universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Identification of candidate genes affecting stem cell proliferation by shRNA screening and caloric restriction by meta-analysis

angestrebter akademischer Grad

Magister/Magistra der Naturwissenschaften (Mag. rer.nat.) bzw.
Magister/Magistra der Pharmazie (Mag.pharm.)

Verfasserin / Verfasser:	Michael Plank
Studienrichtung /Studienzweig (lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	Prof. Hofacker

Wien, am 15.11.2010

Identification of candidate genes affecting stem cell proliferation by shRNA screening and caloric restriction by meta-analysis

Michael Plank

November 19, 2010

Contents

Abstract

Despite major efforts the process of aging is one of the least understood phenomena in biology. This work makes use of two important findings in the field of aging research: First of the conclusion that alterations in the proliferation of stem cells might be linked to the aging process, second that caloric restriction is a powerful intervention to extend life-span and delay aging associated diseases. In the first part we analysed a shRNA based screening experiment to identify genes involved in the proliferation of stem cells and undertook first steps towards establishing a flow cytometry based proliferation assay to validate candidates. Secondly we meta-analysed microarray data on different experiments testing gene expression changes associated with caloric restriction. We identified candidate genes enriched for differential expression in the datasets by employing a binomial-test based value counting approach. By including datasets from different organisms, tissues, ages, etc. we aimed at detecting robust and generalizable candidates. We further used different approaches to assign functional categories and common features in terms of their role in signaling networks to the candidate genes. In general the obtained 163 candidate genes and 340 categories overlap with previous findings in the field such as the *Ghr* gene and categories related to lipid metabolism, insulin signaling, collagen or immunity and therefore suggest biological meaningfulness of the approach. On the other hand also novel and so far mainly neglected functions like xenobiotic metabolism, circadian clock, retinol metabolism and copper ion detoxification emerged, that are promising to follow up on in the future. Some of the significant genes might play major roles as regulators of important signaling pathways, as for example *Nfkbia, Airn* (Igf2R antisense RNA) and the notch co-activator *Zfp64*.

Chapter 1

1. The role of stem cells in aging and caloric restriction

1.1 The role of adult stem cells in aging

Many adult tissues as for example the skin, the intestine or the blood require extensive renewal and replacement of cells throughout life time. The source for the generation of new cells is likely to be adult stem cells which could be isolated from various tissues (Watt 2000) (Whitehead et al. 1999) (Weissman 2000). The renewal process is expected to go through committed progenitor cells which themselves further proliferate and differentiate into the required cells. The important property of self-renewal, i.e. the generation of at least one daughter cell identical to the mother cell is however characteristic only for stem cells.

Very early experiments showing that transplanted hematopoietic stem cells (HSCs) could serial repopulate up to 5 mice suggested a extremely long self-renewal capability of stem cells (Siminovitch et al. 1964). Note however that after about the third serial transplant the host HSCs displayed a competitive advantage over the serially passaged donor cells (Ogden & Mickliem 1976).

Other studies proposed the idea of stem cell aging by indicating that stem cells of aged individuals produce less progeny or progeny biased towards proliferation to certain differentiated cell types (Wright et al. 2003). HSCs of aged individuals for example seem to be biased towards the myeloid lineage, while less lymphoid progenitor cells are produced (Rossi et al. 2005). Consistent with decreased numbers or function of HSCs is the well-known increased incidence of anaemia in the elderly (Lipschitz et al. 1981).

Enwere et al. (Enwere et al. 2004) reported decreased olfactory neurogenesis in aged mice. Maslov et al. (Maslov et al. 2004) compared neural stem cell populations in the subventricular zones of the brains of young (2-4 months) and old (24-26 months) mice and detected an about twofold reduction in the older mice. The number of neurospheres recovered in culture from old relative to young animals differed to a similar extent.

Further bone marrow mesenchymal stem cells isolated from older donors show decreased production of progenitor cells and are limited in their differentiation potential. They also have been shown to age in vitro (Baxter et al. 2004).

Evidence if numbers of stem cells decrease with age or not is contradictory for satellite cells (Gibson & E. Schultz 1983) (Conboy et al. 2003) (Brack et al. 2005) and some studies on hematopoietic stem cells even reported an increase in their number (Rossi et al. 2005) (Pearce et al. 2007). However these studies were based on cell surface markers to identify stem cell populations, while a loss of function does become evident e.g. in transplantation assays (Ogden & Mickliem 1976).

One of the most striking experiments in respect to the impact of aging on stem or progenitor cells was done by Conboy et al. (Conboy et al. 2005) showing that circulatory coupling of old and young mice transferred both satellite cells and hepatocytes in the old mouse to a more youthful state with profound changes on their gene expression levels. This suggests that changes occurring with age in theses cells can be reversed by the exposure to one or some serum factors. Note however that findings on satellite cells are not necessarily transferable to all stem cells.

The age associated changes in stem cells may be attributed to accumulating DNA-damage, changes in their niches, telomere shortening, cell senescence e.g. cause by increased p53 activity and / or other reasons (Sharpless & DePinho 2007). Rossi et al. (Rossi et al. 2007) demonstrated loss of functional capacity of hematopoietic cells

in different DNA damage repair defective mouse mutants with age under stress. They further showed that DNA damage accumulates with age in wild-type stem cells. Regarding cellular senescence it is interesting to note that p16INK4a-deficient mice show a significantly lower decline in subventricular zone proliferation, olfactory bulb neurogenesis and the frequency and self-renewal potential of multipotent progenitors. The protein product of p16INK4a is a cycline dependent kinase inhibitor linked to senescence. However no significant changes in this respect were found in progenitor function in the dentate gyrus or enteric nervous system (Molofsky et al. 2006). Further it has been suggested that Bmi-1 prevents the premature senescence of neural stem cells by repressing p16INK4a and p19Arf, a p53 activator (Molofsky et al. 2005). Nonetheless despite a constant expression of Bmi-1 p16INK4a and p19Arf are found to steadily increase in expression throughout life (Bruggeman et al. 2005) (Molofsky et al. 2006). In another study it was found that deletion of the cell cycle inhibitor p21, which gets activated by telomere shortening, can prolong the life-span of telomerase deficient mice. At the same time the proliferation of intestinal progenitor cells and repopulation capacity and self-renewal of hematopoietic stem cells was restored (Choudhury et al. 2007).

With respect to replicative senescence it is interesting that for mice expressing an active form of p53 and showing a premature aging phenotype it has been proposed that this is caused by replicative senescence of stem cells (de Magalhães & Faragher 2008) (Tyner et al. 2002). Similarly Halaschek-Wiener and Brooks-Wilson (Halaschek-Wiener & Brooks-Wilson 2007) argue for a role of stem cell exhaustion in Hutchinson-Gilford progeria (HGP), one of the most severe premature aging disorders. Possibly consistent with this idea may be the growth retardation of HGP patients in their first years of life (Cox & Faragher 2007). Similarly for two other important premature aging syndroms, Cockayne and Werner syndrom, this retardation is also found in early life and puberty respectively (Henning et al. 1995) (Martin & Oshima 2000). However cellular senescence in these diseases is probably not limited to stem cells and stem cell exhaustion in HGP might also be driven by increased apoptosis. Therefore it might rather be the inability of stem cells to ensure tissue homeostasis due to increased senescence and apoptosis of other cells, than specific alterations in the stem cells themselves.

In summary, even though it is not clear if or for which stem cell types there is a decrease in their amount with age, there is growing evidence for functional changes in these cells. The term "stem cell hypothesis of aging" has been coined and also tries to explain age associated conditions like atherosclerosis, type 2 diabetes and frailty (Sharpless & DePinho 2007).

1.2 The influence of caloric restriction on stem cells

Since caloric restriction (CR) is a powerful intervention to extend life-span and delay aging associated diseases in a wide range of organisms (see "3.1.1 The potential of caloric restriction to delay aging") it is obvious to assume an influence of CR on stem cells if you accept the stem cell hypothesis of aging. However not many studies have been conducted in this direction so far.
One of the few was done by Kumar et al. (Kumar et al. 2009) reporting a significant increase in the proliferation rate of neuronal progenitor cells in the brain of caloric restricted rats. Another study demonstrated that lowering glucose concentrations in the medium for culturing mesenchymal stem cells lowered apoptosis and increased the proliferation rate as well as the number and size of fibroblastic colonies in the colony-forming unit assay (Stolzing et al. 2006). Interestingly studies by Yoshida et al. (Yoshida et al. 2006) and Schmuck et al. (Schmuck et al. 2010) described a decrease in hematopoietic progenitor cells and adipose tissue derived mesenchymal stem cells respectively with CR in vivo.

Therefore, even though effects of caloric restriction on adult stem cells have been observed the nature of its influence is but poorly understood. In this work we addressed both the underlying genetic mechanisms of stem cell proliferation and CR by two different approaches. In the next chapter we present a shRNA library screening approach to identify key players involved in stem cell proliferation and first attempts towards confirming promising candidates. Since cell culture work with adult stem cells is not well established we employed an embryonic stem cell line for these experiments. Even though results obtained on this system still have to be tested for their applicability in adult stem cells the important self-renewal capability is common between both embryonic and adult stem cells and shared underlying mechanisms are expected.
In chapter 3, which accounts for the major part of this work, we meta-analysed existing gene expression data to determine genes altered in their expression due to CR.
The two parts therefore start off from two different sides, one experimentally addressing the stem cell hypothesis of aging, the other computationally exploring the life-span extending effect of caloric restriction. However both demonstrate data-driven approaches to increase the knowledge and generate hypotheses about the riddle of aging.

Chapter 2

Determining genes implicated in stem cell proliferation

2.1 Finding candidate genes responsive to oxidative stress and associated to proliferation of stem cells by shRNA library screening

The following screening experiment and preliminary analysis were performed by J.P. de Magalhaes and G. Jansens and are only described in brief here.

2.1.1 Experimental design

To find candidate genes which are involved in the proliferation or ability of embryonic stem cells to survive under oxidative stress the following experiment was performed in our group: 6 replicates of cells of the mouse embryonic stem cell line CCE were virally transfected by adding a mixture of lentiviruses containing DNA representing a part of the Hannon-Elledge shRNA whole-genome library (6144 shRNAs) (Chang et al. 2006). Since it contained more than one shRNA per gene, around 2000 to 3000 genes were targeted. The genes targeted by this so called "focus library" were chosen with focus on cancer research (i.e. targeting genes involved in signaling, cell cycle, etc., as retrieved from gene ontology (GO) databases, and such genes where a phenotype was expected from their knock-down). The shRNA sequences were predicted computationally and most had not yet been validated experimentally. The mixture of plasmids containing these different shRNAs was obtained from S. Elledge. Viruses were produced as described in "2.2.2 Materials and Methods", but with this complex mixture of plasmids instead of one single type of plasmid. The transfection was done as described in 2.2.2.

1 week after the transfection DNA was isolated from an aliquot of the cells while the rest of them were kept in culture. PCR with limited cycle number was performed on the isolated DNA using primers binding to the flanking regions of the shRNA encoding DNA and expected to yield amplification products of the different shRNA encoding sequences (in the following also simply called "shRNA sequences" or "shRNA genes") in proportion to the amount this sequence was present in the population. Cy3 was incorporated during the PCR so that the product was labelled with green fluorescent dye. By culturing the cells for 1 week before the start of the assay it was expected that cells rendered in-viable by the effect of a shRNA were already largely diminished and shRNAs found in the following assay were indeed affecting proliferation rate rather than cell survival. 3 of the replicates were cultured as described in "2.2.2 Materials and Methods" (control), the other 3 were subjected to oxidative stress by addition of hydrogen peroxide. After 2 weeks DNA was extracted and PCR performed as above, but using Cy5 instead of Cy3 for red fluorescent labelling of the PCR product.
A microarray experiment was performed, adding the PCR products from the beginning of the experiment and from after 2 weeks to a custom made spotted cDNA microarray platform, containing two probes per shRNA (strictly speaking one of them is a concatenation of twice the same sequence as the other, however they are referred to as "identical probes" in the following) in the library.
The green and red signal were detected and $\ln(\frac{E_r}{E_g})$ (in the following also called "ln-ratio") calculated, where E is the signal of emission (g in green and r in red).
The logic of this experiment was that the ratio of shRNAs knocking-down genes that have a positive effect on

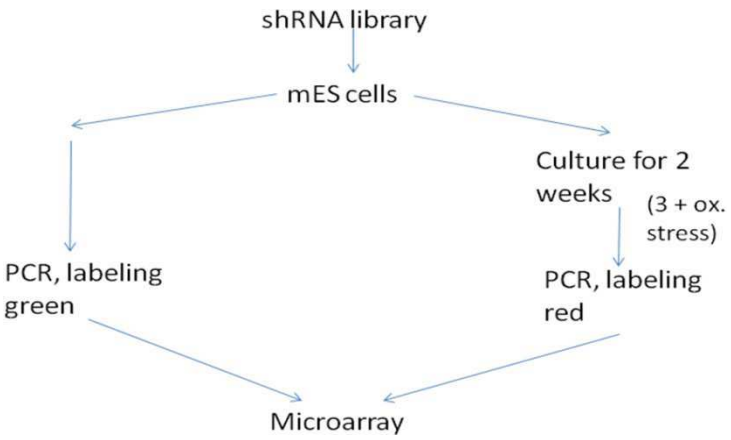


Figure 2.1: Outline of the screening experiment to find genes associated with stem cell proliferation and handling of oxidative stress; "3 + ox. stress": 3 of the 6 samples were subjected to oxidative stress; see text for details

proliferation will diminish due to this effect, while of shRNAs knocking down genes with a negative effect on proliferation will increase. Therefore the effect of the shRNA controls in which amount this shRNA will be present in the population after two weeks.
For genes involved in oxidative stress however the ratio of shRNA after two weeks to shRNA in the beginning will be different between stressed samples and controls. An outline of the experiment is shown in fig. 2.1.

2.1.2 Preliminary analysis

For a preliminary analysis to find genes differentially expressed due to different survival under oxidative stress the average value of $\ln(\frac{E_r}{E_g})$ over three controls was subtracted from the average for the stressed samples. Genes corresponding to probes with high values were assumed to have a negative, with low values to have a positive effect on stress resistance.
E.g. assuming the ln-ratio is positive for stressed and unstressed cells, but higher for stressed, the difference is positive since the ratio of the shRNA increased more in the stressed cells. I.e. cells survive oxidative stress better when the corresponding gene is knocked-down and the gene is therefore assumed to have a negative effect on stress resistance. To determine genes for which knock-down had either a beneficial or detrimental effect on proliferation [1](independent of oxidative stress) the mean value of the ln-ratio was calculated. Genes corresponding to high values indicated a negative, to a low value a positive effect on proliferation.

Since results appeared to be much clearer for the testing of proliferation than oxidative stress it was decided to test the following candidates for their effect on proliferation: *Wnk2, Map3k13* and *Dr1* for which shRNAs were enriched in the screen and *Psma1, Zfp828, Tcf23* and *Pak1* for which shRNAs were depleted in the screen.

2.1.3 Previous attempts to experimentally validate candidates

The following approach was used by G. Jansens to test the effect of these candidates on stem cell proliferation: CCE cells were transfected with the plasmid pHAGE containing the sequence of a candidate or control shRNA as described in "2.2.2 Materials and Methods". In control lines the shRNA targets the firefly gene (FFL) which is not present in mouse. Cells were plated at equal concentrations and allowed to grow for 4 days without splitting. (Splitting (subculturing) is avoided in these proliferation experiments since it is considered a source of variation). Then a single cell suspension was obtained by trypsinization and cells counted using an automatic cell counter (Casey). The experiment was repeated with a growth period of 3 instead of 4 days. It was calculated which percentage of the initial cell number was present after 3 or 4 days respectively. During this period the expected red fluorescence from turboRFP encoded on pHAGE was found in all cell lines expect for the ones where pHAGE contained the shRNA targeting *Oct4* or *Psma1*. This suggests that these tow lines are either outgrown by

[1]To be precise at this point we cannot distinguish if the value was e.g. lower due to a prolonged cell cycle time, due to a lower survival rate or another cause. Therefore we define proliferation here as what is measured, when comparing the number of cells generated after a certain time to a starting number of cells.

untransfected cells which are left in the population or they silence the transcript for the shRNA and turboRFP. As a result no significant difference in proliferation was found between the control FFL-cell line and any of the other lines. Also the tendency for many lines was not consistent between experiments and often not consistent with the prediction from the screen.
Therefore we decided on two ways to improve finding candidates truly involved in cell proliferation:

1. Improving the candidate selection by a more sophisticated analysis of the screening data to find candidates more likely to be linked to proliferation

2. Improving the method for validating candidates: The problem so far was that for meaningful results the cells still have to be in their exponential proliferation phase when counted. Splitting the cells during this procedure would however disturb the analysis since it can only be done with limited accuracy (i.e. the number of cells dieing during trypsinization may vary). Therefore if subculturing was to be avoided, cells could not be allowed to proliferate longer than 3 or 4 days even though a longer proliferation time would most likely lead to more significant results if cells could be kept in exponential growth. Therefore we decided to do an assay where shRNA lines are mixed with wild-type (wt) cells as an internal standard and monitor their ratio over a longer time. When having an internal standard the matter of inaccurate splitting is not expected to be a problem any more since the error appears to the same extent for both lines. This approach will be described in "2.2 Experimental validation of candidate genes by proliferation assays"

2.1.4 Statistical procedure

As explained the main criterion for selecting candidates implicated in stem cell proliferation or handling of oxidative stress should be the difference found in the microarray experiment in the beginning to end ratio or ratio between stressed and non-stressed samples of DNA coding for the shRNA targeting a particular gene. Further criteria were the association of a gene to gene ontology (GO) terms considering these GO terms´ enrichment among potential candidates and potential role in proliferation.
Because of the large number of genes tested compared to the small number of replicates we decided not to use a t-test for the analysis of differential detection of PCR product between beginning and end of the experiment: Considering the number of genes chances are high that for some genes values measured for the amount at the beginning are very close together as well as for the ones in the end by random chance. This would suggest high statistical significance even if there is only a small difference between the means of beginning and end and might therefore lead to false positives with no true difference between the means of the population.
Instead we preferred an analysis that for each probe counts the number of times the $\ln(\frac{E_r}{E_g})$ exceeds a certain positive or negative threshold and obtains the probability that this or a higher number would be found by chance. Therefore in contrast to the t-test this test is based on a fold-change criterion. The false discovery rate (FDR) for all probes is then estimated using a scrambling approach. A disadvantage of this method compared to a combination of a t-test and an effect size (fold-change) cutoff is that we do not account for the dispersion of measured values, i.e. if there is a high or low variation. An advantage is the insensitivity of this test to outliers compared to a t-test (since no mean values are calculated).

2.1.5 Finding genes associated with stem cell proliferation

In a first step we concentrated on finding shRNAs over- / underrepresented after two weeks ignoring the fact that some samples were under oxidative stress and the others were not. This is supposed to detect candidates for genes associated with stem cell proliferation as detailed above. Later we used further information like functional categories associated with the genes or their role in the network of candidate genes to select the candidates for experimental testing.

2.1.5.1 Finding genes over-/ underrepresented after 2 weeks

2.1.5.1.1. Excluding low-signal data and annotation The starting point for this analysis were background subtracted normalized intensities from the two color microarrays.
To remove data for which no sufficient amount of shRNA coding DNA integrated into cellular genomes, for which the PCR product did not bind with sufficient affinity to the probe or for which the signal at t = 0 was consistently low for other reasons we removed probes for which the signal of the green channel (in the following: "green signal") was $\leq$ 200 (arbitrary units) in at least 3 of 6 replicates. (The 6 microarrays are considered

"replicates" in this approach even though the samples on 3 of them were exposed to stress and of 3 were not; the maximum value for the green signal was around 295 000, the median around 1300.)

The program over200_annot.pl (supplement1) extracts those probes from a file (all_arrays.txt in supplement1) for which the signal is > 200 for at least 4 of the 6 signals at the beginning of the experiment (green channel). After this selection 8845 of the original 12 288 probes were left. For these probes the gene symbol, gene name, NCBI Entrez Gene ID and NCBI accession number is added from another file (Mm.ALL.bc.txt in supplement1) by the same program. The file matching these annotations to the probe names was downloaded from Codex (http://cancan.cshl.edu/cgi-bin/Codex/Codex.cgi) earlier, but the download was not available any more at the time of this analysis. Annotations for some of the probes could not be found in the mentioned file. Therefore the probe names not found were uploaded to the old version of codex[2] (Aug 2009), which in contrast to the new version allows searches for probe names. [3] Annotation was obtained and added to the probes for which it was not found before. The 24 probes for which annotation still could not be found were discarded from the analysis.
Probes matching more than one shRNA sequence were removed from the analysis since we wanted to avoid obtaining candidates for which the measured expression value was actually caused by another shRNA. The number of probes excluded during this procedure was 214.

2.1.5.1.2. Collapsing probes targeting the same shRNA Since there were two probes per shRNA on the microarray (prefixes: HH_ and mmFocus_) the two (if both passed the intensity threshold) were collapsed by calculating the mean for each replicate. This is done by collapse_two-probes.pl (supplement1).
In the next step the file was converted to a .xls and mean value and standard deviation (STDEV) for the $\ln(\frac{E_r}{E_g})$ of each experiment over all probes calculated by the corresponding Excel functions. (Means were -0.09 to -0.04, standard deviations 0.98 to 1.16.)
Even though there were different shRNAs targeting the same gene for some genes, these were not collapsed since different shRNAs were expected to perform differently. Collapsing might therefore obscure the effect of the better shRNA by averaging with values of the worse.

The file containing probes selected by the signal intensity criteria mentioned above, annotated and collapsed can be found in supplement1: two-col.txt.

2.1.5.1.3. Finding shRNAs over- and underrepresented after 2 weeks A shRNA gene was termed overrepresented if the $\ln(\frac{E_r}{E_g})$ was above a certain threshold for a certain number of replicates and underrepresented if this number of replicates was below a certain threshold. (If in the following we talk about gene X being over- / underrepresented this means the shRNA targeting this gene was over- / underrepresented.) As threshold for each replicate mean + standard deviation (STDEV) over all probes and mean − STDEV respectively were chosen. In different runs those probes for which (at least) 4, 5 or 6 of 6 values for $\ln(\frac{E_r}{E_g})$ were above / below the mentioned thresholds were selected by the program mult_aboveSTDEV.pl (supplement1). The occurrences of the number of different probes for the same gene were counted with probes_per_gene.pl (supplement1).

2.1.5.1.4. Estimation of p-values and false discovery rate The probability p to find any probe above / below mean +/- STDEV was calculated by dividing the mean number of probes found per sample by the number of probes tested (p = 0.13 for any probe found above mean +/- STDEV, p = 0.14 for below mean +/- STDEV). The probability P to find a probe at least 4, 5 or 6 times respectively above / below mean +/- STDEV (called "4of6", "5of6" and "6of6" criterion) by random chance was calculated using the binomial distribution:

$$P = 1 - \sum_{x=0}^{k-1} \binom{n}{x} * p^x * (1-p)^{(n-x)} \tag{2.1}$$

with

p= probability to find a random gene above / below mean +/- STDEV (see above),

k= 4, 5 or 6 respectively,

[2] http://katahdin.cshl.org:9331/rnai/repository/scripts/newmain.pl
[3] By the time of writing the old codex is not online anymore. Therefore the file obtained for the not found probes is attached in supplement1: codex_found.txt

	# candidates	P-value	FDR
overrep.: 4of6	117	3.84E-03	0.158
overrep.: 5of6	23	2.29E-04	0.050
overrep.: 6of6	6	5.76E-06	0.005
underrep.: 4of6	216	4.95E-03	0.100
underrep.: 5of6	60	3.18E-04	0.024
underrep.: 6of6	10	8.62E-06	0.003

Table 2.1: Number, P-values and FDRs of candidate shRNAs found over- or underrepresented after 2 weeks at different criteria. P-values were calculated using the binomial distribution and FDRs by comparing the found to the expected number of candidates.

n= 6.

P corresponds to the P-value for finding a probe at the given criterion. Multiplying this probability with the number of probes in the assay (giving the number expected to be found for this criterion by chance) and dividing it by the found number for each criterion gives the FDR. The number and P-values for probes found at each condition and corresponding FDRs are shown in table 2.1. The number of over- or underrepresented shRNA candidates closely resembles the number of candidate target genes, since only very few genes (7 for the 4of6 overrepresented, 8 for 4of6 underrepresented, 1 for 5of6 over- and underrepresented each and 0 for the others) met the criteria with more than one shRNA.

Since we aimed at a FDR <0.05 the 5of6 criterion appears to be the appropriate one to chose the candidates to experimentally validate.

2.1.5.2 Relationship of proliferation associated candidate genes to aging

The initial idea of finding genes involved in stem cell proliferation or stress response was motivated by finding genes involved in aging (see "1 The role of stem cells in aging and caloric restriction"). Therefore we tested if our candidates could be found in GenAge (http://genomics.senescence.info) (de Magalhães & Toussaint 2004), a database of genes associated with human longevity or that modulate aging in model organisms. A list of all those genes and their human homologs was downloaded and is_gene_in_list_mod_caseinsens.pl (supplement 1) was used to search for our candidates selected by the 5of6 criterion in this list. Since mouse homologues were not available, we made use of the rule of thumb that the mouse homologue of a human gene annotated as XXX11 would be Xxx11 and the identifiers would therefore be equal in a case-insensitive search. We are aware that this might miss genes in a few special cases.
There was no overlap found between our candidates with the 5of6 criteria and the genes listed in GenAge.

2.1.5.3 Functional analysis: Finding over- / underrepresented functional categories

We employed and compared different ways to find functional categories common to shRNAs associated with stem cell proliferation. One analysis was done using a binomial test employing custom made Perl code, the others were based on the freely available GSEA and DAVID tools.

2.1.5.3.1 Finding enriched GO-categories by a binomial test The first functional analysis was done by searching for gene ontology (GO) terms that were represented significantly higher among over- / underrepresented genes than expected by chance.
GO analysis was done on shRNA genes detected to be over- or underrepresented (in the following called "over- or underrepresented genes") by a method similar to the one described above. However to avoid counting genes represented by two shRNAs twice collapsing was done using combine-select-highest_withTest.pl (supplement 1). This program first collapses signals corresponding to identical shRNAs by calculating the mean, then selects of shRNAs targeting the same gene only the shRNA with the average $\ln(\frac{E_r}{E_g})$ over all replicates which is furthest from 0 (i.e. the shRNA that is most over- / underrepresented). This is because the silencing effect differs from shRNA to shRNA and this approach selects the one with the most marked effect. The program prints warnings if the mean ln-ratio of one shRNA is strongly in the other direction than another shRNA for the same gene. Specifically if for a gene the average $\ln(\frac{E_r}{E_g})$ for a shRNAs is above mean+STDEV and for another it is below

mean-STDEV or vice versa a warning is printed. After manual inspection all probes with warnings were removed. Starting from this file over- and underrepresented genes were determined as above ("2.1.5.1 Finding shRNAs over- and underrepresented after 2 weeks") for the same criteria as described above.

To add GO categories to the corresponding gene a list mapping GO identifiers to all genes was downloaded from NCBI[4] (25/08/2009) and all non-mouse genes were discarded. Since in this file each gene was repeatedly listed for each GO identifier a new file was created with one gene and all its GO identifiers per row. All GO identifiers were added to the list of probes for over- and for underrepresented genes. A small number (10 for over-, 34 for underrepresented) of probes could not be found in the GO-list (and also not searching the database by hand).
It was counted how many overrepresented and how many underrepresented genes were found for each GO identifier and how many for the complete list of all genes after collapsing. Only GO identifiers with at least 3 corresponding genes over- / underrepresented were used for further analysis. These steps were performed by GO_masterprog.pl (supplement 1).

The probability P that an equal or higher number of genes than the actual is found over- or underrepresented for a GO identifier by chance was calculated using a binomial test:

$$P = 1 - \sum_{x=0}^{k-1} \binom{n}{x} * p^x * (1-p)^{(n-x)}$$

where

k is the number of times a GO identifier was found associated with the over-/underrepresented genes,

n is the number of times the GO identifier was found associated with all genes and

p the probability that GO identifiers are found over-/underrepresented.

Therefore p is calculated by dividing the sum of the number of times all GO identifiers are found associated with over- / underrepresented genes by the sum of the number of times they are found associated with all genes after collapsing.
The GO terms were added to the corresponding GO identifiers by using addGO_terms_mult-files.pl (in supplement1).

To assess the significance of the found GO terms and find an appropriate cutoff for P considering multiple hypothesis testing we scrambled the ln-ratios of each replicate with respect to each other replicate manually after we had filtered out low intensity data. The analysis was repeated as with the unscrambled files. Different cutoff values for P were tested to find reasonably low FDRs (FDR is the number of GO identifiers found significant at the chosen P on scrambled divided by the number on actual data; FDR_calc2_over-undercount-in1file.pl in supplement1). Since we scrambled only once the FDR is a rough estimate. The GO identifiers and terms for the 4of6 criterion at the P-value of 0.005 (FDR = 0.08 and 0.06 for over- and underrepresented genes respectively) are shown in table 2.2. Note that some of the GO terms appear for both over- and underrepresented genes. This may biologically make sense depending on which genes of the GO terms are represented in each and how they interact with each other.

2.1.5.3.2. Using GSEA to find enriched gene sets

2.1.5.3.2.1. Introduction to GSEA GSEA (Gene Set Enrichment Analysis) is a program that evaluates microarray data at the level of gene sets. It is freely available at http://www.broadinstitute.org/gsea. The goal of GSEA is to determine whether members of a gene set S tend to occur toward the top (or bottom) of a dataset ranked in a certain way, in our case by ln-ratio. Gene sets are defined based on prior biological knowledge, e.g. genes encoding products in the same metabolic pathway, located in the same cytogenetic band, or sharing the same GO category. A variety of gene sets to test for can be found at the Molecular Signature Database (MSigDB).
The GSEA algorithm comprises the three following steps:

[4] ftp://ftp.ncbi.nih.gov/gene/DATA/gene2go.gz

overrep.		underrep.	
GO:0000287	magnesium ion binding	GO:0000287	magnesium ion binding
GO:0003674	molecular function	GO:0001843	neural tube closure
GO:0003676	nucleic acid binding	GO:0003676	nucleic acid binding
GO:0004672	DNA binding	GO:0003677	DNA binding
GO:0004674	protein kinase activity	GO:0003700	transcription factor activity
GO:0004713	protein serine/threonine kinase activity	GO:0003713	transcription coactivator activity
GO:0004721	protein tyrosine kinase activity	GO:0004842	ubiquitin-protein ligase activity
GO:0004725	phosphoprotein phosphatase activity	GO:0005515	protein binding
GO:0005509	protein tyrosine phosphatase activity	GO:0005622	intracellular
GO:0005515	calcium ion binding	GO:0005634	nucleus
GO:0005524	protein binding	GO:0005829	cytosol
GO:0005737	ATP binding	GO:0005839	proteasome core complex
GO:0005739	cytoplasm	GO:0006350	transcription
GO:0005794	Golgi apparatus	GO:0008270	zinc ion binding
GO:0006468	protein amino acid phosphorylation	GO:0045449	regulation of transcription
GO:0006810	transport	GO:0046872	metal ion binding
GO:0006915	apoptosis	GO:0051603	proteolysis involved in cellular protein catabolic process
GO:0007165	signal transduction		
GO:0007243	protein kinase cascade		
GO:0007275	multicellular organismal development		
GO:0007399	nervous system development		
GO:0016301	kinase activity		
GO:0016740	transferase activity		
GO:0030145	manganese ion binding		
GO:0030154	cell differentiation		

Table 2.2: GO-identifiers and terms enriched after two weeks for over-/underrepresentation at FDR$<$ 0.08 and 0.06 respectively.

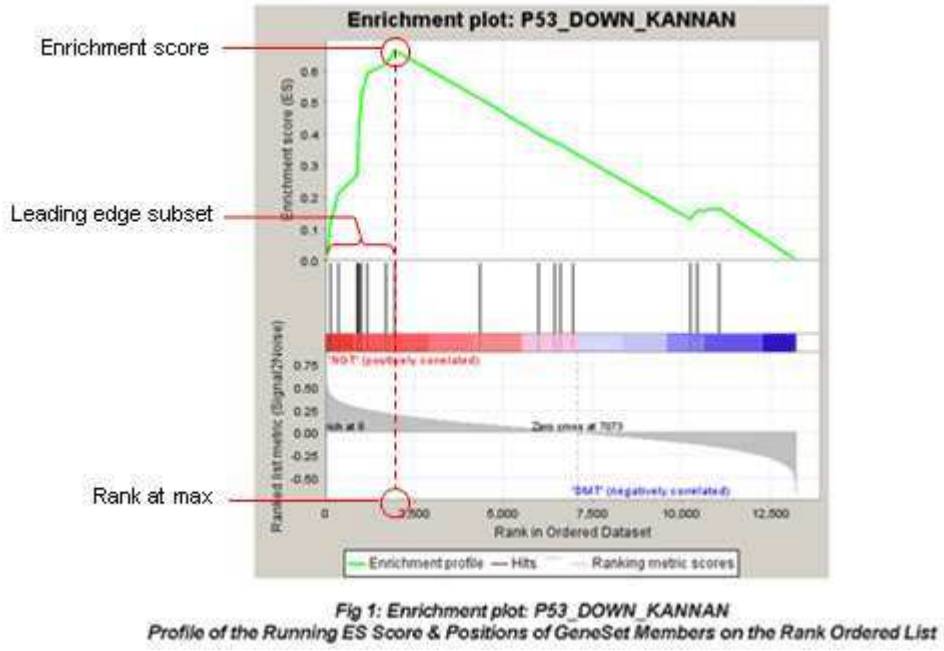


Fig 1: Enrichment plot: P53_DOWN_KANNAN
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

Figure 2.2: Example of the running sum method used by GSEA; bottom: ranking of genes according to signal-to-noise ratio (in our case ln-ratio); middle: genes in the tested gene set are shown by vertical lines at their rank position; top: running sum; Enrichment score is the maximum deviation from 0; picture from http://www.broadinstitute.org/gsea.

1. An enrichment score is calculated by walking down the ranked list of genes, increasing a running-sum statistic when a gene of a gene set is encountered and decreasing it when encountering genes not in the gene set. The enrichment score is the maximum deviation from zero found in the random walk. See fig. 2.2.

2. A P-value is estimated by comparing the enrichment score to an enrichment score calculated from a permutation of the ranked list of genes.

3. Since normally more than one gene set is tested multiple hypothesis testing is done. For this the enrichment score is normalized by dividing by the number of genes in the given gene set and a false discovery rate is estimated by comparing the normalized enrichment score to normalized enrichment scores calculated from a permuted list of genes (Subramanian et al. 2005).

2.1.5.3.2.2. GSEA to find gene sets enriched in proliferation associated genes We started from a file where probes for each gene were collapsed to the probe with the mean of $\ln(\frac{E_r}{E_g})$ furthest from 0 as described above. To create a ranked gene list the means of the ln-ratio over the 6 replicates for each gene were calculated and written in a tab-delimited table with the corresponding gene symbols. For GSEA to be able to recognize the gene symbols all letters had to be changed to capitals. For the resulting file the .txt extension was changed to .rnk.
As gene sets we downloaded msigdb.v2.5.symbols.gmt from MSigDB (Subramanian et al. 2005) which comprised all available gene sets (7/9/09).

The ranked gene list .rnk file and the gene sets were loaded into the GSEA desktop application and the analysis was run using the GseaPreranked tool. The "collapse dataset to gene symbols" option was set to false, otherwise default settings were used.
As result we found no gene set enriched for underrepresented (enrichment score < 0) genes below a FDR of 0.05. For overrepresented genes (enrichment score > 0) we found 5 sets for FDR $<$0.05: PHOSPHORIC_ESTER_HYDROLASE_ACTIVITY, PROTEIN_AMINO_ACID_DEPHOSPHORYLATION, PHOSPHOPROTEIN_PHOSPHATASE_ACTIVITY, DEPHOSPHORYLATION and KERATINOCYTEPATHWAY).

	User Genes	Genome
In Pathway	3-1	40
Not In Pathway	297	29960

Figure 2.3: Example for a contingency table created by the DAVID Functional annotation tool; from the DAVID Introduction file (http://david.abcc.ncifcrf.gov/helps/functional_annotation.html#EXP2)

2.1.5.3.3. Using DAVID to find enriched biological themes and pathways

2.1.5.2.3.1. Introduction to DAVID The Functional Annotation Tool of DAVID (Database for Annotation, Visualization and Integrated Discovery) is based on a procedure similar to Fisher´s exact test. A 2x2 contingency table containing how many of the genes of interest and how many of the given background (genome) associate with a functional term (or pathway) and how many do not is created (see fig. 2.3). To be conservative 1 is subtracted from the number of genes of interest associated with the term. The probability of a number of at least this many genes associated with the category given the marginal distribution is calculated.

Functional terms here do not only include GO terms, but are also based on protein–protein interactions, protein functional domains, disease associations, biological pathways, sequence features, homology, gene functional summaries, gene tissue expression and literature. The annotation categories can be flexibly included or excluded from the analysis by the user.

2.5.1.3.3.2. DAVID to find enriched biological themes and pathways We made use of the Database for Annotation, Visualization and Integrated Discovery (DAVID) to find enriched biological themes and pathways in our candidates for proliferation associated genes. In particular we used the Functional Annotation algorithm accessible at http://david.abcc.ncifcrf.gov/summary.jsp.
We separately uploaded the overrepresented and underrepresented candidates for the 4of6 criterion. As a background for the analysis we loaded all genes represented on the microarray. We ran the program and obtained the Functional Annotation Clusters once for the default themes and once by selecting all pathway options only.
For overrepresented genes we obtained categories related to phosphate, ATP and phosphorylation, for underexpressed the proteasome below a FDR of 5% when searching for default categories. Searching for pathways MAPK signaling was found for overrepresented genes below a FDR of 5%, the proteasome again for underrepresented.

2.5.1.3.4. Comparison of results from GO analysis, GSEA and DAVID While the numbers of significant categories found with GSEA and DAVID are of comparable size the GO-terms found by the binomial analysis is clearly higher. This might partially be due to the slightly more relaxed FDR-cutoff (0.08 and 0.06) used owing to the discrete nature of the cutoff thresholds (4of6, 5of6 or 6of6), but is most likely due of inherent differences between the methods.

The results of the GSEA and DAVID analysis both emphasize the role of phosphate in signalling for overrepresented genes, however, while in GSEA mainly results in terms related to dephosphorylation, DAVID detects phosphorylation. The binomial analysis detects among others categories related to both phosphorylation and dephosphorylation. The most concrete pathway, MAPK signaling, was discovered by DAVID analysis focusing on pathways only.
While GSEA gives no significant category for underrepresented genes DAVID only detects the proteasome at a FDR < 0.05. Again among other categories the binomial test also lists the "proteasome core complex", "proteolysis involved in cellular protein catabolic process" and "ubiquitin-protein ligase activity".

2.1.5.4 Mapping candidate genes to the STRING network

STRING is a database of physical and functional protein interactions and can be employed to build a network from a gene list based on this information. We used STRING 8.3 at default settings on a combined list of genes over- or underrepresented at the 4of6 criterion. See supplement 1 for a figure of the network. It can clearly be seen that while many proteins are not or weakly connected there are two distinct dense parts of the network, one built around *Tcf4, Pparg* and including edges to *Hdac2* and *Hdac3* and another around *Psma1* and *Psma5*, strongly linked to *Pak1*. We assumed that a high degree of a gene in the network represents further evidence for the importance of this gene in mechanisms related to stem cell proliferation.

2.1.5.5 Decision on which candidates to test experimentally

Since only about 10 candidates could be experimentally tested for the effect of their knock-down on the proliferation rate the most promising ones had to be chosen. We first demanded that the candidates were over- / underrepresented at the 5of6 criterion (FDR <0.05) giving lists of 23 and 60 genes respectively. For the further decision we took into account if a gene was also significant at the 6of6 criterion or significant at the 5of6 criterion with more than one probe, if it was associated with "meaningful" functional categories, especially if they were enriched in the functional analyses and if the gene was highly connected in the network of the candidate genes. As a meaningful GO-category we understand one that describes a distinct cellular process, not a function that can be found in many different pathways. Enriched meaningful functional categories were "cell differentiation", "apoptosis" or such related to proteasome function. If the category was not enriched we required that a possible link between the category and proliferation existed as for example for the GO-term "positive regulation of cell proliferation" or categories related to the cell cyclus, etc. Therefore at this point we departed from a purely data-driven candidate selection approach.
For overrepresented candidates we selected *Rnf31, Pkn2, Map4k5, Csnk1a1* and *Ppp3r2* since they all fulfilled the 6of6 criterion, *Clk1* because it was found significant by two probes and *Map3k1* for its central role in the network (6 connections) and its functional association with "apoptotic mitochondrial changes".
For candidates for which the shRNA was underrepresented after 2 weeks we chose *Edd1, Hdac3, Phf17, Sqstm1, Mbd2* and *Zxda* since they all were significant at the 6of6 criterion and associated with meaningful functional categories. *Psma5* was chosen, because it was found significant for two probes, for its role in proteasome function and high degree (7 connections) in the network.
We made sure not to select genes that had already been selected in the preliminary analysis ("2.1.2 Preliminary analysis") and for which plasmids had already been obtained. Interestingly only few of the candidates selected there appeared also promising in this procedure. *Wnk2* was detected at the 6of6 criterion for overrepresented, *Tcf23* and *Pak1* for underrepresented shRNAs. *Pak1* appeared to be a good candidate also in this approach due to its high degree in the network.

2.1.5.6 Determining the expression of candidate genes in early embryonic stages and stem cell lines

In a last step we checked the expression of the selected candidates in early embryonic stages and stem cell lines according to public datasets to assess if their knock-down could be the reason for slower growth of these cells or if the gene of interest is not even expressed in stem cells. Note that our original microarray screen did not test the expression of the shRNA target genes but only the level of shRNAs. Changes in their amount could also be random or due to off-target effects. In a first step we tested the expression in the Theiler Stage 4 (TS4) (Blastocyst, Inner cell mass apparent, 2-4 days post coitum (dpc)) and TS5 (Blastocyst (zona-free), 3-5.5 dpc) embryonic stages according to the Mouse Genome Informatics website (http://www.informatics.jax.org/expression.shtml) (Bult et al. 2008) (Smith et al. 2007). In the next we checked the number of expressed sequence tags (ESTs) at the Unigene website (http://www.ncbi.nlm.nih.gov/unigene) (Pontius, Wagner, Schuler 2003) for our genes in the blastocyst stage and if not found there in the morula and other embryonic tissues. We also checked the candidate list for their expression values in the microarray datasets GDS2666 and GDS2667, GDS2668 and GDS2669 as well as GDS2905 and GDS2906 at the Gene Expression Omnibus (GEO). GDS2666 and GDS2667 (Hailesellasse et al. 2007) compare the gene expression in cells of the embryonic stem cell line R1 at different time points towards differentiation to embryoid bodies, GDS2668 and GDS2669 do the same for line J1 (Hailesellasse et al. 2007). GDS2905 and GDS2906 compare gene expression in J1 stem cells and embryoid bodies.
If the expression of a gene (more precisely: its percentile rank within the sample) was at a low level for t = 0 / for undifferentiated cells and the level at other time points / in the embryoid body were clearly higher we considered this gene as not expressed in stem cell lines, if it was at background level for most of the time points / also for the embryoid body we did not directly assumed this gene not expressed in embryonic stem cells without further hints from other analyses.

Expression information for none of the genes in our narrower candidate list except for *Phf17* was found at the Moue Genome Informatics website. *Phf17* was indicated to be expressed at TS4. The results for the other expression analyses (from Unigene and GEO) are shown in table 2.3.

For all genes except *Ppp3r2* and *Zxda* there was at least one evidence of expression in embryonic stem cells, either by ESTs or microarray data. Even though the data do not prove that *Ppp3r2* and *Zxda* are not expressed in stem cells we excluded these genes from the list of our candidates since none of our analyses gave evidence for

	Unigene: Transcripts per million in blastocyst	further Unigene results		GDS2666 and GDS2667(line R1)	GDS2668 and GDS2669 (line J1)	GDS2905 and GDS2906(line J1)
Csnk1a1	71			>75%	>75%	>75%
Map4k5	486			>25%	>25%	>25%
Pkn2	100			>75%	>75%	>75%
Ppp3r2	0	0 in embryonic tissue		most low	most low	most low
Rnf31	0	13 in embryonic tissue; 3 in cleavage stage or morula		>75%	>75%	most low
Clk1	28			>75%	>75%	>75%
Sqstm1	271			>75%	>75%	>75%
Psma5	142			>75%	>75%	>75%
Phf17	185			>75%	>75%	>75%
Mbd2	14	human: 0		undifferentiated low	undifferentiated low	undifferentiated low
Edd1	no unigene entry			>75%	>75%	>75%
Hdac3	14			>75%	>75%	>75%
Map3k1	0	13 in embryonic tissue; 0 in cleavage stage or morula		>50%	>25%	>50%
Zxda	no unigene entry			most low	most low	most low
Oct4	285			>75%	>75%	>50%
Psma1	371			>75%	>75%	not found

Table 2.3: Expression levels of selected candidate genes in the blastocyst or if not detected there in the morula and embryonic tissues according to Unigene and percentile position in certain GEO datasets comparing embryonic cells to differentiated cells; "human": information for human homologue; "most low" means that the gene was lowly (<<25%) expressed in both differentiated and stem cells; "undifferentiated low" means the expression of the gene was low in stem cells and clearly higher in differentiated cells; genes in red were excluded from the analysis; > x% means all replicates of at least one probe targeting this gene were detected at a higher percentile than x

Candidates for overrepresented genes:	
Rnf31	ring finger protein 31
Map3k1	mitogen activated protein kinase kinase kinase 1
Csnk1a1	casein kinase 1, alpha 1
Pkn2	protein kinase N2
Clk1	CDC-like kinase 1
Map4k5	mitogen activated protein kinase kinase kinase kinase 5

Candidates for underrepresented genes:	
Edd1	E3 ubiquitin protein ligase, HECT domain containing, 1
Hdac3	histone deacetylase 3
Phf17	PHD finger protein 7
Psma5	proteasome (prosome, macropain) subunit, alpha type 5
Sqstm1	sequestosome 1
Hdac2	histone deacetylase 2
Mbd2	Methy-CpG binding domain protein 2

Table 2.4: Candidate genes for which the shRNAs targeting these genes was significantly over- or underrepresented after two weeks which were chosen for experimental validation.

their expression in embryonic stem cells. To compensate for the elimination of these two genes we included *Hdac2* into our list since it performed well for our selection criteria and the above analyses suggested its expression in embryonic stem cells (e.g. it was consistently above the 50th percentile for GDS2668, etc).
For a final list of candidate genes see table 2.4.

2.1.6 Finding candidate genes involved in differential proliferation under stress compared to non-stress conditions

As detailed above the original aim of the shRNA screening assay was not to detect shRNAs affecting stem cell proliferation in general, but such over- or underrepresented in cells grown under stress vs. non-stress conditions. In this approach we searched for shRNAs for which stressed samples exceeded a certain difference to the mean fold-change of $\ln(\frac{E_r}{E_g})$ over all genes while unstressed did not. Or, in simpler words, we searched for shRNAs without effect under normal, but with detrimental or beneficial effect under stress conditions. This means that they make cells more susceptible or protect them from stress.
For the analysis of the effects of shRNAs under stress we started with data processed as described above, i.e. after removal of probes with low signal intensity at t = 0 and collapsing of probes targeting the same shRNA sequence.
We determined probes which had a signal above mean + STDEV for at least two stressed samples and below for at least two controls (called overrepresented) or below mean - STDEV for at least two stressed and above for at least two controls (called underrepresented). To determine false discovery rates (FDRs) we scrambled the values obtained for the probes within each sample. Since we only aimed at a rough estimation of the FDR this scrambling was only done once. FDRs were estimated by comparing the number of genes found after scrambling to the number found for the unscrambled data.
Since the FDR for this analysis turned out to be too high we also tried different criteria: We varied the required number of stressed samples that had to be above / below mean +/- STDEV and of controls that at the same time had to be below / above mean -/+ STDEV. Instead of mean +/- STDEV we tried mean +/- 1.5 STDEV and mean +/- 2 STDEV as alternative thresholds. The number of shRNAs found over- and underrepresented with the different criteria and their FDR are shown in table 2.5.

None of the selected thresholds and no criteria allowed us to find shRNAs over- or underrepresented with stress at a FDR < 0.10, except for the one overrepresented gene at threshold = mean+1.5 and the c3s3 criterion, which would most likely give a higher FDR if scrambling was done several times. This might indicate that 3 replicates are too few for the experimental design and the number of shRNAs tested here.
We therefore decided to focus on testing candidate genes for association with stem cell proliferation instead of

	overrepresented			underrepresented		
	experiment	scramb.	FDR	experiment	scramb.	FDR
	control > mean + STDEV, stressed < mean + STDEV			control < mean + STDEV, stressed > mean + STDEV		
c2s2	237	213	0.90	288	242	0.84
c2s3	24	7	0.29	35	17	0.49
c3s2	108	138	1.28	112	148	1.32
c3s3	11	4	0.36	10	12	1.09
	control > mean + 2 STDEV, stressed < mean + 2 STDEV			control < mean – 2 STDEV, stressed > mean – 2 STDEV		
c2s2	17	6	0.35	55	15	0.27
c2s3	0	0	0/0	0	0	0/0
c3s2	13	6	0.46	24	14	0.58
c3s3	0	0	0/0	0	0	0/0
	control > mean + STDEV, stressed < mean +1.5 STDEV			control < mean – STDEV, stressed > mean – 1.5 STDEV		
c2s2	47	52	1.11	90	45	0.50
c2s3	2	1	0.50	7	5	0.71
c3s2	19	30	1.58	26	34	1.31
c3s3	1	0	0.00	3	3	1.00
	control > mean + STDEV, stressed < mean +2 STDEV			control < mean – STDEV, stressed > mean – 2 STDEV		
c2s2	6	7	1.67	28	13	0.46
c2s3	0	0	0/0	0	0	0/0
c3s2	1	4	4.00	10	11	1.10
c3s3	0	0	0/0	0	0	0/0

Table 2.5: Number of shRNAs found with ln-ratios as indicated for the given number of control and stressed replicates (e.g. c2s3: two control, 3 stressed replicates).

for association with stress response.

2.2 Experimental validation of candidate genes by proliferation assay

2.2.1 Introduction

Even though our primary interest in the shRNA screen was to find genes associated with stress response in embryonic stem cells the much higher statistical significance for the analysis for only proliferation (while ignoring the fact that 3 of the samples were stressed) made us decide to concentrate on validation of candidates for proliferation. The reason that more genes were found significant by the proliferation assay is most likely the higher number of replicates (n=6) compared to the analysis of stressed samples (n=3).

Previous analyses had been done by G. Jansens by plating cells on 6-well plates and comparing the number of cells plated to the number of cells after about 3-5 days. The fold change of cells for the 9 shRNA-transfected lines over this period was compared to that of untransfected cells using 3 replicates for each. These 9 lines included one expressing Firefly (FFL) shRNA as a negative and *Oct4* and *Psma1* shRNA as positive controls.
No significant changes in the proliferation rate between the lines could be detected.

2.2.2 Materials and Methods

2.2.2.1 Cloning of plasmids

Cloning of shRNA sequences into pHAGE was done with contribution of E. Hesketh of our lab.
Cloning was done to transfer sequences coding for candidate shRNAs (see table 2.4) from pSM2 (Silva et al. 2005) as kindly provided by the Elledge lab into the plasmid pHAGE-Mir2 (H. Pan et al. 2008), which is in the following called pHAGE for simplicity. The shRNA sequence was cloned behind the Human Elongation Factor 1 alpha promoter (EF1a promoter) in a microRNA environment. The pHAGE plasmid contains turboRFP as a fluorescent marker, constitutively expressed on the same transcript as the shRNA hairpin and was reported to be superior in the knock-down effect (Elledge lab, personal communication). The plasmid contains genes for ampicillin and puromycin resistance for selection in bacteria and eukaryotic cells respectively. By restriction with MluI and HpaI pSM2 and pHAGE gave the shRNA sequence and the pHAGE-backbone without shRNA sequence respectively with compatible restriction sites. We called pHAGE after inserting a shRNA targeting gene X pHAGE-X.

Transformation

The One Shot TOP10 Chemically Competent E. coli transformation kit (Invitrogen) was used to transform originally obtained plasmids or ligation products according to the manufacturer´s instructions. Negative controls

from ligation reactions (see below) were included as negative controls for the transformation.

Bacterial cultures

E.coli containing pSM2-plasmids with the shRNAs of interest were inoculated in LB medium with 50 µg/ml chloramphenicol. E.coli with (modified) pHAGE plasmids were inoculated in LB medium with 100 µg/ml ampicillin. Bacteria were grown for about 16h at 37°C, shaking at 170 rpm.

Plasmid preparation

Plasmids were extracted using the QIAprep Spin Miniprep Kit (QIAGEN) according to manufacturer´s instructions.

Measurement of DNA concentrations

DNA concentrations were measured via Nanodrop (Thermo Scientific).

Restrictions

Different pSM2 plasmids, each containing a specific shRNA, were digested with HpaI and MluI restriction endonucleases (New England Biolabs (NEB)) in a double digest to obtain shRNA sequences. To obtain the plasmid backbone pHAGE was digested with the same combination of enzymes. The backbone is called pHAGE-HpaI_MluI in the following. For details on restriction setups see table 2.6.
Digestion reactions were heat inactivated at 65°C for 20 min and cooled on ice for 10 min. A 5 µl aliquot of the pHAGE-HpaI_MluI digest was run on a 1% agarose gels to confirm complete digestion. A 10 µl aliquot of the pSM2 digest was run on a 1.5% gel.

Dephosphorylation

The pHAGE-HpaI_MluI plasmid backbone was dephosphorylated by addition of 0.5 U CIP (calf intestinal phosphatase; NEB) per 1 µg DNA and incubation at 37°C for 1.5 h.

DNA preciptiation

To reduce the volume pHAGE-HpaI-MluI was precipitated by adding 10 µl 3M NaAc and 250 µl EtOH to 100 µl. The mixture was incubated at -20°C for at least 20 min and centrifuged at 4°C and 14000 rpm for 15 min. The supernatant was taken off and the pellet washed by addition of 500 µl EtOH and centrifugation at 4°C and 14000 rpm for 10 min. The supernatant was taken off, the pellet dried and resuspended in 30 µl TE-buffer.

Gel extraction

The dephosphorylated vector backbone was run on 1% agarose gels, the band at the expected size (around 9kb) was cut out and gel extracted using the QIAquick Gel Extraction Kit (QIAGEN) according to manufacturer´s instructions.
Clean-up of digestions to obtain shRNAs was not required since E.coli taking up reannealed pSM2 plasmids would not grow under the ampicillin selection which was performed on bacteria transformed with the pHAGE-HpaI-MluI – shRNA ligation (see below).

Ligation

5 µl of the pSM2 digestion reaction and 100 ng of the gel extracted, dephosphorylated pHAGE-HpaI-MluI backbone were mixed with 1µl T4-ligase buffer, 1 µl T4-ligase (NEB) and filled up with water to a 10 µl reaction volume. Ligation was carried out at 16°C over night.
Negative controls for ligation reactions contained water instead of the pSM2 digestion reaction.

Bacteria transformed with the ligation product were grown on LB-agar plates containing 100 µg/ml ampicillin, then in liquid culture as described above. Plasmids were extracted as described above.

Restriction analysis

Restriction analysis on 400 ng aliquots of cloned plasmids was performed with MluI and HpaI. Digests were run on 1% agarose gels to confirm successful ligations.

Sequencing

Restrictions

Preparation of pHAGE-backbone		
pHAGE	20 µg	4h @ 37°C
buffer 4	10 µl	
HpaI	0.25 U/µl	
MluI	0.25 U/µl	
water	up to 100 µl	
Preparation of shRNAs		
pSM2	1.5 µg	4h @ 37°C
buffer 4	3 µl	
BSA	1 µg/ml	
SalI-HF	0.25 U/µl	
NotI-HF	0.25 U/µl	
water	up to 30 µl	
Restriction analysis		
pHAGE-shRNA	400 ng	2.5 h @ 37°C
buffer 4	2 µl	
HpaI	0.25 U/µl	
MluI	0.25 U/µl	
water	up to 20 µl	

Table 2.6: Setup of restrictions for cloning of pHAGE-shRNA plasmids

The inserts of cloned plasmids were Sanger sequenced by the University of Sheffield Core Genomics Facility sequencing service. The primer sequence used was 5'-CACGAGATGGCTGTGGCCAAG-3'. The resulting sequence was aligned with the expeced sequence as provided by the Elledge group using the Needle-algorithmus offered by the EBI (http://www.ebi.ac.uk/Tools/emboss/align/index.html) (Needleman & Wunsch 1970). If the sequences matched over the complete shRNA the sample was accepted as cloned correctly. The shRNAs targeting the following genes were successfully cloned: *Edd1, Hdac3, Map3k1, Mbd2, Pkn2* and *Map4k5*. Even though for all others bacterial colonies were also obtained after transformation none of the plasmids sequenced so far contained the correct sequence.

2.2.2.2 ES cell culture

Mouse embryonic stem cells of the CCE line at around 50-70 passages were grown in ES-DMEM, which contains per 500 ml:

- 410 ml KO-DMEM (knock-out Dulbecco's modified Eagle's medium)(Gibco)
- 75 ml HyClone fetal bovine serum (FBS) (ES-qualified) (Thermo Scientific)
- 5 ml GlutaMAX 200 mM (Gibco)
- 5 ml Non-essential amino acids (Gibco)
- 2.5 ml Penicillin/Streptomycin (50 U/ml Pen, 50 ug/ml Strep)
- 1 ml β-mercaptoethanol 50 mM (Gibco)
- 50 µl leukemia inhibitory factor (LIF) 50 mM

Cells were grown in in T25 cell culture flasks or 6-well plates (Greiner) in a volume of 5 or 1 ml ES-DMEM respectively in a 37°C and 5% CO_2 incubator. Cells were split (see below) about every other days and medium

changed every day in between. Cells were regularly checked for signs of differentiation or infection under an inverted light microscope.

Splitting

Cells were split at about 80% confluence: Medium was taken off, cells were washed twice with phosphate buffered saline (PBS; pH 7.2; Gibco) prewarmed to 37°C, trypsinized with about 100 (per well of a 6-well plate) to 300 µl (T25 flask) 0.05% trypsin-EDTA (Invitrogen) for about 2 min at 37°C and resuspended in ES-DMEM by pipetting up and down several times. About 1/8 to 1/6 of this suspension was transferred to a new flask / well that had been covered with 0.1% gelatin (Millipore) for at least 20 min and which was removed immediately before. Flasks / wells were filled up to 5 / 1 ml with ES-DMEM and shaken gently.

Freezing

For storage cells were trypsinized as described above, resuspended in about 3 ml ES-DMEM and centrifuged at 1000 rpm for 5 min. They were resuspended in 1ml pre-cooled freezing medium (50% FBS, 40% ES-DMEM, 10% DMSO) and frozen in pre-cooled cryo-tubes at -80°C.

Thawing

Frozen cells were thawed quickly at 37°C and the cell suspension in 1 ml freezing medium transferred into about 5 ml KO-DMEM (Gibco). Cells were centrifuged for 5 min at 1000 rpm, resuspended in an appropriate amount of ES-DMEM and plated on gelatinized cell culture flasks / 6-well plates.

2.2.2.3 Transfection

Transfection of packaging cell line

The 293T packaging (producer) cell line was transfected with vectors encoding virus particles and pHAGE-shRNA by lipofection with the TransIT-293 Transfection Reagent (Mirus) according to manufacturer´s instructions. We aimed at a cell density of 70% before transfection. We transfected plasmids at ratios of pHAGE-shRNA : PM2 : Rev : Tat : VSVG = 10 : 1 : 1 : 1 : 2, where PM2, Rev, Tat and VSVG stand for a expression plasmids coding for viral Gag-Pol, Rev, Tat and G-protein of the vesicular stomatitis virus (VSVG).
Medium was changed the next day to DMEM-F12 (Gibco) with 10% FBS, penicillin and streptomycin. One day later if cells appeared to be red due to the expression of turboRFP and (nearly) confluent apart from some plaques the supernatant was taken off and used for transfection of ES cells. The supernatant contained replication-incompetent lentivirus as described by Pan (H. Pan et al. 2008).

Viral transfection of embryonic stem cells

To virally transfect ES cells the supernatant from producer cells was centrifuged at 1000 rpm for 3 min and the supernatant taken to get rid of remaining 293T cells. 10 mg/ml polybrene was diluted 1:10 with PBS and 9 µl of this were mixed with viral supernatant of one well of a 6-well plate (2 ml).
ES cells were trypsinized and resuspended in ES-DMEM. 100 000 cells (in 2 ml ES-DMEM) according to counting with Coulter Counter Z1 (Beckman Coulter) were mixed with the viral supernatant in a gelatinized 6-well plate. The plate was centrifuged at 2000 rpm at 25°C for 50 min. Cells were incubated at 37°C over night. Then the medium was changed to ES-DMEM the next day and to ES-DMEM with 2 mg/ml puromycin the day after. Cells were then cultured as described keeping them on ES-DMEM with 2 mg/ml puromycin for about one week till sufficient fluorescence intensities were reached.
About 3 days after the end of antibiotic selection transfected cells were mixed with untransfected ones as described below. The 3 day interval was chosen to on the one hand allow cells to recover from the stress induced by puromycin selection, but on the other hand to not allow too much loss of fluorescence by either silencing of the transgene or outgrowth of untransfected cells remaining after selection. The extended culturing time after transfection is also a means not to detect shRNA function that renders cells in-viable instead of such slowing their proliferation in the following assay.
Different cell lines were created this way each containing one kind of pHAGE-shRNA vector for all candidate shRNAs we successfully cloned: *Edd1, Hdac3, Map3k1, Mbd2, Pkn2* and *Map4k5* (see "2.2.2.1 Cloning of plasmids"). As a negative control pHAGE-FFL was used since shRNA targeting FFL does not have a target in murine cells. As positive controls we used pHAGE-*Oct4* and pHAGE-*Psma1* which had previously been shown in our lab to significantly reduce stem cell proliferation.

2.2.2.4 Proliferation assay by flow cytometry

Since splitting comes inherently with a relatively high error in the number of viable cells transferred to the new plate we decided to use a different assay which employs untransfected cells as an internal standard and therefore allows splitting. This method is supposed to be robust to slightly different treatment of samples, for example that plating cells at different densities may lead to different differentiation rates of stem cells. When mixing transfected cells with untransfected cells the differentiation which is not due to the effect of the siRNA is expected to be the same for both and proliferation ratios between them are therefore comparable even if different replicates were not plated at exactly the same density. Also the ratio of cells dying due to the splitting procedure is expected to be the same for both.
To use untransfected cells as an internal standard is possible because the plasmid containing the shRNA also contains a gene for turboRFP which allows to distinguish transfected from untransfected cells. Furthermore the shRNA and the fluorescent protein are expressed on the same transcript so that silencing of the shRNA would automatically lead to loss of the fluorescence even though the kinetics of loss of the knock-down effect and fluorescence might be somewhat different.

Mixing of cells

To compare growth rates of transfected cells to that of an internal standard of untransfected cells we aimed at mixing them after trypsinization and resuspension at a ratio of 1:1. We aimed at obtaining a mixture of about 700 000 cells. The concentrations of cells in the resuspensions were determined by counting with a Coulter Counter Z1 (Beckman Coulter). For this resuspended cells were diluted 1:20 in PBS. The lower threshold for particle size was set to 0.8 µm. Mixtures were obtained in triplicate.

Flow cytometry

For flow cytometry cells were trypsinized as described and resuspended in about 2 ml of KO-DMEM. To obtain a single cell suspension cells were pipetted up and down vigorously several times. Flow cytometry was done on FACSCALIBUR (Becton, Dickinson (BD)) controlled by the Cell Quest Pro software. In a first run a side scatter threshold separating presumably intact cells from debris was identified and the same threshold applied in all further runs. 10000 cells above this threshold were measured per sample. The parameters side scatter (SSC), forward scatter (FSC) and FL2 fluorescence (i.e. red fluorescence) were recorded. Before and after each run the instrument was flushed with FACS rinse (BD) and water.
Flow cytometry data were analysed with WinMDI version 2.9. On a dot plot of SSC vs. FSC the cell population containing presumed living, single cells and excluding dead cells and debris was gated. The same gate was applied for different samples measured on the same day, but the best gate was selected at every day of measurement so that they might differ slightly between time points. For the gated cells on a histogram displaying cell counts vs. fluorescence intensity levels positive and negative populations were separated at the minimum between both peaks. The intensity value for the border between the peaks was chosen once and kept for all further analyses and always coincided well with the minimum between the peaks. The percentage of positive to negative cells was given back by the program.

2.2.2.5 Proliferation assay by fluorescence microscopy

Despite the lack of a proper negative control (see "2.2.3.1 Flow cytometry results") flow cytometry showed a much stronger decrease of fluorescent cells in the cell line transfected with pHAGE-Edd1 than in all other cell lines. Therefore the fluorescence level of cells transfected with pHAGE-Edd1 and pHAGE-FFL was observed over two weeks by S. Silva of our group using fluorescence microscopy.

2.2.3 Results

2.2.3.1 Flow cytometry results

Results from one-color flow cytometry obtained from mixes of cell lines with pHAGE-shRNA plasmids and untransfected cells were inconclusive. Cells with shRNAs targeting Oct4 and Psma1, which were expected to have a strongly negative effect on stem cell proliferation did not show any significant difference to other lines in many occasions. This could possibly be attributed to the fact that high transformation levels were never reached for plasmids coding for these shRNAs at the time of mixing. This is probably due to the averse effects of these shRNAs on the cells. On the other hand we noted changes of the fluorescence ratio of lines transfected with

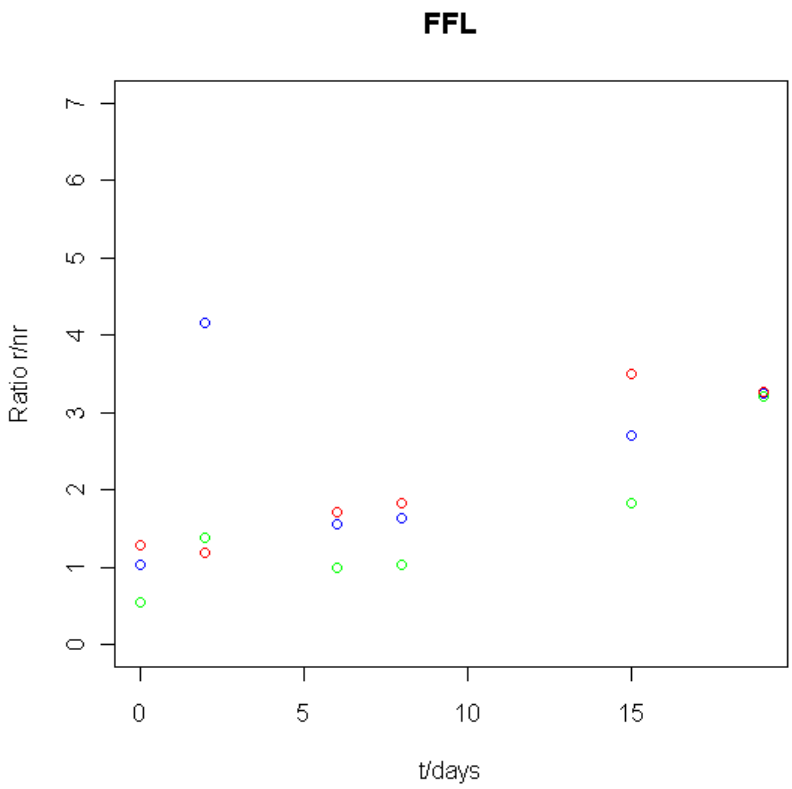


Figure 2.4: Example illustrating the trend of increased proliferation rate in the pHAGE-FFL line. The ratio of red (transfected) to non-red (untransfected) cells is depicted on the y-axis. Different colors indicate different replicates. In another experiment the same line showed a trend towards a decreased proliferation rate.

pHAGE-FFL for which no effect was expected. These changes appeared to be not random fluctuations, but a decrease of fluorescence in one, an increase over time (p-value for null hypothesis that no change: 0.05) in another experiment (see fig. 2.4). A decrease of fluorescence could be explained by silencing of the turboRFP gene and by general negative effects of transformation and an active RNAi machinery on proliferation rate. However we did not find a reasonable explanation why proliferation should be increased in the transfected cells.

One concern about this approach was that untransfected cells could not be distinguished from transfected cells that silenced the turboRFP transgene. Furthermore comparing fluorescent to non-fluorescent cells is sensitive to possible day-to-day fluctuations in the sensitivity of the flow cytometer.
Therefore replacing the turboRFP gene in pHAGE-FFL by GFP and employing cells transfected with this vector as new internal standard might solve this problem and allow comparing fluorescent with fluorescent cells. Mixing the candidate lines with a green fluorescent line instead of a untransfected line has the advantage, that the same effect of the transformation process and an active RNAi machinery is expected in both lines in the mixture. Further if day-to-day fluctuations in the sensitivity of the flow cytometer are laser (color) independent these would affect both cell lines in the same way. Therefore the ratio between the number red and green fluorescent cells should stay constant in cases where the shRNAs in the corresponding vectors have no or both the same effect on proliferation.
This kind of experiments were not finished at the time of this writing.

2.2.3.2 Fluorescence microscopy results

Despite the lack of a proper negative control flow cytometry showed a much stronger decrease of fluorescent cells in the cell line transfected with pHAGE-Edd1 than in all other cell lines (see fig. 2.5). This finding could be verified by S. Silva of our group by following the fluorescence loss of the pHAGE-Edd1 line compared to the

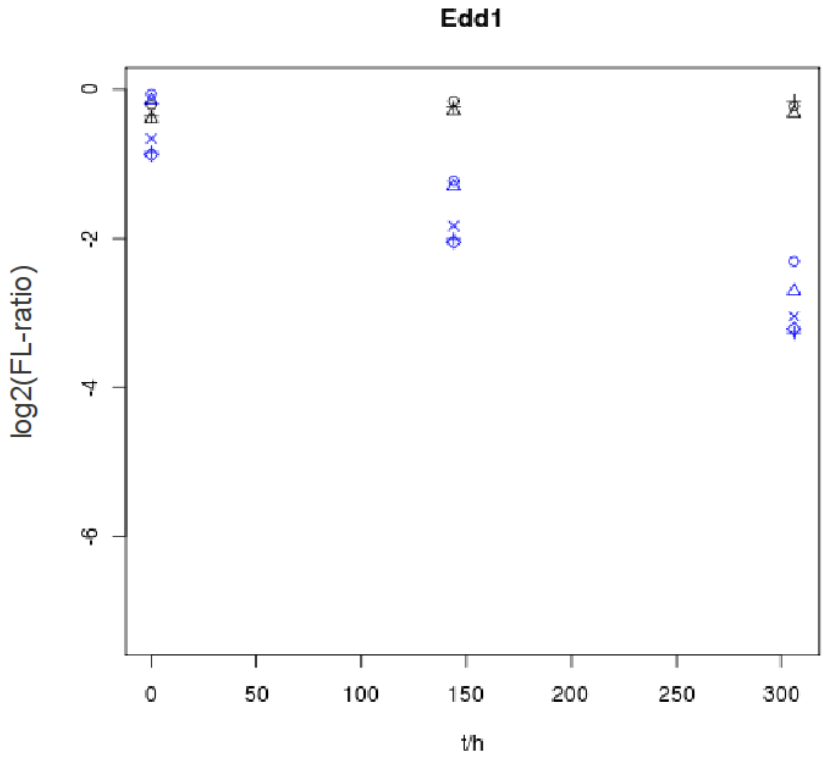


Figure 2.5: Fluorescence ratio (log2-transformed) of Edd1 (blue) and FFL (black) cell lines over time; different measurement of the FFL line than shown in fig. 2.4; different symbols represent different replicates

pHAGE-FFL line using fluorescence microscopy (unpulished).

2.3 Summary

From a shRNA library screen we could identify 23 / 60 shRNA genes for which $\ln(\frac{E_r}{E_g})$ was above mean + STDEV / below mean - STDEV over all shRNAs for 5 of 6 replicates, where E_g is the amount of shRNA coding DNA in the population at the beginning and E_r at the end of two weeks of growth according to microarray analysis. This corresponds to FDRs < 0.05. By their association to (enriched) functional categories, the number of probes by which they were found and their degree in the network of all genes targeted by these 83 shRNA we selected 13 candidates for which to validate their role in stem cell proliferation.
Unfortunately further work is still necessary in establishing a flow cytometry based assay in which the fluorescent to non-fluorescent cell ratio of pHAGE-FFL transfected and untransfected cells stays at a stable level. One possible way to improve the control may be the use of cells transfected with pHAGE-GFP-FFL as internal standard instead of untransfected cells.
The detection of significantly higher loss of fluorescence in pHAGE-Edd1 than in pHAGE-FFL transfected cells simply by fluorescence microscopy suggests success in selecting at least one or some promising candidates.

Chapter 3

Meta-analysis of caloric restriction datasets

3.1 Introduction

3.1.1 The potential of caloric restriction to delay aging

Caloric restriction (CR; also called calorie restriction or dietary restriction) is defined as the reduction of caloric intake below ad libitum level without malnutrition (ad libitum: an organism eats as much as it wants). It has been described to extend (mean and median) life-span in a wide range of organisms from yeast (Lin et al. 2002) to *C. elegans* (Klass 1977) and *D.melanogaster* (Loeb, & Northrop 1917) to rodents (McCay et al. 1989) and some dog breeds (Kealy et al. 2002). The length by which life-span can be extended by CR differs between organisms: 3-fold extension was found in yeast, 2-3 fold in worms, 2-fold in flies and still 30-60% in rodents. In general life-span extension is more pronounced in females (Fontana et al. 2010). Studies on primates are still ongoing, but intermediary results from a study on rhesus monkeys indicated that they lived on average 32 years on CR while controls lived 25 years (Bodkin et al. 2003). The degree of food restriction in CR studies of mammals is normally around 10-50% below ad libitum level (Fontana et al. 2010).
CR is the only known non-genetic intervention that robustly extends life-span in mammals (Bishop & Guarente 2007a). In addition to life-span extension it has been shown to delay signs of aging and the onset and progression of age-related diseases like cardiovascular disease and stroke (Mattson & Wan 2005), cancer (Klebanov 2007), neurodegenerative diseases (Maswood et al. 2004) and diabetes (Anson et al. 2003) as well as to reduce sarcopenia and grey matter atrophy of the brain (Anderson et al. 2009) (Colman et al. 2009). One study reported that around 30% of rats on CR did not show any obvious organ pathology at the time of death compared to 6% of mice fed ad libitum (Shimokawa et al. 1993). [1]
Notably it has been shown that caloric restriction exerts its beneficial effects even in older animals (Spindler 2005) (Rae 2004). Effects on life-span in *Drosophila* seem to occur immediately after the switch to the low-calorie dietary regime (Mair & Dillin 2008) (Giannakou et al. 2008).
Despite the effect of CR in many species it does not appear to extend the lifespan of the housefly (Cooper et al. 2004). It was also reported that no aging delaying effect of CR was found in some mouse strains (Forster et al. 2003). In particular, CR does not appear to extend average lifespan in wild-derived mice, even though it protects against cancer to a certain degree as observed in other mouse strains (Harper et al. 2006).
A possible explanation for the life extending effect of CR in terms of evolution is that it may be preferable for animals under conditions of limited food to delay growth and reproduction and enter a stage of low energy requirement (like the Dauer stage in *C.elegans*) or to shift energy allocation towards body maintenance. As detailed below there is growing evidence for conserved pathways working as anti-aging systems. Not surprisingly reduced fertility was observed in animals under CR (Fontana et al. 2010).
Other frequently observed side effect of CR are decreased wound healing (Reed et al. 1996) and immune functions rendering CR animals more susceptible to infections, although the age-dependent decay of some immune functions appears to be slowed down by CR (Kristan 2008).

[1] Side note: Curiously fasting was shown to reduce the adverse effects of chemotherapy, seemingly by conferring increased stress resistance to normal cells while not protecting cancer cells (Raffaghello et al. 2008) (Safdie et al. 2009)

Alternative dietary regiments except reducing overall food intake without malnutrition have been tested in their potential to delay aging and extend life-span. One of these is protein restriction, where a certain amount of the protein content of the normal diet is replaced with carbohydrates and fat, i.e. not altering the calorie level (López-Torres & Barja 2008). Different studies on protein restriction obtained different results as to its ability to extend life-span (Goodrick 1978) (Leto et al. 1976) (Miller & Payne 1968) (Min & Tatar 2006) (Yu et al. 1985). The majority of these studies indicated the existence of a life-span extending effect however another study even showed an increase in mortality under this diet (Ross & Bras 1973). Restrictions in only individual amino-acids like tryptophan (Segall & Timiras 1976) or methionine (Orentreich et al. 1993) are also tested.
Some studies in *Drosophila* and C. *elegans* demonstrated that the smell of food alone can reduce the effect of CR (Smith et al. 2008) (Libert et al. 2007).
Another dietary setup involving the reduction of calories is intermittent fasting. In contrast to classical CR where the amount of calories is continuously low here periods of low caloric diet alter with periods of ad libitum intake. In studies on intermittent fasting the degree of restriction is often similar to that in CR and time-spans of fasting and ad libitum feeding are similar, normally in the range of days to a few weeks. Even though studies reported reduced tumor formation in mouse tumor-models (Cleary et al. 2007) (Bonorden et al. 2009) and health benefits in humans (Halberg et al. 2005) (Heilbronn et al. 2005) effects on life-span are still unclear. These alternative dietary regiments will however not be the subject of this study.
A number of compounds are currently studied in the hope to find CR-mimetics, drugs that invoke similar effects as CR. Among these are 2-deoxy-d-glucose (Ingram et al. 2006), rapamycin (Harrison et al. 2009), resveratrol (Howitz et al. 2003) (Wood et al. 2004) and the diabetes drug metformin (Anisimov et al. 2003).

3.1.1.1 Physiological changes induced by CR

CR induces alterations in the physiology of many organ systems in mammals however it is not clear which of these changes are causal for the effect of CR (Koubova & Guarente 2003). As expected one important physiological change associated with CR is high insulin-sensitivity, which is particularly noteworthy since aging is generally accompanied by elevated insulin-resistance (Anderson & Weindruch 2010).
The reduction of body weight under CR is usually proportional to the level of CR (i.e. 30% food restriction leads to ∼30% weight loss). The tissue displaying most loss of weight is normally white adipose tissue (Anderson & Weindruch 2010). This is accompanied by size-reduction of adipocytes in mice. Due to the negative correlation of fat mass to adiponectin levels the level of this hormone rises during CR in the adipose tissue and so does its serum concentration (Zhu et al. 2004), especially of the high molecular weight form (Shinmura et al. 2007). This hormonal change comes along with increased fatty acid oxidation in fat tissue and reduced lipid accumulation in other tissues (Zhu et al. 2007). Further positive effects of adiponectin, in particular in mouse models for diabetes are known, like increased insulin-sensitivity and reduced hyperglycemia, hypertriglyceridemia and adipose tissue macrophage levels (Wang et al. 2006).
Further hormonal changes include the reduction of triiodothyronine, testosterone and insulin. Reductions of blood cholesterol, C-reactive protein, blood pressure and intima-media thickness of the carotid arteries, which are risk factors for cardiovascular disease were likewise observed (Fontana & Klein 2007) (Fontana et al. 2004). An overview of tissue-specific changes with CR is given in table 3.1.

A study on Rhesus monkey muscle tissue using immunogold electron microscopy and biochemical assays reported significantly reduced oxidative damage (reduced 4-hydroxy-2-nonenal-, nitrotyrosine- and carbonyl-modified proteins) in the CR group (Zainal et al. 2000). A reduction in inflammation (Anderson et al. 2009) and core body temperature (Mattison et al. 2003) was observed as well.
Another physiological effect of CR observed in rats is the reduced accumulation of advanced glycation endproducts (AGEs) (Teillet et al. 2000). AGEs are created by the combination of glucose and proteins and accumulating with age (Bunn et al. 1978). Notably another study found that a diet enriched in preformed AGEs abolished the beneficial effects of CR (Cai et al. 2008).

3.1.1.2 The genetic basis of CR

Little is understood by now about the changes on molecular levels going on during CR. However some findings in the last years are starting to shed light on its mechanisms.
A way to gain knowledge about which processes occurring during aging on the molecular level are prevented or counteracted by CR is to test which gene expression changes with aging in ad libitum (AL) animals are not found

Tissue	Effects of CR	References
Liver	Increase in gluconeogenesis and glycogenolysis	Weindruch, 1988
	Decrease in glycolysis	
Muscle	Increase in mitochondrial biogenesis and respiration	Koubova, 2003; Nisoli, 2005; Weindruch, 1988
	Increase in β-oxidation of fatty acids	
	Increase in protein turnover	
Fat	Decrease in storage of triglycerides	Weindruch, 1988; Martin, 2007
	Decrease in secreted leptin	
	Increase in secreted adiponectin	
Pancreatic β-cells	Decrease in secreted insulin	Weindruch, 1988
Brain	Decrease in pituitary secretion of growth hormone, thyroid hormone, gonadotropins	Weindruch, 1988; Mobbs, 2001
	Increase in adrenal secretion of corticoids	
Whole organism	Increase in insulin sensitivity and decrease in blood glucose	Weindruch, 1988; Nisoli, 2005
	Increase in metabolism	

Table 3.1: Effects of CR on individual tissues and the whole mammalian organism. From Bishop, 2007.

under CR.
It is not yet clear if CR acts by reversing age associated transcriptional changes, since some studies reported global or partial prevention of age-related changes by CR, while others did not find a significant such effect (Lee et al. 1999) (Kayo et al. 2001) (Dhahbi et al. 2006) (Park & Prolla 2005). It seems however save to assume that CR at least counteracts changes in some aging related transcriptional modules (Swindell 2009). In particular alterations in the expression of components of the electron transport chain, which in an across-species study was found to be the only age-related alteration occurring in flies, worms, mice and humans (Zahn et al. 2007), are opposed by CR (Anderson & Weindruch 2007). It is generally important to note that (mitochondrial) energy metabolism is dysregulated with age and that energy metabolism pathways are affected by the alterations due to CR, especially in heart, skeletal muscle and white adipose tissue in mammals.
It was observed that respiratory capacity per isolated mitochondrion is lower in mitochondria of older mice (18 vs. 3 months old) and -probably as a compensatory mechanism- the number of mitochondria is increased in older animals (observed in skeletal muscle) (Figueiredo et al. 2009). Most likely this is closely linked with elevated levels of oxidative damage that may be a cause of the aging phenotype.
An alteration of metabolic state is invoked by CR which involves a shift from fat anabolism to catabolism and changes in the production of reactive oxygen species (ROS). Notably uncoupling protein [2] UCP3 which is presumably important for lowering ROS levels is overexpressed in CR (Asami et al. 2008).
A common regulatory system for the expression of uncoupling proteins, elements of fatty acid metabolism and transport (e.g. by the transporter CPT1) may be provided by AMPK-signalling ((Anderson & Weindruch 2010); see "3.1.1.2" and "3.1.1.2").
Another common way of detecting genes related to the life-span prolonging effect of CR is by searching for genes that alter (increase or decrease) this effect when mutated, deleted, knocked-down or overexpressed.
In this way many proteins that were already known to extend life-span when altered in their expression level or function were linked to CR. In particular decreased insulin / insulin-like signalling, decreased TOR and / or increased AMPK and increased activity of sirtuins were among the genetic alterations to extend life-span (Bishop & Guarente 2007a). Evidence of relation of these and some other (mainly nutrient sensing) pathways to CR in different model organisms will be discussed, starting with yeast and then examining in how far homologous mechanisms in higher animals exist.

3.1.1.2.1 Genes involved in CR mediated life-span extension in yeast

Life-span in yeast can be measured in two different ways: replicative life-span is the number of daughter cells a mother cell can produce before senescing and chronological life-span is the duration of viability of stationary phase cells. It has been suggested that replicative lifespan is a better model of ageing for mitotically active animal

[2] uncoupling proteins are proteins that lower the proton gradient over the inner mitochondrial membrane

cells and that chronological lifespan is a better model for postmitotic animal cells (Bishop & Guarente 2007a). Both moderate (0.5% glucose medium) and severe CR (0.05% glucose) increase replicative life-span in yeast.
In yeast moderate CR (0.5% glucose instead of 2%) has been shown to increase replicative life-span through a pathway dependent on shifting metabolism from anaerobe to aerobic (Lin et al. 2002). Contributing evidence to this finding is the fact that deletion of cytochrome C1 (*CYT1*) or *LAT1* (a pyruvate dehydrogenase subunit) which in both cases suppresses respiration abolishes the life-span increase with moderate CR. In addition overexpression of *LAT1* increases yeast life-span under 2%, but not under 0.5% glucose conditions. The anaerobe to aerobic shift increases the NAD^+/NADH ratio which has been shown to be necessary and sufficient for an increase in life-span. Interestingly high levels of NAD^+ activate the (histone) deacetylase *SIR2* and its homologues, which are known to drive life-span extension (Lin et al. 2004). If however the triple deletion of *SIR2* and its homologues *HST1* and *HST2* is sufficient to suppresses longevity caused by moderate CR is still a matter of debate (Longo & Kennedy 2006). In yeast recombination between rDNA repeats can lead to excision of self-replicating extrachromosomal rDNA circles, which accumulate in the aging mother-cell, a process that is toxic for the cells (Sinclair & Guarente 1997). The ability of Sir2 to suppress recombination (by leading to higher density chromatin packing) and therefore limiting this process is one important mechanism by which it extends life-span (Lin et al. 2000). Even though this process was not found to occur in other organisms *Sir2* homologues are still linked to longevity in higher organisms (Guarente 2005).
The mechanism of severe (0.05% glucose) CR seems to be distinct from that of moderate CR and has been reported to neither involve the electron transport chain nor *SIR2* or its homologues (Tsuchiya et al. 2006). Unlike for moderate CR *SIR2* deletion does not seem to abolish the effects of severe CR (Lamming et al. 2005), but on the contrary to even enhance them (Kaeberlein et al. 2004) and severe CR does not invoke such a strong increase in the NAD^+/NADH ratio (Easlon et al. 2007).
Instead the *Akt* homologue *SCH9* and *TOR1* have been proposed to be involved in the process, since their deletion leads to life-span extension that cannot further be improved by severe CR (Kaeberlein et al. 2005). Both proteins act in the *S. cervisiae* amino acid sensing pathway and transcription factor Gis1 was reported to be essential for the life-span extension by reduced Tor1-signalling (Wei et al. 2009) (Fabrizio et al. 2001). In general mutations activating the severe CR response also prolong chronological life-span in stationary yeast cells with no access to nutrients, which is not true for genes extending replicative life-span under moderate CR (Powers et al. 2006).
It is interesting that the increase of life-span both under moderate and severe CR seems to require the pyruvate dehydrogenase subunit Lat1 especially since a functional electron transport chain is not required in severe CR (Easlon et al. 2007).
It is not yet clear if indeed two different pathways are underlying moderate and severe CR in yeast. If so, the fact that worms and mice under CR also show increased respiration (Nisoli et al. 2005) might indicate that the mechanism of moderate CR in yeast more closely resembles that in higher organisms, whereas severe CR might rather resemble survival mechanisms triggered by famine (Bishop & Guarente 2007a).
Another nutrient sensing pathway linked to life-span regulation in several studies is the Ras-AC-PKA pathway (Fabrizio et al. 2001) (Medvedik et al. 2007). This pathway is largely homologous to the insulin / insulin-like growth factor signalling pathway in higher organisms (Fontana et al. 2010).
Downstream effects of reduced activity of the Tor1/Sch9 and the Ras-AC-PKA are the activation of oxidative stress protective enzymes like Mn-SOD (superoxide dismutase) via transcription factors as Gis1 (Wei et al. 2008). This would suggest an easy explanation for the anti-aging effect of reduced signalling via these pathways, especially since it was found that superoxide levels rise during yeast aging. However overexpression of both superoxide dismutases or catalase only lead to a minor increase in life-span (Fabrizio et al. 2001) (Fabrizio et al. 2005), so that their increased activity is most likely only one effect of CR.
Another downstream effect of reduced signalling via both pathways mentioned is the expression of *PNC1*, which by increasing NAD^+/NADH and reducing nicotinamide in turn activates Sir2 (Medvedik et al. 2007) (Kaeberlein et al. 2007).

3.1.1.2.2 Genes involved in CR mediated life-span extension in metazoa

Probably the most important genes associated with life-span in C.elegans are genes of the insulin signalling pathway, especially the insulin receptor homologue *daf-2* and *FOXO* homologue *daf-16* acting downstream in this pathway. Mutants in *daf-2* are well-established to be long-lived, however this longevity is abolished in doube-mutants with *daf-16* (Kenyon et al. 1993). The fact that CR was shown to increase life-span in *daf-16* mutants to a similar extent than in wild type worms suggests that CR does not act via the insulin signalling pathway

in worms (Houthoofd et al. 2003) (Lakowski & Hekimi 1998). However a more recent study assaying different CR-regiments concluded that *daf-16* is necessary in some and not in others (Greer & Brunet 2009). Interestingly it is necessary for such regiments in which also AMPK is required. However deletion of the homologous protein in Drosophila, *dFOXO*, shortens life-span and these flies continue to respond to CR (Giannakou et al. 2008) (Min et al. 2008). Another forkhead family transcription factor, PHA-4, has been found to be required for life-span extension by CR in C. elegans (Panowski et al. 2007). This gene is an orthologue of the mammalian FOXA genes that are involved in the production of glucagon and in gluconeogenesis during fasting.
Insulin / insulin-like growth factor signalling was also found to control life-span in flies and mammals (Kenyon 2005). The signalling factors in Drosophila are called Drosophila insulin like peptides (dilps) and the gene expression level of one of the seven known dilps, *dilp5,* can be modulated by diet (Min et al. 2008). The *chico* gene is a homologue to insulin receptor substrate genes and the *chico1* mutation both increases life-span and reduces insulin signalling (Clancy et al. 2001). CR was found to gradually increase life-span with increasing levels of food restriction up to a certain point where it starts to decrease probably due to starvation. Observing this dose-response curve in *chico1* mutants showed that it was shifted towards higher nutrient levels compared to the wild type (Clancy et al. 2002). Therefore an overlap between the mechanisms of CR and reduced insulin signalling was suggested, even though a CR response that is normal apart for the mentioned shift in a mutant background would argue against the role of the mutated gene in CR (Bishop & Guarente 2007a).
Experimental results in mice of testing the link between CR and the growth hormone (GH) – insulin-like growth factor 1 (IGF1) axis, disruption of which leads to increased life-span (Flurkey et al. 2001), are confusing. On the one hand mice with a reduced production of GH due to a mutation in *Prop1* show an increased life-span (Brown-Borg et al. 1996) that could be further prolonged by CR (Bartke et al. 2001), on the other hand longevity due to disruption of the GH receptor (Coschigano et al. 2000) was not further extended by CR (Bonkowski et al. 2006). The first finding argues against, the second for an overlap between genes involved in the CR response and the GH-IGF1 axis. A decrease in GH was linked to elevated levels of antioxidant enzymes and stress response (Brown-Borg 2007). It was also found that IGF1 levels in the blood were lowered by CR in mice (18% restriction, 24 weeks) (Huffman et al. 2008), whereas no changes were detected in humans (20%, 1 year) unless dietary protein levels were strongly reduced (Fontana et al. 2008).
Heterozygous mutations in IGF1-receptor (Suh et al. 2008) and polymorphisms related to reduced plasma IGF1 levels (Bonafè et al. 2003) are overrepresented among long-lived humans. Also human genetic variants of *daf-16* homologous FOXO genes were also associated with life-span (Kuningas et al. 2007).
Another regulatory system most likely involved in CR-dependent life-span extension is built around AMPK. A very simplified view of this network is shown in fig. 3.1.

AMPK, a important protein for sensing energy levels in worms and a homologue in yeast have been shown to be implicated in longevity (Apfeld et al. 2004) (Ashrafi et al. 2000). Deletion of a AMPK subunit gene (*aak-2*) in worms did not alter the effect of CR on life-span (Curtis et al. 2006), which however may be attributed to redundancy of this protein.
AMPK directly activates PGC-1α by phosphorylation and also through its indirect positive influence on the NAD^+/NADH ratio which in turn enhances the activity of SIRT1, the enzyme that deacetylates and thereby activates PGC-1α (Cantó et al. 2009). PGC-1α , a master-regulator of nuclear encoded mitochondrial genes, itself was found to be upregulated with CR in skeletal muscle (Civitarese et al. 2007). Overexpression of PGC-1α also promotes signalling through HIF-1α (O'Hagan et al. 2009) which is downregulated in adipose tissue of mice upon CR (Yoshikazu Higami et al. 2006) and a C.elegans homologue of which is associated with CR and longevity (Chen et al. 2009).
AMPK can also activate eNOS in response to adiponectin (Kondo et al. 2009) which is implicated in mitochondrial biogenesis and SIRT1 expression in CR. Consistently eNOS knock-out mice were found not to undergo the normal metabolic shift associated with CR (Nisoli et al. 2005) and the life extending effect of CR is abolished in mice in which eNOS is inhibited (which also prevents activation of SIRT1). SIRT1 in turn is an activator of PGC-1α, which is consistent with the reported upregulation of PGC-1α coinciding with the upregulation of eNOS upon CR in many tissues (Nisoli et al. 2005).
NAMPT is a protein involved in the depletion of nicotinamide and therefore similar to yeast PNC1. As PNC1 it is expected to favour activation of SIRT1 by changing the NAD^+/NADH ratio and decreasing nicotinamide levels.
Homologues of yeast *SIR2* also play roles in CR in metazoa. *Sir2* in Drosophila is required for longevity caused by CR (Rogina & Helfand 2004) and *Sirt1* in mammals for the increase in spontaneous movement observed in animals under CR, suggesting a neuronal implication (Chen et al. 2005). Knock-out mice of *Sirt1* are short lived

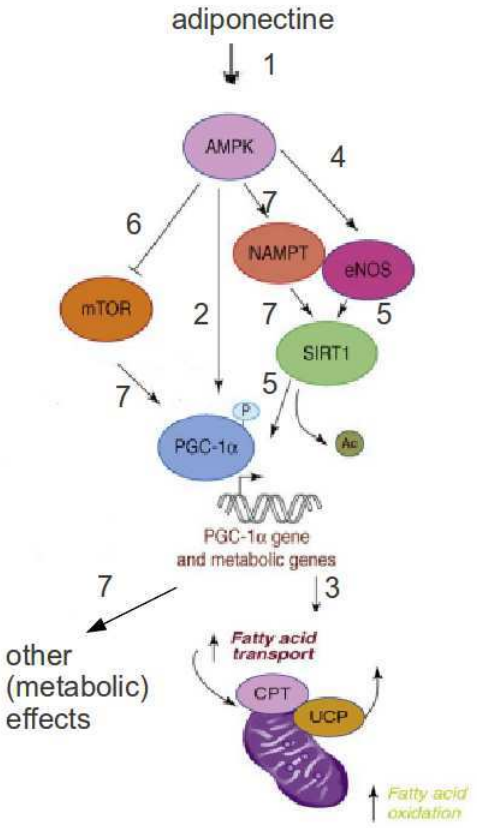


Figure 3.1: Simplistic model of the AMPK signalling pathway with a central role in CR; adapted from Anderson & Weindruch, 2010; Ac: acetyl group; numbers indicate references for the interaction: 1: Civitarese, et al. 2006, 2: Canto, et al. 2009, 3: Andrews, et al. 2008, 4: Kondo, et al. 2009, 5: Nisoli, et al. 2005, 6: Gwinn, et al. 2008, 7: Anderson & Weindruch, 2010

and do not respond to CR (Boily et al. 2008), are however of limited informative value due to the vast number of pathologies caused by this knock-out. The levels of SIRT1 are increased in mouse fat tissue during CR (Cohen et al. 2004), but there is disagreement on the impact of CR on SIRT1 in liver and skeletal muscle (Chen et al. 2008) (Cohen et al. 2004) (Shinmura et al. 2008). SIRT1 activation in mice on a diet rich in fat supports lipid oxidation and the expression of genes of the electron transport chain (Feige et al. 2008).
Even though the implication of *sir2-1*, the only of four *SIR2* homologues in worms tested for its role in CR, remains controversial (Lamming et al. 2005) *sir2* homologues were reported as life-span regulators also in invertebrates (Longo & Kennedy 2006)(Tissenbaum & Guarente 2001).
AMPK, a sensor of cellular energy levels, inhibits mTOR (Complex I) via TSC2 or raptor (Gwinn et al. 2008). It should however be emphasized that mTOR also receives inputs from the insulin / Igf-pathway. As in yeast reduced TOR-signalling leads to an extension in life-span of worms and flies that cannot be further enhanced by CR so that an overlap in the mechanisms is likely (Vellai et al. 2003). A hint towards the mechanism of life-span extension by reduction in TOR-signalling may be that it decreases ribosomal biogenesis. This is interesting, since lower expression of certain ribosomal genes is associated with longevity in yeast and worms (M. Kaeberlein et al. 2005). Disruption of the mTOR pathway in mice leads to longevity associated with reduced insulin resistance and age-related pathologies (Bartke 2005) (Harrison et al. 2009) (Selman et al. 2009). However since in mice the increased expression of genes of the electron transport chain, as observed in skeletal muscle in CR, appears unlikely when mTOR activity is reduced Anderson proposed different tissue-specific effects of CR on mTOR, with decreased signalling in liver, but not some other tissues. Downstream mTOR positively regulates the expression of PGC-1α (Anderson & Weindruch 2010) and importantly it inhibits autophagy. Autophagy is the process of digestion of cellular components by so called phago-lysosomes and was reported to be necessary for life-span extension (Hansen et al. 2008). Other targets indirectly transcriptonally regulated by mTOR-signalling in mouse are heat shock proteins, proteins involved in ER-stress and apoptosis and in xenobiotics metabolism (Amador-Noguez et al. 2007). The detoxification process of xenobiotics became a target of CR related research after it was discovered that long-lived fly, worm and mouse mutants in the insulin / IGF-signalling showed altered expression of genes of this system and proved largely resistant to xenobiotics. Furthermore upregulation of transcription factors involved in xenobiotics metabolism invoked longevity of worms and flies (Piper et al. 2008) (Tullet et al. 2008) (McElwee et al. 2007).
A further important change with age that is counteracted by CR is the increased activity of the tumor-suppressor p53 (Edwards et al. 2007). Even though it is not clear how this relates to CR and aging, a link between p53 and mitochondrial metabolism is provided by the fact that deficiency of p53 in mice leads to a reduction in mitochondrial content, a switch from respiration to anaerobic metabolism and increased ROS levels (Matoba et al. 2006) (Saleem et al. 2009).

3.1.1.2.3 The role of neurons in CR

Some curious recent findings have linked the life-span extending CR response to neurons in invertebrates: In *Drosophila* and *Caenorhabditis elegans* it was observed that the odour of food is sufficient to reduce the longevity resulting form CR and in *Drosophil*a the mutation of OR83B, a neuronal chemoreceptor, was reported to increase life-span and render CR less efficient in this mutant background (Libert et al. 2007) (Smith et al. 2008). Further it was shown that neuron-specific overexpression of human *UCP2* in flies leads to longevity (Fridell et al. 2005). Even though the link between UCP2 and CR is largely unknown it is interesting that in humans UCP2 is involved in nutrient sensing and that a related fly protein, UCP5, is necessary in neurons to adapt to low nutrient levels (Sánchez-Blanco et al. 2006). Deletion of 3 of 7 *Drosophila* insulin like peptides in neuroendocrine brain cells resulted in longevity (Grönke et al. 2010). In *C.elegans* the transcription factor gene *skn-1* was shown to play a role in ASI neurons in CR-related increased respiration and life-span extension (N. A. Bishop & Leonard Guarente 2007b). These two neurons are important in regulating fat metabolism in adult worms in response to nutrient levels and energy status and CR-related longevity is not invoked in worms in which the ASI neurons are ablated (Bargmann & Horvitz 1991b) (Bargmann & Horvitz 1991a).
It is intriguing to assume that CR related longevity in metazoans may be caused in a similar way as in yeast with a central role for energy sensing neurons. Direct sensing of extracellular glucose concentrations e.g. by G-protein coupled receptors in yeast would be replaced by the input of sensory neurons in higher organisms. Intracellular energy levels may be detected in a similar way involving AKT- and TOR-homologues, supplemented by systemic signals from other cells in metazoans. The output would differ in the way that yeast cells would only respond to nutrient levels in a cell-intrinsic way, while neurons in higher organisms have to send appropriate signals to other cells (Bishop & Guarente 2007a).

The brain region corresponding to the energy sensing neurons in invertebrates is the hypothalamus in mammals which senses and responds to energy availability by nervous and hormonal signals. Indeed many homologues of the genes described as involved in longevity in lower organisms have implications in energy sensing in the hypothalamus (e.g. TOR, AMPK, AKT) (Bishop & Guarente 2007a). Note that growth hormone (GH) mentioned above is a signal triggered by the hypothalamus via the pituitary gland.
Even though there is no direct evidence of the role of the hypothalamus in CR one study has provided a link between the hypothalamus and life-span: Uncoupling protein *UCP2* was overexpressed specifically in the so called orexigenic hypocretin (appetite-stimulating) neurons of the hypothalamus of mice. This did not only lead to a core body-temperature reduction and mild hyperphagia, but also to an increase in mean and maximum life-span (12% in males and 20% in females) (Conti et al. 2006).

3.1.1.2.4 Rationale for an unbiased cross-tissue analysis

Even though microarray data comparing samples from individuals of different age indicated that aging related gene expression changes are mainly tissue-specific it has also been shown that the rate of aging of all tissues tested seems to be coordinated, which agrees with the idea of a set of common underlying changes in all tissues (Zahn et al. 2007). In this case besides all the tissue-specific changes a common aging delaying effect of CR on all tissues would also be expected.
As detailed the knowledge about important players and pathways as effectors of CR is growing. However to understand the underlying mechanisms many more components of the complete picture will have to be detected. Especially an explanation of which mechanisms downstream of nutrient sensing pathways lead to life-span extending processes is largely unknown. Since much research was focused so far on candidates known to be involved in nutrient sensing it seems to be advisable to also include unbiased high-throughput studies. Studies so far conducted and deposited to this end used microarrays.

3.1.2 Meta-analysis of microarray data

Meta-analysis is here defined as the quantitative review and synthesis of the results of related but independent studies (Normand, 1999). Meta-analyses can be used to assess the variability between studies or more commonly –as here is the case- to facilitate finding genes differentially expressed between two conditions by integrating different studies.
Microarray results are well-known to be associated with a relatively low signal-to-noise ratio and finding significant results is made difficult by the large number of variables compared to the relatively low number of replicates.
Since microarrays became a more and more common tool over the last years there are results for several microarray analyses available for many biological questions, even though the experimental setup of the individual studies may be more or less different. These differences can however not only be seen as a problem in comparing the analyses, but also as a chance since genes found differentially expressed under similar, but not identical conditions can be considered more reliable in their association with the tested variable, since they are affected under different circumstances. Therefore the "generalizeability" (Ramasamy et al. 2008) of a candidate gene is shown when it is found in more than one tissue, organism, strain, diet composition, for different durations of CR and ages of animals, but also microarray platforms and even different ways of handling samples in different laboratories. Meta-analyses are therefore likely to eliminate false-positives of individual studies. To determine genes showing that kind of robustness is the aim of our meta-analysis. It is a matter of debate if mechanistic candidate genes for CR are expected to be generalizable across tissues, but as detailed above we argue there should be at least some.
Besides that meta-analyses eliminate the idiosyncrasies of the different analyses, they are also a valuable tool to increase statistical power and find genes with small, but consistent differential expression that are not found in the individual analyses.

3.1.2.1 Methods for meta-analysis of microarray experiments

Several meta-analysis techniques have been applied to microarray data (Rhodes et al. 2002) (Rhodes et al. 2004) (Choi et al. 2003) (Choi et al. 2004) (Lottaz et al. 2006) (Smid et al. 2003) (Stuart et al. 2003) (Parmigiani et al. 2004) (Warnat et al. 2005) (Yang et al. 2005) (Aggarwal et al. 2006) (DeConde et al. 2006) (Wang et al. 2006) (Zintzaras & Ioannidis 2008).
According to Ramasay (Ramasamy et al. 2008) the statistical approaches can be classified by the single-study statistics they use for combining the studies: Ranks, p-values, effect sizes or counts, i.e. the number of studies

in which a significance threshold is passed.
Three typical methods out of the first three categories were reviewed by Hong and Breitling (F. Hong & Breitling 2008): A t-based approach, a non-parametric rank product method and Fisher´s inverse chi-square method using P-values from either the t-based or rank-product method. These and other approaches are briefly introduced in the next sections:
In the following T stands for treatment and C for control condition and i = 1,. . . ,I numbers individual datasets. n_iT and n_iC are the number of replicates for the i-th dataset of the treatment and control condition. T_{ij} / C_{ij} represents the (logged) gene expression of a given gene for study i and replicate j. The terms "dataset" and "study" are used interchangeably in this sub-chapter.

3.1.2.1.1 Combining effect-sizes: t-based (hierarchical modeling) approach A standardized mean difference for a given gene in study i can be calculated as an effect-size measure $d_i = \frac{\bar{T}_i - \bar{C}_i}{Sp}$ where Sp indicates the estimated variation. By means of an effect size model the overall (i.e. over all studies) effect size and the corresponding variance can be estimated (Hong & Breitling 2008) (DerSimonian & Laird 1986). A z-score can be derived from these to calculate the standardized average treatment effect for each gene across datasets. Permutation z-scores are calculated by column-wise permutation within each study. These can be used to estimate a false discovery rate (FDR) (by dividing the mean number of genes found by scrambling by the number found for the real data for a given z-score) and a P-value representing the probability that a gene is found more differentially expressed by scrambling than in the real analysis. (P values could also be calculated from the standard normal distribution, but scrambling better accounts for small sample size and avoids violation of the assumption of normality). This t-test based method was for example used by Choi (Choi et al. 2003).

3.1.2.1.2 Combining ranks: Rank product approach In this approach fold-changes are calculated for each gene in each study, pairwise for each treatment with each control replicate for one-channel arrays. For two channel arrays the fold changes are calculated as treatment to control ratios for each array. These fold changes are ranked and r_{gik} denotes the rank of the fold-change of gene g in study i and pairwise comparison k. Then for each gene the rank-product is calculated as $RP_g = (\prod_i \prod_k r_{gik})^{\frac{1}{K}}$

with $K = K_1 + K_2$. To assess the significance of these values rank-products are calculated in the same way after scrambling data within each array several times. Similarly as above p-values for a certain rank product are computed as the average ratio of genes with a rank at least this high in the scrambled data and FDRs by dividing the number of genes with a rank at least this high in the scrambled data by that in the actual data.
To test for overexpression with treatment fold-changes are calculated by dividing the treatment by the control expression value, for underexpression the other way round.
Another method meta-analyzing data by their rank was proposed and implemented in the bioconductor package OrderedList by Lottaz (Lottaz et al. 2006).

3.1.2.1.3 Combining p-values: Fisher´s inverse chi-square method Fisher´s inverse chi-square method (also called Fisher´s sum of logs method; (Fisher 1925)) calculates a combined statistic $S = -2\log(\prod_i P_i)$ with i= 1,...,n from the p-values of the individual studies. S follows a chi-square distribution with 2n degrees of freedom under the joint null-hypothesis and therefore allows the calculation of a combined p-value. Since the t-based and rank-product approach can also be used on single datasets, single study p-values from these methods can be used to calculate the combined statistic. The Fisher´s inverse chi-square method has to be applied testing for over- and underexpression separately.
Variations of this method include weighting single study p-values by their reliability (Good 1955) or calculating the combined statistic only from single-study p-values below a certain cutoff (truncated product method; (Zaykin et al. 2002)). The FDR can for example be controlled by introducing experiment specific p-value cutoffs according to e.g. the Benjamini-Hochberg method (Pyne et al. 2006) (Benjamini & Hochberg 1995).

Such a p-value based meta-analysis approach was presented by Rhodes et al. determining p-values by comparison of the actual with scrambled data (Rhodes et al. 2002).
In the first step the p-value for each gene in each study was calculated by a random permutation t-test, i.e. they obtained the p-value as the fraction of t-statistics obtained by randomly permuting sample labels that are greater than the actual t-statistic.
They then determined a p summary statistic for each gene in each possible combination of studies, i.e. comparing study A to study B, but also comparing studies A and B to C or B to C, etc.. Summary statistics were calculated

for each gene appearing in all studies from the individual-study p-values and were the higher, the smaller all p-values and vice versa. The summary statistic p-values were again obtained by comparing the summary statistics from the actual data to such from data scrambling p-values over genes in each study.
To determine an appropriate summary statistic p-value cutoff accounting for multiple testing genes were ranked and a q-value (FDR) was defined as the p-value divided by the fraction of genes with a lower or equal p-value. This is sensible since a FDR is the number of genes that would be found by chance divided by those actually found, which is the same as dividing the probability of finding a gene by chance (FDR) by the fraction of genes found.
Finally the lowest q-value of all combinations was taken for each gene.
This approach has the advantages that using scrambling no assumptions like normal distribution of data need to be made and that p-values of individual studies are combined without the need of setting a threshold on them.
The problem however is that calculating summary statistics for each combination of studies is computationally intensive. It is feasible for meta-analyses like this one, including 4 studies, but might not be for larger ones.
By working with p-values it is not possible in this method to estimate the mean magnitude of differential expression.

3.1.2.1.4 Limitations of methods combining effect-sizes, p-values and ranks All three of the presented methods (at least if no truncation for single study p-values is used in Fisher´s chi-square approach), as well as other methods combining ranks, p-values or effect-sizes do not seem very likely to detect genes differentially expressed in only a subset of datasets with large variations as they might appear in a combination of a cross-platform, cross-species and cross-organism approach. For example they seem not apt to detect a gene differentially expressed in some tissues, but not in others from datasets from different tissues. This is because the effect-size estimate over all studies and the between-study variance in the t-test based approach, the rank-product in the rank-product approach and the combined statistic in the Fisher´s inverse chi-square method are sensitive to the (few) cases where the gene is not differentially expressed.
On the other hand combining only some of the ranks, p-values or effect-sizes (e.g. only such found significant) and ignoring others may be hard to justify.
To overcome this problem thresholding on the single-study statistic and counting how often the threshold is passed would be useful. This is the procedure applied by vote / value counting approaches (Ramasamy et al. 2008). The disadvantage of these approaches is that statistical values have to be classified as to if they are above or below a chosen rank-, effect-size, or p-value-cutoff and all further information is lost. Therefore the big advantage of counting a gene as only differentially expressed or not in each study, which prevents strong contribution of studies where a gene is clearly non significant is at the same time the probably biggest disadvantage of not allowing studies to contribute with different weights for that gene. Therefore if a gene is found extremely significant in one study it will only contribute with one count, as does a gene with significance close to the set threshold.

3.1.2.1.5 Value-counting approaches Rhodes et al. (Rhodes et al. 2004) presented one such value-counting approach termed "comparative meta-profiling". The aim of this analysis was to find a meta-signature common to different kinds of cancer and therefore to develop a strategy that does not detect genes only differentially expressed in one or very few datasets, but find those differentially expressed in more datasets than expected by chance. By this they hoped to find a meta-signature typical for cancer per se, not a certain type of cancer.
Comparing statistical measures for each dataset rather than gene expression measures was supposed to help overcoming the challenges of comparing data from different microarray platforms. In the first step differential expression in individual datasets was assayed by a t-test. The genes of each set were sorted by the p-value and a Q-value calculated as the number of expected differentially expressed genes (p-value) divided by the number of actually differentially expressed genes (number of genes in the ranking with lower or equal p-value). The Q-value was used for comparing the datasets.
For both over- and underexpression the number of datasets in which each gene was present below a Q-value threshold of 0.1 was counted and the number of genes in each possible number of datasets tallied $(N_0, N_1, N_2, ..., N_s)$. (S is the total number of datasets). The same steps were repeated on datasets with scrambled Q-values, obtaining a tally $(E_0, E_1, E_2, ..., E_S)$. A minimium meta-false discovery rate was calculated as $mFDR_{min} = Min(\frac{(E_i+1)}{N_i})$ for i= 0..S.
If the $mFDR_{min} > 0.1$ the analysis was repeated with the Q-value threshold lowered by 50% until a $mFDR_{min} \leq 0.1$ is reached or the number of genes below the Q-value threshold is 0 for at least 2 datasets. In the second case the meta-analysis is defined not to have found a significant overlap between differentially expressed genes in the

datasets. This procedure assures that the highest possible, but still sufficiently low Q-value threshold is chosen. If a $mFDR_{min} \leq 0.1$ is found, genes enriched for over- / underexpression (meta-signature) were defined as the number of genes appearing in at least i datasets below the Q-value threshold, where i is the same used for calculating this $mFDR_{min}$.
The major drawback of this approach is that it is unlikely to detect genes only tested in a subset of the datasets. This is because the number of datasets in which a gene has to be found below a certain Q-value is determined by considering all genes also such that were tested in a different number of datasets. An alternative value-counting approach to overcome this problem uses a binomial test to both take the number of times the single-study statistic for a gene exceeds a threshold and the number of studies its gene-expression was measured into account (de Magalhães et al. 2009).

Since the sources for our datasets were very diverse, i.e. different tissues, organisms, ages, durations of CR, microarray platforms, etc. we decided to employ a value counting approach. Because the microarray experiments were performed over the course of some years, while annotation of the genomes of model organisms improved and therefore probes for newly discovered genes were included on the platforms over time (and for other reasons) we expected that not each gene was represented in a similar number of studies so that we found the binomial approach best suited for our meta-analysis.
Another advantage of using a value-counting approach is that we could include datasets for which only lists of differentially expressed genes were available (Ramasamy et al. 2008).
Ramasamy´s concern that the results of value-counting approaches are rather granular compared to those obtained by other techniques was not considered a major problem, since ranking the final results was of less importance to us than classifying them as significant or not.
Last but not least Magalhaes showed that in a situation with similar aims (i.e. finding genes robustly differentially expressed in different organisms, tissues, etc.) a binomial value counting approach performed better than Fisher´s chi-square method in terms of the number of genes identified. For the top genes of both approaches there was strong overlap (de Magalhães et al. 2009).

3.1.3 Other meta-analyses of gene expression data for CR

3 important meta-analyses of caloric restriction gene expression data were existent at the time of this writing: Hong 2010, Swindell 2008a (further analysed in Swindell 2008b) and Swindell 2009. These will be briefly introduced here and their results compared to ours in the discussion-section ("3.4.2 Comparison with results from other meta-analyses").

3.1.3.1 Swindell, 2008a

In "Comparative analysis of microarray data identifies common responses to caloric restriction among mouse tissues" Swindell created 23 contrasts comparing caloric restriction to control samples from 13 studies on mouse (Swindell 2008a). For two studies only information in supplemental data were used. In the data used the age of mice at time of killing were 4 to 31 months (or unknown for two studies), duration of CR 2 days to 24 months (or unknown for one study), the level of CR 10-66% (or unknown for one study) and data were from 10 different tissues.

Method: Swindell started off with raw data, processed them by normalization by Robust Multichip Average (RMA) (Irizarry et al. 2003), determined differentially expressed genes using the Bioconductor Limma package (Smyth 2004) and adjusted P-values by the Benjamini-Hochberg method (Benjamini & Hochberg 1995). A significance level of 0.05 was used to identify differentially expressed genes in each study. The number of different tissues in which a gene was differentially expressed was counted. This study therefore emphasizes robustness of differential expression over different tissues. The approach is a value counting approach with the problem of ignoring that different genes may have been tested in different numbers of studies.

A differential expression signature was created for each dataset by assigning -1 to downregulated, 0 to non-significantly differentially expressed and 1 to upregulated genes. A similarity score for each pair of datasets was calculated by

$s = \frac{n_{+,+}+n_{-,-}}{n_{+,+}+n_{-,-}+n_{+,-}+n_{-,+}+Min[(n_{+,0}+n_{-,0}),(n_{0,+}+n_{0,-})]}$, where $n_{+,+}$represents a gene significantly upregulated in both sets, etc. The similarity score was used for clustering datasets.
The significance of the overlap between two datasets was assessed by scrambling of the assigned +1, -1, 0 marks. The test statistic $T = n_{+,+}+n_{-,-}$ for the real data was compared to the null-distribution of T from the scrambled

data. The calculated P-value was adjusted by the Benjamini-Hochberg method and the threshold set at p = 0.05.

Functional analysis was performed based on GO-terms by a method implemented in the GOstats package (Falcon & Gentleman 2007): GO-terms overrepresented among differentially expressed genes were determined and pooled for contrasts of the same tissues. The number of tissues for which a GO-term was found overrepresented was counted. Additionally GO-terms overrepresented among the genes identified as differentially expressed in 5 or more tissues were determined.

Only for liver-datasets genes were determined that were significantly differentially expressed in at least 3 datasets and differentially expressed with aging in the other direction in at least 1 out of 5 independent liver-datasets on aging.
Results: Swindell found that CR in most cases had an effect on less than 5% of genes, with the maximum found to be 23% in one study.
Clustering showed that the datasets in first instance clustered according to tissue type, but also different datasets from the same study were likely to cluster (even when from different tissues). The intersection between differentially expressed gene sets was around 30% or less, however commonly greater than expected by chance.

Among all tissue types examined, CR most commonly led to upregulation of genes involved in lipid metabolism and metal ion response, and downregulation of genes associated with immunity and protein folding.
16 genes were found over- and 12 underexpressed in 5 or more different tissues. Among the overexpressed were two metallothionein genes (*Mt1* and *Mt1*) involved in stress response (Thirumoorthy et al. 2007) and two period homologues (*Per1* and *Per2*) recognized for their role in manipulating the biological clock, but that also exhibit tumor suppression activity (Cheng Chi Lee 2006). Two procollagen (*Col1a1* and *Col3a1*) genes were found among the underexpressed. GO-terms enriched among these were nitric oxide mediated signal transduction (GO:0007263), zinc ion homeostasis (GO:0006882) and circadian rhythm (GO:0007623) for over- and response to heat (GO:0009408), unfolded protein (GO:0006986), biotic stimuli (GO:0009607), chemical stimuli (GO:00042221) and response to pest, pathogen and parasite (GO:0009613) for underrepresented genes.
Igf1 and *mTOR* each were only found differentially expressed in three contrasts and *Sirt1* in none.

GO-terms enriched among genes differentially expressed with CR and in the opposite direction for aging in liver were electron transport (GO:0006118) and cellular metabolism (GO:0044237).

3.1.3.1.1 Further analysis by Swindell, 2008b Swindell´s publication "Genes regulated by caloric restriction have unique roles within transcriptional networks" (Swindell 2008b) is a continuation of the study presented in Swindell, 2008a, in which 16 genes were identified as consistently up- and 12 as downregulated.
Overrepresentation of transcription factor binding sites in the genes enriched for differential expression with CR were determined by sequence analysis of the 500 bp upstream promoter region using the CisView database (http://lgsun.grc.nia.nih.gov/cisview/) (Sharov et al. 2006).
Furthermore a co-expression analysis was performed each: In brief, co-expression of each gene was determined from a large number of microarray measurements by Pearson correlation coefficients for each pair of gene. For each gene the magnitude of its absolute correlation coefficients indicated its connectivity strength. Local (strong) connectivity patterns were calculated as an average over the top absolute correlation coefficients for each gene, non-local (weak) connectivity patterns as the correlation coefficient at a certain high percentile.
Results: Enriched transcription factor binding sites in mouse were:

- for overexpressed genes:
  - TF_MIF, TF_STAT, TF_ZIC, TF_HEN1, TF_HNF4, TF_SREBP, TF_OLF1, ADD_MTF1A, ADD_MTF1B, MIT_051TATA, TF_MYB, ADD_PAX8 for metallothioneins
- for underexpressed genes:
  - ADD_PAX8, TF_NFY, TF_MAZR, TF_MZF, MIT_013LEF for immunity related genes
  - TF_MAF, TF_MYB, TF_MEIS, TF_NFKB for collagen related genes

Swindell also showed that in mice the connectivity of genes determined as enriched for downregulation with CR was high for local network regions, however for those for upregulation it was low for both local and non-local network regions.

3.1.3.2 Swindell, 2009

In his 2009 study "Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse" Swindell meta-analysed microarray data on CR of 17 different mouse tissues from 40 experiments (Swindell 2009). Most of the datasets used in this study plus some additional were also used in our meta-analysis. GSE11845 was not used in our study, since it is based on intermittent fasting, not classical CR. The LIMMA package for linear model analysis (Smyth 2004) was employed to determine the p-values for differential expression of each probe within the datasets. Fisher´s inverse chi-square approach was used for each gene to first combine different datasets of the same tissue (if more than one dataset present) and again to combine the p-values obtained from this over all tissues. Due to the large number of genes found this way a threshold for the number of tissues in which a gene had to be differentially expressed was set for further analysis (GO-analysis, mapping to KEGG-pathways). This introduces a value-counting component into the analysis.
Co-expression analysis was performed similar to that in Swindell, 2008b and genes clustered by their co-expression into modules of 2, 3, 5, 10, 20 and 40 genes. Each module was then scored for the differential expression of the genes contained based on their single-study p-values and the significance assessed by scrambling.
Results: Overall 29.7% (6330) of the genes were up- and 27.6% (5884) downreguated over different tissues. The gene significantly upregulated in most tissues was *Sgk1*. As in Swindell´s previous meta-analysis (Swindell 2008) *Mt2* was found up- and *Serpinh1* downregulated when combining evidence from different tissues.
Genes most strongly increased by CR across tissues were associated with the KEGG-pathways fatty acid metabolism, citrate cycle, PPAR signalling, oxidative phophorylation, amino acid degradation and metabolism, circadian rhythm, renal cell carcinoma, fatty acid elongation in mitochondria and the insulin signalling pathway . Genes commonly down regulated by CR were associated with focal adhesion, antigen processing and presentation, ECM-receptor interaction, DNA replication, MAPK signalling, cell communication, VEGF signalling and natural killer cell mediated cytotoxicity ($P < 0.01$).
A total of 3, 5, 22, 39 and 28 significant CR-responsive modules with 3, 5, 10, 20 and 40 genes, respectively, were identified.

3.1.3.3 Hong, 2010

In "Revealing system-level correlations between aging and calorie restriction using a mouse transcriptome" Hong performed GO-, co-expression and transcription factor binding site analyses (Hong, S. et al. 2010).
Datasets from 6 different studies, comprising 5 tissues were used. Within single studies differentially expressed genes were identified by unpaired two-class analysis using significance analysis for microarray (SAM) (Tusher et al. 2001). No analysis was conducted to detect enrichment of differentially expressed genes over the studies, but all genes found differentially expressed in any study were considered as "CR-transcriptome". The number of times a GO-category was found associated with the genes differentially expressed with CR was compared to the number it was found associated with any of the genes in the study using chi-square analysis. Co-expression analysis was based on correlation coefficients calculated from 131 microarrays from GEO and transcription factor binding site analysis was performed using TRANSFAC (Hinrichs et al. 2006). The relevance of the determined transcription factors was assessed by testing if they were significantly co-expressed with genes found differentially expressed with CR.
Results: GO-terms found enriched in the CR-transcriptome (up- and downregulated genes) were immune response, lipid metabolism, response to stimulus, cell proliferation, glucose catabolism, cholesterol metabolism, angiogenesis, cell adhesion, cell cycle, electron transport, muscle development, cytoskeleton organization, chemotaxis, amino acid metabolism and as for compartments extracellular space, lysosome, mitochondrion and endoplasmic reticulum. The co-expression modules from the aging transcriptome showed strong correlations with the CR-results in both metabolism (e.g., citrate cycle and lipid metabolism) and the immune response. Binding sites for 12 transcription factors were found overrepresented in upregulated genes (v-Myb, HNF-4α, TAL1, E4BP4, HLF, CCAAT box, FOXO1, MAZ, VBP, Tal-1alpha:E47, HNF-3β (FOXA2), Max) and 5 in downregulated (IRF-1, Pax, PAX6, YY1, NKX3A) however non of these was significantly co-expressed with its target genes.

3.1.4 Overview of our study – value-counting approach

In order to better understand the individual steps of our meta-analysis described further below a short overview of the concept is given here: In large our meta-analysis follows the 7 step approach proposed by Ramasamy (Ramasamy et al. 2008):

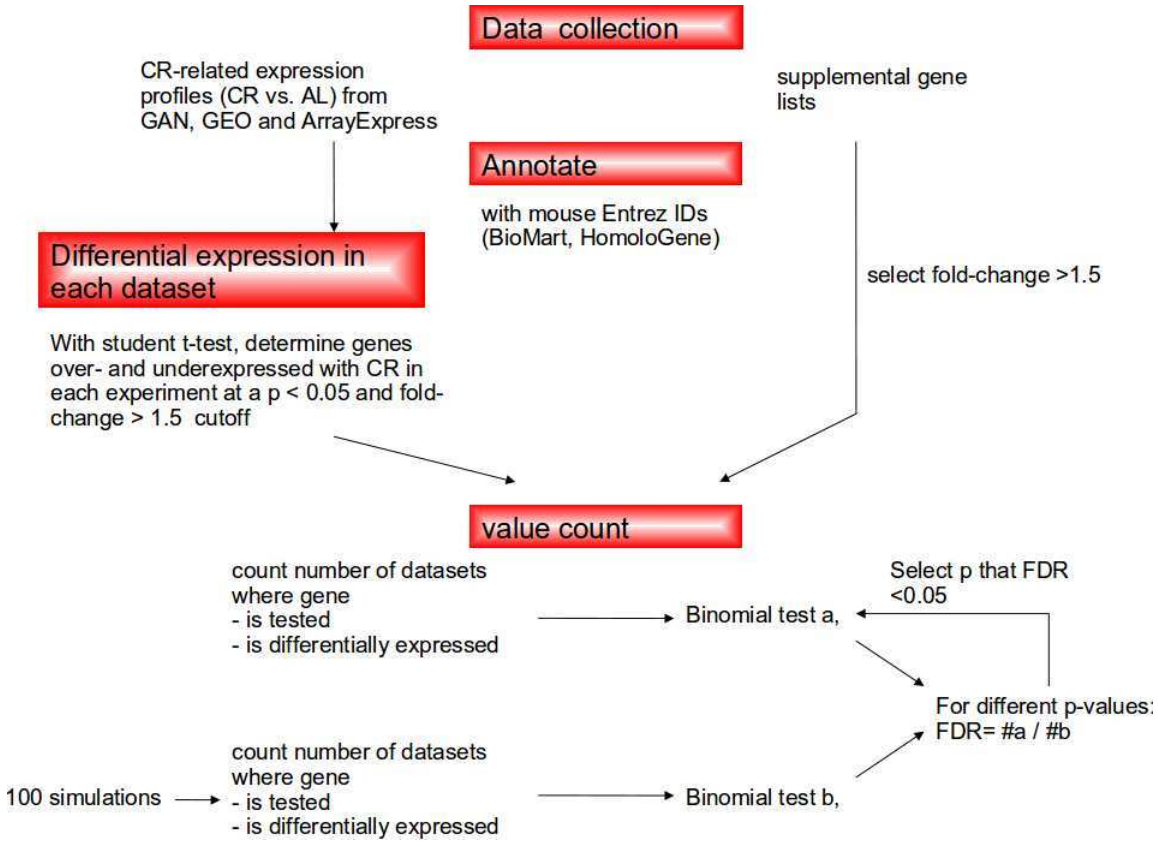


Figure 3.2: Simplified overview of the meta-analysis work-flow. See text for details.

1. Identify suitable microarray studies
2. Extract the data from studies
3. Prepare the individual datasets
4. Annotate the individual datasets
5. Resolve the many-to-many relationship between probes and genes
6. Combine the study-specific estimates
7. Analyze, present, and interpret results

The statistical analysis of our meta-analysis is based on a value-counting approach, i.e. we counted the number of times a gene is found over- / underexpressed in different datasets and determine the probability that this is due to random chance using a binomial test. The threshold for the p-values of the binomial test is determined by repeating the analysis on scrambled data and chosen so that the associated false discovery rate (FDR= mean number of genes significant at this cutoff after scrambling / number of significant genes on unscrambled data) is acceptably low. The principle of the study is therefore similar to that in Magalhaes 2009.
Datasets for our meta-analysis are mainly created from probe-level microarray data from which CR – AL pairs for the same co-variates are extracted. Differential expression for each gene is determined by an unpaired student t-test. Since for a non-negligible number of studies expression data could not be obtained we also included the information on differential expression for lists of genes determined by the original studies.
See fig. 3.2 for a simplified work-flow of the meta-analysis.

3.1.5 Aims of our study

As for other meta-analyses of microarray data this study aims to find genes which are detected as significantly differentially expressed with the increased sample size after combining studies, but are not found in the individual studies. For example Choi et. al (Choi et al. 2003) define integration driven discovery (IDD) as finding a gene differentially expressed in the meta-analysis, but in none of the underlying studies. The integration driven discovery rate (IDR) is the number of such genes divided by the total number of discoveries and is about 44-63% in their study. IDD-genes are therefore such with small but consistent differential expression for which the sample sizes in individual studies was too low for them to be detected to be significant. Therefore the statistical power of the meta-analysis is increased compared to the single studies (the false negative rate is lower) at the same false positive rate. On the other hand the higher statistical power would also increase the significance threshold and therefore reduce the Type I error.
Admittedly in a value-counting approach the level of differential expression in the original study must be high enough that the gene is found differentially expressed in the first place, however thresholds in our study for defining a gene as differentially expressed are more relaxed here than in the original studies.
By including data on a wide range of organisms, tissues and other co-variates we eliminate idiosyncrasies between studies and aim to detect genes differentially expressed with CR under different conditions (even though a sufficient number of detections can also be reached from one frequent organism or tissue). It was shown by Dhabi (Dhahbi et al. 2004) that different genes change their expression after different time-spans of CR. Since we also include data from experiments using a wide range of time-spans our analysis is likely to identify genes that change their expression quickly and stably.
The genes enriched for over- / underexpression serve as candidate genes for further studies, can be examined for an already known role in CR or aging or can be searched for enrichment of transcription factor binding sites. The network of genes can be extended by determining genes co-expressed with them.
Information on functional categories associated with CR can then be retrieved by both detecting enrichment of such categories among the candidate genes or by repeating the described analysis on functional terms instead of genes. A term would in this case be considered over- / underexpressed if the associated gene is over- / underexpressed.

3.2 Materials and methods

3.2.1 Microarray studies used in the meta-analysis

To obtain high-throughput data on caloric restriction we searched the databases "Gene Expression Omnibus" (GEO; from NCBI), "ArrayExpress" (from EBI) and "Gene Aging Nexus" (GAN) for the terms "caloric restriction", "calorie restriction" and "dietary restriction". We further checked other meta-analyses of CR for further datasets for which we requested expression data from the authors of the studies.
For studies for which gene expression data from none of these sources was available we attempted to retrieve published lists of genes differentially expressed according to the statistical criteria in the original study.
The only high-throughput data found were from microarray experiments. Since almost no non-mammalian data were among the studies found and mammalian data are more likely to resemble the situation in humans we decided to focus this meta-analysis only on data from mammals. Data were furthermore excluded if we could not extract data from one group being on CR a corresponding one on AL or high caloric, but otherwise comparable diet with no other differences between the groups. CR is here defined as restriction in the amount of calories consumed without malnutrition. One study comparing humans before and after bariatric surgery (GSE9157) was excluded since it was not clear how much nutrient uptake was restricted by this measure and if it could therefore be defined as CR. Another study on humans (GSE11975) comparing gene expression data from people during diet and the following weight-maintainance period was also excluded since the dietary setups could not clearly be defined as AL vs CR. Finally datasets were not used if the experiment was accompanied by the application of drugs or infection of the animals (GSE15344).
We further checked that the microarray platforms used in all studies were a unbiased representation of the transcriptome and not e.g. representing only selected pathways.

3.2.1.1 Studies for which expression data could be obtained

For the 23 studies shown in table 3.2, expression datasets could be obtained. That means the preprocessed (i.e. background subtracted and normalized) microarray signals for the conditions of interest were given for all probes on the array except when excluded for low quality.

	Subsets used	organism	duration	(end) age	tissue	amount of food
GEO						
GDS1261; (Tsuchiya, 2004)	Ames dwarf and normal mice	Mus musculus	4 months	6 months	liver	90% of the AL intake* of animals of the same genotype for 1 wk, to 80% for the next week, and to 70% for the rest (*average amount consumed daily by AL mice during the preceding week);
GDS1808; (Dhahbi, 2005)	CR8-AL and LTCR-AL; have CON in common	Mus musculus	CR8: 2month; LTCR: 17 months	22 months	liver	CON 93kcal/wk; LTCR: 52.2; CR8: 77 for 2 weeks, 52.2 for 6 weeks
GDS2612; (Edwards, 2007)	25 months old	Mus musculus	23.5 months	~25 months	skeletal muscle	CON 84kcal/wk, CR 26% less (62kcal/wk)
GDS2681; (Someya, 2007)	15 months old; excluded: 4 months: CR missing	Mus musculus	3 months	15 months	cochlea	CON 84kcal/wk, CR 26% less (62kcal/wk)
GDS2961 + GDS2962; (Lustig, 2007)	6, 16 and 24 months old; excluded: 1 months old: CR missing	Mus musculus	11, 41 and 83 weeks	6.5, 13.5, 24 months	thymus	Up to 13 weeks of age, 100% regular feed, followed by 90% fortified feed for 1 week, 75% for 1 week, then 60% fortified feed after that until the age at which the mice were sacrificed
GDS355 + GDS356; (Kayo, unpublished)	>30 months old; excluded: 5 months old; CR missing	Mus musculus	?	> 30 months	kidney	?
GSE11244; (Estep, 2009)	FHC-CR, TAL-CR; have CR in common	Mus musculus	14 days	9.5 months	liver	true ad libitum: as much as wanted (about 125kcal/wk); CR: 73kcal/wk; fixed high cal: 110 kcal/wk
GSE11291; (Barger, 2008)	3 tissues; excluded: 5 months: CR missing	Mus musculus	16 months	30 months	Heart, neocortex, gastrocnemius	CON: 84 kcal/week, CR: 63 kcal/week
GSE14202; (Padovani, 2009)	exercise and non-exercise	Mus musculus	6 weeks	4 months	mammary gland	30% restriction

GSE18297; (Saito, unpublished)	1 week or 1 month CR; 5, 10, 20, 30% food restriction; same controls for different restriction levels	Rattus norvegicus	one week or one month	1.5 or ~2 months	liver	5, 10, 20, 30% restriction
GSE6110; (Chen, 2007)	24 months old; excluded: 4 months old: CR missing	Rattus norvegicus	22.5 months	25 months	kidney	CR begins at 10 wk, 10% restriction until 15 wk where it is increased to 25 and to 40% at 4 months
GSE6718; (Linford, 2007)	2 tissues; excluded: 4 months	Rattus norvegicus	20 months	24 months	Heart and Adipose Tissue	60% of AL
GSE7502; (Sharov, 2008)	2 tissues; ages: 6, 16, 24 months; excluded: 1 month: CR missing	Mus musculus	2.5, 12.5, 20.5 months	6, 16, 24 months	Testis and Ovary	40% restriction
GSE8426; (Xu, 2007)	5 tissues; 6, 16, 24 months; excluded: ~1mo samples: CR missing	Mus musculus	2.5, 12.5, 20.5 months	6, 16, 24 months	Cerebellum, Hippocampus, Spinal Cord, Striatum, Cortex	at 14 weeks of age at 10% restriction, and then changed to 25% at 15 weeks and 40% restriction at 16 weeks onward
GSE9917; (Larrouy, 2008)	no subsets	Homo sapiens	4-wk very-low-calorie diet, a 3–6-wk low-calorie diet, and a 4-wk weight-maintenance	~27-48 years	skeletal muscle	4 weeks: 3.3 MJ/d, 3-6 weeks: 4–5 MJ/d, 4 week: 5.8 MJ/d
GSE17309; (Fernández, unpublished)	no subsets	Sus scrofa	211 days	~7 months ?	liver	25% restriction
GSE12853; (Connor, 2010)	timepoint 1d before realimentation	Bos taurus	12 weeks; 8 weeks realimentation	11 months	liver	60-70% of AL

GSE241; (Massaro, 2004)	2, 4, 12h CR; excluded: other timepoints: miss controls	Mus musculus	2-96h; timepoints >12h miss controls; there is a 14/15d timepoint in the paper, but not in the file	adult	lung	reduction by 66%
GSE9121; (Pohjanvirta, 2008)	adipose tissue: 10 d CR; liver: 4 and 10 d CR (controls for timepoints pooled)	Rattus norvegicus	liver: 4 or 10 days; WAT: 4 days	~3-4 months	liver, adipose tissue	restricted: 4 day: 18–12–9–6 g; 10 day: ad libitum–16–14–11–8–6–4–4–2–1 g
GSE904; (Becker, unpublished)	~170 d old; excluded: 17d old: no CR	Mus musculus	?	?	liver	?
GAN						
Expression Profile of Aging and CR Retardation, Neocortex; (Lee, 2000)	30 months old, Neocortex; excluded: Hippocampus: CR missing, 5 months: CR missing	Mus musculus	28 months	30 months	neocortex	26% less than AL
Array Express						
E-MEXP-748; (Selman, 2006)	4 tissues	Mus musculus	16 days	~4 months	liver, skeletal muscle, colon, hypothalamus	reduced to 90% of AL mice at 14 wk, 80% at 15 wk, and 70% at 16 wk of age
Provided by						
Hu; (Wu, P., 2008)	no subsets	Mus musculus	4 months	8 months	forebrain	70% less than AL

Table 3.2: Microarray studies and their characteristics used in the meta-analysis. Fields underlied in red were excluded at later steps. The "subsets used"-column gives subsets of the dataset used for the meta-analysis and which were excluded and why; AL: ad libitum, CR: caloric restriction, LTCR: long term CR, CR8: 8 weeks CR, CON: control, WAT: white adipose tissue

Two of the studies were excluded in the course of the meta-analysis as described later. From each study one to fifteen datasets / subsets were extracted, so that we obtained a total of 61 datasets. Data in each subsets consisted of AL and CR samples from animals of the same age and CR setup and of the same tissue. The only co-variate for which we did not split data into different datasets was sex, since we did not expect a large

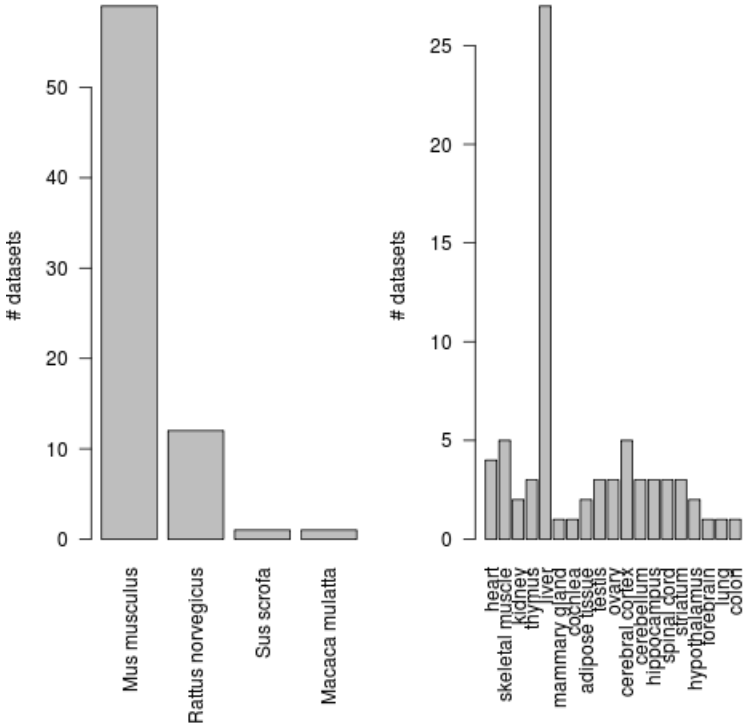
difference in the effect of caloric restriction between male and female animals and we did not want to reduce replicate numbers of each dataset more than necessary. Also the number of subsets of individual studies should not get too large, since this study would gain too much influence in the meta-analysis.The vast majority of 48 datasets was from mouse (Mus musculus), 12 from rat (Rattus norvegicus) and one from pig (Sus scrofa). These include different strains of mice and rats. The biggest group for the tissue co-variate was liver (18) and brain was represented by many different tissues. In the list of 19 different tissues 6 are represented by only one dataset. The duration of CR ranged from less than one day (5 datasets) to 23.5 months and the ages at which tissues were obtained from 1.5 to over 30 (exact age unknown) months for mouse and 1.5 to 24 months for rats. Histograms of the distribution of datasets over these co-variables after including datasets for which expression measurements could not be obtained are shown in fig. S.1.

3.2.1.2 Studies for which expression measurements could not be obtained

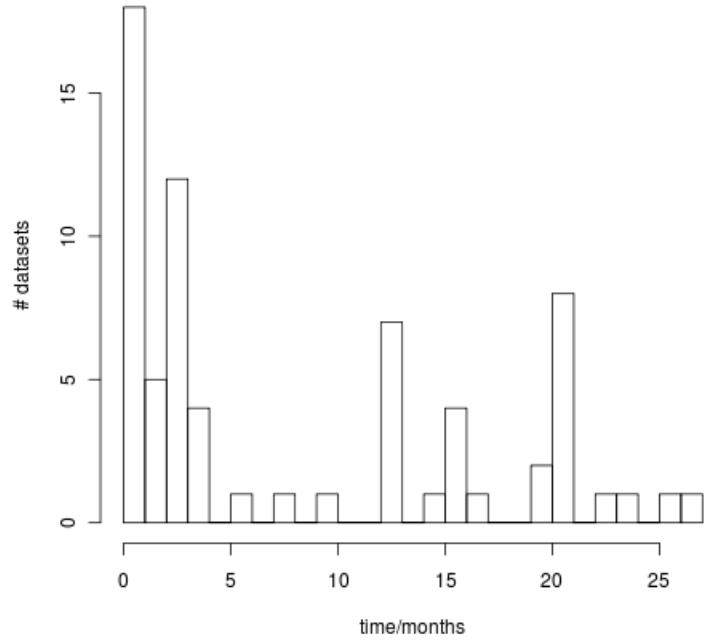
For the following studies the microarray signal intensities for all probes was not available, but rather lists of genes found differentially expressed by the statistical method used in the original study. For some of them p-values and / or effect-sizes were given. We requested expression data from the (corresponding) authors of these studies, but were not able to obtain them. Some of the studies were eventually not used for the reasons described.
(Fu et al. 2006): Genes differentially expressed according to a t-test assuming equal variances at a Benjamini-Hochberg FDR adjusted p-value <0.05 in heart, liver and hypothalamus could be obtained from the supplementary materials of the corresponding publication. The data are from 4-6 months old male mice in which CR animals were restricted to 60% of caloric intake of AL-animals for 2.5 to 4.5 months.
(Wu, P. et al. 2009): A list of differentially expressed genes in the hypothalamus of caloric-restricted vs. ad libitum fed animals was kindly provided to us by the authors. This dataset had to be excluded later on due to annotation problems (see: "3.2.3 Processing gene lists from studies for which expression data were not obtained).
(Higami et al. 2004): Data from Higami were not used since only selected genes differentially expressed with CR could be found in the paper or its supplement. Allowing lists of genes selected for particular criteria would introduce bias to our work.
(Cao et al. 2001): Data for genes differentially expressed with CR at a 1.7-fold change criterion were listed in the corresponding publication. Only data for genes differentially expressed in CR, but not differentially expressed with age in the opposite direction in the same study were used for reasons described in "3.2.2.5 Excluding genes differentially expressed with age". Furthermore we only used data for long term CR, but excluded data for short time CR since the control used in the paper was not age matched. Data were given for liver of 7 and 27 months old female mice of the long lived strain C3B10RF1 which had been on CR for 6 or 26 months respectively.
(Dhahbi et al. 2004): CR data from livers of male mice of the long lived F1 hybrid strain B6C3F1 were obtained from the corresponding publication. Data were obtained for 2, 4 and 8 weeks as well as 27 months of CR. CR of 77kcal/week for 2 weeks and 52.2 kcal/week afterwards (except mice on 2 week CR, which were one week on 77kcal/week and one week on 52.2 kcal/week) compared to 93 kcal/week for control animals was induced at an age so that mice were 34 months old at time of killing. Data from CR-mice were compared to data from 34 months old controls and a 1.5-fold change was considered significant.
(Corton et al. 2004): CR data from livers of mice on a SV129 background, caloric-restricted for 5 weeks were available in the supplement of the corresponding publication. Calories were reduced to 90% of the AL group for one week and 65% for another 4 weeks. All data of mice treated with chemicals were ignored. The threshold for significance was set at $p <= 0.001$ with Bonferroni correction and a at least 1.5-fold change in expression was required.
(Lu et al. 2007): Data from Lu were excluded since mice were treated with TPA, a diacylglycerol mimetic and tumor promoting substance.
Data from Wong, 2002 comparing gene expression in the liver of male C57BL/6 ad libitum fed mice to such restricted to 60-70% of their caloric intake could not be obtained from the corresponding publication or supplementary data. A link in the paper that is supposed to direct to the expression data was not functional.
(Kayo et al. 2001): Kayo provided data on differential gene expression in skeletal muscle of rhesus monkeys on CR for 9 years and sampled at an age of around 20 years. The threshold was selected so that the average fold-change had to exceed 1 standard error from a 1.3-fold change.

Eleven gene lists were created from these studies in addition to the 63 created by analysing gene expression measurements by ourselves. After combining these data more than half of the now 74 datasets were from mice and more than one third from liver. The distribution of the number of datasets over different co-variates is shown

in fig. 3.2.



CR-duration (mouse+rat)



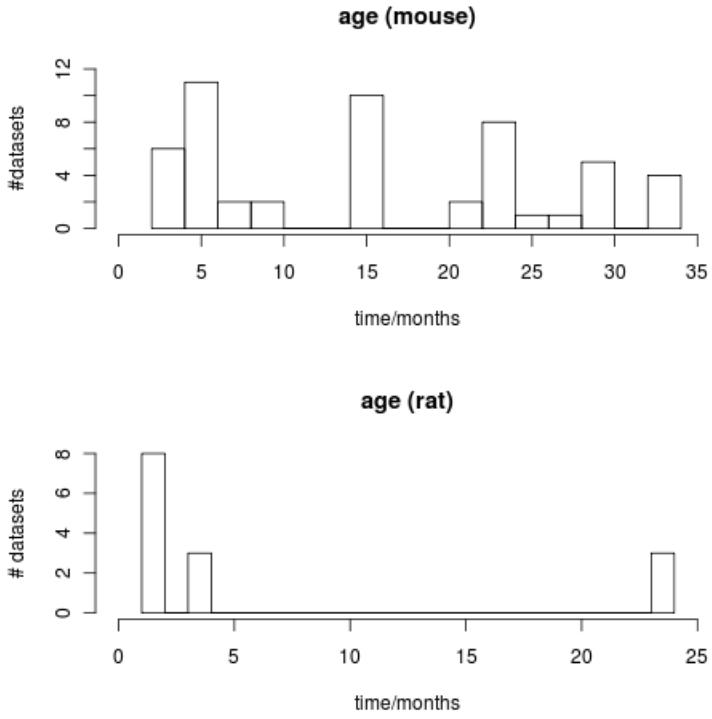


Fig. 3.2. Datasets used for the meta-analysis in terms of co-variate organisms, tissues, duration of CR (mice and rats only) and age of sampling for (mice and rats). Datasets for which the value for a co-variate is not given are not shown.

3.2.2 Analysing gene expression data from complete datasets

3.2.2.1 Obtaining and assembling microarray data files

3.2.2.1.1 Obtaining files from GEO

3.2.2.1.1.1 Downloading files and selecting samples GDS (GEO dataSet) and GSE (GEO series) files were obtained from NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/; (Barrett et al. 2009)) and processed in a similar way using R (R Development Core Team 2009). In supplement2 metaan_R_GDS.txt and metaan_R_GSE.txt examples for selecting and processing samples from normal (as opposed to Ames dwarf) mice for GDS1261 and heart tissue samples for GSE11291) are attached. All GEO files exist in a preprocessed form, i.e. they are background subtracted and normalized. The files were downloaded from GEO and converted to ExpressionSet objects using the GEOquery Bioconductor package (Gentleman et al. 2004) (Sean & Meltzer 2007). Samples that differ only in calorie intake (i.e. caloric restriction vs. control conditions) but keeping all other observed variables constant (e.g. only from one tissue type or age group) were selected by pattern matching on the "description"- or "title"-variable (or in rare cases also other variables like "age") of the ExpressionSet created from GDS or GSE files respectively.
The only other variable (except calorie intake) for which we did not split the data in different files according to the value of the variable was the sex of the animals. We did not make a difference between samples from males or females, but checked that the distribution between male and female for CON and CR within each dataset did not differ significantly.

In GSE9121 data for liver samples after 4 or 10 days of caloric restriction were given together with data for their controls. We created data sets for 4 and 10 days of CR, however used a combination of the controls for both time points for both of them to increase statistical power. We considered this justified since we did not expect major gene expression changes due to a 6 day difference of age for rats which are 11-15 weeks old. In all other cases CR datasets of one time point would only be compared to CON datasets of the same time point.
If available the Entrez Gene ID, Ref. Seq. Transcript ID and Gene Bank accession number for each probe were

obtained from the GPL-file for the microarray platform used in the corresponding experiment. A tab-delimited file was created appending the expression value columns to these three annotation columns. The column names for all control samples were "CON" and CR-samples were "CR".
Most experiments of this study used one-color microarrays, so that separate values for controls and caloric restriction (CR)-samples were given. For the only microarray where CON and CR samples were measured on the same (two-color) array, GSE9917, data were given as ratios of Cy5:Cy3. Relevant values were extracted and stored in this form. Unfortunately this dataset had to be excluded subsequently due to annotation problems (see: "3.2.2.2 Mapping non-mouse Entrez IDs to mouse Entrez IDs").

3.2.2.1.1.2 Binding annotation with different number of lines to expression values In most cases the number of probes in the GSE files matched the number of probes in the corresponding GPL-file since the GSE files contained all probes including those with low signal etc. In the case of GSE7502 probes were excluded from the GSE-file, but the annotation contained all probes of the array. Therefore a table containing the annotation and another containing the expression values were saved in files and expression values bound to their annotation via the common identifier "ID" using Perl (vlookup_mod4_3.pm and use_vlookup_mod4.pl in supplement 2).

3.2.2.1.1.3 Combining files separated into different list elements In the cases of GSE7502 and GSE904 the GSE-file was downloaded to R as a list with two instead of only one element, because different microarray platforms (GPL2552 and GPL4358) or parts of the same platform (GPL738 and GPL782) were used. For GSE7502 different versions of a microarray chip were used. The same number of CON-samples as CR-samples was tested on each platform, so that we did not expect a bias from the use of different platforms.
In this case we bound the annotation in the corresponding GPL-file to each list element, wrote these tables to files and further processed them in Perl by combining values with identical Entrez ID or if no Entrez ID was given for a probe by GeneBank accession number. Since by manual inspection we did not find any case in which a different number of probes for one gene existed in one than the other array (and would therefore lead to lines with values for only some samples, which would most likely lead to discarding the line in further analysis) we did not care which probe of one gene is linked with which probe for the same gene on the other platform. Due to the way collapsing of probes targeting the same gene was done later it would not cause trouble if columns of different probes targeting the same gene were linked here (see "3.2.2.3 Collapsing probes targeting the same gene").
This linking was done with all samples relevant for our analysis and samples corresponding to the same CON – CR pair (e.g. one pair for 6 months caloric restricted animals versus their controls another for 16 months restricted animals versus their controls) were extracted to one file each manually using Excel.
For GSE904 the list of probes in the first file was continued in a second one. Each file was therefore treated as an independent one and then both combined by binding the rows together. In the case of GSE8426 four list elements were obtained since the probes were distributed to two different platforms (GPL738 and GPL782) and for each of those the samples were distributed to two files each. Therefore the samples divided to different files were combined by binding the columns together as for GSE7502 and the probes from the different platforms by binding the rows as for GSE904 after adding the corresponding annotation.

3.2.2.1.1.4 Detecting and reversing transformation of data Since the values in some of the datasets were transformed (mainly log-transformed), but were not in others and we wanted to calculate comparable effect-size measures for all the datasets the transformation of transformed data was reversed: To determine if values in the GEO data files were transformed the value of the value_type field of GDS files or the data_processing field of GSM files corresponding to GSE files were obtained. The value_type "count" tells that there was no transformation done on the data, the value_type "transformed count" indicates some kind of transformation. The data_processing field gives information by which algorithm/software the data were processed so that in doubt it can be found out if this method applies transformation. Furthermore the mean of all samples was calculated for each probe and the median value of these means used as an estimate. E.g. if it was above 10 this supported that there was no log-transformation. For further indications we used the histogram of the sample means, which e.g. indicate log-transformation if values below 0 appear. (For GDS files these further criteria were only used if value_type did not give back "count"). In doubt we checked if it was likely to obtain the given values the way described in the GEO-files from the raw data without log-transformation or contacted the authors.

3.2.2.1.1.5 Handling non-globally normalized data We aimed at creating files with untransformed values which were between-array normalized by global normalization, i.e. adjusting the median (or mean) of all

signals to the same value for all arrays. We did not expect that different ways of normalization would critically impact our p-value and effect size calculation, however there were cases when normalization was intermingled with log-transformation, so that log-transformed could not be reversed easily. For GDS2961/GDS2962 and GSE8426 data were first log10 transformed, then normalized to a mean of 0 by subtraction and then the z-score of the probe in the distribution of all probe signals was calculated (z-score normalization; (Cheadle et al. 2003)).
Since in this case also "RAW" values were given in the separate GSM-files, these files were downloaded and the "RAW" (background subtracted and within-array normalized) values of each sample divided by the mean expression value over all probes and multiplied by the grand mean (mean of the means over all samples relevant for our analysis), which resembles global normalization.
In GSE11244 the Cy3-signal was normalized to the Cy5-signal of Stratagene Universal Mouse Reference RNA in a two-color hybridization and the result log2 transformed. We reversed the log2 transformation by raising the value to the power of 2, but we accepted this way of normalization and expected similar p-values and effect sizes than for globally normalized arrays, even though the values in this file were lower (distributed around a mean of 1).

3.2.2.1.1.6 Combining datasets corresponding to the same experiment In cases where two files for a single experiment existed (the probes of one microarray were divided to these files; e.g. GDS2961 + GDS2962 and GDS355 + GDS356) we combined the files before continuing the analysis in Perl.

3.2.2.1.2 Obtaining data from Gene Aging Nexus (GAN) In GAN (http://gan.usc.edu/public/index.jsp; (Pan et al. 2007)) "Expression Profile of Aging and CR retardation, Hippocampus" and "Expression Profile of Aging and CR retardation, Neocortex" were the only studies on CR not found in GEO. Unfortunately the Hippocampus entry only contained data from controls and was therefore of no use for us. The Neocortex data were downloaded manually for 30 months old animals only since for 5 months old no CR group existed. Column names were changed to "CON" and "CR".

3.2.2.1.3 Obtaining data from ArrayExpress E-MEXP748 was the only file that had to be obtained from ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/; (Parkinson et al. 2009)), since no dataset corresponding to this study was found in GEO. We used the ArrayExpress Bioconductor package (Kauffmann et al. 2009) for obtaining these data. In contrast to most GEO-files the annotation in the .adf file was not in the same order as the probes in the file containing the expression values, so that the columns of the files could not be directly bound together. Instead both files were sorted according to the column containing the probe IDs before binding annotation columns to the expression values.

3.2.2.1.4 Obtaining and processing data directly from authors For all CR microarray studies we knew about that could not be found in one of the databases we contacted the (corresponding) author and requested the data. Unfortunately Hu was the only one to provide these (Wu et al. 2008). (For another study (Wu, P. et al. 2009), for which they could not supply the original data we obtained a list of differentially expressed genes. We tried to include it in "3.2.3 Processing gene lists from studies for which expression data were not obtained", but had to drop it due to annotation problems).
For all studies for which we could not obtain expression data we searched for lists of differentially expressed genes in the corresponding publications and supplementary materials (see 3.2.3).

3.2.2.2 Annotating data with identifiers common between all data files

3.2.2.2.1 Aim and overview To integrate different datasets we needed the same kind of annotation for all of them. The annotation found in the gene expression databases (e.g. the GPL files in GEO) varies between database entries. Many of our datasets were annotated with Entrez IDs, for others e.g. only GeneBank accession numbers and Unigene IDs were available.
Since by far most of the datasets in this analysis were from mouse we aimed at displaying our results annotated with mouse Entrez IDs. We expected Entrez IDs to facilitate the mapping between different organisms, e.g. compared to Unigene IDs. We therefore conducted a gene-centered rather than a transcript centered analysis (which would be done i.e. when using Unigene IDs) and accepted to loose information from probes targeting sequences that do not correspond to annotated genes (or expressed sequence tags (ESTs)) or for which no homology mapping between the organism of the study and mouse existed (as of April 2010).

For this annotation with identifiers common between all data files we needed at several stages a program that looked up the given annotation in another file that matches this annotation to another one. We used use_vlookup_mod4.pl together with vlookup_mod4_3.pm (supplement 2).
vlookup_mod4_3.pm is a subroutine which takes character strings (common identifier) of a specified column of file 1, searches an exact match of this string in a user-specified column of file 2 and adds the value in another specified column of the same line to file1 (fig. 3.3).

If the strings you search for are comma-separated lists of elements the user can specify if he/she wants to search for the complete string or each element of the string individually. In the second case all found strings are combined to a comma-separated one. If the same string is found twice it will occur only once in this list. This list is added to file1 as the found string.
If a common identifier matches to more than one value in the second file, the user can choose if he/she wants to combine all found elements in a comma-separated list, create a new line for each or treat this situation as if nothing was found.
For all values in file 1 for which no corresponding value in file 2 is found the user can specify other columns of common identifiers several times. The program can be run on multiple files at once.

3.2.2.2.2 Adding Entrez IDs to mouse datasets where missing For some mouse data sets Entrez IDs were not available in the platform annotation. Annotation files matching mouse GenBank accession numbers and MGI Automatic Gene Symbol (or if appropriate other identifiers like Ensembl Gene ID, Unigene ID, RefSeq DNA ID, etc.) to Entrez IDs were downloaded from Ensembl (BioMart: Ensembl Genes 57: Mus musculus genes (NCBIM37); April 2010) (http://www.ensembl.org/; (Hubbard et al. 2009)). Entrez IDs were added by searching them in the annotation file by looking up which one matched the GenBank accession number (GB_ACC) in our data files and if not found, other identifiers. For this process we used use_vlookup_mod4.pl together with vlookup_mod4_3.pm. For probes annotated with more than one GB_ACC we obtained all available Entrez IDs. In later steps however we preferred signals mapped to Entrez IDs unambiguously to those with more than one Entrez ID (see: "3.2.2.3 Handling probes targeting more than one gene").
A certain number of lines in the datafiles (8926 of 19200) were lost during this process, e.g. for genes, which were not yet annotated with Entrez IDs.

3.2.2.2.3 Mapping non-mouse Entrez IDs to mouse Entrez IDs For datafiles from species other than mouse we added the Entrez IDs of the homologue mouse gene by searching for the given non-mouse Entrez IDs to obtain uniform annotation for all files. To do this we downloaded the HomoloGene data file (ftp://ftp.ncbi.nih.gov/pub/HomoloGene/current : homologene.data from 08/08/09) matching annotation of homologue genes between different organisms (Sayers et al. 2010).
Files were created containing only *Mus musculus* or the organism of interest´s data. We used use_vlookup_mod4.pl with vlookup_mod4_3.pm to first match all *Mus musculus* Entrez IDs with the annotation of the organism of interest into the same line using the homology group ID as common identifier and then again to add the mouse Entrez ID to the organism of interest´s datafile using this organism´s Entrez ID as common identifier. Again if more than one identifier was matched to a probe in the original file a comma-separated string of all correspondingly found mouse Entrez IDs was added. In later steps however we preferred signals mapped to Entrez IDs unambiguously to those with more than one Entrez ID (see: "3.2.2.3 Handling probes targeting more than one gene" and fig. 3.4)

Since we did not want any non-mouse gene in our analysis with homology to more than one mouse gene and therefore creating ambiguity, we deleted all homology groups comprising more than one gene in the mouse annotation file using only_one_allowed.pl (supplement 2). We however accepted if Entrez IDs of the organism of interest were in more than one homology group, i.e. if more than one of them matched to only one mouse Entrez ID.

This procedure was highly ineffective for the bos taurus dataset GSE12853 and would have lost all but one probe. Therefore we annotated this file in a different way that is described below ("3.2.2.2 Special annotation procedure for GSE12853"). For GSE6110 rat Entrez IDs were not given in the data file and the given Unigene IDs were not part of the HomoloGene files. Therefore the mouse Entrez ID was added first by looking up the corresponding rat Entrez ID in a rat BioMart file (BioMart: Ensembl Genes 57:/ Rattus norvegicus genes (RGSC3.4)) (as described for mouse in "3.2.2.2 Adding Entrez IDs to mouse datasets where missing") and

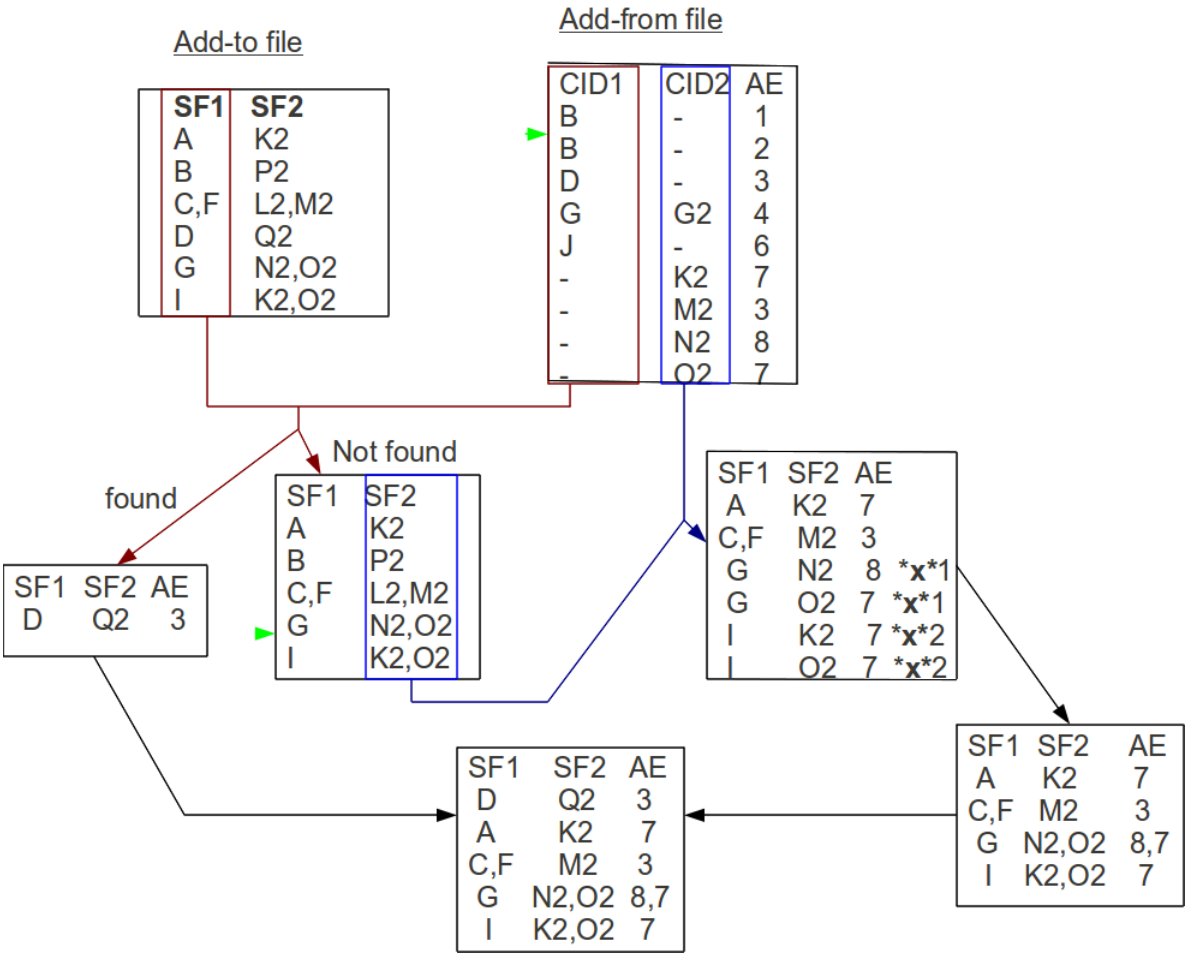


Figure 3.3: Workflow of the vlookup_mod4_3.pm subroutine with options used in this study. SF are columns containing values to search for, CID (common identifier) represents the corresponding value in the file to add from. AE is the element to add. The general flow is from top to bottom; First SF1 from the "Add-to" file is looked up in the "Add-from" file. For lines for which SF1 is not found SF2 is looked up. The results of both searches are combined; green arrowheads indicate examples for special situations: 1, common identifiers matched to different values to add: these values will be ignored; 2, comma-separated lists of elements to search for: individual elements of comma-separated lists are searched; *x*[number] is a marker for multiple lines created from one probe; these will be combined again to their corresponding probes; see text for details.

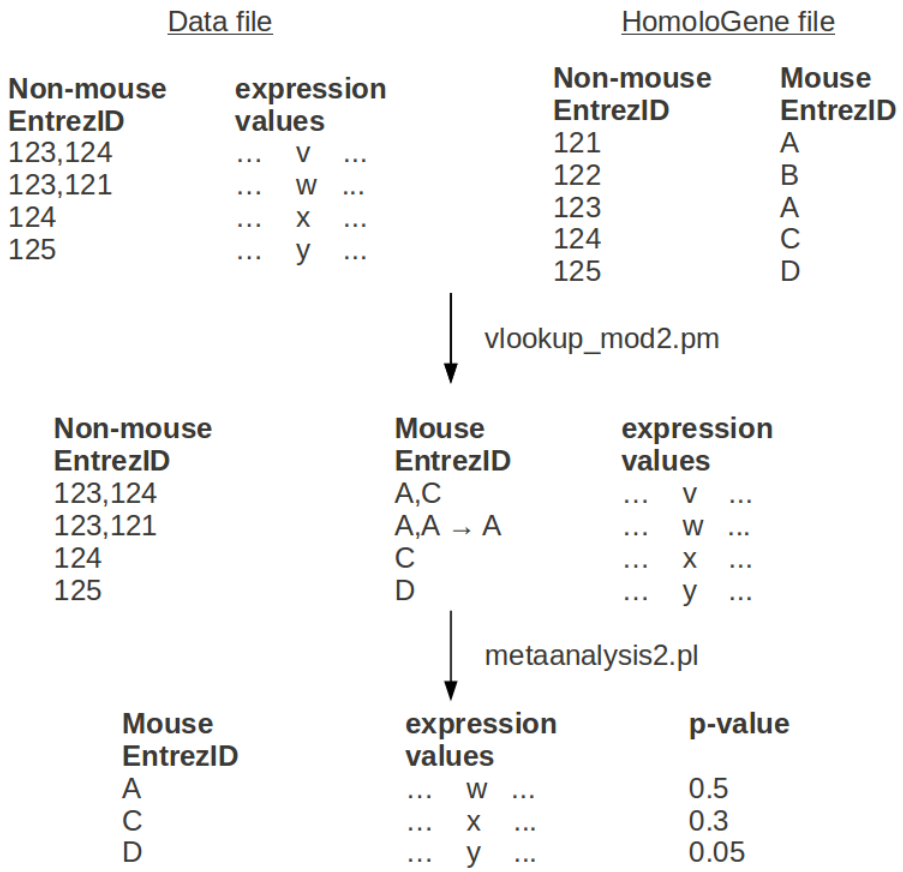


Figure 3.4: Example for the pipeline for adding mouse Entrez IDs to non-mouse expression values and processing of data. The way values are displayed does not resemble their real format. For clarity non-mouse and mouse Entrez IDs were depicted in different formats. Mouse Entrez IDs are added to the datafiles via non-mouse Entrez IDs as common identifiers employing the subroutine vlookup_mod4_3.pm. AA–>A indicates that two identical identifiers in one line are merged. Lines corresponding to the same gene are collapsed and p-values calculated as described in the text using meta-analysis_v3.2.pl. "Expression values" represents lists of expression values both for CON and CR in each line.

using this to find the mouse Entrez ID in the HomoloGene file. Similarly for GSE9917 the annotation in the file (GB_ACC) was not given in the file obtained from HomoloGene. Therefore we downloaded the BioMart annotation file for homo sapiens (BioMart: Ensembl Genes 57: / Homo sapiens genes (GRCh37)) intending to first add the human Entrez ID, which then should have been used to find the mouse Entrez ID. However only two of the GB_ACCs given in the datafile could be found in the BioMart file. We also were not able to obtain further annotation from the authors and therefore had to exclude the dataset.
Annotation of GSE17309 was more complex since this contained data from sus scrofa, for which no homology file was available at HomoloGene. Therefore we obtained homology information from BioMart in which however only Ensembl IDs (Ensembl Gene ID, Ensembl Transcript ID, etc.) were available. Therefore we also obtained the necessary BioMart annotation files on mouse and pig and first mapped the given pig-identifiers to pig Ensembl Gene ID, from there to mouse Gene Ensembl ID and finally to mouse Entrez ID. All but 357 of original 24123 (mainly poorly annotated) probes were lost during this process.

3.2.2.2.4 Special annotation procedure for GSE12853 Since we were not able to map mouse Entrez IDs to the steer data GSE12853 for the given annotation (GB_ACC, probe ID and Gene name) directly via files from HomoloGene, we tried to obtain Bos Taurus Entrez IDs first and map these to mouse Entrez IDs similar to what is described above for GSE6110 and GSE9917. However for no probe we first found bovine Entrez ID and then also the corresponding mouse Entrez ID. This was probably to the poor annotation of Bos Taurus GB_ACCs and Gene names with Entrez IDs.
To overcome this problem the authors (Erin Connor et al.) kindly provided us with further and more recent annotation. See fig. 3.5 for the annotation process using this file: Since this annotation only contained nucleotide RefSeq IDs and the HomoloGene file only protein RefSeq IDs a file was downloaded from BioMart matching bos taurus nucleotide to protein RefSeqIDs. The nucleotide RefSeq IDs were added to the HomoloGene file containing bos taurus protein RefSeqIDs mapped to mouse Entrez IDs (vlookup_homologeneSteer.txt) using a modification of vlookup_mod4_3.pm and the file was now called vlookup_vlookup_homologeneSteer.txt. (The modification of the program was necessary since RefSeqIDs contained version numbers in one file, but not the other).
The annotation provided by the authors was mapped to the experiment data via probe IDs specific for this experiment so that the data were annotated with nucleotide RefSeqID and GB_ACC identifiers (vlookup_vlookup _GSE12853.txt). Finally mouse Entrez IDs were added to this file from vlookup_vlookup_homologeneSteer.txt via the common nucleotide RefSeq ID identifier. Even with this procedure not more than one gene could be annotated. The same is true when searching for GB_ACC additionally to nucleotide RefSeq ID. Therefore the dataset was excluded form the analysis.

3.2.2.3 Processing datasets, performing a t-test and calculating effect sizes

After annotation the dataset files were further processed and t-test p-values and effect-sizes for the CR – CON comparison were calculated. These steps were done in batch-mode for all datasets using meta-analysis_v3.2.pl (supplement 2).

3.2.2.3.1 Handling missing values and annotation In each individual microarray experiment lines that contained more than 30% missing values or for which no Entrez ID annotation was found were eliminated. To facilitate subsequent analysis all remaining missing values were replaced by the row mean, i.e. calculated from values for control and CR samples. This procedure in general lowers the chance to find this gene differentially expressed, reflecting the doubts about it due to the missing value.

3.2.2.3.2 Collapsing probes targeting the same gene Probes targeting transcripts of the same gene (and i.e. having the same Entrez ID) were collapsed by using the mean over each probe (employing the Statistics::Descriptive CPAN package by Colin Kuskie and Shlomi Fish). That probes with higher value therefore contribute more strongly to the final result is justified by the assumption that probes with higher values are more reliable since they probably bind transcripts with higher affinity and their signal to noise ratio is higher. This procedure is therefore more conservative than e.g. the one used by Swindell (Swindell, 2009) which selects the most differentially expressed probe.

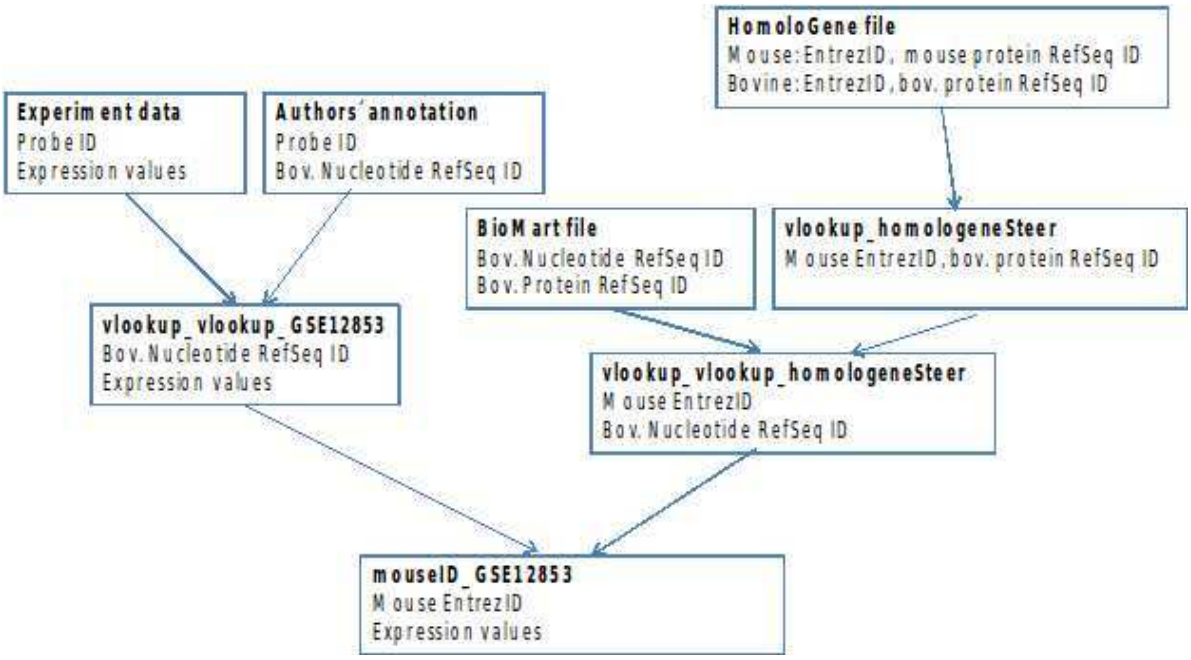


Figure 3.5: Annotation procedure for GSE12853. File descriptions or names are in bold print. See text for details.

Entrez IDs	expression value
1	a
1,2	b
3,4	c
3,5	d

--> collapsing -->

EntrezIDs	expression value
1	a
2	b
3	mean (c,d)
4	c
5	d

Figure 3.6: Illustration of collapsing of probes. Left table: before, right: after collapsing; each line represents a probe. The expression value of Entrez ID "1" is "a" since this is the only unambiguous mapping of a probe to "1". Since no probe is mapped unambiguously to "3", its expression value is the mean over the values of both probes mapping to it. Symbols do not resemble true formats.

3.2.2.3.3 Handling probes targeting more than one gene If probes mapped to more than one Entrez ID we ignored them if other probes existed which only mapped to this Entrez ID, but collapsed them if no such probes existed (Fig. 3.6). We preferred unambiguous probes since the expression values for a gene would not be disturbed by the expression values of other genes. On the other hand we preferred to use ambiguous values if there was no other option to loosing genes from our analysis, especially such with high homology to others, so that no unique probe for them existed.
This approach is therefore more conservative than expanding every entry to all its identifiers before collapsing as suggested by Ramasamy (Ramasamy et al. 2008).

3.2.2.3.4 Performing a t-test and calculating effect sizes For each gene the p-value of an unpaired student t-test assuming equal variances was calculated using the Statistics::Distributions CPAN package by Michael Kospach and Matthias Trautner Kromann. As an effect size measure we calculated the fold change by dividing the mean of CR by the mean of CON values.
For the two datasets that consisted of only one replicate (GSE904 and data from Hu on forebrain (Wu et al. 2008)) we only calculated effect-sizes.

3.2.2.4 Quality control

3.2.2.4.1 Extracting quality control parameters A file containing certain characteristic values for each dataset was built to control the quality of the original data and the quality of processing (The full table is found in tab. 3.3):

"Probes before processing" is the number of probes after obtaining and annotating files for individual experiments. "Genes after processing" is the number of genes for an experiment after processing it as described in "3.2.2.3 Processing datasets, performing a t-test and calculating effect sizes". "CON-samples" and "CR-samples" are the numbers of microarray samples (replicates) for control and CR animals. "mean_CON" and "mean_CR" are the mean expression values over all probes of control and CR samples respectively and "STDEV_CON" and "STDEV_CR" the corresponding standard deviations over the probes (not to be mixed up with STDEVs over replicates). "percent overexpressed" and "percent underexpressed" give the number of genes over- and underexpressed at a p-value <0.05 according to the t-test and "effect size at 1-percentile" and "effect-size at 99-percentile" are the 1. and 99. percentile of the effect size. The experiment names include the GEO-, ArrayExpress or abbreviated GAN-accession of the the record, they were created from and the selected experimental conditions if more there was more than one in the study.
The quality was checked by searching the list for outliers. The number of probes on the arrays was between about 9000 and 45000 and was lowered to a number of genes after processing which was about half of it, presumably mainly due to different probes targeting the same gene. Replicate numbers were between 1 an 11 and about the same for control and CR samples in each dataset. The average expression value was between about 100 and 10000. Exceptions are the two datasets of the GSE11244 study which is because the expression values were normalized to internal standard RNA values. In the dataset of the GSE6110 study expression values were normalized to 1 so that the average is also lower than for the other studies here. The STDEV over the signal of all probes normally was about two to three times the average signal. Between 1 and 25% of genes were found differentially expressed at a p-value cutoff of 0.05 with the effect-size at the 1-percentile being about 0.9 to 0.5 (i.e. downregulation by 1/10 to $\frac{1}{2}$) and at the 99-percentile about 1.1 to 3.0 with the exception of GSE904 for which these values were more extreme. This is probably due to the fact that there is only one replicate and outliers have a higher impact on these values. For the other study with only one replicate by Hu this is not the case, probably because the study was done on pooled samples which reduces variation.

3.2.2.4.2 Comparison to genes found differentially expressed in the original study To further check against any flaws in our analysis we checked the genes found against those published as differentially expressed in each corresponding study. 100% overlap was however not expected since the studies often used different statistical approaches from ours.
To compare findings we downloaded lists of genes described as differentially expressed from tables in the corresponding publications or their supplementary materials.
For Dhahbi, 2005 and Edwards, 2007 no list of differentially expressed genes from the original study could be found. For some other studies only differentially expressed genes for some conditions (e.g. ages) could be obtained. In these cases we tried to extract information on genes found differentially expressed form the text of the original publication. Studies for which no information at all about genes differentially expressed with caloric restriction was given are GSE904, GSE6110, GSE18297, GSE14202, GSE17309, GDS355+GDS356 and GDS2612.
This analysis was done taking about 4 random genes published to be differentially expressed and checking them against the p-values and effect sizes found in our study. Considering that the statistical approach between the original and our study differed we only required about 2 or 3 of these genes to be statistically significant or nearly statistically significantly expressed in the same direction and accepted 1 or 2 genes not found significant in our analysis.
We investigated the case more closely if genes were found statistically significant in the other direction (up or down) than in the original study, which was the case for GDS1808 (Dhahbi, 2005) where 4 of 10 genes mentioned in the paper were found statistically significantly differentially expressed in the other direction in our study. The authors of the original study kindly provided us the original data and these were consistent with our findings calculated from the GEO data rather than the results presented in the publication. Therefore we kept our results for further steps.
All other studies were of satisfying consistency with our results.

File	probes before processing	genes after processing	CON-samples	CR-samples	mean_CON	STDEV_CON	mean_CR	STDEV_CR	over-expressed (p<0.05)	under-expressed (p<0.05)	effect size at 1-percentile	effect-size at 99-percentile
EMEXP-748_liver.csv	19445	10538	3	3	138.2	498.3	137.5	498.6	10.1	10.7	0.5	2.2
EMEXP-748_hypothalamus.csv	19445	10538	3	3	188.6	397.8	189.1	396.3	4.8	5.3	0.6	1.6
EMEXP-748_muscle.txt	19445	10538	3	3	163.6	498.6	164.3	500.5	7.9	7.2	0.5	2.0
EMEXP-748_colon.csv	19445	10538	3	3	171.5	425.1	171.4	423.7	5.8	5.2	0.6	1.6
GSE11244fixed.high.cal.txt	10264	3215	11	8	11.3	80.2	10.7	73.7	15.2	17.1	0.6	1.6
GSE11244true.ad.libitum.txt	10264	3215	4	8	10.5	65.3	10.7	73.7	12.3	13.0	0.6	1.8
GAN_EPACRR_Neocortex_30mo.txt	9257	9257	3	3	657.7	1243.0	690.7	1461.9	15.2	12.8	0.5	1.9
GSE6110_aged.txt	9942	7655	4	4	1.0	0.4	1.0	0.4	9.8	23.3	0.5	1.4
GDS1261Amesdwarf.txt	12488	10533	8	8	914.5	2297.2	919.9	2296.6	3.1	3.9	0.7	1.3
GDS1261normal.txt	12488	10533	7	8	913.5	2322.9	933.6	2376.8	7.5	8.2	0.6	1.4
GDS1808CR8.txt	12488	10533	4	4	948.1	2508.8	975.8	2644.3	4.5	4.4	0.6	1.5
GDS1808LTCR.txt	12488	10533	4	4	948.1	2508.8	1006.5	2854.2	10.5	11.7	0.6	1.7
GDS241_2h.txt	12488	10533	2	2	1431.8	3431.6	1599.6	4348.2	4.9	3.9	0.5	1.9
GDS241_4h.txt	12488	10533	2	2	1454.7	3460.1	1397.0	3177.7	3.4	3.3	0.7	1.9
GDS241_12h.txt	12488	10533	2	2	1538.7	4132.3	1499.2	3989.3	2.7	2.5	0.5	2.0
GDS2612.txt	22690	15253	5	5	852.8	2420.2	892.4	2685.3	11.4	9.8	0.6	1.8
GDS2681.txt	45101	23339	3	3	3042.1	12186.7	2969.3	9568.1	9.8	17.5	0.2	1.8
GDS2961_2962_16months.txt	16896	8265	10	10	1276.2	1813.2	1277.0	1844.1	0.9	3.2	0.7	1.2
GDS2961_2962_24months.txt	16896	8265	10	8	1278.4	1919.5	1278.8	1742.3	8.4	5.9	0.7	1.8
GDS2961_2962_6months.txt	16896	8265	10	9	1284.1	1931.2	1280.7	1881.7	5.7	5.6	0.8	1.4
GDS355_356.txt	13179	7089	5	5	5056.8	10789.8	5168.5	11212.6	5.4	5.0	0.6	1.6
GSE11291gastroecnemius.txt	45101	23339	5	5	1196.9	4280.1	1281.8	4634.1	23.1	9.7	0.6	3.0
GSE11291heart.txt	45101	23339	5	5	1320.3	4979.4	1313.9	5063.8	15.3	11.7	0.6	2.6
GSE11291neocortex.txt	45101	23339	5	5	832.3	2094.9	853.3	2244.6	12.7	15.0	0.6	1.6
GSE14202exercise.txt	22690	15253	9	9	889.6	2727.3	867.2	2515.0	5.4	6.1	0.6	1.3
GSE14202no_exercise.txt	22690	15253	10	9	889.3	2674.6	878.3	2587.6	4.7	7.7	0.7	1.3
GSE8426Cerebellum16months.txt	16896	8265	10	10	884.4	1293.5	890.5	1245.5	0.7	0.2	0.8	1.3
GSE8426Cerebellum24months.txt	16896	8265	9	8	877.0	1233.0	884.7	1228.0	2.1	0.3	0.8	1.5
GSE8426Cerebellum6months.txt	16896	8265	10	9	885.5	1197.5	883.0	1235.1	0.0	0.1	0.8	1.2
GSE8426Cortex16months.txt	16896	8265	10	10	903.6	983.2	898.3	981.6	0.3	0.4	0.9	1.1
GSE8426Cortex24months.txt	16896	8265	10	8	903.0	1051.1	905.0	1005.1	1.3	0.5	0.9	1.2
GSE8426Cortex6months.txt	16896	8265	10	9	902.2	990.1	901.7	1010.0	0.5	0.8	0.9	1.1
GSE8426Hippocampus16months.txt	16896	8265	10	10	893.5	1201.1	878.3	1093.5	20.7	9.2	0.8	1.3
GSE8426Hippocampus24months.txt	16896	8265	10	8	885.0	1169.0	879.3	1145.6	0.6	0.9	0.9	1.1
GSE8426Hippocampus6months.txt	16896	8265	10	9	883.7	1123.4	882.9	1127.0	0.7	0.4	0.9	1.1
GSE8426Spinal.cord16months.txt	16896	8265	10	10	878.9	1497.2	875.8	1487.1	2.3	2.4	0.9	1.1
GSE8426Spinal.cord24months.txt	16896	8265	10	7	874.1	1493.6	873.3	1607.2	1.7	5.3	0.8	1.2
GSE8426Spinal.cord6months.txt	16896	8265	10	8	873.7	1358.2	874.8	1393.0	3.0	3.5	0.8	1.2
GSE8426Striatum16months.txt	16896	8265	10	10	882.5	1438.5	878.9	1482.2	0.1	1.1	0.9	1.1
GSE8426Striatum24months.txt	16896	8265	10	8	916.8	1345.8	888.9	1443.6	0.1	0.2	0.5	1.2
GSE8426Striatum6months.txt	16896	8265	10	9	890.9	1472.6	893.9	1373.9	0.8	0.0	0.9	1.2
GSE18297_1month_10perc.txt	10912	8941	2	2	100.3	335.5	128.8	424.4	5.1	0.8	0.5	2.5
GSE18297_1month_20perc.txt	10912	8941	2	2	100.3	335.5	130.4	425.1	4.8	0.7	0.6	2.5
GSE18297_1month_30perc.txt	10912	8941	2	2	100.3	335.5	129.1	422.1	0.8	0.5	0.5	2.5
GSE18297_1month_5perc.txt	10912	8941	2	2	100.3	335.5	146.5	472.7	1.3	0.8	0.5	2.7
GSE18297_1week_10perc.txt	10912	8941	2	2	104.6	358.2	125.6	414.5	3.9	1.0	0.6	2.5
GSE18297_1week_20perc.txt	10912	8941	2	2	104.6	358.2	133.2	441.1	10.9	1.4	0.6	2.6
GSE18297_1week_30perc.txt	10912	8941	2	2	104.6	358.2	99.7	336.6	1.7	2.5	0.5	2.4
GSE18297_1week_5perc.txt	10912	8941	2	2	104.6	358.2	88.4	299.3	1.5	2.7	0.5	2.2
GSE6718adipose_tissue.txt	17034	11834	5	7	283.7	1054.2	306.7	1080.4	20.2	11.5	0.7	1.8
GSE6718heart.txt	17034	11834	7	7	334.3	1404.0	342.2	1419.0	5.2	2.5	0.9	1.2
vlookup_table_GSE9121_liv4.txt	17034	11834	9	4	631.0	2681.8	581.3	2607.0	9.9	29.2	0.5	1.3
vlookup_table_GSE9121_adip.txt	17034	11834	4	4	733.0	2641.8	683.3	2538.9	1.7	5.8	0.7	1.2
vlookup_table_GSE9121_liv10.txt	17034	11834	9	5	631.0	2681.8	681.0	2786.8	25.6	11.9	0.7	1.6
GSE7502testis6months.txt	25577	10217	2	2	9390.7	29877.7	9243.4	28859.3	3.2	1.7	0.8	1.2
GSE7502ovary24months.txt	21264	12163	2	2	5145.5	16535.8	5107.2	17335.3	2.6	3.3	0.7	1.3
GSE7502testis16months.txt	25577	10217	2	2	9094.2	29409.3	9115.9	29452.9	3.3	4.0	0.8	1.3
GSE7502testis24months.txt	25577	10217	2	2	9126.6	28631.7	9017.7	29227.0	4.0	5.6	0.7	1.3
GSE7502ovary6months.txt	21264	12163	2	2	5671.3	17624.7	5523.2	19762.5	3.1	5.7	0.6	1.3
GSE7502ovary16months.txt	21264	12163	2	2	5225.2	17119.5	5289.9	17062.5	4.3	4.8	0.7	1.3
GSE17309.txt	357	265	4	4	566.6	1906.9	538.5	1970.5	1.5	3.8	0.7	1.2
oneReplicate_tabGSE904.txt	16896	11345	1.0	1.0	399.7	947.1	408.8	735.7	--	--	0.4	6.1
oneReplicate_CR PS RAW DATA.xls (Hu)	8882	8237	1.0	1.0	4183.3	13641.8	4088.7	12427.3	--	--	0.5	2.4

Table 3.3: Table listing characteristics for each dataset for quality control. See text for details.

3.2.2.5 Excluding genes differentially expressed with age

Since CR is a mechanism known to counteract the effects of aging it is expected that some of the gene expression changes by CR in older organisms are due to the reversal of changes normally occurring with age, e.g. while the expression of a certain gene goes down with age in ad libitum fed animals, it does not in CR fed animals. This gene would then be found differentially expressed between old CR and AL-animals (Fig. 3.7). We aim at distinguishing those genes from others found differentially expressed with CR which are supposed to provide a mechanistic explanation for the effects of CR rather than are a consequence of it.

We therefore decided to remove all genes found differentially expressed between older and younger AL animals from the genes differentially expressed with CR in the opposite direction in the older animals. This was done using exclude_on_criteria_v2.1.pl (supplement 2). The younger animals (mice or rats) chosen for the comparison were normally about 4 months old and from the same study. If the study did only contain old animals no genes differentially expressed with age were excluded, since it is nearly impossible to find other studies on changes with aging under the same conditions (same strain, age, etc). We chose about 4 months old animals as a control even if younger animals were available, so that the results were not disturbed by changes between non-fully grown and adult animals.
The t-test for the old vs. young comparison was done the same way and applying the same cutoffs as the one for the CR vs. AL comparison, following the logic that if the gene is significantly differentially expressed in CR only because of ameliorating changes in age dependent gene expression, then these age dependent changes should be significant at the same threshold, given that the sample sizes are similar. This was the case for all datasets. The percentages of genes differentially expressed in opposite direction with age and CR (at $p < 0.05$ and effect size $>$ 1.5-fold) and therefore excluded is given in table 3.4.

For more information on the number of genes differentially expressed with CR, age and with CR and age in opposite direction see "3.2.5 Relationship between differential expression with CR and age".

3.2.3 Processing gene lists from studies for which expression data were not obtained

For studies for which we could not obtain expression data, but only lists of genes differentially expressed according to the statistical test in the original study, we downloaded these lists plus any statistical parameter (e.g. effect-size) if available. Since these lists were extracted from publications or especially corresponding supplementary material these studies are also called "supplemental studies" and corresponding genes "supplemental genes" in the following.
Using expression data is preferable to these data since they are obtained by different statistical methods and criteria and data on non-differentially expressed genes are missing.
Annotation to mouse Entrez IDs was done as for the raw data. For the only non-mouse dataset on rhesus monkey transcripts, assayed on a human microarray platform (Kayo, 2001), human Entrez IDs were added first using annotation from BioMart and those mapped to mouse Entrez IDs using mapping files from HomoloGene. Mouse Entrez IDs were added to all other files by looking up mouse Entrez IDs corresponding to the GBACCs (and if given Gene Symbols) in BioMart mapping files using use_vlookup_mod4.pl with vlookup_mod4_3.pm (See "3.2.2.2 Annotating data with identifiers common to all data files" and sub-chapters for details).
For lists of genes differentially expressed in hypothalamus provided by Hu ((Wu, P. et al. 2009); note: these are different data than the raw data provided by Hu on forebrain (Wu et al. 2008) mentioned above) the only given identifiers were Gene Names and Affymetrix probe IDs. We were not able to map any of them to Entrez IDs. Therefore datasets from this study had to be excluded from further analysis.
Since p-values -if given- were determined by different statistical tests, some of them multiple testing corrected, others not, we replaced -or added- them as $p = 0.001$, i.e. a significant p-value and therefore only evaluated the genes by their effect-sizes and the fact that they were stated as over- or underexpressed. (If effect sizes for underexpresed genes were given as negative values, e.g. -2 we converted them to the corresponding values between 0 and 1, e.g. $-\frac{1}{(-2)} = 0.5$. We accepted that supplement genes may have been chosen with stricter or less strict statistical criteria than in our analysis. For attempts to assimilate our statistical criteria to those used for supplemental data see "3.2.4.2 Combining expression data prepared from raw data and supplemental lists of differentially expressed genes".

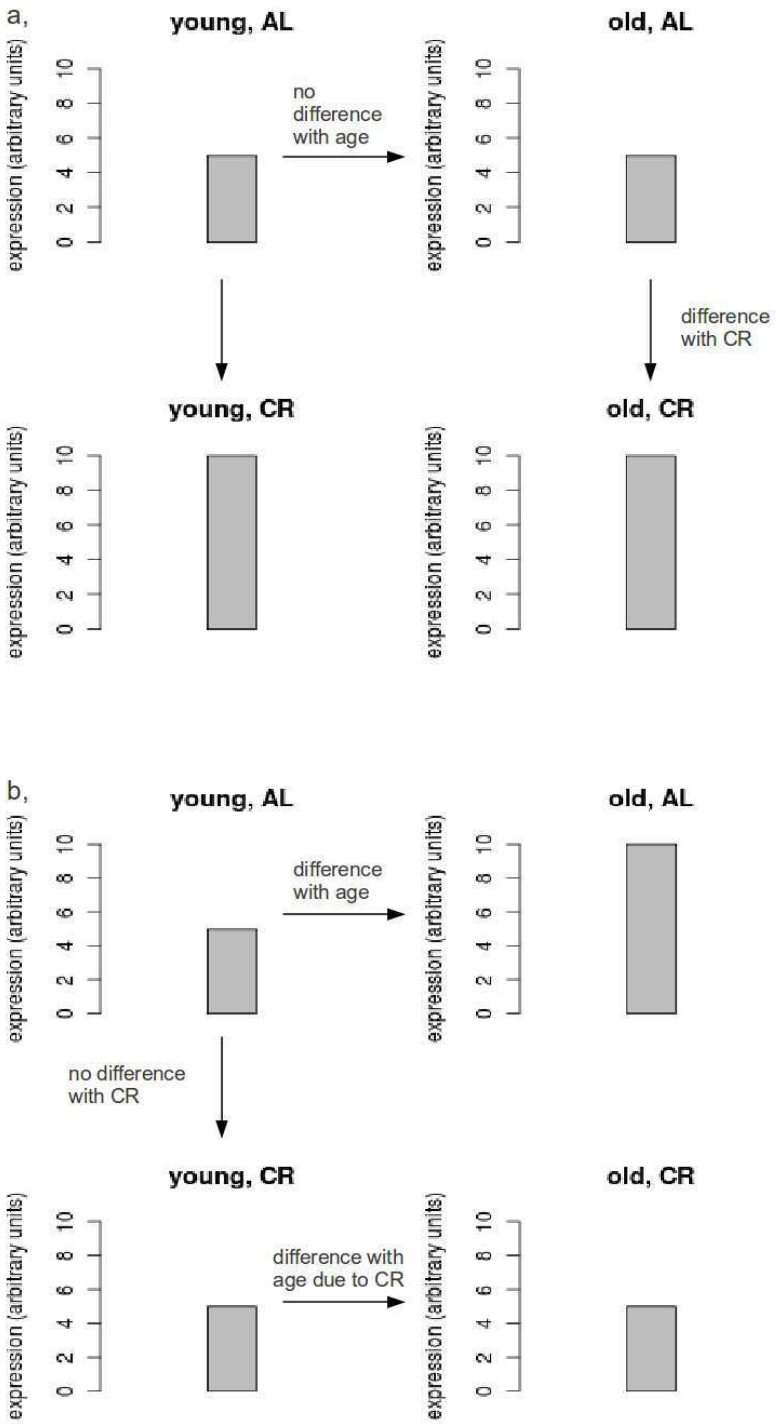


Figure 3.7: (Dummy figure) Example demonstrating the reasoning, why genes differentially expressed with age were excluded. a, No differential expression with age, but with CR; this gene is expected to contribute as mechanistic reason to the effect of CR; b, Difference between old AL and old CR, because the gene is differentially expressed with age under AL, but not CR conditions; this gene is expected to be differentially expressed only as an effect of CR

CR data	total genes	% excluded
GSE8426Cortex24months.txt	8265	0
GSE8426Spinal.cord16months.txt	8265	0
GSE8426Hippocampus24months.txt	8265	0
GSE8426Striatum24months.txt	8265	0
GSE8426Cerebellum24months.txt	8265	0
GSE8426Striatum16months.txt	8265	0
GSE8426Spinal.cord24months.txt	8265	0
GSE8426Cortex16months.txt	8265	0
GSE8426Hippocampus16months.txt	8265	0.01
GSE8426Cerebellum16months.txt	8265	0.04
GSE7502testis16months.txt	10217	0.04
GSE7502ovary16months.txt	12163	0.04
GSE6718heart.txt	11834	0.08
GSE7502ovary24months.txt	12163	0.18
GSE7502testis24months.txt	10217	0.26
GSE11291neocortex.txt	23339	0.32
GAN_Expr_Profile_Aging_CR_Retardation_Neocortex_30months.txt	9257	0.86
GDS2681.txt	23339	1.00
GDS355_356.txt	7089	1.18
GDS2961_2962_24months.txt	8265	1.56
GSE11291gastrocnemius.txt	23339	1.73
GDS2612.txt	15253	1.86
GSE11291heart.txt	23339	1.90
GSE6110.txt	7653	2.21
GSE6718wat.txt	11834	4.69

Table 3.4: Number of genes excluded because of differential expression with age in opposite direction. For all datasets for which data on younger AL-animals existed genes were excluded that were differentially expressed with age in AL-animals in opposite direction of differential expression with CR in the older animals. The total number of genes in the dataset and percentage of genes excluded are given.

3.2.4 Estimating the significance of the number of studies in which genes were differentially expressed

To determine if a gene was found differentially expressed in more studies than expected by chance we first counted for each gene in how many studies its expression was measured and in how many it was found over- or underexpressed at a p-value of p <0.05 and a fold-change of at least 1.5 (see "3.2.4.1 Determining t-test p-value and effect-size cutoff" on how these cutoffs were chosen). We obtained the probability of finding a gene over- / underexpressed at least this often by random chance (binomial p-value) from the cumulative binomial distribution:

$$P = 1 - \sum_{x=0}^{k-1} \binom{n}{x} * p_s^x * (1 - p_s)^{(n-x)} \tag{3.1}$$

For this we used the success probability (p_s) calculated by dividing the number of genes appearing over- / underexpressed in all studies by the total number of appearances of genes in all studies (i.e. a gene differentially expressed / tested more than once is counted for each time it was differentially expressed / tested).[3]
To find an appropriate cutoff for the binomial p-value we repeated the binomial test 100 times on scrambled data. By dividing the mean of the number of genes found with scrambling below a certain binomial p-value by the number of genes found below it on the real data we obtained a FDR estimate. We calculated the FDR for some different binomial p-values and decided on a cutoff of 0.0005 which corresponds to a FDR of 0.041 for over- and 0.062 for underexpressed genes. These calculations were done using CR_binomial_UN_scrambled_v3.1.pl (supplement 2).
Two important decisions had to be made for the binomial test:

1. How to choose the t-test p-value and effect size cutoff.
2. How to combine the genes from supplemental data with those for which the t-test was performed.

3.2.4.1 Determining t-test p-value and effect-size cutoff

In order to determine which genes to consider over- and underexpressed we needed cutoff values for the t-test p-value and / or the effect-size. Note that the criterion for the final results of our analysis is not the t-test combined with effect-sizes, but are the p-values of a binomial test performed on the number of studies in which a gene is found under- / overexpressed by the t-test and / or effect-sizes in relation to the total number of studies in which the gene was tested. Therefore there was no need to select the t-test p-value cutoff in the common way, e.g. as p <0.05 after multiple testing correction. Instead the binomial test is expected to buffer the choice of the t-test p-value and effect-size cutoffs, i.e. if the thresholds are set relaxed, the success probability (p_s in formula 3.1) in the binomial test will be higher, so the number of times a gene is found differentially expressed (k) in relation to tested (n) has to be high to be significant for the tested gene. If on the other hand strict cutoffs are selected p_s will be low, so that the k may be smaller in relation to n and still be significant in the binomial test. Nonetheless, as considering extreme cases shows, the choice of these cutoffs is not completely deliberate.
If choosing extremely relaxed cutoffs p_s might get so low that $\binom{n}{k}$ with low n (e.g. $\binom{4}{3}$) may not be meaningful and not significant, while e.g. $\binom{7}{6}$ will be significant, therefore discriminating against genes tested less often and increasing false negative rates. It would be preferable to find cutoffs so that $\binom{n}{k}$ with low (but not too low) n are meaningful. If however extremely strict cutoffs are chosen finding a gene differentially expressed in only one or two studies might suffice for considering it significant. This would contradict the aim of the meta-analysis. It would allow false positives in the original studies to also become false positives in the meta-analysis which is to be avoided. This may be one of the reasons why rather relaxed cutoffs were chosen in Magalhaes, 2009.
The choice of t-test p-value and effect-size thresholds is therefore a way to determine if the significant results of the meta-analysis should rather be such that were found very reliable in only a few studies or such that were found under more relaxed conditions in a higher number of studies. For our aims the emphasis is on the second point which suggests the use of relatively relaxed cutoffs. However as mentioned above the analysis should still be sensitive enough to detect genes only tested in a relatively low number of studies.
A means to control for false positives is the FDR, calculated by dividing the number of genes below a certain

[3] A more accurate mathematical procedure would include using the hypergeometric instead of binomial distribution. However as n is small compared to the total number of genes the use of the binomial distribution is justified.

binomial p-value cutoff found on scrambled data by the number found on the real data. To assay different cutoff criteria we examined the FDRs for three given binomial p-value cutoffs (0.0001, 0.0005 and 0.001) for different t-test p-value and effect-size pairs (0.05,2; 0.1,2; 0.05,1.5; 0.1,1.5; 0.1,1), to see which settings in general lead to higher or lower FDRs and to maximise the number of genes found differentially expressed at a given FDR (i.e. Type II error for a given Type I error rate). As depicted in fig.. there is no clear trend over the different binomial p-values that either the strict or relaxed ones of our cutoffs are preferential as to their FDR or number of genes found. This supports the argumentation that the binomial test is rather robust to the chosen t-test p-value and effect-size cutoff.

Importantly however we had to consider that this study includes lists of genes obtained from publications or supplements, for which the analysis was not done by ourselves and therefore statistical tests with different cutoffs were applied. If we wish that all studies contribute with a similar weight to the meta-analysis, we had to make sure cutoffs were chosen in our study that resemble those as closely as possible. We expected that to achieve this, the percentage of genes found over- or underexpressed should be similar. For the studies obtained from literature we calculated these numbers from the numbers of genes given as differentially expressed and the total number of genes on each particular array. We used the number of genes given as differentially expressed before annotation since there was no way of estimating the total number of genes in the datasets after annotation (i.e. how many of the total genes would have been lost during annotation). We assumed a similar probability of loss of a gene during annotation for the complete dataset as for the differentially expressed genes.
For our own study the percentage of over- and underexpressed genes in each dataset was calculated from the number of genes found differentially expressed and the total number of genes after annotation.

We used datasets with more than one replicate (all but two datasets), because a t-test is not possible on datasets with only one replicate.

Results are shown in fig. 3.8. It was found that the percentages of genes from literature found over- or underexpressed resemble percentages we obtained with rather strict cutoff settings.
Even though the numbers of genes differentially expressed at a t-test p-value cutoff of 0.05 and an effect-size cutoff of 2 would better fit the results of the supplemental data, we chose 0.05 for the p-value and 1.5 for the effect size. This is because of the argument above that with very low success probabilities a gene can be found significant in the binomial test even when only over-/underexpressed in very few studies. The aim of the meta-analysis is however to find genes consistently differentially expressed over several studies (and conditions). For the 0.05, 2 selection the average percentage of differentially expressed genes (= success probability * 100) is around 0.85% before and 1.0% after including supplemental data, for the 0.05, 1.5 selection about 2.0% before and 2.4% after adding supplemental data. (Success probabilities rise when including supplemental data since these only consist of differentially expressed genes.)
Two datasets (data from Hu on Hypothalamus and GSE904) were based on unreplicated samples. Therefore no t-test could be performed and i.e. no p-value cutoff used for selecting over- or underexpressed genes. Since genes would be selected in a less strict way if only using the same effect-size cutoff as for the other data, we decided that a stricter effect-size threshold should be chosen. The way to find an appropriate threshold was to choose it in a way that a similar number of genes would be found over-/underexpressed as in the other studies. However the percentage of genes found differentially expressed for different thresholds in the two datasets was always much higher for GSE904, especially for overexpressed genes. The fact that data from Hu were from pools of 3 hypothalami, while the data in GSE904 were not pooled suggests higher reliability of the first and the use of different cutoffs for the two datasets. We decided on an effect-size cutoff of 1.7 for the data from Hu, at which 2.1% of genes are overexpressed and 1.4 underexpressed and of 4.0 for GSE904, at which 3.9% are over- and 0.6 are underexpressed.

3.2.4.2 Combining expression data prepared from raw data and supplemental lists of differentially expressed genes

The issue of how to combine lists of genes on the one hand created by calculating effect sizes and t-test p-values from expression data and on the other hand obtained directly as a list of differentially expressed genes from publications and supplements ("supplemental genes") is not trivial.
It is essential for the binomial analysis not only to have data of differentially expressed genes, but also on non-differentially expressed ones, so that both the "number of successes" (k) and of "trials" (n) (in 3.1) for each gene can be given. The data of non-differentially expressed genes were however not available from published lists. There are several possibilities to combine the data, all with their own drawbacks:

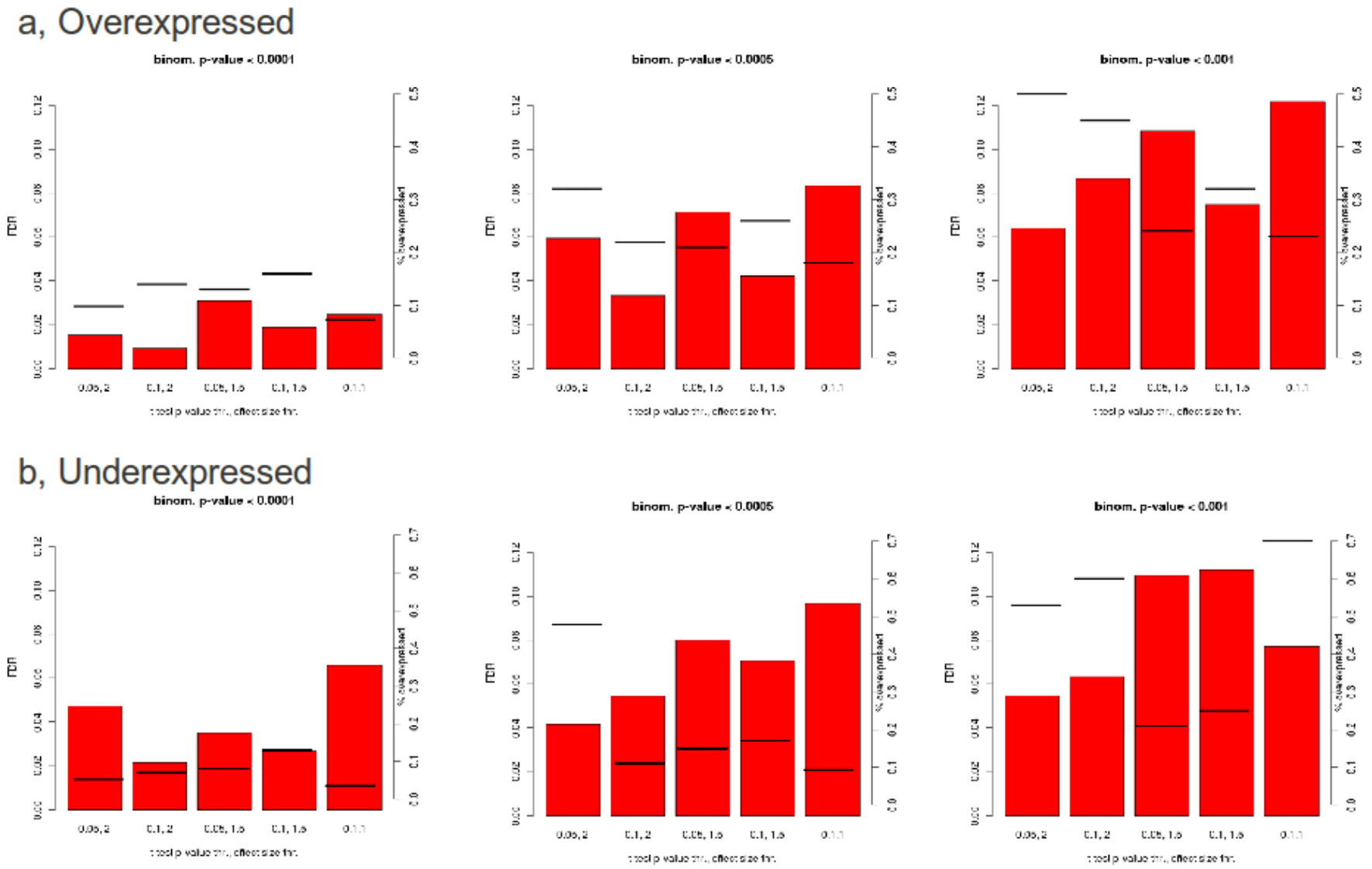


Figure 3.8: FDRs (red columns) and percentages of genes found (a,) over- or (b,) underexpressed in significantly more studies than expected by chance. Different columns show, when genes are selected at different t-test p-value and effect size cutoffs (0.05,2; 0.1,2; 0.05,1.5; 0.1,1.5; 0.1,1). Different panels show different binomial test p-value cutoffs. Created from studies with at least 2 replicates

1. The differentially expressed supplemental genes are added to the genes from raw data, ignoring that other unknown genes were tested in the supplemental studies. This is therefore an analysis as if no other genes than those given in the lists were studied. The success probability p_s in the binomial test is the number of over- / underexpressed genes in these combined data divided by the size of the data. The probability p_s is therefore greater than the probability of finding a gene differentially expressed in a microarray study. The binomial probability cannot be interpreted as the probability of finding the gene at least this often differentially expressed by chance, when tested in the given number of studies. As a consequence the binomial p-value rises for genes not in the supplemental data, when including supplemental genes. However the FDR is estimated by scrambling the same data, so that the binomial p-value cutoff will be higher at the same FDR compared to when supplemental data are not included.

2. Since the total number of genes tested in supplemental studies can be found in the literature, the analysis in 1. can be modified: Instead of p_s as described above, a p_s can be used which is the number of genes found over- / underexpressed divided by the number of genes tested in all studies. Therefore p_s is smaller than the frequency of differentially expressed genes in the combined list, which is used for scrambling. This interpretation of p_s accounts for the fact that more genes were tested in the supplemental data than those given as differentially expressed. However since we do not know which genes were tested in the supplemental studies and found non-differentially expressed, their binomial p-value would be calculated too low using the lower p_s (i.e. this approach is inconsistent in that p_s accounts for the unknown non-differentially expressed genes, but n cannot). However, as in 1., choosing the p-value cutoff from a FDR calculated by scrambling will ameliorate the problem of generally underestimated binomial p-values.
   So far scrambling was always done on a list, that is enriched in differentially expressed genes, because it does not contain the unknown non-differentially expressed genes from the supplemental data. Considering the existence of these genes might lead to generally lower FDRs. Two possibilities are:

3. Lists from supplemental data are filled up with Entrez ID substitutes with non significant effect sizes. Because they are not significant, they will have binomial p-values = 1 in the analysis and act as non-significant genes in the scrambling process. However this is an approach assuming that all unknown genes in the supplemental data are different from the genes in the other studies. In reality probably most are the same as in the other studies. Therefore this approach increases the number of non-differentially expressed genes, but it does not account for the fact that these might be the same as other genes in the analysis. Therefore this approach does not fit for this situation.

4. In order to overcome the problem of 3. the incomplete lists of supplemental data has to be filled with random Entrez IDs, already existing in the lists of obtained data, with non-significant effect-sizes. However this would introduce randomness already at the level of unscrambled data and is therefore to be avoided.

The problems in 3. and 4. show that it is not feasible to include the non-differentially expressed genes of the supplemental studies as long as they are not known.
We decided that the accuracy and interpretation of the binomial p-values is of minor importance for our study, as long as the FDR can be correctly estimated (in contrast to the drawbacks of 3. and 4.) and used to decide on an appropriate p-value cutoff. Since 2. assumes that more than the given number of tests were done (decreasing p_s), but cannot increase the number of times certain genes were tested (increasing n) we choose 1. as more consistent within itself and defined p_s as the probability of a gene being differentially expressed within the combined data.

3.2.5 Relationship between differential expression with CR and age

The importance of CR comes from its ability to extend life-span in several organisms. Therefore we examined the relationship between differential expression with CR and age. In particular we tested if more genes are differentially expressed in opposite direction with age and CR than expected by chance in each dataset. We argued that a gene will be differentially expressed between AL and CR in old animals if differential expression with age in AL animals is ameliorated by CR. Differential expression will be in opposite directions in this case. See "3.2.2.5 Excluding genes differentially expressed with age" for a detailed explanation. Note that this approach does not draw conclusions from negative results, as it is the case when looking for genes that are differentially expressed with age under AL, but not under CR conditions.
For each gene in an annotated dataset we did a t-test for caloric restriction vs. ad libitum fed for old animals

and another t-test for young vs. old AL-animals. For $p < 0.05$ (without multiple testing correction) and a fold-change of at lest 1.5 we extracted genes found differentially expressed in both tests and for which the direction of differential expression was opposite, so that therefore the expression in an old animal under CR resembles that of a young animal. (See "3.2.4.1 Determining t-test p-value and effect-size cutoff" for an explanation why binomial test procedures are robust for the choice of cutoff values.) The genes obtained here are the same as those excluded from the meta-analysis of CR in "3.2.2.5 Excluding genes differentially expressed with age". See this section for details.
We found that for our settings depending on the dataset between 0 and 67% of genes differentially expressed with CR were differentially expressed in opposite direction with age. (However only up to 4.69% of all studied genes were differentially expressed with CR and age in opposite directions. See "4.2.2.5. Excluding genes differentially expressed with age" and table 3.4 therein).
We calculated the probability of obtaining an overlap at least this great by random chance by using the cumulative binomial distribution, taking the number of genes over- / underexpressed with CR and differentially expressed in opposite direction with age as successful trials, the number of genes over- / underexpressed with CR alone as trials and the probability of a gene being under- / overexpressed with age as the probability of success. For these calculations we used exclude_on_criteria_v2.1.pl (supplement).
P-values obtained were <0.005 in most cases (table. 3.5). The number of studies is not high enough to draw conclusions e.g. in which tissues expression changes are most ameliorated with age, etc. since each tissue was only tested a few times and there are other variables that vary between the studies. Nonetheless the data suggest that there is indeed a CR-effect on the level of gene expression for all tissues except some brain tissues. However it should be noted that the number of genes changing expression with age in these brain tissues is generally low so that there is little need for CR action. Interestingly the CR-effect is also less marked in the two ovary datasets. A possible interpretation might be that many genes changing expression in these datasets may do so in a tissue-specific programmed way which is not counteracted by CR.

Note however that this short section on the relationship between differential expression with CR and age is only meant to give a rough idea what can be achieved from such a study and must still be done in a more in depth way.

3.2.6 Functional analyses

3.2.6.1 Determining functional categories enriched in the meta-analysis datasets - GO-analysis

In contrast to determining functional categories in which determined candidate genes were enriched (as in "3.2.6.2 Putting genes found differentially expressed with CR into functional categories - DAVID-analysis") we here asked if the functional (gene ontology (GO)) categories themselves, as the basic units of the binomial test, were found overrepresented (for over- or underexpression) more often than expected by random chance. We therefore compared the number of times a GO-category is found associated with an over- / underexpressed gene in the datasets to the number of times it is found associated with any gene:
A table matching GO-IDs to genes was prepared the following way: A file mapping each Entrez ID to corresponding GO-IDs with one GO-ID per line was downloaded from the NCBI FTP[4] (19/07/10). We created a file mapping each Entrez ID to a comma-separated list of all corresponding GO-IDs using GOparser_modified.pl and CR_GO_UN_scrambled_v1.2.pl (supplement 2). Independently from this we created a list of only those genes appearing in the datafiles using CR_binomial_UN_scrambled_v3.1.pl (supplement 2) and added the GOs to this list with vlookup_mod4.3.pm (supplement 2).
We counted a GO-ID each time it appeared associated with any over- / underexpressed genes in any dataset (counting it twice if the same gene associated with this GO was found in different studies). Then we counted the number of times it appeared associated with any gene studied. We performed a binomial test on those numbers (see 3.1), calculating probabilities (p-values) that a gene would be found over- / underexpressed at least this often (k) by random chance. The number of trials (n) was the total number the GO appeared associated with any gene in the datasets and the success probability (p_s) the ratio of GO-IDs associated with over- / underexpressed genes to GO-IDs associated with any gene. The cutoffs for over- / underexpression were the same as in "3.2.4.1 Determining t-test p-value and effect-size cutoff". This process was done using CR_GO_UN_scrambled_v1.2.pl.
FDRs as a criterion for deciding on cutoffs for the binomial p-value were calculated by dividing the mean number

[4] ftp://ftp.ncbi.nih.gov/gene/DATA/gene2go.gz

CR data	total genes	CR up	aging down	CR up, aging down	CR up, aging down / aging down	p-value		CR down	aging up	CR down, aging up	CR down, aging up/ aging up	p-value
GSE11291heart.txt	23339	2015	673	239	0.36	<0.001		974	660	204	0.31	<0.001
GSE6718WAT.txt	11834	876	616	374	0.61	<0.001		316	545	181	0.33	<0.001
GDS2612.txt (skeletal muscle)	15253	693	427	183	0.43	<0.001		448	230	101	0.44	<0.001
GSE11291neo- cortex.txt	23339	431	267	31	0.12	<0.001		1175	293	43	0.15	<0.001
GSE7502testis24mo.txt	10217	45	37	10	0.27	<0.001		64	85	17	0.20	<0.001
GDS2961_2962_24mo.txt (thymus)	8265	304	115	63	0.55	<0.001		97	278	66	0.24	<0.001
GSE11291gastro- cnemius.txt	23339	4150	685	351	0.51	<0.001		1119	368	52	0.14	<0.001
GDS2681.txt (cochlea)	23339	1445	111	50	0.45	<0.001		3651	275	184	0.67	<0.001
GSE6110.txt (kidney)	7653	176	900	59	0.07	<0.001		354	566	110	0.19	<0.001
GDS355_356.txt (kidney)	7089	196	168	28	0.17	<0.001		193	200	56	0.28	<0.001
GAN_Expr_Profile_Aging_ CR_Retardation_ Neocortex_ 30mo.txt	9257	683	169	31	0.18	<0.001		637	229	49	0.21	<0.001
GSE7502testis16mo.txt	10217	31	6	2	0.33	<0.001		10	57	2	0.04	0.001
GSE8426Hippocampus16mo.txt	8265	23	1	1	1.00	0.003		2	0	0	/	/
GSE6718heart.txt	11834	7	203	1	0.00	0.11		12	574	9	0.02	<0.001
GSE7502ovary16mo.txt	12163	32	458	3	0.01	0.12		57	302	2	0.01	0.42
GSE7502ovary24mo.txt	12163	26	651	2	0.00	0.41		74	302	20	0.07	0.00
GSE8426Spinal.cord16mo.txt	8265	2	6	0	0.00	1		2	3	0	0.00	1
GSE8426Cerebellum24mo.txt	8265	8	7	0	0.00	1		3	270	0	0.00	1
GSE8426Spinal.cord24mo.txt	8265	1	39	0	0.00	1		15	55	0	0.00	1
GSE8426Cerebellum16mo.txt	8265	3	8	0	0.00	1		7	35	3	0.09	<0.001
GSE8426Cortex24mo.txt	8265	3	0	0	/	/		0	0	0	/	/
GSE8426Hippocampus24mo.txt	8265	0	1	0	0.00	/		1	0	0	/	/
GSE8426Striatum24mo.txt	8265	0	0	0	/	/		0	2	0	0.00	/
GSE8426Striatum16mo.txt	8265	0	1	0	0.00	/		2	0	0	/	/
GSE8426Cortex16mo.txt	8265	0	1	0	0.00	/		1	1	0	0.00	1.00

Table 3.5: Number of genes changing expression with CR, aging and 'aging and CR in opposite direction'. Ratios of genes changing in opposite direction of all genes changing with age and probabilities (p-value), that at least this many would be found by chance are also shown. "/" indicates cases in which a binomial test is not reasonable since either the number of trials or the success probability is 0.

of GOs found over- / underexpressed by scrambling 100 times by the number found for the unscrambled data. We selected a binomial p-value threshold of 0.001 which corresponded to a FDR of 0.023 for GOs for over- and 0.029 for GOs for underexpressed genes.
While the enrichment analysis on candidate genes (see next section) tries to classify the genes found in the meta-analysis and therefore to find a possible explanation, why they might have been found, this one might find categories important for the mechanism of CR which might exhibit their action through different members of this category in different circumstances. E.g. while gene A might be overexpressed with CR in liver, gene B of the same category might be overexpressed with CR in kidney.

3.2.6.2 Putting genes found differentially expressed with CR into functional categories - DAVID-analysis

Since the relative large lists of genes differentially expressed with CR are hard to interpret, we used the Functional Annotation Tool of the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al. 2003) to put them into functional categories (21/07/10)(see also "2.1.5.3 Introduction to DAVID").
We separately uploaded the lists of genes enriched for overexpression and underexpression (binomial p-value $<$ 0.0005) and a list of all genes used in the studies in the form of mouse Entrez IDs and ran the analysis under default settings.
We obtained the "Functional Annotation Chart", a list of functional categories enriched in the input genes, and "Functional Annotation Clustering", clusters of those categories according to the genes they have in common. We acquired them by running the program first on the databases (e.g. for GO-terms, pathways, diseases, tissues etc.) selected by default and then specifically for KEGG and BIOCARTA pathways.

3.2.7 Determining tissues contributing to enrichment of genes for over- or underexpression

As already mentioned there are several covariates, varying between the different datasets in our meta-analysis, as for example organism and strain, age, CR regime and duration of CR. Our meta-analysis provides an opportunity to explore how genes overrepresented for over- or underexpression are associated with those variables. Of particular interest is the covariate tissue. This is on the one hand because the meta-analysis aimed at finding genes differentially expressed with CR under multiple conditions and, due to the high number of datasets from liver, it is a concern that genes may be found significant, even though only differentially expressed in liver. On the other hand it may also be interesting if genes only found differentially expressed in one tissue in this meta-analysis indeed exert a tissue-specific CR effect. In fact the liver would be a good candidate for harbouring tissue specific effects of CR due to its important role in metabolism.
We pursued the following approach to shed light on the tissue expression of the genes found significant in the meta-analysis:
We used create_table.pl (supplement 2) to create a table with the genes in the rows and the datasets in the columns and each field displaying the t-test p-value and effect-size of the gene in this dataset. Using mark_fields.pl (supplement 2) on the part of the table that contained significant results of the meta-analysis, we indicated fields with t-test p-values and effect-size values that corresponded to over- or underexpression according to the relaxed thresholds used in the meta-analyses ($p < 0.05$, effect-size: 1.5 fold-change). The identified fields were manually color coded red for over- and green for underexpression and the column-header was color coded according to the tissue the corresponding dataset was obtained from. We then identified genes that were over- (for genes enriched for overexpression) or underexpressed (for genes enriched for underexpression) in only one, two or more than two different tissues (Fig. 3.10).

3.2.8 DAVID-analysis on presumably tissue-independent and liver-specific candidates

Because we found that a large number of genes in our final result were differentially expressed only in liver datasets or in datasets from only liver and one other tissue, we repeated the DAVID-analysis on genes differentially expressed at least in 3 different tissues to find the functional categories behind genes important for the mechanism of CR in a truly tissue independent manner. We also ran DAVID on these candidates only differentially expressed in liver to find truly liver-specific mechanisms of CR.

3.2.9 Co-expression analysis of CR-associated genes

Besides determining functional categories of genes associated with a certain trait it is often useful to determine genes significantly more strongly co-expressed with the genes of interest than with other genes. These detected genes may therefore be important upstream regulators or downstream targets of the studied process.
The co-expression analysis of the genes associated with CR was done with software developed by S. van Damm of our group (unpublished). In brief, from a large number of microarray datasets on mouse in GEO for each gene, similarity scores to the expression of all other genes were calculated and genes ranked by these scores. The top 5% of genes with highest similarity for each gene were considered co-expressed with this gene.
Each mouse gene g_i was then tested for overrepresentation in the number of times it was found co-expressed (i.e. in the top 5%-list) with a certain-subset of genes, compared to the number of times it was co-expressed with all mouse genes. In our case this subset was once genes enriched for overexpression with CR and once for underexpression. More precisely a binomial test 3.1 was done with the number of tests (n) being the number of genes in the subset and the number of hits (k) being the number of times g_i is co-expressed with genes of this subset. The success probability (p_s) of g_i being co-expressed with any gene was p_s = number of times g_i is co-expressed with any gene / number of all genes.
The genes were ranked by their p-values of the binomial test and a FDR estimated (as in (Rhodes et al. 2002)) as the number of genes found divided by the number of genes expected at each p-value, which is the ratio of genes found with smaller or equal p-value divided by the p-value itself.
Since a large number (1576 and 1069; given in supplement 2) of genes were found co-expressed with genes enriched for over- and underexpression we performed DAVID-analysis under default settings on them.

3.2.10 Transcription factors regulating expression of candidate genes

To detect enriched transcription factor (TF) binding sites in our candidate genes we used WebMOTIFS[5] (Romer et al. 2007). This program acts as an interface to the motif discovery programs MEME (Bailey & Elkan 1994), AlignACE (Hughes et al. 2000), MDscan (Liu et al. 2002), Weeder (Pavesi et al. 2004) and THEME (Macisaac et al. 2006). The downside of using this program was that input genes had to be given as RefSeq-IDs. The conversion process lead to loss of about 20 genes each for over- and underexpressed candidates. However we expect that the lost genes represented rather poorly annotated ones, so that not much information was expected form them anyway. Sequence motifs were searched between 1000 bp downstream to 200 bp upstream with an expected motif length of <12 bp, strict significance filtering and trying all initial hypotheses for the search in THEME.

3.2.11 Detecting overlap with CR-essential genes, their orthologues and interaction partners

Genes experimentally identified to be essential for the effect of CR to induce life-span extension in different model organisms were recently extracted from literature and summarized in the database GeneDR by D. Wuttke of our group (unpublished). Essential here means that manipulation of the transcription levels of the genes (e.g. knock-out by deletion, knock-down via RNAi or transposition, or overexpression) significantly modified the effect of CR on life-span extension.
The only mouse gene known to be essential for CR-induced life-span extension in this database was *Ghr* (Growth hormone receptor; Entrez ID: 14600) and this gene was found enriched for downregulation in our meta-analyses. The following further comparisons between the results of the meta-analysis and genes in GeneDR, undertaken by D. Wuttke, are only described in brief:

1. The results of the meta-analysis were also compared to murine orthologues of genes essential for CR in *S. cerevisiae* and *C.elegans*.

2. A network of murine CR-essential gene orthologues and *Ghr* was built according to information on physical protein-protein and genetic interactions retrieved and integrated from IntAct (Hermjakob et al. 2004), DIP (Xenarios et al. 2000), MINT (Zanzoni et al. 2002), BIND (Bader et al. 2001), BioGRID (Stark et al. 2006), MPACT (Güldener et al. 2006), DroID (Jingkai Yu et al. 2008), Reactome (Stein 2004), HPRD (Prasad et al. 2009), PDZBase (Beuming et al. 2005), CORUM (Ruepp et al. 2008), iRefIndex (Razick

[5] http://fraenkel.mit.edu/webmotifs

et al. 2008), PhosphoSitePlus (Hornbeck et al. 2004), PhosphoGRID (Stark et al. 2010), I2D (Brown & Jurisica 2007), InteroPorc (Michaut et al. 2008), InterologFinder (Wiles et al. 2010), MiMI (Jayapandian et al. 2007) and PINA (Wu, J. et al. 2009), extended by direct interaction partners and analyzed using Cytoscape (Shannon et al. 2003). The specificity of an interaction partner was defined as the number of this protein's interactions with CR-essential genes as percentage of its total number of interactions. A p-value for the specificity was calculated using a binomial test 3.1, calculating the by chance probability for this many interactions with CR-essential genes (k) at the given number of interactions (n). Interaction partners significantly overlapping with results of the meta-analysis were extracted.

3.2.12 Testing the association of individual datasets to the meta-signature of CR

Genes differentially expressed under a certain condition are often defined as the signature of this condition. Genes enriched for differential expression in these datasets can be called the corresponding meta-signature (Rhodes et al. 2004). To test how well the individual datasets in our analysis associate with the final meta-signature we employed a chi-square test. To create contingency tables for each dataset specifying how many genes are in the meta-signature and how many are not and how many genes are differentially expressed and how many not we used metasignature_test_v1.2.pl (supplement 2).
The chi-square test therefore assesses if genes of each dataset are significantly more likely to be differentially expressed, when they are in the meta-signature. To check that the p-value of the chi-square test indicates genes to be more, not less likely to be differentially expressed, when they are in the meta signature we calculated

$$\frac{\text{"\#diff. exp., in meta-signature"}}{\text{"\#not diff. exp.,in meta-signature"}} \Big/ \frac{\text{"\#diff. exp., not in meta-signature"}}{\text{"\#not diff. exp., not in meta-signature"}}$$

and checked that the result was >1.

3.3 Results

3.3.1 Genes enriched in the number of studies they are found over- / underexpressed

97 and 65 genes were found over- and underexpressed respectively in more datasets than expected by chance below a threshold of the binomial p-value of 0.0005. (In the following these are called "genes enriched for over- / underexpression" or sometimes simply "over- / underexpressed genes"). The full lists of genes are displayed in table 3.6 and 3.7.

MGI Symbol	MGI Description	Entrez ID	total	overexp.	underexp.	p_overexp.
Mt2	metallothionein 2 Gene	17750	59	14	5	1.85E-10
Adh1	alcohol dehydrogenase 1 (class I) Gene	11522	42	12	0	3.50E-10
Per2	period homolog 2 (Drosophila) Gene	18627	44	12	1	6.38E-10
Por	P450 (cytochrome) oxidoreductase Gene	18984	61	13	0	3.41E-9
Inmt	indolethylamine N-methyltransferase Gene	21743	33	10	4	5.51E-9
Dbp	D site albumin promoter binding protein Gene	13170	34	10	4	7.63E-9
Nat8	N-acetyltransferase 8 (GCN5-related, putative) Gene	68396	26	9	0	8.53E-9
Ehhadh	enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase Gene	74147	39	10	0	3.30E-8
Mt1	metallothionein 1 Gene	17748	61	12	2	3.54E-8
Cyp2j6	cytochrome P450, family 2, subfamily j, polypeptide 6 Gene	13110	30	9	0	3.56E-8

Abcg5	ATP-binding cassette, sub-family G (WHITE), member 5 Gene	27409	30	9	0	3.56E-8
Fam107a	family with sequence similarity 107, member A Gene	268709	22	8	0	3.73E-8
Klf15	Kruppel-like factor 15 Gene	66277	32	9	0	6.68E-8
Sds	serine dehydratase Gene	231691	25	8	0	1.18E-7
Fkbp5	FK506 binding protein 5 Gene	14229	59	11	1	2.34E-7
Zbtb16	zinc finger and BTB domain containing 16 Gene	235320	19	7	0	2.46E-7
Angptl4	angiopoietin-like 4 Gene	57875	37	9	2	2.64E-7
Usp2	ubiquitin specific peptidase 2 Gene	53376	60	11	0	2.79E-7
Cobll1	Cobl-like 1 Gene	319876	28	8	0	3.17E-7
Fmo3	flavin containing monooxygenase 3 Gene	14262	29	8	0	4.28E-7
Cyp7a1	cytochrome P450, family 7, subfamily a, polypeptide 1 Gene	13122	39	9	2	4.30E-7
Ablim3	actin binding LIM protein family, member 3 Gene	319713	21	7	1	5.42E-7
Nr1i3	nuclear receptor subfamily 1, group I, member 3 Gene	12355	40	9	0	5.43E-7
Cyp4a14	cytochrome P450, family 4, subfamily a, polypeptide 14 Gene	13119	32	8	0	9.80E-7
Sult1d1	sulfotransferase family 1D, member 1 Gene	53315	45	9	3	1.57E-6
Herpud1	homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1 Gene	64209	45	9	2	1.57E-6
LOC 100047583	similar to apolipoprotein D	100047583	5	4	0	1.96E-6
Ctgf	connective tissue growth factor Gene	14219	35	8	0	2.05E-6
Slc37a4	solute carrier family 37 (glucose-6-phosphate transporter), member 4 Gene	14385	35	8	0	2.05E-6
Tenc1	tensin like C1 domain-containing phosphatase Gene	209039	60	10	0	2.41E-6
Wee1	WEE 1 homolog 1 (S. pombe) Gene	22390	37	8	2	3.22E-6
Klf9	Kruppel-like factor 9 Gene	16601	51	9	0	4.70E-6
Ppara	peroxisome proliferator activated receptor alpha Gene	19013	40	8	1	5.99E-6
Trp53i13	transformation related protein 53 inducible protein 13 Gene	216964	29	7	1	6.10E-6
Irs2	insulin receptor substrate 2 Gene	384783	29	7	1	6.10E-6
Fam195a	family with sequence similarity 195, member A Gene	68241	20	6	0	7.23E-6
Acot4	acyl-CoA thioesterase 4 Gene	171282	30	7	0	7.78E-6
Ntf3	neurotrophin 3 Gene	18205	42	8	0	8.79E-6
Tmem218	transmembrane protein 218 Gene	66279	21	6	0	9.91E-6
Aldh1a1	aldehyde dehydrogenase family 1, subfamily A1 Gene	11668	56	9	2	1.05E-5
Gm6957	predicted gene 6957 Gene	629219	13	5	0	1.09E-5
Pim3	proviral integration site 3 Gene	223775	57	9	0	1.21E-5
Klf9	Kruppel-like factor 9 Gene	70273	14	5	0	1.67E-5
Aqp6	aquaporin 6 Gene	11831	23	6	2	1.77E-5

Cyp2b13	cytochrome P450, family 2, subfamily b, polypeptide 13 Gene	13089	23	6	1	1.77E-5
Decr2	2-4-dienoyl-Coenzyme A reductase 2, peroxisomal Gene	26378	24	6	0	2.30E-5
Cry1	cryptochrome 1 (photolyase-like) Gene	12952	49	8	0	2.87E-5
Tsc22d3	TSC22 domain family, member 3 Gene	14605	26	6	0	3.77E-5
Cbr1	carbonyl reductase 1 Gene	12408	38	7	0	4.04E-5
Rgs16	regulator of G-protein signaling 16 Gene	19734	27	6	2	4.75E-5
Hacl1	2-hydroxyacyl-CoA lyase 1 Gene	56794	27	6	0	4.75E-5
Sult1c2	sulfotransferase family, cytosolic, 1C, member 2 Gene	69083	27	6	1	4.75E-5
Gys2	glycogen synthase 2 Gene	232493	27	6	0	4.75E-5
Cyp2e1	cytochrome P450, family 2, subfamily e, polypeptide 1 Gene	13106	39	7	0	4.82E-5
Plin5	perilipin 5 Gene	66968	17	5	1	4.83E-5
Cpt1a	carnitine palmitoyltransferase 1a, liver Gene	12894	53	8	1	5.16E-5
Igfbp2	insulin-like growth factor binding protein 2 Gene	16008	40	7	1	5.72E-5
Arrdc2	arrestin domain containing 2 Gene	70807	40	7	0	5.72E-5
4833417 J20Rik	4833417J20Rik RIKEN cDNA 4833417J20 gene	74604	4	3	0	6.24E-5
4432414 F05Rik	4432414F05Rik RIKEN cDNA 4432414F05 gene	77027	4	3	0	6.24E-5
Agxt2l1	alanine-glyoxylate aminotransferase 2-like 1 Gene	71760	18	5	0	6.55E-5
St3gal5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5 Gene	20454	41	7	1	6.74E-5
Slc25a25	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 25 Gene	227731	41	7	0	6.74E-5
Lpin1	lipin 1 Gene	14245	29	6	1	7.30E-5
Gpr146	G protein-coupled receptor 146 Gene	80290	31	6	0	1.08E-4
Adcy1	adenylate cyclase 1 Gene	432530	11	4	0	1.15E-4
Ifrd1	interferon-related developmental regulator 1 Gene	15982	45	7	0	1.25E-4
Mat1a	methionine adenosyltransferase I, alpha Gene	11720	60	8	0	1.28E-4
Acot12	acyl-CoA thioesterase 12 Gene	74156	32	6	0	1.31E-4
Nfkbia	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha Gene	18035	61	8	0	1.44E-4
Epb4.1	erythrocyte protein band 4.1 Gene	269587	61	8	1	1.44E-4
Hsd17b2	hydroxysteroid (17-beta) dehydrogenase 2 Gene	15486	46	7	5	1.44E-4
Sun2	Sad1 and UNC84 domain containing 2 Gene	223697	34	6	1	1.86E-4
Mgp	matrix Gla protein Gene	17313	48	7	1	1.89E-4
Aldh1a7	aldehyde dehydrogenase family 1, subfamily A7 Gene	26358	35	6	2	2.19E-4
Sult3a1	sulfotransferase family 3A, member 1 Gene	57430	23	5	1	2.32E-4
Niacr1	niacin receptor 1 Gene	80885	13	4	0	2.38E-4

BC089597	cDNA sequence BC089597 Gene	216454	13	4	0	2.38E-4
Dusp1	dual specificity phosphatase 1 Gene	19252	36	6	0	2.57E-4
Klf10	Kruppel-like factor 10 Gene	21847	36	6	0	2.57E-4
Rhbdd2	rhomboid domain containing 2 Gene	215160	51	7	0	2.79E-4
Sult1a1	sulfotransferase family 1A, phenol-preferring, member 1 Gene	20887	37	6	0	3.01E-4
Decr1	2,4-dienoyl CoA reductase 1, mitochondrial Gene	67460	37	6	0	3.01E-4
Cd163	CD163 antigen Gene	93671	14	4	1	3.27E-4
Plcxd3	phosphatidylinositol-specific phospholipase C, X domain containing 3 Gene	239318	14	4	0	3.27E-4
Bnip3	BCL2/adenovirus E1B interacting protein 3 Gene	100042570	14	4	0	3.27E-4
Fzd1	frizzled homolog 1 (Drosophila) Gene	14362	38	6	2	3.50E-4
Per1	period homolog 1 (Drosophila) Gene	18626	38	6	1	3.50E-4
Enpep	glutamyl aminopeptidase Gene	13809	25	5	0	3.51E-4
Sall1	sal-like 1 (Drosophila) Gene	58198	25	5	0	3.51E-4
Slc25a42	solute carrier family 25, member 42 Gene	73095	25	5	1	3.51E-4
Zfp354a	zinc finger protein 354A Gene	21408	54	7	0	4.00E-4
Pla2g12a	phospholipase A2, group XIIA Gene	66350	39	6	1	4.04E-4
Map3k6	mitogen-activated protein kinase kinase kinase 6 Gene	53608	26	5	0	4.25E-4
Rbp7	retinol binding protein 7, cellular Gene	63954	26	5	3	4.25E-4
Rhobtb1	Rho-related BTB domain containing 1 Gene	69288	26	5	0	4.25E-4
Crym	crystallin, mu Gene	12971	15	4	0	4.37E-4
Plin4	perilipin 4 Gene	57435	15	4	0	4.37E-4
LOC 100044830	similar to acyl-CoA thioesterase	100044830	15	4	0	4.37E-4
Smoc1	SPARC related modular calcium binding 1 Gene	64075	55	7	0	4.48E-4
Tob1	transducer of ErbB-2.1 Gene	22057	40	6	0	4.66E-4

Table 3.6: Genes found overexpressed in more datasets than expected by chance below the threshold of the binomial p-value of 0.0005. The total number of datasets the gene was studied in, the number of datasets in which it was over- and underexpressed and the binomial p-value for enrichment for overexpression are shown.

MGI Symbol	MGI Description	EntrezID	total	overexp.	underexp.	p_under exp.
Slc6a6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6 Gene	21366	60	1	12	7.66E-9
Car3	carbonic anhydrase 3 Gene	12350	49	0	11	8.86E-9
Cyp2j5	cytochrome P450, family 2, subfamily j, polypeptide 5 Gene	13109	25	0	8	4.64E-8
Dhcr7	7-dehydrocholesterol reductase Gene	13360	49	0	10	1.11E-7
Arntl	aryl hydrocarbon receptor nuclear translocator-like Gene	11865	63	3	11	1.41E-7
Zfp64	zinc finger protein 64 Gene	22722	34	1	8	6.52E-7
Srebf1	sterol regulatory element binding transcription factor 1 Gene	20787	60	1	10	8.13E-7
Es31	esterase 31 Gene	382053	25	1	7	9.14E-7

Gck	glucokinase Gene	103988	41	1	8	2.98E-6
Col15a1	collagen, type XV, alpha 1 Gene	12819	32	1	7	5.58E-6
G0s2	G0/G1 switch gene 2 Gene	14373	33	3	7	6.95E-6
Insig1	insulin induced gene 1 Gene	231070	33	1	7	6.95E-6
C9	complement component 9 Gene	12279	36	1	7	1.28E-5
Phlda1	pleckstrin homology-like domain, family A, member 1 Gene	21664	39	1	7	2.23E-5
Hspa5	heat shock protein 5 Gene	14828	69	0	9	2.27E-5
Irgm1	immunity-related GTPase family M member 1 Gene	15944	28	0	6	3.00E-5
Dpp9	dipeptidylpeptidase 9 Gene	224897	28	0	6	3.00E-5
Alas2	aminolevulinic acid synthase 2, erythroid Gene	11656	58	3	8	4.28E-5
Tmem 132d	transmembrane protein 132D Gene	243274	4	0	3	4.34E-5
Irf7	interferon regulatory factor 7 Gene	54123	30	1	6	4.56E-5
Fabp5	fatty acid binding protein 5, epidermal Gene	16592	59	3	8	4.85E-5
Tnfsf10	tumor necrosis factor (ligand) superfamily, member 10 Gene	22035	19	0	5	4.89E-5
Acly	ATP citrate lyase Gene	104112	60	2	8	5.49E-5
Scly	selenocysteine lyase Gene	50880	31	1	6	5.54E-5
C4bp	complement component 4 binding protein Gene	12269	20	0	5	6.40E-5
Ifi27l2a	interferon, alpha-inducible protein 27 like 2A Gene	76933	20	0	5	6.40E-5
Casc5	cancer susceptibility candidate 5 Gene	76464	11	1	4	7.14E-5
Serpinh1	serine (or cysteine) peptidase inhibitor, clade H, member 1 Gene	12406	63	4	8	7.83E-5
Ifih1	interferon induced with helicase C domain 1 Gene	71586	33	0	6	8.03E-5
1110051 M20Rik	RIKEN cDNA 1110051M20 gene Gene	228356	33	0	6	8.03E-5
Ttll12	tubulin tyrosine ligase-like family, member 12 Gene	223723	21	0	5	8.25E-5
Aqp8	aquaporin 8 Gene	11833	34	1	6	9.56E-5
Cldn1	claudin 1 Gene	12737	34	1	6	9.56E-5
Nr1d1	nuclear receptor subfamily 1, group D, member 1 Gene	217166	34	3	6	9.56E-5
Ghr	growth hormone receptor Gene	14600	65	0	8	9.82E-5
R3hdm2	R3H domain containing 2 Gene	71750	49	0	7	1.02E-4
Hipk2	homeodomain interacting protein kinase 2 Gene	15258	36	0	6	1.33E-4
Rsc1a1	regulatory solute carrier protein, family 1, member 1 Gene	69994	13	0	4	1.49E-4
Cyp2f2	cytochrome P450, family 2, subfamily f, polypeptide 2 Gene	13107	37	0	6	1.56E-4
Cxcl9	chemokine (C-X-C motif) ligand 9 Gene	17329	37	0	6	1.56E-4
Hsd3b2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2 Gene	15493	24	0	5	1.63E-4
Mup4	major urinary protein 4 Gene	17843	24	0	5	1.63E-4
Extl1	exostoses (multiple)-like 1 Gene	56219	24	1	5	1.63E-4

Sc5d	sterol-C5-desaturase (fungal ERG3, delta-5-desaturase) homolog (S. cerevisae) Gene	235293	38	0	6	1.82E-4
G6pdx	glucose-6-phosphate dehydrogenase X-linked Gene	14381	54	2	7	1.92E-4
Scrt1	scratch homolog 1, zinc finger protein (Drosophila) Gene	170729	25	0	5	2.00E-4
Ptprj	protein tyrosine phosphatase, receptor type, J Gene	668629	14	1	4	2.05E-4
Psmb8	proteasome (prosome, macropain) subunit, beta type 8 (large multifunctional peptidase 7) Gene	16913	39	0	6	2.11E-4
Slc10a2	solute carrier family 10, member 2 Gene	20494	39	0	6	2.11E-4
Actg1	actin, gamma, cytoplasmic 1 Gene	11465	55	1	7	2.15E-4
Comt1	catechol-O-methyltransferase 1 Gene	12846	55	2	7	2.15E-4
Ntn3	netrin 3 Gene	18209	15	0	4	2.75E-4
2900086 B20Rik	RIKEN cDNA 2900086B20 gene	73074	15	0	4	2.75E-4
Stac3	SH3 and cysteine rich domain 3 Gene	237611	15	0	4	2.75E-4
Mmp15	matrix metallopeptidase 15 Gene	17388	27	0	5	2.93E-4
Gtf2ird1	general transcription factor II I repeat domain-containing 1 Gene	57080	27	0	5	2.93E-4
Phf19	PHD finger protein 19 Gene	74016	27	0	5	2.93E-4
Inhbe	inhibin beta E Gene	16326	42	2	6	3.20E-4
Col3a1	collagen, type III, alpha 1 Gene	12825	59	1	7	3.35E-4
Cdc42ep2	CDC42 effector protein (Rho GTPase binding) 2 Gene	104252	28	1	5	3.50E-4
1110054 M08Rik	RIKEN cDNA 1110054M08 gene	68841	16	1	4	3.60E-4
2810051 F02Rik	RIKEN cDNA 2810051F02 gene	72704	7	0	3	3.61E-4
Gm13768	predicted gene 13768	627525	7	0	3	3.61E-4
Gm7450	predicted gene 7450	665017	7	0	3	3.61E-4
LOC 677259	similar to Ornithine decarboxylase (ODC)	677259	7	0	3	3.61E-4
LOC 100045005	similar to Deltex3	100045005	7	0	3	3.61E-4
Dnase1l2	deoxyribonuclease 1-like 2 Gene	100047816	7	0	3	3.61E-4
LOC 100048733	similar to WAP four-disulfide core domain 2	100048733	7	0	3	3.61E-4
D0H4S114	DNA segment, human D4S114 Gene	27528	60	0	7	3.72E-4
Litaf	LPS-induced TN factor Gene	56722	60	0	7	3.72E-4
Pdia3	protein disulfide isomerase associated 3 Gene	14827	62	0	7	4.56E-4
Ly6e	lymphocyte antigen 6 complex, locus E Gene	17069	62	2	7	4.56E-4
Hspb7	heat shock protein family, member 7 (cardiovascular) Gene	29818	30	1	5	4.89E-4

Table 3.7.: Genes found underexpressed in more datasets than expected by chance below the threshold of the binomial p-value of 0.0005. The total number of datasets the gene was studied in, the number of datasets in which it was over- and underexpressed and the binomial p-value for enrichment for underexpression are shown

Besides providing researchers with a list of well-known genes, for some giving a first hint towards association

with CR, for others contributing to the already existing evidence for such association, the aim of our meta-analysis is also to find interesting behaviour of sequences with unknown function, often annotated as ESTs or pseudogenes. To this end we found LOC100047583 (Entrez ID: 100047583, similar to apolipoprotein D), 4833417J20Rik (74604, RIKEN cDNA 4833417J20 gene) and 4432414F05Rik (77027, RIKEN cDNA 4432414F05 gene) among the genes enriched for overexpression, which are classified as protein coding genes, but are on RefSeq status "model" or without any, miss annotation on the reference assembly and generally seem to be studied little (22/07/10). Also found among the genes enriched for overexpression was the pseudogene LOC100044830 (100044830, similar to acyl-CoA thioesterase).
Similarly among genes enriched for underexpression we detected 1110051M20Rik (67829, meanwhile replaced by 228356, RIKEN cDNA 1110051M20 gene), 2900086B20Rik (73074, RIKEN cDNA 2900086B20 gene), 1110054M08Rik (68841, RIKEN cDNA 1110054M08 gene), LOC677259 (677259, similar to Ornithine decarboxylase (ODC)), LOC100045005 (100045005, similar to Deltex3) and LOC100048733 (100048733, similar to WAP four-disulfide core domain 2). These findings might assign interesting functions as transcribed genes to these sequences, however note that the detection of expression of (pseudo)genes similar to other genes might also result from the lack of specificity of the microarray probe to distinguish between the two sequences. We also found 2810051F02Rik (72704, RIKEN cDNA 2810051F02 gene) among the genes enriched for underexpression, which is meanwhile replaced by the validated NCBI entry "antisense Igf2r RNA" (Airn, 104103), which might therefore be an interesting non-coding RNA contributing to the mechanism of CR.

Table 3.8[6] presents the 10 genes most significantly enriched for over- / underexpression, a description of their function and indications of known relationships with CR.

Most of these genes are somehow associated with candidate GOs as found in the functional analysis ("3.3.2 Functional categories of genes differentially expressed with CR"), especially circadian clock, lipid metabolism and xenobiotic metabolism. Some of these genes have important regulatory functions in these categories, in particular *Per2* as master-regulator and *Dbp* as another transcription factor regulating the circadian clock and *Srebf1* as a transcription factor regulating sterol metabolism.
Transcriptional levels of *Per2* oscillate diurnally in the suprachiasmatic nucleus (SCN) of the hypothalamus and are supposedly set by light (Lamont et al. 2007). The timing of oscillators in peripheral tissues is controlled by the SCN when food is available ad libitum. If feeding is however temporally limited the time of feeding is a more important regulator for peripheral oscillators (Girotti et al. 2009). If additionally the level of food intake is altered also the timing of clock gene expression in the SCN changes, arguing for metabolic regulation. Therefore both the changed amount of food, but also the fact that CR might also change the timing of food availability compared to AL might have an important influence on changed expression levels of clock genes. *Srebf1* is an interesting candidate, since it has been linked to the mechanism, by which resveratrol could increase life-span in obese mice (Wang, G. et al. 2009). Its expression levels also have been already studied in the context of CR, showing that, while its liver specific expression does not change in the first week of CR (Mulligan et al. 2008), its levels are influenced by CR and refeeding in adipose tissue (Stelmanska et al. 2004).
Zfp64, as a little understood co-activator in the notch pathway, also has the potential to be an interesting candidate concerning the mechanism of CR.
All of the top 10 genes enriched for overexpression were overexpressed in more than 3 different tissues, while many of the underexpressed were only found underexpressed in one or two tissues. This may however also have to do with the fact that they were generally underexpressed in less datasets than the overexpressed were overexpressed. That *Gck* was found underexpressed in liver only makes sense, since this gene is assumed to be liver and beta-cell specific and pancreas was not tested in our datasets.
It is also noteworthy that many of the top overexpressed genes were found underexpressed in a considerable number of datasets and vice versa, even though among all significant genes the number of datasets of opposite differential expression is rather low (on average around 1). This might mean that the top genes are highly regulated.

[6]Table 3.8: The 10 genes most significantly enriched for over- and underexpression and description of their function; it is given which enriched functional category, as determined in the functional analysis ("3.3.2. Functional categories of genes differentially expressed with CR") they are related to. (This does not necessarily mean that they are directly classified with a GO-term exactly like this). The number of different tissues they are over- / underexpressed in is shown. Information not from stated references is from www.genecards.org; references: 1: (Waddington Lamont et al. 2007), 2: (Girotti et al. 2009), 3: (Kranendonk et al. 2008), 4: (H. Saito et al. 2008), 5: (Sakamoto et al. 2008), 6: (Guang-Li Wang et al. 2009), 7: (Mulligan et al. 2008), 8: (Stelmanska et al. 2004)

a,

Gene Symbol	Gene Name	Function	related candidate GOs	#tissues	comment	ref.
Mt2	metallothionein 2 Gene	binds various metals	cellular copper ion homeostasis	7	most significant gene; also reported by Swindell, 2008 and 2009; underexpressed in 5 tissues	
Adh1	alcohol dehydrogenase 1 (class I) Gene	metabolizes besides ethanol also retinol, etc.		5		
Per2	period homolog 2 (Drosophila) Gene	master regulator of circadian clock	circadian clock	6	transcriptional levels oscillate diurnally	1,2
Por	P450 (cytochrome) oxidoreductase Gene	transfers electrons from NADPH to among others P450 and heme oxigenase	xenobiotic metabolism	4		3
Inmt	indolethylamine N-methyltransferase Gene	N-methylation of indoles (endogenous and xenobiotic)	xenobiotic metabolism	4	underexpressed in 4 datasets	
Dbp	D site albumin promoter binding protein Gene	transcription factor that modulates clock-output genes	circadian clock	4	clock-controlled gene; underexpressed in 4 datasets	4
Nat8	N-acetyltransferase 8 (GCN5-related, putative) Gene	not yet clear		3		
Ehhadh	enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase Gene	part of the peroxisomal beta-oxidation pathway	lipid metabolism	4		
Mt1	metallothionein 1 Gene	binds various metals	copper ion binding	4	also reported by Swindell, 2008; underexpressed in 2 datasets	
Cyp2j6	cytochrome P450, family 2, subfamily j, polypeptide 6 Gene	arachidonic and linoleic acid and retinoid metabolism	lipid metabolism, retinol metabolism	4		

b,

Gene Symbol	Gene Name	Function	related candidate GOs	#tissues	comment	ref.
Slc6a6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6 Gene	transports both taurine and beta-alanine		2	most significant gene; overexpressed in 1 dataset	
Car3	carbonic anhydrase 3 Gene	catalyze the reversible hydration of carbon dioxide		only in liver		
Cyp2j5	cytochrome P450, family 2, subfamily j, polypeptide 5 Gene	arachidonic acid epoxygenase	lipid metabolism	2		
Dhcr7	7-dehydrocholesterol reductase Gene	Production of cholesterol by reduction of C7-C8 double bond of 7-dehydrocholesterol	lipid metabolism; cholesterol metabolism	3		
Arntl	aryl hydrocarbon receptor nuclear translocator-like Gene	heterodimer with Clock is transcription factor that regulates Per1 and other clock-gens	circadian clock	4	overexpressed in 4 datasets	
Zfp64	zinc finger protein 64 Gene	coactivator of Notch; regulates differentiation		4	overexpressed in 1 dataset	5
Srebf1	sterol regulatory element binding transcription factor 1 Gene	transcription factor that regulates genes involved in sterol biosynthesis	lipid metabolism, sterol metabolism	2	resveratrol inhibits expr. of SREBP1 in cell model of steatosis; change in Srebf-1 levels in adip. tissue during CR and refeeding; overexp. in 2 datasets	6 – 8
Es31	esterase 31 Gene	hydrolysis of esters and amide bonds; involved in detoxification of xenobiotics and maybe in lipid metabolism	xenobiotic metabolism	2	overexpressed in 1 dataset	
Gck	glucokinase Gene	catalyzes initial step of glucose utilization by the beta-cell and liver; effective when glucose is abundant		only in liver	overexpressed in 1 dataset	
Col 15a1	collagen, type XV, alpha 1 Gene	structural protein, especially stabilizing microvessels and muscle cells		4	overexpressed in 4 datasets	

Table 3.8: see footnote 6

3.3.2 Functional categories of genes differentially expressed with CR

3.3.2.1 GO-terms enriched in studies in which associated genes are found over- / underexpressed - GO-analysis

187 and 153 GO-terms were found enriched for studies in which their associated genes were over- and underexpressed respectively according to the analysis described in "3.2.6.1 Determining functional categories enriched in the meta-analysis datasets - GO-analysis" (binomial p-value < 0.001). These GO-terms are shown in table 3.9 and 3.10.

GO term	GO	total	overexp.	underexp.	p_overexp.
lipid metabolic process	GO:0006629	8255	352	216	8.01E-24
rhythmic process	GO:0048511	899	73	27	6.52E-19
monooxygenase activity	GO:0004497	2803	147	96	8.69E-18
circadian rhythm	GO:0007623	1025	72	45	2.15E-15
detoxification of copper ion	GO:0010273	181	26	8	3.77E-13
retinol metabolic process	GO:0042572	298	33	5	5.46E-13
cellular_component	GO:0005575	219270	5771	4906	6.54E-13
molecular_function	GO:0003674	232675	6087	4986	4.46E-12
NADPH-hemoprotein reductase activity	GO:0003958	149	22	0	1.34E-11
microsome	GO:0005792	10612	366	316	1.57E-11
acyl-CoA metabolic process	GO:0006637	749	51	9	6.73E-11
oxidoreductase activity	GO:0016491	20263	630	469	1.21E-10
nitric oxide mediated signal transduction	GO:0007263	307	30	9	1.35E-10
oxidation reduction	GO:0055114	19926	620	461	1.49E-10
acetaldehyde biosynthetic process	GO:0046186	42	12	0	2.03E-10
retinoic acid metabolic process	GO:0042573	456	37	11	2.27E-10
extracellular region	GO:0005576	42731	1227	1102	2.51E-10
fatty acid metabolic process	GO:0006631	3408	143	83	3.03E-10
catalytic activity	GO:0003824	28555	850	617	3.43E-10
biological_process	GO:0008150	237351	6151	5091	4.41E-10
cellular zinc ion homeostasis	GO:0006882	306	29	10	5.55E-10
metabolic process	GO:0008152	22860	694	523	6.80E-10
tyrosine-ester sulfotransferase activity	GO:0017067	82	15	3	1.08E-9
nitric oxide catabolic process	GO:0046210	61	13	0	1.92E-9
flavin-containing monooxygenase activity	GO:0004499	132	18	1	3.44E-9
amine N-methyltransferase activity	GO:0030748	33	10	4	3.49E-9
iron ion binding	GO:0005506	4581	175	145	3.74E-9
alkane 1-monooxygenase activity	GO:0018685	54	12	1	4.82E-9
benzaldehyde dehydrogenase (NAD^+) activity	GO:0018479	91	15	4	4.85E-9
retinoid metabolic process	GO:0001523	185	21	2	5.25E-9
carboxylesterase activity	GO:0004091	1170	63	30	5.60E-9
late recombination nodule	GO:0005715	26	9	0	5.63E-9
2,4-dienoyl-CoA reductase (NADPH) activity	GO:0008670	61	12	0	2.10E-8
intrinsic to endoplasmic reticulum membrane	GO:0031227	480	34	10	3.47E-8
palmitoyl-CoA hydrolase activity	GO:0016290	368	29	3	3.52E-8
ethanol catabolic process	GO:0006068	81	13	0	7.03E-8
DNA photolyase activity	GO:0003913	84	13	1	1.10E-7
ethanol binding	GO:0035276	118	15	2	1.74E-7
negative regulation of lipoprotein lipase activity	GO:0051005	37	9	2	1.76E-7
MDM2 binding	GO:0070215	88	13	1	1.92E-7
aryl sulfotransferase activity	GO:0004062	136	16	4	2.03E-7
lyase activity	GO:0016829	4802	173	128	2.04E-7
drug metabolic process	GO:0017144	249	22	6	2.13E-7
steroid metabolic process	GO:0008202	2076	89	73	2.33E-7
dodecenoyl-CoA delta-isomerase activity	GO:0004165	155	17	0	2.38E-7
cholesterol 7-alpha-monooxygenase activity	GO:0008123	39	9	2	2.88E-7
regulation of bile acid biosynthetic process	GO:0070857	39	9	2	2.88E-7

positive regulation of bile acid biosynthetic process	GO:0070859	39	9	2	2.88E-7
cellular response to cholesterol	GO:0071397	39	9	2	2.88E-7
extracellular space	GO:0005615	18369	548	484	2.91E-7
electron carrier activity	GO:0009055	3478	132	96	3.79E-7
L-serine ammonia-lyase activity	GO:0003941	52	10	0	3.93E-7
L-threonine ammonia-lyase activity	GO:0004794	52	10	0	3.93E-7
aromatase activity	GO:0070330	711	41	31	4.05E-7
cellular metal ion homeostasis	GO:0006875	161	17	4	4.11E-7
transporter activity	GO:0005215	7159	239	192	4.78E-7
regulation of cholesterol metabolic process	GO:0090181	222	20	3	5.45E-7
peroxisome	GO:0005777	4072	149	76	5.72E-7
acyl-CoA thioesterase activity	GO:0016291	265	22	1	6.17E-7
fatty acid (omega-1)-hydroxylase activity	GO:0008393	32	8	0	6.84E-7
icosanoid biosynthetic process	GO:0046456	32	8	0	6.84E-7
behavioral response to ethanol	GO:0048149	186	18	2	6.97E-7
myeloid progenitor cell differentiation	GO:0002318	151	16	0	8.49E-7
pyridoxal phosphate binding	GO:0030170	1959	83	60	9.06E-7
glucose-6-phosphate transport	GO:0015760	134	15	5	9.18E-7
histone phosphorylation	GO:0016572	152	16	3	9.28E-7
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	GO:0016712	534	33	26	1.18E-6
ethanol oxidation	GO:0006069	120	14	2	1.27E-6
glucose-6-phosphate transmembrane transporter activity	GO:0015152	35	8	0	1.43E-6
negative regulation of chemokine production	GO:0032682	35	8	0	1.43E-6
cytosolic calcium ion transport	GO:0060401	35	8	0	1.43E-6
positive regulation of cardiac muscle contraction	GO:0060452	35	8	0	1.43E-6
extracellular matrix constituent secretion	GO:0070278	35	8	0	1.43E-6
positive regulation of G0 to G1 transition	GO:0070318	35	8	0	1.43E-6
cholesterol catabolic process	GO:0006707	176	17	2	1.44E-6
steroid hormone receptor activity	GO:0003707	1859	79	40	1.45E-6
cellular homeostasis	GO:0019725	60	10	0	1.57E-6
multicellular organismal homeostasis	GO:0048871	60	10	0	1.57E-6
methionine adenosyltransferase activity	GO:0004478	123	14	0	1.71E-6
S-adenosylmethionine biosynthetic process	GO:0006556	123	14	0	1.71E-6
heme binding	GO:0020037	3600	132	111	2.15E-6
3-chloroallyl aldehyde dehydrogenase activity	GO:0004028	266	21	5	2.44E-6
ligand-dependent nuclear receptor activity	GO:0004879	1889	79	41	2.60E-6
nerve development	GO:0021675	204	18	3	2.64E-6
positive regulation of cholesterol esterification	GO:0010873	78	11	3	2.65E-6
thiolester hydrolase activity	GO:0016790	314	23	1	2.97E-6
insulin-like growth factor binding	GO:0005520	886	45	28	3.45E-6
symporter activity	GO:0015293	3572	130	100	3.63E-6
regulation of fatty acid oxidation	GO:0046320	82	11	1	4.37E-6
long-chain fatty acid metabolic process	GO:0001676	443	28	7	4.75E-6
response to glucocorticoid stimulus	GO:0051384	1021	49	16	6.09E-6
neurotrophin receptor binding	GO:0005165	42	8	0	6.20E-6
9-cis-retinoic acid metabolic process	GO:0042905	56	9	2	7.13E-6
alcohol dehydrogenase (NAD) activity	GO:0004022	159	15	2	7.79E-6
linoleic acid metabolic process	GO:0043651	87	11	1	7.84E-6
glycogen (starch) synthase activity	GO:0004373	57	9	1	8.28E-6
optic cup morphogenesis involved in camera-type eye development	GO:0002072	88	11	2	8.76E-6
lauric acid metabolic process	GO:0048252	46	8	2	1.26E-5

9-cis-retinoic acid biosynthetic process	GO:0042904	207	17	4	1.28E-5
carbon-carbon lyase activity	GO:0016830	60	9	1	1.28E-5
amino acid binding	GO:0016597	1110	51	16	1.29E-5
nitrate transmembrane transporter activity	GO:0015112	23	6	2	1.35E-5
nitrate transport	GO:0015706	23	6	2	1.35E-5
fatty acid beta-oxidation	GO:0006635	834	41	7	1.91E-5
carnitine O-palmitoyltransferase activity	GO:0004095	172	15	2	2.00E-5
cholesterol homeostasis	GO:0042632	1160	52	22	2.05E-5
leg morphogenesis	GO:0035110	36	7	0	2.06E-5
retinol dehydrogenase activity	GO:0004745	330	22	0	2.09E-5
growth factor activity	GO:0008083	4826	162	112	2.16E-5
water transport	GO:0006833	508	29	22	2.20E-5
positive regulation of lipid metabolic process	GO:0045834	65	9	2	2.48E-5
protein homotetramerization	GO:0051289	1146	51	23	2.93E-5
15-hydroxyprostaglandin dehydrogenase (NADP+) activity	GO:0047021	38	7	0	2.98E-5
prostaglandin-E2 9-reductase activity	GO:0050221	38	7	0	2.98E-5
water channel activity	GO:0015250	363	23	19	3.06E-5
ectoplasm	GO:0043265	27	6	0	3.63E-5
progesterone receptor signaling pathway	GO:0050847	85	10	1	3.79E-5
lactosylceramide alpha-2,3-sialyltransferase activity	GO:0047291	41	7	1	4.99E-5
regulation of cell growth	GO:0001558	1295	55	38	5.13E-5
brown fat cell differentiation	GO:0050873	1236	53	34	5.44E-5
succinate transmembrane transporter activity	GO:0015141	107	11	4	5.60E-5
succinate transport	GO:0015744	107	11	4	5.60E-5
NADP or NADPH binding	GO:0050661	1152	50	21	6.44E-5
neutrophil homeostasis	GO:0001780	75	9	1	7.85E-5
drug binding	GO:0008144	1877	72	41	1.03E-4
arachidonic acid monooxygenase activity	GO:0008391	96	10	2	1.08E-4
cytoplasmic sequestering of NF-kappaB	GO:0007253	97	10	0	1.18E-4
phosphatidate phosphatase activity	GO:0008195	373	22	11	1.27E-4
male germ-line stem cell division	GO:0048133	63	8	0	1.30E-4
endocrine pancreas development	GO:0031018	429	24	9	1.44E-4
polysaccharide binding	GO:0030247	430	24	21	1.49E-4
arachidonic acid metabolic process	GO:0019369	352	21	12	1.52E-4
protein homooligomerization	GO:0051260	2177	80	50	1.72E-4
negative regulation of astrocyte differentiation	GO:0048712	208	15	10	1.73E-4
amine sulfotransferase activity	GO:0047685	23	5	1	1.85E-4
positive regulation of adiponectin secretion	GO:0070165	13	4	0	1.98E-4
nicotinic acid receptor activity	GO:0070553	13	4	0	1.98E-4
3-hydroxyacyl-CoA dehydrogenase activity	GO:0003857	308	19	1	2.01E-4
interleukin-6-mediated signaling pathway	GO:0070102	124	11	3	2.11E-4
response to steroid hormone stimulus	GO:0048545	496	26	14	2.18E-4
sulfate assimilation	GO:0000103	168	13	5	2.31E-4
4-nitrophenol metabolic process	GO:0018960	37	6	0	2.32E-4
3'-phosphoadenosine 5'-phosphosulfate binding	GO:0050656	37	6	0	2.32E-4
sulfation	GO:0051923	37	6	0	2.32E-4
pancreatic ribonuclease activity	GO:0004522	214	15	1	2.36E-4
positive regulation of collagen biosynthetic process	GO:0032967	193	14	5	2.64E-4
short-chain fatty acid metabolic process	GO:0046459	53	7	2	2.66E-4
inductive cell-cell signaling	GO:0031129	25	5	0	2.81E-4
nucleolar fragmentation	GO:0007576	54	7	0	2.99E-4
glutathione transferase activity	GO:0004364	890	39	17	3.15E-4
cellular amino acid metabolic process	GO:0006520	741	34	15	3.27E-4
enoyl-CoA hydratase activity	GO:0004300	245	16	1	3.30E-4

glucose homeostasis	GO:0042593	1562	60	42	3.48E-4
negative regulation of epidermal growth factor receptor activity	GO:0007175	56	7	0	3.76E-4
response to testosterone stimulus	GO:0033574	178	13	1	4.04E-4
lipid catabolic process	GO:0016042	2862	98	78	4.12E-4
detection of mechanical stimulus involved in equilibrioception	GO:0050973	57	7	1	4.20E-4
nerve growth factor binding	GO:0048406	75	8	1	4.40E-4
regulation of insulin secretion	GO:0050796	607	29	27	4.54E-4
sodium ion transport	GO:0006814	3788	124	94	4.59E-4
thiosulfate transmembrane transporter activity	GO:0015117	58	7	2	4.68E-4
malate transmembrane transporter activity	GO:0015140	58	7	2	4.68E-4
secondary active transmembrane transporter activity	GO:0015291	58	7	2	4.68E-4
thiosulfate transport	GO:0015709	58	7	2	4.68E-4
malate transport	GO:0015743	58	7	2	4.68E-4
urea transport	GO:0015840	181	13	3	4.73E-4
cholesterol esterification	GO:0034435	42	6	3	4.74E-4
proline racemase activity	GO:0018112	28	5	0	4.90E-4
endosomal lumen acidification	GO:0048388	59	7	0	5.20E-4
mitochondrial inner membrane	GO:0005743	12407	355	162	5.42E-4
FMN binding	GO:0010181	388	21	11	5.54E-4
ligand-regulated transcription factor activity	GO:0003706	97	9	4	5.58E-4
negative regulation of thymocyte apoptosis	GO:0070244	78	8	2	5.74E-4
aconitate hydratase activity	GO:0003994	118	10	1	5.79E-4
glycerol transport	GO:0015793	118	10	3	5.79E-4
ammonia assimilation cycle	GO:0019676	98	9	0	6.02E-4
aldehyde dehydrogenase (NAD) activity	GO:0004029	446	23	8	6.06E-4
sensory perception of chemical stimulus	GO:0007606	163	12	11	6.18E-4
response to muscle activity	GO:0014850	119	10	1	6.19E-4
positive regulation of fatty acid beta-oxidation	GO:0032000	141	11	2	6.34E-4
nucleotide-binding oligomerization domain containing 1 signaling pathway	GO:0070427	80	8	0	6.80E-4
photoreceptor outer segment	GO:0001750	928	39	17	6.93E-4
response to stress	GO:0006950	4085	131	113	7.07E-4
NF-kappaB binding	GO:0051059	214	14	2	7.41E-4
chaperone-mediated protein folding	GO:0061077	64	7	1	8.53E-4
negative regulation of B cell apoptosis	GO:0002903	65	7	2	9.37E-4
photoreceptor activity	GO:0009881	324	18	6	9.85E-4

Table 3.9: GO-terms enriched in the number of studies in which their associated genes were found overexpressed. The total number of times genes were found associated with each GO-term, the numbers in which they were over- and underexpressed and the binomial p-value for the enrichment of overexpression are shown.

GO term	GO	total	overexp.	underep.	p_underexp
sterol biosynthetic process	GO:0016126	1091	29	59	5.57E-10
plasma membrane	GO:0005886	68511	1690	1722	6.14E-9
beta-alanine transmembrane transporter activity	GO:0001761	60	1	12	6.32E-9
beta-alanine transport	GO:0001762	60	1	12	6.32E-9
taurine transmembrane transporter activity	GO:0005368	60	1	12	6.32E-9
taurine:sodium symporter activity	GO:0005369	60	1	12	6.32E-9
taurine transport	GO:0015734	60	1	12	6.32E-9

taurine binding	GO:0030977	60	1	12	6.32E-9
cholesterol biosynthetic process	GO:0006695	1022	31	53	1.59E-8
innate immune response	GO:0045087	3356	85	125	1.80E-8
response to sterol depletion	GO:0006991	68	3	12	2.80E-8
steroid biosynthetic process	GO:0006694	2298	65	93	2.97E-8
extracellular region	GO:0005576	42731	1227	1102	3.84E-8
microsome	GO:0005792	10612	366	316	7.15E-8
7-dehydrocholesterol reductase activity	GO:0047598	49	0	10	9.43E-8
response to virus	GO:0009615	1706	46	73	1.06E-7
positive regulation of transcription via sterol regulatory element binding	GO:0035104	92	2	13	1.17E-7
pheromone binding	GO:0005550	164	9	17	1.52E-7
ISG15-protein conjugation	GO:0032020	132	2	15	2.42E-7
lipid biosynthetic process	GO:0008610	4030	121	139	2.50E-7
collagen fibril organization	GO:0030199	856	24	44	3.11E-7
regulation of heart rate by chemical signal	GO:0003062	60	1	10	6.95E-7
sterol response element binding	GO:0032810	60	1	10	6.95E-7
glucose 6-phosphate metabolic process	GO:0051156	250	12	20	8.76E-7
positive regulation of glycolysis	GO:0045821	148	3	15	1.07E-6
3-beta-hydroxy-delta5-steroid dehydrogenase activity	GO:0003854	283	7	21	1.59E-6
fatty acid biosynthetic process	GO:0006633	2216	72	84	1.74E-6
citrate metabolic process	GO:0006101	309	8	22	1.80E-6
cell cortex part	GO:0044448	41	1	8	2.63E-6
detection of glucose	GO:0051594	41	1	8	2.63E-6
endoplasmic reticulum	GO:0005783	30121	749	778	2.64E-6
antigen processing and presentation	GO:0019882	1004	10	46	3.77E-6
complement activation, classical pathway	GO:0006958	853	15	41	3.96E-6
cellular response to mycophenolic acid	GO:0071506	74	0	10	5.01E-6
negative regulation of steroid biosynthetic process	GO:0010894	110	5	12	5.79E-6
creatine metabolic process	GO:0006600	133	0	13	8.12E-6
creatinine metabolic process	GO:0046449	133	0	13	8.12E-6
positive regulation of cholesterol biosynthetic process	GO:0045542	197	5	16	8.60E-6
modification-dependent protein catabolic process	GO:0019941	97	1	11	9.66E-6
sugar binding	GO:0005529	5415	112	167	1.19E-5
iron ion binding	GO:0005506	4581	175	145	1.24E-5
circadian rhythm	GO:0007623	1025	72	45	1.37E-5
cholesterol metabolic process	GO:0008203	2037	69	75	1.56E-5
20-alpha-hydroxysteroid dehydrogenase activity	GO:0047006	68	1	9	1.76E-5
allantoin metabolic process	GO:0000255	144	1	13	1.92E-5
glucokinase activity	GO:0004340	86	1	10	1.95E-5
activation of signaling protein activity involved in unfolded protein response	GO:0006987	69	0	9	1.98E-5
FasL biosynthetic process	GO:0045210	39	1	7	1.99E-5
monooxygenase activity	GO:0004497	2803	147	96	2.02E-5
syndecan binding	GO:0045545	105	3	11	2.06E-5
immune response	GO:0006955	4261	110	135	2.29E-5
defense response to virus	GO:0051607	451	11	25	2.94E-5
extracellular space	GO:0005615	18369	548	484	3.11E-5

positive regulation of fatty acid biosynthetic process	GO:0045723	177	8	14	4.10E-5
positive regulation of triglyceride biosynthetic process	GO:0010867	201	3	15	4.35E-5
choline binding	GO:0033265	156	10	13	4.45E-5
glucose binding	GO:0005536	464	17	25	4.65E-5
cytokine receptor activity	GO:0004896	1148	30	47	4.94E-5
selenocysteine lyase activity	GO:0009000	31	1	6	5.03E-5
positive regulation of histone deacetylation	GO:0031065	204	5	15	5.15E-5
steroid delta-isomerase activity	GO:0004769	117	3	11	5.66E-5
regulation of transforming growth factor beta receptor signaling pathway	GO:0017015	232	4	16	6.35E-5
carbohydrate phosphorylation	GO:0046835	283	11	18	6.56E-5
collagen biosynthetic process	GO:0032964	63	4	8	6.94E-5
collagen	GO:0005581	626	16	30	7.52E-5
steroid metabolic process	GO:0008202	2076	89	73	8.26E-5
extracellular matrix	GO:0031012	3665	104	116	8.37E-5
growth hormone receptor activity	GO:0004903	65	0	8	8.71E-5
growth hormone receptor signaling pathway	GO:0060396	65	0	8	8.71E-5
cranial suture morphogenesis	GO:0060363	192	5	14	9.87E-5
isoleucine metabolic process	GO:0006549	195	2	14	1.16E-4
naphthalene metabolic process	GO:0018931	37	0	6	1.42E-4
trichloroethylene metabolic process	GO:0018979	37	0	6	1.42E-4
acetyl-CoA biosynthetic process	GO:0006085	175	9	13	1.43E-4
positive regulation of programmed cell death	GO:0043068	152	4	12	1.44E-4
C-5 sterol desaturase activity	GO:0000248	38	0	6	1.66E-4
cholesterol biosynthetic process via lathosterol	GO:0033490	38	0	6	1.66E-4
lathosterol oxidase activity	GO:0050046	38	0	6	1.66E-4
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	GO:0016712	534	33	26	1.70E-4
NADP biosynthetic process	GO:0006741	54	2	7	1.72E-4
integral to membrane	GO:0016021	133096	3156	3104	1.78E-4
protein disulfide isomerase activity	GO:0003756	335	3	19	1.87E-4
positive regulation of homocysteine metabolic process	GO:0050668	55	2	7	1.94E-4
proteinaceous extracellular matrix	GO:0005578	8619	238	239	1.94E-4
defense response to Gram-positive bacterium	GO:0050830	1018	24	41	1.99E-4
calcium ion transport	GO:0006816	3675	94	114	2.03E-4
regulation of angiogenesis	GO:0045765	423	8	22	2.10E-4
misfolded protein binding	GO:0051787	183	2	13	2.23E-4
cellular response to glucose starvation	GO:0042149	137	2	11	2.32E-4
membrane	GO:0016020	176738	4222	4083	2.35E-4
second-messenger-mediated signaling	GO:0019932	75	2	8	2.40E-4
endoplasmic reticulum lumen	GO:0005788	1061	25	42	2.40E-4
basement membrane	GO:0005604	2976	76	95	2.56E-4
NADPH oxidase complex	GO:0043020	188	1	13	2.89E-4
protein secretion	GO:0009306	239	2	15	2.95E-4
purinergic nucleotide receptor activity, G-protein coupled	GO:0045028	376	7	20	2.97E-4

heme binding	GO:0020037	3600	132	111	3.00E-4
collagen type III	GO:0005586	59	1	7	3.02E-4
aromatase activity	GO:0070330	711	41	31	3.02E-4
oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water	GO:0016717	240	13	15	3.08E-4
integral to plasma membrane	GO:0005887	10052	278	272	3.26E-4
catechol O-methyltransferase activity	GO:0016206	60	2	7	3.35E-4
phosphoinositide 3-kinase cascade	GO:0014065	79	3	8	3.43E-4
negative regulation of epinephrine secretion	GO:0032811	121	3	10	3.49E-4
nickel ion binding	GO:0016151	168	1	12	3.64E-4
epinephrine secretion	GO:0048242	62	2	7	4.11E-4
hexokinase activity	GO:0004396	171	3	12	4.27E-4
polyspecific organic cation transmembrane transporter activity	GO:0015354	82	0	8	4.43E-4
positive regulation of activated T cell proliferation	GO:0042104	303	3	17	4.47E-4
mRNA modification	GO:0016556	331	5	18	4.47E-4
response to ethanol	GO:0045471	1337	42	49	4.50E-4
mitotic cell cycle G2/M transition DNA damage checkpoint	GO:0007095	172	2	12	4.50E-4
organic cation transmembrane transporter activity	GO:0015101	125	1	10	4.53E-4
cellular response to interferon-alpha	GO:0035457	63	2	7	4.53E-4
fructose 2,6-bisphosphate metabolic process	GO:0006003	223	11	14	4.60E-4
JAK-STAT cascade	GO:0007259	795	12	33	4.67E-4
cell adhesion	GO:0007155	14943	360	388	4.73E-4
taurine metabolic process	GO:0019530	277	12	16	4.74E-4
negative regulation of cell-matrix adhesion	GO:0001953	149	10	11	4.77E-4
leukemia inhibitory factor receptor activity	GO:0004923	46	0	6	4.84E-4
establishment or maintenance of transmembrane electrochemical gradient	GO:0010248	46	0	6	4.84E-4
epinephrine transport	GO:0048241	46	0	6	4.84E-4
water channel activity	GO:0015250	363	23	19	5.04E-4
regulation of insulin secretion	GO:0050796	607	29	27	5.26E-4
proton-dependent oligopeptide secondary active transmembrane transporter activity	GO:0005427	47	0	6	5.45E-4
cerebellar Purkinje cell layer development	GO:0021680	281	2	16	5.53E-4
5-aminolevulinate synthase activity	GO:0003870	106	6	9	5.58E-4
protein import into nucleus, translocation	GO:0000060	310	14	17	5.78E-4
cholinesterase activity	GO:0004104	107	8	9	5.98E-4
substrate-bound cell migration	GO:0006929	66	0	7	6.03E-4
polysaccharide binding	GO:0030247	430	24	21	6.50E-4
positive regulation of natural killer cell proliferation	GO:0032819	33	0	5	7.13E-4
response to interleukin-15	GO:0070672	33	0	5	7.13E-4
left-handed Z-DNA binding	GO:0003692	68	0	7	7.23E-4
elevation of cytosolic calcium ion concentration	GO:0007204	2698	73	85	7.54E-4
osteoblast differentiation	GO:0001649	1264	26	46	7.55E-4
dopamine transport	GO:0015872	158	2	11	7.78E-4

dopamine transmembrane transporter activity	GO:0005329	111	2	9	7.80E-4
cytolysis	GO:0019835	721	24	30	7.88E-4
regulation of neuron differentiation	GO:0045664	529	10	24	7.91E-4
cyclin binding	GO:0030332	263	6	15	7.96E-4
chemokine activity	GO:0008009	1024	31	39	8.00E-4
negative regulation of female receptivity	GO:0007621	184	6	12	8.17E-4
female pregnancy	GO:0007565	531	14	24	8.33E-4
positive regulation of prostaglandin biosynthetic process	GO:0031394	266	5	15	8.92E-4
membrane attack complex	GO:0005579	137	3	10	9.25E-4
phosphatidylcholine biosynthetic process	GO:0006656	353	9	18	9.38E-4
regulation of natriuresis	GO:0003078	35	1	5	9.40E-4
V1B vasopressin receptor binding	GO:0031895	35	1	5	9.40E-4
multicellular organismal water homeostasis	GO:0050891	35	1	5	9.40E-4
acyl carrier activity	GO:0000036	114	2	9	9.44E-4
organic cation transport	GO:0015695	138	1	10	9.78E-4
blood vessel development	GO:0001568	2009	43	66	9.98E-4

Table 3.10: GO-terms enriched in the number of studies in which their associated genes were found underexpressed. The total number of times genes were found associated with each GO-term, the numbers in which they were over- and underexpressed and the binomial p-value for the enrichment of underexpression are shown.

Such a large number of significant GO-terms is difficult to interpret as to their role in CR. Therefore we focused on categories represented by similar GO-terms (at different levels of specificity) and GO-terms that were found with lowest p-values or were already known to be associated with CR. The possible use of these lists therefore exceeds what is described here by allowing to also investigate the relevance in respect to CR of all the other GO-terms not explicitly described here as.

The top GO-term for overexpressed genes with a highly significant p-value of $p < 10^{-23}$ is "lipid metabolic process". Also other, more specific GO-terms related to lipid metabolism like "acyl-CoA metabolic process" or "fatty acid metabolic process" were found. Some similar functional categories ("fatty acid metabolic process", "lipid metabolism", etc.) were also obtained in the DAVID analysis ("3.2.6.2 Putting genes found differentially expressed with CR into functional categories – DAVID-analysis") with low p-values, however not significant after Benjamini-Hochberg correction (p-values before / after correction: ~0.005 / ~0.3). Interestingly 3 of the 6 genes associated with "fatty acid metabolic process" in the DAVID-analysis were also associated with peroxisomes. These 6 genes are

- enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase,
- 2-hydroxyacyl-CoA lyase 1,
- acyl-CoA thioesterase
- carnitine palmitoyltransferase 1a, liver,
- acyl-CoA thioesterase 12 and
- peroxisome proliferator activated receptor alpha, knock-out of which was reported to protect mice from high-fat-diet induced insulin resistance (Cha et al. 2007).

Interestingly Hong (Hong, S. et al. 2010) found genes of the GO-category "lipid metabolism" enriched for downregulation with aging in a meta-analysis of microarray data on aging.

For underexpressed genes the top GO-term is "sterol biosynthetic process" with a p-value of $< 10^{-9}$. Also related to lipid synthesis "cholesterol biosynthetic process" and "lipid biosynthetic process" itself are among the top GO-terms. Interestingly also "response to sterol depletion" is detected, represented by insulin induced gene 1 Gene (Entrez ID 231070) among the significant genes. Also among the most significant GO-terms for upregulated genes are "rhythmic process" and "circadian rhythm", the second of which was also found for downregulated genes. A

GO term	GO	total	overexp.	underexp.	p_overexp.	p_underexp
monooxygenase activity	GO:0004497	2803	147	96	8.69E-18	2.02E-5
circadian rhythm	GO:0007623	1025	72	45	2.15E-15	1.37E-5
microsome	GO:0005792	10612	366	316	1.57E-11	7.15E-8
extracellular region	GO:0005576	42731	1227	1102	2.51E-10	3.84E-8
iron ion binding	GO:0005506	4581	175	145	3.74E-9	1.24E-5
steroid metabolic process	GO:0008202	2076	89	73	2.33E-7	8.26E-5
extracellular space	GO:0005615	18369	548	484	2.91E-7	3.11E-5
aromatase activity	GO:0070330	711	41	31	4.05E-7	3.02E-4
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	GO:0016712	534	33	26	1.18E-6	1.70E-4
heme binding	GO:0020037	3600	132	111	2.15E-6	3.00E-4
water channel activity	GO:0015250	363	23	19	3.06E-5	5.04E-4
polysaccharide binding	GO:0030247	430	24	21	1.49E-4	6.50E-4
regulation of insulin secretion	GO:0050796	607	29	27	4.54E-4	5.26E-4

Table 3.11: GO-terms enriched in the number of studies both in which their associated genes were found over- and underexpressed. The total number of times genes were found associated with each GO-term, the numbers in which they were over- and underexpressed and the binomial p-values for the enrichment of over- and undererexpression are shown.

link between circadian rhythm and both CR and aging has already been noticed in several instances (see e.g. (Froy & Miskin 2010)).

Several categories related to immune response were found for downregulated genes: "innate immune response", "antigen processing and presentation", "complement activation, classical pathway".

Even though the GO-term "xenobiotic metabolism" itself was not enriched among our candidate genes, enzyme activities related to this process were represented by monooxygenase activity (for up- and downregulated genes) and NADPH-hemoprotein reductase activity (up). Some of the genes found in categories related to oxidation and reduction fall into this category. Xenobiotic metabolism (see e.g. (Gourley & C. J. Kennedy 2009)) and in particular monooxygenases (Schmucker et al. 1991) have been previously associated with CR, even thought their exact role remains unclear.

"Positive regulation of collagen biosynthetic process" was among the enriched terms for over- and "collagen", "collagen type III", "collagen fibril organization" and "collagen biosynthetic process" for underexpressed genes. It has been shown previously that caloric restriction to a certain degree prevents collagen accumulation and collagen aging (see (Frey 2004)).

The findings of "growth hormone receptor activity" and "growth hormone receptor signaling pathway" for downregulated genes and "regulation of insulin secretion" for both up- and down-, as well as "insulin-like growth factor binding" for upregulated genes argues for involvement of the growth factor and insulin / IGF signalling pathways in CR.

"Retinol metabolism", which was found enriched for upregulated genes, has been linked to CR in a broader sense by a study reporting the decrease of retinol during fasting in humans (Söderlund et al. 2003).

Of the top 10 categories for overexpression to our knowledge no known link exists between CR and "copper ion detoxification". The GO-category "beta-alanine transmembrane transporter activity", found for downregulated genes, contains only 1 gene, *Slc6a6*. 5 other of the top 10 GO-categories for underexpressed genes were also found due to this single gene, found downregulated 12 of 60 times it was studied. To our knowledge this gene has not yet been associated with CR.

Out of the GO-terms shown there are 13 which meet the selection criteria for both over- and underexpressed genes. These are shown in table 3.11.

Since these terms are relatively broad it seems acceptable that their activities are changed by upregulation of some of their members and downregulation of others. Interestingly "steroid metabolic process" appears among those, while "steroid biosynthesis" is one of the top GO-terms for underexpressed genes and is only found at a

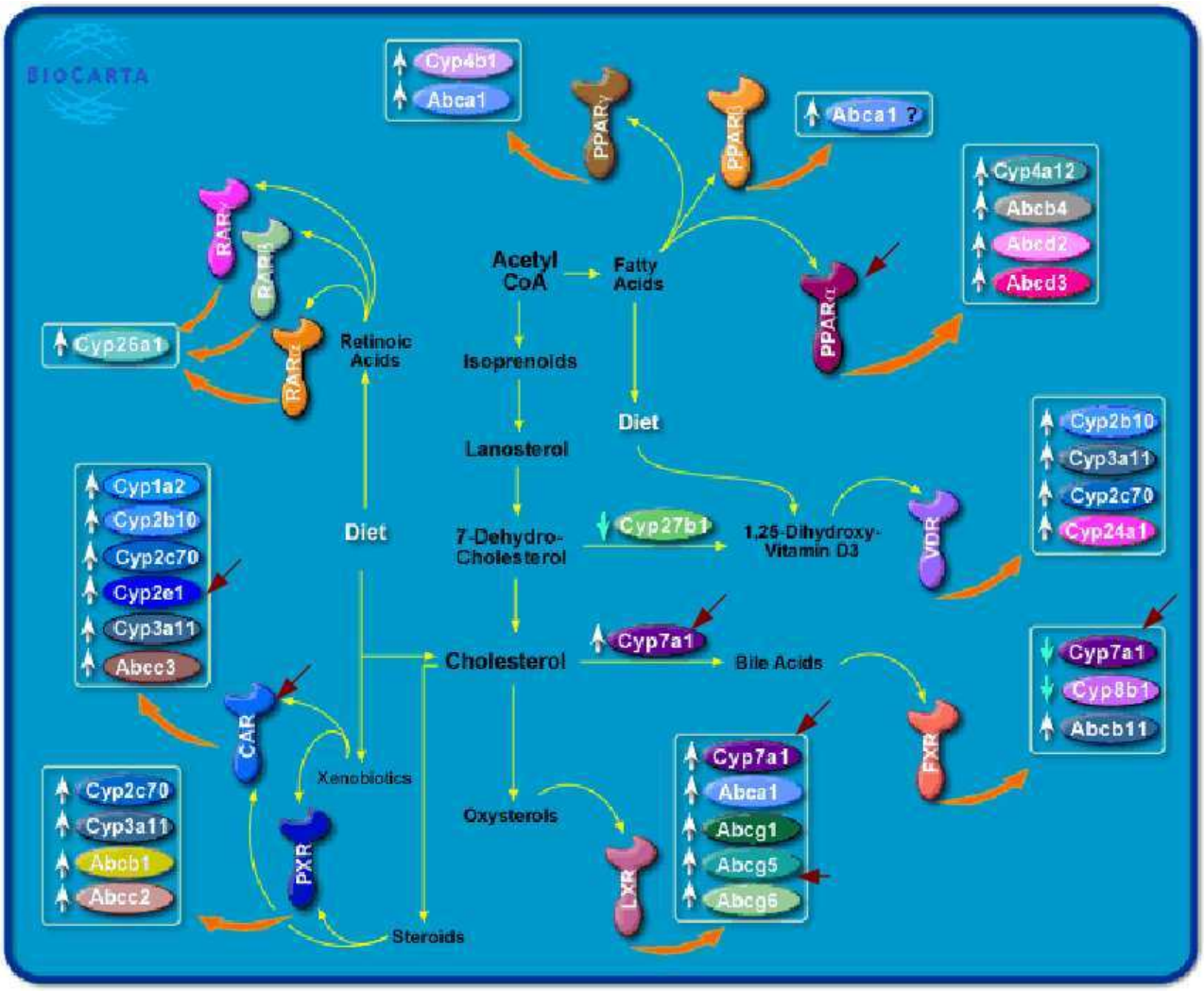
binomial p-value of 0.1 for overexpressed genes, i.e. much less emphasized. This suggests that while genes involved in steroid metabolism can be both up- or downregulated by CR, the ones responsible for the biosynthesis tend more towards downregulation.
Note that "steroid hormone receptor activity" and "response to steroid hormone stimulus" appear among the significant GO-terms for upregulated genes. This suggests that the alteration of steroid hormone levels and the effect of this alteration on cells is an important mechanism of CR.
Even though a single GO-category related to sterol / cholesterol metabolism is not found for both up- and downregulated genes, there are different such categories in both cases (e.g. "cholesterol 7-alpha-monooxygenase activity" and "regulation of cholesterol metabolic process" for over- and "sterol biosynthetic process" and "cholesterol biosynthetic process" for underexpressed genes).

3.3.2.2 Functional classification of genes enriched in the number of studies they are found over- / underexpressed - DAVID-analysis

We used the DAVID Functional Annotation tool to group genes enriched in studies in which they were found over- / underexpressed into functional categories. We obtained groups of such (often similar) categories clustered according to genes which they had in common (functional annotation clusters).
These clusters for the overexpressed genes containing at least one category with a Benjamini-Hochberg FDR below 0.05 contained categories related to sulfotransferase-activity, NAD(P) involving processes, oxidoreductases -of which a large fraction was also associated with endoplasmatic reticulum- and to biological rhythms. Even though not significant after multiple-testing correction the finding of the GO-term "response to nutrient levels" at a Benjamini-Hochberg corrected FDR of 0.16 acts as a prove of concept for successfully detecting functional categories determined by feeding levels. This term was represented by the genes: ATP-binding cassette, subfamily G (WHITE), member 5 (Entrez ID: 27409), alcohol dehydrogenase 1 (class I)(11522), angiopoietin-like 4 (57875), matrix Gla protein (17313), peroxisome proliferator activated receptor alpha (19013) and solute carrier family 37 (glucose-6-phosphate transporter), member 4 (14385).
The only functional annotation cluster with categories below a Benjamini-Hochberg FDR of 0.05 for underexpressed genes was related to endoplasmic reticulum.
A problem about the DAVID procedure under default options seems to be that so many hypotheses are tested that extremely low p-values are necessary for categories to remain significant after Benjamini-Hochberg correction. The number of significant functional categories was much lower than that found in the GO-analysis.
One significant Biocarta and 3 KEGG (Kanehisa et al. 2010) pathways were found below a Benjamini-Hochberg FDR of 0.05 for genes enriched for overexpression, none for those enriched with underexpression. (The analysis only for Biocarta and KEGG pathways tests less hypotheses as for all default categories and allows therefore pathways to be significant that were not, when testing more hypotheses). The Biocarta pathway "Nuclear Receptors in Lipid Metabolism and Toxicity" is shown in fig. 3.9, the illustrations of the KEGG pathways "PPAR signaling pathway", "Arachidonic acid metabolism" and "Retinol metabolism in animals" can be found in supplement 2.

3.3.2.3 Overlap between GO-analysis on original data and DAVID functional analysis on result genes

There is strong overlap between the functional categories found using DAVID on the genes found in the meta-analysis and meta-analysing GO-terms themselves. For example the significant DAVID functional clusters related to sulfotransferase-activity, NAD(P) involving processes, oxidoreductases and biological rhythms are represented by some of the most highly significant GO-terms, e.g. "tyrosine-ester sulfontransferase activity", "NADPH-hemprotein reductase activity", "oxidoreductase activity", "rhythmic process", "circadian rhythm" and others.
"Endoplasmic reticulum" which is found in the DAVID analysis for underexpressed genes is also found significant for the GO-analysis, even though not among the very top genes. "Sterol metabolism" is found among the top GO-terms and also among the top DAVID categories, even when not significant after Benjamini-Hochberg correction.
Note that a profound difference between meta-analysis on the level of GO-terms and DAVID-analysis on the significant results of meta-analysis on gene level is that a single gene found in many datasets can lead to significance of its GO-terms, while a GO-term has to be associated with different significant genes to be significant in the DAVID-analysis. GO-analysis is in theory able to detect functional categories associated with CR, even though no single gene of the category is itself significantly enriched for over- or underexpression. A strong overlap



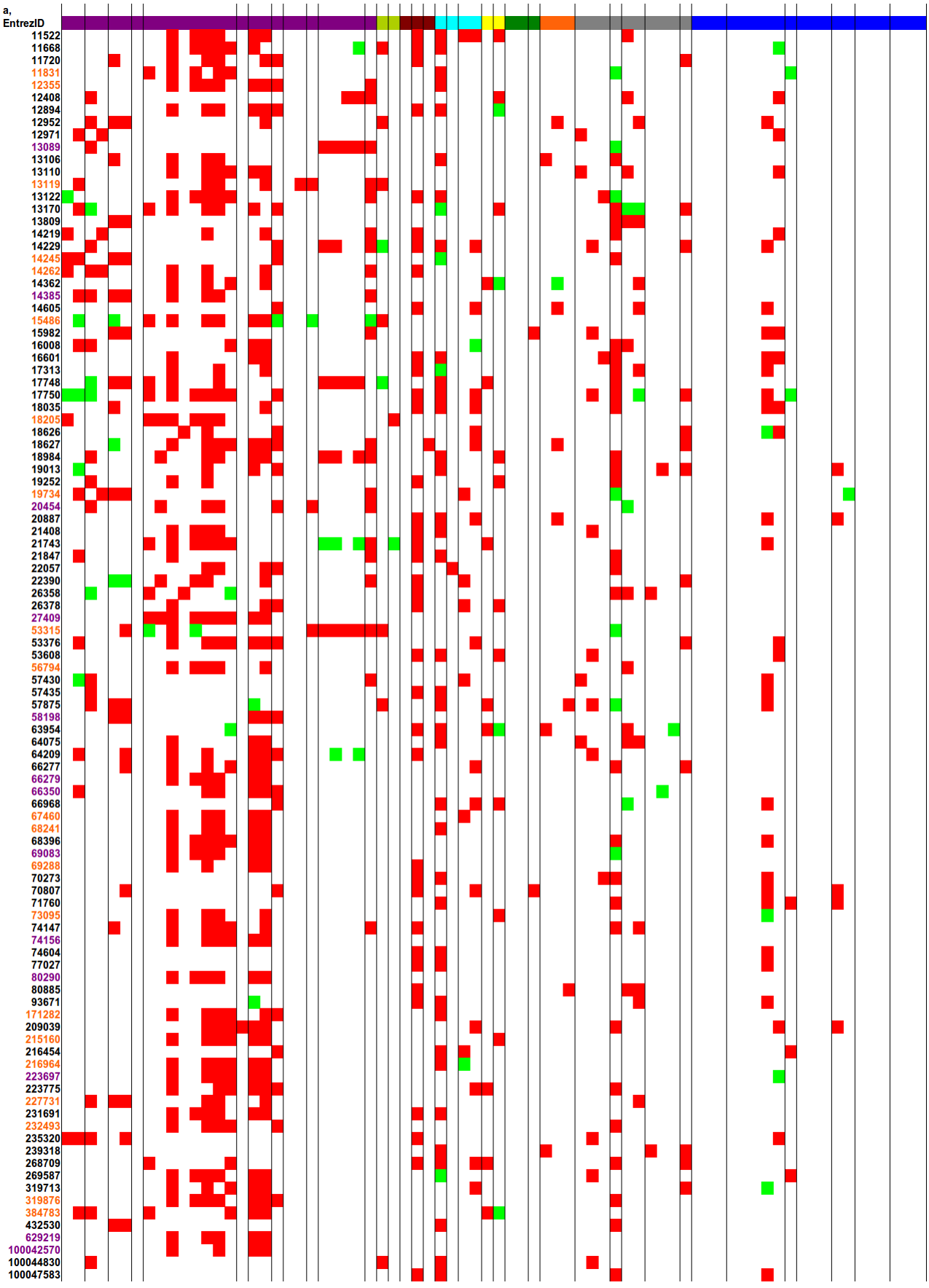
relevant legend:

Figure 3.9: Biocarta pathway "Nuclear Receptors in Lipid Metabolism and Toxicity", found associated with genes enriched for overexpression by the DAVID functional analysis tool. Genes enriched for overexpression are indicated by red arrows. For further information see http://www.biocarta.com/genes/index.asp.

between the GO-terms and the DAVID categories however implies a strong overlap of the GO-terms with genes found significantly enriched, since DAVID is based on these genes.

3.3.3 Tissues contributing to enrichment of a gene for over- or underexpression

As described in "3.2.7 Determining tissues contributing to enrichment of genes for over- or underexpression", we determined if the enrichment of a gene for over- / underexpression was due to its over- / underexpression in one, two or more than two tissues. Complete matrices showing the tissue specific differential-expression profiles of these genes are shown in Fig. 3.10.
It can be seen, that different datasets contribute to a different extent to the number of genes found enriched for over- / underexpression, especially liver-datasets (particularly GSE18297) can be found to contribute more and brain-tissue datasets less strongly. This is surprising in the sense that the brain-datasets contributing least are from GSE8426, a study among the highest in terms of the number of replicates.
13% and 16% of genes enriched for over- and underexpression respectively were found over- or underexpressed only in liver and 34% and 49% in less than three tissues (and mainly in liver and one other tissue). Since liver-specific signatures might mask tissue-independent ones we performed functional analysis (using DAVID) besides for the complete list of significant genes also for the list subtracted of genes over- / underexpressed in less than 3 tissues. Looking for liver-specific signatures we did the analysis for genes only over- / underexpressed in liver. The procedure is described in "4.7.1. Putting genes found differentially expressed with CR into functional categories".


a,
EntrezID
11522
11668
11720
11831
12355
12408
12894
12952
12971
13089
13106
13110
13119
13122
13170
13809
14219
14229
14245
14262
14362
14385
14605
15486
15982
16008
16601
17313
17748
17750
18035
18205
18626
18627
18984
19013
19252
19734
20454
20887
21408
21743
21847
22057
22390
26358
26378
27409
53315
53376
53608
56794
57430
57435
57875
58198
63954
64075
64209
66277
66279
66350
66968
67460
68241
68396
69083
69288
70273
70807
71760
73095
74147
74156
74604
77027
80290
80885
93671
171282
209039
215160
216454
216964
223697
223775
227731
231691
232493
235320
239318
268709
269587
319713
319876
384783
432530
629219
100042570
100044830
100047583

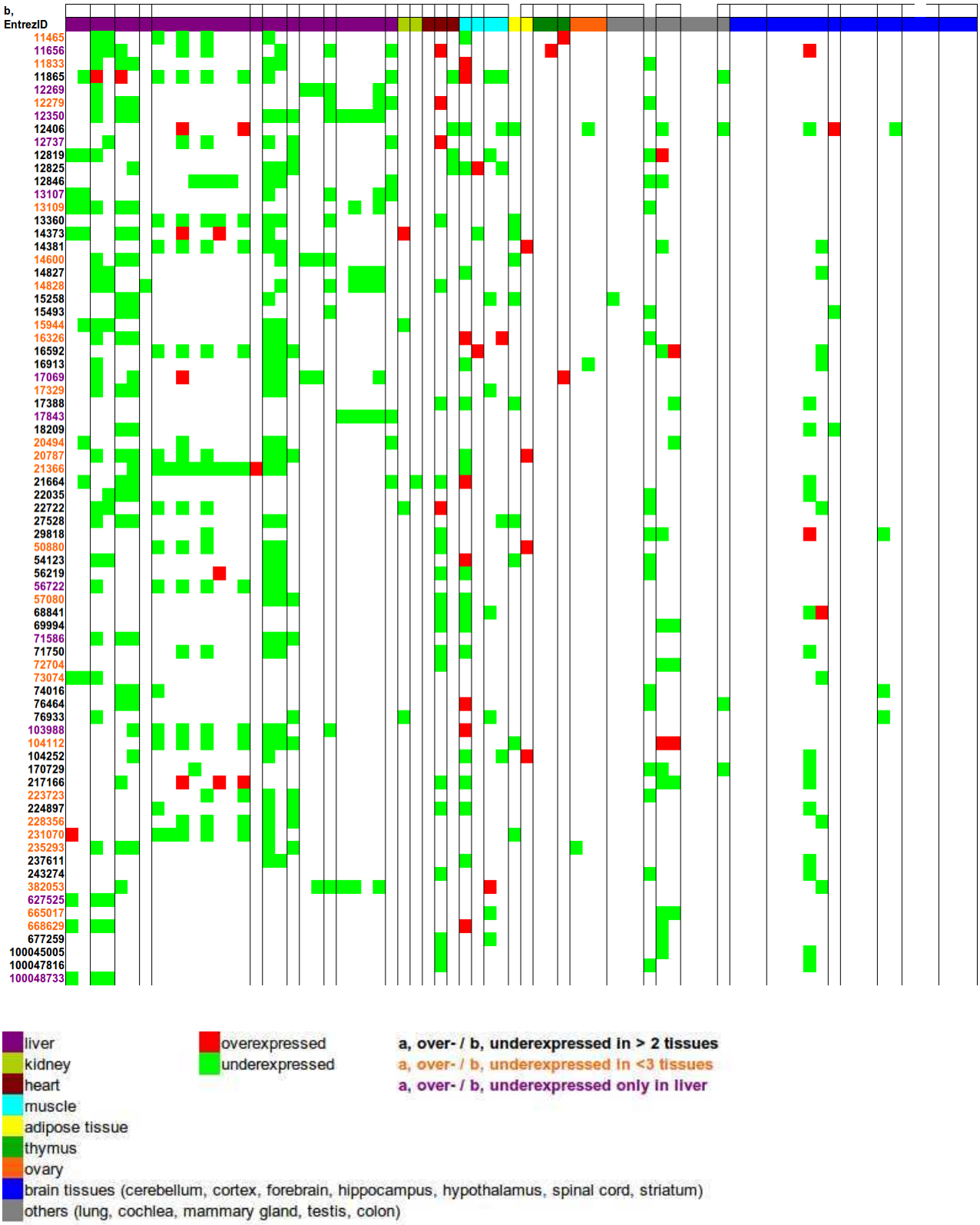

b,
EntrezID
11465
11656
11833
11865
12269
12279
12350
12406
12737
12819
12825
12846
13107
13109
13360
14373
14381
14600
14827
14828
15258
15493
15944
16326
16592
16913
17069
17329
17388
17843
18209
20494
20787
21366
21664
22035
22722
27528
29818
50880
54123
56219
56722
57080
68841
69994
71586
71750
72704
73074
74016
76464
76933
103988
104112
104252
170729
217166
223723
224897
228356
231070
235293
237611
243274
382053
627525
665017
668629
677259
100045005
100047816
100048733
liver
kidney
heart
muscle
adipose tissue
thymus
ovary
brain tissues (cerebellum, cortex, forebrain, hippocampus, hypothalamus, spinal cord, striatum)
others (lung, cochlea, mammary gland, testis, colon)
overexpressed
underexpressed
a, over- / b, underexpressed in > 2 tissues
a, over- / b, underexpressed in <3 tissues
a, over- / b, underexpressed only in liver

Fig. 3.10: Matrix view of genes (rows) vs. datasets (columns) for genes enriched for a, overexpression and b, underexpression. Red fields indicate over-, green underexpression ($p < 0.05$, effect size > 1.5-fold). See legend for colour coding of different tissues in the top row. For brain tissues and "others" vertical lines separate datasets from different tissues, for all other tissues datasets from different studies. Font colors in the first column: black: the gene is (a,) over- / (b,) underexpressed in at least 3 different tissues; orange: the gene is (a,) over- / (b,) underexpressed in less than 3 different tissues; purple: the gene is only differentially expressed in tissues liver.

3.3.4 Results of the analysis of non-liver and liver-only datasets

To determine genes differentially expressed on the one hand in a tissue tissue-independent manner, on the other hand liver specifically we repeated the DAVID-analysis first on genes found over- / underexpressed in at least 3 different tissues then on such found over- / underexpressed in liver-datasets only. For the second no categories were found at a Benjamini-Hochberg FDR < 0.05. Categories determined for genes enriched for over- / underexpression in at least 3 different tissues compared to these found for candidates resulting from the all-tissue meta-analysis are shown in tab. 3.12.

While functional categories related to sulfotransferase, vesicular transport, retinol and arachidonic acid metabolism were enriched for overexpressed genes and to endoplasmatic reticulum for underexpressed genes, these were not found enriched among the genes found over- / underexpressed in at least 3 different tissues. This does however not necessarily mean that these categories cannot be associated with CR cross-tissues, but might mean that by restricting to genes over- / underexpressed in > 2 different tissues the statistical power is simply too reduced to detect this association.

On the other hand this analysis showed that categories related to "NADP" and "circadian rhythm" were also found for only genes differentially expressed in at least 3 different tissues and can therefore be assumed to be truly tissue-independent. Interestingly two categories, "metal binding" and "vesicular transport" that were not significantly enriched among all genes were found significant for genes differentially expressed in at least 3 different tissues.

Note that some categories (like lipid metabolism) detected by the GO-analysis were not found by the DAVID-analysis on all-tissue candidates and it is therefore not possible to draw conclusions about their tissue-specificity or tissue-independence by this method.

3.3.5 Co-expression analysis of CR-associated genes

Genes enriched in the co-expression with genes overrepresented for up- / downregulation are given in supplement. 2. Since a large number of genes (1576 for over- and 1069 for underexpression) were found, we performed DAVID-analysis under default settings on them.

Interestingly we found that the functional categories obtained for upregulated genes were en large the same as for downregulated genes. Some of the most significant functional categories retrieved for both up- and downregulated genes were related to extracellular space, lipid metabolism, amino acid catabolism, inflammation / immunity, peroxisomes, steroid / sterol / cholesterol metabolism, endopeptidase inhibitor activity, lipoprotein particles, response to hormons, mitochondria, xenobiotics metabolism / cytochrome P450, blood coagulation.

Therefore, after we had already detected some functional categories that appeared associated both with genes overrepresented for over- and underexpression, we found this overlap even more pronounced on the level of their interaction partners. This might also have to do with the increased statistical power due to the large number of genes in this test. It suggests that pathways important for the effect of CR are upregulated in some and downregulated in other genes.

3.3.6 Transcription factors regulating expression of candidate genes

Transcription factor (TF) binding sites enriched for our candidate genes were searched using WebMotifs which acts as an interface to different TF-binding site detection softwares. The only one that obtained significant results was THEME which uses reported transcription factor binding sites and optimizes them to fit best fit to our data. The optimized sequences found significantly enriched for overexpressed genes were derived from binding sites for CBFB_NFYA (CCAAT-binding transcription factor subunit B), CUT and PBC domains (Fig. 3.10). According

	all genes	diff. exp. in >2 diff. tissues
sulfotransferase	+	
endoplasmatic reticulum	-	
circadian rhythm	+	+
xenobiotic metabolism / oxidoreductase activity	+	
arachidonic acid metabolism	+	
retinol	+	
metal binding		+
NADP	+	+
vesicular fraction	+	
vesicular traffic		+

Table 3.12: This table shows under which conditions certain functional categories are enriched for genes overexpressed (+) or underexpressed (-) according to DAVID-analysis on all candidate genes from the meta-analysis and on such over- / underexpressed in more than two different tissues.

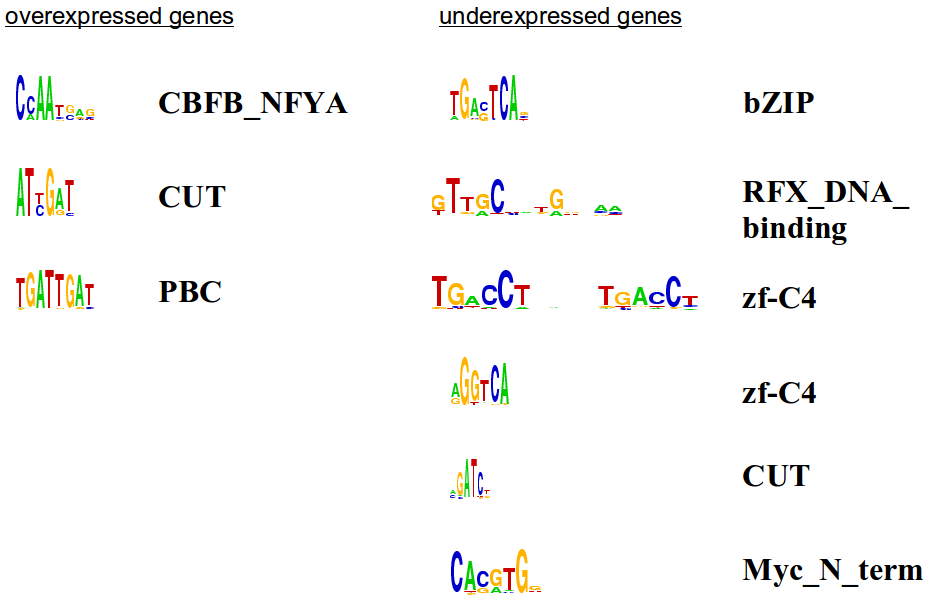


Figure 3.10: WebLogos (Crooks et al. 2004) of binding sites and corresponding TF-domains / domain families enriched in our candidate over- and underexpressed genes.

to PFAM CBFB_NFYA binds to a CCAAT motif in the promoters of a wide variety of genes, including type I collagen (pfam.sanger.co.uk).
For underexpressed genes we detected binding sites for bZIP (Basic Leucine Zipper), RFX (Regulatory Factor binding to X box), zf-C4 (Zinc finger, C4 type/Nuclear Hormone Receptor; for which two optimized sequences were found), CUT and Myc_N-term (Myc animo-terminal region) (Fig. 3.10). Myc forms a heterodimer with Max, and this complex regulates cell growth through direct activation of genes involved in cell replication. An especially interesting candidate domain is zf-C4 since it appears in steroid hormone receptors (according to PFAM). It therefore fits well with our functional analysis in which steroid metabolism and regulation by steroid hormones were recurrent topics.

3.3.7 Overlap with CR-essential genes, their orthologues and interaction partners

The only mouse gene in the database for genes experimentally identified to be essential for CR, GeneDR, is *Ghr* (Growth hormone receptor; Entrez ID: 14600). It was shown that mutating this gene cancels out the life-span extension effect of CR (Coschigano et al. 2003) (Bonkowski et al. 2006). In our meta-analysis this genes was enriched for underexpression, which is both a convincing argument for the biological meaningfulness of our results

and for the implication of *Ghr* in the mechanism of CR.
Further 4 of our candidates have CR-essential gene orthologues in lower model orgainsms: Of the genes enriched for overexpression these were *Irs2* (insulin-receptor substrate 2; an ortholog of *chico* in *Drosophila melanogaster*) and *Mat1α* (methionine adenosyltransferase I, alpha; the ortholog of *sams-1* in *Caenorhabditis elegans*) and for those enriched for downregulation *Gck* (Glucokinase) and *Sc5d* (sterol-C5-desaturase) which are orthologues of *HXK2* and *ERG3* in *S. cerevisiae*, respectively (Clancy et al. 2002) (Hansen et al. 2005) (Lin, S. J. et al. 2000) (Tang et al. 2008). Note than the detection of genes associated with CR in these organisms in a meta-analysis of mammalian datasets suggests at least some degree of conservation in the mechanism of CR from yeast to mammals.
Additional 42 genes were direct interaction partners of murine CR-essential gene orthologues as determined by the procedure described in "3.2.11 Detecting overlap with CR-essential genes, their orthologues and interaction partners". The complete list of these genes with their specificity measure and p-value is shown in table 3.13.

Moreover, 3 of these 47 genes were also implicated in aging according to the GenAge database (de Magalhães & Toussaint 2004): *Ghr, Irs2* and *Arntl* (aryl hydrocarbon receptor nuclear translocator-like Gene), an important circadian clock transcription factor (Coschigano et al. 2000) (Kondratov et al. 2006) (Taguchi et al. 2007).

3.3.8 Association of individual datasets to the meta-signature of CR

The p-values obtained in the chi-square test assessing the association between each dataset and the meta-signature of CR as described in "3.2.12 Testing the association of individual datasets to the meta-signature of CR" are given in table 3.14. The test was not done for datasets obtained form literature and supplements, since they only provide differentially expressed genes.

It can be seen, that many datasets show a strong association with the meta-signature. This is especially true for liver datasets, while for many of the brain-tissue datasets no gene in the meta-signature was found differentially expressed. However the correlation between the strength of the association and the study from which the datasets came from seems relatively strong. Therefore, for tissues that only contain datasets from one or a few studies (e.g. most brain tissues are from GSE8426) it is hard to conclude if they are especially well / weakly represented by the meta-signature or if the corresponding study (studies) show strong / weak association(s) for other reasons. Because liver was tested by many individual studies and for most low p-values in the chi-square test were obtained, it appears save to conclude that at least the effect of CR on liver is well represented by our meta-signature.

3.4 Discussion

3.4.1 Summary and interpretation

CR is the most promising non-genetic intervention to extend life-span and delay aging associated diseases in a range of organisms. To understand the genetic basis of CR we aimed at determining robust changes in gene expression linked to CR by meta-analysing microarray data on CR with wide variation in different experimental variables. To on the one hand find genes differentially expressed under different conditions, but on the other hand to also allow transcription levels not to be affected or to be affected in opposite direction under a few circumstances we chose a value-counting approach. To account for the fact that different genes were tested in a different number of datasets we chose a binomial test.

As microarray analyses themselves also this meta-analysis of microarray data in the first place provides a source of candidate genes and functional categories that may be implicated in the CR-process. The found genes and categories can be broadly divided into such providing further evidence for genes and functions already associated with CR and such not yet tested for their role in CR. Genes and categories for which we are aware of their relation to CR will be discussed in the following as will the most outstanding novel ones. For all others we refer you to the complete lists as provided in tables 3.6, 3.7, 3.9, 3.10.

It is interesting to note that considering all experiments less genes were found under- than overexpressed. Even though this decreases the success probability (p_s) in the binomial test (eqation3.1)[7], also less genes / GOs were found enriched in studies in which they / their associated genes were found over- than underexpressed. This result is somewhat expected if you assume that CR induces a transcriptional response, e.g. to more strongly pronounce alternative metabolic pathways.

[7] A lower p_S requires a lower number of hits (k) for the same number of trials (n) to give the same binomial p-value

Entrez ID	Gene Symbol	MGI Description	specificity (%)	specificity p-value	comment
11833	Aqp8	aquaporin 8 Gene	41	1.48E-06	
11831	Aqp6	aquaporin 6 Gene	37	3.58E-06	
232493	Gys2	glycogen synthase 2 Gene	17	5.61E-05	
384783	Irs2	insulin receptor substrate 2 Gene	13	8.07E-05	CR-associated ortholog, aging-associated
15982	Ifrd1	interferon-related developmental regulator 1 Gene	27	0.01	
14381	G6pdx	glucose-6-phosphate dehydrogenase X-linked Gene	13	0.03	
58198	Sall1	sal-like 1 (Drosophila) Gene	25	0.03	
29818	Hspb7	heat shock protein family, member 7 (cardiovascular) Gene	100	0.04	
11668	Aldh1a1	aldehyde dehydrogenase family 1, subfamily A1 Gene	8	0.05	
12846	Comt1	catechol-O-methyltransferase 1 Gene	15	0.08	
11865	Arntl	aryl hydrocarbon receptor nuclear translocator-like Gene	11	0.09	aging-associated
22390	Wee1	WEE 1 homolog 1 (S. pombe) Gene	7	0.12	
70807	Arrdc2	arrestin domain containing 2 Gene	25	0.14	
15258	Hipk2	homeodomain interacting protein kinase 2 Gene	9	0.14	
26358	Aldh1a7	aldehyde dehydrogenase family 1, subfamily A7 Gene	7	0.15	
18035	Nfkbia	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha Gene	6	0.15	
57080	Gtf2ird1	general transcription factor II I repeat domain-containing 1 Gene	14	0.24	
67460	Decr1	2,4-dienoyl CoA reductase 1, mitochondrial Gene	13	0.26	
235293	Sc5d	sterol-C5-desaturase (fungal ERG3, delta-5-desaturase) homolog (S. cerevisae) Gene	6	0.27	CR-associated ortholog
100042570	Bnip3	BCL2/adenovirus E1B interacting protein 3 Gene	11	0.29	
235320	Zbtb16	zinc finger and BTB domain containing 16 Gene	5	0.31	
269587	Epb4.1	erythrocyte protein band 4.1 Gene	5	0.33	
223697	Sun2	Sad1 and UNC84 domain containing 2 Gene	5	0.34	
14600	Ghr	growth hormone receptor Gene	6	0.39	CR-associated, aging-associated
14828	Hspa5	heat shock protein 5 Gene	4	0.40	
103988	Gck	glucokinase Gene	5	0.41	CR-associated ortholog
12406	Serpinh1	serine (or cysteine) peptidase inhibitor, clade H, member 1 Gene	7	0.44	
14229	Fkbp5	FK506 binding protein 5 Gene	4	0.47	
13170	Dbp	D site albumin promoter binding protein Gene	6	0.48	
19013	Ppara	peroxisome proliferator activated receptor alpha Gene	5	0.50	
11465	Actg1	actin, gamma, cytoplasmic 1 Gene	4	0.55	
215160	Rhbdd2	rhomboid domain containing 2 Gene	4	0.60	
18627	Per2	period homolog 2 (Drosophila) Gene	4	0.62	
18626	Per1	period homolog 1 (Drosophila) Gene	4	0.66	
14827	Pdia3	protein disulfide isomerase associated 3 Gene	3	0.67	
20787	Srebf1	sterol regulatory element binding transcription factor 1 Gene	3	0.78	
104112	Acly	ATP citrate lyase Gene	3	0.78	
668629	Ptprj	protein tyrosine phosphatase, receptor type, J Gene	2	0.81	
54123	Irf7	interferon regulatory factor 7 Gene	2	0.85	
13360	Dhcr7	7-dehydrocholesterol reductase Gene	2	0.87	
15493	Hsd3b2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2 Gene	2	0.88	
71586	Ifih1	interferon induced with helicase C domain 1 Gene	2	0.88	
69288	Rhobtb1	Rho-related BTB domain containing 1 Gene	1	0.98	
11720	Mat1a	methionine adenosyltransferase I, alpha Gene	1	0.98	CR-associated ortholog
13809	Enpep	glutamyl aminopeptidase Gene	1	1.00	
80885	Niacr1	niacin receptor 1 Gene	1	1.00	
73074	Cxcl9	RIKEN cDNA 2900086B20 gene	1	1.00	

Table 3.13: Genes found in the meta-analysis that are interaction partners of genes experimentally associated with CR. See text ("3.2.11 Detecting overlap with CR-essential genes, their orthologues and interaction partners") for definition of specificity and specificity p-value. Analysis by D. Wuttke.

tissue	dataset	chi-square p-value
liver	GDS1261Amesdwarf.txt	1.7E-13
	GDS1261normal.txt	1.2E-15
	GDS1808LTCR.txt	1.8E-35
	GDS1808CR8.txt	3.0E-17
	GSE11244true.ad.libitum.txt	1.3E-25
	GSE11244fixed.high.cal.txt	3.7E-31
	GSE17309.txt	0.52
	GSE18297_1week_5perc.txt	2.5E-31
	GSE18297_1month_5perc.txt	0.01
	GSE18297_1week_20perc.txt	3.1E-78
	GSE18297_1month_30perc.txt	0.01
	GSE18297_1week_30perc.txt	1.5E-67
	GSE18297_1month_10perc.txt	3.3E-97
	GSE18297_1month_20perc.txt	2.2E-81
	GSE18297_1week_10perc.txt	8.5E-47
	oneReplicate_GSE904.txt	0.13
	GSE9121_liv4.txt	1.2E-59
	GSE9121_liv10.txt	3.1E-85
	EMEXP-748_liver.csv	3.1E-36
	prep_Cao_7months.csv	suppl.
	prep_Cao_27months.csv	suppl.
	prep_Corton2004.csv	suppl.
	prep_Dhabi_temporal2weeks.csv	suppl.
	prep_Dhabi_temporal4weeks.csv	suppl.
	prep_Dhabi_temporal8weeks.csv	suppl.
	prep_Dhabi_temporal27mo.csv	suppl.
	prep_Fu_liver.csv	suppl.
kidney	GDS355_356.txt	9.5E-04
	GSE6110.txt	0.51
heart	GSE6718heart.txt	0 in MS
	GSE11291heart.txt	1.7E-20
	prep_Fu_heart.csv	suppl.
skeletal muscle	GSE11291gastrocnemius.txt	2.2E-04
	prep_Kayo 2001.csv	suppl.
	GDS2612.txt	0.03
	EMEXP-748_muscle.txt	4.2E-23

tissue	dataset	chi-square p-value
adipose tissue	GSE9121_adip.txt	3.9E-27
	GSE6718wat.txt	0.28
thymus	GDS2961_2962_16months.txt	0 in MS
	GDS2961_2962_6months.txt	0 in MS
	GDS2961_2962_24months.txt	0.80
ovary	GSE7502ovary16months.txt	1.2E-03
	GSE7502ovary6months.txt	2.2E-04
	GSE7502ovary24months.txt	0.32
lung	GDS241_2h.txt	0.76
	GDS241_12h.txt	0.00
	GDS241_4h.txt	0.57
cochlea	GDS2681.txt	0.04
mammary gland	GSE14202exercise.txt	8.3E-14
	GSE14202no_exercise.txt	1.4E-9
testis	GSE7502testis16months.txt	0.04
	GSE7502testis24months.txt	0.82
	GSE7502testis6months.txt	0 in MS
colon	EMEXP-748_colon.csv	6.3E-30
cerebellum	GSE8426Cerebellum24months.txt	0 in MS
	GSE8426Cerebellum6months.txt	0 in MS
	GSE8426Cerebellum16months.txt	0 in MS
cortex	GSE8426Cortex6months.txt	0 in MS
	GSE8426Cortex24months.txt	0 in MS
	GSE8426Cortex16months.txt	0 in MS
	GSE11291neocortex.txt	2.3E-15
	GAN_Expr_Profile_Aging_CR_Retardation_Neocortex_30months.txt	0.22
forebrain	oneReplicate_CR PS RAW DATA.txt	0.26
hippocampus	GSE8426Hippocampus6months.txt	0 in MS
	GSE8426Hippocampus16months.txt	0 in MS
	GSE8426Hippocampus24months.txt	0 in MS
hypothalamus	EMEXP-748_hypothalamus.csv	7.2E-7
	prep_Fu_hypothalamus.csv	suppl.
spinal cord	GSE8426Spinal.cord6months.txt	0 in MS
	GSE8426Spinal.cord16months.txt	0 in MS
	GSE8426Spinal.cord24months.txt	0 in MS
striatum	GSE8426Striatum24months.txt	0 in MS
	GSE8426Striatum16months.txt	0 in MS
	GSE8426Striatum6months.txt	0 in MS

Table 3.14: Association of individual datasets with the meta-signature of CR. Datasets are sorted according to tissue; datasets of different studies are separated from those of another study within tissue entries; "0 in MS": none of the genes in the meta-signature was found differentially expressed in this dataset; "suppl.": dataset from literature or supplement.

We also detected that many of the top functional categories appear enriched for both over- and underexpressed genes, e.g. categories related to lipid, steroid, sterol / cholesterol metabolism, circadian clock and xenobiotic metabolism. We expect that especially in these rather broad categories overexpression of some and underexpression of other genes might lead to the same outcome (e.g. if an activator of a gene is up- and a suppressor downregulated this both leads to activation of the gene). Of course this assumption has to be further validated by closer examination of the underlying signalling networks.

The appearance of many lipid metabolism and sterol biosynthesis related GO-terms among the ones of highest significance fits well with the idea of different metabolic states of AL and CR animals. It is not at all surprising that lipid metabolism and related categories emerge as results since it is expected that animals with significantly reduced caloric intake rather catabolize than anabolize fat. Besides the intuitive understanding that caloric restriction alters lipid metabolism there is plenty of literature linking lipid metabolism with possible mechanisms of CR. For an overview see e.g. (Puca et al. 2008). It has also been reported that CR prevents age related changes in cholesterol metabolism (Martini et al. 2008). *Sc5d* (sterol-C5-desaturase) was one of the candidates for downregulation involved in sterol metabolism and is a homologue of *ERG3*, which is important for life-span extension by CR in *S. cerevisiae*. Also finding the endoplasmic reticulum as a category significant for both over- and underexpressed genes is in agreement with this idea since this is an important compartment for lipid synthesis (Hong, S. et al. 2010).

Our functional analysis detected categories related to the growth hormone and insulin / IGF-signalling pathways, mutations in which have effects on longevity and the life-span extending effect of CR. *Ghr* (growth hormon receptor) is the only known mouse gene that cancels out the life-span extending effect of CR upon mutation (Bonkowski et al. 2006). This gene was enriched for underexpression in our analysis. *Irs2* (insulin-receptor substrate 2) was found for overexpression and is an ortholog of *chico* in *Drosophila melanogaster*, which was experimentally associated with aging and CR. In this respect one of our most interesting candidates enriched for underexpression is *Airn* (antisense Igf2r RNA), which might be a ncRNA with an important role in the regulation of insulin / IGF-signalling. Note that this gene until recently was annotated as a RIKEN cDNA gene and that therefore others of our candidate genes with unknown function might also promise interesting roles in the CR mechanism. In general the role of ncRNAs in the context of CR is widely unknown.

We determined categories related to circadian rhythm and xenobiotic metabolism both for over- and underexpressed genes, which had both already been associated with CR (Froy & Miskin 2010) (Gourley & Kennedy 2009) (Schmucker et al. 1991), however for which deeper understanding of their role in CR remains elusive. Two of our candidate genes, *Arntl* (aryl hydrocarbon receptor nuclear translocator-like Gene) and *Dbp* (albumin D site-binding protein), are important circadian clock transcription factors of which the first was already associated with the aging process, while *Dbp* has not yet received much attention with respect to aging or CR.

One of the major side effects of CR is the repression of immune functions and an important physiological change with aging is increased inflammation and alterations in collagen deposition. Therefore it is noteworthy that our meta-analysis also established relations between CR and these functional categories.

A process less well established as to its role in CR is retinol metabolism and to our knowledge no reports on copper ion detoxification exist in respect to CR. Still both processes were found among the most significantly enriched for genes overexpressed with CR. Especially since many of the functional categories detected are meaningful in the light of existing knowledge we also believe in the relevance of these terms.

Note that even though not found in the context of an enriched functional category *Nfkbia,* which was found enriched for overexpression is such a central molecule in NfkB-signalling, that it might by itself render this pathway important for the mechanism of CR. *Zfp64* as a little understood co-activator in the notch pathway also has the potential to be an interesting candidate concerning the mechanism of CR.

When extending the number of genes by obtaining genes significantly co-expressed with the determined candidates and therefore increasing the power of the approaches determining underlying functional categories, we noted that basically all these categories were found for both over- and underexpressed genes. Some of the additional categories found this way were "mitochondria" and "peroxisomes" as subcellular locations, "response to hormones" and others. "Xenobiotic metabolism" was found explicitly as a GO-term as well as categories related to *P450*.

Due to the overrepresentation of liver-datasets in our analysis we cannot claim that all genes found in the meta-analysis over all tissues are associated with CR in a tissue-independent manner. However is seems save to assume that out of these genes those found over- / underexpressed in at least three different tissues are truly tissue-independent. Nonetheless, even when tissue-specific, we expect that genes found in the (all-tissue) meta-analysis are robustly associated with CR due to the large variation in different co-variates (e.g. organism,

duration of CR, ...) between the original studies. Of the functional categories found in the DAVID-analysis of the all-tissue candidates "circadian rhythm" and "NADP" related categories can be strongly assumed to be tissue-independent, since they were also found significantly enriched among genes found overexpressed in at least 3 different tissues.

3.4.2 Comparison with results from other meta-analyses

The other meta-analyses on CR presented in "3.1.3 Other meta-analyses of gene expression data for CR" were somewhat different from ours as far as the aim was concerned. While our focus was on determining genes with a mechanistic effect in CR other studies set out to find any genes differentially expressed with CR, no matter if due to the role of the gene in the mechanism of CR or due to the effect of CR on the expression of the gene. Hong (Hong, S. et al. 2010) even explicitly reported genes and modules for which differential expression was opposite of the change found with aging. Their expression changes are more likely to be an effect than a cause of the mechanism of CR. Even though in this kind of analysis there is of course no way to determine if a gene really mechanistically contributes to CR we expected to make this more likely by excluding genes, which we suspected were only found differentially expressed with CR in old animals due to the lack of the normal expression change with age as an effect of CR (see "3.2.2.5 Excluding genes differentially expressed with age"). Even though we could only do this for studies on old animals that also provided microarray data from young AL animals this is one of the major differences of our analysis to these of others.

A summary of other meta-analyses of CR microarray data in comparison with our meta-analysis is shown in table 3.15.

Since our study is more recent than the other ones mentioned, we were able to include more datasets into the meta-analysis. This makes especially a difference compared to Swindell, 2008a and Hong, 2010, while Swindell, 2009 included a comparable number of studies. Importantly while all meta-analyses (in at least part of the study) used data from different tissues all but ours focused only on data from mouse. In this respect we have to admit that also the fast majority of datasets in our study was from mice and that in some cases data-loss during annotation with mouse gene identifiers limited the contribution of non-mouse studies. While we expect that the use of different organisms strengthened the robustness of our findings we cannot claim all determined candidates to be organism-independent.

Our meta-analysis was not so focused on tissue-independence of the findings as Swindell, 2008a. While Swindell accepted to loose information by only counting if a gene was differentially expressed in any dataset of a certain tissue and ignoring in how many of these datasets it was detected, we counted occurrences of differential expression independently of the tissue arguing that variability in other covariates introduced sufficient robustness.

As for the statistical procedure we used a value-counting approach as did Swindell, 2008a. Since this study counted the number of tissues in which a gene was over- / underexpressed, but did not account for the number of datasets in which a gene was studied a bias for detecting genes studied more often is introduced. We tried to overcome this problem by employing a binomial test. Swindell, 2009 used Fisher´s inverse chi-square approach which is, since it is based on p-values, relatively sensitive to single datasets not fitting a certain differential expression trend in other datasets. This might e.g. lead to not detecting a gene that is robustly differentially expressed over many studies in animals up to a certain age, but not any more in very old animals. Since it is not sure if CR exerts its effect over all the life, every tissue, etc. it seems to be reasonable to want to find such a gene significant. Therefore we chose a value-counting approach which is not sensitive to these cases. Hong, 2010 simply pooled genes found in different studies and then e.g. searched for enriched functional categories. Therefore this can be understood as a meta-analysis on the level of e.g. the functional categories, but not on gene level.

Surprisingly many genes were found differentially expressed in Swindell, 2009. For many of the top genes there was contradicting evidence (upregulation in some, downregulation in other datasets) rather than indicating some of them as non-significant. The number of non-significant results was generally very low. It appears likely that the high number of significant results in the individual studies, rather than much higher power of the Fisher´s chi-square over the value-counting approach lead to the large number of detected genes.

As Swindell (Swindell, 2008a) we found *Per1, Per2, Mt1, Mt2, Fkbp5, Sult1a1* (and additionally *Sult1c2, Sult1d1* and *Sult3a1*), *Ppara* and *Nfkbia* enriched for overexpression and *Col3a1* (but not *Col1a1*, however *Col5a1*), for underexpression. We did not find *Hsp10* for underexpression, but *Hsp5* and *Hsp7*, not *Ifi27*, but *Ifi27l2α* (interferon, alpha-inducible protein 27 like 2A Gene). The overlap with the genes he found overrepresented for overexpression was therefore much bigger than with those he found for underexpression. Note that finding similar

meta-analysis	Swindell, 2008a	Swindell, 2009	Hong, 2010	this meta-analysis
number of studies	13	21	6	23
number of tissues	10	17	5	19
organism(s)	mouse	mouse	mouse	mouse, rat, pig, rhesus monkey
meta-analysis technique	value counting	Fisher´s inverse chi-square	pooling diff. exp. genes	value counting
number of significant genes	28	12114	N.A. (pool: 586)	175
comment	seperately for liver	seperately for liver, heart and muscle		seperately for liver and all but liver

Table 3.15: Comparison of different meta-analyses of microarray studies on CR. Since Hong, 2010 only pooled the data from different studies and performed analyses on those, this can be understood as a meta-analysis on the level of underlying categories, but not on gene level.

genes could result from not unambiguously matching probes as well as from that the genes may have similar functions.

Overall there was good agreement between the functional categories determined in our and the other meta-analyses. Especially all of them reported lipid metabolism or similar categories to be among of the most significant findings. Apart from that Swindell, 2009 mentioned "circadian rhythm" as another important result. As for subcellular localization the lysosome, mitochondria and endoplasmatic reticulum were enriched among genes differentially expressed with CR. On the other hand the studies also displayed differences to one another. Apparently no other study than ours assigned an important role to copper-ion detoxification and retinol metabolism.

3.4.3 Perspective

This meta-analysis provides a large number of candidate genes that are robustly differentially expressed with CR and functional categories associated with such genes. These genes and categories range from such already extensively studied for their role in CR, which suggests that our results are biologically meaningful, to such that received less attention and some that were not at all associated with CR before. For further studies on the relationship of these categories with CR the candidates associated within them, their co-expressed genes and transcription factors regulating their expression can serve as a starting point.
Meta-analyses are already a powerful and inexpensive method to draw information from already existing data. We expect that meta-analyses on high throughput studies will become even more valuable once e.g. next generation sequencing and proteomics data are added to the microarray data already deposited in public databases.
For meta-analyses on CR increasing availability of studies on invertebrates might allow a better understanding of evolutionary conserved pathways acting during CR.
Meta-analyses like this would be more powerful if raw data from all studies performed would be provided in databases or at least by the researchers upon request, so that there is no need to include supplemental data, requiring many compromises in the approach.

Acknowledgements

I thank my colleagues Shona Wood, Thomas Craig, Daniel Wuttke, Emily Hesketh, Sipko van Dam, Andrew Holmes, Alex Freitas and Yang Li for the pleasant atmosphere in our group. Additionally I acknowledge Daniel´s, Sipko´s and Emily´s contributions to this project and Sofia Silva´s continuation of the experimental work. I thank all members of the School of Biological Sciences (University of Liverpool) for their advice on experimental and computational methods. Further I appreciate all the helpful advice from the Perl and R community forums. I thank Prof. Ivo Hofacker without whom my work on this project would not have been possible for co-supervision at the University of Vienna. Especially I thank Dr. Pedro de Magalhaes for the opportunity to work in his lab, all our discussions and his ideas and expertise contributed to this work.

References

Aggarwal, A. et al., 2006. Topological and functional discovery in a gene coexpression meta-network of gastric cancer. Cancer Research, 66(1), 232-241.

Amador-Noguez, D. et al., 2007. Alterations in xenobiotic metabolism in the long-lived Little mice. Aging Cell, 6(4), 453-470.

Anderson, R.M., Shanmuganayagam, D. & Weindruch, R., 2009. Caloric restriction and aging: studies in mice and monkeys. Toxicologic Pathology, 37(1), 47-51.

Anderson, R.M. & Weindruch, R., 2010. Metabolic reprogramming, caloric restriction and aging. Trends in Endocrinology and Metabolism: TEM, 21(3), 134-141.

Anderson, R.M. & Weindruch, R., 2007. Metabolic reprogramming in dietary restriction. Interdisciplinary Topics in Gerontology, 35, 18-38.

Anisimov, V.N., Semenchenko, A.V. & Yashin, A.I., 2003. Insulin and longevity: antidiabetic biguanides as geroprotectors. Biogerontology, 4(5), 297-307.

Anson, R.M. et al., 2003. Intermittent fasting dissociates beneficial effects of dietary restriction on glucose metabolism and neuronal resistance to injury from calorie intake. Proceedings of the National Academy of Sciences of the United States of America, 100(10), 6216-6220.

Apfeld, J. et al., 2004. The AMP-activated protein kinase AAK-2 links energy levels and insulin-like signals to lifespan in C. elegans. Genes & Development, 18(24), 3004-3009.

Asami, D.K. et al., 2008. Effect of aging, caloric restriction, and uncoupling protein 3 (UCP3) on mitochondrial proton leak in mice. Experimental Gerontology, 43(12), 1069-1076.

Ashrafi, K. et al., 2000. Sip2p and its partner snf1p kinase affect aging in S. cerevisiae. Genes & Development, 14(15), 1872-1885.

Bader, G.D. et al., 2001. BIND–The Biomolecular Interaction Network Database. Nucleic Acids Research, 29(1), 242-245.

Bailey, T.L. & Elkan, C., 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology, 2, 28-36.

Bargmann, C.I. & Horvitz, H.R., 1991a. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in C. elegans. Neuron, 7(5), 729-742.

Bargmann, C.I. & Horvitz, H.R., 1991b. Control of larval development by chemosensory neurons in Caenorhabditis elegans. Science (New York, N.Y.), 251(4998), 1243-1246.

Barrett, T. et al., 2009. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Research, 37(Database issue), D885-890.

Bartke, A. et al., 2001. Extending the lifespan of long-lived mice. Nature, 414(6862), 412.

Bartke, A., 2005. Minireview: role of the growth hormone/insulin-like growth factor system in mammalian aging. Endocrinology, 146(9), 3718-3723.

Baxter, M.A. et al., 2004. Study of telomere length reveals rapid aging of human marrow stromal cells following in vitro expansion. Stem Cells (Dayton, Ohio), 22(5), 675-682.

Benjamini, Y; Hochberg, Y, B., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289-300.

Berggren Söderlund, M., Fex, G. & Nilsson-Ehle, P., 2003. Decreasing serum concentrations of all-trans, 13-cis retinoic acids and retinol during fasting and caloric restriction. Journal of Internal Medicine, 253(3), 375-380.

Beuming, T. et al., 2005. PDZBase: a protein-protein interaction database for PDZ-domains. Bioinformatics (Oxford, England), 21(6), 827-828.

Bishop, N.A. & Guarente, L., 2007a. Genetic links between diet and lifespan: shared mechanisms from yeast to humans. Nature Reviews. Genetics, 8(11), 835-844.

Bishop, N.A. & Guarente, L., 2007b. Two neurons mediate diet-restriction-induced longevity in C. elegans. Nature, 447(7144), 545-549.

Bodkin, N.L. et al., 2003. Mortality and morbidity in laboratory-maintained Rhesus monkeys and effects of long-term dietary restriction. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 58(3), 212-219.

Boily, G. et al., 2008. SirT1 regulates energy metabolism and response to caloric restriction in mice. PloS One, 3(3), e1759.

Bonafè, M. et al., 2003. Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: cues for an evolutionarily conserved mechanism of life span control. The Journal of Clinical Endocrinology and Metabolism, 88(7), 3299-3304.

Bonkowski, M.S. et al., 2006. Targeted disruption of growth hormone receptor interferes with the beneficial actions of calorie restriction. Proceedings of the National Academy of Sciences of the United States of America, 103(20), 7901-7905.

Bonorden, M.J.L. et al., 2009. Intermittent calorie restriction delays prostate tumor detection and increases survival time in TRAMP mice. Nutrition and Cancer, 61(2), 265-275.

Brack, A.S., Bildsoe, H. & Hughes, S.M., 2005. Evidence that satellite cell decrement contributes to preferential decline in nuclear number from large fibres during murine age-related muscle atrophy. Journal of Cell Science, 118(Pt 20), 4813-4821.

Brown-Borg, H.M. et al., 1996. Dwarf mice and the ageing process. Nature, 384(6604), 33.

Brown-Borg, H.M., 2007. Hormonal regulation of longevity in mammals. Ageing Research Reviews, 6(1), 28-45.

Brown, K.R. & Jurisica, I., 2007. Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biology, 8(5), R95.

Bruggeman, S.W.M. et al., 2005. Ink4a and Arf differentially affect cell proliferation and neural stem cell self-renewal in Bmi1-deficient mice. Genes & Development, 19(12), 1438-1443.

Bult, C.J. et al., 2008. The Mouse Genome Database (MGD): mouse biology and model systems. Nucleic Acids Research, 36(Database issue), D724-728.

Bunn, H.F., Gabbay, K.H. & Gallop, P.M., 1978. The glycosylation of hemoglobin: relevance to diabetes mellitus. Science (New York, N.Y.), 200(4337), 21-27.

Cai, W. et al., 2008. Oral glycotoxins determine the effects of calorie restriction on oxidant stress, age-related diseases, and lifespan. The American Journal of Pathology, 173(2), 327-336.

Cantó, C. et al., 2009. AMPK regulates energy expenditure by modulating NAD^+ metabolism and SIRT1 activity. Nature, 458(7241), 1056-1060.

Cao, S.X. et al., 2001. Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice. Proceedings of the National Academy of Sciences of the United States of America, 98(19), 10630-10635.

Cha, D.R. et al., 2007. Peroxisome proliferator-activated receptor-alpha deficiency protects aged mice from insulin resistance induced by high-fat diet. American Journal of Nephrology, 27(5), 479-482.

Chang, K., Elledge, S.J. & Hannon, G.J., 2006. Lessons from Nature: microRNA-based shRNA libraries. Nature Methods, 3(9), 707-714.

Cheadle, C. et al., 2003. Analysis of microarray data using Z score transformation. The Journal of Molecular Diagnostics: JMD, 5(2), 73-81.

Chen, D. et al., 2008. Tissue-specific regulation of SIRT1 by calorie restriction. Genes & Development, 22(13), 1753-1757.

Chen, D. et al., 2005. Increase in activity during calorie restriction requires Sirt1. Science (New York, N.Y.), 310(5754), 1641.

Chen, D., Thomas, E.L. & Kapahi, P., 2009. HIF-1 modulates dietary restriction-mediated lifespan extension via IRE-1 in Caenorhabditis elegans. PLoS Genetics, 5(5), e1000486.

Choi, J.K. et al., 2004. Integrative analysis of multiple gene expression profiles applied to liver cancer study. FEBS Letters, 565(1-3), 93-100.

Choi, J.K. et al., 2003. Combining multiple microarray studies and modeling interstudy variation. Bioinformatics (Oxford, England), 19 Suppl 1, i84-90.

Civitarese, A.E. et al., 2007. Calorie restriction increases muscle mitochondrial biogenesis in healthy humans. PLoS Medicine, 4(3), e76.

Civitarese, A.E. et al., 2006. Role of adiponectin in human skeletal muscle bioenergetics. Cell Metabolism, 4(1), 75-87.

Clancy, D.J. et al., 2002. Dietary restriction in long-lived dwarf flies. Science (New York, N.Y.), 296(5566), 319.

Clancy, D.J. et al., 2001. Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. Science (New York, N.Y.), 292(5514), 104-106.

Cleary, M.P. et al., 2007. Prevention of mammary tumorigenesis by intermittent caloric restriction: does caloric intake during refeeding modulate the response? Experimental Biology and Medicine (Maywood, N.J.), 232(1), 70-80.

Cohen, H.Y. et al., 2004. Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. Science (New York, N.Y.), 305(5682), 390-392.

Colman, R.J. et al., 2009. Caloric restriction delays disease onset and mortality in rhesus monkeys. Science (New York, N.Y.), 325(5937), 201-204.

Conboy, I.M. et al., 2003. Notch-mediated restoration of regenerative potential to aged muscle. Science (New York, N.Y.), 302(5650), 1575-1577.

Conboy, I.M. et al., 2005. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. Nature, 433(7027), 760-764.

Conti, B. et al., 2006. Transgenic mice with a reduced core body temperature have an increased life span. Science (New York, N.Y.), 314(5800), 825-828.

Cooper, T.M. et al., 2004. Effect of caloric restriction on life span of the housefly, Musca domestica. The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology, 18(13), 1591-1593.

Corton, J.C. et al., 2004. Mimetics of caloric restriction include agonists of lipid-activated nuclear receptors. The Journal of Biological Chemistry, 279(44), 46204-46212.

Coschigano, K.T. et al., 2003. Deletion, but not antagonism, of the mouse growth hormone receptor results in severely decreased body weights, insulin, and insulin-like growth factor I levels and increased life span. Endocrinology, 144(9), 3799-3810.

Coschigano, K.T. et al., 2000. Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. Endocrinology, 141(7), 2608-2613.

Crooks, G.E. et al., 2004. WebLogo: a sequence logo generator. Genome Research, 14(6), 1188-1190.

Curtis, R., O'Connor, G. & DiStefano, P.S., 2006. Aging networks in Caenorhabditis elegans: AMP-activated protein kinase (aak-2) links multiple aging and metabolism pathways. Aging Cell, 5(2), 119-126.

DeConde, R.P. et al., 2006. Combining results of microarray experiments: a rank aggregation approach. Statistical Applications in Genetics and Molecular Biology, 5, Article15.

Dennis, G. et al., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology, 4(5), p.P3.

DerSimonian, R. & Laird, N., 1986. Meta-analysis in clinical trials. Controlled Clinical Trials, 7(3), 177-188.

Dhahbi, J.M. et al., 2004. Temporal linkage between the phenotypic and genomic responses to caloric restriction. Proceedings of the National Academy of Sciences of the United States of America, 101(15), 5524-5529.

Dhahbi, J.M. et al., 2006. Gene expression and physiologic responses of the heart to the initiation and withdrawal of caloric restriction. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 61(3), 218-231.

Easlon, E. et al., 2007. The dihydrolipoamide acetyltransferase is a novel metabolic longevity factor and is required for calorie restriction-mediated life span extension. The Journal of Biological Chemistry, 282(9), 6161-6171.

Edwards, M.G. et al., 2007. Gene expression profiling of aging reveals activation of a p53-mediated transcriptional program. BMC Genomics, 8, 80.

Enwere, E. et al., 2004. Aging results in reduced epidermal growth factor receptor signaling, diminished olfactory neurogenesis, and deficits in fine olfactory discrimination. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 24(38), 8354-8365.

Fabrizio, P. et al., 2001. Regulation of longevity and stress resistance by Sch9 in yeast. Science (New York, N.Y.), 292(5515), 288-290.

Fabrizio, P. et al., 2005. Sir2 blocks extreme life-span extension. Cell, 123(4), 655-667.

Falcon, S. & Gentleman, R., 2007. Using GOstats to test gene lists for GO term association. Bioinformatics (Oxford, England), 23(2), 257-258.

Feige, J.N. et al., 2008. Specific SIRT1 activation mimics low energy levels and protects against diet-induced metabolic disorders by enhancing fat oxidation. Cell Metabolism, 8(5), 347-358.

Figueiredo, P.A. et al., 2009. Aging impairs skeletal muscle mitochondrial bioenergetic function. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 64(1), 21-33.

Fisher, R.A., 1925. Statistical Methods for Research Worker, Edinburg and London: Oliver and Boyd.

Flurkey, K. et al., 2001. Lifespan extension and delayed immune and collagen aging in mutant mice with defects in growth hormone production. Proceedings of the National Academy of Sciences of the United States of America, 98(12), 6736-6741.

Fontana, L. & Klein, S., 2007. Aging, adiposity, and calorie restriction. JAMA: The Journal of the American Medical Association, 297(9), 986-994.

Fontana, L. et al., 2004. Long-term calorie restriction is highly effective in reducing the risk for atherosclerosis in humans. Proceedings of the National Academy of Sciences of the United States of America, 101(17), 6659-6663.

Fontana, L., Partridge, L. & Longo, V.D., 2010. Extending healthy life span–from yeast to humans. Science (New York, N.Y.), 328(5976), 321-326.

Fontana, L. et al., 2008. Long-term effects of calorie or protein restriction on serum IGF-1 and IGFBP-3 concentration in humans. Aging Cell, 7(5), 681-687.

Forster, M.J., Morris, P. & Sohal, R.S., 2003. Genotype and age influence the effect of caloric intake on mortality in mice. The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology, 17(6), 690-692.

Frey, J., 2004. Collagen, ageing and nutrition. Clinical Chemistry and Laboratory Medicine: CCLM / FESCC, 42(1), 9-12.

Fridell, Y.C. et al., 2005. Targeted expression of the human uncoupling protein 2 (hUCP2) to adult neurons extends life span in the fly. Cell Metabolism, 1(2), 145-152.

Froy, O. & Miskin, R., 2010. Effect of feeding regimens on circadian rhythms: implications for aging and longevity. Aging, 2(1), 7-27.

Fu, C. et al., 2006. Tissue specific and non-specific changes in gene expression by aging and by early stage CR. Mechanisms of Ageing and Development, 127(12), 905-916.

Gentleman, R.C. et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biology, 5(10), R80.

Giannakou, M.E., Goss, M. & Partridge, L., 2008. Role of dFOXO in lifespan extension by dietary restriction in Drosophila melanogaster: not required, but its activity modulates the response. Aging Cell, 7(2), 187-198.

Gibson, M.C. & Schultz, E., 1983. Age-related differences in absolute numbers of skeletal muscle satellite cells. Muscle & Nerve, 6(8), 574-580.

Girotti, M., Weinberg, M.S. & Spencer, R.L., 2009. Diurnal expression of functional and clock-related genes throughout the rat HPA axis: system-wide shifts in response to a restricted feeding schedule. American Journal of Physiology. Endocrinology and Metabolism, 296(4), E888-897.

Good,I.J., 1955. On the weighted combination of significance tests. J. R. Stat.Soc., 2, 264–265.

Goodrick, C.L., 1978. Body weight increment and length of life: the effect of genetic constitution and dietary protein. Journal of Gerontology, 33(2), 184-190.

Gourley, M.E. & Kennedy, C.J., 2009. Energy allocations to xenobiotic transport and biotransformation reactions in rainbow trout (Oncorhynchus mykiss) during energy intake restriction. Comparative Biochemistry and Physiology. Toxicology & Pharmacology: CBP, 150(2), 270-278.

Greer, E.L. & Brunet, A., 2009. Different dietary restriction regimens extend lifespan by both independent and overlapping genetic pathways in C. elegans. Aging Cell, 8(2), 113-127.

Grönke, S. et al., 2010. Molecular evolution and functional characterization of Drosophila insulin-like peptides. PLoS Genetics, 6(2), e1000857.

Guarente, L., 2005. Calorie restriction and SIR2 genes–towards a mechanism. Mechanisms of Ageing and Development, 126(9), 923-928.

Güldener, U. et al., 2006. MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Research, 34(Database issue), D436-441.

Gwinn, D.M. et al., 2008. AMPK phosphorylation of raptor mediates a metabolic checkpoint. Molecular Cell, 30(2), 214-226.

Hailesellasse Sene, K. et al., 2007. Gene function in early mouse embryonic stem cell differentiation. BMC Genomics, 8, 85.

Halaschek-Wiener, J. & Brooks-Wilson, A., 2007. Progeria of stem cells: stem cell exhaustion in Hutchinson-Gilford progeria syndrome. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 62(1), 3-8.

Halberg, N. et al., 2005. Effect of intermittent fasting and refeeding on insulin action in healthy men. Journal of Applied Physiology (Bethesda, Md.: 1985), 99(6), 2128-2136.

Hansen, M. et al., 2008. A role for autophagy in the extension of lifespan by dietary restriction in C. elegans. PLoS Genetics, 4(2), e24.

Hansen, M. et al., 2005. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a Caenorhabditis elegans genomic RNAi screen. PLoS Genetics, 1(1), 119-128.

Harper, J.M., Leathers, C.W. & Austad, S.N., 2006. Does caloric restriction extend life in wild mice? Aging Cell, 5(6), 441-449.

Harrison, D.E. et al., 2009. Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. Nature, 460(7253), 392-395.

Heilbronn, L.K. et al., 2005. Glucose tolerance and skeletal muscle gene expression in response to alternate day fasting. Obesity Research, 13(3), 574-581.

Hermjakob, H. et al., 2004. IntAct: an open source molecular interaction database. Nucleic Acids Research, 32(Database issue), D452-455.

Higami, Y. et al., 2006. Energy restriction lowers the expression of genes linked to inflammation, the cytoskeleton, the extracellular matrix, and angiogenesis in mouse adipose tissue. The Journal of Nutrition, 136(2), 343-352.

Higami, Y. et al., 2004. Adipose tissue energy metabolism: altered gene expression profile of mice subjected to long-term caloric restriction. The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology, 18(2), 415-417.

Hinrichs, A.S. et al., 2006. The UCSC Genome Browser Database: update 2006. Nucleic Acids Research, 34(Database issue), D590-598.

Hong, F. & Breitling, R., 2008. A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments. Bioinformatics (Oxford, England), 24(3), 374-382.

Hong, S. et al., 2010. Revealing system-level correlations between aging and calorie restriction using a mouse transcriptome. Age (Dordrecht, Netherlands), 32(1), 15-30.

Hornbeck, P.V. et al., 2004. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. Proteomics, 4(6), 1551-1561.

Houthoofd, K. et al., 2003. Life extension via dietary restriction is independent of the Ins/IGF-1 signalling pathway in Caenorhabditis elegans. Experimental Gerontology, 38(9), 947-954.

Howitz, K.T. et al., 2003. Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. Nature, 425(6954), 191-196.

Hubbard, T.J.P. et al., 2009. Ensembl 2009. Nucleic Acids Research, 37(Database issue), D690-697.

Huffman, D.M. et al., 2008. Effect of exercise and calorie restriction on biomarkers of aging in mice. American Journal of Physiology. Regulatory, Integrative and Comparative Physiology, 294(5), R1618-1627.

Hughes, J.D. et al., 2000. Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. Journal of Molecular Biology, 296(5), 1205-1214.

Ingram, D.K. et al., 2006. Calorie restriction mimetics: an emerging research field. Aging Cell, 5(2), 97-108.

Irizarry, R.A. et al., 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics (Oxford, England), 4(2), 249-264.

Jayapandian, M. et al., 2007. Michigan Molecular Interactions (MiMI): putting the jigsaw puzzle together. Nucleic Acids Research, 35(Database issue), D566-571.

Kaeberlein, M., Burtner, C.R. & Kennedy, B.K., 2007. Recent developments in yeast aging. PLoS Genetics, 3(5), e84.

Kaeberlein, M. et al., 2004. Sir2-independent life span extension by calorie restriction in yeast. PLoS Biology, 2(9), E296.

Kaeberlein, M. et al., 2005. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. Science (New York, N.Y.), 310(5751), 1193-1196.

Kanehisa, M. et al., 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Research, 38(Database issue), D355-360.

Kauffmann, A. et al., 2009. Importing ArrayExpress datasets into R/Bioconductor. Bioinformatics (Oxford, England), 25(16), 2092-2094.

Kayo, T. et al., 2001. Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. Proceedings of the National Academy of Sciences of the United States of America, 98(9), 5093-5098.

Kealy, R.D. et al., 2002. Effects of diet restriction on life span and age-related changes in dogs. Journal of the American Veterinary Medical Association, 220(9), 1315-1320.

Kenyon, C. et al., 1993. A C. elegans mutant that lives twice as long as wild type. Nature, 366(6454), 461-464.

Kenyon, C., 2005. The plasticity of aging: insights from long-lived mutants. Cell, 120(4), 449-460.

Keshava Prasad, T.S. et al., 2009. Human Protein Reference Database–2009 update. Nucleic Acids Research, 37(Database issue), D767-772.

Klass, M.R., 1977. Aging in the nematode Caenorhabditis elegans: major biological and environmental factors influencing life span. Mechanisms of Ageing and Development, 6(6), 413-429.

Klebanov, S., 2007. Can short-term dietary restriction and fasting have a long-term anticarcinogenic effect? Interdisciplinary Topics in Gerontology, 35, 176-192.

Kondo, M. et al., 2009. Caloric restriction stimulates revascularization in response to ischemia via adiponectin-mediated activation of endothelial nitric-oxide synthase. The Journal of Biological Chemistry, 284(3), 1718-1724.

Kondratov, R.V. et al., 2006. Early aging and age-related pathologies in mice deficient in BMAL1, the core componentof the circadian clock. Genes & Development, 20(14), 1868-1873.

Koubova, J. & Guarente, L., 2003. How does calorie restriction work? Genes & Development, 17(3), 313-321.

Kranendonk, M. et al., 2008. Impairment of human CYP1A2-mediated xenobiotic metabolism by Antley-Bixler syndrome variants of cytochrome P450 oxidoreductase. Archives of Biochemistry and Biophysics, 475(2), 93-99.

Kristan, D.M., 2008. Calorie restriction and susceptibility to intact pathogens. Age (Dordrecht, Netherlands), 30(2-3), 147-156.

Kumar, S. et al., 2009. Interactive effect of excitotoxic injury and dietary restriction on neurogenesis and neurotrophic factors in adult male rat brain. Neuroscience Research, 65(4), 367-374.

Kuningas, M. et al., 2007. Haplotypes in the human Foxo1a and Foxo3a genes; impact on disease and mortality at old age. European Journal of Human Genetics: EJHG, 15(3), 294-301.

Lakowski, B. & Hekimi, S., 1998. The genetics of caloric restriction in Caenorhabditis elegans. Proceedings of the National Academy of Sciences of the United States of America, 95(22), 13091-13096.

Lamming, D.W. et al., 2005. HST2 mediates SIR2-independent life-span extension by calorie restriction. Science (New York, N.Y.), 309(5742), 1861-1864.

Lee, C.C., 2006. Tumor suppression by the mammalian Period genes. Cancer Causes & Control: CCC, 17(4), 525-530.

Lee, C.K. et al., 1999. Gene expression profile of aging and its retardation by caloric restriction. Science (New York, N.Y.), 285(5432), 1390-1393.

Leto, S., Kokkonen, G.C. & Barrows, C.H., 1976. Dietary protein life-span, and physiological variables in female mice. Journal of Gerontology, 31(2), 149-154.

Libert, S. et al., 2007. Regulation of Drosophila life span by olfaction and food-derived odors. Science (New York, N.Y.), 315(5815), 1133-1137.

Lin, S.J., Defossez, P.A. & Guarente, L., 2000. Requirement of NAD and SIR2 for life-span extension by calorie restriction in Saccharomyces cerevisiae. Science (New York, N.Y.), 289(5487), 2126-2128.

Lin, S. et al., 2004. Calorie restriction extends yeast life span by lowering the level of NADH. Genes & Development, 18(1), 12-16.

Lin, S. et al., 2002. Calorie restriction extends Saccharomyces cerevisiae lifespan by increasing respiration. Nature, 418(6895), 344-348.

Lipschitz, D.A., Mitchell, C.O. & Thompson, C., 1981. The anemia of senescence. American Journal of Hematology, 11(1), 47-54.

Liu, X.S., Brutlag, D.L. & Liu, J.S., 2002. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nature Biotechnology, 20(8), 835-839.

Loeb, J. & Northrop, J. H., 1917. On the influence of food and temperature upon the duration of life. J. Biol. Chem., 32, 103–121.

Longo, V.D. & Kennedy, B.K., 2006. Sirtuins in aging and age-related disease. Cell, 126(2), 257-268.

López-Torres, M. & Barja, G., 2008. Lowered methionine ingestion as responsible for the decrease in rodent mitochondrial oxidative stress in protein and dietary restriction possible implications for humans. Biochimica Et Biophysica Acta, 1780(11), 1337-1347.

Lottaz, C. et al., 2006. OrderedList–a bioconductor package for detecting similarity in ordered gene lists. Bioinformatics (Oxford, England), 22(18), 2315-2316.

Lu, J. et al., 2007. Different gene expression of skin tissues between mice with weight controlled by either calorie restriction or physical exercise. Experimental Biology and Medicine (Maywood, N.J.), 232(4), 473-480.

Macisaac, K.D. et al., 2006. A hypothesis-based approach for identifying the binding specificity of regulatory proteins from chromatin immunoprecipitation data. Bioinformatics (Oxford, England), 22(4), 423-429.

de Magalhães, J.P., Curado, J. & Church, G.M., 2009. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics (Oxford, England), 25(7), 875-881.

de Magalhães, J.P. & Toussaint, O., 2004. GenAge: a genomic and proteomic network map of human ageing. FEBS Letters, 571(1-3), 243-247.

Mair, W. & Dillin, A., 2008. Aging and survival: the genetics of life span extension by dietary restriction. Annual Review of Biochemistry, 77, 727-754.

Martini, C. et al., 2008. Omega-3 as well as caloric restriction prevent the age-related modifications of cholesterol metabolism. Mechanisms of Ageing and Development, 129(12), 722-727.

Maslov, A.Y. et al., 2004. Neural stem cell detection, characterization, and age-related changes in the subventricular zone of mice. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 24(7), 1726-1733.

Maswood, N. et al., 2004. Caloric restriction increases neurotrophic factor levels and attenuates neurochemical and behavioral deficits in a primate model of Parkinson's disease. Proceedings of the National Academy of Sciences of the United States of America, 101(52), 18171-18176.

Matoba, S. et al., 2006. p53 regulates mitochondrial respiration. Science (New York, N.Y.), 312(5780), 1650-1653.

Mattison, J.A. et al., 2003. Calorie restriction in rhesus monkeys. Experimental Gerontology, 38(1-2), 35-46.

Mattson, M.P. & Wan, R., 2005. Beneficial effects of intermittent fasting and caloric restriction on the cardiovascular and cerebrovascular systems. The Journal of Nutritional Biochemistry, 16(3), 129-137.

McCay, C.M., Crowell, M.F. & Maynard, L.A., 1989. The effect of retarded growth upon the length of life span and upon the ultimate body size. 1935. Nutrition (Burbank, Los Angeles County, Calif.), 5(3), 155-171; discussion 172.

McElwee, J.J. et al., 2007. Evolutionary conservation of regulated longevity assurance mechanisms. Genome Biology, 8(7), R132.

Medvedik, O. et al., 2007. MSN2 and MSN4 link calorie restriction and TOR to sirtuin-mediated lifespan extension in Saccharomyces cerevisiae. PLoS Biology, 5(10), e261.

Michaut, M. et al., 2008. InteroPORC: automated inference of highly conserved protein interaction networks. Bioinformatics (Oxford, England), 24(14), 1625-1631.

Miller, D.S. & Payne, P.R., 1968. Longevity and protein intake. Experimental Gerontology, 3(3), 231-234.

Min, K. & Tatar, M., 2006. Restriction of amino acids extends lifespan in Drosophila melanogaster. Mechanisms of Ageing and Development, 127(7), 643-646.

Min, K. et al., 2008. Drosophila lifespan control by dietary restriction independent of insulin-like signaling. Aging Cell, 7(2), 199-206.

Molofsky, A.V. et al., 2005. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16Ink4a and p19Arf senescence pathways. Genes & Development, 19(12), 1432-1437.

Molofsky, A.V. et al., 2006. Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. Nature, 443(7110), 448-452.

Mulligan, J.D., Stewart, A.M. & Saupe, K.W., 2008. Downregulation of plasma insulin levels and hepatic PPARgamma expression during the first week of caloric restriction in mice. Experimental Gerontology, 43(3), 146-153.

Needleman, S.B. & Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology, 48(3), 443-453.

Nisoli, E. et al., 2005. Calorie restriction promotes mitochondrial biogenesis by inducing the expression of eNOS. Science (New York, N.Y.), 310(5746), 314-317.

Normand, S.L., 1999. Tutorial in biostatistics-meta-analysis: formulating, evaluating, combining, and reporting. Stat. Med., 18, 321-359.

Ogden, D.A. & Mickliem, H.S., 1976. The fate of serially transplanted bone marrow cell populations from young and old donors. Transplantation, 22(3), 287-293.

O'Hagan, K.A. et al., 2009. PGC-1alpha is coupled to HIF-1alpha-dependent gene expression by increasing mitochondrial oxygen consumption in skeletal muscle cells. Proceedings of the National Academy of Sciences of the United States of America, 106(7), 2188-2193.

Orentreich, N. et al., 1993. Low methionine ingestion by rats extends life span. The Journal of Nutrition, 123(2), 269-274.

Pan, F. et al., 2007. Gene Aging Nexus: a web database and data mining platform for microarray data on aging. Nucleic Acids Research, 35(Database issue), D756-759.

Pan, H. et al., 2008. Dual-promoter lentiviral system allows inducible expression of noxious proteins in macrophages. Journal of Immunological Methods, 329(1-2), 31-44.

Panowski, S.H. et al., 2007. PHA-4/Foxa mediates diet-restriction-induced longevity of C. elegans. Nature, 447(7144), 550-555.

Parkinson, H. et al., 2009. ArrayExpress update–from an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Research, 37(Database issue), D868-872.

Park, S. & Prolla, T.A., 2005. Lessons learned from gene expression profile studies of aging and caloric restriction. Ageing Research Reviews, 4(1), 55-65.

Parmigiani, G. et al., 2004. A cross-study comparison of gene expression studies for the molecular classification of lung cancer. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 10(9), 2922-2927.

Pavesi, G. et al., 2004. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Research, 32(Web Server issue), W199-203.

Pearce, D.J. et al., 2007. Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging. Stem Cells (Dayton, Ohio), 25(4), 828-835.

Piper, M.D.W. et al., 2008. Separating cause from effect: how does insulin/IGF signalling control lifespan in worms, flies and mice? Journal of Internal Medicine, 263(2), 179-191.

Pontius JU, Wagner L, Schuler GD, 2003. UniGene: a unified view of the transcriptome. In The NCBI Handbook. Bethesda (MD).

Powers, R.W. et al., 2006. Extension of chronological life span in yeast by decreased TOR pathway signaling. Genes & Development, 20(2), 174-184.

Puca, A.A., Chatgilialoglu, C. & Ferreri, C., 2008. Lipid metabolism and diet: possible mechanisms of slow aging. The International Journal of Biochemistry & Cell Biology, 40(3), 324-333.

Pyne, S., Futcher, B. & Skiena, S., 2006. Meta-analysis based on control of false discovery rate: combining yeast ChIP-chip datasets. Bioinformatics (Oxford, England), 22(20), 2516-2522.

Rae, M., 2004. It's never too late: calorie restriction is effective in older mammals. Rejuvenation Research, 7(1), 3-8.

Raffaghello, L. et al., 2008. Starvation-dependent differential stress resistance protects normal but not cancer cells against high-dose chemotherapy. Proceedings of the National Academy of Sciences of the United States of America, 105(24), 8215-8220.

Ramasamy, A. et al., 2008. Key issues in conducting a meta-analysis of gene expression microarray datasets. PLoS Medicine, 5(9), e184.

Razick, S., Magklaras, G. & Donaldson, I.M., 2008. iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics, 9, 405.

R Development Core Team, 2009. R: A language and environment for statistical computing, Vienna, Austria: R Foundation for Statistical Computing.

Reed, M.J. et al., 1996. Enhanced cell proliferation and biosynthesis mediate improved wound repair in refed, caloric-restricted mice. Mechanisms of Ageing and Development, 89(1), 21-43.

Rhodes, D.R. et al., 2002. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Research, 62(15), 4427-4433.

Rhodes, D.R. et al., 2004. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proceedings of the National Academy of Sciences of the United States of America, 101(25), 9309-9314.

Rogina, B. & Helfand, S.L., 2004. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. Proceedings of the National Academy of Sciences of the United States of America, 101(45), 15998-16003.

Romer, K.A., Kayombya, G. & Fraenkel, E., 2007. WebMOTIFS: automated discovery, filtering and scoring of DNA sequence motifs using multiple programs and Bayesian approaches. Nucl. Acids Res., gkm376.

Rossi, D.J. et al., 2007. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. Nature, 447(7145), 725-729.

Rossi, D.J. et al., 2005. Cell intrinsic alterations underlie hematopoietic stem cell aging. Proceedings of the National Academy of Sciences of the United States of America, 102(26), 9194-9199.

Ross, M.H. & Bras, G., 1973. Influence of protein under- and overnutrition on spontaneous tumor prevalence in the rat. The Journal of Nutrition, 103(7), 944-963.

Ruepp, A. et al., 2008. CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Research, 36(Database issue), D646-650.

Safdie, F.M. et al., 2009. Fasting and cancer treatment in humans: A case series report. Aging, 1(12), 988-1007.

Saito, H. et al., 2008. Regulatory mechanism governing the diurnal rhythm of intestinal H+/peptide cotransporter 1 (PEPT1). American Journal of Physiology. Gastrointestinal and Liver Physiology, 295(2), G395-402.

Sakamoto, K. et al., 2008. Zfp64 participates in Notch signaling and regulates differentiation in mesenchymal cells. Journal of Cell Science, 121(Pt 10), 1613-1623.

Saleem, A., Adhihetty, P.J. & Hood, D.A., 2009. Role of p53 in mitochondrial biogenesis and apoptosis in skeletal muscle. Physiological Genomics, 37(1), 58-66.

Sánchez-Blanco, A., Fridell, Y.C. & Helfand, S.L., 2006. Involvement of Drosophila uncoupling protein 5 in metabolism and aging. Genetics, 172(3), 1699-1710.

Sayers, E.W. et al., 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 38(Database issue), D5-16.

Schmucker, D.L. et al., 1991. Caloric restriction affects liver microsomal monooxygenases differentially in aging male rats. Journal of Gerontology, 46(1), B23-27.

Schmuck, E.G., Mulligan, J.D. & Saupe, K.W., 2010. Caloric restriction attenuates the age-associated increase of adipose-derived stem cells but further reduces their proliferative capacity. Age (Dordrecht, Netherlands). Available at: http://www.ncbi.nlm.nih.gov/pubmed/20628827 [Accessed September 17, 2010].

Sean, D. & Meltzer, P.S., 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics (Oxford, England), 23(14), 1846-1847.

Segall, P.E. & Timiras, P.S., 1976. Patho-physiologic findings after chronic tryptophan deficiency in rats: a model for delayed growth and aging. Mechanisms of Ageing and Development, 5(2), 109-124.

Selman, C. et al., 2009. Ribosomal protein S6 kinase 1 signaling regulates mammalian life span. Science (New York, N.Y.), 326(5949), 140-144.

Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498-2504.

Sharov, A.A., Dudekula, D.B. & Ko, M.S.H., 2006. CisView: a browser and database of cis-regulatory modules predicted in the mouse genome. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes, 13(3), 123-134.

Sharpless, N.E. & DePinho, R.A., 2007. How stem cells age and why this makes us grow old. Nature Reviews. Molecular Cell Biology, 8(9), 703-713.

Shimokawa, I. et al., 1993. Diet and the suitability of the male Fischer 344 rat as a model for aging research. Journal of Gerontology, 48(1), B27-32.

Shinmura, K., Tamaki, K. & Bolli, R., 2008. Impact of 6-mo caloric restriction on myocardial ischemic tolerance: possible involvement of nitric oxide-dependent increase in nuclear Sirt1. American Journal of Physiology. Heart and Circulatory Physiology, 295(6), H2348-2355.

Shinmura, K. et al., 2007. Cardioprotective effects of short-term caloric restriction are mediated by adiponectin via activation of AMP-activated protein kinase. Circulation, 116(24), 2809-2817.

Silva, J.M. et al., 2005. Second-generation shRNA libraries covering the mouse and human genomes. Nature Genetics, 37(11), 1281-1288.

Siminovitch, L., TILL, J.E. & MCCULLOCH, E.A., 1964. DECLINE IN COLONY-FORMING ABILITY OF MARROW CELLS SUBJECTED TO SERIAL TRANSPLANTATION INTO IRRADIATED MICE. Journal of Cellular Physiology, 64, 23-31.

Sinclair, D.A. & Guarente, L., 1997. Extrachromosomal rDNA circles–a cause of aging in yeast. Cell, 91(7), 1033-1042.

Smid, M., Dorssers, L.C.J. & Jenster, G., 2003. Venn Mapping: clustering of heterologous microarray data based on the number of co-occurring differentially expressed genes. Bioinformatics (Oxford, England), 19(16), 2065-2071.

Smith, C.M. et al., 2007. The mouse Gene Expression Database (GXD): 2007 update. Nucleic Acids Research, 35(Database issue), D618-623.

Smith, E.D. et al., 2008. Age- and calorie-independent life span extension from dietary restriction by bacterial deprivation in Caenorhabditis elegans. BMC Developmental Biology, 8, 49.

Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology, 3, Article3.

Spindler, S.R., 2005. Rapid and reversible induction of the longevity, anticancer and genomic effects of caloric restriction. Mechanisms of Ageing and Development, 126(9), 960-966.

Stark, C. et al., 2006. BioGRID: a general repository for interaction datasets. Nucleic Acids Research, 34(Database issue), D535-539.

Stark, C. et al., 2010. PhosphoGRID: a database of experimentally verified in vivo protein phosphorylation sites from the budding yeast Saccharomyces cerevisiae. Database: The Journal of Biological Databases and Curation, 2010, bap026.

Stein, L.D., 2004. Using the Reactome database. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al, Chapter 8, Unit 8.7.

Stelmanska, E., Korczynska, J. & Swierczynski, J., 2004. Tissue-specific effect of refeeding after short- and long-term caloric restriction on malic enzyme gene expression in rat tissues. Acta Biochimica Polonica, 51(3), 805-814.

Stolzing, A., Coleman, N. & Scutt, A., 2006. Glucose-induced replicative senescence in mesenchymal stem cells. Rejuvenation Research, 9(1), 31-35.

Stuart, J.M. et al., 2003. A gene-coexpression network for global discovery of conserved genetic modules. Science (New York, N.Y.), 302(5643), 249-255.

Subramanian, A. et al., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545-15550.

Suh, Y. et al., 2008. Functionally significant insulin-like growth factor I receptor mutations in centenarians. Proceedings of the National Academy of Sciences of the United States of America, 105(9), 3438-3442.

Swindell, W.R., 2008. Comparative analysis of microarray data identifies common responses to caloric restriction among mouse tissues. Mechanisms of Ageing and Development, 129(3), 138-153.

Swindell, W.R., 2009. Genes and gene expression modules associated with caloric restriction and aging in the laboratory mouse. BMC Genomics, 10, 585.

Swindell, W.R., 2008. Genes regulated by caloric restriction have unique roles within transcriptional networks. Mechanisms of Ageing and Development, 129(10), 580-592.

Taguchi, A., Wartschow, L.M. & White, M.F., 2007. Brain IRS2 signaling coordinates life span and nutrient homeostasis. Science (New York, N.Y.), 317(5836), 369-372.

Tang, F. et al., 2008. A life-span extending form of autophagy employs the vacuole-vacuole fusion machinery. Autophagy, 4(7), 874-886.

Teillet, L. et al., 2000. Food restriction prevents advanced glycation end product accumulation and retards kidney aging in lean rats. Journal of the American Society of Nephrology: JASN, 11(8), 1488-1497.

Thirumoorthy, N. et al., 2007. Metallothionein: an overview. World Journal of Gastroenterology: WJG, 13(7), 993-996.

Tissenbaum, H.A. & Guarente, L., 2001. Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans. Nature, 410(6825), 227-230.

Tsuchiya, M. et al., 2006. Sirtuin-independent effects of nicotinamide on lifespan extension from calorie restriction in yeast. Aging Cell, 5(6), 505-514.

Tullet, J.M.A. et al., 2008. Direct inhibition of the longevity-promoting factor SKN-1 by insulin-like signaling in C. elegans. Cell, 132(6), 1025-1038.

Tusher, V.G., Tibshirani, R. & Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98(9), 5116-5121.

Vellai, T. et al., 2003. Genetics: influence of TOR kinase on lifespan in C. elegans. Nature, 426(6967), 620.

Waddington Lamont, E. et al., 2007. Restricted access to food, but not sucrose, saccharine, or salt, synchronizes the expression of Period2 protein in the limbic forebrain. Neuroscience, 144(2), 402-411.

Wang, G. et al., 2009. Resveratrol inhibits the expression of SREBP1 in cell model of steatosis via Sirt1-FOXO1 signaling pathway. Biochemical and Biophysical Research Communications, 380(3), 644-649.

Wang, Y. et al., 2006. Inferring gene regulatory networks from multiple microarray datasets. Bioinformatics (Oxford, England), 22(19), 2413-2420.

Wang, Y. et al., 2006. Post-translational modifications of the four conserved lysine residues within the collagenous domain of adiponectin are required for the formation of its high molecular weight oligomeric complex. The Journal of Biological Chemistry, 281(24), 16391-16400.

Warnat, P., Eils, R. & Brors, B., 2005. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. BMC Bioinformatics, 6, 265.

Watt, F.M., 2000. Epidermal stem cells as targets for gene transfer. Human Gene Therapy, 11(16), 2261-2266.

Wei, M. et al., 2008. Life span extension by calorie restriction depends on Rim15 and transcription factors downstream of Ras/PKA, Tor, and Sch9. PLoS Genetics, 4(1), e13.

Wei, M. et al., 2009. Tor1/Sch9-regulated carbon source substitution is as effective as calorie restriction in life span extension. PLoS Genetics, 5(5), e1000467.

Weissman, I.L., 2000. Stem cells: units of development, units of regeneration, and units in evolution. Cell, 100(1), 157-168.

Whitehead, R.H. et al., 1999. Clonogenic growth of epithelial cells from normal colonic mucosa from both mice and humans. Gastroenterology, 117(4), 858-865.

Wiles, A.M. et al., 2010. Building and analyzing protein interactome networks by cross-species comparisons. BMC Systems Biology, 4, 36.

Wood, J.G. et al., 2004. Sirtuin activators mimic caloric restriction and delay ageing in metazoans. Nature, 430(7000), 686-689.

Wright, L.S. et al., 2003. Gene expression in human neural stem cells: effects of leukemia inhibitory factor. Journal of Neurochemistry, 86(1), 179-195.

Wu, J. et al., 2009. Integrated network analysis platform for protein-protein interactions. Nature Methods, 6(1), 75-77.

Wu, P. et al., 2009. Systematic gene expression profile of hypothalamus in calorie-restricted mice implicates the involvement of mTOR signaling in neuroprotective activity. Mechanisms of Ageing and Development, 130(9), 602-610.

Wu, P. et al., 2008. Calorie restriction ameliorates neurodegenerative phenotypes in forebrain-specific presenilin-1 and presenilin-2 double knockout mice. Neurobiology of Aging, 29(10), 1502-1511.

Xenarios, I. et al., 2000. DIP: the database of interacting proteins. Nucleic Acids Research, 28(1), 289-291.

Yang, X., Bentink, S. & Spang, R., 2005. Detecting common gene expression patterns in multiple cancer outcome entities. Biomedical Microdevices, 7(3), 247-251.

Yoshida, K. et al., 2006. Caloric restriction prevents radiation-induced myeloid leukemia in C3H/HeMs mice and inversely increases incidence of tumor-free death: implications in changes in number of hemopoietic progenitor cells. Experimental Hematology, 34(3), 274-283.

Yu, B.P., Masoro, E.J. & McMahan, C.A., 1985. Nutritional influences on aging of Fischer 344 rats: I. Physical, metabolic, and longevity characteristics. Journal of Gerontology, 40(6), 657-670.

Yu, J. et al., 2008. DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics, 9, 461.

Zahn, J.M. et al., 2007. AGEMAP: a gene expression database for aging in mice. PLoS Genetics, 3(11), e201.

Zainal, T.A. et al., 2000. Caloric restriction of rhesus monkeys lowers oxidative damage in skeletal muscle. FASEB J., 14(12), 1825-1836.

Zanzoni, A. et al., 2002. MINT: a Molecular INTeraction database. FEBS Letters, 513(1), 135-140.

Zaykin, D.V. et al., 2002. Truncated product method for combining P-values. Genetic Epidemiology, 22(2), 170-185.

Zhu, M. et al., 2007. Adipogenic signaling in rat white adipose tissue: modulation by aging and calorie restriction. Experimental Gerontology, 42(8), 733-744.

Zhu, M. et al., 2004. Circulating adiponectin levels increase in rats on caloric restriction: the potential for insulin sensitization. Experimental Gerontology, 39(7), 1049-1059.

Zintzaras, E. & Ioannidis, J.P.A., 2008. Meta-analysis for ranked discovery datasets: theoretical framework and empirical demonstration for microarrays. Computational Biology and Chemistry, 32(1), 38-46.

Zusammenfassung

Trotz größerer Anstrengungen is der Alterungsprozess eines der am wenigsten verstandenen Phänomene der Biologie. Diese Arbeit bedient sich zweier bedeutenden Erkentnisse der Altersforschung: Zum einen der Schlussfolgerung, dass Veränderungen der Stammzell-Proliferation mit dem Alterungsprozess gekoppelt sein könnten, zum anderen dass Calorische Restriktion eine wirksame Maßnahme zur Verlängerung der Lebensspanne und zur Verzögerung alters-assoziierter Krankheiten darstellt. Im ersten Teil dieser Arbeit analysieten wir ein shRNA-basiertes Screening-Experiment um Gene zu identifizieren, die eine Rolle in der Stammzell-Proliferation spielen und unternahmen erste Schritte zur Etablierung eines Durchfluss-Cytometrie basierten Proliferations-Tests um Kandidaten zu validieren. Zweitens meta-analysierten wir Microarray-Daten aus verschiedenen Experimenten, die die Änderungen der Genexpression in Folge von Calorischer Restriktion untersuchten. Wir identifizierten mit Hilfe einer Binomial-Test basierenden Abzähl-Methode ("value counting approach") Kandidatengene, die hinsichtlich differentieller Expression in den Datensätzen angereichert waren. Wir zielten durch die Verwendung von Datensätzen von verschiedenen Organismen, Geweben, Altern, usw. darauf ab robuste und generalisierbare Kandidaten zu finden. Wir verwendeten ferner verschiedene Vorgehensweisen um den Kandidaten zugrunde liegende funktionelle Kategorien und Gemeinsamkeiten hinsichtlich ihrer Rolle in Signaltransduktions-Netzwerken zu detektieren. Im Ganzen überlappen die 163 gefundenen Kandidaten-Gene und 340 Kategorien mit früheren Erkenntnissen auf diesem Gebiet, wie zum Beispiel das Ghr Gen und Kategorien aus dem Bereich Lipid-Stoffwechsel, Insulin-Signalwege, Kollagen oder Immunität und suggerieren daher einen biologischen Bedeutunggehalt unserer Methode. Andererseits traten auch neue und bisher vernachlässigte Funktionen wie Fremdstoff-Metabolismus, Biorhythmus, Retinol-Metabolismus und Kupfer-Ionen-Entgiftung zum Vorschein, welche vielversprechende Gegenstände zukunftiger Forschung sein könnten. Einige der signifikanten Gene spielen möglicherweise eine tragende Rolle als Regulatoren wichtiger Signalwege, wie z.B. *Nfkbia, Airn* (Igf2R antisense RNA) und der Notch Co-Aktivator *Zfp64*.

Curriculum Vitae

Personal Details

- Michael Plank
- Schneitweger Str. 49, 93128 Regenstauf, Germany
- Email: michael84p@web.de
- Phone: 0043 6802124093
- Family status: Unmarried

Education

- Aug 2009 - Aug 2010: Diploma Thesis[8] at the School of Biological Sciences, University of Liverpool (co-supervised at the University of Vienna): Identification of candidate genes affecting stem cell proliferation by shRNA screening and caloric restriction by meta-analysis
- Since Oct 2006: Studies of Molecular Biology, University of Vienna
  - Specialization: Genetics, Cell Biology, Molecular Medicine
- Oct 2004-Jun 2006: Studies of Biochemistry, University of Regensburg (Vordiplom, grade: 2[9])
- May 2003: Abitur at Johann-Michael-Fischer-Gymnasium Burglengenfeld (grade: 1.2[10])

Civil Service

- Aug 2003-March 2004: Caritas Hospital St. Joseph, Regensburg

Relevant work experience

- Aug 2009 - Aug 2010: Diploma thesis in Dr de Magalhaes group, University of Liverpool; field: Aging research
- July-Aug 2009: Work placement in Prof Weckwerth´s group, University of Vienna; field: Quantitative mass spectrometry
- May-June 2009: Work placement in Prof Getoff´s group, University of Vienna; field: Free radicals and hormons

[8] the Diploma degree is equivalent to a Bachelor's + Master´s degree

[9] The Vordiplom is the first part of this Diploma study and corresponds roughly to a Bachelor-Degree; grades: 1 (very good) - 5 (failed)

[10] grades: 1 (very good) - 6 (failed)

- Apr-May 2009: Work placement in Prof Marian´s group, Medical University, Vienna; field: β-catenin and human adenomas
- March-Apr 2009: Work placement in Prof Martinez´s group, Institute for Molecular Biotechnology of the Austrian Academy of Sciences; field: miRNA processing
- Feb-March 2009: Work placement in Prof Barta´s group, University of Vienna; field: Splicing factors in plants
- July-Oct 2008: Work placement in Dr Leake´s and Dr Wadhams' groups, Department of Physics and Department of Biochemistry, University of Oxford, field: Single molecule microscopy
- Nov 2007-Apr 2008: Part-time technician in Prof Mittelsten-Scheid´s group, Gregor Mendel Institute, Vienna, field: Plant epigenetics
- July-Sept 2007: Work placement in Prof Mittelsten-Scheid´s group, Gregor Mendel Institute, Vienna
- May-Aug 2006: Student helper at the Institute for human genetics, University hospital, Regensburg, field: Mutation analysis
- July-Sept 2005: Work placement at "rent-a-scientist", Regensburg, field: Microbiology

Teaching experience

Co-supervision of a final year project with Dr de Magalhaes, University of Liverpool

Membership

Biochemical Society

Publication

Plank, M., Wadhams, G. H., & Leake, M. C. (2009). Millisecond timescale slimfield imaging and automated quantification of single fluorescent protein molecules for use in probing complex biological processes. Integrative Biology: Quantitative Biosciences from Nano to Macro, 1(10), 602-612. doi:10.1039/b907837a