Supporting Information

Schembri *et al.* 10.1073/pnas.0806383106

SI Methods

RNA Isolation and Size Fractionation. RNA was extracted from bronchial airway epithelial cells and fractioned into low molecular weight (LMW) (<200 nt) and high molecular weight (HMW) (>200 nt) fractions, by using the miRNeasy mini kit (QIAGEN) according to manufacturer's protocol. LMW yields ranged from 0.7 to 2.0 μg, and HMW yields ranged from 7 to 15 μg of per sample. Integrity of the RNA and purity of the RNA size fractionation were confirmed by using a 10% Tns-Borate Urea Polyacrylamide Gel Electrophoresis (TBU PAGE) gel.

MicroRNA Microarray Hybridization and Data Acquisition. Three hundred nanograms of LMW RNA were labeled by using the FlashTag labeling kit (Genisphere, Inc.). MicroRNAs were first poly(A) tailed and then directly ligated to a fluorescent dendrimer (a branched DNA structure carrying ≈15 molecules of Oyster-550 dye (DeNovo Biolabels). After labeling, 5 μl of 10% BSA and 28.5 μl of 2× enhanced hybridization buffer (Genisphere) were added to the sample to make the hybridization mix. This was applied to Invitrogen NCode miRNA microarrays containing 1053 miRNAs from 6 species (467 human miRNAs) printed in triplicate. The arrays were incubated overnight (18 h) in a humidified chamber at 52°C and the following morning, serial 15-min washes were done with 2× SSC/0.2% SDS at 52°C, 2× SSC at room temperature, and 0.2× SSC at room temperature. After drying, the arrays were scanned on a GenePix 4000 scanner (Axon/Molecular Devices).

mRNA Microarray Hybridization and Data Acquisition. One microgram of the HMW fraction from each total RNA sample was used as starting material for mRNA expression analysis. By using random hexamers incorporating a T7 promoter, double-stranded cDNA was synthesized from 500 ng of the HMW RNA from which the majority of the ribosomal RNA had first been removed by using a RiboMinus Human/Mouse Transcriptome Isolation Kit (Invitrogen). cRNA was generated from the double-stranded cDNA template by an in vitro transcription reaction and purified by using the Affymetrix sample cleanup module. cDNA was regenerated by a random-primed reverse transcription using a dNTP mix containing dUTP and then fragmented with uracil DNA glycosylase and apurinic/apyrimidic endonuclease-1. Fragmented material was then end labeled with a biotinylated dideoxynucleotide, using terminal transferase. Five and a half micrograms of the fragmented, biotinylated cDNA were added to a hybridization mixture, loaded on a Human Exon 1.0 ST GeneChip, and hybridized for 16 h at 45°C and 60 rpm. After hybridization, the array was washed and stained according to the standard Affymetrix protocol. The stained array was scanned using an Affymetrix GeneChip Scanner 3000. These scans were used to generate intensity calculations on each pixel value (CEL files) for each array.

The ≈230,000 "core" exon probe sets on the Exon array that map to ≈17,800 empirically supported transcripts (RefSeq and full-length GenBank mRNAs) with a high degree of confidence were used to derive transcript-level expression estimates, using the model-based iterPLIER algorithm as implemented in the ExACT software package (Affymetrix). The gene annotations used for each probe set were from the annotation file obtained from Affymetrix (http://www.affymetrix.com).

Analysis of miRNAs Detected in Bronchial Epithelium. Principal component analysis was performed by using the Spotfire DecisionSite software (TIBCO Spotfire) to determine if bronchial airway samples separate on the basis of smoking status across the 232 microRNAs detected across these samples. PCA reveals modest separation based on smoking status across the first and second principal components with few exceptions (Fig. S3).

Enrichment of miRNA Targets Among Genes Previously Identified as Changing with Smoking in Vivo. To determine miRNAs that mediate the smoking-induced mRNA changes in a population of smokers ($n = 36$) and nonsmoker ($n = 27$) from a previous study (1), we first identified genes differentially expressed between current and never smokers, using a Student's t test. We identified 343 probe sets differentially expressed in smokers (false discovery rate, FDR < 0.05). Our goal was to determine whether mRNA data alone can predict miRNAs with expression that is altered by smoking. Using a hypergeometric test (http://www.broad.mit.edu/gsea/msigdb/annotate.jsp), we computed the enrichment of miRNA targets among the mRNAs that were differentially expressed in smokers. We obtained lists of miRNA targets from MSigDb (2). Approximately 26% of the mRNAs that were differentially expressed in smokers were predicted to be targets of just 20 miRNAs (Table S2). This led us to test the hypothesis using data from this study that the expression of miRNAs might also be perturbed in smokers.

Seven of the top enriched miRNAs from this analysis (miR-218, miR-15a, miR-106a/106b, miR-19b, miR-128a/128b, miR-130a, and miR-125b) were also identified as differentially expressed in the airways of smokers vs. nonsmokers in miRNA measurements on an independent set of samples used in this study (see Fig. 1). The miRNA expression profiling performed for this study, combined with the mRNA data analysis described above, provides concordant evidence that these miRNAs may play a role in mediating mRNA changes in response to cigarette smoke exposure in bronchial epithelium.

Validating Potential mRNA-miRNA Regulatory Relationships via Transient Transfection of miRNA in H1299 Cells. H1299 cells, an immortalized human lung carcinoma cell line, were cultured in RPMI 1640 media with 10% FBS and antibiotics (50 μg/ml penicillin/streptomycin) and plated at a density of 2.5×10^5 cells in 6-well plates 24 h before transfection. For mir-218 over-expression experiments, cells were transfected with either a 30 nM concentration of a short, synthetic RNA oligonucleotide sequence corresponding to mir-218 ($n = 5$) (Ambion) or a Cy3-labeled synthetic oligonucleotide negative control, containing a scrambled sequence ($n = 5$) (Ambion) using the lipofectamine 2000 transfection system (Invitrogen) according to manufacturer's protocol. For mir-218 knock-down experiments, cells were similarly transfected with 30 nM of anti-mir-218 ($n = 6$) (Ambion) or a Cy3-labeled anti-mir negative control ($n = 6$), as described above. The transfection efficiency was estimated to be close to 100% in both experiments as determined by the frequency of Cy3-positive cells in the negative control transfections. Cells were harvested at 48 h post transfection and RNA was isolated from these cells using the miRNeasy mini kit according to manufacturer's protocol. One microgram of total RNA from cells transfected with either the negative control ($n = 3$) or mir-218 ($n = 3$) was labeled and hybridized to Human Exon ST 1.0 arrays, with transcript-expression estimates determined as described above.

Validating Potential mRNA-miRNA Regulatory Relationships via Transient Transfection of miRNA in NHBE Cells. Primary human bronchial epithelial cells (NHBE) were purchased from Cell Applications and plated at ≈30% confluence in 6-well plates and cultured in optimized growth media (Cell Applications). For mir-218 knock-up and knock-down experiments. NHBE cells (6 replicates per condition) were transfected mir-218, anti-mir-218, or a negative control containing a Cy3-labeled scrambled sequence (Ambion) at 30 nM using the PrimeFect siRNA Transfection reagent (Lonza) according to manufacturer's protocol. The transfection efficiency was estimated to be close to 100% in both experiments as determined by the frequency of Cy3-positive cells in the negative control transfectants. Cells were harvested at 48 h post transfection and RNA was isolated from these cells using the miRNeasy mini kit according to the manufacturer's protocol.

Effects of CSC on DNA Synthesis in Normal Human Bronchial Epithelial Cells. We performed [^{3}H]thymidine incorporation assays to determine the effects of CSC on DNA synthesis. Briefly, normal human bronchial epithelial (NHBE) cells were plated in 24-well plates at a density of 40,000 cells per well. Replicate cultures of NHBE cells were treated with different concentrations of CSC (0–60 μg/ml) or with DMSO for controls. The final concentration of DMSO in the cultures was 0.2%. Twenty-four hours after the addition of CSC, 1 μCi of [3H]thymidine was added to each well and the cells were returned to the incubator for an additional 24 h. The medium was then removed and the cells were washed 3 times with 2 ml of 5% trichloroacetic acid (TCA). The TCA-insoluble genomic DNA was dissolved in 0.3 ml of 0.3 N NaOH and the incorporated radioactivity was determined by scintillation counting.

Enrichment of miRNA Targets Among Genes Identified as Changing in NHBE Exposed to Cigarette Smoke. To determine the effect of cigarette smoke on targets of mir-218 (the miRNA that changed most in smokers in vivo), gene set enrichment analysis (GSEA) (2) was performed on an mRNA expression dataset obtained from the Gene Expression Omnibus (GSE10700). This dataset contained gene-expression profiles of NHBE cells exposed to an American brand of "light" cigarette smoke or a reference cigarette (2R4F, University of Kentucky). We sought to determine whether mir-218 targets are enriched among the genes most up-regulated in response to cigarette smoke exposure in these cells. The study assayed the expression of mRNA in NHBE cells at 2, 4, 8, and 24 h postexposure to cigarette smoke. We observed that starting at 4 h postexposure, mir-218 targets were up-regulated significantly in NHBE cells exposed to cigarette smoke relative to nonsmoke-exposed controls (Table S4), which indicates that the targets of mir-218 are induced in response to acute cigarette smoke exposure and suggests that the down-regulation of mir-218 in these cells is a possible mechanism for these gene-expression differences.

Transcription Factor Analysis of Predicted miRNA Targets. To determine whether genes altered by smoking were regulated by common transcription factors, the ROVER (Relative OVERabundance of *cis*-elements) (3) algorithm was used. This algorithm compares 2 groups of promoters (in this case, differentially expressed genes in smokers vs. a set of randomly selected background genes) to determine the relative overabundance of binding sites for a particular transcription factor in one group vs. the other.

Real-Time PCR Validation of miRNA and mRNA Targets. Quantitative real-time PCR was used to confirm the differential expression of a select number of miRNAs in vivo and target mRNAs in vitro. For miRNAs, 10 ng of LMW RNA was used in a Taq-Man assay (Applied Biosystems) as per manufacturer's protocol to confirm the differential expression of mir-218, mir-128b, mir-500, and mir-181d, normalized to expression of U6. Mir-218 levels were also confirmed in the in vitro transfection and CSC exposure experiments with expression normalized to U6 levels. For mRNA, 25 ng of starting HMW RNA was used for qRT-PCR of 5 putative mir-218 targets including MAFG, ECOP, NAC-1, NFE2L1, and LASP1, all normalized to expression of GAPDH (see below for primer sequences) using the SYBER green protocol (Applied Biosystems). For both miRNA and mRNA quantitation from in vivo and H1299 samples, each PCR was run in duplicate. PCR were run in triplicate for miRNA and mRNA quantitation from all NHBE samples. Forty cycles of amplification were used and data acquisition was carried out with both ABI Prism 7700 Sequence Detector and StepOnePlus Real-Time PCR systems (Applied Biosystems). Data from all qRT-PCR experiments were analyzed using the comparative Ct method (4), and replicates were filtered by using a q test, with outliers being removed at 90% confidence (5).

Primer Sequences for miRNA Target PCR Validation

MAFG:
Forward, CACCAGCGTCATCACAATAGTAAAG;
Reverse, CCCTACGATCGGGCATCC.
NAC-1:
Forward, CTCTTCTACCAACGATCCCCGT;
Reverse, GCATTCATCTCGCTCTCCTTGA.
ECOP:
Forward, CTGCAGGCCTCCATCTAAATGA;
Reverse, GCAAAACAAATGACCTGGAGGA.
LASP1:
Forward, GCATGCTTCCATTGCGAGA;
Reverse, TGCTTGAGGCGAAGGTTTTC.
NFE2L1:
Forward, TTTGCCATAAAGACTCCGTGTAAC;
Reverse, AGGAGAAAAATCCCAAGTGTTCAC.

1. Beane J, *et al*. (2007) Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 8:R201.
2. Subramanian A, *et al*. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550.
3. Haverty PM, Hansen U, Weng Z (2004) Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. *Nucleic Acids Res* 32:179–188.
4. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408.
5. Shoemaker JP, Garland CW, Steinfeld JI (1974) *Experiments in Physical Chemistry* (McGraw–Hill, New York), pp 34–39.

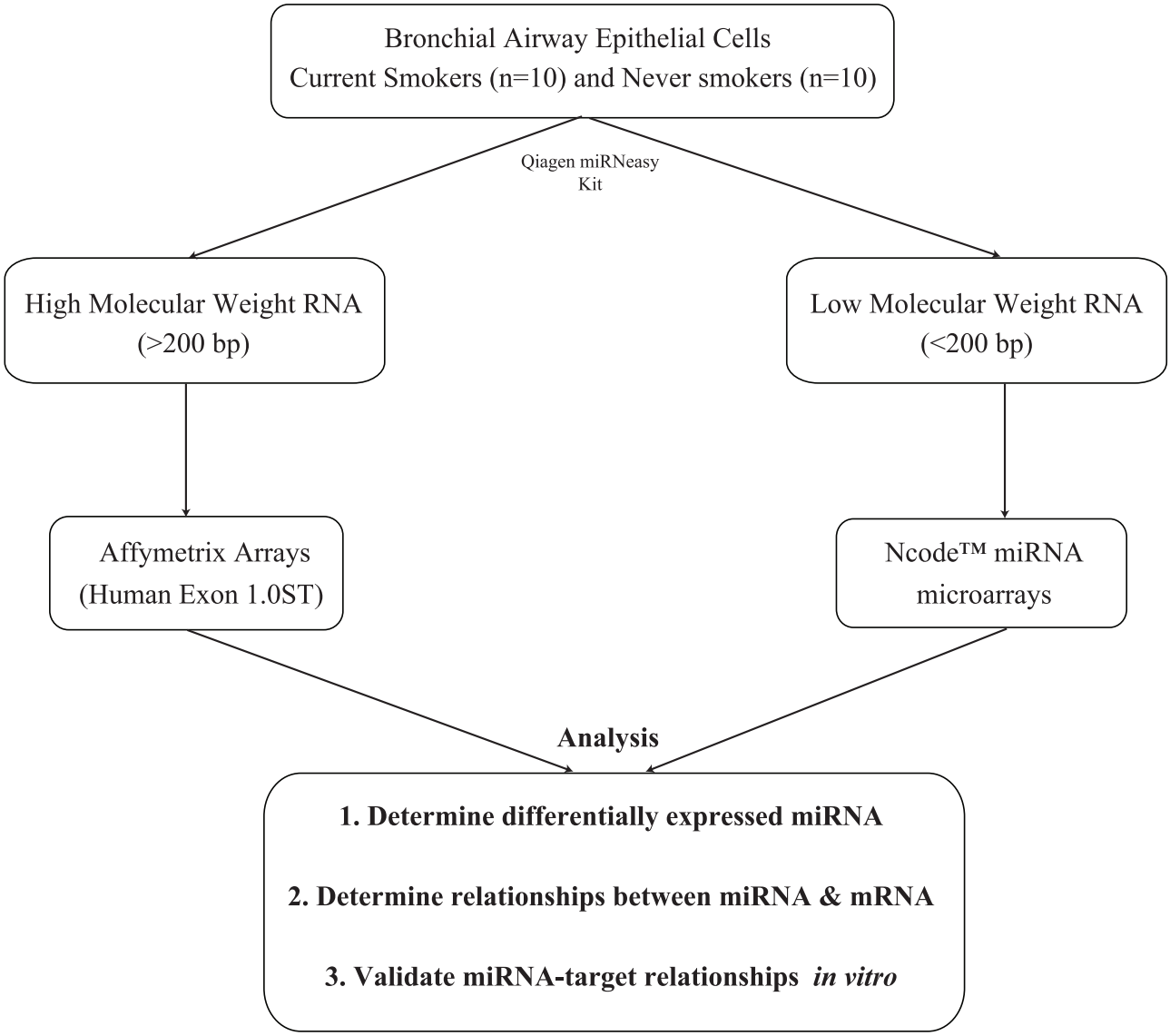

Fig. S1. Study design for in vivo microRNA and mRNA microarray experiments.

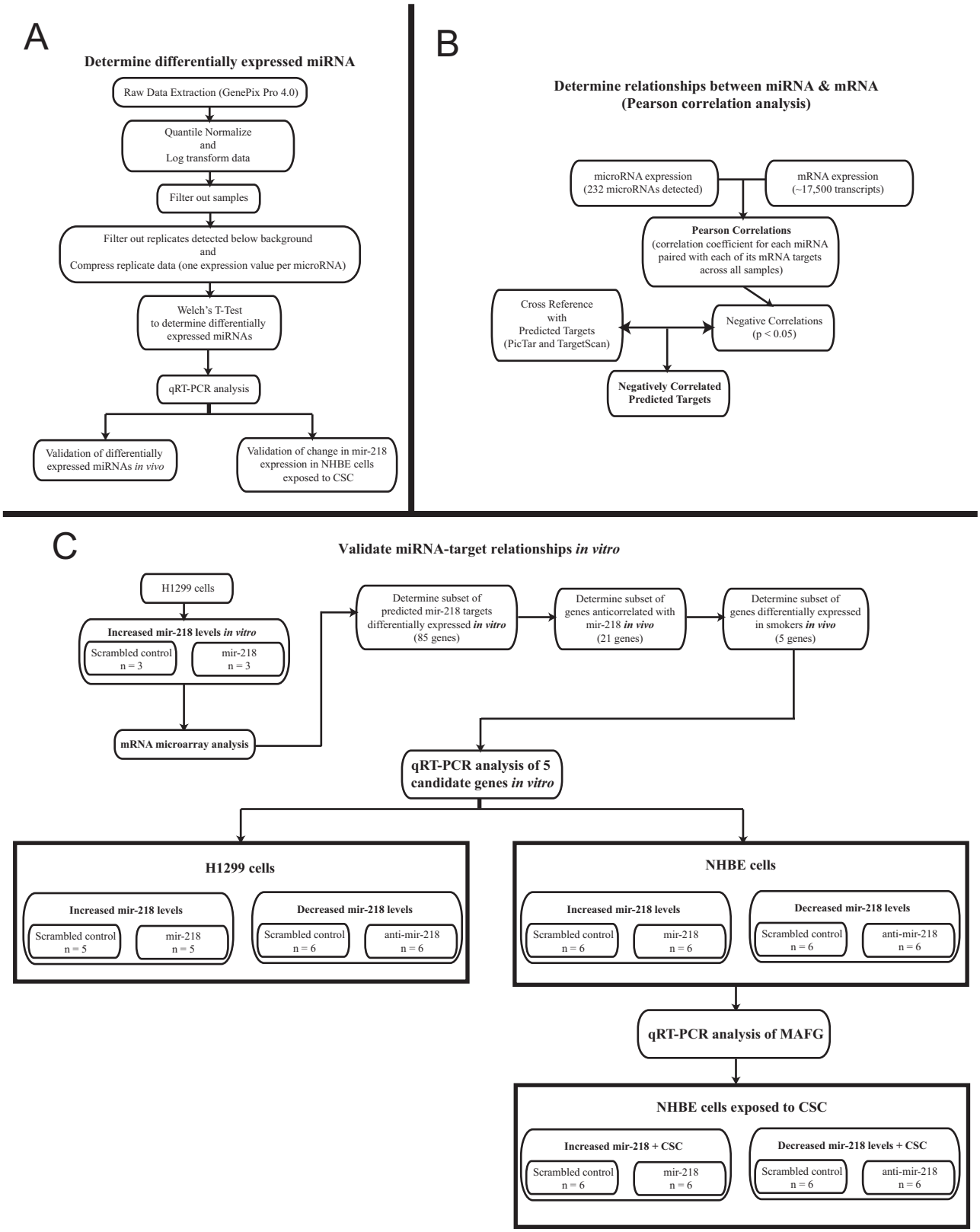

Fig. S2. Analytical flow for (*A*) miRNA microarray data preprocessing and statistical analysis, (*B*) miRNA–mRNA correlation analysis, and (*C*) in vitro validation of miRNA–mRNA relationships.

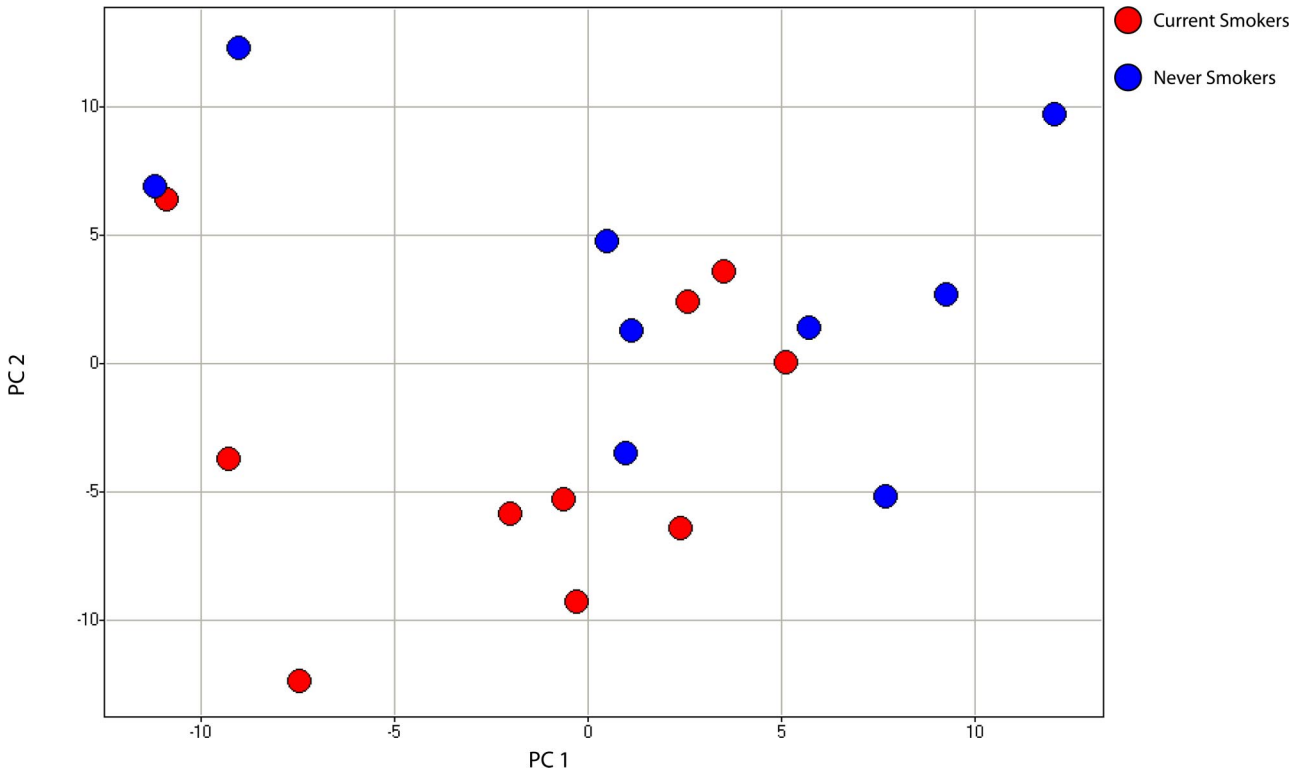

Fig. S3. Principal component analysis of smoker and nonsmoker airway samples across the expression levels of 232 miRNAs detected in a majority of subjects.

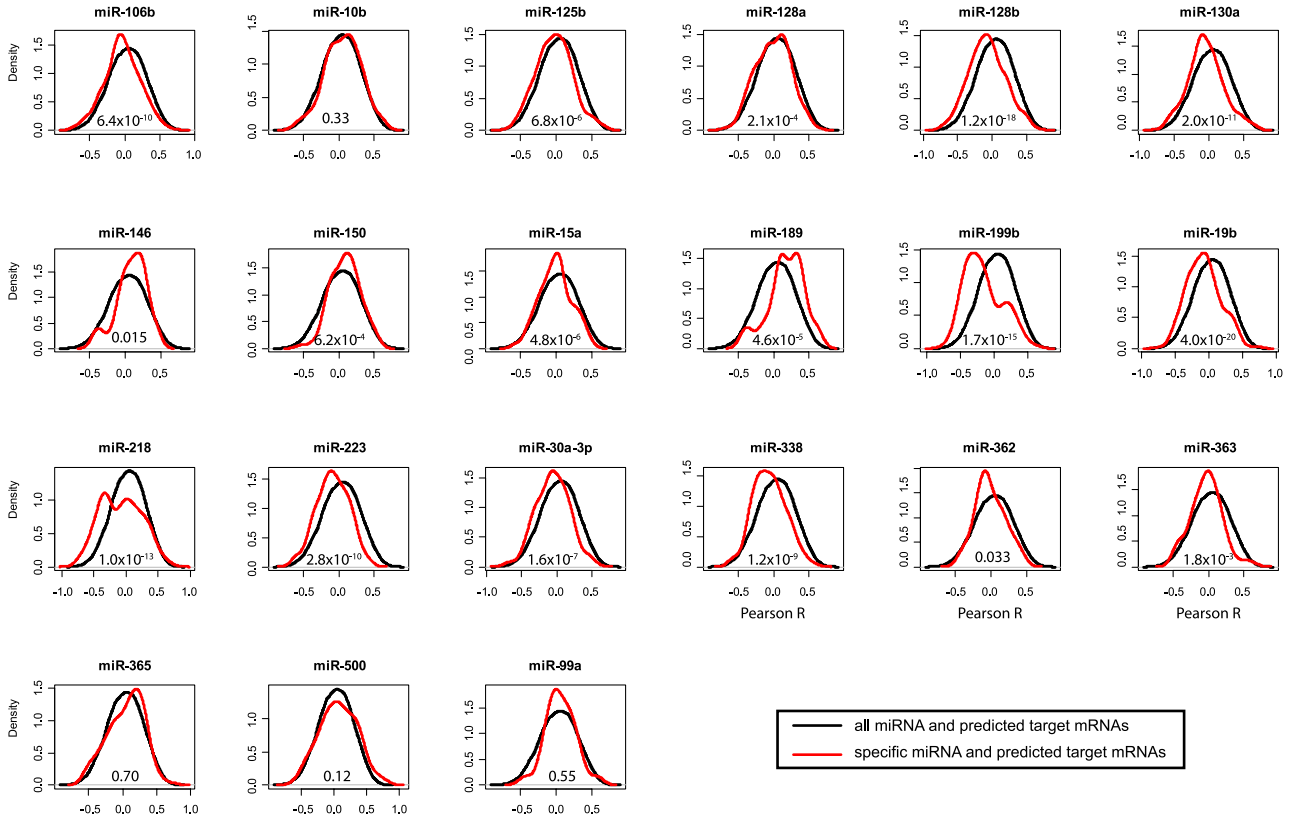

Fig. S4. Density plots of the distribution of pairwise Pearson correlations from 21 of the 28 differentially expressed miRNAs and their predicted targets (shown in red) vs. 232 miRNAs and their predicted targets (shown in black). Several miRNAs show a significant level of anticorrelation with their predicted targets when compared with all possible pairwise correlations (Welch's *t* test; $P < 0.05$). Six of the miRNAs shown above are displayed in Fig. 3. Seven miRNAs did not have any predicted targets in either the PicTar or the TargetScan databases and therefore are not shown.

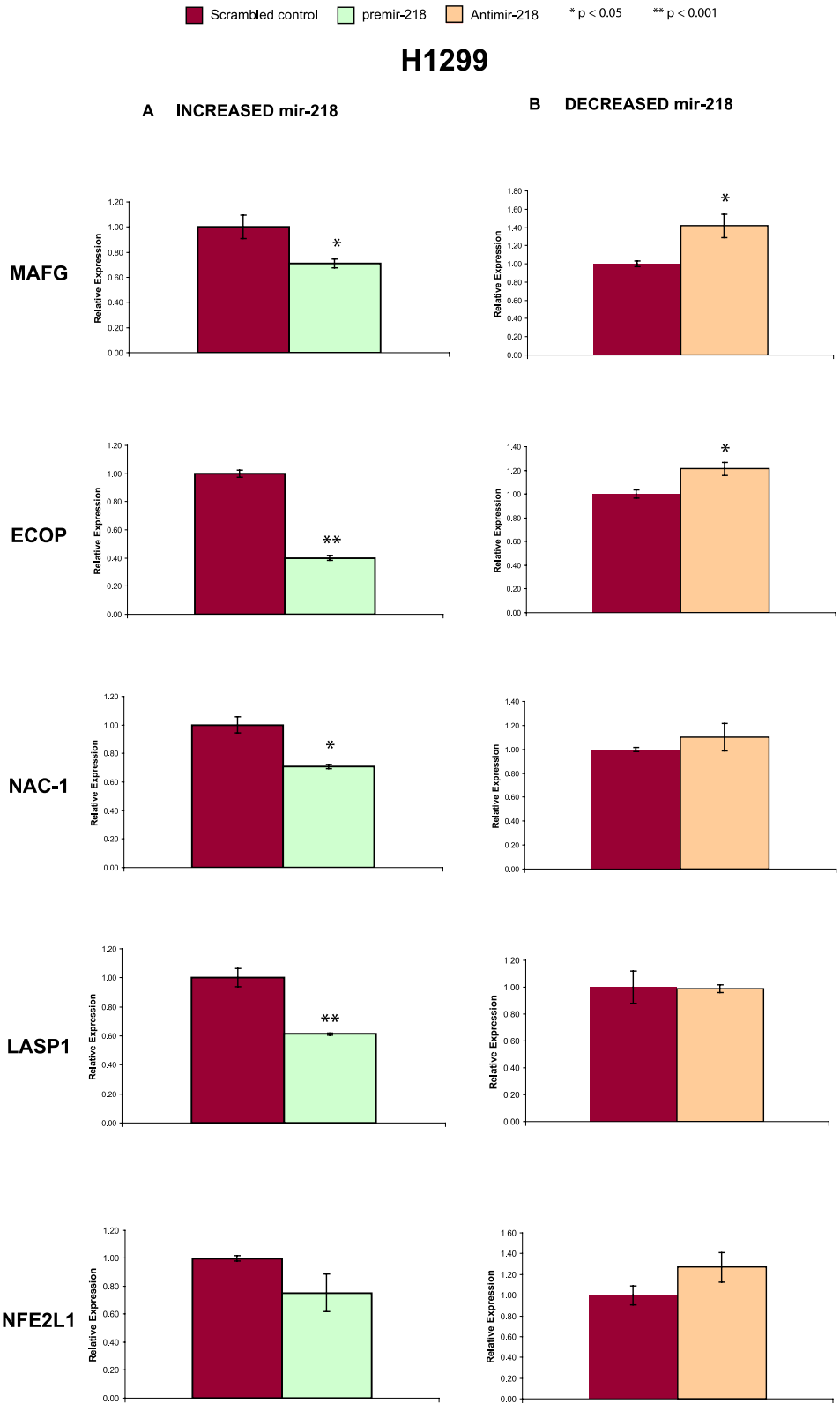

Fig. S5. Relative expression of predicted mir-218 targets in H1299 cells when mir-218 levels are increased (*A*, $n = 5$ per condition) or decreased (*B*, $n = 6$ per condition). Target expression is inversely correlated with mir-218 expression levels, with significant down-regulation seen in the presence of mir-218. Error bars indicate standard error. *P* values indicate Student's *t* test between the gene expression in mir-218 or anti-mir-218 vs. control transfectants, respectively.

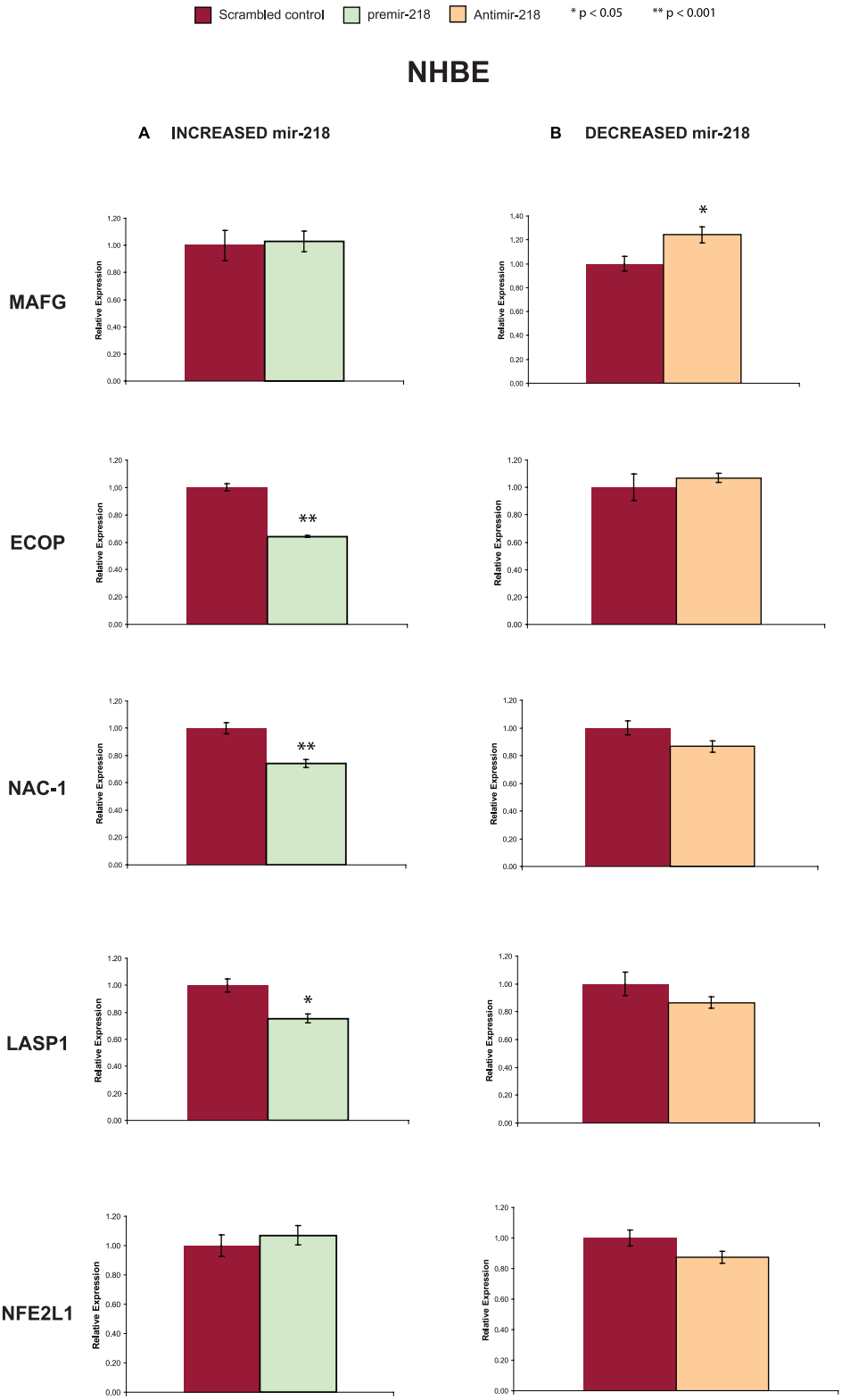

Fig. S6. Relative expression of predicted mir-218 targets in primary normal human bronchial epithelial cells when mir-218 levels are increased (*A*, $n = 6$ per condition) or decreased (*B*, $n = 6$ per condition). Target expression is down-regulated with increased mir-218 expression levels in 3 of 5 putative targets, whereas expression is marginally up-regulated with decreased mir-218 levels in 2 of 5 targets. Error bars indicate standard error. *P* values indicate Student's *t* test between the gene expression in mir-218 or anti-mir-218 vs. control transfectants, respectively.

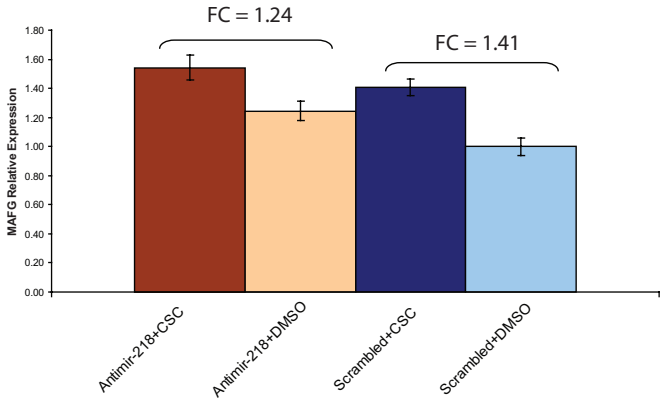

Fig. S7. Reduced expression of mir-218 in anti-mir-transfected NHBE also attenuates the CSC-dependent induction of MAFG but to a lesser extent (which does not reach statistical significance).

Table S1. Mean and standard deviations of normalized, log-transformed expression of the 28 differentially expressed miRNAs in current and never smokers (see Figure 1). Also provided are fold changes, *P* values, and FDR of differential expression of each miRNA between current and never smokers

	Never smokers		Current smokers				
miRNA	Mean	SD	Mean	SD	Fold change (log_2)	*P* value	FDR
mir-218	7.15	1.02	5.30	0.76	−1.85	4.59*E*-04	0.06
mir-128b	6.21	0.57	4.77	0.90	−1.44	5.10*E*-04	0.06
mir-30a-3p	7.44	0.48	6.11	0.93	−1.32	1.22*E*-03	0.09
miR-106b	9.80	0.68	8.00	1.32	−1.81	2.28*E*-03	0.13
miR-125b	11.93	1.22	10.22	1.03	−1.70	4.60*E*-03	0.21
miR-18a*	4.60	0.73	5.57	0.66	0.98	6.64*E*-03	0.21
miR-500	5.07	1.13	3.60	0.89	−1.47	6.87*E*-03	0.21
miR-106a	10.16	0.81	9.00	0.91	−1.17	7.40*E*-03	0.21
miR-189	5.27	0.49	5.82	0.24	0.55	9.59*E*-03	0.25
miR-652	7.34	0.46	5.98	1.30	−1.37	1.24*E*-02	0.29
miR-362	5.28	0.62	4.49	0.72	−0.80	1.66*E*-02	0.29
miR-181d	7.02	0.77	7.94	0.72	0.92	1.66*E*-02	0.29
miR-337	5.65	0.61	6.42	0.71	0.77	1.74*E*-02	0.29
miR-130a	7.87	0.83	6.87	0.84	−1.00	1.76*E*-02	0.29
miR-223	9.02	1.31	7.47	1.29	−1.55	1.88*E*-02	0.29
miR-146	10.27	1.20	8.55	1.56	−1.72	2.22*E*-02	0.31
miR-128a	6.55	0.47	5.74	0.92	−0.81	2.27*E*-02	0.31
miR-99a	11.38	0.71	10.72	0.40	−0.66	3.03*E*-02	0.36
miR-19b	10.23	0.80	9.38	0.80	−0.85	3.10*E*-02	0.36
miR-17–3p	5.74	0.59	5.08	0.59	−0.67	3.24*E*-02	0.36
miR-338	4.51	0.43	4.06	0.45	−0.45	3.52*E*-02	0.36
miR-150	9.60	1.05	8.60	0.91	−1.00	3.82*E*-02	0.36
miR-15a	11.18	0.67	10.34	0.94	−0.84	3.88*E*-02	0.36
miR-199b	5.76	1.09	4.81	0.74	−0.96	4.30*E*-02	0.36
miR-10b	5.83	0.67	5.14	0.74	−0.69	4.52*E*-02	0.36
miR-365	4.24	1.00	5.21	0.95	0.97	4.56*E*-02	0.36
miR-625	5.21	0.68	4.41	0.99	−0.80	4.59*E*-02	0.36
miR-363	5.89	0.62	5.21	0.69	−0.68	4.76*E*-02	0.36

Table S2. MicroRNAs with at least 5 targets that are statistically overrepresented among the genes most altered in the bronchial epithelium of smokers from a prior study (1)

Rank	*P* value	Gene set name (MSigDb)
1	1.98*E*-04	GTGCCTT, MIR-506
2	5.18*E*-03	AAGCACA, MIR-218
3	5.41*E*-03	CTTTGTA, MIR-524
4	8.33*E*-03	AGCACTT, MIR-93, MIR-302A, MIR-302B, MIR-302C, MIR-302D, MIR-372, MIR-373, MIR-520E, MIR-520A, MIR-526B, MIR-520B, MIR-520C, MIR-520D
5	8.40*E*-03	TGGTGCT, MIR-29A, MIR-29B, MIR-29C
6	9.14*E*-03	CACTGTG, MIR-128A, MIR-128B
7	9.79*E*-03	TTTGTAG, MIR-520D
8	1.07*E*-02	TTGCACT, MIR-130A, MIR-301, MIR-130B
9	1.13*E*-02	TGCTGCT, MIR-15A, MIR-16, MIR-15B, MIR-195, MIR-424, MIR-497
10	2.08*E*-02	GCACTTT, MIR-17–5P, MIR-20A, MIR-106A, MIR-106B, MIR-20B, MIR-519D
11	2.09*E*-02	ACCAAAG, MIR-9
12	2.17*E*-02	TGCTTTG, MIR-330
13	2.39*E*-02	CTCAGGG, MIR-125B, MIR-125A
14	2.42*E*-02	TCATCTC, MIR-143
15	2.60*E*-02	TTTGCAC, MIR-19A, MIR-19B
16	2.87*E*-02	TGCACTT, MIR-519C, MIR-519B, MIR-519[a]
17	3.12*E*-02	ATGTACA, MIR-493
18	3.52*E*-02	TGCCTTA, MIR-124A
19	3.67*E*-02	TGCACTG, MIR-148A, MIR-152, MIR-148B
20	3.73*E*-02	AAGCCAT, MIR-135A, MIR-135B

Target gene sets for human miRNAs were determined using MSigDB (http://www.broad.mit.edu/gsea/msigdb/index.jsp). MiRNAs are listed as individual miRNAs or as a cluster or family of miRNAs where relevant. Targets for the top 10 miRNA/miRNA families account for 17% of the differentially expressed mRNAs in smokers, whereas targets for the top 20 account for 26% of the differentially expressed mRNAs. Targets for all 222 miRNAs listed in MSigDB account for 45% of the differentially expressed smoking-induced mRNAs.

Table S3. Functional analysis of targets of select differentially expressed miRNAs

miRNA	Anticorrelated targets	Gene Ontology functional categories	KEGG pathways
mir-218	62	Regulation of actin filament length, intracellular signaling cascade, regulation of anti-apoptosis	Tight junction
mir-128b	36	I-kappaB kinase/NF-kappaB cascade	MAPK signaling pathway
mir-19b	31	Intracellular protein transport, cell–cell signaling, DNA methylation, cell motility, cell morphogenesis	Various metabolic and biosynthesis pathways
mir-106b	28	Regulation of transcription, DNA dependent, development, neurogenesis	Wnt signaling pathway, purine metabolism, regulation of actin cytoskeleton
mir-130a	27	Regulation of transcription, DNA dependent	MAPK signaling pathway
mir-199b	23	Wnt receptor signaling pathway, organogenesis, cell growth, morphogenesis, differentiation	Hedgehog signaling pathway, Wnt signaling pathway

We examined the correlation between the expression of miRNAs that are differentially expressed in the bronchial epithelium of current and never smokers and their predicted targets. The number of predicted mRNA targets per miRNA showing significantly anticorrelated expression (Pearson correlation, P value <0.05) is shown. Targets were identified in the PicTar or the TargetScan databases. Gene Ontology and KEGG pathways that are overrepresented ($P < 0.05$) within each set of anticorrelated mRNAs are listed.

Table S4. Gene Set Enrichment Analysis (GSEA) on expression targets of mir-218 in NHBE cells exposed to two different types of cigarette smoke, by using data obtained from an independent mRNA expression dataset (GSE10700)

	Direction of mir-218 target expression in smoke-exposed NHBE cells	FDR, *P* value
American brand of ″light″ cigarettes (h)		
2	Down	FDR < 0.25, $P < 0.01$
4	Up	FDR < 0.25
8	Up	FDR < 0.25, $P < 0.01$
24	Up	FDR < 0.25, $P < 0.01$
2R4F (University of Kentucky) (h)		
2	Up	Not significant
4	Up	Not significant
8	Up	FDR < 0.25, $P < 0.05$
24	Up	FDR < 0.25

Results suggest that mir-218 targets are significantly up-regulated in cells exposed to cigarette smoke after 4 h of exposure. Targets for mir-218 were determined as described in the main text *Methods*. *P* values and FDR values were determined by GSEA.