. For the Affymetrix probe set Id *65521_at*, experiments with organism *Homo sapiens (HS)*, experiment classification *senescence (SEN)* and experiment type *qPCR (QP)* were performed.

ments that were performed on the gene. Additionally, tags are links, that lead to the experiments corresponding to the experiment type, experiment classification or organism.

Additional application features

Favourite lists representing genes and miRNAs of special interest for researchers can be composed and stored in GiSAO.db. Additional information can be added to these lists and they can be compared to check for common entries. Furthermore, GiSAO.db offers a comprehensive search function which takes gene or miRNA identifiers as input and returns all available data about the specified gene or miRNA including expression data, annotation data, experimental data

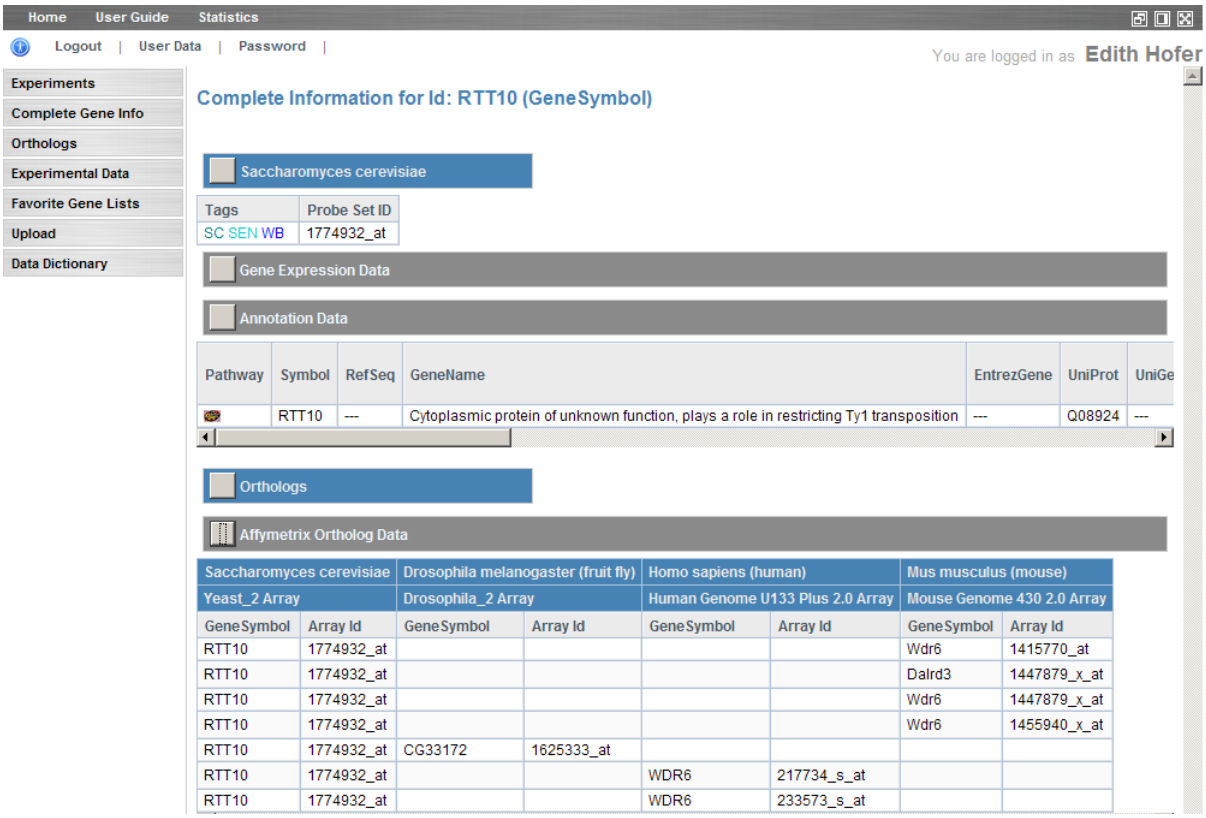

Figure 2.6: Search result for the queried gene RTT10. All the information stored in the database, i.e. tags, gene expression data, annotation data and ortholog data are displayed.

and favourite lists (Fig. 2.6). In order to enhance the data stored in GiSAO.db, links to external databases are available. Gene identifiers are referenced to their respective databases such as RefSeq, Entrez Gene or UniProt and miRNAs are linked to the miRBase database. Moreover, GiSAO.db uses a web service provided by the Kyoto Encyclopedia of Genes and Genomes (KEGG) to access pathways.

Data input and export

GiSAO.db offers two different possibilities to input data into the database: manual data input through web forms and data upload from files. Data of verified orthologs and follow-up experiments can be entered using specific upload forms. However, experimental data may also be uploaded from files just as expression data, annotation data, Affymetrix ortholog data and data of favourite gene lists. These files are parsed and the data are then stored in the corresponding database tables. As Affymetrix and Exiqon regularly update the annotation data of their microarray chips and Affymetrix frequently updates ortholog data, GiSAO.db provides update functions

to keep the data in the database up-to-date.
Usually, several files describing the experiments and their outcome, such as protocols, are associated with experimental data. To facilitate the storage of these files a Java applet which offers simultaneous file upload for multiple files has been incorporated into the application.
As GiSAO.db provides an upload feedback report, the user is able to check whether the upload was successful and view error messages in case something went wrong.
An export mechanism which enables further processing of data from the database in external tools has been included into GiSAO.db. Lists of expression values, favourite genes and orthologs can be written to plain text files, pdf files or files in comma separated values (csv) format.

Data dictionary

Several data input forms of the GiSAO.db web application contain fields which require an input that is usually a value from a limited list of options. To avoid inconsistencies and spare the user from typing the input, selection fields are used. In order to still allow free-text input, the concept of a data dictionary has been introduced. A data dictionary offers a flexible, simplified and fast way to enter data in a web application. It extends the functionality of selection fields that often have hard coded or statically predefined options by allowing the user to add new options. Furthermore, it compensates the disadvantage that data entered via a text field may cause redundant database entries due to different spellings or spelling mistakes. The implemented data dictionary stores its data in a database table where each entry specifies the value which can be selected and its domain. It is possible to disable a value so that it is no longer provided as an option in a selection. In GiSAO.db, the input of orthologs or experimental data is facilitated by data dictionary selection fields for experiment types, experiment classifications or gene identifier types.

User interface

Due to the application of cascading style sheets, a consistent colour scheme could be introduced to the web pages of the GiSAO.db web application. Moreover, AJAX controls and effects were integrated into the web pages using the script.aculo.us framework to improve the usability of the user interface.

- **Autocomplete fields** offer suggestions according to the characters the user has typed into the text field. The suggestions are fetched from the database and displayed in a list below the text field without a page reload. Autocomplete fields accelerate the process of filling out forms and help to prevent spelling mistakes. In the GiSAO.db web application autocomplete fields are mainly applied for input of gene identifiers (Fig. 2.7).

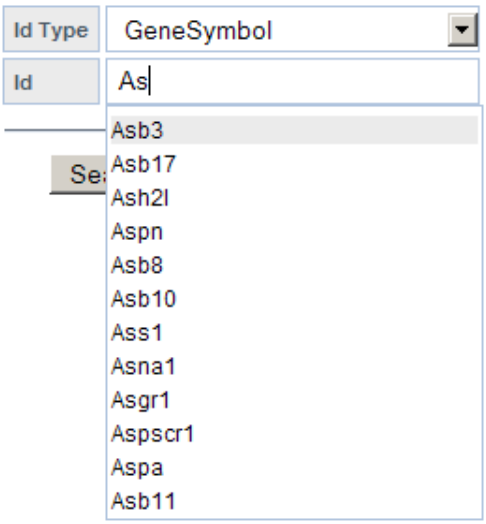

Figure 2.7: Autocomplete field. Gene symbols starting with the characters entered by the user are suggested.

- **Sortable lists** allow the user to arrange a list in the desired order. Such a list is used to define the order of experiment classes on the page where gene expression values are displayed.
- **Drag-and-drop fields** are used to move elements from one list to another list. In the GiSAO.db web application drag-and-drop fields facilitate the comparison of favourite gene lists.
- **Visual effects** make text appear on a web page after a mouse click without reloading the whole page. Information, that supports the user e.g. in filling out forms, is hidden when the web page is loaded and appears or disappears after a click on the designated icon (Fig. 2.8).

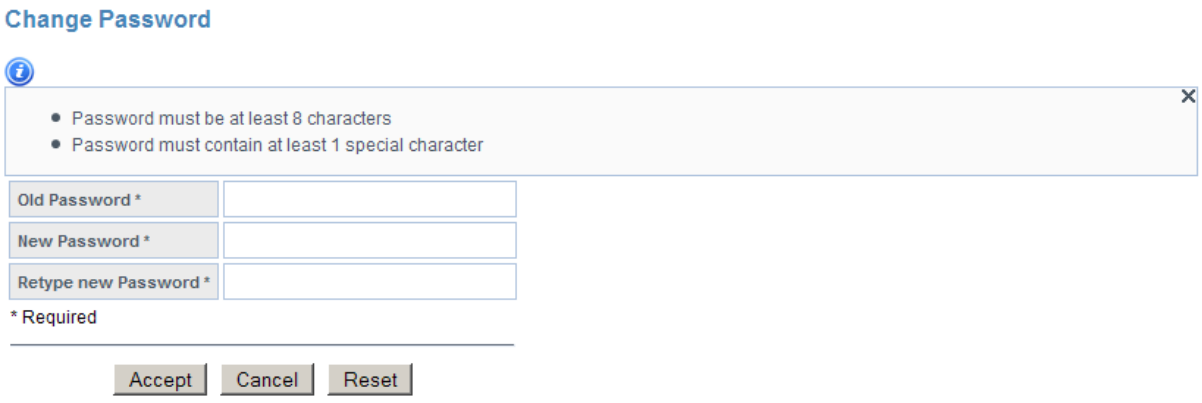

Figure 2.8: AJAX visual effect. Explanatory information appears after a click on the icon.

User access

Three user types with different access rights are defined in the authentication and authorization system for GiSAO.db:

- **Guest** can view all the data stored in the database but is not allowed to add, edit or delete any data.
- **User** can add data of follow-up experiments as well as verfied ortholog data and may import genes to favourite gene lists. Furthermore, the user is allowed to edit and delete her own data or data belonging to a member of the user's institute.
- **Administrator** has rights to edit and delete all data stored in the database and can enter and upload all possible types of data into the database including expression data and annotation data.

2.2 Integrative data analysis

The GiSAO.db database stores gene expression profiles obtained by Affymetrix microarray experiments which explore cellular ageing. In order to identify genes which play a role in ageing and determine their function, integrative data analysis was performed based on these gene expression profiles.
Statistical analysis, pathway analysis and clustering was carried out using the expression patterns of different ageing models (Section 2.2.1). Additionally, pathway and promoter analysis as well as the determination of gene interactions was conducted on gene expression patterns reflecting UVB induced senescence (Section 2.2.2).

2.2.1 Comparison of gene expression profiles of various ageing models

Expression profiles of selected ageing experiments performed on human cell models were compared to determine conserved gene expression patterns and pathways that are linked to senescence.
Two different types of experiments were investigated: premature senescence invoked by oxidative stress and other senescence models.
In the oxidative stress experiments, human umbilical vein endothelial cells (HUVEC), primary prostate stromal cells (PrSC), renal proximal tubular epithelial cells (RPTEC), human mesenchymal stem cells (MSC) and $CD8^+$ T-lymphocytes were exposed to pro-oxidant conditions. Some cells were grown in an environment with low oxygen content whereas other cells were treated with sublethal doses of tert-Butyl hydroperoxide (tBHP).
The senescence experiments comprise in vitro experiments studying senescence which is not caused by oxidative stress and in vivo samples of senescent cells. In vitro samples were taken from senescent cultures of HUVEC and RPTEC as well as human diploid fibroblasts (HDF) with induced mitochondrial dysfunction.
The in vivo samples were $CD28^-$ and $CD28^+$ T- lymphocytes, MSCs from young and old donors, and prostate stromal fibroblasts (PPF) treated with transforming growth factor beta (TGF-β) to induce cellular degeneration.

Statistical analysis

Both experiment types, oxidative stress and senescence, comprised, in all, 47 independent samples. From these samples 17 oxidative stress experiment pairs and 18 senescence experiment pairs were assembled. For each probe set and each sample pair, the fold change between the

samples was computed as the quotient of the expression values. Probe sets with a fold change >+/- 1.5 were defined as differentially expressed (DE). Using a value counting method, it was determined how often each probe set was differentially expressed in the oxidative stress experimental group and in the senescence experimental group.

Statistical analysis was performed with R [R Development Core Team, 2007] and the qvalue package [Storey and Tibshirani, 2003]. As demonstrated by de Magalhães et al. [de Magalhães et al., 2009b], the p-value for each probe set was calculated with the cumulative binomial distribution (CBD) (Eqn. 2.1).

$$P_i(X \geq k) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j} \tag{2.1}$$

The CBD gives the probability P for a probe set i to be as often or more often differentially expressed than the times k it was actually differentially expressed in n experiments.

For each probe set i, the probability p that i was differentially expressed in an experimental group, was defined as the average of differentially expressed probe sets in an experimental group divided by the number of all probe sets on the microarrray.

According to these calculations, the probability of a probe set to be differentially expressed with oxidative stress induced senescence was 0.11116383 and the probability of a probe set to be differentially expressed with senescence was 0.11895646.

The used Affymetrix microarray chips contain 54675 probe sets and the p-value was computed for each probe set. To correct the p-values for multiple testing, a method to control the false discovery rate (FDR) was used. The FDR is the proportion of expected false positives within all probe sets that were declared significant [Storey et al., 2004]. In order to calculate the q-value for each probe set, Storey's FDR approach with the bootstrapping method was used [Storey and Tibshirani, 2003]. Since the obtained p-values are very low, the q-values were made more accurate for small p-values [Storey, 2002].

For both experimental groups, the q-value cut-off was set to 0.005, resulting in p-value cut-offs of 0.0002235 for the oxidative stress experimental group and 0.000578618 for the senescence experimental group. In the oxidative stress group, the q-values of 587 probe sets which were differentially expressed in at least 9 out of 17 experiments were below the cut-off (Appendix D). 1423 probe sets with a q-value less than the cut-off were differentially expressed in at least 9 out of 18 senescence experiments (Appendix D).

Comparing the two probe set lists, 180 common probe sets which were differentially expressed in at least 9 experiments in each experimental group, were identified (Appendix D).

Pathway analysis

Pathway analysis was performed with PathwayExplorer [Mlecnik et al., 2005] separately for significant probe sets in the oxidative stress experimental group, significant probe sets in the senescence experimental group and significant probe sets in both groups. The expression ratios of the probe sets were mapped to all available KEGG [Kanehisa and Goto, 2000], GenMapp [Dahlquist et al., 2002] and BioCarta [BioCarta, 2010] pathways.

As a result, a list containing the corresponding pathways, sorted by the number of probe sets that could be mapped on the pathway, was created for each probe set list. The PathwayExplorer tool provides a p-value calculated using Fisher's exact test for each pathway. Fisher's exact test determined if the proportion of the probe sets mapped on the pathway among the probe sets in the selected probe set list was significantly larger than the proportion of probe sets that were mapped on the pathway among those probe sets on the microarray which did not belong to the selected probe sets [Mlecnik et al., 2005].

For those probe sets which were significant in both experimental groups, p53 signalling, cell cycle, cell proliferation and apoptosis pathways were ranked at the top (Tab. 2.1). Seven probe sets out

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
p53 signaling pathway	7	215	3.26	$<$0.001	KEGG
Cell cycle	7	258	2.71	$<$0.001	KEGG
cell proliferation	7	659	1.06	0.006	GenMapp
apoptosis	7	748	0.94	0.012	GenMapp
cytokinesis	5	304	1.64	0.003	GenMapp
chromosome	5	476	1.05	0.021	GenMapp
lipid biosynthesis	4	404	0.99	0.045	GenMapp
Cholesterol Biosynthesis	3	30	10	$<$0.001	GenMapp
Nitrogen metabolism	3	65	4.62	0.001	KEGG
Fructose and mannose metabolism	3	122	2.46	0.008	KEGG
Apoptosis	3	125	2.4	0.008	KEGG
G1 to S cell cycle control	3	159	1.89	0.016	GenMapp
Pyrimidine metabolism	3	227	1.32	0.04	KEGG
IL 18 Signaling Pathway	2	9	22.22	$<$0.001	BioCarta
Synthesis and degradation of ketone bodies	2	18	11.11	0.002	KEGG

Table 2.1: Top 15 pathways of 180 probe sets which were significant in both experimental groups. The Ids of these probe sets were mapped to KEGG, GenMapp and BioCarta pathways and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

of 180 were mapped on each of these pathways. Moreover, the pathways for cholesterol biosynthesis, nitrogen metabolism and IL 18 signaling were significantly over-represented in this probe set list with a p-value $<=$ 0.001.

In the oxidative stress experimental group, the apoptosis pathway was the top ranked pathway

with 17 mapped probe sets out of 587 probe sets (Tab 2.2). Pathways for cell cycle, chromosome,

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
apoptosis	17	748	2.27	0.003	GenMapp
Cell cycle	15	258	5.81	<0.001	KEGG
chromosome	14	476	2.94	0.001	GenMapp
lipid biosynthesis	13	404	3.22	<0.001	GenMapp
enzyme activator activity	12	548	2.19	0.017	GenMapp
cytokinesis	10	304	3.29	0.002	GenMapp
mRNA metabolism	10	459	2.18	0.028	GenMapp
Pyrimidine metabolism	9	227	3.96	0.001	KEGG
fatty acid metabolism	9	279	3.23	0.003	GenMapp
ATP-dependent helicase activity	9	353	2.55	0.015	GenMapp
mRNA processing	9	427	2.11	0.043	GenMapp
G1 to S cell cycle control	8	159	5.03	<0.001	GenMapp
D4-GDI Signaling Pathway	7	41	17.07	<0.001	BioCarta
DNA Replication	7	76	9.21	<0.001	GenMapp
p53 signaling pathway	7	215	3.26	0.009	KEGG

Table 2.2: Top 15 pathways of 587 probe sets which were significant in the oxidative stress experimental group. The Ids of these probe sets were mapped to KEGG, GenMapp and BioCarta pathways and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

lipid biosynthesis and enzyme activator activity were listed beneath. Other pathways that were significantly over-represented in this list of probe sets with a p-value <= 0.001 were pyrimidine metabolism, G1 to S cell cycle control, D4-GDI signaling and DNA replication.

In the list of probe sets which were significant in the senescence experimental group, again the apoptosis pathway with 39 mapped probe sets out of 1423 probe sets, followed by pathways for MAPK signaling, extracellular matrix, p53 signaling and cell proliferation were the top pathways (Tab. 2.3). Additionally, enzyme inhibitor activity, cell cycle, cytokine-cytokine receptor interaction, cell surface receptor linked signal transduction and cell growth pathways were significantly over-represented with a p-value <= 0.001.

Pathways for apoptosis, cell cycle and p53 signalling were among the top 15 pathways in all three probe set lists. Cholesterol biosynthesis, nitrogen metabolism, fructose and mannose metabolism, IL 18 signaling, synthesis and degradation of ketone bodies pathways solely occur in the top pathways of significant probe sets in both experimental groups. Pathways that appear only in the top 15 pathways of significant probe sets of the oxidative stress experimental group are mRNA metabolism, fatty acid metabolism, ATP-independent helicase activity, mRNA processing, D4-GDI signaling and DNA replication. Senescence specific top ranked pathways are MAPK signaling, extracellular matrix, focal adhesion, enzyme inhibitor, cytokine-cytokine receptor interaction, cell surface receptor linked signal transduction, cell growth, Wnt signaling pathway and

myometrial relaxation and contraction.

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
apoptosis	39	748	5.21	<0.001	GenMapp
MAPK signaling pathway	36	795	4.53	0.001	KEGG
extracellular matrix	33	680	4.85	0.001	GenMapp
p53 signaling pathway	32	215	14.88	<0.001	KEGG
cell proliferation	32	659	4.86	0.001	GenMapp
Focal adhesion	29	777	3.73	0.035	KEGG
enzyme inhibitor activity	28	458	6.11	<0.001	GenMapp
Cell cycle	27	258	10.47	<0.001	KEGG
Cytokine-cytokine receptor interaction	27	516	5.23	0.001	KEGG
enzyme activator activity	25	548	4.56	0.005	GenMapp
cell surface receptor linked signal transduction	24	321	7.48	<0.001	GenMapp
cell growth	24	330	7.27	<0.001	GenMapp
Wnt signaling pathway	24	509	4.72	0.004	KEGG
Myometrial Relaxation and Contraction Pathways	23	470	4.89	0.003	GenMapp
Apoptosis	22	258	8.53	<0.001	KEGG

Table 2.3: Top 15 pathways of 1423 probe sets which were significant in the senescence experimental group. The Ids of these probe sets were mapped to KEGG, GenMapp and BioCarta pathways and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

Clustering

Hierarchical as well as k-means clustering was performed with the Genesis tool [Sturn et al., 2002]. Hierarchical clustering was applied for probe sets as well as for samples using the absolute expression values of all samples of both experimental groups. Expression profiles of 9492 probe sets were used for clustering because these profiles showed a variation between samples with a coefficient of variation >1.1. The distance between clusters was calculated using the average linkage clustering method with euclidean distance as distance measure. Hierarchical clustering of samples showed the similarity of the expression profiles between the samples (Fig. 2.9). K-means

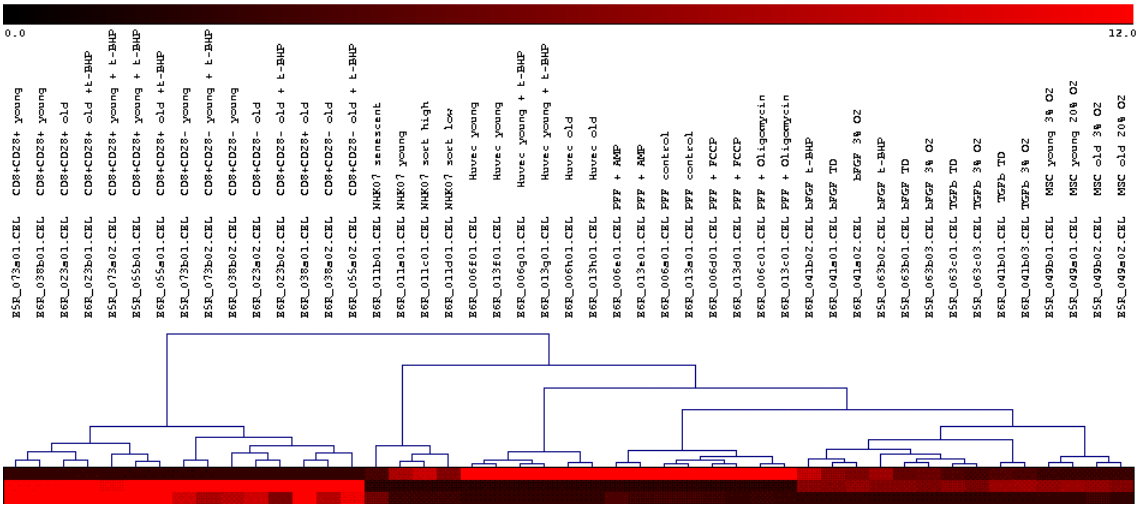

Figure 2.9: Samples of both experimental groups were clustered hierarchically according to their gene expression patters, using the euclidean distance as similarity measure.

clustering was performed on significant probe sets of both experimental groups. The number of clusters was determined by the figure of merit and the euclidean distance was used as similarity measure. As a result, 1830 differentially expressed probe sets were assigned to 12 clusters, where each cluster contained probe sets with a similar expression pattern (see cluster 10 in Fig. 2.10).

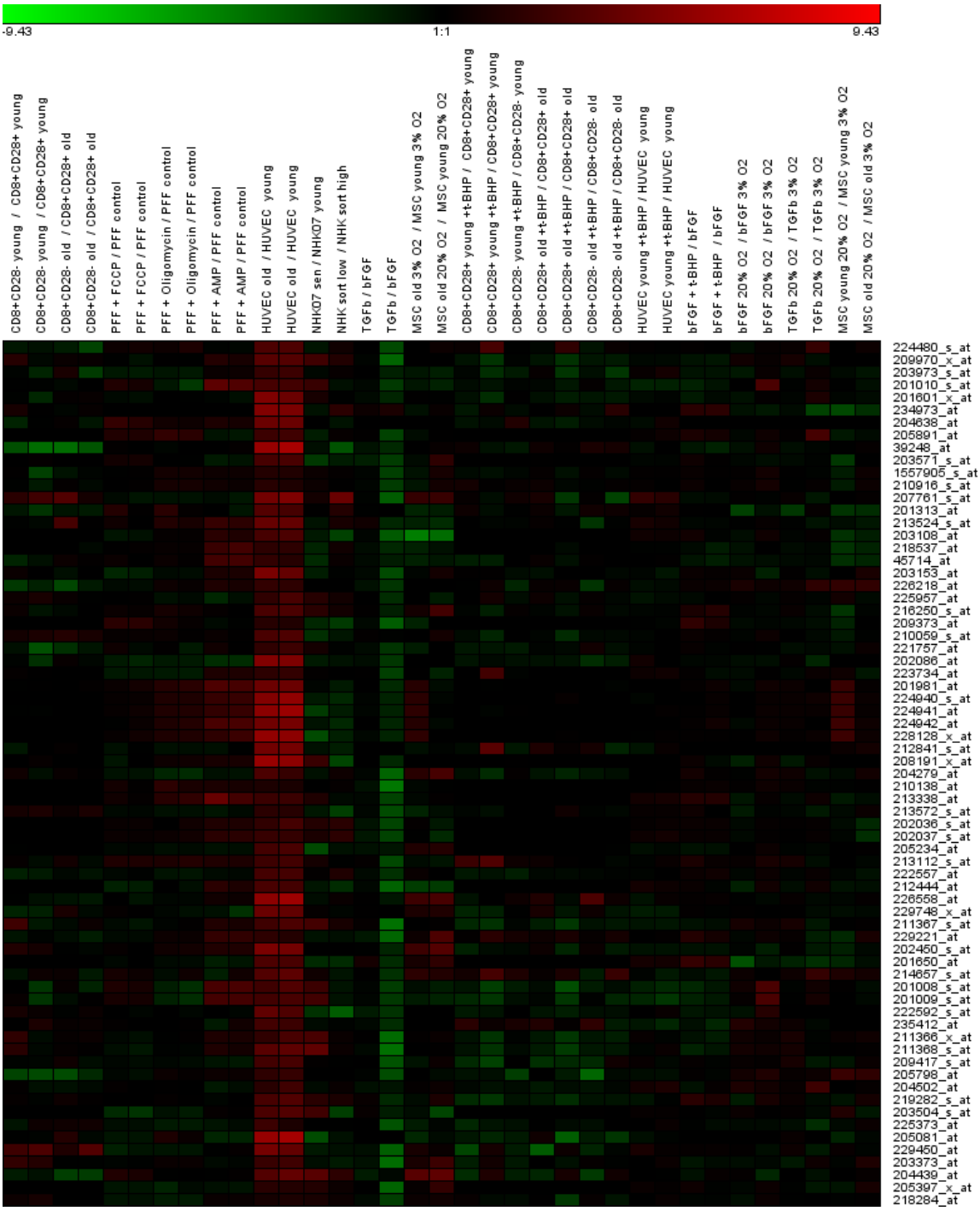

Figure 2.10: K-means clustering. Cluster no. 10 of 12 clusters obtained by k-means clustering of 1830 probe sets which were differentially expressed in at least 9 experiments in each experimental group.

2.2.2 Analysis of UVB induced senescent gene expression profiles

Ultraviolet B (UVB) radiation is emitted by the sun and causes premature extrinsic skin ageing. It has been shown that after UVB treatment, cells acquire typical senescence properties like cell cycle arrest, cell surface increase and β-galactose activity [Chainiaux et al., 2002]. To study the effect of UVB light on skin, human diploid fibroblasts (HDF) were exposed to eight sublethal doses of UVB. HDFs were treated with UVB twice a day on four consecutive days, starting on the day after seeding. The experiment was carried out three times and mRNA samples from UVB treated HDFs and control HDFs were harvested twice on day 9(a, b) of the first experiment, on day 1 (after the first exposure to UVB), day 7 and day 9 of the second experiment and on day 1, day 4, day 7, day 9 and day 15 of the third experiment. Gene expression values in all samples of experiments 1 and 2 were determined by microarray experiments, whereas gene expression values of 96 genes of interest in the samples of experiment 3 were measured using qPCR. For microarray expression data, the fold change of each probe set at each time point of experiments 1 and 2 was calculated as the ratio of the expression values of UVB treated HDFs and control HDFs. The log2 fold change threshold for differential expression was set to +/- 1.5.

Pathway analysis

Pathway analysis was performed with PathwayExplorer [Mlecnik et al., 2005]. Log2 expression ratios of differentially expressed probe sets were mapped to all available KEGG [Kanehisa and Goto, 2000], GenMapp [Dahlquist et al., 2002] and BioCarta [BioCarta, 2010] pathways separately for each time point in experiment 1 and 2.

The result was a pathway list for each time point, where the pathways were sorted by the number of mapped probe sets. Additionally, a p-value, calculated using Fisher's exact test, was provided for each pathway by PathwayExplorer. Fisher's exact test determined if the proportion of differentially expressed probe sets which could be mapped on the pathway was significantly greater than the proportion of probe sets that were mapped on the pathway and were not differentially expressed [Mlecnik et al., 2005].

While the top ten pathways of day 1 in experiment 2 (Tab. 2.6), did not intersect with the highest ranked pathways of any other time point of the two experiments, the top 15 pathway lists of day 9(a) (Tab. 2.4) and day 9(b) (Tab. 2.5) in experiment 1 and day 7 (Tab. 2.7) and day 9 (Tab. 2.8) in experiment 2 had 13 pathways in common. The top four pathways were the same, although in a varying order: chromosome pathway, extracellular matrix pathway and chromosome organization and biogenesis pathway from GenMapp and KEGG's cell cycle pathway.

For example, 61 out of 1947 probe sets which were differentially expressed on day 9 in experiment

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
chromosome	70	476	14.71	<0.001	GenMapp
Cell cycle	55	258	21.32	<0.001	KEGG
extracellular matrix	45	680	6.62	<0.001	GenMapp
chromosome organization and biogenesis	40	530	7.55	<0.001	GenMapp
cytokinesis	37	304	12.17	<0.001	GenMapp
DNA packaging	35	460	7.61	<0.001	GenMapp
cell proliferation	35	659	5.31	<0.001	GenMapp
chromatin	35	264	13.26	<0.001	GenMapp
DNA Replication	31	76	40.79	<0.001	GenMapp
enzyme inhibitor activity	31	458	6.77	<0.001	GenMapp
G1 to S cell cycle control	30	159	18.87	<0.001	GenMapp
Focal adhesion	29	777	3.73	0.03	KEGG
p53 signaling pathway	28	215	13.02	<0.001	KEGG
chromatin assembly/disassembly	26	226	11.5	<0.001	GenMapp
Inositol phosphate metabolism	21	346	6.07	<0.001	KEGG

Table 2.4: Top 15 Pathways of 1402 differentially expressed probe sets on day 9(a), UVB experiment 1. Probe set Ids were mapped to KEGG, GenMapp and BioCarta pathways, and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
chromosome	77	476	16.18	<0.001	GenMapp
extracellular matrix	55	680	8.09	<0.001	GenMapp
Cell cycle	54	258	20.93	<0.001	KEGG
chromosome organization and biogenesis	45	530	8.49	<0.001	GenMapp
DNA packaging	42	460	9.13	<0.001	GenMapp
chromatin	42	264	15.91	<0.001	GenMapp
cell proliferation	41	659	6.22	<0.001	GenMapp
cytokinesis	39	304	12.83	<0.001	GenMapp
enzyme inhibitor activity	32	458	6.99	<0.001	GenMapp
chromatin assembly/disassembly	31	226	13.72	<0.001	GenMapp
ATP-dependent helicase activity	30	353	8.5	<0.001	GenMapp
G1 to S cell cycle control	29	159	18.24	<0.001	GenMapp
DNA Replication	29	76	38.16	<0.001	GenMapp
p53 signaling pathway	24	215	11.16	<0.001	KEGG
Myometrial relaxation and contraction pathway	23	470	4.89	0.031	GenMapp

Table 2.5: Top 15 Pathways of 1754 differentially expressed probe sets on day 9(b), UVB experiment 1. Probe set Ids were mapped to KEGG, GenMapp and BioCarta pathways, and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
actin binding	4	699	0.57	0.035	GenMapp
Adherens junction	3	279	1.08	0.013	KEGG
mRNA processing	3	381	0.79	0.03	GenMapp
Cadherin-mediated cell adhesion	3	249	1.2	0.01	KEGG
Erk1/Erk2 Mapk Signaling pathway	2	127	1.57	0.021	BioCarta
Corticosteroids and cardioprotection	2	57	3.51	0.005	BioCarta
Chromatin Remodeling by hSWI/SNF ATP-dependent Complexes	2	62	3.23	0.005	BioCarta
TGF Beta Signaling Pathway	2	162	1.23	0.033	GenMapp
Chondroitin/Heparan sulfate biosynthesis	2	85	2.35	0.01	KEGG
NFkB activation by Nontypeable Hemophilus influenzae	2	72	2.78	0.007	BioCarta

Table 2.6: Top 10 Pathways of 96 differentially expressed probe sets on day 1, UVB experiment 2. Probe set Ids were mapped to KEGG, GenMapp and BioCarta pathways, and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
extracellular matrix	75	680	11.03	$<$0.001	GenMapp
chromosome	72	476	15.13	$<$0.001	GenMapp
Cell cycle	61	258	23.64	$<$0.001	KEGG
chromosome organization and biogenesis	44	530	8.3	$<$0.001	GenMapp
G1 to S cell cycle control	41	159	25.79	$<$0.001	GenMapp
DNA packaging	41	460	8.91	$<$0.001	GenMapp
chromatin	40	264	15.15	$<$0.001	GenMapp
p53 signaling pathway	39	215	18.14	$<$0.001	KEGG
cytokinesis	38	304	12.5	$<$0.001	GenMapp
Focal adhesion	37	777	4.76	0.019	KEGG
DNA Replication	36	76	47.37	$<$0.001	GenMapp
enzyme inhibitor activity	34	458	7.42	$<$0.001	GenMapp
cell proliferation	33	659	5.01	0.014	GenMapp
chromatin assembly/disassembly	31	226	13.72	$<$0.001	GenMapp
Cytokine-cytokine receptor interaction	30	516	5.81	0.002	KEGG

Table 2.7: Top 15 Pathways of 1815 differentially expressed probe sets on day 7, UVB experiment 2. Probe set Ids were mapped to KEGG, GenMapp and BioCarta pathways, and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

Pathway	Mapped Ids	Total Ids	Mapped Ids (%)	p - value	Source
chromosome	71	476	14.92	<0.001	GenMapp
extracellular matrix	67	680	9.85	<0.001	GenMapp
Cell cycle	61	258	23.64	<0.001	KEGG
chromosome organization and biogenesis	47	530	8.87	<0.001	GenMapp
cytokinesis	42	304	13.82	<0.001	GenMapp
G1 to S cell cycle control	41	159	25.79	<0.001	GenMapp
DNA packaging	41	460	8.91	<0.001	GenMapp
cell proliferation	38	659	5.77	0.003	GenMapp
chromatin	38	264	14.39	<0.001	GenMapp
DNA Replication	35	76	46.05	<0.001	GenMapp
Cytokine-cytokine receptor interaction	33	516	6.4	0.001	KEGG
enzyme inhibitor activity	32	458	6.99	<0.001	GenMapp
chromatin assembly/disassembly	30	226	13.27	<0.001	GenMapp
p53 signaling pathway	28	215	13.02	<0.001	KEGG
Myometrial Relaxation and Contraction Pathways	27	470	5.74	0.011	GenMapp

Table 2.8: Top 15 Pathways of 1947 differentially expressed probe sets on day 9, UVB experiment 2. Probe set Ids were mapped to KEGG, GenMapp and BioCarta pathways, and the resulting pathway list was sorted by the number of mapped probe set Ids.
Mapped Ids (%) is the percentage of mapped probe set Ids among all probe set Ids of the pathway.

2 could be mapped on the KEGG cell cycle pathway which consists of 258 probe sets (Fig. 2.11). In this case, the majority of the mapped probe sets was down-regulated.

Promoter analysis

Promoter Analysis was performed on 6 gene sets, containing genes whose corresponding probe sets were differentially expressed on day 9(a) and day 9(b) in experiment 1 and on day 1 (after the first UVB exposure), day 7 and day 9 of experiment 2 and additionally for 96 genes of interest. The genes of interest were selected manually according to their expression profiles obtained from the microarray experiments and literature research. The promoter sequences of these genes were examined regarding the following transcription factors: aryl hydrocarbon receptor / aryl hydrocarbon receptor nuclear translocator complex (AHR/ARNT), E2 binding factor (E2F), early growth response 2 protein (EGR2), GATA binding protein 6 (GATA6), nuclear factor kappa-light chain enhancer of activated B cells (NF-κB), tumor protein 53 (p53), retinoid acid receptor beta (RARB), SRY (sex-determining region Y)-box 11 (SOX11) and vitamin D receptor / retinoid X receptor complex (VDR/RXR). These transcription factors were identified based on their expression profiles obtained from the two UVB experiments and literature search.

Over-representation analysis was carried out using the programs Pscan [Zambelli et al., 2009]

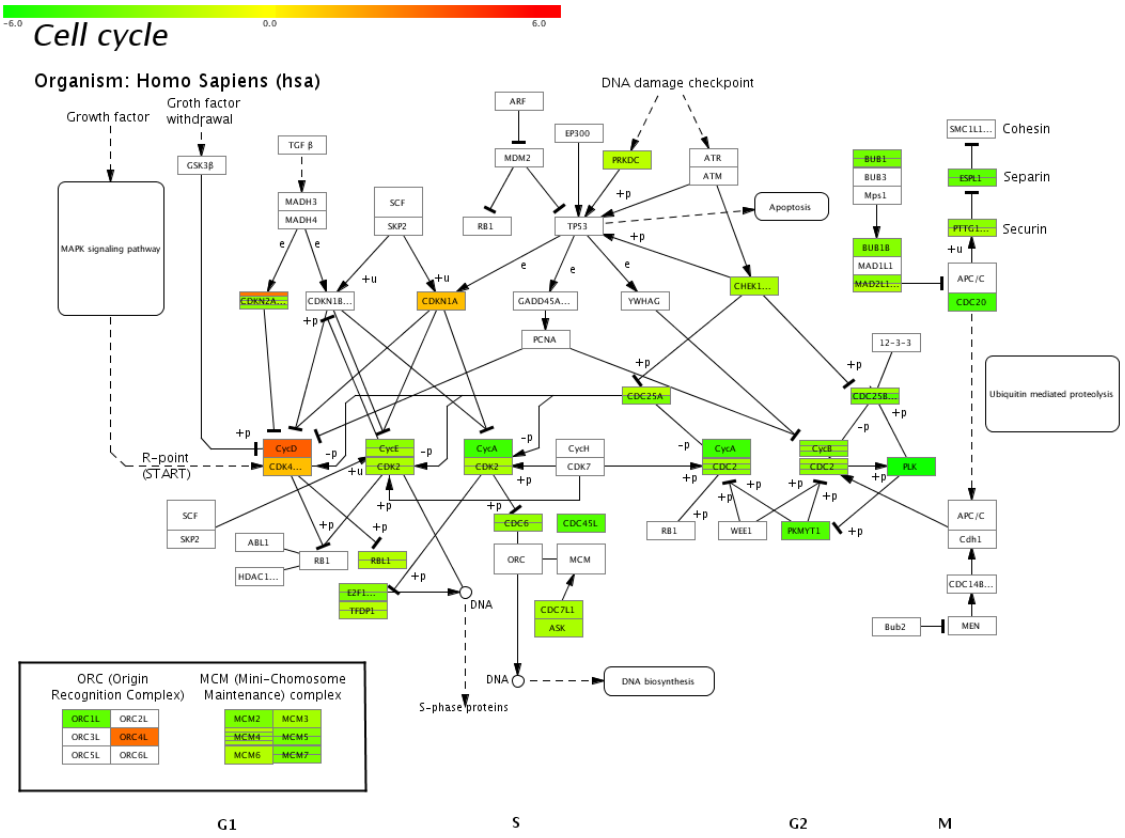

Figure 2.11: Probe sets which were differentially expressed on day 9, UVB experiment 2, were mapped on the KEGG cell cycle pathway. In the pathway, genes which correspond to the probe sets are represented by their symbols. Green coloured boxes denote down-regulated genes, red coloured boxes denote up-regulated genes.

and ORA [Hackl, 2010].

Pscan analysis was performed using TRANSFAC [Wingender et al., 2000] position weight matrices (PWM). TRANSFAC 7.0 offers PWM for transcription factors AHR/ARNT, E2F, EGR2, GATA6, NF-κB and p53. For each gene, the sequence 1000 base pairs (bp) upstream of the transcription start site (TSS) is scanned for binding motifs of the before mentioned transcription factors. The p-value is calculated using a z-test and corrected for multiple testing with the Bonferroni correction. A p-value $<$0.05 denotes that potential binding sites of the transcription factor are significantly over-represented in the promoters of the gene set.

In the promoter sequences of manually selected genes of interest, binding sites for NF-κB and p53 were significantly over-represented (Tab. 2.9). Binding motifs of transcription factors E2F and EGR2 were over-represented in the promoters of the differentially expressed genes of all time points of experiment 1 (Tab. 2.10 and Tab. 2.11) and experiment 2 (Tab. 2.12, Tab. 2.13 and Tab. 2.14). Promoter regions of genes which were differentially expressed on day 9(a) and day 9(b) of

TF	Matrix	adj. p-value	sample size
NF-KB	V$NFKB_Q6	0.015177	101
p53	V$P53_01	0.035357	101
NF-KB	V$NFKAPPAB65_01	0.039752	101

Table 2.9: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 96 genes of interest. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction. *Sample size* is the number of promoters found for the gene set.

TF	Matrix	adj. p-value	sample size
E2F	V$E2F_03	5.23E-07	1085
E2F	V$E2F_02	5.81E-06	1085
E2F	V$E2F_01	1.72E-01	1085
AHR/ARNT	V$AHRARNT_02	0.013988	1085
EGR2	V$EGR2_01	0.015915	1085

Table 2.10: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 1402 differentially expressed probe sets on day 9(a) of UVB experiment 1. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction.
Sample size is the number of promoters found for the differentially expressed probe sets.

TF	Matrix	adj. p-value	sample size
E2F	V$E2F_03	2.65E-16	1432
E2F	V$E2F_02	1.18E-10	1432
E2F	V$E2F_01	2.89E-04	1432
EGR2	V$EGR2_01	3.05E-02	1432
AHR/ARNT	V$AHRARNT_02	0.000499	1432
AHR/ARNT	V$AHRARNT_01	0.039869	1432

Table 2.11: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 1754 differentially expressed probe sets on day 9(b) of UVB experiment 1. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction.
Sample size is the number of promoters found for the differentially expressed probe sets.

TF	Matrix	adj. p-value	sample size
EGR2	V$EGR2_01	0.009075	88
E2F	V$E2F_03	0.017643	88

Table 2.12: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 96 differentially expressed probe sets on day 1 of UVB experiment 2. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction.
Sample size is the number of promoters found for the differentially expressed probe sets.

TF	Matrix	adj. p-value	sample size
E2F	V$E2F_02	3.85E-10	1369
E2F	V$E2F_03	1.65E-08	1369
E2F	V$E2F_01	1.85E-02	1369
EGR2	V$EGR2_01	0.000213628	1369
NF-KB	V$NFKAPPAB65_01	0.00765967	1369
AHR/ARNT	V$AHRARNT_01	0.0127853	1369

Table 2.13: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 1815 differentially expressed probe sets on day 7 of UVB experiment 2. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction.
Sample size is the number of promoters found for the differentially expressed probe sets.

TF	Matrix	adj. p-value	sample size
E2F	V$E2F_03	9.83E-12	1498
E2F	V$E2F_02	9.42E-12	1498
E2F	V$E2F_01	7.68E-03	1498
EGR2	V$EGR2_01	7.31E-02	1498
AHR/ARNT	V$AHRARNT_02	0.010054	1498
AHR/ARNT	V$AHRARNT_01	0.021456	1498

Table 2.14: Significantly over-represented TFBS determined by Pscan in the promoter sequences of 1974 differentially expressed probe sets on day 9 of UVB experiment 2. 1000 bp upstream of the transcription start site were scanned for TFBS and the p-value was calculated with a z-test and corrected for multiple testing using Bonferroni correction.
Sample size is the number of promoters found for the differentially expressed probe sets.

experiment 1 and on day 7 and day 9 of experiment 2 showed a significantly increased number of potential binding sites for AHR/ARNT. Additionally, in the promoter sequences of the differentially expressed genes of day 7, experiment 2, potential NF-κB binding sites were over-represented.

ORA scans sequences 4500 bp upstream and 500 bp downstream of the transcription start site. It provides PWM for the transcription factors AHR/ARNT, E2F, EGR2, GATA6, NF-κB, p53 and RXR/VDR. Fisher's exact test is used for the calculation of the p-value, which is then corrected for multiple testing with the Benjamini-Hochberg false discovery rate (FDR) control method. If the adjusted p-value of a transcription factor is $<$0.05, binding sites for the factor are significantly over-represented in the promoter sequences of the input gene set.
Binding sites for E2F were significantly over-represented in all six gene sets (Tab. 2.15, Tab. 2.16, Tab. 2.17, Tab. 2.18, Tab. 2.19 and Tab. 2.20). Furthermore, p53 binding motifs were over-represented in the promoter sequences of the 96 genes of interest. Additionally, p53 and NF-κB motifs were over-represented in the set of differentially expressed genes of day 9(a), experiment 1. NF-κB binding sites were also over-represented in the promoter sequences of differentially expressed genes on day 1, experiment 2. Moreover, promoter sequences of day 1 contained a

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	MA0024	148	26	3.76E-06	0.0002
p53	M00034	148	11	4.31E-06	0.0002
E2F	M00050	148	26	1.23E-05	0.0002
E2F	M00024	148	20	0.0002	0.0021
p53	MA0106	148	12	0.0017	0.0108

Table 2.15: Significantly over-represented TFBS determined by ORA in the promoter sequences of 96 genes of interest. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction.
Dataset entries is the number of promoters found for the gene set.

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	M00516	710	161	1.20E-18	2.12E-16
E2F	MA0024	710	109	4.48E-16	5.27E-14
E2F	M00050	710	110	1.35E-14	9.55E-13
E2F	M00024	710	78	1.78E-08	8.97E-07
NF-kappaB	M00208	710	69	0.0006	0.0089
p53	M00034	710	19	0.0029	0.0221

Table 2.16: Significantly over-represented TFBS determined by ORA in the promoter sequences of 1402 differentially expressed probe sets on day 9(a) of UVB experiment 1. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction.
Dataset entries is the number of promoters found for the gene set.

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	M00516	747	166	2.85E-18	1.00E-15
E2F	MA0024	747	104	9.66E-13	8.50E-11
E2F	M00050	747	105	1.96E-11	1.15E-09
E2F	M00024	747	80	3.54E-08	1.56E-06
Egr-2	M00246	747	142	0.0004	0.0066
AhR:Arnt	M00235	747	56	0.003	0.0262

Table 2.17: Significantly over-represented TFBS determined by ORA in the promoter sequences of 1754 differentially expressed probe sets on day 9(b) of UVB experiment 1. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction.
Dataset entries is the number of promoters found for the gene set.

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	M00516	128	33	2.17E-06	8.24E-05
NF-kappaB	M00051	128	20	6.09E-05	0.0008
Egr-2	M00246	128	35	0.0001	0.0014
E2F	MA0024	128	20	0.0002	0.0025
E2F	M00050	128	20	0.0006	0.0049
E2F	M00024	128	16	0.002	0.0109
NF-kappaB	M00054	128	20	0.0074	0.0283
AhR:Arnt	M00237	128	15	0.0093	0.0337

Table 2.18: Significantly over-represented TFBS determined by ORA in the promoter sequences of 96 differentially expressed probe sets on day 1 of UVB experiment 2. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction. *Dataset entries* is the number of promoters found for the gene set.

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	M00516	764	161	1.23E-15	4.33E-13
E2F	MA0024	764	111	1.12E-14	1.97E-12
E2F	M00050	764	111	7.15E-13	8.39E-11
E2F	M00024	764	74	4.29E-06	0.0003

Table 2.19: Significantly over-represented TFBS determined by ORA in the promoter sequences of 1815 differentially expressed probe sets on day 7 of UVB experiment 2. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction. *Dataset entries* is the number of promoters found for the gene set.

TF	Matrix	Dataset entries	TFBS in dataset	p-value	adj. p-value
E2F	M00516	797	179	5.22E-20	1.83E-17
E2F	MA0024	797	113	3.48E-14	4.06E-12
E2F	M00050	797	115	4.61E-13	4.03E-11
E2F	M00024	797	75	1.01E-05	0.0004

Table 2.20: Significantly over-represented TFBS determined by ORA in the promoter sequences of 1947 differentially expressed probe sets on day 9 of UVB experiment 2. 4500 bp upstream and 500 bp downstream of the transcription start site were scanned for TFBS and the p-value was calculated using Fisher's exact test and corrected for multiple testing using Bonferroni correction. *Dataset entries* is the number of promoters found for the gene set.

significantly over-represented number of binding sites for EGR2 and AHR/ARNT, which were also over-represented in the promoters of differentially expressed genes of day 9(b), experiment 1.

Transcription factor binding site (TFBS) prediction was performed using Genomatix MatInspector [Cartharius et al., 2005] and Match [Kel et al., 2003]. The differentially expressed gene lists of day 9(a) and day 9(b) of experiment 1 and day 1, day 7 and day 9 of experiment 2 were splitted into sets of up- and down-regulated genes. Both programs scanned the promoter sequences of the genes in the gene sets retrieved from the Genomatix ElDorado genome database. For each gene, the sequence 500 bp upstream of the first transcription start site and 100 bp downstream of the last transcription start site were scanned for potential TFBS.

MatInspector provides PWM families for the transcription factors AHR/ARNT, E2F, EGR2, GATA6, NF-κB, p53, RARB, SOX11 and VDR/RXR. Since the binding motifs of RARB and VDR/RXR are very similar, the same PWM family is used to search for the binding sites of both transcription factors. The tool reports binding sites whose matrix similarity score exceeds the optimized matrix threshold provided by MatInspector for each PWM. MatInspector, for example, predicts 241 binding sites for NF-κB in the promoter sequences of the 96 selected genes of interest (Appendix D). Furthermore, binding sites for EGR2 were the most frequent binding sites in the promoter regions of each gene set, followed by binding sites for E2F as well as RARB and VDR/RXR (Tab. 2.21).

	EGR2	**E2F**	**RARB VDR/RXR**	**SOX11**	**GATA6**	**NF-κB**	**AHR/ARNT**	**P53**
GOI* (174)	1053	604	396	331	207	241	207	215
day 1 (255)	1623	1037	549	489	405	343	351	187
day 7 $\downarrow$ (650)	4535	3230	1593	1184	701	1035	1191	542
day 7 $\uparrow$ (719)	3967	2350	1705	1615	1043	995	880	674
day 9 $\downarrow$ (760)	5409	3799	1835	1364	789	1169	1438	704
day 9 $\uparrow$ (736)	4157	2403	1744	1693	1099	995	914	662
day 9 (1/1) $\downarrow$ (592)	4004	2916	1511	1075	664	876	1036	522
day 9 (1/1) $\uparrow$ (491)	2821	1604	1230	1069	645	703	572	498
day 9 (1/2) $\downarrow$ (880)	6388	4249	2102	1616	975	1373	1560	813
day 9 (1/2) $\uparrow$ (570)	3223	1901	1420	1260	762	779	667	579

Table 2.21: Genomatix MatInspector TFBS. For each transcription factor the number of binding sites found by Genomatix MatInspector in the promoter sequences of the gene sets are listed.
The number in brackets next to the gene set identifier are the number of promoter sequences of the genes in the gene set.
* **GOI**: genes of interest
$\downarrow$: down-regulated
$\uparrow$: up-regulated

The least frequent binding sites for the promoters of all gene sets except for 96 manually selected genes of interest were TFBS for p53. Furthermore, transcription factors with binding sites in the

majority of the promoter sequences of a gene set were determined (Tab. 2.22). TFBS for E2F

	TF	# Sequences	p-value
GOI*	E2F	160 / 174	2.41E-20
GOI*	EGR2	150 / 174	2.02E-17
day 1	E2F	246 / 255	8.00E-41
day 7 ↓	E2F	612 / 650	4.03E-85
day 7 ↓	EGR2	608 / 650	4.53E-101
day 7 ↓	RARB, VDR/RXR	573 / 650	0.00842825
day 7 ↑	E2F	634 / 719	1.20E-58
day 9 ↓	E2F	718 / 760	1.01E-101
day 9 ↓	EGR2	718 / 760	1.39E-125
day 9 ↓	RARB, VDR/RXR	664 / 760	0.0252167
day 9 ↑	E2F	646 / 760	5.31E-58
day 9 (1/1) ↓	E2F	559 / 592	2.64E-79
day 9 (1/1) ↓	EGR2	549 / 592	2.32E-87
day 9 (1/1) ↓	RARB, VDR/RXR	527 / 592	0.00181869
day 9 (1/1) ↑	E2F	432 / 491	5.98E-40
day 9 (1/1) ↑	RARB, VDR/RXR	434 / 491	0.0134264
day 9 (1/2) ↓	E2F	827 / 880	4.27E-113
day 9 (1/2) ↓	EGR2	819 / 880	5.79E-132
day 9 (1/2) ↑	E2F	498 / 570	8.15E-44
day 9 (1/2) ↑	RARB, VDR/RXR	502 / 570	0.0150379

Table 2.22: Genomatix MatInspector common TFBS. Binding sites of the listed transcription factors (TF) could be found in at least 86% of the promoter sequences of a gene set. The p-value provided by the MatInspector software is the probability of observing a match between a matrix of the matrix family and the scanned sequence in as many or more sequences in a random sequence sample with the same amount of sequences as the examined sequence set [MatInspector, 2010].
 # Sequences: number of sequences in the gene set with at least one binding site of the specified transcription factor / total number of sequences in the gene set
* **GOI**: genes of interest
↓: down-regulated
↑: up-regulated

could be found in more than 86% percent of the examined promoter regions of each gene set. EGR2 binding sites were located in more than 86% of the promoter sequences of the genes of interest and of all the sets containing down-regulated genes. Finally, RARB and RXR/VDR binding sites were determined in at least 86% of the promoter sequences of the down-regulated gene sets, except for the day 9(b), experiment 1, and in the up-regulated gene sets of day 9(a) and day 9(b) of experiment 1.

Match uses TRANSFAC 6.0 position weight matrices for the transcription factors AHR/ARNT, E2F, EGR2, NF-κB and p53. All binding sites exceeding a matrix similarity score cut-off, which minimizes the number of false positive TFBS, are reported by Match. Match, for instance, returns 9 potential NF-κB binding sites in the promoter sequences of the 96 genes of interest (Tab. 2.23).

In all gene sets except for day 7, experiment 2, down-regulated and day 9(a), experiment 1, down-

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence (+ Strand)
NF-κB	B2M	66	+	1	0.997	GGGAAttccc
NF-κB	B2M	66	-	1	0.997	gggaaTTCCC
NF-κB	B2M	413	-	1	0.975	cgggaaaGTCCCtc
NF-κB	B2M	415	-	0.986	0.992	ggaaaGTCCC
NF-κB	EGR2	150	-	1	0.976	ttggaaaGTCCCag
NF-κB	EGR2	151	-	1	0.997	tggaaaGTCCCa
NF-κB	EGR2	152	-	0.986	0.992	ggaaaGTCCC
NF-κB	NME5	618	-	1	0.997	ccggaaaGTCCCcg
NF-κB	NME6	620	-	0.986	0.992	ggaaaGTCCC

Table 2.23: Match predicts 9 NF-κB binding sites in the promoter sequences of the 96 genes of interest. 500 bp upstream and 100 bp downstream of the transcription start site were scanned for TFBS.

regulated, most binding sites were found for EGR2 (Tab. 2.24). E2F binding sites represented the highest number of binding sites found by Match in the promoter sequences of the genes which were down-regulated on day 7, experiment 2 and on day 9(a), experiment 1. In all gene sets, except for day 1 of experiment 2, AHR/ARNT binding sites were the least frequent binding sites.

	EGR2	p53	NF-κB	E2F	AHR/ARNT
GOI* (174)	22	16	9	2	0
day 1 (255)	53	14	0	6	7
day 7 $\downarrow$ (650)	67	17	18	71	1
day 7 $\uparrow$ (719)	61	32	30	19	1
day 9 $\downarrow$ (760)	83	17	13	80	1
day 9 $\uparrow$ (736)	63	29	17	22	2
day 9 (1/1) $\downarrow$ (592)	39	9	16	56	2
day 9 (1/1) $\uparrow$ (491)	32	18	14	13	1
day 9 (1/2) $\downarrow$ (880)	86	19	23	74	3
day 9 (1/2) $\uparrow$ (570)	44	24	12	21	1

Table 2.24: Match TFBS. The number of binding sites found by Match in the promoter sequences of the gene sets for each transcription factor are listed. The number in brackets next to the gene set identifier are the number of promoter sequences of the gene set.
*** GOI**: genes of interest
$\downarrow$: down-regulated
$\uparrow$: up-regulated

Gene-gene interaction network

Gene expression values obtained by qPCR of 96 genes of interest in samples of day 1, day 4, day 7, day 9 and day 15 of experiment 3 were used for the construction of a gene-gene interaction network. In gene expression data, correlations can be used to determine interactions [Zhang et al., 2009] and therefore the statistical framework R [R Development Core Team, 2007] was used to calculate pairwise correlation between all 96 genes. Based on the qPCR expression profiles the correlations and their significance were computed with the Pearson correlation method [Bland, 2006] and a t-test.

Using the Cytoscape tool [Shannon et al., 2003], a network which visualizes the interactions between genes was constructed from the statistically significant correlations with a p-value <0.05. The resulting network had 96 nodes, each node representing a gene, and 571 edges corresponding to the gene-gene interactions.

The Cytoscape plug-in Cluego [Bindea et al., 2009] was used to determine molecular function GO terms [Ashburner et al., 2000] of the genes of interest. GO terms of the levels 3-8 were retrieved if at least two genes of interest could be associated with the GO term and if those genes made up at least three percent of the total genes that are associated to this term (Fig. 2.12).

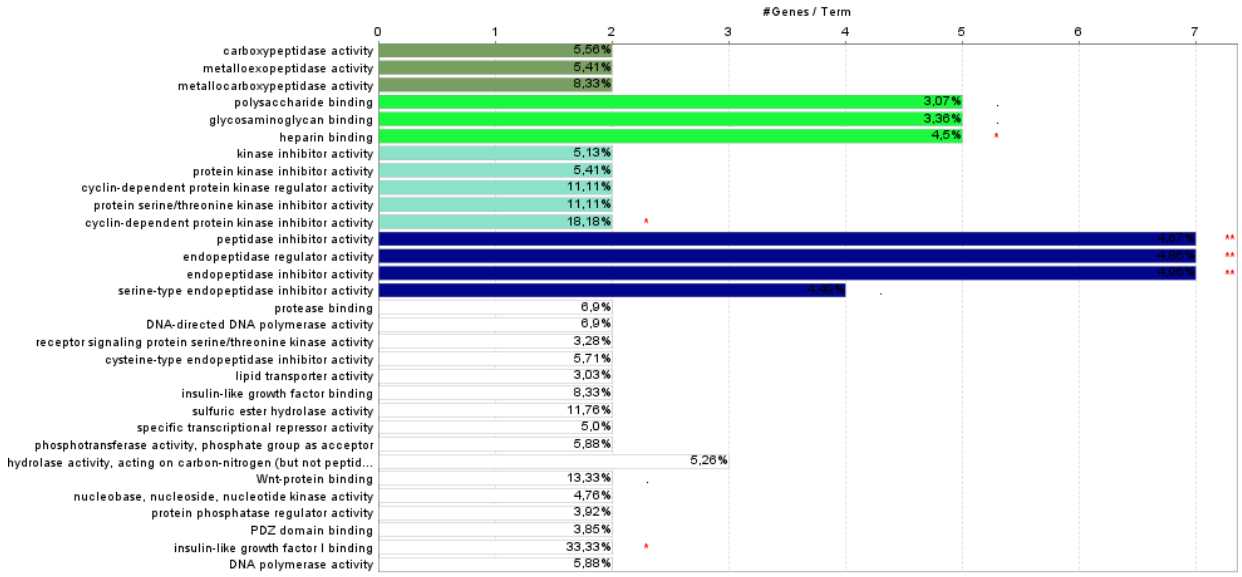

Figure 2.12: Molecular function GO terms of the levels 3-8, which are associated with the 96 genes of interest. Every term has at least two associated genes of interest and these associated genes make up at least 3% of the total genes associated with this term.

Finally, GO terms were mapped on the gene-gene interaction network with the GOlorize Cytoscape plug-in [Garcia et al., 2007], to visualize gene - GO term associations shown as colours (Fig. 2.13).

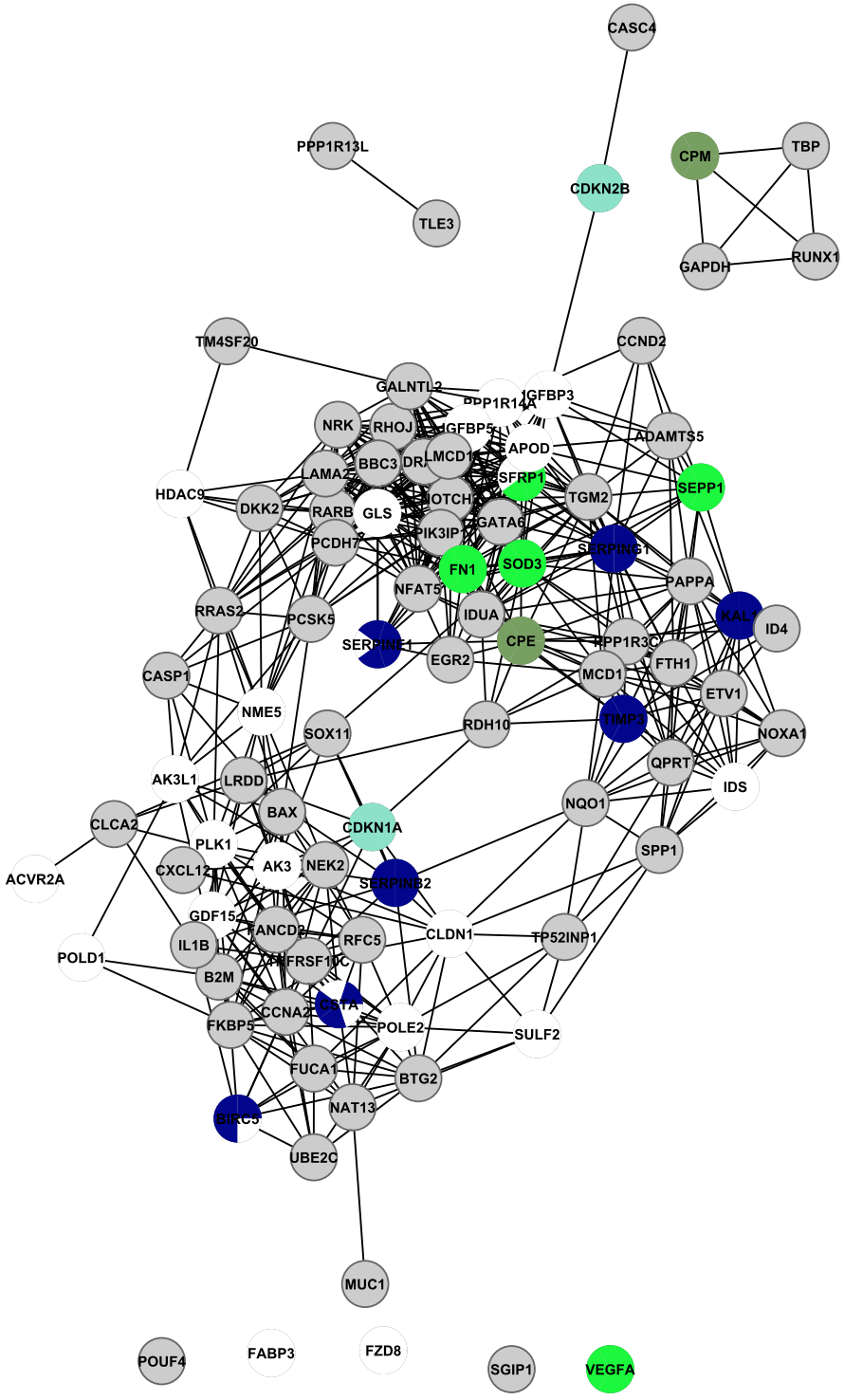

Figure 2.13: Gene-gene interaction network where nodes and edges represent genes and gene interactions. The width of the edges corresponds to the correlation coefficient, and the node colours correspond to the molecular function GO terms of the genes. Grey nodes do not have any molecular function GO term associated, for the other nodes, see Fig. 2.12 for the colour legend.

Chapter 3

Discussion

3.1 Development of GiSAO.db

In this thesis GiSAO.db, a web-based database system for storage and retrieval of data, such as gene expression values and orthologs, of genes involved in senescence, apoptosis and oxidative stress, was designed and developed.
GiSAO.db data, web application, database content, resemblances and differences to similar databases as well as the applied software technology are discussed in this section.

The database schema of GiSAO.db has been designed specifically for the storage of normailzed gene expression data obtained from microarray experiments. It has been shown that miRNAs play a role in the process of ageing [Bates et al., 2009][Grillari and Grillari-Voglauer, 2010][Chen et al., 2010] and therefore normalized miRNA expression data of microarray experiments were integrated into GiSAO.db additionally.
Within the database, a gene or miRNA is represented by the Id of the corresponding spot on the analysed microarray chip and all gene- or miRNA-related data in the database (e.g. expression values or experimental data) are linked to this/these probe (set) Id(s). As probe (set) Ids are usually not very expressive, annotation data comprising gene and miRNA identifiers, such as gene symbol or the miRNA name, are available to facilitate the identification of a gene or miRNA in the web application. Furthermore, for each gene in GiSAO.db GO terms, which describe functions, locations and biological processes of gene products, are provided.
In the GiSAO.db web application gene expression values and miRNA expression ratios are represented by colour coded boxes which are used as a visual support to interpret the results of a

microarray experiment by facilitating the determination of prominently expressed genes, up- or down-regulated miRNAs and comparison of expression values or ratios of different experiments.
Genetic ageing research in humans is often supported by studying ageing in model organisms which have a much shorter life span than humans, e.g. Mus musculus, Saccharomyces cerevisae, Caenorhabditis elegans and Drosophila melanogaster. Most of the research results obtained from the model organisms can be adopted for humans as those organisms share numerous orthologous genes with humans and orthologous genes often have the same function [Lodish et al., 2007]. Thus ortholog data were integrated into GiSAO.db to compare the research results from different species and to determine human genes with a previously unknown connection to ageing.
GiSAO.db stores two types of ortholog data; ortholog data provided by Affymetrix and verified ortholog data which are entered manually with a specification of the source. Affymetrix provides numerous orthologs for each chip but quite a few of these orthologs are only predicted. Therefore, GiSAO.db offers the possibility to enter validated ortholog data.
In order to verify microarray experiments or perform further investigations, often additional experiments are carried out using biomolecular methods like qPCR or Western blot. GiSAO.db provides facilities to store and access such experimental data via a search function and tags. In gene lists, tags display the organism used for an experiment as well as the experiment type and classification of experiments that were performed on a gene and provide links to these experiments.
Genes or miRNAs of interest can be assembled to favourite lists which can be compared among different research groups. Sophisticated search functions provide easy and comprehensive access to the data in the database. Moreover, GiSAO.db facilitates access to further information concerning the database data via links to external databases.
Asynchronous data upload and update functions were implemented so that the server is not blocked and GiSAO.db can be used while the upload takes place. A report gives feedback on upload success or failure. To enable further processing of data stored in GiSAO.db, data export functions are available. The data can be exported in several file formats which are suitable for direct import in data processing programs such as MS Excel.
A data dictionary concept was introduced to facilitate manual data input. A data dictionary is an enhanced selection field, which assists in controlled, yet customizable and quick submission of data.
Web 2.0 components like autocomplete fields, sortable lists and drag and drop fields were adopted to improve the usability of the web application.

GiSAO.db has been implemented using Java EE which is a platform for the development of robust, secure and platform independent web applications. To accelerate the development of the

web application, a basic code skeleton was built from UML models with the AndroMDA code generation frame work. Model driven architecture reduces recurring coding tasks and facilitates the extension of the web application's architecture. Presently, application data are stored in a reliable and secure Oracle RDBMS.
Java Server Pages were used to create web pages with dynamic content and by using the Struts framework, the presentation was separated from the application logic which facilitates maintenance and extension of the web application. Finally, the AJAX technology was used to add Web 2.0 functionality to GiSAO.db. AJAX facilitates the development of desktop-like web applications as it offers asynchronous data loading without a complete page reload.

GiSAO.db offers annotation data for the Affymetrix microarray chips Human Genome U133 Plus 2.0 Array, Mouse Genome 430 2.0 Array, Yeast Genome 2.0 Array, C. elegans Genome Array and Drosophila Genome Array. Moreover, annotation data for two custom made human Exiqon miRNA microarrays are provided. The database contains Affymetrix orthologs between Homo sapiens, Mus musculus, Saccharomyces cerevisae, Caenorhabditis elegans and Drosophila melanogaster as well as numerous verified orthologs. Additionally, data of several follow-up experiments are stored. Currently, GiSAO.db contains gene expression data of 11 experiments comprising in all 111 Affymetrix microarrays of three different species: Homo sapiens, Mus musculus and Saccharomyces cerevisiae. Seven human miRNA experiments with 40 Exiqon microarrays in total are stored in the database.

Human Aging Genomic Resources (HAGR) [de Magalhães et al., 2009a], *Aging Gene Database* [genesDB, 2010], *Gene Aging Nexus* (GAN) [Pan et al., 2007] and the *Atlas of Gene Expression in Mouse Aging Project* (AGEMAP) [Zahn et al., 2007] are publically available databases that also store ageing-related data. Among these databases HAGR and the *Aging Gene Database* also store ortholog data. HAGR and GAN additionally provide annotation data. However, just GAN and AGEMAP [Zahn et al., 2007] store gene expression profiles obtained from microarray experiments studying ageing. GAN contains data of 6 species (Homo sapiens, Mus musculus, Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans and Rattus norvegicus) whereas AGEMAP only contains mouse data.
To the best of our knowledge, GiSAO.db is the only database which contains ageing-related gene expression data sets obtained from microarray experiments together with ortholog data and additionally stores miRNA microarray data as well as data of follow-up experiments.

3.2 Integrative data analysis

Microarray gene expression patterns stored in GiSAO.db were analysed and combined with other types of data, e.g. genomic data or annotations, to investigate mechanisms of ageing. These analyses are discussed in the following sections.

3.2.1 Comparison of gene expression profiles of various ageing models

Genome wide gene expression profiles obtained from two groups of microarray experiments studying either oxidative stress induced senescence or other senescence models were compared to determine genes which are consistently differentially expressed in ageing using a similar approach as de Magalhães et al. [de Magalhães et al., 2009b].

Within each experimental group sample pairs were assembled and differential expression between samples was defined using a fold change cut-off of +/- 1.5. To determine if a probe set was associated with oxidative stress or senescence, the differential expression frequency of each gene in both experimental groups was counted. The significance of these associations was determined by the means of a p-value based on the value counting method. The p-value was calculated using the cumulative binomial distribution, which provided the probability for a probe set to be as often or more often differentially expressed than actually observed. Since the p-value was computed simultaneously for each of the ten thousands of probe sets on the used microarray, it was necessary to correct the resulting p-values for multiple testing by controlling the false discovery rate (FDR). As opposed to the false positive rate which gives the proportion of probe sets that are wrongly declared significant among all probe sets which are not significant, the FDR is the share of expected false positives among all probe sets which were defined as significant [Storey and Tibshirani, 2003]. Therefore the q-value representing the FDR is more conclusive than the p-value which is a measure of the false positive rate. The FDR cut-off was set to 0.005 for both experimental groups, yielding a p-value cut-off of 0.0002235 for the oxidative stress experimental group and 0.000578618 for the senescence experimental group. 587 probe sets which were differentially expressed in at least 9 out of 17 sample pairs in the oxidative stress experimental group passed the q-value cut-off as well as 1423 probe sets which were differentially expressed in at least 9 put of 18 senescence experiments. The two resulting probe set lists shared 180 probe sets. Statistical analysis was performed using the well established statistical framework R [R Development Core Team, 2007] and the qvalue package [Storey and Tibshirani, 2003] for FDR calculation.

In order to further investigate the biological context of the genes corresponding to the probe sets

in the resulting lists, pathway analysis was performed using the sophisticated PathwayExplorer tool [Mlecnik et al., 2005]. The analysis was conducted separately for each probe set list and the common probe sets to determine pathways which play a role in oxidative stress induced senescence, other senescence forms or both. The probe sets and their expression ratios were mapped onto all available KEGG [Kanehisa and Goto, 2000], GenMapp [Dahlquist et al., 2002] and BioCarta [BioCarta, 2010] pathways and the resulting pathway lists were sorted by the number of probe sets which could be mapped on a pathway.
Pathways for apoptosis, cell cycle and p53 signalling were among the top 15 pathways in all three probe set lists, whereby the apoptosis pathway was ranked at the top of all lists and the cell cycle pathway was on the second place for the significant probe sets of the oxidative stress experimental group and the common probe sets. Chromosome, lipid biosynthesis and enzyme activator activity pathways were also listed at the top in the oxidative stress experimental group. In the list of common probe sets, pathways for cell proliferation and p53 signalling were among the highest ranked pathways. In the senescence experimental group MAPK signalling, extracellular matrix p53 signalling and cell cycle proliferation were top pathways.
The mature cluster analysis application Genesis [Sturn et al., 2002] was used for hierarchical clustering and k-means clustering. Hierarchical sample and probe set clustering was performed to show similarities between the expression patterns of samples and probe set respectively. Expression profiles of the samples of both experimental groups which showed a variation between samples (coefficient of variation $>$1.1) were used for clustering since unvarying expression patterns do not provide any information. Additionally, k-means clustering was carried out for the expression ratios of probe sets with a q-value $<$0.005 to pool probe sets with a similar fold change pattern.

3.2.2 Analysis of UVB induced senescent gene expression profiles

Gene expression profiles obtained from microarray and qPCR experiments investigating UVB induced senescence were analysed.
Microarray experiments were performed for samples of 5 time points from two biological experiments. One sample was taken at the beginning of the UVB treatment and the four remaining samples were taken after the treatment was completed. Differential expression was defined using a 1.5 log2 fold change cut-off between sample and untreated control.
Pathway analysis was performed separately for differentially expressed probe sets of each time point by mapping the probe sets and their expression ratios onto KEGG [KEGG, 2010], GenMapp [GenMAPP, 2010] and BioCarta [BioCarta, 2010] pathways. The resulting pathway lists were sorted by the number of probe sets which ould be mapped on a pathway. For all time points after

the treatment, the top 15 pathways had 13 pathways in common and the top 4 pathways were the same, although in a different order: *chromosome pathway*, *extracellular matrix pathway*, *chromosome organization and biogenesis pathway* and *cell cycle pathway*.

In order to find regulators of senescence, promoter analysis was performed for differentially expressed genes at each time point with the focus on predetermined transcription factors. Significantly over-represented transcription factor binding sites were determined with the tools Pscan [Zambelli et al., 2009] and ORA [Hackl, 2010] to find out which transcription factors are most likely to regulate genes at the various time points. Pscan analysis yielded significantly over-represented binding sites for transcription factors E2F and EGR2 in the promoter sequences of genes differentially expressed at each time point and AHR/ARNT at the time points after the treatment. The analysis using the ORA tool revealed significantly over-represented transcription factor binding sites for E2F in differentially expressed genes at all time points. The differing results obtained by Pscan and ORA probably result from the varying regions that were scanned; Pscan searched 1000 bp upstream of the transcription start (TSS) site while ORA scanned the sequence 4500 bp upstream and 500 bp downstream of the TSS. Additionally, the two tools apply matrices from different TRANSFAC database versions and ORA uses also Jaspar PWM and custom PWM.

Transcription factor binding sites for the predetermined transcription factors were predicted in the promoter sequences of differentially expressed genes at all time points using Genomatix MatInspector [Cartharius et al., 2005] and Match [Kel et al., 2003]. Genomatix MatInspector reported binding sites for AHR/ARNT, E2F, EGR2, GATA6, NF-κB, p53, RARB, SOX11 and VDR/RXR as well as transcription factors which have binding sites in at least 86% of the promoters in a gene set. E2F binding sites can be found in at least 86% of the promoters of each gene set.

Match provided binding sites for the transcription factors AHR/ARNT, E2F, EGR2, NF-κB and p53. Far less binding sites were reported by Match than by Genomatix MatInspector. On the one hand, different PWM may be the reason; Match uses TRANSFAC position weight matrices whereas MatInspector uses its own matrices, on the other hand a far more stringent matrix similarity cutoff to minimize the false positive reported binding sites was used for the Match analysis. However, the resulting binding sites of both programs are just predictions; their functionality has to be confirmed experimentally.

Promoter analysis was additionally performed for 96 genes of interest which were selected based on their expression patterns obtained from the UVB microarray experiments and on literature search. Pscan reported over-represented transcription binding sites of NF-κB and p53 in the promoters of these genes whereas ORA determined binding sites for E2F and p53 as over-represented. In more than 86% of the sequences of the gene set Genomatix MatInspector found E2F and EGR2 binding sites.

Finally, expression profiles of these genes of interest were obtained by qPCR experiments with one sample taken at the beginning of the UVB treatment, and the other samples taken on 4 time points after the treatment. These expression patterns were used to construct a network which visualizes the interactions between genes or gene products. Since gene interactions can be represented by correlations between genes [Zhang et al., 2009], the statistical framework R was used to compute pairwise correlations. From these correlations a network was constructed using Cytoscape [Shannon et al., 2003] to visualize gene-gene interactions. However, the co-expression network does not provide whether an interaction between two genes is direct or indirect. To enrich the network with annotations, over-represented molecular function GO terms were determined in the gene set with the Cytoscape plugin ClueGO [Bindea et al., 2009]. The GOlorize plugin [Garcia et al., 2007] was used to map these GO terms onto genes in the network and represent them as colours.

3.3 Conclusion

GiSAO.db is a web-based database application which allows researchers to store and retrieve ageing-related gene data. The rich user interface facilitates data input and database queries. Data access is controlled by an authorization and authentication system which manages user data.

By comparing gene expression profiles of various ageing models stored in GiSAO.db numerous genes, which were consistently differentially expressed in either oxidative stress induced senescence, other forms of senescence or both, were determined as well as relevant pathways of these genes.

Analysis of gene expression patterns reflecting UVB induced senescence revealed several pathways which might be related to senescence. Possible regulators of senescence were determined and their binding sites were predicted in genes whose expression is affected by UVB treatment. Finally, interactions between several genes which could be associated to UVB induced senescence were visualized in a network.

Chapter 4

Methods

This chapter gives an overview of the methods that were applied to achieve the aims of the thesis. In the first part, software technologies used for the development of GiSAO.db are introduced. The second part focuses on methods and tools used for the analysis of data stored in GiSAO.db.

4.1 Software technology

This section describes software development technologies, frameworks and tools that were used to develop the GiSAO.db web application and database.

4.1.1 Java Enterprise Edition (Java EE)

Java is an object-oriented general purpose programming language [Gosling et al., 2000]. It is designed as a high-level language which provides automatic storage management using a garbage collector. In order to run a Java program, Java code is first compiled into a machine-understandable byte code which is then executed within the Java virtual machine (JVM). The JVM handles loading and linking of objects, program optimization and program execution.
As Java byte code is platform independent, it can be executed on any machine regardless of hardware or operating system. Furthermore, the Java technology offers security features to prevent unauthorized intrusion into applications [Gosling and McGilton, 1996]. Java promotes distributed computing as it supports HTML, XML, network protocols like HTTP or TCP/IP, and web services [Ullenboom, 2007]. For these reasons Java is very well suited for the development of web applications.
Java EE is a specification for developing distributed, portable and secure web applications [Jen-

drock et al., 2008]. The Java EE platform is intended for the implementation of complex enterprise applications which handle large amounts of data from a variety of sources. Applications implemented with the Java EE technology are fast, scalable and highly available. Additionally, such programs show improved performance, are less complex and require less development time. The Java EE platform provides a comprehensive set of application programming interfaces (APIs) for transaction management, database connectivity management, security and communication [Mukhar et al., 2005]. Moreover, it offers definitions which facilitate the deployment of the software. The application model of the Java EE platform splits application logic into components according to their function [Jendrock et al., 2008]. These components are encapsulated in multiple layers (Fig. 4.1):

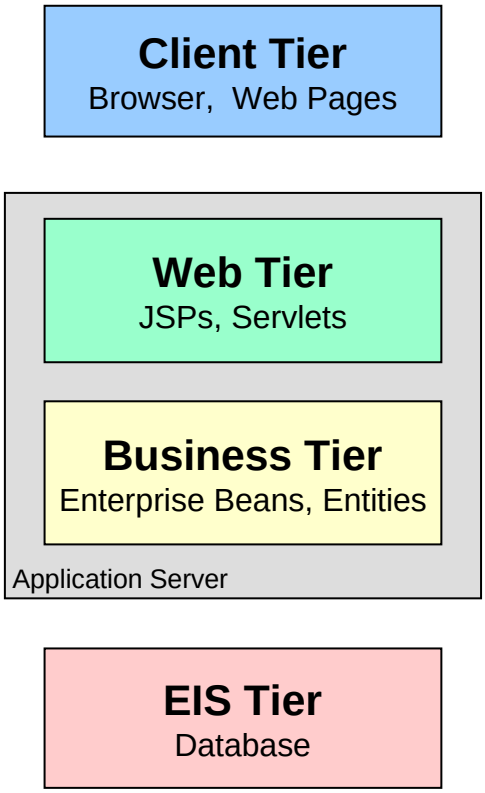

Figure 4.1: Multi-tier architecture. The client tier presents the data, the web tier processes requests from the client tier as well as data received from the business tier which represents the application logic [Jendrock et al., 2008]. Data are usually stored in a database.

- **Client tier** is responsible for presentation and user interaction.
- **Web tier** handles client tier requests and prepares data received from the business tier for presentation.

- **Business tier** implements the business processes of the application.
- **Enterprise information system (EIS) tier** is usually a database where the application data are persisted.

The division of the application into tiers facilitates maintenance because tiers can be modified or even exchanged without effecting other tiers. Moreover, a multi-tier application is scalable, as computational power can be specifically added at any part of the application, and robust, as there is not just one point of failure [Mukhar et al., 2005].
Java EE components of the different tiers may exist on different machines. The browser which displays dynamic web pages runs on the client machine. However, a Java EE client can also be for instance a Java GUI application or a Java console application. Web tier and business tier reside on the application server, whereas the EIS is usually located on a separate database server. The application server provides the connection between the business logic and the client application, and additionally offers infrastructure, like security, resource location or network communication, for Java EE components.
The components are managed in containers which provide a runtime environment as well as interfaces for the services offered by the application server. Containers are responsible for loading and managing components, and for starting application server services. They allow developers to concentrate on the business logic without bothering about basic essentials such as data access or communication between EJB components. Two different containers are provided by the application server: a container for web-tier components and a container for business-tier components. In order to deploy a Java EE application on the server, web-tier components and business-tier components are packaged separately with deployment descriptors and dependent library components into a web application archive (WAR) file, respectively an enterprise archive (EAR) file [Kodali et al., 2006]. Several application servers are compliant with the Java EE specification, e.g. JBoss Application Server [JBoss, 2010], GlassFish [GlassFish, 2010], Java Open Application Server [JOnAS, 2010] or IBM WebSphere [WebSphere, 2010]. These application servers offer an integrated web container such as Apache Tomcat [Tomcat, 2010].

Business tier

The business tier contains code that implements the functionality of the web application [Jendrock et al., 2008]. Data received from the client or the database are processed in the business tier. Data manipulations are carried out and the resulting data are either written to the database or returned to the user. The business tier components Java beans and entities will be elaborated in detail in the following sections.

Enterprise Java beans (EJB) are server-side components that encapsulate the business logic of a Java EE application. Usually, the business logic of an application is a composition of several EJBs. EJBs are distributed components as beans of an application can reside within different servers [Burke and Monson-Haefel, 2006]. Moreover, Enterprise Java beans are portable and may be reused in other applications. The EJB container loads beans when they are needed, invokes operations and supports transactions [Mukhar et al., 2005]. A major drawback of earlier versions of the EJB technology was its complexity. To overcome this inconvenient shortcoming EJB 3.0 was released [DeMichiel and Keith, 2006]. The API of EJB 3.0 is greatly simplified compared to previous versions and dispensable interfaces were eliminated [Burke and Monson-Haefel, 2006]. Meta data, e.g. for security or transactions, are now defined mainly using annotations within the bean class. Additionally, entity beans have been superseded by entities of the newly established Java persistence API (JPA).

There are two types of Java Enterprise beans:

- **Session beans** manage business processes. They interact with entities to create a task flow. In order to carry out an application task, the client calls the session bean's method and then the session bean performs the task on the server [Jendrock et al., 2008]. A session bean has exactly one client and cannot be shared amongst clients. Additionally, session beans are not persistent but may access data in the database. For this reason, session beans are well suited to insert, modify and read business process data [Burke and Monson-Haefel, 2006]. Two variants of session beans exist:
  - **Stateful session beans** are extensions of the client application. A stateful session bean is bound to one client for the entire lifecycle until the bean is removed by the client or the client session is closed. A stateful session bean remembers its state, which is represented by client specific instance variables, between method calls. This conversational state represents the communication between the client and the session bean and is shared between all bean methods which can read and write data to the state. To keep the conversational state, every method invocation of a client is served by the same instance of the stateful session bean. For this reason modifications to the bean's state acquired in a method can influence the outcome of subsequent method calls. Stateful session beans are usually used to manage processes, bean interactions and data access to conduct a complex set of tasks for one client e.g. a shopping cart.
  - **Stateless session beans** maintain a client specific state only until the method is finished and not between method invocations. Since a stateless session bean does not

keep a conversational state, a method call is independent of preceding method calls. Thus, in each method all steps of a task have to be performed and all the data necessary for the method execution have to be passed to the method by using parameters. When a method invocation is finished, the stateless session bean can serve a new client. Therefore, an instance of a stateless session bean can serve several clients which leads to better scalability of the application. Furthermore, as stateless session beans do not need to remember the conversational state, they are lightweight and fast. This type of session beans is designated for processes such as report generation, method processing or validation of credit card data.

- **Message driven beans** process messages asynchronously. A message driven bean listens for Java message service (JMS) messages which can be sent by any Java EE component. When the bean receives a message, the message is handled e.g. by calling a session bean to process the message data [Jendrock et al., 2008]. Message driven beans are stateless and therefore one bean can process messages from several clients. Furthermore, they are not persistent although they may access and update database data. Due to the asynchronous invocation of message driven beans, the server is not blocked and other tasks can be performed on the server while the message is being processed. A message driven bean can handle numerous messages simultaneously as many instances of a message driven bean can exist in the container at the same time.

Java Persistence offers support for persisting Java objects. It comprises the Java persistence API (JPA) and the Java persistence query language (JPQL) [Jendrock et al., 2008].

- **JPA** simplifies storage of Java entities in a database. Entities represent real-world objects like things, people or places [Burke and Monson-Haefel, 2006]. An entity can, for instance, be a customer, a car or a store. Entities are plain old Java objects (POJOs) with fields and methods describing attributes and behaviour of real-world objects. As opposed to session beans or message driven beans, entities are persistent. The data associated with an entity are stored in the database so that they can be accessed at any time even if the application has been restarted or a system error has occurred. Usually, relational databases are used to store data of Java EE applications. Therefore, the Java EE specification provides object-relational mapping to store the state of a Java entity bean in database tables. An entity corresponds to a table in the database and a row in that table corresponds to an instance of the entity. Database columns represent the attributes of the objects and a primary key is used to identify the different entities.
  The framework allows the definition of relationships between entities [Jendrock et al., 2008].

An instance of an entity may possess a relationship to an instance of another entity (one-to-one relationship) or can be related to several instances of other entities (one-to-many relationship). Moreover, many-to-many relationships can be defined where the instances of an entity are related to several instances of another entity.

- **JPQL** has an SQL-like syntax but operates on Java objects instead of database tables [Burke and Monson-Haefel, 2006]. In JPQL query definitions, properties and relations of entities are referenced in place of database tables and columns to which the entities are mapped to. The queries are then translated into SQL queries of the database schema and executed on the database. Queries defined in JPQL are independent of the database implementation and are thus portable. JPQL is a simple, powerful and flexible query language which is executed fast at runtime. Due to its object-orientation the language is better readable and more compact than SQL. As JPQL is portable between different databases, it is not able to exploit the features of individual database implementations. To solve this problem, an API which is used to write native SQL queries has been introduced.

Transactions encapsulate a set of tasks which are executed together, e.g. processing a flight reservation and the corresponding payment. If several components try to modify the same information at the same time or the system crashes while a transaction takes place, the integrity of the data is lost because the data are not up to date or only partially updated. To keep data in the database up-to-date and accurate, Java EE automatically manages transactions. A transaction is performed atomically, meaning that the transaction is only successful if every single task has been completed. In case of success the modified data are stored in the database. If a task fails, all the modifications of previous tasks performed within the scope of the transaction are rolled back.

Web tier

The web tier receives requests from the client and forwards it to the business tier. Dynamic web pages are created according to the response from the business tier and returned to the client.
The two commonly used Java EE web components Servlets and Java Server Pages are presented in the following sections.

Java Server Pages (JSP) is a technology to develop web pages with dynamic content [JSP, 2010]. JSPs were developed to increase the efficiency of web page creation by separating the presentation from the data [Bergsten, 2004]. The separation of presentation and content allows the developer to edit the layout of the web page without impacting content generation. A JSP web page is composed of HTML code with embedded JSP code. The markup language HTML is used

for formatting the static part of a web page, while the dynamic content in a web page is created with JSP tags or small scripts. JSP tags provide an interface to functionality and objects [Brown et al., 2001]. They encapsulate complex code, e.g. for controlling the flow of the application or accessing Java beans. Many tags have already been developed and can be freely used, or developers may as well create custom tags for their own code. Tags promote code reuse, speed up the development process of a web page and make the code lucid and easy to maintain. Usage of tags does not require programming knowledge, therefore even non-Java programmers can work with them on web pages. Additionally, scripting elements allow the developer to add pieces of code to a JSP page. Such scripts include expressions for evaluating objects, comments, method and variable declarations as well as short code segments.
When a user accesses a JSP for the first time after its creation or modification, it is compiled to a Java Servlet. A Servlet is a Java class that handles client requests. Servlets read client data from the user, which are e.g. entered via web forms, as well as data received from the browser such as media type information [Hall and Brown, 2004]. In the next step, Servlets process the data by, for instance, forwarding them to the business tier. Finally, the Servlet creates the web document, usually in HTML format, which includes the result of the data processing. The document is then sent together with implicit information like updated cookie data to the client. As web browsers communicate with web servers using the HTTP protocol, Servlets receive data via HTTP requests from the client and send data via HTTP responses back to the client. Servlets offer a comprehensive infrastructure to parse and decode data from web forms and to deal with cookies. Moreover, they are able to store data between requests to maintain sessions and to cache previous calculations. The lifecycle of a Servlet is managed by the web container which also assigns requests to the corresponding Servlets.
Servlets are not just compiled Java Server Pages, they can also be written by developers to complement JSPs. While Servlets are suitable for data processing operations, JSPs are used for data presentation.

Apache Struts is an open source framework for the development of Java web applications [Struts, 2010]. The framework is based on technologies like Servlets, Java Server Pages, Java beans and XML and combines these components to facilitate the development of web applications [Husted et al., 2003]. Struts implements the Model-View-Controller (MVC) web application architecture (Fig. 4.2). The *model* represents the code of the application functionality, the *view* is the presentation code and the *controller* handles the communication between the *view* and the *model*. The *model* and the *view* do not interact directly.
The *controller* defines the flow of the web application by specifying the way in which the appli-

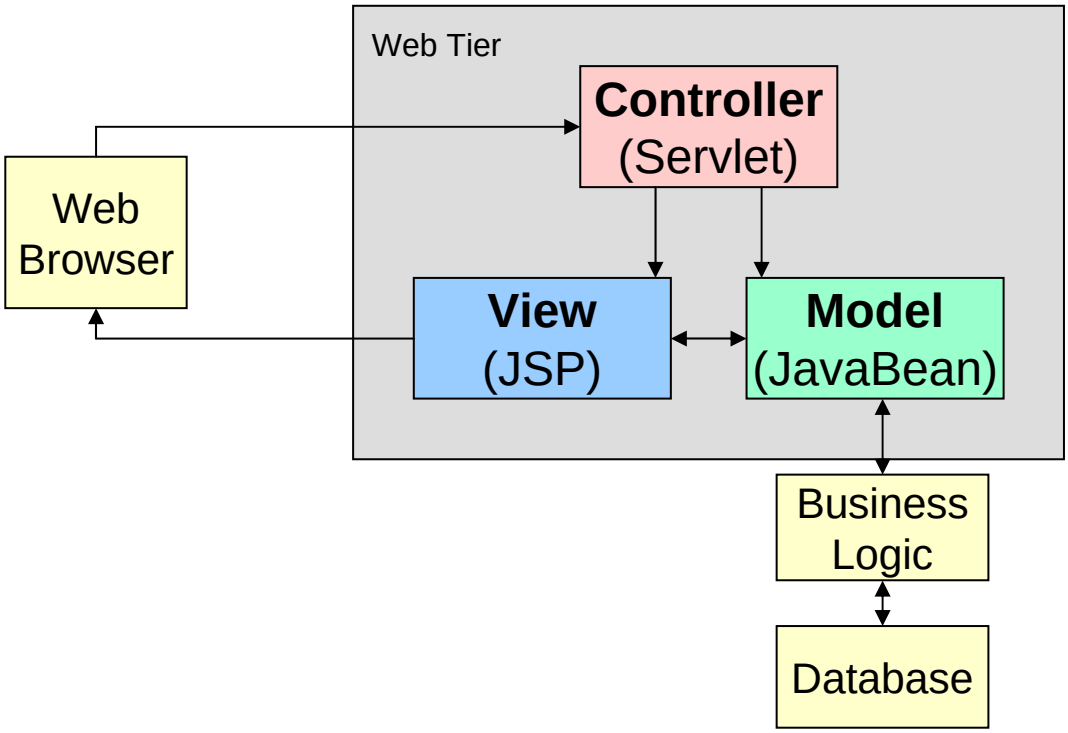

Figure 4.2: Model-view-controller architecture. The *model* comprises the application functionality, the *view* represents the presentation and the *controller* manages the communication between *model* and *view* [Struts, 2010].

cation communicates with the user. A controller component of the web application is the action Servlet. When the client sends a request, the container forwards it to the action Servlet and the Servlet uses an action mapping file to look up which action class needs to be executed. The different action classes contain the business logic of the application. They perform calculations, update sessions and request data or access Java beans of the business tier to process or retrieve data. Depending on the result of the action, the Servlet either forwards the response to another action class, renders a JSP page containing the generated data or sends the request back to the client with instructions, such as error messages or directives. Input data from the *view* such as web form data are stored in action forms. An action form is a Java bean which has properties that correspond to fields of the web form. This bean is then accessed by action classes for data processing. To determine the correctness of user input, action forms can be validated by checking e.g. if the phone number has enough digits or if a character has been entered to a numeric field. Struts provides JSP tag libraries for commonly used standard features e.g. creating an input field [Brown et al., 2001]. Four libraries are included in Struts 1.0: the HTML library, the bean library, the logic library and the template tag library. The HTML library offers tags for generating dynamic

HTML web forms such as input fields which are mapped to attributes of action forms. To access and manipulate Java beans within a Java server page, the bean library is used. The presentation flow of a JSP can be controlled with the logic library which offers comparison tags as well as iterators over data types like collections and arrays. Finally, the template tag library facilitates the construction of web pages with a common format from individual JSP components.
Summarizing, the Struts framework supports the development of robust web applications. The MVC architecture separates the presentation from the business logic and therefore makes a Struts application easy to maintain and extend as modifications or even the exchange of one part of the application do not effect the other parts. Struts reduces the development time of web applications as the developer does not need to care about the separation of presentation and business logic. Moreover, Struts offers internationalization features which adjust the content of the web page according to the user's origin.

4.1.2 Relational database management system (RDBMS)

A relational database management system (RDBMS) is used to store application specific data. A database management system (DBMS) consists of the database and applications to create, manage and administer the database [Sumathi and Esakkirajan, 2007].
The aim of a DBMS is to achieve

- **Data availability**: the data are offered in a meaningful format and are easily accessible to a large variety of users.
- **Data integrity**: the data stored in the database are correct.
- **Data independence**: the data are managed and stored separately from the programs that access the data.
- **Data security**: the data can only be retrieved from the database by authorized users and conflicts between users who try to access the same data simultaneously must be prevented.

A database is a collection of related data [Prigmore, 2008]. Database data are persistent, as they are stored until they are deleted or overwritten, and structured to facilitate data access. Databases contain not only the actual data but also meta data which describe the stored data. The concept of a relational database was first described by E.F. Codd in 1970 [Codd, 1970]. A relational database is a set of interconnected tables where a table represents a set of data objects. The rows correspond to the objects whereas the table columns correspond to the object attributes. To identify each table row primary keys are used. Foreign keys are used to define different kinds of

relationships between tables like one-to-one, one-to-many or many-to-many. A foreign key is a primary key from one table which is inserted as an attribute into another table.
In order to avoid redundancies in the database schema which may cause insertion, update or deletion irregularities, a database needs to be normalized [Codd, 1990]. Normalization is the process of reorganizing the information in the database by splitting relations following defined criteria. Currently, there are 6 normal forms representing increasing normalization levels [Sumathi and Esakkirajan, 2007]. For average databases, the third normal form is considered to be sufficiently well structured as too many relations may result in a performance loss of the database.
In order to access and manage the database and contained data, the structured query language (SQL) can be used. SQL comprises three different subsets:

- Data Definition Language (DDL) for modifications of the database schema. DDL statements are used for creation, modification and deletion of database tables and their relationships.
- Data Manipulation Language (DML) for data processing. DML offers commands to insert, update, delete and retrieve data from the database.
- Data Control Language (DCL) for management of data access. DCL instructions are used to grant or revoke user access rights.

SQL is not a full programming language but a data sublanguage as it only provides instructions to create and process database data and meta data but lacks commands for conditional logic and loops [Kroenke, 2005]. Data sublanguage instructions can be embedded into code written in programming languages such as Java or carried out using applications of the DBMS. The structured query language is a nonprocedural data language meaning that SQL statements specify only the necessary input as well as the required data output and the DBMS decides the way in which the data are retrieved [Beaulieu, 2009]. The majority of RDBMS implement SQL but each system uses its own dialect.
Some of the most popular relational database management systems are MySQL [MySQL, 2011], PostgreSQL [PostgreSQL, 2011], Oracle DBMS [Oracle, 2011], Microsoft SQL Server [Microsoft, 2010] and IBM DB2 [IBM, 2010].
Java applications can access a database using the Java database connectivity (JDBC) API. A JDBC driver opens a connection to the data source, sends an SQL query and processes the result [JDBC, 2010]. JDBC drivers are RDBMS specific. To exchange the database behind a Java application no alterations to the application are necessary; only the JDBC driver has to be replaced with the JDBC driver of the new database.

4.1.3 Model driven architecture (MDA)

In 2001, the usage of models in software development was formalized through the definition of MDA by the Object Management Group [OMG, 2010a]. A model is a textual and graphical representation of the structure, function and behaviour of an application. Model driven architecture uses models to separate ”business and application logic from underlying platform technology” [OMG, 2010b]. Furthermore, the MDA approach deals with the various phases of the entire development lifecycle of an application such as design, implementation, component assembly, testing, deployment and maintenance [Truyen, 2006].
The development of an application using MDA involves four consecutive steps (Fig. 4.3) [Siegel, 2001] [Miller and Mukerji, 2003].

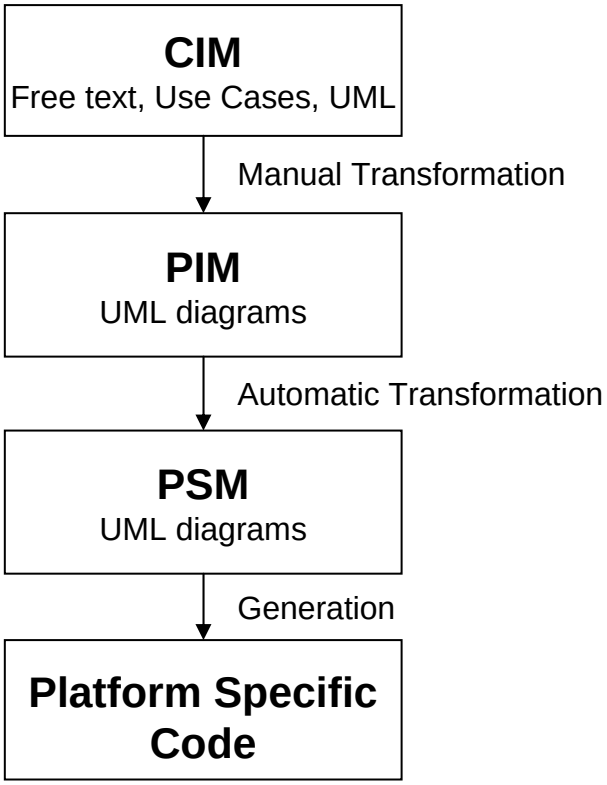

Figure 4.3: Model driven architecture. The platform independent model (PIM) can be retrieved from the computation independent model (CIM) and is transformed automatically into the platform specific model (PSM). Finally, the platform specific code is generated from the PSM.

1. **Computation independent model (CIM)** describes the system's requirements without any information technology details. The CIM can be represented as free text, use case diagrams, UML diagrams, etc. This step is optional but the CIM is useful for understanding the problem

and facilitates the communication between domain experts and application developers. The requirements of the computation independent model should be traceable to the models of the following steps (PIM and PSM) and vice versa.

2. **Platform independent model (PIM)** describes the business functionality of the application. The model is independent of the application's platform e.g. Java EE, Spring or .NET. It is created using the Unified Modeling Language (UML). The structure of the application is modelled with class or object diagrams whereas the behaviour is defined with activity or sequence diagrams.

3. **Platform specific model (PSM)** is derived from the PIM to specialize for a certain platform. The PSM is also expressed in UML which provides extensions like stereotypes or tagged values to define an environment tailored to a certain platform.

4. **Platform specific code** is then generated from the PSM. Finally, this code can be extended by the developer.

Model driven architecture facilitates the definition of machine readable models of applications. The use of MDA allows the developer to concentrate on the implementation of the business logic because MDA takes care of generating the basic code skeleton of the application. Models may be reused to develop applications on other platforms or with newer technologies. As a model provides an easily readable overview of the application structure, maintenance is simpler in MDA applications compared to traditional applications. Moreover, the model driven approach speeds up the development process because it spares the developer from writing repetitive code.

AndroMDA

AndroMDA is an open source code generation framework which follows the MDA approach [AndroMDA, 2010a]. The framework takes models as input and builds custom code components from these models using plug-ins such as cartridges or translation libraries. Generated code can be created in any programming language but AndroMDA is mainly used for Java EE projects.
To create an application, a model of the application logic is created by the developer using an UML tool. As the UML model is platform independent, it must not contain items related to any platform such as programming language specific data types. Instead, generic data types are required by AndroMDA. The model is saved in the XML Metadata Interchange (XMI) format which is then read by AndroMDA. The Velocity template engine [Velocity, 2010] integrated in AndroMDA uses cartridges to transform the PIM into a PSM and then generate platform specific code. A cartridge

is a set of code template files and helper classes. For each class in the model, the corresponding cartridge templates are identified and used for code generation. One or, if necessary, several files with code can be created for a class in the model.
AndroMDA provides predefined cartridges for different technologies like Hibernate, Spring, JSF, EJB or Struts, and the developer can write custom cartridges or extend existing ones. The EJB3 cartridge is used to develop the business tier of a web application. It generates Java business logic and persistence code, such as session beans or entities, which is conform to the EJB3 standard. AndroMDA produces two types of files: code that is overwritten in every code generation run and code that is generated only once when the code is retrieved from the model the first time. Code files which are constantly overwritten are not meant to be modified by the user and automatically adapt to changes in the UML model. The implementation of the application functionality is therefore included into code files that are generated only once. Furthermore, AndroMDA is capable of creating a database schema which may be used to initialize the database of the application.

4.1.4 Web page presentation

The following technologies were used to design the web page layout of the GiSAO.db web application.

Cascading style sheets (CSS)

CSS, which was developed by the World Wide Web Consortium (W3C) [Schmitt et al., 2008], is a markup standard set to define the presentation of web documents. Cascading style sheets facilitate the application of a consistent style to all the web pages of an application and are used together with HTML to design the layout of a web page. HTML is a structural language that is very well suited for the description of page parts, like tables, paragraphs or hyperlinks, but also offers presentation options such as setting text size or background colour [Meyer, 2006]. However, when HTML is used not only for structuring a page but also for designing a page, the code is barely readable due to numerous markup tags. Moreover, the style of each HTML element on each page of the application has to be defined separately.
For these reasons, cascading style sheets have been developed. CSS allow the definition of styles of elements in an external file and to name these styles. The name of the style is then assigned to all HTML tags that are supposed to have the same look. If the developer wants to change the appearance of the web application, the style has to be changed only once in the CSS file and not in every HTML document. Not only a web developer can create style sheets, some browsers allow the user to write style sheets which e.g. make the text larger or place the favourite picture in the

document background. In case of conflicting style sheet instructions caused by competing style sheets of the developer or of the developer and the user, CSS offer cascade rules, that define which style is finally applied to an element. When CSS are used, the code is neater and easier to maintain as less markup tags are used. Additionally, the file size is reduced and the load time is decreased. Furthermore, cascading style sheets provide richer presentation features than HTML like setting the text colours or the background of any HTML element.

Asynchronous Javascript and XML (AJAX)

AJAX is a collection of technologies which are combined to improve the usability of web applications. The aim of using AJAX is the creation of dynamic web sites that are not just a set of HTML pages but which act similar to traditional desktop applications [Murray, 2005].
Technologies incorporated by AJAX [Garrett, 2005] are

- **XMLHttpRequest** for asynchronous data transfer from the server to the client
- **XML and XSLT** for exchange and manipulation of data
- **DOM (Document Object Model)** to change objects on the web site
- **CSS and XHTML** for the presentation
- **Javascript** to connect all technologies

AJAX changes the way the web browser and the server communicate with each other. On a conventional web site each request to the server results in a complete reload of the page which means that the user has to wait until that page has been completely loaded in the browser. Web applications which are based on AJAX can send requests to the server and when the data are returned from the server, parts of the web page are updated without a complete page reload. AJAX requests are asynchronous, meaning that they do not block the browser. The user can still use the web site while the request is being processed in the background [Gehtland et al., 2006].

Script.aculo.us [script.aculo.us, 2010] is a Javascript framework for developing AJAX web applications based on the Prototype framework [Prototype, 2010]. Script.aculo.us provides GUI elements like drag and drop fields, autocompletion fields or in place editing fields. In addition script.aculo.us offers visual effects which can be added to HTML elements.

4.1.5 Authentication and authorization

To manage the user access rights to a web application, methods for authentication and authorization are required. The authentication and authorization system (AAS) is a Java application which

was developed at the Institute for Genomics and Bioinformatics [Zeller, 2003][Maurer et al., 2005]. It stores user accounts in a database and offers a web interface to edit user settings as well as application settings.
Software libraries enable the integration of AAS into web applications so that it is not necessary to store access rights data in the application database. Authorization rights can be configured through access control lists (ACLs) which define read, write and delete rights of users. ACLs allow the assignment of application access rights to single users as well as to entire groups of users. These rights can be defined for individual parts of the application separately in various levels e.g. a group has only read rights or a user has no delete rights.
The authentication and authorization system provides a custom JSP tag library to control the access to sections of web pages. The user can only view the respective part of the web page if logged in and previously got assigned sufficient access rights.
The AAS was developed to manage users of several web applications. It offers single-sign-on functionality which means that a user has to log in only once and is then able to switch between all managed web applications for which she has access rights. Furthermore, the authentication and authorization system contributes to the database security as it prevents unauthorized access.

4.1.6 Development tools

Appropriate tools facilitate and accelerate the development of web applications. Amongst others the following tools were used for the construction of GiSAO.db.

Eclipse

The projects of the Eclipse open source community aim at providing a flexible and extendable development platform as well as application and runtime frameworks to manage software within the complete development lifecycle [Eclipse, 2011]. The main project is the Eclipse software development kit (SDK) with an integrated development environment (IDE) that offers built-in Java support. Several editors for different file types, such as Java, HTML and JSP, are included in Eclipse IDE [Eclipse, 2006]. These editors facilitate coding by providing syntax and keyword highlighting, display of compiler problems and code completion. Comprehensive search functions are integrated in Eclipse as well as functions to navigate between source code files. The created applications can be compiled and executed within Eclipse. Additionally, the IDE offers debugging features for fast error detection.
Each developer can create a custom workspace to manage all resources [Holzner, 2004]. To keep an overview of all the projects in the workspace, Eclipse provides different views which for

instance display only Java classes or allow navigation between different projects.
The Eclipse IDE can be extended with various plug-ins for other programming languages, e.g. C, Perl or PHP, or for tasks such as documentation or testing. MyEclipse is a plug-in for the development of Java EE applications which offers database tools, a visual tool for web design as well as the possibility to connect to application servers [MyEclipse, 2011].

MagicDraw

MagicDraw is a Java based computer-aided software engineering (CASE) tool for design and analysis of object oriented systems [MagicDraw, 2010]. The main focus of MagicDraw is the visualization of UML models. However, MagicDraw also supports the generation of code in various programming languages, such as Java, C++ or C from UML models, as well as the modelling of database schemas. Additionally, it provides reverse engineering facilities to create UML models from existing source code.
Various types of diagrams are offered by MagicDraw to model a system [NoMagic, 2011]. One of the most important diagrams is the class diagram which is used to define attributes of objects, operations for editing objects and associations between objects. For modelling the behaviour of a system, MagicDraw provides use case diagrams, activity diagrams which describe the system processes as well as communication diagrams and sequence diagrams which define object interactions.

Oracle SQL Developer

Oracle SQL Developer is a freely available tool for database management [SQLDeveloper, 2011]. The tool has a graphical user interface to browse databases and the contained tables. Database data and meta data can be viewed, inserted, edited and deleted. It is possible to execute, edit and debug SQL instructions and SQL scripts, export data and create database status reports. Although SQL Developer is an Oracle product, it allows the management of various third-party databases such as MySQL, Microsoft SQL Server and IBM DB2. Moreover, third-party databases can be migrated to the Oracle DBMS using SQL Developer.

4.2 Integrative data analysis

This section provides information about methods and tools that were used for the analysis of ageing-related gene expression data to gain previously unknown biological insights.

4.2.1 Statistical analysis

High-throughput technologies such as microarrays have led to a sharp increase in the amount of biological data. These data pose various challenges for data analyzers [Lee, 2010]:

- **Noise** in high-throughput data results from biological sources, such as variations in cells, as well as from technical reasons like sample preparation or imperfections in the technique.
- **Large number of parameters but small sample size** due to limited biological material and high experimental costs. For instance, the expression values of thousands of genes is measured in one microarray experiment but usually only a few experiments using independent samples are performed.
- **Multiple Testing** of a statistical hypothesis in parallel leads to a large number of false positives. For example, if a gene is classified as differentially expressed with a significance level of 0.05, 5% are false positive results. In case of thousands of genes on a microarray, 5% yields a large number of genes.

Statistical methods which consider the above mentioned obstacles are used to extract relevant information from a huge quantity of biological data.

R

R is a computer language for statistical data analysis [R Development Core Team, 2007]. It is a high-level programming language that resembles the S language developed at the Bell Laboratories and provides constructs like functions, loops and conditionals. Moreover, R is a framework which offers a large number of statistical methods such as hypothesis tests or linear modelling, as well as graphical methods to plot data in the form of e.g. histograms or dot plots [Verzani, 2005]. The R environment can be used to perform calculations as well as data manipulations. It is suitable for data storage and has implemented operators for array and matrix calculations.
The platform independent R software is open source and can be extended easily. Although a large number of statistical techniques are implemented in the R base environment, numerous packages containing additional functions are available. The Bioconductor project provides software for computational biology and bioinformatics, especially for the analysis of high-throughput

data [Gentleman et al., 2004]. Bioconductor offers for instance packages for preprocessing of microarray data and identification of differentially expressed genes. Another Bioconductor module is the qvalue package with functions to correct for multiple testing by controlling the false discovery rate (FDR) [Storey and Tibshirani, 2003]. The input is a list of p-values from which the q-values are computed. The q-value is a measure of the false discovery rate, which gives the expected number of null hypothesis that are true but were rejected in the hypothesis test among all the null hypothesis that were rejected in the test. Additionally, graphical representation of q-values and p-values, like histograms or graphs that plot q-values against the corresponding p-values, are offered by the Bioconductor qvalue package.

4.2.2 Pathway analysis

A biological pathway describes a biological process with molecule relationships that fulfil a specific function [Cary et al., 2005]. Entities, like DNA, RNA or proteins, and interactions between them, such as activation, binding or synthesis, make up a pathway [Viswanathan et al., 2008]. Diagrams are used to display a pathway graphically. A biological pathway is represented by a two dimensional network with the molecules as nodes and the interactions between them as edges [Tao et al., 2004].

So far, numerous pathways of various biological processes involved in e.g. metabolism, signalling, gene regulation or diseases have been discovered. To make these pathways available to the public, several databases, such as KEGG Pathway Database [Kanehisa and Goto, 2000], BioCarta [BioCarta, 2010] or GenMAPP [Dahlquist et al., 2002], have been established. These databases contain pathway information and visualizations. The pathway information is usually derived from literature research and the pathway diagrams are either drawn manually or generated with graphic tools.

An approach to analyse gene expression data is the integration with previously acquired biological data [Mlecnik et al., 2005]. By correlating gene expression data with pathway data, gene expression profiles can be visualized and interpreted in the context of biological processes.

To detect and examine connections between genes, the gene expression values of a data set are mapped to pathway diagrams obtained from pathway databases. For a better insight into biological processes at the time point or state of the experiment, it is possible to extract those pathways which are over-represented in a specified gene set.

Integrating colour-coded expression values, which reflect the expression array patterns, into pathway diagrams supports visual analysis of gene expression data [Nakao et al., 1999].

PathwayExplorer

PathwayExplorer [Mlecnik et al., 2005] is a platform independent Java web application that maps gene or protein expression profiles onto pathways provided by KEGG, BioCarta and GenMAPP.
A text file containing the gene expression values is uploaded [Mlecnik et al., 2005]. The data set can be filtered to obtain for instance differentially expressed genes or those genes with sufficient data points. The resulting subset can then be mapped either to a single pathway or to all available pathways. In the case in which the data set is mapped to all available pathways, a ranked list with all mapped pathways is displayed. The list can be sorted based on several criteria, e.g. number of unique genes of the dataset that where mapped to the pathway or the p-value of a Fisher's exact test. Fisher's exact test is used to determine if the ratio between genes, which are differentially expressed and could be mapped on the pathway, and genes, that are not differentially expressed and mapped on the pathway, is significant.
The PathwayExplorer application also produces graphical output. For every pathway, on which genes of the dataset were mapped, an image showing the pathway with colour-coded boxes that represent the mapped genes is created. The colours of the boxes correspond to the gene expression values. The pathway graphs can be exported in PNG, JPEG or vector graphics (SVG) file format.

4.2.3 Cluster analysis

Clustering is a technique that groups together items which have similar properties [Brazma and Vilo, 2001]. Genes can be clustered according to the patterns of their expression values in different samples [Jiang et al., 2004]. A cluster contains co-expressed genes whereas genes with differing expression patterns are in different clusters. Cluster analysis supports the interpretation of large scale gene expression data since co-expressed genes may be co-regulated and often have similar functions or are involved in the same processes.
Moreover, not only genes but also samples can be clustered to detect correlations between samples. To determine the likeness of gene expression patterns similarity distance measures, e.g. the Euclidean distance or Pearson's correlation coefficient, are used [D'haeseleer, 2005]. Popular algorithms for clustering gene expression data are hierarchical clustering, k-means and self-organizing maps.
Hierarchical clustering creates nested clusters which are graphically displayed as a tree [Jiang et al., 2004]. There are two approaches for hierarchical clustering: agglomerative or divisive. For the agglomerative version, the algorithm defines each gene as a cluster. Gradually, the nearest clusters are aggregated until all genes reside in a supercluster. The divisive approach begins with

a supercluster and splits it until each gene has its own cluster. The result of hierarchical clustering is a dendrogram with branches that reveal the resemblance between the clusters.
K-means clustering sorts the genes into a predefined number k of clusters. First, k cluster centers are determined randomly and every gene is assigned to the cluster of the nearest cluster center. Then the cluster centers are recalculated to minimize the distances between the genes assigned to a cluster and the cluster center. Again, all the genes are assigned to the nearest cluster center and the cluster centers are newly calculated. This is repeated until the genes in the clusters do not change anymore.

Genesis

Genesis is a stand-alone Java application for cluster analysis of gene expression data [Sturn et al., 2002]. The application offers a large number of similarity distance measures including Euclidean distance, Manhattan distance or Spearman's rank correlation coefficients. Hierarchical clustering, k-means clustering, self organising maps, principal component analysis, correspondence analysis and support vector machines are the clustering algorithms provided by Genesis. Hierarchical clustering can be performed for genes, experiments and both. Three different methods to measure the distance between hierarchical clusters are offered by Genesis:

- Average linkage clustering: the distance between two clusters is defined as the mean distance between all possible element pairs, where each element belongs to a different cluster [Keith, 2008b].
- Complete linkage clustering: the inter-cluster distance is the distance between those two elements, each from a different cluster, which are furthest from each other.
- Single linkage clustering: the distance between clusters is the distance between the two nearest elements where each element belongs to a separate cluster.

Genesis facilitates the application of different clustering methods on the same data set and the comparison of the clustering results. Furthermore, a figure of merit can be computed for a data set to estimate the number of clusters which has to be prespecified for some clustering algorithms such as k-means.
Genesis takes a flat file with gene expression values as input. Gene expression values can be normalized and filters can be used to extract genes of interest. Genesis displays gene expression values in a matrix with genes in rows and experiments in columns. Moreover, the application provides intuitive visualizations of clustering results.

4.2.4 Promoter analysis

Gene expression is mainly regulated by transcription factors [Latchman, 1990]. Transcription factors (TF) are proteins which activate or repress the transcription of a gene by binding to specific DNA sequences, the so called transcription factor binding sites (TFBS). TFBS are DNA sequences which are approximately five to fifteen base pairs (bp) long [Bulyk, 2003]. Sequence motifs are often degenerate. The degree of degeneracy affects the level of transcription, i.e. the amount of gene products. Binding site degeneracy also implicates that there are a huge number of potential binding sites in the human genome. However, functional TFBS are usually located within a gene's promoter regions, which reside in the proximity of the transcription start site (TSS) of a gene, but binding sites can also be located in introns or far away from the gene upstream as well as downstream.

By determining binding sites of transcription factors in the promoter region of a gene, potential regulators of the gene can be identified. Due to degeneracy of binding sites, it is not possible to detect TFBS by searching for strict sequence identity. Therefore, position weight matrices (PWMs), also known as position specific scoring matrix (PSSM), are used to predict binding sites in DNA sequences [Stormo, 2000]. A PWM describes a transcription factor binding site. The matrix has four rows, one for each base, and the number of columns correspond to the length of the TFBS (Fig. 4.4). The matrix values represent the chance of the bases to be located at a certain

	1	2	3	4	5	6	7	8	9	10	11	12
A	5.81	0	0	0	22.49	4.03	1.41	0	0	1.21	0.78	13.1
C	2.92	0	0	0	0	12.56	0	5.94	6.9	20.68	21.71	4.81
G	9.86	22.49	22.49	22.49	0	0	0	0	0	0	0	2.6
T	3.31	0	0	0	0	5.31	21.08	16.55	15.59	0	0	1.98

Figure 4.4: Example of a position weight matrix. The matrix from the TRANSFAC database [TRANSFAC, 2010] represents binding sites for the transcription factor NF-κB.

position of the binding site. These values are calculated from the frequency of the bases at each position of experimentally verified binding sites of a transcription factor. To detect binding sites of a certain transcription factor in the promoter sequence of a gene, the promoter is scanned for matches with the position weight matrix corresponding to the transcription factor. The similarity between the matrix and a DNA sequence is determined with a score that is computed as the sum of the matrix values which correspond to the investigated sequence. As a PWM gives the complete distribution of the nucleotides for every position, it provides for every possible sequence the likelihood of being a transcription factor binding site [Keith, 2008a]. In order to decide if the

sequence is a potential binding site, a suitable threshold for the score in terms of sensitivity and specificity has to be defined. Although TFBS can be computationally predicted by using PWM, their functionality has to be experimentally verified using biotechnologies such as chromatin immunoprecipitation (ChIP) or electrophoretic mobility shift assay (EMSA) [Cartharius et al., 2005]. Several public databases such as TRANSFAC [Wingender et al., 2000] and Jaspar [Portales-Casamar et al., 2010] provide PWMs.

Over-representation analysis

To determine common regulators of a set of genes over-representation analysis can be performed.

1. A set of genes which are expected to be regulated by the same transcription factors are assembled. The assumption that a common regulator exists is based on common features of these genes such as co-expression or orthology [Keith, 2008a].

2. The promoter sequences of these genes are searched to identify transcription factor binding sites which are significantly over-represented in the gene set [Zambelli et al., 2009]. If binding sites of a certain transcription factor occur in the investigated sequences more often than in some random sequence set, it can be assumed that the transcription factor in question actually regulates the genes of the input gene set.

In order to perform over-representation analysis various publicly available web based tools like Pscan [Pscan, 2010] or ORA [Hackl, 2010] are available.

Pscan is an application to detect significantly over- or under-represented motifs in a set of sequences [Zambelli et al., 2009]. Human, mouse, yeast, fruit fly and thale cress sequences can be explored for motifs.
Pscan provides promoter sequences from RefSeq, TAIR as well as SGD and accepts the corresponding Ids as input. The user can specify the length of the analysed region which may be up to 1000 base pairs at maximum upstream of the transcription factor binding site and 50 base pairs at the longest downstream. Moreover the user chooses whether to use TRANSFAC 7.0, Jaspar or custom PWM. The tool determines for every sequence a score which correlates to the likelihood of a transcription factor to bind to this sequence. The analysis result is a list of those position weight matrices, which yielded matches in the input sequences, ranked by their p-values. The p-value is obtained by a z-test and corrected for multiple testing using Bonferroni correction. The z-test evaluates for each matrix the probability that the likelihood score is the same for the input gene set and a random gene set.

ORA is a tool not only for determining over-represented transcription factor binding sites but also for finding over-represented miRNA target sites and over-represented gene ontology terms in a given set of data [Hackl, 2010]. ORA can analyse mouse, human and rat data and uses TRANSFAC 6.0 matrices as well as Jaspar PWM and custom PWM. It takes RefSeq Ids as input. Two different significance tests are provided: Fisher's exact test and cumulative binomial distribution. Since the input can be up to 1000 RefSeq Ids, methods for multiple testing corrections are provided: False discovery rate, Bonferroni and Bonferroni step down.
Over-representation is calculated by comparing the number of found items, i.e. transcription factor binding sites of a certain transcription factor, in the input data set with all genes in the RefSeq database. The output is a text file with a list of transcription factors, for which binding sites could be determined in the sequences, sorted by their multiple-testing corrected p-values.

Transcription factor binding site prediction

In order to scan a sequence for binding sites of a specific potential transcription factor the programs MatInspector [MatInspector, 2010] and Match [Match, 2010] are available. Both tools use position weight matrices for the prediction of TFBS and a similar scoring mechanism to evaluate how well the sequence and the matrix match. Two scores which are computed in a similar way by MatInspector and Match are used to determine the quality of a match between a PWM and a sequence [Kel et al., 2003] [Quandt et al., 1995]:

- **Core similarity score (CSS)** is calculated from the core positions of the matrix. The core positions are the most conserved consecutive matrix positions. For the determination of the CSS MatInspector takes the four most conserved consecutive positions, while Match considers the five most conserved consecutive positions.
- **Matrix similarity score (MSS)** is computed with regard to all positions of the matrix.

The scores, which are both calculated using the same formula, reach from zero to one. A score of one is a perfect match, meaning that each nucleotide of the scanned sequence is equal to the most conserved nucleotide at the corresponding matrix position. For each position weight matrix thresholds are defined for matrix similarity score and core similarity score to determine a match. In the first step the CSS of a matrix for each position of the input sequence is calculated. In case the core similarity score of the sequence position is above the CSS threshold, the MSS is calculated. If the matrix similarity also exceeds the MSS threshold, it is called a match.
The CSS is a selection criterion as only core matches are considered further. This reduces the total amount of matches and thereby improves the performance of the tools.

MatInspector is a web based tool to identify potential transcription factor binding sites in promoter regions [Cartharius et al., 2005]. Matrices in the MatInspector position weight matrix library are automatically generated and evaluated. They are based on nucleotide distribution matrices from databases like TRANSFAC or binding sites found in publications. A position weight matrix is only added to the MatInspector library if it is of high quality, i.e. the matrix is retrieved from at least four verified binding sites and produces less than five matches in a random 1000 bp sequence.
MatInspector provides an optimized threshold for the matrix similarity score. It is less likely that a fixed cut-off for the matrix similarity score is reached for long PWM and for matrices with many conserved positions and thus too many false negative matches would be reported. On the other hand, very short and little conserved matrices result in too many false positive matches. Therefore, a threshold to compensate different matrix sizes and conservation was introduced. The threshold is computed for each PWM individually and allows in 10000 bp of non regulatory sequences only three matches.
To filter out redundant matches, matrix families are used. Redundant matches are caused by several similar PWMs that correspond to a single transcription factor. These similar matrices lead to several matches at the same position, shifted matches, matches which are overlapping or matches which vary only in length. For that reason, similar matrices are grouped to form a matrix family and in case of match redundancy only the match with the highest score in the family is reported.
MatInspector takes promoter sequences of any length which are retrieved from the ElDorado genome database [ElDorado, 2010] as input.
In the next step, the user specifies the PWM families corresponding to the transcription factors of interest. The result of a MatInspector analysis is a list of potential transcription factor binding sites denoting the input sequence name, matrix family, matrix name, optimized threshold, position, strand and sequence of the binding site as well as core similarity and matrix similarity.
MatInspector is accompanied by several tools. One tool facilitates the design of regulatory sequences for experiments that determine binding site functionality by removing unnecessary potential binding sites. Moreover, it is possible to determine common TFBS in a set of sequences and to identify binding sites which are localized at the same position in aligned promoter sequences of ortholog genes.

Match is a TFBS prediction web application which is directly connected to the TRANSFAC 6.0 position weight matrix library [Kel et al., 2003]. The user can enter either a fixed threshold for the core similarity score and the matrix similarity score or select a predefined threshold depending on her objectives. The tool provides three score thresholds:

- Threshold to minimize the false negative rate
- Threshold to minimize the false positive rate
- Threshold to minimize the sum of false negatives and false positives

Match takes DNA sequences as input and scans these sequences with TRANSFAC PWMs which correspond to the user specified transcription factors. The result of the search is a file that lists all potential transcription factor binding sites which exceeded the chosen cut-off. A list entry contains the Id of the sequence and the matrix, the position of the match with the strand, the matrix similarity score and the core similarity score, the sequence of the match and the transcription factors associated with the matrix. Additionally, Match provides a graphical representation of the locations of the predicted binding sites.

4.2.5 Gene co-expression network

A gene co-expression network is an undirected graph which shows the degree of relations between several genes based on expression values [Bhattacharyya and Bandyopadhyay, 2009]. The network nodes represent the genes and the edges correspond to the co-expression relationships between the genes. The construction of a co-expression network encompasses two steps:

1. The similarity of genes is determined by the correlation between their expression profiles obtained from samples in various conditions. For a gene co-expression network, the correlation between every possible gene pair of a set of genes is determined with a correlation measure, e.g. the Pearson correlation coefficient [Ruan et al., 2010].
2. Gene relationships, which exceed a predefined correlation threshold, are then used to construct a network with a visualization tool.

The aim of co-expression networks is to detect gene interactions and gene neighbourhood relations which cannot be determined with other analysis methods, e.g. clustering. However, from a co-expression network it can not be concluded whether a gene interaction is direct or indirect.

Cytoscape

Cytoscape is a Java based platform for visualization and analysis of networks, in particular molecular and genetic interaction networks as well as biological pathways [Shannon et al., 2003]. The core of Cytoscape is a two-dimensional network graph with nodes, that represent molecules or genes, and edges, that represent the interactions between these molecules or genes. In order to

automatically determine the network layout, which defines the location of nodes and edges, Cytoscape provides various algorithms, such as hierarchical layout, and circular layout. Properties can be assigned to nodes and edges to integrate supplementary data, e.g. gene expression data or annotations. The appearance of nodes and edges can be adjusted to their properties through e.g. node colour, node size, edge thickness or edge style. Cytoscape supports the display of selected subnetworks to simplify a large, complex network. Nodes and edges can be chosen according to several criteria like the name or any of their properties. Additionally, Cytoscape offers sophisticated filters for the reduction of network complexity such as differential expression filter or minimum neighbour filter. Further functionality is provided by plug-ins. A large number of plug-ins exist for tasks like querying networks, integrating data in networks or adding Gene Ontology (GO) information [Cline et al., 2007].

ClueGO is a Cytoscape plug-in which facilitates the interpretation of a large number of genes [Bindea et al., 2009]. The plug-in identifies, organizes and visualizes biological terms of genes.
The tool takes a gene list or a gene network as input and retrieves KEGG and BioCarta pathways as well as GO terms corresponding to the genes. From these terms a functionally organized annotation network is built. The network displays the biological terms as nodes and their relationships as edges. The connections between terms are determined by the similarity of the genes which correspond to the terms. The size of the nodes is determined by the significance of the term in the gene set. ClueGO groups together functionally related terms and the corresponding nodes are dyed in the same colour. The plug-in makes it possible to define the specificity of the network and provides statistical tests to calculate the significance of terms or groups of terms.
Moreover, two clusters of genes can be compared and the functional differences can be displayed graphically. ClueGO can be used together with the Cytoscape GOlorize plug-in to map the retrieved biological terms to an existing gene network.

GOlorize is a Cytoscape plug-in for sophisticated network visualization [Garcia et al., 2007]. In order to display biological functions of genes in networks, the plug-in uses annotation data such as GO terms for defining the network layout. Annotations can be split into classes according to their functionality. Colours are assigned to the different classes and the network nodes representing genes or proteins are coloured corresponding to the class the gene/protein belongs to, whereby a node can belong to several classes. Moreover, nodes of the same classes can be grouped together visually.

Appendix A

Bibliography

Article References

[Adams and White, 2004] Adams J. M. and White M.; Biological ageing: a fundamental, biological link between socio-economic status and health? *Eur J Public Health*; 14(3):331–334; Sep 2004.

[Ames et al., 1993] Ames B. N., Shigenaga M. K. and Hagen T. M.; Oxidants, antioxidants, and the degenerative diseases of aging. *Proc Natl Acad Sci U S A*; 90(17):7915–7922; Sep 1993.

[Ashburner et al., 2000] Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M., Davis A. P., Dolinski K., Dwight S. S., Eppig J. T., Harris M. A., Hill D. P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J. C., Richardson J. E., Ringwald M., Rubin G. M. and Sherlock G.; Gene ontology: tool for the unification of biology. *Nat Genet*; 25(1):25–29; May 2000. doi: 10.1038/75556. URL `http://dx.doi.org/10.1038/75556`.

[Bates et al., 2009] Bates D. J., Liang R., Li N. and Wang E.; The impact of noncoding RNA on the biochemical and molecular mechanisms of aging. *Biochim Biophys Acta*; 1790(10):970–979; Oct 2009. doi: 10.1016/j.bbagen.2009.03.028. URL `http://dx.doi.org/10.1016/j.bbagen.2009.03.028`.

[Bhattacharyya and Bandyopadhyay, 2009] Bhattacharyya M. and Bandyopadhyay S.; Integration of Co-expression Networks for Gene Clustering. *Advances in Pattern Recognition, International Conference on*; 0:355–358; 2009. doi: http://doi.ieeecomputersociety.org/10.1109/ICAPR.2009.55.

[Bindea et al., 2009] Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W.-H., Pagès F., Trajanoski Z. and Galon J.; ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*; 25 (8):1091–1093; Apr 2009. doi: 10.1093/bioinformatics/btp101. URL http://dx.doi.org/10.1093/bioinformatics/btp101.

[Brazma and Vilo, 2001] Brazma A. and Vilo J.; Gene expression data analysis. *Microbes Infect*; 3(10):823–829; Aug 2001.

[Bulyk, 2003] Bulyk M. L.; Computational prediction of transcription-factor binding site locations. *Genome Biol*; 5(1):201; 2003. doi: 10.1186/gb-2003-5-1-201. URL http://dx.doi.org/10.1186/gb-2003-5-1-201.

[Campisi, 2003] Campisi J.; Cellular senescence and apoptosis: how cellular responses might influence aging phenotypes. *Exp Gerontol*; 38(1-2):5–11; 2003.

[Campisi and d'Adda di Fagagna, 2007] Campisi J. and d'Adda diFagagna F.; Cellular senescence: when bad things happen to good cells. *Nat Rev Mol Cell Biol*; 8(9):729–740; Sep 2007. doi: 10.1038/nrm2233. URL http://dx.doi.org/10.1038/nrm2233.

[Cartharius et al., 2005] Cartharius K., Frech K., Grote K., Klocke B., Haltmeier M., Klingenhoff A., Frisch M., Bayerlein M. and Werner T.; MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*; 21(13):2933–2942; Jul 2005. doi: 10.1093/bioinformatics/bti473. URL http://dx.doi.org/10.1093/bioinformatics/bti473.

[Cary et al., 2005] Cary M. P., Bader G. D. and Sander C.; Pathway information for systems biology. *FEBS Lett*; 579(8):1815–1820; Mar 2005. doi: 10.1016/j.febslet.2005.02.005. URL http://dx.doi.org/10.1016/j.febslet.2005.02.005.

[Chainiaux et al., 2002] Chainiaux F., Magalhaes J.-P., Eliaers F., Remacle J. and Toussaint O.; UVB-induced premature senescence of human diploid skin fibroblasts. *Int J Biochem Cell Biol*; 34(11):1331–1339; Nov 2002.

[Chen et al., 2010] Chen L.-H., Chiou G.-Y., Chen Y.-W., Li H.-Y. and Chiou S.-H.; microRNA and aging: A novel modulator in regulating the aging network. *Ageing Res Rev*; Aug 2010. doi: 10.1016/j.arr.2010.08.002. URL http://dx.doi.org/10.1016/j.arr.2010.08.002.

[Chuaqui et al., 2002] Chuaqui R. F., Bonner R. F., Best C. J. M., Gillespie J. W., Flaig M. J., Hewitt S. M., Phillips J. L., Krizman D. B., Tangrea M. A., Ahram M., Linehan W. M., Knezevic V. and Emmert-Buck M. R.; Post-analysis follow-up and validation of microarray experiments.

Nat Genet; 32 Suppl:509–514; Dec 2002. doi: 10.1038/ng1034. URL http://dx.doi.org/10.1038/ng1034.

[Cline et al., 2007] Cline M. S., Smoot M., Cerami E., Kuchinsky A., Landys N., Workman C., Christmas R., Avila-Campilo I., Creech M., Gross B., Hanspers K., Isserlin R., Kelley R., Killcoyne S., Lotia S., Maere S., Morris J., Ono K., Pavlovic V., Pico A. R., Vailaya A., Wang P.-L., Adler A., Conklin B. R., Hood L., Kuiper M., Sander C., Schmulevich I., Schwikowski B., Warner G. J., Ideker T. and Bader G. D.; Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc*; 2(10):2366–2382; 2007. doi: 10.1038/nprot.2007.324. URL http://dx.doi.org/10.1038/nprot.2007.324.

[Codd, 1970] Codd E.; A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*; 13:377–387; 1970.

[Dahlquist et al., 2002] Dahlquist K. D., Salomonis N., Vranizan K., Lawlor S. C. and Conklin B. R.; GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet*; 31(1):19–20; May 2002. doi: 10.1038/ng0502-19. URL http://dx.doi.org/10.1038/ng0502-19.

[de Magalhães et al., 2009] deMagalhães J. P., Budovsky A., Lehmann G., Costa J., Li Y., Fraifeld V. and Church G. M.; The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging Cell*; 8(1):65–72; Feb 2009a. doi: 10.1111/j.1474-9726.2008.00442.x. URL http://dx.doi.org/10.1111/j.1474-9726.2008.00442.x.

[de Magalhães et al., 2009] deMagalhães J. P., Curado J. and Church G. M.; Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*; 25 (7):875–881; Apr 2009b. doi: 10.1093/bioinformatics/btp073. URL http://dx.doi.org/10.1093/bioinformatics/btp073.

[D'haeseleer, 2005] D'haeseleer P.; How does gene expression clustering work? *Nat Biotechnol*; 23(12):1499–1501; Dec 2005. doi: 10.1038/nbt1205-1499. URL http://dx.doi.org/10.1038/nbt1205-1499.

[Finkel and Holbrook, 2000] Finkel T. and Holbrook N. J.; Oxidants, oxidative stress and the biology of ageing. *Nature*; 408(6809):239–247; Nov 2000. doi: 10.1038/35041687. URL http://dx.doi.org/10.1038/35041687.

[Friedman et al., 2000] Friedman N., Linial M., Nachman I. and Pe'er D.; Using Bayesian networks to analyze expression data. *J Comput Biol*; 7(3-4):601–620; 2000. doi: 10.1089/106652700750050961. URL http://dx.doi.org/10.1089/106652700750050961.

[Garcia et al., 2007] Garcia O., Saveanu C., Cline M., Fromont-Racine M., Jacquier A., Schwikowski B. and Aittokallio T.; GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout and coloring. *Bioinformatics*; 23(3):394–396; Feb 2007. doi: 10.1093/bioinformatics/btl605. URL http://dx.doi.org/10.1093/bioinformatics/btl605.

[Gentleman et al., 2004] Gentleman R. C., Carey V. J., Bates D. M., Bolstad B., Dettling M., Dudoit S., Ellis B., Gautier L., Ge Y., Gentry J., Hornik K., Hothorn T., Huber W., Iacus S., Irizarry R., Leisch F., Li C., Maechler M., Rossini A. J., Sawitzki G., Smith C., Smyth G., Tierney L., Yang J. Y. H. and Zhang J.; Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*; 5(10):R80; 2004. doi: 10.1186/gb-2004-5-10-r80. URL http://dx.doi.org/10.1186/gb-2004-5-10-r80.

[Griffiths-Jones et al., 2006] Griffiths-Jones S., Grocock R. J., vanDongen S., Bateman A. and Enright A. J.; miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*; 34(Database issue):D140–D144; Jan 2006. doi: 10.1093/nar/gkj112. URL http://dx.doi.org/10.1093/nar/gkj112.

[Grillari and Grillari-Voglauer, 2010] Grillari J. and Grillari-Voglauer R.; Novel modulators of senescence, aging, and longevity: Small non-coding RNAs enter the stage. *Exp Gerontol*; 45(4):302–311; Apr 2010. doi: 10.1016/j.exger.2010.01.007. URL http://dx.doi.org/10.1016/j.exger.2010.01.007.

[Harman, 1956] Harman D.; Aging: a theory based on free radical and radiation chemistry. *J Gerontol*; 11(3):298–300; Jul 1956.

[Jiang et al., 2004] Jiang D., Tang C. and Zhang A.; Cluster Analysis for Gene Expression Data: A Survey. *IEEE Trans. on Knowl. and Data Eng.*; 16(11):1370–1386; 2004. ISSN 1041-4347. doi: http://dx.doi.org/10.1109/TKDE.2004.68.

[Kanehisa and Goto, 2000] Kanehisa M. and Goto S.; KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*; 28(1):27–30; Jan 2000.

[Kel et al., 2003] Kel A. E., Gössling E., Reuter I., Cheremushkin E., Kel-Margoulis O. V. and Wingender E.; MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res*; 31(13):3576–3579; Jul 2003.

[Latchman, 1990] Latchman D. S.; Eukaryotic transcription factors. *Biochem J*; 270(2):281–289; Sep 1990.

[Lepperdinger et al., 2008] Lepperdinger G., Berger P., Breitenbach M., Frohlich K.-U., Grillari J., Grubeck-Loebenstein B., Madeo F., Minois N., Zwerschke W. and Jansen-Durr P.; The use of

genetically engineered model systems for research on human aging. *Front Biosci*; 13:7022–7031; 2008.

[Maurer et al., 2005] Maurer M., Molidor R., Sturn A., Hartler J., Hackl H., Stocker G., Prokesch A., Scheideler M. and Trajanoski Z.; MARS: microarray analysis, retrieval, and storage system. *BMC Bioinformatics*; 6:101; 2005. doi: 10.1186/1471-2105-6-101. URL http://dx.doi.org/10.1186/1471-2105-6-101.

[Mlecnik et al., 2005] Mlecnik B., Scheideler M., Hackl H., Hartler J., Sanchez-Cabo F. and Trajanoski Z.; PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res*; 33(Web Server issue):W633–W637; Jul 2005. doi: 10.1093/nar/gki391. URL http://dx.doi.org/10.1093/nar/gki391.

[Nakao et al., 1999] Nakao , Bono , Kawashima , Kamiya , Sato , Goto and Kanehisa ; Genome-scale Gene Expression Analysis and Pathway Reconstruction in KEGG. *Genome Inform Ser Workshop Genome Inform*; 10:94–103; 1999.

[O'Brien et al., 2005] O'Brien K. P., Remm M. and Sonnhammer E. L. L.; Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res*; 33(Database issue):D476–D480; Jan 2005. doi: 10.1093/nar/gki107. URL http://dx.doi.org/10.1093/nar/gki107.

[Ogata et al., 1999] Ogata H., Goto S., Sato K., Fujibuchi W., Bono H. and Kanehisa M.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*; 27(1):29–34; Jan 1999.

[Pan et al., 2007] Pan F., Chiu C.-H., Pulapura S., Mehan M. R., Nunez-Iglesias J., Zhang K., Kamath K., Waterman M. S., Finch C. E. and Zhou X. J.; Gene Aging Nexus: a web database and data mining platform for microarray data on aging. *Nucleic Acids Res*; 35(Database issue): D756–D759; Jan 2007. doi: 10.1093/nar/gkl798. URL http://dx.doi.org/10.1093/nar/gkl798.

[Portales-Casamar et al., 2010] Portales-Casamar E., Thongjuea S., Kwon A. T., Arenillas D., Zhao X., Valen E., Yusuf D., Lenhard B., Wasserman W. W. and Sandelin A.; JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res*; 38(Database issue):D105–D110; Jan 2010. doi: 10.1093/nar/gkp950. URL http://dx.doi.org/10.1093/nar/gkp950.

[Quandt et al., 1995] Quandt K., Frech K., Karas H., Wingender E. and Werner T.; MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res*; 23(23):4878–4884; Dec 1995.

[Reed, 2000] Reed J. C.; Mechanisms of apoptosis. *Am J Pathol*; 157(5):1415–1430; Nov 2000.

[Ruan et al., 2010] Ruan J., Dean A. K. and Zhang W.; A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Syst Biol*; 4:8; 2010. doi: 10.1186/1752-0509-4-8. URL http://dx.doi.org/10.1186/1752-0509-4-8.

[Shannon et al., 2003] Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B. and Ideker T.; Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*; 13(11):2498–2504; Nov 2003. doi: 10.1101/gr.1239303. URL http://dx.doi.org/10.1101/gr.1239303.

[Sies, 1993] Sies H.; Strategies of antioxidant defense. *Eur J Biochem*; 215(2):213–219; Jul 1993.

[Sies, 1997] Sies H.; Oxidative stress: oxidants and antioxidants. *Exp Physiol*; 82(2):291–295; Mar 1997.

[Storey, 2002] Storey J. D.; A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*; 64(3):479–498; 2002. doi: 10.1111/1467-9868.00346. URL http://dx.doi.org/10.1111/1467-9868.00346.

[Storey and Tibshirani, 2003] Storey J. D. and Tibshirani R.; Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*; 100(16):9440–9445; Aug 2003. doi: 10.1073/pnas.1530509100. URL http://dx.doi.org/10.1073/pnas.1530509100.

[Storey et al., 2004] Storey J. D., Taylor J. E. and Siegmund D.; Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal Of The Royal Statistical Society Series B*; 66(1):187–205; 2004. URL http://ideas.repec.org/a/bla/jorssb/v66y2004i1p187-205.html.

[Stormo, 2000] Stormo G. D.; DNA binding sites: representation and discovery. *Bioinformatics*; 16(1):16–23; Jan 2000.

[Sturn et al., 2002] Sturn A., Quackenbush J. and Trajanoski Z.; Genesis: cluster analysis of microarray data. *Bioinformatics*; 18:207–208; 2002.

[Tao et al., 2004] Tao Y., Liu Y., Friedman C. and Lussier Y. A.; Information Visualization Techniques in Bioinformatics during the Postgenomic Era. *Drug Discov Today Biosilico*; 2(6):237–245; Nov 2004. doi: 10.1016/S1741-8364(04)02423-0. URL http://dx.doi.org/10.1016/S1741-8364(04)02423-0.

[Viswanathan et al., 2008] Viswanathan G. A., Seto J., Patil S., Nudelman G. and Sealfon S. C.; Getting started in biological pathway construction and analysis. *PLoS Comput Biol*; 4(2):e16; Feb 2008. doi: 10.1371/journal.pcbi.0040016. URL http://dx.doi.org/10.1371/journal.pcbi.0040016.

[Weinert and Timiras, 2003] Weinert B. T. and Timiras P. S.; Invited review: Theories of aging. *J Appl Physiol*; 95(4):1706–1716; Oct 2003. doi: 10.1152/japplphysiol.00288.2003. URL http://dx.doi.org/10.1152/japplphysiol.00288.2003.

[Wilmoth, 2000] Wilmoth J. R.; Demography of longevity: past, present, and future trends. *Exp Gerontol*; 35(9-10):1111–1129; Dec 2000.

[Wingender et al., 2000] Wingender E., Chen X., Hehl R., Karas H., Liebich I., Matys V., Meinhardt T., Prüss M., Reuter I. and Schacherer F.; TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res*; 28(1):316–319; Jan 2000.

[Zahn et al., 2007] Zahn J. M., Poosala S., Owen A. B., Ingram D. K., Lustig A., Carter A., Weeraratna A. T., Taub D. D., Gorospe M., Mazan-Mamczarz K., Lakatta E. G., Boheler K. R., Xu X., Mattson M. P., Falco G., Ko M. S. H., Schlessinger D., Firman J., Kummerfeld S. K., Wood W. H., Zonderman A. B., Kim S. K. and Becker K. G.; AGEMAP: a gene expression database for aging in mice. *PLoS Genet*; 3(11):e201; Nov 2007. doi: 10.1371/journal.pgen.0030201. URL http://dx.doi.org/10.1371/journal.pgen.0030201.

[Zambelli et al., 2009] Zambelli F., Pesole G. and Pavesi G.; Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. *Nucleic Acids Res*; 37(Web Server issue):W247–W252; Jul 2009. doi: 10.1093/nar/gkp464. URL http://dx.doi.org/10.1093/nar/gkp464.

[Zhang et al., 2009] Zhang J., Li J. and Deng H.-W.; Identifying gene interaction enrichment for gene expression data. *PLoS One*; 4(11):e8064; 2009. doi: 10.1371/journal.pone.0008064. URL http://dx.doi.org/10.1371/journal.pone.0008064.

Book References and Manuals

[Beaulieu, 2009] Beaulieu A.; *Learning SQL*. O'Reilly Media; 2009.

[Bergsten, 2004] Bergsten H.; *JavaServer Pages*. O'Reilly Media; 3rd edition; 2004.

[Bland, 2006] Bland M.; *An Introduction to Medical Statistics*. Oxford University Press; 3 edition; 2006.

[Brown et al., 2001] Brown S., Burdick R., Falkner J., Galbraith B., Johnson R., Kim L., Kochmer C., Kristmundsson T., Li S., Malks D., Nelson M., Palmer G., Sullivan B., Taylor G., Timney J., Tyagi S., Damme G. V. and Wilkinson S.; *Professional JSP*. Wrox; 2nd edition; 2001.

[Burke and Monson-Haefel, 2006] Burke B. and Monson-Haefel R.; *Enterprise JavaBeans 3.0*. O'Reilly; 2006.

[Codd, 1990] Codd E. F.; *The relational model for database management: version 2*. Addison-Wesley Longman Publishing Co., Inc.; Boston, MA, USA; 1990. ISBN 0-201-14192-2.

[Gehtland et al., 2006] Gehtland J., Galbraith B. and Almaer D.; *AJAX Pragmatisch Programmieren*. Hanser Verlag; 2006.

[Gosling et al., 2000] Gosling J., Joy B., Steele G. and Bracha G.; *The Java Language Specification*. Addison-Wesley; 2000.

[Hall and Brown, 2004] Hall M. and Brown L.; *Core Servlets und JavaServer Pages*. Markt+Technik Verlag; 2nd edition; 2004.

[Holzner, 2004] Holzner S.; *Eclipse*. O'Reilly Media; 2004.

[Husted et al., 2003] Husted T. N., Dumoulin C., Franciscus G. and Winterfeldt D.; *Struts in Action*. Manning Publications; 2003.

[Jendrock et al., 2008] Jendrock E., Ball J., Carson D., Evans I., Fordin S. and Haase K. *The Java™EE 5 Tutorial*. Sun Microsystems; 10 2008. URL `http://download.oracle.com/javaee/5/tutorial/doc/`.

[Keith, 2008] Keith J. M.; *Bioinformatics Volume I: Data, Sequence Analysis and Evolution*; volume 1. Humana Press Inc.; Hatfield, UK; 2008a.

[Keith, 2008] Keith J. M.; *Bioinformatics Volume II: Structure, Function And Applications*; volume 2. Humana Press Inc.; Hatfield, UK; 2008b.

[Kodali et al., 2006] Kodali R. R., Wetherbee J. R. and Zadrozny P.; *Beginning EJB 3 Application Development: From Novice to Professional*. Apress; 2006.

[Kroenke, 2005] Kroenke D. M.; *Database Processing*. Prentice Hall; 2005.

[Lee, 2010] Lee J. K.; *Statistical Bioinformatics for Biomedical and Life Science Researchers*. Wiley-Blackwell; 2010.

[Lodish et al., 2007] Lodish H., Berk A., Kaiser C. A., Krieger M., Scott M. P., Bretscher A., Ploegh H. and Matsudaira P.; *Molecular Cell Biology*. W. H. Freeman; Basingstoke, UK & New York, USA; 6 edition; 2007.

[Meyer, 2006] Meyer E. A.; *CSS - The Definitive Guide*. O'Reilly Media; 2006.

[Mukhar et al., 2005] Mukhar K., Zelenak C., Weaver J. L. and Crume J.; *Beginning Java EE 5: From Novice to Professional*. Apress; 2005.

[Prigmore, 2008] Prigmore M.; *An Introduction to Databases with Web Applications*. Prentice Hall; 2008.

[R Development Core Team, 2007] R Development Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria; 2007. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

[Schena, 2000] Schena M.; *Microarray Biochip Technology*. Eaton Publishing; Natick, USA; 2000.

[Schmitt et al., 2008] Schmitt C., Dominey T., Li C., Marcotte E., Orchard D. and Tramell M.; *Professional CSS: Cascading Style Sheets for Web Design*. Wrox; 2008.

[Sumathi and Esakkirajan, 2007] Sumathi S. and Esakkirajan S.; *Fundamentals of Relational Database Management Systems*. Springer; 2007. ISBN 1860-9503.

[Ullenboom, 2007] Ullenboom C.; *Java ist auch eine Insel*. Galileo Computing; 2007.

[Verzani, 2005] Verzani J.; *Using R for Introductory Statistics*. Chapman and Hall/CRC; 2005.

Unpublished References

[DeMichiel and Keith, 2006] DeMichiel L. and Keith M.; JSR 220: Enterprise JavaBeansTM 3.0. URL `http://jcp.org/jsr/detail/220.jsp`. 2006.

[Eclipse, 2006] Eclipse ; Eclipse Platform Technical Overview. URL `http://www.eclipse.org/articles/Whitepaper-Platform-3.1/eclipse-platform-whitepaper.pdf`. White Paper; 2006.

[Garrett, 2005] Garrett J. J.; Ajax: A New Approach to Web Applications. URL `http://www.adaptivepath.com/ideas/essays/archives/000385.php`. Essay; February 2005.

[Gosling and McGilton, 1996] Gosling J. and McGilton H.; The Java Language Environment. URL `http://java.sun.com/docs/white/langenv/`. White Paper; 1996.

[JSP, 2010] JSP ; JavaServer Pages[tm] Technology. URL `http://java.sun.com/products/jsp/whitepaper.html`. White Paper; 2010.

[Miller and Mukerji, 2003] Miller J. and Mukerji J.; MDA Guide Version 1.0.1. URL `http://www.omg.org/mda/mda_files/MDA_Guide_Version1-0.pdf`. Guide; June 2003.

[Murray, 2005] Murray G.; Asynchronous JavaScript Technology and XML (Ajax) With the Java Platform. URL http://java.sun.com/developer/technicalArticles/J2EE/AJAX/. Technical Article; June 2005.

[NoMagic, 2011] NoMagic ; Magic Draw User's Manual. URL https://www.magicdraw.com/files/manuals/MagicDrawUserManual.pdf. User Guide; 2011.

[Pitzl, 2007] Pitzl M.; Design and Development of a Database and Retrieval System for Research in Cellular Aging. Master's thesis; Graz University of Technology; 2007.

[Siegel, 2001] Siegel J.; Developing in OMG's Model-Driven Architecture. URL http://www.eecg.toronto.edu/~jacobsen/courses/ece1770/reader/developing-in-mda.pdf. White Paper; November 2001.

[Truyen, 2006] Truyen F.; The Fast Guide to Model Driven Architecture. URL http://www.omg.org/mda/mda_files/Cephas_MDA_Fast_Guide.pdf. White Paper; 2006.

[Zeller, 2003] Zeller D.; Design and development of a user managment system for molecular biology database systems. Master's thesis; Graz University of Technology; 2003.

Web Link References

[Affymetrix, 2010] Affymetrix. Last visited on 03/06/2010. URL http://www.affymetrix.com.

[AndroMDA, 2010] AndroMDA. Last visited on 12/04/2010. URL http://www.andromda.org.

[BioCarta, 2010] BioCarta. Last visited on 27/04/2010. URL http://www.biocarta.com/.

[Eclipse, 2011] Eclipse. Last visited on 10/03/2011. URL http://www.eclipse.org/.

[ElDorado, 2010] ElDorado. Last visited on 10/06/2010. URL http://www.genomatix.de/products/ElDorado/.

[EntrezGene, 2010] EntrezGene. Last visited on 10/06/2010. URL http://www.ncbi.nlm.nih.gov/gene.

[Exiqon, 2010] Exiqon. Last visited on 09/06/2010. URL http://www.exiqon.com/.

[FlyBase, 2011] A Database of Drosophila Genes and Genomes. Last visited on 29/03/2011. URL http://flybase.org/.

[genesDB, 2010] Aging Genes Database. Last visited on 19/08/2010. URL http://uwaging.org/genesdb/index.php.

[GlassFish, 2010] GlassFish. Last visited on 27/04/2010. URL Lastvisitedon16/04/2010.

[Hackl, 2010] Over Representation Analysis. Last visited on 29/04/2010. URL http://genome.tugraz.at/ORA/.

[HomoloGene, 2010] HomoloGene. Last visited on 04/06/2010. URL http://www.ncbi.nlm.nih.gov/homologene.

[IBM, 2010] IBM DB2 Software. Last visited on 08/03/2011. URL http://www-01.ibm.com/software/data/db2/.

[JBoss, 2010] JBoss Application Server. Last visited on 27/04/2010. URL http://www.jboss.org/jbossas.

[JDBC, 2010] Getting Started with the JDBC API. Last visited on 20/04/2010. URL http://java.sun.com/javase/6/docs/technotes/guides/jdbc/getstart/GettingStartedTOC.fm.html.

[JOnAS, 2010] Java Open Application Server. Last visited on 27/04/2010. URL http://wiki.jonas.ow2.org/xwiki/bin/view/Main/.

[MagicDraw, 2010] MagicDraw. Last visited on 10/03/2011. URL http://www.magicdraw.com/.

[Match, 2010] Match. Last visited on 29/04/2010. URL http://www.gene-regulation.com/pub/programs.html#match.

[MatInspector, 2010] MatInspector. Last visited on 30/04/2010. URL http://www.genomatix.de/matinspector.html.

[MGI, 2011] Mouse Genome Informatics. Last visited on 10/03/2011. URL http://www.informatics.jax.org/.

[Microsoft, 2010] Microsoft SQL Server. Last visited on 08/03/2011. URL http://www.microsoft.com/sqlserver.

[MyEclipse, 2011] MyEclipse. Last visited on 10/03/2011. URL http://www.myeclipseide.com/.

[MySQL, 2011] MySQL. Last visited on 08/03/2011. URL http://www.mysql.de/.

[NRN, 2010] Proliferation, Differentiation and Cell Death during Cellular Aging. Last visited on 19/08/2010. URL http://www.iba.oeaw.ac.at/nfn/index.php.

[OMG, 2010] Object Management Group. Last visited on 15/04/2010. URL http://www.omg.org/.

[OMG, 2010] OMG Model Driven Architecture. Last visited on 15/04/2010. URL http://www.omg.org/mda/.

[Oracle, 2011] Oracle Database. Last visited on 08/03/2011. URL http://www.oracle.com/us/products/database/index.html.

[PostgreSQL, 2011] PostgreSQL. Last visited on 08/03/2011. URL http://www.postgresql.org/.

[Prototype, 2010] Prototype. Last visited on 15/04/2010. URL http://prototypejs.org/.

[Pscan, 2010] Pscan. Last visited on 29/04/2010. URL http://159.149.109.9/pscan/.

[RefSeq, 2010] RefSeq. Last visited on 10/06/2010. URL http://www.ncbi.nlm.nih.gov/RefSeq/.

[RGD, 2010] Rat Genome Database. Last visited on 10/03/2011. URL http://rgd.mcw.edu/.

[script.aculo.us, 2010] script.aculo.us. Last visited 12/04/2010. URL http://script.aculo.us/.

[SGD, 2010] Saccharomyces Genome Database. Last visited on 10/03/2011. URL http://www.yeastgenome.org/.

[SQLDeveloper, 2011] Oracle SQL Developer. Last visited on 10/03/2011. URL http://www.oracle.com/technetwork/developer-tools/sql-developer/overview/index.html.

[Struts, 2010] Apache Struts. Last visited on 21/04/2010. URL http://struts.apache.org/.

[TAIR, 2011] The Arabidopsis Information Resource. Last visited on 05/03/2011. URL http://www.arabidopsis.org/.

[Tomcat, 2010] Apache Tomcat. Last visited on 21/04/2010. URL http://tomcat.apache.org/.

[TRANSFAC, 2010] TRANSFAC 7.0 Public 2005. Last visited on 29/04/2010. URL http://www.gene-regulation.com/pub/databases.html.

[UniGene, 2011] UniGene: An Organized View of the Transcriptome. Last visited on 10/03/2011. URL http://www.ncbi.nlm.nih.gov/unigene.

[UniProt, 2010] UniProt. Last visited on 10/06/2010. URL http://www.uniprot.org/.

[Velocity, 2010] The Apache Velocity Project. Last visited on 16/04/2010. URL http://velocity.apache.org/.

[WebSphere, 2010] WebSphere Application Server. Last visited on 27/04/2010. URL http://www-01.ibm.com/software/webservers/appserv/was/.

List of Figures

1.1 Relationship between oxidative stress, senescence and apoptosis 2

2.1 GiSAO.db: overview 7
2.2 GiSAO.db: database core schema for gene information storage 8
2.3 GiSAO.db: gene expression values 9
2.4 GiSAO.db: ortholog search result 10
2.5 GiSAO.db: tags 11
2.6 GiSAO.db: search function 12
2.7 GiSAO.db: autocomplete field 14
2.8 GiSAO.db: AJAX visual effect 14
2.9 Hierarchical sample clustering 21
2.10 K-means clustering 22
2.11 Cell cycle pathway 27
2.12 Molecular function GO terms 35
2.13 Gene-gene interaction network 36

4.1 Multi-tier architecture 45
4.2 Model-view-controller architecture 51
4.3 Model driven architecture 54
4.4 Position weight matrix 64

List of Tables

2.1 Oxidative stress and senescence pathways 18
2.2 Oxidative stress pathways 19
2.3 Senescence pathways 20
2.4 Pathways of DE probe sets on day 9(a) of UVB experiment 1 24
2.5 Pathways of DE probe sets on day 9(b) of UVB experiment 1 24
2.6 Pathways of DE probe sets on day 1 of UVB experiment 2 25
2.7 Pathways of DE probe sets on day 7 of UVB experiment 2 25
2.8 Pathways of DE probe sets on day 9 of UVB experiment 2 26
2.9 Pscan over-represented TFBS of genes of interest 28
2.10 Pscan over-represented TFBS of DE probe sets on day 9(a) of UVB experiment 1 28
2.11 Pscan over-represented TFBS of DE probe sets on day 9(b) of UVB experiment 1 28
2.12 Pscan over-represented TFBS of DE probe sets on day 1 of UVB experiment 2 . . 28
2.13 Pscan over-represented TFBS of DE probe sets on day 7 of UVB experiment 2 . . 29
2.14 Pscan over-represented TFBS of DE probe sets on day 9 of UVB experiment 2 . . 29
2.15 ORA over-represented TFBS of genes of interest 30
2.16 ORA over-represented TFBS of DE probe sets on day 9(a) of UVB experiment 1 . 30
2.17 ORA over-represented TFBS of DE probe sets on day 9(b) of UVB experiment 1 . 30
2.18 ORA over-represented TFBS of DE probe sets on day 1 of UVB experiment 2 . . . 31
2.19 ORA over-represented TFBS of DE probe sets on day 7 of UVB experiment 2 . . . 31
2.20 ORA over-represented TFBS of DE probe sets on day 9 of UVB experiment 2 . . . 31
2.21 Genomatix MatInspector TFBS 32
2.22 Genomatix MatInspector common TFBS 33
2.23 Match NF-κB TFBS of genes of interest 34
2.24 Match TFBS 34

Appendix B

Glossary

AAS	Authentication and Authorization System
ACL	Access Control Lists
AGEMAP	Atlas of Gene Expression in Mouse Aging Project
AHR	Aryl Hydrocarbon Receptor
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator
bp	Base Pair
CASE	Computer-Aided Software Engineering
CBD	Cumulative Binomial Distribution
CIM	Computation Independent Model
ChIP	Chromatin Immunoprecipitation
CSS	Cascadian Style Sheets
CSS	Core Similarity Score
CSV	Comma Separated Values
DBMS	Database Management Systems
DCL	Data Control Language
DDL	Data Definition Language
DE	Differentially expressed
DML	Data Manipulation Language
DNA	Deoxyribonucleic Acid
DOM	Document Object Model

EAR	Enterprise Archive
EGR2	Early Growth Response 2
EIS	Enterprise Information System
EJB	Enterprise Java Beans
EMSA	Electrophoretic Mobility Shift Assay
E2F	E2 Binding Factor
FDR	False Discovery Rate
GAN	Gene Aging Nexus
GATA6	GATA Binding Protein 6
GiSAO.db	Genes involved in Senescence, Apoptosis and Oxidative Stress Database
GO	Gene Ontology
GUI	Graphical User Interface
HAGR	Human Aging Genomic Resources
HDF	Human Diploid Fibroblasts
HTML	Hypertext Markup Language
HTTP	Hypertext Transport Protocol
HUVEC	Human Umbilical Vein Epithelial Cells
IDE	Integrated Development Environment
Java EE	Java Enterprise Edition
JDBC	Java Database Connectivity
JMS	Java Messaging Service
JPA	Java Persistence API
JPEG	Joint Photographic Experts Group
JPQL	Java Persistence Query Language
JSF	JavaServer Faces
JSP	JavaServer Pages
JVM	Java Virtual Machine
KEGG	Kyoto Encyclopedia of Genes and Genomes
MDA	Model Driven Architecture
MDB	Message Driven Bean
miRNA	microRNA
MSC	Mesenchymal Stem Cells
MSS	Matrix Similarity Score
MVC	Model View Controller
NF-κB	Nuclear Factor kappa-Light Chain-Enhancer of activated B Cells

OMG	Object Management Group
ORA	Over Representation Analysis
PFF	Primary Foreskin Fibroblasts
PIM	Platform Independent Model
PNG	Portable Network Graphics
POJO	Plain Old Java Object
PPF	Primary Prostatic stromal Fibroblasts
PrSC	Primary Prostate Stromal Cells (PrSC)
PSM	Platform Specific Model
PSSM	Position Specific Scoring Matrix
PWM	Position Weight Matrix
p53	Tumor Protein 53
qPCR	Real-time quantitative Polymerase Chain Reaction
RARB	Retinoid Acid Receptor Beta
RDBMS	Relational Database Management System
RNA	Ribonucleic Acid
RPTEC	Renal Proximal Tubular Epithelial Cells
RXR	Retinoid X Receptor
SDK	Software Development Kit
SGD	Saccharomyces Genome Database
SOM	Self-organizing Maps
SOX11	SRY (Sex Determining Region Y)-Box 11
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TAIR	The Arabidopsis Information Resource
tBHP	tert-Butyl Hydroperoxide
TCP/IP	Transmission Control Protocol / Internet Protocol
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TGF-β	Transforming Growth Factor Beta
TSS	Transcription Start Site
UML	Unified Modeling Language
UVB	Ultraviolet B
VDR	Vitamin D Receptor
WAR	Web Application Archive

W3C	World Wide Web Consortium
XHTML	Extensible HyperText Markup Language
XMI	Extensible Markup Language Metadata Interchange
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformation

Appendix C

Acknowledgments

This work has been supported by the National Research Network (NRN) of the Austrian Science Fund (FWF) and the Austrian Academy of Sciences.
I would like to express my gratitude to my supervisor Zlatko Trajanoski for supporting me throughout my thesis.
Further, I want to thank Gerhard Thallinger for his guidance and help.
Sincere thanks go to the members of the Institute for Genomics and Bioinformatics and the Division for Bioinformatics at Innsbruck Medical University for their assistance and friendship.
I would also like to thank Günter Lepperdinger and Pidder Jansen-Duerr who introduced me to the interesting field of ageing and their group members Gerhard Laschober and Ruth Greussing who provided data and very helpful advice.
I am indebted to my family for their unfailing support, and my companion in life Bernhard, for his love, encouragement and understanding.

Appendix D

Supplementary Information

Comparison of gene expression profiles of various ageing models

Probe sets associated with senescence and oxidative stress

The q-value of the listed probe sets is $<$0.005 in the oxidative stress experimental group (OS) as well as in the senescence experimental group (SEN). Every probe set is differentially expressed in at least 9 experiments in each experimental group. For each probe set and both experimental groups, the number of experiments where the probe set was differentially expressed (# Sen DE, # OS DE), the p-value and the q-value are displayed.

Probe set Id	Symbol	# Sen DE	SEN p-value	SEN q-value	# OS DE	OS p-value	OS q-value
204205_at	APOBEC3G	16	1.94E-13	0.00156	10	2.66E-06	0.00293
202016_at	MEST	16	1.94E-13	0.00156	9	2.72E-05	0.00293
201008_s_at	TXNIP	14	2.17E-10	0.00156	11	2.08E-07	0.00293
201009_s_at	TXNIP	14	2.17E-10	0.00156	11	2.08E-07	0.00293
219304_s_at	PDGFD	14	2.17E-10	0.00156	9	2.72E-05	0.00293
202503_s_at	KIAA0101	13	4.56E-09	0.00156	14	2.15E-11	0.00293
201890_at	RRM2	13	4.56E-09	0.00156	12	1.28E-08	0.00293
242245_at	SYDE2	13	4.56E-09	0.00156	12	1.28E-08	0.00293
201170_s_at	BHLHB2	13	4.56E-09	0.00156	11	2.08E-07	0.00293
216607_s_at	CYP51A1	13	4.56E-09	0.00156	11	2.08E-07	0.00293
204439_at	IFI44L	13	4.56E-09	0.00156	10	2.66E-06	0.00293
226388_at	TCEA3	13	4.56E-09	0.00156	10	2.66E-06	0.00293
227236_at	TSPAN2	13	4.56E-09	0.00156	10	2.66E-06	0.00293
201010_s_at	TXNIP	13	4.56E-09	0.00156	9	2.72E-05	0.00293
201195_s_at	SLC7A5	13	4.56E-09	0.00156	9	2.72E-05	0.00293
219529_at	CLIC3	13	4.56E-09	0.00156	9	2.72E-05	0.00293

Probe set Id	Symbol	# Sen DE	SEN p-value	SEN q-value	# OS DE	OS p-value	OS q-value
223220_s_at	PARP9	13	4.56E-09	0.00156	9	2.72E-05	0.00293
225606_at	BCL2L11	13	4.56E-09	0.00156	9	2.72E-05	0.00293
228582_x_at	MALAT1	13	4.56E-09	0.00156	9	2.72E-05	0.00293
205822_s_at	HMGCS1	12	7.43E-08	0.00156	12	1.28E-08	0.00293
208078_s_at	TCF8	12	7.43E-08	0.00156	12	1.28E-08	0.00293
1555938_x_at	VIM	12	7.43E-08	0.00156	11	2.08E-07	0.00293
1557527_at	RUNX1	12	7.43E-08	0.00156	11	2.08E-07	0.00293
203799_at	CD302	12	7.43E-08	0.00156	11	2.08E-07	0.00293
203973_s_at	CEBPD	12	7.43E-08	0.00156	10	2.66E-06	0.00293
204026_s_at	ZWINT	12	7.43E-08	0.00156	10	2.66E-06	0.00293
221521_s_at	Pfs2	12	7.43E-08	0.00156	10	2.66E-06	0.00293
224559_at	MALAT1	12	7.43E-08	0.00156	10	2.66E-06	0.00293
227099_s_at	LOC387763	12	7.43E-08	0.00156	10	2.66E-06	0.00293
228003_at	RAB30	12	7.43E-08	0.00156	10	2.66E-06	0.00293
230664_at	MGC39900	12	7.43E-08	0.00156	10	2.66E-06	0.00293
201596_x_at	KRT18	12	7.43E-08	0.00156	9	2.72E-05	0.00293
201688_s_at	TPD52	12	7.43E-08	0.00156	9	2.72E-05	0.00293
205807_s_at	TUFT1	12	7.43E-08	0.00156	9	2.72E-05	0.00293
206085_s_at	CTH	12	7.43E-08	0.00156	9	2.72E-05	0.00293
209765_at	ADAM19	12	7.43E-08	0.00156	9	2.72E-05	0.00293
221841_s_at	KLF4	12	7.43E-08	0.00156	9	2.72E-05	0.00293
223044_at	SLC40A1	12	7.43E-08	0.00156	9	2.72E-05	0.00293
235593_at	ZFHX1B	12	7.43E-08	0.00156	9	2.72E-05	0.00293
1554020_at	BICD1	11	9.59E-07	0.00156	13	6.10E-10	0.00293
225687_at	C20orf129	11	9.59E-07	0.00156	12	1.28E-08	0.00293
238542_at	ULBP2	11	9.59E-07	0.00156	12	1.28E-08	0.00293
1556821_x_at	DLEU2	11	9.59E-07	0.00156	11	2.08E-07	0.00293
213348_at	CDKN1C	11	9.59E-07	0.00156	11	2.08E-07	0.00293
214657_s_at	TncRNA	11	9.59E-07	0.00156	11	2.08E-07	0.00293
218009_s_at	PRC1	11	9.59E-07	0.00156	11	2.08E-07	0.00293
223079_s_at	GLS	11	9.59E-07	0.00156	11	2.08E-07	0.00293
228069_at	FAM54A	11	9.59E-07	0.00156	11	2.08E-07	0.00293
230494_at	SLC20A1	11	9.59E-07	0.00156	11	2.08E-07	0.00293
205659_at	HDAC9	11	9.59E-07	0.00156	10	2.66E-06	0.00293
214710_s_at	CCNB1	11	9.59E-07	0.00156	10	2.66E-06	0.00293
227609_at	EPSTI1	11	9.59E-07	0.00156	10	2.66E-06	0.00293
238013_at	PLEKHA2	11	9.59E-07	0.00156	10	2.66E-06	0.00293
1555860_x_at		11	9.59E-07	0.00156	9	2.72E-05	0.00293
1556035_s_at	ZNF207	11	9.59E-07	0.00156	9	2.72E-05	0.00293
204334_at	KLF7	11	9.59E-07	0.00156	9	2.72E-05	0.00293
204825_at	MELK	11	9.59E-07	0.00156	9	2.72E-05	0.00293
204908_s_at	BCL3	11	9.59E-07	0.00156	9	2.72E-05	0.00293
205282_at	LRP8	11	9.59E-07	0.00156	9	2.72E-05	0.00293
206108_s_at	SFRS6	11	9.59E-07	0.00156	9	2.72E-05	0.00293
206385_s_at	ANK3	11	9.59E-07	0.00156	9	2.72E-05	0.00293
207563_s_at	OGT	11	9.59E-07	0.00156	9	2.72E-05	0.00293
212446_s_at	LASS6	11	9.59E-07	0.00156	9	2.72E-05	0.00293
218782_s_at	ATAD2	11	9.59E-07	0.00156	9	2.72E-05	0.00293

Probe set Id	Symbol	# Sen DE	SEN p-value	SEN q-value	# OS DE	OS p-value	OS q-value
223062_s_at	PSAT1	11	9.59E-07	0.00156	9	2.72E-05	0.00293
227478_at	SETBP1	11	9.59E-07	0.00156	9	2.72E-05	0.00293
227618_at	FLJ44635	11	9.59E-07	0.00156	9	2.72E-05	0.00293
228390_at		11	9.59E-07	0.00156	9	2.72E-05	0.00293
235476_at	TRIM59	11	9.59E-07	0.00156	9	2.72E-05	0.00293
209773_s_at	RRM2	10	9.97E-06	0.00156	14	2.15E-11	0.00293
210567_s_at	SKP2	10	9.97E-06	0.00156	14	2.15E-11	0.00293
202705_at	CCNB2	10	9.97E-06	0.00156	12	1.28E-08	0.00293
211367_s_at	CASP1	10	9.97E-06	0.00156	12	1.28E-08	0.00293
212143_s_at	IGFBP3	10	9.97E-06	0.00156	12	1.28E-08	0.00293
222118_at	BM039	10	9.97E-06	0.00156	12	1.28E-08	0.00293
201663_s_at	SMC4L1	10	9.97E-06	0.00156	11	2.08E-07	0.00293
201790_s_at	DHCR7	10	9.97E-06	0.00156	11	2.08E-07	0.00293
203362_s_at	MAD2L1	10	9.97E-06	0.00156	11	2.08E-07	0.00293
205321_at	EIF2S3	10	9.97E-06	0.00156	11	2.08E-07	0.00293
208079_s_at	STK6	10	9.97E-06	0.00156	11	2.08E-07	0.00293
217591_at	SKIL	10	9.97E-06	0.00156	11	2.08E-07	0.00293
218039_at	NUSAP1	10	9.97E-06	0.00156	11	2.08E-07	0.00293
226568_at	LOC284611	10	9.97E-06	0.00156	11	2.08E-07	0.00293
226713_at	C3orf6	10	9.97E-06	0.00156	11	2.08E-07	0.00293
1553994_at	NT5E	10	9.97E-06	0.00156	10	2.66E-06	0.00293
201250_s_at	SLC2A1	10	9.97E-06	0.00156	10	2.66E-06	0.00293
201291_s_at	TOP2A	10	9.97E-06	0.00156	10	2.66E-06	0.00293
202450_s_at	CTSK	10	9.97E-06	0.00156	10	2.66E-06	0.00293
207564_x_at	OGT	10	9.97E-06	0.00156	10	2.66E-06	0.00293
213158_at		10	9.97E-06	0.00156	10	2.66E-06	0.00293
213527_s_at	ZNF688	10	9.97E-06	0.00156	10	2.66E-06	0.00293
227404_s_at	EGR1	10	9.97E-06	0.00156	10	2.66E-06	0.00293
227510_x_at	MALAT1	10	9.97E-06	0.00156	10	2.66E-06	0.00293
201292_at	TOP2A	10	9.97E-06	0.00156	9	2.72E-05	0.00293
201650_at	KRT19	10	9.97E-06	0.00156	9	2.72E-05	0.00293
203158_s_at	GLS	10	9.97E-06	0.00156	9	2.72E-05	0.00293
203186_s_at	S100A4	10	9.97E-06	0.00156	9	2.72E-05	0.00293
203554_x_at	PTTG1	10	9.97E-06	0.00156	9	2.72E-05	0.00293
206114_at	EPHA4	10	9.97E-06	0.00156	9	2.72E-05	0.00293
208891_at	DUSP6	10	9.97E-06	0.00156	9	2.72E-05	0.00293
212706_at	RASA4	10	9.97E-06	0.00156	9	2.72E-05	0.00293
213629_x_at	MT1F	10	9.97E-06	0.00156	9	2.72E-05	0.00293
215252_at	DNAJC7	10	9.97E-06	0.00156	9	2.72E-05	0.00293
216252_x_at	FAS	10	9.97E-06	0.00156	9	2.72E-05	0.00293
218894_s_at	FLJ10292	10	9.97E-06	0.00156	9	2.72E-05	0.00293
218921_at	SIGIRR	10	9.97E-06	0.00156	9	2.72E-05	0.00293
218983_at	C1RL	10	9.97E-06	0.00156	9	2.72E-05	0.00293
221892_at	H6PD	10	9.97E-06	0.00156	9	2.72E-05	0.00293
222217_s_at	SLC27A3	10	9.97E-06	0.00156	9	2.72E-05	0.00293
222740_at	ATAD2	10	9.97E-06	0.00156	9	2.72E-05	0.00293
225341_at	MTERFD3	10	9.97E-06	0.00156	9	2.72E-05	0.00293
225485_at	TSGA14	10	9.97E-06	0.00156	9	2.72E-05	0.00293

Probe set Id	Symbol	# Sen DE	SEN p-value	SEN q-value	# OS DE	OS p-value	OS q-value
226666_at	DAAM1	10	9.97E-06	0.00156	9	2.72E-05	0.00293
227212_s_at	PHF19	10	9.97E-06	0.00156	9	2.72E-05	0.00293
228793_at	JMJD1C	10	9.97E-06	0.00156	9	2.72E-05	0.00293
229007_at	LOC283788	10	9.97E-06	0.00156	9	2.72E-05	0.00293
229221_at	CD44	10	9.97E-06	0.00156	9	2.72E-05	0.00293
232369_at	MBNL2	10	9.97E-06	0.00156	9	2.72E-05	0.00293
239487_at	DKFZP564-F0522	10	9.97E-06	0.00156	9	2.72E-05	0.00293
241755_at	UQCRC2	10	9.97E-06	0.00156	9	2.72E-05	0.00293
244871_s_at	USP32	10	9.97E-06	0.00156	9	2.72E-05	0.00293
1555858_at		9	8.42E-05	0.00298	15	5.34E-13	0.00293
1559096_x_at	FBXO9	9	8.42E-05	0.00298	12	1.28E-08	0.00293
201694_s_at	EGR1	9	8.42E-05	0.00298	12	1.28E-08	0.00293
206102_at	PSF1	9	8.42E-05	0.00298	12	1.28E-08	0.00293
222846_at	RAB8B	9	8.42E-05	0.00298	12	1.28E-08	0.00293
202095_s_at	BIRC5	9	8.42E-05	0.00298	11	2.08E-07	0.00293
209408_at	KIF2C	9	8.42E-05	0.00298	11	2.08E-07	0.00293
218723_s_at	RGC32	9	8.42E-05	0.00298	11	2.08E-07	0.00293
220789_s_at	TBRG4	9	8.42E-05	0.00298	11	2.08E-07	0.00293
223229_at	UBE2T	9	8.42E-05	0.00298	11	2.08E-07	0.00293
224566_at	TncRNA	9	8.42E-05	0.00298	11	2.08E-07	0.00293
227044_at		9	8.42E-05	0.00298	11	2.08E-07	0.00293
229715_at		9	8.42E-05	0.00298	11	2.08E-07	0.00293
232338_at	ZNF431	9	8.42E-05	0.00298	11	2.08E-07	0.00293
235117_at	ASB3	9	8.42E-05	0.00298	11	2.08E-07	0.00293
205967_at	HIST1H4C	9	8.42E-05	0.00298	10	2.66E-06	0.00293
209970_x_at	CASP1	9	8.42E-05	0.00298	10	2.66E-06	0.00293
210052_s_at	TPX2	9	8.42E-05	0.00298	10	2.66E-06	0.00293
213645_at	ENOSF1	9	8.42E-05	0.00298	10	2.66E-06	0.00293
215629_s_at	KIAA1799	9	8.42E-05	0.00298	10	2.66E-06	0.00293
226750_at	LARP2	9	8.42E-05	0.00298	10	2.66E-06	0.00293
227164_at	SFRS1	9	8.42E-05	0.00298	10	2.66E-06	0.00293
227961_at	CTSB	9	8.42E-05	0.00298	10	2.66E-06	0.00293
228964_at	PRDM1	9	8.42E-05	0.00298	10	2.66E-06	0.00293
237400_at	ATP5S	9	8.42E-05	0.00298	10	2.66E-06	0.00293
242727_at	ARL8	9	8.42E-05	0.00298	10	2.66E-06	0.00293
1555419_a_at	ASAH1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
1560814_a_at	MGC20481	9	8.42E-05	0.00298	9	2.72E-05	0.00293
201631_s_at	IER3	9	8.42E-05	0.00298	9	2.72E-05	0.00293
201664_at	SMC4L1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
202022_at	ALDOC	9	8.42E-05	0.00298	9	2.72E-05	0.00293
202668_at	EFNB2	9	8.42E-05	0.00298	9	2.72E-05	0.00293
203868_s_at	VCAM1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
204768_s_at	FEN1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
205691_at	SYNGR3	9	8.42E-05	0.00298	9	2.72E-05	0.00293
205842_s_at	JAK2	9	8.42E-05	0.00298	9	2.72E-05	0.00293
210461_s_at	ABLIM1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
213693_s_at	MUC1	9	8.42E-05	0.00298	9	2.72E-05	0.00293

Probe set Id	Symbol	# Sen DE	SEN p-value	SEN q-value	# OS DE	OS p-value	OS q-value
214290_s_at	HIST2H2AA	9	8.42E-05	0.00298	9	2.72E-05	0.00293
218793_s_at	SCML1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
219787_s_at	ECT2	9	8.42E-05	0.00298	9	2.72E-05	0.00293
221750_at	HMGCS1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
221823_at	LOC90355	9	8.42E-05	0.00298	9	2.72E-05	0.00293
222077_s_at	RACGAP1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
224753_at	CDCA5	9	8.42E-05	0.00298	9	2.72E-05	0.00293
225521_at	ANAPC7	9	8.42E-05	0.00298	9	2.72E-05	0.00293
226333_at	IL6R	9	8.42E-05	0.00298	9	2.72E-05	0.00293
226350_at	OPN3	9	8.42E-05	0.00298	9	2.72E-05	0.00293
226936_at	C6orf173	9	8.42E-05	0.00298	9	2.72E-05	0.00293
227606_s_at	AMSH-LP	9	8.42E-05	0.00298	9	2.72E-05	0.00293
228729_at	CCNB1	9	8.42E-05	0.00298	9	2.72E-05	0.00293
229704_at	APRIN	9	8.42E-05	0.00298	9	2.72E-05	0.00293
230505_at	LOC145474	9	8.42E-05	0.00298	9	2.72E-05	0.00293
231769_at	FBXO6	9	8.42E-05	0.00298	9	2.72E-05	0.00293
232194_at	METTL4	9	8.42E-05	0.00298	9	2.72E-05	0.00293
235693_at		9	8.42E-05	0.00298	9	2.72E-05	0.00293
236364_at		9	8.42E-05	0.00298	9	2.72E-05	0.00293
241769_at	ITGAV	9	8.42E-05	0.00298	9	2.72E-05	0.00293
242471_at		9	8.42E-05	0.00298	9	2.72E-05	0.00293

Probe sets associated with oxidative stress

The q-value of the listed probe sets is <0.005 in the oxidative stress experimental group (OS). Every probe set is differentially expressed in at least 9 experiments out of 17 oxidative stress experiments. For each probe set, the number of experiments where the probe set was differentially expressed (# OS DE), the p-value and the q-value are displayed.

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
219282_s_at	TRPV2	14	2.15E-11	0.00293
1555996_s_at	EIF4A2	13	6.10E-10	0.00293
201468_s_at	NQO1	13	6.10E-10	0.00293
1555989_at	DAAM1	12	1.28E-08	0.00293
218252_at	CKAP2	12	1.28E-08	0.00293
201123_s_at	EIF5A	12	1.28E-08	0.00293
203028_s_at	CYBA	12	1.28E-08	0.00293
212366_at	ZNF292	12	1.28E-08	0.00293
225191_at	CIRBP	12	1.28E-08	0.00293
225767_at	LOC284801	12	1.28E-08	0.00293
228273_at	FLJ11029	12	1.28E-08	0.00293
229390_at	LOC441168	12	1.28E-08	0.00293
239577_at		12	1.28E-08	0.00293
235419_at	ERRFI1	12	1.28E-08	0.00293
201310_s_at	C5orf13	11	2.08E-07	0.00293
202589_at	TYMS	11	2.08E-07	0.00293
204127_at	RFC3	11	2.08E-07	0.00293
205193_at	MAFF	11	2.08E-07	0.00293
209211_at	KLF5	11	2.08E-07	0.00293
219049_at	ChGn	11	2.08E-07	0.00293
223255_at	KIAA1333	11	2.08E-07	0.00293
228974_at	ZNF677	11	2.08E-07	0.00293
AFFX-PheX-M_at		11	2.08E-07	0.00293
216237_s_at	MCM5	11	2.08E-07	0.00293
224407_s_at	MASK	11	2.08E-07	0.00293
201792_at	AEBP1	11	2.08E-07	0.00293
202581_at	HSPA1B	11	2.08E-07	0.00293
202688_at	TNFSF10	11	2.08E-07	0.00293
214132_at	ATP5C1	11	2.08E-07	0.00293
214820_at	BRWD1	11	2.08E-07	0.00293
226998_at	NARG1	11	2.08E-07	0.00293
228334_x_at	KIAA1712	11	2.08E-07	0.00293
229391_s_at	LOC441168	11	2.08E-07	0.00293
235412_at	ARHGEF7	11	2.08E-07	0.00293
236114_at	RUNX1	11	2.08E-07	0.00293
AFFX-PheX-3_at		11	2.08E-07	0.00293
201211_s_at	DDX3X	11	2.08E-07	0.00293
205092_x_at	ZBTB1	11	2.08E-07	0.00293
207838_x_at	PBXIP1	11	2.08E-07	0.00293
209679_s_at	LOC57228	11	2.08E-07	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
221848_at	ZGPAT	11	2.08E-07	0.00293
226613_at	TBC1D10A	11	2.08E-07	0.00293
228328_at		11	2.08E-07	0.00293
233014_at	IDI1	11	2.08E-07	0.00293
231975_s_at	FLJ35954	11	2.08E-07	0.00293
241242_at	C9orf10	11	2.08E-07	0.00293
207247_s_at	ZFY	11	2.08E-07	0.00293
211538_s_at	HSPA2	11	2.08E-07	0.00293
221135_s_at	HT001	11	2.08E-07	0.00293
1552701_a_at	COP1	10	2.66E-06	0.00293
1554609_at	MGC12965	10	2.66E-06	0.00293
204092_s_at	STK6	10	2.66E-06	0.00293
204244_s_at	ASK	10	2.66E-06	0.00293
204709_s_at	KIF23	10	2.66E-06	0.00293
207165_at	HMMR	10	2.66E-06	0.00293
211368_s_at	CASP1	10	2.66E-06	0.00293
212977_at	CMKOR1	10	2.66E-06	0.00293
217924_at	C6orf106	10	2.66E-06	0.00293
222040_at	HNRPA1	10	2.66E-06	0.00293
222088_s_at	SLC2A14	10	2.66E-06	0.00293
225902_at	PPIG	10	2.66E-06	0.00293
228751_at	CLK4	10	2.66E-06	0.00293
230461_s_at	MUM1	10	2.66E-06	0.00293
232797_at	ITGAV	10	2.66E-06	0.00293
233899_x_at	ZBTB10	10	2.66E-06	0.00293
235152_at		10	2.66E-06	0.00293
235167_at	DKFZp547E087	10	2.66E-06	0.00293
236402_at	BRAF	10	2.66E-06	0.00293
236752_at	PKP4	10	2.66E-06	0.00293
238712_at		10	2.66E-06	0.00293
243648_at	ZC3H11A	10	2.66E-06	0.00293
AFFX-r2-Bs-thr-3_s_at		10	2.66E-06	0.00293
201774_s_at	CNAP1	10	2.66E-06	0.00293
202338_at	TK1	10	2.66E-06	0.00293
203109_at	UBE2M	10	2.66E-06	0.00293
203422_at	POLD1	10	2.66E-06	0.00293
206917_at	GNA13	10	2.66E-06	0.00293
209093_s_at	GBA	10	2.66E-06	0.00293
210006_at	ABHD14A	10	2.66E-06	0.00293
211366_x_at	CASP1	10	2.66E-06	0.00293
212187_x_at	PTGDS	10	2.66E-06	0.00293
213023_at	UTRN	10	2.66E-06	0.00293
214877_at	CDKAL1	10	2.66E-06	0.00293
219004_s_at	C21orf45	10	2.66E-06	0.00293
219334_s_at	FLJ22833	10	2.66E-06	0.00293
221843_s_at	KIAA1609	10	2.66E-06	0.00293
225762_x_at	LOC284801	10	2.66E-06	0.00293
225786_at	FAM36A	10	2.66E-06	0.00293
227198_at	AFF3	10	2.66E-06	0.00293
228561_at	CDC37L1	10	2.66E-06	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
231069_at		10	2.66E-06	0.00293
231152_at		10	2.66E-06	0.00293
242868_at	EPAS1	10	2.66E-06	0.00293
243829_at	BRAF	10	2.66E-06	0.00293
36711_at	MAFF	10	2.66E-06	0.00293
AFFX-r2-Bs-phe-M_at		10	2.66E-06	0.00293
1552302_at	MGC20235	10	2.66E-06	0.00293
1552703_s_at	COP1	10	2.66E-06	0.00293
1555892_s_at	LOC253039	10	2.66E-06	0.00293
202107_s_at	MCM2	10	2.66E-06	0.00293
202295_s_at	CTSH	10	2.66E-06	0.00293
206548_at	FLJ23556	10	2.66E-06	0.00293
207030_s_at	CSRP2	10	2.66E-06	0.00293
208881_x_at	IDI1	10	2.66E-06	0.00293
210589_s_at	GBA	10	2.66E-06	0.00293
210845_s_at	PLAUR	10	2.66E-06	0.00293
211676_s_at	IFNGR1	10	2.66E-06	0.00293
213302_at	PFAS	10	2.66E-06	0.00293
219854_at	ZNF14	10	2.66E-06	0.00293
224558_s_at	MALAT1	10	2.66E-06	0.00293
224973_at	FAM46A	10	2.66E-06	0.00293
225227_at		10	2.66E-06	0.00293
227265_at		10	2.66E-06	0.00293
234995_at	LOC152185	10	2.66E-06	0.00293
240572_s_at	LOC374443	10	2.66E-06	0.00293
1553984_s_at	DTYMK	10	2.66E-06	0.00293
1562836_at	DDX6	10	2.66E-06	0.00293
1563469_at	ARID5B	10	2.66E-06	0.00293
202687_s_at	TNFSF10	10	2.66E-06	0.00293
204170_s_at	CKS2	10	2.66E-06	0.00293
212514_x_at	DDX3X	10	2.66E-06	0.00293
226444_at	SLC39A10	10	2.66E-06	0.00293
228742_at		10	2.66E-06	0.00293
229735_s_at	NPAL3	10	2.66E-06	0.00293
230036_at	SAMD9L	10	2.66E-06	0.00293
231269_at	DJ467N11.1	10	2.66E-06	0.00293
232978_at	MBNL2	10	2.66E-06	0.00293
233849_s_at	ARHGAP5	10	2.66E-06	0.00293
236974_at	CCNI	10	2.66E-06	0.00293
241627_x_at	FLJ10357	10	2.66E-06	0.00293
1557813_at	SSBP2	10	2.66E-06	0.00293
205449_at	SAC3D1	10	2.66E-06	0.00293
214787_at	MYCPBP	10	2.66E-06	0.00293
221428_s_at	TBL1XR1	10	2.66E-06	0.00293
221763_at	JMJD1C	10	2.66E-06	0.00293
244811_at	PHIP	10	2.66E-06	0.00293
1557522_x_at	BLCAP	10	2.66E-06	0.00293
204868_at	ICT1	10	2.66E-06	0.00293
211752_s_at	NDUFS7	10	2.66E-06	0.00293
212037_at	PNN	10	2.66E-06	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
218166_s_at	HBXAP	10	2.66E-06	0.00293
219470_x_at	CCNJ	10	2.66E-06	0.00293
228758_at	LOC389185	10	2.66E-06	0.00293
232382_s_at	LOC115294	10	2.66E-06	0.00293
234411_x_at	CD44	10	2.66E-06	0.00293
238465_at	MGC33648	10	2.66E-06	0.00293
239493_at	RPL7	10	2.66E-06	0.00293
234881_at		10	2.66E-06	0.00293
238456_at		10	2.66E-06	0.00293
1558401_at		9	2.72E-05	0.00293
1569157_s_at	LOC162993	9	2.72E-05	0.00293
201136_at	PLP2	9	2.72E-05	0.00293
201294_s_at	WSB1	9	2.72E-05	0.00293
203022_at	RNASEH2A	9	2.72E-05	0.00293
206632_s_at	APOBEC3B	9	2.72E-05	0.00293
208116_s_at	MAN1A1	9	2.72E-05	0.00293
208733_at	RAB2	9	2.72E-05	0.00293
209164_s_at	CYB561	9	2.72E-05	0.00293
210172_at	SF1	9	2.72E-05	0.00293
212583_at	AQR	9	2.72E-05	0.00293
212847_at	FUBP1	9	2.72E-05	0.00293
212907_at	SLC30A1	9	2.72E-05	0.00293
213182_x_at	CDKN1C	9	2.72E-05	0.00293
219215_s_at	SLC39A4	9	2.72E-05	0.00293
221555_x_at	CDC14B	9	2.72E-05	0.00293
222316_at	VDP	9	2.72E-05	0.00293
222662_at	LOC286044	9	2.72E-05	0.00293
223103_at	STARD10	9	2.72E-05	0.00293
225834_at	LOC440687	9	2.72E-05	0.00293
226682_at	LOC283666	9	2.72E-05	0.00293
228153_at	IBRDC2	9	2.72E-05	0.00293
229256_at	PGM2L1	9	2.72E-05	0.00293
232573_at	FBXO32	9	2.72E-05	0.00293
235440_at	FLJ39441	9	2.72E-05	0.00293
235592_at	ELL2	9	2.72E-05	0.00293
238909_at	S100A10	9	2.72E-05	0.00293
242691_at		9	2.72E-05	0.00293
244353_s_at	SLC2A12	9	2.72E-05	0.00293
AFFX-r2-Bs-phe-3_at		9	2.72E-05	0.00293
1558102_at	TM6SF1	9	2.72E-05	0.00293
1566901_at	TGIF	9	2.72E-05	0.00293
201626_at	INSIG1	9	2.72E-05	0.00293
201627_s_at	INSIG1	9	2.72E-05	0.00293
201702_s_at	PPP1R10	9	2.72E-05	0.00293
202708_s_at	HIST2H2BE	9	2.72E-05	0.00293
202869_at	OAS1	9	2.72E-05	0.00293
202979_s_at	ZF	9	2.72E-05	0.00293
203828_s_at	IL32	9	2.72E-05	0.00293
204502_at	SAMHD1	9	2.72E-05	0.00293
205681_at	BCL2A1	9	2.72E-05	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
205798_at	IL7R	9	2.72E-05	0.00293
208859_s_at	ATRX	9	2.72E-05	0.00293
209417_s_at	IFI35	9	2.72E-05	0.00293
209512_at	HSDL2	9	2.72E-05	0.00293
209681_at	SLC19A2	9	2.72E-05	0.00293
210778_s_at	MXD4	9	2.72E-05	0.00293
212944_at	MRPS6	9	2.72E-05	0.00293
213372_at	PAQR3	9	2.72E-05	0.00293
216766_at	PRKCE	9	2.72E-05	0.00293
218585_s_at	DTL	9	2.72E-05	0.00293
218918_at	MAN1C1	9	2.72E-05	0.00293
219622_at	RAB20	9	2.72E-05	0.00293
219978_s_at	NUSAP1	9	2.72E-05	0.00293
220941_s_at	C21orf91	9	2.72E-05	0.00293
221276_s_at	SYNC1	9	2.72E-05	0.00293
221985_at	KLHL24	9	2.72E-05	0.00293
222592_s_at	ACSL5	9	2.72E-05	0.00293
225107_at	LOC442518	9	2.72E-05	0.00293
225876_at	NPAL3	9	2.72E-05	0.00293
226158_at	KLHL24	9	2.72E-05	0.00293
226390_at	STARD4	9	2.72E-05	0.00293
226542_at		9	2.72E-05	0.00293
227040_at	LOC387921	9	2.72E-05	0.00293
228160_at	LOC400642	9	2.72E-05	0.00293
228468_at	MASTL	9	2.72E-05	0.00293
228674_s_at	EML4	9	2.72E-05	0.00293
229574_at	TRA2A	9	2.72E-05	0.00293
229685_at	TBC1D14	9	2.72E-05	0.00293
230082_at	LRRFIP2	9	2.72E-05	0.00293
234973_at	SLC38A5	9	2.72E-05	0.00293
235190_at	CALM2	9	2.72E-05	0.00293
236072_at	ATAD1	9	2.72E-05	0.00293
236989_at	EIF3S6	9	2.72E-05	0.00293
238452_at	FCRLM2	9	2.72E-05	0.00293
238781_at	SFRS12	9	2.72E-05	0.00293
242836_at		9	2.72E-05	0.00293
243016_at	TYMS	9	2.72E-05	0.00293
1565703_at	SMAD4	9	2.72E-05	0.00293
1565951_s_at	CHML	9	2.72E-05	0.00293
201755_at	MCM5	9	2.72E-05	0.00293
202464_s_at	PFKFB3	9	2.72E-05	0.00293
202779_s_at	UBE2S	9	2.72E-05	0.00293
203836_s_at	MAP3K5	9	2.72E-05	0.00293
204194_at	BACH1	9	2.72E-05	0.00293
205139_s_at	UST	9	2.72E-05	0.00293
205436_s_at	H2AFX	9	2.72E-05	0.00293
208744_x_at	HSPH1	9	2.72E-05	0.00293
209163_at	CYB561	9	2.72E-05	0.00293
209421_at	MSH2	9	2.72E-05	0.00293
209840_s_at	LRRN3	9	2.72E-05	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
210424_s_at	GOLGA8A	9	2.72E-05	0.00293
210793_s_at	NUP98	9	2.72E-05	0.00293
211323_s_at	ITPR1	9	2.72E-05	0.00293
211748_x_at	PTGDS	9	2.72E-05	0.00293
212746_s_at	KIAA0470	9	2.72E-05	0.00293
212890_at	MGC15523	9	2.72E-05	0.00293
215011_at	RNU17D	9	2.72E-05	0.00293
217949_s_at	VKORC1	9	2.72E-05	0.00293
218317_x_at	GIYD2	9	2.72E-05	0.00293
220148_at	ALDH8A1	9	2.72E-05	0.00293
222429_at	DBNL	9	2.72E-05	0.00293
223276_at	NID67	9	2.72E-05	0.00293
223407_at	C16orf48	9	2.72E-05	0.00293
223577_x_at	MALAT1	9	2.72E-05	0.00293
225541_at	RPL22L1	9	2.72E-05	0.00293
225899_x_at	LOC440395	9	2.72E-05	0.00293
226922_at	RANBP2	9	2.72E-05	0.00293
226923_at	SCFD2	9	2.72E-05	0.00293
228193_s_at	RGC32	9	2.72E-05	0.00293
228570_at	BTBD11	9	2.72E-05	0.00293
229075_at		9	2.72E-05	0.00293
229393_at	L3MBTL3	9	2.72E-05	0.00293
233292_s_at	ANKHD1	9	2.72E-05	0.00293
235787_at	CDC37L1	9	2.72E-05	0.00293
235926_at	ANAPC5	9	2.72E-05	0.00293
236703_at		9	2.72E-05	0.00293
239228_at	CSNK2A1	9	2.72E-05	0.00293
242110_at	ARHGAP5	9	2.72E-05	0.00293
244054_at	SKIL	9	2.72E-05	0.00293
1555247_a_at	RAPGEF6	9	2.72E-05	0.00293
1558837_a_at		9	2.72E-05	0.00293
1559401_a_at	ZNF609	9	2.72E-05	0.00293
1570021_at	LOC360030	9	2.72E-05	0.00293
200832_s_at	SCD	9	2.72E-05	0.00293
201435_s_at	EIF4E	9	2.72E-05	0.00293
201601_x_at	IFITM1	9	2.72E-05	0.00293
201673_s_at	GYS1	9	2.72E-05	0.00293
201712_s_at	RANBP2	9	2.72E-05	0.00293
203209_at	RFC5	9	2.72E-05	0.00293
204243_at	RLF	9	2.72E-05	0.00293
204270_at	FLJ13941	9	2.72E-05	0.00293
204331_s_at	MRPS12	9	2.72E-05	0.00293
204545_at	PEX6	9	2.72E-05	0.00293
204857_at	MAD1L1	9	2.72E-05	0.00293
207713_s_at	C20orf18	9	2.72E-05	0.00293
210048_at	NAPG	9	2.72E-05	0.00293
217370_x_at	FUS	9	2.72E-05	0.00293
217936_at	ARHGAP5	9	2.72E-05	0.00293
220934_s_at	MGC3196	9	2.72E-05	0.00293
221919_at	HNRPA1	9	2.72E-05	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
222872_x_at	FLJ22833	9	2.72E-05	0.00293
223025_s_at	AP1M1	9	2.72E-05	0.00293
223839_s_at	SCD	9	2.72E-05	0.00293
224315_at	DDX20	9	2.72E-05	0.00293
225539_at	ZNF295	9	2.72E-05	0.00293
225747_at	FLJ32452	9	2.72E-05	0.00293
225766_s_at	TNPO1	9	2.72E-05	0.00293
226018_at	Ells1	9	2.72E-05	0.00293
226541_at	FBXO30	9	2.72E-05	0.00293
228812_at		9	2.72E-05	0.00293
229151_at	SLC14A1	9	2.72E-05	0.00293
231252_at	FLJ23861	9	2.72E-05	0.00293
231793_s_at	CAMK2D	9	2.72E-05	0.00293
232092_at	MCART1	9	2.72E-05	0.00293
233406_at	KIAA0256	9	2.72E-05	0.00293
235067_at	MKLN1	9	2.72E-05	0.00293
235848_x_at		9	2.72E-05	0.00293
239334_at	FAM62B	9	2.72E-05	0.00293
239630_at		9	2.72E-05	0.00293
239835_at	KBTBD8	9	2.72E-05	0.00293
239979_at	EPSTI1	9	2.72E-05	0.00293
240141_at		9	2.72E-05	0.00293
241388_at		9	2.72E-05	0.00293
241837_at	ARID5B	9	2.72E-05	0.00293
242539_at	MGC42174	9	2.72E-05	0.00293
243659_at	ELL2	9	2.72E-05	0.00293
AFFX-PheX-5_at		9	2.72E-05	0.00293
1555878_at	RPS24	9	2.72E-05	0.00293
1561640_at	PSMD5	9	2.72E-05	0.00293
201660_at	ACSL3	9	2.72E-05	0.00293
204824_at	ENDOG	9	2.72E-05	0.00293
205241_at	SCO2	9	2.72E-05	0.00293
205249_at	EGR2	9	2.72E-05	0.00293
206613_s_at	TAF1A	9	2.72E-05	0.00293
209482_at	POP7	9	2.72E-05	0.00293
209919_x_at	GGT1	9	2.72E-05	0.00293
210075_at	02.Mär	9	2.72E-05	0.00293
210716_s_at	RSN	9	2.72E-05	0.00293
212930_at	ATP2B1	9	2.72E-05	0.00293
215043_s_at	SMA5	9	2.72E-05	0.00293
215567_at	C14orf111	9	2.72E-05	0.00293
217954_s_at	PHF3	9	2.72E-05	0.00293
218070_s_at	GMPPA	9	2.72E-05	0.00293
218399_s_at	CDCA4	9	2.72E-05	0.00293
218529_at	CD320	9	2.72E-05	0.00293
219164_s_at	C14orf103	9	2.72E-05	0.00293
220368_s_at	KIAA2010	9	2.72E-05	0.00293
221826_at	ANGEL2	9	2.72E-05	0.00293
224582_s_at	NUCKS1	9	2.72E-05	0.00293
225197_at	PRO0149	9	2.72E-05	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
225870_s_at	TRAPPC5	9	2.72E-05	0.00293
225898_at	WDR54	9	2.72E-05	0.00293
227031_at	SNX13	9	2.72E-05	0.00293
227751_at	PDCD5	9	2.72E-05	0.00293
230850_at	FMNL3	9	2.72E-05	0.00293
230937_at	LOC442200	9	2.72E-05	0.00293
235698_at	ZFP90	9	2.72E-05	0.00293
235791_x_at	CHD1	9	2.72E-05	0.00293
236059_at	FLJ23342	9	2.72E-05	0.00293
239106_at	CA5BL	9	2.72E-05	0.00293
239437_at		9	2.72E-05	0.00293
240600_at	AP3B1	9	2.72E-05	0.00293
243709_at	FLJ90709	9	2.72E-05	0.00293
243788_at	PHF11	9	2.72E-05	0.00293
1554229_at	LOC153222	9	2.72E-05	0.00293
1559025_at	09.Sep	9	2.72E-05	0.00293
1563497_at	USP25	9	2.72E-05	0.00293
201284_s_at	APEH	9	2.72E-05	0.00293
202331_at	BCKDHA	9	2.72E-05	0.00293
202928_s_at	PHF1	9	2.72E-05	0.00293
203189_s_at	NDUFS8	9	2.72E-05	0.00293
206976_s_at	HSPH1	9	2.72E-05	0.00293
210950_s_at	FDFT1	9	2.72E-05	0.00293
222569_at	UGCGL1	9	2.72E-05	0.00293
223109_at	TRUB2	9	2.72E-05	0.00293
223585_x_at	KBTBD2	9	2.72E-05	0.00293
224489_at	LOC284058	9	2.72E-05	0.00293
225830_at	PDZK8	9	2.72E-05	0.00293
225878_at	KIF1B	9	2.72E-05	0.00293
226370_at	KLHL15	9	2.72E-05	0.00293
226942_at	PHF20L1	9	2.72E-05	0.00293
228655_at		9	2.72E-05	0.00293
230970_at	SSH2	9	2.72E-05	0.00293
232628_at	FAM13A1	9	2.72E-05	0.00293
235390_at	P18SRP	9	2.72E-05	0.00293
235557_at	LOC150763	9	2.72E-05	0.00293
235860_at	KCMF1	9	2.72E-05	0.00293
236310_at	ZNF37B	9	2.72E-05	0.00293
238032_at	DHRS3	9	2.72E-05	0.00293
238496_at	WHSC1L1	9	2.72E-05	0.00293
239070_at		9	2.72E-05	0.00293
AFFX-LysX-3_at		9	2.72E-05	0.00293
1559946_s_at	RUVBL2	9	2.72E-05	0.00293
200822_x_at	TPI1	9	2.72E-05	0.00293
204080_at	TOE1	9	2.72E-05	0.00293
212436_at	TRIM33	9	2.72E-05	0.00293
213672_at	MARS	9	2.72E-05	0.00293
216960_s_at	ZNF133	9	2.72E-05	0.00293
219433_at	BCOR	9	2.72E-05	0.00293
222133_s_at	PHF20L1	9	2.72E-05	0.00293

Probe set Id	Symbol	# OS DE	OS p-value	OS q-value
222580_at	ZNF644	9	2.72E-05	0.00293
223515_s_at	COQ3	9	2.72E-05	0.00293
224956_at	182-FIP	9	2.72E-05	0.00293
225565_at	CREB1	9	2.72E-05	0.00293
225743_at	RPUSD3	9	2.72E-05	0.00293
225821_s_at	KIAA1327	9	2.72E-05	0.00293
226934_at	CPSF6	9	2.72E-05	0.00293
228764_s_at	MGC5987	9	2.72E-05	0.00293
230120_s_at	LOC400966	9	2.72E-05	0.00293
235689_at	MTFMT	9	2.72E-05	0.00293
239946_at	KIAA0922	9	2.72E-05	0.00293
200776_s_at	LOC151579	9	2.72E-05	0.00293
229914_at	FLJ38717	9	2.72E-05	0.00293
231319_x_at	KIF9	9	2.72E-05	0.00293
232709_at		9	2.72E-05	0.00293
236629_at	C1orf69	9	2.72E-05	0.00293
213213_at	DATF1	9	2.72E-05	0.00293

Probe sets associated with senescence

The q-value of the listed probe sets is <0.005 in the senescence experimental group (SEN). Every probe set is differentially expressed in at least 9 of 18 senescence experiments. For each probe set, the number of senescence experiments where the probe set was differentially expressed (# Sen DE), the p-value and the q-value are displayed.

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
209610_s_at	SLC1A4	18	2.28E-17	0.00156
212810_s_at	SLC1A4	17	3.06E-15	0.00156
222912_at	ARRB1	17	3.06E-15	0.00156
212811_x_at	SLC1A4	16	1.94E-13	0.00156
225647_s_at	CTSC	15	7.74E-12	0.00156
207761_s_at	DKFZP586A0522	15	7.74E-12	0.00156
226142_at	GLIPR1	15	7.74E-12	0.00156
228708_at	RAB27B	15	7.74E-12	0.00156
218145_at	TRIB3	15	7.74E-12	0.00156
204748_at	PTGS2	14	2.17E-10	0.00156
201397_at	PHGDH	14	2.17E-10	0.00156
220576_at	PGAP1	14	2.17E-10	0.00156
225842_at	PHLDA1	14	2.17E-10	0.00156
229450_at	IFIT3	14	2.17E-10	0.00156
226225_at	MCC	14	2.17E-10	0.00156
203542_s_at	KLF9	14	2.17E-10	0.00156
212675_s_at	KIAA0582	14	2.17E-10	0.00156
200742_s_at	TPP1	14	2.17E-10	0.00156
221539_at	EIF4EBP1	14	2.17E-10	0.00156
221911_at	ETV1	14	2.17E-10	0.00156
229872_s_at	LOC440667	14	2.17E-10	0.00156
231907_at	ABL2	14	2.17E-10	0.00156
212677_s_at	RAB1A	14	2.17E-10	0.00156
201340_s_at	ENC1	13	4.56E-09	0.00156
201867_s_at	TBL1X	13	4.56E-09	0.00156
204529_s_at	TOX	13	4.56E-09	0.00156
204612_at	PKIA	13	4.56E-09	0.00156
210299_s_at	FHL1	13	4.56E-09	0.00156
226610_at	PRR6	13	4.56E-09	0.00156
238599_at	IRAK1BP1	13	4.56E-09	0.00156
202934_at	HK2	13	4.56E-09	0.00156
209189_at	FOS	13	4.56E-09	0.00156
215177_s_at	ITGA6	13	4.56E-09	0.00156
221011_s_at	LBH	13	4.56E-09	0.00156
201487_at	CTSC	13	4.56E-09	0.00156
204256_at	ELOVL6	13	4.56E-09	0.00156
204749_at	NAP1L3	13	4.56E-09	0.00156
208926_at	NEU1	13	4.56E-09	0.00156
216236_s_at	SLC2A3	13	4.56E-09	0.00156
217996_at	PHLDA1	13	4.56E-09	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
222862_s_at	AK5	13	4.56E-09	0.00156
227224_at	RALGPS2	13	4.56E-09	0.00156
227486_at	NT5E	13	4.56E-09	0.00156
228033_at	E2F7	13	4.56E-09	0.00156
239614_x_at	GLS	13	4.56E-09	0.00156
242051_at		13	4.56E-09	0.00156
213524_s_at	G0S2	13	4.56E-09	0.00156
225646_at	CTSC	13	4.56E-09	0.00156
235408_x_at	ZNF117	13	4.56E-09	0.00156
1553613_s_at	FOXC1	13	4.56E-09	0.00156
202679_at	NPC1	13	4.56E-09	0.00156
204070_at	RARRES3	13	4.56E-09	0.00156
209459_s_at	ABAT	13	4.56E-09	0.00156
213067_at	MYH10	13	4.56E-09	0.00156
224851_at	CDK6	13	4.56E-09	0.00156
226757_at	IFIT2	13	4.56E-09	0.00156
242560_at	FANCD2	13	4.56E-09	0.00156
1555788_a_at	TRIB3	13	4.56E-09	0.00156
204222_s_at	GLIPR1	13	4.56E-09	0.00156
211272_s_at	DGKA	13	4.56E-09	0.00156
215307_at		13	4.56E-09	0.00156
217914_at	TPCN1	13	4.56E-09	0.00156
218953_s_at	MGC3265	13	4.56E-09	0.00156
227530_at	AKAP12	13	4.56E-09	0.00156
228726_at	SERPINB1	13	4.56E-09	0.00156
231993_at		13	4.56E-09	0.00156
209217_s_at	WDR45	13	4.56E-09	0.00156
227140_at	INHBA	13	4.56E-09	0.00156
201313_at	ENO2	12	7.43E-08	0.00156
202086_at	MX1	12	7.43E-08	0.00156
202291_s_at	MGP	12	7.43E-08	0.00156
204279_at	PSMB9	12	7.43E-08	0.00156
204457_s_at	GAS1	12	7.43E-08	0.00156
207826_s_at	ID3	12	7.43E-08	0.00156
208442_s_at	ATM	12	7.43E-08	0.00156
210095_s_at	IGFBP3	12	7.43E-08	0.00156
214731_at	CTTNBP2NL	12	7.43E-08	0.00156
218486_at	KLF11	12	7.43E-08	0.00156
223578_x_at	PRO1073	12	7.43E-08	0.00156
227449_at	EPHA4	12	7.43E-08	0.00156
229899_s_at	HSUP1	12	7.43E-08	0.00156
242273_at		12	7.43E-08	0.00156
201079_at	SYNGR2	12	7.43E-08	0.00156
202528_at	GALE	12	7.43E-08	0.00156
202847_at	PCK2	12	7.43E-08	0.00156
203139_at	DAPK1	12	7.43E-08	0.00156
203358_s_at	EZH2	12	7.43E-08	0.00156
203498_at	DSCR1L1	12	7.43E-08	0.00156
204011_at	SPRY2	12	7.43E-08	0.00156
204589_at	NUAK1	12	7.43E-08	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
210001_s_at	SOCS1	12	7.43E-08	0.00156
228702_at	FLJ43663	12	7.43E-08	0.00156
229744_at	SSFA2	12	7.43E-08	0.00156
230179_at	LOC285812	12	7.43E-08	0.00156
230930_at	LOC338620	12	7.43E-08	0.00156
231775_at	TNFRSF10A	12	7.43E-08	0.00156
233020_at	SEC22L1	12	7.43E-08	0.00156
241763_s_at	FBXO32	12	7.43E-08	0.00156
201341_at	ENC1	12	7.43E-08	0.00156
201540_at	FHL1	12	7.43E-08	0.00156
201739_at	SGK	12	7.43E-08	0.00156
202887_s_at	DDIT4	12	7.43E-08	0.00156
202912_at	ADM	12	7.43E-08	0.00156
203851_at	IGFBP6	12	7.43E-08	0.00156
204596_s_at	STC1	12	7.43E-08	0.00156
205803_s_at	TRPC1	12	7.43E-08	0.00156
209383_at	DDIT3	12	7.43E-08	0.00156
217967_s_at	C1orf24	12	7.43E-08	0.00156
219118_at	FKBP11	12	7.43E-08	0.00156
221973_at		12	7.43E-08	0.00156
221974_at	SNRPN	12	7.43E-08	0.00156
222931_s_at	THNSL1	12	7.43E-08	0.00156
223196_s_at	SESN2	12	7.43E-08	0.00156
225496_s_at	SYTL2	12	7.43E-08	0.00156
227326_at	MXRA7	12	7.43E-08	0.00156
227345_at	TNFRSF10D	12	7.43E-08	0.00156
228754_at	GRIP2	12	7.43E-08	0.00156
228774_at	C9orf81	12	7.43E-08	0.00156
231798_at	NOG	12	7.43E-08	0.00156
232344_at	RASA1	12	7.43E-08	0.00156
242751_at	PRDX6	12	7.43E-08	0.00156
1555950_a_at	DAF	12	7.43E-08	0.00156
200907_s_at	KIAA0992	12	7.43E-08	0.00156
202388_at	RGS2	12	7.43E-08	0.00156
204472_at	GEM	12	7.43E-08	0.00156
204595_s_at	STC1	12	7.43E-08	0.00156
205505_at	GCNT1	12	7.43E-08	0.00156
213112_s_at	SQSTM1	12	7.43E-08	0.00156
213850_s_at	SFRS2IP	12	7.43E-08	0.00156
226136_at		12	7.43E-08	0.00156
227354_at	PAG1	12	7.43E-08	0.00156
229374_at	EPHA4	12	7.43E-08	0.00156
203386_at	TBC1D4	12	7.43E-08	0.00156
207057_at	SLC16A7	12	7.43E-08	0.00156
209160_at	AKR1C3	12	7.43E-08	0.00156
217998_at	PHLDA1	12	7.43E-08	0.00156
217999_s_at	PHLDA1	12	7.43E-08	0.00156
203157_s_at	GLS	12	7.43E-08	0.00156
203708_at	PDE4B	12	7.43E-08	0.00156
206342_x_at	IDS	12	7.43E-08	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
210511_s_at	INHBA	12	7.43E-08	0.00156
212311_at	KIAA0746	12	7.43E-08	0.00156
219773_at	NOX4	12	7.43E-08	0.00156
221019_s_at	COLEC12	12	7.43E-08	0.00156
225919_s_at	C9orf72	12	7.43E-08	0.00156
235629_at	FN1	12	7.43E-08	0.00156
235944_at	HMCN1	12	7.43E-08	0.00156
238622_at	RAP2B	12	7.43E-08	0.00156
202625_at	LYN	12	7.43E-08	0.00156
205463_s_at	PDGFA	12	7.43E-08	0.00156
1555812_a_at	ARHGDIB	12	7.43E-08	0.00156
200743_s_at	TPP1	12	7.43E-08	0.00156
227529_s_at	AKAP12	12	7.43E-08	0.00156
217853_at	TNS3	12	7.43E-08	0.00156
203628_at	IGF1R	11	9.59E-07	0.00156
203919_at	TCEA2	11	9.59E-07	0.00156
204415_at	G1P3	11	9.59E-07	0.00156
209031_at	IGSF4	11	9.59E-07	0.00156
213164_at	MRPS6	11	9.59E-07	0.00156
213459_at	RPL37A	11	9.59E-07	0.00156
214995_s_at	KIAA0907	11	9.59E-07	0.00156
224724_at	SULF2	11	9.59E-07	0.00156
226876_at	MGC45871	11	9.59E-07	0.00156
228846_at	MXD1	11	9.59E-07	0.00156
229072_at		11	9.59E-07	0.00156
232914_s_at	SYTL2	11	9.59E-07	0.00156
238931_at	LOC284009	11	9.59E-07	0.00156
244766_at	LAT1-3TM	11	9.59E-07	0.00156
1553995_a_at	NT5E	11	9.59E-07	0.00156
1555827_at	CCNL1	11	9.59E-07	0.00156
1555847_a_at	LOC284454	11	9.59E-07	0.00156
200962_at	TBC1D8	11	9.59E-07	0.00156
203404_at	ARMCX2	11	9.59E-07	0.00156
204162_at	KNTC2	11	9.59E-07	0.00156
205081_at	CRIP1	11	9.59E-07	0.00156
208690_s_at	PDLIM1	11	9.59E-07	0.00156
209016_s_at	KRT7	11	9.59E-07	0.00156
209891_at	SPBC25	11	9.59E-07	0.00156
209911_x_at	HIST1H2BD	11	9.59E-07	0.00156
210524_x_at		11	9.59E-07	0.00156
213865_at	DCBLD2	11	9.59E-07	0.00156
215071_s_at	HIST1H2AC	11	9.59E-07	0.00156
217979_at	TSPAN13	11	9.59E-07	0.00156
221778_at	KIAA1718	11	9.59E-07	0.00156
225415_at	DTX3L	11	9.59E-07	0.00156
227223_at	RNPC2	11	9.59E-07	0.00156
227467_at	RDH10	11	9.59E-07	0.00156
228948_at	EPHA4	11	9.59E-07	0.00156
230746_s_at	STC1	11	9.59E-07	0.00156
235652_at	SCML1	11	9.59E-07	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
235803_at	CRLF3	11	9.59E-07	0.00156
236769_at	LOC158402	11	9.59E-07	0.00156
239108_at	MLSTD1	11	9.59E-07	0.00156
1552622_s_at	MGC13098	11	9.59E-07	0.00156
202600_s_at	NRIP1	11	9.59E-07	0.00156
203989_x_at	F2R	11	9.59E-07	0.00156
204497_at	ADCY9	11	9.59E-07	0.00156
204620_s_at	CSPG2	11	9.59E-07	0.00156
204863_s_at	IL6ST	11	9.59E-07	0.00156
205771_s_at	AKAP7	11	9.59E-07	0.00156
206744_s_at	ZMYM5	11	9.59E-07	0.00156
208003_s_at	NFAT5	11	9.59E-07	0.00156
208998_at	UCP2	11	9.59E-07	0.00156
209706_at	NKX3-1	11	9.59E-07	0.00156
212624_s_at	CHN1	11	9.59E-07	0.00156
213506_at	F2RL1	11	9.59E-07	0.00156
214196_s_at	TPP1	11	9.59E-07	0.00156
219587_at	TTC12	11	9.59E-07	0.00156
221731_x_at	CSPG2	11	9.59E-07	0.00156
223125_s_at	C1orf21	11	9.59E-07	0.00156
223199_at	MKNK2	11	9.59E-07	0.00156
223484_at	NMES1	11	9.59E-07	0.00156
225081_s_at	CDCA7L	11	9.59E-07	0.00156
225421_at	ACY1L2	11	9.59E-07	0.00156
226085_at	CBX5	11	9.59E-07	0.00156
226707_at	NAPRT1	11	9.59E-07	0.00156
227917_at		11	9.59E-07	0.00156
228497_at	SLC22A15	11	9.59E-07	0.00156
229515_at	PAWR	11	9.59E-07	0.00156
235028_at		11	9.59E-07	0.00156
235106_at	MAML2	11	9.59E-07	0.00156
238021_s_at	LOC388279	11	9.59E-07	0.00156
240690_at	DKFZp761P0423	11	9.59E-07	0.00156
242467_at	CSNK1A1	11	9.59E-07	0.00156
1558173_a_at	LUZP1	11	9.59E-07	0.00156
201212_at	LGMN	11	9.59E-07	0.00156
202206_at	ARL7	11	9.59E-07	0.00156
203387_s_at	TBC1D4	11	9.59E-07	0.00156
204014_at	DUSP4	11	9.59E-07	0.00156
205027_s_at	MAP3K8	11	9.59E-07	0.00156
206693_at	IL7	11	9.59E-07	0.00156
208893_s_at	DUSP6	11	9.59E-07	0.00156
210715_s_at	SPINT2	11	9.59E-07	0.00156
211668_s_at	PLAU	11	9.59E-07	0.00156
212190_at	SERPINE2	11	9.59E-07	0.00156
212444_at		11	9.59E-07	0.00156
216598_s_at	CCL2	11	9.59E-07	0.00156
218409_s_at	DNAJC1	11	9.59E-07	0.00156
219352_at	HERC6	11	9.59E-07	0.00156
221757_at	MGC17330	11	9.59E-07	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
221865_at	C9orf91	11	9.59E-07	0.00156
223195_s_at	SESN2	11	9.59E-07	0.00156
223712_at	PCBD2	11	9.59E-07	0.00156
226348_at		11	9.59E-07	0.00156
228423_at	LOC389237	11	9.59E-07	0.00156
229071_at	LOC388327	11	9.59E-07	0.00156
229307_at	ANKRD28	11	9.59E-07	0.00156
230147_at	F2RL2	11	9.59E-07	0.00156
230508_at	DKK3	11	9.59E-07	0.00156
235266_at	ATAD2	11	9.59E-07	0.00156
235836_at	MXRA7	11	9.59E-07	0.00156
236545_at	PPP3CA	11	9.59E-07	0.00156
236907_at	PABPC1	11	9.59E-07	0.00156
239049_at		11	9.59E-07	0.00156
242121_at	RNF12	11	9.59E-07	0.00156
244341_at	MAK3	11	9.59E-07	0.00156
244753_at	ACTN4	11	9.59E-07	0.00156
1554016_a_at	FLJ13154	11	9.59E-07	0.00156
200879_s_at	EPAS1	11	9.59E-07	0.00156
201107_s_at	THBS1	11	9.59E-07	0.00156
201580_s_at	TXNDC13	11	9.59E-07	0.00156
201858_s_at	PRG1	11	9.59E-07	0.00156
202669_s_at	EFNB2	11	9.59E-07	0.00156
202838_at	FUCA1	11	9.59E-07	0.00156
204224_s_at	GCH1	11	9.59E-07	0.00156
204338_s_at	RGS4	11	9.59E-07	0.00156
205047_s_at	ASNS	11	9.59E-07	0.00156
206237_s_at	NRG1	11	9.59E-07	0.00156
212979_s_at	KIAA0738	11	9.59E-07	0.00156
215813_s_at	PTGS1	11	9.59E-07	0.00156
218590_at	PEO1	11	9.59E-07	0.00156
219326_s_at	B3GNT1	11	9.59E-07	0.00156
219675_s_at	UXS1	11	9.59E-07	0.00156
221864_at	MGC13024	11	9.59E-07	0.00156
222572_at	PPM2C	11	9.59E-07	0.00156
223217_s_at	NFKBIZ	11	9.59E-07	0.00156
224480_s_at	MGC11324	11	9.59E-07	0.00156
225798_at	JAZF1	11	9.59E-07	0.00156
225929_s_at	C17orf27	11	9.59E-07	0.00156
226460_at	KIAA1450	11	9.59E-07	0.00156
228244_at	BLOC1S3	11	9.59E-07	0.00156
228416_at	ORC4L	11	9.59E-07	0.00156
228461_at	LOC442040	11	9.59E-07	0.00156
228574_at	DKFZp762A217	11	9.59E-07	0.00156
230109_at	PDE7B	11	9.59E-07	0.00156
232031_s_at	KIAA1632	11	9.59E-07	0.00156
235405_at	GSTA4	11	9.59E-07	0.00156
235542_at	MGC22014	11	9.59E-07	0.00156
236356_at	NDUFS1	11	9.59E-07	0.00156
238940_at	KLF12	11	9.59E-07	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
240383_at	UBE2D3	11	9.59E-07	0.00156
242767_at	LMCD1	11	9.59E-07	0.00156
244546_at	CYCS	11	9.59E-07	0.00156
201141_at	GPNMB	11	9.59E-07	0.00156
201360_at	CST3	11	9.59E-07	0.00156
203108_at	GPRC5A	11	9.59E-07	0.00156
203908_at	SLC4A4	11	9.59E-07	0.00156
204678_s_at	KCNK1	11	9.59E-07	0.00156
210534_s_at	EPPB9	11	9.59E-07	0.00156
213032_at	NFIB	11	9.59E-07	0.00156
213703_at	LOC150759	11	9.59E-07	0.00156
214701_s_at	FN1	11	9.59E-07	0.00156
226492_at	SEMA6D	11	9.59E-07	0.00156
227955_s_at		11	9.59E-07	0.00156
228057_at	DDIT4L	11	9.59E-07	0.00156
235457_at	MAML2	11	9.59E-07	0.00156
237411_at	ADAMTS6	11	9.59E-07	0.00156
1557242_at	LOC285835	11	9.59E-07	0.00156
201578_at	PODXL	11	9.59E-07	0.00156
202975_s_at	RHOBTB3	11	9.59E-07	0.00156
204797_s_at	EML1	11	9.59E-07	0.00156
205634_x_at	LENG4	11	9.59E-07	0.00156
206059_at	ZNF91	11	9.59E-07	0.00156
209940_at	PARP3	11	9.59E-07	0.00156
213260_at	FOXC1	11	9.59E-07	0.00156
213572_s_at	SERPINB1	11	9.59E-07	0.00156
219210_s_at	RAB8B	11	9.59E-07	0.00156
238478_at	BNC2	11	9.59E-07	0.00156
203571_s_at	C10orf116	11	9.59E-07	0.00156
204484_at	PIK3C2B	11	9.59E-07	0.00156
227611_at	TARSL2	11	9.59E-07	0.00156
228128_x_at	PAPPA	11	9.59E-07	0.00156
229638_at	IRX3	11	9.59E-07	0.00156
204203_at	CEBPG	11	9.59E-07	0.00156
209199_s_at	MEF2C	11	9.59E-07	0.00156
201169_s_at	BHLHB2	10	9.97E-06	0.00156
201925_s_at	DAF	10	9.97E-06	0.00156
202479_s_at	TRIB2	10	9.97E-06	0.00156
202748_at	GBP2	10	9.97E-06	0.00156
203716_s_at	DPP4	10	9.97E-06	0.00156
203764_at	DLG7	10	9.97E-06	0.00156
204784_s_at	MLF1	10	9.97E-06	0.00156
206448_at	ZNF365	10	9.97E-06	0.00156
208892_s_at	DUSP6	10	9.97E-06	0.00156
209184_s_at	IRS2	10	9.97E-06	0.00156
209238_at	STX3A	10	9.97E-06	0.00156
210512_s_at	VEGF	10	9.97E-06	0.00156
213222_at	PLCB1	10	9.97E-06	0.00156
213226_at	EXOSC9	10	9.97E-06	0.00156
216041_x_at	GRN	10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
216894_x_at	CDKN1C	10	9.97E-06	0.00156
219024_at	PLEKHA1	10	9.97E-06	0.00156
220739_s_at	CNNM3	10	9.97E-06	0.00156
221484_at	B4GALT5	10	9.97E-06	0.00156
224428_s_at	CDCA7	10	9.97E-06	0.00156
225123_at	SESN3	10	9.97E-06	0.00156
227004_at	CDKL5	10	9.97E-06	0.00156
227856_at	FLJ39370	10	9.97E-06	0.00156
228157_at	ZNF207	10	9.97E-06	0.00156
235197_s_at	OSTM1	10	9.97E-06	0.00156
238623_at	MAP3K4	10	9.97E-06	0.00156
241036_at	HPS3	10	9.97E-06	0.00156
242134_at		10	9.97E-06	0.00156
45714_at	HCFC1R1	10	9.97E-06	0.00156
1553185_at	RASEF	10	9.97E-06	0.00156
1559883_s_at	SAMHD1	10	9.97E-06	0.00156
200974_at	ACTA2	10	9.97E-06	0.00156
201392_s_at	IGF2R	10	9.97E-06	0.00156
201752_s_at	ADD3	10	9.97E-06	0.00156
202393_s_at	KLF10	10	9.97E-06	0.00156
202481_at	DHRS3	10	9.97E-06	0.00156
204444_at	KIF11	10	9.97E-06	0.00156
204793_at	GPRASP1	10	9.97E-06	0.00156
205884_at	ITGA4	10	9.97E-06	0.00156
209255_at	KIAA0265	10	9.97E-06	0.00156
209304_x_at	GADD45B	10	9.97E-06	0.00156
210868_s_at	ELOVL6	10	9.97E-06	0.00156
212099_at	RHOB	10	9.97E-06	0.00156
212307_s_at	OGT	10	9.97E-06	0.00156
215191_at	FBXL11	10	9.97E-06	0.00156
215719_x_at	FAS	10	9.97E-06	0.00156
217997_at	PHLDA1	10	9.97E-06	0.00156
218273_s_at	PPM2C	10	9.97E-06	0.00156
218724_s_at	TGIF2	10	9.97E-06	0.00156
219038_at	MORC4	10	9.97E-06	0.00156
221766_s_at	FAM46A	10	9.97E-06	0.00156
222696_at	AXIN2	10	9.97E-06	0.00156
226269_at	GDAP1	10	9.97E-06	0.00156
229881_at	KLF12	10	9.97E-06	0.00156
230958_s_at		10	9.97E-06	0.00156
234989_at	TncRNA	10	9.97E-06	0.00156
235061_at	PPM1K	10	9.97E-06	0.00156
235463_s_at	LASS6	10	9.97E-06	0.00156
236600_at	SPG20	10	9.97E-06	0.00156
238002_at	GOLPH4	10	9.97E-06	0.00156
239251_at	RTN4	10	9.97E-06	0.00156
241722_x_at		10	9.97E-06	0.00156
AFFX-HUMRGE/M10098_5_at		10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
1552733_at	KLHDC1	10	9.97E-06	0.00156
1557270_at		10	9.97E-06	0.00156
1564796_at	EMP1	10	9.97E-06	0.00156
1569312_at	ZNF146	10	9.97E-06	0.00156
201422_at	IFI30	10	9.97E-06	0.00156
201427_s_at	SEPP1	10	9.97E-06	0.00156
202497_x_at	SLC2A3	10	9.97E-06	0.00156
202498_s_at	SLC2A3	10	9.97E-06	0.00156
203408_s_at	SATB1	10	9.97E-06	0.00156
203946_s_at	ARG2	10	9.97E-06	0.00156
204783_at	MLF1	10	9.97E-06	0.00156
205128_x_at	PTGS1	10	9.97E-06	0.00156
205756_s_at	F8	10	9.97E-06	0.00156
207231_at	DZIP3	10	9.97E-06	0.00156
207813_s_at	FDXR	10	9.97E-06	0.00156
207980_s_at	CITED2	10	9.97E-06	0.00156
208190_s_at	LISCH7	10	9.97E-06	0.00156
209754_s_at	TMPO	10	9.97E-06	0.00156
209969_s_at	STAT1	10	9.97E-06	0.00156
210538_s_at	BIRC3	10	9.97E-06	0.00156
211085_s_at	STK4	10	9.97E-06	0.00156
214085_x_at	HRB2	10	9.97E-06	0.00156
214252_s_at	CLN5	10	9.97E-06	0.00156
215646_s_at	CSPG2	10	9.97E-06	0.00156
216870_x_at	DLEU2	10	9.97E-06	0.00156
217127_at	CTH	10	9.97E-06	0.00156
217165_x_at	MT1F	10	9.97E-06	0.00156
217208_s_at	DLG1	10	9.97E-06	0.00156
218701_at	LACTB2	10	9.97E-06	0.00156
218726_at	DKFZp762E1312	10	9.97E-06	0.00156
219209_at	IFIH1	10	9.97E-06	0.00156
219250_s_at	FLRT3	10	9.97E-06	0.00156
219493_at	SHCBP1	10	9.97E-06	0.00156
219684_at	IFRG28	10	9.97E-06	0.00156
221207_s_at	NBEA	10	9.97E-06	0.00156
221922_at	GPSM2	10	9.97E-06	0.00156
222787_s_at	FLJ11273	10	9.97E-06	0.00156
223681_s_at	INADL	10	9.97E-06	0.00156
225626_at	PAG1	10	9.97E-06	0.00156
226218_at	IL7R	10	9.97E-06	0.00156
226239_at	FLJ90024	10	9.97E-06	0.00156
226939_at	CPEB2	10	9.97E-06	0.00156
227124_at		10	9.97E-06	0.00156
227156_at	TNRC8	10	9.97E-06	0.00156
227458_at	CD274	10	9.97E-06	0.00156
227462_at	LRAP	10	9.97E-06	0.00156
228171_s_at	PLEKHG4	10	9.97E-06	0.00156
228480_at	VAPA	10	9.97E-06	0.00156
232235_at	C18orf4	10	9.97E-06	0.00156
234994_at	KIAA1913	10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
235274_at		10	9.97E-06	0.00156
235547_at	PFAAP5	10	9.97E-06	0.00156
236046_at	FLJ44896	10	9.97E-06	0.00156
238458_at	EFHA2	10	9.97E-06	0.00156
238587_at	STS-1	10	9.97E-06	0.00156
239413_at	Cep152	10	9.97E-06	0.00156
239866_at		10	9.97E-06	0.00156
240146_at	CAPZA2	10	9.97E-06	0.00156
240307_at	FUBP1	10	9.97E-06	0.00156
241879_at		10	9.97E-06	0.00156
1568611_at	P4HA2	10	9.97E-06	0.00156
200906_s_at	KIAA0992	10	9.97E-06	0.00156
202205_at	VASP	10	9.97E-06	0.00156
202350_s_at	MATN2	10	9.97E-06	0.00156
202859_x_at	IL8	10	9.97E-06	0.00156
203397_s_at	GALNT3	10	9.97E-06	0.00156
203823_at	RGS3	10	9.97E-06	0.00156
204221_x_at	GLIPR1	10	9.97E-06	0.00156
204326_x_at	MT1X	10	9.97E-06	0.00156
204407_at	TTF2	10	9.97E-06	0.00156
204597_x_at	STC1	10	9.97E-06	0.00156
204638_at	ACP5	10	9.97E-06	0.00156
205200_at	CLEC3B	10	9.97E-06	0.00156
205559_s_at	PCSK5	10	9.97E-06	0.00156
206133_at	BIRC4BP	10	9.97E-06	0.00156
206554_x_at	SETMAR	10	9.97E-06	0.00156
206584_at	LY96	10	9.97E-06	0.00156
206686_at	PDK1	10	9.97E-06	0.00156
206756_at	CHST7	10	9.97E-06	0.00156
208527_x_at	HIST1H2BE	10	9.97E-06	0.00156
208961_s_at	KLF6	10	9.97E-06	0.00156
209258_s_at	CSPG6	10	9.97E-06	0.00156
209276_s_at	GLRX	10	9.97E-06	0.00156
209822_s_at	VLDLR	10	9.97E-06	0.00156
212621_at	KIAA0286	10	9.97E-06	0.00156
214169_at	UNC84A	10	9.97E-06	0.00156
215092_s_at	NFAT5	10	9.97E-06	0.00156
218383_at	C14orf94	10	9.97E-06	0.00156
219315_s_at	C16orf30	10	9.97E-06	0.00156
219806_s_at	FN5	10	9.97E-06	0.00156
219938_s_at	PSTPIP2	10	9.97E-06	0.00156
221833_at	SIAH1	10	9.97E-06	0.00156
221986_s_at	KLHL24	10	9.97E-06	0.00156
222736_s_at	TMEM38B	10	9.97E-06	0.00156
224496_s_at	MGC10744	10	9.97E-06	0.00156
224707_at	ORF1-FL49	10	9.97E-06	0.00156
225412_at	FLJ14681	10	9.97E-06	0.00156
225611_at	MAST4	10	9.97E-06	0.00156
225855_at	EPB41L5	10	9.97E-06	0.00156
226051_at	SELM	10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
226264_at	SUSD1	10	9.97E-06	0.00156
227674_at	ZNF585A	10	9.97E-06	0.00156
227722_at	RPS23	10	9.97E-06	0.00156
227911_at	ARHGAP28	10	9.97E-06	0.00156
228490_at	ABHD2	10	9.97E-06	0.00156
229963_at	NGFRAP1L1	10	9.97E-06	0.00156
230206_at	DOCK5	10	9.97E-06	0.00156
230256_at	C1orf104	10	9.97E-06	0.00156
230292_at	RCBTB2	10	9.97E-06	0.00156
231862_at	CBX5	10	9.97E-06	0.00156
233303_at	UBE2D3	10	9.97E-06	0.00156
235171_at		10	9.97E-06	0.00156
235609_at	BRIP1	10	9.97E-06	0.00156
236524_at	TM2D1	10	9.97E-06	0.00156
237563_s_at	LOC440731	10	9.97E-06	0.00156
1557852_at	PHC2	10	9.97E-06	0.00156
1559332_at	CRIM1	10	9.97E-06	0.00156
1565651_at		10	9.97E-06	0.00156
203505_at	ABCA1	10	9.97E-06	0.00156
203725_at	GADD45A	10	9.97E-06	0.00156
203817_at	GUCY1B3	10	9.97E-06	0.00156
204352_at	TRAF5	10	9.97E-06	0.00156
204646_at	DPYD	10	9.97E-06	0.00156
205885_s_at	ITGA4	10	9.97E-06	0.00156
205891_at	ADORA2B	10	9.97E-06	0.00156
205966_at	TAF13	10	9.97E-06	0.00156
207574_s_at	GADD45B	10	9.97E-06	0.00156
209305_s_at	GADD45B	10	9.97E-06	0.00156
209561_at	THBS3	10	9.97E-06	0.00156
210298_x_at	FHL1	10	9.97E-06	0.00156
211317_s_at	CFLAR	10	9.97E-06	0.00156
212092_at	PEG10	10	9.97E-06	0.00156
213416_at	CERKL	10	9.97E-06	0.00156
213836_s_at	WIPI49	10	9.97E-06	0.00156
218543_s_at	PARP12	10	9.97E-06	0.00156
218613_at	PSD3	10	9.97E-06	0.00156
219073_s_at	OSBPL10	10	9.97E-06	0.00156
219279_at	DOCK10	10	9.97E-06	0.00156
220260_at	TBC1D19	10	9.97E-06	0.00156
220892_s_at	PSAT1	10	9.97E-06	0.00156
224847_at	CDK6	10	9.97E-06	0.00156
225373_at	C10orf54	10	9.97E-06	0.00156
225529_at	CENTB5	10	9.97E-06	0.00156
225931_s_at	C17orf27	10	9.97E-06	0.00156
226279_at	PRSS23	10	9.97E-06	0.00156
226419_s_at	SFRS1	10	9.97E-06	0.00156
226558_at		10	9.97E-06	0.00156
227059_at	GPC6	10	9.97E-06	0.00156
227566_at	HNT	10	9.97E-06	0.00156
228617_at	BIRC4BP	10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
228762_at	LFNG	10	9.97E-06	0.00156
229371_at	C8orf40	10	9.97E-06	0.00156
230703_at	C14orf32	10	9.97E-06	0.00156
230886_at		10	9.97E-06	0.00156
231057_at	MTMR2	10	9.97E-06	0.00156
231890_at		10	9.97E-06	0.00156
232412_at	FBXL20	10	9.97E-06	0.00156
235919_at		10	9.97E-06	0.00156
236229_at	LOC400721	10	9.97E-06	0.00156
236492_at	PPP2R2A	10	9.97E-06	0.00156
238449_at	DKFZp547E087	10	9.97E-06	0.00156
238824_at		10	9.97E-06	0.00156
242191_at	AE01	10	9.97E-06	0.00156
1555526_a_at	06.Sep	10	9.97E-06	0.00156
1558199_at	FN1	10	9.97E-06	0.00156
1568609_s_at	PDE4DIP	10	9.97E-06	0.00156
1569472_s_at	TTC3	10	9.97E-06	0.00156
205347_s_at	TMSL8	10	9.97E-06	0.00156
205381_at	LRRC17	10	9.97E-06	0.00156
205397_x_at	SMAD3	10	9.97E-06	0.00156
207357_s_at	GALNT10	10	9.97E-06	0.00156
209406_at	BAG2	10	9.97E-06	0.00156
209598_at	PNMA2	10	9.97E-06	0.00156
209760_at	KIAA0922	10	9.97E-06	0.00156
210102_at	LOH11CR2A	10	9.97E-06	0.00156
210869_s_at	MCAM	10	9.97E-06	0.00156
211376_s_at	C10orf86	10	9.97E-06	0.00156
214587_at	COL8A1	10	9.97E-06	0.00156
216048_s_at	RHOBTB3	10	9.97E-06	0.00156
218717_s_at	LEPREL1	10	9.97E-06	0.00156
219506_at	C1orf54	10	9.97E-06	0.00156
221679_s_at	ABHD6	10	9.97E-06	0.00156
225583_at	UXS1	10	9.97E-06	0.00156
227074_at	DR1	10	9.97E-06	0.00156
227290_at		10	9.97E-06	0.00156
227415_at	DGKH	10	9.97E-06	0.00156
227487_s_at	SERPINE2	10	9.97E-06	0.00156
227802_at	LOC441022	10	9.97E-06	0.00156
228158_at		10	9.97E-06	0.00156
228191_at		10	9.97E-06	0.00156
228242_at		10	9.97E-06	0.00156
228566_at	P15RS	10	9.97E-06	0.00156
229285_at	RNASEL	10	9.97E-06	0.00156
229838_at	NUCB2	10	9.97E-06	0.00156
230722_at	BNC2	10	9.97E-06	0.00156
232060_at		10	9.97E-06	0.00156
235088_at	LOC201725	10	9.97E-06	0.00156
235320_at	ARL6	10	9.97E-06	0.00156
239336_at	THBS1	10	9.97E-06	0.00156
1554097_a_at	LOC554202	10	9.97E-06	0.00156

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
201538_s_at	DUSP3	10	9.97E-06	0.00156
202037_s_at	SFRP1	10	9.97E-06	0.00156
202510_s_at	TNFAIP2	10	9.97E-06	0.00156
202627_s_at	SERPINE1	10	9.97E-06	0.00156
203385_at	DGKA	10	9.97E-06	0.00156
203504_s_at	ABCA1	10	9.97E-06	0.00156
203579_s_at	SLC7A6OS	10	9.97E-06	0.00156
203741_s_at	ADCY7	10	9.97E-06	0.00156
203951_at	CNN1	10	9.97E-06	0.00156
204925_at	CTNS	10	9.97E-06	0.00156
206396_at	SLC1A1	10	9.97E-06	0.00156
218000_s_at	PHLDA1	10	9.97E-06	0.00156
221577_x_at	GDF15	10	9.97E-06	0.00156
222719_s_at	PDGFC	10	9.97E-06	0.00156
223734_at	OSAP	10	9.97E-06	0.00156
223789_s_at	GTPBP2	10	9.97E-06	0.00156
224942_at	PAPPA	10	9.97E-06	0.00156
224964_s_at	GNG2	10	9.97E-06	0.00156
226278_at	DKFZp313A2432	10	9.97E-06	0.00156
226810_at	C6orf155	10	9.97E-06	0.00156
227396_at		10	9.97E-06	0.00156
229011_at		10	9.97E-06	0.00156
231067_s_at	AKAP12	10	9.97E-06	0.00156
232790_at		10	9.97E-06	0.00156
235311_at	FKBP14	10	9.97E-06	0.00156
201981_at	PAPPA	10	9.97E-06	0.00156
202036_s_at	SFRP1	10	9.97E-06	0.00156
206172_at	IL13RA2	10	9.97E-06	0.00156
206371_at	FOLR3	10	9.97E-06	0.00156
209460_at	ABAT	10	9.97E-06	0.00156
210517_s_at	AKAP12	10	9.97E-06	0.00156
217564_s_at	CPS1	10	9.97E-06	0.00156
219441_s_at	LRRK1	10	9.97E-06	0.00156
222150_s_at	LOC54103	10	9.97E-06	0.00156
224940_s_at	PAPPA	10	9.97E-06	0.00156
224941_at	PAPPA	10	9.97E-06	0.00156
227997_at	IL17RD	10	9.97E-06	0.00156
229441_at	FZD4	10	9.97E-06	0.00156
230831_at	FRMD5	10	9.97E-06	0.00156
235299_at		10	9.97E-06	0.00156
239629_at	CFLAR	10	9.97E-06	0.00156
210041_s_at	PGM3	10	9.97E-06	0.00156
213664_at	SLC1A1	10	9.97E-06	0.00156
213895_at	EMP1	10	9.97E-06	0.00156
224838_at	FOXP1	10	9.97E-06	0.00156
225362_at	LOC159090	10	9.97E-06	0.00156
228335_at	CLDN11	10	9.97E-06	0.00156
39582_at	CYLD	10	9.97E-06	0.00156
1552619_a_at	ANLN	9	8.42E-05	0.00298
1555486_a_at	FLJ14213	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
1555978_s_at	LOC440476	9	8.42E-05	0.00298
1559094_at	FBXO9	9	8.42E-05	0.00298
1559190_s_at	RDH13	9	8.42E-05	0.00298
200632_s_at	NDRG1	9	8.42E-05	0.00298
200678_x_at	GRN	9	8.42E-05	0.00298
201282_at	OGDH	9	8.42E-05	0.00298
202499_s_at	SLC2A3	9	8.42E-05	0.00298
203017_s_at	SSX2IP	9	8.42E-05	0.00298
203037_s_at	MTSS1	9	8.42E-05	0.00298
203213_at	CDC2	9	8.42E-05	0.00298
203276_at	LMNB1	9	8.42E-05	0.00298
203755_at	BUB1B	9	8.42E-05	0.00298
204507_s_at	PPP3R1	9	8.42E-05	0.00298
204510_at	CDC7	9	8.42E-05	0.00298
205171_at	PTPN4	9	8.42E-05	0.00298
205339_at	SIL	9	8.42E-05	0.00298
206011_at	CASP1	9	8.42E-05	0.00298
208070_s_at	REV3L	9	8.42E-05	0.00298
209267_s_at	SLC39A8	9	8.42E-05	0.00298
210559_s_at	CDC2	9	8.42E-05	0.00298
211991_s_at	HLA-DPA1	9	8.42E-05	0.00298
212141_at	MCM4	9	8.42E-05	0.00298
212188_at	KCTD12	9	8.42E-05	0.00298
212822_at	HEG1	9	8.42E-05	0.00298
213046_at	PABPN1	9	8.42E-05	0.00298
213906_at	MYBL1	9	8.42E-05	0.00298
214012_at	ARTS-1	9	8.42E-05	0.00298
215388_s_at	CFHL1P	9	8.42E-05	0.00298
217627_at	ZNF573	9	8.42E-05	0.00298
217966_s_at	C1orf24	9	8.42E-05	0.00298
218308_at	TACC3	9	8.42E-05	0.00298
218755_at	KIF20A	9	8.42E-05	0.00298
220987_s_at	NUAK2	9	8.42E-05	0.00298
222439_s_at	THRAP3	9	8.42E-05	0.00298
222608_s_at	ANLN	9	8.42E-05	0.00298
223307_at	CDCA3	9	8.42E-05	0.00298
223940_x_at	MALAT1	9	8.42E-05	0.00298
225875_s_at	NPAL3	9	8.42E-05	0.00298
227100_at	B3GTL	9	8.42E-05	0.00298
227517_s_at	GAS5	9	8.42E-05	0.00298
228049_x_at		9	8.42E-05	0.00298
228937_at	FLJ38725	9	8.42E-05	0.00298
229943_at	RFP2	9	8.42E-05	0.00298
230180_at	DDX17	9	8.42E-05	0.00298
230560_at	STXBP6	9	8.42E-05	0.00298
231042_s_at	CAMK2D	9	8.42E-05	0.00298
231817_at	USP53	9	8.42E-05	0.00298
231863_at	ING3	9	8.42E-05	0.00298
242059_at	ETNK1	9	8.42E-05	0.00298
242389_at	CROP	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
244447_at	KLF10	9	8.42E-05	0.00298
39248_at	AQP3	9	8.42E-05	0.00298
1553575_at		9	8.42E-05	0.00298
1554408_a_at	TK1	9	8.42E-05	0.00298
1554478_a_at	FLJ20718	9	8.42E-05	0.00298
1555852_at	TAP1	9	8.42E-05	0.00298
201666_at	TIMP1	9	8.42E-05	0.00298
202269_x_at	GBP1	9	8.42E-05	0.00298
202270_at	GBP1	9	8.42E-05	0.00298
202534_x_at	DHFR	9	8.42E-05	0.00298
203373_at	SOCS2	9	8.42E-05	0.00298
203586_s_at	ARF4L	9	8.42E-05	0.00298
203665_at	HMOX1	9	8.42E-05	0.00298
204780_s_at	FAS	9	8.42E-05	0.00298
205443_at	SNAPC1	9	8.42E-05	0.00298
205518_s_at	CMAH	9	8.42E-05	0.00298
205608_s_at	ANGPT1	9	8.42E-05	0.00298
206561_s_at	AKR1B10	9	8.42E-05	0.00298
208960_s_at	KLF6	9	8.42E-05	0.00298
209147_s_at	PPAP2A	9	8.42E-05	0.00298
209193_at	PIM1	9	8.42E-05	0.00298
209581_at	HRASLS3	9	8.42E-05	0.00298
209744_x_at	ITCH	9	8.42E-05	0.00298
209806_at	HIST1H2BK	9	8.42E-05	0.00298
209862_s_at	PIG8	9	8.42E-05	0.00298
212196_at	IL6ST	9	8.42E-05	0.00298
212730_at	DMN	9	8.42E-05	0.00298
212841_s_at	PPFIBP2	9	8.42E-05	0.00298
216028_at	DKFZP564C152	9	8.42E-05	0.00298
218485_s_at	SLC35C1	9	8.42E-05	0.00298
218662_s_at	HCAP-G	9	8.42E-05	0.00298
218795_at	ACP6	9	8.42E-05	0.00298
218935_at	EHD3	9	8.42E-05	0.00298
219000_s_at	DCC1	9	8.42E-05	0.00298
219117_s_at	FKBP11	9	8.42E-05	0.00298
219148_at	PBK	9	8.42E-05	0.00298
220940_at	KIAA1641	9	8.42E-05	0.00298
222640_at	DNMT3A	9	8.42E-05	0.00298
223222_at	SLC25A19	9	8.42E-05	0.00298
223319_at	GPHN	9	8.42E-05	0.00298
223464_at	OSBPL5	9	8.42E-05	0.00298
224836_at	TP53INP2	9	8.42E-05	0.00298
225129_at	CPNE2	9	8.42E-05	0.00298
225613_at	MAST4	9	8.42E-05	0.00298
226583_at	FLJ40142	9	8.42E-05	0.00298
226718_at	AMIGO1	9	8.42E-05	0.00298
228370_at	SNRPN	9	8.42E-05	0.00298
228667_at	AGPAT4	9	8.42E-05	0.00298
228829_at	ATF7	9	8.42E-05	0.00298
229189_s_at		9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
229193_at	CROP	9	8.42E-05	0.00298
229490_s_at	IQGAP3	9	8.42E-05	0.00298
229553_at	PGM2L1	9	8.42E-05	0.00298
229748_x_at		9	8.42E-05	0.00298
230083_at	USP53	9	8.42E-05	0.00298
230097_at	GART	9	8.42E-05	0.00298
230300_at		9	8.42E-05	0.00298
230305_at		9	8.42E-05	0.00298
230906_at	GALNT10	9	8.42E-05	0.00298
231577_s_at	GBP1	9	8.42E-05	0.00298
235113_at	PPIL5	9	8.42E-05	0.00298
235315_at	TSC22D1	9	8.42E-05	0.00298
238902_at	LOC115294	9	8.42E-05	0.00298
239208_s_at		9	8.42E-05	0.00298
240165_at		9	8.42E-05	0.00298
242907_at	GBP2	9	8.42E-05	0.00298
52940_at	SIGIRR	9	8.42E-05	0.00298
AFFX-r2-Bs-lys-5_at		9	8.42E-05	0.00298
1557081_at	RBM25	9	8.42E-05	0.00298
1558522_at	MPP6	9	8.42E-05	0.00298
200897_s_at	KIAA0992	9	8.42E-05	0.00298
201389_at	ITGA5	9	8.42E-05	0.00298
201810_s_at	SH3BP5	9	8.42E-05	0.00298
202284_s_at	CDKN1A	9	8.42E-05	0.00298
202409_at	LOC492304	9	8.42E-05	0.00298
202870_s_at	CDC20	9	8.42E-05	0.00298
202973_x_at	FAM13A1	9	8.42E-05	0.00298
203016_s_at	SSX2IP	9	8.42E-05	0.00298
203603_s_at	ZFHX1B	9	8.42E-05	0.00298
203680_at	PRKAR2B	9	8.42E-05	0.00298
203867_s_at	NLE1	9	8.42E-05	0.00298
204015_s_at	DUSP4	9	8.42E-05	0.00298
204066_s_at	CENTG2	9	8.42E-05	0.00298
204506_at	PPP3R1	9	8.42E-05	0.00298
204686_at	IRS1	9	8.42E-05	0.00298
204796_at	EML1	9	8.42E-05	0.00298
204900_x_at	SAP30	9	8.42E-05	0.00298
205259_at	NR3C2	9	8.42E-05	0.00298
205398_s_at	SMAD3	9	8.42E-05	0.00298
205802_at	TRPC1	9	8.42E-05	0.00298
207029_at	KITLG	9	8.42E-05	0.00298
207417_s_at	ZNF177	9	8.42E-05	0.00298
207426_s_at	TNFSF4	9	8.42E-05	0.00298
208534_s_at	RASA4	9	8.42E-05	0.00298
208663_s_at	TTC3	9	8.42E-05	0.00298
208710_s_at	AP3D1	9	8.42E-05	0.00298
209278_s_at	TFPI2	9	8.42E-05	0.00298
209714_s_at	CDKN3	9	8.42E-05	0.00298
209782_s_at	DBP	9	8.42E-05	0.00298
210785_s_at	C1orf38	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
211115_x_at	SIP1	9	8.42E-05	0.00298
211302_s_at	PDE4B	9	8.42E-05	0.00298
211343_s_at	COL13A1	9	8.42E-05	0.00298
211596_s_at	LRIG1	9	8.42E-05	0.00298
211602_s_at	TRPC1	9	8.42E-05	0.00298
211653_x_at	AKR1C2	9	8.42E-05	0.00298
212022_s_at	MKI67	9	8.42E-05	0.00298
212543_at	AIM1	9	8.42E-05	0.00298
212758_s_at	TCF8	9	8.42E-05	0.00298
212812_at		9	8.42E-05	0.00298
212913_at	MSH5	9	8.42E-05	0.00298
213006_at	CEBPD	9	8.42E-05	0.00298
213035_at	ANKRD28	9	8.42E-05	0.00298
213338_at	RIS1	9	8.42E-05	0.00298
213908_at	LOC339005	9	8.42E-05	0.00298
214255_at	ATP10A	9	8.42E-05	0.00298
214933_at	CACNA1A	9	8.42E-05	0.00298
217838_s_at	EVL	9	8.42E-05	0.00298
218475_at	HTF9C	9	8.42E-05	0.00298
218729_at	LXN	9	8.42E-05	0.00298
218747_s_at	TAPBPL	9	8.42E-05	0.00298
219918_s_at	ASPM	9	8.42E-05	0.00298
221561_at	SOAT1	9	8.42E-05	0.00298
222680_s_at	DTL	9	8.42E-05	0.00298
223194_s_at	C6orf85	9	8.42E-05	0.00298
223257_at	KIAA1333	9	8.42E-05	0.00298
224828_at	CPEB4	9	8.42E-05	0.00298
224840_at	FKBP5	9	8.42E-05	0.00298
224860_at	C9orf123	9	8.42E-05	0.00298
224901_at	SCD5	9	8.42E-05	0.00298
225177_at	RAB11FIP1	9	8.42E-05	0.00298
225214_at	PSMB7	9	8.42E-05	0.00298
225582_at	KIAA1754	9	8.42E-05	0.00298
226200_at	VARSL	9	8.42E-05	0.00298
226275_at	MXD1	9	8.42E-05	0.00298
226456_at	MGC24665	9	8.42E-05	0.00298
226905_at	MGC45871	9	8.42E-05	0.00298
226982_at	ELL2	9	8.42E-05	0.00298
227020_at	YPEL2	9	8.42E-05	0.00298
227062_at	TncRNA	9	8.42E-05	0.00298
227176_at	FLJ40126	9	8.42E-05	0.00298
227501_at	WSB1	9	8.42E-05	0.00298
227755_at		9	8.42E-05	0.00298
227793_at	LOC158257	9	8.42E-05	0.00298
227868_at	LOC154761	9	8.42E-05	0.00298
228287_at	ING5	9	8.42E-05	0.00298
228393_s_at	ZNF302	9	8.42E-05	0.00298
229800_at	DCAMKL1	9	8.42E-05	0.00298
230629_s_at	EP400	9	8.42E-05	0.00298
230733_at	MRCL3	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
230793_at	LRRC16	9	8.42E-05	0.00298
231035_s_at	OTUD1	9	8.42E-05	0.00298
231727_s_at	AD023	9	8.42E-05	0.00298
232021_at	LOC283464	9	8.42E-05	0.00298
232535_at		9	8.42E-05	0.00298
233401_at	NRCAM	9	8.42E-05	0.00298
235085_at	DKFZp761P0423	9	8.42E-05	0.00298
235289_at	EIF5A2	9	8.42E-05	0.00298
236075_s_at	ZNF232	9	8.42E-05	0.00298
236599_at	SERPINE2	9	8.42E-05	0.00298
236834_at	SCFD2	9	8.42E-05	0.00298
239944_at		9	8.42E-05	0.00298
240282_at	WDR1	9	8.42E-05	0.00298
242029_at	FNDC3B	9	8.42E-05	0.00298
242463_x_at	ZNF600	9	8.42E-05	0.00298
243747_at	ZNF599	9	8.42E-05	0.00298
244433_at		9	8.42E-05	0.00298
244623_at		9	8.42E-05	0.00298
37152_at	PPARD	9	8.42E-05	0.00298
AFFX-r2-Bs-lys-M_at		9	8.42E-05	0.00298
1554153_a_at	PHF21A	9	8.42E-05	0.00298
1555976_s_at	LOC440476	9	8.42E-05	0.00298
1556346_at	COTL1	9	8.42E-05	0.00298
1558290_a_at	LOC441378	9	8.42E-05	0.00298
1559232_a_at	SLC33A1	9	8.42E-05	0.00298
201171_at	ATP6V0E	9	8.42E-05	0.00298
201482_at	QSCN6	9	8.42E-05	0.00298
201506_at	TGFBI	9	8.42E-05	0.00298
201795_at	LBR	9	8.42E-05	0.00298
202241_at	TRIB1	9	8.42E-05	0.00298
202263_at	CYB5R1	9	8.42E-05	0.00298
202419_at	FVT1	9	8.42E-05	0.00298
202458_at	PRSS23	9	8.42E-05	0.00298
202478_at	TRIB2	9	8.42E-05	0.00298
203216_s_at	MYO6	9	8.42E-05	0.00298
203231_s_at	ATXN1	9	8.42E-05	0.00298
203414_at	MMD	9	8.42E-05	0.00298
203865_s_at	ADARB1	9	8.42E-05	0.00298
203881_s_at	DMD	9	8.42E-05	0.00298
204160_s_at	ENPP4	9	8.42E-05	0.00298
204396_s_at	GRK5	9	8.42E-05	0.00298
204418_x_at	GSTM2	9	8.42E-05	0.00298
204759_at	RCBTB2	9	8.42E-05	0.00298
204774_at	EVI2A	9	8.42E-05	0.00298
204971_at	CSTA	9	8.42E-05	0.00298
206307_s_at	FOXD1	9	8.42E-05	0.00298
207992_s_at	AMPD3	9	8.42E-05	0.00298
208025_s_at	HMGA2	9	8.42E-05	0.00298
208579_x_at	H2BFS	9	8.42E-05	0.00298
209447_at	SYNE1	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
211000_s_at	IL6ST	9	8.42E-05	0.00298
211571_s_at	CSPG2	9	8.42E-05	0.00298
212224_at	ALDH1A1	9	8.42E-05	0.00298
212442_s_at	LASS6	9	8.42E-05	0.00298
212828_at	SYNJ2	9	8.42E-05	0.00298
212989_at	TMEM23	9	8.42E-05	0.00298
213029_at	NFIB	9	8.42E-05	0.00298
213286_at	ZFR	9	8.42E-05	0.00298
213328_at	NEK1	9	8.42E-05	0.00298
214305_s_at	SF3B1	9	8.42E-05	0.00298
214375_at	PPFIBP1	9	8.42E-05	0.00298
216250_s_at	LPXN	9	8.42E-05	0.00298
218128_at	NFYB	9	8.42E-05	0.00298
218170_at	ISOC1	9	8.42E-05	0.00298
218522_s_at	BPY2IP1	9	8.42E-05	0.00298
218694_at	ARMCX1	9	8.42E-05	0.00298
218772_x_at	TMEM38B	9	8.42E-05	0.00298
219014_at	PLAC8	9	8.42E-05	0.00298
219104_at	RNF141	9	8.42E-05	0.00298
219131_at	UBIAD1	9	8.42E-05	0.00298
219132_at	PELI2	9	8.42E-05	0.00298
219308_s_at	AK5	9	8.42E-05	0.00298
221593_s_at	RPL31	9	8.42E-05	0.00298
221756_at	MGC17330	9	8.42E-05	0.00298
222587_s_at	GALNT7	9	8.42E-05	0.00298
223339_at	ATPIF1	9	8.42E-05	0.00298
223622_s_at	HYI	9	8.42E-05	0.00298
223672_at	SGIP1	9	8.42E-05	0.00298
223769_x_at	HYI	9	8.42E-05	0.00298
224944_at	TMPO	9	8.42E-05	0.00298
225330_at	IGF1R	9	8.42E-05	0.00298
225592_at	NRM	9	8.42E-05	0.00298
225816_at	PHF17	9	8.42E-05	0.00298
226021_at	RDH10	9	8.42E-05	0.00298
226037_s_at	TAF9L	9	8.42E-05	0.00298
226272_at		9	8.42E-05	0.00298
226384_at	PPAPDC1B	9	8.42E-05	0.00298
226550_at		9	8.42E-05	0.00298
227114_at	DKFZp547C195	9	8.42E-05	0.00298
227607_at	AMSH-LP	9	8.42E-05	0.00298
228333_at		9	8.42E-05	0.00298
228955_at		9	8.42E-05	0.00298
229359_at		9	8.42E-05	0.00298
229491_at	LOC133308	9	8.42E-05	0.00298
229538_s_at	IQGAP3	9	8.42E-05	0.00298
229692_at	API5	9	8.42E-05	0.00298
229850_at		9	8.42E-05	0.00298
229865_at	FNDC3B	9	8.42E-05	0.00298
230345_at		9	8.42E-05	0.00298
231929_at		9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
232164_s_at	EPPK1	9	8.42E-05	0.00298
232889_at	LOC389281	9	8.42E-05	0.00298
233543_s_at	FLJ13614	9	8.42E-05	0.00298
235046_at		9	8.42E-05	0.00298
235276_at	EPSTI1	9	8.42E-05	0.00298
235309_at		9	8.42E-05	0.00298
235653_s_at	THAP6	9	8.42E-05	0.00298
236449_at	CSTB	9	8.42E-05	0.00298
237459_at	PCTK2	9	8.42E-05	0.00298
238320_at	TncRNA	9	8.42E-05	0.00298
238462_at	STS-1	9	8.42E-05	0.00298
238673_at		9	8.42E-05	0.00298
239147_at	DKFZp313G1735	9	8.42E-05	0.00298
239331_at		9	8.42E-05	0.00298
241017_at	RPL31	9	8.42E-05	0.00298
241359_at		9	8.42E-05	0.00298
241838_at		9	8.42E-05	0.00298
242201_at	PMS2L5	9	8.42E-05	0.00298
242429_at	ZNF567	9	8.42E-05	0.00298
242750_at	MMAA	9	8.42E-05	0.00298
242856_at		9	8.42E-05	0.00298
244025_at		9	8.42E-05	0.00298
91703_at	EHBP1L1	9	8.42E-05	0.00298
1552643_at	ZNF626	9	8.42E-05	0.00298
1557685_at	DDEF1	9	8.42E-05	0.00298
1558143_a_at	BCL2L11	9	8.42E-05	0.00298
1566482_at	LOC400047	9	8.42E-05	0.00298
1566557_at	FLJ90757	9	8.42E-05	0.00298
1569022_a_at	PIK3C2A	9	8.42E-05	0.00298
200965_s_at	ABLIM1	9	8.42E-05	0.00298
202345_s_at	FABP5	9	8.42E-05	0.00298
202742_s_at	PRKACB	9	8.42E-05	0.00298
202761_s_at	SYNE2	9	8.42E-05	0.00298
202897_at	PTPNS1	9	8.42E-05	0.00298
203128_at	SPTLC2	9	8.42E-05	0.00298
203492_x_at	PIG8	9	8.42E-05	0.00298
203493_s_at	PIG8	9	8.42E-05	0.00298
203753_at	TCF4	9	8.42E-05	0.00298
203821_at	HBEGF	9	8.42E-05	0.00298
203921_at	CHST2	9	8.42E-05	0.00298
204088_at	P2RX4	9	8.42E-05	0.00298
204363_at	F3	9	8.42E-05	0.00298
204663_at	ME3	9	8.42E-05	0.00298
204720_s_at	DNAJC6	9	8.42E-05	0.00298
204862_s_at	NME3	9	8.42E-05	0.00298
205379_at	CBR3	9	8.42E-05	0.00298
205648_at	WNT2	9	8.42E-05	0.00298
206025_s_at	TNFAIP6	9	8.42E-05	0.00298
207304_at	ZNF45	9	8.42E-05	0.00298
209015_s_at	DNAJB6	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
209292_at	ID4	9	8.42E-05	0.00298
209829_at	C6orf32	9	8.42E-05	0.00298
211981_at	COL4A1	9	8.42E-05	0.00298
212070_at	GPR56	9	8.42E-05	0.00298
212136_at	ATP2B4	9	8.42E-05	0.00298
212607_at	AKT3	9	8.42E-05	0.00298
213069_at	HEG1	9	8.42E-05	0.00298
213435_at	SATB2	9	8.42E-05	0.00298
213449_at	POP1	9	8.42E-05	0.00298
213652_at	PCSK5	9	8.42E-05	0.00298
216035_x_at	TCF7L2	9	8.42E-05	0.00298
217452_s_at	B3GALT2	9	8.42E-05	0.00298
218537_at	HCFC1R1	9	8.42E-05	0.00298
219737_s_at	PCDH9	9	8.42E-05	0.00298
220651_s_at	MCM10	9	8.42E-05	0.00298
221127_s_at	RIG	9	8.42E-05	0.00298
221423_s_at	YIPF5	9	8.42E-05	0.00298
222044_at	C20orf67	9	8.42E-05	0.00298
222557_at	STMN3	9	8.42E-05	0.00298
222853_at	FLRT3	9	8.42E-05	0.00298
223315_at	NTN4	9	8.42E-05	0.00298
223614_at	DKFZp761D112	9	8.42E-05	0.00298
223805_at	OSBPL6	9	8.42E-05	0.00298
225724_at	PSMA3	9	8.42E-05	0.00298
225879_at	TSEN54	9	8.42E-05	0.00298
225957_at	LOC153222	9	8.42E-05	0.00298
226150_at	PPAPDC1B	9	8.42E-05	0.00298
226157_at	TFDP2	9	8.42E-05	0.00298
226210_s_at	MEG3	9	8.42E-05	0.00298
226670_s_at	C20orf119	9	8.42E-05	0.00298
226869_at		9	8.42E-05	0.00298
227051_at		9	8.42E-05	0.00298
227088_at	PDE5A	9	8.42E-05	0.00298
227249_at	NDE1	9	8.42E-05	0.00298
227605_at	SCYE1	9	8.42E-05	0.00298
227649_s_at	SRGAP2	9	8.42E-05	0.00298
228062_at	NAP1L5	9	8.42E-05	0.00298
228885_at	MAMDC2	9	8.42E-05	0.00298
228927_at	ZNF397	9	8.42E-05	0.00298
229549_at	OPN1SW	9	8.42E-05	0.00298
229693_at	LOC388335	9	8.42E-05	0.00298
229886_at	FLJ32363	9	8.42E-05	0.00298
230218_at		9	8.42E-05	0.00298
231234_at	CTSC	9	8.42E-05	0.00298
231513_at	KCNJ2	9	8.42E-05	0.00298
231779_at	IRAK2	9	8.42E-05	0.00298
231807_at	KIAA1217	9	8.42E-05	0.00298
231956_at	KIAA1618	9	8.42E-05	0.00298
232038_at	C6orf170	9	8.42E-05	0.00298
232238_at	ASPM	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
232589_at		9	8.42E-05	0.00298
232865_at	AFF4	9	8.42E-05	0.00298
233223_at	NEDD9	9	8.42E-05	0.00298
236009_at		9	8.42E-05	0.00298
236016_at	SMARCE1	9	8.42E-05	0.00298
238469_at	C6orf155	9	8.42E-05	0.00298
238860_at	C6orf130	9	8.42E-05	0.00298
239043_at	ZNF404	9	8.42E-05	0.00298
239231_at		9	8.42E-05	0.00298
239587_at		9	8.42E-05	0.00298
240145_at		9	8.42E-05	0.00298
241484_x_at		9	8.42E-05	0.00298
241820_at	RIF1	9	8.42E-05	0.00298
242052_at	BICD1	9	8.42E-05	0.00298
37892_at	COL11A1	9	8.42E-05	0.00298
1553962_s_at	RHOB	9	8.42E-05	0.00298
1554250_s_at	TRIM50B	9	8.42E-05	0.00298
1555870_at	C1orf188	9	8.42E-05	0.00298
1555920_at	CBX3	9	8.42E-05	0.00298
1556641_at	FLJ37228	9	8.42E-05	0.00298
1557905_s_at	CD44	9	8.42E-05	0.00298
1560296_at	DST	9	8.42E-05	0.00298
1567224_at	HMGA2	9	8.42E-05	0.00298
1568619_s_at	LOC162073	9	8.42E-05	0.00298
200766_at	CTSD	9	8.42E-05	0.00298
201262_s_at	BGN	9	8.42E-05	0.00298
201325_s_at	EMP1	9	8.42E-05	0.00298
201425_at	ALDH2	9	8.42E-05	0.00298
201539_s_at	FHL1	9	8.42E-05	0.00298
201581_at	TXNDC13	9	8.42E-05	0.00298
201818_at	FLJ12443	9	8.42E-05	0.00298
201869_s_at	TBL1X	9	8.42E-05	0.00298
202341_s_at	TRIM2	9	8.42E-05	0.00298
202382_s_at	GNPDA1	9	8.42E-05	0.00298
203153_at	IFIT1	9	8.42E-05	0.00298
203217_s_at	ST3GAL5	9	8.42E-05	0.00298
204084_s_at	CLN5	9	8.42E-05	0.00298
204137_at	TM7SF1	9	8.42E-05	0.00298
204140_at	TPST1	9	8.42E-05	0.00298
204493_at	BID	9	8.42E-05	0.00298
205407_at	RECK	9	8.42E-05	0.00298
205540_s_at	RRAGB	9	8.42E-05	0.00298
206777_s_at	CRYBB2	9	8.42E-05	0.00298
206907_at	TNFSF9	9	8.42E-05	0.00298
208230_s_at	NRG1	9	8.42E-05	0.00298
209035_at	MDK	9	8.42E-05	0.00298
209146_at	SC4MOL	9	8.42E-05	0.00298
209277_at	TFPI2	9	8.42E-05	0.00298
209373_at	MALL	9	8.42E-05	0.00298
209409_at	GRB10	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
209537_at	EXTL2	9	8.42E-05	0.00298
209758_s_at	MFAP5	9	8.42E-05	0.00298
210059_s_at	MAPK13	9	8.42E-05	0.00298
210732_s_at	LGALS8	9	8.42E-05	0.00298
210916_s_at	CD44	9	8.42E-05	0.00298
211208_s_at	CASK	9	8.42E-05	0.00298
211825_s_at	EWSR1	9	8.42E-05	0.00298
211980_at	COL4A1	9	8.42E-05	0.00298
212959_s_at	GNPTAB	9	8.42E-05	0.00298
213764_s_at	MFAP5	9	8.42E-05	0.00298
214250_at	NUMA1	9	8.42E-05	0.00298
214505_s_at	FHL1	9	8.42E-05	0.00298
214702_at	FN1	9	8.42E-05	0.00298
214723_x_at	KIAA1641	9	8.42E-05	0.00298
214927_at	ITGBL1	9	8.42E-05	0.00298
215775_at	THBS1	9	8.42E-05	0.00298
217785_s_at	YKT6	9	8.42E-05	0.00298
218051_s_at	FLJ12442	9	8.42E-05	0.00298
218934_s_at	HSPB7	9	8.42E-05	0.00298
219501_at	FLJ10094	9	8.42E-05	0.00298
220615_s_at	MLSTD1	9	8.42E-05	0.00298
221142_s_at	PECR	9	8.42E-05	0.00298
221882_s_at	TMEM8	9	8.42E-05	0.00298
223271_s_at	CTDSPL2	9	8.42E-05	0.00298
223287_s_at	FOXP1	9	8.42E-05	0.00298
223374_s_at	B3GALT3	9	8.42E-05	0.00298
223395_at	ABI3BP	9	8.42E-05	0.00298
224755_at	SMBP	9	8.42E-05	0.00298
225140_at	KLF3	9	8.42E-05	0.00298
225579_at	PQLC3	9	8.42E-05	0.00298
225640_at	LOC401504	9	8.42E-05	0.00298
225742_at	MDM4	9	8.42E-05	0.00298
225871_at	STEAP2	9	8.42E-05	0.00298
226534_at	KITLG	9	8.42E-05	0.00298
226908_at	LRIG3	9	8.42E-05	0.00298
226912_at	ZDHHC23	9	8.42E-05	0.00298
227408_s_at	SNX25	9	8.42E-05	0.00298
227525_at	GLCCI1	9	8.42E-05	0.00298
228314_at		9	8.42E-05	0.00298
228613_at	RAB11FIP3	9	8.42E-05	0.00298
229010_at	CBL	9	8.42E-05	0.00298
229512_at		9	8.42E-05	0.00298
230000_at	C17orf27	9	8.42E-05	0.00298
230165_at	SGOL2	9	8.42E-05	0.00298
230424_at	C5orf13	9	8.42E-05	0.00298
234863_x_at	FBXO5	9	8.42E-05	0.00298
235147_at	LOC440928	9	8.42E-05	0.00298
235287_at	CDK6	9	8.42E-05	0.00298
235746_s_at	PLA2R1	9	8.42E-05	0.00298
238722_x_at	NAPE-PLD	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
241893_at		9	8.42E-05	0.00298
243507_s_at	FLJ25067	9	8.42E-05	0.00298
243982_at	BTBD5	9	8.42E-05	0.00298
244669_at	LOC441164	9	8.42E-05	0.00298
1557080_s_at	ITGBL1	9	8.42E-05	0.00298
201185_at	HTRA1	9	8.42E-05	0.00298
201615_x_at	CALD1	9	8.42E-05	0.00298
201668_x_at	MARCKS	9	8.42E-05	0.00298
202948_at	IL1R1	9	8.42E-05	0.00298
203354_s_at	PSD3	9	8.42E-05	0.00298
203424_s_at	IGFBP5	9	8.42E-05	0.00298
203827_at	WIPI49	9	8.42E-05	0.00298
203879_at	PIK3CD	9	8.42E-05	0.00298
204807_at	TMEM5	9	8.42E-05	0.00298
204979_s_at	SH3BGR	9	8.42E-05	0.00298
205234_at	SLC16A4	9	8.42E-05	0.00298
205741_s_at	DTNA	9	8.42E-05	0.00298
205925_s_at	RAB3B	9	8.42E-05	0.00298
207264_at	KDELR3	9	8.42E-05	0.00298
208107_s_at	LOC81691	9	8.42E-05	0.00298
208191_x_at	PSG4	9	8.42E-05	0.00298
209216_at	WDR45	9	8.42E-05	0.00298
211340_s_at	MCAM	9	8.42E-05	0.00298
211713_x_at	KIAA0101	9	8.42E-05	0.00298
212386_at	TCF4	9	8.42E-05	0.00298
213206_at	GOSR2	9	8.42E-05	0.00298
213765_at	MFAP5	9	8.42E-05	0.00298
213959_s_at	KIAA1005	9	8.42E-05	0.00298
214151_s_at	PIGB	9	8.42E-05	0.00298
214581_x_at	TNFRSF21	9	8.42E-05	0.00298
214724_at	DIXDC1	9	8.42E-05	0.00298
217502_at	IFIT2	9	8.42E-05	0.00298
218284_at	SMAD3	9	8.42E-05	0.00298
218665_at	FZD4	9	8.42E-05	0.00298
218986_s_at	FLJ20035	9	8.42E-05	0.00298
219703_at	MNS1	9	8.42E-05	0.00298
221012_s_at	TRIM8	9	8.42E-05	0.00298
221139_s_at	CSAD	9	8.42E-05	0.00298
221489_s_at	SPRY4	9	8.42E-05	0.00298
222020_s_at	HNT	9	8.42E-05	0.00298
222621_at	DNAJC1	9	8.42E-05	0.00298
222802_at		9	8.42E-05	0.00298
224839_s_at	GPT2	9	8.42E-05	0.00298
224959_at	SLC26A2	9	8.42E-05	0.00298
225411_at	FLJ14681	9	8.42E-05	0.00298
225571_at	LIFR	9	8.42E-05	0.00298
226833_at	FLJ32499	9	8.42E-05	0.00298
227037_at	LOC201164	9	8.42E-05	0.00298
227070_at	GLT8D2	9	8.42E-05	0.00298
227347_x_at	HES4	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
227657_at	RNF150	9	8.42E-05	0.00298
227897_at	RAP2B	9	8.42E-05	0.00298
228946_at	PDZK6	9	8.42E-05	0.00298
228998_at	TNRC6B	9	8.42E-05	0.00298
229893_at	FRMD3	9	8.42E-05	0.00298
230425_at	EPHB1	9	8.42E-05	0.00298
230559_x_at	FGD4	9	8.42E-05	0.00298
230645_at	FRMD3	9	8.42E-05	0.00298
231904_at	U2AF1	9	8.42E-05	0.00298
235202_x_at	IKIP	9	8.42E-05	0.00298
235706_at	CPM	9	8.42E-05	0.00298
238841_at	PTPDC1	9	8.42E-05	0.00298
239761_at	GCNT1	9	8.42E-05	0.00298
242343_x_at	ZNF518	9	8.42E-05	0.00298
1552812_a_at	SENP1	9	8.42E-05	0.00298
1557275_a_at		9	8.42E-05	0.00298
200706_s_at	LITAF	9	8.42E-05	0.00298
201645_at	TNC	9	8.42E-05	0.00298
201801_s_at	SLC29A1	9	8.42E-05	0.00298
202741_at	PRKACB	9	8.42E-05	0.00298
205542_at	STEAP1	9	8.42E-05	0.00298
205920_at	SLC6A6	9	8.42E-05	0.00298
206942_s_at	PMCH	9	8.42E-05	0.00298
209505_at	NR2F1	9	8.42E-05	0.00298
211527_x_at	VEGF	9	8.42E-05	0.00298
212382_at	TCF4	9	8.42E-05	0.00298
212521_s_at	PDE8A	9	8.42E-05	0.00298
213295_at	CYLD	9	8.42E-05	0.00298
214023_x_at	TUBB-PARALOG	9	8.42E-05	0.00298
222668_at	KCTD15	9	8.42E-05	0.00298
222717_at	SDPR	9	8.42E-05	0.00298
224002_s_at	FKBP7	9	8.42E-05	0.00298
224719_s_at	GRCC10	9	8.42E-05	0.00298
224817_at	NEURL	9	8.42E-05	0.00298
224837_at	FOXP1	9	8.42E-05	0.00298
225653_at	TGFBRAP1	9	8.42E-05	0.00298
226719_at		9	8.42E-05	0.00298
226771_at	ATP8B2	9	8.42E-05	0.00298
229830_at	PDGFA	9	8.42E-05	0.00298
230748_at	SLC16A6	9	8.42E-05	0.00298
231403_at	TRIO	9	8.42E-05	0.00298
235508_at	PML	9	8.42E-05	0.00298
237169_at	TNC	9	8.42E-05	0.00298
242979_at		9	8.42E-05	0.00298
1554795_a_at	FBLIM1	9	8.42E-05	0.00298
207265_s_at	KDELR3	9	8.42E-05	0.00298
210026_s_at	CARD10	9	8.42E-05	0.00298
210138_at	RGS20	9	8.42E-05	0.00298
212816_s_at	CBS	9	8.42E-05	0.00298
213358_at	KIAA0802	9	8.42E-05	0.00298

Probe set Id	Symbol	# SEN DE	SEN p-value	SEN q-value
220161_s_at	EPB41L4B	9	8.42E-05	0.00298
224204_x_at	ARNTL2	9	8.42E-05	0.00298
230121_at		9	8.42E-05	0.00298
237466_s_at	HHIP	9	8.42E-05	0.00298
241387_at	PTK2	9	8.42E-05	0.00298

Analysis of UVB induced senescent gene expression profiles

MatInspector NF-κB TFBS

The table lists the NF-κB binding sites detected by Genomatix MatInspector in the promoter sequences of the manually selected genes of interest.
500 bp upstream and 100 bp downstream of the transcription start site were scanned for TFBS.

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	ACVR2A	625	+	1	0.961	ggcggtttTTCCc
NF-κB	ACVR2A	792	-	1	0.937	cctgggtcTTCCt
NF-κB	ADAMTS5	80	+	1	0.911	gcaggaacTTCCc
NF-κB	ADAMTS5	82	-	1	0.939	gtGGGAagttcct
NF-κB	APOD	127	-	1	0.898	agGGGAgatacca
NF-κB	APOD	212	+	0.75	0.865	aggGGGCtgccca
NF-κB	APOD	812	-	0.75	0.842	aaGGCAgtttcca
NF-κB	B2M	70	+	1	1	caGGGAattccca
NF-κB	B2M	71	-	1	1	ttGGGAattccct
NF-κB	B2M	418	+	1	0.893	acGGGAaagtccc
NF-κB	B2M	420	-	1	0.995	gaGGGActttccc
NF-κB	B2M	539	+	1	0.957	ggcgggcaTTCCt
NF-κB	BBC3	147	-	1	0.935	gacgggcgTTCCa
NF-κB	BBC3	317	+	1	0.933	caGGGAaaccccc
NF-κB	BBC3	319	-	1	0.997	ccgggggtTTCCc
NF-κB	BTG2	520	+	1	0.893	cgGGGAaagtccg
NF-κB	BTG2	522	-	1	0.98	cccggactTTCCc
NF-κB	CASC4	152	-	1	0.919	ctagggagTTCCt
NF-κB	CASC4	152	-	1	0.919	ctagggagTTCCt
NF-κB	CDKN1A	70	+	0.81	0.91	gtggggctTTTCt
NF-κB	CDKN1A	346	+	1	0.906	gaGGGAagtgccc
NF-κB	CDKN1A	348	-	1	0.921	gagggcacTTCCc
NF-κB	CDKN1A	70	+	0.81	0.91	gtggggctTTTCt
NF-κB	CDKN1A	346	+	1	0.906	gaGGGAagtgccc
NF-κB	CDKN1A	348	-	1	0.921	gagggcacTTCCc
NF-κB	CDKN2B	665	-	1	0.968	ccgggcttTTCCt
NF-κB	CDKN2B	889	+	0.75	0.865	aagGGCAtgccca
NF-κB	CDKN2B	890	-	0.75	0.865	ctgGGCAtgccct
NF-κB	CDKN2B	665	-	1	0.968	ccgggcttTTCCt
NF-κB	CDKN2B	889	+	0.75	0.865	aagGGCAtgccca
NF-κB	CDKN2B	890	-	0.75	0.865	ctgGGCAtgccct
NF-κB	CLCA2	338	-	1	0.87	tagtgattTTCCt
NF-κB	CLCA2	575	-	1	0.892	ctGGGAcctgcct
NF-κB	CLDN1	381	-	1	0.833	tgGGGAccacccg
NF-κB	CPM	767	+	1	0.832	gtcGGGAgaccct
NF-κB	CPM	803	+	0.75	0.865	ctgGTGAtgcccc
NF-κB	CPM	868	+	0.75	0.83	caTGGActtcccg
NF-κB	CPM	767	+	1	0.832	gtcGGGAgaccct
NF-κB	CPM	803	+	0.75	0.865	ctgGTGAtgcccc

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	CPM	868	+	0.75	0.83	caTGGActtcccg
NF-κB	CPM	767	+	1	0.832	gtcGGGAgaccct
NF-κB	CPM	803	+	0.75	0.865	ctgGTGAtgcccc
NF-κB	CPM	868	+	0.75	0.83	caTGGActtcccg
NF-κB	CSTA	280	-	1	0.873	actagaatTTCCa
NF-κB	CXCL12	705	+	1	0.932	gcgGGGAggccct
NF-κB	CXCL12	705	+	1	0.932	gcgGGGAggccct
NF-κB	CXCL12	705	+	1	0.932	gcgGGGAggccct
NF-κB	DKK2	48	+	1	0.968	tagggcttTTCCa
NF-κB	DRAM	148	-	1	0.838	gccGGGAggccca
NF-κB	DRAM	446	-	1	0.998	cgggggatTTCCg
NF-κB	DRAM	544	-	1	0.923	gcgGGGAcgcccg
NF-κB	DRAM	848	-	1	0.901	agGGGAggctccg
NF-κB	EGR2	157	-	1	0.997	ctGGGActttcca
NF-κB	ETV1	84	+	1	0.975	gtgggtctTTCCa
NF-κB	FABP3	625	+	1	0.937	cctggggcTTCCt
NF-κB	FKBP5	57	+	1	0.994	gtGGGAtttcccc
NF-κB	FKBP5	58	-	1	0.957	tgGGGAaatccca
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FN1	166	-	1	0.933	accgggagTTCCg
NF-κB	FN1	309	+	1	0.855	taaGGGAgtcccg
NF-κB	FN1	311	-	1	0.846	ctcGGGActccct
NF-κB	FN1	828	+	1	0.898	cgGGGAcagcccg
NF-κB	FN1	885	+	1	0.863	agGGGAccgtccc
NF-κB	FN1	887	-	1	0.857	atGGGAcggtccc
NF-κB	FN1	958	+	1	0.845	agaGGGAacccca
NF-κB	FN1	1105	+	1	0.871	aaGGGAtttttcc
NF-κB	FTH1	557	+	1	0.986	gtcggggtTTCCt
NF-κB	FUCA1	576	-	1	0.842	aaGGGActtaact
NF-κB	FZD8	302	-	1	0.923	cagggcccTTCCg
NF-κB	FZD8	397	+	1	0.918	aaGGGAgacccca
NF-κB	GALNTL2	145	+	1	0.85	aaGGGActtgcaa
NF-κB	GALNTL2	353	+	1	0.943	ttgggataTTCCt
NF-κB	GDF15	177	-	1	0.987	catgggatTTCCt
NF-κB	GDF15	502	-	1	0.897	ctGGGActgacca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	821	+	0.75	0.865	aagTGGAtgccca
NF-κB	HDAC9	400	+	0.75	0.865	aagTGGAtgccca
NF-κB	ID4	25	-	0.75	0.859	tcgCGGAtaccct
NF-κB	ID4	337	-	1	0.832	gcgGGGAgtccgg

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	IDS	85	+	0.75	0.865	ccgGGTAtgccca
NF-κB	IDS	86	-	0.75	0.859	gtgGGCAtacccg
NF-κB	IDS	229	+	1	0.855	aaaGGGAgtccct
NF-κB	IDS	231	-	1	0.846	aaaGGGActccct
NF-κB	IDS	253	+	1	0.895	aaGGGActttctc
NF-κB	IDS	85	+	0.75	0.865	ccgGGTAtgccca
NF-κB	IDS	86	-	0.75	0.859	gtgGGCAtacccg
NF-κB	IDS	229	+	1	0.855	aaaGGGAgtccct
NF-κB	IDS	231	-	1	0.846	aaaGGGActccct
NF-κB	IDS	253	+	1	0.895	aaGGGActttctc
NF-κB	IGFBP3	19	-	1	0.838	ttGGGAcatttct
NF-κB	IGFBP3	19	-	1	0.838	ttGGGAcatttct
NF-κB	IGFBP5	329	+	1	0.944	aagGGGAgccccc
NF-κB	IGFBP5	330	-	0.75	0.877	aggGGGCtcccct
NF-κB	IL1B	332	+	1	0.891	gtGGGAaaatcca
NF-κB	IL1B	334	-	1	0.974	actggattTTCCc
NF-κB	IL1B	1132	-	0.75	0.833	gaGTGActtcccc
NF-κB	KAL1	52	-	1	0.873	attcgaatTTCCt
NF-κB	KAL1	77	-	1	0.943	ccaggcttTTCCc
NF-κB	KAL1	133	-	1	0.899	tgGGGAttgaccc
NF-κB	KAL1	151	+	1	0.92	gtGGGAcgcccct
NF-κB	KAL1	152	-	0.75	0.832	cagGGGCgtccca
NF-κB	LAMA2	304	+	1	0.883	taGGGAttttaca
NF-κB	LMCD1	460	+	1	0.855	agaGGGAgtcccg
NF-κB	LMCD1	462	-	1	0.846	agcGGGActccct
NF-κB	LRDD	187	-	1	0.914	aggggctgTTCCc
NF-κB	LRDD	187	-	1	0.914	aggggctgTTCCc
NF-κB	LRDD	187	-	1	0.914	aggggctgTTCCc
NF-κB	MUC1	134	-	1	0.833	gtcGGGAagccca
NF-κB	MUC1	134	-	1	0.833	gtcGGGAagccca
NF-κB	MUC1	134	-	1	0.833	gtcGGGAagccca
NF-κB	NAT13	412	+	1	0.91	gatggaggTTCCg
NF-κB	NME5	625	-	1	0.992	cgGGGActttccg
NF-κB	NQO1	233	+	0.75	0.865	aagGGCAtgccca
NF-κB	NQO1	234	-	0.75	0.865	gtgGGCAtgccct
NF-κB	NQO1	233	+	0.75	0.865	aagGGCAtgccca
NF-κB	NQO1	234	-	0.75	0.865	gtgGGCAtgccct
NF-κB	NQO1	233	+	0.75	0.865	aagGGCAtgccca
NF-κB	NQO1	234	-	0.75	0.865	gtgGGCAtgccct
NF-κB	PCDH7	17	+	1	0.921	cagggcacTTCCg
NF-κB	PCDH7	155	-	1	0.872	gtctggatTTCCg
NF-κB	PCDH7	195	+	1	0.923	ttgGGGAcgcccg
NF-κB	PCDH7	17	+	1	0.921	cagggcacTTCCg
NF-κB	PCDH7	155	-	1	0.872	gtctggatTTCCg
NF-κB	PCDH7	195	+	1	0.923	ttgGGGAcgcccg
NF-κB	PIK3IP1	553	+	1	0.915	ctGGGAagctccc
NF-κB	PIK3IP1	555	-	1	0.927	ctgggagcTTCCc
NF-κB	PLK1	61	-	1	0.914	tcaggtgaTTCCt
NF-κB	PPP1R13L	1141	-	1	0.948	ctgggggcTTCCt
NF-κB	PPP1R14A	94	-	1	0.834	acGGGAtttcgcc

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	PPP1R14A	492	+	1	0.909	ccGGGAgcccccg
NF-κB	PPP1R14A	828	+	0.75	0.832	gggGGCAgtcccg
NF-κB	PPP1R14A	829	-	1	0.941	ccGGGActgcccc
NF-κB	PPP1R3C	333	+	0.75	0.832	ccgCGGAgtccca
NF-κB	QPRT	573	-	1	0.912	gccggcccTTCCc
NF-κB	RDH10	213	+	0.75	0.859	gagGGGCtacccg
NF-κB	RDH10	329	+	1	0.92	gcgggctcTTCCc
NF-κB	RHOJ	964	-	1	0.922	cgaggggcTTCCt
NF-κB	RUNX1	395	-	1	0.946	acaggcctTTCCg
NF-κB	RUNX1	395	-	1	0.946	acaggcctTTCCg
NF-κB	SEPP1	45	-	1	0.847	atGGGAtttcaca
NF-κB	SERPINB2	102	-	1	0.832	ctaGGGAgaccca
NF-κB	SERPINE1	371	-	1	0.833	ctGGGActtgctg
NF-κB	SERPINE1	672	-	1	0.855	gaGGGAgtttgct
NF-κB	SERPING1	483	-	1	0.897	atGGGActgaccg
NF-κB	SERPING1	622	+	1	0.833	ggGGGActctcta
NF-κB	SERPING1	234	-	1	0.897	atGGGActgaccg
NF-κB	SERPING1	373	+	1	0.833	ggGGGActctcta
NF-κB	SERPING1	741	-	1	0.912	gctggcccTTCCt
NF-κB	SERPING1	835	+	1	0.829	gaGGGAattcgct
NF-κB	SERPING1	483	-	1	0.897	atGGGActgaccg
NF-κB	SERPING1	622	+	1	0.833	ggGGGActctcta
NF-κB	SERPING1	234	-	1	0.897	atGGGActgaccg
NF-κB	SERPING1	373	+	1	0.833	ggGGGActctcta
NF-κB	SERPING1	741	-	1	0.912	gctggcccTTCCt
NF-κB	SERPING1	835	+	1	0.829	gaGGGAattcgct
NF-κB	SFRP1	1018	-	1	0.865	gcgGGGAtgtcca
NF-κB	SGIP1	653	+	1	0.897	aaGGGAttggccg
NF-κB	SOD3	136	+	1	0.83	ctGGGActctgca
NF-κB	SULF2	8	+	1	0.823	cgGGGAgttctcg
NF-κB	SULF2	462	-	1	0.923	gtGGGAttcgccc
NF-κB	SULF2	8	+	1	0.823	cgGGGAgttctcg
NF-κB	SULF2	462	-	1	0.923	gtGGGAttcgccc
NF-κB	TM4SF20	118	-	1	0.948	aaaggtatTTCCt
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg

TF	Symbol	Position	Strand	Core sim.	Matrix sim.	Sequence
NF-κB	UBE2C	117	+	1	0.895	ctGGGAttttcag
NF-κB	UBE2C	911	+	1	0.846	tccGGGActccca
NF-κB	UBE2C	913	-	1	0.855	ggtGGGAgtcccg

Appendix E

Publications

Laschober G, Ruli D, **Hofer E**, Muck C, Carmona-Gutierrez D, Ring J, Hutter E, Ruckenstuhl C, Micutkova L, Brunauer R, Jamnig A, Trimmel D, Herndler-Brandstetter D, Sampson N, Breitenbach M, Fröhlich KU, Grubeck-Loebenstein B, Berger P, Wieser M, Grillari-Voglauer R, Thallinger GG, Grillari J, Trajanoski Z, Madeo F, Lepperdinger G, Jansen-Dürr P. **Identification of evolutionarily conserved genetic regulators of cellular aging**. Aging Cell 2010, 9:1084-1097. PMID: 20883526

Hofer E, Laschober GT, Hackl M, Thallinger GG, Lepperdinger G, Grillari J, Jansen-Dürr P, Trajanoski Z. **GiSAO.db: a database for ageing research**. BMC Genomics 2011, 12:262. PMID: 21609420

Aging Cell (2010) **9,** pp1084–1097 Doi: 10.1111/j.1474-9726.2010.00637.x

Identification of evolutionarily conserved genetic regulators of cellular aging

Gerhard T. Laschober,[1,*] Doris Ruli,[2,*] Edith Hofer,[3] Christoph Muck,[1] Didac Carmona-Gutierrez,[2] Julia Ring,[2] Eveline Hutter,[1] Christoph Ruckenstuhl,[2] Lucia Micutkova,[1] Regina Brunauer,[1] Angelika Jamnig,[1] Daniela Trimmel,[1] Dietmar Herndler-Brandstetter,[1] Stefan Brunner,[1] Christoph Zenzmaier,[1] Natalie Sampson,[1] Michael Breitenbach,[4] Kai-Uwe Fröhlich,[2] Beatrix Grubeck-Loebenstein,[1] Peter Berger,[1] Matthias Wieser,[5] Regina Grillari-Voglauer,[5] Gerhard G. Thallinger,[3] Johannes Grillari,[5] Zlatko Trajanoski,[3,6] Frank Madeo,[2] Günter Lepperdinger[1] and Pidder Jansen-Dürr[1]

[1]Institute for Biomedical Aging Research, Austrian Academy of Sciences, Rennweg 10, A-6020 Innsbruck, Austria
[2]Institute for Molecular Biosciences, University of Graz, Humboldtstrasse 50, 8010 Graz, Austria
[3]Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria
[4]Institute for Cell Biology, Salzburg University, Salzburg, Austria
[5]Aging and Immortalization Research, Department of Biotechnology, University of Natural Resources and Applied Life Sciences, Vienna, Austria
[6]Biocenter, Section for Bioinformatics, Innsbruck Medical University, Innsbruck, Austria

Summary

To identify new genetic regulators of cellular aging and senescence, we performed genome-wide comparative RNA profiling with selected human cellular model systems, reflecting replicative senescence, stress-induced premature senescence, and distinct other forms of cellular aging. Gene expression profiles were measured, analyzed, and entered into a newly generated database referred to as the GiSAO database. Bioinformatic analysis revealed a set of new candidate genes, conserved across the majority of the cellular aging models, which were so far not associated with cellular aging, and highlighted several new pathways that potentially play a role in cellular aging. Several candidate genes obtained through this analysis have been confirmed by functional experiments, thereby validating the experimental approach. The effect of genetic deletion on chronological lifespan in yeast was assessed for 93 genes where (i) functional homologues were found in the yeast genome and (ii) the deletion strain was viable. We identified several genes whose deletion led to significant changes of chronological lifespan in yeast, featuring both lifespan shortening and lifespan extension. In conclusion, an unbiased screen across species uncovered several so far unrecognized molecular pathways for cellular aging that are conserved in evolution.

Key words: aging; evolution; replicative lifespan; replicative senescence; senescence; yeast.

Introduction

A major goal in current aging research is to understand genetic regulation of aging in humans. However, many recent insights into genetic determinants of aging have been gained from studies with short-lived eukaryotic model systems of aging [for recent review, see (Partridge, 2009)], which, unlike human beings, are amenable to molecular genetic analysis in the laboratory. Extrapolation of results obtained in lower model organisms identified cellular pathways such as the insulin/IGF pathway (Holzenberger, 2004; Katic & Kahn, 2005), which are conserved in evolution and appear to modulate aging in a variety of organisms including mammals (Bartke, 2005). Although it is very likely that these very same mechanisms also modulate aging in humans, this has not yet been formally proven. Moreover, extrapolating findings in lower model organisms to more complex organisms, such as humans, bears the inherent risk that some pathways, which are less important or absent there, may be overlooked.

To circumvent this potential problem, we applied an alternative strategy, namely to employ human model systems of cellular aging and senescence as the primary screening system. There is now increasing evidence that the appearance of cellular senescence contributes to the aging of tissues, such as the skin (Dimri *et al.*, 1995; Ressler *et al.*, 2006), the vascular system (Vasile *et al.*, 2001; Minamino *et al.*, 2002; reviewed by Erusalimsky, 2009), and the kidney (Melk & Halloran, 2001). We performed genome-wide RNA profiling on several experimental models for cellular senescence, including stress-induced premature senescence (SIPS), replicative senescence, and other distinct forms of cellular aging, as proposed recently (Lepperdinger *et al.*, 2008; Hackl *et al.*, 2010) and collected the data in a novel database referred to as the GiSAO (genes involved in senescence, apoptosis, and oxidative stress) database. Comparison of the differential gene expression patterns of the models should in turn unveil distinct novel genetic regulators of cellular aging. Potentially,

Correspondence
Dr. Pidder Jansen-Dürr, Institute for Biomedical Aging Research, Rennweg 10, A-6020 Innsbruck, Austria. Tel.: +43 512 583919 44; fax: +43 512 583919 8; e-mail: p.jansen-duerr@oeaw.ac.at
*These authors made equal contributions to the manuscript.

Acceptance for publication *14 September 2010*

the information obtained from these experiments, when filtered accordingly, should provide molecular information about highly conserved gene expression patterns related to cellular aging. Because of the genome-wide approach, the data collected from cell types as diverse as those employed herein should also uncover recurrent cellular alterations indicative for hitherto unknown biological aging mechanisms. As a further step in evaluation of the functional role of the individually selected genes, those that were also found to be highly conserved during phylogeny were taken to functional tests in yeast deletion strains thereby monitoring the respective lifespan of the mutant organisms.

Results

Comparative RNA profiling of aging and senescence in human cellular models

Accumulation of damage is considered a key factor driving cellular aging, and oxidative stress is a common denominator of stress that induces cellular senescence *in vitro* in a variety of cell types (Dumont *et al.*, 2000). To identify key transcriptional events preceding senescence induced by oxidative stress, a comprehensive set of human cellular models were subjected to oxidative stress at doses known to elicit a senescence response. The corresponding model systems are referred to as experimental group 1 (EG1) (Table 1). For comparison, other model systems of cellular aging were also analyzed collectively referred to as experimental group 2 (EG2) (Table 1). On the one hand, we used models of *in vitro* cellular senescence not triggered by oxidative stress, namely replicative senescence of human umbilical vein endothelial cells (HUVEC) and renal proximal tubular epithelial cells (RPTEC). Similarly, we applied primary foreskin fibroblasts (PFF) in which premature senescence was induced by mitochondrial dysfunction (Passos *et al.*, 2007), which is largely independent of oxidative stress (Stockl *et al.*, 2007). On the other hand, three alternative models of cellular aging (included in experimental group 2, see Table 1) were also analyzed: first, we compared $CD28^-$ and $CD28^+$ $CD8^+$ T lymphocytes, based on the assumption that the increased percentage of $CD28^+$ T cells result from high level cell proliferation and hence represent a form of *in vivo* senescence (Rufer *et al.*, 2003; Lazuardi *et al.*, 2009). Furthermore, we also analyzed differences in the gene expression profiles of mesenchymal stem cells (MSC) obtained from young and old human donors, another system that permits the impact of human donor age on cellular phenotype to be assessed (Fehrer *et al.*, 2007). Finally, we also analyzed gene expression profiles of prostate stromal cells (PrSC), which were driven to transdifferentiate into premature senescent myofibroblasts by TGFβ1, a process considered a hallmark of the aging human prostate, leading to benign prostatic hyperplasia (Untergasser *et al.*, 2005). RNA was prepared from all samples and analyzed on Affymetrix arrays for whole-genome expression profiling. For validation purposes, the expression levels of 22 genes, which were found to be differentially expressed in a high number of individual experiments, were determined by quantitative reverse transcription–PCR (qPCR). Good concordance between data of the microarray and qPCR analyses was obtained. Linear regression yielded an *r*-value of 0.68 and 362 of 380 individual experimental data pairs sitting within the 95% prediction interval (Fig. S1).

Table 1 Overview microarray experiments: experimental model systems: HUVEC, human umbilical vein epithelial cells; PFF, primary foreskin fibroblasts; PrSC, primary prostatic stromal fibroblasts; RPTEC, renal proximal tubular epithelial cells; CD8, $CD8^+$ T lymphocytes; MSC, mesenchymal stem cells; reagents: t-BHP (*tert*-butylhydroperoxide), FCCP (carbonyl cyanide p-(trifluoromethoxy)phenylhydrazone), AMP (adenosine monophosphate); number of array hybridizations/samples; number of array experiments for senescence and oxidative stress; EMBL-EBI Array Express accession number

	Controls	Experimental group 1 (oxidative stress)		Experimental group 2 (cellular aging)		
Cell type	No. of arrays	Treatment	No. of arrays	Treatment	No. of arrays	Array Express accession number
HUVEC	6	t-BHP	2	RS	2	E-MEXP-2283
PFF	8	ND	ND	FCCP	2	E-MEXP-2285
				AMP	2	
				Oligomycin	2	
RPTEC	4	ND	ND	RS	1	E-MEXP-2683
				High/low ROS	1	
PrSC	10	t-BHP	2	TGF-β (trans-differentiation)	2	E-MEXP-2167
		20% vs. 3% O_2	4			
CD8	15	t-BHP	7	$CD28^-/CD28^+$	4	E-MEXP-2345
MSC	4	20% vs. 3% O_2	2	Young vs. old donor	2	E-MEXP-1506
Total	47		17		18	

RS, replicative senescence; ND, not determined.

Identification of transcriptional signatures of oxidative stress and cellular aging

For computational comparison, normalized whole-genome expression profiles from 47 independent experimental samples were established (Table 1), and hierarchical clustering of the gene expression profiles derived from all experimental human cell models was performed (Fig. S2). Differential expression values were calculated and compiled for further analysis by assembling 35 pairs of individual experiments with regard to cellular aging or oxidative stress (Fig. 1; see also Table S1). Considerable resemblance with respect to sets of differentially expressed genes was observed between HUVEC and RPTEC that have undergone replicative senescence, *in vivo* aged CD8 T cells and MSC, as well as in a compilation of these experimental model systems (Fig. S3).

Several methods have been established to identify genes that are differentially expressed in a homologous set of biological replicates. However, complex datasets such as those described herein, which comprise transcriptional profiles from different cell types derived from differently aged, nonrelated human indi-

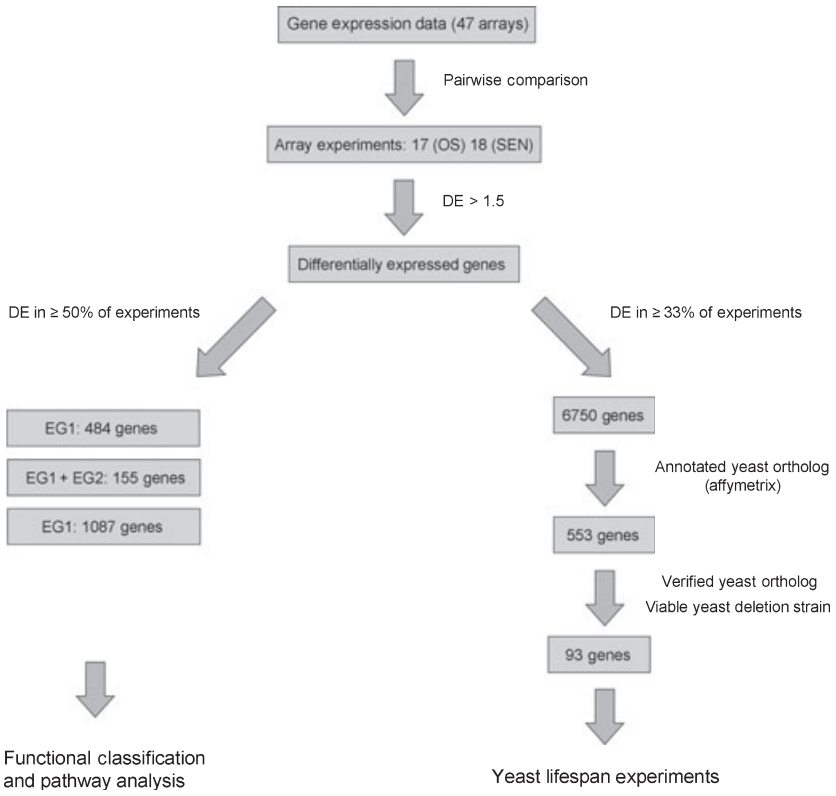

Fig. 1 Workflow for candidate gene identification and functional classification. First, gene expression data from 47 microarrays were processed to identify differentially expressed (DE) genes in experimental group 1 (EG1) and experimental group 2 (EG2). The selection of genes for functional classification and pathway analysis, starting from a total of 1566 genes, is shown (left arm). In the right arm of the diagram, the strategy to identify yeast genes for testing in lifespan analysis is depicted.

viduals, often show highly variable transcript levels with regard to the expression of individual genes (Lu *et al.*, 2004; Rodwell *et al.*, 2004; Tan *et al.*, 2005; Lener *et al.*, 2006). This can be either because of variant genetic background or the redundant use of closely related homologous genes. Therefore, we employed a value counting method.

The number a particular gene was found under- or overexpressed was counted, and the list of all genes was sorted according to occurrence. Using this methodology, we identified 586 probe sets representing 484 genes across EG1 with differential expression (DE) higher than ±1.5-fold in more than 50% of individual pair-wise comparisons, and 1423 probe sets representing 1087 genes across EG2, again with the same criteria of DE > ±1.5 in > 50% of experiments, whereas 180 probe sets representing 155 genes were present in both data sets. With respect to a cumulative binomial distribution (CBD), we additionally calculated *P*-values together with the corresponding *q*-values (Table S2) yielding a probability of 0.12 for any probe set to be differentially expressed with senescence and a probability of 0.11 for a probe set to be differentially expressed with oxidative stress. The *q*-values were ≤ 0.0030 in both EG1 and EG2, and *P*-values were ≤ 0.000028 in EG1 and ≤ 0.000085 in EG2 (Table S2). When evaluating both experimental groups along these lines, we revealed a set of 335 candidate genes with DE > ±1.5 in at least 18 pair-wise comparisons (Table S3).

The highest overall score (with most occurrences of DE) was observed for *PAF* (PCNA associated factor) (Table 2A) (Yu *et al.* 2001), a gene recently reported to be involved in DNA repair in HeLa cells (Turchi *et al.*, 2009). Another top scoring gene was SKP2 (Table 2A), encoding a regulatory subunit of an ubiquitin ligase complex, which was recently identified as playing a key role in setting the threshold for stress-induced cellular senescence (Lin *et al.*, 2010). Several additional genes were identified with high scores, such as SLC1A4, a member of the solute carrier family 1 (Table 2A), which so far have not been linked to the senescent phenotype. Of note, four genes involved in the regulation of transcription (BHLHB2, TCF8, RUNX1, and TCEA3) were included in the top 15 genes common to both experimental groups (Table 2A).

When both experimental groups were treated separately, the highest scoring genes (with most occurrences of DE) for experimental group 1 (oxidative stress) were genes involved in regulating cell proliferation, such as RRM2, cyclin B2, EGR1, and IGFBP-3 (Table 2B), consistent with the induction of rapid growth arrest by oxidative stress (Chen & Ames, 1994). In addition, several genes were identified for which a role in stress-induced senescence was not known before, such as the translation initiation factors EIF4A and EIF5A (Table 2B). The highest scoring genes (with most occurrences of DE) for experimental group 2 (cellular aging) are shown in Table 2C. Surprisingly, the

Table 2 (A) Genes that were found differentially expressed with highest occurrence in both experimental groups, (B) group 1 (oxidative stress), (C) group 2 (cellular aging)

Symbol	Description	Unigene	GenBank	EG1 up	EG1 down	EG2 up	EG2 down	EG1 + EG2DE
(A)								
KIAA0101	KIAA0101/PAF	Hs.81892	NM_014736	7	7	6	7	27
APOBEC3G	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	Hs.660143	NM_021822	0	10	8	8	26
MEST	Mesoderm-specific transcript homolog (mouse)	Hs.270978	NM_002402	4	5	8	8	25
RRM2	Ribonucleotide reductase M2 polypeptide	Hs.226390	BE966236	7	5	4	9	25
SYDE2	Synapse defective 1, Rho GTPase, homolog 2 (*C. elegans*)	Hs.718601	N90719	7	5	5	8	25
TXNIP	Thioredoxin interacting protein	Hs.715525	AA812232	2	9	7	7	25
BHLHB2	Basic helix-loop-helix domain containing, class B, 2	Hs.719093	NM_003670	5	6	10	3	24
BICD1	Bicaudal D homolog 1 (*Drosophila*)	Hs.505202	BC010091	10	3	7	4	24
CYP51A1	Cytochrome P450, family 51, subfamily A, polypeptide 1	Hs.417077	U40053	10	1	10	3	24
HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	Hs.397729	NM_002130	9	3	9	3	24
SKP2	S-phase kinase-associated protein 2 (p45)	Hs.23348	BC001441	5	9	3	7	24
SLC1A4	Solute carrier family 1 (neutral amino acid transporter), member 4	Hs.654352	W72527	6	1	11	6	24
TCF8	Transcription factor 8 (represses interleukin 2 expression)	Hs.282113	NM_030751	6	6	6	6	24
C20orf129	Chromosome 20 open reading frame 129	Hs.472716	BC001068	11	1	7	4	23
CD302	CD302 antigen	Hs.130014	NM_014880	7	4	6	6	23
IFI44L	Interferon-induced protein 44-like	Hs.389724	NM_006820	3	7	8	5	23
PDGFD	Platelet derived growth factor D	Hs.352298	NM_025208	5	4	6	8	23
RUNX1	Runt-related transcription factor 1 (acute myeloid leukemia 1)	Hs.675708	BU789637	2	9	9	3	23
TCEA3	Transcription elongation factor A (SII), 3	Hs.446354	AI675780	2	8	4	9	23
TSPAN2	Tetraspanin 2	Hs.310458	AI743596	4	6	7	6	23

Symbol	Description	Unigene	GenBank	EG1 up	EG1 down	EG1 DE
(B)						
KIAA0101	KIAA0101/PAF	Hs.81892	NM_014736	7	7	14
RRM2	Ribonucleotide reductase M2 polypeptide	Hs.226390	BC001886	8	6	14
SKP2	S-phase kinase-associated protein 2 (p45)	Hs.23348	BC001441	5	9	14
TRPV2	Transient receptor potential cation channel, subfamily V, member 2	Hs.279746	NM_015930	4	10	14
BICD1	Bicaudal D homolog 1 (Drosophila)	Hs.505202	BC010091	10	3	13
EIF4A2	Eukaryotic translation initiation factor 4A, isoform 2	Hs.518475	AI332397	10	3	13
NQO1	NAD(P)H dehydrogenase, quinone 1	Hs.406515	NM_000903	9	4	13
C20orf129	Chromosome 20 open reading frame 129	Hs.472716	BC001068	11	1	12
CASP1	Caspase 1, apoptosis-related cysteine peptidase (interleukin 1, beta, convertase)	Hs.2490	U13699	2	10	12
CCNB2	Cyclin B2	Hs.194698	NM_004701	5	7	12
CIRBP	Cold-inducible RNA-binding protein	Hs.634522	AL565767	8	4	12
CKAP2	Cytoskeleton-associated protein 2	Hs.444028	NM_018204	11	1	12
CYBA	Cytochrome *b*-245, alpha polypeptide	Hs.513803	NM_000101	3	9	12
DAAM1	Disheveled-associated activator of morphogenesis 1	Hs.19156	AA890373	8	4	12
EGR1	Early growth response 1	Hs.326035	NM_001964	6	6	12
EIF5A	Eukaryotic translation initiation factor 5A	Hs.534314	NM_001970	11	1	12
ERRFI1	ERBB receptor feedback inhibitor 1	Hs.658370	AW612461	7	5	12
FBXO9	F-box protein 9	Hs.216653	AK095307	8	4	12
HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1 (soluble)	Hs.397729	NM_002130	9	3	12
IGFBP3	Insulin-like growth factor binding protein 3	Hs.450230	BF340228	4	8	12

Symbol	Description	Unigene	GenBank	EG2 up	EG2 down	EG2 DE
(C)						
SLC1A4	Solute carrier family 1 (neutral amino acid transporter), member 4	Hs.654352	BF340083	11	7	18
ARRB1	Arrestin, beta 1	Hs.503284	BE207758	8	9	17
APOBEC3G	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	Hs.660143	NM_021822	8	8	16
MEST	Mesoderm-specific transcript homolog (mouse)	Hs.270978	NM_002402	8	8	16
CTSC	Cathepsin C	Hs.128065	AI246687	10	5	15
METTL7A	Methyltransferase like 7A	Hs.716437	NM_014033	10	5	15
GLIPR1	GLI pathogenesis-related 1 (glioma)	Hs.205558	AV682252	7	8	15
RAB27B	RAB27B, member RAS oncogene family	Hs.25318	BF438386	11	4	15

Table 2 (*Continued*)

Symbol	Description	Unigene	GenBank	EG2 up	EG2 down	EG2 DE
TRIB3	Tribbles homolog 3 (Drosophila)	Hs.516826	NM_021158	10	5	15
ABL2	v-abl Abelson murine leukemia viral oncogene homolog 2	Hs.159472	AK025877	9	5	14
EIF4EBP1	Eukaryotic translation initiation factor 4E binding protein 1	Hs.411641	AB044548	8	6	14
ETV1	ets variant gene 1	Hs.22634	BE881590	9	5	14
IFIT3	Interferon-induced protein with tetratricopeptide repeats 3	Hs.714337	AI075407	7	7	14
KLF9	Kruppel-like factor 9	Hs.150557	AI690205	3	11	14
MCC	Mutated in colorectal cancers	Hs.593171	BE967311	3	11	14
PDGFD	Platelet derived growth factor D	Hs.352298	NM_025208	6	8	14
PHGDH	Phosphoglycerate dehydrogenase	Hs.487296	NM_006623	4	10	14
PHLDA1	Pleckstrin homology-like domain, family A, member 1	Hs.602085	AK026181	10	4	14
PTGS2	Prostaglandin-endoperoxide synthase 2	Hs.196384	NM_000963	8	6	14
RAB1A	RAB1A, member RAS oncogene family	Hs.709257	BG530481	4	10	14

top 15 genes in this group were distinct from the top scorers obtained in experimental group 1. Moreover, most of the genes identified in this approach were not previously associated with cellular senescence. Hence, this list of genes may contain novel information about potential unidentified regulators of cellular aging, probably including genes that are relevant for cellular aging *in vivo*. The regulation of selected genes was also addressed at the protein level. On the one hand, selected genes were analyzed across all cellular model systems. For example, protein levels were determined for the conserved candidate gene EZH2 (DE = 19; Table S2), coding for a histone methyltransferase that plays a major role in epigenetic regulation of senescence (Bracken *et al.*, 2007.). Consistent with data obtained by RNA profiling, EZH2 protein level was considerably reduced in senescent RPTEC and HUVEC, as well as in PFF after treatment with adenosine monophosphate (AMP) and carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone (FCCP) but not oligomycin (Fig. 2A). EZH2 levels were not significantly regulated in MSC, PrSC, and CD8 T lymphocytes (Fig. 2B, C) consistent with the results of RNA profiling. Additional protein analyses were performed for relevant proteins selectively in particular model systems. Thus, we confirmed by Western blot upregulation of IGFBP-3 in trans-differentiated PrSC and MSC grown in a hyperoxic atmosphere (Fig. 2B). In additional experiments, we confirmed the upregulation of IGFBP-3 and IGFBP-6 in cellular supernatants by ELISA and the downregulation of IGF1R in $CD8^{+}CD28^{-}$ T lymphocytes by antibody assisted flow cytometry (data not shown).

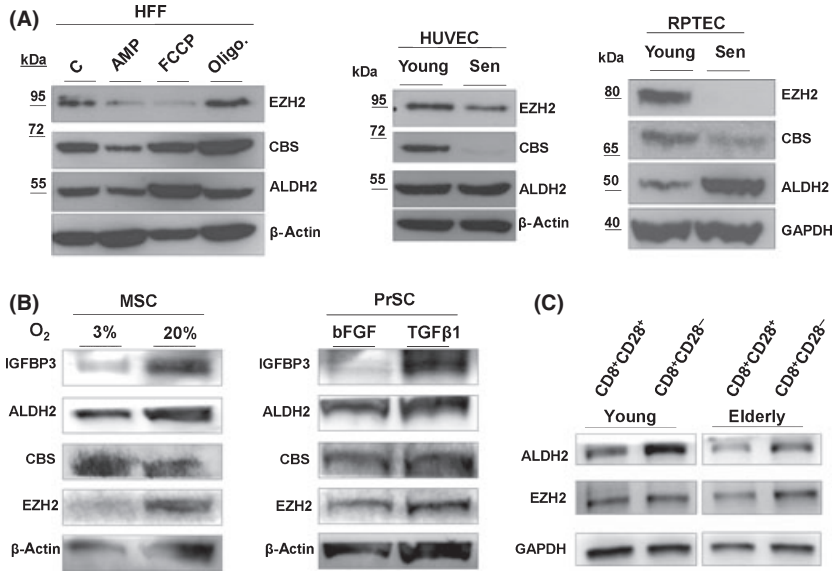

Fig. 2 Western blot analysis. Cell lysates were prepared from various model systems employed in this study: young and senescent human umbilical vein endothelial cells and renal proximal tubular epithelial cells, respectively; PFF treated with FCCP, AMP, and oligomycin for three days (A); mesenchymal stem cells cultured at atmospheric conditions of 3% or 20% oxygen; prostate stromal cells treated with either basic fibroblast growth factor (bFGF) or transforming growth factor beta 1 (TGFβ1) (B); CD8 T cells either isolated from a young or an elderly donor and sorted with respect to CD28 (C); Proteins were analyzed by immunoblotting using antibodies to EZH2, ALDH2, cystathionine beta-synthase, and IGFBP3, as indicated. For loading control, antibodies to β-actin and GAPDH were used as indicated.

Pathway analysis

To analyze the large amount of gene expression data in a user-friendly way and for optimal data mining, a database was constructed referred to as the GiSAO database ('Genes Involved in Senescence, Apoptosis and Oxidative stress' (Lepperdinger *et al.*, 2008); for details, see 'Experimental procedures'). Currently, more than a hundred whole-genome array data sets fulfilling MIAME criteria from cellular aging model systems were deposited by this consortium in the GiSAO database (https://gisao.genome.tugraz.at/gisao_web/), also including the datasets of miRNA expression published recently (Hackl *et al.*, 2010). To identify principal biological processes associated with age-related transcriptional changes, we first evaluated top-ranked genes via Pathway Explorer (Mlecnik *et al.*, 2005), a web-accessible tool that allows life scientists to infer the biological meaning behind large gene lists. Pathway analysis for genes that are DE in both experimental groups highlighted the p53 signaling pathway, apoptosis, and cell cycle regulation as the most relevant pathways. The general importance of pathways controlling cell proliferation and cell death for cellular senescence is well established (Mallette & Ferbeyre, 2007). Thus, these observations validate our experimental approach. The analysis identified additional pathways that were significantly overrepresented ($P < 0.001$) in the group of differentially regulated genes, which have not previously been linked to cellular senescence, such as cholesterol biosynthesis, IL-18 signaling, and nitrogen metabolism (Table 3).

Pathway analysis was also separately applied to experimental groups 1 and 2. For genes differentially expressed in experimental group 1 (oxidative stress), the highest ranked clusters were related to the regulation of apoptosis and cell cycle, in particular the G1/S transition. Additionally, chromosome maintenance, lipid biosynthesis, D4-GDI signaling, and DNA replication scored highly significant ($P < 0.001$). Among the genes differentially expressed in experimental group 2, top clusters related to apoptosis, MAPK signaling, and extracellular matrix. In addition, highly significant scores ($P < 0.001$) were observed for p53 signaling, enzyme inhibitor activity, cell cycle regulation, cell surface receptor linked signal transduction, and cell growth. In this group, we also noted a high representation of the Wnt signaling pathway and cytokine–cytokine receptor interactions, two pathways previously linked with cellular senescence (Ye *et al.*, 2007; Acosta et al. 2008; Coppe *et al.*, 2010b); however, in both cases, statistical significance was rather weak.

Table 3 Pathway analysis: functional classification of genes that are differentially expressed in experimental groups 1, 2 or both, using Pathway Explorer (Mlecnik *et al.*, 2005; available online: https://pathwayexplorer.genome.tugraz.at/), gene number attributed to pathway, *P*-values were calculated from complete expression value dataset (54 675 probe sets) with Fisher's exact test, where gene numbers are given, all *P*-values are < 0.05

Pathway	Gene number EG1	Gene number EG2	Gene number EG1 + EG2
Apoptosis	17	39	7
Cell cycle	15	27	7
p53 signaling pathway	7	32	7
Cell proliferation	*	32	7
Chromosome	14	*	5
Lipid biosynthesis	13	*	4
Cytokinesis	10	*	5
Pyrimidine metabolism	9	*	3
G1 to S cell cycle control	8	*	3
mRNA metabolism	10	*	*
ATP-dependent helicase activity	9	*	*
Fatty acid metabolism	9	*	*
mRNA processing	9	*	*
D4-GDI signaling pathway	7	*	*
DNA replication	7	*	*
MAPK signaling pathway	*	36	*
Extracellular matrix	*	33	*
Focal adhesion	*	29	*
Enzyme inhibitor activity	*	28	*
Cytokine–cytokine receptor interaction	*	27	*
Cell growth	*	24	*
Cell surface receptor linked signal transduction	*	24	*
Wnt signaling pathway	*	24	*
Cholesterol biosynthesis	*	*	3
Fructose and mannose metabolism	*	*	3
IL 18 signaling pathway	*	*	2
Nitrogen metabolism	*	*	3
Synthesis and degradation of ketone bodies	*	*	2

*No genes were found for particular pathway or $P > 0.05$.
Pathway data source: KEGG pathway database (http://www.genome.jp/kegg/pathway.html), GenMapp (http://www.genmapp.org/), Biocarta (http://www.biocarta.com/genes).

Functional validation of candidate genes by lifespan analysis in yeast mutants

Next, potential homologues of the human candidate genes were identified in the genome of budding yeast, *Saccharomyces cerevisiae*. To this end, an extended list of human genes was compiled with differential expression (DE > 1.5) in at least 6 of 17 experiments in experimental group 1 or at least 6 of 18 experiments in experimental group 2 (Fig. 1). Both upregulated and downregulated genes were considered. Starting with a total of 6750 human genes, 553 yeast orthologs were identified, of which only the nonessential genes were considered for further analysis. For the top ranking human genes, 93 nonessential yeast orthologs were identified (Table S4). Subsequently, functional tests regarding long-term proliferation were performed, and the viability of the respective mutant yeast strain was validated. The deletion mutants of the corresponding *S. cerevisiae* homologues were obtained from the EUROSCARF knockout strain collection. A panel of yeast mutant cells in stationary phase was analyzed by chronological lifespan experiments (MacLean *et al.*, 2001; Piper, 2006) involving 22-day culturing in stationary phase as described (Fabrizio *et al.*, 2004; Herker

et al., 2004). As control, wild-type strains were also analyzed, which have a mean lifespan of about 11 days under these conditions (Fabrizio *et al.*, 2004; Herker *et al.*, 2004). Twenty-two days after the start of the experiment, the wild-type strain completely ceased growth and was all found dead. The 93 selected yeast mutant strains displayed significantly different survival rates (Table 4), ranging from drastic lifespan shortening to lifespan extension relative to wild-type cells. For a more detailed analysis, the lifespan data were repeated for the nine most short-lived strains containing mutations in ATG18, GCM5, KGD1, LYS69, MSW1, NCR1, TIM1, RAD27, and SHM1. Lifespan shortening relative to the wild-type was highly reproducible (Fig. 3A), with the disruption of KGD1, MSW1, and TIM1 having the greatest effect on lifespan, suggesting that these genes play important roles for survival in stationary culture. Similarly, yeast mutants with extended lifespan were also identified (Table 4), and a significant extension of lifespan was obtained for the mutants ΔCYS4, ΔALD4, and ΔPDX3 (Fig. 3B). Whereas yeast strains mutated in UBC12 and PAS3 showed a trend for extended lifespan, this did not reach statistical significance. The *CYS4* gene encodes a cystathionine beta-synthase involved in the first step of cysteine biosynthesis while Ald4p is the major aldehyde dehydrogenase isoform. *PDX3*, encoding pyridoxine (pyridoxamine) 5′-phosphate oxidase, is involved in the salvage pathway of pyridoxal 5′-phosphate. As a first step to address the relevance of the human orthologs in our cellular senescence models, expression and regulation of ALDH2 and cystathionine beta-synthase (CBS) was analyzed by Western blot. ALDH2 protein level was upregulated in senescent RPTEC and in FCCP-treated PFF (Fig. 2), but remained unchanged in the other experimental models (Fig. 2 and data not shown). Cystathionine beta-synthase protein level was downregulated with senescence in both HUVEC and RPTEC, as well as in AMP-treated PFF (Fig. 2), consistent with the data obtained by RNA profiling. In contrast, CBS was not significantly regulated in the other cellular model systems (data not shown). Together, the data suggest that the levels of ALDH2 and CBS gene products are indeed regulated by the various treatments, in most cases reflecting the differential expression values as elucidated by our initial genomic screening method. The pattern of regulation is complex in both cases, and more work is required to understand the exact contribution of these genes to cellular senescence.

Table 4 Effects on chronological aging observed in 93 single deletion strains: deletion strains were assigned to five categories, depending on the effects on survival during aging when compared to the wild-type

Survival during chronological aging (compared to WT)	Single deletion of
Strongly reduced (< 50% of wt)	*KGD1*, *SOD2*, *RPL31A*, *NCR1*, *LYS9*, *SHM1*, *GCN5*, *HAP3*, *PIM1*, *PRY1*, *ATG18*, *CYC1*, *RPL13B*, *MSW1*, *RAD27*, *YDC1*, *TIF3*
Slightly reduced (> 50% and < 85%)	*CLB1*, *CLB2*, *APM1*, *SYM1*, *IDP2*, *GSY2*, *ALT2*, *SIZ1*, *TIF1*, *TIF2*, *NPL3*, *PTC5*, *RNR3*, *ERG24*, *MRT4*, *MSH6*, *NUP170*, *HAT1*, *INP53*, *SSO1*, *TWF1*, *SGT2*, *LHS1*, *RPL37A*, *ARF1*, *DHP5*, *CHL1*, *TRX1*, *UBC8*, *CHD1*, *SPE2*, *HRT3*, *MRE11*, *MSH2*, *DBP1*, *SSN3*, *SSF1*
Not affected (same as wt ±15%)	*HTA2*, *AIP1*, *HXK1*, *SER1*, *YPK2*, *UBC11*, *ENO1*, *RPL22B*, *RPL43B*, *PEX13*, *REV3*, *SYG1*, *MSH5*, *ERP6*, *TPK1*, *PER1*, *URA2*, *SLH1*, *CHK1*, *ALD5*, *CAF40*
Slightly increased (> 115% and < 150%)	*MAD2*, *PCH2*, *YTA7*, *PHO89*, *UGA1*, *DIE2*, *CAF17*, *HMT1*, *ALG5*, *SSA3*, *LSB6*
Strongly increased (> 150%)	*CYS4*, *DNF1*, *ALD4*, *PDX3*, *RNH201*, *UBC12*, *PRS3*

Discussion

The here presented analysis has identified gene expression signatures of cellular aging, which are conserved between different human tissues. The experimental design was based on the assumptions that (i) genes that are differentially regulated in human cellular aging and senescence are trustworthy candidates for modulators of aging and that (ii) the ability of a certain gene to restrict lifespan in yeast suggests that this gene has a role in modulating the rate of aging that potentially extends beyond yeast. This analysis revealed several genes and molecular pathways already known to be involved in aging, thereby validating the screen. In addition, a reasonable number of novel genes and a few novel pathways were identified that have not been linked to aging before. Based on these findings, new experimental approaches to study human aging will become possible.

Model systems for cellular aging

According to current hypotheses, aging at the cellular level contributes significantly to organismic aging, and for a better understanding of human aging, cellular model systems are indispensible (Campisi, 2005; Toussaint *et al.*, 2002a, 2002b). Several model systems for cellular aging have been developed, and a representative selection of these models was applied for the present study: replicative senescence, a process of cellular aging *in vitro* that is primarily driven by telomere erosion, is a classic experimental system considered relevant for the aging of tissues with high regenerative potential (von Figura *et al.*, 2009). In addition to replicative senescence, most if not all primary human cells can be driven into premature senescence by various stressors, in a process which does not necessarily involve telomere shortening. It is assumed that SIPS is relevant for aging of both mitotic and postmitotic tissues (Toussaint *et al.*, 2002a, 2002b). The importance of SIPS for *in vivo* aging is supported by the fact that SIPS can be induced *in vitro* by factors that are well-known risk factors for age-associated degeneration of the corresponding tissue *in vivo*. For example, vascular endothelial cells undergo SIPS in response to oxidative stress, high glucose, shear stress, and incubation with oxidized LDL, all of which are known risk factors for age-associated vascular dysfunction and disease (Erusalimsky & Kurz, 2006; Goligorsky *et al.*, 2009). Three additional model systems of cellular aging, with a particular

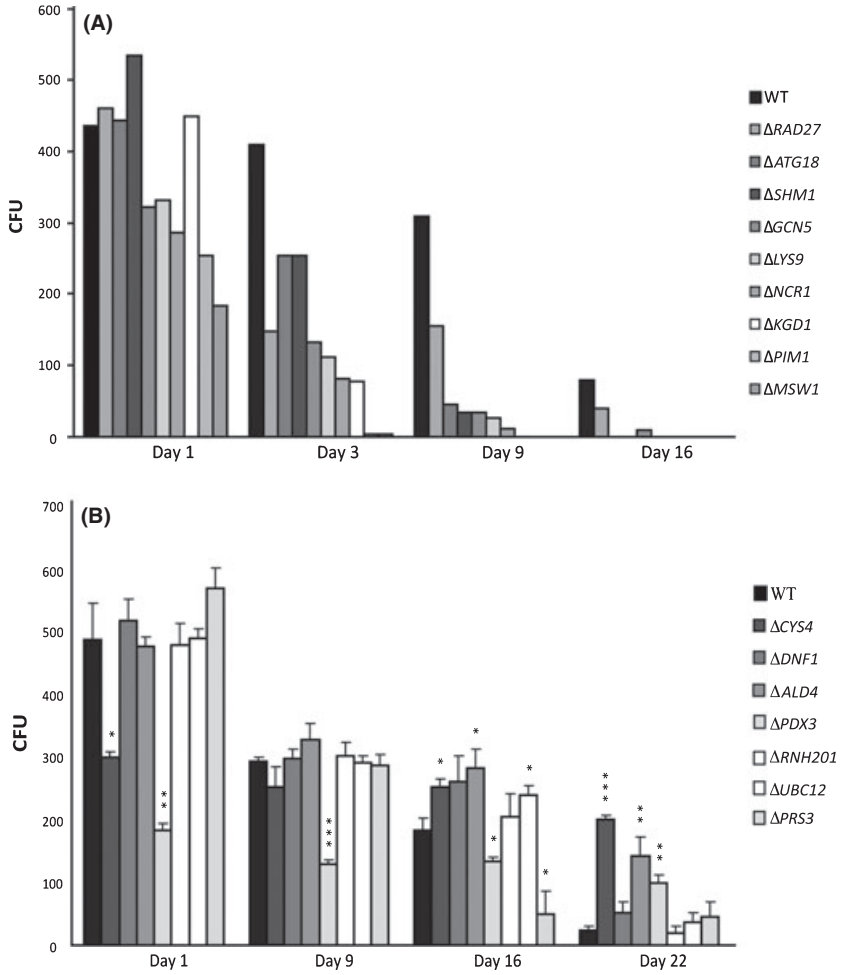

Fig. 3 Functional analysis of selected candidate genes in yeast. Lifespan data for the nine yeast mutants with shortest lifespan (A) and for the seven mutants with lifespan extension (B) are shown. (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

focus on aging *in vivo*, were included in the analysis: first, we compared $CD28^-$ and $CD28^+$ $CD8^+$ T lymphocytes, described as one of the rare models for *in vivo* senescence (Rufer *et al.*, 2003; Lazuardi *et al.*, 2009); second, we compared MSC freshly obtained from young and old human donors (Fehrer *et al.*, 2007); finally, PrSC were included, which can be transdifferentiated into premature senescent myofibroblasts, a process considered a hallmark of the aging human prostate (Untergasser *et al.*, 2005). These cellular model systems were developed by the participating laboratories and are not easily available elsewhere. Information on genetic regulators of cellular aging obtained from these systems is considered particularly valuable, given the scarcity of knowledge on molecular genetic processes involved in human aging.

The GiSAO database

Other resources have been created to systematically compare the influence of the genome on aging processes. Of note, the Human Ageing Genomic Resources (HAGR) (de Magalhaes *et al.*, 2005) represent a systematic collection of genomic data relevant for aging processes. It consists of the database GenAge featuring genes associated with aging and longevity in short-lived model organisms and humans, and AnAge, a database of aging in animals, featuring information on aging processes in over 4000 animal species (de Magalhaes *et al.*, 2009a, 2009b). The GiSAO database established in the current communication focuses on genetic regulation of aging at the cellular level, based on human gene expression data, and allows the identification of evolutionarily conserved regulators of cellular aging. In our view, information contained in the GiSAO database conveniently complements the data on organismic aging available in the HAGR. We have compared data sets in both studies. Indeed, several genes with a high score in the meta-analysis described earlier (de Magalhaes *et al.*, 2009a, 2009b), including TXNIP, S100A4, and MT1F, featured prominently in our analysis described here (Table S3). In turn, 15 genes taken from Table S3, including IGFBP-3, PTGS2, EGR1, FOS, ATM, DDIT3 (DE ≥ 20) are also highlighted as potential regulators of aging in the HAGR database. Hence, future attempts to link both resources are warranted.

Confirmation of known association of genes/pathways with aging

Several of the pathways highlighted in the current analysis have been associated with aging in cellular senescence and/or animal models before. For example, apoptosis regulation (Campisi, 2003), MAP kinase signaling (Maruyama *et al.*, 2009), p53 signaling (Zuckerman *et al.*, 2009), control of cell proliferation and cell cycle (Chen, 2000), cytokine–cytokine receptor interaction (Coppe *et al.*, 2010a), and Wnt signaling (Ye *et al.*, 2007) are thought to contribute to cellular senescence and aging. In addition, several of the genes identified in the unbiased screen described here were found to play important roles as regulators of senescence. Thus, Skp2 (Lin *et al.*, 2010) and EZH2 (Bracken *et al.*, 2007) are known to contribute to the senescence response in various cell types. Work in the consortium established the genes coding for IGFBP-3 (Kim et al., 2007; Muck *et al.*, 2008), glutaminase (Unterluggauer *et al.*, 2008), and IGFBP-6 (Micutkova *et al.*, submitted) as functional regulators of cellular senescence. The prominent appearance of these pathways and genes in our screening results serves as positive control and indicates that the screening procedure was adequate.

Identification of novel genes/pathways related to aging

Some interesting human genes, e.g. genes related to cholesterol biosynthesis, came up in our screen, which have not been related to aging before. Similarly, genes related to the regulation of focal adhesion, extracellular matrix, mRNA metabolism, fatty acid metabolism, and lipid biosynthesis were not known to be associated with aging processes. Although many of these genes could not be analyzed in yeast because there are no obvious orthologs in the yeast genome, these observations will stimulate future research on the role of the respective genes/pathways both in short-lived model organisms of aging and in human cellular aging.

Using extension of the chronological lifespan in yeast, which depends far less on cell-to-cell communication compared to higher organisms, as an additional screening tool, we expected to identify primarily basic cell autonomous mechanisms of aging. In this category, three genes particularly stood out, as follows: (i) ALD4, the major (mitochondrial) aldehyde dehydrogenase isoform, which is involved in the degradation of various amino acids and fatty acids in yeast, whereas the human homolog ALDH2 acts in the major oxidative pathway of alcohol metabolism; (ii) CYS4, cystathionine beta-synthase, that carries out the first committed step in cysteine biosynthesis, whereas the human homolog CBS catalyzes the conversion of homocysteine to cystathionine, the first step in the trans-sulfuration pathway; (iii) Pyridoxine (pyridoxamine) phosphate oxidase (PDX3), known to play a role in the synthesis of pyridoxal 5′-phosphate in yeast, similar to the human homolog pyridoxamine 5′-phosphate oxidase (PNPO), which catalyzes the terminal, rate-limiting step in the synthesis of pyridoxal 5′-phosphate, also known as vitamin B6. Interestingly, CBS requires vitamin B6 as a cofactor for the conversion of homocysteine to cystathionine in human cells, suggesting that CBS and PNPO influence aging via the same pathway. Importantly, none of these pathways have been linked to aging before, and future studies are required to address the role of these proteins in multi-cellular aging model systems including humans. As a first step to validate their relevance in human cellular senescence, expression levels were analyzed in cellular extracts where antibodies were available. We found differential regulation of ALDH2 and CBS protein levels in several of our model systems (Fig. 2). Accordingly, further studies are warranted to determine their role in cellular senescence.

Experimental procedures

Cell isolation, cultivation, and characterization

Human diploid fibroblasts from foreskin were seeded in 10-cm dishes at a density of 200,000 cells per dish and treated in parallel with the following compounds: untreated, vehicle treated (ethanol), oligomycin (8 μmol L^{-1} final concentration), FCCP (3 μmol L^{-1} final concentration), and AMP (3 μmol L^{-1} final concentration). Every day, medium was replaced and inhibitory compounds freshly added. This treatment leads to premature senescence within 14 days (data not shown, see also (Stockl *et al.*, 2006, 2007; Zwerschke *et al.*, 2003). RNA was prepared after 72 h of incubation and analyzed by RNA profiling as described earlier.

Human umbilical vein endothelial cells were isolated from human umbilical veins and cultured in Endothelial Cell Basal Medium (Lonza, Basel, Switzerland) supplemented with endothelial cell growth medium (EGM) Single Quots (Lonza), containing hEGF 0.5 mL, hydrocortisone 0.5 mL, GA-1000 0.5 mL, BBE 2.0 mL, and FBS 10.0 mL. The cells were subcultured by trypsination with trypsin-EDTA (Gibco Life Technologies, Vienna, Austria), seeded on cell culture dishes coated with 0.2% gelatine and grown at 37°C at ambient atmosphere containing 5% CO_2. Cells were passaged at a ratio of 1:5 in regular intervals. At later passages, the splitting ratio was reduced to 1:3 and 1:2, respectively. Cells were passaged before reaching 70–80% confluence. Population doublings (PDL) were estimated using the following equation: $n = (\log10\ F - \log10\ I)/0.301$ (where n is the population doublings, F, number of cells at the end of one passage, and I, number of cells that were seeded at the beginning of one passage). After 50 PDL, the cells reached growth arrest, and the senescent phenotype was verified by staining for senescence-associated beta galactosidase, which was positive for ≥ 95% of cells.

Human primary PrSC were established and trans-differentiation induced by TGF-beta1 treatment, as described (Untergasser *et al.*, 2005). Briefly, PrSC were derived from prostate cancer patients who have not received hormonal therapy (n = 3, age 65–72) after obtaining written informed consent. After radical prostatectomy and inspection by the pathologist, two tissue

wedges showing no histological signs of malignancy were removed from the transition zone. These explants were minced into organoids of ~ 1 mm^3 and seeded on uncoated plastic material in stromal cell growth medium containing insulin, human basic fibroblast growth factor, 5% fetal calf serum, and gentamycin and amphotericin-B as antibiotics (SCGM; Lonza). Explants were maintained at 37°C in a humidified atmosphere of 5% CO_2. These conditions produce a homogeneous fibroblast cell population after 7 days of culture. When cells reached 70% confluence, they were split at a 1:3 ratio using trypsin-EDTA to expand the population. In all experiments, cells of passage 2–4 were used directly from culture (not previously frozen).

Renal proximal tubular epithelial cells were cultivated as recently reported (Wieser *et al.*, 2008). In brief, within 24 h after surgery, tissue from the renal cortex was fragmented and incubated at 37°C for 15–20 min in DMEM/Ham's F12 (1:1) (Biochrom KG, Berlin, Germany) containing 1 mg mL^{-1} collagenase type IV (PAN-BioTech GmbH, Aidenbach, Germany) and 1 mg mL^{-1} trypsin-inhibitor (Sigma, Vienna, Austria). After being passed through a 105-μm nylon mesh, the filtrate was centrifuged, washed twice with phosphate-buffered saline (PBS), resuspended in medium, and dispensed into roux-flasks (Nunc, Wiesbaden, Germany). 24 hours thereafter medium was changed. The initial passage of confluent cells after 3–5 days was considered as PDL zero. Cells were passaged (1:2 to 1:4) at confluence, using 0.25% trypsin/0.02% EDTA, which was inactivated with 1 mg mL^{-1} trypsin-inhibitor. Cumulative PDL was calculated as a function of passage number and split ratio (4). Medium consisted of DMEM/Ham's F12 (1:1) supplemented with 4 mM L-glutamine, 10 mM HEPES buffer, 5 pM tri-iodothyronine, 10 ng mL^{-1} recombinant human EGF, 3.5 μg mL^{-1} ascorbic acid, 5 μg mL^{-1} transferrin, 5 μg mL^{-1} insulin, 25 ng mL^{-1} prostaglandin E1, 25 ng mL^{-1} hydrocortisone, and 8.65 ng mL^{-1} sodium selenite (all from Sigma). After 24 PDL, the cells reached growth arrest, and the senescent phenotype was verified by staining for senescence-associated beta galactosidase, which was positive for ≥ 95% of cells. Cells at intermediate passage could be divided into two groups concerning their redox status, as monitored by dichlorofluorescein diacetate (DCFDA) staining. To address the significance of this distinction, RPTECs were sorted into $DCFDA^{bright}$ and $DCFDA^{dim}$ subpopulations, which were analyzed separately.

Mesenchymal stem cells were isolated from the iliac crest of systemically healthy individuals (young donor, age 5; elderly donor, age 56), which had been harvested for reconstructive bone surgery of defects within other areas of the body as described previously (Fehrer *et al.*, 2007). Briefly, a small biopsy of substantia spongiosa osseum, which otherwise would have been discarded based on necessary bone for molding and re-contouring prior to insertion into the recipient site, was taken to further investigation under an Institutional Review Board-approved protocol after having obtained the parents' and the respective patient's written consent. After surgery, the bone was transferred into minimal essential medium (MEM) supplemented with 20% heat-inactivated fetal calf serum, 100 units mL^{-1} penicillin, and 100 μg mL^{-1} streptomycin (growth medium) for transportation from the operation theater to the clean room at room temperature. The biopsies were fragmented, and marrow cells were isolated from pieces (20–100 mm^3) by centrifugation (400 *g*, 1 min). After centrifugation, the remaining pieces were treated with collagenase (2.5 mg mL^{-1} in MEM) for 2–3 h at 37°C, 20% O_2, and 5% CO_2. Thereafter, the specimen was again centrifuged (400 *g*, 1 min). Cells were resuspended and loaded on a Ficoll-Paque Plus® gradient and centrifuged at 2500 *g* for 30 min. Cells were harvested from the interphase (density < 1.075 g mL^{-1}), washed, and collected by centrifugation (1500 *g*, 15 min). Cells were cultured at a density of 0.2–0.5 × 10^6 cells cm^{-2} at either 20% or 3% O_2 in combination with 5% CO_2 and 37°C (HeraCell240 – Heraeus, Thermo Scientific, Vienna, Austria; Thermo Electron Forma Series II, 3110). After 24 h, the nonadherent cell fraction was removed by washing twice with PBS. After the primary culture had reached approximately 30–50% confluence, cells were washed twice with PBS and subsequently treated with 0.05% trypsin/1 mM EDTA for 3–5 min at 37°C. Cells were harvested, washed in MEM, and further expanded at a density of 50 cells cm^{-2}.

Isolation of $CD8^+CD28^+$ and $CD8^+CD28^-$ T cells from peripheral blood of apparently healthy young (< 35 year, n = 6, mean age 29, range 26–35) and elderly (> 65 year, n = 10, mean age 72, range 66–87) donors was performed by preparing peripheral blood mononuclear cells (PBMCs) by Ficoll-Paque PLUS (Amersham Biosciences) density gradient centrifugation as approved by the Ethics committee of Innsbruck Medical University. $CD8^+$ T cells were negatively selected from the obtained PBMC fraction by applying the magnetic separation protocol $CD8^+$ T cell isolation kit II (depleting CD4, CD14, CD16, CD19, CD36, CD56, CD123, TCRγ/δ, and CD235a; Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's instructions. Subsequently, purified $CD8^+$ T cells were stained with an allophycocyanin (APC)-conjugated αCD28 monoclonal antibody (mAb) and split into $CD8^+CD28^+$ and $CD8^+CD28^-$ T-cell populations using αAPC MicroBeads (Miltenyi Biotec) by passing the cell suspension through a positive selection column (LS; Miltenyi Biotec) mounted in a magnetic field. The $CD8^+CD28^-$ T-cell fraction was then re-incubated with αAPC MicroBeads and run over a fresh LS-column to increase purity. For phenotypic analysis, purified T-cell fractions were labelled with a combination of mAbs (αTCRαβ-FITC, αCD16-PE, αCD4-PerCP, αCD8-PE-Cy7, αCD28-APC, and αCD3-APC-Cy7; all BD Biosciences, Heidelberg, Germany) and analyzed on a FACSCanto II (BD Biosciences) revealing that the described isolation protocol yields population homogeneities of > 95%.

Applying oxidative stress by tBHP treatment

Human umbilical vein endothelial cells, PrSC, RPTEC, and $CD8^+$ T lymphocytes were treated by exposure to sublethal doses of tBHP (Unterluggauer *et al.*, 2003). Induction of senescence by

repeated exposure to sublethal levels of tBHP takes 10–14 days (Unterluggauer *et al.*, 2003, and data not shown). To identify early transcriptional events involved in the senescence response, cells were analyzed 3 days after starting the treatment.

RNA Isolation, whole-genome array analysis of mRNA expression, and quantitative PCR

RNA was isolated using either Tri Reagent (Sigma-Aldrich) or performing homogenization in 4.2 M guanidinium thiocyanate before phenol extraction and ethanol precipitation (Chomczynski & Sacchi, 1987), followed by LiCl precipitation at a final concentration of 4.5 M. Whole-genome expression analysis was carried out on GeneChip HG-U133 Plus 2.0 Arrays (Affymetrix) by MFT Services, Tuebingen, Germany.

Bioinformatics, clustering, and visualization of array data

Array raw data were normalized via CARMAweb (Rainer *et al.*, 2006) using the gcrma algorithm (Wu *et al.*, 2004). Hierarchical clustering of samples and genes (Euclidian distance, average linkage) was performed after filtering out genes with low variance on a subset of 20 000 genes with the MeV program package (Saeed *et al.*, 2006) available online: http://www.tm4.org/mev/).

To evaluate the probability of observing an elevated number of under- or overexpressed gene occurrences, the distinct number of occurrences in which the gene is under- or overexpressed in a respective group of experiments was determined. The resulting gene list was ranked according to occurrences with respect to either age/senescence-related model systems or oxidative stress–derived data sets as well as a gene list of highly under- or overrepresented candidates present in both cases.

Statistical analysis to identify significant probe sets was performed as described by de Magalhaes *et al.* (2009a,b): For each probe set, a *P*-value was calculated with the cumulative binomial distribution (CBD):

$$P(X \geq k) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

providing the probability P for a probe set to be as often or more often differentially expressed than the times k, it was actually differentially expressed in n experiments. The threshold for differential expression of a probe set was defined as a fold change between samples $\geq \pm 1.5$. Applying statistical analysis based on CBD, the probability *P* that any probe set is differentially expressed with senescence or oxidative stress was defined as the average of differentially expressed probe sets in the experimental group senescence or oxidative stress divided by the number of all probe sets per whole-genome analysis.

The *q*-value was calculated for each probe set using Storey's false discovery rate (FDR) approach with the bootstrapping method (Storey *et al.*, 2004). The robust parameter was used to make the *q*-values more accurate for small *P*-values (Storey, 2002). Statistical computation was carried out using the statistical framework R (R-Team, 2007) version 2.5.1. Bioconductor *q*-value package version 1.10.0 (Storey & Tibshirani, 2003) was used to calculate the FDR.

Differentially expressed genes were grouped into protein families associated with characterized pathways applying Pathway Explorer ((Mlecnik *et al.*, 2005), available online: https://pathwayexplorer.genome.tugraz.at/), *P*-values were calculated from the complete expression value dataset (54 675 probe sets) with Fisher's exact test, pathway data sources: KEGG pathway database (http://www.genome.jp/kegg/pathway.html), GenMapp (http://www.genmapp.org/), Biocarta (http://www.biocarta.com/genes/).

Database

GiSAO.db (https://gisao.genome.tugraz.at) is a web-based database system for storing and retrieving data of genes involved in senescence, apoptosis, and oxidative stress. The application is based on a three-tier architecture consisting of a web interface, business logic, and a database. The web interface enables data input and presentation. It was implemented by using Struts framework (http://struts.apache.org/) with Java Server Pages (http://java.sun.com/products/jsp/). The business logic, which is responsible for data processing is an Enterprise JavaBeans 3 (http://java.sun.com/products/ejb/) application deployed on JBoss (http://www.jboss.org/jbossas/) application server. The data are stored in an Oracle database, a relational database management system.

The GiSAO database contains normalized gene expression values obtained from experiments evaluated with the aid of Affymetrix arrays. Gene expression values of each experiment can be displayed and compared with the gene expression values of two or more microarray experiments. Additionally, experimental data of follow-up experiments regarding candidate genes, such as qPCR or Western Blot analysis, are entered into the GiSAO.db. Furthermore, GiSAO.db contains two types of orthologue data: orthologue data provided and updated by Affymetrix from cross reference tables linking *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens* facilitate comparative genomic analyses, as well as orthology data computed and entered manually. Besides mRNA expression profiles, also data from proteome analysis and further validation with respect to functional analyses together with information from public resources are available for distinct candidate genes. Moreover, external links lead to orthologs of HomoloGene (http://www.ncbi.nlm.nih.gov/homologene) and InParanoid (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). The gene IDs are linked to their respective database, such as Entrez Gene or RefSeq. GiSAO.db also provides gene annotation (Gene Symbol, Gene Name, etc.) and GO terms (http://www.geneontology.org/). Finally, KEGG pathways (http://www.genome.jp/kegg/) can be displayed and data can be exported in various formats.

Identification of human orthologues in *Sacharomyces cerevisiae*

Either previously assigned cross references distinctly referring to homologous gene identifiers from other species provided in the array annotation tables accompanying the Affymetrix gene chip or data from reciprocal-best-fit-protein-homology searches (Moreno-Hagelsieb & Latimer, 2008) were employed to define the closest homologous gene pair between man and yeast. Only orthologous genes, which were found non-essential after genomic deletion in yeast laboratory strains (http://mips.helmholtz-muenchen.de/proj/eurofan/eurofan_1/b0/search/simpleSearch.html) were selected for further functional analyses.

Lifespan analysis of gene disruption mutants of *Sacharomyces cerevisiae*

Experiments were carried out in BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) as the wild-type strain and respective null mutants, obtained from Euroscarf. Strains were grown at 28°C on SC medium containing 0.17% yeast nitrogen base (Difco), 0.5% $(NH_4)_2SO_4$ and 30 mg L^{-1} of all amino acids (except 80 mg L^{-1} histidine and 200 mg L^{-1} leucine), 30 mg L^{-1} adenine, and 320 mg L^{-1} uracil with 2% glucose. For all experiments, yeast cells were grown at 28°C and 145 rpm. For chronological lifespan experiments, cultures were inoculated at an OD_{600} of 0.1, and aliquots were taken to perform survival plating at indicated time points.

Statistics

All values were expressed as means ± standard deviation of the mean. Statistical differences of experimental scores were evaluated using Student's *t* tests. Differences were considered significant when P was < 0.05.

Acknowledgments

We thank Michael Neuhaus for excellent technical assistance. This work was supported by the Austrian Science Fund (FWF, NFNS93) and the Austrian Ministry of Science and Research (GEN-AU project BIN). N.S. is a recipient of a Lise Meitner Scholarship and D.H. is supported by a European FLARE fellowship.

Author contributions

G.T.L., E.H., C.M., L.M., R.B., A.J., D.T., D.H., S.B., C.Z., N.S. and M.W. performed the cell culture work and contributed RNAs for the microarray analysis; D.R., D.C., J.R., and C.R. performed lifespan analysis in yeast; G.T.L., G.L., E.H., G.G.T., and Z.T. did the bioinformatic analysis and constructed the GiSAO database; G.T.L., M.B., K.F., B.G., P.B., R.G., J.G., Z.T., F.M., G.L., and P.J. conceived the experiments, interpreted the results, and wrote the manuscript.

References

Acosta JC, O'Loghlen A, Banito A, Raguz S, Gil J (2008) Control of senescence by CXCR2 and its ligands. *Cell cycle* **7**, 2956–2959.

Bartke A (2005) Minireview: role of the growth hormone/insulin-like growth factor system in mammalian aging. *Endocrinology* **146**, 3718–3723.

Bracken AP, Kleine-Kohlbrecher D, Dietrich N, Pasini D, Gargiulo G, Beekman C, Theilgaard-Mönch K, Minucci S, Porse BT, Marine JC, Hansen KH, Helin K (2007) The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev.* **2007**, 21.

Campisi J (2003) Cellular senescence and apoptosis: how cellular responses might influence aging phenotypes. *Exp. Gerontol.* **38**, 5–11.

Campisi J (2005) Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors. *Cell* **120**, 513–522.

Chen QM (2000) Replicative senescence and oxidant-induced premature senescence. Beyond the control of cell cycle checkpoints. *Ann. N Y Acad. Sci.* **908**, 111–125.

Chen Q, Ames BN (1994) Senescence-like growth arrest induced by hydrogen peroxide in human diploid fibroblast F65 cells. *Proc. Natl. Acad. Sci. U S A* **91**, 4130–4134.

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.

Coppe JP, Desprez PY, Krtolica A, Campisi J (2010a) The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* **5**, 99–118.

Coppe JP, Patil CK, Rodier F, Krtolica A, Beausejour CM, Parrinello S, Hodgson JG, Chin K, Desprez PY, Campisi J (2010b) A human-like senescence-associated secretory phenotype is conserved in mouse cells dependent on physiological oxygen. *PLoS ONE* **5**, e9188.

Dimri GP, Lee X, Basile G, Acosta M, Scott G, Roskelley C, Medrano EE, Linskens M, Rubelj I, Pereira-Smith O, Peacocke M, Campisi J (1995) A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl. Acad. Sci. U S A* **92**, 9363–9367.

Dumont P, Burton M, Chen QM, Gonos ES, Frippiat C, Mazarati JB, Eliaers F, Remacle J, Toussaint O (2000) Induction of replicative senescence biomarkers by sublethal oxidative stresses in normal human fibroblast. *Free Radic. Biol. Med.* **28**, 361–373.

Erusalimsky JD (2009) Vascular endothelial senescence: from mechanisms to pathophysiology. *J. Appl. Physiol.* **106**, 326–332.

Erusalimsky JD, Kurz DJ (2006) Endothelial cell senescence. *Handb Exp Pharmacol* **176** (Pt 2), 213–248.

Fabrizio P, Battistella L, Vardavas R, Gattazzo C, Liou LL, Diaspro A, Dossen JW, Gralla EB, Longo VD (2004) Superoxide is a mediator of an altruistic aging program in Saccharomyces cerevisiae. *J. Cell Biol.* **166**, 1055–1067.

Fehrer C, Brunauer R, Laschober G, Unterluggauer H, Reitinger S, Kloss F, Gully C, Gassner R, Lepperdinger G (2007) Reduced oxygen tension attenuates differentiation capacity of human mesenchymal stem cells and prolongs their lifespan. *Aging cell* **6**, 745–757.

von Figura G, Hartmann D, Song Z, Rudolph KL (2009) Role of telomere dysfunction in aging and its detection by biomarkers. *J. Mol. Med.* **87**, 1165–1171.

Goligorsky MS, Chen J, Patschan S (2009) Stress-induced premature senescence of endothelial cells: a perilous state between recovery and point of no return. *Curr. Opin. Hematol.* **16**, 215–219.

Hackl M, Brunner S, Fortschegger K, Schreiner C, Micutkova L, Muck C, Laschober GT, Lepperdinger G, Sampson N, Berger P *et al.*

(2010) miR-17, miR-19b, miR-20a, and miR-106a are down-regulated in human aging. *Aging cell* **9**, 291–296.

Herker E, Jungwirth H, Lehmann KA, Maldener C, Frohlich KU, Wissing S, Buttner S, Fehr M, Sigrist S, Madeo F (2004) Chronological aging leads to apoptosis in yeast. *J. Cell Biol.* **164**, 501–507.

Holzenberger M (2004) The role of insulin-like signalling in the regulation of ageing. *Horm. Res.* **62**(Suppl 1), 89–92.

Katic M, Kahn CR (2005) The role of insulin and IGF-1 signaling in longevity. *Cell. Mol. Life Sci.* **62**, 320–343.

Kim KS, Seu YB, Baek SH, Kim MJ, Kim KJ, Kim JH, Kim JR (2007) Induction of cellular senescence by insulin-like growth factor binding protein-5 through a p53-dependent mechanism. *Mol. Biol. Cell* **18**, 4543–4552.

Lazuardi L, Herndler-Brandstetter D, Brunner S, Laschober GT, Lepperdinger G, Grubeck-Loebenstein B (2009) Microarray analysis reveals similarity between CD8+CD28- T cells from young and elderly persons, but not of CD8+CD28+ T cells. *Biogerontology* **10**, 191–202.

Lener T, Moll PR, Rinnerthaler M, Bauer J, Aberger F, Richter K (2006) Expression profiling of aging in the human skin. *Exp. Gerontol.* **41**, 387–397.

Lepperdinger G, Berger P, Breitenbach M, Frohlich KU, Grillari J, Grubeck-Loebenstein B, Madeo F, Minois N, Zwerschke W, Jansen-Durr P (2008) The use of genetically engineered model systems for research on human aging. *Front Biosci* **13**, 7022–7031.

Lin HK, Chen Z, Wang G, Nardella C, Lee SW, Chan CH, Yang WL, Wang J, Egia A, Nakayama KI *et al.* (2010) Skp2 targeting suppresses tumorigenesis by Arf-p53-independent cellular senescence. *Nature* **464**, 374–379.

Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, Yankner BA (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891.

MacLean M, Harris N, Piper PW (2001) Chronological lifespan of stationary phase yeast cells; a model for investigating the factors that might influence the ageing of postmitotic tissues in higher organisms. *Yeast (Chichester, England)* **18**, 499–509.

de Magalhaes JP, Costa J, Toussaint O (2005) HAGR: the Human Ageing Genomic Resources. *Nucleic Acids Res.* **33**, D537–D543.

de Magalhaes JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, Church GM (2009a) The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging cell* **8**, 65–72.

de Magalhaes JP, Curado J, Church GM (2009b) Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*, **25**, 875–881.

Mallette FA, Ferbeyre G (2007) The DNA damage signaling pathway connects oncogenic stress to cellular senescence. *Cell cycle* **6**, 1831–1836.

Maruyama J, Naguro I, Takeda K, Ichijo H (2009) Stress-activated MAP kinase cascades in cellular senescence. *Curr Med Chem* **16**, 1229–1235.

Melk A, Halloran PF (2001) Cell senescence and its implications for nephrology. *J. Am. Soc. Nephrol.*, **12**, 385–393.

Mlecnik B, Scheideler M, Hackl H, Hartler J, Sanchez-Cabo F, Trajanoski Z (2005) PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res.* **33**, W633–W637.

Minamino T, Miyauchi H, Yoshida T, Ishida Y, Yoshida H, Komuro I (2002) Endothelial cell senescence in human atherosclerosis: role of telomere in endothelial dysfunction. *Circulation* **105**, 1541–1544.

Moreno-Hagelsieb G, Latimer K (2008) Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **24**, 319–324.

Muck C, Micutkova L, Zwerschke W, Jansen-Durr P (2008) Role of insulin-like growth factor binding protein-3 in human umbilical vein endothelial cell senescence. *Rejuvenation Res* **11**, 449–453.

Partridge L (2009) Some highlights of research on aging with invertebrates, 2009. *Aging cell* **8**, 509–513.

Passos JF, Saretzki G, Ahmed S, Nelson G, Richter T, Peters H, Wappler I, Birket MJ, Harold G, Schaeuble K *et al.* (2007) Mitochondrial dysfunction accounts for the stochastic heterogeneity in telomere-dependent senescence. *PLoS Biol.* **5**, e110.

Piper PW (2006) Long-lived yeast as a model for ageing research. *Yeast* **23**, 215–226.

R Development Core Team (2007) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org [accessed on 16 August 2010].

Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z (2006) CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res.* **34**, W498–W503.

Ressler S, Bartkova J, Niederegger H, Bartek J, Scharffetter-Kochanek K, Jansen-Dürr P, Wlaschek M (2006) p16INK4A is a robust in vivo biomarker of cellular aging in human skin. *Aging Cell.* **5**, 379–389.

Rodwell GE, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, Xiao W, Mindrinos M, Crane E, Segal E *et al.* (2004) A transcriptional profile of aging in the human kidney. *PLoS Biol.* **2**, e427.

Rufer N, Zippelius A, Batard P, Pittet MJ, Kurth I, Corthesy P, Cerottini JC, Leyvraz S, Roosnek E, Nabholz M *et al.* (2003) Ex vivo characterization of human CD8+ T subsets with distinct replicative history and partial effector functions. *Blood* **102**, 1779–1787.

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J (2006) TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193.

Stockl P, Hutter E, Zwerschke W, Jansen-Durr P (2006) Sustained inhibition of oxidative phosphorylation impairs cell proliferation and induces premature senescence in human fibroblasts. *Exp. Gerontol.* **41**, 674–682.

Stockl P, Zankl C, Hutter E, Unterluggauer H, Laun P, Heeren G, Bogengruber E, Herndler-Brandstetter D, Breitenbach M, Jansen-Durr P (2007) Partial uncoupling of oxidative phosphorylation induces premature senescence in human fibroblasts and yeast mother cells. *Free Radic. Biol. Med.* **43**, 947–958.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, **64**, 479–498.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc.Natl.Acad.Sci.U.S.A*, **100**, 9440–9445.

Storey JD, Taylor JE, Siegmund D (2004) Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, Series B*, **66**, 187–205.

Tan Q, Christensen K, Christiansen L, Frederiksen H, Bathum L, Dahlgaard J, Kruse TA (2005) Genetic dissection of gene expression observed in whole blood samples of elderly Danish twins. *Hum. Genet.* **117**, 267–274.

Toussaint O, Royer V, Salmon M, Remacle J (2002a) Stress-induced premature senescence and tissue ageing. *Biochem. Pharmacol.* **64**, 1007–1009.

Toussaint O, Remacle J, Dierick JF, Pascal T, Frippiat C, Zdanov S, Magalhaes JP, Royer V, Chainiaux F (2002b) From the Hayflick mosaic to the mosaics of ageing. Role of stress-induced premature senescence in human ageing. *Int. J. Biochem. Cell Biol.* **34**, 1415–1429.

Turchi L, Fareh M, Aberdam E, Kitajima S, Simpson F, Wicking C, Aberdam D, Virolle T (2009) ATF3 and p15PAF are novel gatekeepers of genomic integrity upon UV stress. *Cell Death Differ.* **16**, 728–737.

Untergasser G, Madersbacher S, Berger P (2005) Benign prostatic hyperplasia: age-related tissue-remodeling. *Exp. Gerontol.* **40**, 121–128.

Unterluggauer H, Hampel B, Zwerschke W, Jansen-Durr P (2003) Senescence-associated cell death of human endothelial cells: the role of oxidative stress. *Exp. Gerontol.* **38**, 1149–1160.

Unterluggauer H, Mazurek S, Lener B, Hütter E, Eigenbrodt E, Zwerschke W, Jansen-Dürr P (2008) Premature senescence of human endothelial cells induced by inhibition of glutaminase. *Biogerontology* **2008**, 9.

Vasile E, Tomita Y, Brown LF, Kocher O, Dvorak HF (2001) Differential expression of thymosin beta-10 by early passage and senescent vascular endothelium is modulated by VPF/VEGF: evidence for senescent endothelial cells in vivo at sites of atherosclerosis. *FASEB J.* **15**, 458–466.

Wieser M, Stadler G, Jennings P, Streubel B, Pfaller W, Ambros P, Riedl C, Katinger H, Grillari J, Grillari-Voglauer R (2008) hTERT alone immortalizes epithelial cells of renal proximal tubules without changing their functional characteristics. *Am J Physiol Renal Physiol* **295**, F1365–F1375.

Wu Z *et al.* (2004) A model-based background adjustment for oligonucleotideexpression arrays. *J. Am Stat. Assoc.* **99**, 909–917.

Ye X, Zerlanko B, Kennedy A, Banumathy G, Zhang R, Adams PD (2007) Downregulation of Wnt signaling is a trigger for formation of facultative heterochromatin and onset of cell senescence in primary human cells. *Mol. Cell* **27**, 183–196.

Yu P, Huang B, Shen M, Lau C, Chan E, Michel J, Xiong Y, Payan DG, Luo Y (2001) p15(PAF), a novel PCNA associated factor with increased expression in tumor tissues. *Oncogene* **20**, 484–489.

Zuckerman V, Wolyniec K, Sionov RV, Haupt S, Haupt Y (2009) Tumour suppression by p53: the importance of apoptosis and cellular senescence. *J. Pathol.* **219**, 3–15.

Zwerschke W, Mazurek S, Stockl P, Hutter E, Eigenbrodt E, Jansen-Durr P (2003) Metabolic analysis of senescent human fibroblasts reveals a role for adenosine monophosphate in cellular senescence. *Biochem. J.* **376**, 403–411.

Supporting Information

Additional supporting information may be found in the online version of this article:

Fig. S1 Validation of microarray analysis: expression values of 22 genes exhibiting differential expression of > 1.5-fold in more than 18 of 37 experimental pairs were assessed by quantitative reverse transcription–PCR.

Fig. S2 Hierarchical clustering of genome-wide gene expression profiles derived from all experimental human cell models (47 individual datasets).

Fig. S3 Comparison of numbers of differentially expressed probe sets (DE > 1.5) between individual members of EG2.

Table S1 Experimental systems: HUVEC (human umbilical vein epithelial cells), PFF (primary foreskin fibroblasts), PrSC (primary prostate stromal cells), RPTEC (renal proximal tubular epithelial cells), CD8 ($CD8^+$ T lymphocytes), MSC (mesenchymal stem cells), Reagents: bFGF: basic fibroblast growth factor, TGFβ: transforming growth factor-β 1, t-BHP (tert-butylhydroperoxide), FCCP (carbonyl cyanide 4-(trifluoromethoxy)phenylhydrazone), AMP (adenosine monophosphate), DCFDA (2′,7′-dichlorofluorescein diacetate), number of up- and downregulated probe sets for each array experiment (DE > 1.5, total probe set number: 54 675 for Affymetrix U133 Plus 2.0 array)

Table S2 Genes that are differentially expressed (DE > 1.5) in ≥ 50% of the experiments for experimental groups 1 and 2, numbers for up- and downregulation in EG1 and EG2, *P*-values (cumulative binomial distribution), *q*-values (Storeys FDR method)

Table S3 Three hundred and thirty-four genes differentially expressed (DE > 1.5) with the highest combined score for both experimental groups (≥ 18), provided as Excel file

Table S4 Highest scoring genes, which were found differentially expressed (DE > 1.5) in both experimental groups, for which an orthologous nonessential gene in yeast could be identified (according to orthology information provided by Affymetrix Inc

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Hofer et al. BMC Genomics 2011, 12:262
http://www.biomedcentral.com/1471-2164/12/262

BMC Genomics

DATABASE Open Access

GiSAO.db: a database for ageing research

Edith Hofer[1,2], Gerhard T Laschober[3], Matthias Hackl[4], Gerhard G Thallinger[1], Günter Lepperdinger[3], Johannes Grillari[4], Pidder Jansen-Dürr[3] and Zlatko Trajanoski[2*]

Abstract

Background: Age-related gene expression patterns of *Homo sapiens* as well as of model organisms such as *Mus musculus*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* are a basis for understanding the genetic mechanisms of ageing. For an effective analysis and interpretation of expression profiles it is necessary to store and manage huge amounts of data in an organized way, so that these data can be accessed and processed easily.

Description: GiSAO.db (Genes involved in senescence, apoptosis and oxidative stress database) is a web-based database system for storing and retrieving ageing-related experimental data. Expression data of genes and miRNAs, annotation data like gene identifiers and GO terms, orthologs data and data of follow-up experiments are stored in the database. A user-friendly web application provides access to the stored data. KEGG pathways were incorporated and links to external databases augment the information in GiSAO.db. Search functions facilitate retrieval of data which can also be exported for further processing.

Conclusions: We have developed a centralized database that is very well suited for the management of data for ageing research. The database can be accessed at https://gisao.genome.tugraz.at and all the stored data can be viewed with a guest account.

Background

Accumulated cell damage is one of the main perpetrators of ageing. The damage is caused by a variety of different factors and conditions, including somatic mutations, mitochondrial dysfunction and oxidative stress [1,2]. If damage of cellular components (proteins, nucleic acids, lipids, etc.) remains permanently and is not corrected by repair systems (e.g. DNA repair or the elimination of damaged organelles and proteins), then cellular senescence and/or apoptosis is occuring. Cellular senescence and apoptosis contribute to a characteristic ageing phenotype as well as to the development of age-related diseases [3,4]. Since the underlying mechanisms of cellular ageing, leading to senescence or perhaps to apoptosis, have not yet been fully revealed, it is indispensable to identify and study genes and miRNAs which are involved in the ageing process. An effective way to determine these genes and gene-regulatory miRNAs are genome-wide studies of expression patterns, as it is well known that expression profiles of organisms change with age [2,5]. Microarrays are well suited for this task as they are a high-throughput method for determining the expression of tens of thousands of genes in parallel [6]. Results of microarray experiments are usually validated by applying low-throughput methods for measuring gene expression, e.g. qPCR or Northern blots [7], or confirmed with protein assays like Western blots.

Genetic research into human ageing is supported by investigation of ageing in various model organisms, such as *Mus musculus, Saccharomyces cerevisiae, Caenorhabditis elegans* and *Drosophila melanogaster*. These organisms have a much shorter lifespan than humans and can be easily genetically manipulated for experimental purposes. The results obtained from model organisms can be transferred to a certain extent to *Homo sapiens*, since these organisms share orthologous genes [8]. However, in order to effectively analyse data generated in various experiments using different organisms, it is necessary to structure and manage this data in an organized way.

Therefore, several publicly available databases which store ageing specific gene information were developed: the Human Aging Genomic Resources (HAGR) [9], the Gene Aging Nexus (GAN) [10], the Aging Gene Database

* Correspondence: zlatko.trajanoski@i-med.ac.at
[2]Division for Bioinformatics, Biocenter, Innsbruck Medical University, Schöpfstrasse 45, 6020 Innsbruck, Austria
Full list of author information is available at the end of the article

[11], the Atlas of Gene Expression in Mouse Aging Project (AGEMAP) [12], and the NetAge database [13]. However, to the best of our knowledge, there is no database which contains microarray gene expression data together with orthologous genes, ageing-related microarray miRNA expression data as well as data of follow-up experiments. We have therefore initiated the development of a database GiSAO.db (**G**enes **i**nvolved in **s**enescence, **a**poptosis and **o**xidative stress) to support ongoing and future studies in experimental ageing research.

Construction and content

GiSAO.db is a database for storing and managing expression data of genes involved in senescence, apoptosis and oxidative stress. It is connected to a web application which provides an easy and controlled access to this data. Specifically, the database is capable of storing four data types: expression data, annotation data, orthologous data and data of follow-up experiments (Figure 1).

Expression data

Normalized gene and miRNA expression values obtained from microarray experiments investigating ageing reside in GiSAO.db. It is possible to store gene expression data from Affymetrix one-colour microarrays as well as miRNA expression data from Exiqon two-colour microarrays in the database.

Annotation data

For genes, several gene identifiers are available as annotation in GiSAO.db: GeneSymbol, Refseq Id, Gene Name, EntrezGene Id, UniProt Id, UniGene Id, SGD Id, MGI Id, FlyBase Id, RGD Id and AGI Id. These identifiers were obtained together with Gene Ontology (GO) [14] terms from Affymetrix which provides annotation data of each spot on a microarray chip [15]. In case of miRNAs, the miRNA name and the miRBase Id [16] are stored as annotations in the database.

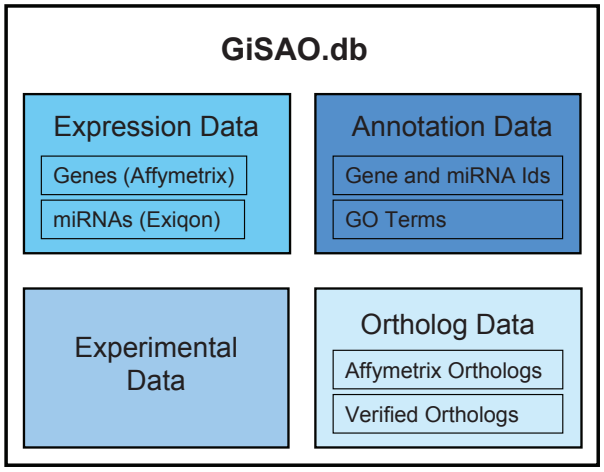

Figure 1 GiSAO overview. Four different types of data are stored in the database: gene expression data obtained from Affymetrix microarray experiments and miRNA expression data from Exiqon microarray experiments, annotations for genes and miRNAs including GO terms, orthologs data provided by Affymetrix as well as manually entered verified orthologs data and experimental data of follow-up experiments.

Orthologs data

Two types of orthologous data are included in GiSAO.db: orthologs provided by Affymetrix and verified orthologs data. Pairs of orthologs probe sets from different Affymetrix microarray chips are stored. Moreover, verified orthologous gene pairs between different species retrieved from other sources, such as literature or orthologs databases can be entered manually.

Experimental data

Finally, GiSAO.db provides facilities to store data of follow-up experiments. An experiment is specified by properties such as type (e.g. qPCR, Western blot), classification (e.g. senescence, inflammation), organism and cell type. All genes that were investigated can be linked to the experiment and antibodies or primers can be defined. Moreover, it is possible to specify references, and protocols as well as result files may be uploaded and attached to experiments or their associated genes.

Database content

GiSAO.db provides annotation data for five Affymetrix microarrays: Human Genome U133 Plus 2.0 Array, Mouse Genome 430 2.0 Array, Yeast Genome 2.0 Array, C. elegans Genome Array and Drosophila Genome Array. Furthermore, annotations for two custom made human Exiqon miRNA microarrays are available. The database contains orthologs provided by Affymetrix between *Homo sapiens, Mus musculus, Saccharomyces cerevisiae, Caenorhabditis elegans* and *Drosophila melanogaster*. Currently GiSAO.db stores gene expression values of 11 experiments comprising 111 Affymetrix microarrays of three different species: *Homo sapiens, Mus musculus* and *Saccharomyces cerevisae*. Additionally there are 7 human miRNA experiments with 40 Exiqon microarrays stored in the database. Moreover, numerous verified orthologs and data of several follow-up experiments are available in GiSAO.db.

Normalization of expression data obtained from Affymetrix microarray experiments was performed using the gcrma algorithm [17] in CARMAweb [18]. Independently of the particular experiments, data of a certain cell type, e.g. HUVEC or PFF, were normalized together.

The statistical framework R [19] and the Bioconductor package limma [20] were used for background correction

of miRNA expression data with the normexp algorithm as well as lowess normalization [21].

Implementation

The GiSAO.db database system was developed using the object-oriented and platform independent Java programming language [22]. Based on the Java Enterprise Edition (Java EE) platform [23], a three-tier application composed of a relational database, business logic and presentation layer was implemented.

In order to control data access and manage user data, the web application offers an integrated authentication and authorization system [24].

Utility

User Interface

The database system offers a user-friendly web interface which facilitates data input and retrieval. Results of microarray experiments can be viewed in detail as both the expression value of each spot for one-colour microarray chips, and the expression ratio between the colour channels for two-colour microarrays are displayed. Expression values and ratios are represented by colour-coded boxes which facilitate the determination of highly expressed genes, up- or down-regulated miRNAs and the comparison of expression values and ratios of different microarrays (Figure 2). The values and ratios can be displayed in a logarithmic or decimal scale, and a threshold can be defined to show only those genes or miRNAs whose expression values or ratios exceed the defined cut-off.

Pairs of orthologous genes can be retrieved using a simple search function which returns Affymetrix orthologs as well as verified orthologs. For experimental data of follow-up experiments, the application provides a flexible query mechanism which accepts organism, experiment classification and experiment type as parameters. Additionally, tags are displayed in gene lists to provide basic information about the different experiments performed on a gene at first glance. Tags are essentially shortcuts describing experiment classification, experiment type and organism, referencing experimental data (Figure 2). Genes or miRNAs which are of special interest for users can be assembled to *favourite lists* and furnished with additional information. These lists can be compared to check for common entries (Figure 3).

A comprehensive search function that takes gene and miRNA Ids as parameters provides access to all data about a gene or miRNA in GiSAO.db. The search yields expression data, annotation data, orthologs, experimental data tags and favourite lists of the specified gene or miRNA (Figure 4). An export mechanism which enables further processing of data from the database in external tools is seamlessly included into GiSAO.db. Lists of expression values, favorite genes and orthologs can be written to plain text files, PDF files or files in comma separated values (CSV) format.

Identifiers of genes and miRNAs as well as GO terms are provided to attribute a meaning to the probe (set) identifiers of the microarray spots. To enhance stored components with additional information, links to external databases are offered. Gene Ids are linked to their respective databases, e.g. RefSeq, Entrez Gene or UniProt, and to the ortholog databases HomoloGene [25] and InParanoid [26]. Moreover, pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [27] can be accessed in GiSAO.db via a web service provided by KEGG.

Data Submission

There are two ways of entering data in the database: manual data input using web forms or data upload from files. Forms are enhanced by a data dictionary concept, which extends the functionality of select fields, facilitates data input, and prevents inconsistencies in the database caused by spelling mistakes or duplicate entries.

Verified orthologs and experimental data can be entered in GiSAO.db using specific upload forms.

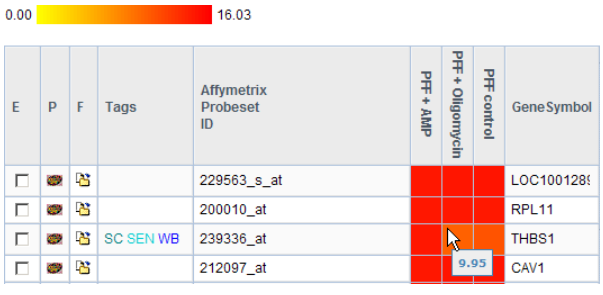

Figure 2 Display of expression values obtained from Affymetrix one-colour arrays. The values are colour coded to facilitate the identification of prominent expression values and comparison between values of a gene of different arrays. Additionally, tags are displayed for a quick overview of experimental data. In this case, for the gene THBS a follow-up experiment was performed with the organism *Homo sapiens* (HS), experiment classification *oxidative stress* (OS) and experiment type *qPCR* (QP).

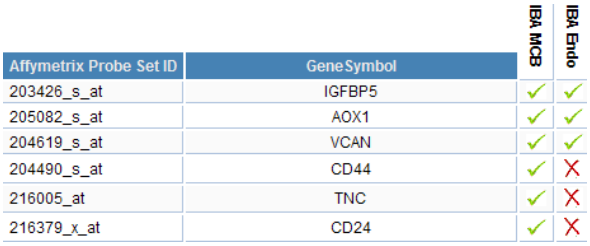

Figure 3 Favourite gene list comparison. Genes of interest can be collected in lists and two or more favourite lists can be compared to determine common genes.

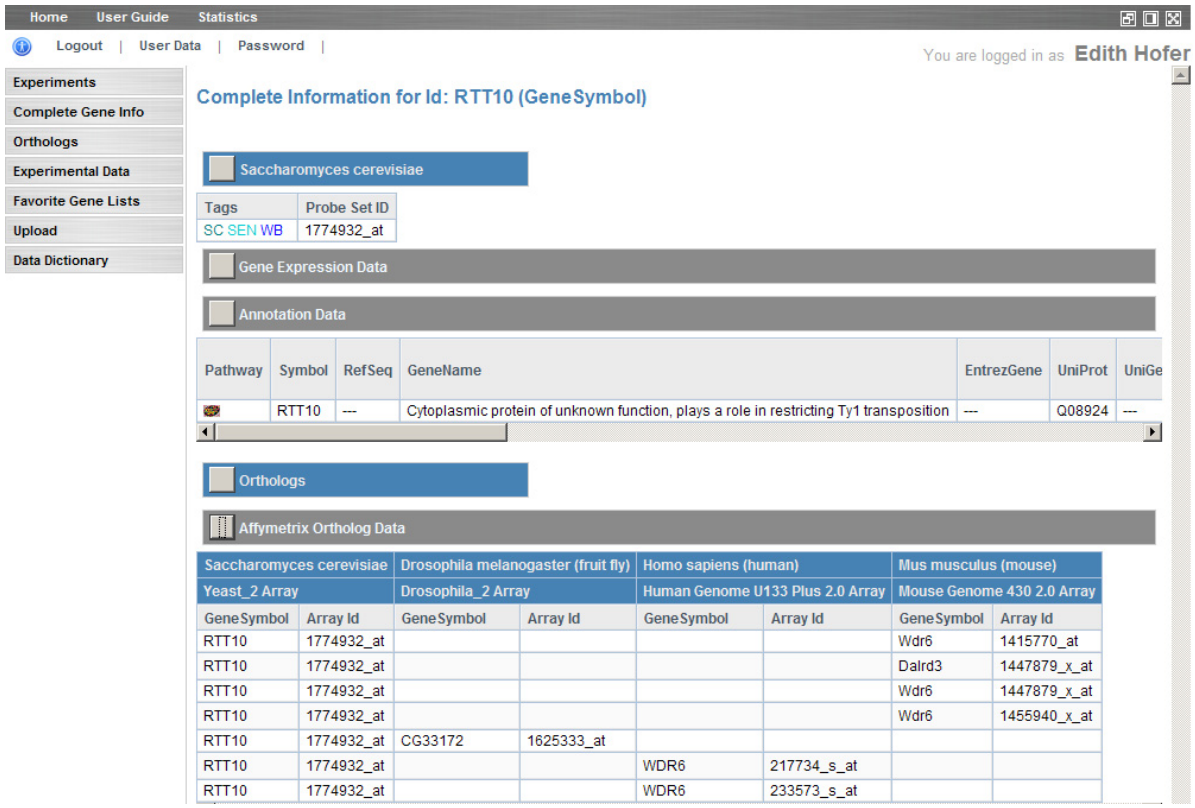

Figure 4 Gene search result. For the queried gene all the information stored in the database, i.e. tags, gene expression data, annotation data and orthologs data are listed.

Experimental data can be additionally uploaded from files as well as favorite lists, expression data, annotation data and Affymetrix ortholog data to accelerate the data input process. Affymetrix and Exiqon annotation data of microarray chips as well as Affymetrix ortholog data are updated on a regular basis by the producers. To adopt these changes in GiSAO.db, update functions have been implemented.

Moreover, several protocols or result files of follow-up experiments can be uploaded at once by using a Java applet. For each file upload, a feedback report is created which allows the user to check whether the upload was successful and view error messages in case something went wrong.

Authentication and authorization

To protect the data in the database fine grained access rights were defined in the integrated authentication and authorization system [24]. Three different user roles can access GiSAO.db: administrator, user and guest. An "administrator" can add, edit and delete all data. A "user" is allowed to add data, edit and delete his/her own data or data belonging to a member of the user's institute. Finally, a user assigned the "guest" role may view all the data stored in the database but has no rights to add, edit or delete any data.

Case study

Expression data of 47 human Affymetrix microarrays stored in GiSAO.db were analyzed. The experiments were performed on different cell types and were divided into two different groups: premature senescence induced by oxidative stress and other senescence models. The derived expression profiles were compared to determine conserved patterns in the various cellular aging models. As a result 484 genes associated with oxidative stress-induced senescence, 1087 genes associated with other senescence models, and 155 genes which seem to play a role in both experimental settings were determined [28]. This information guided the selection of 93 candidate genes which were tested for their ability to modulate lifespan in a unicellular model system (yeast chronological lifespan) to study organismic ageing [28]. Thereby, several new pathways which may be important in cellular senescence were identified [28]. Additionally, GiSAO.db was used supporting another study investigating the contribution of miRNAs in ageing [21].

Up to date 87 expression data sets stored in GiSAO.db have also been published in the public repository ArrayExpress [29]: 51 Affymetrix arrays from 6 experiments (E-MEXP-2283, E-MEXP-2285, E-MEXP-2167, E-MEXP-2345, E-MEXP-1506 and E-MEXP-2683) as well as 36 Exiqon arrays from 6 experiments (E-MEXP-2386, E-MEXP-2425, E-MEXP-2393, E-MEXP-2398, E-MEXP-2455, E-MEXP-2459).

Discussion

GiSAO.db is a database for the storage and management of ageing-related data. The core of GiSAO.db consists of normalized gene expression and miRNA expression data retrieved from microarray experiments. Annotation data like gene or miRNA identifiers as well as GO terms are available to interpret the expression profiles. As many of the orthologs provided by Affymetrix are only predicted ones, a manual curation of verified orthologs was implemented. The orthologs in the database facilitate cross-species comparison of expression profiles and the detection of evolutionary conserved expression patterns. Data of follow-up experiments, e.g. qPCR or Western blot experiments complement the microarray expression data. For a quick overview on these follow-up experiments performed for a specific gene, tags which serve as links to experimental data can be added.

Additionally, links to external gene, miRNA and ortholog databases are offered as well as KEGG pathways. Data upload and update is performed asynchronously, meaning that GISAO.db can be used while the upload takes place. In web forms data dictionary fields support controlled, yet customizable data input to keep the database content consistent. Search functions deliver data from the database which can then be exported in various file formats that are suitable for direct import in programs like MS Excel that are used for further processing of the data.

Furthermore, genes or miRNAs of interest can be grouped into favorite lists which can then be compared among different research groups. A sophisticated authentication and authorization system prevents undesired manipulation of data, yet allows all users to view the entire content of the database. By using a three-tier architecture, maintenance and extension of the application is facilitated and the various layers, e.g. the underlying database system, may also be exchanged. The usability of the web application is greatly enhanced by Web 2.0 functionality which was added using AJAX technology.

Conclusions

We have developed GiSAO.db, a system for storage and management of ageing-related gene data. An intuitive user interface provides fast and organised access to these data. Additionally, references to external databases are offered to elaborate the data in the database. These features in combination with the stored gene and miRNA expression data, annotation data, orthologs data and data of follow-up experiments make it a powerful tool for genetic ageing research.

Availability

GiSAO.db is available at http://gisao.genome.tugraz.at. All the data stored in the database can be viewed with a guest account. Username and password for guest users are provided on the login page of the application.

Acknowledgements

The authors thank Martina Pitzl for the initial development of GiSAO.db. This work was supported by the Austria Science Fund (NFN Project S93 Proliferation, Differentiation and Apoptosis in Aging) and the GEN-AU project BIN from the Austrian Ministry for Science and Research.

Author details

[1]Institute for Genomics and Bioinformatics, Graz University of Technology, Petersgasse 14, 8010 Graz, Austria. [2]Division for Bioinformatics, Biocenter, Innsbruck Medical University, Schöpfstrasse 45, 6020 Innsbruck, Austria. [3]Institute for Biomedical Aging Research, Austrian Academy of Sciences, Rennweg 10, 6020 Innsbruck, Austria. [4]Aging and Immortalization Research, Department of Biotechnology, University of Natural Resources and Applied Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria.

Authors' contributions

EH designed and implemented the application and drafted the manuscript. GGT contributed to conception, design, and implementation of the application. PJD contributed to the concept of the database, was involved in data upload and data quality control issues, and contributed to writing the manuscript. GL, MH, and GL contributed to the concept of the database and were involved in data upload. PJD and ZT were responsible for the overall project coordination. All authors gave final approval of the version to be published.

Received: 16 February 2011 Accepted: 24 May 2011
Published: 24 May 2011

References

1. Dilman VM: **Age-associated elevation of hypothalamic, threshold to feedback control, and its role in development, ageing, and disease.** *Lancet* 1971, **1**:1211-1219.
2. Weinert BT, Timiras PS: **Invited review: Theories of aging.** *J Appl Physiol* 2003, **95**:1706-1716.
3. Campisi J, d'Adda dF: **Cellular senescence: when bad things happen to good cells.** *Nat Rev Mol Cell Biol* 2007, **8**:729-740.
4. Campisi J: **Cellular senescence and apoptosis: how cellular responses might influence aging phenotypes.** *Exp Gerontol* 2003, **38**:5-11.
5. Chen LH, Chiou GY, Chen YW, Li HY, Chiou SH: **microRNA and aging: a novel modulator in regulating the aging network.** *Ageing Res Rev* 2010, **9(Suppl 1)**:S59-S66.
6. Schena M: *Microarray Biochip Technology* Natick: Eaton Publishing; 2000.
7. Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR: **Post-analysis follow-up and validation of microarray experiments.** *Nat Genet* 2002, **32**:509-514.
8. Kuningas M, Mooijaart SP, van Heemst D, Zwaan BJ, Slagboom PE, Westendorp RG: **Genes encoding longevity: from model organisms to humans.** *Aging Cell* 2008, **7**:270-280.
9. de Magalhaes JP, Budovsky A, Lehmann G, Costa J, Li Y, Fraifeld V, Church GM: **The Human Ageing Genomic Resources: online databases and tools for biogerontologists.** *Aging Cell* 2009, **8**:65-72.
10. Pan F, Chiu CH, Pulapura S, Mehan MR, Nunez-Iglesias J, Zhang K, Kamath K, Waterman MS, Finch CE, Zhou XJ: **Gene Aging Nexus: a web**

database and data mining platform for microarray data on aging. *Nucleic Acids Res* 2007, **35**:D756-D759.
11. **Aging Gene Database.** [http://uwaging.org/genesdb/index.php].
12. Zahn JM, Poosala S, Owen AB, Ingram DK, Lustig A, Carter A, Weeraratna AT, Taub DD, Gorospe M, Mazan-Mamczarz K, Lakatta EG, Boheler KR, Xu X, Mattson MP, Falco G, Ko MS, Schlessinger D, Firman J, Kummerfeld SK, Wood WH, Zonderman AB, Kim SK, Becker KG: **AGEMAP: a gene expression database for aging in mice.** *PLoS Genet* 2007, **3**:e201.
13. Tacutu R, Budovsky A, Fraifeld VE: **The NetAge database: a compendium of networks for longevity, age-related diseases and associated processes.** *Biogerontology* 2010, **11**:513-522.
14. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
15. Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun S, Kulp D, Siani-Rose MA: **NetAffx: Affymetrix probesets and annotations.** *Nucleic Acids Res* 2003, **31**:82-86.
16. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**:D140-D144.
17. Wu Z, Irizarry R, Gentleman R, Martinez-Murillo F, Spencer F: **A Model-Based Background Adjustment for Oligonucleotide Expression Arrays.** *Journal of the American Statistical Association* **99**:909.
18. Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z: **CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis.** *Nucleic Acids Res* 2006, **34**:W498-W503.
19. **The R Project for Statistical Computing.** [http://www.r-project.org/].
20. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:Article3.
21. Hackl M, Brunner S, Fortschegger K, Schreiner C, Micutkova L, Muck C, Laschober GT, Lepperdinger G, Sampson N, Berger P, Herndler-Brandstetter D, Wieser M, Kuhnel H, Strasser A, Rinnerthaler M, Breitenbach M, Mildner M, Eckhart L, Tschachler E, Trost A, Bauer JW, Papak C, Trajanoski Z, Scheideler M, Grillari-Voglauer R, Grubeck-Loebenstein B, Jansen-Dürr P, Grillari J: **miR-17, miR-19b, miR-20a, and miR-106a are down-regulated in human aging.** *Aging Cell* 2010, **9**:291-296.
22. Gosling J, Joy B, Steele G, Bracha G: *The Java Language Specification.* 3 edition. Amsterdam: Addison-Wesley; 2005.
23. **The Java EE 5 Tutorial.** [http://download.oracle.com/javaee/5/tutorial/doc/].
24. Maurer M, Molidor R, Sturn A, Hartler J, Hackl H, Stocker G, Prokesch A, Scheideler M, Trajanoski Z: **MARS: microarray analysis, retrieval, and storage system.** *BMC Bioinformatics* 2005, **6**:101.
25. **HomoloGene.** [http://www.ncbi.nlm.nih.gov/homologene].
26. O'Brien KP, Remm M, Sonnhammer EL: **Inparanoid: a comprehensive database of eukaryotic orthologs.** *Nucleic Acids Res* 2005, **33**:D476-D480.
27. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27-30.
28. Laschober GT, Ruli D, Hofer E, Muck C, Carmona-Gutierrez D, Ring J, Hutter E, Ruckenstuhl C, Micutkova L, Brunauer R, Jamnig A, Trimmel D, Herndler-Brandstetter D, Brunner S, Zenzmaier C, Sampson N, Breitenbach M, Frohlich KU, Grubeck-Loebenstein B, Berger P, Wieser M, Grillari-Voglauer R, Thallinger GG, Grillari J, Trajanoski Z, Madeo F, Lepperdinger G, Jansen-Dürr P: **Identification of evolutionarily conserved genetic regulators of cellular aging.** *Aging Cell* 2010, **9**:1084-1097.
29. **ArrayExpress.** [http://www.ebi.ac.uk/arrayexpress/].

doi:10.1186/1471-2164-12-262
Cite this article as: Hofer *et al.*: GiSAO.db: a database for ageing research. *BMC Genomics* 2011 **12**:262.