TITLE (<100 char)

Rat Deconvolution as Knowledge Miner for Immune Cell Trafficking from Toxicogenomics Databases

Katsuhisa Morita*, [1], Tadahaya Mizuno*, †, [2], Iori Azuma[1], and Hiroyuki Kusuhara

[1]Graduate School of Pharmaceutical Sciences, the University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan

[2]Graduate School of Pharmaceutical Sciences, the University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan, tadahaya@gmail.com

* Equal contribution

† to whom correspondence should be addressed

Keywords: ALT/legacy data/transcriptome/ (alphabet order, <5 words, not in title)

ABSTRACT

Toxicogenomics databases are useful for understanding biological responses in individuals because they are derived from well-controlled experiments and include a diverse spectrum of biological responses. Although these databases contain no information regarding immune cells in the liver, which are important in the progression of liver injury, deconvolution that estimates cell-type proportions from bulk transcriptome could add information regarding immune cell trafficking to the database. However, deconvolution has been mainly applied to humans and mice and less often to rats, which are the main target of toxicogenomics databases. Here, we developed a deconvolution method for rats and established a methodology to obtain information regarding immune cells from toxicogenomics databases. The contributions of this work are three-fold. First, we obtained the gene expression profiles of various rat immune cells necessary for deconvolution and constructed a dataset; second, we compared the accuracy of models based on human and mouse datasets and showed the impact of species differences on deconvolution; third, we showed that rat deconvolution could retrieve information regarding immune cell trafficking from toxicogenomics databases.

INTRODUCTION

Immune cell trafficking is one of the most important factors in the progression of liver injury, such as drug-induced liver injury (DILI) (Yang & Tonnesseen, 2019; Jaeschke *et al*, 2012; Hossain & Kubes, 2019). While the importance of various immune cells, such as Kupffer cells and NKT cells, has been demonstrated individually, their contribution to the progression of liver injury and their relationship with each other have remained largely unknown (Liu *et al*, 2006; Laskin *et al*, 1995; Yang *et al*, 2019; Liu *et al*, 2004). One possible explanation for this situation is that it is difficult to compare accumulated findings due to the complexity of the experimental systems, such as the fact that liver injury models that have been constructed using various methods including compound administration, diet, and surgery, as well as different time points for evaluation (Cai *et al*, 2017; Gehring *et al*, 2006; Graubardt *et al*, 2017; Chen *et al*, 2017).

Large-scale toxicity databases of compounds (toxicogenomics databases), such as Open TG-GATEs and DrugMatrix, are useful for understanding biological responses in individuals (Igarashi *et al*, 2015, https://ntp.niehs.nih.gov/data/drugmatrix). This is because compound administration is relatively easy to control and less confounded compared with other model preparation approaches such as diet and surgery, making it easier to compare findings. In addition, considering that compounds generally have multiple effects, including those unrecognized, it is also expected that a diverse spectrum of biological responses can be covered due to nonarbitrary perturbations by various compounds (Morita *et al*, 2020; Nemoto *et al*, 2021). Open TG-GATEs, for example, contain transcriptome data, pathological images, and blood biochemistry of tissues collected at various time points from rats treated with various compounds at various concentrations, while there is no information regarding immune cells in tissue in these databases.

Deconvolution is a data analysis method to estimate the ratio of cells in a specimen from bulk transcriptome data (Abbas *et al*, 2009; Im & Kim, 2023). Applying this method to the transcriptome data of a toxicogenomics database, could add information regarding immune cell trafficking as a new layer to the database and contribute to a deeper understanding of immune cell relationships in liver injury. However, deconvolution has been mainly applied to humans and mice, and there have been few applications to rats, which are the main target of large-scale toxicity databases (Chen *et al*, 2018b; Altboum *et al*, 2014; Wang *et al*, 2021; Petitprez *et al*, 2020). In particular, although deconvolution generally requires a gene expression dataset of various immune cells for training, to our knowledge, such a dataset does not exist and all reports applying deconvolution to rats have used human immune cells as a substitute (Gil Del Alcazar *et al*, 2022; Wang *et al*, 2021).

Here, we developed a rat deconvolution method, examined species differences in the

deconvolution method, and established a methodology to obtain immune cell information from valuable toxicogenomics databases (**Fig 1**).

RESULTS

Extension of Application of Deconvolution to Rats

Deconvolution requires gene expression profiles of the sample to be analyzed (mixture profiles) and gene expression profiles of the target cells with the specific marker genes (signature matrix) created from reference profiles. Given that $M \in \mathbb{R}^{n \times p}$ are the mixture profiles, $B \in \mathbb{R}^{k \times n}$ is the signature matrix, and $F \in \mathbb{R}^{n \times k}$ indicates the scores corresponding to the cell composition, deconvolution calculates F in the following equation: $M = F \times B$, where n, p, and k are the number of samples, genes, and the estimated cell types.

To extend the application of deconvolution to rats, we obtained gene expression profiles of representative immune cell types for the rat-derived reference profiles. Neutrophils, monocytes, CD4 T cells, CD8 T cells, natural killer cells, and B cells were selected and sorted from the spleen or blood of rats. The immune cell types used in this study are equivalent to LM6, the well-known reference profiles for humans composed of the representative 6 immune cell types prepared by Chen et al. (Chen *et al*, 2018a). Spleen and blood, whose immune cell composition is known to be different, and the liver derived from rats treated with liver toxicants were obtained and subjected to both transcriptome analysis and flow cytometry simultaneously for validation. As liver toxicants, we selected acetaminophen and methylene dianiline, which are known to induce neutrophil and monocyte migration 24 h after treatment, respectively (Mitchell *et al*, 1973; Bailie *et al*, 1994).

We compared the ratio of immune cells measured by flow cytometry and those estimated by applying deconvolution using the prepared rat-specific reference profiles to the counterpart mixture profiles. Regarding spleen and blood, the results showed that a positive correlation was observed for CD4 T, CD8 T, NK cells, neutrophils, and B cells between the evaluated tissues, whereas the estimate for monocytes varied (**Fig 2A**). The spleen contains resident macrophages with profiles similar those to monocytes (Davies *et al*, 2013; Ingersoll *et al*, 2011), which could interfere with deconvolution and be the cause the variation. We performed the same analysis on liver samples subjected to chemical perturbations and found a high positive correlation, especially for neutrophils and monocytes (**Fig 2B**).

We also examined the extrapolation using human and mouse signature matrices derived from existing reference profiles. The accuracy was very low especially for CD4 T and CD8 T cells, indicating that there are species differences in the expression of the gene cluster of each reference immune cell and that the rat gene expression profiles are required to create a signature matrix for rat deconvolution (**Figs EV1 and 2**).

Based on these findings, we concluded that deconvolution in rat gene expression profiles had been established. Notably, in neutrophils and monocytes, the accumulation by

perturbation of the compound can be readily predicted with deconvolution. However, the predictivity may be low for other immune cells including NK cells, which remain to be investigated in future works (**Fig EV3**).

Validation of Rat-specific Deconvolution to a Public Toxicogenomics Database

Open TG-GATEs is a representative rat toxicogenomics database in which more than 150 compounds were administered to rats at multiple concentrations, and blood biochemical values, liver and kidney gene expression profiles, and liver pathology specimens were obtained at multiple time points (Igarashi *et al*, 2015) . In the present, we used this database for further analysis.

We validated whether the rat-specific deconvolution could be applied to the liver gene expression profiles of Open TG-GATEs, which shows generalized performance of our approach to external and public databases. From the compounds in the database, we targeted compounds causing ALT elevation, which indicates liver toxicity, within 24 h and analyzed these data with the rat-specific deconvolution. Considering that there was no existing information regarding the relationships with the focused immune cells, four compounds were selected as candidates for validation. Colchicine and bortezomib were selected to verify whether the estimated NK and B cell fluctuations could be confirmed. In addition, galactosamine and thioacetamide were selected to compare whether the estimated fluctuations, including those in the early time of 3 h, could be confirmed (**Fig 3A**).

The candidate compounds were administered to rats under the same conditions as those of Open TG-GATEs, and the ratio of immune cells in the liver was measured by flow cytometry. Then, we compared the ratio of immune cells measured by flow cytometry and the estimated values of immune cells by rat-specific deconvolution. Although no clear correlation was observed for NK and B cells, a positive correlation was observed for neutrophils, monocytes, CD4 T cells, and CD8 T cells (**Fig 3B**). It is noteworthy that we confirmed consistency between the observed and estimated results for tiopronin treatment at 3 h: tiopronin induced a weak but significant increase in ALT but no change in the neutrophil ratio (**Fig EV4**). The tiopronin data here indicate that rat-specific deconvolution has the potential to detect these slight differences. One of the explanations for the difficulty in estimating NK and B cells is the possibility of large subpopulations of these cells in the liver. Hepatic NK cells have subpopulations that exhibit different profiles than circulating NK cells in the blood (Mikulak *et al*, 2019). Similarly, B cells have a liver-resident subpopulation that exhibits a specific response during inflammation (Curry *et al*, 2003; Racanelli *et al*, 2001). It is possible that these subsets with close, but different, profiles in the liver, affect the accuracy of deconvolution. Our flow cytometry showed that many NK and B cells were resident in the

liver specimens compared with the fewer other immune cells, which may support the impact of liver-resident NK and B cells.

We also examined the extrapolation using human and mouse-derived signature matrices obtained from existing reference profiles. Overall, there were cell types that became less predictable (**Fig EV5, Dataset EV1**). This result supports the previous conclusion and the importance of preparing rat-derived reference profiles for rat deconvolution.

These results suggest that it is possible to estimate and analyze immune cell trafficking except for B and NK cells in the liver from toxicogenomics databases, which are often organized with rats.

Identification of Clusters of Compounds That Have Different Immune Cell Trafficking

Although there are a few studies on immune cell trafficking as a response to a single compound or a few compounds, there are no reports comparing the responses to a large number of compounds. By applying the established deconvolution to existing large toxicogenomics databases, it is possible to gain aggregated knowledge of the differences in immune cell trafficking as biological responses to variable perturbations. As in the validation study, Open TG-GATEs database was subjected to deconvolution, followed by stratification, and analysis of biological phenomena.

Based on ALT elevation ($\geq$ 65), we targeted 16 compounds (**Fig EV6**, **Table 1**). For the four types of cells (neutrophils, monocytes, CD4 T cells, and CD8 T cells) that had been accurately estimated, as shown in **Fig 3**, the estimated scores of each cell at each time point were used as the features. After normalization, the compounds were mapped and visualized in two dimensions with meta-visualization method using 6 dimensional reduction methods as detailed in methods section (**Figs 4A, EV7**). The compounds were divided into four clusters based on hierarchical clustering (**Fig 4B**). Note that the result of this clustering is different from the result of clustering of the liver gene expression profiles (**Figs 4C, and D, EV8**). This indicates that the analysis based on immune cell trafficking captures a different perspective from that of gene expression profiles themselves and adds an interpretable new layer.

Characterization of the Compound Clusters

Do the clusters based on immune cell trafficking found in the data-driven analysis in **Fig 4** have any biological meaning? First, to gain insight into the behavior of immune cell trafficking and ALT as the index of liver injury for each cluster, we plotted the time course of their changes (**Fig 5A**). Note that the color filling around the line indicates the 95% confidence interval of all samples belonging to each cluster and depends on the number of

compounds. **Fig 5A** shows that compounds belonging to cluster 3 have a strong effect on immune cell trafficking. Cluster 2 is the second most effective cluster, and interestingly, the monocytes at 6 and 9 h after administration show more accumulation than those of cluster 3. Compounds belonging to clusters 1 and 4 have much weaker effects than those of clusters 2 and 3, and cluster 1 is slightly stronger than cluster 4. When these clusters were classified using random forest, ALT did not contribute significantly to the model, suggesting that these clusters do not simply indicate the degree of liver injury (**Dataset EV2**). Not surprisingly, because the clusters are different, these behaviors differ from those resulting from clustering based on the transcriptome (**Fig EV9**).

Cluster 3, the most effective cluster, contains two compounds, LPS, which is known as a strong inflammation inducer, and cycloheximide, a protein synthesis inhibitor targeting the translational process (Ferluga & Allison, 1978; Ledda-Columbano *et al*, 1992). Because the number of compounds in this cluster is small and easy to compare with existing knowledge, we focused on this cluster and evaluated its consistency with the existing biological knowledge. We calculated the gene ontology enrichment score back to the gene expression data of each compound to investigate the commonalities among the effects of the compounds. We searched for gene ontology terms that discriminate these compounds from those belonging to other clusters with statistical tests. As a result, the type I interferon signaling pathway was detected as a characteristic gene ontology term (**Fig 5B**). Although existing knowledge of the inflammatory response induced by cycloheximide is limited compared to that of LPS, a representative inflammation inducer, in a literature survey, cycloheximide was reported to stimulate the production of type I interferon RNA (Ringold *et al*, 1984).

Similarly, in cluster 4, to which tacrine and nitrofurazone belong, platelet aggregation was detected as a characteristic gene ontology term, particularly in the early time points (**Fig 5C**). Platelets are known to release signaling factors related to liver injury and regeneration, and their association with acute liver failure and liver disease has been reported (Morris & Chauhan, 2022) . Their importance has also been suggested in mouse models of drug-induced liver injury (Chauhan *et al*, 2020; Miyakawa *et al*, 2015) . A literature survey indicated that the two compounds inhibit platelet aggregation (Rossi & Levin, 1973; Slevin *et al*, 2018). In addition, the biochemical values of blood in the database showed a slight trend toward an increase in platelet concentration at 9 h compared with the other clusters (**Fig EV10**). The two compounds can influence liver injury by affecting pathways and mechanisms related to platelets.

Other characteristic other gene ontology terms for each cluster were found by statistical tests with conservative correction for multiplicity (**Fig EV11**, **Dataset EV3**).

Biological knowledge, extracted by a stratification analysis using estimated immune cell trafficking, could not be extracted by using the liver transcriptome alone (**Figs EV12 and 13**). In summary, the clusters obtained by immune cell trafficking are expected to have biological meanings that are consistent with existing biological knowledge, rather than overfitting the data.

DISCUSSION

Importance of establishing deconvolution method in rats

Although there have been studies that applied deconvolution to rat tissues, to our knowledge, there are no rat immune cell transcriptome datasets appropriate for reference profiles of deconvolution in public databases, and therefore, LM22, a human dataset, is often employed (Wang *et al*, 2021; Gil Del Alcazar *et al*, 2022). Here, we have confirmed that the use of mouse or human-derived reference profiles for deconvolution of rat specimens resulted in a decrease in accuracy for several immune cell types compared with rat-derived reference profiles, and that species differences did exist. This indicates that the rat-derived reference profiles obtained in this research are necessary for accurate deconvolution of rat specimens. Note that single-cell RNA-seq data of rat specimens, including rat liver, are currently available (Qaisar *et al*, 2021) . However, the characteristics of the data, which would affect the accuracy of the deconvolution, are different from the majority of accumulated legacy transcriptome datasets in public databases.

Deconvolution can now be applied to rats, for which large-scale toxicogenomics data are available in public databases, and work as a knowledge miner for immune cell trafficking in response to various perturbations achieved by compounds. In the present study, we tested data obtained from Open TG-GATEs because of its rich time-series data, while analysis of data from DrugMatrix, another large toxicogenomics database containing a larger number of compounds than Open TG-GATEs, is also an interesting target. By aggregating extracted knowledge, we could deepen our understanding of immune cell trafficking. Notably, rats are often used to assess the safety of drugs and chemicals. Thus, the combination of rat-specific deconvolution and toxicity databases would contribute not only to understanding the mechanisms of immune cell trafficking in liver injury but also to safety assessment.

Types of immune cells that can be estimated in rat deconvolution

One of the limitations of this study is the restricted number of immune cell types that can be estimated for the experimental and analytical reasons. First, knowledge of rat immunology is much lower than that of mice, and the established isolation protocols for rat immune-relative cells are restricted based on our preliminary literature survey. This is an interesting contrast to the field of toxicology, where rats are more often employed than mice. Second, as shown in **Fig 3**, it is difficult to estimate NK and B cells when rat-specific deconvolution is applied to the external data. In reference-based deconvolution, the signature matrix is created so that the co-correlation among immune cells to be analyzed is low. Although the v-support vector regression used in CIBERSORT and Elastic Net used in the present study are relatively resistant to multicollinearity compared with simple linear models, model

instability is inevitable in the presence of multicollinearity(Dormann *et al*, 2013). A possible solution to this problem is to incorporate the other cells belonging to the tissue of interest in the model. Notably, the importance of modeling considering tissue specificity in the deconvolution method has been discussed, although it has not been rigorously demonstrated. In this regard, the dataset we provide through this study could act as a dataset to evaluate deconvolution on liver tissue.

To increase the number of immune cell types for rat deconvolution, the utilization of scRNA-seq data is one idea (Schelker *et al*, 2017). However, care should be taken in using of scRNA-seq data for reference-based deconvolution because the shape of the scRNA-seq data differs from that of bulk transcriptome data accumulated in public databases. In this regard, unsupervised and semi-supervised deconvolution methods, which are independent or less dependent on the reference profile, respectively, and have been developed in recent years, are good options (Gaujoux & Seoighe, 2012; Dimitrakopoulou *et al*, 2018; Li & Wu, 2019; Tang *et al*, 2020). Comparison of deconvolution method paradigms is a future work for utilizing large toxicogenomics databases as data sources for immune cell trafficking (Avila Cobos *et al*, 2020).

Contamination of blood into the liver transcriptome

Blood contamination is one of the important discussion topics in the utilization of deconvolution for knowledge miner for immune cell trafficking from public databases. We measured the immune responses and transcriptomes using samples collected after liver perfusion. By contrast, the transcriptomes in the database were only collected after blood was released from the inferior vena cava. This may have resulted in contamination of the samples with blood-derived immune cells, and these differences may have affected the accuracy comparisons. In fact, in some samples, the number of neutrophils in the blood in the database is proportional to the score estimated by deconvolution (**Fig EV14**). However, it is difficult to determine whether this is the actual behavior of immune cells in the liver or a significant blood influence.

How can immune cells infiltrating tissues and blood-derived immune cells be distinguished and evaluated? Intravascular staining method is a possible experimental solution to this problem (Anderson *et al*, 2014). In this method, staining discriminates between tissue-localized cells and blood-born cells by intravenous injection of staining antibodies. However, the liver differs from organs such as the lungs, where the intravascular staining method is generally applied, in that it has fenestrae, and care should be taken when applying intravascular staining method (Gracia-Sancho *et al*, 2021). Anderson et al referred to the specific character of CD69+ cells regarding the relationship between the liver and blood. In

addition, there are few reports about applying intravascular staining method in studies of tissue injury (Anderson *et al*, 2014). The vascular wall discriminating hepatocytes and blood flow is disrupted in liver injury, and whether intravascular staining method can be applied to investigate liver injury remains to be determined (DeLeve *et al*, 1999). It is important to grasp the biological background of the source specimens when using legacy data because it directly affects the interpretation of the data analysis outcomes. Thus, it is essential to understand the relationship between the immune cells infiltrating in the liver and those from blood with a liver injury background.

MATERIALS AND METHODS

Data Preparation

- TG-GATEs dataset

  All raw files were downloaded from the home page of Open TG-GATEs (https://toxico.nibiohn.go.jp) (Igarashi *et al*, 2015). Raw signal intensities for each probe set as they are contained in the CEL files, were analyzed using MAS5.0, which is implemented in the software package Bioconductor (http://bioconductor.org) with R (version 3.6.3).

- Rat RNA sequencing data

  FASTQ files of the rat RNA sequencing data were prepared according to the RNA isolation and sequencing sections. Quality control of all reads was performed using PINSEQ++ (version 1.2.4) with the indicated parameters (trim_left = 5, trim_tail_right = 5, trim_qual_right = 30, ns_max_n = 20, min_len = 30) (https://peerj.com/preprints/27553/). The expression of transcripts was quantified using Salmon (version 1.6.0) with the indicated parameters (validationMappings, gcBias, seqBias) and a decoy-aware index created using Salmon and Rattus_norvegicus.Rnor_6.0.cdna.all was obtained from EMBL-EBI (https://www.ebi.ac.uk/) (Patro *et al*, 2017). Transcripts per kilobase million data were obtained using tximport, which is implemented in the software package Bioconductor with R (version 4.1.3) from quant.sh files created by Salmon.

Animals

Male Sprague Dawley rats (6 weeks old) were purchased from CLEA Japan (Tokyo, Japan) for RNA sequencing experiments or Oriental Yeast Co. (Tokyo, Japan) to validate the immune response in the present study. All animals were maintained under standard conditions with a reverse dark-light cycle and were treated humanely. Food and water were available ad libitum.

The procesures reported in this article were performed in accordance with the guidelines provided by the Institutional Animal Care Committee (Graduate School of Pharmaceutical Sciences, the University of Tokyo, Tokyo, Japan).

Rat Liver Injury Model

- RNA-sequence

  To induce liver injury, acetaminophen (H0190, Tokyo Chemical Industry Co., Tokyo, Japan) or 4,4′-methylene dianiline (M0220, Tokyo Chemical Industry Co.) was administered at 3 g/kg or 300 mg/kg, respectively by gavage. As a vehicle control group, 0.5 (w/v)% methylcellulose 400 solution (133-17815, Fujifilm Wako Pure Chemical Corporation, Osaka, Japan) was given at 10 μL/g. Only those rats in the experimental groups that were subjected to RNA sequencing had fasted between 21:00 and 9:00, which was the administration time.

- Open TG-GATEs Validation Study

  To validate the immune responses in the liver, metformin at 1 g/kg by gavage (M2009, Tokyo Chemical Industry Co.), tiopronin at 1 g/kg by gavage (T2614, Tokyo Chemical Industry Co.), galactosamine at 1 g/kg by gavage (G0007, Tokyo Chemical Industry Co.), thioacetamide at 45 mg/kg by gavage (T0187, Tokyo Chemical Industry Co.), colchicine at 15 g/kg by gavage (C0380, Tokyo Chemical Industry Co.), or bortezomib at 0.6 mg/kg by i.v. through a tail vein (B5741, Tokyo Chemical Industry Co.) was administered. As a vehicle control group, 0.5 (w/v)% methylcellulose 400 solution at 10 μL/g by gavage or saline at 2.5 mL/kg by i.v. was given. In this validation experiment, no fasting was performed following the Open TG-GATE protocol.

Sample Collection and Preparation

- Liver

  Rats were euthanized and the liver and blood were harvested after compound administration. Briefly, the superior vena cava was clipped using a clamp, and the blood was collected through an inferior vena cava into a 1.5-mL tube containing 1 μL heparin (Yoshindo, Toyama, Japan). The collected blood sample was centrifuged for 15 min at 1,700 g, and the supernatant was subjected to Fuji Drychem NX500sV (Fujifilm Corporation, Tokyo, Japan) to quantify the concentrations of alanine aminotransferase, aspartate aminotransferase, and total bilirubin. The liver perfusion was performed through an inferior vena cava using a plastic cannula-type puncture needle (SR-OT1832C, 18G 32 mm) (Terumo corporation, Tokyo, Japan) and 5 mM HEPES (345-06681, Fujifilm Wako Pure Chemical Corporation) /5 mM EDTA (6381-92-6, Dojindo

Laboratories, Kumamoto, Japan) Hanks' balanced salt solution (17461-05, Nacalai Tesque, Kyoto, Japan). Finally, 1/9 of the liver's largest lobe or 1/6 of all minced liver was subjected to RNA isolation or flow cytometry, respectively. For flow cytometry, the sample was dissociated using gentleMACS (Miltenyi Biotec, North Rhine-Westphalia, Germany) according to the manufacturer's instructions. Except where noted, phosphate-buffered saline containing 2% fetal bovine serum was used as "wash buffer" thereafter. After using wash buffer to eliminate hepatocytes, the sample was centrifuged at 50 g at 4 ° C for 3 min and the pellets were discarded. To eliminate the hepatic erythrocytes, 5 mL of ACK buffer was added to the sample, and the sample was incubated in a 37 °C water bath for 6 min. ACK buffer was prepared by adding 8,024 mg of NH_4Cl (A2037, Tokyo Chemical Industry Co.), 10 mg of $NHCO_3$ (166-03275, Fujifilm Wako Pure Chemical Corporation), and 3.772 mg of EDTA $2Na2H_2O$ (6381-92-6, Dojindo Laboratories) into 1 L of pure water. The sample was washed with the wash buffer three times, and then the sample was subjected to flow cytometry.

- Spleen

  An untreated was placed under anesthesia, and blood was collected from the heart by cardiopuncture. The blood was washed once with wash buffer once. After removing the supernatant, 15 mL of ACK buffer was added to the sample and then incubated in a 37 °C water bath for 6 min to eliminate erythrocytes. After incubation, 25 mL of the wash buffer was added to the sample, which was then centrifuged at 400 g at 4 °C for 5 min. The incubation and wash operations were repeated twice. After the last wash, the sample was washed with the wash buffer three times, and then subjected to RNA isolation and flow cytometry or cell sorting.

- Blood

  An untreated rat was euthanized, and the spleen was collected. Single-cell suspensions from the spleen were obtained by gently forcing the tissues through a 70 μm nylon mesh grid in RPMI-1640 (06261-65, Nacalai Tesque). After washing with the wash buffer, 15 mL of ACK buffer was added into the sample, and then incubated in a 37 °C water bath for 6 min to eliminate the splenic erythrocytes. The sample was washed with the wash buffer three times, and then subjected to RNA isolation and flow cytometry or cell sorting.

RNA Isolation and RNA Sequencing

Total RNA was prepared using ISOGEN II (311-07361, Nippon Gene Co., Tokyo, Japan) and purified using an RNeasy Plus Mini Kit (74136, Qiagen, Limburg, Netherland) with gDNA elimination using an RNase-Free DNase set (79254, Qiagen) according to the manufacturer's protocols. RNA was used to prepare RNA-Seq libraries with a NEBNext Ultra

II Directional RNA Library Prep Kit for Illumina (E7760L, New England Biolabs, MA, USA). The libraries were sequenced for single-end reads using a Novaseq 6000 (Illumina, CA, USA). Sequence reads from each cDNA library were processed as described in the Data Preparation section.

Flow Cytometry

Flow cytometric analysis was performed with FACSAria III (BD Biosciences, NJ, USA), and data were analyzed with FlowJo software (BD Biosciences). The antibodies used are listed in **Table 2**.

Cell sorting

Cell sorting was performed with a FACSAria III (BD Biosciences, NJ, USA). CD4 T cells (CD45+/CD11b-/CD3+/CD4+), CD8 T cells (CD45+/CD11b-/CD3+/CD8a+), NK cells (CD45+/CD11b-/CD3-/CD161+), and B cells (CD45+/CD11b-/CD3-/CD45RA+) were obtained from the rat spleen and monocytes (CD45+/CD11b+/CD3-/CD45RA-/CD161-/SSC-low) and neutrophils (CD45+/CD11b+/CD3-/CD45RA-/CD161-/CD43int/SSC-high) were obtained from the rat blood. The antibodies used are listed in **Table 2**. Cells were suspended in wash buffer containing 5% 7-AAD solution (559925, BD Pharmgen) 5 min before analysis to exclude dead cells. The sorted cells were at once pooled in RPMI-1640 containing 10% fetal bovine serum. After centrifugation at 400 g at 4 °C for 5 min, RNA was isolated and sequenced from the cells in the pellets collected.

Deconvolution

We converted transcript IDs to RGD gene symbols, and median values were selected for duplicate gene names. Only genes in common with the bulk tissue gene expression matrix were retained. Among the retained genes, up to 200 genes were retained as markers with an absolute fold-change > 1.5 for other cell types. Finally, N genes collected from K cell markers defined above were included in the analysis, and the cell type-specific expression matrix $\boldsymbol{X} \in \boldsymbol{R}^{N\times K}$ was defined. Consider a measured bulk gene expression matrix $\boldsymbol{Y} \in \boldsymbol{R}^{N\times M}$ for N genes across M samples, each containing K different cell types. The goal of deconvolution is to estimate cell type-specific expression $\boldsymbol{X} \in \boldsymbol{R}^{N\times K}$ and cell type-specific proportion $\boldsymbol{P} \in \boldsymbol{R}^{K\times M}$, and can be written as:

$$\boldsymbol{Y} \approx \boldsymbol{XP}. \tag{1}$$

Elastic Net is a regularized regression model with combined L1 and L2 penalties. We can estimate the cell type-specific expression $\widehat{\boldsymbol{P}}$ via:

$$\widehat{\boldsymbol{P}} = \underset{\boldsymbol{P}}{\arg\min}\ \{\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{P}\|_2^2 + \lambda \sum_{j=1}^{M} \left[\frac{1}{2}(1-\alpha)\boldsymbol{P}_j^2 + \alpha\left|\boldsymbol{P}_j\right|\right]\}, \quad (2)$$

where λ and α are hyperparameters, and we set $\lambda = 1$ and $\alpha = 0.05$ which are based on published paper using Elastic Net for deconvolution (Altboum *et al*, 2014). The essential codes are available in GitHub (https://github.com/mizuno-group/Deconv). Analysis was performed using Python (version 3.9.12) with open-source packages (numpy, version 1.22.4; pandas, version 1.1.5; scipy, version 1.8.0; scikit-learn, version 1.0.2; statsmodels, version 0.13.2; umap, 0.5.3)

Other Species' Signature Matrices

We used LM6 from the CIBERSORTx home page (https://cibersortx.stanford.edu/) as the human-derived signature matrix (Chen *et al*, 2018a). For the mouse-derived signature matrix, we collected mouse immune cell transcriptomes from the public database and summarized them to create reference profiles. To convert the gene names, the correspondence tables for rat gene symbols (RGD) to human gene symbols (HGNC) and mouse gene symbols (MGI) were downloaded from the BioMart Ensemble database (http://asia.ensembl.org/index.html). Genes that corresponded to RGD genes that do not exist were omitted from subsequent analysis. The locations and immune cell types of the data are listed in the **Dataset EV4**.

Single Sample Gene Set Enrichment Analysis

The GO annotation file (rgd.gaf) and GO tree structure (go.owl) were downloaded from the GO Consortium website (http://geneontology.org). The ontology at the fifth depth from the top of the tree structure and the corresponding gene set were extracted and used for the following analysis. We subjected RNA sequence data to single-sample gene enrichment analysis (ssGSEA) with following the ssGSEA algorithms (https://github.com/broadinstitute/ssGSEA2.0). Note that the gene expression profiles of the Open-TGGATEs database have approximately three duplicate samples for each administration condition. Figures about the ssGSEA include all those duplicates.

Clustering Analysis

Dimensionality reduction and clustering were performed using immune cell trafficking or transcriptome of the liver at 3, 6, 9, and 24 h after administration as features. All features were reduced in dimensionality by locally linear embedding, multidimensional scaling, spectral embedding, principal component analysis, t distributed stochastic neighbor embedding, and uniform manifold approximation and projection. The dimensionally reduced features were combined and visualized using a meta-visualization method (Ma *et al*, 2023).

The combined meta-distance matrix was subjected to hierarchical clustering and visualized with a heat map. Note that only CD4 T cells, CD8 T cells, monocytes, and neutrophils were used for immune cell trafficking features, and all features were converted to a *z*-score for the corresponding control samples. Additionally, note that transcriptome features were log-transformed and converted to a *z*-score for each compound feature. The detailed code used is found on GitHub (https://github.com/mizuno-group/RatDeconvolution). Analysis was performed using Python (version 3.9.12) with open-source packages (numpy, version 1.22.4; pandas, version 1.1.5; scipy, version 1.8.0; scikit-learn, version 1.0.2; statsmodels, version 0.13.2; umap, 0.5.3)

Authors Contributions

Katsuhisa Morita: Data curation, Formal analysis, Methodology, Software, Investigation, Writing – Original Draft, Visualization.

Tadahaya Mizuno: Conceptualization, Data curation, Investigation, Resources, Supervision, Project administration, Writing – Original Draft, Writing – Review & Editing, Funding acquisition.

Iori Azuma: Data curation, Investigation.

Hiroyuki Kusuhara: Writing – Review & Editing

Correspondence

Tadahaya Mizuno

Availability of data and materials

Code, models, and data are available at https://github.com/mizuno-group/RatDeconvolution. RNA-seq data will be shown in GEO datasets.

Financial Support

This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) (grant number 21K06663) from the Japan Society for the Promotion of Science, JSPS KAKENHI Grant Number 16H06279 (PAGS), and Takeda Science Foundation.

Acknowledgments

We appreciate Dr. Hiroyuki Kusuhara and Dr. Hisamitsu Hayashi for helpful discussion about this study. We thank all those who contributed to the construction of the datasets employed in the present study such as LM6, Open TG-GATEs, and so on.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

REFERENCES

Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z & Clark HF (2009) Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One* 4: e6098

Altboum Z, Steuerman Y, David E, Barnett-Itzhaki Z, Valadarsky L, Keren-Shaul H, Meningher T, Mendelson E, Mandelboim M, Gat-Viks I, *et al* (2014) Digital cell quantification identifies global immune cell dynamics during influenza infection. *Mol Syst Biol* 10: 720

Anderson KG, Mayer-Barber K, Sung H, Beura L, James BR, Taylor JJ, Qunaj L, Griffith TS, Vezys V, Barber DL, *et al* (2014) Intravascular staining for discrimination of vascular and tissue leukocytes. *Nat Protoc* 9: 209–22

Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P & De Preter K (2020) Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* 11: 5650

Bailie MB, Hewett JA, Schultze AE & Roth RA (1994) Methylene dianiline hepatotoxicity is not leukocyte-dependent. *Toxicol Appl Pharmacol* 124: 25–30

Cai S-Y, Ouyang X, Chen Y, Soroka CJ, Wang J, Mennone A, Wang Y, Mehal WZ, Jain D & Boyer JL (2017) Bile acids initiate cholestatic liver injury by triggering a hepatocyte-specific inflammatory response. *JCI insight* 2: e90780

Chauhan A, Sheriff L, Hussain MT, Webb GJ, Patten DA, Shepherd EL, Shaw R, Weston CJ, Haldar D, Bourke S, *et al* (2020) The platelet receptor CLEC-2 blocks neutrophil mediated hepatic recovery in acetaminophen induced acute liver failure. *Nat Commun* 11: 1939

Chen B, Khodadoust MS, Liu CL, Newman AM & Alizadeh AA (2018a) Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* 1711: 243–259

Chen S, Hoffman RA, Scott M, Manson J, Loughran P, Ramadan M, Demetris AJ & Billiar TR (2017) NK1.1+ cells promote sustained tissue injury and inflammation after trauma with hemorrhagic shock. *J Leukoc Biol* 102: 127–134

Chen Z, Quan L, Huang A, Zhao Q, Yuan Y, Yuan X, Shen Q, Shang J, Ben Y, Qin FX-F, *et al* (2018b) seq-ImmuCC: Cell-Centric View of Tissue Transcriptome Measuring Cellular Compositions of Immune Microenvironment From Mouse RNA-Seq Data. *Front Immunol* 9: 1286

Curry MP, Golden-Mason L, Doherty DG, Deignan T, Norris S, Duffy M, Nolan N, Hall W, Hegarty JE & O'Farrelly C (2003) Expansion of innate CD5pos B cells expressing high levels of CD81 in hepatitis C virus infected liver. *J Hepatol* 38: 642–50

Davies LC, Jenkins SJ, Allen JE & Taylor PR (2013) Tissue-resident macrophages. *Nat Immunol* 14: 986–95

DeLeve LD, McCuskey RS, Wang X, Hu L, McCuskey MK, Epstein RB & Kanel GC (1999)

Characterization of a reproducible rat model of hepatic veno-occlusive disease. *Hepatology* 29: 1779–91

Dimitrakopoulou K, Wik E, Akslen LA & Jonassen I (2018) Deblender: a semi-/unsupervised multi-operational computational method for complete deconvolution of expression data from heterogeneous samples. *BMC Bioinformatics* 19: 408

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, *et al* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop)* 36: 27–46

Ferluga J & Allison AC (1978) Role of mononuclear infiltrating cells in pathogenesis of hepatitis. *Lancet (London, England)* 2: 610–1

Gaujoux R & Seoighe C (2012) Semi-supervised Nonnegative Matrix Factorization for gene expression deconvolution: a case study. *Infect Genet Evol* 12: 913–21

Gehring S, Dickson EM, San Martin ME, van Rooijen N, Papa EF, Harty MW, Tracy TF & Gregory SH (2006) Kupffer cells abrogate cholestatic liver injury in mice. *Gastroenterology* 130: 810–22

Gil Del Alcazar CR, Trinh A, Alečković M, Rojas Jimenez E, Harper NW, Oliphant MUJ, Xie S, Krop ED, Lulseged B, Murphy KC, *et al* (2022) Insights into Immune Escape During Tumor Evolution and Response to Immunotherapy Using a Rat Model of Breast Cancer. *Cancer Immunol Res* 10: 680–697

Gracia-Sancho J, Caparrós E, Fernández-Iglesias A & Francés R (2021) Role of liver sinusoidal endothelial cells in liver diseases. *Nat Rev Gastroenterol Hepatol* 18: 411–431

Graubardt N, Vugman M, Mouhadeb O, Caliari G, Pasmanik-Chor M, Reuveni D, Zigmond E, Brazowski E, David E, Chappell-Maor L, *et al* (2017) Ly6Chi Monocytes and Their Macrophage Descendants Regulate Neutrophil Function and Clearance in Acetaminophen-Induced Liver Injury. *Front Immunol* 8: 626

Hossain M & Kubes P (2019) Innate immune cells orchestrate the repair of sterile injury in the liver and beyond. *Eur J Immunol* 49: 831–841

Igarashi Y, Nakatsu N, Yamashita T, Ono A, Ohno Y, Urushidani T & Yamada H (2015) Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res* 43: D921-7

Im Y & Kim Y (2023) A Comprehensive Overview of RNA Deconvolution Methods and Their Application. *Mol Cells* 46: 99–105

Ingersoll MA, Platt AM, Potteaux S & Randolph GJ (2011) Monocyte trafficking in acute and chronic inflammation. *Trends Immunol* 32: 470–7

Jaeschke H, Williams CD, Ramachandran A & Bajt ML (2012) Acetaminophen hepatotoxicity and repair: the role of sterile inflammation and innate immunity. *Liver Int* 32: 8–20

Laskin DL, Gardner CR, Price VF & Jollow DJ (1995) Modulation of macrophage functioning

abrogates the acute hepatotoxicity of acetaminophen. *Hepatology* 21: 1045–50

Ledda-Columbano GM, Coni P, Faa G, Manenti G & Columbano A (1992) Rapid induction of apoptosis in rat liver by cycloheximide. *Am J Pathol* 140: 545–9

Li Z & Wu H (2019) TOAST: improving reference-free cell composition estimation by cross-cell type differential analysis. *Genome Biol* 20: 190

Liu Z-X, Govindarajan S & Kaplowitz N (2004) Innate immune system plays a critical role in determining the progression and severity of acetaminophen hepatotoxicity. *Gastroenterology* 127: 1760–74

Liu Z-X, Han D, Gunawan B & Kaplowitz N (2006) Neutrophil depletion protects against murine acetaminophen hepatotoxicity. *Hepatology* 43: 1220–30

Ma R, Sun ED & Zou J (2023) A spectral method for assessing and combining multiple data visualizations. *Nat Commun* 14: 780

Mikulak J, Bruni E, Oriolo F, Di Vito C & Mavilio D (2019) Hepatic Natural Killer Cells: Organ-Specific Sentinels of Liver Immune Homeostasis and Physiopathology. *Front Immunol* 10: 946

Mitchell JR, Jollow DJ, Potter WZ, Davis DC, Gillette JR & Brodie BB (1973) Acetaminophen-induced hepatic necrosis. I. Role of drug metabolism. *J Pharmacol Exp Ther* 187: 185–94

Miyakawa K, Joshi N, Sullivan BP, Albee R, Brandenberger C, Jaeschke H, McGill MR, Scott MA, Ganey PE, Luyendyk JP, *et al* (2015) Platelets and protease-activated receptor-4 contribute to acetaminophen-induced liver injury in mice. *Blood* 126: 1835–43

Morita K, Mizuno T & Kusuhara H (2020) Decomposition profile data analysis of multiple drug effects identifies endoplasmic reticulum stress-inducing ability as an unrecognized factor. *Sci Rep* 10: 13139

Morris SM & Chauhan A (2022) The role of platelet mediated thromboinflammation in acute liver injury. *Front Immunol* 13: 1037645

Nemoto S, Morita K, Mizuno T & Kusuhara H (2021) Decomposition Profile Data Analysis for Deep Understanding of Multiple Effects of Natural Products. *J Nat Prod* 84: 1283–1293

Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419

Petitprez F, Levy S, Sun C-M, Meylan M, Linhard C, Becht E, Elarouci N, Tavel D, Roumenina LT, Ayadi M, *et al* (2020) The murine Microenvironment Cell Population counter method to estimate abundance of tissue-infiltrating immune and stromal cell populations in murine samples using gene expression. *Genome Med* 12: 86

Qaisar N, Arowosegbe A, Derr AG, Kucukural A, Satish B, Racicot R, Guo Z, Trombly MI & Wang JP (2021) Type I IFN-Driven Immune Cell Dysregulation in Rat Autoimmune Diabetes. *ImmunoHorizons* 5: 855–869

Racanelli V, Sansonno D, Piccoli C, D'Amore FP, Tucci FA & Dammacco F (2001) Molecular characterization of B cell clonal expansions in the liver of chronically hepatitis C virus-infected patients. *J Immunol* 167: 21–9

Ringold GM, Dieckmann B, Vannice JL, Trahey M & McCormick F (1984) Inhibition of protein synthesis stimulates the transcription of human beta-interferon genes in Chinese hamster ovary cells. *Proc Natl Acad Sci U S A* 81: 3964–8

Rossi EC & Levin NW (1973) Inhibition of primary ADP-induced platelet aggregation in normal subjects after administration of nitrofurantoin (furadantin). *J Clin Invest* 52: 2457–67

Schelker M, Feau S, Du J, Ranu N, Klipp E, MacBeath G, Schoeberl B & Raue A (2017) Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat Commun* 8: 1–12

Slevin M, Iemma RS, Zeinolabediny Y, Liu D, Ferris GR, Caprio V, Phillips N, Di Napoli M, Guo B, Zeng X, *et al* (2018) Acetylcholine Inhibits Monomeric C-Reactive Protein Induced Inflammation, Endothelial Cell Adhesion, and Platelet Aggregation; A Potential Therapeutic? *Front Immunol* 9: 2124

Tang D, Park S & Zhao H (2020) NITUMID: Nonnegative matrix factorization-based Immune-TUmor MIcroenvironment Deconvolution. *Bioinformatics* 36: 1344–1350

Wang J, Kumar S & Liu G (2021) Bulk Gene Expression Deconvolution Reveals Infiltration of M2 Macrophages in Retinal Neovascularization. *Invest Ophthalmol Vis Sci* 62: 22

Yang R & Tonnesseen TI (2019) DAMPs and sterile inflammation in drug hepatotoxicity. *Hepatol Int* 13: 42–50

Yang W, Tao Y, Wu Y, Zhao X, Ye W, Zhao D, Fu L, Tian C, Yang J, He F, *et al* (2019) Neutrophils promote the development of reparative macrophages mediated by ROS to orchestrate liver repair. *Nat Commun* 10: 1076

FIGURE LEGENDS

Figure 1

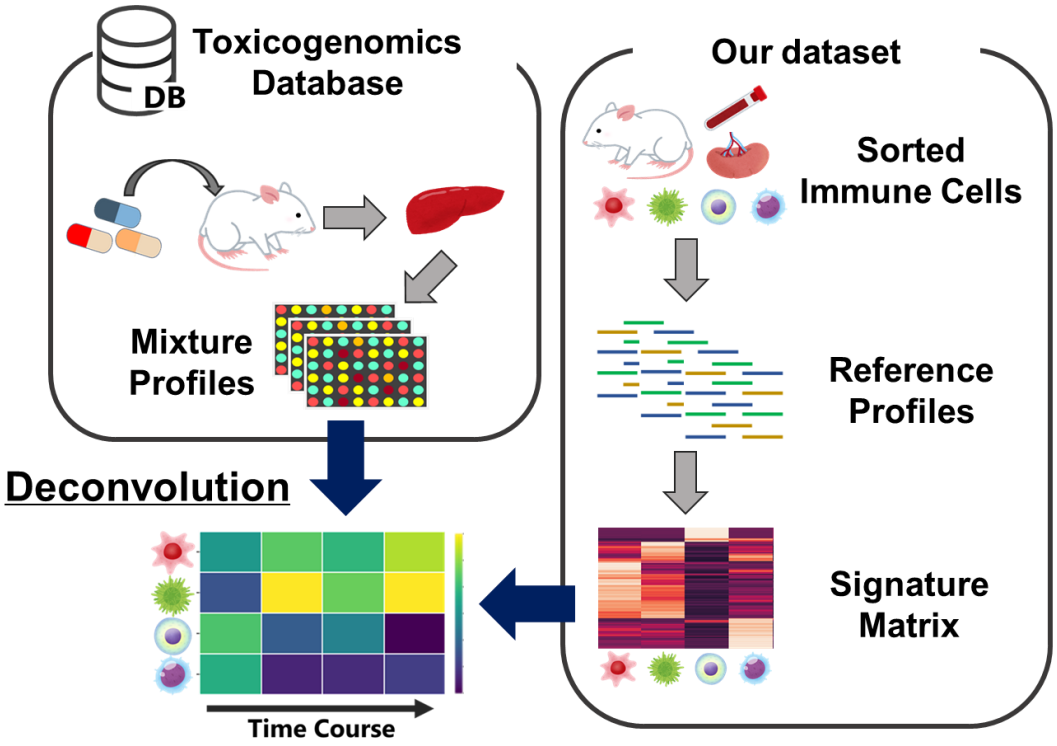

Figure 1. Overview of the present study. The rat-specific ”signature matrix” comprising marker genes that are enriched in each immune cell type was created from our sorted immune cell transcriptome (“reference profiles”). Deconvolution using the prepared signature matrix was applied to the rat liver transcriptome (“mixture profiles”), and estimated immune cell trafficking was subjected to validation and stratification.

Figure 2

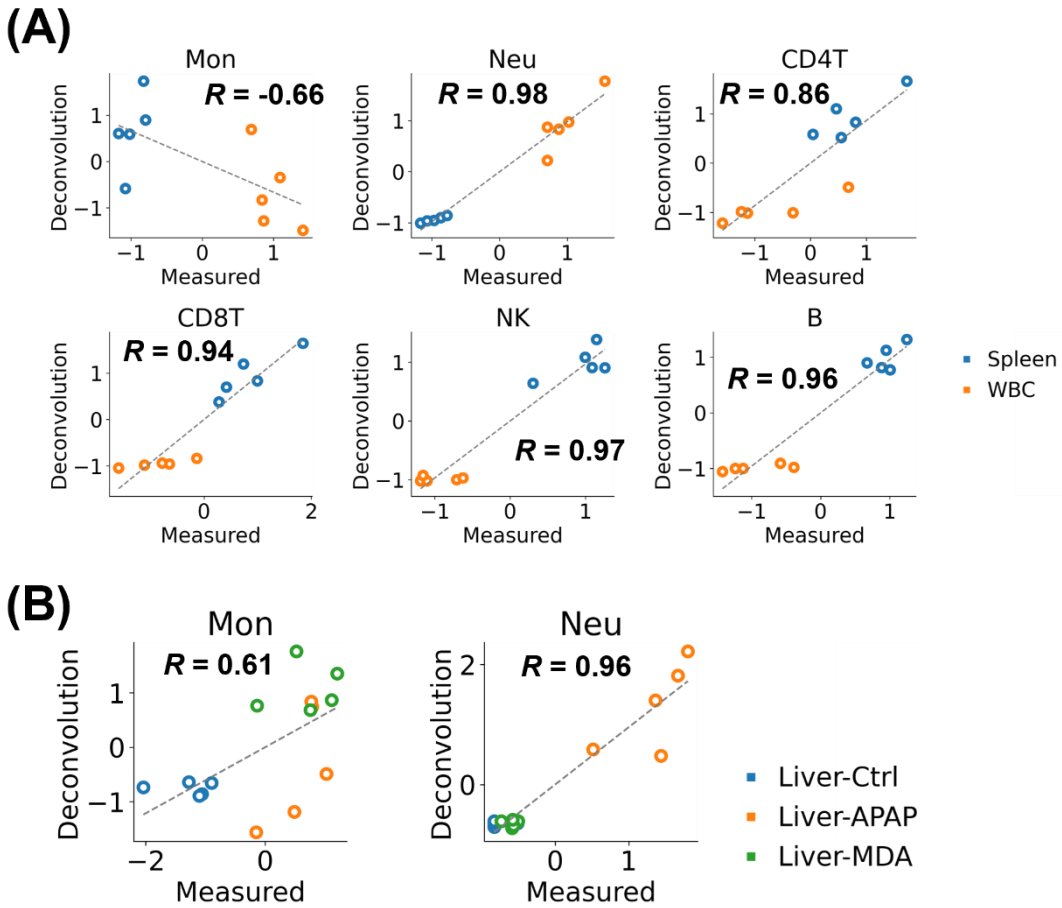

Figure 2. Comparison of estimated and measured immune cell trafficking in rat tissues. (A) Comparison in rat spleen and whole blood cells (WBCs). (B) Comparison using the compounds administered rat liver. Values were measured by FACS after cell isolation and red blood cell lysis. Estimated values were calculated by deconvolution using a rat-specific signature matrix. Note that these values were converted to *z*-score between the compared samples. Ctrl, control; APAP, acetaminophen; MDA, methylene dianiline.

Figure 3

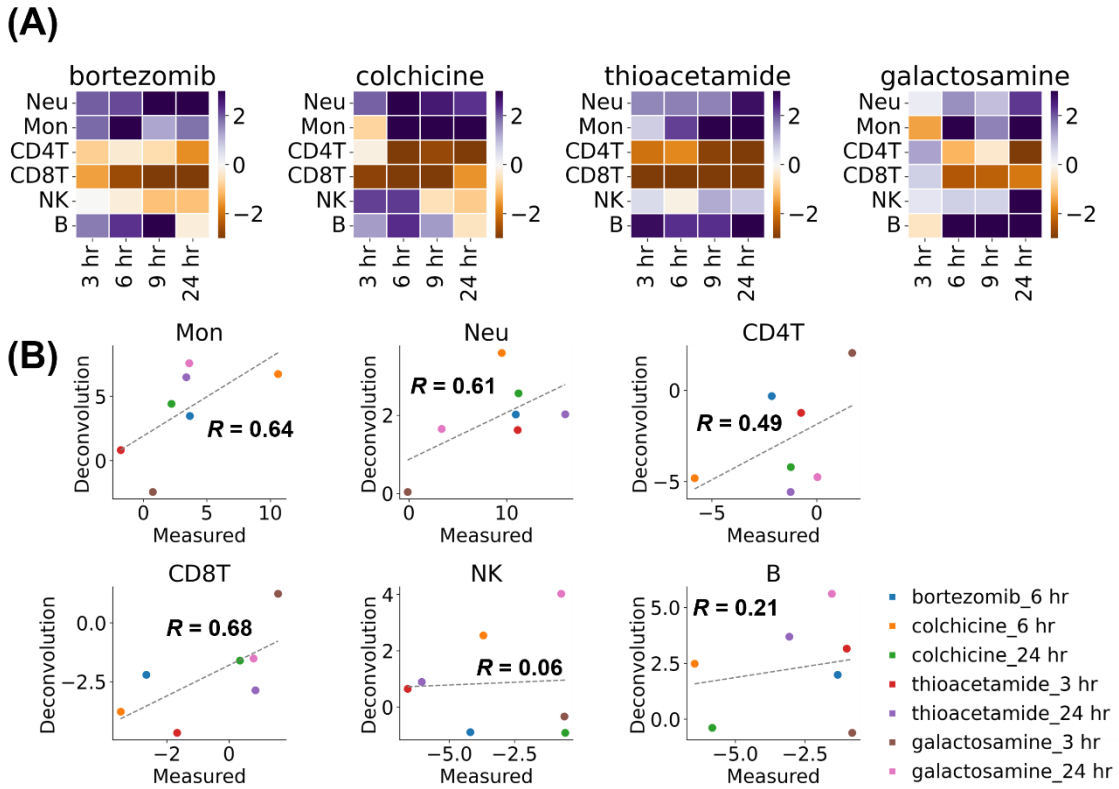

Figure 3. (A) Heatmaps of estimated immune cell trafficking in the liver. Estimated values at 3, 6, 9, and 24 h after administration of each compound used in the validation are shown. Note that these values are converted to *z*-score for the corresponding control samples in the Open TG-GATEs database or obtained by ourselves. (B) Comparison of estimated and measured immune cell trafficking in rat liver after administering each compound. Note that these values are converted to *z*-scores between the compared samples. Note that each point is the median value in the experiment or in the Open-TGGATEs database (n = 3~4).

Figure 4

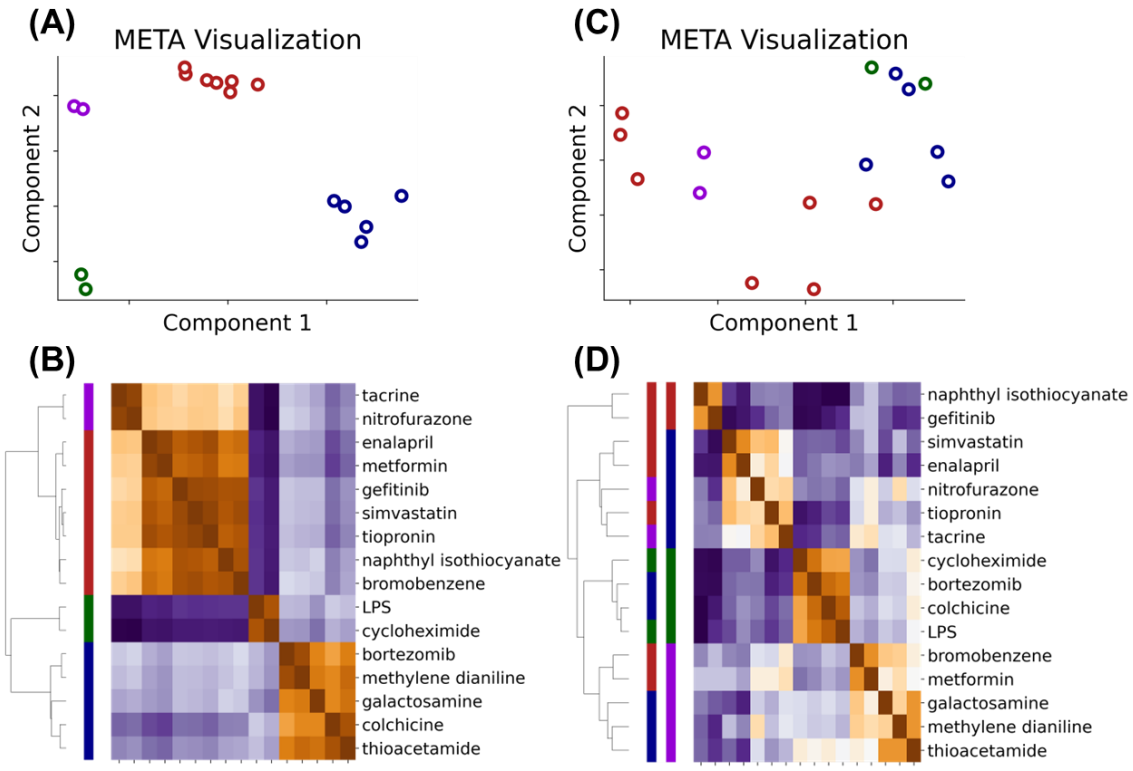

Figure 4. Clustering analysis of 16 compounds that induced significant ALT elevation in the Open TG-GATEs database. Dimensionality reduction and clustering were performed using (A), (B) immune cell trafficking (C), and (D) transcriptome of rat liver at 3, 6, 9, and 24 hours after administration. All features were reduced in dimensionality by locally linear embedding, multidimensional scaling, spectral embedding, principal component analysis, t distributed stochastic neighbor embedding, and uniform manifold approximation and projection. (A), (C) The dimensionally reduced features were combined and visualized by the meta-visualization method. (B), (D) The combined meta-distance matrix was subjected to hierarchical clustering and visualized with a heatmap. The color bar closest to the heatmap corresponds to the stratified compound clusters of the heatmap, and the leftmost color bar in (D) is the color corresponding to the compound clusters stratified by (B). Note that only CD4 T cells, CD8 T cells, monocytes, and neutrophils were used for immune cell trafficking features and all features were converted to *z*-score for the corresponding control samples. Note also that transcriptome features were log-transformed and converted to *z*-score between each compound feature.

Figure 5

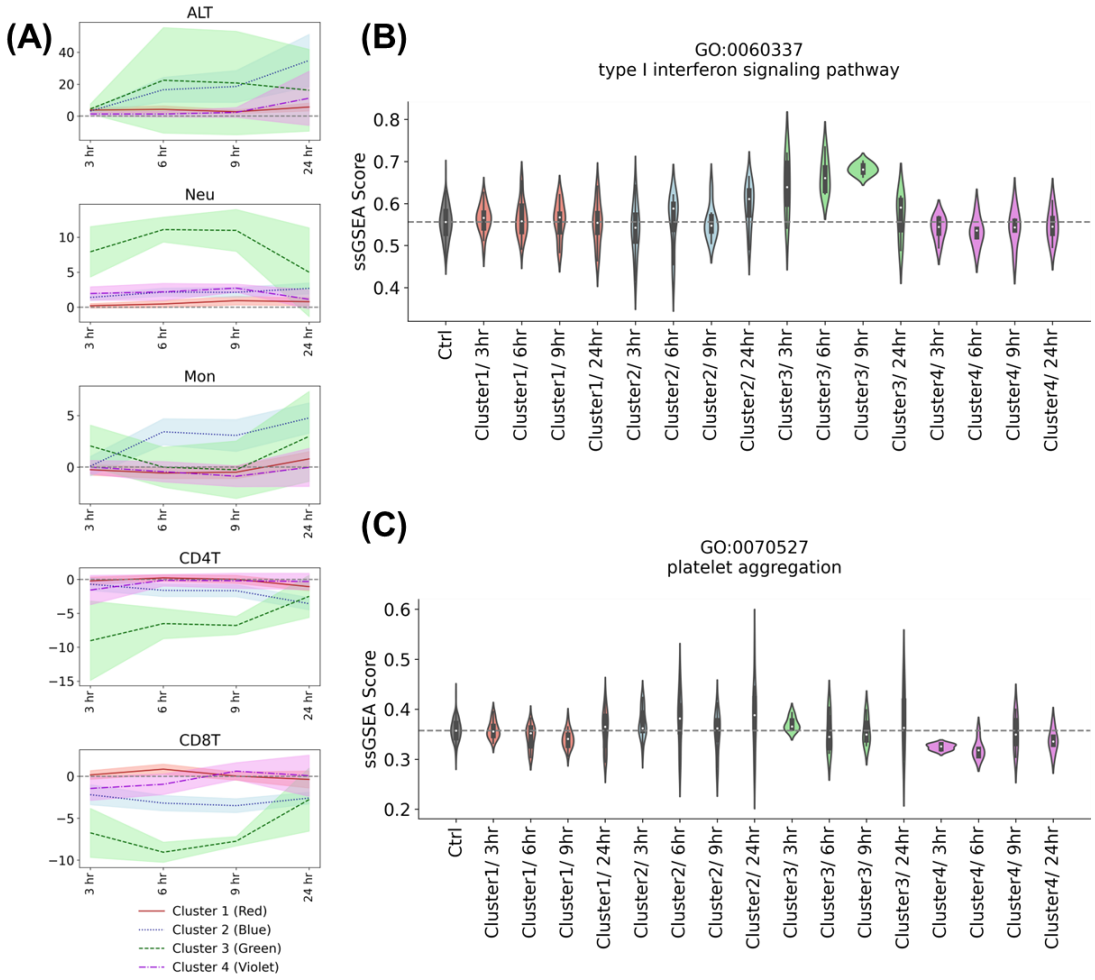

Figure 5. (A) Time course of changes in immune cell trafficking and ALT in stratified compound clusters. The line is the average value for each cluster, and the filled color around it indicates the 95% confidence interval. Each color corresponds to the color of the compound stratified by analysis of immune cell trafficking. Note that the scale of the x-axis does not correspond to the actual time values. (B), (C) Time course of single-sample gene set enrichment analysis (ssGSEA) scores of the liver transcriptome. Gene ontology terms of (B) type I interferon signaling pathway and (C) platelet aggregation.

TABLES and THEIR LEGENDS

Table 1

Compound used in stratification analysis. Cluster numbers stratified by analysis are attached for immune-based and transcript-based.

Compound Name	Immune Cluster	Transcript Cluster
Tacrine	4	2
Nitrofurazone	4	2
Enalapril	1	2
Metformin	1	4
Gefitinib	1	1
Simvastatin	1	2
Tiopronin	1	2
Naphthyl Isothiocyanate	1	1
Bromobenzene	1	3
LPS	3	3
Cycloheximide	3	3
Bortezomib	2	4
Methylene dianiline	2	4
Galactosamine	2	4
Colchicine	2	3
Thioacetamide	2	4

Table 2

List of antibodies and staining reagents used in sorting and flow cytometry of immune cells.

Marker	Detection	Cat. No.	Conjugate	Clone	Isotype
CD11b	FITC	561684	FITC	WT.5	Mouse IgA, κ
CD45	APC-Cy7	561586	APC-Cy7	OX-1	Mouse IgG1, κ
CD45RA	BV421	740043	BV421	OX-33	Mouse IgG1, κ
CD3	BV605	563949	BV605	1F4	Mouse IgM, κ
CD4	BV786	740912	BV786	OX-35	Mouse IgG2a, κ
CD8	BV711	740724	BV711	OX-8	Mouse IgG1, κ
CD43	PE	202812	PE	W3/13	Mouse IgG1, κ
CD161	APC	FAB9766R			
7AAD	7-AAD	559925			-

Figure Expanded View

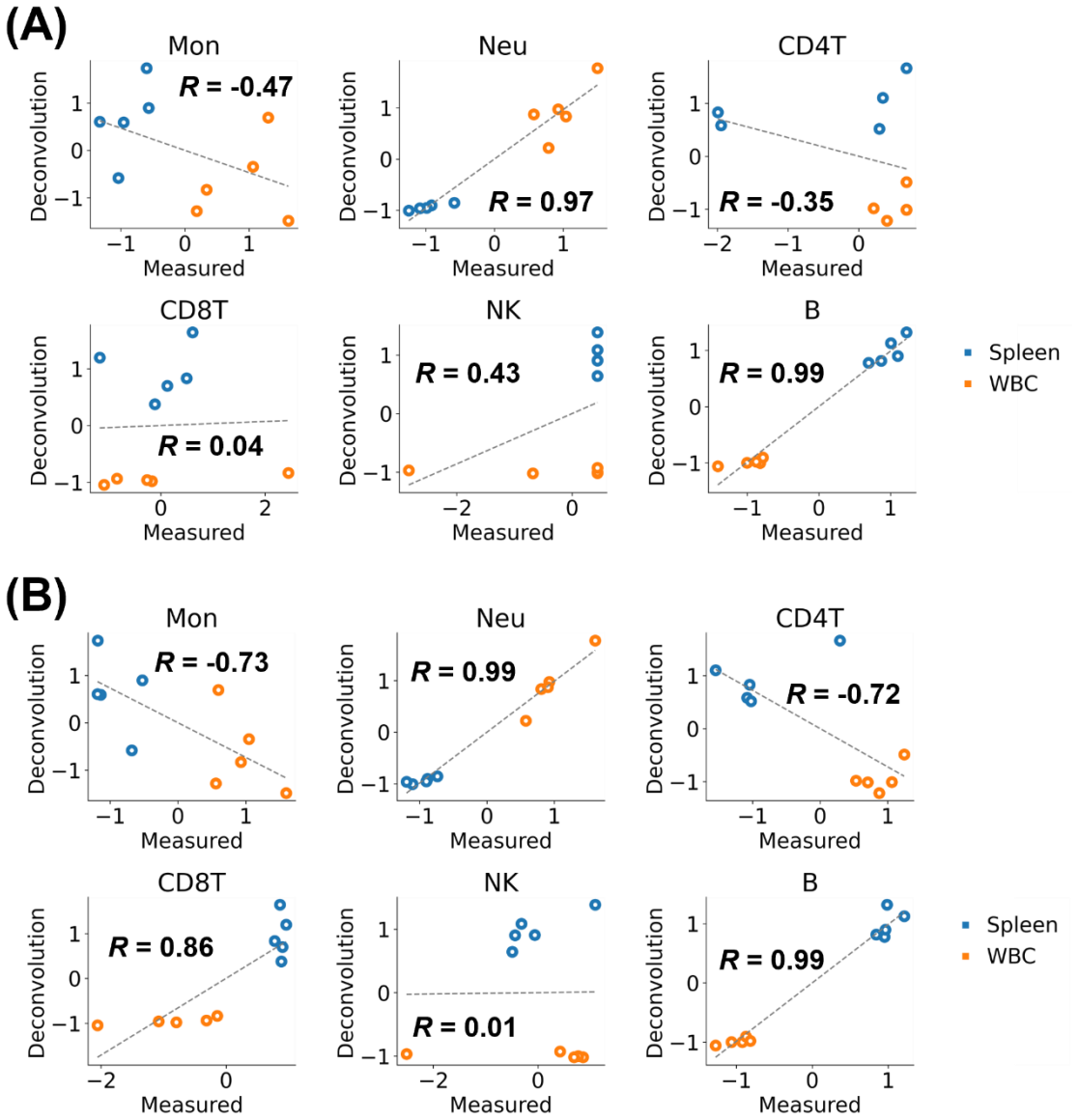

Figure EV1

Comparison of estimated and measured immune cell trafficking in the rat spleen and whole blood cells (WBCs). Values were measured by flow cytometry after cell isolation and red blood cell lysis. Estimated values were calculated by deconvolution using (A) a human-derived signature or (B) mouse-derived signature matrices. Note that these values were converted to *z*-score between the compared samples.

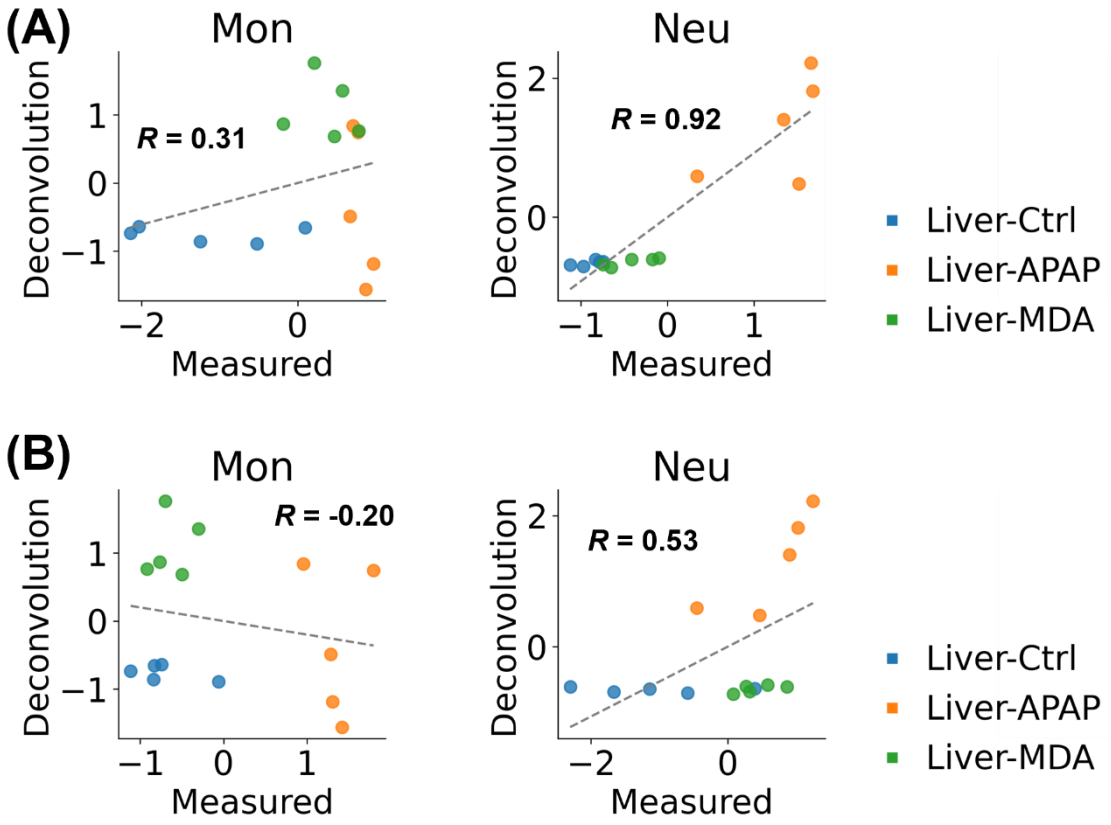

Figure EV2

Comparison of estimated and measured immune cell trafficking in rat liver after compounds were administered. values were measured by FACS after cell isolation and red blood cell lysis. Estimated values were calculated by deconvolution using (A) a human-derived signature matrix or (B) mouse-derived signature matrices. Note that these values were converted to *z*-score between the compared samples. Ctrl, control; APAP, acetaminophen; MDA, methylene dianiline.

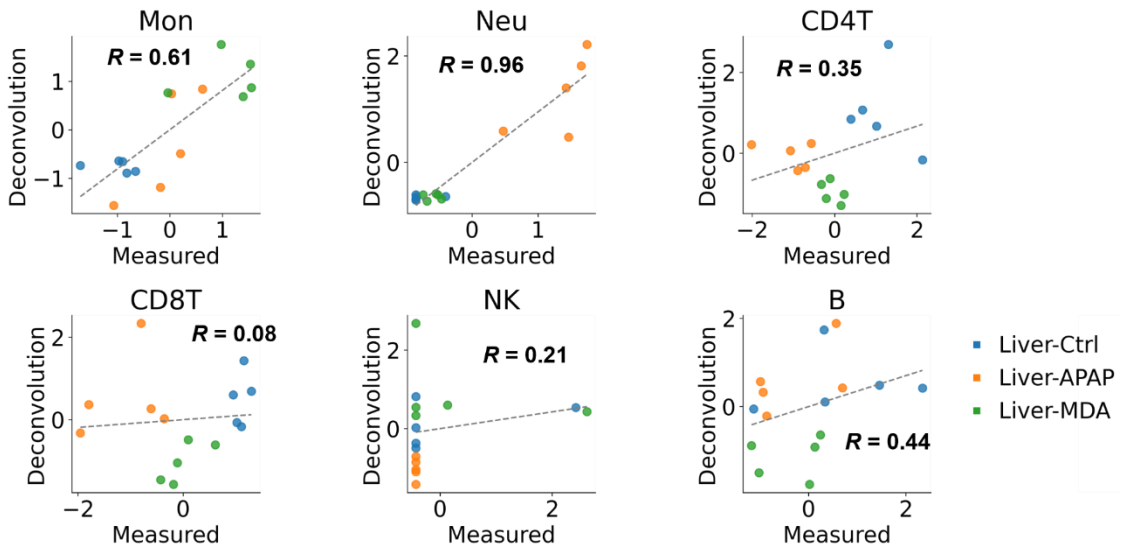

Figure EV3

Comparison of estimated and measured immune cell trafficking in rat liver after compounds were administered. Values were measured by FACS after cell isolation and red blood cell lysis. Estimated values were calculated by deconvolution using a rat-specific signature matrix. Note that these values were converted to *z*-score between the compared samples. Ctrl, control; APAP, acetaminophen; MDA, methylene dianiline.

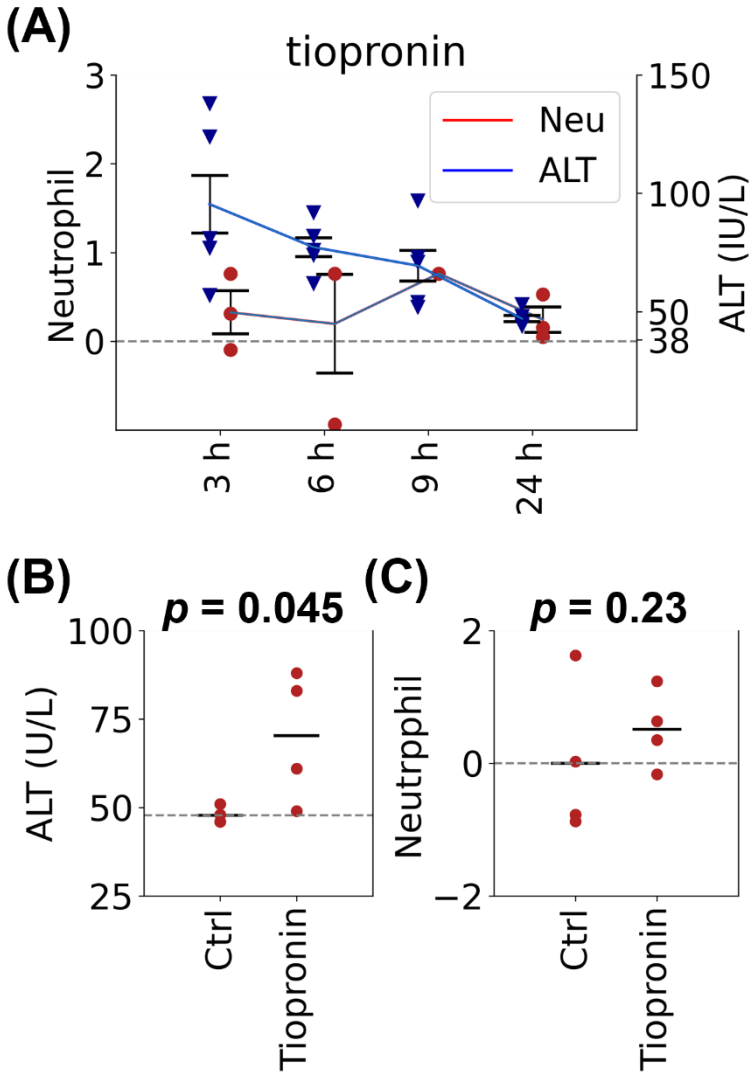

Figure EV4

(A) Time course of changes in neutrophils estimated in the liver and ALT measured in the blood of rat administered tiopronin as determined from the Open TG-GATEs database. Value of (B) ALT in blood and (C) neutrophils in the liver at 3 h after tiopronin administration. Note that neutrophils values are converted to *z*-score for the corresponding control samples. The tests of significance were conducted using an unpaired sample greater-sided Welch *t* test.

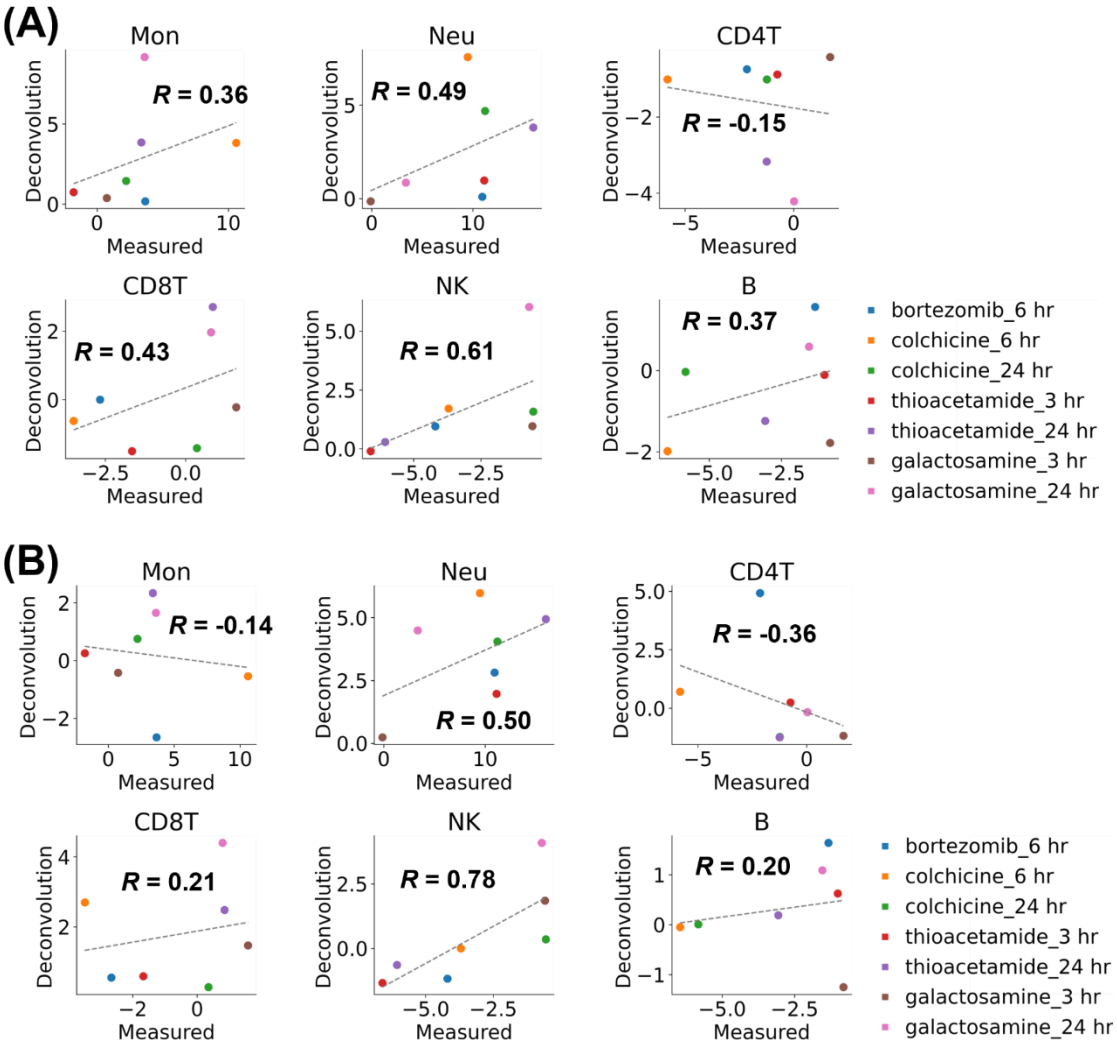

Figure EV5

Comparison of estimated and measured immune cell trafficking in the rat liver after administering each compound. Values were calculated by deconvolution using (A) human-derived signature matrix or (B) mouse-derived signature matrices. Note that these values were converted to *z*-score between the compared samples. Note that each point is the median value in the experiment or in the Open-TGGATEs database (n = 3~4).

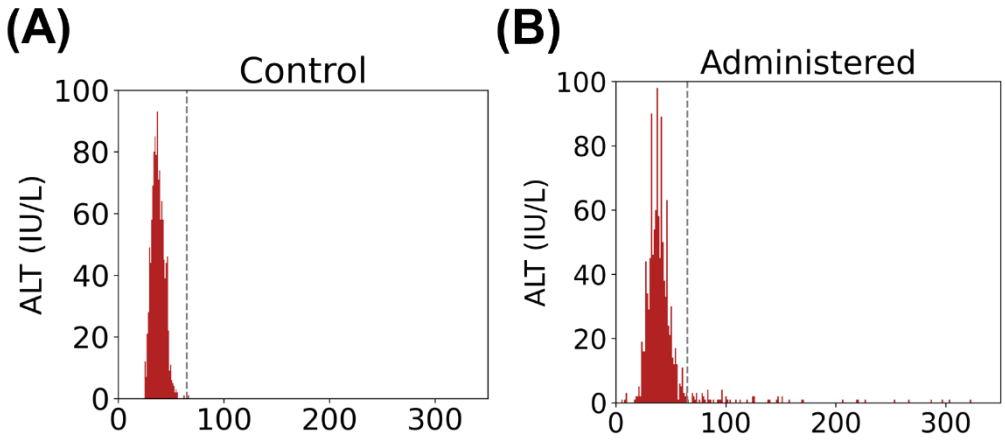

Figure EV6

ALT distribution after (A) administration of control solvent or (B) of compound from the Open TG-GATEs database. The vertical dotted line is the threshold for targeting the compounds used for stratification.

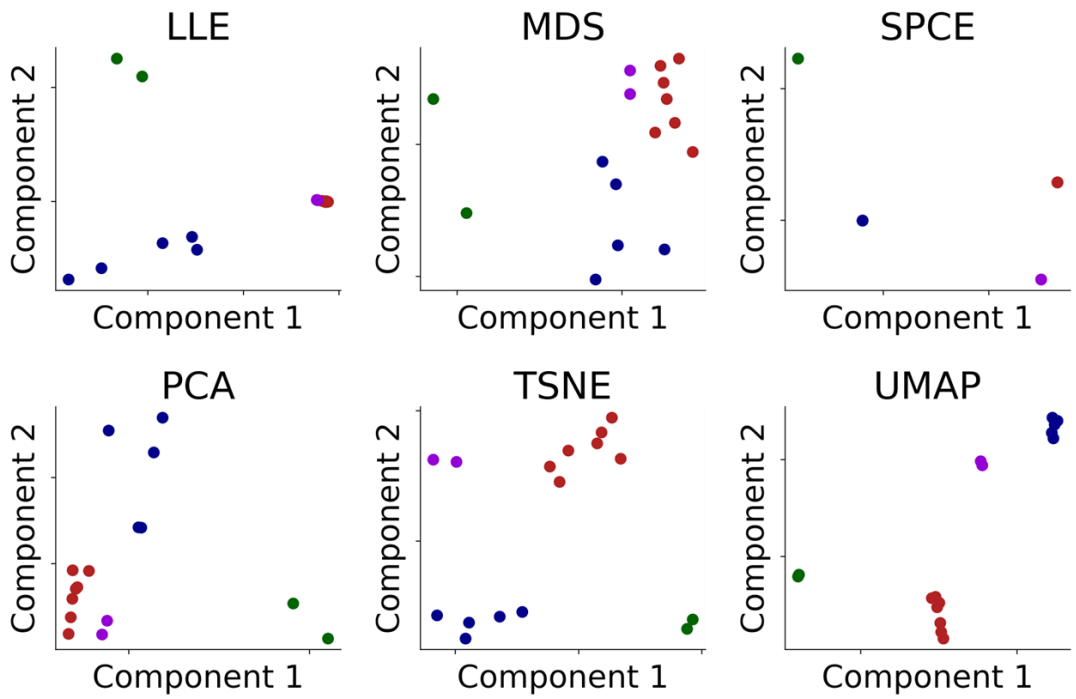

Figure EV7

Visualization of dimensionally reduced immune cell trafficking features from each method subjected to meta-visualization. Each color corresponds to the color of the stratified compound. LLE, locally linear embedding; MDS, multidimensional scaling; SPCE, spectral embedding; PCA, principal component analysis; TSNE, *t* distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection.

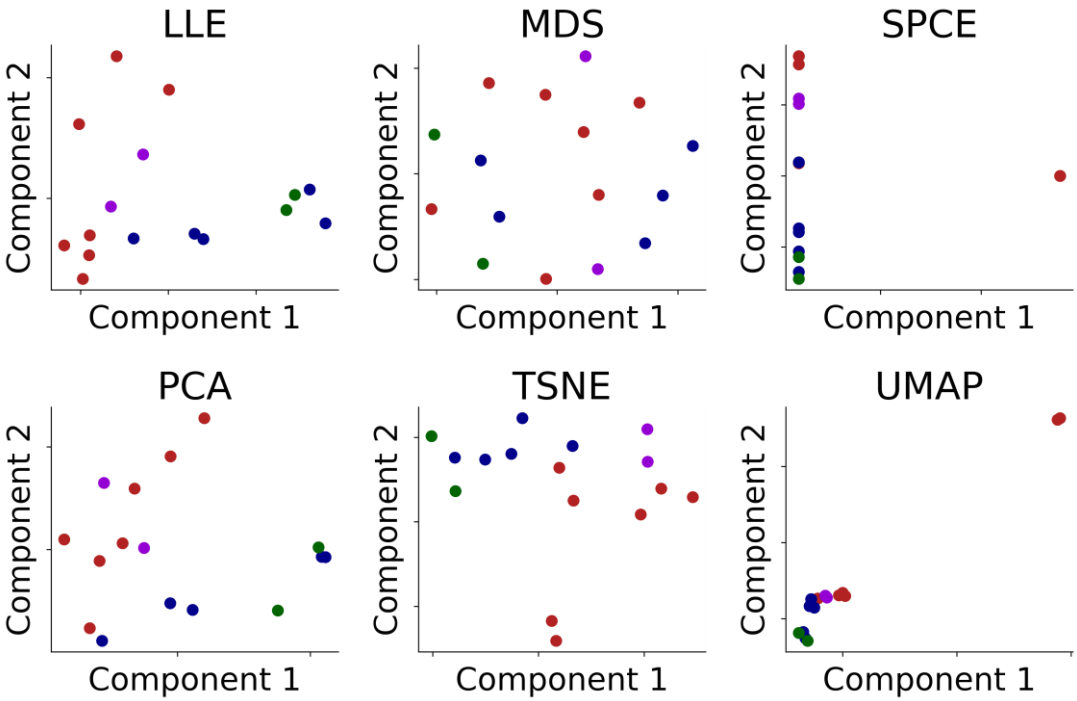

Figure EV8

Visualization of dimensionally-reduced transcriptome features from each method subjected to meta-visualization. Each color corresponds to the color of the compounds stratified by immune cell trafficking analysis. LLE, locally linear embedding; MDS, multidimensional scaling; SPCE, spectral embedding; PCA, principal component analysis; TSNE, *t* distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection.

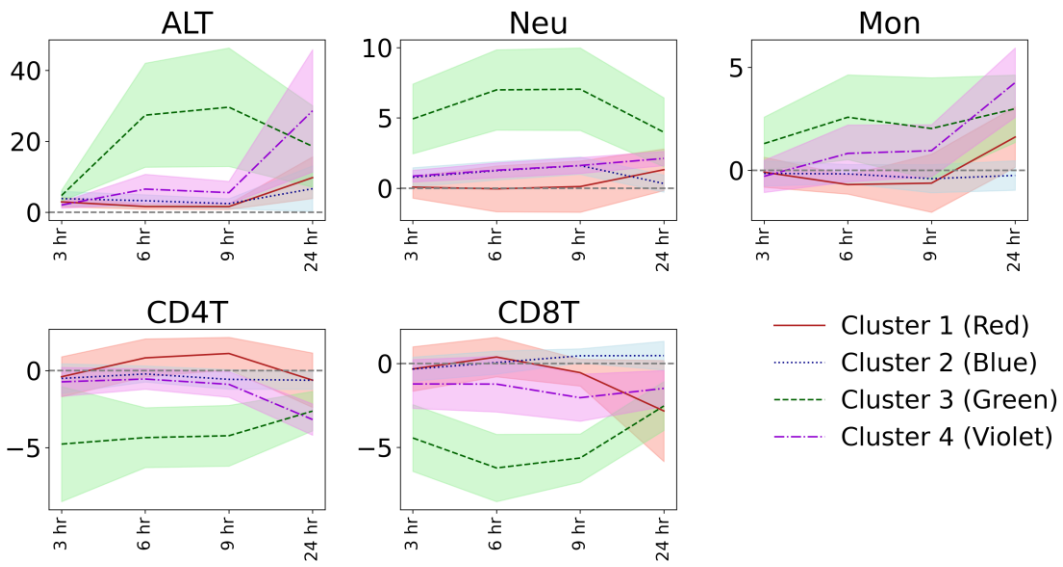

Figure EV9

Time course of changes in immune cell trafficking and ALT in stratified compound clusters. The line is the average value for each cluster, and the filled color around it indicates the 95% confidence interval. Each color corresponds to the color of the compound stratified by analysis of the transcriptome. Note that the scale of the x-axis does not correspond to the actual times.

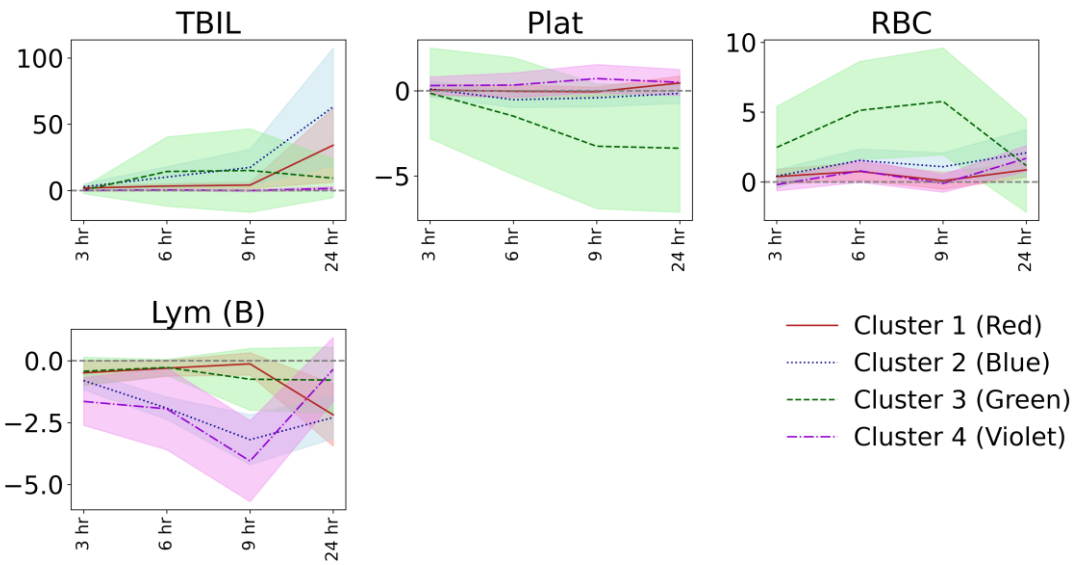

Figure EV10

Time course of changes in blood biochemistry. The line is the average value for each cluster, and the filled color around it indicates the 95% confidence interval. Each color corresponds to the color of the compound stratified by analysis of immune cell trafficking. Note that the scale of the x-axis does not correspond to the actual times. TBIL, total bilirubin; Plat, Platelets; RBC, red blood cells; LymB, lymphocytes in the blood.

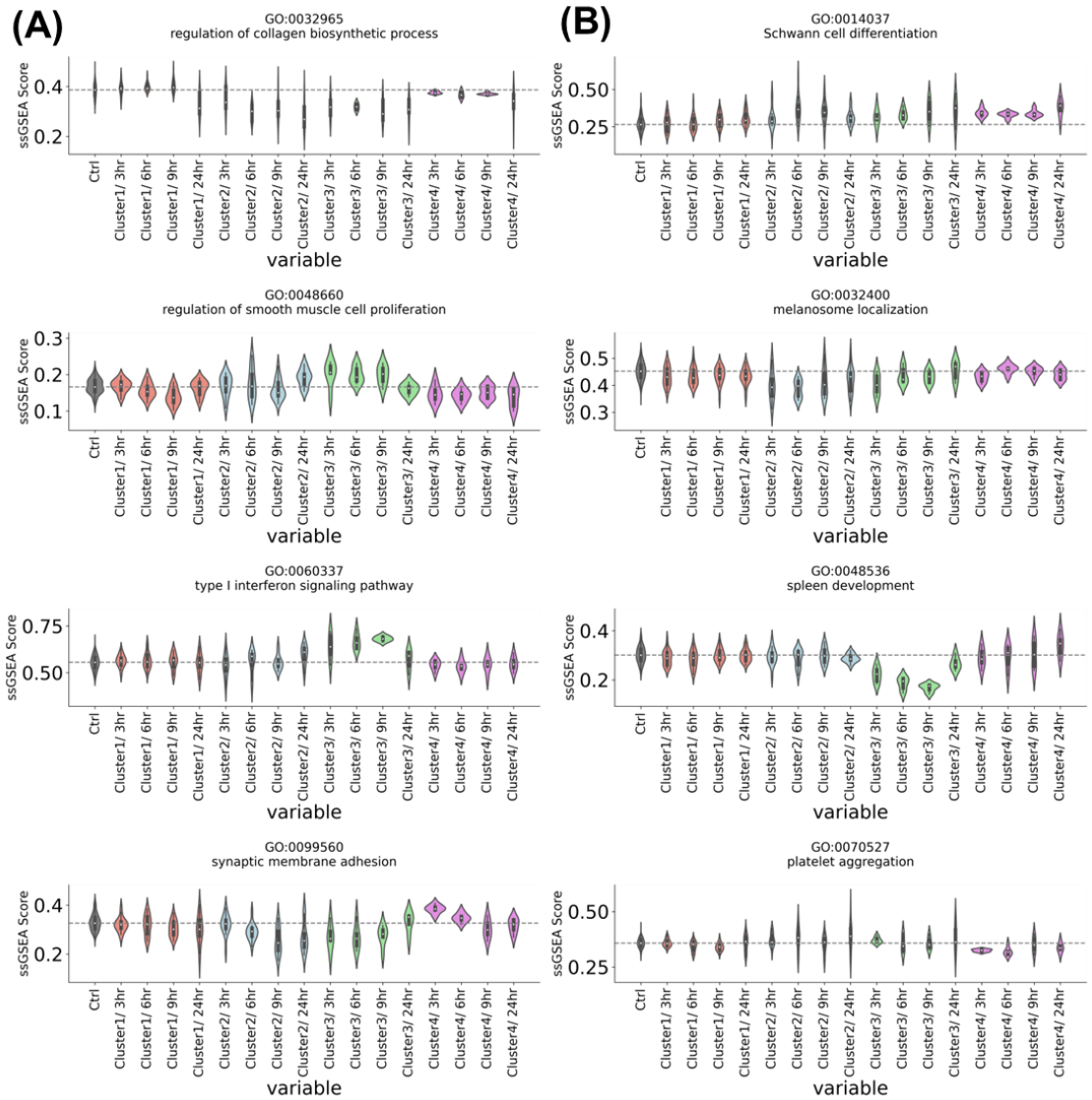

Figure EV11

Time course of changes in single sample gene set enrichment analysis (ssGSEA) scores of the liver transcriptomes. For each cluster, calculated from the immune cell trafficking, from the top to the bottom, the gene ontology terms with the lowest p-value in the comparison test with the other clusters are shown. Terms with (A) highest and (B) lowest ssGSEA score in that time compared with other clusters.

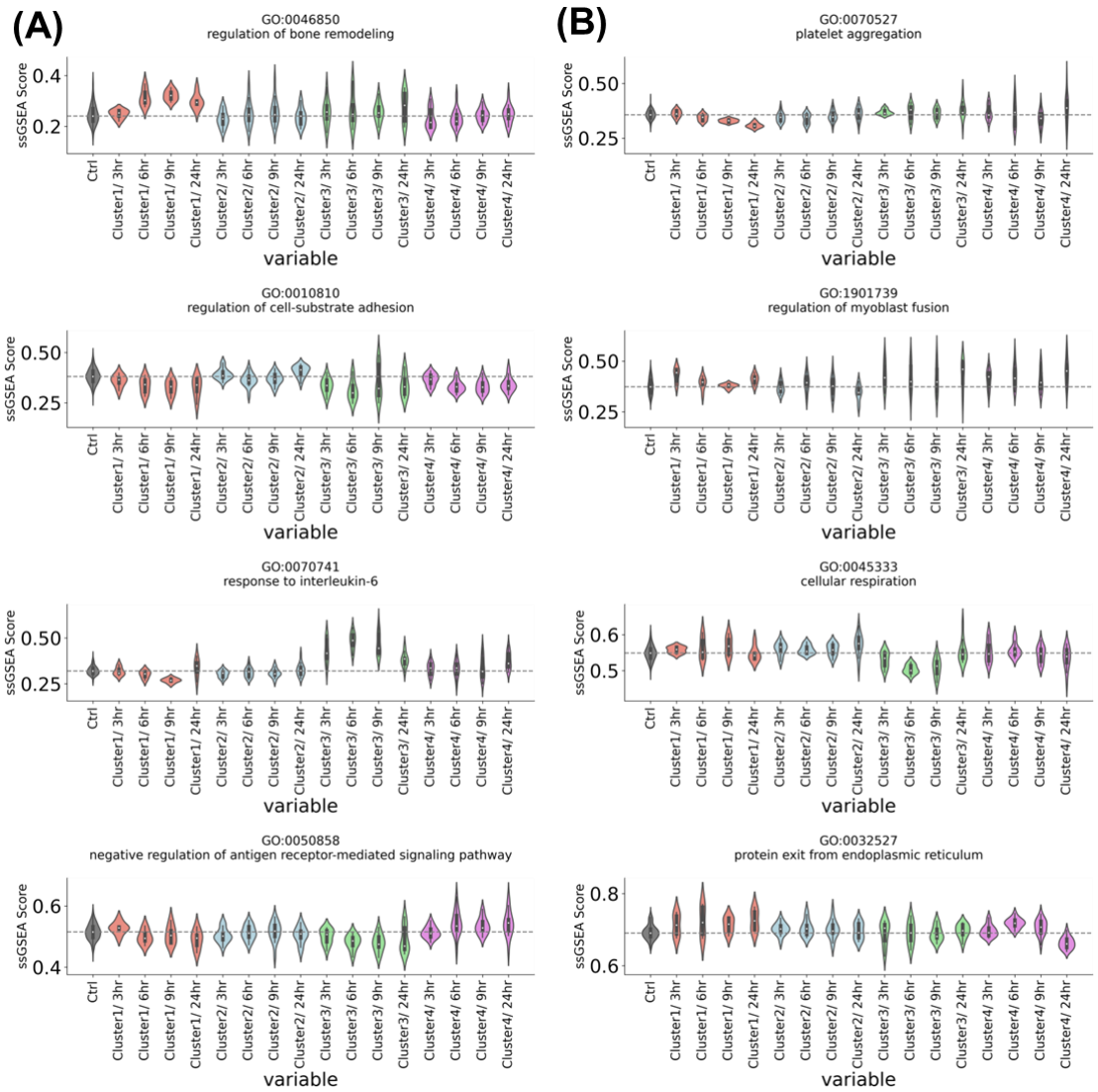

Figure EV12

Time course of changes in single sample gene set enrichment analysis (ssGSEA) scores of the liver transcriptomes. For each cluster, calculated from the liver transcriptomes, from the top to the bottom, the gene ontology terms with the lowest p-value in the comparison test with the other clusters are shown. Terms with (A) highest and (B) lowest ssGSEA score in that time compared with other clusters.

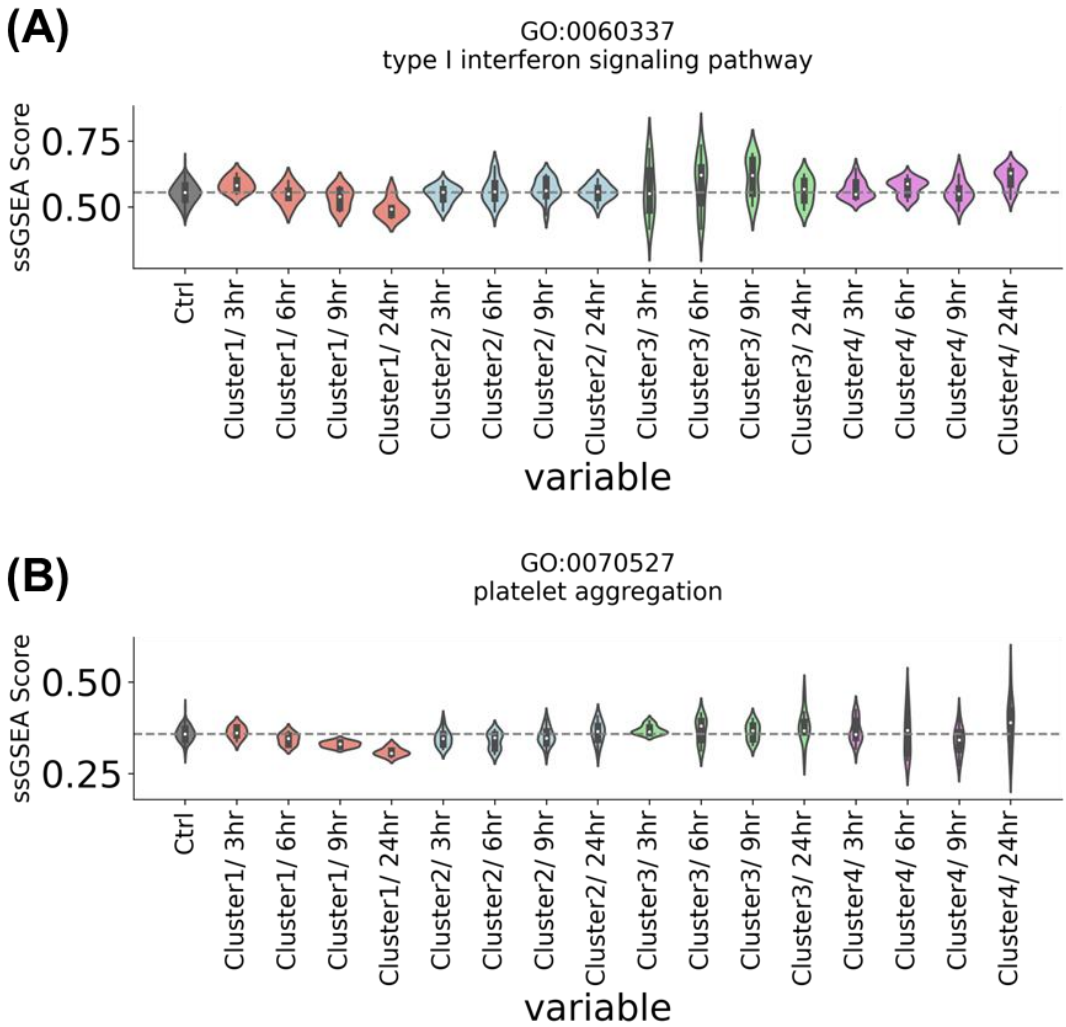

Figure EV13

Time course of changes in single sample gene set enrichment analysis (ssGSEA) scores of the liver transcriptomes. For gene ontology terms of (A) interleukin-1-mediated signaling pathway and (B) platelet aggregation.

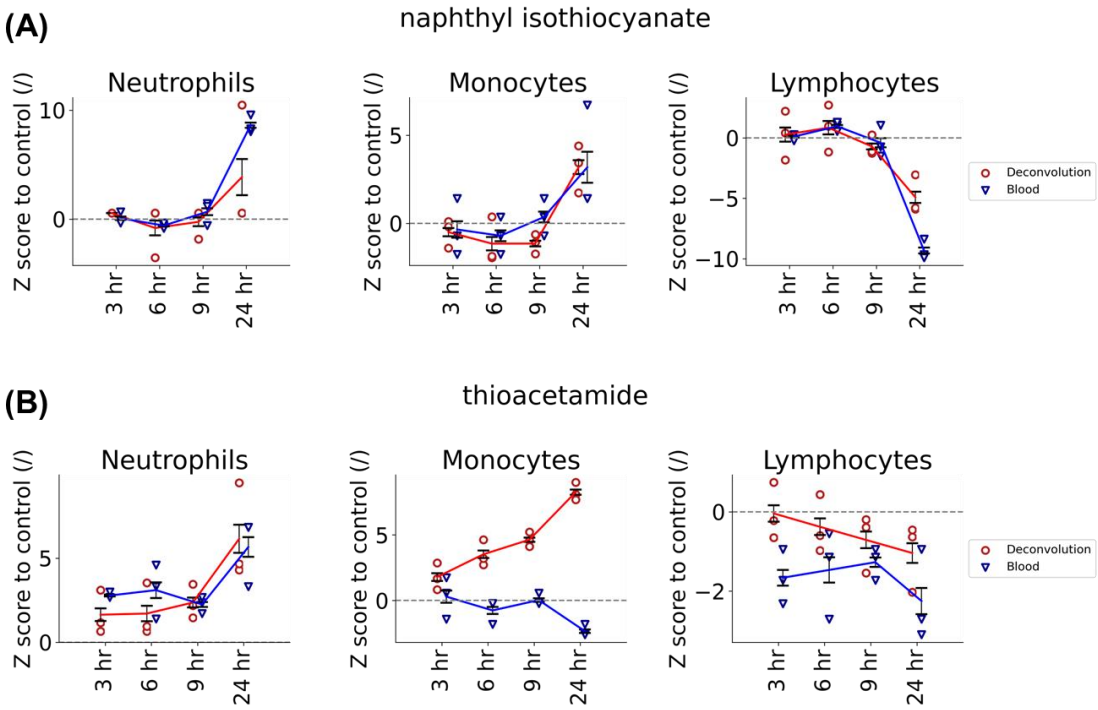

Figure EV14

Time course of changes in estimated neutrophils, monocytes, and lymphocytes in the liver using deconvolution and that of measured in the blood of rats administered (A) naphthyl isothiocyanate or (B) thioacetamide as determined from the Open TG-GATEs database. Note that all values were converted to *z*-score for the corresponding control samples. Estimated values for lymphocytes are estimated from sum of NK, B, CD4 T, and CD8 T cells.