Supplementary materials and methods:

DNA extraction, sample preparation and hybridization protocol (BEC biopsy cohort)

Bronchial biopsies were taken from segmental divisions of the main bronchi and frozen in Tissue-Tek (VWR, Radnor, PA) at -80°C. On the day of analysis, samples were thawed at room temperature and cut from the blocks when they were semi-solid. Samples were lysed in 600 µl RLT-plus using an IKA Ultra Turrax T10 Homogenizer. Total genomic DNA was extracted using AllPrep DNA/RNA Mini kit, according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). Concentrations of DNA were assessed using a Nanodrop-1000 and on a Labchip GX (Perkin Elmer, Waltham, MA). Eluted DNA was precipitated in ethanol, reconstituted in elution buffer and treated with sodium bisulfite. This was followed by a PCR-free whole genome amplification (Truseq DNA PCR-free workflow), after which the treated DNA was hybridized to the Infinium HumanMethylation450 BeadChip array (450k array). After hybridization, allele-specific single-base extension incorporated a fluorescent label for detection of methylated and unmethylated sites. The conversion and hybridization were performed according to the manufacturer's instructions (Illumina, San Diego, CA). Every beadchip was processed with a control of DNA from a blood sample of a single female, which was used to assess the efficiency of the normalization procedure.

RNA extraction, sample preparation and high-throughput sequencing (bronchial epithelial cell biopsy (BEC cohort)

RNA samples were further processed using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina, San Diego, CA), using an automated procedure in a Caliper Sciclone NGS Workstation (PerkinElmer, Waltham, MA). In this procedure, all cytoplasmic and mitochondrial rRNA was removed (RiboZero Gold kit). The obtained

cDNA fragment libraries were loaded in pools of multiple samples unto an Illumina HiSeq2500 sequencer using default parameters for paired-end sequencing (2 × 100 bp). The trimmed fastQ files were aligned to build b37 of the human reference genome using HISAT (version 0.1.5) allowing for 2 mismatches. Before gene quantification SAMtools (version 1.2) was used to sort the aligned reads. The gene level quantification was performed by HTSeq (version 0.6.1p1) using Ensembl version 75 as gene annotation database. Quality control (QC) metrics were calculated for the raw sequencing data, using the FastQC tool (version 0.11.3) (Andrews 2010). Alignments of 220 subjects were obtained. QC metrics were calculated for the aligned reads using Picard-tools (version 1.130) (http://picard.sourceforge.net) CollectRnaSeqMetrics, MarkDuplicates, CollectInsertSize-Metrics and SAMtools flagstat. In addition, we checked for concordance between sexlinked (XIST and Y-chromosomal genes) gene expression and reported sex. All samples were concordant.

Quality control, sample filtering, probe filtering and normalization (BEC cohort)

Raw intensity values were read from IDAT-files and converted to β values using the minfi-package. We identified samples with insufficient quality using the R-package MethylAid, based on five diagnostic filter variables with the following thresholds: MU = 10.5, OP = 11.75, BS = 12.5, HC = 13.25, and DP = 0.95. Genotypes from 65 single nucleotide polymorphisms (SNP) probes on the 450k array were compared to previously acquired genotypes from blood samples and discordant samples were discarded as swapped or contaminated samples. We excluded probes which had intensities indistinguishable from background levels (>1% of samples at P = 0.05), probes which were cross-reactive, probes which had SNPs segregating at the interrogation or extension site (MAF > 10%), and type I probes which displayed high

intensity signals. In addition, we excluded all X-linked and Y-linked probes. After probe-filtering, raw β-values were background-corrected and normalized using the dasen method from the wateRmelon-package.

Supplementary table 1: Pyrosequencing primers

Primer for amplification were designed using Pyromark Assay Design 2.0 software **(**Qiagen, Hilden, Germany) under recommended conditions.

Gene	*COL4A3*
CpG site	cg11797365
F primer	ATTTTTGATTTTTAGGTATTGGATAAGTTT bio
R primer	CCCCCACTTCCATCCCTCCTC
S primer	ACTTCACCCAAATCCC
Sequence analysed	YAAACAAACA ACYAACRAAA AAAAACAAAC YAAAATACCC CYAAACAAAT AAAATACYCYAAACTAAAAA AAAAAA
F primer: forward primer; R primer: reverse primer; S primer: sequence primer; bio: biotinylation.	

Supplementary table 2: ChIP- Seq primers list

Primer name	**Sequence**	**Purpose**
Untr4	Human negative control primer set	Negative control
h *ANO7*_+16k	GCTTGAGGTGATTGAGGTGTG	Positive control
h *ANO7*_+16k	CAAGGCCCTCTGTGACACTC	
h *COL4A3*_+569	TGTCCAGTGCCTAGGAGTCAG	Test site
h *COL4A3*_+569	TCCCTGGCGATTATCTATGTG	
h *AHDC1*_-462	CTCCCTCCTCCTGCCTTCT	Test site
h AHDC1_-462	GCGGACCAGACAGACAGAC	
h *AP1S3*_+46	GCTTAGACCATGGCTGCTTC	Test site
h *AP1S3*_+46	GAGGAGGAGAAGGGAGGAGA	
ANO7: Anoctamin-7; AHDC1: AT-hook DNA binding motif containing 1; AP1S3: Adapted related protein complex 1 subunit sigma 3; Untr4: Untranscribed genomic 4		

Supplementary table 3: Univariate analysis: *COL4A3* expression is independent of age, weight, atopy, and ICS use

Variable	Spearman's ρ co-efficient	p-value
Disease status	-0.250	0.002
Sex	0.160	0.055
Atopy	0.104	0.212
Weight	-0.085	0.308
ICS	-0.082	0.326
Age	-0.014	0.869

Supplementary table 4: Multivariate analysis: *COL4A3* expression correlates with disease status and sex but not with age, atopy and use of ICS

Variable	Estimate	Std Error	p-value	Lower 95%	Upper 95%
Intercept	0.831	0.421	1.972	0.051	-0.002
disease status	-0.103	0.048	-2.123	0.036	-0.198
Sex (female)	0.065	0.034	1.893	0.061	-0.003
age	0.000	0.002	0.174	0.862	-0.004
ICS	0.133	0.101	1.319	0.189	-0.066
Atopy	-0.123	0.405	-0.303	0.763	-0.923

Supplementary table 5: DNA Methylation values at cg11797365 for ICS stratification

	Statistics	Healthy	Asthma, no ICS	Asthma, ICS users
cg11797365	Mean	0.17	0.19	0.17
	Std Dev	0.03	0.05	0.04
	Number of subjects	70	26	44

Supplementary table 6: Quantitative ChIP-Seq analysis

Methylation (%) at cg11797365	BEAS-2Bs	NHBEs	p-value
Mean (SD)	6.4 (1.0)	10.7 (1.7)	0.0003
n	9	10	

Supplementary table 7: Regression analysis: No direct association of gene expression and levels of methylation

Variable	Estimate	Std Error	p-value
Intercept	0.110	0.004	5.19E-77
z-value (COL4A3)	-0.001	0.003	0.832
Data Set [1-0]	0.061	0.005	2.1E-28

Data set [1-0]: 0- Nicodemus Johnson cohort; 1- BEC cohort

Supplementary table 8: Regression analysis of gene expression, levels of DNA methylation and outcome of asthma

Variable	Estimate	Std Error	p-value	Lower 95%	Upper 95%
Intercept	-1.821	0.710	0.010	-3.273	-0.476
cg11797365	11.904	5.326	0.025	2.057	23.117
Data Set [1-0]	-1.505	0.450	0.001	-2.417	-0.645
age	0.030	0.011	0.007	0.008	0.052
Sex [Female]	0.294	0.158	0.062	-0.013	0.606
z-value (COL4A3)	-0.350	0.178	0.050	-0.716	-0.015

Data set [1-0]: 0- Nicodemus Johnson cohort; 1- BEC cohort