

RESEARCH ARTICLE

The sensitivity of transcriptomics BMD modeling to the methods used for microarray data normalization

Roman Mezencev^{1*}, Scott S. Auerbach^{2*}

1 Center for Public Health and Environmental Assessment, Office of Research and Development, US EPA, Washington DC, United States of America, **2** Division of the National Toxicology Program, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, Durham, NC, United States of America

* Mezencev.Roman@epa.gov (RM); auerbachs@niehs.nih.gov (SSA)



Abstract

Whole-genome expression data generated by microarray studies have shown promise for quantitative human health risk assessment. While numerous approaches have been developed to determine benchmark doses (BMDs) from probeset-level dose responses, sensitivity of the results to methods used for normalization of the data has not yet been systematically investigated. Normalization of microarray data converts raw hybridization signals to expression estimates that are expected to be proportional to the amounts of transcripts in the profiled specimens. Different approaches to normalization have been shown to greatly influence the results of some downstream analyses, including biological interpretation. In this study we evaluate the influence of microarray normalization methods on the transcriptomic BMDs. We demonstrate using *in vivo* data that the use of alternative pipelines for normalization of Affymetrix microarray data can have a considerable impact on the number of detected differentially expressed genes and pathways (processes) determined to be treatment responsive, which may lead to alternative interpretations of the data. In addition, we found that normalization can have a considerable effect (as much as ~30-fold in this study) on estimation of the minimum biological potency (transcriptomic point of departure). We argue for consideration of alternative normalization methods and their data-informed selection to most effectively interpret microarray data for use in human health risk assessment.

OPEN ACCESS

Citation: Mezencev R, Auerbach SS (2020) The sensitivity of transcriptomics BMD modeling to the methods used for microarray data normalization. PLoS ONE 15(5): e0232955. <https://doi.org/10.1371/journal.pone.0232955>

Editor: Lars Kaderali, Universitätsmedizin Greifswald, GERMANY

Received: October 10, 2019

Accepted: April 25, 2020

Published: May 15, 2020

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All original data are held in a public repository under accession numbers indicated within the paper. All relevant derivative data are within the paper and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Whole-genome expression data generated by microarray studies are a promising resource for human health risk assessment. Analysis of these data can provide insight into mechanisms of biological processes, enable prediction of adverse outcomes of chemical exposures, and support estimation of points of departure (PODs) for derivation of toxicity values (reviewed in [1]).

Genomic dose-response studies (GDRS) done in an *in vivo* setting have been shown to identify gene set (e.g., pathway)-level PODs that approximate those identified using much more resource intensive guideline toxicity assessments [2,3]. In addition, mechanistic interpretation of GDRS data can yield a deeper understanding of molecular effects produced by tested substances in a manner that can support Adverse Outcome Pathway (AOP) development and human relevance determination [4].

One approach to analysis of GDRS data is embodied in the BMDEExpress software [5] which carries out a three-step process to identify gene set-level biological potency estimates referred to as benchmark dose (BMD) values. The three-step process includes (i) identification of probe sets (genes) responsive to chemical exposure, (ii) dose response model fitting of treatment responsive probe sets, and (iii) summarization of gene-level BMDs as gene set (e.g., pathway) level BMDs. Pathway level BMD and their lower confidence limits (BMDLs) are subsequently interpreted in the context of the lowest doses at which biological changes occur (i.e., biological effect point of departure [BEPOD] and biological point of departure lower bound [BEPOD/L], in the case of the lowest BMD and associated BMDL, respectively).

At each level of the analysis, parameter selection (e.g., choosing a minimum fold change) is made that can dramatically impact the amount of information that is carried through the analysis. For the assessment discussed here we have used a set of parameters identified by the National Toxicology Program (NTP) [6].

Before BMD modeling of microarray expression data can take place, the raw fluorescence signals must be processed. This processing, frequently referred to as “normalization”, mathematically transforms raw signals to expression estimates that are proportional to the amounts of corresponding transcripts in the profiled specimens. The need for normalization relates to a complex set of processes that can introduce non-biological variability, along with complex relationships between input quantities of mRNAs and signal intensities. These processes include, but are not limited to reverse transcription, labeling and hybridization on microarrays. Normalization of raw data from Affymetrix microarrays has been the focus of a remarkable amount of research due to the significant effect it can have when interpreting microarray data and the popularity of the Affymetrix platform. Numerous normalization methods are currently available, including MAS5.0.0, RMA, GCRMA, and PLIER. Each of these methods includes background adjustment (separates the specific signal from the non-specific signal), normalization and probe summarization steps [7]. Due to differences in assumptions underlying each of these methods [8], selection of normalization methods can significantly impact results of downstream analyses, such as identification of differentially expressed genes [9], clustering of genes or specimens [10], development of gene expression-based classifiers and building gene networks [11]. Previously published GDRS employed the most commonly used method for normalization of Affymetrix microarrays RMA (Robust Multi-array Average) [3,12–15]. Here we investigate the influence of different normalization methods on genomic dose-response modeling.

In this article we demonstrate the influence of different normalizations of expression microarrays on the findings from 5-day *in vivo* genomic dose-response studies of crude 4-methylcyclohexanemethanol (crude MCHM) [16], neat 4-methylcyclohexanemethanol (MCHM) [16], N,N-dimethyl-p-toluidine (DMPT) [17] and p-toluidine [17] in liver tissues, and propylene glycol phenyl ether (PPH) [16] in kidney tissues of orally-exposed rats. Our results indicate that normalization remarkably impacts the number of detected differentially expressed genes and responsive gene sets. In addition, we demonstrate that the different normalization methods can lead to changes in model fitting, which alter individual probe set BMD values. Estimates of minimum biological effect potency (BEPOD/L) were found to be robust to the effects of different normalization methods for some but not all chemicals. In

some cases, potentially impactful differences between BEPOD/L values were observed with different normalization methods, hence this variable should be considered carefully when performing genomic dose response analysis.

Materials and methods

Gene expression data

All expression data used in this study were generated by profiling liver specimens of rats using the Affymetrix Rat Genome 230 2.0 Array that contains more than 31,000 probe sets analyzing transcripts from 15,575 annotated unique genes (NetAffx Annotation File, Release 36; URL: <https://www.thermofisher.com/order/catalog/product/900505>).

Raw expression data for livers of male Harlan Sprague Dawley or male F344/N rats exposed orally for 5 days to crude 4-methylcyclohexanemethanol (crude-MCHM), or neat 4-methylcyclohexanemethanol (MCHM), and for kidneys of rats exposed orally to propylene glycol phenyl ether (PPH) were accessed from the Gene Expression Omnibus (GEO) as series GSE75655, GSE75657 and GSE75656, respectively. Hepatic transcriptomics data for F344/N rats exposed orally to N,N-dimethyl-p-toluidine (DMPT) or p-toluidine for 5 days [17] were accessed from the GEO as series GSE100502. All.cel files available for each replicated exposure level and vehicle controls were included in this analysis.

Crude MCHM represents a mixture of six components in addition to the major component 4-methylcyclohexanemethanol (MCHM) used as a coal cleaning liquid. Propylene glycol phenyl ether (PPH) is an industrial chemical used as a latex coalescent and a solvent for textile dyes. Toxicological significance of MCHM and PPH is associated with a recent spill and large scale contamination of drinking water in West Virginia [18]. N,N-dimethyl-p-toluidine (DMPT) is an accelerant for methyl methacrylate monomers in medical devices that has been shown to induce liver carcinogenesis in male and female F344/N rats and B6C3F1 mice in a 2-year oral exposure study. p-toluidine, structurally-related to DMPT, is reportedly a liver carcinogen in mice [17].

Data processing

The raw fluorescence signals was processed using the following methods: MAS5.0.0 (Microarray Affymetrix Suite version 5.0), RMA (Robust Multichip Analysis), GCRMA (GeneChip Robust Multichip Analysis) and PLIER (Probe Logarithmic Intensity Error Estimation) [7,19]. Raw data in.cel file format were imported into Expression Console Build 1.3.1.187 (Affymetrix, Santa Clara, CA, USA) and processed using the MAS5.0 and RMA methods with default configurations. Processing using the PLIER method was performed with PM-MM background correction and quantile normalization. Processing using the GCRMA method was implemented in R version 3.5.1 (<https://www.R-project.org>) using the “GCRMA: Background Adjustment Using Sequence Information” R package version 2.52.0 (Bioconductor version 3.7). PLIER16 values were calculated from PLIER values by adding 16 to each signal intensity value as a simple variance-stabilizing transformation. Low-quality microarrays were identified by visual inspection of the Relative Log Expression (RLE) boxplots generated by Expression Console (Affymetrix). RLE values are calculated for each probeset as the ratio between expression of this probeset in a given microarray and the median expression of this probeset across all the arrays in the dataset [20]. Low-quality microarrays were removed from the datasets and the raw data were re-normalized by the GCRMA, RMA and PLIER methods (MAS5.0 uses a per-chip approach and re-normalization was not needed). Normalized data used in this study for BMD modelling are available as [S1 Dataset](#) (also at the URL: <https://catalog.data.gov/organization/epa-gov>).

Differential expression analysis and transcriptomics BMD modeling

Normalized data were imported into BMDExpress for Windows, version 2.20 build 0167 BETA (<https://github.com/auerbachs/BMDExpress-2/releases>) [21]. For all chemicals, differentially expressed genes were determined from normalized datasets with all probesets included. To examine the influence of non-informative probesets, the analysis was also performed following removal of the probesets that showed "Absent" MAS5.0 absolute detection calls across all specimens from MAS5.0 and PLIER16-normalized data (MAS5.0_noA calls and PLIER16_noA calls methods).

Differentially expressed probesets were detected using the Williams trend test ($p < 0.05$; absolute fold change ≥ 1.5). Benchmark response of 1 SD was used for each feature. Best models among linear, 2nd degree polynomial, Hill, power and exponential models (degrees 2–5) were selected based on the lowest AIC. For more details on parameters used in BMDExpress, see [S1 File](#). Gene set-level BMD values were determined by mapping probes that met BMD filtering criteria ([S1 File](#)) to GO: Biological Process ontologies and Reactome Pathways [22]. The most sensitive GO:BP (Gene Ontology Biological Process) or Reactome pathways were identified as those with lowest median BMD or BMDL values calculated from all mapped probesets and reported as BEPOD or BEPOD/L, respectively.

Visualization and statistics

Comparison of differentially expressed genes was visualized using Venn diagrams (http://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/calculate_venn.html). All other plots were produced using TIBCO Spotfire Analyst version 7.8.0 (TIBCO Software Inc, Palo Alto, CA 94304, USA). Agreement between probe-level BMD or BMDL values determined by different normalization methods was assessed by Pearson correlation using TIBCO Spotfire Analyst version 7.11.10 (TIBCO Software Inc, Palo Alto, CA 94304, USA). Statistical significance of differences among probeset-level best model fit BMD and BMDL values for PPH determined by 7 normalization methods were tested using Kruskal-Wallis test and the differences were considered statistically significant for two-tail p -values < 0.05 . Most sensitive GO:BP gene sets were summarized for visualization using Revigo tool [23].

Results

Assessment of normalization

We explored the influence of 7 different microarray normalization methods and two different collections of gene sets on estimation of BEPOD/L values for five test articles using public data generated by genomic dose-response studies. Methodological differences between 7 used normalization methods are summarized in (Table 1 in [S1 File](#)).

Visual inspection of RLE plots for raw MCHM and PPH data identified low-quality microarrays that differed from other microarrays in medians and distributions of their RLE values. These microarrays were removed from datasets and differentially expressed genes (probe sets) and BMD values were determined from the remaining expression data ([S1 File](#), Figs 1–10 in [S1 File](#)).

Effect of normalization on the identification of differentially expressed probe sets and active gene sets

Different normalization methods identified remarkably different numbers of differentially expressed probe sets (DEPSs). The highest numbers of DEPSs for most chemicals were identified by using the MAS5.0 and PLIER methods, while PLIER16 and PLIER16-noAcalls identified considerably less probe sets corresponding to differentially expressed genes ([Fig 1A](#)).

Likewise, the GCRMA method produced more DEPSs than RMA with the single exception of p-toluidine (Fig 1A).

The use of MAS5.0 and PLIER also produced the highest numbers of DEPSs with acceptable model fits (AMFs; i.e. DEPSs where the global goodness of fit p -value >0.1 and the BMDU/BMDL ratio <40), which would be therefore included in the gene set BMD analysis (Fig 1B). Consistent with this observation, these two methods produced highest numbers of GO:BP and Reactome Pathways, from which gene-set-level BMDs and ultimately BEPOD values could be determined (Fig 1C and 1D). In contrast, the RMA, GCRMA, PLIER16 and PLIER16_noAcalls methods generally identified fewer DEPSs for which acceptable model fits, and therefore all identified fewer “active” GO:BP and Reactome pathways that could be used for estimation of BEPOD (Fig 1C and 1D).

Comparison of overlaps among DEPSs with acceptable model fits across the 7 normalization methods was also performed for all data sets. In all chemical-organ sets, the percent of overlapping DEPSs with AMFs was quite small ($<10\%$) when all normalizations were compared, suggesting that normalization is likely to have a large impact on the qualitative interpretation (e.g., mode of action, AOP assessment) of the toxicogenomic effects. To obtain a sense of how great the effect normalization has on the identification of DEPS with AMF, an “intersection” set for all normalizations per treatment-chemical pair was identified and used as a comparator for each normalization in a given chemical-treatment set (Fig 2A). This analysis showed that normalization methods such as MAS5.0 and PLIER consistently identified a greater number of DEPSs with acceptable model fits compared with other normalization methods. To more accurately quantify the relative increase and DEPSs with AMFs, a fold increase over intersection for each normalization in a chemical-organ pair was calculated (Fig 2B). MAS5.0 (median ~ 9.5) and PLIER (median ~ 7) showed the greatest fold increase in DEPSs with BMDs whereas PLIER16 normalizations showed the smallest fold increases of intersection (Fig 2B). These findings are consistent with the relatively larger number of active GO:BP and Reactome pathways that are identified when using MAS5.0 and PLIER normalized data (Fig 1C and 1D). In addition, these results are consistent with analysis of overlaps among differentially expressed genes identified by different normalization methods (S1 File, Fig 11 in S1 File).

Effect of normalization on the distribution of probeset BMD and BMDL

Distributions of probeset-level BMD and BMDL values showed differences across normalization methods (Fig 3A and 3B). For example, differences among distributions of best model fit BMD and BMDL values for PPH were found statistically significant (Kruskal-Wallis $p < 0.0001$ for BMD and BMDL). Median probeset-level BMD value for the MAS5.0 method (BMD = 637.4 mg/kg-day) was higher than median BMD values determined by all other normalization methods (range: 341.6–583.9 mg/kg-day); however, the lowest probeset-level BMDs for PLIER16 and PLIER16_noA (71.1 mg/kg-day) were found substantially higher than corresponding values for MAS5.0 and other normalization methods (range: 0.36–19.3 mg/kg-day) (Fig 3A). Notably, median probe-set level BMD values were highest for MAS5.0 normalization across all chemical-treatment pairs (Fig 3C). Similar manifestations are present in the gene set analysis where the GO:BP median BMD value distributions were higher when using MAS5.0 and generally lowest with PLIER16 (Fig 3D and 3E)

Effect of normalization on the lowest gene set BMD and BMDL (BEPOD and BEPOD/L) determination

Selection of different normalization methods influenced determination of lowest median gene set-level BMD (BEPOD) and BMDL (BEPOD/L) for different chemicals to a different extent.

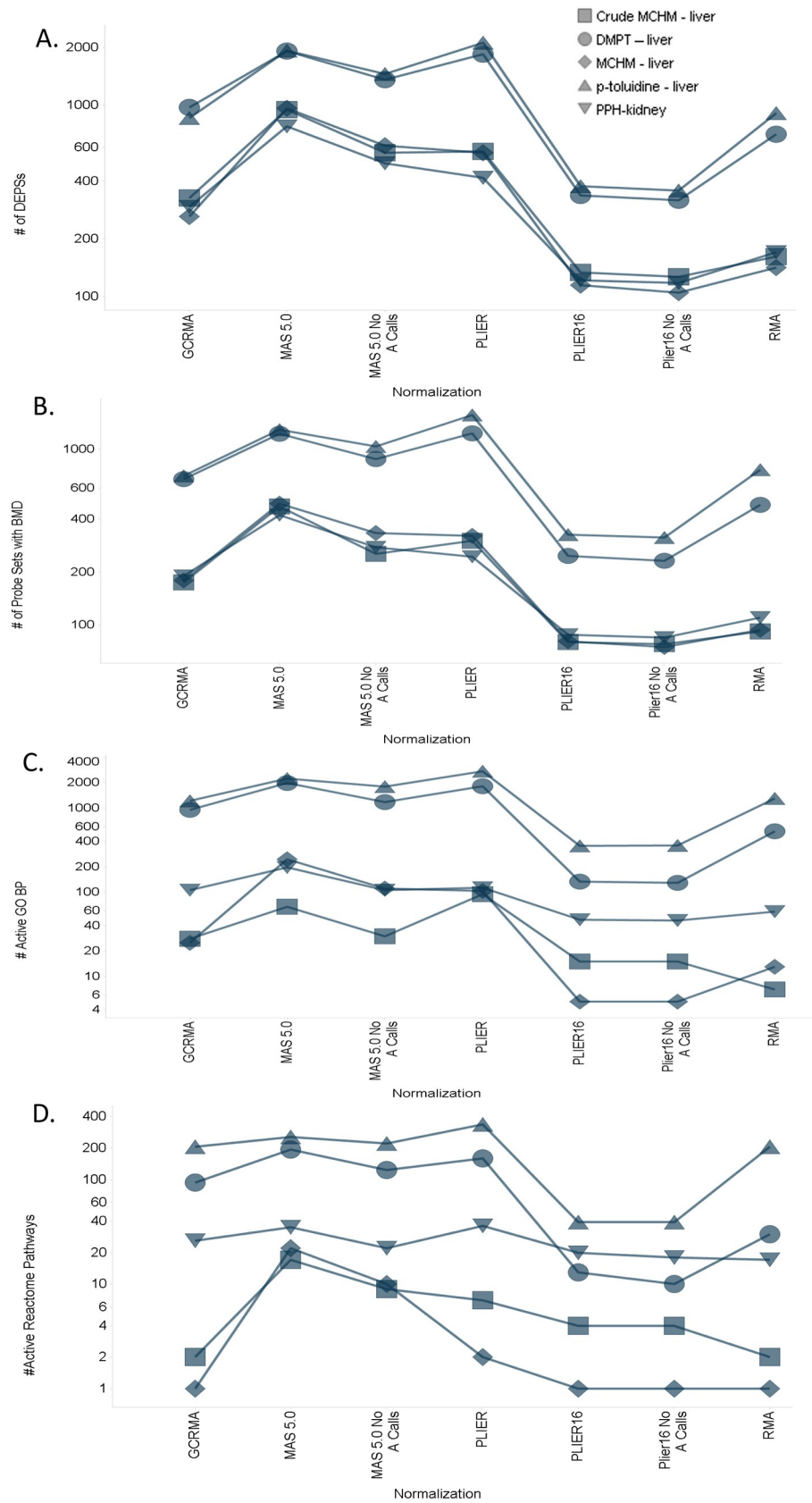


Fig 1. Effect of different normalizations on the number of (A) differentially expressed probe sets (DEPSs), (B) probe sets with acceptable BMD estimates, (C) active GO Biological Processes and (D) active Reactome Pathways.

<https://doi.org/10.1371/journal.pone.0232955.g001>

Within a gene set type for any one treatment, the BEPOD/L variation ranged from 1.1 fold (Crude MCHM, Reactome BMDL) to up to 30.4 fold (P-toluidine, GO:BP BMDL) due to different normalizations (Fig 4 and S1 File, Table 2 in S1 File). The mean and median variations in BEPOD/L values for a chemical treatment-gene set pair due to different normalizations were 6.2- and 3-fold, respectively. In general, MAS5.0 normalization tended to yield the lowest BEPOD/L values and PLIER16 the highest (S1 File, Fig 12 in S1 File). Further, GO:BP gene sets tended to produce lower BEPOD/Ls compared to the Reactome Pathway set, likely due to the relatively larger number curated gene sets included in the GO:BP.

Effect of normalization on the identity of the gene sets with the lowest BMD and BMDL (BEPOD and /L)

Influence of normalization methods on the identity of the most sensitive GO Biological Processes and Reactome pathways varied across datasets (Fig 5A and 5B). For instance, the same three most sensitive GO:BP gene sets (GO:0002933, GO:0070988, and GO:0070989) were found for all normalization methods in the case of the MCHM data set (Fig 5A and S1 File, Table 3 in S1 File). In contrast, a variety of different GO:BP gene sets were identified as the most sensitive when using different normalizations of the p-toluidine dataset. While “MAS5.0_noA calls” and “PLIER16_noA calls” detected the same most sensitive pathways as their corresponding parental methods, MAS5.0 and PLIER16, other normalization methods identified seemingly biologically unrelated most sensitive GO:BP gene sets such as, e.g. GO:00043588 “Skin development” (RMA), GO:0044246 “Regulation of multicellular organismal metabolic process” (GCRMA), GO:0007040 “Lysosome organization” (MAS5.0), and GO:0048266 “Behavioral response to pain” (PLIER) (Fig 5C). Further, median BMD values corresponding to these pathways were also appreciably different (S1 File, Table 2 in S1 File).

Agreement of Gene Set BMD and BMDL values across normalizations

To determine the effect of normalization on overall agreement of the BMD values, Pearson correlation between both probe set BMD values and median BMD values for GO:BP and Reactome pathways in each of the five experiments was performed. In nearly all cases, BMD values from probe sets (PS) or gene sets exhibited positive correlation, however the range of Pearson correlation coefficients (PCCs) varied from as high as 1, with average PCCs for any pair of normalizations mostly falling between 0.2 to 0.4 (Fig 6A). In most cases, PS BMDs and median BMDs had higher agreement for GO:BP than for Reactome pathways. Evaluation of individual pairings of normalizations across experiments demonstrated the MAS5.0 group of normalizations and PLIER16 group normalization show high intragroup similarity across all PSs and gene set BMDs (Fig 6B). Other pairings such as RMA-MAS5.0, PLIER16-MAS5.0, PLIER--MAS5.0 show considerably worse agreement. Notably, the agreement of BMD values from Reactome pathways was found to be more negatively affected by changes in normalization than BMD values for GO:BP gene sets.

Discussion

Gene expression profiling has been identified as a promising method to address challenges in chemical risk assessment that have traditionally relied on data generated by time and resource consuming animal studies. This promise is especially true for short-term, in vivo, dose-

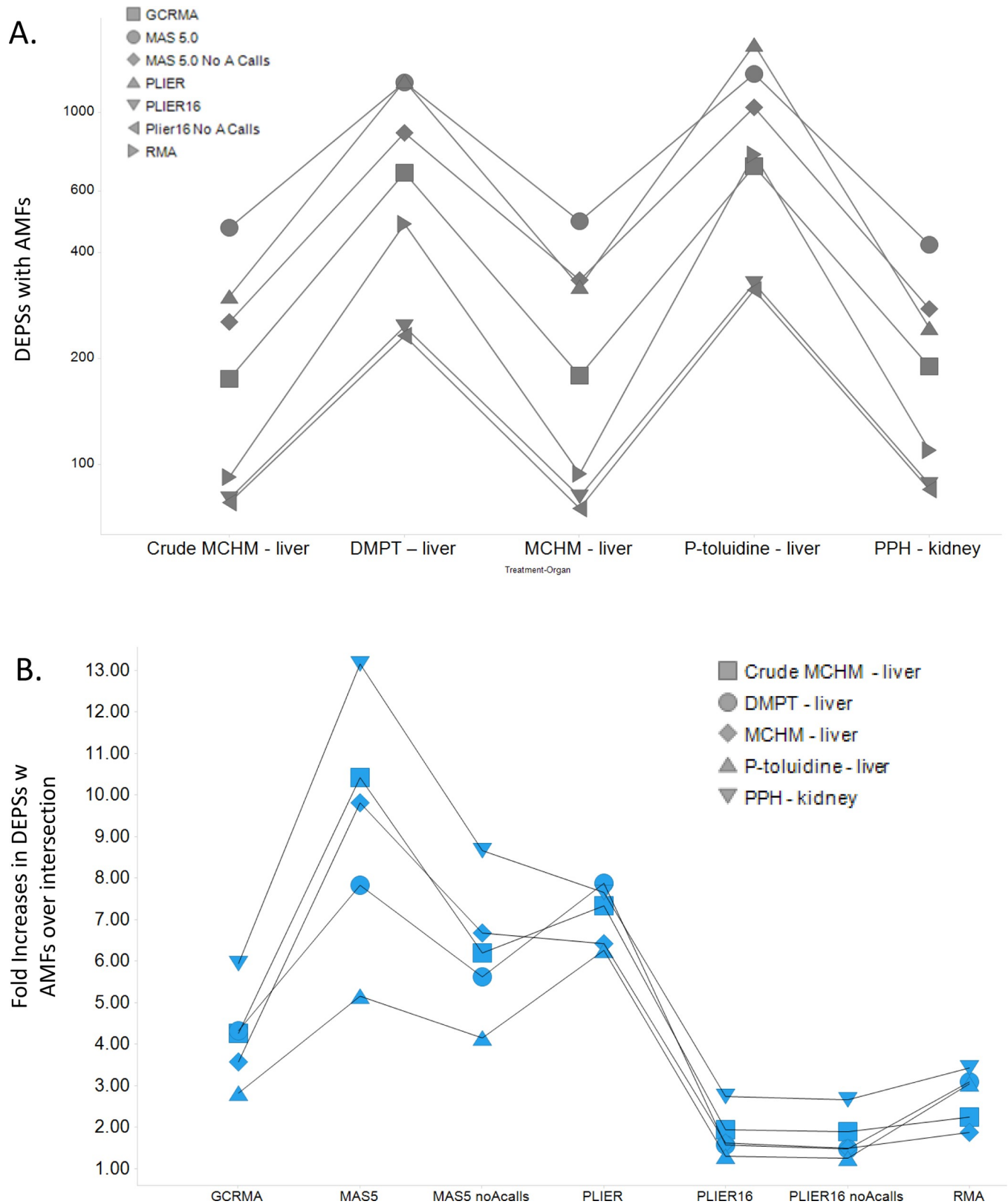


Fig 2. (A) Effect of different normalizations on the number of differentially expressed probe sets (DEPs) with acceptable model fits (AMFs). (B) Fold increases in DEPs over an intersection of all normalizations for any individual chemical-treatment pairs.

<https://doi.org/10.1371/journal.pone.0232955.g002>

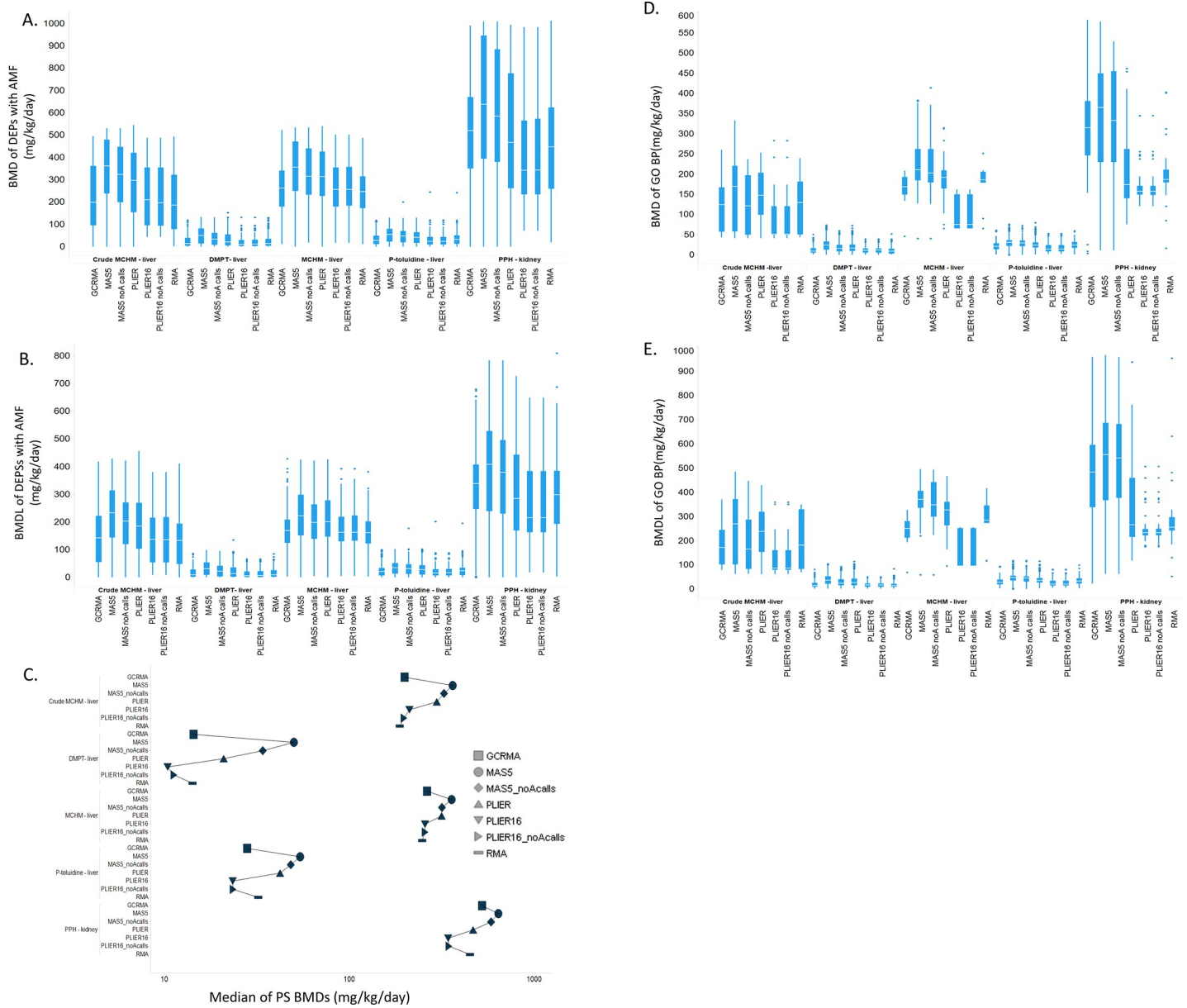


Fig 3. Impact of normalization on the overall distribution of probe set-level BMD (A) and BMDL (B) values and GO:BP median BMD (D) and BMDL (E) values in each of the 5 experiments.

<https://doi.org/10.1371/journal.pone.0232955.g003>

response microarray studies that demonstrated potential to produce transcriptomics benchmark doses in good agreement with traditional toxicological studies [1,2].

Unlike RNA-seq, where methods for analysis are still evolving, gene expression microarrays are an established tool in transcriptomics studies. Considering the relative maturity of data analysis of expression microarrays, they will likely continue to play a substantial role in whole-genome expression studies particularly in cases where the data will be used for critical decision making (e.g., in regulatory context) [1].

Analysis of microarray expression data needs to address a non-biological variability introduced by sample preparation, labeling, hybridization, fluorescence reading and other technical

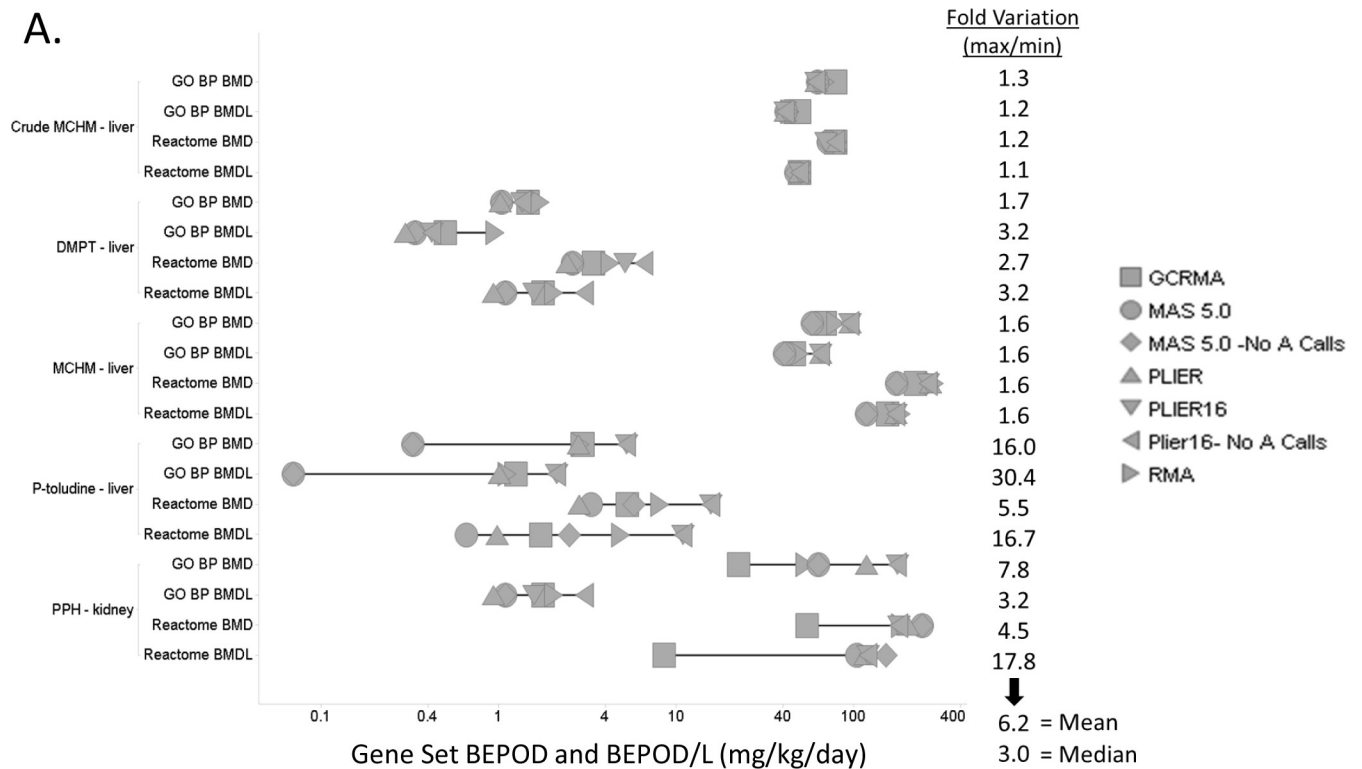


Fig 4. Effect of normalization on the lowest median BMD (BEPOD) and BMDL (BEPOD/L) values for GO:BP and Reactome pathways. The distribution of BEPOD and BEPOD/L values across 7 normalizations are shown for each of the 5 experiments. Fold variation: max BEPOD to min BEPOD (or max BEPOD/L to min BEPOD/L) ratio. Mean and median fold variations are shown for BEPOD/L values.

<https://doi.org/10.1371/journal.pone.0232955.g004>

issues through normalization of raw microarray data. Normalization procedures attempt to remove non-biological variability from data by exploiting and enforcing known or assumed invariances of the data using different approaches [26]. Not surprisingly, numerous normalization methods have been developed for different platforms of expression microarrays [7,19,27–29]. These methods can produce considerably different results of downstream studies utilizing gene expression data, which was previously shown for some specific types of analyses, such as class discovery and gene co-expression analyses [9–11]. This study is the first to systematically evaluate the effect of microarray normalization on the transcriptomic BMD modeling. Considering the potential influence of normalization methods due to their different underlying assumptions and performance, the absence of insight on microarray normalization, especially in the context of quantitative toxicogenomics, may be perceived as a gap impairing the use of microarray data in chemical risk assessment.

The purpose of this study was to examine sensitivity of transcriptomics BMD modeling to some frequently used methods for microarray data normalization. Our results demonstrate that the use of different normalization methods produces considerably different lists of differentially expressed genes. Further, we found that the influence of microarray normalization methods on the results of transcriptomics BMD modeling is limited on some, but considerable on other datasets. We show that the fold differences between BMDL values determined from microarray data normalized by different methods can be as low as 1.1 -fold, but also as big as 30.4-fold. These differences are potentially noteworthy, because transcriptomic and apical BMD values can differ by a factor of ten, when transcriptomic values are determined using RMA normalization [3]. Historically, genomic BMD analysis using Affymetrix microarrays

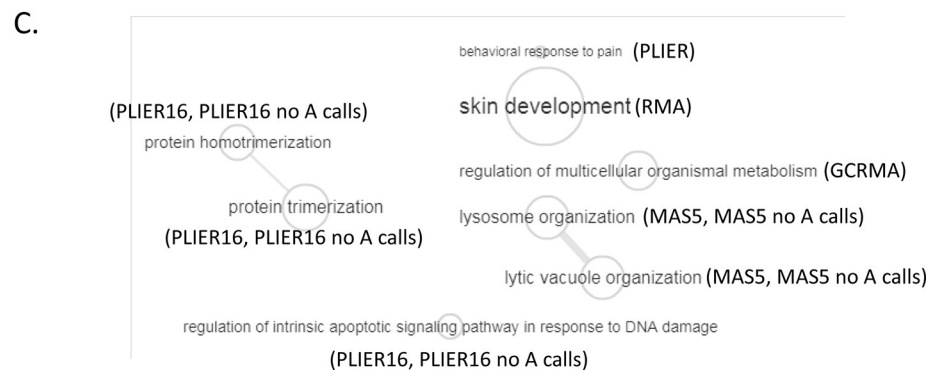
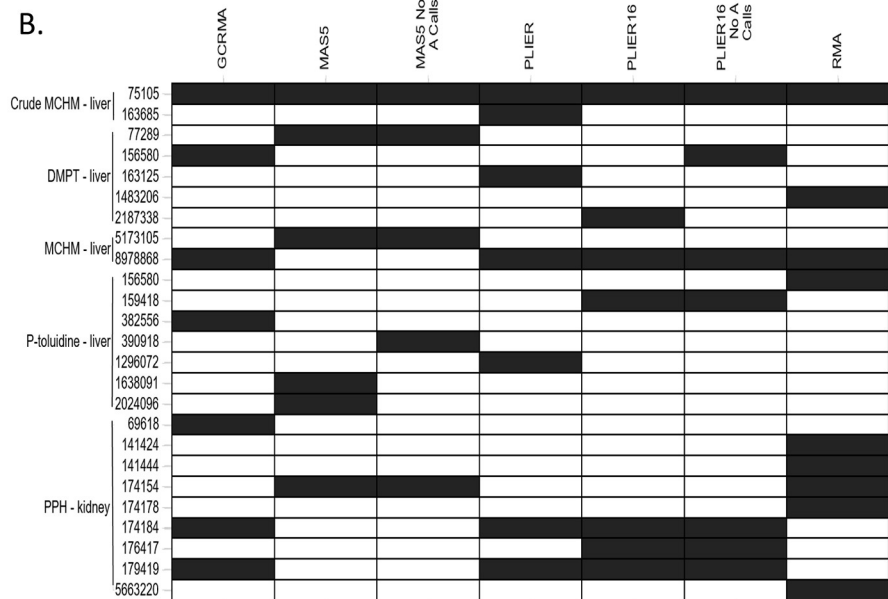
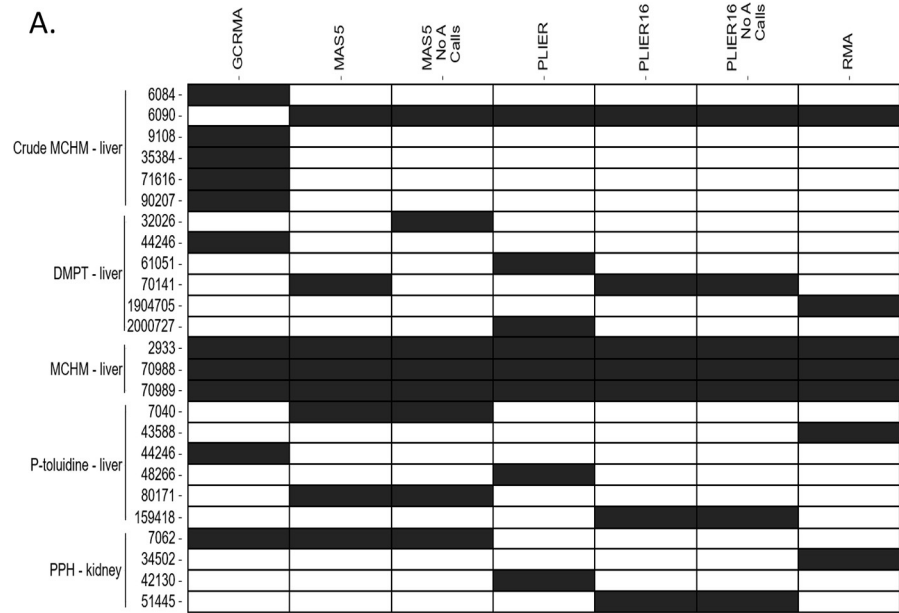


Fig 5. GO Biological Processes (A) and Reactome Pathways (B) that were identified as most sensitive when different normalizations were employed. A black cell indicates that the gene set was identified as the gene set with the lowest median BMD value. (C) Visualization of most sensitive GO:BP terms for p-toluidine/liver data across different normalizations. Most sensitive GO:BP terms for this chemical/tissue pair were found to be most diverse across normalizations. REVIGO tool [23] was used to remove redundant terms (conservative removal; allowed similarity = 0.9) and connect highly similar GO:BP terms. Similarity measure: SimRel score [24]. The size of bubbles indicates the frequency of the GO term in the Gene Ontology Database (GOA) for *Rattus norvegicus* [25].

<https://doi.org/10.1371/journal.pone.0232955.g005>

has relied exclusively on the RMA method; however, the findings presented here suggest that the appropriateness of the choice of normalization methods should likely be considered on a case-by-case basis. Consideration on a case-by-case basis instead of choosing a single method as suitable for all data reflects the fact, that normalization methods differ in their fundamental assumptions, and these assumptions are satisfied by real data to varying degrees. This consideration may include behavior of normalized data (e.g., through RLE and MA plots), number of genes and gene sets that could be used for determination of BEPOD and BEPOD/L values, emphasis on sensitivity/specificity and possibly also examination of quality of probe sets that mapped to the most sensitive gene sets and subsequently projected into BEPOD and BEPOD/L values. If decision among several normalization methods cannot be made, normalizations that provide lowest BEPOD estimates for specific datasets will be interesting in the context of quantitative risk assessment as the most protective. In our study, the MAS5.0 and MAS5.0_noA methods provided lowest (or next to lowest) BEPOD and BEPOD/L estimates for three datasets, GCRMA for one dataset, and all normalization methods were equivalent for one dataset.

Prior comparisons of microarray normalization methods have not produced a clear winner. While some investigators found GCRMA to perform as well as or better than other methods [30], others reported good performance of PLIER and its superiority over MAS5.0 [31], and some studies favored RMA over other methods [32]. In gene expression correlation studies, MAS5.0 reportedly outperformed the GCRMA, RMA and Li-Wong methods [11]. Normalization methods differ in their precision and accuracy and the most precise methods have been shown to be generally less accurate, while more accurate methods tend to have low precision [19]. For instance, RMA offers higher precision than the MAS5.0 method, which introduces high variability, particularly into low-intensity probes. Nevertheless, MAS5.0 provides linear relationship between signal and transcript concentration even at low transcript concentrations when the RMA method introduces bias [19], and the use of MAS5.0 alongside with detection calls substantially improves its performance to detect differentially expressed genes [33]. Furthermore, relative importance of precision and accuracy seems to depend on the purpose of transcriptomics analysis. For instance, classification and clustering problems benefit from more precise methods, because in these analyses variability accumulates from all probe sets and quickly impairs their results. While we do not attempt to propose a guideline regarding the use of microarray normalization methods, we argue that different normalization methods need to be considered as a part of quantitative toxicogenomic studies and that sensitivity of reported BEPOD and BEPOD/L values to different normalization, or justification of appropriateness of the selected normalization method need to be provided. It should be also noted that normalization methods less sensitive to violations of a broader range of assumptions may be developed in future. This would benefit dose-response gene expression studies, in which normalization methods need to address likely substantial differences among transcriptomes at low and high exposure levels, which violate assumption common to several normalization methods, that most genes are not differentially expressed. Consequently, the RMA pipeline, which

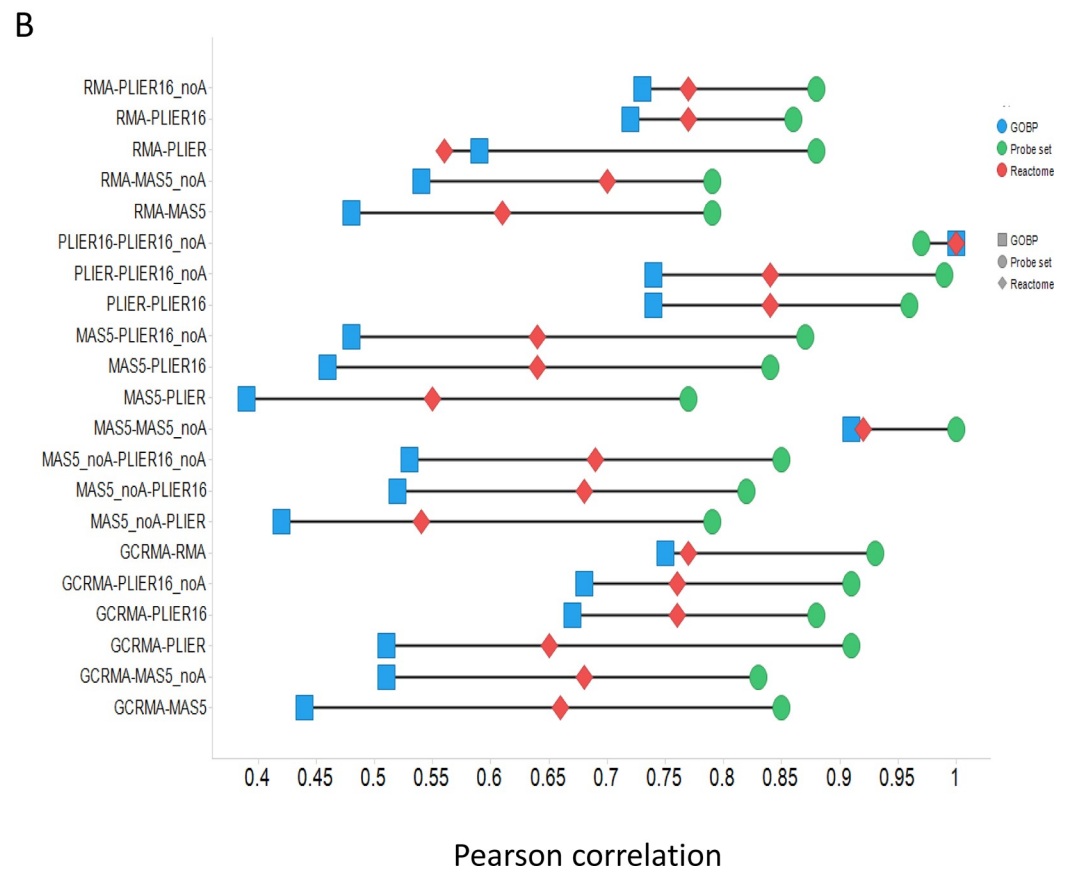
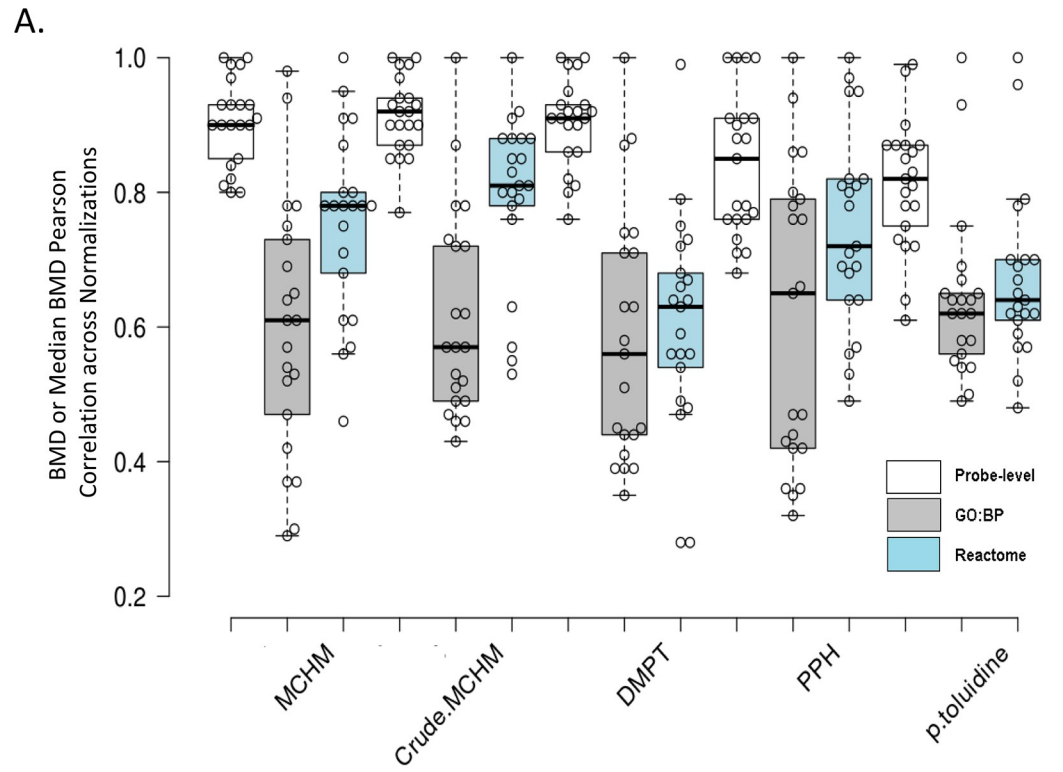


Fig 6. Agreement of BMD values generated by different normalizations. (A) The distribution and 1.5×IQR (interquartile range) of the Pearson correlation coefficients (PCCs) reflecting agreement of probe set-level or gene set-level BMD values between different pairs of normalizations. For each experiment, PCCs for 21 pairs of normalizations were aggregated and presented separately for best model fit probeset level BMDs, median GO:BP BMDs and median Reactome pathway BMDs. Produced by BoxPlotR (<http://boxplot.tyerslab.com>). (B) Average PCC values are shown for 21 different pairings.

<https://doi.org/10.1371/journal.pone.0232955.g006>

has been almost exclusively used for processing of Affymetrix microarray data in quantitative toxicogenomics, should not be applied indiscriminately.

Supporting information

S1 Dataset.

(ZIP)

S1 File.

(PDF)

S1 Fig. Plot GO BP median BMD (F) values under different normalizations.

(PPTX)

S2 Fig. Patterns of effect on BEPOD/L values as a result of different normalizations of the data from the 5 data sets.

(PPTX)

Author Contributions

Conceptualization: Roman Mezencev.

Data curation: Roman Mezencev, Scott S. Auerbach.

Formal analysis: Roman Mezencev, Scott S. Auerbach.

Investigation: Roman Mezencev, Scott S. Auerbach.

Methodology: Roman Mezencev, Scott S. Auerbach.

Project administration: Roman Mezencev.

Resources: Roman Mezencev.

Software: Roman Mezencev, Scott S. Auerbach.

Validation: Roman Mezencev.

Visualization: Roman Mezencev, Scott S. Auerbach.

Writing – original draft: Roman Mezencev, Scott S. Auerbach.

Writing – review & editing: Roman Mezencev, Scott S. Auerbach.

References

1. Mezencev R, Subramaniam R (2019) The use of evidence from high-throughput screening and transcriptomic data in human health risk assessments. *Toxicology and Applied Pharmacology* 380: 114706. <https://doi.org/10.1016/j.taap.2019.114706> PMID: 31400414
2. Thomas RS, Waters MD (2016) Transcriptomic dose-response analysis for mode of action and risk assessment. In: Waters MD, Thomas RS, editors. *Toxicogenomics in predictive carcinogenicity*. Cambridge, UK: The Royal Society of Chemistry. pp. 154–184.

3. Farmahin R, Williams A, Kuo B, Chepelev NL, Thomas RS, et al. (2017) Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. *Arch Toxicol* 91: 2045–2065. <https://doi.org/10.1007/s00204-016-1886-5> PMID: 27928627
4. Webster AF, Lambert IB, Yauk CL (2016) Toxicogenomics case study: Furan. In: Waters MD, Thomas RS, editors. *Toxicogenomics in predictive carcinogenicity*. Cambridge, UK: The Royal Society of Chemistry. pp. 390–422.
5. Phillips JR, Svoboda DL, Tandon A, Patel S, Sedykh A, et al. (2018) BMDEExpress 2: enhanced transcriptomic dose-response analysis workflow. *Bioinformatics* 35: 1780–1782.
6. Program NT (2018) NTP Research Report on National Toxicology Program Approach to Genomic Dose-Response Modeling. In: Sciences NloEH, editor. Research Triangle Park, NC, USA: National Toxicology Program. pp. 1–42.
7. Bolstad B (2008) Preprocessing and normalization for Affymetrix GeneChip expression microarrays. In: Stafford P, editor. *Methods in microarray normalization*. Boca Raton, FL, USA: CRC Press. pp. 41–59.
8. Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, Kimmel M (2015) Microarray experiments and factors which affect their reliability. *Biology Direct* 10: 46. <https://doi.org/10.1186/s13062-015-0077-2> PMID: 26335588
9. Stafford P, Tak Y (2008) Biological interpretation for microarray normalization selection. In: Stafford P, editor. *Methods in microarray normalization*. Boca Raton, FL, USA: CRC Press. pp. 151–172.
10. Freyhult E, Landfors M, Onskog J, Hvidsten TR, Ryden P (2010) Challenges in microarray class discovery: a comprehensive examination of normalization, gene selection and clustering. *BMC Bioinformatics* 11: 503. <https://doi.org/10.1186/1471-2105-11-503> PMID: 20937082
11. Lim WK, Wang K, Lefebvre C, Califano A (2007) Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* 23: i282–288. <https://doi.org/10.1093/bioinformatics/btm201> PMID: 17646307
12. Thomas RS, Allen BC, Nong A, Yang L, Bermudez E, et al. (2007) A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. *Toxicol Sci* 98: 240–248. <https://doi.org/10.1093/toxsci/kfm092> PMID: 17449896
13. Thomas RS, Clewell HJ 3rd, Allen BC, Wesselkamper SC, Wang NC, et al. (2011) Application of transcriptional benchmark dose values in quantitative cancer and noncancer risk assessment. *Toxicol Sci* 120: 194–205. <https://doi.org/10.1093/toxsci/kfq355> PMID: 21097997
14. Thomas RS, Wesselkamper SC, Wang NCY, Zhao QJ, Petersen DD, et al. (2013) Temporal Concordance Between Apical and Transcriptional Points of Departure for Chemical Risk Assessment. *Toxicological Sciences* 134: 180–194. <https://doi.org/10.1093/toxsci/kft094> PMID: 23596260
15. Dean JL, Zhao QJ, Lambert JC, Hawkins BS, Thomas RS, et al. (2017) Editor's Highlight: Application of Gene Set Enrichment Analysis for Identification of Chemically Induced, Biologically Relevant Transcriptomic Networks and Potential Utilization in Human Health Risk Assessment. *Toxicol Sci* 157: 85–99. <https://doi.org/10.1093/toxsci/kfx021> PMID: 28123101
16. Program NT (2016) West Virginia Chemical Spill: 5-DayRat Toxicogenomic Studies July 2016 NTP Update. Research Triangle Park, NC, USA: National Institute of Environmental Health Sciences. 1–5 p.
17. Dunnick JK, Shockley KR, Morgan DL, Brix A, Travlos GS, et al. (2017) Hepatic transcriptomic alterations for N,N-dimethyl-p-toluidine (DMPT) and p-toluidine after 5-day exposure in rats. *Arch Toxicol* 91: 1685–1696. <https://doi.org/10.1007/s00204-016-1831-7> PMID: 27638505
18. Whelton AJ, McMillan L, Novy CLR, White KD, Huang X (2017) Case study: the crude MCHM chemical spill investigation and recovery in West Virginia USA. *Environmental Science: Water Research & Technology* 3: 312–332.
19. Irizarry RA, Wu Z, Jaffee HA (2006) Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* 22: 789–794. <https://doi.org/10.1093/bioinformatics/btk046> PMID: 16410320
20. Bolstad BM, Collin F, Simpson KM, Irizarry RA, Speed TP (2004) Experimental Design and Low-Level Analysis of Microarray Data. *International Review of Neurobiology*: Academic Press. pp. 25–58.
21. Yang L, Allen BC, Thomas RS (2007) BMDEExpress: a software tool for the benchmark dose analyses of genomic data. *BMC Genomics* 8: 387. <https://doi.org/10.1186/1471-2164-8-387> PMID: 17961223
22. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, et al. (2018) The Reactome Pathway Knowledgebase. *Nucleic Acids Res* 46: D649–D655. <https://doi.org/10.1093/nar/gkx1132> PMID: 29145629
23. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* 6: e21800. <https://doi.org/10.1371/journal.pone.0021800> PMID: 21789182

24. Mazandu GK, Chimusa ER, Mulder NJ (2017) Gene ontology semantic similarity tools: survey on features and challenges for biological knowledge discovery. *Briefings in bioinformatics* 18: 886–901. <https://doi.org/10.1093/bib/bbw067> PMID: 27473066
25. Huntley R, Dimmer E, Barrell D, Binns D, Apweiler R (2009) The Gene Ontology Annotation (GOA) Database. *Nature Precedings*.
26. Kreil DP, Russell RR There is no silver bullet—a guide to low-level data transforms and normalisation methods for microarray data.
27. Liu X, Li N, Liu S, Wang J, Zhang N, et al. (2019) Normalization Methods for the Analysis of Unbalanced Transcriptome Data: A Review. *Frontiers in Bioengineering and Biotechnology* 7.
28. Cheng L, Wang X, Wong P-K, Lee K-Y, Li L, et al. (2016) ICN: a normalization method for gene expression data considering the over-expression of informative genes. *Molecular BioSystems* 12: 3057–3066. <https://doi.org/10.1039/c6mb00386a> PMID: 27452923
29. Cheng L, Lo L-Y, Tang NL, Wang D, Leung K-S (2016) CrossNorm: a novel normalization strategy for microarray data in cancers. *Scientific reports* 6: 18898. <https://doi.org/10.1038/srep18898> PMID: 26732145
30. Gharaibeh RZ, Fodor AA, Gibas CJ (2008) Background correction using dinucleotide affinities improves the performance of GCRMA. *BMC Bioinformatics* 9: 452–452. <https://doi.org/10.1186/1471-2105-9-452> PMID: 18947404
31. Therneau TM, Ballman KV (2008) What Does PLIER Really Do? *Cancer Informatics* 6: 423–431. PMID: 19259420
32. Millenaar FF, Okyere J, May ST, van Zanten M, Voesenek LACJ, et al. (2006) How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* 7: 137–137. <https://doi.org/10.1186/1471-2105-7-137> PMID: 16539732
33. Pepper SD, Saunders EK, Edwards LE, Wilson CL, Miller CJ (2007) The utility of MAS5 expression summary and detection call algorithms. *BMC Bioinformatics* 8: 273–273. <https://doi.org/10.1186/1471-2105-8-273> PMID: 17663764