DISSERTATION

Statistical analysis of concentration-dependent high-dimensional gene expression data

Submitted to

the Department of Statistics

of the University of Dortmund

in Fulfillment of

the Requirements for the Degree of

Doktor der Naturwissenschaften

By

Marianna Grinberg

Dortmund, June 2017

Referees:

Prof. Dr. Jörg Rahnenführer

Prof. Dr. Guido Knapp

Prof. Dr. J. Hengstler

Date of Oral Examination: June 28, 2017

Contents

1 Introduction

Understanding the behavior of genes as a response to external influences, such as radiation or chemicals, on a fundamental level is one of the great challenges of modern biology. In specific, the investigation of chemically-induced toxicity is of major importance since it is crucial for the identification of biomarkers and the development of drugs. One approach to accomplish this objective utilizes toxicogenomics which is based upon the combination of toxicology and the analysis of genome-wide gene expression data. This research field uses the technology of microarrays which allows the simultaneous measurement of the expression of tens of thousands of genes. Nowadays, toxicogenomics has evolved to an established practice in the still emerging field of chemical hazard identification. It comprises the analysis of large-scale gene expression data in order to identify and characterize different modes of action associated with certain expression changes. Such deregulations, which occur as a response to chemical exposure, provide initial evidence of the involved toxic mechanisms. Based upon it, the key aspect is to detect those genes which improve the understanding of molecular mechanisms on a protein level. It is this particular understanding of the linkage between the entirety of all genes, transcriptome, proteome and eventually metabolome which qualifies for the assessment of biological processes within the human organism. Hence, especially the pharmaceutical sector applies the methods used within toxicogenomics for the research on drugs and, here, particularly the correct dosage is of vital importance.

Often, concentrations that cause gene alterations are associated with adverse effects. According to the saying ”The dose makes the poison” (Latin: ”sola dosis facit venenum”) which goes back to Paracelsus (founder of toxicology), who said ”All things are poison and nothing is without poison; only the dose makes a thing not a poison”, the dosis is decisive for the effect of a compound. The principle is based on the finding that all substances can cause toxic effects if consumed in high (excessive) quantities. For instance, a high salt consumption can lead to renal insufficiency. Still, sodium chloride is not considered as a toxic substance since it is commonly consumed in moderate amounts. In general, most chemicals, especially in forms of drugs, cause

only toxic reactions when overdosed extremely. To ensure the desired effect, the right dosage is essential. Because of this, dose finding and dose selection are ubiquitous topics in many fields, such as pharmacology, pharmacokinetics, toxicology or clinical research. Methods for modeling dose-response relationships are used to measure the effectiveness and toxicity of a compound. Often, dose-response studies are conducted to determine the lowest effective concentration at which first signs of cytotoxicity become detectable. In this context, Jiang (2013) has proposed to estimate the Absolute Lowest Effective Concentration (ALEC), which is the concentration at which a fixed and pre-specified effect level is reached exactly (point estimate), by fitting a log-logistic model to the data.

In the framework of this thesis, the model-based approach is applied to gene expression data to detect concentrations with critical changes in gene expression. Typically, only measured concentrations are considered as potential candidates for alert concentrations. Based on the assumption that the response dependency of the dose can be described by a sigmoidal function, a four-parameter log-logistic (4pLL) model is fitted to the data. Two alert concentrations referring to critical compound concentrations are estimated from the fitted average trend and compared with those of the classical naïve approach where for each measured concentration separately it is tested if the critical effect level is exceeded. The results are evaluated in a simulation study and in a real dose-response study.

Modeling gene expression data is only one topic of the thesis. Besides this, the work deals with two further issues that often arise in the context of gene expression analysis: The identification and characterization of genes associated with certain modes of action and the detection of biomarker candidates in the *in vitro* system for the prediction of toxicity *in vivo*. To better understand the key principles of transcriptome changes, a genome-wide gene expression analysis is performed. Special attention is drawn to statistical challenges arising from working with large data sets. Besides the curse of dimensionality (many more variables than observations) and the small number of replicates, the statistical analysis is faced with additional complexity including batch effects and implausible concentration progressions. To address this issue in a general manner, a pipeline involving several curation steps and a systematic strategy for the identification of consensus genes is proposed.

Thus, the main objective of this thesis is to gain a better understanding to whether a model-based approach yields more accurate results in terms of predicting critical concentrations than the classical one which is used in this work for the analysis of large-scale toxicogenomics data sets.

The structure of the thesis is as follows: In Chapter 2 the biological basics relevant to the understanding of the gene reactions investigated later in this work are presented. In the context of gene expression analysis, the Affymetrix GeneChip Technology and the RMA+ algorithm for pre-processing Affymetrix microarray data are described. In addition, a thorough description of the used data is given.

In Chapter 3 the statistical methods applied for data analysis within this work are described. This includes methods of descriptive analysis for large-scale gene expression data sets as well as methods for analyzing concentration-dependent expression progressions in the context of dose-finding studies. In the context of differential expression analysis the `Limma` t-test is outlined. Within a model-based approach for detecting critical expression changes, the (absolute) lowest effective concentration (A)LEC, derived from a dose-response model, is introduced. Methods for constructing confidence intervals for the (A)LEC and the effect level are presented. Moreover, the t_{4pLL}-test for the detection of critical expression changes in dose-response analysis is introduced. All methods are based on the application of the four-parameter log-logistic (4pLL) model.

The data is analyzed within the Chapters 4-6 using the aforementioned methods. In Chapter 4 and Chapter 5 the data of human and rat hepatocytes (*in vitro*) and rat liver cells (*in vivo*) is evaluated. Chapter 6 deals with simulated and real dose-response data. The thesis finishes with a comprehensive conclusion and an outlook on further research in Chapter 7.

2 Biological background

This chapter serves as an introduction to microarray analysis, beginning with the fundamentals of molecular genetics and ending with the generation of gene expression data. Section 2.1 gives a brief insight into the biological basics. Herein, the central dogma of molecular genetics according to which the genetic information is transferred from DNA to RNA to protein, is explained. High density oligonucleotide array technologies allow the simultaneous measurement of the expression of tens of thousands of genes. The Affymetrix GeneChip Technology is one of the most commonly applied methods for generating gene expression data. Section 2.2 describes the Affymetrix GeneChip array design and elucidates the principles of the photolithographic process for synthesizing DNA on microarray. The Affymetrix microarray data has to be pre-processed before it can be used for statistical analysis. There exist a number of pre-processing algoritms among which RMA is the most widely used pre-processing method for Affymetrix microarray data. Section 2.3 introduces an extended version of the RMA algorithm, the RMA+ algorithm, which is used for the normalization of the Affymetrix gene expression arrays analyzed within this work. The data sets used for analysis are introduced in Section 2.4.

2.1 Central dogma of molecular genetics

The deoxyribonucleic acid (DNA) is a nucleic acid that contains the genetic information which is essential for the development of all organisms. The DNA contains the genetic instructions that are necessary for the production of ribonucleic acid molecules whose essential function are the implementation of the information into proteins. The segments of the DNA that carry this information are called *genes*. The other sequences of the DNA, the non-coding segments, are either responsible for the regulation of the functional processes or just sequences with so far unknown functions. Within cells, DNA is organized in long structures called *chromosomes*. A chromosome is a single piece of coiled DNA containing many genes, regulatory elements and other nucleotide sequences. Humans have a diploid set of homologous chromosomes

consisting of 22 autosomes and one set of haploid chromosomes consisting of two gonosomes, one chromosome from each parent. The combination of all gene variants, i.e. the whole set of genes, is known as *genotype*.

The DNA consists of two strands which are spirally wrapped around one another forming the structure of a double helix. The so-called *nucleotides* form the molecular backbone of a DNA strand. They consist of a sugar molecule (desoxyribose), a phosphate group and one of four organic bases: Guanine (G), adenine (A), thymine (T) and cytidine (C). The two strands are connected via hydrogen bonds which are formed during the binding process of two complementary bases, A $\leftrightarrow$ T and C $\leftrightarrow$ G. The ribonucleic acid (RNA) differs from the deoxyribonucleic acid in its sugar molecule (ribose). The base thymine is replaced by uracil as complementary base to adenine. With the help of the RNA the genetic information is decoded and translated into proteins. The process in which the genotype is realized into its phenotype is known as *gene expression*. In Figure 2.1 the DNA and RNA strands are shown.

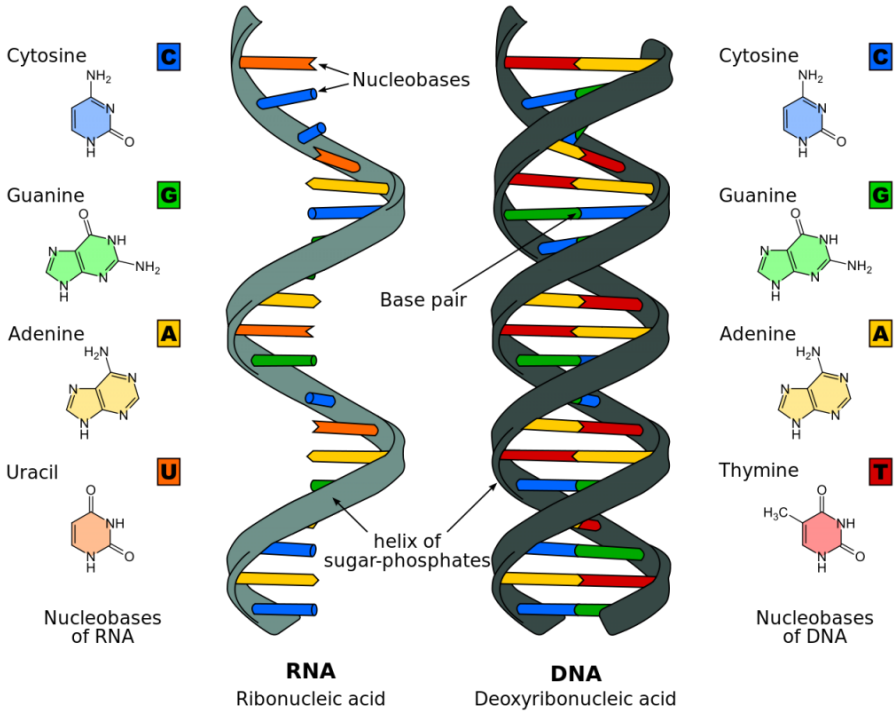


Figure 2.1: *Illustration of the DNA and RNA structure. The deoxyribonucleic acid (DNA, right panel) consists of two long strands which are coiled around one another in a double helix. The molecular backbone of each single strand is composed of deoxyribonucleotides. The nucleotides are differentiated by four bases: Adenine (A) $\leftrightarrow$ thymine (T) and cytosine (C) $\leftrightarrow$ guanine (G). The ribonucleic acid (RNA, left panel) is, in contrast, to the DNA single-stranded, contains ribose instead of deoxyribose and replaces the base thymine by uracil (BK101, 2017).*

The array of a base sequence defines the sequence of an amino acid which in turn defines the structure of a protein. Proteins are chains that are linked together by amino acids which differ in their length and array. They are synthesized according to their base order in the DNA. One of 20 possible amino acids is encoded by a base triplet (*codon*). The assignment of the base triplet to its respective amino acid is specified by the genetic code, see Figure 2.2. Some amino acids are encoded by more than one codon.

The protein biosynthese is one of the most important life processes in cells of living organisms. It consists of two subprocesses, the *transcription* and *translation* process (see Figure 2.3). During the transcription process the DNA sequence is transcribed into complementary mRNA. In contrast to the double-stranded DNA, the RNA is single-stranded. As the DNA consists of coding and non-coding sequences, the *exons* and *introns*, the transcribed non-coding sequences are excised from the pre-mRNA (preliminary messengerRNA) while the exons are retained. This process, known as *splicing*, plays a decisive role in regulating gene expression. However, due the accidental deletion of single exons during the splicing process, mRNA molecules of the same pre-mRNA may differ from one another. The variations resulting from the different composition of the exons in the mRNA might alter the protein structure. Thus, the splicing permits a wide variation of possible nucleotide combinations in the mRNA. These slightly modified proteins are called *isoformes*. Although isoformes are encoded by the same gene, they might execute different functions. The structural differences can either prevent the gene from functioning properly or just result in silent mutations, that means meaningless protein variants. Splicing is the reason why the human genome consists of much more proteins than genes. Moreover, allelic differences in mRNA splicing are often associated with genetic disease susceptibility.

Once the mRNA chain is generated, the synthesis of proteins can start. This process, known as translation, is the key process of the protein biosynthesis. The base sequence of the mRNA is translated sequentially into the corresponding amino acid sequence. Amino acids are bonded together by peptide bonds. Two amino acids join together to form dipeptides, more than ten amino acids form polypeptides, and more than 100 amino acids build proteins.

2.2 Affymetrix GeneChip Technology

Microarray technologies belong to the group of high throughput technologies which are used to generate expression data of tens of thousands of genes simultaneously. In the early 1990s, the US company Affymetrix developed the world's first commercial high-density chip for the analysis of

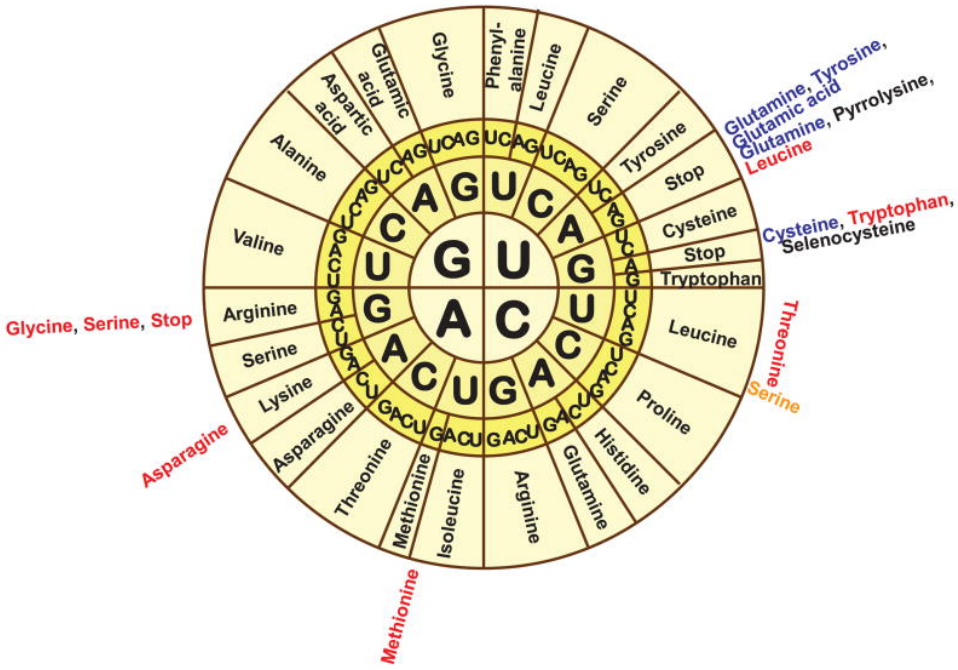


Figure 2.2: *The genetic code: The wheel is read from the inside out with each triplet coding for one particular amino acid. The sequence AUG encodes the start codon and the sequences UAA, UAG, and UGA encode the stop codons (Lobanov et al., 2010).*

gene expression data, the so-called GeneChip. Since then, microarrays are increasingly applied in the field of biomedical research. In practice, two kinds of microarrays are used, one based on cDNA (*complementary* DNA) and one based on oligonucleotides. They mainly differ in the way how the base sequences are synthesized on the chip. Affymetrix makes use of the latter method which synthesizes the single-stranded oligonucleotides on the chip base by base by a photolythographic procedure. Technologies using cDNA chips, in contrast, synthesize the complementary DNA sequence as a whole. The latter type of technologies are not part of this work and are therefore not discussed any further. Within this work, only Affymetrix microarray data is used for analysis. Thus, only their GeneChip technology is described in detail.

DNA microarrays allow to capture gene specific sequences in a cell. Conclusions on the phenotype can be drawn by means of sequential composition. Tens of thousands of such RNA transcripts can be captured and measured simultaneously using the Affymetrix's GeneChip. Depending on the analyzed organism, different GeneChips are used for transcription. The Human Genome U133 Plus 2.0 GeneChip, for example, is used for transcribing the human genome. The chip covers over 50 000 transcripts coding for more than 20 000 genes. Data from rat cells can be analyzed using the Rat Expression Set 230 Array GeneChip or its extended version 230 2.0, each comprising more than 15 000 and 30 000 transcripts and variants from

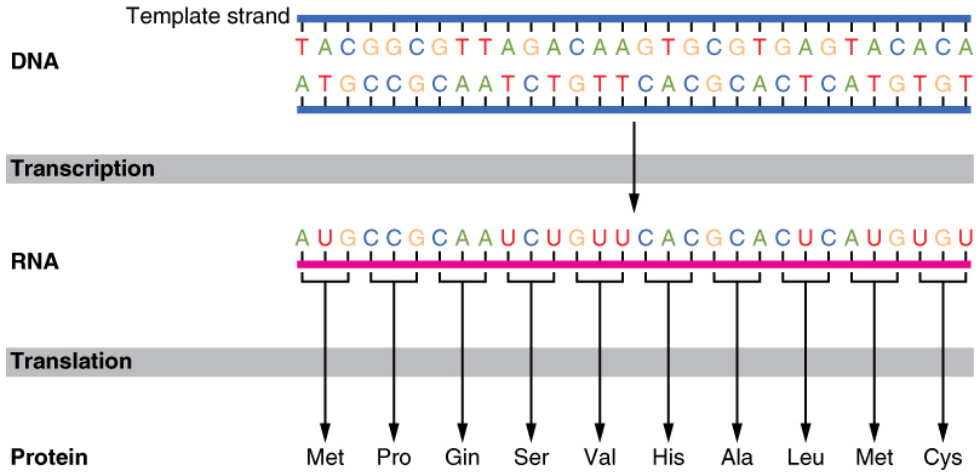


Figure 2.3: *From DNA to protein: The DNA sequence is transcribed into the corresponding RNA sequence by transcribing the base sequence of the gene into its complementary RNA nucleotide sequence. A base triplett encodes for a particular amino acid and the chain of amino acids defines the structure of a protein (oerpub/epubjs-demo book, 2017).*

over 10 000 and 13 000 genes, respectively. All those chips consist of hundreds of thousands of microscopically small *probe cells*, each containing millions of copies of a base sequence artificially synthesized of 25 nucleotides. This oligonucleotide sequence is complementary to the base sequence of the target mRNA. 11-20 of such oligonucleotide probes represent one specific transcript. To detect non-specific hybridizations, and to ensure high accuracy and reproducibility of the data, Affymetrix makes use of a paired design to match and mismatch transcripts (see Figure 2.4). The first probe is referred to as a perfect match (PM) probe and perfectly matches the target sequence, i.e. it is completely complementary to the target mRNA. Each PM is paired with a mismatch probe (MM) that is created by replacing the middle (13th) base by its complement. Thus, at this position it should come to no or a substantially weaker binding. The oligonucleotide probes referring to one probe set differ from one another in terms of their base sequences, such that 11-20 different exon regions of a gene are covered. To avoid spatial effects, the probe pairs of a particular probe set are spread all over the chip.

First of all mRNA is extracted from the tissue of interest. Then it is reverse transcribed into complementary DNA (cDNA). This newly created cDNA serves as template for the amplification of the mRNA. The resulting cRNA molecules are fragmented, labeled with a fluorescent dye and are fixed onto the array such that they can hybridize with their complementary probes on the array. The more the probes on the array coincide with the cRNA molecule, the higher the required temperature to disconnect the match. With a temperature increase non-specific hybridizations can be reduced. It is not uncommon that cRNA fragments hybridize with probes they are not

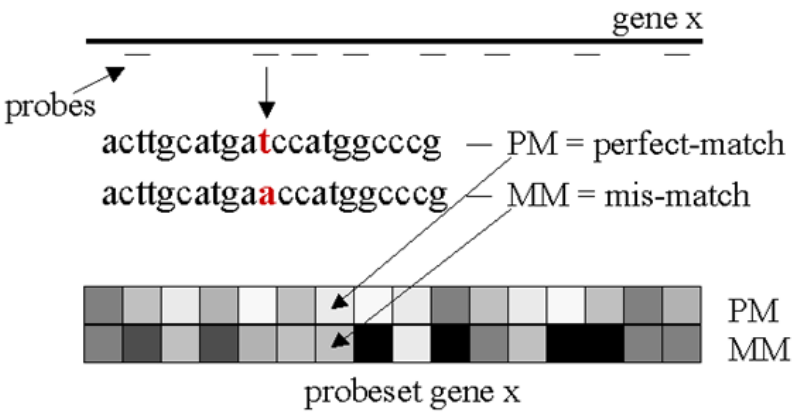


Figure 2.4: *GeneChip expression array design. Structure of a probe set with 16 probe pairs. One probe pair consists of a perfect-match (PM) and a mismatch (MM) of which each consists of 21 oligonucleotides. Both sequences are identical except for the 13th base which is replaced by the corresponding complementary base. The PM probe is completely complementary to the base sequence of the target mRNA, whereas the MM probes serve as control. A probe set is represented by multiple probe pairs (PBworks, 2016).*

intended to hybridize. The MM probes serves as controls for measuring background signals. In the case that only sub-sequences of the cRNA are complementary to the probes, less hydrogen bonds are formed during the binding process such that this kind of bindings can be released easier than the intended ones. As soon as the fluorescently stained cRNAs have interacted with their complementary oligonucleotides a light signal is provided. The unbound cRNA fragments are washed out and the hybridization pattern can be read off from the light distribution. A high definition laser scanner scans the intensity of the fluorescent signals which is used as a measure for the quantity of hybridized target RNA. The intensity values are combined into one *raw expression value* per probe set and stored as CEL files (see Figure 2.5 for procedure overview).

2.3 Data Preprocessing

As previously described, multiple probe pairs quantify one probe set which represents a gene. A gene in turn can be encoded by multiple probe sets. The step, in which the scanned data is reduced from probe level to gene level, is referred to as pre-processing. There exist a number of pre-processing algorithms which all summarize single probe set intensities to one representative expression value. Robust Multiarray Analysis (RMA) proposed by Irizarry et al. (2003a) is one of the most commonly used pre-processing algorithms for Affymetrix microarray data. In a three-stage process the data is background-corrected, normalized and finally summerized to

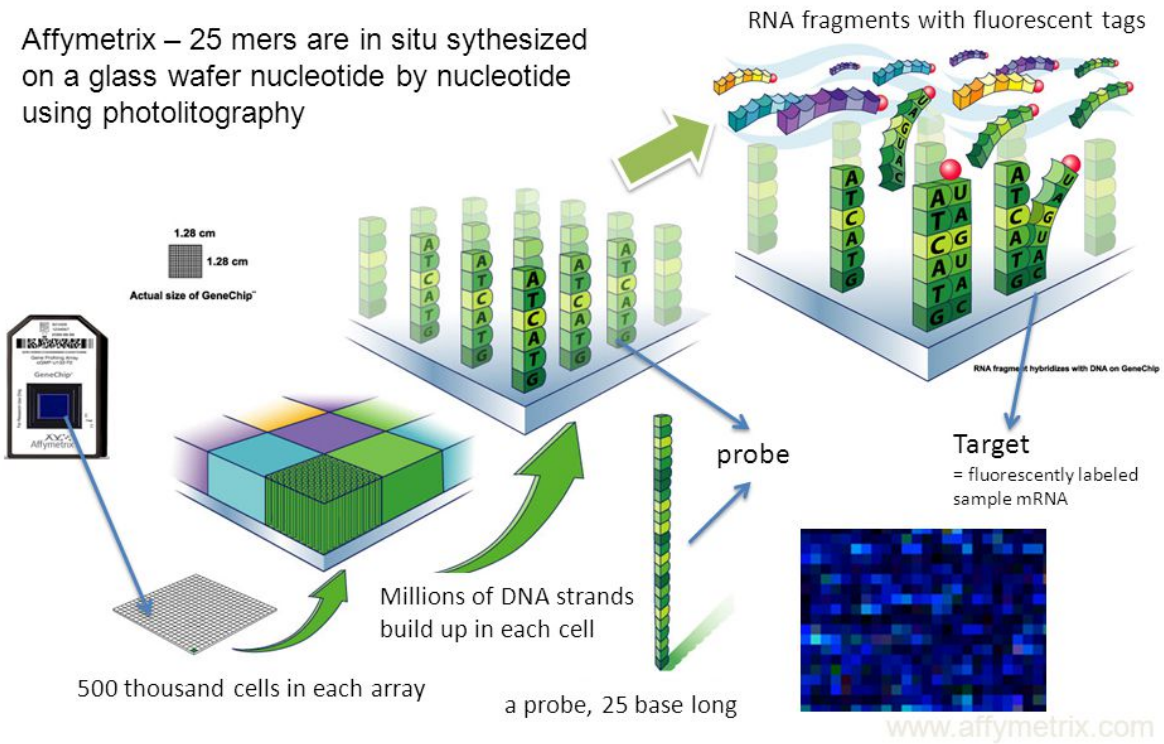


Figure 2.5: *Affymetrix microarrays: Photolithographic synthesis of oligonucleotides on microarrays. A chip consists of hundreds of thousands of microscopically small probe cells. Each cell contains millions of copies of oligonucleotide sequences which serve as template for the hybridization of the probes with their fluorescently labeled mRNA targets. The fluorescent signals are read by a high definition laser scanner and are combined into one raw expression value per probe set (Affymetrix, 2017).*

one value. However, the simultaneous analysis of data requires the simultaneous pre-processing of the respective microarrays. This interdependency of multiple microarrays has one obvious disadvantage: The inclusion of new microarrays implicate the re-pre-processing of the original data set which has to be pre-processed together with the new microarrays, and this process changes again the gene expressions of the original data. Thus, separate pre-processed data sets are not comparable. To ensure the comparability of results across different arrays wihtout changing the expression values of previously pre-processed microarrays Harbron et al. (2007) propose an extended version of the RMA algorithm, which they refer to as RMA+ algorithm. The idea is based on the calculation of reference parameters estimated from a reference set of microarrays which are stored and used for the normalization process of future microarrays. In this manner, the key properties of the RMA algorithm are maintained and new microarrays can be normalized in addition to the already pre-processed ones without re-estimating the reference parameters. This extension of the RMA algorithm allows the joint analysis of arrays analyzed in

different batches. The RMA+ algorithm is being implemented in the package `RefPlus` for the open source statistical software R (Chang et al., 2016).

The RMA+ algorithm makes use of the *Extrapolation Strategy* which splits the data into two sets of microarrays: One set is used as the reference set and the other one as the future set. The reference parameters are estimated from the reference set and are applied to the future set. They are obtained from fitting an RMA model to the reference set. This is accomplished by the aforementioned three-step procedure. Background correction is performed on each array individually and is therefore not discussed here. The reader is referred to Irizarry et al. (2003b) for a detailed explanation of the background correction procedure. After the intensity values have been background-corrected, a normalization step is required to achieve comparability across all arrays. As the slightest differences in the test execution, be it in the target preparation or in the hybridization procedure, might lead to a wide dispersion of the intensity values between arrays, RMA uses quantile normalization to normalize the probe intensities to a common set of quantiles, such that the intensities of all arrays have the same distribution. In the last step, the background-corrected and normalized probe set intensities are summarized to one expression value.

Let i denote the microarray and j the probe of a probe set, then the $\log_2$ background-corrected and normalized intensity N_{ij} of probe j on array i is given by:

$$N_{ij} = P_j + I_i + \epsilon_{ij}, \tag{2.1}$$

where P is the effect of the j^{th} probe, I is the expression of the probe set on array i and ϵ_{ij} indicates the error term. The expression value is estimated for each probe set separately by using Tukey's median polish which is an algorithm for calculating a robust average over all probes and arrays.

The probe set intensities of the reference set are stored together with their estimated quantiles and probe effects. The future microarrays undergo the same three-step procedure as the reference set: Background-correction, quantile normalization and aggregation to one expression value. But this time the microarrays are normalized to the reference quantiles. Assuming that the probe effects of the future set equal those of the reference set, the probe set intensity $\tilde{I}_f$ of a future array f is estimated from the model in (2.1) and given by

$$\tilde{I}_f = \text{median}_{j \in \text{Probes}}(N_{fj} - P_j),$$

where N_{fj} indicates the background-corrected and normalized intensity of probe j on array f and P_j represents the effect of the j^{th} reference probe.

Figure 2.6 compares the RMA algorithm with its extended version, the RMA+ algorithm.

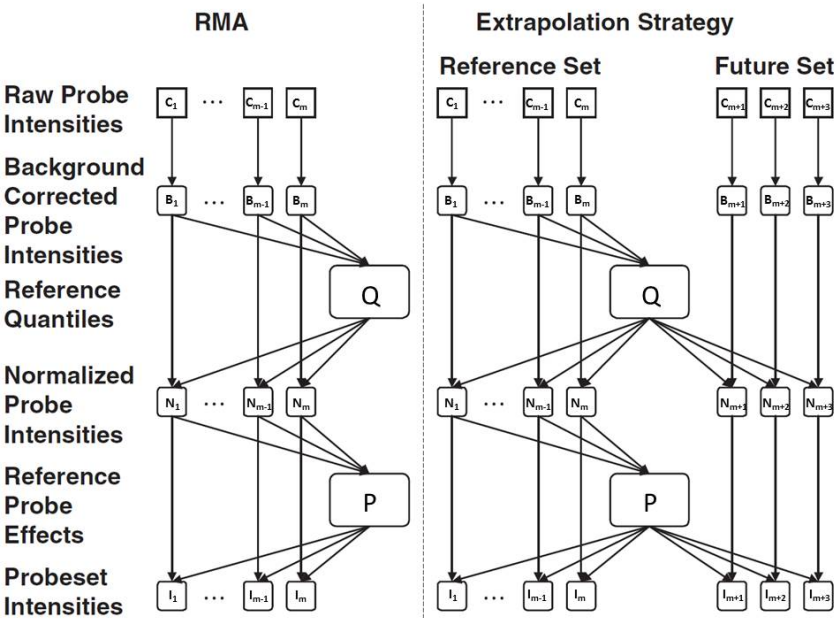


Figure 2.6: *Illustration of the RMA and RMA+ algorithm. Both algorithms uses background correction, quantile normalization and a linear model fit to the normalized data to obtain an intensity value for each probe set. In contrast to the RMA algorithm which uses the information of the complete microarray set the RMA+ algorithm makes use of the extrapolation strategy which splits the data into two sets of microarrays, the reference- and the future set. The reference set is used for the estimation of the reference parameters which are stored and used for the future set. The reference parameter are obtained from fitting a RMA model to the reference set. (Slightly modified version of a figure provided by Harbron et al. (2007)).*

2.4 Data sets

Within the scope of this work, several data sets were used for analysis. All analyses were performed on the basis of Affymetrix gene expression data. For the normalization of the arrays, the Robust Multi-Array Average (RMA+) algorithm was applied. As reference, different normalization parameters were used which depended on the used model organism. The respective parameters were obtained from fitting a RMA model to previously analyzed data of the same GeneChip. The estimated parameters were stored and applied to the currently analyzed arrays.

After normalization, the difference in gene expression between treated- and corresponding untreated samples was calculated for each test condition separately. The subtraction procedure was based on averaged replicate values. As gene expression data is measured on a $\log_2$-scale, the difference between logarithmized average values corresponds to the logarithmized fold change FC_i of gene i:

$$\begin{aligned}
\log\left(\mathrm{FC}_i\right) &= \log\left(\frac{\bar{x}_i^{Exp}}{\bar{x}_i^{\mathrm{Ctrl}}}\right) \\
&= \log\left(\sqrt{\prod_{j=1}^{n} x_{ij}^{Exp}}\right) - \log\left(\sqrt{\prod_{j=1}^{n} x_{ij}^{Ctrl}}\right) \\
&= \log\left[\left(x_{i1}^{Exp} \cdot \ldots \cdot x_{in}^{Exp}\right)^{\frac{1}{n}}\right] - \log\left[\left(x_{i1}^{Ctrl} \cdot \ldots \cdot x_{in}^{Ctrl}\right)^{\frac{1}{n}}\right] \\
&= \frac{1}{n}\left[\sum_{j=1}^{n} \log\left(x_{ij}^{Exp}\right)\right] - \frac{1}{n}\left[\sum_{j=1}^{n} \log\left(x_{ij}^{Ctrl}\right)\right],
\end{aligned}$$

where x_{ij}^{Exp} denotes the gene expression for gene i, $i = 1, \ldots, n_{\mathrm{PS}}$, and array j, $j = 1, \ldots, n$, and x_{ij}^{Ctrl} the corresponding control value. The terms $\bar{x}_i^{Exp}$ and $\bar{x}_i^{Ctrl}$ indicate the geometric mean of the exposed samples and the controls, respectively. The fold change is used as a measure for the exposure-related effect of a compound. All analyses base on the fold change values of a gene. However, the term *gene expression* is used in that context. Table 2.1 provides an overview of the data sets used for analysis.

2.4.1 TG-GATEs database

TGP (**T**he **T**oxicogenomics **P**roject) is a project funded by both the Japanese government and the private sector. The National Institute of Biomedical Innovation (NIBIO, 2017), the National

Institute of Health Sciences (NIHS, 2017) as well as the pharmaceutical industry contributed to its establishment. Between 2002 and 2006, gene array data was generated within the project testing $\approx$ 150 compounds, including hepatotoxic and non-hepatotoxic ones, in primary human and rat hepatocytes as well as rat liver and kidney cells *in vivo*. That data was used to generate a large-scale toxicogenomic database. The TG-GATEs (**T**oxico**G**enomics Project-**G**enomics **A**ssisted **T**oxicity **E**valuation **S**ystem) database was then finally created by integrating further options into the existing database system. The extended database offered possibilities of performing targeted analyses for the prediction of toxicity of the test compounds. TGP2 (**T**he **T**oxicogenomics Informatics **P**roject 2) was a follow-up project of TGP that was initiated in 2007 by the same founder. In the period from 2007 to 2011 30 safety biomarkes were detected within the framework of this project by using the TG-GATEs database. The new data gained by TGP2 was included into TG-GATEs. Open TG-GATEs is a publicly available database for the use of non-profit purposes. The raw microarray data (CEL files) for all analyzed compounds and conditions can be downloaded from the Open TG-GATEs website (http://toxico.nibiohn.go.jp/english/index.html). The portal has been developed to give scientists the opportunity to use the research results of TGP and TGP2. The user is free to use the data for both scientific and private purposes, including the publication of results and the disclosure of the information to third parties. The database compiles Affymetrix HG U133 Plus 2.0 gene expression microarray data on 170 compounds. The search for data is enabled via compound name or pathological findings by organ. Access to phenotype data is provided as well. The documents contain information about the experimental setup, the histopathological findings and the research results of the TGP project which are supplied as PDF file and can be viewed directly or downloaded from the homepage. Currently, the documents are only available in Japanese (NIBIOHN, 2017).

Table 2.1: *Overview of the data sets used in the analyses.*

Database	Data	Test system
TG-GATEs	Primary human hepatocytes	*in vitro*
	Primary rat hepatocytes	*in vitro*
	Rat liver hepatocytes	*in vivo*
NRW	Primary rat hepatocytes	*in vitro*
	Rat liver hepatocytes	*in vivo*
UKN1	Human embryonic stem cells	*in vitro*

Primary human hepatocytes

The primary human hepatocytes were treated with different test compounds using three concentrations (Low, Middle, High) and three incubation periods (2h, 8h, and 24h). For cytotoxic compounds the highest tested concentration was chosen such that it represented approximately the EC_{10} (the concentration that produces 10% reduction of the maximal effect). Each concentration was assessed using two replicate experiments. Table 2.2 provides an overview of the experimental design. A subset of compounds was tested under all conditions (n=52), while most of the compounds were tested only under some of the conditions. The compounds were tested either for only one or two exposure periods, or with only two concentrations, as shown in Table B.1 in the Appendix. Experiments without replication, as well as experiments including cytokines and LPS (lipopolysaccharide), were excluded from the analyses. Cytokines are proteins involved in the regulation of proliferation and differentiation processes in cells. Seven of the tested compounds were cytokines and due to their molecular functionality excluded. Table 2.2 shows the number of compounds tested under the indicated condition with and without cytokines in brackets.

Table 2.2: *Matrix of the compounds tested in primary human hepatocytes. The table provides the numbers of compounds tested under the indicated condition, for each combination of concentration and exposure period, before and after (in brackets) excluding cytokines and LPS (lipopolysaccharide) from the analyses.*

	2h	8h	24h	Overlap
Low	53 (48)	82 (75)	81 (75)	52 (48)
Middle	53 (48)	153 (146)	157 (151)	52 (48)
High	53 (48)	153 (146)	153 (148)	52 (48)
Overlap	53 (48)	82 (75)	77 (72)	52 (48)

Primary rat hepatocytes

The cultured rat hepatocytes were tested in duplicates with the indicated compounds using a low, middle and high concentration for the incubation periods 2h, 8h, 24h. For a detailed overview of the data the reader is referred to Table B.2 in the Appendix. The highest tested concentration was again chosen close to cytotoxic levels. Table 2.3 shows the number of compounds tested at the indicated concentration and time sets. The same exclusion criteria as those for the human hepatocytes were applied to the data from the rat model, for both the *in vitro* and *in vivo* test system.

Table 2.3: *Matrix of the compounds tested in primary rat hepatocytes. The table provides the numbers of compounds tested under the indicated conditions for each combination of concentration and exposure period, before and after (in brackets) excluding cytokines and LPS (lipopolysaccharide) from the analyses.*

	2h	8h	24h	Overlap
Low	140 (138)	140 (138)	145 (143)	140 (138)
Middle	140 (138)	140 (138)	140 (138)	140 (138)
High	138 (137)	139 (138)	138 (137)	138 (137)
Overlap	138 (138)	139 (137)	138 (138)	52 (48)

Rat liver hepatocytes

Rat liver samples were treated with the compounds listed in Tables B.3 and B.4 which are given in the Appendix. Each compound was tested at three concentrations (Low, Middle, High) and sacrificed at different time periods after exposure. This time eight incubation times (3h, 6h, 9h, 24h, 4 days, 8 days, 15 days, 29 days) were investigated, using three replicates in each experiment. Table 2.4 summarizes the compounds with available data with respect to the indicated test condition.

Table 2.4: *Matrix of the compounds tested in rat liver cells. The tables provide the numbers of compounds tested under the indicated conditions for each combination of concentration and exposure period, before and after (in brackets) excluding cytokines and LPS (lipopolysaccharide) from the analyses.*

	3h	6h	9h	24h	Overlap
Low	153 (151)	153 (151)	153 (151)	157 (155)	153 (151)
Middle	152 (150)	153 (151)	153 (151)	157 (155)	152 (150)
High	151 (149)	151 (149)	151 (149)	153 (151)	150 (148)
Overlap	150 (148)	150 (148)	150 (148)	152 (150)	149 (147)

	4 days	8 days	15 days	29 days	Overlap
Low	141 (141)	141 (141)	141 (141)	141 (141)	141 (141)
Middle	141 (141)	141 (141)	141 (141)	141 (141)	141 (141)
High	143 (143)	143 (143)	139 (139)	127 (127)	127 (127)
Overlap	141 (141)	141 (141)	138 (138)	126 (126)	126 (126)

The reader is referred to the Tables B.1-B.4 in the Appendix which provide a detailed compound-specific summary for the three model organisms. The tables give full and abbreviated compound names as well as the concentration in µM (µg/mL, µg/kg) and the number of

independent replicates of gene array data available after incubation with a low, middle and high concentration for the indicated exposure period.

2.4.2 UKN1 test system

Embryonic stem cell (ESC)-based systems have been developed to recapitulate *in vitro* the differentiation of stem cells into neuronal cells. Stem cells have the property of pluripotency, i.e. the ability to differentiate into all types of cells. During the differentiation process different mechanisms such as cell proliferation, migration and apoptosis are induced. Cultures of differentiating human embryonic stem cells (hESC) offer the opportunity to observe, study and control the early steps of human development. External stimulus influences, such as drug exposure, can interfere with early developmental stages. In that context, different human ESC (hESC)-based *in vitro* systems have been developed to recapitulate the different phases of early tissue specification and neural development. The UKN1 test system is one of them and was developed to model the stage of differentiation of neuroepithelial precursor cells (NEP) from hESC. Figure 2.7 visualizes the test system's treatment protocol. In that system the cells were exposed to the test compounds within 6 days. In the present study two compounds, valproic acid (VPA) and methylmercury (MeHg), were tested to detect chemically-induced gene expression alterations. VPA is used as an anti-epileptic drug and known to cause neural tube defects, just as MeHg. Earlier analyses have shown that exposure-related effects strongly depend on the concentration of the test compound. Therefore, the compounds were tested with different concentrations covering non-toxic to toxic concentrations. The highest concentration was chosen according to a benchmark concentration representing the EC_{10}. For more details on the test systems the reader is referred to Krug et al. (2013).

VPA chronic concentration study

The VPA concentration study was conducted to investigate the development of human embryonic stem cells (hESC) to neuroectoderm. The cells were treated *in vitro* with valproic acid (VPA) using eight different concentrations (25-1000 µM). Each concentration was assessed using three replicate experiments (see Table 2.5). The compound was exposed to the cells during the entire differentiation process. In addition, six untreated measurements were available. Replicates of controls were averaged before subtracting from corresponding exposed samples (paired design). The study was carried out within the framework of the European Commission-funded research

consortium (ESNATS) which targets the prediction of toxicity of drug candidates for the use of embryonic stem cell-based novel alternative tests.

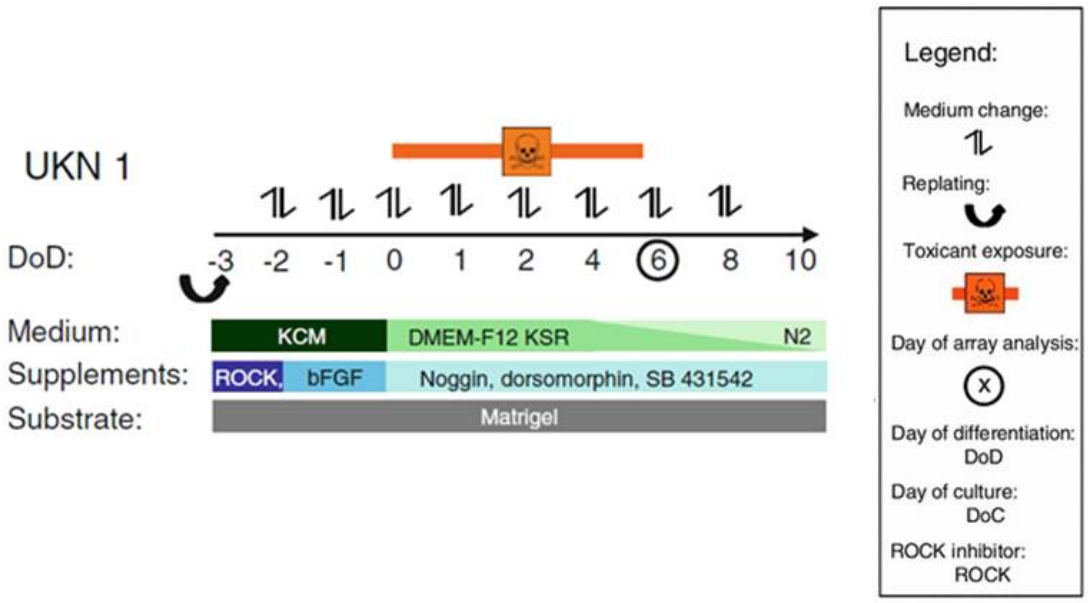


Figure 2.7: *Overview of the UKN1 test system's treatment protocol. The test system recapitulates the differentiation process of human embryonic stem cells (hESC) to neuroepithelial precursor cells (NEP). The cells were treated with valproic acid (VPA). The bars below the test system provide information on replating, medium change, toxicant exposure and day of array analysis, as indicated in the legend to the right (Krug et al., 2013).*

Table 2.5: *Overview of the number of replicates used in the VPA chronic concentration study.*

Concentration in µM								
0	25	150	350	450	550	650	800	1000
6	3	3	3	3	3	3	3	3

2.4.3 NRW database

The NRW data set comprises 30 compounds that have been tested in rats *in vivo* (Ellinger-Ziegelbauer et al., 2008) and 29 compounds that have been tested in cultivated rat hepatocytes (Schug, 2011). With the exception of one compound (Phenobarbital) that was only exposed to rat cells *in vivo*, the other 29 test compounds were the same. Male Wistar rats were used as model organism for both test systems.

In vivo rat cell culture

Primary rat hepatocytes with *in vivo* rat liver data were treated with each compound in 3 replicates and sacrificed after 6h, 12h, 24h, 48h, 3 days, 7 days and 14 days after exposure. Incubation periods of 6h, 12h and 48h were not considered any further as only one compound (Acetaminophen) was tested for these time periods. Table B.5 in the Appendix contains an overview of the number of replicates used in the *in vivo* experiments. The rat liver cells were obtained from five animals treated daily with each compound. The treatment of animals was performed in different experimental series. Therefore, 'experimental series'- and 'exposure period'-matched controls were subtracted from the corresponding treated samples. The concentrations used for the individual compounds during the entire incubation period are indicated in Table B.5 as well. For transcriptional analysis the *rae230a* array was used which comprises 15 923 probe sets, corresponding to approximately 10 045 annotated genes. As no reference parameters are provided for this chip, the arrays of the treated samples were normalized to the complete set of control arrays. The study was originally conducted to predict the toxicity class of unknown compounds. A classifier that separated genotoxic from non-genotoxic carcinogens was built from a set of training compounds (n=13) and applied to an independent set of validation compounds (n=16). The training and validation sets together form the database for the *in vivo* analyses. The classification into the annotated categories is given in the Appendix in Table B.5.

In vitro rat cell culture

Cultured rat hepatocytes were also tested with three replicates using three concentrations (Low, Middle and High) and one incubation period (24h). The highest concentration represents the EC_{20}. Similar to the *in vivo* experiments, not all concentration- and time sets are complete. Some of the compounds were tested only for two conditions as shown in Table B.6 in the Appendix. The experiments were organised in 6-well-dishes. Per concentration, 3-wells were incubated with the test compound and 3-wells were used as controls. According to that test design, the controls were 6-well dish- matched subtracted from the corresponding exposed samples. Transcriptional analysis was performed by using the *rat2302*-GeneChip 31 099 probe sets encoding 13 685 genes. For comparability reasons, the analyses was restricted to those transcripts that have been measured on the *rae230a*-GeneChip which was used for the *in vivo* experiments. For the normalization of the entire set of expression arrays the RMA+ algorithm was used which provided reference parameters for future data sets, such as the TG-GATEs rat data sets.

3 Statistical methods

This chapter provides an overview of the statistical methods used for data analysis in this thesis. First, multivariate methods for pattern recognition are introduced. The basic principle of principal component analysis is explained and the heatmap is introduced as a visualization method for high-dimensional data. In Section 3.3 the `Limma` t-test, which is used for the analysis of high-dimensional gene expression data, is outlined. In the context of dose-finding studies, several indices for the description of concentration-dependent progressions are presented (Section 3.4). Section 3.5 deals with the theory of dose-response models. Within this section, the four-parameter-log-logistic model (4pLL) together with its estimate for the Absolute Lowest Effective Concentration (ALEC) is presented. Moreover, it is elucidated how to construct confidence intervals for the effect level (response) and to how calculate the lowest effective concentration (LEC) by means of hypothesis testing.

3.1 Principal component analysis

Principal component analysis (PCA) is a multivariate procedure for the detection of structures in large data sets. The method targets to reduce the dimensionality of data by projecting the data into a lower-dimensional space while aiming for preservation of information. The approach implies the construction of uncorrelated linear combinations representing the principal components which are sorted according to the proportion of their explained variance in descending order. Thereby the total variance serves as a measure for the information content. Let $\boldsymbol{X}^\top = [X_1, \ldots, X_p]$ be a p-dimensional vector of an $n \times p$ data matrix with n observations and p variables. The idea is to transform the coordinate system that is spanned by the random variables $X_1, \ldots, X_p$ by rotation into a new coordinate system, the vector subspace $\mathbb{R}^k$ ($k < p$). The objective is to find a set of linearly uncorrelated components that are orthogonal to each other and sorted in order of their magnitude, i.e. such that the first principal component explains most of the data variability, the second component contains the next most information and the last components provide the

slightest information. If a sufficiently large proportion of total variance is covered by the first k principal components, they prove to be sufficient to reproduce the data variability with little loss of information. The remaining $p-k$ components make only an insignificant contribution to the overall variance and can therefore be neglected. The coordinate system spanned by the k principal components allows a simplified depiction of the data structure or their covariance matrix, respectively, in a k-dimensional space.

Consider now the vector $\boldsymbol{X}^\top = [X_1, \ldots, X_p]$ with covariance matrix $\boldsymbol{\Sigma}$ and eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p > 0$. Further, let $A = (a_1, \ldots, a_n)$ be an orthogonal $p \times p$ matrix, i.e. $\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{I_p}$. Then the principal components $Y_1, \ldots, Y_p$ are obtained by the transformation of $\boldsymbol{X}^\top \rightarrow \boldsymbol{Y}^\top$ given by

$$\begin{aligned} Y_1 &= \boldsymbol{a_1}^\top \boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ Y_2 &= \boldsymbol{a_2}^\top \boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\ &\vdots \qquad\qquad \vdots \\ Y_p &= \boldsymbol{a_p}^\top \boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p, \end{aligned} \tag{3.1}$$

where $\boldsymbol{a_1}^\top = (a_{11}, \ldots, a_{1p}), \ldots, \boldsymbol{a_p}^\top = (a_{p1}, \ldots, a_{pp})$ are the vectors of weights. Since $\boldsymbol{A}$ is orthogonal the transformation corresponds to a rotation of the n points in the p-dimensional space.

The first principal component minimizes the sum of the Euclidean distances between the projected data points and the original ones and maximizes the variance of the projections. Hence, the variance of $Y_1 = \boldsymbol{a_1}^\top \boldsymbol{X}$ must be maximized subject to $\boldsymbol{a_1}^\top \boldsymbol{a_1} = 1$. A method for the optimization of a function subject to a constraint is the method of Lagrange multipliers. The function L to be maximized is given by

$$L(\boldsymbol{a_1}, \lambda_1) = \underbrace{\boldsymbol{a_1}^\top \boldsymbol{\Sigma}\, \boldsymbol{a_1}}_{\text{fct. to max.}} - \lambda_1 \underbrace{(\boldsymbol{a_1}^\top \boldsymbol{a_1} - 1)}_{\text{constraint}}, \tag{3.2}$$

where $\lambda_1 \in \mathbb{R}$ denotes a Lagrange multiplier. The function in (3.2) is differentiated with respect to $\boldsymbol{a_1}$ and λ_1 and subsequently set equal to zero:

$$\left.\begin{aligned} 1)\, &\tfrac{\partial L}{\partial \lambda_1} = 1 - \boldsymbol{a_1}^\top \boldsymbol{a_1} \\ 2)\, &\tfrac{\partial L}{\partial \boldsymbol{a_1}} = 2\,\boldsymbol{\Sigma}\, \boldsymbol{a_1} - 2\lambda_1\, \boldsymbol{a_1} \end{aligned}\right\} \stackrel{!}{=} 0. \tag{3.3}$$

The system of equations in (3.3) shows that the vector $\boldsymbol{a_1}$ has to satisfy $\boldsymbol{a_1}^\top \boldsymbol{a_1} = \mathbf{1}$ and $\boldsymbol{\Sigma}\, \boldsymbol{a_1} = \lambda_1\, \boldsymbol{a_1}$. The second term equals an eigenequation with the Langrange multiplier as eigenvalue meaning that $\boldsymbol{a_1}$ is the normalized eigenvalue of the covariance matrix $\boldsymbol{\Sigma}$ with eigenvalue λ_1. Hereafter, let $\boldsymbol{e_i}$ denote the eigenvector to eigenvalue i. The first principal component is obtained by the projection of $\boldsymbol{X}^\top$ onto $\boldsymbol{e_1}$ which is the eigenvector corresponding to the largest eigenvalue λ_1. The remaining principal components are constructed recursively to the preceding ones such that the projections onto the k^{th} principal component have maximal variance and are orthogonal to the first $k-1$ principal components, i.e. $\boldsymbol{a_i}^\top \boldsymbol{a_k} = 0$, $1 \leq i < k$. Thus, the second principal component has to maximize

$$L(\boldsymbol{a_1}, \boldsymbol{a_2}, \mu, \lambda_2) = \underbrace{\boldsymbol{a_2}^\top \boldsymbol{\Sigma} \boldsymbol{a_2}}_{\text{fct. to max.}} - \mu \underbrace{\boldsymbol{a_2}^\top \boldsymbol{a_1}}_{\text{constraint}} - \lambda_2 \underbrace{(\boldsymbol{a_2}^\top \boldsymbol{a_2} - 1)}_{\text{constraint}}, \tag{3.4}$$

where $\mu \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$ indicate the multipliers. Differentiating function (3.4) with respect to all parameters and setting the partial derivatives equal to zero, delivers the second most important linear combination $\boldsymbol{e_2}^\top \boldsymbol{X}$ which maximizes $\mathrm{Var}\left(\boldsymbol{e_2}^\top \boldsymbol{X}\right)$ subject to the required constraints. The remaining components are constructed analogously, i.e. the i^{th} linear combination maximizes $\mathrm{Var}\left(\boldsymbol{e_i}^\top \boldsymbol{X}\right)$ with respect to both constraints, $\boldsymbol{e_i}^\top \boldsymbol{e_i} = \mathbf{1}$ and $\mathrm{Cov}\left(\boldsymbol{e_i}^\top \boldsymbol{X}, \boldsymbol{e_k}^\top \boldsymbol{X}\right) = 0$ for $k < i$. By induction, it can be shown that the first k principal components correspond to the first k eigenvalues.

Replacing the coefficients $\boldsymbol{a_i}$, $i = 1, \ldots, p$ from equation (3.1) with the normalized and orthogonal eigenvectors $\boldsymbol{e_i}$ leads to the best i-dimensional approximations

$$Y_i = \boldsymbol{e_i}^\top \boldsymbol{X} = e_{i1} X_1 + e_{i2} X_2 + \ldots + e_{ip} X_p, \qquad i = 1, \ldots, p$$

with the following properties

i) $\mathrm{Var}\,(Y_i) = \boldsymbol{e_i}^\top \boldsymbol{\Sigma}\, \boldsymbol{e_i} = \lambda_i \boldsymbol{e_i}^\top \boldsymbol{e_i} = \lambda_i, \qquad i = 1, \ldots, p$

ii) $\mathrm{Cov}\,(Y_i, Y_k) = \boldsymbol{e_i}^\top \boldsymbol{\Sigma}\, \boldsymbol{e_k} = 0, \qquad i, k = 1, \ldots, p,$

subject to the constraint $\boldsymbol{e_i}^\top \boldsymbol{e_i} = 1$.

The covariance matrix $\boldsymbol{\Sigma}$ can be rewritten as $\boldsymbol{\Sigma} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^\top$ by means of Spectral Decomposition (SD), where $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal entries are the eigenvalues $\lambda_1, \ldots, \lambda_p$ and $\boldsymbol{U} = [\boldsymbol{e_1}, \ldots, \boldsymbol{e_p}]$ is a orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{\Sigma}$. For a

detailed description of the theory of SD, the reader is referred to Johnson and Wichern (1998). The matrix $\boldsymbol{U}$ has the property: $\boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$.

The vectors $\boldsymbol{e_1}, \ldots, \boldsymbol{e_p}$ provide an orthogonal basis, i.e. $\boldsymbol{U}$ corresponds to the vector subspace into which $\boldsymbol{X}^\top$ is projected: $\boldsymbol{Y}^\top = \boldsymbol{X}^\top\boldsymbol{U}$. As a rotation is obtained by the multiplication of a scalar with a orthogonal matrix, the transformation $\boldsymbol{X}^\top \rightarrow \boldsymbol{Y}^\top$ corresponds to the rotated coordinate system.

The vector $\boldsymbol{e_i}^\top = (e_{i1}, \ldots, e_{ik}, \ldots, e_{ip})$ can be interpreted as the score vector of the principal component whose components e_{ik}, $i, k = 1, \ldots, p$, indicate how good the k^{th} variable is approximated by the i^{th} principal component. The first k principal components explain a proportion of

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

of the total variance.

The equation in (3.6) for the correlation of Y_i und X_k can be justified by choosing $\boldsymbol{a_k}^\top = \underbrace{[0, \ldots, 0, 1, 0, \ldots, 0]}_{k^{\text{th}}\text{position}}$ such that

$$X_k = \boldsymbol{a_k}^\top \boldsymbol{X}, \qquad \text{Cov}\,(X_k, Y_i) = \text{Cov}\left(\boldsymbol{a_k}^\top \boldsymbol{X}, \boldsymbol{e_i}^\top \boldsymbol{X}\right) = \boldsymbol{a_k}^\top \boldsymbol{\Sigma}\, \boldsymbol{e_i} = \boldsymbol{a_k}^\top \lambda_i\, \boldsymbol{e_i} = \lambda_i\, e_{ik}. \quad (3.5)$$

The variance of Y_i and X_k satisfy: $\text{Var}\,(Y_i) = \lambda_i$ and $\text{Var}\,(X_k) = \sigma_k$, respectively. It follows

$$\rho_{Y_i, X_k} = \frac{\text{Cov}\,(Y_i, X_k)}{\sqrt{\text{Var}\,(Y_i)}\sqrt{\text{Var}\,(X_k)}} = \frac{\lambda_i e_{ik}}{\sqrt{\lambda_i}\sqrt{\sigma_k}} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_k}}, \qquad i, k = 1, \ldots, p. \quad (3.6)$$

If the variables are not measured in the same scale unit, it is recommended to perform the PCA on the basis of the correlation matrix rather than the covariance matrix, since the variables are not comparable with each other. With respect to microarray analysis, standardization of the data is not required as the variables representing the expression values are measured on the same scale ($\log_2$). Usually, PCA is used as a prestep in a comprehensive analysis to obtain a first overview of the data. Often, the first two components are sufficient to reveal features such as batches or outliers.

3.2 Heatmap

Heatmaps are a commonly used graphical method for the visualization of gene expression data where rows and columns represent the genes of interest with respect to the analyzed arrays. A heatmap compresses large amounts of information into a compact display area and, hence, allows the visual detection of coherent patterns. A matrix containing the expression values of genes is color encoded according to the values' order of magnitude and displayed as color image. Usually, heat colors are used for illustrating the data which is the reason for the name of the heatmap. Many software systems such as the statistical software `R` use the heat colors as default. Generally, different color schemes can be used to visualize the colormap. In the context of gene expression data, it is useful to map the range of values to colors ranging from blue to red with blue colors indicating low expression levels and red colors high expression values.

Coherent patterns of color are generated by hierarchical clustering. The rows and columns of the data matrix are permuted such that objects with similar expression profiles are clustered together. Cluster relationships are indicated by dendrograms generated for both axes. The resulting patterns indicate functional relationships between the arrays and genes (Wilkinson and Friendly, 2009). Heatmaps can be produced by using the `R` standard package `stats` (R Core Team, 2015). It is up to the user to decide which agglomeration rule and metric should be used for the cluster analysis. The *complete linkage method* is used by default to reorder the dendrograms. Alternatively, both dendrograms can be reordered by a prescribed vector of values or be completely omitted. By default, the Euclidean distance is used as distance measure for the calculation of pairwise distances. The `gplot` package implements an extended version of the standard `R` function which offers a number of additional features such as a *color key* illustrating the range of values together with their distribution, or a side bar that may be used to classify the objects with respect to any characteristics (Warnes et al., 2015).

Figure 3.1 shows an example for a heatmap, where rows represent genes and columns indicate samples. For illustration purposes, gene expression data of human embryonic stem cells (hESCs) after VPA (valproic acid) exposure was used. Originally, the cells were treated with eight different concentrations (25-1000 µM with n=30) using three replicates (for more details on the study see Section 2.4). For generating the exemplary heatmap only a subset of the measured concentrations was used (150 µM, 550 µM, 800 µM, 1000 µM with n=12 samples). The absolute gene expression levels ($\log_2$-scaling) of the ten transcripts with highest variance across the

12 samples were color-coded for display. The samples have been classified with respect to concentration level and replicate number. Up- and downregulation is indicated by blue and red coloring.

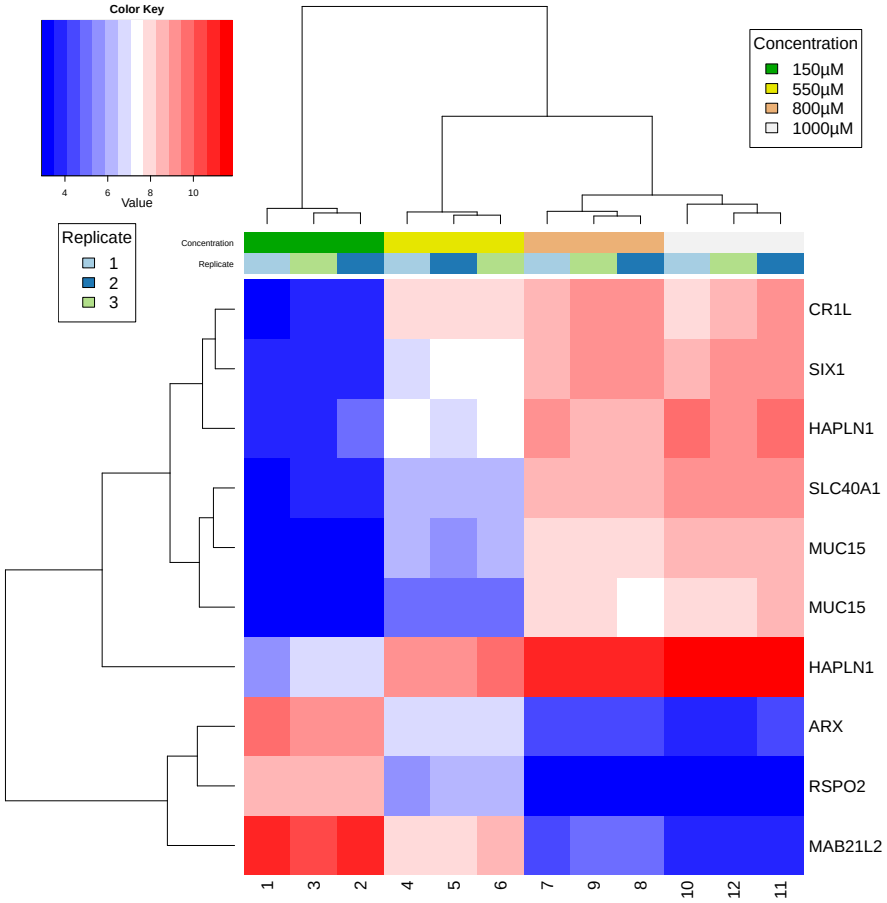


Figure 3.1: *Example of a heatmap. Each row represents a gene, while each column stands for a sample. Red color indicates up- and blue color downregulated genes as indicated by the color key in the upper left. The rows and columns are reordered according to their respective dendrograms which are generated by hierarchical clustering. The samples have been classified according to concentration level and replicate number which are indicated by the column side bar and the legends in the upper right and upper left.*

3.3 Limma: Linear Models for Microarray Data

`Limma` is an `R/Bioconducter` package which is used for the analysis of gene expression data (Ritchie et al., 2015). The `Limma` t-test is a moderated t-test for detecting gene expression changes which are associated with a particular treatment condition. The simple t-test is altered in the sense that the information of the complete set of genes is used in the estimation of the gene-wise variances in place of the ordinary variances. The basic idea of `Limma` is to shrink the gene-wise residual sample variances towards a common value by an empirical Bayes approach. The effect of sharing information has, especially in case of small sample sizes, the advantage of more accurate estimators and therefore less unbiased test results.

Consider a microarray experiment with n arrays where $\boldsymbol{y_g}^\top = (y_{g1}, \ldots, y_{gn})$ denotes the response vector which contains the expression values of gene g. This means that the response of gene g corresponds to the g^{th} row of the expression matrix. Let $\boldsymbol{X}$ be the design matrix with rows representing the arrays and $\boldsymbol{\alpha_g}$ the vector of coefficients. Let $\boldsymbol{C}$ denote the contrast matrix whose rows correspond to the coefficients and whose columns contain the contrasts. In case that only one contrast is of interest, a contrast vector is defined. The contrast matrix allows to adjust for covariates, batch- or interaction effects. In terms of a linear model

$$\mathrm{E}(\boldsymbol{y_g}) = \boldsymbol{X}\boldsymbol{\alpha_g}.$$

The variance of the response vector $\boldsymbol{y_g}$ is given by

$$\mathrm{var}(\boldsymbol{y_g}) = \boldsymbol{W}_g\sigma_g^2,$$

where $\boldsymbol{W_g}$ is a positive definite weight matrix which allows the incorporation of unequal variances for more accurate test results, and σ_g^2 indicates for gene g the residual variance of the model with d_g residual degrees of freedom. Samples can be individually weighted according to their reliability (Smyth, 2005). The contrasts of interest are defined by $\boldsymbol{\beta_g} = \boldsymbol{C}^\top\boldsymbol{\alpha_g}$. Given the expression- and design matrix the `lmFit()` function fits a linear model to gene g. The same model is applied to the other genes of the microarray experiment. The `coefficient` component contains the estimated coefficients for $\boldsymbol{\alpha_g}$. Given the fitted model, the `contrast.fit()` function estimates the coefficients and standard errors for $\boldsymbol{\beta_g}$. Unlike the design matrix which has to be full-ranked, the contrast matrix is allowed to be linearly

dependent. That means that the number of contrasts does not necessarily have to be the same as that of the coefficients. It is quite possible that the contrasts correspond to a subset of the original coefficients. As the coefficients themselves are usually of no further interest, but certain contrasts of them, they are re-calculated together with their standard deviations and their covariance matrix into the corresponding contrast objects. In the special case that the design matrix consists of a single n-dimensional column of ones $\boldsymbol{X}^\top = (1, \dots 1)$, the `Limma` t-test equals a simple t-test but adjusted for the gene-wise variance estimator which comprises both, the information of the gene alone and the pooled one of all genes. The contrast matrix is redundant in that case and can be omitted. The simplest experimental design is a microarray experiment with two treatment groups comparing *experiment* and *control* RNA. If the entries in the gene expression matrix correspond to the $\log_2$-fold changes, i.e. the differences in gene expression between experiment and control, the coefficient vector $\boldsymbol{\alpha_g}$ consists only of one element α_g. The coefficient estimator corresponds to the mean $\log_2$-fold changes of the g^{th} gene. In that special case, the coefficient α_g corresponds to the contrast of interest, namely the treatment effect.

As mentioned above, the coefficient estimator $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{g}}$, the estimator s_g^2 of the residual variance σ_g^2 and the estimated covariance matrix of $\widehat{\boldsymbol{\alpha}}_{\boldsymbol{g}}$: $\widehat{\text{var}}(\hat{\boldsymbol{\alpha}}_g) = \boldsymbol{W_g} s_g^2$ are re-calculated in terms of the contrasts $\hat{\boldsymbol{\beta}}_g = \boldsymbol{C}^\top \widehat{\boldsymbol{\alpha}}_{\boldsymbol{g}}$. Given the contrast matrix $\boldsymbol{C}$ and the weight matrix $\boldsymbol{W_g}$, the covariance matrix of $\boldsymbol{\beta}_g$ is estimated by

$$\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_g) = \underbrace{\boldsymbol{C}^\top \boldsymbol{W}_g \boldsymbol{C}}_{U_g} s_g^2.$$

Unlike the responses $\boldsymbol{y}_g$, for which no distributional assumptions are made, the contrast estimators are assumed to be at least approximately normally distributed with mean $\boldsymbol{\beta}_g$ and covariance matrix $\boldsymbol{C}^\top \boldsymbol{W}_g \boldsymbol{C} \sigma_g^2$. Let $U_g = \boldsymbol{C}^\top \boldsymbol{W}_g \boldsymbol{C}$ be the matrix whose diagonal entries correspond to the unscaled variances of $\hat{\boldsymbol{\beta}}_g$, then the j^{th} diagonal entry u_{gj} denotes the variance of $\hat{\beta}_{gj}$ and the ordinary t-statistics for the j^{th} contrast is given by

$$t_{gj} = \frac{\hat{\beta}_{gj}}{s_g \sqrt{u_{gj}}}, \tag{3.7}$$

which is t-distributed with d_g degrees of freedom.

Thus, the j^{th} component of the estimated contrast vector $\hat{\boldsymbol{\beta}}_g$ satisfies

$$\hat{\beta}_{gj} \big| \beta_{gj}, \sigma_g^2 \sim \mathcal{N}(\beta_{gj}, u_{gj} \sigma_g^2)$$

and

$$s_g^2|\sigma_g^2 \sim \frac{\sigma_g^2}{d_g}\chi_{d_g}^2,$$

where s_g^2 is approximately χ^2-distributed with d_g degrees of freedom.

The empirical Bayes approach uses a hierarchical model to estimate the posterior residual variances. The basic idea is to incorporate the common information provided by all genes together into the gene-wise sample variance. The residual variance σ_g^2 is assumed to be a priori inverse χ^2-distributed with d_0 degrees of freedom

$$\frac{1}{\sigma_g^2} \sim \frac{1}{d_0 s_0^2}\chi_{d_0}^2, \tag{3.8}$$

where s_0^2 is a further hyperparameter which is estimated from the data, similar to d_0. It can be shown that the χ^2-distribution in (3.8) can be rewritten as gamma distribution:

$$\frac{d_0 s_0^2}{\sigma_g^2} \sim \chi_{d_0}^2 \Rightarrow \frac{1}{\sigma_g^2} \sim \Gamma\left(k = \frac{d_0}{2}, \theta = \frac{2}{d_0 s_0^2}\right).$$

The estimators for both parameters, s_0^2 and d_0, are obtained from the observed sample variances s_g^2. Hence, the inverse posterior estimator for the residual variance $\frac{1}{\tilde{s}_g}$ corresponds to the mean of the gamma distribution which is the a posteriori distribution of $\frac{1}{\sigma_g^2}$ (Rempel, 2015):

$$\frac{1}{\sigma_g^2}\middle|(s_0^2, d_0, s_g^2) \sim \Gamma\left(k = \frac{d_0 + d_g}{2}, \theta = \frac{2}{d_0 s_0^2 + d_g s_g^2}\right).$$

It follows

$$\frac{1}{\tilde{s}_g^2} = \mathrm{E}\left(\frac{1}{\sigma_g^2}\middle|s_0^2, d_0, s_g^2\right) = \frac{d_0 + d_g}{d_0 s_0^2 + d_g s_g^2} \Rightarrow \tilde{s}_g^2 = \frac{d_0 s_0^2 + d_g s_g^2}{d_0 + d_g}.$$

For more details the reader is referred to Rempel (2015).

Substituting the prior standard deviation s_g in the ordinary t-statistic, given in (3.7), by the posterior standard deviation $\tilde{s_g}$, results in the moderated t-statistic

$$\tilde{t}_{gj} = \frac{\hat{\beta}_{gj}}{\tilde{s}_g\sqrt{u_{gj}}},$$

which is t-distributed with $(d_0 + d_g)$ degrees of freedom under the null hypothesis $H_0 : \beta_{gj} = 0$. The degrees of freedom are augmented due to the additional information which is borrowed from

the other genes to adjust the individual genes in terms of the global variance. In case of $d_0 = 0$, $\tilde{t}_{gj} = t_{gj}$ holds (Ritchie et al., 2015).

3.4 Statistics for concentration-dependent analyses

In concentration-dependent gene expression studies a convincing concentration progression is a criterion for data quality. Genes that are deregulated by a compound at a certain tested concentration are usually also deregulated at the next higher concentration. Therefore, genes with a deviating expression profile, i.e. genes with a non-monotonous concentration progression, may be indicative of low-data quality and, hence, should be treated with caution. In order to improve the data reliability, Grinberg et al. (2014) has introduced two indices for the progression analysis of gene alterations over increasing concentration levels, the *progression profile index* and the *progression profile error indicator*. Both statistics return exclusivity indices for the comparison of adjacent concentrations. They are calculated for each compound and incubation time point separately. Mathematically, the indices are defined as the probability of being not deregulated at a certain concentration level subject to the condition of being deregulated at an other concentration level.

Let C_1 and C_2 denote two concentration levels and $G_{C_1}^{\text{Diff}}$ and $G_{C_2}^{\text{Diff}}$ the events of being differentially expressed at C_1 and C_2. The complement $\overline{G_{C_1}^{\text{Diff}}}$ indicates the event of being not differentially expressed at C_1. The conditional probability of $\overline{G_{C_1}^{\text{Diff}}}$ given $G_{C_2}^{\text{Diff}}$ is then defined as the ratio of the probability of the intersection of the events $\overline{G_{C_1}^{\text{Diff}}}$ and $G_{C_2}^{\text{Diff}}$, and the probability of the event $G_{C_2}^{\text{Diff}}$:

$$\text{P}\left(\overline{G_{C_1}^{\text{Diff}}} | G_{C_2}^{\text{Diff}}\right) = \frac{\text{P}\left(\overline{G_{C_1}^{\text{Diff}}} \cap G_{C_2}^{\text{Diff}}\right)}{\text{P}\left(G_{C_2}^{\text{Diff}}\right)}. \tag{3.9}$$

This quantity is estimated by replacing the events with the corresponding relative proportions of genes that are deregulated.

3.4.1 Progression profile index

The progression profile index is defined as the ratio of two proportions, the proportion of genes that are deregulated exclusively at the higher concentration C_2, and the proportion of genes that are deregulated in total at C_2. In formula in (3.9) this corresponds to the situation $C_1 < C_2$. Values close to zero indicate that only a few additional genes are deregulated at the next higher

concentration, whereas values close to one indicate many additional genes deregulated at the higher concentration.

3.4.2 Progression profile error indicator

The progression profile error indicator is defined vice versa to the progression profile index, namely as the ratio of the proportion of genes that are deregulated exclusively at the lower concentration, and the proportion of genes that are deregulated in total at the lower concentration. In terms of formula (3.9), it holds $C_1 > C_2$. Values close to one indicate that a high fraction of genes are deregulated exclusively at a lower but not at the respective higher concentration. Values close to zero indicate the revers case. Compounds with values above 0.5 are considered as indicative of an implausible concentration progression.

3.4.3 Modified progression profile error indicator

The modified progression profile error indicator is an adjustment of the progression profile error indicator and has been introduced for the case that only a few genes are altered in total. As a certain amount of false positive genes is to be expected, a tolerance limit, i.e. a minimum amount of differentially expressed genes, should be set before including the respective genes in the calculations of the progression profile error indicator. Therefore the number of genes deregulated in total is incorporated in the calculation of that index. The progression profile error indicator is altered in the sense if the value of the index is larger than 0.5 and the number of genes deregulated at the respective lower concentration is below 20, the value of the index is set to zero. The interpretation of the modified index is the same as for the progression profile error indicator.

3.4.4 Selection value

To systematically analyze stereotypic versus compound-specific gene expression responses, the selection value principle has been introduced in Grinberg et al. (2014). A stereotypic response means that an expression alteration is induced by many compounds, while a specific expression response is induced by individual compounds or small numbers of compounds. For a gene, the selection value determines the number of compounds that induces a change in its expression. Compounds are ranked gene-wise in order of magnitude, in case of upregulated genes compounds are ranked from high to low fold changes and in case of downregulated genes from low to high

values. The selection value x for a gene (Sv x) defines the rank of the compound, indicating that the gene is induced by at least x compounds. The threshold for the critical change is pre-specified. In case of small replicate numbers, it is recommended to consider higher thresholds to keep the number of false positive genes as low as possible. The probability of false positive alerts decreases with increasing fold change. The higher the selection value the less compound-specific is the response. For a given fold change the so-called Sv 20 genes refer to those genes which respond to at least 20 compounds reflecting a stereotypical response. By contrast, a compound-specific response is here specified by Sv 3 genes, i.e. genes which are deregulated by at least three compounds. Note, that genes of higher selection values always overlap with genes of lower selection values, i.e. Sv 20 genes are a subset of Sv 3 genes.

Based on the selection value concept a consensus Sv x signature of genes comprises the Sv x gene lists of all individual test conditions. That means, the consensus Sv x list includes all those genes that show for at least one of the tested conditions a change in expression. Consensus genes are often used for the comparison of different model organisms, test systems, or data sets.

3.4.5 Overlap ratio

The overlap ratio is introduced to approach the question whether the overlap of genes between two test conditions, condition 1 and condition 2, corresponds to a randomly expected result. The ratio quantifies to which degree genes in the overlap are overrepresented, whereby a value of 1.0 indicates a random overlap and values higher than 1.0 are indicative of an overlap which is higher than expected by chance in case of independence. A ratio of 2.0, for example, indicates that twofold more genes are in the overlap than randomly expected. The overlap ratio is defined as follows:

$$\text{Overlap ratio} = \frac{O \cdot n_{\text{Gene universe}}}{n_{\text{Condition 1}} \cdot n_{\text{Condition 2}}},$$

where $n_{\text{Gene universe}}$ represents the total number of genes on the array (array $\hat{=}$ sample), $n_{\text{Condition 1}}$ represents the total number of genes that are altered under the influence of test condition 1, $n_{\text{Condition 2}}$ indicates the total number of genes differentially expressed under test condition 2, and O represents the number of genes in the overlap. Significance of overrepresentation is calculated by the Fisher test. The basic idea of the overlap ratio was first presented in Shinde et al. (2017).

3.5 Dose-response theory

To this date, differential expression analysis was only performed using the classical naïve approach, where for each measured concentration separately it is tested if the critical effect level is exceeded (*Limma* t-test). This procedure has the disadvantage that only measured concentrations can be considered as potential candidates for alert concentrations . But in practice, it is highly unlikely that such a deregulation is first triggered at exactly one of the measured concentrations. To this end, a model-based method is introduced which allows arbitrary positive values as alert levels.

Dose-response models are used in various application fields, such as pharmacology, pharmacokinetics, toxicology and clinical research. Typically, dose-response data exhibit a monotonic relationship between dose and response which can be modeled by a parametric regression model. Often, a sigmoidal dose-response trend is observed in the data which is characterized by S-shaped curves. Besides, there are other curves which can describe the response dependency of a dose, e.g. J-shaped or inverted U-shaped curves. These kind of curves are used for describing dose-response dependencies with a so-called *hormesis effect* which is, in toxicology, associated with low dosis effects and high dosis inhibitions. In the context of gene expression data, such curve progressions might be a hint for the use of cytotoxic concentrations which trigger cell death as response to compound exposure. However, this work addresses only the modeling of log-logistic functions which are by far the most commonly used models for describing dose-response relationships of toxicological background. Besides the log-logistic model, which exists in different parameterizations, there are other models, based for example on the log-normal- or Weibull distribution, that can be equally used to describe sigmoidal dose-response dependencies (Ritz, 2010).

All methods introduced in this section are based on an application of the four-parameter log-logistic model (4pLL). The 4pLL model is fitted to the data in order to estimate the Absolute Lowest Effective Concentration (ALEC) for a fixed and pre-specified effect level which is the concentration at which a pre-specified expression change is observed. The ALEC is derived from the fitted average trend (Jiang, 2013). Due to the fact that the critical concentration (ALEC) results from a simple point estimator, the uncertainty of the effect level is entirely neglected. But as it is vital to provide confidence intervals, the method introduced by Jiang (2013) is

enhanced by means of a thorough confidence interval estimator in this thesis. The hereby resulting concentration value is defined as the Lowest Effective Concentration (LEC).

3.5.1 Four-parameter log-logistic model (4pLL)

Given the parameter vector $\phi = \left(\phi^{(b)}, \phi^{(c)}, \phi^{(d)}, \phi^{(e)}\right)^\top$, the four-parameter log-logistic model (4pLL) is given by

$$\begin{aligned} y = f(x, \boldsymbol{\phi}) &= \phi^{(c)} + \frac{\phi^{(d)} - \phi^{(c)}}{1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}} \\ &= \phi^{(c)} + \frac{\phi^{(d)} - \phi^{(c)}}{1 + (x/\phi^{(e)})^{\phi^{(b)}}}, \end{aligned} \tag{3.10}$$

where x denotes the concentration and y the response. Under the assumption that the response dependency of the dose x can be described by a sigmoidal function, the 4pLL model is a suitable model to describe such a trend. In contrast to experiments where the response is for instance of physiological or biochemical nature or death (mortality rate), the studies used in the present work were conducted to investigate gene expression alterations induced by several test compounds. In terms of gene expression data the response y corresponds to the fold change ($\log_2$-scale), i.e. the difference in gene expression between treatment and control. Another term used in that context is *effect level*. Figure 3.2 shows an example of a 4pLL model for an increasing dose-response curve.

The function f reflects the mean difference between treatment and control (fold change) in relation to a concentration x. The parameters $\phi^{(c)}$ and $\phi^{(d)}$ specify the lower and upper horizontal asymptotes. The slope of f is determined by the parameter $\phi^{(b)}$, where a positive sign indicates a decreasing curve and a negative one an increasing curve. Higher values of $\phi^{(b)}$ are associated with a steeper curve progression. In case of an increasing curve, the lower limit corresponds to the average level of control values and the upper limit to the average level measured at the highest tested concentration. In case of a decreasing curve the upper and lower limit are interchanged. According to Ritz (2010) it is not recommended to scale continuous data to a range of values between 0 and 1, unless there is a good reason to do so. The parameter $\phi^{(e)}$ denotes the effective concentration EC_{50} which is the concentration that induces 50% reduction of the maximal effect. Different parameterizations can be used for the effective concentration $\phi^{(e)}$. The use of the logarithmized parameter $\phi^{(e)}$ provides more accurate parameter estimates (Ritz, 2010). The EC_{50} is then re-parameterized with the following relationship to the original parameter:

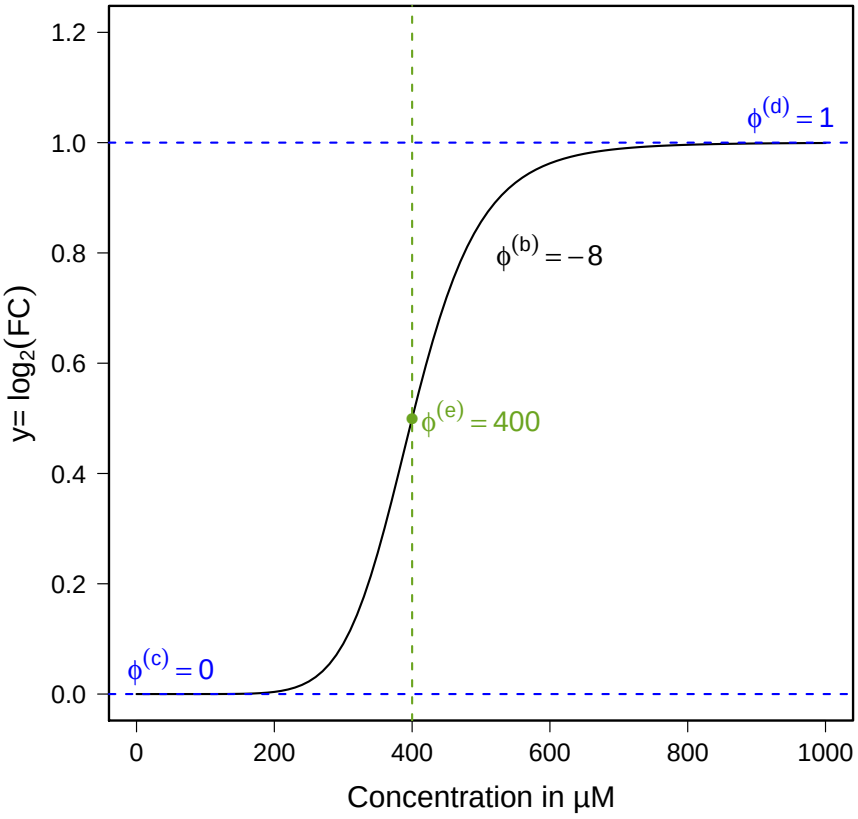


Figure 3.2: *Example of a four-parameter log-logistic (4pLL) model. The blue dashed lines indicate the lower and upper limit, respectively. The dashed green line displays the* EC_{50}*. The negative sign of the slope indicates an increasing curve.*

$\phi^{(e)} = \exp\left(\phi^{(e)^*}\right)$. In general, a dose-response model can be fitted by the `R` function `drm` which is implemented in the `drc` package. The 4pLL model function with $\log\left(\phi^{(e)}\right)$ can be specified by the argument `fct = LL2.4()` (Ritz and Streibig, 2005).

3.5.2 Re-parametrization of the EC_{50}

The choice of the parameterization depends on the sample size and the parameters to estimate. Due to the non-linearity of the function in (3.10), f is approximated according to the least square method with the Gauss-Newton algorithm which assumes that the parameter estimates are approximately normally distributed. However, for a small sample size ($n <$ 20-30) the normality assumption is questionable. The violation of the distributional assumptions can lead to biased estimators. The use of another parameterization can circumvent this problem. By definition, the parameter $\phi^{(e)}$ is a concentration and therefore needs to be a positive value which implicates a right skewed distribution of its estimate. The smaller the data set, the more

pronounced is the skewness and the more difficult it is to approximate the distribution by a normal distribution. Therefore, it is more appropriate to estimate the logarithm-transformed parameter $\phi^{(e)^*} = \log\left(\phi^{(e)}\right)$ rather than the original parameter $\phi^{(e)}$, since $\phi^{(e)^*}$ permits values on the entire real axis which makes the normality assumption more plausible. The EC_{50} together with its confidence interval is then obtained by back-transformation of the corresponding log-scaled parameter estimators.

The following function has to be minimized:

$$\text{RSS} = \sum_{j=1}^{n}(y_j - f(x_j, \boldsymbol{\phi}))^2,$$

with n indicating the number of data pairs (x_j, y_j). Note, that the algorithm used is an iterative method for approximating the parameter estimators and therefore does not guarantee to find the global minimum of the function, that means, depending on the starting values of the iteration process, the algorithm can result in a local rather than global minimum. By default, the starting values are estimated from a self starter function which is specified by the user. In case of the four-parameter log-logistic model the function `fct = LL2.4()` determines the initial values of the four parameters $\phi^{(b)}, \phi^{(c)}, \phi^{(d)}$ and $\phi^{(e)}$. The upper and lower limits $\phi^{(c)}$ and $\phi^{(d)}$ are set to the minimum and maximum response value $\pm$ 0.001. The initial values of the parameters $\phi^{(b)}$ and $\phi^{(e)}$ are set as follows: A critical value is defined by the average value of the upper and lower limit. In a stepwise procedure, the mean response value of two neighboring concentration levels is calculated and compared with the critical value. If the critical value lies between the two response values, the search algorithm stops, otherwise the first concentration value is replaced by the next higher concentration value and the calculations are repeated. The initial value of $\phi^{(e)}$ is then obtained by taking the average value of the two given concentration levels and the starting value of $\phi^{(b)}$ is calculated by subtracting the two given response values from each other (value with a negative sign is taken).

Alternatively, an optional vector of starting values can be specified by the user. Moreover, convergence problems can occur as a result of different parameterizations. For more details the reader is referred to Ritz (2010).

3.5.3 The ALEC and its confidence interval

For a parametric regression model function $y = f(x, \boldsymbol{\phi})$ the ALEC estimates the Absolute Lowest Effective Concentration for a pre-specified critical effect level λ and is defined as the inverse function of f:

$$f(\text{ALEC}, \boldsymbol{\phi}) = \lambda \Rightarrow \text{ALEC} = f^{-1}(\lambda, \boldsymbol{\phi}).$$

In the parameterization (3.10) the ALEC is defined as a function $h(\phi)$ for a given λ:

$$\text{ALEC} = h(\boldsymbol{\phi}) = \phi^{(e)} \left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right)^{1/\phi^{(b)}}. \tag{3.11}$$

The ALEC can only be estimated for an effect level λ which lies between the lower and upper limit $\phi^{(c)}$ and $\phi^{(d)}$.

A confidence interval for the ALEC can be estimated by using the delta method to approximate the variance of $h(\boldsymbol{\phi})$:

$$\text{Var}(h(\boldsymbol{\phi})) \approx \nabla h(\boldsymbol{\phi})^T \cdot \Sigma \cdot \nabla h(\boldsymbol{\phi}), \tag{3.12}$$

where Σ corresponds to the variance-covariance matrix of the parameter vector $\boldsymbol{\phi}$ and $\nabla h(\boldsymbol{\phi})$ to the gradient of h:

$$\nabla h(\boldsymbol{\phi}) = \begin{pmatrix} \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(b)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(c)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(d)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(e)^*}} \end{pmatrix} = h(\boldsymbol{\phi}) \begin{pmatrix} -\frac{1}{\phi^{(b)^2}} \log\left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right) \\ \frac{1}{\phi^{(b)}(\lambda - \phi^{(c)})} \\ \frac{1}{\phi^{(b)}(\phi^{(d)} - \lambda)} \\ 1 \end{pmatrix}.$$

A detailed derivation of $\nabla h(\boldsymbol{\phi})$ is provided in Section A.1 in the Appendix. Differentiating term (3.11) with respect to the parameter vector $\hat{\boldsymbol{\phi}} = \left(\hat{\phi}^{(b)}, \hat{\phi}^{(c)}, \hat{\phi}^{(d)}, \hat{\phi}^{(e)} \right)^T$ and inserting the estimated ALEC together with its estimated variance, into (3.13) results in the following $(1 - \alpha)$ confidence interval

$$\widehat{\text{ALEC}} \pm t_{\nu,(1-\alpha/2)} \sqrt{\widehat{\text{var}}(\widehat{\text{ALEC}})}, \tag{3.13}$$

where $t_{\nu,(1-\alpha/2)}$ is the $(1 - \alpha/2)$ quantile of a t-distribution with $\nu = n - 4$ degrees of freedom for n observations. However, Jiang (2013) recommends to calculate the confidence interval for the logarithmized ALEC estimator and then to back-transform the log-scaled confidence interval

to the original dose scale. That ensures positive estimators for the lower interval limit. To apply this procedure, the expressions in (3.11), (3.12) and (3.13) are re-parameterized into the terms in (3.14), (3.15) and (3.16), respectively:

$$\log(\text{ALEC}) = \log(h(\boldsymbol{\phi})) = \log(\phi^{(e)}) + \frac{1}{\phi^{(b)}} \log\left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}}\right), \tag{3.14}$$

$$\text{Var}(h(\boldsymbol{\phi})) \approx \nabla \log(h(\boldsymbol{\phi}))^T \cdot \Sigma \cdot \nabla \log(h(\boldsymbol{\phi})), \tag{3.15}$$

$$\exp(\log(\widehat{\text{ALEC}})) \pm t_{\nu,(1-\alpha/2)} \sqrt{\widehat{\text{var}}(\log(\widehat{\text{ALEC}}))}. \tag{3.16}$$

Besides the delta method, there are other methods which can be used for the construction of confidence intervals like the profile likelihood method which is based on a re-parameterization of function (3.10) or the bootstrap method that is based on resampling techniques. For a detailed description of the two methods see Jiang (2013).

3.5.4 The effect level and its confidence interval

Given formula (3.10), let the function $f(\cdot)$ define the effect level. The confidence interval for a fixed effect level λ can then be computed analogously to the confidence interval of the ALEC. Approximating the following term

$$\text{Var}(f(x, \boldsymbol{\phi})) \approx \nabla f(x, \boldsymbol{\phi})^T \cdot \Sigma \cdot \nabla f(x, \boldsymbol{\phi}) \tag{3.17}$$

with the delta method provides an estimator for the variance $\widehat{\text{var}}(\widehat{f(x,\boldsymbol{\phi})})$. Differentiating the function in (3.10) with respect to the parameter vector $\boldsymbol{\phi}$ gives the gradient of f:

$$\nabla f(x,\boldsymbol{\phi}) = \begin{pmatrix} \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(b)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(c)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(d)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(e)}} \end{pmatrix} = \begin{pmatrix} -\frac{\left(\phi^{(d)}-\phi^{(c)}\right)\left(\log\left(\frac{x}{\phi^{(e)}}\right)\right)\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\left[1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2} \\ 1-\left[\frac{1}{1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}\right] \\ \frac{1}{1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}} \\ \frac{\phi^{(b)}\left(\phi^{(d)}-\phi^{(c)}\right)\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\phi^{(e)}\left[1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2} \end{pmatrix}. \tag{3.18}$$

For a detailed derivation of $\nabla f(x,\boldsymbol{\phi})$ the reader is referred to Section A.2 in the Appendix. Inserting the estimated parameters $\hat{\phi}^{(b)}, \hat{\phi}^{(c)}, \hat{\phi}^{(d)}$ and $\hat{\phi}^{(e)}$ into the functions (3.10) and (3.18) gives estimators for f and its first order derivatives ∇f for a specific concentration x. Estimating the variance-covariance matrix of the four parameters provides $\hat{\Sigma}$. By applying (3.17) to the estimated parameters, the $(1-\alpha)$ confidence interval for the effect level is obtained:

$$\widehat{f(x,\boldsymbol{\phi})} \pm t_{\nu,(1-\alpha/2)}\sqrt{\widehat{\text{var}}(\widehat{f(x,\boldsymbol{\phi})})}, \tag{3.19}$$

where $t_{\nu,(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of a t-distribution with $\nu = n-4$ degrees of freedom for n observations.

Figure 3.3 illustrates the ALEC estimator together with its 95%-confidence interval, with respect to both concentration and response (fold change). The horizontal line visualizes the 95%-confidence interval obtained by formula (3.13) and the vertical line results from the application of formula (3.19). The red dashed line marks the critical effect level which is 1.5-fold in the present case.

3.5.5 Test statistic for the effect level

Consider the 4pLL model function (3.10) with its parameter vector $\boldsymbol{\phi} = \left(\phi^{b}, \phi^{(c)}, \phi^{(d)}, \phi^{(e)}\right)^\top$. Let $\widehat{\text{var}}(\widehat{f(x,\boldsymbol{\phi})})$ denote the estimate of the variance of $f(x,\boldsymbol{\phi})$ which is approximated according

to formula (3.17) and let λ be the effect level of interest. The test statistic for testing the hypothesis $H_0 : f(x, \boldsymbol{\phi}) = \lambda$ is given by

$$t_{4pLL} = \frac{|\widehat{f(x, \boldsymbol{\phi})} - \lambda|}{\sqrt{\widehat{\text{var}}(\widehat{f(x, \boldsymbol{\phi})})}}. \tag{3.20}$$

H_0 is rejected at level α if the observed value of t_{4pLL} exceeds the $(1 - \alpha/2)$ quantile of a t-distribution with ν degrees of freedom where $\nu = n - 4$ and n is the number of observations.

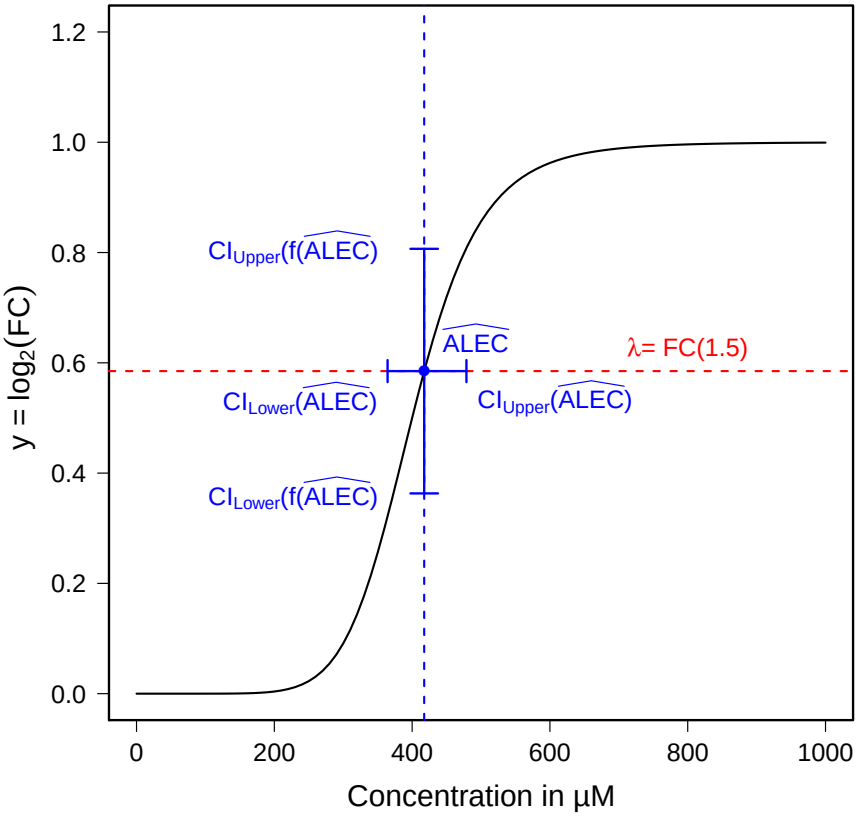


Figure 3.3: *Illustration of the ALEC estimator and its corresponding 95%-confidence intervals, with respect to concentration (horizontal line) and response (vertical line). Dashed lines indicate estimators and solid lines confidence intervals.*

3.5.6 Lowest Effective Concentration (LEC)

The LEC estimates the *L*owest *E*ffective *C*oncentration at which a given fold change is exceeded significantly. For estimating the LEC, a grid search within the permissable parameter space is performed. The used algorithm is presented in pseudo-code in Algorithm 1. Let λ denote the critical effect level and x the test concentration. In this case the t_{4pLL}-test in (3.20) is used to test

Algorithm 1 Search algorithm for the LEC estimator.

Require: $ps = ev$, $start = ALEC$, $finish = 1000$, $\epsilon = 0.01$
$erg \leftarrow t.test_{4pLL}(x = finish)$
if erg is not significant **then**
  Stop and
  return "Maximal number of iterations is reached"
  $start \leftarrow 1$
else
  $lower \leftarrow start$
  $upper \leftarrow finish$
end if
repeat
  $new \leftarrow (lower + upper)/2$
  $erg \leftarrow t.test_{4pLL}(x = new)$
  if erg is significant **then**
    $upper \leftarrow new$
  else
    $lower \leftarrow new$
  end if
until $|upper - lower| < \epsilon$
return new

whether the fold change at the given concentration is different from the pre-specified effect level λ. The null hypothesis is tested at each concentration separately. If the obtained statistic leads to the rejection of $H_0 : f(x, \boldsymbol{\phi}) = \lambda$ and the critical effect level lies below (upregulation) the lower- or above (downregulation) the upper confidence limit, then the tested concentration exceeds the critical effect level at level α significantly. To avoid unnecessarily long computation times, the test concentrations within the parameter space are well-chosen. The lower and upper limit of the parameter space are defined by the ALEC and the highest test concentration which can be specified by the user. The algorithm starts at the highest test concentration and iteratively determines the next test concentration. If no significant result for the highest tested concentration is obtained, the algorithm stops, otherwise the parameter space is halved. The center of the lower and upper limit is chosen as the new test concentration. If the new test concentration exceeds the critical effect level significantly, the parameter space within which the search takes place is restricted to its lower half, otherwise to its upper half. The concentration search continues until the algorithm converges which is the case if the difference between the lower and upper limit is smaller than a given range specified by ϵ. In case of convergence, the algorithm returns an estimator for the LEC, otherwise a warning message that the estimated value lies outside the valid range. The ALEC and LEC estimator are visualized together with their 95%-confidence

intervals in Figure 3.4. The ALEC defines the concentration at which the given fold change (dashed red line) is reached exactly and the LEC refers to the concentration at which the given threshold is exceeded significantly.

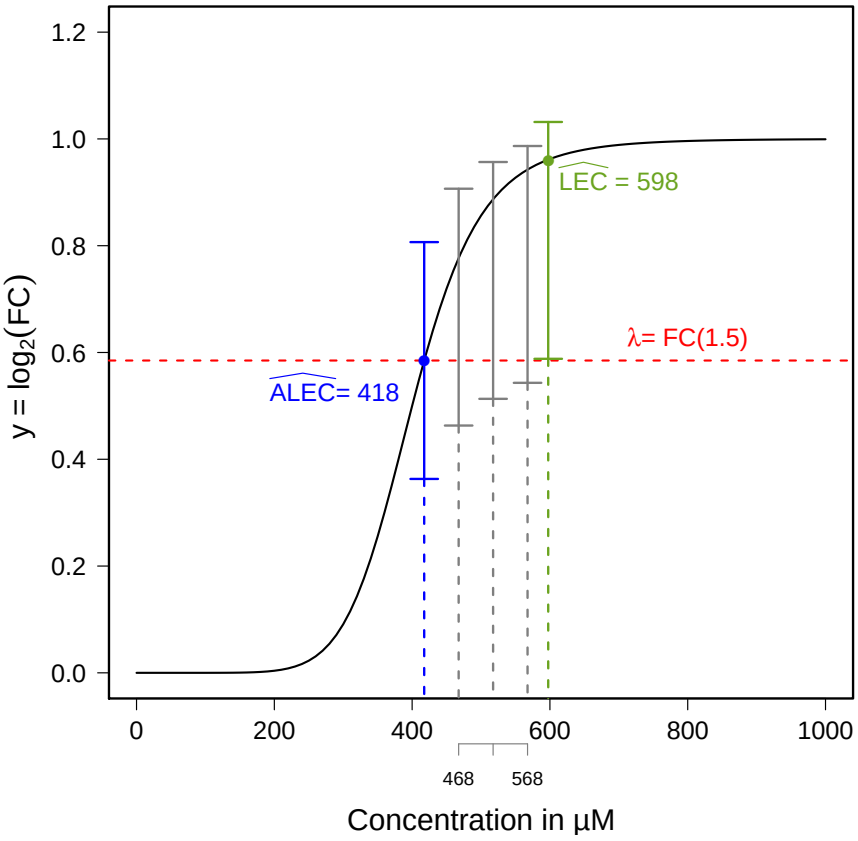


Figure 3.4: *Illustration of the ALEC estimator (in blue) and LEC estimator (in green). The ALEC defines the concentration at which the 1.5-fold threshold is reached exactly (point estimate) and the LEC refers to the concentration at which the given threshold is exceeded significantly (CI-based estimate). For the LEC estimator a grid search is performed within the respective limits.*

3.5.7 Measures of toxicity

The 4pLL method introduced in Section 3.5.4 is an alternative to the classical naïve *Limma* approach where for each measured concentration separately it is tested whether the given threshold is exceeded significantly. In this case the Lowest Observed Effective Concentration (LOEC), also referred to as CI-based estimate, indicates the lowest concentration at which the average value of the fold change exceeds the critical effect level significantly. The *Limma* t-test from Section 3.3 is used to test the hypothesis.

By contrast, the Lowest Effective Concentration (LEC) refers to the lowest concentration at which the fitted fold change exceeds the given threshold significantly (CI-based estimate). Therefore, the LEC estimator is also defined by hypothesis testing, but this time the test procedure is performed within a pre-specified concentration range by means of an iterative algorithm which uses the $t_{4\text{pLL}}$-test. The used algorithm is presented in pseudo-code in Algorithm 1 and the used test statistic in (3.20).

The Absolute Lowest Observed Effective Concentration (ALOEC) defines the lowest concentration at which the given fold change is exceeded by its average fold change value. In contrary to the LOEC (point estimate), no significance testing is performed. Both estimators, the ALOEC and LOEC, are restricted to measured concentration levels. This explains the letter "O" in the names of the two *Limma* estimators. The "O" indicates that for these two alert concentrations only measured values are potential candidates for estimates.

The estimates for the LEC and the ALEC can take any continuous value in the given concentration range. The ALEC (point estimate) is defined as the concentration at which the critical effect level is reached exactly by the fitted fold change value. Table 3.1 summarizes the four parameters with respect to their estimation method.

Table 3.1: *Methods for estimating alert concentrations from concentration - gene expression studies. Rows indicate the cut-off criteria and columns the methods for estimating critical concentrations. An alert means either that a given fold change value is reached exactly (4pLL) or exceeded by the average value (Limma) (FC) or that additionally the corresponding p-value is below a given cutpoint (FC & p-value). The p-value results from the t-test from the $4pLL$ model-based approach or from the Limma t-test.*

	4pLL	**Limma**
FC	**ALEC** (**A**bsolute **L**owest **E**ffective **C**oncentration)	**ALOEC** (**A**bsolute **L**owest **O**bserved **E**ffective **C**oncentration)
FC & *p*-value	**LEC** (**L**owest **E**ffective **C**oncentration)	**LOEC** (**L**owest **O**bserved **E**ffective **C**oncentration)
Criteria	Fitted fold change value ↪ Allow arbitrary positive values	Average fold change value ↪ Restricted to measured concentration levels

4 Toxicogenomics directory of chemically exposed human hepatocytes

This chapter adresses the statistical challenges which arise from the task to analyze a large-scale gene expression data set, in this case on the basis of the TG-GATEs data set (NIBIOHN, 2017). The database is one of the largest toxicogenomic databases to date and was generated by treating cultivated primary human and rat hepatocytes, as well as rat liver and kidney samples with more than 150 compounds using different concentration and time points. However, despite the advantages of such an extensive database, one of the main challenges of toxicogenomics is to identify artifacts and to eliminate errors. To tackle this problem, several curation steps, including batch correction, assessment of data reproducibility and exclusion of implausible data, are applied. The curated database is then used to analyze the structure of the chemically induced gene expression alterations . The results reported in this chapter have been published in Grinberg et al. (2014). All calculations were performed with the statistical software R version 3.3.2 (R Core Team, 2015).

4.1 Batch effects

Besides the curse of dimensionality, the greatest challenges while working with such large data sets is to identify and remove substantial errors. Experimental errors occur inevitably but can be controlled to a certain extent. In order to do so, various steps of a curation process must be performed. In the first step, a principal component analysis (PCA) is carried out in order to visualize the different gene expression alterations across all compounds and replicates. The PCA is based on the 100 probe sets with highest fold change (absolute values) across all compounds. The analysis is performed for all concentration and time sets separately. Figure 4.1 displays

the data of the high concentration and 24h time set. The corresponding figures for all other test conditions are shown in the Appendix (Figures C.1-C.3). Each point represents one experiment, where the color coding indicates the controls (dark green) and the exposed samples (light green). The percentages of the variances covered are indicated on the axes. Plotting of the first two principal components reveals two main clusters. Non-exposed samples are located on the left side of the PCA plot, while exposed samples move along the first principal component which explains 40.4% of the data variability. Controls are subdivided into two batches, but cluster at least within the two batches closely together (Figure 4.1 A). In Figure 4.1 B replicate samples of the same compounds are connected by lines to visualize the reproducibility between replicates. Most of the replicate pairs are located next to each other indicating a high replicate reliability. Due to the low technical variability, replicate values are averaged for further analyses (Figure 4.1 C). As a next step, control-treatment samples are compared by connecting lines (Figure 4.1 D). The resulting pattern reveals that the individual control-treatment pairs are located within the same cluster. This gives reason to believe that the two batches arise from experimental discrepancies. Subtracting controls from the corresponding compound exposed samples results in a pattern without any clusters. The aforementioned batch effect is removed by simple control subtraction. No further procedures are required to correct for batch effects (Figure 4.1 E).

4.2 Reproducibility

As the PCA provides only a visual overview of the technical variability between compound specific replicates, the reproducibility between them is analyzed in more detail. For quantifying the distance between replicate pairs, Euclidean distances are calculated and compared to the Euclidean distances of control-treatment pairs. The results are visualized by histograms. Figure 4.2 illustrates the results for the 24h, high concentration subset. The histogram in the left-hand panel shows the distances between the replicates tested at that condition. Except for some outliers, the distribution is approximately normally distributed with mean 2.1 and standard deviation 0.5. The 95%-quantile of the distribution, indicated by the red line in the histogram, determines the threshold for the acceptable variability range. Distances of larger values are considered as outlier candidates and are visualized with lines in the PCA plot on the right-hand side of the figure. For the given subset, 14 compounds (9.5%) are identified as outliers. Comparing the distances between replicate pairs with those between control-treatment sample pairs shows that even the 5% largest observed distances exhibit a relatively low degree of variability (in median 4.9-fold

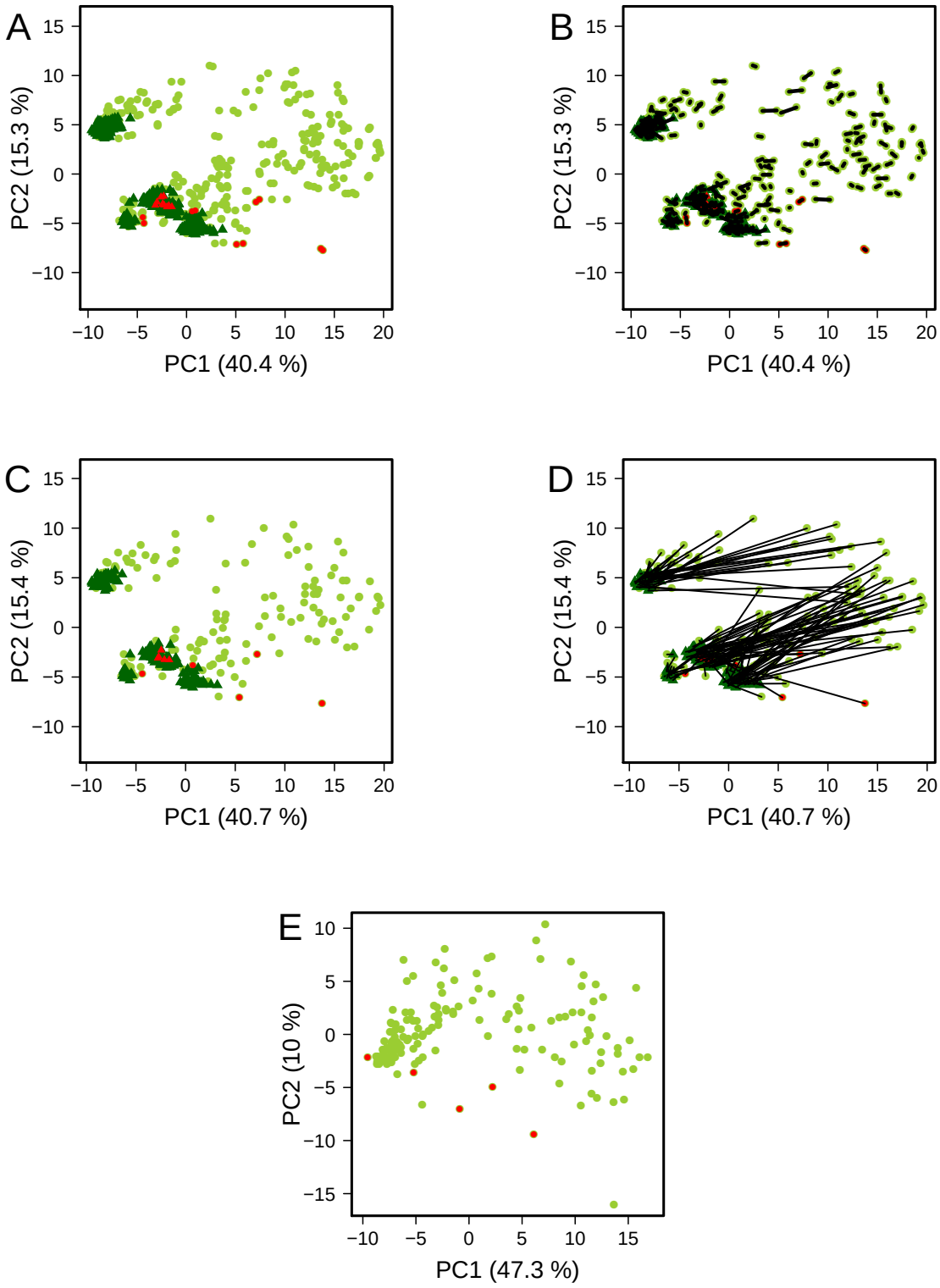


Figure 4.1: *Principle component analysis of gene expression data obtained from human hepatocytes after incubation with 148 chemicals (green symbols) and 7 cytokines (red symbols). Data of the high concentration and 24h incubation is shown. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

lower) in relation to the much larger control-treatment effect. Similar results are obtained for the other concentration and time sets (see Figures C.4-C.6). Distances between control-treatment pairs are visualized in Figure C.7 in the Appendix.

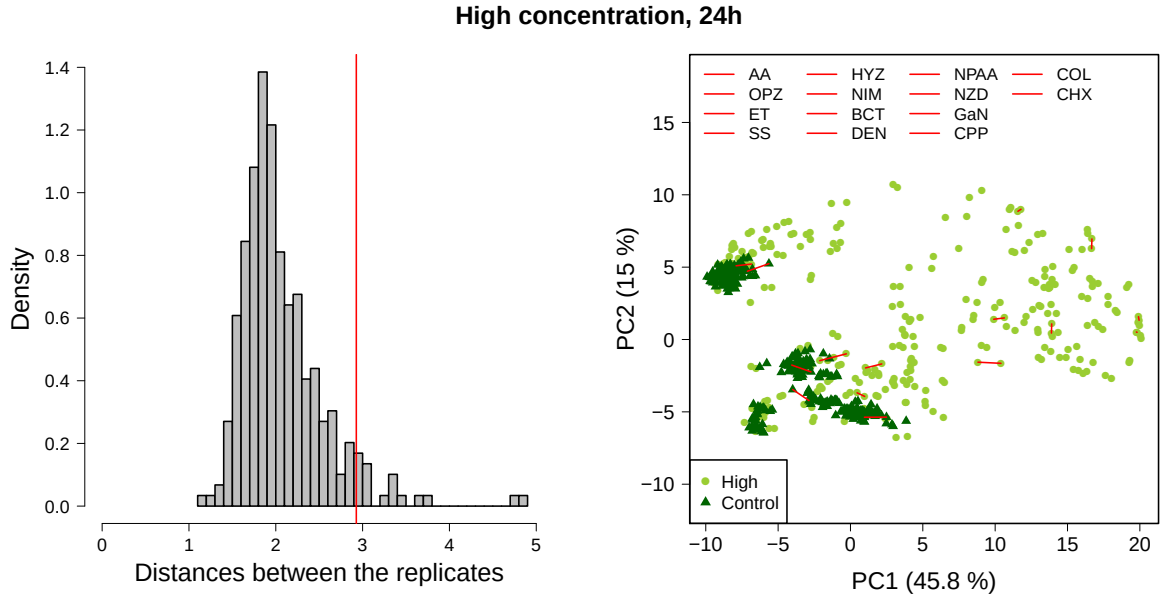


Figure 4.2: *Reproducibility between replicates. The left panel shows the frequency distribution of the Euclidean distance between all pairs of replicates. The red line indicates the 5% largest observed distances between replicates. The right panel shows the PCA plot with connecting lines between the 5% largest observed distances, representing 14 (9.5%) of the compounds tested in the 24h, high concentration subset. The variability of the worst replicates is still relatively small in relation to the much larger compound effects shown by connecting lines in Figure 4.1 D.*

4.3 Number of deregulated genes

As a next step, the number of deregulated genes is calculated for all concentration and time sets. This is shown in Figure 4.3. The barplots list the number of genes with at least 1.5-, 2.0- and 3.0-fold expression change, separately for each test condition. All results are presented cumulated. The barplot in the lower right-panel for the 24h, high concentration subset reveals that the compound cycloheximide (CHX) is responsible for most of the gene expression alterations. After incubation of CHX, 887 genes show a change of at least threefold, 2547 and 5124 genes of at least 2.0- and 1.5-fold, respectively. Similar numbers are observed for downregulated genes. In contrast, triazolam (TZM) causes, under the same test conditions, a change of at least 1.5-fold in only 37 genes (6 up and 31 down) and in only one gene of at least twofold. As a whole, Figure 4.3 shows that the number of induced genes differs strongly between compounds.

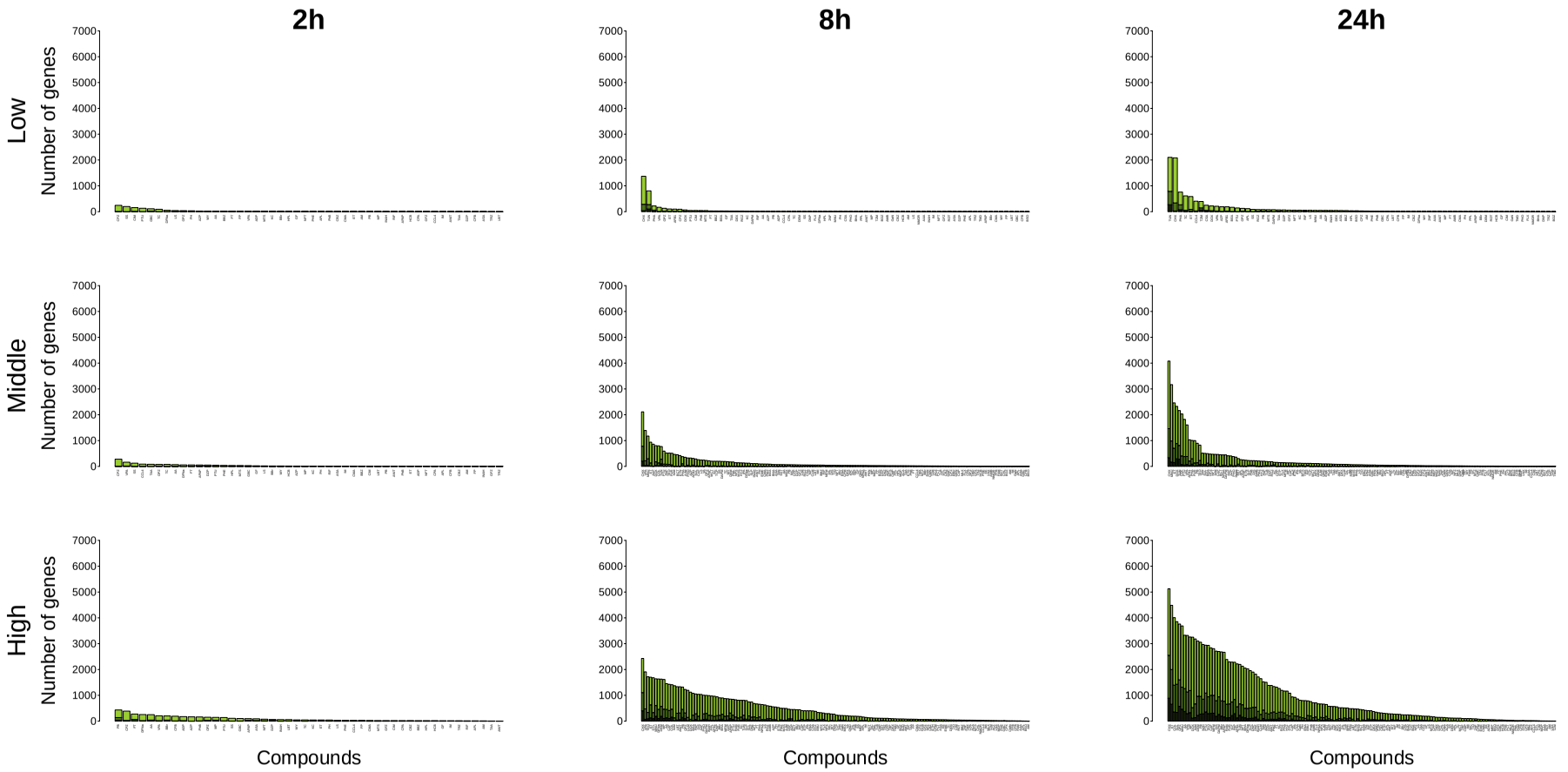


Figure 4.3: *Number of significantly upregulated genes. The x-axis lists all chemicals that were tested at the indicated concentration for the corresponding period. The y-axis gives the number of upregulated genes with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. Dark green: more than 1.5-fold upregulated; light green: more than twofold upregulated; black: more than threefold upregulated. The corresponding data for the downregulated genes is shown in Figure C.8 in the Appendix.*

While most of the compounds cause only very few gene expression alterations, only a small number of compounds induce high fold changes. The number of induced genes increases with increasing concentration and time of incubation. 48 compounds have been tested at all nine test conditions, 11 of them have been characterized as *weak compounds* due to the fact that they deregulate (twofold up or down compared to control) less than 20 genes in total across all concentration and time sets.

In addition, an exclusivity analysis is performed for the up- and downregulated genes, meaning that the 100 top-ranking genes with the highest fold change across all 148 compounds have been assigned to the compound with the most extreme fold change. The same 100 genes were used for the PCA plots in Figure 4.1. The analysis is performed for the up- and downregulated genes separately. Figure 4.4 visualizes the results for the strongest upregulated genes and Figure C.9 in the Appendix shows the results for the corresponding downregulated genes. The dark and light green bars highlight the absolute number of genes deregulated by the compound and the number of genes which are deregulated twofold up or down compared to control, respectively. The exclusivity analysis reveals that only a small fraction of compounds contribute to the 100 strongest deregulated genes. In case of the 24h, high concentration subset, for example, only 32 of the 148 compounds tested, causes upregulations and even fewer compounds ($n = 23$) induce downregulations. The results strengthen the assumption that the majority of the compounds induces only weak expression changes.

4.4 Concentration progression

As mentioned earlier, experimental errors cannot be fully avoided, but can be identified and curated to some extent. It requires some curation steps to improve the data quality, the removal of batch effects is the first step of many and forms the basis for further analyses. The detection of batch effects is important as otherwise relevant mode of actions might remain undetected due to strong batch effects. In the second and third step the data reproducibility across replicates and the increase of gene deregulations across concentration and time sets is assessed. The next step in the curation procedure is the identification of compounds which exhibit an implausible concentration progress. For this, two indices, the *progression profile index* and the *progression profile error indicator*, are introduced for the comparison of adjacent concentrations (Section 3.4).

The progression profile error indicator value is calculated for each compound and specifies the proportion of additional genes at the next lower dose, i.e. the error indicator indicates the fraction

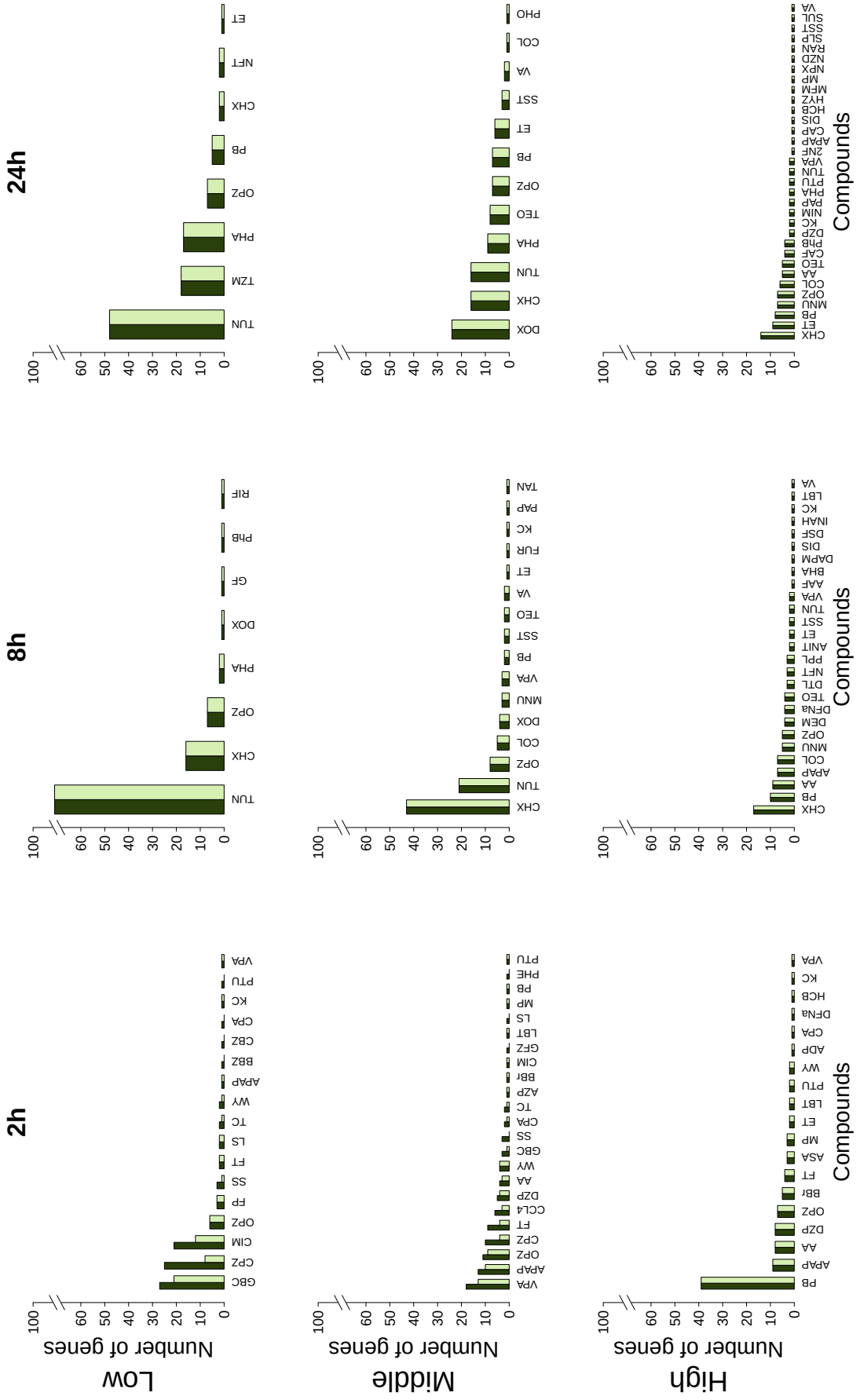


Figure 4.4: *Exclusivity analysis of the upregulated genes. First, this analysis determines the 100 strongest upregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The corresponding analysis for the downregulated genes is shown in Figure C.9 in the Appendix.*

of genes deregulated exclusively at a lower compared to a respective higher concentration, resulting here in the comparison of the low versus middle and middle versus high concentration for each of the three exposure periods 2h, 8h and 24h. Compounds that have been tested with only two concentrations were excluded from the calculations.

The progression profile index which indicates the proportion of genes that are deregulated exclusively at the next higher compared to the respective lower concentration, serves as additional information. Genes are considered as deregulated if they exhibit a change in gene expression of at least twofold. For the progression profile index high values are desirable, while low values are more desirable for the progression profile error indicator. If the latter is the case, a monotonous concentration progression from the low to the middle and from the middle to the high concentration can be assumed.

But as false positive genes are almost unavoidable, the total number of deregulated genes should not be disregarded when calculating the error indicator. If a certain number of non-monotonous genes falls below a pre-specified level, then they should not be included in the calculations of the the error indicator. In the present study, the threshold is set to 20, meaning that in case of less than 20 deregulated genes the respective error indicator value is automatically set to zero. This results in the modified progression profile error indices (Section 3.4).

The principle is illutrated in Figure 4.5, exemplified by the four compounds valproic acid (VPA), propranolol (PPL), triazolam (TZM) and allyl alcohol (AA). The compounds were chosen to illustrate specific cases. Figure 4.5 shows the results of the 24h, high concentration subset. The left panel shows the individual expression profiles of the four compounds at the low, middle and high concentration. The corresponding Venn diagrams which count the genes that are upregulated twofold by the indicated compound are provided in the middle panel. The right panel shows the plots for the respective profilers with the four compounds highlighted in green. Each symbol represents an individual compound. Color coding indicates if more (black) or less (grey) than 20 genes are deregulated by the indicated compound. The triangles represent the compounds that will be later excluded from the analysis.

VPA shows a monotonous concentration progression and therefore yields a relatively high value for the progression profile index for both concentration comparisons, low versus middle and middle versus high. This positions the compound in the upper right corner of the progression profile index plot. Similar values are obtained for the compound AA which clusters closely to VPA. TZM, in contrast, induces genes only at the low but not at the next higher concentrations. This places TZM in the lower left corner of the progression profile index plot.

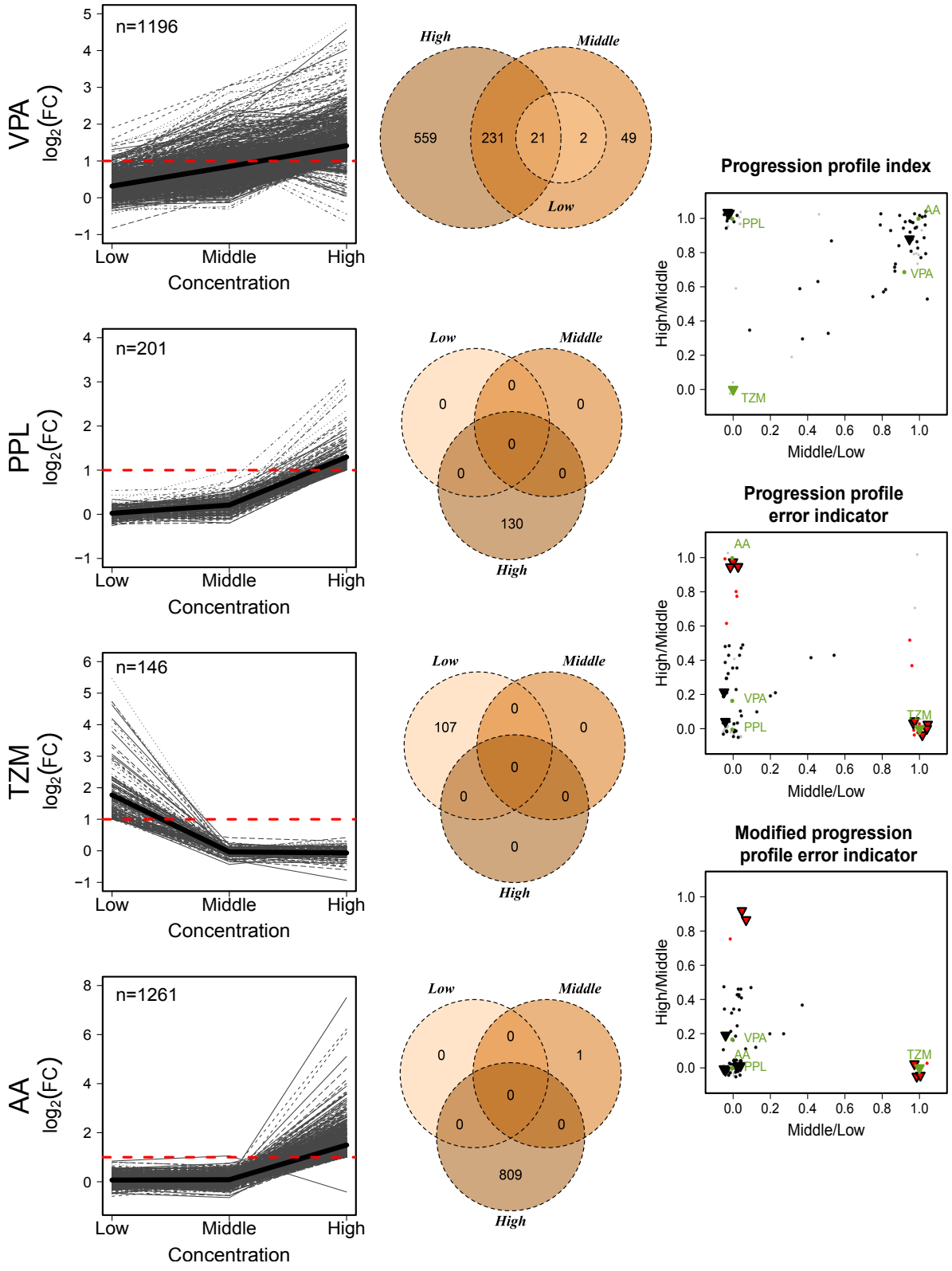


Figure 4.5: *Concentration progression analysis. Principles of the progression profile index and the progression profile error indicator illustrated for the four compounds: Valproic acid (VPA), propranolol (PPL), triazolam (TZM) and allyl alcohol (AA). Only the upregulated genes after 24h exposure are shown here. The left panel shows the levels of the individual upregulated genes at three concentrations. The panel in the middle summarizes the upregulated genes by Venn diagrams. The right panel shows the resulting positions of the four compounds in the respective profile plots indicated in green.*

The progression profile error indices for the four compounds reveal that VPA, PPL and AA yield relatively low values for at least one pair of adjacent concentrations. All three compounds are located on the left-hand side of the progression profile error indicator plot and move up along the y-axis. An exception is TZM which is located in the lower right corner of the plot, indicating that genes that are upregulated with the low concentration are not upregulated with the middle concentration. After adjusting for the number of altered genes, the compound AA moves from the upper left to the lower left corner in the plot of the modified progression profile error indicator. The relocation of the compound's position indicates that the compound induces in less than 20 genes a twofold change at the middle concentration.

The progression profile indices for all compounds are depicted in Figure 4.6. Rows indicate the exposure period and columns the direction of deregulation. The majority of the compounds cluster to the upper right corner of the plot, independently from time point and direction (up or down). This indicates that the number of additional genes that are induced by the indicated compounds increase with increasing concentration. Those compounds that cluster in the upper left or lower right corner exhibit a distinct progression only from the low to the middle or the middle to the high concentration, respectively. Weak compounds which cause no expression changes are located in the lower left corner of the plot.

Figure 4.7 shows the corresponding plots for the progression profile error indices. Compounds deregulating more than 20 genes and yielding an error indicator value above 0.5 for at least one concentration comparison are marked in red. Basically, two main clusters can be observed for all incubation time points. The majority of the compounds cluster to the left-hand side of the plot indicating a plausible concentration progression from the low to the middle concentration. The other compounds cluster to the right-hand side of the plot. This reflects the situation where a high fraction of genes is deregulated exclusively at the low but not at the middle concentration. Compounds that cluster in the lower or upper half of the plot indicate a plausible or implausible progression, respectively, in the comparison of the high versus middle concentration. After applying the modified progression profile error indicator to the compounds, almost all cluster to the lower left corner of the plot indicating a plausible concentration progression (Figure C.10).

On the basis of the modified progression profile error indices a *progression error profile* is defined for each compound which documents a compound's concentration progression for each time period. A profile is created for the up- and downregulated genes separately. For a detailed description of how the profile is defined the reader is referred to Grinberg et al. (2014).

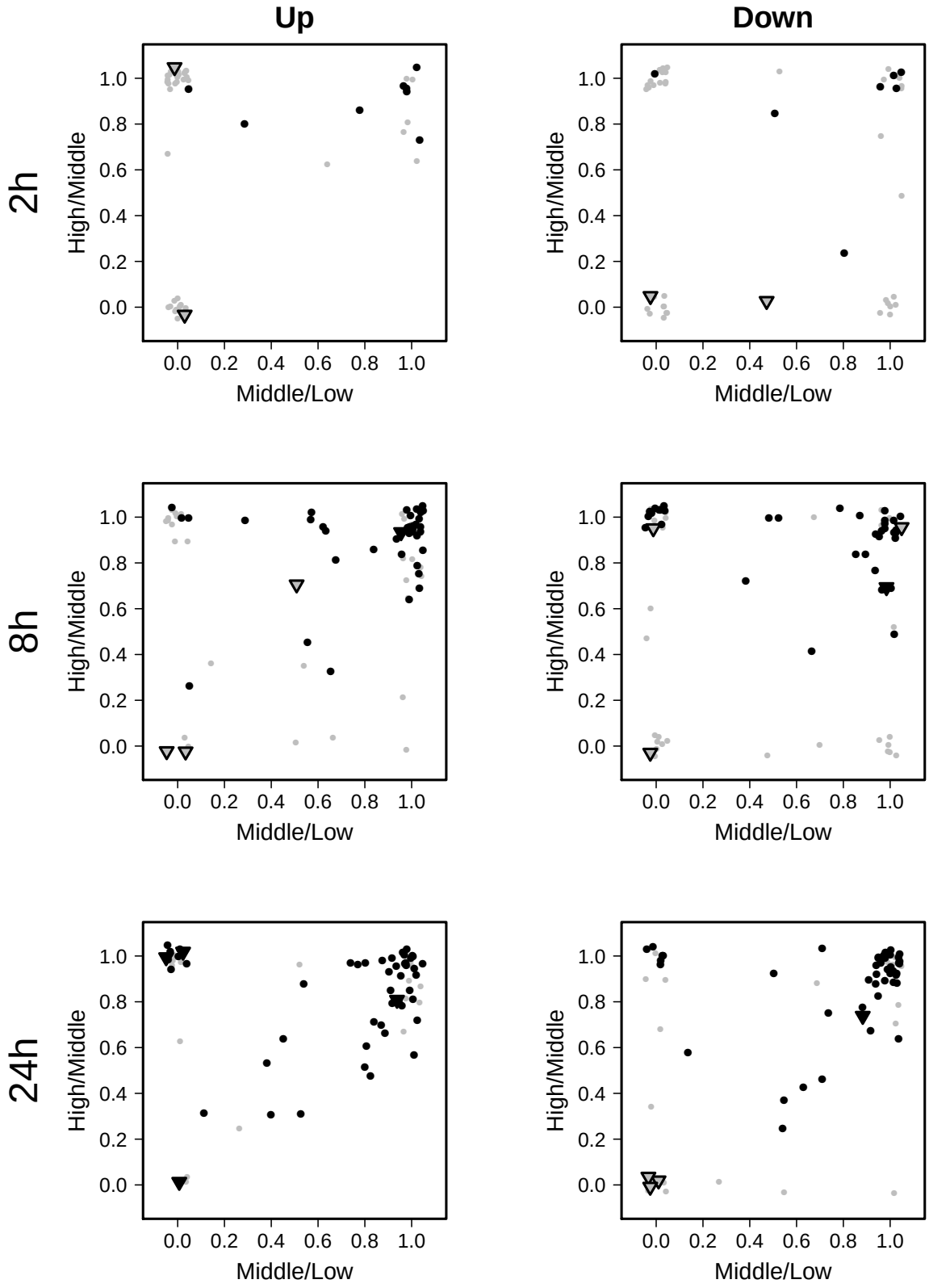


Figure 4.6: *Progression profile indices for all compounds tested at three concentrations (low, middle, high) after three exposure periods (2h, 8h and 24h). A high value means that a high fraction of genes is deregulated exclusively at a higher concentration compared to a respective lower concentration. These calculations were performed comparing the low versus the middle (x-axis) and the middle versus the high (y-axis) concentration. Each symbol represents an individual compound. The triangles represent the later excluded compounds. Black or grey symbols indicate that more than or less than (or equal to) 20 genes, respectively, are deregulated in total.*

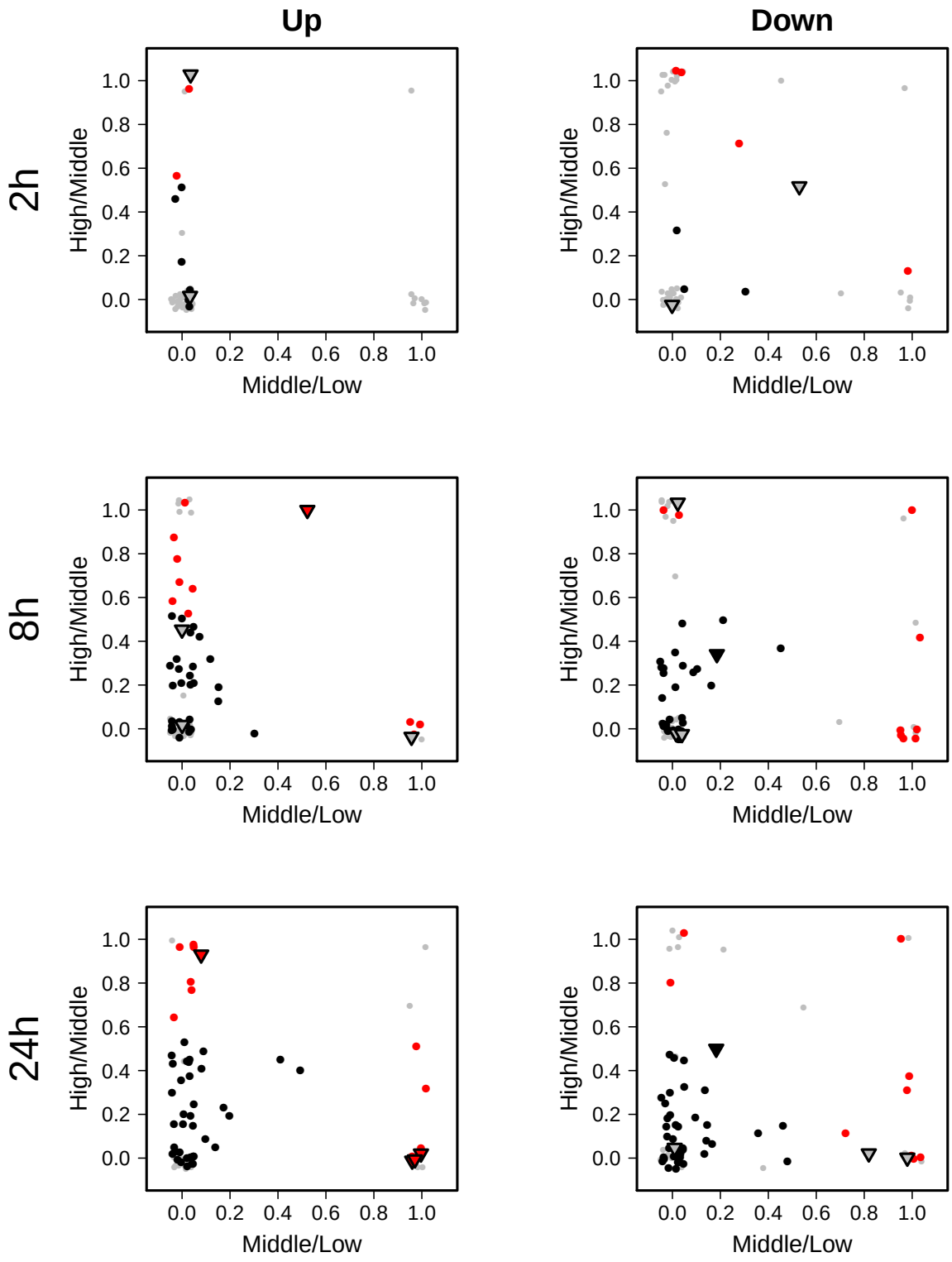


Figure 4.7: *Progression profile error indicator. A high value means that a high fraction of genes is deregulated exclusively at a lower compared to a respective higher concentration. Each symbol represents an individual compound. The triangles represent the excluded compounds. Grey symbols indicate that less than or equal to 20 genes are deregulated in total. Black symbols indicate that more than 20 genes are deregulated in total and both values are smaller than or equal to 0.5. Red symbols indicate that more than 20 genes are deregulated in total and that at least one of the error indicator values is greater than 0.5.*

The individual progression error profiles are used as tools for the identification of compounds with a non-monotonous concentration progression. The application of the profiles leads to the exclusion of five compounds (carbon tetrachloride (CCL4), doxorubicin (DOX), triazolam (TZM), tetracycline (TC), ticlopidine (TCP)). Comparisons across all concentration and time sets have been considered in the exclusion procedure and the combination of all indices have contributed to the exclusion decision. The curated database serves as basis for all further analyses.

In general, only a fraction of the test compounds exhibits strong expression responses, while most of the compounds induce low fold changes. In case of the 24h, high concentration subset, for example, the top 32 (up) and 23 (down) compounds which contribute to the strongest up- and downregulated genes, respectively, yield on average error indicator values below 0.5, independently from the given time period. High error indices are obtained mainly for compounds with weak expression patterns.

4.5 Stereotypic versus compound-specific gene expression responses

The selection value principle is introduced to differentiate between stereotypical and compound-specific responses (Section 3.4). The selection value specifies for a gene the minimum number of compounds that induce an expression change. Mathematically, the compounds are ranked in order of fold change, from high to low fold change (upregulation) or from low to high fold change (downregulation), respectively, and then the compound with rank x determines selection value x (Sv x). The lower the selection value, the more compound-specific the response, and in contrast, the higher the selection value, the more stereotypical the response. Figure 4.8 gives an overview of the observed selection values (Sv 1 to Sv 50) for all concentration and time sets. The selection values are determined for fold changes of at least 1.5-, 2.0- and 3.0-fold. The barplots list for each test condition the number of genes which are upregulated by at least one (Sv 1), two (Sv 2), three (Sv 3),..., and fifty (Sv 50) compounds. The corresponding results for the downregulated genes are shown in Figure C.11 in the Appendix. The number of genes that show a change of at least 1.5-, 2.0- and 3.0-fold for at least one or more compounds increases with increasing time- and concentration progression. The same progression is observed for the downregulated genes. Most of the compounds induce high or low fold changes, respectively, when tested close to cytotoxic concentrations and long incubation periods.

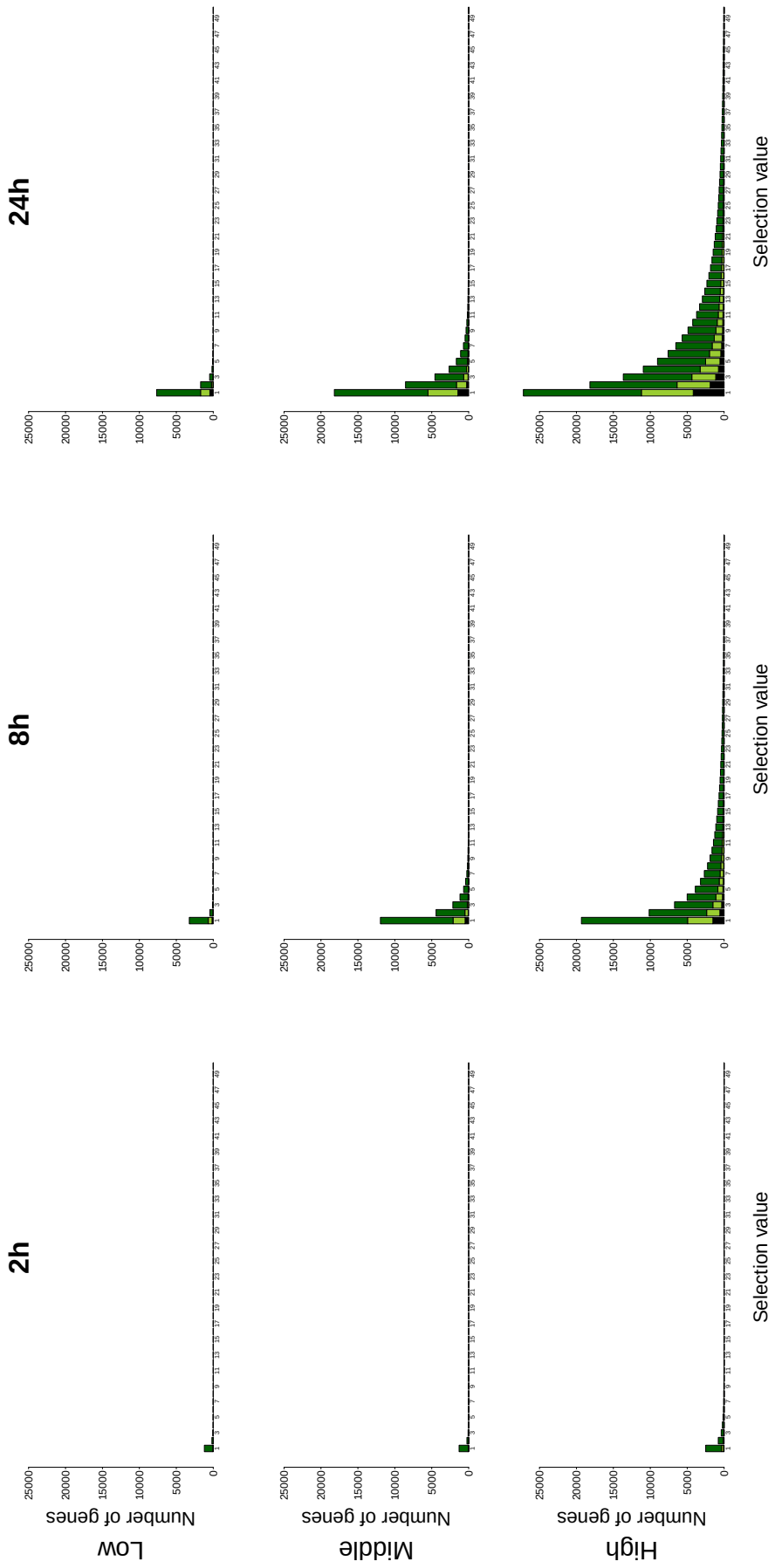


Figure 4.8: *Selection values for the upregulated genes. A selection value of e.g. three means that at least three compounds upregulate (> threefold) the indicated gene. The corresponding analysis for the downregulated genes is shown in Figure C.11 in the Appendix.*

As mentioned above, both response types are of interest, a compound-specific expression response as well as a stereotypical one. While a stereotypical gene expression response is observed for many compounds, a more specific response is only induced by a single or just few compounds. To analyze stereotypical responses, a consensus list is defined including genes influenced by at least 20 compounds (> threefold). These genes are referred to as Sv 20 genes. 20 compounds are considered as sufficient to represent a stereotypical expression pattern as previous results have shown that only 32 of the test compounds induced strong expression responses, while the rest induced only weak reactions. Individual responses are analyzed with Sv 3 genes. Even though Sv 1 genes are more intuitive for studying compound-specific responses, Sv 3 genes are more reliable with respect to higher data stability. Due to the low replicate number ($n = 2$) Sv 3 genes have a lower probability of containing false positive genes than Sv 1 genes. Therefore, Sv 3 genes represent a good compromise between individuality and reliability.

Figure 4.9 gives an overview of the number of genes that are deregulated by at least 1, 3, 5 and 20 compounds. The Venn diagrams summarize the results for the comparison of adjacent selection value genes for the 24h, high concentration subset. Rows differentiate between up- and downregulation. The Sv 20 list compromises 31 up- and 179 downregulated genes. The number increases to 531, 1101 and 4135 for Sv 5, Sv 3 and Sv 1 in case of the upregulated genes and to 857, 1713 and 4479, respectively, in case of the downregulated genes.

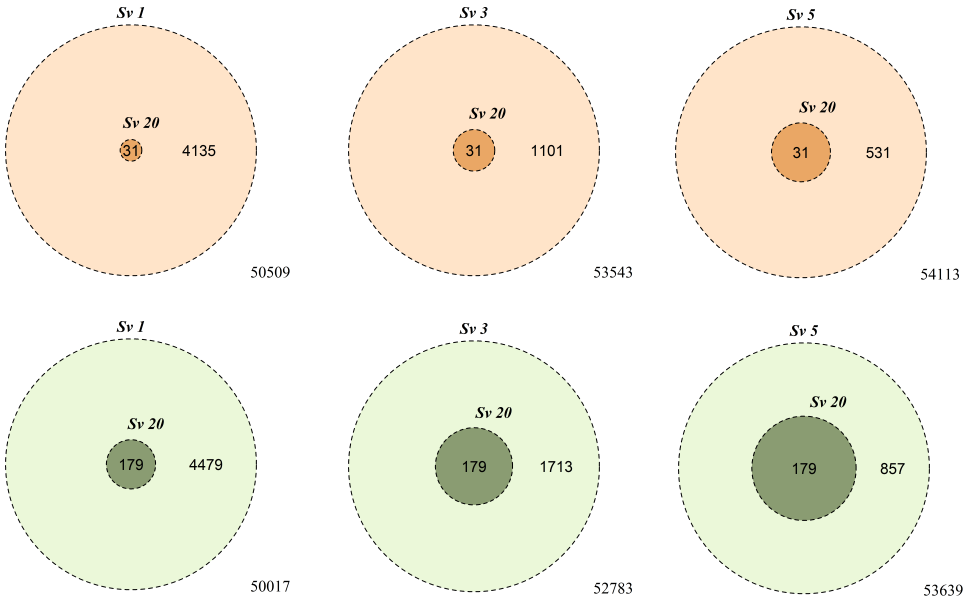


Figure 4.9: *Overview of the numbers of the selection value genes Sv 1, Sv 3, Sv 5 and Sv 20. Red and green color indicate up- and downregulation. For example 31 genes are upregulated (> threefold) by at least 20 compounds (Sv 20).*

4.6 Unstable baseline genes

Genes that are altered solely due to the process of isolation and cultivation stress of cells are called *unstable baseline genes*. These gene expression alterations occur independently from compound exposure, simply as response to stress, and represent therefore a pure *in vitro* artefact. The unstable baseline genes result from the comparison of primary human hepatocytes cultivated in collagen sandwich (CS) for 1, 2, 3, 5, 7, 10 and 14 days ($n = 19$) and freshly isolated primary human hepatocytes ($n = 3$). Genes showing a change of at least threefold at one or more time points are considered as unstable.

It is quite possible that a set of genes is influenced by both, compound exposure and isolation and cultivation stress, as seen in Figure 4.10, which shows the overlaps between the unstable baseline genes (CS) and the Sv 3 and Sv 20 genes, respectively, for the 24h, high concentration subset. 10% to 15% of both response groups, the stereotypical (Sv 20) and compound-specific (Sv 3) alterations, are induced under both, chemical and stress exposure. These genes might give rise to false-positive findings or cloud true findings as a consequence of opposing effects.

4.7 Further analyses

Apart from the presented analyses some further analyses were performed which are not discussed here in detail. To address the question if genes exposed to chemicals *in vitro* respond similarly to *in vivo* exposure, the Sv 20 genes were analyzed with respect to liver disease associated genes. Therefore, the overlap between Sv 20 genes and genes differentially expressed in human liver diseases, such as cirrhosis, hepatocellular cancer or non-alcoholic fatty liver disease was analyzed. The latter data results from a publicly genome-wide data set of human liver tissue. Approximately 20% of the stereotypical stress response genes showed an overlap with liver disease genes. Moreover, the compounds were clustered with respect to genotoxicity, human hepatotoxicity and BSEP inhibiting capacity. Unsupervised clustering of the 24h, high concentration subset compounds resulted in three clusters that could be assigned to biological motifs: proliferation, cytochrome P450 (CYP) and stress response (see Figure C.12 in the Appendix).

Furthermore, GO group analysis was performed for the 31 up- and 179 downregulated Sv 20 genes. Upregulation of stereotypical stress response genes was most associated with xenobiotic metabolism and downregulation with cell cycle progression. GO analysis for the more

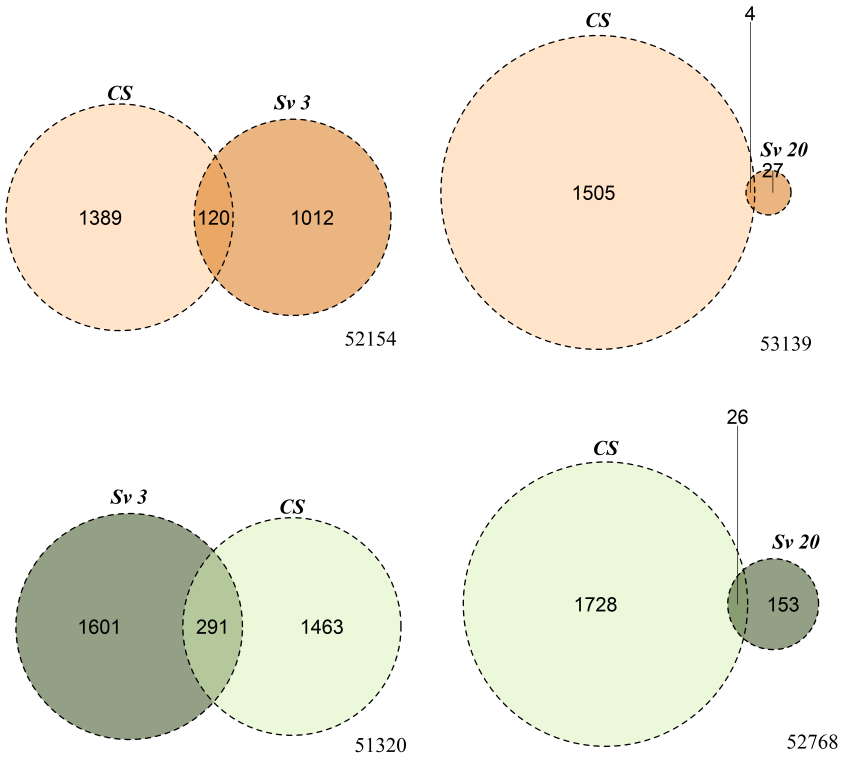


Figure 4.10: *Overlap between unstable baseline genes (CS) and the Sv 20 (Sv 3) genes. Red and green color indicate up- and downregulation. For example, 4 of the 31 Sv 20 genes belong to the unstable baseline genes, meaning that their expression levels are altered by the hepatocyte isolation and cultivation procedure.*

compound-specific expression responses (Sv 3 genes) revealed a wide spectrum of biological motifs including more specific mechanisms.

All in all, the results have been documented in a toxicotranscriptomics directory which is publicly available under http://wiki.toxbank.net/toxicogenomics-map/. The directory provides for each probe set from the HG-U133 Plus 2.0 chip the following information: (1) Is the gene threefold up- or downregulated, and if so, by how many and which compounds? For this purpose, the selection value principle was introduced to answer the question if the change corresponds to a stereotypical gene expression response or rather to a more specific expression response. (2) Is the gene also deregulated in human liver disease? (3) Does the gene belong to a group of unstable baseline genes, i.e. does isolation process and cultivation stress alone cause expression alterations? Moreover, GO group analysis identified about 2000 genes that could be associated with biological functions.

The directory offers a basis for the choice of candidate genes for biomarker evaluation. A long-term goal is to identify biomarkers in *in vitro* systems for the prediction of toxicity *in vivo*.

5 Consensus gene signature of rat hepatocytes tested in *in vitro* and in *in vivo* test systems

The previous chapter has shown one should differentiate between certain concentration ranges: (1) range of tolerance with no expression alterations and range of deregulation (2) with cytotoxic effects and (3) without cytotoxic effects. Cytotoxic concentrations induce cell death events in addition to expression changes. With respect to biomarker detection, it is important to identify the range of concentration in which deregulations are observable. However, expression alterations rarely occur individually, but rather in sets of highly correlated genes. The concentration-dependent analyses of the Open TG-GATEs data set of cultivated human hepatocytes showed a stereotypical expression response in a subset of genes which was induced by many compounds. However, the direct comparison of chemically-induced genes between cultivated human hepatocytes and human liver samples *in vivo* is ethically not justifiable. To investigate whether *in vitro* deregulated genes respond similarly under conditions of *in vivo* exposure, rat data is used for analysis. The following study targets the identification of consensus genes, which show a comparable stress response in rat hepatocytes. Compound-specific gene alterations are analyzed with respect to similar patterns in *in vivo* and in *in vitro* experiments. For this purpose, the NRW and the TG-GATEs database (Sections 2.4.1 and 2.4.3), comprising gene expression data of rat hepatocytes cultured *in vitro* and rat livers exposed *in vivo*, are used. In this chapter a guideline for the identification of consensus genes is developed, evaluated and discussed.

5.1 Data structure of the NRW database

The NRW database includes Affymetrix files of cultured rat hepatocytes incubated with 30 compounds that have been tested in cultured rat hepatocytes and rat liver cells *in vivo*. *In vitro* a

24h exposure period was tested using three concentrations (Low, Middle, High) and *in vivo* four time periods using one concentration level were analyzed (1 day, 3 days, 7 days and 14 days). For more details on the data the reader is referred to Section 2.4. Tables B.5-B.6 in the Appendix provide a compound-specific overview of the test compounds.

To display the transcriptome data structure across *in vivo* and *in vitro* samples a principle component analysis for all concentration and time sets was performed, based on the 100 probe sets with the highest variance across all samples. Figure 5.1 shows the corresponding plot of the first two principal components which results in two main clusters. The two clusters are separated by the first component - the *in vitro* samples cluster to the left and move along the second principal component, and, the *in vivo* samples cluster to the right-hand side of the plot. These results confirm the findings of Schug et al. (2013) who have shown that genes deregulated by compound exposure *in vitro* respond differently under conditions of *in vivo* exposure.

Moreover, the PCA plot shows that the cluster of *in vitro* samples is subdivided into several sub-clusters, the *in vivo* samples cluster closely together with the exposed samples to the right of the controls. In contrast, the *in vitro* data shows no distinct separation between compound exposed samples and controls. Merely, the samples at the beginning of the incubation period (T0) cluster clearly apart from the samples that were tested after 24h of incubation. This underlines that expression alterations already occur when hepatocytes are solely isolated and cultivated (Grinberg et al., 2014).

To assess the data quality of the NRW database the two test systems were analyzed separately with respect to batch effects, reproducibility across replicates and implausible concentration progressions. For this, the pipeline introduced in Chapter 4 was applied to curate the NRW database. To understand the data structure, a PCA analysis has been performed separately for each concentration and time set. Figure 5.2 shows for the data of the *in vitro* experiments the PCA plot for the low, middle and high concentration set. Figure 5.3 illustrates the structure of the *in vivo* data for the time points 1 day, 3 days, 7 days and 14 days. Columns refer to the individual test conditions and rows to the different analysis steps. Panel A provides an overview of all samples tested at the indicated condition. Panel B illustrates the degree of variability by connecting lines between triplicates. *In vivo*, the distances between sample triplicates is relatively small in comparison to the distances between the corresponding sample triplicates *in vitro* which exhibit a much higher variability. To quantify the data variability, the Euclidean distances between all pairs of triplicates were calculated and illustrated in Figure 5.4. The distances vary among the *in vitro* experiments much more than among the *in vivo* experiments.

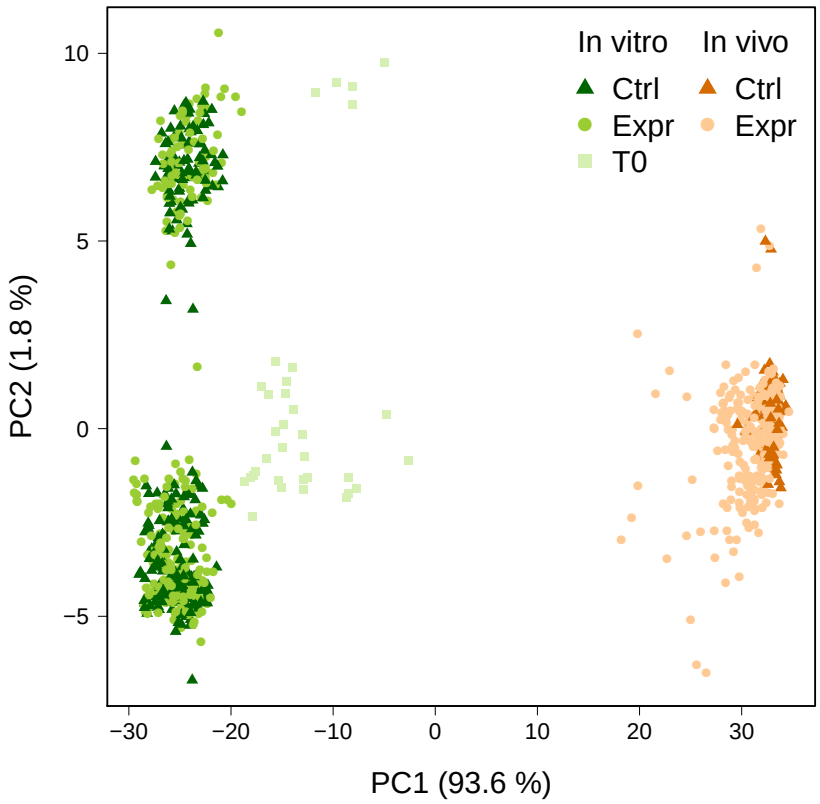


Figure 5.1: *Principle component analysis of gene expression data obtained from cultured rat primary hepatocytes (in vitro) and from rat liver samples (in vivo) after incubation with 29 (in vitro) and 30 (in vivo) chemicals. The PCA plot is based on the 100 probe sets with highest variance across all samples and was generated to display the transcriptome data structure across in vivo and in vitro replicates. Each point represents one experiment. In vitro hepatocytes were harvested at the beginning of the exposure period and 24h later after exposure to the test compounds (exposed) or solvent (controls).*

Compounds with distances outside the tolerated variability range, which is determined by the 95%-quantile of the distribution, have been classified as outlier candidates. *In vitro*, up to three triplicate sample pairs exceed the cutoff of the 5% largest observed distances. *In vivo*, at each incubation time point, two compound exposed triplicate pairs have been identified as outliers, whereas all control triplicate-distances lie within the tolerance range.

For simplicity, mean values of the triplicates were calculated (Figures 5.2 C and 5.3 C, respectively). Connecting lines between controls and corresponding compound exposed samples (panel D) shows that the compound effects are larger *in vivo* than *in vitro*. In the last step, controls were subtracted from the corresponding compound exposed samples (panel E). Panel E shows that the above described clusters have been removed by this procedure. As shown before (Chapter 4), simple control subtraction is sufficient to remove batch effects.

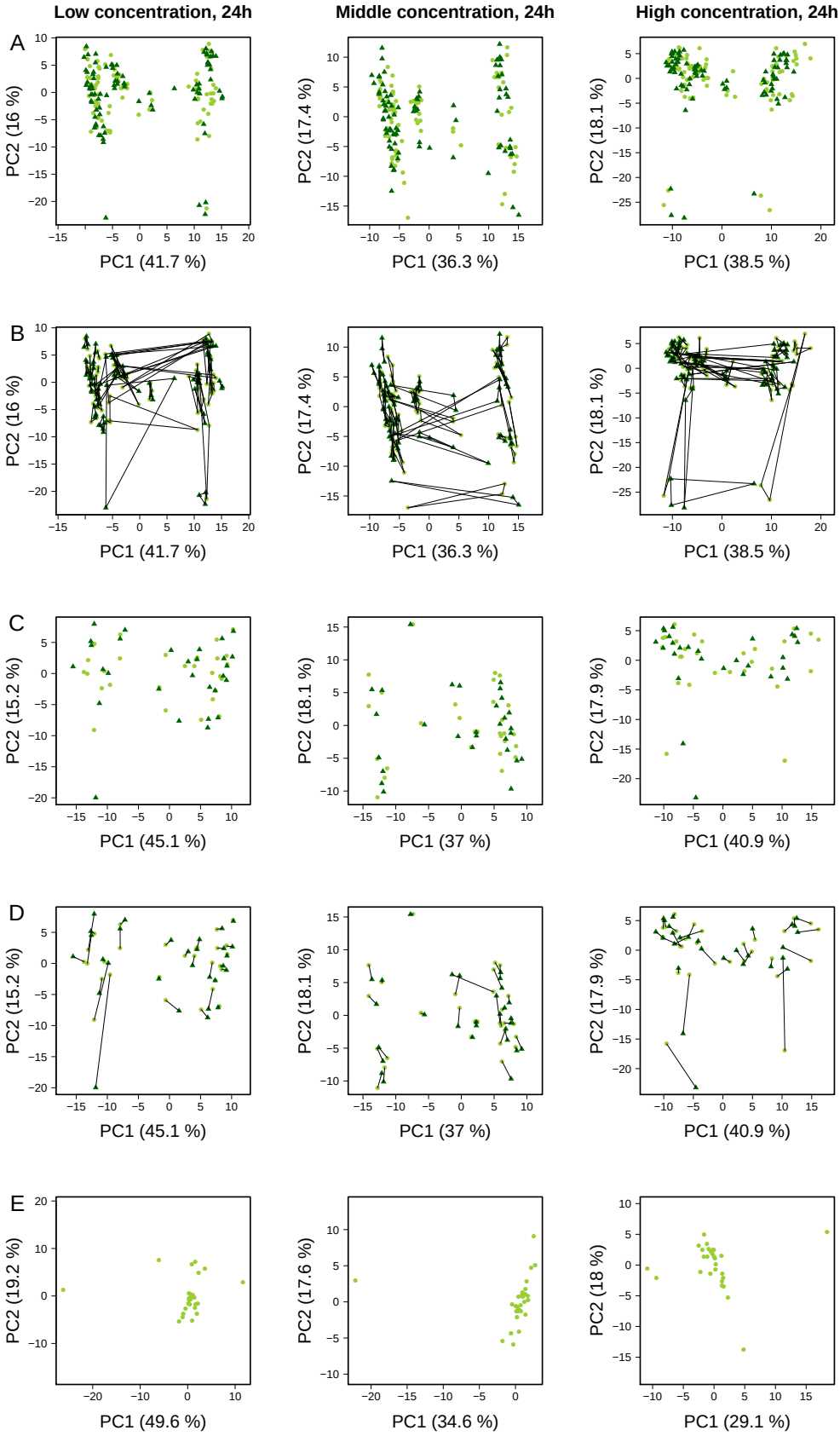


Figure 5.2: *NRW (in vitro): Data of the low, middle and high concentration and the incubation time point 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

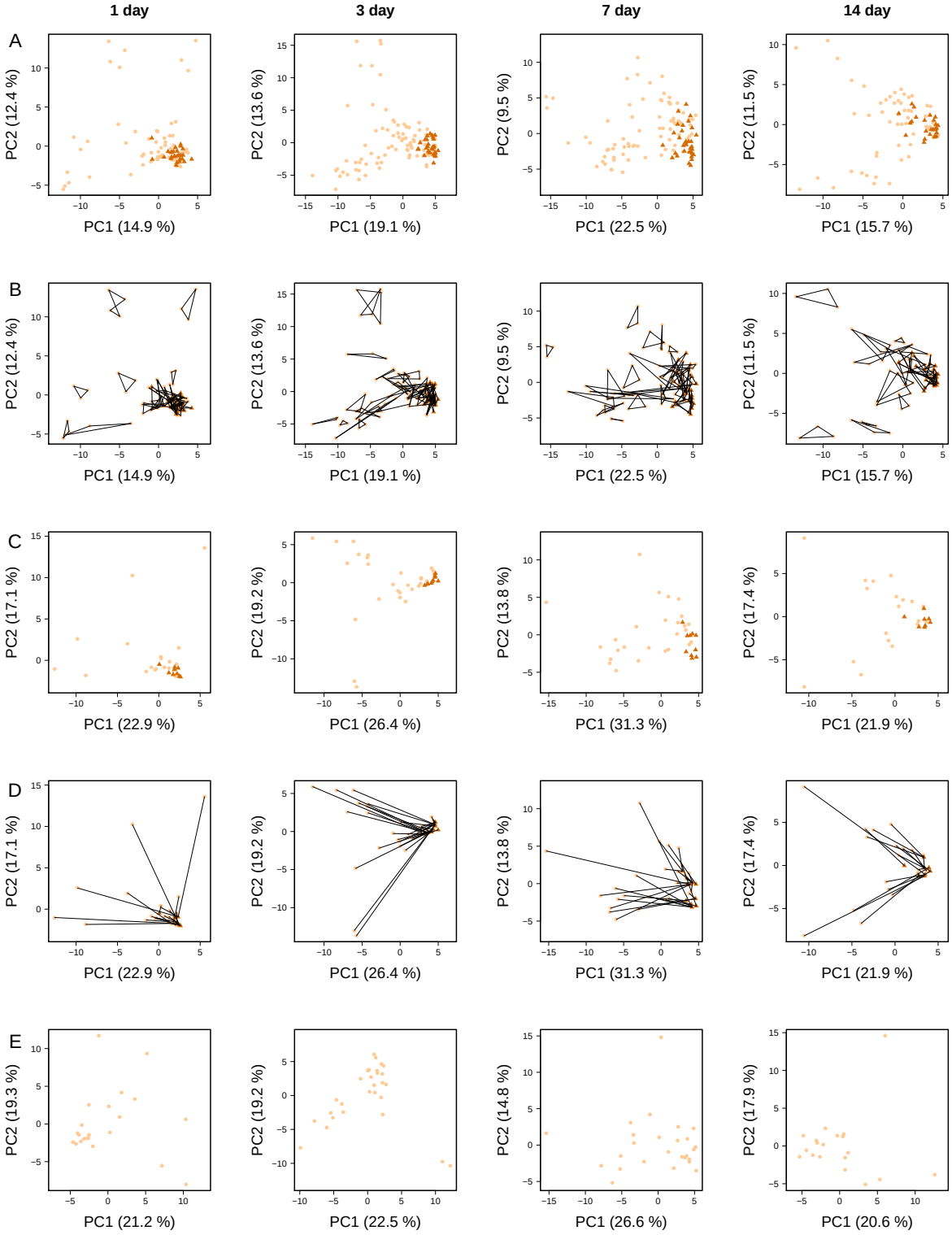


Figure 5.3: *NRW (in vivo): Data of the incubation time points 1 day, 3 days, 7 days and 14 days. A. Overview of all samples and replicates. The dark and light orange symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark orange) and corresponding compound exposed samples (light orange). E. Subtraction of the controls from the corresponding compound exposed samples.*

In vitro, the experiments were organised in 6-well-dishes, i.e. 3-wells were incubated with the test compound and 3-wells were used as controls. For this reason, 6-well dish-matched controls were subtracted. In contrast, control subtraction *in vivo* proved to be more challenging. As the treatment of animals was performed in different experimental series, this factor might be, besides the factor *exposure period*, decisive in the context of batch removal. To decide how to subtract controls, an analysis of variance (ANOVA) with two main effects was performed to test whether the model parameters *experimental series* and *exposure period* have an influence on gene expression. The analysis was performed gene-wise. The resulting p-values for the two model parameters are shown in the frequency distribution in Figure C.13 in the Appendix. The smaller the p-value, the stronger is the influence of the model parameter. Since most of the genes yield p-values around zero, both parameters should be considered in the process of subtraction. Therefore, *experimental series*- and *exposure period*-matched controls were subtracted from the corresponding treated samples.

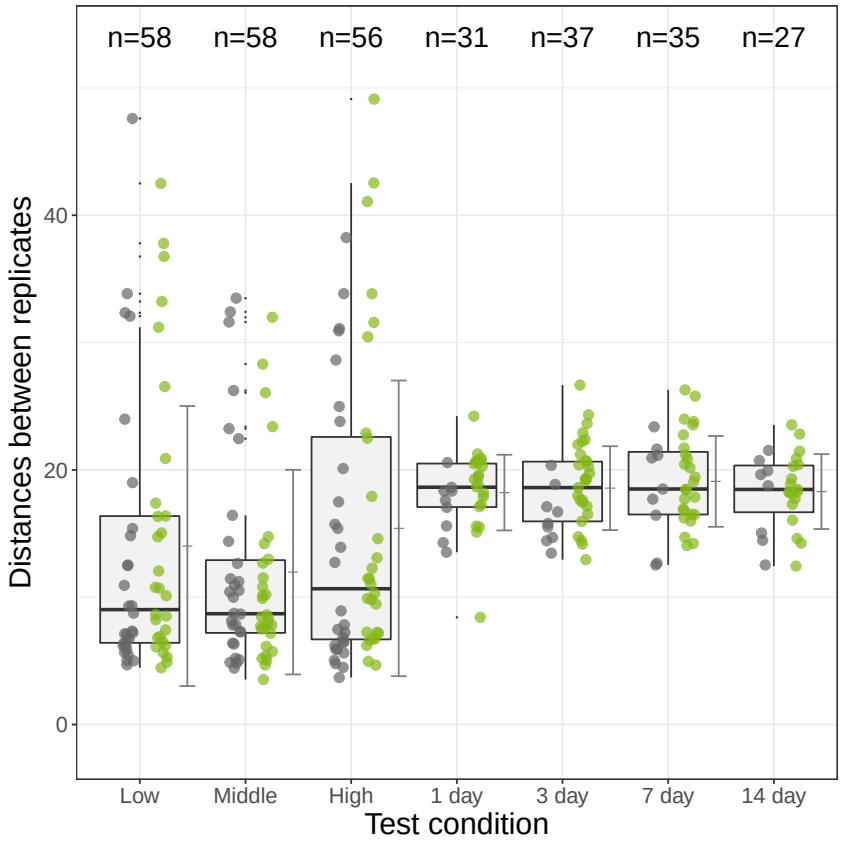


Figure 5.4: *Reproducibility between replicates. Boxplots of the Euclidean distance between all pairs of triplicates for all test conditions. The grey and the green points illustrate the distances between the control and compound replicate samples, respectively.*

The control-subtracted expression values ($\log_2$-fold changes) serve as base data for all further analyses. To obtain an overview of the underlying compound effects, the number of gene deregulations induced by compounds was counted. Changes of at least 1.5-, 2.0- and 3.0-fold were reported separately for each test system and condition. The barplots in Figure 5.5 and Figure 5.6 summarize the cumulated results for the *in vitro* test system and for the *in vivo* experiments, respectively. The upper and lower panels of the figures refer to the up- and downregulated genes, respectively. *In vitro*, the number of induced genes has increased concentration-dependent with most of the genes being deregulated at the highest tested concentration. However, it should be noted, that 8 of 29 compounds (27.5%) have induced only low fold changes (< 1.5-fold). Due to weak expression changes, these compounds have been characterized as *low profile compounds*. This group of compounds deregulates less than 30 genes in total. The exclusivity analysis in Figure 5.7 confirms the aforementioned results. It has identified the 100 strongest deregulated genes across all compounds (top-ranking genes with the highest/lowest fold change values) and has assigned each of them to the compound with the most extreme fold change. These compounds are *match-winners* per gene. Low profile compounds are scored with little to no genes in the exclusivity analysis. In contrast, the *in vivo* data has shown no clear relationship between exposure period and number of gene deregulations (Figure 5.6). Four of 30 compounds (13.3%) have induced less than 15 genes *in vivo* (> 1.5 with adjusted $p \leq 0.01$) and have been consequently classified as low profile compounds. In the exclusivity analysis in Figure 5.8 these compounds have scores not higher than one. Together, the *in vitro* and *in vivo* experiments comprise 11 low profile compounds (AA, Aap, AfB1, CFX, DCB, ETH, MDA, Mcarb, Nif, Praz, Prop). The cut-off criterion for being named low profile compound depends on the size of the used chip, which includes in case of the *in vitro* experiments 31 099 probe sets and in case of the liver samples 15 923 probe sets. The minimum number of deregulated genes, 30 *in vitro* and 15 *in vivo*, represents approximately 1% of the measured probe sets in total.

The results of the foregoing analyses have shown that the data reliability of the *in vivo* experiments is much higher than the reliability of the *in vitro* experiments (Figures 5.2-5.4). For this reason, the *in vivo* data are considered the gold standard to which the *in vitro* results are compared to. But before a direct comparison of differentially expressed genes is performed between the two test systems, the *in vitro* data is further analyzed under the particular aspect of data plausibility. To do so, a concentration-dependent analysis of gene alterations is performed.

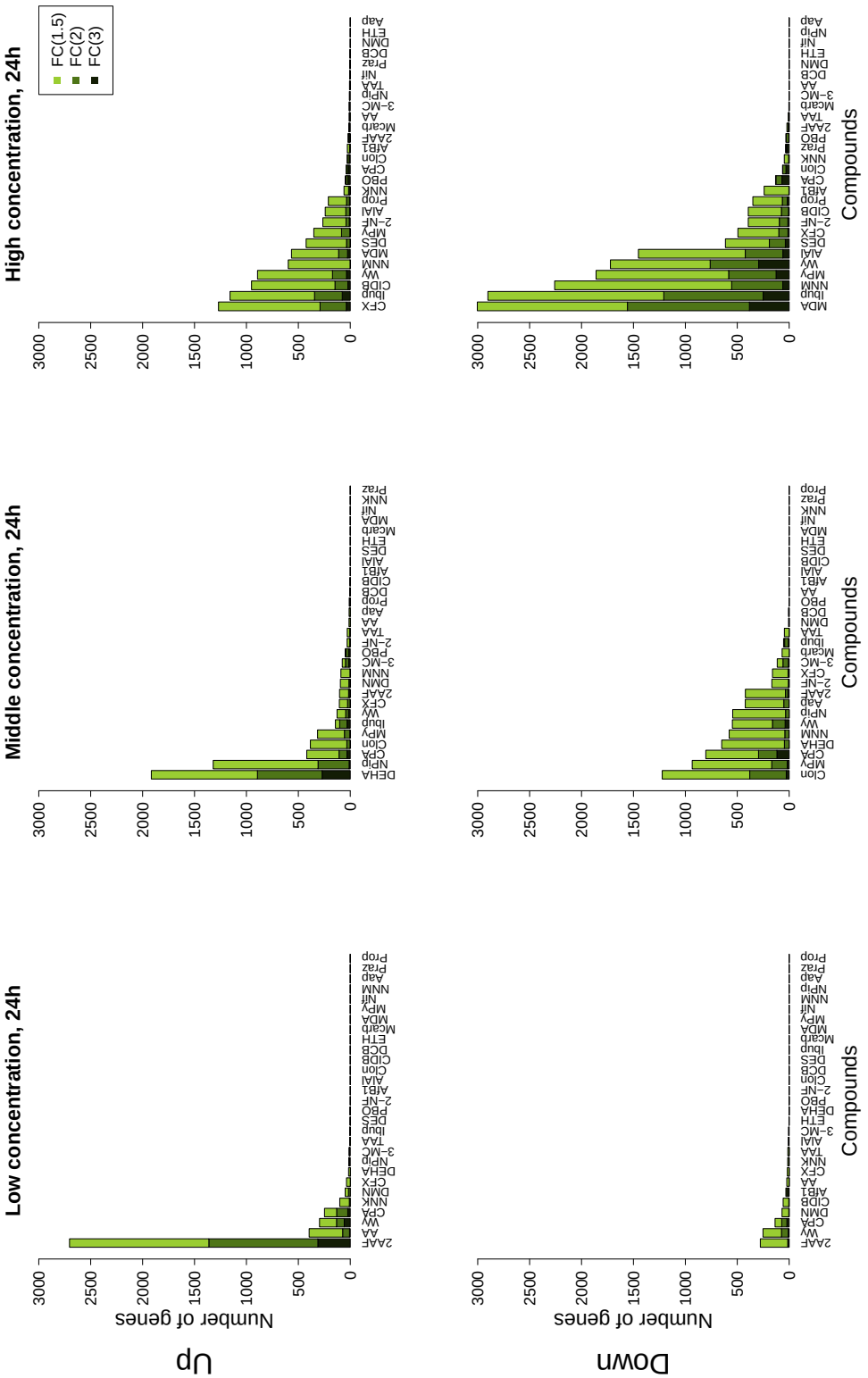


Figure 5.5: *NRW (in vitro): Number of significantly deregulated genes. On the x-axis all chemicals that were tested at the indicated concentration for 24h are listed. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. Light green: more than 1.5-fold deregulated; middle green: more than twofold deregulated; dark green: more than threefold deregulated.*

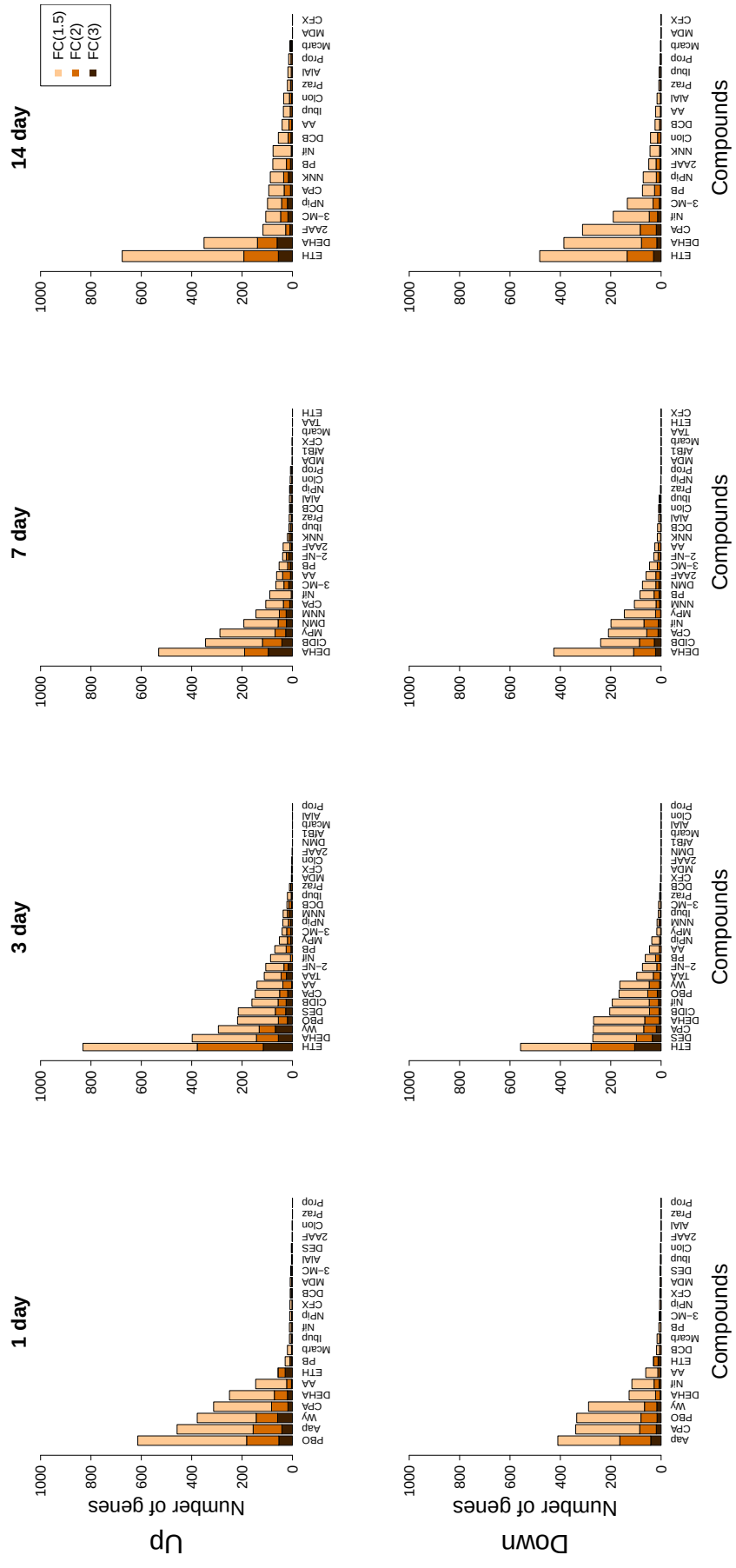


Figure 5.6: *NRW (in vivo): Number of significantly deregulated genes. On the x-axis all chemicals that were tested at the indicated concentration for 24h are listed. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. Light orange: more than 1.5-fold deregulated; middle orange: more than twofold deregulated; dark orange: more than threefold deregulated.*

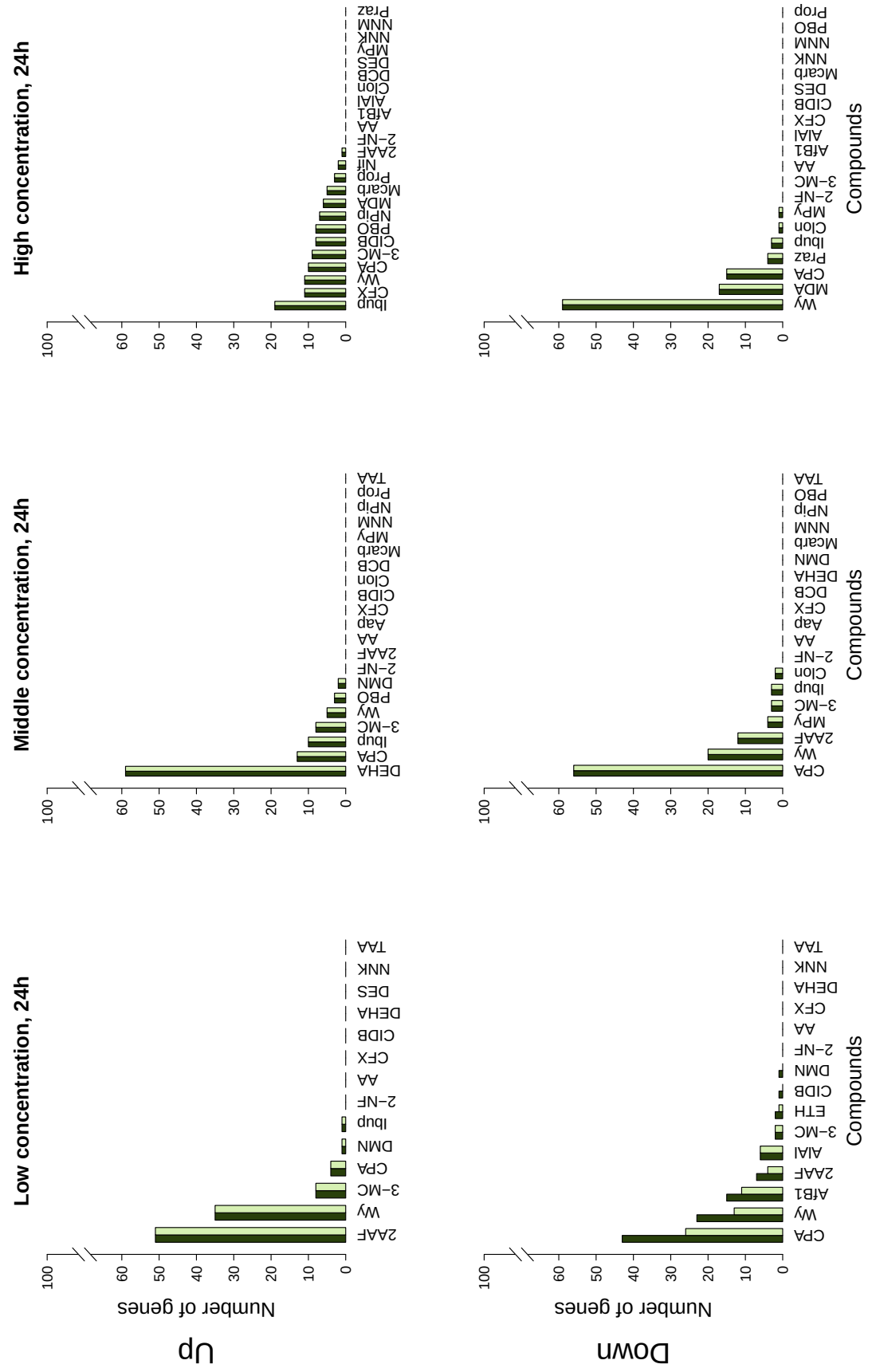


Figure 5.7: *NRW (in vitro): Exclusivity analysis of the up- and downregulated genes (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light-colored barplots indicate whether the top ranking genes meet the criteria for significance (> 1.5-fold with adjusted $p \leq 0.01$). The x-axis lists only those compounds that contributed at least once to the strongest genes in the ranking of all probe sets ($n = 31\,099$).*

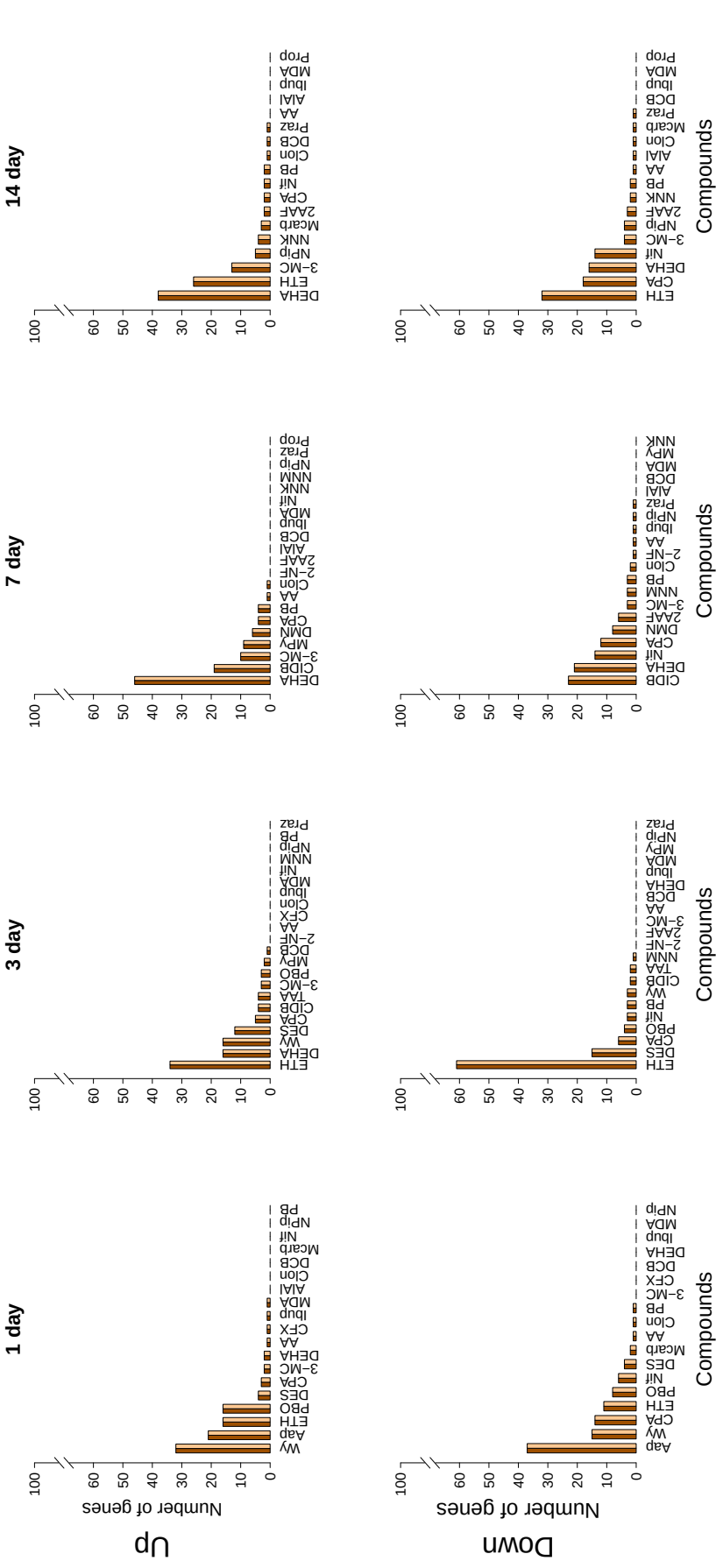


Figure 5.8: *NRW (in vivo): Exclusivity analysis of the up- and downregulated genes (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light-colored barplots indicate whether the top ranking genes meet the criteria for significance (> 1.5-fold with adjusted $p \leq 0.01$). The x-axis lists only those compounds that contributed at least once to the strongest genes in the ranking of all probe sets ($n = 15\,923$).*

The results are illustrated by Venn diagrams, which show the overlap between the low, middle and high concentration. The concept is demonstrated for four examples in Figure 5.9. Genes with a fold change of at least 1.5 and a false discovery rate (FDR) adjusted p-value smaller than 0.01 are considered differentially expressed. Orange and green colored Venn diagrams count significantly up- and downregulated genes. 2-NF and CFX represent compounds with a plausible concentration progression, where with each concentration step the number of additional genes increases. The compound AIAI exemplifies another frequently observed constellation: For such a compound, genes are only deregulated with the highest tested concentration. An implausible concentration progression can be observed for DEHA, where genes are up- or downregulated with the middle, but not with the low or high concentration. Such non-monotonous concentration progressions may be indicative of experimental errors or sample mix-ups. Due to such limitations imposed by the quality of the used data, the selection value principle, which assigns each gene the number of compounds that up- or downregulate this gene after exposure, has been introduced in Chapter 4. Therein, the use of selection value 3 genes (Sv 3) is recommended for the assessment of compound-specific responses in order to ensure a higher data stability.

The Venn diagrams for all compounds are given in the Appendix. The Venn diagrams for the concentration-dependent cultivation experiments are provided in Figure C.24. Those for the time-dependent *in vivo* experiments are shown in Figure C.25. Since four time points were measured *in vivo*, pairwise comparisons of adjacent time periods have been performed for the *in vivo* experiments.

The selection value analysis in Figure 5.10 provides selection values in the range from 1 to 15 indicating that at most 15 compounds have induced gene alterations. 15 compounds represent a relatively large fraction, because only 19 compounds belong to the group of high profile compounds (deregulating more than 15 or 30 genes *in vivo* or *in vitro*, respectively). But the number of deregulated genes decreases enormously from the selection value 6 on. The number of Sv 3 genes varies between 200 and 500 genes for the two test systems.

5.2 Consensus signature

Since it is of high interest to figure out whether a gene chemically-induced in *in vitro* is also induced in *in vivo*, a comparative analysis of genes differentially expressed in both test systems is performed. For this purpose, the overlap between differentially expressed liver genes and chemically deregulated genes *in vitro* is analyzed. To enable a direct comparison of genes

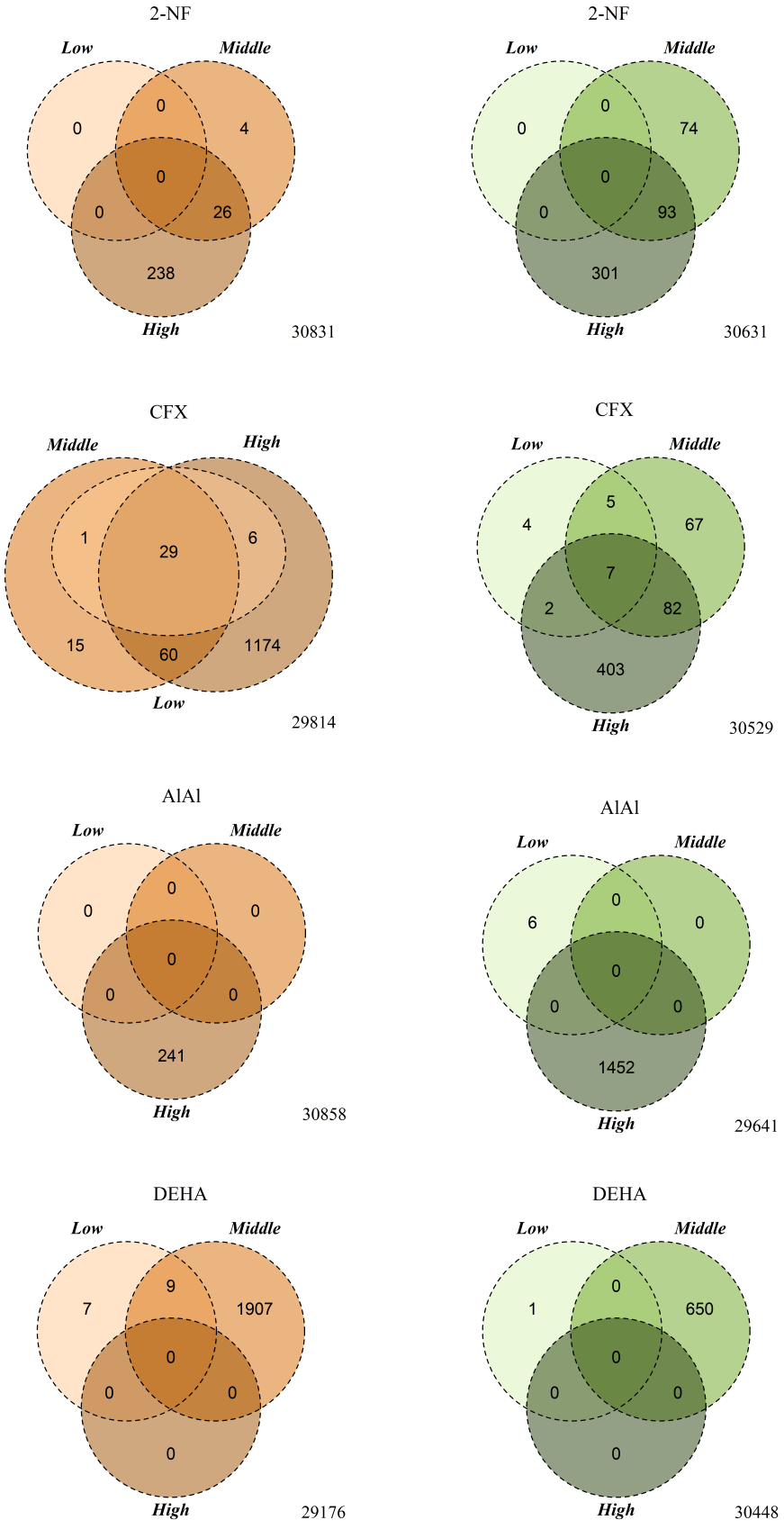


Figure 5.9: *Concentration dependency in the NRW in vitro database. Only representative examples are shown (all compounds are shown in Figure C.24 in the Appendix). Orange colored Venn diagrams show the overlap between genes that are upregulated at the low, middle and high concentrations (> 1.5-fold with adjusted $p \leq 0.01$); green colored Venn diagrams summarize the downregulated genes (< $\frac{2}{3}$-fold with adjusted $p \leq 0.01$).*

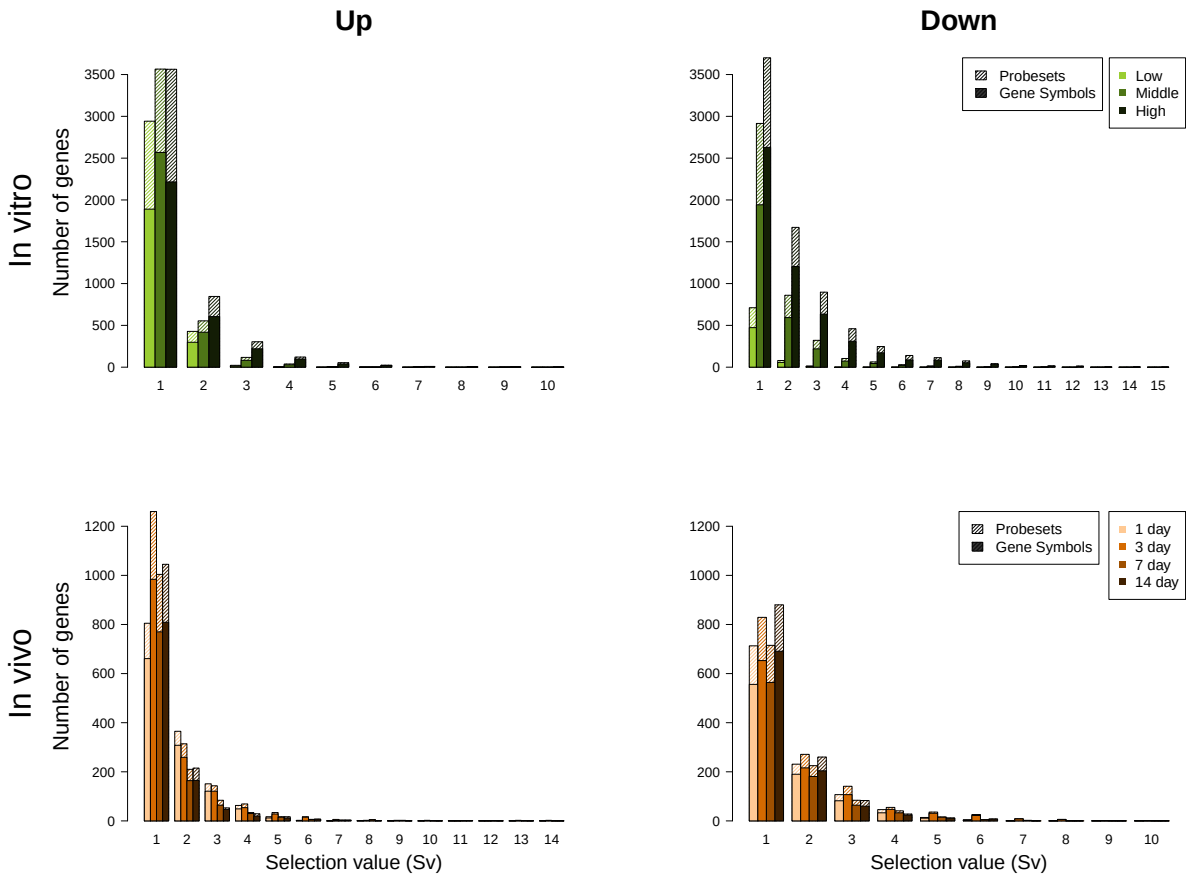


Figure 5.10: *Selection values for the up- and downregulated genes. The upper panel shows the results for the in vitro experiments and the lower panel shows the results for the rat liver samples. A selection value of e.g. three means that at least three compounds up- or downregulate the indicated gene (> 1.5-fold with adjusted $p \leq 0.01$).*

between both test systems, which were analyzed on different chips, the analysis is restricted to those genes that have been measured on the *rae230a*-GeneChip, which was used for the *in vivo* experiments. This results in 10 044 genes for the NRW database. To improve the robustness of the results, the analysis is performed on the basis of Sv 3 genes. First, a consensus signature of Sv 3 genes is compiled for the *in vitro* and *in vivo* experiments. The consensus Sv 3 list combines the Sv 3 gene lists of the individual test conditions which have been predefined for the *in vitro* concentrations and *in vivo* time points. That means, the *in vitro* Sv 3 consensus list comprises the genes that have been up- or downregulated by at least three compounds after exposure with a low, middle or high concentration. The *in vivo* Sv 3 consensus list summerizes those genes that have been deregulated by at least three compounds after 1 day, 3 days, 7 days or 14 days of exposure. For generating the lists, the fold change cut-off was set to 1.5 and the p-value cut-off to 0.01 after false discovery adjustment for multiple testing. This results in 369 up- and 1072 downregulated consensus genes for the *in vitro* experiments and in 354 up- and 326 downregulated consensus

genes for the *in vivo* experiments. The overlap covers 77 up- and 98 downregulated genes (see Figure 5.11). The genes in overlap are further named *consensus Sv 3 NRW genes*. To quantify to which degree genes in the *in vitro-in vivo* overlap are overrepresented, the overlap ratio is calculated (see Section 3.4). The principle of the overlap ratio is illustrated in Figure 5.11 (upper left panel). An overlap ratio of 1.0 corresponds to a randomly expected overlap. The overlap ratio for the upregulated genes is 5.9 ($p < 0.001$) indicating that 5.9-fold more genes are in the overlap than randomly expected. The downregulated genes yield an overlap ratio of 2.8 ($p < 0.001$). The p-values result from the Fisher's exact test. Even though the overlap is larger then expected under stochastic independence, a relatively high fraction of genes is test system-specific.

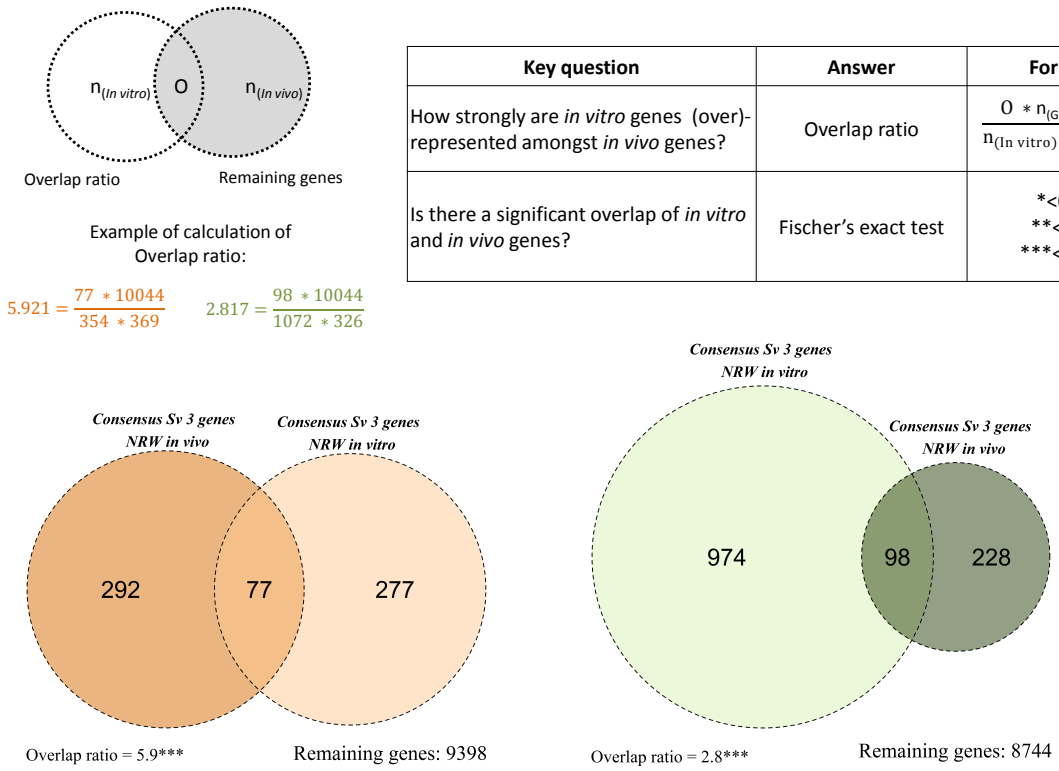


Figure 5.11: *Overlap of genes deregulated by the same compounds in in vitro and in in vivo. The Venn diagram in the upper left panel illustrates the principle of the overlap ratio. The lower panels show the overlap between the in vitro and in vivo consensus Sv 3 genes for the NRW data set. The overlap ratio is calculated by the formula given in the table in the upper right panel.*

Figure 5.12 illustrates for the 77 up- and 98 downregulated *consensus Sv 3 NRW genes* the corresponding overall selection values (upper and lower panel). The overall selection value provides for a gene the number of compounds that deregulate this gene under at least one of the test conditions (repeatedly occurring compounds are counted only once).

Figure 5.12: *NRW: Overall selection values of the consensus Sv 3 NRW genes (genes that are deregulated by at least three compounds in both test systems, in vitro and in vivo, for at least one test condition). Compounds are counted once. The upper panel shows the barplots for the 77 upregulated consensus genes and the lower panel the corresponding ones for the 98 downregulated consensus genes. The shaded barplots indicate the number of individual in vitro and in vivo compounds and the fully colored barplots indicate the number of common compounds.*

Figure 5.12 shows both, the number of compounds common to both test systems and the number of compounds specific for the two test systems. Genes are ordered by their relevance with respect to their selection value in order to find a subset of as few genes as necessary to depict a compound sensitivity as large as possible. Up to nine compounds could be identified that induce common gene alterations *in vitro* and *in vivo*.

In contrast, Figure 5.13 represents the number of genes for the 30 test compounds that are induced by these compounds under at least one of the test conditions, whereas the analysis is restricted to the *consensus Sv 3 NRW genes*. If a gene is deregulated under several conditions, the gene is counted only once. Both common and distinct gene deregulations are shown. Compounds are listed alphabetically on the x-axis.

5.3 Data structure of the TG-GATEs database

To validate the results of the NRW database, another data set is analyzed, the TG-GATEs data set which compiles Affymetrix files of rat liver cells tested *in vivo* and *in vitro*. For the *in vitro* data, liver tissue was used to isolate and cultivate primary rat hepatocytes for 2h, 8h and 24h. The cultured rat hepatocytes were tested with several compounds (up to 145 compounds depending on the test condition) using a low, middle and high concentration. For the *in vivo* data rat liver samples were used and sacrificed after 3h, 6h, 24h, 4 days, 8 days, 15 days and 29 days after exposure. Up to 155 compounds (depending on the test condition) were tested with three different concentration levels. For the analysis only subsets of the test conditions are used (i.e. particular combinations of concentration and time sets). *In vitro*, only the data of the low (145 compounds), middle (140 compounds) and high concentration (138 compounds) after 24h of incubation is used, and, *in vivo* only the highest tested concentration after 24h of exposure is investigated. All other test conditions are excluded from the analysis due to non-removable batch effects.

The structure and quality of the data is assessed along the curation pipeline which was introduced in Chapter 4. An overview of all cultivated hepatocytes is given in Figure C.14 and of all hepatocytes in liver in Figure C.15 (Appendix). PCA analysis has been performed, based on the 100 probe sets with highest variance across all compounds, to visualize the *in vitro* and *in vivo* response to chemicals. Figures C.16-C.18 in the Appendix illustrate the number of altered genes for the concentration-dependent *in vitro* experiments. Changes of at least 1.5-, 2.0- and 3.0-fold have been counted. Figure C.19 in the Appendix shows the corresponding barplots for

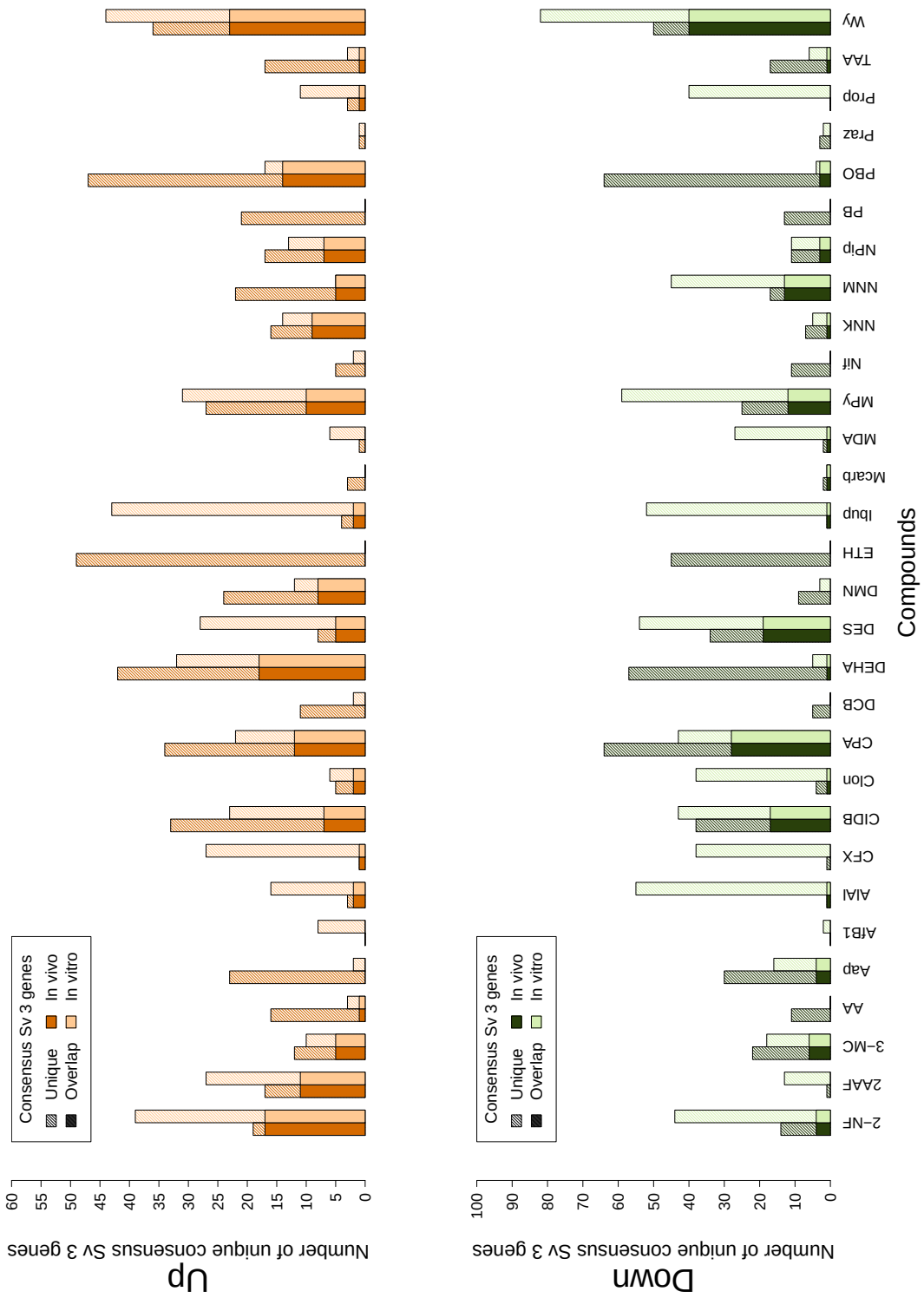


Figure 5.13: *NRW: Number of consensus Sv 3 NRW genes that are deregulated by the indicated compounds in vitro and in vivo for at least one test condition (shaded barplots). The fully colored barplots indicate the number of common consensus genes, i.e. the Sv 3 genes that are deregulated in both test systems, the in vitro and in vivo test system. The upper panel shows the barplots for the upregulated consensus genes and the lower panel shows the barplots for the downregulated consensus genes.*

the *in vivo* data. The analysis reveals that a relatively high fraction of compounds induce only low fold changes. The exclusivity analysis shows that only a few compounds contribute to the 100 strongest genes, while most of the compounds have no effect (see Figures C.20-C.23 in Appendix). The fact that the *in vitro* experiments were tested with only two replicates limits the validity of statistical tests. Therefore genes are considered differentially expressed when the mean difference to controls is at least threefold. The liver hepatocytes, on the contrary, were tested with three replicates which enables the detection of gene alterations by means of hypothesis testing. For this, the `Limma` t-test is applied. The p-values are adjusted for multiple testing by controlling the false discovery rate (FDR) according to the Benjamini-Hochberg (BH) procedure. Changes of at least threefold with adjusted p-value smaller than 0.01 are defined as significantly deregulated. Selection values between 1 and 60 are observed for the *in vitro* and *in vivo* subsets (see Figures C.26-C.27 in Appendix). In comparison to the reactions *in vitro* less compound effects are seen *in vivo*.

Similar to the NRW data set, the selection value concept has been applied to the TG-GATEs database to identify compound-specific gene inductions. For the aforementioned reasons (more reliable results), a consensus signature consisting of Sv 3 genes is compiled for the *in vitro* and *in vivo* comparisons. Figure 5.14 (upper panel) reveals that the consensus Sv 3 *in vitro* list comprises 574 up- and 1210 downregulated genes (across all concentration subsets) and the consensus Sv 3 *in vivo* list for the 24h, high concentration subset summarizes those genes which have been induced by at least three compounds (513 up, 414 down). The overlap of the consensus Sv 3 genes between the *in vitro* and *in vivo* test systems is 140 (up) and 186 (down), respectively. Hereinafter these genes are referred to as *consensus Sv 3 TGD genes*. Figure C.28 in the Appendix summarizes the corresponding overall selection values for the *consensus Sv 3 TGD genes*, i.e. compound inductions across all test conditions are counted. For the compounds studied in the TGD database, Figure C.29 (Appendix) provides an overview of the number of *consensus Sv 3 TGD genes* which are altered by them.

Furthermore, the overlap between the NRW and TGD database is analyzed with respect to similar expression patterns. The Venn diagrams in Figure 5.14 (lower panel) count 23 up- and 22 downregulated genes, respectively, in the overlap. This corresponds to 29.8% (22.4%) of the *consensus Sv 3 NRW genes*. The corresponding overlap ratios indicate a relatively high degree of consensus between the two databases, even though only five compounds are common in both data sets.

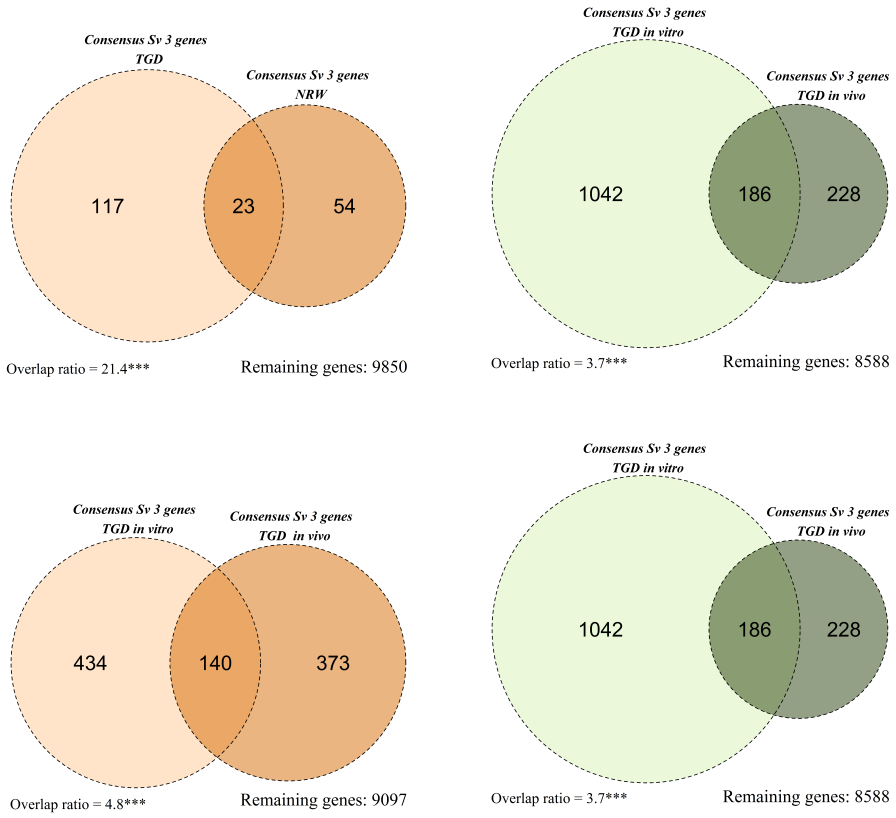


Figure 5.14: *Overlap between the in vitro and in vivo consensus Sv 3 genes. Rows indicate the data set and columns indicate the direction of deregulation. The upper panel shows the overlap within the TGD data set and the lower panel shows the overlap between the TGD and NRW data set. The left-hand panel demonstrates the Venn diagrams for the upregulated genes (NRW: > 1.5-fold with adjusted $p \leq 0.01$ and TGD: > threefold) and the right-hand panel illustrates the Venn diagrams for the downregulated genes.*

The *consensus Sv 3 genes* of the NRW-TGD overlap (Figure 5.14) is further used for the *in vivo* biomarker identification. The particular goal is to identify the smallest possible number of genes, which, in combination, respond to as many test compounds as possible. The *in vitro-in vivo* overlap of these genes indicates that the involved mechanisms are not pure *in vitro* artifacts. The consensus genes are ranked in order of selection value, i.e. the gene with the highest selection value is ranked first. The following genes are determined recursively to the previous ones such that as many as possible individual compounds are covered cumulatively by the selected genes. If no new compounds are added with further genes, the curve of covered compounds is saturated. Figure 5.15 demonstrates such a cumulative curve for the NRW data set (upper panel). The x-axis lists the number of biomarker genes and the y-axis lists the absolute number of covered compounds. Aldh1a1, for example, is upregulated by nine compounds. Including the gene Gdf15, four further compounds are covered in addition to the current ones, resulting in 13 covered compounds in total. Five consensus genes are required to reach the saturation point of the curve.

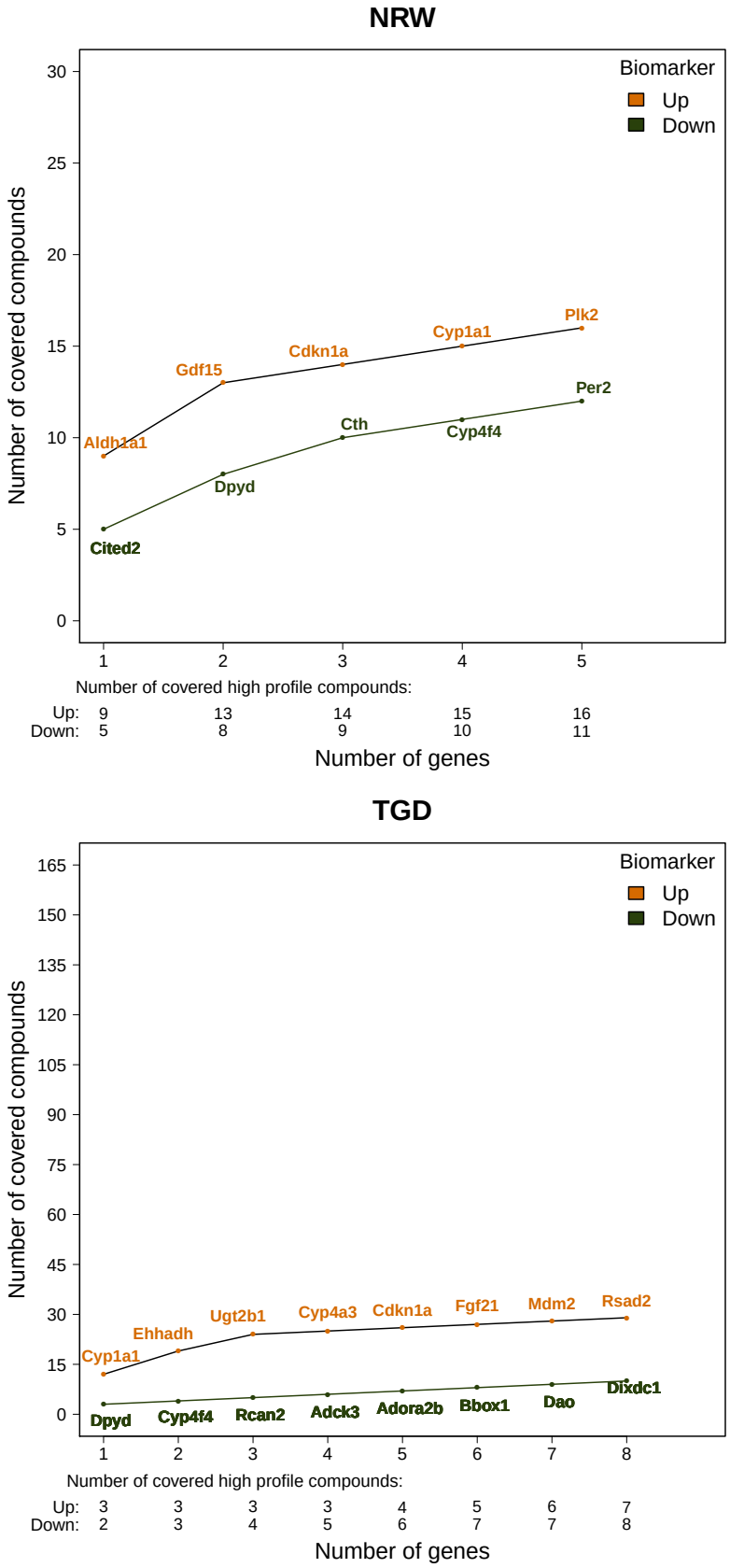


Figure 5.15: *Plot showing the cumulative percentage of covered compounds in the NRW (upper panel) and the TGD (lower panel) data set. Indicated genes represent biomarker genes which are induced in both, the in vitro and in vivo data set, by most of the compounds (same compounds). Biomarkers result from ranking the consensus Sv 3 NRW or TGD genes, respectively, in order of selection value. Orange and green color indicate the up- and downregulated biomarker genes, respectively.*

The lower panel of Figure 5.15 shows the subset of up- and downregulated biomarkers that cover, in combination, 29 and 10 compounds, respectively, which correspond to 17.5% and 6.1% of the compounds studied in the TGD data set ($n = 165$). It can be noted, that from the fourth biomarker on only one compound per gene is added to the ones covered already. In case of the downregulated biomarker genes this single-wise compound increase can be observed from the first gene onwards. At this point it is worth mentioning, that a relatively high fraction of compounds, 151 of 165, induce no or only very few genes and are therefore classified as low profile compounds.

In combination, the NRW and TGD data set comprise 189 compounds of which 160 (84.6%) are low profile compounds. The curve showing the number of covered compounds in the combined database is given in Figure 5.16. Saturation is achieved with 11 biomarker genes.

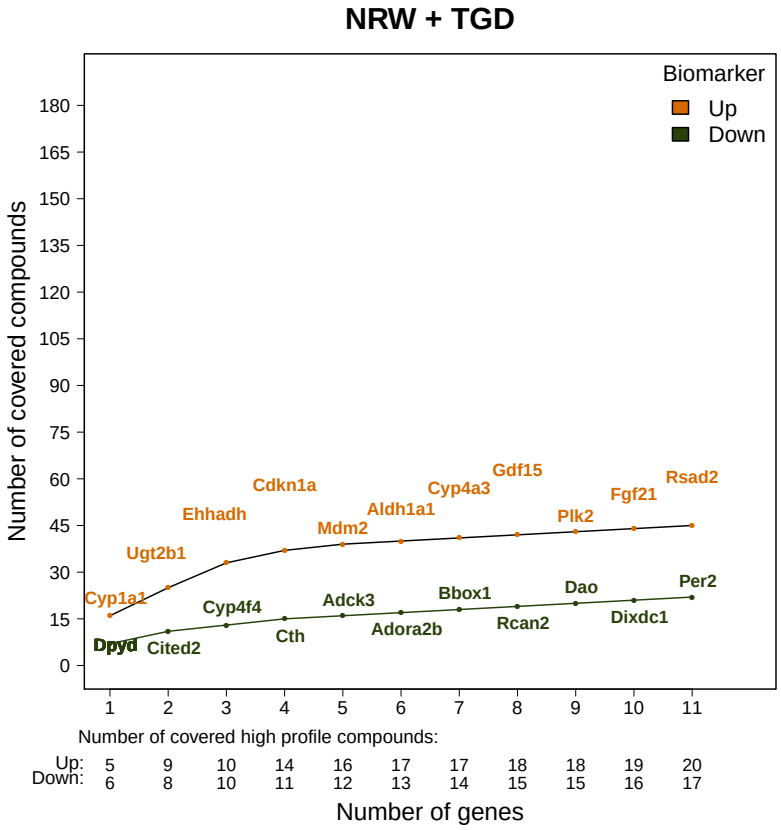


Figure 5.16: *Plot showing the cumulative percentage of covered compounds in both data sets combined, the NRW and TGD data set. Indicated genes represent biomarker genes which are induced in both the in vitro and in vivo combined data set by most of the compounds (same compounds). Biomarkers result from ranking the consensus Sv 3 genes from both data sets in order of selection value. Orange and green color indicate the up-and downregulated biomarker genes, respectively.*

The first five upregulated genes (Cyp1a1, Ugt2b1, Ehhadh, Cdkn1a, Mdm2) cumulatively contribute most to the covered compounds. They cover 39 of 189 compounds, i.e. 20.8% of all compounds, while each of the next six genes adds only one new compound. A similar coverage increase can be observed for the downregulated genes, where only the first four genes (Dpyd, Cited2, Cyp4f4, Cth) add more than one new compound to the already existing compounds.

Cumulatively, the up- and downregulated biomarker genes cover 23.8% and 11.6% of the compounds, respectively, such that 76.2% and 88.4% of the compounds remain uncovered, of which 71.4% and 82.0%, respectively, belong to low profile compounds anyway. Considering the fact that with the exception of five mutual compounds, predominantly different compounds were tested in the NRW and TGD data set, the coverage of compounds is relatively high since only 4.8% (up) and 6.4% (down) of the high profile compounds remain uncovered.

The overall design of the foregoing analysis is summarized in Figure 5.17. In summary, the analysis targeted the question, if a set of genes can be identified which responds similarly *in vitro*

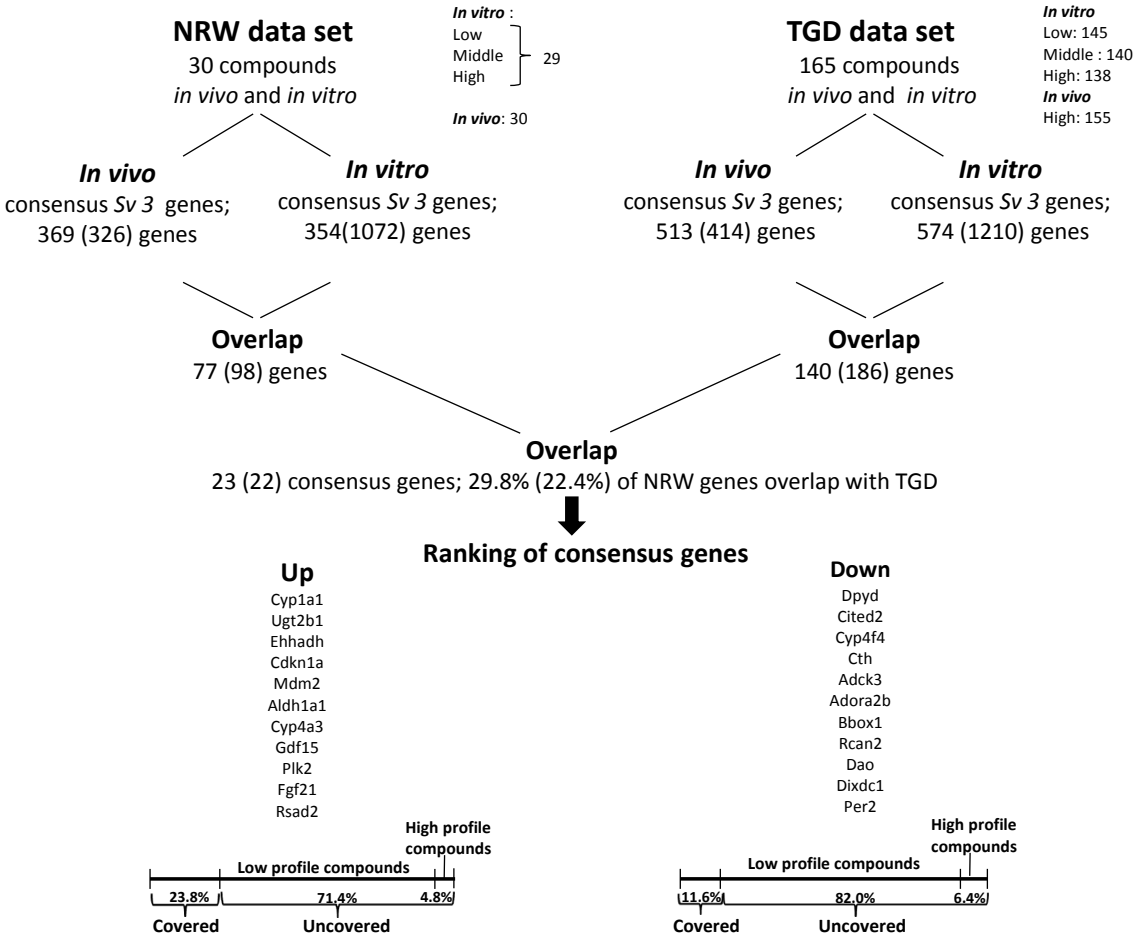


Figure 5.17: *Overview of the study design and the analytical procedure for the in vivo biomarker identification.*

and *in vivo* after chemical induction. To approach this question, two independent data sets, the NRW and TGD databases, were analyzed with respect to a consensus response of rat hepatocytes exposed to 189 different compounds. Database-specific consensus gene lists for the *in vitro-in vivo* comparison, based on the Sv 3 concept, were compiled and examined in terms of overlaps. 11 biomarker genes covering in combination 45 up- and 22 downregulating compounds were identified.

6 Statistical analysis of dose-expression data

In this chapter, in contrast to the two previous ones where a classical method was used for the detection of alert concentrations, an alternative model-based approach is proposed. In Chapters 4 and 5 only few treatment conditions were tested, usually three concentrations or time points. Due to the limited number of measurements, differential expression analysis was performed condition-wise, such that the dose-response relationship across different concentration or time points was completely neglected. That means in particular, that for each test condition separately, it was tested if the critical effect level was exceeded. Then an overlap analysis of deregulated genes between adjacent condition levels was performed. In this chapter a concentration or time series with more than three measurements is considered. Given these conditions, a regression approach for the modeling of gene expression data can be used. The model-based approach fits parametric models for repeated treatment-dependent gene responses using fold changes as a measure of the association. The model is then used to derive an estimate of the concentration (effective dose) that corresponds to a pre-specified effect level (response). Uncertainties of the estimates are indicated by 95%-confidence intervals. This chapter presents the results of a comparative analysis of two estimation methods which are used for the detection of critical compound concentrations. The analysis is performed on both simulated and real experimental dose-response data. All analyses are based on the assumption that the data exhibits a sigmoidal relationship between dose and response and, hence, the four parameter log-logistic (4pLL) model is an adequate approximation of the data.

6.1 Simulation study and setup

The simulation study was performed to evaluate and compare the 4pLL model approach with the classical naïve *Limma* approach with respect to their estimated alert concentrations. The data was

simulated to estimate a) the Absolute Lowest (Observed) Effective Concentration (AL(O)EC) at which the effect level of interest is reached exactly or exceeded by the average value and b) the Lowest (Observed) Effective Concentration (L(O)EC) at which the effect level is exceeded significantly. The two estimates in case a) result from a simple point estimator, while those in case b) take the uncertainty of the respective effect levels into consideration (CI-based estimates). In the following, the two point estimates (ALEC vs. ALOEC) and the two CI-based estimates (LEC vs. LOEC) are compared with each other. The respective alert concentrations are estimated according to the methods in the Sections 3.3 and 3.5. For an overview of the four estimates the reader is referred to Table 3.1. The critical effect level λ was set to a fold change value of 1.5.

Simulation setup

The setup of the simulation study matches the experimental design of the real VPA chronic dose-response study which was introduced in Section 2.4. The measured concentrations are 0, 25, 150, 350, 450, 550, 800 and 1000 µM where the concentration 0 refers to the control values. For each concentration, three replicate experiments were performed.

Simulated dose-expression data was generated from the 4pLL model function (3.10) in Section 3.5. Different gene expression profiles were simulated by fitting 4pLL models with various parameter sets. From the set of possible gene expression patterns four patterns were selected for the simulation study. The four expression profiles were chosen such that four different scenarios were covered by the chosen curve progressions. In the simulation study only increasing dose-response relations were considered, i.e. the lower limit $\phi^{(c)}$ was set to zero in all four scenarios. The results obtained for increasing curve progressions are transferable to decreasing curve progressions by changing the sign of the log-fold change values from positive to negative. Hence, increasing curve progressions represent upregulated gene expression profiles and decreasing curves represent downregulated expression profiles. The following four scenarios were analyzed:

Scenario I: The true parameters were set to $\phi^{(b)} = -4, \phi^{(c)} = 0, \phi^{(d)} = 0.58, \phi^{(e)} = 200$ such that the fitted curve never exceeds the threshold $\log_2(1.5) = 0.585$. Since the given threshold is not exceeded, the ALEC value cannot be calculated.

Scenario II: The true ALEC is equal to 500 µM. In this scenario the parameters were chosen such that the fitted curve clearly exceeds the given threshold and the upper limit of the curve is not reached. The parameters were set to $\phi^{(b)} = -3, \phi^{(c)} = 0, \phi^{(d)} = 4, \phi^{(e)} = 900$.

Scenario III: The scenario represents the case where a saturated sigmoid curve is given and the true ALEC value coincides with a measured concentration level, here 550 µM. The true parameters were set to $\phi^{(b)} = -5, \phi^{(c)} = 0, \phi^{(d)} = 1.5, \phi^{(e)} = 600$.

Scenario IV: The scenario is characterized by a S-shaped curve with an upper asymptote close to the given threshold. The ALEC is equal to 680 µM, which is between two measured values. The scenario represents a compromise between the two curves fitted in Scenario I and II. Thus, the parameter values were set to $\phi^{(b)} = -5$, $\phi^{(c)} = 0$, $\phi^{(d)} = 0.9$ and $\phi^{(e)} = 600$.

The four simulated scenarios are illustrated in Figure 6.1. The true parameters were used to calculate the true ALEC value and to generate simulated data. For each concentration level k response points were generated which were assumed to be normally distributed with a mean equal to $f(x_i, \boldsymbol{\phi})$ and a standard deviation equal to a given value σ_i, $i = 1, \ldots, 8$, where i indicates the i^{th}-concentration. Let y_{ij}^{sim}, $i = 1, \ldots, 8$, and, $j = 1, \ldots, k$, denote the simulated expression value of concentration i and replicate j, then $y_{ij}^{sim} \sim \mathcal{N}(f(x_i, \boldsymbol{\phi}), \sigma_i^2)$, where $f(x_i, \boldsymbol{\phi})$ corresponds to the true 4pLL model which was determined by the pre-specified parameters $\phi^{(b)}, \phi^{(c)}, \phi^{(d)}$ and $\phi^{(e)}$ of the given scenario. The standard deviations between the simulated replicates were obtained from the respective standard deviations in the real data set. They were calculated from the VPA chronic concentration study (real data example) gene-wise and for each concentration separately, i.e. the values for σ_i, $i = 1, \ldots, 8$, were chosen equal to the empirical standard deviations of the eight triplicate pairs of the gene which was randomly selected from the set of all measured genes (54 675 in total).

For each scenario 1000 expression profiles with k, $k \in \{3, 6, 10\}$, replicates per concentration were generated. The *Limma* method used the information of all 1000 genes for the empirical Bayes adjustment of the gene-wise variance estimates. The simulation study with $k = 3$ replicates was carried out to reflect the realistic data example. To evaluate, if the sample size has an effect on the model performance of the applied methods, the number of replicates per concentration was increased from 3 to 6 to 10 replicates. The two methods were evaluated with respect to their estimate accuracy. The estimated alert concentrations of the two methods were compared in terms of their total number of alerts (n), median (Med), interquartile range (IQR) and standard deviation (SD), as well as their deviation from the true alert concentration. Distributions of the estimated alert concentrations are shown. In addition, 95%-confidence intervals (CIs) for the ALEC estimators were calculated (according to formula (3.12) in Section 3.5.3) and compared regarding their lengths and coverage probability. The term *false positive alert* is used in cases in which the estimated concentration value is below the true ALEC value.

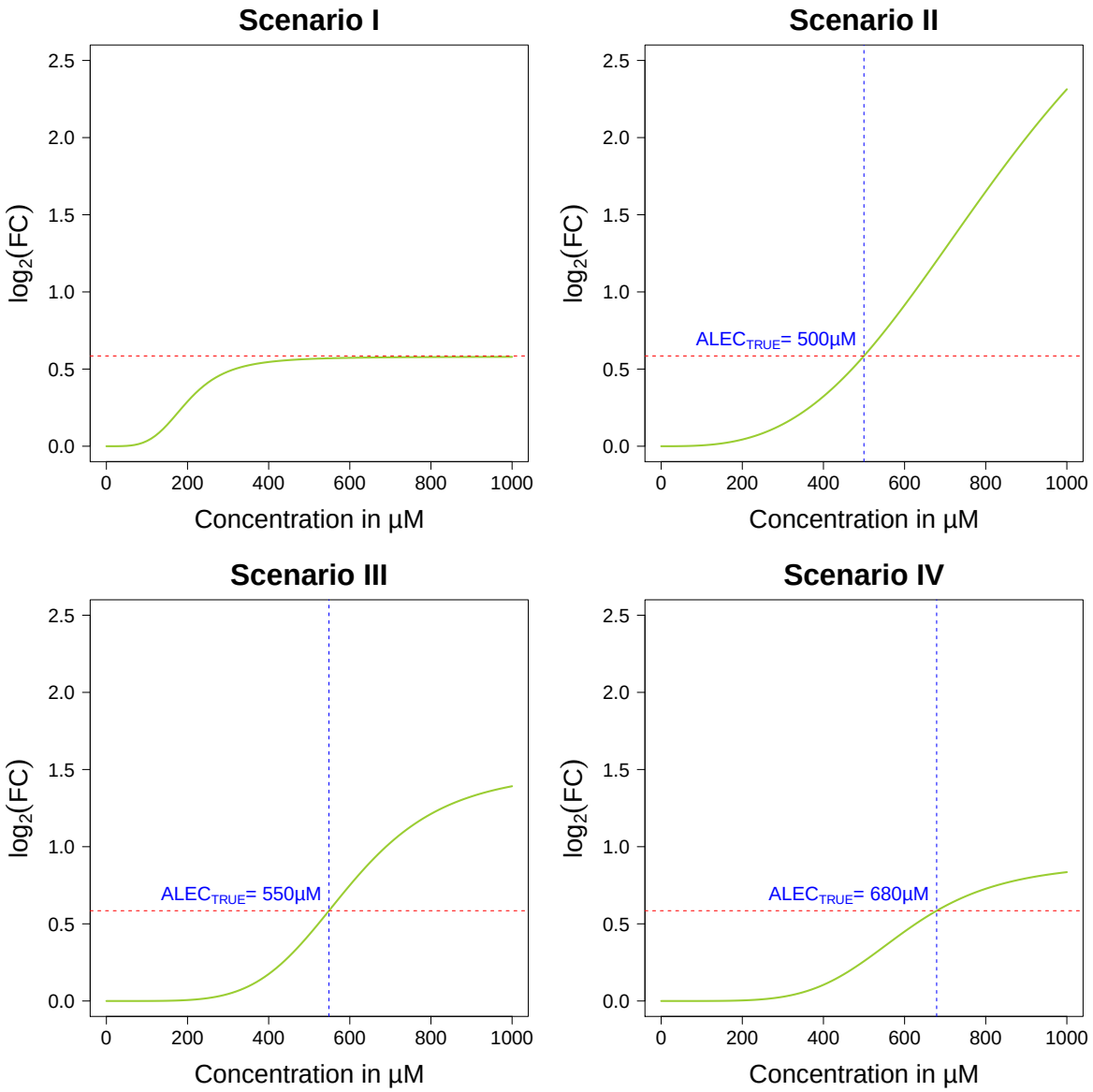


Figure 6.1: *Illustration of Scenarios I-IV: The green curves represent the true expression profiles fitted from the 4pLL model. Dashed lines indicate the critical effect level λ (red) and its true ALEC value (blue).*

6.2 Results of the simulation study

6.2.1 Comparison of the distributions

Figure 6.2 shows the distributions of the estimated alert concentrations for the four simulated scenarios. Rows indicate the scenario and columns the method of estimation. The histograms in the left panel of the figure show the distributions of the estimated ALECs (4pLL). The distributions of the ALOEC estimators (*Limma*) are displayed by the barplots in the right panel

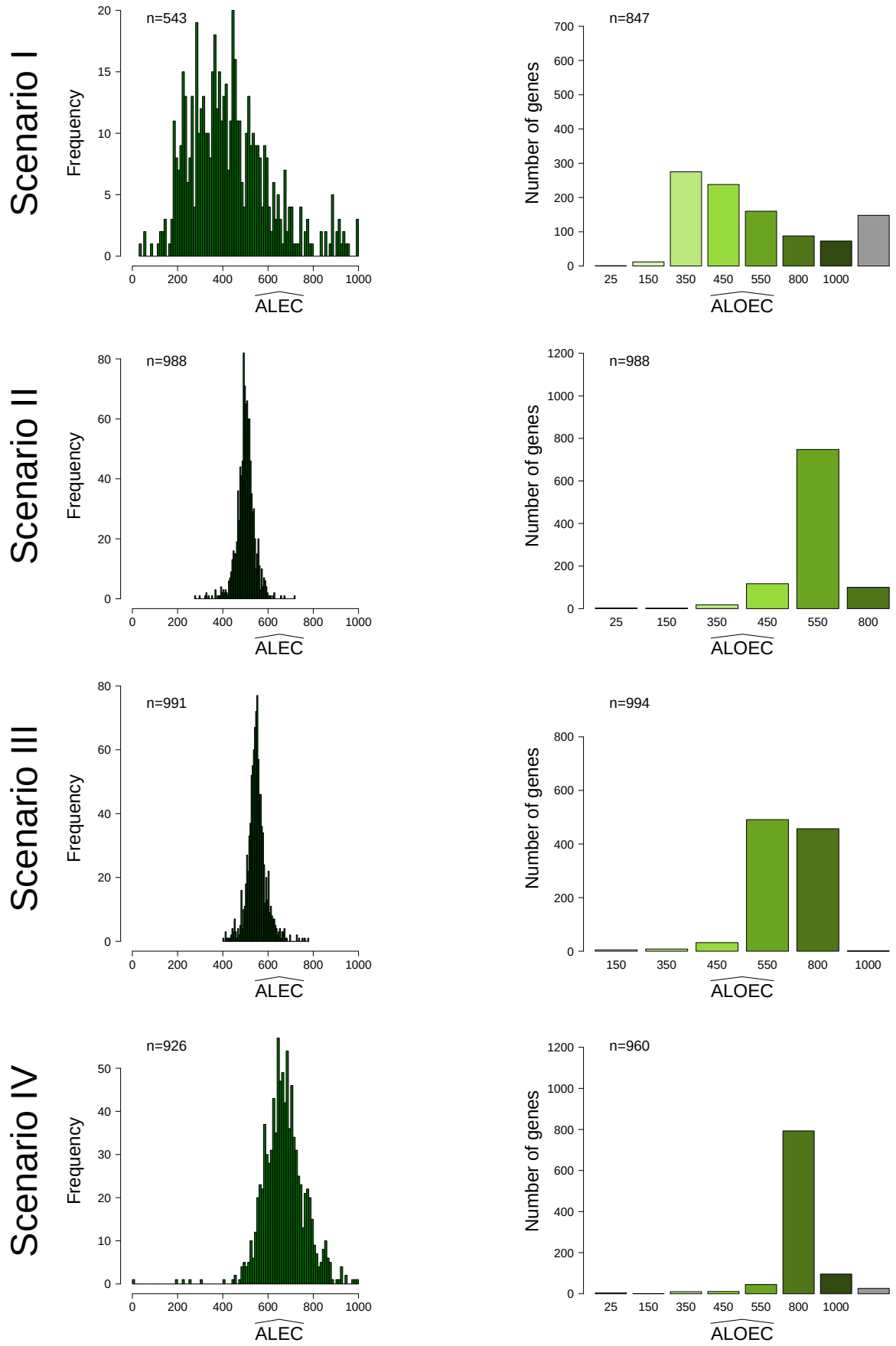


Figure 6.2: *Distributions of the estimated alert concentrations for Scenarios I-IV with $k = 3$ replicates. Rows indicate the scenario and columns the method of estimation. The left panel shows the distributions of the $\widehat{ALEC}$s (4pLL) and the right panel the distributions of the $\widehat{ALOEC}$s (Limma). The number of estimates $\leq$ 1000 µM is indicated by n. Grey colored bars indicate the number of no alerts.*

of the figure. Note, that the estimated alerts shown in Figure 6.2 represent the lowest (observed) concentration levels at which the critical effect level of 1.5-fold is reached exactly ($\widehat{\text{ALEC}}$) or exceeded by the average value ($\widehat{\text{ALOEC}}$), i.e. without significance testing.

The number of successful simulations was lower than 1000 when the 4pLL model did not converge successfully or when the respective ALEC estimator was larger than the highest approved test concentration, which was 1000 µM in the simulation study. For *Limma*, rarely it was observed that for one concentration the upper threshold ($\log_2(1.5)$) was exceeded and for another concentration the estimated value fell below the lower threshold ($-\log_2(1.5)$). From these few cases only ALOEC estimators were affected. The respective genes were excluded from the analysis. Table 6.1 lists, for each scenario, the numbers of excluded genes.

Table 6.1: *Total number of excluded genes in Scenarios I-IV. The given numbers indicate for how many genes the Limma approach has noticed a crossing of the upper threshold (0.585) at one concentration and a crossing of the lower threshold (-0.585) at another concentration. Such genes were excluded from further analysis. Only ALOEC estimators were affected from these cases.*

	Scenario I	**Scenario II**	**Scenario III**	**Scenario IV**
k=3	5	12	6	14
k=6	0	3	0	0
k=10	0	0	0	0

Scenario I, in which the true dose-expression curve does not exceed the given threshold, represents the most challenging situation of all four scenarios. Although, no alerts are expected, both methods trigger alerts which are, in this case, interpreted as false positives. The 4pLL method detects 543 alerts, in median at a concentration level of 406.6, and the *Limma* method falsely identifies 847 alerts, in median at 450 µM. Thus, the number of false positive alerts obtained with the *Limma* approach is clearly higher than the one observed for the $\widehat{\text{ALEC}}$s. In contrast to the other three scenarios, where the $\widehat{\text{ALEC}}$s follow approximately a normal distribution, the distribution of the $\widehat{\text{ALEC}}$s in Scenario I is slightly right skewed.

In Scenario II, where the true expression curve exceeds the given fold change at 500 µM, the median of the estimated ALECs is 502.2 µM. The distribution of the $\widehat{\text{ALEC}}$s is, with a standard deviation of 40.8, more concentrated than the distribution of the $\widehat{\text{ALEC}}$s in Scenario I. The distribution of the $\widehat{\text{ALOEC}}$s has a median of 550 µM and a standard deviation of 96.7 µM, whereas the interquartile range is zero, meaning that the 25th- and 75th percentiles coincide.

In Scenario III, where the true ALEC value is equal to 550 µM (measured concentration), both methods hit on average the true value. Nevertheless, *Limma* gives an alert at 800 µM almost

as often as at 500 µM. In addition, the estimates of *Limma* vary more than those of the 4pLL approach (SD_{Limma} = 136.4 vs. SD_{4pLL} = 43.1).

In Scenario IV, which reflects the situation in which the true ALEC value is 680 µM and lies between two measured concentration levels, the largest differences between the two estimates are visible. The 4pLL method is, on average, with a median of 666.6 µM, closer to the true ALEC value than the *Limma* method with a median of 800 µM. Similar to the previous three scenarios the distribution of the ALOEC estimates exhibits a higher standard deviation than the distribution of the ALEC estimates (SD_{Limma} = 131.3 vs. SD_{4pLL} = 92.7). Table B.7 in the Appendix summarizes the results of Figure 6.2.

Figure 6.3 shows the distributions of the estimated LECs (4pLL) and LOECs (*Limma*). Here, the estimated alerts refer to the lowest (observed) concentration level at which the given fold change is exceeded significantly ($p \leq 0.05$). The p-values result from the respective t-tests adjusted for the two estimation methods.

In all four scenarios the number of successfully converged LEC estimators is only a fraction of the number of successfully converged point estimates (see Figure 6.2).

In Scenario I, *Limma* provides in 153 cases a false positive result, while the 4pLL method triggers only in 36 cases a false positive alert.

In Scenario II, the 4pLL approach did not converge in 684 of 1000 cases. In this scenario, where the true ALEC value is equal to 500 µM, all converged LEC estimators ($n = 316$), except of one, take values in the range between 500 and 1000 µM. On average, an alert is given at 785.1 µM. *Limma*, however, provides a CI-based estimator for the LOEC in all cases but one. This is due to the fact that the curve has not reached its saturation point at the highest tested concentration and the uncertainty about the further path of the curve leads to large variance estimates for the given fold change values. This results in wide confidence intervals, which influence the estimation of the LECs. For this reason, the 4pLL method detects a significant change in expression in only a fraction of the genes (Table 6.2). In contrast, the *Limma* method is not affected by this problem as its estimation strategy does not depend on the curve progression. Thus, *Limma* provides for all genes a LOEC estimator. The median 50% of the $\widehat{\text{LOEC}}$s take values of the two next higher levels measured after the true 500 µM (550 µM or 800 µM). The median concentration coincides with the 75%-quantile which is 800 µM. In those cases, in which an estimator is obtained, both methods overestimates the true ALEC value. The false positive rate is below 1% for both methods.

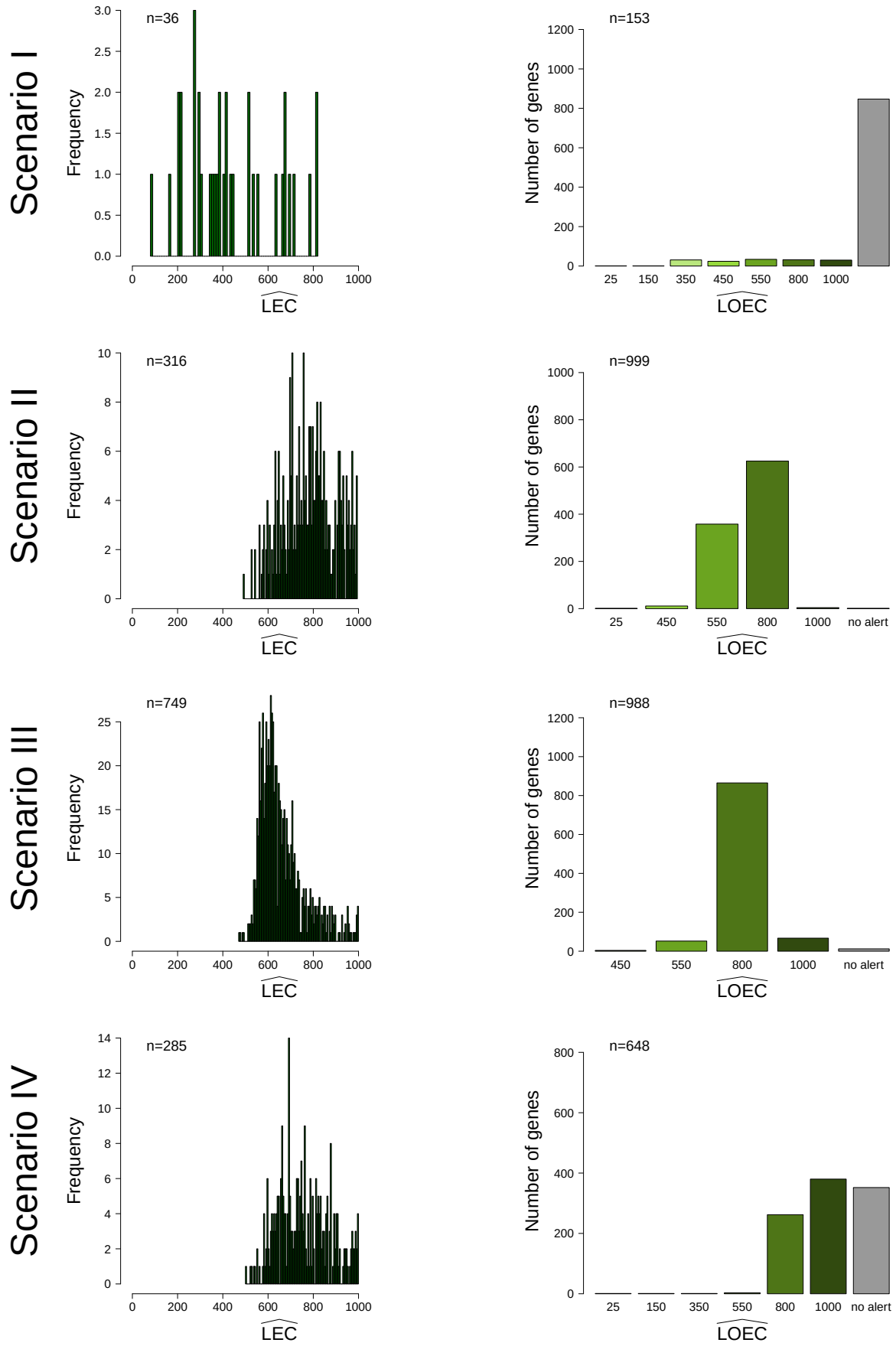


Figure 6.3: *Distributions of the estimated alert concentrations for Scenarios I-IV with $k = 3$ replicates. Rows indicate the scenario and columns the methods of estimation. The left panel shows the distributions of the $\widehat{LEC}$s (4pLL) and the right panel the distribution of the $\widehat{LOEC}$s (Limma). The number of estimates $\leq$ 1000 µM is indicated by n. Grey colored bars indicate the number of no alerts.*

In Scenario III, where a saturated sigmoid curve is given, the largest number of converged LEC estimators is observed ($n = 749$), i.e. in 251 cases (25.1%) the 4pLL model failed to converge. The *Limma* approach, in contrast, did not notice a significant alert in 12 of 1000 cases. Neglecting the number of false positive alerts (n_{4pLL}= 35 vs. n_{Limma}= 4), the 4pLL method provides more accurate estimates than the *Limma* approach. Even though the true ALEC value corresponds to 550 µM, a measured concentration level, the model-based (4pLL) estimates are on average closer to the true ALEC value than the classical estimates (Med_{4pLL} = 632 vs. Med_{Limma} = 800). The $\widehat{LECs}$ take values predominantly in the upper concentration range, starting from the true value of 550 µM. The LOEC estimates vary around 800 µM with a standard deviation of 91.3, which is similar to the standard deviation of the $\widehat{LECs}$ (SD_{4pLL} = 100.7), but with an interquartile range of 0, which is much smaller than the interquartile range of the $\widehat{LECs}$ (IQR_{4pLL} = 113.9).

Table 6.2: *Summary statistics for the distributions of the estimated alert concentrations for Scenarios I-IV. The following parameters are presented: The total number of alerts (n), the median (Med), the interquartile range (IQR) and the standard deviation (SD). The method for estimating the alerts is subscripted after the corresponding parameter. An alert was given when the given fold change value of 1.5 was exceeded significantly ($p \leq 0.05$). The table refers to the Figures 6.3, C.31 and C.33.*

	n_{4pLL}	n_{Limma}	Med_{4pLL}	Med_{Limma}	IQR_{4pLL}	IQR_{Limma}	SD_{4pLL}	SD_{Limma}
				Scenario I				
k=3	36	153	398.200	no alert	284.300	0	197.026	226.859
k=6	55	213	351.600	no alert	258.800	0	203.559	237.007
k=10	50	137	540.500	no alert	326.500	0	217.300	206.052
				Scenario II				
k=3	316	999	785.100	800	163.200	250	114.745	126.586
k=6	427	1000	780.300	550	162.100	250	105.958	125.794
k=10	495	1000	785.700	550	162.200	250	102.972	117.702
				Scenario III				
k=3	749	988	632.000	800	113.900	0	100.691	91.290
k=6	861	995	620.700	800	95.500	0	92.705	77.510
k=10	929	1000	609.200	800	69.000	0	80.034	69.939
				Scenario IV				
k=3	285	648	745.200	1000	175.100	400	115.488	164.588
k=6	419	833	779.600	800	153.700	200	105.302	150.982
k=10	570	910	772.200	800	135.800	200	97.921	130.159

In Scenario IV, where the true ALEC value is 680 µM, the algorithm of the 4pLL method has successfully converged in 285 cases. The distribution of the successfully converged LEC

estimates has a median of 745.2 µM and a standard deviation of 115.5. The *Limma* method, on the other hand, notices a significant crossing of the threshold in 648 cases. *Limma* assumes to find the true ALEC value predominantly at 800 µM and 1000 µM. The 648 LOEC estimates vary around a median of 1000 µM with a standard deviation of 164.6. The 4pLL method provides estimates in the range of 500 and 1000 µM in all 285 cases. Although *Limma* yields less false positive alerts ($n = 6$), the method detects, on average, the true alert with a delay. The 4pLL method, on the other hand, returns more false positive results, 84 in total, but is generally closer to the truth. Nevertheless, it should be noted, that both methods missed the alert in a relatively high fraction of genes. The 4pLL algorithm failed in 715 cases and *Limma* in 352 cases. The reason for this might be that the given scenario is unfavourable for both methods. Although the S-shaped curve reaches its saturation point within the given concentration range, most of the estimated confidence intervals capture, however, the critical effect level λ due to the fact that the distance between the upper asymptote of the curve ($\phi^{(d)}$ = 0.9) and the threshold is very small ($\approx$ 0.3). Similar results are obtained with *Limma*, where the computed average fold change values do not exceed the threshold significantly. To this end, this scenario allows the estimation of an accurate point estimator, but as soon as the uncertainty of the estimates is taken into consideration, the alert can only be proven in a fraction of genes. In Table 6.2 the results of Figure 6.3 are summarized. Table 6.3 reports the total number of false positive results obtained for both criteria, the less stringent criteria (FC $\geq$ 1.5) and the more stringent criteria (FC $\geq$ 1.5 $\&$ $p \leq$ 0.05).

Both figures, Figure 6.2 and Figure 6.3, refer to the simulation study, in which $k = 3$ replicates per concentration were generated. The corresponding distributions for $k = 6$ and $k = 10$ simulated replicates are presented in the Appendix (Figures C.30-C.33). Figures C.30 and C.32 refer to the situation, in which an alert was given when the given fold change value of 1.5 was reached exactly (4pLL) or exceeded by the average value (*Limma*), both indicating the point estimate, and Figures C.31 and C.33 refer to the situation, in which an alert was identified by means of hypothesis testing (CI-based estimate).

The increase of sample size has a larger impact on the model-based estimates than on the classical ones. The number of successfully converged ALEC/LEC estimators has increased with the number of replicates. Similar to, but not quite as pronounced as for the 4pLL estimates, the number of alerts missed by the classical method, has decreased. Notable improvements were obtained for Scenario III (saturated sigmoid curve), where the number of successfully converged LEC estimators has increased from 749 to 929. In general, increasing the sample size, affects the

estimates from the hypothesis driven approach ($\widehat{\text{LEC}}$ and $\widehat{\text{LOEC}}$) more than the point estimates ($\widehat{\text{ALEC}}$ and $\widehat{\text{ALOEC}}$). However, in comparison to the original replicate number of $k = 3$ the distributions for $k = 6$ and $k = 10$ simulated replicates do not show large differences in median (Med), interquartile range (IQR) or standard deviation (SD). Table 6.2 provides an overview of the summary statistics for the CI-based estimates. The corresponding summary statistics for the point estimates are given in Table B.7 in the Appendix. The only remarkable difference can be observed in Scenario II and IV for the $\widehat{\text{LOEC}}$s , where the increase of sample size results in an earlier alert detection. The median of the $\widehat{\text{LOEC}}$s has shifted to the left: The median alert concentration has taken the value of the next lower concentration which is 550 µM in Scenario II (true ALEC = 500 µM) and 800 µM in Scenario IV (true ALEC = 680 µM).

Table 6.3: *Total number of false positive alerts. A false positive alert means that the indicated method yields an estimate of a value below the true ALEC value. In case of Scenario I, where no ALEC value was provided, a false positive alert has been noted, if any alert was triggered. An alert was given when the given fold change value of 1.5 was reached exactly ($\widehat{\text{ALEC}}$) or exceeded by the average value ($\widehat{\text{ALOEC}}$) (upper table) or exceeded significantly ($p \leq 0.05$) (lower table).*

	Scenario I		**Scenario II**		**Scenario III**		**Scenario IV**	
	4pLL	Limma	4pLL	Limma	4pLL	Limma	4pLL	Limma
	$\widehat{\text{ALEC}}$	$\widehat{\text{ALOEC}}$	$\widehat{\text{ALEC}}$	$\widehat{\text{ALOEC}}$	$\widehat{\text{ALEC}}$	$\widehat{\text{ALOEC}}$	$\widehat{\text{ALEC}}$	$\widehat{\text{ALOEC}}$
k=3	543	847	476	140	523	45	520	71
k=6	503	886	507	93	505	16	498	29
k=10	519	815	484	41	520	6	526	12

	Scenario I		**Scenario II**		**Scenario III**		**Scenario IV**	
	4pLL	Limma	4pLL	Limma	4pLL	Limma	4pLL	Limma
	$\widehat{\text{LEC}}$	$\widehat{\text{LOEC}}$	$\widehat{\text{LEC}}$	$\widehat{\text{LOEC}}$	$\widehat{\text{LEC}}$	$\widehat{\text{LOEC}}$	$\widehat{\text{LEC}}$	$\widehat{\text{LOEC}}$
k=3	36	153	1	12	35	4	84	6
k=6	55	213	0	7	26	1	72	2
k=10	50	137	0	0	23	0	88	0

From the distributions of the estimates, it can be concluded that, firstly, the convergence of the LEC estimates critically depends on the sample size. Secondly, the 4pLL method has outperformed the *Limma* approach in Scenario I (no ALEC value provided) and Scenario III (saturated sigmoid curve) in terms of less false positive alerts and more accurate estimates. In Scenario II, where an unsaturated curve progression is given, the highest uncertainties for the LEC estimates are observed. In Scenario IV, where the critical effect level is exceeded only slightly, both methods yield biased estimates.

6.2.2 Comparison of the quantile distributions

Figure 6.4 displays the distribution of the quantiles which were calculated from the distribution of the $\widehat{\text{ALECs}}$ representing the $\widehat{\text{ALOECs}}$. Given the distribution of the ALEC estimators, it can be calculated which quantiles of this distribution the $\widehat{\text{ALOECs}}$ correspond to. A detailed description and derivation of the quantiles is given in Section A.3 in the Appendix.

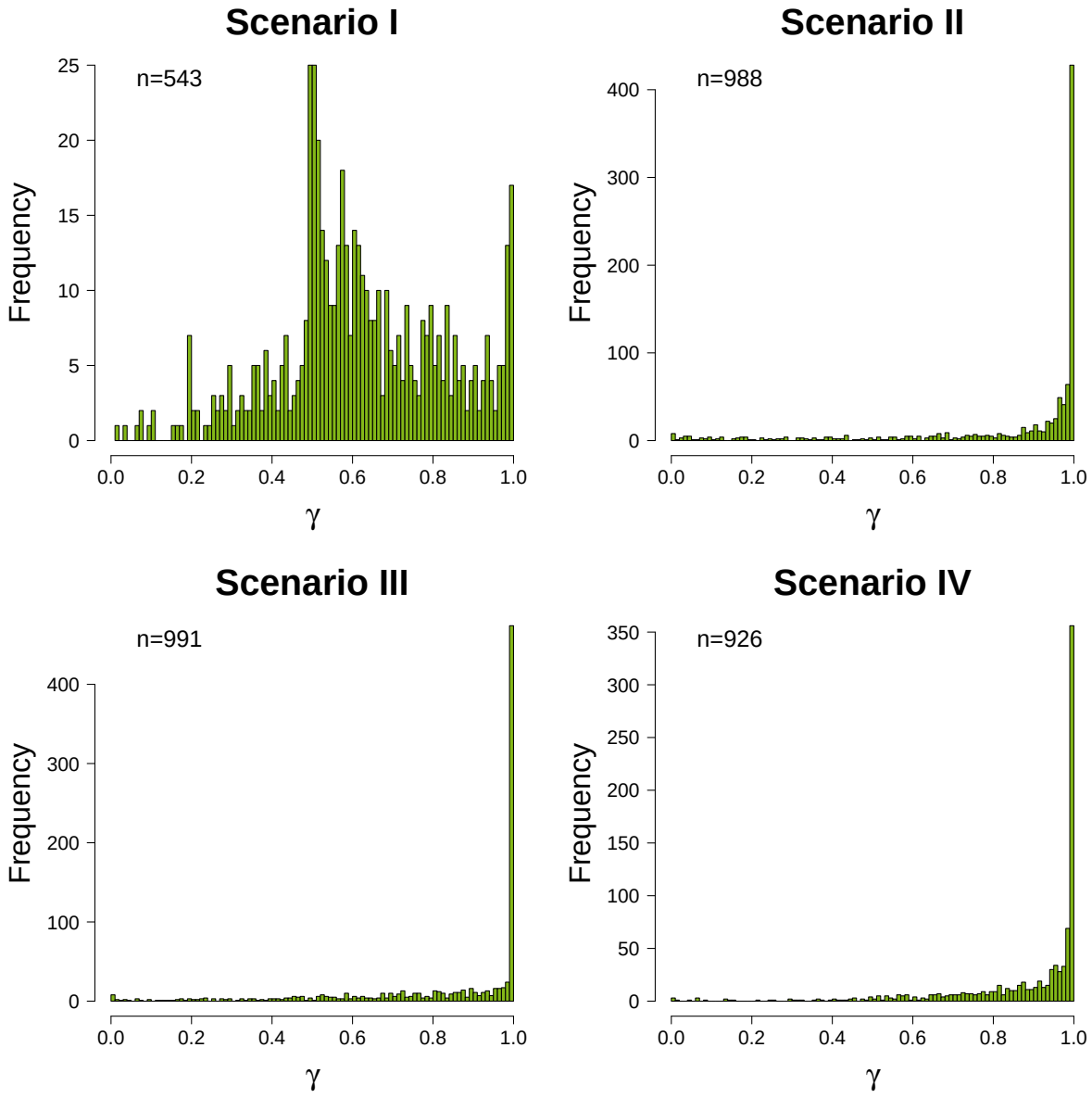


Figure 6.4: *Distributions of the quantiles calculated from the distribution of the $\widehat{\text{ALECs}}$ representing the $\widehat{\text{ALOECs}}$. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the indicated scenario. Values close to zero indicate that the Limma method detects alerts at lower concentrations than the 4pLL method, values close to one indicate the reverse case.*

Values close to zero indicate that *Limma* detects alerts at lower concentration levels than the 4pLL method, while values close to one indicate the reversed case, namely that the 4pLL method notices expression changes at lower concentrations than *Limma*. In Scenario I, 25% of the alerts detected by both methods do not differ in terms of their estimated concentration values. In 50% of the cases, the $\widehat{\text{ALOEC}}$ values are slightly higher than the $\widehat{\text{ALEC}}$ values.

In the Scenarios II-IV the distributions of the quantiles are left-skewed with median values close to one. In none of the three scenarios the 25%-quantile falls below the value of 0.7 indicating that in 75% of the cases lower concentration levels are obtained for the $\widehat{\text{ALEC}}$s than for the $\widehat{\text{ALOEC}}$s .

Similar results can be observed for the distribution of the quantiles which were calculated from the distribution of the $\widehat{\text{LEC}}$s representing the $\widehat{\text{LOEC}}$s (calculated as above, but this time the calculations refer to the distributions of the CI-based estimators). Figure 6.5 shows the corresponding histograms for the Scenarios I-IV. In all four scenarios, except Scenario II, the median of the observed values is close to one. In 75% of the cases, the quantiles take values above 0.6. In Scenario III the smallest interquartile range of 0.08 can be observed indicating a relatively low dispersion of the values. In agreement with Table 6.2, the histogram in Figure 6.5 shows that the lower values of the $\widehat{\text{LEC}}$s are closer to the true ALEC value than the values of the $\widehat{\text{LOEC}}$s. That means, that in this particular case the model-based approach outperforms the classical procedure in terms of more accurate estimates. In Scenario II, however, the distribution of the quantiles shows two peaks, one peak at zero and the other peak at one. Hence, the *Limma* method detects an alert in half of the cases at much lower concentration levels than the 4pLL method and in one sixth of the cases at much higher concentrations. Compared to the other three scenarios, where a lower dispersion is observed, the quantiles obtained in that scenario vary the most with an interquartile range of 0.85 and a median value of 0.02.

All in all, it can be said that in all scenarios, except Scenario II (unsaturated sigmoid curve), where the reverse case is observed, the 4pLL approach generally indicates alerts at lower concentrations than the *Limma* method.

6.2.3 Comparison of the deviations

Next, the differences between the estimated alert concentrations and the respective true ALECs were calculated for both methods. Deviations were only computed for successfully converged simulations. Due to the fact, that in Scenario I no ALEC value was provided, deviations from the

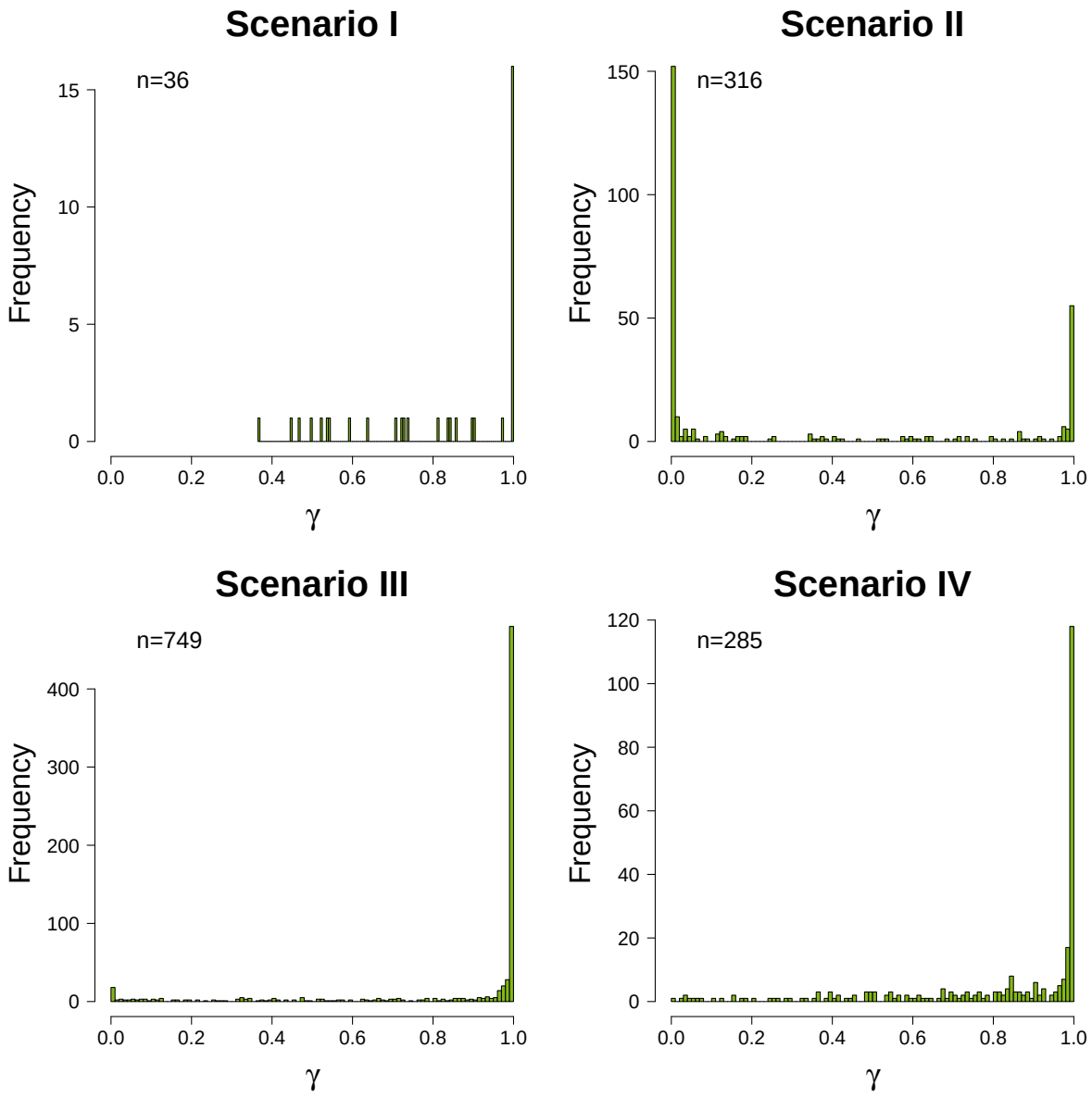


Figure 6.5: *Distribution of the quantiles calculated from the distribution of the $\widehat{LECs}$ representing the $\widehat{LOECs}$. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the indicated scenario. Values close to zero indicate that the Limma method detects alerts at lower concentrations than the 4pLL method, values close to one indicate the reverse case.*

ALEC value could not be computed. Therefore, Scenario I was excluded from this analysis. In general, *no alert* solutions were omitted in the calculations. The results are illustrated by boxplots in Figure 6.6. The boxplots in the upper panel show the deviations of the point estimates from the true ALEC value, and, the boxplots in the lower panel those of the CI-based estimates. The results refer to the simulation study with $k = 3$ replicates. The corresponding results for $k = 6$ and $k = 10$ replicates are shown in the Appendix (Figures C.34-C.35).

In Scenarios II-IV, both methods provide CI-based estimates of values higher than the respective true alert concentrations. The medians of the $\widehat{\text{LOEC}}$ deviations take values between 250 µM and 321 µM. The deviations of the $\widehat{\text{LEC}}$s, on the other hand, yield average values in the range of 66 µM-285 µM. In all three scenarios, Scenario II, III and IV, the mean deviations of the $\widehat{\text{LEC}}$s are smaller and closer to zero than the deviations obtained for the $\widehat{\text{LOEC}}$s. However, it should be taken into account, that the total number of simulations, for which *Limma* has provided estimates of measurable values, is in all scenarios higher than the total number of the corresponding LEC estimators.

Tables B.8-B.9 in the Appendix provide summary statistics for the boxplots shown in Figure 6.6. The results for $k = 6$ and $k = 10$ replicates are documented in the tables as well. The analysis has shown that the distributions of the $\widehat{\text{ALEC}}$s have changed in terms of their interquartile range, which is reduced by half for $k = 10$. The average $\widehat{\text{ALEC}}$s deviate less from the true ALEC value when the replicate number is high, other than the $\widehat{\text{ALOEC}}$s, for which the increase of sample size does not reveal any noteworthy differences.

The reverse case can be observed for the CI-based estimates. In this present situation, the increase of sample size shows an effect on the *Limma* results: In Scenario II and IV, the median of the $\widehat{\text{LOEC}}$ deviations has shifted from the upper to the lower quantile (see Figures C.34-C.35 in Appendix). Apart from these changes, no further noteworthy differences were observed, neither for the LOEC nor the LEC estimators. For more details the reader is referred to the Tables B.8-B.9 in the Appendix.

6.2.4 Direct comparison of the two estimates

In addition, the estimates of the 4pLL method were directly compared with those of the *Limma* approach. Therefore, the ALEC and LEC estimates were analyzed with respect to their concordant *Limma* estimates. The results are illustrated by boxplots in the Figures 6.7-6.10. The distributions of the model-based estimates are split into subsets according to the discrete estimates of the classical approach. Comparisons were made only for simulations for which the 4pLL models have successfully converged. The estimates of the 4pLL aprroach are depicted on the y-axis, which is restricted to the values 0-1000 µM. Alert concentrations of higher values (> 1000 µM) were not used for generating the boxplots, but were highlightened as extreme data points at the top of the single boxes. The number of valid alerts is indicated by n at the bottom of the boxplots. The estimates of the classical *Limma* approach are depicted on the x-axis with

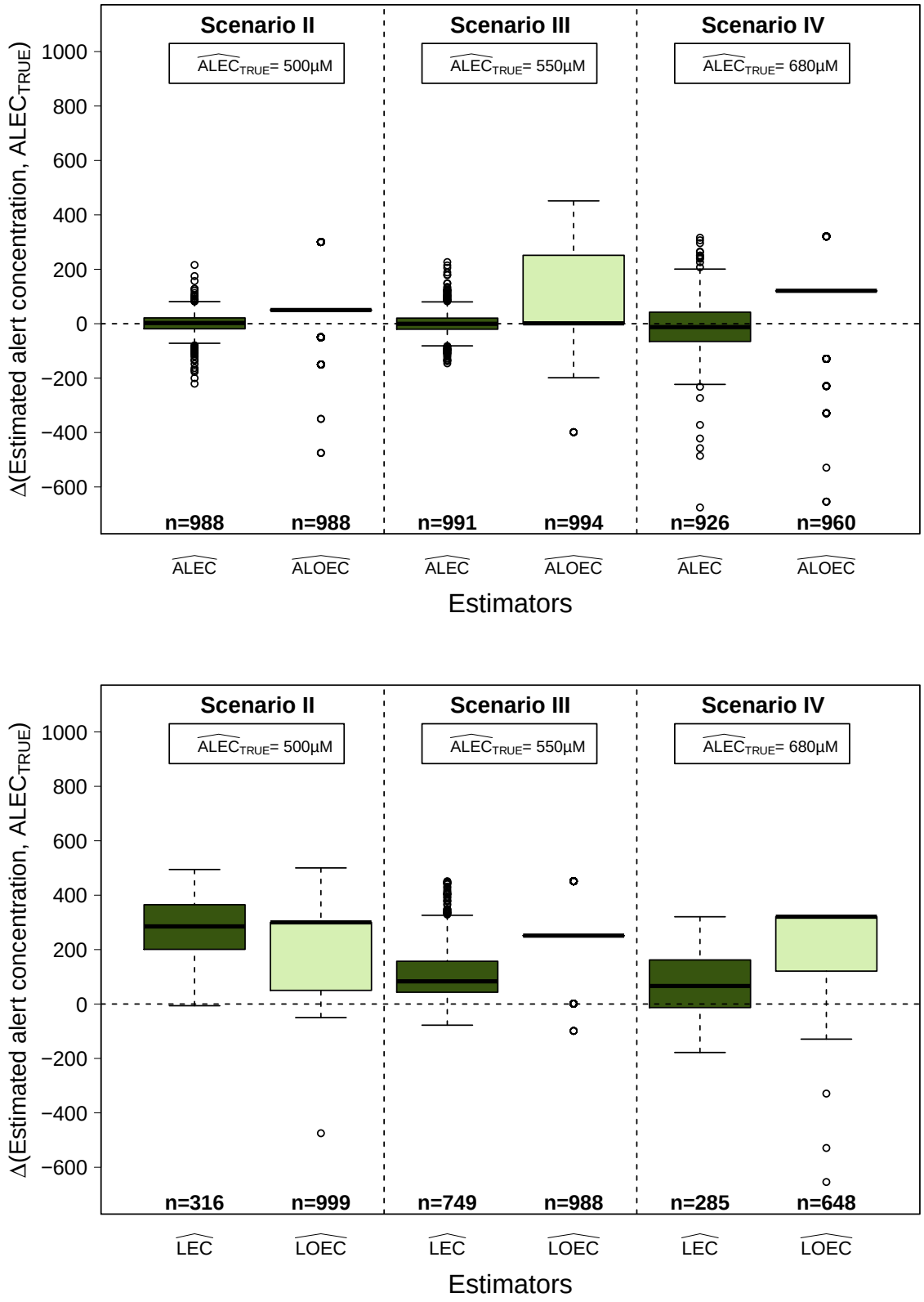


Figure 6.6: *Boxplots illustrating the distributions of the differences between the estimated alert concentrations and the respective true ALECs of the Scenarios II-IV (the difference is indicated by Δ). Scenario I was excluded from the analysis since no deviations could be computed (no ALEC value was provided). The upper panel shows the deviations of the point estimates from the true ALECs and the lower panel shows the deviations of the CI-based estimates ($p \leq 0.05$). The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the indicated scenario.*

coordinate values equal to the measured concentration levels (0, 25, 150, 350, 450, 550, 800 and 1000 µM). *No alert* means that no expression change of at least 1.5-fold was noticed by the indicated method. The black points within the single plots indicate the seven measured concentrations. The uncertainties of the 4pLL point estimators were also taken into account when comparing the two methods with each other. For this, the 95%-confidence intervals (CIs) for the $\widehat{\text{ALECs}}$ and $\widehat{\text{LECs}}$ were calculated and evaluated with respect to the ALOEC and LOEC point estimators, respectively. The results of Scenarios I-IV are discussed individually. The layout of the Figures 6.7-6.10 is the same for all four scenarios: The first row shows the boxplots of the estimated ALECs (left panel) and of their estimated 95%-CIs (right panel) subdivided with respect to the estimated ALOECs. The second row shows the same plots for the CI-based estimates, the $\widehat{\text{LECs}}$ with respect to the $\widehat{\text{LOECs}}$. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the condition of the indicated scenario. Remember that the true expression profiles together with their true ALEC values for $\lambda = \log_2(1.5)$ are shown in Figure 6.1.

Scenario I. In Scenario I the true dose-expression curve falls slightly below the given threshold and thus never meets the 1.5-fold threshold. Nevertheless, both methods sometimes trigger alerts. Figure 6.7 shows that the median of all model-based estimates is below the respective *Limma* estimates (left panel). This indicates earlier alerts with the model-based approach in more than 50% of the cases. The use of the 4pLL method for estimating the ALECs results, in Scenario I, in the widest confidence intervals with median lengths of 500 µM.

The boxplots in the lower panel of Figure 6.7, showing the $\widehat{\text{LECs}}$ in direct comparison to the LOEC estimators, show that the number of alerts decreases enormously when taking confidence intervals into consideration (from 543 to 36 alerts). In 129 cases, the *Limma* approach detects expression changes that do not show any significant change with the 4pLL method (false positive alerts), while the 4pLL method triggers only in 12 other cases a false positive alert (which is not noticed by *Limma*). Thus, in this regard, the 4pLL method outperforms the *Limma* method. These results are in line with the aforementioned results (reported in Table 6.3).

Scenario II. Scenario II was set up to simulate the situation in which the true curve clearly exceeds the given threshold at 500 µM. Figure 6.8 summarizes the results for this scenario. Most of the ALOECs ($n = 748$) are estimated at 550 µM which is the next higher concentration level measured after 500 µM. In contrary, for this subset of genes the 4pLL method provides ALEC estimates of lower values which vary around the true ALEC value with a standard deviation

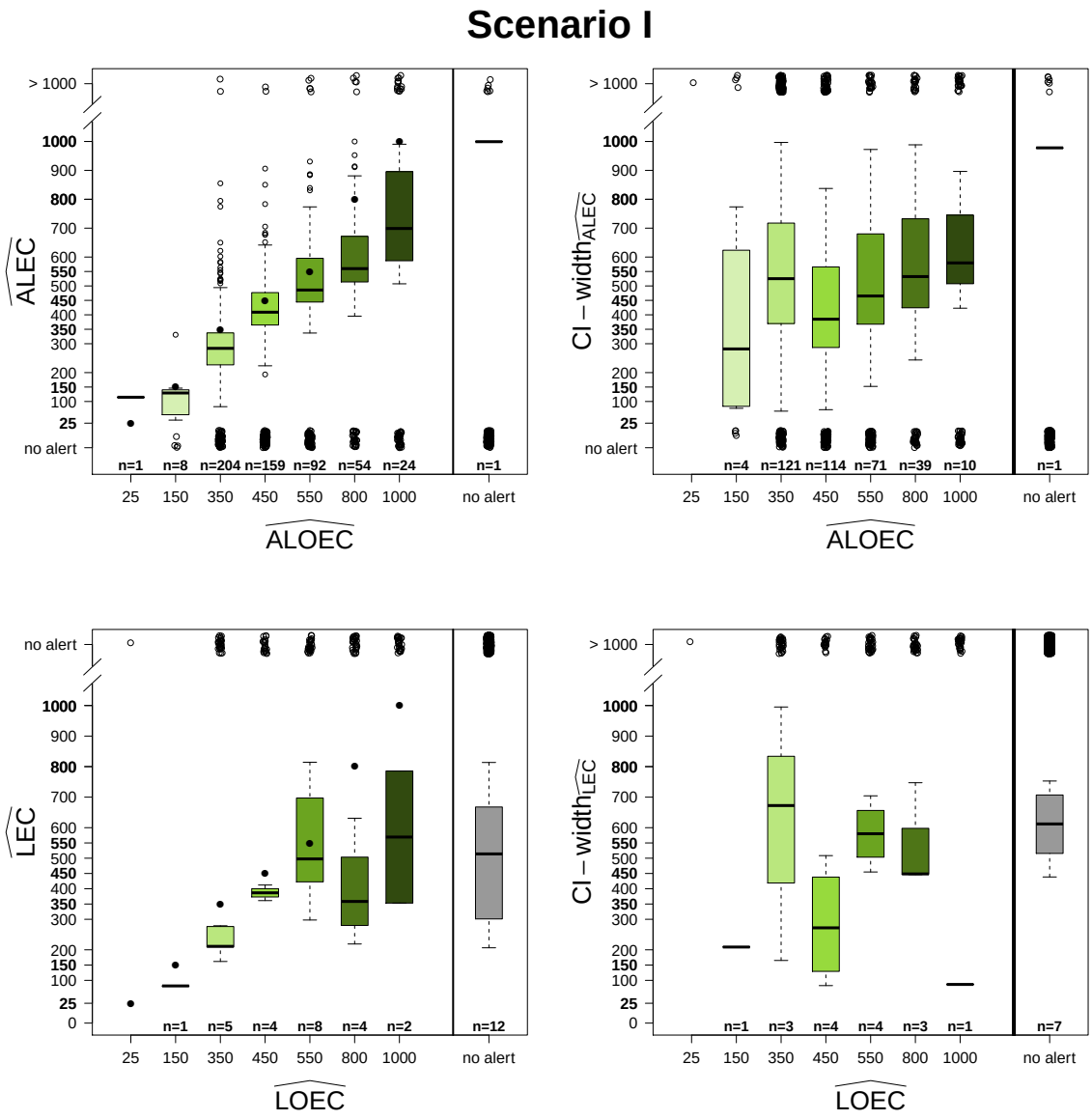


Figure 6.7: *Boxplots comparing the alert concentrations obtained with the 4pLL method with those obtained with the Limma method. Rows indicate the cut-off criteria: The upper row shows the ALEC estimators and their 95%-CIs with respect to the $\widehat{\text{ALOEC}}$s. The lower row shows the same split for the LEC and LOEC estimators. The results of the 4pLL method are depicted on the y-axis, the estimates of the Limma method on the x-axis. Black points indicate equal values on the x- and y-axis. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the conditions of Scenario IV. The blue dashed line indicates the true ALEC value.*

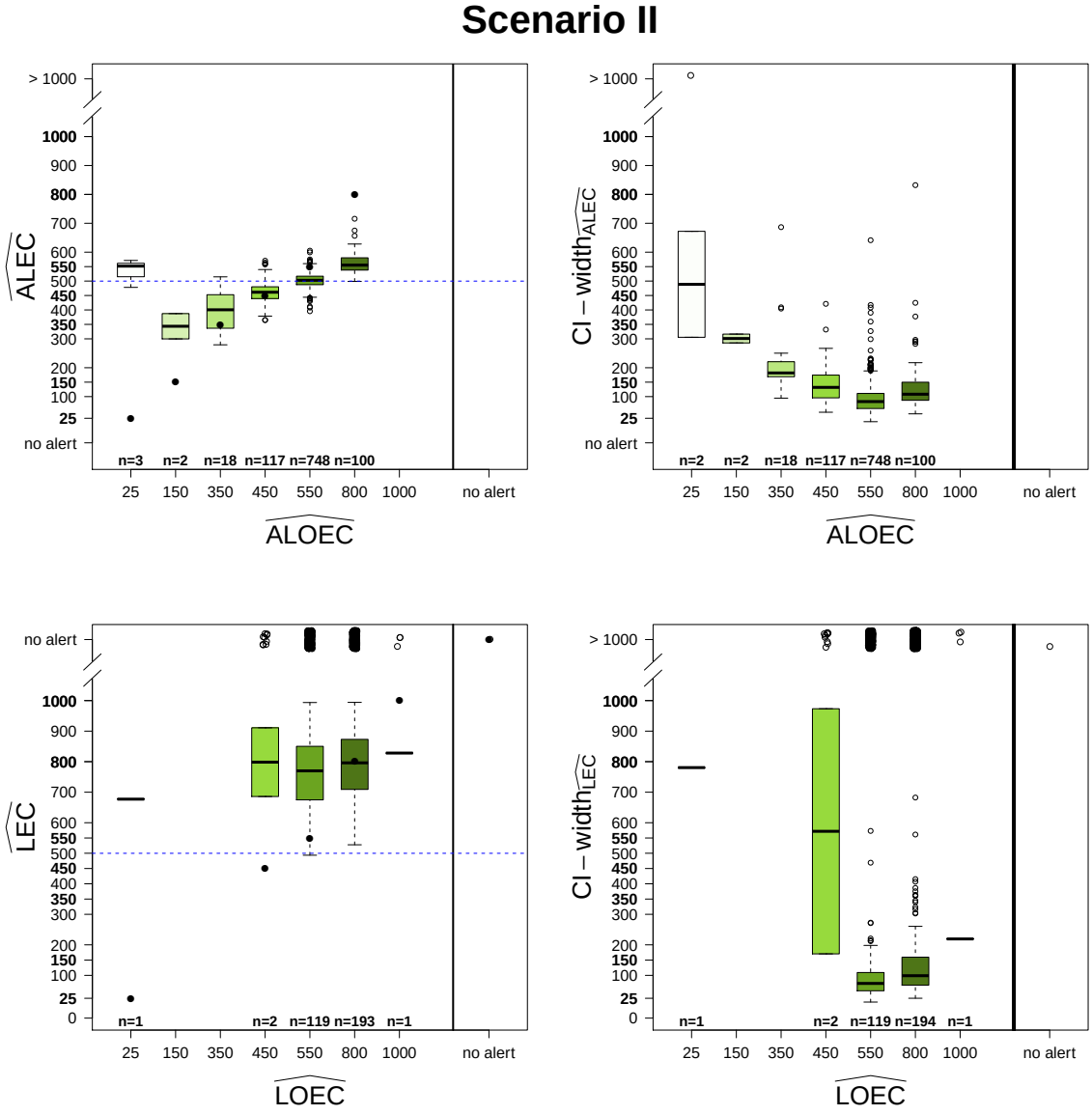


Figure 6.8: *Boxplots comparing the alert concentrations obtained with the 4pLL method with those obtained with the Limma method. Rows indicate the cut-off criteria: The upper row shows the ALEC estimators and their 95%-CIs with respect to the $\widehat{\text{ALOEC}}$s. The lower row shows the same split for the LEC and LOEC estimators. The results of the 4pLL method are depicted on the y-axis, the estimates of the Limma method on the x-axis. Black points indicate equal values on the x- and y-axis. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the conditions of Scenario IV. The blue dashed line indicates the true ALEC value.*

of 25 µM. Moreover, for this subgroup of ALEC estimates the narrowest confidence intervals are obtained (lengths ranging from 25 µM to 100 µM).

140 genes that have been deregulated at early concentrations (25 µM to 450 µM), according to *Limma*, turn out to be false positive alerts, i.e. $\widehat{\text{ALOEC}}$s with values lower than 500 µM. For most of these genes ($n = 117$) *Limma* notices a change at 450 µM, the next lower value measured before the true 500 µM. For this group of genes the 4pLL approach notices expression changes at much higher concentrations.

Considering both criteria, fold change and p-value, shows that a change in gene expression is confirmed to be significant in 16.7% of the cases (lower panel in the figure). For all these genes, except of one, *Limma* provides $\widehat{\text{LOEC}}$s of 550 µM and 800 µM. The 4pLL method, in contrast, provides estimators for the LEC, which are on average much higher than or equal to the $\widehat{\text{LOEC}}$s and, thus, deviate more from the true ALEC value (500 µM) than the $\widehat{\text{LOEC}}$s. The corresponding 95%-CIs for the $\widehat{\text{LEC}}$s exhibit median lenghts of about 100 µM. In this respect, it should be noted that in this scenario the 4pLL method fails in most of the cases due to inestimable confidence intervals, while the *Limma* method provides satisfying results.

Scenario III. In Scenario III the true ALEC value coincides with the measured concentration level of 550 µM. The boxplots in Figure 6.9 show the distributions of the alert concentrations in relation to each other. 489 of the estimated ALOECs exhibit values equal to 550 µM and 456 of those point estimates at 800 µM, the next higher concentration level after 550 µM. The $\widehat{\text{ALEC}}$s, on the other hand, vary closely around the true ALEC value (upper panel).

The boxplots for the CI-based estimates reveal that 45 genes with fold change values of at least 1.5 at $\widehat{\text{ALOEC}}$ values of 150 µM, 350 µM and 450 µM, respectively, do not exceed the given threshold significantly due to the *Limma* t-test (i.e. $p > 0.05$). 749 of the previously 991 point estimates meet the criteria for significance (FC ≥ 1.5 & $p \leq 0.05$). Most of the significant upregulations are detected by *Limma* at 800 µM and by the model-based approach at approximately 600 µM. The number of false positive alerts, i.e. estimates with values below 550 µM, reduces from 45 to 4 and from 523 to 35 in the case of *Limma* and the 4pLL approach, respectively, when taking p-values into account (Table 6.3). The 95%-confidence intervals obtained for the $\widehat{\text{ALEC}}$s and $\widehat{\text{LEC}}$s are mostly small in this scenario.

Of all four scenarios, Scenario III provides the most preferable curve shape (saturated sigmoid curve) for estimating the LECs with the 4pLL method. The results show that in this scenario the LEC estimates clearly outperform the LOEC estimates.

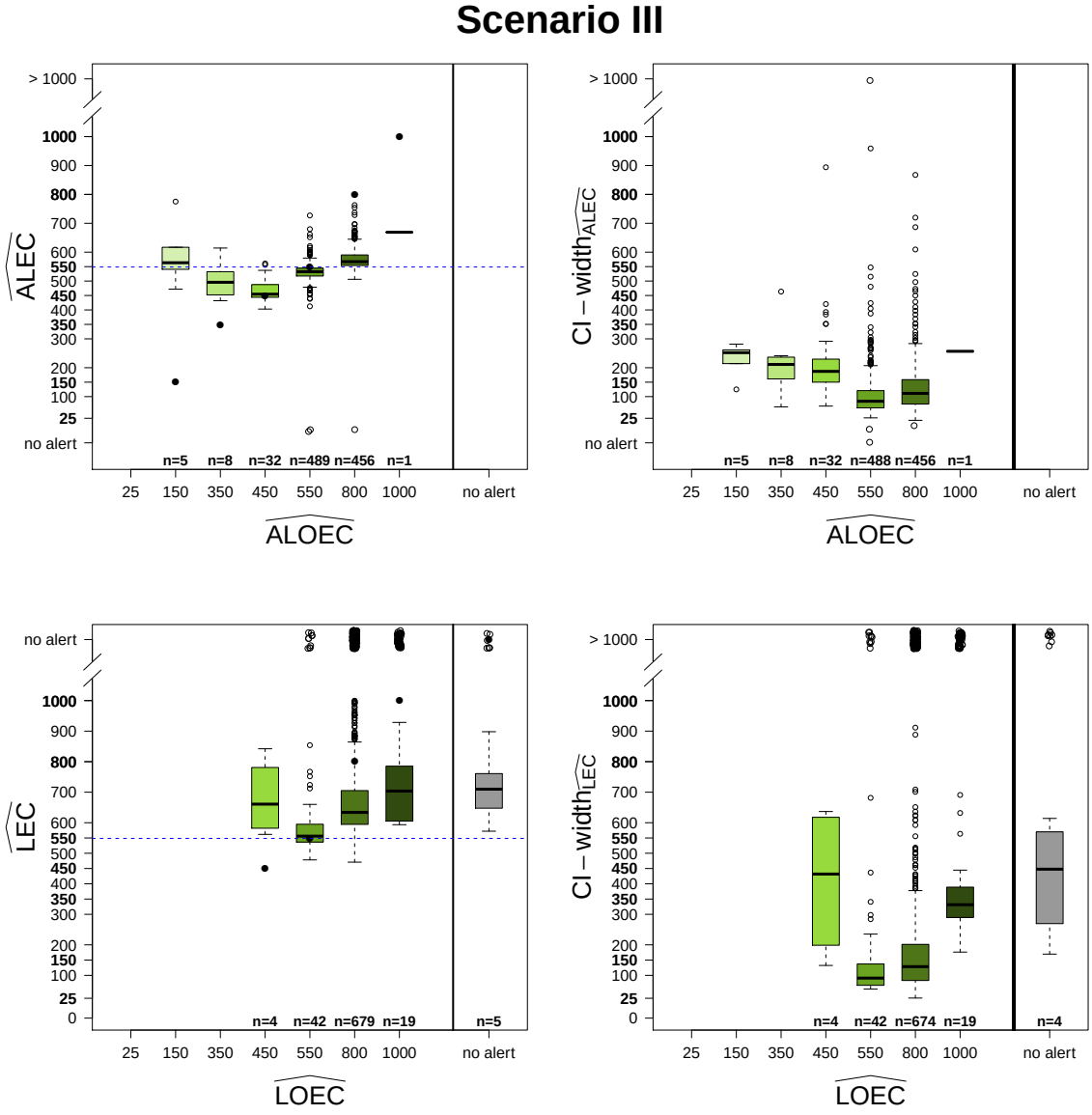


Figure 6.9: *Boxplots comparing the alert concentrations obtained with the 4pLL method with those obtained with the Limma method. The rows indicate the cut-off criteria: The upper row shows the ALEC estimators and their 95%-CIs with respect to the $\widehat{\text{ALOEC}}$s. The lower row shows the same split for the LEC and LOEC estimators. The results of the 4pLL method are depicted on the y-axis, the estimates of the Limma method on the x-axis. Black points indicate equal values on the x- and y-axis. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the conditions of Scenario III. The blue dashed line indicates the true ALEC value.*

Scenario IV. In Scenario IV the true ALEC value is 680 µM. The results are illustrated by the boxplots in Figure 6.10. Due to *Limma*, most of the genes (n = 768) show a change in gene expression at 800 µM, the next higher value measured after the true ALEC value (upper panel). 748 of 926 $\widehat{\text{ALOEC}}$s meet, on average, the true ALEC value exactly. Restricting the analysis to those genes for which the t_{4pLL}-test yields p-values of less than 0.05 results in 285 estimators of which almost all take values of above 680 µM (lower panel). In addition, LEC estimators are obtained for genes that show no significant change according to *Limma* ($n = 41$). 4pLL modeling of these expression profiles results in the widest $\widehat{\text{LEC}}$ confidence intervals so far. Similar lengths have only been observed for the point estimates in Scenario I.

All in all, Scenario IV is unfavorable for both methods since both approaches have difficulties to provide accurate estimates.

In addition, it was counted how often the true ALEC value was captured by the 95%-confidence intervals (CIs) of the $\widehat{\text{ALEC}}$s. Therefore, the coverage probability (CP) which indicates the proportion of covered cases was computed. In the calculations only CIs of lengths smaller than 1000 µM were considered. The corresponding numbers are provided in Table 6.4. In case of Scenario I the coverage probability could not be computed since no ALEC value was provided. The coverage probabilities were computed for $k = 3$, $k = 6$ and $k = 10$ replicates.

The number of successfully computed CIs (widths $\leq$ 1000 µM) has increased with increasing sample size. The observed coverage probabilities are between 0.8 and 0.86 and, thus, in all cases relatively small (Table 6.4).

Table 6.4: *Coverage probabilities for the Scenarios II-IV. The coverage probability (CP) is defined as the proportion of how often the 95%-confidence interval of the $\widehat{\text{ALEC}}$s cover the true ALEC value. In case of Scenario I the coverage probability could not be computed since no ALEC value was provided. The total number of confidence intervals is denoted by n. Only confidence intervals of lengths smaller than 1000 µM are considered.*

	Scenario II		**Scenario III**		**Scenario IV**	
	n	CP	n	CP	n	CP
k=3	987	0.827	990	0.861	904	0.803
k=6	996	0.809	997	0.844	956	0.827
k=10	1000	0.818	996	0.830	963	0.849

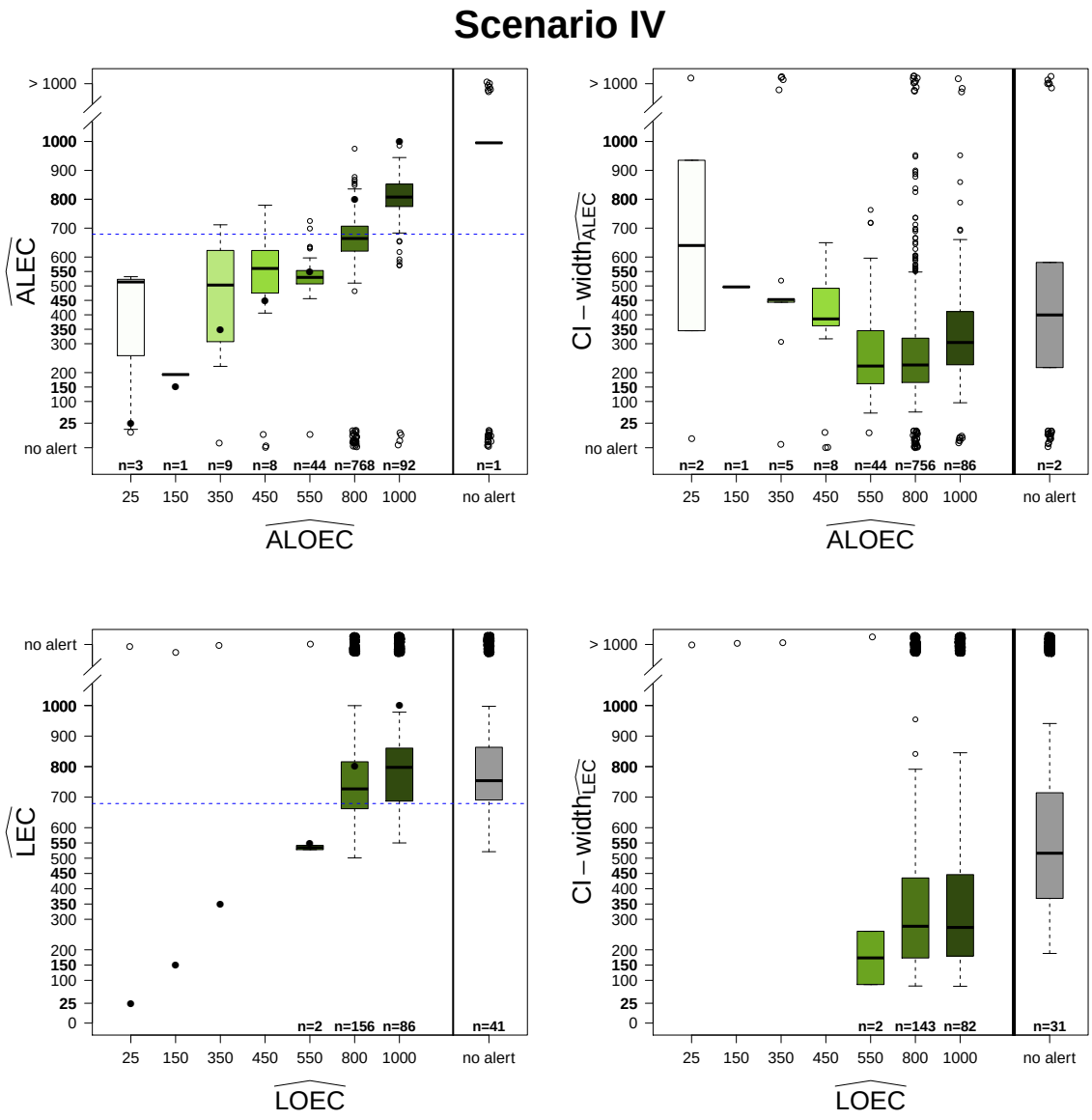


Figure 6.10: *Boxplots comparing the alert concentrations obtained with the 4pLL method with those obtained with the Limma method. The rows indicate the cut-off criteria: The upper row shows the ALEC estimators and their 95%-CIs with respect to the $\widehat{\text{ALOEC}}$s. Lower row shows the same split for the LEC and LOEC estimators. The results of the 4pLL method are depicted on the y-axis, the estimates of the Limma method on the x-axis. Black points indicate equal values on the x- and y-axis. The alert concentrations were estimated from the simulated data with $k = 3$ replicates per concentration under the conditions of Scenario IV. The blue dashed line indicates the true ALEC value.*

6.3 Results of a real data study

The VPA chronic study, which was introduced in Section 2.4, is used as real data example. The study was conducted to investigate concentration-dependent gene expression changes in response to chemical exposure. The cells were treated *in vitro* with valproic acid (VPA) using eight different concentrations (25, 150, 350, 450, 550, 650, 800, 1000 µM) and three replicates per concentration.

To obtain an overview of the data, principal component analysis (PCA) has been performed, based on the 100 probe sets with highest variance across all replicates, to visualize the different gene expression profiles of all experiments (Figure 6.11). Each point represents one experiment, where the color indicates the concentration and the form indicates the replicate. The percentages of the variances covered are plotted on the axes.

Plotting of the first two principal components shows that the three replicates for each concentration all cluster closely together, and the concentrations can be clearly distinguished. The treated samples move in the direction of the first principal component, which explains almost 90% of the data variability, and hence, represents a convincing concentration progression. Only the concentration of 650 µM, which is in the range of beginning cytotoxicity, shows a high variability between the three replicates (upper panel, purple color coding). Therefore, this concentration is excluded from further analysis. The lower panel in Figure 6.11 thus shows that the data quality could be improved by excluding the measurements with 650 µM.

The flowchart in Figure 6.12 illustrates the approach to analyze the data. The goal is to detect concentrations with critical changes in gene expression. Analogously to the procedure in the simulation study, the two methods for estimating critical concentrations, the 4pLL model approach and the *Limma* method, are compared with each other.

On the one hand, the 4pLL model approach is applied to estimate the concentration at which the fitted curve intersects the given threshold (ALEC). On the other hand, the *Limma* method is used to obtain the lowest observed concentration at which the mean fold change exceeds the threshold (ALOEC). The same analysis is performed to detect the lowest concentration at which the critical effect level is exceeded significantly, i.e. by means of hypothesis testing (LEC and LOEC). The t_{4pLL}-test is used for estimating the LEC and the *Limma* t-test for specifying the LOEC estimator. The critical effect level is set to 1.5-fold. The analysis is not performed on the entire set of genes, which includes 54 675 probe sets from the *HG-U133* GeneChip, but is

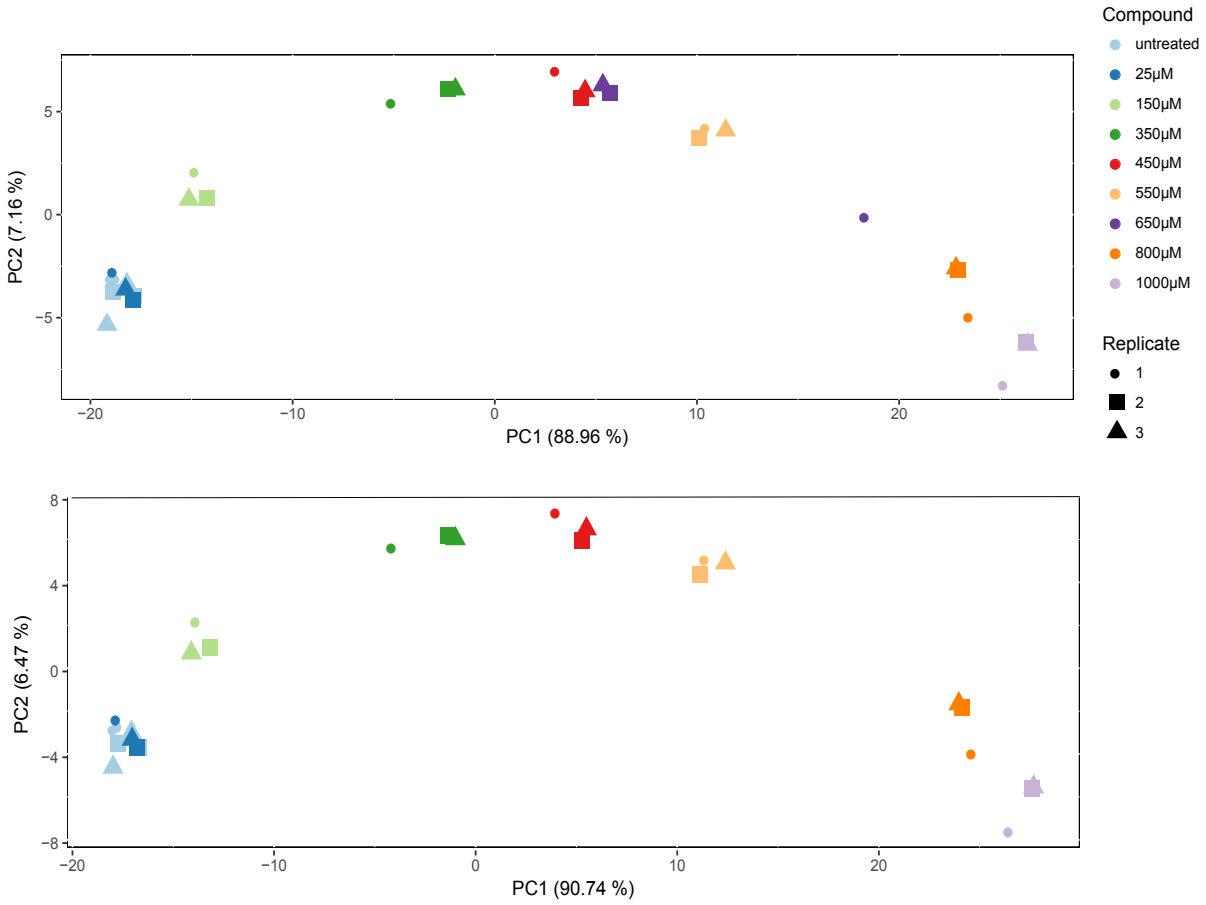


Figure 6.11: *Principal component analysis (PCA), based on the 100 probe sets with highest variance across all replicates, was performed for the VPA chronic concentration study to visualize the data structure across all concentrations and experimental replicates. The upper panel shows the PCA plot for the complete VPA study (including 650 µM) and the lower panel illustrates the concentration progression of the data after excluding experiments with concentration 650 µM. Each point represents one experiment, where the color indicates the concentration and the form indicates the replicate. The percentages of the variances covered are plotted on the axes.*

restricted to those genes that show a significant change in gene expression for at least one of the concentration levels. For this, an analysis of variance (ANOVA) was performed in advance. The 9460 genes with a significant result ($p < 0.001$) were kept for the analysis. In the following, the estimated alert concentrations are compared with respect to their distributions (Figure 6.13).

The left-hand histograms of Figure 6.13 show the distributions of the alert concentrations detected with the 4pLL method and the barplots on the right-hand side of the figure show the distributions of the classical estimates (*Limma*). In the upper row the distributions of the point estimates are presented, while those of the CI-based estimates are displayed in the lower panel. The number of successfully converged estimators (estimates $\leq$ 1000 µM) is denoted by n and given in the upper right-hand corner of the respective distribution. *No alerts* and alerts of values

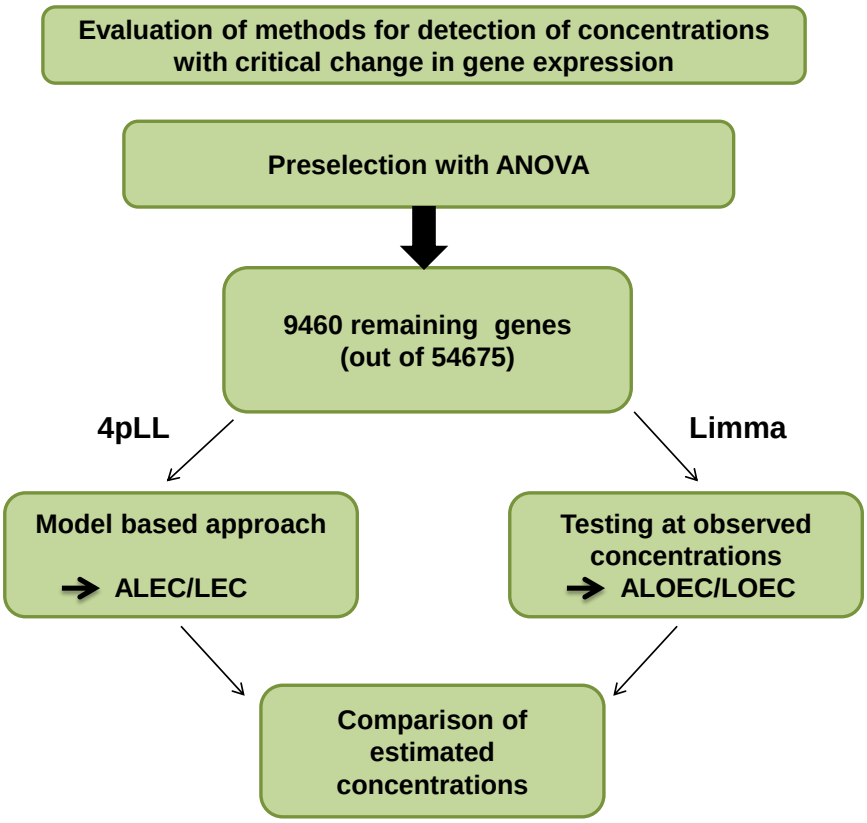


Figure 6.12: *Flowchart of the analytical procedure for the detection of critical changes in gene expression.*

higher than 1000 µM are not shown in the histograms. In the evaluation of the *Limma* results, only estimates with unique solutions are considered, i.e. expression patterns exceeding both the upper and lower limit were excluded before. In case of the ALOEC estimates 75 genes were affected from the exclusion criteria and in case of the LOEC estimates 7 genes were excluded.

For 6756 (71.4%) expression profiles an ALEC estimator could be computed. For a few more genes, 7191 (76.0%) in total, the *Limma* method has detected a 1.5-fold change in gene expression. In more than 2000 cases both methods have not noticed an alert. According to *Limma*, most of the genes ($n = 2112$) have been deregulated at 800 µM. 25% of the $\widehat{\text{ALOEC}}$s have values below 450 µM. The middle 50% of the $\widehat{\text{ALOEC}}$s lie between 450 µM and 1000 µM, while the $\widehat{\text{ALEC}}$s vary around a median value of 520.5 µM with an interquartile range of 324.7 µM and a standard deviation of 220.8 µM. 75% of the $\widehat{\text{ALEC}}$s exhibit values below 695.5 µM.

The lower panel in Figure 6.13 shows the results of the CI-based estimates. The number of *no alert* signals has doubled for the *Limma* method (increase from $n_{\widehat{\text{ALOEC}}} = 2194$ to $n_{\widehat{\text{LOEC}}} = 4745$) and almost tripled for the 4pLL method (increase from $n_{\widehat{\text{ALEC}}} = 2704$ to $n_{\widehat{\text{LEC}}} = 7449$), while

the number of LEC and LOEC estimates has enormously decreased from 6756 to 2011 (4pLL) and 7191 to 4708 (*Limma*) estimates, respectively. The median of the $\widehat{\text{ALOEC}}$s has shifted from 800 µM to *no alert*, by disregarding the *no alert* signals the median of the $\widehat{\text{LOEC}}$s is still 800 µM. The median of the ALEC estimates has shifted slightly from 521 µM to 505 µM ($\widehat{\text{LEC}}$).

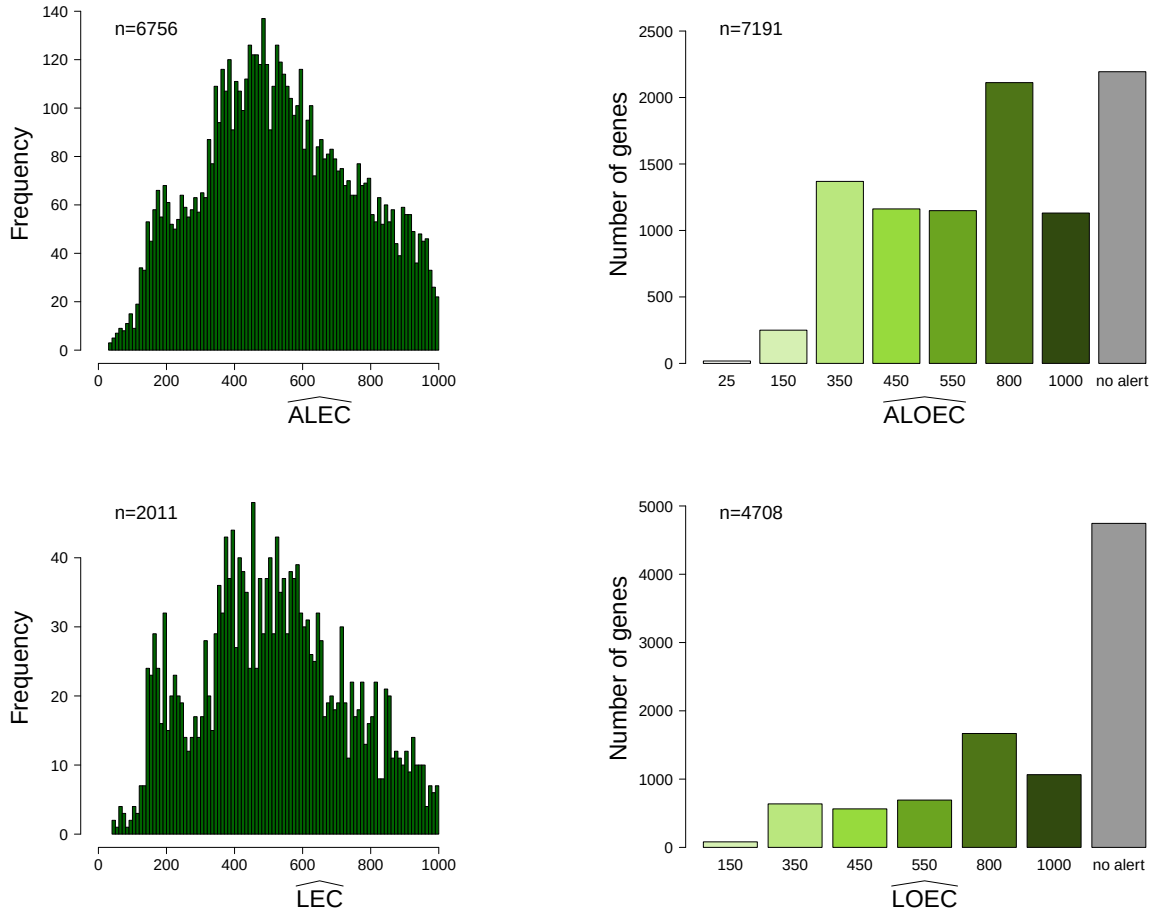


Figure 6.13: *Distributions of the estimated alert concentrations for the VPA chronic concentration study. Rows indicate the cut-off criteria and columns the methods of estimation. The left panel shows the distributions of the ALEC and LEC estimators (4pLL) and the right panel shows the distributions of the ALOEC and LOEC estimators (Limma). The number of estimates* $\leq$ *1000 µM is indicated by* n.

Figure 6.14 displays the distribution of the quantiles which were calculated from the distribution of the $\widehat{\text{ALEC}}$s and $\widehat{\text{LEC}}$s representing the $\widehat{\text{ALOEC}}$s and $\widehat{\text{LOEC}}$s, respectively. Remember, that, given the distribution of the 4pLL estimates, it can be calculated which quantiles of that distribution the *Limma* estimates correspond to. The histogram in the left panel illustrates the results for the point estimators and the histogram in the right panel shows the results for the significant alert concentrations. The calculation of the quantiles is based only on the successfully converged ALEC and LEC estimators, respectively (i.e. estimates $\leq$ 1000 µM). The left-hand

histogram exhibits a peak at one, and thus, indicates that most of the $\widehat{\text{ALOEC}}$s take values which are higher than the values of the $\widehat{\text{LOEC}}$s. The histogram on the right-hand side of the figure shows the distribution with one pronounced peak at one and one smaller peak at zero. That indicates, that, in comparison to the *Limma* approach, the 4pLL method has detected a change in gene expression in 705 cases at lower concentrations and in 167 cases at higher concentrations. The other 1139 quantile values are distributed uniformly within the interval $(0, 1)$.

It can be stated, that for both cut-off criteria, the stringent (FC ≥ 1.5 & $p \leq 0.05$) and the less stringent one (FC ≥ 1.5), the case study shows a tendency towards lower $\widehat{\text{LEC}}$ values than $\widehat{\text{LOEC}}$ values. That means, that the 4pLL method generally indicates alerts at lower concentrations when compared with the alerts of the classical *Limma* approach.

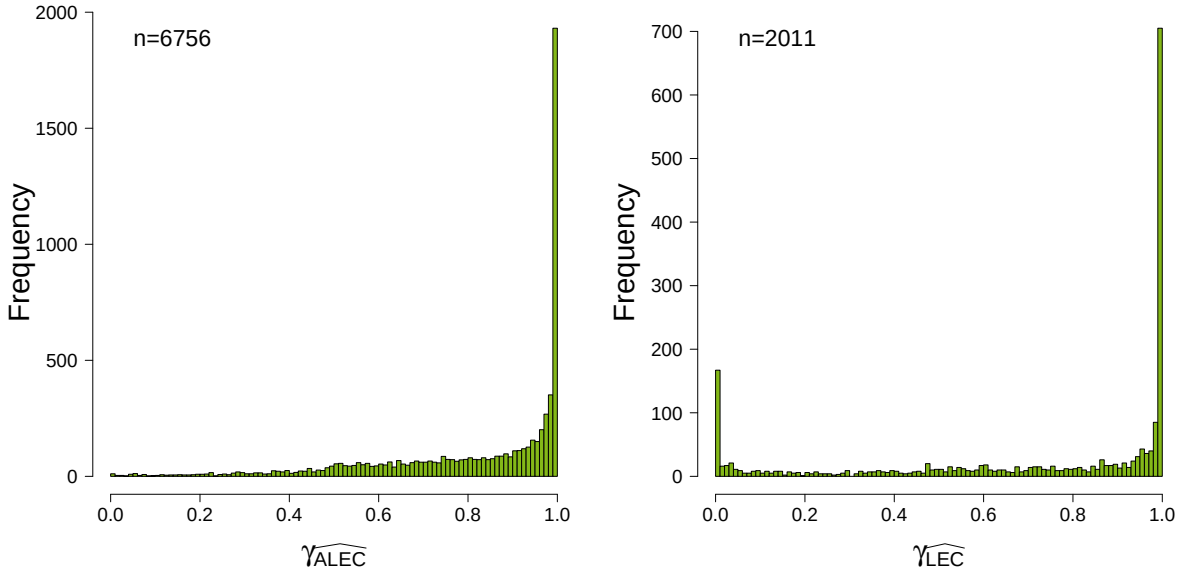


Figure 6.14: *Distributions of the quantiles calculated from the distribution of the $\widehat{\text{ALEC}}$s and $\widehat{\text{LEC}}$s representing the $\widehat{\text{ALOEC}}$s (left panel) and $\widehat{\text{LOEC}}$s (right panel), respectively. The alert concentrations were estimated from the VPA chronic concentration study. Values close to zero indicate that the Limma method detects alerts at lower concentrations than the 4pLL method, values close to one indicate the reverse case.*

The boxplots in Figure 6.15 show the model-based estimates in direct comparison to the classical estimates. The estimates obtained with the 4pLL approach are depicted on the y-axis, while the estimates of the *Limma* approach are displayed on the x-axis. The estimated ALECs and LECs, and their 95%-confidence intervals, are subdivided with respect to the estimated ALOECs and LOECs, respectively. Boxplots have only been generated for estimates for which the 4pLL model has successfully converged (estimates ≤ 1000 µM). *No alert* detections as well as confidence interval lengths of above 1000 µM are highlightened as extreme data points at the

top of the single boxplots. Similar to the figure setups before, the plots in the upper row refer to the distributions of the point estimators and the plots in the lower panel compare the distributions of the CI-based estimates.

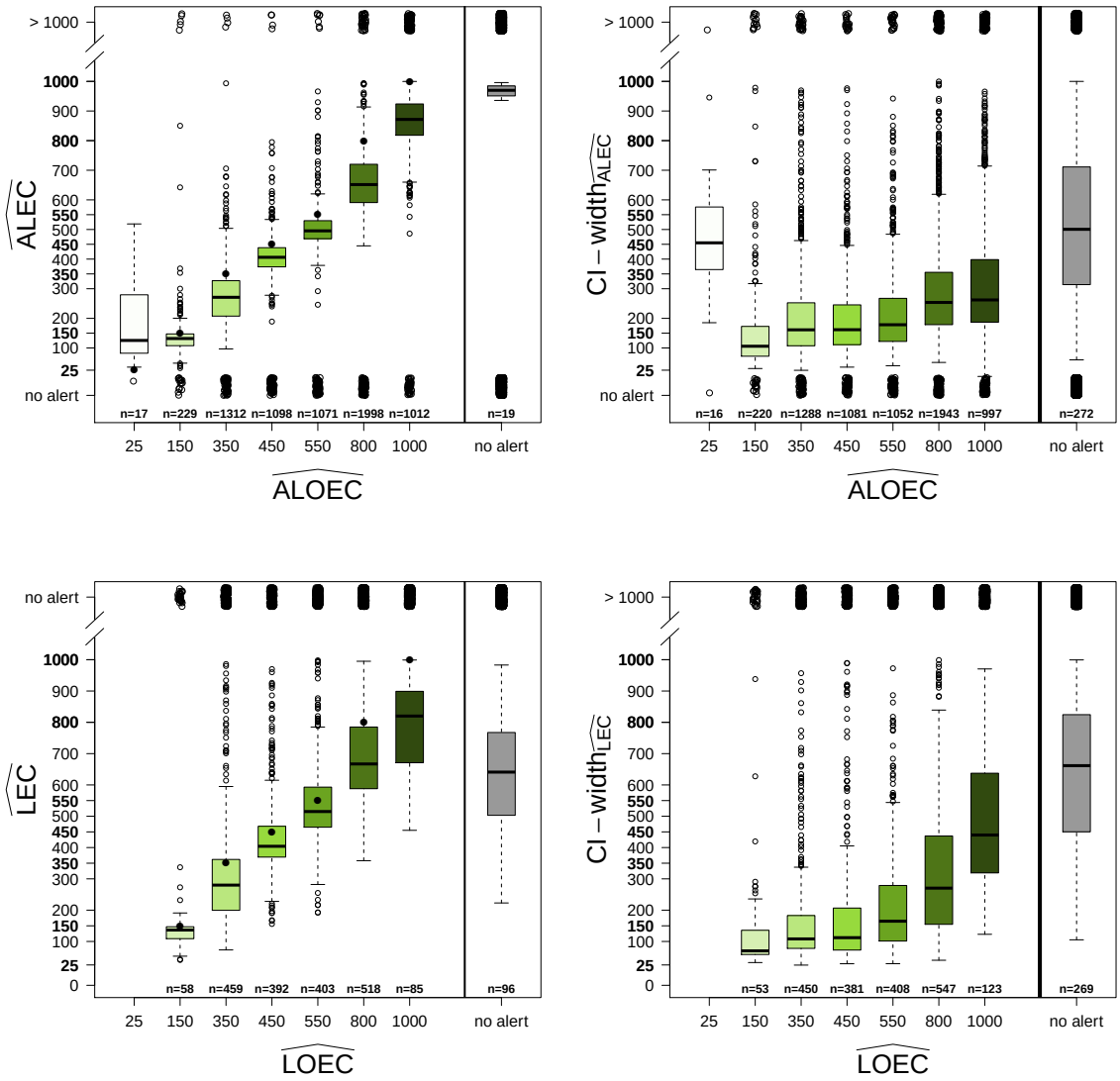


Figure 6.15: *Boxplots comparing the alert concentrations obtained with the 4pLL method with those ones obtained with the Limma method. Rows indicate the cut-off criteria: Upper row shows the ALEC estimators and their 95%-CIs with respect to the $\widehat{\text{ALOEC}}$s. Lower row shows the same split but for the LEC and LOEC estimators. The results of the 4pLL method are depicted on the y-axis, the estimates of the Limma method on x-axis. The alert concentrations were estimated from the VPA chronic concentration study.*

The boxplots in the upper left-hand corner reveal that for almost all observed concentrations the respective boxes are below the indicated black points, indicating that 75% of the $\widehat{\text{ALEC}}$s exhibit values below the values of the respective classical estimates ($\widehat{\text{ALOEC}}$s). The only exception is the lowest measured concentration (25 µM) where the entire boxplot lies above the

value of 25 µM. That means, that this group of genes ($n = 16$) exhibits, according to *Limma*, expression changes at much lower concentrations than according to the 4pLL method, which has first noticed changes at levels between 30 µM and 550 µM. However, the widths of the respective 95%-confidence intervals are broader than those obtained for the higher concentrations (> 25 µM). This gives reason to assume a high insecurity in the estimation of the ALECs. A similar degree of uncertainty can be observed for the ALEC estimates that show *no alert* according to *Limma*. Those ALEC estimates ($n = 19$) exhibit confidence intervals of similar large interquartile ranges (300 µM-700 µM). In contrast, the $\widehat{\text{ALECs}}$ observed for the concentrations 150, 350, 450, 550, 800 and 1000 µM yield confidence intervals of lower widths (on average 100 µM-200 µM).

When considering the LEC estimates (lower panel of Figure 6.15), it is important to mention, that the model-based approach fails in most of the cases in terms of convergence. For the few cases, in which the 4pLL model has converged (21.3%), the boxplots in the lower left-hand corner show, that the relation between the $\widehat{\text{LECs}}$ and $\widehat{\text{LOECs}}$ remains largely the same, except for some minor shifts in the distribution of the CI-based estimates. Early alerts (alerts at 25 µM) are not proved to be significant, but the *Limma* method provides more *no alert* signals than before (for the $\widehat{\text{ALOECs}}$). In 75% of the cases the model-based approach yields concentrations of lower values than the *Limma* method. The uncertainties in the parameter estimates increase with increasing concentrations. Thus, the widest confidence intervals are obtained for high concentration levels ($\geq$ 550 µM).

Note, that if the four-parameter-log-logistic (4pLL) model is misspecified, i.e. if the parametric assumptions do not hold and the chosen model cannot capture the structure in the data, the estimates might be biased. Hence, the results should be interpreted with caution.

7 Summary

The present work focused on the following three topics that often arise in the context of gene expression analysis: Firstly, the identification and characterization of different modes of action associated with certain expression changes, secondly, the identification of *in vitro* biomarkers for the prediction of toxicity *in vivo* and thirdly, the identification of critical concentrations at which pre-defined effect levels are exceeded.

For identifying molecular changes on a genome-wide scale as a response to chemical exposure within the same species and between *in vitro* and *in vivo* systems the open-source Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (TG-GATEs) was used. The database covers more than 150 compounds and compiles Affymetrix files of rat liver samples and *in vitro* cultivated human and rat hepatocytes. The cells were treated with each compound using different concentrations, most of them at three distinct concentrations, and for different time periods.

Regarding the first thesis topic, the main task was to identify general principles of the toxico-transcriptome in human hepatocytes. At first, the database had to be curated. This is a crucial and necessary step to reduce errors in the detection process, such as false positives, false negatives or undetected effects. In the context of concentration-dependent gene expression analysis, batch effects, limited numbers of replicates and implausible concentration progressions can cause such misleading results. A principal component analysis of the investigated data, based on the 100 strongest deregulated genes across all compounds, revealed several clusters which could be removed by simple control subtraction (compound-matched). The results suggest that the batch effects occurred as a consequence of experimental variability. The reproducibility between replicates was assessed by the comparison of the distances between replicates and control-treatment pairs. The analysis has shown that the distances between replicates was small in relation to the much larger compound induced distances. Compounds with implausible concentration progressions, i.e. with a high fraction of genes deregulated at a lower but not at the respective higher concentration, were excluded from the analysis. For this curation step the *progression profile error indicator* was defined to detect deviations from monotonous concentration progressions.

This curation procedure was summarized into a guideline which is now recommended as a necessary step before analyzing high-dimensional toxicogenomic data. The curated data was used to categorize genes according to the following key features: (1) *Stereotypical vs. compound-specific stress response*: To differentiate between these two kinds of responses, the selection value (Sv) was introduced, which specifies for a gene the minimum number of compounds that induces an expression change. Since it was observed that only 32 of 148 compounds have induced marked effects and a single compound can lead to false positive results, genes that were deregulated by at least three and at most 19 compounds, were used to describe compound-specific responses (Sv 3), while genes altered by 20 or more compounds were categorized as stereotypical genes (Sv 20). Conversely, this means that Sv 20 genes are always a subset of the Sv 3 genes. (2) *Liver disease-associated genes*: It could be shown that the reported stereotypical stress response genes overlapped with genes deregulated in human liver diseases, such as steatohepatitis, liver cirrhosis and hepatocellular cancer. (3) *Unstable baseline genes*: A group of unstable baseline genes was identified which were deregulated not by a compound but simply due to the procedure of isolating and cultivating. (4) *Biological function*: More than 2000 genes were associated with biological functions. The results of this analysis were stored in a toxicotranscriptomics directory that is now publicly available under `http://wiki.toxbank.net/toxicogenomics-map/`. It can be employed by toxicologists to obtain basic information of chemically-induced expression responses in human hepatocytes.

Regarding the second thesis topic, detecting biomarkers in *in vitro* systems for the prediction of *in vivo* toxicity, the next step should be to validate the aforementioned results in terms of comparable responses in *in vivo* systems. The latter includes the step to investigate whether the genes that are chemically induced in cultivated hepatocytes are also deregulated in human liver tissue. Since the data investigated stems from human hepatocytes, such an analysis cannot be performed, for obvious ethical reasons. To remedy this problem, a comparative analysis between cultivated rat hepatocytes and rat liver samples was performed to assess the relevance of *in vitro* responses for the *in vivo* situation. To this end, two databases, the NRW and TG-GATEs data sets, comprising *in vivo* and *in vitro* gene expression data of rat hepatocytes for a total number of 189 different compounds, were used, of which only 5 compounds were present in both data sets. The data curation pipeline introduced in this work was applied to both data sets. Data variability among the *in vitro* experiments was observed to be much higher compared to the *in vivo* data. Due to the heterogeneous data structure in *in vitro* data, the *in vivo* data was considered the gold standard to which the *in vitro* results were compared. Concentration progression analysis

revealed inconsistencies in parts of both data sets. To ensure reliable results in the detection of mutual gene deregulations, the Sv 3 concept was applied to generate a consensus signature of commonly regulated genes. To this end, consensus Sv 3 lists were generated to analyze the *in vivo-in vitro* response within and between the two databases.

Consensus analysis of the NRW and TG-GATEs database, revealed an *in vitro-in vivo* overlap which was more than random. Despite the significant overlap, a large fraction of genes was still in the non-overlapping region inidcating *in vivo* and *in vitro* specific responses. Considering this aspect, a direct deduction from *in vitro* to *in vivo* systems is difficult to accomplish. This is in line with the findings of previous studies which revealed substantial differences between the two test systems (Schug et al., 2013).

However, the comparison of the two data sets showed a relatively high degree of similarity, considering the fact that only five mutual compounds were tested. The consensus genes of the NRW-TG-GATEs overlap were further analyzed with respect to their response to different test compounds. They were ordered with respect to the selection value in order to find a subset of as few genes as necessary to depict a compound sensitivity as large as possible. This resulted in 11 up- and downregulated biomarker genes covering most of the high profile compounds. These genes allow for a prediction of toxicity in *in vivo* experiments.

Hence, differential expression analysis was only performed using the classical naïve approach, where for each measured concentration separately it was tested if the critical effect level was exceeded. This procedure has an inherent fundamental flaw in the sense that it is highly unlikely that such a deregulation is first triggered at exactly one of the measured concentrations. To this end, a model-based method was introduced in this thesis.

Based on the assumption that the response dependency of the dose can be described by a sigmoidal function, Jiang (2013) fitted a four-parameter log-logistic (4pLL) model to dose-response data to estimate the Absolute Lowest Effective Concentration (ALEC) for a fixed and pre-specified effect level. But since the ALEC results solely from a simple point estimator, the uncertainty of the effect level is entirely neglected. However, for obvious statistical reasons, it is vital to provide confidence intervals. Thus, in this thesis, the method was enhanced by means of a thorough confidence interval estimator. The critical effect level (fold change) is exceeded significantly if the entire confidence interval lies above the predefined threshold of a given fold change. The hereby resulting concentration value was introduced as the Lowest Effective Concentration (LEC). For significance testing a specific test statistic, the t_{4pLL}-test statistic, derived from the 4pLL model, was proposed.

In order to validate the 4pLL method and assess the general feasibility of model-based test methods, various gene expression profiles were simulated. In the first case the critical fold change was never met. In this case, the proposed approach (LEC) performed better than the classical approaches (ALOEC and LOEC) as well as the point estimator based method (ALEC), i.e. less false positive signals were triggered. In cases in which the expression pattern followed a pronounced sigmoidal shape distinctly crossing the threshold, the new method clearly outperformed the others. However, if the threshold was only slightly exceeded, both methods, the model-based and the classical one, had difficulties to provide accurate estimates. In cases of unsaturated curve progressions, the classical approach yielded satisfying results, while the model-based approach failed in most of the cases due to inestimable confidence intervals. By applying the method to actual experimental data, the general trend was observed that the model-based approaches yield alerts at lower concentrations than the classical approach.

The use of the 4pLL approach is recommended under the circumstances that the expression profile can be justifiably assumed to obey a saturated sigmoidal path. Given this prerequisite, the proposed method is preferential for many reasons. Firstly, the model-based approach benefits from its independence of observed measurements allowing arbitrary positive values as alert levels. Secondly, both estimates, the point and CI-based estimate, can be estimated in the case of incomplete dose-response curves due to the fact that the critical effect-level is defined independently from the lower and upper asymptote of the curve. Thirdly, the modeling of continuous gene expression profiles allows the calculation of confidence intervals for the estimated alert concentrations. In addition, the model-based approach incorporates the entire information about the dose-response relationship which is neglected by the classical procedure.

On the other hand, the new method provides biased estimates if the given prerequisites do not apply to the data. In order to obtain reasonable estimates, it is recommended to test explicitly for sigmoidal functions, more precisely for deviations from sigmoidal curve progressions (Schmoyer, 1984). As the model choice is decisive for the assessment of estimation uncertainty, the search for appropriate model candidates is worth further examination. The proposed method can be extended and applied to other parametric dose-response models, such as the Log-normal- or Weibull model. Fitting different parametric models to the data is one way, another is to fit non-parametric models. The latter procedure has proven to be beneficial in case the parametric form of the dose-response curve cannot be specified. Commonly used non-parametric methods include kernel regression (Müller and Schmitt (1988) and Staniswalis and Cooper (1988)) or local linear regression (Kelly and Rice (1990) and Zhang et al. (2013)). On the one hand, non-parametric

methods can capture structure in the data that a misspecified parametric model cannot, but on the other hand non-parametric techniques often result in estimates with high variance.

Thus, both methods, parametric and non-parametric methods, have their advantages and disadvantages. When the parametric assumption holds, the parametric model yields the highest possible efficiency but when the assumption is violated, the corresponding parameter estimates will be biased with increased variances. Non-parametric models are more robust compared with parametric models but less efficient. As a compromise, a semi-parametric approach can be taken inheriting both efficiency and robustness from the two methods. The idea of a mixture model is to use a linear combination of the two fits to retain the advantages of parametric and non-parametric models. Yuan and Yin (2011) propose an estimator of a dose-response curve which is a weighted average of the parametric and non-parametric curve estimates. The weight is chosen by minimizing the mean integrated square error (MISE) such that a higher weight is given to the estimate that fits the data better. In case of a correctly specified parametric model, the semi-parametric estimate assigns more weight to the parametric estimate, in case of a misspecification a higher weight is given to the non-parametric estimate. Furthermore, they distinguish between a global and a local semi-parametric estimate. The global method assigns a constant weight to the parametric estimate according to the global fit of the parametric model, while the local method allows the weight to vary according to the local fit of the two models.

Nottingham and Birch (2000) linearly combine a logistic regression (parametric method) with a local linear regression (non-parametric method) by using a mixture parameter. The proposed method, known as model-robust quantal regression (MRQR), stabilizes the fit of an inadequate parametric model by incorporating useful information from the non-parametric model. Alternative model-robust procedures are presented in Olkin and Spiegelman (1987), Mays et al. (2000) and Pickle et al. (2008), among others. Robinson et al. (2010) presents a semi-parametric approach for the case when no replication is available. The proposed techniques all base on a convex combination of a parametric and a non-parametric model.

Moreover, Rahman et al. (1997) and Wooldridge (1992) have suggested test procedures for testing a functional form against non-parametric alternatives. Further tests on semi-parametric models have been discussed in Davidson and MacKinnon (1981), Yatchew (1992) and Eubank and Spiegelman (1990). To date, no unified approaches or guidelines have been developed in the gene expression context that address the issue of model selection. Along with this, the reported results have reinforced the general aspect of giving model insecurity in estimation processes much more importance in the future.

Bibliography

Affymetrix (2017): Biology for a better world. URL `http://images.slideplayer.com/15/4665315/slides/slide_17.jpg`. Retrieved March 28, 2017.

BK101 (2017): Basic Knowledge 101. URL `http://www.basicknowledge101.com/categories/dna.html`. Retrieved March 28,2017.

Chang, K.-M., Harbron, C., and South, M. (2016): *RefPlus: A function set for the Extrapolation Strategy (RMA+) and Extrapolation Averaging (RMA++) methods.*. R package version 1.44.0.

Davidson, R. and MacKinnon, J. (1981): Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 49 (3), 781–793.

Ellinger-Ziegelbauer, H., Gmuender, H., Bandenburg, A., and Ahr, H. (2008): Prediction of a carcinogenic potential of rat hepatocarcinogens using toxicogenomics analysis of short-term in vivo studies. *Mutation Research*, 637 (1-2), 23–39.

Eubank, R. L. and Spiegelman, C. H. (1990): Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, 85 (410), 387–392.

Grinberg, M., Stöber, R., Edlund, K., Rempel, E., Godoy, P., Reif, R., Widera, A., Madjar, K., Schmidt-Heck, W., Marchan, R., Sachinidis, A., Spitkovsky, D., Hescheler, J., Carmo, H., Arbo, M., van de Water, B., Wink, S., Vinken, M., Rogiers, V., Escher, S., Hardy, B., Mitic, D., Myatt, G., Waldmann, T., Mardinoglu, A., Damm, G., Seehofer, D., Nüssler, A., Weiss, T., Oberemm, A., Lampen, A., Schaap, M., Luijten, M., van Steeg, H., Thasler, W., Kleinjans, J. S., Stierum, R., Leist, M., Rahnenführer, J., and Hengstler, J. (2014): Toxicogenomics directory of chemically exposed human hepatocytes. *Archives of Toxicology*, 88, 2261–2287.

Harbron, C., Chang, K.-M., and South, M. (2007): Refplus: an r package extending the rma algorithm. *Bioinformatics*, 23 (18), 2493–2494.

Irizarry, R., Bolstad, B., F., C., Cope, L., Hobbs, B., and Speed, T. (2003a): Summaries of affymetrix genechip probe level data. *Nucleic Acids Research*, 31 (4), 1–8.

Irizarry, R., Hobbs, B., Collin, F., Beazer-Barclay, Y., Antonellis, K., Scherf, U., and Speed, T. (2003b): Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4 (2), 249–264.

Jiang, X. (2013): *Estimation of effective concentrations from in vitro dose-response data using the log-logistic model*. Ph.D. thesis, Medical Faculty of Ruprecht-Karls-University in Heidelberg.

Johnson, R. A. and Wichern, D. W. (1998): *Applied Multivariate Statistical Analysis*. Upper Saddle River, New Jersey: Prentice Hall, 4th edition.

Kelly, C. and Rice, J. (1990): Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics*, 46, 1071–1085.

Krug, A., Kolde, R., Gaspar, J., Rempel, E., Balmer, N., Meganathan, K., Vojnits, K., Baquiet', M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S., Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J., Hothorn, L., Bremer, S., van Thriel, C., Krause, K.-H., Hengstler, J., Rahnenführer, J., Leist, M., and Sachinidis, A. (2013): Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Archives of Toxicology*, 87, 123–143.

Lobanov, A., Turanov, A., Hatfield, D., and Gladyshev, V. (2010): Dual functions of codons in the genetic code. *Critical Reviews in Biochemistry and Molecular Biology*, 45 (4), 257–265.

Mays, J., Birch, J., and Einsporn, R. (2000): An overview of model-robust regression. *Journal of Statistical Computation and Simulation*, 66 (1), 79–100.

Müller, H. G. and Schmitt, T. (1988): Kernel and probit estimates in quantal bioassay. *Journal of the American Statistical Association*, 83, 750–759.

NIBIO (2017): National Institute of Biomedical Innovation. URL `https://www.nibiohn.go.jp/english/index.html`.

NIBIOHN (2017): National Institutes of Biomedical Innovation, Health and Nutrition. URL `http://toxico.nibiohn.go.jp/english/index.html`.

NIHS (2017): National Institute of Health Sciences. URL `http://www.nihs.go.jp/english/index.html`.

Nottingham, Q. J. and Birch, J. B. (2000): A semiparametric approach to analysing doseŰresponse data. *Statistics in Medicine*, 19, 389–404.

oerpub/epubjs-demo book (2017): Enhanced eBooks in the browser. URL `http://oerpub.github.io/epubjs-demo-book/content/m46032.xhtml`. Retrieved March 28, 2017.

Olkin, I. and Spiegelman, C. H. (1987): A semiparametric approach to density estimation. *Journal of the American Statistical Association*, 82 (399), 858–865.

PBworks (2016): Online Team Collaboration. URL `http://compbio.pbworks.com/w/page/16252906/Microarray%20Normalization%20and%20Expression%20Index`. Retrieved March 28, 2017.

Pickle, S. M., Robinson, T. J., Birch, J. B., and Anderson-Cook, C. (2008): A semi-parametric approach to robust parameter design. *Journal of Statistical Planning and Inference*, 138, 114–131.

R Core Team (2015): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rahman, M., Gokhale, D., and Ullah, A. (1997): A note on combining parametric and nonparametric regression. *Communications in Statistics-Simulation and Computation*, 26, 519–529.

Rempel, E. (2015): *Analyse der hochdimensionalen toxikologischen Expressionsdaten*. Ph.D. thesis, Faculty of Statistics of the Technical University of Dortmund.

Ritchie, M., Phipson, B., Wu, D., Hu, Y., Law, C., Shi, W., and Smyth, G. (2015): limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43.

Ritz, C. (2010): Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry*, 29 (1), 220–229.

Ritz, C. and Streibig, J. C. (2005): Bioassay analysis using r. *Journal of Statistical Software*, 12 (5). URL `http://www.bioassay.dk`.

Robinson, T. J., B., B. J., and Starnes, A. (2010): A semi-parametric approach to dual modeling when no replication exists. *Journal of Statistical Planning and Inference*, 140, 2860–2869.

Schmoyer, R. L. (1984): Sigmoidally constrained maximum likelihood estimation in quantal bioassay. *Journal of the American Statistical Association*, 79, 448–453.

Schug, M. (2011): *Entwicklung eines in vitro Systems zur Untersuchung von substanzinduzierten Genexpressionsänderungen bei primären Hepatozyten der Ratte*. Ph.D. thesis, Faculty of Chemistry of the Technical University of Dortmund.

Schug, M., Stöber, R., Heise, T., Mielke, H., Gundert-Remy, U., Godoy, P., Reif, R., Blaszkewicz, M., Ellinger-Ziegelbauer, H., Ahr, H., Selinski, S., Günther, G., Marchan, R., Blaszkewicz, M., Sachinidis, A., Nüssler, A., Oberemm, A., and Hengstler, J. (2013): Pharmacokinetics explain in vivo/in vitro discrepancies of carcinogen-induced gene expression alterations in rat liver and cultivated hepatocytes. *Archives of Toxicology*, 87 (2).

Shinde, V., Hoelting, L., Srinivasan, S., Meisig, J., Meganathan, K., Jagtap, S., Grinberg, M., Liebing, J., Bluethgen, N., Rahnenführer, J., Rempel, E., Stoeber, R., Schildknecht, S., Förster, S., Godoy, P., van Thriel, C., Gaspar, J., Hescheler, J., Waldmann, T., Hengstler, J., Leist, M., and Sachinidis, A. (2017): Definition of transcriptome-based indices for quantitative characterization of chemically disturbed stem cell development: introduction of the stop-tox ukn and stop-tox ukk tests. *Archives of Toxicology*, 91, 839–864.

Smyth, G. (2005): Limma: Linear models for microarray data. *In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Gentleman, R. and Carey, V. and Dudoit, S. and Irizarry, R. and Huber, W (eds.), Springer, New York, 179–184.

Staniswalis, J. and Cooper, V. (1988): Kernel estimates of dose response. *Biometrics*, 44, 1103–1119.

Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., and Venables, W. (2015): *gplots: Various R Programming Tools for Plotting Data*. URL `http://CRAN.R-project.org/package=gplots`. R package version 2.17.0.

Wilkinson, L. and Friendly, M. (2009): The history of the cluster heat map. *The American Statistician*, 63 (2), 179–184.

Wooldridge, M. (1992): A test for functional form against nonparametric alternatives. *Econometric theory*, 8, 452–475.

Yatchew, A. (1992): Non-parametric regression tests based on least square. *Econometric Theory*, 8, 435–451.

Yuan, Y. and Yin, G. (2011): Dose-response curve estimation: A semiparametric mixture approach. *Biometrics*, 67, 1543–1554.

Zhang, H., Holden-Wiltse, J., Wang, J., and Liang, H. (2013): A strategy to model nonmonotonic dose-response curve and estimate ic50. *PLoS ONE*, 8 (8).

List of Figures

List of Tables

A Derivation

A.1 Derivation of $\nabla h(\boldsymbol{\phi})$

Let $h(\cdot)$ denote the inverse function of $f(\cdot)$, the four-parameter log-logistic model in (3.10)

$$h(\boldsymbol{\phi}) = f^{-1}(\lambda, \boldsymbol{\phi}).$$

The ALEC for a pre-specified effect level λ is then given by

$$\text{ALEC} = h(\boldsymbol{\phi}) = \phi^{(e)} \left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right)^{1/\phi^{(b)}}, \tag{A.1}$$

where $\phi^{(e)} = \exp\left(\phi^{(e)^*}\right)$ and $\phi^{(e)^*}$ denotes the logarithmized half-maximal effective concentration. The function in (A.1) can be rewritten:

$$h(\boldsymbol{\phi}) = \exp\left\{\log(h(\boldsymbol{\phi}))\right\} = \exp\left\{ \frac{1}{\phi^b} \log\left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right) + \underbrace{\log\left(\phi^{(e)}\right)}_{=\phi^{(e)^*}} \right\}. \tag{A.2}$$

The gradient of $\nabla h(\boldsymbol{\phi})$ is calculated by using the chain rule that states for two functions g_1 and g_2:

$$\nabla h(\boldsymbol{\phi}) = \nabla g_2(g_1(\boldsymbol{\phi})) \cdot \nabla g_1(h(\boldsymbol{\phi})), \tag{A.3}$$

where $g_2 = \exp\{\cdot\}$ and $g_1 = \log(\cdot)$. Thus, g_2 is a function of g_1, which is a function of h, which is itself a function of the parameter vector $\boldsymbol{\phi}$. As the exponential function is its own derivative, the equation in (A.3) can be simplified as follows:

$$\nabla h(\boldsymbol{\phi}) = h(\boldsymbol{\phi}) \cdot \nabla \underbrace{\log\left(h(\boldsymbol{\phi})\right)}_{g_1}. \tag{A.4}$$

The function $h(\boldsymbol{\phi})$ can be considered as prefactor which is placed out the vector of first order derivatives. Hence, it is sufficient to calculate the partial derivatives of the inner function g_1 in (A.2), which is $g_1 = \frac{1}{\phi^b} \log\left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right) + \phi^{(e)^*}$. Differentiating g_1 subject to the four parameters $\phi^{(b)}, \phi^{(c)}, \phi^{(d)}$ and $\phi^{(e)^*}$ yields

$$\begin{aligned}
\frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(b)}} &= -\frac{1}{\phi^{(b)^2}} \log\left(\frac{\phi^{(d)} - \lambda}{\lambda - \phi^{(c)}} \right), \\
\frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(c)}} &= \frac{1}{\phi^{(b)}} \cdot \frac{\lambda - \phi^{(c)}}{\phi^{(d)} - \lambda} \cdot (-1) \cdot \frac{\phi^{(d)} - \lambda}{(\lambda - \phi^{(c)})^2} \cdot (-1) = \frac{1}{\phi^{(b)}(\lambda - \phi^{(c)})}, \\
\frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(d)}} &= \frac{1}{\phi^{(b)}} \cdot \frac{\lambda - \phi^{(c)}}{\phi^{(d)} - \lambda} \cdot \frac{1}{\lambda - \phi^{(c)}} = \frac{1}{\phi^{(b)}(\phi^{(d)} - \lambda)},
\end{aligned} \tag{A.5}$$

$$\frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(e)*}} = 1.$$

The gradient ∇g_1 is the vector of the first order derivatives in (A.5). Setting ∇g_1 into formula (A.4), gives

$$\nabla h(\boldsymbol{\phi}) = \begin{pmatrix} \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(b)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(c)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(d)}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(e)*}} \end{pmatrix} = h(\boldsymbol{\phi}) \begin{pmatrix} -\frac{1}{\phi^{(b)2}} \log\left(\frac{\phi^{(d)}-\lambda}{\lambda-\phi^{(c)}}\right) \\ \frac{1}{\phi^{(b)}(\lambda-\phi^{(c)})} \\ \frac{1}{\phi^{(b)}(\phi^{(d)}-\lambda)} \\ 1 \end{pmatrix}.$$

A.2 Derivation of $\nabla f(x, \phi)$

Let $f(\cdot)$ denote the four-parameter log-logistic model for a dose-response data (x, y):

$$\begin{aligned} f(x, \boldsymbol{\phi}) &= \phi^{(c)} + \frac{\phi^{(d)} - \phi^{(c)}}{1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}} \\ &= \phi^{(c)} + (\phi^{(d)} - \phi^{(c)}) \cdot [1 + \underbrace{\exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}}_{\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}]^{-1} \end{aligned} \tag{A.6}$$

The 4pLL model function in (A.6) is differentiated with respect to $\phi^{(b)}, \phi^{(c)}, \phi^{(d)}$ and $\phi^{(e)}$:

$$\begin{aligned} \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(b)}} &= -\frac{(\phi^{(d)} - \phi^{(c)}) \cdot \left[\log(x) - \log(\phi^{(e)})\right] \cdot \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}}{\left[1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}\right]^2} \\ &= -\frac{(\phi^{(d)} - \phi^{(c)}) \cdot \left[\log(x) - \log(\phi^{(e)})\right] \cdot \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\left[1 + \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(c)}} &= 1 - \left[\frac{1}{1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}}\right] = 1 - \left[\frac{1}{1 + \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}\right] \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(d)}} &= \frac{1}{1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}} = \frac{1}{1 + \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}} \\ \frac{\partial h(\boldsymbol{\phi})}{\partial \phi^{(e)}} &= \frac{-(\phi^{(d)} - \phi^{(c)}) \cdot \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\} \cdot (-\phi^{(b)})}{\phi(e) \cdot \left[1 + \exp\left\{\phi^{(b)}\left[\log(x) - \log(\phi^{(e)})\right]\right\}\right]^2} \\ &= \frac{\phi^{(b)} \cdot (\phi^{(d)} - \phi^{(c)}) \cdot \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\phi^{(e)}\left[1 + \left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2}. \end{aligned} \tag{A.7}$$

The gradient ∇f containing the partial derivatives from the calculations in (A.7) is then given by

$$\nabla f(x,\boldsymbol{\phi}) = \begin{pmatrix} \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(b)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(c)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(d)}} \\ \frac{\partial f(x,\boldsymbol{\phi})}{\partial \phi^{(e)}} \end{pmatrix} = \begin{pmatrix} -\frac{\left(\phi^{(d)}-\phi^{(c)}\right)\left(\log\left(\frac{x}{\phi^{(e)}}\right)\right)\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\left[1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2} \\ 1-\left[\frac{1}{1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}\right] \\ \frac{1}{1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}} \\ \frac{\phi^{(b)}\left(\phi^{(d)}-\phi^{(c)}\right)\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}}{\phi^{(e)}\left[1+\left(\frac{x}{\phi^{(e)}}\right)^{\phi^{(b)}}\right]^2} \end{pmatrix}.$$

A.3 Derivation of $\gamma = \mathrm{F}(t_\nu)$

Let $\widehat{\mathrm{LEC}}$ and $\widehat{\mathrm{LOEC}}$ be the estimated alert concentrations obtained from the 4pLL and *Limma* model approach, respectively. According to section 3.5.4 a confidence interval for the LEC can be constructed as follows

$$\widehat{\mathrm{LEC}} \pm t_{\nu,\gamma}\sqrt{\widehat{\mathrm{var}}(\widehat{\mathrm{LEC}})}, \tag{A.8}$$

where $t_{\nu,\gamma}$ corresponds to the γ-quantile of a t-distribution with $\nu = n-4$ degrees of freedom for n observations.

Given the distribution of the $\widehat{\mathrm{LEC}}$s, it can be calculated which quantiles of this distribution the $\widehat{\mathrm{LOEC}}$s correspond to. For this, the formula in (A.9) must be transformed with respect to γ. The LOEC estimator is set equal to the upper confidence limit which is calculated according to formula (A.8) and γ is then obtained with the following equivalent transformation

$$\begin{aligned} &\widehat{\mathrm{LOEC}} = \widehat{\mathrm{LEC}} + t_{\nu,\gamma}\sqrt{\widehat{\mathrm{var}}(\widehat{\mathrm{LEC}})} \\ &\Leftrightarrow t_{\nu,\gamma} = \frac{\widehat{\mathrm{LOEC}} - \widehat{\mathrm{LEC}}}{\sqrt{\widehat{\mathrm{var}}(\widehat{\mathrm{LEC}})}} \\ &\Leftrightarrow \gamma = \mathrm{F}(t_\nu) = \mathrm{F}\left(\frac{\widehat{\mathrm{LOEC}} - \widehat{\mathrm{LEC}}}{\sqrt{\widehat{\mathrm{var}}(\widehat{\mathrm{LEC}})}}\right), \end{aligned} \tag{A.9}$$

where F indicates the distribution function of the t-distribution with ν degrees of freedom. Analogously, the $\widehat{\mathrm{ALOEC}}$s are re-calculated in terms of the respective quantiles which result from the distribution of the $\widehat{\mathrm{ALEC}}$s.

B Tables

Table B.1: *TG-GATEs: Matrix of the compounds exposed to primary human hepatocytes. The table gives full and abbreviated compound names as well as the concentration in µM (µg/mL, µg/kg) and the number of independent replicates of gene array data available after incubation with a low, middle and high concentration for 2h, 8h and 24h.*

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
2-NF	2-nitrofluorene	4 µM	20 µM	100 µM	0	0	0	2	2	2	2	2	2
AA	Allyl alcohol	3 µM	14 µM	70 µM	2	2	2	2	2	2	2	2	2
AAA	Acetamide	400 µM	2000 µM	10000 µM	0	0	0	2	2	2	2	2	2
AAF	Acetamidofluorene	NA	10 µM	50 µM	0	0	0	0	2	2	0	2	2
ACA	Acarbose	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
ACZ	Acetazolamide	NA	120 µM	600 µM	0	0	0	0	2	2	0	2	2
ADM	Alpidem	NA	3 µM	10 µM	0	0	0	0	0	0	0	2	2
ADP	Adapin	3 µM	15 µM	75 µM	2	2	2	2	2	2	2	2	2
AFB1	Aflatoxin B1	0 µg/mL	1 µg/mL	6 µg/mL	0	0	0	2	2	2	2	2	2
AJM	Ajmaline	NA	60 µM	300 µM	0	0	0	0	2	2	0	2	2
AM	Amiodarone	0 µM	1 µM	7 µM	2	2	2	2	2	2	2	2	2
AMB	Amphotericin B	0 µM	0 µM	2 µM	0	0	0	2	2	2	2	2	2
AMT	Amitriptyline	NA	3 µM	15 µM	0	0	0	0	2	2	0	2	2
ANIT	Naphthyl isothiocyanate	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
APAP	Acetaminophen	200 µM	1000 µM	5000 µM	2	2	2	2	2	2	2	2	2
APL	Allopurinol	6 µM	28 µM	140 µM	2	2	2	2	2	2	2	2	2
ASA	Aspirin	120 µM	600 µM	3000 µM	2	2	2	2	2	2	2	2	2
AZP	Azathioprine	3 µM	15 µM	73 µM	2	2	2	2	2	2	2	2	2
BBr	Benzbromarone	4 µM	20 µM	100 µM	2	2	2	2	2	2	2	2	2
BBZ	Bromobenzene	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
BCT	Bucetin	NA	60 µM	300 µM	0	0	0	0	2	2	0	2	2
BDZ	Bendazac	8 µM	40 µM	200 µM	0	0	0	2	2	2	2	2	2
BEA	Bromoethylamine	20 µM	100 µM	500 µM	0	0	0	2	2	2	2	2	1
BHA	Butylated hydroxyanisole	8 µM	40 µM	200 µM	0	0	0	2	2	2	2	2	2
BPR	Buspirone	NA	6 µM	30 µM	0	0	0	0	0	0	0	2	2
BSO	Buthionine sulfoximine	400 µM	2000 µM	10000 µM	0	0	0	2	2	2	2	2	2
BZD	Benziodarone	NA	8 µM	40 µM	0	0	0	0	2	2	0	2	2
CAF	Caffeine	NA	1500 µM	7500 µM	0	0	0	0	2	2	0	2	2
CAP	Captopril	NA	1600 µM	8000 µM	0	0	0	0	2	2	0	2	2
CBZ	Carbamazepine	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CCL4	Carbon tetrachloride	300 µM	1500 µM	7500 µM	2	2	2	2	2	2	2	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
CFB	Clofibrate	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CHL	Chlorpheniramine	NA	18 µM	90 µM	0	0	0	0	2	2	0	2	2
CHX	Cycloheximide	4 µM	20 µM	100 µM	0	0	0	2	2	2	2	2	2
CIM	Cimetidine	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CLM	Chlormadinone	NA	8 µM	40 µM	0	0	0	0	2	2	0	2	2
CMA	Coumarin	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CMN	Chlormezanone	NA	50 µM	250 µM	0	0	0	0	2	2	0	2	2
CMP	Chloramphenicol	NA	90 µM	450 µM	0	0	0	0	2	2	0	2	2
COL	Colchicine	NA	800 µM	4000 µM	0	0	0	0	2	2	0	2	2
CPA	Cyclophosphamide	80 µM	400 µM	2000 µM	2	2	2	2	2	2	2	2	2
CPM	Clomipramine	NA	2 µM	10 µM	0	0	0	0	2	2	0	2	2
CPP	Chlorpropamide	NA	150 µM	750 µM	0	0	0	0	2	2	0	2	2
CPX	Ciprofloxacin	NA	5 µM	25 µM	0	0	0	0	2	2	0	2	2
CPZ	Chlorpromazine	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
CSA	Cyclosporine A	0 µM	1 µM	6 µM	0	0	0	2	2	2	2	2	2
CZP	Clozapine	NA	10 µM	50 µM	0	0	0	0	0	0	0	2	2
DAPM	Methylene dianiline	24 µM	120 µM	600 µM	0	0	0	2	2	2	2	2	2
DEM	Diethyl maleate	60 µM	300 µM	1500 µM	0	0	0	2	2	2	2	2	2
DEN	Nitrosodiethylamine	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
DEX	Dexamethasone	12 µM	60 µM	300 µM	0	0	0	2	2	2	2	2	2
DFNa	Diclofenac	16 µM	80 µM	400 µM	2	2	2	2	2	2	2	2	2
DIL	Diltiazem	NA	30 µM	150 µM	0	0	0	0	2	2	0	2	2
DIS	Disopyramide	NA	700 µM	3500 µM	0	0	0	0	2	2	0	2	2
DNP	2,4-dinitrophenol	4 µM	20 µM	100 µM	0	0	0	2	2	2	2	2	2
DNZ	Danazol	NA	7 µM	35 µM	0	0	0	0	2	2	0	2	2
DOX	Doxorubicin	0 µM	2 µM	10 µM	0	0	0	2	2	2	2	2	2
DSF	Disulfiram	NA	12 µM	60 µM	0	0	0	0	2	2	0	2	2
DTL	Dantrolene	NA	2 µM	10 µM	0	0	0	0	2	2	0	2	2
DZP	Diazepam	10 µM	50 µM	250 µM	2	2	2	2	2	2	2	2	2
EBU	Ethambutol	NA	800 µM	4000 µM	0	0	0	0	2	2	0	2	2
EE	Ethinylestradiol	NA	3 µM	15 µM	0	0	0	0	2	2	0	2	2
EME	Erythromycin ethylsuccinate	NA	1 µM	5 µM	0	0	0	0	2	2	0	2	2
ENA	Enalapril	NA	400 µM	2000 µM	0	0	0	0	2	2	0	2	2
ET	Ethionine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
ETH	Ethionamide	NA	120 µM	600 µM	0	0	0	0	2	2	0	2	2
ETN	Ethanol	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
ETP	Etoposide	NA	66 µM	330 µM	0	0	0	0	2	2	0	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
FAM	Famotidine	NA	140 µM	700 µM	0	0	0	0	2	2	0	2	2
FFB	Fenofibrate	NA	6 µM	30 µM	0	0	0	0	2	2	0	2	2
FLX	Fluoxetine hydrochloride	4 µM	8 µM	20 µM	0	0	0	2	2	2	2	2	2
FP	Fluphenazine	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
FT	Flutamide	2 µM	10 µM	50 µM	2	2	2	2	2	2	2	2	2
FUR	Furosemide	NA	500 µM	2500 µM	0	0	0	0	2	2	0	2	2
GaN	Galactosamine	400 µM	2000 µM	10000 µM	0	0	0	2	2	2	2	2	2
GBC	Glibenclamide	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
GF	Griseofulvin	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
GFZ	Gemfibrozil	4 µM	20 µM	100 µM	2	2	2	2	2	2	2	2	2
HCB	Hexachlorobenzene	1 µM	6 µM	30 µM	2	2	2	2	2	2	2	2	2
hHGF	Hepatocyte growth factor, human	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	2	2	2
hIFNA	Interferon alpha, human	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	0	0	0
hIL1B	Interleukin 1 beta, human	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	2	2	2
hIL6	Interleukin 6, human	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	2	2	2
HPL	Haloperidol	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
HYZ	Hydroxyzine	NA	8 µM	40 µM	0	0	0	0	2	2	0	2	2
IBU	Ibuprofen	NA	30 µM	150 µM	0	0	0	0	2	2	0	2	2
IM	Indomethacin	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
IMI	Imipramine	NA	3 µM	15 µM	0	0	0	0	2	2	0	2	2
INAH	Isoniazid	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
IPA	Iproniazid	NA	240 µM	1200 µM	0	0	0	0	2	2	0	2	2
KC	Ketoconazole	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
LBT	Labetalol	6 µM	28 µM	140 µM	2	2	2	2	2	2	2	2	2
LNX	Lornoxicam	NA	3 µM	15 µM	0	0	0	0	2	2	0	2	2
LPS	Lipopolysaccharide	12 µg/mL	60 µg/mL	300 µg/mL	0	0	0	2	2	2	2	2	1
LS	Lomustine	5 µM	24 µM	120 µM	2	2	2	2	2	2	2	2	2
MCT	Monocrotaline	NA	18 µM	90 µM	0	0	0	0	2	2	0	2	2
MDP	Methyldopa	NA	10 µM	50 µM	0	0	0	0	2	2	0	2	2
MEF	Mefenamic acid	NA	30 µM	150 µM	0	0	0	0	2	2	0	2	2
MEX	Mexiletine	NA	60 µM	300 µM	0	0	0	0	2	2	0	2	2
MFM	Metformin	NA	200 µM	1000 µM	0	0	0	0	2	2	0	2	2
MLX	Meloxicam	NA	10 µM	50 µM	0	0	0	0	2	2	0	2	2
MNU	N-methyl-N-nitrosourea	1200 µM	6000 µM	10000 µM	0	0	0	2	2	2	2	2	2
MP	Methapyrilene	24 µM	120 µM	600 µM	2	2	2	2	2	2	2	2	2
MTS	Methyltestosterone	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
MTZ	Methimazole	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
MXS	Moxisylyte	NA	80 µM	400 µM	0	0	0	0	2	2	0	2	2
NFT	Nitrofurantoin	5 µM	25 µM	125 µM	2	2	2	2	2	2	2	2	2
NFZ	Nitrofurazone	NA	10 µM	50 µM	0	0	0	0	2	2	0	2	2
NIC	Nicotinic acid	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
NIF	Nifedipine	NA	30 µM	150 µM	0	0	0	0	2	2	0	2	2
NIM	Nimesulide	NA	66 µM	330 µM	0	0	0	0	2	2	0	2	2
NMOR	N-nitrosomorpholine	400 µM	2000 µM	10000 µM	0	0	0	2	2	2	2	2	2
NPAA	Phenylanthranilic acid	NA	40 µM	200 µM	0	0	0	0	2	2	0	2	2
NPX	Naproxen	NA	120 µM	600 µM	0	0	0	0	2	2	0	2	2
NZD	Nefazodone	NA	6 µM	30 µM	0	0	0	0	0	0	0	2	2
OPZ	Omeprazole	24 µM	120 µM	600 µM	2	2	2	2	2	2	2	2	2
PAP	Papaverine	NA	12 µM	60 µM	0	0	0	0	2	2	0	2	2
PB	Phenobarbital	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
PCT	Phenacetin	NA	120 µM	600 µM	0	0	0	0	2	2	0	2	2
PEN	Penicillamine	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
PH	Perhexiline	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
PHA	Phalloidin	16 µg/mL	80 µg/mL	400 µg/mL	0	0	0	2	2	2	2	2	2
PhB	Phenylbutazone	16 µM	80 µM	400 µM	2	2	2	2	2	2	2	2	2
PHE	Phenytoin	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
PHO	Phorone	200 µM	1000 µM	5000 µM	0	0	0	2	2	2	2	2	0
PML	Pemoline	NA	15 µM	75 µM	0	0	0	0	2	2	0	2	2
PMZ	Promethazine	NA	7 µM	35 µM	0	0	0	0	2	2	0	2	2
PPL	Propranolol	6 µg/kg	30 µg/kg	100 µg/kg	0	0	0	2	2	2	2	2	2
PTU	Propylthiouracil	160 µM	800 µM	4000 µM	2	2	2	2	2	2	2	2	2
QND	Quinidine	NA	10 µM	50 µM	0	0	0	0	2	2	0	2	2
RAN	Ranitidine	NA	800 µM	4000 µM	0	0	0	0	2	2	0	2	2
RGZ	Rosiglitazone maleate	12 µg/kg	60 µg/kg	300 µg/kg	0	0	0	2	2	2	2	2	2
RIF	Rifampicin	3 µM	14 µM	70 µM	2	2	2	2	2	2	2	2	2
ROT	Rotenone	0 µM	0 µM	2 µM	0	0	0	2	2	2	2	2	2
SLP	Sulpiride	NA	1000 µM	5000 µM	0	0	0	0	2	2	0	2	2
SS	Sulfasalazine	6 µM	30 µM	150 µM	2	2	2	2	2	2	2	2	2
SST	Simvastatin	NA	6 µM	30 µM	0	0	0	0	2	2	0	2	2
SUL	Sulindac	NA	600 µM	3000 µM	0	0	0	0	2	2	0	2	2
TAA	Thioacetamide	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
TAC	Tacrine	NA	16 µM	80 µM	0	0	0	0	2	2	0	2	2
TAN	Tannic acid	NA	1 µM	5 µM	0	0	0	0	2	2	0	2	2
TBF	Terbinafine	NA	3 µM	15 µM	0	0	0	0	2	2	0	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
TC	Tetracycline	1 µM	5 µM	25 µM	2	2	2	2	2	2	2	2	2
TCP	Ticlopidine	NA	4 µM	20 µM	0	0	0	0	2	2	0	2	2
TEO	Theophylline	NA	2000 µM	10000 µM	0	0	0	0	2	2	0	2	2
TGFB1	Transforming growth factor beta 1	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	2	2	2
TIO	Tiopronin	NA	400 µM	2000 µM	0	0	0	0	2	2	0	2	2
TLB	Tolbutamide	NA	400 µM	2000 µM	0	0	0	0	2	2	0	2	2
TMD	Trimethadione	400 µM	2000 µM	10000 µM	0	0	0	2	2	2	2	2	1
TMX	Tamoxifen	NA	5 µM	25 µM	0	0	0	0	2	2	0	2	2
TNF	Tumor necrosis factor	2 ng/mL	10 ng/mL	50 ng/mL	0	0	0	2	2	2	2	2	2
TRZ	Thioridazine	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
TUN	Tunicamycin	0 µg/mL	2 µg/mL	10 µg/mL	0	0	0	2	2	2	2	2	2
TZM	Triazolam	0 µM	2 µM	10 µM	0	0	0	2	2	2	2	2	2
VA	Vitamin A	NA	2 µM	8 µM	0	0	0	0	2	2	0	2	2
VLF	Venlafaxine	NA	240 µM	1200 µM	0	0	0	0	0	0	0	2	2
VPA	Valproic acid	200 µM	1000 µM	5000 µM	2	2	2	2	2	2	2	2	2
WY	WY-14643	6 µM	30 µM	150 µM	2	2	2	2	2	2	2	2	2

Table B.2: *TGD (in vitro): Matrix of the compounds exposed to rat hepatocytes. The table gives full and abbreviated compound names as well as the concentration in µM (µg/mL, µg/kg) and the number of independent replicates of gene array data available after incubation with a low, middle and high concentration for 2h, 8h and 24h.*

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
AA	Allyl alcohol	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
AAF	Acetamidofluorene	2 µM	10 µM	50 µM	2	2	2	2	2	2	2	2	2
ACA	Acarbose	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
ACZ	Acetazolamide	24 µM	120 µM	600 µM	2	2	2	2	2	2	2	2	2
ADM	Alpidem	10 µM	NA	NA	0	0	0	0	0	0	2	0	0
ADP	Adapin	3 µM	15 µM	75 µM	2	2	2	2	2	2	2	2	2
AJM	Ajmaline	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
AM	Amiodarone	0 µM	1 µM	7 µM	2	2	2	2	2	2	2	2	2
AMT	Amitriptyline	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
ANIT	Naphthyl isothiocyanate	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
APAP	Acetaminophen	1000 µM	3000 µM	10000 µM	2	2	2	2	2	2	2	2	2
APL	Allopurinol	6 µM	28 µM	140 µM	2	2	2	2	2	2	2	2	2
ASA	Aspirin	120 µM	600 µM	3000 µM	2	2	2	2	2	2	2	2	2
AZP	Azathioprine	0 µM	1 µM	4 µM	2	2	2	2	2	2	2	2	2
BBr	Benzbromarone	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
BBZ	Bromobenzene	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
BCT	Bucetin	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
BDZ	Bendazac	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
BEA	Bromoethylamine	20 µM	100 µM	500 µM	2	2	2	2	2	2	2	2	2
BPR	Buspirone	70 µM	NA	NA	0	0	0	0	0	0	2	0	0
BSO	Buthionine sulfoximine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
BZD	Benziodarone	1 µM	5 µM	25 µM	2	2	2	2	2	2	2	2	2
CAF	Caffeine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
CAP	Captopril	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
CBP	Carboplatin	120 µM	600 µM	3000 µM	2	2	2	2	2	2	2	2	2
CBZ	Carbamazepine	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CCL4	Carbon tetrachloride	1000 µM	3000 µM	10000 µM	2	2	0	2	2	2	2	2	0
CFB	Clofibrate	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CHL	Chlorpheniramine	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
CHX	Cycloheximide	4 µM	20 µM	100 µM	2	2	2	2	2	2	2	2	2
CIM	Cimetidine	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
CLM	Chlormadinone	2 µM	8 µM	40 µM	2	2	2	2	2	2	2	2	2
CLT	Cephalothin	120 µM	600 µM	3000 µM	2	2	2	2	2	2	2	2	2
CMA	Coumarin	12 µM	60 µM	300 µM	2	2	2	2	2	2	2	2	2
CMN	Chlormezanone	10 µM	50 µM	250 µM	2	2	2	2	2	2	2	2	2
CMP	Chloramphenicol	18 µM	90 µM	450 µM	2	2	2	2	2	2	2	2	2
COL	Colchicine	200 µM	1000 µM	5000 µM	2	2	2	2	2	2	2	2	2
CPA	Cyclophosphamide	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
CPM	Clomipramine	2 µM	8 µM	40 µM	2	2	2	2	2	2	2	2	2
CPP	Chlorpropamide	30 µM	150 µM	750 µM	2	2	2	2	2	2	2	2	2
CPX	Ciprofloxacin	1 µM	5 µM	25 µM	2	2	2	2	2	2	2	2	2
CPZ	Chlorpromazine	1 µM	4 µM	20 µM	2	2	2	2	2	2	2	2	2
CSA	Cyclosporine A	0 µM	1 µM	6 µM	2	2	2	2	2	2	2	2	2
CSP	Cisplatin	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
CZP	Clozapine	120 µM	NA	NA	0	0	0	0	0	0	2	0	0
DEM	Diethyl maleate	60 µM	300 µM	1500 µM	2	2	2	2	2	2	2	2	2
DEN	Nitrosodiethylamine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
DFNa	Diclofenac	16 µM	80 µM	400 µM	2	2	2	2	2	2	2	2	2
DIL	Diltiazem	10 µM	50 µM	250 µM	2	2	2	2	2	2	2	2	2
DIS	Disopyramide	100 µM	500 µM	2500 µM	2	2	2	2	2	2	2	2	2
DNZ	Danazol	1 µM	7 µM	35 µM	2	2	2	2	2	2	2	2	2
DOX	Doxorubicin	0 µM	0 µM	2 µM	2	2	2	2	2	2	2	2	2
DSF	Disulfiram	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
DTL	Dantrolene	0 µM	2 µM	10 µM	2	2	2	2	2	2	2	2	2
DZP	Diazepam	5 µM	25 µM	125 µM	2	2	2	2	2	2	2	2	2
EBU	Ethambutol	160 µM	800 µM	4000 µM	2	2	2	2	2	2	2	2	2
EE	Ethinylestradiol	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
EME	Erythromycin ethylsuccinate	3 µM	15 µM	75 µM	2	2	2	2	2	2	2	2	2
ENA	Enalapril	80 µM	400 µM	2000 µM	2	2	2	2	2	2	2	2	2
ET	Ethionine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
ETH	Ethionamide	24 µM	120 µM	600 µM	2	2	2	2	2	2	2	2	2
ETN	Ethanol	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
ETP	Etoposide	14 µM	70 µM	350 µM	2	2	2	2	2	2	2	2	2
FAM	Famotidine	28 µM	140 µM	700 µM	2	2	2	2	2	2	2	2	2
FFB	Fenofibrate	1 µM	6 µM	30 µM	2	2	2	2	2	2	2	2	2
FP	Fluphenazine	1 µM	6 µM	30 µM	2	2	2	2	2	2	2	2	2
FT	Flutamide	3 µM	15 µM	75 µM	2	2	2	2	2	2	2	2	2
FUR	Furosemide	100 µM	500 µM	2500 µM	2	2	2	2	2	2	2	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
GaN	Galactosamine	400 μM	2000 μM	10000 μM	2	2	2	2	2	2	2	2	2
GBC	Glibenclamide	2 μM	12 μM	60 μM	2	2	2	2	2	2	2	2	2
GF	Griseofulvin	1 μM	6 μM	30 μM	2	2	2	2	2	2	2	2	2
GFZ	Gemfibrozil	4 μM	20 μM	100 μM	2	2	2	2	2	2	2	2	2
GMC	Gentamicin	1 mg/mL	6 mg/mL	30 mg/mL	2	2	2	2	2	2	2	2	2
HCB	Hexachlorobenzene	1 μM	3 μM	15 μM	2	2	2	2	2	2	2	2	2
HPL	Haloperidol	2 μM	10 μM	50 μM	2	2	2	2	2	2	2	2	2
HYZ	Hydroxyzine	6 μM	30 μM	150 μM	2	2	2	2	2	2	2	2	2
IBU	Ibuprofen	40 μM	200 μM	1000 μM	2	2	2	2	2	2	2	2	2
IM	Indomethacin	12 μM	60 μM	300 μM	2	2	2	2	2	2	2	2	2
IMI	Imipramine	4 μM	20 μM	100 μM	2	2	2	2	2	2	2	2	2
INAH	Isoniazid	400 μM	2000 μM	10000 μM	2	2	2	2	2	2	2	2	2
IPA	Iproniazid	80 μM	400 μM	2000 μM	2	2	2	2	2	2	2	2	2
KC	Ketoconazole	1 μM	3 μM	15 μM	2	2	2	2	2	2	2	2	2
LBT	Labetalol	6 μM	28 μM	140 μM	2	2	2	2	2	2	2	2	2
LNX	Lornoxicam	1 μM	3 μM	15 μM	2	2	2	2	2	2	2	2	2
LPS	LPS	0 μg/mL	1 μg/mL	NA	2	2	0	2	2	0	2	2	0
LS	Lomustine	5 μM	24 μM	120 μM	2	2	2	2	2	2	2	2	2
MCT	Monocrotaline	4 μM	18 μM	90 μM	2	2	2	2	2	2	2	2	2
MDP	Methyldopa	2 μM	10 μM	50 μM	2	2	2	2	2	2	2	2	2
MEF	Mefenamic acid	6 μM	30 μM	150 μM	2	2	2	2	2	2	2	2	2
MEX	Mexiletine	1 μM	3 μM	15 μM	2	2	2	2	2	2	2	2	2
MFM	Metformin	40 μM	200 μM	1000 μM	2	2	2	2	2	2	2	2	2
MLX	Meloxicam	2 μM	10 μM	50 μM	2	2	2	2	2	2	2	2	2
MP	Methapyrilene	1 μM	3 μM	15 μM	2	2	2	2	2	2	2	2	2
MTS	Methyltestosterone	2 μM	8 μM	40 μM	2	2	2	2	2	2	2	2	2
MTZ	Methimazole	400 μM	2000 μM	10000 μM	2	2	2	2	2	2	2	2	2
MXS	Moxisylyte	24 μM	120 μM	600 μM	2	2	2	2	2	2	2	2	2
NFT	Nitrofurantoin	5 μM	25 μM	125 μM	2	2	2	2	2	2	2	2	2
NFZ	Nitrofurazone	12 μM	60 μM	300 μM	2	2	2	2	2	2	2	2	2
NIC	Nicotinic acid	400 μM	2000 μM	10000 μM	2	2	2	2	2	2	2	2	2
NIF	Nifedipine	10 μM	50 μM	250 μM	2	2	2	2	2	2	2	2	2
NIM	Nimesulide	3 μM	15 μM	75 μM	2	2	2	2	2	2	2	2	2
NPAA	Phenylanthranilic acid	8 μM	40 μM	200 μM	2	2	2	2	2	2	2	2	2
NPX	Naproxen	80 μM	400 μM	2000 μM	2	2	2	2	2	2	2	2	2
NZD	Nefazodone	70 μg/kg	NA	NA	0	0	0	0	0	0	2	0	0
OPZ	Omeprazole	5 μM	24 μM	120 μM	2	2	2	2	2	2	2	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
PAN	Puromycin aminonucleoside	100 µM	500 µM	2500 µM	2	2	2	2	2	2	2	2	2
PAP	Papaverine	4 µM	20 µM	100 µM	2	2	2	2	2	2	2	2	2
PB	Phenobarbital	1000 µM	3000 µM	10000 µM	2	2	2	2	2	2	2	2	2
PCT	Phenacetin	24 µM	120 µM	600 µM	2	2	2	2	2	2	2	2	2
PEN	Penicillamine	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
PH	Perhexiline	0 µM	2 µM	10 µM	2	2	2	2	2	2	2	2	2
PHA	Phalloidin	4 µg/mL	20 µg/mL	100 µg/mL	2	2	2	2	2	2	2	2	2
PhB	Phenylbutazone	16 µM	80 µM	400 µM	2	2	2	2	2	2	2	2	2
PHE	Phenytoin	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
PHO	Phorone	20 µM	100 µM	500 µM	2	2	2	2	2	2	2	2	2
PML	Pemoline	3 µM	15 µM	75 µM	2	2	2	2	2	2	2	2	2
PMZ	Promethazine	3 µM	16 µM	80 µM	2	2	2	2	2	2	2	2	2
PTU	Propylthiouracil	160 µM	800 µM	4000 µM	2	2	2	2	2	2	2	2	2
QND	Quinidine	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
RAN	Ranitidine	160 µM	800 µM	4000 µM	2	2	2	2	2	2	2	2	2
RIF	Rifampicin	3 µM	14 µM	70 µM	2	2	2	2	2	2	2	2	2
SLP	Sulpiride	200 µM	1000 µM	5000 µM	2	2	2	2	2	2	2	2	2
SS	Sulfasalazine	4 µM	20 µM	100 µM	2	2	2	2	2	2	2	2	2
SST	Simvastatin	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
SUL	Sulindac	80 µM	400 µM	2000 µM	2	2	2	2	2	2	2	2	2
TAA	Thioacetamide	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
TAC	Tacrine	8 µM	40 µM	200 µM	2	2	2	2	2	2	2	2	2
TAN	Tannic acid	0 µM	2 µM	10 µM	2	2	2	2	2	2	2	2	2
TBF	Terbinafine	1 µM	3 µM	15 µM	2	2	2	2	2	2	2	2	2
TC	Tetracycline	1 µM	5 µM	25 µM	2	2	2	2	2	2	2	2	2
TCP	Ticlopidine	2 µM	12 µM	60 µM	2	2	2	2	2	2	2	2	2
TEO	Theophylline	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
TIO	Tiopronin	1 µM	5 µM	25 µM	2	2	2	2	2	2	2	2	2
TLB	Tolbutamide	80 µM	400 µM	2000 µM	2	2	2	2	2	2	2	2	2
TMD	Trimethadione	400 µM	2000 µM	10000 µM	2	2	2	2	2	2	2	2	2
TMX	Tamoxifen	0 µM	1 µM	3 µM	2	2	2	2	2	2	2	2	2
TNF	TNFa	2 ng/mL	10 ng/mL	50 ng/mL	2	2	2	2	2	2	2	2	2
TRI	Triamterene	1 µM	6 µM	30 µM	2	2	2	2	2	2	2	2	2
TRZ	Thioridazine	0 µM	2 µM	10 µM	2	2	2	2	2	2	2	2	2
TUN	Tunicamycin	2 µg/mL	10 µg/mL	50 µg/mL	2	2	2	2	2	2	2	2	2
TZM	Triazolam	0 µM	2 µM	10 µM	2	2	2	2	2	2	2	2	2
VA	Vitamin A	0 µM	2 µM	8 µM	2	2	2	2	2	2	2	2	2

Continued on next page

Compound		Concentration			2h			8h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
VLF	Venlafaxine	1200 μM	NA	NA	0	0	0	0	0	0	2	0	0
VPA	Valproic acid	400 μM	2000 μM	10000 μM	2	2	2	2	2	2	2	2	2
WY	WY-14643	8 μM	40 μM	200 μM	2	2	2	2	2	2	2	2	2

Table B.3: *TGD (in vivo): Matrix of the compounds exposed to rat liver cells. The table gives full and abbreviated compound names as well as the concentration in µM (µg/mL, µg/kg) and the number of independent replicates of gene array data available after incubation with a low, middle and high concentration for 3h, 6h, 9h and 24h.*

Compound		Concentration			3h			6h			9h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
2-NF	2-nitrofluorene	100 mg/kg	300 mg/kg	1000 mg/kg	0	0	0	0	0	0	0	0	0	3	3	3
3-MC	3-methylcholanthrene	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AA	Allyl alcohol	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AAF	Acetamidofluorene	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ACA	Acarbose	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ACZ	Acetazolamide	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ADP	Adapin	100 mg/kg	300 mg/kg	NA	3	3	0	3	3	0	3	3	0	3	3	0
AFB1	Aflatoxin B1	1 mg/kg	3 mg/kg	10 mg/kg	0	0	0	0	0	0	0	0	0	3	3	5
AJM	Ajmaline	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AM	Amiodarone	200 mg/kg	600 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AMB	Amphotericin B	1 mg/kg	2 mg/kg	NA	3	0	0	3	3	0	3	3	0	3	3	0
AMT	Amitriptyline	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ANIT	Naphthyl isothiocyanate	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
APAP	Acetaminophen	300 mg/kg	600 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
APL	Allopurinol	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ASA	Aspirin	450 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AZP	Azathioprine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BBr	Benzbromarone	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BBZ	Bromobenzene	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BCT	Bucetin	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BDZ	Bendazac	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BEA	Bromoethylamine	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BHA	Butylated hydroxyanisole	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BOR	Bortezomib	0 mg/kg	0 mg/kg	1 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BSO	Buthionine sulfoximine	150 mg/kg	450 mg/kg	1500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BZD	Benziodarone	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CAF	Caffeine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CAP	Captopril	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CBP	Carboplatin	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CBZ	Carbamazepine	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CCL4	Carbon tetrachloride	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			3h			6h			9h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
CFB	Clofibrate	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CHL	Chlorpheniramine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CHX	Cycloheximide	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
CIM	Cimetidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CLM	Chlormadinone	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CLT	Cephalothin	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMA	Coumarin	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMN	Chlormezanone	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMP	Chloramphenicol	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
COL	Colchicine	2 mg/kg	5 mg/kg	15 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPA	Cyclophosphamide	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPM	Clomipramine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPP	Chlorpropamide	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPX	Ciprofloxacin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPZ	Chlorpromazine	45 mg/kg	150 mg/kg	NA	3	3	0	3	3	0	3	3	0	3	3	0
CSA	Cyclosporine A	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CSP	Cisplatin	0 mg/kg	1 mg/kg	3 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DAPM	Methylene dianiline	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DDAVP	Desmopressin acetate	20 µg/kg	200 µg/kg	2000 µg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DEM	Diethyl maleate	80 mg/kg	240 mg/kg	800 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DEN	Nitrosodiethylamine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DEX	Dexamethasone	1 mg/kg	5 mg/kg	50 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DFNa	Diclofenac	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DIL	Diltiazem	80 mg/kg	240 mg/kg	800 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DIS	Disopyramide	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DNP	2,4-dinitrophenol	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DNZ	Danazol	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DOX	Doxorubicin	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DSF	Disulfiram	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DTL	Dantrolene	25 mg/kg	75 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DZP	Diazepam	25 mg/kg	75 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EBU	Ethambutol	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EE	Ethinylestradiol	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EME	Erythromycin ethylsuccinate	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ENA	Enalapril	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ET	Ethionine	25 mg/kg	80 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ETH	Ethionamide	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			3h			6h			9h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
ETN	Ethanol	400 mg/kg	1200 mg/kg	4000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ETP	Etoposide	10 mg/kg	100 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FAM	Famotidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FFB	Fenofibrate	10 mg/kg	100 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FLX	Fluoxetine hydrochloride	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FP	Fluphenazine	2 mg/kg	6 mg/kg	20 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FT	Flutamide	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FUR	Furosemide	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GaN	Galactosamine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GBC	Glibenclamide	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GEF	Gefitinib	200 mg/kg	600 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GF	Griseofulvin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GFZ	Gemfibrozil	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GMC	Gentamicin	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HCB	Hexachlorobenzene	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HPL	Haloperidol	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HYZ	Hydroxyzine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IBU	Ibuprofen	60 mg/kg	200 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IM	Indomethacin	5 mg/kg	15 mg/kg	50 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IMA	Imatinib, methanesulfonate salt	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IMI	Imipramine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
INAH	Isoniazid	200 mg/kg	600 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IPA	Iproniazid	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
KC	Ketoconazole	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
LBT	Labetalol	45 mg/kg	150 mg/kg	450 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
LNX	Lornoxicam	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
LPS	LPS	10 µg/kg	300 µg/kg	3000 µg/kg	3	3	3	3	3	3	3	3	3	3	3	3
LS	Lomustine	1 mg/kg	2 mg/kg	6 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MCT	Monocrotaline	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MDP	Methyldopa	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MEF	Mefenamic acid	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MEX	Mexiletine	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MFM	Metformin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MLX	Meloxicam	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MNU	N-methyl-N-nitrosourea	10 mg/kg	30 mg/kg	100 mg/kg	0	0	0	0	0	0	0	0	0	3	3	3
MP	Methapyrilene	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MTS	Methyltestosterone	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			3h			6h			9h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
MTZ	Methimazole	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MXS	Moxisylyte	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NFT	Nitrofurantoin	100 mg/kg	300 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NFZ	Nitrofurazone	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIC	Nicotinic acid	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIF	Nifedipine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIM	Nimesulide	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NMOR	N-nitrosomorpholine	100 mg/kg	300 mg/kg	600 mg/kg	0	0	0	0	0	0	0	0	0	3	3	3
NPAA	Phenylanthranilic acid	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NPX	Naproxen	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
OPZ	Omeprazole	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PAN	Puromycin aminonucleoside	12 mg/kg	40 mg/kg	120 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PAP	Papaverine	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PB	Phenobarbital	100 mg/kg	150 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PCT	Phenacetin	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PEN	Penicillamine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PH	Perhexiline	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PHA	Phalloidin	NA	NA	1 mg/kg	0	0	3	0	0	3	0	0	3	0	0	3
PhB	Phenylbutazone	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PHE	Phenytoin	600 mg/kg	1200 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PHO	Phorone	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PML	Pemoline	8 mg/kg	25 mg/kg	75 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PMZ	Promethazine	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PPL	Propranolol	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PTU	Propylthiouracil	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
QND	Quinidine	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RAN	Ranitidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RGZ	Rosiglitazone maleate	20 mg/kg	100 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RIF	Rifampicin	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ROT	Rotenone	5 mg/kg	15 mg/kg	50 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SLP	Sulpiride	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SS	Sulfasalazine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SST	Simvastatin	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SUL	Sulindac	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TAA	Thioacetamide	5 mg/kg	15 mg/kg	45 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TAC	Tacrine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TAN	Tannic acid	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			3h			6h			9h			24h		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
TBF	Terbinafine	75 mg/kg	250 mg/kg	750 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TC	Tetracycline	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TCP	Ticlopidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TEO	Theophylline	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TIO	Tiopronin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TLB	Tolbutamide	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TMD	Trimethadione	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TMX	Tamoxifen	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TNF	TNFa	1 µg/kg	3 µg/kg	10 µg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TRI	Triamterene	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TRZ	Thioridazine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TUN	Tunicamycin	30 µg/kg	100 µg/kg	300 µg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TZM	Triazolam	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
VA	Vitamin A	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
VPA	Valproic acid	45 mg/kg	150 mg/kg	450 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
WY	WY-14643	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Table B.4: *TGD (in vivo): Matrix of the compounds exposed to rat liver cells. The table gives full and abbreviated compound names as well as the concentration in µM (µg/mL, µg/kg) and the number of independent replicates of gene array data available after incubation with a low, middle and high concentration for 4 days, 8 days, 15 days and 29 days.*

Compound		Concentration			4 days			8 days			15 days			29 days		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
AA	Allyl alcohol	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AAA	Acetamide	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AAF	Acetamidofluorene	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ACA	Acarbose	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ACZ	Acetazolamide	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ADP	Adapin	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AJM	Ajmaline	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AM	Amiodarone	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AMB	Amphotericin B	0 mg/kg	0 mg/kg	1 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AMT	Amitriptyline	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ANIT	Naphthyl isothiocyanate	2 mg/kg	5 mg/kg	15 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
APAP	Acetaminophen	300 mg/kg	600 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
APL	Allopurinol	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ASA	Aspirin	45 mg/kg	150 mg/kg	450 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
AZP	Azathioprine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
BBr	Benzbromarone	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BBZ	Bromobenzene	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BCT	Bucetin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BDZ	Bendazac	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BEA	Bromoethylamine	2 mg/kg	6 mg/kg	20 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BHA	Butylated hydroxyanisole	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
BZD	Benziodarone	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CAF	Caffeine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CAP	Captopril	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CBP	Carboplatin	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CBZ	Carbamazepine	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CCL4	Carbon tetrachloride	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CFB	Clofibrate	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CH+DS-Na	1% cholesterol + 0.25% sodium cholate	NA	NA		0	0	3	0	0	3	0	0	3	0	0	3
CHL	Chlorpheniramine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound Abbr.	Name	Concentration Low	Middle	High	4 days Low	Middle	High	8 days Low	Middle	High	15 days Low	Middle	High	29 days Low	Middle	High
CIM	Cimetidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CLM	Chlormadinone	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CLT	Cephalothin	300 mg/kg	1000 mg/kg	2000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMA	Coumarin	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMN	Chlormezanone	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CMP	Chloramphenicol	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
COL	Colchicine	1 mg/kg	2 mg/kg	5 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPA	Cyclophosphamide	2 mg/kg	5 mg/kg	15 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPM	Clomipramine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPP	Chlorpropamide	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPX	Ciprofloxacin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CPZ	Chlorpromazine	5 mg/kg	15 mg/kg	45 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CSA	Cyclosporine A	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
CSP	Cisplatin	0 mg/kg	0 mg/kg	1 mg/kg	3	3	3	3	3	3	3	3	3	3	3	1
DAPM	Methylene dianiline	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DDAVP	Desmopressin acetate	2 µg/kg	20 µg/kg	200 µg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DEN	Nitrosodiethylamine	3 mg/kg	10 mg/kg		3	3	3	3	3	3	3	3	1	3	3	0
DFNa	Diclofenac	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DIL	Diltiazem	80 mg/kg	240 mg/kg	800 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DIS	Disopyramide	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DNP	2,4-dinitrophenol	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DNZ	Danazol	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DOX	Doxorubicin	0 mg/kg	0 mg/kg	1 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
DSF	Disulfiram	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DTL	Dantrolene	25 mg/kg	75 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
DZP	Diazepam	25 mg/kg	75 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EBU	Ethambutol	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EE	Ethinylestradiol	1 mg/kg	3 mg/kg	10 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
EME	Erythromycin ethylsuccinate	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ENA	Enalapril	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ET	Ethionine	25 mg/kg	80 mg/kg	250 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ETH	Ethionamide	30 mg/kg	100 mg/kg		3	3	3	3	3	3	3	3	3	3	3	0
ETN	Ethanol	400 mg/kg	1200 mg/kg	4000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ETP	Etoposide	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FAM	Famotidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FFB	Fenofibrate	10 mg/kg	100 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			4 days			8 days			15 days			29 days		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
FLX	Fluoxetine hydrochloride	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FP	Fluphenazine	2 mg/kg	6 mg/kg	20 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FT	Flutamide	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
FUR	Furosemide	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GBC	Glibenclamide	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GF	Griseofulvin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GFZ	Gemfibrozil	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
GMC	Gentamicin	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HCB	Hexachlorobenzene	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HPL	Haloperidol	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
HYZ	Hydroxyzine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IBU	Ibuprofen	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IM	Indomethacin	1 mg/kg	2 mg/kg		3	3	3	3	3	3	3	3	3	3	3	0
IMI	Imipramine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
INAH	Isoniazid	50 mg/kg	100 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
IPA	Iproniazid	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
KC	Ketoconazole	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
LBT	Labetalol	45 mg/kg	150 mg/kg	450 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
LNX	Lornoxicam	0 mg/kg	1 mg/kg		3	3	3	3	3	3	3	3	3	3	3	0
LS	Lomustine	1 mg/kg	2 mg/kg	6 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MCT	Monocrotaline	3 mg/kg	10 mg/kg		3	3	3	3	3	3	3	3	3	3	3	0
MDP	Methyldopa	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MEF	Mefenamic acid	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MEX	Mexiletine	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MFM	Metformin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MLX	Meloxicam	3 mg/kg	10 mg/kg		3	3	3	3	3	3	3	3	2	3	3	0
MP	Methapyrilene	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MTS	Methyltestosterone	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MTZ	Methimazole	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
MXS	Moxisylyte	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NFT	Nitrofurantoin	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NFZ	Nitrofurazone	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIC	Nicotinic acid	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIF	Nifedipine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NIM	Nimesulide	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
NPAA	Phenylanthranilic acid	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
NPX	Naproxen	6 mg/kg	20 mg/kg		3	3	3	3	3	3	3	3	3	3	3	0

Continued on next page

Compound		Concentration			4 days			8 days			15 days			29 days		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
OPZ	Omeprazole	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PAN	Puromycin aminonucleoside	4 mg/kg	12 mg/kg		3	3	3	3	3	3	3	3	0	3	3	0
PAP	Papaverine	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PB	Phenobarbital	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PCT	Phenacetin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PEN	Penicillamine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PH	Perhexiline	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PHA	Phalloidin	NA	NA		0	0	3	0	0	3	0	0	0	0	0	0
PhB	Phenylbutazone	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PHE	Phenytoin	60 mg/kg	200 mg/kg	600 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PML	Pemoline	8 mg/kg	25 mg/kg	75 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PMZ	Promethazine	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PPL	Propranolol	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
PTU	Propylthiouracil	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
QND	Quinidine	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RAN	Ranitidine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RGZ	Rosiglitazone maleate	2 mg/kg	20 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
RIF	Rifampicin	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
ROT	Rotenone	5 mg/kg	15 mg/kg	50 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SLP	Sulpiride	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SS	Sulfasalazine	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SST	Simvastatin	40 mg/kg	120 mg/kg	400 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
SUL	Sulindac	5 mg/kg	15 mg/kg	50 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TAA	Thioacetamide	5 mg/kg	15 mg/kg	45 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TAC	Tacrine	3 mg/kg	10 mg/kg	30 mg/kg	3	3	3	3	3	3	3	3	3	3	3	1
TAN	Tannic acid	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TBF	Terbinafine	75 mg/kg	250 mg/kg	750 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TC	Tetracycline	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TCP	Ticlopidine	30 mg/kg	100 mg/kg	300 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TEO	Theophylline	20 mg/kg	60 mg/kg	200 mg/kg	3	3	3	3	3	3	3	3	3	3	3	2
TIO	Tiopronin	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TLB	Tolbutamide	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TMD	Trimethadione	50 mg/kg	150 mg/kg	500 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TMX	Tamoxifen	6 mg/kg	20 mg/kg	60 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TRI	Triamterene	15 mg/kg	50 mg/kg	150 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
TRZ	Thioridazine	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Continued on next page

Compound		Concentration			4 days			8 days			15 days			29 days		
Abbr.	Name	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
TZM	Triazolam	100 mg/kg	300 mg/kg	1000 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
VA	Vitamin A	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
VPA	Valproic acid	45 mg/kg	150 mg/kg	450 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3
WY	WY-14643	10 mg/kg	30 mg/kg	100 mg/kg	3	3	3	3	3	3	3	3	3	3	3	3

Table B.5: *NRW (in vivo): Matrix of the compounds exposed to rat liver cells. The table gives full and abbreviated compound names as well as the concentration in µM and the number of independent replicates of gene array data available after incubation for 6h, 12h, 48h, 1 day, 3 days, 7 days and 14 days. *For the compound PB no further information is provided.*

Compound			Time point								Category	
Abbr.	Name	Concentration	06h	12h	48h	1 day	3 days	7 days	14 days	Series	Class	Batch
2-NF	2-Nitrofluorene	44 mg/(kg day)	0	0	0	0	3	3	0	1C	Genotoxic carcinogen	Training
2AAF	2-Acetylaminofluorene	1.2 mg/(kg day)	0	0	0	3	3	3	3	1B	Genotoxic carcinogen	Validation
3-MC	3-Methylcholanthrene	25 mg/(kg day)	0	0	0	3	3	3	3	1M	Non-hepatocarcinogen	Validation
AA	Acetamide	3000 mg/(kg day)	0	0	0	3	3	3	3	1D	Non-genotoxic carcinogen	Validation
Aap	Acetaminophen	4.25 g/(kg day)	3	3	2	4	0	0	0	1J	Non-genotoxic carcinogen	Validation
AfB1	Aflatoxin B1	0.24 mg/(kg day)	0	0	0	0	3	3	0	1D	Genotoxic carcinogen	Training
AlAl	Allyl alcohol	36 mg/(kg day)	0	0	0	3	3	3	3	1F	Non-hepatocarcinogen	Validation
CFX	Cefuroxime	250 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Training
CIDB	C.I Direct Black	146 mg/(kg day)	0	0	0	0	3	3	0	1C	Genotoxic carcinogen	Training
Clon	Clonidine	0.1 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Validation
CPA	Cyproterone acetate	100 mg/(kg day)	0	0	0	3	3	3	3	1A	Non-genotoxic carcinogen	Validation
DCB	1,4-Dichlorobenzene	300 mg/(kg day)	0	0	0	3	3	3	3	1A	Non-hepatocarcinogen	Validation
DEHA	Dehydroepiandrosterone	600 mg/(kg day)	0	0	0	3	3	3	3	1D	Non-genotoxic carcinogen	Validation
DES	Diethylstilbestrol	10 mg/(kg day)	0	0	0	2	3	0	0	1H	Non-genotoxic carcinogen	Training
DMN	Dimethylnitrosamine	4 mg/(kg day)	0	0	0	0	3	3	0	1C	Genotoxic carcinogen	Training
ETH	Ethionine	200 mg/(kg day)	0	0	0	3	3	3	3	1H	Non-genotoxic carcinogen	Validation
Ibup	Ibuprofen	94 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Validation
Mcarb	Methylcarbamate	400 mg/(kg day)	0	0	0	3	3	3	3	1D	Non-genotoxic carcinogen	Validation
MDA	Methylendianiline	50 mg/(kg day)	0	0	0	3	3	3	3	1A	Genotoxic carcinogen	Validation
MPy	Methapyrilene HCl	60 mg/(kg day)	0	0	0	0	3	3	0	1D	Non-genotoxic carcinogen	Training
Nif	Nifedipine	3 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Training
NNK	4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone	20 mg/(kg day)	0	0	0	0	0	3	3	1F	Genotoxic carcinogen	Validation
NNM	N-Nitrososmorpholine	3.5 mg/(kg day)	0	0	0	0	3	3	0	1F	Genotoxic carcinogen	Training
NPip	N-Nitrosopiperidine	20 mg/(kg day)	0	0	0	3	3	3	3	1N	Genotoxic carcinogen	Validation
PB*	NA	NA	0	0	0	3	3	3	3	1B	NA	Validation
PBO	Piperonylbutoxide	1200 mg/(kg day)	0	0	0	3	3	0	0	1H	Non-genotoxic carcinogen	Training
Praz	Prazosin	1 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Validation
Prop	Propranolol	40 mg/(kg day)	0	0	0	3	3	3	3	1O	Non-hepatocarcinogen	Training
TAA	Thioacetamid	19.2 mg/(kg day)	0	0	0	0	3	3	0	1D	Non-genotoxic carcinogen	Training
Wy	Wy-14643	60 mg/(kg day)	0	0	0	3	3	0	0	1D	Non-genotoxic carcinogen	Training

Table B.6: *NRW (in vitro): Matrix of the compounds exposed to rat hypotocytes. The table gives full and abbreviated compound names as well as the concentration in µM and the number of independent replicates of gene array data available after incubation with a low, middle and high concentration for 24h.*

Compound		Concentration			24h			
Abbr.	Name	Low	Middle	High	Low	Middle	High	Class
2-NF	2-Nitrofluorene	4.76 µM	14.38 µM	43.13 µM	3	3	3	Genotoxic
2AAF	2-Acetylaminofluorene	0.11 µM	0.33 µM	0.99 µM	3	3	3	Genotoxic
3-MC	3-Methylcholanthrene	0.64 µM	1.92 µM	5.76 µM	3	3	2	Non-carcinogen
AA	Acetamide	111.1 µM	333.3 µM	1000 µM	3	3	3	Non-genotoxic
Aap	Acetaminophen	111 µM	333 µM	1000 µM	3	3	3	Non-genotoxic
AfB1	Aflatoxin B1	0.00031 µM	0.00093 µM	0.0028 µM	3	3	3	Genotoxic
AlAl	Allyl alcohol	3.08 µM	9.23 µM	27.69 µM	3	3	3	Non-carcinogen
CFX	Cefuroxime	75.45 µM	226.36 µM	679.09 µM	3	3	3	Non-carcinogen
CIDB	C.I Direct Black	1.92 µM	5.76 µM	17.28 µM	3	3	3	Genotoxic
Clon	Clonidine	0.17 µM	303 µM	909 µM	3	3	3	Non-carcinogen
CPA	Cyproterone acetate	1.11 µM	33 µM	100 µM	3	3	3	Non-genotoxic
DCB	1,4-Dichlorobenzene	111 µM	333 µM	1000 µM	3	3	3	Non-carcinogen
DEHA	Dehydroepiandrosterone	48.5 µM	88.9 µM	NA	3	3	NA	Non-genotoxic
DES	Diethylstilbestrol	1.35 µM	4.05 µM	12.15 µM	3	3	3	Non-genotoxic
DMN	Dimethylnitrosamine	39.19 µM	333 µM	1000 µM	3	3	3	Genotoxic
ETH	Ethionine	11.1 µM	33.3 µM	100 µM	3	3	3	Non-genotoxic
Ibup	Ibuprofen	102 µM	306.67 µM	920 µM	3	3	3	Non-carcinogen
Mcarb	Methylcarbamate	111.1 µM	333.3 µM	1000 µM	3	3	3	Non-genotoxic
MDA	Methylendianiline	1.09 µM	3.27 µM	9.81 µM	3	3	3	Genotoxic
MPy	Methapyrilene HCl	1.92 µM	14.78 µM	44.39 µM	3	3	3	Non-genotoxic
Nif	Nifedipine	6.94 µM	200 µM	600 µM	3	3	3	Non-carcinogen
NNK	4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanone	88 µM	333 µM	1000 µM	3	3	3	Genotoxic
NNM	N-Nitrososmorpholine	16.93 µM	50.78 µM	152.33 µM	3	3	3	Genotoxic
NPip	N-Nitrosopiperidine	49 µM	333 µM	1000 µM	3	3	3	Genotoxic
PBO	Piperonylbutoxide	3.52 µM	10.56 µM	31.69 µM	3	3	3	Non-genotoxic
Praz	Prazosin	0.00086 µM	0.00257 µM	0.0077 µM	3	3	3	Non-carcinogen
Prop	Propranolol	4.76 µM	14.29 µM	42.86 µM	3	3	3	Non-carcinogen
TAA	Thioacetamid	8.74 µM	26.22 µM	78.67 µM	3	3	3	Non-genotoxic
Wy	Wy-14643	60.05 µM	180.14 µM	540.43 µM	3	3	3	Non-genotoxic

Table B.7: *Summary statistics for the distributions of the estimated alert concentrations for Scenarios I-IV. The following parameters are presented: The total number of alerts (n), the median (Med), the interquartile range (IQR) and the standard deviation (SD). The method for estimating the alerts is subscripted after the corresponding parameter. An alert was given when the given fold change value of 1.5 was reached exactly (4pLL) or exceeded by the average value (Limma) (point estimate). The table refers to the Figures 6.2, C.30 and C.32.*

	n_{4pLL}	n_{Limma}	Med_{4pLL}	Med_{Limma}	IQR_{4pLL}	IQR_{Limma}	SD_{4pLL}	SD_{Limma}
Scenario I								
k=3	543	847	406.600	450	229.000	450	177.856	307.333
k=6	503	886	431.300	450	221.400	350	171.852	275.917
k=10	519	815	462.800	550	209.500	550	169.098	307.119
Scenario II								
k=3	988	988	502.200	550	40.100	0	40.789	96.717
k=6	996	997	499.600	550	30.400	0	31.084	69.232
k=10	1000	1000	500.400	550	22.200	0	22.251	53.692
Scenario III								
k=3	991	994	548.200	550	40.800	250	43.118	136.465
k=6	997	1000	549.600	550	31.100	250	32.698	128.133
k=10	996	1000	549.200	550	25.000	250	25.010	126.099
Scenario IV								
k=3	926	960	666.600	800	107.400	0	92.728	131.259
k=6	969	987	676.600	800	76.600	0	73.027	86.586
k=10	966	993	674.700	800	62.000	0	55.499	57.018

Table B.8: *Summary statistics for the distributions of the differences between the estimated alert concentrations and the respective true ALECs of the Scenarios II-IV. Scenario I was excluded from the analysis since no deviations could be computed (no ALEC value was provided). The following parameters are presented: The total number of alerts (n), the median (Med), the interquartile range (IQR) and the standard deviation (SD). The method for estimating the alerts is subscripted and indicated after the corresponding parameter. An alert was given when the given fold change value of 1.5 was reached exactly (4pLL) or exceeded by the average value (Limma) (point estimate). The table refers to the upper panels of the Figures 6.6, C.34 and C.35.*

	n_{4pLL}	n_{Limma}	Med_{4pLL}	Med_{Limma}	IQR_{4pLL}	IQR_{Limma}	SD_{4pLL}	SD_{Limma}
Scenario I								
No ALEC								
Scenario II								
k=3	988	988	2.383	50.170	40.070	0.000	40.789	96.717
k=6	996	997	-0.257	50.170	30.410	0.000	31.084	69.232
k=10	1000	1000	0.531	50.170	22.180	0.000	22.251	53.692
Scenario III								
k=3	991	994	-0.482	1.358	40.810	250.042	43.118	136.465
k=6	997	1000	0.937	1.358	31.120	250.042	32.698	128.133
k=10	996	1000	0.537	1.358	24.930	250.042	25.010	126.099
Scenario IV								
k=3	926	960	-12.480	120.900	107.450	0.000	92.728	115.675
k=6	969	987	-2.445	120.900	76.530	0.000	73.027	74.730
k=10	966	993	-4.340	120.900	62.030	0.000	55.499	46.622

Table B.9: *Summary statistics for the distributions of the differences between the estimated alert concentrations and the respective true ALECs of the Scenarios II-IV. Scenario I was excluded from the analysis since no deviations could be computed (no ALEC value was provided). The following parameters are presented: The total number of alerts (n), the median (Med), the interquartile range (IQR) and the standard deviation (SD). The method for estimating the alerts is subscripted and indicated after the corresponding parameter. An alert was given when the given fold change value of 1.5 was exceeded significantly ($p \leq 0.05$). The table refers to the lower panels of the Figures 6.6, C.34 and C.35*

	n_{4pLL}	n_{Limma}	Med_{4pLL}	Med_{Limma}	IQR_{4pLL}	IQR_{Limma}	SD_{4pLL}	SD_{Limma}
Scenario I								
No ALEC								
Scenario II								
k=3	316	999	285.200	300.200	163.200	250.030	114.745	125.684
k=6	427	1000	280.500	50.170	162.200	250.030	105.958	125.794
k=10	495	1000	285.900	50.170	162.300	250.030	102.972	117.702
Scenario III								
k=3	749	988	83.350	251.400	113.870	0.000	100.691	80.644
k=6	861	995	72.020	251.400	95.490	0.000	92.705	71.981
k=10	929	1000	60.580	251.400	69.050	0.000	80.034	69.939
Scenario IV								
k=3	285	648	66.150	320.900	175.080	200.000	115.488	113.294
k=6	419	833	100.500	120.900	153.710	200.000	105.302	98.027
k=10	570	910	93.140	120.900	135.780	200.000	97.921	87.823

C Figures

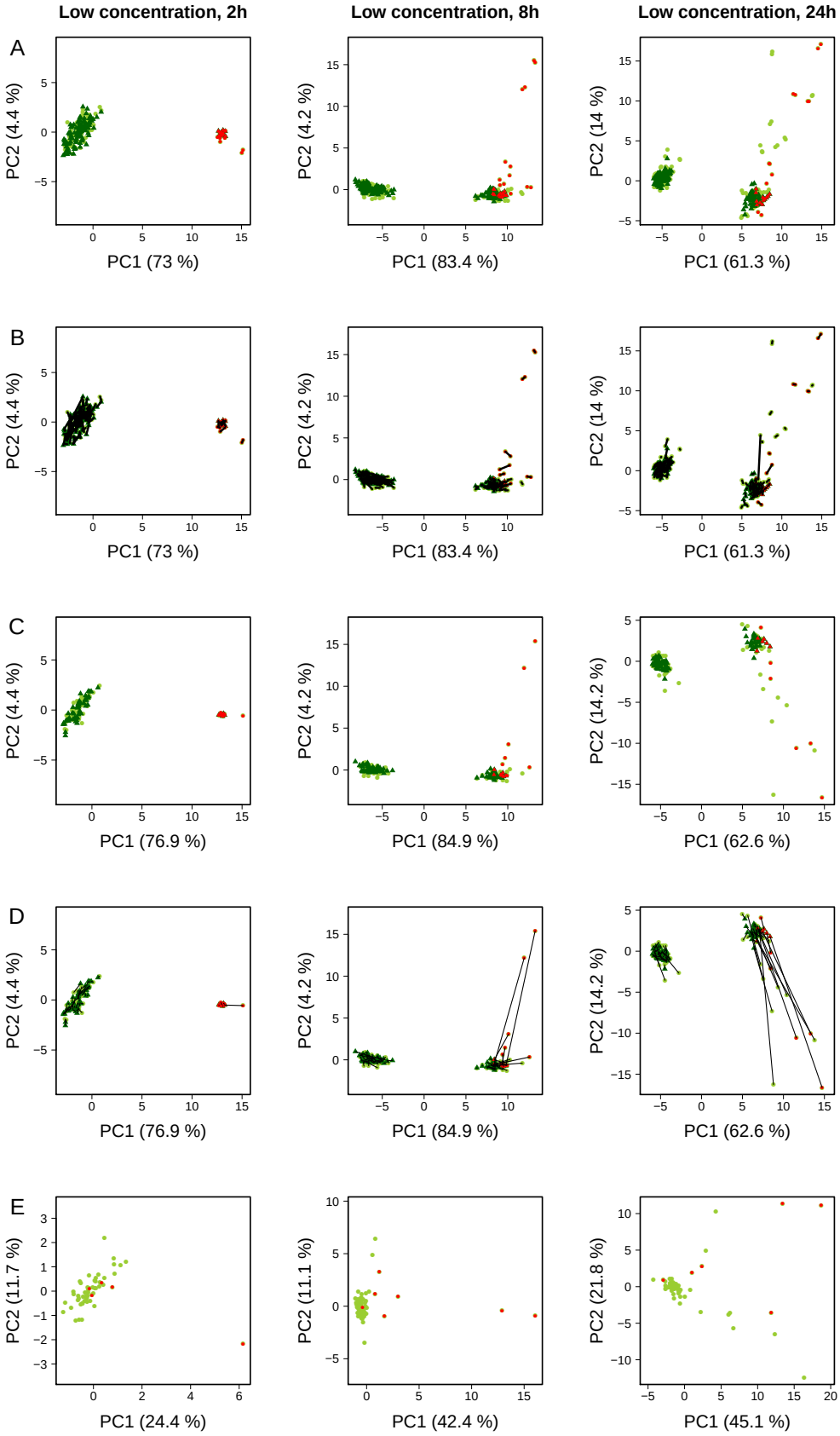


Figure C.1: *Corresponding data to Figure 4.1. Data of the low concentration and the incubation time points 2h, 8h and 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

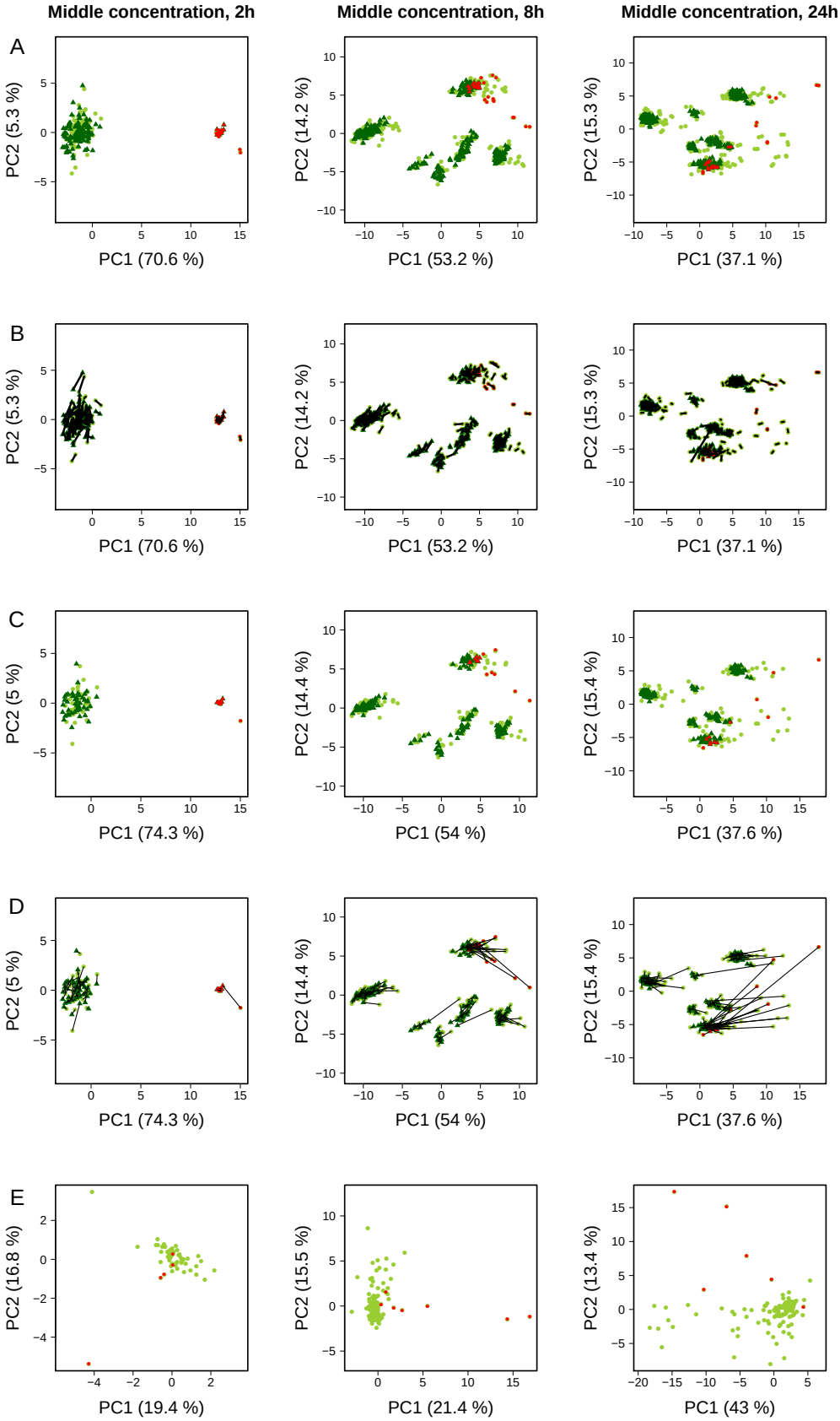


Figure C.2: *Corresponding data to Figure 4.1. Data of the middle concentration and the incubation time points 2h, 8h and 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

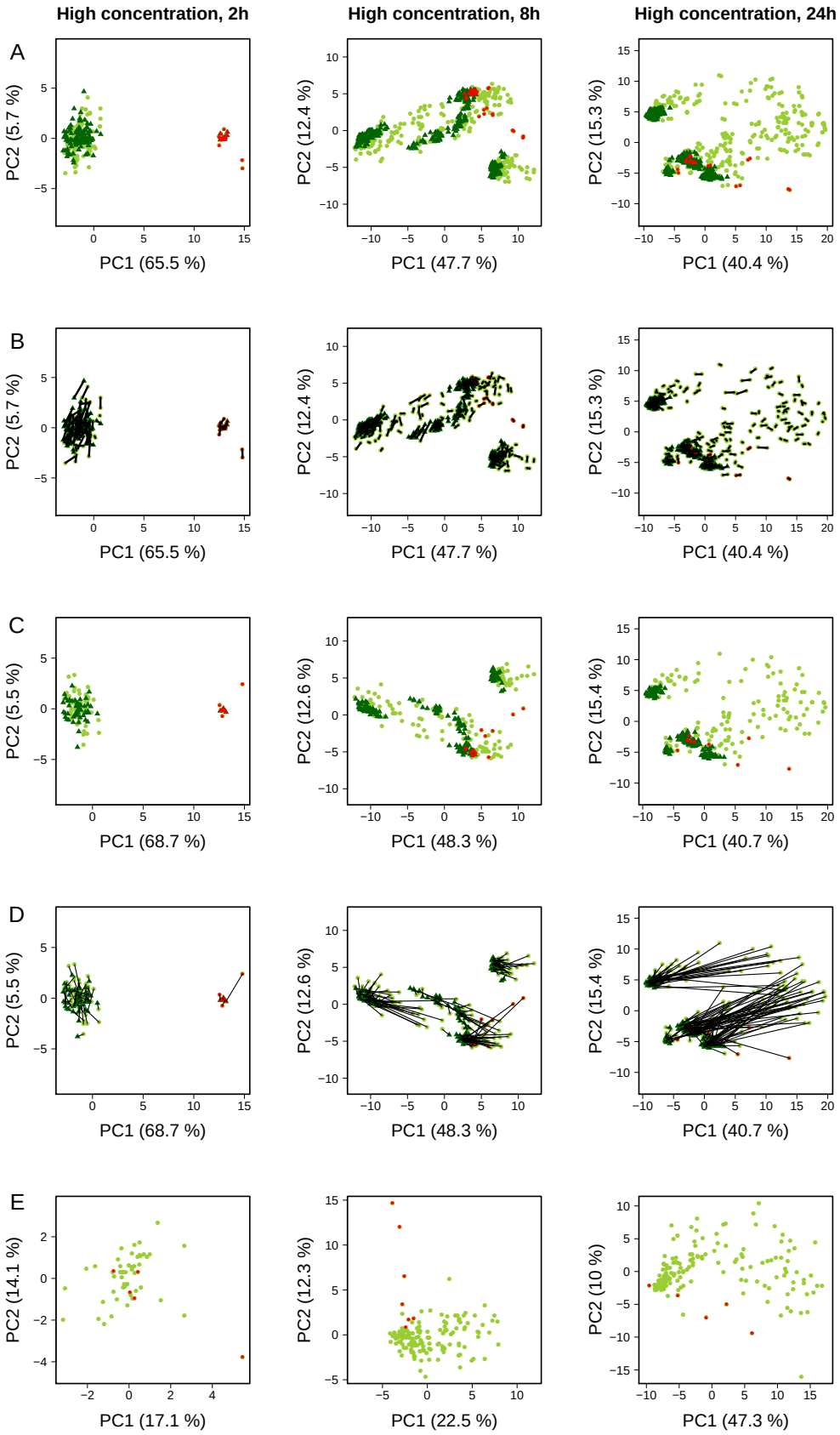


Figure C.3: *Corresponding data to Figure 4.1. Data of the high concentration and the incubation time points 2h, 8h and 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrates the controls and exposed samples, respectively. B. Connecting lines between replicates illustrate the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

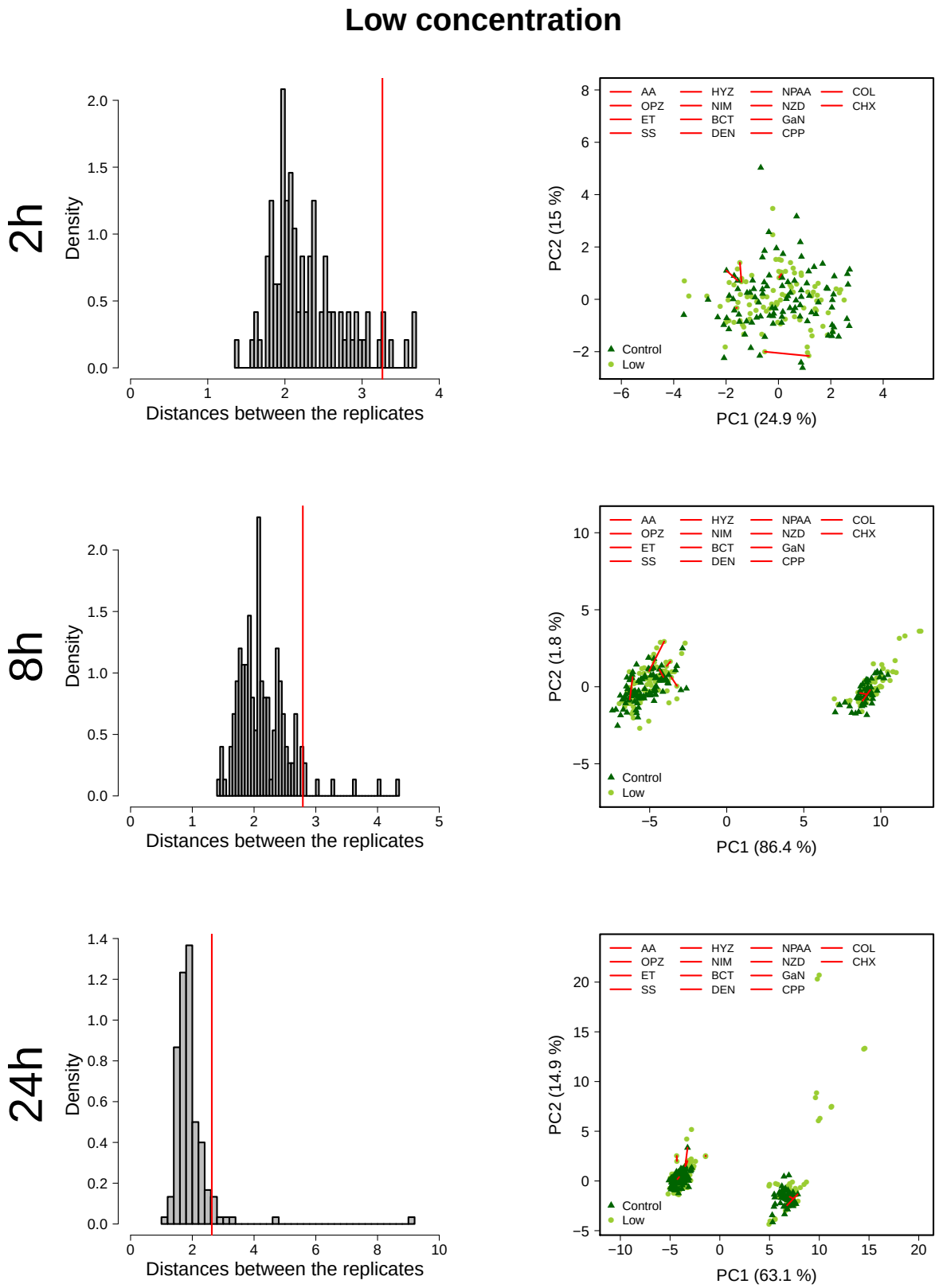


Figure C.4: *Reproducibility between replicates (low concentration). Left panel shows the frequency distribution of the Euclidean distance between all pairs of replicates. The red line indicates the 5% largest observed distances between replicates. The right panel shows the PCA plot with connecting lines between the 5% largest observed distances.*

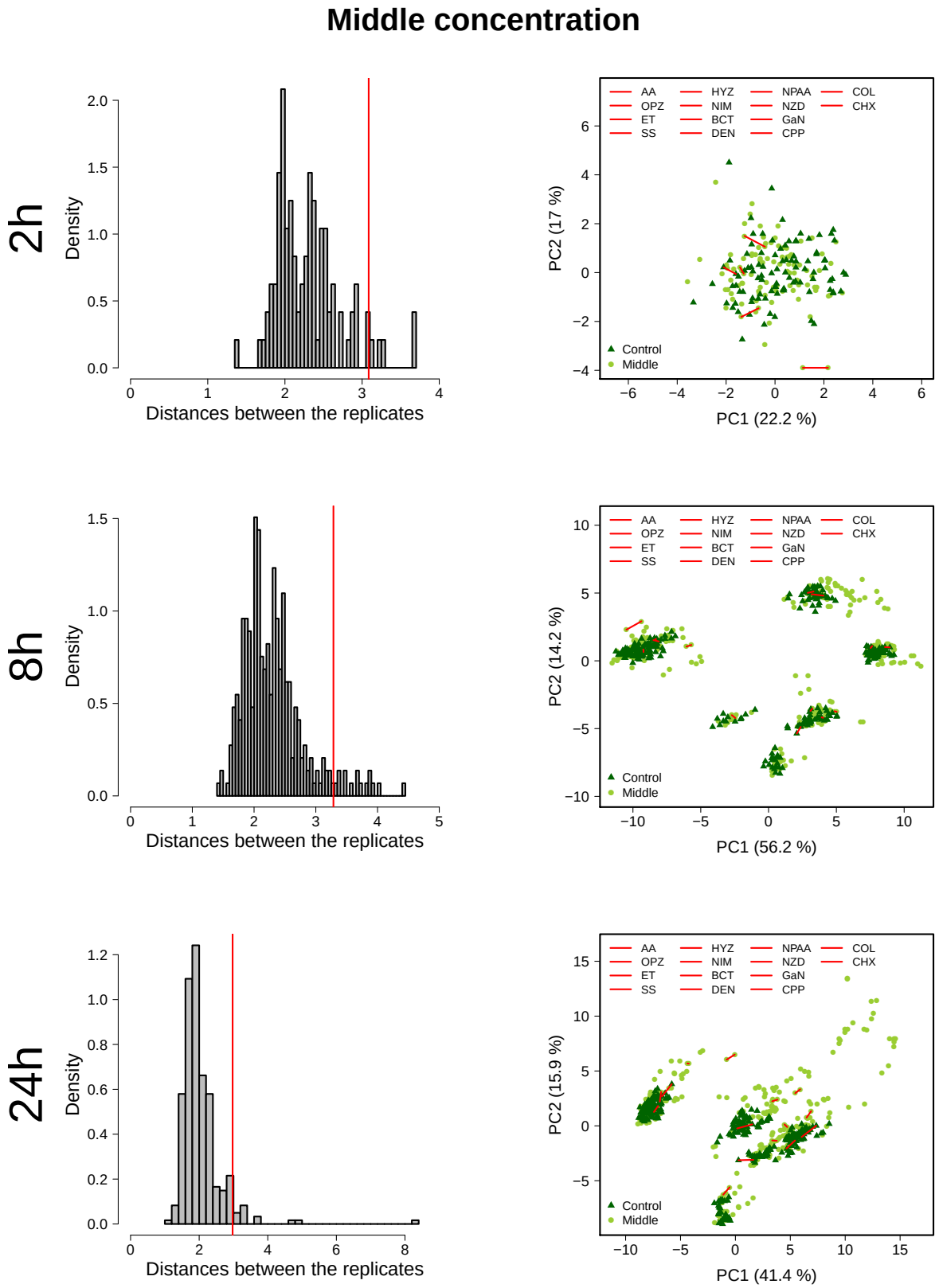


Figure C.5: *Reproducibility between replicates (middle concentration). Left panel shows the frequency distribution of the Euclidean distance between all pairs of replicates. The red line indicates the 5% largest observed distances between replicates. The right panel shows the PCA plot with connecting lines between the 5% largest observed distances.*

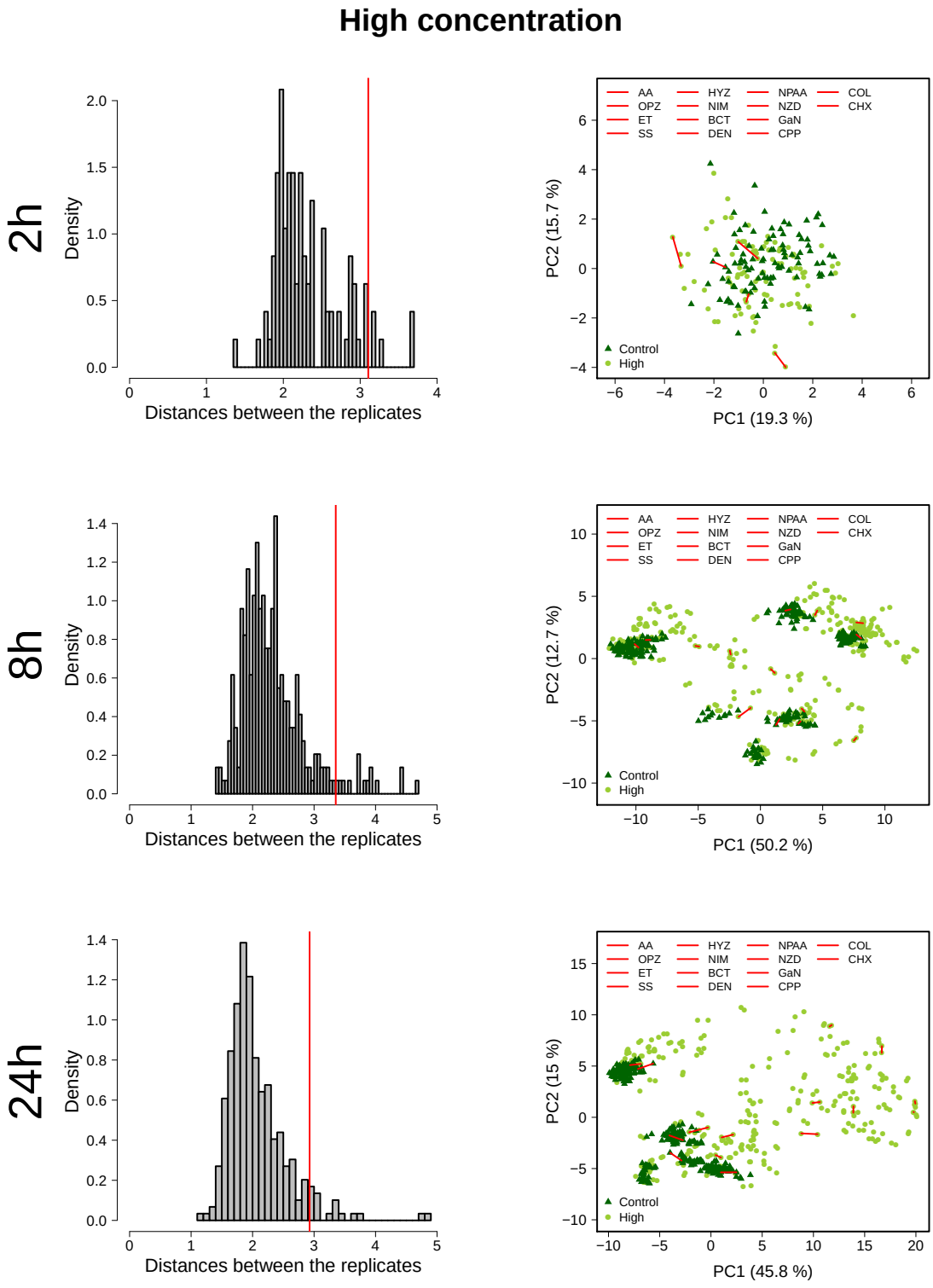


Figure C.6: *Reproducibility between replicates (high concentration). Left panel shows the frequency distribution of the Euclidean distance between all pairs of replicates. The red line indicates the 5% largest observed distances between replicates. The right panel shows the PCA plot with connecting lines between the 5% largest observed distances.*

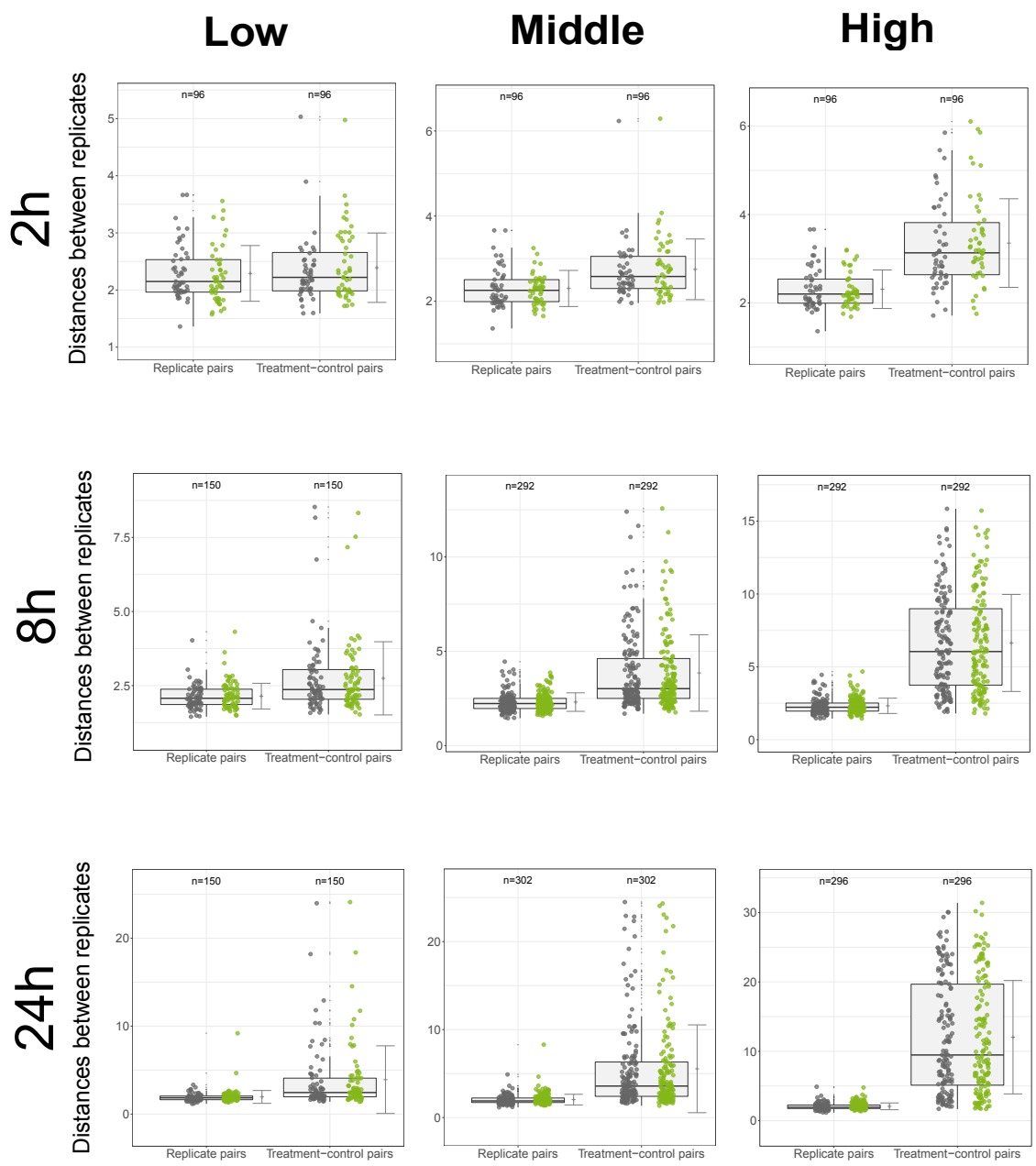


Figure C.7: *Reproducibility between replicates. Boxplots of the Euclidean distances between all pairs of replicates for all test conditions. The grey and green points in each left boxplot illustrate the distances between the control and exposed replicates; those in each right boxplot show the distances between control-treatment pairs (matched pairs).*

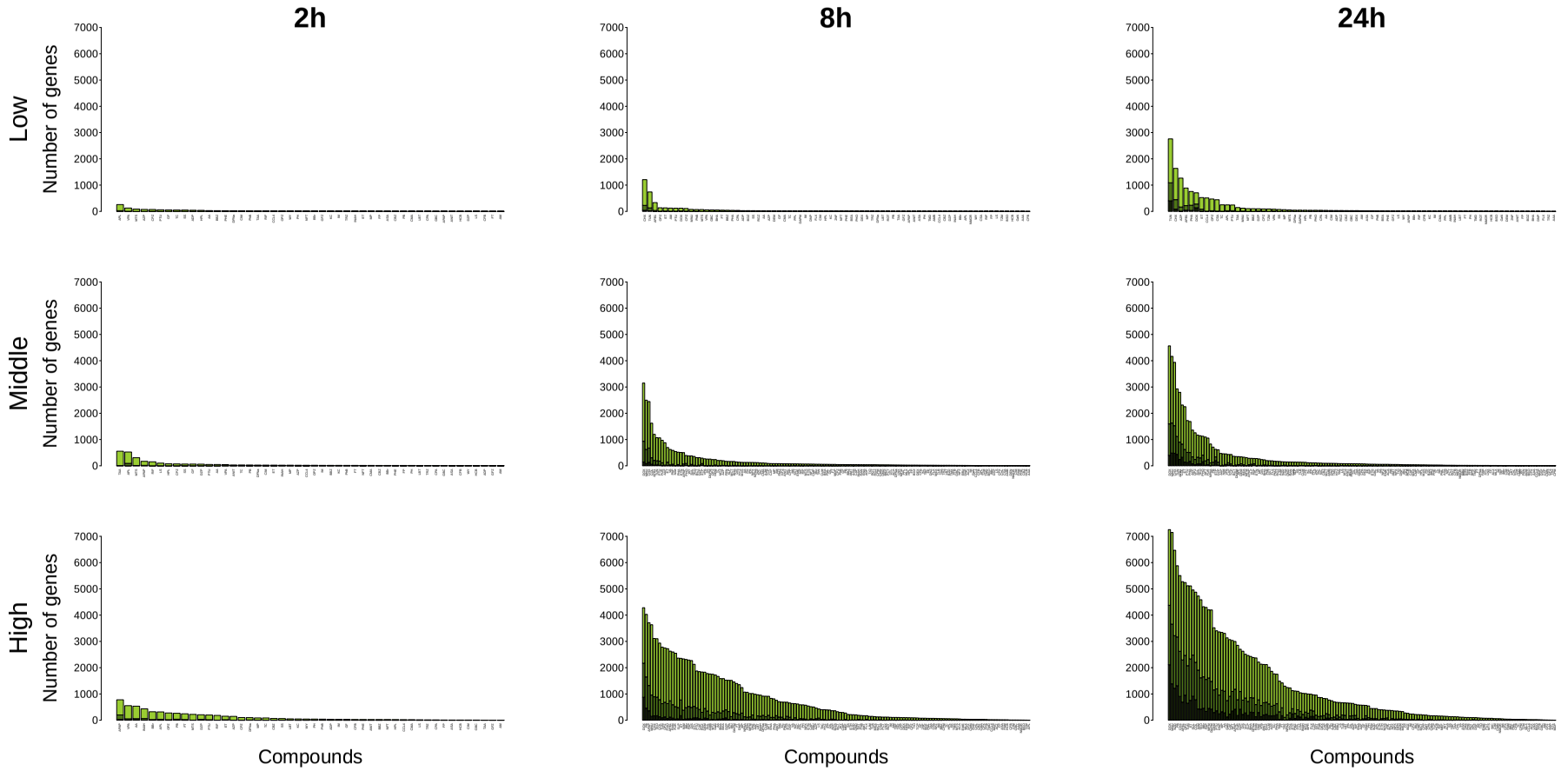


Figure C.8: *Number of significantly downregulated genes. The x-axis lists all chemicals that were tested at the indicated concentration for the corresponding period. The y-axis gives the number of downregulated genes with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. The corresponding data for the upregulated genes is shown in Figure 4.3. Dark green: more than 1.5-fold downregulated; light green: more than twofold downregulated; black: more than threefold downregulated.*

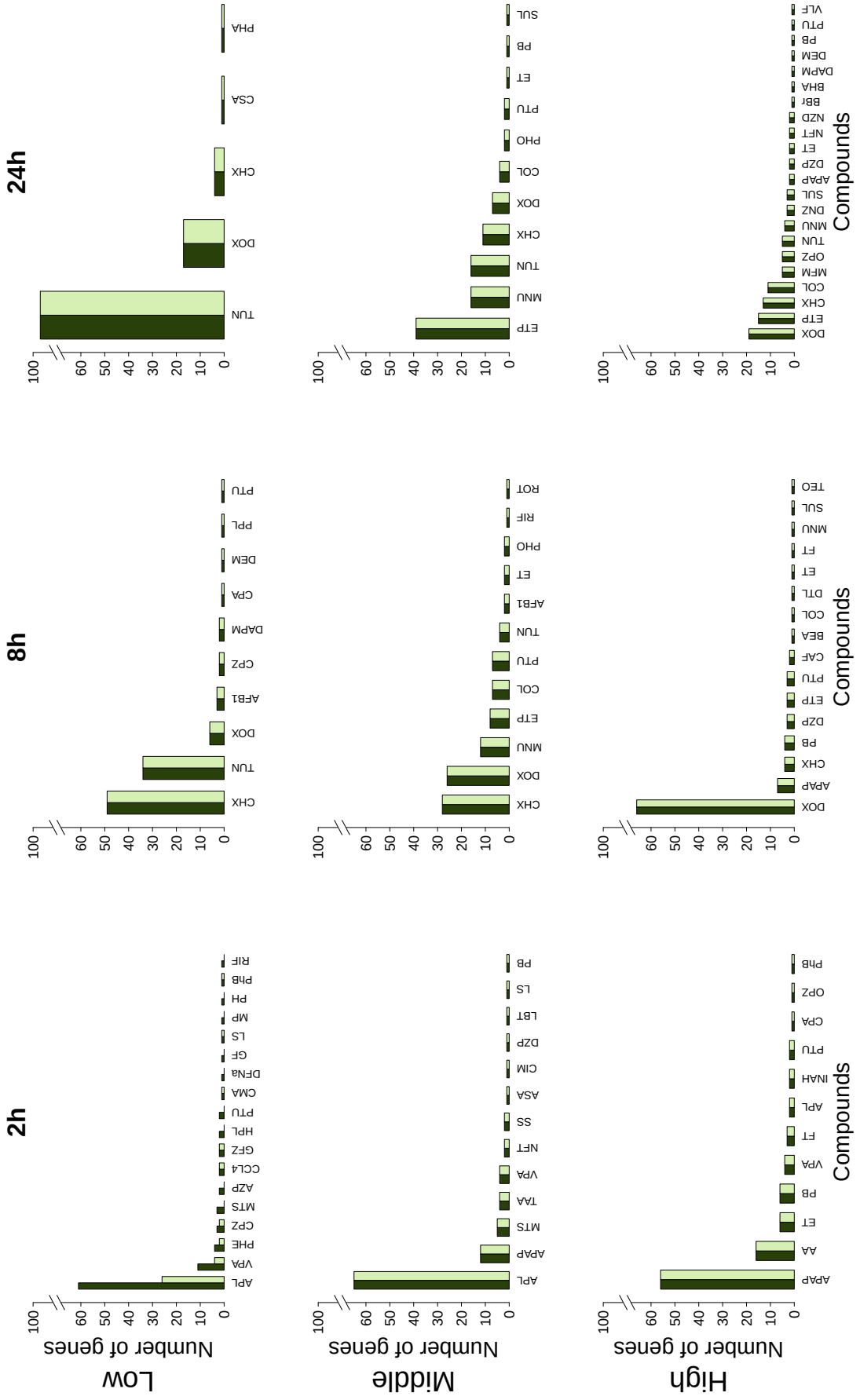


Figure C.9: *Exclusivity analysis of the downregulated genes. This analysis first determines the 100 strongest downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The corresponding analysis for the upregulated genes is shown in Figure 4.4.*

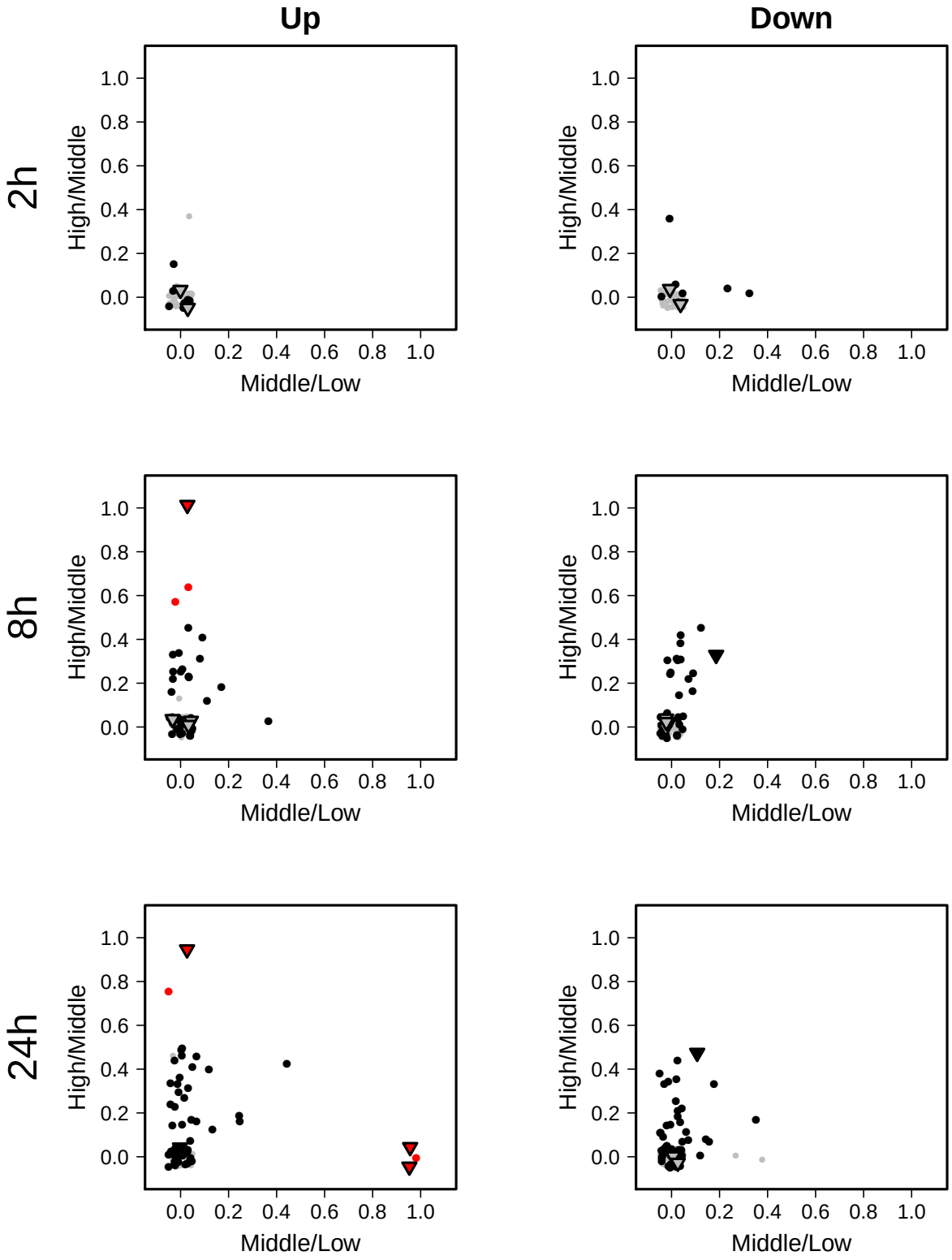


Figure C.10: *Modified progression profile error indicator. The progression profile error indicator values have been modified such that the values were set to zero if they were greater than 0.5 but the indicated compound has deregulated less than (or equal to) 20 genes at the respective lower concentration. A high value means that a high fraction of genes is deregulated exclusively at a lower compared to a respective higher concentration. Each symbol represents an individual compound. The triangles present the excluded compounds. Grey symbols: less than or equal to 20 genes are deregulated in total; black symbols: more than 20 genes are deregulated in total and both values are smaller than or equal to 0.5; red symbols: more than 20 genes are deregulated in total and at least one of the error indicator values is greater than 0.5.*

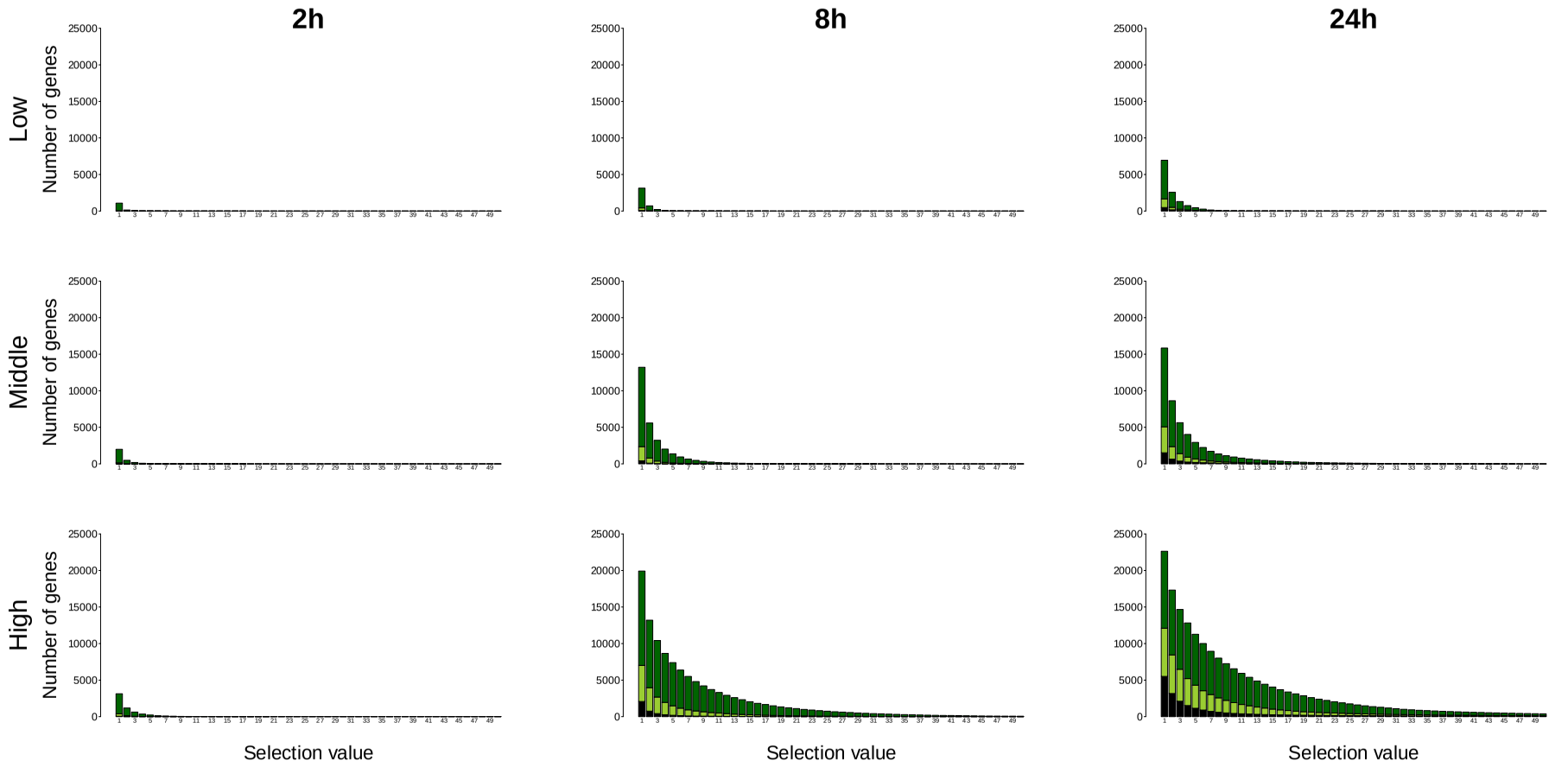


Figure C.11: *Selection values for the downregulated genes. A selection value of e.g. three means that at least three compounds downregulate (> threefold) the indicated gene. The corresponding analysis for the upregulated genes is shown in Figure 4.8.*

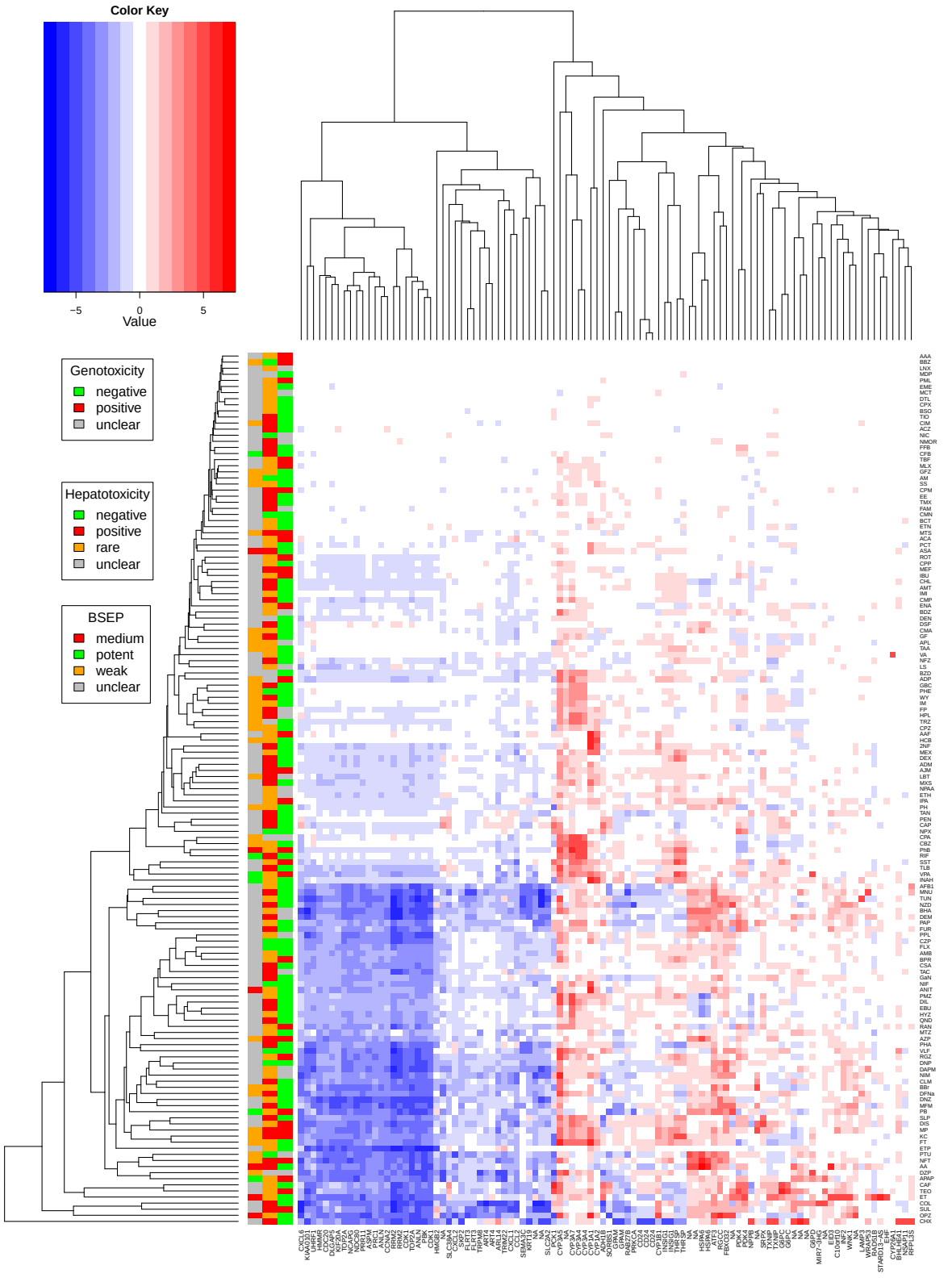


Figure C.12: *Unsupervised clustering of the 100 most deregulated genes across all compounds tested at the highest concentration for 24h of incubation. The lines represent the compounds, while each column stands for a gene. Red color indicates up and blue color downregulated genes as indicated by the code in the upper left. Moreover, the compounds have been classified with respect to their genotoxicity, human hepatotoxicity, and BSEP inhibiting capacity. These properties are indicated in the columns left of the heatmap. Unsupervised clustering results in three clusters that can be associated with biological motifs, proliferation, cytochrome P450 (CYP), and stress response.*

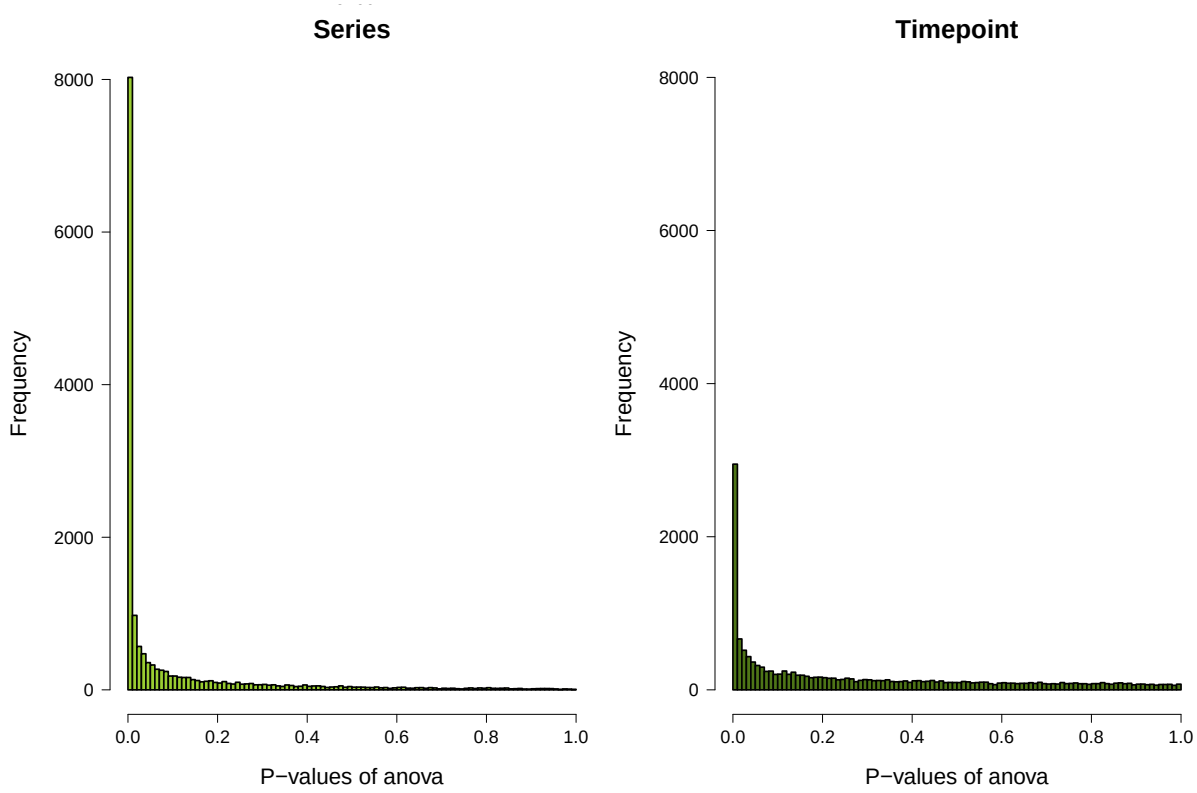


Figure C.13: *P-values of ANOVA (analysis of variance): The analysis was performed for the in vivo NRW experiments to test whether the model parameters experimental series and exposure period have an influence on gene expression. The analysis was performed gene-wise. The left panel shows the results for the factor experimental series and the right panel shows the results for the factor exposure period. Small p-values are indicative of strong parameter influence.*

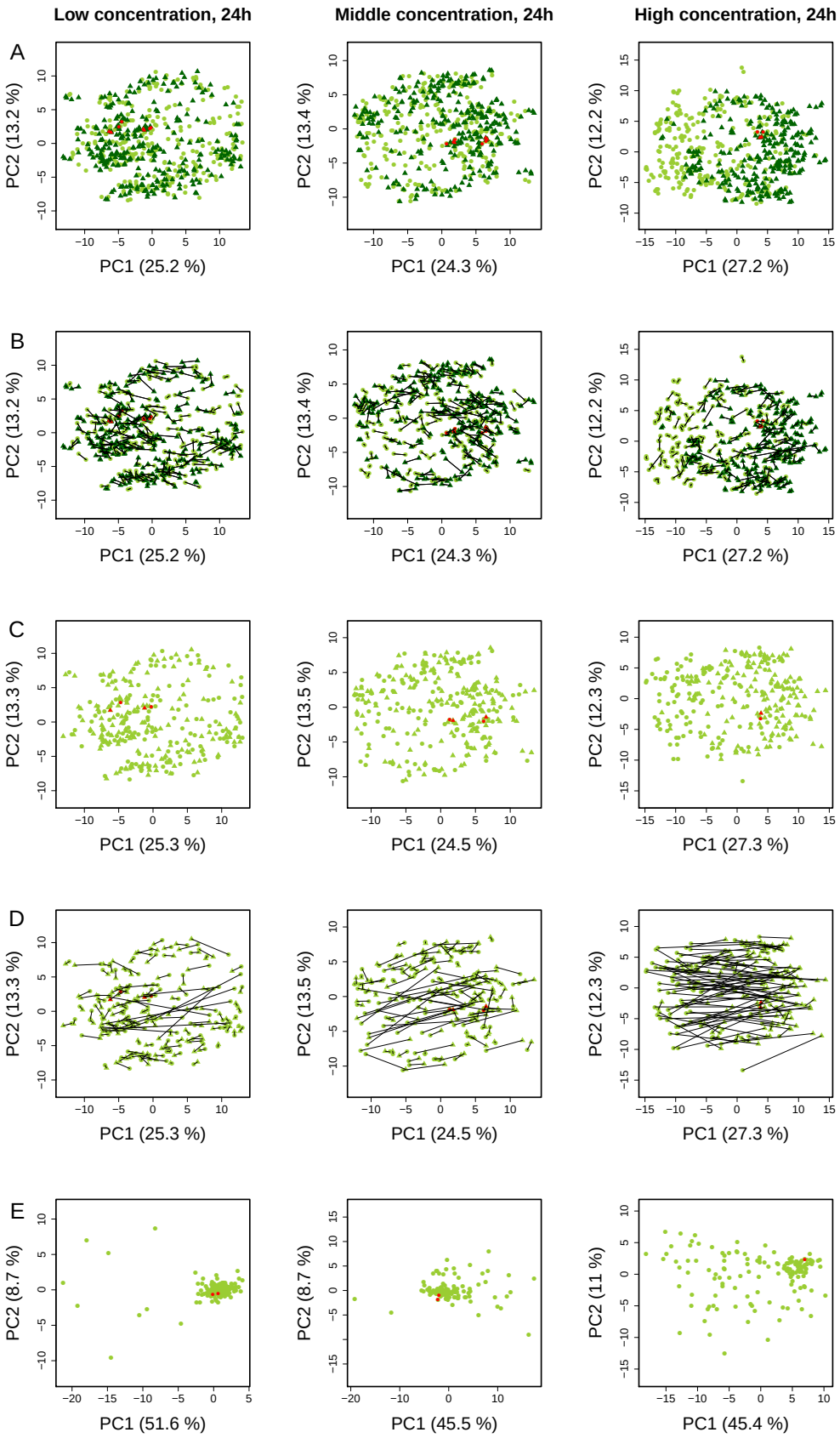


Figure C.14: *TGD (in vitro): Data of the low, middle and high concentration and the incubation time point 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

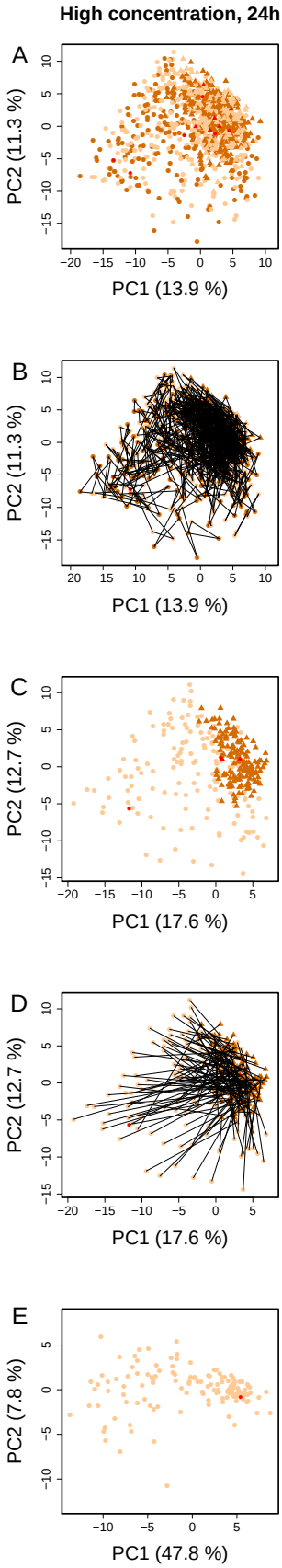


Figure C.15: *TGD (in vivo): Data of the high concentration and the incubation time point 24h. A. Overview of all samples and replicates. The dark and light green symbols illustrate the controls and exposed samples, respectively. B. Connecting lines between replicates illustrates the degree of variability. C. Mean values of the replicates. D. Connecting lines between controls (dark green) and corresponding compound exposed samples (light green). E. Subtraction of the controls from the corresponding compound exposed samples.*

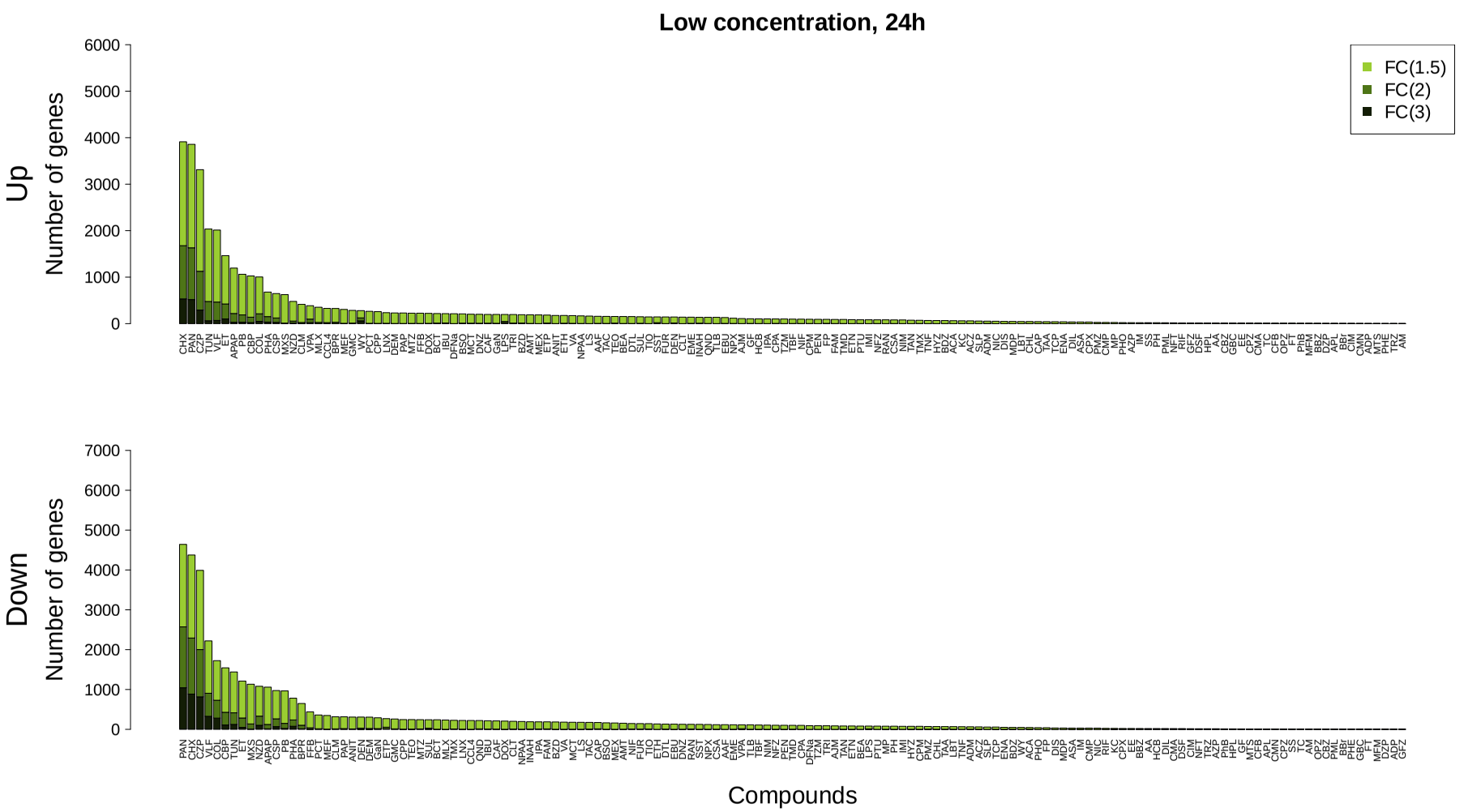


Figure C.16: *TGD (in vitro): Number of significantly deregulated genes. On the x-axis all chemicals are listed that were tested at the low concentration for 24h. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. Light green: more than 1.5-fold deregulated; middle green: more than twofold deregulated; dark green: more than threefold deregulated.*

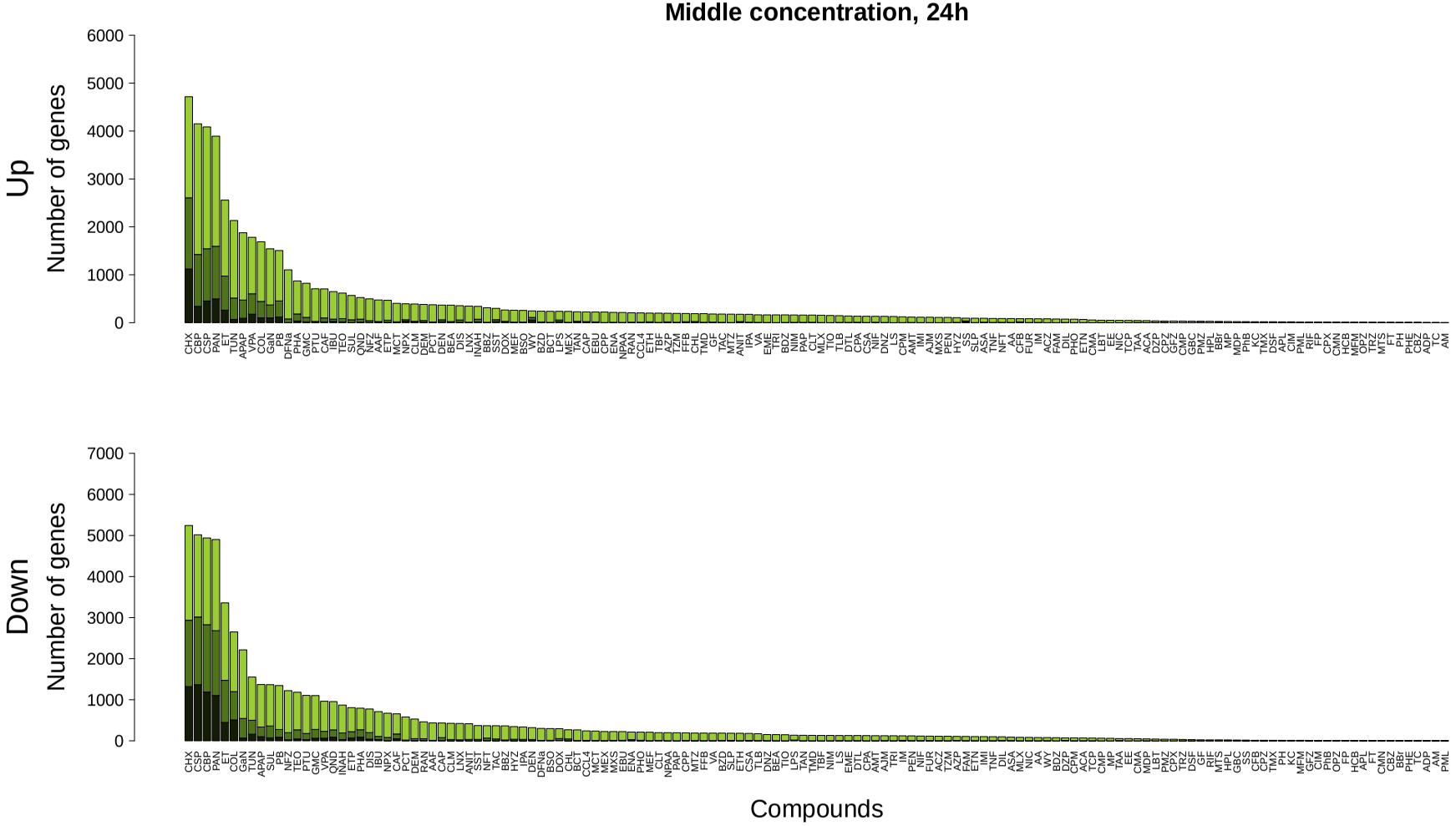


Figure C.17: *TGD (in vitro): Number of significantly deregulated genes. On the x-axis all chemicals are listed that were tested at the middle concentration for 24h. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. Light green: more than 1.5-fold deregulated; middle green: more than twofold deregulated; dark green: more than threefold deregulated.*

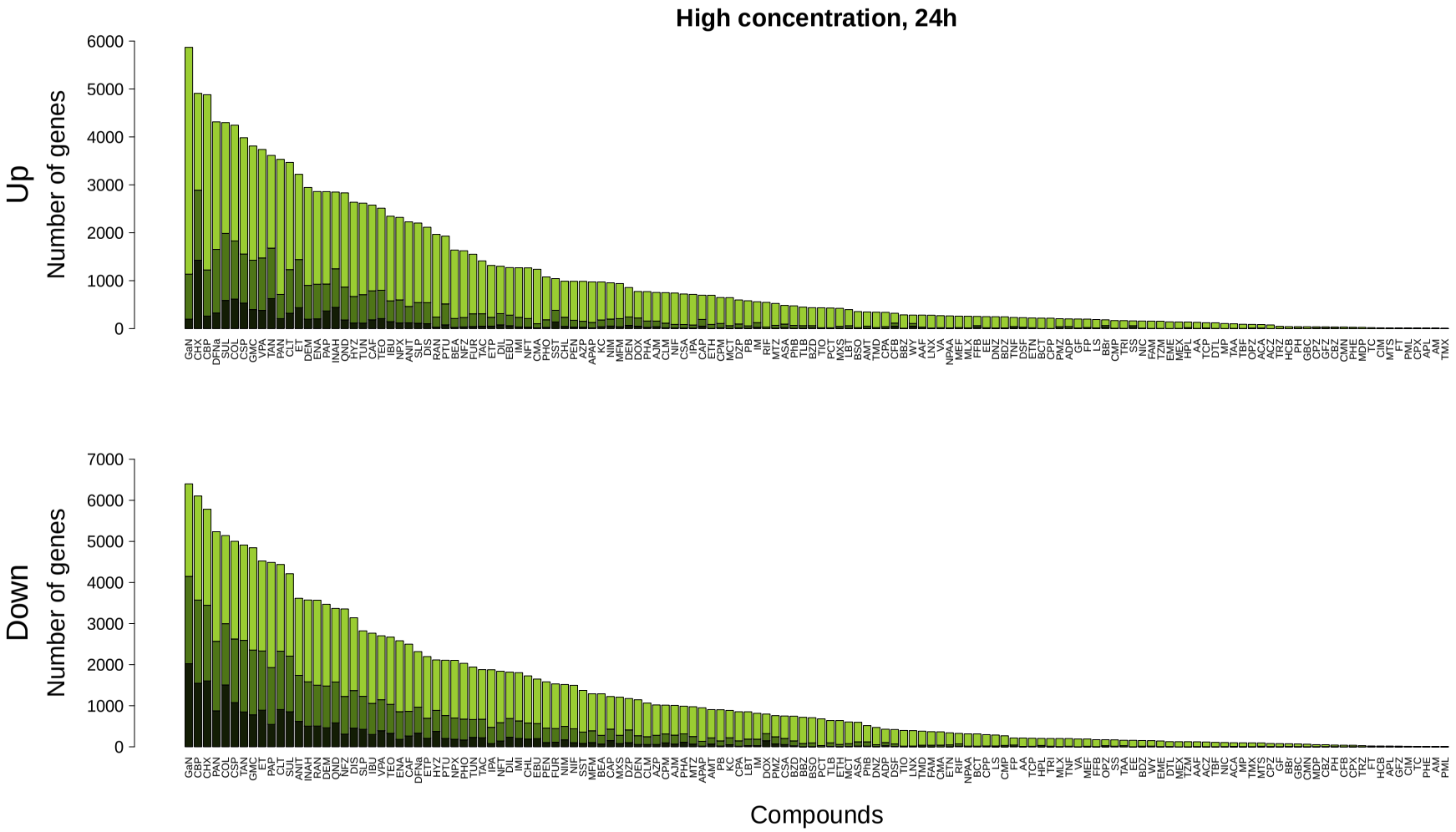


Figure C.18: *TGD (in vitro): Number of significantly deregulated genes. On the x-axis all chemicals are listed that were tested at the high concentration for 24h. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. Light green: more than 1.5-fold deregulated; middle green: more than twofold deregulated; dark green: more than threefold deregulated.*

Figure C.19: *TGD (in vivo): Number of significantly deregulated genes. On the x-axis all chemicals are listed that were tested at the high concentration for 24h. The y-axis gives the number of up- and downregulated genes (upper and lower panel) with at least 1.5-, 2.0- and 3.0-fold change. The result shows that the number of deregulated genes differs strongly between the chemicals. Light orange: more than 1.5-fold deregulated; middle orange: more than twofold deregulated; dark orange: more than threefold deregulated.*

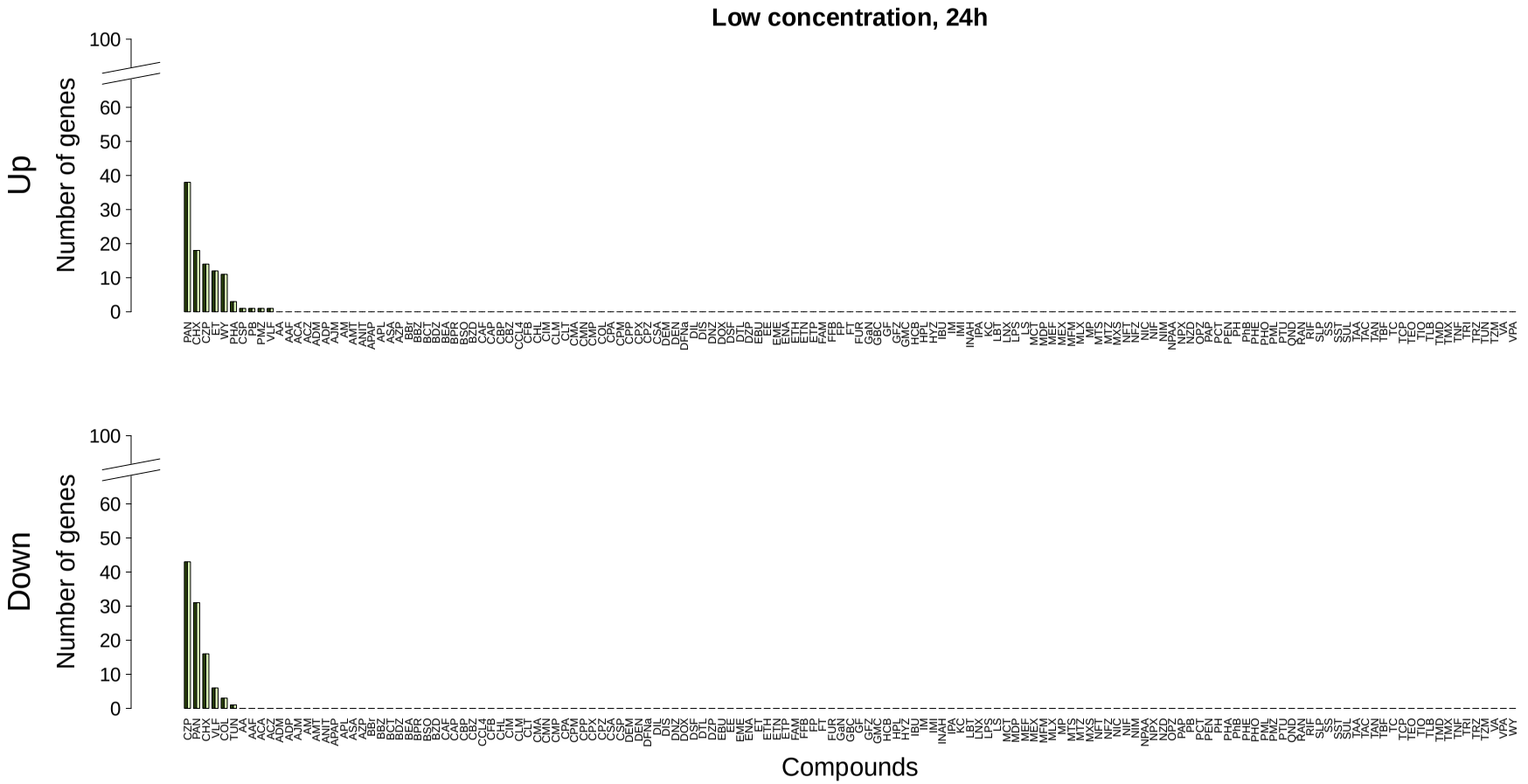


Figure C.20: *TGD (in vitro): Exclusivity analysis of the genes up- and downregulated at the low concentration after 24h exposure (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light green colored barplots indicate whether the top ranking genes meet the criteria for significance (> |3|-fold).*

Figure C.21: *TGD (in vitro): Exclusivity analysis of the genes up- and downregulated at the middle concentration after 24h exposure (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light green colored barplots indicate whether the top ranking genes meet the criteria for significance (> |3|-fold).*

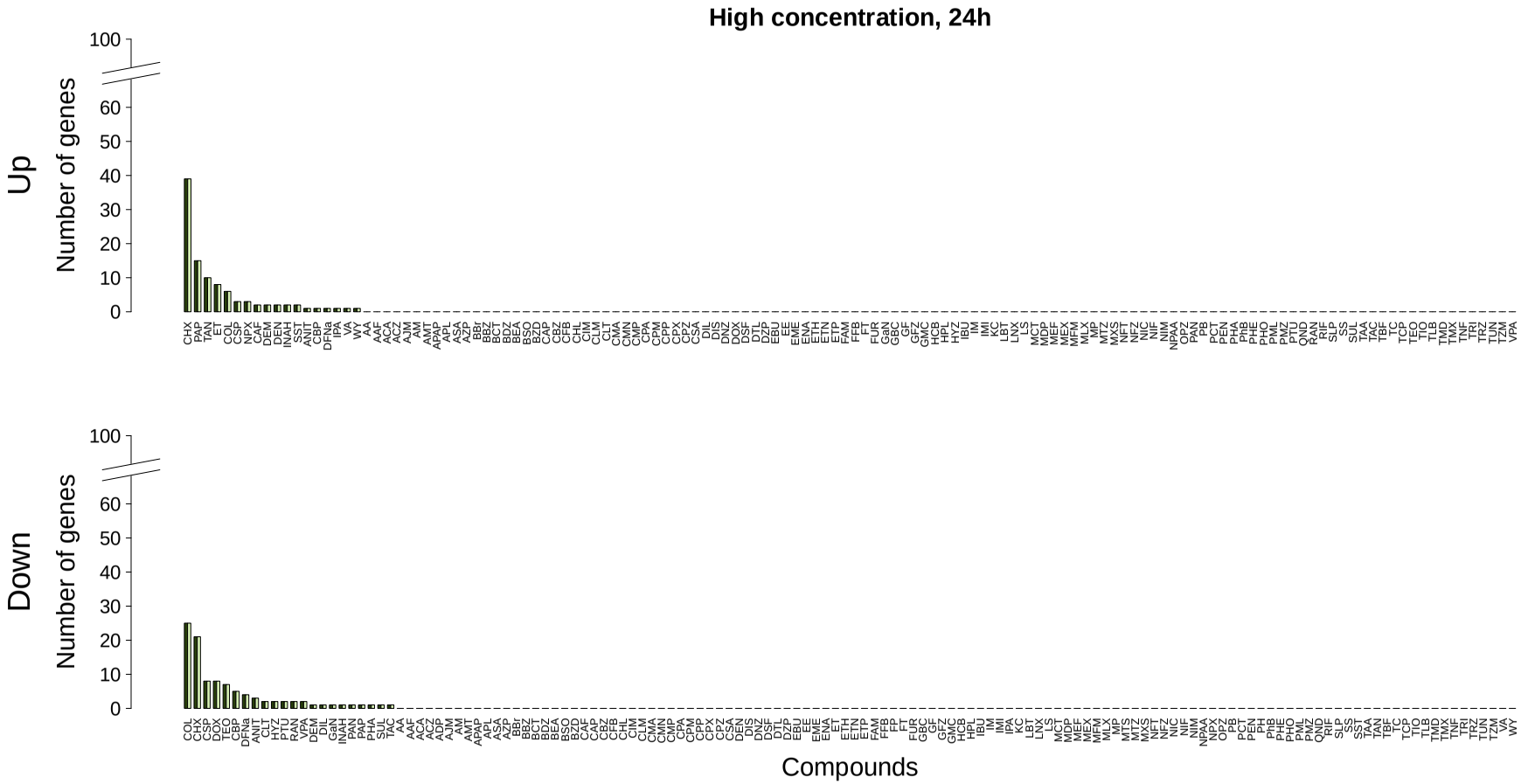
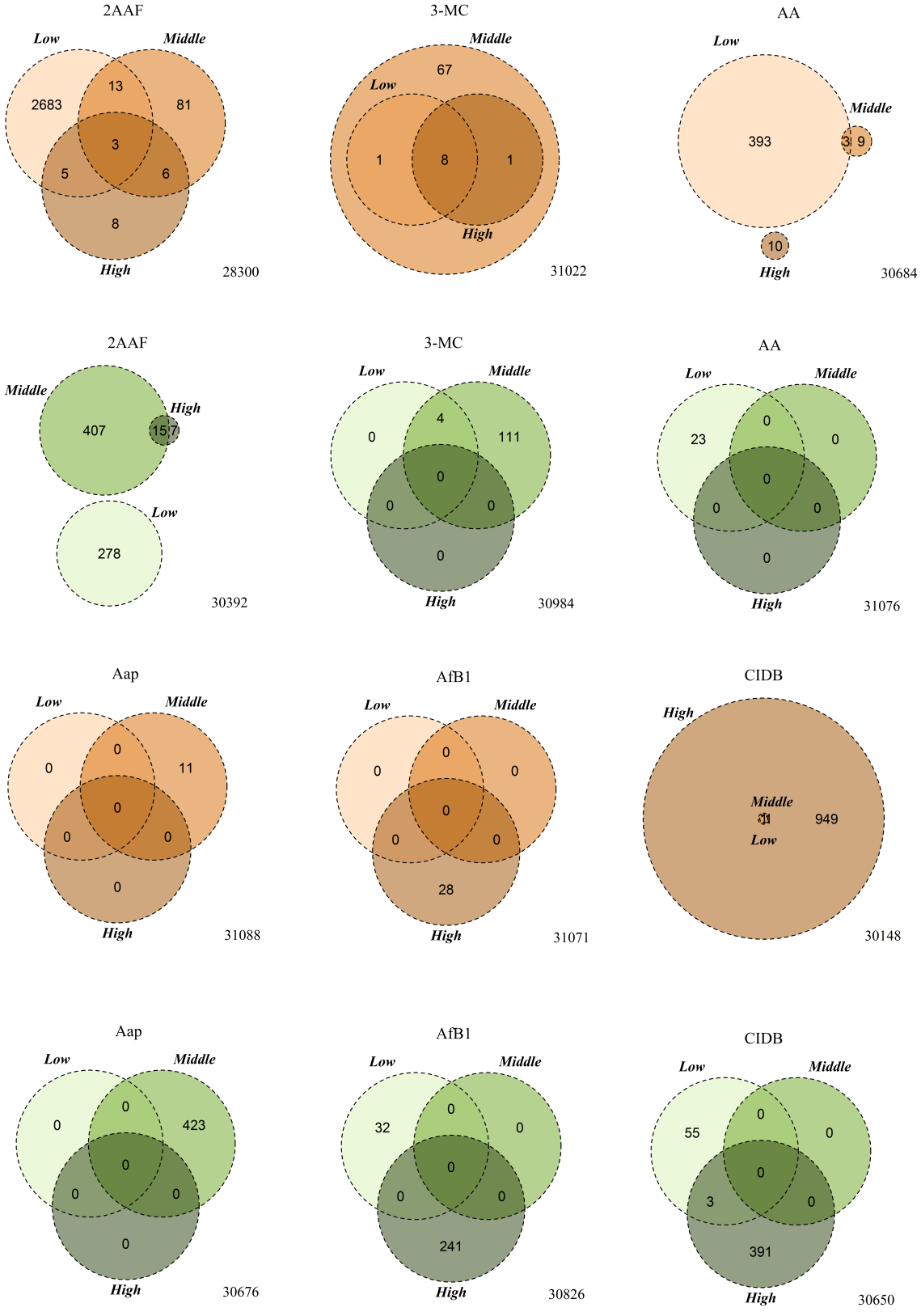
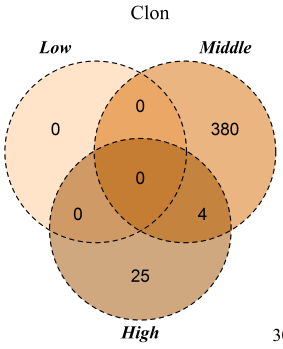


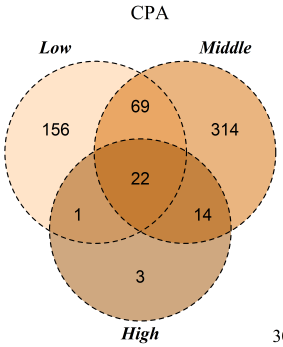
Figure C.22: *TGD (in vitro): Exclusivity analysis of the genes up- and downregulated at the high concentration after 24h exposure (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light green colored barplots indicate whether the top ranking genes meet the criteria for significance (> |3|-fold).*

Figure C.23: *TGD (in vivo): Exclusivity analysis of the genes up- and downregulated at the high concentration after 24h exposure (upper and lower panel). This analysis first determines the 100 strongest up- or downregulated genes across all compounds. Next, these genes are assigned to the compound with the most extreme fold change. The light green colored barplots indicate whether the top ranking genes meet the criteria for significance (> |3|-fold with adjusted $p \leq 0.01$).*

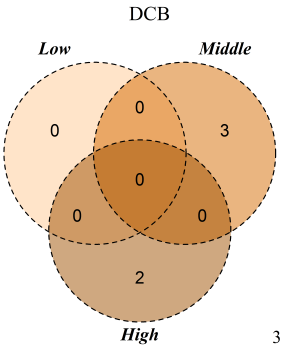

2AAF
Low
Middle
2683
13
81
3
5
6
8
High
28300
3-MC
Middle
67
Low
1
8
1
High
31022
AA
Low
Middle
393
3
9
10
High
30684
2AAF
Middle
High
407
15
7
Low
278
30392
3-MC
Low
Middle
0
4
111
0
0
0
0
High
30984
AA
Low
Middle
23
0
0
0
0
0
0
High
31076
Aap
Low
Middle
0
0
11
0
0
0
0
High
31088
AfB1
Low
Middle
0
0
0
0
0
0
28
High
31071
CIDB
High
Middle
11
949
Low
30148
Aap
Low
Middle
0
0
423
0
0
0
0
High
30676
AfB1
Low
Middle
32
0
0
0
0
0
241
High
30826
CIDB
Low
Middle
55
0
0
3
0
391
High
30650



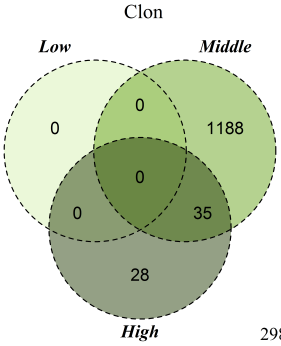
30690



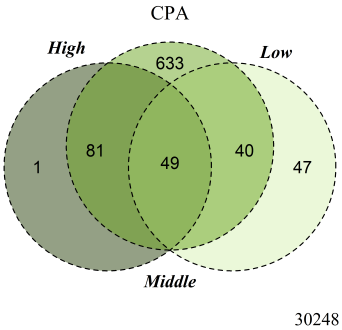
30520



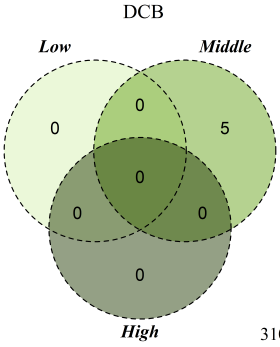
31094



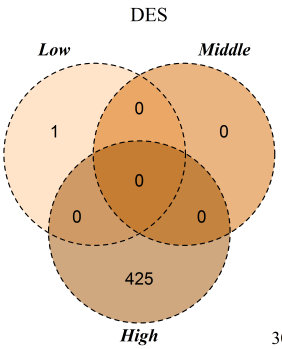
29848



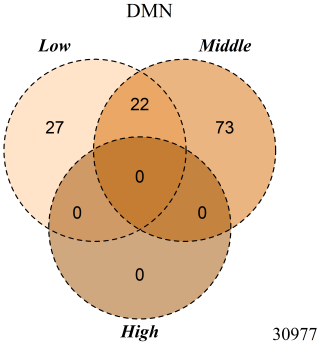
30248



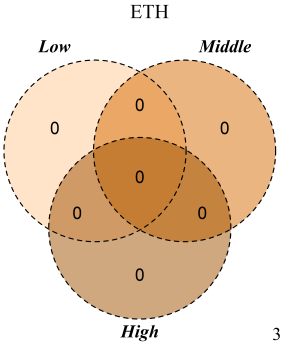
31094



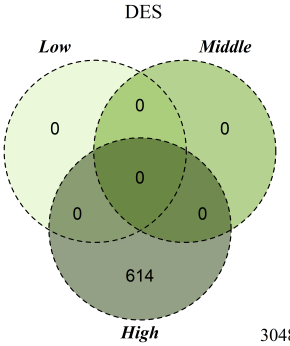
30673



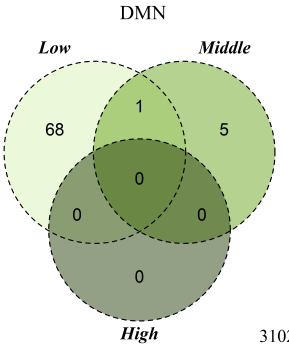
30977



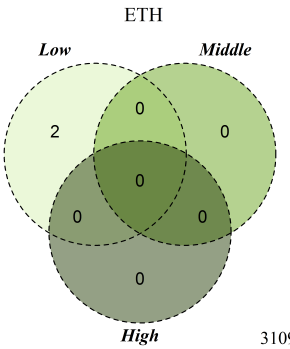
31099



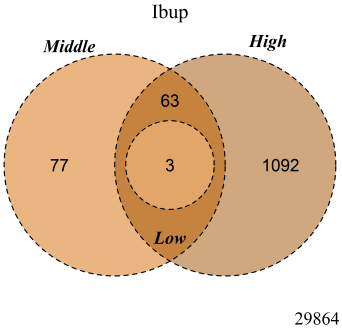
30485

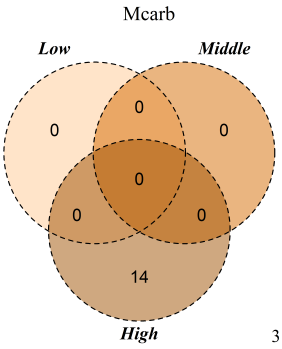


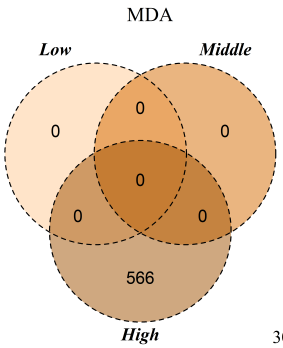
31025



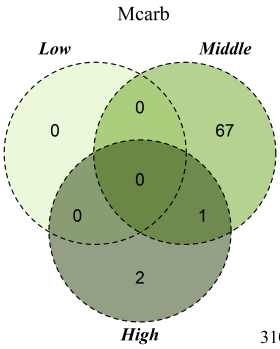
31097

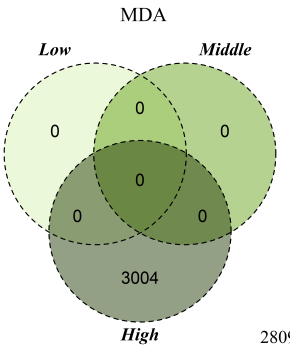

Ibup
Middle
High
63
77
3
1092
Low
29864

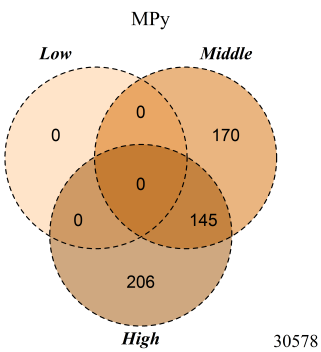

Mcarb
Low
Middle
0
0
0
0
0
0
14
High
31085

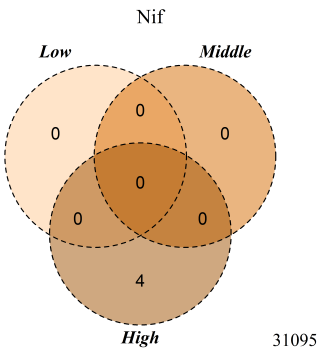

MDA
Low
Middle
0
0
0
0
0
0
566
High
30533

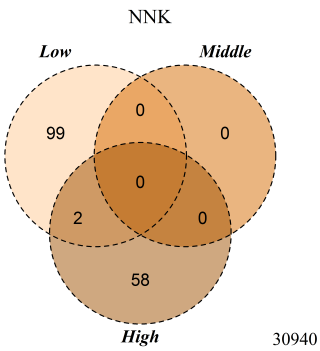

Ibup
Low
Middle
0
0
19
0
0
35
2867
High
28178


Mcarb
Low
Middle
0
0
67
0
0
1
2
High
31029

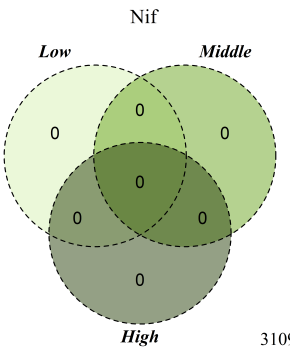

MDA
Low
Middle
0
0
0
0
0
0
3004
High
28095

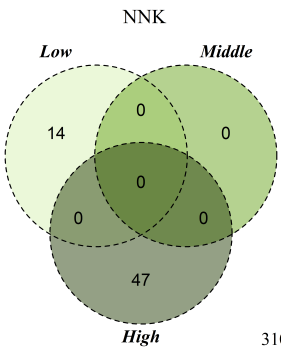

MPy
Low
Middle
0
0
170
0
0
145
206
High
30578


Nif
Low
Middle
0
0
0
0
0
0
4
High
31095


NNK
Low
Middle
99
0
0
0
2
0
58
High
30940


MPy
Low
Middle
0
0
256
0
0
678
1182
High
28983


Nif
Low
Middle
0
0
0
0
0
0
0
High
31099


NNK
Low
Middle
14
0
0
0
0
0
47
High
31038

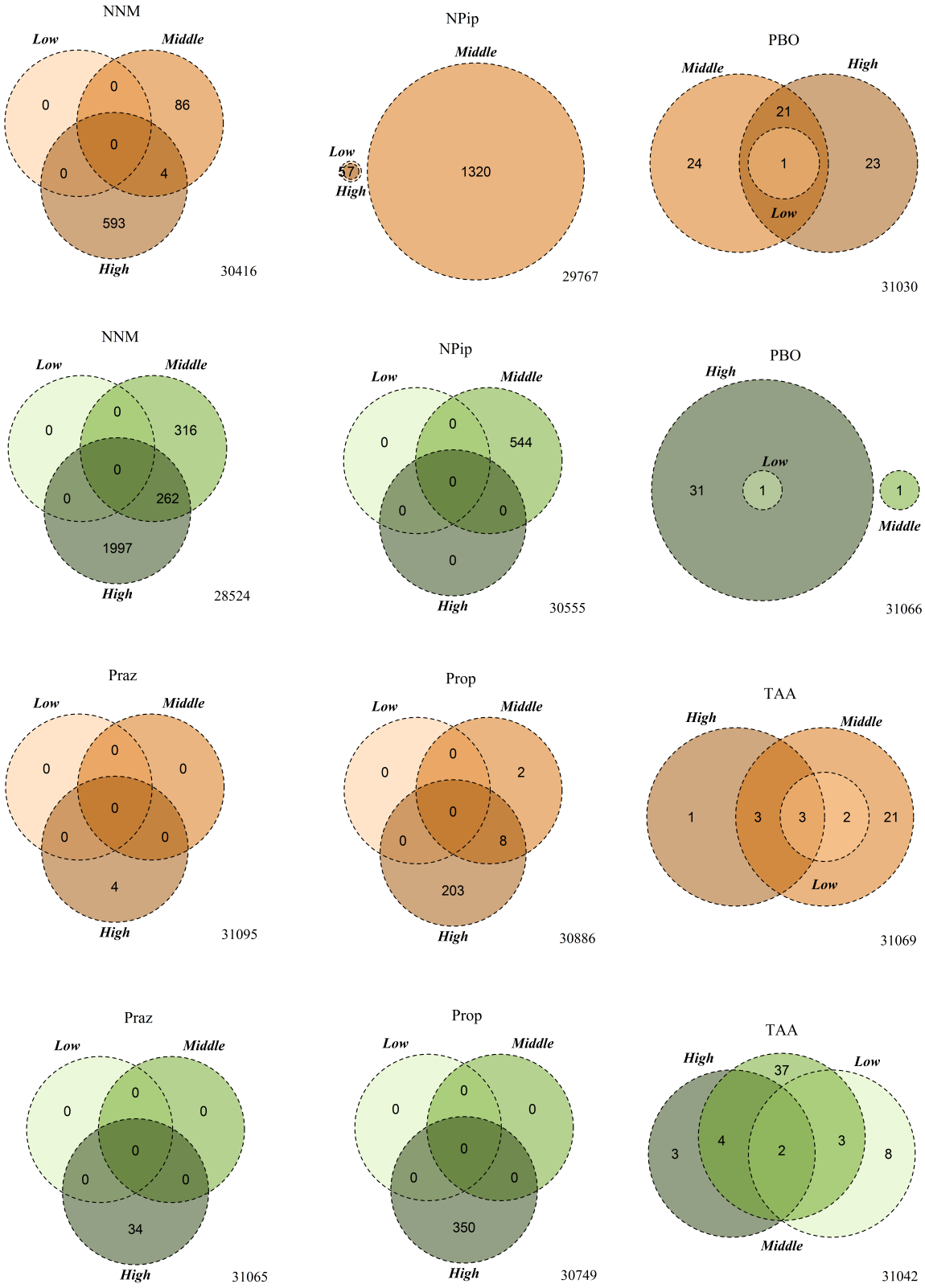
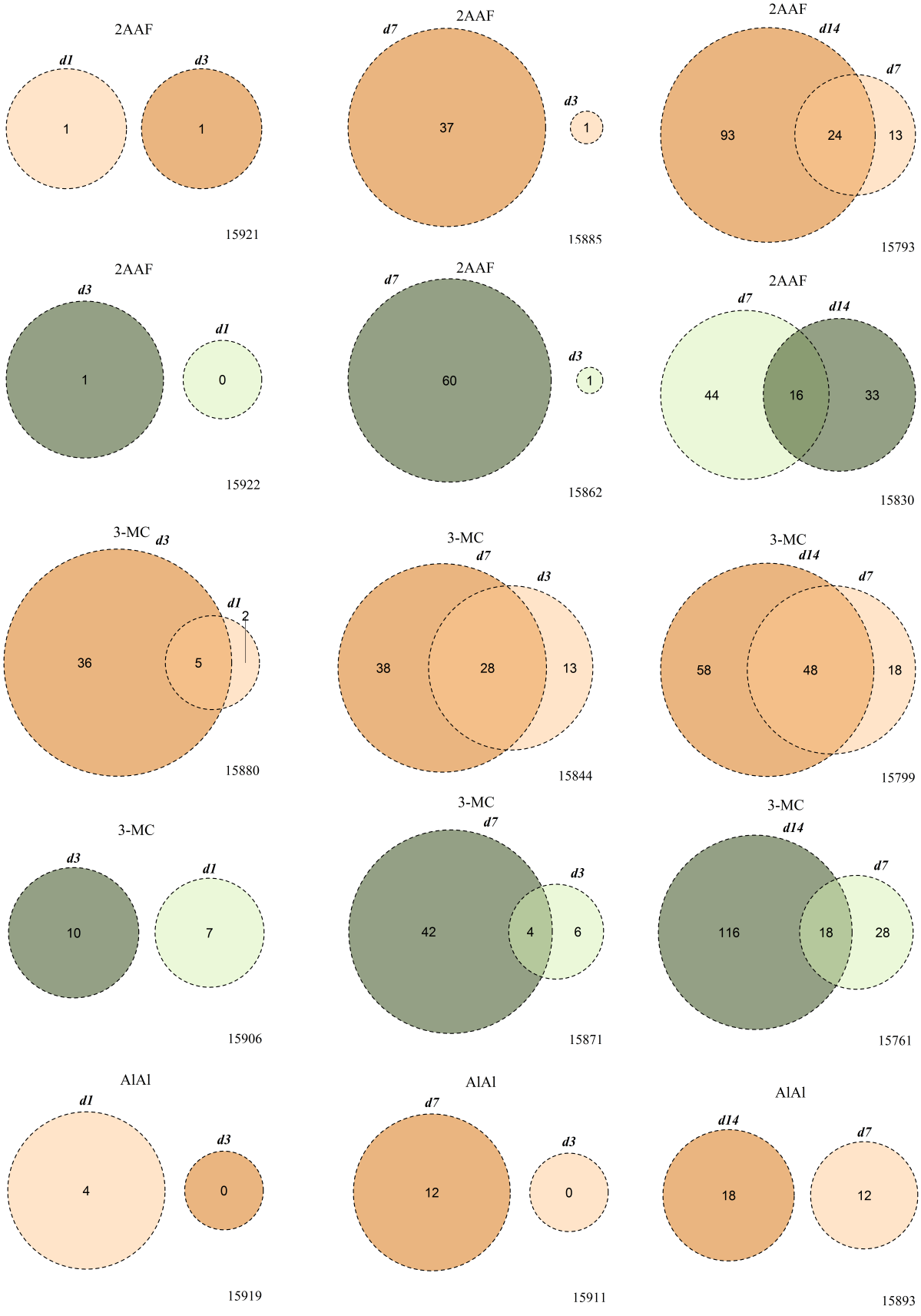
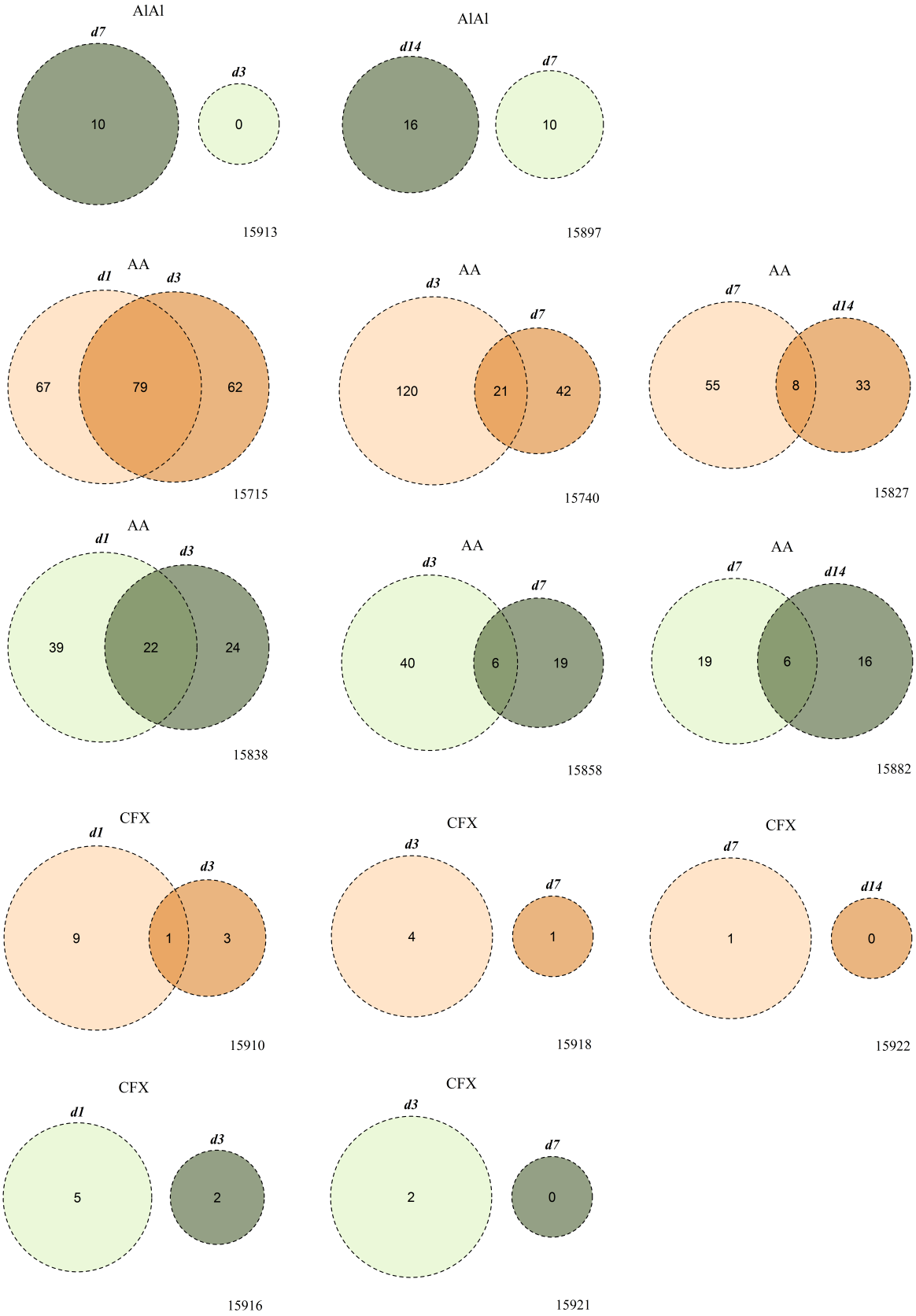
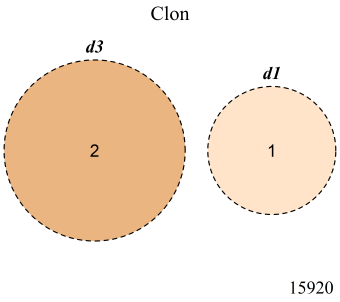
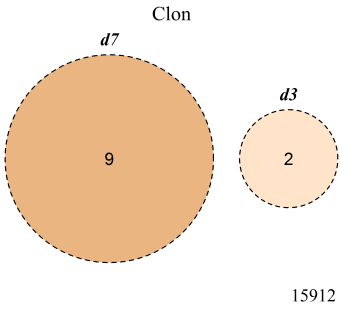


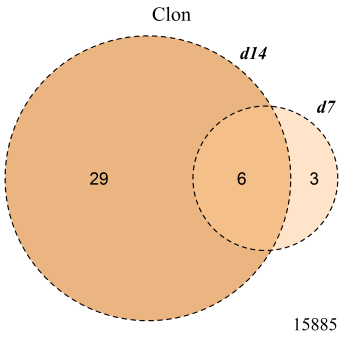
Figure C.24: *Concentration dependency in the NRW in vitro database. Corresponding Venn diagrams to Figure 5.9 summarizing the concentration progressions of all further compounds besides 2-NF, CFX, AlAl, DEHA which are already shown in Figure 5.9. Orange colored Venn diagrams show the overlap between genes that are upregulated at the low, middle and high concentration (> 1.5-fold with adjusted $p \leq 0.01$); green colored Venn diagrams summarize the downregulated genes (< $\frac{3}{2}$-fold with adjusted $p \leq 0.01$).*

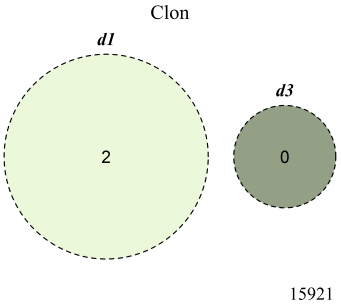

2AAF
d1
d3
1
1
15921
2AAF
d7
d3
37
1
15885
2AAF
d14
d7
93
24
13
15793
2AAF
d3
d1
1
0
15922
2AAF
d7
d3
60
1
15862
2AAF
d7
d14
44
16
33
15830
3-MC
d3
d1
36
5
2
15880
3-MC
d7
d3
38
28
13
15844
3-MC
d14
d7
58
48
18
15799
3-MC
d3
d1
10
7
15906
3-MC
d7
d3
42
4
6
15871
3-MC
d14
d7
116
18
28
15761
AlAl
d1
d3
4
0
15919
AlAl
d7
d3
12
0
15911
AlAl
d14
d7
18
12
15893

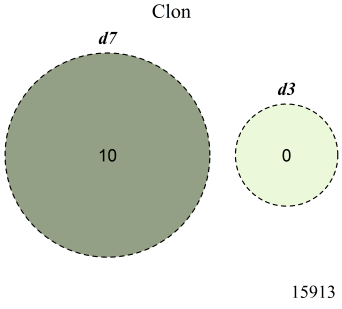

AlAl
d7
d3
10
0
15913
AlAl
d14
d7
16
10
15897
AA
d1
d3
67
79
62
15715
AA
d3
d7
120
21
42
15740
AA
d7
d14
55
8
33
15827
AA
d1
d3
39
22
24
15838
AA
d3
d7
40
6
19
15858
AA
d7
d14
19
6
16
15882
CFX
d1
d3
9
1
3
15910
CFX
d3
d7
4
1
15918
CFX
d7
d14
1
0
15922
CFX
d1
d3
5
2
15916
CFX
d3
d7
2
0
15921


Clon
d3
d1
2
1
15920

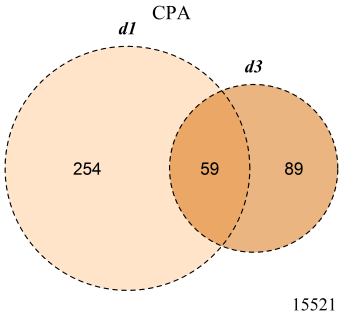

Clon
d7
d3
9
2
15912

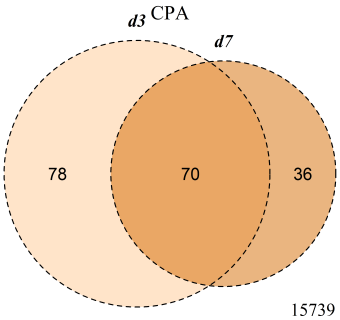

Clon
d14
d7
29
6
3
15885

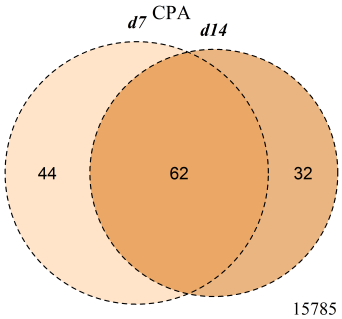

Clon
d1
d3
2
0
15921

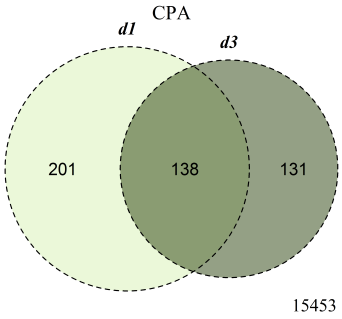

Clon
d7
d3
10
0
15913

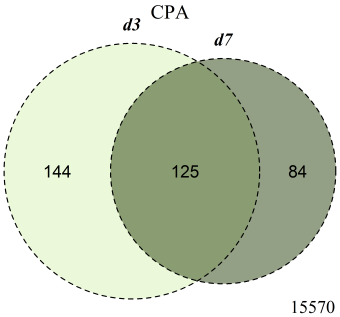

Clon
d14
d7
38
4
6
15875

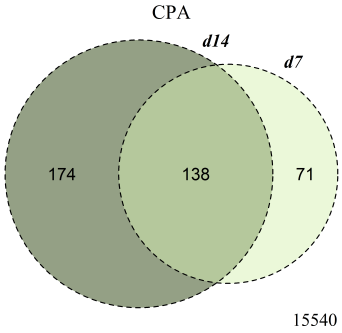

CPA
d1
d3
254
59
89
15521

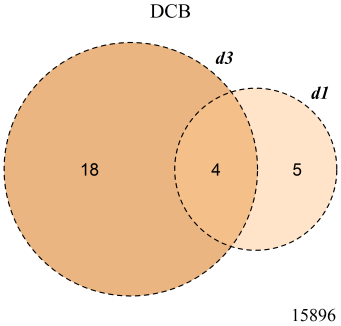

CPA
d3
d7
78
70
36
15739

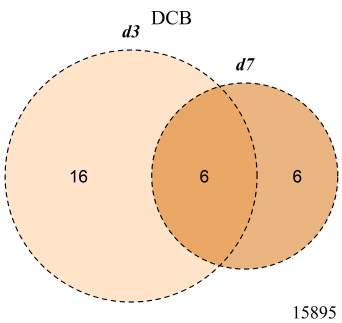

CPA
d7
d14
44
62
32
15785

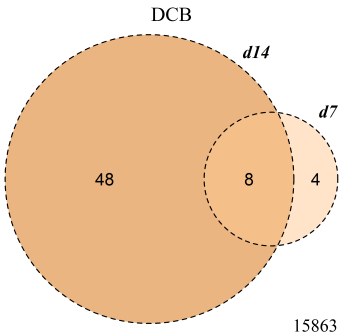

CPA
d1
d3
201
138
131
15453

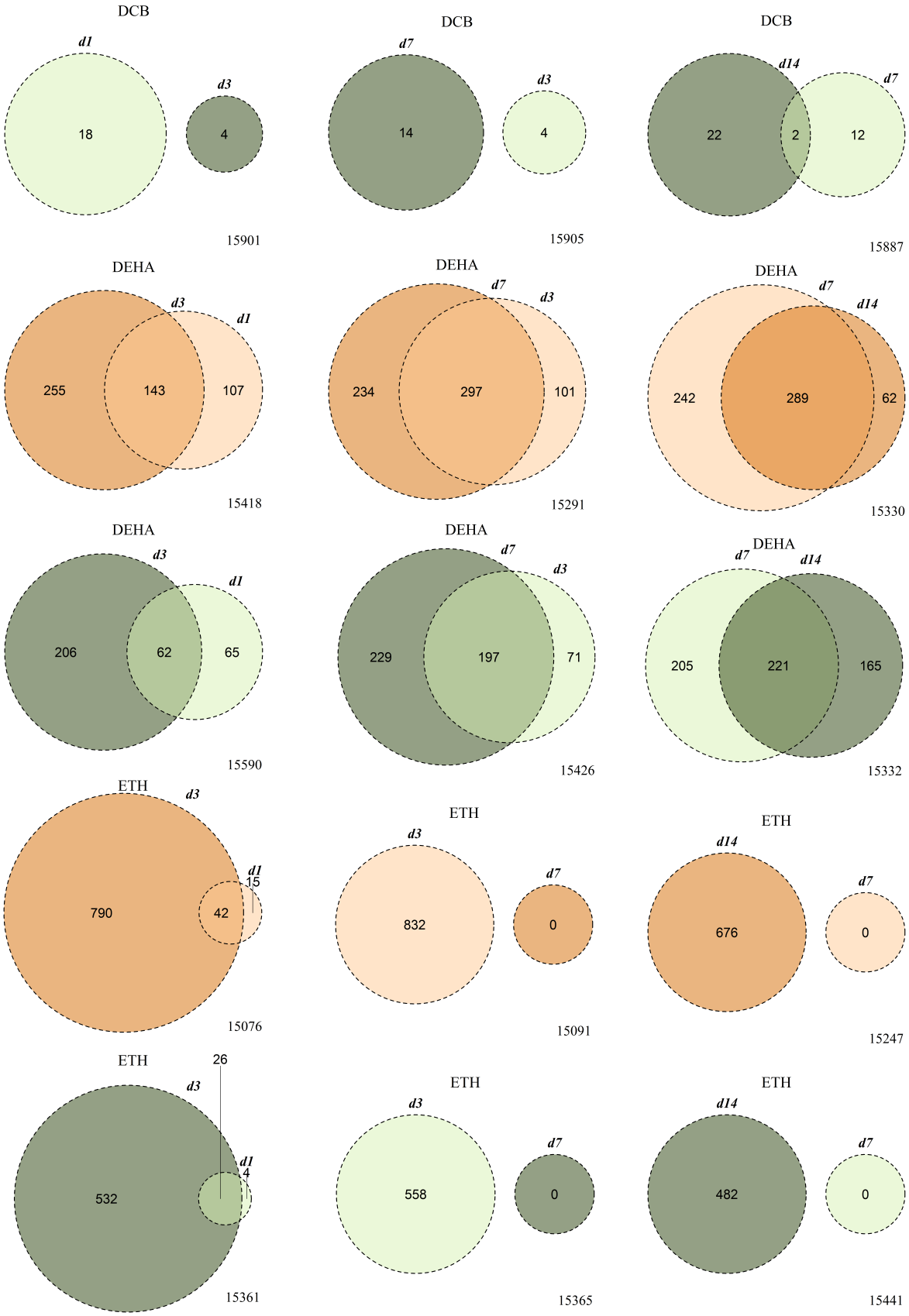

CPA
d3
d7
144
125
84
15570

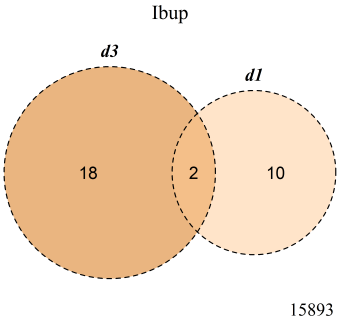

CPA
d14
d7
174
138
71
15540


DCB
d3
d1
18
4
5
15896

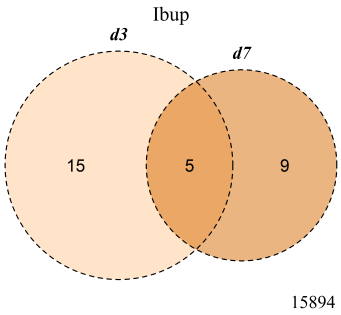

DCB
d3
d7
16
6
6
15895


DCB
d14
d7
48
8
4
15863

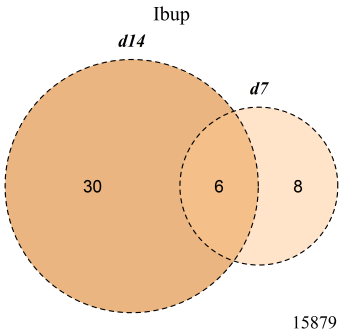

DCB
d1
d3
18
4
15901
DCB
d7
d3
14
4
15905
DCB
d14
d7
22
2
12
15887
DEHA
d3
d1
255
143
107
15418
DEHA
d7
d3
234
297
101
15291
DEHA
d7
d14
242
289
62
15330
DEHA
d3
d1
206
62
65
15590
DEHA
d7
d3
229
197
71
15426
DEHA
d7
d14
205
221
165
15332
ETH
d3
d1
790
42
15
15076
ETH
d3
d7
832
0
15091
ETH
d14
d7
676
0
15247
ETH
26
d3
d1
532
4
15361
ETH
d3
d7
558
0
15365
ETH
d14
d7
482
0
15441



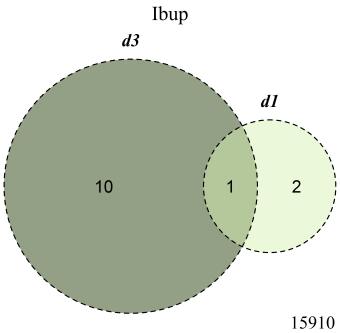
15893



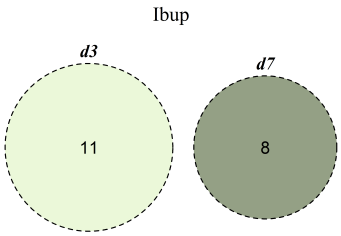
15894



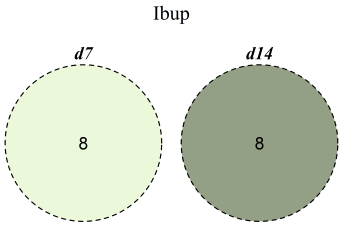
15879



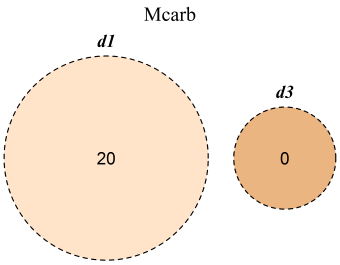
15910



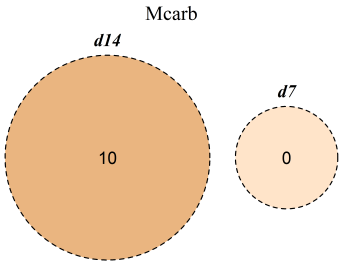
15904



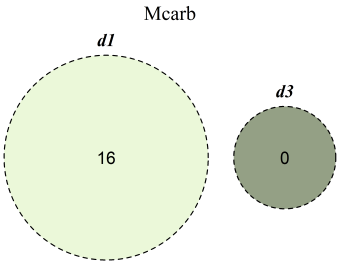
15907



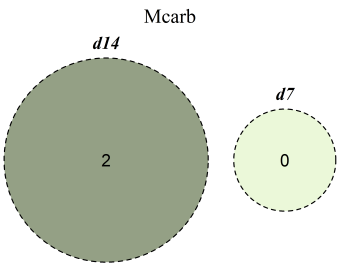
15903



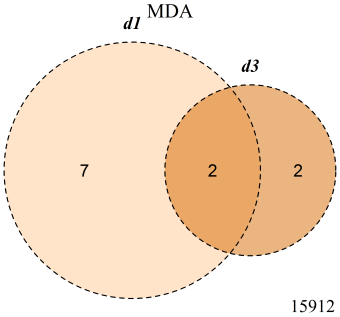
15913



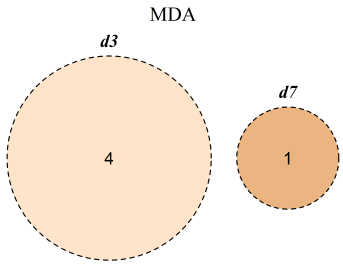
15907



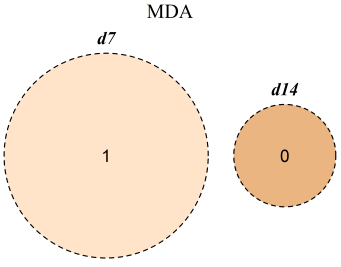
15921



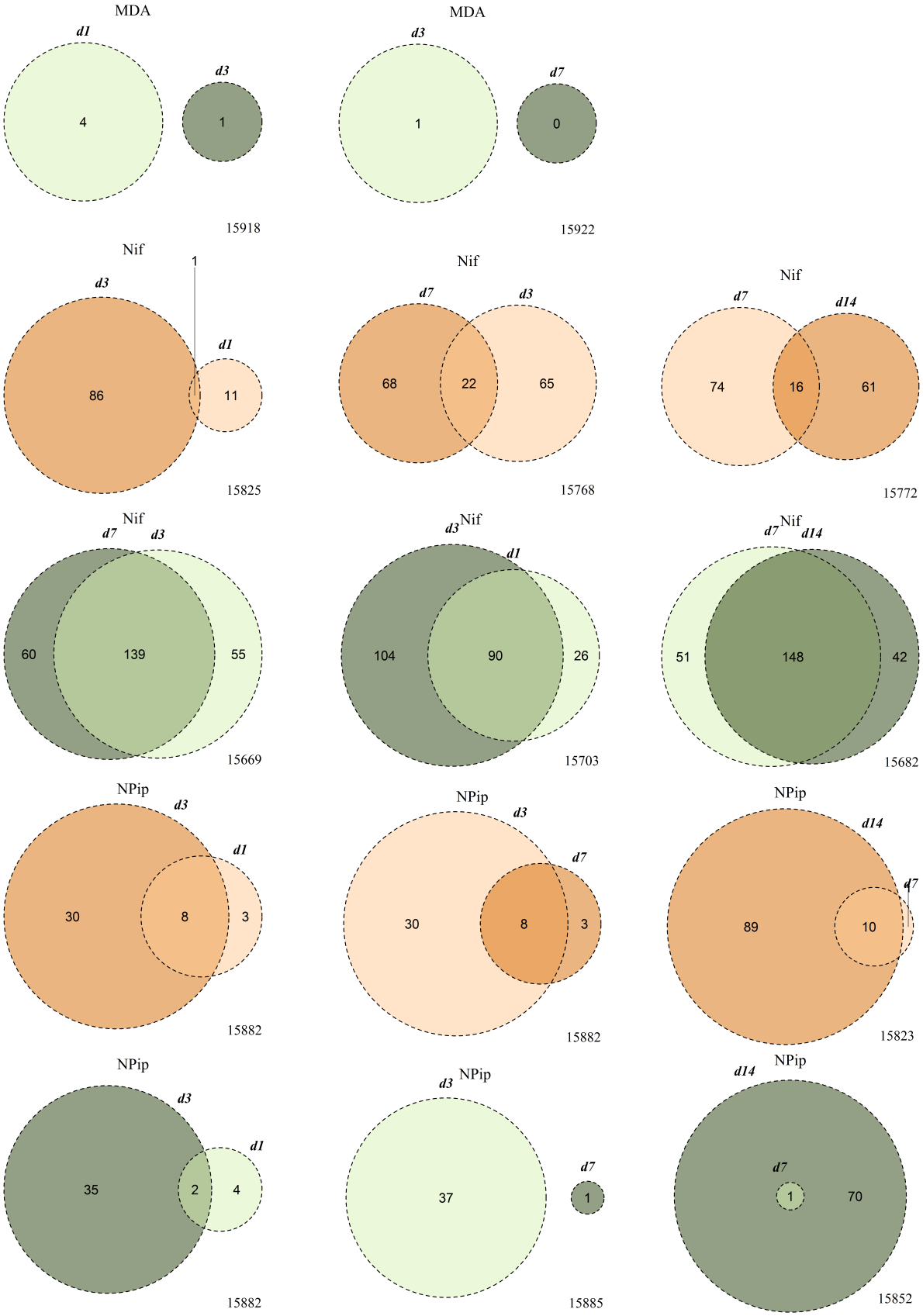
15912

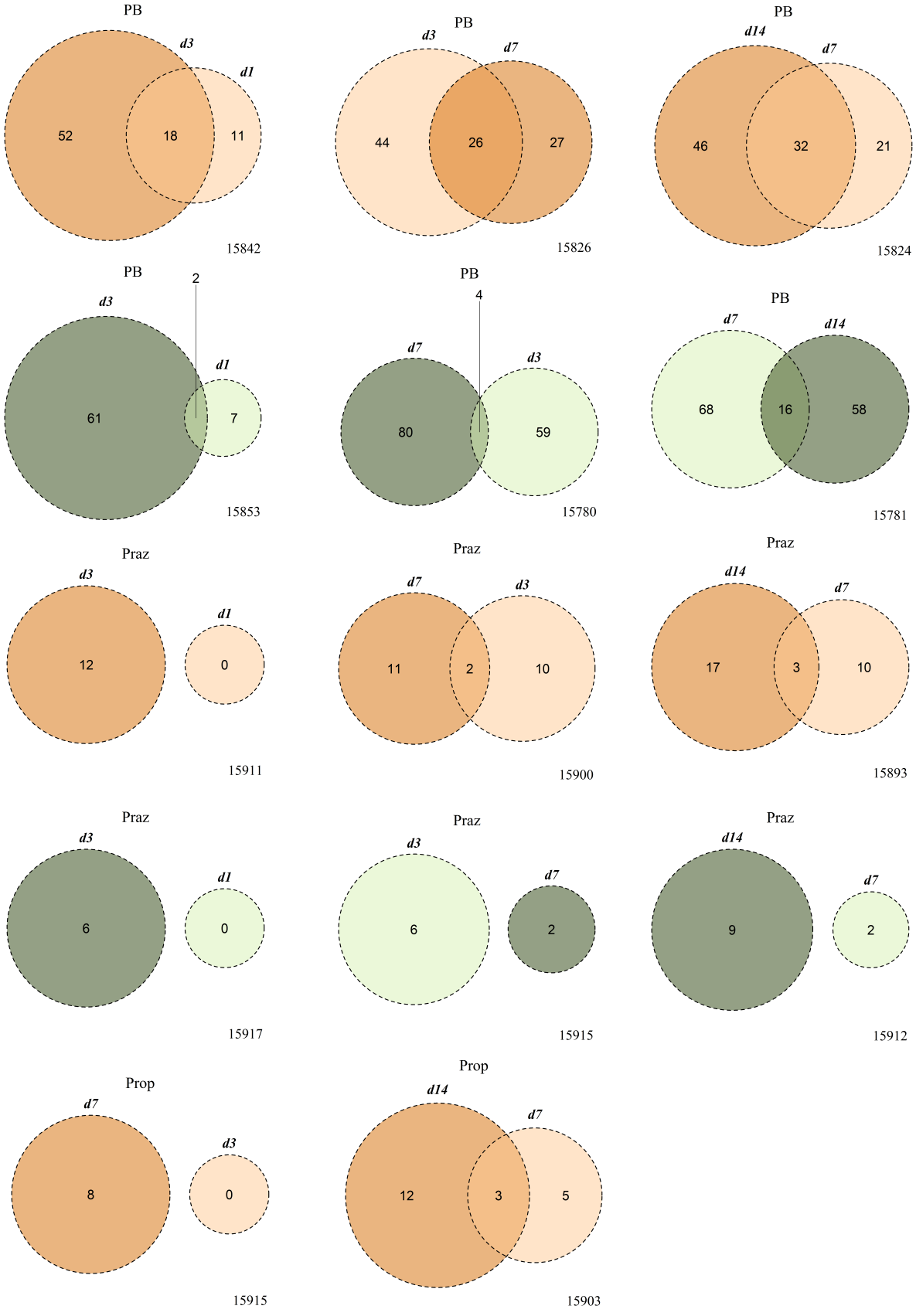


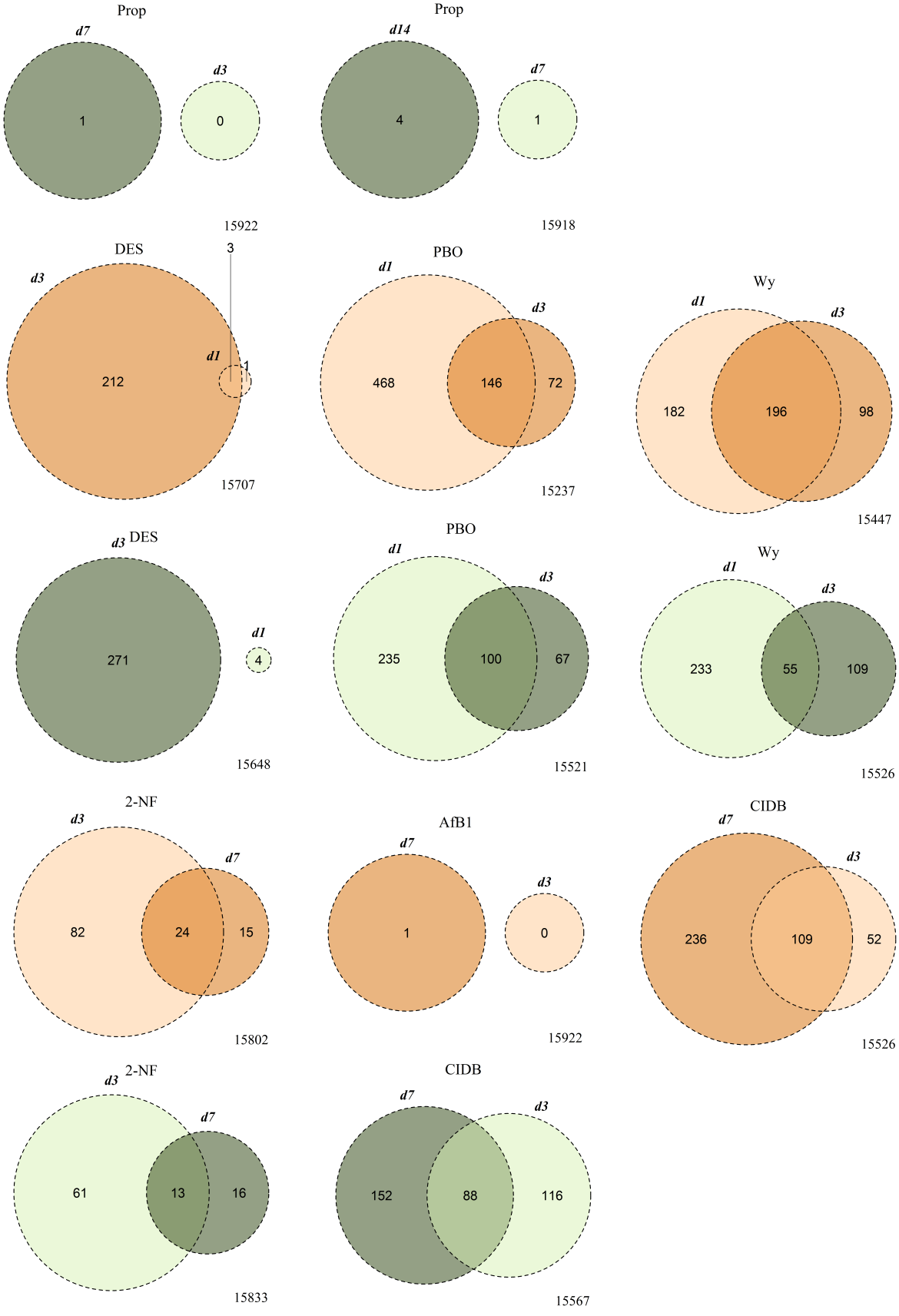
15918



15922


MDA
d1
d3
4
1
15918
MDA
d3
d7
1
0
15922
Nif
1
d3
d1
86
11
15825
Nif
d7
d3
68
22
65
15768
Nif
d7
d14
74
16
61
15772
Nif
d7
d3
60
139
55
15669
Nif
d3
d1
104
90
26
15703
Nif
d7
d14
51
148
42
15682
NPip
d3
d1
30
8
3
15882
NPip
d3
d7
30
8
3
15882
NPip
d14
d7
89
10
15823
NPip
d3
d1
35
2
4
15882
NPip
d3
d7
37
1
15885
NPip
d14
d7
1
70
15852


PB
d3
d1
52
18
11
15842
PB
d3
d7
44
26
27
15826
PB
d14
d7
46
32
21
15824
PB
2
d3
d1
61
7
15853
PB
4
d7
d3
80
59
15780
PB
d7
d14
68
16
58
15781
Praz
d3
d1
12
0
15911
Praz
d7
d3
11
2
10
15900
Praz
d14
d7
17
3
10
15893
Praz
d3
d1
6
0
15917
Praz
d3
d7
6
2
15915
Praz
d14
d7
9
2
15912
Prop
d7
d3
8
0
15915
Prop
d14
d7
12
3
5
15903


Prop
d7
d3
1
0
15922
Prop
d14
d7
4
1
15918
DES
d3
d1
212
3
1
15707
PBO
d1
d3
468
146
72
15237
Wy
d1
d3
182
196
98
15447
d3 DES
d1
271
4
15648
PBO
d1
d3
235
100
67
15521
Wy
d1
d3
233
55
109
15526
2-NF
d3
d7
82
24
15
15802
AfB1
d7
d3
1
0
15922
CIDB
d7
d3
236
109
52
15526
d3 2-NF
d7
61
13
16
15833
CIDB
d7
d3
152
88
116
15567

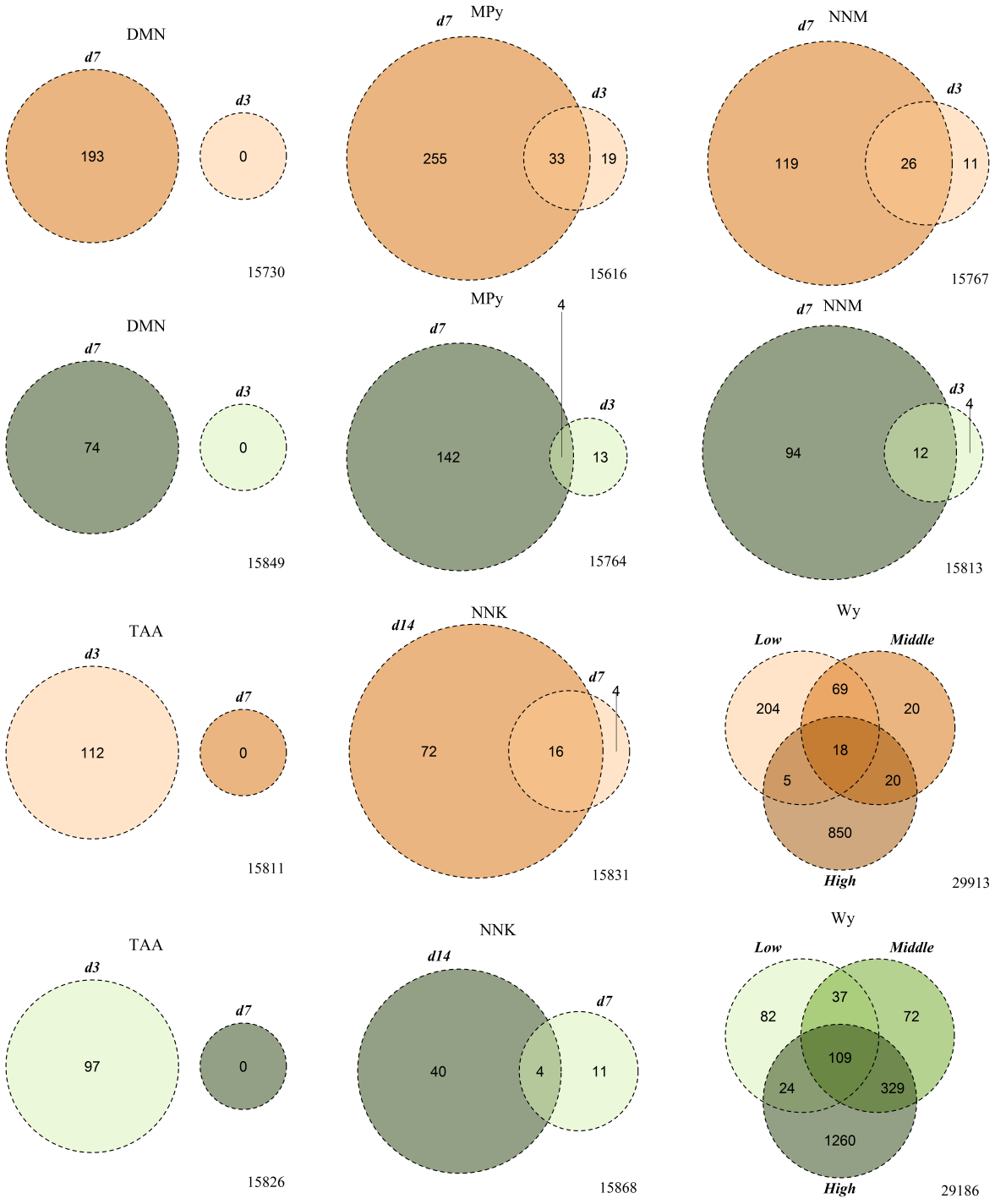


Figure C.25: *Concentration dependency in the NRW in vivo database. Venn diagrams summarizing the concentration progressions of all compounds tested in the NRW in vivo data set. Orange colored Venn diagrams show the overlap between genes that are upregulated at adjacent time periods (> 1.5-fold with adjusted $p \leq 0.01$); green colored Venn diagrams summarize the downregulated genes (< $\frac{3}{2}$-fold with adjusted $p \leq 0.01$). Due to the lack of space in Figure C.24 the concentration progression of the compound Wy tested in vitro at the low, middle and high concentration is shown at the end of this figure.*

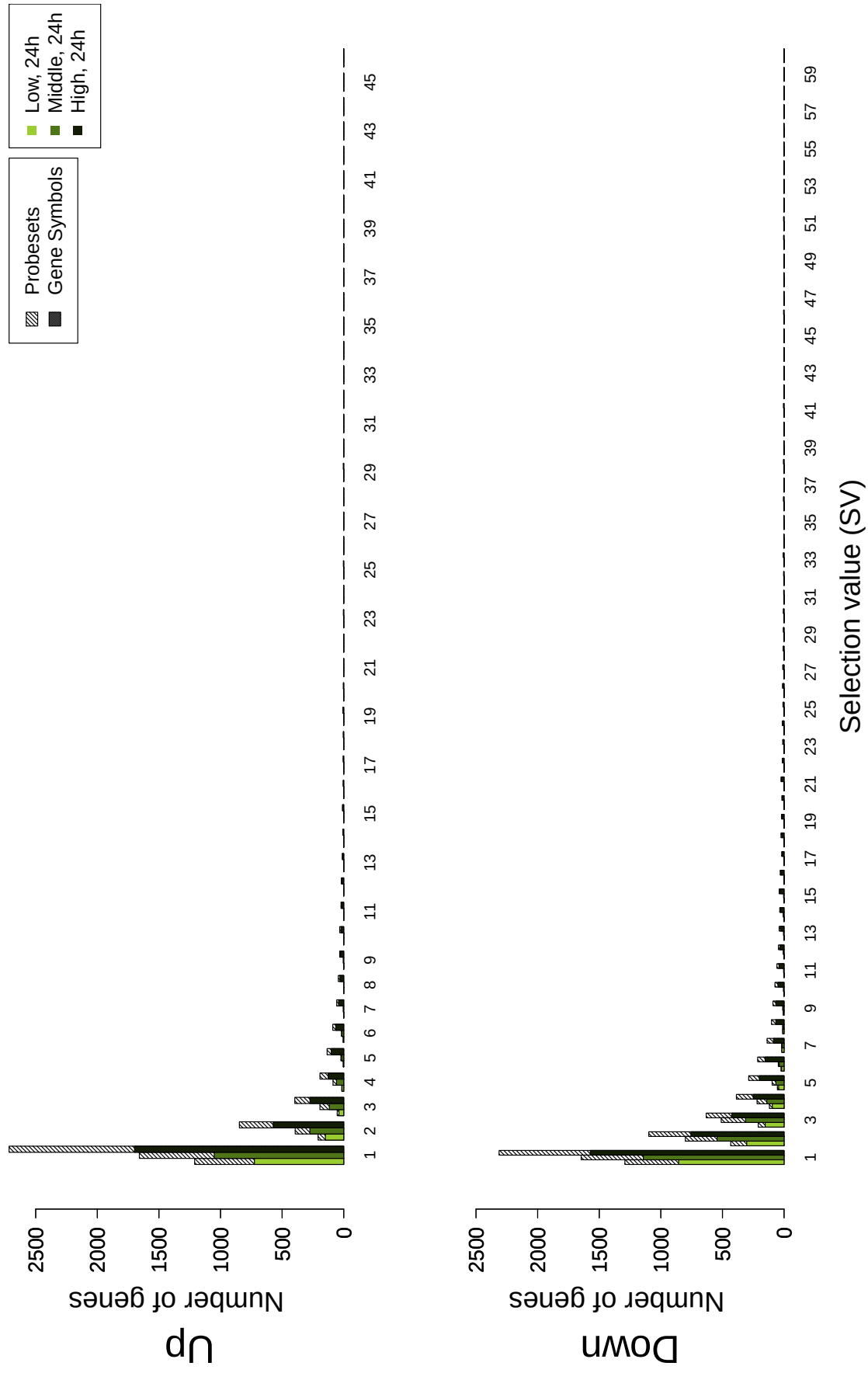


Figure C.26: *TGD (in vitro): Selection values for the up- and downregulated genes (upper and lower panel). A selection value of e.g. three means that at least three compounds up- or downregulate (> |3|-fold) the indicated gene.*

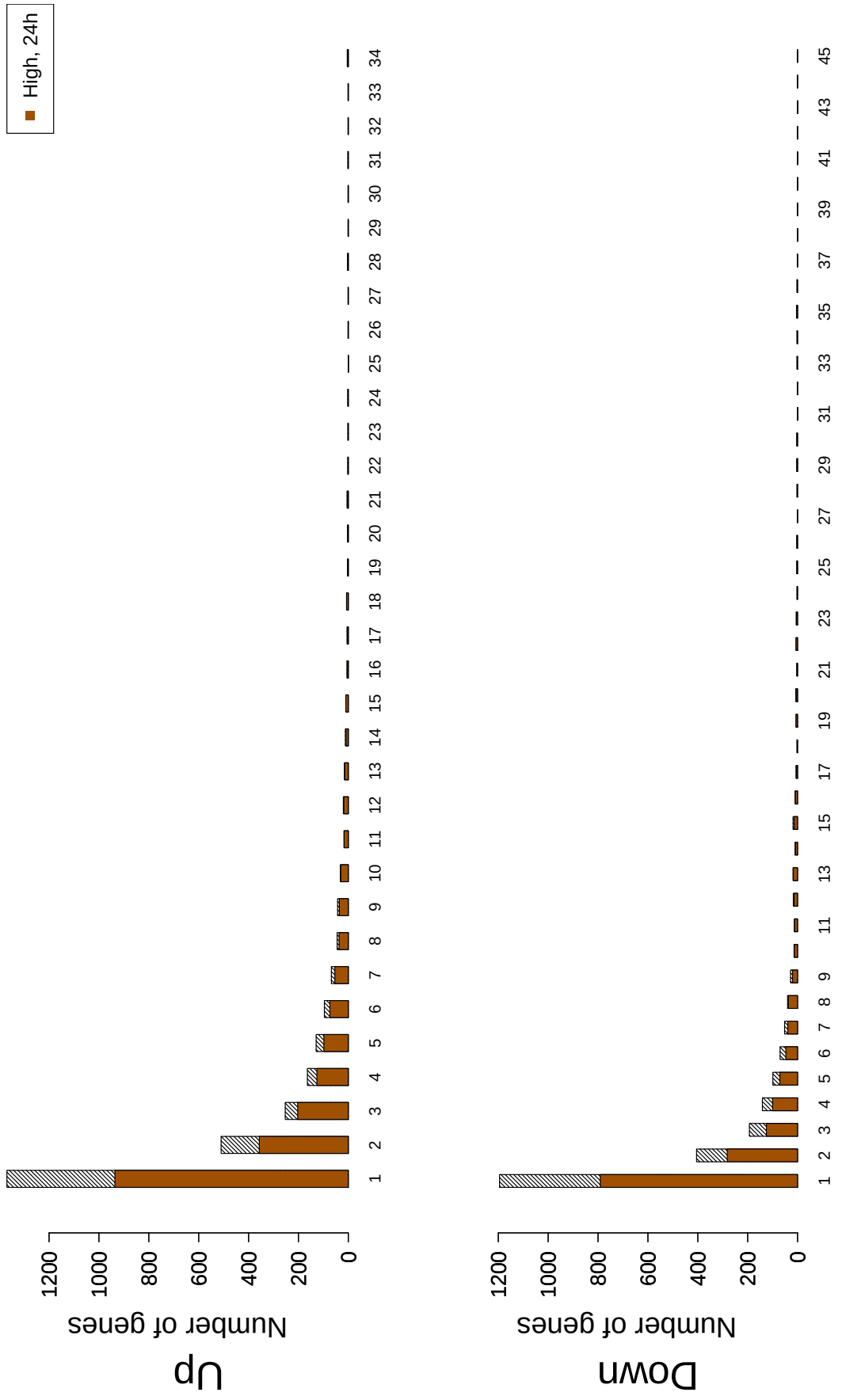


Figure C.27: *TGD (in vivo): Selection values for the up- and downregulated genes (upper and lower panel). A selection value of e.g. three means that at least three compounds up- or downregulate (> |3|-fold with adjusted $p \leq 0.01$) the indicated gene.*

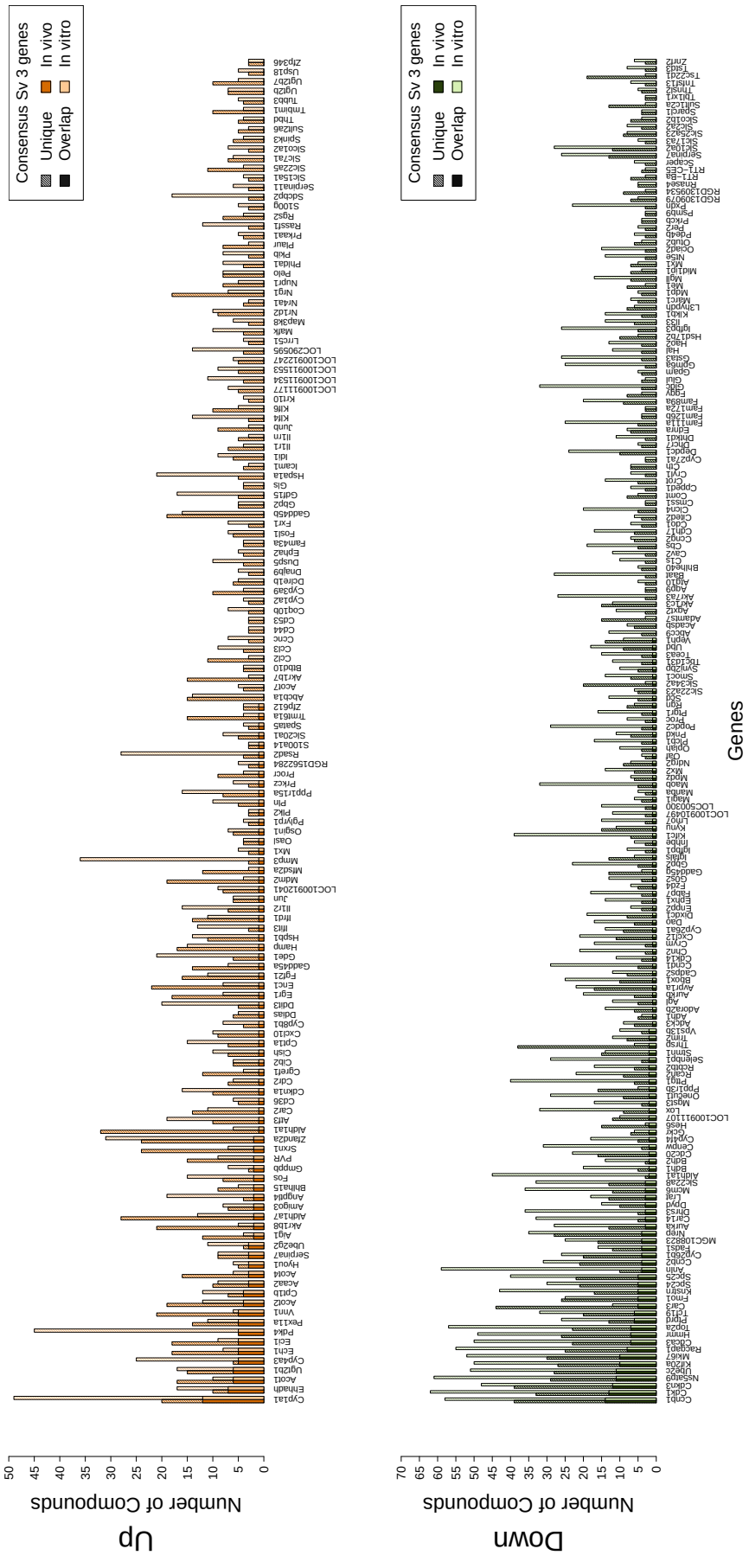


Figure C.28: *TGD: Overall selection values of the consensus Sv 3 TGD genes (genes that are deregulated by at least three compounds in both test systems, in vitro and in vivo, for at least one test condition). Compounds are counted once. The upper panel shows the barplots for the 140 upregulated consensus genes and the lower panel shows the barplots for the 186 downregulated consensus genes. The shaded barplots indicate the number of individual in vitro and in vivo compounds and the fully colored barplots indicate the number of common compounds.*

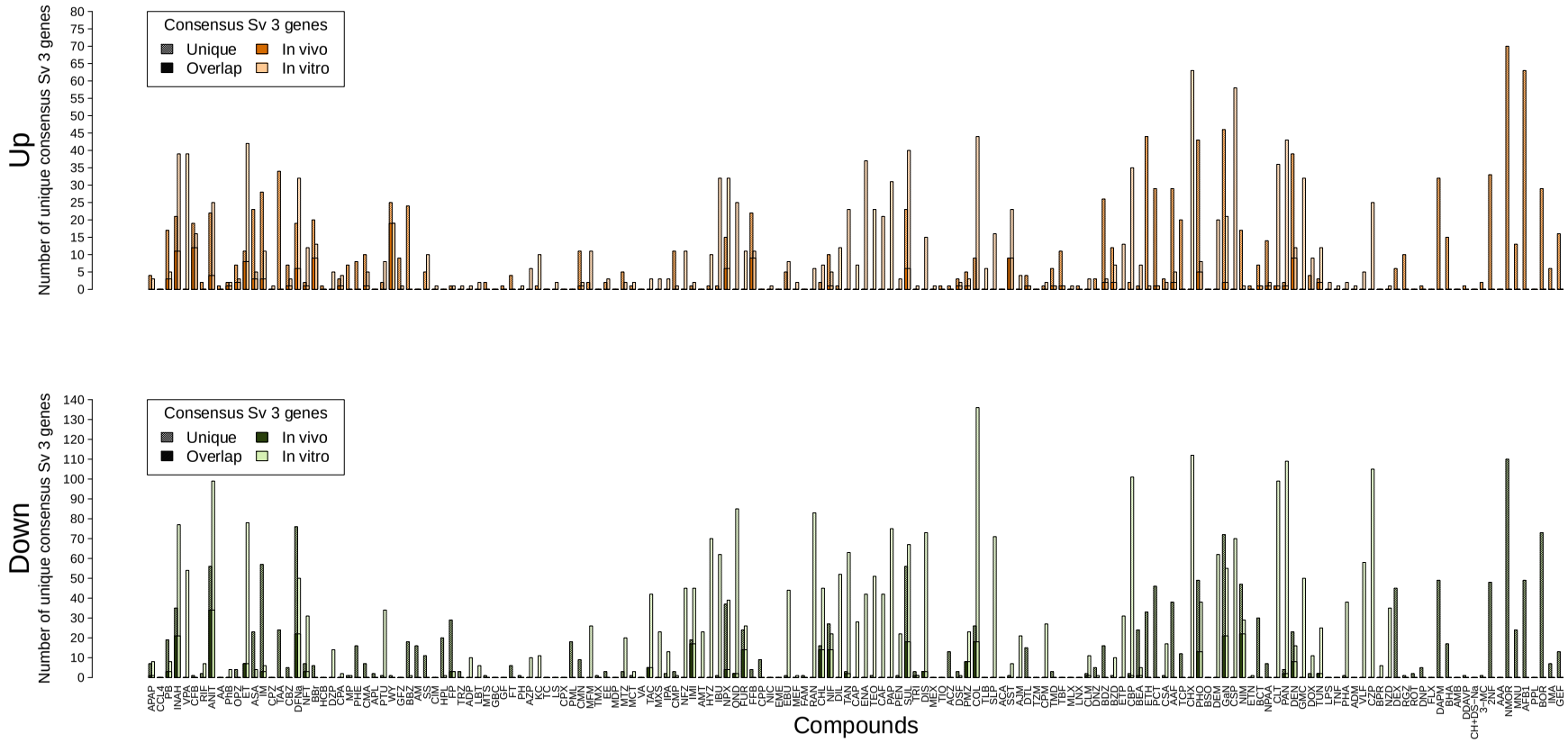


Figure C.29: *TGD: Number of consensus Sv 3 TGD genes that are deregulated by the indicated compounds in in vitro and in in vivo for at least one test condition (shaded barplots). The fully colored barplots indicate the number of common consensus genes, i.e. the Sv 3 genes that are deregulated in both test systems, the in vitro and in vivo test system. The upper panel shows the barplots for the upregulated consensus genes and the lower panel shows the barplots for the downregulated consensus genes.*

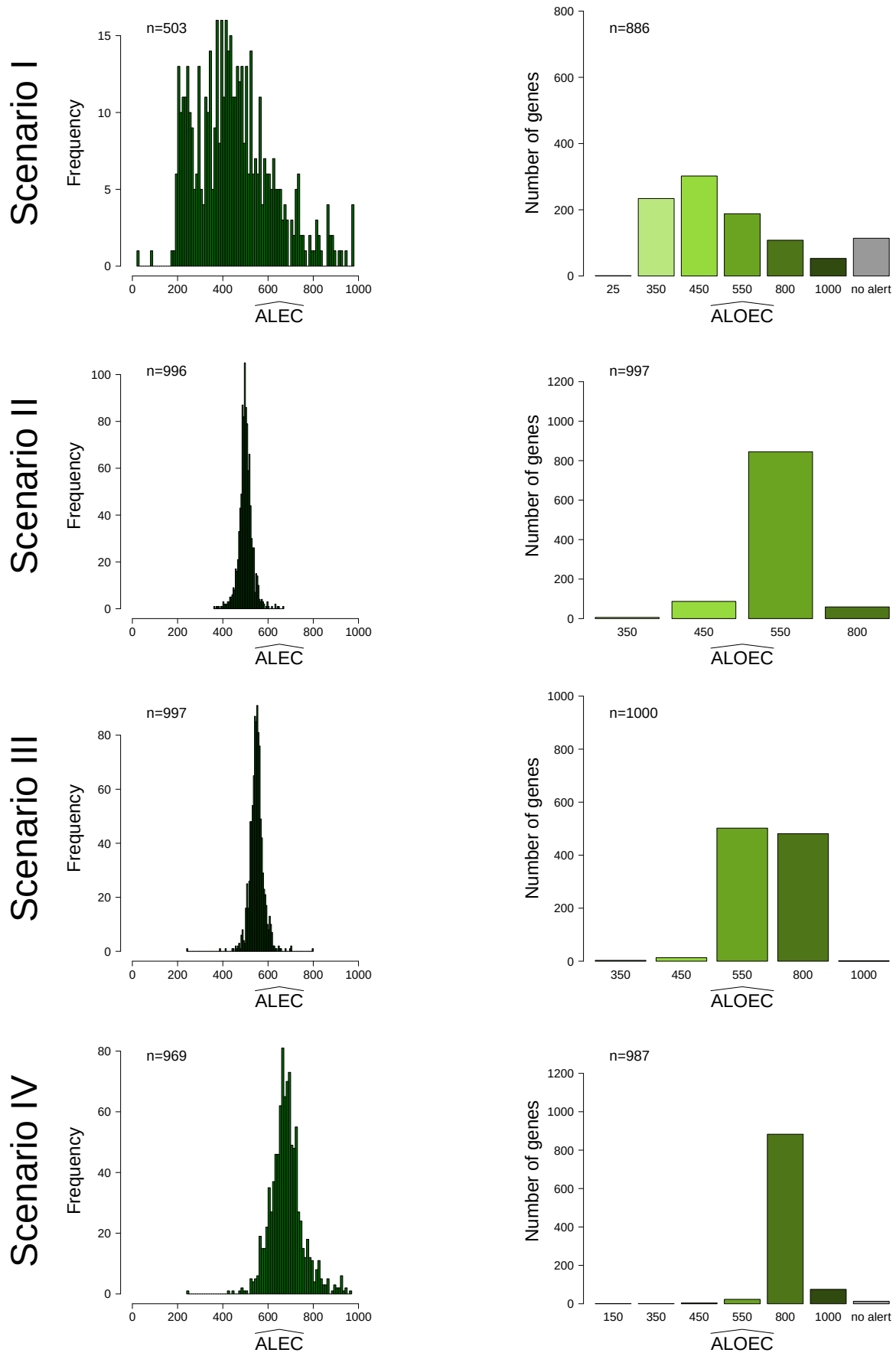


Figure C.30: *Distributions of the estimated alert concentrations for Scenarios I-IV with* $k = 6$ *replicates. Rows indicate the scenario and columns the methods of estimation. The left panel shows the distributions of the* $\widehat{ALEC}$*s (4pLL) and the right panel the distribution of the* $\widehat{ALOEC}$*s (Limma). The number of estimates* $\leq$ *1000 is indicated by* n*. Grey colored bars indicate the number of no alerts.*

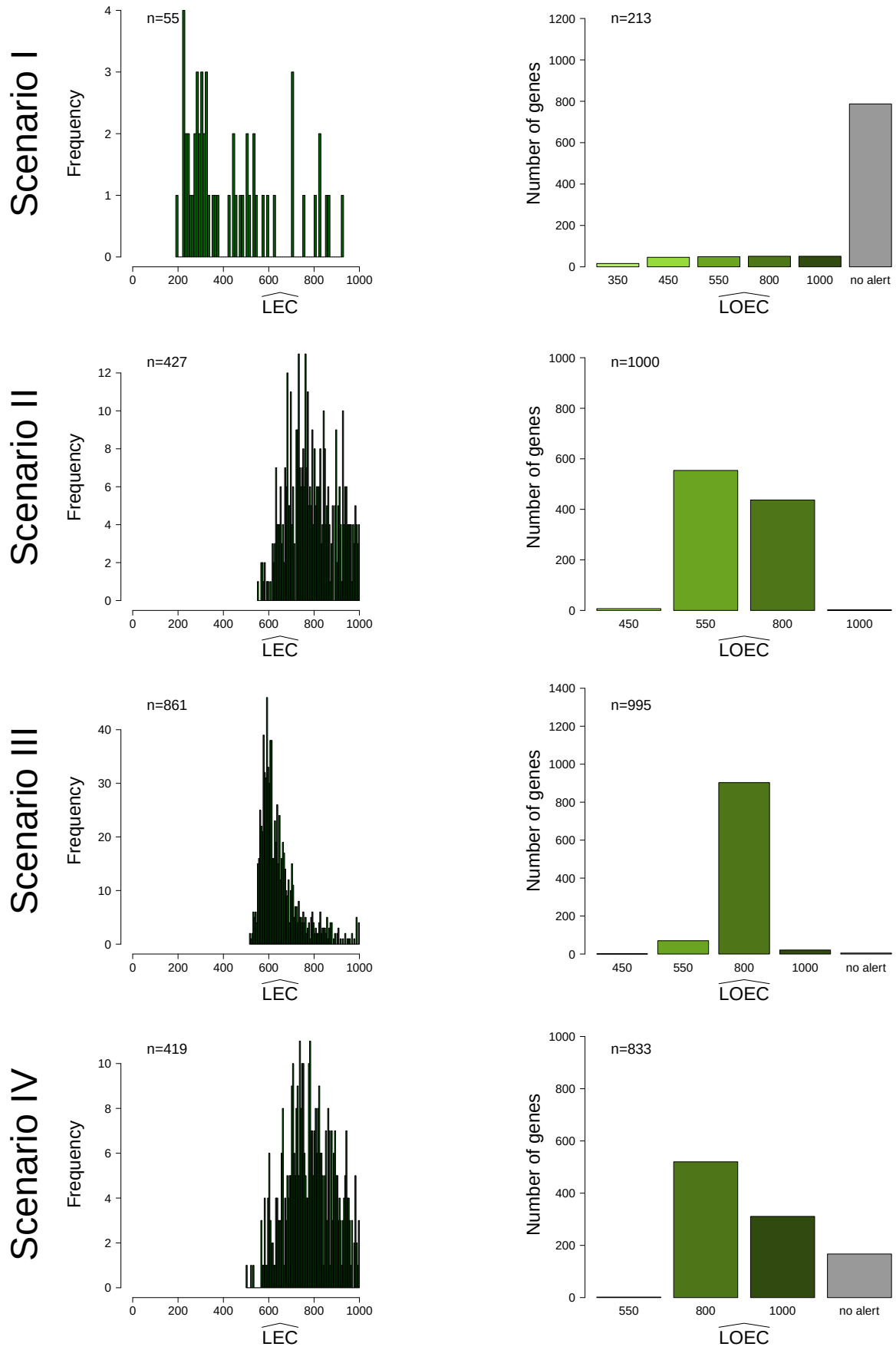


Figure C.31: *Distributions of the estimated alert concentrations for Scenarios I-IV with $k = 6$. Rows indicate the scenario and columns the methods of estimation. The left panel shows the distributions of the $\widehat{LEC}$s (4pLL) and the right panel the distribution of the $\widehat{LOEC}$s (Limma). The number of estimates $\leq$ 1000 µM is indicated by n. Grey colored bars indicate the number of no alerts.*

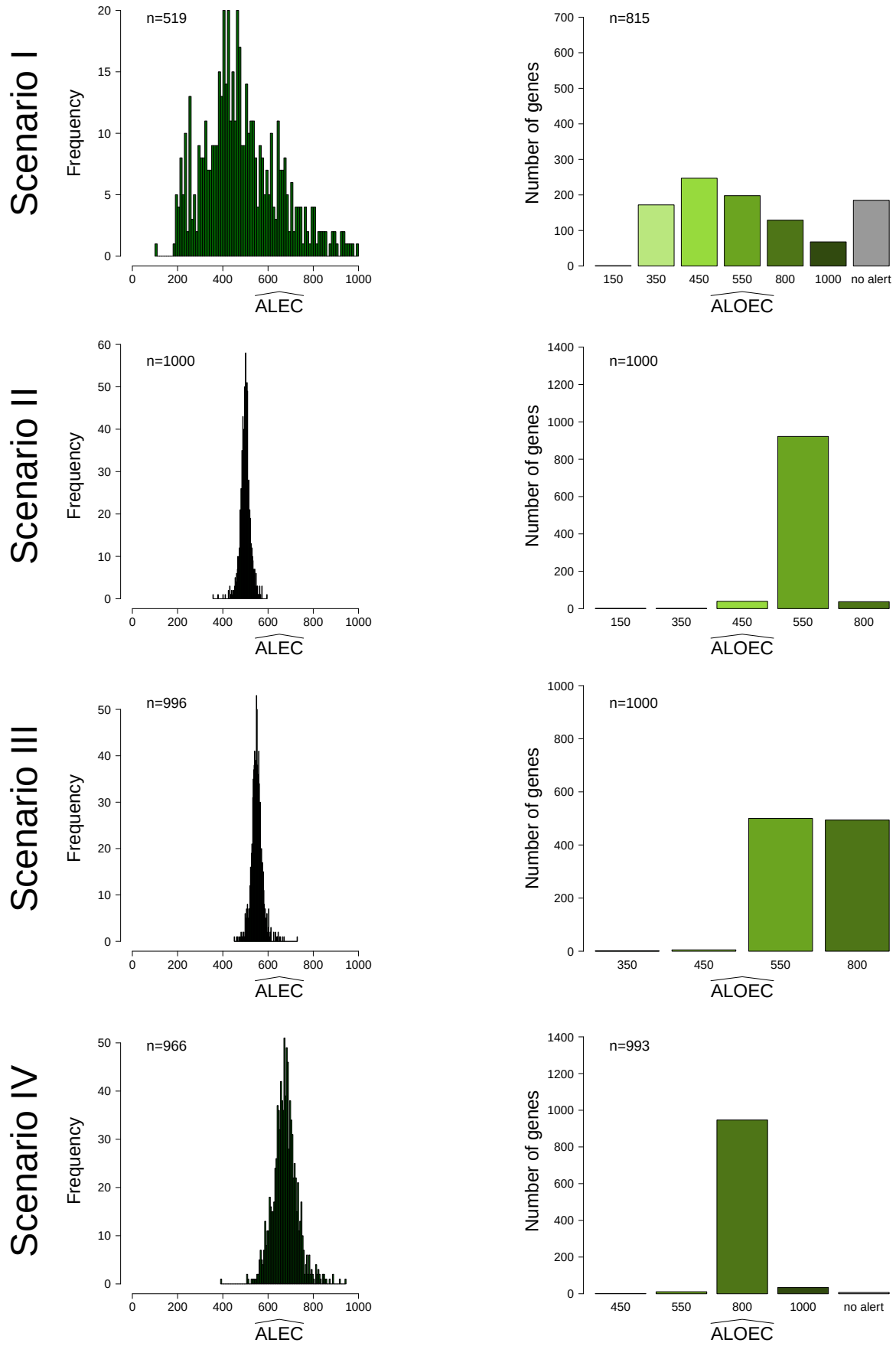


Figure C.32: *Distributions of the estimated alert concentrations for Scenarios I-IV with* $k = 10$ *replicates. Rows indicate the scenario and columns the methods of estimation. The left panel shows the distributions of the* $\widehat{ALEC}$*s (4pLL) and the right panel the distribution of the* $\widehat{ALOEC}$*s (Limma). The number of estimates* $\leq$ *1000 µM is indicated by* n*. Grey colored bars indicate the number of no alerts.*

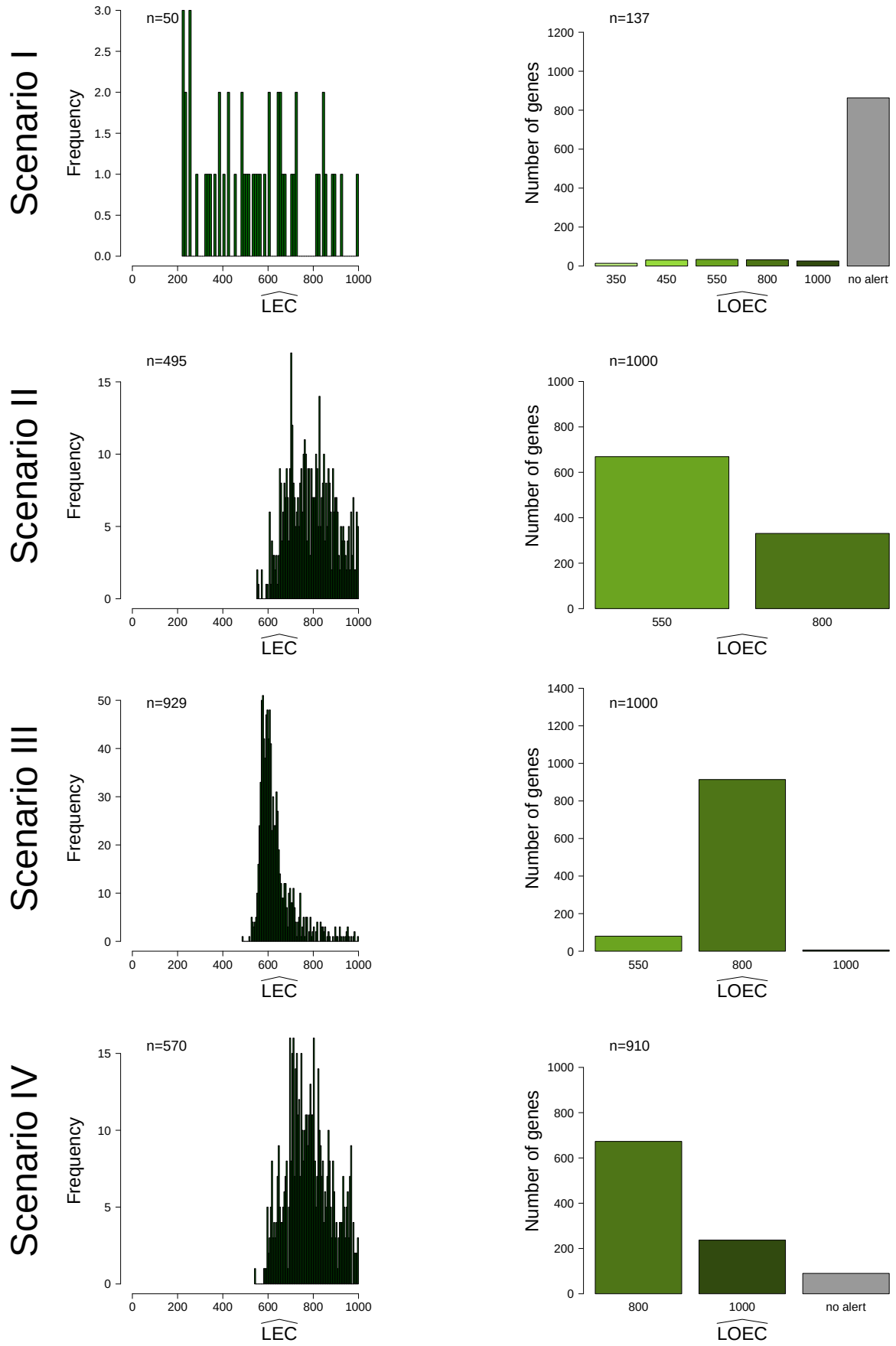


Figure C.33: *Distributions of the estimated alert concentrations for Scenarios I-IV with* $k = 10$ *replicates. Rows indicate the scenario and columns the methods of estimation. The left panel shows the distributions of the* $\widehat{LECs}$ *(4pLL) and the right panel the distribution of the* $\widehat{LOECs}$ *(Limma). The number of estimates* $\leq$ *1000 µM is indicated by* n*. Grey colored bars indicate the number of no alerts.*

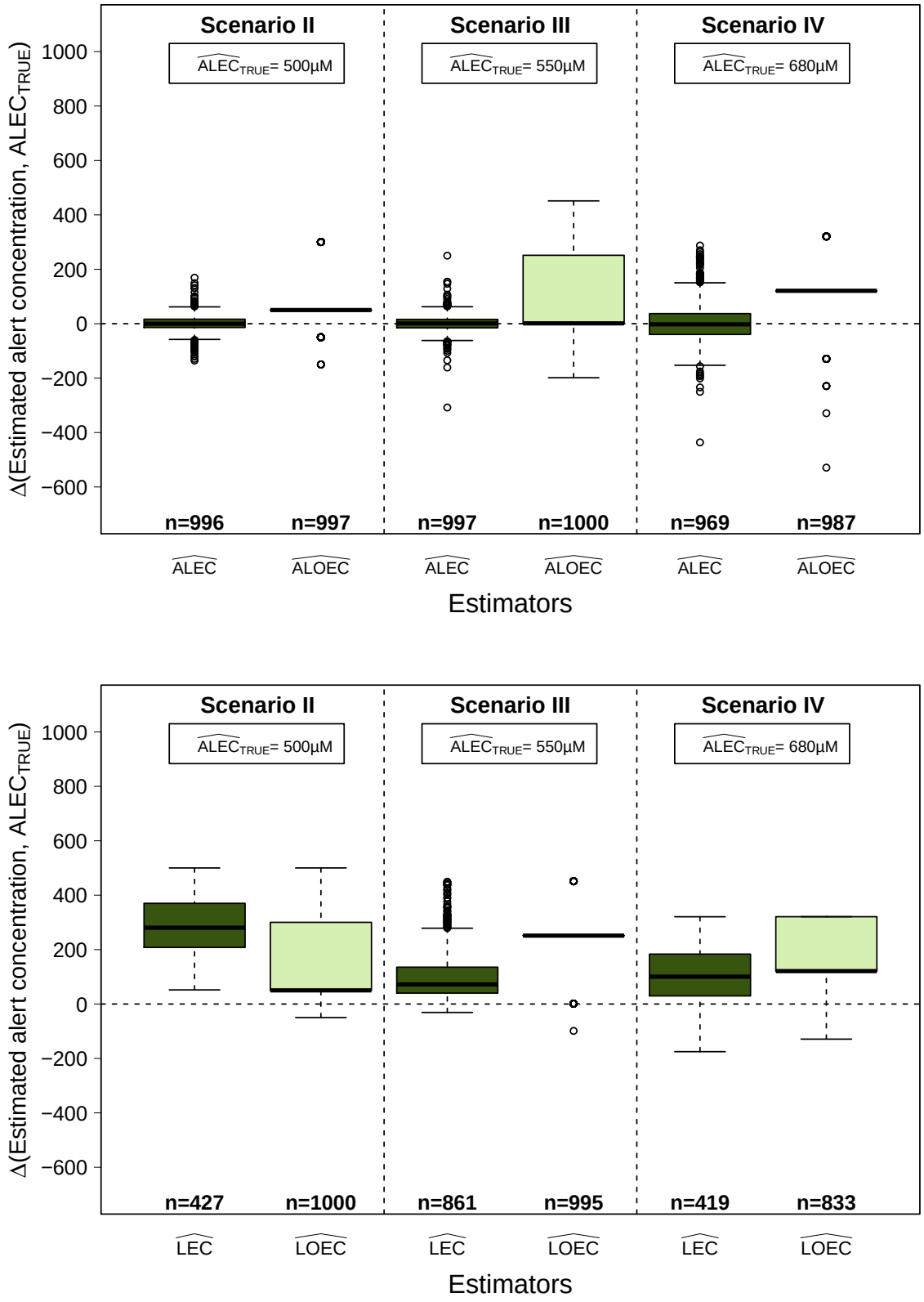


Figure C.34: *Boxplots illustrating the distributions of the differences between the estimated alert concentrations and the respective true ALECs of the Scenarios II-IV (the difference is indicated by Δ). Scenario I was excluded from the analysis since no deviations could be computed (no ALEC value was provided). The upper panel shows the deviations of the point estimates from the true ALECs and the lower panel shows the deviations of the CI-based estimates ($p \leq 0.05$). The alert concentrations were estimated from the simulated data with $k = 6$ replicates per concentration under the indicated scenario.*

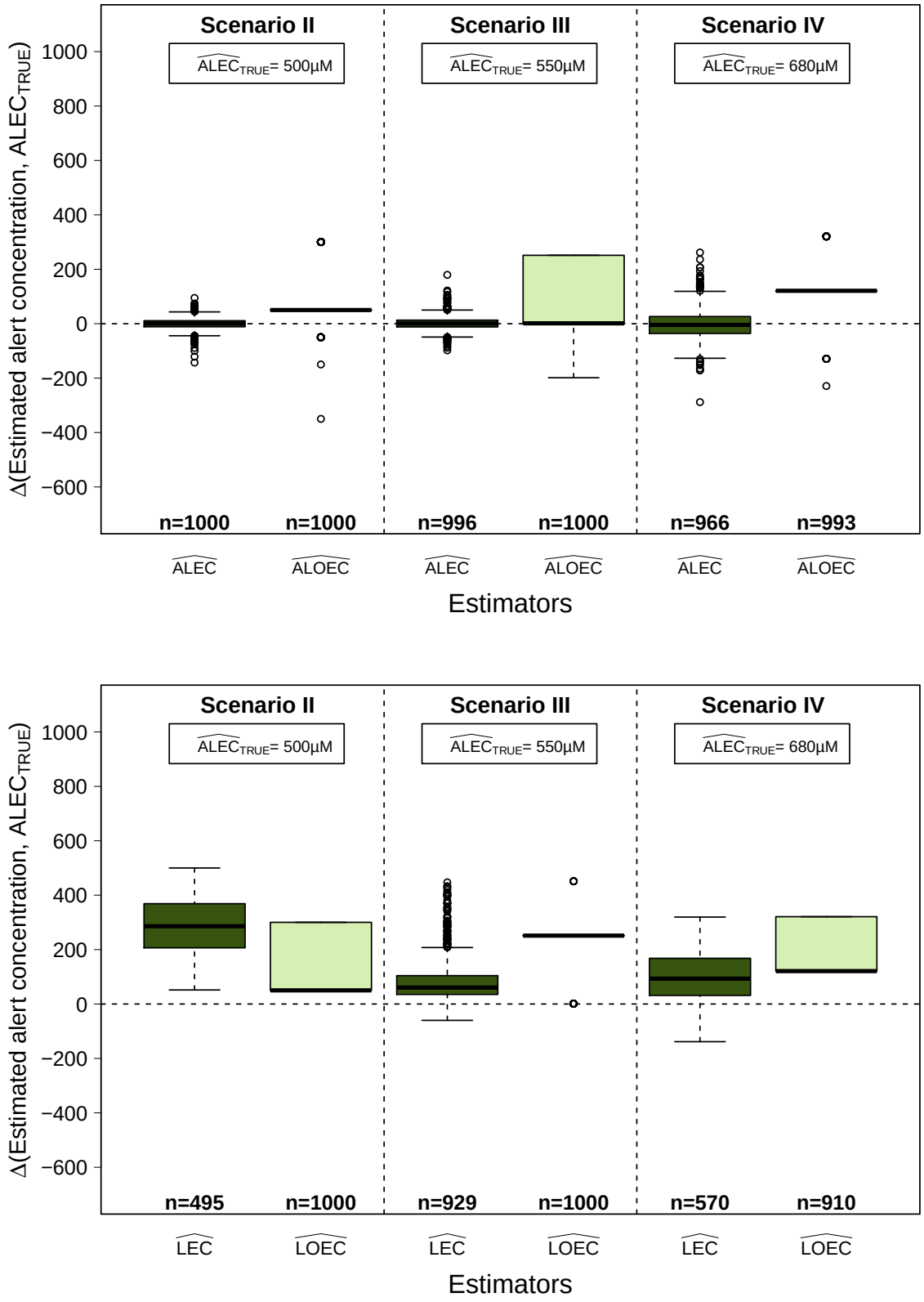


Figure C.35: *Boxplots illustrating the distributions of the differences between the estimated alert concentrations and the respective true ALECs of the Scenarios II-IV (difference is indicated by Δ). Scenario I was excluded from the analysis since no deviations could be computed (no ALEC value was provided). The upper panel shows the deviations of the point estimates from the true ALECs and the lower panel the deviations of the CI-based estimates ($p \leq 0.05$). The alert concentrations were estimated from the simulated data with $k = 10$ replicates per concentration under the indicated scenario.*