A mouse tissue atlas of small noncoding RNA

Alina Isakova[a] , Tobias Fehlmann[b] , Andreas Keller[b,c] , and Stephen R. Quake[a,d,e,1]

[a]Department of Bioengineering, Stanford University, Stanford, CA 94305; [b]Chair for Clinical Bioinformatics, Saarland University, 66123 Saarbrücken, Germany; [c]Department of Neurology, School of Medicine, Stanford University, Stanford, CA 94305; [d]Department of Applied Physics, Stanford University, Stanford, CA 94305; and [e]Chan Zuckerberg Biohub, San Francisco, CA 94158

Small noncoding RNAs (ncRNAs) play a vital role in a broad range of biological processes both in health and disease. A comprehensive quantitative reference of small ncRNA expression would significantly advance our understanding of ncRNA roles in shaping tissue functions. Here, we systematically profiled the levels of five ncRNA classes (microRNA [miRNA], small nucleolar RNA [snoRNA], small nuclear RNA [snRNA], small Cajal body-specific RNA [scaRNA], and transfer RNA [tRNA] fragments) across 11 mouse tissues by deep sequencing. Using 14 biological replicates spanning both sexes, we identified that ~30% of small ncRNAs are distributed across the body in a tissue-specific manner with some also being sexually dimorphic. We found that some miRNAs are subject to "arm switching" between healthy tissues and that tRNA fragments are retained within tissues in both a gene- and a tissue-specific manner. Out of 11 profiled tissues, we confirmed that brain contains the largest number of unique small ncRNA transcripts, some of which were previously annotated while others are identified in this study. Furthermore, by combining these findings with single-cell chromatin accessibility (scATAC-seq) data, we were able to connect identified brain-specific ncRNAs with their cell types of origin. These results yield the most comprehensive characterization of specific and ubiquitous small RNAs in individual murine tissues to date, and we expect that these data will be a resource for the further identification of ncRNAs involved in tissue function in health and dysfunction in disease.

miRNA | noncoding | sex dimorphism | tissue specificity

Small noncoding RNAs (ncRNAs) are a large family of endogenously expressed transcripts, 18 to 200 nt long, that play a crucial role in regulating cell function (1, 2). Seen mainly as "junk" RNA of unknown function two decades ago, today small ncRNAs are believed to be involved in nearly all developmental and pathological processes in mammals (2–4). While the exact function of many ncRNAs remain unknown, numerous studies have revealed the direct involvement of various small ncRNAs in regulation of gene expression at the levels of posttranscriptional mRNA processing (5–7) and ribosome biogenesis (8). Aberrant expression of small ncRNAs, in turn, has been associated with diseases such as cancer, autoimmune disease, and several neurodegenerative disorders (9, 10).

Mammalian cells express several classes of small ncRNA, including microRNA (miRNA) (11), small interfering RNAs (siRNA), small nucleolar RNAs (snoRNA) (12), small nuclear RNA (snRNA) (13), PIWI-interacting RNA (piRNA) (14), and tRNA-derived small RNAs (tRFs) (15), with some being shown to be expressed in a tissue- (16), cell type- (17), or even cell state-specific manner (18). Through their interactions with ribosomes and mRNA, these small noncoding molecules shape the dynamic molecular spectrum of tissues (4, 17). Despite extensive knowledge of ncRNA biogenesis and function, much remains to be explored about tissue- and sex-specific small ncRNA expression. Given the emerging role of ncRNAs as biomarkers (19, 20) and potent therapeutic targets (21), a comprehensive reference atlas of tissue small ncRNA expression would represent a valuable resource not only for fundamental but also for clinical research.

The first attempts to catalog tissue-specific mammalian small ncRNAs began a decade ago with characterization of miRNA

levels (16, 22, 23). 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 following decade of miRNA research. Efforts to characterize tissue-specific noncoding transcripts have recently resumed with the advent of RNA sequencing (RNA-seq), which greatly advanced the discovery of novel and previously undetected miRNA (24, 25). However, no prior study encompasses a spectrum of mammalian tissues from both female and male individuals, nor includes the other noncoding RNA types that have recently been identified to carry out tissue- and cell type-specific functions (26, 27).

Here, we describe a comprehensive atlas of small ncRNA expression across 11 mammalian tissues. Using multiple biological replicates ($n = 14$) from individuals of both sexes, we mapped tissue-specific as well as broadly transcribed small ncRNA attributed to five different classes and spanning a large spectrum of expression levels. Our data reveal that tissue specificity extends to ncRNA types other than miRNA and provide insights on the tissue-dependent distribution of miRNA arms and tRNA fragments. We have also discovered that certain miRNAs are broadly sexually dimorphic, while other show sex bias in the context of specific tissues. Finally, integrating our ncRNA expression measurements with the scATAC-seq data (28) enabled us to map cell type specificity of small ncRNA expressed in the adult mouse brain.

Results

Small ncRNA Expression Atlas of Mouse Tissues. We profiled the expression of small ncRNA across 10 tissues from adult female ($n = 10$) and 11 from adult male ($n = 4$) C57BL/6J mice (Fig. 1A and Dataset S1). We generated a dataset comprising a total of 140 small ncRNA sequencing libraries from brain, lung, heart, muscle, kidney, pancreas, liver, small intestine, spleen, bone

---

Significance

We report a systematic unbiased analysis of small RNA molecule expression in 11 different tissues of the model organism mouse. We discovered uncharacterized noncoding RNA molecules and identified that ~30% of total noncoding small RNA transcriptome are distributed across the body in a tissue-specific manner with some also being sexually dimorphic. Distinct distribution patterns of small RNA across the body suggest the existence of tissue-specific mechanisms involved in noncoding RNA processing.
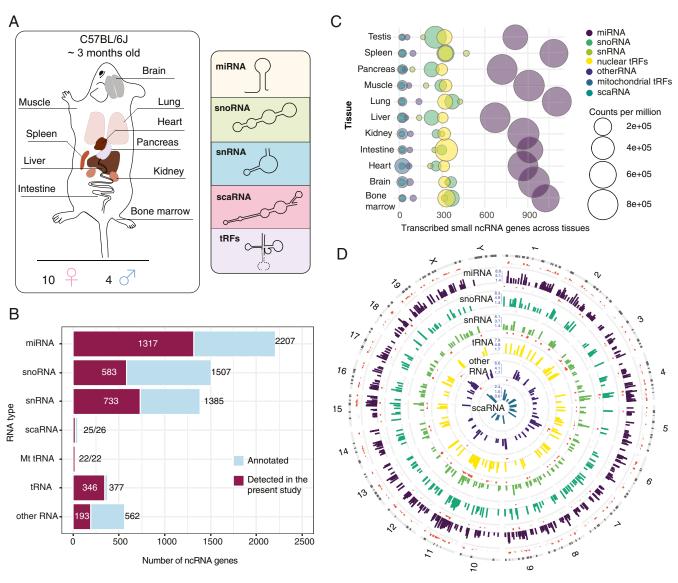

Fig. 1. Small ncRNA expression across mouse tissues. (*A*) Tissues and ncRNA classes profiled in the current study. Ten somatic tissues were collected from adult mice (*n* = 14). Testes were collected from male mice (*n* = 4). (*B*) ncRNAs identified in the current study. Numbers indicate detected and total annotated within GENCODE M20 miRNA, snoRNA, snRNA, scaRNA, Mt_tRNA, as well as high-confidence tRNA listed in GtRNAdb. (*C*) Coverage of ncRNA types within the profiled tissues. ncRNA was considered transcribed in a tissue if detected at >1 cpm. (*D*) Genomic map of small RNAs (sRNAs) expression across mouse genome. The bars show the log-transformed normalized expression count of ncRNAs. The red and gray bars around each circle represent the variance of each sRNA across 10 mouse tissues. Red denotes highly (the SD of expression above 25% of the mean value), and gray, low, variable ncRNAs (SD below 25% of the mean value).

marrow, and testes RNA. Each library yielded ~5 to 20 million reads mapping to the mouse genome, out of which, on average ~7 million mapped to the exons of small ncRNA genes (*Materials and Methods*), resulting in a total of ~100 million ncRNA reads per tissue (*SI Appendix*, Fig. S1A). Using the GENCODE M20 (29), GtRNAdb (30), and miRBase (31) mouse annotations, we mapped the expression of distinct small ncRNA classes: miRNA, snRNA, snoRNA, scaRNA, tRF, and other small ncRNA in profiled tissues (Fig. 1*B*). Among all of the tissues, we identified a total of 1,317 distinct miRNA, 733 snRNA, 583 snoRNA, 25 scaRNA, 346 tRNA, 22 mitochondrial tRNA, and 193 other miscellaneous small ncRNA genes, which correspond to 60%, 53%, 39%, 96%, 92%, 100%, and 34% of annotated transcripts of each respective class (Fig. 1*B*). miRNA was the most abundant small ncRNA type in our libraries, followed by snoRNA, snRNA, and tRFs (Fig. 1*C* and *SI Appendix*, Fig. S1B).

snoRNA and snRNA are believed to yield incomplete recovery in small RNA-seq experiments due to their secondary structure (32). With respect to protein coding genes, the detected tRFs were intergenic, snoRNAs were of intronic origin, snRNAs and scaRNAs were intronic and intergenic (63/35% and 64/36%, respectively), and miRNAs were transcribed from either introns (53%), exons (11%), or intergenic regions (11%) (*SI Appendix*, Fig. S1C). The number of distinct ncRNA greatly varied across tissues; for example, lymphoid tissues (lung, spleen, and bone marrow) contained the largest number of distinct ncRNA, while pancreas and liver contained the fewest (Fig. 1*C* and *SI Appendix*, Fig. S1B). Furthermore, within the profiled tissues, we detected 95.1% of miRNA precursors denoted by miRBase v22 database as high-confidence transcripts (31). Using these data, we have reconstructed genome-wide expression map of various

small ncRNA types across 11 murine tissues (Fig. 1*D* and Dataset S2).

Tissue-Specific Expression of Small ncRNA. We first assessed the differences in the levels of small ncRNAs across profiled tissues at the gene level, based on the expression of all assayed RNA types. Dimensionality reduction via *t*-distributed stochastic neighbor embedding (*t*-SNE) (*Materials and Methods*) on ncRNA genes revealed a robust clustering of samples according to tissue types (Fig. 2*A*). For each ncRNA, we have computed the tissue specificity index (TSI), as described previously in ref. 33. We observed that ~17% of all detected ncRNA were present in only one tissue (TSI = 1) (*SI Appendix*, Fig. S2*A*), while the remaining ncRNA were either ubiquitously expressed or found in some but absent in other tissues. We next ran a differential gene expression (DGE) analysis on all detected ncRNAs across 11 tissues (*Materials and Methods*) and found that out of 3,219 detected genes, 897 (28%) contribute to the tissue-specific signature of ncRNA expression (at a false discovery rate [FDR] < 1%) (Fig. 2 *B* and *C* and Dataset S3). Interestingly, we found brain to contain the highest number of unique transcripts not present in other tissues (~400) (Fig. 2 *B* and *C* and Dataset S3) even though lung, spleen, and bone marrow expressed the widest spectrum of detected genes (Fig. 1*C*). We found miRNA to be the main contributor of tissue specificity (reflected by the fraction of each specific RNA type with TSI > 0.9) (Fig. 2 *B* and *C*); however, we have also identified hundreds of ncRNAs of other types that are expressed in a tissue-specific fashion (*SI Appendix*, Fig. S2).

Tissue-Specific snoRNAs. We found that snoRNA alone is capable of separating the majority of profiled tissues based on their transcript levels (*SI Appendix*, Fig. S2*B*), with over 200 snoRNA showing tissues-specific patterns (*SI Appendix*, Fig. S2*C* and Dataset S3). For example, we discovered that maternally imprinted *AF357428* (also known as MBII-78), *AF357341* (MBII-19), and *Gm25854*, transcribed from a 10-kb region of chromosome 12qF1, are up-regulated in the brain and muscle. Interestingly, two other snoRNA, *Gm22962* and *Gm24598*, followed the same tissue-specificity pattern, despite being transcribed from other chromosomes (9qC and XqA7.1, respectively). While present at low levels, we also identified *Snora35* (MBI-36) and *Snord116* (MBII-85), known to be involved in neurodevelopmental disorders (34), to be brain exclusive (TSI = 1). We observed high levels of *Snord17* and *Snord15a* in the spleen and bone marrow, and lower levels in other tissues. These snoRNAs have been previously reported among up-regulated genes in bacterial infection of soft tissues (35), suggesting the association of these transcripts with immune cells. We found several snoRNAs present mainly in the pancreas, such as *Snord123* (*SI Appendix*, Fig. S2*C*), located 3 kb upstream of the pancreatic cancer-associated *Sema5a* gene, and *Gm22888* (Fig. 2*B*) located within the introns of *Ubap2*. We also identified a large number of other snoRNA, the exact function of which is still unknown, to be enriched in either one or multiple tissues (*SI Appendix*, Fig. S2*C*), among which are *Snord53*, *Gm24339*, *Gm26448*, *Snora73a*, *Snord104* in lymphoid tissues, and *Snord34*, *Gm24837* in testes. Finally, we show that some snoRNAs, such as *Snord70* and *Snord66*, which are often used as normalization controls in qPCR-based assays (36, 37), are also expressed in a tissue-biased manner (*SI Appendix*, Fig. S2*C*).

Tissue-Specific Expression of *Rny*, *Terc*, and Other ncRNA. Analyzing the levels of other ncRNA classes, we found that the brain contains high levels of *Rny1* compared to other tissues (Fig. 2*B*). We also observed that the levels of another transcript from the same class, *Rny3*, are elevated in pancreas, brain, and kidney (*SI Appendix*, Fig. S2*D*). The precise biological function of *Rny1* and *Rny3* is so far undefined, although they have been suggested to maintain RNA stability (38).

Interestingly, we detected the presence of telomerase RNA component (*Terc*) in analyzed somatic tissues, with the highest levels seen in the bone marrow and spleen. Together with previous reports that identify telomerase activity in hematopoietic cells (39) and show *Terc*+ cells to secrete inflammatory cytokines (39, 40), our data suggest that *Terc* is specific to cells of hematopoietic origin. Among other ncRNA types that we found to be differentially expressed across profiled tissues are snRNA and scaRNA, both known to be involved in the regulation of splicing events (41). We observed that, similarly to primate orthologs (42), mouse Rnu11 and Scarna6 are preferentially found in lymphoid tissues, while three snRNAs of unverified function, *Gm25793*, *Gm22448*, and *Gm23814*, are specific to the brain (*SI Appendix*, Fig. S2*D*).

Tissue-Specific miRNA. We found ~400 miRNAs differentially expressed across profiled tissues (Dataset S3). Out of these 400, nearly one-quarter are specific to the brain, with some being uniquely expressed within the tissue (*SI Appendix*, Fig. S3*A*). We identified both well-described brain-specific miRNAs, such as *mir-9*, *mir-124*, *mir-219*, *mir-338* (16, 24, 33), as well as those which are missing from existing catalogs, such as *mir-666*, *mir-878*, *mir-433*, etc. (Fig. 2*B* and *SI Appendix*, Fig. S3*A*). Examples of other miRNAs previously unknown to be tissue-specific include *mir-499* in the heart, *mir-3073b* in the kidney, and *mir-215* and *mir-194* in the intestine (*SI Appendix*, Fig. S3*A*). We also observed multiple miRNAs present in several tissues but absent in others, reflecting the cellular composition of the tissues. Surprisingly, we also found a few miRNAs, such as *mir-134*, *mir-182*, *mir-376c*, *mir299a*, *mir-3061*, and *mir-7068* (*SI Appendix*, Fig. S3*A*) to be shared solely between muscle, brain, and pancreas, which, in turn, do not contain any evident common cell types that are absent in other tissues. Independently, unsupervised clustering of the top 400 most differentially expressed ncRNA in our dataset also revealed that two out of three identified clusters comprise ncRNA genes that are up-regulated in the above-mentioned three tissues (*SI Appendix*, Fig. S3*B*). Altogether, these findings suggest that certain small ncRNAs are involved in maintaining a specific function within the brain, pancreas, and muscle, which could, for example, be ion transport or exocytosis.

Tissue-Specific Arm Selection of miRNA. Assessing the overall abundance of 5p or 3p arms of miRNA across tissues, we found no significant bias in strand selection (*SI Appendix*, Fig. S4*A*). For many miRNAs, we generally observed the dominance of either 3p or 5p arm, while for some we also detected high levels of both arms present in one or multiple tissues (Fig. 2 *C* and *D* and Dataset S4). Nonetheless, we found that ~5% of all miRNAs switch their arm preference between tissues. Some of them, like *mir-337*, *mir-106b*, and *mir-26b*, are represented by both arms in certain tissues but only by one of the arms in other (Fig. 2*D* and Dataset S4). More striking examples of complete arm switching from one tissue to another are *mir-141* and *mir-350* (Fig. 2*D*). miR-141-5p but not -3p is present in the brain, and -3p but not -5p in the testes, while both arms are found in the pancreas and intestine. In the case of *mir-350*, both arms are detected in the bone marrow, brain, and spleen, while only the 5p arm is present in the heart, lung, and muscle, and the 3p arm in the pancreas (Fig. 2*D*). This highlights the complexity of tissue-dependent miRNA biogenesis and indicates that the phenomena of miRNA arm switching, so far only observed in cancer, extends to healthy mammalian tissues (43–45).

Ubiquitous ncRNA Transcripts. We detected many ubiquitously expressed ncRNAs across tissues (*SI Appendix*, Fig. S4*B*). Among these are ncRNAs known to be expressed in a large number of cell types, such as *let-7d-3p*, *miR-320-3p* (25), ncRNAs
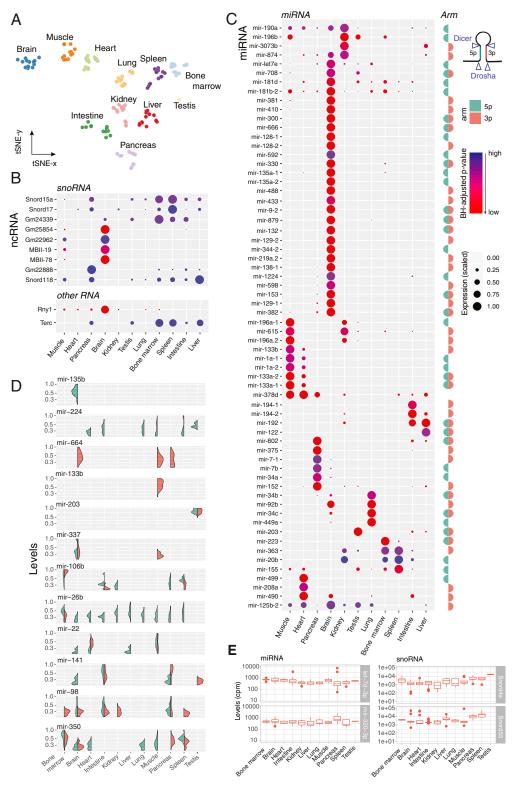
Fig. 2. Tissue-specific patterns of small ncRNA expression. (*A*) *t*-SNE projection of ncRNA expression patterns performed on ~4,000 ncRNA genes of various classes detected in 11 mouse tissues. (*B*) Dot plot of tissue-specific snoRNA, *Rny1* and *Terc*, identified as the most tissue-specific (Benjamini–Hochberg [BH]-adjusted *P* value < 0.01 in LRT test). The size of the dot represents scaled log-transformed normalized counts. (*C*) Dot plot of tissue-specific miRNA. Only a subset of miRNA passing the specificity threshold (BH-adjusted *P* value < 0.01 in LRT test) is shown. "Arm" column denotes whether *5p-*, *3p-*, or both arms are passing the specificity threshold. (*D*) Levels of miRNA arms detected across tissues. Representative miRNAs, for which we consistently detect either one of the arms, both, or switched arm between tissues. The *y* axis represents normalized scaled counts. (*E*) Examples of ubiquitous ncRNA present in all tissues.

the cell type specificity of which is still unknown, such as *Snord4a* and *Snord55*, as well as those known to be expressed in the cell types that are abundant in all tissues (like endothelial *miR-151-5p*) (*SI Appendix*, Fig. S4B).

Novel miRNAs. We have recently demonstrated that the current miRbase annotation of mammalian miRNA remains incomplete but can be readily expanded with the help of emerging small RNA-seq data (46). To search for novel miRNA in our data, we first processed all unmapped reads using miRDeep2 (47) and selected 473 genomic regions harboring a putative miRNA gene supported by at least 10 sequencing reads. To refine this list, we employed three parallel strategies: 1) we searched for the presence of the putative miRNA in 141 public Argonaute CLIP-seq (AGO-CLIP) datasets from various mouse cell and tissue types; 2) we performed a literature and database search for prior mentions of the putative miRNA; and 3) we ran an RT-qPCR validation of selected candidates. Analysis of AGO-CLIP data showed evidence for 214 out of 473 candidates (total, >5 counts). We also found that 87 out of 473 novel miRNA were previously reported within other studies (48–52) (Fig. 3A). Importantly, 52 novel miRNAs identified by this and previous studies were not present in the AGO-CLIP data (Dataset S5). The RT-qPCR quantification of two miRNAs selected from this list, *17_11530* and *7_16137*, in turn, confirmed the existence of these transcripts (*SI Appendix*, Fig. S5A). On the other hand, we identified novel miRNAs (*17_8620*, *4_6440*, *9_15723*) that are supported by AGO-CLIP data, prior reports, or both, but for which we could not confirm the existence through RT-qPCR (*SI Appendix*, Fig.

S5A). We also found a novel miRNA that, among the three validation methods, was only verified through RT-qPCR. Interestingly, the genomic coordinates of this miRNA, *14_6588*, matched the coordinates of another, annotated one, *mir-802a*. Unlike *mir-802a*, however, *14_6588* is transcribed from the negative DNA strand and is only present in the brain (*SI Appendix*, Fig. S5B). Altogether, by comparing the miRNA levels measured through RT-qPCR with the tissue transcript abundance identified by small RNA-seq, we validated 12 novel miRNAs that were either also reported by others (Fig. 3B and *SI Appendix*, Fig. S5A) or uniquely identified in the present study (Fig. 3 C and D and *SI Appendix*, Fig. S5A).

We found that the majority of putative miRNAs are present in only one tissue (312), but a small number (4) are found in all 11 tissues (Fig. 3E). 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 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 (*SI Appendix*, Fig. S5C). Differential expression analysis on putative novel miRNAs identified six miRNAs to be also expressed in a sex-specific manner. Strikingly, all six were male-dominant, with one of them even found to be consistently up-regulated in two tissues, male muscle and pancreas (*SI Appendix*, Fig. S5D).

Tissue-Resident tRNA Fragments. About a quarter of our small RNA-seq libraries consisted of tRFs—fragments of either mature
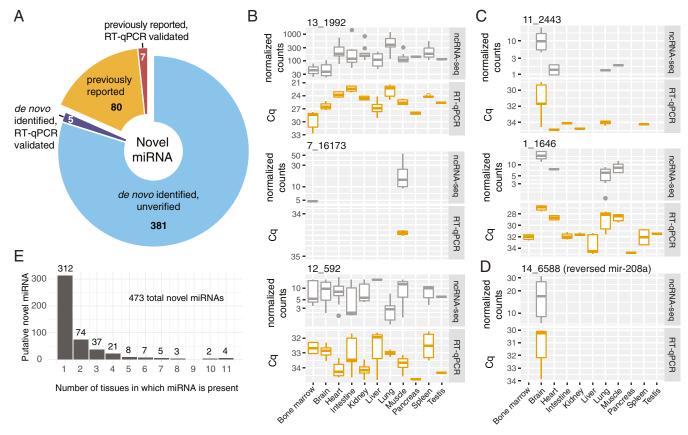

Fig. 3. miRNAs uniquely detected in the present study. (*A*) Pie chart of predicted and verified miRNA. (*B*) Examples of RT-qPCR verified miRNAs identified by this and previous studies. The levels of miRNA were determined from small RNA-seq data (DESeq2 normalized counts) and through RT-qPCR (Cq, adjusted for the sample-to-sample variability using cel-mir-39 spike in control). (*C*) Same as C but for miRNAs detected uniquely within the present study. (*D*) RT-qPCR verified miRNA, 14_6588, transcribed from the negative strand of mir-208a. (*E*) Tissue specificity of putative miRNA.

Isakova et al.

of precursor tRNA molecules enzymatically cleaved by angiogenin (Ang), Dicer, RNaseZ, and RNaseP (7, 53) (Fig. 4A).

"Exact tRNA multimapping" of these fragments to the mouse genome revealed the presence of tRFs of various sizes. Interestingly, consistent with a previous report on tRFs in human cell lines (7), we observed a major difference in the size of fragments originating from either nuclearly or mitochondrially encoded tRNA (ntRNA and mtRNA, respectively). While the majority of ntRNA fragments were 33 nt long, mtRNA fragments spanned a large size range of 18 to 54 nt (Fig. 4B and *SI Appendix*, Fig. S6A). This distinct pattern of fragment sizes reflected the bias in the amounts of tRF types originating from nt- and mtRNA (Fig. 4C and *SI Appendix*, Fig. S6B). We observed that the distribution of nuclear tRFs was largely skewed toward 5′tR-halves, generated by the cleavage in the anticodon loops of mature tRNA. However, within mitochondrial tRFs, we identified a more uniform representation of cleaved fragments. Furthermore, we found that the relative abundance of tRF types, within both ntRNA and mtRNA space, is not constant, but varies across tissues (Fig. 4 C and D). The tissue-type differences are also present across different tRNA isoacceptors and even its anticodons (Fig. 4D and *SI Appendix*, Fig. S6B). In the case of nuclear tRFs, the vast majority of fragments in each tissue were attributed to glycine, glutamine, valine, and lysine tRNA, with the intestine containing the largest amounts of the respective 5′tR-halves. Since the abundance of these specific fragments has been shown to correlate with the levels of functional angiogenin in the cell (54), we speculate that the biological explanation of the intestine yielding high levels of tRFs is due to the activity of Ang4, one of the five Ang proteins in mouse, highly expressed in Paneth and Goblet cells of the intestinal epithelium (55, 56).
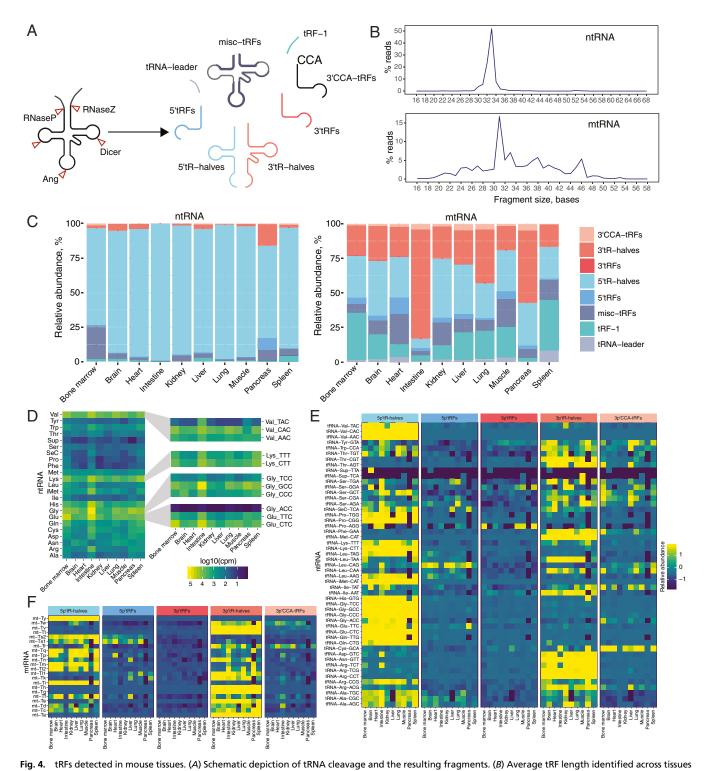
For many ntRFs, the distribution between 3′- and 5′-, tRF and tR-halves was surprisingly shifted toward one form, i.e., one fragment type was present at higher amounts than others (Fig. 4 E and F). For the majority of fragments, we found 5′tR- or 3′tR-halves to be the most dominant fragment type. However, in the rare cases, we found 5′- or 3′tRFs to dominate other fragments in a tissue-specific manner. An example of such a fragment is 5′tRF Glu-TTC, which we found to be enriched in the pancreas, compared to other tissues that mostly contained Glu-TTC 5′tR-halves. mtRFs followed a similar trend of fragment shift. We found 5′tRFs of proline-transferring mt-Tp in the heart and 5′tRFs of asparagine-transferring mt-Tn in the liver, while within other tissues we detected different fragment types of these tRNAs (Fig. 4F).

miRNAs Are Expressed in a Sex-Specific Manner.

Several groups have observed a sex bias in the levels of miRNA in blood, cancer tissues, and human lymphoblastoid cell lines (57, 58). To investigate whether this phenomenon extends to healthy tissues, we compared the ncRNA levels within each tissue coming from either female or male mice. Among ~6,000 genes assigned to various ncRNA classes, we identified several miRNAs to be differentially expressed between females and males (at FDR < 0.01) (Fig. 5A). Some of them are globally sexually dimorphic, while the majority are sex-biased only within a specific tissue. In each somatic tissue, except pancreas, we identified at least two miRNAs differentially expressed between sexes (log2Fold-Change > 1, normalized counts > 100, FDR < 0.01) (*SI Appendix*, Fig. S7 A and B). Kidney and lung contained the highest number of sex-biased miRNAs (27 and 18, respectively), while only two were detected in the heart, five in the muscle, and seven in the brain (Fig. 5B and *SI Appendix*, Fig. S7A). Three out of eight female-dominant miRNAs: *mir-182*, *mir-148a*, and *mir-145a*, were also shown previously to be estrogen regulated (59), while another miRNA, *mir-340*, was reported to be down-regulated in response to elevated androgen levels (60). Interestingly, we also found that four out of five male-specific miRNAs

in the brain are transcribed from a 5-kb region of imprinted Dlk1-Dio3 locus on chromosome 12 (Fig. 5C).

Given the innate ability of miRNA to lower the levels of target mRNA (61), we hypothesized that the levels of protein-coding transcripts targeted by sex-biased miRNA would also differ across male and female tissues. To test this hypothesis, we correlated the expression of sex-biased miRNAs with the levels of their respective target mRNAs across profiled tissues (*Materials and Methods*). Among the 60 anticorrelated targets ($r_s < −0.8$, FDR < 0.1) we identified two genes previously shown to be sexually dimorphic (*SI Appendix*, Fig. S7C). Specifically, we found *miR-423*, up-regulated in male lung and bone marrow, 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*).

ncRNA-Based Tissue Classification.

It is natural to wonder whether the observed variation in ncRNA expression across tissues (Fig. 2 B and C) would be sufficient to accurately predict the tissue type based solely on small RNA-seq data. To address this question, we set out to construct an algorithm that can learn the characteristics of a healthy tissue from the data reported in the current study and make predictions on new datasets. We limited our analysis to miRNA, since high-throughput tissue data for other ncRNA types is not available. We trained a support vector machine (SVM) model (62) on datasets generated in this study, each containing the expression scores for 1,973 miRNAs (*SI Appendix*, Fig. S8A). As a validation dataset, we used available miRNA-seq data released by the ENCODE consortium for multiple mouse tissues (63). 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 (Dataset S6). 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, and muscle samples, yielding an overall accuracy of 0.96 (*Materials and Methods*) (see Fig. 6A for a full list of accurately predicted tissues as well as for false positives/negatives). 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 instead of assigning them to bone marrow (Fig. 6A). Strikingly, in this case, our model accurately predicted the hematopoietic composition of the organ, known to shift from the liver at the embryonic stages to the bone marrow in adulthood (64), rather than the tissue type itself. Furthermore, we identified hematopoiesis-associated *miR-150* and *miR-155* (65) to have highest weights among the features defining the bone marrow in our model (*SI Appendix*, Fig. S8B).

We next asked how the identified tissue expression patterns compare to those of individual cell types. To investigate this, we correlated our data with the miRNA data generated for primary mammalian cells by FANTOM5 consortium (25). Comparing mouse samples first, we found that FANTOM5 embryonic and neonatal cerebellum tissues strongly correlated with our brain samples ($r_s = 0.89$ to 0.9), while erythroid cells had the strongest correlation with spleen and bone marrow ($r_s = 0.93$) (*SI Appendix*, Fig. S8C). 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 (*SI Appendix*, Fig. S8D). Spearman correlation coefficients reflected the cell-type composition of tissues (*SI Appendix*, Fig. S8E). 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 correlated best with neural stem cells, spinal cord, and pineal and pituitary glands ($r_s = 0.49$) (*SI Appendix*, Fig. S8E).

GENETICS

Fig. 4. tRFs detected in mouse tissues. (*A*) Schematic depiction of tRNA cleavage and the resulting fragments. (*B*) Average tRF length identified across tissues for either ntRNA-derived (*Top*) or mtRNA-derived (*Bottom*) fragments. (*C*) Fragment type abundance across tissues. (*D*) Heatmap of tRF levels in tissues. For each of the 23 tRNA types, the sum among its tRFs is plotted. (*E*) Heatmap of relative abundance of ntRNA fragment types (5′-tR-halves, 3′-tR-halves, 5′-tRFs, 3′-tRFs, or 3′-CCA-tRFs) for each tRNA isoacceptor across 10 somatic tissues. Relative abundance is represented by row-wise scaled fractionated scores of tRFs computed by *unitas*. (*F*) Same as *E* but for mtRNA.

Integration of Small RNA-Seq and scATAC-Seq Data. Finally, to deconvolute the complex noncoding tissue profiles and identify the cell types that contribute the observed tissue-specific ncRNAs, we integrated the sequencing data generated within our study with a previously published single-cell ATAC-seq atlas—a catalog of single-cell chromatin accessibility profiles across various cell types (28). First, we compared the expression scores predicted through ATAC-seq measurements with our estimates of ncRNA expression, derived from small RNA-seq. Using the top 400 ncRNAs identified in our analysis as tissue-specific, we correlated average ATAC-seq activity scores and ncRNA levels across eight tissues for which both types of data were available. We
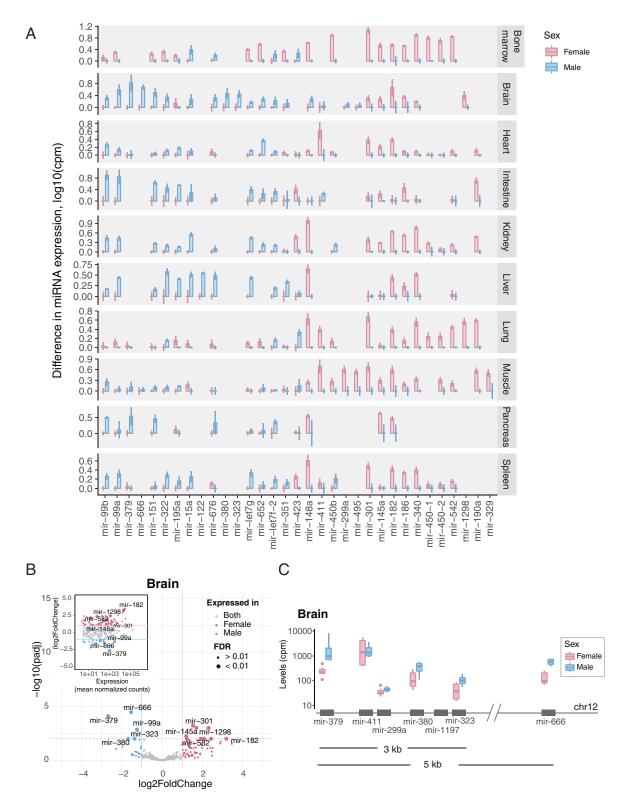
Fig. 5. Sex-specific miRNA expression. (*A*) Sex-dimorphic miRNAs identified in this study. The *y* axis represents the log$_{10}$ of the difference between mean miRNA levels computed for males and females in each tissue (in counts per million [cpm]). Error bars denote SD of miRNA levels across a tissue within each sex. (*B*) Volcano plot showing miRNAs differentially expressed between the female and the male brain. (*C*) Expression and genomic location of male-biased miRNA in the brain.

observed a strong correlation of both measurements for the brain, liver, and heart, and a weaker correlation for kidney and mixed scores between spleen, bone marrow, and lung (Fig. 6*B*). We found, however, that within each tissue we could attribute the

cell type of origin to a number of identified tissue-specific ncRNA. For example, in agreement with previous studies, we could identify that muscle-specific *mir-133a-2* is expressed in cardiomyocytes, while *mir-148a* is expressed in hepatocytes and duct
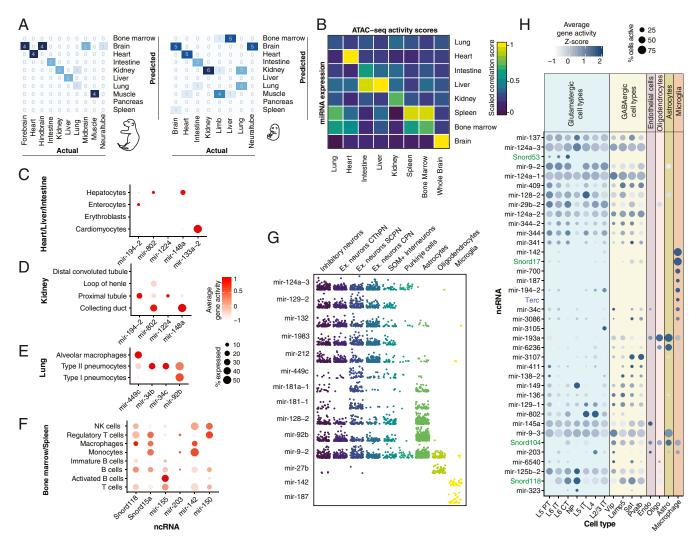
Fig. 6. (*A*) Confusion matrices obtained from SVM tissue classifier for postnatal and embryonic datasets. (*B*) Correlation between small RNA-seq scores derived for top 400 tissue-specific ncRNA genes with the respective activity scores obtained from mouse ATAC atlas. (*C–F*) Average gene activity scores of tissue-specific ncRNAs within each cell type resident to the respective tissue. Gene activity scores for the ncRNAs of interest were retrieved from the mouse ATAC atlas. (*G*) Log$_{10}$ gene activity scores of brain-specific miRNA across individual cells. (*H*) Gene activity scores computed for brain-specific ncRNA from the droplet-based scATAC-seq data (10XGenomics; *Materials and Methods*).

cells (Fig. 6 *C* and *D*) and *mir-194–2* comes from the enterocytes in the gut (24, 25). In addition, we found that in the lung, *mir-449c* in expressed in alveolar macrophages, *mir-34b* and *mir-34c* in type II pneumocytes (Fig. 6*E*), and *mir-155* is found in B cells and even correlates with its maturation status (Fig. 6*F*). Among brain-specific ncRNA, we identified *mir-187* and *mir-142* to be expressed in microglia, *mir-27b* in oligodendrocytes, and *mir-124a-3*, *mir-1983*, *mir-212*, and others in neurons (Fig. 6*G*). For the majority of brain-specific ncRNA identified in our study, however, due to resolution limitations of the mouse ATAC atlas data, we were unable to unambiguously define the cell type of origin. To overcome that, we analyzed a complimentary single-cell ATAC-seq dataset, generated specifically from the mouse adult brain (obtained from 10X Genomics; *Materials and Methods*) and mapped the activity of the brain-specific ncRNA to 15 cell types in the adult mouse brain annotated by the Allen Brain Atlas (66). This analysis revealed that among the brain-specific ncRNA, many are potentially expressed solely in neurons, with some even being predominantly present in either glutamatergic (*Snord53*, *mir-802*) or GABAergic (*mir-3107*) neurons (Fig. 6*H*). Among glia-specific ncRNA, we identified

Snord17 and *mir-700* in macrophages as well as *mir-193a* and *mir-6236* in astrocytes and oligodendrocytes.

Discussion

Small ncRNA plays an indispensable role in shaping cellular identity in health and disease by orchestrating vital cellular processes and altering the expression of protein-coding genes (12). Recent efforts in profiling of the most studied types of small ncRNA, miRNA, across cells and tissues demonstrated the existence of tissue- and cell type-specific short noncoding transcripts (16, 24, 25, 33). In this work, we show that this phenomenon extends beyond one ncRNA class and involves not only tissue-specific but also sex-specific ncRNA expression. The present resource demonstrates that each healthy mammalian tissue carries a unique noncoding signature, contributed by well-understood RNA types as well as by less studied ones.

By analyzing the expression of several classes of ncRNA we discovered that nearly ~900 transcripts contribute to the unique noncoding tissue profile. Moreover, we identified that in addition to variable transcription levels and posttranscriptional modifications, noncoding tissue specificity is achieved through an

unknown mechanism of selective RNA retention. While at this point we are unable to judge the functional significance of this phenomenon, we discovered that, even between healthy tissues, certain miRNA undergo so-called "arm switching"—a process previously thought to be strictly pathogenic in mammals (43, 45, 67). Among other ncRNA class, tRFs, we observed a selective enrichment of certain fragment types over others, happening in both gene- and tissue-specific manners. Taken together with previous observations (7, 68, 69), this finding raises additional questions regarding the biogenesis pathways of tRFs as well as their tissue-specific function.

Within our study, we also report several tissue-specific miR-NAs not identified in previous studies (Fig. 3, *SI Appendix*, Fig. S4, and Dataset S7). The validation process of the identified miRNAs brought to light several important observations. First, we noted that the AGO-CLIP, while often used as a "gold standard" of miRNA validation (24), in fact, does not support the existence of many miRNA independently detected within RNA-seq datasets or directly validated through RT-qPCR. The gap between AGO-CLIP and small RNA-seq data in terms of data quality, diversity, and depth suggests that validating against AGO-CLIP data may not be the optimal approach for miRNA discovery. Instead, one could search for evidence of miRNA expression within publicly available RNA-seq data as a first thresholding step (70). Second, it is important to consider that the genomic location of a novel miRNA might match with that of a previously annotated one, while the molecule itself could be transcribed from an opposite DNA strand. We observed this phenomenon on the de novo identified miRNA 14_6588, whose coordinates strictly overlap with *mir-802a* and that is only present in the brain.

miRNA has been previously used to train classifiers capable of differentiating cancer/tissue types (71). Our work demonstrates that machine learning algorithms applied to quantitative miRNA expression estimates also detect changes related to the cell-type composition of tissues, such as the shift in hematopoietic cell abundance in the postnatal compared to the fetal liver. Given the emerging evidence of ncRNA stability in the blood and its rapid propagation throughout the body within extracellular vesicles (72), we anticipate that the current space of markers used to noninvasively monitor development (73) could be further expanded to small ncRNA.

Small ncRNAs have been long known to regulate the development and function of the brain. Despite the tremendous progress of neuroscience in understanding the regulation of coding genes, surprisingly little is known about cell type-specific small ncRNA in the brain. Even within the available tissue-level ncRNA resources, the brain remains one of the most underrepresented tissue. We believe this is mostly due to technical limitations of small RNA sequencing, which has yet to be applied to single neurons and, so far, still relies on the robust enrichment of certain cell types. Given the extensive molecular heterogeneity of cell types in the brain, one would expect the diversity of ncRNAs in this tissue to be high. Our study finds that the brain, in fact, contains the largest number of unique mammalian ncRNA transcripts that are absent in other tissues. However, our knowledge of cell-specific ncRNA expression is not complete, and thus for the majority of these identified RNA we could not call the cell type of origin based in the data generated within previous studies offering cell type resolution (24, 25). Taking an alternative route and integrating our tissue-level ncRNA measurement with single-cell chromatin accessibility profiles turned out to be surprisingly informative and allowed us to infer the activity of ncRNA within individual neuronal and glial types. While the validation of these cell-specific transcripts through a direct measurement remains highly desirable, the provided ncRNA estimates indicate that ncRNA is another contributor to complexity in the architecture of nervous system.

We found that the lung contains the largest number of distinct small ncRNA among 11 profiled tissues. However, in the case of the lung, open chromatin data did not provide sufficient resolution for us to infer the cell types of origin for the majority of the transcripts. This inability to fully explain the roots of tissue complexity points to the need for further characterization of the ncRNAs content of specific cell types or even, similarly to mRNA, that of single cells (17, 74). This atlas, meanwhile, will hopefully stimulate future small ncRNA studies and serve as a powerful resource of ncRNA tissue identity for fundamental and clinical research.

Materials and 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 mo old, were used (Dataset S1).

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

Library Preparation and Sequencing. Short RNA libraries were prepared following the Illumina TruSeq Small RNA Library Preparation kit (catalog #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 from 135 to 250 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) following ENCODE small RNA-seq pipeline (63), with minor modifications. We used STAR v2.5.1 (75) with the following parameters: –outFilterMismatchNoverLmax 0.04–outFilterMatchNmin 16–outFilterMatchNminOverLread 0–outFilterScoreMinOverLread 0–alignIntronMax 1–outMultimapperOrder Random–clip3pAdapterSeq TGGAATTCTC–clip3pAdapterMMp 0.1. We allowed incremental mismatch: no mismatches in the reads ≤25 bases, 1 mismatch in 26 to 50 bases, and 2 in 51 to 75 bases. Spliced alignment was disabled. We additionally filtered out reads "soft-clipped" at the 5′-end but kept 3′-clipped ones to account for miRNA isoforms and tRNA modifications. We used GENCODE M20 (29) and miRBase v22 (31) annotations to count the number of ncRNA transcripts. For snoRNAs, snRNAs, scaRNA, miscRNA, or miRNA quantification, reads were assigned to the respective genes using *featureCounts v 1.6.1* (76) with the following parameters: -a Mus_musculus.M20.gtf -M –primary -s 1. Read spanning two overlapping exons were excluded. To account for the multimappers, we used -M -primary option, which only counts a "primary" alignment reported by STAR (either a location with the best mapping score or, in the case of equal multimapping score, the genomic location randomly chosen as "primary"). This quantification approach largely agreed with the results obtained through mapping and quantification against the short nucleotide library (77) (*SI Appendix*, Fig. S9). However, it proved to be more inclusive for the reads uniquely mapping within the miRNA exon but missing one base at the 5′ prime end of the molecule and more strict in counting reads mapping elsewhere in the genome, for which the levels were consistently overestimated by the other method. All reads mapping to miRNA arms and to stem loops were used to quantify miRNA expression at the gene level. For tRF quantification, for each library, we first extracted reads mapped by STAR to the GENCODE-annotated tRNA within the mouse genome (30, 78). We then ran *unitas* (78) on these reads and used fractionated scores to compute the differences in tRF abundance across tissues.

Unsupervised Clustering and Dimensionality Reduction Analysis. Raw counts were normalized and log-transformed using *DESeq2* package. Batch effects were corrected using *limma* R package (79). 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*-SNE with the following parameters: perplexity = 20 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.

TSI. To compute the tissue specificity index, we used the formula described previously in ref. 33:

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

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 coefficients of correlation between samples generated in the current study and the mouse miRNA data generated by FANTOM5 consortium (25), we used *DESeq2*-normalized scores of 2,207 annotated miRNAs. To compare miRNA expression between mouse tissues and human cell types, we generated a curated list of miRNA orthologs, each of which contained a maximum of two mismatches per mature miRNA. In total, 531 miRNAs passed this criteria and were used to compute Spearman correlation coefficients shown in *SI Appendix*, Fig. S8.

Differential Expression Analysis with DESeq2. We used a likelihood-ratio test (LRT) implemented in *DESeq2* (80) to compute the significance of each gene in tissue-specific expression. Briefly, LRT compares whether the tissue type parameter, removed in the "reduced" (~ Batch + Sex), compared to "full" model (~ Batch + Sex +Tissue, in DESeq2), explains a significant amount of variation in the data. Statistical significance of the test (*P* values) was calculated by comparing the difference in deviance between the "full" and "reduced" model to χ^2 distribution. *P* values obtained from the LRT test were corrected using Benjamini–Hochberg procedure to obtain an FDR estimate of tissue specificity scores for each gene. Gene clusters in *SI Appendix*, Fig. S3B were computed on 250 differentially expressed genes (P_{adj} < 1e-90 and base mean > 3) using *DEGreport* R package (81). miRNAs differentially expressed between female and male tissues were computed based on uniquely mapping counts (excluding multimappers), using Wald test within *DESeq2*. To test for the NULL hypothesis, we performed a permutation test in which we randomly reassigned 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) (*SI Appendix*, Fig. S7A). We used Benjamini–Hochberg-corrected *P* values (FDR) to assess the statistical significance of the computed DE scores (Fig. 5 and *SI Appendix*, Fig. 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, DI-ANA, miRanda, or mirDB databases (82, 83). Only targets 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 (84) (Dataset S7). Spearman correlation coefficients were computed between fragments per kilobase of transcript per million mapped reads retrieved from the ENCODE mRNA expression tables and *DESeq2*-normalized miRNA counts across 10 profiled tissues using corr.test() function from "psych" R package, and thresholded above Benjamini–Hochberg-adjusted *P* value of 0.1 and Spearman correlation coefficient ($-0.8 < r_s < 0.8$).

Identification of Identified Candidate miRNA. Candidate miRNAs were identified using miRDeep2 software (47). Only miRNAs supported by >5 reads were reported in this study. AGO-CLIP data were mapped to the mouse genome using STAR (same as for small RNA-seq libraries described in *Data Processing*) and the reads falling within the putative miRNA coordinates were counted using *featureCounts*. We counted a putative miRNA as "supported" if it had >5 AGO-CLIP counts.

To search for the previous mentions of identified miRNAs, we looked up their sequences in miRCarta (85) and used the Google search engine to query

the literature. Candidate miRNAs were ranked by novoMiRank scores, which we computed as described in ref. 85.

For independent validation, we performed RT-qPCR using custom Small RNA TaqMan probes (Life Technologies; catalog #4398987) designed on the star consensus sequence reported by miRDeep2. We used 0.5 ng of total RNA per tissue sample supplied with Cel-mir-39 spike-in (Qiagen; catalog #339390) to perform the reactions in a final volume of 20 μL.

We analyzed tissue and sex specificity of identified 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.

miRNA-Based Classifier. We trained the radial kernel SVM model on 136 samples corresponding to different tissue types (*SI Appendix*, Fig. S8A) using *e1070* (86) R package. We used *z* scores of *DESeq2* normalized counts obtained in this study as the train dataset and those obtained from ENCODE miRNA-seq data as the test dataset (Dataset S6). 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:

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

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. The reported accuracy is computed as a mean over the 10 testing sets in which nine 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.

Comparison with scATAC-Seq Data. To compute and plot the correlations of small RNA-seq with scATAC-seq (Fig. 6 *B–G*), we used Cicero "activity scores" reported in Cusanovich et al. (28). Cicero scores were computed as described in ref. 87. Briefly, Cicero activity score represents the summarized score of chromatin accessibility of all sites linked to a given gene, which include proximal sites to the gene's transcription start site (within 500 bp of an annotated TSS) or distal sites linked to them. Cicero scores were loaded in *Seurat v3*, normalized, scaled, and averaged per cell or tissue type. To compute the accessibility scores for the brain-specific ncRNA in Fig. 6*H*, we used Cicero to derive gene activity scores from scATAC-seq data generated by 10XGenomics for the mouse adult brain (https://www.10xgenomics.com/10x-university/single-cell-atac/) with the chromium single-cell ATAC platform, and demultiplexed and preprocessed with the single-cell ATAC Cell Ranger platform. Using *Seurat v3*, we clustered the cells and merged them with Allen Brain Atlas single-cell RNA-seq data (66) for the further transfer of cell annotation labels. We computed the activity scores for brain-specific ncRNA identified through small RNA-seq using *cicero* (87). We used Spearman correlation of top 400 tissue-specific genes to compute the relationship between small RNA-seq and ATAC-seq activity scores reported in Fig. 6*B*.

Data Availability. The datasets generated and analyzed in the study have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) repository (GSE119661) (88). All study data are included in the article and *SI Appendix*.

1. D. P. Bartel, Metazoan microRNAs. *Cell* **173**, 20–51 (2018).
2. T. R. Cech, J. A. Steitz, The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
3. M. Esteller, Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861–874 (2011).
4. J. Gebetsberger, L. Wyss, A. M. Mleczko, J. Reuther, N. Polacek, A tRNA-derived fragment competes with mRNA for ribosome binding and regulates translation during stress. *RNA Biol.* **14**, 1364–1373 (2017).
5. D. Becker et al., Nuclear pre-snRNA export is an essential quality assurance mechanism for functional spliceosomes. *Cell Rep.* **27**, 3199–3214.e3 (2019).
6. L. F. R. Gebert, I. J. MacRae, Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* **20**, 21–37 (2019).
7. A. G. Telonis et al., Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. *Oncotarget* **6**, 24797–24822 (2015).
8. S. L. Reichow, T. Hamma, A. R. Ferré-D'Amaré, G. Varani, The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35**, 1452–1464 (2007).
9. A. Keller et al., Toward the blood-borne miRNome of human diseases. *Nat. Methods* **8**, 841–843 (2011).

10. M. Somel et al., MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. Genome Res. 20, 1207–1218 (2010).
11. M. Ha, V. N. Kim, Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. 15, 509–524 (2014).
12. A. G. Matera, R. M. Terns, M. P. Terns, Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. 8, 209–220 (2007).
13. T. Kiss, Biogenesis of small nuclear RNPs. J. Cell Sci. 117, 5949–5951 (2004).
14. H. Ishizu, H. Siomi, M. C. Siomi, Biology of PIWI-interacting RNAs: New insights into biogenesis and function inside and outside of germlines. Genes Dev. 26, 2361–2373 (2012).
15. P. Kumar, J. Anaya, S. B. Mudunuri, A. Dutta, 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, 78 (2014).
16. P. Landgraf et al., A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129, 1401–1414 (2007).
17. O. R. Faridani et al., Single-cell sequencing of the small-RNA transcriptome. Nat. Biotechnol. 34, 1264–1266 (2016).
18. V. V. Sherstyuk et al., Genome-wide profiling and differential expression of microRNA in rat pluripotent stem cells. Sci. Rep. 7, 2787 (2017).
19. S. Anfossi, A. Babayan, K. Pantel, G. A. Calin, Clinical utility of circulating non-coding RNAs—an update. Nat. Rev. Clin. Oncol. 15, 541–563 (2018). .
20. F. J. Slack, A. M. Chinnaiyan, The role of non-coding RNAs in oncology. Cell 179, 1033–1055 (2019).
21. H. L. A. Janssen et al., Treatment of HCV infection by targeting microRNA. N. Engl. J. Med. 368, 1685–1694 (2013).
22. R. A. Ach, H. Wang, B. Curry, Measuring microRNAs: Comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol. 8, 69 (2008).
23. Y. Liang, D. Ridzon, L. Wong, C. Chen, Characterization of microRNA expression profiles in normal human tissues. BMC Genomics 8, 166 (2007).
24. M. N. McCall et al., Toward the human cellular microRNAome. Genome Res. 27, 1769–1781 (2017).
25. D. de Rie et al.; FANTOM Consortium, An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat. Biotechnol. 35, 872–878 (2017).
26. J. Jehn et al., 5′ tRNA halves are highly expressed in the primate hippocampus and sequence-specifically regulate gene expression. RNA 26, 694–707 (2019).
27. J. M. Rimer et al., Long-range function of secreted small nucleolar RNAs that direct 2′-O-methylation. J. Biol. Chem. 293, 13284–13296 (2018).
28. D. A. Cusanovich et al., A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 174, 1309–1324.e18 (2018).
29. A. Frankish et al., GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 47, D766–D773 (2019).
30. P. P. Chan, T. M. Lowe, GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. 44, D184–D189 (2016).
31. A. Kozomara, S. Griffiths-Jones, miRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–D73 (2014).
32. V. Boivin et al., Simultaneous sequencing of coding and noncoding RNA reveals a human transcriptome dominated by a small number of highly expressed noncoding genes. RNA 24, 950–965 (2018).
33. N. Ludwig et al., Distribution of miRNA expression across human tissues. Nucleic Acids Res. 44, 3865–3877 (2016).
34. R. C. Gallagher, B. Pils, M. Albalwi, U. Francke, Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader–Willi syndrome. Am. J. Hum. Genet. 71, 669–678 (2002).
35. R. A. Brady, V. M. Bruno, D. L. Burns, RNA-seq analysis of the host response to Staphylococcus aureus skin and soft tissue infection in a mouse model. PLoS One 10, e0124877 (2015).
36. P.-Y. Chen et al., FGF regulates TGF-β signaling and endothelial-to-mesenchymal transition via control of let-7 miRNA expression. Cell Rep. 2, 1684–1696 (2012).
37. A. Emde et al., Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: A new mechanism for ALS. EMBO J. 34, 2633–2651 (2015).
38. M. P. Kowalski, T. Krude, Functional roles of non-coding Y RNAs. Int. J. Biochem. Cell Biol. 66, 20–29 (2015).
39. S. J. Morrison, K. R. Prowse, P. Ho, I. L. Weissman, Telomerase activity in hematopoietic cells is associated with self-renewal potential. Immunity 5, 207–216 (1996).
40. H. Liu, Y. Yang, Y. Ge, J. Liu, Y. Zhao, TERC promotes cellular inflammatory response independent of telomerase. Nucleic Acids Res. 47, 8084–8095 (2019).
41. A. G. Matera, Z. Wang, A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 15, 108–121 (2014).
42. L. Pipes et al., The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. Nucleic Acids Res. 41, D906–D914 (2013).
43. L. Chen et al., miRNA arm switching identifies novel tumour biomarkers. EBioMedicine 38, 37–46 (2018).
44. K. Pinel, L. A. Diver, K. White, R. A. McDonald, A. H. Baker, Substantial dysregulation of miRNA passenger strands underlies the vascular response to injury. Cells 8, 83 (2019).
45. F. Kern et al., miRSwitch: Detecting microRNA arm shift and switch events. Nucleic Acids Res. 48, W268–W274 (2020).
46. J. Alles et al., An estimate of the total number of true human miRNAs. Nucleic Acids Res. 47, 3353–3364 (2019).
47. M. R. Friedländer, S. D. Mackowiak, N. Li, W. Chen, N. Rajewsky, miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52 (2012).
48. J. M. Dhahbi et al., Deep sequencing identifies circulating mouse miRNAs that are functionally implicated in manifestations of aging and responsive to calorie restriction. Aging (Albany NY) 5, 130–141 (2013).
49. T. Fehlmann et al., A high-resolution map of the human small non-coding transcriptome. Bioinformatics 34, 1621–1628 (2018).
50. R. Javed et al., miRNA transcriptome of hypertrophic skeletal muscle with overexpressed myostatin propeptide. BioMed Res. Int. 2014, 328935 (2014).
51. R. P. R. Metpally et al., Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. Front. Genet. 4, 20 (2013).
52. L. S. Sundaram, "Toxoplasma gondii-mediated host cell transcriptional changes lead to metabolic alterations akin to the Warburg effect," PhD thesis, University of Cambridge, Cambridge, UK (2017).
53. Y. S. Lee, Y. Shibata, A. Malhotra, A. Dutta, A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev. 23, 2639–2649 (2009).
54. S. P. Thomas, T. T. Hoang, V. T. Ressler, R. T. Raines, Human angiogenin is a potent cytotoxin in the absence of ribonuclease inhibitor. RNA 24, 1018–1027 (2018).
55. R. A. Forman et al., The goblet cell is the cellular source of the anti-microbial angiogenin 4 in the large intestine post Trichuris muris infection. PLoS One 7, e42248 (2012).
56. L. V. Hooper, T. S. Stappenbeck, C. V. Hong, J. I. Gordon, Angiogenins: A new class of microbicidal proteins involved in innate immunity. Nat. Immunol. 4, 269–273 (2003).
57. L. Guo, Q. Zhang, X. Ma, J. Wang, T. Liang, miRNA and mRNA expression analysis reveals potential sex-biased miRNA expression. Sci. Rep. 7, 39812 (2017).
58. P. Loher, E. R. Londin, I. Rigoutsos, IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. Oncotarget 5, 8790–8802 (2014).
59. C. M. Klinge, Estrogen regulation of microRNA expression. Curr. Genomics 10, 169–183 (2009).
60. C. E. Fletcher, D. A. Dart, C. L. Bevan, Interplay between steroid signalling and microRNAs: Implications for hormone-dependent cancers. Endocr. Relat. Cancer 21, R409–R429 (2014).
61. H. Guo, N. T. Ingolia, J. S. Weissman, D. P. Bartel, Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466, 835–840 (2010).
62. C. Cortes, V. Vapnik, Support-vector networks. Mach. Learn. 20, 273–297 (1995).
63. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
64. M. H. Baron, J. Isern, S. T. Fraser, The embryonic origins of erythropoiesis in mammals. Blood 119, 4828–4837 (2012).
65. U. Bissels, A. Bosio, W. Wagner, MicroRNAs are shaping the hematopoietic landscape. Haematologica 97, 160–167 (2012).
66. E. S. Lein et al., Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176 (2007).
67. M. Lin et al., Comprehensive identification of microRNA arm selection preference in lung cancer: miR-324-5p and -3p serve oncogenic functions in lung cancer. Oncol. Lett. 15, 9818–9826 (2018).
68. J. M. Dhahbi et al., 5′ tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. BMC Genomics 14, 298 (2013).
69. U. Sharma et al., Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science 351, 391–396 (2016).
70. C. Backes et al., Prioritizing and selecting likely novel miRNAs from NGS data. Nucleic Acids Res. 44, e53 (2016).
71. M. Sherafatian, Tree-based machine learning algorithms identified minimal set of miRNA biomarkers for breast cancer diagnosis and molecular subtyping. Gene 677, 111–118 (2018).
72. R. Bhome et al., Exosomal microRNAs (exomiRs): Small molecules with a big role in cancer. Cancer Lett. 420, 228–235 (2018).
73. T. T. M. Ngo et al., Noninvasive blood tests for fetal development predict gestational age and preterm delivery. Science 360, 1133–1136 (2018).
74. C. Trapnell, Defining cell types and states with single-cell genomics. Genome Res. 25, 1491–1498 (2015).
75. A. Dobin et al., STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
76. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
77. Y. Lu, A. S. Baras, M. K. Halushka, miRge 2.0 for comprehensive analysis of microRNA sequencing data. BMC Bioinformatics 19, 275 (2018).
78. D. Gebert, C. Hewel, D. Rosenkranz, unitas: The universal tool for annotation of small RNAs. BMC Genomics 18, 644 (2017).
79. M. E. Ritchie et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).
80. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
81. L. Pantano, DEGreport: Report of DEG analysis (R package, Version 1.18.1). lpantano. github.io/DEGreport/. Accessed 1 January 2020.
82. V. Agarwal, G. W. Bell, J.-W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. eLife 4, e05005 (2015).
83. M. D. Paraskevopoulou et al., DIANA-microT web server v5.0: Service integration into miRNA functional analysis workflows. Nucleic Acids Res. 41, W169–W173 (2013).
84. E. Pennisi, Genomics. ENCODE project writes eulogy for junk DNA. Science 337, 1159–1161 (2012).
85. C. Backes et al., miRCarta: A central repository for collecting miRNA candidates. Nucleic Acids Res. 46, D160–D167 (2018).
86. D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, e1071: Misc functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien (2017). https://cran.r-project.org/web/packages/e1071/e1071.pdf. Accessed 1 January 2020.
87. H. A. Pliner et al., Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol. Cell 71, 858–871.e8 (2018).
88. A. Isakova, S. Quake, A mouse tissue atlas of small noncoding RNA. GEO (Gene Expression Omnibus). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE119661. Deposited 7 September 2018.

GENETICS