Supplementary material

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

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

[a] Department of Respiratory Medicine, Academic Medical Centre, Amsterdam, The Netherlands

[b] Department of Biostatistics, Academic Medical Centre, Amsterdam, The Netherlands

[c] Department of Respiratory Medicine, Admiraal De Ruyter Hospital, Goes, The Netherlands

[d] Department of Pathology, VU Medical Centre, Amsterdam, The Netherlands

[e] Department of Pathology, São Paulo University Medical School (USP), São Paulo, Brazil

[f] Department of Experimental Immunology, Academic Medical Centre, Amsterdam, The Netherlands

[g] Department of Genome Analysis, Academic Medical Centre, Amsterdam, The Netherlands

METHODS

Subjects

All subjects were recruited by the Department of Respiratory Medicine of the Academic Medical Centre Amsterdam and gave written informed consent prior to enrolment. The inclusion criteria for asthma patients were: aged 18-50 years; non-smoking or stopped >12 months with ≤5 pack years; episodic chest symptoms; no exacerbations within 6 weeks prior to participation; steroid-naïve or stopped using steroids by any dosing route ≥8 weeks prior to participation; no pulmonary diseases other than asthma; no pulmonary medication except inhaled short-acting β2-agonists as rescue therapy; airway hyperresponsiveness defined by a methacholine PC_{20} ≤8 mg/mL; post-bronchodilator FEV_1 >70% of predicted; atopy defined by a positive skin prick test. The inclusion criteria for healthy controls were similar, except that they had no chest symptoms, no pulmonary diseases, no airway hyperresponsiveness (methacholine PC_{20} >8 mg/mL), and a negative skin prick test. This study was approved by the local Medical Ethics Committee and registered at the Netherlands Trial Register (NTR1306).

Airway function and skin prick test

Spirometry was performed according to European Respiratory Society recommendations [1]. PC_{20} was measured by methacholine bronchoprovocation test using the standardized tidal volume method with a maximum methacholine dose of 16 mg/mL [2]. FEV_1 reversibility was determined after inhalation of 400µg Salbutamol. The skin prick test included 12 common aeroallergen extracts (ALK, Hørsholm, Denmark).

Bronchoscopy

Fibreoptic bronchoscopy was performed according to international recommendations [3] and a previously published study [4]. Local anaesthea was provided by Lignocaine spray in the

nose and throat prior to the bronchoscopy. During the procedure, additional Lignocaine solution was instilled into lung segments to dampen the cough reflex. A cup forceps (KW2411S; Pentax, Breda, The Netherlands) was used to collect 4 endobronchial biopsies per participant at RB 7-9. Special care was taken in positioning the forceps laterally to the bronchial carina in order to minimize the amount of connective tissue and at the same time maximize the yield of ASM in the biopsies.

Preparation of biopsies for RNA-Seq

The collected endobronchial biopsy specimens were put into RNAlater (Qiagen, Venlo, The Netherlands) for overnight incubation, embedded in Tissue-Tek (Sakura Finetek, Alphen aan den Rijn, The Netherlands) and frozen in liquid nitrogen-cooled isopentane. To facilitate the isolation of RNA and to ensure a maximum yield, the whole biopsy specimen was cut into 20µm sections in a RNase-free cryostat. All cryosections from the 4 frozen biopsies per participant were pooled and put directly in TRIzol (Invitrogen, Carlsbad, CA, USA) to isolate RNA. RNase-free glycogen (UltraPure Glycogen; Invitrogen) was added to the RNA isolation-protocol as a carrier to maximize the yield of RNA. Amplified cDNA was prepared with the Ovation RNA-Seq System (NuGEN, San Carlos, CA, USA) and purified using the Agencourt RNAClean XP kit (Beckman Coulter, Brea, CA, USA). GS FLX Titanium Rapid Library Multiplex Identifier (MID) adaptors (Roche, Penzberg, Germany) were added to allow multiplexing of samples. cDNA concentration was measured on the Qubit fluorometer (Invitrogen) using the dsDNA BR Assay Kit (Invitrogen). The amount of cDNA molecules was measured on the FLUOstar OPTIMA (BMG Labtech, Ortenberg, Germany).

RNA-Seq and sequence data analysis

Enriched DNA library beads were loaded on a 2-region PicoTiterPlate with 2 million beads per region (GS FLX Titanium Sequencing Kit XLR70; Roche, Penzberg, Germany). The GS FLX+ System (454/Roche) was used for RNA-Seq with the number of nucleotide cycles set at 200 [5]. Genes that were significantly up- or downregulated were uploaded to the web-

based Ingenuity Pathway Analysis (IPA) software application [6] for gene network analysis. The IPA network score for each gene network displayed as the negative log of the p-value of that specific network, gives the likelihood that the set of genes in this network could be explained by chance alone. Therefore, networks with a score ≥2 have at least 99% confidence that it is not generated by chance. Statistical analyses were performed using SPSS 18 (IBM Corporation, Armonk, New York, USA).

Sequencing reads were mapped against the human genome (hg19) [7] using the GS Reference Mapper (Roche). Multiplex Identifiers (MIDs) were specified using the RLMIDs scheme for Rapid Library preparation kit MID adaptors with the Number of Errors allowed set to 1. The expected depth was set at 0 and the minimum read length at 20bp. Duplicate reads were excluded from computation of the consensus for a contig and frequency statistics. A minimum overlap length of 40bp and minimum overlap identity of 90% were used for alignment. Mapping of the reads by GS Reference Mapper follows a sequence census approach [8], which is in its concept similar to the serial analysis of gene expression (SAGE) method. In short, each sequence read is aligned with the reference genome. However, a sequence read may align to multiple locations of the reference genome due to their short bp length compared to the reference. To identify the actual site of origin in the reference genome, the sequence read is put together with other sequence reads that contain overlapping DNA sequences to undergo multiple alignments. This will result in the formation of a contig [9], which is a contiguous sequence constructed from many clone sequences. The eventual bp length of a contig will thus be determined by the length of the sequence tags with overlapping DNA sequences. With an overall longer bp length than the individual sequence tags with which it was constructed, the contig will point towards the actual site of origin when aligned to the reference genome. In our current study, the median read length of the contigs is 261bp. With 88% of the sequence reads mapped to the reference genome resulting in the identification of about 11.000 unique genes based on contigs with a median length of 261bp, the coverage is approximately 24x in the current study [10].

Validation of RNA-Seq data by qPCR

To validate the sequencing data, quantitative PCR (qPCR) was performed on selected genes that were differentially expressed between asthma and controls. In short, PCR conditions were optimized in rt-PCR and subsequently in the qPCR (StepOnePlus; Applied Biosystems, Foster City, CA, USA). Samples were run in duplicate and the product was analyzed by melting temperature and occasionally on gel. qPCR data is expressed as the ratio of quantification cycles (Cq) of the genes compared to those of GAPDH [11]. The number of sequence reads per gene per subject was correlated with the Cq-ratio.

Primers sets for WARS, SCGB1A1, and BCL2 were developed flanking intron-exon borders to rule out genomic contamination. For WARS (75bp product) we used: forward: 5'-gccattgaccaggatccttact-3' and reverse: 5'-cagcaggggctggtttaggat-3' (Invitrogen), for SCGB1A1 (150bp): forward: 5'-ccctcctcatggacacaccct-3'; reverse: 5'-tttccatgagcttaatgatgctttctc-3', and for BCL2 (106bp product): forward 5'-cgggagatgtcgcccctggt-3' and reverse: 5'-caaaggcatcccagcctccgt-3'. GAPDH was quantified for normalization: forward: 5'-tcatctctgccccctctgc-3'; reverse: 5'-gagtccttccaggataccaa-3'. The qPCR-programme was as follows: hot start (95ºC), 3'' (95ºC), cycle: 15'' (95ºC), 30'' (WARS: 55ºC; SCGB1A1: 57ºC; BCL2: 60 ºC; GAPDH: 55ºC), 45'' (42ºC). mRNA from NCI-H292 cells (ATCC: CRL-1848) exposed to IFN-γ (100pg/mL for 24h) expressed WARS, SCGB1A1, and BCL2 and was used in the PCRs.

REFERENCES

1. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P, Jensen R, Johnson DC, MacIntyre N, McKay R, Navajas D, Pedersen OF, Pellegrino R, Viegi G, Wanger J; ATS/ERS Taskforce. Standardisation of spirometry. *Eur Respir J* 2005; 26: 319-338.
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. Busse WW, Wanner A, Adams K, Reynolds HY, Castro M, Chowdhury B, Kraft M, Levine RJ, Peters SP, Sullivan EJ. Investigative bronchoprovocation and bronchoscopy in airway diseases. *Am J Respir Crit Care Med* 2005; 172: 807-816.
4. 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.
5. Weterman MA, Sorrentino V, Kasher PR, Jakobs ME, van Engelen BG, Fluiter K, de Wissel MB, Sizarov A, Nürnberg G, Nürnberg P, Zelcer N, Schelhaas HJ, Baas F. A frameshift mutation in LRSAM1 is responsible for a dominant hereditary polyneuropathy. *Hum Mol Genet* 2012; 21: 358-370.
6. Ingenuity Systems; http://www.ingenuity.com
7. UCSC Genome Bioinformatics. http://genome.ucsc.edu
8. Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics* 2008; 5: 255-264.
9. Staden R. A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res* 1980; 8: 3673-3694.

10. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 1988; 2: 231-239.
11. Real-time PCR Data Markup Language, MIQE guidelines. http://www.rdml.org/miqe.php