Eur Respir J 2008; 32: 121–128
DOI: 10.1183/09031936.00172107

Identification of transcripts overexpressed during airway epithelium differentiation

B. Chhin*,#,f, J.T. Pham*,#,f, L. El Zein*,#, K. Kaiser¶, O. Merrot+ and P. Bouvagnet*,#,§

ABSTRACT: Human airway epithelium, the defence at the forefront of protecting the respiratory tract, evacuates inhaled particles by a permanent beating of epithelial cell cilia. When deficient, this organelle causes primary ciliary dyskinesia, and, despite numerous studies, data regarding ciliated cell gene expression remain incomplete. The aim of the present study was to identify genes specifically expressed in human ciliated respiratory cells *via* transcriptional analysis.

The transcriptome of dedifferentiated epithelial cells was subtracted from that of fully redifferentiated cells using complementary DNA representational difference analysis. In order to validate the results, gene overexpression in ciliated cells was confirmed by real-time PCR, and by comparing the present list of genes overexpressed in ciliated cells to lists obtained in previous studies.

A total of 53 known and 12 unknown genes overexpressed in ciliated cells were identified. The majority (66%) of known genes had never previously been reported as being involved in ciliogenesis, and the 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 primary ciliary dyskinesia by linkage analysis.

In conclusion, the present study revealed sequences of new cilia-related genes, new transcript isoforms and novel genes which should be further characterised to aid understanding of their function(s) and their probable disorder-related involvement.

KEYWORDS: Airway epithelium, cilia, human, representational difference analysis, transcriptome

The airway epithelium is a pseudo-stratified layer, consisting of specialised 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 damage induced by inhaled toxins, pathogens and particles. It constitutes a physical barrier against environmental aggression, through secreted factors that mediate the host immune system and through mucociliary clearance. On respiratory cells, ciliary 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 exhibit numerous similarities, but they differ depending on their motility or sensory function. An increasing interest in respiratory epithelia has led researchers to elucidate genes acting in ciliogenesis.

Proteomic analyses have been used to identify components located in the axonemes or centrioles of cilia in humans or flagella in other well-known organisms [1–5]. Comparative genomic searches led to the detection of genes conserved in the genome of ciliated organisms *versus* nonciliated organisms [6, 7]. In order to reveal genes specifically expressed during flagellar regeneration or ciliogenesis, several studies have been 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 disease, retinal dystrophy, neurosensory impairment, Bardet–Biedl syndrome, oral-facial-digital syndrome type 1 and PCD, demonstrating that these genes should be considered in deciphering the aetiology of ciliopathies [13].

Using a different approach, to discover genes specifically expressed in human ciliated respiratory cells that could be responsible for human disorders, a method referred to as representational difference analysis (RDA) was used. This method, first described by LISITSYN *et al.* [14], represents a process of subtraction coupled to amplification and was initially applied to the detection of differences between two genomes. Subsequently, HUBANK and SCHATZ [15] adapted RDA for use with complementary DNA (cDNA)

AFFILIATIONS
*Laboratory of Cardiogenetics, Laboratory Team 4171, and,
¶Parasitology, University of Lyon 1,
#Laboratory of Cardiogenetics, Methodology Research Team 0107, National Institute for Health and Medical Research,
+Ear, Nose and Throat Service, Croix-Rousse Hospital, Hospices Civils de Lyon,
§Laboratory of Cardiogenetics, East Centre of Biology and Pathology, Groupe Hospitalier Est, Hospices Civils de Lyon, Lyon, France.
fBoth authors contributed equally to this study.

CORRESPONDENCE
P. Bouvagnet, Laboratoire Cardiogénétique, CBPE, Groupe Hospitalier Est, 59 boulevard Pinel, 69677 Bron, France.
Fax: 33 427855900
E-mail: Patrice.Bouvagnet@recherche.univ-lyon1.fr

Received:
December 19 2007
Accepted after revision:
February 08 2008

SUPPORT STATEMENT
B. Chhin was supported by a stipend from the French Association against Myopathies (Evry, France) between 2004 and 2007. The Laboratory of Cardiogenetics (Hospices Civils de Lyon, Lyon, France) is supported by grants from Scientific Interest Group – Rare Diseases 2003 and Hospital Project of Clinical Research regional 2003 (Paris, France), and the Laboratory of Cardiogenetics (University of Lyon 1, Lyon, France) by grants from the ProKartagener Foundation 2005 (Geneva, Switzerland) and Renaud Febvre Foundation (Les Etards, France).

STATEMENT OF INTEREST
None declared.

European Respiratory Journal
Print ISSN 0903-1936
Online ISSN 1399-3003

in order to isolate genes expressed differentially in two cell populations.

In the present study, the sequential culture system described by JORISSEN *et al.* [16], in which epithelial cells covering the turbinates of the nasal cavity are dedifferentiated in flat nonciliated cells and then redifferentiated in ciliated cells, was utilised. The transcriptome of flat nonciliated cells was subtracted from that of re-ciliated cells in order to characterise transcripts specific to ciliated cells. Differentially expressed genes were cloned and sequenced, resulting in the identification of *bona fide* and predicted genes. Moreover, genomic fragments that lay in intergenic intervals were cloned, suggesting the existence of new putative genes. The increased expression of some known and predicted genes during ciliogenesis was confirmed by real-time PCR validation studies.

MATERIALS AND METHODS

Cell culture

Human respiratory cells from normal subjects were obtained from nasal turbinates, which were removed and discarded, thereby providing access to the ethmoidal sinus (at the Ear Nose and Throat Service, Croix-Rousse Hospital, Lyon, France). Patients were operated on for tumours located in the ethmoidal region and showed no respiratory disease. Cells were grown using the immerged cell culture method described by JORISSEN *et al.* [16]. Briefly, ciliated cells were isolated by pronase digestion and expanded in collagen-coated 25-cm^2 flasks to dedifferentiate in nonciliated cells at 37°C under 5% CO_2. When they reached 80–90% confluence, collagen was digested and cells were suspended in flasks with rotation (80 revolutions·min^{-1}) at 37°C to redifferentiate in the form of ciliated vesicles. Nonciliated cells were collected at 80–90% confluence, when they stopped proliferating, and vesicles were collected when they were fully covered by cilia.

Isolation of mRNA and complementary DNA synthesis

RNA was extracted from nonciliated and ciliated cells using Extract-All® (Eurobio, Courtaboeuf, France), following the manufacturer's instructions. Poly(A)$^+$ mRNA was separated from total RNA using the Dynabeads Oligo(dT)$_{25}$ purification kit (Dynal Biotech, Oslo, Norway) and its quality was assessed on an agarose gel. cDNA was synthesised from 2.85 μg poly(A) mRNA by oligo-deoxythymidine (dT) priming using SuperScript II Reverse Transcriptase as recommended by the manufacturer (Invitrogen, Grand Island, NY, USA). Double-stranded cDNA was prepared in an 80-μL total volume, containing 20 μL cDNA template, 400 μM deoxyribonucleoside triphosphates, 5 U DNA ligase (New England Biolabs, Ipswich, MA, USA), 24 U DNA polymerase (Invitrogen) and 1 U RNase H (Invitrogen). The reaction was performed for 2 h at 16°C and then supplemented by 6 U T4 DNA polymerase (Invitrogen) for an additional 30-min incubation.

The absence of genomic DNA contamination was confirmed by PCR using α-tubulin primers, which could amplify either a 320-bp cDNA fragment or a 468-bp genomic DNA fragment (protocol 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 was digested with *Dpn*II (New England Biolabs) to generate tester (ciliated cells) and driver (nonciliated cells) cDNA representations. In order to facilitate purification of the digested representations, primers pair sets were biotinylated and removed using a Streptavidin M-280 kit (Dynal Biotech), following the manufacturer's recommendations. The first subtractive hybridisation tester:driver cDNA ratio was 1:50. In the second and third rounds of subtractive hybridisation, the ratio was increased to 1:500 and 1:250,000, and mung bean nuclease digestion of PCR products was omitted (detailed protocol available on request).

Cloning and DNA sequencing

The products of the third round of the 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 QIAquick gel extraction kit (Qiagen, Germantown, MD, USA). Purified products were shotgun cloned into a *Bam*HI-digested dephosphorylated pBlueScript® II KS+ vector (Stratagene, San Diego, CA, USA) and used to transform DH5α One Shot competent cells (Invitrogen), according to the manufacturer's protocol. Bacteria were plated on Luria–Bertani medium/ampicillin plates and colonies screened for inserts by *Sac*II and *Xho*I double digestion, following conventional plasmid extraction.

Cloned products were sequenced using the M13–20 primer. The sequencing reaction was set up using the plasmid asa template and the Big Dye® Terminator v1.1 cycle sequencing kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. Sequence analysis was performed on a 3100 automated ABI sequencing apparatus (Applied Biosystems), and sequences were aligned using Staden 1.7.0 for Windows from the Staden Package [17] after extraction of primer and vector sequences.

Sequence analysis

Sequences were formatted using the FAST-All program [18] and compared to the public human genomic databases, Ensembl [19] and National Center for Biotechnology Information [20], with the nucleotide–nucleotide basic local alignment search tool (BLASTN).

Real-time PCR

Nonciliated and ciliated cells were collected, centrifuged to remove cell medium and washed in PBS (pH 7.4). The cell pellet was stored at -80°C until processing. Purified mRNA was prepared using the Chemagic mRNA direct kit (Chemagen, Baesweiler, Germany) following the manufacturer's recommendations. DNA contamination was removed with a DNase I treatment (Invitrogen), and mRNA was quantified using an ND-1000 spectrophometer (NanoDrop, Wilmington, DE, USA). mRNA (10 μg) was reverse transcribed to generate cDNA using the Transcriptor First Strand cDNA synthesis kit (Roche Applied Science, Rotkreuz, Switzerland) and anchored oligo(dT)$_{18}$, according to the manufacturer's recommendations.

TABLE 1 Genes overexpressed in human ciliated respiratory cells: complementary DNA fragments mapping to exonic regions

NCBI gene ID	Gene symbol	Description	Location	Maximum 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 defence				
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, 9 kDa	7p21.3	99
58472	*SQRDL*	Sulphide quinone reductase-like (yeast)	15q15	97
29796	*UCRC*	Ubiquinol-cytochrome c reductase complex 7.2 kDa	22cen–q12.3	95
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/sulphate symporters), member 4	7q33	100
11254	*SLC6A14*	Solute carrier family 6 (amino acid transporter), member 14	Xq23–q24	98
Cell signalling 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

NCBI: National Center for Biotechnology Information [20]; mt: mitochondrial; *S. cerevisiae*: *Saccharomyces cerevisiae*. #: between insert and genomic sequence.

Real-time PCR was carried out in a LightCycler System® using the FastStart DNA Master SYBR Green I kit (Roche Applied Science). Reference and target gene primers were obtained from QuantiTect Primer Assays (Qiagen), which contained validated primers sets for the reduced glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), axonemal dynein intermediate chain 1 (*DNAI1*), glycoprotein nmb (*GPNMB*), retinitis pigmentosa GTPase regulator (*RPGR*), chromosome 7 open reading frame 49 (*C7orf49*) and chromosome 14 open reading frame 166 (*C14orf166*) genes. PCR reactions were set up in a total volume of 20 μL, containing 2 μL SYBR Green FastStart reaction mix, 2.4 mM $MgCl_2$, 2 μL 10× primers mix and 2 μL cDNA. The 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 65–95°C to confirm the specificity of the PCR products.

Gene expression level was determined with the comparative threshold method, using the level of the housekeeping gene *GAPDH* as a reference value [21]. The threshold cycle of PCR at which amplified product was first detected (Ct) was determined for the real-time PCR. The corrected Ct (ΔCt) for each determination was then used to calculate the relative n-fold

TABLE 2 Genes overexpressed in human ciliated respiratory cells: complentary DNA fragments mapping to intronic regions

NCBI gene ID	Gene symbol	Description	Location	Maximum 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

NCBI: National Center for Biotechnology Information [20]. #: between insert and genomic sequence; ¶: mapped to two genes (*FAM82B* and *CPNE3*); +: at the exon–intron junction.

differential expression of a specific gene in a ciliated cell compared with a nonciliated cell sample and expressed as the ratio of the $2^{-\Delta Ct}$ values.

Statistical analysis

Data from triplicate experiments are presented as mean±SD. For each target gene, the $2^{-\Delta Ct}$ values of ciliated and nonciliated cells were analysed using an unpaired t-test with the significance set at a p-value of <0.05 for a one-tailed test.

RESULTS

The goal of the present study was to identify genes specifically expressed in human ciliated respiratory cells. To this end, the transcriptome of dedifferentiated epithelial cells was subtracted from that of fully redifferentiated cells. The cDNA RDA procedure used in the present study was closely based on the protocol described by HUBANK and SCHATZ [15], which permits the enrichment of transcripts specifically expressed in a cell type through iterative cycles of amplification/subtraction.

Determination of differentially expressed genes

Of the bacterial clones, 25% contained an insert. All clones with an insert (n=78) were sequenced, including four chimeric clones that each contained two cDNA fragments. Altogether, 82 individual cDNA fragments were identified. Four of these cDNA fragments were recovered several times, and corresponded to the following genes: *GPNMB* (nine clones), zinc finger protein 236 (*ZNF236*; six clones), ribosomal protein large P0 (*RPLP0*; four clones) and ribosomal protein L14 (*RPL14*; two clones). Finally, 65 unique cDNA fragments were identified and mapped to the genomic human sequence by BLASTN. Of these, 53 (82%) cDNA fragments mapped to exonic sequences of known genes, seven (11%) to intronic regions or the boundary of an intron–exon junction and five (7%) to intergenic regions.

A list of the 53 cDNA fragments corresponding to exonic regions is presented in table 1. Genes are 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 also detected. The mitochondrial cluster includes seven genes encoding mitochondrial components. Among these, six are nuclear genes and one is mitochondrial (the mitochondrially encoded cytochrome b gene (*MT-CYB*)). Components of channels, transporters or related proteins are listed in one group, which includes two solute carrier family genes. Genes for cell signalling and signal transduction, such as those encoding spermatogenesis associated 13 (*SPATA13*) and testis enhanced gene transcript (*TEGT*; Bax inhibitor 1), are present.

TABLE 3 Complementary DNA fragments mapping to intergenic regions

Sequence	Description	Location	Maximum identity# %
A	1,939 bp to 5' side: Kruppel-like factor 1 (erythroid; *KLF1*)	19p13	100
	2,130 bp to 3' side: glutaryl-coenzyme A dehydrogenase isoform b precursor (*GCDH*)		
B	13,018 bp to 5' side: apolipoprotein A-V (*APOA5*)	11q23	98
	15,879 bp to 3' side: apolipoprotein A-IV precursor (*APOA4*)		
C	58,077 bp to 5' side: hypothetical protein LOC220134 (*C18orf24*)	18q21.1	97
	213,331 bp to 3' side: mitogen-activated protein kinase 4 (*MAPK4*)		
D	146,030 bp to 5' side: phosphatidylinositol-specific phospholipase C (*PLCXD3*)	5p13.1	100
	74,896 bp to 3' side: 3-oxoacid CoA transferase 1 precursor (*OXCT1*)		
E	757,367 bp to 5' side: methionine adenosyltransferase II, beta isoform 1 (*MAT2B*)	5q34	92
	3,300,072 bp to 3' side: similar to odd Oz/ten-m homologue 2 isoform 5 (*ODZ2*)		

#: between insert and genomic sequence.

TABLE 4 Relative changes in expression of selected genes

	Relative fold change#
DNAI1	3696.61
GPNMB	7.89
RPGR	83.44
C7orf49	2.94
C14orf166	4.92

DNAI1: axonemal dynein intermediate chain 1 gene; *GPNMB*: glycoprotein nmb gene; *RPGR*: retinitis pigmentosa GTPase regulator gene; *C7orf49*: chromosome 7 open reading frame 49 gene; *C14orf166*: chromosome 14 open reading frame 166. #: ratio in ciliated to nonciliated cells.

Several cytoskeletal genes, such as the actin gamma 1 (*ACTG1*), keratin and tropomyosin genes were evidenced. Despite the fact that ciliated cells do not replicate, several cell proliferation genes were found, including the non-metastatic cells 2, protein (NM23B) expressed in gene (*NME2*). Among the remaining genes of table 1, only one, the beta-tubulin gene (*TUBB*), had been previously demonstrated to be implicated in cilia. Finally, two genes corresponding to predicted open reading frames were identified, *C7orf49* and *C14orf166*.

The cDNA fragments with sequences 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 (*XPO6*) includes intronic and exonic sequences. The six other fragments come from intronic regions. One fragment maps to a genomic region in which two genes overlap on opposite strands (family with sequence similarity 82 (*FAM82B*) and copine III (*CPNE3*)). Since the present cloning strategy was not oriented, it is not known whether this fragment is derived from one or the other gene.

The five cDNA fragments that mapped to intergenic regions are presented in table 3. The first two cDNA fragments (A and B) could represent a new 5′ or 3′ exon since the distance from the closest gene is at most 16 kb. By contrast, the last two cDNA fragments (D and E) are so isolated (minimum distance of 75 kb from the nearest gene) that they can only be segments of new genes.

Real-time PCR validation studies

First, in order to validate the present culture system, expression of *DNAI1* and *RPGR*, two genes which are known to be overexpressed in ciliated cells, was analysed. *DNAI1* encodes a component of outer dynein arms in ciliary axonemes and is essential for ciliary function, whereas *RPGR* is specifically expressed in tissues containing cells with cilia or cilia-like organelles, such as retina, lung, cochlea and epithelial cells lining bronchi and sinuses [22]. For data analysis, relative quantification of target gene transcripts was performed for each sample, with normalisation to *GAPDH* expression, since it is an endogenous unregulated gene transcript. By real-time PCR, it was found that these two genes exhibit a 3,696- and 83-fold relative increase, respectively (table 4).

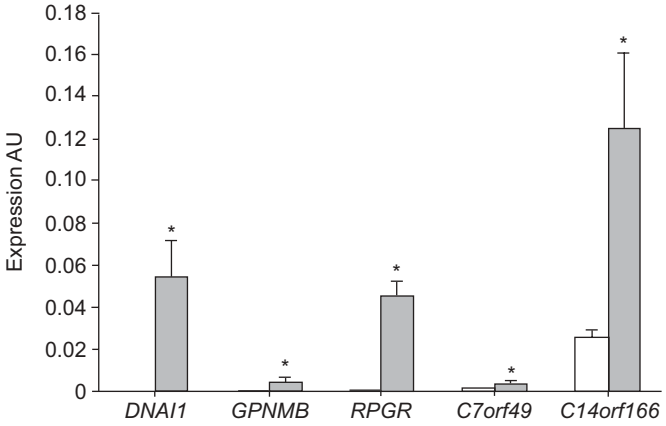

FIGURE 1. Validation of the expression data for a number of the identified human ciliated cell genes: expression in nonciliated (□), and ciliated cells (■). Data are presented as mean±SEM. AU: arbitrary unit; *DNAI1*: axonemal dynein intermediate chain 1 gene; *GPNMB*: glycoprotein nmb gene; *RPGR*: retinitis pigmentosa GTPase regulator gene; *C7orf49*: chromosome 7 open reading frame 49 gene; *C14orf166*: chromosome 14 open reading frame 166. *: $p<0.05$ *versus* nonciliated cells.

Secondly, in order to validate the list of genes, the expression of one gene (*GPNMB*) and two open reading frames, namely *C7orf49* and *C14orf166*, were evaluated by real-time PCR in nonciliated and ciliated cells. The three selected genes exhibited an expression in ciliated cells that was significantly higher than in nonciliated cells (fig. 1). *GPNMB* displayed a relative 7.89-fold increase, a change consistent with the 3.01–12.17-fold change reported by ROSS *et al.* [12]. *C7orf49* and *C14orf166* showed increased expression in human respiratory ciliated cells, with 2.94- and 4.92-fold relative changes, respectively (table 4).

Comparison with other studies on cilia

In order to confirm the present data, the genes list was compared to other lists obtained by various methods from the ciliated cells of various organisms (table 5). Ten out of 53 genes showing increased transcript synthesis in the present study had been reported once in other studies: annexin A1 (*ANXA1*); eukaryotic translation initiation factor 1 (*EIF1*); *GPNMB*; nebulette (*NEBL*); *NME2*; *RPL14*; secretoglobin family 1A member 1 (uteroglobin) (*SCGB1A1*); *SPATA13*; thymosin beta 4, X-linked (*TMSB4X*); and ubiquitin A-52 residue ribosomal protein fusion product 1 (*UBA52*) [1, 3, 11, 12]. Two additional genes were mentioned in the studies of both OSTROWSKI *et al.* [1] and PAZOUR *et al.* [3]: ribosomal protein SA (*RPSA*); and *ACTG1*. *TUBB* was also detected twice [4, 8]. Finally, a computer-predicted gene, referred to as *C14orf166*, was mentioned three times [6–8].

In agreement with the study of PAZOUR *et al.* [3], it was found that FK506 binding protein 1A, 12kDa gene (*FKBP1A*) showed increased expression during ciliogenesis [3]. By contrast, ROSS *et al.* [12] noted decreased expression (-2.33-fold) of this gene. Three genes that showed increased expression during ciliogenesis in the present study showed a reversed pattern in the study of ROSS *et al.* [12]: serpin peptidase inhibitor, clade B (ovalbumin), member 2 (*SERPINB2*; -2.82-fold), tropomyosin 1 (*TPM1*; -3.31/-3.73-fold) and *TPM3* (-3.49/-3.21-fold) [12].

TABLE 5 Genes overexpressed in human ciliated respiratory cells: genes reported in previous studies

Gene	Description	Reference	ID in reference
ACTG1	Actin, gamma 1	[3]	C_1310009 in Cr
		[1]	X03284 in Hs
ANXA1	Annexin A1	[1]	Multiple annexin 1 in Hs
C14orf166	Chromosome 14 open reading frame 166	[6, 8]	168283 in Cr
		[7]	CG31249 in Dm
EIF1	Eukaryotic translation initiation factor 1	[3]	C_190059 in Cr
FKBP1A	FK506 binding protein 1A	[3]	C_230098 in Cr
		[12]#	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]	C_130042 in Cr
		[1]	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]	170055 in Cr
		[4]	158210 in Cr
UBA52	Ubiquitin A-52	[3]	C_1610014 in Cr

Cr: *Chlamydomonas reinhardtii*; Hs: *Homo sapiens*; Dm: *Drosophila melanogaster*. #: decreased expression.

TABLE 6 Genes overexpressed in human ciliated respiratory cells: complementary DNA fragments mapping to loci related to known ciliary structure defects and diseases

Defect	Location#	Reference	Genes at these loci
Familial studies			
ODA	19q13–19qter	[24]	*CXCL17*¶
IDA	15q13.3–15.1	[25]	*SQRDL*¶
IDA	X	[26]	*SLC6A14*¶, *TMSB4X*
Genomic analysis			
PCD/SI	3p	[23]	*RPL14*, *RPL32*¶, *RPSA*
DAD	7p	[23]	*GPNMB*, *NDUFA4*¶
PCD	10p	[23]	*DDIT4*¶, *NEBL*
SI	11q	[23]	*POLD4*¶, *SCGB1A1*, sequence B¶,+
SI	13q	[23]	*SPATA13*
PCD	15q	[23]	*TPM1*
SI/DAD	17q	[23]	*ACTG1*, *ATP5H*¶, *CDC27*¶, *EIF1*, *KRT14*¶, *NME2*

ODA: outer dynein arm; IDA: inner dynein arm; PCD: primary ciliary dyskinesia; SI: situs inversus; DAD: dynein arm defect; *CXCL17*: chemokine (CXC) ligand 17 gene; *SQRDL*: sulphide quinone reductase-like (yeast) gene; *SLC6A14*: solute carrier family 6 (amino acid transporter), member 14 gene; *TMSB4X*: thymosin, beta 4, X-linked gene; *RPL14*: ribosomal protein L14 gene; *RPSA*: ribosomal protein SA gene; *GPNMB*: glycoprotein nmb gene; *NDUFA4*: NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa gene; *DDIT4*: DNA-damage-inducible transcript 4 gene; *NEBL*: nebulette gene; *POLD4*: polymerase (DNA-directed), delta 4 gene; *SCGB1A1*: secretoglobin, family 1A, member 1 (uteroglobin) gene; *SPATA13*: spermatogenesis associated 13 gene; *TPM1*: tropomyosin 1 gene; *ACTG1*: actin, gamma 1 gene; *ATP5H*: ATP synthase, H^+ transporting, mitochondrial F0 complex, subunit d gene; *CDC27*: cell division cycle 27 homologue (*S. cerevisiae*) gene; *EIF1*: eukaryotic translation initiation factor 1 gene; *KRT14*: keratin 14 (epidermolysis bullosa simplex, Dowling–Meara, Koebner) gene; *NME2*: non-metastatic cells 2, protein (NM23B) expressed in gene. #: potential loci indicated for genomic analysis; ¶: not previously reported in studies aimed at identifying specific components of cilia/flagella; +: see table 3.

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 from the present series map to chromosomal regions which may contain a causal gene as determined by a positional cloning approach [23–26]. These 22 genes are presented in table 6. Half of these have never previously been reported in studies aimed at characterising components specific to cilia/flagella. Among these 22 genes, four are located in chromosomal regions showing definite linkage in familial PCD: chemokine (CXC) ligand 17 (*CXCL17*), sulphide quinone reductase-like (yeast) (*SQRDL*), solute carrier family 6 (amino acid transporter), member 14 gene and *TMSB4X* [24–26].

DISCUSSION

In the present study, RDA was used to identify genes differentially expressed in human ciliated respiratory cells. Subtraction of the nonciliated cell representation from that of ciliated cells resulted in the detection of 53 genes and 12 new coding sequences of known genes (n=7) or putative new genes (n=5). These 65 cDNA fragments are truly overexpressed in ciliated cells since a subset of three cDNAs were tested using real-time PCR analysis and all three were expressed at a significantly higher level in ciliated than nonciliated cells. In addition, 18 out of the 53 cDNAs had already been reported in other studies aimed at deciphering transcripts or proteins specific to ciliated cells. The remaining 35 cDNAs were newly identified genes overexpressed in ciliated cells. Of the sequences, 22 were located at loci related to known ciliary structure defects and diseases.

RDA is a method that is long and difficult to set up, but which is recognised as providing no false positives. Consistent with this observation, all of the cDNAs tested by real-time PCR had their overexpression confirmed, and 15 of the cDNAs of the present study had previously been reported in other studies as being overexpressed in ciliated cells. One limitation of this method is that there is selection of short cDNA fragments, meaning that there is no hope of obtaining a complete list of all overexpressed transcripts. Although microarrays provide a systematic view of transcript expression and can only detect preselected sequences, RDA has the power to detect additional expressed sequences either as novel transcript variants of known genes or even as new genes.

In table 1, a set of genes that could be linked to cells with cilia/flagella are reported. Indeed, seven genes encoding proteins of the mitochondrial respiratory chain and oxidative phosphorylation system were identified. In ciliated cells, increased ATP production is presumably necessary for intracellular and intraflagellar transport and ciliary beating. Mutations in *ACTG1*, a major component of sensory ciliated cells of the cochlea, have been described as causing dominant deafness [27]. Finally, *NME2* was demonstrated to be involved in spermiogenesis and flagellar movement [28], and *TMSB4X* was found to be highly represented in lung parenchyma and unrelated tissue types relative to the bronchial epithelium in a previous study [11].

For the real-time PCR analysis, two positive controls that have previously been described as showing increased expression in ciliated tissues, *DNAI1* and *RPGR*, were used. Two other genes, *GPNMB* and *C14orf166*, have been reported, in others studies, to show increased expression in ciliated cells, but their function in ciliogenesis remains to be elucidated. GPNMB, a transmembrane glycoprotein, was hypothesised to be involved in growth delay and reduction of metastatic potential [29], but its role in ciliogenesis remains elusive. The *C14orf166* and *C7orf49* genes were predicted by bioinformatic searches of the human genome. *C14orf166* is frequently detected during ciliogenesis since it has been reported by three other studies. This gene encodes protein involved in the functional regulation of human ninein in the centrosome structure [6–8, 30]. It would be interesting to obtain complete data on *C14orf166* and elucidate the biological function of *C7orf49*.

Finally, five totally new genes (~10% of the whole set) were detected. Further work is warranted to characterise in detail 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 ciliary dysfunction in the early embryo.

ACKNOWLEDGEMENTS

The authors warmly thank S. Picot for the contribution of the Parasitology Laboratory (University of Lyon 1, Lyon, France) to the real-time PCR experiments.

REFERENCES

1 Ostrowski LE, Blackburn K, Radde KM, *et al.* A proteomic analysis of human cilia: identification of novel components. *Mol Cell Proteomics* 2002; 1: 451–465.
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: 909–919.
3 Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. *J Cell Biol* 2005; 170: 103–113.
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: 1090–1098.
5 Broadhead R, Dawe HR, Farr H, *et al.* Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 2006; 440: 224–227.
6 Li JB, Gerdes JM, Haycraft CJ, *et al.* Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 2004; 117: 541–552.
7 Avidor-Reiss T, Maer AM, Koundakjian E, *et al.* Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 2004; 117: 527–539.
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 USA* 2005; 102: 3703–3707.
9 Blacque OE, Perens EA, Boroevich KA, *et al.* Functional genomics of the cilium, a sensory organelle. *Curr Biol* 2005; 15: 935–941.

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: 411–421.
11 Lonergan KM, Chari R, Deleeuw RJ, *et al.* Identification of novel lung genes in bronchial epithelium by serial analysis of gene expression. *Am J Respir Cell Mol Biol* 2006; 35: 651–661.
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: 169–185.
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: 946–951.
15 Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* 1994; 22: 5640–5648.
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: 343–356.
17 UK Medical Research Council Laboratory of Molecular Biology, Staden Package. http://staden.sourceforge.net/ Date last updated: July 5, 2006.
18 Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988; 85: 2444–2448.
19 European Molecular Biology Laboratory–European Bioinformatics Institute, Wellcome Trust Sanger Institute, *Ensembl. www.ensembl.org Date last updated: October 23, 2007. Date last accessed: December*, 6: 2007.
20 National Center for Biotechnology Information, National Center for Biotechnology Information. NCBI Build 36.2. www.ncbi.nlm.nih.gov/ Date last updated: September 14, 2006. Date last accessed: December 6, 2007.
21 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001; 29: e45.
22 Iannaccone A, Breuer DK, Wang XF, *et al.* 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.
23 Blouin JL, Meeks M, Radhakrishna U, *et al.* Primary ciliary dyskinesia: a genome-wide linkage analysis reveals extensive locus heterogeneity. *Eur J Hum Genet* 2000; 8: 109–118.
24 Meeks M, Walne A, Spiden S, *et al.* A locus for primary ciliary dyskinesia maps to chromosome 19q. *J Med Genet* 2000; 37: 241–244.
25 Jeganathan D, Chodhari R, Meeks M, *et al.* 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: 233–240.
26 Krawczynski MR, Witt M. PCD and RP: X-linked inheritance of both disorders? *Pediatr Pulmonol* 2004; 38: 88–89.
27 Rendtorff ND, Zhu M, Fagerheim T, *et al.* 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: 1097–1105.
28 Munier A, Serres C, Kann ML, *et al.* Nm23/NDP kinases in human male germ cells: role in spermiogenesis and sperm motility? *Exp Cell Res* 2003; 289: 295–306.
29 Weterman MA, Ajubi N, van Dinter IM, *et al. nmb*, a novel gene, is expressed in low-metastatic human melanoma cell lines and xenografts. *Int J Cancer* 1995; 60: 73–81.
30 Howng SL, Hsu HC, Cheng TS, *et al.* A novel nineinteraction protein, CGI-99, blocks ninein phosphorylation by GSK3β and is highly expressed in brain tumors. *FEBS Lett* 2004; 566: 162–168.