Tissue expression patterns identify mouse cilia genes

Timothy S. McClintock, Chad E. Glasser, Soma C. Bose, and Daniel A. Bergman

Department of Physiology, University of Kentucky, Lexington, Kentucky

Submitted 15 June 2007; accepted in final form 29 October 2007

McClintock TS, Glasser CE, Bose SC, Bergman DA. Tissue expression patterns identify mouse cilia genes. Physiol Genomics 32: 198-206, 2008. First published October 30, 2007; doi:10.1152/physiolgenomics.00128.2007.-In mammals, cilia are critical for development, sensation, cell signaling, sperm motility, and fluid movement. Defects in cilia are causes of several congenital syndromes, providing additional reasons to identify cilia-related genes. We hypothesized that mRNAs selectively abundant in tissues rich in highly ciliated cells encode cilia proteins. Selective abundance in olfactory epithelium, testes, vomeronasal organ, trachea, and lung proved to be an expression pattern uniquely effective in identifying documented cilia-related genes. Known and suspected cilia-related genes were statistically overrepresented among the 99 genes identified, but the majority encoded proteins of unknown function, thereby predicting new cilia-related proteins. Evidence of expression in a highly ciliated cell, the olfactory sensory neuron, exists for 73 of the genes. In situ hybridization for 17 mRNAs confirmed expression of all 17 in olfactory sensory neurons. Most were also detected in vomeronasal sensory neurons and in neighboring tissues rich in ciliated cells such as respiratory epithelium. Immunoreactivity for one of the proteins identified, Spa17, colocalized with acetylated tubulin in the cilia layer of the olfactory epithelium. In contrast, the ciliary rootlet protein, Crocc, was located in discrete structures whose position was consistent with the dendritic knobs of the olfactory sensory neurons. A compilation of >2,000 mouse genes predicted to encode ciliarelated proteins revealed a strong correlation (R = 0.99) between the number of studies predicting a gene's involvement in cilia and documented evidence of such involvement, a fact that simplifies the selection of genes for further study of the physiology of cilia.

olfaction; bioinformatics; coiled-coil domain; gene expression

NEARLY ALL VERTEBRATE CELLS have a single primary cilium. While its function in many cell types is a mystery, available evidence suggests the broad conclusion that the primary cilium is important for signaling and cell polarity (27, 30). A few cell types have elaborated and expanded on this single cilium, generating specialized structures that perform critical functions of several broad types: sensation, development, fluid movement, sperm motility, and cell signaling. The functional significance of cilia in tissues is reflected in the severity and diversity of pathologies caused by defects in cilia. These include anosmia, retinitis pigmentosa and retinal degeneration, polycystic kidney disease, diabetes, neural tube defects and neural patterning defects, chronic sinusitus and bronchiectasis, obesity, heterotaxias, polydactyly, and infertility (3). This list includes defects in highly specialized cilia, exemplified by retinitis pigmentosa and anosmia, as well as defects in primary cilia, exemplified by polycystic kidney disease (23, 24). Defects in cilia are therefore underlying causes of several congenital disease syndromes with pleiotropic symptoms. These include Alstrom syndrome, Bardet-Biedl syndrome, Kartagener syndrome, Meckel-Gruber syndrome, Senior-Loken syndrome, orofacial digital syndrome 1, Joubert syndrome, and primary cilia dyskinesia (3). Investigation of mammalian cilia genes will enhance not only our knowledge of ciliary function but also of dysfunctions that contribute to significant human health problems.

The function of cilia in cell signaling is evident in the specialization of cilia for sensation. In the mammalian olfactory epithelium, for example, olfactory cilia are elaborated during the maturation of olfactory sensory neurons (OSNs). This event is a defining feature that distinguishes a mature OSN from its immature antecedent (8, 28). These nonmotile cilia are the sites where odorants interact with odorant receptors, triggering G protein-dependent generation of cAMP and subsequent electrical activity in the OSN. Not surprisingly, the bioinformatics of expression profiling of OSN mRNAs consistently identifies cilia/flagella and ciliogenesis/spermatogenesis as overrepresented biological processes in OSNs (26, 29). This link between OSNs and sperm, and our observation that mRNAs linked to cilia were detected in both OSNs and in cells of the neighboring respiratory epithelium, another tissue enriched in highly ciliated cells, led to the hypothesis that tissue expression pattern could be used as an ab initio approach to identify genes whose products are important to cilia.

The ability of bioinformatics and expression profiling to identify cilia-related genes has been documented several times. Comparison of the genomes of ciliated and nonciliated organisms predicts hundreds of cilia-related genes in Chlamydomonas reinhardtii and Drosophila melanogaster (2, 18). The presence of X box promoter elements, likely targets of the cilia-related transcription factors of the Rfx family, are also predictors of cilia-related genes (10, 34). Proteomics on ciliaenriched samples from C. reinhardtii, Tetrahymena aurelia, and cultured human respiratory epithelial cells predict cilia proteins (21, 22, 31). Expression profiling to compare ciliated vs. nonciliated cells in C. elegans or mRNAs that increased during replacement of flagella in C. reinhardtii predict additional cilia-related genes (5, 32). Although previous efforts have been substantial, we reasoned that additional cilia-related genes remain to be identified. This is especially true in mammals, which were the focus of only one of these earlier studies. We therefore compared our data on gene expression in mouse OSNs with gene expression patterns across >60 mouse tissues to predict mouse cilia-related genes (26, 33). We identified 99 putative cilia-related genes, many of which had not previously been linked to cilia and most of which encode proteins of unknown function.

MATERIALS AND METHODS

Bioinformatics. Mouse tissue expression patterns in which mRNAs were most abundant in olfactory epithelium, testes, vomeronasal

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: T. S. McClintock, Dept. of Physiology, Univ. of Kentucky, Lexington, KY 40536-0298 (e-mail: mcclint@uky.edu).

organ, trachea, and lung were identified via SymAtlas, a repository of microarray data on mRNA abundance in many mouse tissues (33). Seven such genes were identified. The correlation function resident in SymAtlas was applied to these seven mRNAs to identify additional mRNAs sharing their expression patterns. Each of the seven lists of correlated mRNAs was ordered by the strength of correlation and then truncated before the first mRNA whose abundance pattern failed to have two of the five target tissues as the tissues with the largest amounts of that mRNA. This threshold corresponded to correlation coefficients that ranged from 0.90 to 0.99 for the seven mRNAs tested. We then combined the seven correlations, tracking how many correlation sets included each mRNA and excluding genes whose mRNAs were found in only one correlation set. As a filter for evidence of expression in highly ciliated cells, we then used expression profiling data generated from separating the olfactory epithelium into purified OSNs, termed the OSN⁺ sample, and a sample of all other cells (the OSN⁻ sample), expecting to delete any genes whose mRNAs lacked evidence of expression in OSNs (26). The filter criterion for exclusion was an OSN^+ sample-to- OSN^- sample (OSN^+/OSN^-) ratio <0.5. Ratios higher than this predict expression in OSNs with accuracies ranging from 83% (ratios of 1.2–0.5) to 99% (ratios \geq 1.3). However, none of the mRNAs failed this criterion. All either had evidence of expression in OSNs or were not represented in the filter criterion data. Previous evidence linking candidate genes to cilia was detected by searching publications on each gene via PubMed and by searching in http://www.ciliome.com in April, 2007 (15). Documented cilia-related genes were identified by searching PubMed using the keyword strategy cili* OR flagel* for the period January, 1996, to April, 2007. Thousands of titles and abstracts were inspected manually to identify those that documented genes directly linked to cilia. Protein domains were identified by use of the simple modular architecture research tool (SMART; http://smart.embl-heidelberg.de/) (17). Kaleidagraph version 3.6 was used to plot data and for curve fitting.

In situ hybridization. Wild-type C57Bl/6J mice of both sexes aged 21 days postnatal through 28 days postnatal (aged P21–P28) were used. All procedures involving animals were approved by the institutional animal care and use committee of the university. In situ hybridizations were performed as described previously (26, 29, 39). Briefly, mice were transcardially perfused with ice-cold 4% paraformaldehyde. The dorsal regions of the snout and anterior cranium were dissected, cryoprotected, stored at -80° C, and cut in 10- to 14-µm

coronal sections on a cryostat. One to two digoxygenin-labeled riboprobes, typically \sim 500 bp in length, were prepared for each mRNA and hybridized in 50% formamide in 10 mM Tris·HCl (pH 8.0), 10% dextran sulfate, $1 \times$ Denhardt's solution, 600 mM NaCl, 0.25% SDS, 1 mM EDTA, and 200 µg/ml yeast tRNA at 65°C (1 ng/µl per riboprobe) on cryosections mounted on slides. Detection was done using an alkaline phosphatase-conjugated antibody to digoxygenin and hydrolysis of nitro blue tetrazolium chloride/5bromo-4-chloro-3'-indolyphosphate *p*-toluidine. Sense and antisense probes were always tested simultaneously, and the sense controls were invariably negative. Digital wide-field images were obtained at room temperature using a Spot 2e camera and Spot software version 4.0.6 on a Nikon Diaphot 300 inverted microscope with a $4 \times /0.13$ numerical aperture (NA) Plan objective and a $40 \times /0.75$ NA Plan Fluor objective. Images were processed in Adobe Photoshop version 7.0 by adjusting size, brightness, and contrast and using the dodging tool to eliminate shadows (only in open areas that lacked tissue) to improve the consistency of illumination. Images were then combined and labeled using Deneba Canvas version 8.0.

Immunohistochemistry. Cryostat sections at 10-µm thickness on slides (Fisher Superfrost Plus) were prepared from wild-type C57Bl/6J mice of both sexes, aged P21-P28, and fixed in 4% paraformaldehyde as described for in situ hybridization above. Slides were dried with a hair dryer for 20 s and washed in PBS for 5 min, PBS containing 1% Triton X-100 for 30 min, and PBS for 5 min. Blocking was then done with 2% BSA and 0.4% Triton X-100 in PBS for 1 h in a hydrated chamber at room temperature. Primary antibodies were a mouse monoclonal against acetylated tubulin (Sigma no. T6793) at 1/1,000 dilution and rabbit polyclonal antisera against Spa17 (16) at 1/200 and Crocc (Root6) (38) at 1/5,000. The specificity of these antibodies and their utility for immunohistochemistry are well documented (16, 37, 38). Slides were incubated in primary antibodies overnight in the blocking solution at 4°C in a hydrated chamber. Washing was in PBS for 10 min, PBS containing 0.5% Tween 20 for 10 min, and PBS for 10 min. Slides were incubated in cyanine-3 (Cy3)-conjugated anti-rabbit secondary antibody in PBS for 1 h in a hydrated chamber at room temperature in the dark. Washing was in PBS for 10 min, PBS containing 0.5% Tween 20 for 10 min, and PBS for 10 min. Slides were then incubated in FITC-conjugated anti-rabbit secondary antibody in PBS for 1 h in a hydrated chamber at room temperature in the dark. Washing was in PBS for 10 min twice in the



Mean microarray signal intensity

Fig. 1. Tissue expression patterns identify mRNAs most abundant in tissues rich in highly ciliated cells. Shown are the 3 most divergent patterns among the 7 mRNAs used for the correlation analysis that identified putative cilia-related genes. The 20 tissues with the highest signal intensities for each mRNA are listed. Testis, olfactory epithelium, vomeronasal organ, trachea, and lung are rich in ciliated cells and are listed first regardless of signal intensity. *A*: Ccdc40 represents mRNAs where signal intensities were highest in the olfactory epithelium. *B*: 4932425124Rik represents mRNAs with low abundance overall but greater expression in tissues with highly ciliated cells. *C*: 2310007A19Rik represents the most common pattern observed, mRNAs that were abundant in tissues rich in highly ciliated cells. Data obtained from SymAtlas (33). Microarray signal intensities reported as gcRMA processed data.

EXPRESSION PATTERNS IDENTIFY CILIA GENES

Table 1. Predicted mouse cilia-related genes, identified by tissue expression patterns

Gene ID	Gene Symbol	Frequency	Ciliome Database?	PubMed ID	ISH	OSN ⁺ /OSN ⁻ Ratio
69329	1700003M02Rik	7	ves		OSN. RE	6.2
66353	2310007A19Rik	7	<i>ye</i> s		0010,112	7.6
77048	4921537D05Rik	7				27
71845	4922501L.14Rik	7			OSN VSN RE	8
75040	4930504H06Rik	7	Ves			53
68283	9530077C05Rik	7	yes			27
223690	Ankrd54	7				2.7
77600	C330001K17Dik	7	VAC			5.7
70604	Dnaih14	7	yes			4.0
75563	Dnaj014	7	VAC	16406424		5.5
220820	Gm770	7	yes	10490424		0.4 nd
330830	$\operatorname{cmf1m}(7624)$ a st	7				iiu nd
75050	giii 11107624_a_at	7		*17020004		nd
75050	K1127	7		*17089004		nd
329104	LOC329164	7				nd
70144	LICH3	7				1.3
/6668	Man1b	/				6.1
/5533	Nme5	/	yes	1(07000)		2.2
69310	Pacrg	7	yes	162/8296		6
21689	Tektl	-	yes	11089920		3.9
74407	Ttc25	7			OSN, RE	110.2
78802	Ttc30a1	7				1.2
237930	Ttll6	7		*15890843		9.9
68625	1110020C03Rik	6				3.5
69315	1700001L19Rik	6	yes		OSN, VSN, RE	12.7
75472	1700009P17Rik	6			OSN, VSN, RE	5.9
227736	1700019L03Rik	6	yes			nd
216516	4930562D19Rik	6			OSN, VSN, RE	4.4
228778	6820408C15Rik	6				1.1
244653	Hydin	6	yes	12719380		nd
67661	Ift172	6	yes	16061793		2.9
50525	Spag6	6	yes	12167721		nd
68777	Tmem53	6	yes			nd
22092	Tsga2	6	yes	16354795		4.5
71860	Wdr16	6	yes			3
66758	Zfp474	6				5.1
73721	1110017D15Rik	5		15242845	OSN, VSN, RE	6.3
67085	1700024G13Rik	5				nd
76420	1700027A23Rik	5			OSN, VSN, RE	10.6
73287	1700040L02Rik	5			OSN, VSN, RE	7.4
74895	4930455F23Rik	5	ves			2.1
75172	4930528G09Rik	5	5			49.6
70977	4931407G18Rik	5				nd
71233	4933434I06Rik	5	ves	15385169	OSN VSN RE	12.5
74934	Armc4	5	ves	16809635		nd
67331	Atp8b3	5	905	10007055		9.2
218232	AW456874	5	Vec			2
210252	Cede113	5	yes		OSN VSN PE	64
51038	Cede30	5	yes		OSN VSN PE	5.7
105822	Code65	5	yes		USIN, VSIN, KE	10.1
207175	Cetted	5	yes	12002050		10.1
20/1/3	Effect	5		12602036		2.4
/18//		5	yes	150/0855		2.9
238/26	Gm276	5				nd
169//	Lrrc23 (Lrpb7)	5	yes			3.5
20686	Spal7	5	yes	15257753		3.7
72236	Tsnaxip1	5	yes			nd
76670	Ttc18	5	yes			3.2
74052	Ttc21a	5	yes			10.7
269701	Wdr66	5	yes			1.2
69234	Zfp688	5				2.4
67483	1700028P14Rik	4				5.9
66706	4733401H18Rik	4				0.8
69312	4930418G15Rik	4	yes			9.1
70882	Armc3	4		16809635		3.4
330323	C330043M08Rik	4				nd
382137	D630004A14Rik	4				nd
331595	Gm777	4				4.3
67694	Ift74 (Ccdc2)	4	yes	15955805	OSN, VSN, RE	2.3
27356	Insl6	4				nd

Continued

Table 1.—*Continued*

Gene ID	Gene Symbol	Frequency	Ciliome Database?	PubMed ID	ISH	OSN ⁺ /OSN ⁻ Ratio
53885	Nphp1	4	yes	12872123		2.4
114602	Zmynd10	4	5			34.5
66061	0610012D17Rik	3	ves			2.3
71837	1700003E16Rik	3	2			nd
75556	1700026D08Rik	3	yes			nd
71864	1700026J04Rik	3	2			24.7
73376	1700061J05Rik	3				8
235345	4833427G06Rik	3				nd
70966	4931415C17Rik	3				nd
207686	A330021E22Rik	3	yes		OSN, VSN	16.3
68644	Abhd14a	3	2			4.1
232566	C730024G19Rik	3				1.1
241116	Ccdc108	3				nd
102124	E130303B06Rik	3	yes			0.9
13516	Epyc (Dspg3)	3	2			nd
381738	Gm1060	3				13.5
210762	Gm70	3				nd
none	gnf1m11188_a_at	3				nd
69325	1700012B09Rik	2			OSN, VSN, RE	43.2
654809	4930428O21Rik	2				3.4
217887	BC022687	2				3
327951	Cyb5d1	2	yes			nd
71200	Dydc2	2	yes			5.5
15223	Foxj1	2		9739041		5.3
18415	Hspa41	2		16923965		3.7
71827	Lrrc34	2			OSN, RE	5.2
54562	Lrrc6	2	yes			4
104362	Meig1	2				6.6
73075	Ppilő	2			OSN, RE	4.2
19725	Rfx2	2		*12403718		nd
66832	Rshl2	2		*2941441		1.1

Frequency, how often the mRNA appeared in the 7 tissue expression pattern correlation gene lists; Ciliome database, whether the gene is listed at http://www.ciliome.com; PubMed ID, a representative publication linking the encoded protein to cilia (*published evidence is indirect or circumstantial); ISH, in situ hybridization results; OSN, olfactory sensory neuron; RE, respiratory epithelium; VSN, vomeronasal sensory neurons; OSN^+/OSN^- ratio, predicts likelihood of expression in OSNs at 99% for ratios ≥ 1.3 and at 83% for ratios between 1.2 and 0.5 (26); nd, not determined.

dark. Hoechst-33258 at a 1/1,000 dilution was applied in PBS for 5 min in the dark, followed by 5 min in PBS and a rinse with distilled water. Slides were dried at room temperature in the dark for 1 h and then mounted with Vectashield (Vector Laboratories) and a coverslip.

Data supplement. A summary of 2,128 mouse genes linked to cilia by previous high-throughput approaches (15), by published studies of individual gene products, and by the work described herein was compiled as an Excel workbook file (Supplemental File 1; supplemental data are available at the online version of this article). The present Entrez gene IDs and gene symbols were verified, and gene ontology annotation was added to the file. Each gene was also searched against PubMed [gene symbol AND (cili* OR flagel*)] for published evidence linking the mouse gene or putative orthologs to cilia or flagella. This searchable file therefore lists identifying information about each gene, functional information about the encoded protein, the number of high-throughput studies linking the gene to cilia, and PubMed ID numbers for at least one study documenting relationships to cilia.

RESULTS AND DISCUSSION

Identification of mouse cilia-related genes by tissue expression pattern. We reasoned that the core elements of cilia and cilia support structures (such as the axoneme, basal body, and intraflagellar transport particles) are enriched in highly ciliated cell types, loosely defined as cells that elaborate large or multiple cilia. We identified seven mRNAs that shared expression in mature OSNs and a subset of cells in neighboring respiratory epithelium and found that they shared a tissue expression pattern of high abundance in tissues with numerous highly ciliated cells: olfactory epithelium, testes, vomeronasal organ, trachea, and lung (Fig. 1). We used these mRNAs to identify 99 mouse genes whose mRNAs had highly correlated tissue expression patterns (Table 1). Of these, 39 are the subject of papers archived in PubMed, and of these 39, 18 are directly linked to cilia or flagella by these publications. This proportion is significant even if we make the cautious prediction that 25% of all genes are linked to cilia (binomial test; H₀:cilia gene fraction = 0.25; $B^* = 3.05$, P < 0.001). The frequency with which mRNAs were detected in the seven tissue expression pattern correlations was not correlated with the proportion with proven links to cilia as determined by articles archived in PubMed. This finding argues that even those mRNAs whose abundance patterns were highly similar to only two of the seven test mRNAs are not less likely to be important to cilia.

Confirmation of expression in OSN, a highly ciliated cell. If the mRNAs we identified do indeed encode proteins important to cilia, then these mRNAs must be expressed in the ciliated cells of these tissues. We tested this by in situ hybridization in the olfactory epithelium where the only highly ciliated cells, mature OSNs, can be unambiguously identified in tissue sections because of the position of their cell bodies (Fig. 2A). Published evidence already indicates that 73 of the mRNAs are expressed in OSNs (Table 1; OSN⁺/OSN⁻ ratios >0.8) (26).

EXPRESSION PATTERNS IDENTIFY CILIA GENES



We tested 17 mRNAs, including three that have proven links to cilia (4933434I06Rik, Ift74, and 1110017D15Rik). All 17 mRNAs were detected in the mature OSN layer of the epithelium. Examples are shown in Fig. 2, B-O, and the data are summarized in Table 1 [column titled ISH (in situ hybridization)]. Thirteen of the mRNAs were also detected in the neurons of the vomeronasal organ (not shown, but summarized in Table 1). These tissue sections also contain neighboring regions with other highly ciliated cells, such as respiratory epithelium and nasal gland linings. Of the 17 mRNAs, only A330021E33Rik failed to unequivocally label subsets of cells in these epithelia. Examples of labeling of respiratory epithelium and nasal glands can be seen in the low-magnification images of Fig. 2, B'-O'. Even though the ciliated cells in respiratory epithelium and nasal glands cannot be unambiguously identified by in situ hybridization because their cell bodies are intermingled with other cell types, the presence of the mRNAs in these epithelia is additional evidence supporting the conclusion drawn directly from the localization of the mRNAs to OSNs and vomeronasal sensory neurons. We conclude that our tissue expression pattern approach identified mRNAs expressed primarily by highly ciliated cells.

The cell and tissue expression patterns of the 99 mRNAs we identified were consistent with the expectation that many of the encoded proteins reside in cilia or cilia support structures. However, some may instead have functions in other cellular compartments, such as nuclei or vesicular trafficking compartments, yet still be important to cilia. None of the novel proteins we linked to cilia is as yet a target of useful antibodies. However, our analyses predicted that Spa17 is present in olfactory cilia. Spa17 has previously been localized to sperm flagella and cilia of the respiratory tract and reproductive organs and is known to be expressed by OSNs, but it has not been localized to the cilia of OSNs (6, 12, 16). Not surprisingly, we observed that Spa17 immunoreactivity in the olfactory epithelium overlaps completely with immunoreactivity for acetylated tubulin, a marker for cilia (Fig. 2P). In contrast, immunoreactivity for the ciliary rootlet protein, Crocc, did not (Fig. 2Q). Crocc is the major component of ciliary rootlets that extend basally from the basal bodies of cilia (37, 38). We observed Crocc immunoreactivity in round structures just below immunoreactivity for acetylated tubulin, consistent with its expected location in the dendritic knobs from which olfactory cilia arise. Immunoreactivity for Spa17 and Crocc in neighboring respiratory epithelium gave similar patterns. Spa17 overlapped completely with acetylated tubulin labeling of cilia, and Crocc immunoreactivity was found directly below the cilia layer (data not shown). These data provide additional reasons to expect that many of the 99 genes we identified will prove to be important to cilia.

Tissue expression pattern predictions overlap with other predictive approaches. As described above, several other methods have been used to predict cilia-related genes with some degree of success. We hypothesized that our method would overlap significantly with these previous efforts. Indeed, 39 of our 99 genes are probable orthologs of genes previously predicted to be related to cilia (15). This proportion is greater than a random sampling of genes should yield (binomial test; H_0 :cilia gene fraction = 0.25; $B^* = 6.37$, P < 0.0001). Of these 39, 27 as yet have no direct evidence of importance to cilia. The mouse genes we identified that are probable orthologs of genes previously linked to cilia were not limited to orthology with mammalian genes. Just six were probable orthologs of genes previously linked to cilia only in humans. Instead, most were orthologs of nonmammalian genes, or of both human and nonmammalian genes.

The basic logic of our approach, an expression pattern largely restricted to tissues with large numbers of highly ciliated cells, appeared to be uniquely consistent with cilia genes. We were unable to find other patterns involving other tissues where cilia have important functions, such as pituitary, retina, kidney, and cartilage/bone (11, 14, 36), that had even a few mRNAs whose abundance was so highly correlated (coefficients >0.9). This tight correlation corresponded with rough estimates of the proportion of highly ciliated cells in the tissues, with testes, olfactory epithelium, vomeronasal organ, trachea, and lung having substantial proportions, whereas these other tissues have relatively few highly ciliated cells, or in the case of the retina and its photoreceptors, have cells whose cilia have diverged into a very specialized structure. The frequency with which mRNAs were detected in the seven tissue expression pattern correlations was not correlated with the proportion of them listed in http://www.ciliome.com (15). This finding is consistent with the expectation that our tissue expression pattern approach does not share biases inherent in previous approaches.

The reliability of our search strategy was better than that of previous methods, judging by the frequency of genes shared with other predictive studies. Genes identified by multiple high-throughput approaches would seem more likely to be true cilia-related genes than those detected only once. The frequency of identification of predicted cilia genes is listed at http://www.ciliome.com, which summarizes previous work in this area (Table 1) (15). Only 22% of the genes listed were identified by multiple studies. For the genes we identified, the proportion was 48%. Overall, the available evidence suggests that the proportion of true cilia-related genes in our list will prove to be substantial.

Many candidates are genes of unknown function. Like previous bioinformatics approaches to identify cilia-related genes,

Fig. 2. Predicted cilia-related mRNAs locate to ciliated cells. A: guide to the cell body layers of the pseudostratified olfactory epithelium, using in situ hybridization for the mature olfactory sensory neuron (OSN)-specific mRNA, olfactory marker protein, to identify the mature OSN layer (mOSN). Mature OSNs are the only cell bodies in this central layer of the epithelium. Apical is up. Sus, sustentacular cell layer; iOSN, immature OSN layer; Basal, basal cell layer; dashed line, basal lamina. *B* and *D*–*O*: In situ hybridization shows labeling in OSNs for each mRNA. Apical is up. *C*: example of the lack of labeling in sense controls. *B'* and D'-O': low-magnification views of the left one-half of coronal sections show mRNA detection in both the thick olfactory epithelium (OE) and the thin respiratory epithelium (RE) within the bilaterally symmetric nasal cavity. The medially located septum (Se) is at *left* in each image, and the olfactory turbinates occupy most of the remaining space. Dorsal is up. *P*: immunoreactivity for Spa17 (red) overlapped significantly (yellow) with immunoreactivity for acetylated tubulin (green), locating Spa17 in the cilia layer of the epithelium. Blue, nuclei. *Inset*: magnified view of immunofluorescence for Spa17 alone. *O*: immunoreactivity for Crocc (red)-labeled round structures consistent with dendritic knobs lying underneath the cilia layer, identified by immunoreactivity for acetylated tubulin (green). *Inset*: magnified view of these structures. Blue, nuclei. Scale bars: *A* and *B*–*O*, 10 µm; *C* and *B'*–*O'*, 200 µm; *P*–*Q*, 20 µm.

our tissue expression pattern approach was not biased by gene type. We detected genes involved in signaling, transcriptional regulation, protein chaperone function, cytoskeletal structure, and cytoskeletal dynamics, all of which are functions needed by cilia. However, what was remarkable about the genes as a group was how little we know about them. For 60 of them, we found no evidence of published work on them or their products. Because our approach was unbiased, these genes could encode products involved in virtually any aspect of cilia function or support of cilia, such as the axonome, basal body, protein trafficking, and transcription of other cilia-related genes. Expression patterns limited to tissues enriched in motile cilia (testes, trachea, lung) were unsuccessful in identifying a different group of highly correlated mRNAs that could be associated specifically with motility. Our search pattern included tissues with both motile and nonmotile cilia, so we conclude that the genes we identified are likely to encode components of processes shared by both motile and nonmotile cilia. These processes are not necessarily restricted to cilia or the basal body but could instead be processes necessary to support cilia that happen at other locations such as the nucleus or the cell body.

Several of the most interesting proteins encoded by the 99 genes we identified have never been studied. Proteins with coiled-coil domains appear to be important to cilia (22). This domain mediates protein interactions, especially with other coiled-coil domain proteins, predicting that additional coiledcoil domain proteins might be important to cilia (7). Ccdc39, Ccdc65, Ccdc113, and 1700003M02Rik are all predicted to possess coiled-coil domains, and all have putative orthologs in other species that have been linked to cilia by at least one other high-throughput approach (15). Similarly, tetratricopeptide repeat domain (Ttc) proteins and WD (tryptophan aspartic acid) repeat domain proteins have been linked to cilia (1, 4, 9, 13, 20, 25). We identified five such genes, three of which have putative orthologs that have been tentatively linked to cilia in other species: Ttc18, Ttc21a, and Wdr66. Last, 0610012D17Rik encodes a novel protein that is predicted to contain a Tctex-1 domain. Tctex-1 is a dynein light chain that interacts with rhodopsin, implicating it in the trafficking of rhodopsin to the photoreceptor outer segment (35). 0610012D17Rik may therefore be a component of a more general mechanism of trafficking proteins to cilia.

A mouse ciliome resource. Our prediction that the identification of mammalian cilia genes is incomplete was supported not only by our prediction of additional mouse cilia genes but also by published work on genes that were not detected by high-throughput approaches. We detected several proteins with published links to cilia, such as Foxi1 (19), whose genes were absent from http://www.ciliome.com (Table 1) (15). A literature search for mouse and human cilia-related genes revealed 87 genes that have documented importance to cilia but were not detected by high-throughput approaches. Surprisingly, 62% of the mouse genes with documented importance to cilia were not predicted by previous high-throughput methods, nor were probable orthologs of these genes. Combining the genes identified herein (Table 1) with those identified in our literature search and those listed in http://www.ciliome.com identifies 2,128 mouse genes that have been linked to cilia in some fashion (Supplemental File 1).

We expect this mouse ciliome resource to be neither complete, as discussed above, nor wholly accurate. For example, a few genes in our results may be selectively abundant in olfactory epithelium, vomeronasal organ, testes, trachea, and lung for reasons unrelated to cilia. More broadly based bioinformatics and expression profiling approaches are even more likely to make this type of false inclusion error. Similarly, the proteomics approach suffers from limited sensitivity, may misidentify some proteins, and is prone to include proteins from contaminating subcellular fractions. In fact, >75% of the genes predicted as important to cilia by high-throughput approaches were identified by only one of the eight previous studies (Fig. 3), a statistic that argues that many cilia gene predictions are erroneous. Users should therefore keep in mind that the mouse ciliome resource suffers from omissions of cilia-related genes and from inclusion of genes not truly important to cilia and may be complicated by the identification of genes important to cilia in some organisms and tissues but not others.

Even with these imperfections, the compilation of documented and predicted cilia-related genes provides a valuable resource. As the pace of prediction of cilia-related genes has far outstripped the ability to test these predictions, significant opportunities to investigate proteins critical for cilia function and dysfunction have arisen. Supplemental File 1 provides a rational basis for identifying which proteins are the best candidates for as yet undocumented roles in cilia or the support of cilia. We hypothesized that the number of studies linking a gene to cilia would correlate with the frequency of genes independently validated as encoding cilia proteins. Indeed, we found a significant correlation within the exponential relationship between these factors (Fig. 3). Genes identified by multiple approaches are therefore the most promising candidates



Fig. 3. Frequency of prediction correlates with documented evidence. Experimental evidence of a gene's involvement in cilia is correlated with the no. of predictive high-throughput studies that linked the gene to cilia. This relationship is described by an exponential function (equation and correlation coefficient displayed). Nos., the no. of genes associated with each data point. Nos. on the abscissa derived from http://www.ciliome.com.

for future studies to test whether the encoded proteins are sufficient and necessary for cilia function.

Conclusions

The tissue expression pattern we discovered proved to be significantly better than chance at identifying known and suspected cilia-related genes. It identified 99 mRNAs, of which 48 have some link to cilia, ranging from documented involvement in cilia to prediction by previous high-throughput studies. Most of the 99 genes encode proteins that have never been studied, suggesting that much remains to be learned about cilia. The correlation of documented importance to cilia with the frequency of prediction of links to cilia indicates that our approach was more reliable than previous approaches, but this came at the expense of limiting the number of genes identified. The mouse ciliome resource we compiled from nine different sources contains 2,128 mouse genes with documented or predicted importance to cilia. Users of this resource should understand that additional mammalian cilia-related genes remain to be discovered, and that some proportion of the predictions will prove to be erroneous. Genes identified in multiple studies are the most promising for future study of cilia function and dysfunction.

ACKNOWLEDGMENTS

We thank Drs. Michael O'Rand and Tiansen Li for the gift of antisera against Spa17 and Crocc, respectively.

Present address of D. A. Bergman: Dept. of Biomedical Sciences, Grand Valley State Univ., 130 Padnos Hall, Allendale, MI 49401.

GRANTS

This work was supported by National Institute of Deafness and Other Communications Disorders Grants R01-DC-002736 and R01-DC-007194 and Boyarsky Professorship funds to T. S. McClintock.

REFERENCES

- Ahmed NT, Mitchell DR. ODA16p, a Chlamydomonas flagellar protein needed for dynein assembly. *Mol Biol Cell* 16: 5004–5012, 2005.
- 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* 117: 527–539, 2004.
- Badano JL, Mitsuma N, Beales PL, Katsanis N. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* 7: 125–148, 2006.
- Blacque OE, Li C, Inglis PN, Esmail MA, Ou G, Mah AK, Baillie DL, Scholey JM, Leroux MR. The WD repeat-containing protein IFTA-1 is required for retrograde intraflagellar transport. *Mol Biol Cell* 17: 5053– 5062, 2006.
- 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* 15: 935–941, 2005.
- Bumm K, Grizzi F, Franceschini B, Koch M, Iro H, Wurm J, Ceva-Grimaldi G, Dimmler A, Cobos E, Dioguardi N, Sinha UK, Kast WM, Chiriva-Internati M. Sperm protein 17 expression defines 2 subsets of primary esthesioneuroblastoma. *Hum Pathol* 36: 1289–1293, 2005.
- Burkhard P, Stetefeld J, Strelkov SV. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 11: 82–88, 2001.
- 8. Cuschieri A, Bannister LH. The development of the olfactory mucosa in the mouse: electron microscopy. *J Anat* 119: 471–498, 1975.
- Efimenko E, Blacque OE, Ou G, Haycraft CJ, Yoder BK, Scholey JM, Leroux MR, Swoboda P. Caenorhabditis elegans DYF-2, an orthologue of human WDR19, is a component of the intraflagellar transport machinery in sensory cilia. *Mol Biol Cell* 17: 4801–4811, 2006.

- Efimenko E, Bubb K, Mak HY, Holzman T, Leroux MR, Ruvkun G, Thomas JH, Swoboda P. Analysis of xbx genes in C. elegans. *Development* 132: 1923–1934, 2005.
- Fliegauf M, Horvath J, von Schnakenburg C, Olbrich H, Muller D, Thumfart J, Schermer B, Pazour GJ, Neumann HP, Zentgraf H, Benzing T, Omran H. Nephrocystin specifically localizes to the transition zone of renal and respiratory cilia and photoreceptor connecting cilia. *J Am Soc Nephrol* 17: 2424–2433, 2006.
- Grizzi F, Chiriva-Internati M, Franceschini B, Bumm K, Colombo P, Ciccarelli M, Donetti E, Gagliano N, Hermonat PL, Bright RK, Gioia M, Dioguardi N, Kast WM. Sperm protein 17 is expressed in human somatic ciliated epithelia. J Histochem Cytochem 52: 549–554, 2004.
- Haycraft CJ, Swoboda P, Taulman PD, Thomas JH, Yoder BK. The C. elegans homolog of the murine cystic kidney disease gene Tg737 functions in a ciliogenic pathway and is disrupted in osm-5 mutant worms. *Development* 128: 1493–1505, 2001.
- Haycraft CJ, Zhang Q, Song B, Jackson WS, Detloff PJ, Serra R, Yoder BK. Intraflagellar transport is essential for endochondral bone formation. *Development* 134: 307–316, 2007.
- Inglis PN, Boroevich KA, Leroux MR. Piecing together a ciliome. Trends Genet 22: 491–500, 2006.
- Lea IA, Widgren EE, O'Rand MG. Association of sperm protein 17 with A-kinase anchoring protein 3 in flagella. *Reprod Biol Endocrinol* 2: 57, 2004.
- Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 32: D142–D144, 2004.
- 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* 117: 541–552, 2004.
- Murphy DB, Seemann S, Wiese S, Kirschner R, Grzeschik KH, Thies U. The human hepatocyte nuclear factor 3/fork head gene FKHL13: genomic structure and pattern of expression. *Genomics* 40: 462–469, 1997.
- Ogawa K, Inaba K. Ap58: a novel in situ outer dynein arm-binding protein. *Biochem Biophys Res Commun* 343: 385–390, 2006.
- 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* 1: 451–465, 2002.
- Pazour GJ, Agrin N, Leszyk J, Witman GB. Proteomic analysis of a eukaryotic cilium. J Cell Biol 170: 103–113, 2005.
- Pazour GJ, Baker SA, Deane JA, Cole DG, Dickert BL, Rosenbaum JL, Witman GB, Besharse JC. The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance. J Cell Biol 157: 103–113, 2002.
- Pazour GJ, Dickert BL, Vucica Y, Seeley ES, Rosenbaum JL, Witman GB, Cole DG. Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J Cell Biol* 151: 709–718, 2000.
- Pedersen LB, Rompolas P, Christensen ST, Rosenbaum JL, King SM. The lissencephaly protein Lis1 is present in motile mammalian cilia and requires outer arm dynein for targeting to Chlamydomonas flagella. *J Cell Sci* 120: 858–867, 2007.
- Sammeta N, Yu TT, Bose SC, McClintock TS. Mouse olfactory sensory neurons express 10,000 genes. J Comp Neurol 502: 1138–1156, 2007.
- Scholey JM, Anderson KV. Intraflagellar transport and cilium-based signaling. *Cell* 125: 439–442, 2006.
- Schwarzenbacher K, Fleischer J, Breer H. Formation and maturation of olfactory cilia monitored by odorant receptor-specific antibodies. *Histochem Cell Biol* 123: 419–428, 2005.
- Shetty RS, Bose SC, Nickell MD, McIntyre JC, Hardin DH, Harris AM, McClintock TS. Transcriptional changes during neuronal death and replacement in the olfactory epithelium. *Mol Cell Neurosci* 30: 90–107, 2005.
- Singla V, Reiter JF. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* 313: 629–633, 2006.
- 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 4: 909–919, 2005.

206

- 32. 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 102: 3703–3707, 2005.
- 33. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101: 6062–6067, 2004.
- Swoboda P, Adler HT, Thomas JH. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol Cell* 5: 411–421, 2000.
- Tai AW, Chuang JZ, Bode C, Wolfrum U, Sung CH. Rhodopsin's carboxy-terminal cytoplasmic tail acts as a membrane receptor for cytoplasmic dynein by binding to the dynein light chain Tctex-1. *Cell* 97: 877–887, 1999.
- Watanabe D, Saijoh Y, Nonaka S, Sasaki G, Ikawa Y, Yokoyama T, Hamada H. The left-right determinant Inversin is a component of node monocilia and other 9+0 cilia. *Development* 130: 1725–1734, 2003.
- 37. Yang J, Gao J, Adamian M, Wen XH, Pawlyk B, Zhang L, Sanderson MJ, Zuo J, Makino CL, Li T. The ciliary rootlet maintains long-term stability of sensory cilia. *Mol Cell Biol* 25: 4129–4137, 2005.
- Yang J, Liu X, Yue G, Adamian M, Bulgakov O, Li T. Rootletin, a novel coiled-coil protein, is a structural component of the ciliary rootlet. *J Cell Biol* 159: 431–440, 2002.
- Yu TT, McIntyre JC, Bose SC, Hardin D, Owen MC, McClintock TS. Differentially expressed transcripts from phenotypically identified olfactory sensory neurons. J Comp Neurol 483: 251–262, 2005.

