Biologists and Bioinformaticians team up and explore the interplay between UV radiation, P53 modification and gene expression

Oskar Bruning[1#], Wendy Bruins[2#], Martijs J. Jonker[1], Edwin Zwart[2], Tessa van der Hoeven-Dillerop[1], Jeroen L.A. Pennings[2], Han Rauwerda[1], Annemieke de Vries[2], and Timo M. Breit[1]

1 MicroArray Department (MAD), Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam (UvA), Amsterdam, the Netherlands (www.microarray.nl).
2 Laboratory of Toxicology, Pathology and Genetics (TOX), National Institute of Public Health and the Environment (RIVM), Bilthoven, the Netherlands.
# These authors contributed equally to this work

Email: bruning@science.uva.nl / mjonker@science.uva.nl

Abstract:
Transcription factor P53 plays a crucial role as DNA-damage sensor preventing accumulation of genetic lesions and thus tumor development. Proper functioning of this protein requires post-translational modification, predominantly phosphorylation. Biologists and bioinformaticians worked closely together to set up a robust microarray study in mouse embryonic fibroblast (MEF) cells, elucidating the relevance of a specific P53 phosphorylation site (S389) for the expression of many genes in response to UV radiation. The goal of this study is to analyze:
(I) the effect of absence of p53.S389 phosphorylation on the basal gene-expression levels of p53-dependent genes.
(II) the transcriptome response of wild-type MEFs to UV irradiation in time
(III) the effect of absence of p53.S389 phosphorylation on UV response in time.
The P53 dependent genes were identified in-vitro rather than in-silico, by including a P53 knock out (KO) cell line in the experiment. The wild type (WT) response and the KO response were used as "reference profiles" relative to which the gene expression of the mutant, with disrupted S389 functioning (SA), could be compared in a full factorial design (Fig. 1). The data was analyzed with two stage linear models, taking into account variance components in the experiment. The models were reparameterized using linear contrasts of model coefficients in order to answer the three research questions effectively.

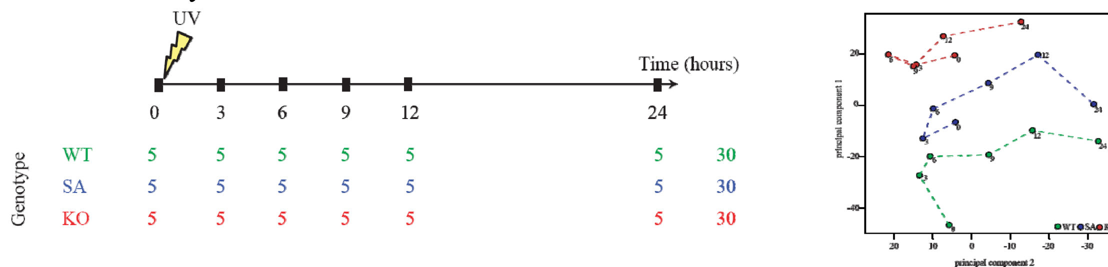

Fig. 1. Experimental design (left) for analyzing the transcriptional effect of disrupted S389 phosphorylation of P53, and principal components analyses (PCA; right) of the data. WT indicates wild type cells, SA indicates cells with disrupted S389 phosphorylation, and KO indicates cells with a knocked out P53 gene.

For instance, after normalization the first research question was answered using

$$Y_{ijktgr} = \mu + A_i + D_j + (AD)_{ij} + r_{ijktgr}$$

$$r_{ijktgr} = G_g + \alpha_{kg} + \beta_{tg} + S_{ig} + \varepsilon_{ijktgr}$$

modeling array effects (A, i = 1,...,90), dye effects (D, j = 1,2), genotype effects and (α, k = 1,2,3), time effects (β, t = 1..6) as fixed effects, and spot effects (S) as random, for all g = 1,…,21,766 genes and r =

1,…,5 independent replicates. Gene level basal differences between genotypes were quantified by inferring

$$H_0: \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_{1g} \\ \alpha_{2g} \\ \alpha_{3g} \end{bmatrix} = 0$$

where WT is indicated by k = 1, SA by k = 2 and KO by k = 3. The two other research questions were analyzed by contrasting the time-effects in the wild type data, and by contrasting the genotype × time-effects respectively. Contrast analyses allow fine-tuning the statistical model to answer specific research questions, whilst estimating a robust gene specific error and taking into account experimental variance components. Emperical bayes methods can still be included, and the results are directly biologically interpretable. For instance, the results of analyzing the first research question enabled the biologists to classify the genes that were affected by phosphorylation, into P53 dependent genes and P53 independent genes (Fig. 2A). Further analyses showed that these genes were either repressed or induced by P53.

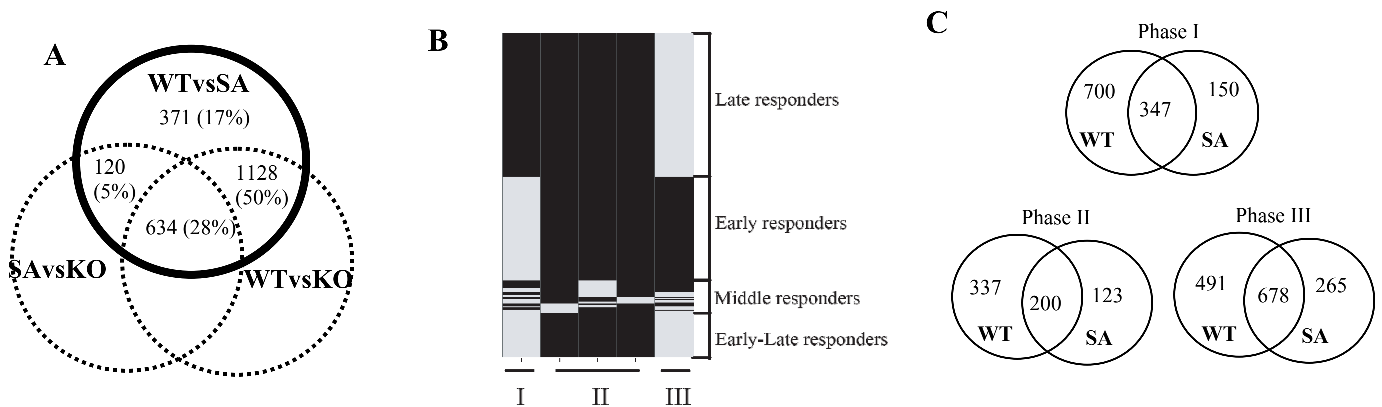

Fig. 2. Biological classification of P53 related genes determined after analyzing research question I (Fig. 2A), research question II (Fig. 2B), and research question III (Fig. 2C). See text for details

Analyzing the second research question, the WT response to UV, yielded a classification of the genes into early, late, middle and early-late responders, and showed a "triphasic" transcriptional response to UV radiation (in grey in the heatmap, Fig 2B). Analyzing the third research question enabled the identification of WT and SA specific responding genes for each time point. All gene classes were further interpreted with overrepresentation analyses and gene set enrichment analyses, using gene specific information from the gene ontology, kegg, and the molecular signature database.

In conclusion, close collaboration between biologists and bioinformaticians allows extracting biologically relevant information from complex multidimensional data sets, possibly not recognized by bioinformaticians only. It was for instance recognized that S389 phosphorylation is involved in repression and induction of P53 dependent genes, which specifically could be revealed after measuring the transcriptional response of the KO. In addition, the transcriptional response to UV was triphasic, and S389 phosphorylation has a strong effect on basal gene expression of many genes, but the effect on UV-response is more subtle. Analyzing linear contrasts is a robust and versatile tool as it allows constructing tailor-made models to answer specific research questions, whilst estimating a robust gene specific error and taking into account variance components.