SUPPLEMENTAL MOVIE LEGENDS

Supplemental Movie S1 (related to Fig. 1). p73KO mice show respiratory distress, characterized by continuous coughing and sneezing. WT mice are completely silent. Video recording of 6 wk old p73KO and WT littermates processed with identical settings.

Supplemental Movie S2 (related to Fig. 2)**.** Representative fluorescent bead movements over live TAp73WT trachea, live TAp73KO trachea and dead WT trachea (Diffusion control).

LIST OF SUPPLEMENTAL TABLES

Supplemental Table S1

- Alignment statistics of RNA-seq murine MTEC samples
- Alignment statistics of small RNA (sRNA) murine MTEC samples
- Quality control of microarray human arrays from published study GSE22142

Supplemental Table S2

- Summary DEG (differentially expressed genes)
- Summary cilium genes mouse
- Cilia-related genes mouse
- Read counts RNA-seq murine MTEC
- DEG RNA-seq murine MTEC
- Read counts sRNA-seq murine MTEC
- DEG sRNA-seq murine MTEC

Supplemental Table S3

- ChIP alignment Qc
- Peak summary
- Peak regions
- Gene types
- Peak distances
- Motif info
- TAp73 alpha annotated peaks
- TAp73 beta annotated peaks

- TAp73 union annotated peaks

Supplemental Table S4

- Overview DEG human
- Sequences of p73-enriched regions with p73 motifs whose nearest gene is DEG
- Sequences of p73-enriched regions with p73 motifs whose nearest gene is DEG and GO term "cilium"
- Sequences of p73-enriched regions with p73 motifs validated whose nearest gene is DEG
- Overview of cilium genes human
- Human genes associated with cilia-related gene ontology terms

Supplemental Table S5

- List of antibodies used
- List of primers for qRT-PCR and targeted ChIP assays used
- LUC reporter sequences used, derived from DEG genes that contain TP73 binding motifs

SUPPLEMENTAL MATERIALS AND METHODS

Mice

The p73KO mice with a deletion of exons 5 and 6 of the *TP73* gene were a generous gift from Dr. Frank McKeon (University of Houston, TX) (Yang et al. 2000). The TAp73KO mice with a deletion of exons 2 and 3 of the *TP73* gene were a generous gift from Dr. Tak Mak (Princess Margaret Cancer Centre, Toronto, Canada) (Tomasini et al. 2008). The p73KO mice were enriched for 6 generations on the SV129 background and the TAp73KO mice were maintained on pure C57Bl/6J background. Corresponding WT mice were littermates. The strains were barrier-maintained in the mouse facilities of Stony Brook University (p73KO) and Goettingen University (TAp73KO) through heterozygous interbreeding. The study was approved by the respective Animal Care Committees of both Institutions and animal use was in full compliance with institutional guidelines.

Pulmonary function

Pulmonary function measurements were performed using the forced oscillation technique with the automated FlexiVent system (Scireq Inc., Canada). Mice were anesthetized with a mixture of ketamine and xylazine. The trachea of anesthetized mice was exposed and cannulated using an 18G cannula. Data acquisition was performed using flexiWare 7.1 according to the manufacturer's instructions.

Stereology

Lungs were prepared from uniform random sampling for stereological analysis as described elsewhere (Nyengaard and Gundersen 2006; Weibel et al. 2007). The mean chord length (mean linear intercept) was calculated using the Visiopharm Integrator

System package (Version 4.2.2, Visiopharm, Denmark) on an Olympus BX51 microscope equipped with an 8-position slide loader.

Mucociliary transport assay

Freshly harvested vital tracheae were longitudinally cut with a microtome (VT1200S, Leica, Germany) and immersed in a solution containing 98mM NaCl, 2mM KCl, 1mM $CaCl_2$, 2mM $MgCl_2$, 5mM glucose, 5mM sodium pyruvate, 10mM Hepes; 230mOsm, pH7.8. Fluorescent microspheres (FluoSpheres®, 0.17 µm, PS-Speck, Life Technologies) were added to the bath solution. Areas of 280x280 µm (512x512 pixels) at 10-20 µm above the mucosal surface were imaged using a custom-built confocal line illumination microscope (Junek et al. 2009) at 61Hz for 2000 frames. Particle positions were detected for each frame and tracked using functions previously described (Crocker and Grier 1996). Absolute full-trajectory velocities along the focal plane were pooled within one measurement. Averaged velocities of independent measurements from different tissue positions and different preparations were combined into the groups 'WT' versus 'TAp73KO' or 'p73KO', as well as the 'diffusion control' group from measurements using nonliving tracheae. We measured separate TAp73 WT/KO and global p73 WT/KO cohorts, performed on different days. In each WT/KO cohort, the ages were matched. Diffusion over dead tracheae of WT mice from theTAp73 and p73 cohorts were merged since they are identical. To obtain particle velocity distributions, individual trajectory velocities from all measurements within a respective group were included. All data evaluation was performed with custom software written in Matlab (MathWorks, USA). Representative videos can be found in the Supplemental Movie S2.

Histology and immunostaining

Mice were euthanized with CO_2. Tracheae, main bronchi and lungs were removed, washed in PBS, and fixed overnight in 10% formalin/PBS. Paraffin-embedded 3µm sections were processed and stained with H&E for histology. Immunofluorescence (IF) and immunoperoxidase (IHC) stainings were performed as described (Holembowski et al. 2014). MTECs on supporting membranes were fixed in either 3.2% PFA/3% sucrose in PBS, pH 7.4, or in methanol-acetone (1:1), depending on the antigen, and processed for immunostaining as described (You et al. 2002). Fixed sample were incubated with primary antibodies for 16h at 4°C (Supplemental Table S5). For IF, nuclei were counterstained with DAPI (Sigma) before mounting with Fluoromount-G (Southern Biotech).

Samples were viewed on a Zeiss AxioScope A1 and a Zeiss confocal microscope LSM 510 (with a 63x/1.4 NA or a 100x/1.4 NA objective). Images were acquired and quantified using digital cameras (AxioCam MRm) with the AxioVision Rel. 4.8 Software (Zeiss). In addition, samples were imaged on a Nikon N-SIM microscope (Superresolution Structured Illumination Microscopy) with a 100x/1.49 NA objective, equipped with an electron-multiplying CCD camera (iXon3 897, Andor Technology) and NIS-Elements image analysis software (Nikon).

Quantitation of ciliary stages

The ciliation process in MCCs is subdivided into 4 stages, based on staining for the ciliary axoneme marker Ac α-tubulin and the centriole/basal body marker γ-tubulin (Vladar and Stearns 2007). Approximately 2-3d post-ALI, elongated primary cilia become visible, marking stage I. In stages II and III centrioles replicate en masse and migrate toward the apical plasma membrane. In stage IV axoneme growth occurs.

MTECs from WT and p73KO tracheae at D2, 4, 7 and 14 post-ALI were immunostained for Ac α-tub. The percent of unciliated, partly ciliated (stages I-IV) and fully ciliated MCCs were determined by confocal microscopy. Fully differentiated MCCs are defined by the presence of at least 20 fully extended Ac α-tub-positive axonemes per cell. Images of at least 250 cells/genotype/ALI day were counted from 3 indep. MTEC preparations, thus >750 total cells for each condition.

Transmission (TEM) and scanning (SEM) electron microscopy

TEM and SEM were performed using standard methods as described (Voronina et al. 2009; Love et al. 2010). For p73KO TEM, tracheas were fixed in 2% PFA and 2% glutaraldehyde, postfixed in 2% osmium tetroxide, dehydrated and embedded in Durcupan resin. Ultrathin sections of 80 nm were cut with an ultramicrotome (Ultracut E, Reichert–Jung) and placed on formvar-coated slot copper grids. Sections were counterstained with uranyl acetate and lead citrate. Imaging was done with on a Tecnail2 BioTwinG2 transmission electron microscope with a CCD digital camera system (XR-60, Advanced Microscopy Techniques). For TAp73KO TEM, lungs were fixed by immersion using 2 % glutaraldehyde in 0.1 M cacodylate buffer at pH 7.4 overnight at 4°C. After postfixation in 1% osmium tetroxide and preembedding staining with 1% uranyl acetate, tissue samples were dehydrated and embedded in Agar 100. Thin sections (80 nm) were counterstained uranylacetat in methanol and examined using a Philips CM 120 BioTwin transmission electron microscope (Philips Inc.Eindhoven, The Netherlands). Images were taken with a TemCam F416 CMOS camera (TVIPS, Gauting, Germany).

For SEM, tracheae and MTECs at ALI D14 were fixed in 2% PFA and 2% glutaraldehyde, dehydrated and transferred to 100% hexamethyldisilazane (HMDS,

Electron Microscopy Sciences, USA) through a graded series of ethanol-HMDS mixtures. The specimens were air-dried, mounted on SEM stubs, and sputter-coated with gold before examination with a LEO1550 scanning electron microscope (Zeiss, Germany) at 10 kV using a backscatter electron detector (Robinson SPI Supplies).

MTEC cultures and lentiviral constructs and infection

Primary cultures of mouse tracheal epithelial cells (MTECs) were established on membranes using ALI conditions as described (You et al. 2002). In brief, tracheae were excised from 2 mo old WT and TAp73KO/p73KO mice, and epithelial cells were harvested by pronase digestion, seeded onto Transwell polycarbonate or polyester permeable membranes (6.5 mm diameter, Corning), and allowed to proliferate to confluence in growth media. Subsequently, the air-liquid interface (ALI) condition was created by removing apical chamber media and switching the bottom chamber to differentiation media with 2% NuSerum (BD Biosciences).

A 1,007-bp fragment of the human FoxJ1 promoter region was amplified by PCR from pRRLsin-FoxJ1 promoter-GFP (kind gift from Steve Brody) using primers containing PacI and EcoRI restriction sites, followed by a restriction digest with PacI and EcoRI. The pEF1alpha-IRES-EGFP lentiviral vector (Burke et al. 2014) was digested with PacI and EcoRI to remove the EF1alpha promoter followed by replacement with the FoxJ1 promoter fragment to yield the FoxJ1-IRES-EGFP lentiviral vector. All Flag-tagged inserts (mouse TAp73, FoxJ1, Rfx2, Rfx3) were then cloned into the Sfi1 restriction site downstream of the FoxJ1 promoter. The final plasmids were sequence confirmed.

Lentiviral expression restricted to MCC-fated cells is based on FoxJ1 basal promoter activity and the existence of two positive feed forward loops. *i)* p73KO MCCs still have a

basal level of (TAp73-independent) endogenous FoxJ1 expression which, importantly, is too low to drive ciliogenesis (see Supplemental Figure S7E). Yet, these cells have a basal level of endogenous FoxJ1 promoter activity. This implies that the lentiviral FoxJ1 promoter is also active at this basal level, producing ectopic TAp73 that in turn acts as an amplifier and drives endogenous FoxJ1 expression to much higher levels over the threshold that is required for ciliogenesis. *ii)* Moreover, FoxJ1 can activate its own promoter (Didon et al. 2013, Venugopalan et al. 2011). Together these effects set up and perpetuate a positive feedback loop. In this way this system can be exploited to steer MCC-lineage specific ectopic expression.

For viral infection of MTECs, isolated epithelial cells were mixed with lentiviral supernatants in the presence of 10µg/ml protamine sulfate (Sigma) at the time of seeding and repeated once 24 hrs later. For Supplemental Fig. S6D, WT MTECs were lentivirally infected at day 3 after seeding as described (Stubbs et al. 2012). To generate lentiviral supernatants, HEK293T cells were transfected with lentiviral constructs using X-tremeGene (Roche) and maintained in 10% FBS DMEM.

Human airway epithelial cells (HAECs), derived from normal large airways resected during surgery (for fibrosis), were recovered and cultivated as ALI as we previously described (Lize et al. 2010). The protocol was approved by the ethics committee of the University of Saarland and informed consent had been obtained from all patients.

Quantitative ciliary rescue experiments

WT and p73KO MTECs, infected with lentiviral FoxJ1-driven mouse TAp73α or FoxJ1 at time of seeding or not infected, were scored at D4, and D14 post-ALI. The percentage of scored cells was obtained by dividing the number of fully differentiated cells (based on

the presence of at least 20 fully extended Ac α-tub-positive cilia per cell) by the number of total cells. At least 750 random infected cells were counted for each genotype and data point.

Immunoblot analysis

Lung pieces were homogenized in RIPA buffer (20mM Tris-HCl pH7.5, 150mM NaCl, 9.5mM EDTA, 1% Triton X100, 0.1% SDS, 1% sodium desoxycholate) supplemented with 3M urea and protease inhibitors (Complete Mini EDTA-free, Roche). Protein concentration was determined (Pierce BCA kit, ThermoScientific) and proteins were separated by electrophoresis prior to transfer onto a nitrocellulose membrane and immunostaining with specific antibodies listed in Supplemental Table S5.

Quantitative RT-PCR

RNA from lung tissues was isolated using TRIzol/Chloroform (Life Techn.) as recommended by the manufacturer. Reverse transcription of mRNAs from total RNA preparation was performed using the M-MuLV reverse transcriptase (New England Biolabs) and a mixture of Oligo-dT and random nonamers. cDNA was analyzed in triplicates by standard realtime PCR (Chromo4, Biorad) using 75mM Tris-HCl pH 8.8, 50mM (NH4)2SO4, 0.5% Tween-20, 1.5mM Trehalose, 0.125% Triton X-100, 1:80,000 Sybr Green I, 0.2mM dNTP mix and 20 U/mL Taq-polymerase (Primetech). For Fig. 6D, RNA e\was extracted from WT MTECs infected with empty vector or mouse Mcidas (Stubbs et al. 2012) using RNeasy kit (Qiagen). cDNA was prepared using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Primers (250nM final concentration) are listed in Supplemental Table S5. The relative expression of target genes was normalized to the housekeeping gene 36B4 (ribosomal protein, large, P0).

Targeted ChIP analysis

For targeted ChIP analysis of TP73, chromatin was harvested from Saos2 cells transiently overexpressing TAp73α, TAp73β cells or empty vector. Chromatin immunoprecipitation (ChIP) was performed as described by (Denissov et al. 2011). Briefly, vector-expressing and TAp73α / TAp73β-expressing Saos2 cells were cross-linked with 1% formaldehyde for 30 min. Cross-linking was stopped by adding glycine. After cell lysis cells were sonicated two times for 10 min using a BioRuptor (Diagenode, high power setting, 30 sec on/off) in the presence of 0.3% SDS, followed by centrifugation to remove unsheared chromatin. An aliquot of sheared chromatin was de-crosslinked and served as input control. For immunoprecipitation, chromatin was incubated with Protein A/G plus agarose beads and 2 µg of TAp73 isoform-specific antibody (ab14430, Abcam) or IgG (ab2410, rabbit anti-human, Abcam) overnight, rotating at 4°C. ChIP samples were extensively washed using 4 different wash buffers, eluted from the beads, and de-crosslinked for 4 h at 65°C in the presence of 0.2 M NaCl. Subsequently, the DNA was purified using the Mini Elute PCR Purification Kit (Qiagen). Ultimately, purified DNA was resuspended in 100 µl of nuclease-free water and analyzed by real-time qPCR. Analysis of the myoglobin (MB, control) promoter by real-time qPCR served as internal negative control, validating the specificity of the ChIP assay. Fold enrichment was calculated over the recovery from the control IgG-ChIP.

Luciferase reporter assays

For Fig. 5G, Saos2 cells were transfected with pcDNA3-based empty vector or with TAp73α or TAp73β isoforms in the presence of a Luc reporter with a minimal promoter plus a DNA sequence containing the putative wildtype TAp73-binding motifs identified by ChIP-seq (called 'WT'). Luc constructs with the same motif but lacking the predicted TAp73 binding motifs served as mutant controls ('Mut'). For Supplemental Fig. S6E, Saos2 cells transfected with pcDNA3-based empty plasmid, or with E2F4 alone, Mcidas alone, or E2F4 and Mcidas combined in the presence of a Luc reporter with a minimal promoter plus a DNA sequence containing the putative wildtype E2F4-binding motifs identified by ChIP-seq (called 'WT'). Luc constructs with the same motif but lacking the predicted E2F4 binding motifs served as mutant controls ('Mut'). See also Supplemental Table S5 for DNA sequences. A Renilla TK luciferase construct was transfected in parallel for control of transfection efficiencies. Twenty-four hours after transfection, cells were harvested in 1x passive lysis buffer (Promega). Dual luciferase assay was conducted as described by (Dyer et al. 2000), using firefly buffer (15mM K_2HPO_4, 4mM EGTA (pH 8.0), 15mM $MgSO_4$, 4mM ATP, 1.25mM DTT, 0.1mM CoA and 80μM Luciferin) and Renilla buffer (1.1M NaCl, 2.2mM Na_2EDTA, 0.22M K_2HPO_4 (pH 5.1), 0.5mg/mL BSA, 1.5mM NaN_3 and 1.5μM Coelenterazine). The activities of all firefly luciferase constructs were normalized to the respective Renilla activity in each sample.

Statistical analysis (qPCR, targeted ChIP, luciferase)

Normal distribution of the samples was assessed by Shapiro–Wilk test and the equality of variances within the groups was analyzed by F-test or Levene's test depending on normality. Unpaired, one-tailed student *t*-test with Welch corrections for unequal variances was then applied to the collected data. Significance was assumed for $p< 0.05$.

Asterisks indicate: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.005$. Error bars represent standard error of the mean (SEM).

Genomic analyses

RNA-sequencing and small RNA-sequencing

In order to detect TP73-dependent transcriptional changes in murine multiciliated tracheal cells, gene expression for primary wildtype and p73KO MTEC cultures were measured over 4 consecutive time points of differentiation (ALI D0, 4, 7, and 14). RNA for RNA-seq and small RNA-seq experiments was isolated using RNeasy kit (Qiagen) according to the manufacturer's protocol. RNA-seq and small RNA-seq libraries were prepared using the TruSeq® RNA Sample Preparation v2 kit and the TruSeq® Small RNA Preparation kit from Illumina according to the manufacturer's instructions. The library quality was checked using an Agilent 2100 Bioanalyzer and a Qubit dsDNA HS Assay Kit. Sample concentrations were measured using a Qubit dsDNA HS Assay Kit and adjusted to 2 nM before sequencing (50 bp single end) on a HiSeq 2000 instrument (Illumina) using TruSeq SR Cluster Kit v3-cBot-HS and TruSeq SBS Kit v3-HS according to the manufacturer's instructions. Sequencing was performed on three independent biological replicates per condition (i.e. time point and genotype). For each genotype, replicates consisted of pools of mice, i.e. replicate 1-3 consisted of 3, 3, and 5 mice, respectively. In some cases outlier samples were removed from further analysis based on the initial sample quality evaluation, as described in the 'quality control' section.

Computational analysis

To understand the molecular role of TP73 in multiciliogenesis across organisms and cell types, we integrated and analyzed several published and unpublished high-throughput sequencing and microarray datasets. Before describing the details of the computational

analysis of the various datasets, we would like to first provide a rationale for the integration and analysis of these datasets.

Cilium annotation

A primary interest of our computational analyses was to understand TP73's role in the regulation of motile multiciliogenic processes. In order to annotate genes that are involved in ciliary biogenesis, maintenance and function or whose products constitute structural parts of cilia, we generated a set (union of unique gene entries) of 'cilium' genes using three publically available resources:

1. Genes associated with gene ontology (GO) terms that include 'cilium', 'cilia', or 'cilio' in their identifier (e.g. cilium, ciliary part, epithelial cilium movement). Genes were annotated using the libraries GO.db_3.0.0, org.Hg.eg.db_3.0.0, and org.Mm.eg.db_3.0.0.
2. SysCilia Gold Standard version 1 (van Dam et al. 2013)
3. Manual annotation for Ccdc65 (Austin-Tse et al. 2013; Horani et al. 2013; Werner-Peterson and Sloboda 2013; Niederberger 2014); Rfx2 (Bisgrove et al. 2012; Chung et al. 2012; Chung et al. 2014; Kwon et al. 2014) and Spag1 (Knowles et al. 2013).

Information on 'cilium'-related genes can be found in Supplemental Table S2 for mouse and Supplemental Table S4 for human.

MTEC RNA-seq analysis

RNA-seq data of murine tracheal epithelial cells (MTECs) at post-ALI D0, 4, 7 and 14 served to understand gene expression differences between WT and p73KO MTECs over

time. In contrast to all other high-throughput datasets used throughout this study, this data was generated in house (see section 'RNA-sequencing').

Read alignment

Murine RNA-seq data was aligned to the Mus musculus mm10 genome using STAR aligner (version 2.3.0e_r291) with default options and automatically generating alignment files in BAM format (Dobin et al. 2013). Read counts for all genes and all exons (Ensembl annotation v72) were obtained using FeaturesCount (version 1.4.6) (Liao et al. 2014).

Differential expression analysis

Read counts, as generated in the 'Read alignment' section, were used to find differentially expressed genes (DEGs) between WT and KO samples at the different time points (ALI D0, 4, 7 and 14) using DESeq2 (Love et al. 2014). Results of the differential expression analysis can be found in Supplemental Table S2. Genes with an adjusted p-value smaller than 0.1 and an absolute log2 fold change (logFC) higher than 0.6 were considered to be DEGs (thresholds were selected by running parameter testing, as explained in section 'Parameter testing'). A heatmap representation of all DE RNAs over all analysed time points can be found in Supplemental Fig. S4C, D. In addition, a heatmap representation of all DEG RNAs that are also 'cilium'-related can be found in Fig. 4. The association of DEG RNAs during MTEC differentiation to human Saos2 TAp73 binding sites and motifs is described in section 'Peak DEG comparison'.

Functional enrichment analysis

For the identification of enriched gene ontology (GO) categories, the web-service WebGestalt was used. GO category enrichment was assessed by calculating the fold-

change between the observed and the expected number of genes of a given GO category, where terms were deemed enriched if they had adjusted p-values < 0.1. To allow for an easy comparison of different biological conditions, an in-house R script grouped the results and visualized them by heatmaps. The final result consisted of a heatmap of the union of all top 20 most enriched terms calculated for the DEGs for each time point (Supplemental Fig. S5A, B). The functional gene enrichment analysis was repeated for DEGs that are also bound by TP73, based on their overlap with annotated TP73 peaks (see section 'Peak DEG comparison') (Supplemental Fig. S5C-E).

MTEC sRNA-seq analysis

Similar to the analysis of RNA-seq data, small RNA (sRNA) sequencing data of WT and p73KO MTECs over time (ALI D0, 4, 7, and 14) served to understand TP73's regulatory role in sRNA expression during multiciliogenesis. Mapping, prediction of novel miRNAs, quality control, and differential expression (DE) analysis were carried out using a slightly modified in-house version of Oasis1.0 (Capece et al. 2015). The results of the DE analyses can be found in Supplemental Table S2. sRNAs were considered to be significantly DE when their adjusted p-value < 0.1. A heatmap representation of all DE sRNAs over all time points can be found in Supplemental Fig. S4D. The association of DE sRNAs during MTEC differentiation to human Saos2 TA73 binding sites and motifs is described in section 'Peak DEG comparison'. GSE75717 is the GEO accession number for RNA-seq and sRNA-seq data sets.

TAp73 ChIP-seq analysis

To understand if TAp73 directly binds and regulates the differentially expressed genes in MTECs, we used our published ChIP-seq data (GEO dataset GSE15780) of TAp73α or TAp73β isoform overexpressing human osteosarcoma Saos2 cells (Koeppel et al. 2011).

To allow for optimal peak calling and gene assignment, the data was re-analyzed using in-house workflows, encompassing quality assurance, alignment, peak calling, and analysis summarization (see section 'Quality control' for further information, and see the following subsections for the rest). Although Saos2 cells are nonciliated human cancer cells, we reasoned that TAp73 binding is highly conserved. Indeed, as revealed by our results, this assumption was correct, since TAp73 directly binds to a large cohort of ciliogenesis genes in these cells.

Read alignment

The TAp73 ChIP-seq samples were aligned to the Homo sapiens hg19 genome using Bowtie2 (version 2.0.2) with default parameters (Langmead and Salzberg 2012). Reads were subsequently filtered using two alternative options (1) high quality uni- and multi-mapped reads and (2) good quality uni-mapped reads:

1. High quality uniquely and multi-mapped reads were obtained by filtering out reads with low quality (i.e. when the field MAPQ != [0, 2, 3, 4]).
2. Good quality uniquely mapped reads were obtained by filtering out reads with MAPQ scores [0, 1]. This step removes all true multi-map reads (reads that align to several genomic locations with the same score).

The obtained Sequence Alignment/Map (SAM) files were converted into sorted Binary Alignment/Map (BAM) format using Samtools (Li et al. 2009). Throughout the analysis, ChIP data was analyzed using high quality uni- and muli-mapped reads, based on our previous experience with high quality reads (Halder et al. 2016).

TP73 peak detection

To detect significantly enriched binding sites for TAp73α and TAp73β, we calculated the average fragment length for merged biological replicate BAM files (TAp73α replicates 1

and 2; TAp73β replicates 1 and 2) using chequeR (Halder et al. 2016) (see also section 'Quality control'). Subsequently, potential TAp73 binding sites (peaks) were detected using MACS2 (v2.0.9) (Zhang et al. 2008), providing TAp73α and TAp73β merged BAM files as treatment and an input file as control. MACS2 "callpeaks" was run with parameters "--q=0.05" for the selected FDR, "--nomodel" to skip the shift size estimation step, and "--shiftsize" to give the previously calculated shift sizes for each sample. The output from peak annotation was a BED file specifying the coordinates of the peaks that were called, consisting of 20,538 TAp73α and 29,721 TAp73β peaks. Since it was not the scope of this work to differentiate between TAp73α and TAp73β, BED files of the binding sites were imported into R using the rtracklayer package, and the regions were reduced with GenomicRanges to combine multiple overlapping BED regions into single regions, resulting in a total of 36,651 TAp73 peaks (Supplemental Table S3).

Peak annotation

To link potential TAp73 binding sites to gene expression, we next linked peaks to their nearest gene using custom in-house Perl script. To limit the mis-assignment of peaks to far-away genes, we set a conservative peak/target gene distance threshold, using only annotated peaks for further analyses (e.g. peak-DEG comparisons) that are maximally 5 kilo base pairs (kbp) upstream of a gene or 1 kbp downstream of a gene. Peaks within a gene were further annotated as being located in the 5'-UTR, exon, intron or 3'-UTR of the gene. A summary of peak-related statistics (including the genomic distribution of peaks) and annotation can be found in Supplemental Table S3.

Motif discovery and annotation

The HOMER software suite (v.4.6) (Heinz et al. 2010) was used for the de-novo discovery of enriched motifs. The discovery protocol consisted of providing the regions

of interest (BED format) to the script findMotifsGenome.pl with parameters "--len 16" and "--size given" in order to find motifs throughout the entire peak regions of size 16 bp. A summary of the most enriched motifs for TAp73α, TAp73β, and their merging (TAp73 union) is included in Supplemental Table S3.

The *de novo* motif for TP73 (merged peak set) was found to be structurally similar to the known TP53/TP63 motif. This motif was subsequently used to find motif-containing peaks by rerunning the motif analysis findMotifsGenome.pl on TP73 peaks. This analysis resulted in a file containing TP73 peaks, the motif score, and the offset of the motif target from the peak start. The motif-target scores and offsets (recalculated from the peak centers) are visualized in Fig. 5C, D.

Peak DEG comparison

In order to compare human TP73 peaks with mouse MTEC DEG results, human genes that contain a peak or contain a peak and have a TP73 motif (see section 'motif discovery and annotation') were retrieved and converted from human to mouse ENSEMBL identifiers (IDs). The conversion of human to mouse IDs can be split into two approaches, (1) the annotation of RNA-seq results and (2) the annotation of sRNA-seq results:

1. For RNA-seq data, homologous mouse and human genes were annotated using the R library biomaRt (Durinck et al. 2009).
2. For the comparison of sRNAs between human and mouse, we first excluded piwi-RNAs and novel miRNAs due to their lack of homologous gene IDs in humans. With the exception of miRNAs, human sRNAs were compared to mouse sRNAs using biomaRt as described for RNA-seq data. For the annotation of homologous human and mouse miRNAs, murine miRNA IDs were directly

compared to human IDs, disregarding “mmu” and “hsa” prefixes and ‘5p’ and ‘3p’ suffixes. The coordinates of the resulting human miRNAs were retrieved, extended by 10 kb upstream of the start and downstream of the end regions, and overlapped with TP73 peaks and motifs. Any human miRNA overlapping with a TP73 peak or motif were annotated as peak and/or motif-related, resulting in the peak and/or motif-related annotation of the equivalent mouse miRNAs.

The information about peak and motif association to genes (RNA-seq) and to sRNAs is included in the corresponding RNA- and sRNA-seq tables of Supplemental Table S2. The peak and motif associations were also used to generate a barplot complementary to the DEG heatmaps (see ‘MTEC RNA-seq analysis’ and ‘sRNA-seq analysis’ for RNA- and sRNA-seq heatmaps), where for each DEG the maximum binary logarithm fold-change (logFC), computed across all time-points, was plotted, and the bars were filled with colour based on whether the DEGs are TP73 bound (orange), TP73 bound and containing TP73 motif (red) or just DE (grey) (Fig. 4 and Supplemental Fig. S4C, D).

Subsequently, genes were categorized as DEG, containing a TP73 binding site, containing a TP73 binding site plus a motif, belonging to the ‘cilium’ group of genes (see section ‘cilium annotation’), or any combination of the aforementioned (e.g. DEGs that are bound by TP73, contain a binding motif, and belong to the class of ‘cilium’ genes). For each combination, a Fisher’s exact test was calculated to determine whether the overlap between the DEGs and the other gene groups (cilia-related, p73 bound or p73-motif-containing) was significant (Fig. 4).

Enhancer prediction

Regions that are bound by TP73, contain TP73 motifs, and are DEGs during MTEC differentiation were short-listed for *in vitro* enhancer assays (Supplemental Table S4;

see section ‘Luc reporter assays’). To this end, the corresponding sequences were extracted using the R package BSgenome.Hsapiens. UCSC.hg19, highlighting TP73 motifs contained therein (Supplemental Table S4).

Microarray analysis

To assess if the development of tracheal multiciliated cells is conserved between mouse and human, we compared MTEC RNA-seq data to published human expression profiles (GSE22142) (Marcet et al. 2011). The human microarray expression data consists of four time points of regenerating primary human airway epithelium cells (ALI D0, 7, 14, 21).

CEL files were analyzed for significant gene expression differences over time using the R package maSigPro (version 1.38.0) with default parameters (p-value <= 0.05, regression fit >= 0.7, cubic polynomial) (Conesa et al. 2006). In brief, data was first normalized via background subtraction, quantile normalization, and summarization. The normalized gene expression was fit using a regression model to establish which probes in the arrays were significantly different across the different time points. The DE human Affymetrix probes were converted into mouse genes using biomaRt and results are summarized in Supplemental Fig. S6D and Supplemental Table S2.

To compare the expression of the key regulatory factors *FoxJ1*, *Rfx2*, *Rfx3* and *Tp73*, human microarray and murine RNA-seq data for the corresponding genes was z-score normalized and visualized in Fig. S6E.

Quality control

RNA-seq, small RNA-seq, and ChIP-seq data were subjected to an in-house quality control workflow. (Capece et al. 2015; Halder et al. 2016). During quality assessment, sequencing data was analyzed for general read quality, genome alignment, average per base coverage, read saturation, replicate correlation, enrichment, and was visually inspected in a genome browser. Only data passing all quality standards was used for further analyses.

In the case of both RNA-seq and sRNA-seq datasets, the percentage of aligned, uniquely aligned and unaligned reads was used to determine possible sequencing biases, while sequencing depth was determined by the average per base coverage for each sample (Supplemental Table S1).

Principal component analysis (PCA) was performed on both RNA-seq and sRNA-seq expression values to ensure that samples clustered into biological groups such as biological replicates, genetic background, and differentiation time. Samples that did not meet our quality standards were removed from further analysis, leading to the removal of RNA-seq KO replicate # 3 and sRNA-seq replicate # 1 samples from the initial three independent biological replicates per group (Supplemental Fig. S4A, B).

ChIP-seq enrichment and signal-to-noise ratio were assessed per sample by calculating normalized strand cross correlation (NSC) and relative strand cross correlation (RSC) coefficients using the in-house developed R package chequeR (Halder et al., 2016). In general, good quality libraries have an NSC>1.05 and an RSC>0.8 (Landt et al. 2012). In adddition, the saturation correlation of each sample and the correlation between

replicates was calculated using the R package MEDIPS (Lienhard et al. 2014) (Supplemental Table S3).

Microarray data quality was assessed using an in-house R script based on the arrayQualityMetrics R package (version 3.22.1) (Kauffmann et al. 2009). The quality of the microarray data was assessed through criteria of (1) distances between arrays, (2) array intensity distributions (based on raw signal intensity, relative log expression, and normalized unscaled standard error) and (3) individual array quality using MA plots. For each criterion, an outlier threshold was calculated from the data, and used to find whether any of the arrays was an outlier for that criterion. The final table consisted of the metric for each array in each criterion and an indication whether the array is an outlier for each criterion. A summary of the QC results for the RNA-seq, sRNA-seq, and microarray datasets can be found in Supplemental Table S1.

Visualization

For data visualization, MTEC RNA-seq and TAp73 ChIP-seq BAM files for all replicates were merged, such that for each time point and genotype a single BAM file was generated. The merged BAM files were converted into WIG files by first using the MEDIPS function 'MEDIPS.createSet' with a 50 bp windows size, and a read extension of 0, using duplicate reads. Subsequently, the files were RPKM normalized and exported as WIG files using the 'MEDIPS.exportWIG' function. The WIG files were then converted into bigWig files using the wigToBigWig UCSC script. The resulting bigWig files were used to generate genome tracks as displayed in Fig. 4 and Supplemental Fig. S6A, B.

For visualization of MCIDAS/E2F4 binding to the *Xenopus laevis TP73* gene, we downloaded ChIP-seq and RNA-seq SRA files (GSE59309) (Ma et al. 2014) and

converted them to FASTQ files using fastq-dump. The FASTQ files were aligned to the *Xenopus laevis* J-strain version 7.1 genome as previously described (Ma et al. 2014). Finally, the BAM files of replicates were merged and converted into bigWig files by using the genomeCoverageBed (with -split -bg -scale options) and wigToBigWig scripts (Fig. 6C).

SUPPLEMENTAL REFERENCES

Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, Pathak N, Liu Y, Panizzi JR, Patel-King RS et al. 2013. Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. *Am J Hum Genet* **93**: 672-686.

Burke MC, Li FQ, Cyge B, Arashiro T, Brechbuhl HM, Chen X, Siller SS, Weiss MA, O'Connell CB, Love D et al. 2014. Chibby promotes ciliary vesicle formation and basal body docking during airway cell differentiation. The Journal of cell biology 207: 123-137.

Bisgrove BW, Makova S, Yost HJ, Brueckner M. 2012. RFX2 is essential in the ciliated organ of asymmetry and an RFX2 transgene identifies a population of ciliated cells sufficient for fluid flow. *Dev Biol* **363**: 166-178.

Capece V, Garcia Vizcaino JC, Vidal R, Rahman RU, Pena Centeno T, Shomroni O, Suberviola I, Fischer A, Bonn S. 2015. Oasis: online analysis of small RNA deep sequencing data. *Bioinformatics* **31**: 2205-2207.

Chung MI, Kwon T, Tu F, Brooks ER, Gupta R, Meyer M, Baker JC, Marcotte EM, Wallingford JB. 2014. Coordinated genomic control of ciliogenesis and cell movement by RFX2. *Elife* **3**: e01439.

Chung MI, Peyrot SM, LeBoeuf S, Park TJ, McGary KL, Marcotte EM, Wallingford JB. 2012. RFX2 is broadly required for ciliogenesis during vertebrate development. *Dev Biol* **363**: 155-165.

Conesa A, Nueda MJ, Ferrer A, Talon M. 2006. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**: 1096-1102.

Crocker JC, Grier DG. 1996. When Like Charges Attract: The Effects of Geometrical Confinement on Long-Range Colloidal Interactions. *Phys Rev Lett* **77**: 1897-1900.

Denissov S, Lessard F, Mayer C, Stefanovsky V, van Driel M, Grummt I, Moss T, Stunnenberg HG. 2011. A model for the topology of active ribosomal RNA genes. *EMBO Rep* **12**: 231-237.

Didon L, Zwick RK, Chao IW, Walters MS, Wang R, Hackett NR, Crystal RG. 2013. RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. Respir Res 14: 70.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**: 1184-1191.

Dyer BW, Ferrer FA, Klinedinst DK, Rodriguez R. 2000. A noncommercial dual luciferase enzyme assay system for reporter gene analysis. *Anal Biochem* **282**: 158-161.

Halder R, Hennion M, Vidal RO, Shomroni O, Rahman RU, Rajput A, Centeno TP, van Bebber F, Capece V, Vizcaino JC et al. 2016. DNA methylation changes in plasticity genes accompany the formation and maintenance of memory. *Nat Neurosci* **19**: 102-110.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.

Holembowski L, Kramer D, Riedel D, Sordella R, Nemajerova A, Dobbelstein M, Moll UM. 2014. TAp73 is essential for germ cell adhesion and maturation in testis. *The Journal of cell biology* **204**: 1173-1190.

Horani A, Brody SL, Ferkol TW, Shoseyov D, Wasserman MG, Ta-shma A, Wilson KS, Bayly PV, Amirav I, Cohen-Cymberknoh M et al. 2013. CCDC65 mutation causes primary ciliary dyskinesia with normal ultrastructure and hyperkinetic cilia. *PLoS One* **8**: e72299.

Junek S, Chen TW, Alevra M, Schild D. 2009. Activity correlation imaging: visualizing function and structure of neuronal populations. *Biophys J* **96**: 3801-3809.

Kauffmann A, Gentleman R, Huber W. 2009. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* **25**: 415-416.

Knowles MR, Ostrowski LE, Loges NT, Hurd T, Leigh MW, Huang L, Wolf WE, Carson JL, Hazucha MJ, Yin W et al. 2013. Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *Am J Hum Genet* **93**: 711-720.

Koeppel M, van Heeringen SJ, Kramer D, Smeenk L, Janssen-Megens E, Hartmann M, Stunnenberg HG, Lohrum M. 2011. Crosstalk between c-Jun and TAp73alpha/beta contributes to the apoptosis-survival balance. *Nucleic acids research* **39**: 6069-6085.

Kwon T, Chung MI, Gupta R, Baker JC, Wallingford JB, Marcotte EM. 2014. Identifying direct targets of transcription factor Rfx2 that coordinate ciliogenesis and cell movement. *Genom Data* **2**: 192-194.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**: 1813-1831.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923-930.

Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. 2014. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* **30**: 284-286.

Lize M, Herr C, Klimke A, Bals R, Dobbelstein M. 2010. MicroRNA-449a levels increase by several orders of magnitude during mucociliary differentiation of airway epithelia. *Cell Cycle* **9**: 4579-4583.

Love D, Li FQ, Burke MC, Cyge B, Ohmitsu M, Cabello J, Larson JE, Brody SL, Cohen JC, Takemaru K. 2010. Altered lung morphogenesis, epithelial cell differentiation and mechanics in mice deficient in the Wnt/beta-catenin antagonist Chibby. *PLoS One* **5**: e13600.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Ma L, Quigley I, Omran H, Kintner C. 2014. Multicilin drives centriole biogenesis via E2f proteins. *Genes & development* **28**: 1461-1471.

Marcet B, Chevalier B, Luxardi G, Coraux C, Zaragosi LE, Cibois M, Robbe-Sermesant K, Jolly T, Cardinaud B, Moreilhon C et al. 2011. Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nature cell biology* **13**: 693-699.

Niederberger C. 2014. Re: zebrafish ciliopathy screen plus human mutational analysis identifies C21orf59 and CCDC65 defects as causing primary ciliary dyskinesia. *J Urol* **191**: 1355.

Nyengaard JR, Gundersen HJ. 2006. Direct and efficient stereological estimation of total cell quantities using electron microscopy. *J Microsc* **222**: 182-187.

Stubbs JL, Vladar EK, Axelrod JD, Kintner C. 2012. Multicilin promotes centriole assembly and ciliogenesis during multiciliate cell differentiation. *Nature cell biology* **14**: 140-147.

Tomasini R, Tsuchihara K, Wilhelm M, Fujitani M, Rufini A, Cheung CC, Khan F, Itie-Youten A, Wakeham A, Tsao MS et al. 2008. TAp73 knockout shows genomic instability with infertility and tumor suppressor functions. *Genes & development* **22**: 2677-2691.

van Dam TJ, Wheway G, Slaats GG, Group SS, Huynen MA, Giles RH. 2013. The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**: 7.

Venugopalan SR, Li X, Amen MA, Florez S, Gutierrez D, Cao H, Wang J, Amendt BA. 2011. Hierarchical interactions of homeodomain and forkhead transcription factors in regulating odontogenic gene expression. J Biol Chem. **286**: 21372-83.

Vladar EK, Stearns T. 2007. Molecular characterization of centriole assembly in ciliated epithelial cells. *The Journal of cell biology* **178**: 31-42.

Voronina VA, Takemaru K, Treuting P, Love D, Grubb BR, Hajjar AM, Adams A, Li FQ, Moon RT. 2009. Inactivation of Chibby affects function of motile airway cilia. *The Journal of cell biology* **185**: 225-233.

Weibel ER, Hsia CC, Ochs M. 2007. How much is there really? Why stereology is essential in lung morphometry. *J Appl Physiol (1985)* **102**: 459-467.

Werner-Peterson R, Sloboda RD. 2013. Methylation of structural components of the axoneme occurs during flagellar disassembly. *Biochemistry* **52**: 8501-8509.

Yang A, Walker N, Bronson R, Kaghad M, Oosterwegel M, Bonnin J, Vagner C, Bonnet H, Dikkes P, Sharpe A et al. 2000. p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* **404**: 99-103.

You Y, Richer EJ, Huang T, Brody SL. 2002. Growth and differentiation of mouse tracheal epithelial cells: selection of a proliferative population. *American journal of physiology Lung cellular and molecular physiology* **283**: L1315-1321.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.