ERJ Express. Published on March 5, 2008 as doi: 10.1183/09031936.00172107

IDENTIFICATION OF TRANSCRIPTS OVEREXPRESSED DURING AIRWAY EPITHELIUM DIFFERENTIATION

Brigitte CHHIN (1, 2)*, Jacqueline PHAM (1, 2)*, Loubna EL ZEIN (1, 2), Karine KAISER (3), Olivier MERROT (4), Patrice BOUVAGNET (1, 2, 5)

Laboratoire Cardiogénétique, EA 4171, Université Lyon 1, Lyon, France (1); Laboratoire Cardiogénétique, ERM 0107, INSERM, Lyon, France (2); Laboratoire de Parasitologie, Université Lyon 1, Lyon, France (3); Service ORL, Hôpital de la Croix-Rousse, Hospices Civils de Lyon, Lyon, France (4); Laboratoire Cardiogénétique, CBPE, Groupe Hospitalier Est, Hospices Civils de Lyon, Lyon, France (5).

* These authors contributed equally to this study.

Address for correspondence :

Patrice BOUVAGNET

Laboratoire Cardiogénétique, CBPE, Groupe Hospitalier Est

59, boulevard Pinel

69677 Bron, France

Tel.: +33 472 12 96 76; Fax: +33 427 85 59 00

Mail: Patrice.Bouvagnet@recherche.univ-lyon1.fr

Short title: AIRWAY EPITHELIUM SPECIFIC TRANSCRIPTS

ABSTRACT

Human airway epithelium, the forefront defence protecting the respiratory tract, evacuates inhaled particles by a permanent beating of epithelial cells cilia. When deficient, this organelle causes Primary Ciliary Dyskinesia (PCD), and despite numerous studies, data regarding ciliated cells gene expression are still incomplete. To identify genes specifically expressed in human ciliated respiratory cells, we performed a transcriptional analysis.

The transcriptome of de-differentiated epithelial cells was subtracted from fully re-differentiated cells using cDNA Representational Difference Analysis (RDA). To validate our results, gene overexpression in ciliated cells was confirmed by real-time PCR (RT-PCR), and by comparing our list of ciliated cell overexpressed genes to list obtained in previous studies.

We identified 53 known and 12 unknown genes overexpressed in ciliated cells. A majority (66%) of known genes was never reported before as being involved in ciliogenesis, while unknown genes represent hypothetical novel transcript isoforms or new genes not yet reported in databases. Finally, several genes identified here were located in genomic regions involved in PCD by linkage analysis.

In conclusion, this study revealed sequences of new cilia-related genes, new transcript isoforms and novel genes which should be further characterized for an understanding of their function(s) and their probable disorder-related involvement.

KEYWORDS

Human, Airway epithelium, Cilia, Representational Difference Analysis, Transcriptome

INTRODUCTION

The airway epithelium is a pseudo-stratified layer, consisting of specialized cell types including basal cells, goblet/secretory cells and ciliated columnar cells. It plays a critical role in airway defence by protecting the respiratory tract from infections and damages induced by inhaled toxins, pathogens and particles. It constitutes a physical barrier against environmental aggression, through secreted factors that mediate host immune system and through mucociliary clearance. On respiratory cells, cilia beat defects cause a disease referred to as Primary Ciliary Dyskinesia (PCD). Cilia are hair-like organelles which can be present on respiratory cells and on many other human cells. Cilia of all types have numerous similarities, but they differ depending on their motility or sensory function. An increasing interest in respiratory epithelium led researchers to elucidate genes acting in ciliogenesis.

Proteomic analyses were used to identify components located in axonemes or centrioles of cilia in human or flagella in other well-known organisms [1-5]. Comparative genomics searches detected genes conserved in the genome of ciliated organisms versus non-ciliated organisms [6, 7]. To reveal genes specifically expressed during flagellar regeneration or ciliogenesis, several studies were carried out using various transcriptional strategies [8-12].

Mutations in several genes revealed by these studies turned out to cause diverse human ciliary diseases such as polycystic kidney, retinal dystrophy, neurosensory impairment, Bardet-Biedl Syndrome (BBS), oral-facial-digital syndrome (OFD1) and PCD, demonstrating that these genes should be considered to decipher the etiology of ciliopathies [13].

In a different approach to discover genes specifically expressed in human ciliated respiratory cells which could be responsible for human disorders, we used a method referred to as Representational Difference Analysis (RDA). This method, first described by Lisitsyn *et al.* [14], is a process of subtraction coupled to amplification and was initially applied to detect differences between two genomes. Then, Hubank and Schatz adapted RDA for use with cDNA to isolate differentially expressed genes between two cell populations [15].

In this study, we utilized the sequential culture system described by Jorissen *et al*. [16] in which epithelial cells covering the turbinates of the nasal cavity are de-diffentiated in flat non-

ciliated cells and then re-differentiated in ciliated cells. The transcriptome of flat non-ciliated cells was subtracted from the transcriptome of re-ciliated cells in order to characterize transcripts which are specific to ciliated cells. Differentially expressed genes were cloned and sequenced resulting in the identification of *bona fide* and predicted-only genes. Moreover, genomic fragments which lay in intergenic intervals were cloned, suggesting the existence of new putative genes. The increased expression of some known and predicted-only genes during ciliogenesis was confirmed by RT-PCR validation studies.

MATERIAL AND METHODS

Cell culture

Human respiratory cells from normal subjects were obtained from nasal turbinates which were removed and discarded in the process giving access to the ethmoidal sinus (from department of Ear Nose and Throat, Hôpital de la Croix-Rousse, Lyon, France). Patients were operated for tumours located in the ethmoidal region and had no respiratory disease. Cells were grown using the immerged cell culture method described by Jorissen [16]. Briefly, ciliated cells were isolated by pronase digestion and expanded in collagen-coated 25 cm^2 flasks to de-differentiate in non-ciliated cells at 37°C, 5% CO_2. When they reached 80-90% confluence, collagen was digested and cells were suspended in flasks with rotation (80 rpm) at 37°C, to re-differentiate in the form of ciliated vesicles. Non-ciliated cells were collected at 80-90% confluence when they stop proliferating and vesicles were collected when they were fully covered by cilia.

Isolation of mRNA and cDNA synthesis

RNA was extracted from non-ciliated and ciliated cells with Extract-All® (Eurobio, France), following manufacturer's instructions. Poly(A)$^+$ mRNA was separated from total RNA by the Dynabeads Oligo$(dT)_{25}$ purification kit (Dynal Biotech, Norway) and its quality was assessed on an agarose gel. cDNA was synthesized from 2.85 µg poly(A) mRNA by oligo(dT) priming using SuperScript II Reverse Transcriptase as recommended by the manufacturer (Invitrogen,

Grand Island, NY). Double-stranded cDNA was prepared in 80 µL total volume containing: 20 µL cDNA template, 400 µM dNTPs, 5 U DNA ligase (New England Biolabs, Ipswich, MA), 24 U DNA polymerase (Invitrogen) and 1 U RNaseH (Invitrogen). Reaction was performed for 2 hours at 16°C and supplemented by 6 units of T4 DNA polymerase (Invitrogen) for 30 additional minutes of incubation.

Absence of genomic DNA contamination was confirmed by PCR with α-tubulin primers which could amplify either a 320 bp fragment on cDNA or a 468 bp fragment on genomic DNA (protocol is available on request).

Generation of Difference Products

cDNA RDA was performed on the basis of the protocol described by Hubank and Schatz [15] with slight modifications. Double-stranded cDNA (2 µg) from the two cell populations were digested with *Dpn*II (New England Biolabs, Ipswich, MA) to generate tester (ciliated cells) and driver (non-ciliated cells) cDNA representations. To facilitate purification of the digested representations, primers pair sets were biotinylated and removed using Streptavidin M-280 kit, following the manufacturer's recommendations (Dynal Biotech, Norway). The first subtractive hybridization tester:driver cDNA ratio was 1:50. In the second and third rounds of subtractive hybridization, the ratio was increased to 1:500 and 1:250000, and Mung bean nuclease digestion of PCR products was omitted. A detailed protocol is available from the authors on request.

Cloning and DNA sequencing

Products of the third-round of PCR were digested with *Dpn*II, and to facilitate their identification, bands of 200, 300, 400 and 600 bp were separately gel-purified using a Quiaquick gel extraction kit (Qiagen, Germantown, MD). Purified products were shotgun cloned into *Bam*HI-digested, dephosphorylated pBlueScript® II KS+ vector (Stratagene, The Netherlands) and used to transform DH5α One Shot competent cells, according to the manufacturer's protocol (Invitrogen, Grand Island, NY). Bacteria were plated on

LB/ampicillin plates and colonies were screened for inserts by *Sac*II and *Xho*I double digestion, after conventional plasmid extraction.

Cloned products were sequenced using M13 -20 primer. Sequencing reaction was set up using plasmid as template, and Big Dye® Terminator v1.1 cycle sequencing kit following the manufacturer's instructions (Applied Biosystems, Foster City, CA). Sequence analysis was performed on a 3100 automated ABI sequencing apparatus and sequences were aligned by the Staden Package after extraction of primer and vector sequences.

Sequence analysis

Sequences were formatted by using FASTA program [17], and compared to human genomic public databases, Ensembl [18] and NCBI [19], with BLASTN.

Real-time PCR (RT-PCR)

Non-ciliated and ciliated cells were collected, centrifuged to remove cell medium and washed in PBS. Cell pellet was stored at -80°C until processing. Purified mRNA was prepared using the Chemagic mRNA direct kit (Chemagen Biopolymer-Technologie AG, Germany) following manufacturer's recommendations. DNA contamination was removed with a DNase I treatment (Invitrogen, Grand Island, NY), and mRNA was quantified using a ND-1000 spectrophometer (NanoDrop, Wilmington, DE). Ten ng of mRNA were reverse transcribed to generate cDNA using Transcriptor First Stand cDNA synthesis kit (Roche Applied Science, France) and anchored oligo$(dT)_{18}$, according to the manufacturer's recommendations.

Real-time PCR was carried out on a LightCycler System® using the FastStart DNA Master SYBR Green I kit (Roche Applied Science). Reference and target genes primers were obtained in QuantiTect Primer Assays (Qiagen, Germantown, MD) which contained validated primers sets for: *GAPDH*, *DNAI1*, *GPNMB*, *RPGR*, *C7orf49* and *C14orf166*. PCR reactions were set up in a total volume of 20 µL containing 2 µL of SYBR Green FastStart reaction mix, 2.4 mM $MgCl_2$, 2 µl of 10X primers mix, and 2 µL cDNA. Temperature cycling profiles were as follows: 95°C for 10 min, 40 cycles of denaturation at 95°C for 10 s, annealing at

55°C for 10 s, and extension at 72°C for 20 s. Melting curve analysis was carried out in the range from 65 to 95°C to confirm the specificity of PCR products.

The gene expression level was performed with the comparative threshold method using as a reference value the housekeeping gene *glyceraldehyde-3-phosphate dehydrogenase*, *GAPDH* [20]. Ct values of the real-time PCR were collected, where Ct was defined as the threshold cycle of PCR at which amplified product was first detected. The corrected ΔCt value for each determination was then used to calculate the relative N-fold differential expression in a specific gene of ciliated cells sample compared to non-ciliated cells and was expressed as the ratio of the $2^{-\Delta Ct}$ values.

Statistical analysis

Data were expressed as mean ± SD from triplicate experiments. For each target gene, $2^{-\Delta Ct}$ values of ciliated and non-ciliated cells were statistically analysed by Student t test and the significance was set at $p<0.05$ for a one-tailed test.

RESULTS

Our goal was to identify genes specifically expressed in human ciliated respiratory cells. To this end, we subtracted the transcriptome of de-differentiated epithelial cells from fully re-differentiated cells. The cDNA Representational Difference Analysis (RDA) procedure used in the present study was closely based on the protocol described by Hubank and Schatz [15] which allows the enrichment of transcripts specifically expressed in a cell type through iterative cycles of amplification/subtraction.

Determination of genes differentially expressed

Twenty-five percent of bacterial clones contained an insert. All clones (78) with an insert were sequenced, including 4 chimeric clones which had 2 cDNA fragments. Altogether, 82 individual cDNA fragments were identified. Four of these cDNA fragments were recovered several times and correspond to the following genes: *GPNMB* (9 clones), *ZNF236* (6), *RPLP0*

(4) and *RPLP14* (2). Finally, 65 unique cDNA fragments were identified and mapped to genomic human sequence by BLASTN. Fifty-three (82%) cDNA fragments mapped to exonic sequences of known genes. Seven (11%) cDNA fragments mapped to intronic regions or at the boundary of intron-exon junction while 5 (7%) cDNA fragments mapped to intergenic regions.

The list of the fifty-three cDNA fragments corresponding to exonic regions is presented in Table 1. Genes were clustered by their described functions. It is notable that nine ribosomal proteins and an initiation factor involved in translation were identified. Genes related to immunity, inflammation and defence were detected. The cluster 'Mitochondria' includes 7 genes encoding mitochondrial components. Among these, six are nuclear genes and one is mitochondrial (*MT-CYB*). Components of channels, transporters or related proteins are listed in one group which includes two SLC family genes. Genes of cell signaling and signal transduction such as *SPATA13* and *TEGT* are present. Several cytoskeleton genes, as A*CTG1*, keratin and tropomyosin genes were evidenced. In spite the fact that ciliated cells do not replicate, several cell proliferation genes were found, including *NME2*. Among the remaining genes of Table 1, only one - the *β-tubulin* gene - was previously demonstrated to be implicated in cilia. Finally, 2 genes corresponding to predicted open reading frames were identified: *C7orf49* and *C14orf166*.

The cDNA fragments with sequence aligning to introns are reported in Table 2. These cDNA fragments are presumably representative of new isoforms of mRNA. Only the cDNA fragment mapping to the *exportin 6* gene includes intronic and exonic sequences. The 6 other fragments are coming from intronic regions. One fragment maps to a genomic region where 2 genes overlap on opposite strands (*FAM82B*/*CPNE3*). Since our cloning strategy was not oriented, we do not know whether this fragment is derived from one or the other gene.

The five cDNA fragments which mapped to intergenic regions are presented in Table 3. The 2 first cDNA fragments (A and B) could represent a new 5' or 3' exon because the distance with the closest gene is at most 16 kilobases. By contrast, the 2 last cDNA fragments (D and

E) are so isolated (minimal distance 75 kilobases to the nearest gene) that they can only be a segment of new genes.

RT-PCR validation studies

First, to validate our culture system, we analysed the expression of *DNAI1* and *RPGR*, two genes which are known to be overexpressed in ciliated cells. *DNAI1* is a component of outer dynein arms in ciliary axoneme and is essential for cilia function, while RPGR is specifically expressed in tissues containing cells with cilia or cilia-like organelle such as retina, lung, cochlea and epithelial cells lining bronchi and sinuses [21]. For data analysis, we calculated in each sample the relative quantification of target gene transcripts normalized with *GAPDH* expression since it is an endogenous unregulated gene transcript. By RT-PCR, we found that these two genes have a 3696 and 83 relative fold increase, respectively.

Second, to validate our list of genes, we evaluated the expression of one gene (*GPNMB*) and two open reading frames, namely *C7orf49* and *C14orf166*, by RT-PCR in non-ciliated and ciliated cells. The three selected genes had an expression in ciliated cells which was significantly higher than in non-ciliated cells (Figure 1). *GPNMB* displayed an increase of 7.89 relative fold changes consistent with the 3.01 to 12.17 fold changes reported by Ross [12]. *C7orf49* and *C14orf166* showed an increased expression in human respiratory ciliated cells with 2.94 and 4.92 relative fold changes, respectively.

Comparison with other studies on cilia

To confirm our data, we compared our genes list to other lists obtained by various methods from ciliated cells of various organisms. Ten out of 53 genes showing an increased transcript synthesis in this work were reported once in other studies: *ANXA1*, *EIF1*, *GPNMB*, *NEBL*, *NME2*, *RPL14*, *SCGB1A1*, *SPATA13*, *TMSB4X* and *UBA52* [1, 3, 11, 12]. Two additional genes were mentioned in both Ostrowski and Pazour studies: *RPSA* and *ACTG1* [1, 3]. *TUBB* gene was also detected two times [4, 8]. Finally, a computer-predicted gene referred to as *C14orf166* gene, was mentioned in three times [6-8].

In agreement with Pazour's study, we found that *FKBP1A* has an increased expression during ciliogenesis [3]. By contrast, Ross *et al*. noted a decreased expression (-2.33) of this gene [12]. Three genes which are reported in our study with an increased expression during ciliogenesis showed reversed and inconsistent patterns in Ross study: *SERPINB2* (-2.82), *TPM1* (-3.31/3.73) and *TPM3* (-3.49/3.21) [12].

The identification of causal genes in PCD and *situs inversus* by positional cloning is difficult because of the potentially numerous genes involved in these diseases. As a consequence, it is important to note that 22 genes of our series map to chromosomal regions which may contain a causal gene as determined by a position cloning approach [22-25]. These 22 genes are presented in Table 5. Half were never reported before in studies aimed at characterizing components specific to cilia/flagella. Among these 22 genes, 4 are located in chromosomal regions with a definite linkage in familial PCD: *CXCL17*, *SQRDL*, *SLC6A14* and *TMSB4X* [23-25].

DISCUSSION

In this work, we used Representational Difference Analysis (RDA) to identify genes differentially expressed in human ciliated respiratory cells. The subtraction of non-ciliated cells representation from ciliated cells representation resulted in the detection of 53 genes and 12 new coding sequences of known genes (7) or of putative new genes (5). These 65 cDNA fragments are truly overexpressed in ciliated cells because a subset of 3 cDNAs were tested by RT-PCR analysis and all the 3 cDNA were significantly expressed at higher level in ciliated than in non-ciliated cells. In addition, 18 out of the 53 cDNAs were already reported in other studies aimed at deciphering transcripts or proteins specific to ciliated cells. The remaining 35 cDNAs are newly identified genes overexpressed in ciliated cells. Twenty-two sequences are located at loci related to known ciliary structure defects and diseases.

RDA is a method which is long and difficult to set-up but it is re-known for providing no false positives. Consistent with this observation, all the cDNAs which were tested by RT-PCR

confirmed their overexpression and 15 of the cDNAs of this study were reported in other studies as being overexpressed in ciliated cells. One limitation of this method is that there is a selection of short cDNA fragments so that there is no hope to obtain a complete list of all overexpressed transcripts. While microarrays provide a systematic view of transcripts expression and can only detect preselected sequences, RDA has the power to detect additional expressed sequences either as novel transcript variant of known genes or even as new genes.

In Table 1, we reported a set of genes which could be linked to cells with cilia/flagella. Indeed, we identified 7 genes encoding proteins of the mitochondrial respiratory chain and the oxidative phosphorylation system. In ciliated cells, increase in ATP production is presumably necessary for intracellular and intraflagellar transport, and cilia beating. Mutations in *ACTG1*, a major component of sensory ciliated cells of the cochlea, were described to cause dominant deafness [26]. Finally, *NME2* was demonstrated to be involved in spermiogenesis and flagellar movement [27], and *TMSB4X* was found in a previous work to be highly represented in lung parenchyma and unrelated tissue types relative to the bronchial epithelium [11].

For RT-PCR analysis, we used two positive controls which were previously described to present an increased expression in ciliated tissues, *DNAI1* and *RPGR* genes. Two other genes, *GPNMB* and *C14orf166*, were reported in others studies to show an increased expression in ciliated cells, but their function in ciliogenesis is not yet elucidated. GPNMB, a transmembrane glycoprotein was hypothesized to be involved in growth delay and reduction of metastatic potential [28], but its role in ciliogenesis is still elusive. *C14orf166* and *C7orf49* genes were predicted by bio-informatics searches on the human genome. *C14orf166* gene is frequently detected during ciliogenesis since it was reported by three other studies. This gene encoded a protein involved in the functional regulation of human ninein in the centrosome structure [6-8, 29]. It would be interesting to complete data on *C14orf166* and elucidate *C7orf49* gene biological function.

Finally, we detected 5 totally new genes (about 10% of the whole set). Further work is warranted to characterize in details these putative new genes in particular sequence B which is

located in a chromosomal region implicated in *situs inversus*, a disturbance of lateralisation which can be secondary of cilia dysfunction in the early embryo.

ACKNOWLEDGEMENTS: B. Chhin was supported by a stipend from Association Française contre les Myopathies (AFM) from 2004 to 2007. Laboratoire Cardiogénétique is supported by grants from GIS-Maladies Rares 2003, PHRC régional 2003, Fondation ProKartagener 2005 and Fondation Renaud FEBVRE. We warmly thank Pr S. Picot for his laboratory's contribution for the RT-PCR experiments (Laboratoire de Parasitologie, Lyon, France).

REFERENCES

1. Ostrowski LE, Blackburn K, Radde KM, Moyer MB, Schlatzer DM, Moseley A, Boucher RC. A proteomic analysis of human cilia: identification of novel components. *Mol Cell Proteomics* 2002; 1:451-65.

2. Smith JC, Northey JG, Garg J, Pearlman RE, Siu KW. Robust method for proteome analysis by MS/MS using an entire translated genome: demonstration on the ciliome of Tetrahymena thermophila. *J Proteome Res* 2005; 4:909-19.

3. Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 2005; 170:103-13.

4. Keller LC, Romijn EP, Zamora I, Yates JR, 3rd, Marshall WF. Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr Biol* 2005; 15:1090-8.

5. Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McKean PG, Gull K. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 2006; 440:224-7.

6. Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, May-Simera H, Li H, Blacque OE, Li L, Leitch CC, Lewis RA, Green JS, Parfrey PS, Leroux MR, Davidson WS, Beales PL, Guay-Woodford LM, Yoder BK, Stormo GD, Katsanis N, Dutcher SK. Comparative

genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 2004; 117:541-52.

7. Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, Zuker CS. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 2004; 117:527-39.

8. Stolc V, Samanta MP, Tongprasit W, Marshall WF. Genome-wide transcriptional analysis of flagellar regeneration in Chlamydomonas reinhardtii identifies orthologs of ciliary disease genes. *Proc Natl Acad Sci U S A* 2005; 102:3703-7.

9. Blacque OE, Perens EA, Boroevich KA, Inglis PN, Li C, Warner A, Khattra J, Holt RA, Ou G, Mah AK, McKay SJ, Huang P, Swoboda P, Jones SJ, Marra MA, Baillie DL, Moerman DG, Shaham S, Leroux MR. Functional genomics of the cilium, a sensory organelle. *Curr Biol* 2005; 15:935-41.

10. Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol Cell* 2000; 5:411-21.

11. Lonergan KM, Chari R, Deleeuw RJ, Shadeo A, Chi B, Tsao MS, Jones S, Marra M, Ling V, Ng R, Macaulay C, Lam S, Lam WL. Identification of novel lung genes in bronchial epithelium by serial analysis of gene expression. *Am J Respir Cell Mol Biol* 2006; 35:651-61.

12. Ross AJ, Dailey LA, Brighton LE, Devlin RB. Transcriptional profiling of mucociliary differentiation in human airway epithelial cells. *Am J Respir Cell Mol Biol* 2007; 37:169-85.

13. Inglis PN, Boroevich KA, Leroux MR. Piecing together a ciliome. *Trends Genet* 2006; 22:491-500.

14. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 1993; 259:946-51.

15. Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* 1994; 22:5640-8.

16. Jorissen M, Willems T, Van der Schueren B, Verbeken E, De Boeck K. Ultrastructural expression of primary ciliary dyskinesia after ciliogenesis in culture. *Acta otorhinolaryngol Belg* 2000; 54:343-56.

17. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 1988; 85:2444-8.

18. Ensembl. http://www.ensembl.org. Date last updated: October 23 2007. Date last accessed: December 6 2007.

19. NCBI Build 36.2. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/. Date last updated: September 14 2006. Date last accessed: December 6 2007.

20. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001; 29:e45.

21. Iannaccone A, Breuer DK, Wang XF, Kuo SF, Normando EM, Filippova E, Baldi A, Hiriyanna S, MacDonald CB, Baldi F, Cosgrove D, Morton CC, Swaroop A, Jablonski MM. Clinical and immunohistochemical evidence for an X linked retinitis pigmentosa syndrome with recurrent infections and hearing loss in association with an RPGR mutation. *J Med Genet* 2003; 40:e118.

22. Blouin JL, Meeks M, Radhakrishna U, Sainsbury A, Gehring C, Sail GD, Bartoloni L, Dombi V, O'Rawe A, Walne A, Chung E, Afzelius BA, Armengot M, Jorissen M, Schidlow DV, van Maldergem L, Walt H, Gardiner RM, Probst D, Guerne PA, Delozier-Blanchet CD, Antonarakis SE. Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8:109-18.

23. Meeks M, Walne A, Spiden S, Simpson H, Mussaffi-Georgy H, Hamam HD, Fehaid EL, Cheehab M, Al-Dabbagh M, Polak-Charcon S, Blau H, O'Rawe A, Mitchison HM, Gardiner RM, Chung E. A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37:241-4.

24. Jeganathan D, Chodhari R, Meeks M, Faeroe O, Smyth D, Nielsen K, Amirav I, Luder AS, Bisgaard H, Gardiner RM, Chung EM, Mitchison HM. Loci for primary ciliary

dyskinesia map to chromosome 16p12.1-12.2 and 15q13.1-15.1 in Faroe Islands and Israeli Druze genetic isolates. *J Med Genet* 2004; 41:233-40.

25. Krawczynski MR, Witt M. PCD and RP: X-linked inheritance of both disorders? *Pediatr Pulmonol* 2004; 38:88-9.

26. Rendtorff ND, Zhu M, Fagerheim T, Antal TL, Jones M, Teslovich TM, Gillanders EM, Barmada M, Teig E, Trent JM, Friderici KH, Stephan DA, Tranebjaerg L. A novel missense mutation in ACTG1 causes dominant deafness in a Norwegian DFNA20/26 family, but ACTG1 mutations are not frequent among families with hereditary hearing impairment. *Eur J Hum Genet* 2006; 14:1097-105.

27. Munier A, Serres C, Kann ML, Boissan M, Lesaffre C, Capeau J, Fouquet JP, Lacombe ML. Nm23/NDP kinases in human male germ cells: role in spermiogenesis and sperm motility? *Exp Cell Res* 2003; 289:295-306.

28. Weterman MA, Ajubi N, van Dinter IM, Degen WG, van Muijen GN, Ruitter DJ, Bloemers HP. nmb, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. *Int J Cancer* 1995; 60:73-81.

29. Howng SL, Hsu HC, Cheng TS, Lee YL, Chang LK, Lu PJ, Hong YR. A novel ninein-interaction protein, CGI-99, blocks ninein phosphorylation by GSK3beta and is highly expressed in brain tumors. *FEBS Lett* 2004; 566:162-8.

Table 1: Genes overexpressed in human ciliated respiratory cells, cDNA fragments mapping to exonic regions. MT, mitochondrial genome. Values of maximum identity between insert and genomic sequences are indicated.

Gene ID	Gene symbol	Description	Location	Max identity
Translation				
6175	RPLP0	Ribosomal protein, large, P0	12q24.2	99%
6133	RPL9	Ribosomal protein L9	4p13	100%
6136	RPL12	Ribosomal protein L12	9q34	97%
9045	RPL14	Ribosomal protein L14	3p22-p21.2	91%
6228	RPS23	Ribosomal protein S23	5q14.2	100%
6156	RPL30	Ribosomal protein L30	8q22	99%
6161	RPL32	Ribosomal protein L32	3p25-p24	99%
6164	RPL34	Ribosomal protein L34	4q25	100%
3921	RPSA	Ribosomal protein SA	3p22.2	97%
10209	EIF1	Eukaryotic translation initiation factor 1	17q21.2	98%
Immunity, inflammation and defense				
301	ANXA1	Annexin A1	9q12-q21.2	100%
967	CD63	CD63 molecule	12q12-q13	99%
284340	CXCL17	Chemokine (C-X-C motif) ligand 17	19q13.2	100%
2280	FKBP1A	FK506 binding protein 1A, 12kDa	20p13	100%
7356	SCGB1A1	Secretoglobin, family 1A, member 1 (uteroglobin)	11q12.3-q13.1	99%
5055	SERPINB2	Serpin peptidase inhibitor, clade B (ovalbumin), member 2	18q21.3	100%
7114	TMSB4X	Thymosin, beta 4, X-linked	Xq21.3-q22	95%
7311	UBA52	Ubiquitin A-52 residue ribosomal protein fusion product 1	19p13.1-p12	98%
Mitochondria				
514	ATP5E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit	20q13.32	92-97%
10476	ATP5H	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	17q25	99%
539	ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	21q22.11	97%
4519	MT-CYB	Mitochondrially encoded cytochrome b	MT	100%
4697	NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa	7p21.3	99%
58472	SQRDL	Sulfide quinone reductase-like (yeast)	15q15	97%
29796	UCRC	Ubiquinol-cytochrome c reductase complex (7.2 kD)	22cen-q12.3	95%

Gene ID	Gene symbol	Description	Location	Max identity
Channels, transporters and related proteins				
1173	AP2M1	Adaptor-related protein complex 2, mu 1 subunit	3q28	100%
56888	KCMF1	Potassium channel modulatory factor 1	2p11.2	99%
26266	SLC13A4	Solute carrier family 13 (sodium/sulfate symporters), member 4	7q33	100%
11254	SLC6A14	Solute carrier family 6 (amino acid transporter), member 14	Xq23-q24	98%
Cell signaling and signal transduction				
54541	DDIT4	DNA-damage-inducible transcript 4	10pter-q26.12	96%
4092	SMAD7	SMAD family member 7	18q21.1	97%
221178	SPATA13	Spermatogenesis associated 13	13q12.12	98%
7009	TEGT	Testis enhanced gene transcript (BAX inhibitor 1)	12q12-q13	100%
Cytoskeleton				
71	ACTG1	Actin, gamma 1	17q25	98%
3861	KRT14	Keratin 14 (epidermolysis bullosa simplex, Dowling-Meara, Koebner)	17q12-q21	100%
3853/140446	KRT6A/KRT6C	Keratin 6A/6C	12q12-q13	98%
3855	KRT7	Keratin 7	12q12-q13	97-100%
10529	NEBL	Nebulette	10p12	100%
7168	TPM1	Tropomyosin 1 (alpha)	15q22.1	97%
7170	TPM3	Tropomyosin 3	1q21.2	94%
Cell proliferation				
996	CDC27	Cell division cycle 27 homolog (S. cerevisiae)	17q12-17q23.2	94%
10969	EBNA1BP2	EBNA1 binding protein 2	1p35-p33	98%
3397	ID1	Inhibitor of DNA binding 1, dominant negative helix-loop-helix protein	20q11	100%
4831	NME2	Non-metastatic cells 2, protein (NM23B) expressed in	17q21.3	97%
57804	POLD4	Polymerase (DNA-directed), delta 4	11q13	99%
90441	ZNF622	Zinc finger protein 622	5p15.1	95%
Other				
78996	C7orf49	Chromosome 7 open reading frame 49	7q33	98%
51637	C14orf166	Chromosome 14 open reading frame 166	14q22.1	99%
26234	FBXL5	F-box and leucine-rich repeat protein 5	4p15.33	100%
10457	GPNMB	Glycoprotein (transmembrane) nmb	7p15	96%
51108	METTL9	Methyltransferase like 9	16p13-p12	98%
203068	TUBB	Tubulin, beta	6p21.33	98%
7776	ZNF236	Zinc finger protein 236	18q22-q23	98%

Table 2: Genes overexpressed in human ciliated respiratory cells, cDNA fragments mapping to intronic regions. One cDNA fragment sequence mapped to 2 genes (*FAM82B* and *CPNE3*). *, at the exon-intron junction. Values of maximum identity between insert and database sequences are indicated.

Gene ID	Gene symbol	Description	Location	Max identity
54875	C9orf39	Chromosome 9 open reading frame 39	9p22.2	98%
51115/8895	FAM82B/CPNE3	Family with sequence similarity 82, member B/Copine III	8q21.3	99%
23464	GCAT	Glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzyme A ligase)	22q13.1	98%
4862	NPAS2	Neuronal PAS domain protein 2	2q11.2	98%
10196	PRMT3	Protein arginine methyltransferase 3	11p15.1	99%
23214	XPO6	Exportin 6*	16p11.2	98%
619279	ZNF704	Zinc finger protein 704	8q21.13	98%

Table 3: cDNA fragments mapping to intergenic regions. Values of maximum identity between insert and database sequences are indicated.

Sequence	Description	Location	Max identity
A	1939 bp at 5' side: Kruppel-like factor 1 (erythroid) (KLF1); 2130 bp at 3' side: glutaryl-Coenzyme A dehydrogenase isoform b precursor (GCDH)	19p13	100%
B	13018 bp at 5' side: apolipoprotein A-V (APOA5); 15879 bp at 3' side: apolipoprotein A-IV precursor (APOA4)	11q23	98%
C	58077 bp at 5' side: hypothetical protein LOC220134 (C18orf24); 213331 bp at 3' side: mitogen-activated protein kinase 4 (MAPK4)	18q21.1	97%
D	146030 bp at 5' side: phosphatidylinositol-specific phospholipase C (PLCXD3); 74896 bp at 3' side: 3-oxoacid CoA transferase 1 precursor (OXCT1)	5p13.1	100%
E	757367 bp at 5' side: methionine adenosyltransferase II, beta isoform 1 (MAT2B); 3300072 bp at 3' side: similar to odd Oz/ten-m homolog 2 isoform 5 (ODZ2)	5q34	92%

Table 4: Present study identified genes also reported in previous studies. Cr, *Chlamydomonas reinhardtii*; Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; *, decreased expression in the study.

Gene	Description	Reference	ID in reference
ACTG1	Actin, gamma 1	[3] [1]	C_1310009 in Cr X03284 in Hs
ANXA1	Annexin A1	[1]	Multiple annexin 1 in Hs
C14orf166	Chromosome 14 open reading frame 166	[6, 8] [7]	168283 in Cr CG31249 in Dm
EIF1	Eukaryotic translation initiation factor 1	[3]	C_190059 in Cr
FKBP1A	FK506 binding protein 1A	[3] [12]*	C_230098 in Cr 2280 in Hs
GPNMB	Glycoprotein (transmembrane) nmb	[12]	10457 in Hs
NEBL	Nebulette	[12]	10529 in Hs
NME2	Non-metastatic cells 2	[3]	C_1230002 in Cr
RPL14	Ribosomal protein L14	[3]	C_870056 in Cr
RPSA	Ribosomal protein SA	[3] [1]	C_130042 in Cr X61156 in Hs
SCGB1A1	Secretoglobin, family 1A, member 1	[11]	7356 in Hs
SERPINB2	Serpin peptidase inhibitor, clade B, member 2	[12]*	5055 in Hs
SPATA13	Spermatogenesis associated 13	[12]	221178 in Hs
TMSB4X	Thymosin, beta 4, X-linked	[11]	7114 in Hs
TPM1	Tropomyosin 1	[12]*	7168 in Hs
TPM3	Tropomyosin 3	[12]*	7170 in Hs
TUBB	Tubulin, beta	[8] [4]	170055 in Cr 158210 in Cr
UBA52	Ubiquitin A-52	[3]	C_1610014 in Cr

Table 5: Present study genes located at loci related with known ciliary structure defects and diseases. ODA, outer dynein arm defect; IDA, inner dynein arm defect; PCD, Primary Ciliary Dyskinesia; SI, *situs inversus*; DAD, dynein arm defects; *, not previously reported in studies aimed at identifying specific components of cilia/flagella.

Defect	Location	Reference	Present study genes at these loci
Familial studies			
ODA	19q13-19qter	[23]	CXCL17*
IDA	15q13.3-15.1	[24]	SQRDL*
IDA	X	[25]	SLC6A14*, TMSB4X
Genomic analysis	Potential loci		
PCD or SI	3p	[22]	RPL14, RPL32*, RPSA
DAD	7p	[22]	GPNMB, NDUFA4*
PCD	10p	[22]	DDIT4*, NEBL
SI	11q	[22]	POLD4*, SCGB1A1, sequence B*
SI	13q	[22]	SPATA13
PCD	15q	[22]	TPM1
SI, DAD	17q	[22]	ACTG1, ATP5H*, CDC27*, EIF1, KRT14*, NME2

Table 6: Expression relative fold changes for some interesting genes. For each gene, calculation was made using relative expressions ratio of ciliated cells on non-ciliated cells.

	Relative Fold Change
DNAI1	3696.61
GPNMB	7.89
RPGR	83.44
C7orf49	2.94
C14orf166	4.92

Figure legend

Figure 1: Validation of the expression data for several identified genes in human ciliated cells. Bar graphs indicate, in arbitrary units, expressions in non-ciliated (white bars) and ciliated cells (hashed bars). Error bars indicate the standard error of the mean. All ciliated values are significantly different from non-ciliated values, $p<0.05$.

Figure 1