Improving design, execution and analysis of transcriptomics experimentation
Bruning, O.

Citation for published version (APA):
Bruning, O. (2015). Improving design, execution and analysis of transcriptomics experimentation

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Serious complications in gene-expression studies with stress perturbation: An example of UV-exposed p53-mutant mouse embryonic fibroblasts

Abstract

Reanalysis of our UV study of p53-mutant mouse embryonic fibroblasts revealed an intriguing orchestration of massive transcriptome responses. However, close scrutiny of the data uncovered an affected mRNA/rRNA ratio, effectively inhibiting valid data analysis. UV-dose range finding showed low-dose UV specific- and high-dose stress-related responses, which represent a plea for UV dose range finding in experimental design.

This chapter has been published as:

Acknowledgements

This work was supported by the Dutch Cancer Society (KWF) grant 2000-2352, NIH/NIEHS (Comparative Mouse Genomics Centers Consortium) grant 1U01 ES11044-02, NIEHS 72-SNP grant 1R01ES015587-01A2, and BSIK grant through the Netherlands Genomics Initiative (NGI) in the context of the BioRange-II program of the Netherlands Bioinformatics Centre (NBIC).

Abbreviations and Acronyms

DEG differentially-expressed gene as compared to time point 0h
GEO Gene Expression Omnibus
h hour
p53KO gene p53 knock-out mutant
p53SA mutated p53 gene, resulting in a substitution of serine to alanine amino-acid residue at protein position 389 and thus absent phosphorylation.
MEF mouse embryonic fibroblast
SAGE serial analysis of gene expression
SOMS self-organizing maps
WT wild-type

Running Title (50 characters)
Serious complications in gene-expression studies.
Introduction

Transcriptomics experimentation has become an important tool in deciphering gene-expression response and regulation. As can be read for instance from the Gene Expression Omnibus (GEO) repository (http://www.ncbi.nlm.nih.gov/geo/), numerous microarray and serial analysis of gene expression (SAGE) experiments have already been performed and many more will follow, also using next-generation transcriptome sequencing. Over the years, the major technical drawbacks of the microarray and SAGE technologies have been successfully tackled, as undoubtedly will be the case with the current and future next-generation sequencing technologies.

Transcriptomics studies that aim for biomarker detection in classification and/or diagnostics are often quite successful, as biological systems robustly and reproducibly respond to stimuli or perturbations. Unfortunately, many mechanism-focused transcriptomics experiments seem to stall at endless lists of genes that are differentially expressed between all imaginable experimental contrasts. Hence, the actual biological knowledge obtained from all these interesting experiments appears to be lagging behind. This might be caused by the complexity of the cellular mechanisms, lack of proficient bioinformatics and statistical expertise in the biology domain, insufficient replication or measurement-points density in the experimental design, limited knowledge about the individual genes and their products, and so on. In this article, we would like to emphasize an often-overlooked aspect: the effect of too much stress in perturbation experiments and the consequences this has on omics experiments and their interpretation. We will explicate our point of view by using a “real-life” example that we encountered while analyzing a p53-focussed transcriptomics experiment.

Bioinformatics analyses of UV-exposed (mutant) p53 MEFs

In 2008, we published an article on the comprehensive analysis of the transcriptome response to UV exposure in wild-type and mutant p53 mouse embryonic fibroblasts (MEFs). Although we obtained several new and important insights regarding this response, we felt unsatisfied by the overall interpretation of the transcriptomics data. We therefore set out to employ creative bioinformatics analyses to further understand the involved cellular mechanisms. Eventually we appeared to be quite successful in our approach and obtained many new insights, the most important ones of which we will here summarize and illustrate in Figure 1 & Supplementary information.

As reference, the experimental design and heatmap of differentially-expressed genes (DEGs) over time in wild-type (WT) are displayed in Figure 1A and E, respectively. All DEGs in time are determined per time point as compared to time point 0h. Simply counting these DEGs revealed interesting observations between the p53 genotypes over time (Fig. 1B):

- The WT genotype displayed a persistent high number of DEGs, whereas the p53KO seems to return to its original state. The p53SA response looks biphasic/
Serious complications in gene-expression studies

mixed (Fig. 1B);

- As different genes were differentially expressed over time, in total 34% of all measured genes in WT were detected as DEG (Table S1);

- The identified DEGs (99%) were unidirectional, either up- or down-regulated (Fig. 1E);

A subsequent eigen-gene profile analysis, to identify the most prominent gene-expression profiles of the separate up- and down-regulated DEGs at each time point as compared to time point 0h revealed some exciting DEG features:

- The far majority of DEGs were present in eigen-gene profile 1, which explained typically over 90% of the variation (Table S1).

- Each eigen-gene profile 1 showed differential gene-expression from time point 0h on (Fig. 1F);

- The eigen-gene profile at each time point peaked almost invariably only at the same time point of the involved DEGs (Fig. 1F);

- The eigen-gene profile 1 profiles of up-regulated DEGs over time were mirror images of those of the down-regulated DEGs (Fig. 1F).

- There was little difference between the eigen-gene profiles 1 of the different genotypes (Fig. S2).

---

Figure 1 – Summary of transcriptome responses to UV exposure in WT and mutant p53 MEFs.

(A) Experimental design of the experiment with 20J UV-C exposure at t = 0h. WT: wild-type MEFs; p53SA: p53S389A mutant MEFs; and p53KO: p53 knock-out MEFs. The colors represent genotypes in B, C, and D. (B) Numbers of UV DEGs over time for all genotypes. (C) rRNA expression profiles of raw signal data over time in WT MEFs. (D) Relative mRNA/rRNA ratio for all genotypes. (E) Heatmap showing DEGs up- or down-regulated in WT per time point compared to time point 0h. The colors represent up- and down-regulated DEGs in E and F. (F) Normalized expression profiles of up- and down-regulated eigen-gene profiles 1 per time point for WT. Cf. Fig. S2 for the p53-mutant genotypes.
Chapter 3

All these and some non-mentioned results, led us to hypothesize about orchestration of this massive transcriptome response to UV exposure in WT and mutant MEFs. This orchestration could be a result of one, or several, general transcription factors that would, distributed over time, turn genes on or off due to promoter sites with decreasing transcription factor binding affinity. One of those transcription factors of course could be p53, in which we theorized that the level of activated p53 complex was increased for specific cellular mechanisms such as DNA repair, cell-cycle arrest and apoptosis as a result of the UV exposure, while binding to many other promoter sites, and thus differential gene expression was merely a byproduct. This is in line with the concept that statistical significance does not equal biological significance.

Potential transcriptomics artifacts

As we were drafting an article to publish our findings, we grew unease with our results, even though similar findings in other organisms have been reported before. With respect to our hesitation, especially the high percentage of DEGs (34%) combined with their symmetric preferential gene-expression profile in all time points and genotypes seemed suspicious.

Although stress often induces pleiotropic effects by affecting mRNA synthesis and even alternative splicing, it is the extend of these effects that could lead to misinterpretation of transcriptome data. Most statistical data analysis methods in transcriptomics have a strict requirement that not too many genes exhibit changed expression in the experiment. To illustrate this by a mind experiment: Assume 1,000 genes having an average expression X; upon exposure, the RNA concentration of 50% of the genes remains the same and of the other 50% it reduces 4 fold; thus the total amount of RNA reduces from 1,000X to 625X; After statistical normalization, the unchanged genes will be detected as DEGs with on average 1.6 fold RNA increase, whereas the 4 fold decreased genes will be detected as DEGs with on average 2.5 fold RNA decrease. Besides faulty fold changes, this also means that many genes are erroneously labeled as DEG, which has major impact on the downstream analysis and biological interpretation. We suspect that such might be the case in the human skin UVB study where ~40% of all genes are identified as DEGs, even though several melanocyte-specific genes could be confirmed by immunohistochemistry.

For our study, we considered three possible artefacts: statistical, technological, and biological. A statistical artefact could be so-called “over-fitting” of the data in normalization procedures. A technological explanation could be a change in the mRNA/rRNA ratio due to the treatment. A biological effect could be a non-specific stress response with many non-relevant DEGs.

We exhaustively checked for a statistical artefact, which seems not to be the case, as we observed most phenomena also in the raw data.
The effect of a changing mRNA/rRNA ratio

As to rRNA measurements; rRNAs do normally have no poly(A) tail, so with the applied oligo(d(T)) protocol no signal should be detected by the rRNA probes on a microarray. Hence, per fixed amount of total RNA, only the mRNA fraction is labeled into cRNA for each sample. Subsequently, a fixed amount of cRNA is hybridized to a microarray for each sample. Unexpectedly, the analysis revealed that all probes on our mouse microarray representing three types of rRNA had clear signals. The presence of these signals can be explained by “leakage” due to non-specific oligo-d(T) binding or by polyadenylation of rRNA, which occurs as a part of the rRNA degradation process 22-24. The rRNA gene-expression profiles all showed a marked response in all genotypes (Fig. 1C and S3). Compared to the untreated control (t = 0h), a clear and sudden increase was present in all treated samples, with a maximum fold change of about 32. So, given the options above, this increase can be caused by: an increase in rRNA concentration; an overall decrease in mRNA concentration; an increase in rRNA polyadenylation/degradation; or any combination of these effects.

To estimate a relative mRNA/rRNA ratio, we evaluated per sample the total RNA yield—as a derivative for rRNA—and cRNA yield—as a derivative for mRNA. We observed a trend of an about twofold increasing mRNA/rRNA ratio from about 6 hours on in WT MEFs (Fig. 1D), which is primarily caused by a decrease in total RNA yield. This was also present in p53SA, but absent in p53KO MEFs. Together, this means that the mRNA/rRNA ratios are affected by the exposure in a time- and genotype-dependent matter. This severely hampers the normalization of the between-samples microarray data and effectively inhibits their analysis. This phenomenon seems to be independent of the above-described boost in rRNA signals at 3 hours.

It is generally known that high-stress induces the decrease of mRNA from most genes 25 as well as rRNA 26. UV exposure studies show that mRNA synthesis changes both in relation to dose 5 as well as time 18. Although for instance Offmann et al 5, apply UV doses (at 12 hours) that induce approximate equal mRNA synthesis reduction (at 16 hours), this only considers newly synthesized mRNA and not mRNA degradation or major rRNA concentration changes.

It is impossible to determine from our experiment what causes the observed changes. The known rRNA polyadenylation before degradation might explain the increased rRNA microarray signals we observed. It also might well be that in WT and SA MEFs after 6 hours, apoptosis occurs and dead cells are washed away in the harvest procedure. Even though we do not know the causes, we still felt that changing mRNA/rRNA ratios will have a profound effect on determining differential gene expression. It is therefore in our opinion quite important to keep track of the rRNA and mRNA concentrations per cell in gene-expression studies.

Dose range finding reveals non-specific stress response

However, before we were to redo our original experiment, we felt that avoiding changes
in rRNA and mRNA concentrations would be better than correcting for them. Thus, we first performed a dose range finding study. After all, 20 J radiation, which is a commonly used UV-C dose in cellular experiments \cite{27, 28}, is quite high.

The results of a limited dose range finding study, with a fixed \( t = 3\)h and \( n = 1\), were very encouraging and alarming at the same time. Firstly, we investigated the rRNA signals (Fig 2A). Due to less input of sample RNA, only two rRNA types showed increasing signal at increasing UV dose. The pre-18S probe, detecting immature 18S transcripts, was increased \( \sim 25 \) times between dose 1.6 J and 12.8 J. The mRNA/rRNA ratio over the dose range was ambiguous due to the lack of replicates.

Several well-known p53 responsive genes showed a clear dose-response curve (Fig. 2B and S4), that is: for the lower doses. These responses are overall in line with the known UV responses. At the higher doses \( \sim 20 \) J is the highest dose– a clear plateau-effect is visible. A self-organizing map (SOM) analysis summarizes the scaled fold-change ratio’s of all DEGs of this experiment (Fig. 2C). The most obvious effect in the SOM is a clear nick in most clusters around 1.6 J in their dose-gene expression ratio curve. This nick marks the transition between two dose-response curves. Considering the p53-responsive genes, led us to hypothesize the existence of a low-dose range with a UV specific response and a high-dose range where a UV non-specific –presumably stress related– response prevails. So, for UV it means that a reliable dose to measure specific cellular responses is below \( \sim 1\) J UV-C radiation, up to twenty times lower than commonly used \( \sim 3-6, 27, 28\). The quite reproducible and also apparently rather specific dose-response curves hold the promise that a transcriptomics experiment using the right conditions could in fact lead to new biological insights.

In addition, we were surprised to discover what happened with the response of many DEGs, if there is a (much) higher dose applied. In most cases the response reached a plateau, became substantially weaker, or even completely reversed with increasing doses. In many cases, there seems little to no response to low-dose UV radiation, but only to the high-doses. In our opinion, these high-dose DEGs probably represent the non-specific stress genes, which are non-relevant to our study, or are caused by a severely altered mRNA/rRNA ratio, which renders the DEGs artificial. This in fact means that the majority of DEGs, which formed the orchestrated eigen-gene profiles in our original experiment, most likely represents a coordinated and massive stress response or consists of artificial DEGs, as a result of an extreme high UV dose.

This emphasizes the important of dose range finding, which obviously also applies to time, to determine the optimal experimental conditions to study the aimed-for cellular phenomena.

It could well be that the large sets of stereotypic-regulated or ”core environmental stress response” genes in yeast that after exposure to many different environmental stress types show perfect reciprocal transcriptome responses \cite{1, 2} similar to our observations, might in fact be by-and-large artificial DEGs due to non-corrected mRNA/rRNA ratio changes.
Serious complications in gene-expression studies

Concluding remarks

Our point of view from these experiments is that the high-stress UV perturbation we applied to MEFs resulted in thousands of stress-response and/or artificial DEGs when we employed established and wide-used bioinformatics data-analysis techniques. Equivalent UV perturbations in other studies resulted in similar findings.

The results of a limited dose range finding study, with a fixed t = 3h and n = 1, were very encouraging and alarming at the same time. Firstly, we investigated the rRNA signals (Fig 2A). Due to less input of sample RNA, only two rRNA types showed increasing signal at increasing UV dose. The pre-18S probe, detecting immature 18S transcripts, was increased ~ 25 times between dose 1.6 J and 12.8 J. The mRNA/rRNA ratio over the dose range was ambiguous due to the lack of replicates.

Several well-known p53 responsive genes showed a clear dose-response curve (Fig. 2B and S4), that is: for the lower doses. These responses are overall in line with the known UV responses. At the higher doses –20 J is the highest dose– a clear plateau-effect is visible. A self-organizing map (SOM) analysis summarizes the scaled fold-change ratio's of all DEGs of this experiment (Fig. 2C). The most obvious effect in the SOM is a clear nick in most clusters around 1.6 J in their dose-gene expression ratio curve. This nick marks the transition between two dose-response curves. Considering the p53-responsive genes, led us to hypothesize the existence of a low-dose range with a UV specific response and a high-dose range where a UV non-specific –presumably stress related– response prevails. So, for UV it means that a reliable dose to measure specific cellular responses is below ~1J UV-C radiation, up to twenty times lower than commonly used.

The quite reproducible and also apparently rather specific dose-response curves hold the promise that a transcriptomics experiment using the right conditions could in fact lead to new biological insights.

In addition, we were surprised to discover what happened with the response of many DEGs, if there is a (much) higher dose applied. In most cases the response reached a plateau, became substantially weaker, or even completely reversed with increasing doses. In many cases, there seems little to no response to low-dose UV radiation, but only to the high-doses. In our opinion, these high-dose DEGs probably represent the non-specific stress genes, which are non-relevant to our study, or are caused by a severely altered mRNA/rRNA ratio, which renders the DEGs artificial. This in fact means that the majority of DEGs, which formed the orchestrated eigen-gene profiles in our original experiment, most likely represents a coordinated and massive stress response or consists of artificial DEGs, as a result of an extreme high UV dose.

This emphasizes the important of dose range finding, which obviously also applies to time, to determine the optimal experimental conditions to study the aimed-for cellular phenomena.

It could well be that the large sets of stereotypic-regulated or “core environmental stress response” genes in yeast that after exposure to many different environmental stress types show perfect reciprocal transcriptome responses similar to our observations, might in fact be by-and-large artificial DEGs due to non-corrected mRNA/rRNA ratio changes.

Figure 2 – Dose range finding

Range finding experiment with increasing UV-dose exposures of WT MEFs. Experimental set-up: Dose range: 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 9.6, 12.8, 16, and 20] UV-C; One replicate per dose; Identical microarrays as in original experiment. Similar dual-color protocol as in with 1 ug sample RNA as input. In total, 2,461 DEGs were identified by using spline fitting.

(A) rRNA expression profiles of raw signal data over dose range. (B) Normalized expression value profiles of well-known p53 dependent genes that were found as DEG. (B) Self-organizing map clustering of normalized expression ratios (red/green signal plus t/t) of the DEGs found for this analysis.
cloud of non-relevant DEGs severely blocked our view of the molecular mechanisms we wanted to study. The settings for our perturbation originate from an era where primarily phenotypic endpoints were studied. The optimal perturbation settings for transcriptomics experimentation may be different. As transcriptomics techniques, be it microarray technology or next-generation sequencing-based, are more sensitive, dose and time range finding experiments could be performed to determine the optimal conditions for a perturbation with respect to the cellular mechanism one wants to study. As many perturbation experiments are performed without these range finding experiments, we are afraid that the effects we observed in our studies might apply to other studies as well. This means effectively that such stress-perturbation transcriptomics studies are by-and-large interpreting general stress/artificial responses as specific cellular response mechanisms to the applied stress, much like we have done in our original study.

An important indication for this situation is the changing mRNA/rRNA ratio. In studies of mRNA decay, the mRNA/rRNA ratio has been considered in the analysis, but this obviously also applies in experiments with massive mRNA or rRNA content changes. By measuring and quantifying this ratio, for instance using appropriate sample spike-ins, major deviations could be a warning for severe RNA changes in the cell. Again, it is important to keep in mind that most statistical processing and analysis methods have as explicit requirement that no severe RNA changes have occurred in the experimental samples. Alternatively, in contrast to the relative concentration, the (semi) quantitative mRNA gene concentration could be analyzed. This would require additional spike-ins for microarray technology, but could maybe easier be achieved in next-generation transcriptome sequencing. In any case, the use of appropriate controls is essential to not be deceived by non-specific responses.

Range finding studies are a good way to avoid the complications related to high-stress associated responses. Often omics-based range finding experiments are perceived to be too expensive, but with a replicate of n=1, the severely dropped microarray costs, and the possibility to reuse microarrays for such experiments, this is not the case anymore. One should accept the fact that different cellular mechanisms often require different dose and/or time ranges to study, as will be the case in our example: DNA-repair mechanisms ask for a low dose, whereas apoptotic mechanisms probably requires a higher dose. These experiments are currently ongoing in our lab in an attempt to further unravel the true UV-specific responses in WT and p53-mutant MEFs.

**Supplemental Information**

Supplemental material can be downloaded from: [http://www.tandfonline.com/doi/suppl/10.4161/trns.1.3.13487#tabModule](http://www.tandfonline.com/doi/suppl/10.4161/trns.1.3.13487#tabModule)
Serious complications in gene-expression studies

References


Chapter 3


34. Storey JD, Dai JY, Leek JT. The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. Biostat 2007; 8:414-32.