Positional Cloning of *cobblestone*, a Mouse Mutant Showing Major Defects in Brain Development, Identifies *Ift88* as a Candidate Gene

Valentin Evsyukov[(1)], Marc A. Willaredt[(1,2)], Kerry L. Tucker[(1,2)]
[1] Interdisciplinary Center for Neurosciences, [2] Institute of Anatomy, University of Heidelberg, Heidelberg, 69120 Germany

The ENU-induced *cobblestone* (*cbs*) mouse mutant exhibits severe defects in fore- and midbrain development. Via genomic mapping, the causative mutation for the *cbs*-phenotype has previously been located on chromosome 14 in the proximity of the gene *Ift88* (*Intraflagellar transport, 88 kDa*). The *Ift88* protein is involved in intraflagellar transport and is required for genesis and maintenance of primary and motile cilia. In order to refine the genetic interval further, we performed fine-mapping analysis of the *cbs* mutant. The candidate region of the *cbs* mutation is shown to be situated in a region of 4.14 Mb on chromosome 14, containing the *Ift88* gene, and excluding dozens of genes present in the roughly-mapped interval. However, neither sequencing of the core *Ift88* promoter region nor of two conserved intronic sequences within *Ift88* revealed any mutations, indicating that the responsible mutation lies in a transcriptional regulatory sequence within or near the *Ift88* gene. Finally, other potential candidate genes in this genetic interval have been identified using synteny analysis on six other vertebrate genomes. This analysis is thus compatible with the idea that the mutation in *cbs* is located on the *Ift88* gene, but it also allows the possibility of other candidate genes that lie near *Ift88* to be involved in the phenotype.

Abbreviations: *cbs*, *cobblestone*; *Ift88*, *Intraflagellar transport, 88 kDa*; SNP, single nucleotide polymorphism; ENU, N-ethyl-N-nitroso-urea; EGFP, enhanced green fluorescent protein

Keywords: cortex, heterotopia, *Ift88*, intraflagellar transport, primary cilia, SNP

Introduction

Cilia are organelles protruding from the surface of nearly all eukaryotic cells. Cilia are usually found either in large clusters or as solitary structures on the cell surface. They are commonly classified into motile and primary cilia. Motile cilia that line structures such as the respiratory tract and the cerebral ventricles beat in an orchestrated fashion, thereby inducing a directional flow of mucous and cerebrospinal fluid, respectively (Davenport and Yoder, 2005). Both primary and motile cilia share a microtubule-based axoneme that emerges from the centriole-derived basal body. Although it is known that primary cilia are important components in chemo- and mechanoperception, their presence on nearly all eukaryotic cells is puzzling. Recently, proteins of several signal transduction pathways such as Sonic hedgehog (SHH) (Huangfu et al., 2003) and Wnt (Corbit et al., 2008) have been shown to localize to the primary cilium. Both pathways are integral to developmental processes of the eukaryotic organism. While it has been widely known that defective

motile cilia are responsible for several human pathologies, such as primary cilia dyskinesia (PCD) and its symptoms (including respiratory infections, anosmia and male infertility; Bisgrove and Yost, 2006), non-motile primary cilia have long been deemed irrelevant for human pathology. This view has been challenged as several proteins have been observed to localize to the primary cilium in patients with polycystic kidney disease (PKD), such as fibrocystin/polyductin, the protein encoded by the *PKHD1* gene, which is responsible for this disease, arguing that it could be a malfunction in primary cilia that is responsible for the particular pathology (Wang et al., 2004). Many other multi-spectrum syndromes have recently been shown to have defects in proteins localized to primary cilia, and thus this organelle is now widely regarded as the cause of a diverse array of human pathologies.

The involvement of cilia in PKD has been supported by numerous mouse models. Many of these models have been shown to revolve around proteins localized to the primary cilium (Davenport and Yoder, 2005). The most widely known of these models is the Oak Ridge Polycystic Kidney ($Tg737^{Orpk}$) mutant. Mice carrying the $Tg737^{Orpk}$ allele exhibit renal cysts (Moyer et al., 1994), defects of the liver and pancreas (Zhang et al., 2005), and skeletal abnormalities such as polydactyly (Lehman et al., 2008). This hypomorphic allele of the *Ift88* gene encodes the intraflagellar transport (IFT) protein IFT88 (Taulman et al., 2001). IFT is the mechanism by which protein cargoes are shuttled in and out of the cilium (Kozminski et al., 1993). IFT is required to build and maintain the structural and functional integrity of the cilium. After passing the ciliary necklace region (a transition area between the basal body and the axoneme), protein complexes are transported between the axoneme and its surrounding membrane via kinesin- (anterograde) and dynein-driven (retrograde transport) motors. The IFT machinery proteins are subdivided into two complexes, A and B, according to their sedimentation in a sucrose gradient (Cole et al., 1998).

Our group has previously performed an N-ethyl-N-nitroso-urea-(ENU)-based mutagenesis screening for mouse mutants exhibiting defects in the development of the nervous system. One of the resulting mutants was the *cobblestone* mutant (*cbs*), named after distinct neuronal heterotopias of the embryonic brain found in the homozygous mutant (Willaredt et al., 2008). Several other major developmental defects are present: the formation of the dorso-medial telencephalon is defective, with the choroid plexus, the cortical hem and the hippocampal anlage also failing to form properly. All homozygous mutants die during embryonic development (Willaredt et al., 2008).

The *cbs* mutation has been generated on a background containing the *tauGFP* transgene, a genetically-modified locus from which all newborn neurons express the green fluorescent protein EGFP (Tucker et al., 2001). In order to identify the gene responsible for the *cbs* mutation, traditional genetic analysis and positional cloning was performed using genetic markers that identify individual chromosomes in the mouse genome. Using single nucleotide polymorphism- (SNP)-based rough mapping, the mutation has been localized on chromosome 14 in a region containing the gene *Ift88* (Willaredt et al., 2008). The *Ift88* gene has a size of almost 100 kb, containing 26 exons. Because a targeted mutation of *Ift88* results in a phenotype that resembles the *cbs* phenotype (Murcia et al., 2000), we have focused upon this gene. Real-time reverse transcription PCR, Northern, and Western blot analyses all have shown *Ift88* mRNA and protein levels to be reduced by 70–80% in the mutant, and a complementation analysis strongly suggested that *Ift88* is in fact the mutant gene (Willaredt et al., 2008). In order to narrow down the candidate region for the *cbs* mutation, fine mapping of the mutation was performed using SNP markers in order to pinpoint its genetic location. The promoter region and two intronic areas of *Ift88* with high predicted regulatory potential

were sequenced, and a shared synteny analysis between mouse and several other species was performed, searching for genes that could potentially influence *Ift88* transcription. Several potential regions for the mutation were thus excluded, while others remain to be considered.

Materials and Methods

Mouse lines

All animal experiments were in compliance with the regulations of Baden-Württemberg, Germany. We employed the strain *cbs* (Willaredt et al., 2008), which was originally derived on a C57BL/6J background, but for purposes of genetic analysis were crossed to mice of the inbred strain CBA/J. Mice were housed in a Specified-Pathogen Free barrier facility with a 12-hour light cycle and fed ad libitum with water and normal mouse chow. For mouse matings, mice were put together at 4 PM and females checked the next day for a vaginal plug. 12:00 PM of this day was considered 0.5 days *post coitum*, for the purpose of staging embryos.

Amplification of DNA for sequencing

The region of interest was sequenced in short overlapping spans with a mean length of about 850 bp. Since sequencing of a given span does not provide accurate sequencing data for its beginning and end, the chosen spans overlap with one another, so as to secure accurate sequencing data for the entire region. The chosen spans are shown in Figure 1.

Amplification was performed in 20 µl reactions on a Mastercycler gradient thermocycler (Eppendorf, Hamburg, Germany). The following enzymes were used for amplification: NEB Taq DNA Polymerase/ThermoPol (Ipswich, MA, USA) buffer for gradient PCR; Roche FastStart High Fidelity PCR System (F. Hoffmann-La Roche AG, Basel, Switzerland) for amplification of mutant DNA; NEB Phusion High-Fidelity DNA Polymerase/HF-buffer and GC-buffer (for amplification of the CpG-island situated at the very beginning of *Ift88*) from the NEB Phusion High-Fidelity PCR Kit for amplification of mutant DNA. Template DNA concentration was 50 ng/µl. Sequencing was performed by GATC Biotech (Constance, Germany).

Primers were designed using Primer3 (Rozen and Skaletsky, 2000). Seven primers were used for amplification of the 2.8 kb region spanning the 2.5 kb directly upstream from *Ift88* and the first coding 300 bp (Table 1). DMSO (3%) was added to PCR reactions. Three additional primers (#A, #B/1, and #B/2) were used for amplification of the two intronic regions of *Ift88* (Table 1).

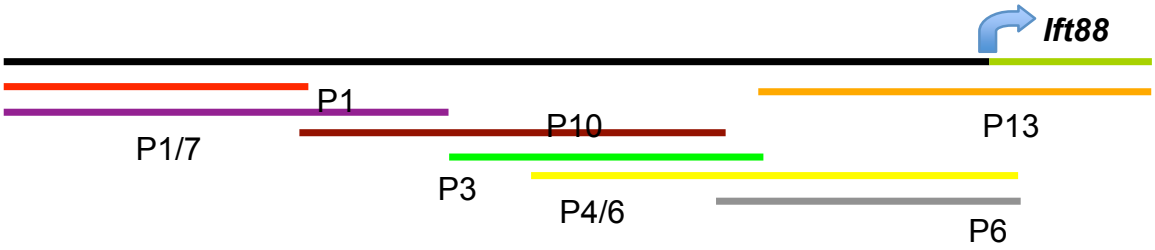

Figure 1: Contig map showing the spans used for sequencing of the sequenced region directly upstream of *Ift88*, and a small part of the gene itself. The start of transcription of *Ift88* is indicated by an arrow. The upper line represents the sequenced section of mouse chromosome 14. The lines below represent the spans used for sequencing of this region. The contig covers a region of 2793 bp, stretching from 58,040,455 bp to 58,043,247 bp on mouse chromosome 14, according to NCBI mouse genome build 37.

Genomic mapping

The *cbs* mutation has been generated on a pure C57BL/6J background, and the mice were crossed to the CBA/J strain, a genetically-different and widely used inbred mouse line. In the mapping analysis, SNP markers that were located close to the *cbs* mutation were shown to be homozygous for the mutation from the strain of origin, namely C57BL/6J. Nearby areas of the genome that did not show such a pattern in embryos that were phenotypically mutant could thus be excluded from the area of interest. Mapping was performed on an ABI PRISM 7000 real-time PCR machine (Applied Biosystems, Foster City, CA, USA) using Applied Biosystems TaqMan SNP genotyping assays for mouse (Table 2) and TaqMan Universal PCR Master Mix, No AmpErase UNG. The SNP assay stocks, delivered at a 40x concentration, were diluted to 20x with 1x TE buffer, pH 8.0.

Ten ng of DNA were used per reaction in a total reaction volume of 25 µl.

Nine phenotypically mutant and one wild type embryo were used for the mapping. The sample names are comprised as follows: the number preceding "♀" refers to the mother, while the second number refers to a specific embryo within the litter, being either homozygous *cbs* mutant ("mut") or phenotypically wildtype ("wt"): 530♀mut6, 378♀mut1, 378♀mut2, 378♀mut4, 318♀mut1, 313♀mut1, 153♀mut1, 313♀mut2, 313♀mut3, 13♀wt1. Most mapping reactions were performed in triplicates, so as to determine the robustness of the method. Cycling and read-out were performed in Applied Biosystems MicroAmp 96-well optical reaction plates. Pure C57BL/6J, CBA/J and artificial F1 (C57BL/6J and CBA/J 1:1 mixture) was used as the reference for genotyping of the embryo samples. Mapping was performed in two independent PCR runs for some of the embryos, using triplicate reactions for most of the samples (13♀wt1 mapping was performed in duplicates except for the downstream mapping in the second run, where only one DNA sample was used; 378♀mut2 mapping was performed in a triplicate for the upstream marker and a duplicate for the downstream marker; 153♀mut mapping was performed in a duplicate for the downstream marker in the second run and in triplicates in the remaining cases; all remaining samples were mapped using triplicate reactions. Varying reaction numbers were due to limited amounts of sample DNA. In No-template controls (NTC) water used instead of DNA.

Table 1: Primer sequences used for amplification of the 2.8 kb region spanning the 2.5 kb directly upstream from *Ift88*, the first coding 300 bp, and the two intronic regions of *Ift88*. *fragments amplified using primer pairs P3, P4/6 and P1/7 have been created in previous work in the laboratory and subsequently ligated into TOPO vectors (Invitrogen, Carlsbad, CA, USA). For this reason, no annealing temperatures are provided here. P13 is covering a CpG-island. NEB Phusion with the corresponding GC-buffer was used for PCR with this primer pair.

Span	Fwd.	Rev.	Anneal. temp.	Length
P1	TCTTCAGACACACCAGAAGAGG	GGAATAGGGGGATGTAGAACG	60,2 °C	771 bp
P1/7 (*)	TCTTCAGACACACCAGAAGAGG	GTCCCATGAAGTGGGTATGG	-	1129 bp
P3 (*)	CTCAGGCTTGTCTGCTACCC	GCTTCAAGTCAACATGCCC	-	801 bp
P4/6 (*)	AAGTTCAGTGCTGGACAATGG	ATCCTAGGCAGGCCAGAGC	-	510 bp
P6	ACACCAAGTAGCCCGAACC	ATCCTAGGCAGGCCAGAGC	61,6 °C	758 bp
P10	CGTTCTACATCCCCCTATTCC	GGTTCGGGCTACTTGGTGT	60,2 °C	1023 bp
P13	GCTTCAAGTCAACATGCCC	CCAAGCCGTGTGACATTAGC	66,0 °C	883 bp
#A	GTCCCTGTTGAAGAGGTCCA	CCATTGCACGTGTTCAAATC	55.7 °C	507 bp
#B/1	ATTTTGTTGGTGGCGTTCTC	TGATGGTTTTCCTGCCTACC	63,1 °C	637 bp
#B/2	GCCACAGACTTAGGACCCAGAACG	GAGGTCCTAACAAGTAAGCCCAGTG	65 °C	1 kb

Assay ID	Context sequence ([CBA/J]/[C57BL/6J])
M__23131531_10	TTGGAAGCATCATTTTACATTCTGA[C/T]TGTATGGCTTCATGGCATCTTTGTC
M__23590757_10	CAACTAGCAGTGTTTGACGTAATCC[A/G]GCCCTTGGCATAAGAGTTCCCTTCCP

Table 2: SNP genotyping assays for mouse. Assay IDs: M__23590757_10, 55,265,184 Mb on chromosome 14, and M__23131531_10, 59,404,366 Mb on chromosome 14.

Applied Biosystems sequence detection software, provided with the ABI PRISM PCR machine, was used.

Ligation into TOPO vectors

Commercially available TOPO cloning kits (Invitrogen, Carlsbad, CA, USA) were used for cloning of PCR products. Depending on whether the particular PCR product was blunt-end or had an A overhang, either the pCR-Blunt II-TOPO or the pCR2.1-TOPO vector, respectively, was used according to the manufacturer's instructions.

Restriction digestion analysis

The total reaction volume was 10 µl, using 5 µl of DNA solution with 10 units of each enzyme (in case of double digests) and the manufacturer-supplied buffers. Deionized water was added to a final volume of 10 µl. All enzymes and buffers were obtained from NEB and Fermentas (Ontario, Canada).

Transformation of bacteria

In order to obtain adequate amounts of plasmid DNA containing the amplified promoter stretches, chemically competent *Escherichia coli* bacteria (Invitrogen OneShot TOP10 and OneShot MAX Efficiency DH5α-T1^R, as well as XL1-Blue) were transformed according to the manufacturer's instructions using the desiredplasmids. Several colonies were subsequently picked for further cultivation in selective LB medium overnight at 37 °C and 225 rpm TH 15 and KS-15) (Edmund Bühler, Hechingen, Germany). Afterwards, plasmid was harvested.

Bioinformatic analysis of *Ift88*

Several common bioinformatics web-databases were used for the analysis of sequenced *Ift88* promoter and intronic regions, as well as for syntenic analysis of the regions containing the *Ift88* gene in different species, functional predictions of different areas of the *Ift88* gene and further information on *Ift88* and its surrounding genes. The following databases were used for retrieving sequence and gene product information on *Ift88* and other genes:
http://genome.ucsc.edu/,
http://www.ensembl.org,
http://www.ncbi.nlm.nih.gov,
http://www.informatics.jax.org.

Other features of the ENSEMBL and UCSC databases were used to predict functional sequence areas such as X-boxes or regulatory potential and to visualize conservation.

Results

The cbs mutation is situated in a region containing Ift88 SNP-based genomic mapping was performed in order to narrow down the candidate region of the *cbs* mutation. The analysis was used to distinguish between the C57BL/6J- (the background carrying the mutation) and the CBA/J-mouse lines (used for cross-breeding for mapping). Because the mutant phenotype is only present in *cbs* homozygotes, heterozygous or CBA/J-

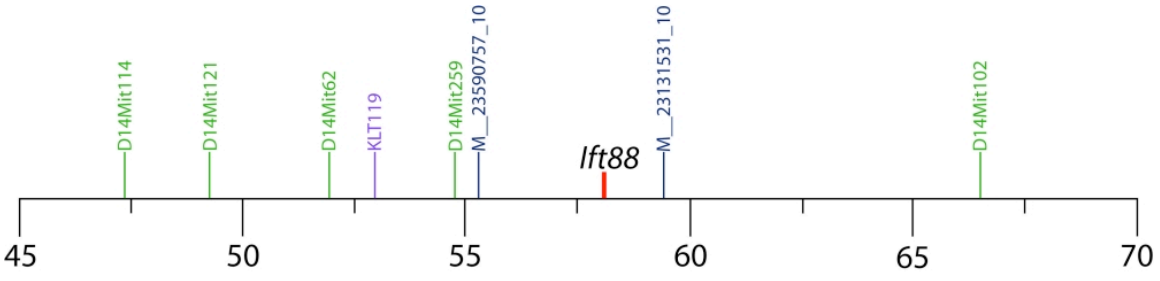

Figure 2: Position of SNP (KLT119 and M__x) and SSLP (D14x) markers on mouse chromosome 14 used for genomic mapping in order to narrow down the localization of the *cbs* mutation. Positional reference: NCBI mouse genome build 37.

homozygous SNP discrimination results made it possible to narrow down the candidate region for the mutation. We had previously narrowed down the mutation to a candidate region of about 11.7 Mb, situated between the SSLP markers D14Mit259 and D14Mit102, as shown in Figure 2. By standard statistical analysis, the mutation should be found in a 4.5 Mb stretch distal to the D14Mit259 marker (Willaredt et al., 2008), lying within a 95% confidence interval (see Silver, 1995, pp. 294-304). However we wished to exclude as much of the 11.7-Mb area as possible using further mapping markers. The Applied Biosystems TaqMan SNP genotyping assays with the product IDs M__23590757_10 and M__23131531_10, which were the nearest-situated genotyping assays to *Ift88* at the time of the experimental work, were used for this purpose. The distance between these markers is approximately 4.14 Mb. Genomic DNA from nine phenotypically mutant embryos and one wild type embryo were used. Comparison of the relative fluorescence intensities and the temporal fluorescence increase of the embryonic DNA with these parameters in the references made it possible to determine the particular genotype of a given embryo. In Figure 3, which shows the clustering of the samples to the references based on the fluorescence of the particular dyes, this particular sample clusters with the C57BL/6J reference group. Figure 4 shows the respective allelic discrimination results for the distal SNP marker. All results were confirmed in additional runs (data not shown).

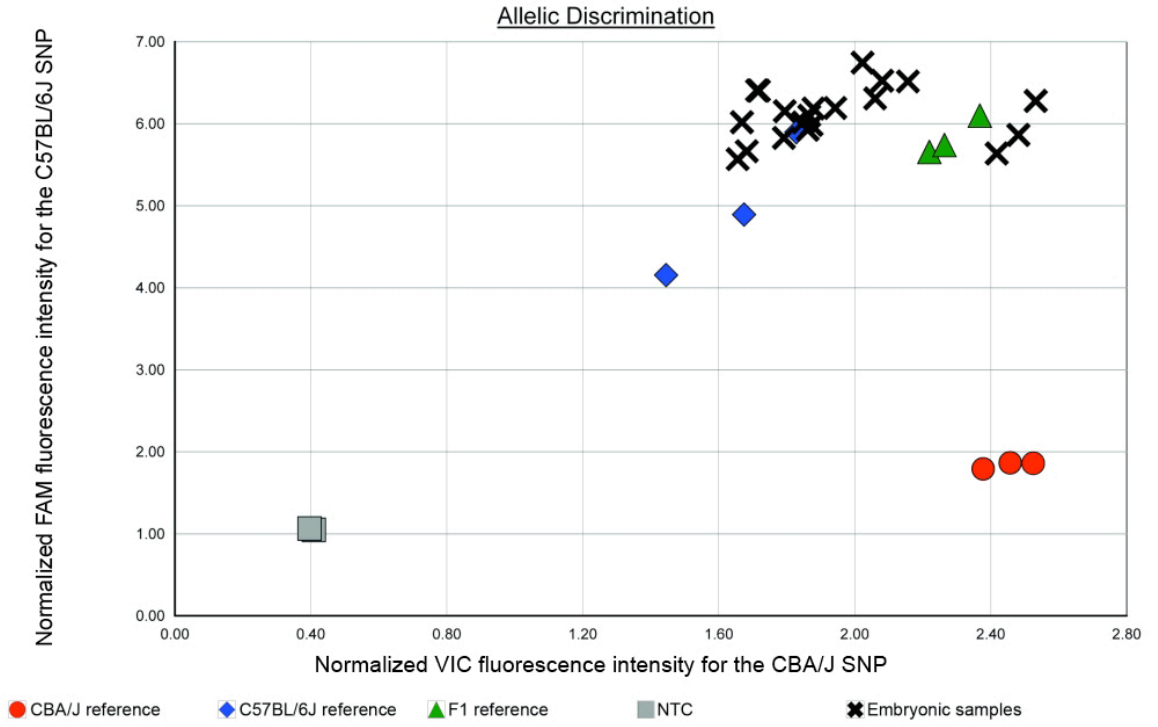

Figure 3: Mapping results for the proximal SNP marker, first PCR run. The three samples clustering with the F1 reference are 530 ♀mut6. The remaining samples are: 318♀mut1, 313♀mut1, 153♀mut1, 313♀mut2, 313♀mut3, 13♀wt1. NTC: no-template control.

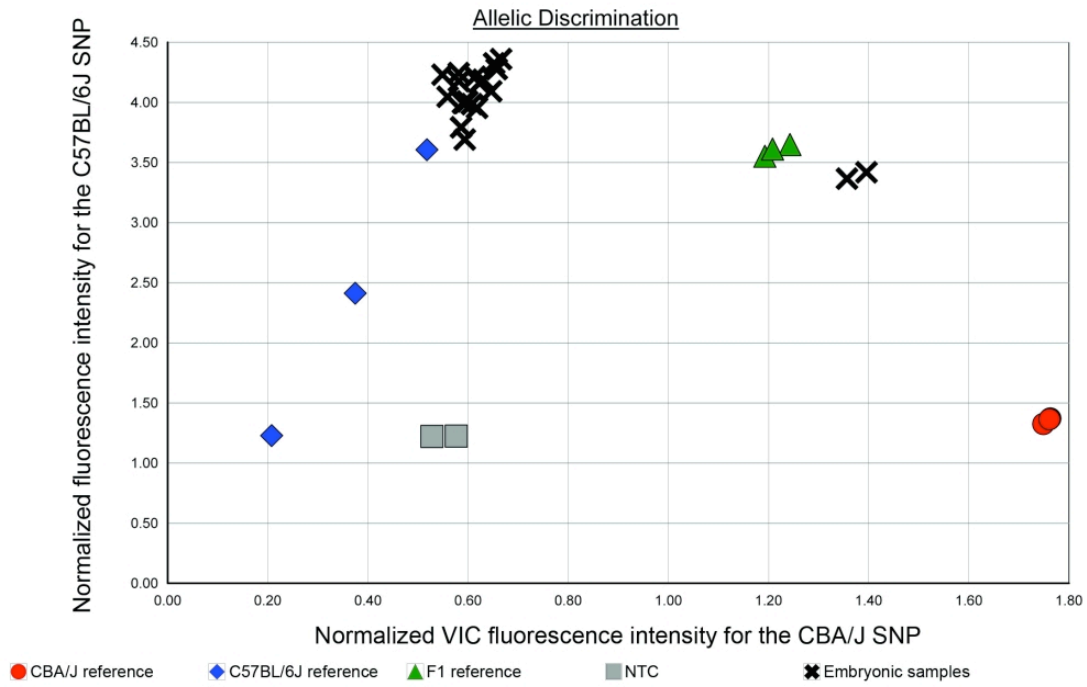

Figure 4: Mapping results for the distal SNP marker, first run. The two samples clustering with the F1 reference are 13□wt1. The remaining samples are: 318♀mut1, 313♀mut1, 153♀mut, 313♀mut2, 313♀mut3, 530♀mut6. NTC: no-template control.

Sequencing of the Ift88 core promoter region did not reveal any mutations

Strong downregulation of *Ift88*-expression in the *cbs* mouse mutant, as observed by Western blot, Northern blot and RT-PCR analyses, and a striking similarity of the *cbs* phenotype to the *Ift88* knockout-phenotype (Murcia et al., 2008), suggested that the ENU-induced mutation is located either in the *Ift88* gene itself or in an area affecting the specific expression of this gene. However, sequencing of the *Ift88*-mRNA revealed no mutations (Willaredt et al., 2008). As a mutation in the promoter region of the gene could well account for changes in its expression strength and pattern, 2500 bp upstream and 300 bp downstream of the first exon were sequenced and compared to the corresponding wildtype-sequence via BLAST-analysis using the NCBI online database. Figure 1 shows the overlapping

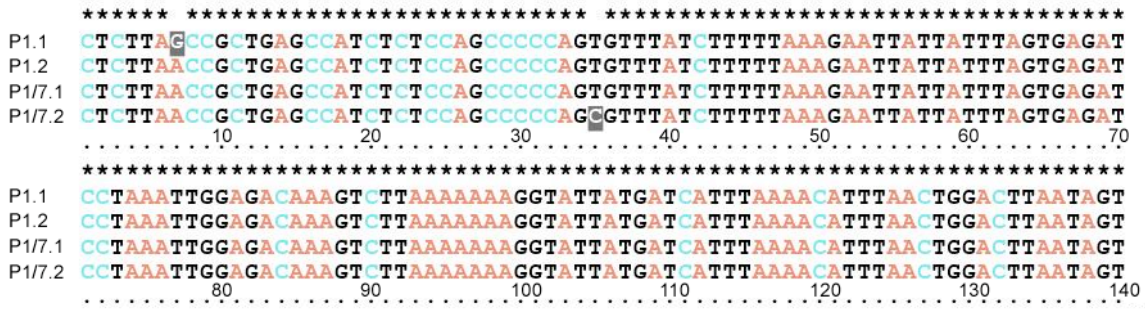

Figure 5: Excerpt of raw sequence from the sequencing of the *Ift88* promoter region. Bacterial colonies were created using either the P1 span PCR product, or the P1/7 span PCR product. Upon transformation and growth of the bacteria with the respective recombinant plasmid containing the PCR product, two colonies for each span were used for sequencing. The asterisks represent a match with the NCBI online database entry using BLAST.

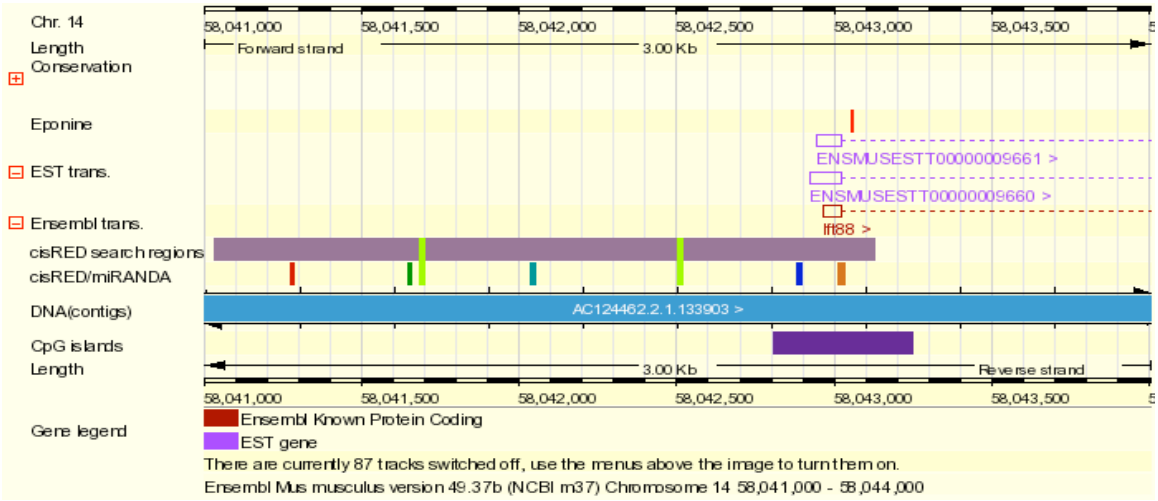

Figure 6: The promoter region of mouse *Ift88* is shown. Using the miRANDA feature of the ENSEMBL online database, five putative regulatory regions were determined, four of which are situated upstream of *Ift88*, and one spanning the end of the first exon and the beginning of the first intron, as indicated by the small colored boxes to the right of the text "cisRED/miRANDA". The two tall lime-green boxes demarcate putative X-box regions that have been previously determined via BLAST analysis against the mouse *Ift88* promoter region using X-box sequences from the *C. elegans* orthologue of *Ift88*, *osm-5*.

spans comprising the chosen promoter region. Most sequencing reactions of the *cbs* mutant DNA produced mismatches with the wildtype NCBI-database DNA using BLAST. None of these mismatches, however, were confirmed in other embryos (Figure 5). Together, these results indicated that the mutation responsible for the *cbs*-phenotype is not situated in the *Ift88* core promoter region.

This result was surprising as several regulatory elements have been predicted to be situated in this region via bioinformatic analysis. Among these regulatory regions putative X-box sequences, DNA motifs associated with the transcriptional regulation of ciliary genes (Efimenko et al., 2005), have been found (Figure 6).

Two conserved intronic regions of Ift88 with predicted regulatory potential are not mutated in the cbs mutant

In addition to sequencing of the *Ift88* promoter region, two conserved intronic elements between exons 3 and 4 (named span #B) and exons 14 and 15 (named span #A) with a high predicted potential to regulate transcriptional activity, based upon bioinformatics analysis using the online

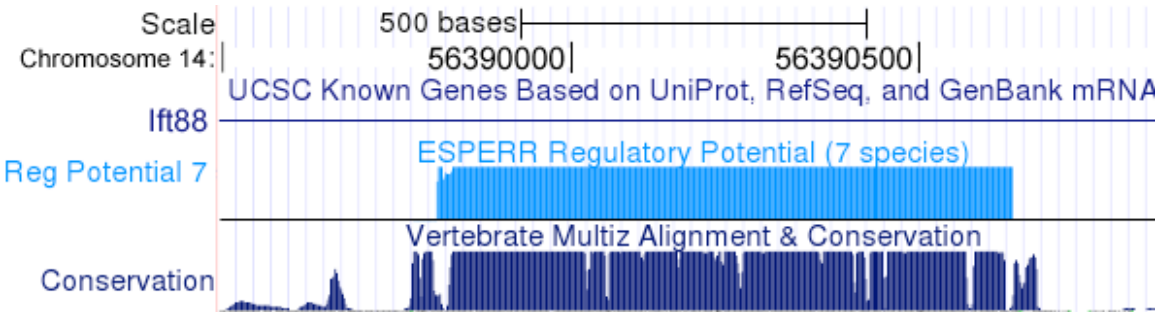

Figure 7: One of the two sequenced intronic regions, #B. Light blue and dark blue bars indicate regions with predicted regulatory potential and conservation, respectively. Reference: NCBI mouse genome build 36. Created using UCSC Genome Browser (http://genome.ucsc.edu/).

UCSC Genome Browser (http://genome.ucsc.edu/), were amplified, cloned, and sequenced in the same way as for the promoter region. These regions could conceivably control the transcriptional levels of *Ift88*, and were therefore of great interest. A graphical representation of span #B is shown in Figure 7. Sequencing of these two intronic regions did not show any mutations.

Synteny analysis of the Ift88 genomic area

The shared synteny between mouse and six other vertebrate species was analyzed: *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Canis familiaris* (dog), *Equus caballus* (horse), *Gallus gallus* (chicken), *Oryzias latipes* (Japanese medaka). The ENSEMBL and NCBI databases were used for addressing the synteny around *Ift88*. To this end, the online ENSEMBL synteny viewer as well as simple comparison "by hand" were applied, i.e. the gene names and their order as represented in the aforementioned databases (see Materials and Methods) were compared between mouse and several chosen species. Within a certain interval the genes around *Ift88* and their order were essentially identical in all examined organisms. The point at which the reference species showed a major deviation in gene succession from the murine chromosome was considered the end of the examined synteny block. Although the mouse synteny with the primate organisms ends earlier than the final designated synteny block, namely at *Mphosph8*, the syntenic allocation with the other species was continued farther until *Atp12a*. The consideration behind this methodology is the high evolutionary stage of human and chimpanzee, that could account for a difference in regulation of *Ift88* expression.

The resulting block of shared synteny containing *Ift88* has a size of 1.7 Mb on mouse chromosome 14 and stretches from the gene *Atp12a* at 57.0 Mb to *Fgf9* at 58.7 Mb (according to NCBI mouse genome build 37). Several of the products of the genes that were found in this region of synteny are reported to be involved in transcriptional regulation: *CenpJ* (Peng et al., 2002), *Pspc1* (Kuwahara et al., 2006), *Zmym5* (Pastorcic, Das, 2007), *Zmym2* (Glocke, Yu, 2008), *N6amt2* (as a DNA methyltransferase, the gene product is very likely to affect gene transcription via epigenetic mechanisms), *Lats2* (Powzaniuk et al., 2004), *Sap18* (Paces-Fessy et al., 2004), and *Fgf9* (White et al., 2007). After selecting those genes in the *Ift88* synteny block that encode proteins involved in transcriptional regulation, we investigated which of these genes have been reported to be expressed in the mouse brain and if they have been previously eliminated by gene targeting in mouse. The knock-out phenotypes were then compared to the *cbs* phenotype. Only two of the candidate genes have been targeted in the mouse: *Lats2* (encoding a serine/threonine protein kinase that localizes to centrosomes during interphase and early and late metaphase) and *Fgf9* (encoding a Shh-dependently expressed growth factor). *Lats2*: Homozygous knockout mice die during embryonic development, showing decreased cell proliferation, chromosomal instability, atrial hyperplasia, ventricular hypoplasia, and a kinked neural tube (McPherson et al., 2004; Yabuta et al., 2007). *Fgf9*: Knock-out of this gene results in a reduced size of the embryos, pulmonary hypoplasia, cardiac dilation, impaired testicular development with male-to-female sex reversal, retinal defects, and neonatal lethality (Colvin et al., 2001; Lin et al., 2006).

Discussion

Although the site of the mutation responsible for the *cbs* phenotype has not yet been determined, the candidate region of the mutation was localized to a stretch of 4.14 Mb on chromosome 14, distal to the SNP marker situated at 55.26 Mb in the current NCBI mouse genome build 37 (Figure 8). Previous rough mapping analysis using SNP

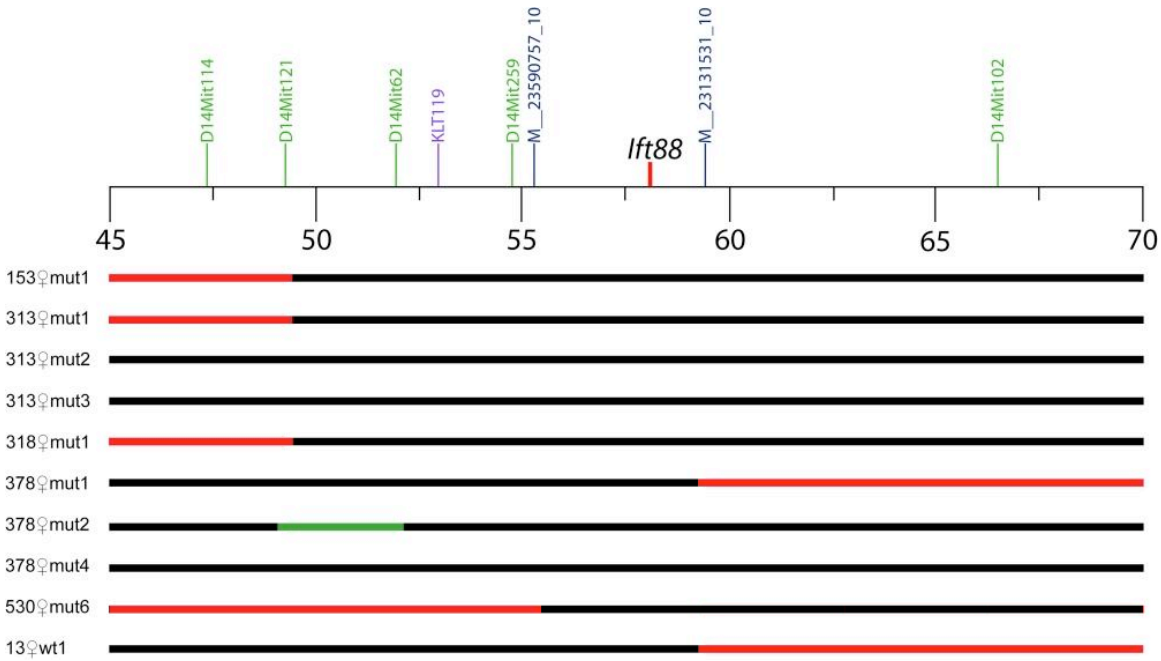

Figure 8: Results of the positional cloning performed to identify the genetic locus responsible for the *cbs* mutation. The lines below the chromosome fragment show the genotyping results for the respective mouse embryos based on allelic discrimination analysis. SSLP (D14Mitx) and SNP (KLT119 and M__x) markers were used to distinguish between the mouse lines C57BL/6 (carrying the mutation) and CBA/J (cross-breeding background). Red areas indicate a CBA/J/C57BL/6J-heterozygous configuration, black areas indicate C57BL/6J homozygous intervals. The green span in 378□mut2 shows a CBA/J-homozygous background.

markers had located the *cbs* mutation on chromosome 14, and subsequent mapping using SSLP markers had placed the mutation in a 4.5 Mb stretch distal to the SSLP marker D14Mit259 (Willaredt et al., 2008), lying within a 95% confidence interval (see Silver, 1995, pp. 294-304). One of the genes in this region is *Ift88*. This gene has previously been knocked out in the mouse, and the resulting phenotype strongly resembled the *cbs* phenotype, including poorly-characterized defects in neural development, polydactyly, and prenatal death (Murcia et al., 2000). Due to the profound disorganization of cerebral structures in the *cbs* mutant, our group has put its focus on examining this particular gene.

These results strongly suggest that the mutation responsible for the *cbs* phenotype is situated in the *Ift88* gene itself or in a regulatory sequence lying in relatively close proximity to this gene. We have performed fine mapping in order to narrow down the candidate region using commercially available Applied Biosystems TaqMan SNP genotyping assays. Several phenotypically-mutant mouse embryos were used for fine mapping, employing molecular markers that can distinguish C57BL/6J (in which the *cbs* mutation lies) from CBA/J (which was used to allow mapping) inbred mouse strains. Two of these mutants were identified as heterozygous for the distal or the proximal SNP marker, respectively, indicating that they are recombinants. Since the *cbs* mutant only shows a phenotype when homozygous, a heterozygous genotyping result for the proximal SNP marker places the *cbs* mutation distal to that marker (Figure 2). By the same logic, a heterozygous genotyping result for the distal marker places the *cbs* mutation proximal to that marker (Figure 8). Because rough mapping analysis has placed the mutation 4.5 Mb distal to the SSLP marker D14Mit259 (Figure 8), the heterozygous mapping result of the 530♀mut6 mutant embryo for the proximal SNP marker, which is situated 0.54 Mb distal from the

D14Mit259 marker, narrows down the mutation to a probable stretch of 4.14 Mb distal to the SNP marker M__23590757_10, which is situated upstream of *Ift88* (Figure 8). Scattering of the cluster data points in Figures 4 and 5 is most likely due to pipetting errors. Because sequencing of the *Ift88* mRNA had not revealed any mutations (Willaredt et al., 2008), the *cbs*-mutation is most likely situated in a regulatory region inside or associated with *Ift88*.

As a diminished expression level could be caused by a mutation within the promoter itself, a region of 2.8 kb spanning the first 2.5 kb upstream of *Ift88* containing the promoter region, and the first 300 bp of the transcribed gene part containing a predicted regulatory element, were sequenced using DNA of *cbs/cbs* mutant embryos and compared with the corresponding wildtype sequences available from the NCBI online database. Surprisingly, this region turned out to be free of mutations, as mismatches were not consistent throughout different embryos, indicating random errors introduced by the Taq polymerase in individual clones. Additionally, two intronic sequences that are conserved among several vertebrate species and have a high predicted regulatory potential were sequenced in genomic DNA extracted from the *cbs* mutant. These areas were named span #A (extends from 56,413,351 Mb to 56,413,858 Mb on mouse chromosome 14), and span #B (extends from 56,389,495 Mb to 56,390,884 Mb on chromosome 14, according to NCBI mouse genome build 36). Again, no mutations were uncovered.

Classical complementation analysis of the *cbs* mutation with a targeted, disrupted allele of *Ift88* reproduced the *cbs* phenotype in compound heterozygotes, strongly suggesting that the two mutations do not complement each other and therefore are situated in the same gene (Willaredt et al., 2008). However, the possibility still exists that the *cbs* mutation is not in the *Ift88* gene itself but in an associated regulatory region in its vicinity, or even in another gene that is involved in the regulation of *Ift88* expression, acting in a *cis* orientation. For this reason the shared synteny between mouse and six other vertebrate species was analyzed, and the resulting block of shared synteny containing *Ift88* was found to consist of a stretch of 1.7 Mb on mouse chromosome 14. Some of the genes in this block present in mouse are either not present in the particular synteny block of some of the species or not annotated, and the exact order of the genes is not identical in all mentioned organisms, which does not prevent the consideration of the region between *Atp12a* and *Fgf9* as a synteny block, since the overall shared synteny is upheld. The point at which the examined species showed several genes that are positioned either farther apart on mouse chromosome 14 or on an entirely different chromosome was considered the end of the designated synteny block. The synteny of human and chimpanzee with mouse ends earlier than that of the remaining reference species. It was nonetheless decided to consider *Atp12a* the end of the block, since this particular region in dog and horse corresponds to that of mouse. In some organisms there were insertions of genes not present in the vicinity of *Ift88* in mouse. Before and after these regions, the names and order of the genes were the same as in mouse. Thus, these regions were not considered to break up the synteny with mouse.

Although the exact position of the mutation responsible for the *cbs* phenotype remains elusive, using genomic mapping techniques we were able to narrow down the candidate region to an area of about 4 Mb on chromosome 14. This region contains the *Ift88* gene, which plays a pivotal role in the development of the mammalian organism. Since the *Ift88* gene product levels are significantly decreased in homozygous *cbs* mutants, this gene may be indeed affected in the *cobblestone* mutant. Having reduced mRNA levels in the mutant, it was likely that the mutation was within the promoter region of *Ift88*. However, sequencing of the first 2.5 kb directly upstream of *Ift88* did not reveal any mutations. Other potential sites

for the mutation are intronic regions of the gene. Therefore, two intronic regions, which are highly conserved in vertebrates and may have a role in the regulation of gene expression, were sequenced but, also, did not contain any mutations. In theory, the mutation could affect correct splicing of the gene. However, no convincing evidence of functional splicing variants of the mouse *Ift88* protein is available to this day. Since the candidate region for the mutation contains other genes as well, a syntenic analysis between several vertebrate species was performed in order to find genes that might somehow affect the expression of *Ift88*. Eight genes, *CenpJ*, *Pspc1*, *Zmym5*, *Zmym2*, *N6amt2*, *Sap18*, *Lats2* and *Fgf9*, were considered particularly interesting and suitable for further investigation. Although the idea of one of the genes in the synteny block other than *Ift88* being mutated may appear far-fetched, the uncertainty of the mutation position in the *cobblestone* mouse and the apparent genetic linkage between the genes surrounding *Ift88* do not make this idea implausible.

Further localization of the mutation requires additional mapping analyses using markers that are situated closer to *Ift88* in order to narrow down the candidate region as much as possible. If by means of fine genomic mapping the *cbs* mutation is shown to be indeed situated in *Ift88* or its close proximity, sequencing of a large DNA region covering *Ift88* should unveil the mutation. Although no human diseases have yet been identified that are directly caused by mutations in *Ift88*, its contribution to the maintenance of cilia and their fundamental role in a wide range of diseases will allow the cbs mouse to serve as a potential model for these diseases. An elucidation of the fine genetic interval, as demonstrated in this study, is a crucial step in the determination of the exact genetic mutation causing the defects in brain development seen in the mutant.

Acknowledgements

The authors would like to thank Prof. Klaus Unsicker for use of the ABI Prism 7000. Support provided by the German Research Foundation (DFG, SFB 488, Teilprojekt B9).

Corresponding Author

Valentin Evsyukov
Interdisciplinary Center for Neurosciences
University of Heidelberg
Im Neuenheimer Feld 307 / 4. OG
69120 Heidelberg
Germany
Email:Evsyukov@stud.uni-heidelberg.de

References

Bisgrove BW, Yost HJ (2006) The roles of cilia in developmental disorders and disease. *Development 133*, 4131-43.

Cole DG, Diener DR, Himelblau AL, Beech PL, Fuster JC, Rosenbaum JL (1998) Chlamydomonas kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in Caenorhabditis elegans sensory neurons. *J Cell Biol 141(4),*993-1008.

Colvin JS, Green RP, Schmahl J, Capel B, Ornitz DM (2001) Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell 104*, 875-89.

Corbit KC, Shyer AE, Dowdle WE, Gaulden J, Singla V, Chen MH, Chuang PT, Reiter JF (2008) Kif3a constrains β-catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms. *Nat Cell Biol 10*, 70-6.

Davenport JR, Yoder BK (2005) An incredible decade for the primary cilium: a look at a once-forgotten organelle. *Am J Physiol Renal Physiol 289*, F1159-69.

Efimenko E, Bubb K, Mak HY, Holzman T, Leroux MR, Ruvkun G, Thomas JH, Swoboda P (2005) Analysis of xbx

genes in C. elegans. *Development 132*, 1923-34.

Gocke CB, Yu H (2008) ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One 3(9)*, e3255.

Huangfu D, Liu A, Rakeman AS, Murcia NS, Niswander L, Anderson KV (2003) Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature 426*, 83-87.

Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL (1993) A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proc Natl Acad Sci USA 90*, 5519-23.

Kuwahara S, Ikei A, Taguchi Y, Tabuchi Y, Fujimoto N, Obinata M, Uesugi S, Kurihara Y (2006) PSPC1, NONO, and SFPQ are expressed in mouse Sertoli cells and may function as coregulators of androgen receptor-mediated transcription. *Biol Reprod 75(3)*, 352-9.

Lehman JM, Michaud EJ, Schoeb TR, Aydin-Son Y, Miller M, Yoder BK (2008) The Oak Ridge Polycystic Kidney mouse: Modeling ciliopathies of mice and men. *Dev Dyn 237*, 1960-71.

Lin Y, Liu G, Wang F (2006) Generation of an Fgf9 conditional null allele. *Genesis 44*, 150-4.

McPherson JP, Tamblyn L, Elia A, Migon E, Shehabeldin A, Matysiak-Zablocki E, Lemmers B, Salmena L, Hakem A, Fish J, Kassam F, Squire J, Bruneau BG, Hande MP, Hakem R (2004) Lats2/Kpm is required for embryonic development, proliferation control and genomic integrity. *EMBO J 23*, 3677-88.

Moyer JH, Lee-Tischler MJ, Kwon HY, Schrick JJ, Avner ED, Sweeney WE, Godfrey VL, Cacheiro NL, Wilkinson JE, Woychik RP (1994) Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. *Science 264*, 1329-33.

Murcia NS, Richards WG, Yoder BK, Mucenski ML, Dunlap JR, Woychik RP (2000) The Oak Ridge Polycystic Kidney (orpk) disease gene is required for left-right axis determination. *Development 127*, 2347-55.

Paces-Fessy M, Boucher D, Petit E, Paute-Briand S, Blanchet-Tournier MF (2004) The negative regulator of Gli, Suppressor of fused (Sufu), interacts with SAP18, Galectin3 and other nuclear proteins. *Biochem J 378(Pt 2)*, 353-62.

Pastorcic M, Das HK (2007) Analysis of transcriptional modulation of the presenilin 1 gene promoter by ZNF237, a candidate binding partner of the Ets transcription factor ERM. *Brain Res 1128(1)*, 21-32.

Pazour GJ, Dickert BL, Vucica Y, Seeley ES, Rosenbaum JL, Witman GB, Cole DG (2000) *Chlamydomonas IFT88* and its mouse homologue, polycystic kidney disease gene *Tg737*, are required for assembly of cilia and flagella. *J Cell Biol 151*, 709-18.

Peng B, Sutherland KD, Sum EY, Olayioye M, Wittlin S, Tang TK, Lindeman GJ, Visvader JE (2002) CPAP is a novel stat5-interacting cofactor that augments stat5-mediated transcriptional activity. *Mol Endocrinol 16(9)*, 2019-33.

Porter ME, Sale WS (2000) The 9 + 2 axoneme anchors multiple inner arm dyneins and a network of kinases and phosphatases that control motility. *J Cell Biol 151*, F37-42.

Powzaniuk M, McElwee-Witmer S, Vogel RL, Hayami T, Rutledge SJ, Chen F, Harada S, Schmidt A, Rodan GA, Freedman LP, Bai C (2004) The LATS2/KPM tumor suppressor is a negative regulator of the androgen receptor. *Mol Endocrinol 18(8)*, 2011-23.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biology 132,* 365-86.

Silver LM (1995) Mouse Genetics: Concepts and Applications. Oxford, England: Oxford University Press.

Taulman PD, Haycraft CJ, Balkovetz DF, Yoder BK (2001) Polaris, a protein involved in left-right axis patterning,

localizes to basal bodies and cilia. *Mol Biol Cell 12*, 589-99.

Tucker KL, Meyer M, Barde YA (2001) Neurotrophins are required for nerve growth during development. *Nat Neurosci 4*, 29-37.

Wang S, Luo Y, Wilson PD, Witman GB, Zhou J (2004) The autosomal recessive polycystic kidney disease protein is localized to primary cilia, with concentration in the basal body area. *J. Am. Soc. Nephrol. 15*,592 -602

White AC, Lavine KJ, Ornitz DM (2007) FGF9 and SHH regulate mesenchymal Vegfa expression and development of the pulmonary capillary network. *Development 134(20)*, 3743–52.

Willaredt MA, Hasenpusch-Theil K, Gardner HAR, Kitanovic I, Hirschfeld-Warneken VC, Gojak CP, Gorgas K, Bradford CL, Spatz J, Wölfl S, Theil T, Tucker KL (2008) A crucial role for primary cilia in cortical morphogenesis. *J Neurosci 28*, 12887-900.

Yabuta N, Okada N, Ito A, Hosomi T, Nishihara S, Sasayama Y, Fujimori A, Okuzaki D, Zhao H, Ikawa M, Okabe M, Nojima H (2007) Lats2 is an essential mitotic regulator required for the coordination of cell division. *Biol Chem 282*, 19259-71.

Zhang Q, Davenport JR, Croyle MJ, Haycraft CJ, Yoder BK (2005) Disruption of IFT results in both exocrine and endocrine abnormalities in the pancreas of Tg737(orpk) mutant mice. *Lab Invest 85*, 45-64.