

# **Reanalysis of mouse-rat comparative gene expression data towards batch effects**

Maciej Bielecki

# Original papers

## Comparison of the transcriptional landscapes between human and mouse tissues

Shin Lin<sup>a,b,1</sup>, Yiing Lin<sup>c,1</sup>, Joseph R. Nery<sup>d</sup>, Mark A. Urich<sup>d</sup>, Alessandra Breschi<sup>e,f</sup>, Carrie A. Davis<sup>g</sup>, Alexander Dobin<sup>g</sup>, Christopher Zaleski<sup>g</sup>, Michael A. Beer<sup>h</sup>, William C. Chapman<sup>c</sup>, Thomas R. Gingeras<sup>g,i</sup>, Joseph R. Ecker<sup>d,j,2</sup>, and Michael P. Snyder<sup>a,2</sup>

(Lin et al., 2014)

**A reanalysis of mouse ENCODE comparative gene expression data [version 1; peer review: 3 approved, 1 approved with reservations]**

Yoav Gilad, Orna Mizrahi-Man

(Gilad, Mizrahi-Man, 2015)

# New dataset

## Data Descriptor: An RNA-Seq atlas of gene expression in mouse and rat normal tissues

Julia F. Söllner<sup>1,2</sup>, German Leparc<sup>1</sup>, Tobias Hildebrandt<sup>1</sup>, Holger Klein<sup>1</sup>, Leo Thomas<sup>3</sup>, Elia Stupka<sup>1</sup> & Eric Simon<sup>1</sup>

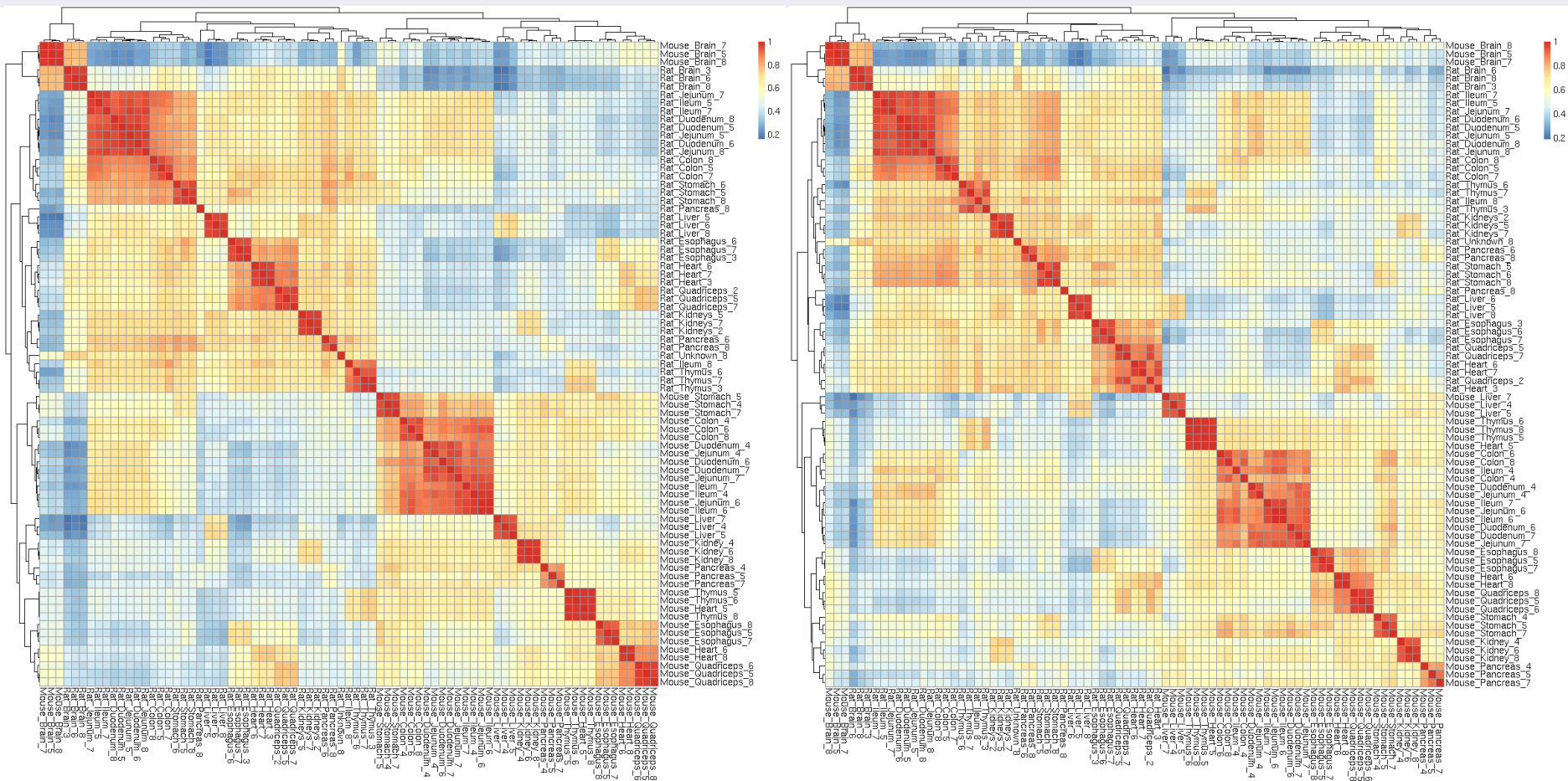
(Söllner et al, 2017)

# Sample count

- Original dataset: 47,531 mouse genes  
32,662 rat genes
- Upon pairing up homologues: 17945 gene pairs  
(38% and 55% passthrough rate respectively)
- Upon filtering: 12561 gene pairs  
(70% of highest-expression genes are used in reanalysis,  
rest discarded)

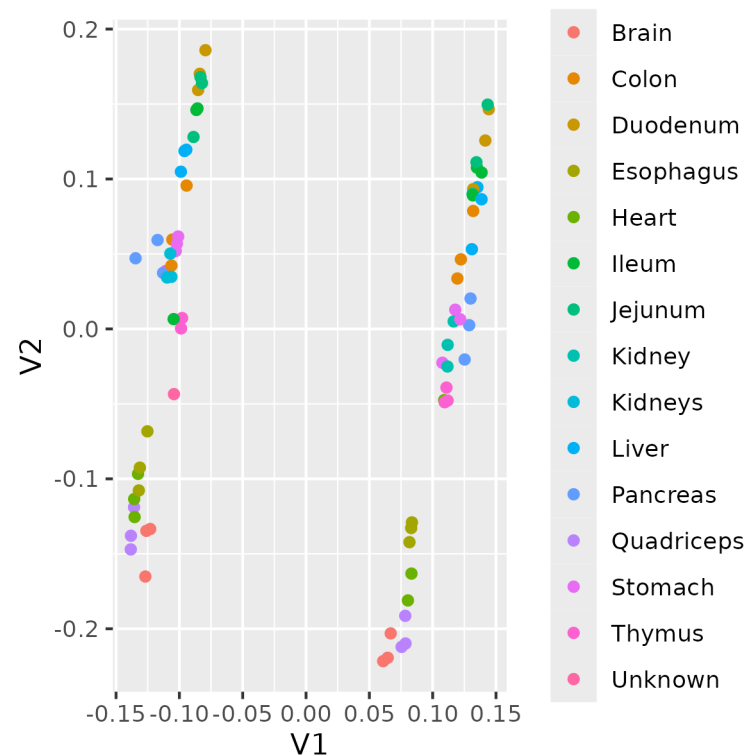
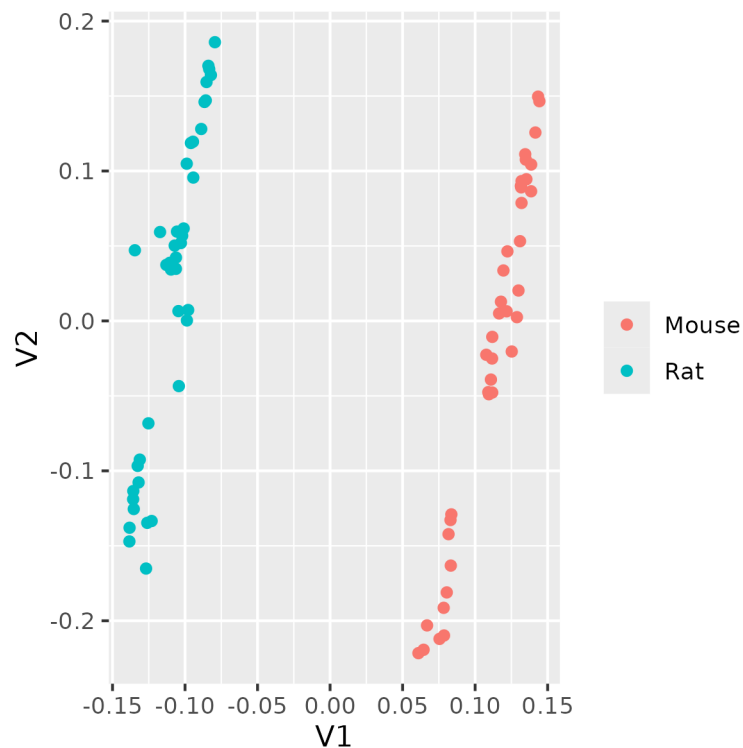
Overall, only 26% of intact dataset is used in reanalysis.

# Pre- and post-reanalysis comparison



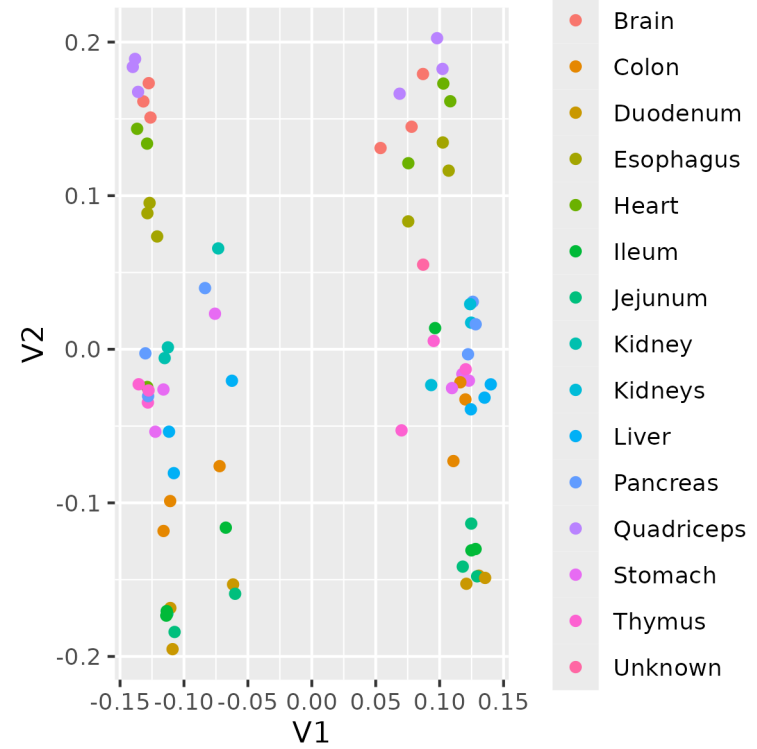
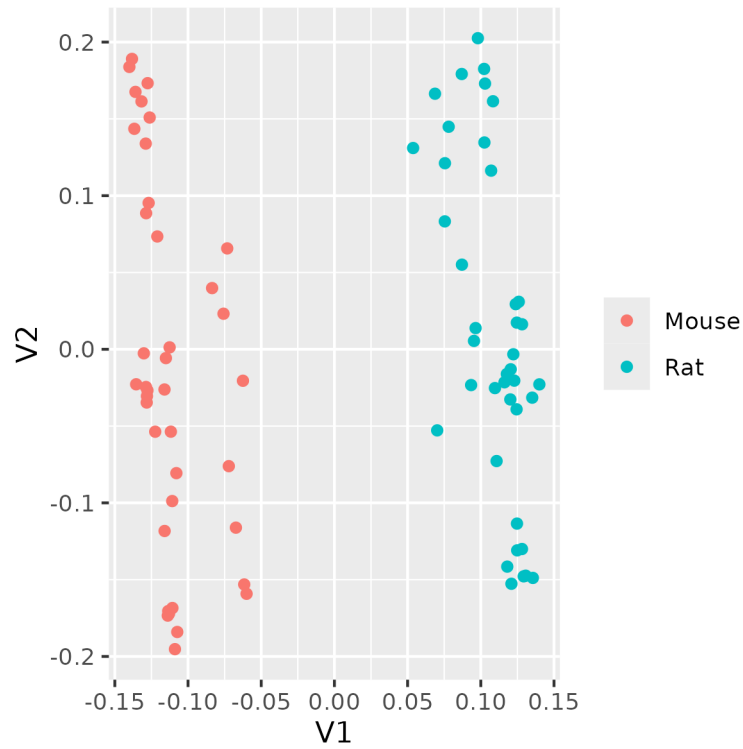
Pearson correlation heatmaps of data before (left) and after (right) reanalysis

# SVD plots: before reanalysis



Top 2 PCs of SVD in data before reanalysis, colored by species (left) and tissue (right)

# SVD plots: after reanalysis



Top 2 PCs of SVD in data after reanalysis, colored by species (left) and tissue (right)

# Conclusions

- Intra-over-inter clustering of samples appears genuine
- The top 2 PCs seem to highly associate with species and tissue, respectively
- Brain samples cluster separately to all others
- Low variability in types of tissue



# What's left?

- Trawl additional databases for more homologue pairs
- Try less-stringent filtering
- Statistical test for more definitive proof of intra-over-inter clustering
- Additional visualizations
- Compiling the final raport

# References

- Gilad Y, Mizrahi-Man O,  
A reanalysis of mouse ENCODE comparative gene expression data.  
*F1000Res* **4**(121), 2015.  
DOI: 10.12688/f1000research.6536.1
- Lin S, Lin Y, Nery JR et al.,  
Comparison of the transcriptional landscapes between human and mouse tissues.  
*Proc Natl Acad Sci U S A* **111**(48), 2014.  
DOI: 10.1073/pnas.1413624111
- Söllner J, Leparac G, Hildebrandt T et al.,  
An RNA-Seq atlas of gene expression in mouse and rat normal tissues.  
*Sci Data* **4**, 2017.  
DOI: 10.1038/sdata.2017.185