Molecular BioSystems

Dynamic Article Links

Cite this: *Mol. BioSyst.*, 2012, **8**, 1339–1349

www.rsc.org/molecularbiosystems

PAPER

A meta-analysis of caloric restriction gene expression profiles to infer common signatures and regulatory mechanisms†

Michael Plank,[a] Daniel Wuttke,[a] Sipko van Dam,[a] Susan A. Clarke[ab] and João Pedro de Magalhães*[a]

Received 22nd June 2011, Accepted 18th January 2012
DOI: 10.1039/c2mb05255e

Caloric restriction, a reduction in calorie intake without malnutrition, retards age-related degeneration and extends lifespan in several organisms. CR induces multiple changes, yet its underlying mechanisms remain poorly understood. In this work, we first performed a meta-analysis of microarray CR studies in mammals and identified genes and processes robustly altered due to CR. Our results reveal a complex array of CR-induced changes and we re-identified several genes and processes previously associated with CR, such as growth hormone signalling, lipid metabolism and immune response. Moreover, our results highlight novel associations with CR, such as retinol metabolism and copper ion detoxification, as well as hint of a strong effect of CR on circadian rhythms that in turn may contribute to metabolic changes. Analyses of our signatures by integrating co-expression data, information on genetic mutants, and transcription factor binding site analysis revealed candidate regulators of transcriptional modules in CR. Our results hint at a transcriptional module involved in sterol metabolism regulated by Srebf1. A putative regulatory role of Ppara was also identified. Overall, our conserved molecular signatures of CR provide a comprehensive picture of CR-induced changes and help understand its regulatory mechanisms.

Introduction

Caloric restriction (CR) consists of reducing the caloric intake of organisms without malnutrition. CR can extend mean and maximum lifespan in a wide range of organisms from yeast to worms and flies to rodents.[1] Results from rhesus monkeys suggest that CR delays mortality.[2] In addition to extending lifespan, CR has been shown to delay signs of aging and the onset and progression of age-related diseases like cardiovascular disease, cancer, neurodegenerative diseases and diabetes.[1]

In mammals, CR induces numerous physiological alterations across organ systems, though it is not clear which of these changes are important for the life-extending effect of CR.[3] One important physiological change associated with CR is high insulin-sensitivity, which is particularly noteworthy since aging is generally accompanied by elevated insulin-resistance.[4] A reduction of body weight under CR has been observed and the tissue displaying most loss of weight is normally white adipose tissue.[4] This is accompanied by size-reduction of adipocytes in mice. Due to the negative correlation of fat mass to adiponectin levels the serum concentration of this hormone rises during CR.[5] This hormonal change comes along with increased fatty acid oxidation in fat tissue and reduced lipid accumulation in other tissues.[6] A reduction in inflammation has also been observed in CR animals.[7]

Decades of research have led to the identification of potentially important genes and pathways as effectors of CR. Decreased growth hormone/insulin/insulin-like signalling, decreased TOR, increased AMPK signalling, and increased activities of sirtuins are among the pathways linked to CR.[8] Several genes that disrupt or cancel out life-extending effects of CR have also been identified.[1,8] In spite of progress in physiological and genetic studies of CR, its underlying mechanisms remain a subject of debate. In particular, it remains unclear which mechanisms downstream of hormonal and energy metabolism alterations lead to lifespan extension. Since research on this topic so far has been largely knowledge-driven it seems logical to also approach CR by unbiased high-throughput studies. Several studies conducted so far used microarrays,[4,9] but a clear picture of CR mechanisms remains lacking.

Meta-analyses, quantitative syntheses of different studies on a subject, are used to increase the sample size and therefore the statistical power beyond that of individual studies.[10,11] Applied to CR, meta-analyses have already demonstrated that it is possible to identify significant transcriptional signatures of CR not found in individual studies.[11,12] These studies have also shown that CR can oppose some (but not all) age-related changes in gene expression, in particular at the level of genes

[a] *Integrative Genomics of Ageing Group, Institute of Integrative Biology, University of Liverpool, Biosciences Building, Room 245, Crown Street, Liverpool L69 7ZB, UK. E-mail: jp@senescence.info; Fax: +44 151 7954408; Tel: +44 151 7954517*
[b] *School of Biosciences, University of Exeter, Exeter, UK*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb05255e

associated with specific aging processes such as the age-related upregulation of genes involved in immune response.[11,12] Herein, we performed a meta-analysis of microarray studies of CR conducted in mammals, including studies in different organisms, strains, tissues, CR protocols, and microarray platforms. We expanded on previous meta-analyses of CR by employing a meta-analysis method developed specifically for studying aging,[10] by performing a CR meta-analysis across species and in an unprecedented number of datasets, and by employing several integrative and data-mining methods to identify candidate regulators of CR-induced changes.

Biological research is increasingly more reliant on integrating different types of data and studying pathways and networks as an integrated system to improve statistical power and decipher the underlying complex biological processes—the so-called 'systems biology paradigm'. Therefore, in addition to identifying genes robustly differentially expressed with CR, we performed functional enrichment analyses and integrated our results with co-expression data and data on genes previously shown to disrupt CR life-extending effects. To shed light on transcriptional modules underlying the differential expression of top genes we employed transcription factor binding site algorithms. Our results emphasize the importance of changes related to lipid (especially sterol) metabolism, and suggest a number of differentially expressed key regulators potentially responsible for the observed changes in transcriptional modules activated or repressed during CR.

Results

Detecting genes consistently over- and underexpressed in CR

To identify conserved molecular signatures of CR in mammals we first obtained 61 microarray datasets from publicly available databases: 48 from mouse, 12 from rat and 1 from pig (Table S1 in ESI†). In addition, 11 lists of genes differentially expressed in CR, including 1 dataset from rhesus monkeys, were obtained from the supplementary material of previous studies for which we lacked the raw data. Then we employed a value counting meta-analysis method[10,23] to detect genes consistently differentially expressed in CR. Succinctly, the number of datasets in which each gene was found differentially expressed according to a *t*-test *p*-value <0.05 and fold change of 1.5 was compared to the number of studies in which the respective genes were tested (see Materials and Methods). In other words, we identified genes more often over- and underexpressed in the microarray studies than expected by chance, akin to.[10]

Overall, 101 and 73 genes were found to be over- and underexpressed, respectively, in more datasets than expected by chance below a threshold of the binomial *p*-value of 0.0005 (FDR estimated by scrambling <0.05; see Materials and Methods for details). *Mt1a*, *Adh1* and *Per2* were the genes most significantly enriched for overexpression with *Slc6a6*, *Car3* and *Cyp2j5* enriched for underexpression. The 10 most significant over- and underexpressed genes are shown in, respectively, Tables 1 and 2 with full lists of genes in Table S2 (ESI†) and online (http://genomics.senescence.info/diet/).

Due to the large number of datasets from liver, we assessed whether genes may be found significant even though only differentially expressed in the liver. Indeed, from the genes enriched for over- and underexpression, 13% and 16%, respectively, were found only in the liver and 34% and 49% in less than three tissues (mainly in liver and one other tissue) (Fig. S3, ESI†). Due to the importance of the liver in regulating metabolism, we kept all liver-specific genes in subsequent analyses, but full results are given in the ESI† and online if researchers wish to redo our analyses.

Functional categories enriched for over- and underexpressed genes

Using Gene Ontology (GO) categories and repeating the meta-analysis at the level of GO-terms (see Materials and Methods),

Table 1 Top genes consistently overexpressed in CR

Gene symbol	Gene name	Function	Related candidate GOs	# Tissues[a]	Ref.[b]
Mt2	Metallothionein 2	Binds various metals	Cellular copper ion homeostasis	7	
Adh1	Alcohol dehydrogenase 1 (class I)	Metabolizes besides ethanol also retinol, *etc.*		5	
Per2	Period homolog 2 (Drosophila)	Master regulator of circadian clock	Circadian clock	6	13,14
Por	P450 (cytochrome) oxidoreductase	Transfers electrons from NADPH to among others P450 and heme oxygenase	Xenobiotic metabolism	4	15
Inmt	Indolethylamine *N*-methyltransferase	*N*-methylation of indoles (endogenous and xenobiotic)	Xenobiotic metabolism	4	
Dbp	D site albumin promoter binding protein	Transcription factor that modulates clock-output genes	Circadian clock	4	16
Nat8	*N*-Acetyltransferase 8 (GCN5-related, putative)	Not yet clear		3	
Ehhadh	Enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase	Part of the peroxisomal beta-oxidation pathway	Lipid metabolism	4	
Mt1	Metallothionein 1	Binds various metals	Copper ion binding	4	
Cyp2j6	Cytochrome P450, family 2, subfamily j, polypeptide 6	Arachidonic and linoleic acid and retinoid metabolism	Lipid metabolism, retinol metabolism	4	

[a] The number of different tissues in which the corresponding gene was found overexpressed. [b] Information about the genes was extracted from the references given or www.genecards.org.[17]

Table 2 Top genes consistently underexpressed in CR

Gene symbol	Gene name	Function	Related candidate GOs	# Tissues[a]	Ref.[b]
Slc6a6	Solute carrier family 6 (neurotransmitter transporter, taurine), member 6	Transports both taurine and beta-alanine		2	
Car3	Carbonic anhydrase 3	Catalyze the reversible hydration of carbon dioxide		Only in liver	
Cyp2j5	Cytochrome P450, family 2, subfamily j, polypeptide 5	Arachidonic acid epoxygenase	Lipid metabolism	2	
Dhcr7	7-Dehydrocholesterol reductase	Production of cholesterol by reduction of the C7–C8 double bond of 7-dehydrocholesterol	Lipid metabolism; cholesterol metabolism	3	
Arntl	Aryl hydrocarbon receptor nuclear translocator-like	Heterodimerize with Clock which together regulate Per1 and other clock output genes	Circadian clock	4	18
Zfp64	Zinc finger protein 64	Coactivator of Notch; regulates differentiation		4	19
Srebf1	Sterol regulatory element binding transcription factor 1	Transcription factor that regulates genes involved in sterol biosynthesis	Lipid metabolism, sterol metabolism	2	20–22
Es31	Esterase 31	Hydrolysis of esters and amide bonds; involved in detoxification of xenobiotics and maybe in lipid metabolism	Xenobiotic metabolism	2	
Gck	Glucokinase	Catalyzes the initial step of glucose utilization by the beta-cell and liver; effective when glucose is abundant		Only in liver	
Col15a1	Collagen, type XV, alpha 1	Structural protein, especially stabilizing microvessels and muscle cells		4	

[a] The number of different tissues in which the corresponding gene was found underexpressed; if only in one tissue the name of the tissue is given. [b] Information about the genes was extracted from the references given or www.genecards.org.[17]

Table 3 Top GO-terms enriched for overexpression in CR

GO-term[a]	GO ID	Total	# overexp.	*p*-Value
Lipid metabolic process	GO:0006629	8255	352	8.01×10^{-24}
Rhythmic process	GO:0048511	899	73	6.52×10^{-19}
Monooxygenase activity	GO:0004497	2803	147	8.69×10^{-18}
Circadian rhythm	GO:0007623	1025	72	2.15×10^{-15}
Detoxification of copper ion	GO:0010273	181	26	3.77×10^{-13}
Retinol metabolic process	GO:0042572	298	33	5.46×10^{-13}
NADPH-hemoprotein reductase activity	GO:0003958	149	22	1.34×10^{-11}
Acyl-CoA metabolic process	GO:0006637	749	51	6.73×10^{-11}
Oxidoreductase activity	GO:0016491	20 263	630	1.21×10^{-10}
Nitric oxide mediated signal transduction	GO:0007263	307	30	1.35×10^{-10}

[a] Very broad and cellular component GO-terms are not shown.

187 and 153 GO-terms were found to be enriched for genes over- and underexpressed, respectively. A selection of non-redundant top GO-terms ($p < 0.0005$; FDR estimated by scrambling <0.05) enriched for genes over- and underexpressed is shown, respectively, in Tables 3 and 4. The full results for GO-terms can be found in Table S3 (ESI†) as well as online (http://genomics.senescence.info/diet/).

The top GO category for overexpressed genes was "lipid metabolic process" and other, more specific GO-terms related to lipid metabolism like "acyl-CoA metabolic process" or "fatty acid metabolic process" were found. For underexpressed genes the top GO-term was "sterol biosynthetic process" with "cholesterol biosynthetic process" and "lipid biosynthetic process" among the top GO-terms. Interestingly, "response to sterol depletion" was significant. Note that also changes related to steroid hormones appear among the significant GO-terms for upregulated genes.

In addition, among the most significant GO-terms for upregulated genes were "rhythmic process" and "circadian rhythm" (Table 3). The latter was also found to be significant for downregulated genes. As expected, several categories related to immune response were found for downregulated genes. Furthermore, categories related to collagen were associated with both over- and underexpressed genes. In line with the hormesis hypothesis that CR invokes a stress response that induces survival in cells, upregulation of "response to stress" genes was also observed. The findings of "growth hormone receptor activity" and "growth hormone receptor signalling 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 support the involvement of the growth factor and insulin/IGF signalling pathways in CR.

"Retinol metabolism", which was found to be enriched for upregulated genes, had been linked to CR in a broader sense

Table 4 Top GO-terms enriched for underexpression in CR

GO-term[a]	GO ID	Total	# underexp.	*p*-Value
Sterol biosynthetic process	GO:0016126	1091	59	5.57×10^{-10}
Beta-alanine transmembrane transporter activity/taurine:sodium symporter activity	GO:0001761	60	12	6.32×10^{-9}
Cholesterol biosynthetic process	GO:0006695	1022	53	1.59×10^{-8}
Innate immune response	GO:0045087	3356	125	1.80×10^{-8}
Response to sterol depletion	GO:0006991	68	12	2.80×10^{-8}
Steroid biosynthetic process	GO:0006694	2298	93	2.97×10^{-8}
7-Dehydrocholesterol reductase activity	GO:0047598	49	10	9.43×10^{-8}
Response to virus	GO:0009615	1706	73	1.06×10^{-7}
Positive regulation of transcription *via* sterol regulatory element binding	GO:0035104	92	13	1.17×10^{-7}
Pheromone binding	GO:0005550	164	17	1.52×10^{-7}

[a] Very broad and cellular component GO-terms are not shown. Different GO-terms related to taurine and beta-alanine transport significant due to the underexpression of one gene (Slc6a6) in 12 datasets were collapsed.

by a study reporting the decrease of retinol during fasting in humans.[24] Also changes in nitric oxide mediated signalling, a category enriched for upregulated genes, were previously associated with CR; it was found that calorically restricted mice show overexpression of eNOS (endothelial nitric oxide synthase) and its knock-out largely reduced CR-induced effects.[25] Of the top 10 categories enriched for overexpressed genes, to our knowledge no functional link has been reported between CR and "copper ion detoxification".

In a complementary approach, we used the DAVID functional analysis tool to determine the overrepresented functional categories in the genes significantly enriched for over- and underexpression. The results obtained in this way (Benjamini–Hochberg corrected FDR < 0.05; note that FDR estimation by scrambling is not feasible with this tool) represent subsets of those obtained by the meta-analysis at the functional level (see Table S3, ESI†). The significant functional clusters found for overexpressed genes were related to sulfotransferase-activity, NAD(P)+ involving processes, oxidoreductases and biological rhythms and for underexpressed genes associated with the endoplasmic reticulum. One significant Biocarta pathway, "Nuclear Receptors in Lipid Metabolism and Toxicity", and three KEGG[26] pathways, "PPAR signaling pathway", "Arachidonic acid metabolism" and "Retinol metabolism in animals" were found for genes overexpressed; none were found for genes underexpressed in CR.

Inferring regulatory modules from CR signatures

Co-expression analysis of CR-associated genes. To understand regulatory modules underlying the observed differential expression patterns, we employed a transcriptome-wide co-expression map to extract genes more strongly co-expressed with genes over- and underexpressed during CR than expected by chance (see Materials and Methods). Genes identified in this way may serve as further candidates involved in CR and combining them with differentially expressed genes improves the statistical power for the detection of underlying functional categories. Moreover, transcription factors (TFs) among co-expressed genes that are themselves differentially expressed in the same direction as the genes they are associated with may be candidate transcriptional regulators.

A large number of genes (1075 for over- and 410 for underexpression) were found to be co-expressed and we performed a functional enrichment analysis using DAVID. Interestingly, we found that many of the genes that are co-expressed with the upregulated genes overlap with the genes that are co-expressed with the downregulated genes (Fig. S4, ESI†), and that functional categories obtained for upregulated genes were by and large the same as for downregulated genes. Some of the most significant functional categories retrieved for both up- and downregulated genes were related to lipid metabolism, circadian clock, inflammation/immunity, steroid/sterol/cholesterol metabolism, response to hormones, mitochondria, and xenobiotic metabolism/cytochrome P450 (Table S4, ESI†). This could indicate that as particular pathways and processes are altered by CR, some of its genes are overexpressed while others are underexpressed.

To identify possible regulators of transcriptional changes with CR we looked for TFs among co-expressed genes. Interestingly, Ppara, the TF most significantly co-expressed with overexpressed genes, is itself a member of the genes enriched for overexpression. Similarly, Dbp, Klf17 and Nr1i3 were found to be both enriched for overexpression and co-expressed with overexpressed genes. Of the 17 TFs co-expressed with downregulated genes, 14 were also among the 42 co-expressed with upregulated genes. However, one of the TFs which was solely co-expressed with the down-regulated seed list, Irf7 (interferon regulatory factor 7), was the only one to be also underexpressed itself. Lists of genes co-expressed with genes enriched for over- and underexpression are shown in Table S5 (ESI†).

Transcription factors regulating expression of differentially expressed genes

To gain a better understanding of which genes regulate and are regulated in response to CR, we analyzed transcription factor binding sites (TFBS) in the signatures. The complete list of TFs with binding sites enriched ($p < 0.01$) in the [−500, 100 bp] region of the genes determined by the meta-analysis according to the Biobase ExPlain tool is shown in Table S6 (ESI†). Overall, we found 29 TFs with binding sites enriched among overexpressed genes and 40 TFs among underexpressed genes. In line with results from our co-expression analysis, some TFs were found to be enriched among both over- and underexpressed genes (Table S6, ESI†). It is also noteworthy that Foxa TFs (Foxa1, Foxa2 and Foxa3) were significant among overexpressed genes (Table S6, ESI†) as Foxa has been linked to CR.[27] We then looked for TFs overlapping with significant genes from the meta-analysis as these are more likely to be biologically important. The only TF with its binding site

enriched that was also significantly enriched for underexpression itself was Srebf1. For genes enriched for overexpression, we only found overlapping TFs with binding sites enriched at the more relaxed cutoff of $p < 0.2$ and these were Nr1i3 and Ppara, two TFs co-expressed with overexpressed genes.

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

Top hits were evaluated in light of previous knowledge by searching for overlaps with genes that alter the life-extending effect of CR when mutated. This list of CR-essential genes is largely based on homologs from non-mammalian model organisms (see Materials and Methods) since the only mouse gene known to be essential for CR-induced lifespan extension is Ghr, the growth hormone receptor.[28]

Four of our hits have significant sequence similarity to CR-essential genes in lower model organisms. *Sc5d* (sterol-C5-desaturase), enriched for underexpression, may be a functional homologue of the CR-essential gene *ERG3* in *Saccharomyces cerevisiae*. *Gck* (glucokinase), enriched for downregulation, is homologous to CR-essential *HXK2* in *S. cerevisiae*. *Mat1a* (methionine adenosyltransferase I, alpha), enriched for upregulation, is homologous to *sams-1* in *Caenorhabiditis elegans*. *Irs2* (insulin-receptor substrate 2), enriched for upregulation, displays homology to *chico* in *Drosophila melanogaster*.[29] *Sc5d* and *Gck* appear to be functional homologs, while in the case of *Mat1a* and *Irs2* a functional conservation is not proven.[29–32]

In addition, 42 of our hits were direct protein interaction partners of murine CR-essential gene orthologues, which were significantly enriched for "circadian rhythm", "oxido-reductase", "glucose metabolic process" and "positive regulation of apoptosis"-related terms when using DAVID. Among these 42 proteins that interact with CR-essential genes are Ppara and Srebf1. The complete list of these genes with their specificity measure and *p*-values is shown in Table S7 (ESI†). Moreover, three of the genes in the network of CR-essential genes overlapping with the meta-analysis hits have been implicated in aging according to the GenAge database:[33] the genes are *Ghr*, *Irs2* and *Arntl*, the latter being an important circadian clock transcription factor.[18]

Discussion

A meta-analysis approach to CR gene expression profiles

CR is the most promising non-genetic intervention to extend lifespan and delay aging-associated diseases in a range of organisms. To understand the genetic and molecular basis of CR in a data-driven way we determined robust changes in gene expression linked to CR by meta-analyzing microarray data in a variety of experimental variables, *e.g.* species, tissue, age, duration and extent of CR, diet composition and microarray platform. We employed a value counting approach for our meta-analysis that emphasizes sensitivity and has been previously shown to outperform other methods in the context of aging studies.[10] Another advantage of using a value counting approach is that we could include datasets for which only lists of differentially expressed genes were available.[10,23] The inclusion of such lists of genes increases our signal and number of significant genes, though it does not significantly alter our results; full results are provided in the ESI† with (Tables S2 and S3) and without (Tables S8 and S9) additional lists of genes.

In addition to the meta-analysis method, our study is also unique in that, contrary to some previous works, we excluded genes we suspected were only found differentially expressed with CR in old animals due to the lack of the normal expression change with age (see Materials and Methods), as we reason that these are more likely to be an effect than a cause for changes induced by CR (Fig. S2, ESI†). Moreover, our study includes a larger number of datasets than previous CR meta-analyses and, to our knowledge, is the first to focus on data from mammals other than mice, though our study still had a high prevalence of mouse datasets. Lastly, the integration of additional types of data, and in particular information on genetic mutants that we systematically collected, means that our study is in a unique position to identify key regulators of CR.

Interestingly, 10 genes overexpressed and 9 underexpressed from our meta-signature have been validated experimentally in different tissues of mice and rats, mostly by direct measurement of mRNA levels by qRT-PCR (Table S10, ESI†). Of note, both Ppara and Srebf1 TFs have been validated in accordance with our predictions.[5] This demonstrates that our method can detect biologically meaningful results.

Genes and functional categories overrepresented for differential expression

Our meta-analysis of microarray data reveals candidate genes and functional categories for a role in CR. The fact that we re-discovered a large number of genes and categories already related to CR further demonstrates that our method can detect biologically-relevant findings and supports the role of these genes as a conserved signature of CR. Besides their use as biomarkers, the novel genes and processes detected could serve as new foci for future studies. 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, it seems safe to assume that out of these genes those found to be over-/underexpressed in at least three different tissues are truly tissue-independent. The complete results are provided in Tables S2 and S3 (ESI†) as well as online (http://genomics.senescence.info/diet/).

The presence of many lipid metabolism and sterol biosynthesis related GO-terms among the ones with highest significance fits well with the idea of different metabolic states of *ad libitum* (AL) level and CR animals and is in agreement with previous meta-analyses on CR.[12] It confirms the idea that there is a hormone-driven shift from glycolysis to lipolysis and gluconeogenesis in CR. There is a large literature linking lipid metabolism to CR (for a review see ref. 34). It has also been reported that CR prevents age-related changes in cholesterol metabolism.[35] Moreover, one of the major effects of CR is the repression of immune functions and an important physiological change with aging is increased inflammation and alterations in collagen deposition. It has been shown previously that CR prevents to a certain degree collagen accumulation and collagen aging.[36] Therefore it is noteworthy that our meta-analysis also

established relations between CR and these functional categories.

Our functional analysis detected categories related to the growth hormone and insulin/IGF-signalling pathways, mutations in which have effects on longevity and the lifespan-extending effect of CR. *Ghr*, enriched for underexpression in our analysis, is the only known mouse gene that when mutated favours longevity and cancels out the lifespan extending effect of CR.[28] This argues for the biological meaningfulness of our results. Another gene enriched for underexpression is *Airn* (antisense Igf2r RNA) which might function as an important ncRNA in the regulation of insulin/IGF-signalling. Note that this gene until recently was annotated as a RIKEN cDNA gene and that therefore other of our top genes with unknown function might also be promising new candidates for further studies. *Irs2* (insulin-receptor substrate 2) was found to be enriched for overexpression and is a homologue of *chico* in *D. melanogaster*, which was experimentally associated with aging and CR.[29] An association between hormonal alterations involving insulin/insulin like growth factor 1 (IGF1)/growth hormone (GH) and aging is well-known and these are likely involved in CR mechanisms.

We detected categories related to circadian rhythm and monooxygenase activity/xenobiotic metabolism both for over- and underexpressed genes. Although both these processes have been associated with CR[37–40] a deeper understanding of their role in CR remains elusive. The results for circadian rhythm changes in CR are in agreement with previous observations in mice (for an overview see ref. 41). Two of our top genes, *Arntl* and *Dbp*, are important circadian clock TFs of which the former has already associated with the aging process.[18] Alterations of the circadian rhythm are often considered side effects of CR, caused by changed times of food consumption. However it cannot be ruled out that this effect is at least partially causal for the life extending effect of CR. Notably, also the circadian clock master regulator *Per2* was among the genes most significantly enriched for overexpression. Transcriptional levels of *Per2* oscillate diurnally in the suprachiasmatic nucleus (SCN) of the hypothalamus and are supposedly set by light.[13] The timing of oscillators in peripheral tissues is controlled by the SCN when food is available AL. If feeding is, however, temporally limited, the time of feeding is a more important regulator for peripheral oscillators.[14] If additionally the level of food intake is altered also the timing of clock gene expression in the SCN changes, possibly impacting metabolic regulation. Therefore both the changed amount of food, and also the fact that CR might change the timing of food availability compared to AL might have an important influence on altered expression levels of clock genes.

To our knowledge no reports on copper ion detoxification exist in respect to CR, which we found among the categories most significantly enriched for genes overexpressed with CR. Another process less well established as to its role in CR is retinoic acid/retinol metabolism. Also note that 9-retinoic acid is a ligand of RXR which forms a complex with peroxisome proliferator activators (PPARs) and therefore establishes a link between retinol metabolism and regulation of lipid metabolism.[26] It is interesting to note that anti-inflammatory effects of retinoic acid are known.[42]

By bringing together, analyzing and interpreting as a whole multiple large-scale gene expression studies using CR animals, our work provides an integrated picture of CR-induced alterations. Fig. 1 summarizes our CR-induced signatures together with current theory of CR-induced physiological changes.

Transcriptional regulation of CR-induced changes and candidate regulators

In addition to identifying processes and genes altered during CR, our goal was to identify candidate regulators of CR-induced changes, for which we employed a whole-genome co-expression map, data on gene mutations that disrupt CR effects plus their interaction partners and TFBS analysis.

Srebf1, a TF regulating sterol metabolism, is a candidate for regulating CR-suppressed gene expression changes. Notably this was not just one of the genes most significantly enriched for underexpression, but also the only TF with binding sites enriched upstream of genes enriched for underexpression. Among the Srebf1 regulated genes are CR-essential orthologous *Gck*[43] and *Sc5d*[44,45] as well as ATP citrate lyase (*Acly*)[46] all of which are themselves enriched for downregulation in our results. Gck is an orthologue of HXK2 in *S. cerevisiae* which catalyses the phosphorylation of glucose prior to glycolysis and was previously associated with CR.[30] Likewise, *Sc5d* (sterol-C5-desaturase), another gene enriched for downregulation, encodes an enzyme involved in sterol metabolism and is a homologue of *ERG3*, which is important for lifespan extension by CR in *S. cerevisiae*.[35] Acly is the primary enzyme responsible for the generation of cytosolic Acetyl-CoA out of citrate, which was found to be crucial in CR-induced metabolic changes.[47] Interestingly, Srebf1 expression is induced by insulin, *via* the activation of the phosphatidylinositol 3-kinase (PI3K) pathway, in hepatocytes.[43] In *Drosophila*, Srebf1 is induced by Akt activation but this is blocked by glucose starvation, inhibition of glycolysis, or AMPK activation and requires TOR activity.[48] Moreover, silencing Srebf1 in flies was shown to block the induction of cell growth by dPI3K. CR could, therefore, reduce TOR signalling and/or reduce insulin/IGF1 signalling and Akt activity with a resulting reduced nuclear accumulation of Srebf1. Srebf1 is also involved in lipogenesis across metazoans and its activation has been suggested to contribute to human metabolic disorders.[49] These results place Srebf1 at the heart of CR-induced signalling and warrant further studies, for example of whether silencing Srebf1 (*e.g.*, in flies) induces CR responses.

Another gene that called our attention was *Ppara*, a member of the steroid hormone receptor superfamily and a TF that plays a key role in the regulation of lipid metabolism.[21] *Ppara* was enriched for overexpression in CR animals, co-expressed with genes enriched for overexpression and interacting with CR-essential orthologs. Ppara binding sites were significantly enriched in the promoter of upregulated genes, even though only at FDR < 0.2. Our results therefore suggest that an alteration of steroid hormone signalling and the effect of this alteration on cells is an important mechanism of CR with Ppara as a candidate regulator.

Lastly, a TF significantly co-expressed with genes enriched for both over- and underexpression, Cebpa, is a regulator of lipid metabolism[50] even though not itself significantly enriched for up- or downregulation. A summary of top hits with strong evidence for a role in CR in this study is depicted in Fig. 2 with

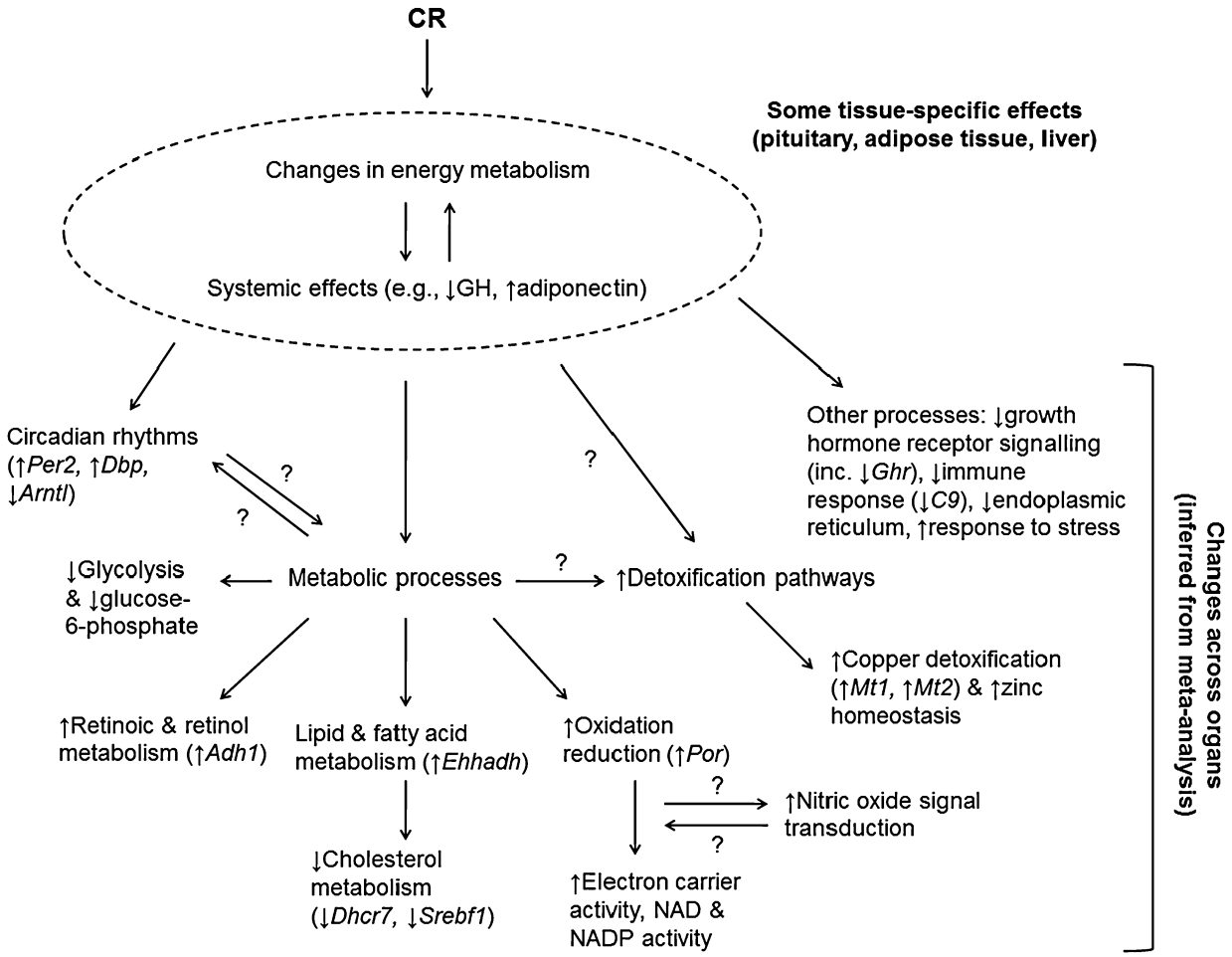

Fig. 1 Overview of CR-induced changes based on the results from our meta-analysis integrated with current theory. It is widely accepted that CR induces systemic changes at the level of various hormones, and growth hormone (GH) in particular. These changes may be due to the impact of CR on energy metabolism and its interplay with tissues responsible for endocrine changes, such as the pituitary where GH is produced. Systemic effects then drive multiple changes that tend to be conserved across different organs. Displayed are the major alterations across organs inferred from our meta-analysis. One crucial axis of CR-induced changes involves alterations in metabolic processes. These changes in metabolism appear in turn to trigger mitochondrial changes at the level of, for instance, oxidation/reduction and electron transport chain. Some important doubts remain, for example in regards to the specific mechanisms triggering the activation of detoxification pathways and nitric oxide signal transduction.

suggestive (albeit not comprehensive) links to other CR-related players and pathways.

Concluding remarks and future prospects

Overall, our meta-analysis allowed us to obtain conserved molecular signatures of CR that may be used as biomarkers of CR as well as help identify genes and pathways mediating CR life-extending effects. Our results emphasize that changes to metabolism play an important role in CR, which to some extent might be attributed to hormonal alterations, *e.g.* of adiponectin, GH, insulin/IGF and glucagon levels, described previously.[4,5] Based on our comprehensive picture of CR-induced changes we suggest a model for CR-induced changes (Fig. 1). Furthermore, the results of this meta-analysis hint at an involvement of circadian rhythm pathways in CR.

One important aspect of our study is that we integrate additional types of data to gain insights into transcriptional regulation under CR. In particular, we suggest the regulation of modules of differential expression by steroid hormone receptors. Other examples of TFs potentially central to CR are given (Fig. 2). We believe that further studies of our top genes and the discovery of their relation to hormonal and other physiological changes in CR will significantly contribute to the understanding of CR.

Materials and methods

Microarray studies used in the meta-analysis

Microarray data on CR were obtained from the Gene Expression Omnibus (GEO),[51] ArrayExpress[52] and Gene Aging Nexus (GAN).[53] Non-deposited datasets were also requested from authors. Only data from mammals were employed, as these are more likely to be relevant to humans. Datasets were not used if the experiment was accompanied by the application of drugs or infection of the animals. Only microarray platforms that were an unbiased representation of the transcriptome were used.

For 21 studies expression datasets could be obtained (Table S1, ESI†). That means the preprocessed (*i.e.* background subtracted

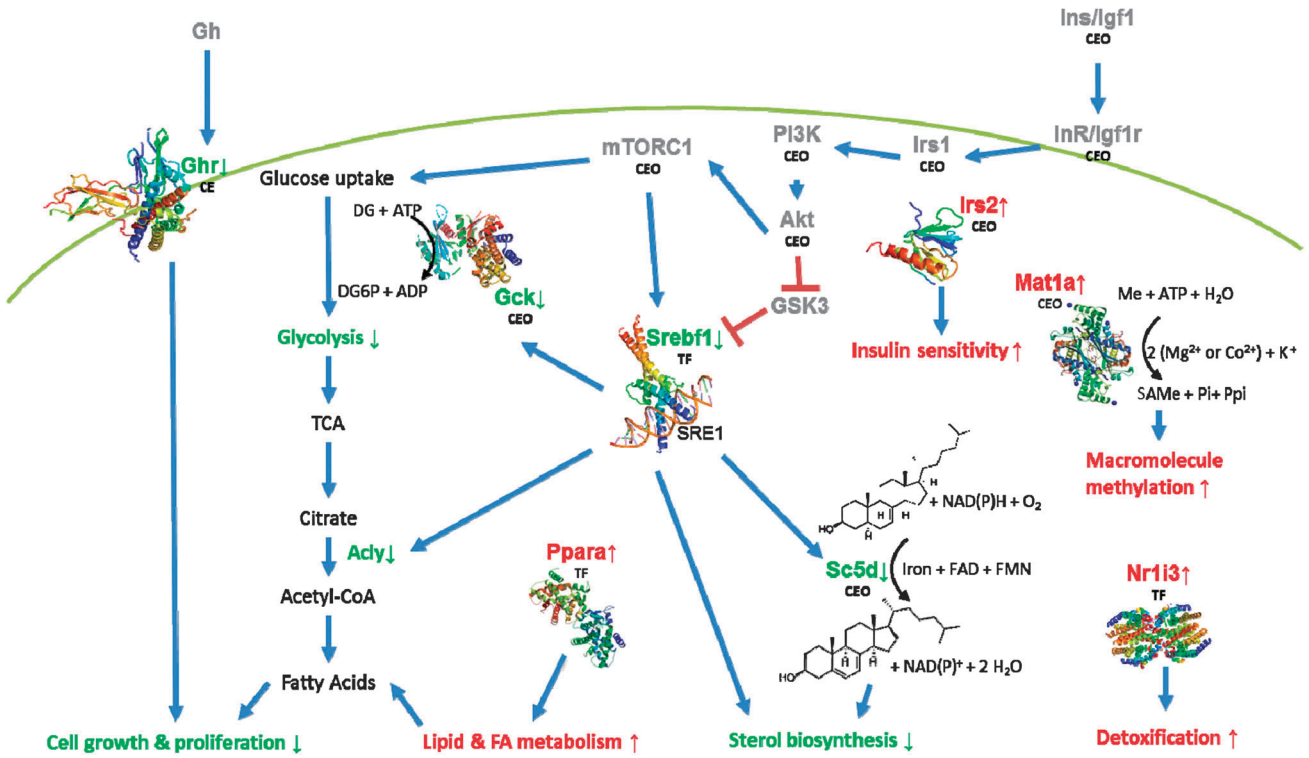

Fig. 2 Summary of genes enriched for differential expression with CR, the importance of which is emphasized by the fact that they are either CR-essential (CE), orthologues of CR-essential genes (CEO) or transcription factors with binding sites enriched in up- or downregulated genes (TF). Important downstream functions are depicted; green indicates enrichment for underexpression while red represents enrichment for overexpression. Ghr = Growth hormone receptor; Gck = Glucokinase; Irs2 = Insulin receptor substrate 2; Sc5d = Lathosterol oxidase (sterol-C5-desaturase); Srebf1 = Sterol regulatory element-binding protein 1 (sterol regulatory element-binding transcription factor 1); Ppara = Peroxisome proliferator-activated receptor alpha; Mat1a = *S*-adenosylmethionine synthase isoform type-1 (methionine adenosyltransferase 1); Nr1i3 = Nuclear receptor subfamily 1, group I, member 3; DG = D-Glucose; DG6P = D-Glucose 6-Phosphate; Me = Methionine; SAMe = *S*-Adenosyl-L-Methionine; Pi = Phosphate, ionic; PPi = diphosphate, ionic; FA = Fatty acid; FAD = Flavin Adenine Dinucleotide; FMN = Flavin Mononucleotide; SRE1 = Sterol Regulatory Element-1. The structure of Irs1 is shown instead of Irs2. Note that the diagram is not comprehensive and thus only suggestive and there may be other important factors that are not shown.

and normalized) microarray signals for the conditions of interest were given for all probes on the array except when excluded for low quality. From each study one to fifteen datasets were extracted. Data in each dataset consisted of *ad libitum* (AL) and CR samples from animals of the same age, CR setup and the same tissue. All studies employed single-color chips. The only co-variate for which we did not split data into different datasets was sex, since we did not want to reduce replicate numbers of each dataset more than necessary. This yielded a total of 61 datasets. Of these, 48 were from mouse (*Mus musculus*), 12 from rat (*Rattus norvegicus*) and 1 from pig (*Sus scrofa*). In total, datasets came from 19 different tissues with the most frequent tissue being the liver. 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. The reduction in calorie intake in the CR group was in most cases between 10 and 40% (for details see Table S1, ESI†).

Eleven lists of genes differentially expressed with CR according to the statistical procedure used in the original study were created for which the raw microarray data were not available.[54–58] Among these was one study focussing on rhesus monkeys. The distribution of the number of datasets over different co-variates after including these gene lists is shown in Fig. S1 (ESI†).

Processing datasets

To be able to integrate the different datasets we annotated the probes in all datasets with their corresponding mouse Entrez ID; since most datasets were from mouse and Entrez IDs were expected to facilitate matching between homologous genes of different organisms. Matching tables between different identifiers were obtained from Ensembl Genes 57.[59] If a probe matched to more than one Entrez ID it was represented by a list of all of them. Probes targeting transcripts of the same gene (and *i.e.* having the same Entrez ID) were collapsed by using the mean over each probe. However, if a given probe was mapped to more than one Entrez ID we discarded its effects on a given Entrez ID for which other probes existed which only mapped to it. For organisms other than mouse the Entrez ID of the homologous mouse gene according to HomoloGene (08/2009)[60] was used. Non-mouse genes homologous to more than one mouse gene were discarded. Probes not matching mouse Entrez IDs were lost during this procedure, which was especially a problem for poorly annotated *Sus scrofa* genes.

In each individual microarray, probes that contained more than 30% missing values or for which no Entrez ID annotation was found were eliminated. All remaining missing values were

replaced using the row average method. The datasets were processed using customized Perl and R scripts.[61]

Since CR is a mechanism that counteracts the effects of aging, some of the gene expression changes induced by CR in older organisms when compared to age-matched AL animals may be due to the retardation of changes normally occurring with age. Since our work focused on the mechanisms of CR, we removed all genes found to be differentially expressed between older and younger AL animals from the genes differentially expressed with CR in the opposite direction in the older animals, if possible. The number of genes excluded from each dataset in this procedure was generally $<5\%$.

Statistical analyses

The datasets were meta-analyzed using the following value counting approach (see Ramasamy *et al.*[23] for an overview of meta-analysis approaches): for each gene in each study, the *p*-value of an unpaired Student's *t*-test was calculated between CR and AL samples. As an effect size measure we calculated the fold change by dividing the mean of CR by the mean of AL values. 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 (n) and in how many it was found over- or underexpressed (k) at a *p*-value of <0.05 and a fold-change of at least 1.5 or lower than 1/1.5. For the two datasets with only one AL and one CR sample, where no *t*-test could be performed, a stricter effect size threshold was applied. For studies for which only lists of genes were available only the effect-size threshold was applied, where possible.

We obtained the probabilities of finding a gene over- or underexpressed in the observed number of studies from the cumulative binomial distribution:

$$P = \sum_{x=k}^{n} \binom{n}{x} \cdot p_s^x \cdot (1 - p_s)^{(n-x)}$$

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.

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 false discovery rate (FDR) estimate. A cutoff of $p = 0.0005$ corresponds to an FDR of about 0.05 for both over- and underexpressed genes.

An overview of the meta-analysis approach is given in Fig. 3.

Functional enrichment analyses

In an analogous approach to that described above, we compared the number of times a Gene Ontology (GO) category was found associated with an over- or underexpressed gene in the datasets to the number of times it is found associated with any gene. A file mapping each Entrez ID to the corresponding GO-IDs was downloaded from the NCBI FTP (ftp://ftp.ncbi.nih.gov/gene/DATA/gene2go.gz; 07/2010).

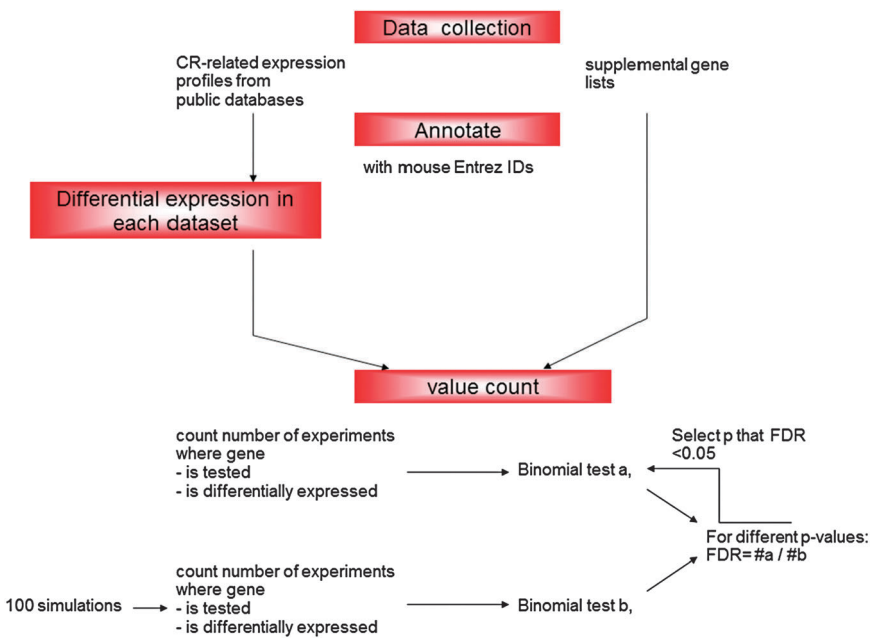

Fig. 3 Overview of the meta-analysis approach. Expression profiles were obtained from public databases and genes differentially expressed in each dataset selected by a Student's *t*-test. Further lists of genes differentially expressed with CR were obtained from original publications and added to the datasets. Genes were annotated with mouse Entrez IDs. The number of datasets in which a gene was differentially expressed (separately for over- and underexpression) was compared to the number of datasets in which the gene was tested and a binomial test used both on the original data and in 100 simulations using scrambled data. The FDR was calculated as the number of genes found on the original data (#a) divided by the mean number on simulated data (#b) at the corresponding *p*-value. The cutoff *p*-value for significant genes was chosen so that the corresponding FDR was below 0.05.

The Database for Annotation, Visualization and Integrated Discovery (DAVID)[62] was also employed to detect enriched functional categories in the top over- and underexpressed genes.

Transcription factors regulating expression of differentially expressed genes

To determine transcription factor binding sites enriched in the [−500, 100 bp]-region of top differentially expressed genes we employed Biobase ExPlain.[63] The F-Match tool was used with a cutoff of $p < 0.01$ and otherwise default conditions. Transcription factors (TFs) corresponding to these binding sites were extracted. We determined the overlap of these TFs with the top hits from the meta-analysis. To identify further potential regulators we also compared TFs with binding sites enriched at the more relaxed cutoff of $p < 0.2$ with the top genes.

Co-expression analysis of CR-associated genes

A whole-genome mouse co-expression map from 1678 experiments was employed that assigns expression similarity scores between pairs of genes (van Dam *et al*., in preparation). The top 5% of genes with highest similarity for each gene were considered co-expressed with a query 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 each gene over- or underexpressed with CR compared to the number of times it was co-expressed with all mouse genes. A binomial test was performed 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 the number of times g_i was co-expressed with any gene divided by the number of all genes.

Genes were ranked by their p-values from the binomial test and FDR values were calculated using a Bonferroni correction by multiplying the p-values by the number of tests, in this case corresponding to the 20 677 genes in the co-expression map. Genes with a p-value of $<10^{-6}$ (FDR < 0.02) were considered significant. Transcription factors among these genes were extracted as they are potential upstream regulators of the differentially expressed genes. DAVID was employed under default settings to detect functional enrichment.

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

Over 100 genes experimentally identified to be essential for the effect of CR to induce lifespan extension in different model organisms were retrieved from the literature (Wuttke *et al*., in preparation) and mouse orthologues derived. A CR-essential gene is defined as a gene in which genetic manipulation significantly reduced or canceled the life-extending effect of CR. A network around CR-essential genes was built according to information on physical protein–protein and genetic interactions retrieved from multiple databases, extended by direct interaction partners and analyzed using Cytoscape.[64] The creation of such networks and definition of CR-essential genes will be described in another work (Wuttke *et al*., in preparation). Interaction partners significantly overlapping with results of the meta-analysis were extracted.

Acknowledgements

We thank Dr Brian Merry and Dr Brian Lane for their helpful advice and Prof. Hu for sharing microarray data with us. The work of DW and MP was supported by the Erasmus programme. DW is also thankful for support from the Bundesministerium für Bildung und Forschung. SVD is funded by a BBSRC grant (BB/H008497/1) to JPM. JPM is also grateful for support from the Ellison Medical Foundation and from a Marie Curie International Reintegration Grant within EC-FP7.

Notes and references

1 L. Fontana, L. Partridge and V. D. Longo, *Science*, 2010, **328**, 321–326.
2 R. J. Colman, R. M. Anderson, S. C. Johnson, E. K. Kastman, K. J. Kosmatka, T. M. Beasley, D. B. Allison, C. Cruzen, H. A. Simmons, J. W. Kemnitz and R. Weindruch, *Science*, 2009, **325**, 201–204.
3 J. Koubova and L. Guarente, *Genes Dev*., 2003, **17**, 313–321.
4 R. M. Anderson and R. Weindruch, *Trends Endocrinol. Metab*., 2010, **21**, 134–141.
5 M. Zhu, J. Miura, L. X. Lu, M. Bernier, R. DeCabo, M. A. Lane, G. S. Roth and D. K. Ingram, *Exp. Gerontol*., 2004, **39**, 1049–1059.
6 M. Zhu, G. D. Lee, L. Ding, J. Hu, G. Qiu, R. de Cabo, M. Bernier, D. K. Ingram and S. Zou, *Exp. Gerontol*., 2007, **42**, 733–744.
7 R. M. Anderson, D. Shanmuganayagam and R. Weindruch, *Toxicol. Pathol*., 2009, **37**, 47–51.
8 N. A. Bishop and L. Guarente, *Nat. Rev. Genet*., 2007, **8**, 835–844.
9 E.-S. Han and M. Hickey, *J. Nutr*., 2005, **135**, 1343–1346.
10 J. P. de Magalhães, J. Curado and G. M. Church, *Bioinformatics*, 2009, **25**, 875–881.
11 W. R. Swindell, *BMC Genomics*, 2009, **10**, 585.
12 S.-E. Hong, H.-S. Heo, D. H. Kim, M.-S. Kim, C. H. Kim, J. Lee, M.-A. Yoo, B. P. Yu, C. Leeuwenburgh and H. Y. Chung, *Age (Dordrecht, Neth.)*, 2010, **32**, 15–30.
13 E. Waddington Lamont, V. L. Harbour, J. Barry-Shaw, L. Renteria Diaz, B. Robinson, J. Stewart and S. Amir, *Neuroscience*, 2007, **144**, 402–411.
14 M. Girotti, M. S. Weinberg and R. L. Spencer, *Am. J. Physiol.: Endocrinol. Metab*., 2009, **296**, E888–E897.
15 M. Kranendonk, C. C. Marohnic, S. P. Panda, M. P. Duarte, J. S. Oliveira, B. S. S. Masters and J. Rueff, *Arch. Biochem. Biophys*., 2008, **475**, 93–99.
16 H. Saito, T. Terada, J. Shimakura, T. Katsura and K.-I. Inui, *Am. J. Physiol.: Gastrointest. Liver Physiol*., 2008, **295**, G395–G402.
17 M. Safran, I. Dalah, J. Alexander, N. Rosen, T. Iny Stein, M. Shmoish, N. Nativ, I. Bahir, T. Doniger, H. Krug, A. Sirota-Madi, T. Olender, Y. Golan, G. Stelzer, A. Harel and D. Lancet, *Database*, 2010, **2010**, baq020.
18 R. V. Kondratov, A. A. Kondratova, V. Y. Gorbacheva, O. V. Vykhovanets and M. P. Antoch, *Genes Dev*., 2006, **20**, 1868–1873.
19 K. Sakamoto, Y. Tamamura, K. ichi Katsube and A. Yamaguchi, *J. Cell Sci*., 2008, **121**, 1613–1623.
20 Y. Wang, T. Joshi, X.-S. Zhang, D. Xu and L. Chen, *Bioinformatics*, 2006, **22**, 2413–2420.
21 J. D. Mulligan, A. M. Stewart and K. W. Saupe, *Exp. Gerontol*., 2008, **43**, 146–153.
22 E. Stelmanska, J. Korczynska and J. Swierczynski, *Acta Biochim. Pol*., 2004, **51**, 805–814.
23 A. Ramasamy, A. Mondry, C. C. Holmes and D. G. Altman, *PLoS Med*., 2008, **5**, e184.
24 M. Berggren Söderlund, G. Fex and P. Nilsson-Ehle, *J. Intern. Med*., 2003, **253**, 375–380.
25 E. Nisoli, C. Tonello, A. Cardile, V. Cozzi, R. Bracale, L. Tedesco, S. Falcone, A. Valerio, O. Cantoni, E. Clementi, S. Moncada and M. O. Carruba, *Science*, 2005, **310**, 314–317.
26 M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe and M. Hirakawa, *Nucleic Acids Res*., 2010, **38**, D355–D360.
27 S. H. Panowski, S. Wolff, H. Aguilaniu, J. Durieux and A. Dillin, *Nature*, 2007, **447**, 550–555.

28 M. S. Bonkowski, J. S. Rocha, M. M. Masternak, K. A. Al Regaiey and A. Bartke, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 7901–7905.
29 D. J. Clancy, D. Gems, E. Hafen, S. J. Leevers and L. Partridge, *Science*, 2002, **296**, 319.
30 S. J. Lin, P. A. Defossez and L. Guarente, *Science*, 2000, **289**, 2126–2128.
31 M. Hansen, A.-L. Hsu, A. Dillin and C. Kenyon, *PLoS Genet.*, 2005, **1**, 119–128.
32 F. Tang, J. W. Watkins, M. Bermudez, R. Gray, A. Gaban, K. Portie, S. Grace, M. Kleve and G. Craciun, *Autophagy*, 2008, **4**, 874–886.
33 J. P. de Magalhães, A. Budovsky, G. Lehmann, J. Costa, Y. Li, V. Fraifeld and G. M. Church, *Aging Cell*, 2009, **8**, 65–72.
34 A. A. Puca, C. Chatgilialoglu and C. Ferreri, *Int. J. Biochem. Cell Biol.*, 2008, **40**, 324–333.
35 C. Martini, V. Pallottini, E. De Marinis, M. Marino, G. Cavallini, A. Donati, S. Straniero and A. Trentalance, *Mech. Ageing Dev.*, 2008, **129**, 722–727.
36 J. Frey, *Clin. Chem. Lab. Med.*, 2004, **42**, 9–12.
37 P. L. Mote, J. M. Grizzle, R. L. Walford and S. R. Spindler, *J. Gerontol.*, 1991, **46**, B95–B100.
38 O. Froy and R. Miskin, *Aging (N. Y.)*, 2010, **2**, 7–27.
39 M. E. Gourley and C. J. Kennedy, *Comp. Biochem. Physiol., Part C: Toxicol. Pharmacol.*, 2009, **150**, 270–278.
40 D. L. Schmucker, R. K. Wang, D. Snyder, H. Strobel and U. Marti, *J. Gerontol.*, 1991, **46**, B23–B27.
41 O. Froy and R. Miskin, *Prog. Neurobiol.*, 2007, **82**, 142–150.
42 R. Reifen, *Proc. Nutr. Soc.*, 2002, **61**, 397–400.
43 D. Azzout-Marniche, D. Bécard, C. Guichard, M. Foretz, P. Ferré and F. Foufelle, *Biochem. J.*, 2000, **350**(Pt 2), 389–393.
44 T. Sugawara, Y. Fujimoto and T. Ishibashi, *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids*, 2001, **1533**, 277–284.
45 J. Fernø, S. Skrede, A. O. Vik-Mo, B. Håvik and V. M. Steen, *BMC Neurosci.*, 2006, **7**, 69.
46 H.-Q. Yin, M. Kim, J.-H. Kim, G. Kong, K.-S. Kang, H.-L. Kim, B.-I. Yoon, M.-O. Lee and B.-H. Lee, *Toxicol. Appl. Pharmacol.*, 2007, **223**, 225–233.
47 Y.-yi Lin, J.-ying Lu, J. Zhang, W. Walter, W. Dang, J. Wan, S.-C. Tao, J. Qian, Y. Zhao, J. D. Boeke, S. L. Berger and H. Zhu, *Cell*, 2009, **136**, 1073–1084.
48 T. Porstmann, C. R. Santos, B. Griffiths, M. Cully, M. Wu, S. Leevers, J. R. Griffiths, Y.-L. Chung and A. Schulze, *Cell Metab.*, 2008, **8**, 224–236.
49 A. K. Walker, R. L. Jacobs, J. L. Watts, V. Rottiers, K. Jiang, D. M. Finnegan, T. Shioda, M. Hansen, F. Yang, L. J. Niebergall, D. E. Vance, M. Tzoneva, A. C. Hart and A. M. Näär, *Cell*, 2011, **147**, 840–852.
50 L. E. Olofsson, M. Orho-Melander, L. William-Olsson, K. Sjöholm, L. Sjöström, L. Groop, B. Carlsson, L. M. S. Carlsson and B. Olsson, *J. Clin. Endocrinol. Metab.*, 2008, **93**, 4880–4886.
51 T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, I. F. Kim, A. Soboleva, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter and R. Edgar, *Nucleic Acids Res.*, 2009, **37**, D885–D890.
52 H. Parkinson, M. Kapushesky, N. Kolesnikov, G. Rustici, M. Shojatalab, N. Abeygunawardena, H. Berube, M. Dylag, I. Emam, A. Farne, E. Holloway, M. Lukk, J. Malone, R. Mani, E. Pilicheva, T. F. Rayner, F. Rezwan, A. Sharma, E. Williams, X. Z. Bradley, T. Adamusiak, M. Brandizi, T. Burdett, R. Coulson, M. Krestyaninova, P. Kurnosov, E. Maguire, S. G. Neogi, P. Rocca-Serra, S.-A. Sansone, N. Sklyar, M. Zhao, U. Sarkans and A. Brazma, *Nucleic Acids Res.*, 2009, **37**, D868–D872.
53 F. Pan, C.-H. Chiu, S. Pulapura, M. R. Mehan, J. Nunez-Iglesias, K. Zhang, K. Kamath, M. S. Waterman, C. E. Finch and X. J. Zhou, *Nucleic Acids Res.*, 2007, **35**, D756–D759.
54 C. Fu, M. Hickey, M. Morrison, R. McCarter and E.-S. Han, *Mech. Ageing Dev.*, 2006, **127**, 905–916.
55 S. X. Cao, J. M. Dhahbi, P. L. Mote and S. R. Spindler, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10630–10635.
56 J. M. Dhahbi, T. Tsuchiya, H.-J. Kim, P. L. Mote and S. R. Spindler, *J. Gerontol., Ser. A*, 2006, **61**, 218–231.
57 J. C. Corton, U. Apte, S. P. Anderson, P. Limaye, L. Yoon, J. Latendresse, C. Dunn, J. I. Everitt, K. A. Voss, C. Swanson, C. Kimbrough, J. S. Wong, S. S. Gill, R. A. S. Chandraratna, M.-K. Kwak, T. W. Kensler, T. M. Stulnig, K. R. Steffensen, J.-A. Gustafsson and H. M. Mehendale, *J. Biol. Chem.*, 2004, **279**, 46204–46212.
58 T. Kayo, D. B. Allison, R. Weindruch and T. A. Prolla, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 5093–5098.
59 T. J. P. Hubbard, B. L. Aken, S. Ayling, B. Ballester, K. Beal, E. Bragin, S. Brent, Y. Chen, P. Clapham, L. Clarke, G. Coates, S. Fairley, S. Fitzgerald, J. Fernandez-Banet, L. Gordon, S. Graf, S. Haider, M. Hammond, R. Holland, K. Howe, A. Jenkinson, N. Johnson, A. Kahari, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, D. Rios, M. Schuster, G. Slater, D. Smedley, W. Spooner, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, S. Wilder, A. Zadissa, E. Birney, F. Cunningham, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, A. Kasprzyk, G. Proctor, J. Smith, S. Searle and P. Flicek, *Nucleic Acids Res.*, 2009, **37**, D690–D697.
60 E. W. Sayers, T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. Dicuccio, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, Z. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Mizrachi, J. Ostell, A. Panchenko, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. Wang, W. John Wilbur, E. Yaschenko and J. Ye, *Nucleic Acids Res.*, 2010, **38**, D5–D16.
61 R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2009.
62 G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane and R. A. Lempicki, *GenomeBiology*, 2003, **4**, P3.
63 A. Kel, N. Voss, R. Jauregui, O. Kel-Margoulis and E. Wingender, *BMC Bioinformatics*, 2006, **7**(Suppl 2), S13.
64 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.