Pathview: a pathway based data integration and visualization tool

Weijun Luo*, Cory Brouwer

Email: luo_weijun@yahoo.com

UNC Charlotte

Bioinformatics Service Division

Dept. Bioinformatics & Genomics

Agenda

- Overview
- Main features
  - Basic visualization
  - Data integration
  - Integrated pathway analysis
- Implementation & impact

Overview

Pathview package

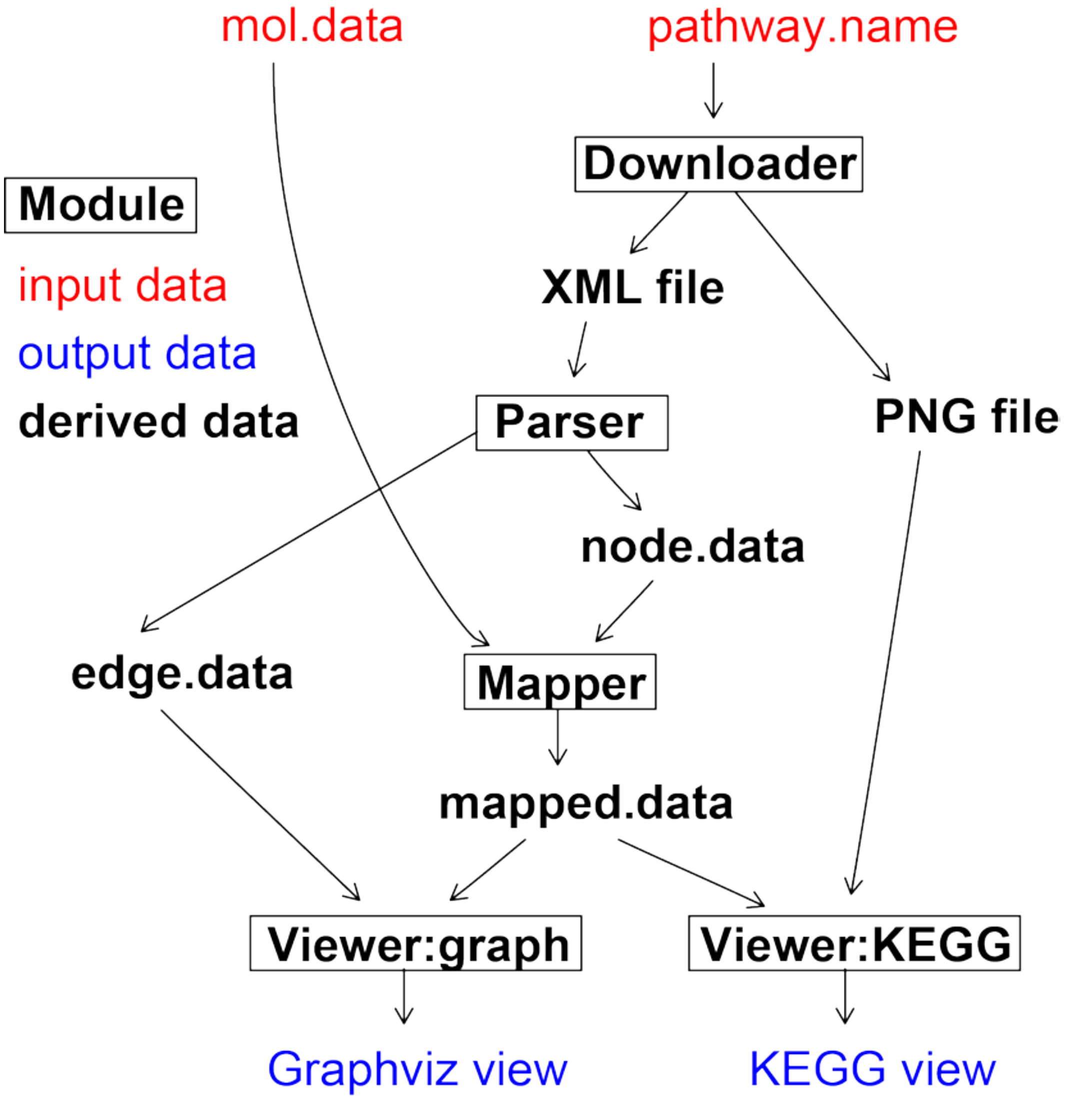
Luo *et al. Bioinformatics* 2013

- Pathway visualization
  - Informative and readable: KEGG based
  - Publication quality: full graphics

- Strong data integration
  - different data types, formats, attributes
  - numerous ID types, species etc.

- Automated/integrated pathway analysis

Design



Module

input data

output data

derived data

mol.data

pathway.name

Downloader

XML file

PNG file

Parser

node.data

edge.data

Mapper

mapped.data

Viewer:graph

Viewer:KEGG

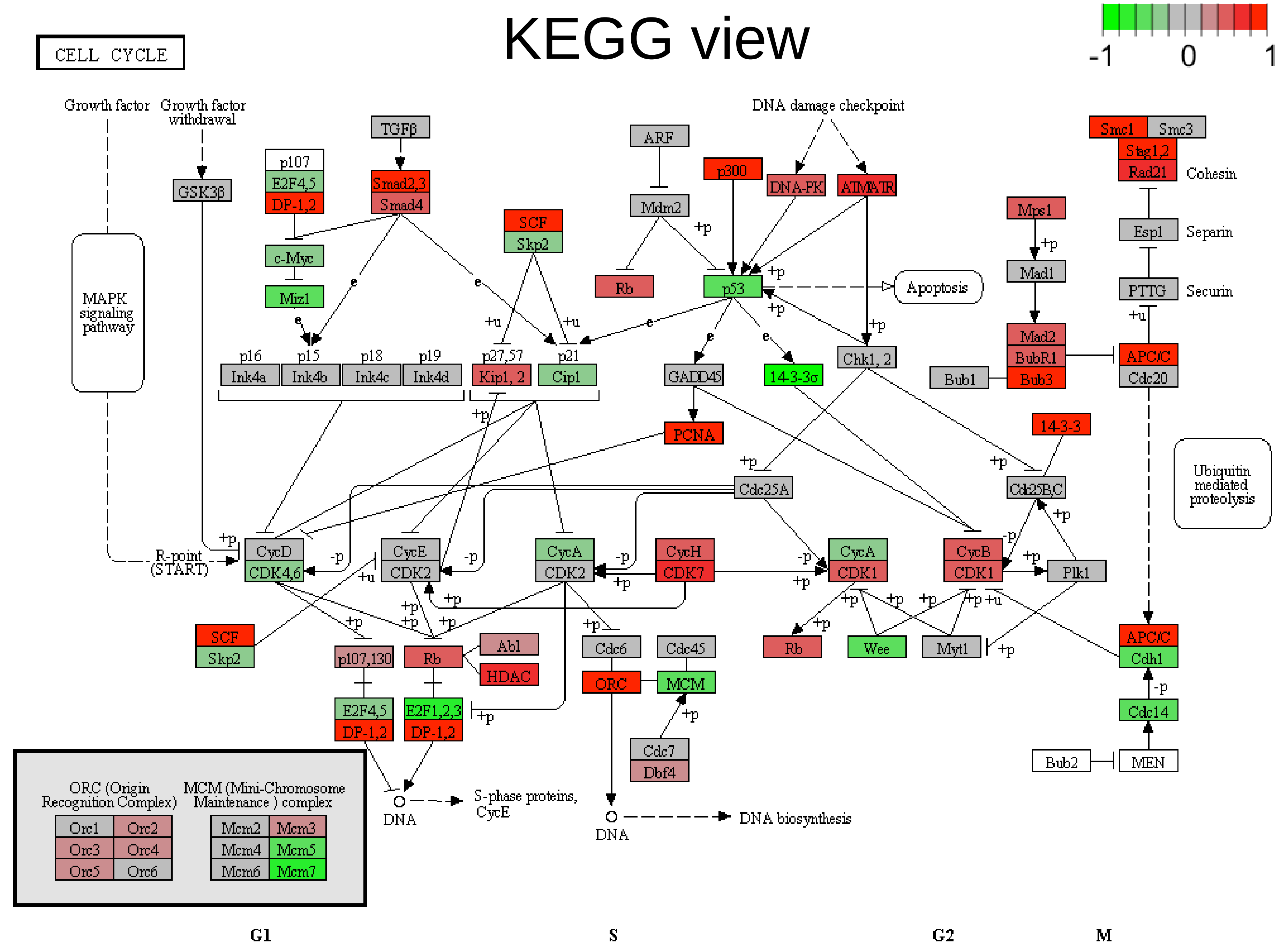
Graphviz view

KEGG view

pathview(gene.data = gse16873.d, pathway.id = "04110")

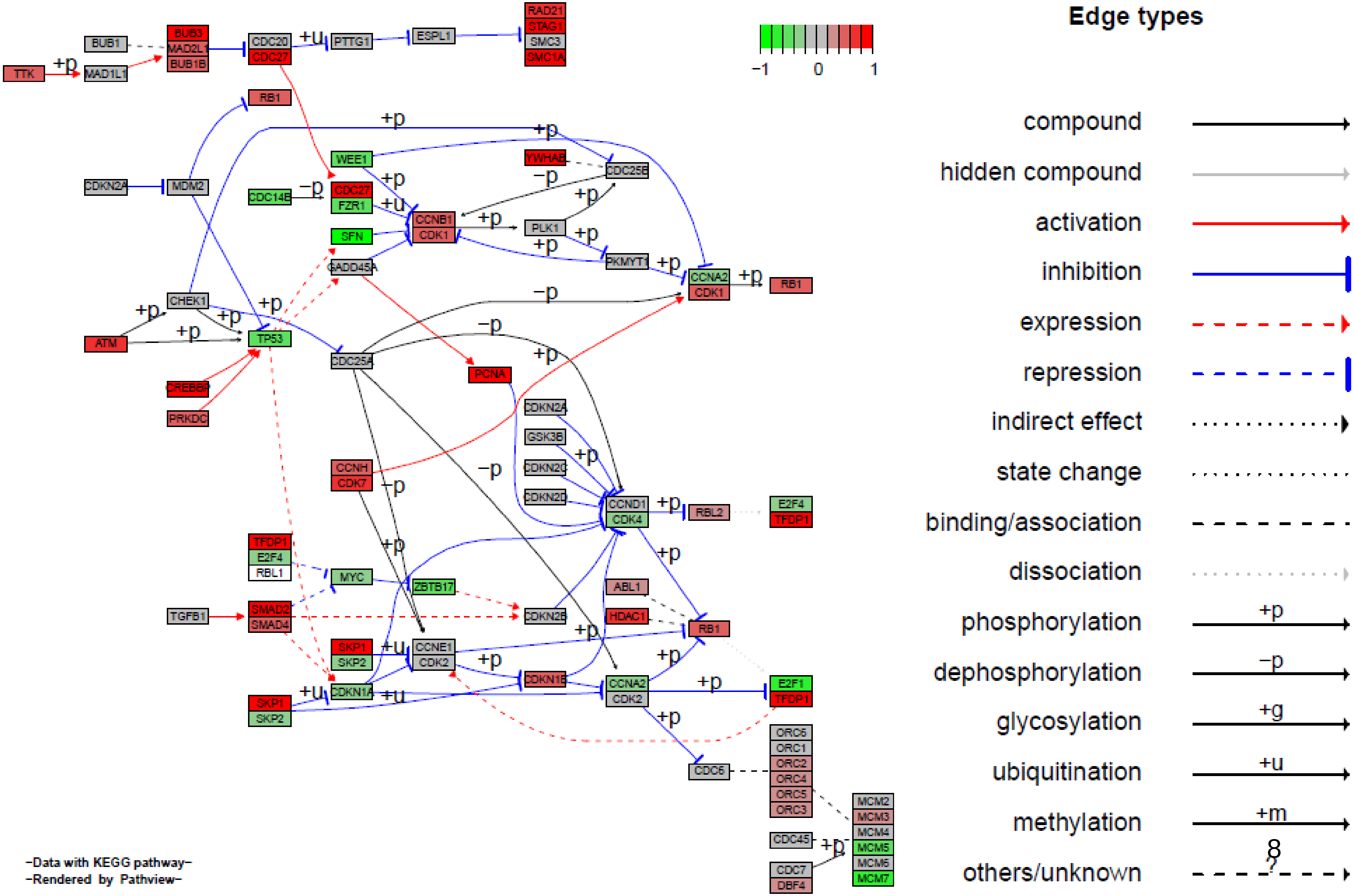
Basic visualization

KEGG view



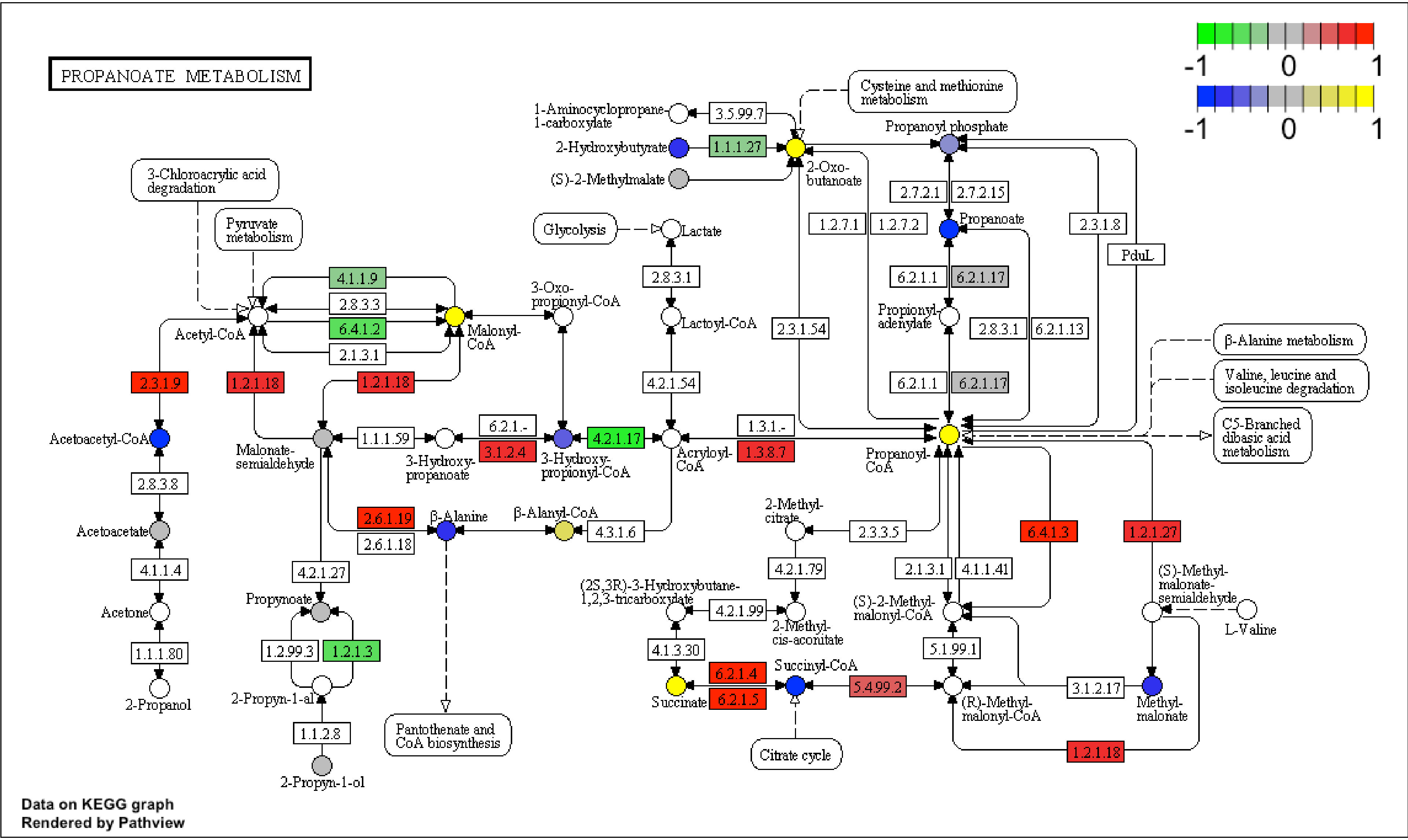
Data on KEGG graph
Rendered by Pathview

Graphviz view



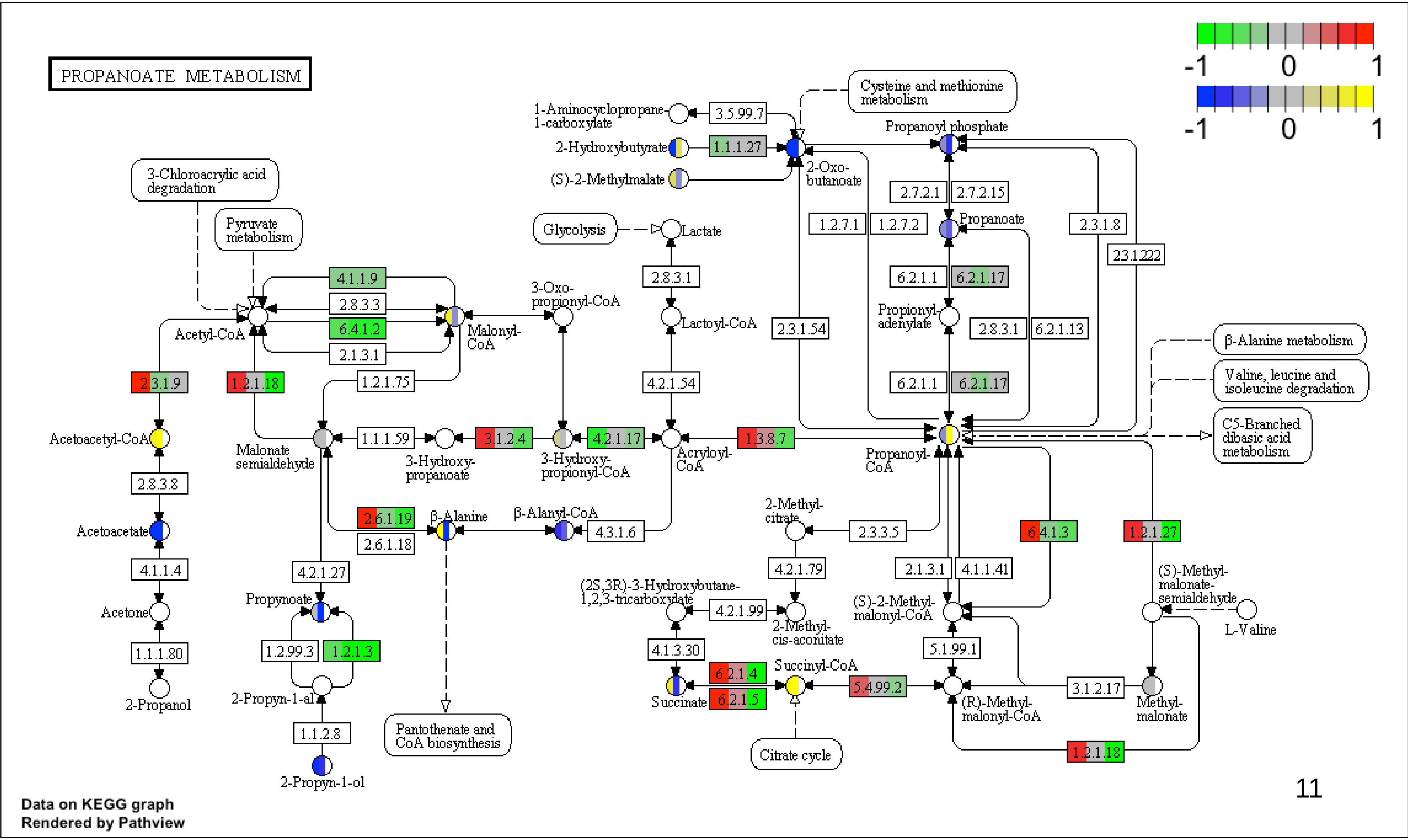
Data integration

KEGG view: gene + compound data

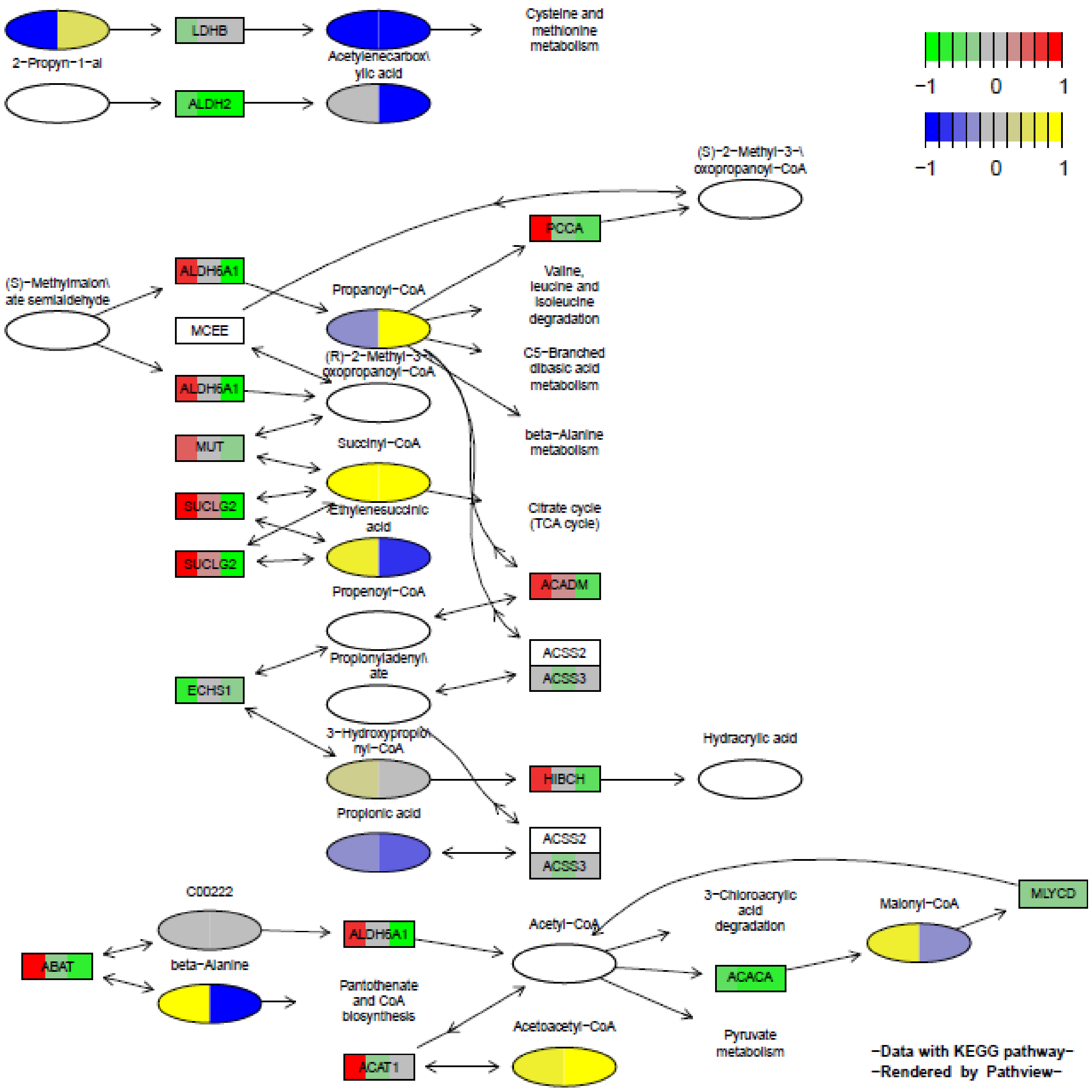


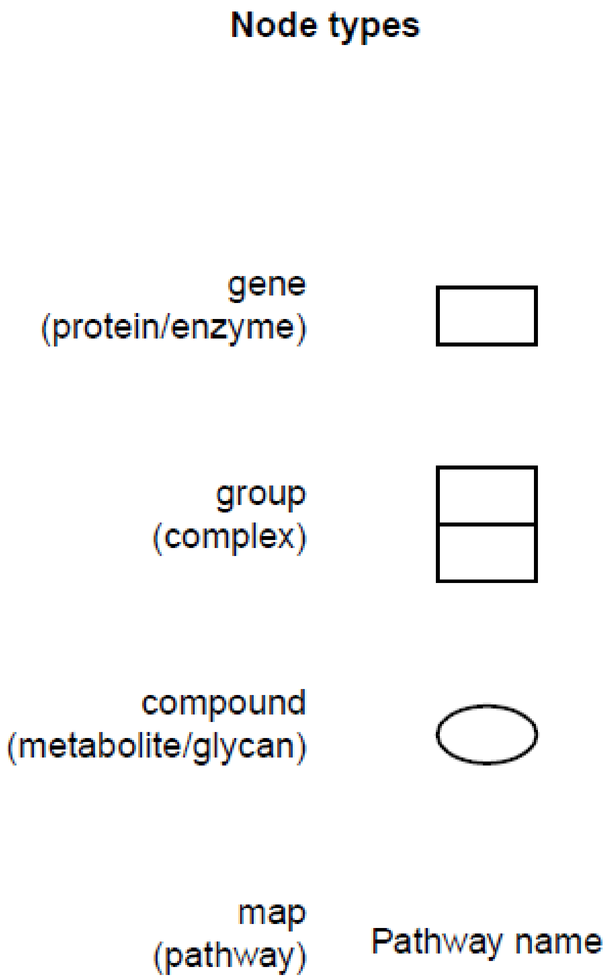
Data on KEGG graph
Rendered by Pathview

KEGG view: multiple samples

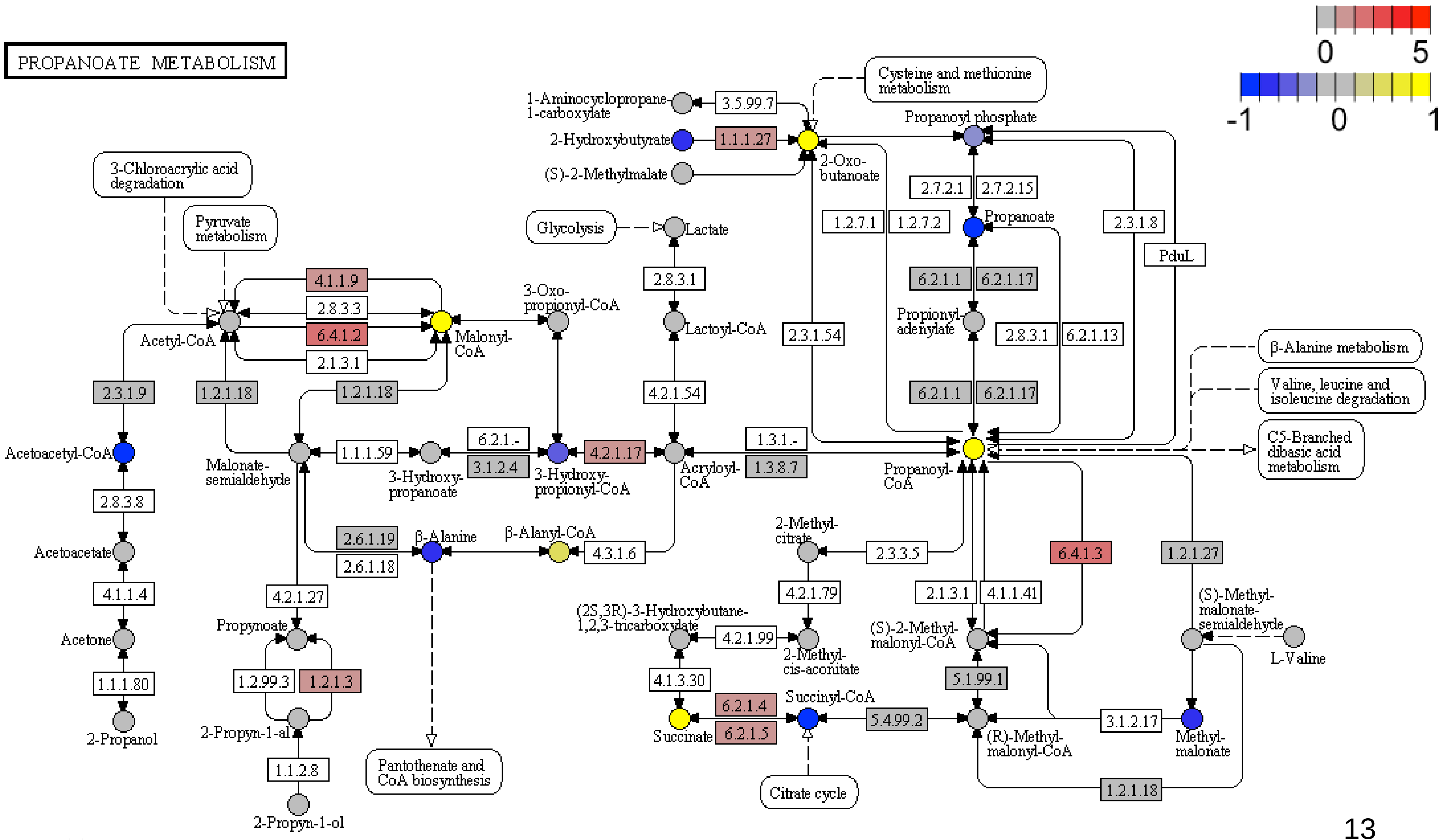


Graphviz view : multiple samples





Data attributes

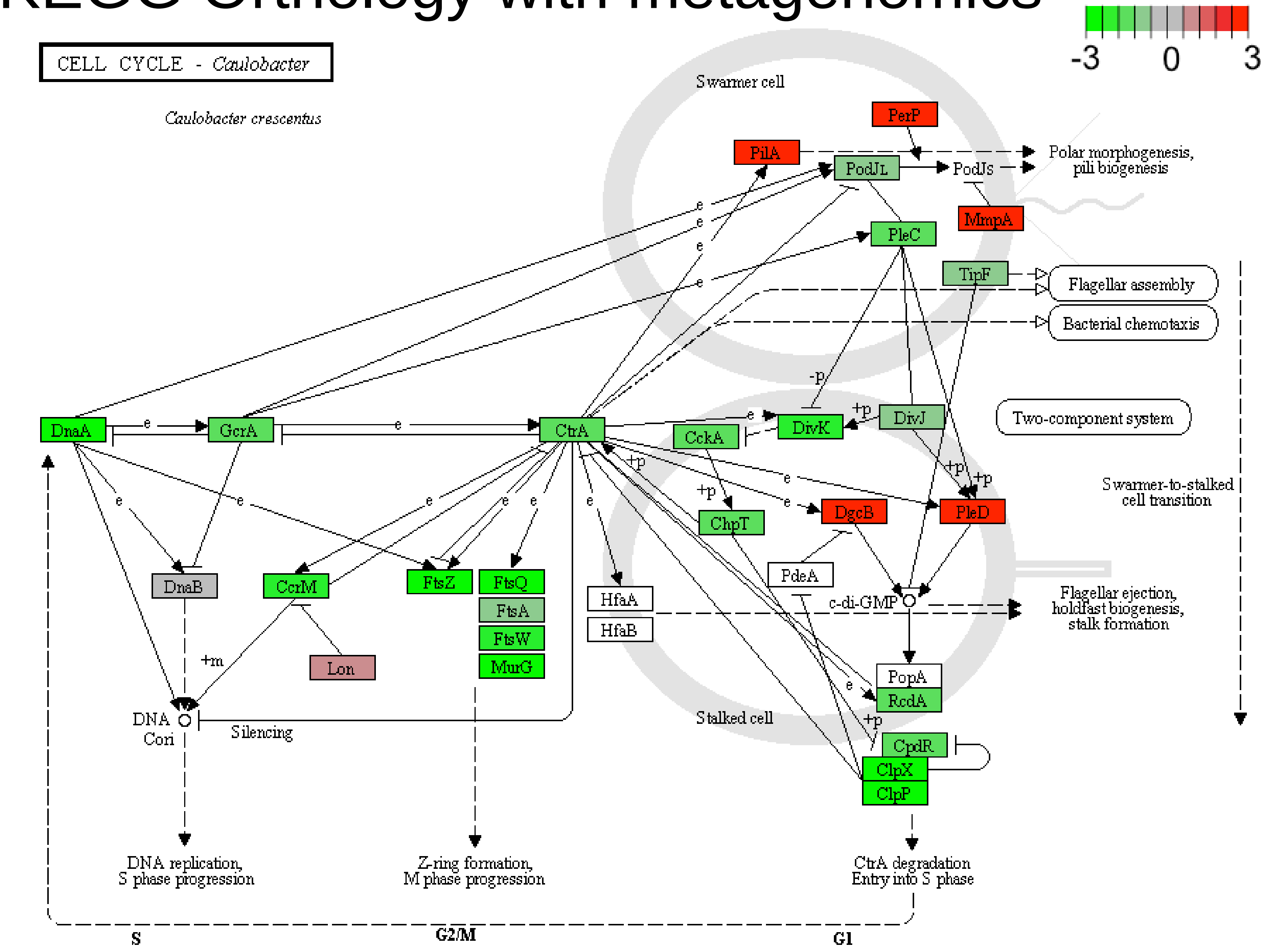


Data on KEGG graph
Rendered by Pathview

Examples of 3000 species supported

KEGG Code	**Scientific Name**	**Common Name**	**KEGG GeneID**	**Entrez GeneID**
hsa	*Homo sapiens*	human	100	100
mmu	*Mus musculus*	mouse	100042069	100042069
rno	*Rattus norvegicus*	rat	100359539	100359539
xla	*Xenopus laevis*	African clawed frog	100037030	100037030
dre	*Danio rerio*	zebrafish	100000978	100000978
dme	*Drosophila melanogaster*	fruit fly	Dmel_CG10078	38753
cel	*Caenorhabditis elegans*	nematode	CELE_AH6.1	191639
ath	*Arabidopsis thaliana*	thale cress	AT1G01210	839481
sce	*Saccharomyces cerevisiae*	budding yeast	YAL038W	851193
eco	*Escherichia coli* K-12 MG1655		b0049	944770

KEGG Orthology with metagenomics



Data on KEGG graph
Rendered by Pathview

Integrated workflows for pathway analysis

Pathway analysis workflows (with Pathview)

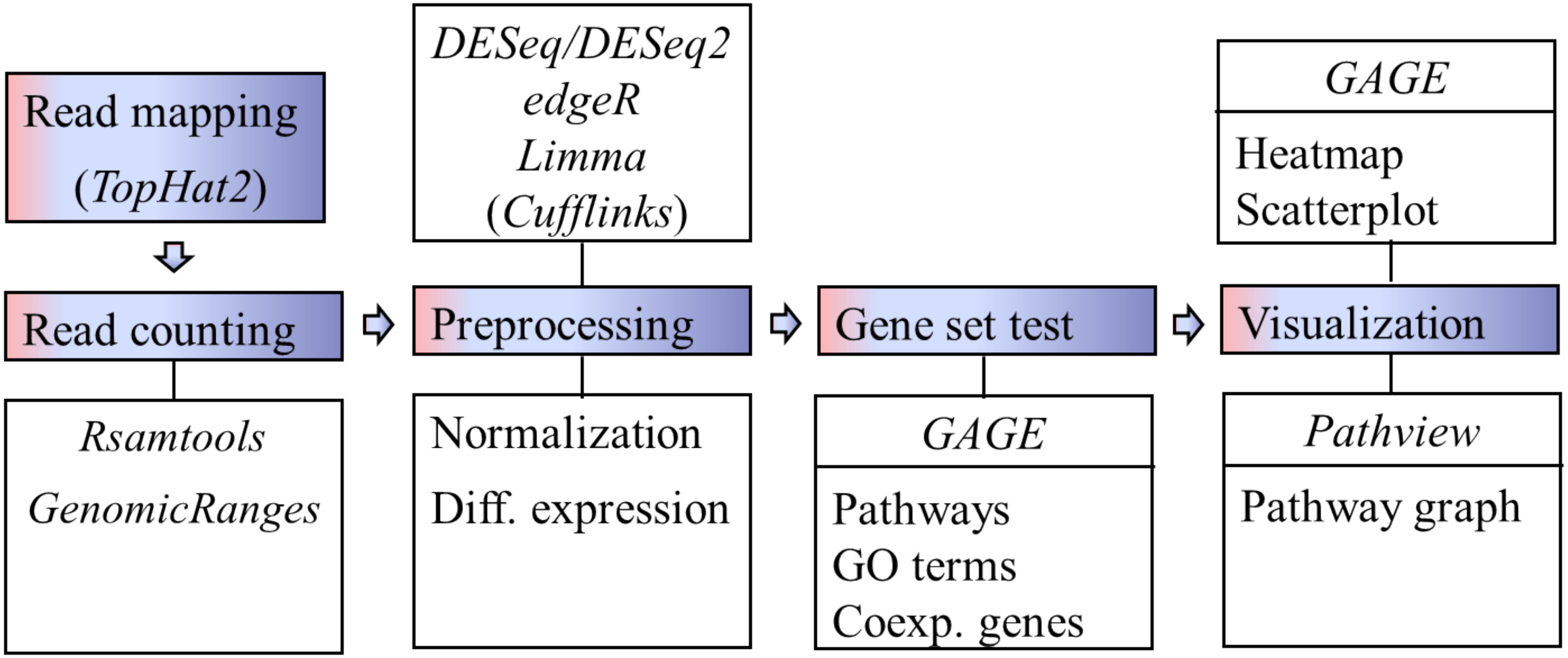
- Transcriptomics
  - Microarray
  - RNA-Seq
- Genomics
  - GWAS
  - CNV
- Metabolomics
- Proteomics
- Epigenomics

- Related tools:
  - GAGE
  - GSEA
  - clusterProfiler
  - opm
  - CompGO
  - GeneProf
  - …

RNA-Seq workflows with GAGE/Pathview

-Tutorial in BioC gage package
-Google: rna-seq pathway

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS



Read mapping (*TopHat2*)

Read counting

Rsamtools
GenomicRanges

Preprocessing

DESeq/DESeq2
edgeR
Limma
(*Cufflinks*)

Normalization
Diff. expression

Gene set test

GAGE
Pathways
GO terms
Coexp. genes

Visualization

GAGE
Heatmap
Scatterplot

Pathview
Pathway graph

Implementation & Impact

Software package: pathview

- Status
  - Current release/devel version: 1.4.1/1.5.2
  - Initial release: BioC 2.12 (R-3.0) (04/2013)
  - Multiple platforms, complete doc and help
  - Actively maintained and improved

- Available:
  - Google: pathview bioconductor

Global Impact

- Bioinformatics Most-Read (June-October, 2013).
- Related tools:
  - R/BioC: clusterProfiler, CompGO, gage, opm
  - Web: GeneProf
- 10,000 downloads past year
- Inquiries/refers:
  - emails, BioC list
  - SEQanswers, Biostars
  - online blogs, tutorials

Course adoptions

JOHNS HOPKINS MEDICINE

Techniques in Glycobiology
Prof. Jennifer Van Eyk, USA

MDI Biological Laboratory

Environmental Genomics
Prof. John Colbourne, USA

DRAGON STAR 龙星计划 PROGRAM

Cancer Bioinformatics
Prof. Ying Xu, China

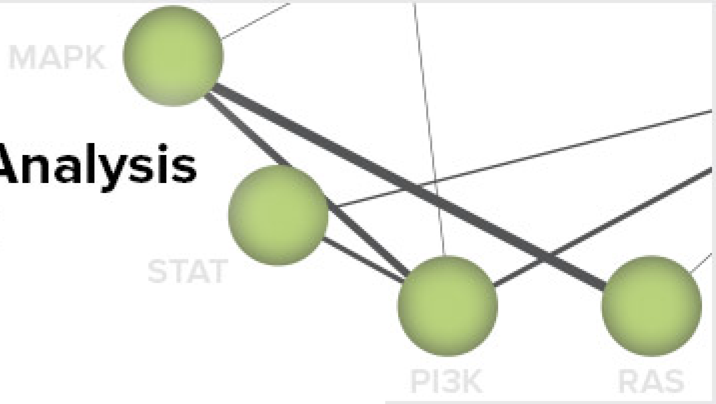
THE UNIVERSITY OF TOKYO

Functional genomics
Prof. Koji Kadota, Japan

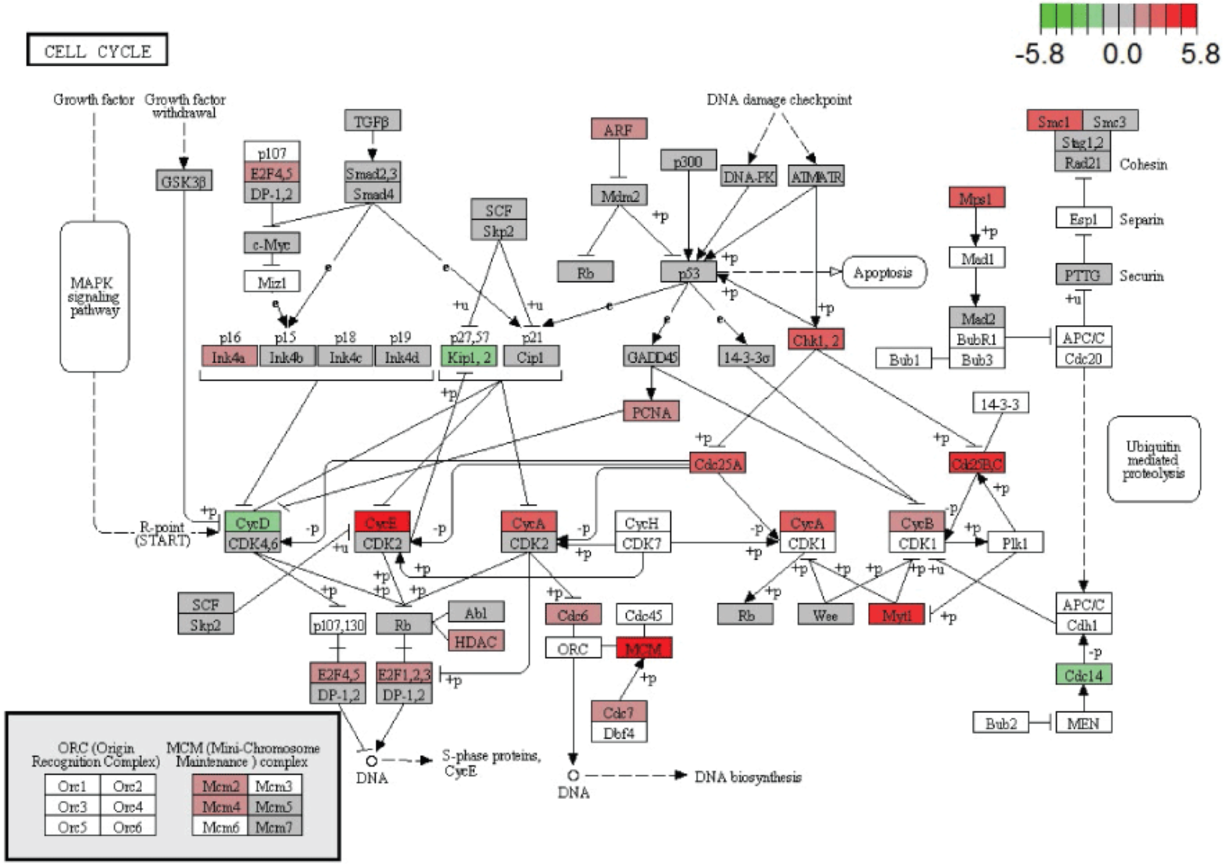
nanoString TECHNOLOGIES®

Multiplexed Cancer Pathway Analysis

nCounter PanCancer Pathways Panel
for Gene Expression



Differential Gene Expression Mapped to Proteins in the Cell Cycle Pathway



Data on KEGG graph
Rendered by Pathview

http://www.nanostring.com/forms/pancancer/download

Future Development

- R/Bioc interface improvement
- Multiple user interfaces
  - Web
  - Galaxy
- More node/edge attributes
- More pathway databases besides KEGG

Acknowledgements

UNC Charlotte
Bioinformatics Service Division
Dept. Bioinformatics & Genomics

R/Bioconductor teams
GAGE/Pathview users

All of you!

Questions/comments?
luo_weijun@yahoo.com

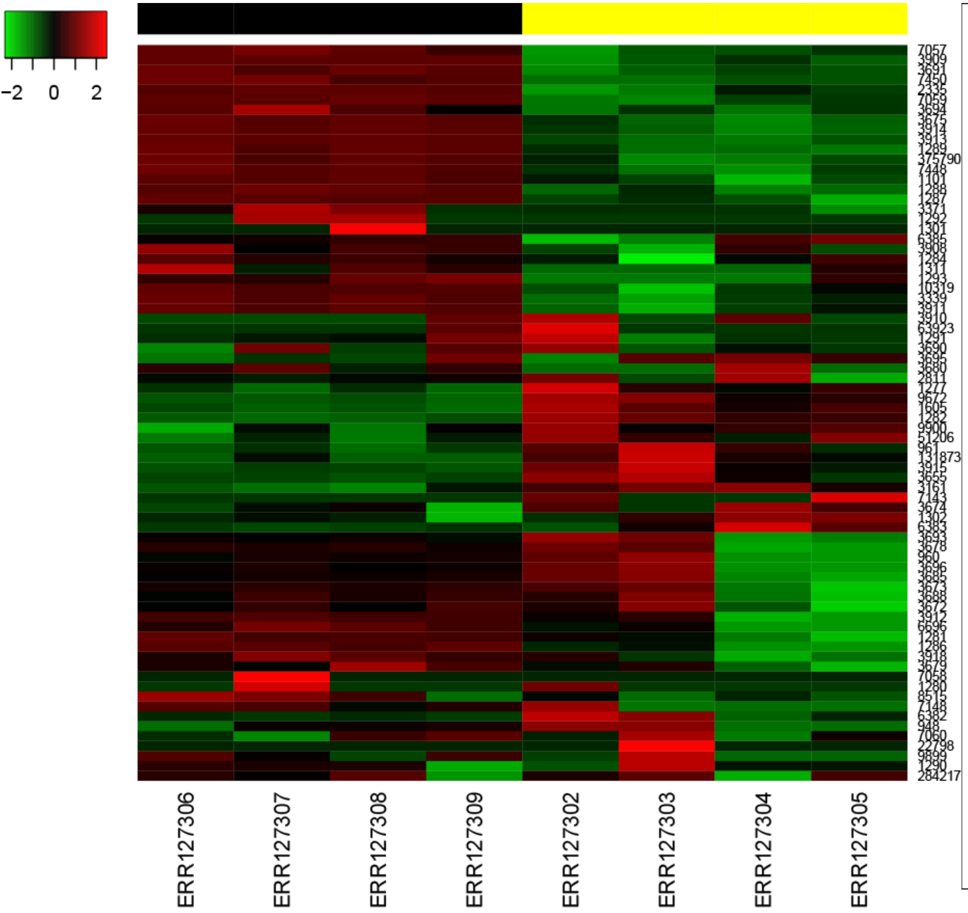
Data integration/processing capacity

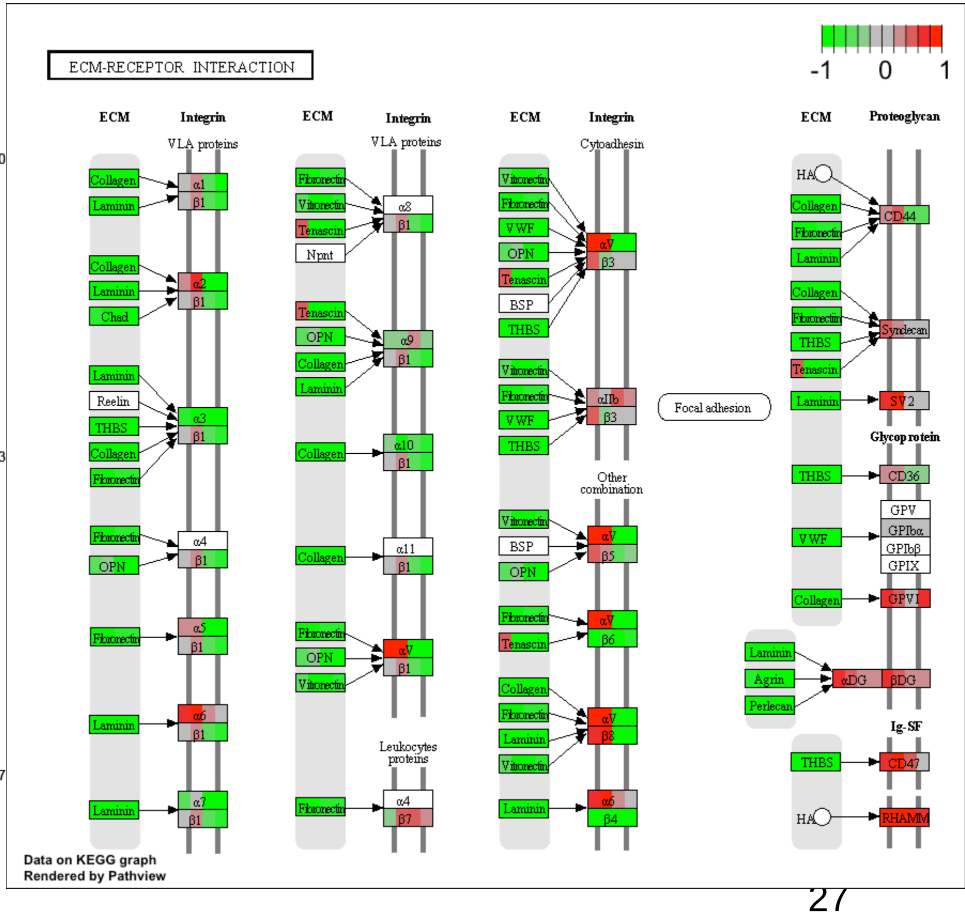
Data Attribute	**Pathview Coverage**
Type	Gene, protein, metabolite, genetics, literature, and others
ID type	12 gene, 21 compound
Species	3000 KEGG species and ortholog
Value	Continuous, discrete, 1 or 2 directions
Format	Matrix, vector

Pathway analysis visualized

Before

Now





Conclusion

Pathview

Luo *et al.* *Bioinformatics* 2013

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

- Informative, readable, publishable graphs
- Strong data integration
- Integrated pathway analysis for omics data
- Global impact, widely adopted