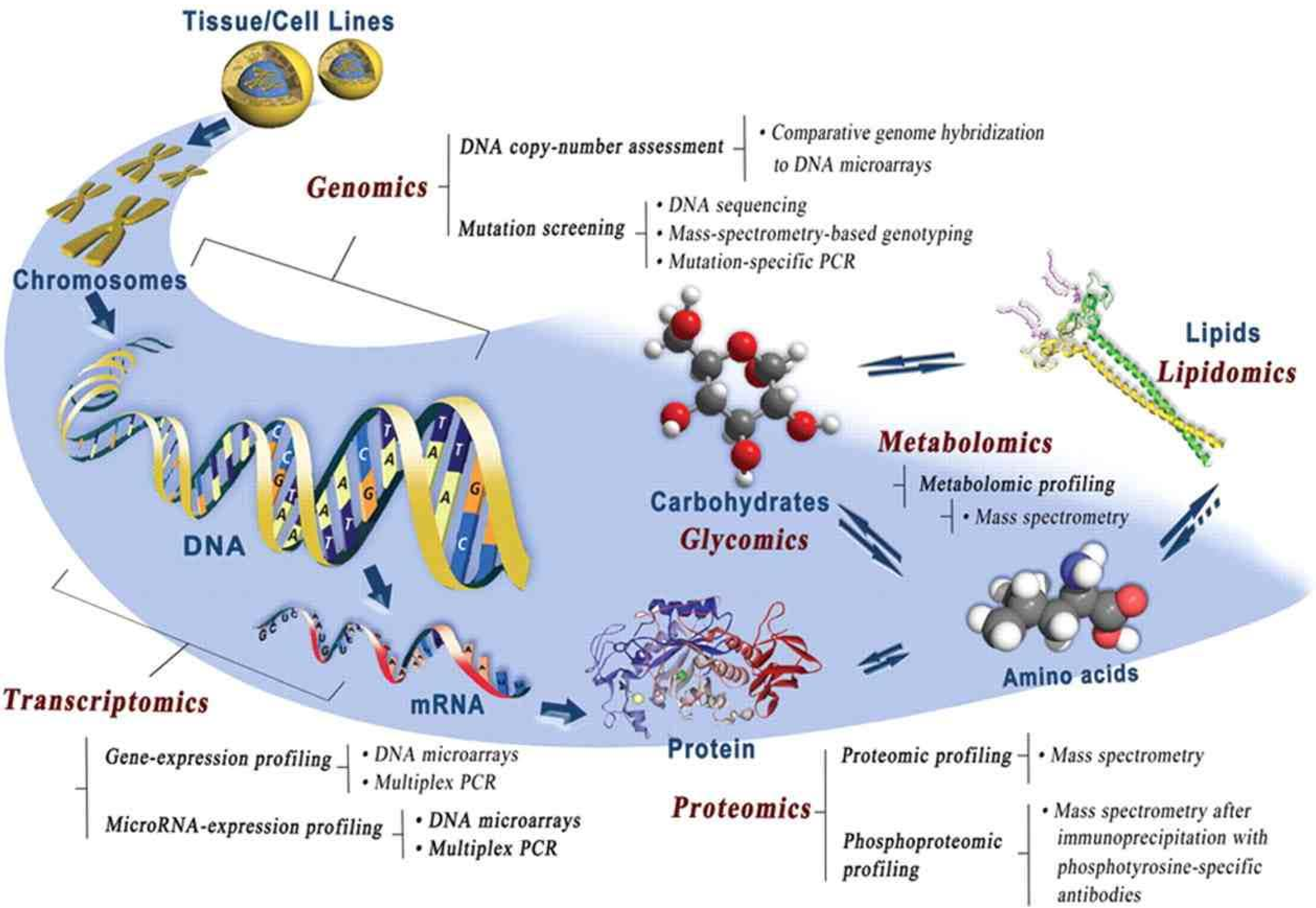
A Systems Biology Approach to Environmental Biology

Philipp Antczak

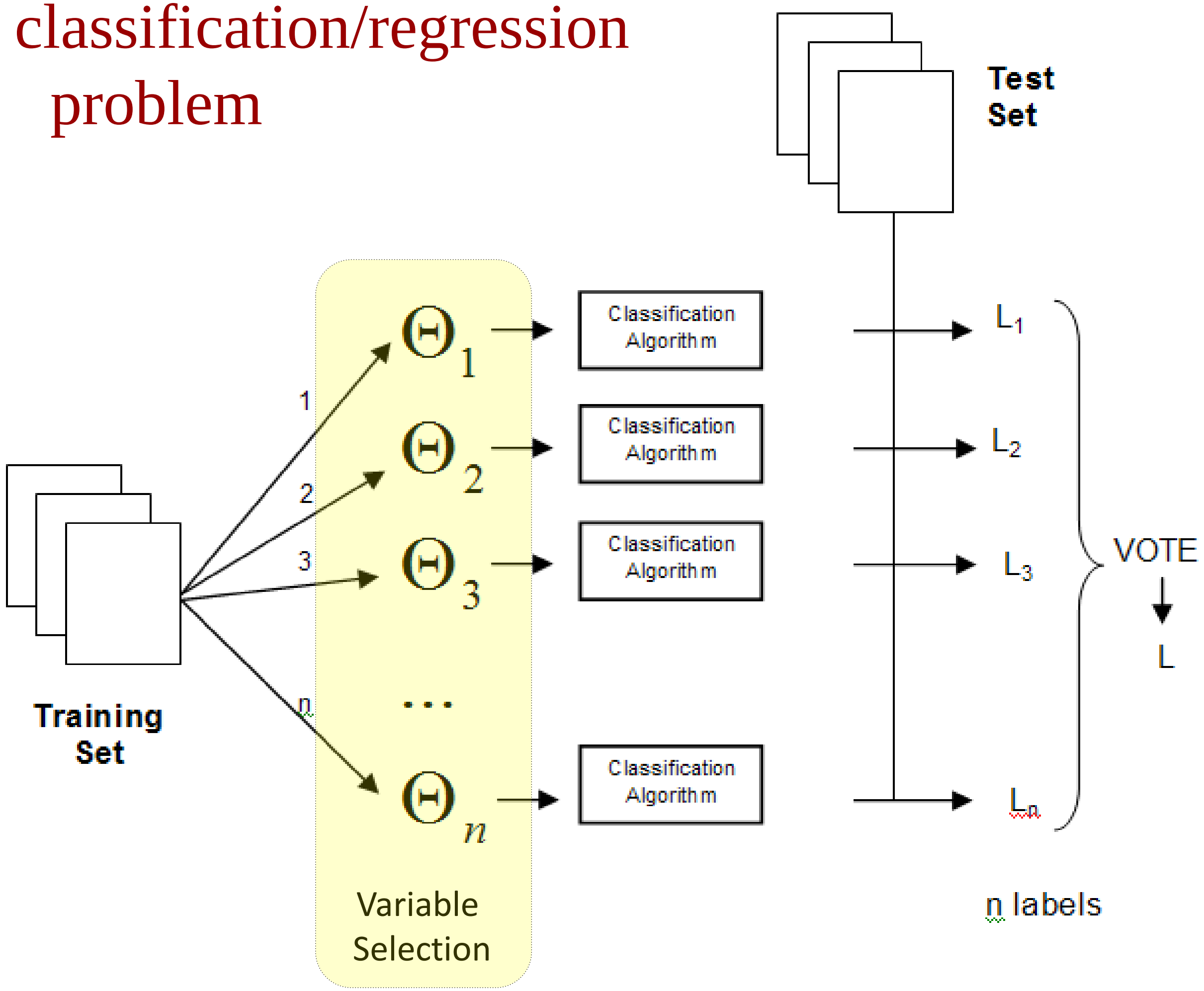
Why Systems Biology?



Phosphorylated p120-catenin expression has predictive value for oral cancer progression J. Clin. Pathol. April 1, 2012 65: 315-319

How do we deal with that much information?

A supervised classification/regression problem

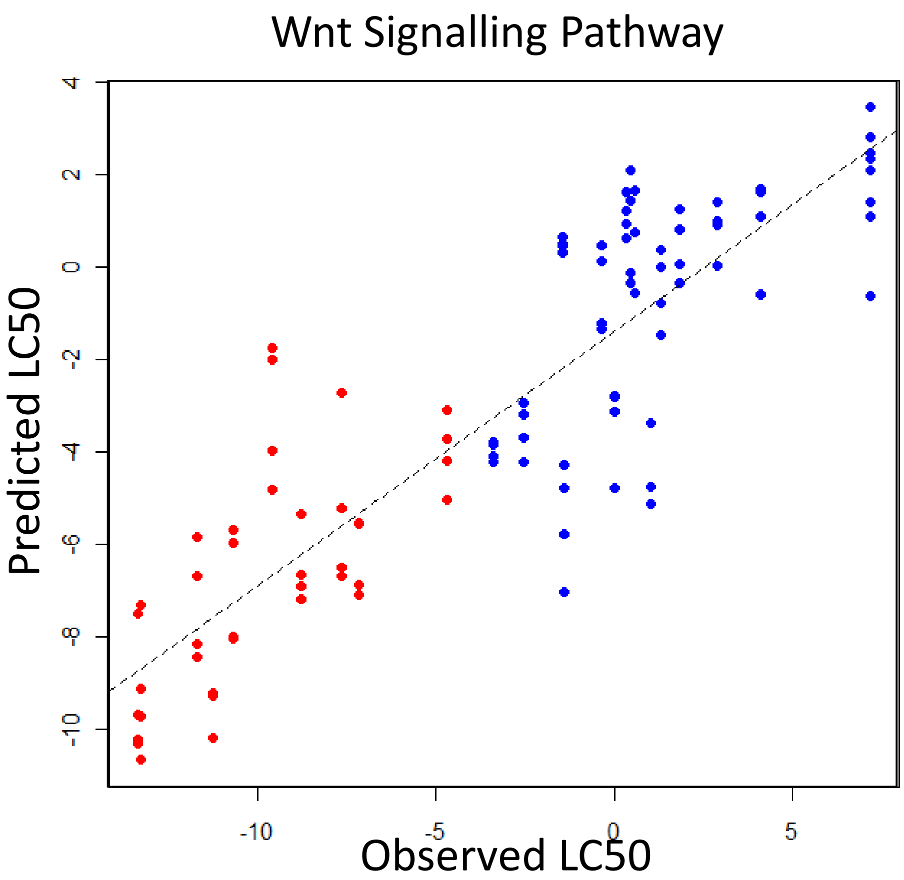


Genetic Algorithms and Bayesian Variable Selection

Trevino, V, & **F Falciani**, 'GALGO: an R package for multivariate variable selection using genetic algorithms.', Bioinformatics vol. 22, no. 9, 2006, pp. 1154-1156.

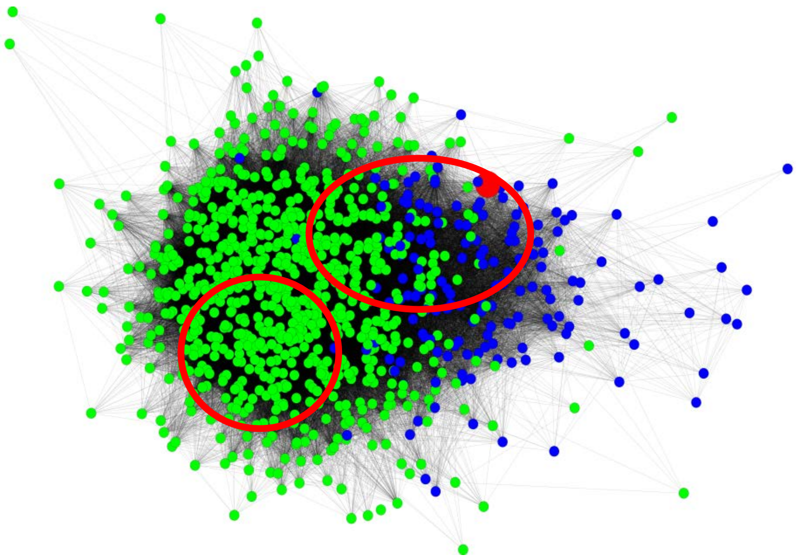
Sha N, Vannucci M, Tadesse MG, Brown PJ, Dragoni I, Davies N, Roberts TC, Contestabile A, Salmon M, Buckley C, **Falciani F**. Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. Biometrics. 2004 Sep;60(3):812-9.

Linking endpoints to molecular response

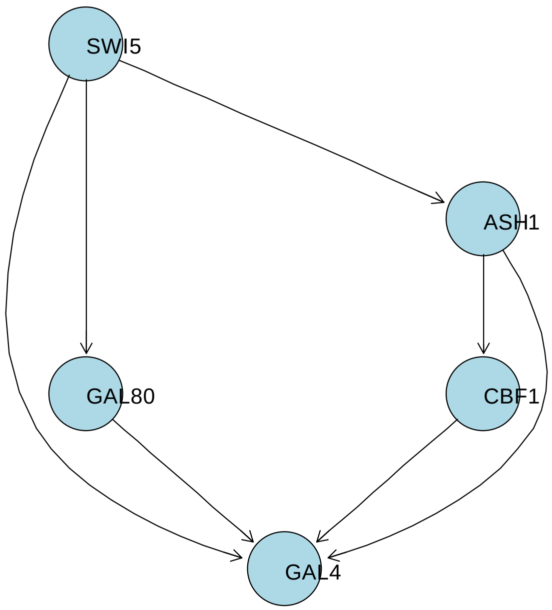


Regulatory networks

Static

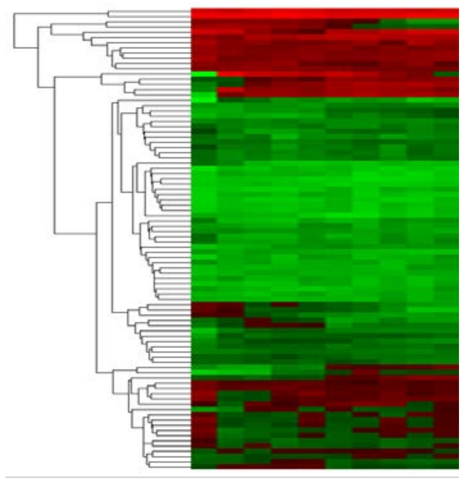


Dynamic

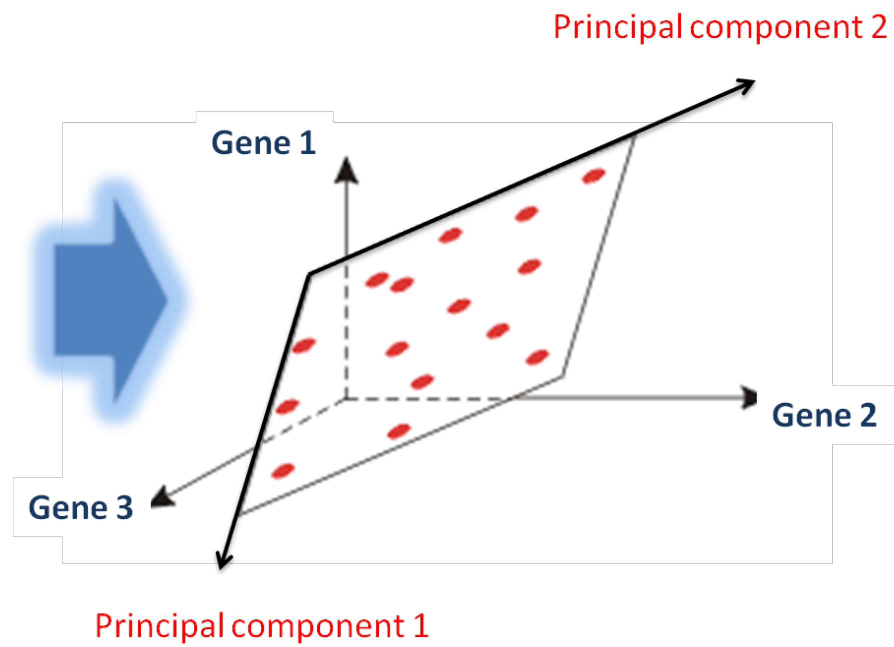


Gene-level analyses can be hard to interpret!

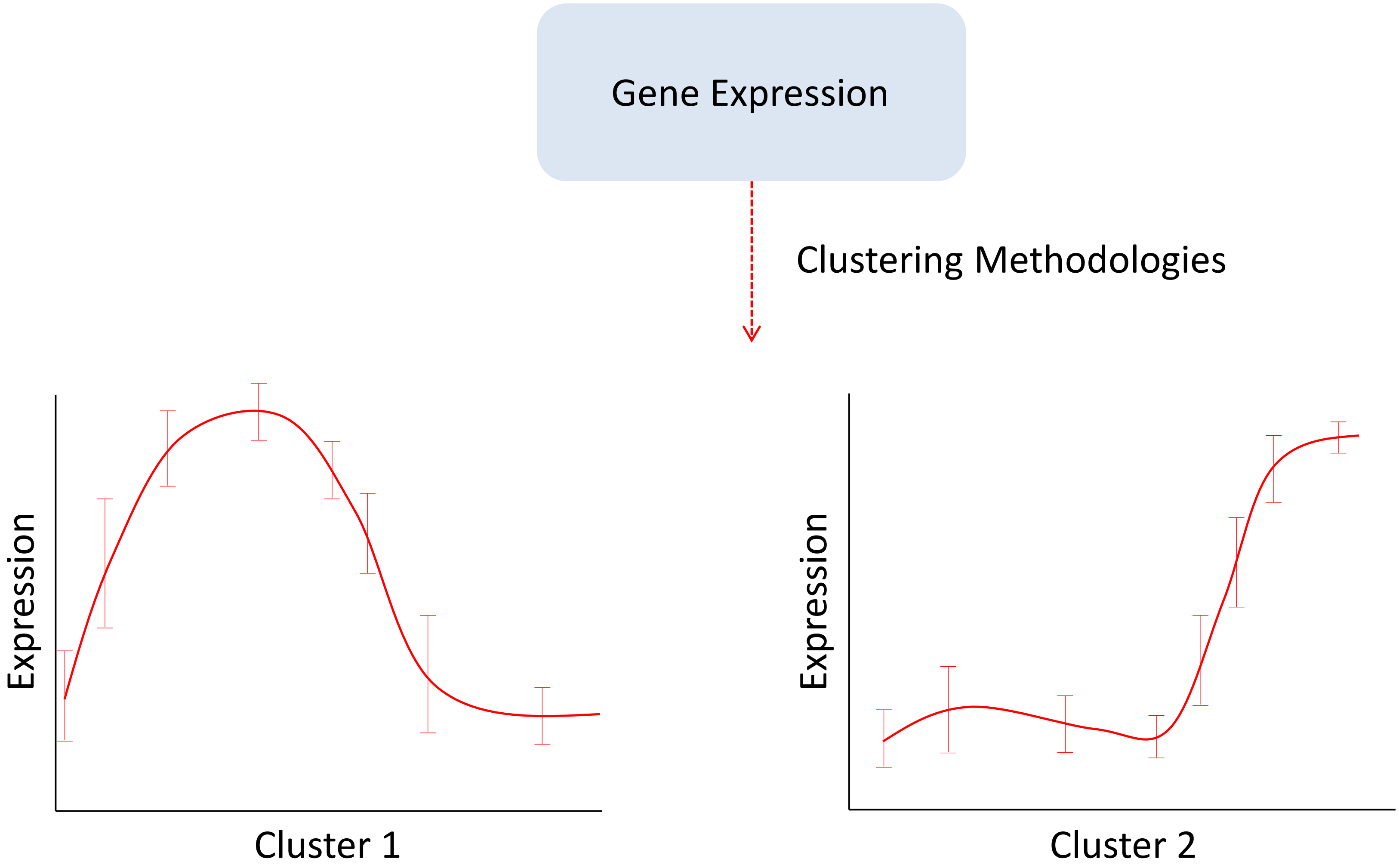
Simplifying the Problem by previous knowledge



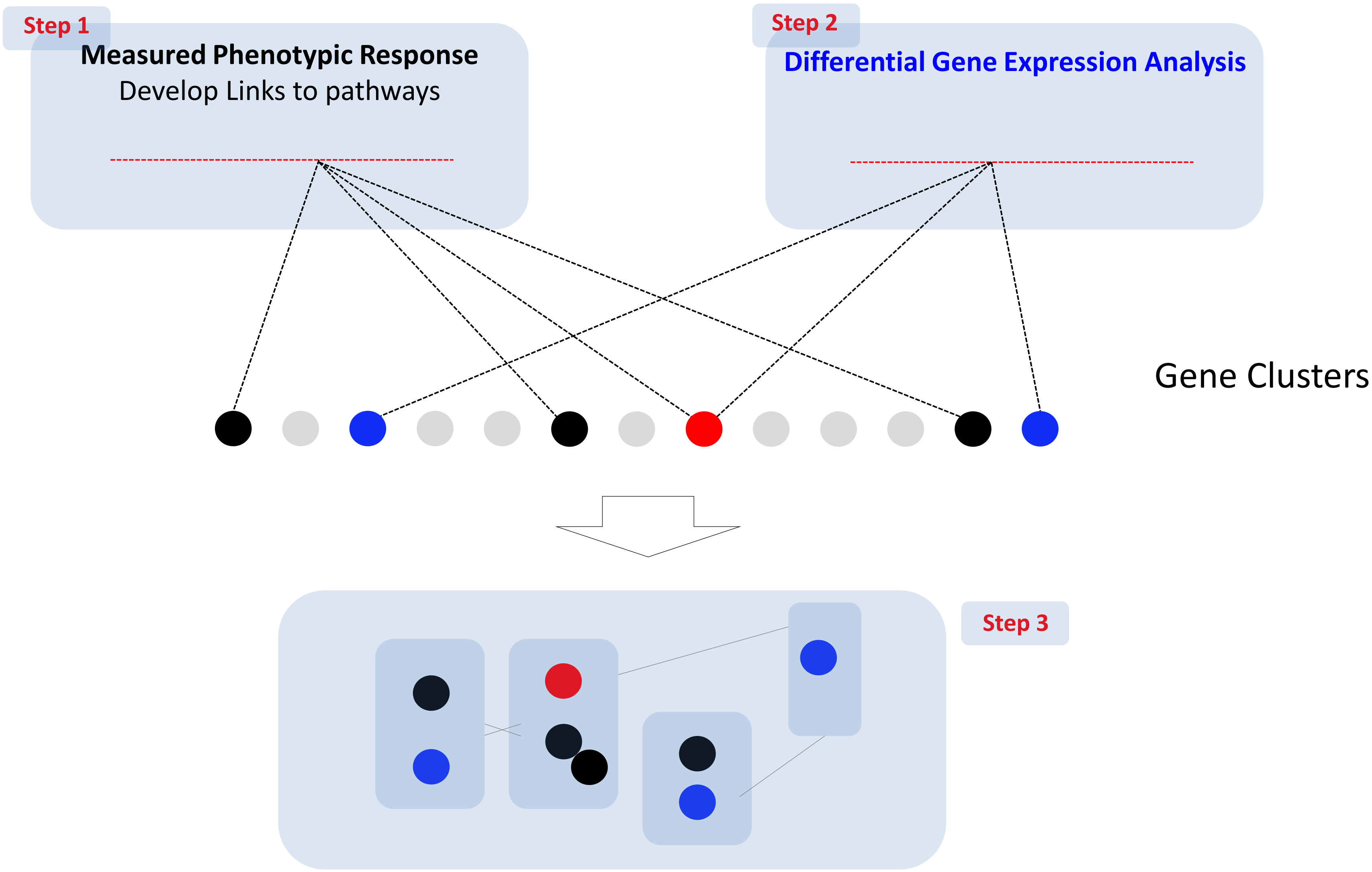
KEGG PATHWAY



Simplifying the problem by expression similarity

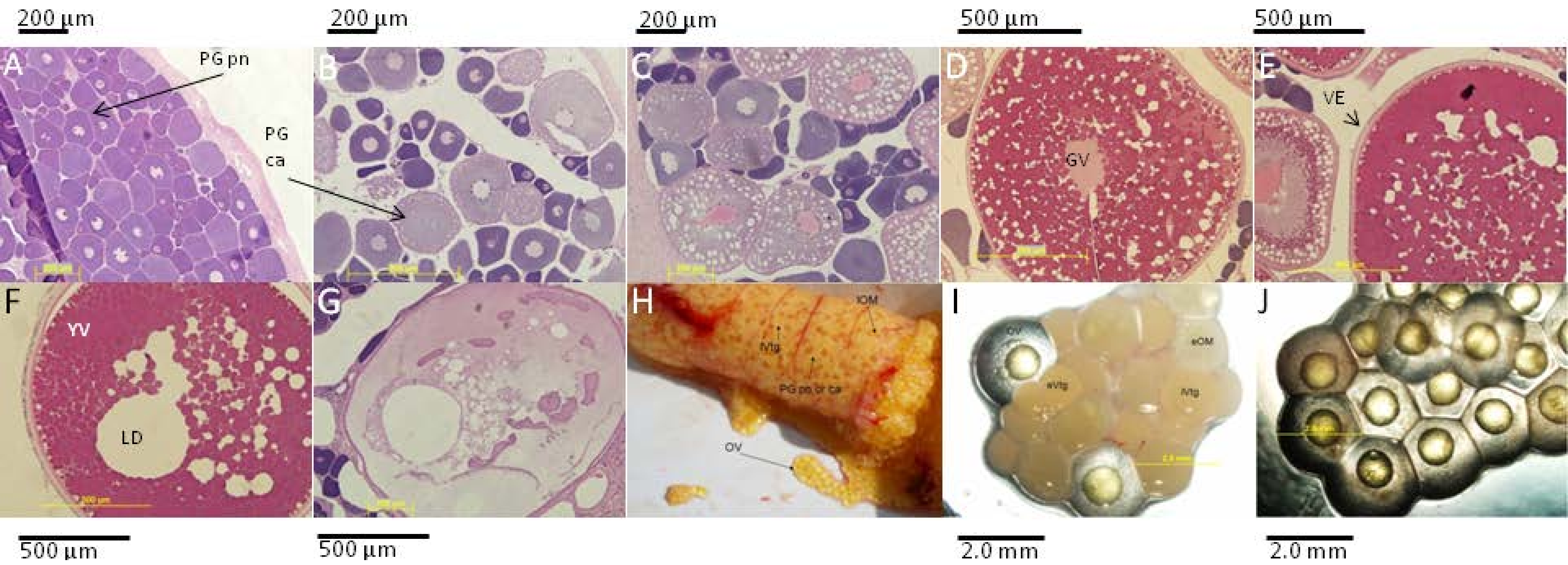


Combination into Workflows

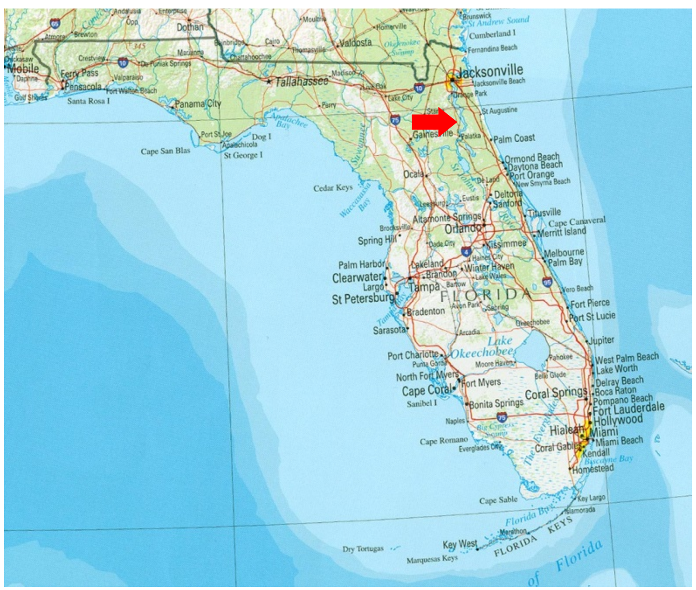


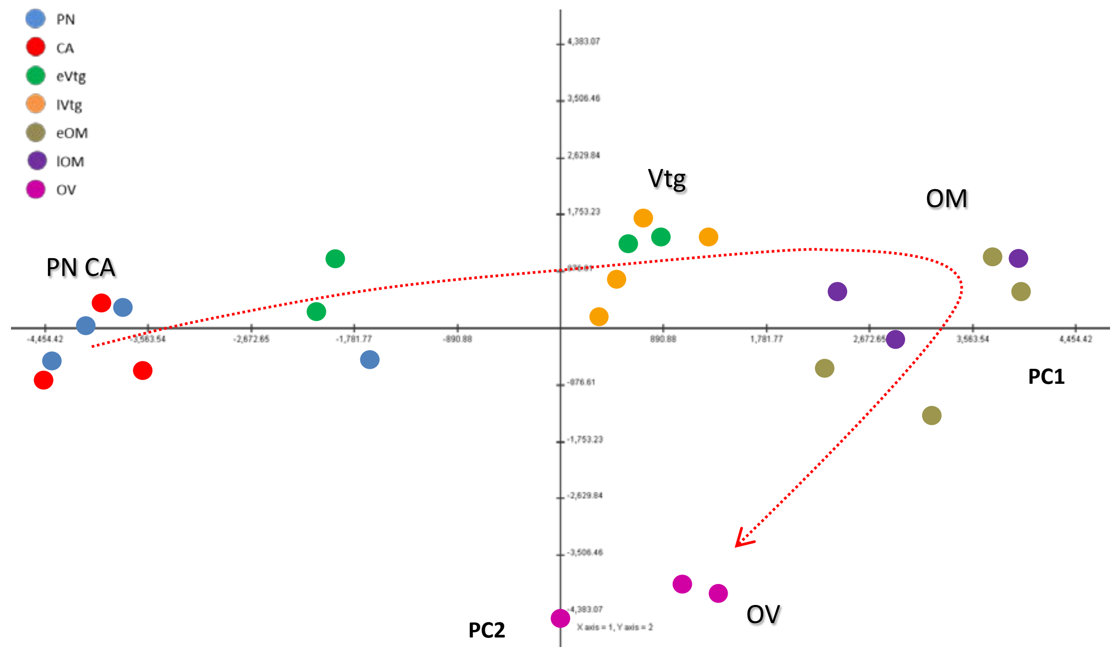
How can we apply these techniques in environmental biology?

Case Study – Ovarian Maturation

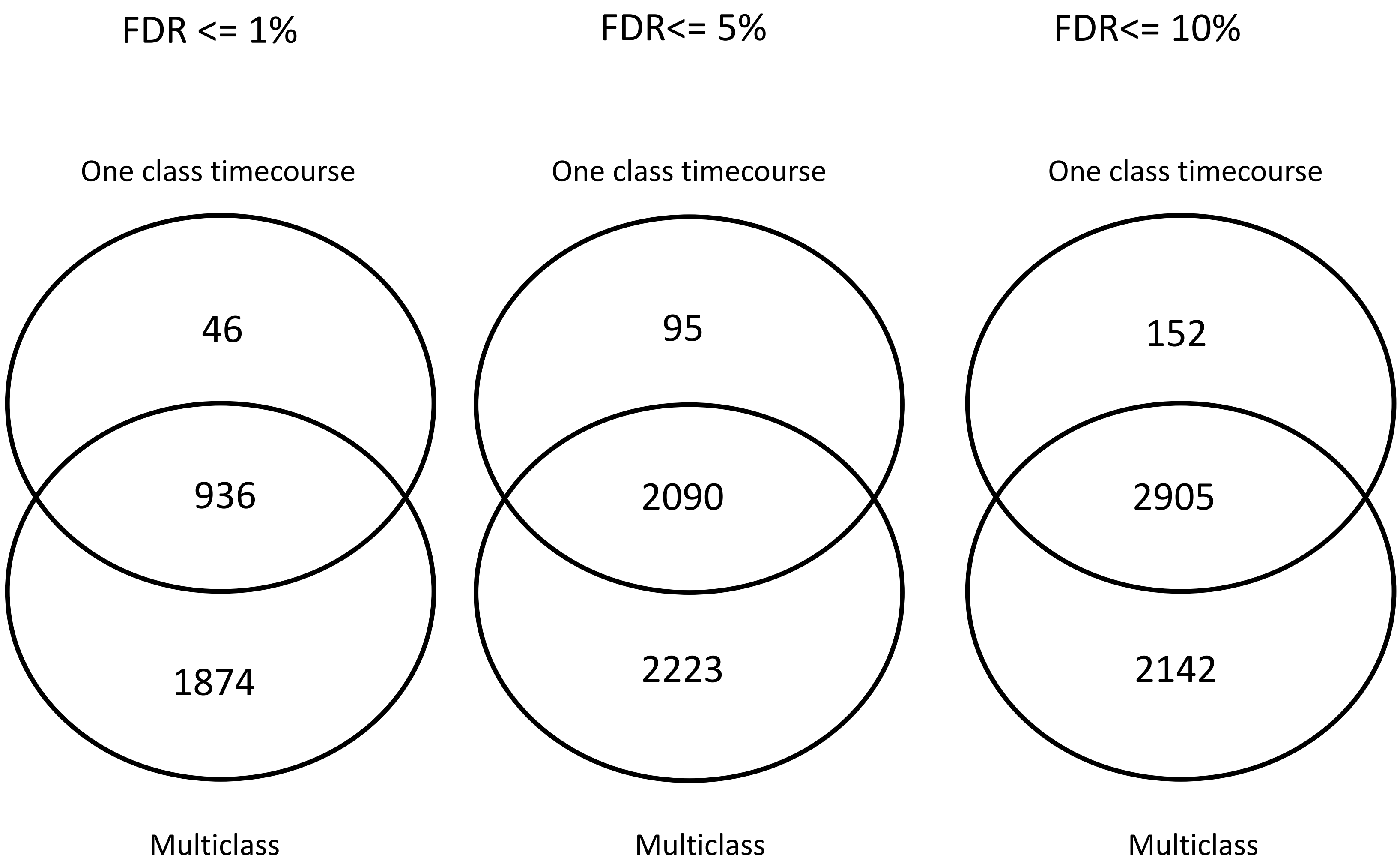


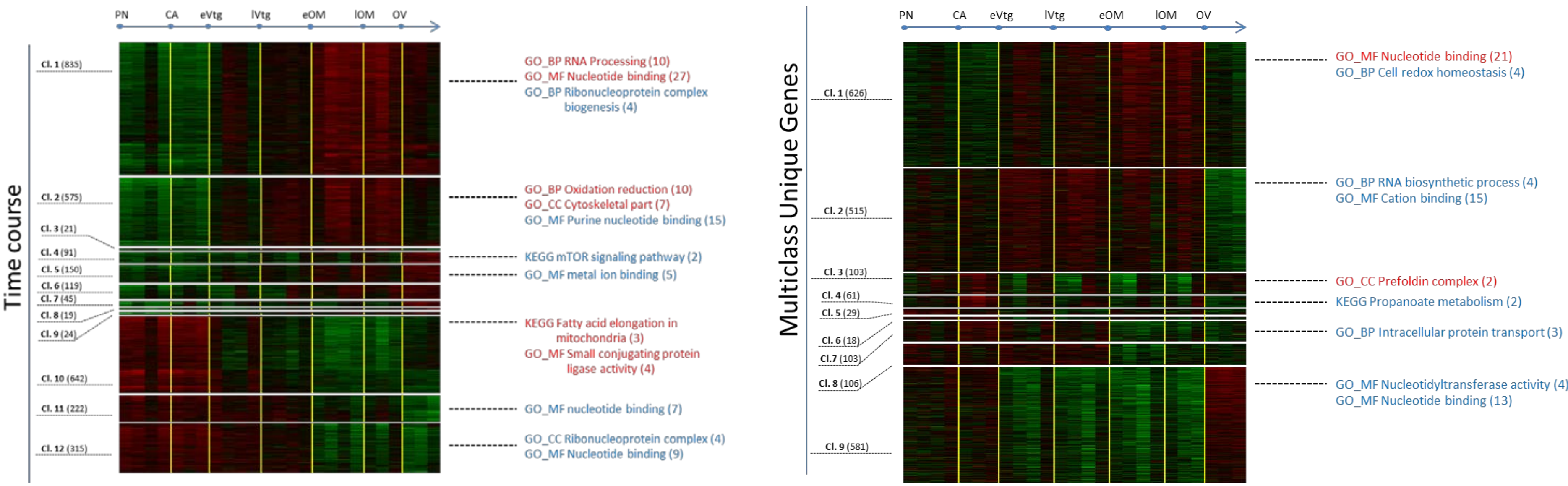
Martyniuk CJ, Prucha MS, Doperalski NJ, Antczak P, Kroll KJ, et al. (2013) Gene Expression Networks Underlying Ovarian Development in Wild Largemouth Bass (*Micropterus salmoides*). PLoS ONE 8(3): e59093. doi:10.1371/journal.pone.0059093



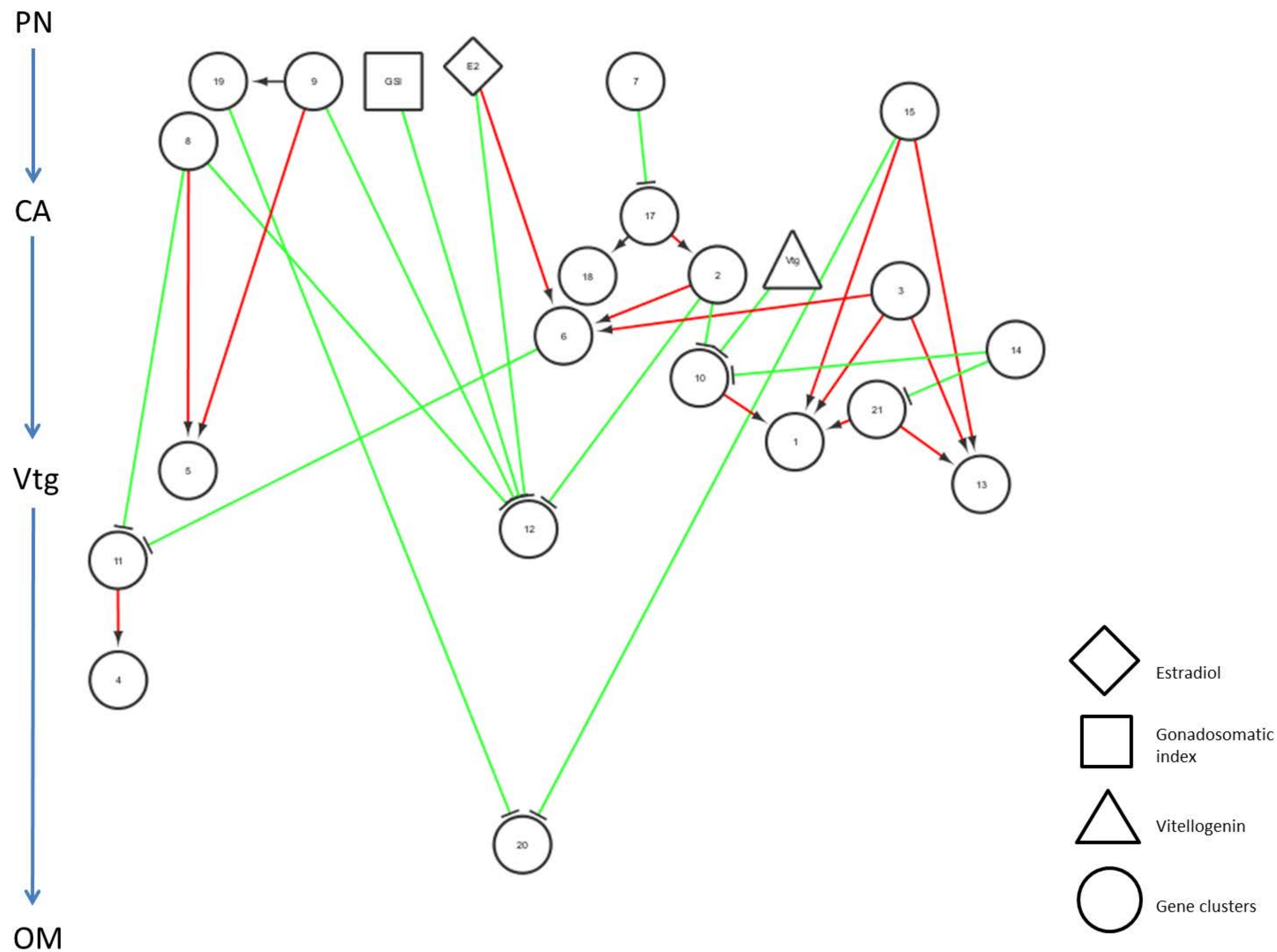


- Vitellogenin (Vtg), estradiol (E2) and Gonadosomatic index (GSI) measurments were taken at the sampling time.


FDR <= 1%
FDR<= 5%
FDR<= 10%
One class timecourse
One class timecourse
One class timecourse
46
936
1874
95
2090
2223
152
2905
2142
Multiclass
Multiclass
Multiclass

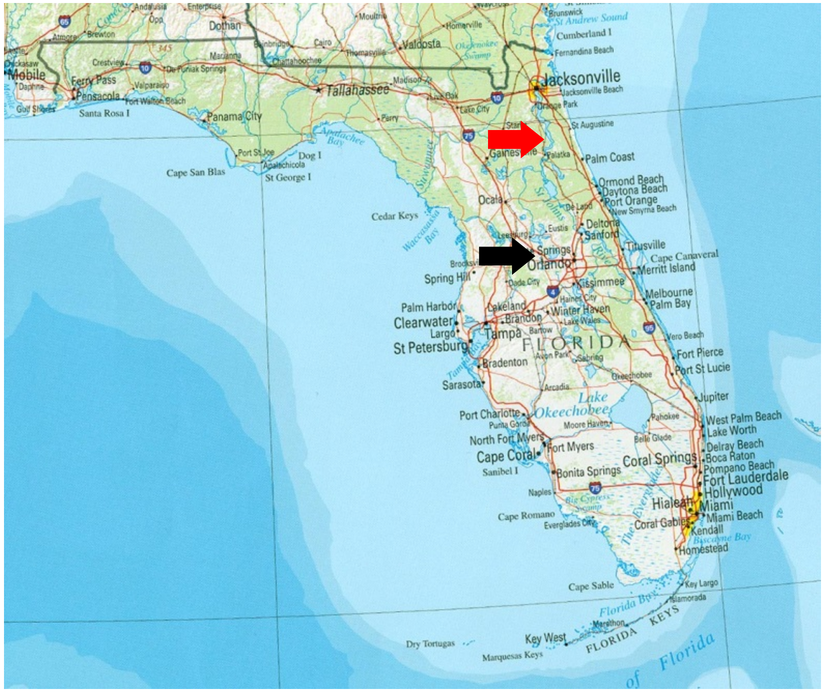


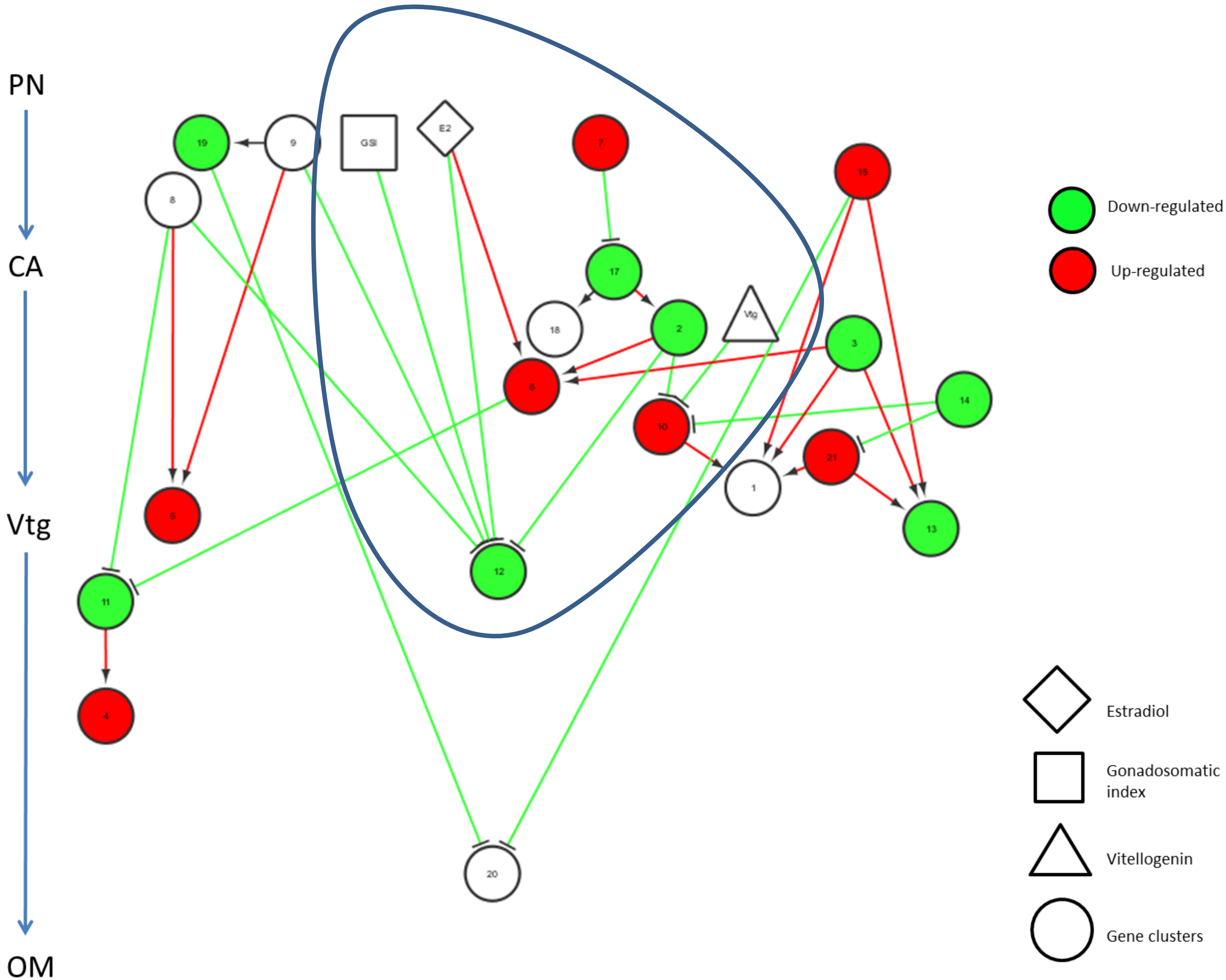
Time course
PN
CA
eVtg
lVtg
eOM
lOM
OV
Cl. 1 (835)
Cl. 2 (575)
Cl. 3 (21)
Cl. 4 (91)
Cl. 5 (150)
Cl. 6 (119)
Cl. 7 (45)
Cl. 8 (19)
Cl. 9 (24)
Cl. 10 (642)
Cl. 11 (222)
Cl. 12 (315)
GO_BP RNA Processing (10)
GO_MF Nucleotide binding (27)
GO_BP Ribonucleoprotein complex biogenesis (4)
GO_BP Oxidation reduction (10)
GO_CC Cytoskeletal part (7)
GO_MF Purine nucleotide binding (15)
KEGG mTOR signaling pathway (2)
GO_MF metal ion binding (5)
KEGG Fatty acid elongation in mitochondria (3)
GO_MF Small conjugating protein ligase activity (4)
GO_MF nucleotide binding (7)
GO_CC Ribonucleoprotein complex (4)
GO_MF Nucleotide binding (9)
Multiclass Unique Genes
PN
CA
eVtg
lVtg
eOM
lOM
OV
Cl. 1 (626)
Cl. 2 (515)
Cl. 3 (103)
Cl. 4 (61)
Cl. 5 (29)
Cl. 6 (18)
Cl.7 (103)
Cl. 8 (106)
Cl. 9 (581)
GO_MF Nucleotide binding (21)
GO_BP Cell redox homeostasis (4)
GO_BP RNA biosynthetic process (4)
GO_MF Cation binding (15)
GO_CC Prefoldin complex (2)
KEGG Propanoate metabolism (2)
GO_BP Intracellular protein transport (3)
GO_MF Nucleotidyltransferase activity (4)
GO_MF Nucleotide binding (13)

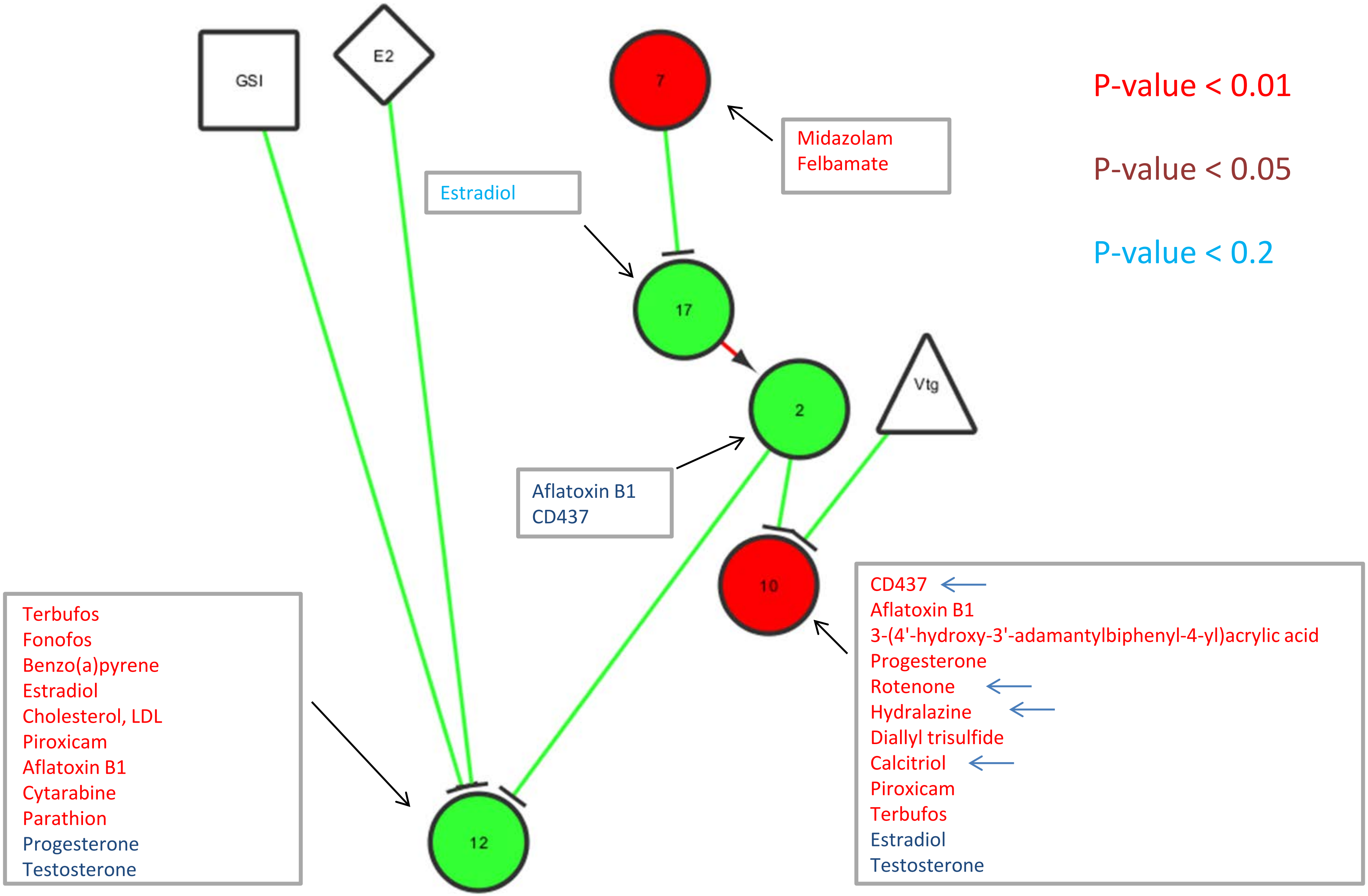


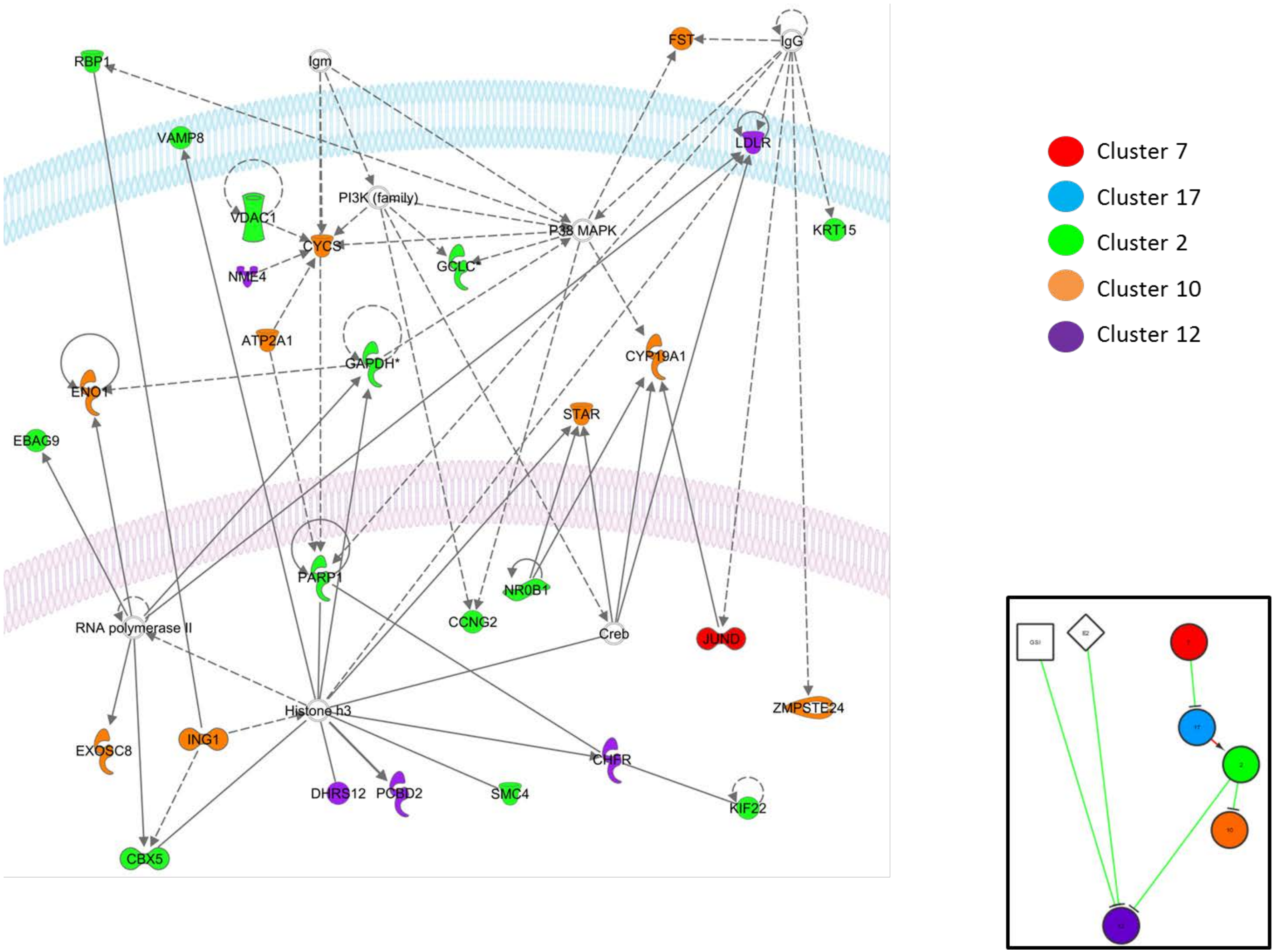
PN
CA
Vtg
OM
19
9
GSI
E2
7
15
8
17
18
2
Vtg
3
6
14
10
21
1
13
5
12
11
4
20
Estradiol
Gonadosomatic index
Vitellogenin
Gene clusters

How does pollution perturb this network?




PN
CA
Vtg
OM
19
9
GSI
E2
7
15
8
17
18
2
Vtg
3
6
14
10
21
1
5
13
12
11
4
20
Down-regulated
Up-regulated
Estradiol
Gonadosomatic index
Vitellogenin
Gene clusters


GSI
E2
7
Midazolam
Felbamate
P-value < 0.01
P-value < 0.05
P-value < 0.2
Estradiol
17
2
Vtg
Aflatoxin B1
CD437
10
CD437
Aflatoxin B1
3-(4'-hydroxy-3'-adamantylbiphenyl-4-yl)acrylic acid
Progesterone
Rotenone
Hydralazine
Diallyl trisulfide
Calcitriol
Piroxicam
Terbufos
Estradiol
Testosterone
Terbufos
Fonofos
Benzo(a)pyrene
Estradiol
Cholesterol, LDL
Piroxicam
Aflatoxin B1
Cytarabine
Parathion
Progesterone
Testosterone
12

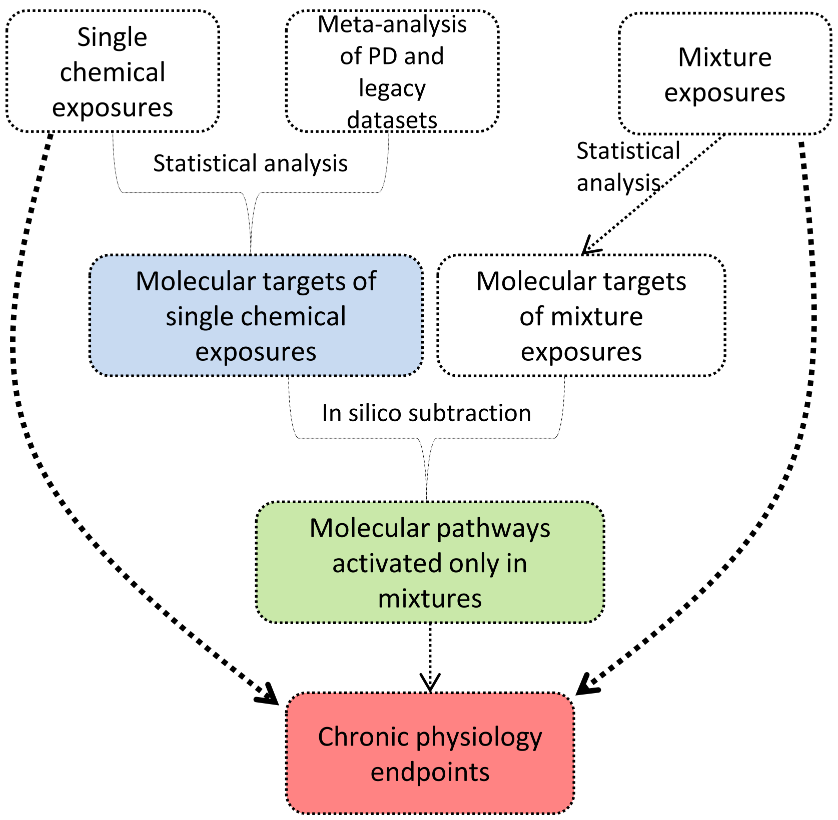

RBP1
Igm
IgG
FST
VAMP8
LDLR
VDAC1
PI3K (family)
P38 MAPK
KRT15
CYCS
NME4
GCLC*
ATP2A1
GAPDH*
CYP19A1
ENO1
STAR
EBAG9
PARP1
NR0B1
CCNG2
Creb
JUND
RNA polymerase II
Histone h3
ZMPSTE24
EXOSC8
ING1
CHFR
DHRS12
PCBD2
SMC4
KIF22
CBX5
Cluster 7
Cluster 17
Cluster 2
Cluster 10
Cluster 12

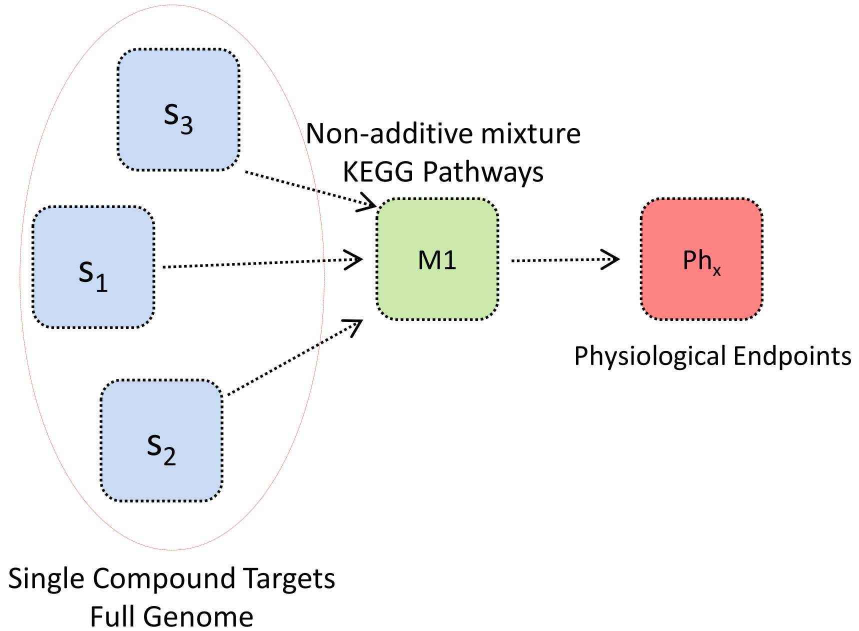
Discovering Adverse Outcome Pathways from molecular data

Example of complexity

DEET
lenacil
bisphenol A
musk ambrette (artificial)
pirimiphos methyl
benzo(ghi)perylene
2,2',3',4,4',5-hexachlorobiphenyl
flutolanil
cyhalothrin-lambda
Benzo(a)pyrene
Permethrin
cyprodinil
naphthalene
tebuconazole
2,4,5,2',5'-pentachlorobiphenyl
indeno(1,2,3-cd)pyrene
BAS 490 F
deltamethrin
Chlorpyrifos
cashmeran
iprodione
anthracene
benzo(b)fluoranthene
pendimethalin
phenanthrene
benzo(k)fluoranthene
aclonifen
galaxolide
ethofumesate
clomazone
etridiazol
Clotrimazole
ethoprop
tonalide (AHTN)
Dimethoate
9,10-anthraquinone
chrysene
metazachlor
terbutryne
tris(chloroethyl)phosphate
pendimethalin
Propoxur
tri-(2-chloroisopropyl)phosphate
Estrone
pyrene
tributyl phosphate
prosulfocarb
bifenox
Diethylstilbestrol
cyproconazole
Chlorpropham
tris(2-butoxyethyl) phosphate
triphenyl phosphate
tris(1,3-dichloro-2-propyl)phosphate
PCB 180
Iopanoic Acid
fluorene
4-deethylatrazine
penconazole
moskene
isopropyl 4,4'-dibromobenzilate
pirimicarb
tri-(2-chloroisopropyl)phosphate
2,4,5,2',4',5'-hexachlorobiphenyl
2,3',4,4',5-pentachlorobiphenyl
4-chloro-N-((4-(1,1-dimethylethyl)phenyl)methyl)-3-ethyl-1-methyl-1H-pyrazole-5-carboxamide
Gemfibrozil
metamitron
cyproconazole
metolachlor
tri-(2-chloroisopropyl)phosphate
dodemorph
pronamide
2-cyano-N-((ethylamino)carbonyl)-2-(methoxyimino)acetamide
dimethenamid
Caffeine
cycloxydim
diflufenican
desmedipham
metribuzin
terbutylazine
acetyl tert-butyl dimethylindan
1,2,5,6-dibenzanthracene
cyproconazole
benz(a)anthracene
fenpropimorph
acenaphthylene
trifloxystrobin
Carbofuran
Linuron
traseolide (ATTI)
fluoranthene
tris(2-ethylhexyl)phosphate
pyrazon
dodemorph
musk xylene
isoproturon
pyrimethanil
metalaxyl
Simazine
Diuron
2,4,4'-trichlorobiphenyl
pencycuron
acenaphthene
musk ketone
Estrone
dichlobanil
phantolid
2,5,2',5'-tetrachlorobiphenyl
Triallate
Fenofibrate
Permethrin
Dimethoate

High Level approach to pAOPs

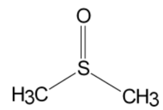




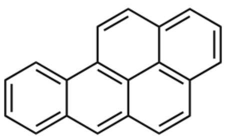
Experimental System

Stickleback (*Gasterosteus aculeatus*):

- widespread
- native UK species
- annual reproductive cycle
- Cefas experience
- microarray and biomarkers developed
- large enough to dissect tissues
- small enough to maintain in the laboratory
- well annotated draft genome sequence



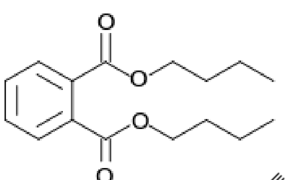
DMSO: 88 mg/l (0.008%)
Solvent



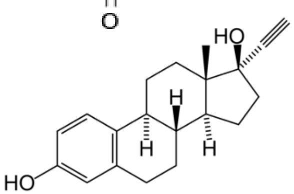
Benzo(a)pyrene: 10μg/l
PAH
LC50: 1200, HEC: 96 μg/l

Cd^{2+}

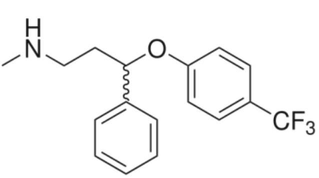
Cadmium: 65μg/l
Heavy metal
LC50: 6500, HEC: 4000 μg/l



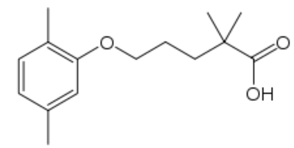
Dibutyl phthalate: 35μg/l
Plasticizer
LC50: 350, HEC : 170 μg/l



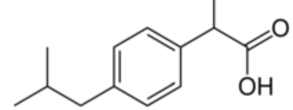
Ethinyl estradiol: 0.06 μg/l
Endocrine disrupter
LC50: 1600, HEC 0.04 μg/l



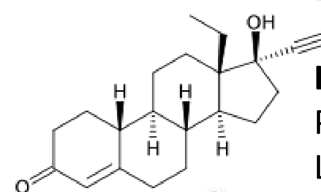
Fluoxetine: 10μg/l
SSRI antidepressant
LC50: 700, HEC 1 μg/l



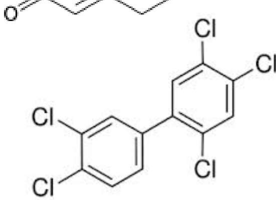
Gemfibrozil: 50 μg/l
Fibrate
LC50: 22000, HEC : 5 μg/l



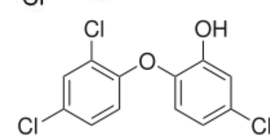
Ibuprofen: 50 μg/l
Painkiller
LC50: 7100, HEC : 28 μg/l



Levonorgestrel: 0.05 μg/l
Progestin
LC50: 6500, HEC : 0.015 μg/l



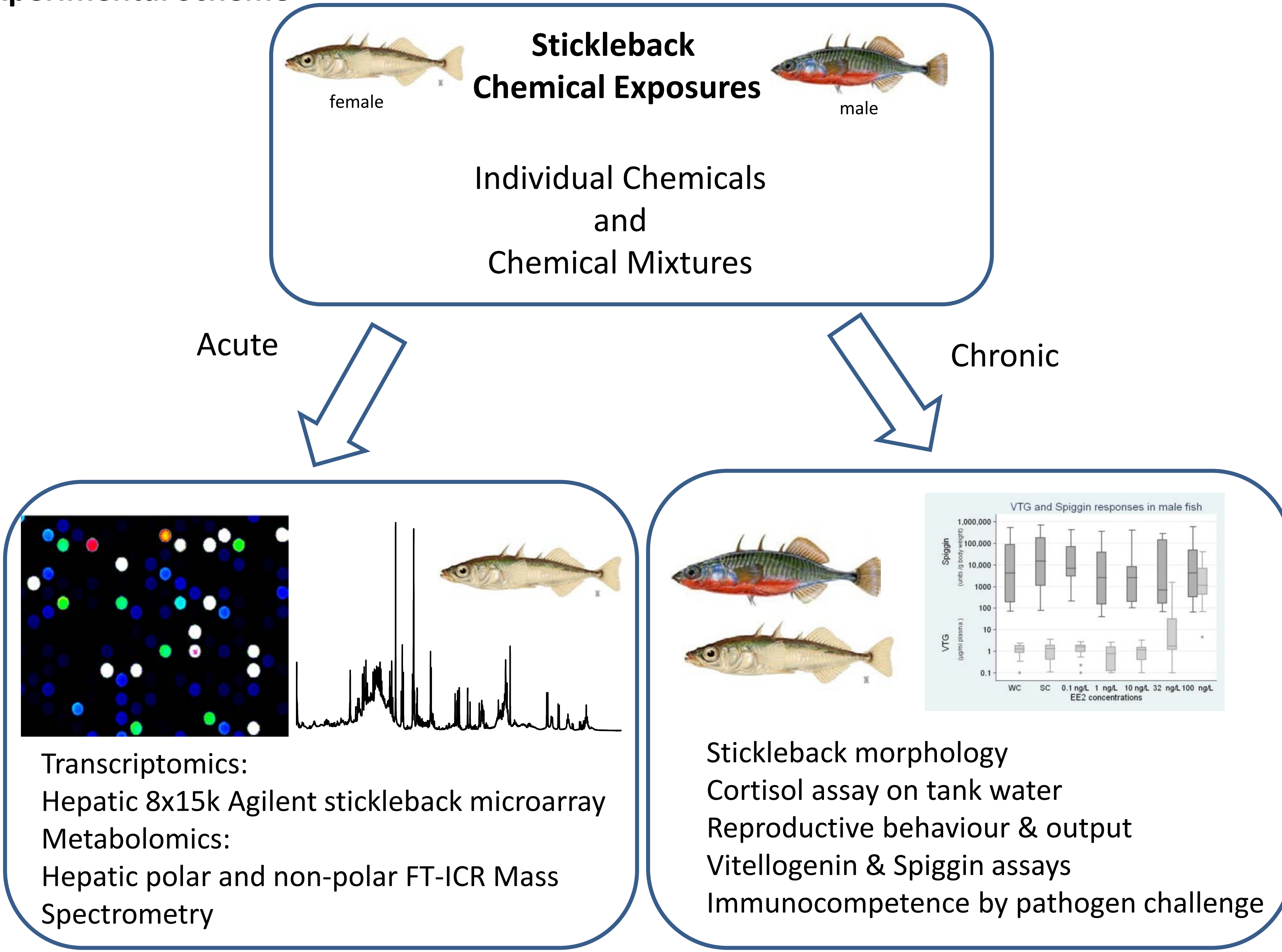
PCB-118: 1 μg/l
PCB
LC50: 15 μg/l, HEC : 123 μg/kg (sed)



Triclosan: 20 μg/l
Antibacterial/fungal
LC50: 260, HEC : 5 μg/l

LC50: Lowest found for stickleback or most sensitive fish species. HEC: Highest environmental concentration found

Experimental Scheme



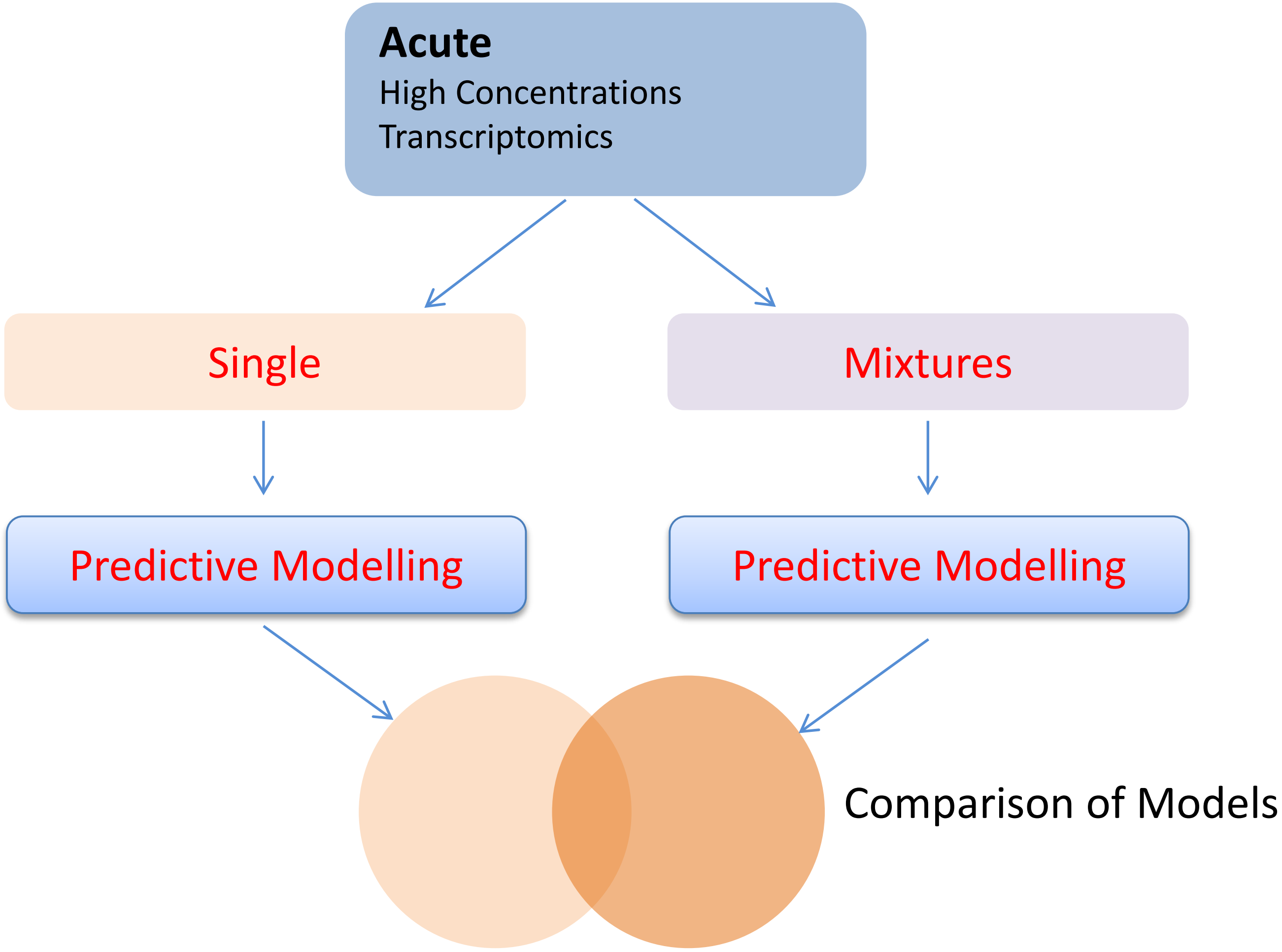
Exposures:

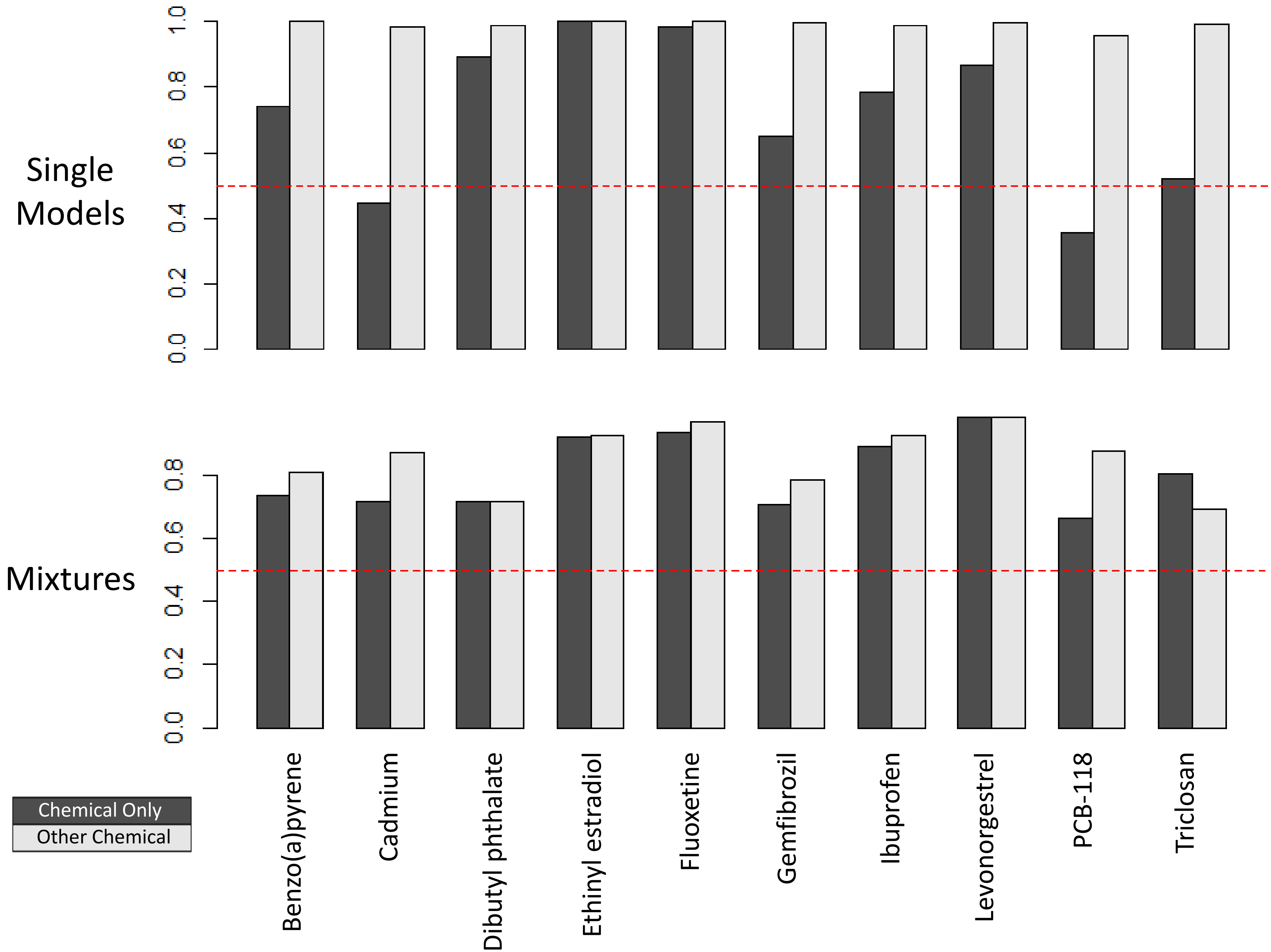
- •Each of 10 compounds singly, plus solvent
- •25 mixtures of 5 components plus solvent, one of all 10
- •10 sticklebacks per tank (mixed male and female)
- •Solvent and water controls
- •Duplicate tanks for each exposure = 80 tanks with 800 fish
- •Acute = 4 day exposure (complete)
- •800 sticklebacks sexed, livers dissected and frozen at -80C
- •Chronic = 4 months (2014)
- •Chemical analysis: Passive samplers (selected tanks; 2013-14)

Exposure	WC	SC	BaP	Cd	DBP	EE2	Fluo	Gem	Ibu	Levo	PCB	Tri	V01	V02	V03	V04	V05	V06	V07	V08	V09	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26
Solvent	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Benzo[a]pyrene	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	1	0	0	1	0	0	0	1	1
Cadmium	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	0	1	0	1	0	1
Dibutyl phthalate	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1
Ethinyl-oestradiol	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	1	1	1	0	1	1	1	0	1	1	0	0	1	0	1	1	0	1	1	1
Fluoxetine	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	0	0	0	0	1	1	0	1
Gemfibrozil	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0	1	0	1	0	0	1	0	0	1
Ibuprofen	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	0	0	1
Levonorgestrel	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1	1	0	1	1	1	1	0	0	1	1	1	1	1	0	0	1	1	1	1
PCB-118	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	1	1	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1
Triclosan	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	0	1	0	1	0	1	0	1	1	1	0	0	1	0	0	0	0	0	1	1

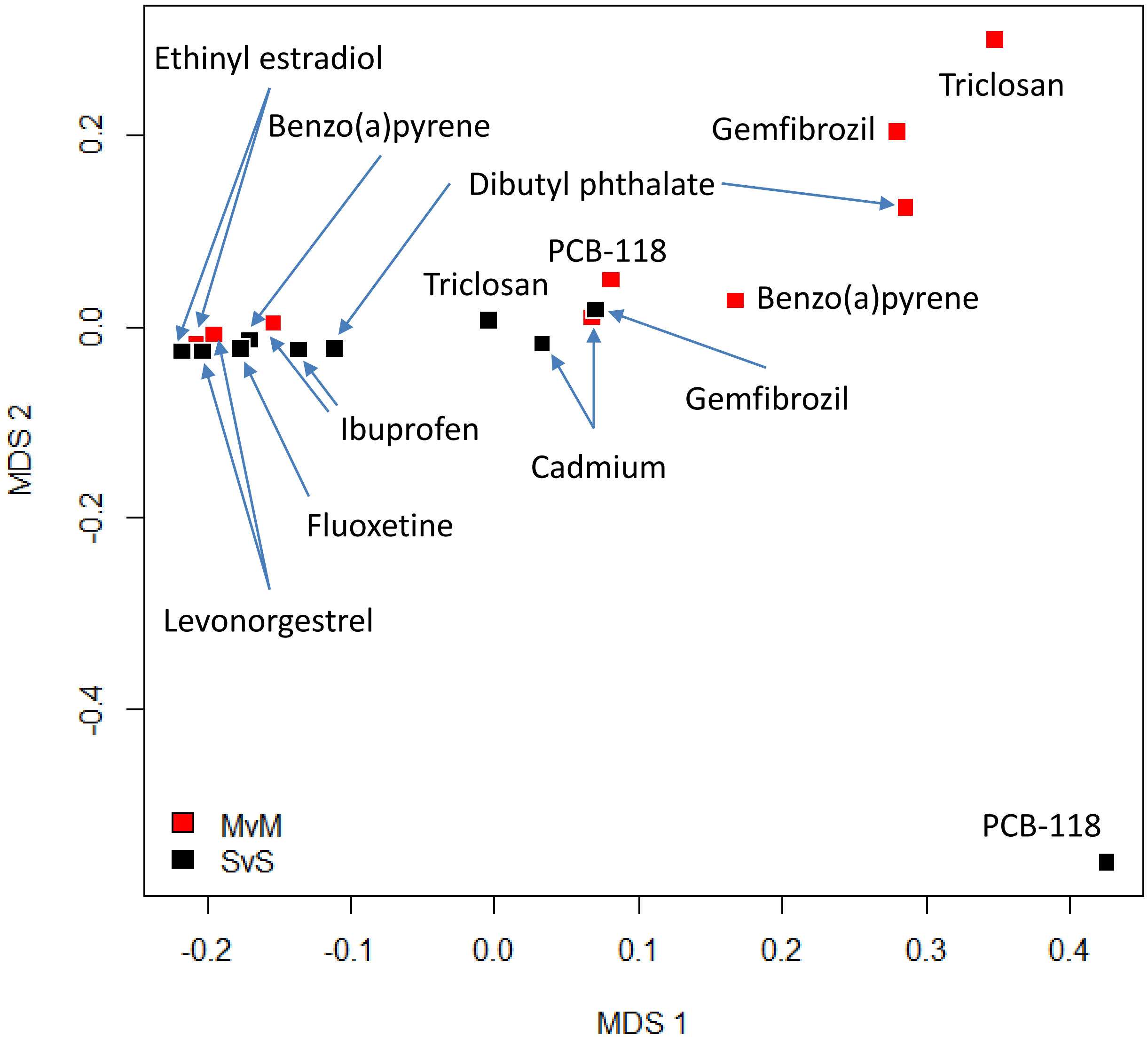
Multi-Step Modelling Procedure

Model 1 – Prediction of Compound presence



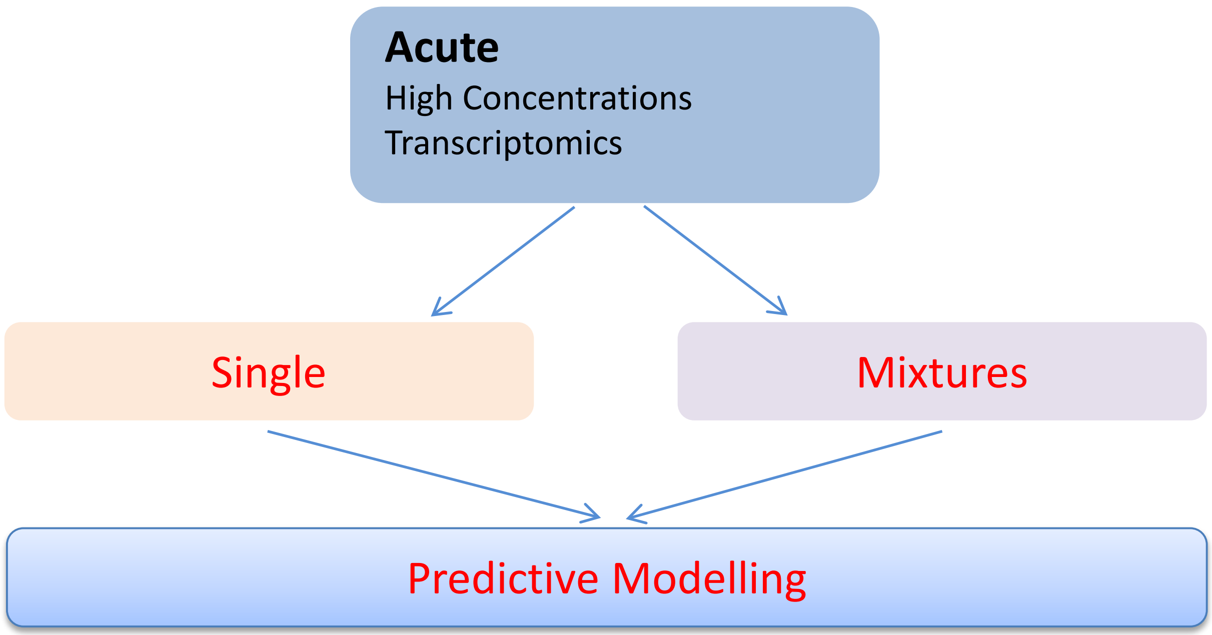

Single Models
Mixtures
0.0
0.2
0.4
0.6
0.8
1.0
Chemical Only
Other Chemical
Benzo(a)pyrene
Cadmium
Dibutyl phthalate
Ethinyl estradiol
Fluoxetine
Gemfibrozil
Ibuprofen
Levonorgestrel
PCB-118
Triclosan

Comparing Model Space



Multi-Step Modelling Procedure

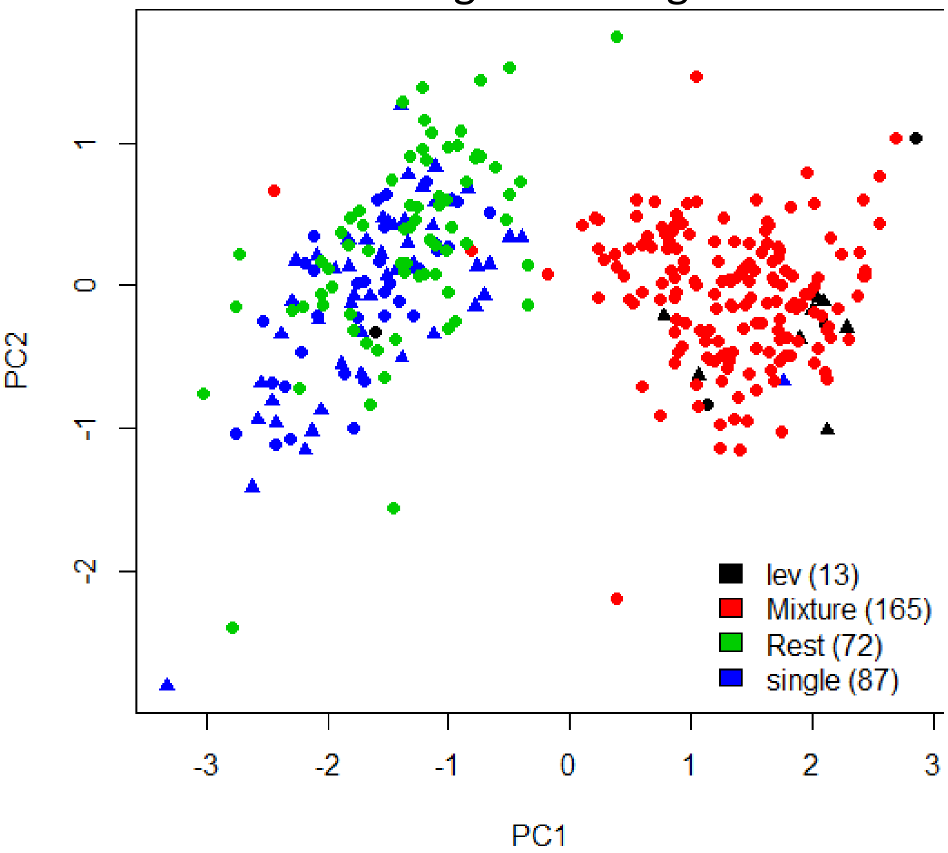
Model 2 – Model Refinement



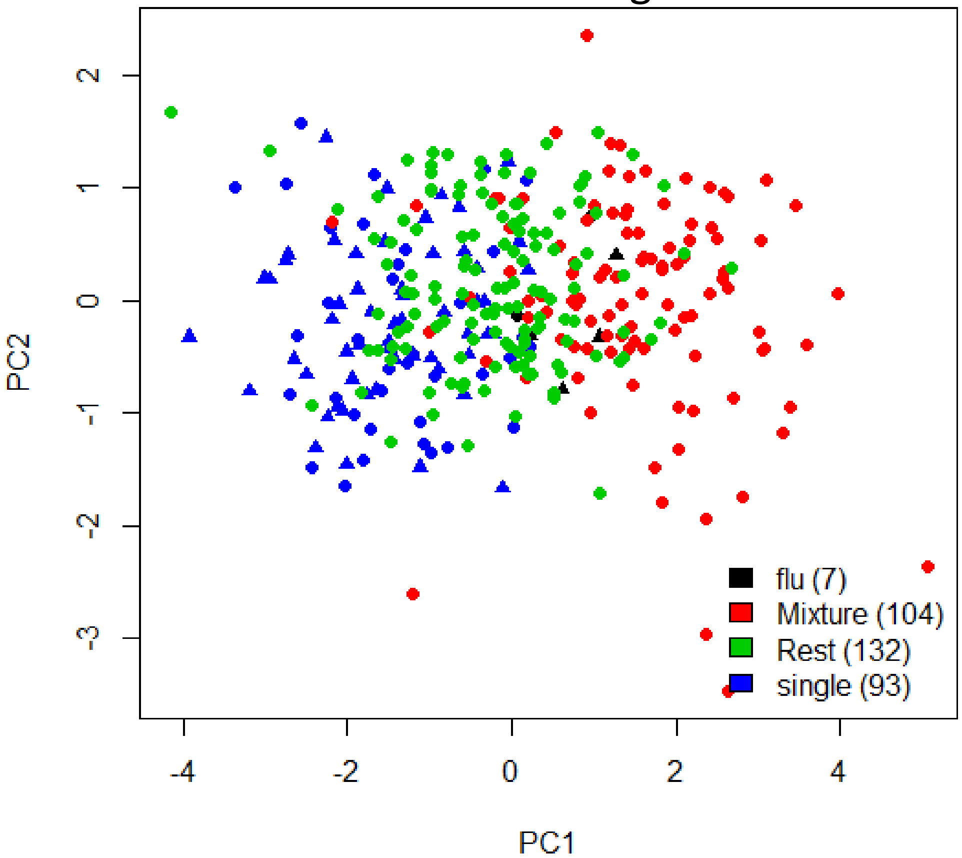
Building Models Predictive in both Single and Mixtures

Predicting exposure to single and mixture exposures

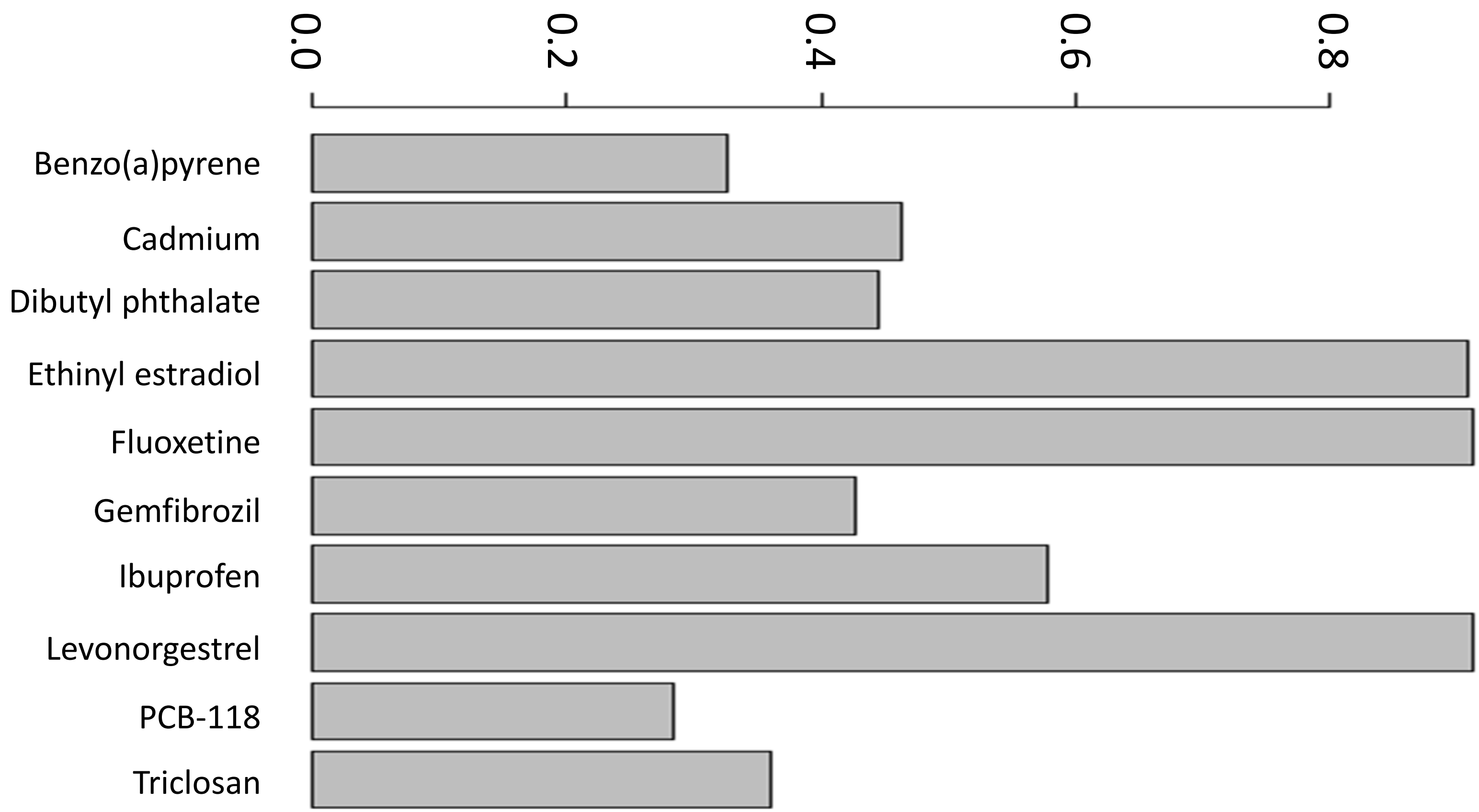
Levonorgestrel – 3 genes



Fluoxetine – 4 genes

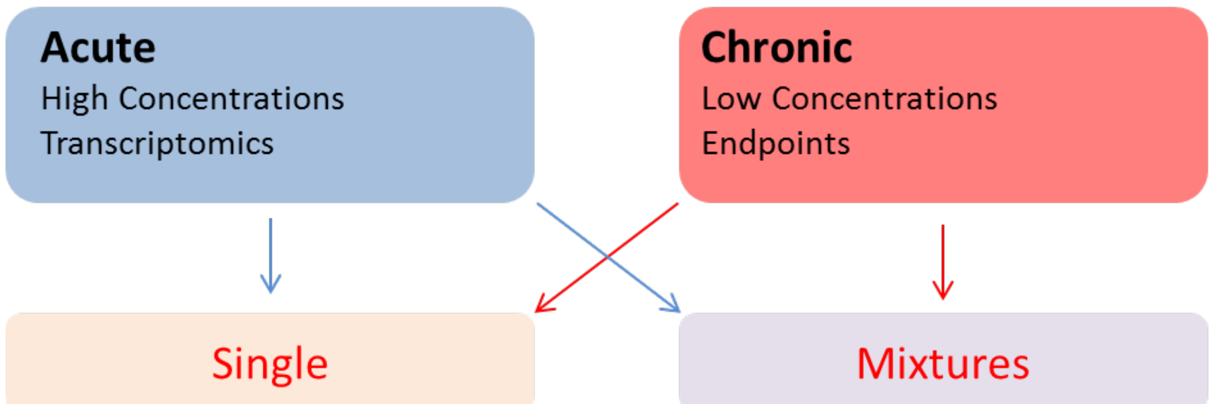


Models predictive of both Single and Mixtures

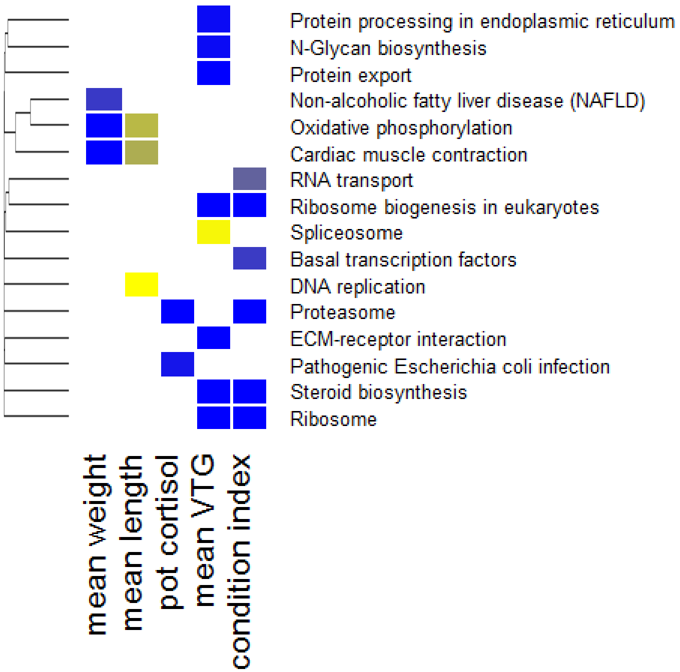


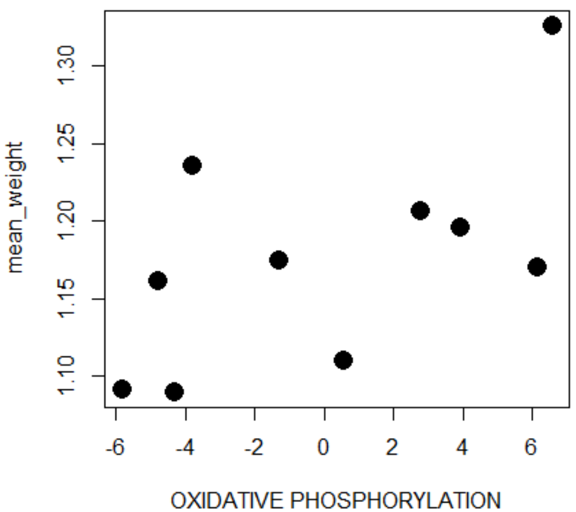
Multi-Step Modelling Procedure

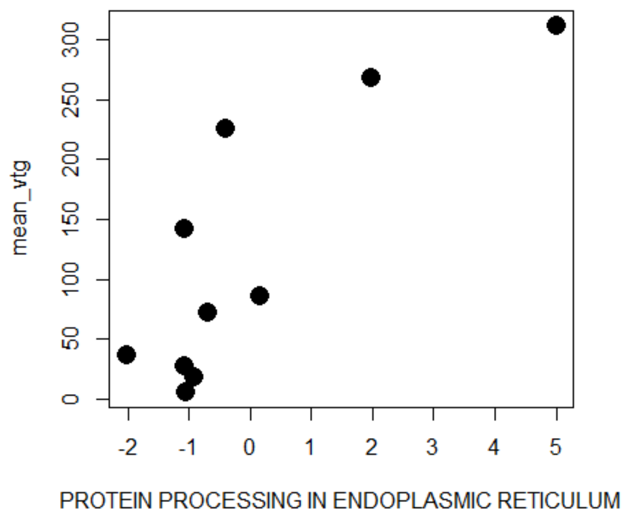
Model 3 – Linking Chronic phenotypes to early molecular response

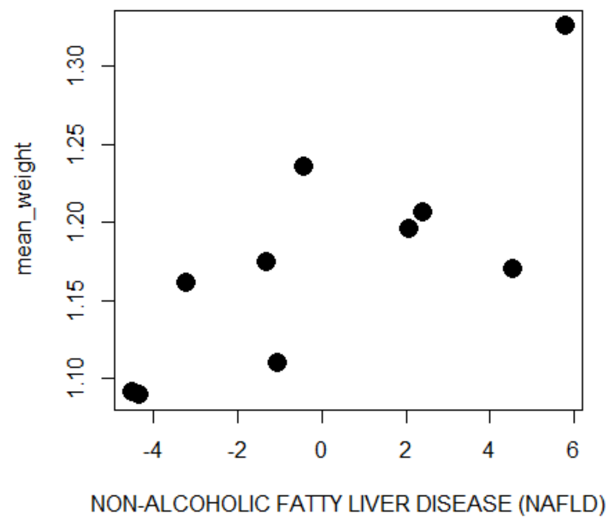


Pathway to Phenotype Association



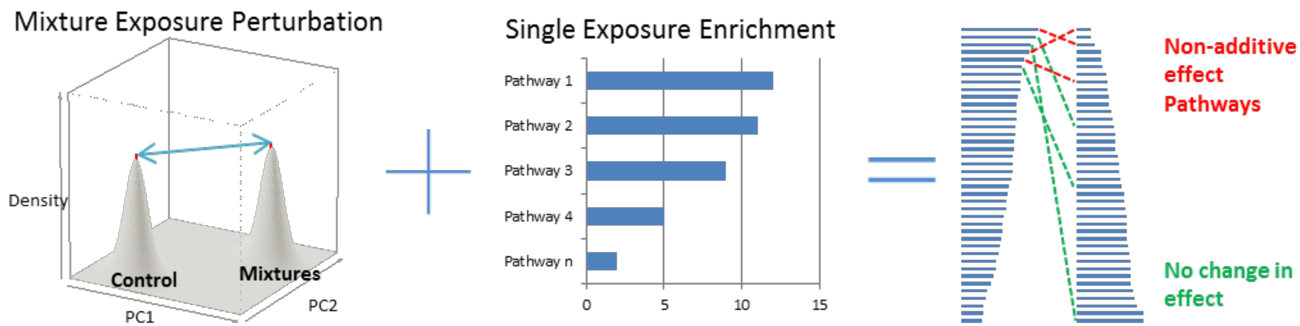




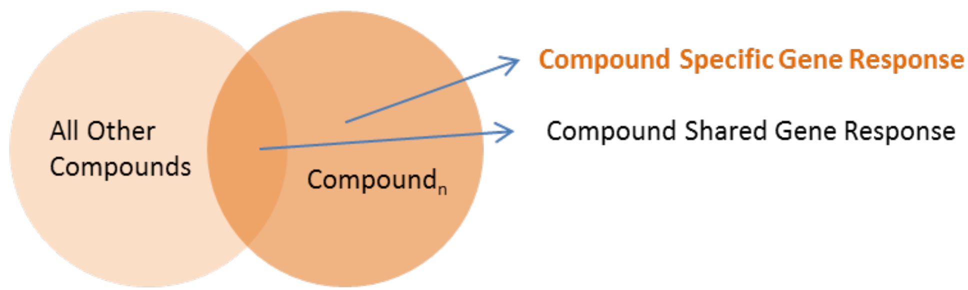


Multi-Step Modelling Procedure

Step 2: Identification of non-additive Pathways



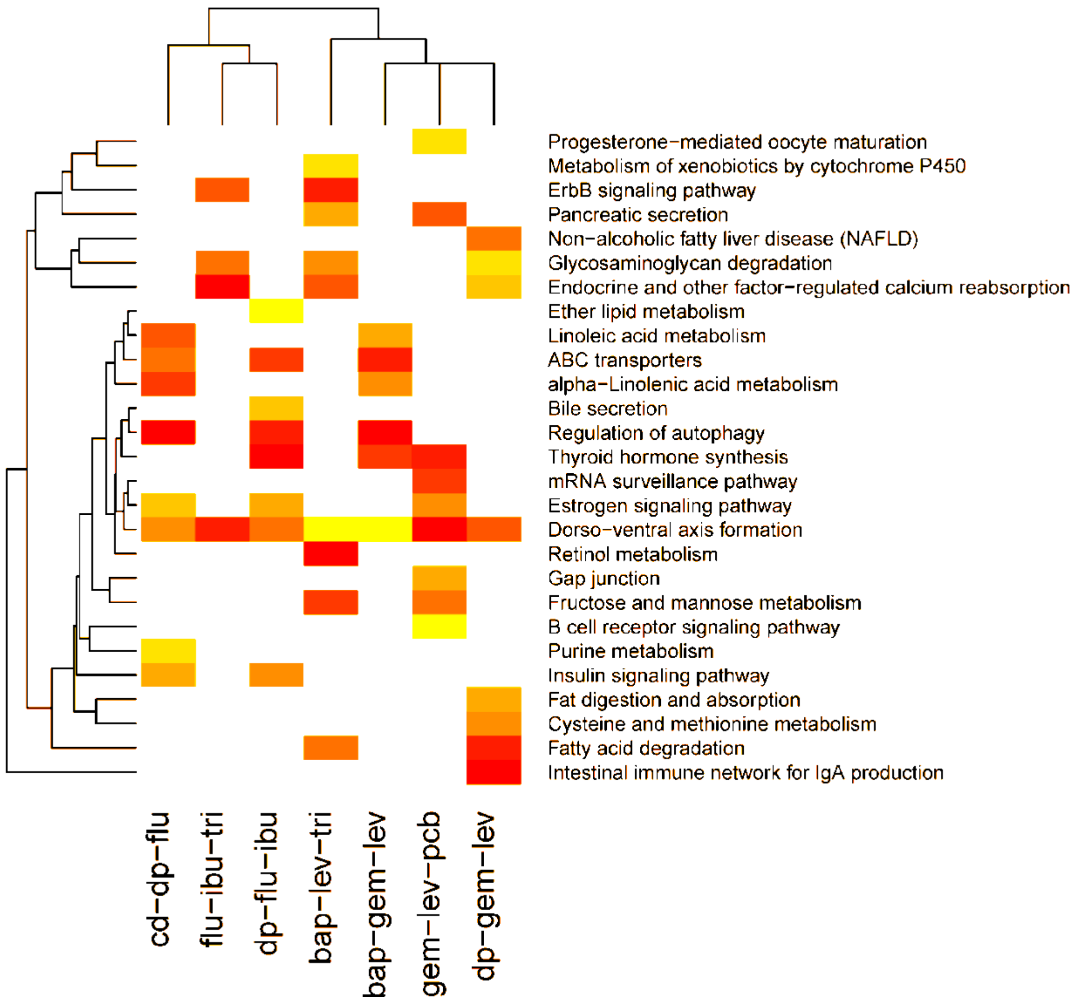
Step 3: Identification of Specific Molecular Response



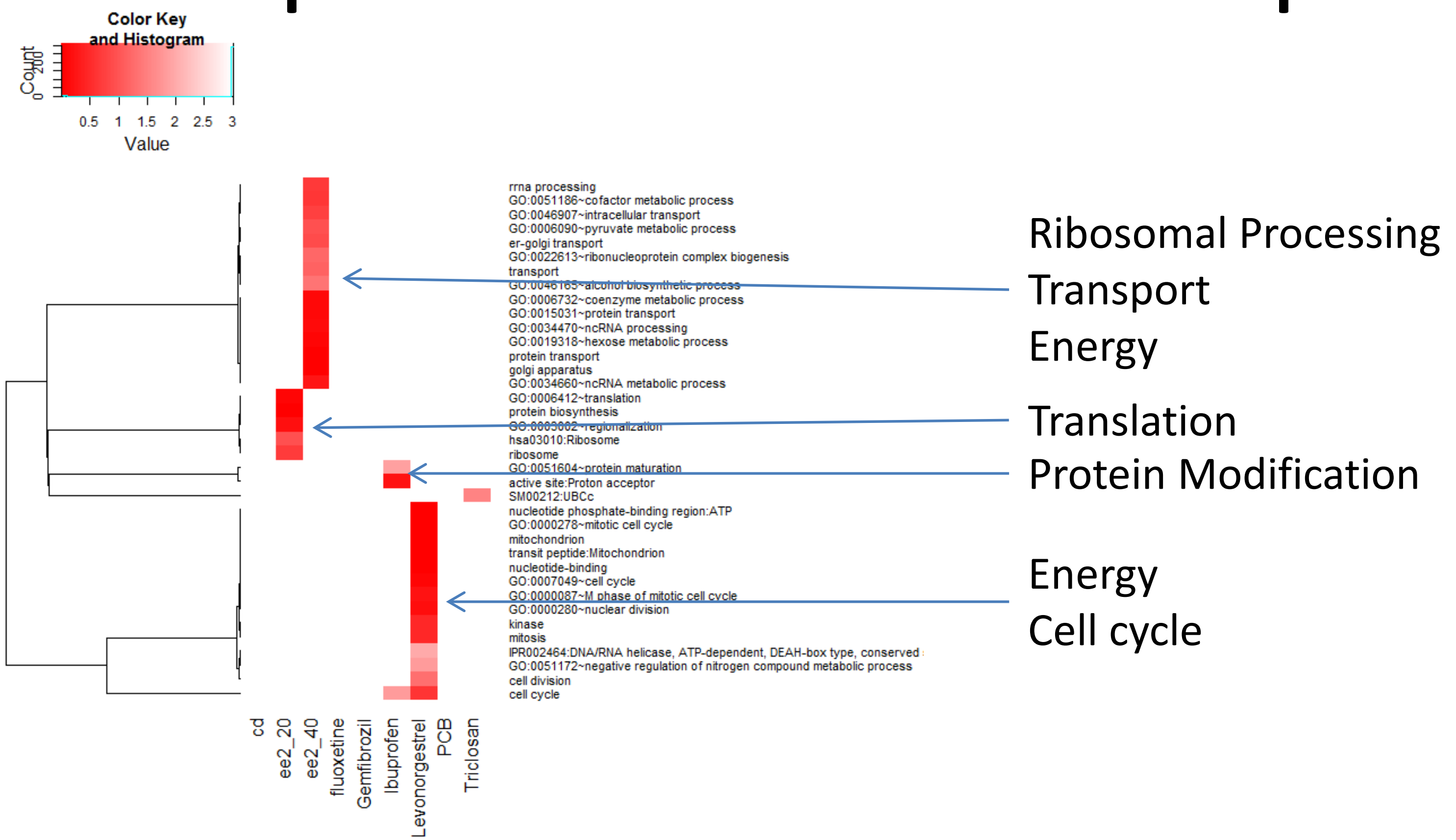
Step 4: Developing statistical relationships between genes and pathways

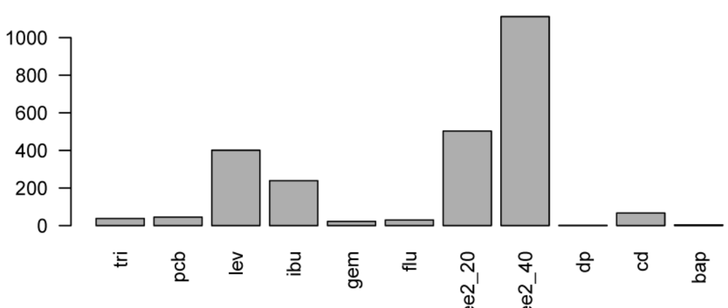
Step 5: Identification of Endpoints associated to non-additive effect

Non additive effect Pathways



Specific Molecular Response





Underlying model

- Genetic Algorithm based optimization technique (GALGO library R)
  - RandomForest regression

$$PC_{i,k} = gene_{Compound1} + gene_{Compound2} + gene_{Compound3} + d + \epsilon$$

Putative mixture AOPs

Identify differentially expressed genes from Single exposures

↓

Identify Pathways only expressed in mixtures

↓

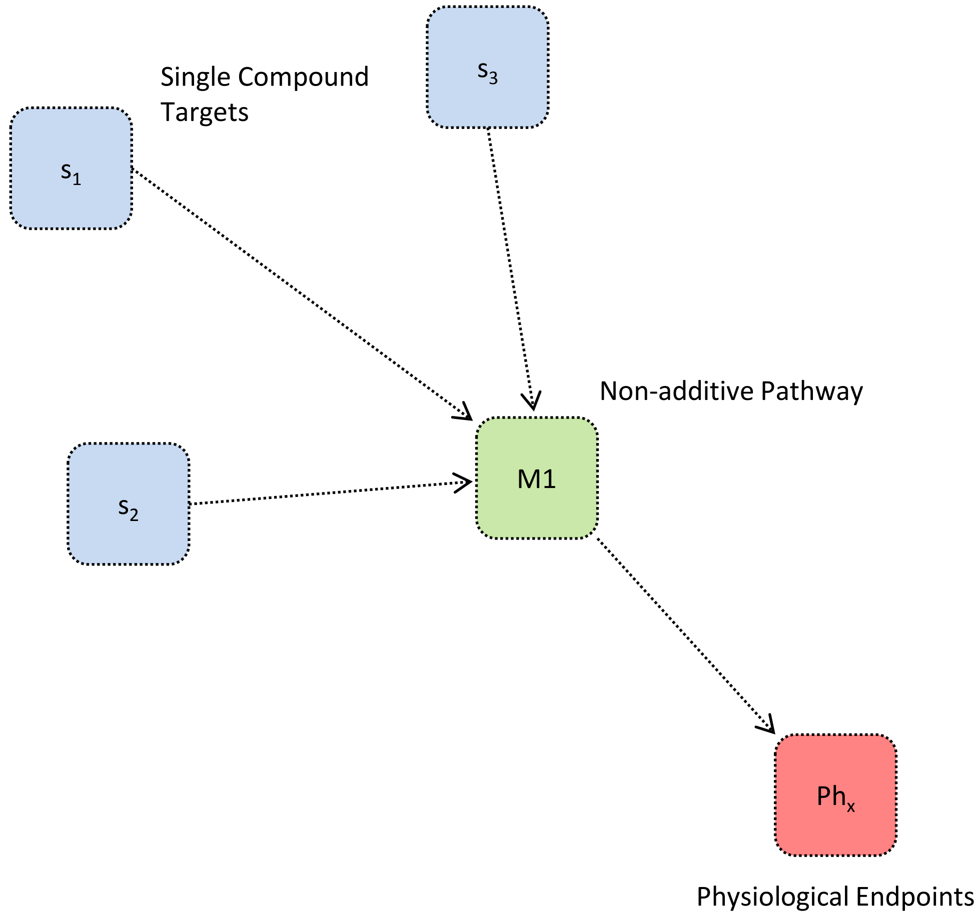
Link genes from single exposures (1 per compound) to pathways

↓

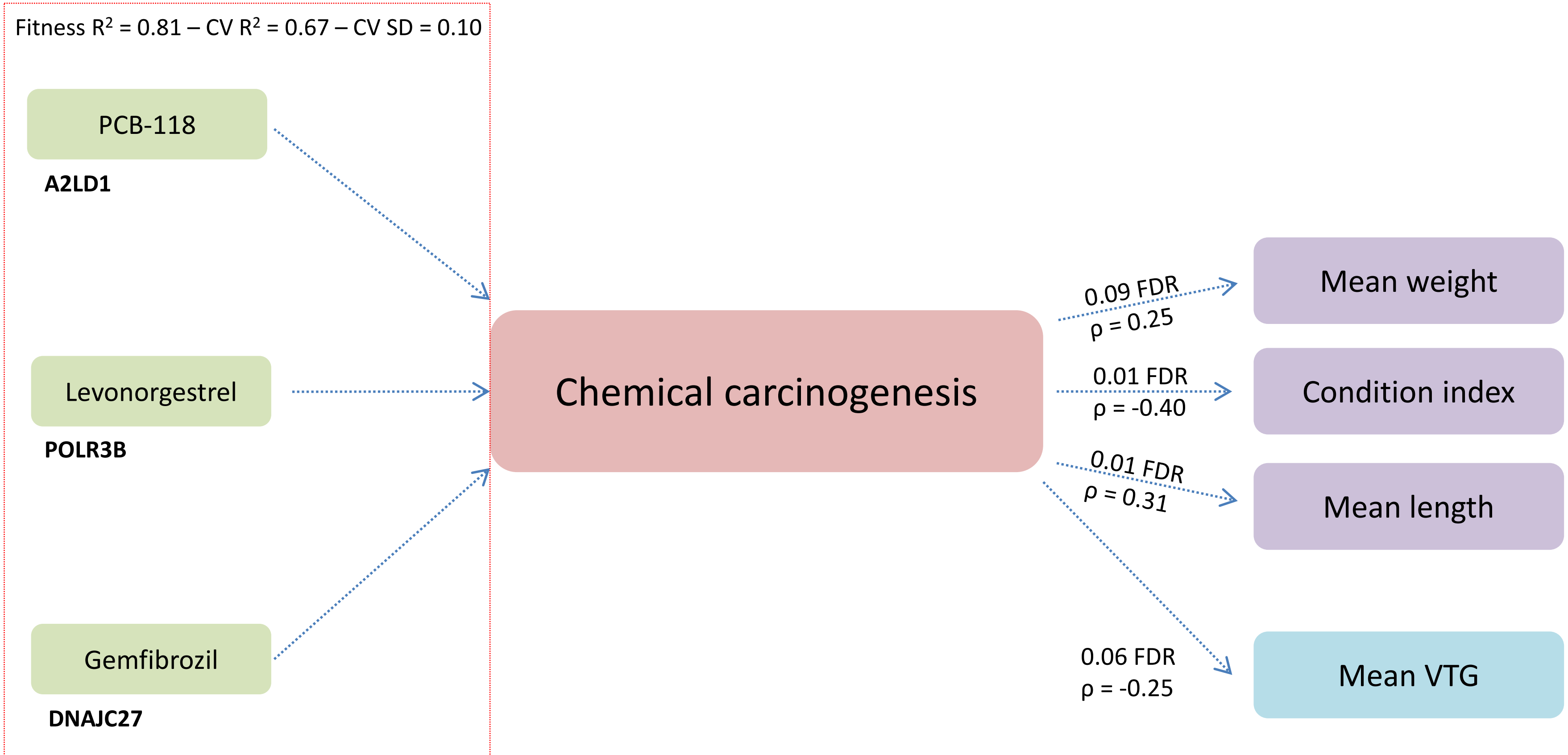
Link physiological endpoints to pathways

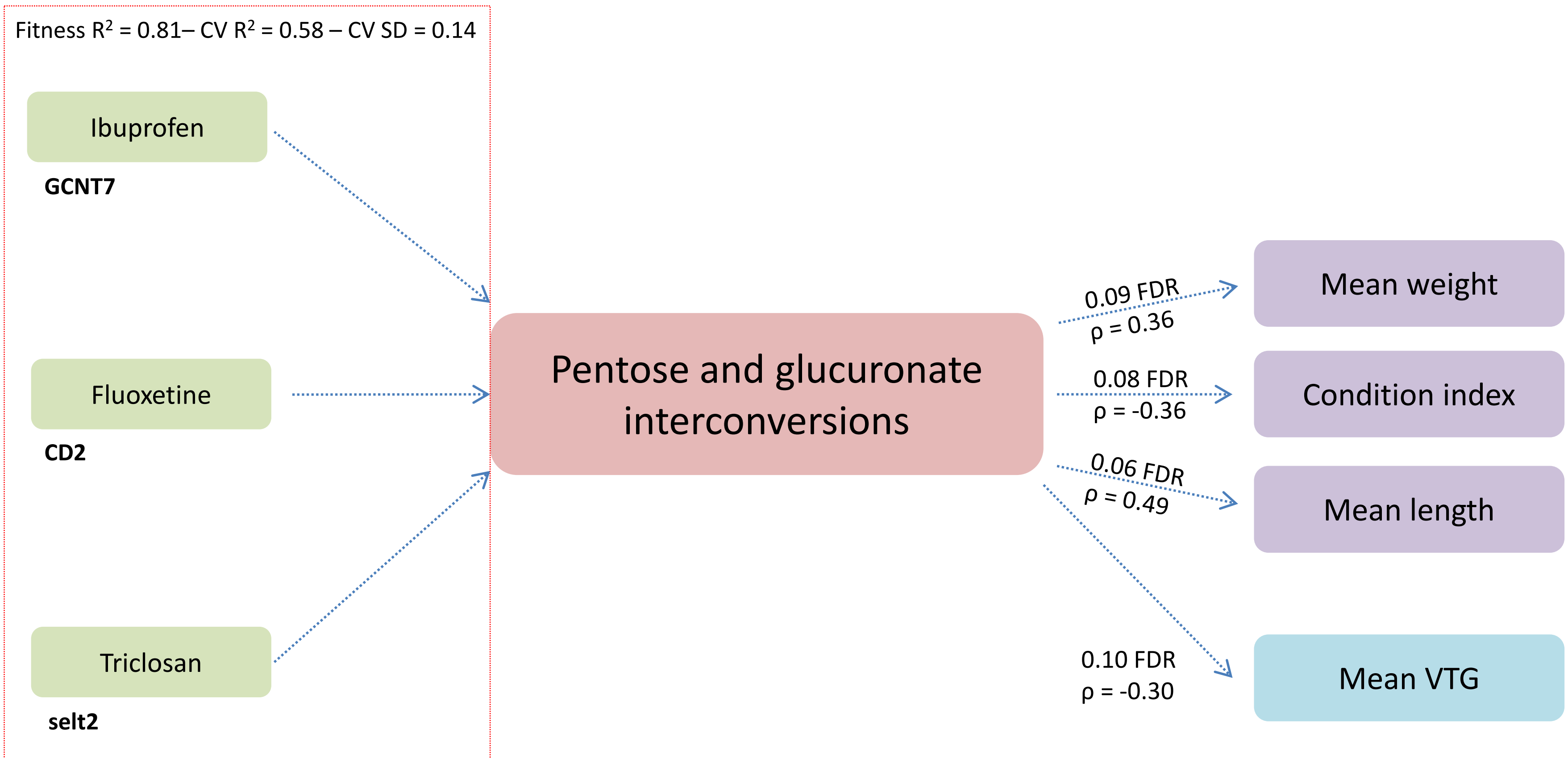
↓

Derive mixture AOP based on best model fit



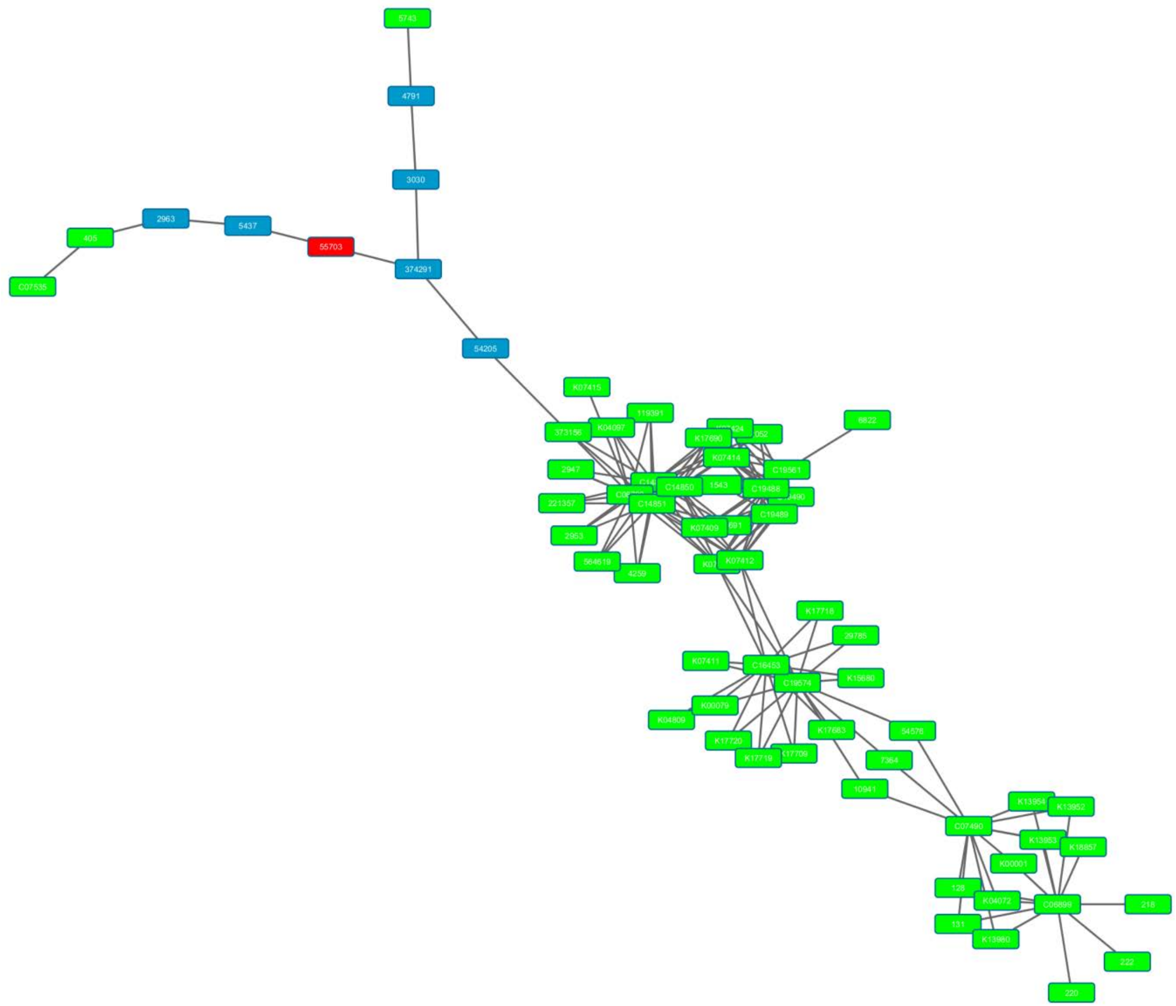
Integrating and identifying pAOPs



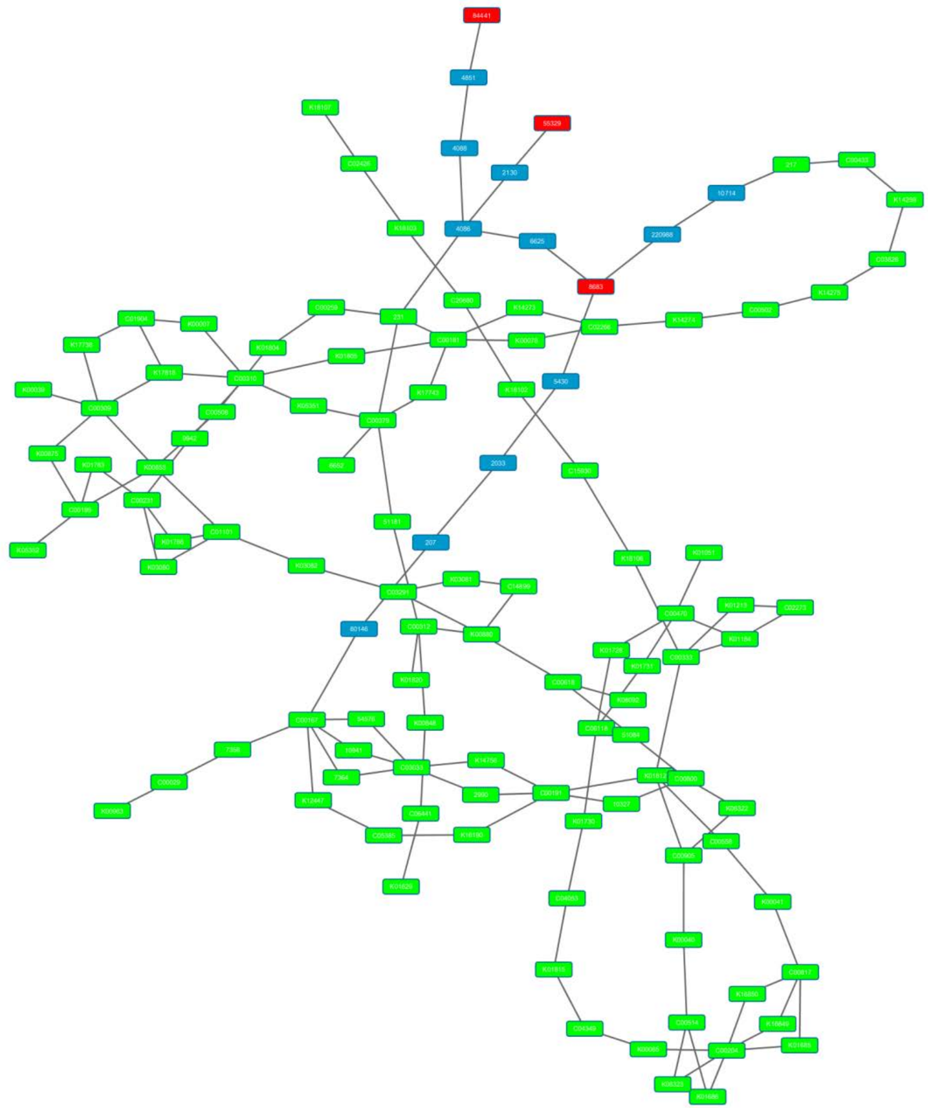


Fitness R^2 = 0.81– CV R^2 = 0.58 – CV SD = 0.14
Ibuprofen
GCNT7
Fluoxetine
CD2
Triclosan
selt2
Pentose and glucuronate interconversions
0.09 FDR
ρ = 0.36
Mean weight
0.08 FDR
ρ = -0.36
Condition index
0.06 FDR
ρ = 0.49
Mean length
0.10 FDR
ρ = -0.30
Mean VTG

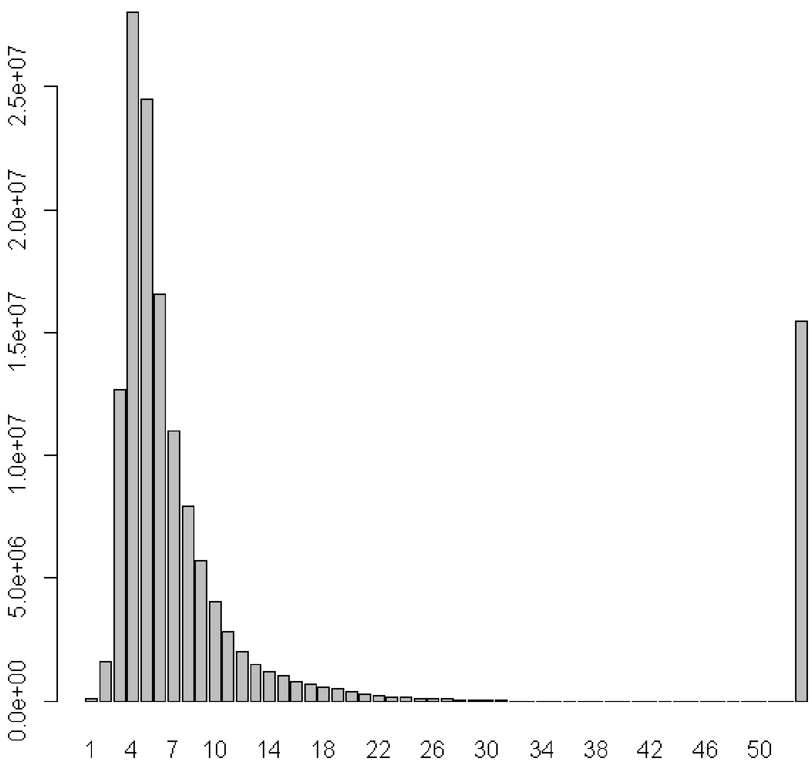
Chemical Carcinogenesis

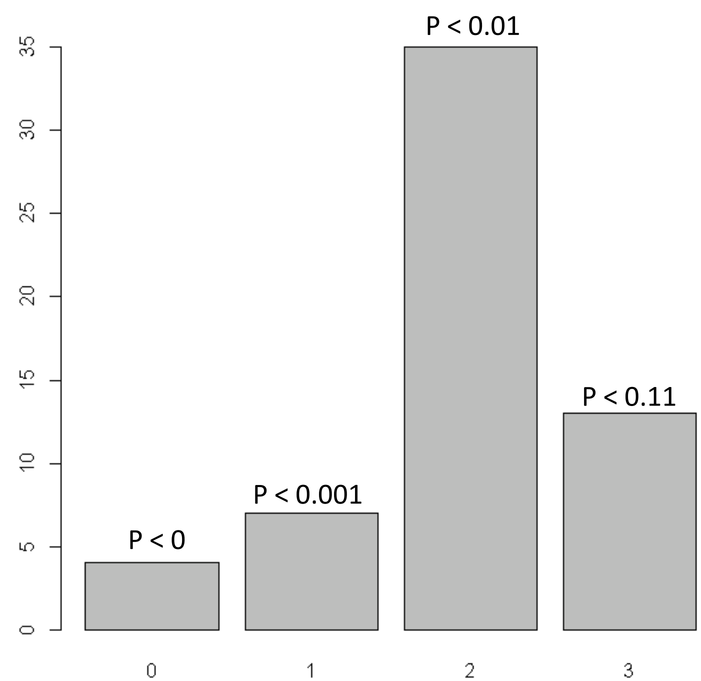


Pentose and glucuronate Interconversions



Shortest Paths within KEGG+MiMI





Summary

- Molecular data can be used as a **predictive tool** to **identify** and **classify samples**
- We are able to **develop** models linking single chemical exposure, non-additive mixture effect and phenotypic endpoints to develop **putative mixture adverse outcome pathways**
- We need to develop more quantitative/predictive Adverse Outcome Pathways to support risk assessment.

The next challenge

- Predictive/quantitative Adverse Outcome Pathways
- Cross species extrapolation of adverse outcome pathways
- Interactions between chemicals and a changing environment
- Robust molecular models for mTIE (molecular toxicity identification and evaluation) across large numbers of compounds
- Mixture AOPs linking single exposures to expected phenotypic effect and population outcome

Acknowledgements

University of Liverpool

Prof. Francesco Falciani

Kim Clarke

Jaanika Kronberg

John Herbert

John Ankers

Peter Davidsen

Cefas

Ioanna Katsiadaki

Marion Sebire

Jessica Tasker

Jenni Prokkola

Brett Lyons

Tim Bean

University of Birmingham

Prof. Mark Viant

Tom White

Prof. Kevin Chipman