Estimating Enzyme Expression and Metabolic Pathway Activity in *Borreliella*-Infected and Uninfected Mice

F. M. Rondel[1*], H. Farooq[1*], R. Hosseini[1], A. Juyal[1], S. Knyazev[2,3], S. Mangul[3,4,5], A. S. Rogovskyy[6], A. Zelikovsky[1*]

[1]Department of Computer Science, Georgia State University, Atlanta, Georgia, USA

[2]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, California, USA

[3]Bioinformatics Interdepartmental Program, University of California, Los Angeles, California, USA

[4]Department of Clinical Pharmacy, USC Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences, University of Southern California, Los Angeles, California, USA

[5]Department of Quantitative and Computational Biology, USC Dornsife College of Letters, Arts and Sciences

[6]Department of Veterinary Pathobiology, School of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas, USA

[*]Email: {frondel1,hfarooq5}@student.gsu.edu, alexz@gsu.edu

April 24, 2024

Keywords: Maximum likelihood, Expectation maximization, Enzyme expression, Metabolic pathway activity, Mice

Abstract: Evaluating changes in metabolic pathway activity is essential for studying disease mechanisms and developing new treatments, with significant benefits extending to human health. Here, we propose EMPathways2, a maximum likelihood pipeline that is based on the expectation-maximization (EM) algorithm, that is capable of evaluating enzyme expression as well as metabolic pathway activity level. We first estimate enzyme expression from RNA-seq data that is used for simultaneous estimation of pathway activity levels using enzyme participation levels in each pathway. We implement the novel pipeline to RNA-seq data from several groups of mice which provides a deeper look at the biochemical changes occurring as a result of bacterial infection, disease, and immune response. Our results show that estimated enzyme expression, pathway activity levels, and enzyme participation levels in each pathway are robust and stable across all samples. Estimated activity levels of a significant number of metabolic pathways strongly correlate with the infected and uninfected status of the respective rodent types.

1 Introduction

Lyme disease, a significant public health concern, is caused by the tick-borne spirochete, *Borreliella burgdorferi* (*Bb*). Specific mammalian hosts respond differently to *Bb* infection, with varying disease manifestations (Mead, 2015; Kugeler et al, 2021). For instance, certain strains of *Mus musculus* exhibit severe arthritis, while others like *Peromyscus leucopus* do not show visible disease symptoms post-infection (Barthold et al, 1990; Crandall et al, 2006; Schwanz et al, 2011). This study focuses on the infection-induced changes in gene expression to understand the potential mechanisms of disease tolerance in *P. leucopus* mice. Comparing transcriptomic responses to *Bb* infection between *P. leucopus* and *Mus musculus* (C3H/HeJ, hereafter referred to as C3H) should shed light on the disease-tolerance capacities of *P. leucopus* mice. Mice have been the experimental tool of choice for the vast majority of immunologists. Studying their immune responses has yielded tremendous insight into the inner workings of the human immune system (Masopust et al, 2017). Humans and mice share approximately 70 percent of the same protein-coding gene sequences (Margolin, 2000).

Therefore, analyzing the activity of metabolic pathways of mice with contrasting immune responses is imperative to gaining a deeper understanding of human immune system. Measuring the functional activity, enrichment, and interaction of metabolic pathways in rodent groups with diametric health conditions is essential for understanding the biochemical and metabolic changes that may occur in humans during stress or disease. Despite many advances of using biomolecules (DNA, RNA, enzymes) to assess the biochemical changes in mice, it remains challenging to quantify how the expression of individual enzymes contributes to the activity of multi-enzyme metabolic pathways. In this study, we analyze differentially active metabolic pathways from RNA sequencing data to generate an efficient model for understanding metabolic pathway activity changes (Subramanian et al, 2005; Efron and Tibshirani, 2007; Mitrea et al, 2013; Shen et al, 2019). Even though advances in high-throughput sequencing have aided the exploration of RNA-Seq data, it is often challenging to analyze metabolic pathway activity changes in organisms with varying health conditions, notably as existing pathway analysis tools (e.g., MinPath, MetaPathways, MEGAN4) often yield variable conclusions about the activity of pathways based on RNA data (Huson et al, 2011; Konwar et al, 2013; Ye and Doak, 2009; Sharon et al, 2011). To overcome the current challenges, we developed a workflow that uses a maximum likelihood-based model and annotations based on the KEGG (Kanehisa and Goto, 2000) database to estimate transcript frequency, enzyme expression, enzyme participation in pathways, and metabolic pathway activity in microbial communities (Rondel et al, 2020, 2021).

In this paper, we test this model using transcriptomic data of mice infected with *Bb*, an agent of Lyme disease, and their uninfected controls. The data describes the infected as well as the uninfected groups of two rodent species - *Mus Musculus* and *Peromyscus leucopus* to elucidate the complex metabolic pathway activity changes between rodents with inherent tolerance to *Bb* infection (*P. leucopus* mice) and those that develop Lyme disease (a laboratory strain of C3H/HeJ (C3H) mice). The proposed methodology is to use a maximum likelihood estimate to infer the pathway activity considering an enzyme's participation. First, we filtered mouse specific metabolic pathways from the KEGG database and merged the expression of enzymes represented by the same group of genes. We adjusted our EM algorithm based pipeline and improved it using enzyme

participation level in each pathway and then used these estimations for more accurate predictions of pathway activity(Rondel et al, 2021). Our contributions include:

- Estimation of metabolic enzyme expression, identification of groups of rodents' enzymes that are represented by the same group of genes
- Estimation of enzyme-in-pathways coefficients and confirmation that they are more stable than for microbial communities in (Rondel et al, 2021). Additionally, we show that these coefficients do not significantly vary across species of infected and uninfected mice
- Differential analysis of metabolic pathway activity in *P. leucopus* and C3H mice uninfected and infected with *Bb*

The rest of the paper is organized as follows. In the next section, we describe the pipeline of our software framework and several EM-based algorithms for estimating enzyme expression and metabolic pathway activity between two rodent species. Further on, we describe our data including sequencing data, and extraction of metabolic enzymes and pathways. Finally, we use our results to provide a statistical validation of the proposed pipeline.

2 Materials and Methods

2.1 Data procurement.

The data for this study was acquired from a previous experiment conducted by (Gaber et al, 2023). It consists of twelve male mice, six *P. leucopus* and six C3H/HeJ (C3H) mice, were split into four groups. Half of these mice were subcutaneously inoculated with *Bb* 297, while the remaining mice were injected with a sterile saline solution as a control group. Following inoculation, blood samples were taken from all twelve mice to confirm the infection. At 70th day of post-inoculation, various tissues were harvested from the mice and cultured to further examine the presence or absence of viable spirochetes. Spleens were harvested from all the mice, preserved, and stored at - 80°C until RNA extraction was performed.

2.2 Pipeline for estimating metabolic pathway activity of C3H and *P. leucopus* mice

In the past, we created a pipeline for estimating metabolic pathway activity levels in a microbial community(Rondel et al, 2021). We explored the differential pathway activity inside of a microbial community under different conditions. Microbial community has diverse species and in some cases hard to interpret due to abundance of species in the samples.
Below, we describe our novel metabolic pathway activities pipeline *EMPathways2* (see Fig 1) that is used for estimating pathways activities in mice. These models are resolved using the EM algorithm.(see Fig 1).

The entire pipeline *EMPathways2* consists of the following five steps:

- The first step is the collection of samples from infected and uninfected rodent groups, which then get sequenced.

- RNA-Seq reads are mapped into reference transcriptomes of C3H and *P. leucopus* mice collected from NCBI reference database. The mapped reads were used by IsoEM2 to generate gene expression data (Mandric et al, 2017).

- We use KEGG to establish the many-to-many correspondence between genes and enzymes (see Sec. 2.3). We estimate enzyme expressions based on gene expression using EM (see Fig 1).

- Unstable enzymes that converge inconsistently were identified, grouped, and collapsed (see Sec. 2.4).

- The feedback loop is based on inferred enzyme expressions and metabolic pathway annotation. It simultaneously estimates enzyme participation coefficients and metabolic pathways activity levels (see Sec. 2.5).

2.3 Mapping between genes, enzymes and pathways for C3H and *P. leucopus* mice

KEGG metabolic pathway database has information on all metabolic pathways that occur in the living organisms. However, the scope of *EMPathways2* is to analyze metabolic pathways in the rodents. We concentrate on 152 metabolic pathways and 2386 enzymes that play a significant role in mouse metabolism which is confirmed by literature referenced in PubMed.

In order to compute metabolic pathway activity levels, EMPathways2 requires an input in a form of a correspondence between genes and enzymes as well as a dictionary of enzymes participating in metabolic pathways. Gene-enzyme as well as enzyme-pathway mappings were extracted from NCBI Entrez Molecular Sequence Database System as well as KEGG PATHWAY database respectively and which provides consolidated access to nucleotide, protein sequence, gene-centered and genomic mapping data. We used KEGG's and NCBI's APIs to collect raw data allowing us to produce a correspondence of genes to enzymes and enzymes to metabolic pathways. We used the collected data to create sets of genes participating in production of every enzyme, as well as sets of enzymes required for functional activity of every metabolic pathway.

2.4 Enzyme grouping

There is a many-to-many correspondence between genes and enzymes which may pose challenges to compute enzymes expression. To resolve this challenge, we use a maximum likelihood EM model to infer enzyme expression from gene expression which converges consistently in vast majority of cases. However, there are enzymes that share some genes as well as enzymes whose genes are entirely a subset of genes used for production of another enzyme. In some of those cases, EM struggles to discern one such enzyme from its genetic relatives and in turn converges inconsistently from one run to another. Those enzymes that fail to converge consistently are labeled unstable and grouped into clusters whose expression as a single entity converges consistently after every EM iteration. After running a few iterations of gene-enzyme EM, we observe clusters of enzymes whose expression varies individually but they are stable in groups. The unstable enzymes individual expressions vary

from one run to another. However, summing them always converge to the same expression in every run.(see Table 1). This instability makes such groups of enzymes indistinguishable to our algorithm. To establish the groups accurately, we run EM and produce enzyme expression values for every enzyme. We establish clusters by evaluating the grouped enzyme expressions which do not converge consistently individually, but the sum of their expressions always converges to the same value. As a result such enzymes must be treated as single entities. After all unstable enzyme groups are found, we collapse them into one (see Figure 2 (A)). The groups are collapsed to a single enzyme with the lowest EC number nomenclature. The collapsed group enzyme is then used to compute metabolic pathway expressions of all related pathways (see Figure 2 (B)). In total, we found and collapsed 59 pairs, three triplets and one quadruple of indistinguishable enzymes. Table 2 gives the list of triplets and a quadruplet found in mice. We have compared the list of collapsed enzymes for microbial communities found in (Rondel et al, 2021) with the list of collapsed enzymes in rodents. We found out that there are 28 pairs common for these two datasets.

2.5 Feedback loop for pathway activity level estimation

Each enzyme is initially assigned a participation coefficient of $1/|w|$, where $|w|$ is the total amount of enzymes in the pathway w. The *Feedback loop for pathway activity* updates the enzyme participation level by fitting expected enzyme expressions to the expressions estimated by *EM for enzyme expression*.

The initial estimate of the participation level of an enzyme e in a pathway w may be far from accurate. However, more accurate estimates of enzyme participation can lead to more accurate estimates for the pathway activity levels. Our algorithm first estimates enzyme expression from gene expression using the *EM for enzyme expression*. The **E-step** and **M-step** are ran in order to compute expected expression and compare it to the new estimate respectively. After computing enzyme expressions, we then filter out enzymes with stable expressions and perform enzyme grouping on enzymes with unstable expressions. Pathway activity levels are in turn computed using the EM for pathway activity level.

Following, we estimate how well the computed activities f_w's fit the enzyme expressions using

the *EM for enzyme participation* depicted in Figure 1.

Together, EM for enzyme participation and EM for pathway activity levels make up the *Feedback loop* for pathway activity level estimation. If the fit is not good enough, then the *Feedback loop for pathway activity level* is applied to update the enzyme participation levels p_{ew}'s with the *EM for enzyme participation* and then f_w's are recomputed according to updated p_{ew}'s.

The E-step. Compute expected p_{ew}^{exp}'s that will make $f_e = f_e^{exp}$ for each $e \in E, w \in W$,

$$p_{ew}^{exp} = p_{ew} \times \frac{f_e}{f_e^{exp}}$$

The M-step. Provide the new estimates by normalization for each $e \in E, w \in W$,

$$p_{ew}^{new} = \frac{p_{ew}^{exp}}{\sum_{e \in E} p_{ew}^{exp}}$$

The algorithm halts when the change in estimates between iterations is small enough:

$$||p^{new} - p|| = \sum_{e \in E, w \in W} (p_{ew}^{new} - p_{ew})^2 \leq \epsilon \ll 1$$

3 Results

We have applied the proposed pipeline *EMPathways2* to rodent RNA-Seq data. For each group of rodents, we compute the mean and the standard deviation for each pathway activity level. We categorize a metabolic pathway as having significantly (resp. slightly) different activity across conditions if its standard deviation intervals do not intersect (resp. its standard deviation intervals intersect but do not contain each other means) for different conditions. Note that if a metabolic pathway has significantly (resp. slightly) different activity, then the probability that the activity is the same is below 0.25% (resp. 5%).

The list of metabolic pathways with significantly different activity across infected and uninfected

C3H (res *P. leucopus*) are in Tables 3,6. We found that four C3H metabolic pathways are expressed with differing activity levels. For example, caffeine metabolism has a significant difference in its activity levels between the infected and uninfected groups. Note that the number of metabolic pathways of *P. leucopus* significantly affected by the infection is much higher than for C3H that can explain why C3H get sick after infection while *P. leucopus* do not show any symptoms.

The list of metabolic pathways with slightly different activity across infected/uninfected C3H (res *P. leucopus*) are in Tables 5,7. Note that the lists of these pathways are very different for different mouse species.

Finally, we check how stable are the enzyme participation coefficients across different mouse species (see Table 4). Note that the average relative standard deviation (RSD) for C3H is 2.7% in contrast to much higher RSD for 8.9% for *P. leucopus*. That can be caused by that fact that C3H mice are genetically identical. Note that the average RSD for enzyme participation coefficients in the microbial community for the same metabolic pathway (ec00620) is 34.8% which is significantly higher (see (Rondel et al, 2021)) than RSD for mice.

4 Discussion

The results of our study highlight the potential of using RNA-Seq data to estimate enzyme expression and metabolic pathway activity in the rodent models of disease. Our modified expectation-maximization based pipeline, EMPathways2, has successfully demonstrated its ability to estimate enzyme expression, enzyme participation in pathways, and metabolic pathway activity levels in both infected and uninfected mice.

These findings further enhance our understanding of the biochemical changes occurring in the host during bacterial infection. The differences in enzyme expression and pathway activity levels between infected and uninfected mice could provide insights into the immune response mechanisms at the metabolic level. This, in turn, can potentially be used to develop new therapeutic strategies for bacterial infections and other diseases.

The variation in pathway activity levels between C3H and *P. leucopus* sheds light on the different

immune responses in these rodent species. The higher number of pathways significantly affected by infection in *P. leucopus* compared to C3H may explain why C3H mice develop Lyme disease symptoms while *P. leucopus* mice do not. This highlights the importance of considering the host species in understanding the disease pathogenesis. It is also interesting to note that the enzyme participation coefficients were more stable in C3H compared to *P. leucopus*. This could be due to the genetic similarities among laboratory mouse strains, as compared to wild mice.

5 Conclusions

In this paper we propose an improved maximum likelihood-based pipeline for the estimation of metabolic pathway activity in mice using the KEGG pathway database. Specifically, the proposed approach uses EM-based algorithms to estimate enzyme expression, enzyme participation levels in pathways, and metabolic pathway activity.

The proposed metabolic pathway analysis was applied to the RNA-Seq data from 12 mice samples collected from C3H and *P. leucopus* with half them infected by *Bb* 297. The key findings of the study are as follows:

- The infection affects metabolism of both mice while for *P. leucopus*, the affect is more significant than for C3H.

- The enzymes participation coefficients vary insignificantly for C3H in contrast to higher variation for *P. leucopus* and much higher variation for microbial communities.

Acknowledgments

The authors would like to thank Dr. Artem S. Rogovskyy for his invaluable expertise and guidance throughout this project.

Authorship Confirmation

Fil Rondel did conceptualization and methodology and writing of the paper. Hafsa Farooq and Roya Hosseini worked on implementation of methodology of metabolic pathway retrieval and did some writing of the draft. Akshay Juyal helped in methodology technically and also did formatting of the draft. Sergey Knyazev worked on conceptualization and preparing the initial implementation of the project. Serghei Mangul and Alex Zelikovsky supervised the project along funding acquisition.

Authors Disclosure

The authors state that the research was conducted without any commercial or financial affiliations that could be interpreted as potential conflicts of interest.

Funding

The work at Georgia State University (GSU) was partially supported by NIH grant 1R21CA241044-01A1, NSF grant IIS-2212508, by the GSU Molecular Basis of Disease Fellowship, and by the GSU Brain and Behavior Fellowship. SM was supported by the National Science Foundation Grants (2041984 and 2135954) and National Institutes of Health Grant R01AI173172. The work at Texas A&M University was partially supported by National Institutes of Health (NIH) grant R03AI135159-02, Department of Veterinary Pathobiology, Texas A&M School of Veterinary Medicine and Biomedical Sciences, and Texas A&M AgriLife.

Data Availability

The data presented in the study are deposited in the GenBank Sequence Read Archive (SRA) depository and the SRA accession numbers are SAMN32740077, SAMN32740078, SAMN32740079, SAMN32740080, SAMN32740081, SAMN32740082, SAMN32740083, SAMN32740084, SAMN32740085, SAMN32740086, SAMN32740087, SAMN32740088.

References

Barthold, S. W., Beck, D. S., Hansen, G. M., et al. Lyme borreliosis in selected strains and ages of laboratory mice. *Journal of Infectious Diseases*, 162(1):133–138, 1990.

Crandall, H., Dunn, D. M., Ma, Y., et al. Gene expression profiling reveals unique pathways associated with differential severity of lyme arthritis. *The Journal of Immunology*, 177(11):7930–7942, 2006.

Efron, B., Tibshirani, R. On testing the significance of sets of genes. 2007.

Gaber, A. M., Mandric, I., Nitirahardjo, C., et al. Comparative transcriptome analysis of peromyscus leucopus and c3h mice infected with the lyme disease pathogen. *Frontiers in Cellular and Infection Microbiology*, 13:1115350, 2023.

Huson, D. H., Mitra, S., Ruscheweyh, H.-J., et al. Integrative analysis of environmental sequences using megan4. *Genome research*, 21(9):1552–1560, 2011.

Kanehisa, M., Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

Konwar, K. M., Hanson, N. W., Pagé, A. P., et al. Metapathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC bioinformatics*, 14:1–10, 2013.

Kugeler, K. J., Schwartz, A. M., Delorey, M. J., et al. Estimating the frequency of lyme disease diagnoses, united states, 2010–2018. *Emerging Infectious Diseases*, 27(2):616, 2021.

Mandric, I., Temate-Tiagueu, Y., Shcheglova, T., et al. Fast bootstrapping-based estimation of confidence intervals of expression levels and differential expression from rna-seq data. *Bioinformatics*, 33(20):3302–3304, 2017.

Margolin, J. Of mice, men, and the genome. *Genome Research*, 10(10):1431–1432, 2000.

Masopust, D., Sivula, C. P., Jameson, S. C. Of mice, dirty mice, and men: using mice to understand human immunology. *The Journal of Immunology*, 199(2):383–388, 2017.

Mead, P. S. Epidemiology of lyme disease. *Infectious Disease Clinics*, 29(2):187–210, 2015.

Mitrea, C., Taghavi, Z., Bokanizad, B., et al. Methods and approaches in the topology-based analysis of biological pathways. *Frontiers in physiology*, 4:278, 2013.

Rondel, F., Hosseini, R., Sahoo, B., et al. Estimating enzyme participation in metabolic pathways for microbial communities from rna-seq data. In *Bioinformatics Research and Applications: 16th International Symposium, ISBRA 2020, Moscow, Russia, December 1–4, 2020, Proceedings 16*, pages 335–343. Springer, 2020.

Rondel, F. M., Hosseini, R., Sahoo, B., et al. Pipeline for analyzing activity of metabolic pathways in planktonic communities using metatranscriptomic data. *Journal of Computational Biology*, 28 (8):842–855, 2021.

Schwanz, L. E., Voordouw, M. J., Brisson, D., et al. Borrelia burgdorferi has minimal impact on the lyme disease reservoir host peromyscus leucopus. *Vector-borne and zoonotic diseases*, 11(2): 117–124, 2011.

Sharon, I., Bercovici, S., Pinter, R. Y., et al. Pathway-based functional analysis of metagenomes. *Journal of Computational Biology*, 18(3):495–505, 2011.

Shen, M., Li, Q., Ren, M., et al. Trophic status is associated with community structure and metabolic potential of planktonic microbiota in plateau lakes. *Frontiers in microbiology*, 10: 482889, 2019.

Subramanian, A., Tamayo, P., Mootha, V. K., et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005.

Ye, Y., Doak, T. G. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS computational biology*, 5(8):e1000465, 2009.

6 Tables

Table 1: A pair of individually unstable enzymes that are stable when summed into a group.

Enzymes	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
EC:3.1.3.12	0.054	0.311	0.251	0.317	0.12
EC:2.4.1.15	0.404	0.147	0.207	0.141	0.338
Sum	0.458	0.458	0.458	0.458	0.458

Table 2: Three triplets and one quadruplet of collapsed enzymes.

Triplet1	**Triplet2**	**Triplet3**	**Quadruplets**
EC:1.1.1.51	EC:6.3.4.13	EC:2.1.3.2	EC:6.3.4.9
EC:1.1.1.213	EC:6.3.3.1	EC:6.3.5.5	EC:6.3.4.10
EC:1.1.1.188	EC:2.1.2.2	EC:3.5.2.3	EC:6.3.4.11
			EC:6.3.4.15

Table 3: C3H pathways with significant different activity level across infected and uninfected groups.

Pathway Name	**ID**	**Infected Mice**	**Uninfected Mice**
		Mean ± Std	**Mean ± Std**
Caffeine metabolism	ec00232	84.48±1.069	82.888 ± 0.357
Mucin type O-glycan biosynthesis	ec00512	0.873 ± 0.666	2.205 ± 0.656
Pentose & glucuronate interconversions	ec00040	273.774 ±0.896	269.624 ± 1.82
Thiamine metabolism	ec00730	49.922 ± 0.297	59.741 ± 0.205

Table 4: The enzyme expression coefficients and relative standard deviations *(%RSD)* for the enzyme participation coefficients in pathway ec00620.

ec00620	Infected C3H	Uninfected C3H	%RSD	Infected P. leucopus	Uninfected P. leucopus	%RSD
EC:1.1.1.1	.110 .107 .113	.106 .109 .112	2.501	.054 .061 .049	.051 .045 .048	10.928
EC:1.5.8.3	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:3.1.3.3	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:2.1.2.10	.034 .034 .035	.035 .034 .034	1.504	.028 .030 .027	.027 .025 .025	7.027
EC:5.1.1.18	.032 .038 .033	.037 .034 .031	8.157	.013 .016 .011	.015 .012 .012	14.740
EC:1.4.3.21	.050 .055 .055	.054 .054 .051	4.019	.028 .029 .019	.027 .025 .024	14.269
EC:2.6.1.52	.059 .058 .060	.059 .061 .060	1.763	.047 .050 .042	.043 .041 .040	8.826
EC:2.1.1.20	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:2.7.1.165	.095 .086 .087	.088 .087 .088	3.696	.077 .067 .074	.061 .060 .059	11.586
EC:1.5.3.1	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:2.3.1.29	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:4.1.2.48	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:1.1.99.1	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:2.3.1.37	.055 .052 .055	.055 .056 .055	2.499	.072 .066 .074	.067 .070 .077	5.909
EC:2.1.2.1	.051 .051 .052	.052 .052 .050	1.591	.038 .041 .035	.036 .034 .033	8.093
EC:1.1.1.95	.050 .048 .050	.050 .050 .050	1.644	.050 .050 .048	.048 .046 .049	3.127
EC:1.1.1.103	.027 .025 .026	.026 .026 .026	2.433	.035 .031 .038	.033 .035 .041	10.039
EC:2.1.4.1	.049 .047 .049	.049 .049 .050	2.013	.049 .049 .044	.047 .046 .051	5.252
EC:4.2.1.22	.050 .048 .050	.050 .050 .050	1.644	.050 .050 .048	.048 .046 .049	3.127
EC:4.4.1.1	.061 .059 .061	.061 .060 .060	1.353	.047 .050 .043	.045 .042 .042	7.112
EC:4.3.1.17	.050 .048 .050	.050 .050 .050	1.644	.050 .050 .048	.048 .046 .049	3.127
EC:1.4.3.4	.062 .070 .064	.070 .067 .068	4.864	.028 .033 .023	.028 .027 .026	11.895
EC:1.4.3.3	.046 .053 .047	.052 .049 .045	6.711	.019 .023 .015	.021 .017 .017	15.771
EC:1.8.1.4	.088 .090 .089	.091 .090 .090	1.152	.042 .049 .034	.043 .039 .040	12.040
EC:2.1.1.5	.050 .048 .050	.050 .050 .050	1.644	.050 .050 .048	.048 .046 .049	3.127
EC:2.1.1.2	.049 .047 .049	.049 .049 .050	2.013	.049 .049 .044	.047 .046 .051	5.252

Table 5: C3H pathways activity level across infected and uninfected groups.

Pathway Name	**ID**	**Infected Mice Mean ± Std**	**Uninfected Mice Mean ± Std**
Ascorbate and aldarate metabolism	ec00053	139.789 ± 0.958	142.04 ± 1.581
Drug metabolism - cytochrome P450	ec00982	104.598 ± 0.85	105.261 ±0.518
Glycine, serine and threonine metabolism	ec00260	50.586 ± 0.807	48.544 ± 2.094
Glycosaminoglycan degradation	ec00531	78.611 ± 0.568	77.616 ± 1.778
Glycosphingolipid biosynthesis-globo & isoglobo series	ec00603	198.785 ± 8.711	202.718 ± 1.443
Selenocompound metabolism	ec00450	141.024 ± 23.292	159.357 ±1.326
Amino sugar and nucleotide sugar metabolism	ec00520	105.101 ± 0.287	104.142 ±1.246
Arginine and proline metabolism	ec00330	102.133 ±0.884	100.602 ± 0.933
Citrate cycle (Krebs cycle)	ec00020	116.843 ± 12.089	124.87 ± 0.702
Fatty acid biosynthesis	ec00061	303.491 ± 5.538	307.308 ± 0.489
Fatty acid elongation	ec00062	67.066 ± 8.073	71.807 ± 0.022
Folate biosynthesis	ec00790	302.951 ± 9.635	287.446 ± 9.319
Glycolysis	ec00010	145.131 ± 6.6	138.049 ± 11.634
Lysine degradation	ec00310	13.663 ± 3.617	8.986 ± 3.171
Mannose type O-glycan biosynthesis	ec00515	136.003 ± 20.316	152.586 ±6.335
Metabolism of xenobiotics by cytochrome P450	ec00980	69.32 ± 0.17	68.617 ±0.827
N-Glycan biosynthesis	ec00510	221.444 ± 2.738	227.992 ± 5.498
O-glycan biosynthesis	ec00514	162.416 ± 1.829	155.666 ± 8.056
Other glycan degradation	ec00511	177.914 ± 1.182	175.957 ± 4.27
Pantothenate and CoA biosynthesis	ec00770	24.598 ± 8.195	27.777 ± 0.525
Pentose phosphate	ec00030	102.537± 0.314	97.822 ± 9.699
Propanoate metabolism	ec00640	212.329 ± 1.465	201.314 ± 9.563
Pyrimidine metabolism	ec00240	172.185 ±4.223	181.393 ± 6.125
Sulfur metabolism	ec00920	33.591 ± 7.459	36.213 ± 2.483

Table 6: *P. leucopus* pathways activity level across infected and uninfected groups.

Pathway Name	**ID**	**Infected Mice Mean ± Std**	**Uninfected Mice Mean ± Std**
Arginine and proline metabolism	ec00330	108.443 ± 3.567	103.845 ± 1.015
D-Amino acid metabolism	ec00470	218.092 ± 0.626	206.601 ± 7.797
Glycerophospholipid metabolism	ec00564	78.228 ± 0.336	77.621 ± 0.172
Glycine, serine and threonine metabolism	ec00260	49.423 ± 0.728	47.543 ± 0.119
One carbon pool by folate	ec00670	66.566 ± 0.204	67.377 ± 0.301
Selenocompound metabolism	ec00450	103.557 ± 25.685	137.99 ± 8.249
Starch and sucrose metabolism	ec00500	64.353 ± 1.33	66.401 ± 0.433
Tryptophan metabolism	ec00380	98.223 ± 0.896	102.88 ± 0.892
ascorbate and aldarate metabolism	ec00780	24.271 ± 0.578	25.417 ± 0.049
Ascorbate and aldarate metabolism	ec00053	131.871 ± 1.17	136.458 ± 0.912
Citrate cycle	ec00020	116.276 ± 10.912	128.679 ± 0.663
Glycosaminoglycan biosynthesis-heparan sulfate/heparin	ec00534	85.392 ± 1.203	90.012 ± 1.656
Glycosaminoglycan biosynthesis-keratan sulfate	ec00533	351.816 ± 1.994	342.511 ± 1.023
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	ec00563	348.609 ± 1.349	353.073 ± 1.766
Linoleic acid metabolism	ec00591	440.035 ± 10.893	423.801 ± 1.7
Other glycan degradation	ec00511	164.744 ± 2.361	135.58 ± 0.722
Pentose phosphate	ec00030	103.646 ± 0.475	104.649 ± 0.247
Pyrimidine metabolism	ec00240	167.062 ± 0.407	179.749 ± 11.62
Valine, leucine and isoleucine biosynthesis	ec00290	77.081 ± 2.466	83.37 ± 2.5
Valine, leucine and isoleucine degradation	ec00280	113.366 ± 4.269	103.142 ± 5.56
Vitamin B6 metabolism	ec00750	56.675± 0.557	52.601 ± 0.395

Table 7: *P.leucopus* pathways activity level across infected and uninfected groups.

Pathway Name	**ID**	**Infected Mice Mean ± Std**	**Uninfected Mice Mean ± Std**
Amino sugar and nucleotide sugar metabolism	ec00520	104.8 ±1.365	102.262 ± 2.796
Arachidonic acid metabolism	ec00590	163.557 ± 0.317	162.903 ± 1.17
Nitrogen metabolism	ec00910	102.949 ± 0.324	101.743 ± 0.897
Folate biosynthesis	ec00790	314.768 ± 6.619	307.406 ± 1.934
Fructose and mannose metabolism	ec00051	30.991 ± 0.403	30.493 ± 0.193
Glutathione metabolism	ec00480	45.435 ± 0.73	44.655 ± 0.569
Glycosphingolipid biosynthesis-lacto & neolacto series	ec00601	29.256 ± 6.267	41.326 ± 6.14
Glyoxylate and dicarboxylate metabolism	ec00630	108.993 ± 11.861	120.784 ± 8.979
Inositol phosphate metabolism	ec00562	39.927± 0.154	39.575 ± 0.588
Porphyrin metabolism	ec00860	278.62 ± 1.556	275.075 ± 6.258
Riboflavin metabolism	ec00740	117.214 ± 8.465	105.181± 5.623
Steroid hormone biosynthesis	ec00140	131.347± 2.431	132.832 ± 0.644
Thiamine metabolism	ec00730	58.599 ± 0.158	58.15 ± 0.715
Tyrosine metabolism	ec00350	70.298 ± 2.634	66.036 ± 2.207
Ubiquinone and other terpenoid-quinone biosynthesis	ec00130	194.363 ± 4.996	201.709 ± 5.371

7 Figure Legends

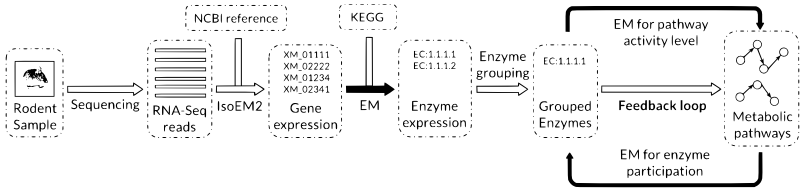

Figure 1: EMPathways2 pipeline for metabolic pathway analysis for the rodent samples. The RNA-Seq data obtained from the rodents are sequenced, then raw reads are mapped into genes. The genes obtained contigs are further mapped into the enzyme-pathway database. Gene expression is obtained using IsoEM2 (Mandric et al, 2017). Then, we estimate enzyme expression using gene expression. Finally, the the pathway activity level and enzyme participation coefficients are estimated in the feedback loop.

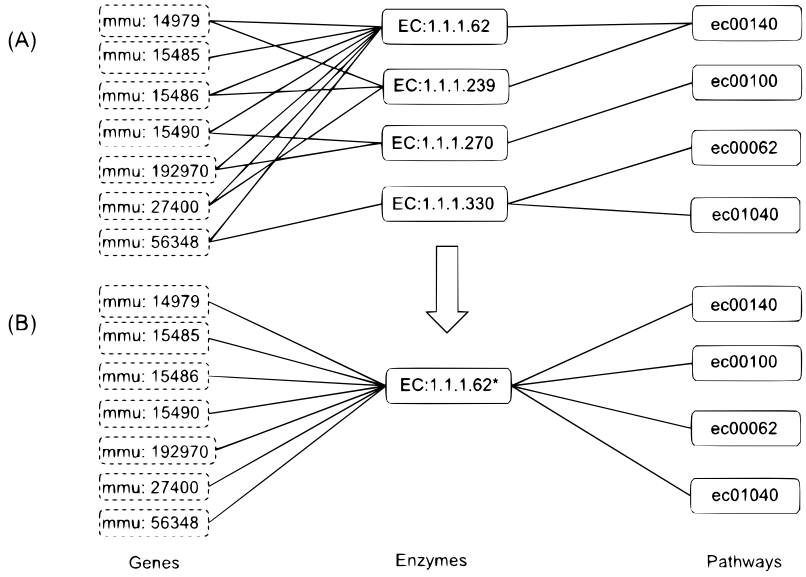

Figure 2: (A) Enzymes that cannot be distinguished from each other must be treated as groups. (B) Enzymes that are unstable are collapsed into a single enzyme with the lowest EC nomenclature number.