



Funded by the Seventh Framework
Programme of the European Union



Project full title:
Hepatic and Cardiac Toxicity Systems modelling

Project acronym:
HeCaTos

Collaborative project
HEALTH.2013.1.3.-1:
Modelling toxic response in case studies for predictive human safety assessment

FP7-HEALTH-2013-INNOVATION-1-602156-HeCaTos

Deliverable Report D12.2:
Report on results of the model analysis and comparison

Edited by
Work package 12

Due date of deliverable: M24
Actual submission date: M25

Start date of project: October, 2013

Duration: 60 months

Maastricht University (UM)

Project co-funded by the European Commission within the 7th Framework Programme (2013-2018)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Contributions to deliverable - Internal review procedure

Deliverable produced by:	Date:
Moritz Beber - Partner MPIMG	September 2015
Ralf Herwig - Partner MPIMG	October 2015
Deliverable internally reviewed by:	Date:
Jos Kleinjans - Partner UM	October 2015

Contents

Publishable Summary.....	3
Objectives	3
Introduction	3
Results.....	4
Modelling expression time series	5
Gene scoring from complex experiments.....	7
Pairwise Pearson correlation of gene scores across all treatments.	9
Pathway scoring from complex experiments	9
Pairwise Pearson correlation of RPR scores for all treatment experiments.....	11
References	12

PUBLISHABLE SUMMARY

An important characteristic of drug treatment experiments that is crucial for analyzing dynamic changes of the cellular systems is the variation that is induced by the drug over time and by different dosages. Thus, by nature most drug experiments are performed with multiple doses and at several time points and a key question is how to cope with this complex experimental set-up, in particular with respect to the identification of differentially expressed genes, proteins and metabolites and associated molecular pathways.

In this deliverable MPIMG has implemented, tested and compared several strategies how to model multi-dosage and time-point data. Methods rely on the modelling pipeline that has been set up in WP2 (cf. Deliverable Report 2.3). The major output of that modelling pipeline is (given a drug treatment experiment) a drug-specific network module that consists of interconnected proteins or pathways that predict the molecular effects of the drug and that can be connected to key molecular events in the construction of adverse outcome pathways (AOPs) and are predictive of the underlying toxic phenotype. An important part in this pipeline is the question how to best score differentially expressed genes (DEGs) and pathways (DEPs) from complex toxicity experiments so that the variations through time and dosages are taken into account.

In D12.2 we have addressed this question, and implemented and tested several gene scoring methods that are able to take into account the model variance imposed by the different time points and doses. We have tested the different methods with benchmark data identified in WP12 (cf. MS.26 milestone report).

OBJECTIVES

The proper modelling of time series experiments is a key aspect of toxicogenomics. In this deliverable we report on the different scoring methods for identifying differentially expressed genes (DEGs) and pathways (DEPs) from multiple dosage and time-series data. We test the methods with public benchmark data and compare their performance with inherent criteria.

INTRODUCTION

The computational modeling pipeline is illustrated in Figure 1. Starting from a compound and available omics data (e.g. transcriptome, proteome, DNA-protein binding) parallel steps are performed for identifying associated biomolecules. One strategy for identifying biomolecules is based on OMICS differential expression analysis for each single experimental time point and dosage regimen, the second on analysis of data using multivariate statistics that include the full time and dose range of the available experiments.

In a further step the biomolecules are associated with interaction data in order to describe the compound effects at the level of biological networks. This is done either with pre-defined pathways or with protein-protein interactions integrated from MPIMG's ConsensusPathDB resource. The OMICS data is used to weight the corresponding nodes in the networks and graph algorithms are used to identify key modules in the networks that agglomerate a lot of the observed changes.

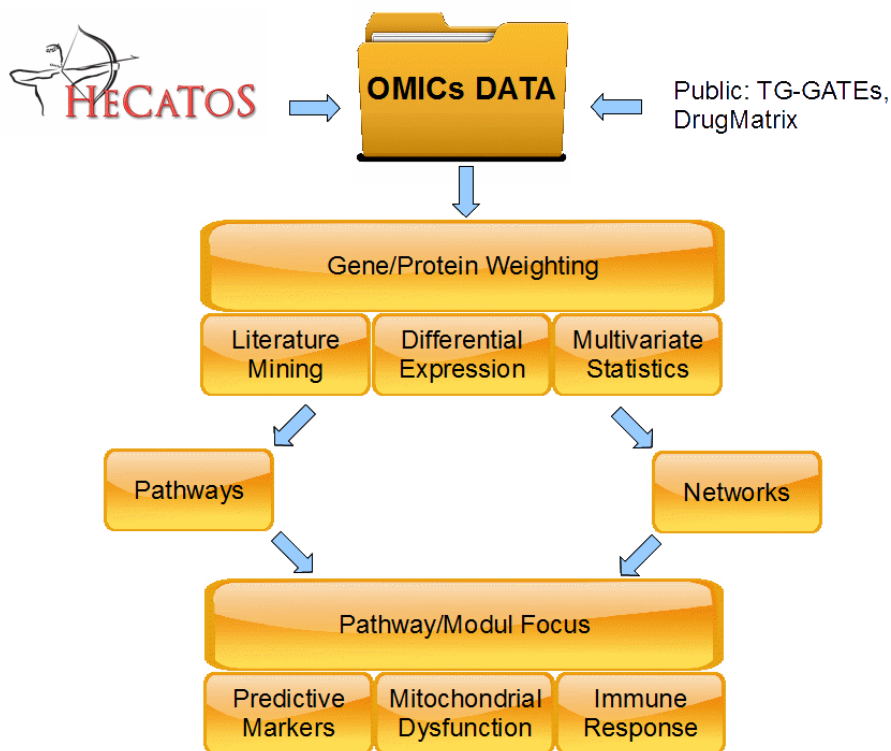


Figure 1: Schema of the modelling pipeline.

RESULTS

Drug data is typically generated at different doses and response of the cells under study is measured at several time points after drug delivery. In order to cope with this set-up and to identify genes that change over time and dosage it is typically not sufficient to analyse a single time-point/dosage experiment but to take into account changes over time and at several dosages. There are several approaches to the modelling of time-dosage drug data, for example STEM (Ernst and Bar-Joseph (2006) BMC Bioinformatics 7:191) that is mainly a visualization-based clustering approach, Bayesian networks (e.g. Kim et al. (2003) Brief Bioinf 4:228-235) or ANOVA (e.g. Ritchie et al. (2015) Nucleic Acids Res 43:e47).

MPIMG has analyzed all data sets that were made available from DrugMatrix and TG-GATES studies (in total 1,152 experiments). Fig. 2 gives an overview on the processed data sets. It displays five different ANOVA models with respect to the number of dosage and time-points. As can be seen from the 1,152 experiments only approx. 50% are available at three dosages and at least three time-points ($179+344+7+1=531$) and could thus be modelled by ANOVA. On the other hand, the rest of the data is available at only few dose or time points so that they are not accessible for models such as ANOVA leading to larger variations in the identification of DEGs.

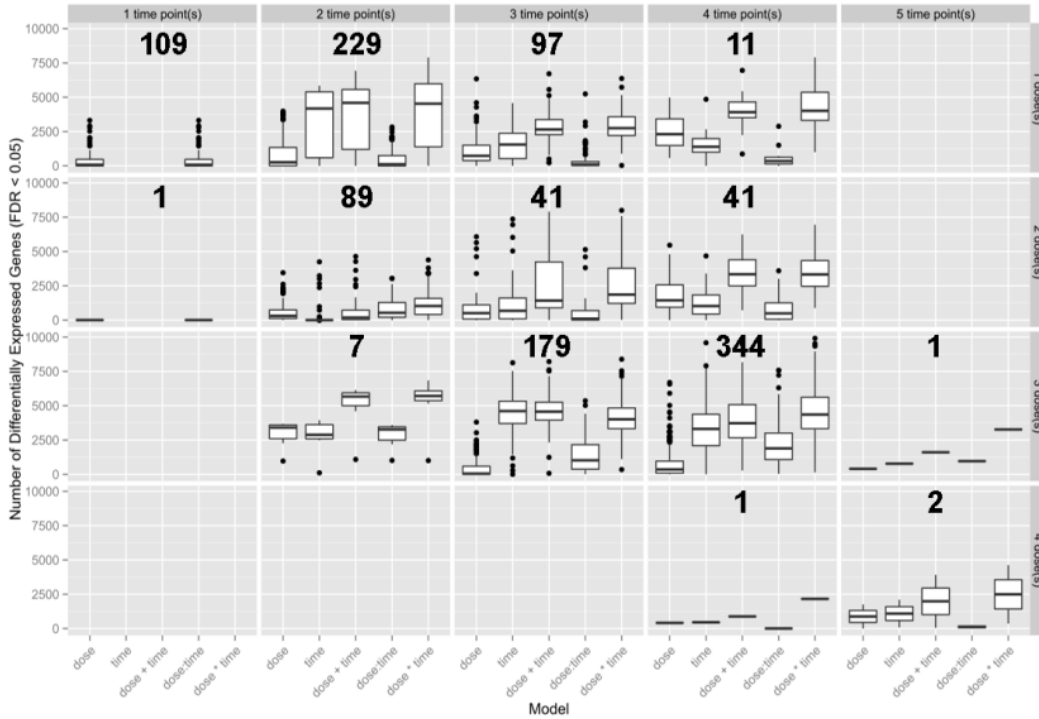


Figure 2: ANOVA models tested with TG-GATES and DrugMatrix data. X-axes describe the five models under consideration: dose - dosage factor only; time - time factor only; dose+time - time and dosage factor; dose:time - interaction factor only; dose*time - full model. Box plots represent number of differentially expressed genes after FDR correction. A) Experiments with respect to time and dose points. B) Experiments with respect to the biological material under consideration.

Modelling expression time series

The treatment experiments contained within the TG-GATES and DrugMatrix studies span $1 \leq x_t \leq 5$ time points and $1 \leq x_d \leq 4$ dosages. They consist of mRNA abundance measured by Affymetrix microarrays.

Linear Continuous Models: One way of modeling gene expression responses is by linear regression trying to estimate the expression level, y_i , of a gene i using time and concentration level as continuous variables.

$$y_i = \beta_{i,t}x_{i,t} + \beta_{i,d}x_{i,d} + \beta_{i,0} + \epsilon$$

Of course, this assumes a linear response of gene expression both in time and concentration level. For a short time frame after treatment, i.e., less than a day, this assumption is valid for many but not all genes. However, there are 447 treatment experiments in the DrugMatrix study that contain only one time point or one tested dose. For those experiments the above linear model approach fails.

Fitted Time-Series: The idea of fitting the continuous variable time with splines and regarding the different concentration levels as a categorical variable has the advantage that it can model non-linear time-series but the time point interpolation suffers from the same problem as the linear model approach in that there may be too few time points.

Categorical Model: An alternative approach is to treat each unique combination of time and dose as its own group and then compare the time-matched gene expression changes between some concentration level and the untreated samples. This is our preferred method since it is independent of the number of time points or concentration levels tested and will also accurately detect gene expression changes when there is a non-linear dependency on time.

As an example, the following table describes a treatment experiment with two time points, two doses, and three replicates each.

cel	time	dose	Factor
1.CEL	2h	control	c.2h.control
2.CEL	2h	control	c.2h.control
3.CEL	2h	control	c.2h.control
4.CEL	2h	therap.	c.2h.therap.
5.CEL	2h	therap.	c.2h.therap.
6.CEL	2h	therap.	c.2h.therap.
7.CEL	2h	toxic	c.2h.toxic
8.CEL	2h	toxic	c.2h.toxic
9.CEL	2h	toxic	c.2h.toxic
10.CEL	8h	control	c.8h.control
11.CEL	8h	control	c.8h.control
12.CEL	8h	control	c.8h.control
13.CEL	8h	therap.	c.8h.therap.
14.CEL	8h	therap.	c.8h.therap.
15.CEL	8h	therap.	c.8h.therap.
16.CEL	8h	toxic	c.8h.toxic
17.CEL	8h	toxic	c.8h.toxic
18.CEL	8h	toxic	c.8h.toxic

Treatment experiment description linking to the CEL files.

The contrast design that leads to the \log_2 fold changes is given below. In this case there are four contrasts of interest and we refer to the whole set as J .

	c.2h.therap. - c.2h.control	c.2h.toxic - c.2h.control	c.8h.therap. - c.8h.control	c.8h.toxic - c.8h.control
c.2h.control	-1	-1	0	0
c.8h.control	0	0	-1	-1
c.2h.therap.	1	0	0	0
c.8h.therap.	0	0	1	0
c.2h.toxic	0	1	0	0
c.8h.toxic	0	0	0	1

The contrasts design used to compute \log_2 fold changes.

Gene scoring from complex experiments

The proper scoring of genes is the basis for the identification of differentially expressed genes (DEGs) and for computing pathway scores from the gene expression response. Gene scores form the basis of the relative pathway response (RPR) scores further downstream in our analysis and should be chosen with care. Our original goal was to evaluate the genetic response to drug treatment. We therefore discard the information at what time and concentration level the response is strongest. Additionally, we wish to carry forward through the analysis information about the probability of differential gene expression. We do this by weighting the score with the p-value and then considering only RPR scores as significant that are above the 95% quantile. In the following, we define and motivate four different scoring schemes:

1. An obvious measure of a gene response is the fold change. The experiment-wise gene score g_i of gene i is then weighted by the p-value of the corresponding $FC_{i,j}$ in contrast j and the maximum over all these scores is taken,

$$g_i = \max[|\log_2(FC_{i,j})\log_{10}(p_{i,j})| \forall j \in J].$$

2. A slight adjustment is to use the q-value of the F-test over all contrasts J as a weighting factor. The q-value is the p-value adjusted for multiple testing using a false discovery rate (FDR).

$$g_i = \max[|\log_2(FC_{i,j})\log_{10}(q_i^{(F)})| \forall j \in J]$$

3. Instead of the pure FC, the effect-size, such as Cohen's d , is a logical measure that is simply the FC divided by the moderated standard deviation of the gene expression.

$$g_i = \max[|\log_2(d_{i,j})\log_{10}(p_{i,j})| \forall j \in J]$$

4. Lastly, the F-statistic is a reasonable measure by itself and can then be weighted with its q-value.

$$g_i = |F_i \log_{10}(q_i^{(F)})|$$

We then show the relationship between these different scoring schemes using three different treatment experiments as examples, as well as tables showing the pairwise correlation between the scores over all treatment experiments (Fig. 3-5).

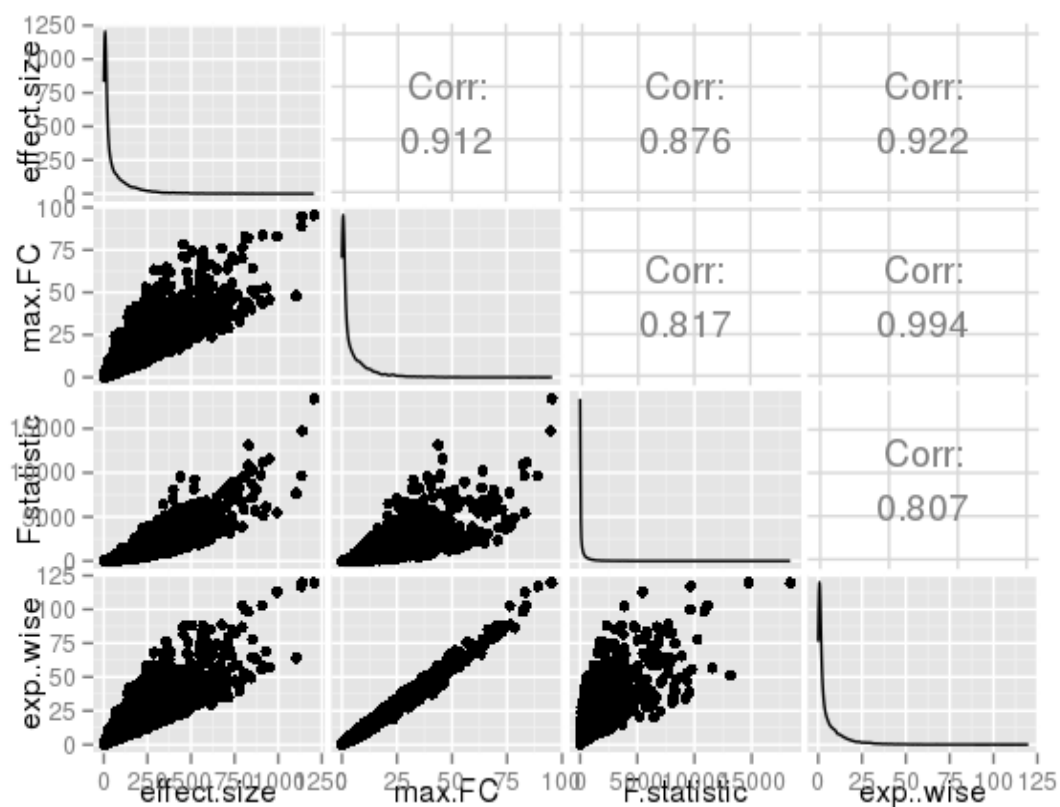


Figure 3: Relationship between gene scores as measured in acetaminophen-treated hepatocytes.

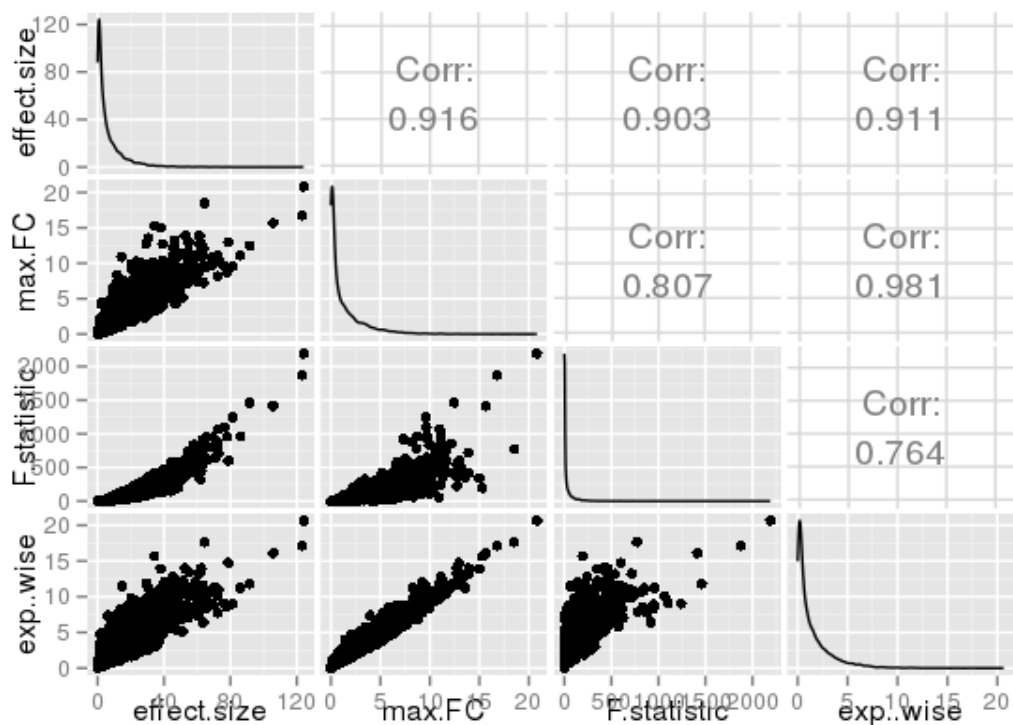


Figure 4: Relationship between gene scores as measured in doxorubicin-treated rat heart tissue.

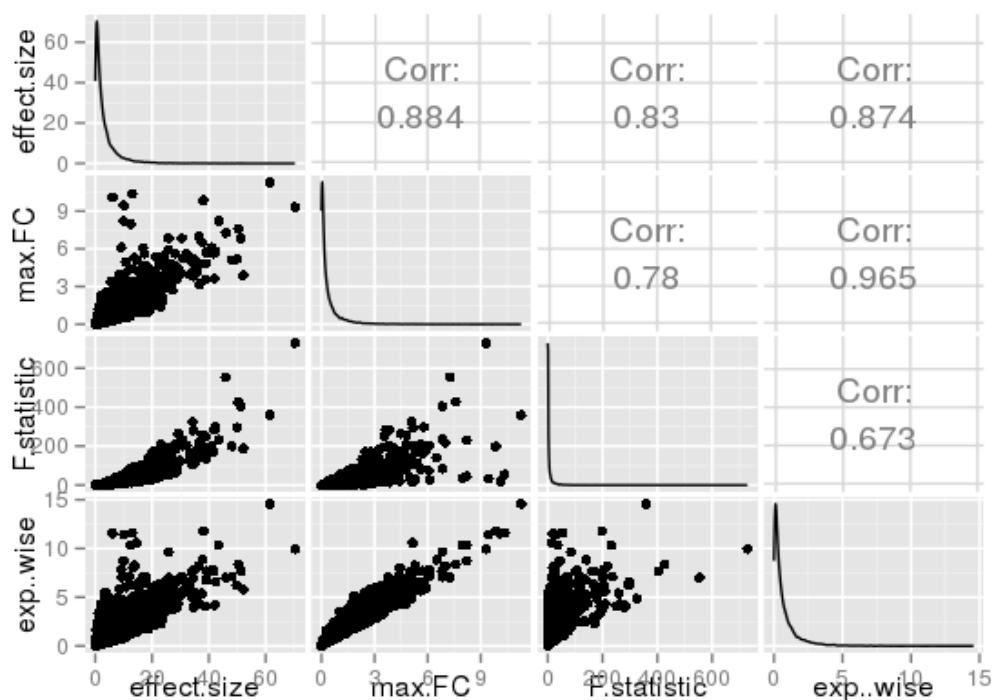


Figure 5: Relationship between gene scores as measured in idarubicin-treated rat heart tissue.

	effect-size	max FC	F-statistic	exp.-wise
effect-size	1.0000	0.8482	0.8407	0.8763
max FC	0.8482	1.0000	0.6797	0.9801
F-statistic	0.8407	0.6797	1.0000	0.6959
exp.-wise	0.8763	0.9801	0.6959	1.0000

Pairwise Pearson correlation of gene scores across all treatments.

As can be seen there are slight variations between the different methods although most of the gene scores correlate quite well. In particular, the F-statistic that takes into account the different experimental groups according to time and dosage, deviates from the other measures.

Pathway scoring from complex experiments

Given the computed gene scores according to each of the described method, a pathway score can be computed as the average gene score of all genes assigned to the pathway. Additionally, relative pathway response (RPR) scores are built by dividing each pathway score by the median pathway score of that particular experiment in order to make RPRs comparable across experiments. This has been used to perform the initial use case described in Delivery Report D12.1.

$$RPR_i = \log_2 \left(\frac{\text{mean}(g)}{\text{median}(RPR)} \right).$$

In the sequel we compute the RPR scores based on the different computations of the gene scores (Fig. 6-8).

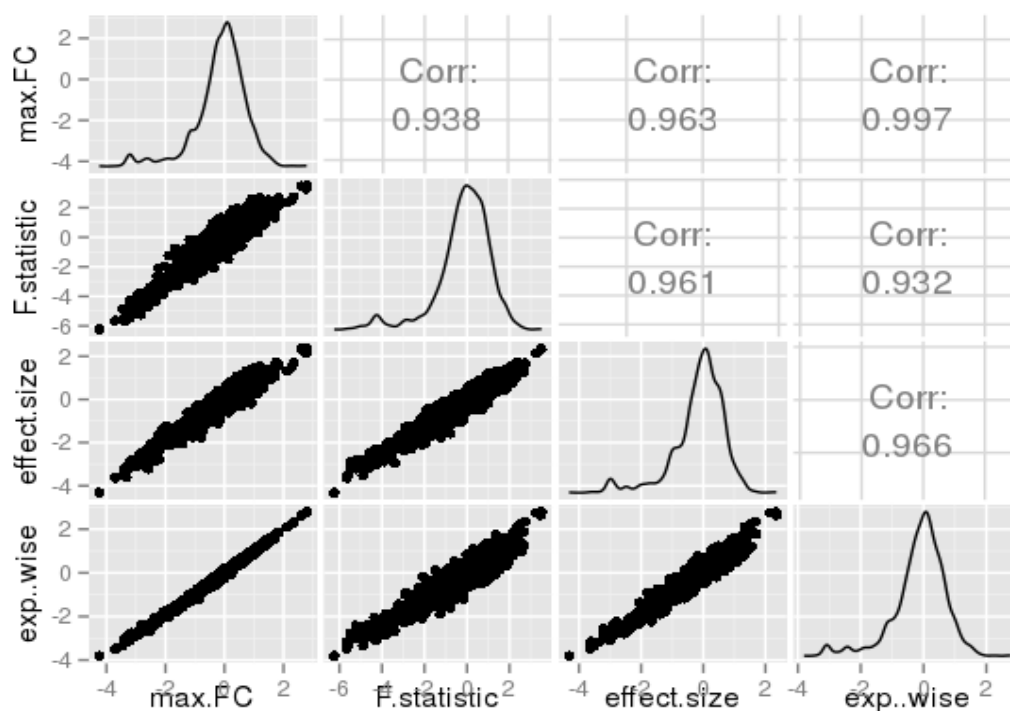


Figure 6: Relationship between RPR scores as measured in acetaminophen-treated hepatocytes.

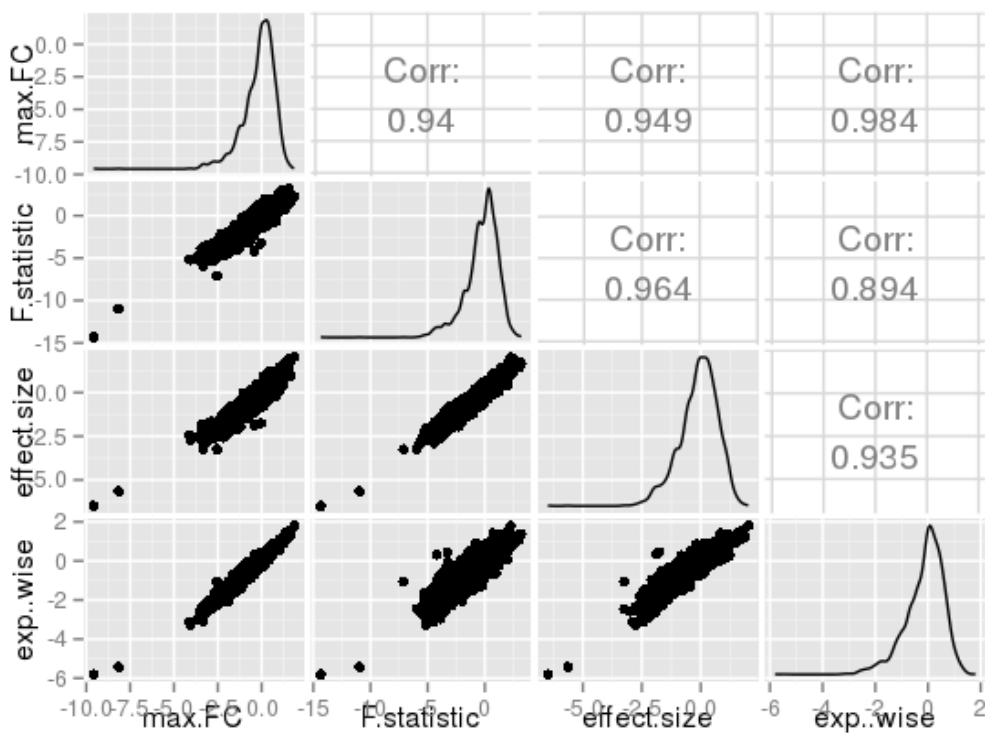


Figure 7: Relationship between RPR scores as measured in doxorubicin-treated rat heart tissue.

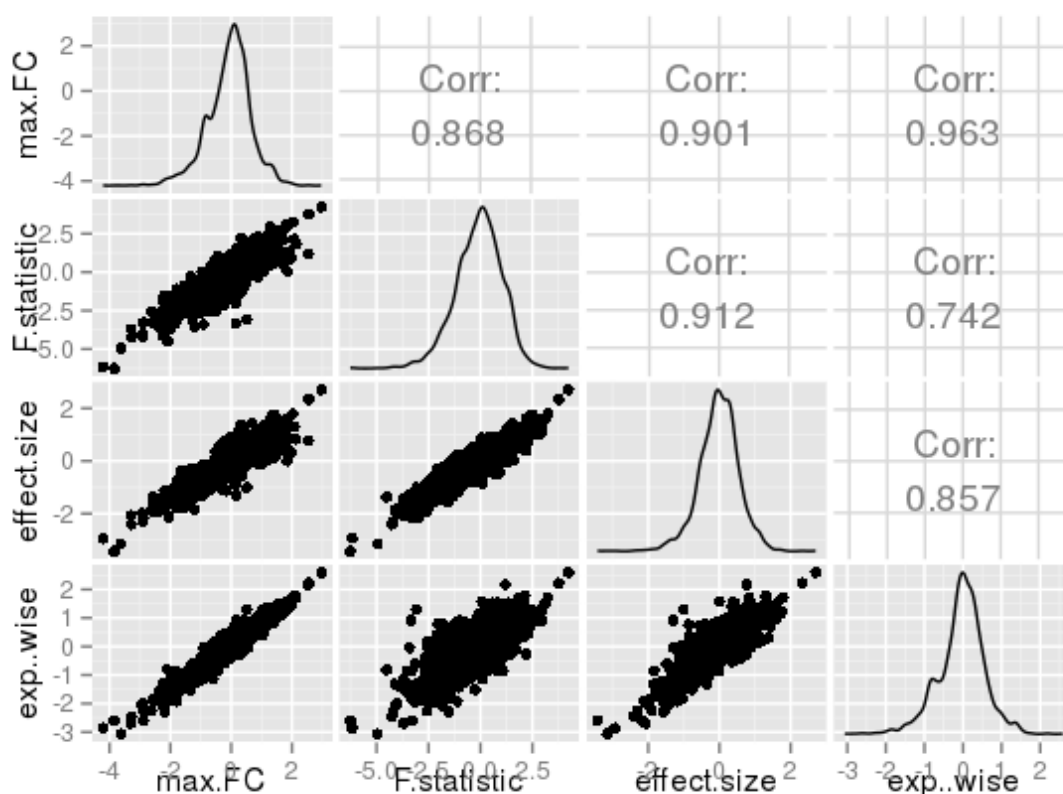


Figure 8: Relationship between RPR scores as measured in idarubicin-treated rat heart tissue.

	max FC	F-statistic	effect-size	exp.-wise
max FC	1.0000	0.8888	0.9028	0.9787
F-statistic	0.8888	1.0000	0.9474	0.8532
effect-size	0.9028	0.9474	1.0000	0.9087
exp.-wise	0.9787	0.8532	0.9087	1.0000

Pairwise Pearson correlation of RPR scores for all treatment experiments.

As can be seen from the tables above the pairwise correlation increases when extrapolating from gene to pathway scoring, in particular the correlation of effect size (a measure for maximal deviation across time points and doses) and F-statistic (a measure for deviation of variability across time points and doses) which increases from 0.8407 to 0.9474 when considering all DrugMatrix treatment experiments. This reflects the stabilizing effect of pathway-based analysis.

As a result of this comparison we have incorporated effect size as scoring method of choice for gene and pathway scoring which can be applied to all experiments shown in Fig. 2.

REFERENCES

- Hendrickx DM, Aerts HJ, Caiment F, Clark D, Ebbels TM, Evelo CT, Gmuender H, Hebels DG, Herwig R, Hescheler J, Jennen DG, Jetten MJ, Kanterakis S, Keun HC, Matser V, Overington JP, Pilicheva E, Sarkans U, Segura-Lepe MP, Sotiriadou I, Wittenberger T, Wittwehr C, Zanzi A, Kleinjans JC. (2015) diXa: a data infrastructure for chemical safety assessment. *Bioinformatics*. 31(9):1505-7.
- Herwig R. (2014) Computational modelling of drug response with application to neuroscience. *Dialogues Clin Neurosci*, 16(4):465-77.
- Kamburov A, Stelzl U, Lehrach H, Herwig R (2013) The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*, 41:D793-D800.