

# Transcriptome Sequencing (RNA-Seq) of Human Endobronchial Biopsies: Asthma *versus* Controls

Ching Yong Yick, MD<sup>a</sup>, Aeilko H Zwinderman, MD, PhD<sup>b</sup>, Peter W Kunst, MD, PhD<sup>c</sup>, Katrien Grünberg, MD, PhD<sup>d</sup>, Thais Mauad, MD, PhD<sup>e</sup>, Annemiek Dijkhuis, MSc<sup>f</sup>, Elisabeth H Bel, MD, PhD<sup>a</sup>, Frank Baas, MD, PhD<sup>g</sup>, René Lutter, PhD<sup>a,f</sup>, and Peter J Sterk, MD, PhD<sup>a</sup>.

<sup>a</sup> Department of Respiratory Medicine, Academic Medical Centre, Amsterdam, The Netherlands

<sup>b</sup> Department of Epidemiology and Bioinformatics, Academic Medical Centre, Amsterdam, The Netherlands

<sup>c</sup> Department of Respiratory Medicine, Admiraal De Ruyter Hospital, Goes, The Netherlands

<sup>d</sup> Department of Pathology, VU Medical Centre, Amsterdam, The Netherlands

<sup>e</sup> Department of Pathology, São Paulo University Medical School (USP), São Paulo, Brazil

<sup>f</sup> Department of Experimental Immunology, Academic Medical Centre, Amsterdam, The Netherlands

<sup>g</sup> Department of Genome Analysis, Academic Medical Centre, Amsterdam, The Netherlands

## Corresponding author

Ching Yong Yick, MD

Department of Respiratory Medicine (F5-260)

Academic Medical Centre

Meibergdreef 9

1105 AZ, Amsterdam, The Netherlands

Telephone: +31-20-5664359

Fax: +31-20-5669001

E-mail: c.y.yick@amc.uva.nl

Running title:

Gene expression of bronchial biopsies

Declaration of all sources of funding:

This study was supported by a research-grant from the Netherlands Asthma Foundation (project number 3.2.09.065). Thais Mauad is funded by CNPQ (Brazilian Research Council).

The authors declare that they have no competing interests.

Word count (Introduction to Conclusion): 3157

## **ABSTRACT**

### **Background:**

The cellular and molecular pathways in asthma are highly complex. Increased understanding can be obtained by unbiased transcriptomic analysis (RNA-Seq). We hypothesized that the transcriptomic profile of whole human endobronchial biopsies differs between asthma patients and controls.

### **Objectives:**

First, we investigated the feasibility to obtain RNA from whole endobronchial biopsies suitable for RNA-Seq. Second, we examined the difference in transcriptomic profiles between asthma and controls.

### **Methods:**

This cross-sectional study compared 4 steroid-free atopic asthma patients and 5 healthy non-atopic controls. Total RNA of 4 biopsies per subject was prepared for RNA-Seq. Comparison of the numbers of reads per gene in asthma and controls was based on the Poisson distribution.

### **Results:**

Forty-six genes were differentially expressed between asthma and controls, including pendrin, periostin, and BCL2. Ten gene networks were found involved in cellular morphology, movement, and development.

### **Conclusion:**

RNA isolated from whole human endobronchial biopsies is suitable for RNA-Seq, showing different transcriptomic profiles between asthma and controls. Novel and confirmative genes were found to be linked to asthma. These results indicate that biological processes in the airways of asthma patients are differently regulated compared to controls, which may be relevant for the pathogenesis and treatment of the disease.

## **KEY WORDS**

Gene expression analysis, transcriptomics

**ABBREVIATIONS**

**ASM:** airway smooth muscle

**FEV<sub>1</sub>:** forced expiratory volume in 1 second

**FVC:** forced vital capacity

**GINA:** Global Initiative for Asthma

**IPA:** Ingenuity Pathway Analysis

**PC<sub>20</sub>:** provocative concentration of methacholine causing a 20 % drop in FEV<sub>1</sub>

## INTRODUCTION

Asthma is associated with a variety of functional changes of the airways including variable airways obstruction [1], increased sensitivity and elevated maximal responses to inhaled irritants [2], and impairment of bronchodilation following deep inspiration [3]. In addition, a process of structural changes called airway remodelling is usually observed in asthma. This includes an increase in airway smooth muscle (ASM) mass, altered deposition of extracellular matrix proteins in- and outside the ASM layer, goblet cell and glandular hyperplasia, and angiogenesis [4]. Therefore, the functional and structural characteristics of the airways in asthma are highly complex, which impedes development of novel, targeted therapy [5].

Disease activity in asthma appears to be associated with the pattern and activity of airways inflammation. It has been shown that increased expression of the Th2-cytokines IL-4 and IL-13 not only promotes eosinophilic inflammation, but also activates mast cells [6] and that microlocalisation of these mast cells predominantly in the ASM layer is a key factor in the pathophysiology in asthma [7]. Activated mast cells release various mediators, which can lead to enhanced ASM contraction and proliferation [6,8]. Interestingly, gene expression analysis has shown that the degree of Th2-cytokine expression in bronchial biopsies varies between asthma patients, being associated with aspects of airway remodelling and with responsiveness to inhaled steroids [9]. This suggests that there is a close interaction between inflammation and remodelling in asthma.

Airway remodelling is not necessarily a chronic repair response to ongoing inflammation as traditionally hypothesized. Recent evidence shows that airway remodelling may be driven by pathways that are partly independent of airway inflammation [10]. In addition, airway structural components themselves may drive the pathophysiologic process of asthma [11]. Given this complexity of the cellular and molecular pathways, increased understanding of the

driving mechanisms in asthma might be best derived from unbiased analysis of bronchial tissue specimens in relation to careful clinical phenotyping [12].

Transcriptomic analysis of the airways by next-generation sequencing (RNA-Seq) allows high-throughput and detailed characterization of gene expression profiles at the tissue level [13]. Particularly in complex disorders such as asthma this technology has the potential to discover gene expression profiles that are characteristic of the disease. To our knowledge, there are no head-to-head comparisons of RNA-Seq profiles from bronchial biopsies between asthma and controls.

We hypothesized that the transcriptomic profile of whole endobronchial biopsies is different between patients with asthma and healthy controls. Therefore, the aim of the present study was: a) to investigate the feasibility to obtain RNA from whole human endobronchial biopsies suitable for RNA-Seq, and b) to examine the difference in transcriptomic profiles of whole human endobronchial biopsies between atopic mild, steroid-free asthma patients and healthy non-atopic controls.

## METHODS

Additional information regarding the study methods is included in the Online Data Supplement.

### *Design*

This cross-sectional study consisted of 2 visits. At visit 1, subjects were screened for eligibility to participate according to the in- and exclusion criteria. This was followed by lung function measurements including spirometry, FEV<sub>1</sub> reversibility, and methacholine bronchoprovocation test. At visit 2, a bronchoscopy was performed with the collection of endobronchial biopsies.

### *Subjects*

The study population included steroid-free patients with atopic asthma (n=4) and healthy non-atopic controls (n=5). Asthma patients had controlled disease according to GINA guidelines [14] and had no exacerbations within 6 weeks prior to participation. Control subjects had no history or symptoms of pulmonary diseases. This study was part of a larger study registered at the Netherlands Trial Register (The role of the airway smooth muscle layer in asthma; NTR1306).

### *RNA-Seq and data analysis*

Four endobronchial biopsies per subject were collected and frozen after overnight incubation in RNAlater (Qiagen, Venlo, The Netherlands). TRIzol (Invitrogen, Carlsbad, CA, US) was used to isolate RNA. Amplified cDNA was prepared with the Ovation RNA-Seq System (NuGEN, San Carlos, CA, US) and cDNA libraries were constructed using SPRIworks Fragment Library System II (Beckman-Coulter, Brea, CA, US). Emulsion-based clonal amplification was performed using the GS FLX Titanium emPCR Kit Lib-L (Roche, Penzberg, Germany) to prepare enriched DNA library beads for sequencing on the GS FLX+ System (454/Roche). A schematic overview of the strict workflow is presented in Figure 1.

Sequencing reads were mapped against the human genome (hg19) [15] using the GS Reference Mapper (Roche). Ingenuity Pathway Analysis (IPA; Redwood City, CA, US) was used to identify gene networks. IPA generated a network score, which gives the likelihood that the set of genes in this network could be explained by chance alone. Networks with a score  $\geq 2$  have at least 99% confidence that it is not generated by chance. Quantitative PCR (qPCR) was performed to validate the sequencing data. RNA-Seq data was uploaded to the Gene Expression Omnibus (GSE38994). To validate the sequencing data, quantitative PCR (qPCR) was performed on selected genes that were differentially expressed between asthma and controls.

#### *Statistical analysis*

Subject characteristics of the study groups were compared using unpaired t-tests or Mann-Whitney U tests. Comparison of the numbers of reads per gene between asthma and controls was based on the Poisson distribution with correction for total number of reads. Cohen's effect size was calculated to quantify the size of the difference in read numbers of individual genes between asthma and controls. Bonferroni correction for multiple testing was applied. A p-value of  $<0.05$  was considered statistically significant.

#### *Sample size and power calculation*

We conservatively assumed that all pathways that are associated with asthma consist of  $\geq 10,000$  genes [16], which represents approximately a third of the 31,227 genes of the human genome (hg19) tested. The average fold change of the gene expressions of asthma *versus* controls was presumed to be about 1.2, which coincides with a Cohen's effect size of about 0.5. A significance level of 0.0001 was used in the statistical analysis to correct for multiple testing. With this significance level and 9 subjects in total with 4 biopsy specimens per subject, we expected to find 48 genes that are truly associated with asthma and 2 falsely significant genes [17,18]. This resulted in a false discovery rate of about 4%.



## RESULTS

### *Subjects*

Table 1 shows the subject characteristics of the asthma patients and healthy controls. Lung function characteristics were largely comparable between the 2 groups. As expected, PC<sub>20</sub> was lower in asthma as compared to controls.

### *Endobronchial biopsy specimens*

From all study participants 4 endobronchial biopsies per subject of approximately 1-2mm<sup>3</sup> were successfully collected. The amount of total RNA isolated from 4 endobronchial biopsy specimens per study subject ranged from 900-9,300ng.

### *RNA-Seq*

Amplified cDNA was prepared from 50ng of isolated RNA, and had an average length of 250bp with an amount ranging from 6,120-9,630ng. cDNA libraries had an average length of 700bp. The amount of cDNA molecules ranged from 0.25-1.58x10<sup>9</sup> molecules/μL. DNA beads with 16-24% enrichment after clonal amplification of the cDNA libraries using emulsion PCR were used for sequencing.

The median read length obtained with the sequencing run was 199 nucleotides with a maximum of 867 nucleotides. Total read numbers of 349,239 and 399,999 were obtained for asthma and controls, respectively. For asthma 87% of the reads could be mapped to the human genome hg19, whilst this was 88% for controls.

The transcriptomes mapped to 10,167 unique genes for asthma and 11,006 for controls (Figure 2). With a medium contig read length of 261bp, the coverage in the current study is approximately 24x [19]. After correction for multiple testing, 46 genes with an effect size of 0.5 were differentially expressed between asthma and controls (Table 2). No significant differences in reference genes were found, including GAPDH, ACTB, B2M, and LDHA.

### *Gene network identification*

The 46 differentially expressed genes between asthma and controls were used to identify gene networks by the IPA software application. Ten gene networks were identified with a network score of at least 2. The highest scoring network had a network score of 32 and was associated with the network functions cellular movement, cell death, and cellular morphology. The second and third ranking networks scoring 21 and 17 were associated with genetic disorder, and cellular movement and development, respectively (Figure 3). The gene B-cell CLL/lymphoma 2 (BCL2), which was one of the 46 differentially expressed genes between asthma and controls, appeared to be a key component in the network with the highest network score (Figure 3a). Other key components in this network were mitogen-activated protein kinase 1 (MAPK 1) also known as ERK 1/2, nuclear factor kappa-B (NF- $\kappa$ B), p38 mitogen-activated protein kinase (p38 MAPK), and transforming growth factor  $\beta$  (TGF- $\beta$ ). Fourteen of the 46 input genes could be located in this network. The second ranking network included signal transducer and activator of transcription 3 (STAT3), periostin (POSTN), and stauferin (STAU2) (Figure 3b), whilst the third network comprised epidermal growth factor receptor (EGFR) and pendrin (SLC26A4) (Figure 3c).

### *Validation of sequencing data*

Significant correlations were found between the number of sequence reads per gene and Cq-ratio for WARS ( $r = -0.652$ ;  $p = 0.05$ ), SCGB1A1 ( $r = -0.671$ ;  $p = 0.05$ ), and BCL2 ( $r = 0.724$ ;  $p = 0.04$ ). These results show that a high sequencing read number corresponded with a low Cq-ratio (high amount of cDNA) and vice versa. Therefore, the qPCR results confirmed the data found by RNA-Seq.

## DISCUSSION

In this study RNA isolated from whole human endobronchial biopsy specimens was successfully sequenced, showing transcriptomic profiles of bronchial tissue that were different between asthma and controls. Therefore, RNA isolated from endobronchial biopsies is suitable for RNA-Seq when processed according to a strict workflow. Analysis of the RNA-Seq data revealed expression of novel genes that have not been linked to asthma and airway inflammation previously. Additionally, gene network identification showed that several of the 46 differentially expressed genes between asthma and controls are key components in gene networks associated with the regulation of one or multiple cellular functions, including cell morphology, movement, and development. These findings indicate that the regulation of biological processes in the airways is evidently different between asthma and healthy controls, which may be relevant for the pathogenesis and future treatment of the disease.

To our knowledge this is the first study to apply RNA-Seq instead of microarrays to examine the transcriptomic profiles of *in vivo* human endobronchial biopsies of asthma patients and healthy controls. This high-throughput next-generation sequencing method enables an unbiased way of gene expression analysis due to the fact that it is not limited to a selection of a number of known genes or sequences as is the case with microarrays [13]. Therefore, next-generation sequencing facilitates the discovery of *de novo* transcripts and identification of novel disease-related genes.

We have found 46 genes in bronchial biopsies that were differentially expressed between asthma and controls. Several of these genes have been shown to exert various effects on mRNA degradation and translational processes affecting cellular functions, such as staußen (STAU2) [20] and tryptophanyl-tRNA synthetase (WARS) [21]. However, the majority of the differentially expressed genes has not been linked to asthma yet. Previous studies using microarrays have shown that the gene expression profile of airway epithelial cells is different in asthma as compared to healthy controls [22,23]. This included SLC26A4, also known as

pendrin, which in our study too was found to be one of the most prominently upregulated genes in asthma. It has been shown that this gene is closely associated with the pathophysiology of asthma [24]. More specifically, stimulation of the IL4/IL13 receptor complex leads to an increased expression of SLC26A4 via the transcription factor STAT6 resulting in increased mucus production and viscosity of the airway surface fluid [25]. Additionally, it has been shown previously that periostin (POSTN), chloride channel regulator 1 (CLCA1), and serpin peptidase inhibitor, clade B, member 2 (SERPINB2) were highly induced in airway epithelial cells [9,22]. Our current study extends these findings, showing that the expression of these three genes was higher in asthma compared to controls, although only periostin reached significance (Table 2). This could be due to the fact that we have sequenced the transcriptome of whole bronchial biopsies instead of airway epithelial cells only.

Additionally, input of the 46 differentially expressed genes into the IPA database yielded 10 gene networks that were related to regulatory functions on the cellular level. Besides the fact that several of the 46 genes take key positions in these networks, it should also be noted that particular cytokines, chemokines, and protein complexes that are well-known for their major roles in inflammatory processes, such as TGF- $\beta$ , IFN- $\alpha$ , p38 MAPK, and NF- $\kappa$ B, are part of these same networks [26,27]. In addition to p38 MAPK and NF- $\kappa$ B, two other major nodes present in the highest IPA scoring gene network are BCL2, and ERK1/2. It has been hypothesized that inhibition of BCL2 leads to an upregulated immune response accompanied with an enhanced release of pro-inflammatory cytokines and enhanced efficacy of necrosis of cells [28]. Our results show a significantly decreased expression of BCL2 in asthma compared to controls. Furthermore, phosphorylation of ERK1/2 has been shown to be increased by RANTES and eotaxin, whereas inhibition of ERK1/2 reduced the level of ASM proliferation induced by chemokines [29]. The second and third highest IPA scoring gene networks (Figure 3b, 3c) contain STAT3 and EGFR amongst others. STAT3 has been shown to be involved as a transcriptional mechanism for thymic stromal lymphopoietin to induce a

pro-inflammatory gene expression in ASM cells [30], whereas blocking EGFR may lead to a reduction in mucus hypersecretion [31]. Taken together, these results indicate that the regulation of biological processes in the airways is evidently different in asthma compared to controls. Furthermore, these three gene networks may give further insight into the axis from pathology and inflammation (p38 MPK, NF- $\kappa$ B, BCL2, and EGFR) to physiology (ERK1/2, and STAT3) of the airways in asthma.

Several methodological issues need to be addressed. The inclusion of well-characterized subjects seems to represent a strength of our study. Only steroid-free asthma patients with controlled disease were included, thereby avoiding the potential bias of steroid usage on the gene expression profile of the airway wall. All asthma patients remained stable and controlled with minimal clinical symptoms during the study period. Although they were allowed to use inhaled short-acting  $\beta$ 2-agonists as rescue therapy, none of these patients used any during the study period. Therefore, the gene expression profiles could not be potentially influenced by medication usage.

We designed a strict standard operating procedure for the workup of the endobronchial biopsies. Furthermore, by combining the Ovation RNA-Seq System and GS FLX+ we could maximize the utility of the potentially low yield of and even partially degraded RNA from *in vivo* obtained human endobronchial biopsies. First, for the successful processing of RNA into amplified cDNA, the required input RNA for the Ovation RNA-Seq System is  $\geq 500$ pg, which was certainly appropriate for the currently obtained amount of isolated RNA (900-9,300ng). Second, amplification is initiated both at the 3'-end and randomly throughout the transcriptome, thereby maximizing the coverage as reads are distributed throughout the whole transcriptome, as well as minimizing the potentially detrimental effects of degraded RNA on gene expression analysis. Finally, our sample size was based on detailed power calculations showing an acceptable false discovery rate of 4%.

Gene network analyses by IPA are based on known pathways. The differentially expressed genes found in the current study may be part of novel pathways that are not yet included in the IPA database even though it is updated weekly with findings from other databases including Entrez Gene, RefSeq, OMIM, and Gene Ontology. We used IPA, because we aimed to clarify the association of the gene expression profile found in asthma patients with the pathophysiology of asthma. As the field of gene sequencing is advancing our knowledge of various biochemical processes, novel pathways that are associated with asthma may be discovered in future studies.

Nevertheless, there are some limitations in our study. First, the amplification by the Ovation RNA-Seq System could introduce bias due to the generation of short cDNA fragments. This may potentially result in less representative consensus sequences to be found and used for mapping against the reference genome. Consequently, fewer genes than actually present in the samples could be detected. However, it is unlikely that this affected our results, because we generated high coverage depth and the use of the GS FLX+ System ensures longer sequence reads than competing technologies.

Second, the biopsy specimens were taken from RB 4-6. As the degree of airway wall remodelling and inflammation in asthma can vary with location in the tracheobronchial tree and with asthma severity [32], it is possible that gene expression profiles of the peripheral airways and in patients with more severe asthma could turn out to be different from those observed in the present study. Subsequent studies focusing on these locations of the airways in asthma patients with more severe disease are needed to complement our current data in order to provide a more comprehensive overview of the gene expression of the asthmatic airways. In addition, temporal variability should be examined, which will be restricted by ethical constraints.

Third, RNA was isolated from whole endobronchial biopsies. We chose to perform whole biopsy analysis, because the contribution of the individual components of the airway wall to the pathophysiology of asthma is still largely unknown [33]. Furthermore, the 46 differentially expressed genes between asthma and controls did not include transcripts of specific inflammatory cells but of various cellular processes instead, indicating that RNA-Seq of whole biopsies was not affected by a difference in inflammatory cells present in the biopsy material.

Based on the currently observed differences in gene expression profiles of whole biopsies it seems warranted to repeat the analysis in selected individual airway components e.g. epithelium and ASM by laser capture microdissection. This will allow linking gene expression profiles of individual airway wall components to airway function and disease activity. High-throughput next-generation sequencing has the potential to achieve this aim. It could not only lead to the discovery of novel genes or pathways that are vital to the pathogenesis of asthma, but may also qualify as a tool for subphenotyping patients in relation to clinical outcome or treatment responsiveness. After adequate external validation in newly recruited patients, new therapeutic targets could arise, which may facilitate developing essentially novel asthma therapies. In addition, by implementing a systems biology approach in future asthma studies [12], the entire biochemical process from translation of DNA to RNA to proteins, to the post-translational modification and regulation of these proteins will be unraveled through the combination of different 'omics' technologies including genomics, transcriptomics, proteomics, and metabolomics.

## **CONCLUSION**

The current study has shown that RNA isolated from whole human endobronchial biopsies was suitable for RNA-Seq. Transcriptomic analysis of whole biopsies revealed 46 genes that were differentially expressed between asthma and controls with the majority not linked to asthma previously. Furthermore, pathway analysis showed that several of the 46

differentially expressed genes were key components in gene networks associated with the regulation of one or multiple cellular functions. These results indicate that biological processes in the airways of asthma patients are differently regulated at the transcriptomic level when compared to healthy controls.



**ACKNOWLEDGEMENTS**

We thank Saheli Chowdhury of the department of Experimental Immunology, Academic Medical Centre, Amsterdam, The Netherlands for the support in designing the primer sets for the quantitative PCRs.

## REFERENCES

1. Reddel H, Jenkins C, Woolcock A. Diurnal variability – time to change asthma guidelines? *BMJ* 1999; 319: 45-47.
2. Sterk PJ, Fabbri LM, Quanjer PhH, Cockcroft DW, O'Byrne PM, Anderson SD, Juniper EF, Malo JL. Airway responsiveness. Standardized challenge testing with pharmacological, physical and sensitizing stimuli in adults. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl* 1993; 16: 53-83.
3. Slat AM, Janssen K, van Schadowijk A, van der Plas DT, Schot R, van den Aardweg JG, de Jongste JC, Hiemstra PS, Mauad T, Rabe KF, Sterk PJ. Bronchial inflammation and airway responses to deep inspiration in asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007; 176: 121-128.
4. Fixman ED, Stewart A, Martin JG. Basic mechanisms of development of airway structural changes in asthma. *Eur Respir J* 2007; 29: 379-389.
5. Mauad T, Bel EH, Sterk PJ. Asthma therapy and airway remodelling. *J Allergy Clin Immunol* 2007; 120: 997-1009.
6. Brightling CE, Bradding P, Pavord ID, Wardlaw AJ. New insights into the role of the mast cell in asthma. *Clin Exp Allergy* 2003; 33: 550-556.
7. Siddiqui S, Hollins F, Saha S, Brightling CE. Inflammatory cell microlocalisation and airway dysfunction: cause and effect? *Eur Respir J* 2007; 30: 1043-1056.
8. Brown JM, Wilson TM, Metcalfe DD. The mast cell and allergic diseases: role in pathogenesis and implications for therapy. *Clin Exp Allergy* 2008; 38: 4-18.
9. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. T-helper type 2-driven inflammation defines major subphenotypes of asthma. *Am J Respir Crit Care Med* 2009; 180: 388-395.

10. Holgate ST, Roberts G, Arshad HS, Howarth PH, Davies DE. The role of the airway epithelium and its interaction with environmental factors in asthma pathogenesis. *Proc Am Thorac Soc* 2009; 6: 655-659.
11. Kaur D, Saunders R, Hollins F, Woodman L, Doe C, Siddiqui S, Bradding P, Brightling C. Mast cell fibroblastoid differentiation mediated by airway smooth muscle in asthma. *J Immunol* 2010; 185: 6105-6114.
12. Auffray C, Adcock IM, Chung KF, Djukanovic R, Pison C, Sterk PJ. An integrative systems biology approach to understanding pulmonary diseases. *Chest* 2010; 137: 1410-1416.
13. Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010; 11: 31-46.
14. GINA report, Global strategy for asthma management and prevention (update December 2009). <http://www.ginasthma.com>
15. UCSC Genome Bioinformatics. <http://genome.ucsc.edu>
16. Loza MJ, McCall CE, Li L, Isaacs WB, Xu J, Chang BL. Assembly of inflammation-related genes for pathway-focused genetic analysis. *PLOS ONE* 2007; 2: e1035.
17. Ferreira JA, Zwinderman AH. On the Benjamini-Hochberg method. *Ann Statist* 2006; 34: 1827-1849.
18. Ferreira JA, Zwinderman AH. Approximate sample size calculations with microarray data: an illustration. *Stat Appl Genet Mol Biol* 2006; 5: Article25.
19. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 1988; 2: 231-239.
20. Villacé P, Marión RM, Ortin J. The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. *Nucleic Acids Res* 2004; 32: 2411-2420.
21. Wakasugi K, Nakano T, Morishima I. Oxidative stress-responsive intracellular regulation specific for the angiostatic form of human tryptophanyl-tRNA synthetase. *Biochemistry* 2005; 44: 225-232.

22. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C, Erle DJ, Yamamoto KR, Fahy JV. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. *Proc Natl Acad Sci U S A* 2007; 104: 15858-15863.
23. Laprise C, Sladek R, Ponton A, Bernier MC, Hudson TJ, Laviolette M. Functional classes of bronchial mucosa genes that are differentially expressed in asthma. *BMC Genomics* 2004; 5: 21.
24. Nakao I, Kanaji S, Ohta S, Matsushita H, Arima K, Yuyama N, Yamaya M, Nakayama K, Kubo H, Watanabe M, Sagara H, Sugiyama K, Tanaka H, Toda S, Hayashi H, Inoue H, Hoshino T, Shiraki A, Inoue M, Suzuki K, Aizawa H, Okinami S, Nagai H, Hasegawa M, Fukuda T, Green ED, Izuhara K. Identification of pendrin as a common mediator for mucus production in bronchial asthma and chronic obstructive pulmonary disease. *J Immunol* 2008; 180: 6262-6269.
25. Nofziger C, Vezzoli V, Dossena S, Schönherr T, Studnicka J, Nofziger J, Vanoni S, Stephan S, Silva ME, Meyer G, Paulmichl M. STAT6 links IL-4/IL-13 stimulation with pendrin expression in asthma and chronic obstructive pulmonary disease. *Clin Pharmacol Ther* 2011; 90: 399-405.
26. Chung KF. p38 mitogen-activated protein kinase pathways in asthma and COPD. *Chest* 2011; 139: 1470-1479.
27. Clarke D, Damera G, Sukkar MB, Tliba O. Transcriptional regulation of cytokine function in airway smooth muscle cells. *Pulm Pharmacol Ther* 2009; 22: 436-445.
28. Sasi N, Hwang M, Jaboin J, Csiki I, Lu B. Regulated cell death pathways: new twists in modulation of BCL2 family function. *Mol Cancer Ther* 2009; 8: 1421-1429.
29. Halwani R, Al-Abri J, Beland M, Al-Jahdali H, Halayko AJ, Lee TH, Al-Muhsen S, Hamid Q. CC and CXC chemokines induce airway smooth muscle proliferation and survival. *J Immunol* 2011; 186: 4156-4163.

30. Shan L, Redhu NS, Saleh A, Halayko AJ, Chakir J, Gounni AS. Thymic stromal lymphopoietin receptor-mediated IL-6 and CC/CXC chemokines expression in human airway smooth muscle cells: role of MAPKs (ERK1/2, p38, and JNK) and STAT3 pathways. *J Immunol* 2010; 184: 7134-7143.
31. Yang J, Li Q, Zhou XD, Kolosov VP, Perelman JM. Naringenin attenuates mucous hypersecretion by modulating reactive oxygen species production and inhibiting NF- $\kappa$ B activity via EGFR-PI3K-Akt/ERK MAPKinase signaling in human airway epithelial cells. *Mol Cell Biochem* 2011; 351: 29-40.
32. Araujo BB, Dolhnikoff M, Silva LF, Elliot J, Lindeman JH, Ferreira DS, Mulder A, Gomes HA, Fernezlian SM, James A, Mauad T. Extracellular matrix components and regulators in the airway smooth muscle in asthma. *Eur Respir J* 2008; 32: 61-69.
33. Yick CY, Ferreira DS, Annoni R, von der Thüsen JH, Kunst PW, Bel EH, Lutter R, Mauad T, Sterk PJ. Extracellular matrix in airway smooth muscle is associated with dynamics of airway function in asthma. *Allergy* 2012; 67: 552-559.

## TABLES

Table 1 – Subject characteristics of asthma and healthy controls

	Asthma	Healthy
Subjects (n)	4	5
Age (years) <sup>*</sup>	25 (21-27)	25 (21-30)
FEV <sub>1</sub> post-bronchodilator (%pred.) <sup>†</sup>	98 (13)	112 (7)
FEV <sub>1</sub> /FVC post-bronchodilator <sup>†</sup>	0.86 (0.07)	0.87 (0.11)
PC <sub>20</sub> (mg/mL) <sup>‡</sup>	0.91 (0.07-12)	> 8

<sup>\*</sup> Mean (min-max)

<sup>†</sup> Mean (SD)

<sup>‡</sup> Geometric mean (95% CI)

FEV<sub>1</sub>: forced expiratory volume in 1 second; FVC: forced vital capacity; PC<sub>20</sub>: provocative concentration of methacholine causing a 20 % drop in FEV<sub>1</sub>; %pred.: % of predicted

**Table 2 – List of differentially expressed genes asthma vs. controls**

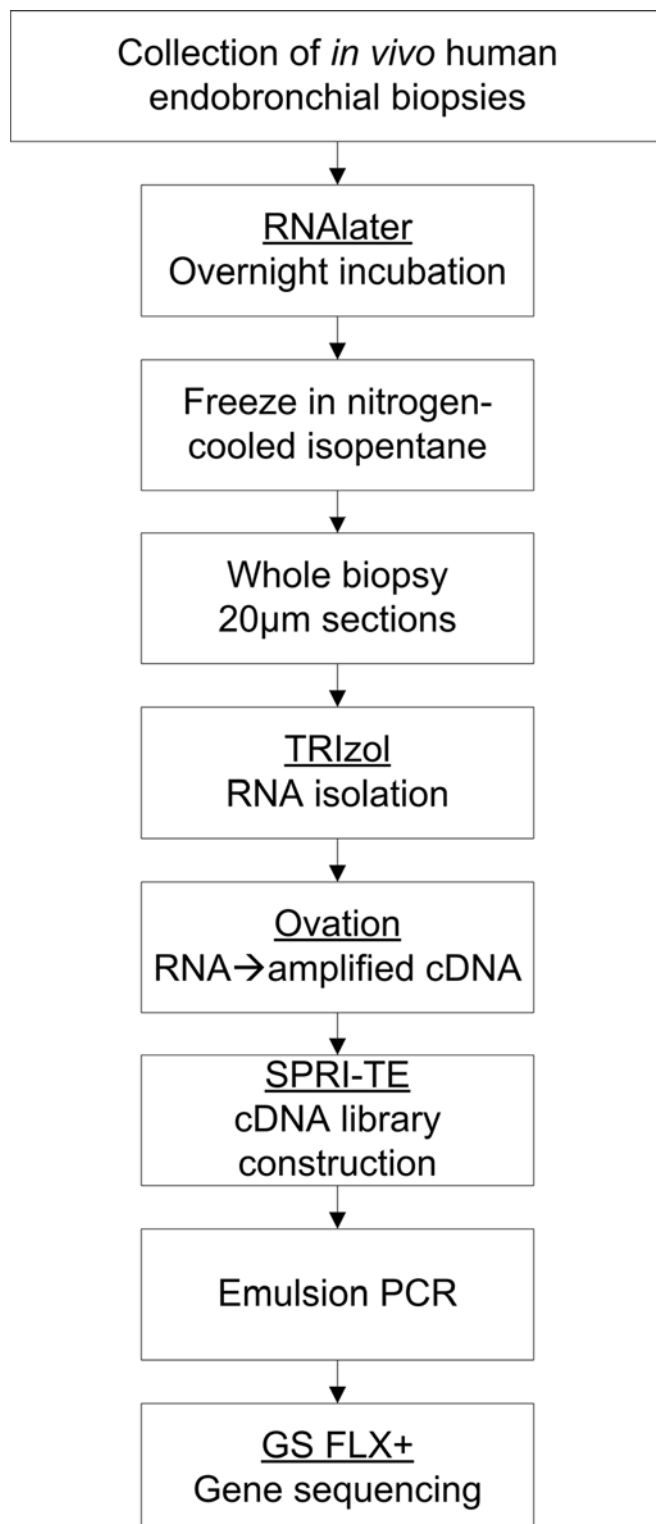
Gene symbol	Gene name	NCBI Gene ID	Read number		In fold change*	p-value†
			Asthma	Control		
SLC26A4	Solute carrier family 26, member 4	5172	24	1	3.40	0.04
DCP1B	DCP1 decapping enzyme homolog B ( <i>S. cerevisiae</i> )	196513	41	3	2.84	<0.01
FKTN	Fukutin	2218	38	6	2.07	<0.01
C9orf84	Chromosome 9 open reading frame 84	158401	42	8	1.88	<0.01
CD96	CD96 molecule	10225	46	10	1.75	<0.01
LOC643650	<i>Uncharacterized</i>	643650	53	12	1.71	<0.01
POSTN	Periostin, osteoblast specific factor	10631	85	20	1.67	<0.01
APCDD1	Adenomatosis polyposis coli down-regulated 1	147495	38	9	1.66	0.04
FAM83B	Family with sequence similarity 83, member B	222584	43	12	1.50	0.03
UBP1	Upstream binding protein 1 (LBP-1a)	7342	43	12	1.50	0.03
WARS	Tryptophanyl-tRNA synthetase	7453	48	14	1.46	0.01
ITGAV	Integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)	3685	71	21	1.44	<0.01
DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	1665	57	18	1.38	<0.01
ACAP2	ArfGAP with coiled-coil, ankyrin repeat and PH domains 2	23527	58	19	1.34	<0.01
HBS1L	HBS1-like ( <i>S. cerevisiae</i> )	10767	51	17	1.32	0.02
LRIG3	Leucine-rich repeats and immunoglobulin-like domains 3	121227	84	32	1.19	<0.01
SCGB1A1	Secretoglobin, family 1A, member 1 (uteroglobin)	7356	353	152	1.07	<0.01
ADARB2	Adenosine deaminase, RNA-specific, B2	105	129	58	1.02	<0.01
STRBP	Spermatid perinuclear RNA binding protein	55342	215	97	1.02	<0.01
DMXL1	Dmx-like 1	1657	77	36	0.98	0.02
STAU2	Staufen, RNA binding protein, homolog 2 ( <i>Drosophila</i> )	27067	89	42	0.97	<0.01
PARP14	Poly (ADP-ribose) polymerase family, member 14	54625	118	62	0.87	<0.01
FGF13	Fibroblast growth factor 13	2258	98	56	0.78	0.05

GCC2	GRIP and coiled-coil domain containing 2	9648	134	77	0.78	<0.01
BICC1	Bicaudal C homolog 1 (Drosophila)	80114	129	80	0.70	0.02
SPEF2	Sperm flagellar 2	79925	176	123	0.58	0.01
PTPRT	Protein tyrosine phosphatase, receptor type, T	11122	268	203	0.50	<0.01
MBNL1	Muscleblind-like (Drosophila)	4154	308	241	0.47	<0.01
HFM1	ATP-dependent DNA helicase homolog (S. cerevisiae)	164045	491	853	-0.33	<0.01
BCL2	B-cell CLL/lymphoma 2	596	236	462	-0.45	<0.01
NOTCH2NL	Notch 2 N-terminal like	388677	305	605	-0.46	<0.01
PDE3A	Phosphodiesterase 3A, cGMP-inhibited	5139	249	530	-0.53	<0.01
ITPR2	Inositol 1,4,5-trisphosphate receptor, type 2	3709	33	99	-0.88	0.03
LYZ	Lysozyme	4069	34	106	-0.91	0.01
RNF180	Ring finger protein 180	285671	13	57	-1.25	0.04
LAMC1	Laminin, gamma 1 (formerly LAMB2)	3915	12	59	-1.37	0.01
SLC9A8	Solute carrier family 9 (sodium/hydrogen exchanger), member 8	23315	51	266	-1.43	<0.01
NUS1	Nuclear undecaprenyl pyrophosphate synthase 1 homolog (S. cerevisiae)	116150	6	41	-1.70	0.03
CTR9	Ctr9, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	9646	5	37	-1.78	0.05
RNASET2	Ribonuclease T2	8635	1	25	-3.00	0.01
CUBN	Cubilin (intrinsic factor-cobalamin receptor)	8029	1	28	-3.11	<0.01
WDR17	WD repeat domain 17	116966	1	28	-3.11	0.01
HBB	Hemoglobin, beta	3043	1	29	-3.14	0.01
DDB2	Damage-specific DNA binding protein 2, 48kDa	1643	1	33	-3.27	<0.01
DYX1C1	Dyslexia susceptibility 1 candidate 1	161582	1	33	-3.27	<0.01
EFCAB3	EF-hand calcium binding domain 3	146779	1	33	-3.27	<0.01

<sup>†</sup> Natural logarithm fold change asthma vs. controls: positive number signifies upregulated in asthma, negative number signifies downregulated in asthma

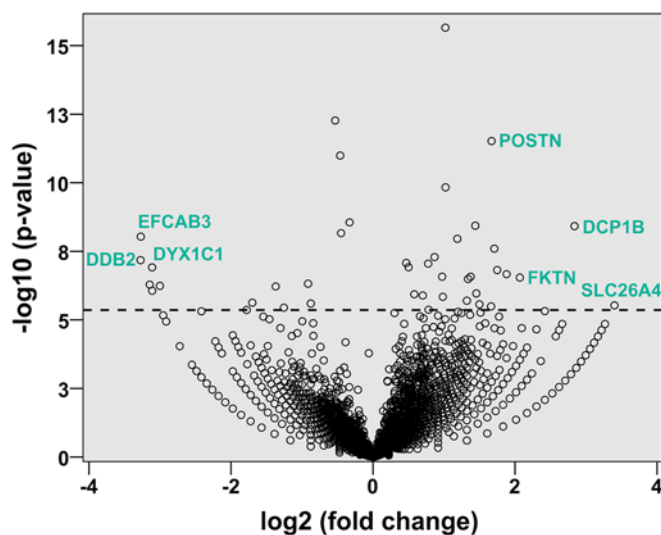
<sup>†</sup> p-value after correction for multiple testing



**FIGURES****Figure 1 – Workflow biopsy specimens**

**Figure 2 – Volcano-plot of genes asthma vs. controls**

The logarithms of the fold changes of individual genes (x-axis) are plotted against the negative logarithm of their p-value to base 10 (y-axis). Positive log<sub>2</sub> (fold change)-values represents upregulation in asthma compared to healthy controls, negative values downregulation. Circles above the dotted line represent differentially expressed genes between asthma and controls with  $p < 0.05$  after correction for multiple testing. DCP1B: DCP1 decapping enzyme homolog B (*S. cerevisiae*); DDB2: damage-specific DNA binding protein 2, 48kDa; DYX1C1: dyslexia susceptibility 1 candidate 1; EFCAB3: EF-hand calcium binding domain 3; FKTN: fukutin; POSTN: periostin; SLC26A4: pendrin.



### Figure 3 – Top 3 gene networks by IPA

The network with the highest network score of 32 (a) including BCL2, MAPK1, NF- $\kappa$ B, p38 MAPK, and TGF- $\beta$  was associated with the network functions cellular movement, cell death, and cellular morphology. The second (b) highest scoring network including STAT3, POSTN, and STAU2, and third (c) network including EGFR, and SLC26A4 were associated with genetic disorder, and cellular movement and development, respectively. The intensity of red- or green-colored shapes indicates the degree of up- or downregulation, respectively.

