

Superomics - Correlating multiple *omics* data with experimental design

Susann Mönchgesang

Leibniz Institute of Plant Biochemistry (IPB)
Bioinformatics and Mass Spectrometry

March 20, 2014



Table of contents

1 Introduction

- Canonical Correlation Analysis

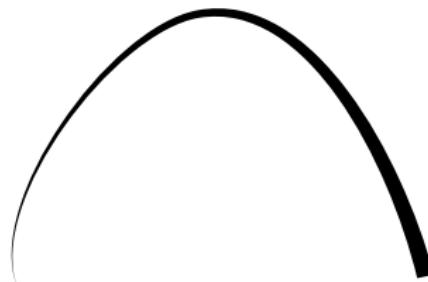
2 spCCA

- Experimental Design
- Results
- Challenges



Canonical Correlation Analysis

max. corr



Metabolomics

$m_1, m_2, m_3, \dots, m_p$

Proteomics

$p_1, p_2, p_3, \dots, p_q$



Penalized Canonical Correlation Analysis

max. corr

$$\begin{aligned}\omega_{m_2} &= 0.8 \\ \omega_{m_3} &= 0.7\end{aligned}$$

$$\begin{aligned}\omega_{p_1} &= 0.7 \\ \omega_{p_q} &= 0.6\end{aligned}$$

Metabolomics

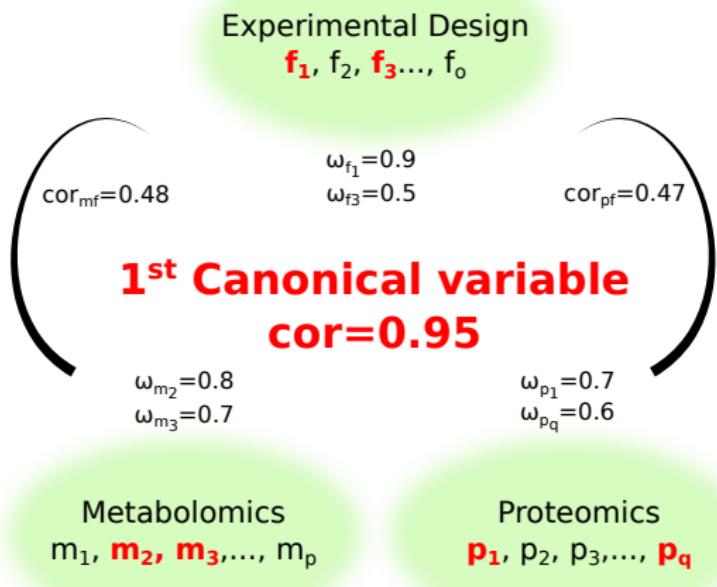
$m_1, m_2, m_3, \dots, m_p$

Proteomics

$p_1, p_2, p_3, \dots, p_q$

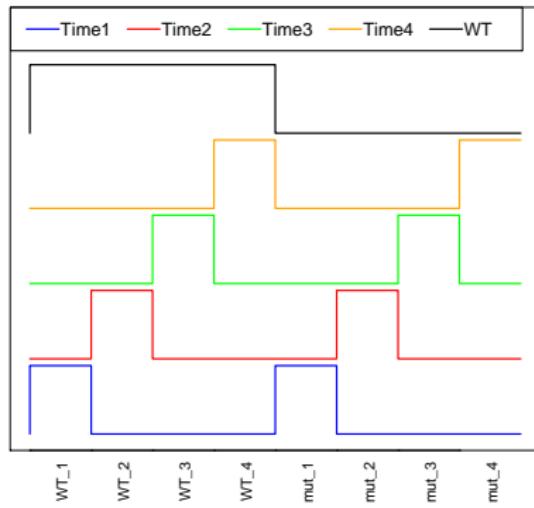


Supervised penalized Canonical Correlation Analysis



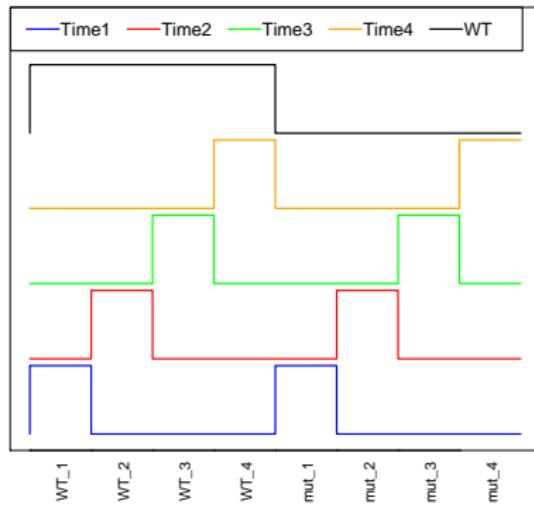
Experimental design as data set

Experimental design



Experimental design as data set

Experimental design

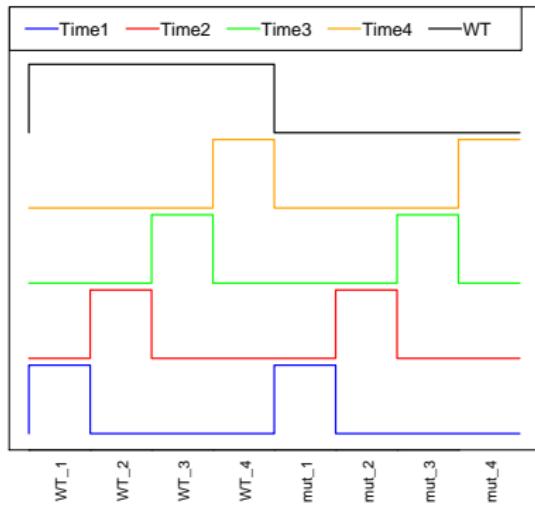


Matrix of binary vectors:

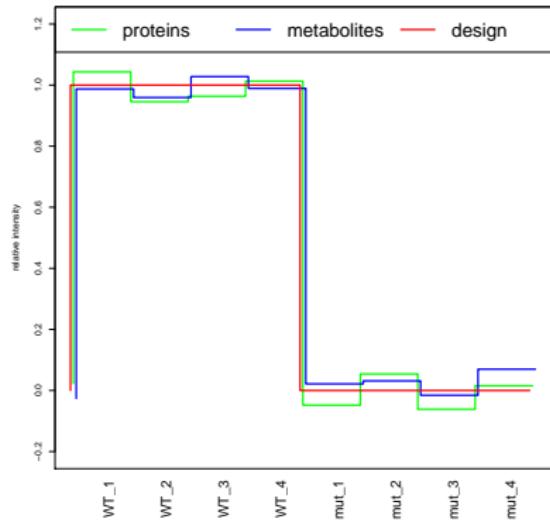
	WT_1	WT_2	WT_3	WT_4	mut_1	mut_2	mut_3	mut_4
WT	1	1	1	1	0	0	0	0
Time4	0	0	0	1	0	0	0	1
Time3	0	0	1	0	0	0	0	1
Time2	0	1	0	0	0	1	0	0
Time1	1	0	0	0	1	0	0	0

Experimental design as data set

Experimental design

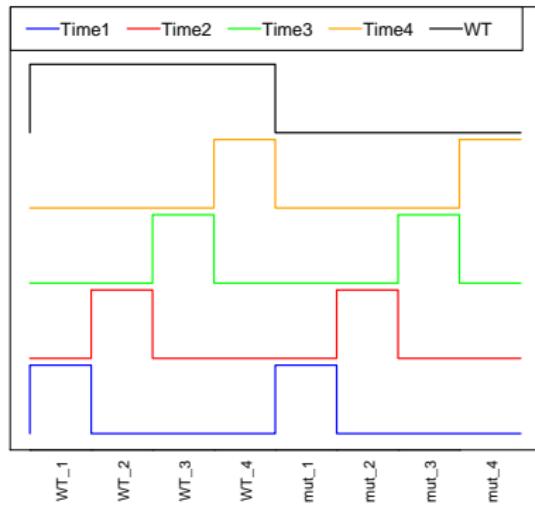


Canonical Variable no 1, corr=0.95

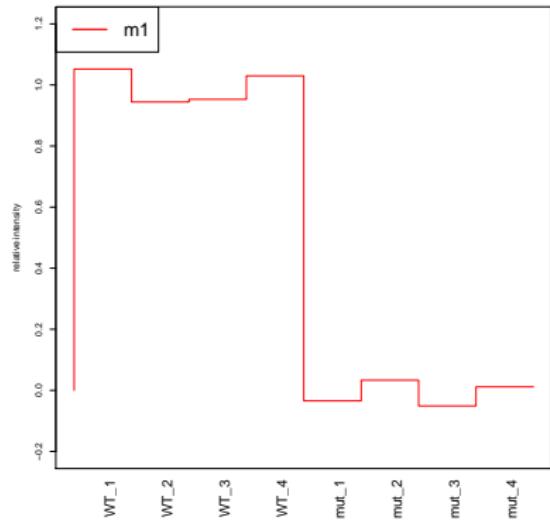


Experimental design as data set

Experimental design

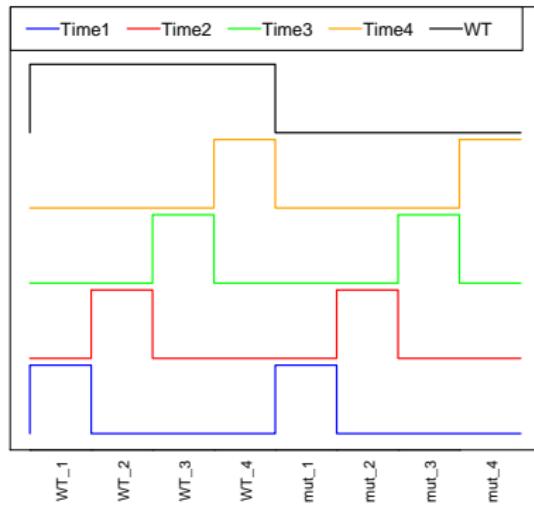


Features indicative for WT

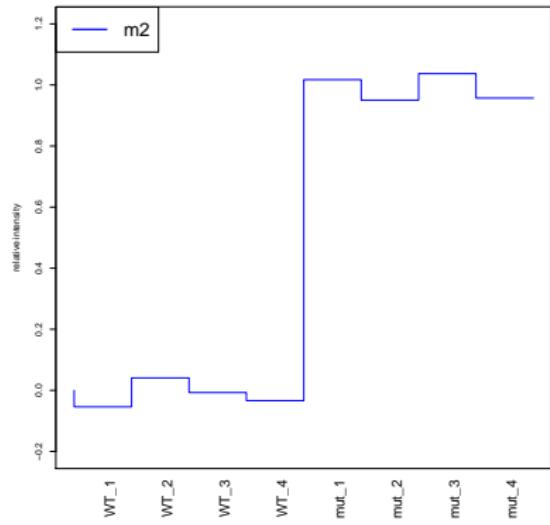


Experimental design as data set

Experimental design

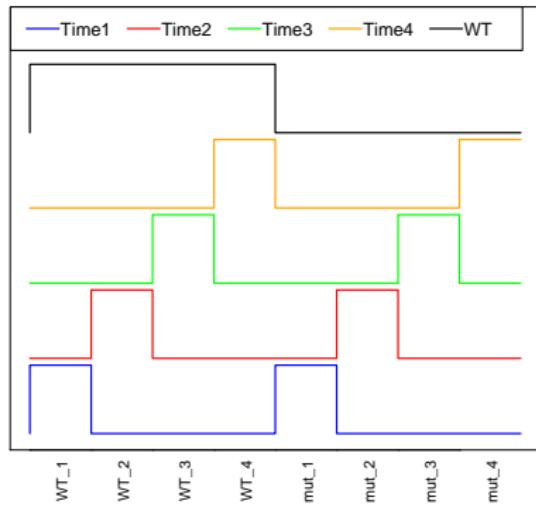


Features indicative for WT

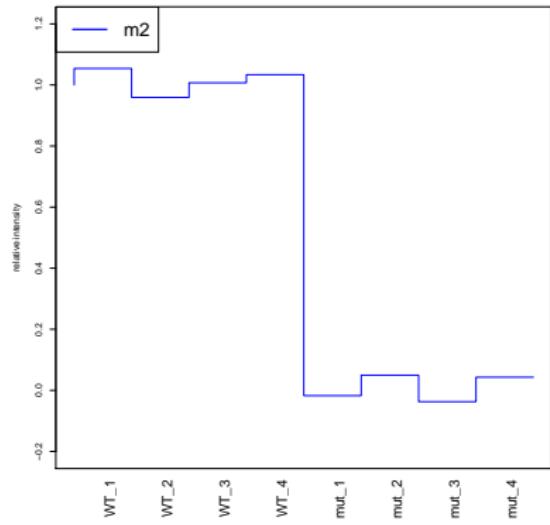


Experimental design as data set

Experimental design

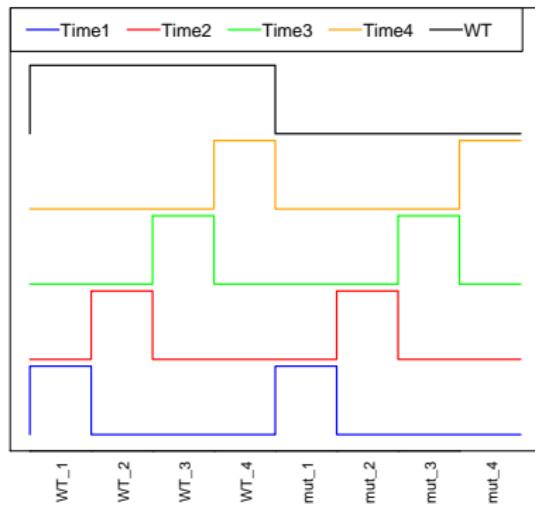


Features indicative for WT

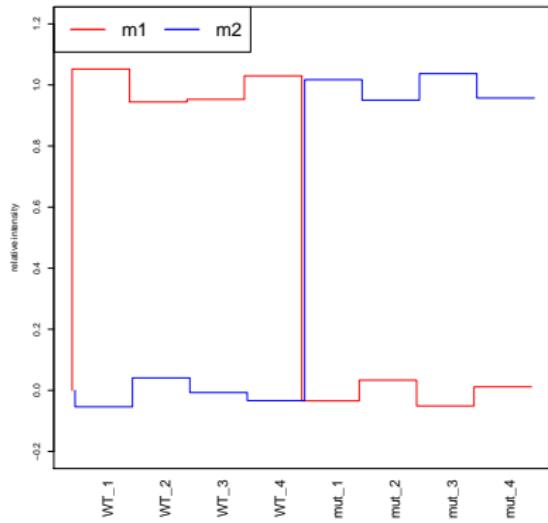


Experimental design as data set

Experimental design



Features indicative for WT

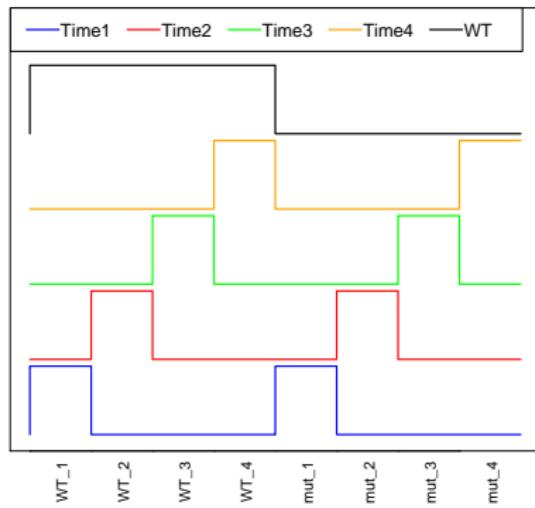


$$\omega_{m1} > 0 ; \omega_{m2} < 0$$

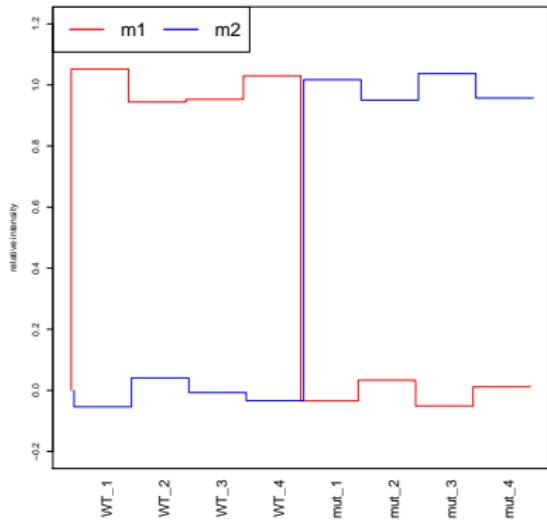


Experimental design as data set

Experimental design



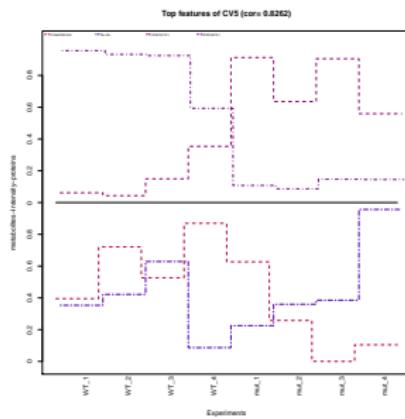
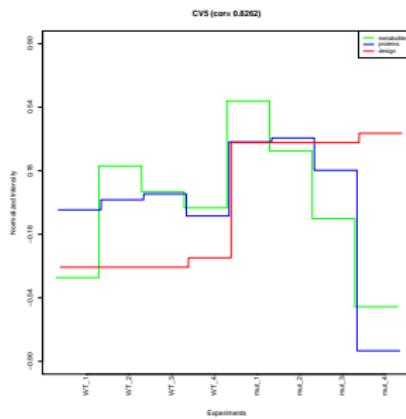
Features indicative for WT



$$\omega_{m1} > 0 ; \omega_{m2} < 0$$

Easier interpretation of Canonical Variables!

Canonical Correlation Variables (CV)



Weights

Exp. Design
WT 0.998
Time4 -0.056

Metabolomics

Met1 **0.400**

Met2 0.367

Met3 -0.361

Met4 -0.351

Met5 -0.346

Met6 -0.327

Met7 -0.283

Met8 -0.224

Met9 -0.195

Met10 -0.159

Met15 -0.004

Proteomics

Pr1 **0.077**

Pr2 -0.077

Pr3 0.077

Pr4 -0.076

Pr5 -0.076

Pr6 -0.076

Pr7 -0.076

Pr8 0.076

Pr9 0.075

Pr10 -0.075

Pr11 ..

Pr12 ..

Pr13 ..

Pr14 ..

Pr15 ..

Pr16 ..

Pr17 ..

Pr18 ..

Pr19 ..

Pr20 ..

Pr21 ..

Pr22 ..

Pr23 ..

Pr24 ..

Pr25 ..

Pr26 ..

Pr27 ..

Pr28 ..

Pr29 ..

Pr30 ..

Pr31 ..

Pr32 ..

Pr33 ..

Pr34 ..

Pr35 ..

Pr36 ..

Pr37 ..

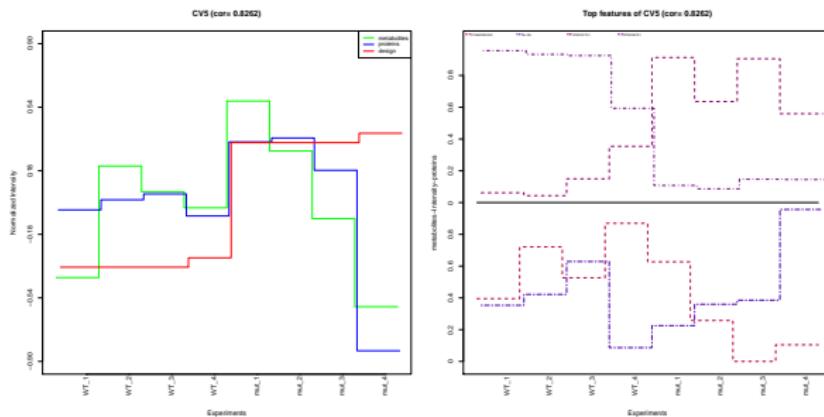
Pr38 ..

Pr39 ..

Pr40 ..



Canonical Correlation Variables (CV)



Weights

Exp. Design
WT 0.998
Time4 -0.056

Metabolomics

Met1 0.400

Met2 0.367

Met3 -0.361

Met4 -0.351

Met5 -0.346

Met6 -0.327

Met7 -0.283

Met8 -0.224

Met9 -0.195

Met10 -0.159

Met11 -0.125

Met12 -0.098

Met13 -0.075

Met14 -0.056

Met15 -0.004

Proteomics

Pr1 0.077

Pr2 -0.077

Pr3 0.077

Pr4 -0.076

Pr5 -0.076

Pr6 -0.076

Pr7 -0.076

Pr8 0.076

Pr9 0.075

Pr10 -0.075

Pr11 -0.075

Pr12 -0.075

Pr13 -0.075

Pr14 -0.075

Pr15 -0.075

Pr16 -0.075

Pr17 -0.075

Pr18 -0.075

Pr19 -0.075

Pr20 -0.075

Pr21 -0.075

Pr22 -0.075

Pr23 -0.075

Pr24 -0.075

Pr25 -0.075

Pr26 -0.075

Pr27 -0.075

Pr28 -0.075

Pr29 -0.075

Pr30 -0.075

Pr31 -0.075

Pr32 -0.075

Pr33 -0.075

Pr34 -0.075

Pr35 -0.075

Pr36 -0.075

Pr37 -0.075

Pr38 -0.075

Pr39 -0.075

Pr40 -0.075

Pr41 -0.075

Pr42 -0.075

Pr43 -0.075

Pr44 -0.075

Pr45 -0.075

Pr46 -0.075

Pr47 -0.075

Pr48 -0.075

Pr49 -0.075

Pr50 -0.075

Pr51 -0.075

Pr52 -0.075

Pr53 -0.075

Pr54 -0.075

Pr55 -0.075

Pr56 -0.075

Pr57 -0.075

Pr58 -0.075

Pr59 -0.075

Pr60 -0.075

Pr61 -0.075

Pr62 -0.075

Pr63 -0.075

Pr64 -0.075

Pr65 -0.075

Pr66 -0.075

Pr67 -0.075

Pr68 -0.075

Pr69 -0.075

Pr70 -0.075

Pr71 -0.075

Pr72 -0.075

Pr73 -0.075

Pr74 -0.075

Pr75 -0.075

Pr76 -0.075

Pr77 -0.075

Pr78 -0.075

Pr79 -0.075

Pr80 -0.075

Pr81 -0.075

Pr82 -0.075

Pr83 -0.075

Pr84 -0.075

Pr85 -0.075

Pr86 -0.075

Pr87 -0.075

Pr88 -0.075

Pr89 -0.075

Pr90 -0.075

Pr91 -0.075

Pr92 -0.075

Pr93 -0.075

Pr94 -0.075

Pr95 -0.075

Pr96 -0.075

Pr97 -0.075

Pr98 -0.075

Pr99 -0.075

Pr100 -0.075

Pr101 -0.075

Pr102 -0.075

Pr103 -0.075

Pr104 -0.075

Pr105 -0.075

Pr106 -0.075

Pr107 -0.075

Pr108 -0.075

Pr109 -0.075

Pr110 -0.075

Pr111 -0.075

Pr112 -0.075

Pr113 -0.075

Pr114 -0.075

Pr115 -0.075

Pr116 -0.075

Pr117 -0.075

Pr118 -0.075

Pr119 -0.075

Pr120 -0.075

Pr121 -0.075

Pr122 -0.075

Pr123 -0.075

Pr124 -0.075

Pr125 -0.075

Pr126 -0.075

Pr127 -0.075

Pr128 -0.075

Pr129 -0.075

Pr130 -0.075

Pr131 -0.075

Pr132 -0.075

Pr133 -0.075

Pr134 -0.075

Pr135 -0.075

Pr136 -0.075

Pr137 -0.075

Pr138 -0.075

Pr139 -0.075

Pr140 -0.075

Pr141 -0.075

Pr142 -0.075

Pr143 -0.075

Pr144 -0.075

Pr145 -0.075

Pr146 -0.075

Pr147 -0.075

Pr148 -0.075

Pr149 -0.075

Pr150 -0.075

Pr151 -0.075

Pr152 -0.075

Pr153 -0.075

Pr154 -0.075

Pr155 -0.075

Pr156 -0.075

Pr157 -0.075

Pr158 -0.075

Pr159 -0.075

Pr160 -0.075

Pr161 -0.075

Pr162 -0.075

Pr163 -0.075

Pr164 -0.075

Pr165 -0.075

Pr166 -0.075

Pr167 -0.075

Pr168 -0.075

Pr169 -0.075

Pr170 -0.075

Pr171 -0.075

Pr172 -0.075

Pr173 -0.075

Pr174 -0.075

Pr175 -0.075

Pr176 -0.075

Pr177 -0.075

Pr178 -0.075

Pr179 -0.075

Pr180 -0.075

Pr181 -0.075

Pr182 -0.075

Pr183 -0.075

Pr184 -0.075

Pr185 -0.075

Pr186 -0.075

Pr187 -0.075

Pr188 -0.075

Pr189 -0.075

Pr190 -0.075

Pr191 -0.075

Pr192 -0.075

Pr193 -0.075

Pr194 -0.075

Pr195 -0.075

Pr196 -0.075

Pr197 -0.075

Pr198 -0.075

Pr199 -0.075

Pr200 -0.075

Pr201 -0.075

Pr202 -0.075

Pr203 -0.075

Pr204 -0.075

Pr205 -0.075

Pr206 -0.075

Pr207 -0.075

Pr208 -0.075

Pr209 -0.075

Pr210 -0.075

Pr211 -0.075

Pr212 -0.075

Pr213 -0.075

Pr214 -0.075

Pr215 -0.075

Pr216 -0.075

Pr217 -0.075

Pr218 -0.075

Pr219 -0.075

Pr220 -0.075

Pr221 -0.075

Pr222 -0.075

Pr223 -0.075

Pr224 -0.075

Pr225 -0.075

Pr226 -0.075

Pr227 -0.075

Pr228 -0.075

Pr229 -0.075

Pr230 -0.075

Pr231 -0.075

Pr232 -0.075

Pr233 -0.075

Pr234 -0.075

Pr235 -0.075

Pr236 -0.075

Pr237 -0.075

Pr238 -0.075

Pr239 -0.075

Pr240 -0.075

Pr241 -0.075

Pr242 -0.075

Pr243 -0.075

Pr244 -0.075

Pr245 -0.075

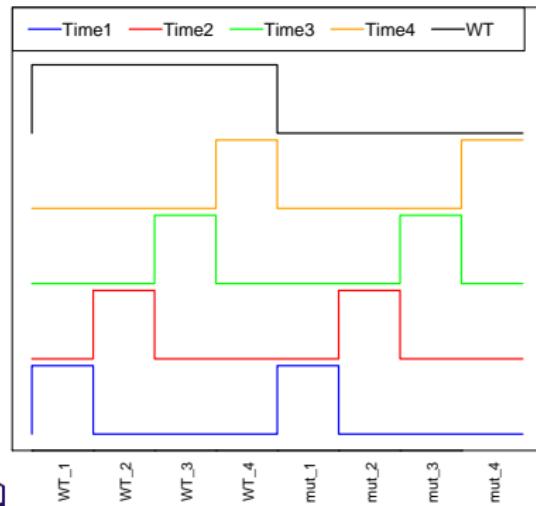
Pr246 -0.075

Pr247 -0.075

Pr248 -0.075

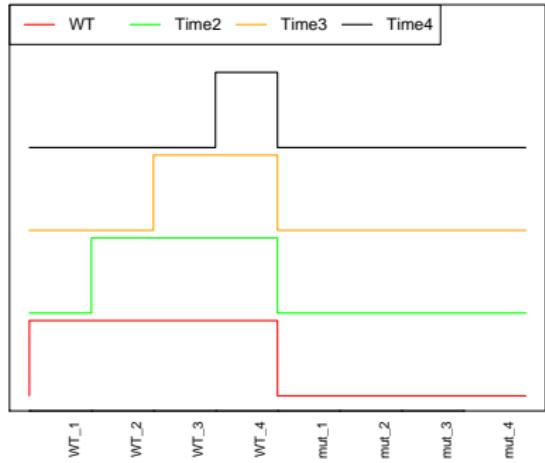
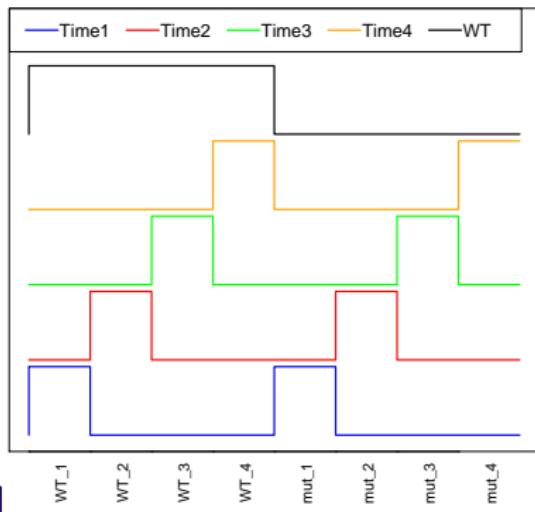
Reconsider experimental design

Did we ask the correct question?



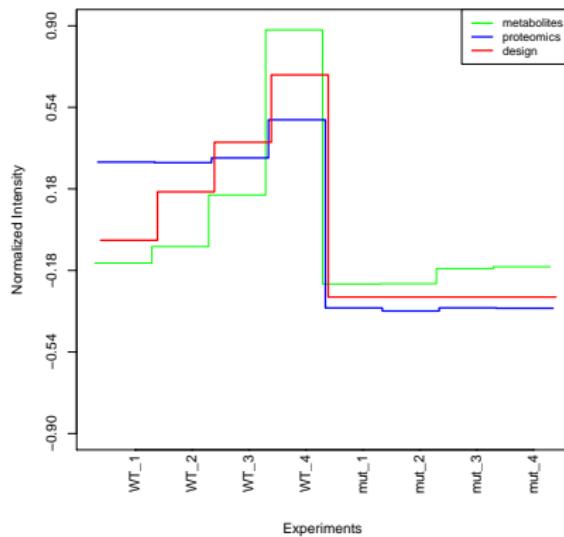
Reconsider experimental design

Did we ask the correct question?



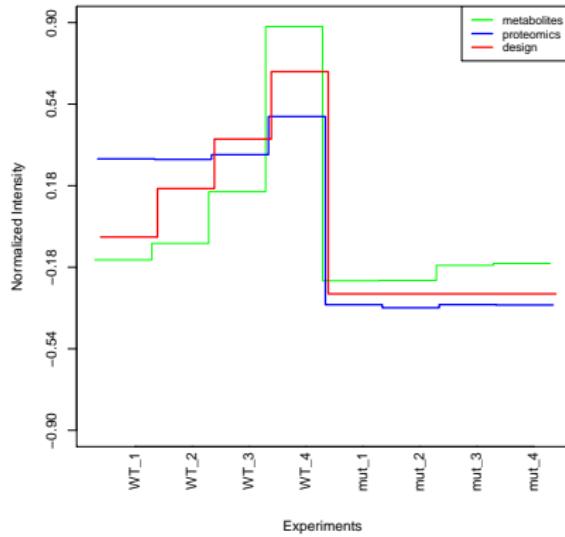
Reconsider experimental design

1. canonical variable with corr= 0.8613

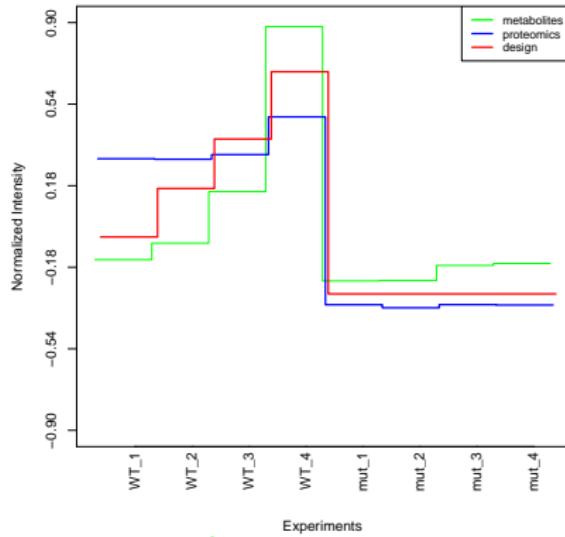


Reconsider experimental design

1. canonical variable with corr= 0.8613



1. canonical variable with corr= 0.8612

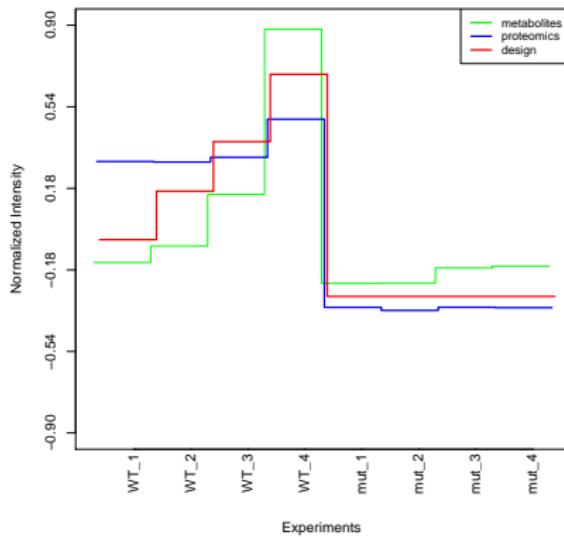


✓ reproducible

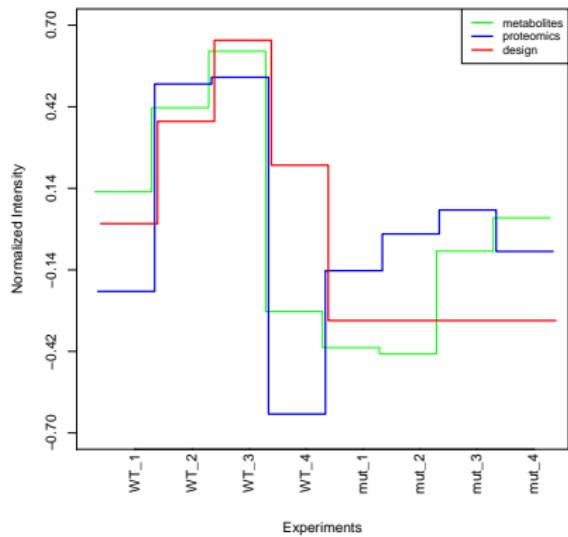


Reconsider experimental design

1. canonical variable with corr= 0.8613



2. canonical variable with corr= 0.5035



✗ only 1 meaningful CV



Challenges

Metabolomics
 $m_1, m_2, m_3, \dots, m_p$

Proteomics
 $p_1, p_2, p_3, \dots, p_q$

Transcriptomics
 $t_1, t_2, t_3, \dots, t_r$

- differently sized data sets
 - prefiltering
 - enforce sparsity



Challenges

- differently sized data sets
 - prefiltering
 - enforce sparsity
- biological interpretation of statistical procedure
 - biologically relevant may not be best correlated



Challenges

- differently sized data sets
 - prefiltering
 - enforce sparsity
- biological interpretation of statistical procedure
 - biologically relevant may not be best correlated
- missing values
 - in *omics* data sets
 - in design: too "univariate"



Challenges

- differently sized data sets
 - prefiltering
 - enforce sparsity
- biological interpretation of statistical procedure
 - biologically relevant may not be best correlated
- missing values
 - in *omics* data sets
 - in design: too "univariate"
- significance level of correlations



Challenges

- differently sized data sets
 - prefiltering
 - enforce sparsity
- biological interpretation of statistical procedure
 - biologically relevant may not be best correlated
- missing values
 - in *omics* data sets
 - in design: too "univariate"
- significance level of correlations
- if > 1 CV, order not 100% reproducible



Acknowledgement

Credits to

- Steffen Neumann
- Andrea Thum
- Ines Lassowskat
- Christoph Böttcher
- Daniel Scheer
- group MSBI



Thank you for your attention

