Supplementary Information

A-to-I editing of miR-200b-3p in airway cells is associated with moderate-to-severe asthma

Kevin M. Magnaye[1,*], Katherine A. Naughton[1], Janel Huffman[1], D. Kyle Hogarth[2], Edward T. Naureckas[2], Steven R. White[2], Carole Ober[1,*]

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637
[2]Department of Medicine, University of Chicago, Chicago, IL 60637

*Corresponding Authors:
Department of Human Genetics
University of Chicago
Chicago, IL 60615
TEL: 773-834-0734
FAX: 773-834-0505
Email: c-ober@genetics.uchicago.edu
kmagnaye@uchicago.edu

Methods

Genotyping and QC

SNPs were genotyped using the Illumina Omni2.5-8v1A, Omni1MDuo, or Human Core arrays. SNPs on each array were excluded with HWE < 0.0001 within each ethnic group (European American, African American), MAF < 0.05, SNP call rate < 0.95, and subject call rates < 0.95. Ancestry principal component analysis (PCA) was performed using 676 ancestral informative markers included on the arrays that overlap with the HapMap release 3.

Small RNA and total RNA extraction and sequencing

RNAs were extracted from bronchial epithelial epithelial cells and purified using the QIAGEN AllPrep DNA/RNA/miRNA Universal Kit (Hilden, Germany), following manufacterer's instructions. Quality and quantity assessment of small RNA and total RNA were measured at the University of Chicago Functional Genomics Core using an Agilent RNA 6000 Pico assay and the Agilent 2100 Bioanalyzer. Small RNA-seq libraries were prepared using Illumina small RNA-SEQ library kits. cDNA libraries were constructed using the Illumina TruSeq RNA Library Prep Kit v2. Small RNA and RNA sequencing were performed at the University of Chicago Genomics Core on either the Illumina HiSeq 2500 or 4000 platforms.

Quality control of small and total RNA-seq

For the small and total RNA-seq data, potential sample contamination and sample swaps were assessed using VerifyBamID (1). No cross-contamination was detected. For the RNA-seq data, two sample swaps between individuals were identified and corrected. Quality control checks were performed on the small RNA and total RNA-seq data using FastQC (2). Small RNA and total RNA sequences were then aligned and annotated to known miRNA and RNA sequences using miRge2.0 (3) and Spliced Transcripts Alignment to a Reference using STAR (4). miRNAs and genes with low count data (<1 CPM in at least 25% of the sample) and those on the X, Y, and mitochondrial chromosomes were removed. Samples containing > 2M mapped miRNA reads or > 8M mapped total RNA reads were kept, retaining 138 and 124 subjects for downstream analyses.

Raw miRNA and mRNA counts were normalized using quantile normalization and the trimmed mean of M-values (TMM) methods, respectively. Mean-variance trend of both datasets was adjusted using variance modeling in voom (5)). Technical sources of variation were identified using principal components analysis (PCA) for the global miRNA and total RNA expression profiles. For differential expression, surrogate variable analysis (6) was performed to predict latent covariates to include in the model. All covariates are described in the Methods of the main paper.

Detection of A-to-I edited sites

The percentages of A-to-I editing were calculated for mapped miRNAs, using the A-to-I analysis module in miRge2.0 (3). Briefly, putative A-to-I edited signals were removed if: 1) they were located at miRNA SNPs with A/G differences; 2) they were in the 455 miRNAs located in repeat

elements; 3) they were in a lowly expressed miRNA (RPM < 1); or 4) their edited sequences aligned to ≥1+ locations in the genome. For the retained reads that belonged to the canonical and edited miRNAs, all nucleotide positions in the canonical miRNA, except the terminal 5 bp were screened for A to G changes based on a binomial test considering the expected sequencing error rate (0.1%), as described (7). The A-to-I editing frequency (% A-to-I) was defined as the proportion of the mapped reads containing the edited nucleotide (G) relative to the total mapped reads at the given location.

Target gene prediction and pathway analysis

Bioinformatic prediction of target genes was performed for the unedited and edited miR-200b-3p using TargetScanHuman5.2 and TargetScanHuman5.2 Custom, respectively (8). For the unedited miR-200b-3p, *in silico* target genes were identified using a search for the unedited "miR-200b-3p". For the edited miR-200b-3p, *in silico* target genes were identified using a search for the seed sequence at positions 2-8 including the edited base, "AAT**G**CTG". Pathway analysis was performed using the TopFunn function within the ToppGene Suite for gene list enrichment analysis (9). Pathway enrichment analyses used a hypergeometric distribution with Bonferroni correction (p-value < 0.05). Venn diagrams of the genes and pathways of the unedited and edited forms of miR-200b-3p were generated using the VennDiagram package in R. Because TopFunn only reports the top 50 pathway enrichments, the unique pathways among the top 50 pathway enrichments of the unedited miR-200b-3p were used for comparison.

Tables

Table S1. Nineteen A-to-I edited sites in bronchial epithelial cells from 138 subjects. The edited adenosine within the mature miRNA sequence is in bold.

miRNA	Position	Sequence (seed sequence)	Average Mapped Counts	Average % A-to-I Reads	# Subjects with Edited miRNA
miR-200b-3p	5	TAAT**A**CTGCCTGGTAATGATGA	75358	0.52	135
miR-191-5p	3	CA**A**CGGAATCCCAAAAGCAGCTG	58512	0.17	27
miR-200b-3p	3	TA**A**TACTGCCTGGTAATGATGA	91380	0.15	22
miR-186-5p	3	CA**A**AGAATTCTCCTTTTGGGCT	24646	0.18	13
miR-148a-3p	3	TC**A**GTGCACTACAGAACTTTGT	154711	0.14	6
miR-411-5p	5	TAGT**A**GACCGTATAGCGTACG	372	5.40	5
miR-497-5p	2	C**A**GCAGCACACTGTGGTTTGT	355	5.21	5
miR-421	14	ATCAACAGACATT**A**ATTGGGCGC	663	0.89	5
miR-21-3p	3	CA**A**CACCAGTCGATGGGCTGT	22650	0.21	4
miR-376c-3p	6	AACAT**A**GAGGAAATTCCACGT	27	34.09	3
miR-381-3p	4	TAT**A**CAAGGGCAAGCTCTCTGT	254	3.09	3
miR-200a-3p	3	TA**A**CACTGTCTGGTAACGATGT	24079	0.19	2
miR-339-3p	15	TGAGCGCCTCGACG**A**CAGAGCCG	822	9.62	2
miR-451a	3	AA**A**CCGTTACCATTACTGAGTT	36137	0.16	2
miR-191-5p	8	CAACGGA**A**TCCCAAAAGCAGCTG	56489	0.15	1
miR-20a-5p	4	TAA**A**GTGCTTATAGTGCAGGTAG	81	3.7	1
miR-561-5p	13	ATCAAGGATCTT**A**AACTTTGCC	225	2.22	1
miR-589-3p	6	TCAGA**A**CAAATGCCGGTTCCCAGA	14	21.43	1
miR-664a-5p	8	ACTGGCT**A**GGGAAAATGATTGGAT	153	3.92	1

Table S2. Top 50 pathway enrichments for the gene targets of unedited miR-200b-3p. Only the 604 gene targets of unedited miR-200b-3p that were expressed in BECs (CPM > 1) were included as input for pathway analysis, as implemented in TopFunn. Thirty-nine unique pathways are listed.

ID	Name	Database	Bonferroni Adjusted P-value	Number of genes from input	Number of genes in annotation
M16763	Neurotrophin signaling pathway	MSigDB C2 BIOCARTA (v7.1)	1.53E-11	27	126
M186	PDGFR-beta signaling pathway	MSigDB C2 BIOCARTA (v7.1)	2.30E-10	26	129
101143	Neurotrophin signaling pathway	BioSystems: KEGG	2.06E-09	24	119
M48	Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	MSigDB C2 BIOCARTA (v7.1)	1.89E-08	19	79
137930	Signaling events mediated by Hepatocyte Growth Factor Receptor (c-Met)	BioSystems: Pathway Interaction Database	3.30E-07	17	72
M295	Genes related to PIP3 signaling in cardiac myocytes	MSigDB C2 BIOCARTA (v7.1)	8.95E-06	15	67
1269443	Signalling by NGF	BioSystems: REACTOME	9.03E-06	43	483
M281	Signaling events mediated by focal adhesion kinase	MSigDB C2 BIOCARTA (v7.1)	1.32E-05	14	59
1457794	Signaling by MET	BioSystems: REACTOME	1.38E-05	15	69
M121	mTOR signaling pathway	MSigDB C2 BIOCARTA (v7.1)	1.38E-05	15	69
1269380	Signaling by EGFR	BioSystems: REACTOME	1.46E-05	36	367
M2404	Mechanism of Gene Regulation by Peroxisome Proliferators via PPARa(alpha)	MSigDB C2 BIOCARTA (v7.1)	2.35E-05	13	52
169348	Signaling events mediated by focal adhesion kinase	BioSystems: Pathway Interaction Database	3.86E-05	13	54

83107	Renal cell carcinoma	BioSystems: KEGG	5.05E-05	14	65
1269654	Transcriptional Regulation by TP53	BioSystems: REACTOME	7.93E-05	35	374
1269507	Signaling by Rho GTPases	BioSystems: REACTOME	9.33E-05	38	430
M13266	Renal cell carcinoma	MSigDB C2 BIOCARTA (v7.1)	1.38E-04	14	70
M164	ErbB1 downstream signaling	MSigDB C2 BIOCARTA (v7.1)	1.45E-04	17	105
M187	Neurotrophic factor-mediated Trk receptor signaling	MSigDB C2 BIOCARTA (v7.1)	1.85E-04	13	61
138001	mTOR signaling pathway	BioSystems: Pathway Interaction Database	1.85E-04	13	61
1270302	Developmental Biology	BioSystems: REACTOME	2.22E-04	69	1081
1269460	NGF signalling via TRKA from the plasma membrane	BioSystems: REACTOME	2.25E-04	35	390
1404799	Endocrine resistance	BioSystems: KEGG	2.34E-04	16	96
1270303	Axon guidance	BioSystems: REACTOME	4.81E-04	43	554
138057	ErbB1 downstream signaling	BioSystems: Pathway Interaction Database	5.57E-04	16	102
694606	Hepatitis B	BioSystems: KEGG	7.02E-04	19	144
1269479	Downstream signal transduction	BioSystems: REACTOME	7.36E-04	32	355
1269478	Signaling by PDGF	BioSystems: REACTOME	1.06E-03	33	379
PW:0000578	scatter factor/hepatocyte growth factor signaling	Pathway Ontology	1.22E-03	6	11
M195	C-MYB transcription factor network	MSigDB C2 BIOCARTA (v7.1)	1.47E-03	14	84
83067	Focal adhesion	BioSystems: KEGG	1.88E-03	22	199
1269650	Generic Transcription Pathway	BioSystems: REACTOME	2.06E-03	57	879

1268855	Diseases of signal transduction	BioSystems: REACTOME	2.21E-03	32	373
P00047	PDGF signaling pathway	PantherDB	2.40E-03	17	127
83105	Pathways in cancer	BioSystems: KEGG	2.70E-03	33	395
1269284	DAP12 signaling	BioSystems: REACTOME	2.86E-03	31	359
83065	Axon guidance	BioSystems: KEGG	3.54E-03	20	175
138073	C-MYB transcription factor network	BioSystems: Pathway Interaction Database	3.71E-03	13	78
1383076	Regulation of TP53 Activity	BioSystems: REACTOME	4.37E-03	19	162
137977	Neurotrophic factor-mediated Trk receptor signaling	BioSystems: Pathway Interaction Database	4.40E-03	11	56
M19118	Keratinocyte Differentiation	MSigDB C2 BIOCARTA (v7.1)	4.88E-03	10	46
M100	SHP2 signaling	MSigDB C2 BIOCARTA (v7.1)	5.29E-03	11	57
M13863	MAPKinase Signaling Pathway	MSigDB C2 BIOCARTA (v7.1)	5.76E-03	13	81
137989	FGF signaling pathway	BioSystems: Pathway Interaction Database	6.01E-03	10	47
852705	MicroRNAs in cancer	BioSystems: KEGG	6.06E-03	27	299
1269283	DAP12 interactions	BioSystems: REACTOME	6.75E-03	31	374
M7253	Focal adhesion	MSigDB C2 BIOCARTA (v7.1)	7.11E-03	21	199
138036	FoxO family signaling	BioSystems: Pathway Interaction Database	7.37E-03	10	48
M136	FoxO family signaling	MSigDB C2 BIOCARTA (v7.1)	8.99E-03	10	49
P00018	EGF receptor signaling pathway	PantherDB	9.17E-03	15	111

Figures

Figure S1. Study overview.

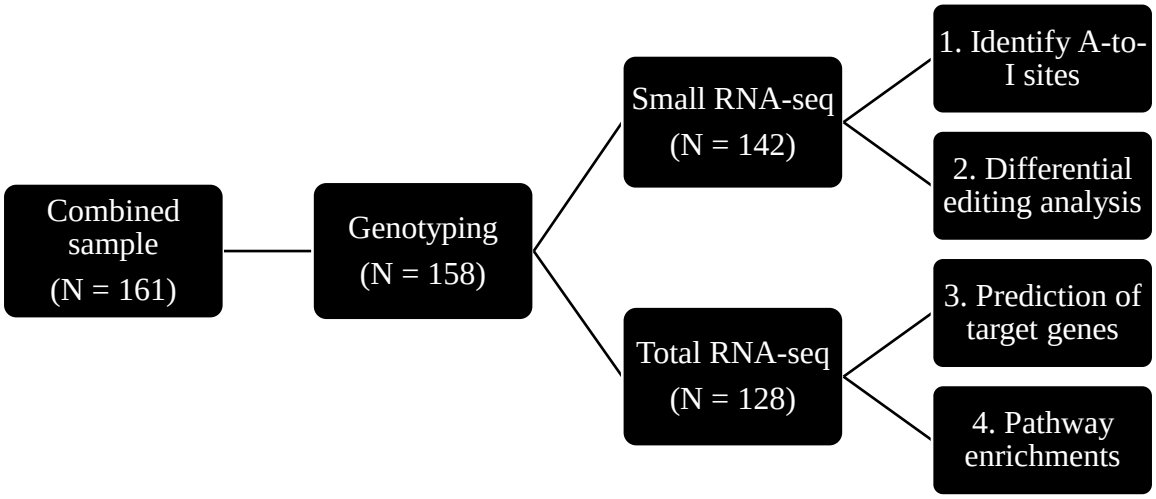


Figure S2. A-to-I editing of position 5 of miR-200b-3p in asthma cases and controls excluding one outlier. Associations of percentage (%) of A-to-I editing of position 5 of miR-200b-3p in bronchial epithelial cells with (**A**) asthma and (**B**) asthma severity groups based on STEP classification scores. One outlier (>3 SD) was included from the control group in Figure 1 of the main text. The number of subjects is shown below risk group. Adjusted p-values are shown above the horizontal bars. The p-values for (**A**) and (**B**) were corrected for four and three tests, respectively, using a Bonferroni correction.

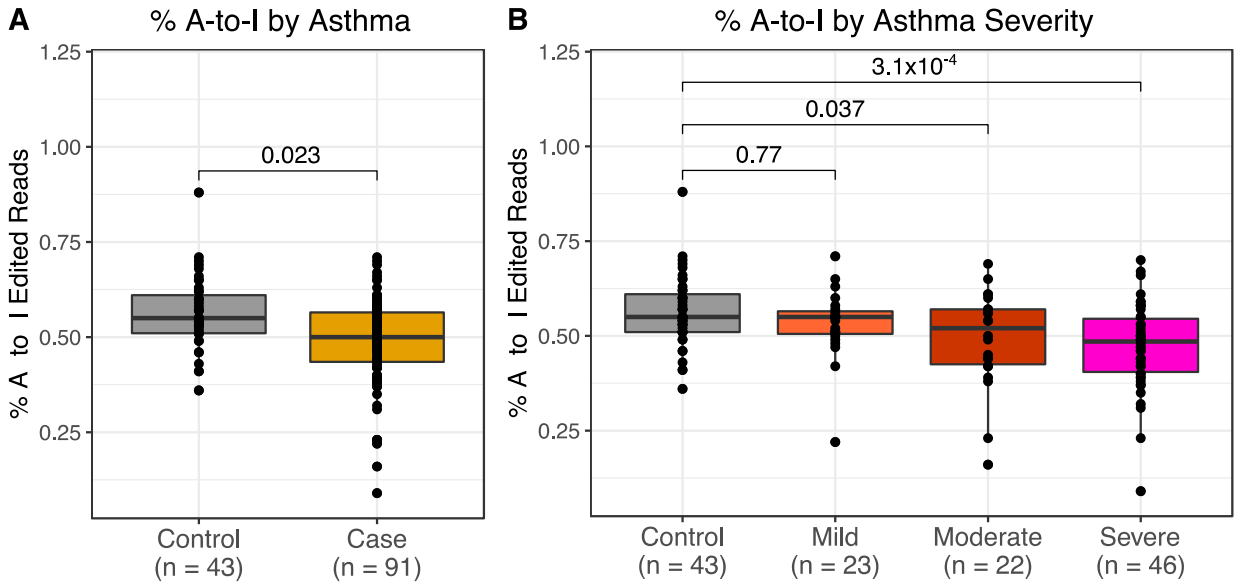


Figure S3. Differential expression analysis of the top four A-to-I edited sites. Comparisons of the percentage (%) of A-to-I editing at the 5th position of miR-200b-3p (top left), the 3rd position of miR-191-5p (top right), the 3rd position of miR-200b-3p (bottom left), and the 3rd position of miR-186-5p (bottom right) were made between 91 asthma cases and 44 non-asthma controls. Adjusted p-values are shown above the horizontal bars. P-values were corrected for four tests using a Bonferroni correction.

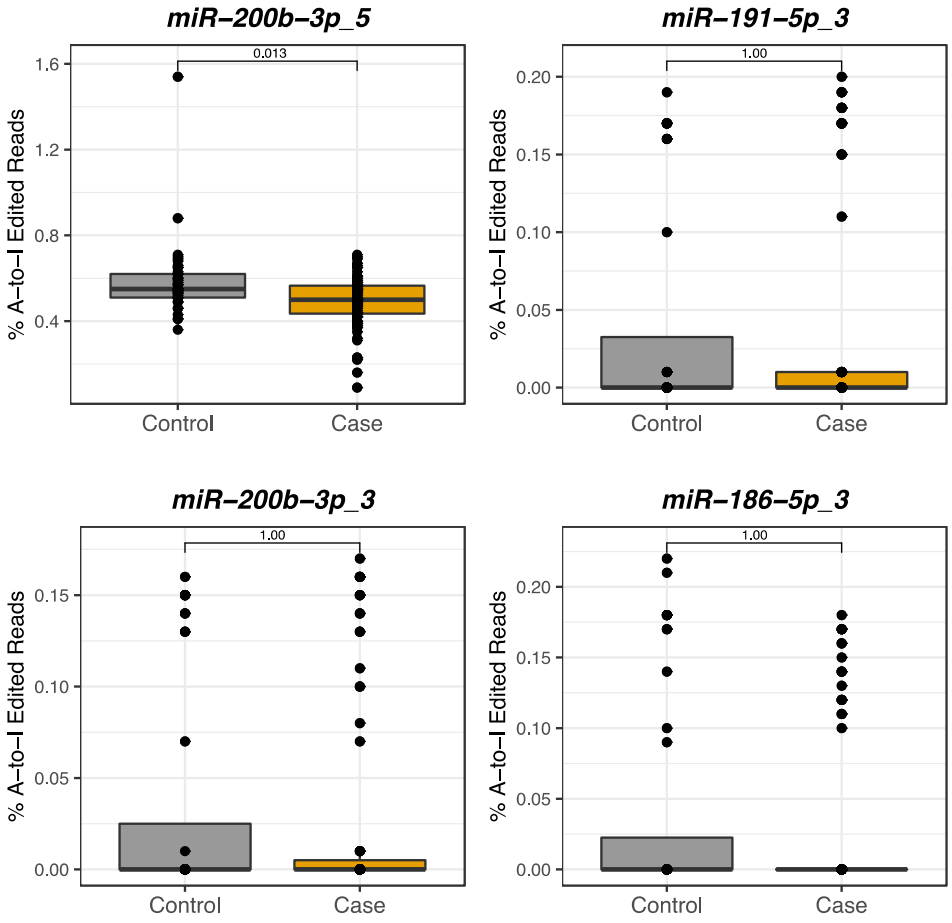
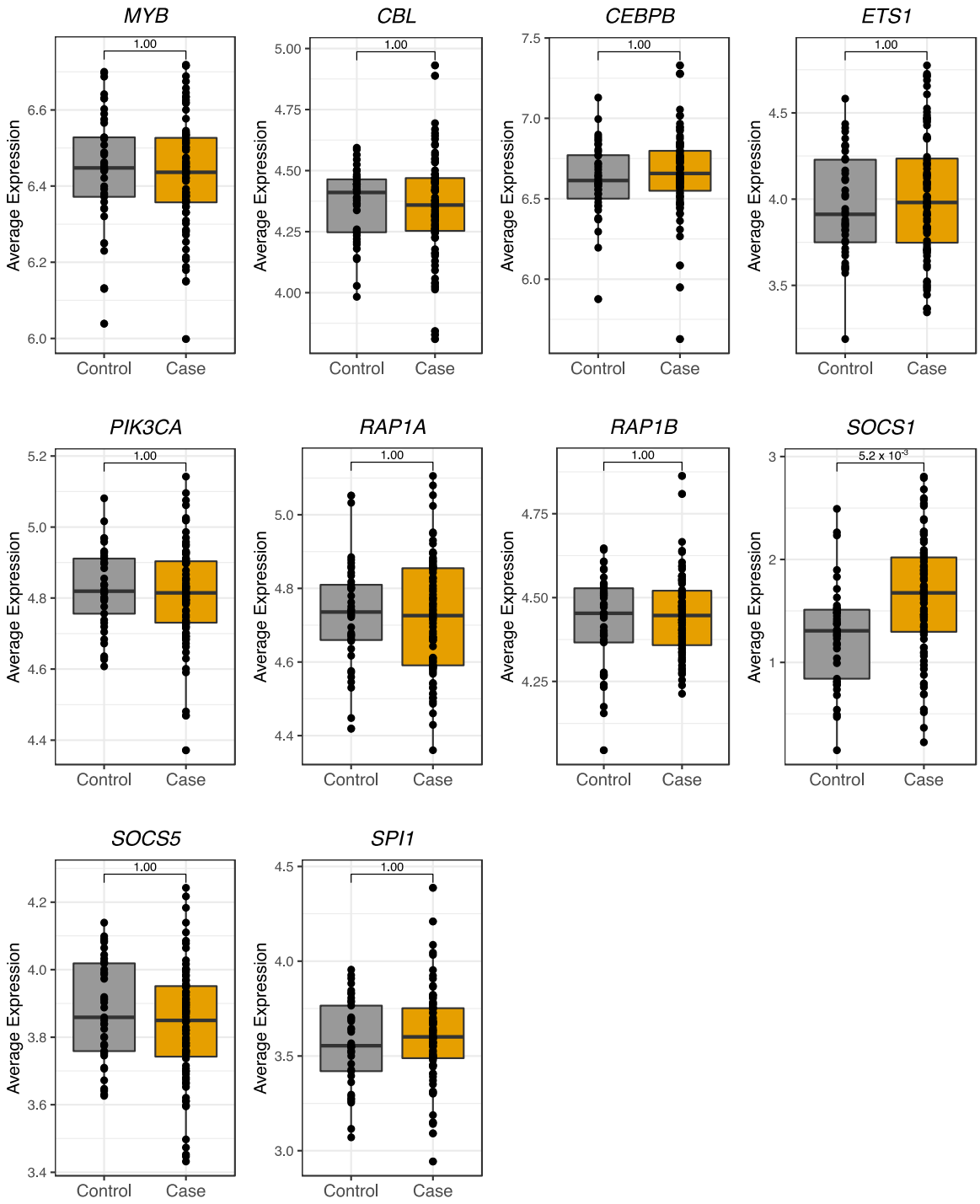


Figure S4. Normalized expression of ten gene targets for the two enriched pathways of the edited miR-200b-3p in 83 asthma cases and 41 controls. Gene targets that were expressed (CPM > 1) in BECs in at least 25% of the 124 subjects with RNA-seq were retained for analysis. Ten gene targets of the "IL-4 mediated signaling events" and "IFN-gamma" pathways are displayed. Adjusted p-values are shown above the horizontal bars. P-values were corrected for ten tests using a Bonferroni correction.



References

1. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, Boehnke M, Kang HM. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012; 91: 839-848.
2. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
3. Lu Y, Baras AS, Halushka MK. miRge 2.0 for comprehensive analysis of microRNA sequencing data. *BMC Bioinformatics* 2018; 19: 275.
4. Dobin A, Gingeras TR. Mapping RNA-seq Reads with STAR. *Curr Protoc Bioinformatics* 2015; 51: 11.14.11-11.14.19.
5. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014; 15: R29.
6. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007; 3: 1724-1735.
7. Alon S, Mor E, Vigneault F, Church GM, Locatelli F, Galeano F, Gallo A, Shomron N, Eisenberg E. Systematic identification of edited microRNAs in the human brain. *Genome Res* 2012; 22: 1533-1540.
8. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015; 4.
9. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 2009; 37: W305-311.