GCB 2012 Satellite Workshop on Systems Biology of Ageing,
September 19, 2012, Jena/Germany

Cross-species analysis of age-related transcriptome data

S. Priebe[1], U. MENZEL[1], R. Guthke[1] and the JenAge consortium[2]

[1]Systems Biology and Bioinformatics Group, Hans-Knöll-Institute, Jena.
[2]JenAge: Jena Centre for Systems Biology of Ageing

Leibniz Institute
for Natural Product Research and Infection Biology
Hans Knöll Institute

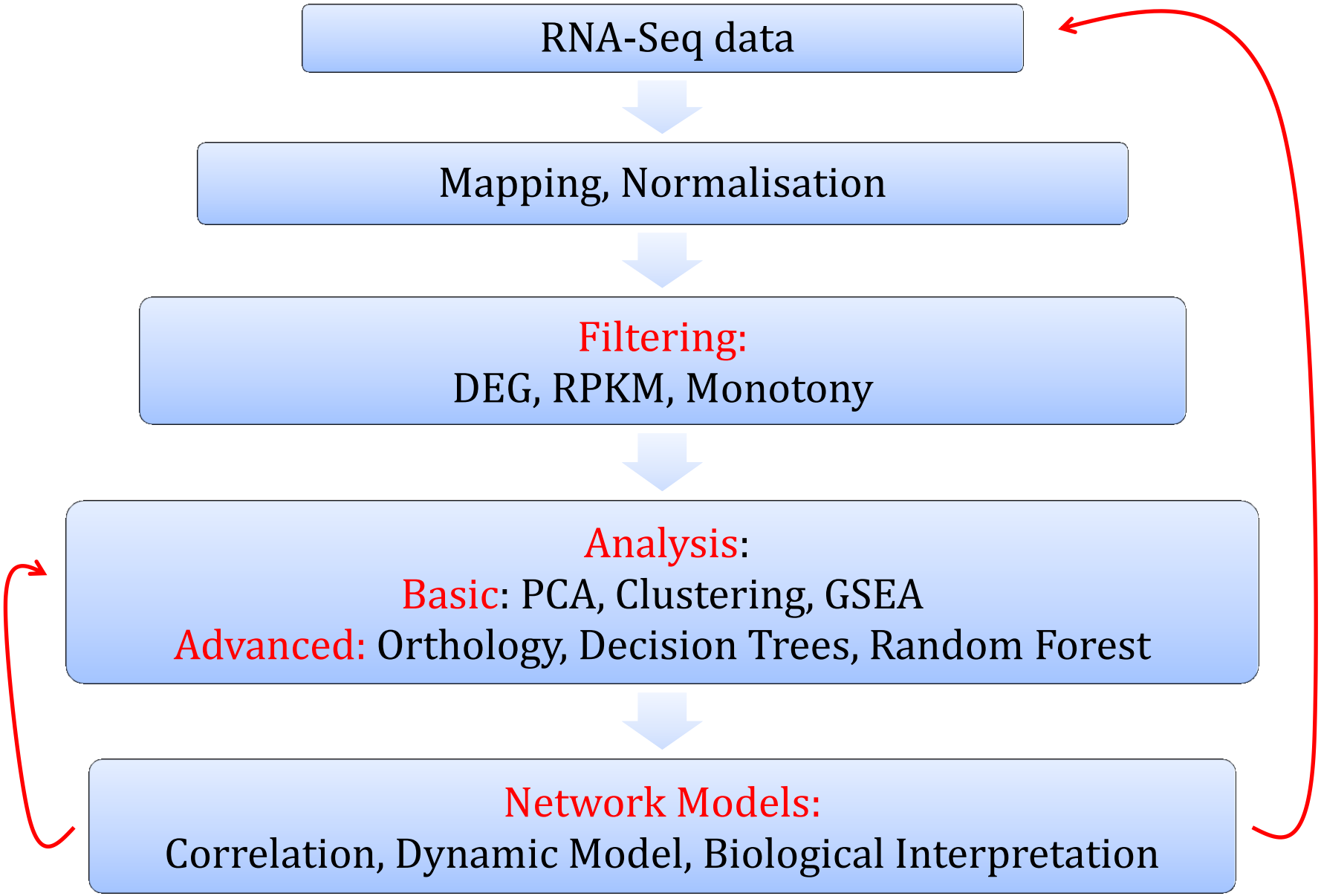
JENAGE

Multi-species approach

	Scientific name	Common name	Lifetime
	*Caenorhabditis elegans*	Worm	2-3 weeks
	*Nothobranchius furzeri*	Killifish	3 months
	*Danio rerio*	Zebrafish	30 months
	*Mus musculus*	Mouse	36-48 months

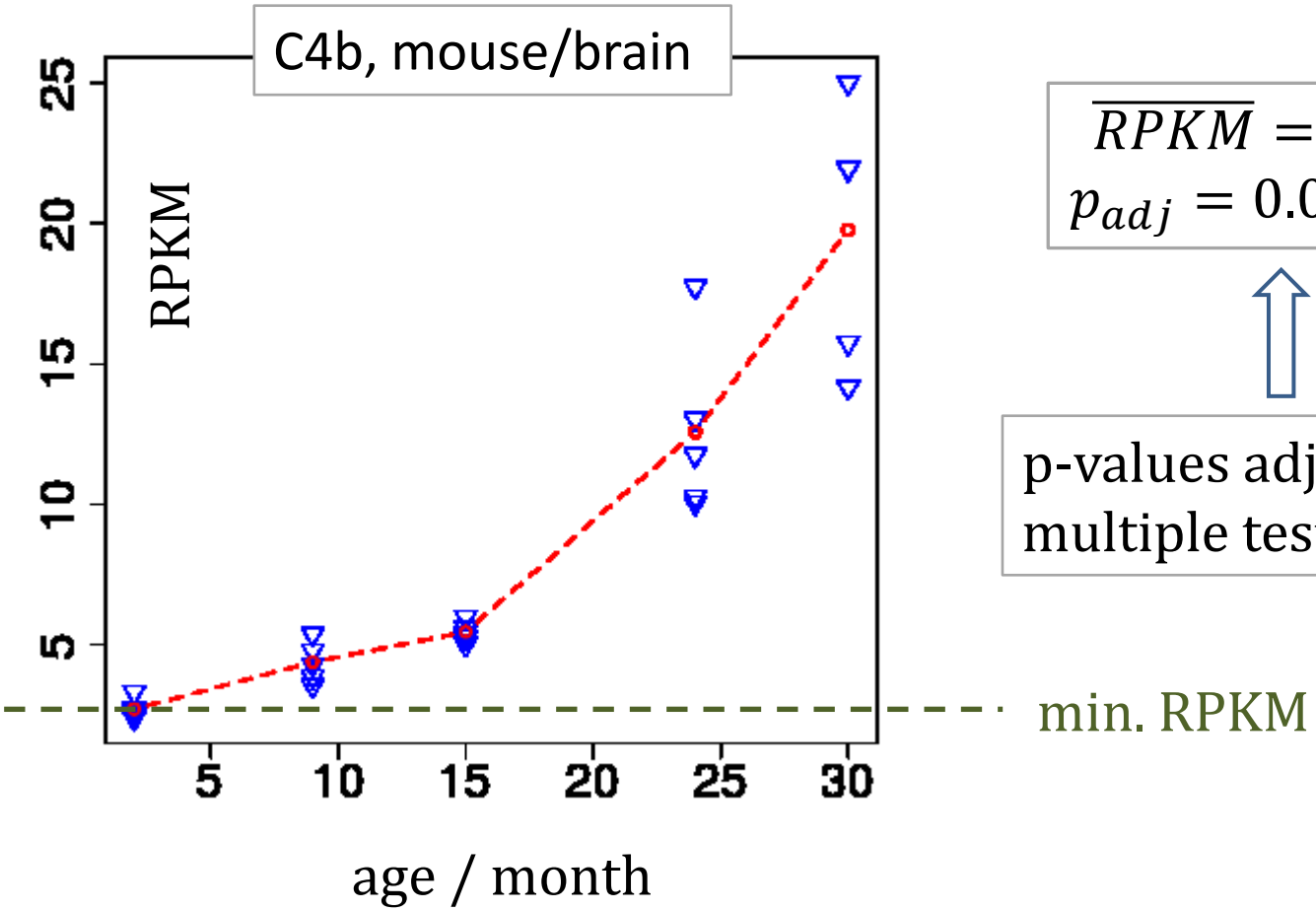
- 4 - 5 age levels
- 2 - 5 replicates at each level

Data flow


RNA-Seq data
Mapping, Normalisation
Filtering:
DEG, RPKM, Monotony
Analysis:
Basic: PCA, Clustering, GSEA
Advanced: Orthology, Decision Trees, Random Forest
Network Models:
Correlation, Dynamic Model, Biological Interpretation

Filtering: 1) DEG, 2) RPKM

1. DEG: Differentially Expressed Genes (over age)
   - edgeR, DESeq, baySeq (NegBin, overdispersion)
2. RPKM: Reads per Kb of exon model per Million mapped reads

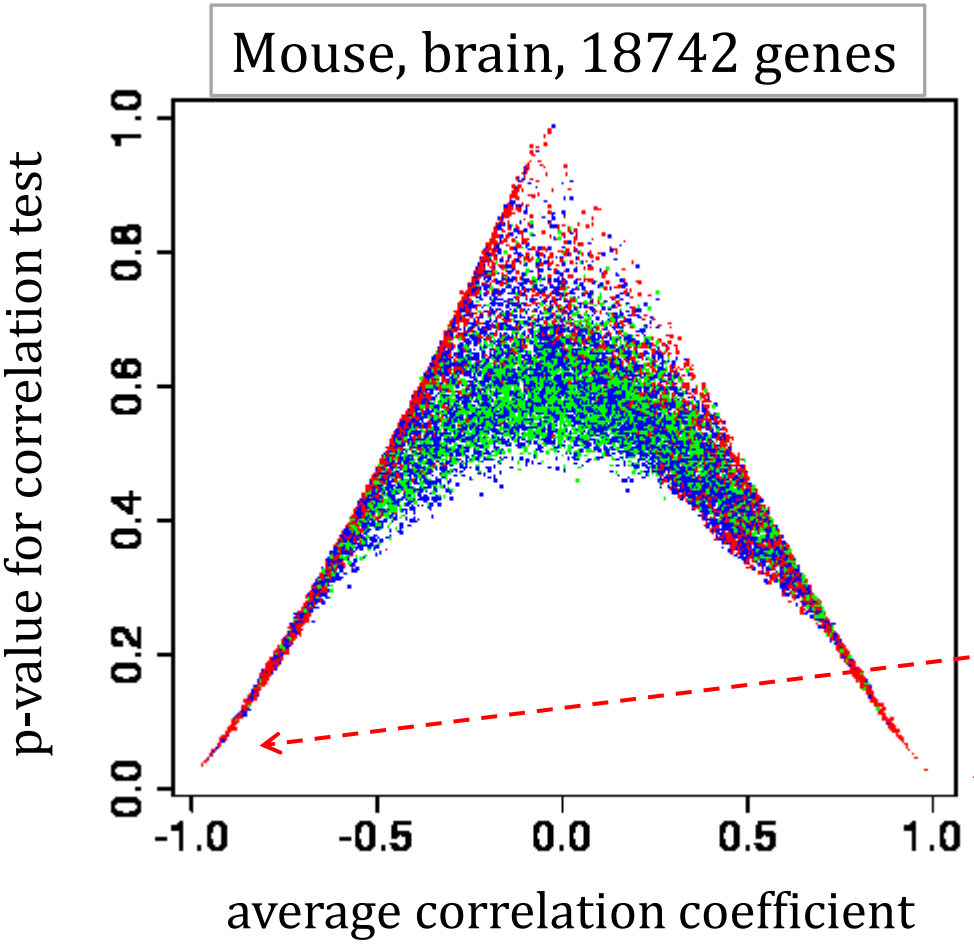


$\overline{RPKM} = 8.98$
$p_{adj} = 0.0012$

p-values adjusted for multiple testing (FDR)

Filtering: 3) Monotony with age

- Genes changing monotonically with age are interesting
- Spearman rank correlation to prototype (Spearman's ρ)
- permutation test to calculate p-values



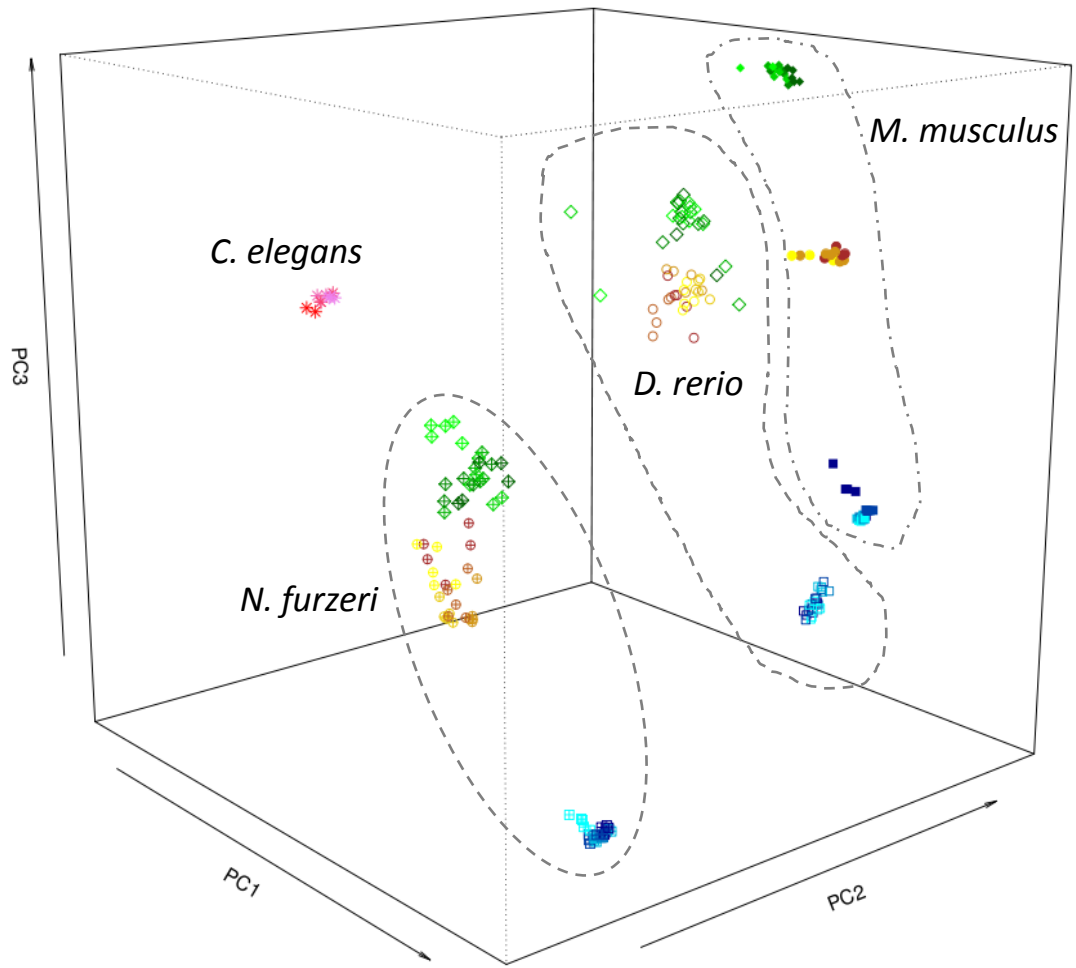
- 18742 genes
- $|RPKM| \geq 1$
- $FDR \leq 0.05$

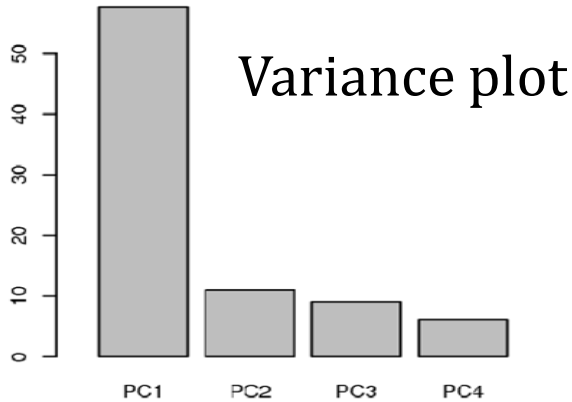
genes with a strongly monotonic slope

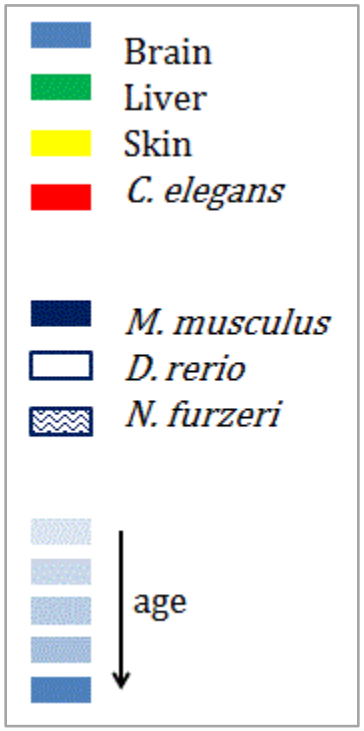
JENAGE

Principal Component Analysis (PCA)

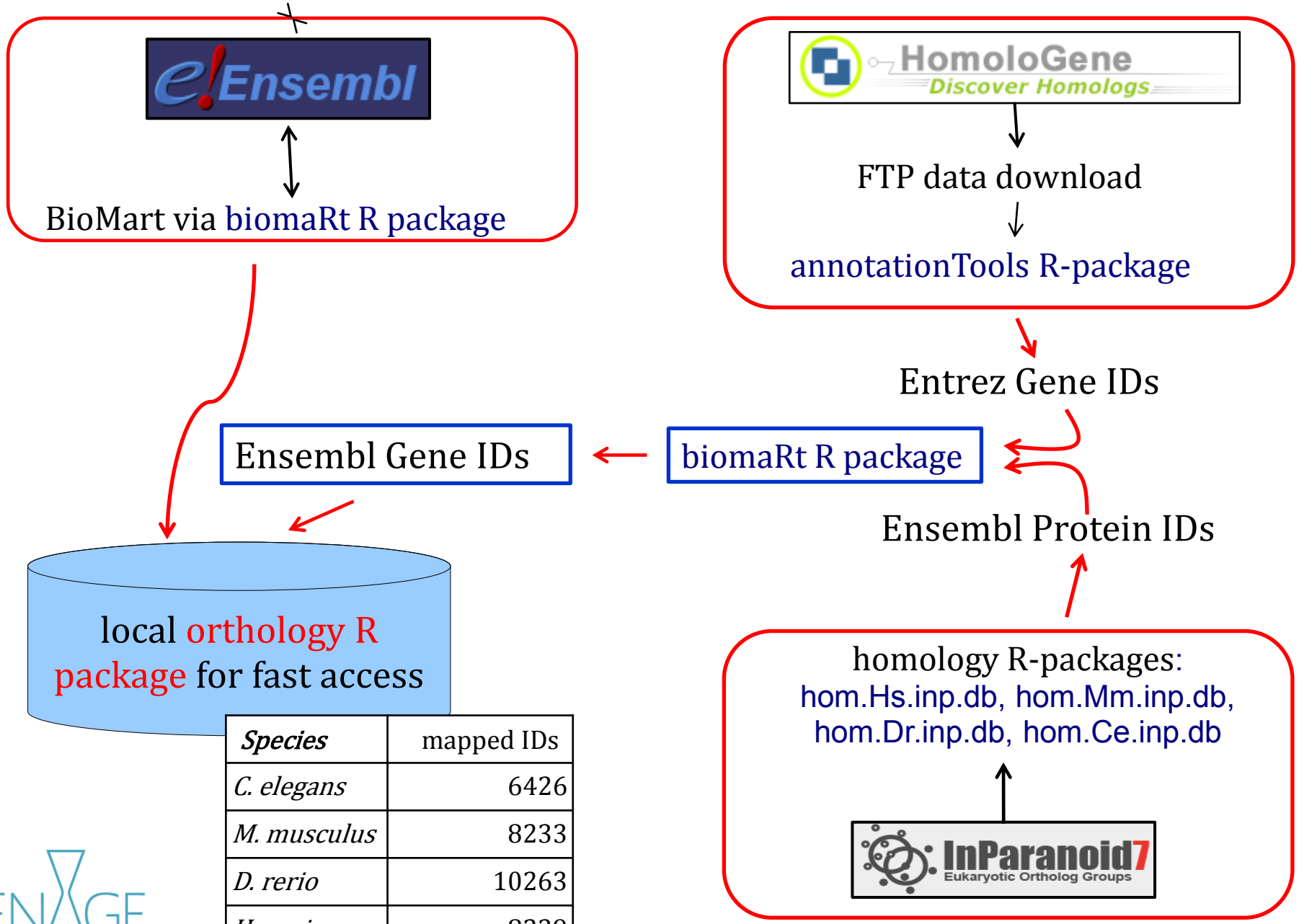
Principal components 1, 2 & 3







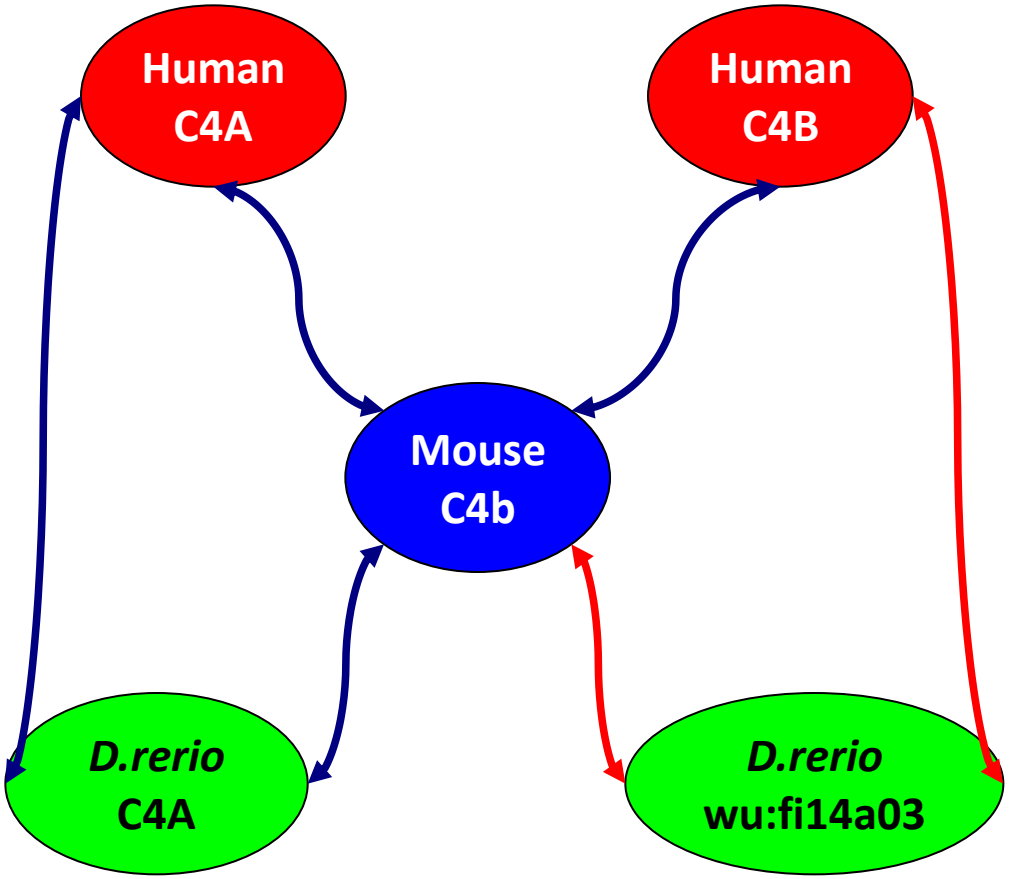
Orthology: New, integrated R-package



Species	mapped IDs
C. elegans	6426
M. musculus	8233
D. rerio	10263
H. sapiens	8239

JENAGE

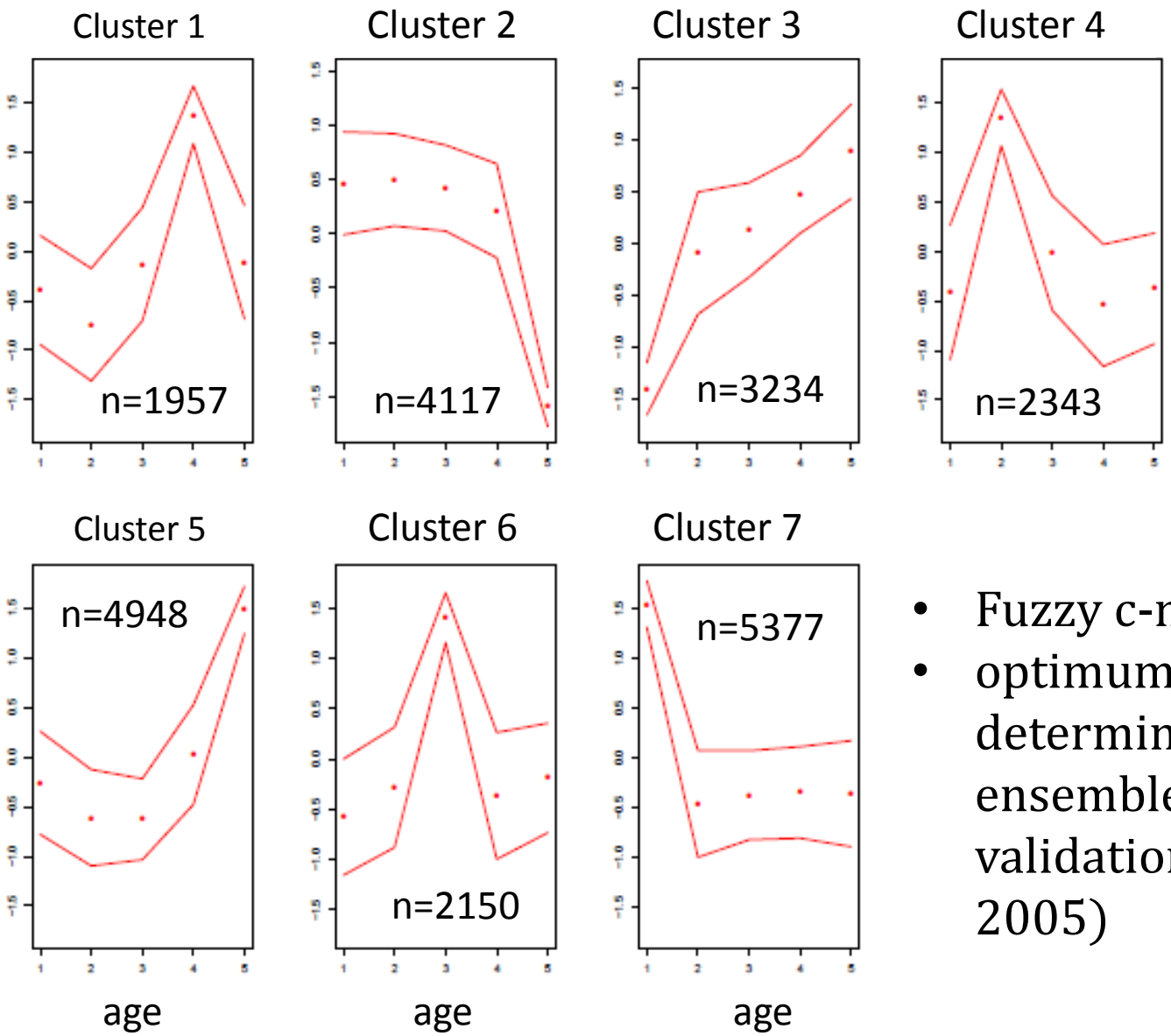
Orthology



color of the arrows indicates database of origin

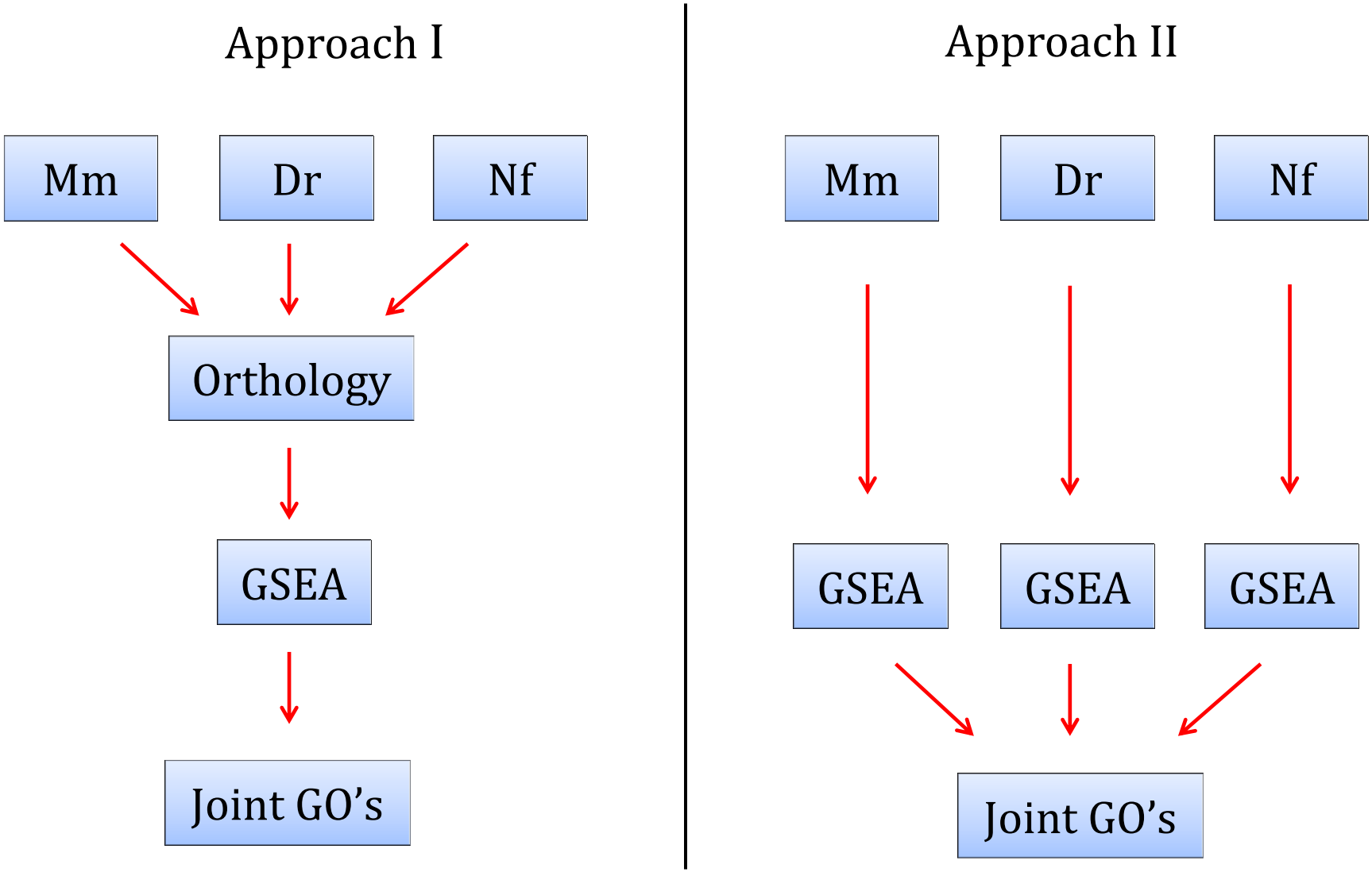
Clustering (across species)

Union of Species: *M. musculus, D. rerio, N. furzeri, C. elegans*



- Fuzzy c-means clustering
- optimum no. of clusters determined using an ensemble of cluster validation indices (Guthke 2005)

Gene Set Enrichment Analysis (GSEA)



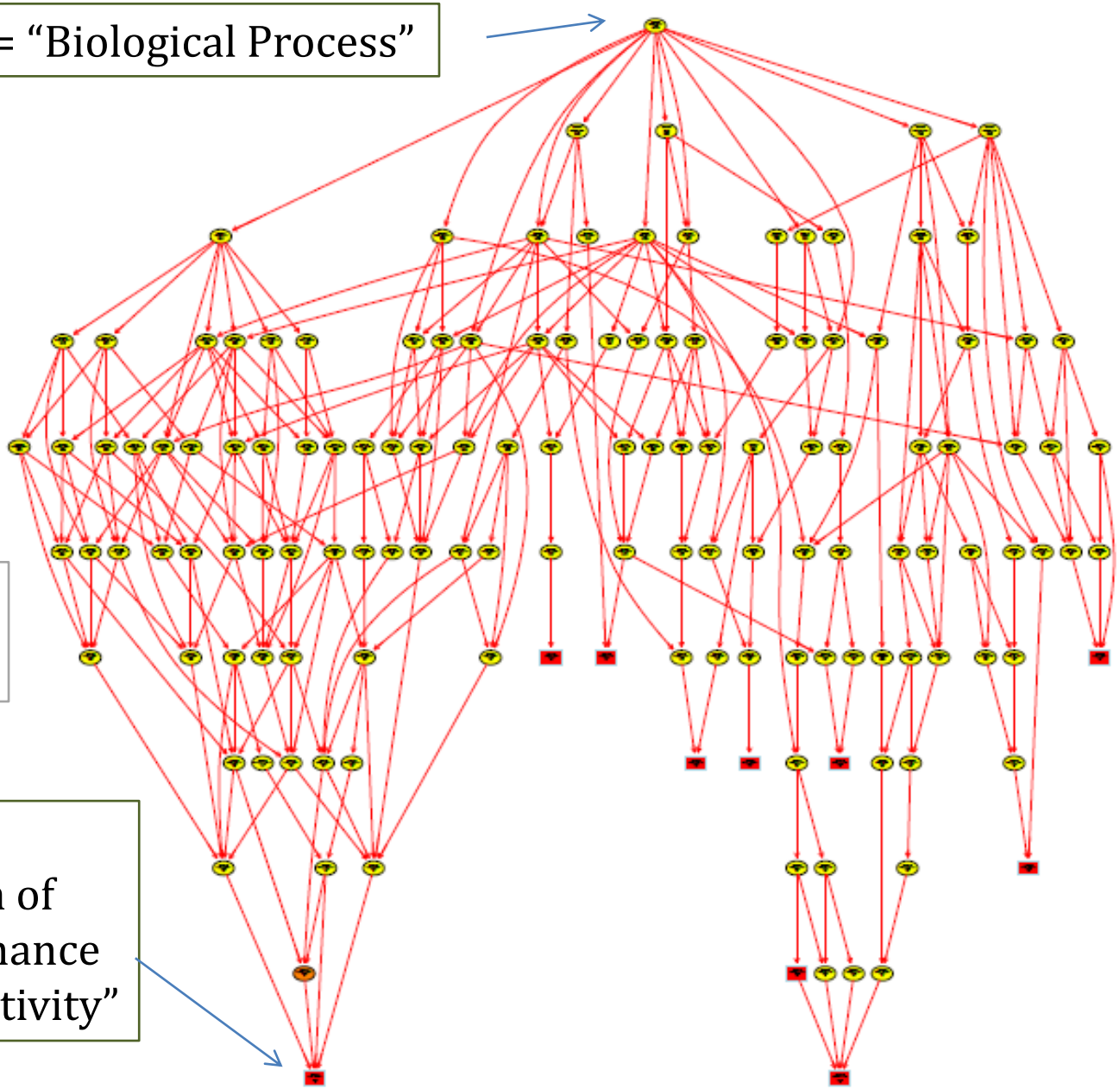
GSEA: Graphical Representation

GO:0008150 = "Biological Process"

Mouse, brain:
176 up-genes
$FDR \leq 0.05$
$\overline{RPKM} \geq 3$
$\bar{\rho} \geq 0.75$

KS-Test with elimination

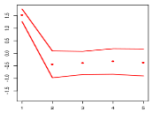
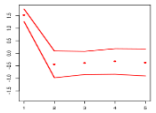
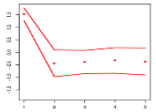
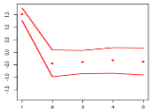
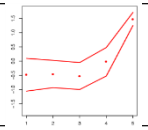
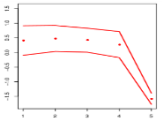
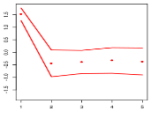
GO:0032211 = "down-regulation of telomere maintenance via telomerase activity"



GSEA: Table of Enriched GO's & Pathways

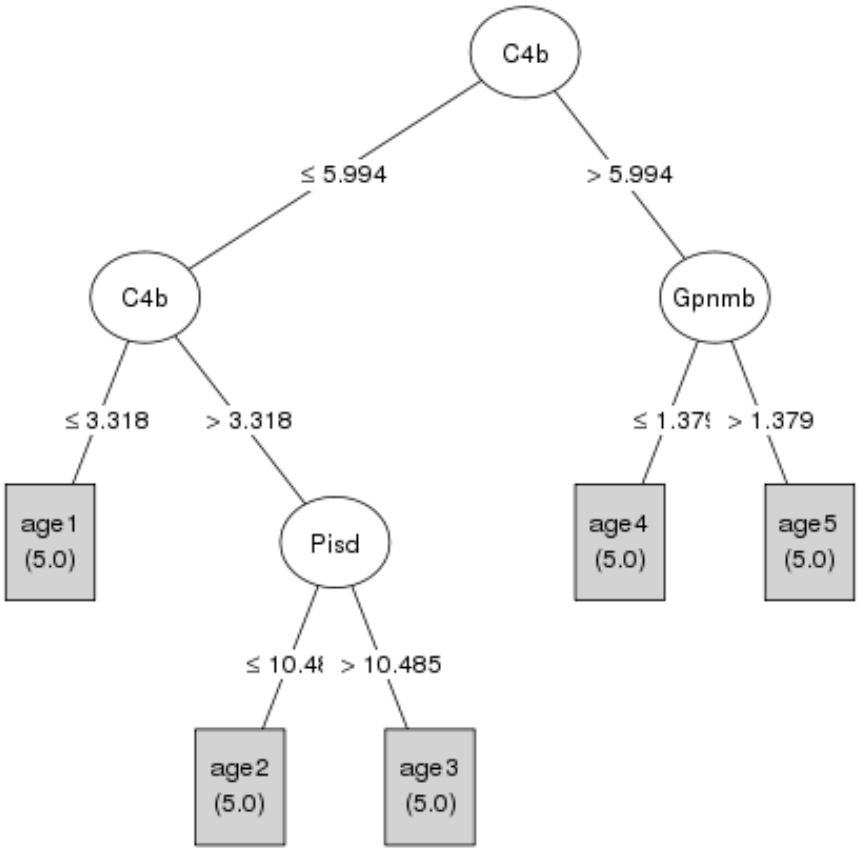
Clustering → GSEA → Intersection

3 species: mouse + 2 fishes, > 200 enriched GO's

Tissue	GO-ID	Description	Cluster
brain	GO:0007399	nervous system development	
brain	GO:0007017	microtubule-based process	
brain	GO:0007169	transmembrane receptor protein tyrosine ...	
brain	KEGG:04512	ECM-receptor interaction	
skin	GO:0042113	B cell activation	
skin	GO:0031012	Extracellular matrix	
liver	GO:0000278	Mitotic cell cycle	

Supervised Machine Learning: Decision Trees

- Classifier, can be used when the number of variables (genes) is higher than the number of observations (transcriptome data sets)
- pinpoints genes that are informative with regard to some attribute, e.g. age, tissue, or species.



Mouse, brain
$|\bar{\rho}| \geq 0.85$
no FDR or RPKM-filter

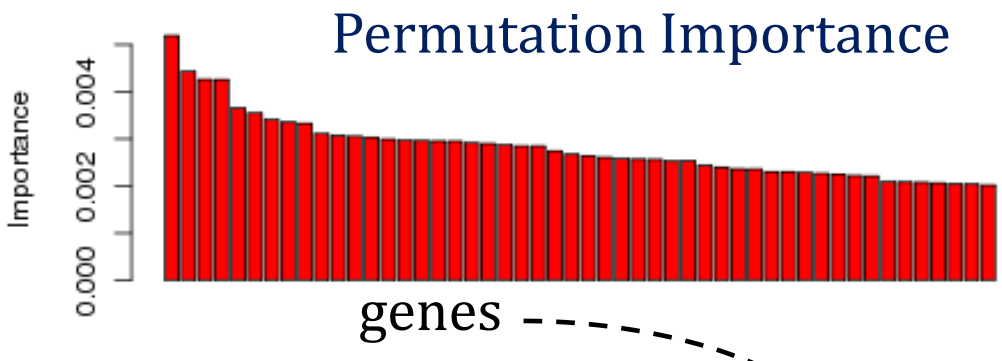
C4b	Complement component 4B
GpnmB	Glycoproteine (transmembrane)
Pisd	Phosphatidylserine decarboxylase

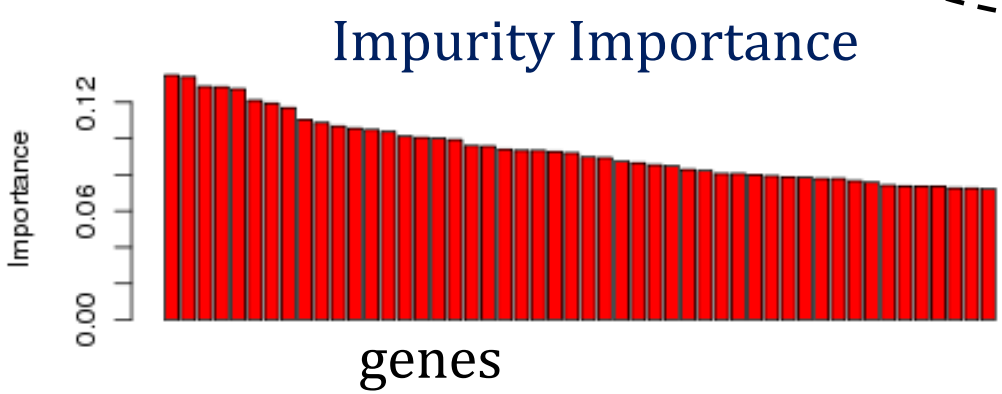
- C4.5 algorithm (Quinlan 1993).
- Overfitting reduced by pruning

Random Forest: Variable Importance

- Ensemble classifier, builds many Decision Trees (Breiman, Cutler)
- random exclusion of a part of variables (and samples) in each tree
- **Variable importance:** measures explanatory power of a variable (gene)

same data as in previous slide



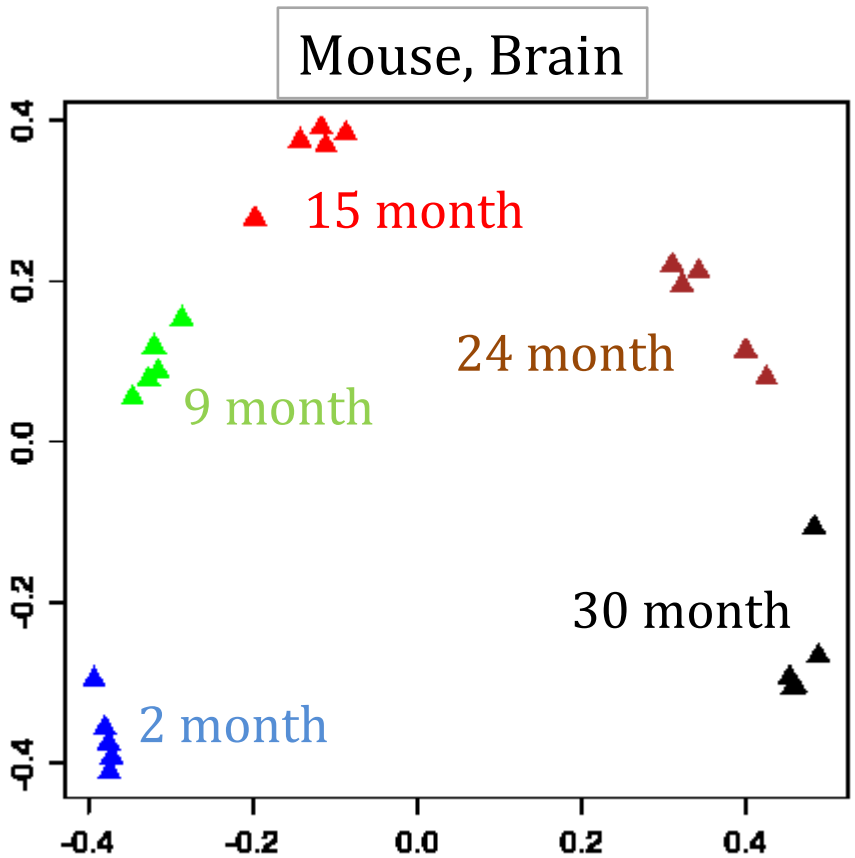


Putative biomarkers:

Apc
C4b
Pisd-ps1
ing4
depdc1a
Pcdh20
Pisd
Plekhb1
Epha3
Tmem167
Pcdhb9
Gpnmb
Prkci

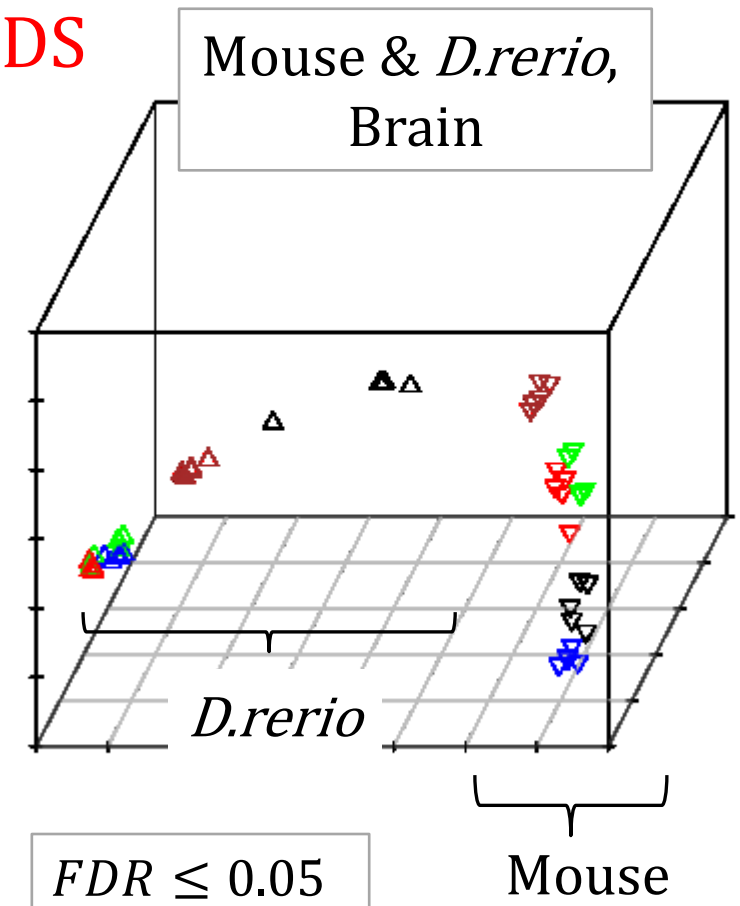
Random Forest: Sample Proximity

- A RF-classifier estimates the pairwise sample proximity (SP)
- samples are "close" if they end up in the same leaf frequently



$|\bar{\rho}| \geq 0.85$
no FDR or RPKM-filter

MDS

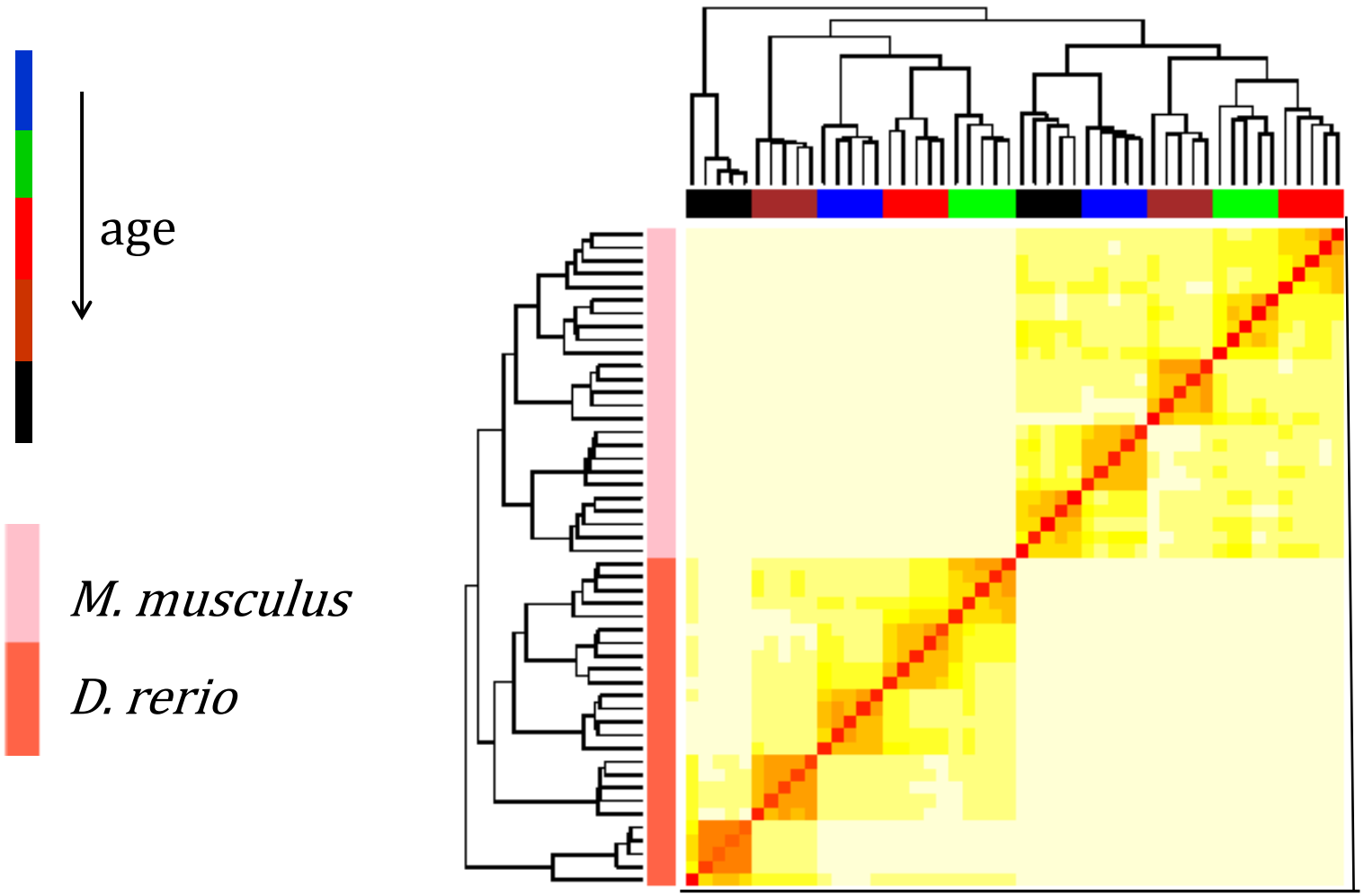


$FDR \leq 0.05$
$|RPKM| \geq 1$
$|\bar{\rho}| \geq 0.3$

JENAGE

Random Forest: Hierarchical clustering

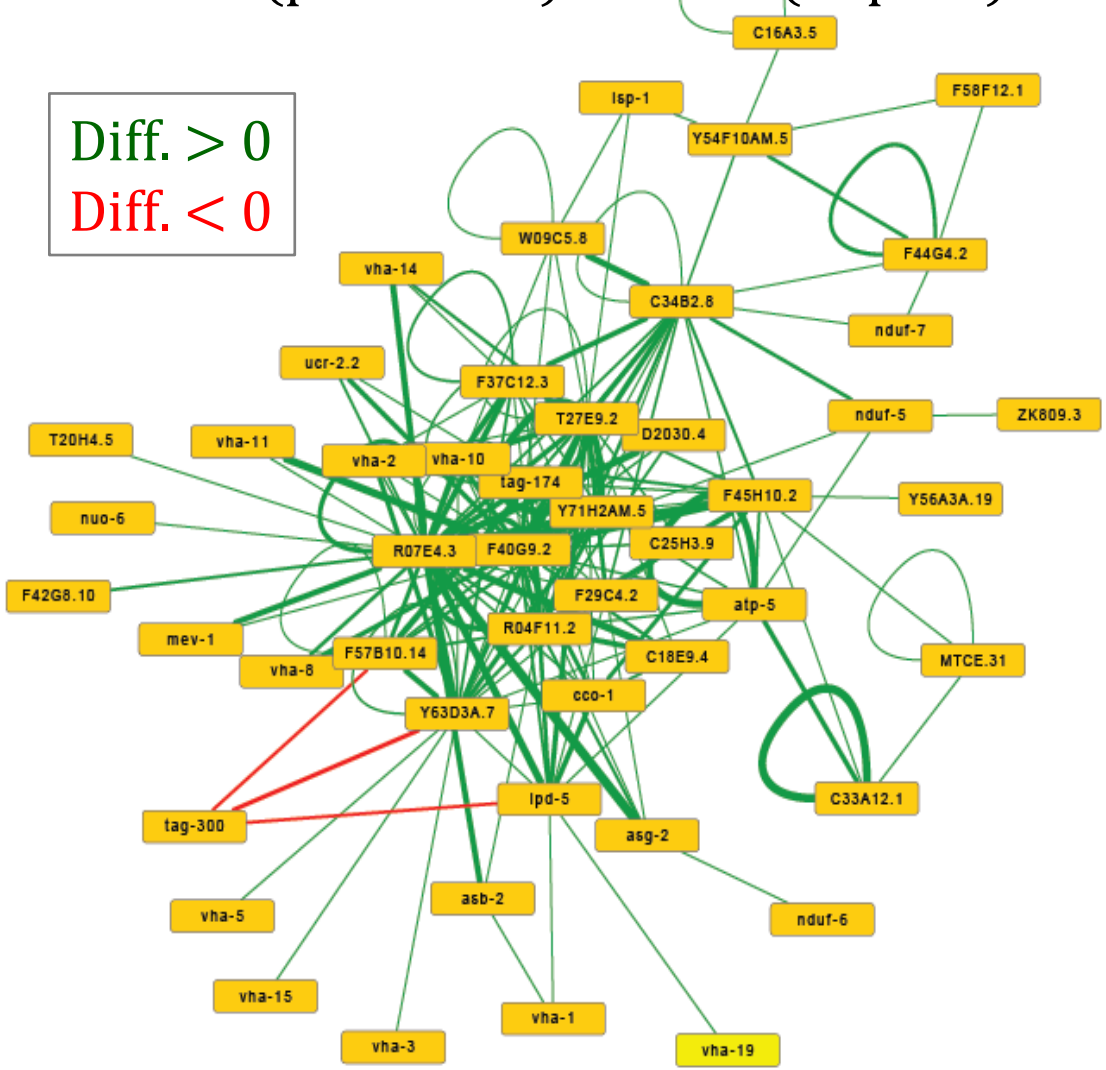
- Hierarchical clustering for *M. musculus* and *D. rerio* based on proximity from Random Forest



Correlation Networks

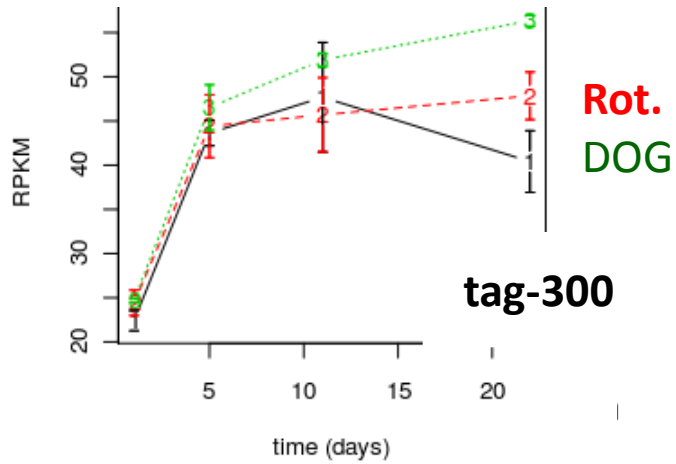
Difference network: *C. elegans*
KEGG-pathway "Oxidative Phosphorylation"
Rotenone (perturbed) - DMSO (unpert.)

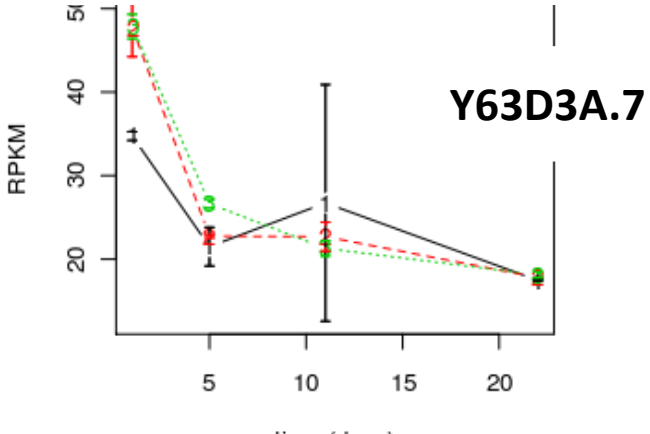
Diff. > 0
Diff. < 0



tag-300/Y110A7A.12 encodes an ortholog of subunit B of the cytoplasmic (V1) domain of vacuolar proton-translocating ATPase

Y63D3A.7 encodes the C. elegans ortholog of the NDUFA2/B8 subunit of the mitochondrial NADH dehydrogenase (ubiquinone) complex (complex I)





Pathways with significantly changed co-expression

In the difference network:

- count fraction of connected genes in a pathway/GO-term
- count fraction of connected genes in the whole network

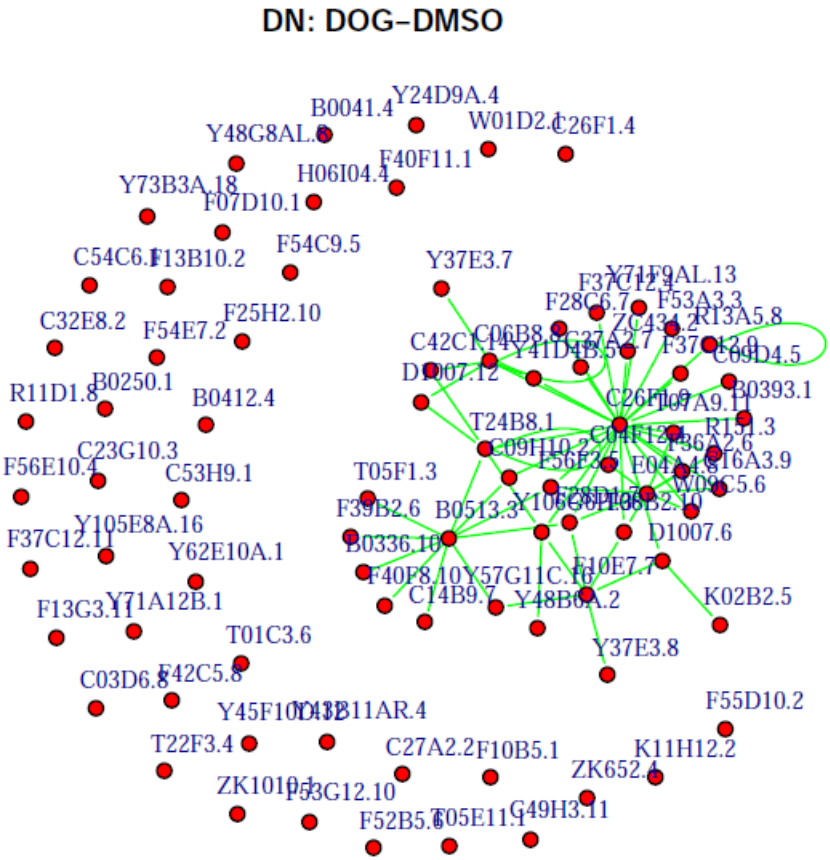
➢ pathways with disproportionately many connected genes

Fisher's Exact Test

fisher.test

Do this for:

- 2881 GO terms and
- 125 KEGG-pathways



Pathways with significantly changed co-expression

- KEGG-pathways whose co-expression is affected most by the perturbation (DOG, glucose restriction):

KEGG-ID	Term	p.adj	Genes	… connected
3010	Ribosome	0.0000	83	41
1100	Metabolic pathways	0.0000	596	214
190	Oxidative phosphorylation	0.0000	100	39
3018	RNA degradation	0.0159	37	9
350	Tyrosine metabolism	0.0396	19	3

Outlook

- Inclusion of human cell lines into the analysis (orthology)
- Perturbed ageing for more species (currently Ce, Mm, Hs only)
- Dynamic models of hormesis connected to ageing
  - mTOR
  - preliminary results achieved
- Identification of relevant biomarkers for ageing, and of targets to support healthy ageing
- Wet-lab validation

Jena Centre for Systems Biology of Ageing

Systems Biology and Bioinformatics Group, Hans-Knöll-Institute (HKI):

- Steffen Priebe
- Uwe Menzel
- Reinhard Guthke

Leibniz Institute
for Natural Product Research and Infection Biology
Hans Knöll Institute

Fritz Lipmann Institute (FLI):

- Alessandro Cellerino
- Christoph Englert
- Stefan Diekmann/Peter Hemmerich
- Matthias Platzer
- Jürgen Sühnel

Leibniz Institute for Age Research
Fritz Lipmann Institute (FLI)

Friedrich Schiller University Jena:

- Udo Hahn
- Christoph Kaleta
- Michael Ristow
- Stefan Schuster

seit 1558

Jena University Hospital:

- Otto Witte

Universitätsklinikum
Jena