

Variation of MicroRNA Expression in the Human Placenta Driven by Population Identity and Sex of the Newborn

Song Guo

Skolkovo Institute of Science and Technology

Shuyun Huang

Partner Institute for Computational Biology Chinese Academy of Sciences and Max Planck Society

Xi Jiang

Partner Institute for Computational Biology Chinese Academy of Sciences and Max Planck Society

Haiyang Hu

Partner Institute for Computational Biology Chinese Academy of Sciences and Max Planck Society

Dingding Han

Partner Institute for Computational Biology Chinese Academy of Sciences and Max Planck Society

Carlos S Moreno

Emory University

Genevieve L Fairbrother

Obstetrics and Gynecology of Atlanta

David A Hughes

University of Bristol

Mark Stoneking

Max-Planck-Institut für evolutionäre Anthropologie

Philipp Khaitovich (✉ khaitovich@eva.mpg.de)

Skolkovo Institute of Science and Technology

Research article

Keywords: Human, placenta, populations, sexual dimorphism, newborn, imprinting, miRNA

Posted Date: September 22nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-48907/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on April 20th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07542-0>.

Abstract

Background

Analysis of lymphocyte cell lines revealed substantial differences in the expression of mRNA and microRNA (miRNA) among human populations. The extent of such population-associated differences in actual human tissues remains largely unexplored. The placenta is one of the few solid human tissues that can be collected in substantial numbers in a controlled manner, enabling quantitative analysis of transient biomolecules such as RNA transcripts. Here, we analyzed microRNA (miRNA) expression in human placental samples derived from 36 individuals representing four genetically distinct human populations: African Americans, European-Americans, South Asians, and East Asians. All samples were collected at the same hospital following a unified protocol, thus minimizing potential biases that might influence the results.

Results

Sequence analysis of the miRNA fraction yielded 938 annotated and 70 novel miRNA transcripts expressed in the placenta. Of them, 82 (9%) of annotated and 11 (16%) of novel miRNAs displayed quantitative expression differences among populations, generally reflecting reported genetic and mRNA-expression-based distances. Several co-expressed miRNA clusters stood out from the rest of the population-associated differences in terms of miRNA evolutionary age, tissue-specificity, and disease-association characteristics. Among three non-environmental influenced demographic parameters, the second largest contributor to miRNA expression variation after population was the sex of the newborn, with 32 miRNAs (3% of detected) exhibiting significant expression differences depending on whether the newborn was male or female. Male-associated miRNAs were evolutionarily younger and correlated inversely with the expression of target mRNA involved in neuron-related functions. In contrast, both male and female-associated miRNAs appeared to mediate different types of hormonal responses. Demographic factors further affected reported imprinted expression of 66 placental miRNAs: the imprinting strength correlated with the mother's weight, but not height.

Conclusions

Our results showed that among 12 assessed demographic variables, population affiliation and fetal sex had a substantial influence on miRNA expression variation among human placental samples. The effect of newborn-sex-associated miRNA differences further led to expression inhibition of the target genes clustering in specific functional pathways. By contrast, population-driven miRNA differences might mainly represent neutral changes with minimal functional impacts.

Background

Phenotypic differences among humans can be attributed to the combined effect of genetic, epigenetic, and environmental factors. The genetic basis for phenotypic variation in human populations has been

extensively studied. Previous studies identified a number of genetic variants, including differences in single-nucleotide polymorphism (SNP) frequencies, copy number variation (CNV), transposable elements (TEs), and DNA methylation, that are associated with human population-specific phenotypic traits, including differential disease susceptibility [1–8].

In addition to genomic analyses, studies focusing on gene expression variation as a complex quantitative trait have played a fundamental role in advancing our understanding of the molecular mechanisms of evolution [9–11]. Most of our current knowledge about expression variation among human populations, however, comes from systematic investigations of transformed lymphoblastoid cell lines (LCLs) rather than native tissues [11–15]. Several such studies focusing on mRNA expression demonstrated that 4.5–29% of expressed genes were differentially expressed among human populations that included Europeans (CEU), Yoruba from sub-Saharan Africa (YRI), and two East Asian populations: Han Chinese (CHB) and Japanese (JPT) [12–14, 16, 17]. Parallel analysis of genetic differences explaining these expression differences identified a large number of *cis*-regulatory variants [12, 15, 16] and also *trans*-acting remote regulatory variants [14–16]. In most cases, these genetic variants might affect the binding of transcription factors (TFs) and hence alter the transcript isoform repertoire [18, 19].

MicroRNAs (miRNAs) also play a role in regulation of gene expression variation. miRNAs are short, 21–23 nucleotide-long hairpin-shaped RNA molecules that act as co-factors binding target sequences within mRNA transcripts, commonly in their 3' untranslated regions, through Watson-Crick complementarity interactions [20–22]. Simultaneously, miRNAs interact with parts of protein complexes, functioning as RNA endonucleases or as mRNA binding proteins that sequester target mRNA from the pool of actively translated transcripts [23]. Accordingly, miRNA expression levels inversely correlate with expression levels of their mRNA targets [24, 25]. Differences in miRNA expression among human populations were examined previously using LCLs derived from CEU and YRI individuals; this study revealed population-associated expression differences for 33 of the 757 detected miRNAs, resulting in downregulation of 55–88% of their expressed target genes [26]. Cancer studies investigating circulating miRNA abundance further indicated differences between individuals of African and non-African descent [27, 28].

However, gene expression variation among human populations measured in cell lines might not be indicative of the variation found in native tissues. Earlier, we reported mRNA expression differences at 6.3% of expressed genes among placental samples, all collected at the same location following the same protocol, from four populations: African-Americans, European-Americans, South Asians, and East Asians [29]. Here, we build upon this work by examining microRNA (miRNA) expression in these same placental samples and how it is influenced by 12 demographic variables for which we have sufficient information. We find that population identity and sex of the newborn contribute the most to miRNA expression variation.

Results

Placental miRNA expression measurements

We analyzed miRNA expression in placenta samples from individuals representing four major human ethnic groups (further referred to as populations): African Americans, European-Americans, South Asians, and East Asians (Fig. 1a). For each population, we sampled ten individuals (Table S1, Additional file 1) from previous study[29]. All samples were collected at the same geographic location (Northside Hospital in Atlanta, Georgia) from residents of the area. We sampled each placenta at five sites within the central *villous parenchyma* region and pooled the dissected samples before the mRNA and miRNA isolation [29]. In addition to population identity, for each sample we collected information for 26 demographic parameters from GSE66622. Among them, 12 parameters (listed in Methods), including delivery type (natural or cesarean), newborn infant's sex, number of previous births, mother's age, and mother's BMI, had sufficient variability to estimate their influence on miRNA expression levels.

We estimated miRNA expression levels using high-throughput transcriptome sequencing (RNA-seq) conducted on the Illumina sequencing platform. For each sample, we obtained an average of 31.4 million reads (Table S2, Additional file 2). Based on these data, we detected 938 miRNAs annotated in miRbase (v22) and 70 novel miRNAs (Fig. 1b; Table S3, Additional file 3; Table S4, Additional file 4) with the total expression count among the 40 individuals greater than 100 reads. Four individuals did not pass data quality criteria and were removed from further analyses (Fig. 1a; Fig S1c, Additional file 5).

Placental miRNA expression variation

Besides individual differences, the most notable contributors to the miRNA expression variation were population identity and sex of the newborn (SON), explaining 11% and 4% of the total variation, respectively (Fig. 1c). Accordingly, 139 miRNAs showed significant expression differences among populations, including 12 novel ones (ANOVA, nominal $p < 0.05$, permutation $p < 0.0001$; Fig. 1b,d,e), while 32 miRNAs, including one novel miRNA, differed depending on SON (ANOVA, nominal $p < 0.01$, permutation $p < 0.05$; Fig. 1f,g).

Population-associated placental miRNA

Further analysis of the 139 miRNAs showing population-associated expression yielded 93 miRNAs with significant expression differences between at least one pair of populations (Student's t-test, Benjamini-Hochberg corrected $p < 0.05$). Visualization of the distances among populations based on the expression of 93 or 139 population-associated miRNAs yielded dendrograms compatible with the genetic relationships among populations (Fig. 2a; Fig S2, Additional file 6 and Fig S3, Additional file 7). Specifically, miRNA expression in African Americans was the most distant from the other populations, while the two Asian populations were most similar to one another. Similarly, miRNA expression in African American population differed most from the other three based on analysis of 1,008 expressed miRNAs (Fig S2a, Additional file 6).

Unsupervised analysis of the 93 population-associated miRNA further identified six co-expressed miRNA clusters (Fig. 2b,c; Table S5, Additional file 8). Characterization of these clusters concerning miRNA evolutionary age, expression tissue-specificity, and disease associations further identified specific miRNA clusters showing significant feature enrichment (two-sided Wilcoxon test, nominal $p < 0.01$). Specifically,

cluster 1 (C1), characterized by elevated expression in European-American samples, contained significantly younger miRNAs than the bulk (two-sided Wilcoxon test, nominal $p < 0.01$) (Fig. 2c,d) and showed the highest miRNA expression tissue-specificity, restricted mainly to the placenta (Fig. 2c,e). Further, cluster 5 (C5), characterized by low expression in African-Americans and elevated expression in Asian populations (Fig. 2c), showed the highest number of miRNA disease associations (Fig. 2f; Fig S4, Additional file 9; Table S6, Additional file 10).

To assess the potential effects of population-associated miRNAs on expression of their target genes, we examined the published mRNA expression dataset derived from a partially overlapping set of placental samples [29] (GSE66622; Table S1, Additional file 1). Only cluster 1 (C1) reveal significant downregulation of predicted targets of population-associated miRNAs (one-side Wilcoxon rank test, $p < 0.05$, correlation $r < -0.5$). The potential targets of C1 miRNAs were enriched in the functional term associated with vasculogenesis and muscle organ development (Table S7, Additional file 11).

Sex-of-the-newborn-associated placental miRNA

Among 32 miRNAs showing expression differences depending on the sex of the newborn (SON-associated miRNA), 14 miRNAs were elevated in pregnancies with a male child (male-associated miRNA) and 18 in pregnancies with a female child (female-associated miRNA) (FDR $< 31\%$, permutation $p < 0.05$; Fig. 3a,b; Table S5, Additional file 8). All SON-associated miRNA expression differences were reproduced in multiple populations, with 24 of the 32 reproduced in all four (Exact binomial test, $p < 0.01$; Fig S5, Additional file 12). Notably, female-associated miRNAs were of significantly older evolutionary origin compared to most male-associated miRNAs (two-sided Wilcoxon test, nominal $p < 0.05$; Fig. 3c). Further, female-associated miRNAs were enriched in imprinted mir-379 cluster (C14MC) implicated in regulation of brain-specific functions [30] (hypergeometric test, Bonferroni corrected $p = 4.58 \times 10^{-6}$; Table S6, Additional file 10). Both female- and male-associated miRNA groups showed, however, the same moderate tissue-specificity (Fig. 3d).

To assess the potential effects of SON-associated miRNA expression, we identified their potential targets in the published mRNA expression dataset derived from a partially overlapping set of placental samples [29] (Table S1, Additional file 1; GSE66622). In total, we classified 46 mRNAs as potential targets of male-associated miRNAs and 65 mRNAs as potential targets for female-associated miRNAs, using a combination of miRNA target predictions and the inverse relationship of miRNA and target expression profiles as selection criteria. Notably, the potential targets of male-associated miRNAs were enriched in functional terms associated with glutamate receptor signaling and endocrine processes (Fig. 3e, Table S7, Additional file 11). By contrast, the potential targets of female-associated miRNAs were enriched in functions linked to steroid hormones, estradiol, and glucocorticoid response, as well as cell differentiation and metabolic processes (Fig. 3e, Table S7, Additional file 11).

Expression of imprinted miRNA

One of the characteristic features of placental miRNA is the prevalence of imprinted expression, a term referring to complete or partial suppression of one of the parental alleles [31]. To assess the extent of

miRNA expression imprinting in our data, we focused on the largest characterized imprinted miRNA cluster, located on chromosome 19 (C19MC) and expressed almost exclusively in the placenta [32]. This cluster contains 67 mature miRNAs, of which 66 were detected in our study (Table S5, Additional file 8). Expression analysis of these 66 miRNAs revealed a significant negative correlation with the mother's BMI (two-sided Wilcoxon test, $p = 2.8 \times 10^{-14}$) and weight ($p = 1.7 \times 10^{-10}$), but not height ($p = 0.27$) (Fig. 4a). This relationship was further apparent at the level of individual miRNAs (Spearman correlation, $p < 0.05$; Fig. 4b).

Discussion

The placenta plays an essential role in fetal development. Thus, understanding the role of factors determining miRNA expression variation in this tissue can shed light on the fundamental mechanisms of human developmental regulation and variability. Our study design helps to address this question by minimizing sampling effects on the results. The placentas were obtained from a single location, all processed according to the same protocol, and all collected at the same time point (birth). Sampling was further averaged in each individual by taking five independently dissected tissue fragments. For each sample, we recorded 26 demographic variables of mother and newborn infants, allowing us to assess their influence of placental miRNA expression variation.

Our results demonstrate that of three investigated non-environmental demographic variables, two substantially influence the expression of common posttranscriptional regulators, miRNAs, in the human placenta: population identity and sex of the newborn. Population has the most substantial influence explaining up to 11% of the total miRNA variance, and the relative miRNA expression divergence among four populations investigated in the study is consistent with their genetic divergence (Two-sided Mantel permutation test, Spearman's correlation coefficients $\rho = 0.771$, $p = 0.08$; Fig S3, Additional file 7) [4]. Since genetic divergence is largely thought to reflect the accumulation of phenotypically neutral mutations [33], it is therefore conceivable that miRNA variation among populations is similarly influenced by the phenotypically neutral changes. This notion aligns with previous work suggesting that mRNA expression divergence includes a substantial proportion of functionally and phenotypically neutral changes [29, 34].

Regulatory effects of population-associated miRNA expression differences estimated using mRNA expression data derived mainly from the same tissue revealed significant excess of expressional repression among predicted targets for one of the six miRNA clusters only. This result appears to contrast the reported widespread population-specific downregulation of miRNA targets described in cell lines [26]. While part of this discrepancy might be due to the limited statistical power of our study, the rest could be caused by unequal extent of the evolutionarily constraint in tissues and cell lines. As other regulators controlling multiple targets, miRNAs are under substantial evolutionary constraint [35, 36]. Assuming that most of randomly arising population-specific miRNA expression differences are non-adaptive, the ones with large regulatory effects are likely to be detrimental and will not be observed in a natural tissue, such

as placenta. The artificial growth conditions of the cell lines could, however, allow the manifestation of large-scale population-associated regulatory effects of miRNA variation.

Several technical factors might have further restricted our ability to detect miRNA-driven regulation of their predicted mRNA targets. Such factors include a mismatch between computational and experimentally verified miRNA target predictions, sequestering of target mRNA out of the translational pool without degradation, and the complex and often a tissue-specific interplay between miRNAs and other regulators [37, 38]. Biologically, our study includes a limited number of populations and biological replicates and certainly does not cover all population-associated aspects of miRNA regulatory effects. Evolutionarily, as mentioned above, the proportion of population-associated miRNA differences leading to functionally meaningful effects might be minor, analogous to genetic and mRNA divergence [4, 29]. It has to be noted, however, that despite these limitations, the fact that our study reveals many population-associated miRNA expression differences indicates the importance of further studies investigating the functional significance of this phenomenon.

Previous investigation of mRNA expression in human placenta reported 41 genes with sex-associated expression, 12 of them (30%) localized on sex chromosomes [39]. The substantial prevalence for sex chromosome localization was not, however, the case for SON-associated differences in miRNA expression: of 32 miRNAs, four (13%) localize on sex chromosomes. Sex-associated differences in miRNA expression were similarly reported in human tissues other than the placenta. Specifically, miRNA analysis across postnatal brain development revealed 40 miRNAs with significant sex-biased expression differences in the prefrontal cortex regions, 93% of them female-biased [40]. Further, investigation of four adult human tissues – brain, colorectal mucosa, peripheral blood, and cord blood – revealing 73 female-biased and 163 male-biased expressed miRNAs [41]. Notably, two of 32 SON-associated miRNAs overlapped with miRNAs showing corresponding sex-biased expression in the adult brain, and two overlapped with miRNAs showing such a bias in the peripheral blood. In addition to human studies, sex-biased miRNA expression was reported in mouse brain [42], mouse liver [43], rat liver [44], developing rat cortex [45], and other mammalian somatic tissues [46]. Previous studies singled out hormonal regulation as the main driving mechanism of miRNA sex-biased expression [43, 47]. In our study, functional analysis of target genes downregulated by SON-associated miRNAs in placenta similarly revealed terms related to hormonal processes, but also in other biological pathways.

In addition to the identification of population and SON effects, our data allowed us to examine a well-characterized phenomenon of imprinted miRNA expression in the human placenta [31]. Previously reported imprinted expression of the miRNA cluster located on chromosome 19 (C19MC) [31, 32] was also evident in our data. Previous work further linked the amplitude of the imprinting effect with the mother's BMI [48]. Our analysis of demographic variables indicated that the relationship between C19MC cluster imprinting and mother's BMI depends on the mother's weight but not the height.

Conclusions

Our results indicate that miRNA expression in the placenta varies substantially due to the population identity and the sex of the newborn. While the majority of population effects might reflect recent evolutionary drift caused by geographical separation, miRNA expression differences associated with given birth to a female child are evolutionarily older and display detectable regulatory effect on genes involved in particular functional processes. Further, the identification of regulatory effects induced by population-associated miRNA expression differences might differ between tissues and cell lines, highlighting the need for further studies of miRNA variation in a population framework.

Methods

Samples

Placental samples were obtained from previously study[29], including ten individuals from each ancestry (African-American, European-American, South Asian-American (India) and East Asian-American ancestry (Korea, China, Vietnam, and Taiwan)) from Northside Hospital in Atlanta, Georgia with the approval of the Northside Hospital Institutional Review Board (NSH #804), as described previously [29]. The sample information with 26 demographic variables was provided under GEO (GSE66622).

Construction of indexed small RNA-Seq Libraries

Total RNA was isolated from human placenta using the Trizol (Invitrogen, USA) according to the manufacturer's directions. RNA quality was determined using Agilent 6000 Nano chip on Agilent Technologies 2100 Bioanalyzer. Samples with RIN > 6 were selected for library construction. Sequencing libraries were constructed in a single batch according to the TruSeq SmallRNA Library Preparation guide (Illumina) with no modification. In each sequencing lane on an Illumina HiSeq X10 platform, we pooled 14 samples with different index sequences and carried out 150-bp paired-ended sequences. In total, 42 samples with two technical replicates were sequenced on three lanes.

Sequence mapping and novel miRNA identification

The adapter sequence at the 3'-end of each read was trimmed using cutadapt v1.13 [49] with parameters -m 17 -M 50 -a TGG AATTCTCGGGTGCCAAGG -A GATCGTCG GACTGTAGAACTCTGAAC. Novel miRNAs for each sample were detected using miRDeep2 algorithm with default parameters based on the human genome (hg38) and miRbase v22 as references [50]. Predicted rRNA/tRNA reads and reads with the miRDeep2 score ≤ 5 , representing the miRNA hairpin properties matching, were removed from the following analyses. Sequencing reads mapped to overlapping miRNA genomic location were merged across all samples. The detected miRNAs with no overlap with known mature miRNA genomic positions were considered as novel.

miRNA expression quantification

The trimmed raw sequences that were at least 17 bp long were mapped allowing no mismatches to the sequences of known and novel mature miRNAs determined as described in the previous section, extended

8nt both up and downstream, using the Bowtie algorithm [51]. To quantify the miRNA expression, only the reads from the R1 strand were considered. Following the protocol described in [52], the expression value of each miRNA was calculated as the sum of the read number mapped the reference mature miRNA sequence. All miRNA reads' count data were log₂ transformed and quantile normalized. As sequences were obtained in three batches, the batch effect was removed by removeBatchEffect (limma) algorithm (Fig S1, Additional file 5). The detailed read mapping information is listed in Table S2, Additional file 2. Based on a principle component analysis of all 42 samples, four samples were considered as outliers based on their dispersion and were removed from further analysis (Fig S1c, Additional file 5). In total, data from 8 African-American, 10 European-American, 8 South Asian-American and 10 East Asian-American individuals were retained.

Estimation of miRNA expression variation

Of the 26 demographic variables provided in previous study[29](GSE66622), 12 had at least five observations for each level. These 12 variables include: maternal body mass index, birth weight of the newborn child, birth length of the child, mother's age, sex of the newborn, birth delivery type (Cesarean or natural), population identity, first pregnancy or not, drinker or not, has pregnancy infection or not, number of pregnancy, and number of kid. For these variables estimated miRNA expression variation using analysis of variance (ANOVA; Table S8, Additional file 13). To assess the variables' effects more precisely, we recalculated the ANOVA-based variation estimates after subtraction of the potential effects of four continuous demographic variables using linear regression. The four continuous variables were: maternal body mass index, birth weight of the newborn child, birth length of the child, and mother's age. We applied an ANOVA test independently for each factor, including sex of the newborn, birth delivery type (Cesarean or natural), and population identity. The explained variance estimates agreed well between the ANOVA conducted for all 12 variables and for individual variables after the subtraction of continuous variable effects (Fig. 1c).

Identification of population-associated miRNA expression differences

To identify miRNAs differentially expressed between populations, we applied an ANOVA test with factors contributing the most to the general variation effects listed in the model in the following order: birth delivery type, sex of the newborn, and population. The expression data used in analysis was adjusted for the continuous demographic variables as described in the previous section. To estimate the significance of results, we applied same test 1,000 times to the dataset with randomly permuted individual labels. The permutation p-value was calculated as the proportion of times in which the number of significantly differentially-expressed miRNAs was greater than or equal to the observed difference in the data.

To further identify the miRNAs differentially expressed between any two human populations, we applied a two-sided pairwise t-test (pairwise.t.test in R multcomp package). Differentially-expressed miRNAs were defined as those with a $p < 0.05$ after Benjamini-Hochberg correction.

Clustering of population-associated placental miRNA

To detect co-expression patterns among 93 population-associated miRNAs, we applied hierarchical clustering method (hclust function in R) on z-transformed miRNA expression values with (1-rho) as the distance measure, where rho is the Spearman correlation coefficient. We chose k equal six based on visual inspecting of clusters obtained using different k values (Fig. 2b).

Construction of population dendrograms

To illustrate the extent of the expression difference among human populations, we built unrooted neighbor joining trees (nj function in R package ape) using the Euclidean distances between the mean expression of miRNA in each population. The expression level of each miRNA was scaled by the mean expression of this miRNA across all samples. The same approach was applied to human genetic data with distances based on the Fst values from 1,000 genome project [4]. Populations used for genetic data included African-Americans, CEPH, Telugu, Han Chinese, Southern Han Chinese, and Kinh Vietnamese to match the populations used in our study. The mean Fst value of the Han Chinese, Southern Han Chinese, and Kinh Vietnamese populations was considered as the value for East Asian population.

Quantification of miRNA expression tissue specificity

We downloaded 87 miRNA expression datasets derived from 12 healthy tissues from miRmine database [53]. Tissue specificity was measured by the tau index [54] using quantile normalized log2-RPM expression data. A tau value can range between 0 for house-keeping genes and 1 for tissue-specific genes.

Dating of human miRNA evolutionary ages

To estimate miRNA evolutionary age, we first downloaded all hairpin sequences for the 23 species listed in Table S9, Additional file 14 from miRBase v22. We then mapped the hairpin sequences to all of the 23 genomes using blastn with parameter -evalue 1e-5. A positive hit was called when the sequence overlapped with another species' known miRNA coordinates using bedtools intersect with default parameters. If a miRNA pair could be blasted reciprocally, it was considered to be a one-to-one miRNA hit. Next, according to Table S9, Additional file 14, we assigned each miRNA to an age group from 0–12, where 0 is the oldest (433 MYA to human) and 12 is the most recent evolutionary age group. We considered the evolutionary age of the pre-miRNA to correspond to the evolutionary age of the mature miRNA.

miRNA target prediction and functional analysis of miRNA targets

We downloaded miRNA predicted targets using miRnAtap [55], requiring a predicted target to be identified by at least three of the five following methods: "pictar", "diana", "targetscan", "miranda" and "mirdb". To get potential targets for each cluster of population-associated miRNAs, we assessed if predicted target genes for cluster of miRNAs were negatively regulated compared to non-target genes (Spearman correlation coefficient $\rho < -0.5$; one-sided Wilcoxon rank test $p < 0.05$). To identify potential targets of SON-associated miRNAs, we required the absolute expression fold change of miRNA targets to be greater than

0.1 and the direction of change to be opposite to the direction of miRNA expression difference. The normalized expression of miRNA targets was obtained from [29].

Gene ontology (GO) and pathway enrichment tests were processed with Metascape [56] with all genes expressed in human placenta used as a background. The GO terms with $-\log_{10}(p) > 2$ were reported as significantly enriched.

miRNA disease association analysis

We used the database TAM2.0 [57] to analyze the functional and disease associations of miRNAs with cancer-related terms masked. All expressed human placenta miRNAs were taken as a background. The terms with nominal $p < 0.05$ were reported as associated diseases.

Statistical analysis and software

All statistical analyses and plots were performed in the R environment (<http://www.r-project.org/>), using packages preprocessCore, multcomp, limma, dendextend, ggpubr, ggsci, ggplot2, gridExtra, reshape2, miRNAatap and cultevo. TAM2.0 was used for miRNA disease association and Metascape was used for the miRNA target enrichment test.

Abbreviations

miRNA: microRNA

BMI: Body mass index

SON: Sex of newborn

RPM: Reads per million

RNA-seq: RNA sequencing

ANOVA: analysis of variance.

MYA: Million years ago

C19MC: Chromosome 19 miRNA cluster

Declarations

Ethics approval and consent to participate

All placentas were collected in October and November of 2006 at Northside Hospital in Atlanta, Georgia, with the approval of the Northside Hospital Institutional Review Board (NSH #804), with the written informed consent of the donors and in accordance with the Declaration of Helsinki agreement.

Consent for publication

Not applicable. No individual personal data are included in this study. Data analyses are presented in aggregate using anonymized miRNA data, with no means of tracing the results and the data back to the individual.

Availability of data and materials

The raw datasets generated during the current study are available in the Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>) with accession number PRJNA602793. It will be released upon publication. Now the link is available in read-only format (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA602793?reviewer=mjtfukonhgrfi2ftgoesd3g4f0>). Detailed sample information can be found both in the PRJNA602793 and in the Gene Expression Omnibus (GEO) under accession number GSE66622 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66622>). The quantified miRNA counts, and normalized expression data can be found under supplementary files (Table S3, Additional file 3; Table S4, Additional file 4). The quantified gene expression data from the previous study[29] were downloaded from GEO under accession number GSE66622 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66622>).

Competing interests

The authors declare no conflict of interest

Funding

This work was financially supported by the National Natural Science Foundation of China (Grants 91331203 to P.K.) and the National One Thousand Foreign Experts Plan (Grant WQ20123100078 to P.K.), D.H. and M.S. are grateful for the support of the Max Planck Society.

Authors' contributions

D.A.H., M.S., and P.K. designed the study. D.A.H., C.M., G.F. collected the samples. S.G. designed and executed the bioinformatics analysis. HY.H provided scripts on miRNA expression. SY.H., X.J. and DD.H constructed the libraries. S.G. wrote the manuscript with assistance from D.A.H., P.K, and M.S. and input from all authors. P.K. and M.S. supervised the project. All authors have read and approved the manuscript.

Acknowledgments

We would like to thank all donors for their participation, and the nursing staff at Northside Hospital in Atlanta, Georgia for their assistance in the sample collection.

References

1. de Magalhaes JP, Matsuda A: **Genome-wide patterns of genetic distances reveal candidate Loci contributing to human population-specific traits.** *Ann Hum Genet* 2012, **76**(2):142-158.
2. Armengol L, Villatoro S, Gonzalez JR, Pantano L, Garcia-Aragones M, Rabionet R, Caceres M, Estivill X: **Identification of copy number variants defining genomic differences among major human groups.** *PLoS One* 2009, **4**(9):e7230.
3. McCarroll SA, Altshuler DM: **Copy-number variation and association studies of human disease.** *Nat Genet* 2007, **39**(7 Suppl):S37-42.
4. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al*: **A global reference for human genetic variation.** *Nature* 2015, **526**(7571):68-74.
5. Guo J, Wu Y, Zhu Z, Zheng Z, Trzaskowski M, Zeng J, Robinson MR, Visscher PM, Yang J: **Global genetic differentiation of complex traits shaped by natural selection in humans.** *Nat Commun* 2018, **9**(1):1865.
6. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C *et al*: **Relative impact of nucleotide and copy number variation on gene expression phenotypes.** *Science* 2007, **315**(5813):848-853.
7. Wang L, Rishishwar L, Marino-Ramirez L, Jordan IK: **Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements.** *Nucleic Acids Res* 2017, **45**(5):2318-2328.
8. Husquin LT, Rotival M, Fagny M, Quach H, Zidane N, McEwen LM, Maclsaac JL, Kobor MS, Aschard H, Patin E *et al*: **Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation.** *Genome Biol* 2018, **19**(1):222.
9. Oleksiak MF, Churchill GA, Crawford DL: **Variation in gene expression within and among natural populations.** *Nature Genetics* 2002, **32**(2):261-266.
10. Whitehead A, Crawford DL: **Variation within and among species in gene expression: raw material for evolution.** *Mol Ecol* 2006, **15**(5):1197-1211.
11. Daca-Roszak P, Swierniak M, Jaksik R, Tyszkiewicz T, Oczko-Wojciechowska M, Zebracka-Gala J, Jarzab B, Witt M, Zietkiewicz E: **Transcriptomic population markers for human population discrimination.** *BMC Genet* 2018, **19**(1):54.
12. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG: **Common genetic variants account for differences in gene expression among ethnic groups.** *Nature Genetics* 2007, **39**(2):226-231.

13. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM: **Gene-expression variation within and among human populations.** *Am J Hum Genet* 2007, **80**(3):502-509.
14. Zhang W, Duan S, Kistner EO, Bleibel WK, Huang RS, Clark TA, Chen TX, Schweitzer AC, Bieme JE, Cox NJ *et al*: **Evaluation of genetic variation contributing to differences in gene expression between populations.** *Am J Hum Genet* 2008, **82**(3):631-640.
15. Price AL, Patterson N, Hancks DC, Myers S, Reich D, Cheung VG, Spielman RS: **Effects of cis and trans genetic ancestry on gene expression in African Americans.** *PLoS Genet* 2008, **4**(12):e1000294.
16. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D *et al*: **Population genomics of human gene expression.** *Nat Genet* 2007, **39**(10):1217-1224.
17. Li J, Liu Y, Kim T, Min R, Zhang Z: **Gene expression variability within and between human populations and implications toward disease susceptibility.** *PLoS Comput Biol* 2010, **6**(8).
18. Zhang W, Duan S, Bleibel WK, Wisel SA, Huang RS, Wu X, He L, Clark TA, Chen TX, Schweitzer AC *et al*: **Identification of common genetic variants that account for transcript isoform variation between human populations.** *Hum Genet* 2009, **125**(1):81-93.
19. Martin AR, Costa HA, Lappalainen T, Henn BM, Kidd JM, Yee MC, Grubert F, Cann HM, Snyder M, Montgomery SB *et al*: **Transcriptome sequencing from diverse human populations reveals differentiated regulatory architecture.** *PLoS Genet* 2014, **10**(8):e1004549.
20. Ambros V: **The functions of animal microRNAs.** *Nature* 2004, **431**(7006):350-355.
21. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116**(2):281-297.
22. Bartel DP: **Metazoan MicroRNAs.** *Cell* 2018, **173**(1):20-51.
23. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD: **Asymmetry in the assembly of the RNAi enzyme complex.** *Cell* 2003, **115**(2):199-208.
24. Bartel DP: **MicroRNAs: target recognition and regulatory functions.** *Cell* 2009, **136**(2):215-233.
25. Fabian MR, Sonenberg N, Filipowicz W: **Regulation of mRNA translation and stability by microRNAs.** *Annu Rev Biochem* 2010, **79**:351-379.
26. Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, Wing C, Duan S, Bleibel WK, Cox NJ *et al*: **Population differences in microRNA expression and biological implications.** *RNA Biol* 2011, **8**(4):692-701.
27. Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S: **A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer.** *PLoS One* 2010, **5**(10):e13735.
28. Rawlings-Goss RA, Campbell MC, Tishkoff SA: **Global population-specific variation in miRNA associated with cancer risk and clinical biomarkers.** *BMC Med Genomics* 2014, **7**:53.
29. Hughes DA, Kircher M, He Z, Guo S, Fairbrother GL, Moreno CS, Khaitovich P, Stoneking M: **Evaluating intra- and inter-individual variation in the human placental transcriptome.** *Genome Biol* 2015, **16**:54.
30. Winter J: **MicroRNAs of the miR379-410 cluster: New players in embryonic neurogenesis and regulators of neuronal function.** *Neurogenesis (Austin)* 2015, **2**(1):e1004970.

31. Malnou EC, Umlauf D, Mouysset M, Cavaille J: **Imprinted MicroRNA Gene Clusters in the Evolution, Development, and Functions of Mammalian Placenta.** *Front Genet* 2018, **9**:706.
32. Noguer-Dance M, Abu-Amero S, Al-Khtib M, Lefevre A, Coullin P, Moore GE, Cavaille J: **The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta.** *Hum Mol Genet* 2010, **19**(18):3566-3582.
33. Kimura M: **The Neutral Theory of Molecular Evolution.** *Cambridge:Cambridge University Press* 1983.
34. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, Wirkner U, Ansorge W, Paabo S: **A neutral model of transcriptome evolution.** *PLoS Biol* 2004, **2**(5):E132.
35. Xu J, Zhang R, Shen Y, Liu G, Lu X, Wu CI: **The evolution of evolvability in microRNA target sites in vertebrates.** *Genome Res* 2013, **23**(11):1810-1816.
36. Wu CI, Shen Y, Tang T: **Evolution under canalization and the dual roles of microRNAs: a hypothesis.** *Genome Res* 2009, **19**(5):734-743.
37. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ *et al*: **miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database.** *Nucleic Acids Res* 2016, **44**(D1):D239-247.
38. Guo Z, Maki M, Ding R, Yang Y, Zhang B, Xiong L: **Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues.** *Sci Rep* 2014, **4**:5150.
39. Sood R, Zehnder JL, Druzin ML, Brown PO: **Gene expression patterns in human placenta.** *Proc Natl Acad Sci U S A* 2006, **103**(14):5478-5483.
40. Ziats MN, Rennert OM: **Identification of differentially expressed microRNAs across the developing human brain.** *Mol Psychiatry* 2014, **19**(7):848-852.
41. Cui C, Yang W, Shi J, Zhou Y, Yang J, Cui Q, Zhou Y: **Identification and Analysis of Human Sex-biased MicroRNAs.** *Genomics Proteomics Bioinformatics* 2018, **16**(3):200-211.
42. Koturbash I, Zemp F, Kolb B, Kovalchuk O: **Sex-specific radiation-induced microRNAome responses in the hippocampus, cerebellum and frontal cortex in a mouse model.** *Mutat Res* 2011, **722**(2):114-118.
43. Hao P, Waxman DJ: **Functional Roles of Sex-Biased, Growth Hormone-Regulated MicroRNAs miR-1948 and miR-802 in Young Adult Mouse Liver.** *Endocrinology* 2018, **159**(3):1377-1392.
44. Kwekel JC, Vijay V, Han T, Moland CL, Desai VG, Fuscoe JC: **Sex and age differences in the expression of liver microRNAs during the life span of F344 rats.** *Biol Sex Differ* 2017, **8**:6.
45. Murphy SJ, Lusardi TA, Phillips JI, Saugstad JA: **Sex differences in microRNA expression during development in rat cortex.** *Neurochem Int* 2014, **77**:24-32.
46. Warnefors M, Mossinger K, Halbert J, Studer T, VandeBerg JL, Lindgren I, Fallahshahrudi A, Jensen P, Kaessmann H: **Sex-biased microRNA expression in mammals and birds reveals underlying regulatory mechanisms and a role in dosage compensation.** *Genome Res* 2017, **27**(12):1961-1973.
47. Dai R, Ahmed SA: **Sexual dimorphism of miRNA expression: a new perspective in understanding the sex bias of autoimmune diseases.** *Ther Clin Risk Manag* 2014, **10**:151-163.

48. Prats-Puig A, Xargay-Torrent S, Carreras-Badosa G, Mas-Pares B, Bassols J, Petry CJ, Girardot M, F DEZ, Ibanez L, Dunger DB *et al*: **Methylation of the C19MC microRNA locus in the placenta: association with maternal and childhood body size.** *Int J Obes (Lond)* 2020, **44**(1):13-22.
49. Martin M: **Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads.** *EMBnet J* 2011, **2011**;17:10.
50. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N: **miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades.** *Nucleic Acids Res* 2012, **40**(1):37-52.
51. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**(3):R25.
52. Hu HY, Yan Z, Xu Y, Hu H, Menzel C, Zhou YH, Chen W, Khaitovich P: **Sequence features associated with microRNA strand selection in humans and flies.** *BMC Genomics* 2009, **10**:413.
53. Panwar B, Omenn GS, Guan YF: **miRmine: a database of human miRNA expression profiles.** *Bioinformatics* 2017, **33**(10):1554-1560.
54. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E *et al*: **Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification.** *Bioinformatics* 2005, **21**(5):650-659.
55. Simpson MPaTI: **miRNAtap: miRNAtap: microRNA Targets - Aggregated Predictions.** 2019.
56. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.** *Nat Commun* 2019, **10**(1):1523.
57. Li J, Han X, Wan Y, Zhang S, Zhao Y, Fan R, Cui Q, Zhou Y: **TAM 2.0: tool for MicroRNA set analysis.** *Nucleic Acids Res* 2018, **46**(W1):W180-W185.

Figures

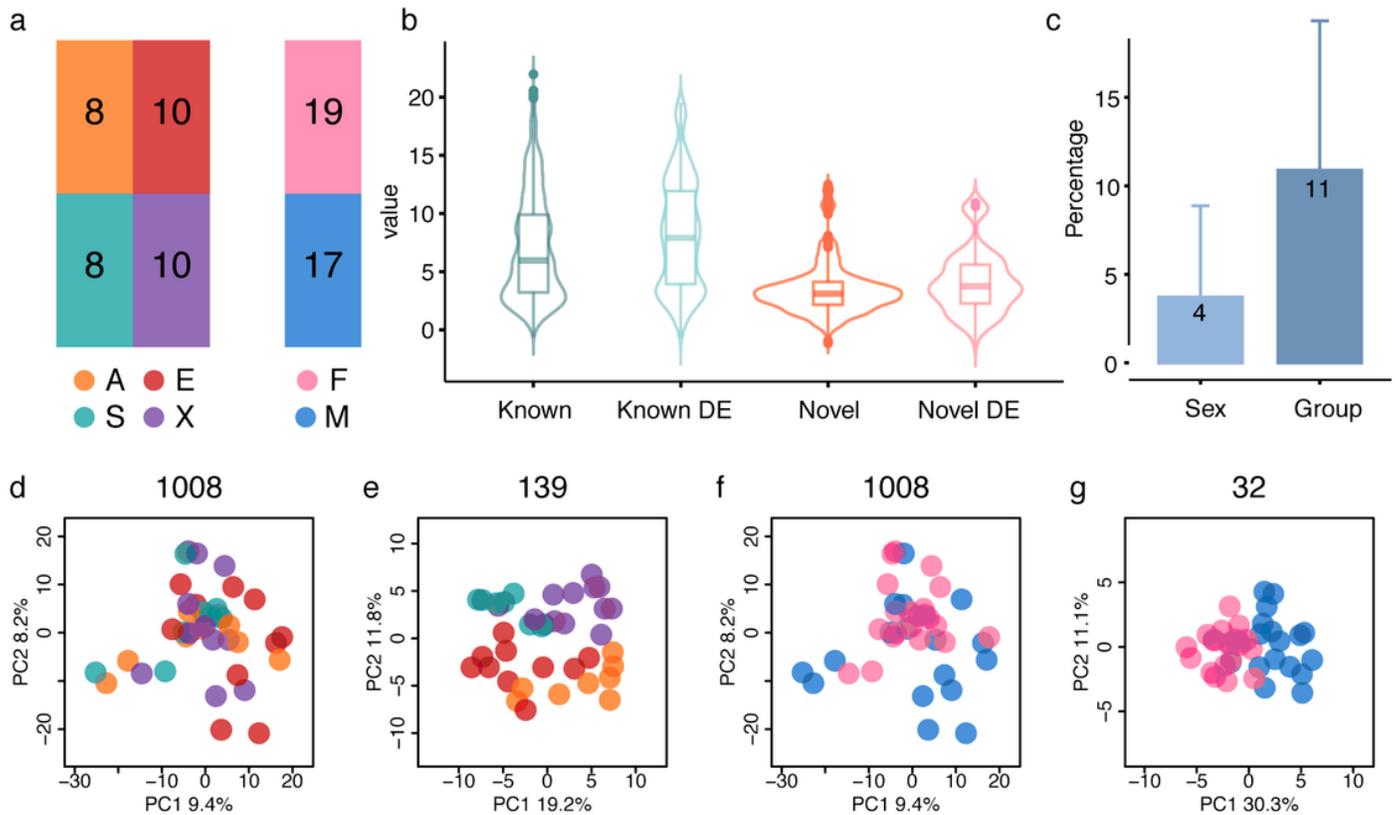


Figure 1

Sample information and miRNA expression distribution. a Schematic illustration of sample numbers according to population ID and newborn sex. The abbreviations here and in the text indicate: A – African Americans; E – European-Americans; S – South Asians; X – East Asians; F - female newborn and M - male newborn. b Violin plot showing miRNA expression distribution. Y-axis shows the quantile normalized log₂-transformed miRNA read count values after removing the batch effect. X-axis labels indicate 938 annotated miRNAs (Known), 127 annotated miRNAs differentially expressed among populations (Known DE), 70 novel miRNAs (Novel), and 12 novel miRNAs differentially expressed among populations (Novel DE). c Percentage of total expression variance explained by newborn sex and population. Bars represents the mean of the variance proportion explained by the factors. Error bars represent the standard deviation of the mean. d-g Principal component analysis plots based on the miRNA expression of all 1,008 miRNAs (d, colored according to population and f, colored according to newborn sex), 139 miRNAs differentially expressed among populations (e), and 32 miRNAs differentially expressed depending on the sex of the newborn (g).

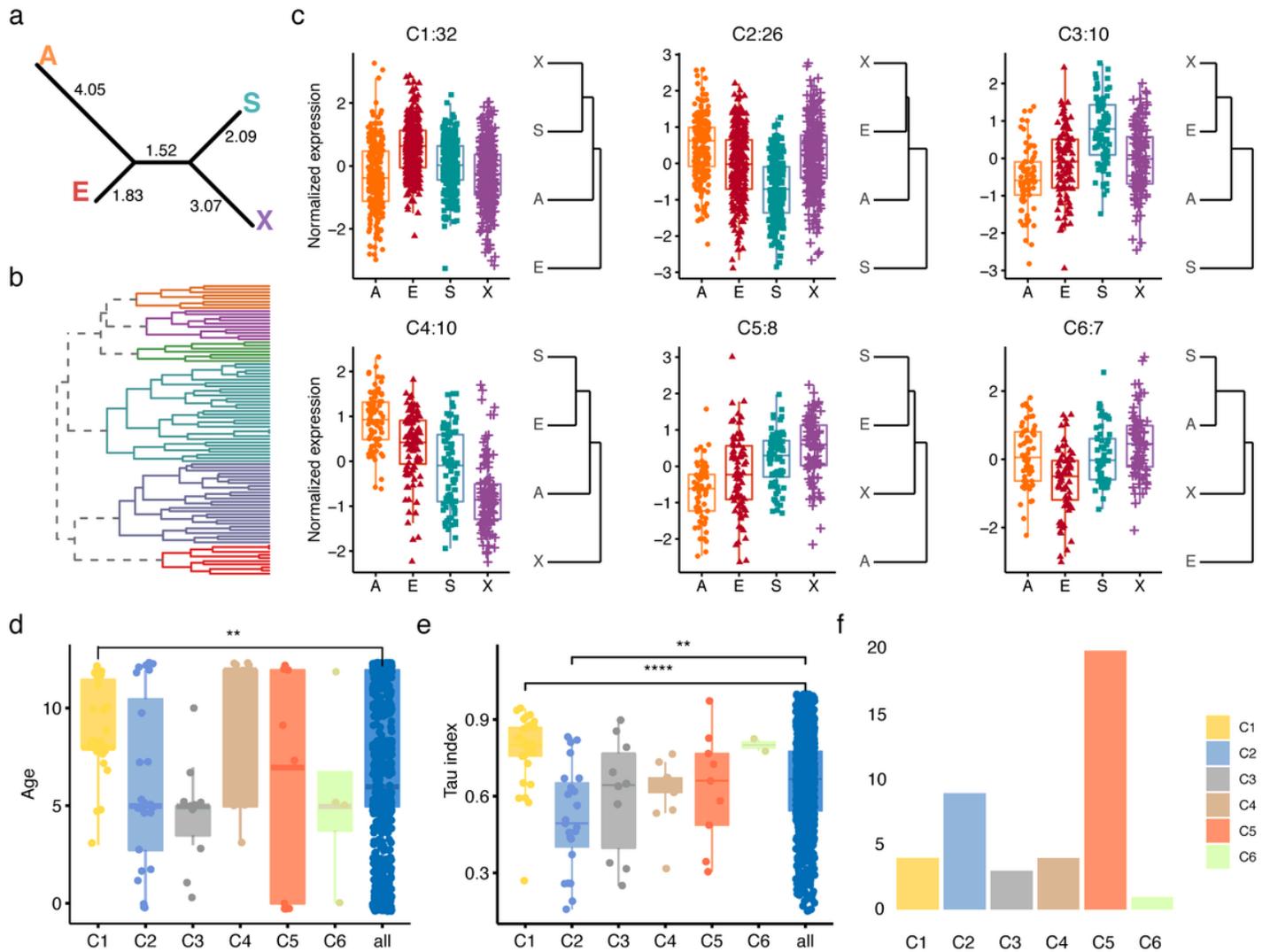


Figure 2

Characterization of miRNAs differentially expressed among human populations. a Dendrogram based on expression levels of 93 population-associated miRNAs. The abbreviations here and in the text indicate: A – African-Americans; E – European-Americans; S – South Asians; X – East Asians. Numbers indicate the branch length. b Hierarchical clustering of 93 population-associated miRNAs based on correlation of their expression profiles. Colors represent six main clusters. c miRNA expression patterns in each of the six clusters. Colors represent populations. Panel titles show the cluster name and the number of miRNAs in the cluster. Y-axis indicates Z-transformed miRNA expression values. The dendrograms on the right of each panel represents the average normalized expression distances among populations based on the expression of cluster miRNAs. d Distribution of miRNA evolutionary age in the six clusters. The age scale extends from 433 Mya (age 0) to human-specific miRNA (age 12). Asterisks indicate the significance of the difference (two-sided Wilcoxon test, ** represents nominal $p < 0.01$). e Distribution of miRNA tissue expression index (Tau) in the six clusters. Large values represent greater expression tissue-specificity.

Asterisks indicate the significance of the difference (two-sided Wilcoxon test, **** represents nominal $p < 0.0001$). f Number of miRNAs associated with disease in each cluster.

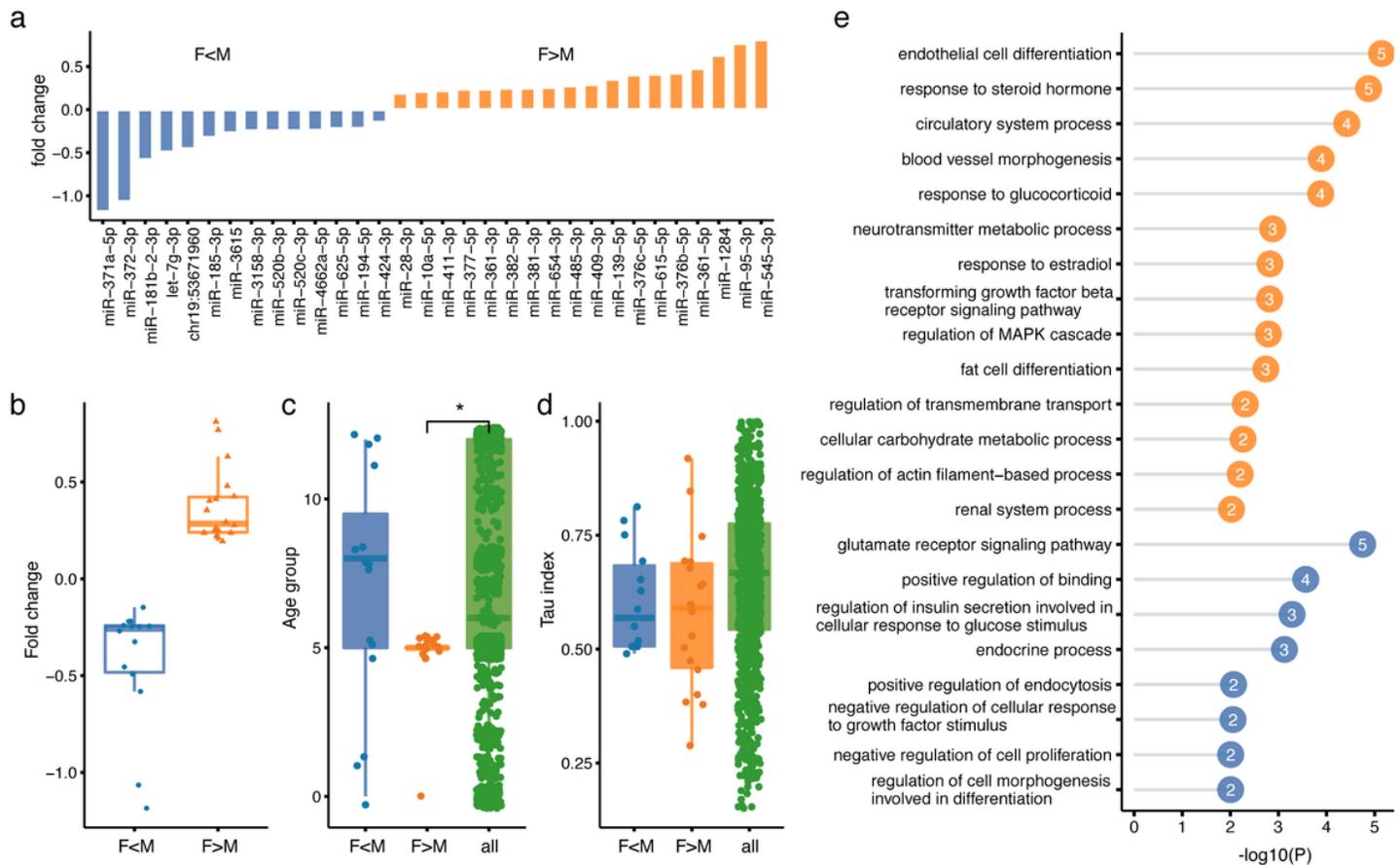


Figure 3

Characterization of miRNAs with newborn sex-associated expression. **a** Bar plot showing individual miRNA expression differences between placental samples from female vs. male newborn. Colors represent male-newborn-associated (F<M, blue) and female-newborn-associated (F>M, orange) miRNAs. Abbreviations: F – female newborn; M – male newborn. **b** Boxplot showing the distributions of miRNA expression fold-change for placental samples from female vs. male newborn infants. The blue and yellow boxes represent miRNAs with male-newborn-associated and female-newborn-associated expression. Each dot represents one miRNA. **c** Distribution of miRNA evolutionary age for male-newborn-associated (blue) and female-newborn-associated (orange) miRNA. The age scale extends from 433 Mya (age 0) to human-specific miRNA (age 12). Asterisks indicate the significance of the difference (two-sided Wilcoxon test, * represents nominal $p < 0.05$). **d** Distribution of miRNA tissue expression index (Tau) for male-newborn-associated (blue) and female-newborn-associated (orange) miRNA. Large values represent greater expression tissue-specificity. **e** GO terms enriched in targets of male-newborn-associated (blue) and female-newborn-associated (orange) miRNAs. X-axis and the number within circles indicate $-\log_{10}$ -transformed p-values.

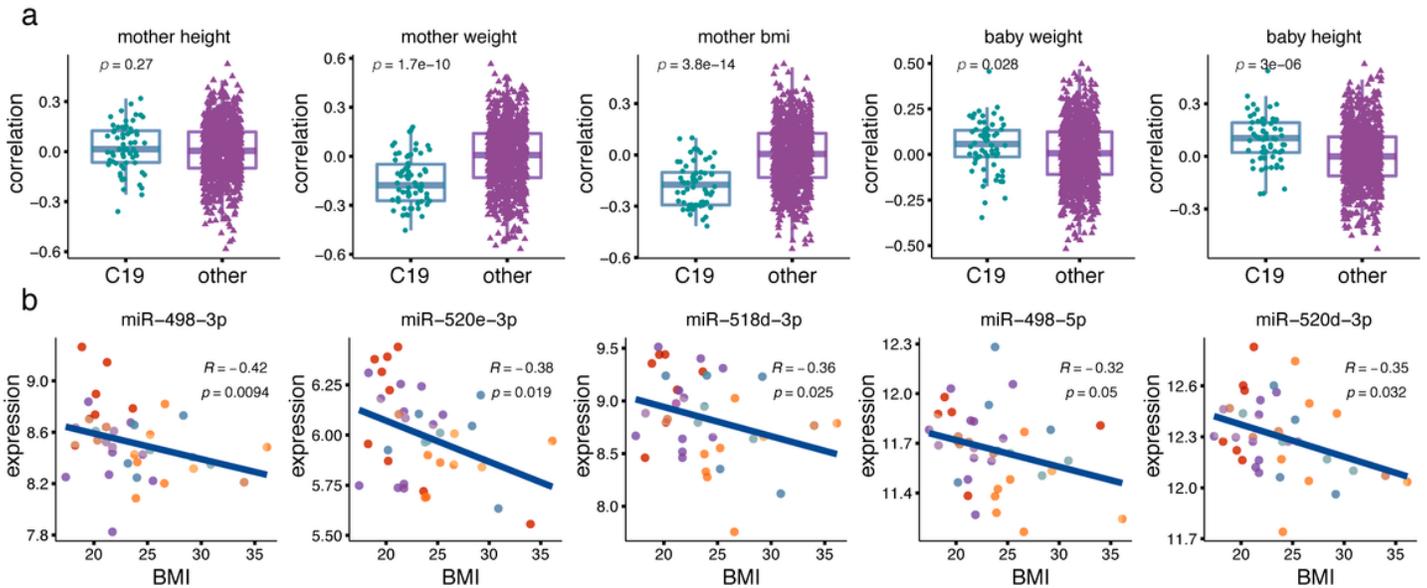


Figure 4

Expression of imprinted miRNAs in the C19MC cluster. a Correlation distribution between imprinted miRNAs located in the C19MC cluster and demographic variables. Panel titles indicate the demographic variable used in the comparison. P-values for a two-sided Wilcoxon test are shown within panels. b Five miRNAs showing a significant expression correlation with mother's BMI. Each dot represents the expression level in a sample. Colors represent human populations as illustrated in Fig 1a. Shaded areas represent confidence intervals. Spearman's correlation coefficients rho (R) and p-values (p) are displayed in the top right corner of each scatter plot.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1TableS1.xlsx](#)
- [Additionalfile2TableS2.xlsx](#)
- [Additionalfile3TableS3.xlsx](#)
- [Additionalfile4TableS4.xlsx](#)
- [Additionalfile5FigureS1.tif](#)
- [Additionalfile6FigureS2.tif](#)
- [Additionalfile7FigureS3.tif](#)
- [Additionalfile14TableS9.xlsx](#)
- [Additionalfile8TableS5.xlsx](#)
- [Additionalfile13TableS8.xlsx](#)

- [Additionalfile9FigureS4.tif](#)
- [Additionalfile12FigureS5.tif](#)
- [Additionalfile10TableS6.xlsx](#)
- [Additionalfile11TableS7.xlsx](#)