



HAL
open science

Novel approach combining transcriptional and evolutionary signatures to identify new multiciliation genes

Audrey Defosset, Dorine Merlat, Laetitia Poidevin, Yannis Nevers, Arnaud Kress, Olivier Poch, Odile Lecompte

► To cite this version:

Audrey Defosset, Dorine Merlat, Laetitia Poidevin, Yannis Nevers, Arnaud Kress, et al.. Novel approach combining transcriptional and evolutionary signatures to identify new multiciliation genes. *Genes*, 2021, 10.3390/genes12091452 . hal-03402882

HAL Id: hal-03402882

<https://hal.science/hal-03402882>

Submitted on 25 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

Novel Approach Combining Transcriptional and Evolutionary Signatures to Identify New Multiciliation Genes

Audrey Defosset ¹ , Dorine Merlat ¹ , Laetitia Poidevin ¹ , Yannis Nevers ^{2,3,4}, Arnaud Kress ¹ , Olivier Poch ¹ and Odile Lecompte ^{1,*} 

- ¹ Department of Computer Science, ICube, UMR 7357, Centre de Recherche en Biomédecine de Strasbourg, University of Strasbourg, CNRS, 67000 Strasbourg, France; adefosset@etu.unistra.fr (A.D.); dorine.merlat@etu.unistra.fr (D.M.); l.poidevin@unistra.fr (L.P.); akress@unistra.fr (A.K.); olivier.poch@unistra.fr (O.P.)
- ² Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland; yannis.nevers@unil.ch
- ³ Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
- ⁴ Swiss Institute for Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland
- * Correspondence: Odile.Lecompte@unistra.fr

Abstract: Multiciliogenesis is a complex process that allows the generation of hundreds of motile cilia on the surface of specialized cells, to create fluid flow across epithelial surfaces. Dysfunction of human multiciliated cells is associated with diseases of the brain, airway and reproductive tracts. Despite recent efforts to characterize the transcriptional events responsible for the differentiation of multiciliated cells, a lot of actors remain to be identified. In this work, we capitalize on the ever-growing quantity of high-throughput data to search for new candidate genes involved in multiciliation. After performing a large-scale screening using 10 transcriptomics datasets dedicated to multiciliation, we established a specific evolutionary signature involving *Otomorpha* fish to use as a criterion to select the most likely targets. Combining both approaches highlighted a list of 114 potential multiciliated candidates. We characterized these genes first by generating protein interaction networks, which showed various clusters of ciliated and multiciliated genes, and then by computing phylogenetic profiles. In the end, we selected 11 poorly characterized genes that seem like particularly promising multiciliated candidates. By combining functional and comparative genomics methods, we developed a novel type of approach to study biological processes and identify new promising candidates linked to that process.

Keywords: multiciliation; rare diseases; comparative genomics; evolution; functional genomics; multi-omics data integration



Citation: Defosset, A.; Merlat, D.; Poidevin, L.; Nevers, Y.; Kress, A.; Poch, O.; Lecompte, O. Novel Approach Combining Transcriptional and Evolutionary Signatures to Identify New Multiciliation Genes. *Genes* **2021**, *12*, 1452. <https://doi.org/10.3390/genes12091452>

Academic Editor: Laura Crisponi

Received: 26 August 2021

Accepted: 18 September 2021

Published: 21 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cilia are ancient and complex organelles found on the surface of most eukaryotic cells, in the form of a single non-motile projection used as a sensory device capable of transducing signals during development and homeostasis (reviewed in [1]). Some specialized cells, however, can generate several dozens of motile cilia on their surface that beat in a coordinated manner to generate a directional fluid flow used to circulate liquids or give movement to particles [2]. In mammals, particularly in humans, multiciliated cells (MCCs) are found in the spinal cord and the brain ventricles, where they direct cerebrospinal fluid flow in the respiratory tract and play an important role in mucus clearance, as well as in both male and female reproductive tracts (reviewed in [3]). MCCs have also frequently been studied in other vertebrate species, such as in *Xenopus laevis* tadpoles that possess a mucociliary epidermis [4], and in *Danio rerio*, where they can be found in the pronephros [5].

Multiciliation is a process found throughout evolution, but its exact presence or absence in all eukaryotic clades has not yet been clearly established. Multiple accounts of

MCCs in Metazoa have been given in the literature, although some unknowns remain; they have been observed in non-bilaterian species, in some Protostomes but not in others, with a clear absence in Ecdysozoa, as well as in most subphyla of Deuterostomia [6] (Figure 1). While MCCs have been observed in various unicellular species such as *Multicilia marina*, *Tetrahymena* and *Paramecium* [7–9] as well as in sperm of certain plants [10,11], it is unclear whether the mechanisms responsible for the generation of multiple motile cilia are the same, though it seems unlikely.

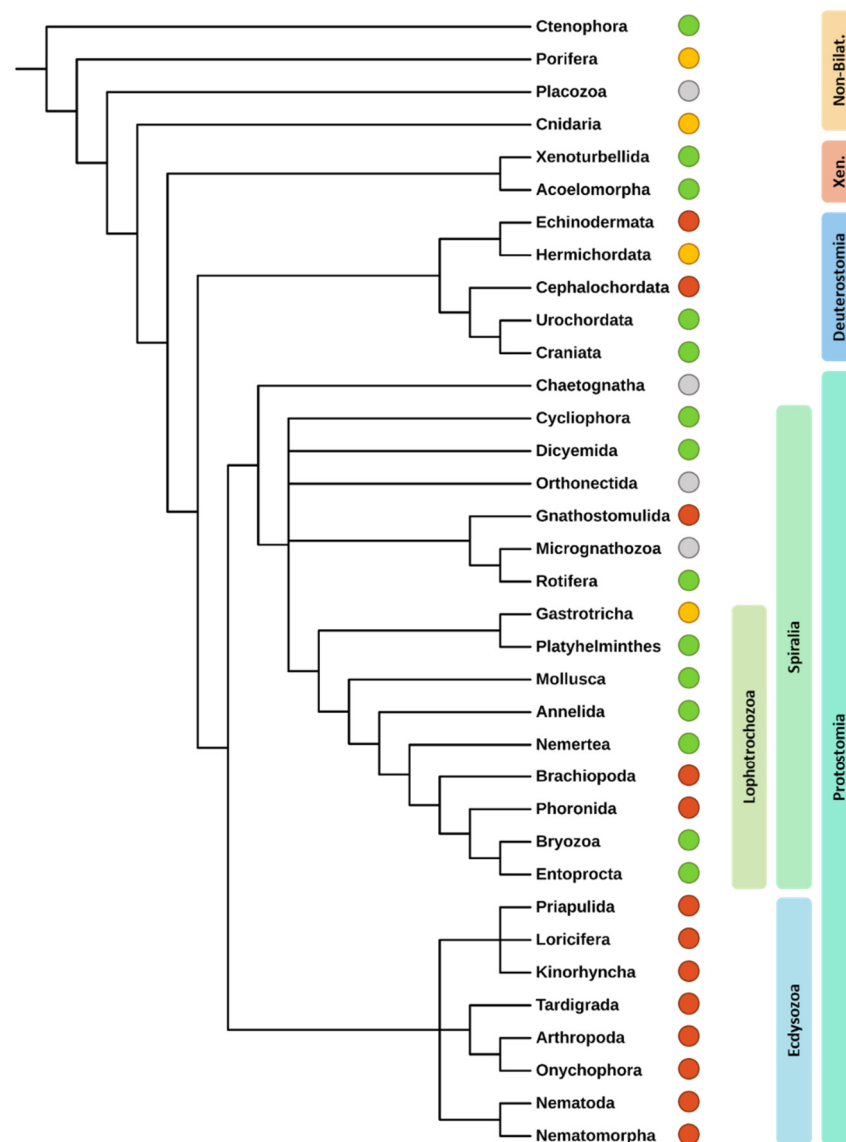


Figure 1. Distribution of MCCs in different metazoan clades. A green circle indicates the presence of MCCs, a red circle indicates their absence. A yellow circle indicates the existence of both multiciliated species and strictly monociliated species in the same branch. Gray circles indicate a lack of information regarding multiciliation. Xen.: xenacoelomorpha; Non-Bilat: non-bilaterian. Phylogeny based on the results of Giribet [12].

Over the years, studies have allowed us to get a better understanding of the transcriptional cascade involved in the differentiation of MCCs in vertebrates (Figure 2, reviewed in [13]). The roles of the Notch signaling pathway are numerous during development; in multiciliated epithelia, various studies have shown its function as a determinant of the cellular fate, either into a secretory cell in the presence of Notch, or into a multiciliated cell in its absence [14–16]. One of the earliest steps of the multiciliated pathway identified

involves a family of micro-RNAs, the *miR-34/449* family, which contains six members in mammals, three of which are hosted by the same gene, *CDC20B*. The role of these miRNAs is to promote the exit from the cell cycle and inhibit Notch1 receptors, thus provoking the sealing of the multiciliated fate in the cell [17]. STK11, a serine/threonine kinase, has recently been identified as a necessary actor to permit differentiation of MCCs, presumably by inhibiting cell proliferation through a signaling cascade involving MARK3, ERK1/2, although its relation with other multiciliated determinants remains to be clarified [18].

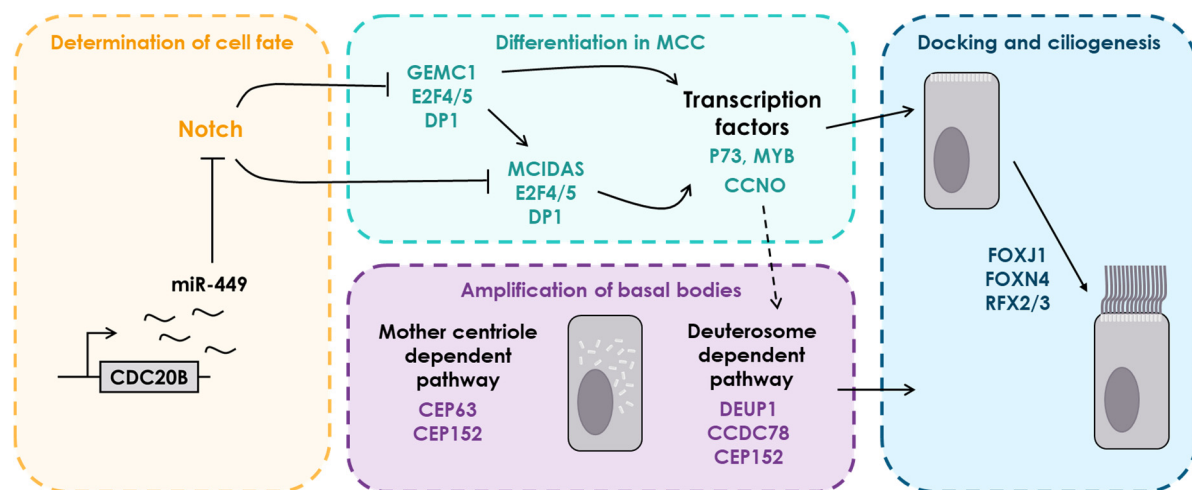


Figure 2. Schematic representation of the main stages of multiciliogenesis, with major genes. Flat-ended arrows depict the repression of a gene, regular arrows depict the activation of a gene.

Once the multiciliated fate has been determined, the inhibition of *Notch* allows the activation of the two central mediators of MCCs differentiation: GEMC1, encoded by *GMNC*, and multicilin/MCIDAS, encoded by *MCIDAS*, members of the geminin protein family, characterized by the presence of coiled-coil domains [19–22]. Both indispensable for the development of MCCs, GEMC1 and MCIDAS bind to E2F4/5 and DP1 to activate various transcription factors necessary for the steps leading to the generation of multiple motile cilia [22,23]. Among MCIDAS and GEMC1 targets are *p73* and *MYB*, two other important regulators of multiciliation also known to activate transcription factors of the multiciliated pathway [24–26].

One of the most important and challenging aspects of multiciliogenesis is the efficient generation and docking of several dozens of basal bodies (BB) from which cilia will later develop. So far, two pathways have been identified in BB amplification: the mother centriole dependent (MCD) pathway, and the deuterosome dependent (DD) pathway [27]. However, this step of multiciliogenesis remains blurry; it was previously thought that the DD pathway accounted for up to 95% of BBs, but recent studies have shown that the absence of either of both amplification pathways did not prevent the generation of procentrioles [28,29].

The canonical MCD pathway depends on two centrosomal proteins, CEP63 and CEP152, to form a ring like structure around the mother centriole from which new centrioles will then stem. The DD pathway relies on CEP63 paralog DEUP1 and CEP152 for de novo BB generation, through a ring-shaped structure called the deuterosome. CEP63 and DEUP1 both associate with CEP152 in a mutually exclusive manner, and in *X. laevis* it was shown that DEUP1 requires the presence of CCDC78 to recruit Cep152 and form new centrioles [30]. To allow for the adequate formation of deuterosomes in terms of number, structure and shape, CCNO, an MDICAS target, is also required, although its exact role remains to be determined [31].

Essential to BB migration, docking and ciliogenesis, FOXJ1 is one of the first transcription factors identified as necessary to the development of MCCs, notably through its

large number of target genes [32]. Since then, several studies have shown that its activity could be influenced by RFX2 and RFX3, and that FOXN4 shares a non-negligible number of targets with FOXJ1. Together, these four transcription factors control the expression of numerous genes involved in BB docking and ciliogenesis.

Despite the growing number of works dedicated to the study of multiciliation, quite a few gray areas remain regarding the transcriptional events responsible for the differentiation of MCCs. So far, only a few genes have been associated with anomalies of multiciliation, responsible of pathologies characterized by chronic respiratory infections from a young age, hydrocephalus and, in all likelihood, sterility [33]. As of yet, only MCIDAS [34], CCNO [35] and FOXJ1 [36] have been identified as causative of reduced generation of motile cilia (RGMC).

To identify new candidate genes involved in multiciliogenesis, we capitalized on the ever-growing quantity of biological data available to realize a large-scale screening of potential genes using several genomics techniques. We analyzed ten transcriptomics datasets of experiments dedicated to multiciliation to look for genes overexpressed in MCCs and combined our results with comparative genomics in an unprecedented study. We identified a specific evolutionary pattern of multiciliation involving a group of bony fish, the Otomorpha (*D. rerio*, *Astyanax mexicanus*, etc.), that allowed us to prioritize target genes that were brought out in our transcriptomics analysis. The integration of both functional and comparative genomics allowed us to identify 11 new candidates that appear particularly promising.

2. Materials and Methods

2.1. Functional Genomics Analysis

2.1.1. Public Transcriptomic Datasets

Eight datasets comprised of results from ten transcriptomics experiments specific to multiciliation (Table 1) were retrieved from the Gene Expression Omnibus (GEO) database of the NCBI [37] for our functional genomics analysis. The selected datasets are either microarray or RNAseq experiments conducted on *X. laevis* or *Mus musculus* and consist in the inactivation of genes involved in the currently known multiciliation pathway.

Table 1. Overview of the different datasets and experiments used for the functional analysis.

GEO Accession	Overall Experimental Design (Multiciliation vs. No Multiciliation)	Species	Experiment Type
GSE32452 [19]	Notch intracellular domain (ICD) + glucocorticoid inducible Multicilin vs. ICD	<i>Xenopus laevis</i>	Microarray
GSE59309 [23]	Inducible Multicilin vs. inducible Multicilin + truncated E2F4	<i>X. laevis</i>	RNASeq
GSE89271 [38]	Inducible Multicilin vs. inducible Multicilin + Foxn4 morpholino	<i>X. laevis</i>	RNASeq
	Inducible Multicilin vs. inducible Multicilin + CRISPR/Cas9 Foxn4 mutant	<i>X. laevis</i>	RNASeq
GSE76342 [39]	Inducible Multicilin vs. inducible Multicilin + CRISPR/Cas9 Foxj1 mutant	<i>X. laevis</i>	RNASeq
	Notch- vs. ICD; ICD vs. ICD + Multicilin; Notch- vs. Notch- + Multicilin-	<i>X. laevis</i>	RNASeq
GSE60365 [40]	Non-targeted shRNA vs. Myb shRNA	<i>Mus musculus</i>	Microarray
GSE75715 [25]	Wild Type vs. p73 knockout	<i>M. musculus</i>	RNASeq
GSE73331 [41]	Wild Type vs. E2F4 knockout	<i>M. musculus</i>	Microarray
GSE116690 [18]	Stk11+ vs. Stk11-	<i>M. musculus</i>	RNASeq

GSE32452. In this experiment, Stubbs et al. compare gene expression between *X. laevis* embryo in which Notch intracellular domain (ICD) was injected and embryos in which both ICD and a glucocorticoid inducible multicilin were injected [19].

GSE59309. In this experiment, authors compare gene expression in epithelial progenitors of *Xenopus* embryos in which multiciliation was induced by multicilin either with wild type E2F4 or a truncated form of E2F4 missing the last 140 amino acids [23].

GSE89271. This dataset includes data from 3 experiments aimed at the characterization of *Foxn4* and its role in multiciliation compared to *Foxj1*. Thus, gene expression of control embryos with glucocorticoid inducible multicilin was compared with three states: (1) embryos injected with glucocorticoid inducible multicilin and *Foxn4* morpholino, (2) embryos injected with glucocorticoid inducible multicilin and CRISPR/Cas9 system aimed at *Foxn4* or (3) embryos injected with glucocorticoid inducible multicilin and CRISPR/Cas9 aimed at *Foxj1* [38].

GSE76342. Quigley and Kintner compare gene expression in epithelial progenitors of *Xenopus* at three timepoints (3, 6 and 9 h) in three pairs of conditions: (1) Notch blocking compared to injection of ICD, (2) injection of ICD compared to injection of ICD and multicilin and (3) *Notch* blocking compared to *Notch* blocking with inactive multicilin [39].

GSE60365. In this experiment, authors compare gene expression in mouse tracheal cells when injected with control shRNA or with *Myb* blocking shRNA [40].

GSE75715. Authors compare wild type individuals with *p73* knockout mice [25].

GSE73331. Mori et al. compare wild type mice with *E2F4* knockout mice [41].

GSE116690. In this dataset, authors compare gene expression in lung cells of wild type mouse with cells of mice in which *STK11* was deleted in lung progenitors using the Cre-lox system [18].

2.1.2. Raw Data Analysis

When available, the overexpression results provided by the authors were used directly (GSE89271, GSE76342, GSE78715 and GSE116690), otherwise, raw data were analyzed with the GEO2R tool of the NCBI [37] (GSE32452, GSE60365, GSE73331) and genes significantly overexpressed in a multiciliated state were retrieved. In the case of the GSE59309 dataset, the raw data were not directly analyzable with GEO2R, we thus followed the same method the authors describe in the manuscript and used the DESeq2 R package [42] to analyze each time point individually. For all datasets, the logFC threshold was set as ≥ 1 and as ≤ 0.05 for the *p*-value.

2.1.3. Overexpressed Genes Clustering

Before comparing overexpressed genes of the different datasets previously cited, we had to homogenize sequences identifiers, as the datasets come from different platforms and species. We thus used orthologous human gene names retrieved with the Mouse Genome Informatics database (<http://www.informatics.jax.org/> (accessed on 26 November 2020)) and Xenbase [43] for mouse and xenopus sequences respectively.

We generated a binary matrix with the results of the 10 experiments indicating the presence (1) or the absence (0) of overexpression for each gene (as rows) in each experiment (as columns). Pairwise distance between gene profiles was calculated using the *dist* function of the *amap* R package, using the 'binary' method (<https://CRAN.R-project.org/package=amap> (accessed on 2 December 2020)). Hierarchical clustering was then carried out using the *hclust* R function with the 'ward.D2' method. Finally, gene clusters were defined through dynamic trimming of the resulting dendrogram using the *cuttreeDynamic* function of the *dynamicTreeCut* R package (<https://CRAN.R-project.org/package=dynamicTreeCut> (accessed on 4 December 2020)) using a depth of 2 and the 'hybrid' cut method.

2.2. Comparative Genomics Analysis

2.2.1. Evolutionary Analysis of Multiciliated Genes

For the evolutionary study of known multiciliated genes, we used the NCBI implementation of BLASTp [44] with default parameters to search for orthologous sequences, using the human sequence as a query. When necessary, gene absence was confirmed using tBLASTn on the latest release of the relevant species' genome, with default parameters and a word length of 3. Multiple sequence alignments of proteins were computed using the PipeAlign2 tool (<http://www.lbgi.fr/pipealign> (accessed on 12 January 2021)) [45], using MAFFT as the main alignment software [46] and manually analyzed. Genomic context

of selected genes was analyzed on the Ensembl platform [47] or the NCBI sequence viewer [48].

2.2.2. Search for Atypically Conserved Genes in Otomorpha

To search proteins with a marked sequence divergence in Otomorpha, we used BLUR [49], a tool designed by the authors with the purpose of comparing groups of species to identified unexpected variations in sequence. We compared Otomorpha fish with species of the clade Euteleostomorpha, another group of bony fish. We used the human proteome as a reference and chose to work with orthologs.

2.3. In Depth Analysis of Target Genes

2.3.1. Generation of Interaction Networks

We used the STRING protein interaction database [50] to construct our interaction networks and retrieved all genes that had interactions with at least 2 genes of our target list. We used a threshold score of 0.7 for interactions between the genes of our list, and 0.9 for interactions between genes of our list and external genes. We used Cytoscape [51] in combination with the GLayer community clustering algorithm [52] implementation in the clusterMaker2 Cytoscape plugin [53].

2.3.2. Computation and Clustering of Phylogenetic Profiles

OrthoInspector 3.0 [54] was used to retrieve orthologs of our target genes in 711 eukaryotic species, with which we generated a matrix containing the presence or absence of each gene in those species. Once the phylogenetic profiles were constructed, we used the *correlation* method of the *amap* R-package. Hierarchical clustering was then carried out using the *hclust* R function with the 'ward.D2' method and trimming was done using the *cuttreeDynamic* function of the *dynamicTreeCut* R package with a depth of 2 and the 'hybrid' cut method (see Section 2.1.3).

3. Results

To expand our current knowledge regarding the generation and maintenance of MCCs, we combined functional genomics and comparative genomics data in order to highlight genes that were likely to be involved in this process. We first looked at expression data of multiciliation oriented transcriptomics data, then used specific evolutionary patterns to pinpoint the most probable target genes.

3.1. Functional Analysis of Multiciliation Oriented Experiments

The first step of our study was the analysis of transcriptomics data from experiments designed to compare the multiciliated state and the non multiciliated state of cells. In this step, we looked for genes that were overexpressed in a multiciliated state in more than one experiment.

3.1.1. Gene Overexpression

To identify genes overexpressed in MCCs, we selected 8 transcriptomics datasets comprising 10 experimental designs aimed at comparing MCCs to cells in which multiciliation was altered by inactivating genes known to be involved in the generation of MCCs (Table 1, see Section 2.1.1). These experiments can be grossly classified in two types depending on when the studied gene acts in the pathway: 'early' genes, responsible for the overall regulation of multiciliation (*STK11*, *p73*, *MCIDAS* and *E2F4*) and 'late' genes, responsible for the activation of multiple motile cilia generation (*Myb*, *Foxn4* and *Foxj1*).

For each dataset, we retrieved genes overexpressed in multiciliated condition using a threshold of 1 for logFC and 0.05 for *p*-value, for a total of 4151 genes differentially expressed in at least one experiment. We compiled these results in a two-dimensional binary matrix with each gene as a row and each experiment as a column, and in each cell, the value 1 is entered if the gene is overexpressed, and 0 if it is not.

3.1.2. Gene Clustering

To further highlight genes of interest, we carried out a clustering of the various expression results using the Jaccard index applied to each gene profile to measure the distance between them, as well as the Ward algorithm [55] to perform the clustering (see Section 2.1.3). We obtained 28 clusters with 10 containing genes overexpressed in only one experiment, for a total of 2278 genes that will not be taken in consideration for the rest of our study (Figure 3).

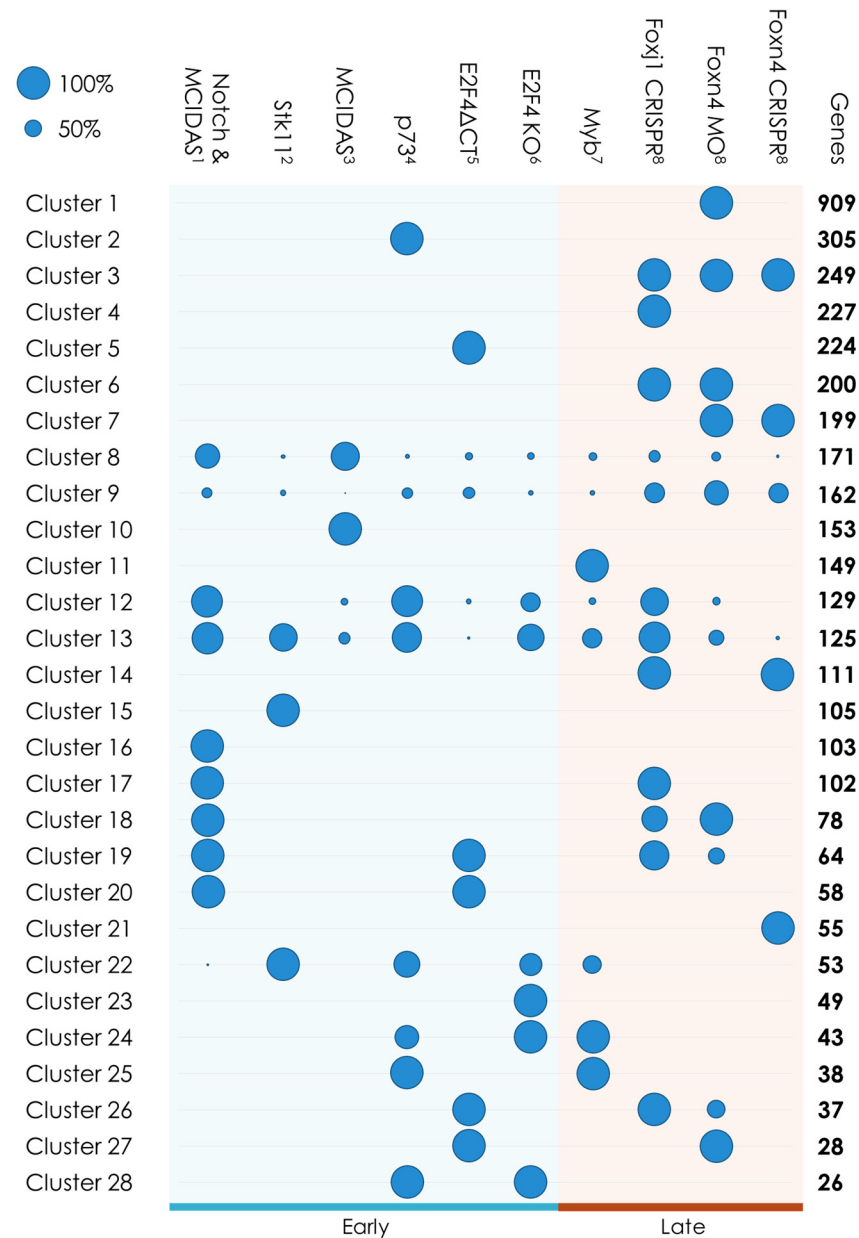


Figure 3. Schematic representation of the clustering of the results from the transcriptomics datasets. Circle size is proportional to the percentage of the cluster's genes found in each dataset. ¹GSE76342; ²GSE116690; ³GSE32452; ⁴GSE75715; ⁵GSE59309; ⁶GSE73331; ⁷GSE60365; ⁸GSE89271. 'Early' genes (*STK11*, *p73*, *MCIDAS* and *E2F4*) are responsible for the overall regulation of multiciliation. 'Late' genes (*Myb*, *Foxn4* and *Foxj1*) activate the generation of motile cilia.

The 18 remaining clusters can be regrouped according to the similarity of their expression profile and Panther [56] was used to perform Gene Ontology (GO) [57] term enrichments on each group of clusters to identify associated biological processes (Table 2).

Table 2. Cluster regrouping according to similarity in profiles. Pertinent GO terms were selected among the ‘biological process’ GO annotation dataset and the results with the lowest *p*-value.

Clusters	Characteristics	Genes	GO Terms	<i>p</i> -Value
3, 6, 7, 14	Foxj1/Foxn4 targets	759	‘establishment of localization’	1.64×10^{-13}
17, 18, 19, 26, 27	MCIDAS/E2F4 complex and Foxj1 or Foxn4 targets	308	‘cilium organization’	8.28×10^{-23}
22, 24, 25, 28	Mouse genes	160	‘cilium movement’	1.32×10^{-9}
20	MCIDAS/E2F4 complex targets	58	‘centrosome cycle’	1.30×10^{-13}
8, 9, 12, 13	Multiciliated clusters	587	‘cilium organization’	1.94×10^{-92}

Clusters 3, 6, 7 and 14 contain ‘late’ genes that seem to be under the influence of Foxn4 and Foxj1 whose roles, as we mentioned before, are largely redundant. GO term enrichments reveal an overrepresentation of genes involved in the establishment of the localization of biological components, which may both be linked to the numerous roles played by the FOX protein family as well as the large need for vesicular transport mechanisms during ciliogenesis [58].

Clusters 17, 18, 19, 26 and 27 hold genes influenced by both the MCIDAS/E2F4 complex and one or both of its targets, Foxn4 and Foxj1. These clusters are largely enriched in genes linked to cilia related GO terms such as ‘cilium organization’ (*p*-value: 8.28×10^{-23}) or ‘cilium assembly’ (*p*-value: 1.26×10^{-20}). Interestingly, 44 out of the 308 genes found in those clusters are not associated with any GO terms in *Homo sapiens* or *M. musculus*.

Some clusters (22, 24, 25 and 28) contain genes influenced by a combination of STK11, p73, E2F4 and Myb, but more importantly, these genes only result from experiments done on the mouse. That can either be due to a specific regulation of multiciliation only found in mice, or to the absence of orthologs of these genes in *X. laevis*.

Cluster 20 contains genes influenced only by a variation in the expression of MCIDAS and E2F4 and that seem mostly involved in multiplication of basal bodies, with enrichment in GO terms such as ‘centriole replication’ (*p*-value: 1.02×10^{-12}), ‘centriole assembly’ (*p*-value: 2.58×10^{-12}) or ‘centrosome duplication’ (*p*-value: 1.22×10^{-11}).

Finally, the last clusters (8, 9, 12, 13) contain genes influenced by a large number of multiciliated factors and, as such, these clusters also contain the majority of genes known to be involved in this process. Thus, MCIDAS, DEUP1, CEP152, PLK4, MYB, CCP110, CDC20B, FOXJ1 and FOXN4 can be found in cluster 8, CCNO and CCDC78 are in cluster 13, and RFX2 and RFX3 are in cluster 12. In addition to a strong enrichment in GO terms linked to cilia, these clusters contain 60 genes for which no GO annotation can be found in human orthologs.

3.2. Comparative Genomics

The first step of our study generated a large list of potential genes linked to multiciliation, however in most cases, the dissociation of ciliation and multiciliation is impossible. To overcome this limitation, we used comparative genomics to identify and exploit a specific evolutionary pattern of multiciliation found in some species of bony fish, the Otomorpha.

3.2.1. Identification of Evolutionary Pattern

To begin our search for a multiciliation-specific evolutionary pattern, we identified orthologs of the following proteins in Metazoa using BlastP and human sequences as a query on the RefSeq protein database [59]: CEP63, DEUP1, MCIDAS, GEMC1, CCNO, CCDC78, E2F4, E2F5, CEP152 and CDC20B. We established the phylogenetic distribution of each gene in selected metazoan organisms to identify a phylogenetic profile linked to MCCs but most of the genes studied have different and seemingly complex evolutionary histories which did not enable us to construct a unique evolutionary signature (Figure 4). Some genes like CCDC78, CCNO, MCIDAS, GMNC and CDC20B seem to have appeared more recently and are mostly limited to Vertebrata, while the others seem to be older and present orthologs in most Metazoa. Interestingly, no orthologs were found in Ecdysozoa

except for E2F4 and E2F5, which is consistent with the absence of MCCs in those species, and CDC20B is absent in a group of bony fish, the Otomorpha. These absences were verified at the genomic level by TBLASTN searches.

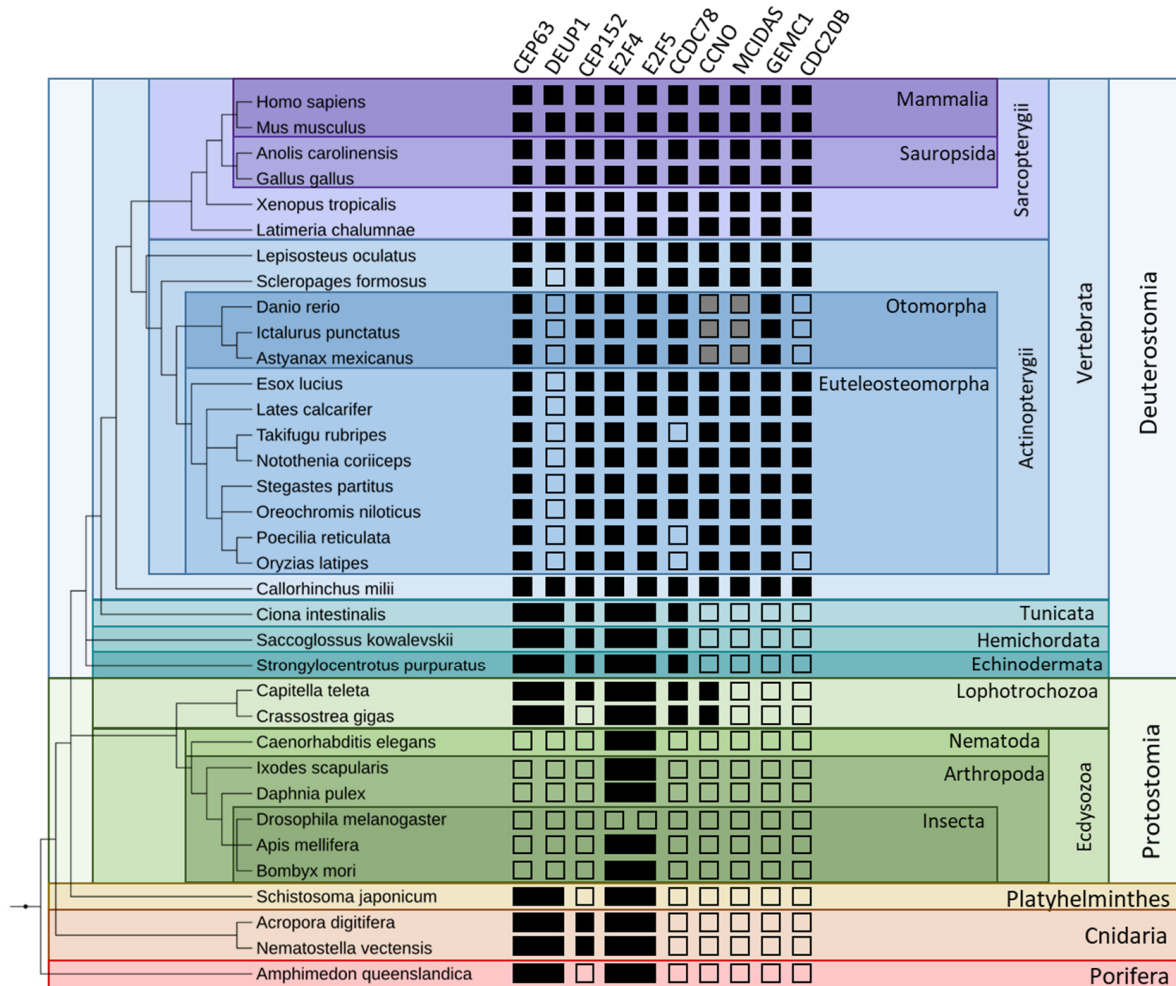


Figure 4. Phylogenetic distribution of multiciliated genes in Metazoa. Only a selected pertinent subset of the studied species is represented. A black square indicates the presence of an ortholog in the concerned species, a blank square indicates the absence of an ortholog. Black rectangles indicate the presence of an ancestral gene before its split in two paralogs. Gray squares indicate the presence of a homologue with an abnormally divergent sequence.

We then focused on the conservation of orthologous sequences through evolution at the subprotein level by analyzing protein multiple sequences alignments. While most of the proteins are relatively well conserved in all taxa studied, two proteins stand out with a particular sequence variation in specific bony fish species of the Otomorpha clade: MCIDAS, which was previously shown in zebrafish [60] and CCNO.

In the case of CCNO, Otomorpha fish proteins present divergences along the whole sequence when compared to mammals or other bony fishes, with the N-terminal portion of the sequence truncated (Figure 5). As for MCIDAS, both the coiled-coil domain, used for interaction with GEMC1, Geminin and MCIDAS itself, and the TIRT domain, used for interaction with E2F4 and E2F5 [23] are relatively well conserved in most species, including Otomorpha. The rest of the protein sequence, however, is largely divergent in the latter species, with an almost absent N-terminal domain, while being overall similar between other bony fish species and Sarcopterygii.



Figure 5. Overview of CCNO multiple sequence alignment. Framed areas show parts of the alignment with the most divergence in Otomorpha fish when compared to other species. Oto.: Otomorpha; Actino.: Actinopterygii.

Lastly, we looked at the genomic localization of multiciliated genes, studying a well-known locus conserved in humans, mice and frogs containing *MCIDAS*, *CCNO* and *CDC20B* [19]. We focused our search on bony fish to see if this syntenic block could be seen beyond tetrapods and found that the co-localization of *MCIDAS*, *CCNO* and *CDC20B* exists in *Latimeria chalumnae*, certain species of Euteleostomorpha fish, as well as in *Callorhynchus milii*, a cartilaginous fish. Interestingly, this gene co-localization is known to be broken in zebrafish [60], but our study showed that this extends to other Otomorpha fish, with *MCIDAS* on one chromosome and *CCNO* on another, and *CDC20B* gone entirely (Figure 6).

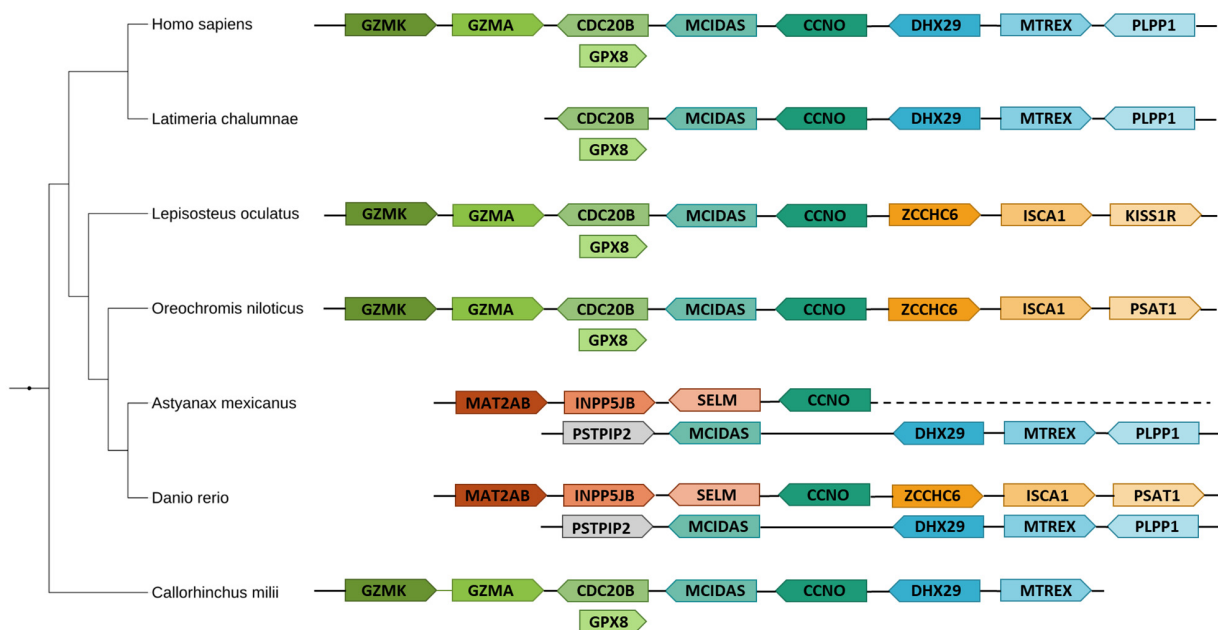


Figure 6. Multiciliated locus in selected vertebrate species. The locus contains *MCIDAS*, *CCNO* and *CDC20B*, three genes essential for multiciliation. This syntenic block is conserved in most vertebrate species except in Otomorpha fish (*Danio rerio*, *Astyanax mexicanus*).

All these results point towards complex evolutionary events that seem to have happened in Otomorpha fish that lead to the loss of *CDC20B* and the *mir-449* family, as well as a possible sequence variation in *CCNO* and *MCIDAS*.

3.2.2. Multi-Level Identification of Differential Conservation

Our evolutionary study showed the existence of gene loss as well as protein sequence divergence in several of the key genes of multiciliation in the Otomorpha group of fish. Based on these results, we searched for other genes that would either be missing in Otomorpha while being present in other bony fish, or genes that would have diverged in an unexpected way in Otomorpha, in order to identify new multiciliated candidates. We used BLUR [49], a tool designed to detect differential conservation in specific species both at the protein and sub-protein level, and compared Otomorpha to Euteleostomorpha using the human proteome as a reference.

Out of the 21,044 proteins found in the human proteome, 1361 present an unexpected behavior in Otomorpha when compared to Euteleostomorpha fish; 634 have no orthologs, 104 are highly likely to present a sequence divergence in Otomorpha, and 623 are mildly likely to present a divergence according to BLUR. Interestingly, *MCIDAS* is found among the highly likely targets. GO term enrichment of the 1361 proteins showed a slight overrepresentation of proteins linked to immunity, suggesting that more than one process might be divergent between Otomorpha and Euteleostomorpha fish and thus making it important to cross these results with the ones stemming from our functional genomics analysis.

3.3. Integration of Functional and Comparative Genomics Results

With the aim of distinguishing genes linked to ciliation and genes specific to multiciliation, we used comparative genomics to identify a multiciliation-specific evolutionary pattern in the form of unexpected gene divergences in Otomorpha fish, to use as a filter on target genes found with our transcriptomics study. The integration of both approaches allowed us to generate a list of 114 genes (Supplementary Table S1) overexpressed in multiciliated conditions, that are either absent in Otomorpha fish or show an unexpected sequence divergence when compared to Euteleostomorpha fish. Among the 114 targets, 41 are absent in Otomorpha, 10 are highly likely to be differentially conserved in Otomorpha and 63 are mildly likely to present an atypical sequence divergence in Otomorpha.

To further analyze our target genes and have a comprehensive picture of multiciliation, we added eight known multiciliated genes to our list (*CCNO*, *GEMC1*, *CEP63*, *CEP152*, *E2F4*, *E3F5*, *CCDC78* and *CDC20B*) and used two characterization approaches: one using functional interaction networks and one using phylogenetic profiling.

3.3.1. Functional Interaction Networks

To characterize our target genes, we used functional interaction networks with the hypothesis that proteins that interact together tend to be involved in the same biological process. With this in mind, we broadened our gene selection to build a sizeable network that allowed us to identify functional clusters. We used the STRING protein interaction database to select additional proteins that interacted with at least two of the proteins of our 122 targets, thus creating a network of 919 genes, including 57 from our list, divided into 10 clusters (Figure 7).

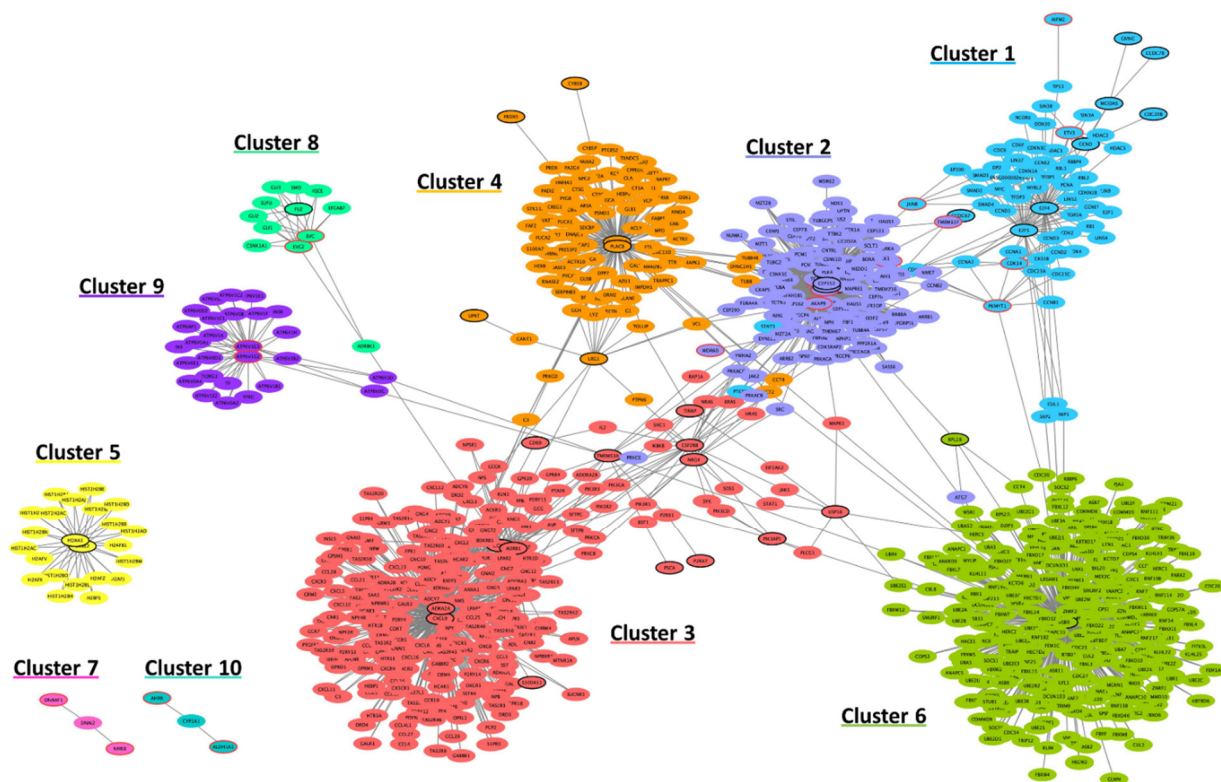


Figure 7. Interaction network computed from the list of 114 potential multiciliated targets as well as genes presenting a functional interaction with at least two genes of that list. These clusters contain 57 of our genes (circled in bold) and a total of 919 genes.

We performed GO enrichment analyses for each cluster and found that six out of ten clusters are either directly or indirectly linked to multiciliogenesis or, more broadly, to ciliogenesis. Cluster 1, while being enriched in proteins related to the regulation of the cell cycle (p -value: 1.53×10^{-34}), contains eight proteins currently known to be involved in multiciliation (GEMC1, CCDC78, MCIDAS, CDC20B, CCNO, E2F4, E2F5 and DEUP1). Clusters 2 and 8 both show an enrichment in proteins localized to the centrosome (p -value: 2.75×10^{-108}) and cilium (p -value: 7.26×10^{-16}) respectively, and as such contains proteins that could be involved in ciliation as well as in multiciliation. Cluster 9 is enriched in GO terms related to autophagy (p -value: 4.22×10^{-32}), a mechanism involved in centrosome number regulation [61]. Finally, clusters 7 and 10, although containing only 3 proteins each, are respectively enriched in the terms ‘cilium movement’ (p -value: 4.18×10^{-7}) and ‘retinoid metabolic process’ (p -value: 1.87×10^{-5}), the latter of which might seem unrelated to multiciliation, but recent studies have shown that retinoic acid is involved in the regulation of multiciliogenesis in *D. rerio* [62]. Those six clusters contain 32 of the 57 target genes, among which 20 are either of unknown function, or are currently unrelated to cilia (Supplementary Table S1).

3.3.2. Phylogenetic Profiling of Multiciliated Targets

We mentioned above that multiciliation is a complex process with an even more complex evolutionary history, with specific losses in some clades such as the Ecdysozoa and a fair number of species for which information is still lacking. With this in mind, we used phylogenetic profiling and clustering to identify genes amongst our targets with profiles that correlate the most with multiciliation. To permit a more comprehensive clustering and a better view of multiciliary evolution, we reused the genes of our previous interaction network analysis.

We used OrthoInspector to generate phylogenetic profiles for our 984 genes in 711 eukaryotic species and clustered them based on the distance between the profiles

(Figure 8). Nine clusters were generated, of which only four are of interest, with a phylogenetic distribution possibly related to cilia or multiciliation. Cluster 5 contains genes absent in non-ciliated fungi and plants as well as in nematodes, which are also non-ciliated. Cluster 7 contains genes present mainly in Metazoa apart from nematodes as well as several genes known to be involved in multiciliation (*CEP152*, *PLK4*, *CCP110*) and cluster 8 contains genes present in chordates, but more interestingly, some of the major genes of multiciliation (*GMNC*, *CCNO*, *MCIDAS*, *CEP63*). Finally, cluster 9 contains genes present in vertebrates but absent in Otomorpha fish, such as *CDC20B*. In total, those 4 clusters contain 514 genes, of which 87 come from our target list of 122 genes, including 76 that have no known role in multiciliation (Supplementary Table S1).

3.3.3. Identification of Most Promising Candidates

Taken together, the results of the various analyses we performed allow us to highlight several candidate genes of unknown function that seem most likely to be involved in multiciliation: *C1orf189* (chromosome 1 open reading frame 189), *C20orf85* (chromosome 20 open reading frame 85), *C5orf24* (chromosome 5 open reading frame 24), *KIAA1841*, *FAM181A* (family with sequence similarity 181 member 1), *IQCK* (IQ domain-containing protein K), *LRRC43* (leucine-rich repeat-containing 43), *DYDC1* (DPY30 domain-containing protein 1), *CFAP47* (cilia and flagella associated protein 47), *ANKRD60* (ankyrin repeat domain-containing protein 60) and *TEX43* (testis expressed protein 43) (Table 3).

Table 3. Most promising target genes highlighted by all methods.

Gene	Transcriptomics Cluster	BLUR	STRING Cluster	Profiling Cluster
<i>C1orf189</i>	13	Absent Otomorpha	-	9
<i>C20orf85</i>	18	Absent Otomorpha	-	9
<i>C5orf24</i>	3	Mildly likely divergence	-	8
<i>KIAA1841</i>	19	Mildly likely divergence	-	5
<i>FAM181A</i>	9	Highly likely divergence	-	8
<i>IQCK</i>	12	Mildly likely divergence	-	7
<i>LRRC43</i>	22	Mildly likely divergence	-	8
<i>DYDC1</i>	8	Mildly likely divergence	-	8
<i>CFAP47</i>	12	Absent Otomorpha	-	5
<i>ANKRD60</i>	17	Absent Otomorpha	-	9
<i>TEX43</i>	18	Absent Otomorpha	-	9

The 11 genes we selected all show a clear lack of characterization: we indeed noted a general absence of GO annotations in their human orthologs, as well as missing information regarding their protein interaction leading to their absence in our STRING analysis network. These genes all belong to phylogenetic profile clusters that could be linked to cilia or multiciliation, with the majority in clusters 8 or 9, both of which contain multiple known genes involved in multiciliation (*CCNO*, *GMNC*, *MCIDAS*, *CDC20B*, *DEUP1*, *CCDC78*, etc.). Five of those genes, *C1orf189*, *FAM181A*, *IQCK*, *DYDC1* and *CFAP47*, are of particular interest, as they were found in the ‘multiciliated clusters’ of our transcriptomics analysis. As a reminder, those clusters contain genes influenced by the most factors, including the main multiciliated genes (*MCIDAS*, *DEUP1*, *CEP152*, *PLK4*, *MYB*, *CDC20B*, *FOXJ1*, *FOXN4*, *CCNO*, etc.). Interestingly, all candidate genes, with the exception of *C5orf24*, have been localized to either ciliated cells, photoreceptors or testis in the Human Protein Atlas [63]. Moreover, *C20orf85* has been linked to lung cancer [64] and *DYDC1* seems to be involved in acrosome biogenesis [65], although *IQCK* has recently been highlighted in several studies regarding Alzheimer’s disease [66,67]. Combining all this information led us to believe these genes to be particularly interesting new targets for future in-depth studies.

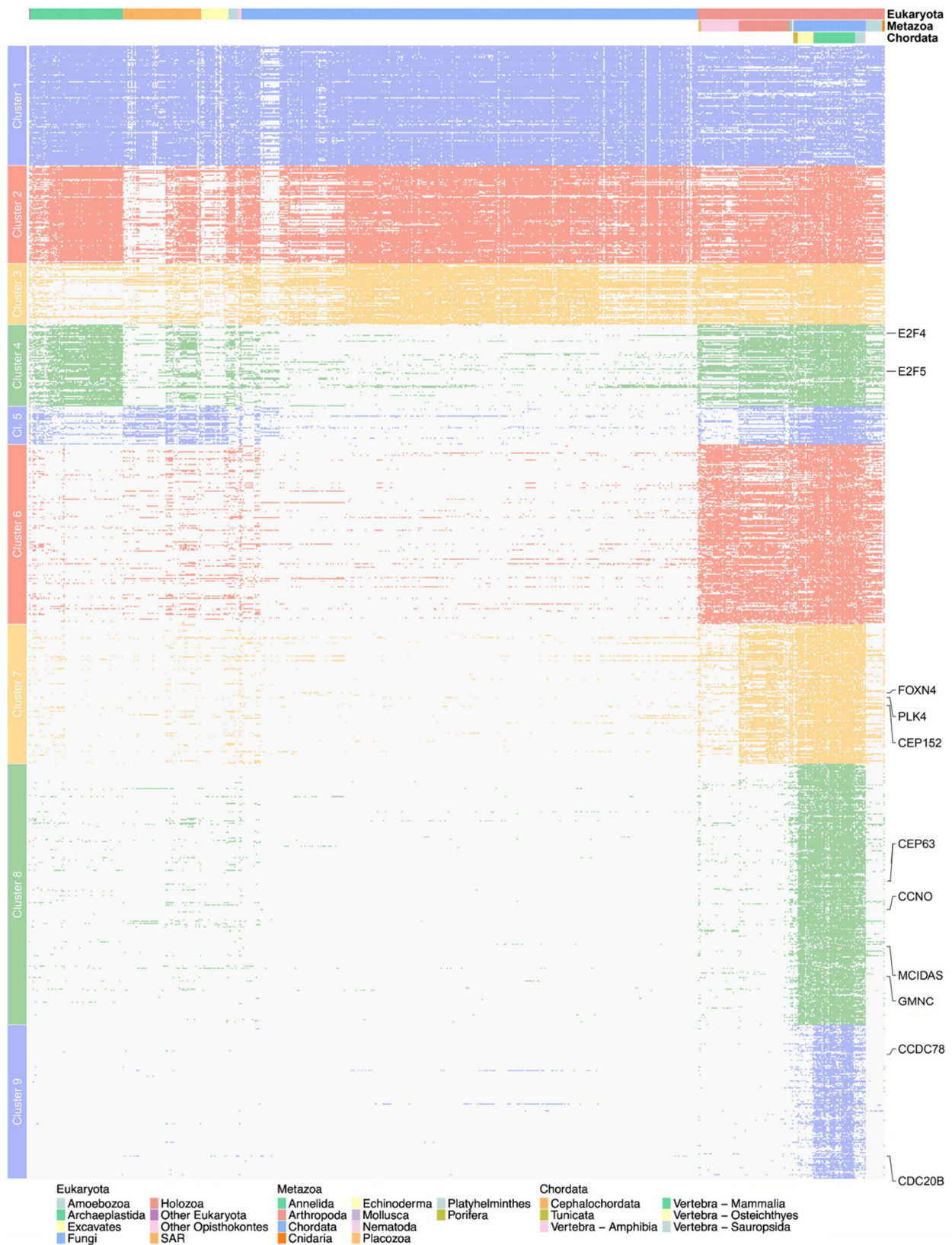


Figure 8. Evolutionary clustering of 984 human genes based on their phylogenetic distribution in 711 eukaryotic species. Rows represent genes and columns represent species, with each colored square indicating the presence of an ortholog of the concerned gene in the chosen species. The three lines above the profiles indicate the main clades in Eukaryota, Metazoa and Chordata respectively.

4. Discussion

Multiciliogenesis is a complex process involving numerous actors and our current knowledge about its mechanisms still leave a lot of gray areas that we need to shed light upon as shown by the existence of only eight genes associated with multiciliation in the GO database. Our goal was to identify new multiciliated genes to expand our understanding, using several genomics methods and combining them.

In this study, we have done a large-scale screening of genes using 10 transcriptomics datasets dedicated to the analysis of multiciliation, which, to our knowledge, has never been done before. This allowed us to identify 1873 genes that are overexpressed in multiciliated conditions in at least two different experiments, including 148 with no GO annotations in humans, but also raised an important and frequent problem in the study of multiciliation, which is its tight link to ciliogenesis and the difficulty to separate both processes. It led us to the use of an original comparative genomics approach as an additional criterion to distinguish between multiciliogenesis and ciliogenesis.

We analyzed a selection of 10 genes known to be involved in multiciliation under the light of evolution, which not only allowed us to add clarity and detail to the current knowledge but also showed the existence of a particular evolutionary pattern linked to multiciliation. Our analysis of the ‘multiciliated locus’ containing *MCIDAS*, *CCNO* and *CDC20B* showed that, while it is not maintained in *Otomorpha* fish, it is found in other Tetrapods, *L. chalumnae* and *C. milii*, suggesting that it might have been present in the beginning of vertebrate evolution. The rupture of the synteny block in *Otomorpha* as well as the absence of *CDC20B* shows the existence of a likely complex evolutionary history in these species. A possible consequence of those variations could be that, in *Otomorpha*, *MCIDAS* has become facultative for the development of MCCs in some organs, as shown recently by Zhou et al. [60]. Another probable impact of sequence divergence of *MCIDAS* and *CCNO* in those species is the apparent reduction of cilia on the surface of their MCCs; observations count less than 16 cilia in pronephric ducts of *D. rerio* [5,68], as opposed to the several hundred generally seen in most other species.

The marked variations in the protein sequences of *MCIDAS* and *CCNO* observed in *Otomorpha* as well as the absence of *CDC20B* led us to search for other genes that would follow the same evolutionary patterns. The signature that we associated with multiciliation is very subtle, much more so than that of cilia, for which a classical profiling method based on gene presence and absence leads to excellent results [69]. This led us to the development of a new and finer strategy. We used BLUR, a program specifically designed to highlight protein and sub-protein variations between two groups of species to compare *Otomorpha* and *Euteleostomorpha*, another group of bony fishes, which showed that, besides multiciliation, several processes diverged between these two clades, such as immunity. This showed the relevance of combining the two approaches, which highlighted a list of 114 genes that were both functionally linked to multiciliation but also presented a specific evolutionary pattern.

The functional characterization of these genes using the STRING database confirmed that a large part of our targets interacted with known ciliary genes, but also highlighted one common problem of in silico studies, which is the shortage of data and annotations in some cases, as shown by the lack of functional interactions detected for 65 of our target genes. This led us to the analysis of the phylogenetic profiles of our target genes, with two thirds of our target list showing a distribution that correlates to either a ciliary or a multiciliary pattern.

In this study, we highlighted a list of 114 potentially multiciliated target genes, of which 11 were especially promising: *C1orf189*, *C20orf85*, *C5orf24*, *KIAA1841*, *FAM181A*, *IQCK*, *LRR43*, *DYDC1*, *CFAP47*, *ANKRD60* and *TEX43*. Highlighted in our various analysis as either presenting an evolutionary signature that correlates with multiciliation, with a divergence in *Otomorpha* and a distribution in ciliated species, or as heavily influenced by the main transcription factors of the multiciliated pathway, these genes are, as of today, poorly characterized. With the exception of *DYDC1*, they are defined as TDARK

according to the system developed by the Illuminating the Druggable Genome Knowledge Management Center (IDG-KMC), i.e., they correspond to targets about which virtually nothing is known. Their lack of annotations makes them perfect targets for further studies and particularly encouraging candidates for multiciliation. Experimental validation is now required for these results, which could not only improve our understanding of the complex process that is multiciliation but could also be helpful as part of variant analyses in pathologies linked to multiciliation, for which very few responsible genes are currently known.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12091452/s1>, Table S1: List of target genes identified by the combined functional and comparative genomics approaches characterized using protein interaction annotations and phylogenetic profiling.

Author Contributions: Conceptualization, A.D., O.L. and O.P.; methodology, A.D.; software, A.D.; validation, A.D. and O.L.; formal analysis, A.D. and O.L.; investigation, A.D., D.M. and L.P.; resources, Y.N. and A.K.; data curation, A.D. and D.M.; writing—original draft preparation, A.D.; writing—review and editing, O.L. and O.P.; visualization, A.D., D.M. and L.P.; supervision, O.L.; project administration, O.L.; funding acquisition, O.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the IdEx Unistra in the framework of the “Investments for the future” program of the French government and Institute funds from the Centre National de la Recherche Scientifique and the Université de Strasbourg.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found in the GEO database (datasets GSE32452, GSE59309, GSE89271, GSE76342, GSE60365, GSE75715, GSE73331, GSE116690).

Acknowledgments: We thank the Bioinformatics and Genomics Est-ICube (BIGEst-ICube) and BISTRO bioinformatics platforms for informatics support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Pala, R.; Alomari, N.; Nauli, S.M. Primary Cilium-Dependent Signaling Mechanisms. *Int. J. Mol. Sci.* **2017**, *18*, 2272. [[CrossRef](#)]
2. Satir, P.; Christensen, S.T. Overview of Structure and Function of Mammalian Cilia. *Annu. Rev. Physiol.* **2007**, *69*, 377–400. [[CrossRef](#)]
3. Spassky, N.; Meunier, A. The Development and Functions of Multiciliated Epithelia. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 423–436. [[CrossRef](#)]
4. Brooks, E.R.; Wallingford, J.B. Multiciliated Cells. *Curr. Biol.* **2014**, *24*, R973–R982. [[CrossRef](#)]
5. Kramer-Zucker, A.G.; Olale, F.; Haycraft, C.J.; Yoder, B.K.; Schier, A.F.; Drummond, I.A. Cilia-Driven Fluid Flow in the Zebrafish Pronephros, Brain and Kupffer’s Vesicle Is Required for Normal Organogenesis. *Development* **2005**, *132*, 1907–1921. [[CrossRef](#)]
6. Nielsen, C. *Animal Evolution: Interrelationships of the Living Phyla*, 3rd ed.; Oxford University Press: Oxford, NY, USA, 2012; ISBN 978-0-19-960602-3.
7. Nikolaev, S.I. Phylogenetic Position of Multicilia Marina and the Evolution of Amoebozoa. *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 1449–1458. [[CrossRef](#)] [[PubMed](#)]
8. Allen, R.D. The Morphogenesis of Basal Bodies and Accessory Structures of the Cortex of the Ciliated Protozoan *Tetrahymena Pyriformis*. *J. Cell Biol.* **1969**, *40*, 716–733. [[CrossRef](#)]
9. Machemer, H. Ciliary Activity and the Origin of Metachrony in *Paramecium*: Effects of Increased Viscosity. *J. Exp. Biol.* **1972**, *57*, 239–259. [[CrossRef](#)]
10. Mizukami, I.; Gall, J. Centriole Replication. II. Sperm Formation in the Fern, *Marsilea*, and the Cycad, *Zamia*. *J. Cell Biol.* **1966**, *29*, 97–111. [[CrossRef](#)] [[PubMed](#)]
11. Hodges, M.E.; Wickstead, B.; Gull, K.; Langdale, J.A. The Evolution of Land Plant Cilia. *New Phytol.* **2012**, *195*, 526–540. [[CrossRef](#)] [[PubMed](#)]
12. Giribet, G. New Animal Phylogeny: Future Challenges for Animal Phylogeny in the Age of Phylogenomics. *Org. Divers. Evol.* **2016**, *16*, 419–426. [[CrossRef](#)]

13. Lewis, M.; Stracker, T.H. Transcriptional Regulation of Multiciliated Cell Differentiation. In *Seminars in Cell & Developmental Biology*; Academic Press: Cambridge, MA, USA, 2021; pp. 51–60. [\[CrossRef\]](#)
14. Deblandre, G.A.; Wettstein, D.A.; Koyano-Nakagawa, N.; Kintner, C. A Two-Step Mechanism Generates the Spacing Pattern of the Ciliated Cells in the Skin of Xenopus Embryos. *Development* **1999**, *126*, 4715–4728. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Liu, Y.; Pathak, N.; Kramer-Zucker, A.; Drummond, I.A. Notch Signaling Controls the Differentiation of Transporting Epithelia and Multiciliated Cells in the Zebrafish Pronephros. *Development* **2007**, *134*, 1111–1122. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Tsao, P.-N.; Vasconcelos, M.; Izvolosky, K.I.; Qian, J.; Lu, J.; Cardoso, W.V. Notch Signaling Controls the Balance of Ciliated and Secretory Cell Fates in Developing Airways. *Development* **2009**, *136*, 2297–2307. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Marcet, B.; Chevalier, B.; Luxardi, G.; Coraux, C.; Zaragosi, L.-E.; Cibois, M.; Robbe-Sermesant, K.; Jolly, T.; Cardinaud, B.; Moreilhon, C.; et al. Control of Vertebrate Multiciliogenesis by MiR-449 through Direct Repression of the Delta/Notch Pathway. *Nat. Cell Biol.* **2011**, *13*, 694–701. [\[CrossRef\]](#)
18. Chu, Q.; Yao, C.; Qi, X.; Stripp, B.R.; Tang, N. STK11 Is Required for the Normal Program of Ciliated Cell Differentiation in Airways. *Cell Discov.* **2019**, *5*, 1–16. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Stubbs, J.L.; Vladar, E.K.; Axelrod, J.D.; Kintner, C. Multicilin Promotes Centriole Assembly and Ciliogenesis during Multiciliate Cell Differentiation. *Nat. Cell Biol.* **2012**, *14*, 140–147. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Zhou, F.; Narasimhan, V.; Shboul, M.; Chong, Y.L.; Reversade, B.; Roy, S. Gmnc Is a Master Regulator of the Multiciliated Cell Differentiation Program. *Curr. Biol.* **2015**, *25*, 3267–3273. [\[CrossRef\]](#)
21. Kyrousi, C.; Arbi, M.; Pilz, G.-A.; Pefani, D.-E.; Lalioti, M.-E.; Ninkovic, J.; Go tz, M.; Lygerou, Z.; Taraviras, S. Mcidas and GemC1 Are Key Regulators for the Generation of Multiciliated Ependymal Cells in the Adult Neurogenic Niche. *Development* **2015**, *142*, 3661–3674. [\[CrossRef\]](#)
22. Terré, B.; Piergiovanni, G.; Segura-Bayona, S.; Gil-Gómez, G.; Yousef, S.A.; Attolini, C.S.-O.; Wilsch-Bräuninger, M.; Jung, C.; Rojas, A.M.; Marjanović, M.; et al. GEMC1 Is a Critical Regulator of Multiciliated Cell Differentiation. *EMBO J.* **2016**, *35*, 942–960. [\[CrossRef\]](#)
23. Ma, L.; Quigley, I.; Omran, H.; Kintner, C. Multicilin Drives Centriole Biogenesis via E2f Proteins. *Genes Dev.* **2014**, *28*, 1461–1471. [\[CrossRef\]](#)
24. Marshall, C.B.; Mays, D.J.; Beeler, J.S.; Rosenbluth, J.M.; Boyd, K.L.; Santos Guasch, G.L.; Shaver, T.M.; Tang, L.J.; Liu, Q.; Shyr, Y.; et al. P73 Is Required for Multiciliogenesis and Regulates the Foxj1-Associated Gene Network. *Cell Rep.* **2016**, *14*, 2289–2300. [\[CrossRef\]](#)
25. Nemajerova, A.; Kramer, D.; Siller, S.S.; Herr, C.; Shomroni, O.; Pena, T.; Gallinas Suazo, C.; Glaser, K.; Wildung, M.; Steffen, H.; et al. TAP73 Is a Central Transcriptional Regulator of Airway Multiciliogenesis. *Genes Dev.* **2016**, *30*, 1300–1312. [\[CrossRef\]](#)
26. Tan, F.E.; Vladar, E.K.; Ma, L.; Fuentealba, L.C.; Hoh, R.; Espinoza, F.H.; Axelrod, J.D.; Alvarez-Buylla, A.; Stearns, T.; Kintner, C.; et al. Myb Promotes Centriole Amplification and Later Steps of the Multiciliogenesis Program. *Development* **2013**, *140*, 4277–4286. [\[CrossRef\]](#)
27. Zhao, H.; Zhu, L.; Zhu, Y.; Cao, J.; Li, S.; Huang, Q.; Xu, T.; Huang, X.; Yan, X.; Zhu, X. The Cep63 Parologue Deup1 Enables Massive de Novo Centriole Biogenesis for Vertebrate Multiciliogenesis. *Nat. Cell Biol.* **2013**, *15*, 1434–1444. [\[CrossRef\]](#)
28. Mercey, O.; Levine, M.S.; LoMastro, G.M.; Rostaing, P.; Brotslaw, E.; Gomez, V.; Kumar, A.; Spassky, N.; Mitchell, B.J.; Meunier, A.; et al. Massive Centriole Production Can Occur in the Absence of Deuterosomes in Multiciliated Cells. *Nat. Cell Biol.* **2019**, *21*, 1544–1552. [\[CrossRef\]](#)
29. Nanjundappa, R.; Kong, D.; Shim, K.; Stearns, T.; Brody, S.L.; Loncarek, J.; Mahjoub, M.R. Regulation of Cilia Abundance in Multiciliated Cells. *Elife* **2019**, *8*, e44039. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Klos Dehring, D.A.; Vladar, E.K.; Werner, M.E.; Mitchell, J.W.; Hwang, P.; Mitchell, B.J. Deuterosome-Mediated Centriole Biogenesis. *Dev. Cell* **2013**, *27*, 103–112. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Funk, M.C.; Bera, A.N.; Menchen, T.; Kualess, G.; Thriene, K.; Lienkamp, S.S.; Dengjel, J.; Omran, H.; Frank, M.; Arnold, S.J. Cyclin O (Cno) Functions during Deuterosome-Mediated Centriole Amplification of Multiciliated Cells. *EMBO J.* **2015**, *34*, 1078–1089. [\[CrossRef\]](#) [\[PubMed\]](#)
32. You, Y.; Huang, T.; Richer, E.J.; Schmidt, J.-E.H.; Zabner, J.; Borok, Z.; Brody, S.L. Role of F-Box Factor Foxj1 in Differentiation of Ciliated Airway Epithelial Cells. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **2004**, *286*, L650–L657. [\[CrossRef\]](#)
33. Berlucchi, M.; de Santi, M.M.; Bertoni, E.; Spinelli, E.; Timpano, S.; Padoan, R. Ciliary Aplasia Associated with Hydrocephalus: An Extremely Rare Occurrence. *Eur. Arch. Oto-Rhino-Laryngol.* **2012**, *269*, 2295–2299. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Boon, M.; Wallmeier, J.; Ma, L.; Loges, N.T.; Jaspers, M.; Olbrich, H.; Dougherty, G.W.; Raidt, J.; Werner, C.; Amirav, I.; et al. MCIDAS Mutations Result in a Mucociliary Clearance Disorder with Reduced Generation of Multiple Motile Cilia. *Nat. Commun.* **2014**, *5*, 4418. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Wallmeier, J.; Al-Mutairi, D.A.; Chen, C.-T.; Loges, N.T.; Pennekamp, P.; Menchen, T.; Ma, L.; Shamseldin, H.E.; Olbrich, H.; Dougherty, G.W.; et al. Mutations in CCNO Result in Congenital Mucociliary Clearance Disorder with Reduced Generation of Multiple Motile Cilia. *Nat. Genet.* **2014**, *46*, 646–651. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Wallmeier, J.; Frank, D.; Shoemark, A.; Nöthe-Menchen, T.; Cindric, S.; Olbrich, H.; Loges, N.T.; Aprea, I.; Dougherty, G.W.; Pennekamp, P.; et al. De Novo Mutations in FOXJ1 Result in a Motile Ciliopathy with Hydrocephalus and Randomization of Left/Right Body Asymmetry. *Am. J. Hum. Genet.* **2019**, *105*, 1030–1039. [\[CrossRef\]](#)

37. Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [[CrossRef](#)]
38. Campbell, E.P.; Quigley, I.K.; Kintner, C. Foxn4 Promotes Gene Expression Required for the Formation of Multiple Motile Cilia. *Development* **2016**, *143*, 4654–4664. [[CrossRef](#)]
39. Quigley, I.K.; Kintner, C. Rfx2 Stabilizes Foxj1 Binding at Chromatin Loops to Enable Multiciliated Cell Gene Expression. *PLoS Genet.* **2017**, *13*, e1006538. [[CrossRef](#)]
40. Pan, J.; Adair-Kirk, T.L.; Patel, A.C.; Huang, T.; Yozamp, N.S.; Xu, J.; Reddy, E.P.; Byers, D.E.; Pierce, R.A.; Holtzman, M.J.; et al. Myb Permits Multilineage Airway Epithelial Cell Differentiation. *Stem Cells* **2014**, *32*, 3245–3256. [[CrossRef](#)]
41. Mori, M.; Hazan, R.; Danielian, P.S.; Mahoney, J.E.; Li, H.; Lu, J.; Miller, E.S.; Zhu, X.; Lees, J.A.; Cardoso, W.V. Cytoplasmic E2f4 Forms Organizing Centres for Initiation of Centriole Amplification during Multiciliogenesis. *Nat. Commun.* **2017**, *8*, 1–11. [[CrossRef](#)]
42. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
43. Karimi, K.; Fortriede, J.D.; Lotay, V.S.; Burns, K.A.; Wang, D.Z.; Fisher, M.E.; Pells, T.J.; James-Zorn, C.; Wang, Y.; Ponferrada, V.G.; et al. Xenbase: A Genomic, Epigenomic and Transcriptomic Model Organism Database. *Nucleic Acids Res.* **2018**, *46*, D861–D868. [[CrossRef](#)]
44. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)] [[PubMed](#)]
45. Plewniak, F.; Bianchetti, L.; Brelivet, Y.; Carles, A.; Chalmel, F.; Lecompte, O.; Mochel, T.; Moulinier, L.; Muller, A.; Muller, J.; et al. PipeAlign: A New Toolkit for Protein Family Analysis. *Nucleic Acids Res.* **2003**, *31*, 3829–3832. [[CrossRef](#)] [[PubMed](#)]
46. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
47. Yates, A.D.; Achuthan, P.; Akanni, W.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; et al. Ensembl 2020. *Nucleic Acids Res.* **2020**, *48*, D682–D688. [[CrossRef](#)] [[PubMed](#)]
48. Rangwala, S.H.; Kuznetsov, A.; Ananiev, V.; Asztalos, A.; Borodin, E.; Evgeniev, V.; Joukov, V.; Lotov, V.; Pannu, R.; Rudnev, D.; et al. Accessing NCBI Data Using the NCBI Sequence Viewer and Genome Data Viewer (GDV). *Genome Res.* **2021**, *31*, 159–169. [[CrossRef](#)]
49. Defosset, A.; Kress, A.; Nevers, Y.; Ripp, R.; Thompson, J.D.; Poch, O.; Lecompte, O. Proteome-Scale Detection of Differential Conservation Patterns at Protein and Subprotein Levels with BLUR. *Genome Biol. Evol.* **2020**, *13*, evaa248. [[CrossRef](#)] [[PubMed](#)]
50. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [[CrossRef](#)]
51. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
52. Su, G.; Kuchinsky, A.; Morris, J.H.; States, D.J.; Meng, F. GLay: Community Structure Analysis of Biological Networks. *Bioinformatics* **2010**, *26*, 3135–3137. [[CrossRef](#)]
53. Morris, J.H.; Apeltsin, L.; Newman, A.M.; Baumbach, J.; Wittkop, T.; Su, G.; Bader, G.D.; Ferrin, T.E. ClusterMaker: A Multi-Algorithm Clustering Plugin for Cytoscape. *BMC Bioinform.* **2011**, *12*, 436. [[CrossRef](#)]
54. Nevers, Y.; Kress, A.; Defosset, A.; Ripp, R.; Linard, B.; Thompson, J.D.; Poch, O.; Lecompte, O. OrthoInspector 3.0: Open Portal for Comparative Genomics. *Nucleic Acids Res.* **2019**, *47*, D411–D418. [[CrossRef](#)]
55. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [[CrossRef](#)]
56. Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools. *Nucleic Acids Res.* **2019**, *47*, D419–D426. [[CrossRef](#)] [[PubMed](#)]
57. Gene Ontology Consortium. The Gene Ontology Resource: Enriching a GOLD Mine. *Nucleic Acids Res.* **2021**, *49*, D325–D334. [[CrossRef](#)] [[PubMed](#)]
58. Long, H.; Huang, K. Transport of Ciliary Membrane Proteins. *Front. Cell Dev. Biol.* **2020**, *7*, 381. [[CrossRef](#)] [[PubMed](#)]
59. O’Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufu, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [[CrossRef](#)]
60. Zhou, F.; Rayamajhi, D.; Ravi, V.; Narasimhan, V.; Chong, Y.L.; Lu, H.; Venkatesh, B.; Roy, S. Conservation as Well as Divergence in Mcidas Function Underlies the Differentiation of Multiciliated Cells in Vertebrates. *Dev. Biol.* **2020**, *465*, 168–177. [[CrossRef](#)]
61. Honda, S.; Shimizu, S. Autophagy Controls Centrosome Number. *Oncotarget* **2017**, *8*, 14277–14278. [[CrossRef](#)]
62. Marra, A.N.; Cheng, C.N.; Adeeb, B.; Addiego, A.; Wesselman, H.M.; Chambers, B.E.; Chambers, J.M.; Wingert, R.A. Iroquois Transcription Factor Irx2a Is Required for Multiciliated and Transporter Cell Fate Decisions during Zebrafish Pronephros Development. *Sci. Rep.* **2019**, *9*, 6454. [[CrossRef](#)]
63. Thul, P.J.; Åkesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm, T.; Asplund, A.; Björk, L.; Breckels, L.M.; et al. A Subcellular Map of the Human Proteome. *Science* **2017**, *356*, eaal3321. [[CrossRef](#)]

64. Hong, K.-M.; Yang, S.-H.; Chowdhuri, S.R.; Player, A.; Hames, M.; Fukuoka, J.; Meerzaman, D.; Dracheva, T.; Sun, Z.; Yang, P.; et al. Inactivation of LLC1 Gene in Nonsmall Cell Lung Cancer. *Int. J. Cancer* **2007**, *120*, 2353–2358. [[CrossRef](#)]
65. Li, S.; Qiao, Y.; Di, Q.; Le, X.; Zhang, L.; Zhang, X.; Zhang, C.; Cheng, J.; Zong, S.; Koide, S.S.; et al. Interaction of SH3P13 and DYDC1 Protein: A Germ Cell Component That Regulates Acrosome Biogenesis during Spermiogenesis. *Eur. J. Cell Biol.* **2009**, *88*, 509–520. [[CrossRef](#)] [[PubMed](#)]
66. Zhao, L.; Zhang, Z.; Rodriguez, S.M.B.; Vardarajan, B.N.; Renton, A.E.; Goate, A.M.; Mayeux, R.; Wang, G.T.; Leal, S.M. A Quantitative Trait Rare Variant Nonparametric Linkage Method with Application to Age-at-Onset of Alzheimer's Disease. *Eur. J. Hum. Genet.* **2020**, *28*, 1734–1742. [[CrossRef](#)]
67. Raybould, R.; Sims, R. Searching the Dark Genome for Alzheimer's Disease Risk Variants. *Brain Sci.* **2021**, *11*, 332. [[CrossRef](#)] [[PubMed](#)]
68. Hansen, A.; Zeiske, E. Development of the Olfactory Organ in the Zebrafish, *Brachydanio Rerio*. *J. Comp. Neurol.* **1993**, *333*, 289–300. [[CrossRef](#)]
69. Nevers, Y.; Prasad, M.K.; Poidevin, L.; Chennen, K.; Allot, A.; Kress, A.; Ripp, R.; Thompson, J.D.; Dollfus, H.; Poch, O.; et al. Insights into Ciliary Genes and Evolution from Multi-Level Phylogenetic Profiling. *Mol. Biol. Evol.* **2017**, *34*, 2016–2034. [[CrossRef](#)]