A MOUSE TISSUE ATLAS OF SMALL NON-CODING RNA

Alina Isakova[1], Stephen R. Quake[1,2,3]

[1]Department of Bioengineering, Stanford University, Stanford, California, USA
[2]Department of Applied Physics, Stanford University, Stanford, California, USA
[3]Chan Zuckerberg Biohub, San Francisco, California, USA

Correspondence
quake@stanford.edu.

SUMMARY

Small non-coding RNAs (sRNAs) play a vital role in a broad range of biological processes both in health and disease. A comprehensive quantitative reference of sRNA expression would significantly advance our understanding of sRNA roles in shaping tissue functions. Here, we systematically profiled the expression of four sRNA classes across eleven mouse tissues by RNA-seq. Using fourteen biological replicates spanning both genders, we identified 3,962 various sRNAs; 473 of these sRNAs are novel and identified for the first time by this work. We found that 40% of these transcripts were distributed across the body in a tissue-specific manner, and this tissue-specificity extends through multiple sRNA classes; furthermore, some sRNAs are also sexually dimorphic. By combining these findings with machine learning, we were able to accurately classify tissue types from sRNA data generated by other studies. These results yield the most comprehensive catalog of specific and ubiquitous small RNAs in individual tissues to date, and we expect that this catalog will be a resource for the further identification of sRNAs involved in tissue-function in health and dysfunction in disease.

INTRODUCTION

Small non-coding RNAs (sRNAs) are a large family of endogenously expressed transcripts, 18 to 40 nucleotides long, that play a crucial role in regulating cell function (Bartel, 2018; Cech and Steitz, 2014). Unknown to researchers two decades ago, today sRNAs are believed to be involved in nearly all developmental and pathological processes in mammals (Cech and Steitz, 2014; Esteller, 2011; He and Hannon, 2004; Ng et al., 2016). The main function of sRNAs in a cell is to tightly regulate gene expression at the levels of post-transcriptional RNA processing and translation. Aberrant expression of small sRNAs, in turn, has been associated with diseases such as cancer, autoimmune disease and several neurodegenerative disorders (Almeida et al., 2011; Liu et al., 2017; Somel et al., 2010).

Mammalian cells express several classes of sRNAs including microRNA (miRNA) (Ha and Kim, 2014), small interfering RNAs (siRNA), small nucleolar RNAs (snoRNA) (Matera et al., 2007), sno-derived RNA (sdRNA) (Taft et al., 2009), PIWI-interacting RNA (piRNA) (Ishizu et al., 2012), tRNA-derived small RNAs (Kumar et al., 2014), with some being shown to be expressed in a tissue- (Dittmar et al., 2006; Landgraf et al., 2007), cell-type-(Faridani et al., 2016; Volinia et al., 2006) or even cell-state-specific manner (Hayes et al., 2014; Palfi et al., 2016; Sherstyuk et al., 2017). Through their interactions with the messenger RNA (mRNA) that code for protein expression, these small non-coding molecules shape the dynamic molecular spectrum of tissues (Faridani et al., 2016; Sharma et al., 2016). Despite extensive knowledge of sRNA biogenesis and function (Bartel, 2018; Jorjani et al., 2016), much remains to be explored about tissue- and gender-specific sRNA expression. Given the emerging role of sRNA as biomarkers (Anfossi et al., 2018) and potent therapeutic targets (Janssen et al., 2013), a comprehensive reference catalog of tissue sRNA expression would represent a highly valuable resource not only for the fundamental but also for the clinical research.

The first attempts to establish a catalog of tissue-specific mammalian sRNAs began a decade ago (Ach et al., 2008; Hsu et al., 2007; Landgraf et al., 2007; Liang et al., 2007). While these pioneering microarray-, qPCR- and Sanger sequencing-based studies mapped only a limited number of highly expressed miRNA, they, nevertheless, established a “gold standard” reference for the upcoming 10 years of miRNA research. Efforts to characterize tissue-specific sRNA have recently resumed with employment of RNA-seq, which greatly advanced the discovery of novel and previously undetected low expressed transcripts (Londin et al., 2015; McCall et al., 2017; de Rie et al., 2017). However, cataloging the tissue-specific patterns of miRNA expression was mostly done using publicly available RNA-seq data originating from various experiments designed to target a tissue or a condition of interest rather than systematically analyzing a spectrum of normal tissues from the same individual (Ludwig et al., 2016; McCall et al., 2017; de Rie et al., 2017). As a result, tissue expression patterns reported by these studies are mainly based on one replicate, suffer from protocol biases (Giraldez et al., 2018) and, still yield an incomprehensive picture of bodily miRNA patterns. In addition, none of

the prior studies encompass a spectrum of mammalian tissues from both female and male individuals, let alone, various non-coding RNA species other than miRNA.

Here, we describe a comprehensive atlas of sRNA expression across eleven normal mouse tissues. Using multiple biological replicates (n=14), we mapped tissue-specific as well as broadly transcribed sRNA attributed to different classes and spanning a large spectrum of expression levels. We have also, for the first time, provided evidence of gender-based differences in sRNA expression across multiple tissues. Finally, we used machine learning to accurately predict the tissue and, in certain cases, even its functionality based on sRNA expression.

RESULTS

sRNA expression atlas of mouse tissues

We profiled the expression of sRNA across eleven tissues from adult female (n=10) and male (n=4) C57BL/6J mice (**Figure 1A, Table S1**). We generated a dataset comprising in total of 142 sRNA sequencing libraries from brain, lung, heart, muscle, kidney, pancreas, liver, small intestine, spleen, bone marrow and testes RNA. Each library yielded ~1-10 million sRNA reads, resulting in the average of 21 million sRNA reads per tissue (**Figure S1A**). Using the ENCODE GRCm38 annotation, we mapped the expression of four distinct sRNA classes: miRNA, snRNA, snoRNA and scaRNA, in profiled tissues. Among all the tissues we identified 1551 distinct pre-miRNA, 941 snRNA, 953 snoRNA and 44 scaRNA, which corresponds to 70.3% 67.9%, 63.2% and 86.3% of ENCODE-annotated transcripts of the respective class (**Figure 1B**). With respect to protein coding genes, the majority of detected snoRNA were of intronic origin, snRNA and scaRNA were intronic and intergenic (63/35% and 64/36% respectively) and miRNA were transcribed from either introns (53%), exons (11%) or intergenic regions (11%) (**Figure S1B**). The number of distinct sRNA greatly varied across tissues, for example, spleen and lung contained the largest number of distinct sRNA (~ 1300) while pancreas and liver – the lowest (~ 500) (**Figure S1C**). Furthermore, within the profiled tissues we detected 95.1% of pre-miRNA denoted by miRBase v22 database as high confidence transcripts (Kozomara and Griffiths-Jones, 2014). Using the obtained data we have reconstructed the most complete genome-wide tissue map of mammalian sRNA expression (**Figure 1C, Table S2**).

Tissue-specific expression of sRNA

We first assessed the differences in sRNA patterns across profiled tissues based on the expression of all four sRNAs classes (**Figure 2A**). Unsupervised clustering of most variable sRNAs across all samples (*Methods*) demonstrated that a large number of sRNA transcripts are shared among brain and pancreas, while not being expressed in other tissues (**Figure 2B and Figure S2**). Similar patterns were observed for bone marrow and spleen. Clustering of the most variable transcripts within each sRNA class allowed us to further identify miRNA to be the

main driver of brain and pancreas similarity, while bone marrow and spleen clustered together within snoRNAs **(Figure S2)**.

Dimensional reduction via t-distributed stochastic neighbor embedding (tSNE) (Laurens Van Der Maaten and Geoffrey Hinton, 2008) (*Methods*) on all sRNA genes revealed a robust clustering of samples according to tissue types (**Figure 2C**). Further analysis, performed on each sRNA class separately (**Figure S3A**), showed that out of all four sRNA classes miRNA separated samples in eleven clearly identifiable clusters corresponding to profiled tissues, while snRNA showed no clear separation by tissue type. snoRNA and scaRNA tSNEs resolved the majority of tissue types, however failed to identify pancreas, intestine and kidney.

For each sRNA we next computed the tissue specificity index, *TSI*, as described previously in (Ludwig et al., 2016) (**Table S2**). We observed that ~ 21% of all detected sRNA were expressed in only one tissue (*TSI* = 1), ~ 19% of all sRNA were highly expressed in one but also present in other tissues (0.95 < *TSI* < 1) while the remaining ~60% were either ubiquitously expressed or had high expression scores in multiple tissues (**Figure S3B**). Brain contains the highest number of tissue-specific transcripts (437 with *TSI* > 0.95) followed by lung (257 with *TSI* > 0.95), bone marrow (161) and spleen (152) (**Figure 2D**). Interestingly, despite of the lowest number of transcript counts across profiled tissues, lung contained the largest number of distinct sRNAs (2532), followed by spleen (2430) and brain (2081) (**Figure 2D**). At the level of individual RNA class, brain, lung, spleen and bone marrow remained the top four tissues harboring the largest number of unique tissue-specific transcripts (**Figure S3C**), except for scaRNA, for which only few transcripts were found to have *TSI*>0.95.

Tissue-specific miRNAs

Comparing the expression of miRBase-annotated (Kozomara and Griffiths-Jones, 2014) pre-miRNAs across all eleven tissues, we identified well-described tissue-specific miRNAs, such as miR-122, miR-375, miR-10a enriched respectively in liver, pancreas and kidney (Landgraf et al., 2007) to be among the most variable pre-miRNA (**Figure 3A and Figure S4**). We also found a large number of highly expressed miRNA previously unknown to preferentially localize within a particular tissue; these include miR-6236 in the bone marrow, miR-194 and miR-215 – in the intestine, miR-381 – in the brain, miR-203 and miR-23b – in testes (**Figure 3A and Figure S4**). In addition, we found several low-expressed miRNAs specific to either one or two tissues (**Figure 3B and 3C**). Moreover, among the top fifty miRNAs with highest expression scores in each tissue, only the minority of transcripts appeared to be tissue specific (*TSI>0.9*) while the majority were also expressed in other ten tissues (**Figure S4**).

We next asked how the identified tissue expression patterns compare to those of individual cell types. To investigate that, we correlated our data to the miRNA-seq data generated for primary mammalian cells by FANTOM5 consortium (de Rie et al., 2017). Comparing mouse samples

first, we found that FANTOM5 embryonic and neonatal cerebellum tissues strongly correlated with our brain samples (r_s=0.89-0.9), while erythroid cells had the strongest correlation with spleen and bone marrow (r_s=0.93) (**Figure S5A**). To perform a comparison with human samples, we focused on the expression scores of 531 orthologs detected in both the current study and the FANTOM5 samples (**Figure S5B**). Spearman correlation coefficients reflected the cell-type composition of tissues (**Figure S5C**). As such, we observed that mouse bone marrow and spleen had the highest correlation with human B-cells, T-cells, dendritic cells and macrophages ($0.5<r_s<0.6$), muscle correlated the most with myoblasts and myotubes (r_s=0.47), while brain – with neural stem cells, spinal cord, pineal and pituitary glands (r_s=0.49) **(Figure S5C)**.

Tissue-specific snoRNAs

Given the ability of snoRNA to separate the majority of profiled tissues based on the expression (**Figure S3A**), we have next focused on identifying the most variable snoRNAs across tissues. Among the verified snoRNA we detected sixty that varied across tissues (**Figure 3D,** *Methods*). Top variable snoRNAs included Snora35 and Snord116, which were highly expressed in the brain and previously shown to be specific to neural tissues (Cavaille et al., 2000). Bone marrow contained the largest fraction of tissue-specific snoRNAs, while a few were also present in spleen, pancreas, liver and testis. Importantly, we found that Snord70 and Snord66, often used as normalization controls in qPCR-based assays (Chen et al., 2012; Emde et al., 2015), are also expressed in a tissue specific manner. Another example of identified tissues-specific snoRNA is Snord123, located 3kb upstream of pancreatic cancer-associated *Sema5a* gene, we found it to be expressed predominantly in the pancreas. (**Figure 3E**). We also discovered several snoRNAs whose existence had been previously predicted but which had not yet been detected experimentally to be specifically expressed across profiled tissues (**Figure 3E**).

miRNAs are expressed in a gender-specific manner

To address a long standing question of gender bias in miRNA expression (Guo et al., 2017; Kolhe et al., 2017), we compared the miRNA expression levels between female and male mice. In each somatic tissue, except pancreas, we identified at least two miRNAs to be differentially expressed (log2FoldChange > 1, normalized counts > 100) at FDR < 0.01 between genders (**Figure 4, Figure S6**). Kidney and lung contained the highest number of gender-biased miRNAs (27 and 18 respectively), while only two were detected in the heart, five in the muscle and seven in the brain **(Figure S7A).** Among the identified gender-biased miRNAs, we found miR-411, miR-186, miR-340, miR-182, miR-183, miR-148a, miR-145a, miR-101b to be systematically overrepresented in female compared to male tissues and miR-379, miR-195a, miR-99a, miR-let-7g, miR-666, miR-15b, miR-151 to be expressed higher in male over female tissues (**Figure S7B**). Interestingly, three out of eight female-dominant miRNAs: miR-182, miR-148a and miR-145a, were also shown previously to be estrogen regulated (Klinge, 2009) while

another miRNA, miRNA-340, was reported to be downregulated in response to elevated androgen levels (Fletcher et al., 2014).

Given the innate ability on miRNA to lower the levels of target mRNA (Guo et al., 2010), we hypothesized that the levels of protein-coding transcripts targeted by gender-biased miRNA would also differ across male and female tissues. To test this hypothesis we correlated the expression of gender-biased miRNAs with the levels of their respective target mRNAs across profiled tissues (*Methods*). Among the anticorrelated targets (r_s < -0.8, FDR < 0.1) we identified two genes previously shown to be sexually dimorphic (**Figure S8C**). Specifically, we found miR-423, upregulated in male tissues, to negatively correlate with its target – estrogen-related receptor gamma (*Esrrg*) (r_s= -0.9, FDR < 0.1), and female-specific miR-340 to negatively correlate with androgen-associated ectodysplasin A2 receptor (*Eda2r*) (Prodi et al., 2008). However, we also found that the majority of the predicted targets were in fact positively correlated with the respective miRNAs (r_s > 0.8, FDR < 0.1), suggesting that the tested miRNAs are not involved in silencing of these targets through degradation.

Novel miRNAs

To search for novel miRNA in our data, unmapped reads were analyzed with the miRDeep2 framework (Friedländer et al., 2012). We identified 473 novel miRNAs supported by at least five sequencing reads, with the majority being present in only one tissue (312), but a small number (4) being found in all eleven tissues (**Figure 5A**). Principal component analysis on the newly identified miRNAs, supported by at least 50 reads, showed a clear separation of brain, lung and muscle from other tissues based on the expression values. Similar to annotated transcripts, novel miRNAs demonstrate a spectrum of tissue specificity with some being ubiquitously expressed, while others are only present in one tissue (**Figure S8A**). Differential expression analysis on putative novel miRNAs identified six miRNAs to be also expressed in a gender-specific manner. Strikingly, all six were male-dominant, with one of them even found to be consistently upregulated in two tissues, male muscle and pancreas (**Figure 5B and S8C**). We speculate that the prevalence of male-specific novel miRNAs identified in our study reflects the inconsistent sampling of both genders by prior murine miRNA research.

miRNA-based tissue classification

We finally asked whether the observed variation in miRNA expression across tissues (**Figure 2B and C**) would be sufficient to accurately predict the tissue type based solely on miRNA-seq data. To address this question we set out to construct an algorithm that can learn tissue characteristics from the data reported in the current study and make predictions on new data sets. We first trained a support vector machine (SVM) model (Cortes and Vapnik, 1995) on 134 data sets generated in this study, each containing the expression scores for 1973 miRNAs (**Figure S9A**). As a validation dataset we used available miRNA-seq data released by the ENCODE consortium for multiple mouse tissues (Dunham et al., 2012). Notably, the ENCODE

datasets contained data generated for the postnatal and embryonic life stages, as opposed to the adult stage profiled in the current study (**Table S4**). Nonetheless, our SVM model accurately classified postnatal forebrain, midbrain, hindbrain and neural tube as brain tissue, as well as accurately inferred the tissue types for heart, intestine, kidney, liver, muscle samples, yielding an overall accuracy of 0.96 (*Methods*). For the embryonic tissues, however, our model was able to only reach an accuracy of 0.69. This was mainly due to inability of the model to correctly classify liver tissues and instead assigning them to bone marrow (**Figure 6A**). Strikingly, in this case our model accurately predicted the hematopoietic function of the organ, known to shift from the liver at the embryonic stages to the bone marrow in adulthood (Baron et al., 2012), rather than the tissue type itself. Furthermore, we identified hematopoiesis-associated miR-150 and miR-155 (Bissels et al., 2012) to have highest weights among the features defining the bone marrow in our model (**Figure S9B**).

The random forest (RF) regression model (Tin Kam Ho, 1998) weighted miRNA differently **(Figure S9C)** and was able to predict the postnatal and embryonic tissues with respective accuracies 1.0 and 0.9, guessing correctly 27/27 and 44/49 samples and misassigning one of the embryonic kidney, lung, liver and intestine samples (**Figure 6B**).

DISCUSSION

Small non-coding RNA plays an indispensable role in shaping cellular identity by altering the levels of protein-coding transcripts (Matera et al., 2007; Qureshi and Mehler, 2012). Recent efforts in profiling the miRNA content of cells and tissues demonstrated the existence of tissue- and cell type-specific short non-coding transcripts. (Landgraf et al., 2007; Londin et al., 2015; Ludwig et al., 2016; McCall et al., 2017; de Rie et al., 2017). In this work, we show that this phenomenon extends beyond one sRNA class and involves not only tissue-specific but also gender-specific sRNA expression. Compared to previous studies, we use a large number of biological replicates (n=14) and focus on deriving a quantitative rather than qualitative reference of sRNA expression across eleven normal murine tissues. We demonstrate that the obtained quantitative data can be used to train machine learning algorithms to recognize tissues or even their functions based on sRNA expression.

We found a large number of new tissue-specific sRNA missed by previous studies (**Figure 3 and S4, Table S2**) because of low sequencing coverage (Landgraf et al., 2007) or due to insignificant expression scores derived from a single biological replicate (McCall et al., 2017). By analyzing the expression of several classes of sRNA we discovered a large number of snoRNA expressed mainly in the brain, pancreas or bone marrow (**Figure 3D**). Taken together with previous observations (Jorjani et al., 2016), this finding raises additional questions regarding the biogenesis pathways of snoRNA as well as its potential specialized functions across distinct tissues.

More than 60% of protein-coding genes in mammalian genome harbor predicted miRNA target sites (Friedman et al., 2008). However, only a handful of them are *bona fide* miRNA targets, while up to 70% are falsely assigned by prediction algorithms (Agarwal et al., 2015; Betel et al., 2010; Dweep and Gretz, 2015). Currently, the validation of miRNA:mRNA interactions is still mainly based on low-throughput labor-intensive approaches such as knock-down or over-expression assays (Thompson et al., 2015) and thus has been only done for a limited number of miRNAs. Growing amounts of both, mRNA- and miRNA-seq data, generated for various cell and tissue types, now provide us with a possibility to narrow down the list of putative targets by identifying those that go down with elevated miRNA levels (Guo et al., 2010). Here, using our tissue miRNA-seq dataset and publicly available mRNA-seq data, we demonstrate the implementation of this approach. By correlating the expression of miRNA:target pairs across tissues, we show that at least half of the predicted targets are not affected by increasing miRNA levels. In parallel, we were able identify targets that show a strong negative correlation with miRNAs. On the example of sexually dimorphic miRNA we further demonstrate that the levels of some protein-coding transcripts indeed decrease with increased miRNA levels (**Figure S6C**), suggesting miRNAs to contribute to gender-biases in gene expression.

Our work, to our knowledge, is the first to demonstrate that the expression of short non-coding transcripts can be used to accurately predict tissue types (**Figure 6**). Within the current study we show that machine learning algorithms applied to quantitative sRNA expression estimates yield robust tissue classifiers applicable to data generated by other groups. Given the emerging evidence of strong miRNA dysregulation in disease (Mendell and Olson, 2012), particularly in cancers (Jansson and Lund, 2012; Peng and Croce, 2016), we anticipate that in the future sRNA-based classifiers could be expanded to recognize affected tissues or various disease types.

sRNAs have been long known to regulate development and functions of the brain (Qureshi and Mehler, 2012). Our study finds that brain, in fact, contains the largest number of unique mammalian sRNA transcripts that are absent in other tissues. We found that lung also contains a large number of tissue-specific sRNAs, and expresses the largest number of distinct sRNA among eleven profiled tissues (**Figure 2D**). Given the complexity of tissues, one would expect that the majority of specific transcripts are residing within a particular cell type or state uniquely present in the tissue. However, our knowledge of cell-specific miRNA expression is not complete and does not yet allow us to identify all the cell types driving the complexity (McCall et al., 2017; de Rie et al., 2017). We directly observed this phenomenon in miR-92b, miR-448 and miR-1298, which we identified as expressed in both, brain and lung tissues (**Figure 3B and S4**). According to the cell-type-based studies, however, these transcripts are specific to neural and stem cells (de Rie et al., 2017), which explains why we see them in the brain, but does not explain their presence in the lung. This inability to fully explain the roots of tissue complexity elucidates the need for further characterization of the sRNAs content of specific cell types or even, similarly to mRNA, that of single cells (Faridani et al., 2016; Trapnell, 2015). This

atlas, meanwhile, will set a solid pillar for future sRNAs studies and will serve as a powerful resource of sRNA tissue identity for fundamental and clinical research.

Acknowledgements

We thank Dylan Henderson for assistance in RNA extraction and library preparation. Norma Neff and Jennifer Okamoto for sequencing expertise. Geoff Stanley and Kiran Kocherlakota for kind advice in tissue dissection and preservation. This study was supported by Howard Hughes Medical Institute and Chan Zuckerberg Biohub. A.I. was supported by the Swiss National Foundation Early PostDoc Mobility Fellowship.

Author Contributions

A.I. designed and performed the experiments and data analysis. A.I. and S.Q. interpreted the data and wrote the manuscript.

Competing interest

The authors declare no conflict of interest.

REFERENCES

Ach, R.A., Wang, H., and Curry, B. (2008). Measuring microRNAs: Comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol. *8*, 69.

Agarwal, V., Bell, G.W., Nam, J.-W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. ELife *4*.

Almeida, M.I., Reis, R.M., and Calin, G.A. (2011). MicroRNA history: Discovery, recent applications, and next frontiers. Mutat. Res. Mol. Mech. Mutagen. *717*, 1–8.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anfossi, S., Babayan, A., Pantel, K., and Calin, G.A. (2018). Clinical utility of circulating non-coding RNAs — an update. Nat. Rev. Clin. Oncol.

Baron, M.H., Isern, J., and Fraser, S.T. (2012). The embryonic origins of erythropoiesis in mammals. Blood *119*, 4828–4837.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol. *11*, R90.

Bissels, U., Bosio, A., and Wagner, W. (2012). MicroRNAs are shaping the hematopoietic landscape. Haematologica *97*, 160–167.

Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.-P., Brosius, J., and Huttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. *97*, 14311–14316.

Cech, T.R., and Steitz, J.A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. Cell *157*, 77–94.

Chen, P.-Y., Qin, L., Barnes, C., Charisse, K., Yi, T., Zhang, X., Ali, R., Medina, P.P., Yu, J., Slack, F.J., et al. (2012). FGF Regulates TGF-β Signaling and Endothelial-to-Mesenchymal Transition via Control of let-7 miRNA Expression. Cell Rep. *2*, 1684–1696.

Cortes, C., and Vapnik, V. (1995). Support-vector networks. Mach. Learn. *20*, 273–297.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-Specific Differences in Human Transfer RNA Expression. PLoS Genet. *2*, e221.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Dweep, H., and Gretz, N. (2015). miRWalk2.0: a comprehensive atlas of microRNA-target interactions. Nat. Methods *12*, 697–697.

Emde, A., Eitan, C., Liou, L.-L., Libby, R.T., Rivkin, N., Magen, I., Reichenstein, I., Oppenheim, H., Eilam, R., Silvestroni, A., et al. (2015). Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: a new mechanism for ALS. EMBO J. *34*, 2633–2651.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat. Rev. Genet. *12*, 861–874.

Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. Nat. Biotechnol.

Fletcher, C.E., Dart, D.A., and Bevan, C.L. (2014). Interplay between steroid signalling and microRNAs: implications for hormone-dependent cancers. Endocr. Relat. Cancer *21*, R409–R429.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. *40*, 37–52.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2008). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Giraldez, M.D., Spengler, R.M., Etheridge, A., Godoy, P.M., Barczak, A.J., Srinivasan, S., De Hoff, P.L., Tanriverdi, K., Courtright, A., Lu, S., et al. (2018). Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. Nat. Biotechnol.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2007). miRBase: tools for microRNA genomics. Nucleic Acids Res. *36*, D154–D158.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guo, L., Zhang, Q., Ma, X., Wang, J., and Liang, T. (2017). miRNA and mRNA expression analysis reveals potential sex-biased miRNA expression. Sci. Rep. *7*.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. *15*, 509–524.

Hayes, J., Peruzzi, P.P., and Lawler, S. (2014). MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol. Med. *20*, 460–469.

He, L., and Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. *5*, 522–531.

Hsu, S.-D., Chu, C.-H., Tsou, A.-P., Chen, S.-J., Chen, H.-C., Hsu, P.W.-C., Wong, Y.-H., Chen, Y.-H., Chen, G.-H., and Huang, H.-D. (2007). miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Res. *36*, D165–D169.

Ishizu, H., Siomi, H., and Siomi, M.C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. *26*, 2361–2373.

Janssen, H.L.A., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patick, A.K., Chen, A., Zhou, Y., et al. (2013). Treatment of HCV Infection by Targeting MicroRNA. N. Engl. J. Med. *368*, 1685–1694.

Jansson, M.D., and Lund, A.H. (2012). MicroRNA and cancer. Mol. Oncol. *6*, 590–610.

Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavolan, M., and Gruber, A.R. (2016). An updated human snoRNAome. Nucleic Acids Res. *44*, 5068–5082.

Klinge, C. (2009). Estrogen Regulation of MicroRNA Expression. Curr. Genomics *10*, 169–183.

Kolhe, R., Hunter, M., Liu, S., Jadeja, R.N., Pundkar, C., Mondal, A.K., Mendhe, B., Drewry, M., Rojiani, M.V., Liu, Y., et al. (2017). Gender-specific differential expression of exosomal miRNA in synovial fluid of patients with osteoarthritis. Sci. Rep. *7*, 2029.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68–D73.

Kuhn, M. (2018). caret: Classification and Regression Training.

Kumar, P., Anaya, J., Mudunuri, S.B., and Dutta, A. (2014). Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. BMC Biol. *12*.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. Cell *129*, 1401–1414.

Laurens Van Der Maaten, and Geoffrey Hinton (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. 2579--2605.

Liang, Y., Ridzon, D., Wong, L., and Chen, C. (2007). Characterization of microRNA expression profiles in normal human tissues. BMC Genomics *8*, 166.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liu, C., Liu, R., Zhang, D., Deng, Q., Liu, B., Chao, H.-P., Rycaj, K., Takata, Y., Lin, K., Lu, Y., et al. (2017). MicroRNA-141 suppresses prostate cancer stem cells and metastasis by targeting a cohort of pro-metastasis genes. Nat. Commun. *8*, 14270.

Londin, E., Loher, P., Telonis, A.G., Quann, K., Clark, P., Jing, Y., Hatzimichael, E., Kirino, Y., Honda, S., Lally, M., et al. (2015). Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc. Natl. Acad. Sci. *112*, E1106–E1115.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E., et al. (2016). Distribution of miRNA expression across human tissues. Nucleic Acids Res. *44*, 3865–3877.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

McCall, M.N., Kim, M.-S., Adil, M., Patil, A.H., Lu, Y., Mitchell, C.J., Leal-Rojas, P., Xu, J., Kumar, M., Dawson, V.L., et al. (2017). Toward the human cellular microRNAome. Genome Res. *27*, 1769–1781.

Mendell, J.T., and Olson, E.N. (2012). MicroRNAs in Stress Signaling and Human Disease. Cell *148*, 1172–1187.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071) (TU Wien).

Ng, K.W., Anderson, C., Marshall, E.A., Minatel, B.C., Enfield, K.S.S., Saprunoff, H.L., Lam, W.L., and Martinez, V.D. (2016). Piwi-interacting RNAs in cancer: emerging functions and clinical utility. Mol. Cancer *15*.

Palfi, A., Hokamp, K., Hauck, S.M., Vencken, S., Millington-Ward, S., Chadderton, N., Carrigan, M., Kortvely, E., Greene, C.M., Kenna, P.F., et al. (2016). microRNA regulatory circuits in a mouse model of inherited retinal degeneration. Sci. Rep. *6*.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A.G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res. *41*, W169–W173.

Peng, Y., and Croce, C.M. (2016). The role of MicroRNAs in human cancer. Signal Transduct. Target. Ther. *1*.

Pennisi, E. (2012). ENCODE Project Writes Eulogy for Junk DNA. Science *337*, 1159–1161.

Prodi, D.A., Pirastu, N., Maninchedda, G., Sassu, A., Picciau, A., Palmas, M.A., Mossa, A., Persico, I., Adamo, M., Angius, A., et al. (2008). EDA2R Is Associated with Androgenetic Alopecia. J. Invest. Dermatol. *128*, 2268–2270.

Qureshi, I.A., and Mehler, M.F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat. Rev. Neurosci. *13*, 528.

Revelle, W. psych: Procedures for Personality and Psychological Research (Northwestern University).

de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M., et al. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat. Biotechnol. *35*, 872–878.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.

Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science *351*, 391–396.

Sherstyuk, V.V., Medvedev, S.P., Elisaphenko, E.A., Vaskova, E.A., Ri, M.T., Vyatkin, Y.V., Saik, O.V., Shtokalo, D.N., Pokushalov, E.A., and Zakian, S.M. (2017). Genome-wide profiling and differential expression of microRNA in rat pluripotent stem cells. Sci. Rep. *7*.

Somel, M., Guo, S., Fu, N., Yan, Z., Hu, H.Y., Xu, Y., Yuan, Y., Ning, Z., Hu, Y., Menzel, C., et al. (2010). MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. Genome Res. *20*, 1207–1218.

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. RNA N. Y. N *15*, 1233–1240.

Thompson, D., Regev, A., and Roy, S. (2015). Comparative Analysis of Gene Regulatory Networks: From Network Reconstruction to Evolution. Annu. Rev. Cell Dev. Biol. *31*, 399–428.

Tin Kam Ho (1998). The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. *20*, 832–844.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Volinia, S., Calin, G.A., Liu, C.-G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl. Acad. Sci. U. S. A. *103*, 2257–2261.

Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. *43*, D146–D152.

FIGURE LEGENDS

Figure 1. sRNA expression across mouse tissues. (A) Tissues and sRNA classes profiled in the current study. Ten somatic tissues were collected from adult mice (n=14). Testes were collected from male individuals (n=4). **(B) sRNAs identified in the current study.** High confidence pre-miRNA were retrieved from miRbase v22**. (C) Genomic map of sRNAs expression across mouse genome.** Circles represent sRNAs. Size of each circle corresponds to the log-transformed normalized expression counts of that sRNA. sRNAs, expression of which fell within the first quartile of all expression values, are depicted in blue, the rest are shown in red. Bars around the circle represent the variance of each sRNA across ten mouse tissues. Green and grey bars denote variability of each sRNAs across tissue types. Green denote highly (the standard deviation of expression above 25% of the mean value) and grey - low variable sRNAs (standard deviation below 25% of the mean value).

Figure 2. Tissue-specific patterns of sRNA expression. (A) Average number of sRNA found in each tissue. **(B) Heatmap** of normalized log-transformed RNA-seq counts of ~300 sRNAs detected in various mouse tissues. To obtain the miRNA expression estimates presented in the heatmap, we have first mapped the raw sequencing reads to the mouse genome using *STAR* (Dobin et al., 2013), then computed and normalized the sRNA counts using *FeatureCounts* (Anders et al., 2015) and *DESeq2* (Love et al., 2014) packages respectively. **(C) t-SNE projection of sRNA expression patterns** performed on ~3000 sRNA genes detected in eleven mouse tissues. **(D) Number of tissue-specific sRNAs detected in each tissue.** Size of the circle corresponds to the average number of normalized counts across each tissue.

Figure 3. Tissue specificity of mi- and snoRNAs. (A) Top 60 most variable miRNAs across profiled tissues. Heatmap of normalized log-transformed RNA-seq counts detected in various mouse tissues. **(B) Examples of miRNAs** previously unknown to be specific to a certain tissue. **(C) Examples of miRNAs** expressed only in two profiled tissues. **(D) Top 60 most variable snoRNAs across profiled tissues. (E) Examples of snoRNAs** previously unknown to be specific to a certain tissue.

Figure 4. Gender-specific miRNA expression. Volcano plots showing miRNAs differentially expressed between female and male brain, kidney, muscle, bone marrow and heart.

Figure 5. Novel miRNA detected in the present study. (A) Distribution of novel miRNAs across mouse tissues. (B) Volcano plots showing novel miRNAs differentially expressed between female and male tissues. **(C) Examples** of tissue-specific novel miRNAs.

Figure 6. miRNA-based tissues classifier. Confusion matrices obtained for **(A) SVM** and **(B) RF** tissue predictions for postnatal and embryonic datasets.

STAR Methods

Subject details

Animals

All procedures followed animal care and biosafety guidelines approved by Stanford University's Administrative Panel on Laboratory Animal Care and Administrative Panel of Biosafety. Wild type C57BL/6J mice, 4 males and 10 females, aged ~3 month old were used (**Table S1**).

Methods details

Tissue handling and RNA extraction

Upon collection, tissue samples were submerged and preserved at -80C in RNAlater stabilization solution (ThermoFisher cat # AM7021) until further processing. Total RNA was isolated from ~ 100 mg of tissue using Qiagen miRNeasy mini kit (cat # 217004) and the Qiagen tissue lyser using 5 mm stainless steel beads. RNA integrity was assesses using Agilent Bioanalyzer using RNA 6000 pico kit (Agilent Technologies cat # 5067-1513).

Library preparation and sequencing

Short RNA libraries were prepared following the Illumina TruSeq Small RNA Library Preparation kit (cat # RS-200-0012, RS-200-0024, RS-200-0036, RS-200-0048) according to the manufacturer's protocol and size-selected using Pippin Prep 3% Agarose Gel Cassette (Safe Science) in a range 135 bp – 210 bp. Samples were pooled in batches of 48 and sequenced using the Illumina NextSeq500 instrument in a single-read, 50 or 75 -base mode.

Data processing

Sequencing reads were demultiplexed by BaseSpace (Illumina). Reads were trimmed from the adaptor sequences and aligned to the mouse genome (GRCm38) using STAR_v2.5.1 (Dobin et al., 2013) with the following parameters --outFilterMismatchNoverLmax 0.05 --outFilterMatchNmin 16 --outFilterMatchNminOverLread 0 --outFilterScoreMinOverLread 0 --alignIntronMax 1 --outMultimapperOrder Random --clip3pAdapterSeq TGGAATTCTC --clip3pAdapterMMp 0.1 . Spliced alignments and hard/soft-clipping were disabled. Reads mapping with insertions or deletions were removed. We used ENCODE GRCm38 and miRBase v22 annotations to count the number of sRNA transcripts. Reads were assigned to one of the annotated biotypes: miRNAs, snoRNAs, snRNAs and scaRNA from ENSEMBL or to pre-miRNAs from miRBase v22 using *featureCounts v 1.6.1* (Liao et al., 2014). We first counted reads from both intronic and exonic regions of the protein coding and lincRNAs attempting to capture small RNAs transcribed from these regions (**Figure S1A and B**), using the following command: f*eatureCounts* -a Mus_musculus.GRCm38.90.gtf -M –primary -s 0. We used a –M -primary option to count reads mapping to multiple genomic locations marked as "primary alignment" by STAR. Next, we only counted uniquely mapping sRNAs using the

following parameters *featureCounts* -a Mus_musculus.GRCm38.90.gtf -s 0. We excluded libraries with fewer than 10,000 mapped reads. All tissue specificity and differential expression analyses in the manuscript were carried out on uniquely mapping counts (i.e., only mapped to one genomic location), except where otherwise noted.

Unsupervised clustering and dimensionality reduction analysis

sRNA raw counts were normalized and log-transformed using *DESeq2* package. Batch effects were corrected using *limma* R package (Ritchie et al., 2015). Hierarchical clustering was performed using $\log_2$ transformed expression values and using complete linkage as distance measure between clusters. We computed Euclidian distances between samples and used these values to perform the *t*-distributed stochastic neighborhood embedding (*t*-SNE) (Laurens Van Der Maaten and Geoffrey Hinton, 2008) with the following parameters: perplexity = 20, initial dimensions of 50 and maximum iteration of 1,000. Transcripts detected in one or more samples with overall $\log_2$ expression scores <1 were excluded from this analysis.

Tissue specificity index

To compute the tissue specificity index we used the formula described previously in (Ludwig et al., 2016):

$$TSI_j = \frac{\sum_{i=1}^{N}(1 - x_{j,i})}{N - 1}$$

Where N is the total number of tissues measured and $x_{j,i}$ is the expression score of tissue i normalized by the maximal expression of any tissue for miRNA j.

Comparison with available miRNA data

To compute Spearman correlation coeffitients between samples generated in the current study and mouse miRNA data generated by FANTOM5 consortium (de Rie et al., 2017) we used *DESeq2*-normalized scores of 2207 annotated miRNAs. To compare the miRNA expression between mouse tissues and human cell types we generated a curated list of orthologous miRNAs that contained maximum two mismatches per ortholog mature miRNA. 531 miRNA passed this criteria and were used to compute Spearman correlation coefficients shown in **Figure S5**.

Differential expression analysis with DESeq2

miRNAs differentially expressed between female and male tissues were computed using *DESeq2* (Love et al., 2014). To test for the NULL hypothesis, we performed a permutation test in which we randomly re-assigned the sex labels to 14 samples across each tissue and

plotted the distribution of *DESeq2* p-values computed for the two groups (i.e. female and male) (**Figure S6**). We used Benjamini-Hochberg-corrected p-values to assess the statistical significance of the computed DE scores (**Figure 4 and S7A**). The differentially expressed miRNAs were visualized on volcano plots, where male- and female-specific miRNAs (adjusted *P*-value < 0.01 and absolute fold change > 1) were labeled accordingly.

Analysis of correlation between miRNA expression and the expression of its targets

Putative miRNA target genes were extracted from TargetScan, DIANA, miRanda, or mirDB databases (Agarwal et al., 2015; Griffiths-Jones et al., 2007; Paraskevopoulou et al., 2013; Wong and Wang, 2015). Only targetes present in two or more databases were used. The gene expression scores of the respective targets in various tissues were extracted from the ENCODE database (Pennisi, 2012) (**Table S6**). Spearman correlation coefficients were computed between FPKM retrieved from the ENCODE mRNA expression tables and *DESeq2*-normalized miRNA counts across ten profiled tissues using corr.test() function from 'psych' R package (Revelle, W.) and threshholded above Benjamini-Hochberg adjusted *P*-value of 0.1 and Spearman correlation coefficient ($-0.8 < r_s < 0.8$).

Identification of candidate novel miRNA.

Candidate novel miRNA were identified using miRDeep2 software (Friedländer et al., 2012). Only miRNAs supported by > 5 reads were reported in this study. We analyzed tissue- and gender- specificity of novel miRNAs based on transcripts supported by at least 50 sequencing reads across all samples. Statistical analysis and data visualization were performed as described above for annotated miRNAs.

Machine learning

We trained the radial kernel SVM and the Random Forest models on 136 samples corresponding to different tissue types (**Figure S9A**) using *e1070* (Meyer et al., 2017) and *caret (Kuhn, 2018)* R packages respectively. We used z-scores of *DESeq2* normalized counts obtained in this study as a train dataset and those obtained from ENCODE miRNA-seq data as test dataset (**Table S4**). We normalized and scaled train and test datasets separately.

To measure the predictive power of each model we used the accuracy measure, calculated as the following:

$$\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total observations}}$$

We tuned the SVM model to derive optimal cost and gamma using tune.svm() function and searching within gamma $\in$ [2^(-10): 2^10] and cost $\in$ [10^(-5):10^3]. We tuned RF model using first random and then grid search, with an evaluation metric set to "Accuracy". The

accuracy was computed using 10-fold cross-validation procedure (Japkowicz and Shah, 2011). The reported accuracy is computed as a mean over the 10 testing sets in which 9-folds are used for training and the held-out fold used as a test set. The R script used to train the models and compute the predictions is included in the supplement.

Data availability

The datasets generated and analyzed in the study are available in the NCBI Gene Expression Omnibus (GEO) under the entry GEO:GSE119661.

Supplementary figures:

Figure S1. sRNA read statistics. (A) Average number of reads mapping to sRNAs per tissue. (B) Genomic localization of identified sRNAs with respect to protein-coding genes. (C) Number of unique transcripts per tissue across four sRNAs classes.

Figure S2. Unsupervised clustering of RNA-seq datasets. Heatmap of normalized log-transformed RNA-seq counts detected in various murine tissues across four sRNAs classes.

Figure S3. Tissue-specific patterns of sRNA expression. (A) t-SNE projection of expression values for each sRNAs class. (B) Tissue specificity index of sRNA. (C) Tissue-specific as well as unique sRNAs transcripts detected in each tissue, separated by scnRNA class. Size of the circle corresponds to the average number of normalized counts across each tissue.

Figure S4. Top 50 highest miRNA expression scores across eleven tissues. Heatmaps show highest expression scores for each tissues alongside the scores of the same miRNA in all other tissues.

Figure S5. Correlation of our data with previously published miRNA datasets generated for mammalian cell- and tissues- types. (A) Spearman correlation coefficients of comparison between expression scores obtained in this study and mouse miRNA-seq data generated by FANTOM5 (de Rie et al., 2017). (B) Number of miRNA orthologs used to compare FANTOM5 human and current mouse datasets. (C) same as A but for human FANTOM5 samples.

Figure S6. Differential expression analysis. NULL hypothesis testing. Distribution of *DESeq2*-derived p-values computed for miRNAs differentially expressed between actual and permuted female and male samples.

Figure S7. Gender-specific expression of miRBase-annotated miRNA. (A) Volcano plots showing annotated miRNAs differentially expressed between female and male intestine, spleen, liver and pancreas. (B) Sexually dimorphic miRNAs. miRNAs found to be consistently upregulated in female tissues shown in dark red and those in male tissues – in dark blue. miR-182, miR-148a and miR-145a are upregulated by estrogen ("E"), while miR-340 is down-regulated by androgen. (C) Spearman correlation of miRNA expression and the expression of respective targets. A list of target mRNAs was obtained from TargetScan, DIANA, miRanda, or mirDB databases (Agarwal et al., 2015; Griffiths-Jones et al., 2007; Paraskevopoulou et al., 2013; Wong and Wang, 2015). Only targets listed in two or more databases were used in this analysis. mRNA expression values for the respective tissues were retrived from the ENCODE database (Pennisi, 2012). Only r_s scores of correlations that passed FDR<0.1 threshhold are shown. Positively correlated miRNA:mRNA scores are depicted in red, while negatively correlated – in blue.

Figure S8. Tissue- and gender-specific expression of novel miRNA. (A) Principal component analysis performed on novel miRNAs detected in one or multiple mouse tissues. (B) Examples of an ubiquetous and a tissue-specific novel miRNA. (C) Volcano plots showing novel miRNA differentially expressed miRNA between female and male brain, spleen, heart, intestine, kidney and liver.

Figure S9. Machine learning statistics. (A) RNA-seq samples used to train machine learning models. Left: number of train datasets corresponding to each tissue, right: examples of feature scores. (B) miRNAs assigned highest weights in defining bone marrow in SVM model. (C) top 20 miRNAs atributed highest weights in defining tissue types by RF model.

Supplementary tables:

Table S1. Animals used in the study.

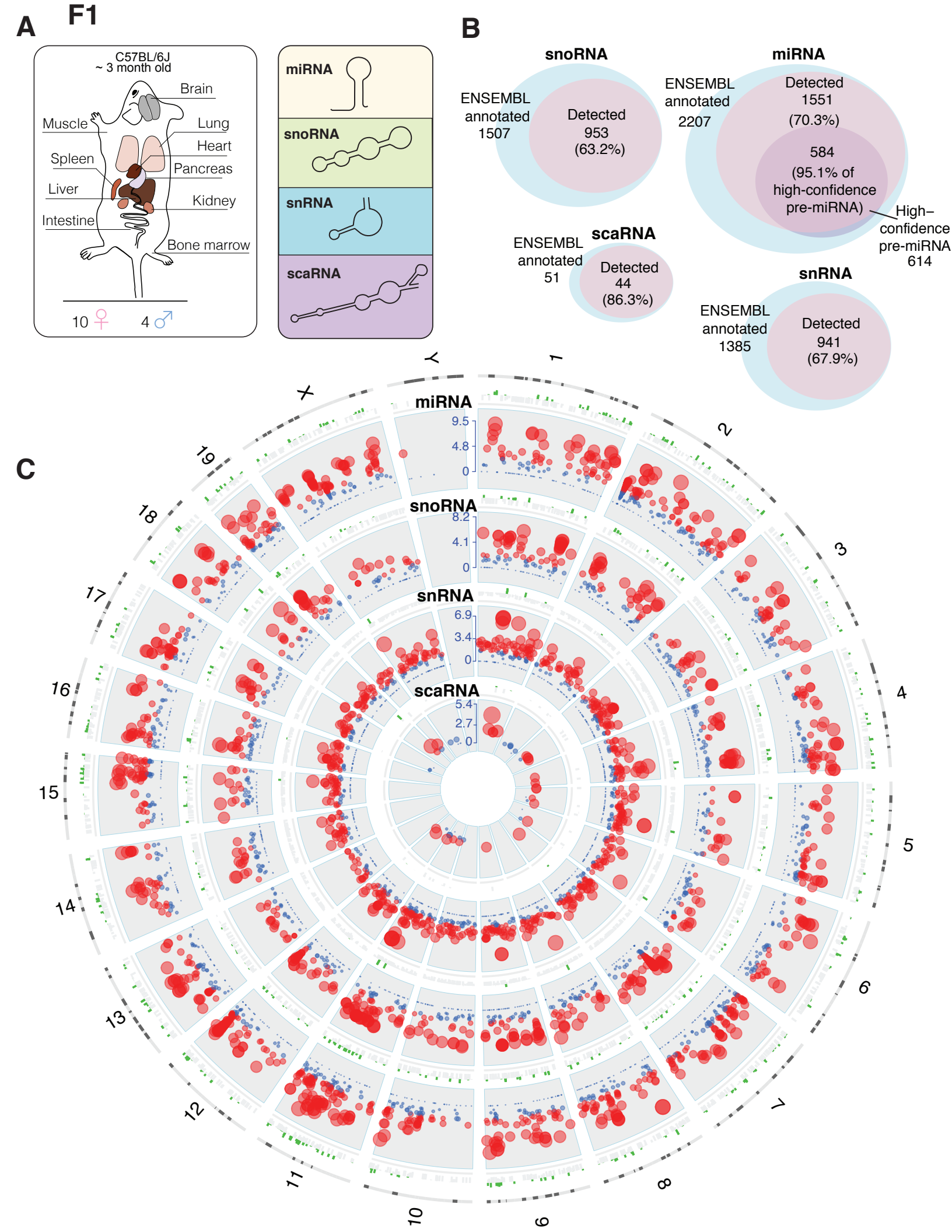
Table S2. Mean normalized expression counts.

Table S3. Novel miRNAs, sequence and expression counts.

Table S4. ENCODE data used for ML predictions.

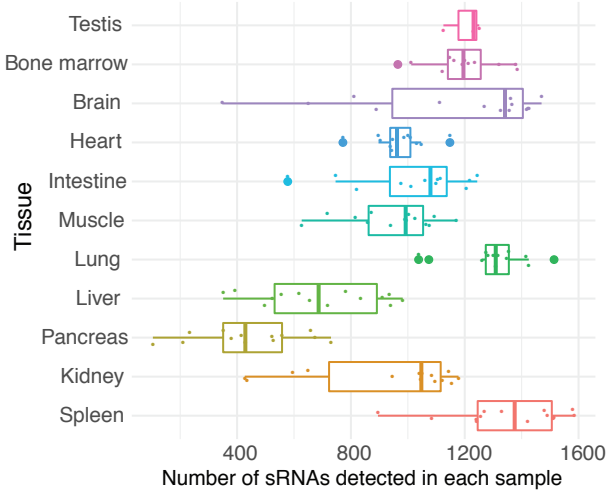
Table S5. Targets of sexually dimorphic miRNAs.

Table S6. ENCODE mRNA datasets used in the study.

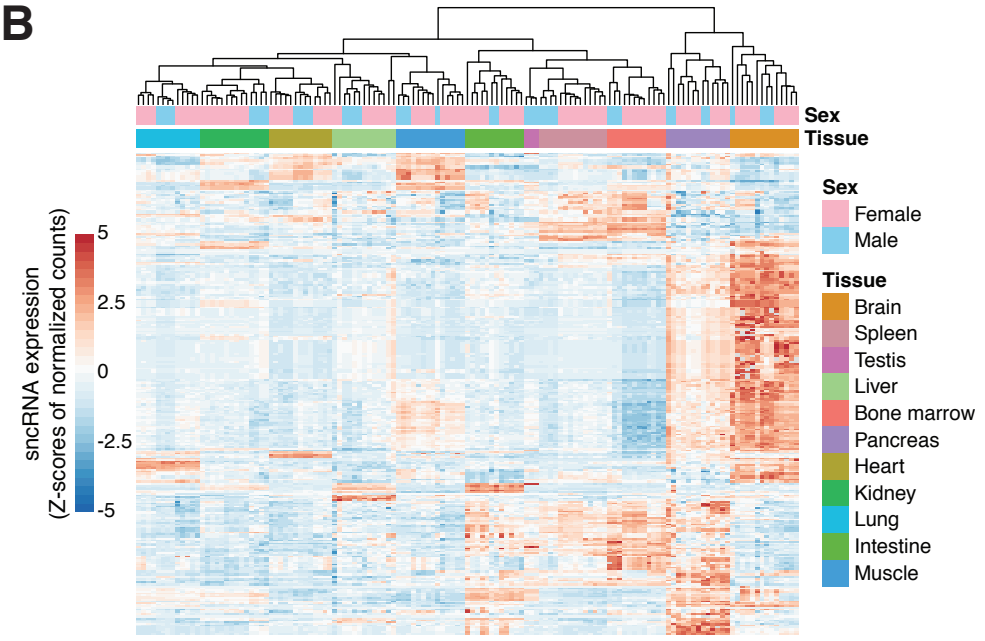

A
F1
C57BL/6J
~ 3 month old
Brain
Muscle
Lung
Heart
Spleen
Pancreas
Liver
Kidney
Intestine
Bone marrow
10 ♀
4 ♂
miRNA
snoRNA
snRNA
scaRNA
B
snoRNA
ENSEMBL annotated 1507
Detected 953 (63.2%)
miRNA
ENSEMBL annotated 2207
Detected 1551 (70.3%)
584 (95.1% of high-confidence pre-miRNA)
High–confidence pre-miRNA 614
scaRNA
ENSEMBL annotated 51
Detected 44 (86.3%)
snRNA
ENSEMBL annotated 1385
Detected 941 (67.9%)
C
miRNA
9.5
4.8
0
snoRNA
8.2
4.1
0
snRNA
6.9
3.4
0
scaRNA
5.4
2.7
0
X
Y
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

F2

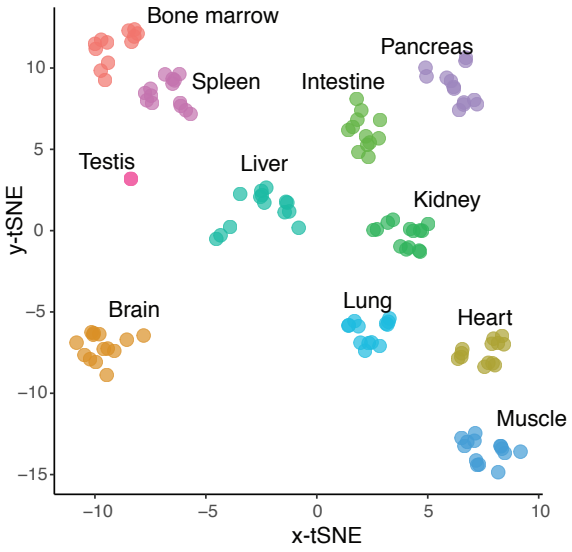
A


Testis
Bone marrow
Brain
Heart
Intestine
Muscle
Lung
Liver
Pancreas
Kidney
Spleen
Tissue
400
800
1200
1600
Number of sRNAs detected in each sample

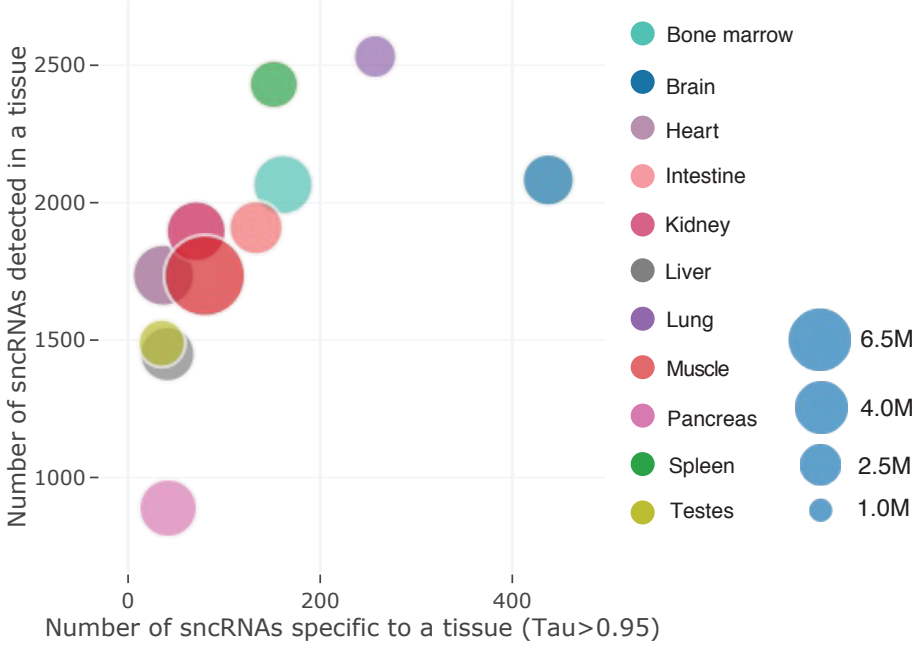
B

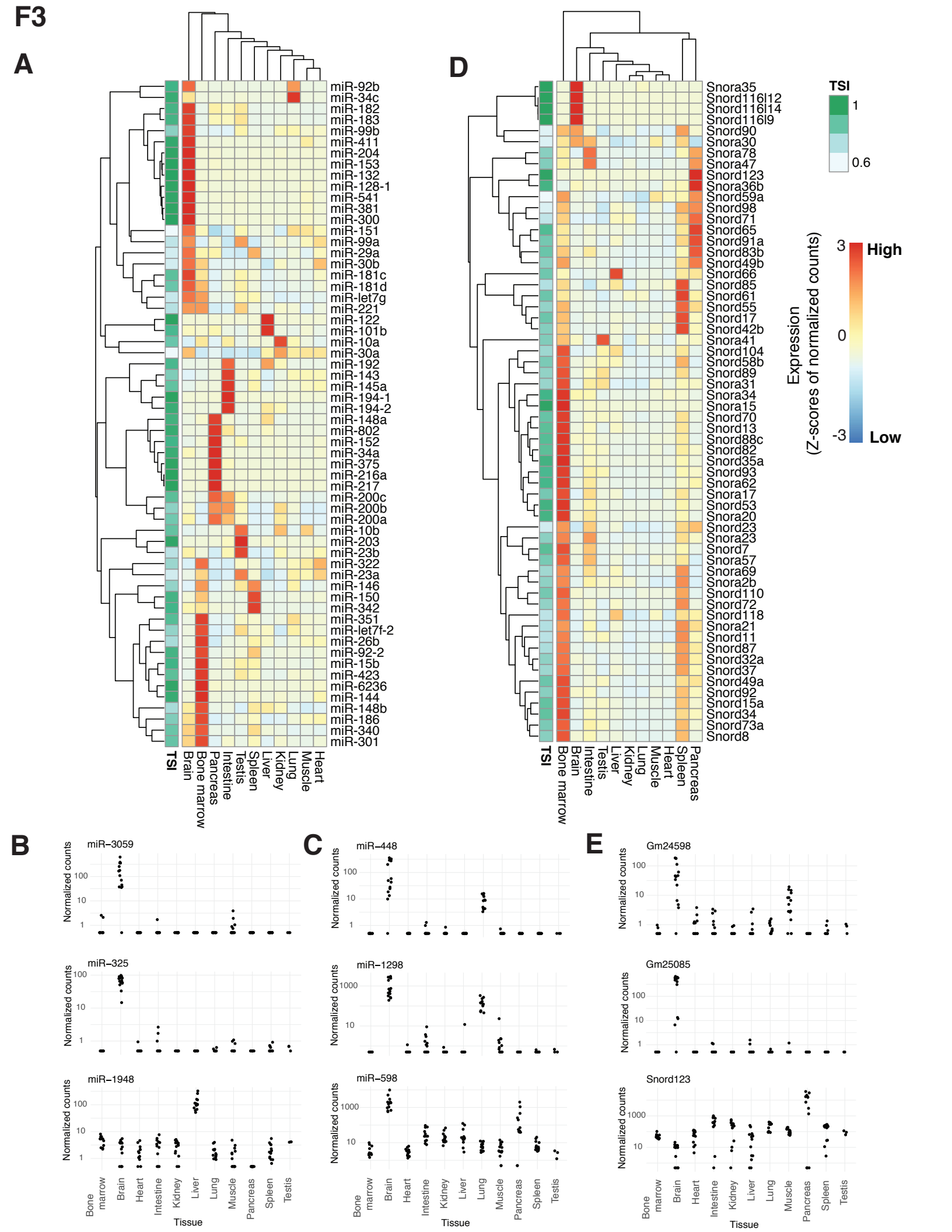

Sex
Tissue
sncRNA expression
(Z-scores of normalized counts)
5
2.5
0
-2.5
-5
Sex
Female
Male
Tissue
Brain
Spleen
Testis
Liver
Bone marrow
Pancreas
Heart
Kidney
Lung
Intestine
Muscle

C

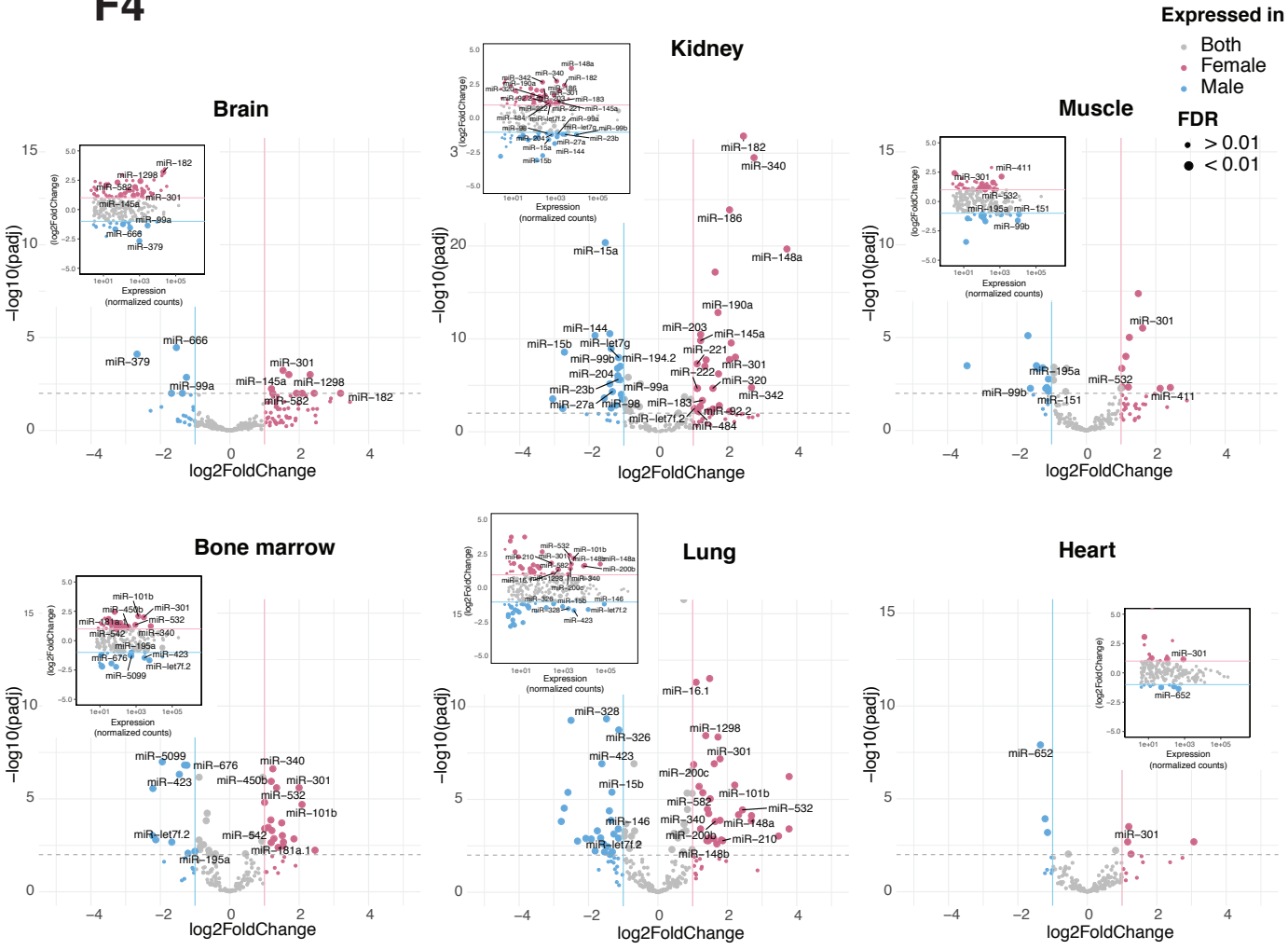

Bone marrow
Spleen
Pancreas
Intestine
Testis
Liver
Kidney
Brain
Lung
Heart
Muscle
y-tSNE
x-tSNE
10
5
0
−5
−10
−15
−10
−5
0
5
10

D


Number of sncRNAs detected in a tissue
2500
2000
1500
1000
0
200
400
Number of sncRNAs specific to a tissue (Tau>0.95)
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testes
6.5M
4.0M
2.5M
1.0M

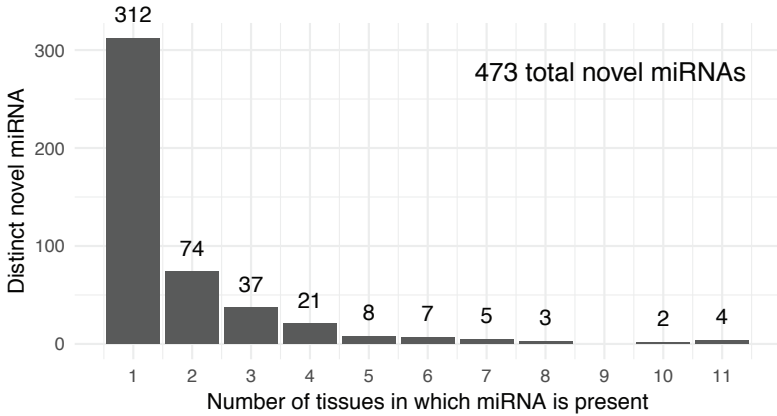

F3
A
miR-92b
miR-34c
miR-182
miR-183
miR-99b
miR-411
miR-204
miR-153
miR-132
miR-128-1
miR-541
miR-381
miR-300
miR-151
miR-99a
miR-29a
miR-30b
miR-181c
miR-181d
miR-let7g
miR-221
miR-122
miR-101b
miR-10a
miR-30a
miR-192
miR-143
miR-145a
miR-194-1
miR-194-2
miR-148a
miR-802
miR-152
miR-34a
miR-375
miR-216a
miR-217
miR-200c
miR-200b
miR-200a
miR-10b
miR-203
miR-23b
miR-322
miR-23a
miR-146
miR-150
miR-342
miR-351
miR-let7f-2
miR-26b
miR-92-2
miR-15b
miR-423
miR-6236
miR-144
miR-148b
miR-186
miR-340
miR-301
TSI
Brain
Bone marrow
Pancreas
Intestine
Testis
Spleen
Liver
Kidney
Lung
Muscle
Heart
D
Snora35
Snord116l12
Snord116l14
Snord116l9
Snord90
Snora30
Snora78
Snora47
Snord123
Snora36b
Snord59a
Snord98
Snord71
Snord65
Snord91a
Snord83b
Snord49b
Snord66
Snord85
Snord61
Snord55
Snord17
Snord42b
Snora41
Snord104
Snord58b
Snord89
Snora31
Snora34
Snora15
Snord70
Snord13
Snord88c
Snord82
Snord35a
Snord93
Snora62
Snora17
Snord53
Snora20
Snord23
Snora23
Snord7
Snora57
Snora69
Snora2b
Snord110
Snord72
Snord118
Snora21
Snord11
Snord87
Snord32a
Snord37
Snord49a
Snord92
Snord15a
Snord34
Snord73a
Snord8
TSI
1
0.6
Expression
(Z-scores of normalized counts)
3
High
0
-3
Low
B
miR−3059
miR−325
miR−1948
C
miR−448
miR−1298
miR−598
E
Gm24598
Gm25085
Snord123
Normalized counts
Tissue
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testis

F4

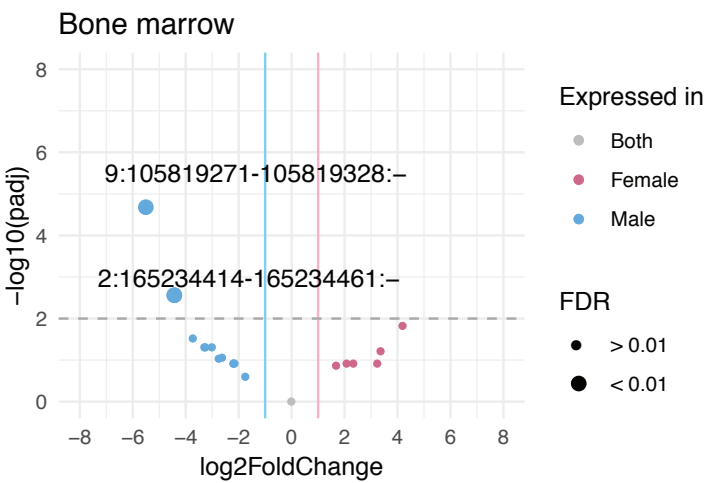

Expressed in
Both
Female
Male
FDR
> 0.01
< 0.01
Brain
Kidney
Muscle
Bone marrow
Lung
Heart
−log10(padj)
log2FoldChange
Expression
(normalized counts)
(log2FoldChange)
miR−182
miR−1298
miR−582
miR−145a
miR−301
miR−99a
miR−666
miR−379
miR−340
miR−186
miR−15a
miR−148a
miR−190a
miR−203
miR−144
miR−15b
miR−let7g
miR−99b
miR−194.2
miR−204
miR−23b
miR−27a
miR−98
miR−183
miR−221
miR−222
miR−320
miR−342
miR−92.2
miR−484
miR−411
miR−532
miR−195a
miR−151
miR−652
miR−101b
miR−450b
miR−542
miR−181a.1
miR−676
miR−423
miR−let7f.2
miR−5099
miR−16.1
miR−328
miR−326
miR−200c
miR−146
miR−200b
miR−210
miR−148b

F5

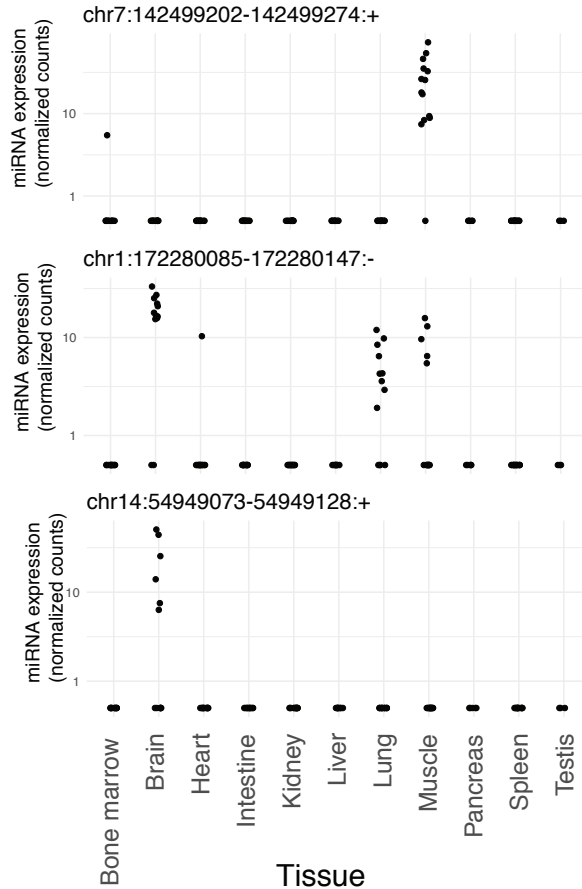
A


473 total novel miRNAs
Distinct novel miRNA
Number of tissues in which miRNA is present
312
74
37
21
8
7
5
3
2
4
0
100
200
300
1
2
3
4
5
6
7
8
9
10
11

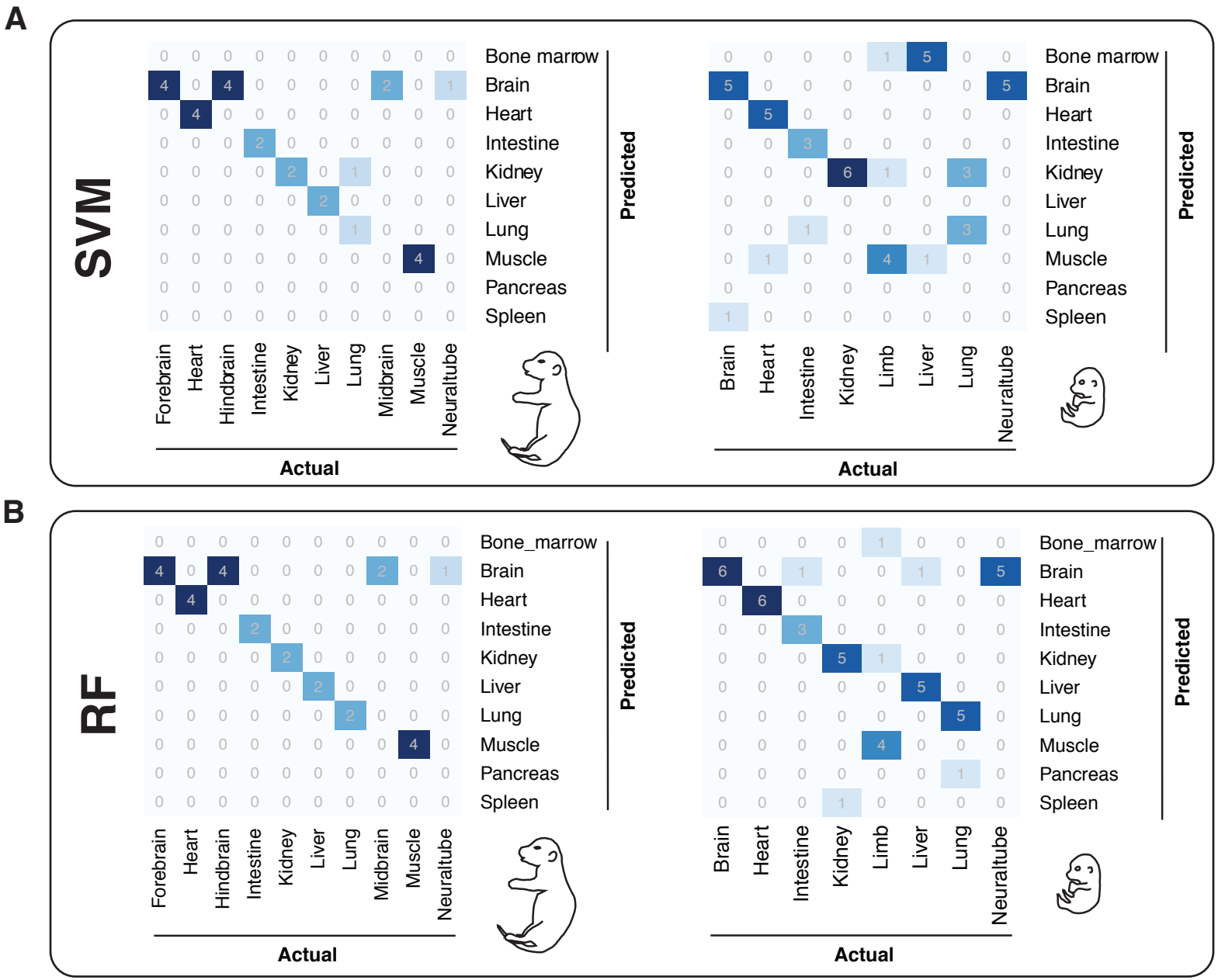
B


Bone marrow
9:105819271-105819328:–
2:165234414-165234461:–
−log10(padj)
log2FoldChange
Expressed in
Both
Female
Male
FDR
> 0.01
< 0.01

C


chr7:142499202-142499274:+
chr1:172280085-172280147:-
chr14:54949073-54949128:+
miRNA expression (normalized counts)
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testis
Tissue

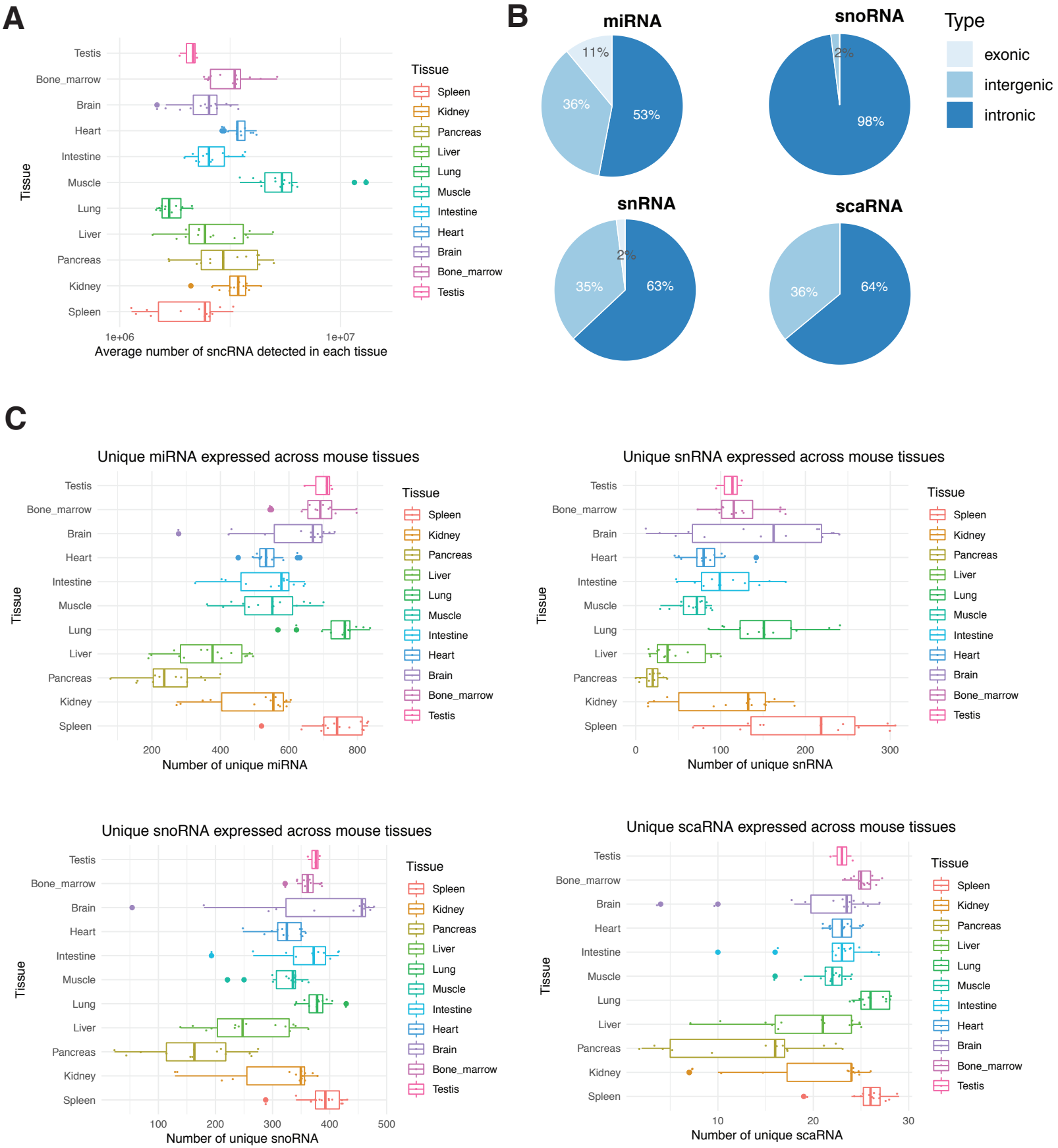
F6


A
SVM
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Forebrain
Heart
Hindbrain
Intestine
Kidney
Liver
Lung
Midbrain
Muscle
Neuraltube
Actual
Predicted
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Brain
Heart
Intestine
Kidney
Limb
Liver
Lung
Neuraltube
Actual
Predicted
B
RF
Bone_marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Forebrain
Heart
Hindbrain
Intestine
Kidney
Liver
Lung
Midbrain
Muscle
Neuraltube
Actual
Predicted
Bone_marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Brain
Heart
Intestine
Kidney
Limb
Liver
Lung
Neuraltube
Actual
Predicted

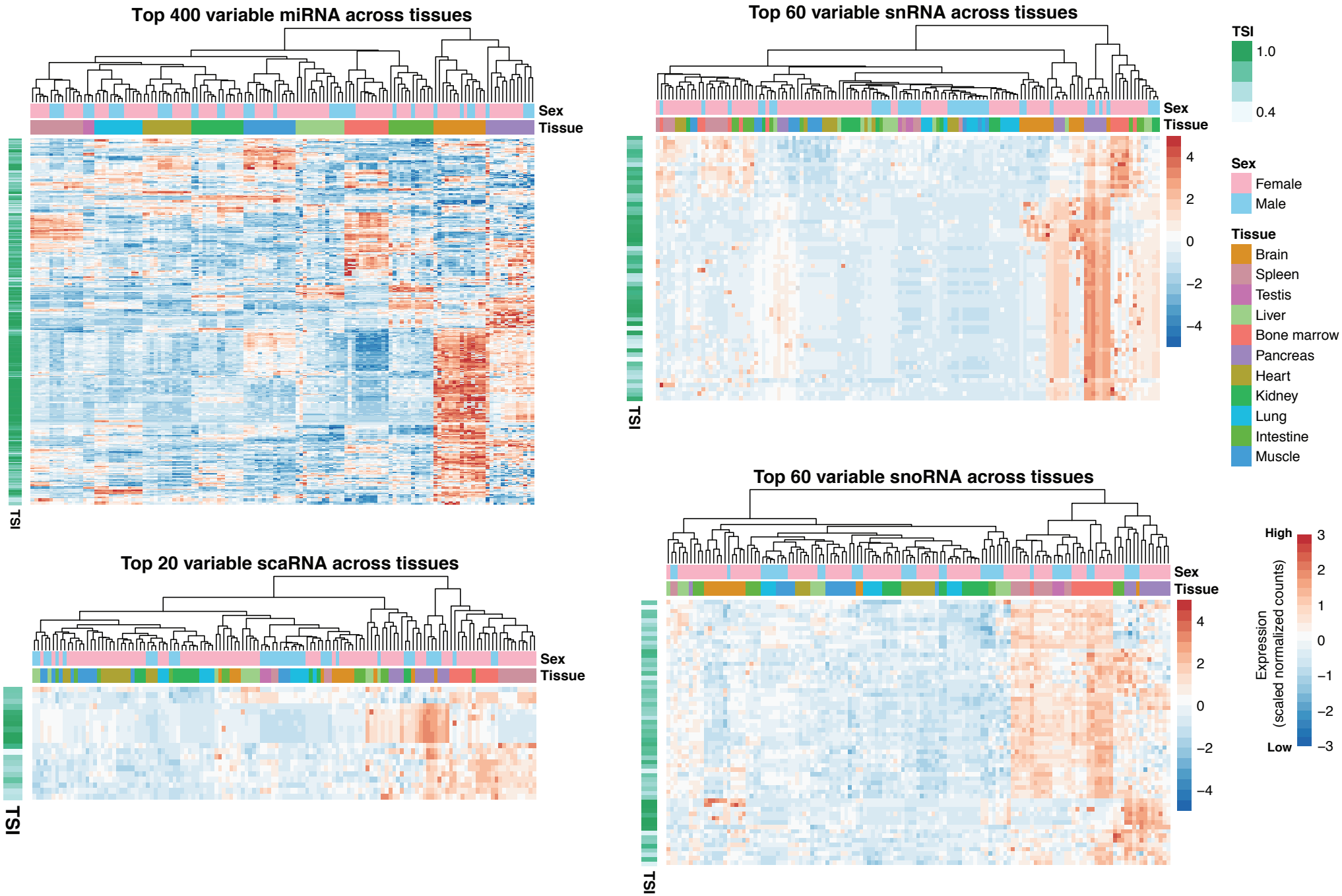
Supplementary figures

S1

Click here to access/download;Supplemental Figure;Isakova.Suppl.figures.pdf

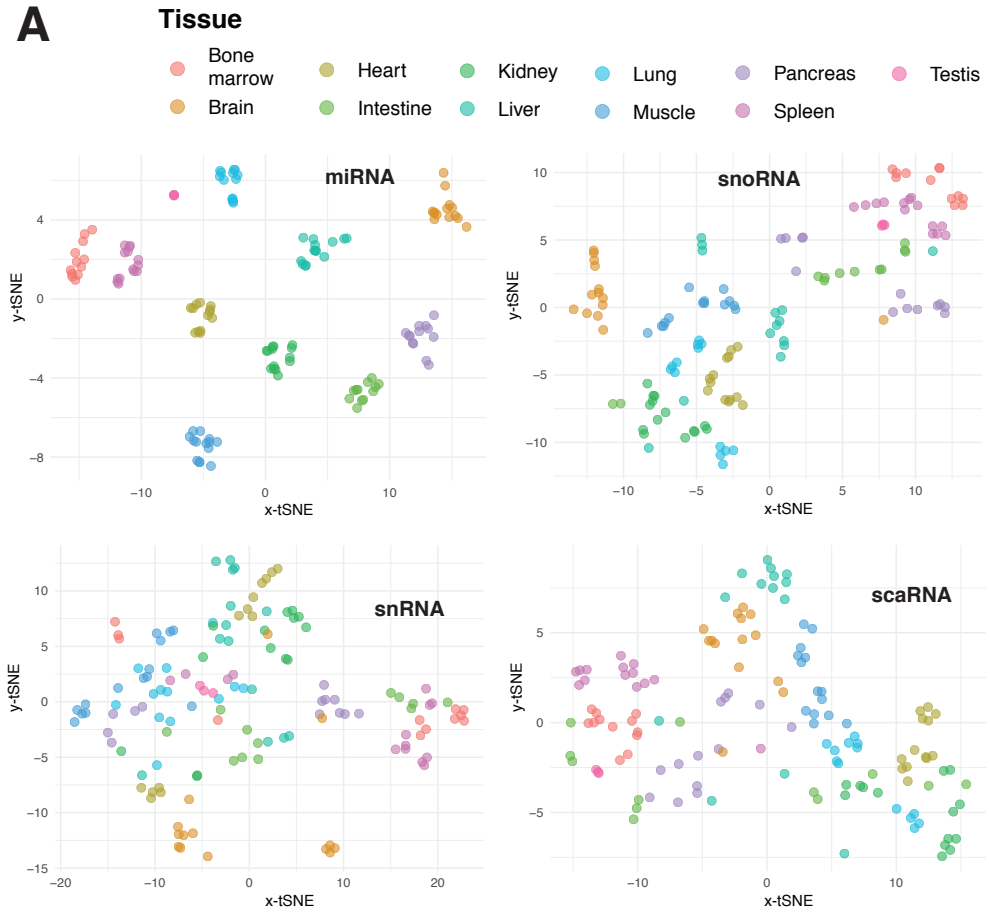


S2

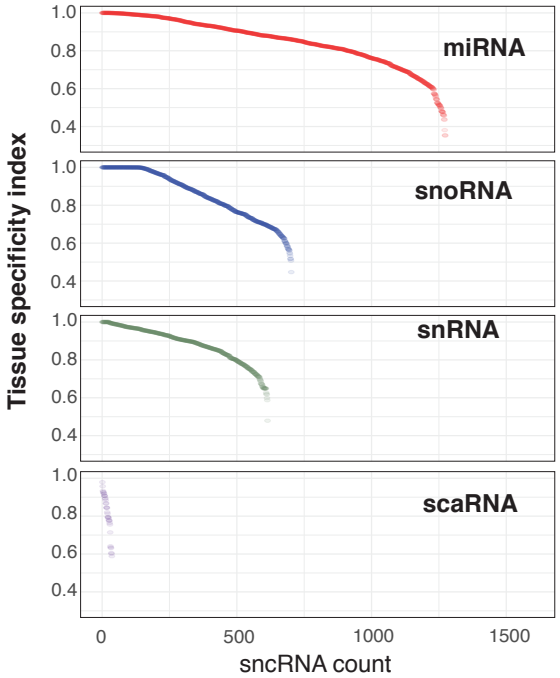

Top 400 variable miRNA across tissues
Sex
Tissue
TSI
Top 60 variable snRNA across tissues
Sex
Tissue
TSI
4
2
0
−2
−4
TSI
1.0
0.4
Sex
Female
Male
Tissue
Brain
Spleen
Testis
Liver
Bone marrow
Pancreas
Heart
Kidney
Lung
Intestine
Muscle
Top 20 variable scaRNA across tissues
Sex
Tissue
TSI
Top 60 variable snoRNA across tissues
Sex
Tissue
TSI
4
2
0
−2
−4
High
Low
Expression
(scaled normalized counts)
3
2
1
0
−1
−2
−3

S3

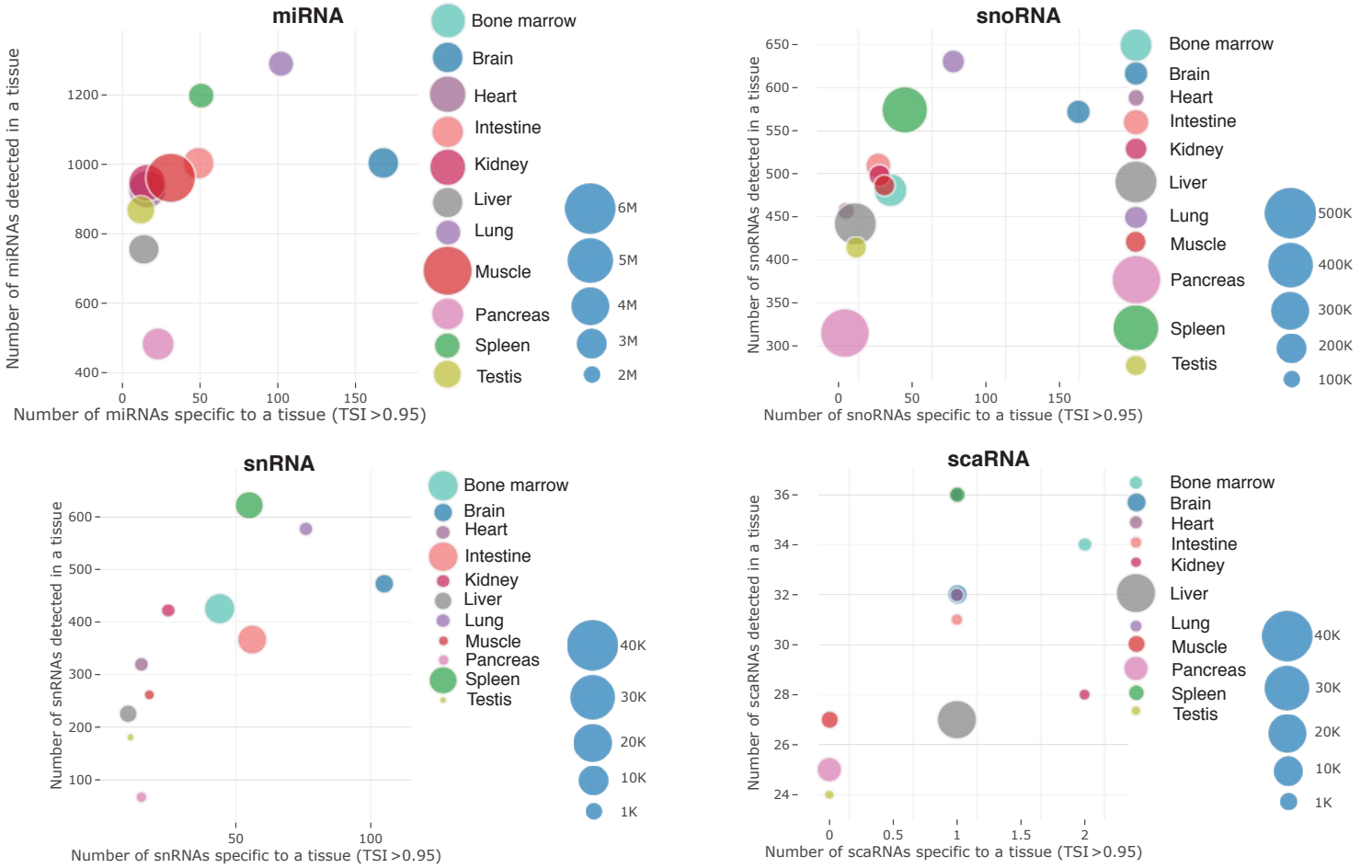
A


Tissue
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testis
miRNA
snoRNA
snRNA
scaRNA
x-tSNE
y-tSNE

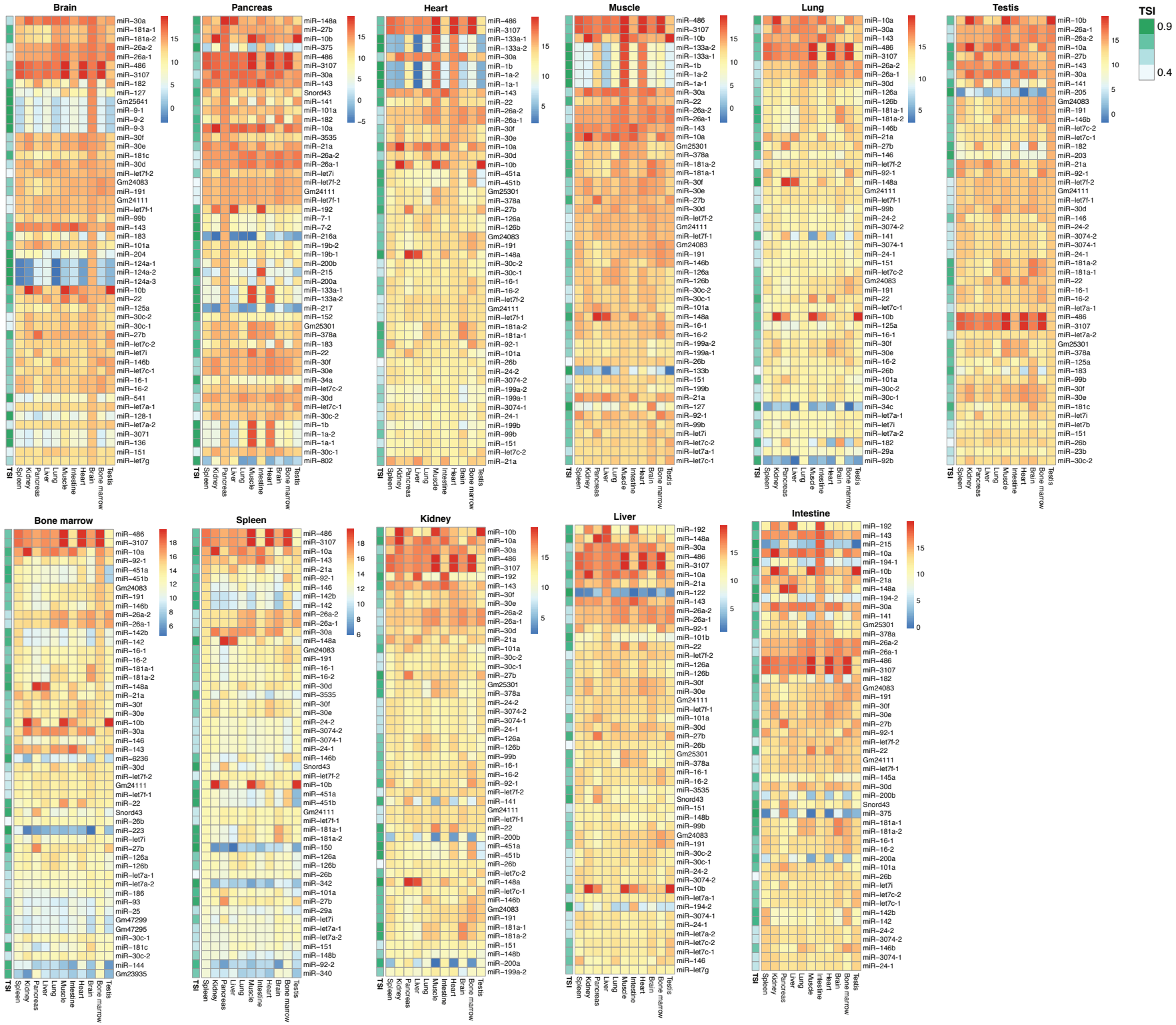
B


miRNA
snoRNA
snRNA
scaRNA
Tissue specificity index
sncRNA count

C

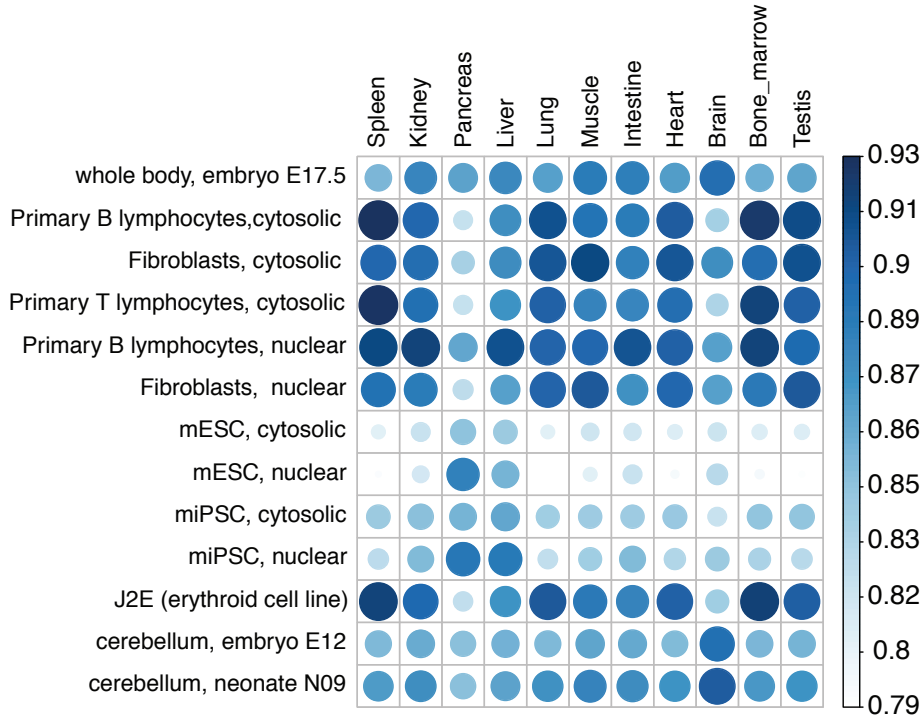

miRNA
Number of miRNAs detected in a tissue
Number of miRNAs specific to a tissue (TSI >0.95)
snoRNA
Number of snoRNAs detected in a tissue
Number of snoRNAs specific to a tissue (TSI >0.95)
snRNA
Number of snRNAs detected in a tissue
Number of snRNAs specific to a tissue (TSI >0.95)
scaRNA
Number of scaRNAs detected in a tissue
Number of scaRNAs specific to a tissue (TSI >0.95)
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testis

S4

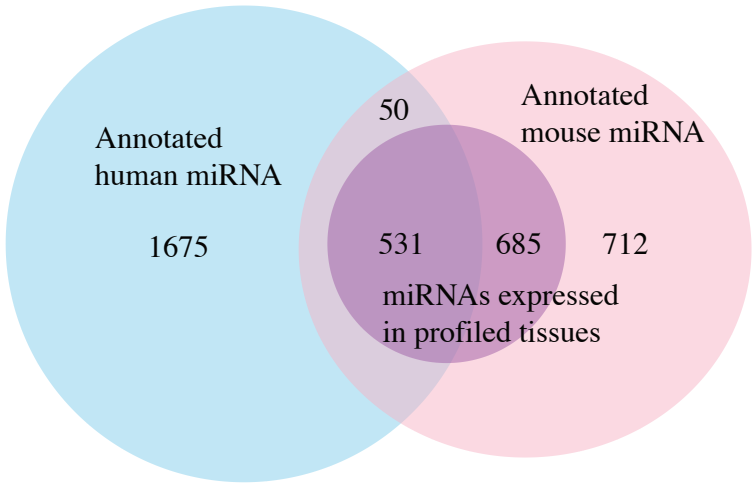

Brain
Pancreas
Heart
Muscle
Lung
Testis
Bone marrow
Spleen
Kidney
Liver
Intestine
TSI
0.9
0.4
TSI Spleen Kidney Pancreas Liver Lung Muscle Intestine Heart Brain Bone marrow Testis

S5

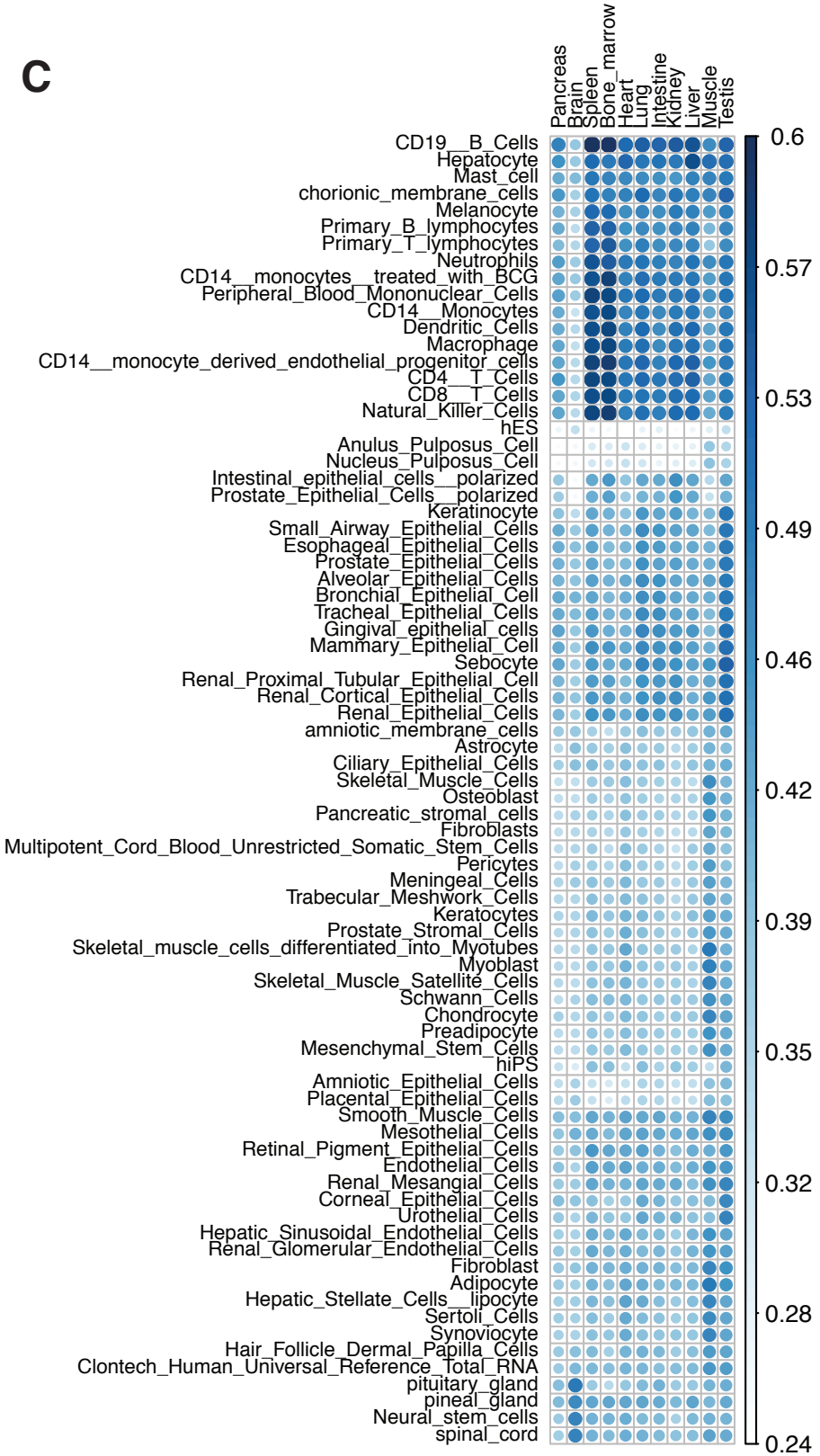
A


Spleen
Kidney
Pancreas
Liver
Lung
Muscle
Intestine
Heart
Brain
Bone_marrow
Testis
whole body, embryo E17.5
Primary B lymphocytes,cytosolic
Fibroblasts, cytosolic
Primary T lymphocytes, cytosolic
Primary B lymphocytes, nuclear
Fibroblasts, nuclear
mESC, cytosolic
mESC, nuclear
miPSC, cytosolic
miPSC, nuclear
J2E (erythroid cell line)
cerebellum, embryo E12
cerebellum, neonate N09
0.93
0.91
0.9
0.89
0.87
0.86
0.85
0.83
0.82
0.8
0.79

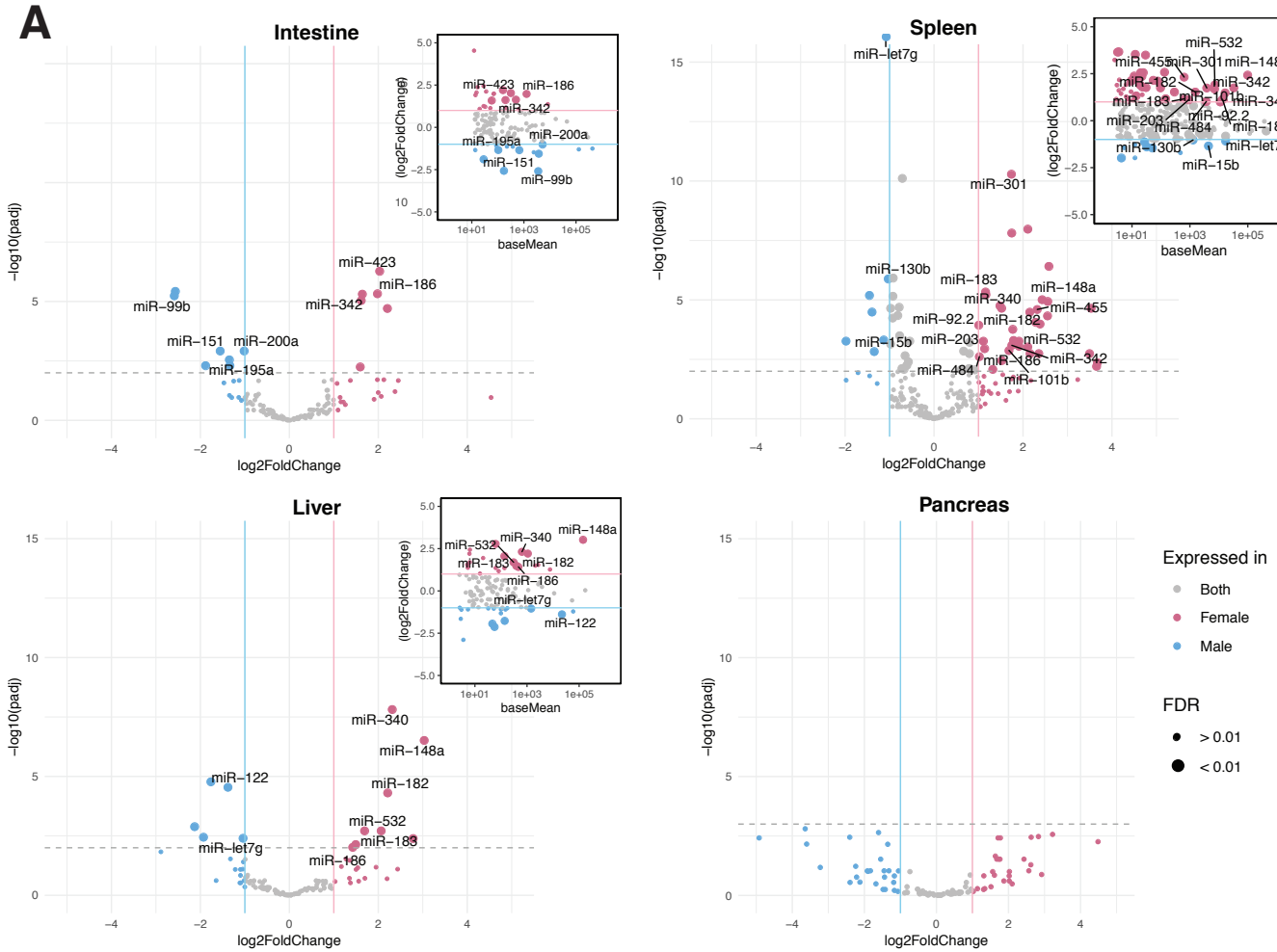
B

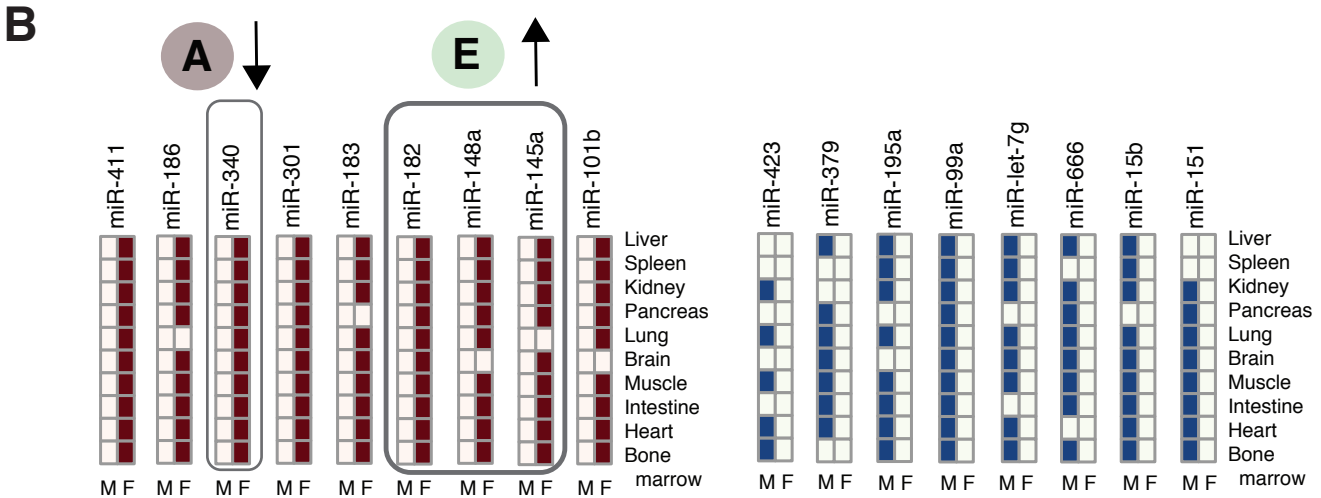

Annotated
human miRNA
1675
50
Annotated
mouse miRNA
531
685
712
miRNAs expressed
in profiled tissues

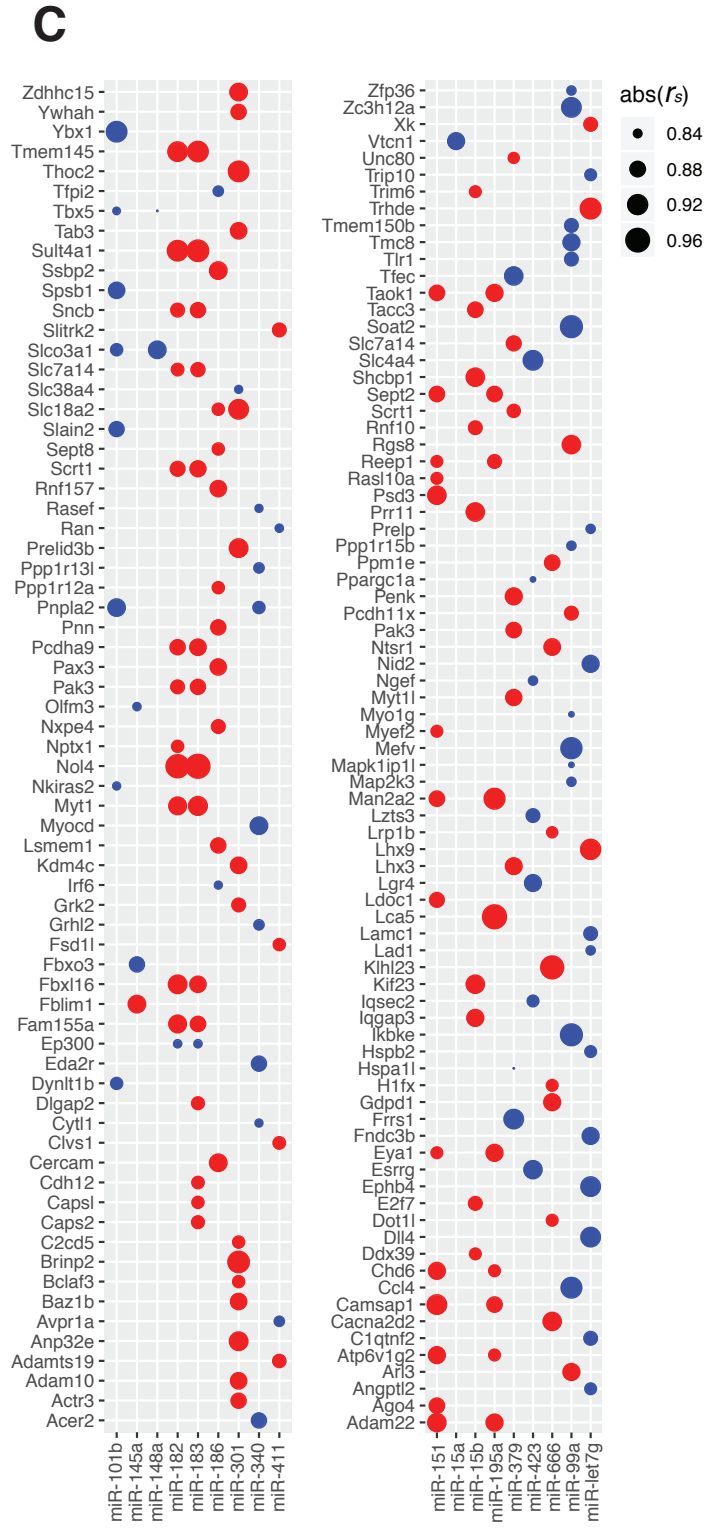
C


Pancreas
Brain
Spleen
Bone_marrow
Heart
Lung
Intestine
Kidney
Liver
Muscle
Testis
CD19__B_Cells
Hepatocyte
Mast_cell
chorionic_membrane_cells
Melanocyte
Primary_B_lymphocytes
Primary_T_lymphocytes
Neutrophils
CD14__monocytes__treated_with_BCG
Peripheral_Blood_Mononuclear_Cells
CD14__Monocytes
Dendritic_Cells
Macrophage
CD14__monocyte_derived_endothelial_progenitor_cells
CD4__T_Cells
CD8__T_Cells
Natural_Killer_Cells
hES
Anulus_Pulposus_Cell
Nucleus_Pulposus_Cell
Intestinal_epithelial_cells__polarized
Prostate_Epithelial_Cells__polarized
Keratinocyte
Small_Airway_Epithelial_Cells
Esophageal_Epithelial_Cells
Prostate_Epithelial_Cells
Alveolar_Epithelial_Cells
Bronchial_Epithelial_Cell
Tracheal_Epithelial_Cells
Gingival_epithelial_cells
Mammary_Epithelial_Cell
Sebocyte
Renal_Proximal_Tubular_Epithelial_Cell
Renal_Cortical_Epithelial_Cells
Renal_Epithelial_Cells
amniotic_membrane_cells
Astrocyte
Ciliary_Epithelial_Cells
Skeletal_Muscle_Cells
Osteoblast
Pancreatic_stromal_cells
Fibroblasts
Multipotent_Cord_Blood_Unrestricted_Somatic_Stem_Cells
Pericytes
Meningeal_Cells
Trabecular_Meshwork_Cells
Keratocytes
Prostate_Stromal_Cells
Skeletal_muscle_cells_differentiated_into_Myotubes
Myoblast
Skeletal_Muscle_Satellite_Cells
Schwann_Cells
Chondrocyte
Preadipocyte
Mesenchymal_Stem_Cells
hiPS
Amniotic_Epithelial_Cells
Placental_Epithelial_Cells
Smooth_Muscle_Cells
Mesothelial_Cells
Retinal_Pigment_Epithelial_Cells
Endothelial_Cells
Renal_Mesangial_Cells
Corneal_Epithelial_Cells
Urothelial_Cells
Hepatic_Sinusoidal_Endothelial_Cells
Renal_Glomerular_Endothelial_Cells
Fibroblast
Adipocyte
Hepatic_Stellate_Cells__lipocyte
Sertoli_Cells
Synoviocyte
Hair_Follicle_Dermal_Papilla_Cells
Clontech_Human_Universal_Reference_Total_RNA
pituitary_gland
pineal_gland
Neural_stem_cells
spinal_cord
0.6
0.57
0.53
0.49
0.46
0.42
0.39
0.35
0.32
0.28
0.24

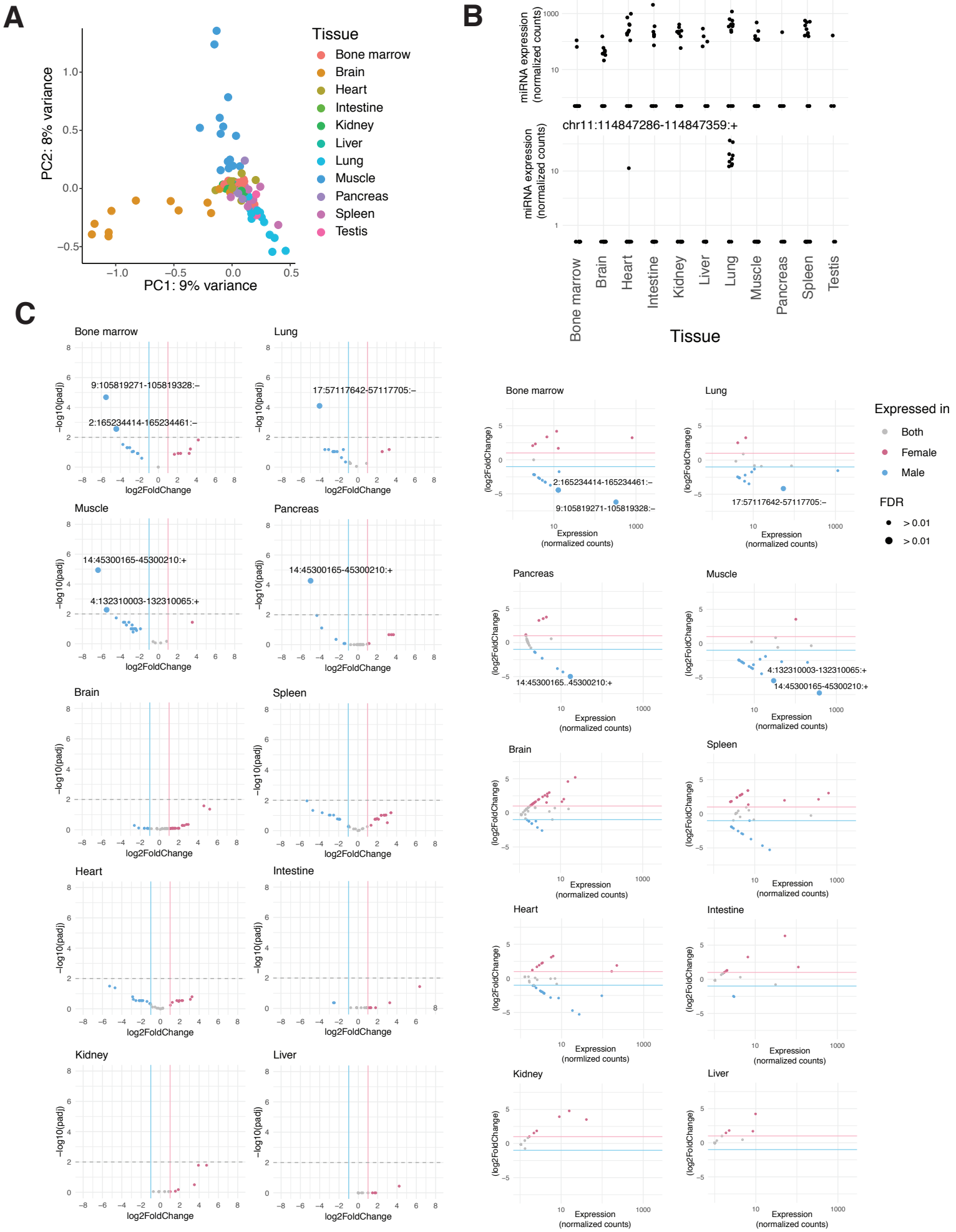
S6


A
Intestine
Spleen
Liver
Pancreas
−log10(padj)
log2FoldChange
(log2FoldChange)
baseMean
miR−423
miR−186
miR−342
miR−99b
miR−151
miR−200a
miR−195a
miR−let7g
miR−301
miR−130b
miR−183
miR−148a
miR−340
miR−455
miR−92.2
miR−182
miR−15b
miR−203
miR−532
miR−484
miR−101b
miR−122
Expressed in
Both
Female
Male
FDR
> 0.01
< 0.01


B
A
E
miR-411
miR-186
miR-340
miR-301
miR-183
miR-182
miR-148a
miR-145a
miR-101b
miR-423
miR-379
miR-195a
miR-99a
miR-let-7g
miR-666
miR-15b
miR-151
Liver
Spleen
Kidney
Pancreas
Lung
Brain
Muscle
Intestine
Heart
Bone marrow
M F


C
abs(r_s)
0.84
0.88
0.92
0.96
Zdhhc15
Ywhah
Ybx1
Tmem145
Thoc2
Tfpi2
Tbx5
Tab3
Sult4a1
Ssbp2
Spsb1
Sncb
Slitrk2
Slco3a1
Slc7a14
Slc38a4
Slc18a2
Slain2
Sept8
Scrt1
Rnf157
Rasef
Ran
Prelid3b
Ppp1r13l
Ppp1r12a
Pnpla2
Pnn
Pcdha9
Pax3
Pak3
Olfm3
Nxpe4
Nptx1
Nol4
Nkiras2
Myt1
Myocd
Lsmem1
Kdm4c
Irf6
Grk2
Grhl2
Fsd1l
Fbxo3
Fbxl16
Fblim1
Fam155a
Ep300
Eda2r
Dynlt1b
Dlgap2
Cytl1
Clvs1
Cercam
Cdh12
Capsl
Caps2
C2cd5
Brinp2
Bclaf3
Baz1b
Avpr1a
Anp32e
Adamts19
Adam10
Actr3
Acer2
miR-101b
miR-145a
miR-148a
miR-182
miR-183
miR-186
miR-301
miR-340
miR-411
Zfp36
Zc3h12a
Xk
Vtcn1
Unc80
Trip10
Trim6
Trhde
Tmem150b
Tmc8
Tlr1
Tfec
Taok1
Tacc3
Soat2
Slc7a14
Slc4a4
Shcbp1
Sept2
Scrt1
Rnf10
Rgs8
Reep1
Rasl10a
Psd3
Prr11
Prelp
Ppp1r15b
Ppm1e
Ppargc1a
Penk
Pcdh11x
Pak3
Ntsr1
Nid2
Ngef
Myt1l
Myo1g
Myef2
Mefv
Mapk1ip1l
Map2k3
Man2a2
Lzts3
Lrp1b
Lhx9
Lhx3
Lgr4
Ldoc1
Lca5
Lamc1
Lad1
Klhl23
Kif23
Iqsec2
Iqgap3
Ikbke
Hspb2
Hspa1l
H1fx
Gdpd1
Frrs1
Fndc3b
Eya1
Esrrg
Ephb4
E2f7
Dot1l
Dll4
Ddx39
Chd6
Ccl4
Camsap1
Cacna2d2
C1qtnf2
Atp6v1g2
Arl3
Angptl2
Ago4
Adam22
miR-151
miR-15a
miR-15b
miR-195a
miR-379
miR-423
miR-666
miR-99a
miR-let7g

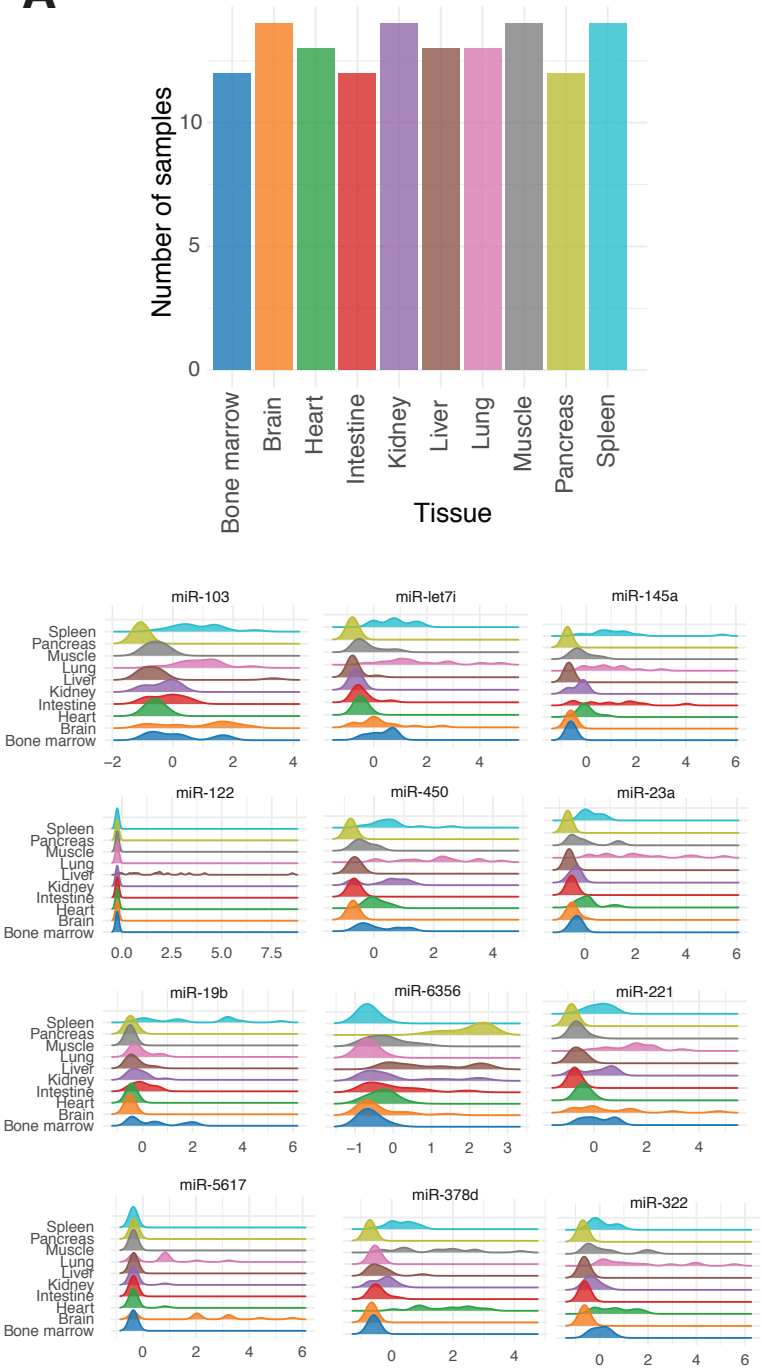
S7



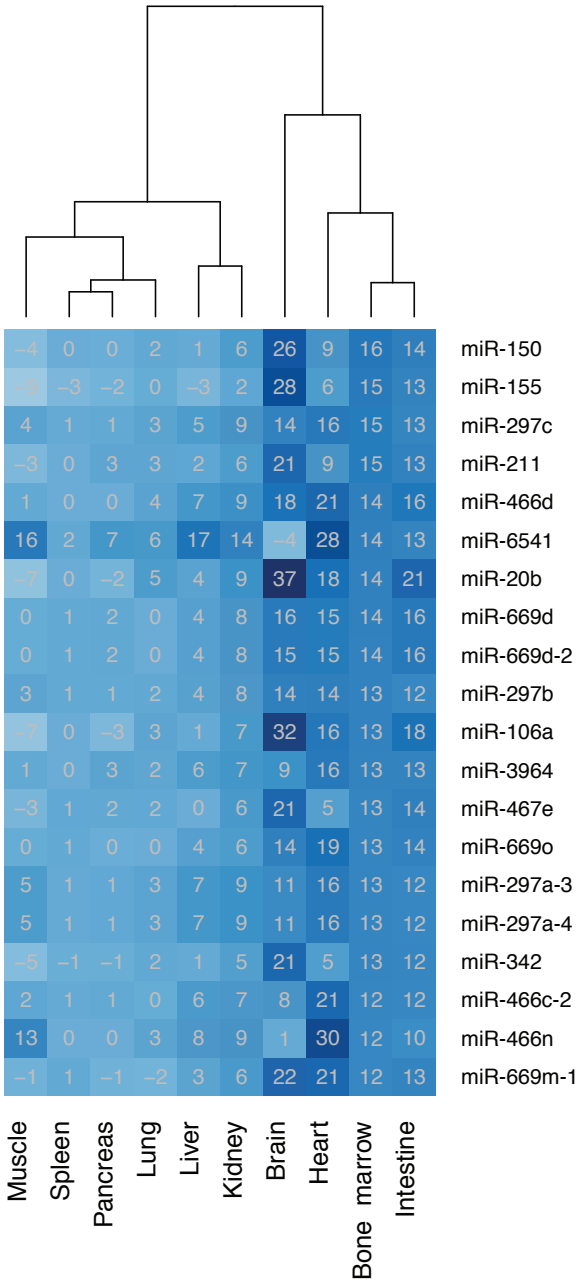
A
PC2: 8% variance
PC1: 9% variance
Tissue
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Testis
B
chr13:84058940-84058982:+
chr11:114847286-114847359:+
miRNA expression (normalized counts)
Tissue
C
Bone marrow
Lung
Muscle
Pancreas
Brain
Spleen
Heart
Intestine
Kidney
Liver
9:105819271-105819328:–
2:165234414-165234461:–
17:57117642-57117705:–
14:45300165-45300210:+
4:132310003-132310065:+
−log10(padj)
log2FoldChange
(log2FoldChange)
Expression (normalized counts)
Expression (normlized counts)
Expressed in
Both
Female
Male
FDR
> 0.01
> 0.01

S8

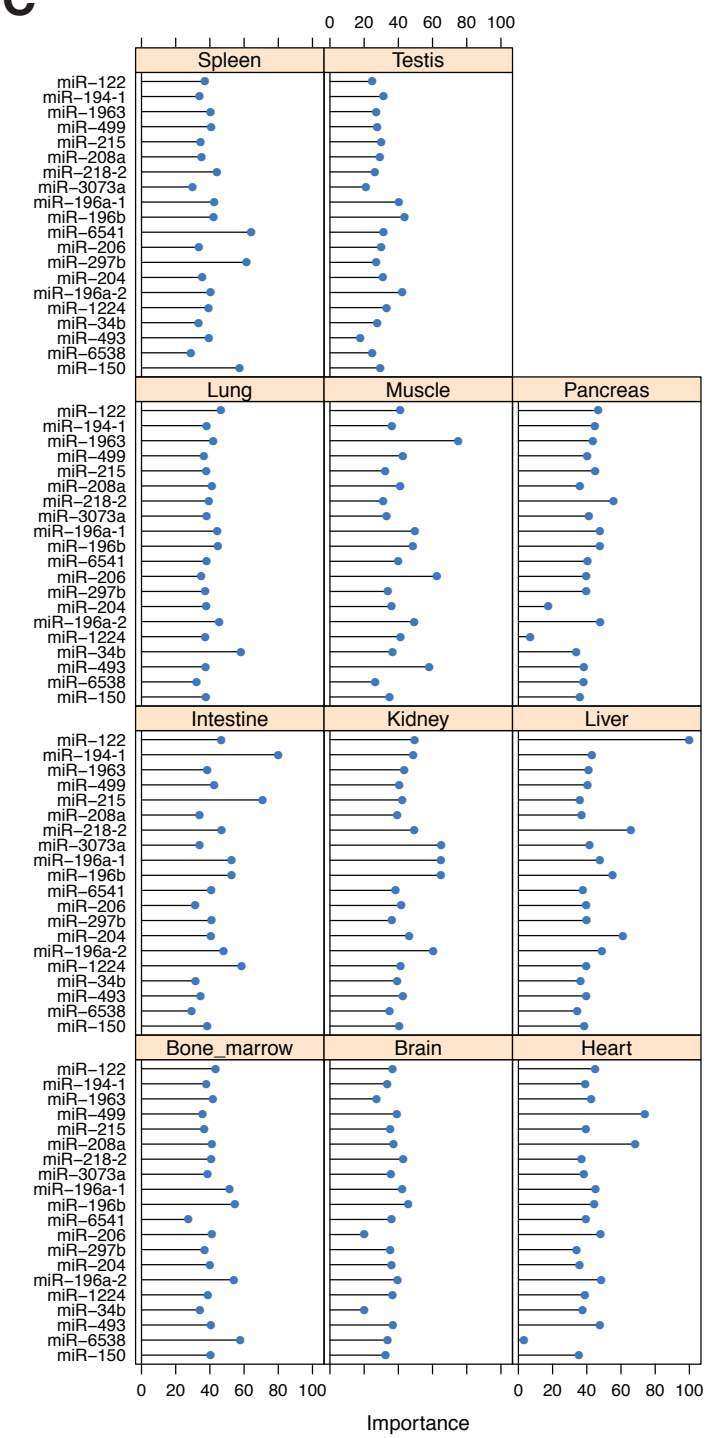
A


Number of samples
0
5
10
Bone marrow
Brain
Heart
Intestine
Kidney
Liver
Lung
Muscle
Pancreas
Spleen
Tissue
miR-103
miR-let7i
miR-145a
miR-122
miR-450
miR-23a
miR-19b
miR-6356
miR-221
miR-5617
miR-378d
miR-322
Spleen
Pancreas
Muscle
Lung
Liver
Kidney
Intestine
Heart
Brain
Bone marrow

B


miR-150
miR-155
miR-297c
miR-211
miR-466d
miR-6541
miR-20b
miR-669d
miR-669d-2
miR-297b
miR-106a
miR-3964
miR-467e
miR-669o
miR-297a-3
miR-297a-4
miR-342
miR-466c-2
miR-466n
miR-669m-1
Muscle
Spleen
Pancreas
Lung
Liver
Kidney
Brain
Heart
Bone marrow
Intestine

C


Spleen
Testis
Lung
Muscle
Pancreas
Intestine
Kidney
Liver
Bone_marrow
Brain
Heart
miR-122
miR-194-1
miR-1963
miR-499
miR-215
miR-208a
miR-218-2
miR-3073a
miR-196a-1
miR-196b
miR-6541
miR-206
miR-297b
miR-204
miR-196a-2
miR-1224
miR-34b
miR-493
miR-6538
miR-150
0
20
40
60
80
100
Importance